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
“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]
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

These ads allow criteo.com to link your visits between sites, even if you never click on the ads.
Concerns About Privacy (2010 – 2011)

Letting Down Our Guard With Web Privacy

By SOMINI SENGUPTA  MARCH 30, 2013
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).

![Diagram showing first-party and third-party cookies]

- 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.

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?

- The onion.com
- tracker.com
How does Third Party Web Tracking Work?

logo

ad

tracker.com

theonion.com
How does Third Party Web Tracking Work?
How does Third Party Web Tracking Work?

![Diagram showing how third-party web tracking works](image)
How does Third Party Web Tracking Work?

Set this cookie: id=789

Browsing profile for user 789:
theonion.com
How does Third Party Web Tracking Work?

Set this cookie: id=789

Browsing profile for user 789: theonion.com
How does Third Party Web Tracking Work?

Browsing profile for user 789: theonion.com

tracker.com

id=789

logo

Ad
How does Third Party Web Tracking Work?

Browsing profile for user 789:
theonion.com

tracker.com

cnn.com
How does Third Party Web Tracking Work?

Browsing profile for user 789: theonion.com

tracker.com

cnn.com
How does Third Party Web Tracking Work?

Browsing profile for user 789: theonion.com
How does Third Party Web Tracking Work?

Browsing profile for user 789: theonion.com

My cookie is: id=789

tracker.com

id=789

The onion.com

CNN

Ad

Logo

Tracker.com

id=789

Tracker.com

Ad

Logo
How does Third Party Web Tracking Work?

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

My cookie is: id=789

tracker.com

id=789
How does Third Party Web Tracking Work?

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

My cookie is: id=789
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&bustys=2710 HTTP/1.1\r\nHost: pixel.quantserve.com\r\nConnection: keep-alive\r\nAccept: image/webp, */*; q=0.8\r\nUser-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36\nReferer: http://www.theonion.com/\r\nAccept-Encoding: gzip, deflate, sdch\r\nAccept-Language: en-US, en; q=0.8\r\nCookie: mc=52a65386-f1de1-00ade-0b26e; d=ENkBpGHD4GYEA35MMIL74MkiyDs1A2MQII1Q
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!)
### Panopticlick

**How Unique — and Trackable — Is Your Browser?**

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

<table>
<thead>
<tr>
<th>Test</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is your browser blocking tracking ads?</td>
<td>✓ yes</td>
</tr>
<tr>
<td>Is your browser blocking invisible trackers?</td>
<td>✓ yes</td>
</tr>
<tr>
<td>Does your browser unblock 3rd parties that promise to honor Do Not Track?</td>
<td>✓ yes</td>
</tr>
<tr>
<td>Does your browser protect from <strong>fingerprinting</strong>?</td>
<td>⌚️ no</td>
</tr>
</tbody>
</table>

*your browser has a unique fingerprint*
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
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
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

- 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.
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
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 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.

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)

Top 20 Cross-Site Trackers on Top 500 Domains

<table>
<thead>
<tr>
<th>Tracker Prevalence (# Domains)</th>
<th>Cross-Site (Personal)</th>
<th>Cross-Site (Anonymous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>facebook.com</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td>google.com</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>quantserve.com</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>twitter.com</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>atdmt.com</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>imrworldwide.com</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>revsci.net</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>advertising.com</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>addthis.com</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>adxns.com</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>serving-sys.com</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>youtube.com</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>addthiscdn.com</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>bluekai.com</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>mediaplex.com</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>207.net</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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?

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

**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*
LocalStorage and Flash Cookies

• Surprisingly little use of these mechanisms!
• Of 524 trackers on Alexa Top 500:
  – Only 5 set unique identifiers inLocalStorage
  – 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

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

Top 20 Cross-Site Trackers on Top 500 Domains

Tracker Prevalence (# Domains)

- Cross-Site (Personal)
- Cross-Site (Anonymous)

<table>
<thead>
<tr>
<th>Tracker</th>
<th>Cross-Site (Personal)</th>
<th>Cross-Site (Anonymous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>189</td>
<td>149</td>
</tr>
<tr>
<td>facebook.com</td>
<td>154</td>
<td>105</td>
</tr>
<tr>
<td>google.com</td>
<td>110</td>
<td>60</td>
</tr>
<tr>
<td>quantserv.com</td>
<td>93</td>
<td>45</td>
</tr>
<tr>
<td>twitter.com</td>
<td>81</td>
<td>44</td>
</tr>
<tr>
<td>adtmt.com</td>
<td>40</td>
<td>34</td>
</tr>
<tr>
<td>imrworldwide.com</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>revsci.net</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>advertising.com</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>addthis.com</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>adnks.com</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>serving-sys.com</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>youtube.com</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>bluekai.com</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>mediaplex.com</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>207.net</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>
What 3\textsuperscript{rd} 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?

Often rely on blacklists, which may be incomplete.

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

[Privacy Badger logo]

https://www.eff.org/privacybadger