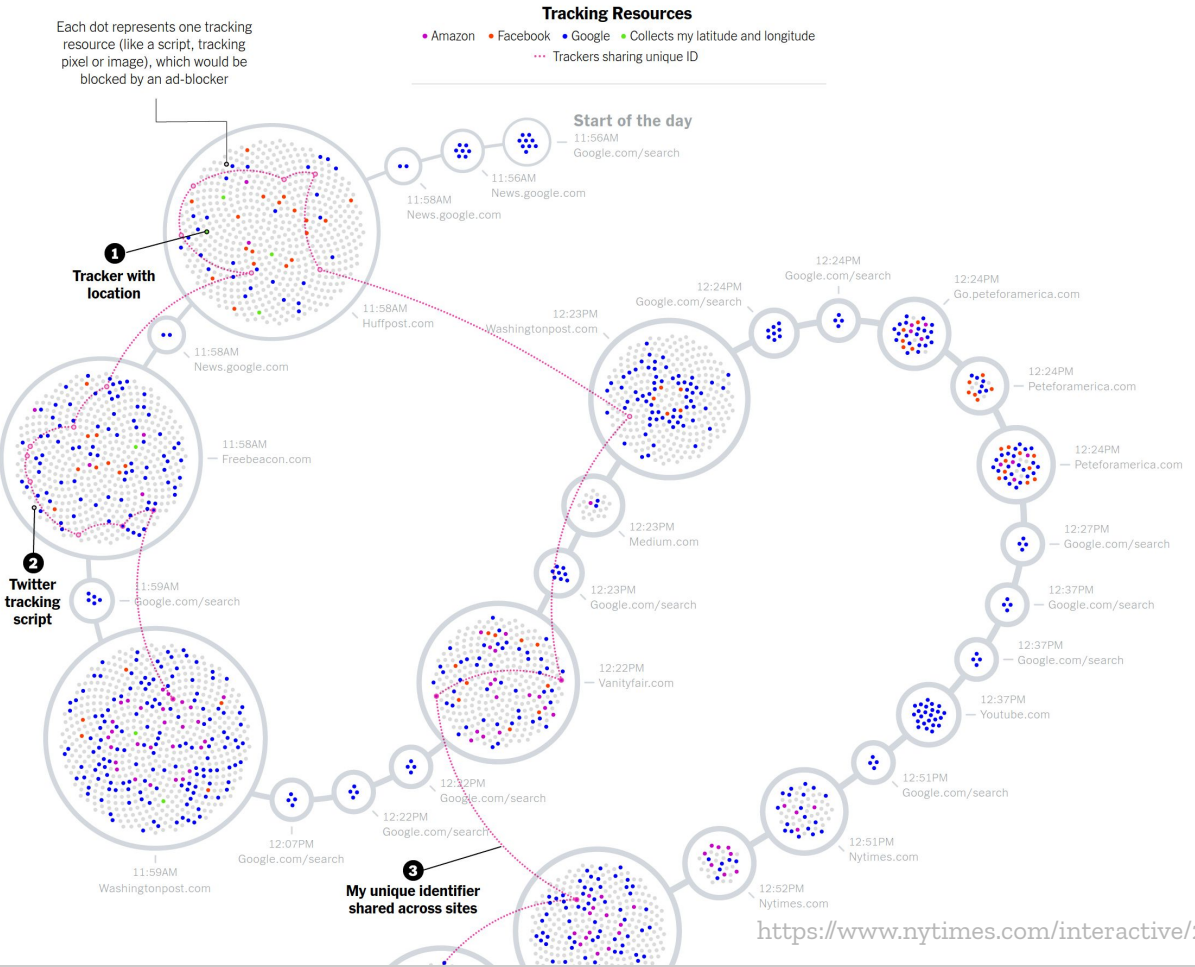
Challenges in building a private web

(without burning it all down)

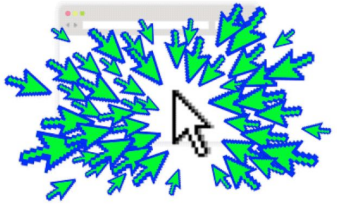
Steven Englehardt

Privacy Engineer

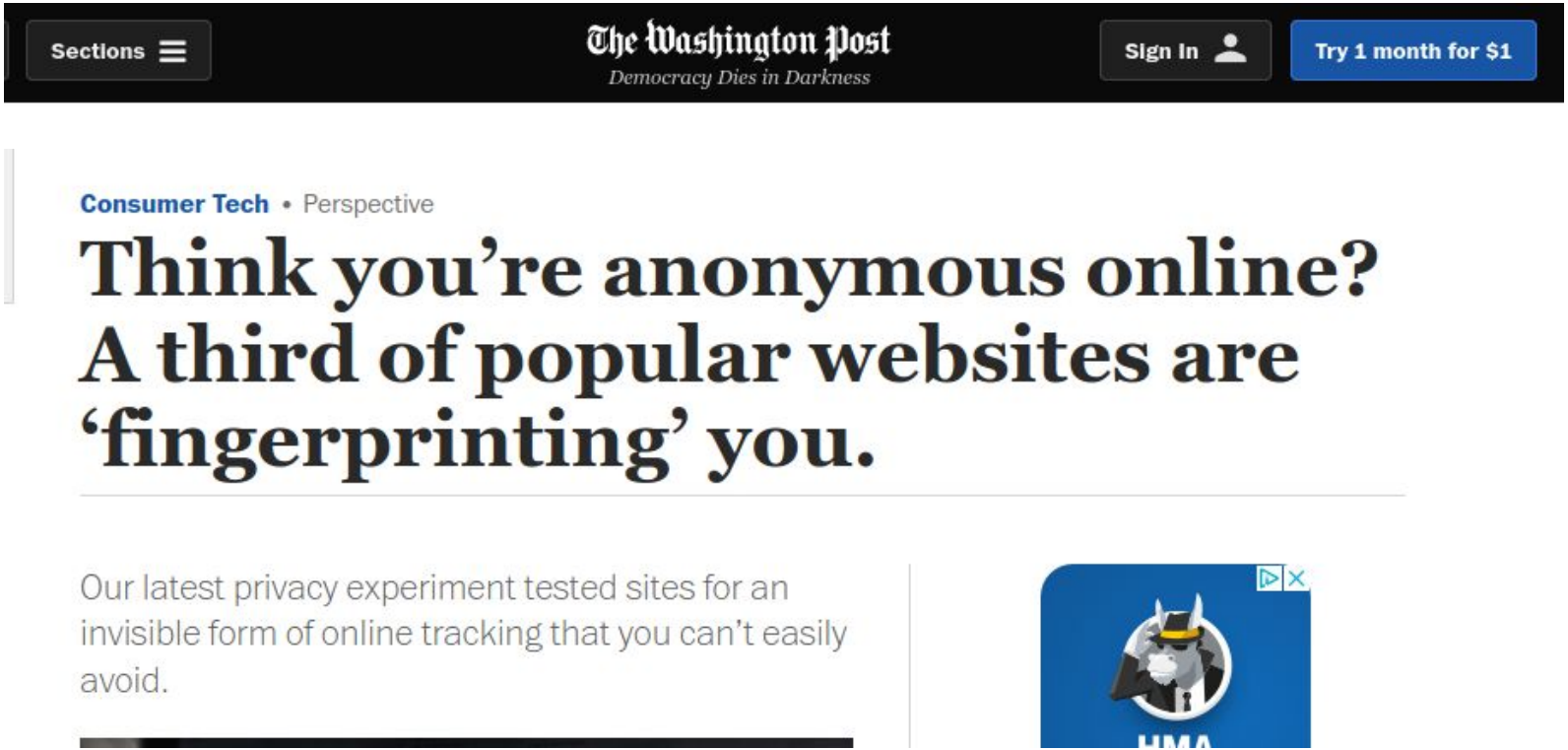


I Visited 47 Sites. Hundreds of Trackers Followed Me.

By Farhad Manjoo
Graphics by Nadieh Bremer

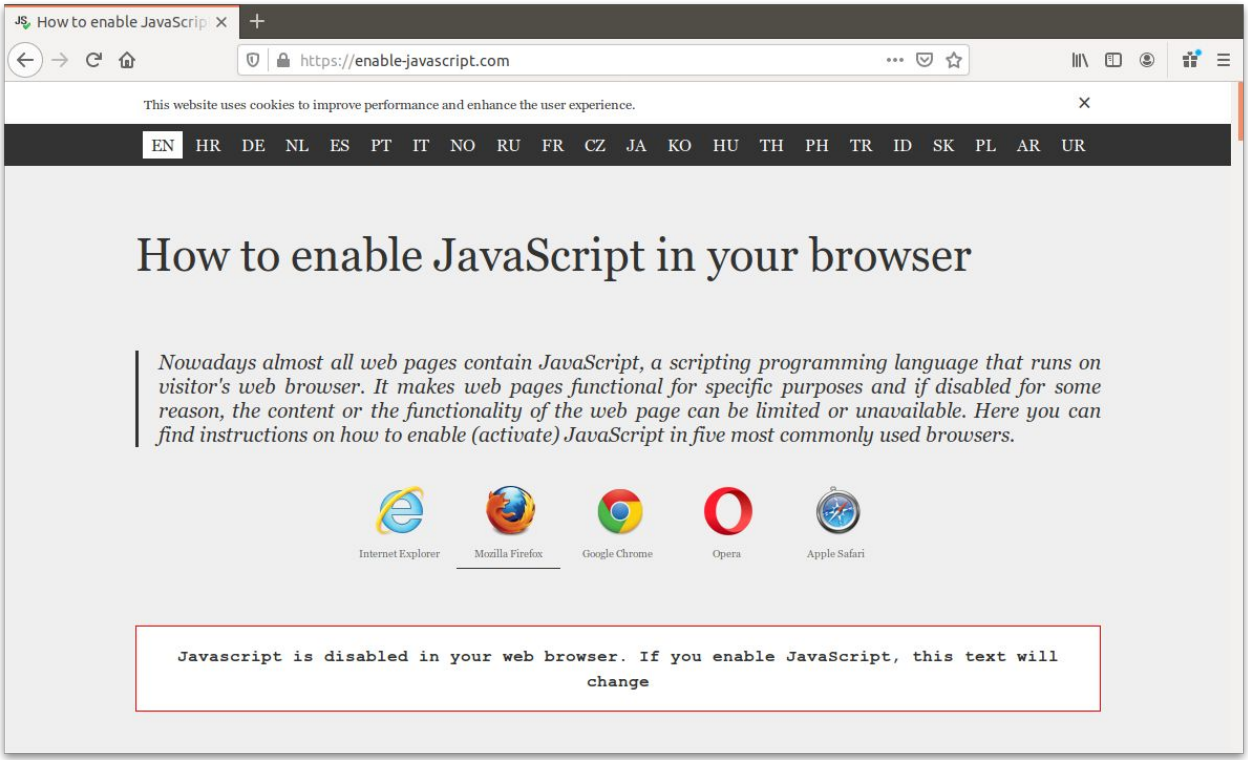


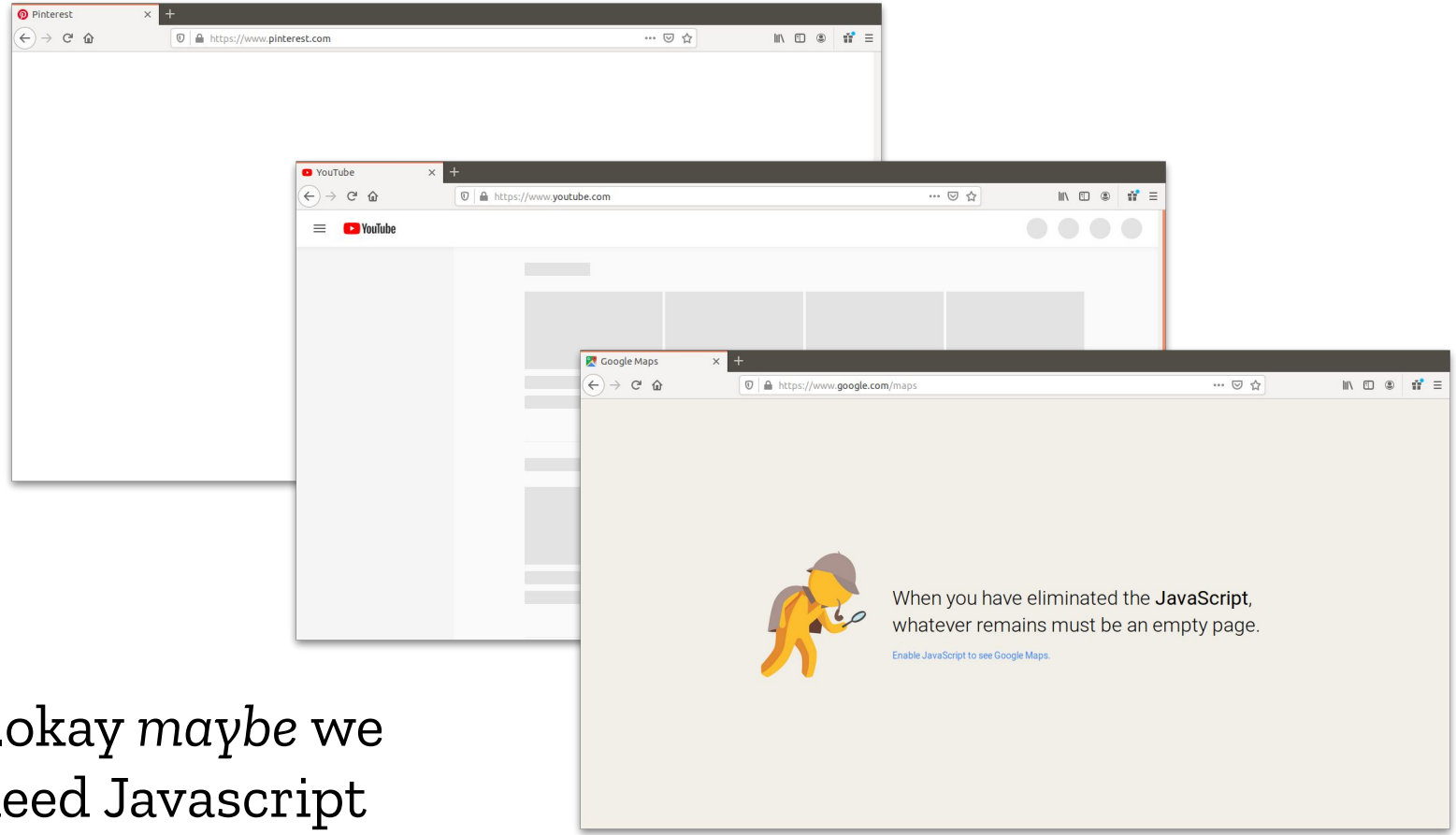
https://www.nytimes.com/interactive/2019/08/23/opinion/data-internet-privacy-tracking.html



https://www.washingtonpost.com/technology/2019/10/31/think-youre-anonymous-online-third-popular-websites-are-fingerprinting-you/

Let's just block it all! Who needs Javascript or cookies?

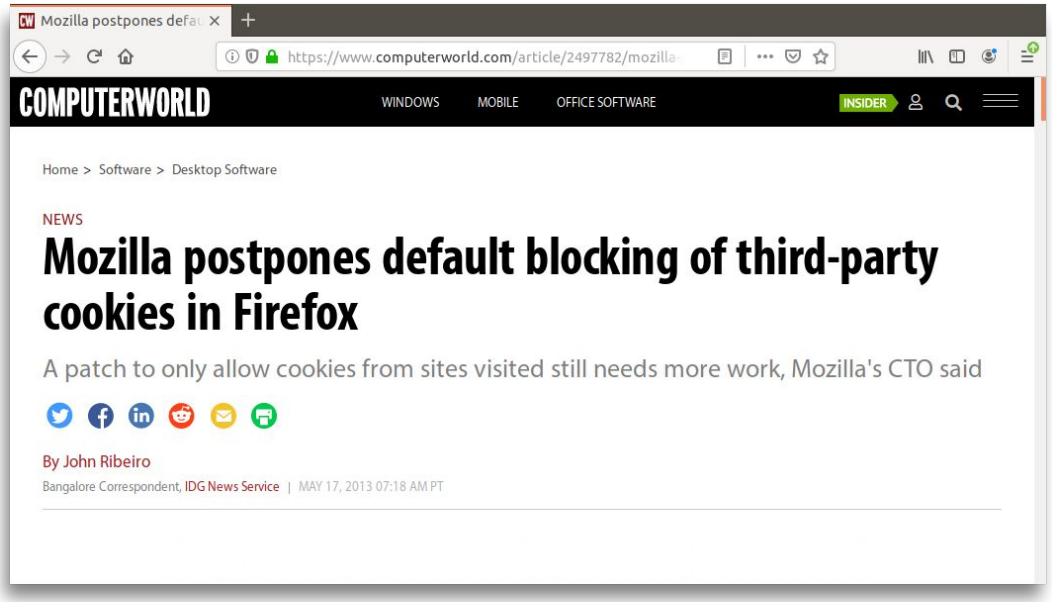




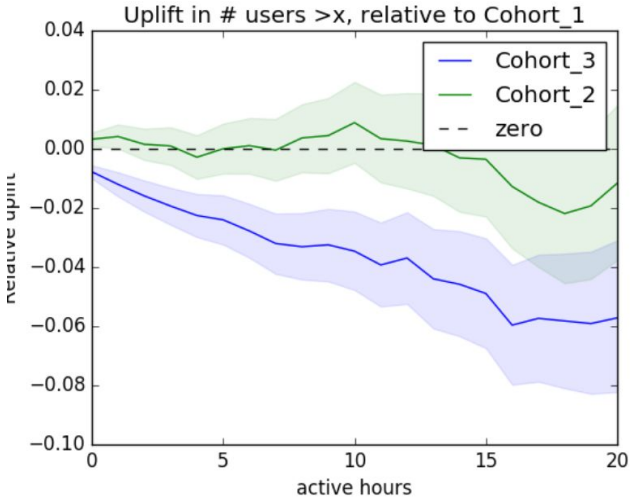
...okay *maybe* we need Javascript

Can we at least block all third-party cookies?

We've tried that back in 2013...



Can we at least block all third-party cookies?



Cohort_3: Block all third-party cookies
Cohort_2: Block cookies from trackers
Cohort_1: Control; No blocking

Compared to no blocking, users who had all third-party cookies blocked (over a 1 - 2 week study period):

- ~1% fewer active users overall
- 2 - 5% fewer users active over 10 hours

The drop was so significant, we ended this part of the study immediately.

https://mozilla.report/post/projects/cookie_restrictions.kp/

The core problem: one identifier, many uses

id=LgzcCiQvIbXtXjtkWrOZ

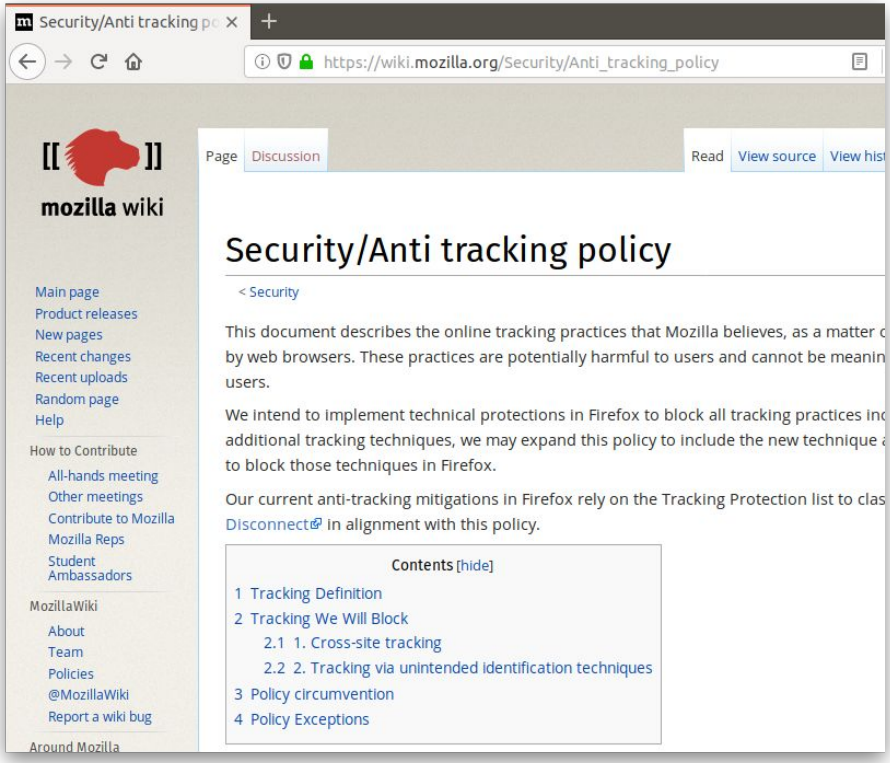
- Cross-site tracking
- Ad performance tracking
- Federated login
- First-party login
- Fraud detection
- Captcha / device reputation
- Analytics
- … and many more

Our approach to anti-tracking...

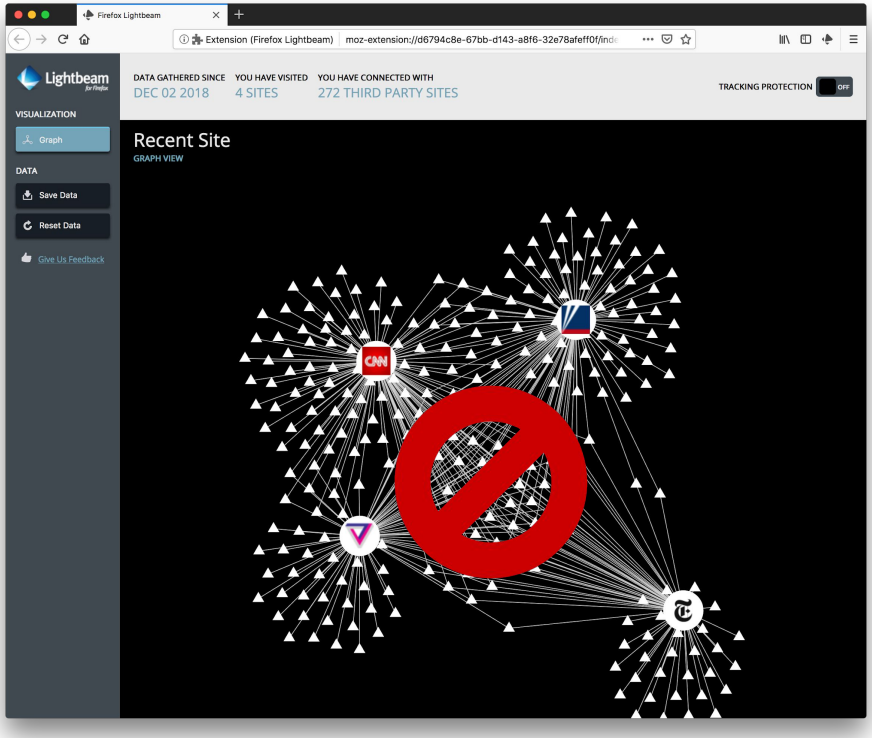
1. No configuration necessary; **private by default**.
2. **Block the trackers**, not necessarily the tracking capability.
3. Don't break experiences **users care about**.

"[T]racking practices that Mozilla believes, as a matter of policy, **should be blocked** by default **by web browsers.**"

1. **Cross-site tracking:**
   a. Cookie-based
   b. URL parameter-based
2. **Unintended identification techniques:**
   a. Browser fingerprinting
   b. Supercookies



Our first step: Block only **tracking** cookies

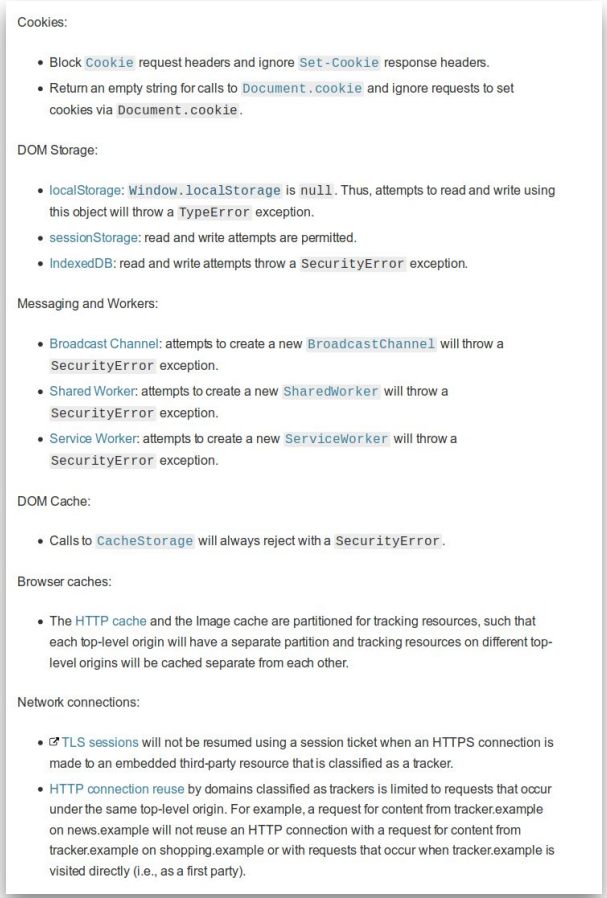


Trackers are identified by Disconnect, based on a review of privacy policies.

It's more than just cookies...

We block all of this →

for all domains on the Disconnect Tracking Protection list



Cookies:

- Block `Cookie` request headers and ignore `Set-Cookie` response headers.
- Return an empty string for calls to `Document.cookie` and ignore requests to set cookies via `Document.cookie`.

DOM Storage:

- localStorage: `Window.localStorage` is `null`. Thus, attempts to read and write using this object will throw a `TypeError` exception.
- sessionStorage: read and write attempts are permitted.
- IndexedDB: read and write attempts throw a `SecurityError` exception.

Messaging and Workers:

- Broadcast Channel: attempts to create a new `BroadcastChannel` will throw a `SecurityError` exception.
- Shared Worker: attempts to create a new `SharedWorker` will throw a `SecurityError` exception.
- Service Worker: attempts to create a new `ServiceWorker` will throw a `SecurityError` exception.

DOM Cache:

- Calls to `CacheStorage` will always reject with a `SecurityError`.

Browser caches:

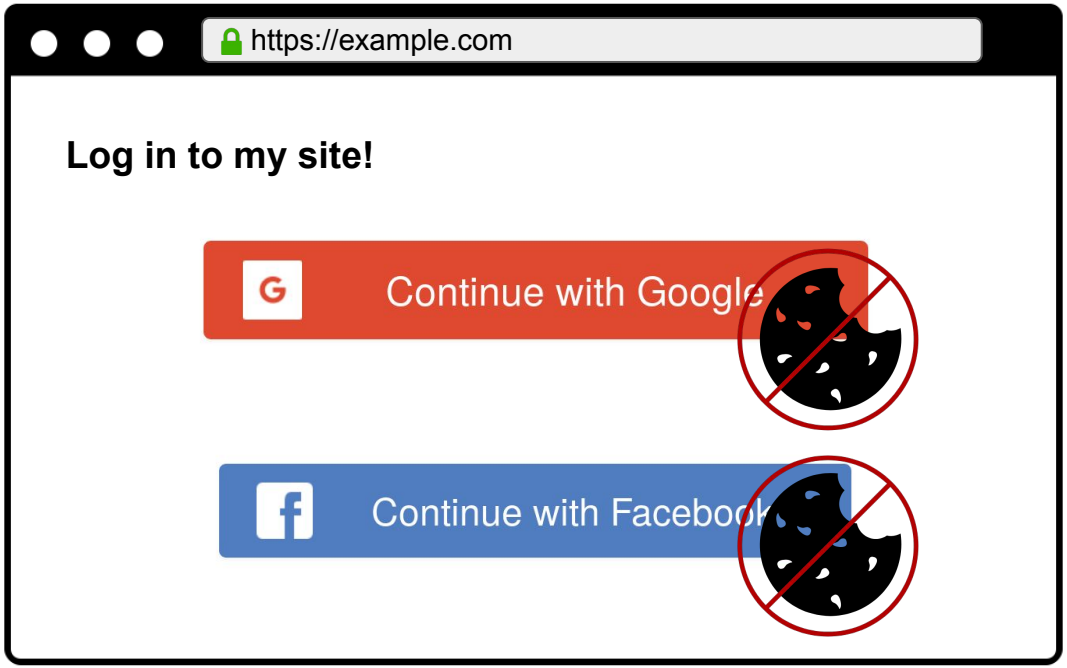
- The HTTP cache and the Image cache are partitioned for tracking resources, such that each top-level origin will have a separate partition and tracking resources on different top-level origins will be cached separate from each other.

Network connections:

- TLS sessions will not be resumed using a session ticket when an HTTPS connection is made to an embedded third-party resource that is classified as a tracker.
- HTTP connection reuse by domains classified as trackers is limited to requests that occur under the same top-level origin. For example, a request for content from tracker.example on news.example will not reuse an HTTP connection with a request for content from tracker.example on shopping.example or with requests that occur when tracker.example is visited directly (i.e., as a first party).

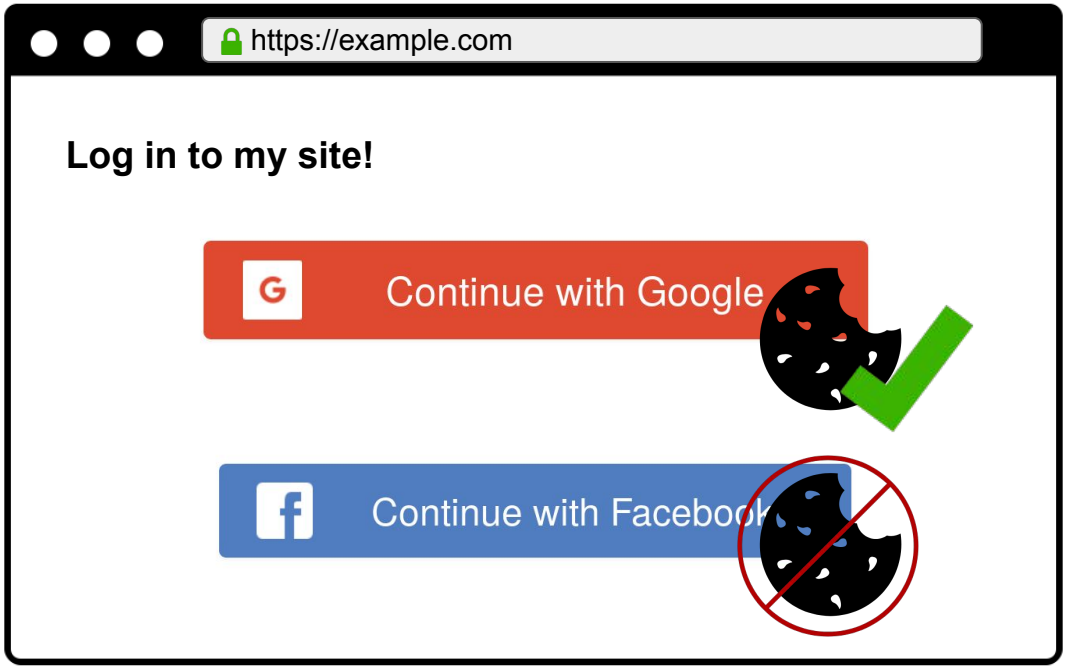
https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Privacy/Storage_access_policy

Some tracking is dual-use



Key insight: users interact with these services!

Cookies permitted after interaction, but **only on** example.com



https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Privacy/Storage_access_policy

Programmatic cookie access: the Storage Access API

Storage Access API methods

The storage API methods are implemented on the `Document` interface:

`Document.hasStorageAccess()`

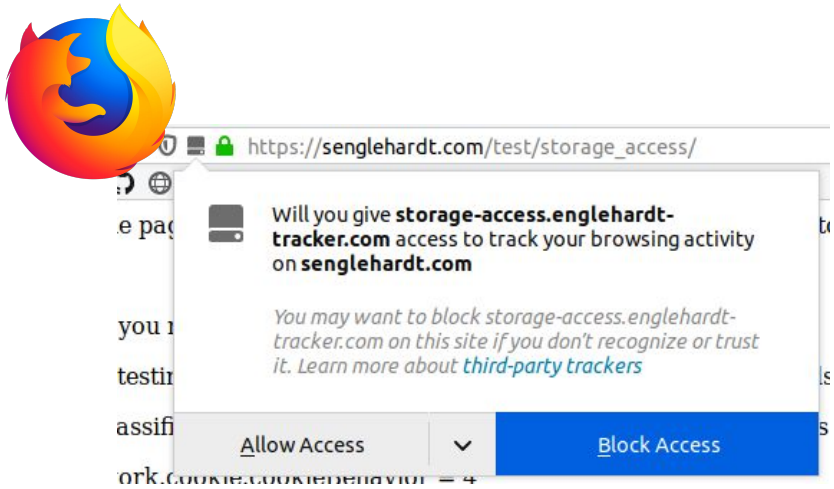
Returns a `Promise` that resolves with a boolean value indicating whether the document has access to its first-party storage.

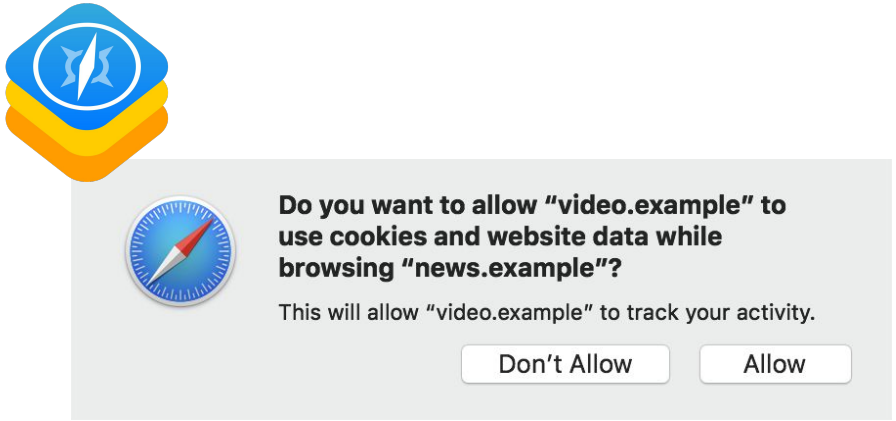
`Document.requestStorageAccess()`

Returns a `Promise` that resolves if the access to first-party storage was granted, and rejects if access was denied.

https://developer.mozilla.org/en-US/docs/Web/API/Storage_Access_API

Programmatic cookie access: the Storage Access API



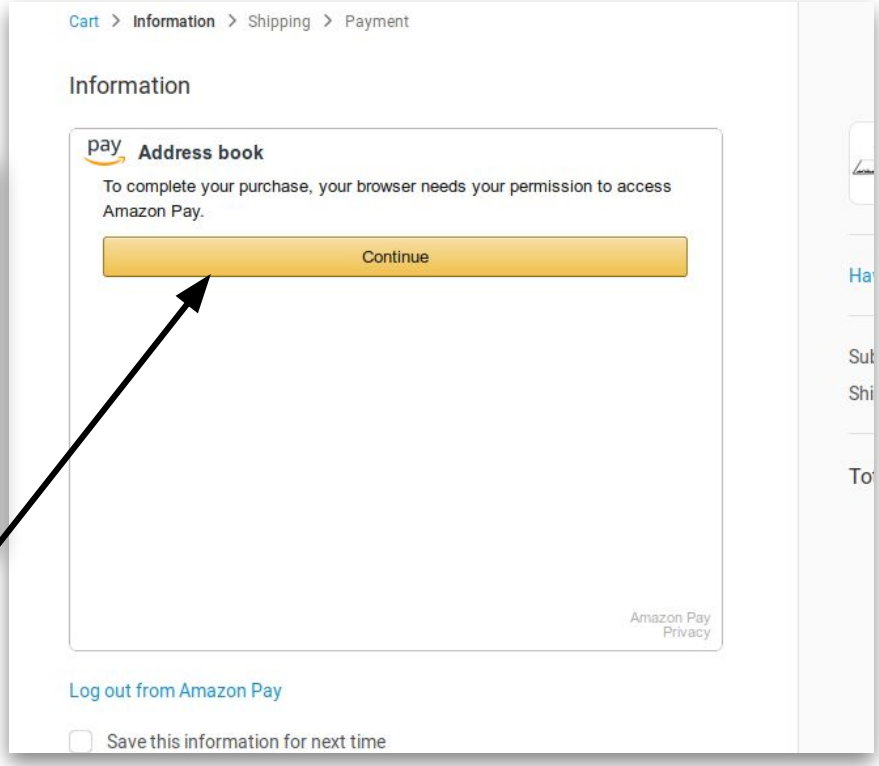


Notable differences in the prompting heuristics and scope of granted access:

- https://developer.mozilla.org/en-US/docs/Web/API/Document/requestStorageAccess#Conditions_for_granting_storage_access
- https://developer.mozilla.org/en-US/docs/Web/API/Storage_Access_API#Safari_implementation_differences

We're seeing adoption of the Storage Access API

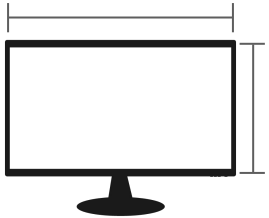
calls Document.requestStorageAccess()



The next step: browser fingerprinting

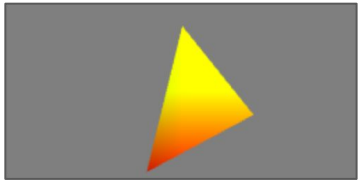
Locale

Screen Size



User Agent

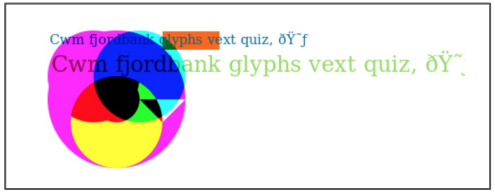
WebGL



Font probing

```
Times New Roman, Arial,
Open Sans, Courier New,
Georgia, Comic Sans, ...
```

HTML Canvas



Fingerprinting use is still growing...

2016

~1.6% of the Alexa top 1M

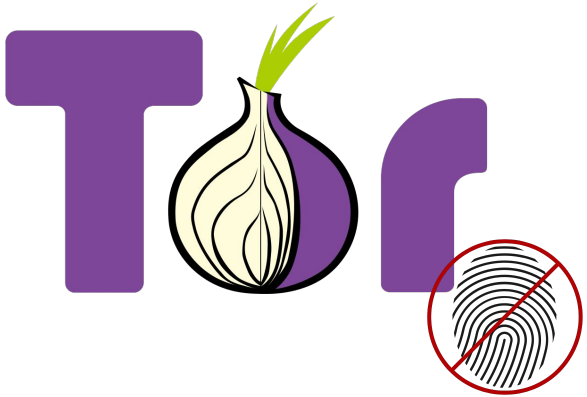
Englehardt & Narayanan; Online Tracking (CCS2016)

2019

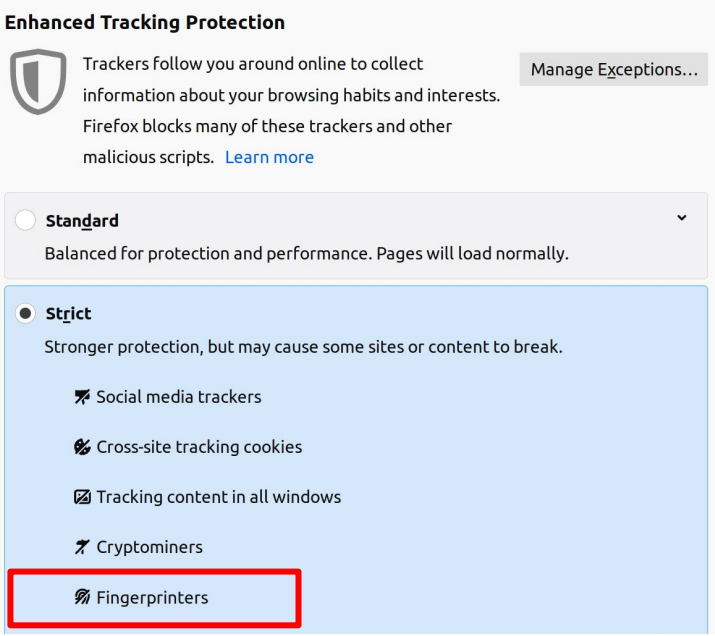
~3.8% of the Alexa top 1M

Unpublished OpenWPM measurements from Feb 2019

Two general approaches to anti-fingerprinting



Tor Browser's anti-fingerprinting in Firefox behind `privacy.resistFingerprinting`



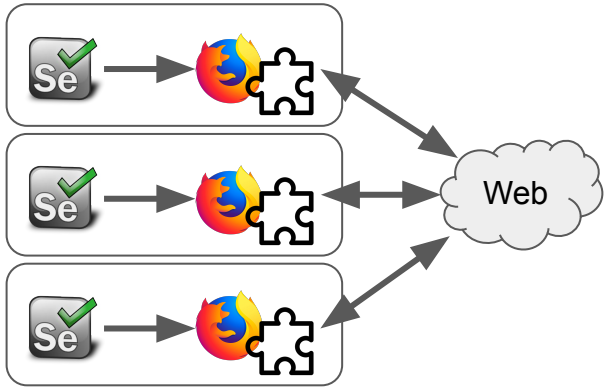
Firefox 70

Changing APIs is hard...

- Image scaling problems from changing `devicePixelRatio`
- Image transparency issues
- Framerate and performance problems from timing changes

https://bugzilla.mozilla.org/show_bug.cgi?id=1507517

Our current approach: blocking fingerprinting scripts

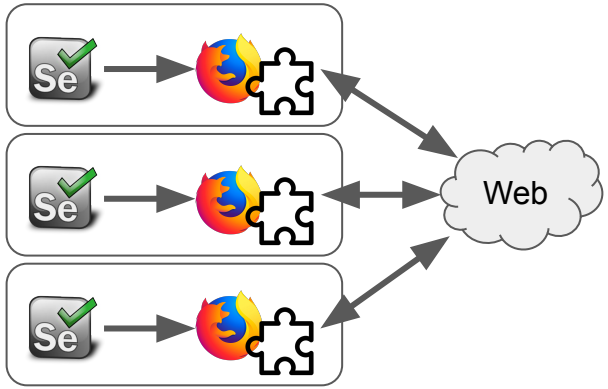


Crawl the web with OpenWPM. Detect fingerprinting scripts.

Share flagged scripts with Disconnect, who does a review to remove false positives.

Domains blocked in Firefox. Eventually by default.

Our current approach: blocking fingerprinting scripts

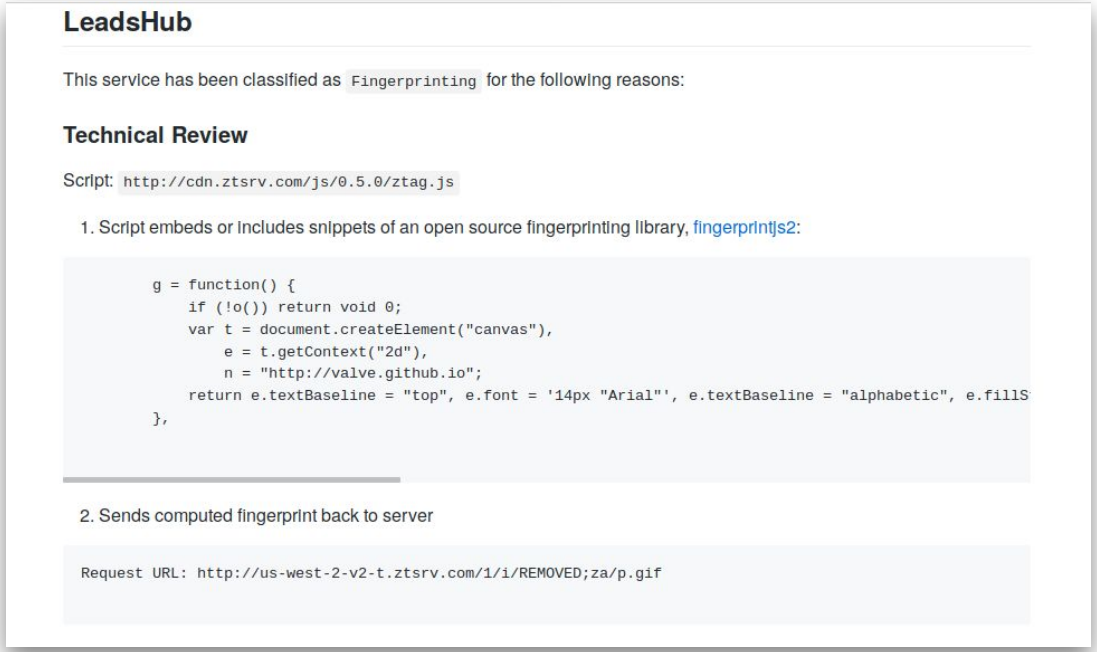


138 verified fingerprinters

75% of the fingerprinting instances we've detected

Disconnect reviews candidate scripts

An example:



LeadsHub

This service has been classified as `Fingerprinting` for the following reasons:

Technical Review

Script: `http://cdn.ztsrv.com/js/0.5.0/ztag.js`

1. Script embeds or includes snippets of an open source fingerprinting library, fingerprintjs2:

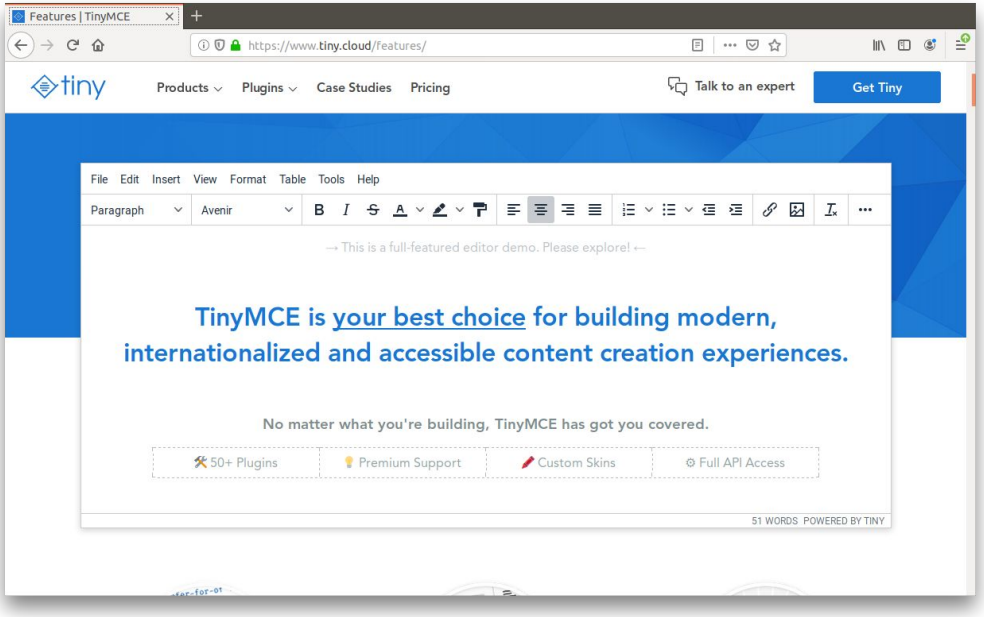
```
g = function() {
    if (!o()) return void 0;
    var t = document.createElement("canvas"),
        e = t.getContext("2d"),
        n = "http://valve.github.io";
    return e.textBaseline = "top", e.font = '14px "Arial"', e.textBaseline = "alphabetic", e.fillS
},
```

2. Sends computed fingerprint back to server

```
Request URL: http://us-west-2-v2-t.ztsrv.com/1/i/REMOVED;za/p.gif
```

https://github.com/disconnectme/disconnect-tracking-protection/blob/master/descriptions.md

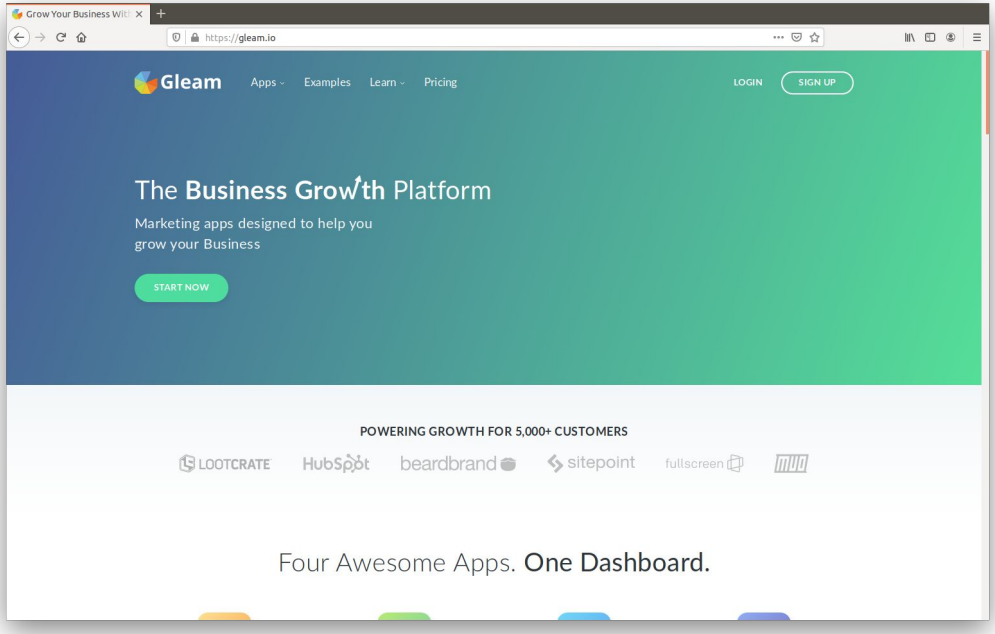
Initial success: changing practices



TinyMCE removed fingerprint2js from their HTML Editor as a result of our blocking

https://bugzilla.mozilla.org/show_bug.cgi?id=1544159

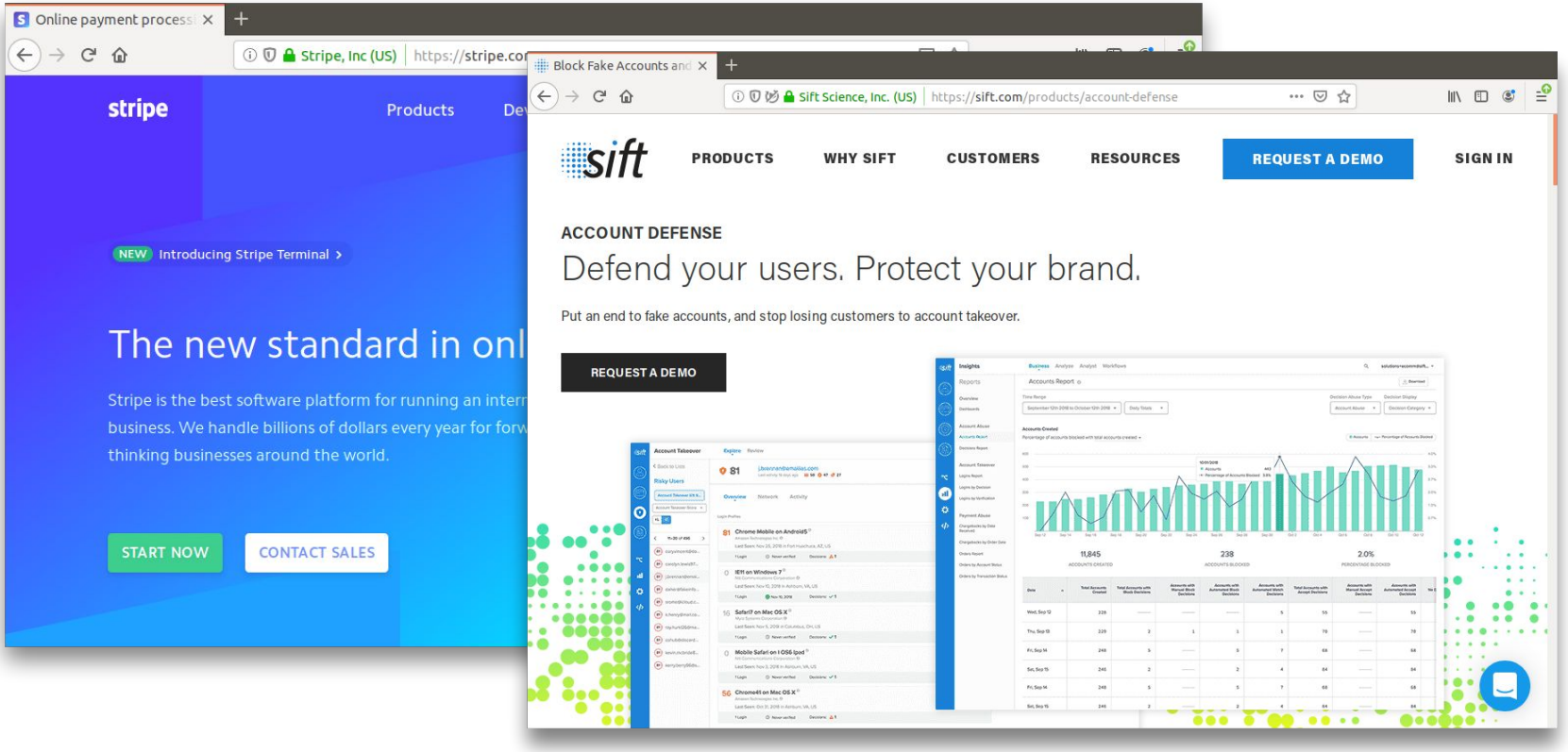
Initial success: changing practices



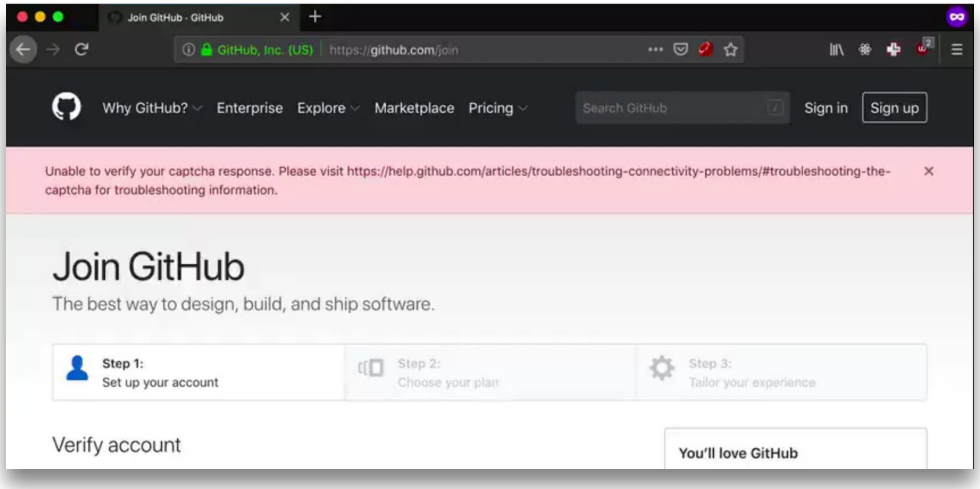
Gleam moved fingerprinting script from gleam.io to fraudjs.io to avoid app breakage

https://bugzilla.mozilla.org/show_bug.cgi?id=1558658

Challenge: Fingerprinting for anti-fraud



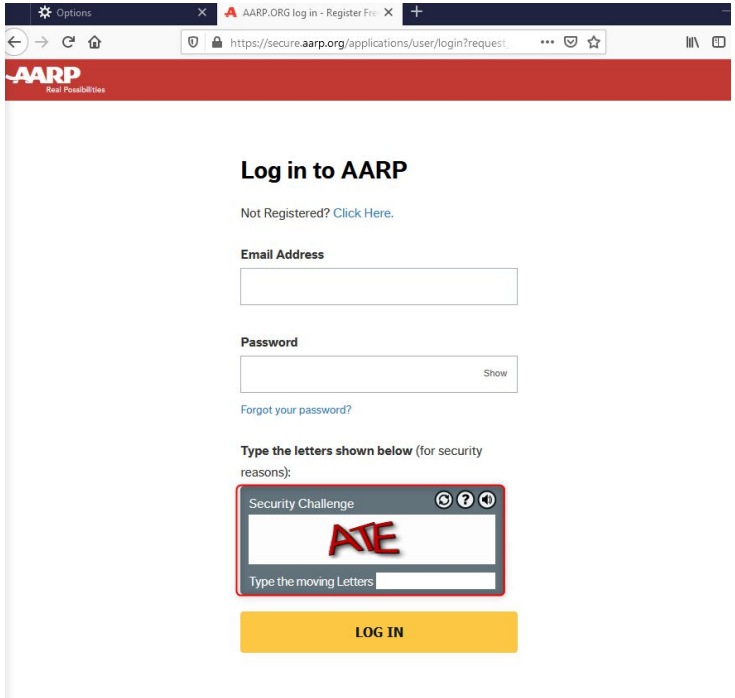
Blocking anti-fraud leads to major site breakage



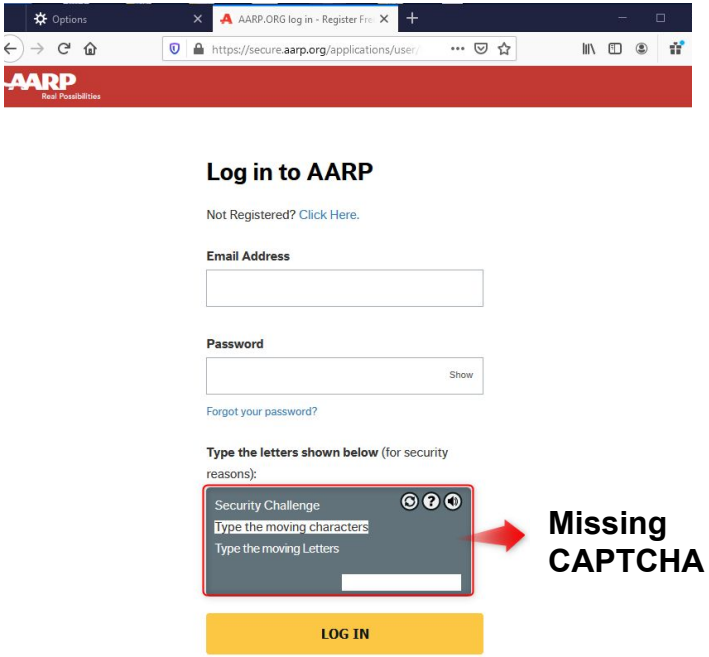
Github account creation broken when Arkose Labs captcha was blocked for fingerprinting

https://twitter.com/richgilbank/status/1090269987888082944

Blocking anti-fraud leads to major site breakage

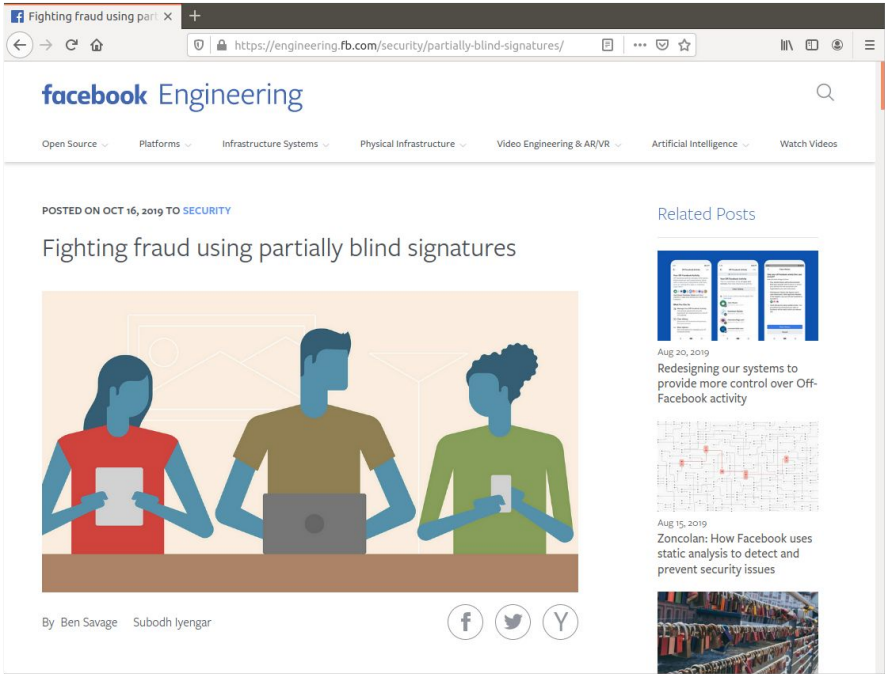


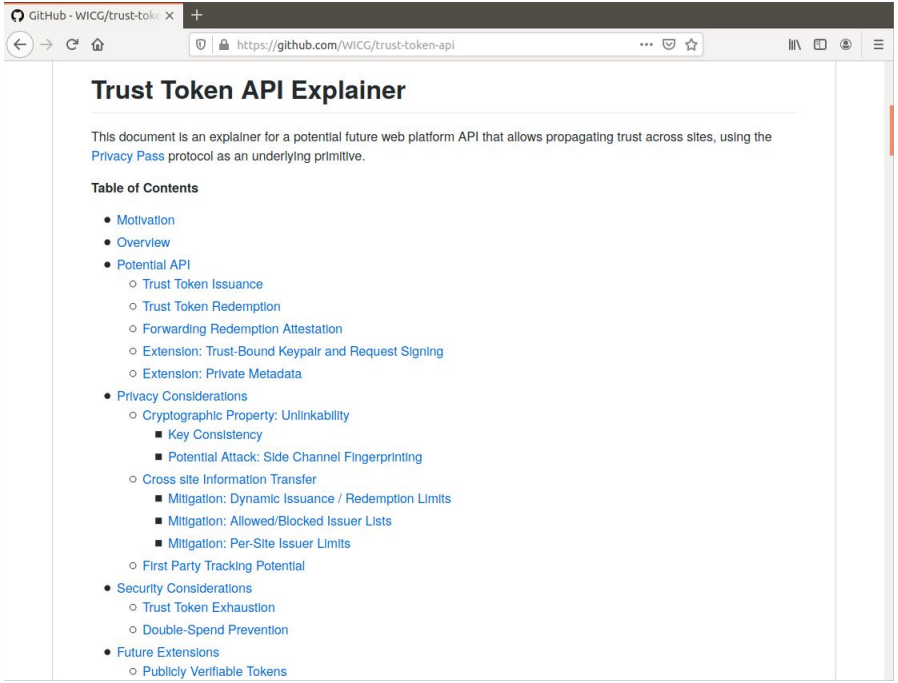
No Blocking



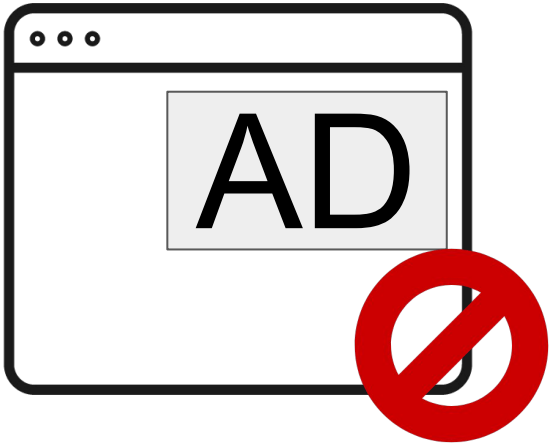
Blocking Fingerprinters

Two possible solutions to safer anti-fraud?

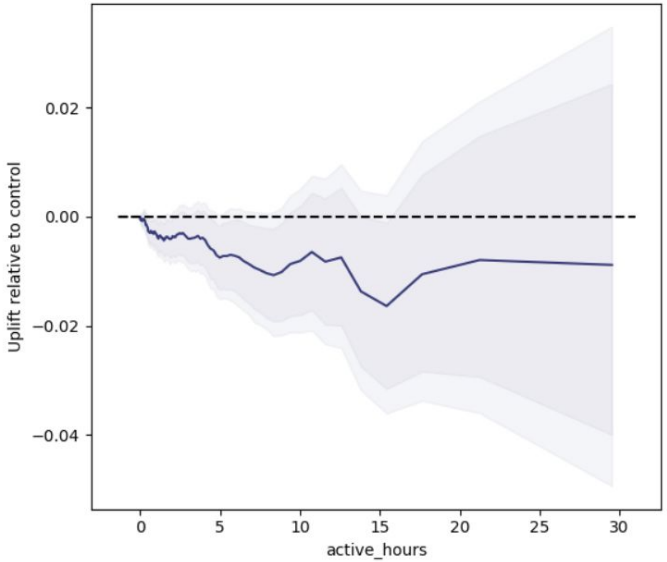




Assume fraud is solved; can we block everything else?



Assume fraud is solved; can we block everything else?



… we tried that in a user study but we still saw 0.2% and 0.6% users leave Firefox because of it.

https://mozilla.report/post/projects/fingerprinting_protections.kp/

Why does blocking non-tracking fingerprinters cause users to leave Firefox?

Challenge: discovering sites broken by our protections

Approach	**Problems**
User Reports	• Noisey • Unreliable
User Studies	• Noisey • No clear way to measure
Manual QA	• Limited scope • Expensive: 1 month of full-time work per 1,000 sites
Automated Crawls	• Limited scope • No clear way to measure

Mozilla Research Grants 2019H2

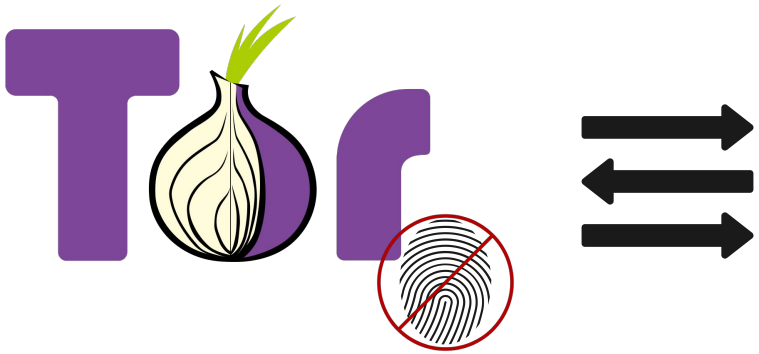
How can we automate the process of discovering broken sites?

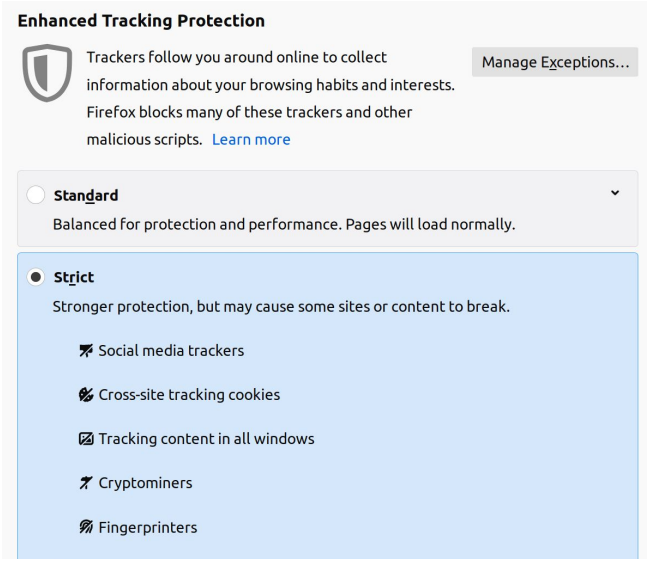
Amount: $40,000

Deadline
Friday, November 22nd at 2:22pm
Pacific Daylight Time (PDT)

https://mozilla-research.forms.fm/mozilla-research-grants-2019h2/forms/7376

A possible step forward for anti-fingerprinting?





Per-frame fingerprinting resistance based on a blocklist

https://bugzilla.mozilla.org/show_bug.cgi?id=1531873

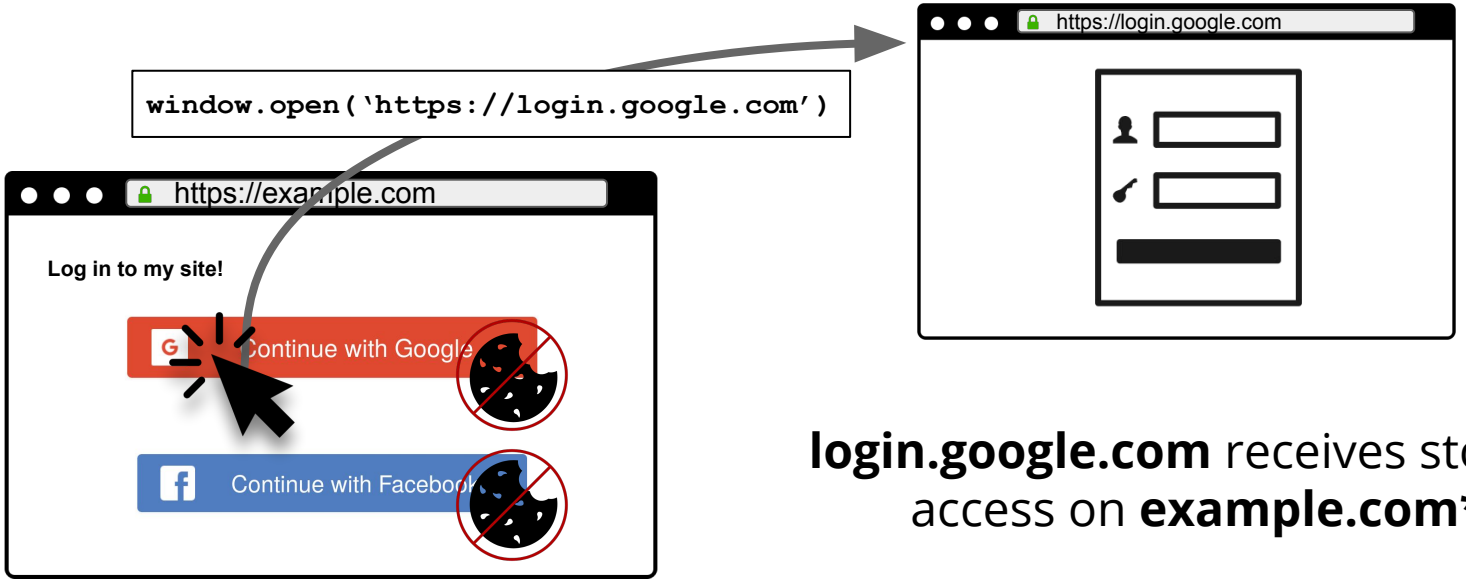
My asks for you:

1. Help us find technical alternatives to a global identifier for:
   a. Federated login
   b. Anti-fraud / device reputation
   c. Advertisement attribution / measurement
2. Find violations of our anti-tracking policies
   a. Name and shame
   b. We can update our blocks
3. Help us explore ways to better discover broken sites
   a. Apply to our grant!

THANK YOU!

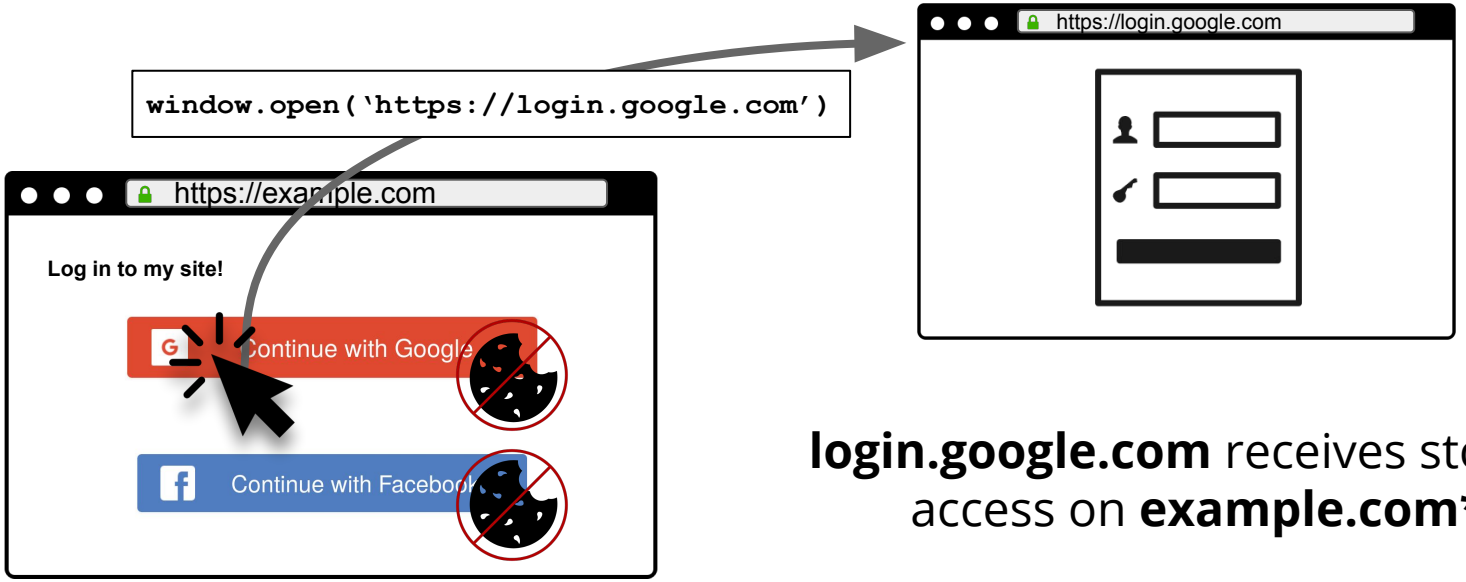
BACKUP SLIDES

Interaction allows interactive embeds to work



login.google.com receives storage access on **example.com***

Interaction allows interactive embeds to work

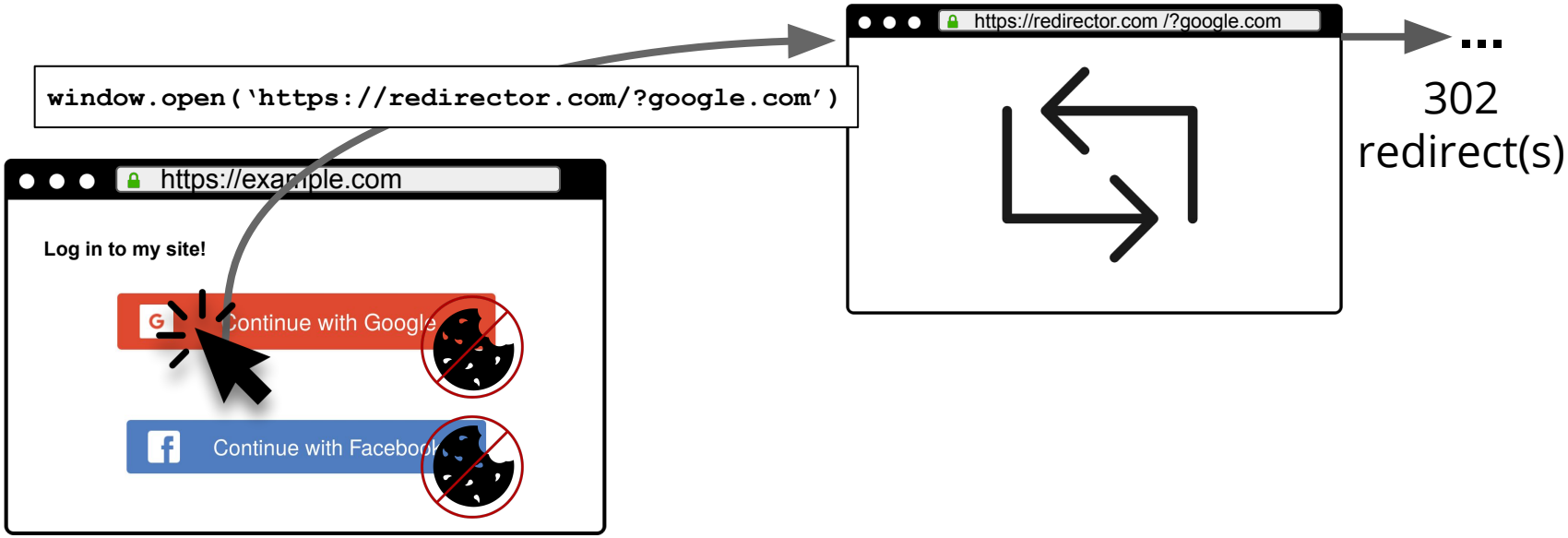


login.google.com receives storage access on **example.com***

** Provided it meets some additional requirements.*

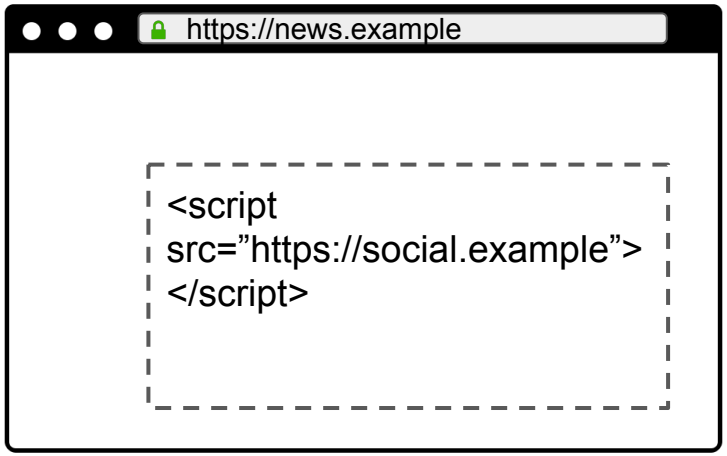
See: https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Privacy/Storage_access_policy

We also handle redirects...



See: https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Privacy/Storage_access_policy

A workaround: Cross-site tracking with query strings



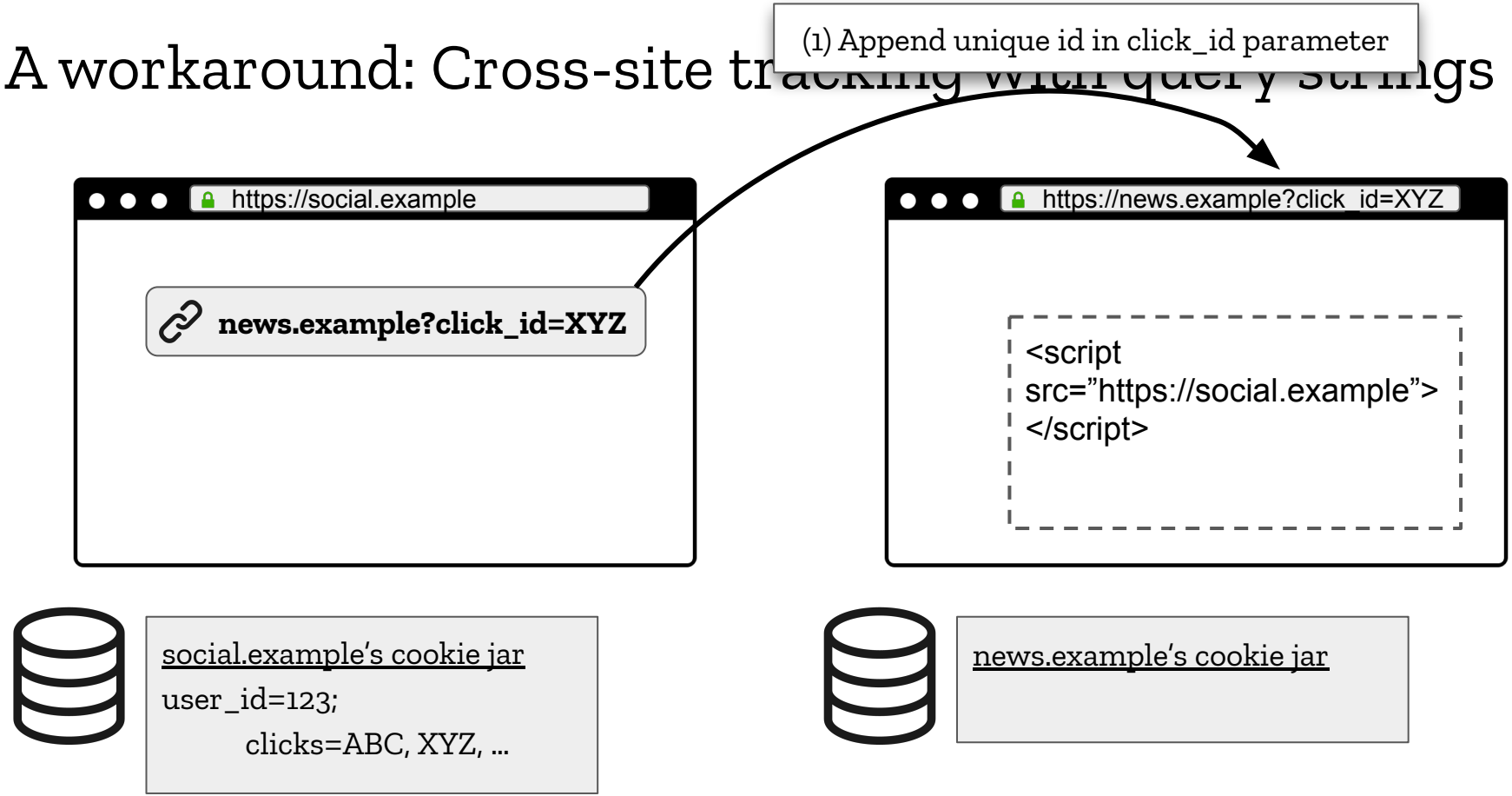
social.example's cookie jar

user_id=123;

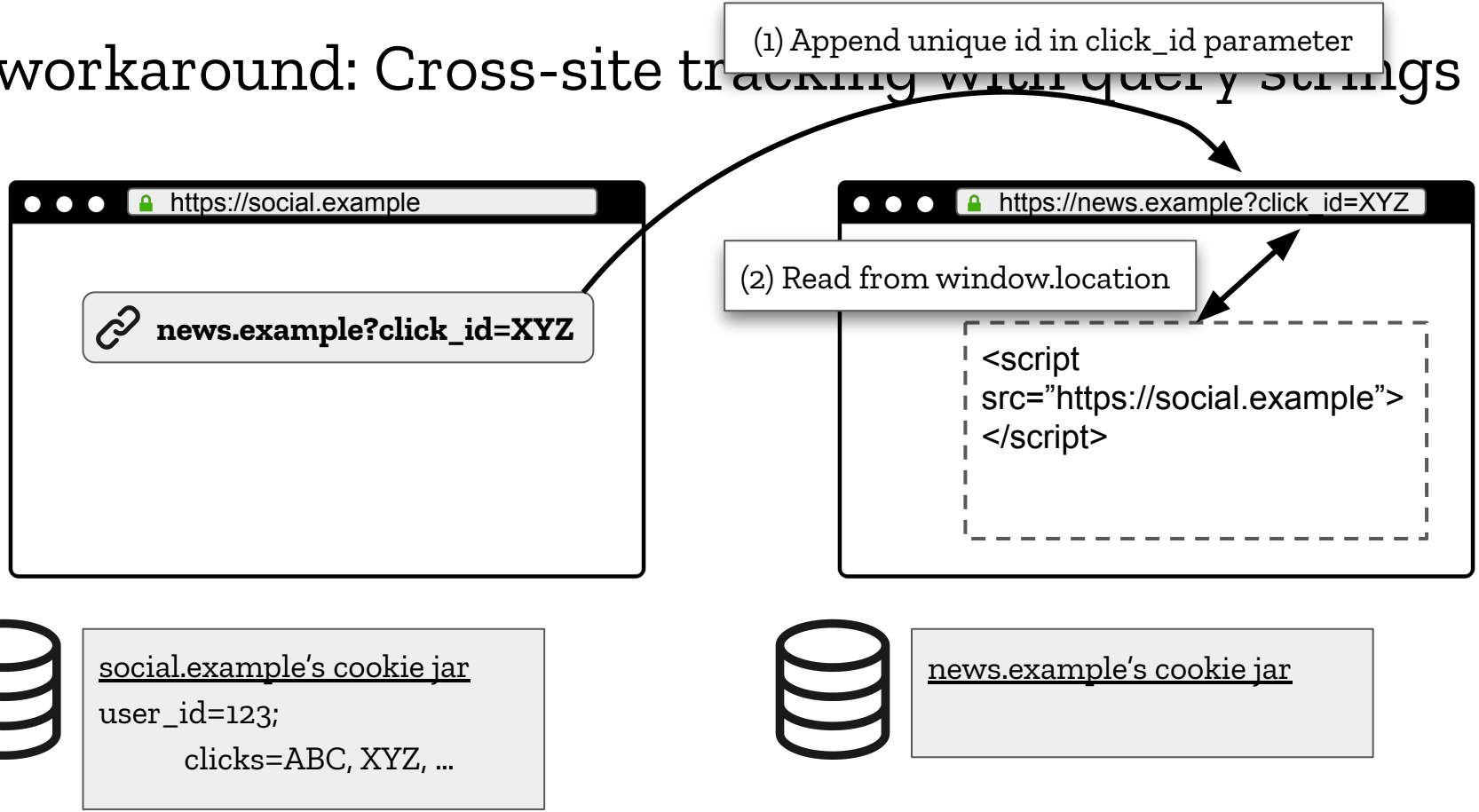
clicks=ABC, ...

news.example's cookie jar

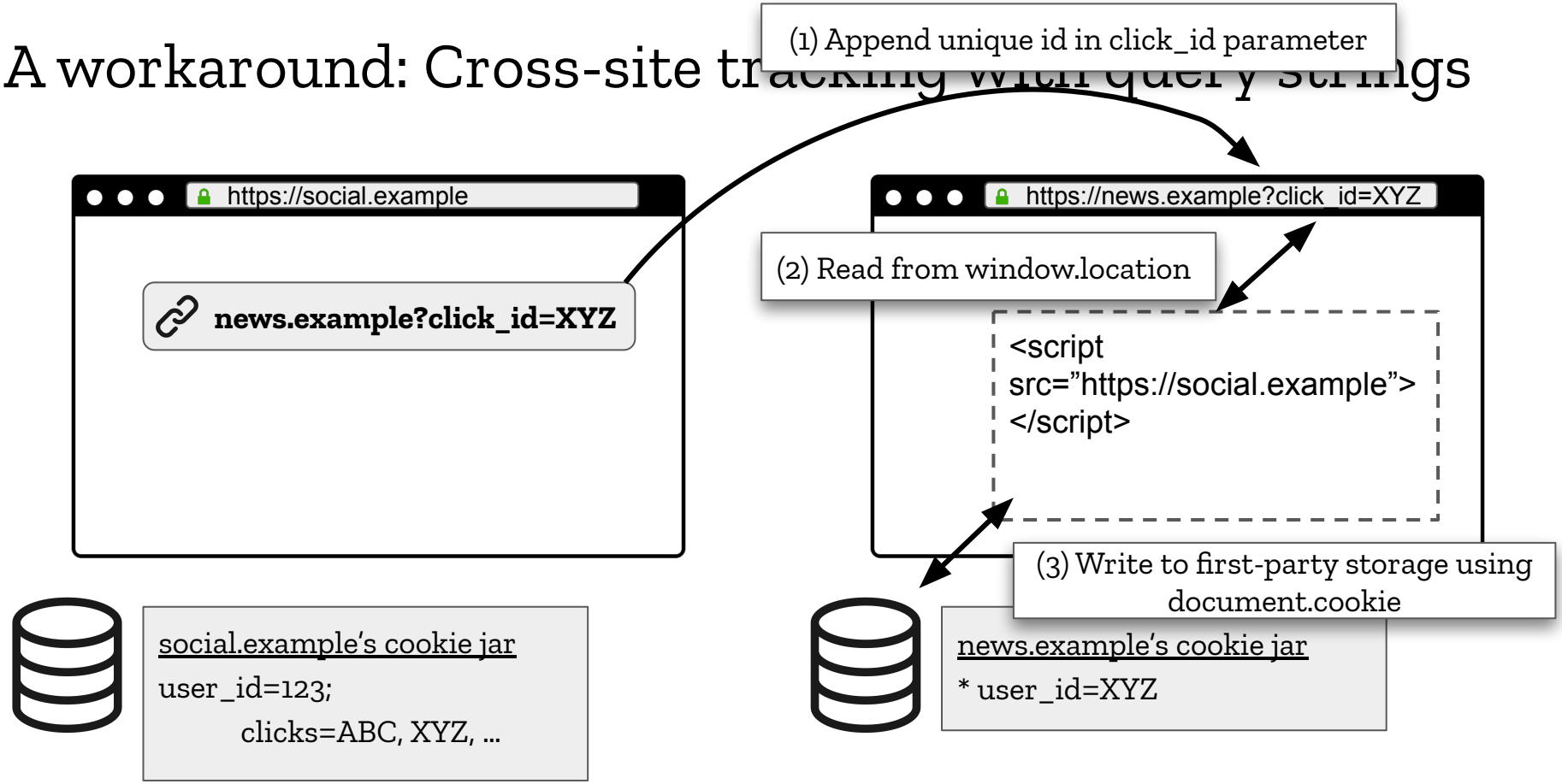
A workaround: Cross-site tracking with query strings



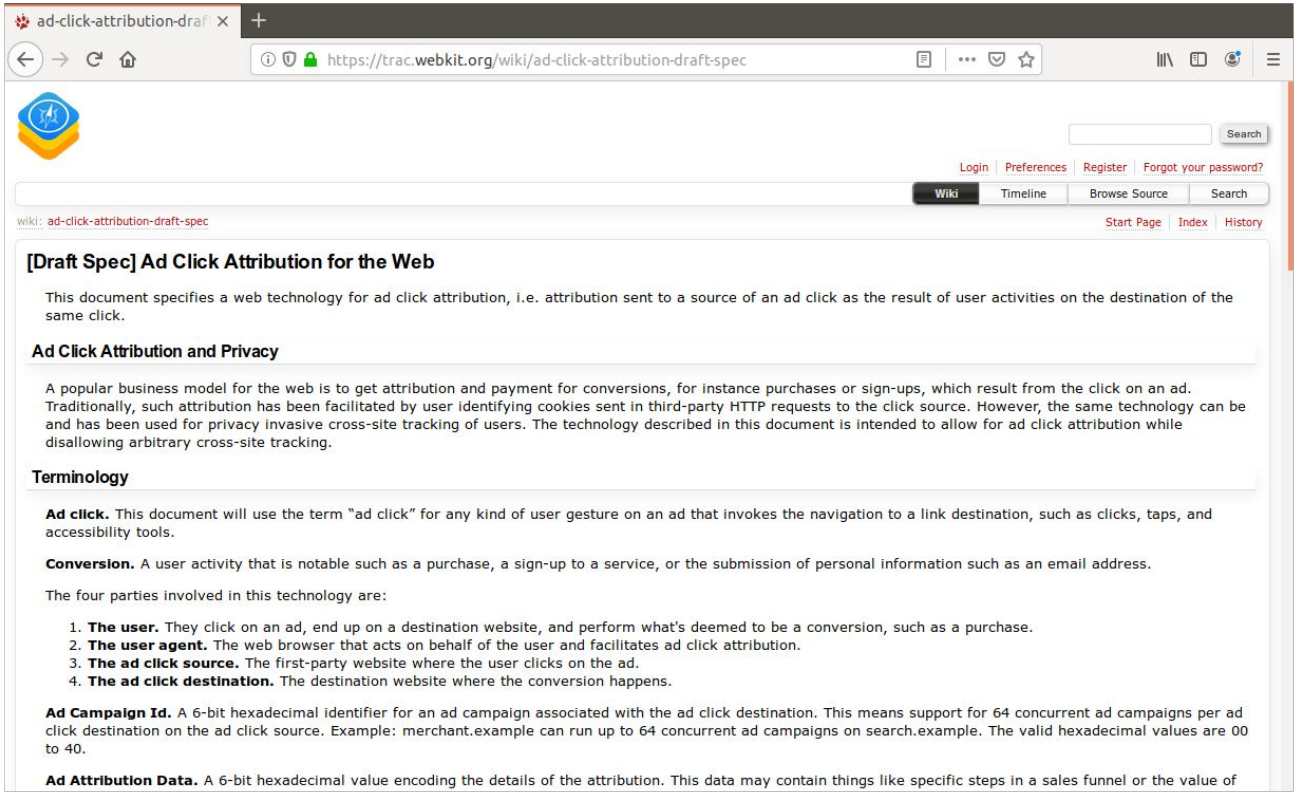
A workaround: Cross-site tracking with query strings



A workaround: Cross-site tracking with query strings



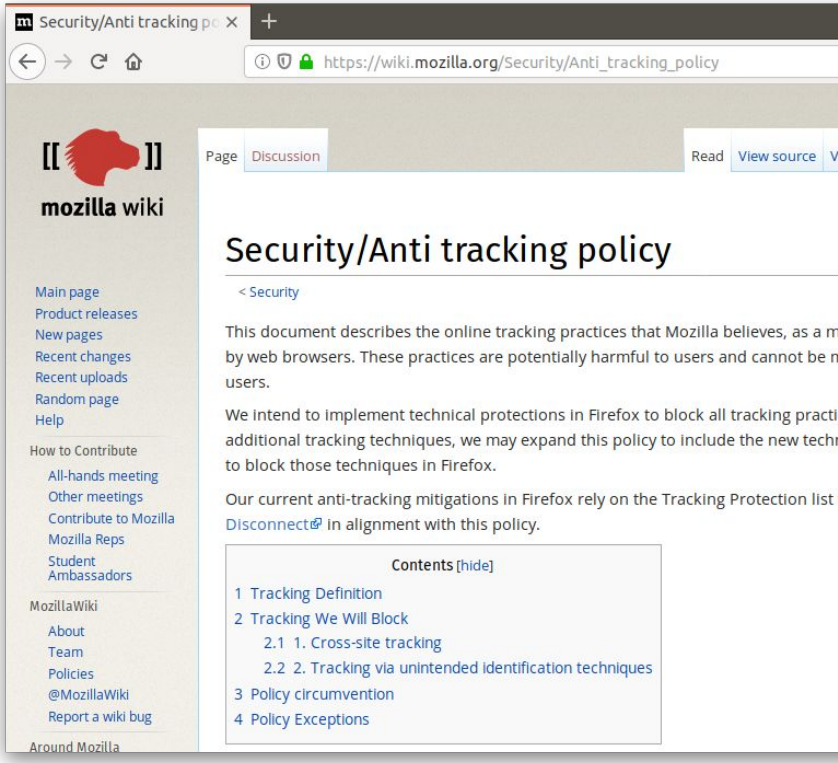
Request: a safer way to do ad measurement



Challenge: Fingerprinting for anti-fraud

Policy exceptions:

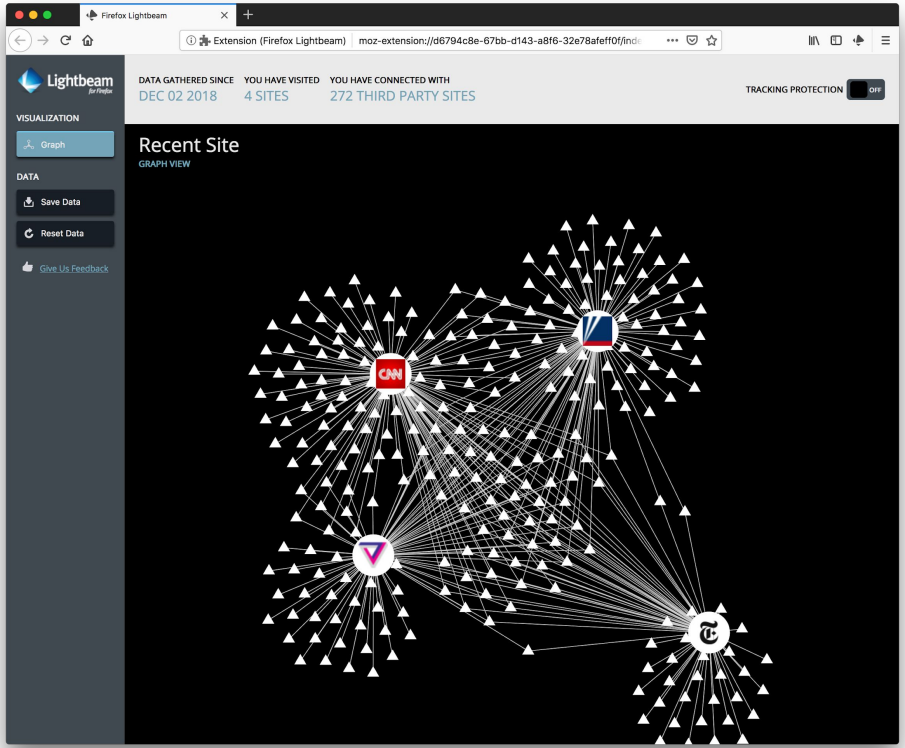
1. Improving client authentication
2. Preventing the creation of fraudulent accounts
3. Preventing the completion of fraudulent purchases.



The web needs **default-on** tracking protection …

4 news sites

272 third parties



The web needs **default-on** tracking protection …
… and not just from third-party cookies

Browser state

IP Address + Device Properties

Identity

Browser state

IP Address + Device Properties

Identity

Tracking vectors completely within browser's control

The tracking landscape

Browser state

IP Address + Device Properties

Identity

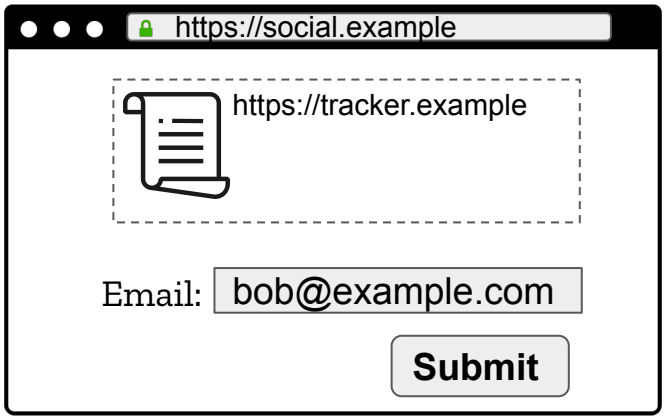
The tracking landscape

Browser state

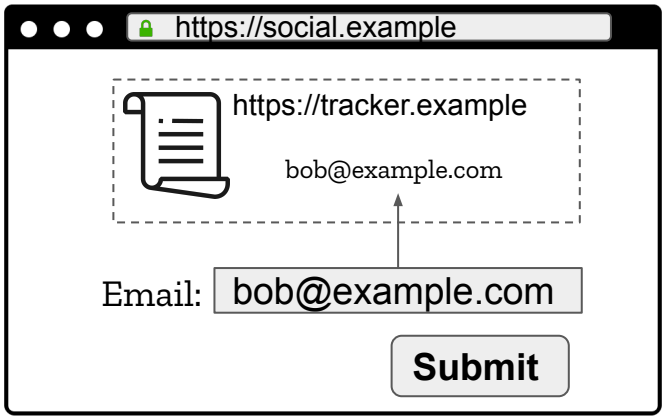
IP Address + Device Properties

Identity

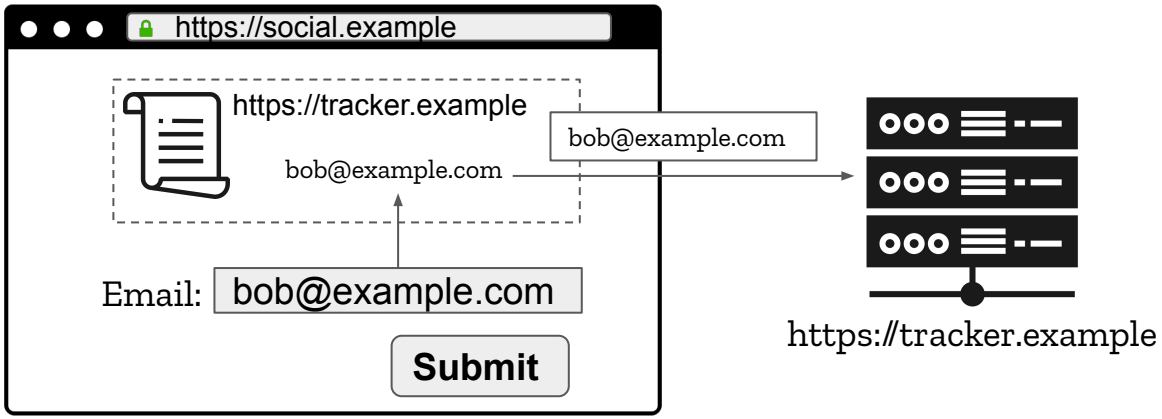
Scripts can collect PII for tracking



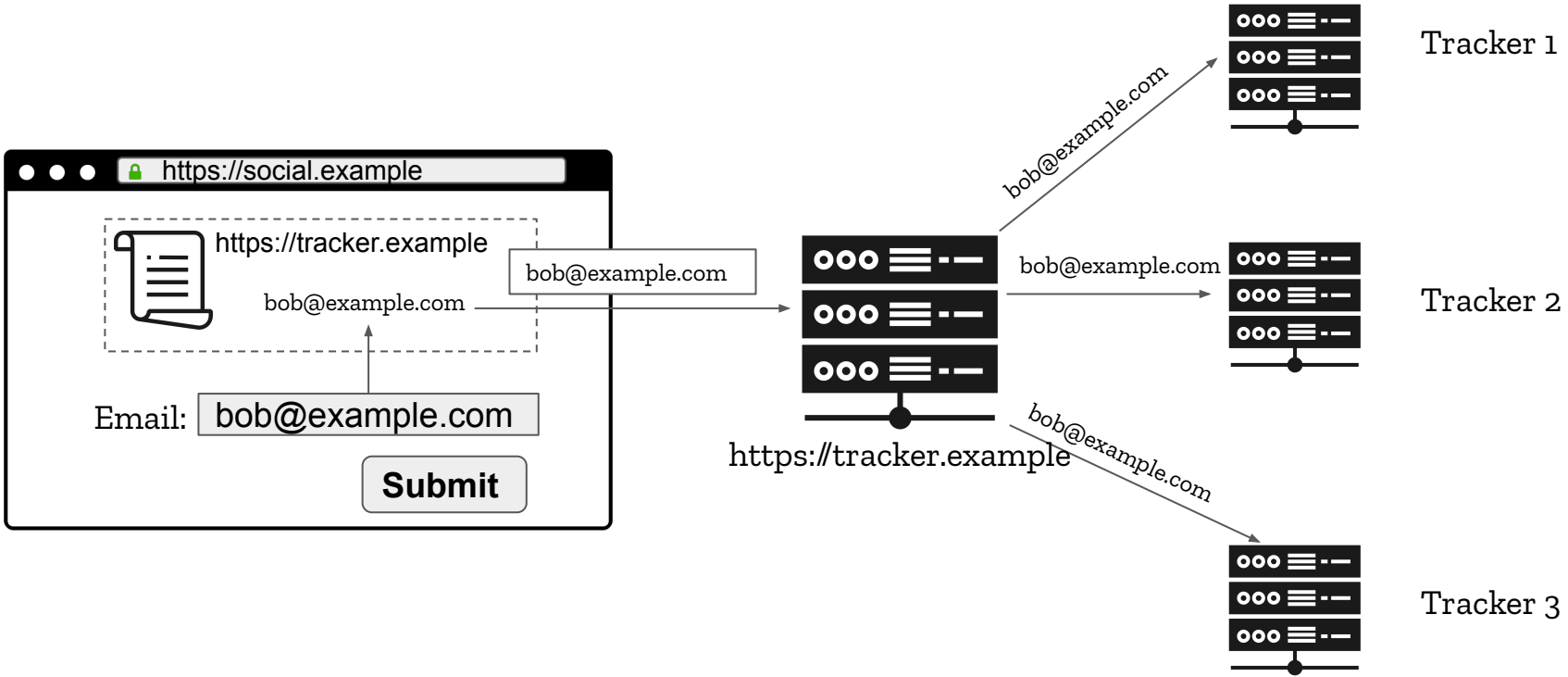
Scripts can collect PII for tracking



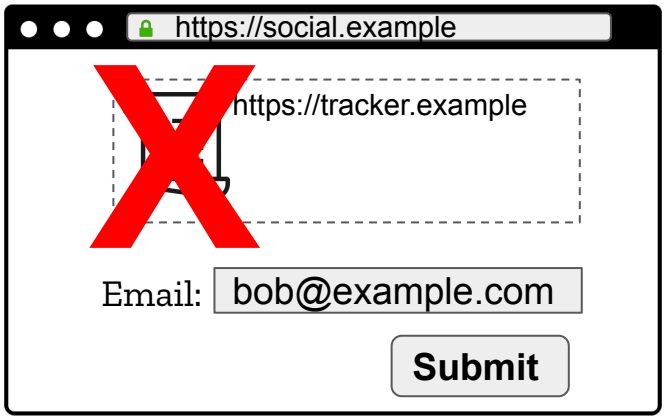
Scripts can collect PII for tracking

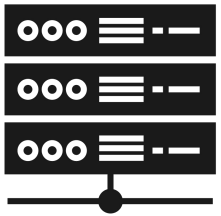


Scripts can collect PII for tracking

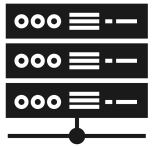


We can block scripts that scrape PII

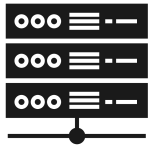




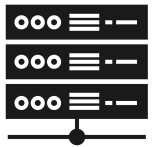
https://tracker.example



Tracker 1

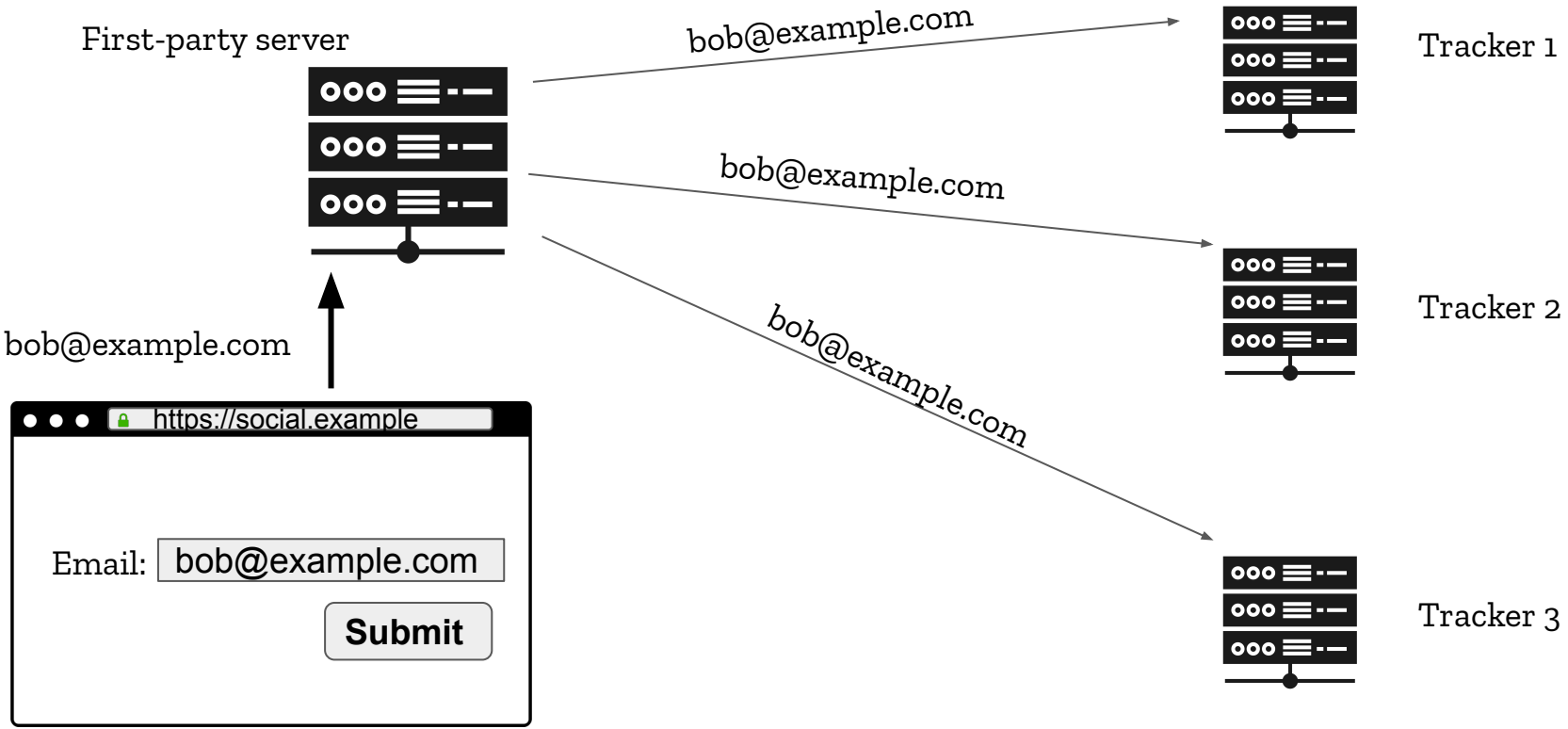


Tracker 2



Tracker 3

In-browser protection options are limited...



The tracking landscape

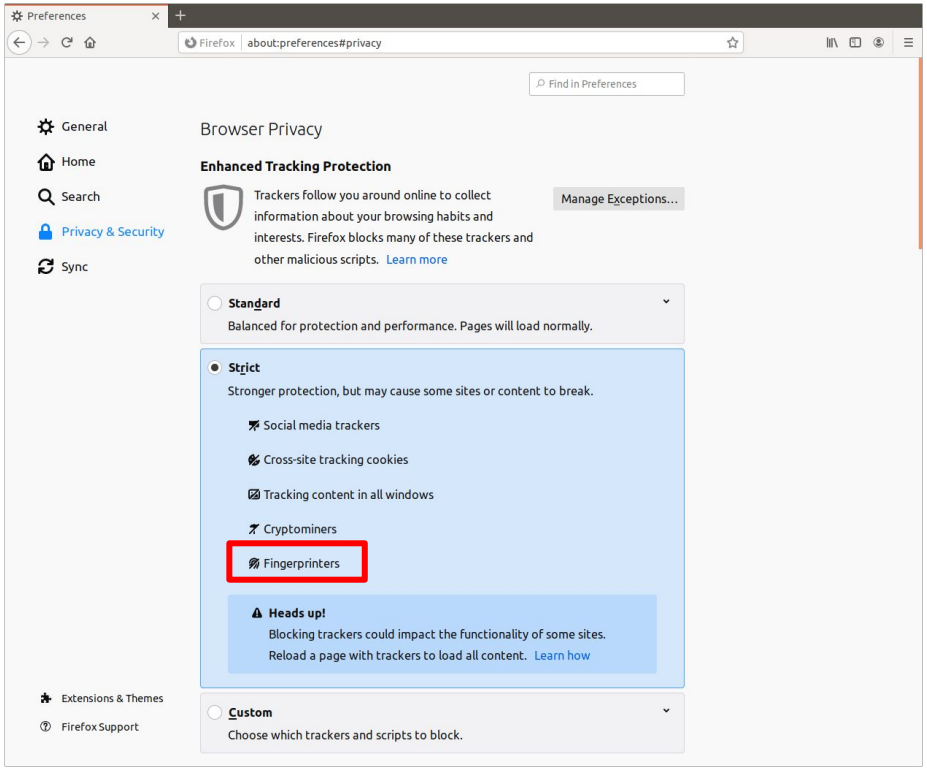
Browser state

IP Address + Device Properties

Identity

Fingerprinting blocking available since Firefox 67

Firefox 70



How can we prevent identity-based tracking?

Request: a safer way to do anti-fraud