Online Tracking

Vitaly Shmatikov

Slides courtesy of Arvind Narayanan
Reading Assignment

◆ “Third-Party Web Tracking: Policy and Technology”
◆ “Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting”
It’s the Internet! Of course they know you’re a dog. They also know your favorite brand of pet food and the name of the cute poodle at the park that you have a crush on!
Third-Party Tracking

Third-party cookies:
- Disabled by default (Safari)
- Can be disabled by user (many browsers)
- Cannot be disabled (Android)
... but there are many other tracking technologies
Behavioral Targeting

- Publishers
  - The New York Times
  - okcupid
  - Google
- Ad network
- Advertisers
  - Coca-Cola
  - Ford
<table>
<thead>
<tr>
<th>Partial List of Ad Networks</th>
<th>Ad Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>24/7 Real Media</td>
<td>33Across</td>
</tr>
<tr>
<td>Adify</td>
<td>Acerno</td>
</tr>
<tr>
<td>Aggregate Knowledge</td>
<td>Acxiom Relevance-X</td>
</tr>
<tr>
<td>Blue Kai</td>
<td>AdJuggler</td>
</tr>
<tr>
<td>Casale Media</td>
<td>AdShuffle</td>
</tr>
<tr>
<td>Coremetrics</td>
<td>ADTECH (AOL)</td>
</tr>
<tr>
<td>eXelate</td>
<td>Bizo</td>
</tr>
<tr>
<td>Fox Audience Network</td>
<td>BTBuckets</td>
</tr>
<tr>
<td>Navegg</td>
<td>Burst Media</td>
</tr>
<tr>
<td>Microsoft</td>
<td>Criteo</td>
</tr>
<tr>
<td>NextAction</td>
<td>Collective Media</td>
</tr>
<tr>
<td>MindSet Media</td>
<td>Criteo</td>
</tr>
<tr>
<td>Outbrain</td>
<td>Effective Measure</td>
</tr>
<tr>
<td>richrelevance</td>
<td>FetchBack</td>
</tr>
<tr>
<td>Snobi</td>
<td>Flashtalking</td>
</tr>
<tr>
<td>Snoobi</td>
<td>FloodBack</td>
</tr>
<tr>
<td>Traffic Marketplace</td>
<td>Flashtalking</td>
</tr>
<tr>
<td>ValueClick Media</td>
<td>FloodBack</td>
</tr>
<tr>
<td>Vizu</td>
<td>FloodBack</td>
</tr>
<tr>
<td>Weborama</td>
<td>FloodBack</td>
</tr>
<tr>
<td>WebTrends</td>
<td>FloodBack</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>FloodBack</td>
</tr>
<tr>
<td>[x+1]</td>
<td>FloodBack</td>
</tr>
</tbody>
</table>
Tracking Is Pervasive

64 independent tracking mechanisms in an average top-50 website
Sticky Tracking

Subverting same origin policy
(publisher also runs an ad network)
ad.hi5.com = ad.yieldmanager.com

Flash cookies

Browser fingerprinting

History sniffing
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
Everything Has a Fingerprint
Fingerprinting Web Browsers

- User agent
- HTTP ACCEPT headers
- Browser plug-ins
- MIME support
- Clock skew
- Installed fonts
- Cookies enabled?
- Browser add-ons
- Screen resolution
Your browser fingerprint **appears to be unique** among the 3,435,834 tested so far.

You see how easily identifiable you might be as you click the web.

Only anonymous data will be collected by this site.

A paper reporting the statistical results of this experiment is now available: *How Unique Is Your Browser?*, Proceedings of the Privacy Enhancing Technologies Symposium (PETS 2010), Springer Lecture Notes in Computer Science.
Panopticlick Example

84% of browser fingerprints are unique
With Flash or Java, 94% are unique
“Don’t Worry, It’s All Anonymous”

◆ Is it?
◆ What’s the difference between
  “anonymous”
  “pseudonymous”
  “identified”
◆ Which technology changed data collection from anonymous to pseudonymous?
How Websites Get Your Identity

Third party is sometimes the site itself

Leakage of identifiers

GET http://ad.doubleclick.net/adj/...
Referer: http://submit.SPORTS.com/...?email=jdoe@email.com
Cookie: id=35c192bcfe0000b1...

Security bugs

Remember XSUH (cross-site URL hijacking)?

Third party buys your identity
Introduction

Syphilis is a bacterial infection that is usually passed on through having sex with someone who is infected. It can also be passed from an infected mother to her unborn child and, in rare cases, can be caught through injecting drugs.

It is extremely rare to catch syphilis through a blood transfusion in the UK as blood donors are carefully screened.

Three stages of disease

Stage 1 (primary syphilis). Symptoms of syphilis begin with a painless but highly infectious sore on the genitals or sometimes around the mouth. If somebody else comes into close contact with the sore, typically during sexual contact, they can also become infected. The sore lasts two to six weeks before disappearing.

Stage 2 (secondary syphilis). Secondary symptoms, such as a skin rash and sore throat, then develop. These symptoms may disappear within a few weeks, after which you experience a latent (hidden) phase with no symptoms, which can last for years. After this, syphilis can progress to its third, most dangerous stage.

Stage 3 (tertiary syphilis). At this stage, it can cause serious damage to the body.

The primary and secondary stages are when you are most infectious to other people. In the latent phase (and usually around two years after becoming infected), syphilis cannot be passed onto others but can still cause symptoms. See Symptoms of syphilis for more information on the
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**
Identity Sniffing

- All social networking sites allow users to join groups
- Users typically join multiple groups
  - Some of these groups are public
- Group-specific URLs are predictable
  - https://www.xing.com/net/[groupID]/forums+
- Intersection of group affiliations acts as a fingerprint
  - Can sometimes infer identity by computing the intersection of group membership lists
One-Click Fraud

Thank you for your patronage! You successfully registered for our premium online services, at an incredible price of 50,000 JPY. Please promptly send your payment by bank transfer to ABC Ltd at Ginko Bank, Account 1234567. Questions? Please contact us at 080-1234-1234.

Your IP address is 10.1.2.3, you run Firefox 3.5 over Windows XP, and you are connecting from Tokyo.

Failure to send your payment promptly will force us to mail you a postcard reminder to your home address. Customers refusing to pay will be prosecuted to the fullest extent of the law. Once again, thank you for your patronage!
One-Click Fraud

- Estimated costs to victims: USD 260 million / year

- What’s going on here?

- Why only Japan?
  - Cultural factors:
    - susceptibility to authoritative language
    - threat of public shaming

Credible because the website does have your real identity!
Instant Personalization

Creepy is the new normal
Do Not Track

Basics

HTTP header
- DNT: 1

Standardization

Browser support in FF4, IE9

Beginning to see adoption (AP, NAI)... or not

Privacy protections

No tracking across sites
- Who is the “third” party?
  Can’t be based on domain
  Example: amazonaws.com, ad.hi5.com ...

No intrusive tracking

Limits on regular log data

Exceptions for fraud prevention, etc.
“But the NAI code also recognizes that companies sometimes need to continue to collect data for operational reasons that are separate from ad targeting based on a user’s online behavior. For example, online advertising companies may need to gather data to prove to advertisers that an ad has been delivered and should be paid for; to limit the number of times a user sees the same ad; or to prevent fraud.”

Translation: we’re going to keep tracking you, but we’ll simply call it “operational reasons.”
Brave New World?

How are these identifiers different from third-party cookies?