The Web Privacy Problem is a Transparency Problem

How OpenWPM and the Transparency Census will bring transparency to the web.

Steven Englehardt
@s_englehardt

webtap.princeton.edu
Web tracking lacks transparency
Web tracking lacks transparency...but we are changing that
Web tracking lacks transparency

...but we are changing that

(and I’ll show you how we already have)
Transparency encourages best practices

Canvas
Fingerprinting
Introduced

May 2012
Transparency encourages best practices

May 2012

Canvas
Fingerprinting
Introduced

Add This

May 2014
Transparency encourages best practices

May 2012

Canvas Fingerprinting Introduced

AddThis

Ligatus

May 2014

Canvas Fingerprinting Measured
Transparency encourages best practices

May 2012

Canvas Fingerprinting Introduced

May 2014

Add This

News Coverage

July 21st 2014

Canvas Fingerprinting Measured
Transparency encourages best practices
Canvas Fingerprinting was a known technique for 2 years.

In just 2 months following our measurement work the largest users of canvas fingerprinting stopped.

Why?
Our measurement work removed information asymmetry between trackers and the rest of the web.
Our measurement work removed **information asymmetry** between trackers and the rest of the web.
Our measurement work removed information asymmetry between trackers and the rest of the web.
Our measurement work removed information asymmetry between trackers and the rest of the web.
Our measurement work removed information asymmetry between trackers and the rest of the web.
Our measurement work removed information asymmetry between trackers and the rest of the web.
Our measurement work removed information asymmetry between trackers and the rest of the web.
Our measurement work removed **information asymmetry** between trackers and the rest of the web.
Our measurement work removed **information asymmetry** between trackers and the rest of the web.
Information asymmetry not just between trackers and users.

“YouPorn contacted us to say…’[the website was] completely unaware that AddThis contained a tracking software…”"
Transparency is effective at returning control to users and publishers
Automated, large-scale measurements can provide this transparency
We’re doing three things to help:

1. Developing OpenWPM

2. Running monthly, 1 million site measurements

3. Building an analysis layer on top of the data
OpenWPM supports browsing with persistent state

- Browser can keep profile through crashes and freezes
  - Cookie setting over a session
  - Cookie synchronization (id sharing)
  - Zombie Cookies
OpenWPM uses a real browser

- Extensions
  - AdBlock Plus, Ghostery, ...
- Privacy Features
  - Block third-party cookies, FF tracking protection, ...
- Support for new web technologies
  - WebRTC, Audio, Video, WebGL
OpenWPM is already used by at least 7 research groups

● At Princeton
  ○ 4 published studies and several ongoing

● Ongoing Research
  ○ Columbia University

● In published studies:
  ○ The Web Privacy Census (UC Berkeley / Berkeley Law)
  ○ Variations in Tracking in Relation to Geographic Location (CMU / RAND)
  ○ Forthcoming WWW’16 study by Nick Nikiforakis (Stony Brook)

● By journalists

● By regulators
The Web Transparency Census

Monthly
1 Million Site Crawl
The Web Transparency Census

Monthly
1 Million Site Crawl

Collecting:
- Javascript Calls
- All javascript files
- HTTP Requests and Responses
- Storage (cookies, Flash, etc)
Supporting a variety of measurements

1. Effectiveness of Privacy Tools
   ● Ghostery
   ● AdBlock Plus
   ● HTTPS Everywhere
Supporting a variety of measurements

1. Effectiveness of Privacy Tools
   - Ghostery
   - AdBlock Plus
   - HTTPS Everywhere

2. Effectiveness Browser Protections
   - DNT
   - Third-party cookie Blocking
   - Firefox Tracking Protection
Supporting a variety of measurements

1. Effectiveness of Privacy Tools
   - Ghostery
   - AdBlock Plus
   - HTTPS Everywhere

2. Effectiveness Browser Protections
   - DNT
   - Third-party cookie Blocking
   - Firefox Tracking Protection

3. Use of javascript for tracking
   - Canvas Fingerprinting
   - Property Enumeration
   - WebRTC Local IP Sniffing
Supporting a variety of measurements

1. Effectiveness of Privacy Tools
   - Ghostery
   - AdBlock Plus
   - HTTPS Everywhere

2. Effectiveness Browser Protections
   - DNT
   - Third-party cookie Blocking
   - Firefox Tracking Protection

3. Use of javascript for tracking
   - Canvas Fingerprinting
   - Property Enumeration
   - WebRTC Local IP Sniffing

4. Tracking Practices
   - Cookie Syncing
   - Cookie Respawning
   - Setting ID cookies
Case Study 1: Canvas Fingerprinting

Case Study 2: WebRTC Local IP Sniffing
2012: Canvas Fingerprinting Introduced

<table>
<thead>
<tr>
<th>Windows:</th>
</tr>
</thead>
<tbody>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OS X:</th>
</tr>
</thead>
<tbody>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linux:</th>
</tr>
</thead>
<tbody>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
<tr>
<td>How quickly daft jumping zebras vex. (Also, pur</td>
</tr>
</tbody>
</table>

Figure 6: 13 ways to render 20px Arial

Source: Mowery and Shacham; Pixel Perfect: Fingerprinting Canvas in HTML5
2012: Canvas Fingerprinting Introduced

Windows:
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur

OS X:
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur

Linux:
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur
- How quickly daft jumping zebras vex. (Also, pur

Original Image:
- How quickly daft jumping zebras vex. (L
- How quickly daft jumping zebras vex. (L
- How quickly daft jumping zebras vex. (L

OSX:
- How quickly daft jumping zebras vex. (O

Windows (XP, Vista, 7):
- How quickly daft jumping zebras vex. (W
- How quickly daft jumping zebras vex. (W
- How quickly daft jumping zebras vex. (W

Windows 8:
- How quickly daft jumping zebras vex. (W

Figure 6: 13 ways to render 20px Arial

Figure 7: Difference maps for a group on text_arial

Source: Mowery and Shacham; Pixel Perfect: Fingerprinting Canvas in HTML5
The Web Never Forgets: Persistent Tracking Mechanisms in the Wild

Gunes Acar¹, Christian Eubank², Steven Englehardt³, Marc Juarez¹
Arvind Narayanan², Claudia Diaz¹

¹KU Leuven, ESAT/COSIC and iMinds, Leuven, Belgium
{name.surname}@esat.kuleuven.be
²Princeton University
{cge,ste,arvindn}@cs.princeton.edu

ABSTRACT
We present the first large-scale studies of three advanced web tracking mechanisms — canvas fingerprinting, evercookies and use of "cookie syncing" in conjunction with evercookies.

1. INTRODUCTION
A 1999 New York Times article called cookies comprehensive privacy invaders and described them as "surveillance cookies".
2014: Canvas Fingerprinting Measured

Source: Acar, Eubank, Englehardt, Juarez, Narayanan, Diaz; The Web Never Forgets: Persistent Tracking Mechanisms in the Wild
2014: Canvas Fingerprinting Measured

1. Write a Firefox patch
2. Write automation with Selenium
3. Write analysis code

Source: Acar, Eubank, Englehardt, Juarez, Narayanan, Diaz; The Web Never Forgets: Persistent Tracking Mechanisms in the Wild
Case Study 1: Canvas Fingerprinting

Case Study 2: WebRTC Local IP Sniffing
1. I saw a tweet that nytimes.com is IP sniffing

WebRTC being used now by embedded 3rd party on nytimes.com to report visitors' local IP addresses.
2. I added code to JS Instrumentation for next crawl

```javascript
// Access to webRTC
instrumentObject(window.mozRTCPeerConnection.prototype
    "mozRTCPeerConnection",
    prototype=true);
```
3. I wrote some analysis code

- Grab all urls that execute
  - `mozRTCPeerConnection.onicecandidate`
  - `mozRTCPeerConnection.createDataChannel`
  - `mozRTCPeerConnection.createOffer`

- Check JS Files to confirm
4. I found several third-parties sniffing local IP

- 121 first-party sites (October 2015)
  - 29 in the top 10k
- 24 unique scripts
- Only 1 of which is blocked by EasyList/EasyPrivacy
Measurement with OpenWPM is much easier

**Canvas Fingerprinting**

1. Write a Firefox patch

2. Write automation with Selenium

3. Write analysis code

**WebRTC Local IP Sniffing**
Measurement with OpenWPM is much easier

Canvas Fingerprinting

1. Write a Firefox patch
2. Write automation with Selenium
3. Write analysis code

WebRTC Local IP Sniffing

1. Write 1 line of JavaScript
Measurement with OpenWPM is much easier

**Canvas Fingerprinting**

1. Write a Firefox patch
2. Write automation with Selenium
3. Write analysis code

**WebRTC Local IP Sniffing**

1. Write 1 line of JavaScript
2. Use OpenWPM
Measurement with OpenWPM is much easier

**Canvas Fingerprinting**

1. Write a Firefox patch
2. Write automation with Selenium
3. Write analysis code

**WebRTC Local IP Sniffing**

1. Write 1 line of JavaScript
2. Use OpenWPM
3. Write analysis code
Where to go from here:

1. Inform the public
2. Provide data for privacy tools
3. Make data more accessible to less technical investigators
We’d like to collaborate with you

1. Submit pull requests for OpenWPM
2. Use OpenWPM to run measurements and release the data
3. Download our data and build analysis on top of it
   a. (Coming soon!)
Help us make the web more transparent!

- **Contribute:**
  - [github.com/citp/OpenWPM](https://github.com/citp/OpenWPM)

- **Collaborate:**
  - [webtap.princeton.edu](http://webtap.princeton.edu)

**Email:** ste@cs.princeton.edu

**Twitter:** @s_inglehardt