Third Party Web Tracking

Amit Datta

CMU
Fall 2016
Outline

• **History and Status Quo**
• How web tracking works and ensuing concerns
• What can we do?
  – User Choice Mechanisms
    • Opt-out cookies + AdChoices
    • Do Not Track
    • Blocking trackers
  – Government Regulation
  – Privacy Preserving Advertising
World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents. Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists, Policy, November's W3 news, Frequently Asked Questions.

What's out there?
Pointers to the world's online information, subjects, W3 servers, etc.

Help
on the browser you are using

Software Products
A list of W3 project components and their current state. (e.g. Line Mode, X11, Viola, NeXTStep, Servers, Tools, Mail robot, Library)

Technical
Details of protocols, formats, program internals etc

Bibliography
Paper documentation on W3 and references.

People
A list of some people involved in the project.

History
A summary of the history of the project.

How can I help?
If you would like to support the web..

Getting code
Getting the code by anonymous FTP, etc.

1992

Source: W3C
That didn’t last long.
Cookies
1994

JavaScript
1995
“A user agent should make every attempt to prevent the sharing of session information between hosts that are in different domains.”

-IETF RFC 2109

1997
Outline

• History and Status Quo
• **How web tracking works and ensuing concerns**
• What can we do?
  – User Choice Mechanisms
    • Opt-out cookies + AdChoices
    • Do Not Track
    • Blocking trackers
  – Government Regulation
  – Privacy Preserving Advertising
stateful tracking

tagging

stateless tracking

fingerprinting
this is blue

link
this is purple

link
HTTP Basics
HTTP Basics

1. User issues URL from a browser http://host:port/path/file
2. Browser sends a request message
   
   GET URL HTTP/1.1
   Host: host:port
   
3. Server maps the URL to a file or program under the document directory.
4. Server returns a response message
   
   HTTP/1.1 200 OK
   
5. Browser formats the response and displays

Client (Browser)

HTTP (Over TCP/IP)

Server (@ host:port)
Set-Cookie

• First interaction between user and server:

User request:

GET /index.html HTTP/1.1
Host: www.example.org

Server response: places cookie in user’s browser

HTTP/1.0 200 OK
Content-type: text/html
Set-Cookie: theme=light
Set-Cookie: sessionToken=abc123; Expires=Wed, 09 Jun 2021 10:18:14 GMT
Set-Cookie

• Subsequent interactions between browser and server:

  Browser sends cookie + other info to server

GET /spec.html HTTP/1.1
Host: www.example.org
Cookie: theme=light; sessionToken=abc123
Third Party Cookies

• There is **no intrinsic difference** between a first-party cookie and a third-party cookie.
• The distinction only exists within the context of a particular visit.

- **First party cookie**
  - website: news.com
  - cookie: news.com

- **Third party cookie**
  - website: news.com
  - cookie: advertising.com
<img>
<script>
<iframe>
GET http://advertising.com/ HTTP/1.1
Referer: http://news.com/
Cookie: id=12345
<table>
<thead>
<tr>
<th>User ID</th>
<th>Time</th>
<th>URL</th>
<th>Page Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>6/18/12</td>
<td><a href="http://foxnews.com/">http://foxnews.com/</a>...</td>
<td>Why Liberals Hate America</td>
</tr>
<tr>
<td></td>
<td>10:01am</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12345</td>
<td>6/18/12</td>
<td><a href="http://youtube.com/">http://youtube.com/</a>...</td>
<td>Squirrels Waterskiing?!</td>
</tr>
<tr>
<td></td>
<td>10:02am</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Scorecard Research, 81 sites (44%)
2. Google Analytics, 78 sites (42%)
3. Quantcast, 63 sites (34%)
4. Google Advertising, 62 sites (34%)
5. Facebook, 45 sites (24%)

(signed up and interacted with 185 sites)
Super Cookies

• Evercookie library

http://samy.pl/evercookie/
Stateless Tracking

User-Agent

HTTP ACCEPT Headers

browser plug-ins

MIME support

clock skew

installed fonts

cookies enabled?

browser add-ons

screen resolution

Sources: [Eckersley10], [Mayer09]
Browser Fingerprinting

• EFF’s Panopticlick

https://panopticlick.eff.org/

Peter Eckersley on browser fingerprinting:

https://panopticlick.eff.org/browser-uniqueness.pdf
Browsing History
Health Information
Financial Information
Shopping History

Hello, my name is...
“it’s all anonymous”
actually, it’s all pseudonymous

The Rock

Dwayne Johnson
Concerns

- sensitive, identifiable information
- lack of transparency
- lack of usable, effective controls
- inadequate market incentives
Cross Device Tracking

• Deterministic
  – Based on user logins
  – Facebook, Google

• Probabilistic
  – Based on location, browsing habits, etc.
  – Drawbridge, Tapad, BlueCava
Outline

• History and Status Quo
• How web tracking works and ensuing concerns
• **What can we do?**
  – User Choice Mechanisms
    • Opt-out cookies + AdChoices
    • Do Not Track
    • Blocking trackers
  – Government Regulation
  – Privacy Preserving Advertising
Opt-out Cookies

- Users opt out by installing opt-out cookies
- Anecdotal reports: usage < 1% in browsers
- Usability: low
THE NETWORK ADVERTISING INITIATIVE'S SELF-REGULATORY CODE OF CONDUCT

2008 NAI PRINCIPLES

≈

75 companies

http://www.networkadvertising.org/choices

102 companies
• Icon + text in display ads
• Explanation for why ad was served
• Landing page for user to set opt-out cookies
• Opt out of behavioral targeting not tracking
• Alexa US top 500 web sites:
  Icon in 9.9% of ads; icon + text in 5.1% of ads
DIGITAL ADVERTISING ALLIANCE

128 companies

http://www.aboutads.info/choices/
• not comprehensive
• not all third-party trackers offer
• vast majority do not participate
• requires updating*
• can accidentally clear*
Do Not Track

• Users specify via browser settings

• Implemented through an HTTP header field
  – DNT: 1

• W3C standardization effort ongoing
Privacy

Tracking

- Use Tracking Protection in Private Windows
- Change Block List

You can also manage your Do Not Track settings.

History

Firefox will:
- Remember where you have been
- Disable site tracking
- Keep websites from linking back to you

You may want:
- History
- Bookmarks
- Open tabs

Change preferences for search engine suggestions...

Do Not Track

- Always apply Do Not Track

Firefox will send a signal that you don't want to be tracked whenever Tracking Protection is on. Learn More

[Options: Cancel, OK]
On startup
- Open the New Tab page
- Continue where I left off.
- Open a specific page or set of pages. Set pages

Appearance
- Get themes
- Reset to default theme
- Show Home button
- Always show the bookmarks bar

Search
- Set which search engine is used when searching from the omnibox.
- Google
- Manage search engines...
- Enable Instant for faster searching (omnibox input may be logged)

Users
- You are currently the only Google Chrome user.
- Add new user...
- Delete this user
- Import bookmarks and settings...

Default browser
- The default browser is currently Google Chrome.

Show advanced settings...
Default browser
The default browser is currently Google Chrome.

Privacy
Google Chrome may use web services to improve your browsing experience. You may optionally disable these services. Learn more

- Use a web service to help resolve navigation errors
- Use a prediction service to help complete searches and URLs typed in the address bar
- Predict network actions to improve page load performance
- Enable phishing and malware protection
- Use a web service to help resolve spelling errors
- Automatically send usage statistics and crash reports to Google
- Send a 'Do Not Track' request with your browsing traffic

Passwords and forms
- Enable Autofill to fill out web forms in a single click. Manage Autofill settings
- Offer to save passwords I enter on the web. Manage saved passwords

Web content
Font size: Medium
Page zoom: 100%
- Pressing Tab on a webpage highlights links, as well as form fields
Settings

Default browser

The default browser is currently Google Chrome.

Privacy

Google Chrome may use web services to improve your browsing experience. You may optionally disable these services. Learn more

- Use a web service.
- Use a prediction system.
- Predict network traffic.
- Enable phishing protection.
- Use a service to display pages faster.
- Automatically fill out login forms.
- Send a ‘Do Not Track’ request.

Passwords and forms

- Enable Autofill to fill out web forms in a single click. Manage Autofill settings
- Offer to save passwords I enter on the web. Manage saved passwords

Web content

- Font size: Medium
- Page zoom: 100%
- Pressing Tab on a webpage highlights links, as well as form fields

Google Chrome may use web services to improve your browsing experience. You may optionally disable these services. Learn more

- Use a web service.
- Use a prediction system.
- Predict network traffic.
- Enable phishing protection.
- Use a service to display pages faster.
- Automatically fill out login forms.
- Send a ‘Do Not Track’ request.

Passwords and forms

- Enable Autofill to fill out web forms in a single click. Manage Autofill settings
- Offer to save passwords I enter on the web. Manage saved passwords

Web content

- Font size: Medium
- Page zoom: 100%
- Pressing Tab on a webpage highlights links, as well as form fields

Enable ‘Do Not Track’ means that a request will be included with your browsing traffic. Any effect depends on whether a website responds to the request, and how the request is interpreted. For example, some websites may respond to this request by showing you ads that aren’t based on other websites you’ve visited. Many websites will still collect and use your browsing data – for example to improve security, to provide content, services, ads and recommendations on their websites, and to generate reporting statistics. Learn more

OK Cancel
tens of millions of users
≈25 third parties honor DNT
Negotiations for 2+ years, no agreement.
Enforcing these signaling mechanisms

✓ observe suspicious behavior

• monitor ad distributions

Sources: [Jang10], [Guha10]
anti-tracking technology

blocking
Privacy Badger

https://www.eff.org/privacybadger
• not comprehensive
• requires updating
• breaks stuff
• requires user knowledge about providers
Evaluation of Blocking tools

Study with FourthParty of 11 blocking tools

• Crawled Alexa US top 500 sites
• For each tool, measured average across all trackers relative to baseline
  – Pages with HTTP request
  – Pages with HTTP Set-Cookie response
  – Cookies added - deleted
Evaluation of Blocking tools

• Findings
  – Most effective tool: community-maintained Fanboy’s lists
  – All top performing tools blocked third-party advertising
  – Block list from TRUSTe least effective + overrides other lists to allow tracking by several sizable third parties
Government Intervention

- FTC enforcement
- FTC proposal
- White House proposal
- Pending legislation

- Draft legislation

- ePrivacy Directive
- Article 29 Working Party opinions
- GDPR

- PIPEDA
Privacy Preserving Advertising

$\text{value}$

privacy

status quo

technical countermeasures, heavy-handed regulation
Privacy Preserving Advertising

$ value

privacy
Barriers to deployment

- Anonymizing proxy, download multiple ads
- Secure hardware, heavy crypto
- Download multiple ads
- Browser extension

Information leaked

- PrivAd: Nothing
- ObliviAd: IP address + User-Agent
- Adnostic: Ad clicks
- RePriv: Ad impressions (to CDN)
- CoP: Interest segments
- Cookies: Recent history, Entire history

Deployability

Privacy
What did we learn today?

• How web tracking works

• Concerns with tracking; how to address them

• FTC enforcement
• FTC proposal
• White House proposal
• Pending legislation
Third Party Web Tracking

Amit Datta

CMU
Fall 2016

Slide deck ack: Anupam Datta, Jonathan Mayer