Swiss National Science Foundation

Vault JS

Online tracking

CS-523 Advanced topics on

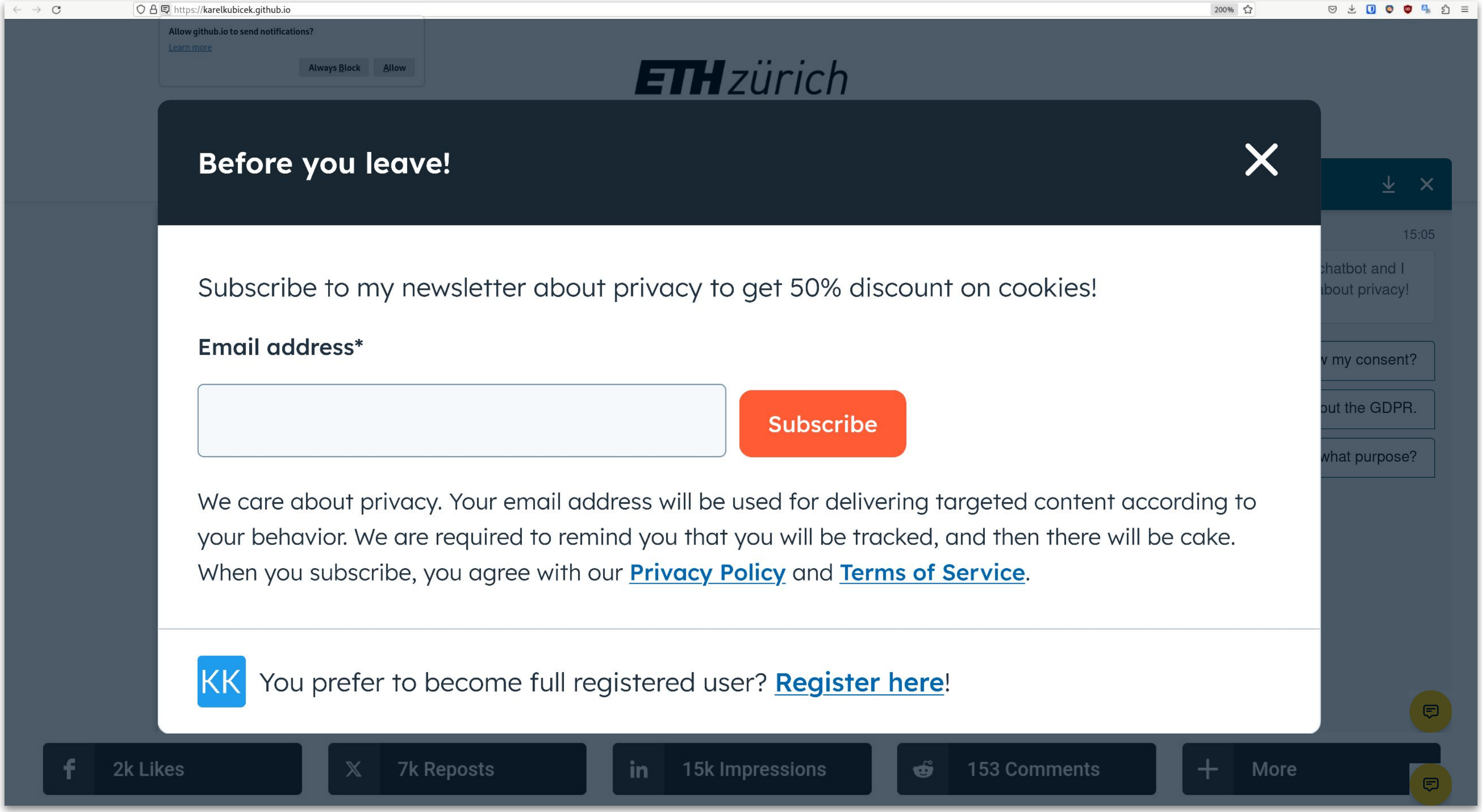
Privacy Enhancing Technologies

Karel Kubicek

karel.kubicek@inria.fr

Some slides inspired from Carmela Troncoso

About me



Before you leave!

Subscribe to my newsletter about privacy to get 50% discount on cookies!

Email address*

We care about privacy. Your email address will be used for delivering targeted content according to your behavior. We are required to remind you that you will be tracked, and then there will be cake. When you subscribe, you agree with our Privacy Policy and Terms of Service.

You prefer to become full registered user? Register here!

About me (for real)

- PhD from ETH Zurich on web tracking compliance
  - Cookies
  - Emails
  - New tracking technologies
- About to start as a researcher in tracking detecting company Vault JS

Outline slide: Motivation

How much online activity are people engaged in?

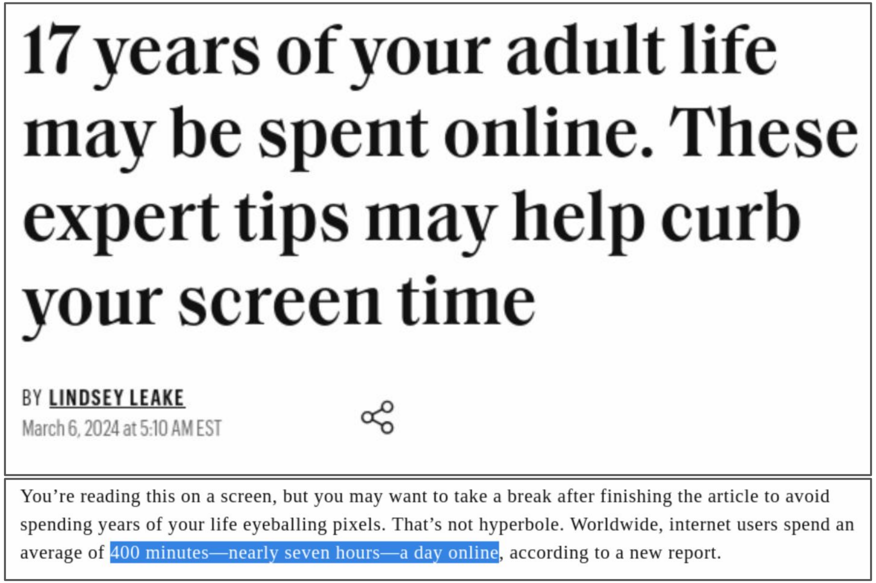
Size of online marketing industry

- martech data (how it grows)?
- total revenue?

Examples of deanonymization (dog joke?)

- dating apps selling info
- TODO: watch last week tonight on data brokers

Motivation



17 years of your adult life may be spent online. These expert tips may help curb your screen time

BY LINDSEY LEAKE

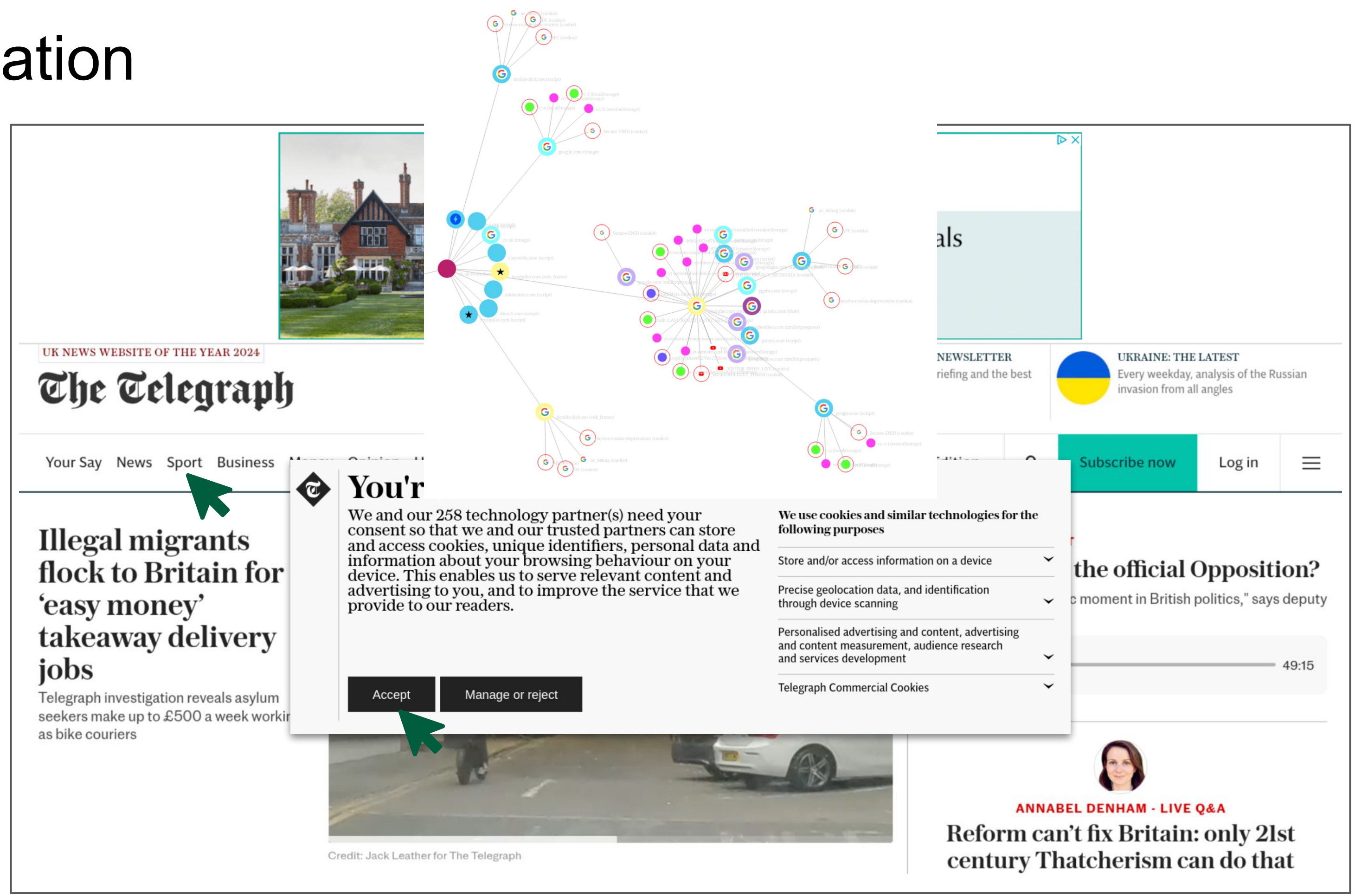
March 6, 2024 at 5:10 AM EST

You're reading this on a screen, but you may want to take a break after finishing the article to avoid spending years of your life eyeballing pixels. That's not hyperbole. Worldwide, internet users spend an average of 400 minutes—nearly seven hours—a day online, according to a new report.

Motivation

“On the Internet, nobody knows you’re a dog.”

Motivation



UK NEWS WEBSITE OF THE YEAR 2024

The Telegraph

Illegal migrants flock to Britain for ‘easy money’ takeaway delivery jobs

Telegraph investigation reveals asylum seekers make up to £500 a week workir as bike couriers

You'r

We and our 258 technology partner(s) need your consent so that we and our trusted partners can store and access cookies, unique identifiers, personal data and information about your browsing behaviour on your device. This enables us to serve relevant content and advertising to you, and to improve the service that we provide to our readers.

We use cookies and similar technologies for the following purposes

- Store and/or access information on a device
- Precise geolocation data, and identification through device scanning
- Personalised advertising and content, advertising and content measurement, audience research and services development
- Telegraph Commercial Cookies

NEWSLETTER

riefing and the best

UKRAINE: THE LATEST

Every weekday, analysis of the Russian invasion from all angles

the official Opposition?

c moment in British politics," says deputy

49:15

Credit: Jack Leather for The Telegraph

ANNABEL DENHAM - LIVE Q&A

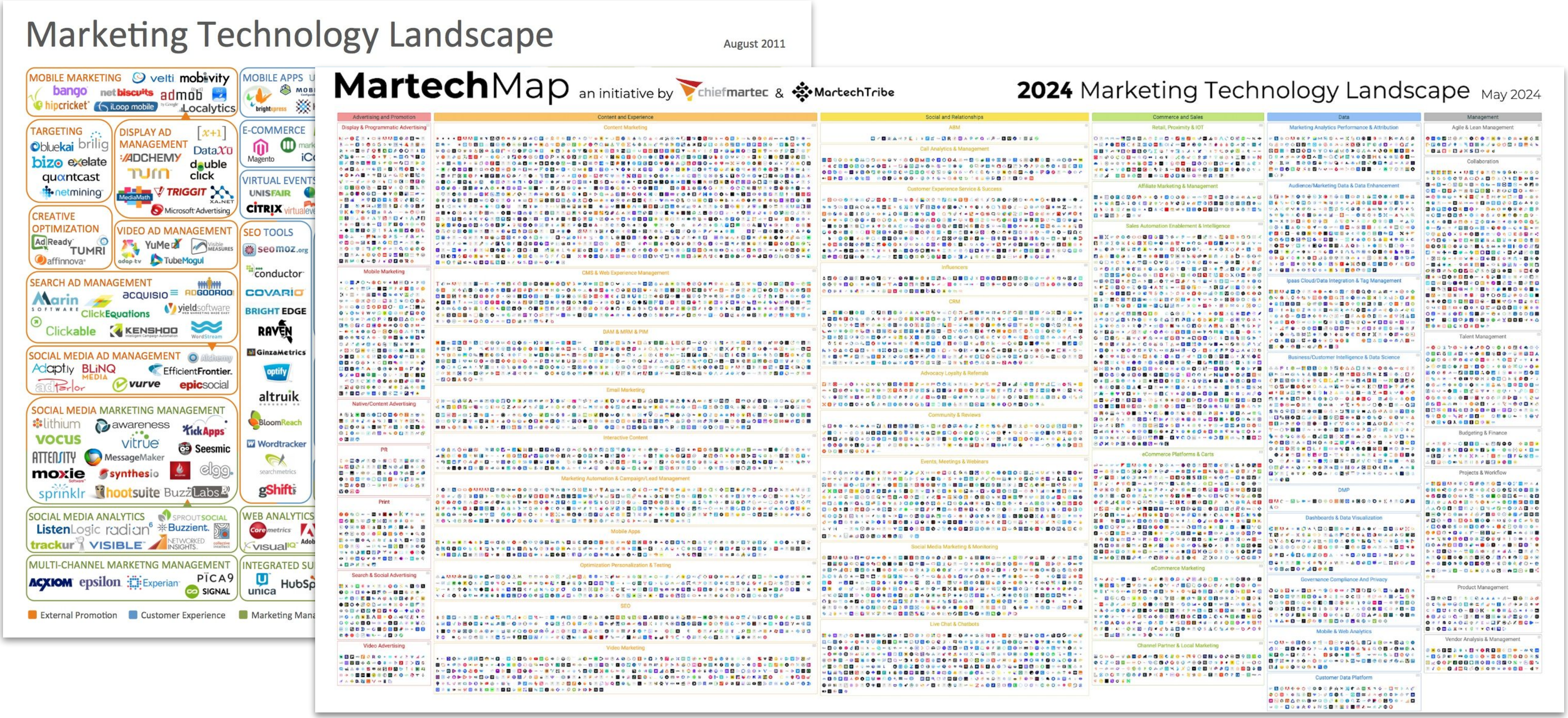
Reform can’t fix Britain: only 21st century Thatcherism can do that

Motivation

“If something is free, you’re not the customer; you’re the product.”

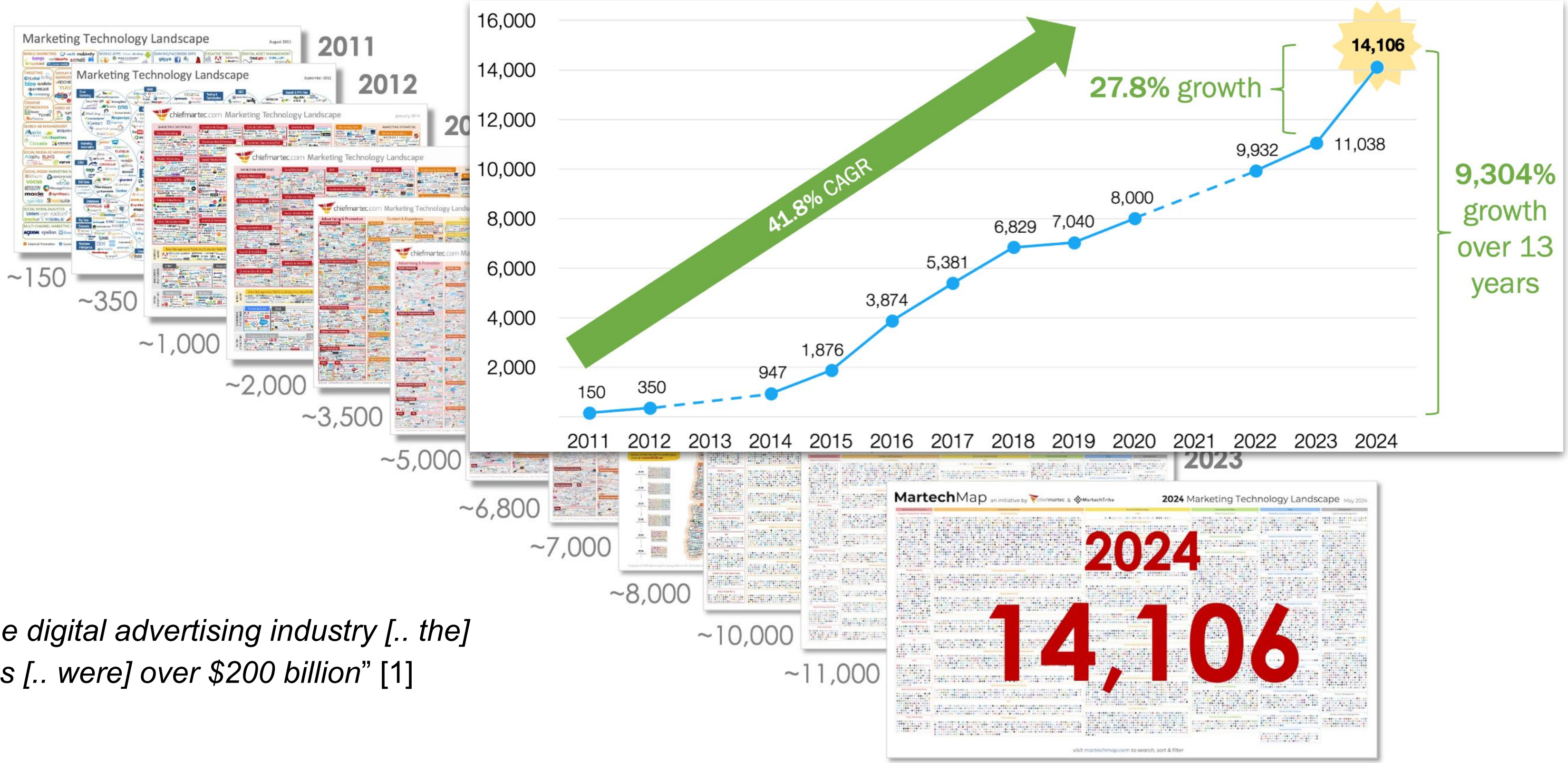
— Bruce Schneier, Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World

Marketing industry



https://chiefmartec.com

Marketing industry

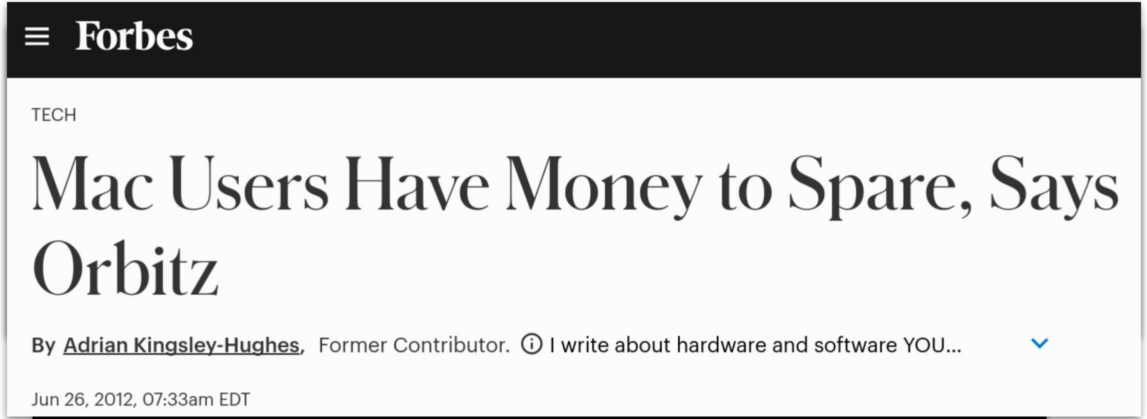


"In 2022, the digital advertising industry [.. the] ad revenues [.. were] over $200 billion" [1]

https://chiefmartec.com

[1] https://www.iab.com/wp-content/uploads/2023/04/IAB_PwC_Internet_Advertising_Revenue_Report_2022.pdf

“I don't care, I do not interact with ads”



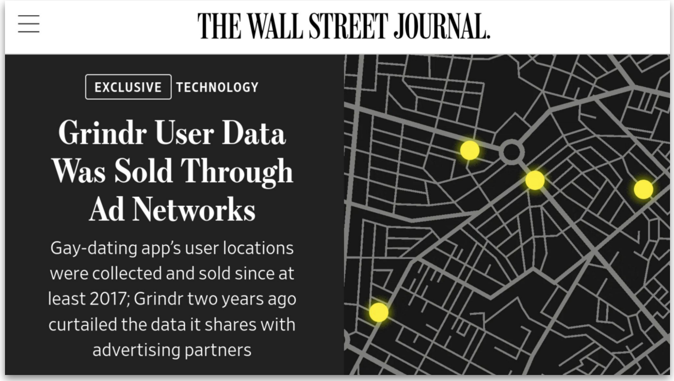
Forbes

TECH

Mac Users Have Money to Spare, Says Orbitz

By Adrian Kingsley-Hughes, Former Contributor. ⓘ I write about hardware and software YOU...

Jun 26, 2012, 07:33am EDT

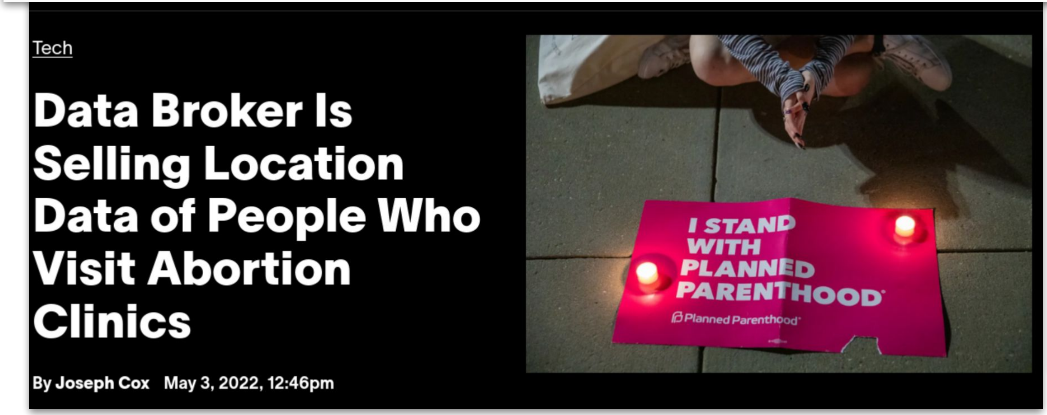


THE WALL STREET JOURNAL.

EXCLUSIVE TECHNOLOGY

Grindr User Data Was Sold Through Ad Networks

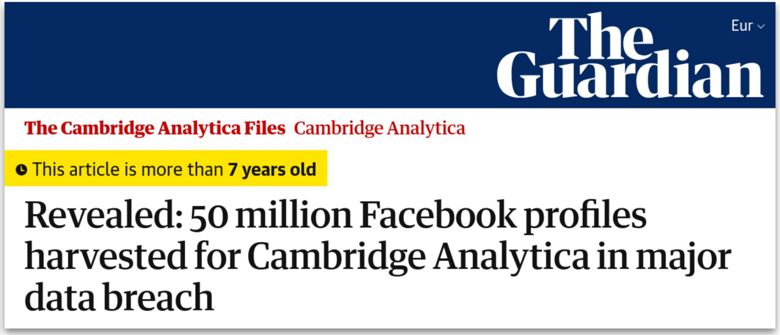
Gay-dating app's user locations were collected and sold since at least 2017; Grindr two years ago curtailed the data it shares with advertising partners



Tech

Data Broker Is Selling Location Data of People Who Visit Abortion Clinics

By Joseph Cox May 3, 2022, 12:46pm



The Guardian

Eur

The Cambridge Analytica Files Cambridge Analytica

This article is more than **7 years old**

Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

Motivation

~~"On the Internet, nobody knows you're a dog."~~

©The New Yorker Collection 1993 Peter Steiner
From cartoonbank.com. All rights reserved.

It's the Internet! Of course they know you're a dog. They also know your favorite brand of pet food and the name of the cute poodle at the park that you have a crush on!

Goals of this lecture

- Understand the technology (how are user profiles built):

  You are the experts - you should help spread the word, help your peers

  - Stateful and stateless tracking
- Give you examples of tracking industry's power, jutifying PETs
- Utility-privacy tradeoffs in online technologies

Web stack

Explain web stack

HTML, CSS, JavaScript

Web servers were meant to be stateless

To keep state, there are multiple technologies to identify user

    Cookies - basic technology, over 80-90% [cite Roesner, 2] of websites use them for tracking

    Other storages: localStorage, sessionStorage, IndexedDB, Cache API

        Lot of JS magic: local variables, data attached to DOM, listeners

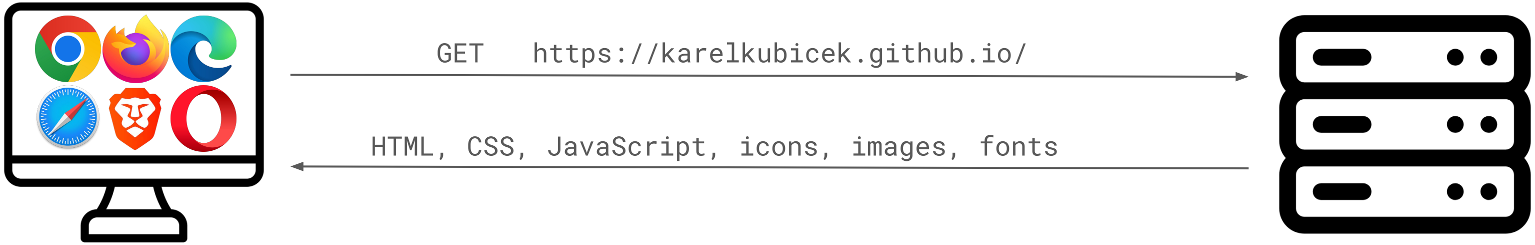
    For syncing: requests parameters, redirects

    Browser fingerprint

    CNAME cloaking

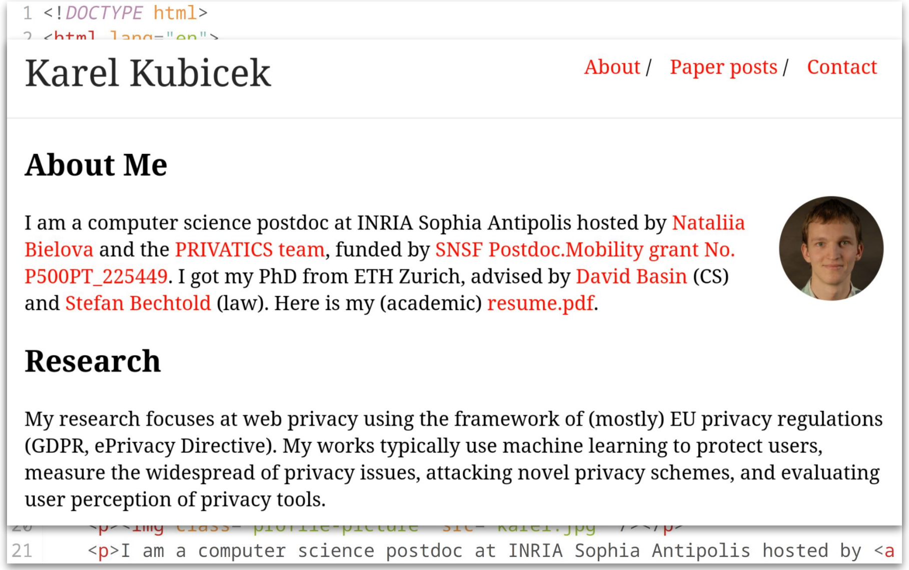
    Form sniffing for email addresses and similar identifiers

Web technologies (reminder)



GET https://karelkubicek.github.io/

HTML, CSS, JavaScript, icons, images, fonts



```
<!DOCTYPE html>
<html lang="en">
```

Karel Kubicek

About Me

I am a computer science postdoc at INRIA Sophia Antipolis hosted by Nataliia Bielova and the PRIVATICS team, funded by SNSF Postdoc.Mobility grant No. P500PT_225449. I got my PhD from ETH Zurich, advised by David Basin (CS) and Stefan Bechtold (law). Here is my (academic) resume.pdf.

Research

My research focuses at web privacy using the framework of (mostly) EU privacy regulations (GDPR, ePrivacy Directive). My works typically use machine learning to protect users, measure the widespread of privacy issues, attacking novel privacy schemes, and evaluating user perception of privacy tools.

```
    <p><img class="profile-picture" src="karel.jpg" /></p>
    <p>I am a computer science postdoc at INRIA Sophia Antipolis hosted by <a
```

HTTP: HyperText Transfer Protocol

HTML: HyperText Markup Language

CSS: Cascading Style Sheets

JavaScript: Logic and interaction

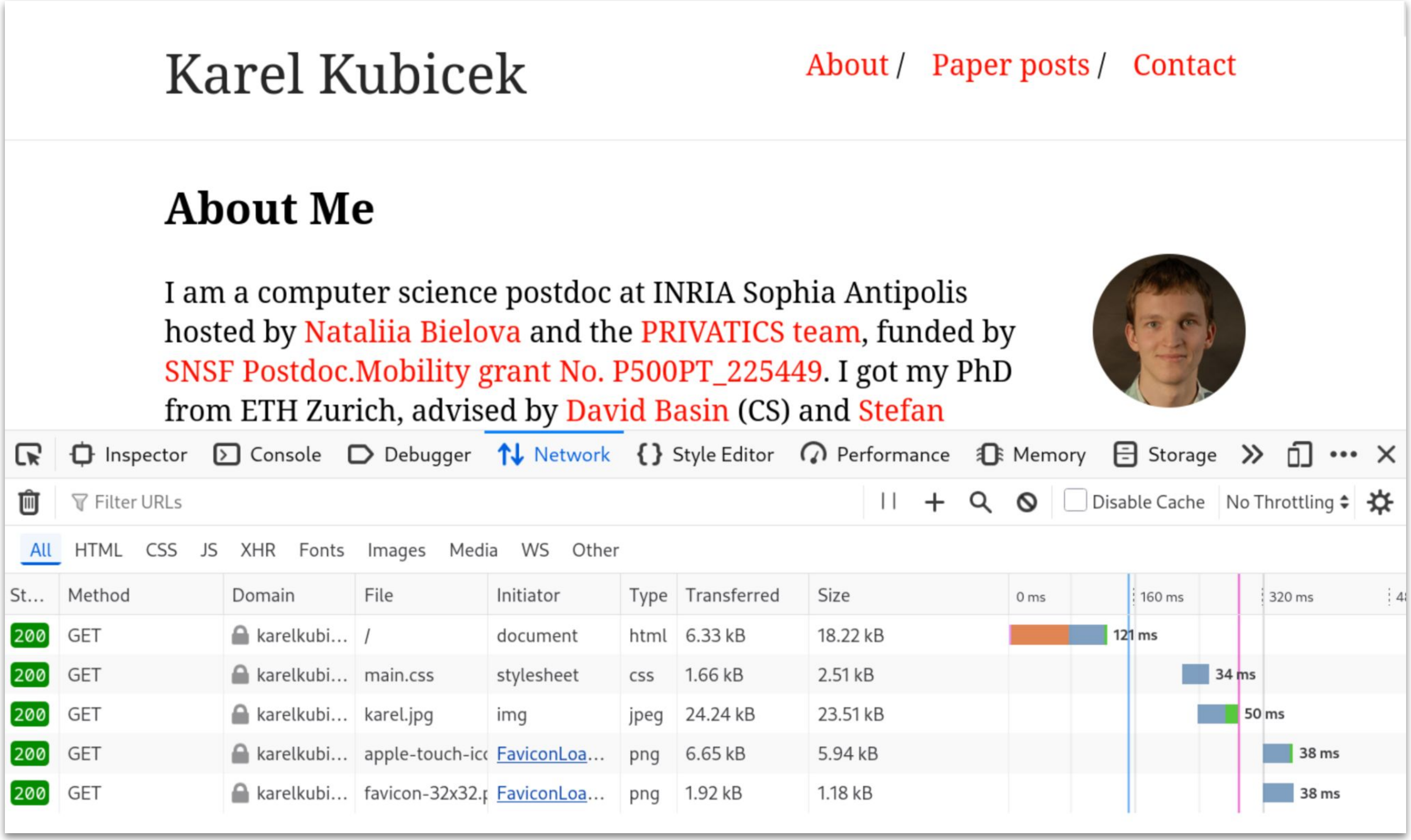
DOM: Document Object Model -

the result of interpreting page code

modifiable by JavaScript

Computer by ratubilqis1986 from Noun Project (CC BY 3.0); Server by Ricons from Noun Project (CC BY 3.0)

DevTools in browser



Karel Kubicek

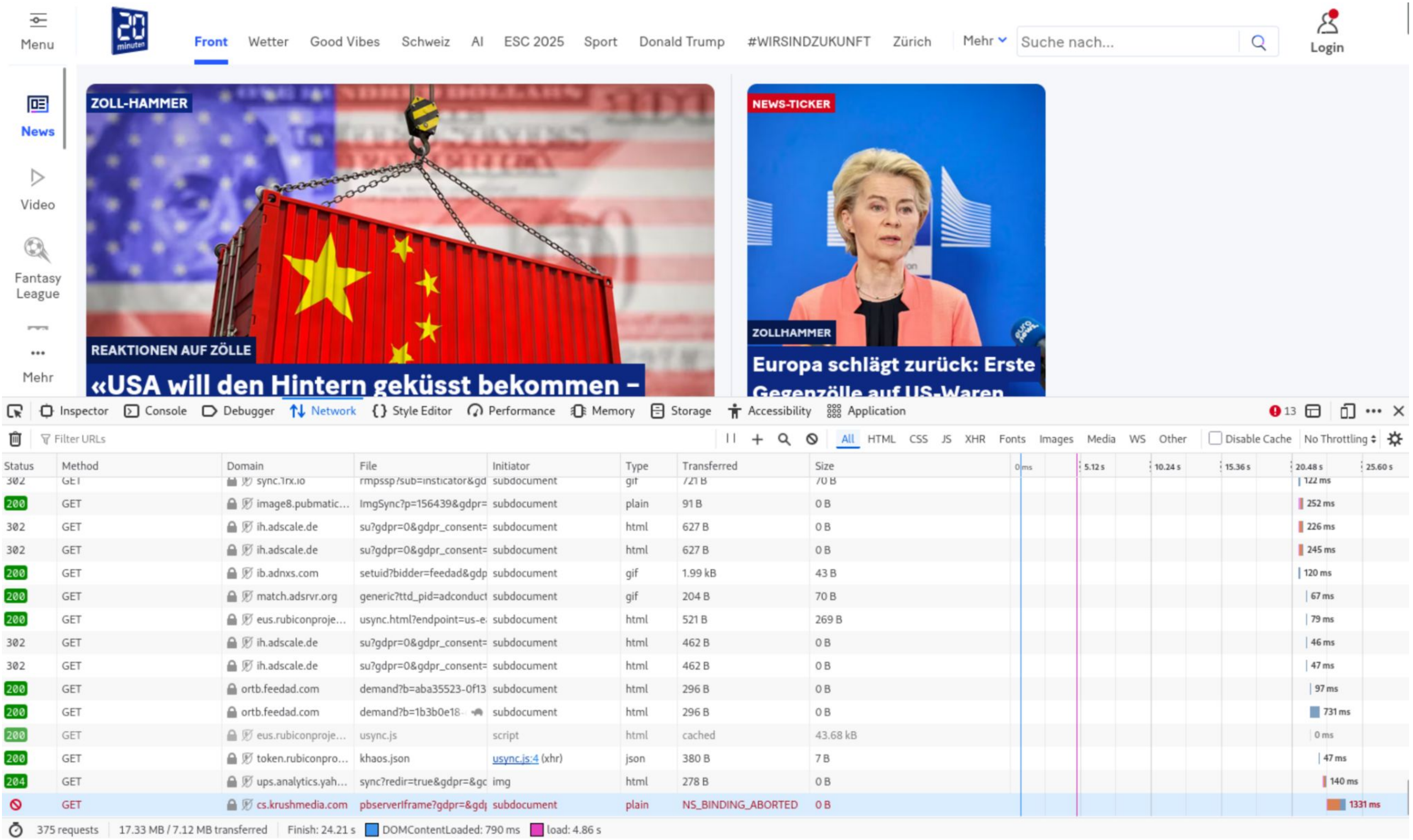
About / Paper posts / Contact

About Me

I am a computer science postdoc at INRIA Sophia Antipolis hosted by Nataliia Bielova and the PRIVATICS team, funded by SNSF Postdoc.Mobility grant No. P500PT_225449. I got my PhD from ETH Zurich, advised by David Basin (CS) and Stefan

St...	Method	Domain	File	Initiator	Type	Transferred	Size	0 ms 160 ms 320 ms 4
200	GET	karelkubi...	/	document	html	6.33 kB	18.22 kB	121 ms
200	GET	karelkubi...	main.css	stylesheet	css	1.66 kB	2.51 kB	34 ms
200	GET	karelkubi...	karel.jpg	img	jpeg	24.24 kB	23.51 kB	50 ms
200	GET	karelkubi...	apple-touch-icc	FaviconLoa...	png	6.65 kB	5.94 kB	38 ms
200	GET	karelkubi...	favicon-32x32.p	FaviconLoa...	png	1.92 kB	1.18 kB	38 ms

DevTools in browser



ZOLL-HAMMER

REAKTIONEN AUF ZÖLLE

«USA will den Hintern geküsst bekommen –

NEWS-TICKER

ZOLLHAMMER

Europa schlägt zurück: Erste Gegenzölle auf US-Waren

Status	Method	Domain	File	Initiator	Type	Transferred	Size	
302	GET	sync.1rx.io	rmpssp?sub=insticator&gd	subdocument	gif	721 B	70 B	122 ms
200	GET	image8.pubmatic...	ImgSync?p=156439&gdpr=	subdocument	plain	91 B	0 B	252 ms
302	GET	ih.adscale.de	su?gdpr=0&gdpr_consent=	subdocument	html	627 B	0 B	226 ms
302	GET	ih.adscale.de	su?gdpr=0&gdpr_consent=	subdocument	html	627 B	0 B	245 ms
200	GET	ib.adnxs.com	setuid?bidder=feedad&gdp	subdocument	gif	1.99 kB	43 B	120 ms
200	GET	match.adsrvr.org	generic?ttd_pid=adconduct	subdocument	gif	204 B	70 B	67 ms
200	GET	eus.rubiconproje...	usync.html?endpoint=us-e	subdocument	html	521 B	269 B	79 ms
302	GET	ih.adscale.de	su?gdpr=0&gdpr_consent=	subdocument	html	462 B	0 B	46 ms
302	GET	ih.adscale.de	su?gdpr=0&gdpr_consent=	subdocument	html	462 B	0 B	47 ms
200	GET	ortb.feedad.com	demand?b=aba35523-0f13	subdocument	html	296 B	0 B	97 ms
200	GET	ortb.feedad.com	demand?b=1b3b0e18-	subdocument	html	296 B	0 B	731 ms
200	GET	eus.rubiconproje...	usync.js	script	html	cached	43.68 kB	0 ms
200	GET	token.rubiconpro...	khaos.json	usync.js:4 (xhr)	json	380 B	7 B	47 ms
204	GET	ups.analytics.yah...	sync?redir=true&gdpr=&gc	img	html	278 B	0 B	140 ms
⃠	GET	cs.krushmedia.com	pbserverIframe?gdpr=&gd	subdocument	plain	NS_BINDING_ABORTED	0 B	1331 ms

375 requests | 17.33 MB / 7.12 MB transferred | Finish: 24.21 s | DOMContentLoaded: 790 ms | load: 4.86 s

Demo

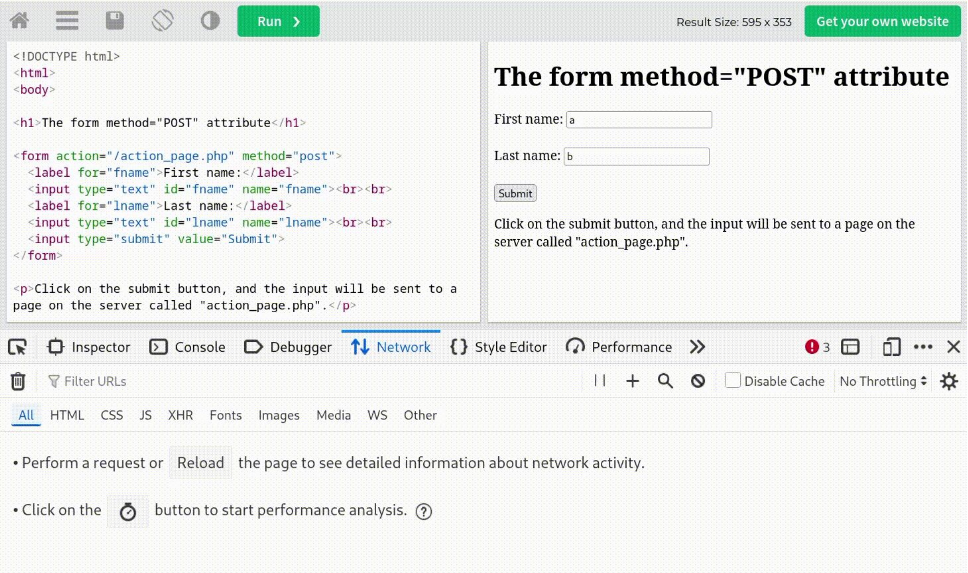
Goals:

- Show different parts of DevTools
- Touch DOM using console
- Show how many third parties are used (from CMP to ads)
- Show various storages, slide for it later

HTTP (slide supplements live demo)

Protocol for requesting and serving (typically) web resources

Browser - server communication



```
<!DOCTYPE html>
<html>
<body>

<h1>The form method="POST" attribute</h1>

<form action="/action_page.php" method="post">
  <label for="fname">First name:</label>
  <input type="text" id="fname" name="fname"><br><br>
  <label for="lname">Last name:</label>
  <input type="text" id="lname" name="lname"><br><br>
  <input type="submit" value="Submit">
</form>

<p>Click on the submit button, and the input will be sent to a
page on the server called "action_page.php".</p>
```

The form method="POST" attribute

First name: a

Last name: b

Submit

Click on the submit button, and the input will be sent to a page on the server called "action_page.php".

- Perform a request or Reload the page to see detailed information about network activity.
- Click on the button to start performance analysis.

Two types of requests:

- GET requests:
  - The majority of requests are GET, invoked by almost everything (images, source files, fonts, typically also trackers)
  - Attributes in URL attributes: `http://example.com?atr=val&id=123`
- POST requests:
  - Invoked by `<form>` submission, attributes hidden in body

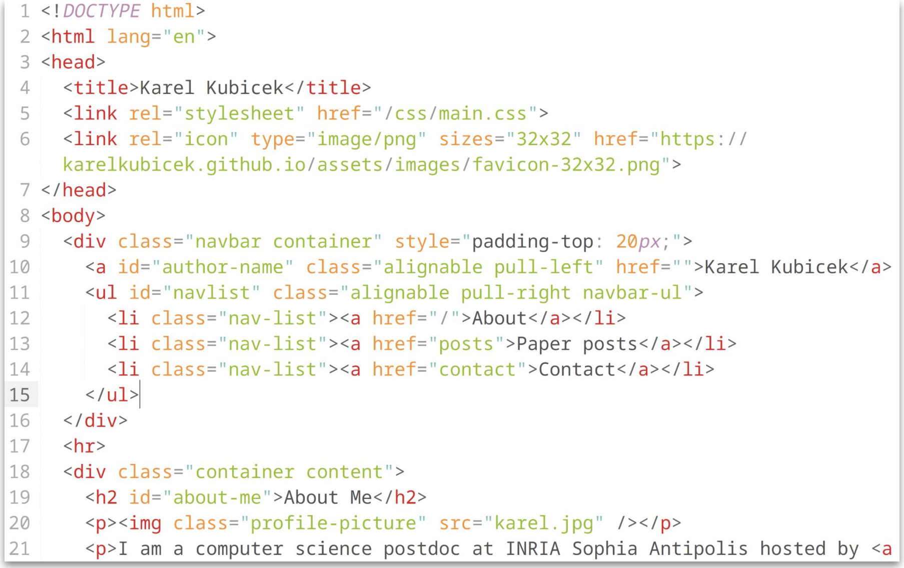
HTML (slide supplements live demo)

Markup language:

- Describes page content and structure
- Invokes loads of other media

Visuals are complemented by CSS

Logic is complemented by JavaScript



```html
<!DOCTYPE html>
<html lang="en">
<head>
  <title>Karel Kubicek</title>
  <link rel="stylesheet" href="/css/main.css">
  <link rel="icon" type="image/png" sizes="32x32" href="https://karelkubicek.github.io/assets/images/favicon-32x32.png">
</head>
<body>
  <div class="navbar container" style="padding-top: 20px;">
    <a id="author-name" class="alignable pull-left" href="">Karel Kubicek</a>
    <ul id="navlist" class="alignable pull-right navbar-ul">
      <li class="nav-list"><a href="/">About</a></li>
      <li class="nav-list"><a href="posts">Paper posts</a></li>
      <li class="nav-list"><a href="contact">Contact</a></li>
    </ul>
  </div>
  <hr>
  <div class="container content">
    <h2 id="about-me">About Me</h2>
    <p><img class="profile-picture" src="karel.jpg" /></p>
    <p>I am a computer science postdoc at INRIA Sophia Antipolis hosted by <a
```

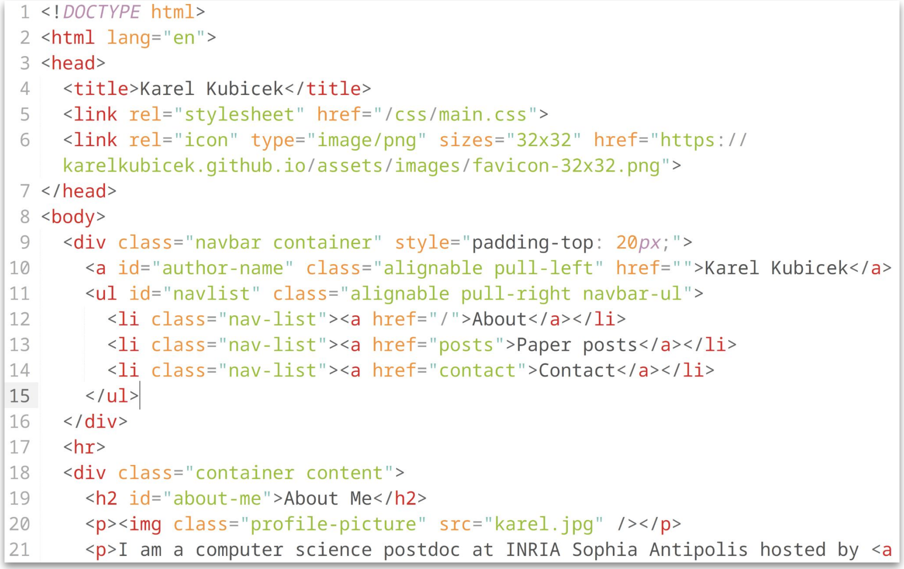
JavaScript (slide supplements live demo)

Programming language:

- Defines logic of website, interactions
- Can make requests or observe them

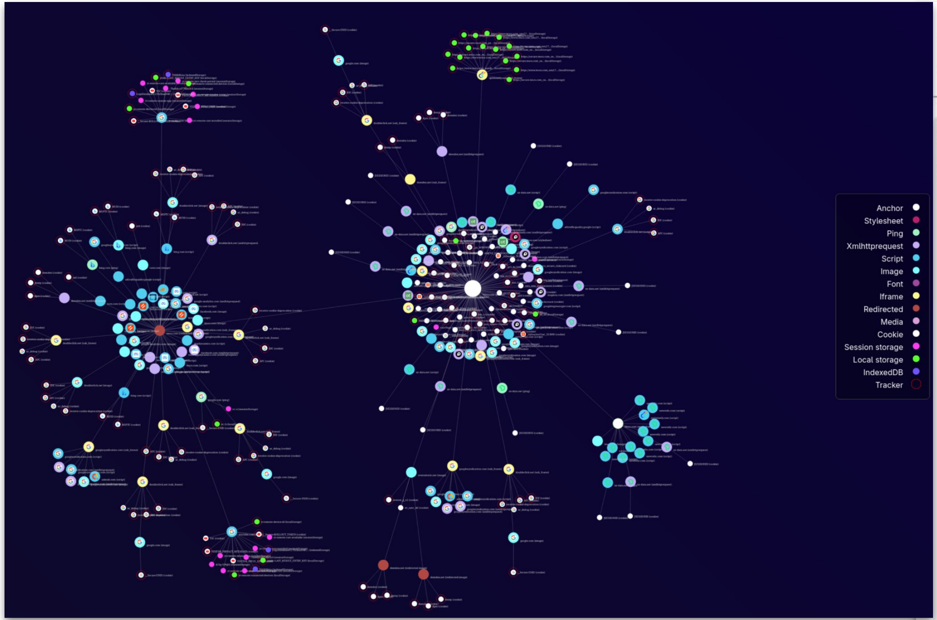
Powerful, can read and manipulate:

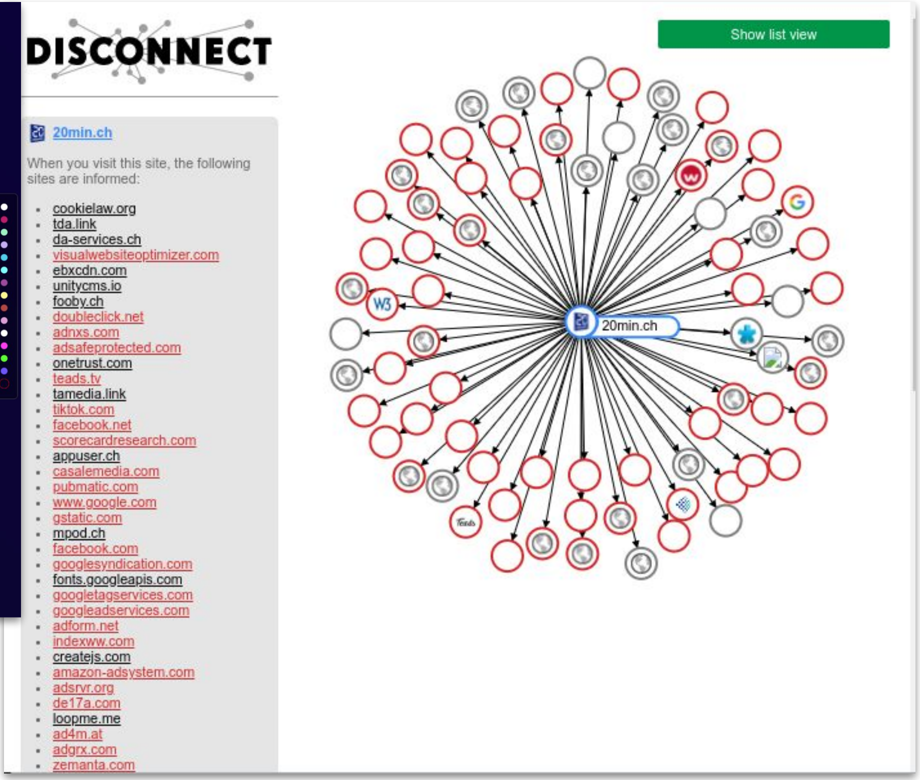
- DOM (how is website rendered)
- Browser storages, API, properties
- Watch almost any event
  (mouse movement, typing, network communication, etc.)



```html
<!DOCTYPE html>
<html lang="en">
<head>
  <title>Karel Kubicek</title>
  <link rel="stylesheet" href="/css/main.css">
  <link rel="icon" type="image/png" sizes="32x32" href="https://
  karelkubicek.github.io/assets/images/favicon-32x32.png">
</head>
<body>
  <div class="navbar container" style="padding-top: 20px;">
    <a id="author-name" class="alignable pull-left" href="">Karel Kubicek</a>
    <ul id="navlist" class="alignable pull-right navbar-ul">
      <li class="nav-list"><a href="/">About</a></li>
      <li class="nav-list"><a href="posts">Paper posts</a></li>
      <li class="nav-list"><a href="contact">Contact</a></li>
    </ul>
  </div>
  <hr>
  <div class="container content">
    <h2 id="about-me">About Me</h2>
    <p><img class="profile-picture" src="karel.jpg" /></p>
    <p>I am a computer science postdoc at INRIA Sophia Antipolis hosted by <a
```

HTTP request chains and parameters





20min.ch

When you visit this site, the following sites are informed:

- cookielaw.org
- tda.link
- da-services.ch
- visualwebsiteoptimizer.com
- ebxcdn.com
- unitycms.io
- fooby.ch
- doubleclick.net
- adnxs.com
- adsafeprotected.com
- onetrust.com
- teads.tv
- tamedia.link
- tiktok.com
- facebook.net
- scorecardresearch.com
- appuser.ch
- casalemedia.com
- pubmatic.com
- www.google.com
- gstatic.com
- mpod.ch
- facebook.com
- googlesyndication.com
- fonts.googleapis.com
- googletagservices.com
- googleadservices.com
- adform.net
- indexww.com
- createjs.com
- amazon-adsystem.com
- adsrvr.org
- de17a.com
- loopme.me
- ad4m.at
- adgrx.com
- zemanta.com

By analyzing the third party trees, we found that the median depth of such trees is one (max eight) [.. and] especially ad networks result in longer tree branches, and that only 7 % of all visited websites never embedded a third party that might pose possible legal problems. [1]

Visual on the left: https://baycloud.com/ (scan of tesco.com)

Urban, T., Degeling, M., Holz, T., & Pohlmann, N. (2020, April). Beyond the front page: Measuring third party dynamics in the field. In *Proceedings of the web conference 2020*.

Browser storages

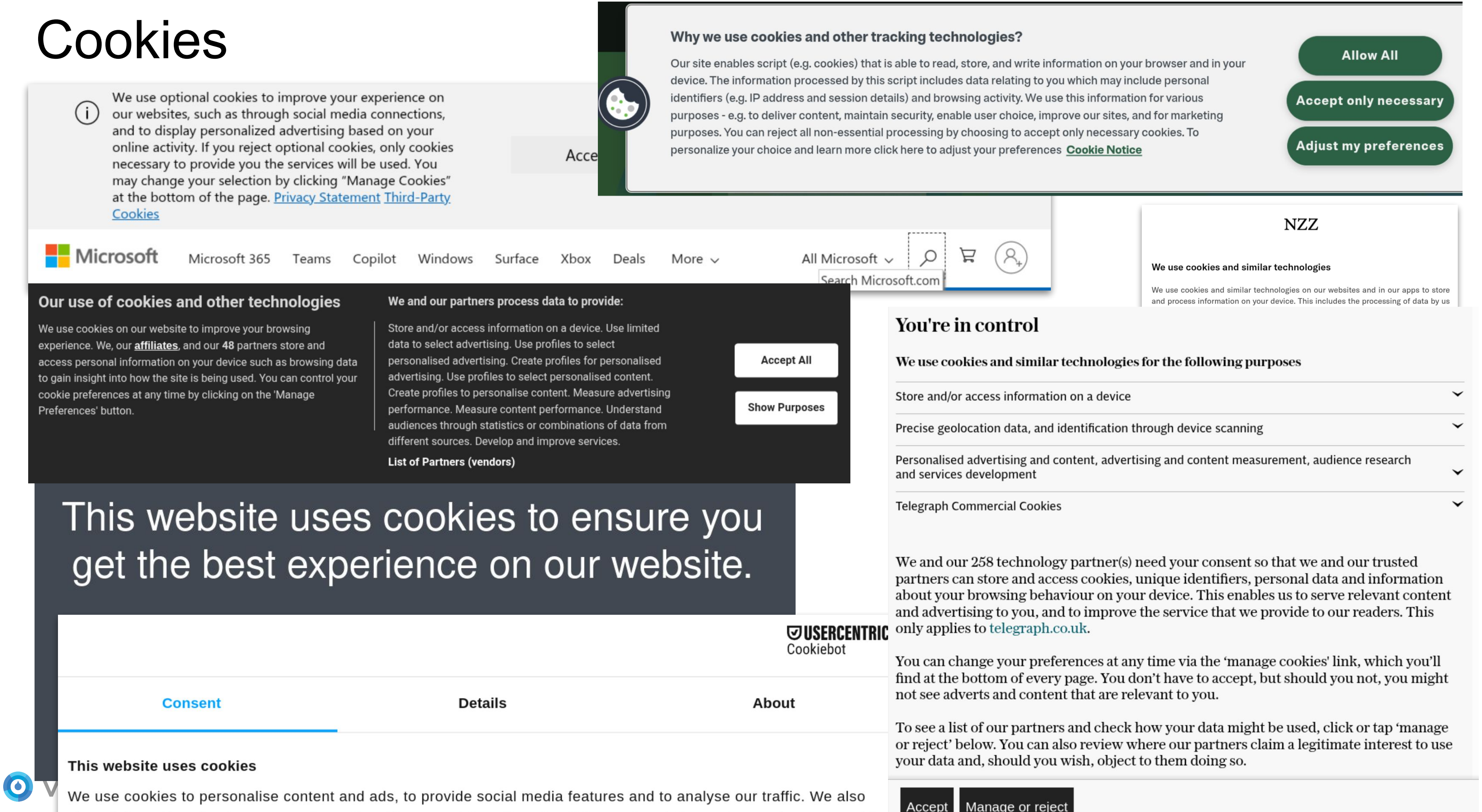
HTTP is stateless, and servers were meant to be as well

Browser storages

HTTP is stateless, and servers were meant to be as well

- Cookies
- Other storages: localStorage, sessionStorage, IndexedDB, Cache API,
  Lot of JS magic: local variables, data attached to DOM, listeners

Cookies



ⓘ We use optional cookies to improve your experience on our websites, such as through social media connections, and to display personalized advertising based on your online activity. If you reject optional cookies, only cookies necessary to provide you the services will be used. You may change your selection by clicking "Manage Cookies" at the bottom of the page. Privacy Statement Third-Party Cookies

Why we use cookies and other tracking technologies?

Our site enables script (e.g. cookies) that is able to read, store, and write information on your browser and in your device. The information processed by this script includes data relating to you which may include personal identifiers (e.g. IP address and session details) and browsing activity. We use this information for various purposes - e.g. to deliver content, maintain security, enable user choice, improve our sites, and for marketing purposes. You can reject all non-essential processing by choosing to accept only necessary cookies. To personalize your choice and learn more click here to adjust your preferences Cookie Notice

NZZ

We use cookies and similar technologies

We use cookies and similar technologies on our websites and in our apps to store and process information on your device. This includes the processing of data by us

Our use of cookies and other technologies

We use cookies on our website to improve your browsing experience. We, our affiliates, and our **48** partners store and access personal information on your device such as browsing data to gain insight into how the site is being used. You can control your cookie preferences at any time by clicking on the 'Manage Preferences' button.

We and our partners process data to provide:

Store and/or access information on a device. Use limited data to select advertising. Use profiles to select personalised advertising. Create profiles for personalised advertising. Use profiles to select personalised content. Create profiles to personalise content. Measure advertising performance. Measure content performance. Understand audiences through statistics or combinations of data from different sources. Develop and improve services.

List of Partners (vendors)

This website uses cookies to ensure you get the best experience on our website.

This website uses cookies

We use cookies to personalise content and ads, to provide social media features and to analyse our traffic. We also

You're in control

We use cookies and similar technologies for the following purposes

Store and/or access information on a device

Precise geolocation data, and identification through device scanning

Personalised advertising and content, advertising and content measurement, audience research and services development

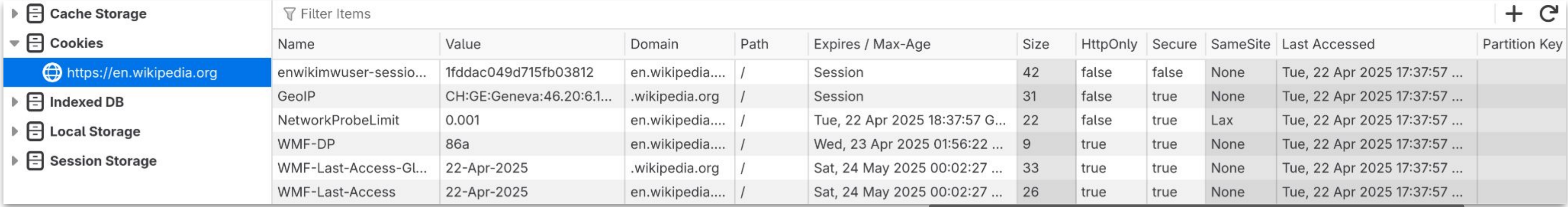
Telegraph Commercial Cookies

We and our 258 technology partner(s) need your consent so that we and our trusted partners can store and access cookies, unique identifiers, personal data and information about your browsing behaviour on your device. This enables us to serve relevant content and advertising to you, and to improve the service that we provide to our readers. This only applies to telegraph.co.uk.

You can change your preferences at any time via the 'manage cookies' link, which you'll find at the bottom of every page. You don't have to accept, but should you not, you might not see adverts and content that are relevant to you.

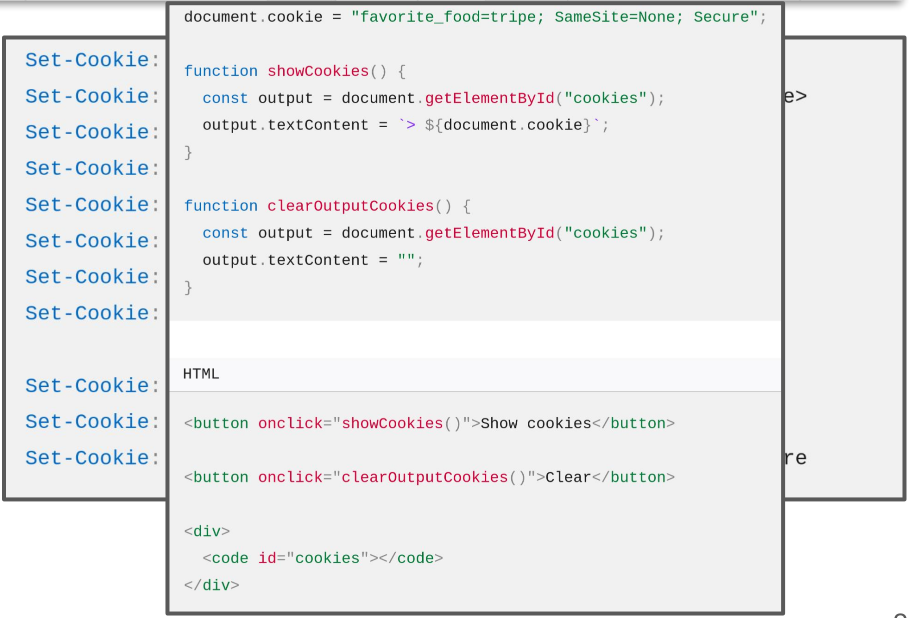
To see a list of our partners and check how your data might be used, click or tap 'manage or reject' below. You can also review where our partners claim a legitimate interest to use your data and, should you wish, object to them doing so.

Cookies



Name	Value	Domain	Path	Expires / Max-Age	Size	HttpOnly	Secure	SameSite	Last Accessed	Partition Key
enwikimwuser-sessio...	1fddac049d715fb03812	en.wikipedia....	/	Session	42	false	false	None	Tue, 22 Apr 2025 17:37:57 ...	
GeoIP	CH:GE:Geneva:46.20:6.1...	.wikipedia.org	/	Session	31	false	true	None	Tue, 22 Apr 2025 17:37:57 ...	
NetworkProbeLimit	0.001	en.wikipedia....	/	Tue, 22 Apr 2025 18:37:57 G...	22	false	true	Lax	Tue, 22 Apr 2025 17:37:57 ...	
WMF-DP	86a	en.wikipedia....	/	Wed, 23 Apr 2025 01:56:22 ...	9	true	true	None	Tue, 22 Apr 2025 17:37:57 ...	
WMF-Last-Access-Gl...	22-Apr-2025	.wikipedia.org	/	Sat, 24 May 2025 00:02:27 ...	33	true	true	None	Tue, 22 Apr 2025 17:37:57 ...	
WMF-Last-Access	22-Apr-2025	en.wikipedia....	/	Sat, 24 May 2025 00:02:27 ...	26	true	true	None	Tue, 22 Apr 2025 17:37:57 ...	

- Key-value pairs (variables)
- 80-90% websites track using cookies ([1] and [2] in 2012 and 2020, resp.)
  I.e., all top 1k websites track you, maybe with exception of `nitter.com`
- Set by request or JavaScript and sent to website with every request matching domain



```
Set-Cookie:
Set-Cookie:                                                          e>
Set-Cookie:
Set-Cookie:
Set-Cookie:
Set-Cookie:
Set-Cookie:
Set-Cookie:

Set-Cookie:
Set-Cookie:
Set-Cookie:                                                          re
```

```
document.cookie = "favorite_food=tripe; SameSite=None; Secure";

function showCookies() {
  const output = document.getElementById("cookies");
  output.textContent = `> ${document.cookie}`;
}

function clearOutputCookies() {
  const output = document.getElementById("cookies");
  output.textContent = "";
}
```

HTML

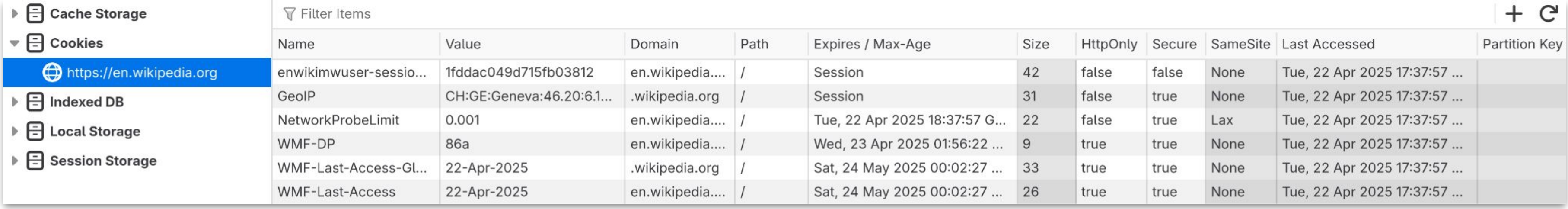
```
<button onclick="showCookies()">Show cookies</button>

<button onclick="clearOutputCookies()">Clear</button>

<div>
  <code id="cookies"></code>
</div>
```

[1] Roesner, Franziska, Tadayoshi Kohno, and David Wetherall. "Detecting and defending against third-party tracking on the web." 9th *USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*. 2012.
[2] Solomos, Konstantinos, et al. "Clash of the trackers: measuring the evolution of the online tracking ecosystem." *Network Traffic Measurement and Analysis Conference (TMA)* (2020)

Cookies



Name	Value	Domain	Path	Expires / Max-Age	Size	HttpOnly	Secure	SameSite	Last Accessed	Partition Key
enwikimwuser-sessio...	1fddac049d715fb03812	en.wikipedia....	/	Session	42	false	false	None	Tue, 22 Apr 2025 17:37:57 ...	
GeoIP	CH:GE:Geneva:46.20:6.1...	.wikipedia.org	/	Session	31	false	true	None	Tue, 22 Apr 2025 17:37:57 ...	
NetworkProbeLimit	0.001	en.wikipedia....	/	Tue, 22 Apr 2025 18:37:57 G...	22	false	true	Lax	Tue, 22 Apr 2025 17:37:57 ...	
WMF-DP	86a	en.wikipedia....	/	Wed, 23 Apr 2025 01:56:22 ...	9	true	true	None	Tue, 22 Apr 2025 17:37:57 ...	
WMF-Last-Access-Gl...	22-Apr-2025	.wikipedia.org	/	Sat, 24 May 2025 00:02:27 ...	33	true	true	None	Tue, 22 Apr 2025 17:37:57 ...	
WMF-Last-Access	22-Apr-2025	en.wikipedia....	/	Sat, 24 May 2025 00:02:27 ...	26	true	true	None	Tue, 22 Apr 2025 17:37:57 ...	

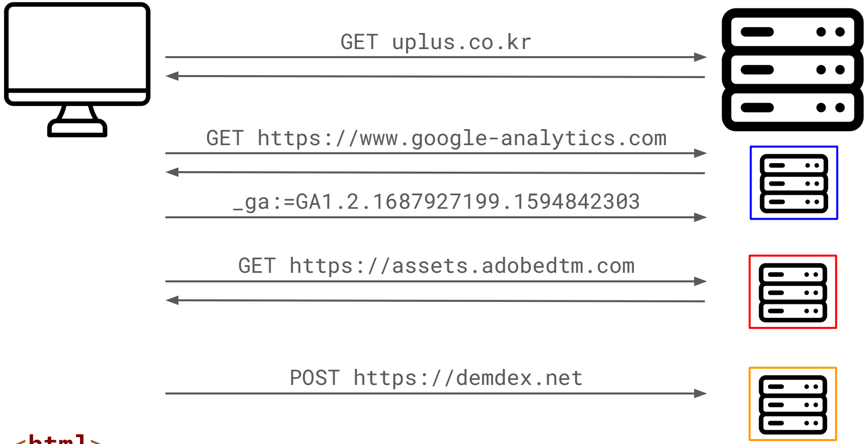
- `Domain` + `Path`: access control mechanism
  - First party: website itself, but also `<script src="tracker.com">`
  - Third party: `<iframe src="tracker.com">`
- `Expiry`: cookie removal time, either a timestamp or when the tab is closed = `Session`
- `HttpOnly`: forbid read/write by JavaScript ("requestOnly")
- `Secure`: can be sent to server only using HTTPS
- `SameSite`:
  - `Strict`: sent only to request with matching domain
  - `Lax`: as `Strict` with exception of sharing to the next one site by user (affiliation links), default
  - `None`: sent everywhere, requires `Secure`

TODO Example cookies

Prepare example cookies:

- Login (first party)
- SSO
- First-party analytics
- Third-party tracker

Cookies exfiltration


GET uplus.co.kr
GET https://www.google-analytics.com
_ga:=GA1.2.1687927199.1594842303
GET https://assets.adobedtm.com
POST https://demdex.net

```html
<html>
  <script src="https://www.google-analytics.com/a.js">

  </script>
  <script src="https://adobedtm.com/satelliteLib.js">




  </script>
</html>
```

“*In total, we found that 97.72% of the websites have first-party cookies that are set by third-party JavaScript, and that on 57.66% of these websites there is at least one such cookie that contains a unique user identifier that is diffused to multiple third parties. Our results highlight the privacy-intrusive capabilities of first-party cookies*” [1]

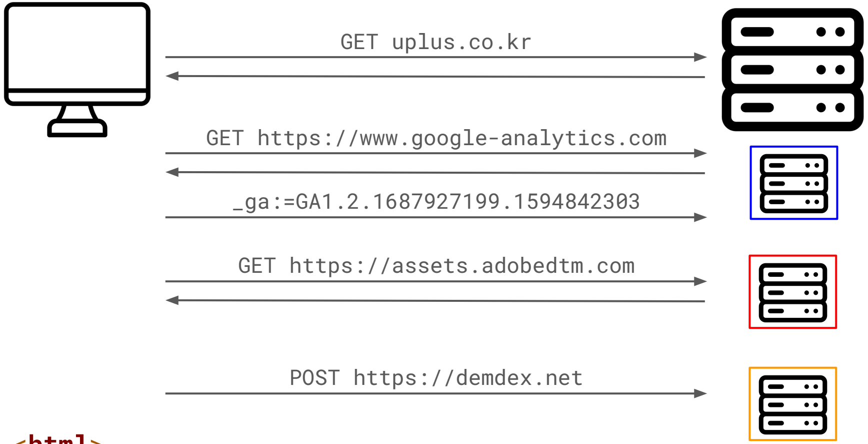
“*Analyzing the browsing histories of 100 volunteers. They found, on average, 60 cookies are synced when a user visits 40 sites. Facebook (facebook.com) and AppNexus (adnxs.com) synced their cookies for 91% of the volunteers.*” [2]

“*They found that 78% of the top 200 websites include 3rd-party scripts which synchronize cookies with at least one other party. These 3rd-party scripts can reconstruct 62-73% of a user’s browsing history.*” [3]

“*They found that 97% of regular web users are exposed to cookie syncing. UserIDs get leaked, on average, to 3.5 different domains. The use of Cookie syncing increases the number of domains that track the user by a factor of 6.75.*” [4]

[1] Chen, Q., Ilia, P., Polychronakis, M., & Kapravelos, A. (2021, April). Cookie swap party: Abusing first-party cookies for web tracking. In *Proceedings of the Web Conference 2021 (WWW).*
[2] Olejnik, L., Minh-Dung, T., & Castelluccia, C. (2013). Selling off privacy at auction. In *Proceedings of the 2013 Network and Distributed System Security Symposium (NDSS).*
[3] Englehardt, S., & Narayanan, A. (2016, October). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM Conference on Computer and Communications Security (CCS).*
[4] Papadopoulos, P., Kourtellis, N., & Markatos, E. (2019, May). Cookie synchronization: Everything you always wanted to know but were afraid to ask. In The World Wide Web Conference (WWW).

Cookies exfiltration (without animation)


GET uplus.co.kr
GET https://www.google-analytics.com
_ga:=GA1.2.1687927199.1594842303
GET https://assets.adobedtm.com
POST https://demdex.net

```
<html>
  <script src="https://www.google-analytics.com/a.js">
    document.cookie="_ga:=GA1.2.1687927199.1594842303";
  </script>
  <script src="https://adobedtm.com/satelliteLib.js">
    gaValue = document.cookie['_ga'];
    var xhr = new XMLHttpRequest();
    xhr.open("POST", "https://demdex.net/", true);
    xhr.send("c_gacid=gaValue"); // sent in POST body
  </script>
</html>
```

“*In total, we found that 97.72% of the websites have first-party cookies that are set by third-party JavaScript, and that on 57.66% of these websites there is at least one such cookie that contains a unique user identifier that is diffused to multiple third parties. Our results highlight the privacy-intrusive capabilities of first-party cookies*” [1]

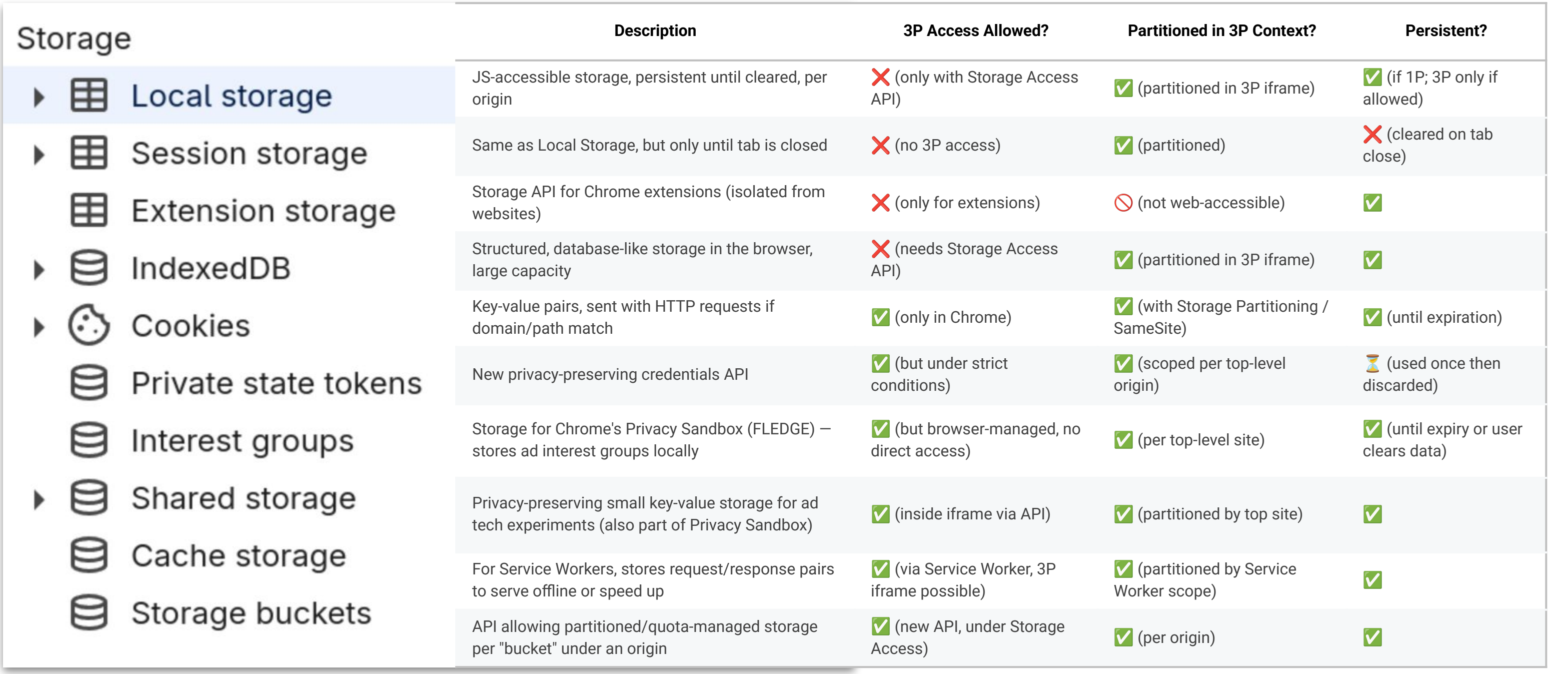
“*Analyzing the browsing histories of 100 volunteers. They found, on average, 60 cookies are synced when a user visits 40 sites. Facebook (facebook.com) and AppNexus (adnxs.com) synced their cookies for 91% of the volunteers.*” [2]

“*They found that 78% of the top 200 websites include 3rd-party scripts which synchronize cookies with at least one other party. These 3rd-party scripts can reconstruct 62-73% of a user’s browsing history.*” [3]

“*They found that 97% of regular web users are exposed to cookie syncing. UserIDs get leaked, on average, to 3.5 different domains. The use of Cookie syncing increases the number of domains that track the user by a factor of 6.75.*” [4]

[1] Chen, Q., Ilia, P., Polychronakis, M., & Kapravelos, A. (2021, April). Cookie swap party: Abusing first-party cookies for web tracking. In *Proceedings of the Web Conference 2021 (WWW).*
[2] Olejnik, L., Minh-Dung, T., & Castelluccia, C. (2013). Selling off privacy at auction. In *Proceedings of the 2013 Network and Distributed System Security Symposium (NDSS).*
[3] Englehardt, S., & Narayanan, A. (2016, October). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM Conference on Computer and Communications Security (CCS).*
[4] Papadopoulos, P., Kourtellis, N., & Markatos, E. (2019, May). Cookie synchronization: Everything you always wanted to know but were afraid to ask. In The World Wide Web Conference (WWW).

Other storages



Storage	Description	3P Access Allowed?	Partitioned in 3P Context?	Persistent?
Local storage	JS-accessible storage, persistent until cleared, per origin	❌ (only with Storage Access API)	✅ (partitioned in 3P iframe)	✅ (if 1P; 3P only if allowed)
Session storage	Same as Local Storage, but only until tab is closed	❌ (no 3P access)	✅ (partitioned)	❌ (cleared on tab close)
Extension storage	Storage API for Chrome extensions (isolated from websites)	❌ (only for extensions)	🚫 (not web-accessible)	✅
IndexedDB	Structured, database-like storage in the browser, large capacity	❌ (needs Storage Access API)	✅ (partitioned in 3P iframe)	✅
Cookies	Key-value pairs, sent with HTTP requests if domain/path match	✅ (only in Chrome)	✅ (with Storage Partitioning / SameSite)	✅ (until expiration)
Private state tokens	New privacy-preserving credentials API	✅ (but under strict conditions)	✅ (scoped per top-level origin)	⏳ (used once then discarded)
Interest groups	Storage for Chrome's Privacy Sandbox (FLEDGE) — stores ad interest groups locally	✅ (but browser-managed, no direct access)	✅ (per top-level site)	✅ (until expiry or user clears data)
Shared storage	Privacy-preserving small key-value storage for ad tech experiments (also part of Privacy Sandbox)	✅ (inside iframe via API)	✅ (partitioned by top site)	✅
Cache storage	For Service Workers, stores request/response pairs to serve offline or speed up	✅ (via Service Worker, 3P iframe possible)	✅ (partitioned by Service Worker scope)	✅
Storage buckets	API allowing partitioned/quota-managed storage per "bucket" under an origin	✅ (new API, under Storage Access)	✅ (per origin)	✅

Also in JavaScript (not persistent):

- local variables
- data attached to DOM
- listeners

Stateless tracking

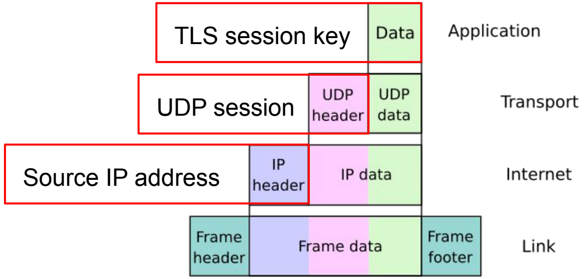
After you remove your cookies

Network stack tracking

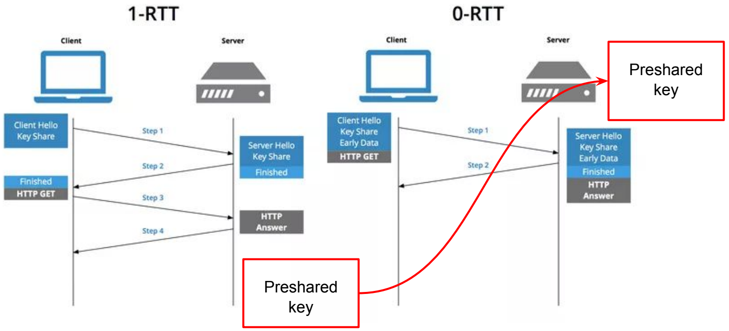
- IPv4: 2^32 ~ 4B addresses
  - You might be hidden behind NAT
- IPv6: 2^128 ~ 42 decimal digits
  - Every device unique

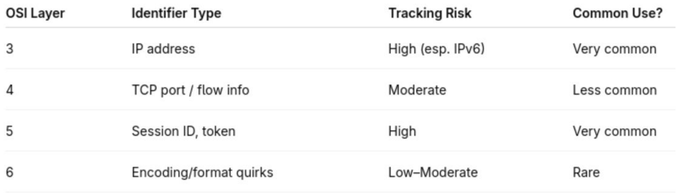
- TCP/UDP session
- TLS session key

- *Internet devices are identifiable by design, it is up to server to honor it*
  - Ex: Google Analytics IP masking
  - Network stack is heavily used for tracking



TLS 1.3 no handshake resumed connection





OSI Layer	Identifier Type	Tracking Risk	Common Use?
3	IP address	High (esp. IPv6)	Very common
4	TCP port / flow info	Moderate	Less common
5	Session ID, token	High	Very common
6	Encoding/format quirks	Low–Moderate	Rare

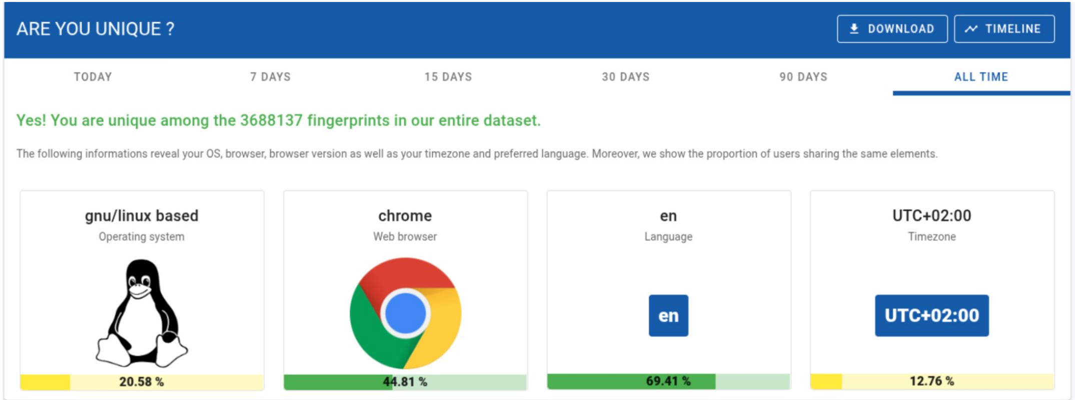
Image sources: https://en.wikipedia.org/wiki/Internet_protocol_suite and https://www.keycdn.com/blog/tls-1-3-support

Browser fingerprinting

- Problem: different devices might support different features
  - Website want to know how to serve you the right content
  - E.g., Windows→executable in .exe, while on Linux→ .deb/.rpm
  - Screen resolution for the right size of content
  - Preferred content language
  - Audio-video codecs
- Usefulness today is questionable

Browser fingerprinting

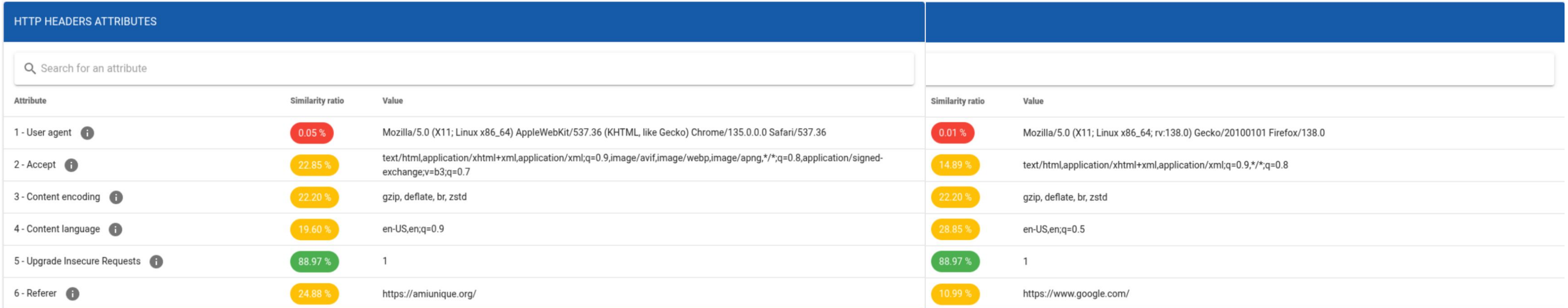
Online testing tools: https://coveryourtracks.eff.org or https://amiunique.org (used here)



Yes! You are unique among the 3688137 fingerprints in our entire dataset.

The following informations reveal your OS, browser, browser version as well as your timezone and preferred language. Moreover, we show the proportion of users sharing the same elements.

gnu/linux based	chrome	en	UTC+02:00
Operating system	Web browser	Language	Timezone
20.58 %	44.81 %	69.41 %	12.76 %



HTTP HEADERS ATTRIBUTES

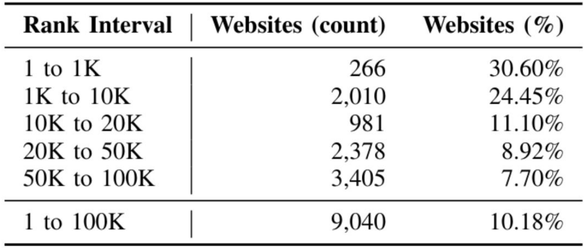
Attribute	Similarity ratio	Value	Similarity ratio	Value
1 - User agent	0.05 %	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/135.0.0.0 Safari/537.36	0.01 %	Mozilla/5.0 (X11; Linux x86_64; rv:138.0) Gecko/20100101 Firefox/138.0
2 - Accept	22.85 %	text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7	14.89 %	text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
3 - Content encoding	22.20 %	gzip, deflate, br, zstd	22.20 %	gzip, deflate, br, zstd
4 - Content language	19.60 %	en-US,en;q=0.9	28.85 %	en-US,en;q=0.5
5 - Upgrade Insecure Requests	88.97 %	1	88.97 %	1
6 - Referer	24.88 %	https://amiunique.org/	10.99 %	https://www.google.com/

Demo of Am I Unique?

Explain:

- HTTP headers attributes
- JS attributes
  - Fonts
  - Canvas
  - WebGL
  - Hardware
  - Browser API permissions

Browser fingerprinting prevalence



Rank Interval	Websites (count)	Websites (%)
1 to 1K	266	30.60%
1K to 10K	2,010	24.45%
10K to 20K	981	11.10%
20K to 50K	2,378	8.92%
50K to 100K	3,405	7.70%
1 to 100K	9,040	10.18%

TABLE IV: Distribution of Alexa top-100K websites that deploy fingerprinting. Results are sliced by site rank.

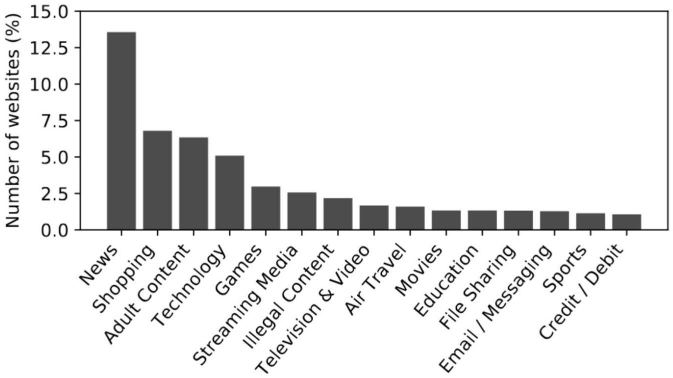


Fig. 4: The deployment of fingerprinting scripts across different categories of websites.

"Overall, we find that more than 10.18% of top-100K websites deploy fingerprinting." [1]

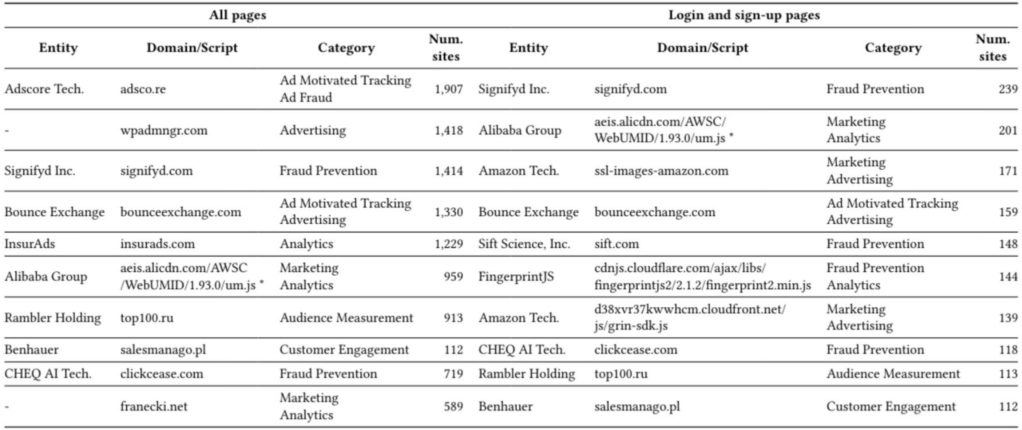
"We found 1,425 cookies respawned using at least one of the studied [browser fingerprinting] features. These cookies were respawned in 1,150 websites that represent 3.83% of the visited websites." [2]

[1] Iqbal, U., Englehardt, S., & Shafiq, Z. (2021, May). Fingerprinting the fingerprinters: Learning to detect browser fingerprinting behaviors. In *2021 IEEE Symposium on Security and Privacy*.
[2] Fouad, I., Santos, C., Legout, A., & Bielova, N. (2022). My Cookie is a phoenix: Detection, measurement, and lawfulness of cookie respawning with browser fingerprinting. In *PETS 2022 Privacy Enhancing Technologies Symposium*.

Privacy-utility trade-off of fingerprinting

Fingerprinting is useful in several scenarios:

- Fraud detection in banking software (against hijacked sessions)
- Bot detection
- Authentication: frictionless authentication or "silent second factor"
    - Google login: if fingerprint is known - no 2FA required



All pages				Login and sign-up pages			
Entity	Domain/Script	Category	Num. sites	Entity	Domain/Script	Category	Num. sites
Adscore Tech.	adsco.re	Ad Motivated Tracking Ad Fraud	1,907	Signifyd Inc.	signifyd.com	Fraud Prevention	239
-	wpadmngr.com	Advertising	1,418	Alibaba Group	aeis.alicdn.com/AWSC/WebUMID/1.93.0/um.js *	Marketing Analytics	201
Signifyd Inc.	signifyd.com	Fraud Prevention	1,414	Amazon Tech.	ssl-images-amazon.com	Marketing Advertising	171
Bounce Exchange	bounceexchange.com	Ad Motivated Tracking Advertising	1,330	Bounce Exchange	bounceexchange.com	Ad Motivated Tracking Advertising	159
InsurAds	insurads.com	Analytics	1,229	Sift Science, Inc.	sift.com	Fraud Prevention	148
Alibaba Group	aeis.alicdn.com/AWSC/WebUMID/1.93.0/um.js *	Marketing Analytics	959	FingerprintJS	cdnjs.cloudflare.com/ajax/libs/fingerprintjs2/2.1.2/fingerprint2.min.js	Fraud Prevention Analytics	144
Rambler Holding	top100.ru	Audience Measurement	913	Amazon Tech.	d38xvr37kwwhcm.cloudfront.net/js/grin-sdk.js	Marketing Advertising	139
Benhauer	salesmanago.pl	Customer Engagement	112	CHEQ AI Tech.	clickcease.com	Fraud Prevention	118
CHEQ AI Tech.	clickcease.com	Fraud Prevention	719	Rambler Holding	top100.ru	Audience Measurement	113
-	franecki.net	Marketing Analytics	589	Benhauer	salesmanago.pl	Customer Engagement	112

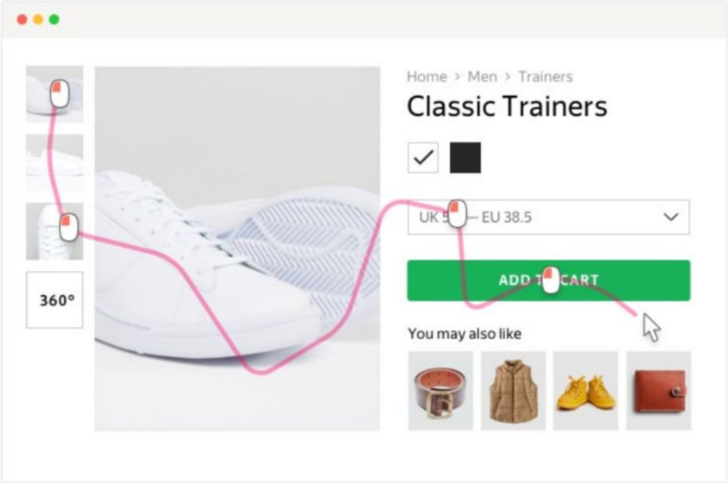
Table 4: The list of primary fingerprinting domains and related entities where at least one fingerprinting attempt was detected during a crawl conducted in August 2023. *Some entities may have multiple associated scripts.

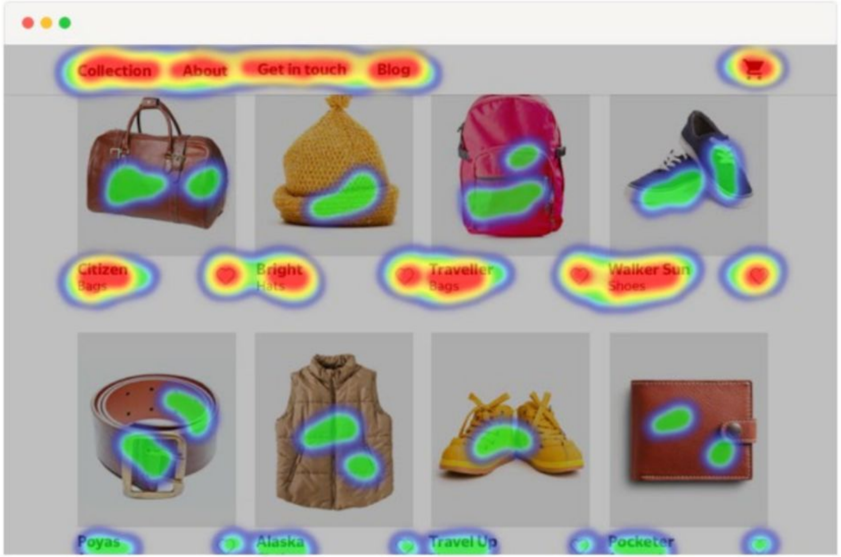
"It is also possible for websites to use fingerprinting for both anti-fraud and advertising [..] simultaneously. For instance, a widely used third-party script on 7% of authentication pages is from [..] sift.com and siftscience.com; these are associated with a single fraud prevention company [17]. However, [..] we noticed that the users' fingerprints were sent to hexagon-analytics.com, which is controlled by the analytics company Hexagon Data [15]" [1]

[1] Senol, A., Ukani, A., Cutler, D., & Bilogrevic, I. (2024). The double edged sword: identifying authentication pages and their fingerprinting behavior. In *ACM Web Conference 2024*.

User activity tracking

- JS event handler to monitor user mouse movements, clicking, and more
- “Watch recordings of your visitors’ sessions. Discover how they browse as if you're looking over their shoulder!” (clicktale.com)

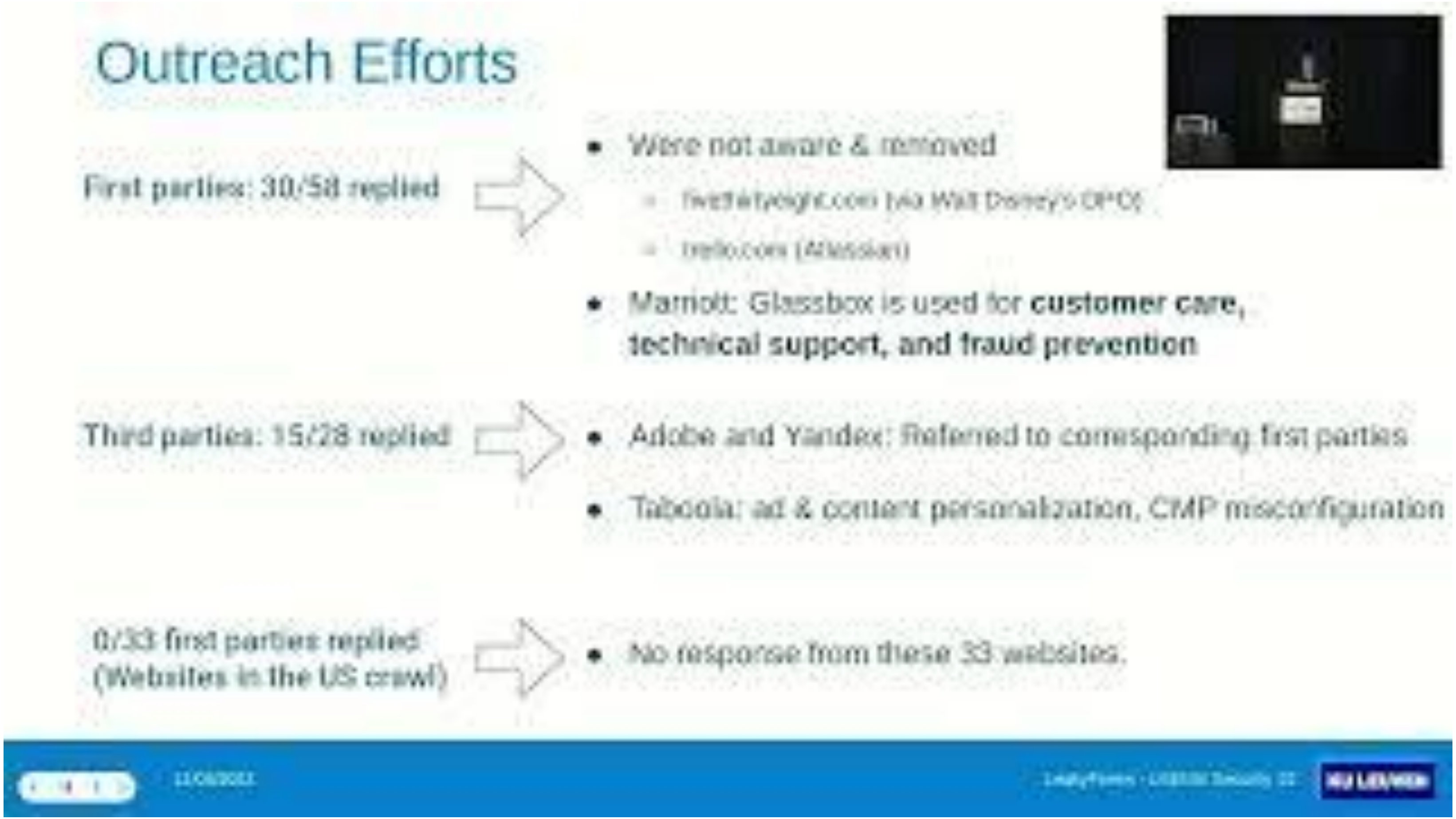




Keystroke and form exfiltration

- JS event handler to get user keystrokes

[1] Senol, A., Acar, G., Humbert, M., & Borgesius, F. Z. (2022). Leaky forms: A study of email and password exfiltration before form submission. In *31st USENIX Security Symposium*.



Outreach Efforts

First parties: 30/58 replied

- Were not aware & removed
  - fivethirtyeight.com (via Walt Disney's DPO)
  - trello.com (Atlassian)
- Marriott: Glassbox is used for **customer care, technical support, and fraud prevention**

Third parties: 15/28 replied

- Adobe and Yandex: Referred to corresponding first parties
- Taboola: ad & content personalization, CMP misconfiguration

0/33 first parties replied (Websites in the US crawl)

- No response from these 33 websites.

[illegible]

LeakyForms - USENIX Security 22

KU LEUVEN

Vault JS

Swiss National
Science Foundation

[1] Senol, A., Acar, G., Humbert, M., & Borgesius, F. Z. (2022). Leaky forms: A study of email and password exfiltration before form submission. In *31st USENIX Security Symposium*.

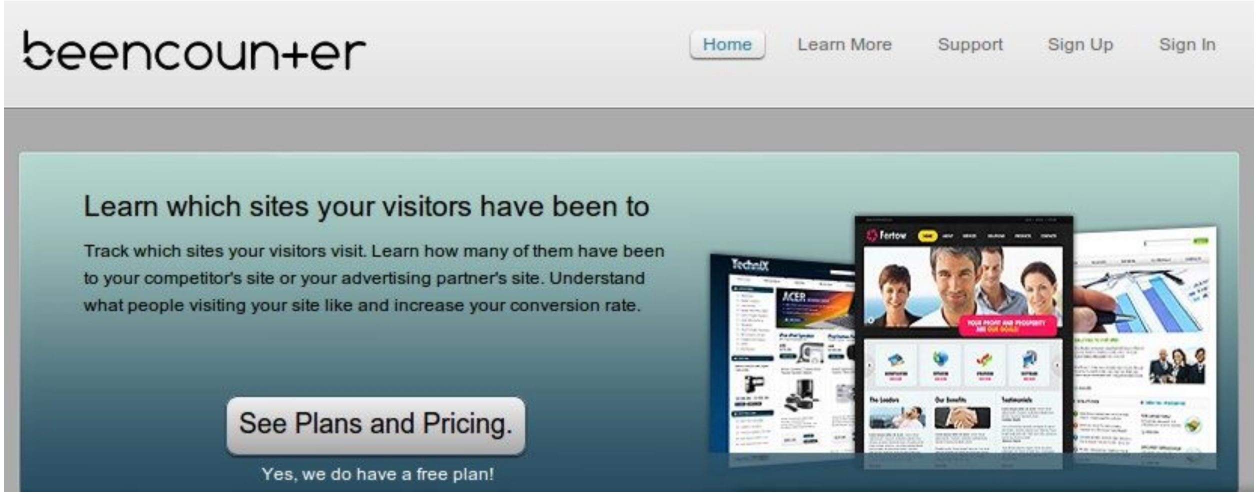
Keystroke and form exfiltration

- JS event handler to get user keystrokes
- Extracting whole form inputs
- Email address or phone number is unique identifier spanning across devices
- Password managers autofill upon request
  - But nevertheless, when you want to login, third-parties can exfiltrate your email
    - Protection: email relay services (Apple Private Relay, Firefox Relay, DuckDuckGo Email Protection, etc.)
  - The same with SSOs, they also leak your birthdate and similar information via "scopes" [2]

[1] Senol, A., Acar, G., Humbert, M., & Borgesius, F. Z. (2022). Leaky forms: A study of email and password exfiltration before form submission. In *31st USENIX Security Symposium*.
[2] Dimova, Y., Van Goethem, T., & Joosen, W. (2023). Everybody's Looking for SSOmething: A large-scale evaluation on the privacy of OAuth authentication on the web. *PETS Proceedings*.

History sniffing

- Checking link color
- CSS `:visited` property
- Timing load (effect of content and DNS cache)



Learn which sites your visitors have been to

Track which sites your visitors visit. Learn how many of them have been to your competitor's site or your advertising partner's site. Understand what people visiting your site like and increase your conversion rate.

Yes, we do have a free plan!

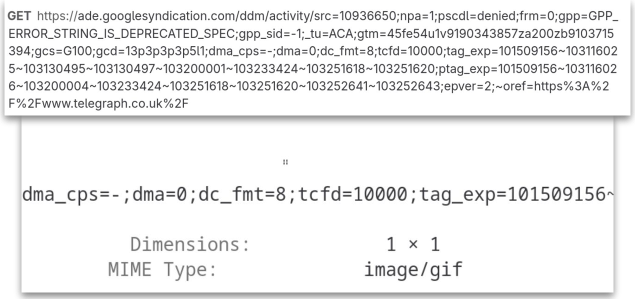
Combined techniques + AdTech

Based on these we can build various composed techniques:

- Tracking pixels (cookie + requests)
    - embedding a pixel is just an easy way to get a third-party load
- Evercookies (persistent cookies using fingerprinting)

Tracking pixels

- 1x1 invisible (transparent or hidden) images
- Sends **URL query parameters**, referrer
  Can set/read 3P cookies, read user agent
- Can be also set from JS (full tracking SDK) = 1P:
  can read DOM, track user actions, access APIs



```
GET https://ade.googlesyndication.com/ddm/activity/src=10936650;npa=1;pscdl=denied;frm=0;gpp=GPP_ERROR_STRING_IS_DEPRECATED_SPEC;gpp_sid=-1;_tu=ACA;gtm=45fe54u1v9190343857za200zb9103715394;gcs=G100;gcd=13p3p3p3p5l1;dma_cps=-;dma=0;dc_fmt=8;tcfd=10000;tag_exp=101509156~103116025~103130495~103130497~103200001~103233424~103251618~103251620;ptag_exp=101509156~103116026~103200004~103233424~103251618~103251620~103252641~103252643;epver=2;~oref=https%3A%2F%2Fwww.telegraph.co.uk%2F
```

```
dma_cps=-;dma=0;dc_fmt=8;tcfd=10000;tag_exp=101509156~

 Dimensions:        1 × 1
MIME Type:          image/gif
```

```
<script>
  var img = new Image();
  img.src = "https://tracker.com/track?uid=xyz&event=pageview";
  document.body.appendChild(img);
</script>
```

- Can be set in an `<iframe>` = 3P

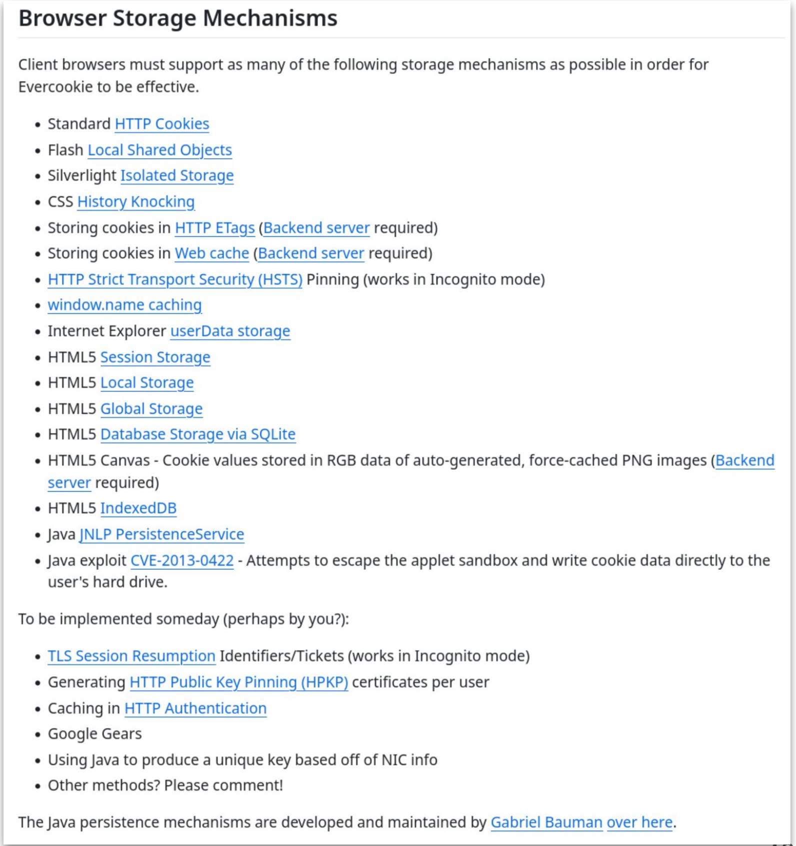
Vault JS Swiss National Science Foundation

Evercookies

JavaScript library (SDK) combining tracking mechanisms to respawn cookies

https://github.com/samyk/evercookie

"We detected respawning by Flash cookies on 10 of the 200 most popular sites and found 33 different Flash cookies were used to respawn over 175 HTTP cookies on 107 of the top 10,000 sites. We also uncovered a new Evercookie vector, IndexedDB that had not been reported before" [1]



Browser Storage Mechanisms

Client browsers must support as many of the following storage mechanisms as possible in order for Evercookie to be effective.

- Standard HTTP Cookies
- Flash Local Shared Objects
- Silverlight Isolated Storage
- CSS History Knocking
- Storing cookies in HTTP ETags (Backend server required)
- Storing cookies in Web cache (Backend server required)
- HTTP Strict Transport Security (HSTS) Pinning (works in Incognito mode)
- window.name caching
- Internet Explorer userData storage
- HTML5 Session Storage
- HTML5 Local Storage
- HTML5 Global Storage
- HTML5 Database Storage via SQLite
- HTML5 Canvas - Cookie values stored in RGB data of auto-generated, force-cached PNG images (Backend server required)
- HTML5 IndexedDB
- Java JNLP PersistenceService
- Java exploit CVE-2013-0422 - Attempts to escape the applet sandbox and write cookie data directly to the user's hard drive.

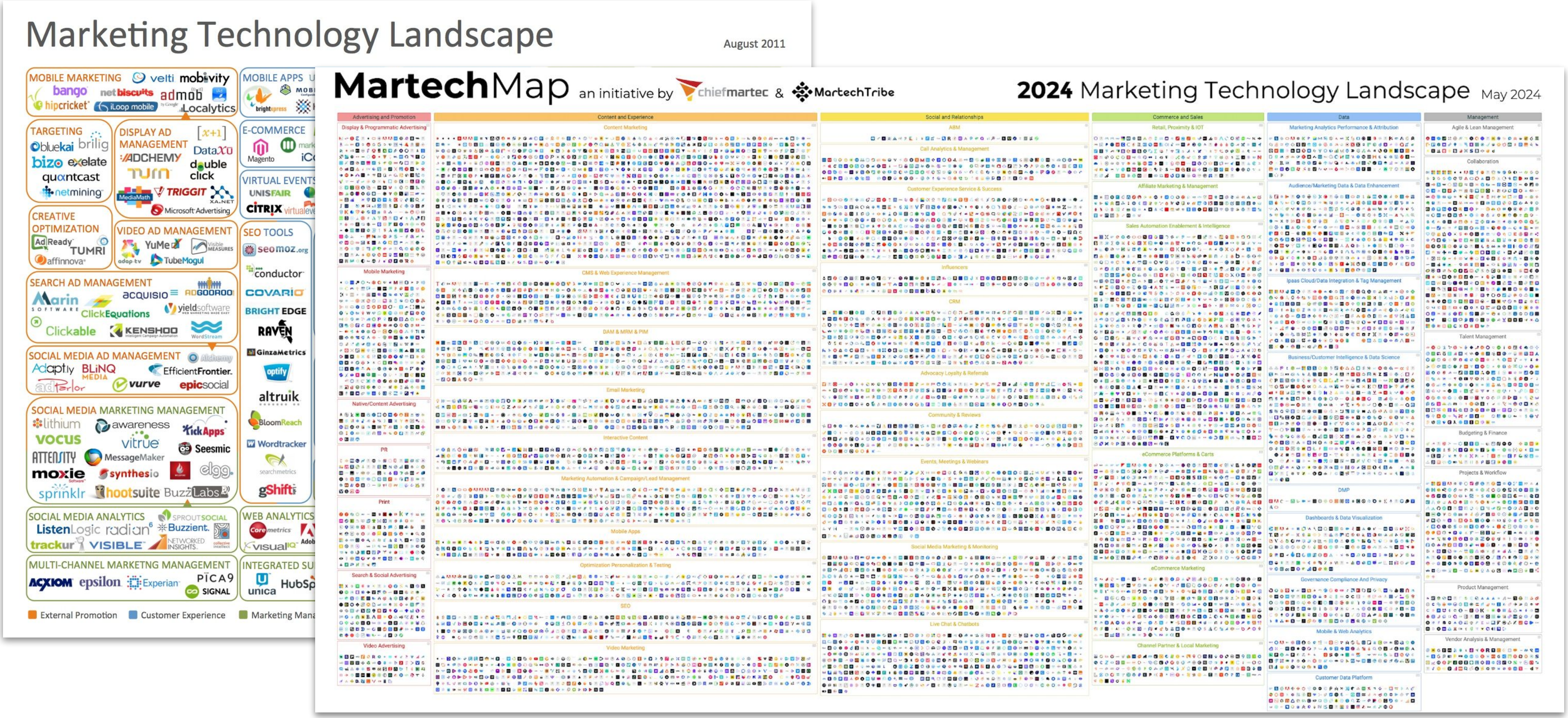
To be implemented someday (perhaps by you?):

- TLS Session Resumption Identifiers/Tickets (works in Incognito mode)
- Generating HTTP Public Key Pinning (HPKP) certificates per user
- Caching in HTTP Authentication
- Google Gears
- Using Java to produce a unique key based off of NIC info
- Other methods? Please comment!

The Java persistence mechanisms are developed and maintained by Gabriel Bauman over here.

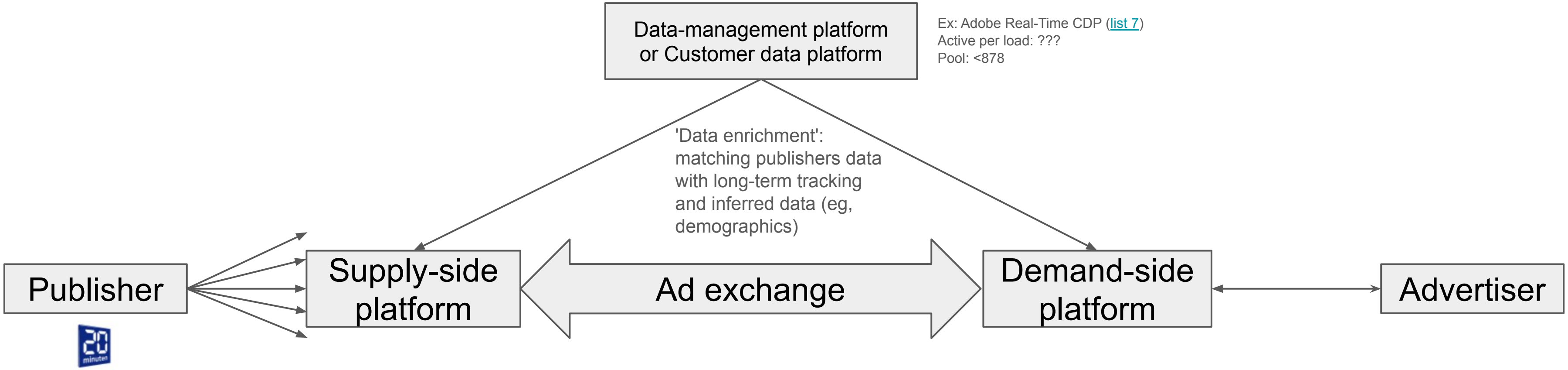
[1] Acar, Gunes, et al. "The web never forgets: Persistent tracking mechanisms in the wild." *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 2014.

Marketing industry



https://chiefmartec.com

Advertising: Real-time bidding


Data-management platform or Customer data platform

Ex: Adobe Real-Time CDP (list 7)
Active per load: ???
Pool: <878

'Data enrichment': matching publishers data with long-term tracking and inferred data (eg, demographics)

Publisher

Supply-side platform

Ad exchange

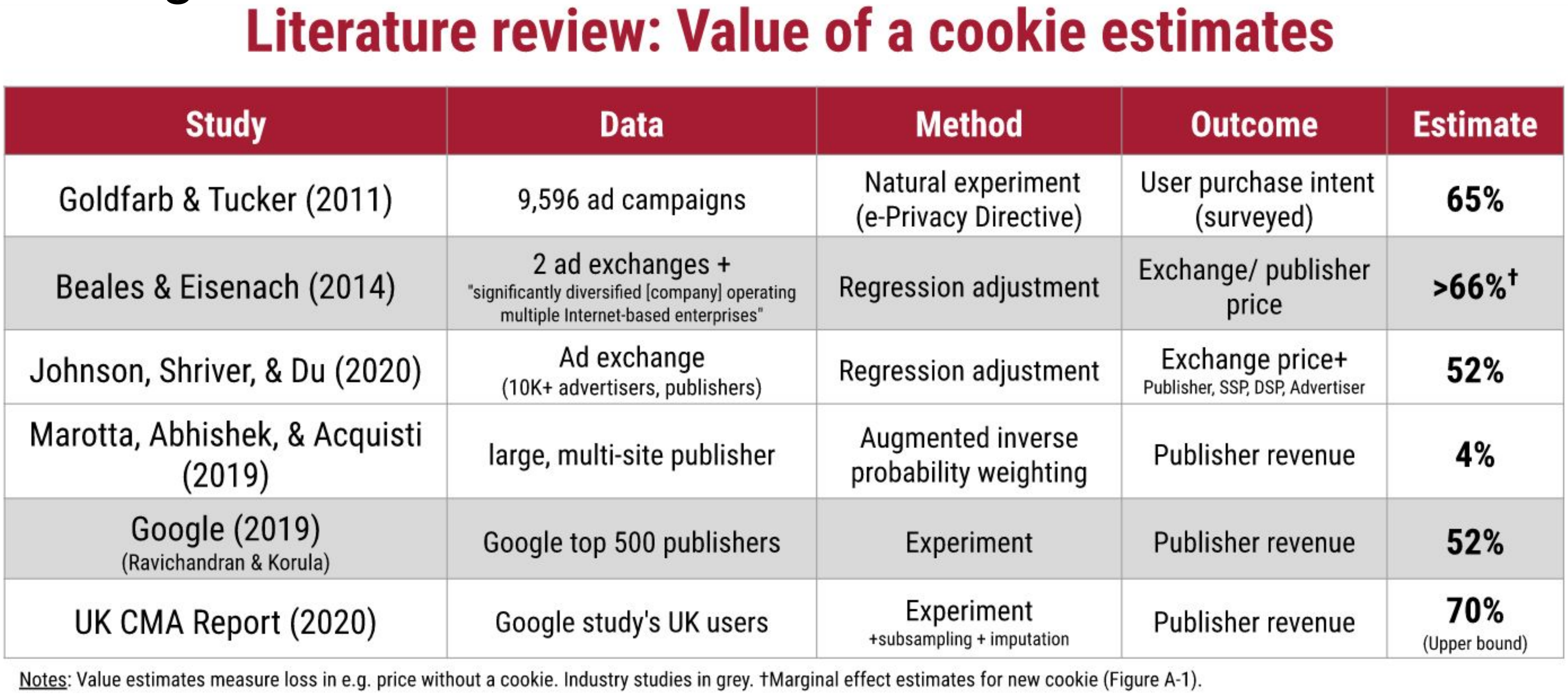
Demand-side platform

Advertiser

Example comp:	Criteo (top 28 list)	2 protocols: OpenRTB by IAB, RTB by Google	Quantcast (list of top 10)	Indiv. brands or ad agencies
Active per load:	3 (e.g.,text, image, and video ads)	6 (e.g., 3 text, 1 image, 1 video ads)	100s	1000s
Pool size:	<878	-	<878	10^{6-7}

Ou, W., Chen, B., Dai, X., Zhang, W., Liu, W., Tang, R., & Yu, Y. (2023). A survey on bid optimization in real-time bidding display advertising. *ACM Transactions on Knowledge Discovery from Data, 18(3)*, 1-31.

Are targetted ads worth?



Literature review: Value of a cookie estimates

Study	Data	Method	Outcome	Estimate
Goldfarb & Tucker (2011)	9,596 ad campaigns	Natural experiment (e-Privacy Directive)	User purchase intent (surveyed)	**65%**
Beales & Eisenach (2014)	2 ad exchanges + "significantly diversified [company] operating multiple Internet-based enterprises"	Regression adjustment	Exchange/ publisher price	**>66%†**
Johnson, Shriver, & Du (2020)	Ad exchange (10K+ advertisers, publishers)	Regression adjustment	Exchange price+ Publisher, SSP, DSP, Advertiser	**52%**
Marotta, Abhishek, & Acquisti (2019)	large, multi-site publisher	Augmented inverse probability weighting	Publisher revenue	**4%**
Google (2019) (Ravichandran & Korula)	Google top 500 publishers	Experiment	Publisher revenue	**52%**
UK CMA Report (2020)	Google study's UK users	Experiment +subsampling + imputation	Publisher revenue	**70%** (Upper bound)

Notes: Value estimates measure loss in e.g. price without a cookie. Industry studies in grey. †Marginal effect estimates for new cookie (Figure A-1).

Johnson, Shriver, & Du (2020): 52% price drop for opt-out users

Countermeasures

Outline slide: Countermeasures

- Browser
  - Chrome Privacy Sandbox
  - Firefox Enhanced Tracking Protection, Total Cookie Protection
  - Safari Intelligent Tracking Prevention
  - Brave
- Extensions
  - Ad blockers
  - Privacy extensions (Ghostery, Privacy Badger, CookieBlock)
- Network level blocking
  - Pi-hole, VPNs

Website breakage - tradeoffs

How are technologies useful for tracking necessary for authentication, etc.?

Voluntary defense

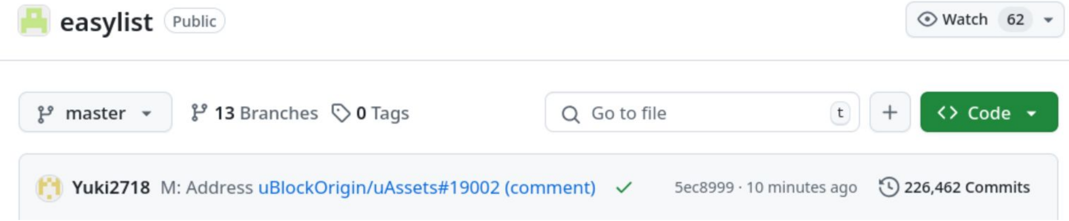
- Platform for website privacy preferences (P3P) [1]
    - User privacy preferences communicated in request headers to servers
    - Proposed in 2002, standardized by W3C, Google and Facebook bypassed it, 2016 discontinued
- Do Not Track (DNT)
    - Binary header field "I do not wish to be tracked"
    - 2012-2019/2025
- Global Privacy Control
    - "Do not sell my data"
    - Mandated by California's CCPA and CPRA and more US states

[1] Cranor, L. F. (2004). P3P: Making privacy policies more useful. IEEE Security & Privacy, 1(6), 50-55.

Browser extensions

Ad blockers and privacy extensions

- Crowd-sourced block lists
  - Advertisement (EasyList)
  - Privacy (EasyPrivacy, AdGuard)
  - Annoyances (EasyCookie)
  - Security filters (malware, intrusion)

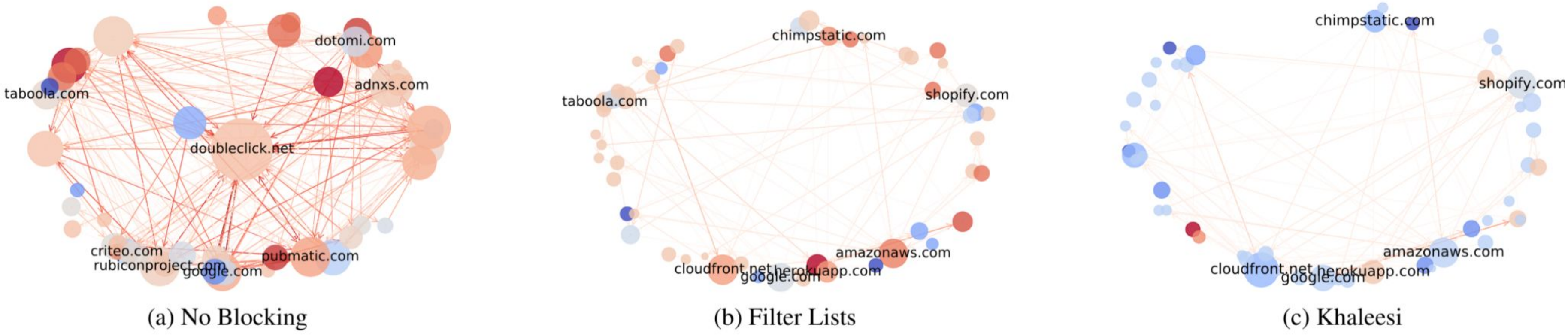


Yuki2718 M: Address uBlockOrigin/uAssets#19002 (comment) ✓ 5ec8999 · 10 minutes ago 226,462 Commits

- ML-based methods:
  - ML suits well the task and often achieves comparable privacy-utility trade offs
  - Issues with adversarial ML methods

Logos by uBlock Origin, Ghostery, Privacy Badger, AdGuard, and AdblockPlus

Browser extensions



(a) No Blocking

(b) Filter Lists

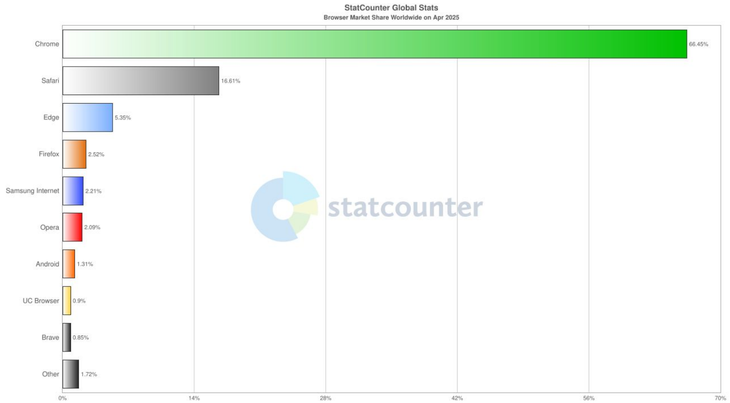
(c) Khaleesi

Figure 4: Request chain graph of redirects between top-50 most popular domains.

Iqbal, U., Wolfe, C., Nguyen, C., Englehardt, S., & Shafiq, Z. (2022). Khaleesi: Breaker of advertising and tracking request chains. In *31st USENIX Security Symposium*.

Browser built-in defenses

- Browsers are primary privacy defense points
  - Tor Browser
  - Brave: Brave Shields
  - DuckDuckGo Browser: App Tracking Protection
  - Firefox: Enhanced Tracking Protection, Total Cookie Protection
  - Safari: Intelligent Tracking Prevention
  - Edge: Tracking Prevention
  - What about Chrome?



Google Ad Manager

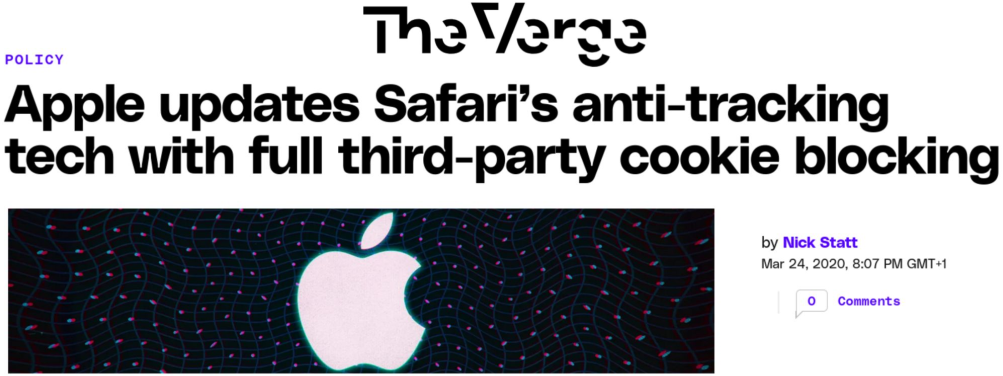
Google Ads

Google Marketing Platform

DoubleClick

By Statcounter - https://gs.statcounter.com/browser-market-share#monthly-202504-202504-bar, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=163203804
Logos by Google

Defense: Third-party cookies discontinuation



The Verge

POLICY

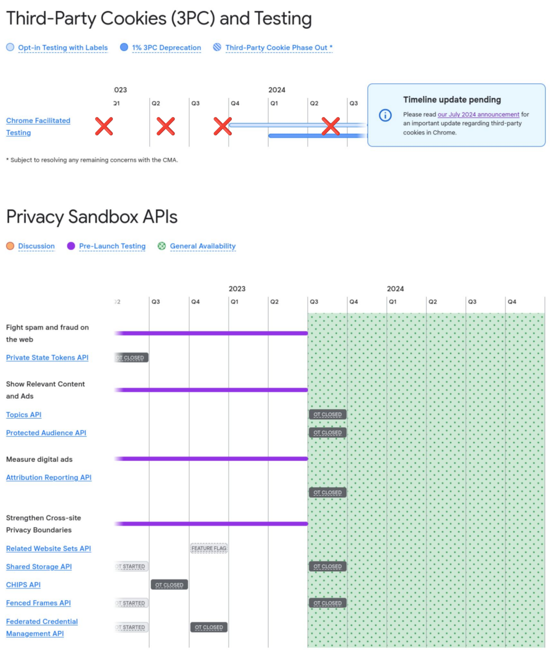
Apple updates Safari's anti-tracking tech with full third-party cookie blocking

by Nick Statt
Mar 24, 2020, 8:07 PM GMT+1

0 Comments

Firefox rolls out Total Cookie Protection by default to more users worldwide

JUNE 14, 2022 MOZILLA



Third-Party Cookies (3PC) and Testing

○ Opt-in Testing with Labels ● 1% 3PC Deprecation ▨ Third-Party Cookie Phase Out *

Timeline update pending

Please read our July 2024 announcement for an important update regarding third-party cookies in Chrome.

* Subject to resolving any remaining concerns with the CMA.

Privacy Sandbox APIs

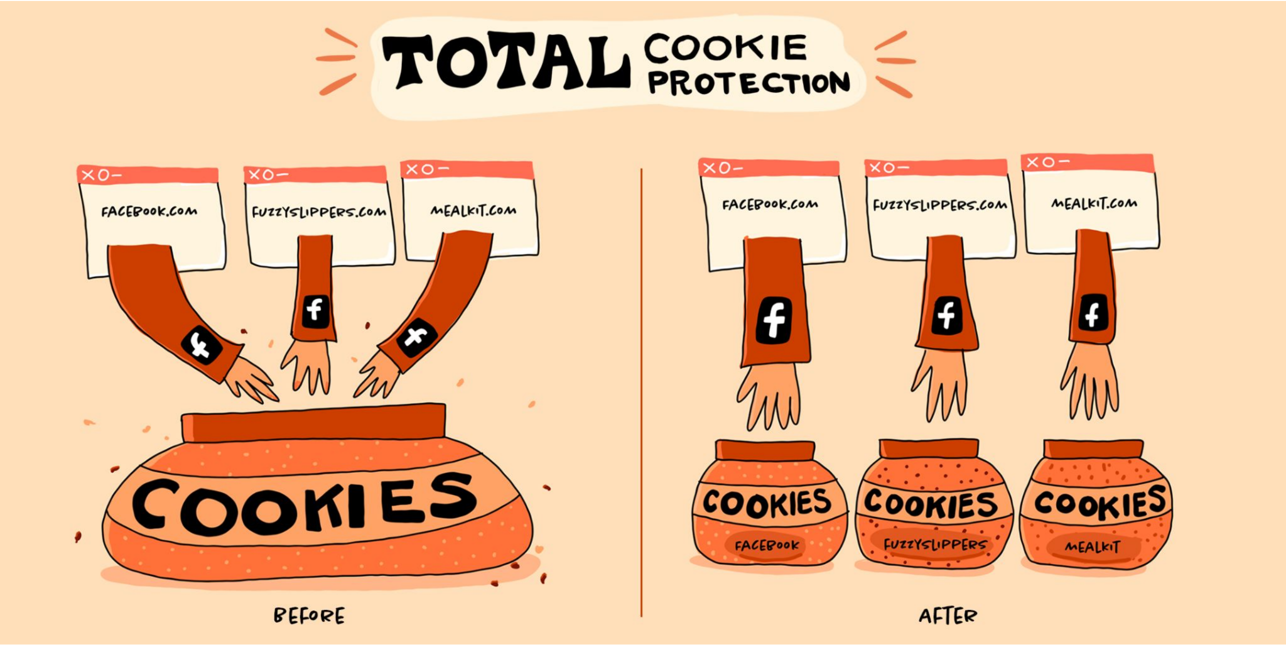
● Discussion ● Pre-Launch Testing ▨ General Availability

See latest browser measures against tracking at: https://www.cookiestatus.com/

Defense: Third-party cookies discontinuation

Third parties loose the simplest method for collecting our histories

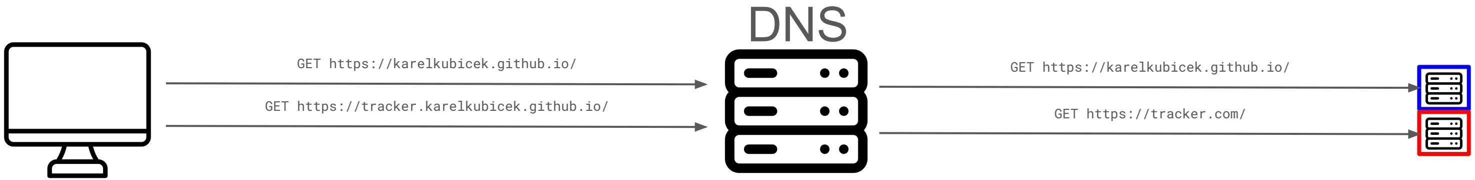
Single-Sign Ons might potentially break (they can still work using redirect params)



https://blog.mozilla.org/security/2021/02/23/total-cookie-protection/

Counter defense: CNAME cloaking

- Allows third parties setting first-party cookie
  - Not as invasive as third-party cookies (lack of connection between visits of different first parties)


DNS
GET https://karelkubicek.github.io/
GET https://tracker.karelkubicek.github.io/
GET https://karelkubicek.github.io/
GET https://tracker.com/

"We perform a historical analysis to study the ecosystem, and find that this form of first-party tracking is becoming increasingly popular and is often used to complement third-party tracking." [1]

"The cloaked subdomains have CNAME records pointing to domains belonging to 32 organizations, which are largely focused on analysis for advertising or marketing purposes" [3]

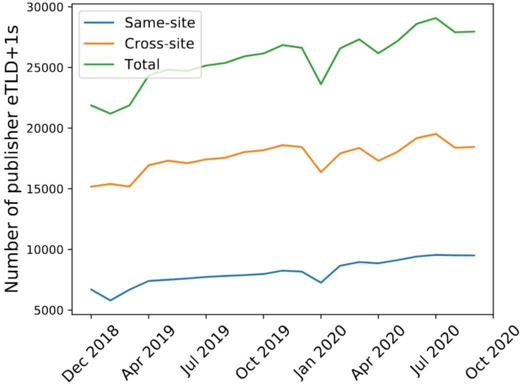
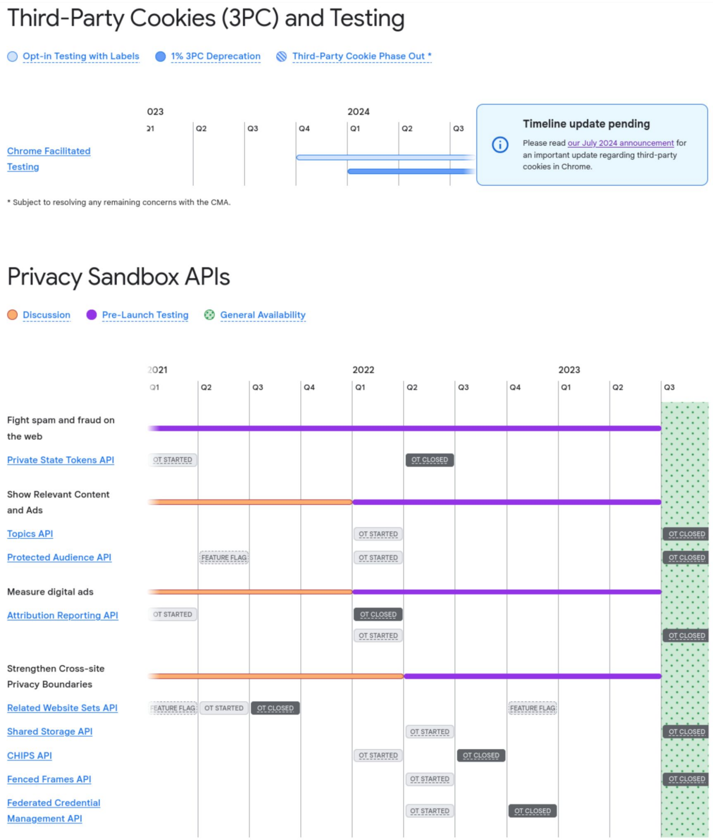


Fig. 5. Number of eTLD+1 domains that include CNAME-based tracking in a same-site and cross-site context.

[1] Dimova, Y., Acar, G., Olejnik, L., Joosen, W., & Van Goethem, T. (2021). The CNAME of the Game: Large-scale Analysis of DNS-based Tracking Evasion. *Proceedings on PETS*.
[2] Dao, H., & Fukuda, K. (2020). Characterizing CNAME Cloaking-based Tracking on the Web. In *TMA*.
[3] https://unit42.paloaltonetworks.com/cname-cloaking/

Chrome Privacy Sandbox

- FLoC
- Topics API
- FLEDGE
- Attribution Reporting API
- Privacy Budget



Third-Party Cookies (3PC) and Testing

○ Opt-in Testing with Labels ● 1% 3PC Deprecation ▨ Third-Party Cookie Phase Out *

Timeline update pending

Please read our July 2024 announcement for an important update regarding third-party cookies in Chrome.

* Subject to resolving any remaining concerns with the CMA.

Privacy Sandbox APIs

● Discussion ● Pre-Launch Testing ● General Availability

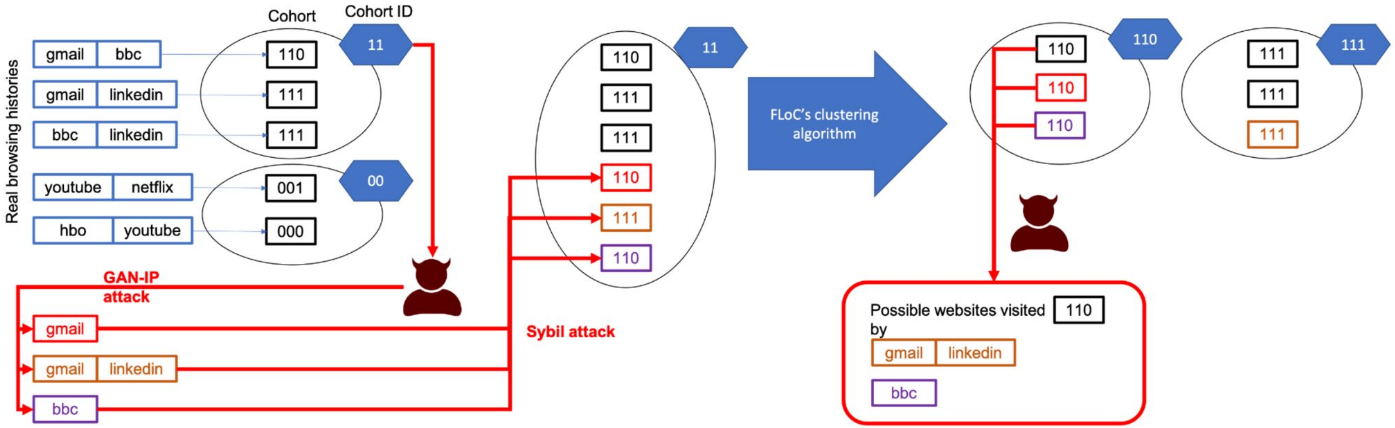
https://privacysandbox.com/

Vault JS

Swiss National Science Foundation

Chrome FLoC (~~Federated~~ Learning of Cohorts)

- First Google's attempt for advertising in post-3P-cookies world
- "Private targeted advertisement"
- Cluster users by browsing history fingerprints (using Locality sensitive hashing)
  - Fingerprints computed locally, clusters globally by "trusted party" (=Google)
  - Clusters of *k* users (to provide *k*-anonymity)
  - Cluster ID shared to advertiser instead of user ID

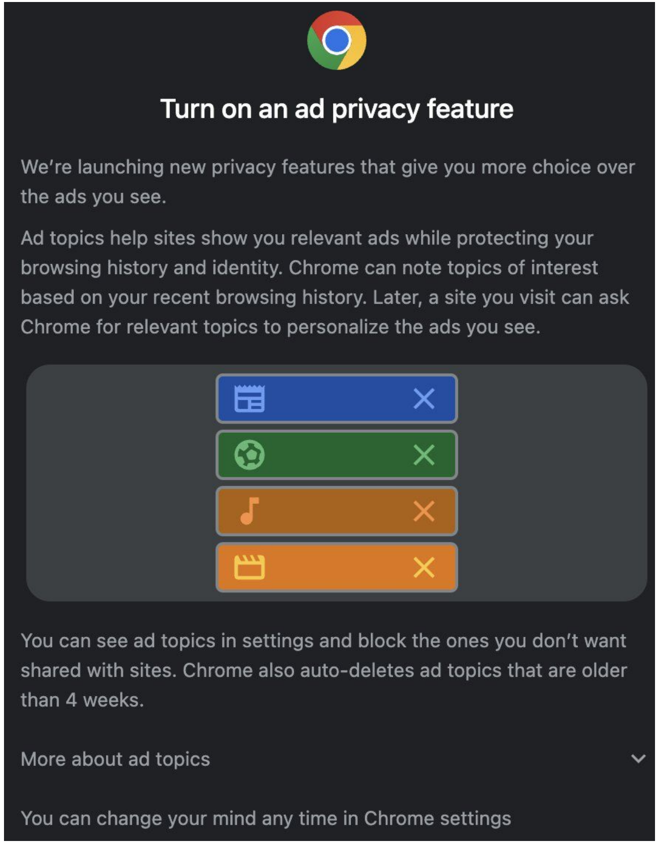


- FLoC discontinued in May 2022

Turati, F., Kubicek, K., Cotrini, C., & Basin, D. Locality-Sensitive Hashing Does Not Guarantee Privacy! Attacks on Google's FLoC and the MinHash Hierarchy System. *PETS 2024*.
Berke, A., & Calacci, D. (2022, November). Privacy limitations of interest-based advertising on the web: A post-mortem empirical analysis of Google's FLoC. In *ACM CCS*.

Chrome Topics API

- Google's second attempt for post-3P-cookies "private targeted advertising"
- Input: browsing history
- Local computation: histogram of topics
- Advertiser gets:
    - Randomly selected of the top 5 topics
    - Or random topic in 5% of cases ("DP")
- Privacy-wise wins over FLoC
    - Still increase browser FP surface



Turn on an ad privacy feature

We're launching new privacy features that give you more choice over the ads you see.

Ad topics help sites show you relevant ads while protecting your browsing history and identity. Chrome can note topics of interest based on your recent browsing history. Later, a site you visit can ask Chrome for relevant topics to personalize the ads you see.

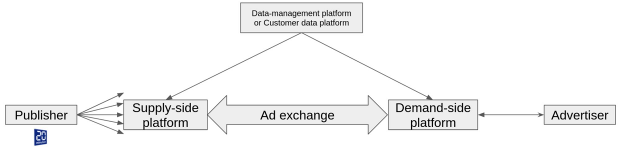
You can see ad topics in settings and block the ones you don't want shared with sites. Chrome also auto-deletes ad topics that are older than 4 weeks.

You can change your mind any time in Chrome settings

https://arstechnica.com/gadgets/2023/09/googles-widely-opposed-ad-platform-the-privacy-sandbox-launches-in-chrome/
https://privacysandbox.google.com/private-advertising/topics?hl=en

Chrome FLEDGE

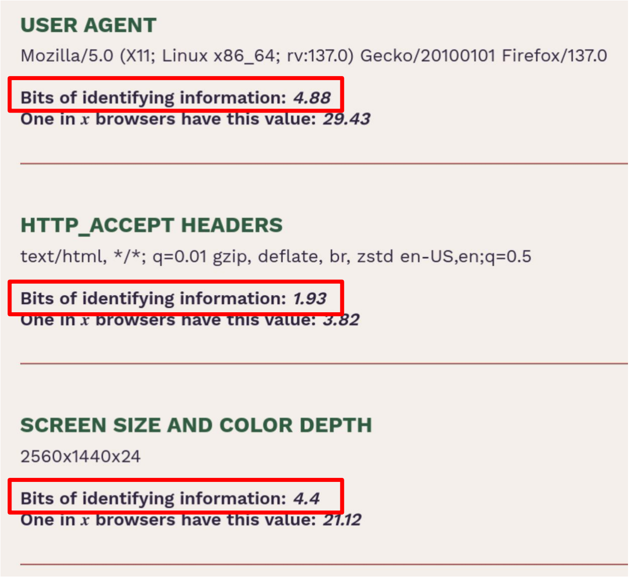
- Alternative to real-time bidding
- Auction runs on client side, using Topics API input
- Check "Interest groups" storage in Chrome
- Can advertisers and publishers trust it?
  - Publisher want to collect attribution, may limit inappropriate ads
  - Advertisers do not want to pay for ads that are not viewed/clicked
  - Uses trusted execution environments (TEEs)
- Attribution Reporting API
  - Results of client-side advertisement FLEDGE+Topics API
  - Attributions are key to measure advertising performance: clicks and views
  - Attributions were typically measured using third-party cookies



https://privacysandbox.google.com/blog/fledge-service-overview#key-value-service and https://privacysandbox.google.com/private-advertising/attribution-reporting?hl=en

Chrome Privacy Budget

- Idea to limit browser fingerprinting
- Measure information leakage of JS calls
  - E.g., screen resolution request→4.4 bits
- Website has *budget* of X bits (X=20)
  - After budget is empty, calls are blocked or get random response
- Likely won't be used in practice 😔



USER AGENT

Mozilla/5.0 (X11; Linux x86_64; rv:137.0) Gecko/20100101 Firefox/137.0

Bits of identifying information: ***4.88***
One in x browsers have this value: ***29.43***

HTTP_ACCEPT HEADERS

text/html, */*; q=0.01 gzip, deflate, br, zstd en-US,en;q=0.5

Bits of identifying information: ***1.93***
One in x browsers have this value: ***3.82***

SCREEN SIZE AND COLOR DEPTH

2560x1440x24

Bits of identifying information: ***4.4***
One in x browsers have this value: ***21.12***

Brave Shields

- Ad and tracker blocking: EasyList, EasyPrivacy, and Disconnect.me lists
- Fingerprinting protection modes:
    - Standard: using list of known FP scripts
    - Aggressive: blocks or spoofs high-entropy attributes (e.g., canvas, audio, WebGL)
- Storage partitioning

Firefox Enhanced Tracking Protection

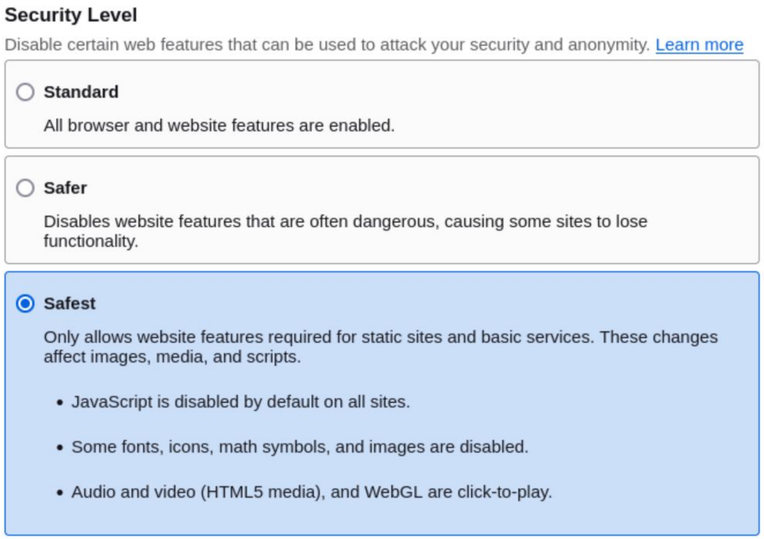
- Tracker blocking: Disconnect.me list
- Fingerprinting protection using list of known FP scripts
- Storage partitioning (Total Cookie Protection) + shorter expiry
- SmartBlock: substitutes tracking scripts to stop tracking without breakage
- Requests: limits CNAME cloaking and DNS prefetching

Safari Intelligent Tracking Protection

- Tracker blocking: on-device ML to recognize cross-site trackers
- Randomizes fingerprintable values
- Block 3P cookies, partition storage, 7-days or 24-hours expiry
- SmartBlock: substitutes tracking scripts to stop tracking without breakage
- Requests: limits CNAME cloaking and request bouncing

Tor Browser

- IP address hidden by Tor network
- All Tor Browsers look same→limits FP
- Limited JavaScript and browser storages
  - According to settings



Security Level

Disable certain web features that can be used to attack your security and anonymity. Learn more

○ **Standard**

All browser and website features are enabled.

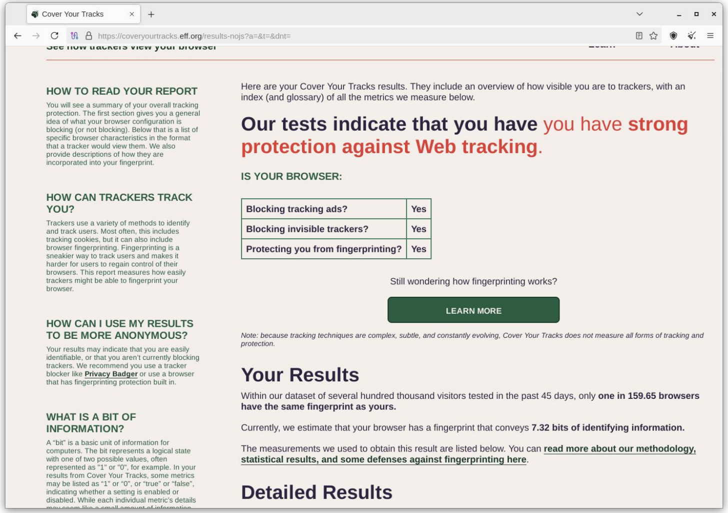
○ **Safer**

Disables website features that are often dangerous, causing some sites to lose functionality.

● **Safest**

Only allows website features required for static sites and basic services. These changes affect images, media, and scripts.

- JavaScript is disabled by default on all sites.
- Some fonts, icons, math symbols, and images are disabled.
- Audio and video (HTML5 media), and WebGL are click-to-play.


Cover Your Tracks
https://coveryourtracks.eff.org/results-nojs?a=&t=&dnt=
See how trackers view your browser

HOW TO READ YOUR REPORT

You will see a summary of your overall tracking protection. The first section gives you a general idea of what your browser configuration is blocking (or not blocking). Below that is a list of specific browser characteristics in the format that a tracker would view them. We also provide descriptions of how they are incorporated into your fingerprint.

HOW CAN TRACKERS TRACK YOU?

Trackers use a variety of methods to identify and track users. Most often, this includes tracking cookies, but it can also include browser fingerprinting. Fingerprinting is a sneakier way to track users and makes it harder for users to regain control of their browsers. This report measures how easily trackers might be able to fingerprint your browser.

HOW CAN I USE MY RESULTS TO BE MORE ANONYMOUS?

Your results may indicate that you are easily identifiable, or that you aren't currently blocking trackers. We recommend you use a tracker blocker like **Privacy Badger** or use a browser that has fingerprinting protection built in.

WHAT IS A BIT OF INFORMATION?

A "bit" is a basic unit of information for computers. The bit represents a logical state with one of two possible values, often represented as "1" or "0", for example. In your results from Cover Your Tracks, some metrics may be listed as "1" or "0", or "true" or "false", indicating whether a setting is enabled or disabled. While each individual metric's details

Here are your Cover Your Tracks results. They include an overview of how visible you are to trackers, with an index (and glossary) of all the metrics we measure below.

Our tests indicate that you have you have strong protection against Web tracking.

IS YOUR BROWSER:

Blocking tracking ads?	Yes
Blocking invisible trackers?	Yes
Protecting you from fingerprinting?	Yes

Still wondering how fingerprinting works?

LEARN MORE

Note: because tracking techniques are complex, subtle, and constantly evolving, Cover Your Tracks does not measure all forms of tracking and protection.

Your Results

Within our dataset of several hundred thousand visitors tested in the past 45 days, only **one in 159.65 browsers have the same fingerprint as yours.**

Currently, we estimate that your browser has a fingerprint that conveys **7.32 bits of identifying information.**

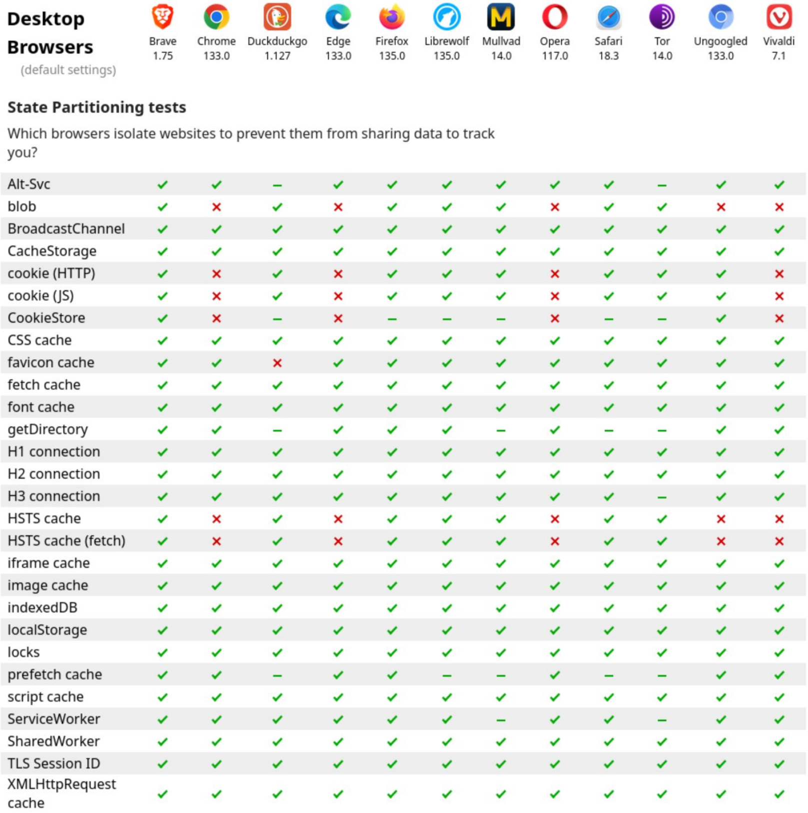
The measurements we used to obtain this result are listed below. You can **read more about our methodology, statistical results, and some defenses against fingerprinting here**.

Detailed Results

Browser comparison

https://privacytests.org/

Note the lack of extension evaluation



Desktop Browsers (default settings)	Brave 1.75	Chrome 133.0	Duckduckgo 1.127	Edge 133.0	Firefox 135.0	Librewolf 135.0	Mullvad 14.0	Opera 117.0	Safari 18.3	Tor 14.0	Ungoogled 133.0	Vivaldi 7.1
State Partitioning tests Which browsers isolate websites to prevent them from sharing data to track you?												
Alt-Svc	✓	✓	–	✓	✓	✓	✓	✓	✓	–	✓	✓
blob	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓	✗	✗
BroadcastChannel	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CacheStorage	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
cookie (HTTP)	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓	✓	✗
cookie (JS)	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓	✓	✗
CookieStore	✓	✗	–	✗	–	–	–	✗	–	–	✓	✗
CSS cache	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
favicon cache	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
fetch cache	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
font cache	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
getDirectory	✓	✓	–	✓	✓	✓	–	✓	–	–	✓	✓
H1 connection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
H2 connection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
H3 connection	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓	✓
HSTS cache	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓	✗	✗
HSTS cache (fetch)	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓	✗	✗
iframe cache	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
image cache	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
indexedDB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
localStorage	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
locks	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
prefetch cache	✓	✓	–	✓	✓	–	–	✓	–	–	✓	✓
script cache	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ServiceWorker	✓	✓	✓	✓	✓	✓	–	✓	✓	–	✓	✓
SharedWorker	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
TLS Session ID	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
XMLHttpRequest cache	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Network-level filtering

Enforce blocklist at DNS level

- Pi-hole
- NextDNS
- Privacy VPN providers (e.g., Mullvad)

Issues:

- Website breakage (coarse-grained blocking)
- Added costs (Pi-hole) or trust in third party (NextDNS)

Advantages:

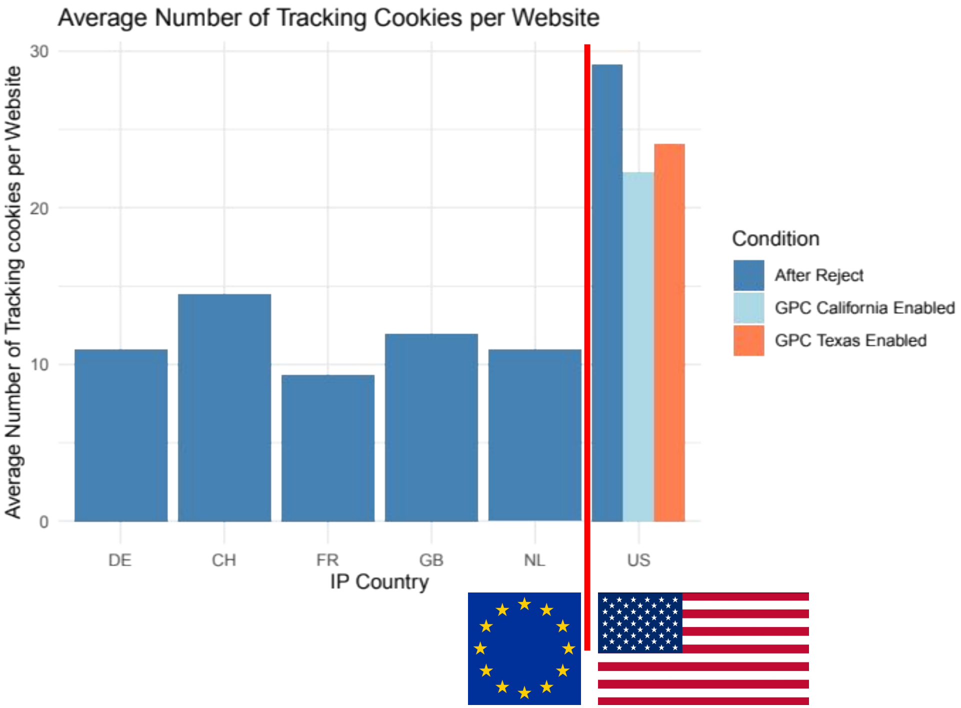
- Works on any device

Regulations

Lecture next week

My opinion:
Tracking is inherent feature of online technologies (think of TCP/IP), regulation has power to bring privacy to masses.

But so far we have good laws (GDPR), but weak enforcement.



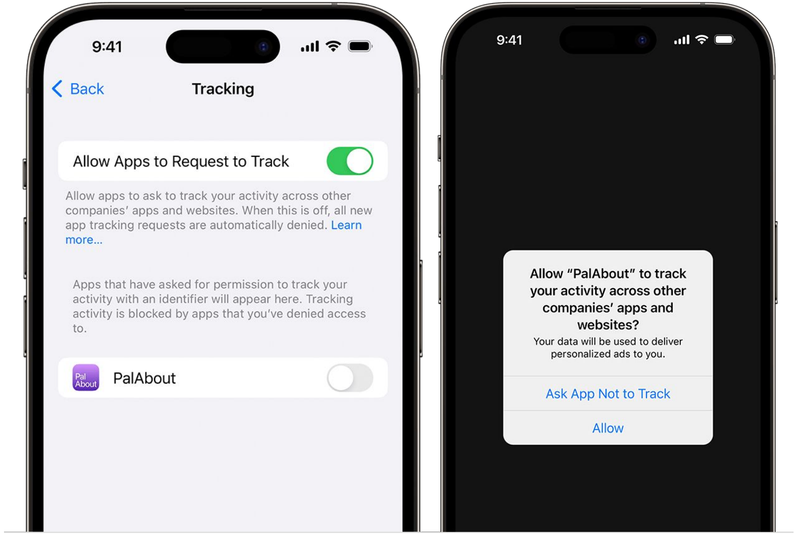
Amit Zac

Ahmed Bouhoula

Outside of web

Phones

- SDKs provide central tracking API
  - iOS option to opt-out
  - Android has no such choice, Defense are custom ROMs:
    - GrapheneOS - blocks all Google services + security hardening
    - CalyxOS - spoof or fake data to Google (less private than GrapheneOS)
    - /e/OS Fairphone and many others (less private than CalyxOS)
- Sensors provide even more risks: https://sensor-js.xyz/



9:41

< Back Tracking

Allow Apps to Request to Track ●

Allow apps to ask to track your activity across other companies' apps and websites. When this is off, all new app tracking requests are automatically denied. Learn more...

Apps that have asked for permission to track your activity with an identifier will appear here. Tracking activity is blocked by apps that you've denied access to.

PalAbout ○

9:41

Allow "PalAbout" to track your activity across other companies' apps and websites?

Your data will be used to deliver personalized ads to you.

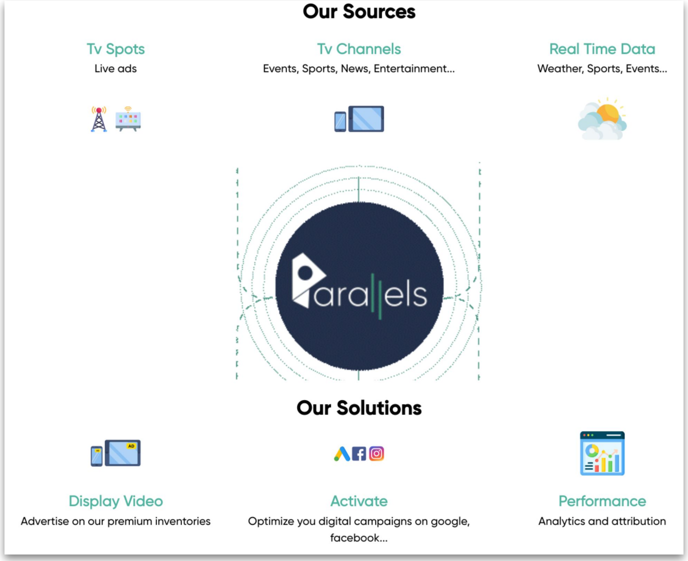
Ask App Not to Track

Allow

https://support.apple.com/en-us/102420

Phone to PC syncing

- Email addresses and phone numbers as universal identifiers across devices
- QR codes (tracking redirects)
    - Potentially can be dynamically generated to connect scanned and scanning devices
- Cross-device tracking



Our Sources

Tv Spots
Live ads

Tv Channels
Events, Sports, News, Entertainment...

Real Time Data
Weather, Sports, Events...

Our Solutions

Display Video
Advertise on our premium inventories

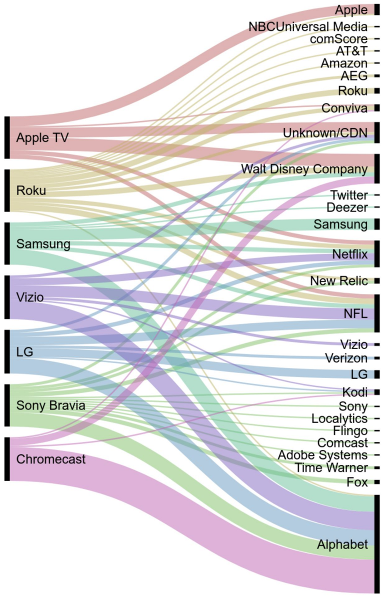
Activate
Optimize you digital campaigns on google, facebook...

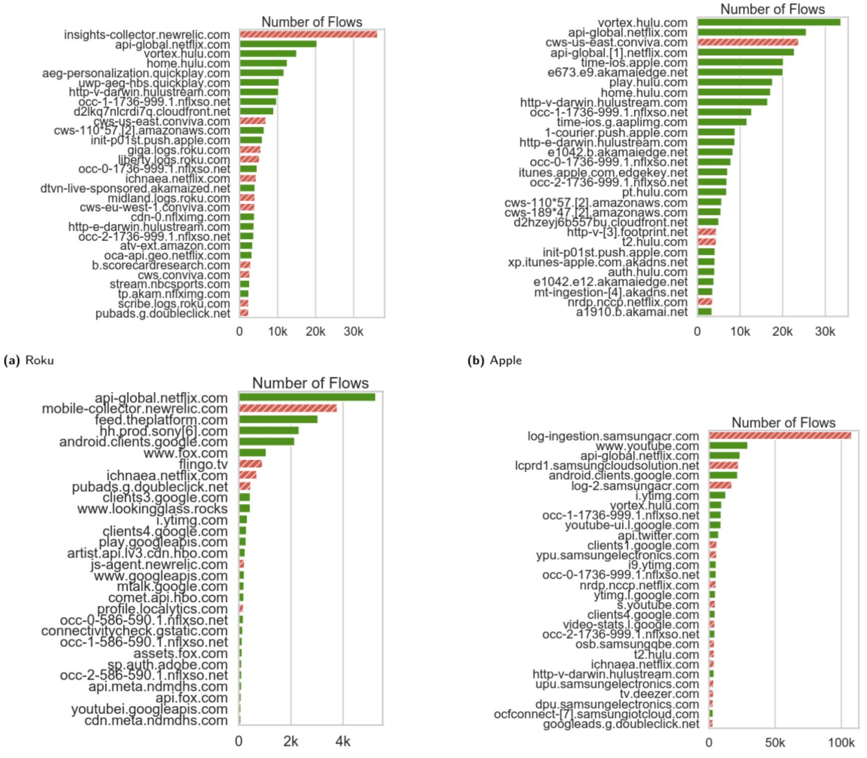
Performance
Analytics and attribution

https://www.silverpush.co/parallels/

Smart TV

- TV prices covered by ads [2]
- Apps live from tracking
- Lack of defense ecosystem (Pi-hole)
- Samsung listening to conversations [3]





(a) Roku

(b) Apple

(c) Sony

(d) Samsung

Fig. 1. Top-30 fully qualified domain names in terms of number of flows per device for a subset of the smart TVs in the "in the wild" dataset. See Appendix C.2 for the other brands. Domains identified as ATS are highlighted with red, dashed bars.

[1] Varmarken, J., Le, H., Shuba, A., Markopoulou, A., & Shafiq, Z. (2020). The TV is smart and full of trackers: Measuring smart TV advertising and tracking. In *PETS 2020*.
[2] https://tech.yahoo.com/general/articles/wondering-why-smart-tv-many-122329643.html
[3] https://www.bbc.com/news/technology-31296188

Internet of Things

- Majority of products come from China
  - Risks of spying on users to industrial espionage
- A Roomba recorded a woman on the toilet. How did screenshots end up on Facebook?
- Somebody’s Watching: Hackers Breach Ring Home Security Cameras
- German parents told to destroy doll that can spy on children
- Millions of Web Camera and Baby Monitor Feeds Are Exposed

Typically even worse protection and awareness than in the case of SmartTVs [1]

Koohang, A., Sargent, C. S., Nord, J. H., & Paliszkiewicz, J. (2022). Internet of Things (IoT): From awareness to continued use. *International Journal of Information Management*.

Closing notes

Conclusion

- Online tracking is the oil of $200B marketing industry
  - They know too much about us and do not want to forget it
  - It is not difficult to buy the data
- Stateful and stateless tracking techniques:
  - (Third-party) Cookies and other various storages
  - Browser fingerprinting, user input sniffing
- PETs exist, but are not enough widespread, either due to:
  - Usability trade-offs (web breakage) or simply because AdTech fights them
- Out of web, the situation is typically worse
  - Interpreted nature of web technologies makes it easier to inspect
  - In mobile, tracking is built in the API
- Technologies are inherently trackable, solution is (in my opinion) regulation

Backup slides

About me (for real)

- PhD from ETH Zurich on web tracking compliance
    - Cookies
    - Emails
    - New tracking technologies
- Privacy job market is difficult:
    - Public sector (PhD ▹ postdoc ▹ professor), regulatory positions (limited in CH)
    - Industry is more interested in tracking than protecting users, exceptions:
      Private search engines: DuckDuckGo, Startpage, Ecosia
      Private browsers: Brave, Mozilla, Safari
      VPN: Proton, Mullvad
      Startups: crawling startups (VaultJS, webXray, etc.), Differential privacy (Tumult Labs)
      Big tech: no power to change their business model

Demo websites

- Mouse and form tracking: https://capturly.com/features/session-replay/demo
- Fingerprinting: https://coveryourtracks.eff.org
- No cookies/IP/fingerprinting tracking: https://potatocrunchcereal.com/cookielesscookies/
-

Scanning tools

Websites:

- https://themarkup.org/blacklight
- https://baycloud.com

Extensions:

- https://disconnect.me/disconnect
- https://addons.mozilla.org/en-US/firefox/addon/lightbeam-chikl/

Browser policies

- Permissions for various sensors
- Access control on execution scopes
  - First vs third party:

    <script src="tracker.com"> is executed as first party

    <iframe src="tracker.com"> is executed as third party
  -

About

Author: Karel Kubicek

karel.kubicek@inria.fr
https://karelkubicek.github.io

Image sources:

The Noun Project: Computer by ratubilqis1986 from Noun Project (CC BY 3.0); Server by Ricons from Noun Project (CC BY 3.0)
Other images cited directly in the slides