Cookies That Give You Away
Evaluating the surveillance implications of web tracking

Steven Englehardt

Dillon Reisman, Christian Eubank, Peter Zimmerman, Arvind Narayanan, and Ed Felten

Jonathan Mayer
Motivation and Attack

Measurement Infrastructure

Results and Lessons
Motivation

- QUANTUMCOOKIE
- XKEYSCORE

- MUTANT BROTH
- QUANTUMINSERT
- BADASS
When the incoming emails stopped arriving, it seemed innocuous at first. But it would eventually become clear that this was no routine technical problem. Inside a row of gray office buildings in Brussels, a major hacking operation was underway. The targets were some of Europe’s largest telecommunications companies.
1. Identify Belgacom employees
   - NOC staff
   - In areas related to maintenance or security
2. Selectors to enable QUANTUM targeting
   - Use of LinkedIn noted
   - Use of Slashdot.org noted
3. MUTANT BROTH used to identify TDI/Selectors coming from identified range/proxy
4. QI capability enhanced to allow shots on LinkedIn
5. QI capability enhanced to allow ‘white listing’ when shooting on proxy
Source: Mayer & Mitchell; Third-Party Web Tracking: Policy and Technology
User

Adversary

Page visits:
- nytimes.com

Cookies:
- e3d0836d-4a3...
- 6076ff86-f7b...
- e7eff9af-8e4...
User

Adversary

Page visits:
- nytimes.com
- imdb.com

Cookies:
- e3d0836d-4a3...
- 6076ff86-f7b...
- e7eff9af-8e4...

Identity:
- Enguerrand Quarton
User

Cookies:
- e3d0836d-4a3...
- 6076ff86-f7b...
- e7eff9af-8e4...

Page visits:
- nytimes.com
- imdb.com

Identity:
- Enguerrand Quarton

Adversary

Youtube.com
Cookie: ‘e3d0836d-4a3…’
User

Page vists:
- nytimes.com
- imdb.com
- arstechnica.com

Cookies:
- e3d0836d-4a3...
- 6076ff86-f7b...
- e7eff9af-8e4...

Identity:
- Enguerrand Quarton

Adversary

Page vists:
- nytimes.com
- imdb.com
- arstechnica.com

Cookies:
- e3d0836d-4a3...
- 6076ff86-f7b...
- e7eff9af-8e4...

Identity:
- Enguerrand Quarton
**User**

- Turn.com
  Cookie: ‘d5j20xnd-23d...’

- Youtube.com
  Cookie: ‘e3d0836d-4a3...’

**Adversary**

- Page visits:
  - nytimes.com
  - imdb.com
  - arstechnica.com

- Cookies:
  - e3d0836d-4a3...
  - 6076ff86-f7b...
  - e7eff9af-8e4...
  - d5j20xnd-23d...

- Identity:
  - Enguerrand Quarton
Assumptions about the user

1. A user’s identifiers stay constant within a page visit

2. A user’s cookies stay constant across page visits
Adversary’s Abilities

1. Can disambiguate traffic between multiple users on the same IP within a page visit

2. Can utilize cookies to link requests across page visits

3. Adversary is passive
Motivation and Attack

Measurement Infrastructure

Results and Lessons
OpenWPM

- Web measurement framework
- Real browser automation
- Fully instrumented
- Support for user profiles
We use two models of browsing

● AOL Query Log
  ○ 25 AOL users with 50 - 100 queries
  ○ Issue queries on Google
  ○ Visit top 5 results

● Alexa Top 500
  ○ Random sample of 200 urls for each user
  ○ Localized to user’s browsing location
An adversary’s restrictions

● Policy restrictions
  ○ NSA: One-End-Foreign

● Geographic restrictions
  ○ Collection outside US at undersea cables
Summary of Measurements

1. Simulate users in 3 locations with 2 browsing models
   a. United States
   b. Ireland
   c. Japan
2. Simulate a restricted adversary
3. Test effectiveness of blocking
4. Average the size of the largest connected component for 25 OpenWPM measurement instances in each case
Motivation and Attack

Measurement Infrastructure

Results and Lessons
Average percentage of first-party sites linked

- Japan: 60%
- Ireland: 64%
- US Alexa: 73%
- US AOL: 62%

Average number of identity leakers

- Japan: 4
- Ireland: 6
- US Alexa: 10

9/25 users > 0
US Linking under “One-End-Foreign”

- **US**: 24% linked, 76% visible
- **Ireland**: 13% linked, 87% visible
- **Japan**: 21% linked, 79% visible

Percentage of Sites
What can users do?

- Enable DNT
- Block 3rd party cookies for non-1st party sites
- Block all 3rd party cookies
- Ghostery
- HTTPS Everywhere
Average of 25 simulated users from a US location browsing 200 sites under AOL model

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTPS Everywhere</td>
<td>46%</td>
</tr>
<tr>
<td>Ghostery</td>
<td>24%</td>
</tr>
<tr>
<td>Block all 3P</td>
<td>30%</td>
</tr>
<tr>
<td>Block some 3P</td>
<td>44%</td>
</tr>
<tr>
<td>DNT = 1</td>
<td>61%</td>
</tr>
<tr>
<td>US AOL (Baseline)</td>
<td>62%</td>
</tr>
</tbody>
</table>
Takeaways

1. Cookies should be treated like identities
2. HTTPS by default
3. Policy debates happen in a vacuum, tools like ours can help fix that

Infrastructure: https://github.com/citp/OpenWPM

Code and Data: https://github.com/englehardt/cookies-that-give-you-away

Icon Credits: User by Jose Campos, database by Stefan Parnarov, Hard Drive by Edward Boatman, programmer by Hadi Davodpour from the Noun Project