Browser fingerprinting
(how did we get here)

Nick Nikiforakis
nick@cs.stonybrook.edu
"On the Internet, nobody knows you're a dog."
I need state!

• HTTP is a stateless protocol
  – The server does not know that two or requests originate from the same user

• No state -> No Personalization
  – No e-banking, e-shops, webmail, etc.

• Solution: Cookies!
<table>
<thead>
<tr>
<th>Domain</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>stonybrook.edu</td>
<td>username=nick;</td>
</tr>
<tr>
<td></td>
<td>Date=30/09/2014;</td>
</tr>
<tr>
<td>google.com</td>
<td>g1=190233213;</td>
</tr>
<tr>
<td></td>
<td>g2=afasfdioujewf;</td>
</tr>
<tr>
<td>slashdot.org</td>
<td>adhese_count=2;</td>
</tr>
<tr>
<td></td>
<td>bcn=e4f5d957-00;</td>
</tr>
</tbody>
</table>
Study Finds News Sites Fail to Aim Ads at Users

By TANZINA VEGA
Published: February 13, 2012

Web sites for newspapers, magazines and television stations might be hungry to make money with digital advertising, but you wouldn’t know it by the way some of them do business online.

A new study released Monday by the Pew Research Center Project for Excellence in Journalism looked at 22 news Web sites and more than 5,300 digital ads. It found that many of the sites had not attracted the same advertisers online as they did on other platforms.

In part, these sites were failing to attract online ads because they were not using technology that would customize ads based on their users’ online behavior. For example, a user searching for tickets to a Broadway show might see ads for that show.

The study, which looked at Web sites for 11 newspapers, four magazines and six television outlets, as well as two online-only sites, focused on premium digital ad placements on home pages or at the top of article pages, which have generally cost more to buy.

“One of the great challenges that faces the financial future of journalism is, how can you begin to charge more for digital advertising?” said Tom Rosenstiel, the director of the center. “The information is really there, but do you know who is looking at it? Do you know how they’re looking at it? Do you know they’re really interested? That’s the kind of information you need to make a big digital advertising business.”
A cookie’s life
3\textsuperscript{rd} Party Tracking

• “Suddenly” all sorts of websites that you’ve never heard about, can create a browsing profile of you and sell it to advertising companies
  – quantserve.com
  – scorecardresearch.com
  – addthis.com
Users reacted...

• 1/3 of users delete first & third-party cookies within a month after they’ve been setup
• Multiple extensions revealing hidden trackers
  – Ghostery
  – Lightbeam
• Private mode of browsers used to avoid traces of cookies from certain websites
Ghostery
Lightbeam
EU Cookie law

“Member States shall ensure that the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.”
“Member States shall ensure that the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.”
Care for a cookie or two?

Here at Bluehost, we want you to have the most relevant customer experience possible. That's why we use cookies – they help remember log-ins and optimize the content you see based on your interests and preferences.
Belgium?
Member States shall ensure that the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.
Money making
PROBLEM = PROBLEM
PROBLEM = Right?
However...

• What if you could track users without the need of cookies or any other stateful client-side identifier?
  – Hidden from users
  – Hard to avoid it / opt-out

Web-based device fingerprinting

• Eckersley showed in 2010 that certain attributes of your browsing environment can be used to accurately track you

• These attributes, when combined, created a quite unique fingerprint of your system?
  – How?
Properties fingerprinted by Panopticlick
Resulting fingerprints

- 94.2% of the users with Flash/Java could be uniquely identified
- Simple heuristic algorithms could track updates of the same browser
Other proposed ways

• Eckersley paved the way of stateless tracking through fingerprinting

• After Eckersley, other researchers proposed ways of fingerprinting browsers, based on:
  – Speed
  – Implementation coverage
  – Rendering of elements
They will know you by your speed...

- Mowery et al. (W2SP 2011) proposed the use of performance benchmarks to tell different JavaScript engines apart
  - Different JavaScript engine -> Different browser

- Collected performance signatures (39 tests) from approx. 1000 users
  - 98.2% correct browser family detection
  - Overall accuracy (versions included): 79.8%
What are we?! Browsers! Browsers! Browsers!

What do we want?! More speed! More speed! More speed!

And when do we want it?! Right now! Right now! Right now!

Browsers!
As well as your features...

- Mulazzani et al. (W2SP 2013) proposed the use of missing functionality in JavaScript engines
  - Different browsers, implement JavaScript standards, at a different rate

<table>
<thead>
<tr>
<th>Browser</th>
<th>Win 7</th>
<th>WinXP</th>
<th>Mac OS X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firefox 3.6.26</td>
<td>3955</td>
<td>3955</td>
<td>3955</td>
</tr>
<tr>
<td>Firefox 4</td>
<td>290</td>
<td>290</td>
<td>290</td>
</tr>
<tr>
<td>Firefox 5</td>
<td>264</td>
<td>264</td>
<td>264</td>
</tr>
<tr>
<td>Firefox 6</td>
<td>214</td>
<td>214</td>
<td>214</td>
</tr>
<tr>
<td>Firefox 7</td>
<td>190</td>
<td>190</td>
<td>190</td>
</tr>
</tbody>
</table>
As well as your artistic talent

- Mowery et al. (W2SP) proposed the use of the HTML5 canvas to detect browser-specific renderings of the same string
  - Write some text in canvas, read it out as an image
  - Different browsers/hardware combinations will create slightly different images

- [http://jsbin.com/ePAheCi/2/edit](http://jsbin.com/ePAheCi/2/edit)
Feds Are Suspects in New Malware That Attacks Tor Anonymity

BY KEVIN POULSEN 08.05.13  3:57 AM

Follow @kpoulsen
EU Cookie law

“Member States shall ensure that the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.”

Does it apply to fingerprinting?
What’s happening out there?

• In mid 2012, all we knew is that fingerprinting is possible and that a small number of companies offer it as a service

• Questions that begged answering:
  – How are they doing it?
  – Could they do more?
  – Who is using them?
  – How are users trying to hide?
    • Is it working?
Manual analysis of 3 fingerprinting companies

1. Find the domains that they use to serve their fingerprinting scripts
2. Find some websites that use them and extract the code
3. De-obfuscate and analyze
4. Compare and classify
Results

• After extracting all features, we created a taxonomy of all fingerprinted features, and compared each company to Panopticlick
• Collectively, Panopticlick was fully covered
Non-trivial extras

• Non-plugin font detection
  – Comparison of text’s width & height

• Native Fingerprinting plugins
  – Accessing highly-specific registry value

• Fingerprint delivery mechanisms

• Proxy detection
Font Detection through JavaScript

<table>
<thead>
<tr>
<th>String</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_DO_NOT_NEED_FLASH</td>
<td>500 x 84</td>
</tr>
<tr>
<td>I_DO_NOT_NEED_FLASH</td>
<td>520 x 84</td>
</tr>
<tr>
<td>I_DO_NOT_NEED_FLASH</td>
<td>580 x 87</td>
</tr>
<tr>
<td>I_DO_NOT_NEED_FLASH</td>
<td>399 x 82</td>
</tr>
</tbody>
</table>
Non-trivial extras

• Non-plugin font detection
  – Comparison of text’s width & height

• Native Fingerprinting plugins
  – Accessing highly-specific registry values

• Fingerprint delivery mechanisms

• Proxy detection
Proxy-detection
Demo

http://www.orbitz.com
Adoption

Dataset A

– Crawled top 10,000 sites, searching for inclusions from the 3 fingerprint providers

– 40 sites discovered
  • Porn & dating sites most prominent
    – Shared credentials & Sybil attacks
  • skype.com the highest ranking one
Adoption

Dataset B
– 3,804 domains from Wepawet
But wait... there’s more!

• Can we find unknown fingerprinting parties?
  – How do we separate a fingerprinting script from a generic analytics script?

• Fonts!
  – Separating feature between normal analytics and fingerprinting
  – Second most identifying feature according to Eckersley
FPDetective

• Fingerprinting-sensitive crawler
  – If fonts are touched, record site

• Detection of font snooping
  – JS-based font probing (Modified browser)
  – Flash-based font probing (decompilation of Flash)
Adoption (revisited)

Dataset A

– Crawled top 10,000 sites, searching for inclusions from the 3 fingerprint providers

– 40 sites discovered
  • Porn & dating sites most prominent
    – Shared credentials & Sybil attacks
  • skype.com the highest ranking one
Adoption (revisited)

Dataset A

- Crawled top 10,000 sites, searching for inclusions from the 3 fingerprint providers
- 40 sites discovered
  - Porn & dating sites most prominent
  - Shared credentials & Sybil attacks
  - skype.com the highest ranking one

- 145 fingerprinting sites in the top Alexa 10K
- DNT does not matter
Status

• Fingerprinting is out there
  – Quite a number of new techniques over Panopticlick

• Large and popular sites are using them

• Could they be doing more?
  – How do the browser internals relate to a browser’s identity?
DIY Fingerprinting

• We decided to try some fingerprinting of our own
• Focus on the two special JS objects that fingerprinters probe the most:
  – navigator
  – screen
• Perform a series of everyday operations and search for differences across browsers
  – Add properties
  – Remove properties
  – Modify properties
Novel methods discovered

• E.g., Natural ordering of properties can give away a browser family, and occasionally, a browser version

navigator.geolocation
navigator.onLine
navigator.cookieEnabled
navigator.vendorSub
navigator.vendor

navigator.appCodeName
navigator.appName
navigator.appVersion
navigator.language
navigator.mimeTypes

navigator.appCodeName
navigator.appName
navigator.appMinorVersion
navigator.cpuClass
navigator.platform
Other methods...

• Family-specific methods & properties
  – screen.mozBrightness
  – navigator.webkitStartActivity
  – screen.logicalXDPI

• Mutability of special objects

• Evolution of functionality

• Miscellaneous
Status

• Fingerprinting is out there
  – Quite a number of new techniques over Panopticlick

• Large and popular sites are using them

• There could be more fingerprinting done by the companies

• How could a user react?
Browser extensions

• Reviewed 11 different browser extensions that spoof a browser’s user-agent
  – 8 Firefox + 3 Chrome
  – More than 800,000 users

• Advice to use such extensions:
  – Previous research in web tracking
  – Underground hacking guides

• How do they stand-up against fingerprinting?
Worse than nothing...

• All of them had one or more of the following:
  – Incomplete coverage of the navigator object
  – Impossible configurations
  – Mismatch between UA header and UA property

• Iatrogenic problem:
  – When installing these, a user becomes more visible and more fingerprintable than before
Case Study

Stats
- 463,293 users
- 187 user reviews
- 4/5 starts
Worse than nothing...

- All of them had one or more of the following:
  - Incomplete coverage of the navigator object
  - Impossible configurations
  - Mismatch between UA header and UA property

- Iatrogenic problem:
  - When installing these, a user becomes more visible and more fingerprintable than before

**Fingerprintable Surface**

- Extension_A
- Extension_B
- Extension_C
Defenses (today)

• The more generic your system is, the better
  – The more exotic plugins and extensions you have installed, the more chances of being singled out

• Fingerprinters can be black-listed

• Disabling Flash and Java will definitely help
  – No explicit font collection

• Virtual machines? Browsers from a stick?
  – Depends on your balance between hassle and privacy
Conclusion

• Web tracking is so much more than cookies
• Fingerprinting is a real problem
• Browsers are so complex that it is really hard to make them seem identical
• Current browser extensions should not be used for privacy reasons
• Long term solutions will most-likely not be pure technical ones
  – Legislation required, like in stateful tracking
In the 1990’s

ON THE INTERNET, NOBODY KNOWS YOU’RE A DOG.

Now

OUR METADATA ANALYSIS INDICATES THAT HE IS DEFINITELY A BROWN LAB.

HE LIVES WITH A WHITE AND BLACK SPOTTED BEAGLE-MIX, AND I SUSPECT THEY ARE HUMPING.