The Invisible Trail: Third-Party Tracking on the Web

Franziska Roesner
Assistant Professor
Computer Science & Engineering
University of Washington
The Invisible Trail: Third-Party Tracking on the Web

Franziska Roesner + many collaborators!

Assistant Professor
Computer Science & Engineering
University of Washington
New technologies bring new benefits...

... but also new risks.
Security & Privacy Research

**Goal:** Improve security & privacy of technologies.

**Security mindset:** Challenge assumptions, think like an attacker.

1. Study existing technologies: attack and measure.
2. Design and build defenses and new technologies.
S&P Challenges Arise Everywhere

Today’s talk: web privacy

Who tracks you as you browse, and how?
Outline

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

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

The New York Times

‘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.

By JENNIFER VALENTINO-DEVRIES, JEREMY SINGER-VINE and ASHKAN SOLTANI

December 24, 2012
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 Approach

ANALYZE
(1) Reverse-engineer trackers’ methods.
(2) Develop tracking taxonomy.

MEASURE
(3) Build automated detection tool.
(4) Measure prevalence in the wild.
(5) Evaluate existing defenses.

BUILD
(6) Develop new defenses.
Web 101: Cookies

Websites store info in cookies in the browser.
– Only accessible to the site that set them.
– Automatically included with web requests.
Web 101: Iframes

Iframes allow one website to include another:

```html
<iframe src="www.washington.edu"> </iframe>
```

“first party”

“third party”
Web 101: First and Third Parties

```
<iframe src="http://www.foo.com">
</iframe>
```

www.bar.com

www.foo.com
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\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=ENkBRgGHG4YEA35MMIL74MKiyDs1A2MQIIQ
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.
Outline

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

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.

TrackingObserver: tracking detection platform http://trackingobserver.cs.washington.edu
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)

Tracking is increasing!

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

- 2011: 383
- 2013: 409
- 2015: 512
Who/what are the top trackers?

Top 20 Cross-Site Trackers on Top 500 Domains

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

("Vanilla" and others)
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%)
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 3\textsuperscript{rd} Party Cookie Blocking

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

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

How to get a cookie set?  
One way: be a first party.  
Facebook, Google, Twitter, etc.
What 3rd Party Cookie Blocking Misses

Top 20 Cross-Site Trackers on Top 500 Domains

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

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>154</td>
</tr>
<tr>
<td>facebook.com</td>
<td>149</td>
<td>105</td>
</tr>
<tr>
<td>google.com</td>
<td>109</td>
<td>93</td>
</tr>
<tr>
<td>quantserve.com</td>
<td>93</td>
<td>81</td>
</tr>
<tr>
<td>twitter.com</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>atdmt.com</td>
<td>44</td>
<td>40</td>
</tr>
<tr>
<td>imrworldwide.com</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>revsci.net</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>advertising.com</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>addthis.com</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>adnks.com</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>serving-sys.com</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>youtube.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>addthiscdn.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bluekai.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mediaplex.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>207.net</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 proposal?
- Private browsing mode?
- Third-party cookie blocking?
- Browser add-ons?

None are perfect, so use a combination:

- Send a "Do Not Track" request with your browsing traffic
- Block third-party cookies and site data
Recommended Browser Add-ons

Privacy Badger (EFF) [https://www.eff.org/privacybadger](https://www.eff.org/privacybadger)

[Privacy Badger](https://www.eff.org/privacybadger)

[Lightbeam for Firefox](https://www.mozilla.org/en-US/lightbeam/)

[Ghostery](https://www.ghostery.com/)
Summary

- Web tracking is complicated and ubiquitous.
- We systematically developed a tracking taxonomy and performed an extensive measurement study.
- Understanding the tracking ecosystem helps us design new tools and defenses.

Thanks to my collaborators! Yoshi Kohno, Adam Lerner, Chris Rovillos, Alisha Saxena, Anna Kornfeld Simpson, David Wetherall
Research Overview: Improving Security & Privacy

Analyze existing systems. e.g.: web tracking, automobiles, QR codes.

Understand mental models. e.g.: smartphone permissions, social media, journalists.

Build new systems. e.g.: web, OS, smartphones, user interface toolkits.

Anticipate future technologies. e.g.: telerobotics, wearables, augmented reality, IoT.

www.franziroesner.com
franzi@cs.washington.edu