Analyzing the Impact of Large Scale Online Tracking Measurement

Steven Englehardt
@s_englehardt
The Princeton Web Census

Monthly
1 Million Site Crawl

Collecting:
- Javascript Calls
- All javascript files
- HTTP Requests and Responses
- Storage (cookies, Flash, etc)
Open Web Privacy Measurement (OpenWPM)

https://github.com/citp/OpenWPM
Insights from the Princeton Web Census

New metric to rank third parties

Trackers impede HTTPS adoption

Consolidation of third-parties

45 out of top 50 3rd parties cookie sync

Online Tracking: A 1-million-site Measurement and Analysis (CCS 2016)
How does measurement of new tracking techniques influence trackers and vendors?

Canvas

WebRTC

Audio

Battery

https://webtransparency.cs.princeton.edu/webcensus/
Canvas Fingerprinting

The Web Never Forgets: Persistent Tracking Mechanisms in the Wild (Acar, et al.)
Canvas Fingerprinting

Figure 6: 13 ways to render 20px Arial

Source: Pixel Perfect: Fingerprinting Canvas in HTML5 (Mowery and Shacham)
Canvas fingerprinting returns in the absence of measurement

**May 2014:** 5% of sites

**Aug 2014:** ~0.1% of sites

**Jan 2016:** 2.6% of sites

Percentage of the Alexa top 100k sites
Using AudioContext for fingerprinting

Used by:
cdn-net.com script

Used by:
pxi.pub and ad-score.com scripts
Using AudioContext for fingerprinting

Live test page: https://audiofingerprint.openwpm.com/
AudioContext fingerprinting the Tor Browser

271 samples from the Tor Browsers
- 7 distinct fingerprints (2 fingerprints account for 80% of samples)
- Overlap with fingerprints from Firefox shows these largely reveal OS of device
Firefox adds a preference to disable Web Audio API

Actual results:

AudioContext, DynamicsCompressor and OscillatorNode are being used in the wild to "fingerprint" web users, as reported by researchers: https://techcrunch.com/2016/05/19/audio-fingerprinting-being-used-to-track-web-users-study-finds/

(online tests here: https://browserprint.info)

Expected results:

An option to disable Web Audio is needed to protect users privacy. It appears this preference was introduced as "media.audio_data.enabled" per this request https://bugzilla.mozilla.org/show_bug.cgi?id=665598 but it's not available anymore in the latest versions of Firefox.
Fingerprinting Protection Mode

Lloyd Dewolf edited this page on Jan 29 · 11 revisions

Fingerprinting methods blocked in Fingerprinting Protection Mode

- Canvas fingerprinting
- WebGL fingerprinting
- AudioContext fingerprinting
- WebRTC IP leakage
- Battery Status fingerprinting (disabled in general, not just when FP mode is turned on)
cberescu commented on May 23, 2016

Using the OfflineAudioContext API created a key for the fingerprint. The mobile support for the OfflineAudioContext is not great so it jumps over if mobile is detected.

Tested also the AudioContext API but does not work all the time.

Code and info about Audio Fingerprint found here: https://audiofingerprint.openwpm.com/
Using Battery Status to Track

Battery Status:
level: 0.11
dischargeTime: 12867

The Leaking Battery, Olejnik et. al. (2015)
Browsers remove BatteryStatus API citing privacy

### Bug 1313580 - Remove web content access to Battery API

<table>
<thead>
<tr>
<th>Status</th>
<th>VERIFIED FIXED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whiteboard</td>
<td></td>
</tr>
<tr>
<td>Keywords</td>
<td>addon-compat, dev-doc-needed, privacy, site-compat</td>
</tr>
<tr>
<td>Product</td>
<td>Core (show info)</td>
</tr>
<tr>
<td>Component</td>
<td>DOM: Device Interfaces (show other bugs) (show info)</td>
</tr>
<tr>
<td>Version</td>
<td>unspecified</td>
</tr>
<tr>
<td>Platform</td>
<td>Unspecified Unspecified</td>
</tr>
<tr>
<td>Importance</td>
<td>-- normal (vote)</td>
</tr>
<tr>
<td>Target Milestone</td>
<td>mozilla52</td>
</tr>
<tr>
<td>Assigned To</td>
<td>Chris Peterson [cpeterson]</td>
</tr>
<tr>
<td>Reported</td>
<td>2016-10-27 23:28 PDT by Chris Peterson [cpeterson]</td>
</tr>
<tr>
<td>Modified</td>
<td>2016-11-02 09:53 PDT (History)</td>
</tr>
<tr>
<td>CC List</td>
<td>7 users (show)</td>
</tr>
<tr>
<td>Flags</td>
<td>ryanvm: in-testsuite-</td>
</tr>
<tr>
<td>QA Whiteboard</td>
<td></td>
</tr>
<tr>
<td>Iteration</td>
<td>---</td>
</tr>
<tr>
<td>Points</td>
<td>---</td>
</tr>
<tr>
<td>Has Regression Range</td>
<td>---</td>
</tr>
</tbody>
</table>

Personas is no longer an option for authentication on BMO. For more details see Persona Deprecated.
Browsers remove BatteryStatus API citing privacy
Browsers remove BatteryStatus API citing privacy

Disable support for Battery Status API

Fix #1885. Battery status is disabled for all users (not just those with Fingerprinting Protection) because Firefox and webkit recently removed support for it. c.f. https://bugs.chromium.org/p/chromium/issues/detail?id=661792.

Auditors: @bbondy

Test Plan: covered by automated test
Battery Status Not Included:
Assessing Privacy in Web Standards

Lukasz Olejnik
University College London
lukasz.w3c@gmail.com
https://lukaszolejnik.com

Steven Englehardt
Princeton University
ste@cs.princeton.edu
http://senglehardt.com

Arvind Narayanan
Princeton University
arvindn@cs.princeton.edu
http://randomwalker.info

Abstract—The standardization process is core to the development of the open web. Until 2013, the process rarely included privacy review and had no formal privacy requirements. But today the importance of privacy engineering has become apparent to standards bodies such as the W3C as well as to browser vendors. Standards groups now have guidelines for privacy assessments, and are including privacy reviews in many new specifications. However, the standards community does not yet have much practical experience in assessing privacy.

In this paper we systematically analyze the W3C Battery Status API to help inform future privacy assessments. We begin by reviewing its evolution — the initial specification, which only cursorily addressed privacy, the discovery of surprising privacy vulnerabilities as well as actual misuse in the wild, followed by the removal of the API from major browser engines, an unprecedented move. Next, we analyze web measurement data from late 2016 and confirm that the majority of scripts used the API for fingerprinting. Finally, we draw lessons from this affair and make recommendations for improving privacy engineering of web standards.

I. INTRODUCTION

The Battery Status API offers an interesting and unusual case study of privacy assessment in the web standardization process. Unlike cookies, which are central to the identity of the user and can track online behavior, the Battery Status API can be used to obtain information that is less personal. It is often used to track whether a user is on a mobile device or not. This information can be used by developers to optimize their applications for different devices. However, the API also reveals to a web vendor when the user’s device may be out of power, which could be used to prevent users from accessing certain content or services. This raises some questions about the privacy implications of the API.

II. BACKGROUND

The W3C is a standards organization that develops web standards. These standards are used by developers to create web applications. The W3C has a process for developing and approving web standards. The process includes public review and feedback, consensus building, and final approval. One of the stages in the process is privacy review. Privacy review involves assessing the impact of a standard on user privacy.

III. METHODOLOGY

We conducted a systematic analysis of the W3C Battery Status API. We reviewed the initial specification, the changes made to the API, and the use of the API by web vendors. We also analyzed web measurement data to determine the extent of the API’s usage.

IV. RESULTS

Our analysis revealed that the Battery Status API is used by many web vendors. It is often used to track whether a user is on a mobile device or not. This information can be used to optimize applications for different devices. However, the API also reveals to a web vendor when the user’s device may be out of power, which could be used to prevent users from accessing certain content or services. This raises some questions about the privacy implications of the API.

V. DISCUSSION

The Battery Status API offers an interesting and unusual case study of privacy assessment in the web standardization process. Unlike cookies, which are central to the identity of the user and can track online behavior, the Battery Status API can be used to obtain information that is less personal. It is often used to track whether a user is on a mobile device or not. This information can be used by developers to optimize their applications for different devices. However, the API also reveals to a web vendor when the user’s device may be out of power, which could be used to prevent users from accessing certain content or services. This raises some questions about the privacy implications of the API.

VI. CONCLUSION

In conclusion, we have shown that the Battery Status API is used by many web vendors. It is often used to track whether a user is on a mobile device or not. This information can be used to optimize applications for different devices. However, the API also reveals to a web vendor when the user’s device may be out of power, which could be used to prevent users from accessing certain content or services. This raises some questions about the privacy implications of the API.

VII. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. 1560960.

VIII. REFERENCES


IX. APPENDICES

See supplementary materials for additional information.

X. BIOGRAPHIES

Lukasz Olejnik is a Ph.D. student at University College London. His research interests include privacy engineering and the web standardization process. Steven Englehardt is an associate professor at Princeton University. His research interests include privacy engineering and the web standardization process. Arvind Narayanan is a professor at Princeton University. His research interests include privacy engineering and the web standardization process.
Summary:

1. **Transparency helps good and bad actors**
   a. Vendors prioritize fixes
   b. Major trackers react to public pressure
   c. Less known trackers start using the technique

2. **Non-technical users may be left behind**
   a. Privacy protection often ends up in products for technical users
      i. Tor Browser
      ii. Brave Browser
   b. Solutions for non-technical users may take years, leaving them at a potential disadvantage
A Path Forward

1. **Take continual measurements**
   a. Princeton Web Census data collected monthly
   b. Data interface under development (ask us for an invite)

2. **Provide up-to-date results**
   a. Make updated fingerprinting script lists available

3. **Research measurement-informed privacy solutions**
   a. Sandboxing trackers and fingerprinters (my talk tomorrow)
   b. Automated fingerprint detection