CSE 484 / CSE M 584: Computer Security and Privacy

Third-Party Tracking on the Web

Spring 2016

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

Thanks to Dan Boneh, Dieter Gollmann, Dan Halperin, Yoshi Kohno, John Manferdelli, John Mitchell, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials ...
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.
Concerns About Privacy (2010 – 2011)

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.
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).
Anonymous Tracking

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

cookie: id=789

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&d=2710 HTTP/1.1
Host: pixel.quantserve.com
Connection: keep-alive
Accept: image/webp,*/*;q=0.8
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
Accept-Language: en-US,en;q=0.8
Cookie: mc=52a65386-f1de1-00ade-0b26e; d=ENkBRgGHG4YEA35MMIL74MKiyDs1A2MQITIQ
```
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!)
Your browser fingerprint appears to be unique among the 3,435,834 tested so far.
Panopticlick Example

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
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
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.
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)

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?

- Idea: Use AOL search logs to create 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%)

---

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 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?

www.bar.com
www.foo.com
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

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

Tracker Prevalence (# Domains)

- doubleclick.net
- facebook.com
- googlesyndication.com
- quantserve.com
- twitter.com
- atdmt.com
- yieldmanager.com
- imnworldwide.com
- revsci.net
- advertising.com
- addthis.com
- adxhs.com
- invitemedia.com
- serving-sys.com
- youtube.com
- bluekai.com
- mediaplex.com
- 207.net
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?

Often rely on blacklists, which may be incomplete.

Recommended

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

https://www.eff.org/privacybadger