ePrivo.eu
An online privacy observatory

Pere Barlet & Ismael Castell
Universitat Politècnica de Catalunya (UPC)

ESNOG 30 – 26 Oct. 2023

Third-party trackers

GOBIERNO DE COALICION >

El PSOE y Sumar alcanzan un acuerdo de Gobierno que incluye la semana laboral de 37,5 horas

El pacto para la legislatura contempla también los impuestos a la banca y las grandes energéticas, el aumento del parque público de vivienda y la ampliación de permisos retribuidos por nacimiento

Pedro Sánchez y Yolanda Díaz, durante la firma del acuerdo, este martes.
Foto: SAMUEL SÁNCHEZ | Vídeo: EPV

PAULA CHOUZA | CARLOS E. CUÉ | JOSÉ MARCOS

Madrid - 24 oct 2023 - 08:43 | Actualizado: 24 OCT 2023 - 21:50 CEST

274

La coalición da un paso decisivo para la investidura de Pedro Sánchez, a la espera de cerrar un pacto con los demás aliados de la mayoría. El PSOE y Sumar han sellado su acuerdo de

POLITICA

La intrahistoria del pacto PSOE-Sumar: "matices de matices" hasta la madrugada y sin concreciones para otros pactos "en cascada"

PREMIUM

Sánchez y Díaz sellan un acuerdo sin mención a la amnistía ni a Cataluña

RAÚL PIÑA
@_raulpina_
Madrid

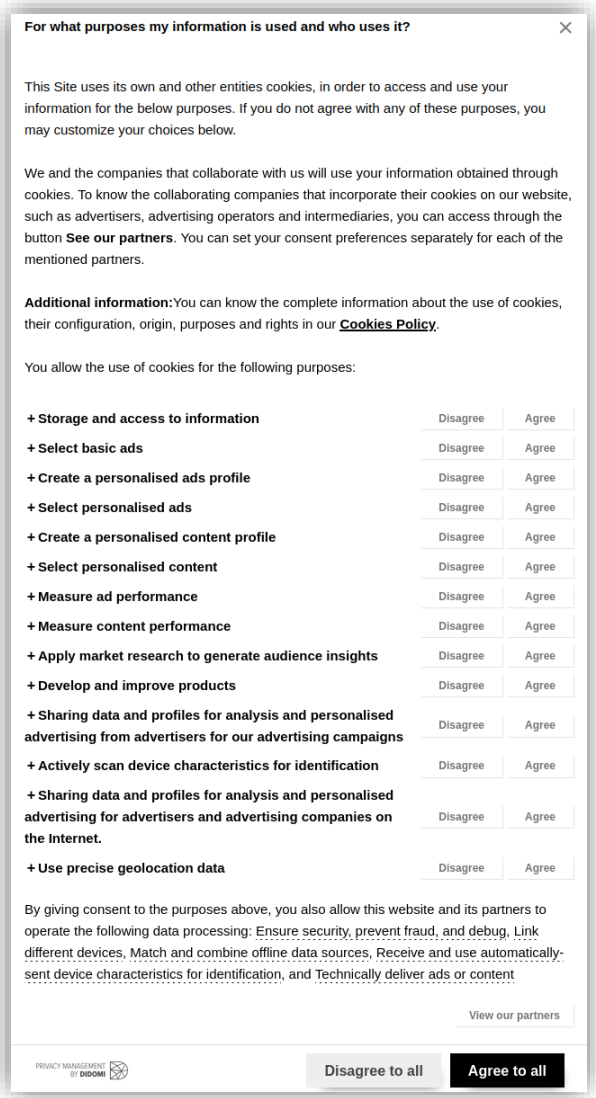
Actualizado Miércoles, 25 octubre 2023 - **09:26**

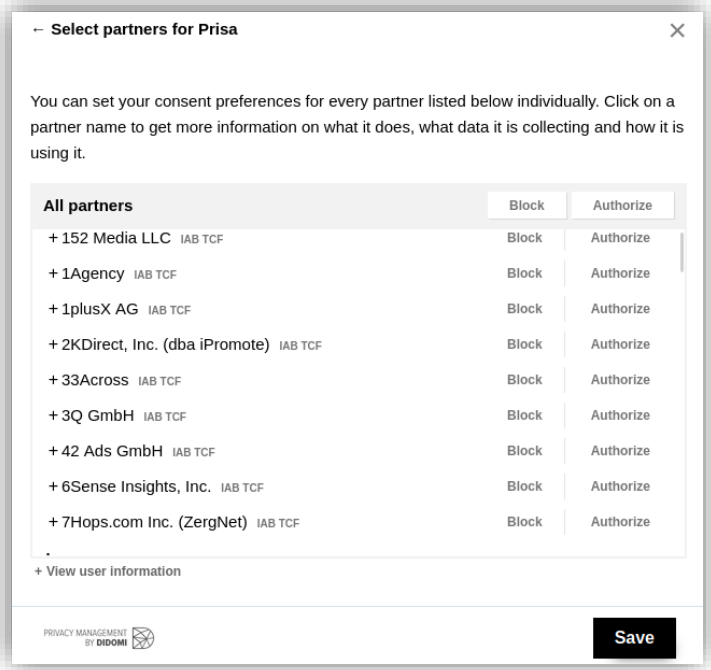
Pedro Sánchez y Yolanda Díaz, en la presentación del pacto de Gobierno PSOE-Sumar. ALBERTO DI LOLLI

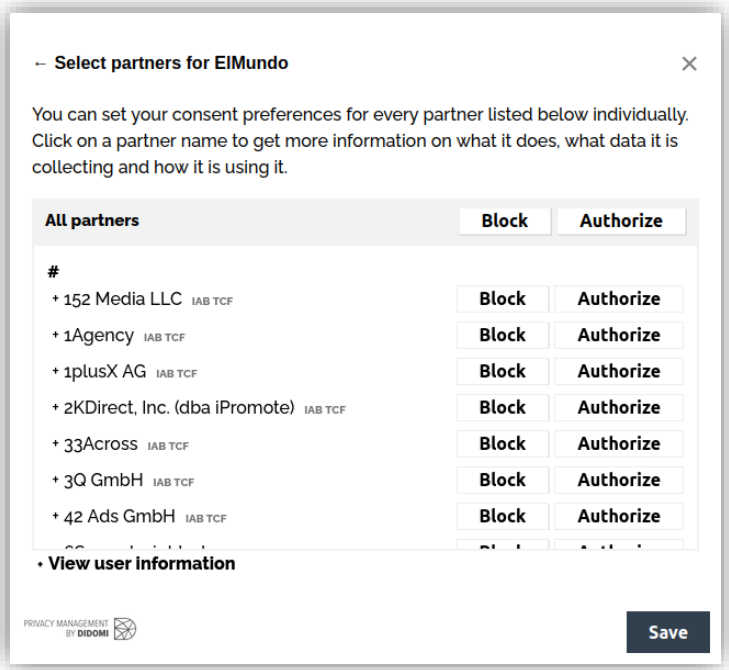
Coalición PSOE y Sumar

Suele decirse que lo que no se nombra no existe. Quizás sea

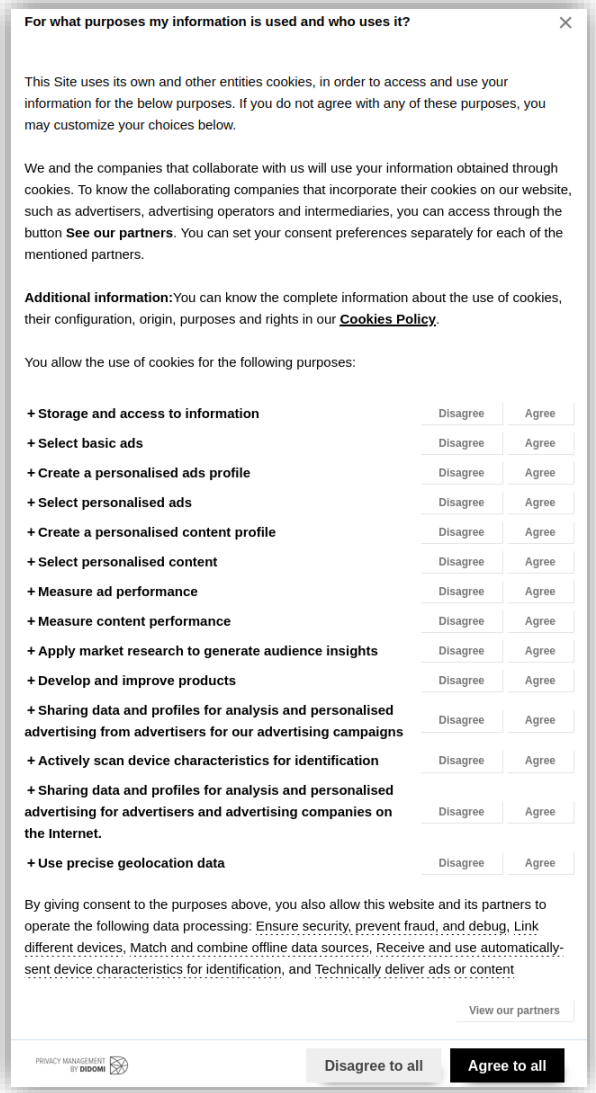
Third-party trackers

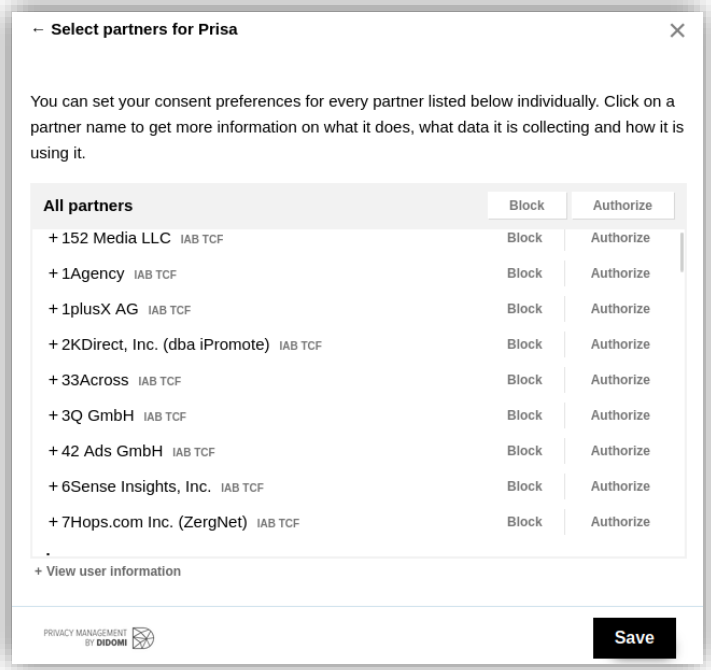


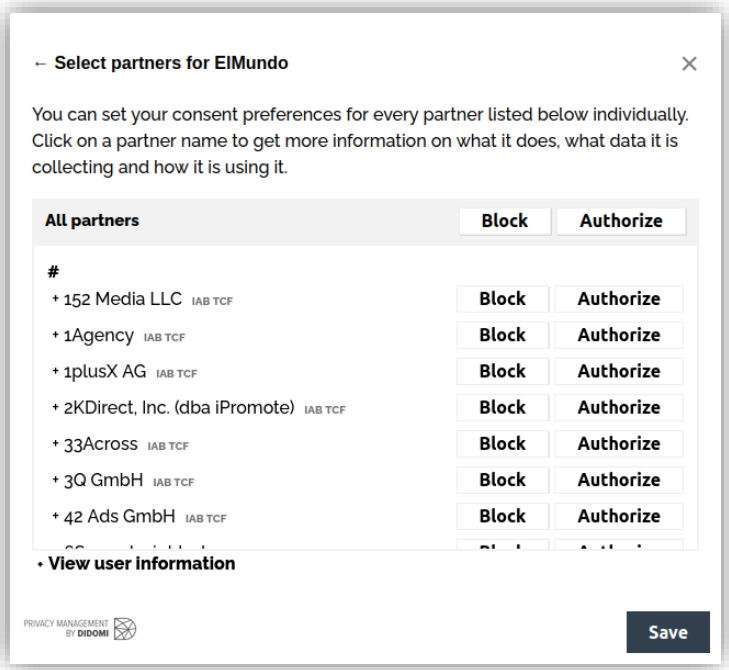




Third-party trackers







1033 “partners”

1128 “partners”

“Partners” (starting with A: 150+)

+A Million AdsIAB TCF

+A.MobIAB TCF

+AA INTERNET-MEDIA LtdIAB TCF

+Aarki, Inc.IAB TCF

+AAX LLCIAB TCF

+Accorp Sp. z o.o.IAB TCF

+AccountInsight LtdIAB TCF

+Active Agent (Virtual Minds GmbH)IAB TCF

+ACTV8, Inc.IAB TCF

+AcuityAds Inc.IAB TCF

+AcxiomIAB TCF

+Ad Alliance GmbHIAB TCF

+ad6mediaIAB TCF

+Adacado Technologies Inc. (DBA Adacado)IAB TCF

+adality GmbHIAB TCF

+Adasta Media S.r.l.IAB TCF

+adbalancer Werbeagentur GmbHIAB TCF

+adbility media GmbHIAB TCF

+AdClear GmbHIAB TCF

+AdColony, Inc.IAB TCF

+AddApptr GmbHIAB TCF

+AdDefend GmbHIAB TCF

+AddrelevanceIAB TCF

+Adelaide Metrics IncIAB TCF

+AdElement Media Solutions Pvt LtdIAB TCF

+Aderize, Inc.IAB TCF

+Adevinta Spain S.L.U.IAB TCF

+Adex (Virtual Minds GmbH)IAB TCF

+ADFI MEDIA SLIAB TCF

+Adform A/SIAB TCF

+AdGear Technologies, Inc.IAB TCF

+AdheseIAB TCF

+adhood.comIAB TCF

+AdikteevIAB TCF

+AdInMo LTDIAB TCF

+ADITION (Virtual Minds GmbH)IAB TCF

+Adjust Digital A/SIAB TCF

+Adjust GmbHIAB TCF

+Adkernel LLCIAB TCF

+Adlane LTDIAB TCF

+Adloox SAIAB TCF

+Adludio Ltd.IAB TCF

+ADMAN - Phaistos Networks, S.A.IAB TCF

+ADman Interactive SLUIAB TCF

+adMarketplace, Inc.IAB TCF

+ADMAXIAB TCF

+AdMaxim LimitedIAB TCF

+admetrics GmbHIAB TCF

+Admixer EU GmbHIAB TCF

+Admo.tv (Clickon)IAB TCF

+Adnami ApsIAB TCF

+adnanny.com SLUIAB TCF

+ADNOW I RENODO MEDIA LTDIAB TCF

+Adnuntius ASIAB TCF

+Adobe Advertising CloudIAB TCF

+Adobe Audience Manager, Adobe Experience PlatformIAB TCF

+Adobe Audience Manager, Adobe Experience Platform

+AdomikIAB TCF

+Adpone SLIAB TCF

+adQueryIAB TCF

+AdQuiver Media SLIAB TCF

+ADRENALEADIAB TCF

+Adrino Mobile Sp. z o.o.IAB TCF

+adrule mobile GmbHIAB TCF

+Ads Interactive Ltd.IAB TCF

+Adserve.zone / Artworx ASIAB TCF

+AdServing Factory srlIAB TCF

+Adsmovil España S.LIAB TCF

+ADSOCYIAB TCF

+Adsolutions BVIAB TCF

+AdsonicaIAB TCF

+AdSpirit GmbHIAB TCF

+adsquare GmbHIAB TCF

+ADSTOURS SASIAB TCF

+AdsWizz Inc.IAB TCF

+Adsyield IncIAB TCF

+Adtarget Teknoloji A.S.IAB TCF

+ADTARGET.ME UABIAB TCF

+Adtelligent Inc.IAB TCF

+AdTheorent, IncIAB TCF

+AdTiming Technology Company LimitedIAB TCF

+Adtriba GmbHIAB TCF

+ADTTRIBUTION IncIAB TCF

+ADUXIAB TCF

+advanced store GmbHIAB TCF

+ADventori SASIAB TCF

+Adventure MediaIAB TCF

+AdverlineIAB TCF

+Adverticum cPlc.IAB TCF

+Advertising.Tech FZ-LLCIAB TCF

+Adverty AB (publ)IAB TCF

+AdViewIAB TCF

+Advisible ABIAB TCF

+Adwatch

+ADWAYS SASIAB TCF

+adWMGIAB TCF

+Adxperience SASIAB TCF

+ADYOULIKE SAIAB TCF

+Adzymic Pte LtdIAB TCF

+Agate Systems LimitedIAB TCF

+agof - daily campaign factsIAB TCF

+agof studiesIAB TCF

+AiCreo S.r.l.IAB TCF

+AIDEM Technologies LTDIAB TCF

+Airnow Media LtdIAB TCF

+Akamai

+AlgoriX Technology Pte. Ltd.,IAB TCF

+AlkimiIAB TCF

+Allegro sp. z o.o.IAB TCF

+Alliance Gravity Data MediaIAB TCF

+Alphonso Inc. IAB TCF

+Always-Computing SASIAB TCF

+Amazon Ad ServerIAB TCF

+Amazon AdvertisingIAB TCF

+Amnet GmbHIAB TCF

+Amobee Inc.IAB TCF

+Amplified IntelligenceTechnologiesIAB TCF

+ANINPRO-CREATIVE, S.L.IAB TCF

+Aniview LTDIAB TCF

+AnonymisedIAB TCF

+Anteriad, LLCIAB TCF

+AntVoiceIAB TCF

+Anzu Virtual Reality LTDIAB TCF

+Apester LtdIAB TCF

+Appier PTE LtdIAB TCF

+Appstock LTD.IAB TCF

+appTV Ltd.IAB TCF

+Appush TechnologiesIAB TCF

+Aptivio IncIAB TCF

+AragoIAB TCF

+Arbeitsgemeinschaft Media-AnalyseIAB TCF

+Arcanor Bilgi Teknolojileri ve Hizmetleri A.Ş.IAB TCF

+ArcSpan Technologies, Inc.IAB TCF

+Arcspire LimitedIAB TCF

+AREIA (ADLOOP)IAB TCF

+ArkeeroIAB TCF

+ARKHEUSIAB TCF

+ARMIS SASIAB TCF

+Arpeely Ltd.IAB TCF

+Arrivalist Co.IAB TCF

+ArtChaos s.r.o.IAB TCF

+Artefact Deutschland GmbHIAB TCF

+Aryel s.r.lIAB TCF

+Ask LocalaIAB TCF

+AttributyIAB TCF

+Audience NetworkIAB TCF

+Audience Solutions S.A.IAB TCF

+Audience Trading Platform Ltd.IAB TCF

+AudienceProject ApsIAB TCF

+AudienceProject Measurement AmzIAB TCF

+Audiencerate LTDIAB TCF

+AudienceRun corpIAB TCF

+Audiens S.r.l.IAB TCF

+Audienzz AGIAB TCF

+AuDigentIAB TCF

+audio content & control GmbHIAB TCF

+Audiohook CorporationIAB TCF

+AUDIOMOB LTDIAB TCF

+AudionIAB TCF

+Automattic AdsIAB TCF

+Avantis Video LtdIAB TCF

+Avocet Systems LimitedIAB TCF

+AWIN AGIAB TCF

+Axel Springer Teaser Ad GmbHIAB TCF

+Axiom Media ConnectIAB TCF

+AXIS CORP LTDIAB TCF

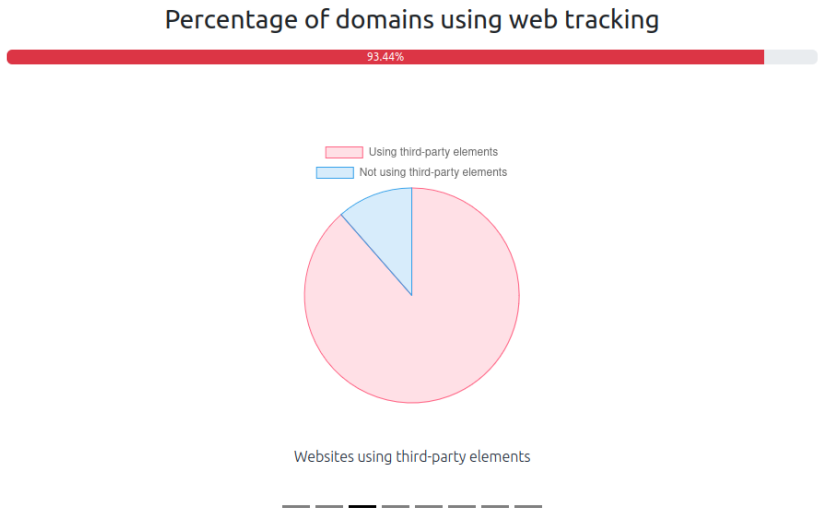
+Axonix LTDIAB TCF

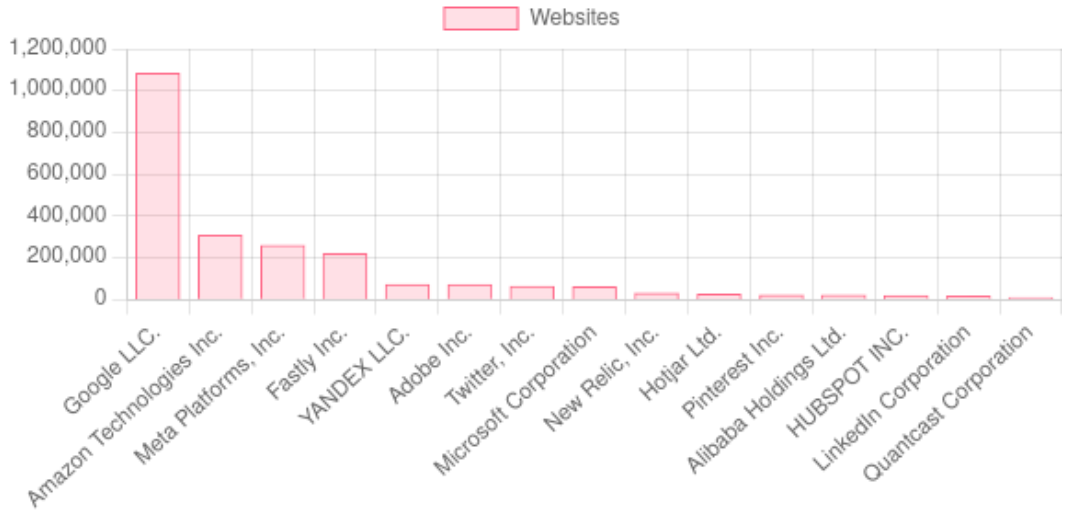
+Azerion Holding B.V.IAB TC

+ >850 more partners!!

Third-party trackers

- Third-party trackers are present in >90% of the websites
- They can easily recover our *browsing history* (at least)
- They can track us even if we delete cookies, use private mode, DNT, VPN, etc.

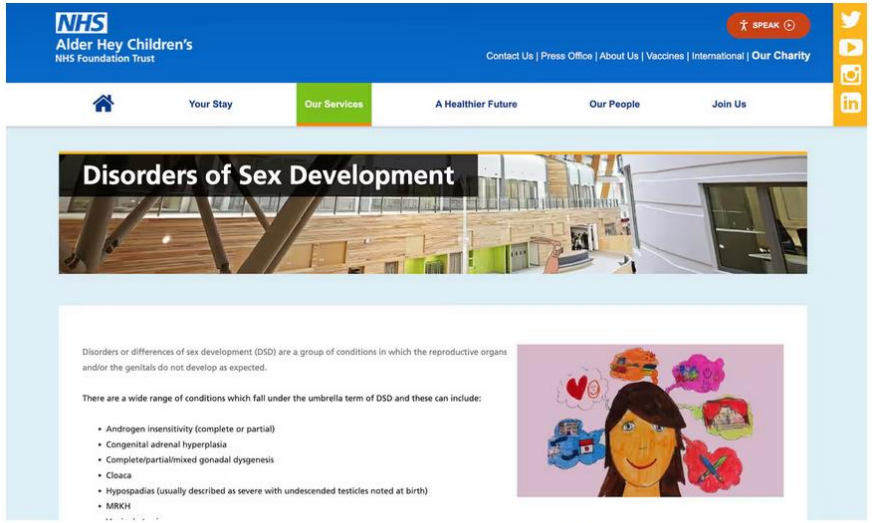




Third-party trackers

NHS data breach: trusts shared patient details with Facebook without consent

The Guardian



A page about sexual development disorders on Alder Hey Children's Hospital's website, which shared details of the browsing with Facebook via the Meta Pixel.

Observer investigation reveals Meta Pixel tool passed on private details of web browsing on medical sites

NHS trusts are sharing intimate details about patients' medical conditions, appointments and treatments with Facebook without consent and despite promising never to do so.

An *Observer* investigation has uncovered a covert tracking tool in the websites of 20 NHS trusts which has for years collected browsing information and shared it with the tech giant in a major breach of privacy.

The data includes granular details of pages viewed, buttons clicked and keywords searched. It is matched to the user's IP address - an identifier linked to an individual or household - and in many cases details of their Facebook account.

Information extracted by Meta Pixel can be used by Facebook's parent company, Meta, for its own business purposes - including improving its targeted advertising services.

Records of information sent to the firm by NHS websites reveal it includes data which - when linked to an individual - could reveal personal medical details.

It was collected from patients who visited hundreds of NHS webpages about HIV, self-harm, gender identity services, sexual health, cancer, children's treatment and more.

It also includes details of when web users clicked buttons to book an appointment, order a repeat prescription, request a referral or to complete an online counselling course. Millions of patients are potentially affected.

Data brokers

- Companies or business units whose primary source of revenue is selling user personal information in data market places [1]

- Dangers for the user
  - Profiles may be (most surely) incorrect
  - There is no easy way to discover and correct it (despite GDPR)
  - There is no easy way to know to whom our info was sold to [2,3]
  - There is no easy way to know how this is impacting us!!

Privacy

The government can't seize your data — but it can buy it

Adam Kovacevich @adamkovac / 2:00 PM UTC • May 21, 2023 Comment

[1] A. Rieke, H. Yu, D. Robinson, and J. van Hoboken, "Data Brokers in an Open Society", Open Society Foundations 2016
[2] https://nrkbeta.no/2020/12/03/my-phone-was-spying-on-me-so-i-tracked-down-the-surveillants/
[3] https://techcrunch.com/2023/05/21/the-government-cant-seize-your-data-but-it-can-buy-it/

Tracking purposes

- Purposes and implications
  - Targeted advertising
  - Price discrimination [4]
  - Insurance coverage [5]
  - Credit rating [6]
  - Phishing/identity theft (e.g. Conti Leaks)
  - Political interests (e.g. Cambridge Analytica)

Issue: September 21, 2021

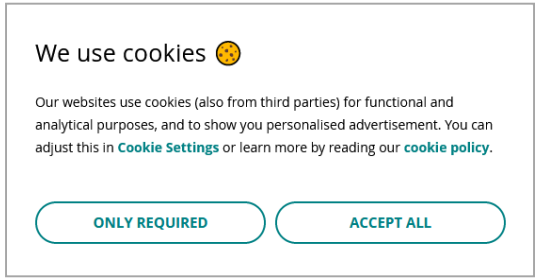
IMF report suggests credit scores could soon be based on web browsing history

Dystopian future.

[4] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris, “Crowd-assisted search for price discrimination in e-commerce: first results,” in Proceedings of the ninth ACM conference on Emerging networking experiments and technologies, ser. CoNEXT ’13
[5] The Economist, “Insurance data: Very personal finance,” 2012. [Online]. Available: http://www.economist.com/node/21556263
[6] Reclaim the Net, [Online] https://reclaimthenet.org/imf-report-suggests-credit-score-browsing-history

Tracking methods

- Stateful methods
  - Cookies
  - HTML5 storage [7]
  - Web cache [8], DNS cache [9], Redirect cache [10]



- Stateless methods (fingerprinting)
  - Combine device properties to form a unique identifier
  - Canvas fingerprinting [11]
  - Audio fingerprinting [12]
  - 3D rendering fingerprinting [13]
  - Mouse fingerprinting [14]

	Device Fingerprinting	OS Fingerprinting	Browser Fingerprinting
IP Address	✓		
Geolocation	✓		
OS version	✓	✓	
Screen resolution	✓	✓	
Timezone	✓		
Fonts	✓	✓	
Browser version	✓	✓	✓
Browser language	✓		✓
Multitouch capabilities	✓		
Pointer capabilities	✓		
Keyboard layout	✓		
Microphone	✓		
Camera	✓		
TCP timestamp	✓		
OS architecture		✓	
System language		✓	
Color depth		✓	
Audio capabilities	✓		
CPU brand	✓		
GPU brand	✓		
System drivers	✓		
Browser instance			✓
Browser plugins			✓
Accepted media formats			✓
Cookies enabled			✓
Supercookies limitations			✓

[7] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz, “The Web Never Forgets: Persistent Tracking Mechanisms in the Wild,” in ACM CCS, 2014.
[8] G. Fleischer, “Implementing Web Tracking,” Proc. Black Hat USA Conf. Briefings, pp. 1–37, Jul. 2002.
[9] A. Klein and B. Pinkas, “DNS Cache-Based User Tracking.” [Online]. Available: https://www.ndss-symposium.org/ndss-paper/dns-cache-based-user-tracking/
[10] E. Bursztein, “Tracking users that block cookies with a http redirect,” Jul. 2011.
[11] K. Mowery and H. Shacham, “Pixel Perfect : Fingerprinting Canvas in HTML 5,” in Proceedings of Web 2.0 Security and Privacy (W2SP), 2012
[12] S. Englehardt and A. Narayanan, “Online Tracking: A 1-million-site Measurement and Analysis,” in ACM CCS, 2016
[13] Y. Cao, S. Li, and E. Wijmans, “(Cross-)Browser Fingerprinting via OS and Hardware Level Features,” NDSS Symposium, 2017
[14] M. Urpí, I. Castell and P. Barlet. “Detecting and Analyzing Mouse Tracking in the Wild,” IEEE EuroS&PW, 2023

Adblockers

- Most popular protection methods
  - Block URLs using pattern lists
  - Difficult to maintain updated and block new methods
  - Easy to evade by periodically changing the hosting domains
  - Adblockers are *too* conservative

- Why are they conservative?
  - “Users want protection that does not break websites” [15]
  - “When extensions break websites, users disable them based on how important the content they are accessing is to them” [16]

- Usually do not block URLs that can impact the website functionality

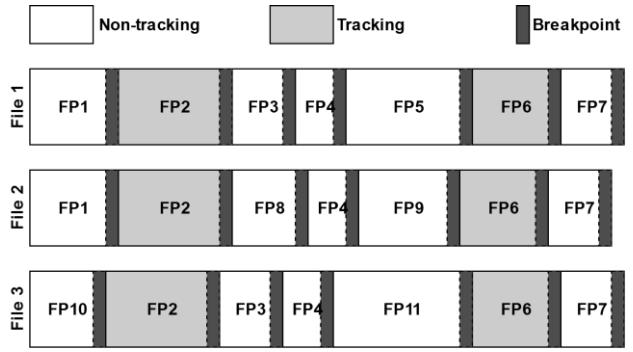
[15] C. E. Wills and D. C. Uzunoglu, “What Ad Blockers Are (and Are Not) Doing,” HotWeb, 2016
[16] G. Merzdovnik, M. Huber, D. Buhov, N. Nikiforakis, S. Neuner, M. Schmiedecker, and E. Weippl, “Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools,” EuroS&P, 2017

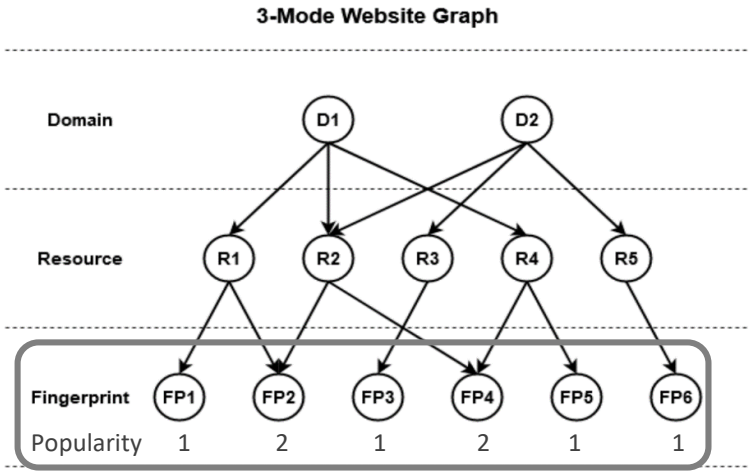
TrackSign

- Objective: Automatically detect *unknown* web trackers
  - Known trackers under unknown URLs
  - Completely new tracking methods
- TrackSign analyzes website code instead of URLs
  - If a piece of website code is mostly present in files which we already know that perform tracking, it is most probably being used for tracking
  - We can automatically label as tracking other files using the same code!
- TrackSign explores two aspects of the code:
  - Popularity: #resources that use this piece of code
  - Dirtiness: %resources known to perform tracking
- The proposal is based on two observations:
  - Limited number of ways to obtain user information
  - New effective tracking methods will become popular at some point

TrackSign

- Challenges
  - How to partition the code?
  - Same file hosted in several servers and linked by many different URLs
- Solution
  - Rabin Fingerprinting for file partitioning
  - A 3-mode network graph (tripartite)
    - Domain, Resource and Signature

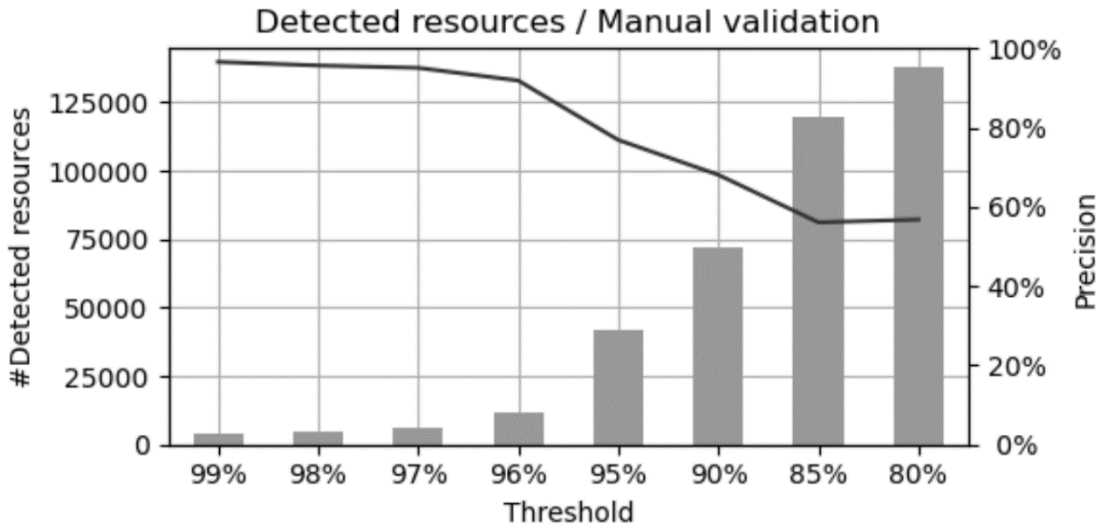




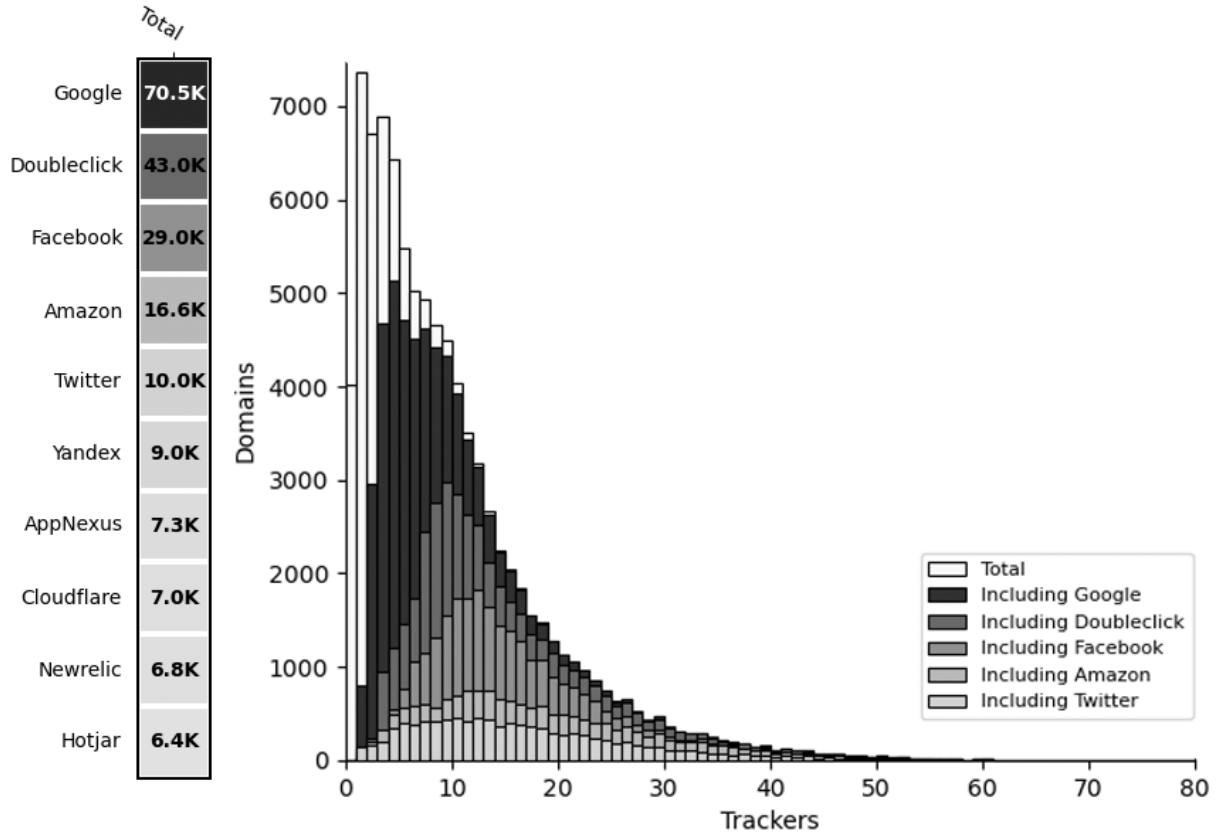
"**TrackSign: Guided Web Tracking Discovery**," *IEEE INFOCOM,* 2021. https://personals.ac.upc.edu/pbarlet/papers/TrackSign.Infocom2021.pdf

TrackSign evaluation

- Top 100.000 most popular websites
  - 5.2M of unique URLs, 933K resources, 73M code signatures
  - Manually validated about 800 files
- 96% threshold maximizes detection while maintaining accuracy (92%)
  - Found more than 12.000 new tracking resources and 300.000 new tracking URLs



TrackSign evaluation

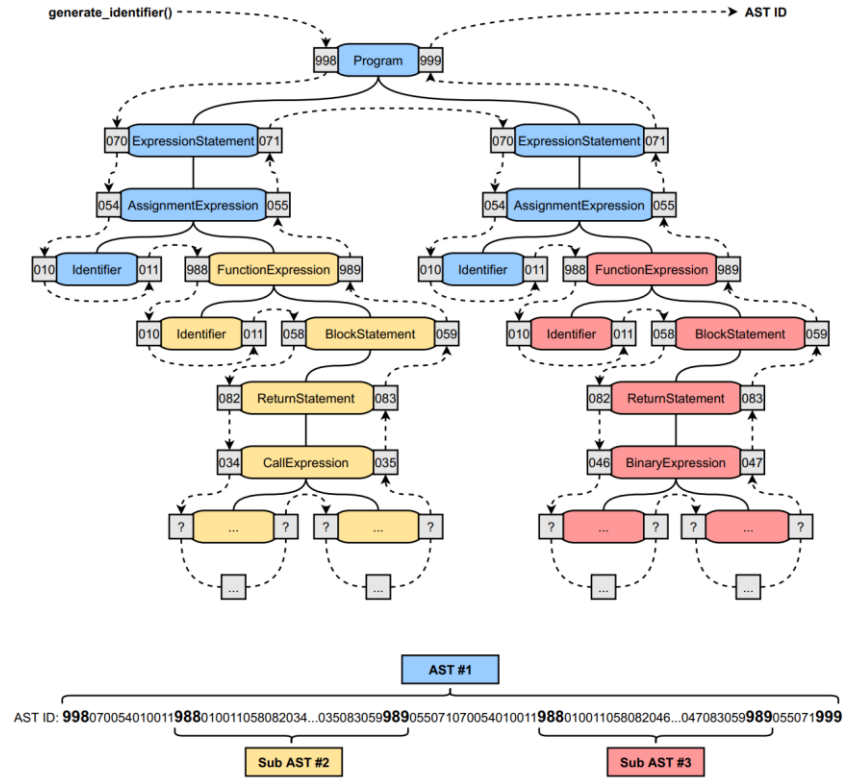


- 95% of websites (86.626 out of 90.637) include web tracking

ASTrack

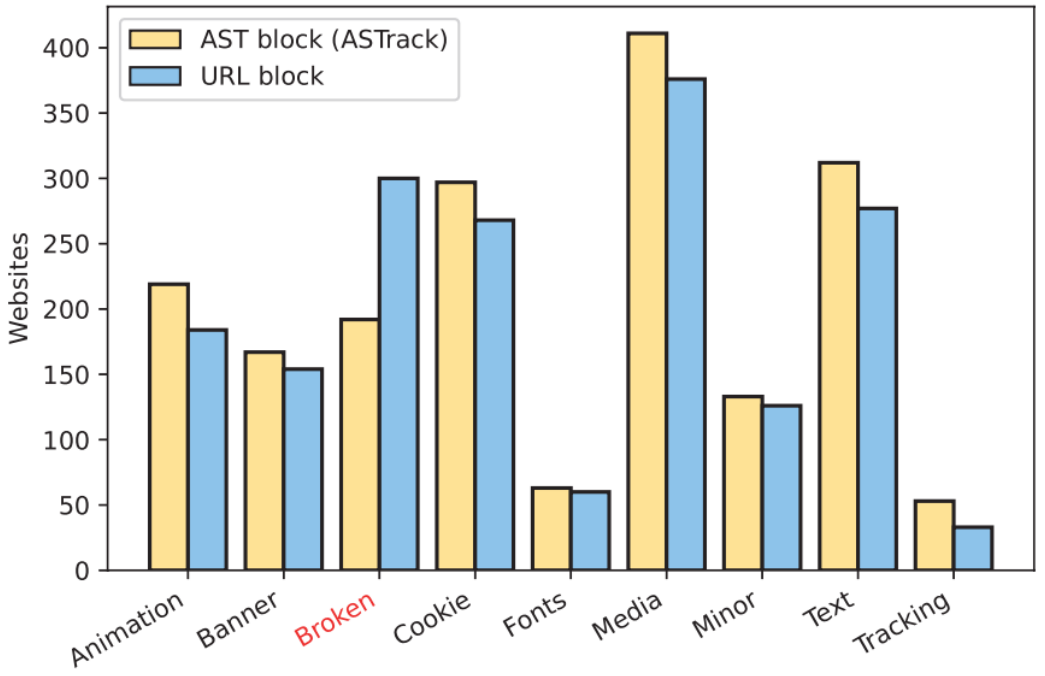
- Analyze code structure instead of shared code
  - We identify shared *functionality* not shared code
  - Robust against obfuscation and minimification
  - We can selectively block tracking AST better preserving functionality!

- ASTrack explores Abstract Syntax Trees (AST)
  - ASTs identified by a simple label chain



ASTrack evaluation

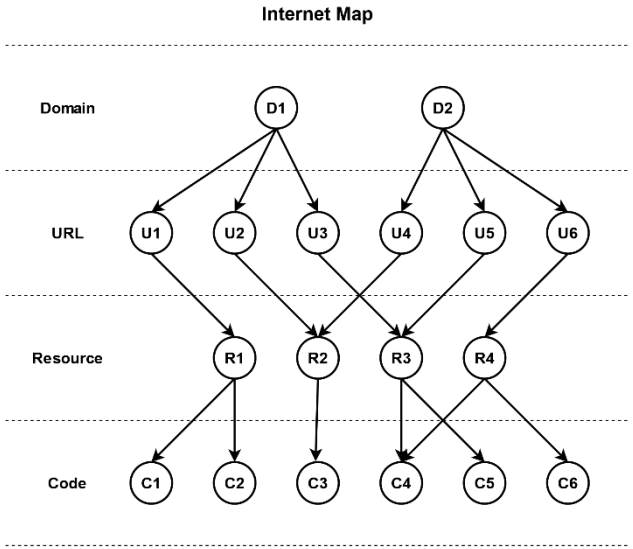
	Static evaluation	Dynamic evaluation
Domains	8,179	
URLs	615,780	
JavaScript URLs	161,593	
Tracking URLs	41,274	
Unique ASTs	7,015,542	
Shared ASTs	2,683,586	
New tracking URLs	3,409	2,183
New tracking JavaScript files	3,109	2,093
Tracking ASTs	49,453	41,114
Precision	98.52%	98.47%



- We block more than 3k URL compared to adblockers, while improving website functionality!

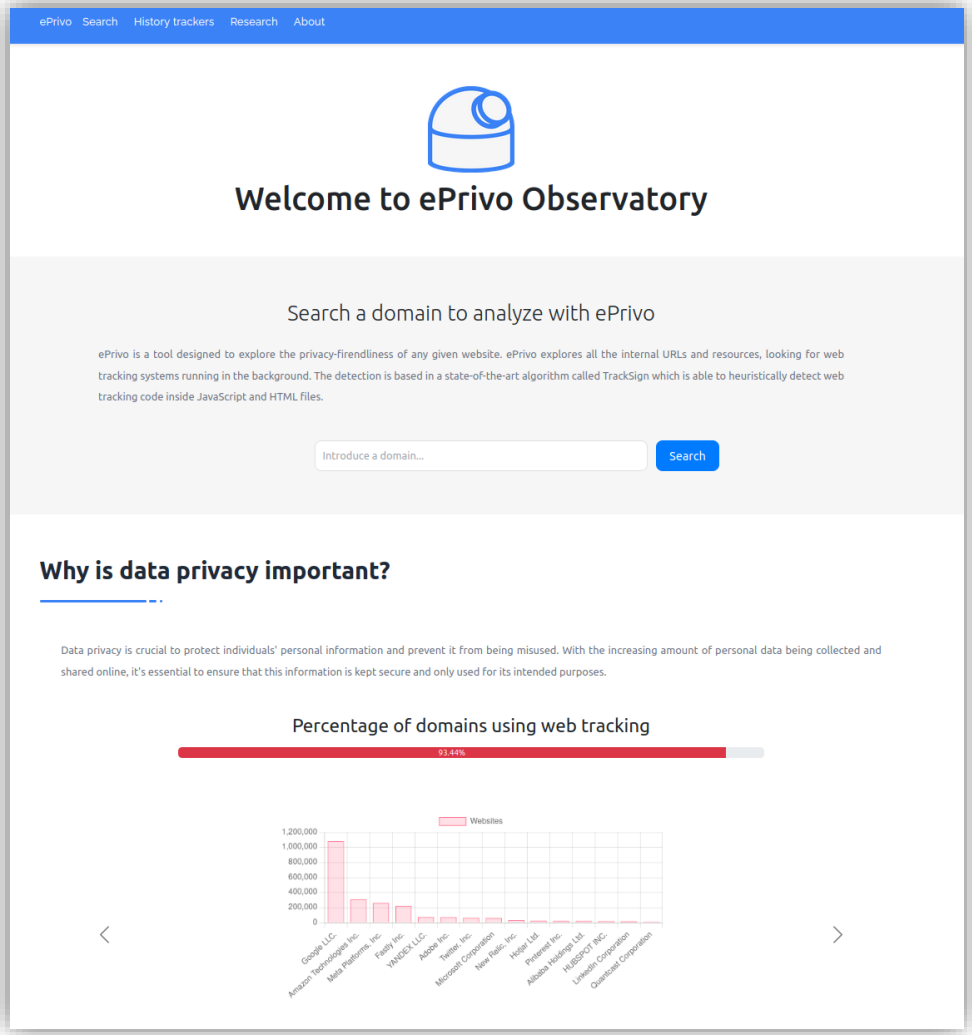
ePrivo.eu

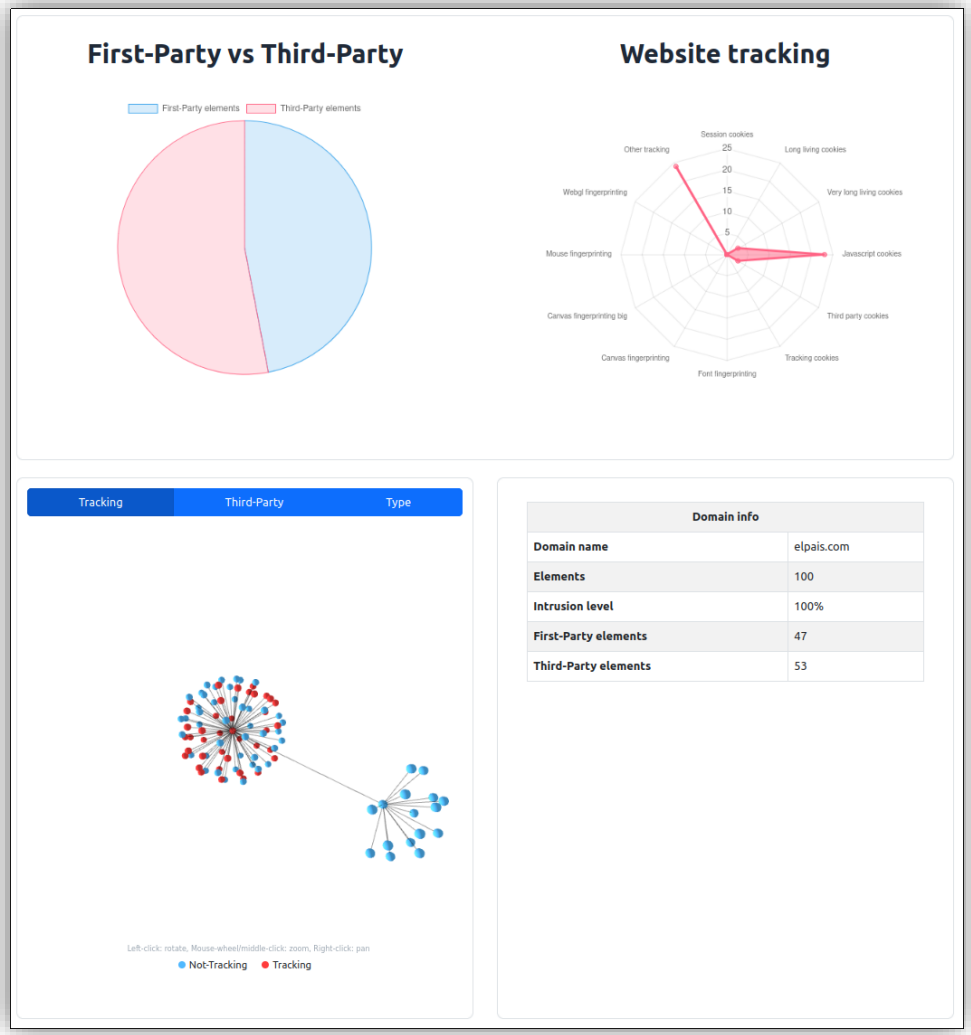
- New online privacy observatory
  - Data collected using TrackSign/ASTrack over 1.5 million websites
- Augmented with a new layer about URL information to obtain privacy information not included inside website code
  - Cookies
  - Certificates
- Also explores prevalence of third-party trackers



"**ePrivo.eu: An Online Service for Automatic Web Tracking Discovery**," *IEEE ACCESS,* 2023. https://upcommons.upc.edu/bitstream/handle/2117/386463/P6.pdf

https://eprivo.eu





Conclusions

- Summary
  - ePrivo.eu can help Internet users realize the huge problem around online privacy
  - ePrivo.eu could also be useful by regulators or even web site owners
  - We can now automatically detect unknown web tracking and discover new methods
  - We can block more than adblockers by keeping website functionality in most cases
- Future work
  - Bring ePrivo.eu into a production state
  - Develop a browser extension that implements TrackSign/ASTrack
- What do we need?
  - It has proven very hard to find funding for this project…
  - We could only develop it as a side project ☹

ePrivo.eu
An online privacy observatory

Pere Barlet & Ismael Castell
{pere.barlet, ismael.castell}@upc.edu
Universitat Politècnica de Catalunya (UPC)

ESNOG 30 – 26 Oct. 2023

Backup slides

Privacy regulations

	GDPR	CCPA	PIPL
Year	2018	2020	2021
Under protection	People living in Europe	People born in California	People living in China
Target	For-profit and non-profit national or private	For-profit companies (with conditions)	For-profit and non-profit national or private
Area of effect	World	State of California	World
Restricted collection scenarios	Yes (6)	No	Yes (7)
Required consent	Yes	No	Yes
Forces service without consent	Yes	No	Yes
Opt-out available	Yes	Partially (data selling processes)	Yes
Access right	Yes	Yes	Yes
Amendment right	Yes	Yes	Yes
Deletion right	Yes	Partially	Yes
Right to be forgotten	Yes	No	Yes
Exemptions for national agencies	No	Yes	Yes
Limits data retention	Yes	No	Yes
Policies on automated algorithms	Yes	No	Yes

ePrivo.eu

- Complemented with automatic labeling of web tracking methods

Fingerprinting

- Font fingerprinting
- Canvas fingerprinting
- Mouse fingerprinting
- WebGL fingerprinting

Cookies

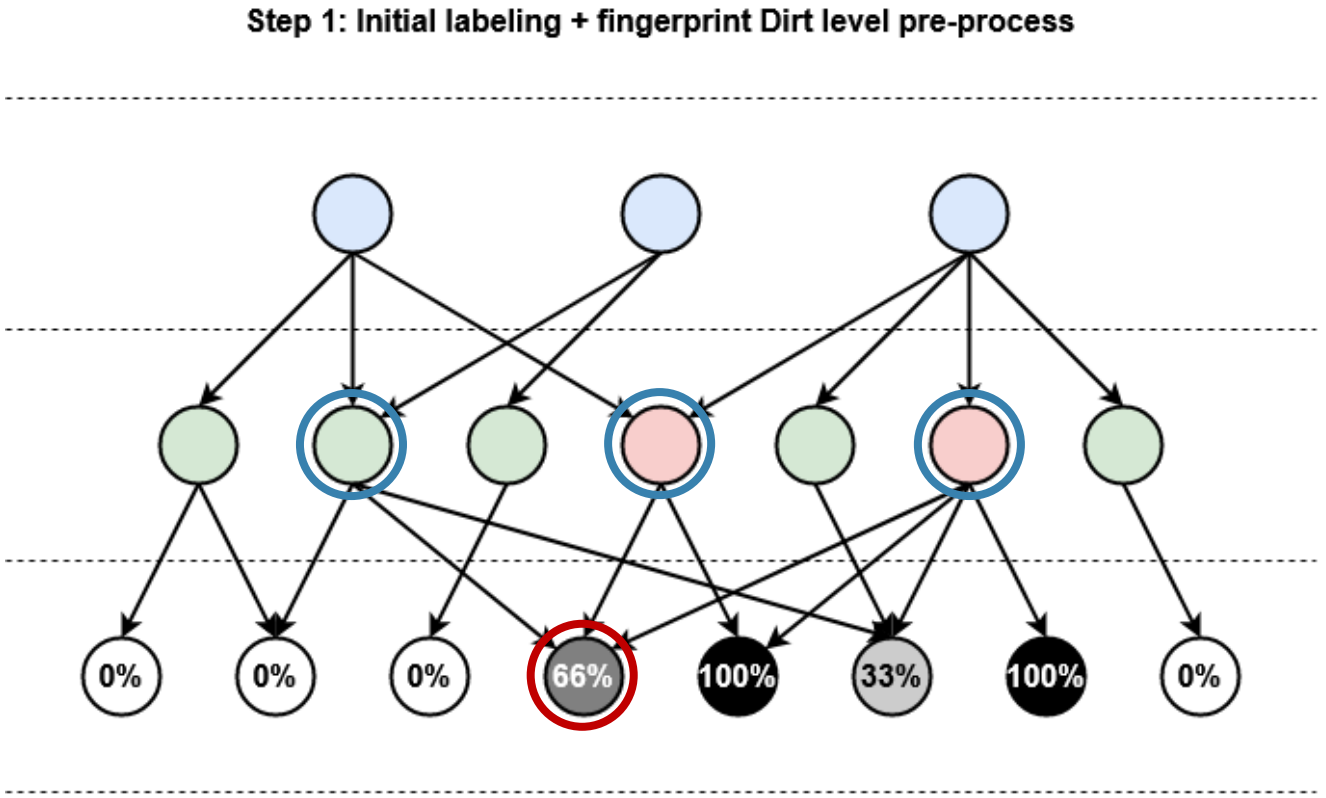
- Session cookies
- Long-living cookies (less than 1 year)
- Very long-living cookies (more than 1 year)
- JavaScript cookies
- Third-party cookies
- Tracking cookies

TrackSign Algorithm

Domain

Resource

Signature

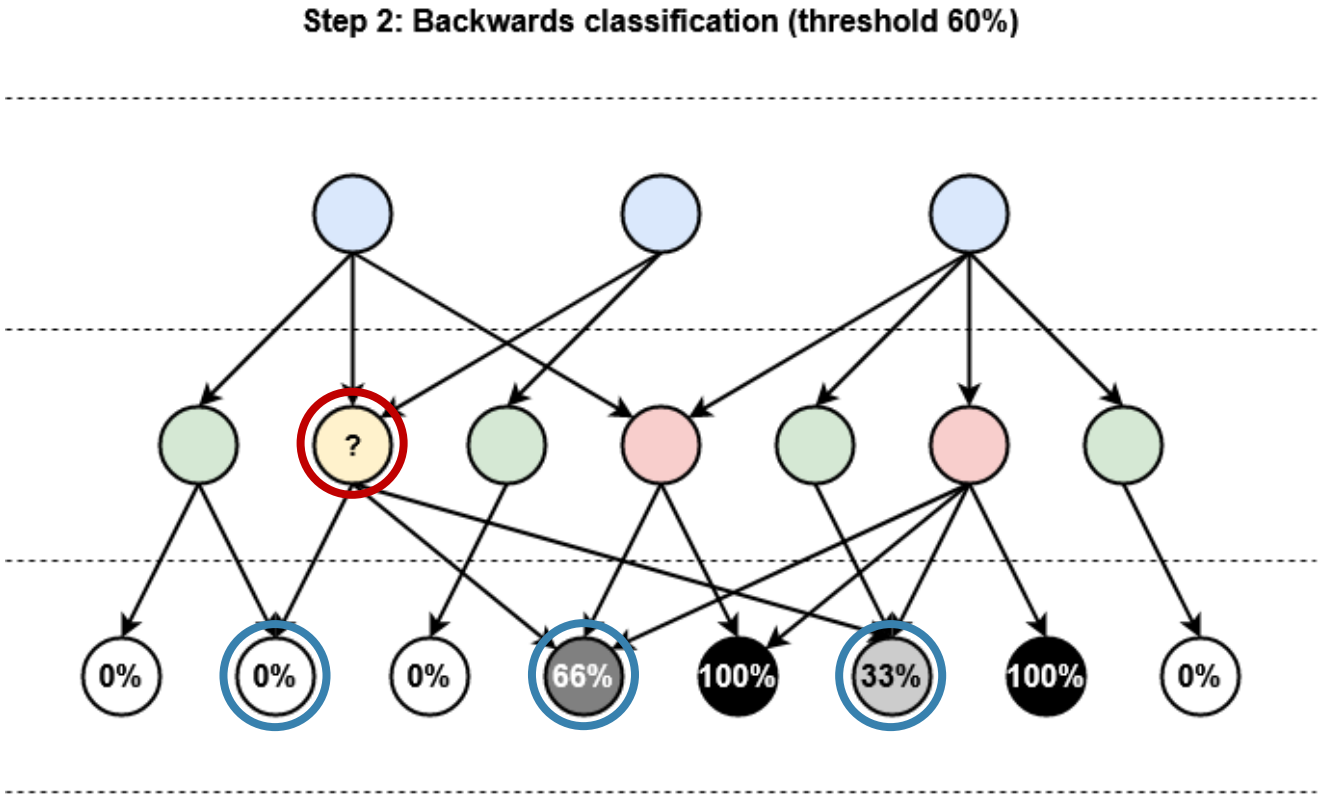


TrackSign Algorithm

Domain

Resource

Signature

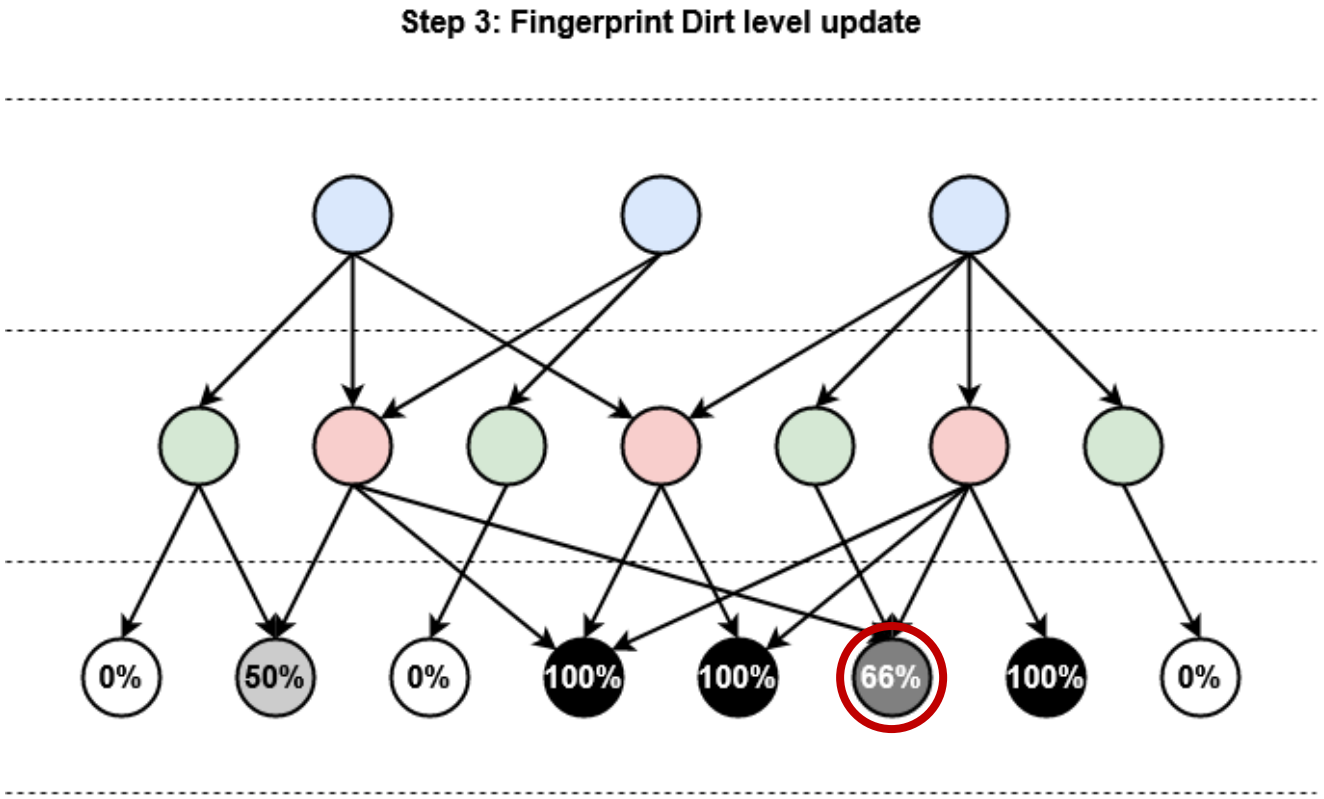


TrackSign Algorithm

Domain

Resource

Signature



TrackSign evaluation

- Completely new web tracking -> popularity level
- Wayback machine
- Popularity 10 -> 737 resources

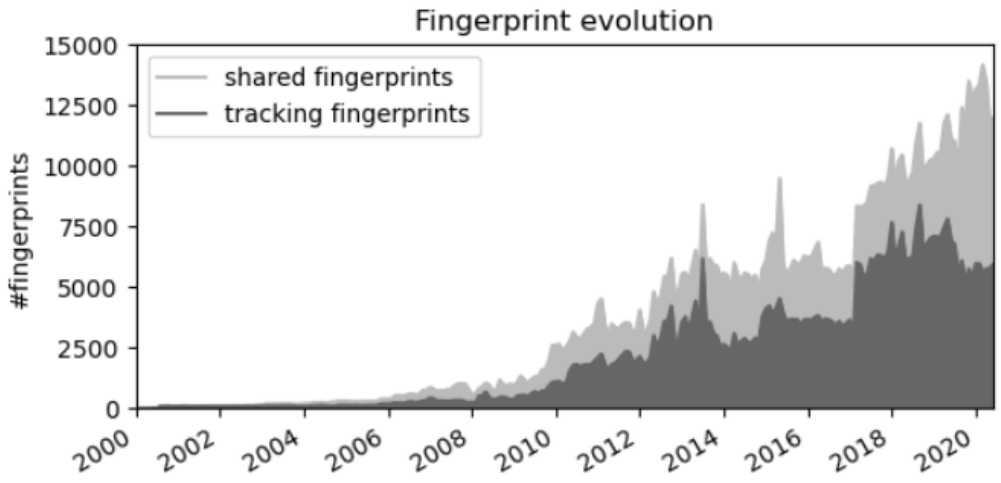
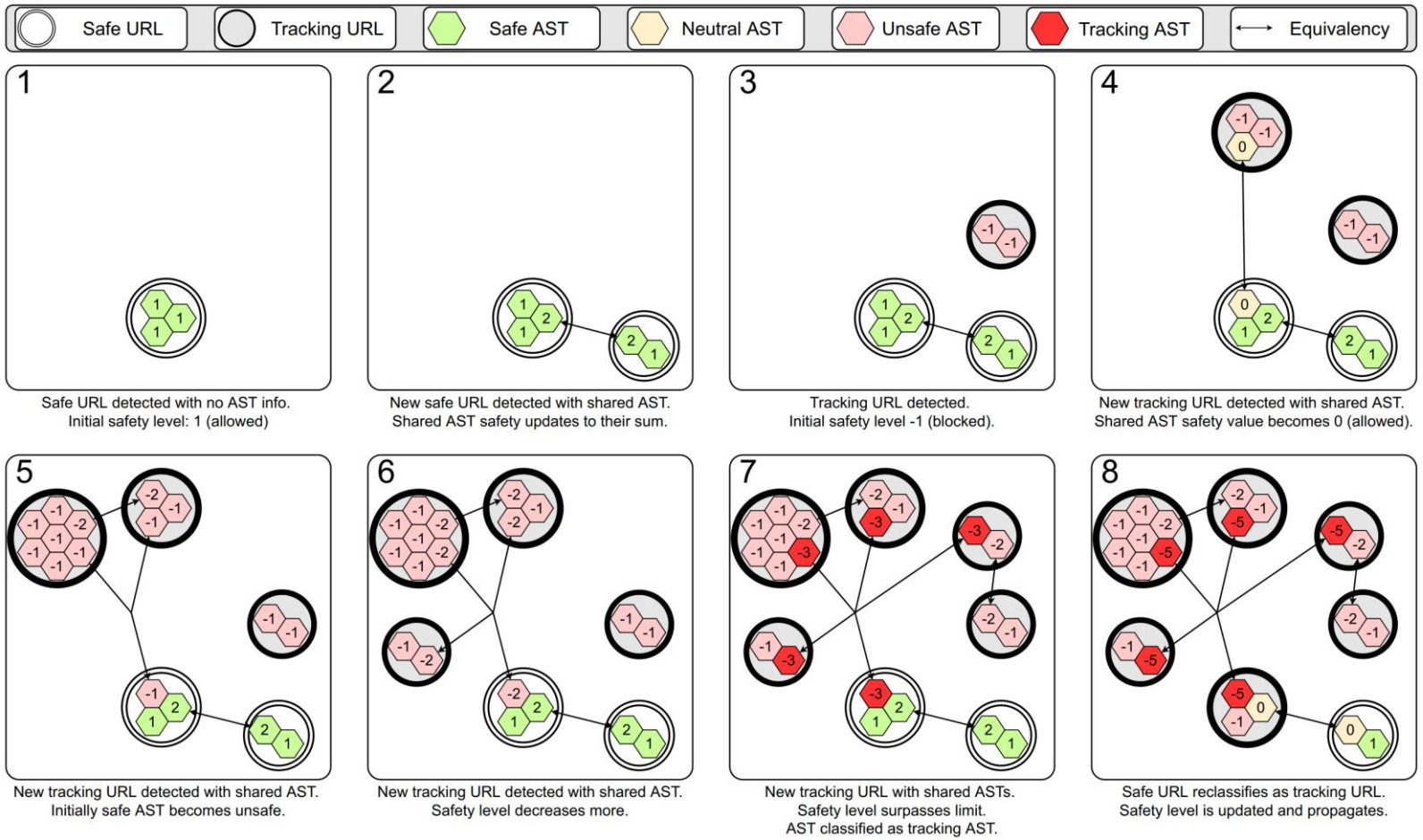


TABLE II
MOST POPULAR SERVICES

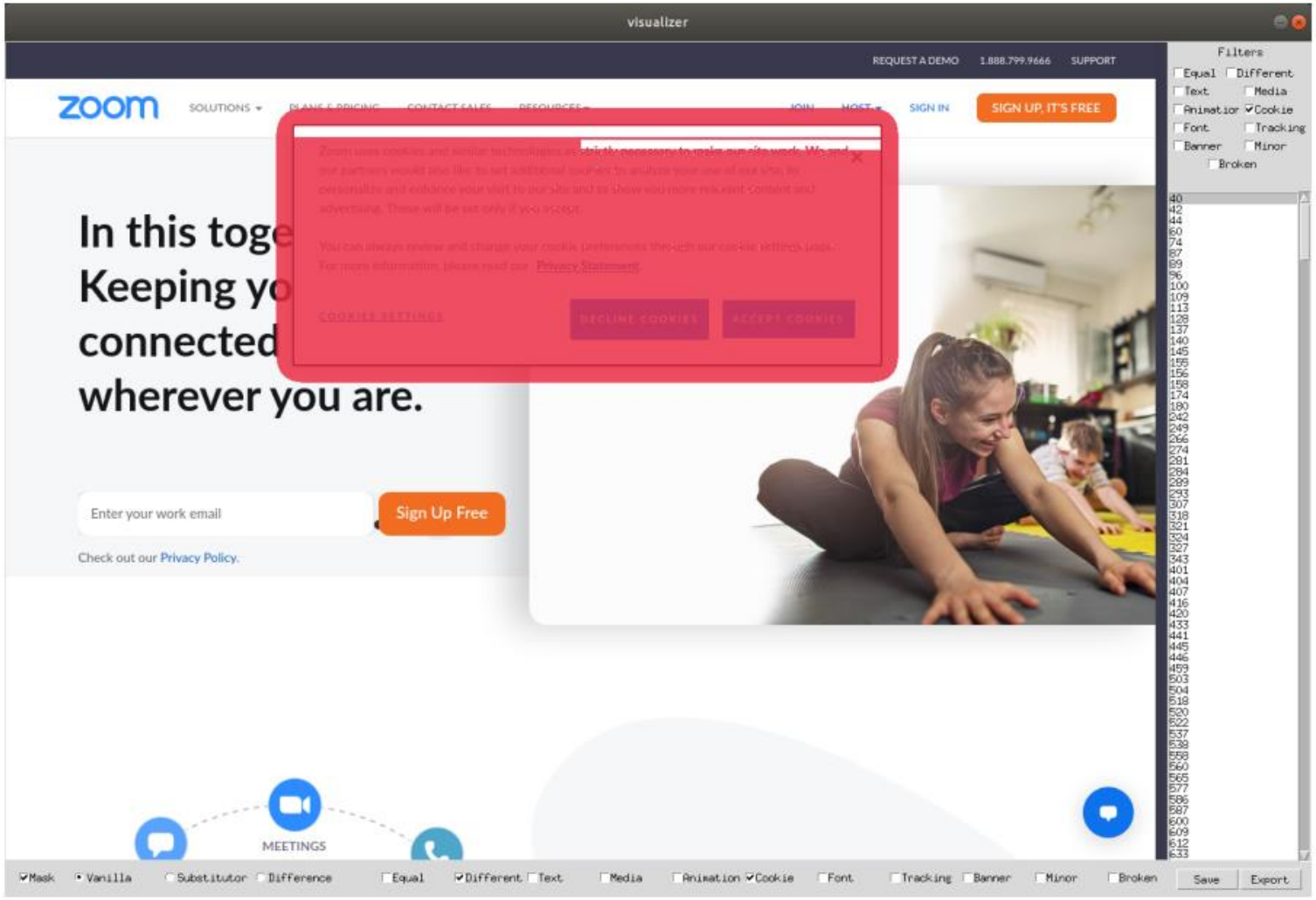
Service	Tracking	Domains	URLs	First detection
Archive.org custom libraries	No	26	127	2000-09
Archive.org analytics	Yes	54	276	2001-04
JQuery library	No	64	465	2003-02
Flash player detection library	Yes	14	14	2007-09
Google analytics (library call)	Yes	24	52	2009-06
Google tracking	Yes	29	2578	2012-04
Facebook events (library call)	Yes	10	10	2015-06
Google Tag Manager (library call)	Yes	31	45	2016-08
Google analytics	Yes	14	14	2017-01
Youtube tracking	Yes	30	403	2017-03
Zoom.us	No	21	21	2020-03

ASTrack algorithm

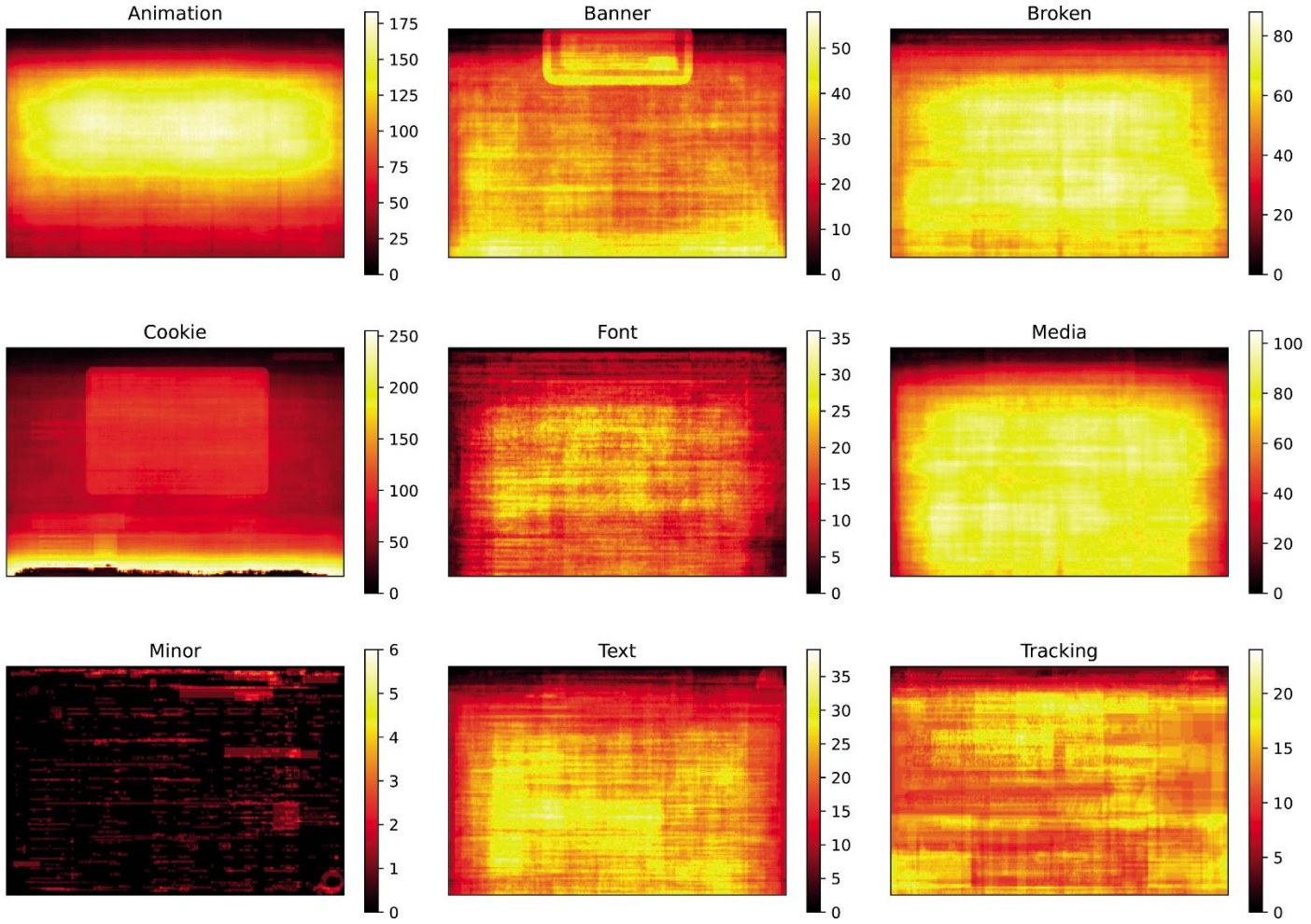


"**ASTrack: Automatic Detection and Removal of Web Tracking Code with Minimal Functionality Loss**," *IEEE INFOCOM,* 2023. https://arxiv.org/pdf/2301.10895.pdf

ASTrack evaluation



ASTrack evaluation



ASTrack evaluation

Obfuscated code matching example #1

```
function a0_0x4172(_0x3dcddc,_0x3f9388){return a0_0x4172=
    function(_0x5a6727,_0x4172bc){_0x5a6727=_0x5a6727-0x19d
    ;var _0x22ac18=a0_0x5a67[_0x5a6727];return _0x22ac18;},
    a0_0x4172(_0x3dcddc,_0x3f9388);}

function u(g,O){return u=function(a,y){a=a-475;var b=m[a];
    return b},u(g,O)}
```

Obfuscated code matching example #2

```
function(_0x210d59,_0x4d8907,_0x1f76d2,_0x4fffef){return
    _0x210d59(_0x4d8907,_0x1f76d2,_0x4fffef);}

function(api,id,key,location){return _callStorageFunction(
    _getItem,arguments,location)}
```