Web Content

CS 249i

Catherine Han — February 10, 2025
catherinehan@cs.stanford.edu

Old School Websites

David Wagner

Professor
Carl J. Penther Chair in Engineering
Computer Science Division
University of California, Berkeley

Research interests. Computer security. I am currently working on security for AI (particularly security for large language models), AI for security, and other topics in computer security. I have worked on software security, electronic voting, wireless security, sensor network security, and applied cryptography.

Projects. My group page is here. I'm part of Berkeley's security research group, a member of the ACTION Institute, and co-director of the KACST-UCB Joint Center on Cybersecurity.

I was previously co-PI or PI for SCRUB, the Intel Science and Technology Center for Secure Computing, and the DHOSA project, and I was a member of the TRUST, ACCURATE, Science of Security, and ISAAC projects.

Publications. My technical papers and publications are all available online.

Teaching. See my past teaching.

`https://people.eecs.berkeley.edu/~daw/`

`https://crypto.stanford.edu/~dabo/`

Publications | Courses | Research

Dan Boneh

dabo@cs.stanford.edu

Professor of Computer Science and Electrical Engineering, Stanford University.
Co-director of the Stanford Computer Security Lab

Address

- Mail: Computer Science Dept., 389 Jane Stanford Way, Stanford, CA 94305
- Office: CoDa Building, office W218.
- Directions to the CoDa building.

Telephone

- Office: (650) 725-3897

Research Interests

My main research focus is applied cryptography and computer security.

~~Old School~~ Modern(?) Websites

David Wagner

Professor
Carl J. Penther Chair in Engineering
Computer Science Division
University of California, Berkeley

Research interests. Computer security. I am currently working on security for AI (particularly security for large language models), AI for security, and other topics in computer security. I have worked on software security, electronic voting, wireless security, sensor network security, and applied cryptography.

Projects. My group page is here. I'm part of Berkeley's security research group, a member of the ACTION Institute, and co-director of the KACST-UCB Joint Center on Cybersecurity.

I was previously co-PI or PI for SCRUB, the Intel Science and Technology Center for Secure Computing, and the DHOSA project, and I was a member of the TRUST, ACCURATE, Science of Security, and ISAAC projects.

Publications. My technical papers and publications are all available online.

Teaching. See my past teaching.

`https://people.eecs.berkeley.edu/~daw/`

`https://crypto.stanford.edu/~dabo/`

Publications | Courses | Research

Dan Boneh

dabo@cs.stanford.edu

Professor of Computer Science and Electrical Engineering, Stanford University.
Co-director of the Stanford Computer Security Lab

Address

- Mail: Computer Science Dept., 389 Jane Stanford Way, Stanford, CA 94305
- Office: CoDa Building, office W218.
- Directions to the CoDa building.

Telephone

- Office: (650) 725-3897

Research Interests

My main research focus is applied cryptography and computer security.

Modern Websites

Enterprise Suite
With automated multi-entity accounting and reporting that cuts down manual work.
Learn more
Powered by quickbooks + mailchimp

The New York Times

Tuesday, February 4, 2025

GIVE THE TIMES

Trump Tariffs Take Effect, and China Retaliates With Penalties of Its Own

China announced retaliatory measures, including tariffs and an investigation of Google, after President Trump's 10 percent tariff on Chinese products took effect.

4 MIN READ

The New York Times

ANALYSIS

What Does Trump Really Want From Canada and Mexico?

President Trump's demands on the United States' neighbors are difficult to measure. That allows him to declare victory when he sees fit.

4 MIN READ

Trump Wields U.S. Power With Unclear Economic Consequences

6 MIN READ

After Tariff Fight With Canada and Mexico, Trump's Next Target Is Europe

5 MIN READ

Modern Websites

What happens on page load?

INTUIT
Enterprise Suite
accounting and reporting
Learn more
Powered by quickbooks + mailchimp

The New York Times

Tuesday, February 4, 2025

GIVE THE TIMES

Trump Tariffs Take Effect, and China Retaliates With Penalties of Its Own

China announced retaliatory measures, including tariffs and an investigation of Google, after President Trump's 10 percent tariff on Chinese products took effect.

4 MIN READ

The New York Times

ANALYSIS

What Does Trump Really Want From Canada and Mexico?

President Trump's demands on the United States' neighbors are difficult to measure. That allows him to declare victory when he sees fit.

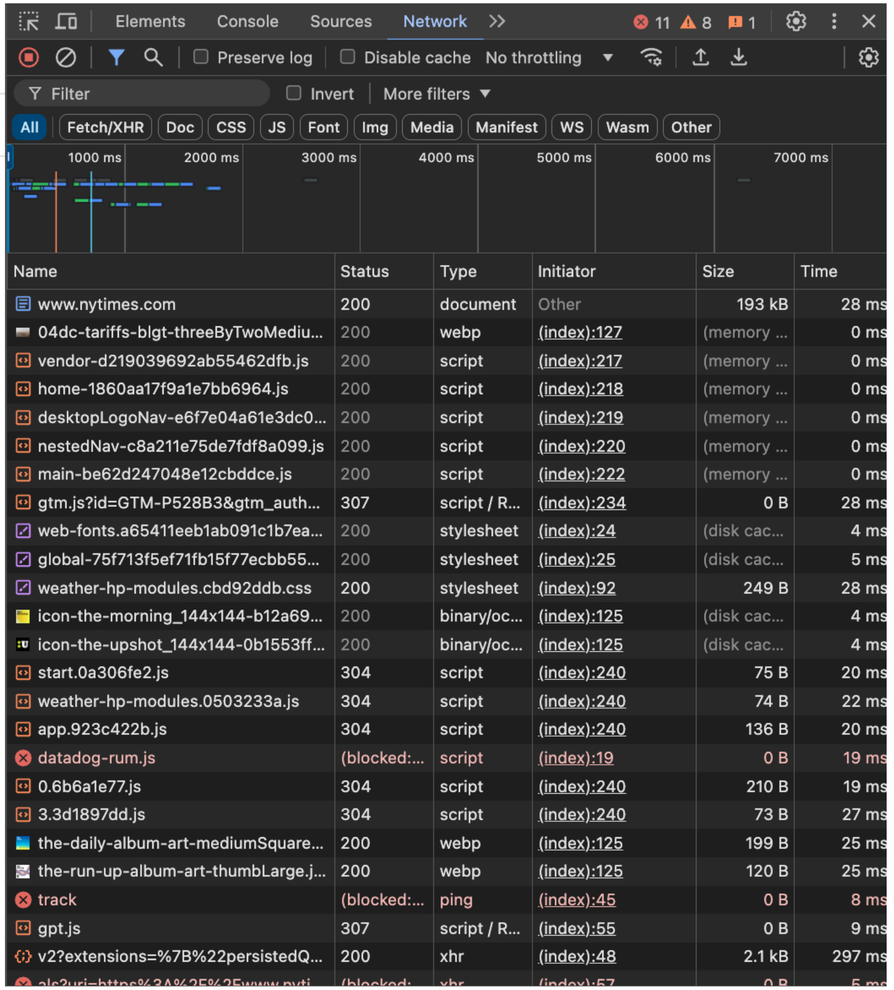
4 MIN READ

Trump Wields U.S. Power With Unclear Economic Consequences

6 MIN READ

After Tariff Fight With Canada and Mexico, Trump's Next Target Is Europe

5 MIN READ



Name	Status	Type	Initiator	Size	Time
www.nytimes.com	200	document	Other	193 kB	28 ms
04dc-tariffs-blgt-threeByTwoMediu...	200	webp	(index):127	(memory ...	0 ms
vendor-d219039692ab55462dfb.js	200	script	(index):217	(memory ...	0 ms
home-1860aa17f9a1e7bb6964.js	200	script	(index):218	(memory ...	0 ms
desktopLogoNav-e6f7e04a61e3dc0...	200	script	(index):219	(memory ...	0 ms
nestedNav-c8a211e75de7fdf8a099.js	200	script	(index):220	(memory ...	0 ms
main-be62d247048e12cbddce.js	200	script	(index):222	(memory ...	0 ms
gtm.js?id=GTM-P528B3&gtm_auth...	307	script / R...	(index):234	0 B	28 ms
web-fonts.a65411eeb1ab091c1b7ea...	200	stylesheet	(index):24	(disk cac...	4 ms
global-75f713f5ef71fb15f77ecbb55...	200	stylesheet	(index):25	(disk cac...	5 ms
weather-hp-modules.cbd92ddb.css	200	stylesheet	(index):92	249 B	28 ms
icon-the-morning_144x144-b12a69...	200	binary/oc...	(index):125	(disk cac...	4 ms
icon-the-upshot_144x144-0b1553ff...	200	binary/oc...	(index):125	(disk cac...	4 ms
start.0a306fe2.js	304	script	(index):240	75 B	20 ms
weather-hp-modules.0503233a.js	304	script	(index):240	74 B	22 ms
app.923c422b.js	304	script	(index):240	136 B	20 ms
datadog-rum.js	(blocked:...	script	(index):19	0 B	19 ms
0.6b6a1e77.js	304	script	(index):240	210 B	19 ms
3.3d1897dd.js	304	script	(index):240	73 B	27 ms
the-daily-album-art-mediumSquare...	200	webp	(index):125	199 B	25 ms
the-run-up-album-art-thumbLarge.j...	200	webp	(index):125	120 B	25 ms
track	(blocked:...	ping	(index):45	0 B	8 ms
gpt.js	307	script / R...	(index):55	0 B	9 ms
v2?extensions=%7B%22persistedQ...	200	xhr	(index):48	2.1 kB	297 ms

Modern Websites

What happens on page load?

INTUIT
Enterprise Suite
accounting and reporting
Learn more
Powered by quickbooks + mailchimp

The New York Times

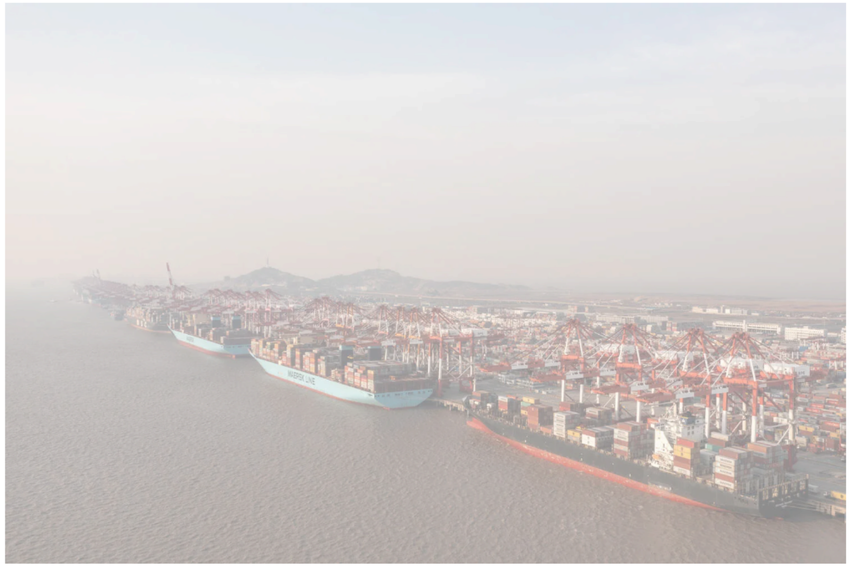
Tuesday, February 4, 2025

GIVE THE TIMES

Trump Tariffs Take Effect, and China Retaliates With Penalties of Its Own

China announced retaliatory measures, including tariffs and an investigation of Google, after President Trump's 10 percent tariff on Chinese products took effect.

4 MIN READ



The New York Times

ANALYSIS

What Does Trump Really Want From Canada and Mexico?

President Trump's demands on the United States' neighbors are difficult to measure. That allows him to declare victory when he sees fit.

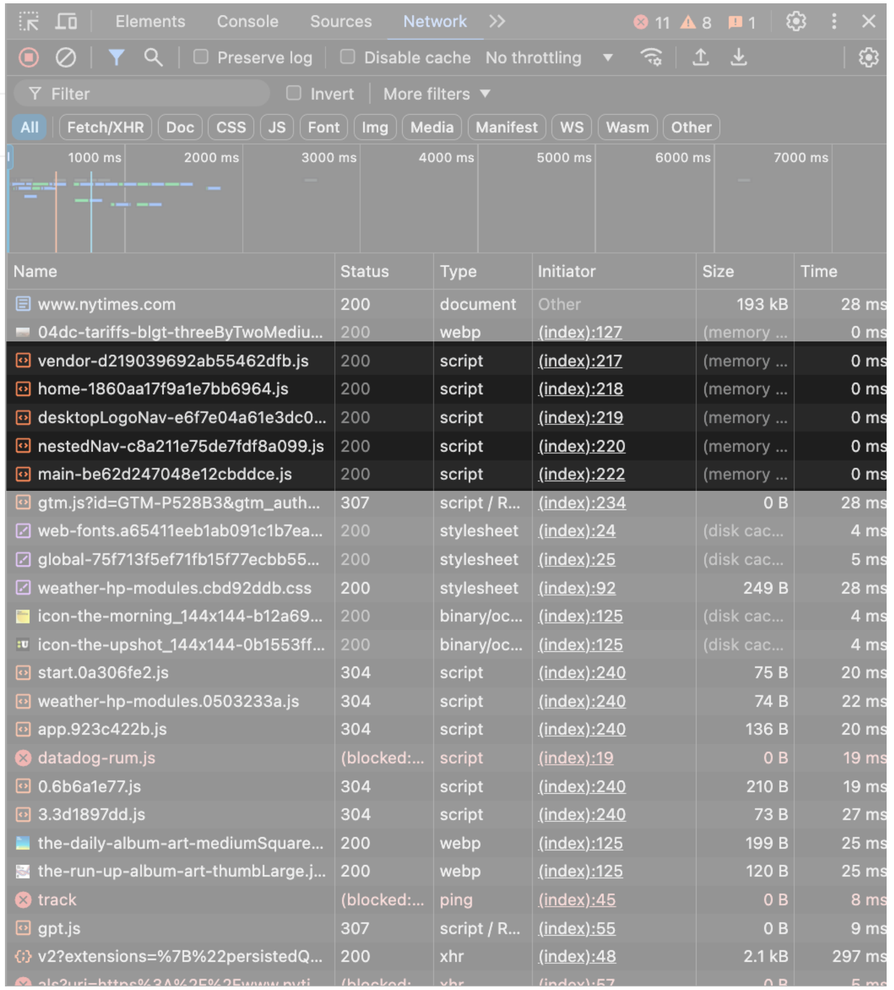
4 MIN READ

Trump Wields U.S. Power With Unclear Economic Consequences

6 MIN READ

After Tariff Fight With Canada and Mexico, Trump's Next Target Is Europe

5 MIN READ



Name	Status	Type	Initiator	Size	Time
www.nytimes.com	200	document	Other	193 kB	28 ms
04dc-tariffs-blgt-threeByTwoMediu...	200	webp	(index):127	(memory ...	0 ms
vendor-d219039692ab55462dfb.js	200	script	(index):217	(memory ...	0 ms
home-1860aa17f9a1e7bb6964.js	200	script	(index):218	(memory ...	0 ms
desktopLogoNav-e6f7e04a61e3dc0...	200	script	(index):219	(memory ...	0 ms
nestedNav-c8a211e75de7fdf8a099.js	200	script	(index):220	(memory ...	0 ms
main-be62d247048e12cbddce.js	200	script	(index):222	(memory ...	0 ms
gtm.js?id=GTM-P528B3&gtm_auth...	307	script / R...	(index):234	0 B	28 ms
web-fonts.a65411eeb1ab091c1b7ea...	200	stylesheet	(index):24	(disk cac...	4 ms
global-75f713f5ef71fb15f77ecbb55...	200	stylesheet	(index):25	(disk cac...	5 ms
weather-hp-modules.cbd92ddb.css	200	stylesheet	(index):92	249 B	28 ms
icon-the-morning_144x144-b12a69...	200	binary/oc...	(index):125	(disk cac...	4 ms
icon-the-upshot_144x144-0b1553ff...	200	binary/oc...	(index):125	(disk cac...	4 ms
start.0a306fe2.js	304	script	(index):240	75 B	20 ms
weather-hp-modules.0503233a.js	304	script	(index):240	74 B	22 ms
app.923c422b.js	304	script	(index):240	136 B	20 ms
datadog-rum.js	(blocked:...	script	(index):19	0 B	19 ms
0.6b6a1e77.js	304	script	(index):240	210 B	19 ms
3.3d1897dd.js	304	script	(index):240	73 B	27 ms
the-daily-album-art-mediumSquare...	200	webp	(index):125	199 B	25 ms
the-run-up-album-art-thumbLarge.j...	200	webp	(index):125	120 B	25 ms
track	(blocked:...	ping	(index):45	0 B	8 ms
gpt.js	307	script / R...	(index):55	0 B	9 ms
v2?extensions=%7B%22persistedQ...	200	xhr	(index):48	2.1 kB	297 ms

Modern Websites

Third-party resources

INTUIT
Enterprise Suite
Learn more
Powered by quickbooks + mailchimp

The New York Times

Tuesday, February 4, 2025

https://securepubads.g.doubleclick.net/tag/js/gpt.js

≠

https://nytimes.com

Trump Tariffs Take Effect, and China Retaliates With Penalties of Its Own

China announced retaliatory measures, including tariffs and an investigation of Google, after President Trump's 10 percent tariff on Chinese products took effect.

4 MIN READ

The New York Times

ANALYSIS

What Does Trump Really Want From Canada and Mexico?

President Trump's demands on the United States' neighbors are difficult to measure. That allows him to declare victory when he sees fit.

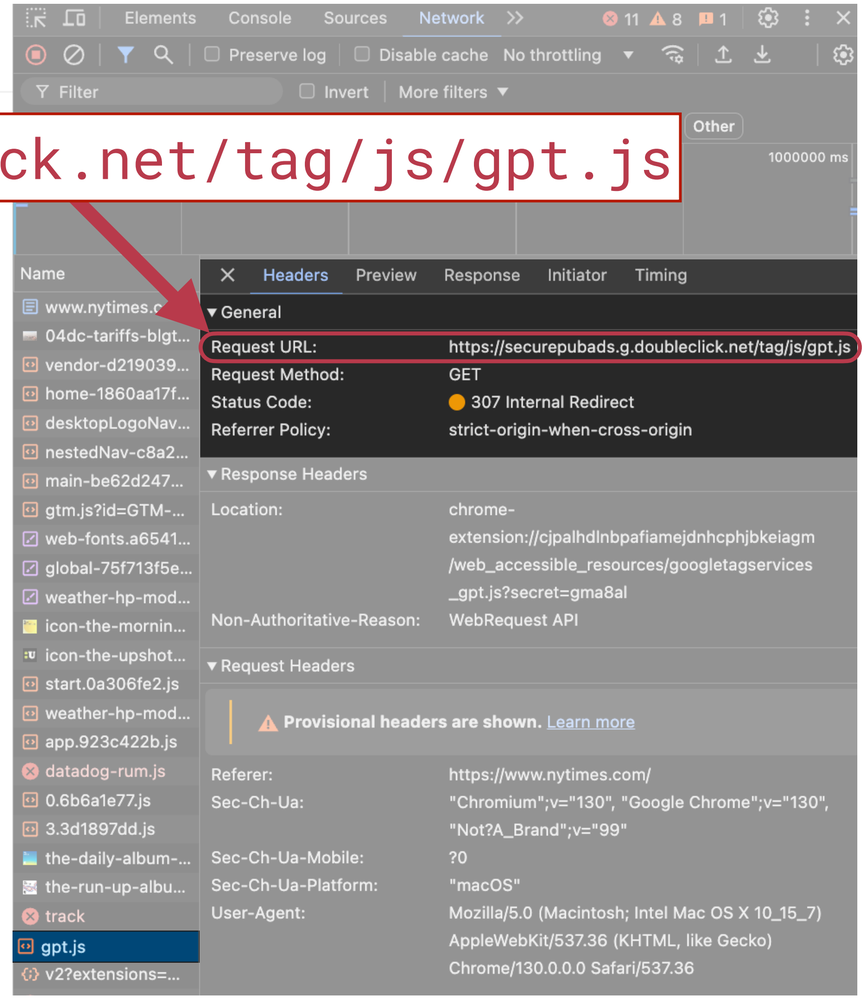
4 MIN READ

Trump Wields U.S. Power With Unclear Economic Consequences

6 MIN READ

After Tariff Fight With Canada and Mexico, Trump's Next Target Is Europe

5 MIN READ



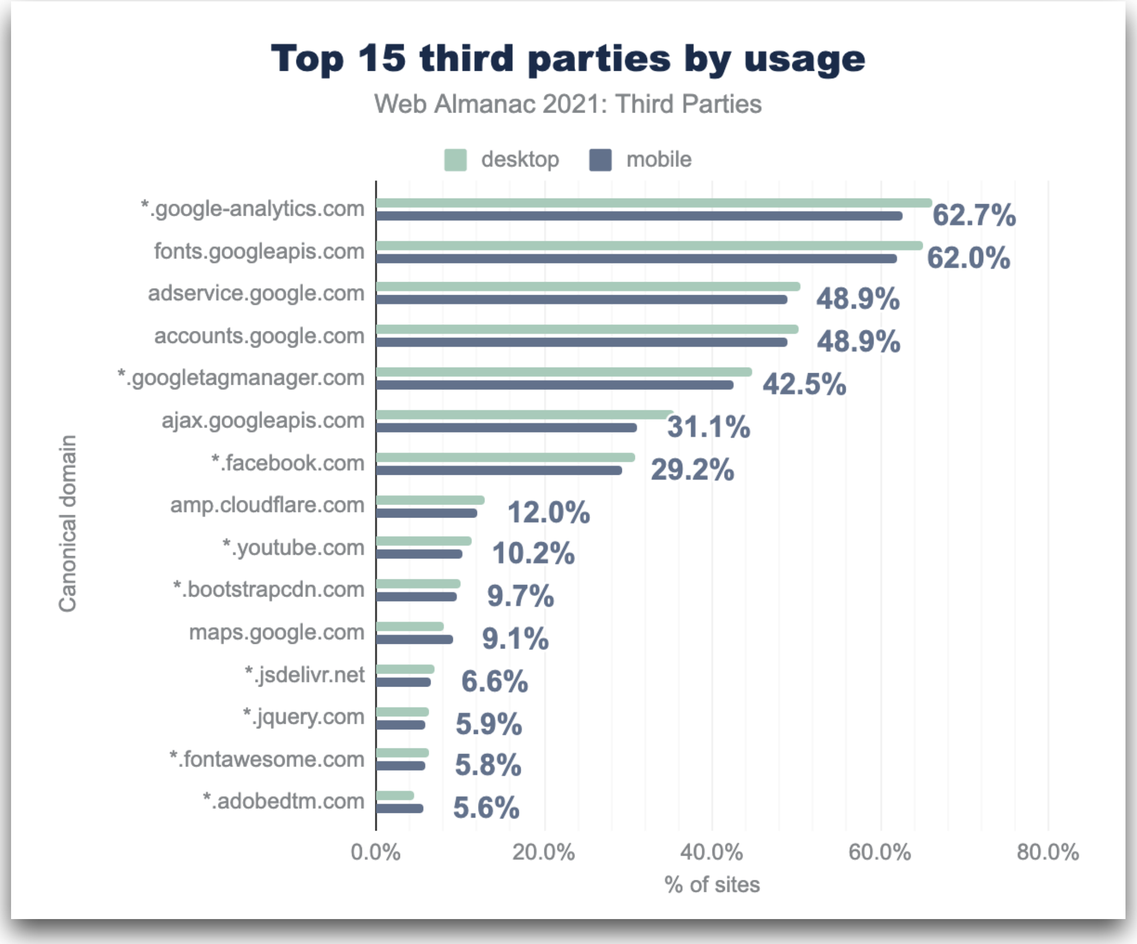
Modern Websites

Third-party resources

Why *third-party* resources?

💡 Diverse content and utilities developed/hosted by other entities

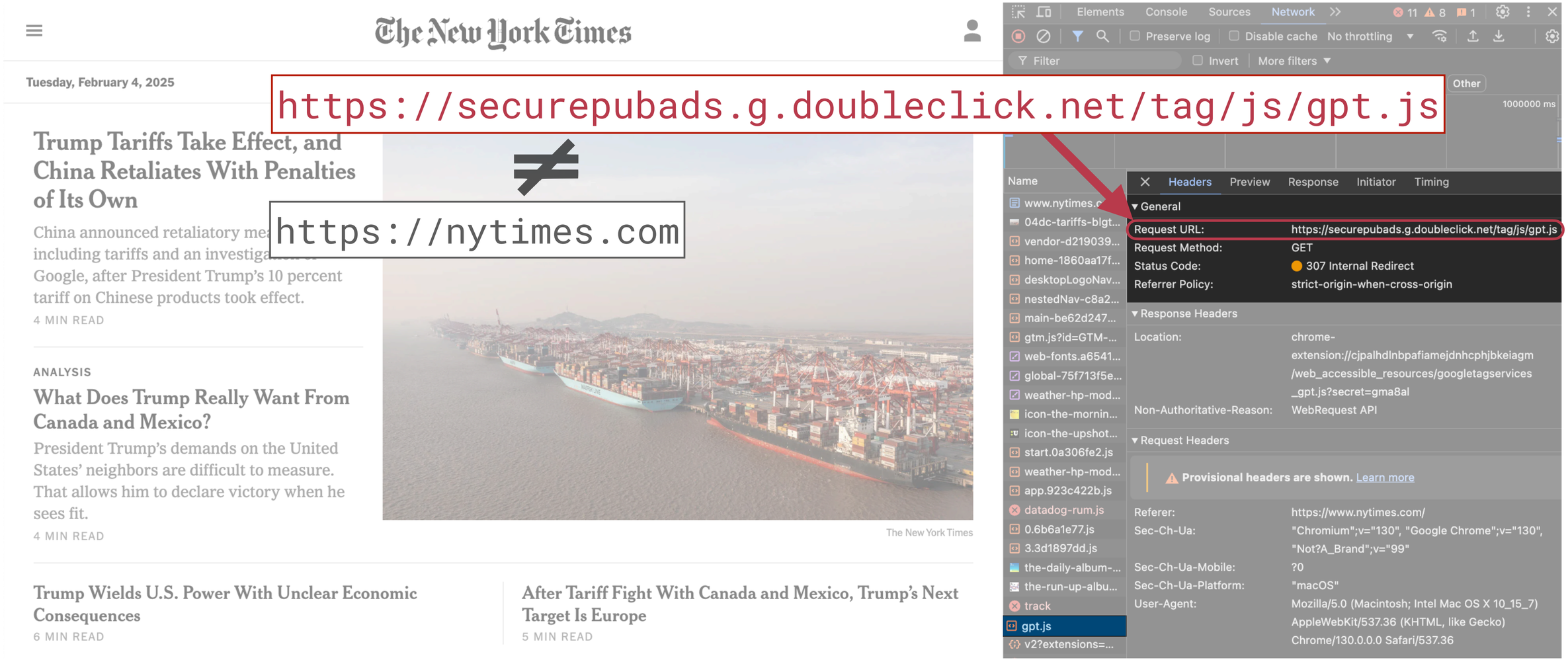
- Libraries including for ads, tracking, and analytics
- Fonts and style sheets
- Media (images, videos, icons)
- Performance (CDNs)



https://almanac.httparchive.org/en/2021/third-parties#fig-12

Modern Websites

Third-party resources



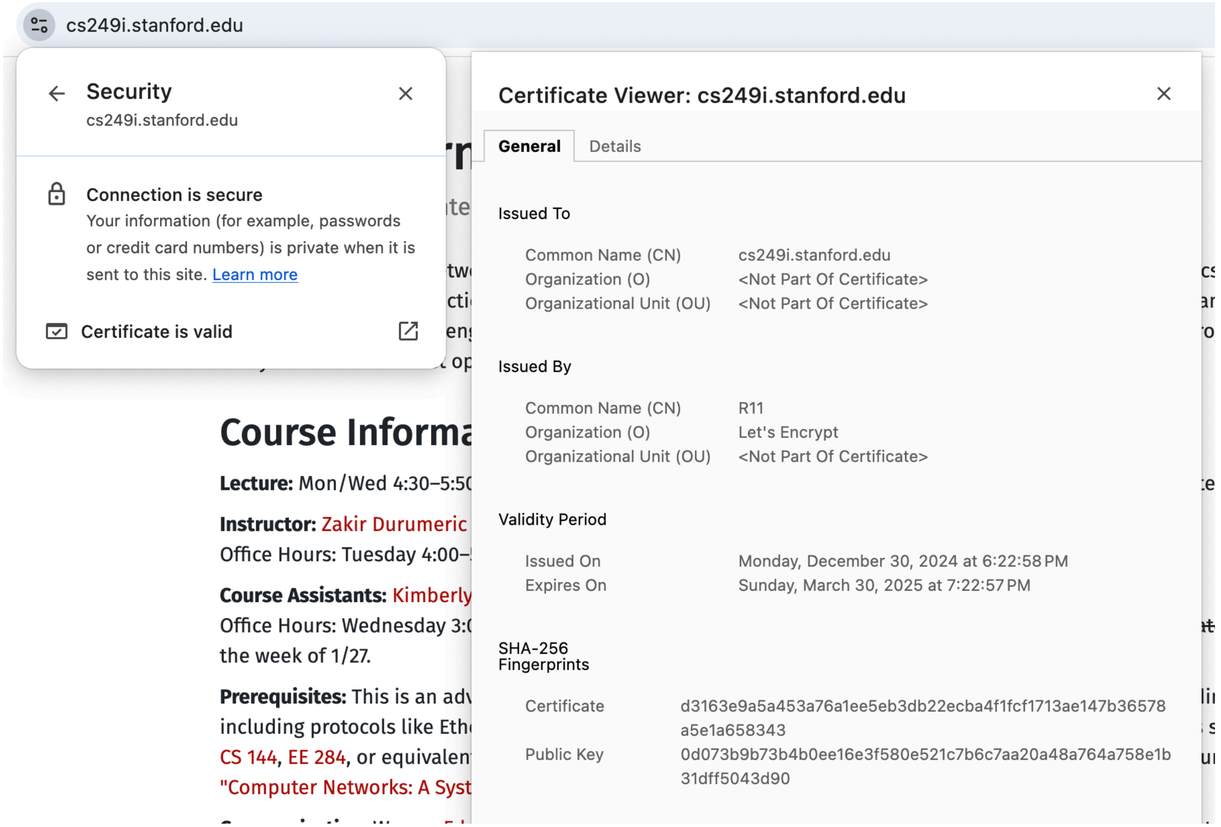
Modern Websites

Third-party resources

- Potential downsides?
  - Performance costs
  - Privacy risks
  - Security risks

“Tangled Web” of dependencies

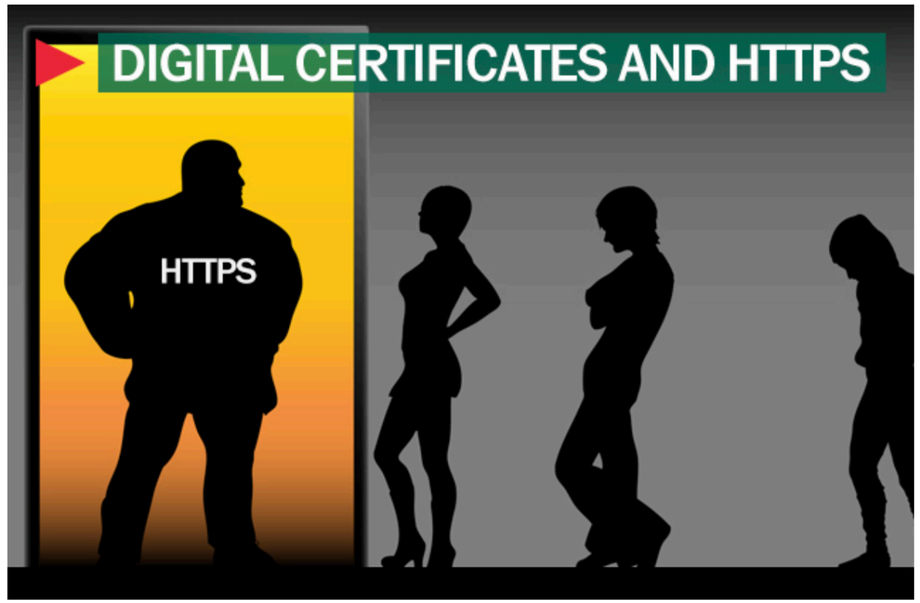
Case study: HTTPS



https://

GlobalSign, 2018

kaspersky daily

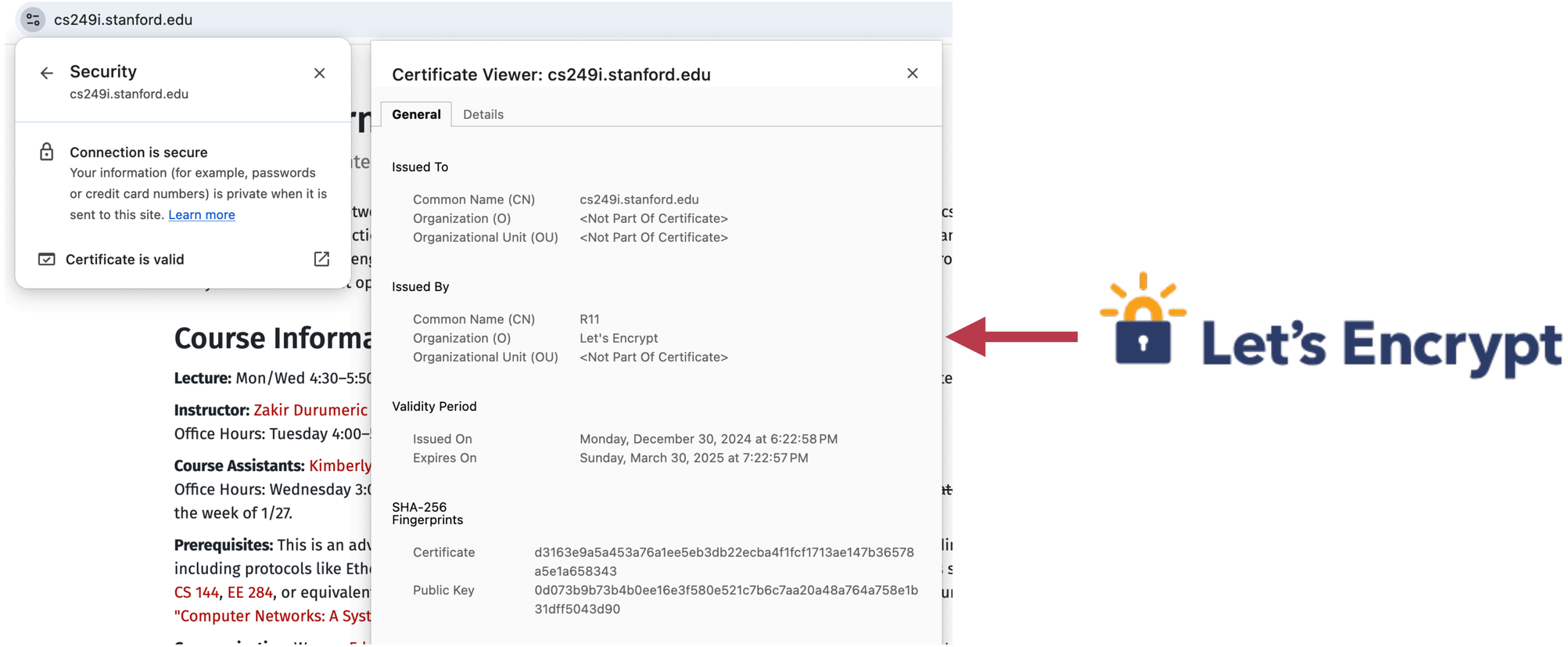


By now, we’ve all been told to look for the little green or grey padlock that, in most browsers, signifies the presence of the secure variety of the hypertext transfer protocol (HTTPS).But what is ‘HTTPS’?

Kaspersky, 2013

“Tangled Web” of dependencies

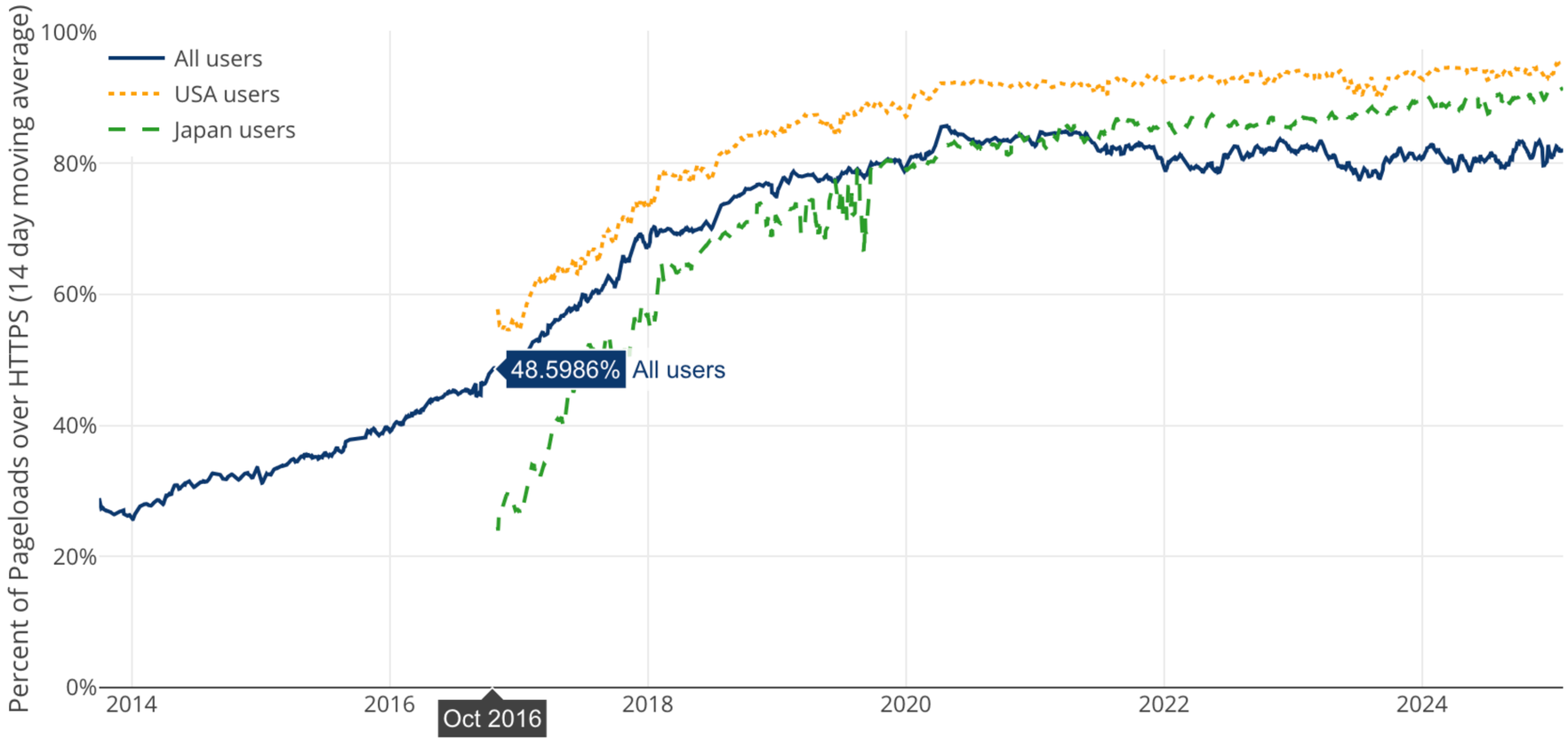
Case study: HTTPS



“Tangled Web” of dependencies

Case study: HTTPS

Let’s Encrypt



https://letsencrypt.org/stats/#percent-pageloads

“Tangled Web” of dependencies

Case study: HTTPS

- > 90% of websites use third-party resources
  - Median: 23 third-party resource loads
- **28%** of websites blocked from upgrading to HTTPS ***because of third-party dependencies*** that are unavailable over HTTPS

🔗 *Weakest link* security across a publisher’s third-party resource dependencies

Security Challenges in an Increasingly Tangled Web

👋

Deepak Kumar[†] Zane Ma[†] Zakir Durumeric[‡†] Ariana Mirian[‡]
Joshua Mason[‡] J. Alex Halderman[‡] Michael Bailey[†]
[†]University of Illinois, Urbana Champaign [‡]University of Michigan
{dkumar11, zanema2, joshm, mdbailey}@illinois.edu {zakir, amirian, jhalderm}@umich.edu

[WWW ’17]

URL	Owner	T10K	T100K	T1M
b.scorecardresearch.com	comScore	27.2%	12.4%	5.3%
*.casalemedia.com	Casale Media	22.1%	10.7%	2.5%
*.baidu.com	Baidu	7.8%	7.9%	1.7%
*.sharethis.com	ShareThis	2.1%	2.6%	1.6%
www.statcounter.com	StatCounter	1.2%	1.3%	1.5%
cdn.turn.com	Turn Inc.	7.8%	4.0%	1.3%
cloudfront-labs.amazonaws.com	Alexa	11.0%	4.1%	1.2%
global.ib-ibi.com	Network Sol.	5.3%	3.1%	1.2%
a.adroll.com	AdRoll	1.9%	1.8%	1.1%
admaym.com	WideOrbit	3.5%	2.0%	0.8%
cdn.rubiconproject.com	Rubicon	6.8%	2.2%	0.6%

Figure 10: **Most Common HTTPS Blocking Domains**—We show the top sites that block the deployment of HTTPS and their prevalence in the top million sites. A site blocks the deployment of HTTPS if hosts a resource that appears on a site over HTTP and is not yet available over HTTPS.

Web Infrastructure Visibility

Research applications

The Chain of Implicit Trust: An Analysis of the Web Third-party Resources Loading

Muhammad Ikram
muhammad.ikram@mq.edu.au
Macquarie University
University of Michigan

Rahat Masood
rahat.masood@data61.csiro.au
UNSW and Data61, CSIRO

Gareth Tyson
g.tyson@qmul.ac.uk
Queen Mary University of London

Mohamed Ali Kaafar
dali.kaafar@mq.edu.au
Macquarie University and Data61, CSIRO

Noha Loizon
noha.loizon@data61.csiro.au
Data61, CSIRO

Roya Ensafi
ensafi@umich.edu
University of Michigan

Bug Fixes, Improvements, ... and Privacy Leaks

A Longitudinal Study of PII Leaks Across Android App Versions

Jingjing Ren*, Martina Lindorfer†, Daniel J. Dubois*,
Ashwin Rao‡, David Choffnes* and Narseo Vallina-Rodriguez§
*Northeastern University †UC Santa Barbara ‡University of Helsinki §IMDEA Networks Institute and ICSI
*{renjj, d.dubois, choffnes}@ccs.neu.edu †martina@iseclab.org ‡ashwin.rao@cs.helsinki.fi §narseo.vallina@imdea.org

Where is the energy spent inside my app?

Fine Grained Energy Accounting on Smartphones with Eprof

Abhinav Pathak
Purdue University
pathaka@purdue.edu

Y. Charlie Hu
Purdue University
ychu@purdue.edu

Ming Zhang
Microsoft Research
mzh@microsoft.com

On the Infrastructure Providers That Support Misinformation Websites

👋 **Catherine Han, Deepak Kumar, Zakir Durumeric**
Stanford University
cathan@stanford.edu, kumarde@stanford.edu, zakird@stanford.edu

TIME

TECH • INTERNET

Cloudflare Is One of the Companies That Quietly Powers the Internet. Researchers Say It's a Haven for Misinformation

4 MINUTE READ

BY CHRIS STOKEL-WALKER AUGUST 26, 2022 9:24 AM EDT

Research applications

Bug Fixes, Improvements, ... and Privacy Leaks

A Longitudinal Study of PII Leaks Across Android App Versions

Jingjing Ren*, Martina Lindorfer†, Daniel J. Dubois*,
Ashwin Rao‡, David Choffnes* and Narseo Vallina-Rodriguez§
*Northeastern University †UC Santa Barbara ‡University of Helsinki §IMDEA Networks Institute and ICSI
*{renjj, d.dubois, choffnes}@ccs.neu.edu †martina@iseclab.org ‡ashwin.rao@cs.helsinki.fi §narseo.vallina@imdea.org

The Chain of Implicit Trust: An Analysis of the Web Third-party Resources Loading

Muhammad Ikram
muhammad.ikram@mq.edu.au
Macquarie University
University of Michigan

Rahat Masood
rahat.masood@data61.csiro
UNSW and Data61, CSIRO

Gareth Tyson

Mohamed Ali Kaafar
dali.kaafar@mq.edu.au
Macquarie University and Data61, CSIRO

Noha Loizon
noha.loizon@data61.csiro.
Data61, CSIRO

How do you obtain better web visibility?

…nergy spent inside my app?

…counting on Smartphones with Eprof

Y. Charlie Hu
Purdue University
ychu@purdue.edu

Ming Zhang
Microsoft Research
mzh@microsoft.com

On the Infrastructure Providers That Support Misinformation Websites

👋 **Catherine Han, Deepak Kumar, Zakir Durumeric**
Stanford University
cathan@stanford.edu, kumarde@stanford.edu, zakird@stanford.edu

TIME

TECH • INTERNET

Cloudflare Is One of the Companies That Quietly Powers the Internet. Researchers Say It's a Haven for Misinformation

4 MINUTE READ

BY CHRIS STOKEL-WALKER AUGUST 26, 2022 9:24 AM EDT

Web Infrastructure Visibility

Business applications

similarweb

Check & Analyze Website Traffic With Similarweb

Reveal the most accurate, up-to-date traffic & engagement data to analyze your competitors' performance and strategies, and to optimize your own

- Enter any domain into the search bar above
- Click the 'Analyze traffic' button
- Get data & insight into the website's traffic stats, global rankings, SEO keywords, referring domains, top search ads, website technologies, and more!

Wappalyzer

Empower your sales and marketing teams

Website profiling
Find out what websites are built with.

Lead generation
Find prospects by the technologies they use.

Market research
Compare market shares and technology trends.

Competitor analysis
Discover who uses competitors' software.

Data enrichment
Technology, company and contact information.

Custom reports
Create lists of websites and contacts.

Website monitoring
Monitor website technology changes.

Browser extension
See the technologies on websites you visit.

CRM integration
See the technologies of your leads.

Email verification
Improve delivery and mailing list quality.

API access
Instant and real-time technology lookups.

Security recon
Reveal web technologies and version numbers.

Web Analytics and Tracking

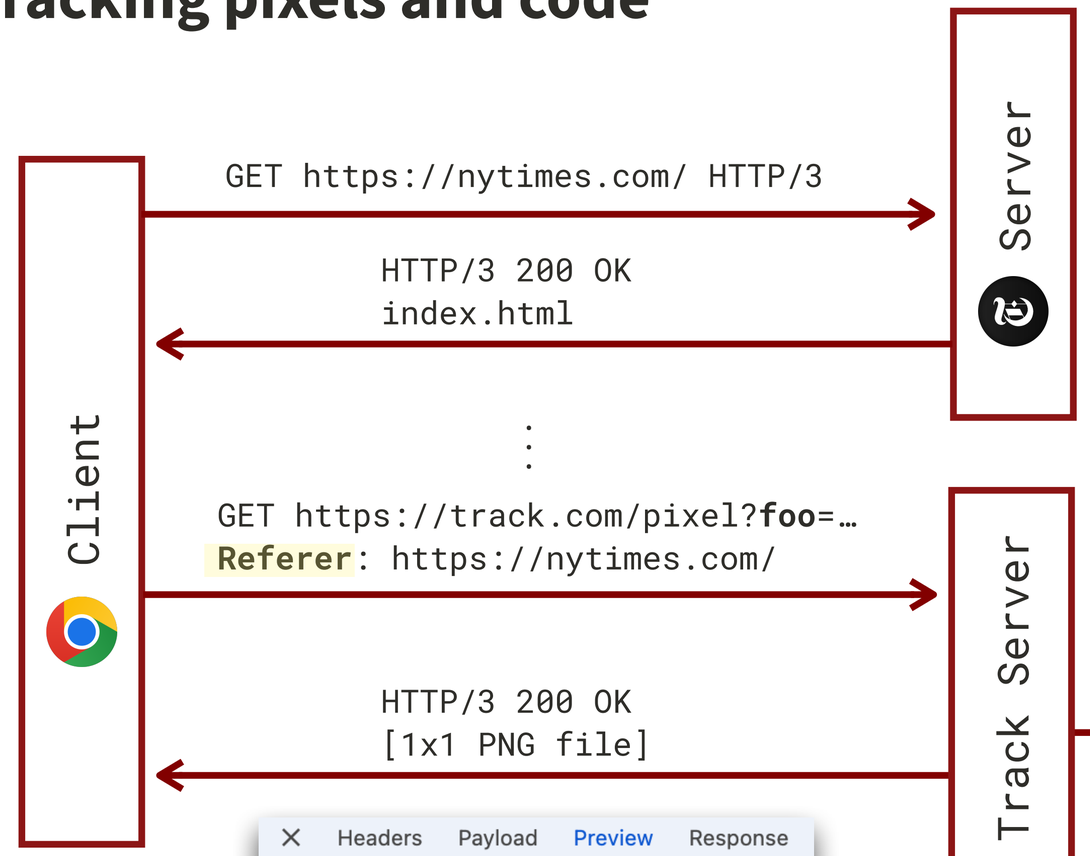
Web Tracking

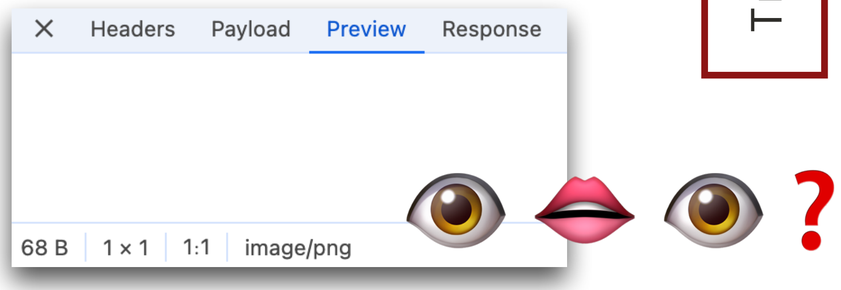
Cookies and pixels and fingerprinting, oh my!

- Techniques
  - Tracking code, pixels
  - Browser cookies
  - Fingerprinting

Tracking via HTTP

Tracking pixels and code



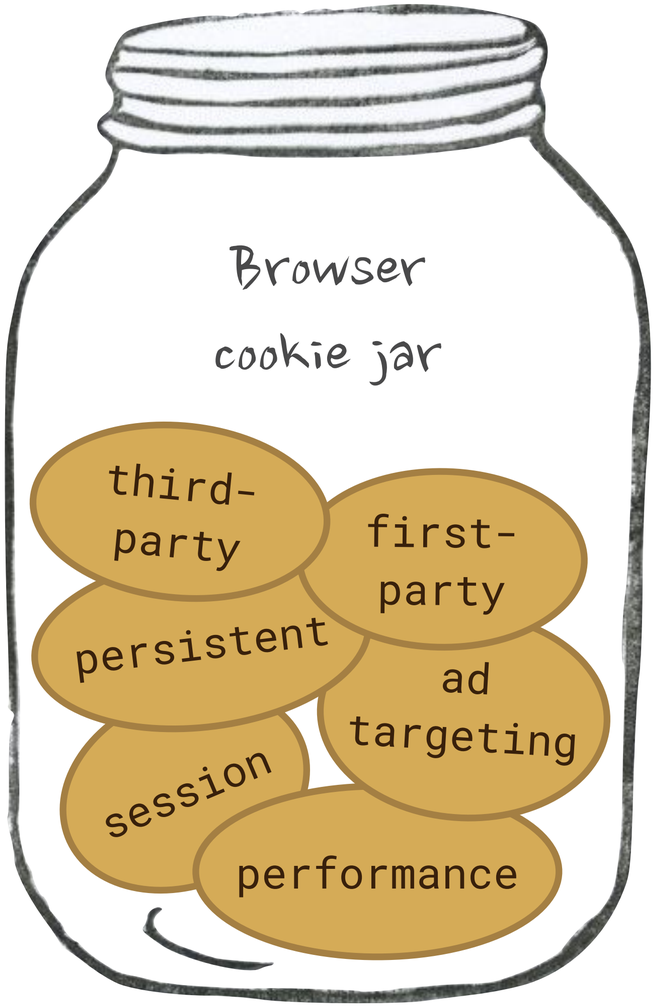


https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/

HTTP Cookies Refresher

Maintaining state in the client-side

"These header fields can be used by HTTP servers to **store state** (called **cookies**) at HTTP user agents, letting the servers maintain a stateful session over the mostly stateless HTTP protocol. Although cookies have many **historical infelicities that degrade their security and privacy**, the Cookie and Set-Cookie header fields are widely used on the Internet." RFC 6265 (2011)



Tracking via HTTP

Setting HTTP cookies

▼ General

Request URL:	https://a.et.nytimes.com/track
Request Method:	POST
Status Code:	200 OK
Remote Address:	52.89.31.9:443
Referrer Policy:	strict-origin-when-cross-origin

▼ Response Headers

Set-Cookie: SIDNY=CBoSMQjr_sGwBhCziJC9BhoSMS10v3
5G8CetLbZ_6Mk5iQa8IPP8yFMqAh5VOP2nu4
sGQgAaQOD8cN0cTBRrwCU2f5gihl7yeaFCsCY
iXnmRjajeBMviw4DqhhLwQjxX9kJltz0jus24nDU
B-9-JKi7_C65efgg=; Domain=.nytimes.com;
Path=/; Secure; HttpOnly; Max-Age=7776000

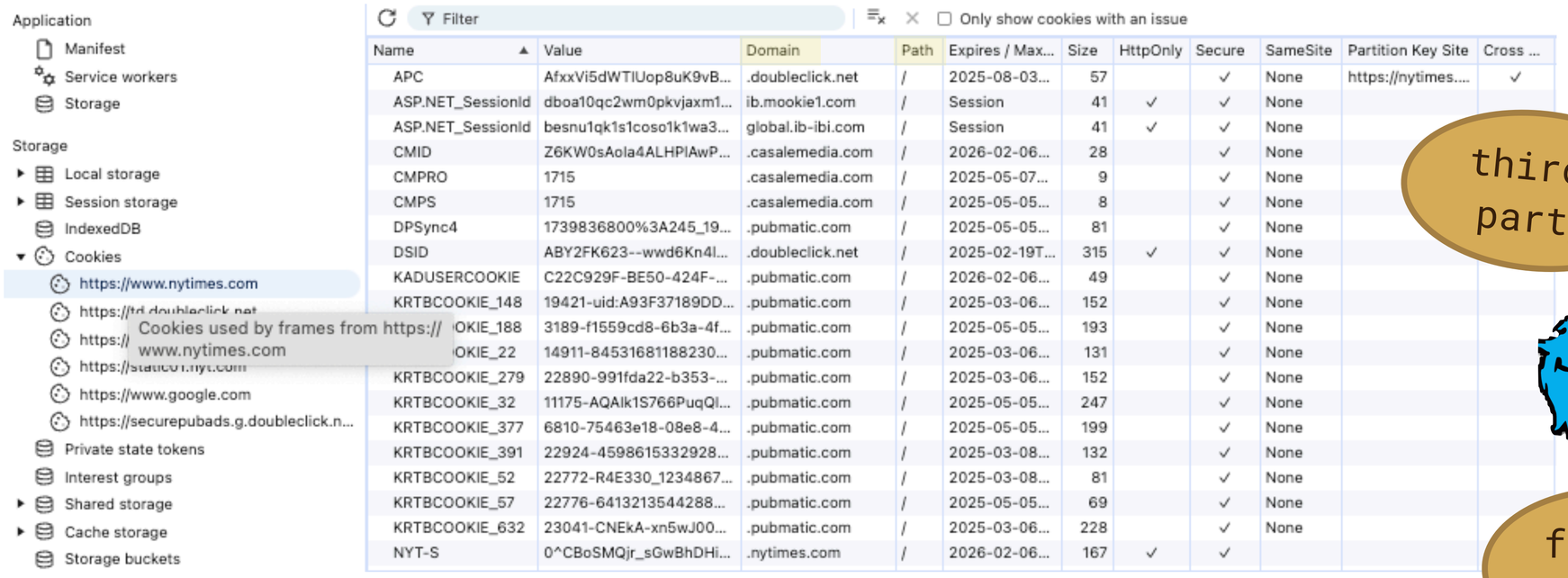
Request Cookies ☐ show filtered out request cookies

Name	Value	Domain	Pa...	Expires /...	Si...
NYT-S	0^CBoS...	.nytimes.com	/	2026-02...	167
SIDNY	CBoSMQj...	.nytimes.com	/	2025-05...	165
__eoi	ID=3a92a...	.nytimes.com	/	2025-08...	82
__gads	ID=0b89...	.nytimes.com	/	2026-0...	89
__gpi	UID=000...	.nytimes.com	/	2026-0...	89

HTTP Cookies Refresher

Maintaining state in the client-side

- Cookie scope: defined by `Domain` and `Path` attributes



What’s so yummy about third-party cookies?

https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies#define_where_cookies_are_sent

Evolution of Cookies

Evercookies (aka Zombie Cookies)

DESCRIPTION

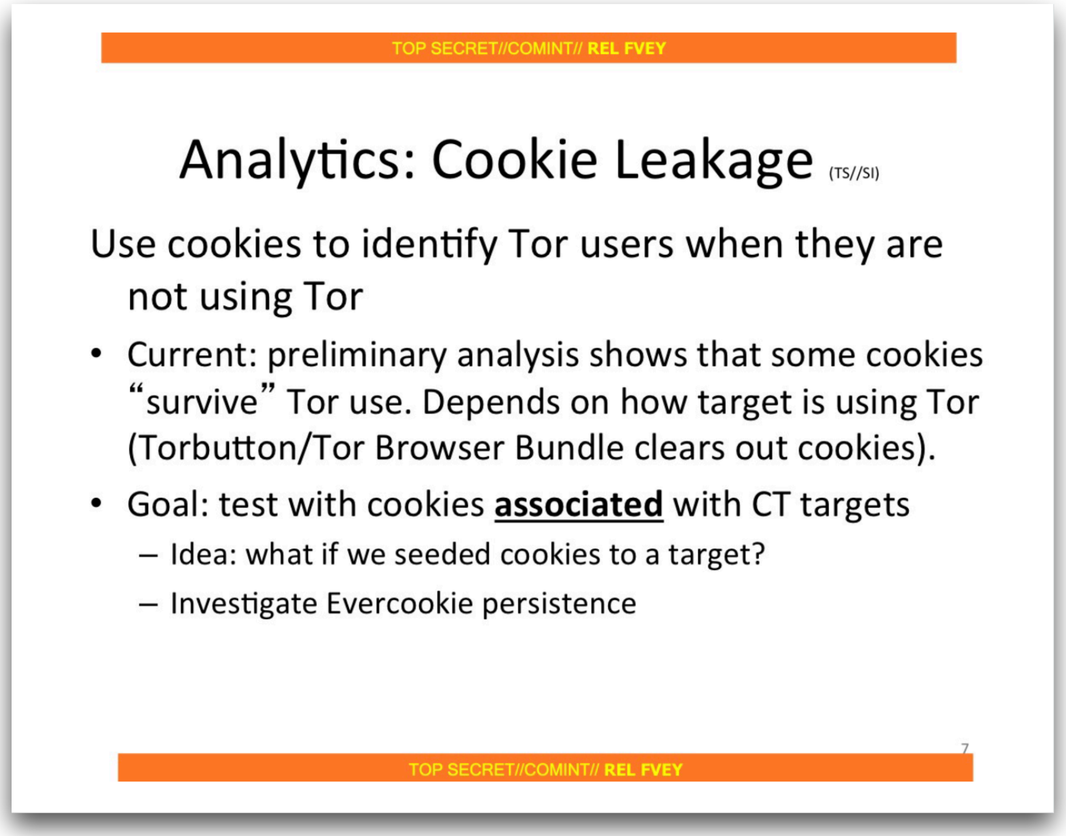
evercookie is a javascript API available that produces extremely persistent cookies in a browser. Its goal is to identify a client even after they've removed standard cookies, Flash cookies (Local Shared Objects or LSOs), and others.

evercookie accomplishes this by storing the cookie data in several types of storage mechanisms that are available on the local browser. Additionally, if evercookie has found the user has removed any of the types of cookies in question, it recreates them using each mechanism available.

https://samy.pl/evercookie/

Stinks

"Tor Stinks" NSA Top Secret Presentation (2012)



https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded

Evolution of Cookies

Stinks

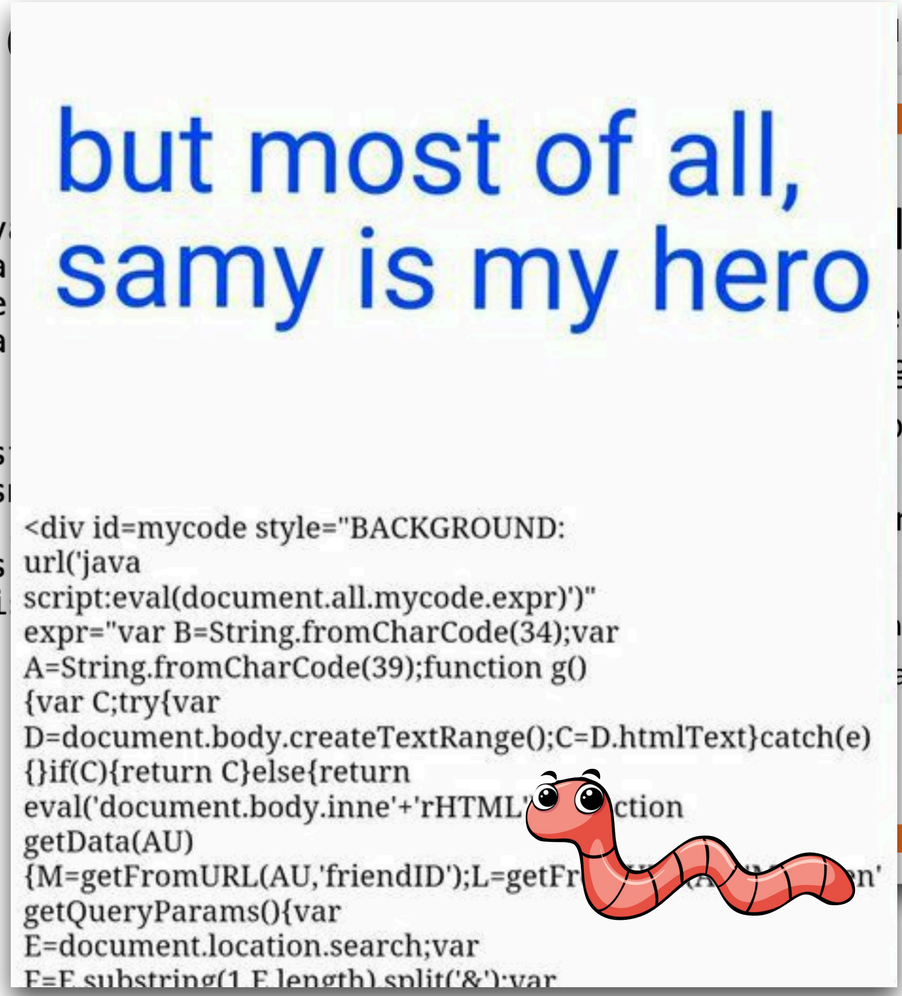
Evercookies (aka Zombie

DESCRIPTION

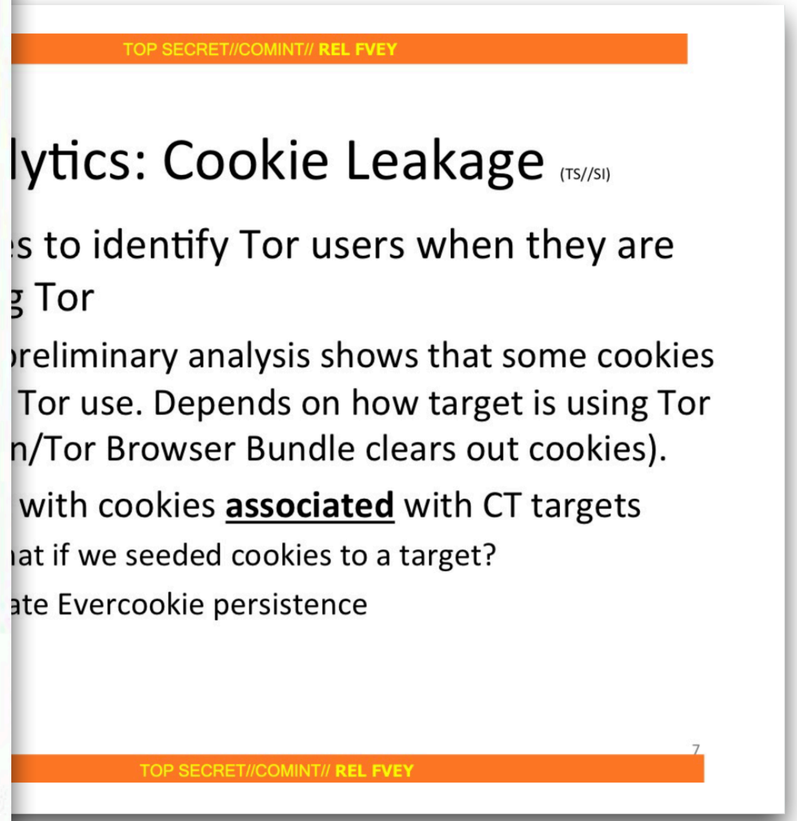
evercookie is a javascript API av
extremely persistent cookies in a
is to identify a client even afte
cookies, Flash cookies (Local Sha
others.

evercookie accomplishes this by s
several types of storage mechanis
the local browser. Additionally,
user has removed any of the types
recreates them using each mechani

https://samy.pl/evercookie/



SA Top Secret Presentation (2012)



https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded

Evolution of Cookies

Same-Origin Policy (SOP) for scripts and documents

- **Same-origin policy**: restricts how a script loaded by one *origin* can interact with a resource from another origin's DOM (HTML, images, etc.)
  - **Origin**: `(protocol, port [if specified], host)`
  - Enforced by client browser
- Example: isolating website A from reading data (via JavaScript execution) from email service B for a user signed into B in the same browser

But wait, SOP doesn't apply to *cookies*…

Recall: cookie scope is not origin-based

Cookies Today

Cookie syncing

Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask

Panagiotis Papadopoulos
Brave Software
panpap@brave.com

Nicolas Kourtellis
Telefonica Research, Spain
nicolas.kourtellis@telefonica.com

Evangelos P. Markatos
FORTH-ICS, Greece
markatos@ics.forth.gr

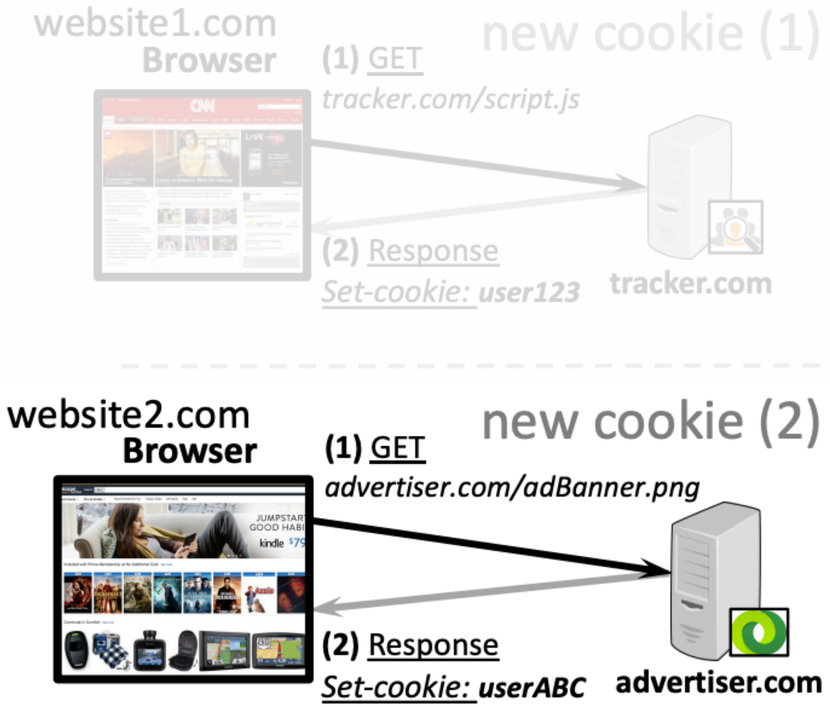
[WWW '19]

💡 — different companies collude to *share* cookie information using ***redirects*** to subvert SOP isolation

Say: `tracker.com` and `advertiser.com` are in cahoots to be the most pervasive ad targeting duo!

Cookies Today

Cookie syncing



💡 — different companies collude to *share* cookie information using ***redirects*** to subvert SOP isolation

Say: `tracker.com` and `advertiser.com` are in cahoots to be the most pervasive ad targeting duo!

Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask

Panagiotis Papadopoulos, Brave Software, panpap@brave.com
Nicolas Kourtellis, Telefonica Research, Spain, nicolas.kourtellis@telefonica.com
Evangelos P. Markatos, FORTH-ICS, Greece, markatos@ics.forth.gr

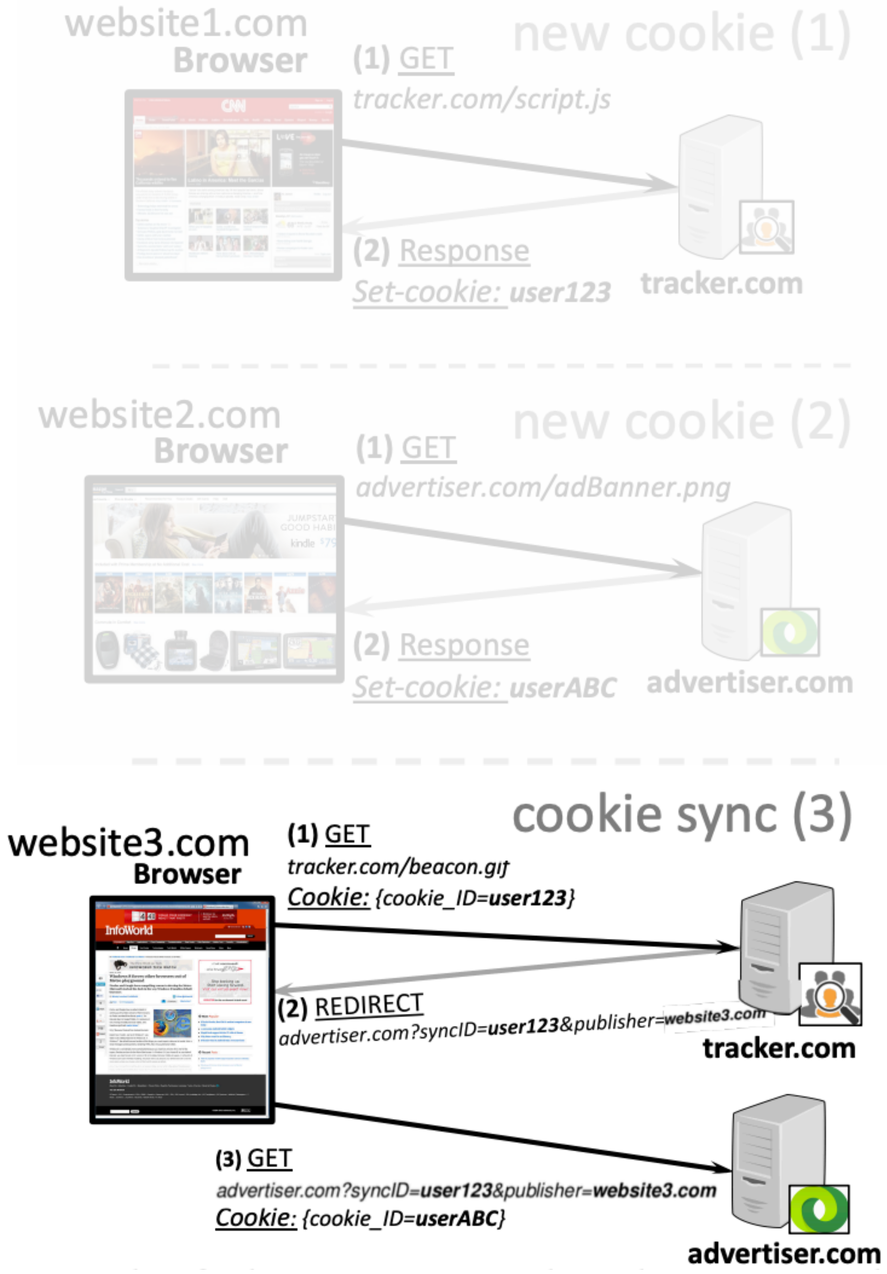
[WWW '19]

Cookies Today

Cookie syncing

💡 — different companies collude to *share* cookie information using ***redirects*** to subvert SOP isolation

Say: `tracker.com` and `advertiser.com` are in cahoots to be the most pervasive ad targeting duo!



Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask

Panagiotis Papadopoulos
Brave Software
panpap@brave.com

Nicolas Kourtellis
Telefonica Research, Spain
nicolas.kourtellis@telefonica.com

Evangelos P. Markatos
FORTH-ICS, Greece
markatos@ics.forth.gr

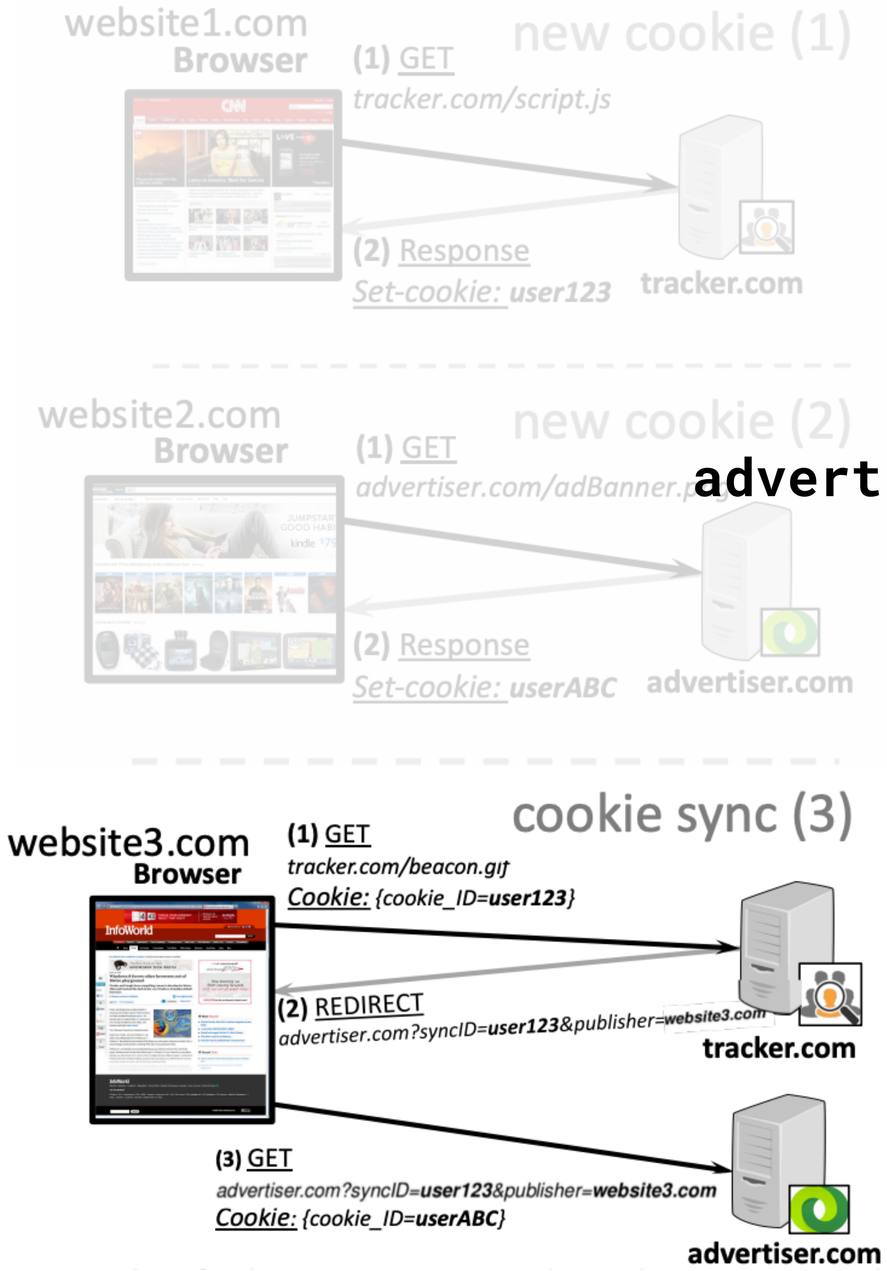
[WWW '19]

Cookies Today

Cookie syncing

💡 — different companies collude to *share* cookie information using ***redirects*** to subvert SOP isolation

Say: `tracker.com` and `advertiser.com` are in cahoots to be the most pervasive ad targeting duo!



Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask

Panagiotis Papadopoulos, Brave Software, panpap@brave.com
Nicolas Kourtellis, Telefonica Research, Spain, nicolas.kourtellis@telefonica.com
Evangelos P. Markatos, FORTH-ICS, Greece, markatos@ics.forth.gr

[WWW '19]

`advertiser.com` 🤝 `tracker.com`

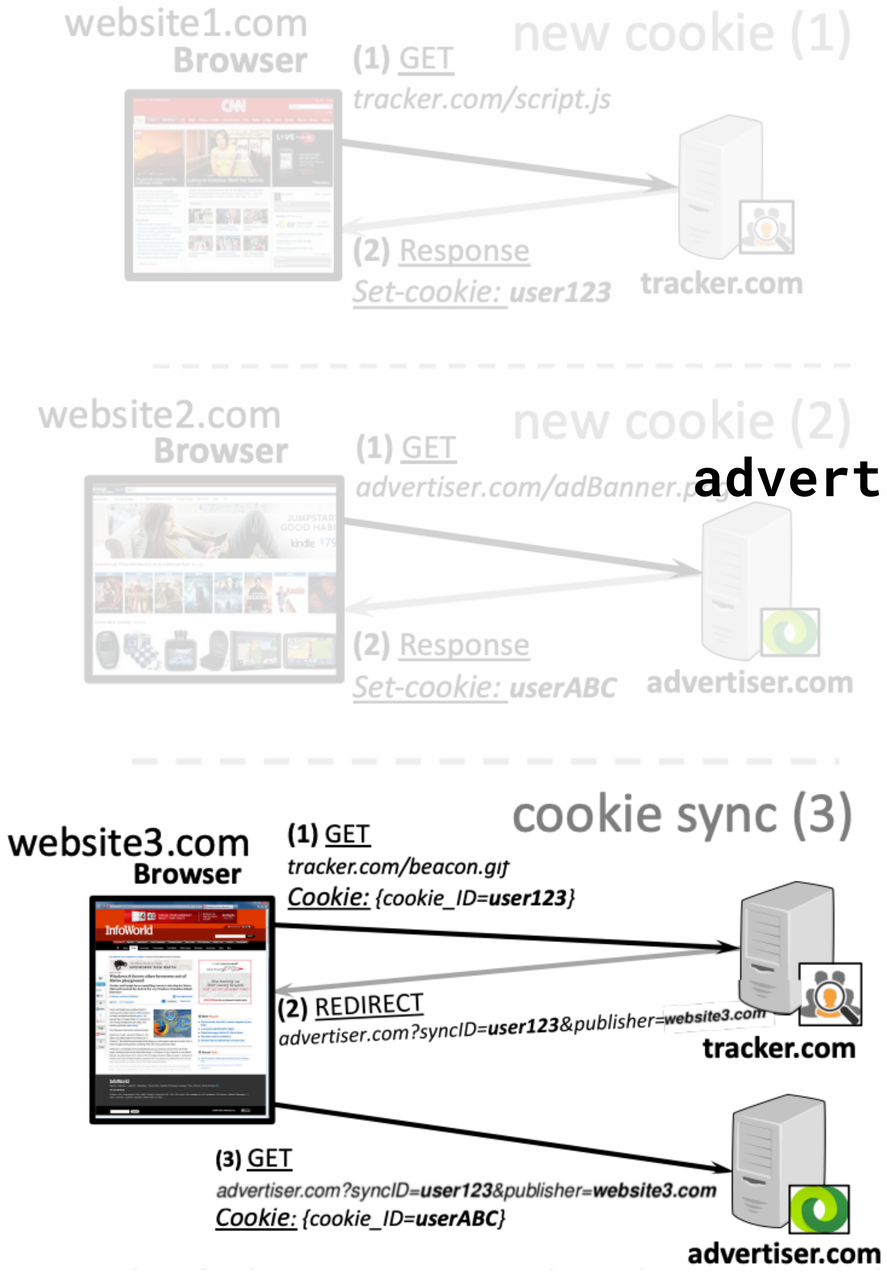
tracking `user123`/ `userABC` across all sites they visit

Cookies Today

Cookie syncing

💡 — different companies collude to *share* cookie information using ***redirects*** to subvert SOP isolation

Say: `tracker.com` and `advertiser.com` are in cahoots to be the most pervasive ad targeting duo!



Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask

Panagiotis Papadopoulos, Brave Software, panpap@brave.com
Nicolas Kourtellis, Telefonica Research, Spain, nicolas.kourtellis@telefonica.com
Evangelos P. Markatos, FORTH-ICS, Greece, markatos@ics.forth.gr

[WWW '19]

2019: **78%** of top 200 websites use CSync!

`advertiser.com` 🤝 `tracker.com`

tracking `user123`/ `userABC` across all sites they visit

No More Third-Party Cookies!

The Verge

Privacy and ads in Chrome are about to become FLoCing complicated

Google is forging ahead with its third-party cookie replacement technology

by **Dieter Bohn**
Mar 30, 2021, 11:00 AM PDT

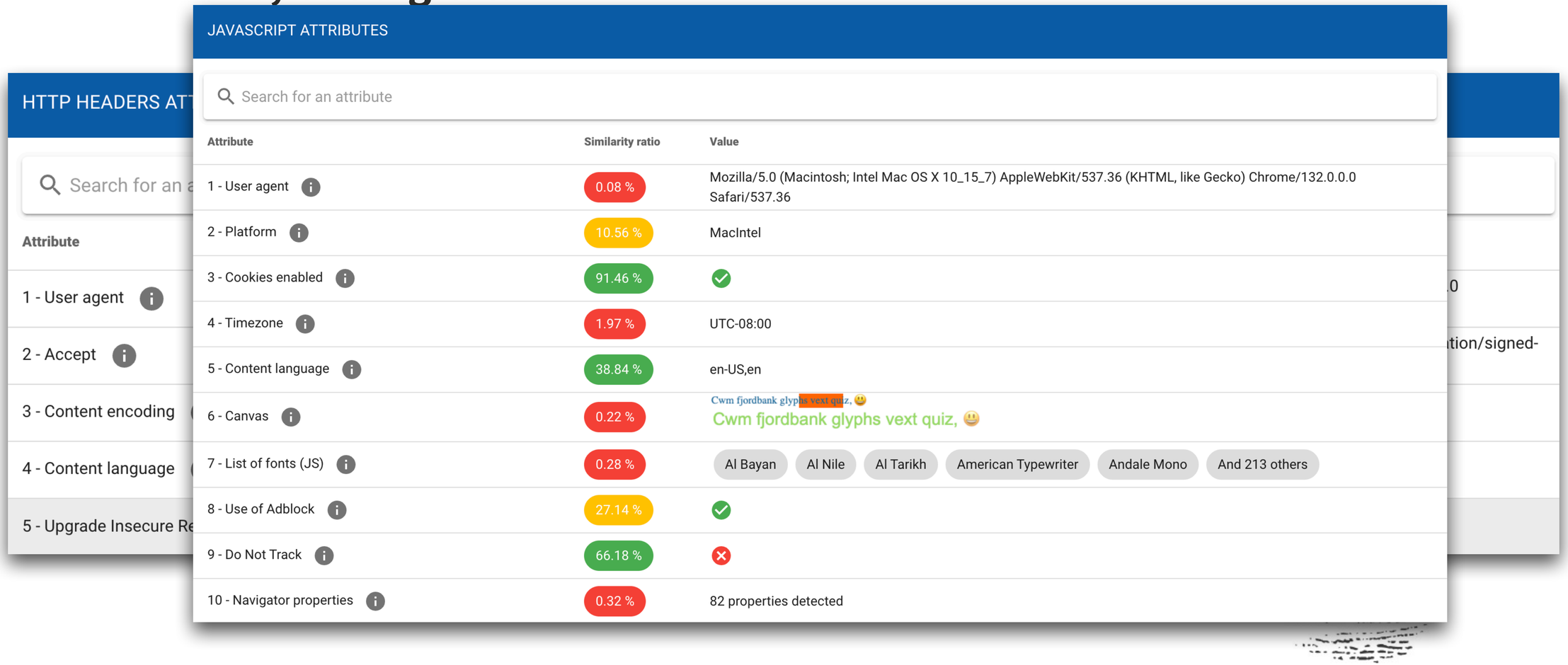
The basic argument from the Chrome team is that erecting a so-called "privacy wall" will entice legitimate ad tech companies into succumbing to the temptation of fingerprinting. Google is hoping that ad tech companies will adopt FLoC as an alternative.

No More Third-Party Cookies!

Oh wait…

Fingerprinting

Like biometrics, but digital?

- 

https://amiunique.org/fingerprint

Fingerprinting

Like biometrics, but digital?

- Combination of device hardware, browser properties, user behavior
- Harder to change, obfuscate, or spoof meaningfully
  - Parallel to biometrics in authentication

Web Tracking

Major players

Security Challenges in an Increasingly Tangled Web

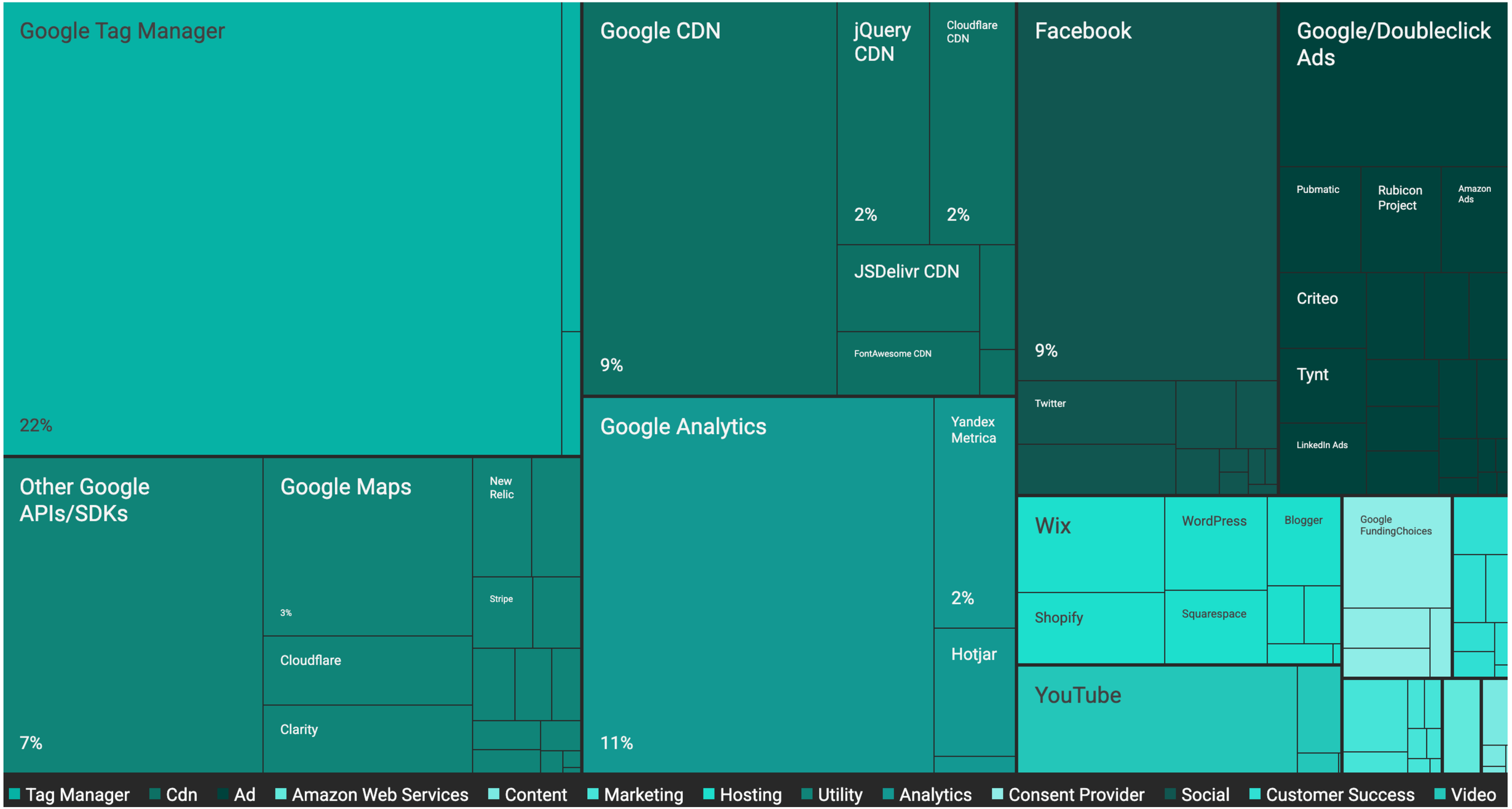
Deepak Kumar† Zane Ma† Zakir Durumeric‡† Ariana Mirian‡
Joshua Mason‡ J. Alex Halderman‡ Michael Bailey†
†University of Illinois, Urbana Champaign ‡University of Michigan
{dkumar11, zanema2, joshm, mdbailey}@illinois.edu {zakir, amirian, jhalderm}@umich.edu

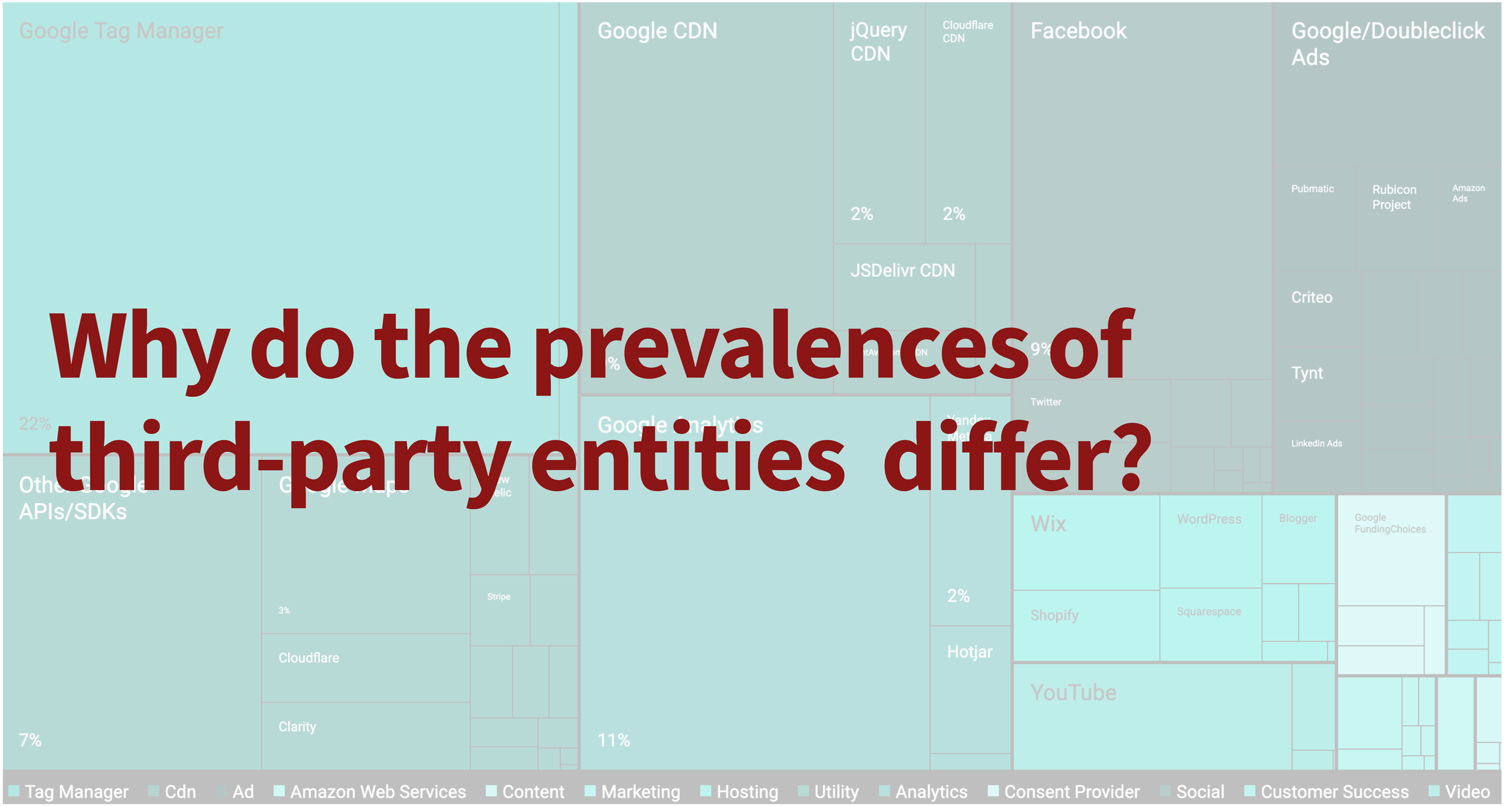
[WWW '17]

- Vantage point of prevalent entities
  - **Google** (82.2% 😱): analytics and advertising
  - **Facebook**: social sharing, tracking, advertising
  - **Cloudflare**, **Akamai**: CDNs

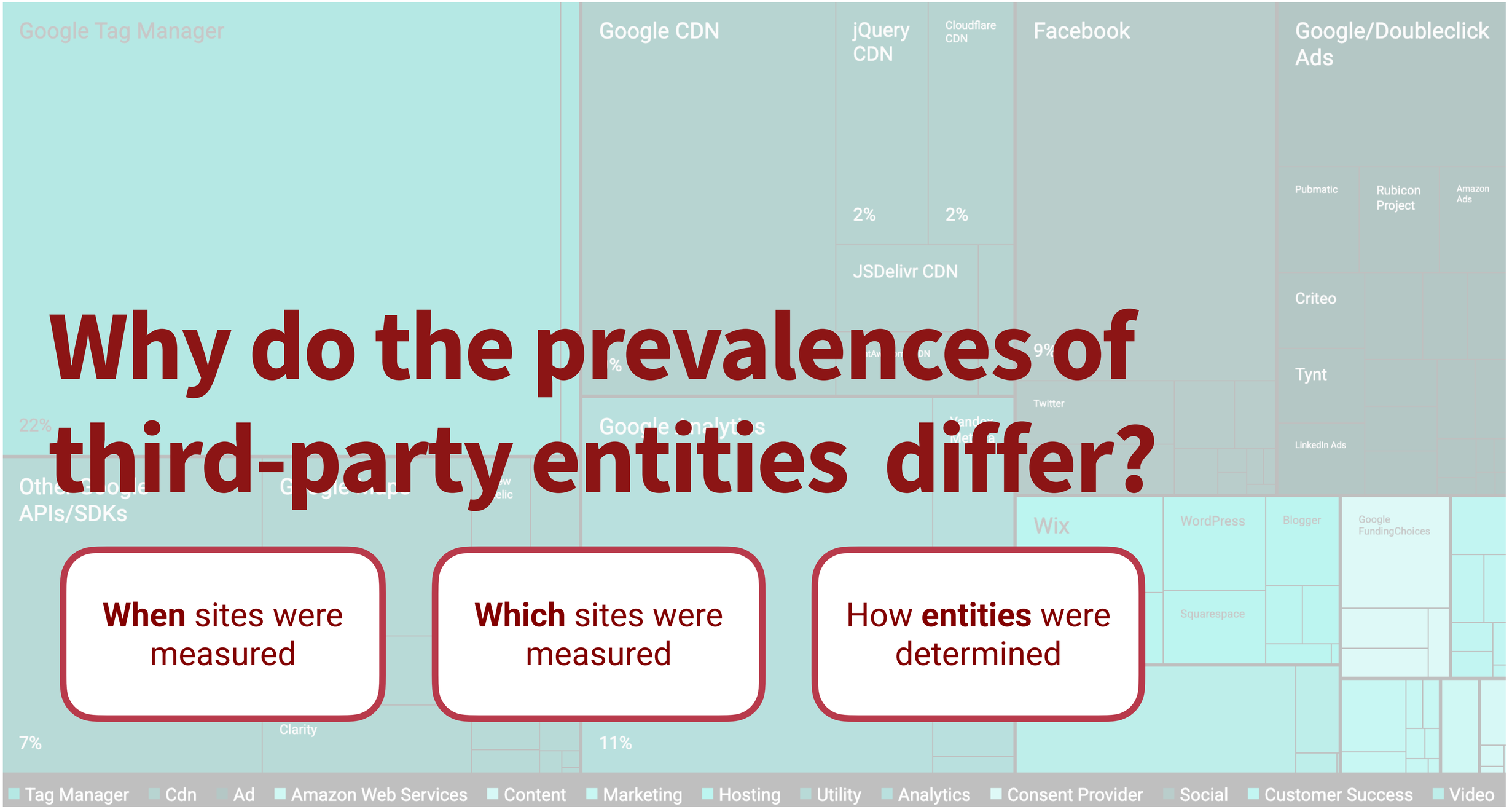
Company (by AS)	Type	% Top 1M
Google	All	82.2
Facebook	Social	34.1
Amazon EC2	Cloud	32.6
Cloudflare	CDN	30.7
Akamai	CDN	20.3

Most prevalent ASes that serve content in top 1M sites


Google Tag Manager
22%
Google CDN
9%
jQuery CDN
2%
Cloudflare CDN
2%
JSDelivr CDN
FontAwesome CDN
Facebook
9%
Twitter
Google/Doubleclick Ads
Pubmatic
Rubicon Project
Amazon Ads
Criteo
Tynt
LinkedIn Ads
Other Google APIs/SDKs
7%
Google Maps
3%
Cloudflare
Clarity
New Relic
Stripe
Google Analytics
11%
Yandex Metrica
2%
Hotjar
Wix
WordPress
Blogger
Shopify
Squarespace
Google FundingChoices
YouTube
Tag Manager
Cdn
Ad
Amazon Web Services
Content
Marketing
Hosting
Utility
Analytics
Consent Provider
Social
Customer Success
Video



Why do the prevalences of third-party entities differ?



Why do the prevalences of third-party entities differ?

When sites were measured

Which sites were measured

How **entities** were determined

Top Lists

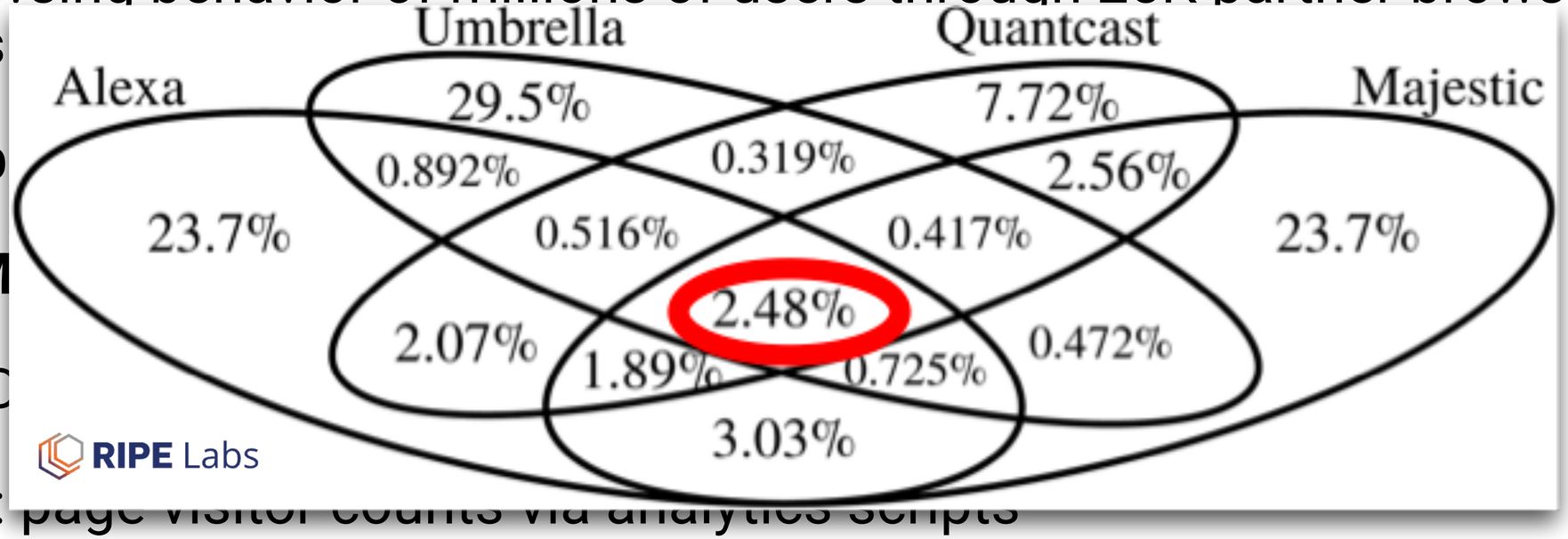
Which sites were measured

- Site “popularity” as a proxy for:
  - Scale of impact for some phenomenon
  - How mainstream a site is
- Web measurements: historically use a “top 1M” list
- Exercise: define a metric for website popularity

Top Lists

Overview of “popularity” definitions

- **Alexa**: browsing behavior of millions of users through 25K partner browser extensions
- **Cisco Umb**
- **Majestic M**
- **SecRank**: D
- **Quantcast**: page visitor counts via analytics scripts
- **Tranco**: aggregates Alexa, Umbrella, and Majestic list over 30-day window to improve stability/robustness (time, adversarial manipulation) [NDSS ’19]



Top Lists

If “popularity” is squishy…

- 2022: Google Chrome publishes Chrome User Experience Report (CrUX)
  - E
    u
    u
  - 🔗 https://github.com/zakird/crux-top-lists
- Exercise: CrUX limitations?

Toppling Top Lists: Evaluating the Accuracy of Popular Website Lists

Kimberly Ruth
Stanford University
kcruth@cs.stanford.edu

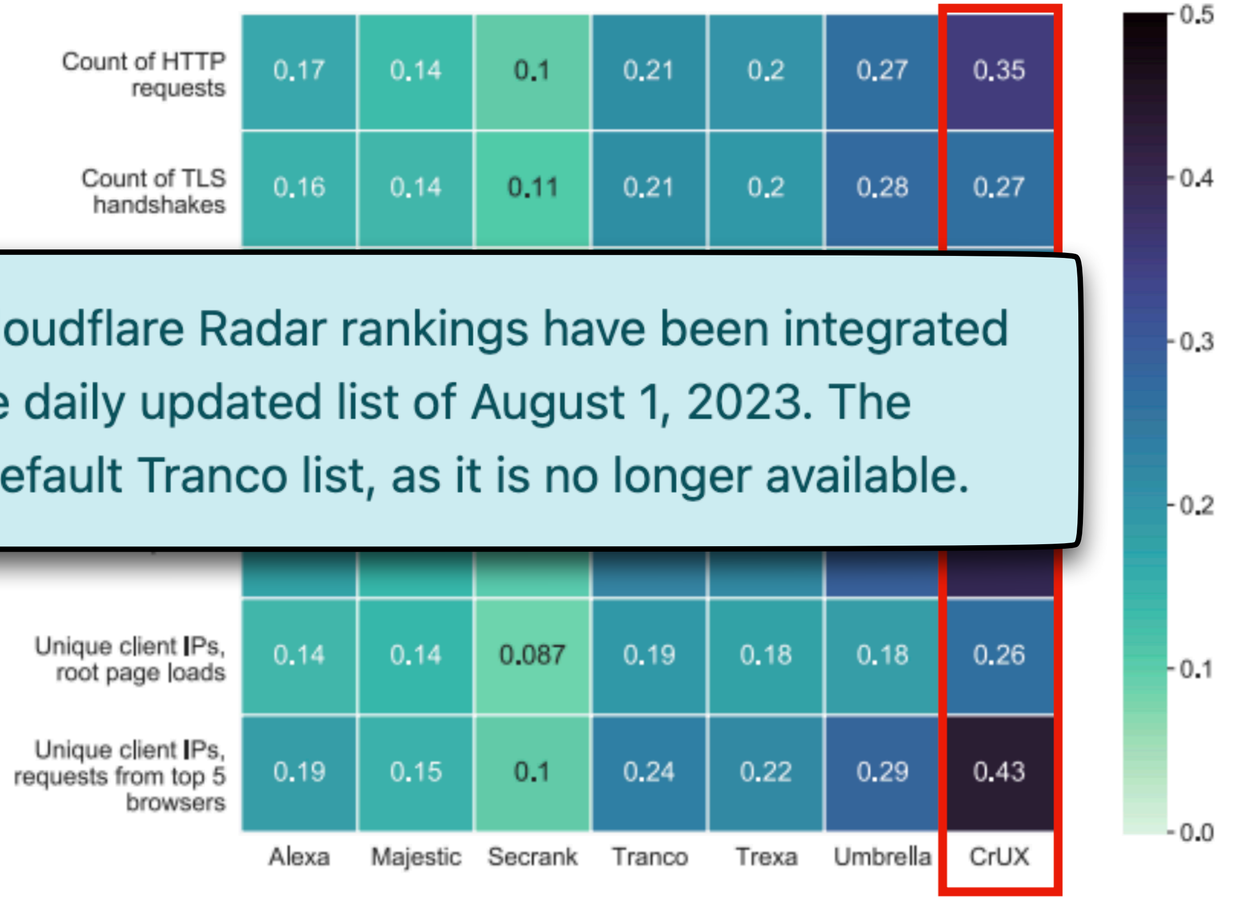
Deepak Kumar
Stanford University
kumarde@cs.stanford.edu

Brandon Wang
Independent Researcher
bmw4@illinois.edu

Luke Valenta
Cloudflare, Inc.
lvalenta@cloudflare.com

Zakir Durumeric
Stanford University
zakir@cs.stanford.edu

The Chrome User Experience Report and Cloudflare Radar rankings have been integrated into the default Tranco list, starting from the daily updated list of August 1, 2023. The Alexa ranking has been removed from the default Tranco list, as it is no longer available.



(a) Jaccard Index

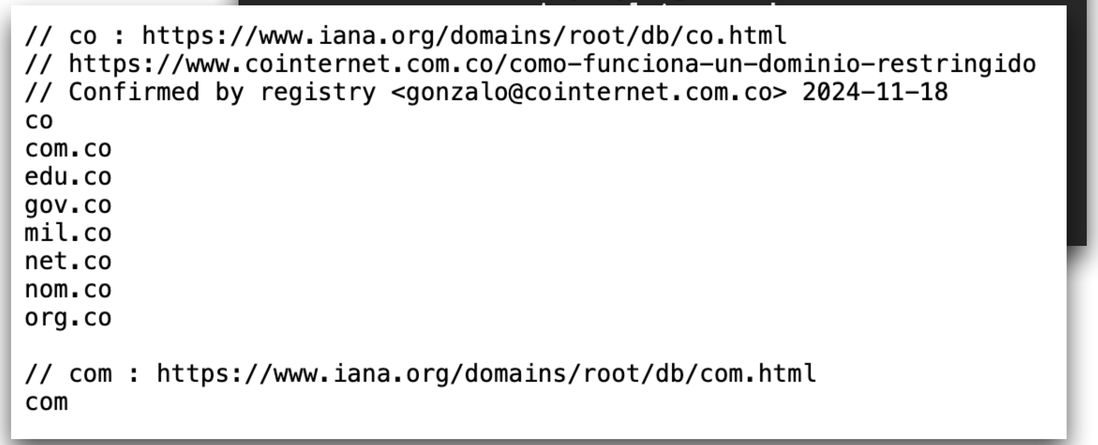
Web Tracking

Understanding entities

How **entities** were determined

- **Entity categorization**: for blocking and measurements
  - EasyList: https://easylist.to/

  - Third Party Web: https://github.com/patrickhulce/third-party-web
  - Public Suffix List (Mozilla): https://publicsuffix.org/learn/

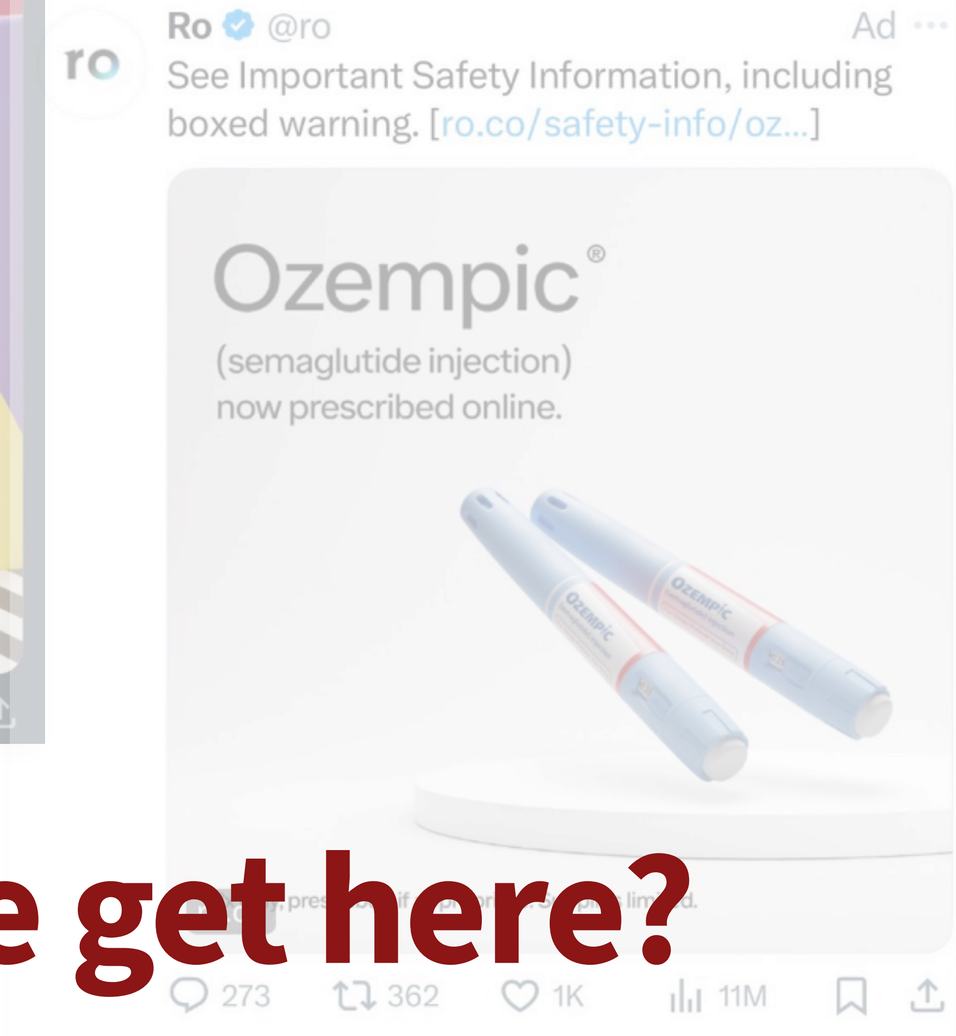


Web Monetization

Crocs @Crocs Ad
한정판 맥도날드 x 크록스 컬렉션을 지금 만나보세요!
x
crocs
crocs.co.kr
33
40
69K

Ro @ro Ad
See Important Safety Information, including boxed warning. [ro.co/safety-info/oz...]
Ozempic®
(semaglutide injection)
now prescribed online.
Rx only, prescribed if appropriate. Supplies limited.
ro.co
273
362
1K
11M

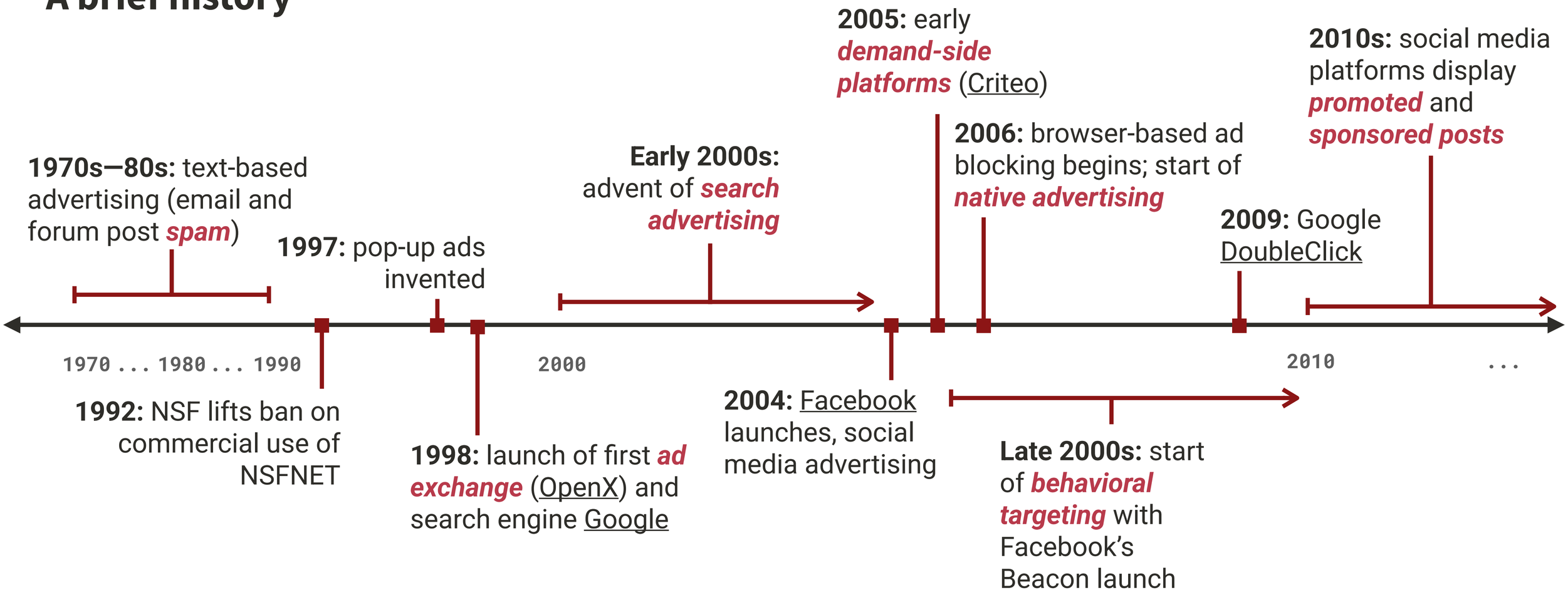
freezebyco
Sponsored
"How do you expect women in their 20s to be able to afford to freeze their eggs?"
WE DON'T — THAT'S WHY WE BUILT A FREE ALTERNATIVE
Match with a family that is seeking donor eggs to grow their family
They pay for your egg freezing and storage, and you split the eggs retrieved with them
Give yourself the gift of options while lifting up another family
SIGN UP



How did we get here?

Online Advertising

A brief history



[Timeline not to scale]

https://en.wikipedia.org/wiki/Timeline_of_online_advertising

Online Advertising

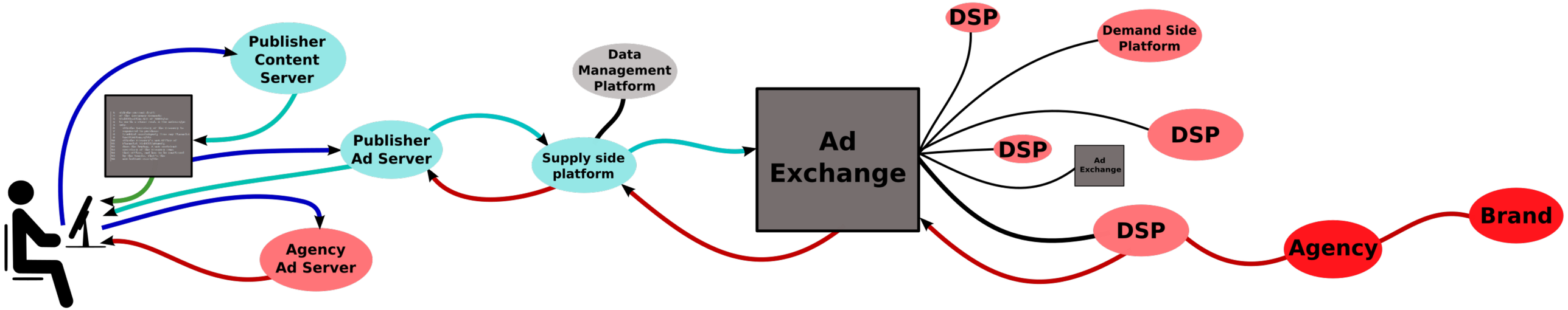
Leveraging web tracking

How do you sustain a business model [where users] don't pay?

- ***Targeted advertising*** based on user browsing behavior (among other things…)
  - 💡 the more targeted the ads…
    - Better conversion of advertiser ad spend (ROI)
    - More advertisers willing to pay a higher price for your ad slots
- How do price negotiation and ad placement work?

Online Advertising

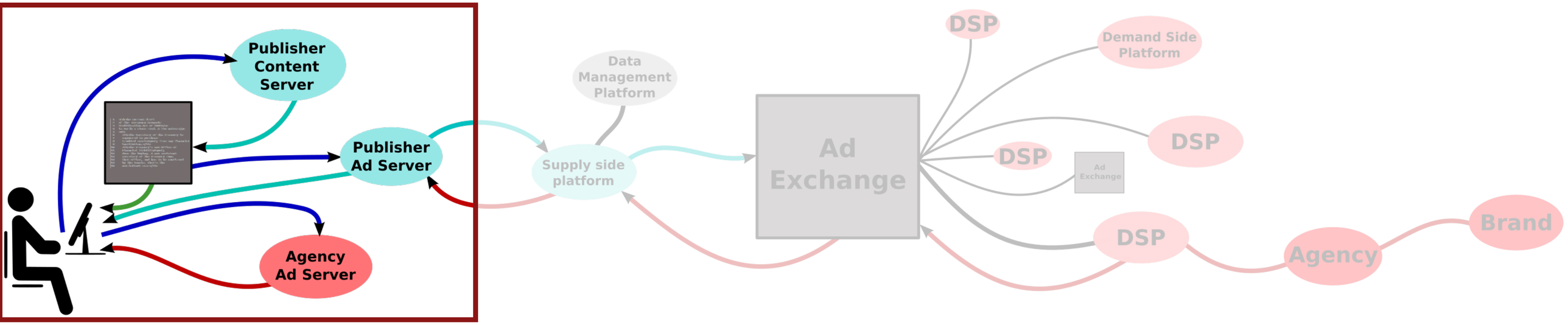
Stakeholders



https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Adservingfull.svg/2880px-Adservingfull.svg.png

Online Advertising

Publishers



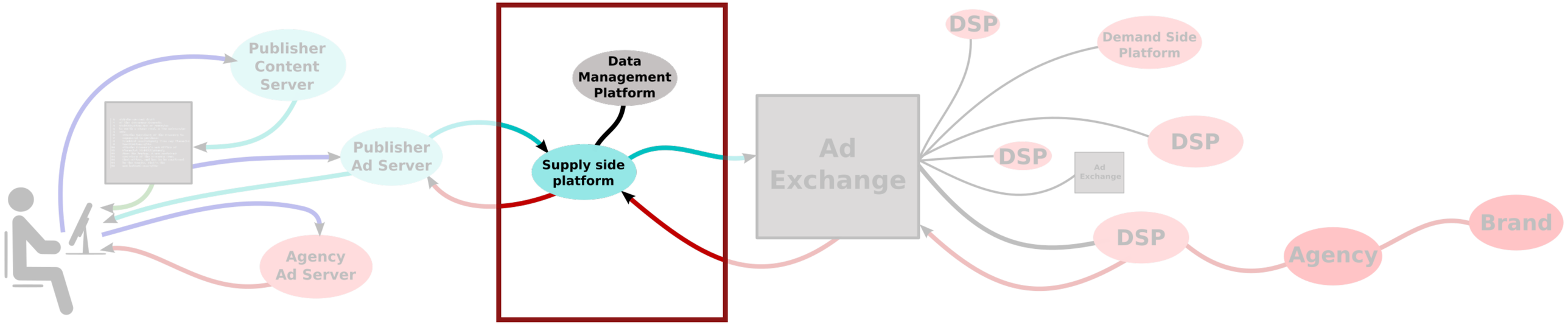
- Publishers: website operators (that “publish” site content)
  - Example: `nytimes.com`, `infowars.com`, `cs249i.stanford.edu`
- Ad slots: (1) dynamically/programmatically bought or (2) directly bought

Online Advertising

Supply-side platforms (SSPs)

DMP data:

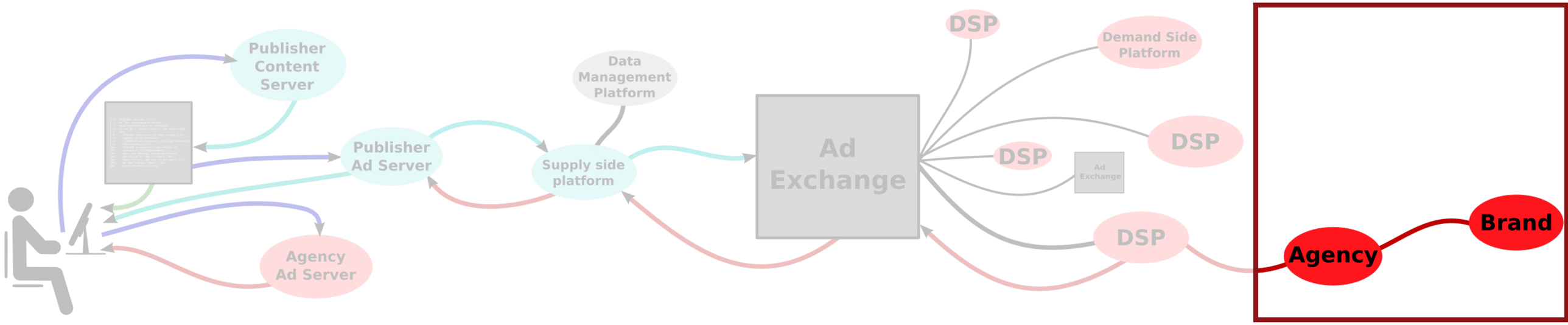
[FOO]



- SSPs service publishers (hence “supply-side”) to maximize the selling price of ad ***impressions***
  - Example: OpenX, Pubmatic, Google Ad Manager, AppNexus/Xandr
- DMPs provide SSPs data to create impression targeting profiles

Online Advertising

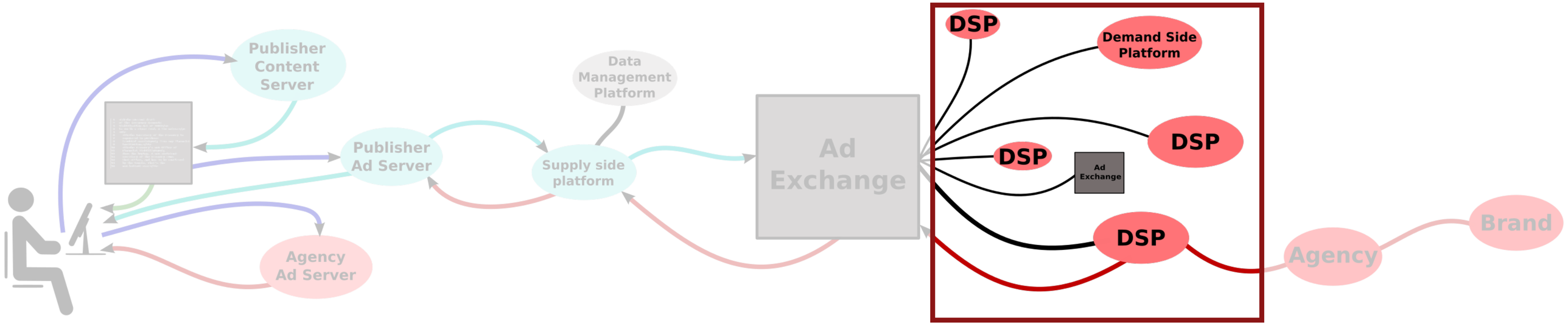
Advertisers



- Advertisers: entities looking to advertise to an online audience
  - Example: Crocs, Ozempic 😐 , political action committees

Online Advertising

Demand-side platforms (DSPs)

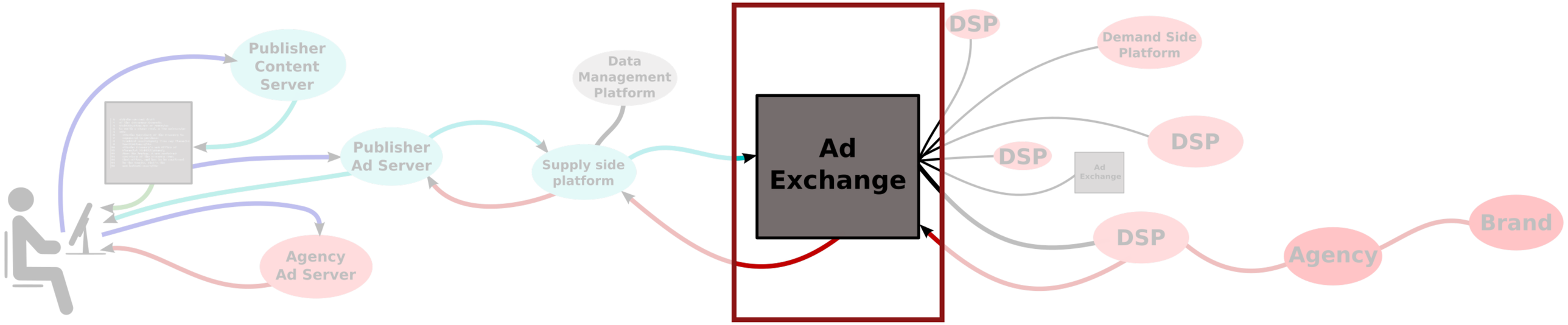


<100ms

- DSPs : advertisers : : SSPs : publishers
- Advertisers use DSPs to purchase ad impressions as cheaply as possible from ***Ad Exchanges*** via ***real-time bidding*** (RTB)
  - Example: Google DoubleClick, Criteo, QuantCast

Online Advertising

Ad exchanges



- Ad exchange: facilitates RTB for supply-side ad slots between demand-side bids based on the *properties* of the ad slot; takes a cut of the winning bid
  - Example: Google DoubleClick, Facebook Exchange, Pubmatic

Real-Time Bidding (RTB)

Programmatic advertising

- Real-time auction process facilitated by ad exchanges for ad inventory
- Pricing models:
  - Cost per mille (CPM): per 1K impressions 👑
  - Cost per click (CPC): per click
  - Cost per acquisition (CPA): per converted purchase
- Methods:
  - Waterfall bidding
  - Header bidding

Real-Time Bidding (RTB)

Request


```
{
  "id": "8652a8680db33faabbf3fa76150f35df50a67060",
  "imp": [
    {
      "id": "121-dt1",
      "banner": {
        "h": 250,
        "w": 300,
        "pos": 1
      },
      "bidfloor": 0.05
    },
    {
      "id": "121-dt2",
      "banner": {
        "h": 728,
        "w": 90,
        "pos": 0
      },
      "bidfloor": 0.12
    }
  ],
```


Impression / ad slot details


```
  "site": {
    "id": "15047",
    "domain": "dailymotion.com",
    "cat": "IAB1",
    "page": "http://www.dailymotion.com/video/xxeauj_www-dramacafe
-tv-hd-yyyy-yy-yyyyyyy-2012-yyyy_shortfilms",
    "publisher": {
      "id": "8796",
      "name": "dailymotion",
      "cat": "IAB3-1",
      "domain": "dailymotion.com"
    }
  },
```


Publisher details


```
  "user": {
    "id": "518c3da3717203f34019b038"
  },
  "device": {
    "ua": "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; Trid
ent/4.0; GTB7.4; (R1 1.6); SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5
.30729; .NET CLR 3.0.30729; Media Center PC 6.0)",
    "ip": "123.145.167.189"
  },
  "at": 1,
  "cur": [
    "USD"
  ]
}
```


User details

Real-Time Bidding (RTB)

Response

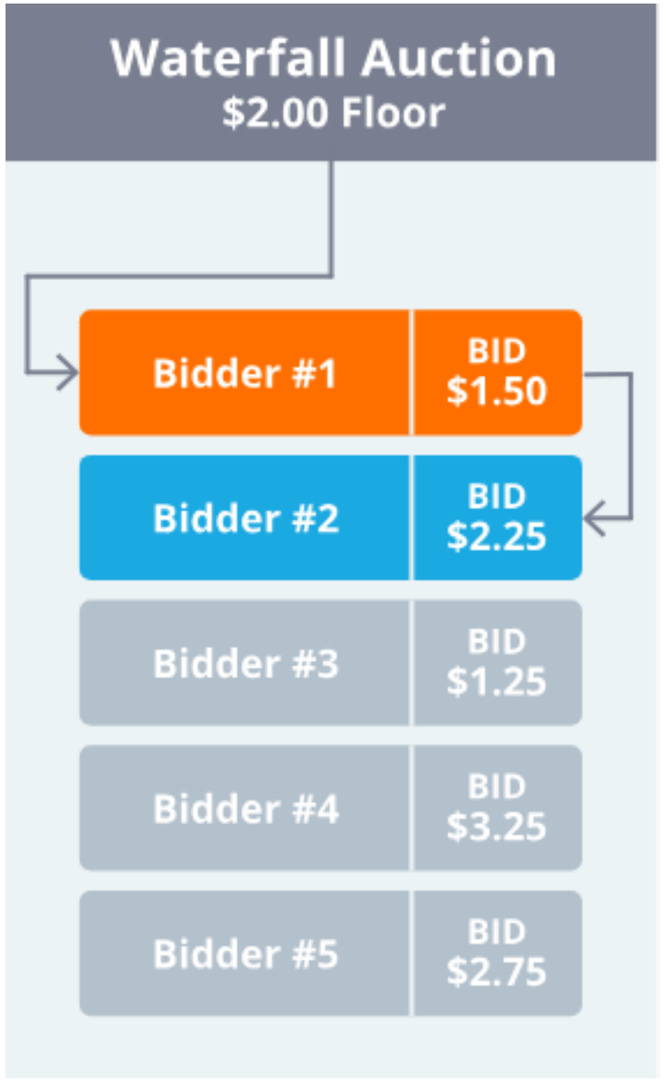

```
{
  "id": "BID-4-ZIMP-4b309eae-504a-4252-a8a8-4c8ceee9791a",
  "seatbid": [
    {
      "bid": [
        {
          "id": "24195efda36066ee21f967bc1de14c82db841f07",
          "impid": "24195efda36066ee21f967bc1de14c82db841f07",
          "price": 1.028428,
          "adid": "52a12b5955314b7194a4c9ff",
          "nurl": "http://ads.com/win/52a12b5955314b7194a4c9ff?won
=${AUCTION_PRICE}",
          "adm": "<iframe src=\"http://ads.com/render/52a12b595531
4b7194a4c9ff?won=${AUCTION_PRICE}\" width=\"728\" height=\"90\" fr
ameborder=\"0\" scrolling=\"no\" marginheight=\"0\" marginwidth=\"
0\" topmargin=\"0\" leftmargin=\"0\"></iframe>",
          "adomain": [
            "ads.com"
          ],
          "cid": "52a12b5955314b7194a4c9ff",
          "crid": "52a12b5955314b7194a4c9ff_1386294105",
          "attr": [],
          "dealid": "DX-1985-010A"
        }
      ],
      "seat": "42"
    },
```

Real-Time Bidding (RTB)

Waterfall bidding (traditional)

- Publishers give SSPs + ad exchanges pre-determined rank
  - Based on historical performance/ROI with publisher
  - Publisher sets floor bid rate
  - Impressions trickle down rank “waterfall” (along with floor rate) until they are sold
- Issues:
  - Slow (serial bidding offers)
  - Anti-competitive (Google has an SSP, DSP, Ad Exchange)

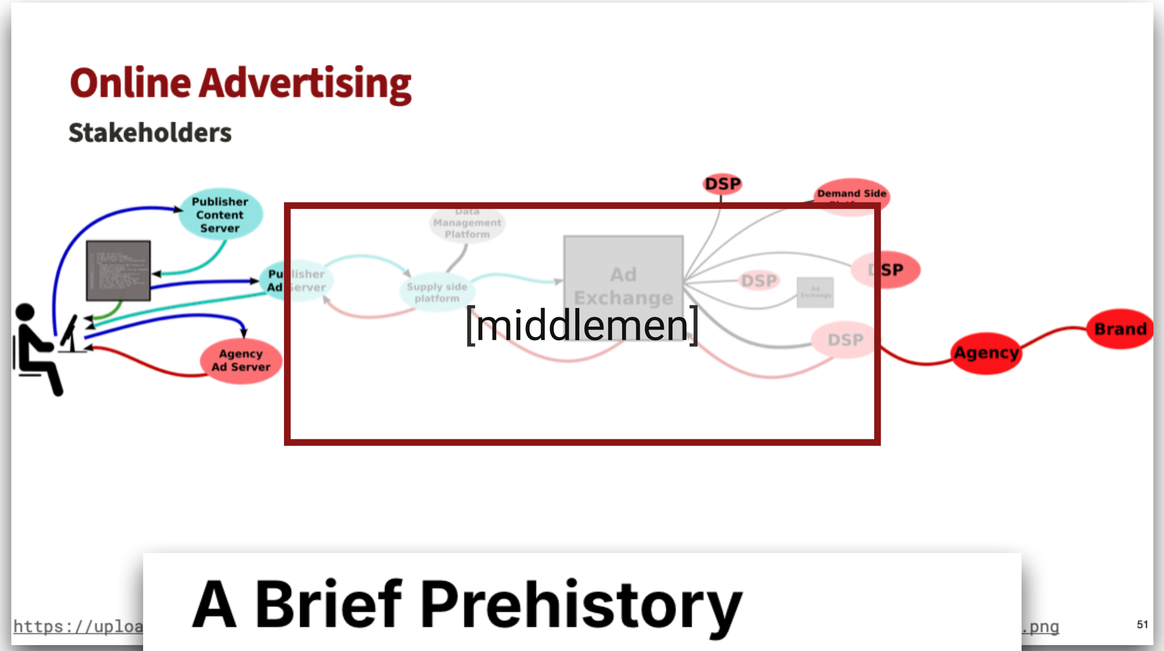
https://prebid.org/about/



Real-Time Bidding (RTB)

Header bidding motivation

- Waterfall bidding: anti-competitive and opaque to advertisers and publishers
- 💡 cut out middleman ad exchanges?
  - Publishers: maximize revenue
  - Advertisers: equal opportunity to bid



A Brief Prehistory

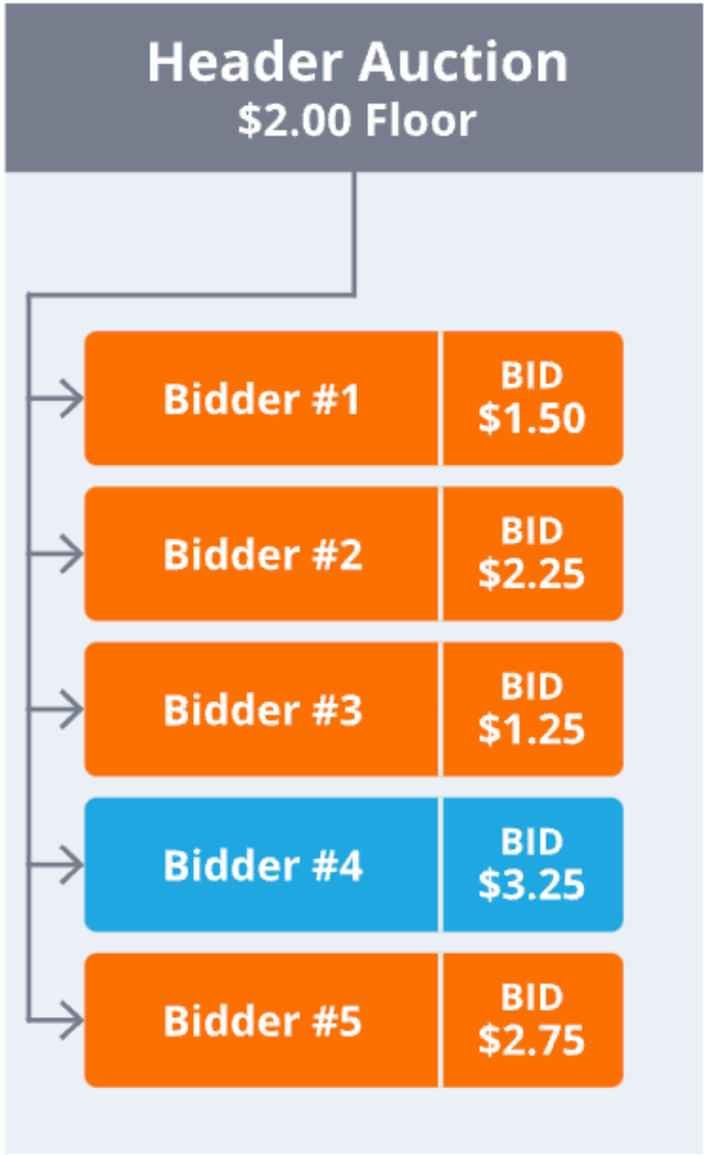
Header bidding is a response to the "waterfall" method, a fragmented and highly inefficient process for implementing programmatic advertising. Instead of offering impressions to one partner at a time, header bidding lets all partners bid simultaneously. With header bidding, publishers can also receive bids that may be unavailable through their primary ad server.

https://prebid.org/about/

Real-Time Bidding (RTB)

Header bidding

- Every bidder (DSP) can bid at the same time, so (theoretically) DSPs should bid closer to the true value of the impression
- Implementation:
  - Client-side (Prebid.js)
    - Finer access to cookies, but page load cost
  - Server-side (at SSP)
    - Requires CSync, latency to RTB pipeline



Real-Time Bidding (RTB)

Header bidding in client-side browser

No More Chasing Waterfalls: A Measurement Study of the Header Bidding Ad-Ecosystem

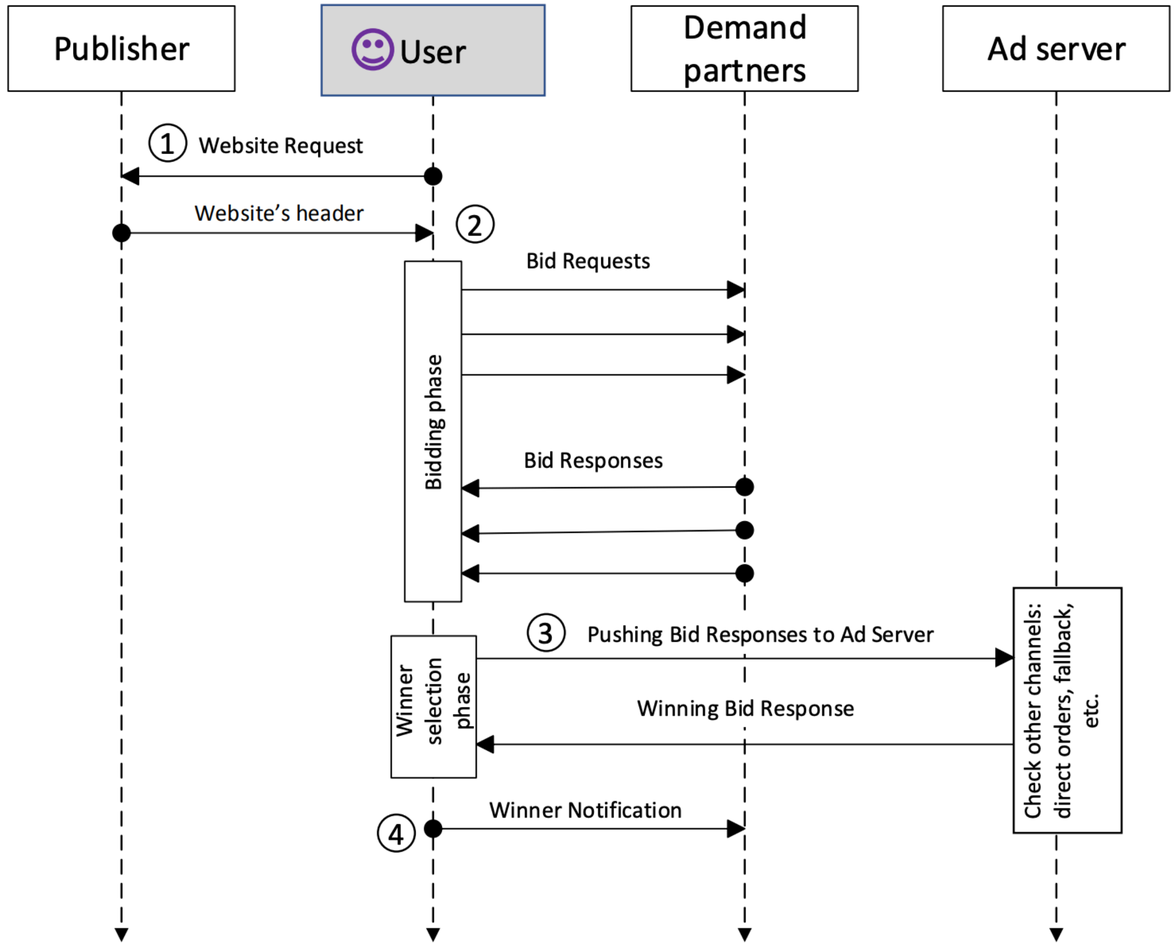
Michalis Pachilakis
FORTH, Greece
mipach@ics.forth.gr

Panagiotis Papadopoulos
Brave Software
panpap@brave.com

Evangelos P. Markatos
FORTH, Greece
markatos@ics.forth.gr

Nicolas Kourtellis
Telefonica Research, Spain
nicolas.kourtellis@telefonica.com

- 2019: ~15% of top websites use header bidding
  - Major DSPs in header bidding also dominate waterfall market
  - Header bidding **latency** is a median **3x higher** than waterfall



Other Monetization

Revenue streams, alternative platforms

Content creators / publishers:

- Donations
  - Mainstream payment processors (PayPal, Stripe)
  - Cryptocurrency exchanges, gift cards (Steam, Amazon)
- Subscriptions and affiliate marketing

Ad ecosystem middlemen:

- Selling user data

Other Monetization

Data brokering

- Consumer data-as-a-service!
  - Beyond targeted advertising
  - Background checks (housing and job applications)
  - Cellphone location, home utility data for DHS deportations
  - Geolocation data for the FBI
- Examples: Acxiom, Experian, Spokeo, Cambridge Analytica

> I give you some examples: police officers home addresses, rape sufferers, domestic violence shelters, genetic disease sufferers, among others

2013 Congressional testimony of Pam Dixon (Executive Director, World Privacy Forum)

The New York Times

Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users

Privacy, Security, Safety

Consumer concerns

- Bad faith (by platforms and ad ecosystem stakeholders)
  - Deception and dark patterns
  - Anti-competition / monopoly power
- Reinforcing systemic oppression with new scale and modalities
- Physical well-being: health misinformation, stalking

BBC

Period trackers 'coercing' women into sharing risky information

15 May 2024

Share Save

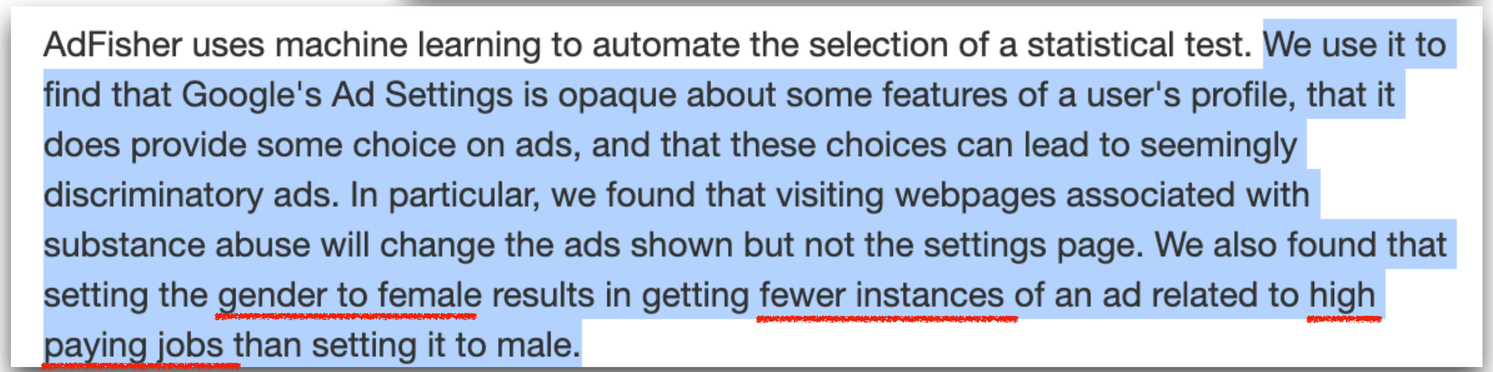
Shiona McCallum & Tom Singleton
Technology reporter

Some privacy experts are concerned that data from menstrual tracking apps could be used to prosecute anyone seeking to terminate a pregnancy.

Amit Datta*, Michael Carl Tschantz, and Anupam Datta

Automated Experiments on Ad Privacy Settings

A Tale of Opacity, Choice, and Discrimination

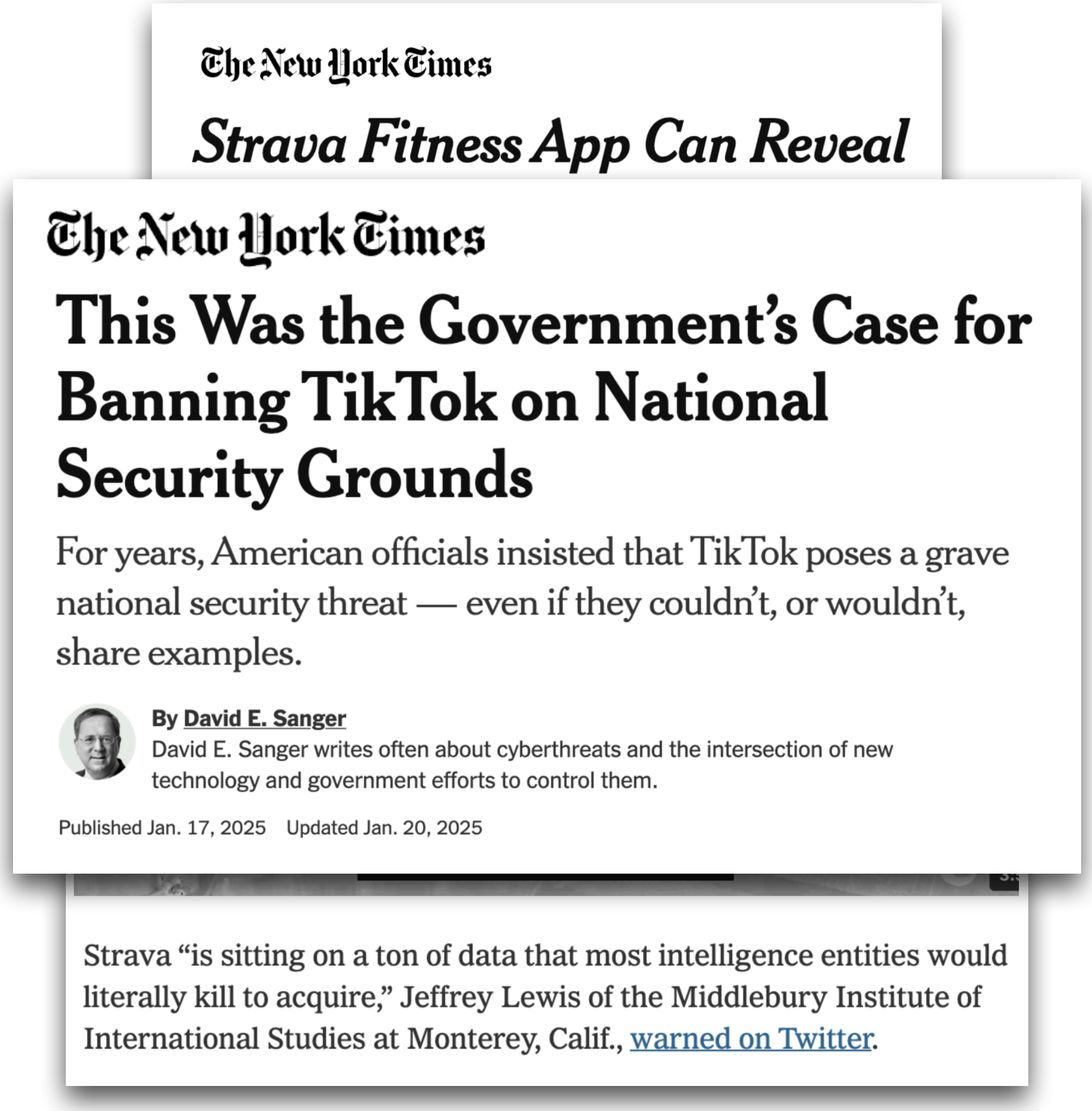


AdFisher uses machine learning to automate the selection of a statistical test. We use it to find that Google's Ad Settings is opaque about some features of a user's profile, that it does provide some choice on ads, and that these choices can lead to seemingly discriminatory ads. In particular, we found that visiting webpages associated with substance abuse will change the ads shown but not the settings page. We also found that setting the gender to female results in getting fewer instances of an ad related to high paying jobs than setting it to male.

Privacy, Security, Safety

Geopolitical concerns

- Nation-state conflicts
  - Operational security (**OPSEC**): extortion of political officials, espionage
  - Information operations (**IO**): disinformation and harassment campaigns
- Government propaganda and surveillance



The New York Times

Strava Fitness App Can Reveal

The New York Times

This Was the Government's Case for Banning TikTok on National Security Grounds

For years, American officials insisted that TikTok poses a grave national security threat — even if they couldn't, or wouldn't, share examples.

By David E. Sanger

David E. Sanger writes often about cyberthreats and the intersection of new technology and government efforts to control them.

Published Jan. 17, 2025 Updated Jan. 20, 2025

Strava "is sitting on a ton of data that most intelligence entities would literally kill to acquire," Jeffrey Lewis of the Middlebury Institute of International Studies at Monterey, Calif., warned on Twitter.

Privacy, Security, Safety

UC San Diego

SCHOOL OF GLOBAL POLICY & STRATEGY

GPS: Is there reason for the U.S. to be legitimately concerned about how TikTok might make users' personal data available to the Chinese government?

VS: I don't buy the argument that we should be alarmed how the Chinese government can get all this data on users from TikTok — because all the other internet platforms active in the U.S. already sell user data to data brokers, and China can obtain it that way. To be clear, I'm sure TikTok is feeding some data to the Chinese government; I'm just saying that it's probably not that much worse than what they could do already using all the other sources of data.

Policy & Regulation

Disclaimer: I am not a lawyer ☝🤓

General Data Protection Regulation (GDPR)

EU setting the standard

"Everyone has the **right to respect for his private** and family life, his home and his correspondence."

- 7 data protection principles — by *design* and by *default*. Of note:
  - Lawfulness, fairness, and transparency
  - Purpose limitation, data minimization
  - Accountability

Consent: freely given, specific, informed, and unambiguous

General Data Protection Regulation (GDPR)

(No) Effect on Third Parties…

Cookie Banners and Privacy Policies: Measuring the Impact of the GDPR on the Web

MICHAEL KRETSCHMER, JAN PENNEKAMP, and KLAUS WEHRLE,
RWTH Aachen University, Germany

[TWEB '21]

We present an extensive summary of key findings based on scientific research regarding the impact of the GDPR. We learn that although the GDPR can directly be linked to an increase in cookie consent notices close to 40% and the volume of privacy policies of around 60%, there is no strong evidence that the GDPR has led to a decrease in online tracking.

Can I Opt Out Yet?
GDPR and the Global Illusion of Cookie Control

Iskander Sanchez-Rola, University of Deusto, Symantec Research Labs
Matteo Dell'Amico, Symantec Research Labs
Platon Kotzias, IMDEA Software Institute, Univ. Politécnica de Madrid
Davide Balzarotti, EURECOM
Leyla Bilge, Symantec Research Labs
Pierre-Antoine Vervier, Symantec Research Labs
Igor Santos, University of Deusto

[Asia CCS '19]

…oth the information presented to users and the …nplemented through cookies; we find that the GDPR has impacted website behavior in a truly global way, both directly and indirectly: USA-based websites behave similarly to EU-based ones, while third-party opt-out services reduce the amount of tracking even for websites which do not put any effort in respecting the new law. On the other hand, we find that tracking remains ubiquitous. In particular, we found cookies that can identify users when visiting more than 90% of the websites in our dataset—and we also encountered a large number of websites that present deceiving information, making it it very difficult, if at all possible, for users to avoid being tracked.

Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites

Jannick Sørensen, Aalborg University Copenhagen, Copenhagen, Denmark, js@cmi.aau.dk
Sokol Kosta, Aalborg University Copenhagen, Copenhagen, Denmark, sok@cmi.aau.dk

[WWW '18]

Our contributions: Having collected HTTP requests and responses 21 times over eight months for a strategic selection of 1250 European and US websites, we can analyze the fluctuations in number of TPs before and after the commencement of GDPR on May 25, 2018. Thereby, we provide empirical evidence for the discussion whether GDPR would lead to fewer TPs. At a general level, we can conclude that the amount of TPs on web pages have slightly declined, but the picture is more complex and contradictory when we study the developments for respectively categories of sites and TPs. This finding adds details to findings provided by [21] and [10]. We cannot support the general assumption that the GDPR has led to fewer TPs, since we cannot find strong evidence for any correlation. As a second contribution, we characterize the differences between websites offered by public and private organizations from the TP perspective, initiating a discussion on public organizations' use of TPs. Finally, we intend to publish the data to allow other research groups further analysis.

Cookie Banners

Illusions of control and the death of usability

Do Cookie Banners Respect my Choice?
Measuring Legal Compliance of Banners from
IAB Europe's Transparency and Consent Framework

Célestin Matte
Université Côte d'Azur, Inria
France
celestin.matte@inria.fr

Nataliia Bielova
Université Côte d'Azur, Inria
France
nataliia.bielova@inria.fr

Cristiana Santos
Research Centre for Justice and Governance
School of Law, University of Minho
cristianasantos@protonmail.com

Cookie Banners and Privacy Policies: Measuring the Impact of the GDPR on the Web

MICHAEL KRETSCHMER, JAN PENNEKAMP, and KLAUS WEHRLE,
RWTH Aachen University, Germany

Can I Opt Out Yet?
GDPR and the Global Illusion of Cookie Control

Iskander Sanchez-Rola
University of Deusto
Symantec Research Labs

Matteo Dell'Amico
Symantec Research Labs

Platon Kotzias
IMDEA Software Institute
Univ. Politécnica de Madrid

Davide Balzarotti
EURECOM

Leyla Bilge
Symantec Research Labs

Pierre-Antoine Vervier
Symantec Research Labs

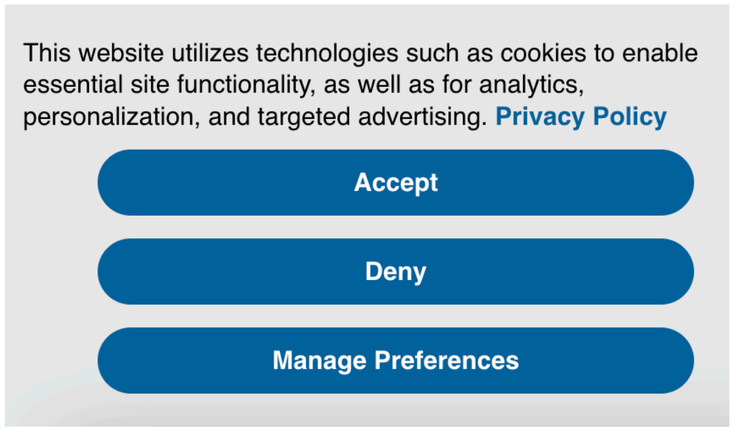
Igor Santos
University of Deusto

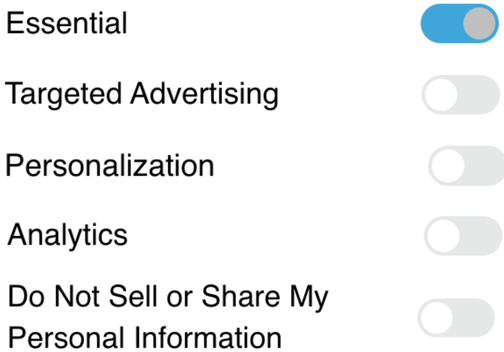
No cookie consent walls — and no, scrolling isn't consent, says EU data protection body

Natasha Lomas – 4:41 AM PDT · May 6, 2020

TechCrunch

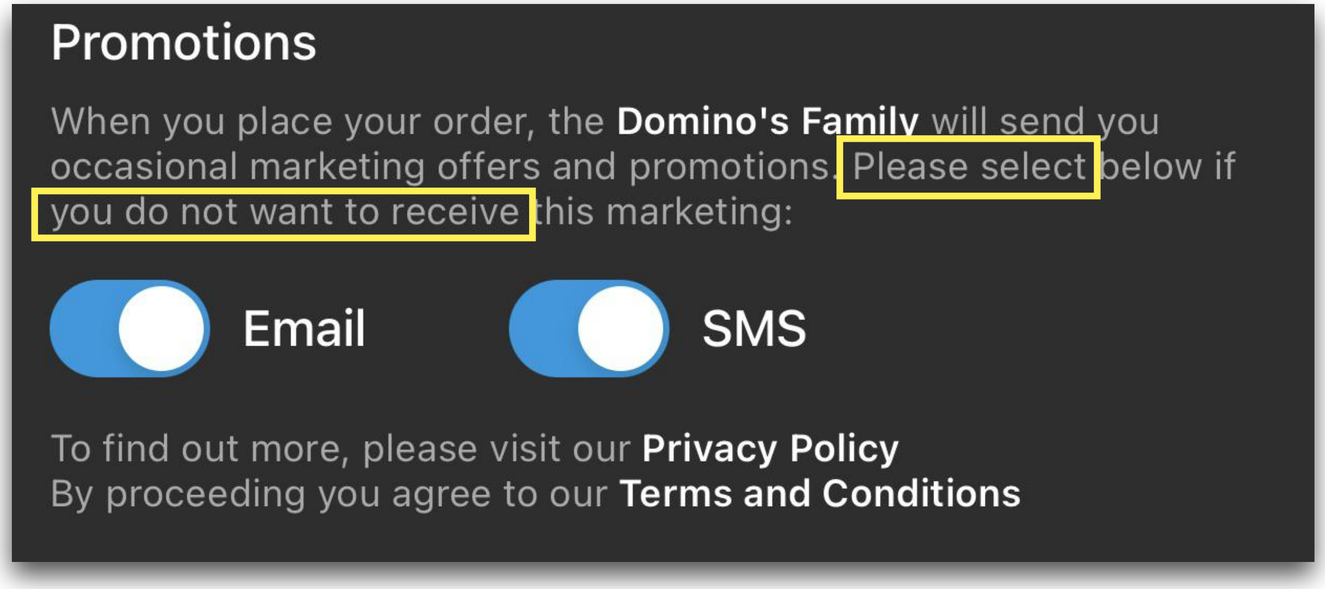
IEEE Xplore®

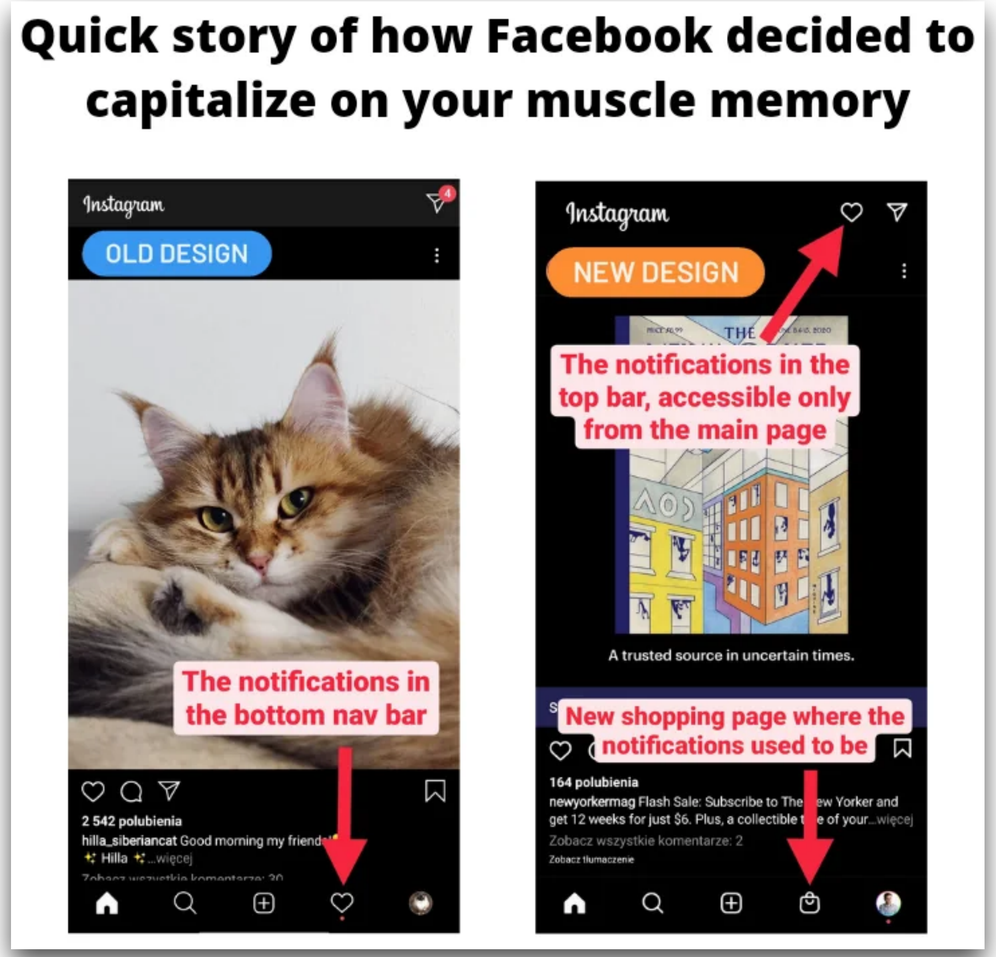




Dark patterns

- Deceptive design patterns to trick users
- Not so GDPR-compliant 😐





https://www.reddit.com/r/assholedesign/
https://www.deceptive.design/hall-of-shame

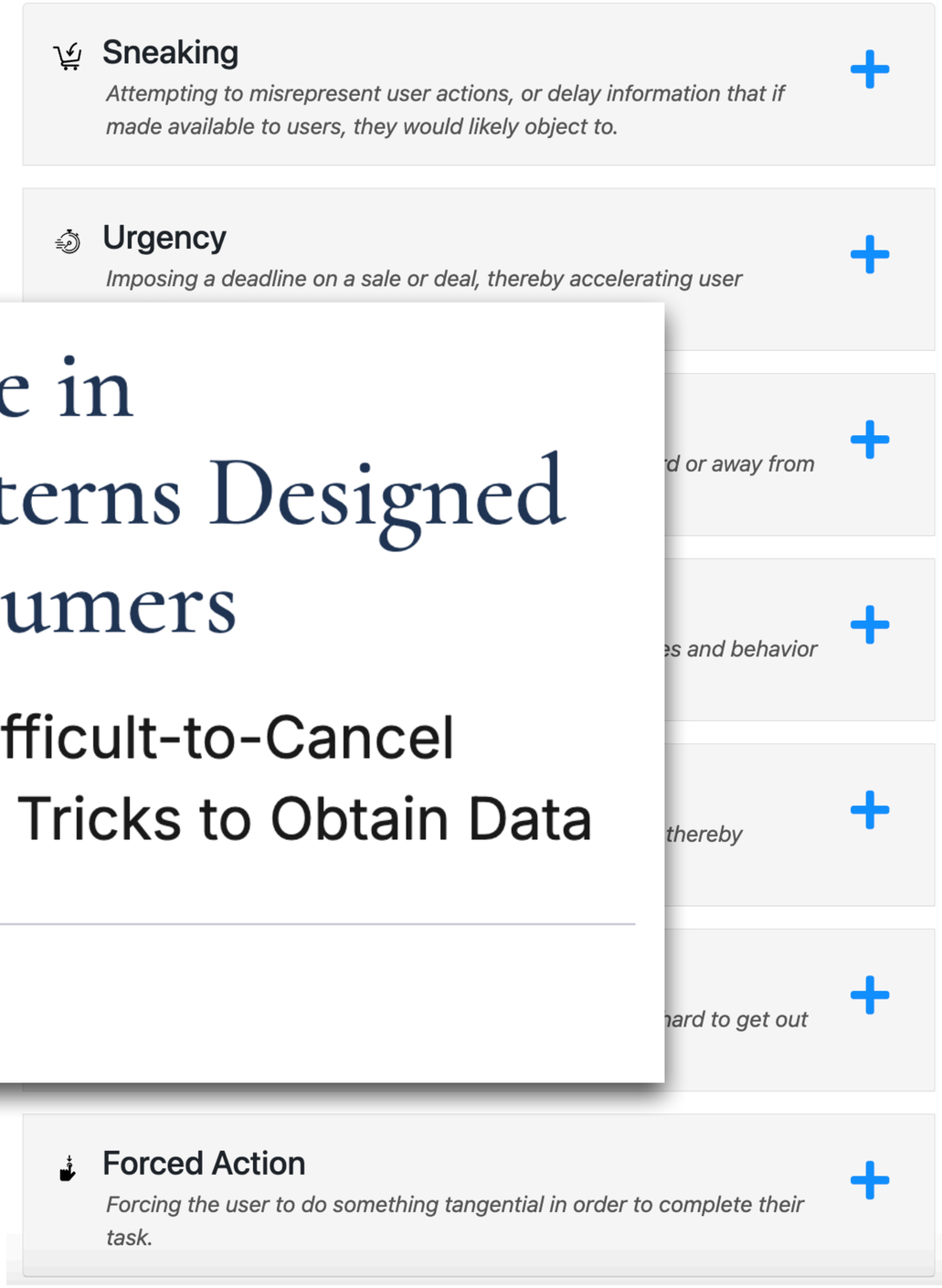
Dark patterns

Dark Patte
Shopping
ARUNESH MA
GUNES ACAR,
MICHAEL J. FR
ELENA LUCHE
JONATHAN M
MARSHINI CH
ARVIND NARA

FTC Report Shows Rise in Sophisticated Dark Patterns Designed to Trick and Trap Consumers

Tactics Include Disguised Ads, Difficult-to-Cancel Subscriptions, Buried Terms, and Tricks to Obtain Data

September 15, 2022


Sneaking
Attempting to misrepresent user actions, or delay information that if made available to users, they would likely object to.

Urgency
Imposing a deadline on a sale or deal, thereby accelerating user

d or away from

es and behavior

thereby

hard to get out

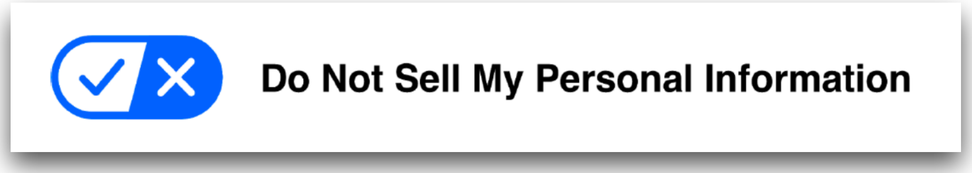
Forced Action
Forcing the user to do something tangential in order to complete their task.

- 1.8K ins
  shoppin
- ⬆️ site p
- Dark pat
- Taxonomy of 15 types of dark patterns

More policy as a treat 😋

California Consumer Privacy Act (CCPA)

Opting out of data brokering



- No federal data privacy law in the US
- CCPA: protecting personal data of California residents via **opt-out**, applying to **for-profit** businesses handling their data
  - Enacted 2018, effective 2020
  - California Privacy Rights Act (CPRA) amends CCPA: better align with GDPR
- **20 states** have passed comprehensive data privacy laws

https://www.whitecase.com/insight-our-thinking/us-data-privacy-guide

DoNotTrack (DNT)

Lessons learned

- DNT: non-standard HTTP header
  - “Please don’t track me 🥺”
  - Deprecated 🪦
- Why did it fail?
  - No teeth for legal authority
- Ironically, a potential data point for fingerprinting 🤡

Stanford University

Do Not Track [edit]

In mid-2010, Mayer and another Stanford researcher Arvind Narayanan argued for Do Not Track in HTTP headers.[12][13] They built Do Not Track prototypes for clients and servers.[14] Working with Mozilla, they wrote the influential Internet Engineering Task Force Internet Draft of Do Not Track.[15][16]

GIZMODO

‘Do Not Track,’ the Privacy Tool Used by Millions of People, Doesn’t Do Anything

By **Kashmir Hill**

Published October 15, 2018 | Comments (0) |

“It is, in many respects, a failed experiment,” said Jonathan Mayer, an assistant computer science professor at Princeton University. “There’s a question of whether it’s time to declare failure, move on, and withdraw the feature from web browsers.”

That’s a big deal coming from Mayer: He spent four years of his life helping to bring Do Not Track into existence in the first place.

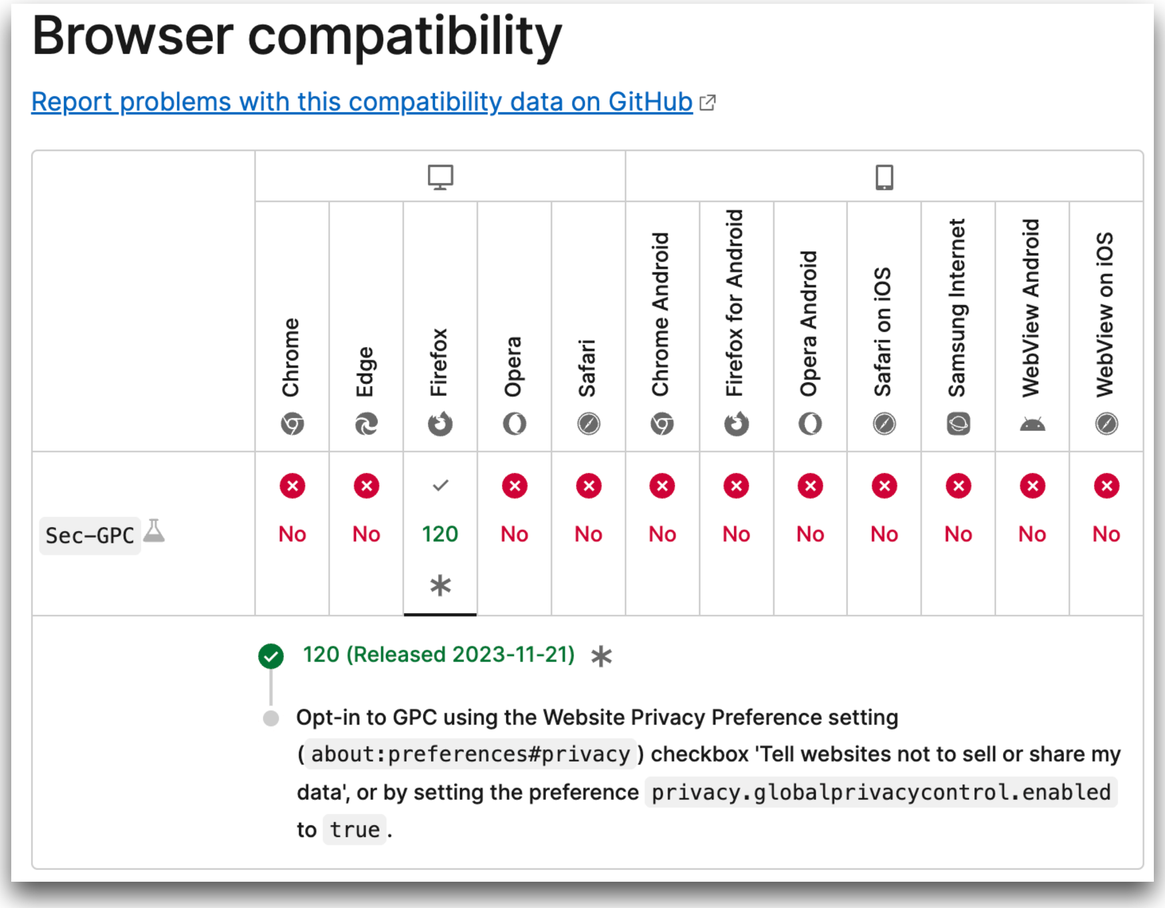
DNT ➡️ Global Privacy Control (GPC)

Opt-outs enabled by CCPA

- Like DNT, but this time, given teeth by CCPA
  - Another HTTP request header
  - Updated Jan 2025

Abstract

This document defines a signal, transmitted over HTTP and through the DOM, that conveys a person's request to websites and services to not sell or share their personal information with third parties. This standard is intended to work with existing and upcoming legal frameworks that render such requests enforceable.



Browser compatibility

Report problems with this compatibility data on GitHub

	Chrome	Edge	Firefox	Opera	Safari	Chrome Android	Firefox for Android	Opera Android	Safari on iOS	Samsung Internet	WebView Android	WebView on iOS
Sec-GPC	No	No	120 *	No	No	No	No	No	No	No	No	No

120 (Released 2023-11-21) *

Opt-in to GPC using the Website Privacy Preference setting (about:preferences#privacy) checkbox 'Tell websites not to sell or share my data', or by setting the preference privacy.globalprivacycontrol.enabled to true .

https://w3c.github.io/gpc/ https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/Sec-GPC#browser_compatibility