CSE 484: Computer Security and Privacy

Privacy + Tracking

Spring 2023

David Kohlbrenner
dkohlbre@cs

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

• Lab 2 is due next week
  • Remember we have a lot of resources/recordings on lab2 stuff!
• Lab 3 will go out shortly(?) after Lab 2 is due
Lab 2 debugging notes

• Step 1: log every page visit in your php script
  • I suggest logging more than the cookie param, maybe also a timestamp/etc
  • This will help you identify if the bot ever hits your page

• Step 2: verify that you can ‘xss yourself’
  • If you visit the error page, does your capture script see a hit?

• Step 3: debug the error page
  • The error page is exactly what the bot is going to visit, so you can debug here
  • Look at the final source of the page (view-source)
  • Check for errors in the javascript console
  • Look at the requests being made in dev-tools/etc
Privacy and web tracking
A topic in flux

• Tracking via cookies

• Tracking via other methods

• Fingerprinting
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

These ads allow *criteo.com* to link your visits between sites, *even if you never click on the ads.*
2022 Marketing Technology Landscape

visit martechmap.com to search, sort & filter
6,521% growth 2011 to 2022

Concerns About Privacy

The New York Times

‘Do Not Track’ Privacy Bill Appears in Congress
By TANZINA VEGA

May 6, 2011, 5:01 pm | 3 Comments

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.

By JENNIFER VALENTINO-DEVRIES,
JEREMY SINGER-VINE and ASHKAN SOLTANI
December 24, 2012
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](image)
Anonymous Tracking

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

criteo.com

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
Host: pixel.quantserve.com
Connection: keep-alive
Accept: image/webp, */*
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36
Referer: http://www.theonion.com
Accept-Encoding: gzip, deflate, sdch
Cookie: mc=52a65386-f1de1-00ade-0b26e; d=ENkBRgGH4GYEA35MMIL74MKiyDs1A2MQI1Q
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)
Other Trackers?

“Personal” Trackers
Personal Tracking

- Tracking is **not anonymous** (linked to accounts).
- Users **directly visit tracker’s site** ➔ evades some defenses.
How prevalent is tracking? (2011)

524 unique trackers on Alexa top 500 websites (homepages)

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.
Who/what are the top trackers?  

(2011)

### Top 20 Cross-Site Trackers on Top 500 Domains

- doubleclick.net
- google.com
- quantserve.com
- twitter.com
- admn.com
- revsci.net
- advertising.com
- adthrive.com
- invitemedia.com
- serving-sys.com
- youtube.com
- bluet Kai.com
- mediaplex.com
- 207.net

Tracker Prevalence (# Domains)

<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>google.com</td>
<td>154</td>
<td>105</td>
</tr>
<tr>
<td>quantserve.com</td>
<td>109</td>
<td>93</td>
</tr>
<tr>
<td>twitter.com</td>
<td>81</td>
<td>45</td>
</tr>
<tr>
<td>admn.com</td>
<td>60</td>
<td>44</td>
</tr>
<tr>
<td>revsci.net</td>
<td>40</td>
<td>34</td>
</tr>
<tr>
<td>advertising.com</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>adthrive.com</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>invitemedia.com</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>serving-sys.com</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>youtube.com</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>bluet Kai.com</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

5/10/2023

CSE 484 - Spring 2023
How has this changed over time?

• The web has existed for a while now...
  - What about tracking before 2011?
  - What about tracking before 2009?

• Solution: time travel!
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, more per site, more coverage

![Rise And Fall of Historical Champion Trackers](chart.png)
Defenses to Reduce Tracking

• Do Not Track?

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 doesn’t protect against network attackers fully.
Defenses to Reduce Tracking

• Do Not Track proposal?
• Private browsing mode?
• Third-party cookie blocking?
3rd party cookies

• Safari and FF (mostly) now block 3rd party cookies
  • https://webkit.org/blog/10218/full-third-party-cookie-blocking-and-more/

• Chrome...
  “By undermining the business model of many ad-supported websites, blunt approaches to cookies encourage the use of opaque techniques such as fingerprinting (an invasive workaround to replace cookies), which can actually reduce user privacy and control. We believe that we as a community can, and must, do better.”

Aug 2022: Remove 3rd party cookies by 2024
How should Google respond?

• Canvas!

• Pretend someone fired all the ad/chrome execs and hired your group instead

• Safari and Firefox have removed ad’s ability to track users via 3rd party cookies, and Google has committed to the same in Chrome by 2024.

• How should google respond to 3rd party cookies being removed?
  • Think about the technical solutions, policy solutions, and even business model solutions available to you!
Cookie Ghostwriting

• flickr.com
  • `<script src=siftscience.com/s.js /></script>`

• S.js runs
  • `Fp = fingerprintjs2`
  • `Setcookie(fp)`
  • `Hexagon-analytics.com/cookiereciever?cookie=fp`  

• Every time you load flickr.com what happens?

Journey to the Center of the Cookie Ecosystem: Unraveling Actors’ Roles and Relationship
Cookie Ghostwriting

• flickr.com patreon.com
  • <script src=siftscience.com/s.js />

• S.js runs
  • Fp = fingerprintjs2
  • Setcookie(fp)
  • Hexagon-analytics.com/cookiereceiver?cookie=fp

• Every time you load flickr.com patreon.com what happens?

Journey to the Center of the Cookie Ecosystem: Unraveling Actors’ Roles and Relationship
Fingerprinting is out there

• Better than a ‘voluntary’ cookie: involuntary, unchangeable id!
  • “Fingerprint”

• Idea: Measure ‘behavior’ of browser
  • Smash into unique ID
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!)
HTML5 Canvas Fingerprinting - Text

Figure 7: Difference maps for a group on text_arial

Mowery and Shacham, 2012
Figure 10: Original render and difference maps for Group 24

Mowery and Shacham, 2012
And its out there!

Figure 4: Different images printed to canvas by fingerprinting scripts. Note that the phrase “Cwm fjordbank glyphs vex tail quiz” in the top image is a perfect pangram, that is, it contains all the letters of the English alphabet only once to maximize diversity of the outcomes with the shortest possible string.
One in 145,235 browsers have the same fingerprint.
Fingerprinting as a security measure

• Blocking bots (e.g. reCAPTCHA)

• Validating users over-time
How should we view tracking and fingerprinting efforts?
“Privacy preserving” personalized ads aka FLoC

- [https://github.com/WICG/turtledove](https://github.com/WICG/turtledove)
  - The browser, not the advertiser, holds the information about what the advertiser thinks a person is interested in.
  - Advertisers can serve ads based on an interest, but cannot combine that interest with other information about the person — in particular, with who they are or what page they are visiting.
  - Web sites the person visits, and the ad networks those sites use, cannot learn about their visitors' ad interests.
“Privacy preserving” personalized ads aka Topics

- https://github.com/patcg-individual-drafts/topics
  - The browser, not the advertiser, holds the information about what the advertiser thinks a person is interested in.
  - Advertisers can serve ads based on an interest, but cannot combine that interest with other information about the person — in particular, with who they are or what page they are visiting.
  - Web sites the person visits, and the ad networks those sites use, cannot learn about their visitors' ad interests.

https://support.google.com/google-ads/answer/11899856?hl=en
“Privacy preserving” personalized ads aka FLEDGE Protected Audience API

- FLEDGE has been renamed to Protected Audience API.
- First Experiment (FLEDGE)
  - This document describes an early prototype for ads serving in the TURTLEDOVE family, appropriate for experimentation before a fully-featured system is ready. It would be the First "Locally-Executed Decision over Groups" Experiment.
  - This first experiment is currently a Chrome Origin Trial. The goal is for us to gain implementer experience, and for the ads ecosystem to evaluate its usability, as soon as it is feasible to do so. We need a robust API to take flight before the removal of third-party cookies shown on Chrome’s Privacy Sandbox timeline.
Privacy is far more than web tracking

• We’ve only started talking about it, in only 1 context.