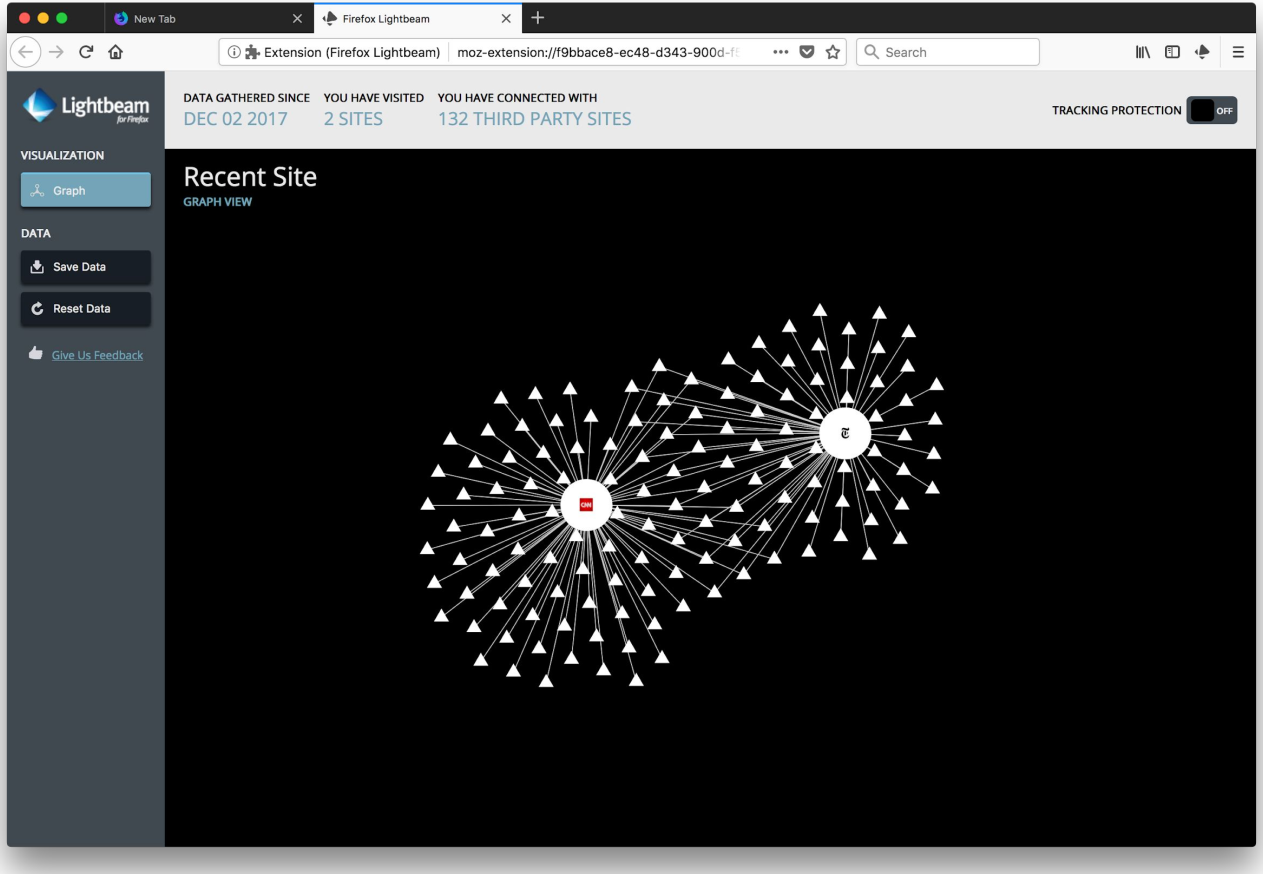
The necessity and challenges of default-on Tracking Protection

Steven Englehardt

moz://a



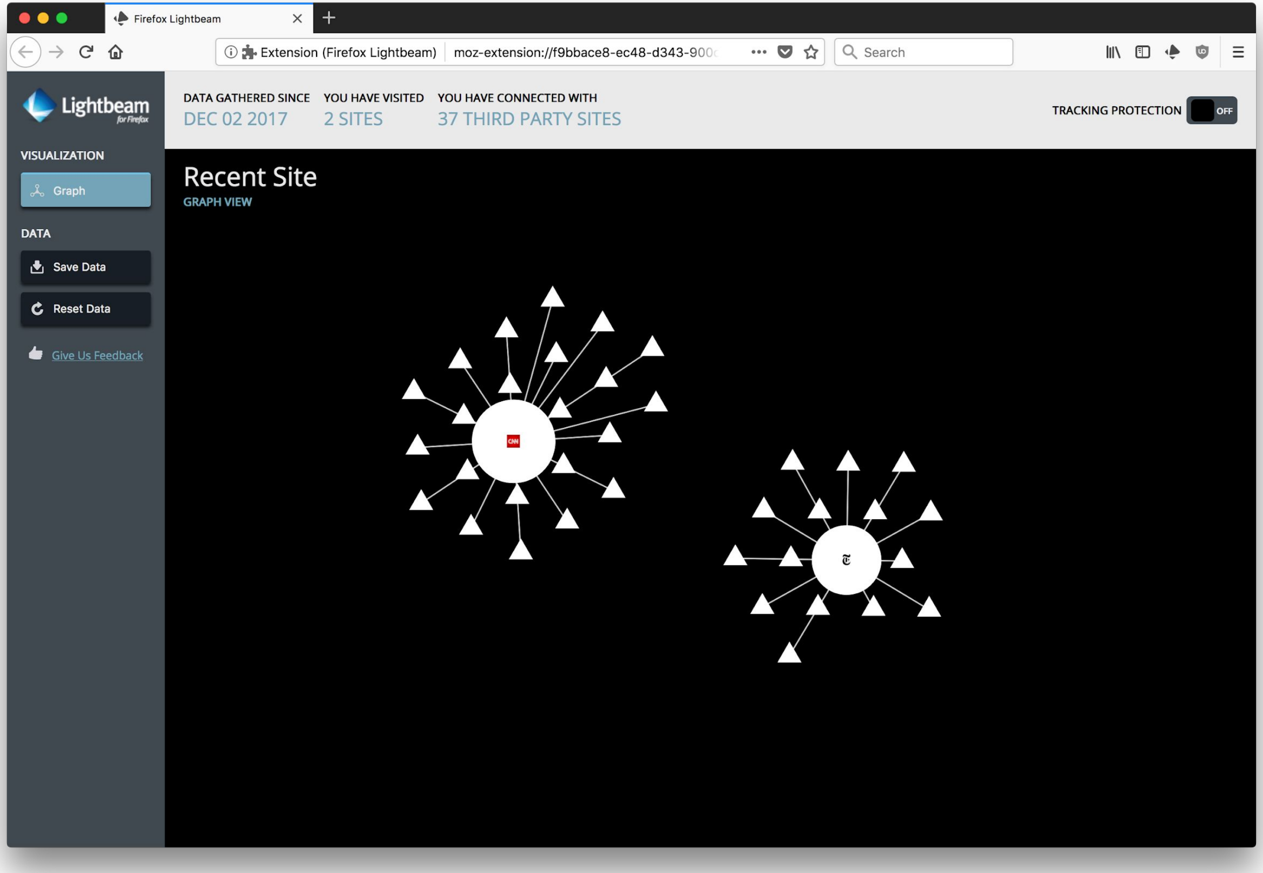
DATA GATHERED SINCE DEC 02 2017

YOU HAVE VISITED 2 SITES

YOU HAVE CONNECTED WITH 132 THIRD PARTY SITES

Recent Site

Just **two page visits** cause requests to **132 distinct hostnames**.



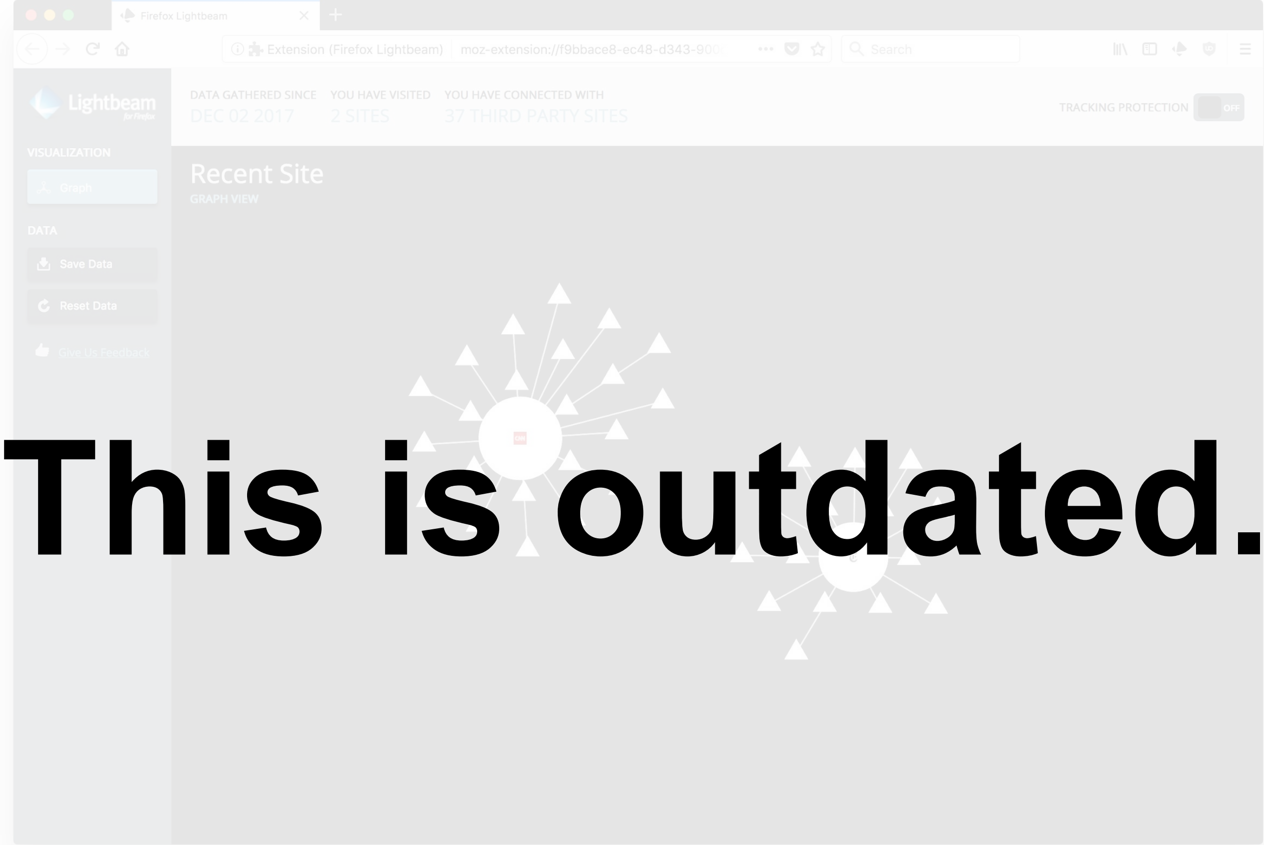
DATA GATHERED SINCE DEC 02 2017

YOU HAVE VISITED 2 SITES

YOU HAVE CONNECTED WITH 37 THIRD PARTY SITES

With uBlock Origin enabled, the number of hostnames requested is down to 37. **Nearly 100 of the hosts loaded were ads, trackers, and analytics.**

This is outdated.



DATA GATHERED SINCE DEC 02 2017

YOU HAVE VISITED 2 SITES

YOU HAVE CONNECTED WITH 37 THIRD PARTY SITES

TRACKING PROTECTION OFF

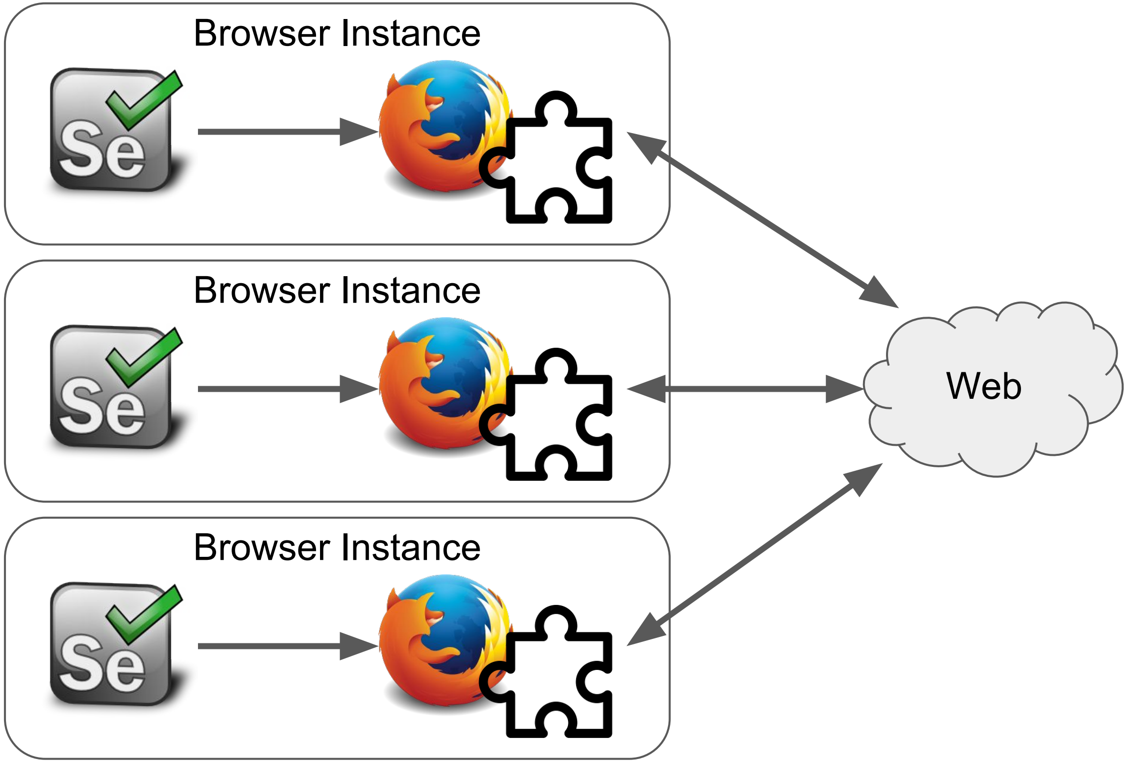
Recent Site

GRAPH VIEW

With uBlock Origin enabled, the number of hostnames requested is down to 37. **Nearly 100 of the hosts loaded were ads, trackers, and analytics.**

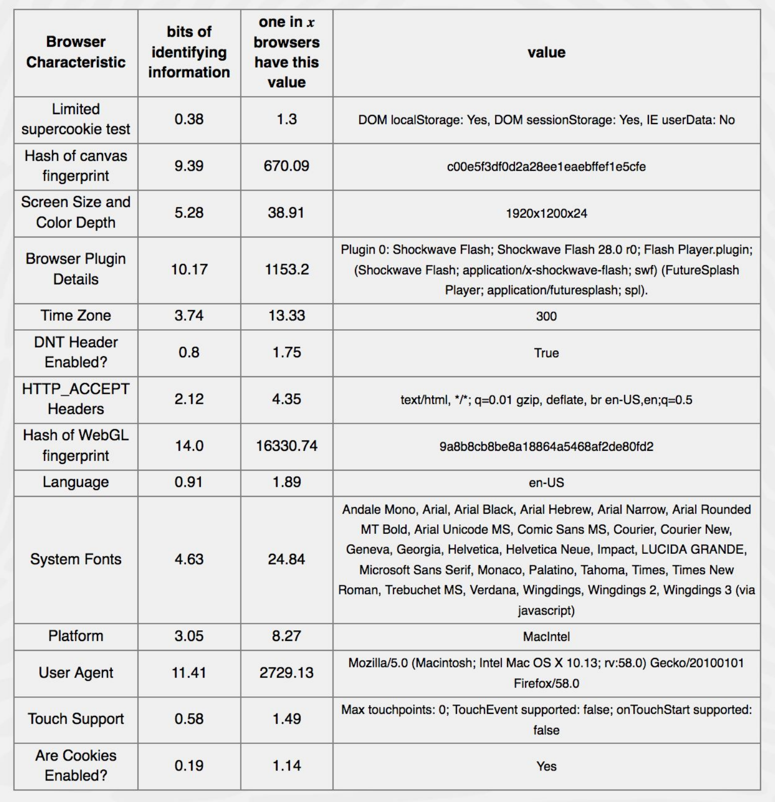
Users are exposed to a diverse set of threats every day

Automated measurement can be used to discover invasive tracking



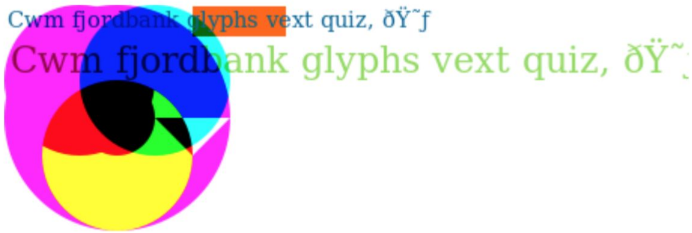
OpenWPM: An automated platform for web privacy measurement (https://github.com/citp/OpenWPM)

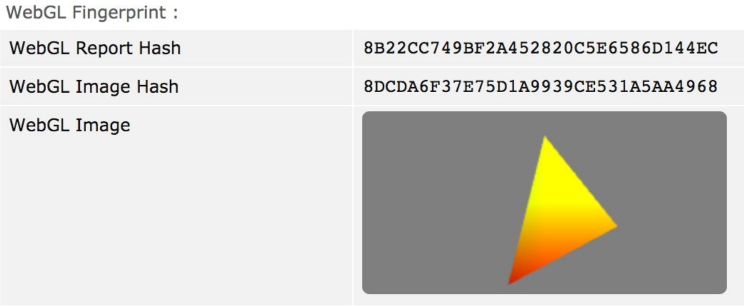
Device fingerprinting: tracking users without state



Browser Characteristic	bits of identifying information	one in *x* browsers have this value	value
Limited supercookie test	0.38	1.3	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No
Hash of canvas fingerprint	9.39	670.09	c00e5f3df0d2a28ee1eaebffef1e5cfe
Screen Size and Color Depth	5.28	38.91	1920x1200x24
Browser Plugin Details	10.17	1153.2	Plugin 0: Shockwave Flash; Shockwave Flash 28.0 r0; Flash Player.plugin; (Shockwave Flash; application/x-shockwave-flash; swf) (FutureSplash Player; application/futuresplash; spl).
Time Zone	3.74	13.33	300
DNT Header Enabled?	0.8	1.75	True
HTTP_ACCEPT Headers	2.12	4.35	text/html, */*; q=0.01 gzip, deflate, br en-US,en;q=0.5
Hash of WebGL fingerprint	14.0	16330.74	9a8b8cb8be8a18864a5468af2de80fd2
Language	0.91	1.89	en-US
System Fonts	4.63	24.84	Andale Mono, Arial, Arial Black, Arial Hebrew, Arial Narrow, Arial Rounded MT Bold, Arial Unicode MS, Comic Sans MS, Courier, Courier New, Geneva, Georgia, Helvetica, Helvetica Neue, Impact, LUCIDA GRANDE, Microsoft Sans Serif, Monaco, Palatino, Tahoma, Times, Times New Roman, Trebuchet MS, Verdana, Wingdings, Wingdings 2, Wingdings 3 (via javascript)
Platform	3.05	8.27	MacIntel
User Agent	11.41	2729.13	Mozilla/5.0 (Macintosh; Intel Mac OS X 10.13; rv:58.0) Gecko/20100101 Firefox/58.0
Touch Support	0.58	1.49	Max touchpoints: 0; TouchEvent supported: false; onTouchStart supported: false
Are Cookies Enabled?	0.19	1.14	Yes

https://panopticlick.eff.org


Cwm fjordbank glyphs vext quiz, ðŸ˜ƒ
Cwm fjordbank glyphs vext quiz, ðŸ˜

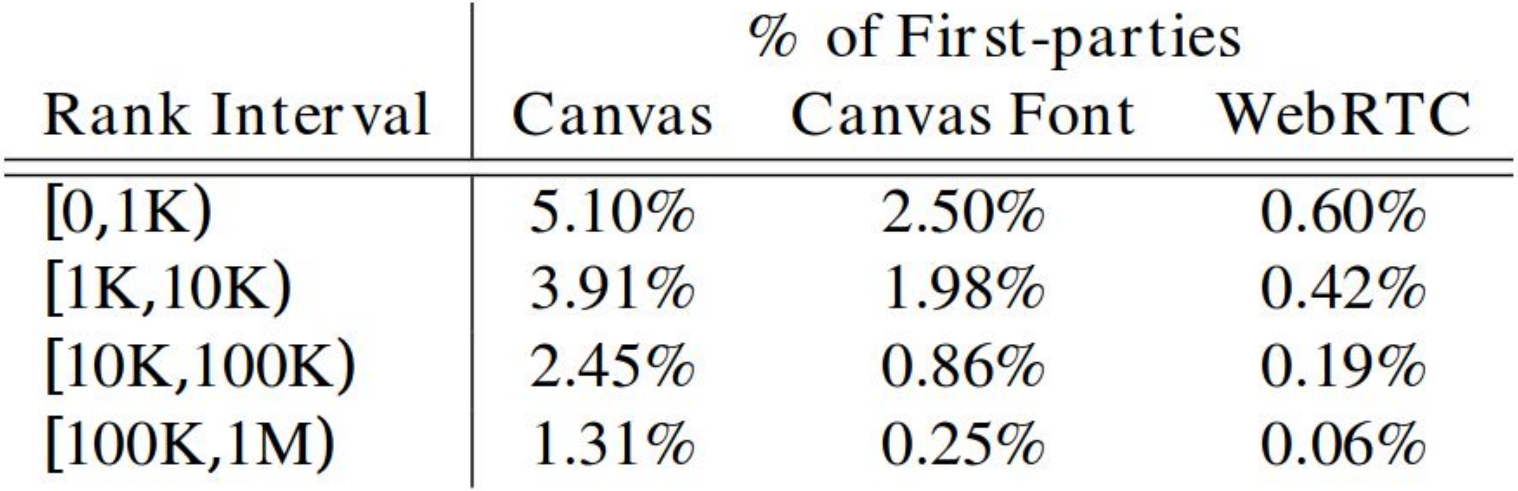


WebGL Fingerprint :

WebGL Report Hash	8B22CC749BF2A452820C5E6586D144EC
WebGL Image Hash	8DCDA6F37E75D1A9939CE531A5AA4968
WebGL Image	

https://browserleaks.com/webgl

Fingerprinting is quite prevalent



	% of First-parties		
Rank Interval	Canvas	Canvas Font	WebRTC
[0,1K)	5.10%	2.50%	0.60%
[1K,10K)	3.91%	1.98%	0.42%
[10K,100K)	2.45%	0.86%	0.19%
[100K,1M)	1.31%	0.25%	0.06%

Englehardt and Narayanan, "Online Tracking: A 1-million-site Measurement and Analysis" (CCS 2016)

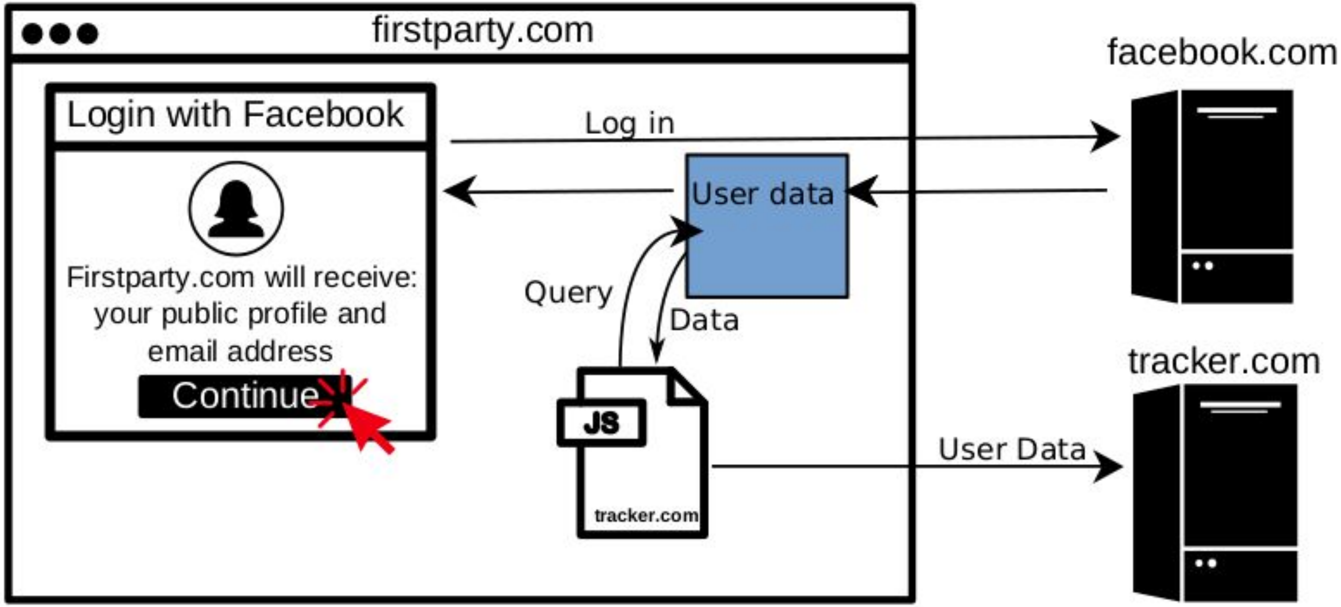
Why identify the device when you can identify the user?

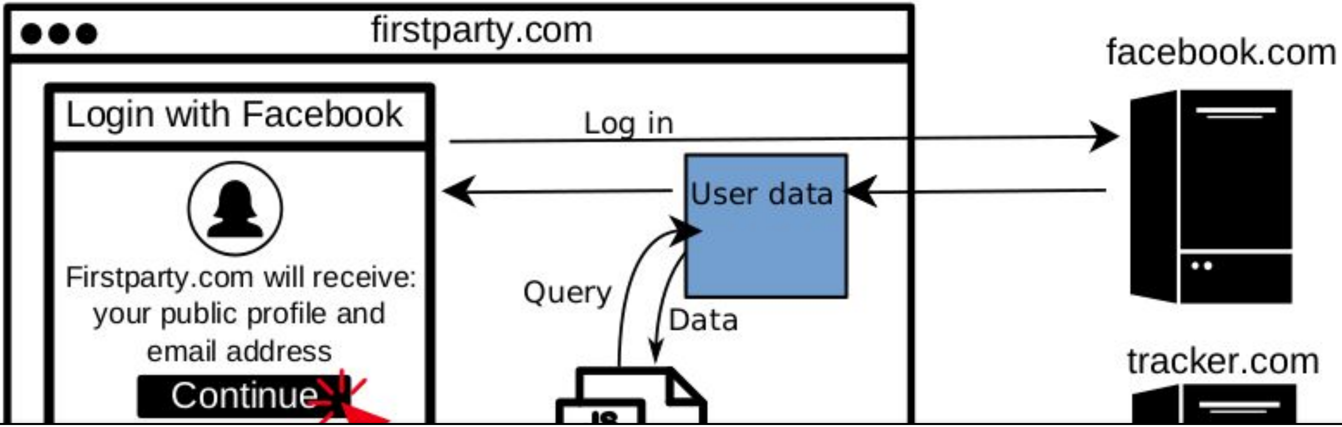
md5(ste@cs.princeton.edu) → b5184f3fb0fe35e4319b729f05017f6e

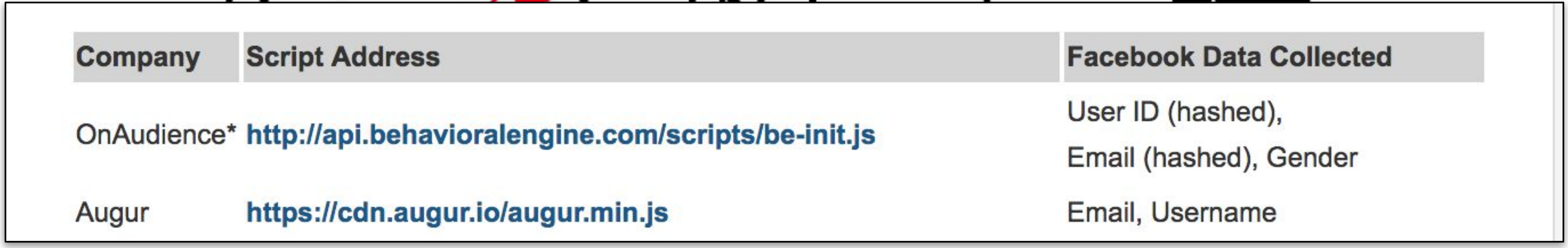
The abuse of social login for web tracking

Login with Facebook

Firstparty.com will receive: your public profile and email address


firstparty.com
Login with Facebook
Firstparty.com will receive: your public profile and email address
Continue
Log in
User data
Query
Data
JS
tracker.com
User Data
facebook.com
tracker.com

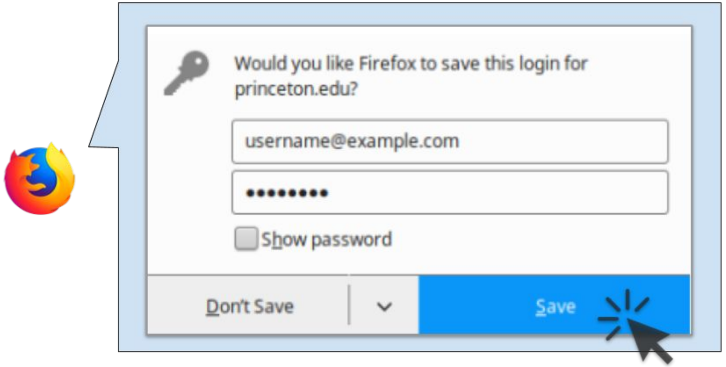




Company	Script Address	Facebook Data Collected
OnAudience*	http://api.behavioralengine.com/scripts/be-init.js	User ID (hashed), Email (hashed), Gender
Augur	https://cdn.augur.io/augur.min.js	Email, Username

Englehardt, Acar, and Narayanan, “No boundaries for Facebook data: third-party trackers abuse Facebook Login” (Freedom to Tinker)

Login manager abuse for web tracking



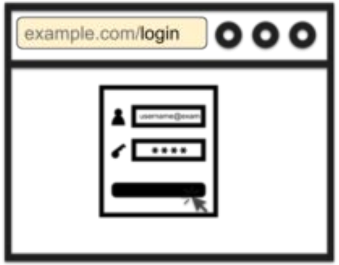
Would you like Firefox to save this login for princeton.edu?

username@example.com

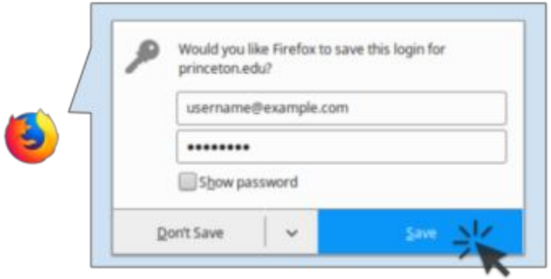
•••••••••

□ Show password

User submits a login or registration form, clicks “Save” to store the credentials.



Third-party script **is not present** on the login page



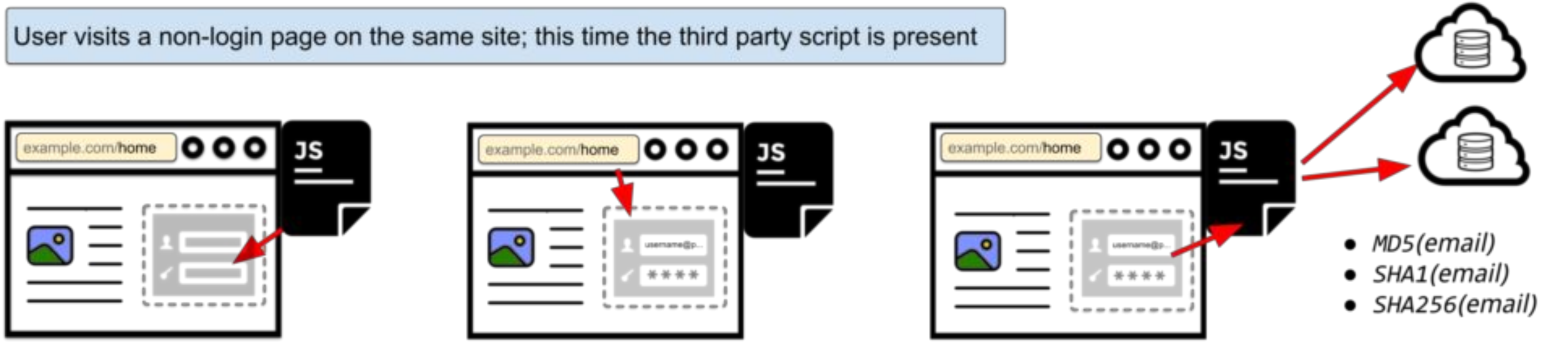
Would you like Firefox to save this login for princeton.edu?

username@example.com

****

□ Show password

User visits a non-login page on the same site; this time the third party script is present



- MD5(email)
- SHA1(email)
- SHA256(email)

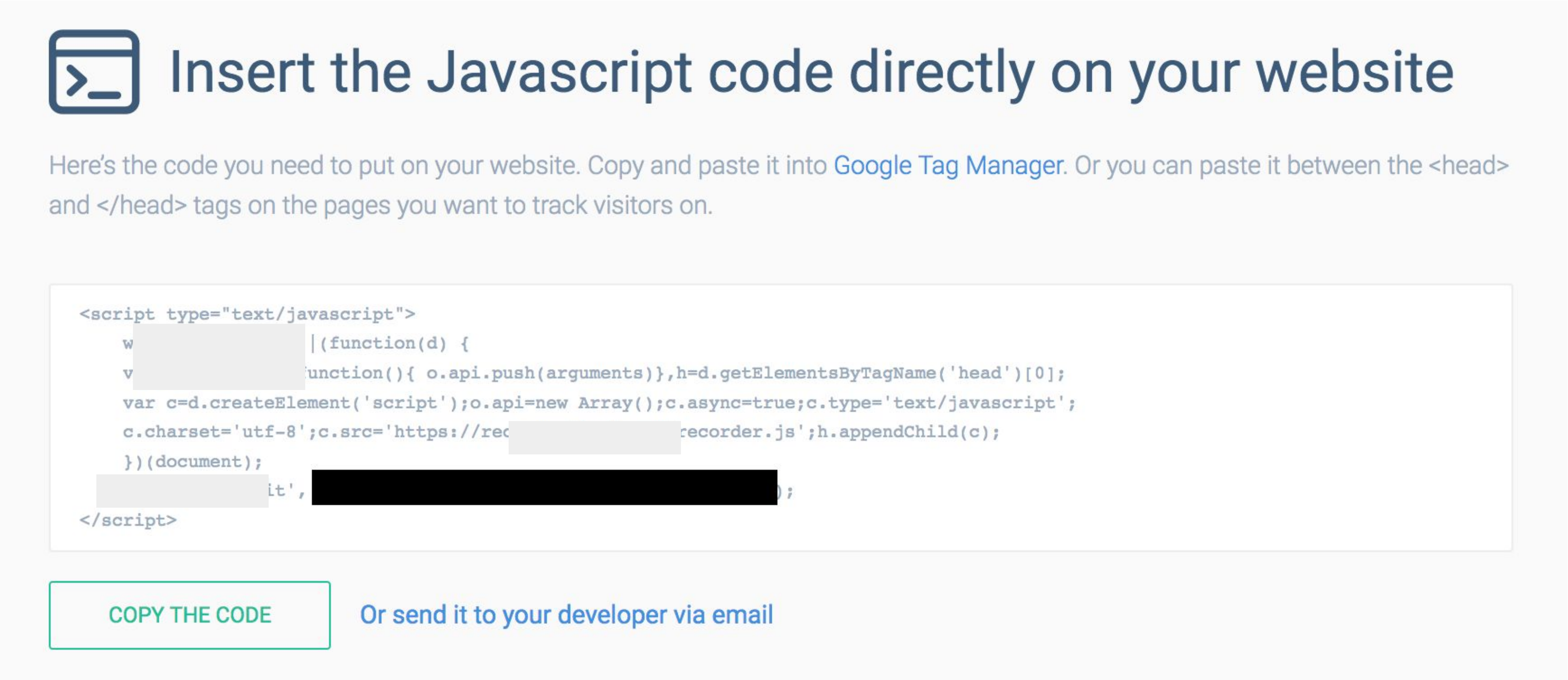
1. Third-party script injects an invisible login form

2. Login manager fills in user's email and password

3. The script reads the email address from the form and sends it hashes to third-party servers

Acar, Englehardt and Narayanan, “No boundaries for user identities: Web trackers exploit browser login managers” (Freedom to Tinker)

The core problem: no security boundaries



Insert the Javascript code directly on your website

Here's the code you need to put on your website. Copy and paste it into Google Tag Manager. Or you can paste it between the <head> and </head> tags on the pages you want to track visitors on.

```
<script type="text/javascript">
    w[redacted]|(function(d) {
    v[redacted]unction(){ o.api.push(arguments)},h=d.getElementsByTagName('head')[0];
    var c=d.createElement('script');o.api=new Array();c.async=true;c.type='text/javascript';
    c.charset='utf-8';c.src='https://rec[redacted]recorder.js';h.appendChild(c);
    })(document);
  [redacted]it',[redacted]);
</script>
```

What can we do?

Build better ad / tracking blockers?

Build better ad / tracking blockers?

+ Great for technical users

- Defunds publishers

- Breaks websites

- Creates no incentive for change

User privacy has long been opt-in on the web.

User privacy has long been opt-in on the web.

It shouldn’t be.

FIREFOX

Changing Our Approach to Anti-tracking

Nick Nguyen | August 30, 2018

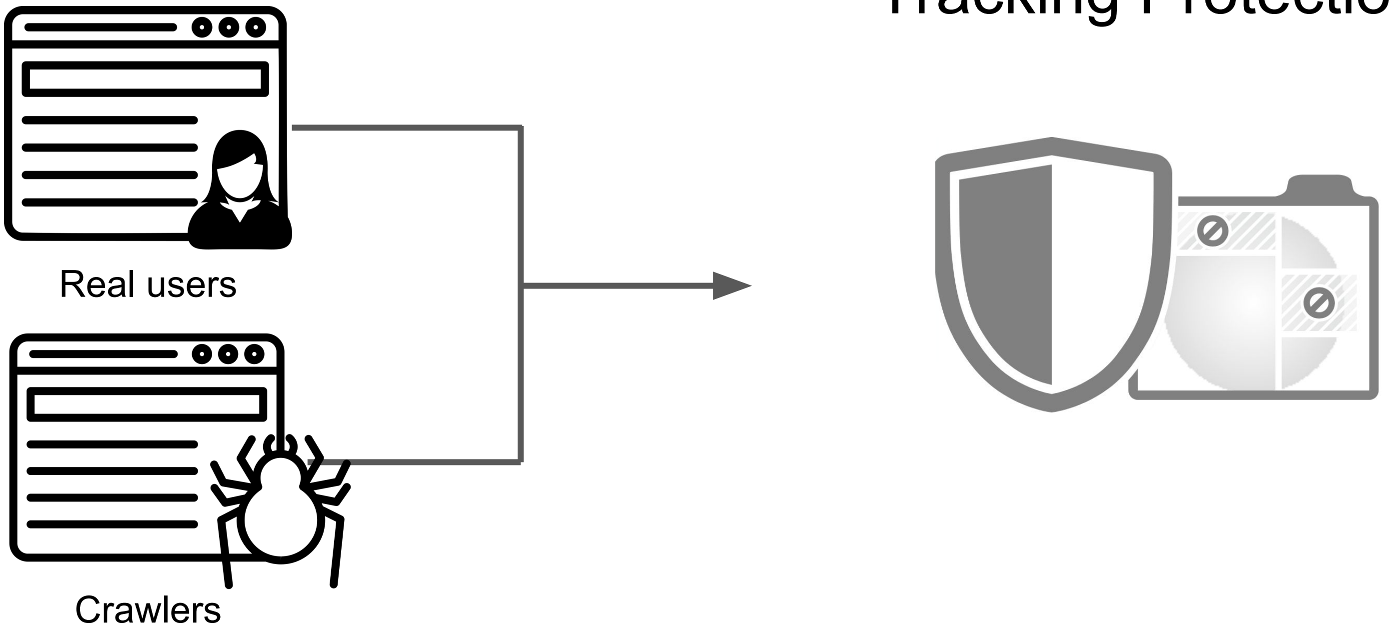
Anyone who isn't an expert on the internet would be hard-pressed to explain how tracking on the internet actually works.

https://blog.mozilla.org/futurereleases/2018/08/30/changing-our-approach-to-anti-tracking/

A possible solution: Apply stronger restrictions to bad actors

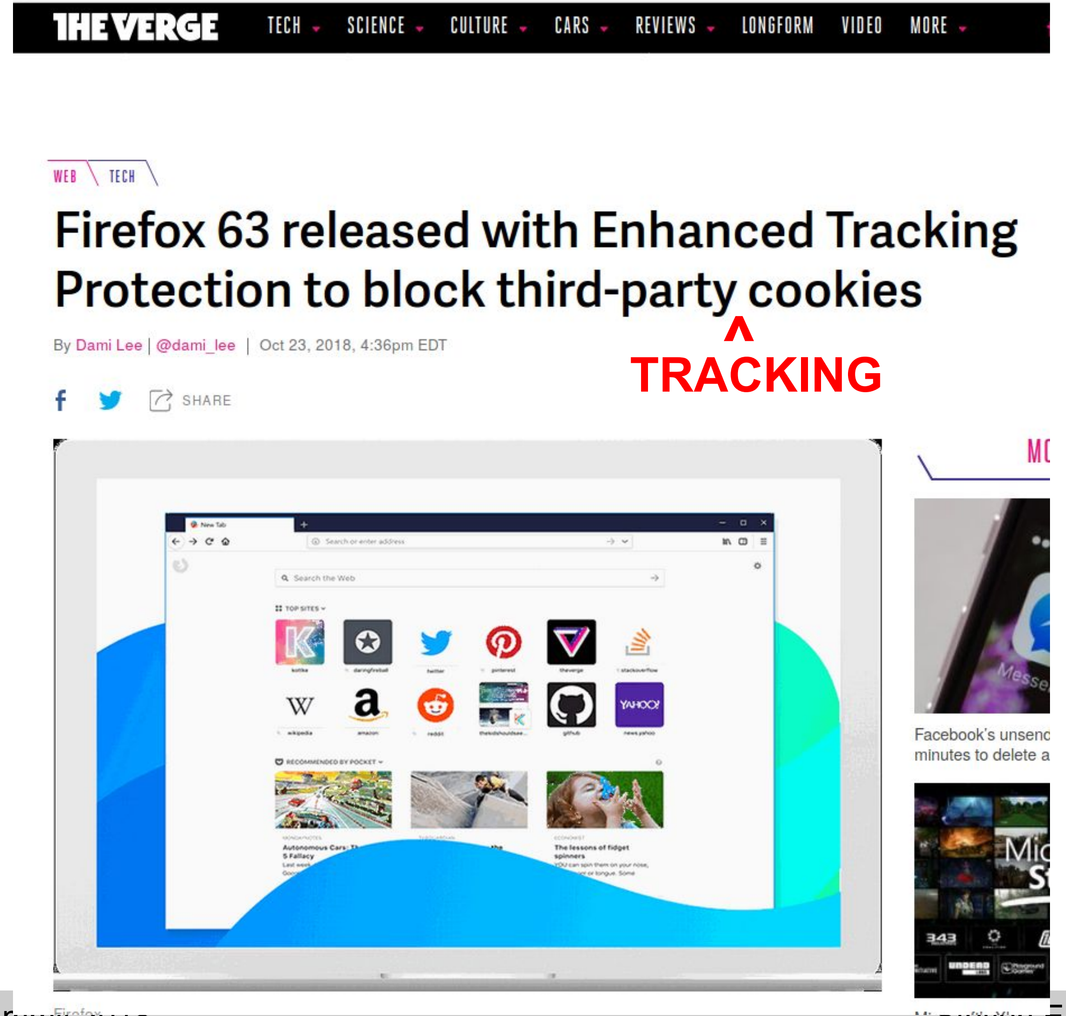
Detect invasive scripts

Selective, default-on Tracking Protection



Real users

Crawlers



WEB \ TECH

Firefox 63 released with Enhanced Tracking Protection to block third-party cookies

^
TRACKING

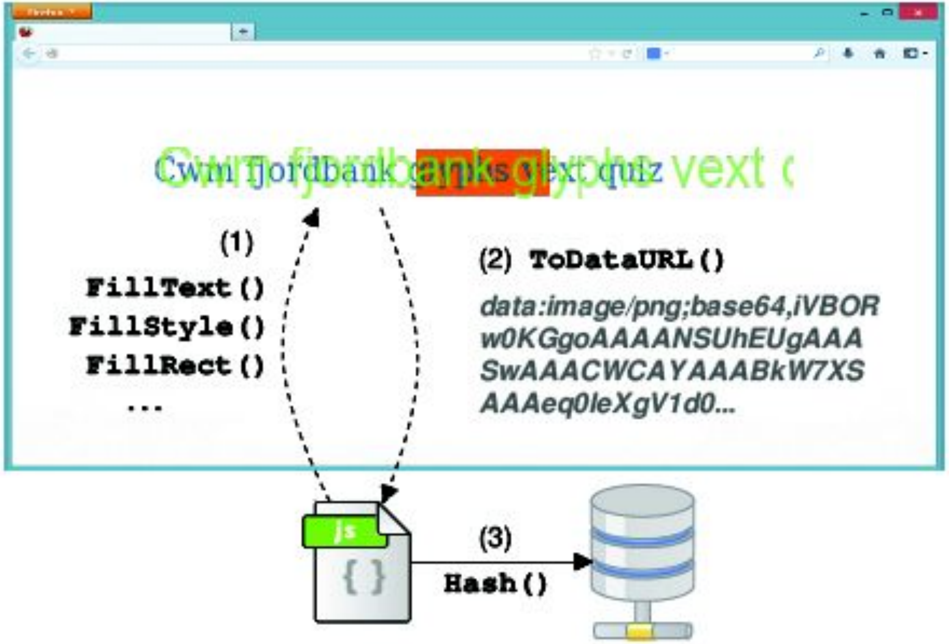
By Dami Lee | @dami_lee | Oct 23, 2018, 4:36pm EDT

Facebook's unsend
minutes to delete a

Challenge: Simple heuristics aren’t good enough

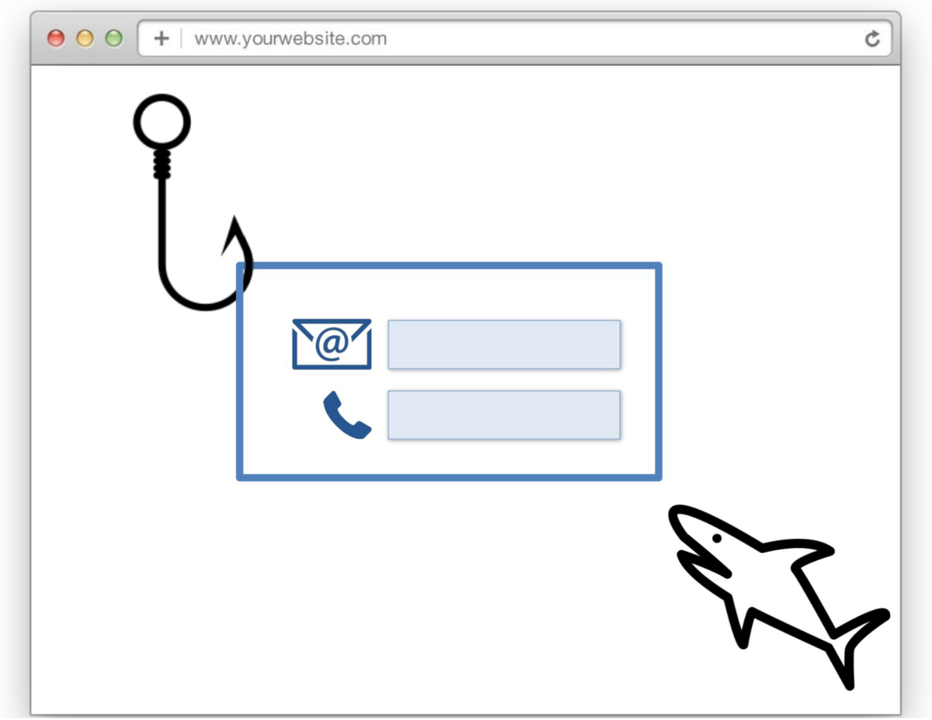
Detection Methodology:

1. Canvas height and width >= 16px
2. Text >= 2 colors OR >= 10 characters
3. Should not call `save`, `restore`, or `addEventListener`. (Used with interactive or animated content)
4. Calls `toDataURL` or `getImageData`.



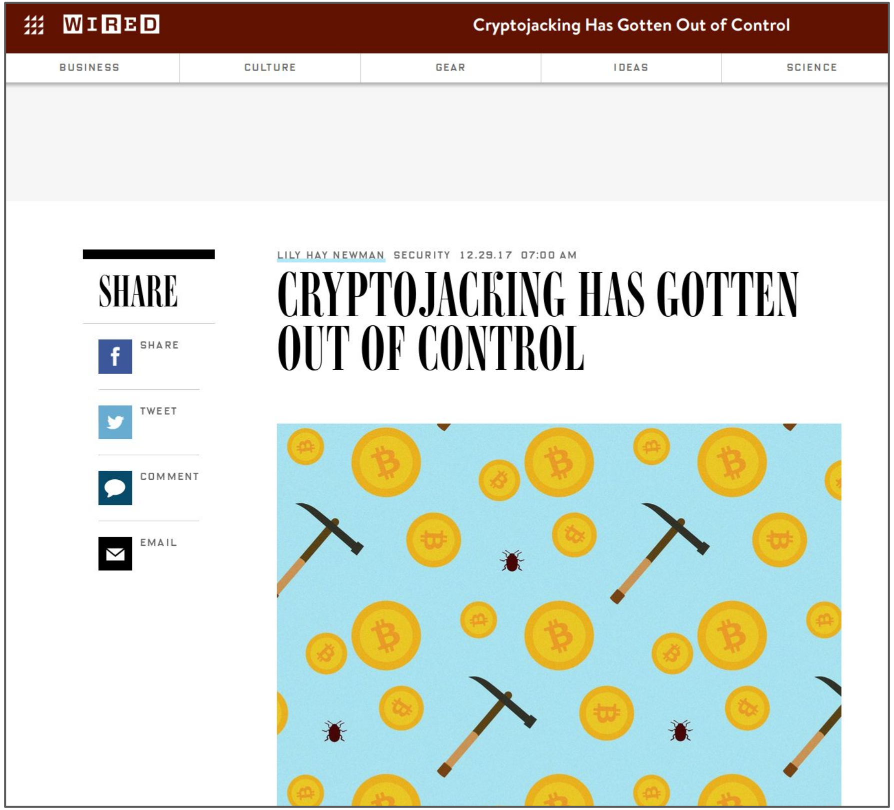
Englehardt and Narayanan, “Online Tracking: A 1-million-site Measurement and Analysis” (CCS 2016)

Challenge: Simple network monitoring isn’t good enough



Search network traffic to find leaked PII

Challenge: Tracking isn't the only threat



WIRED — Cryptojacking Has Gotten Out of Control

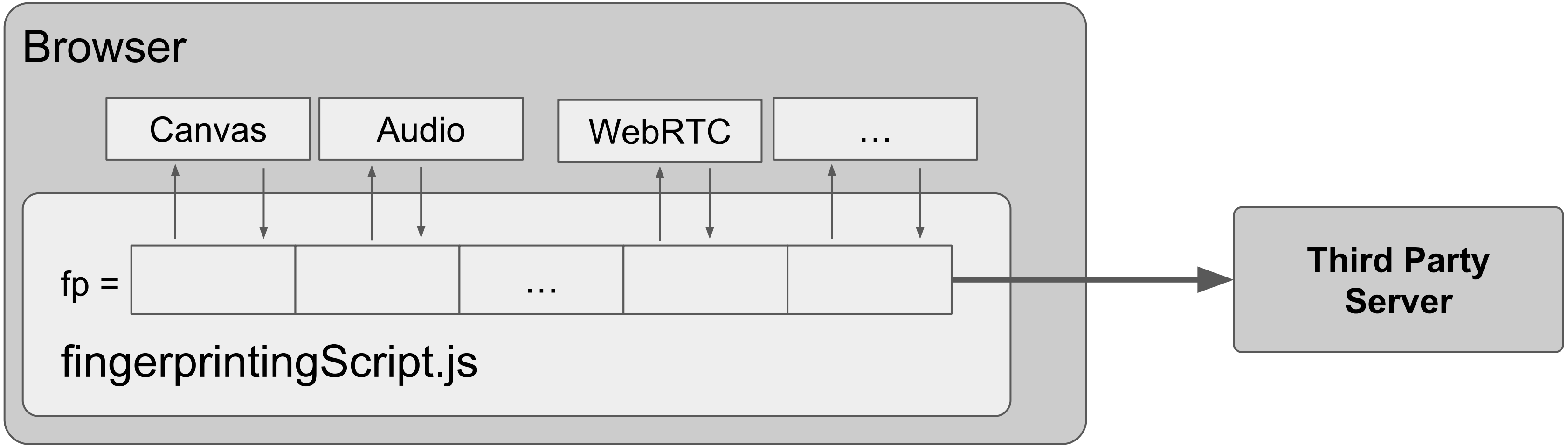
BUSINESS | CULTURE | GEAR | IDEAS | SCIENCE

LILY HAY NEWMAN SECURITY 12.29.17 07:00 AM

CRYPTOJACKING HAS GOTTEN OUT OF CONTROL

Possible solution for fingerprinting:

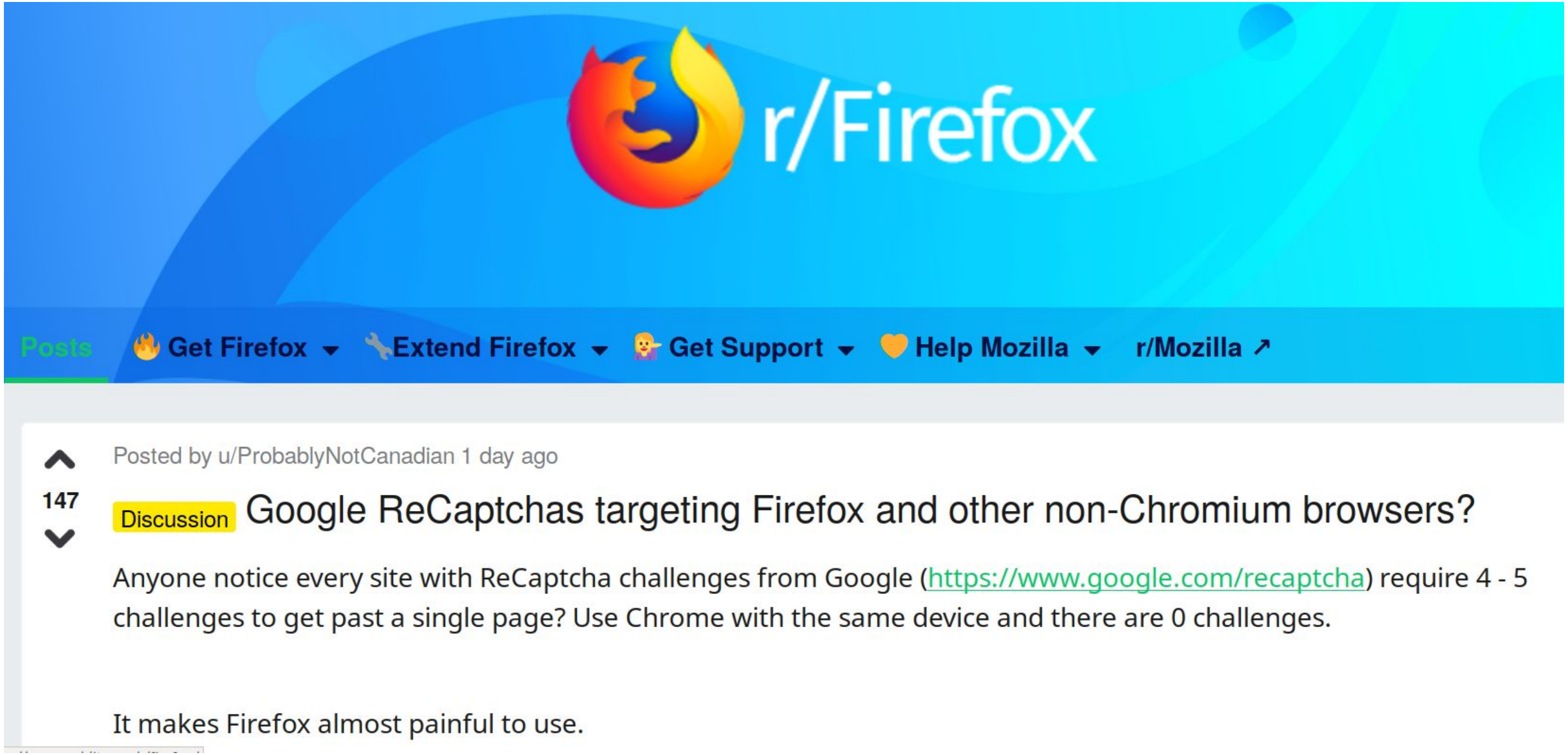
Classify fingerprinters on full API use



Challenge: Tracking techniques are dual-use

- Tracking is used for profile building, bot detection, authentication
- Most major US banks fingerprint visitors
- Captchas fingerprint users to detect bots

Challenge: Tracking techniques are dual-use



Posted by u/ProbablyNotCanadian 1 day ago

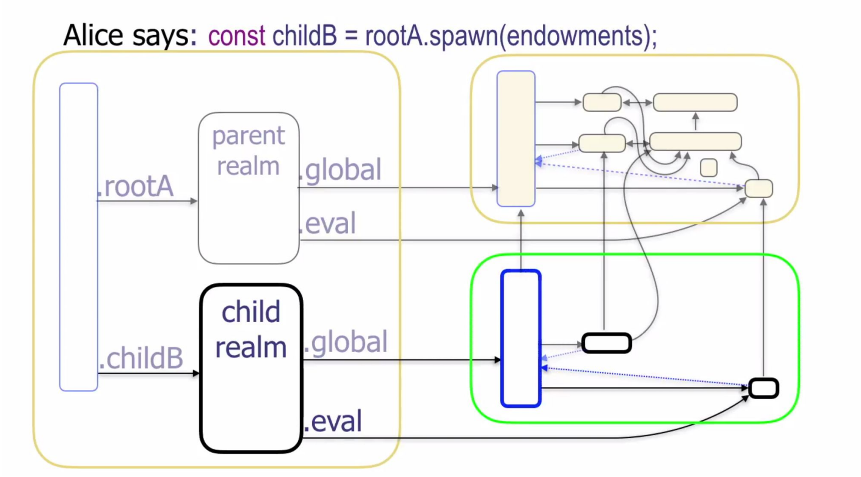
Discussion Google ReCaptchas targeting Firefox and other non-Chromium browsers?

Anyone notice every site with ReCaptcha challenges from Google (https://www.google.com/recaptcha) require 4 - 5 challenges to get past a single page? Use Chrome with the same device and there are 0 challenges.

It makes Firefox almost painful to use.

Challenge: High complexity and cost of confinement

Frozen Realms



(https://github.com/tc39/proposal-frozen-realms)

COWL

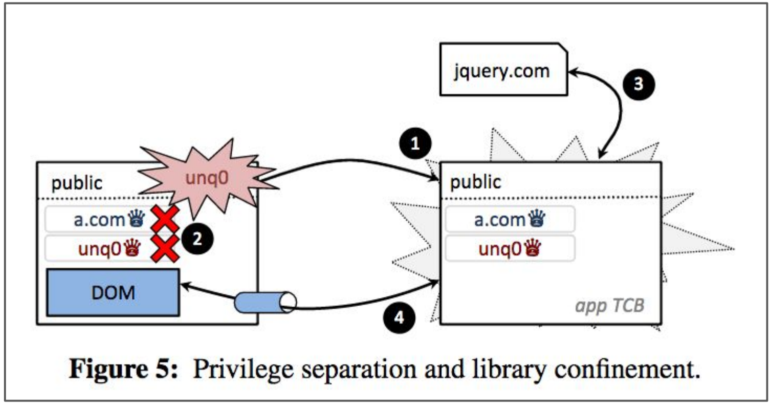
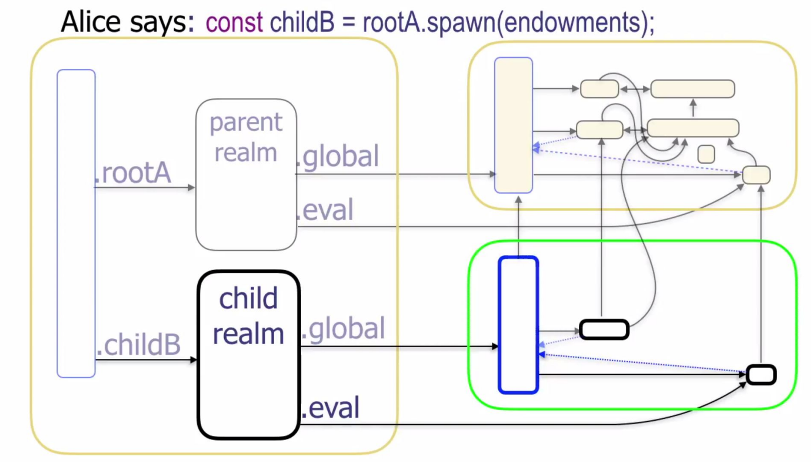


Figure 5: Privilege separation and library confinement.

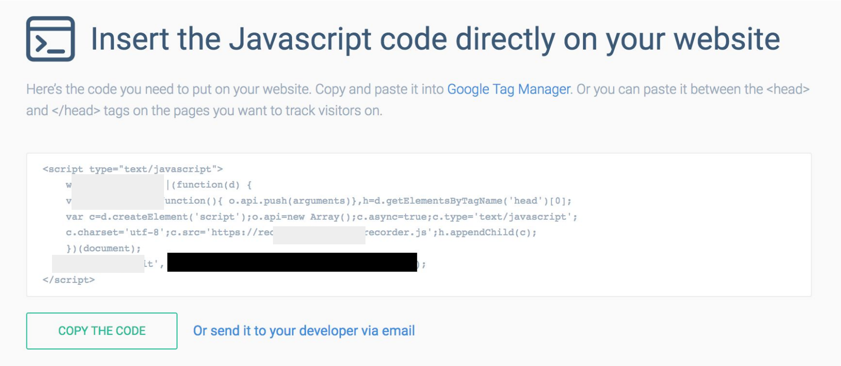
(https://www.usenix.org/node/186158)

Possible solution: Confinement that can be enforced by the browser

Remember:



VS



Insert the Javascript code directly on your website

Here's the code you need to put on your website. Copy and paste it into Google Tag Manager. Or you can paste it between the <head> and </head> tags on the pages you want to track visitors on.

```
<script type="text/javascript">
    w[redacted]|(function(d) {
    v[redacted]unction(){ o.api.push(arguments)},h=d.getElementsByTagName('head')[0];
    var c=d.createElement('script');o.api=new Array();c.async=true;c.type='text/javascript';
    c.charset='utf-8';c.src='https://rec[redacted]recorder.js';h.appendChild(c);
    })(document);
  [redacted]it', [redacted]);
</script>
```

In summary: Firefox is building default-on tracking protection that doesn't break sites. Our challenges:

1. **How can we make fingerprinting and PII exfiltration detection more thorough and robust to adversaries?**
2. **What automated measurement techniques can we use to detect cryptomining?**
3. **Are there alternatives to bot detection, captchas, authentication applications that don't require cross-site tracking?**
4. **Can we automatically apply JS confinement without first-party support?**

Me: https://senglehardt.com | senglehardt@mozilla.com