Third-Party Web Tracking
Technology, Policy, and Politics

Jonathan Mayer

jmayer@stanford.edu
stanford.edu/~jmayer
Obligatory Disclaimers
This is not legal advice.
History and Status Quo
Tracking Methods
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preservation Advertising
1989
World Wide Web

The World Wide Web (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.
That didn’t last long.
“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
<img>
<script>
<iiframe>
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>12345</td>
<td>6/18/12</td>
<td><a href="http://youtube.com/">http://youtube.com/</a>...</td>
<td>Squirrels Waterskiing?!</td>
</tr>
</tbody>
</table>

...
Browsing History

Health Information

Financial Information

Shopping History

...
“it’s all anonymous”
actually, it’s all pseudonymous
• social network or other first party
• intentional leakage
• unintentional leakage
• security exploit
• deanonymization

Source: Narayanan 2011
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)
tracking | 'trakiNG|
noun
1 the collection of a user’s browsing history
• sensitive, identifiable information
• lack of transparency
• lack of usable, effective controls
• inadequate market incentives
History and Status Quo
Tracking Methods
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preserving Advertising
stateful tracking

tagging

stateless tracking

fingerprinting
HTTP cookies
HTTP authentication
HTTP ETags
content cache
IE userData

Flash Local Shared Objects
Silverlight Isolated Storage
TLS session ID & resume
browsing history
window.name

HTML5 protocol & content handlers

HTTP STS
HTML5 session/local/global/database storage
DNS cache

Source: [Aggrawal10]
this is blue

link
this is purple

link
User-Agent

HTTP ACCEPT Headers

browser plug-ins

MIME support

clock skew

installed fonts

cookies enabled?

browser add-ons

screen resolution

Sources: [Eckersley10], [Mayer09]
the technologies that make the web awesome are the very technologies that enable tracking
Lots of empirical questions.
Many Research Designs

1. build custom platform for experiment
2. run experiment
3. write paper
4. goto 1
FourthParty Design

1. build one platform
2. collect as much data as possible
3. run many experiments
4. write many papers
FourthParty Architecture

- easy to use
- shared data, historical data
- works with existing extensions (crawling and more)
- multiplatform
History and Status Quo
Tracking Methods
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preserving Advertising
opt out
THE NETWORK ADVERTISING INITIATIVE'S SELF-REGULATORY CODE OF CONDUCT

≈ 75 companies
• not comprehensive

• not all third-party trackers offer

• vast majority do not participate in NAI

• requires updating*

• can accidentally clear*
opt out = Do Not Target Ads

DECEPTIVE
Do Not Track
• universal
• no updating
• one click
• in the browser
HTTP header
DNT: 1
Tracking

- Tell web sites I do not want to be tracked

History

Firefox will:  Remember history

Firefox will remember your browsing, download, form and search history, and keep cookies from Web sites you visit.

You may want to clear your recent history, or remove individual cookies.

Location Bar

When using the location bar, suggest:  History
Cookies and other website data: Remove All Website Data...

Block cookies: Circle to choose: From third parties and advertisers, Always, Never

Limit website access to location services: Circle to choose: Prompt for each website once each day, Prompt for each website one time only, Deny without prompting

Website tracking: Circle to choose: Ask websites not to track me

Web search: Circle to choose: Prevent search engine from providing suggestions
Chrome

Settings

History
Extensions
Settings
Help

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...
Chrome Settings

Default browser
The default browser is currently Google Chrome.

Privacy

Content settings  Clear browsing data

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  Customize fonts
Page zoom: 100%

Pressing Tab on a webpage highlights links, as well as form fields
Do Not Track

Enabling ‘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
Express settings

You can customize the following settings or choose express settings. If you choose express settings, your PC will occasionally send info to Microsoft and will:

- Automatically install important and recommended updates.
- Help protect your PC from unsafe files and websites.
- Turn on Do Not Track in Internet Explorer.
- Help improve Microsoft software, services, and location services by sending us info.
- Check online for solutions to problems.
- Let apps give you personalized content based on your PC’s location, name, and account picture.
- Turn on sharing and connect to devices on this network.

Learn more about express settings

Privacy statement
tens of millions of users
≈25 third parties honor DNT
Advertising industry White House pledge: Implement by end of 2012
Negotiations for 2+ years, no agreement.
Enforcing a Signaling Mechanism
• observe suspicious behavior
• monitor ad distributions

Sources: [Jang10], [Guha10]
History and Status Quo
Tracking Methods
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preserving Advertising
old view: privacy $\approx$ security
anti-tracking technology
blocking
Source: Adblock Plus
• not comprehensive
• requires updating
• breaks stuff
• requires user knowledge about providers
“blocks many ... technologies that can track and profile you as you browse the Web ... updated weekly”
-Abine TPL

“completely removes all forms of tracking from the internet”
-EasyPrivacy ABP Subscription

“complete control over online tracking”
-PrivacyChoice TrackerBlock

“helps users get good ads, without compromising personal privacy”
-TRUSTe TPL
Average Decrease in Tracking

- Pages with an HTTP Request
- Pages with an HTTP Set-Cookie Response
- Cookies Added - Deleted
(we can’t reasonably expect the average user to sort this out)
new view: privacy ≠ security
relaxed third-party cookie blocking
study: measuring browsers
60k impressions
13 countries
Result 1: Advertising Opt Out Usage (U.S.)
Inspiron™ 15R
It helps me email, study, chat, flirt, and focus all at once

Select Dell systems are available with 2nd generation Intel® Core™ i7 processors

Shop Now >
Queried $\approx 175$ advertising status mechanisms
Cumulative Share of Browsers by Proportion of Opt Outs
Q: revealed preference or usability?
Result 2: Third-Party Cookie Blocking (U.S.)
Third-Party Cookie Blocking

- Chrome: Very low blocking
- Firefox: Low blocking

0.00% - 100.00% blocking ratio
Supports the usability hypothesis
Result 3: Do Not Track (U.S.)
Privacy

Tracking

- Tell web sites I do not want to be tracked

History

Firefox will: Remember history

Firefox will remember your browsing, download, form and search history, and keep cookies from Web sites you visit.

You may want to clear your recent history, or remove individual cookies.

Location Bar

When using the location bar, suggest: History
Technical Details

The DOM properties and methods below can be tested from Javascript, for example:

```javascript
var tpEnabled = window.external.msTrackingProtectionEnabled();
```

The DNT HTTP header requires a small bit of server code in order to be accessible from script:

```javascript
var dntValue = "<%= Request.ServerVariables["HTTP_DNT"] %>";
```

To express your preference not to be tracked in IE9:

1. Click here to add an empty Tracking Protection list
2. When prompted, click the "Add List" button:

To express your preference not to be tracked in IE10:

1. This preference is set by default for you in IE10
2. You can check by going to the Advanced tab under Internet Options and making sure the "Always send Do Not Track header" option is set.
Do Not Track

Firefox 5+

Internet Explorer 9+
As Firefox 5 gains traction, DNT elections hold steady around 6%.
Supports the usability hypothesis
History and Status Quo
Tracking Methods
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preserving Advertising
• FTC enforcement
• FTC proposal
• White House proposal
• Pending legislation

• Draft legislation

• ePrivacy Directive
• Article 29 Working Party opinions
• Pending legislation

• PIPEDA
History and Status Quo
Tracking Methods
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preserving Advertising
engineering conventions
information asymmetries
implementation and switching costs
diminished private utility
inability to internalize
competition barriers
regulatory responses?
History and Status Quo
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preserving Advertising
Questions?

jmayer@stanford.edu
stanford.edu/~jmayer
Tracking Not Required
Privacy-Preserving Third-Party Web Services
Tell me about yourself...
Do you come here often?
1. client-side storage
≈ 4 KB

> 5 MB
(but domains, paths)
2. client-side logic
JavaScript
3. server-side preprocessor
any web platform
example:
advertisement frequency capping
today: ID cookies
step 1: client-side storage
<table>
<thead>
<tr>
<th>Campaign</th>
<th>Count</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarCo</td>
<td>1</td>
<td>today</td>
</tr>
<tr>
<td>Toothpaste Unlimited</td>
<td>4</td>
<td>yesterday</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

“tell me everything”
(gets unique very quickly)
step 2: client-side secrecy
<table>
<thead>
<tr>
<th>Campaign</th>
<th>Count</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarCo</td>
<td>1</td>
<td>today</td>
</tr>
<tr>
<td>Toothpaste Unlimited</td>
<td>4</td>
<td>yesterday</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

“tell me about impressions”
step 3: server-side preprocessing
<table>
<thead>
<tr>
<th>Campaign</th>
<th>Count</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarCo</td>
<td>1</td>
<td>today</td>
</tr>
<tr>
<td>Toothpaste Unlimited</td>
<td>4</td>
<td>yesterday</td>
</tr>
</tbody>
</table>

“tell me about impressions for these ads”
step 4: client-side logic
<table>
<thead>
<tr>
<th>Campaign</th>
<th>Count</th>
<th>Capped</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarCo</td>
<td>1</td>
<td>yes</td>
</tr>
<tr>
<td>Toothpaste Unlimited</td>
<td>4</td>
<td>no</td>
</tr>
</tbody>
</table>
“tell me whether these ads are capped”
step 5: client-side logic
<table>
<thead>
<tr>
<th>Campaign</th>
<th>Count</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarCo</td>
<td>1</td>
<td>today</td>
</tr>
<tr>
<td>Toothpaste Unlimited</td>
<td>4</td>
<td>yesterday</td>
</tr>
</tbody>
</table>

= Toothpaste Unlimited
“tell me my most preferred uncapped ad”
state space = # of candidate campaigns
Website

- list of preference-ranked ads
- arbitrary data

| arbitrary computation |

Browser

- list of ads to display
• frequency capping
• frequency targeting
• behavioral targeting
• retargeting
example:
advertising conversion measurement
<table>
<thead>
<tr>
<th>Campaign</th>
<th>Origin</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarCo</td>
<td>news.com</td>
<td>this week</td>
</tr>
<tr>
<td>Toothpaste Unlimited</td>
<td>sports.com</td>
<td>last week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
social sharing widget personalization
URL: http://gothamnews.com/batman_strikes_again

Friend: Bruce

Photo:

Friend:
Alfred
Bruce

...