The Web Never Forgets: Persistent Tracking Mechanisms in the Wild

Steven Englehardt

Joint work with:
Güneş Acar, Marc Juarez, Christian Eubank, Arvind Narayanan, Claudia Diaz

Expanded version of CCS 2014 Talk
Background
Advertising

- $42.8 billion revenues in 2013 (IAB)
- Relies on tracking

Source: IEEE Spectrum, Photo: Dan Saelinger; Prop Stylist: Dominique Baynes
Source: Mayer & Mitchell; Third-Party Web Tracking: Policy and Technology (and many COS432 lectures)
Tracking with Cookies

Primary tracking mechanism - built into requests

Browser

Request: A.com

Response

Set-Cookie: 12345

Request: A.com

Cookie: 12345

Request: A.com

Cookie: 12345

A.com
Tracking with Cookies

Not good enough?

- can be blocked
- don’t work well on mobile
- 20-40% of the users delete cookies

(ComScore, April 2010)
Advanced Tracking Mechanisms

Cookie Respawning  Fingerprinting  Cookie Syncing
Respawning (Evercookies)

Respawn cookies using obscure storage mechanisms
  ○ e.g. Flash cookies, HTML5 localStorage, ETags
Respawning - Data collection & Analysis

Procedure:

1. Crawl top 10,000 sites with OpenWPM
2. Clear HTTP cookies, keep other location (e.g. LSOs)
3. Recrawl top 10,000 sites (independent visits)
4. Check for regenerated cookies
   a. Check that IDs also in other location to rule out IDs regenerated by other means
# Respawnings

## Findings - Top 10 Using Flash Cookies

<table>
<thead>
<tr>
<th>Global rank</th>
<th>Site</th>
<th>CC</th>
<th>Respawning (Flash) domain</th>
<th>1st/3rd Party</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>sina.com.cn</td>
<td>CN</td>
<td>simg.sinajs.cn</td>
<td>3rd*</td>
</tr>
<tr>
<td>17</td>
<td>yandex.ru</td>
<td>RU</td>
<td>kiks.yandex.ru</td>
<td>1st</td>
</tr>
<tr>
<td>27</td>
<td>weibo.com</td>
<td>CN</td>
<td>simg.sinajs.cn</td>
<td>3rd*</td>
</tr>
<tr>
<td>41</td>
<td>hao123.com</td>
<td>CN</td>
<td>ar.hao123.com</td>
<td>1st</td>
</tr>
<tr>
<td>52</td>
<td>sohu.com</td>
<td>CN</td>
<td>tv.sohu.com</td>
<td>1st</td>
</tr>
<tr>
<td>64</td>
<td>ifeng.com</td>
<td>HK</td>
<td>y3.ifengimg.com</td>
<td>3rd*</td>
</tr>
<tr>
<td>69</td>
<td>youku.com</td>
<td>CN</td>
<td>irs01.net</td>
<td>3rd</td>
</tr>
<tr>
<td>178</td>
<td>56.com</td>
<td>CN</td>
<td>irs01.net</td>
<td>3rd</td>
</tr>
<tr>
<td>196</td>
<td>letv.com</td>
<td>CN</td>
<td>irs01.net</td>
<td>3rd</td>
</tr>
<tr>
<td>197</td>
<td>tudou.com</td>
<td>CN</td>
<td>irs01.net</td>
<td>3rd</td>
</tr>
</tbody>
</table>
Advanced Tracking Mechanisms

Cookie Respawning    Fingerprinting    Cookie Syncing
Panopticlick

How Unique — and Trackable — Is Your Browser?

Is your browser configuration rare or unique? If so, web sites may be able to track you, even if you limit or disable cookies.

Panopticlick tests your browser to see how unique it is based on the information it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only anonymous data will be collected by this site.
Your browser fingerprint **appears to be unique** among the 4,702,890 tested so far.

<table>
<thead>
<tr>
<th>Browser Characteristic</th>
<th>bits of identifying information</th>
<th>one in x browsers have this value</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Agent</td>
<td>9.3</td>
<td>631.77</td>
<td>Mozilla/5.0 (Windows NT 6.1; WOW64; rv:33.0) Gecko/20100101 Firefox/33.0</td>
</tr>
<tr>
<td>HTTP_ACCEPT Headers</td>
<td>3.96</td>
<td>15.52</td>
<td>text/html, <em>/</em> gzip, deflate en-US,en;q=0.5</td>
</tr>
</tbody>
</table>

Plugin 0: Adobe Acrobat; Adobe PDF Plug-In For Firefox and Netscape 10.1.12; nppdf32.dll; (Acrobat Portable Document Format; 

<table>
<thead>
<tr>
<th>System Fonts</th>
<th>22.17+</th>
<th>4702890</th>
</tr>
</thead>
</table>
| GuliMChe, Gungshu, GungshuChe, Harlow Solid Italic, Harrington, High Tower Text, Impact, Intermix Koton, IrisUPC, Iskoola Pota, JasmineUPC, Jokerman, Juice ITC, KaiTi, Kalinga, Kartika, Khmer UI, KodchianUPC, Kokila, Kristen ITC, Kunstler Script, Lao UI, Latha, Leelawadee, Levenim MT, LilyUPC, Lucida Bright, Lucida Calligraphy, Lucida Console, Lucida Fax, Lucida Handwriting, Lucida Sans Unicode, Mageto, Malgun Gothic, Mangal, Marlett, Matura MT Script Capitals, Meiyo, Meiyo UI, Microsoft Himalaya, Microsoft JhengHei, Microsoft New Tai Lue, Microsoft PhagsPa, Microsoft Sans Serif, Microsoft Tai Le, Microsoft Uighur, Microsoft YaHei, Microsoft Yi Baiti, Mincho, Mincho Bold, Myriad ITC, Myriad Pro, Noto Sans, Noto Sans Mono, 

|  |  |  |
|  |  |  |
Browser fingerprinting

FPDetective (Acar et al., CCS’13)

- detecting **font based fingerprinting**
- 16 previously unknown providers
- Flash based, 1.5% 10K sites
- JavaScript based, 0.04% of top 1M

Source: lalit.org
Canvas Fingerprinting

- Canvas and WebGL API
  - Allows sites to draw and render images from javascript
- Depends on:
  - OS
  - browser
  - fonts
  - font library
    - graphics card & driver
    - font rendering (ClearType)

Different than other fingerprinting vectors
Chrome, Window(XP, Vista, 7)

Diffs between renderings

Windows:

How quickly daft jumping zebras vex. (Also, pur
How quickly daft jumping zebras vex. (Also, pur
How quickly daft jumping zebras vex. (Also, pur
How quickly daft jumping zebras vex. (Also, pur
How quickly daft jumping zebras vex. (Also, pur

OS X:

How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu

Linux:

How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu

OSX:

How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4

Windows (XP; Vista; 7):

How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4

Windows 8:

How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4
How quickly daft jumping zebras vex. (4

Source: Mowery & Shacham; Pixel Perfect: Fingerprinting Canvas in HTML5 (W2SP’2012)
Canvas Fingerprinting

1. FillText()
   FillStyle()
   FillRect()

2. ToDataURL()
   data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAA...swAAACWCAAYAAABkW7XSAAAeq0leXg1d0...}

3. Hash()
Canvas Fingerprinting - Experimental Setup

● Crawler based on Selenium, mitmproxy and modified Firefox
  ○ Alexa top 100K (May 1-5, 2014)
  ○ Ran in parallel
  ○ Log canvas access functions (R/W)
  ○ Insert into a SQLite database

● Analysis & false positive removal
## Canvas Fingerprinting
### Findings - Alexa Top 100K

<table>
<thead>
<tr>
<th>Fingerprinting script</th>
<th>Number of including sites</th>
<th>Text drawn into the canvas</th>
</tr>
</thead>
<tbody>
<tr>
<td>ctl.addthis.com/static/r07/core130.js</td>
<td>5282</td>
<td>Cwm fjordbank glyphs vext quiz, ☪</td>
</tr>
<tr>
<td>i.ligatus.com/script/fingerprint.min.js</td>
<td>115</td>
<td><a href="http://valve.github.io">http://valve.github.io</a></td>
</tr>
<tr>
<td>src.kitcode.net/fp2.js</td>
<td>68</td>
<td><a href="http://valve.github.io">http://valve.github.io</a></td>
</tr>
<tr>
<td>admicro1.vcmmedia.vn/fingerprint/figp.js</td>
<td>31</td>
<td><a href="http://admicro.vn/">http://admicro.vn/</a></td>
</tr>
<tr>
<td>amazonaws.com/af-bdaz/bquery.js</td>
<td>26</td>
<td>Centillion</td>
</tr>
<tr>
<td>stat.ringier.cz/js/fingerprint.min.js</td>
<td>4</td>
<td><a href="http://valve.github.io">http://valve.github.io</a></td>
</tr>
<tr>
<td>cya2.net/js/STAT/89946.js</td>
<td>3</td>
<td>ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789+/</td>
</tr>
<tr>
<td>pof.com</td>
<td>2</td>
<td><a href="http://www.plentyoffish.com">http://www.plentyoffish.com</a></td>
</tr>
<tr>
<td>*.rackcdn.com/mongoose.fp.js</td>
<td>2</td>
<td><a href="http://api.gonorthleads.com">http://api.gonorthleads.com</a></td>
</tr>
<tr>
<td>9 others*</td>
<td>9</td>
<td>(Various)</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>5559</strong></td>
<td></td>
</tr>
</tbody>
</table>

(5542 unique)

---

1. Unique refers to the number of unique sites contributing to the total count, excluding duplicates.
Advanced Tracking Mechanisms

Cookie Respawning  Fingerprinting  Cookie Syncing
Cookie Syncing

GET: A.com
Cookie: {uid=12345}

302 Redirect: B.com?pid=A.com&uid=12345

GET: B.com?pid=A.com&uid=12345
Cookie: {uid=XYZ}

A.com

B.com

user XYZ is known as 12345 on A.com

C.com
Cookie Syncing

User

Bob

AppNexus
Bob == Steven
Bob == Peter
Bob == Chris

Bidders

BlueKai.com
Steven

DoubleClick.net
Peter

AdCentral.tk
Chris

Sync
Sync
Sync
Cookie Syncing - Experiments

● Top 3000 sites crawled while accumulating the state using OpenWPM

● Procedure:
  1. Detect identifying cookies and extract identifiers
  2. Search for identifiers in request / redirect URLs
  3. Mark all pairs of third-parties which share IDs
Cookie Syncing - Results

- Olejnik et al. finds **125 companies** syncing cookies (NDSS’14)
- We find **40%** of all tracking IDs are synced
- On average, **3.4 domains** see each ID
  - Turn.com’s ID cookie is seen by **43 domains**
Why do we care?

This can make it much more difficult to start with a fresh profile.
I want to start a new profile

Clear cookies (from the beginning of time)?
→ Oops, forgot to check “Cache”
→ Forced cached PNGs respawns cookies
I want to start a new profile

Clear all local browser state?
→ Your browser fingerprint didn’t change.
→ You visit site again
→ Site links your old history to new cookie using fingerprint
I want to start a new profile

Okay, so a few parties fingerprinted me or respawned cookies -- no big deal, most of my tracking history is gone? → respawned ID is cookie-synced to a large ad exchange → fingerprint linkage may still be shared through business relationships
I want to start a new profile

Well maybe I’ll clear all state and simultaneously change some fonts and browser plugins?
→ You *might* be okay.
→ We observed one instance of ID respawning through (what we assume is) passive, server-side fingerprinting.
Defenses
Opt-out

- opted-out on...
  - Network Advertising Initiative (NAI)
  - European Interactive Digital Advertising Alliance (EDAA)
  - AddThis' own Data Collection Opt-Out website
- no change for canvas fingerprinting and evercookies
- # of IDs involved in cookie sync: reduced by 30%
- # of parties involved in cookie sync: reduced by 5%
Cookie Blocking and DNT

- DNT reduces # of parties and # of IDs in cookie syncing by 3%
- Blocking all third-party cookies reduces by a factor of two
- Ad-blockers?
  - Ghostery cuts HTTP request traffic in half
Defenses - Canvas Fingerprinting

- no option to disable canvas
- Tor Browser returns an empty image
- no significant steps from browser manufacturers
  - pessimistic, “lost cause”
Conclusions
Move along, nothing to see here

- Significant media & public attention
- Top two providers (AddThis, Ligatus) stopped using canvas fingerprinting
  - calling it an “experiment”
- community made anti-canvas fingerprinting extensions
Takeaways

● third-party tracking isn't getting any better
● defense is hard, opt-out is not effective
● fingerprinting is becoming more common
● can automation help?
More info, code and the data:

- [https://webtap.princeton.edu/](https://webtap.princeton.edu/)

---

**The Web never forgets:** Persistent tracking mechanisms in the wild

The Web never forgets: Persistent tracking mechanisms in the wild is the first large-scale study of three advanced web tracking mechanisms - canvas fingerprinting, evercookies and use of "cookie syncing" in conjunction with evercookies.

[securehomes.esat.kuleuven.be](https://securehomes.esat.kuleuven.be)