

Online Tracking

A 1-million-site Measurement and Analysis

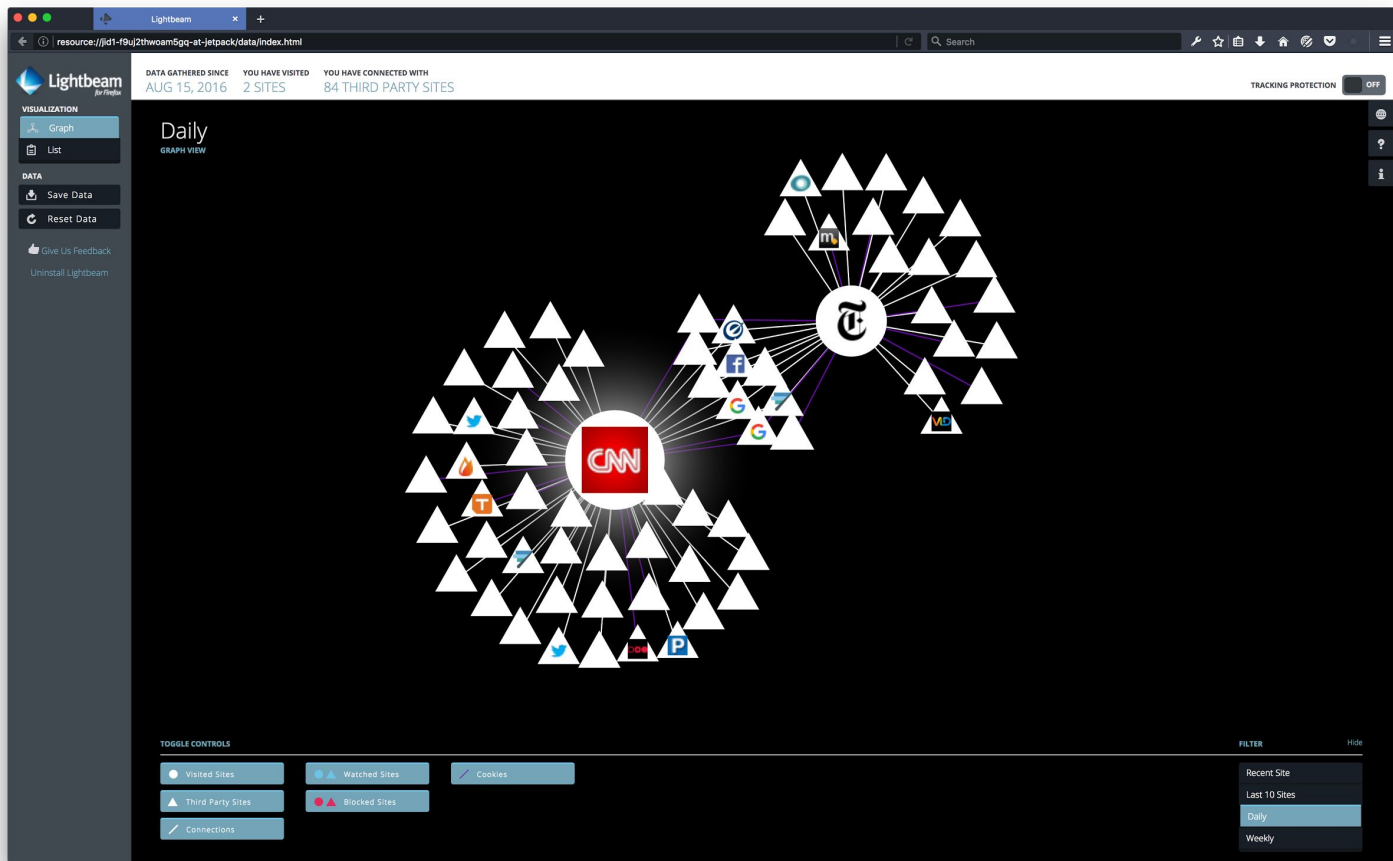
Steven Englehardt
@s_englehardt

Arvind Narayanan
@random_walker

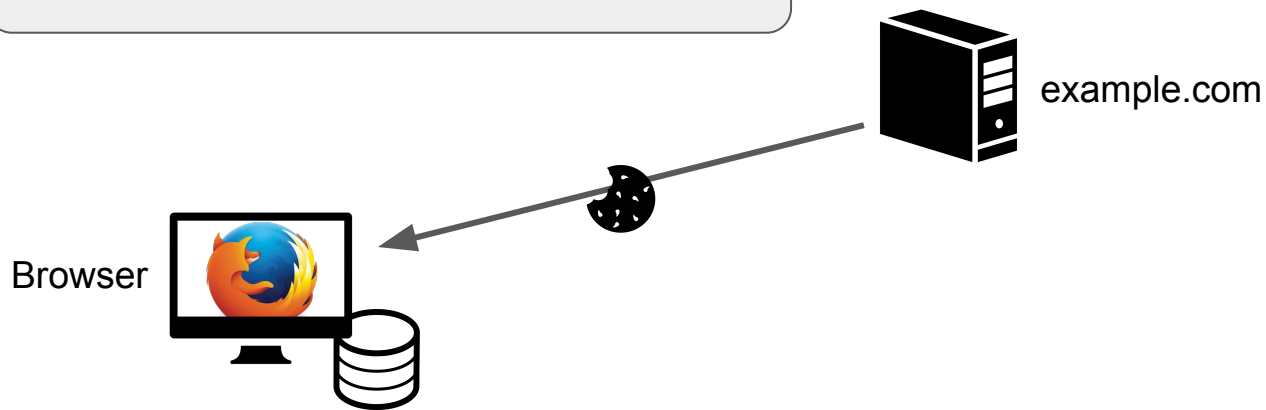


This research was supported by NSF award CNS 1526353, a grant from the Data Transparency Lab, and an Amazon AWS Credits Research Grant.

Visiting 2 websites results in 84 third parties contacted



Tracking with browser state



Tracking with browser state

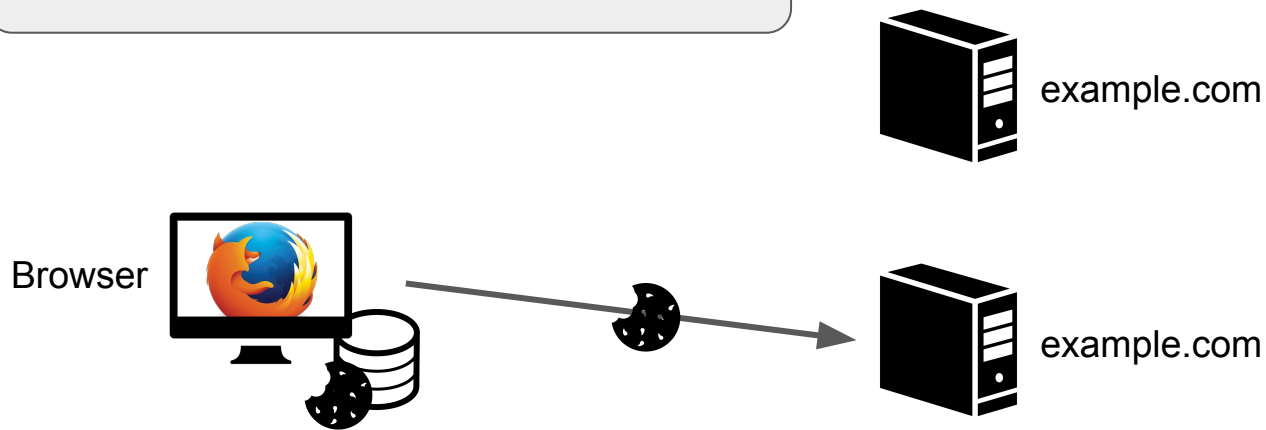


example.com

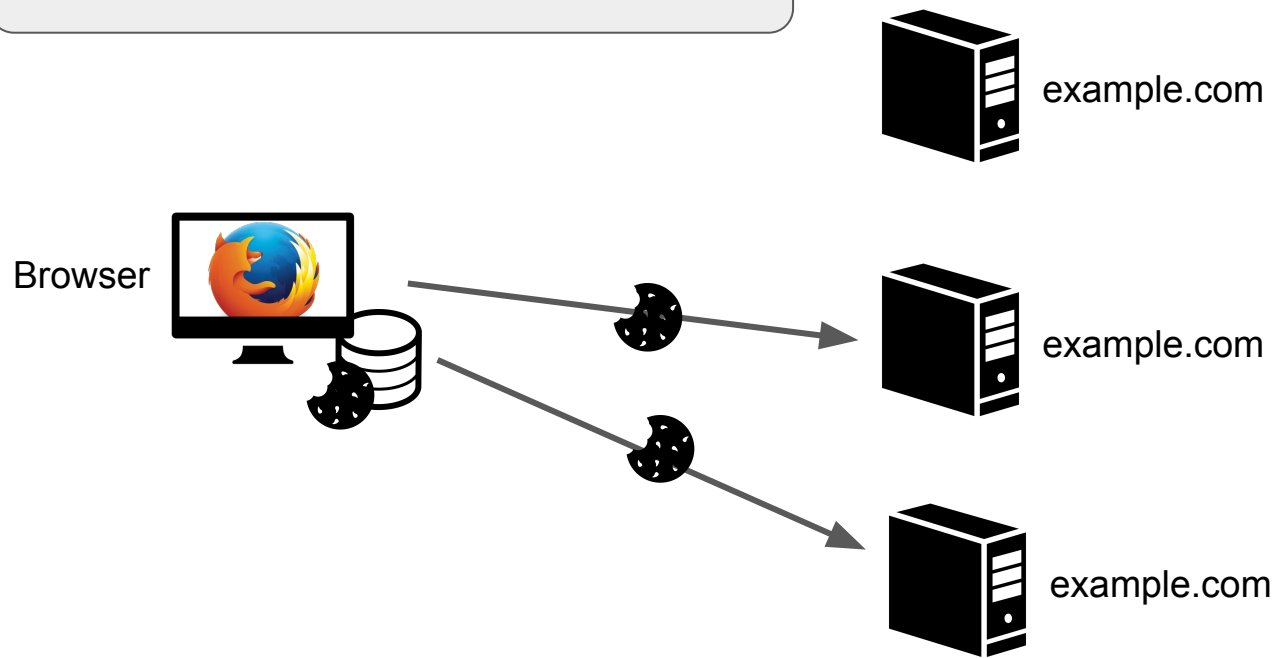
Browser



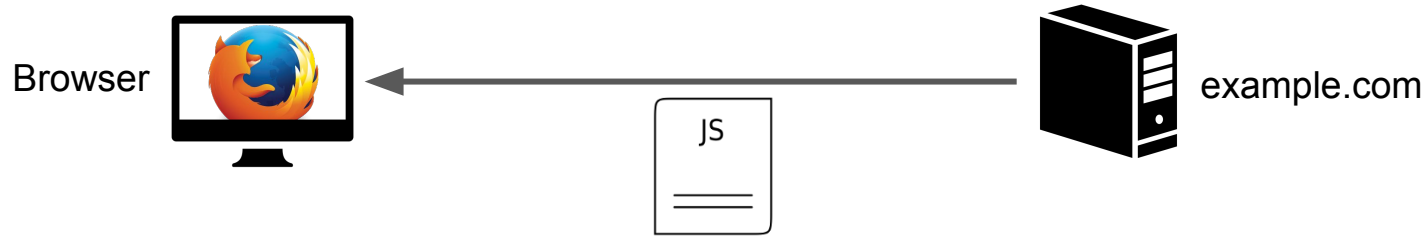
Tracking with browser state



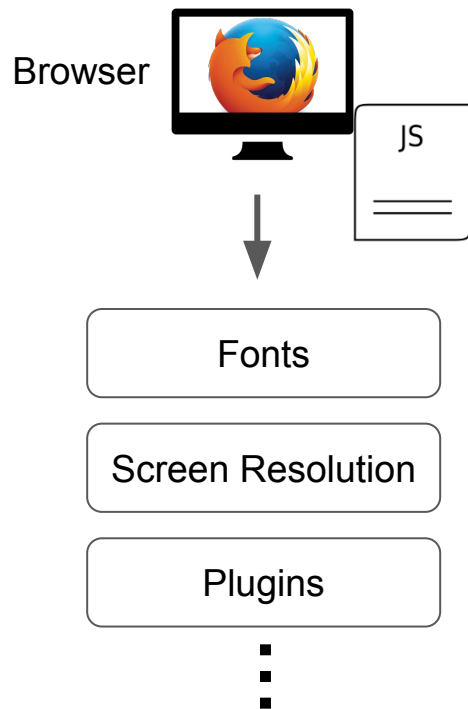
Tracking with browser state



Tracking with fingerprinting

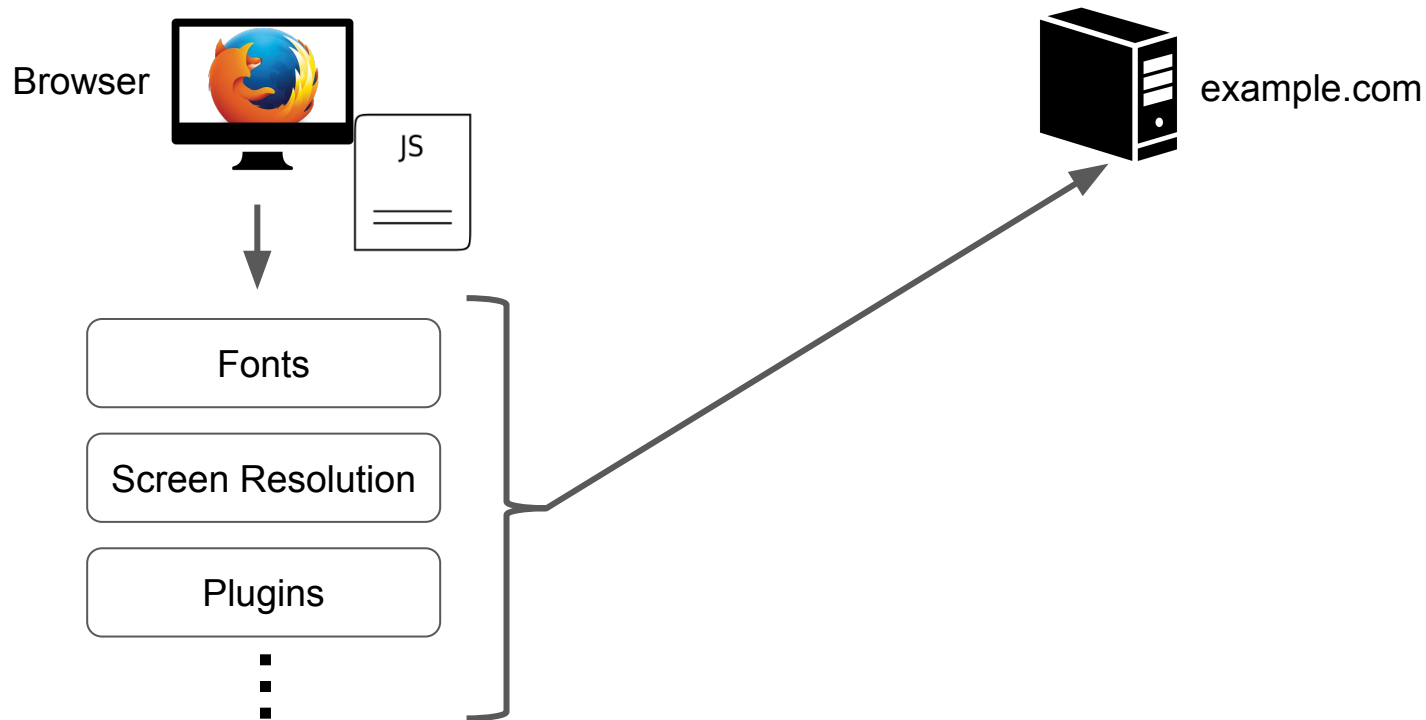


Tracking with fingerprinting



example.com

Tracking with fingerprinting



Open Web Privacy Measurement (OpenWPM)

The screenshot shows the GitHub repository page for `citp / OpenWPM`. At the top, there are buttons for `Unwatch` (49), `Unstar` (435), and `Fork` (67). Below these are tabs for `Code`, `Issues` (45), `Pull requests` (0), `Projects` (0), `Wiki`, `Pulse`, `Graphs`, and `Settings`. A description states: "A web privacy measurement framework <https://webtap.princeton.edu/> — Edit".

Repository statistics are shown in a bar: 480 commits, 4 branches, 12 releases, 13 contributors, and GPL-3.0 license. Below this is a progress bar. Action buttons include `Branch: master`, `New pull request`, `Create new file`, `Upload files`, `Find file`, and `Clone or download`.

A commit by `englehardt` is highlighted: "Merge branch 'master' of github.com:citp/OpenWPM" (Latest commit 3a14416 7 hours ago). Below this is a list of files and their commit messages:

File	Commit Message	Time
<code>automation</code>	Added comments about new commands	15 days ago
<code>test</code>	disabling audiocontext test for travis CI	15 days ago
<code>.gitignore</code>	Merge branch 'master' of github.com:citp/OpenWPM	10 months ago
<code>.travis.yml</code>	Add travis.yml file to run continuous integration tests.	6 months ago
<code>CHANGELOG</code>	Version bump to 0.6.2. Bugfix in previous version	6 months ago
<code>LICENSE</code>	Removing extra whitespace from all infrastructure files	10 months ago
<code>README.md</code>	Modified readme to only use travis status from master branch	15 days ago

<https://github.com/citp/OpenWPM>

The Princeton Web Census

Monthly
1 Million Site Crawl

Collecting:

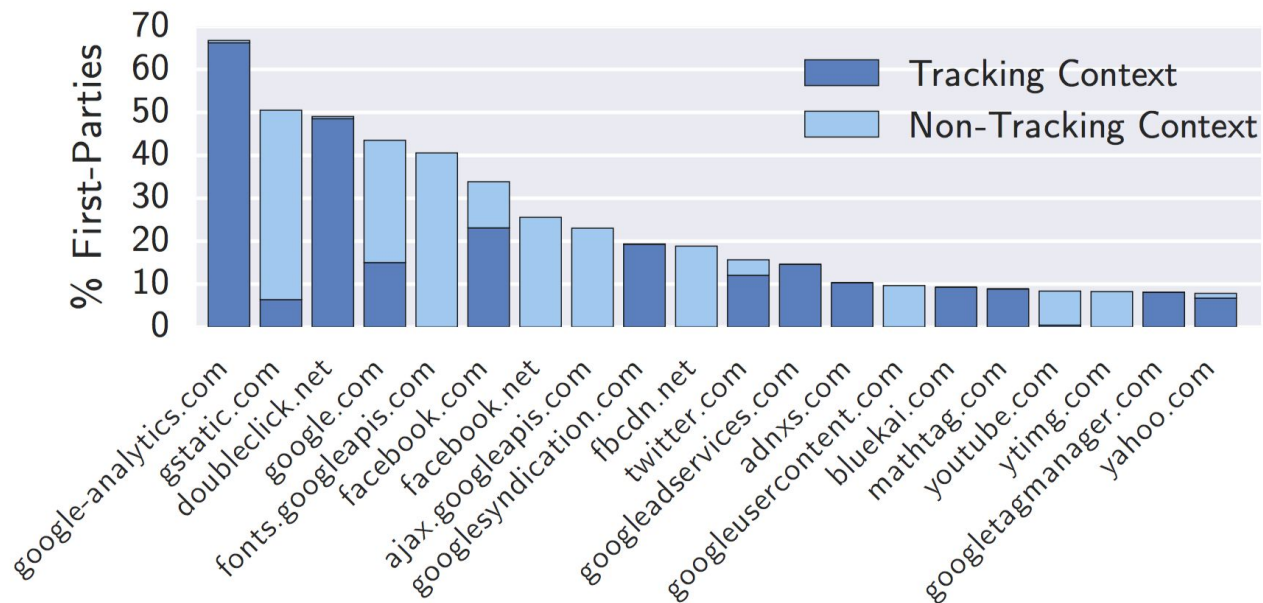
- Javascript Calls
- All javascript files
- HTTP Requests and Responses
- Storage (cookies, Flash, etc)

Measurement is effective because most actors are not malicious

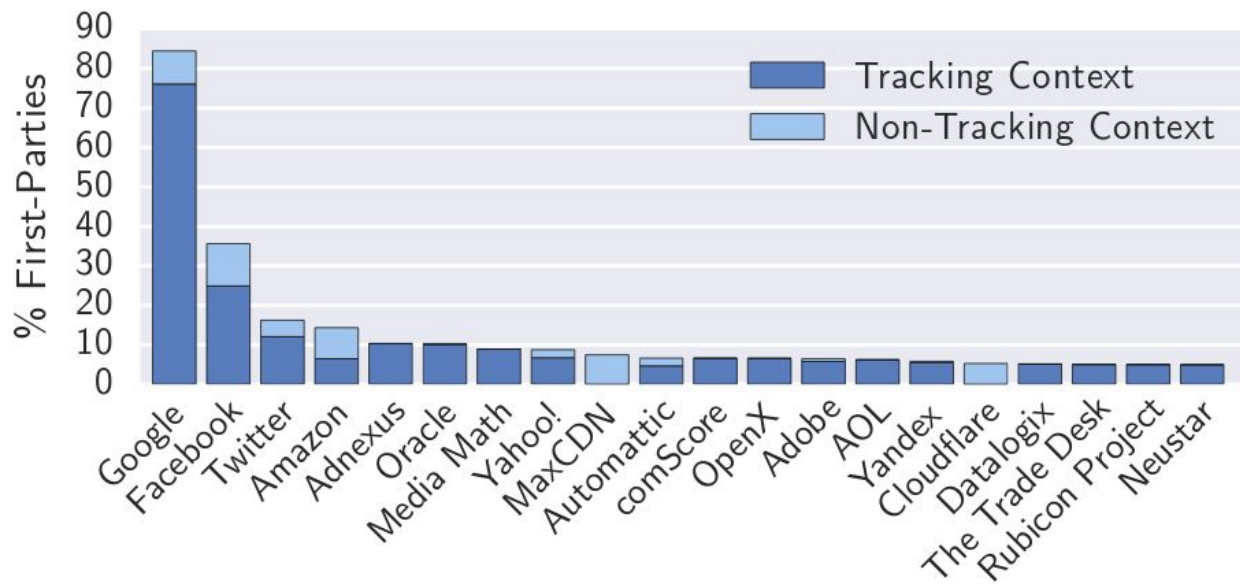
1. Bulk of trackers respond to pressure from publishers, users, and regulators
2. Not trying to avoid detection
3. High risk for malicious actions

Research findings from the Princeton Web Census

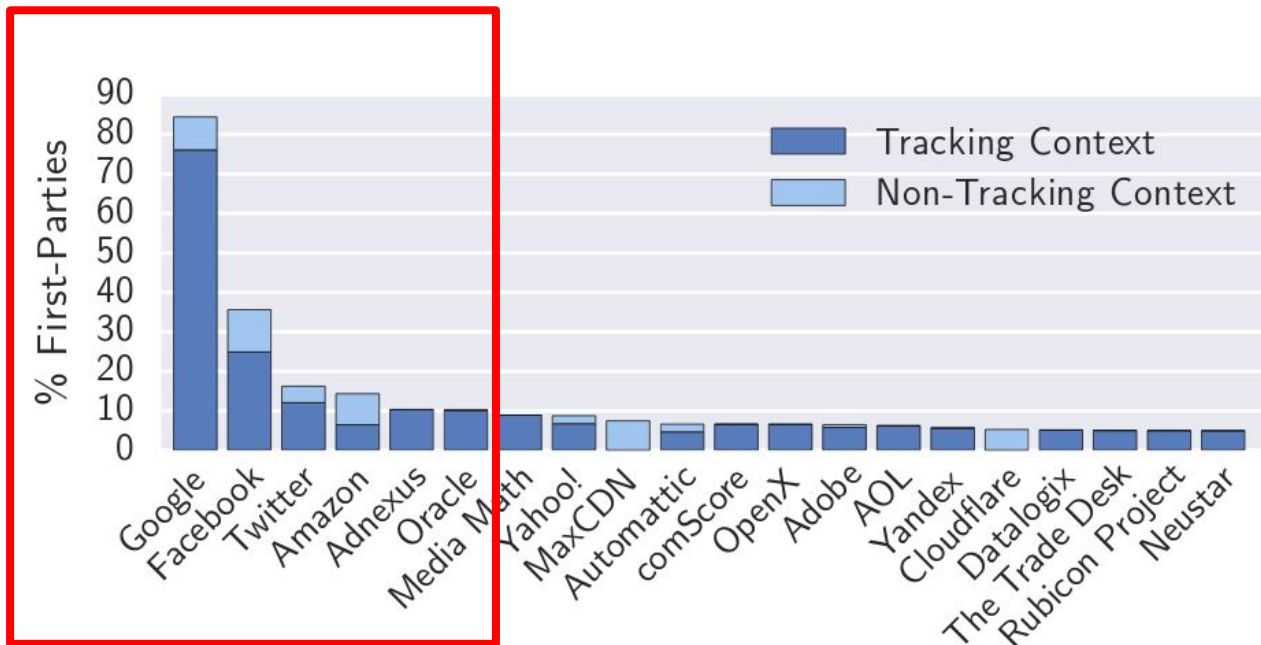
The long tail of third-party tracking



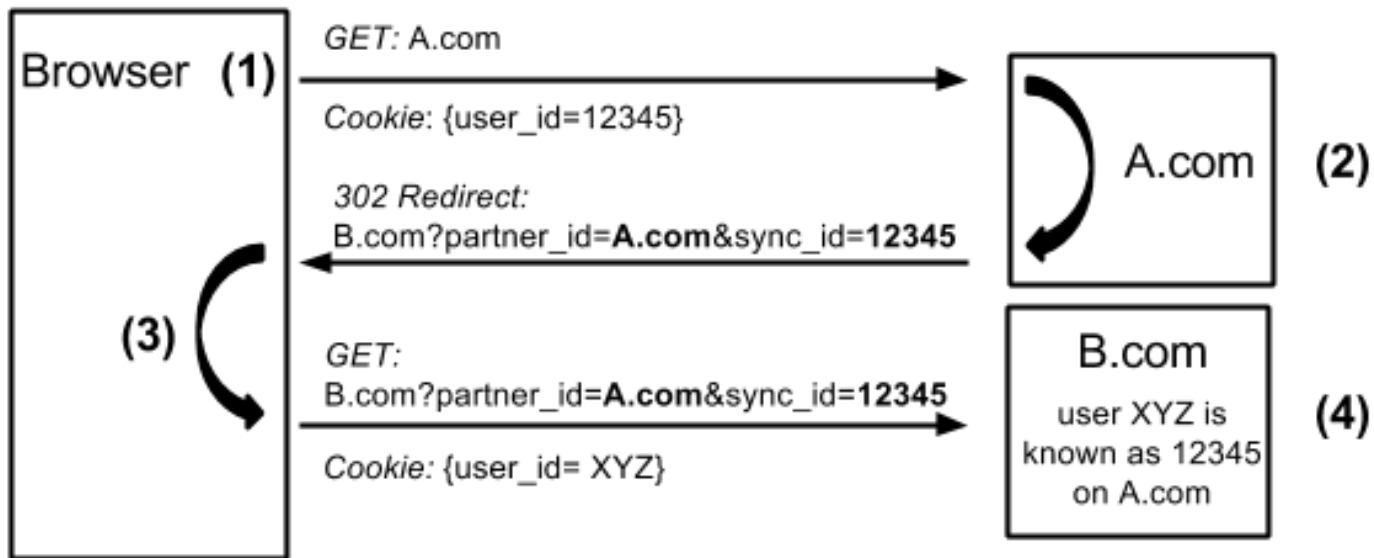
A consolidated tracking ecosystem



Only 6 organizations are present on >10% of sites



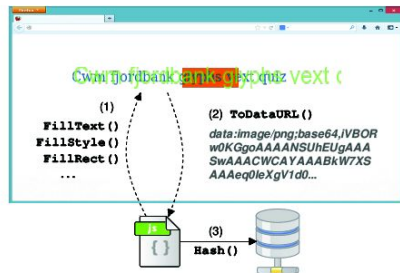
Almost all top third parties cookie sync



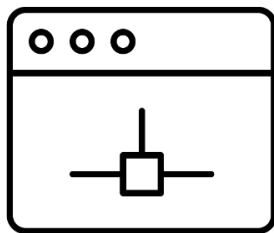
45 of top 50 third parties sync cookies (85% chance any two share an ID)

New browser features used for fingerprinting

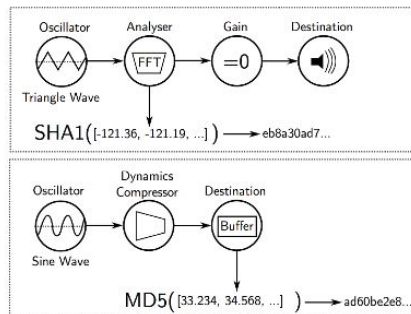
Canvas



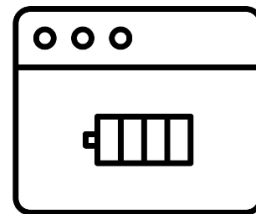
WebRTC



Audio



Battery



<https://webtransparency.cs.princeton.edu/webcensus/>

Detecting Fingerprinting

```
// Measurement Code  
instrumentObject(window.CanvasRenderingContext2D.prototype, ...);  
instrumentObject(window.HTMLCanvasElement.prototype, ...);
```

Detecting Fingerprinting

```
// Measurement Code  
instrumentObject(window.CanvasRenderingContext2D.prototype, ...);  
instrumentObject(window.HTMLCanvasElement.prototype, ...);
```

```
// Canvas Fingerprinting Example  
ctx = canvas.getContext("2d");  
ctx.fillText("hello world", 2, 15);  
ctx.fillStyle = "#f60";  
ctx.fillRect(125, 1, 62, 20);  
fp = canvas.toDataURL();
```

Detecting Fingerprinting

// Measurement Code

```
instrumentObject(window.CanvasRenderingContext2D.prototype, ...);  
instrumentObject(window.HTMLCanvasElement.prototype, ...);
```

// Canvas Fingerprinting Example

```
ctx = canvas.getContext("2d");  
ctx.fillText("hello world", 2, 15);  
ctx.fillStyle = "#f60";  
ctx.fillRect(125, 1, 62, 20);  
fp = canvas.toDataURL();
```

Measurement Logs

(SCRIPT_URL, "getContext", "2d")
(SCRIPT_URL, "fillText", "hello world", 2, 15)
(SCRIPT_URL, "fillStyle", "#f60")
(SCRIPT_URL, "fillRect", 125, 1, 62, 20)
(SCRIPT_URL, "toDataURL", "data: ...")

Detecting Fingerprinting

// Measurement Code

```
instrumentObject(window.CanvasRenderingContext2D.prototype, ...);  
instrumentObject(window.HTMLCanvasElement.prototype, ...);
```

// Canvas Fingerprinting Example

```
ctx = canvas.getContext("2d");  
ctx.fillText("hello world", 2, 15);  
ctx.fillStyle = "#f60";  
ctx.fillRect(125, 1, 62, 20);  
fp = canvas.toDataURL();
```

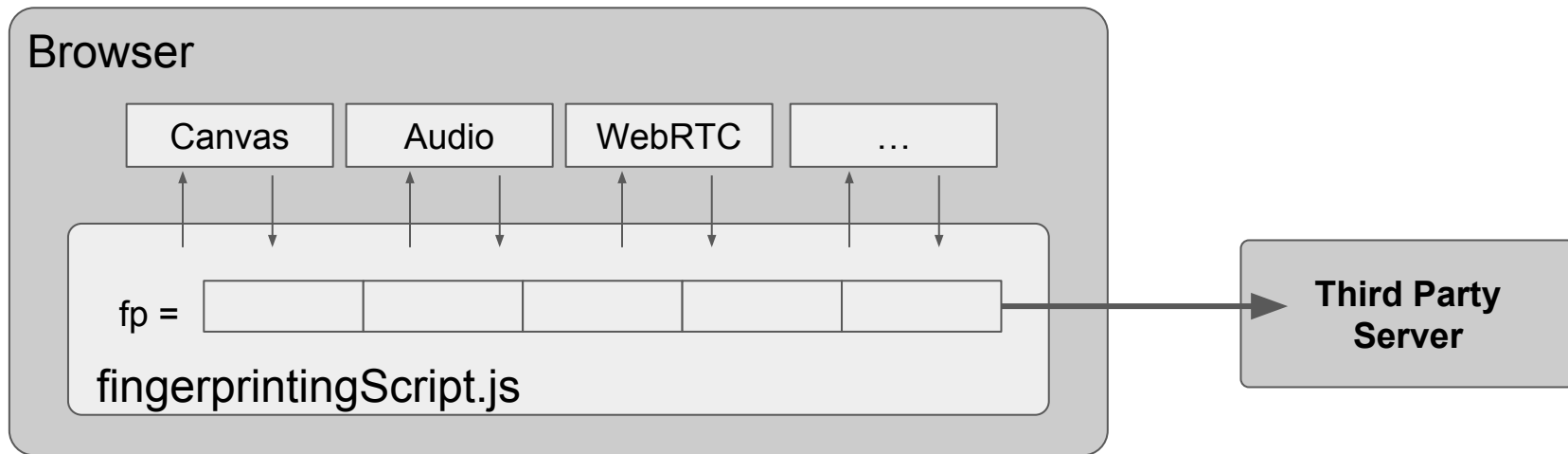
Measurement Logs

(SCRIPT_URL, "getContext", "2d")
(SCRIPT_URL, "fillText", "hello world", 2, 15)
(SCRIPT_URL, "fillStyle", "#f60")
(SCRIPT_URL, "fillRect", 125, 1, 62, 20)
(SCRIPT_URL, "toDataURL", "data: ...")

Post-measurement Analysis

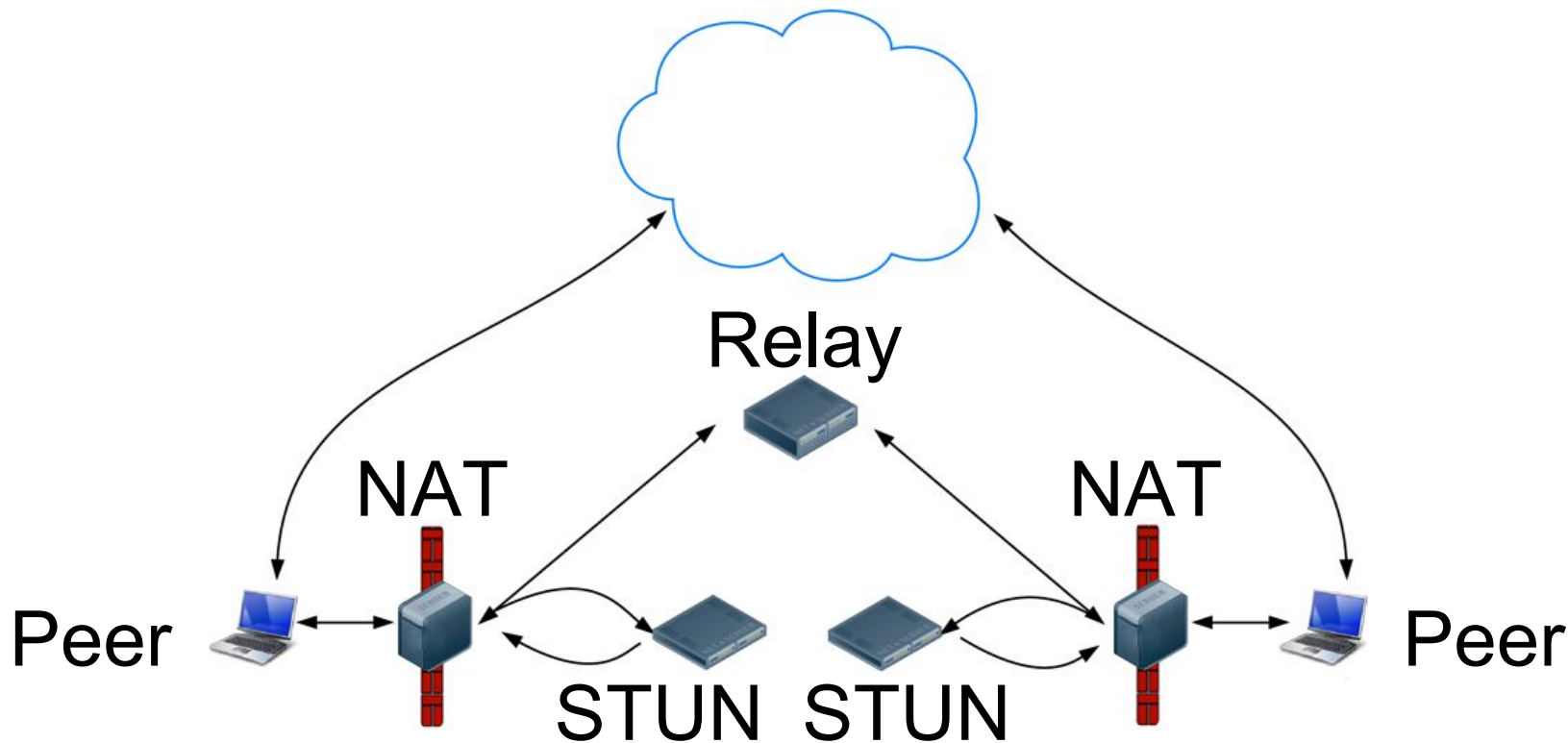
1. Examine API use for fingerprinting
2. Check for tampering / instrumentation inspection

Detecting Fingerprinting



1. Observe a sequence of API calls
2. Techniques clustered together
3. Results of calls combined and sent to server
4. Limited API use beyond that for fingerprinting

Abusing WebRTC candidate generation for tracking



WebRTC `dataChannel` requires no permissions

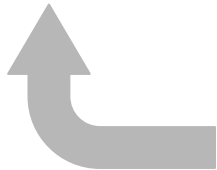
Without user intervention, a tracking script can:

1. Reveal the user's real IP address when behind a VPN
2. Reveal the user's local IP address for each local interface.

WebRTC `dataChannel` requires no permissions

Without user intervention, a tracking script can:

1. Reveal the user's real IP address when behind a VPN
2. Reveal the user's local IP address for each local interface.



More identifying for corporate and university users.

Measuring the use of WebRTC for tracking

Measurement Code:

```
// Access to webRTC
instrumentObject(
    window.RTCPeerConnection.prototype,
    "RTCPeerConnection", true
);
```

Measuring the use of WebRTC for tracking

Measurement Code:

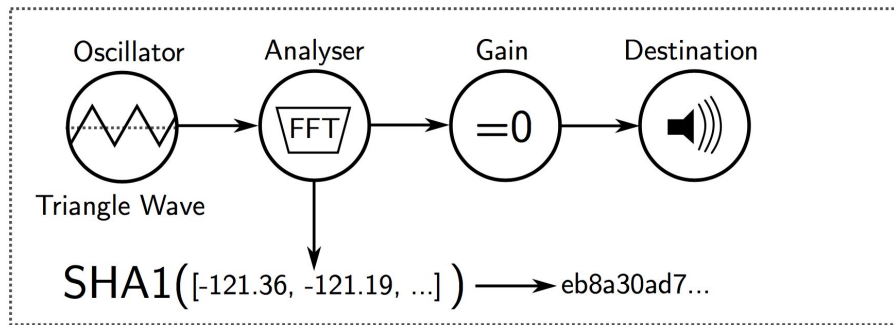
```
// Access to webRTC
instrumentObject(
  window.RTCPeerConnection.prototype,
  "RTCPeerConnection", true
);
```

**~90% of unsolicited dataChannel use
on homepages is for tracking**

57 scripts on 625 sites.

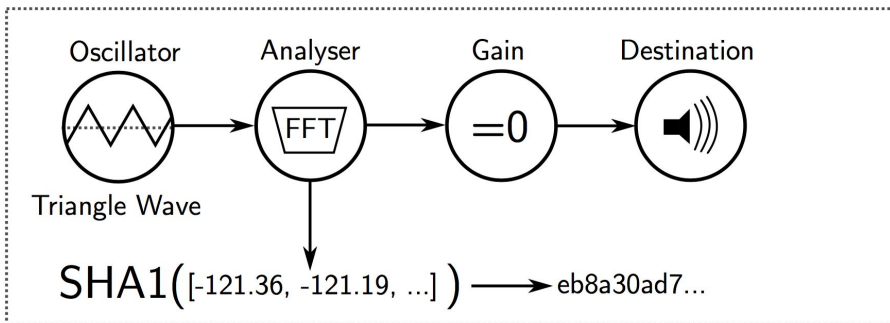
Using AudioContext for fingerprinting

Used by:
cdn-net.com script

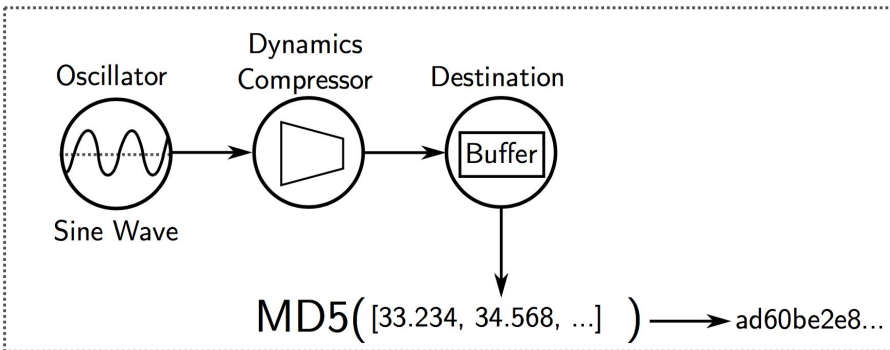


Using AudioContext for fingerprinting


Used by:
`cdn-net.com` script



Used by:
`pxi.pub` and
`ad-score.com` scripts



Implications for Tor Browser



[Login](#) | [Preferences](#) | [Help/Guide](#) | [About Trac](#) | [Register](#)

[View Tickets](#) | [Browse Source](#) | [Roadmap](#) | [Timeline](#) | [Wiki](#) | [Search](#) | [Tags](#)

[← Previous Ticket](#) | [Next Ticket →](#)

#13017 **assigned task**

Opened **2 years ago**
Last modified **3 weeks ago**

Determine if AudioBuffers/OfflineAudioContext are a fingerprinting vector

Reported by:	mikeperry	Owned by:	arthuredelstein
Priority:	Very High	Milestone:	
Component:	Applications/Tor Browser	Version:	
Severity:	Critical	Keywords:	tbb-fingerprinting-os, TorBrowserTeam201610
Cc:	arthuredelstein , isis , mcs , brade	Actual Points:	
Parent ID:		Points:	
Reviewer:		Sponsor:	

Description

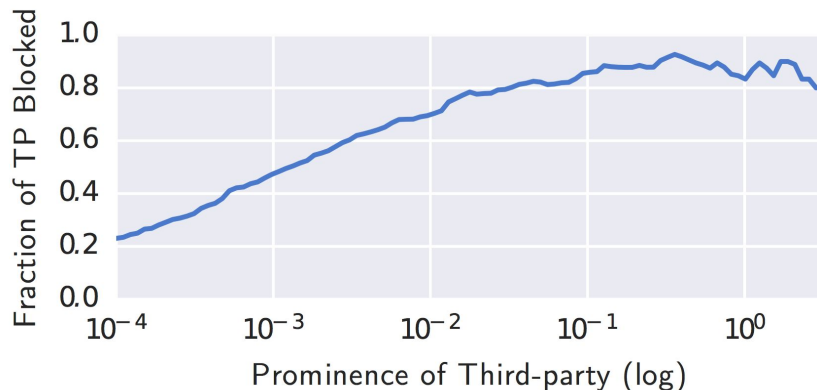
WebAudio allows you to write data to AudioBuffers and perform effects/manipulation/spectral analysis on them, and extract their contents.

If the underlying routines are OS supported, they may be fingerprintable. https://developer.mozilla.org/en-US/docs/Web_Audio_API

Do Privacy Tools Help?

Privacy tools effectively block stateful tracking

- Third-party cookie blocking
 - 32 out of 50,000 sites work around blocking by redirecting the top-level domain
 - Average number of third-parties per site reduced from ~18 to ~13
- Ghostery
 - Average number of third-parties per site reduced from ~18 to ~3
 - Very few third-party cookies are set



Crowdsourced lists miss fingerprinters

EasyList + EasyPrivacy

Technique	Percentage of Scripts	Percentage of Sites

Crowdsourced lists miss fingerprinters

EasyList + EasyPrivacy

Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%

Crowdsourced lists miss fingerprinters

EasyList + EasyPrivacy

Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%
Canvas Font	10%	91%

Crowdsourced lists miss fingerprinters

EasyList + EasyPrivacy

Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%
Canvas Font	10%	91%
WebRTC	5%	6%

Crowdsourced lists miss fingerprinters

EasyList + EasyPrivacy

Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%
Canvas Font	10%	91%
WebRTC	5%	6%
AudioContext	6%	2%

Takeaways

1. Trackers are employing an increasingly diverse set of techniques
2. Measurement heavily influences and controls the adoption of new techniques and tracking norms.
3. Crowdsourced tracking protection misses less popular trackers/techniques
4. Frequent measurement and automated detection provide a path forward

Email: ste@cs.princeton.edu **Twitter:** [@s_englehardt](https://twitter.com/s_englehardt) **Web:** senglehardt.com

Canvas fingerprinting returns in the absence of measurement

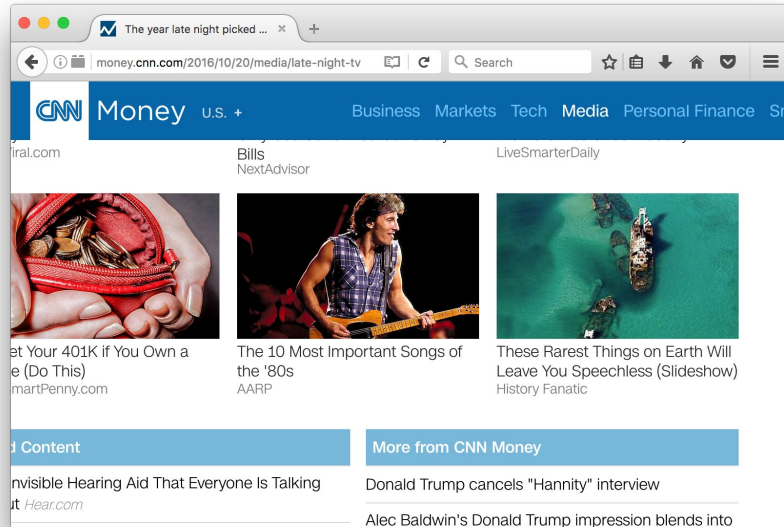
May 2014: 5% of sites *The Web Never Forgets (Acar, et al.)*

Aug 2014: ~0.1% of sites *(Approximate)*

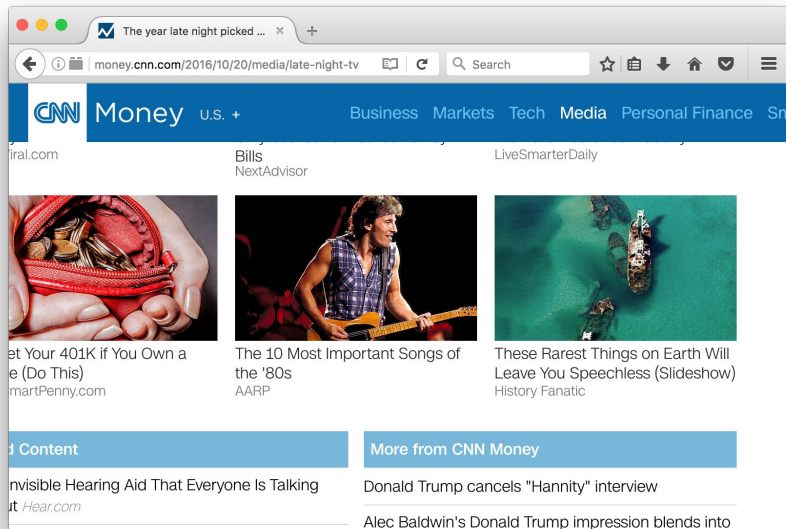
Jan 2016: 2.6% of sites

Percentage of the Alexa top 100k sites

Using Battery Status to Track



Using Battery Status to Track

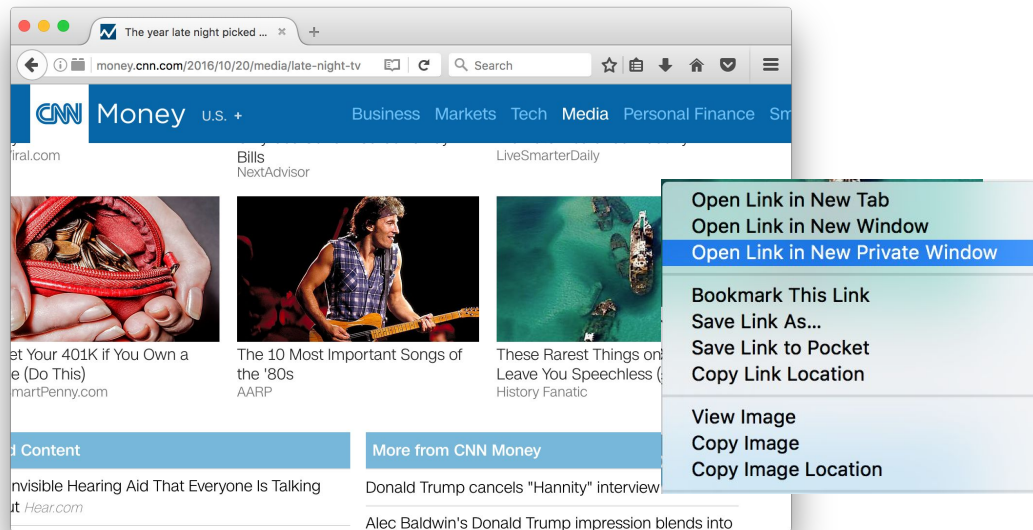


Battery Status:

level: 0.11

dischargeTime: 12867

Using Battery Status to Track

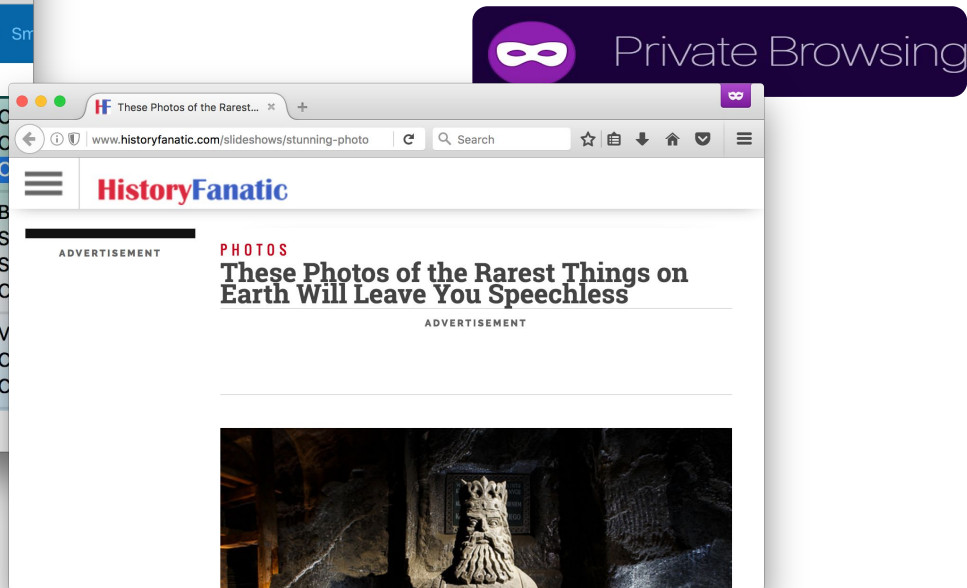
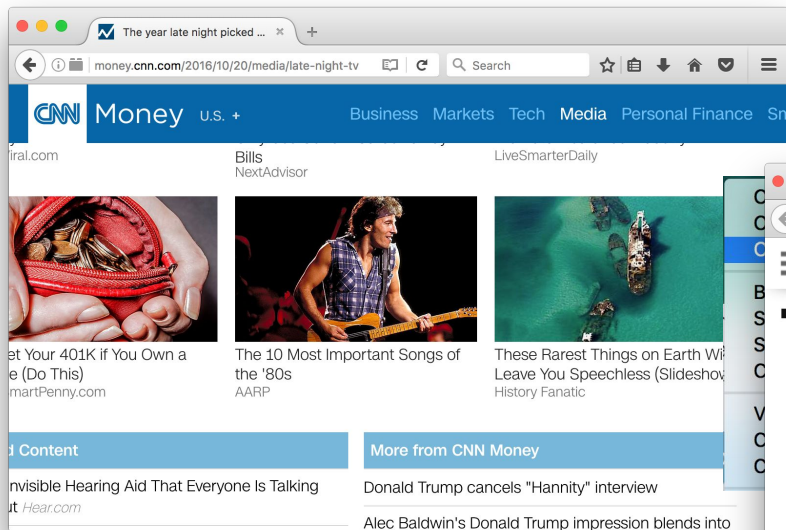


Battery Status:

level: 0.11

dischargeTime: 12867

Using Battery Status to Track

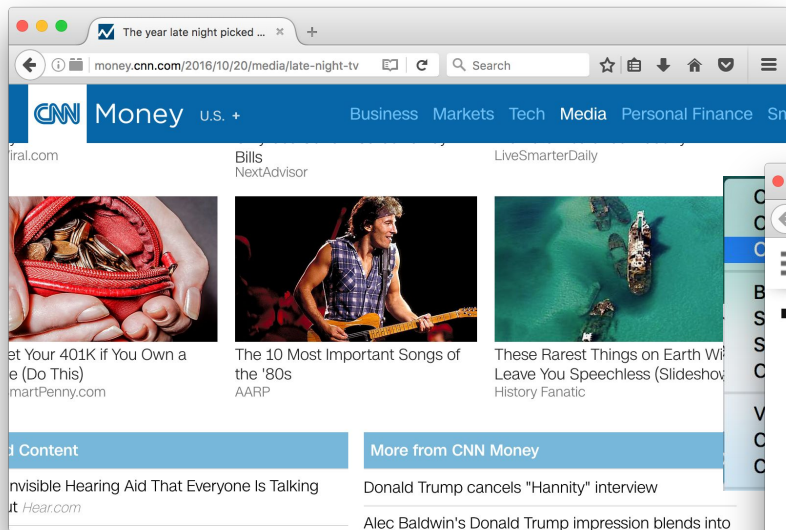


Battery Status:

level: 0.11

dischargeTime: 12867

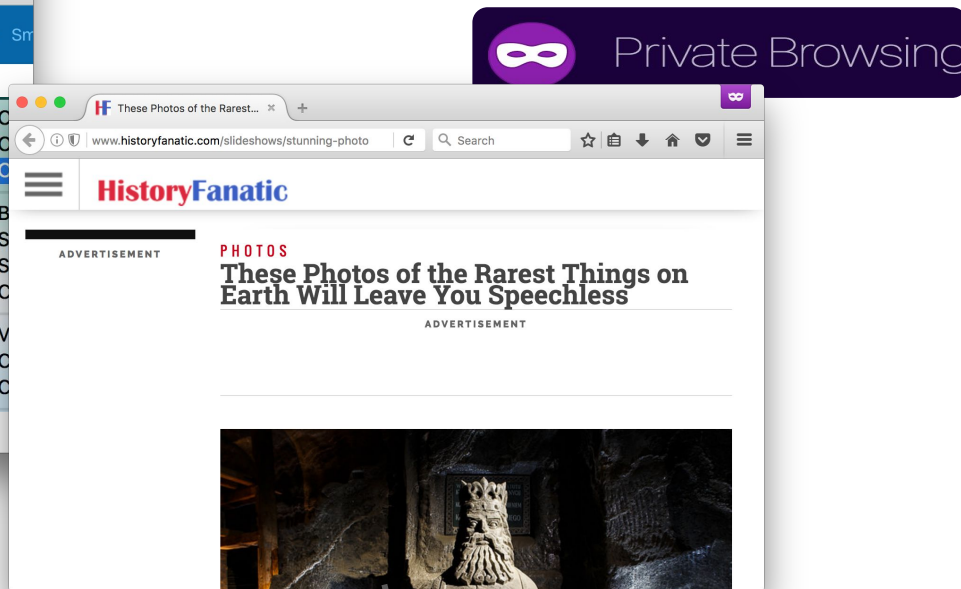
Using Battery Status to Track



Battery Status:

level: 0.11

dischargeTime: 12867

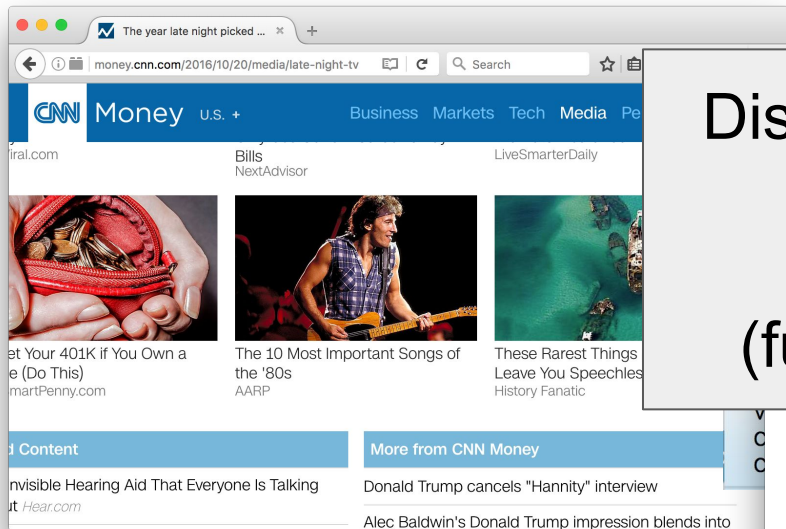


Battery Status:

level: 0.11

dischargeTime: 12867

Using Battery Status to Track



Discovered manually in 2 scripts on
about 22 sites

(full measurement is future work)



Battery Status:

level: 0.11

dischargeTime: 12867

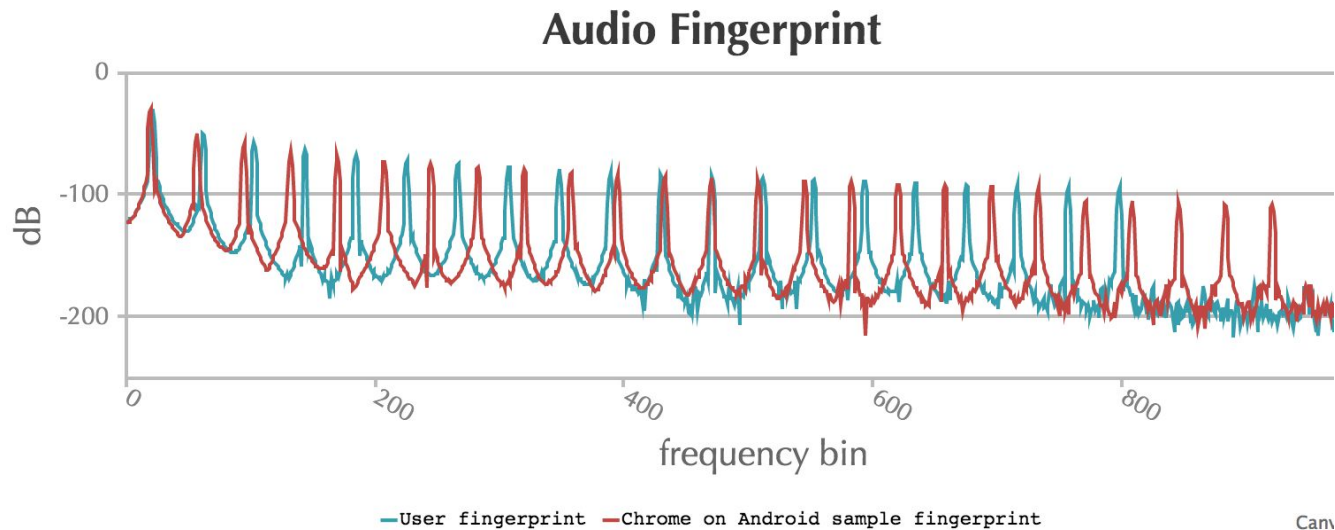


Battery Status:

level: 0.11

dischargeTime: 12867

Using AudioContext for fingerprinting



Live test page: <https://audiofingerprint.openwpm.com/>

Browsers remove BatteryStatus API citing privacy

The image shows two overlapping browser windows. The background window is Mozilla Bugzilla, displaying Bug 1313580: "Remove web content access...". The foreground window is WebKit Bugzilla, displaying Bug 164213: "Remove Battery Status API from the tree".

Bug 1313580 - Remove web content access...

- Status:** VERIFIED FIXED
- Whiteboard:**
- Keywords:** addon-compat, dev-doc-needed, privacy, site-compat
- Product:** Core ([show info](#))
- Component:** DOM: Device Interfaces ([show other bugs](#)) ([show info](#))
- Version:** unspecified
- Platform:** Unspecified Unspecified
- Importance:** -- normal ([vote](#))
- Target Milestone:** mozilla52
- Assigned To:** Chris Peterson [:cpeterson]

Bug 164213 - Remove Battery Status API from the tree

- Status:** RESOLVED FIXED
- Product:** WebKit
- Component:** WebKit Misc.
- Version:** WebKit Nightly Build
- Platform:** Unspecified Unspecified
- Importance:** P2 Normal
- Assigned To:** Alex Christensen
- URL:**
- Keywords:**
- Depends on:**
- Blocks:** [Show dependency tree / graph](#)
- Reported:** 2016-10-30 20:26 PDT by Brady Eidson
- Modified:** 2016-11-02 14:32 PDT ([History](#))
- CC List:** 8 users ([show](#))
- See Also:** [129040](#)

Google settlement for subverting cookie blocking

www.zdnet.com/article/google-pays-17m-to-settle-safari-cookie-privacy-bypass-charge/

EDITION: ▼



SEARCH



WINDOWS 10

CLOUD

INNOVATION

SECURITY

DATA CENTERS

MORE ▼

NEWSLETTER

Google pays \$17m to settle Safari cookie privacy-bypass charge

Settlement ends a two-year investigation into Google's cookie practice



By [Liam Tung](#) | November 19, 2013 -- 10:03 GMT (02:03 PST) | Topic: [Google](#)

Google will pay \$17m to settle claims by dozens of US states that it bypassed privacy settings in Apple's Safari browser designed to block third-party ad cookies.

[The deal](#) with 37 states and the District of Columbia prevents Google from installing

READ THIS



RELATED STORIES

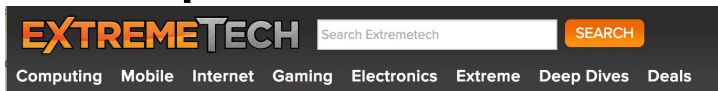


Mobility
Android 7.0 Nougat, M600, and Samsung Note 7 (MobileTechRoundup #379)



Mobility
Hands-on: Tech21 E

Multiple settlements for subverting cookie clearing



HOME > INTERNET > AOL, Spotify, C

AOL, Spotify, C undeletable tra

By Sebastian Anthony on August

12
13
14 JOHN B. KIM, and DAN C. SCHUTZMAN,
15 Individually, on Behalf of Themselves and All
16 Others Similarly Situated,
17 Plaintiffs,
18 v.
19 SPACE PENCIL, INC. DBA KISSMETRICS,
20 BAYPERS.COM, INVULVER.COM, MOO,
21 INC., SITTING, LLC, SHODAZZLE.COM,
22 INC., TRACKS INC., ABOUT ME,
23 FRIENDLY ORCA OMNI MEDIA INC.,
24 HASOFFERS.COM, KONGREGATE INC.,
25 LIVEMEDIA INC., ROCKETTRAIL, LLC,
26 FITNESS KEEPER, INC., SHOMOGZ, INC.,
27 SHARKCASE, LLC, SLIPSHARE.NET,
28 SPOKEO INC., SPOTIFY USA, INC.,
29 VIRTUALLY CONDUIT USA, FLITE, INC.,
30

Anyone who has visited one
damages of up to \$10,000 p
this lawsuit could be worth t

RYAN SINGEL BUSINESS 12.05.10 2:02 AM

ONLINE TRACKING FIRM SETTLES SUIT OVER UNDELETABLE COOKIES

MediaPost

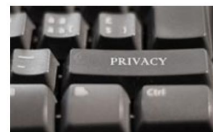
ONLINE MEDIA DAILY

Online tr
pay \$2.4
lawsuit a
ubiquito
tracking
the comp

More tha
go to fun
plaintiffs
filed the
will go to

KISSmetrics Finalizes Supercookies Settlement

by Wendy Davis @wendyndavis, January 18, 2013, 5:24 PM



Analytics company KISSmetrics has finalized the settlement of a class-action lawsuit stemming from its alleged use of "supercookies" to track people online.

The company implemented an agreement calling for it to refrain from using eTags, Flash cookies or other types of hard-to-delete supercookies without first notifying users and allowing them to choose whether to accept the technology, according to

The company also agreed to pay around \$500,000 to the attorneys who brought the case and \$2,500 each to the two consumers who sued: John Kim and Dan Schutzman.

Flash Cookies and Privacy (2009) Soltani, et al.

Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning (2011) Ayenson, et al.