

Research in Privacy on the Web

Walter Rudametkin

Walter.Rudametkin@univ-rennes.fr

<https://rudametw.github.io/teaching/>

Walter Rudametkin

Systemes et logiciels

Sécurité et Vie Privée

Postes

2022: Professeur à l'Université de Rennes 1
IRISA – EPI DiverSE



2014–2022: Maître de Conférences à l'Université de Lille
CRISAL – EPI SPIRALS



2013–2014: Chercheur postdoc l'Inria Rennes
IRISA – DIVERSE



2007–2011: Doctorat CIFRE Bull S.A.S, LIG Adèle



Diplômes

Juin 2021 : Habilitation à Diriger des Recherches

Fév. 2013 : Doctorat Université de Grenoble (LIG, équipe Adèle)

Sep. 2007 : M2R Informatique (UJF)

Sep. 2007 : Ingénieur Ensimag (INP Grenoble)

Juin 2006 : Licence (bac+4) UABC Mexique

THE WEB

Internet: a massive-scale system

- Very high quality and reliable applications
- Billions of users
 - ~3 billion users in 2014 (864 M for Facebook, 350M for YouTube, 283M for Twitter, etc.)
- Massive traffic
 - May 2013: 100 hours upload every minute on YouTube
 - 500M tweets a day
- Very heterogeneous and unpredictable environments
 - **18,796 distinct Android devices**
 - unpredictable network quality

Internet: a wonderful free world

Users have access to massive quantities of services and applications for

free

**THERE AINT
NO SUCH
THING
AS A
FREE
LUNCH**

Armies of highly qualified software, system and network engineers, designers and professional content producers do not work for free

Advertisements

Oral-B

Video will play after ad

Ad · 0:18 · enviedeplus.com/oralbpromo

0:01 / 0:20

Oppenheimer | New Trailer Concept | Experience It In IMAX®

MovieTrailers Entertainme... 581K subscribers

Subscribe 2.6K

197K views 5 months ago #Oppenheimer #ChristopherNolan #OfficialTrailer

#Oppenheimer #OfficialTrailer #ChristopherNolan

Only In Theaters 7-21-23 | Experience It In IMAX®

Show more

206 Comments Sort by

- Oppenheimer | I**
Universal Pictures
29M views · 1 mo
- OPPENHEIMER**
(Universal Studi
Universal Pictures
1.1M views · 1 mo
- A Scientist Rea**
Oppenheimer Tr
Michael Siegel
303K views · 1 mo
- J. Robert Oppen**
become Death,
PenilunePictures
19M views · 11 ye
- Quentin Taranti**
Makes 'Dunkirk'
Bill Simmons
3.9M views · 3 ye
- Oppenheimer**
Tommy Lucas - To
6.6K views · 1 ye
- Asteroid City - C**
In Select Theate
Focus Features
16M views · 2 mo

Le Monde S'abonner

Assurance & Protection
Épargne & Retraite

abeille
ASSURANCES

**UN ASSUREUR PLUS PROCHE,
ÇA M'AIDE À ALLER PLUS LOIN.**

1000 agents généraux partout en France

En savoir plus

Les admirateurs affluent aux obsèques de Silvio Berlusconi à Milan

Florian Grill, un nouveau « président en cohabitation » de la Fédération française de rugby

PUBLICITÉ proposé par **SPOTiCAR**

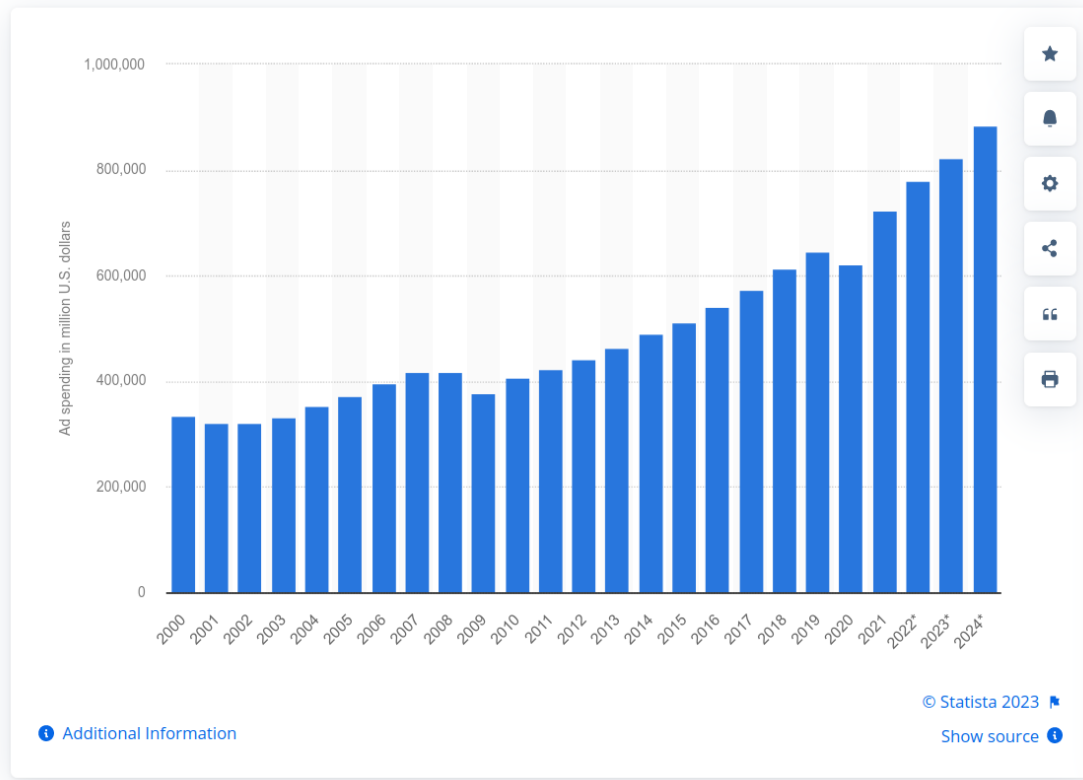
Et si, demain, l'IA rendait la voiture d'occasion plus sûre et plus écologique ?

Advertisements

Advertising & Marketing > Advertising

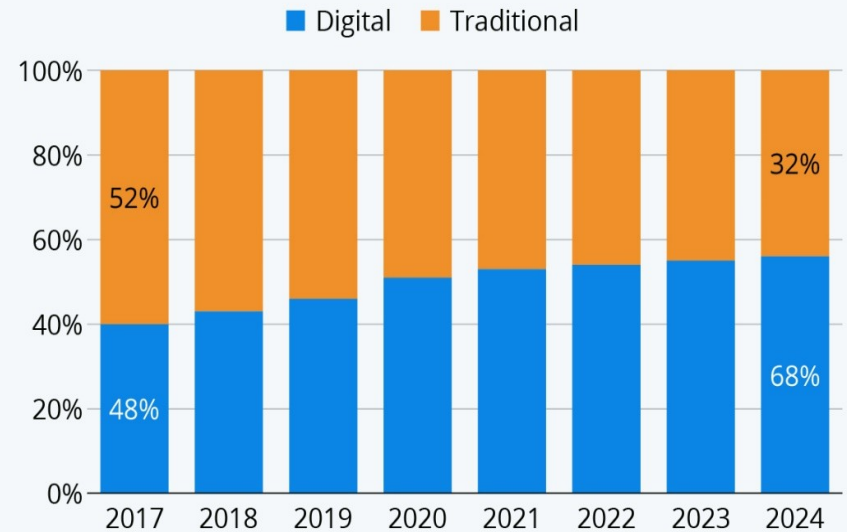
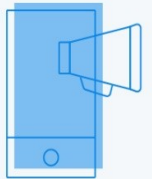
Advertising spending worldwide from 2000 to 2024

(in million U.S. dollars)



Almost Two Thirds of Ad Spending Is Digital

Digital and traditional formats as a share of ad spend in the U.S. (in %)



Source: Statista Advertising & Media Outlook



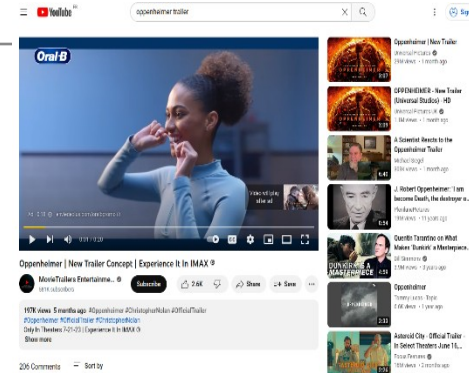
statista

Physical vs Digital Advertising



Physical world

- How many people saw the ad?
- How did they respond to it?
- Did it influence their spending? Increase awareness?



Digital world

- Ad presented to each user
- Number of **impressions**, **interactions** (clicks or taps), **purchases**
- Possible to target each ad to a person

Types of ads

Contextual advertising

- Use the contents of the page to chose the ad

Targeted advertising

- Use the user's profile to deliver an ad

Runner's World website showing contextual advertising. The page features a navigation bar with 'RUNNER'S WORLD', 'GIFT GUIDE', 'SHOES', 'TRAINING', 'SUBSCRIBE', and 'NEWSLETTER'. Below the navigation bar is a horizontal carousel of five articles: '1 Hottest New Shoes for Running and Wearing All Day', '2 The Doctor Is Win', '3 Start Your Winter #RWRunStreak on Thanksgiving Day', '4 The Best Gifts for Runners', and '5 One Energy Drink Could Mess With Your Blood Flow'. Below the carousel is a blue banner for Saucony 'ride ISO' shoes with a 'Shop' button.

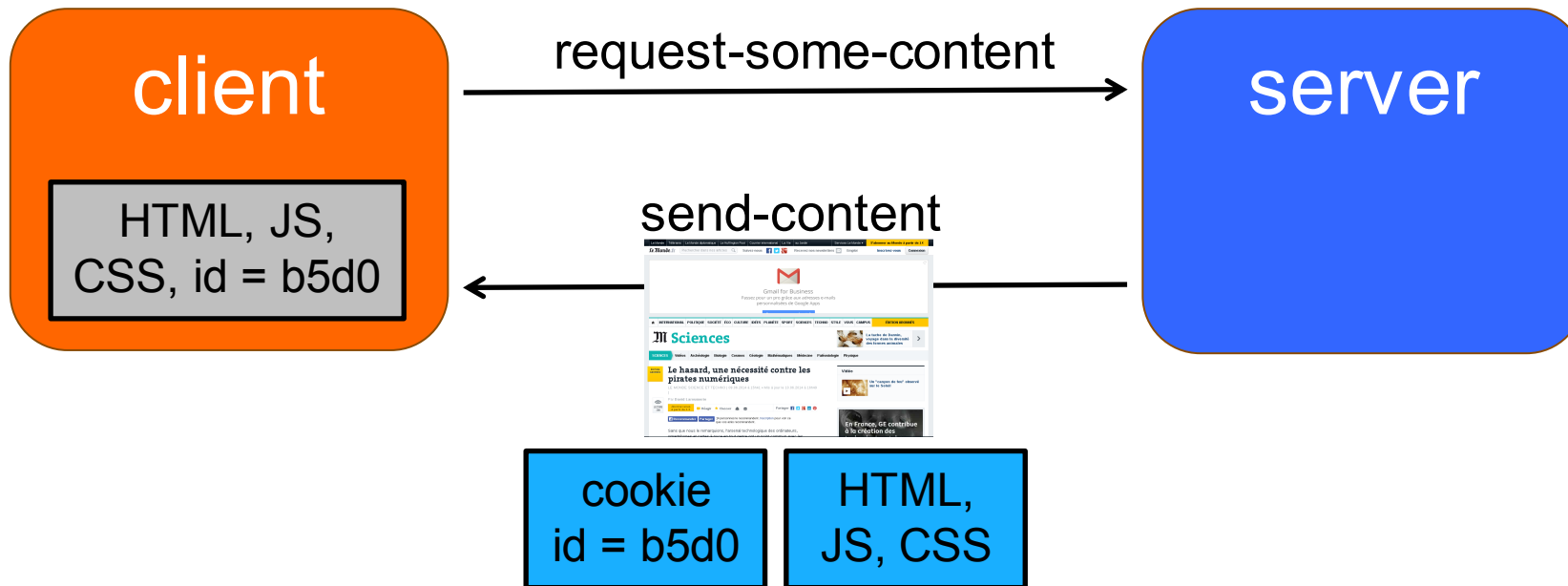
How to Add Speed Workouts to Marathon Training

Running fast improves your form, endurance, and so much more—even if you're not running that pace on race day.

Allociné website showing targeted advertising. The page features a yellow navigation bar with 'ALLOCINÉ', a search bar, and 'NOTEZ DES FILMS'. Below the navigation bar is a horizontal carousel of three movie posters: 'Final Fantasy XVI', 'Penélope Cruz et Luis Tosar engagés dans une course contre la montre face aux injustices', and 'À LA UNE'. The 'À LA UNE' section includes four small movie posters: 'Linda veut du poulet!', 'Lambert Wilson et Grégory Gadebois dans un feel-good movie', 'Tarantino l'adorait, Kubrick la redoutait...', and 'Les (autres) sorties de la semaine'.

A SIMPLE WEB MADE OF COOKIES AND ADS

A Simple Web Request





Gmail for Business

Passez pour un pro grâce aux adresses e-mails personnalisées de Google Apps

M Sciences



La tache de Darwin, voyage dans la diversité des formes animales

ÉDITION ABONNÉS

Le hasard, une nécessité contre les pirates numériques

LE MONDE SCIENCE ET TECHNO | 09.06.2014 à 15h41 • Mis à jour le 10.06.2014 à 16h49

Par David Larousserie



Abonnez-vous à partir de 1 €

Réagir

Classer



Partager f t g+ in p

Recommander Partager

1 personnes le recommandent. Inscription pour voir ce que vos amis recommandent.

Sans que nous le remarquions, l'arsenal technologique des ordinateurs, smartphones et cartes à puce en tout genre ont un point commun avec les

Vidéo



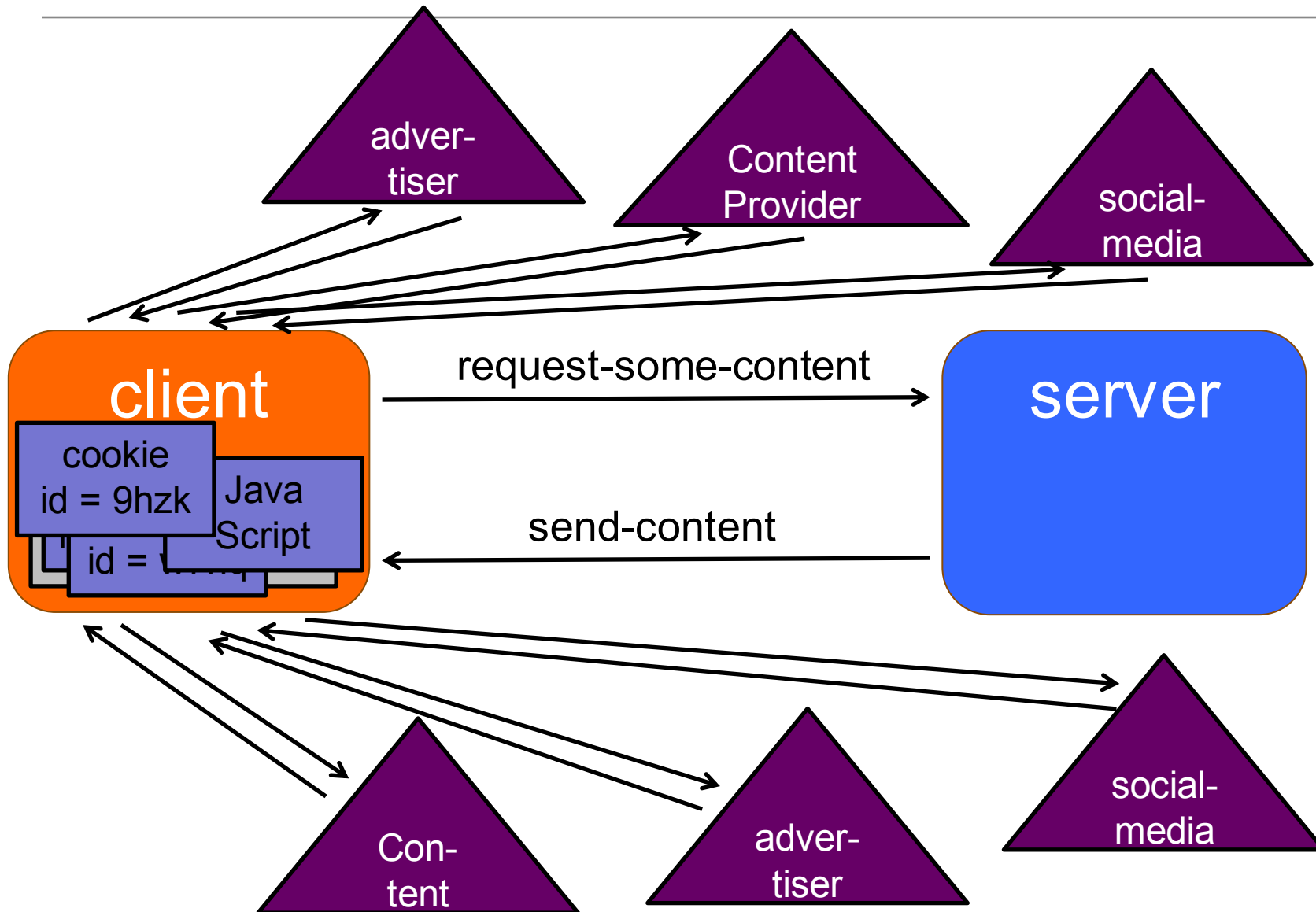
Un "canyon de feu" observé sur le Soleil

En France, GE contribue à la création des

74 third-party services (most invisible)

Accuen Media, Acuity Ads, Adap.tv, Adify, Adroit Digital Solutions, AdScale, ADTECH, Advertising.com, Aggregate Knowledge, AppNexus, AT Internet, Atlas, BidSwitch, Casale Media, Cedexis Radar, Chango, ChartBeat, Connexity, Criteo, Datalogix, DataXu, Digilant, Dotomi, DoubleClick, DoubleClick Bid Manager, DoubleClick Spotlight, EQ Advertising, Eulerian, Experian Marketing Services, eyeReturn Marketing, Ezakus, Facebook Connect, Facebook Exchange (FBX), Facebook Social Plugins, Google Adsense, Google Analytics, Improve Digital, Integral Ad Science, Jumptap, Kameleoon, Ligatus, Lijit, Magnetic, Media Innovation Group, Media Optimizer (Adobe), Media6Degrees, MediaMath, Netmining, Neustar AdAdvisor, OpenX, Optimix Media Delivery, Outbrain, OwnerIQ, PubMatic, PulsePoint, Quantcast, Right Media, Rocker Fuel, Rubicon, ScoreCard Research Beacon, SiteScout, Sizmek, SMART AdServer, SpotXchange, TradeDesk, TubeMogul, Turn, Twitter Button, Veruta, Videology, Video Step, Visual Revenu, Yandex.Metrics, Yieldr

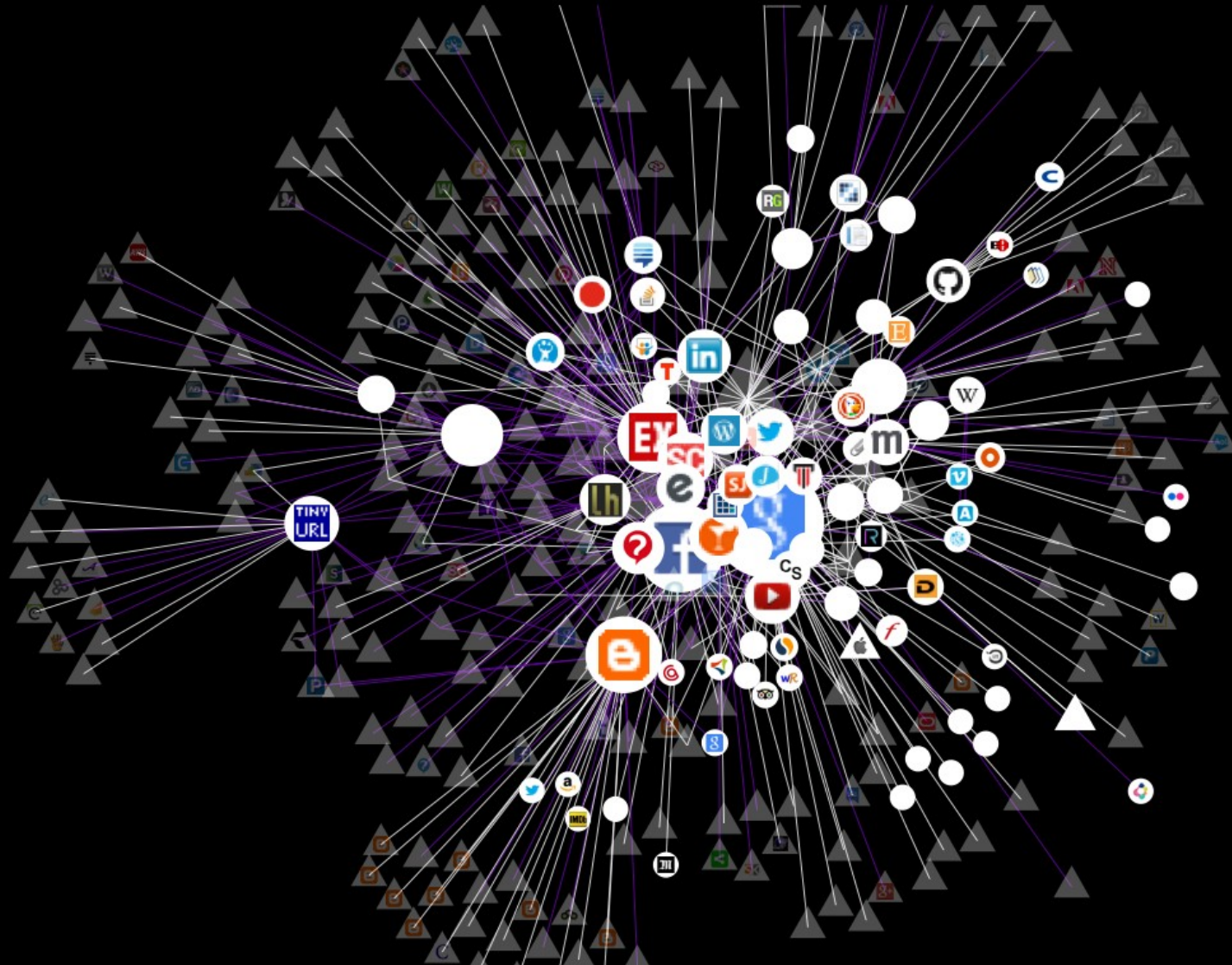
A Simple Web Request



Lightbeam

DATA GATHERED SINCE JAN 15, 2015 YOU HAVE VISITED 132 SITES YOU HAVE CONNECTED WITH 396 THIRD PARTY SITES

Daily
GRAPH VIEW



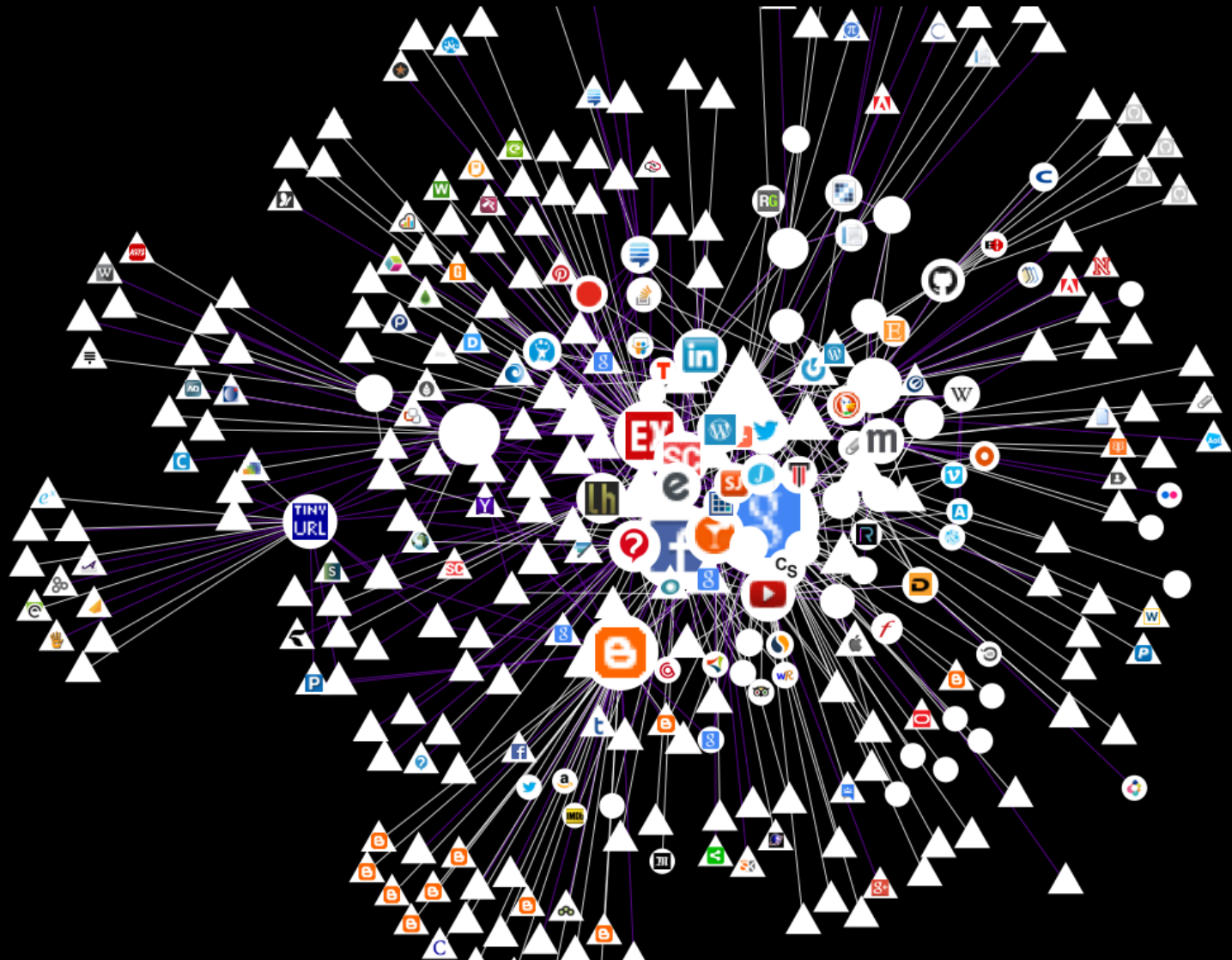
Lightbeam

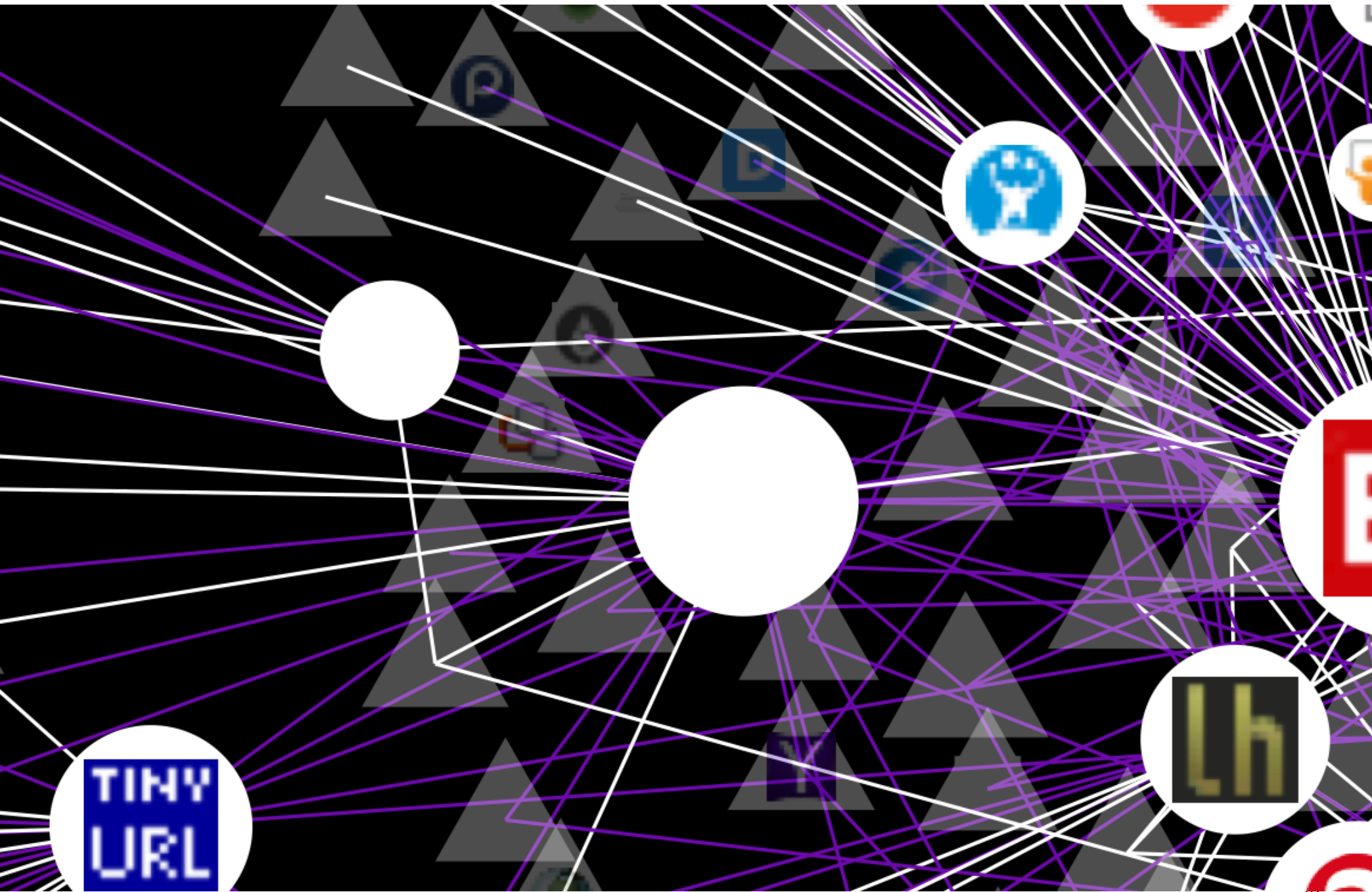
DATA GATHERED SINCE
JAN 15, 2015

YOU HAVE VISITED
132 SITES

YOU HAVE CONNECTED WITH
396 THIRD PARTY SITES

Daily
GRAPH VIEW





Marketing Technology Landscape

The Martech 5000

Total Solutions 8,000

Advertising & Promotion 922

Content & Experience 1,936

Social & Relationships 1,969

Commerce & Sales 1,314

Data 1,258

Management 601

Access all the data of this landscape & more at martech5000.com

2019

7,040 solutions

2018

6,829 solutions

2017

5,391 solutions

2016

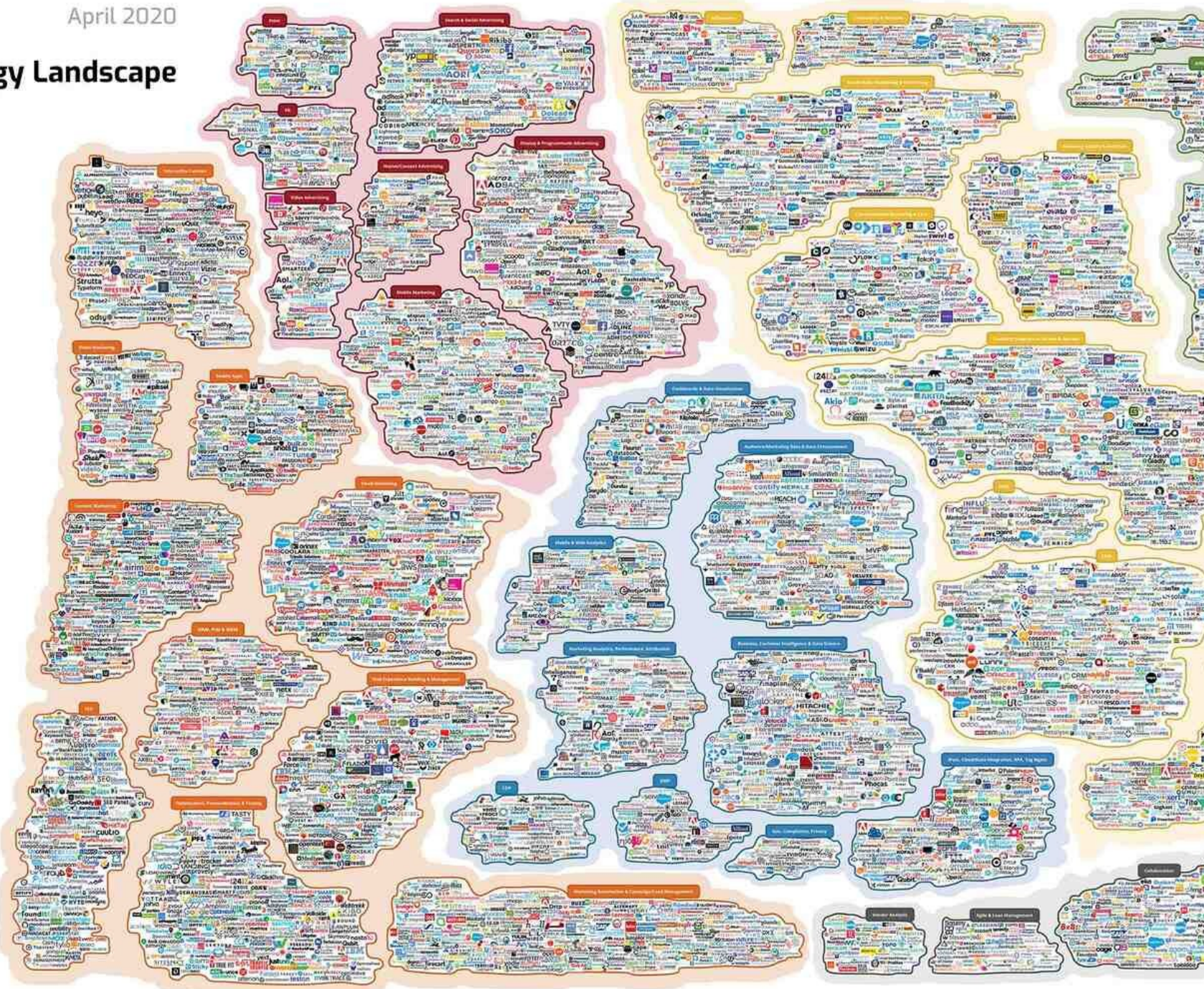
3,874 solutions

2015

1,876 solutions

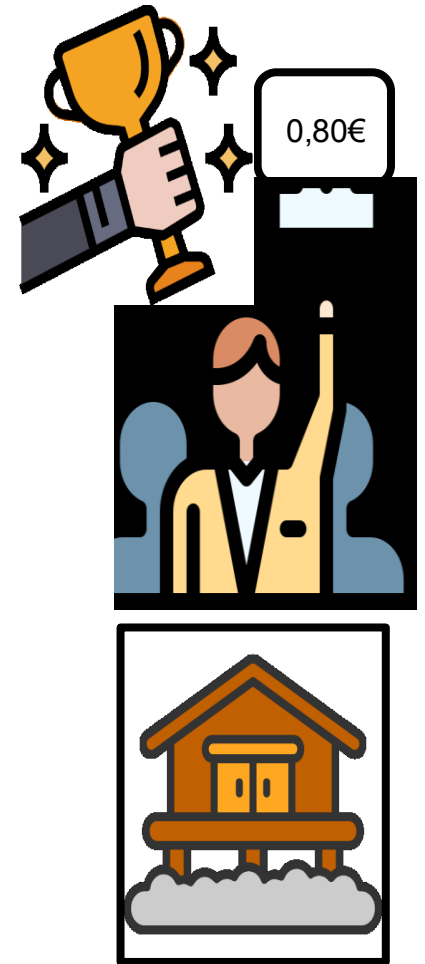
2014

547 solutions

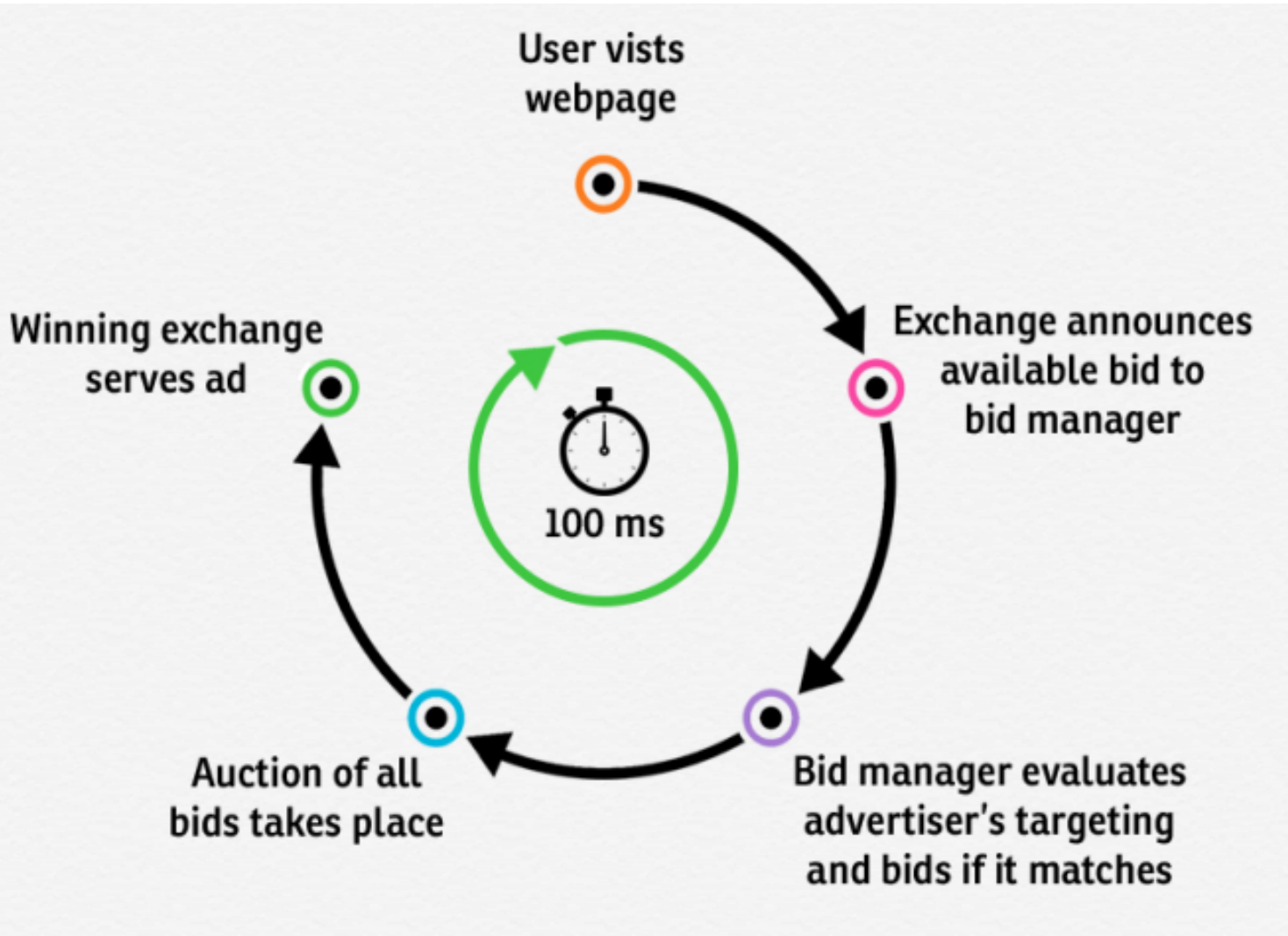


How is an ad delivered?

Real-time bidding



How is an ad TRULY delivered?



- **SSP: Supply Side Platform**, it supplies ad space
- **DSP: Demand Side Platform**, it demands ad space
- **DMP: Data Management Platform**, it provides client data in real time

Attributes can be invasive

The image shows a collage of Facebook targeting interface elements. At the top, two panels show the text "INCLUDE people who match at least ONE of the following" with search boxes containing "Fascism" and "Homosexuality". Below these, a larger panel provides a detailed view of the "Income" targeting options. This panel includes a header with "Income >" and "Suggestions | Browse", and a list of income brackets, each with a "Demographics" link to its right. The list is ordered as follows:

Income >	Demographics
Income > 2. \$50,000 - \$74,999	Demographics
Income > 3. \$75,000 - \$99,999	Demographics
Income > 6. \$150,000 - \$249,999	Demographics
Income > 4. \$100,000 - \$124,999	Demographics
Income > 5. \$125,000 - \$149,999	Demographics
Income > 1. \$40,000 - \$49,999	Demographics

Can we block **third party** cookies?

CHROME

Building a more private web

Aug 22, 2019 - 3 mins read



Justin Schuh
Director, Chrome Engineering

encouraging opaque techniques such as fingerprinting.

Second, blocking cookies without another way to deliver relevant ads significantly reduces publishers' primary means of funding, which jeopardizes the future of the vibrant web.

<https://www.blog.google/products/chrome/building-a-more-private-web/>

BROWSER FINGERPRINTING



Supporting device diversity

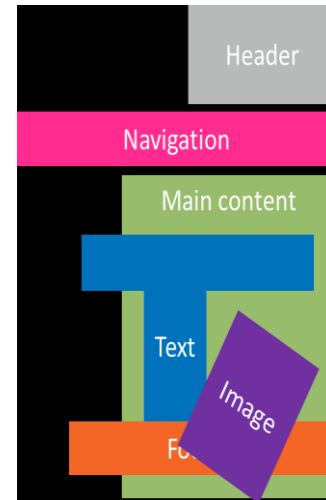
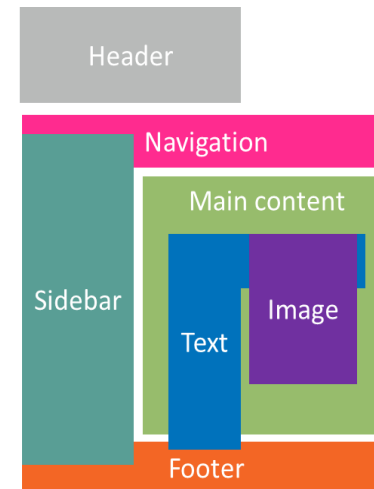
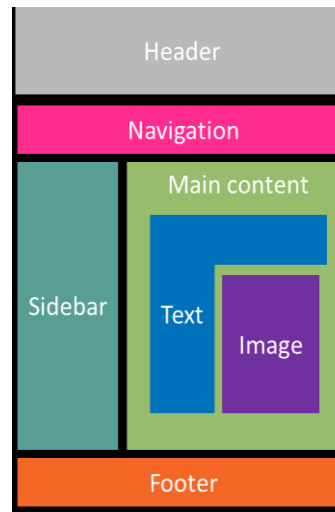
Browsers send device-specific information



Windows 7
1920x1080

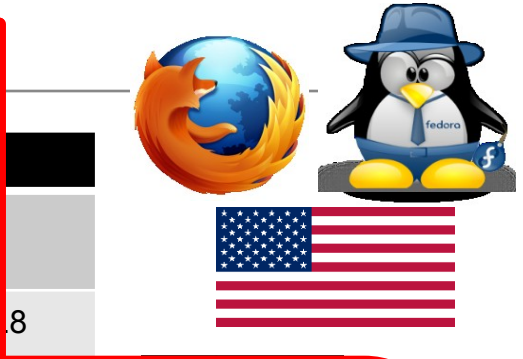


Android v6.0
2560x1440



A browser fingerprint

Attribute
User agent
HTTP accept

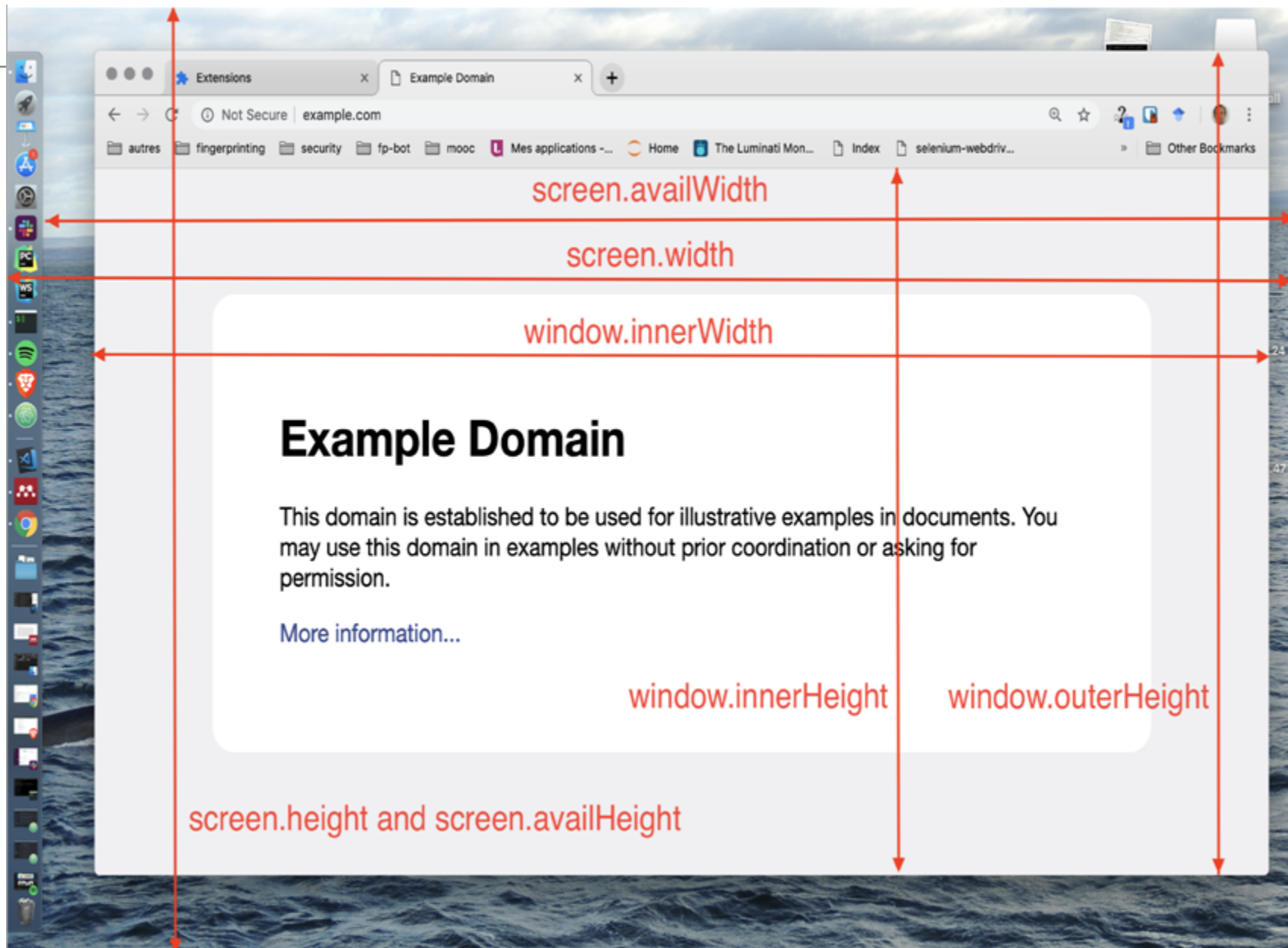


Create a unique fingerprint for each device

OS
Screen resolution
Timezone
DOM Session storage
DOM Local storage
I.E. User data



Screen resolution



<https://amiunique.org> (Launch: 2014)

[AmIUnique](#) [My fingerprint](#) [My history](#) [My extension](#) [Global statistics](#) [FAQ](#) [More ↓](#)

Learn how identifiable you are on the Internet



Help us investigate the diversity of web browsers.

This website aims at studying the diversity of browser fingerprints and providing developers with data to help them design good defenses. Contribute to the efforts by viewing your own browser fingerprint or consult the current statistics of data provided by users around the world!

[View my browser fingerprint](#)

If you click on this button, we will collect your browser fingerprint, we will put a cookie on your browser for a period of 4 months. More details are available in the privacy policy

Research & Dissemination



<http://amiunique.org> (Am I Unique?)

- Website, browser extensions, research papers
- Long-term data acquisition
 - 20K visits per week
 - 5M+ fingerprints
 - 4K extension users

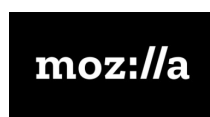


- 1 PhD 2017, 1 PhD 2019, 2 PhDs 2021, 1 PhD 2024, 3 ongoing PhDs

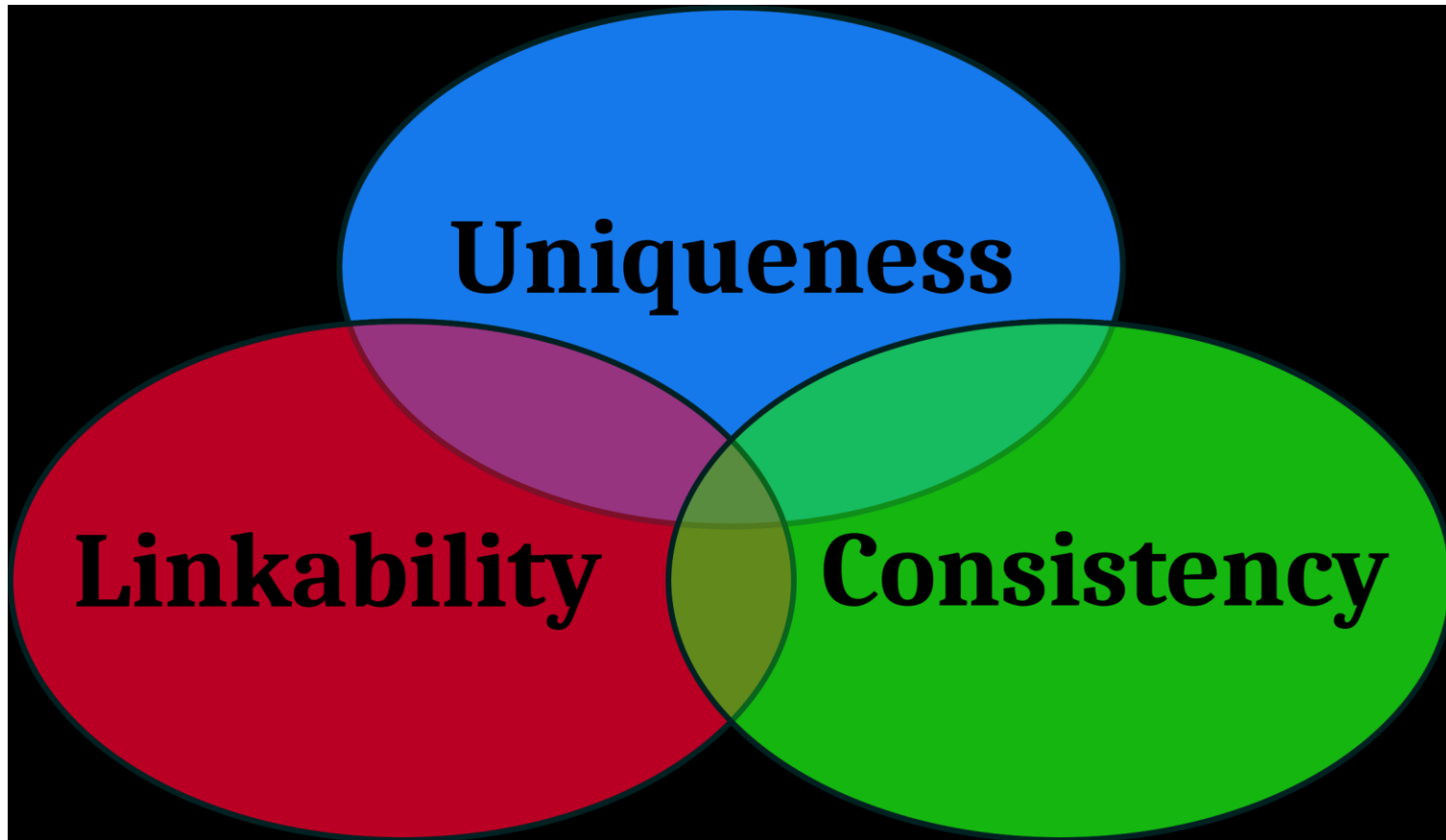
Media coverage and dissemination

- FIC, Framablog, Slashdot, Clubic, fOSSa, NextImpact, Arstechnica, CPDP, ...

Collaborations



Browser Fingerprint Trifecta

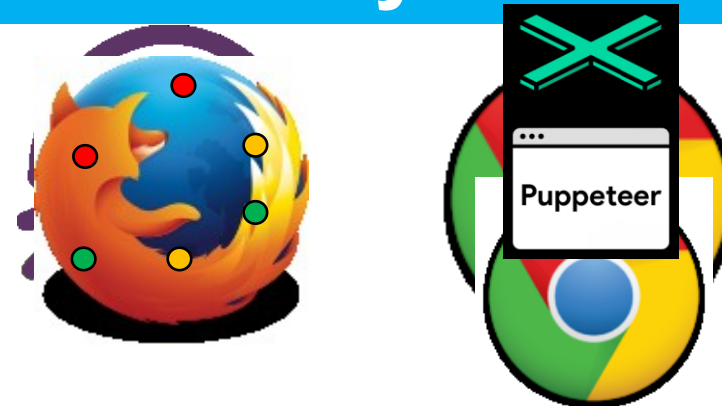


Browser fingerprint properties

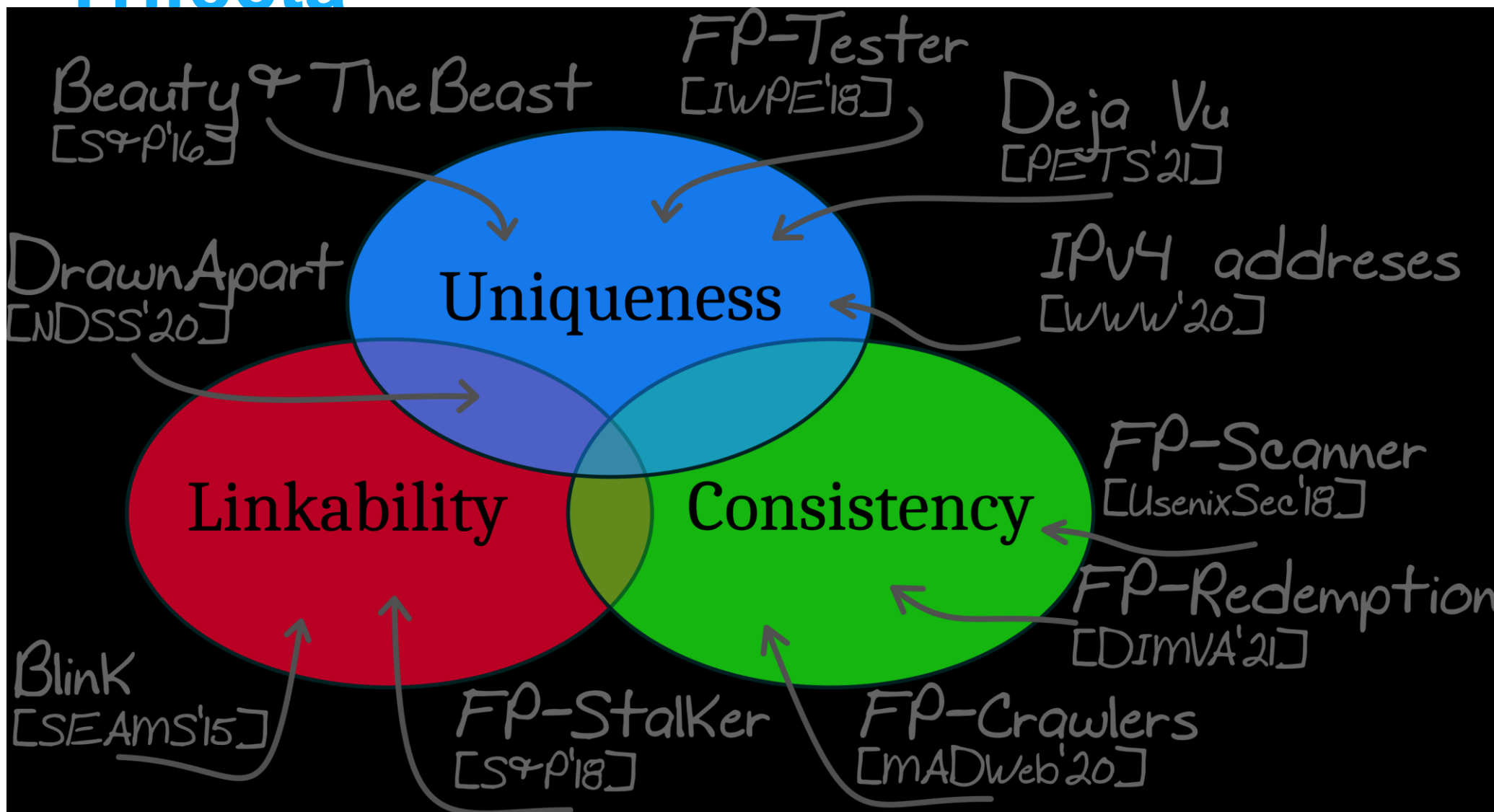
Uniqueness



Linkability



Browser Fingerprint Trifecta



Finding: Mobile fingerprints are also unique [IEEE S&P'16]

- High uniqueness

- Depends on vendor/model



(a) Windows 7



(b) Windows 10



(c) Linux



(d) iOS



(e) Firefox OS



(f) Android 4.3 and before



(g) Android 4.4



(h) Android 5.0



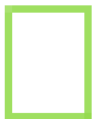
(i) Android on an LG device



(j) Android on a Samsung device



(k) Android on an HTC device



(l) Emoji not supported

- Different attributes

- User agent, Emojis, Canvas

Canvas fingerprint [IEEE S&P'16]

```
canvas = document.createElement("canvas");
canvas.height = 60;
canvas.width = 400;
canvasContext = canvas.getContext("2d");
canvas.style.display = "inline";
canvasContext.textBaseline = "alphabetic";
```



Cwm fjordbank glyphs vext quiz, 😊

Cwm fjordbank glyphs vext quiz, 😊

1

```
canvasContext.fillStyle = "#f60";
canvasContext.fillRect(125, 1, 62, 20);
```

2

```
canvasContext.fillStyle = "#069";
canvasContext.font = "11pt no-real-font-123";
canvasContext.fillText("Cwm fjordbank glyphs vext quiz, \ud83d\ude03", 2, 15);
```

3

```
canvasContext.fillStyle = "rgba(102, 204, 0, 0.7)";
canvasContext.font = "18pt Arial";
canvasContext.fillText("Cwm fjordbank glyphs vext quiz, \ud83d\ude03", 4, 45);
canvasData = canvas.toDataURL();
```

Canvas fingerprint [IEEE S&P'16]

Cwm fjordbank glyphs vext quiz, 😊

Cwm fjordbank glyphs vext quiz, 😊

Finding: Fingerprints can be linked

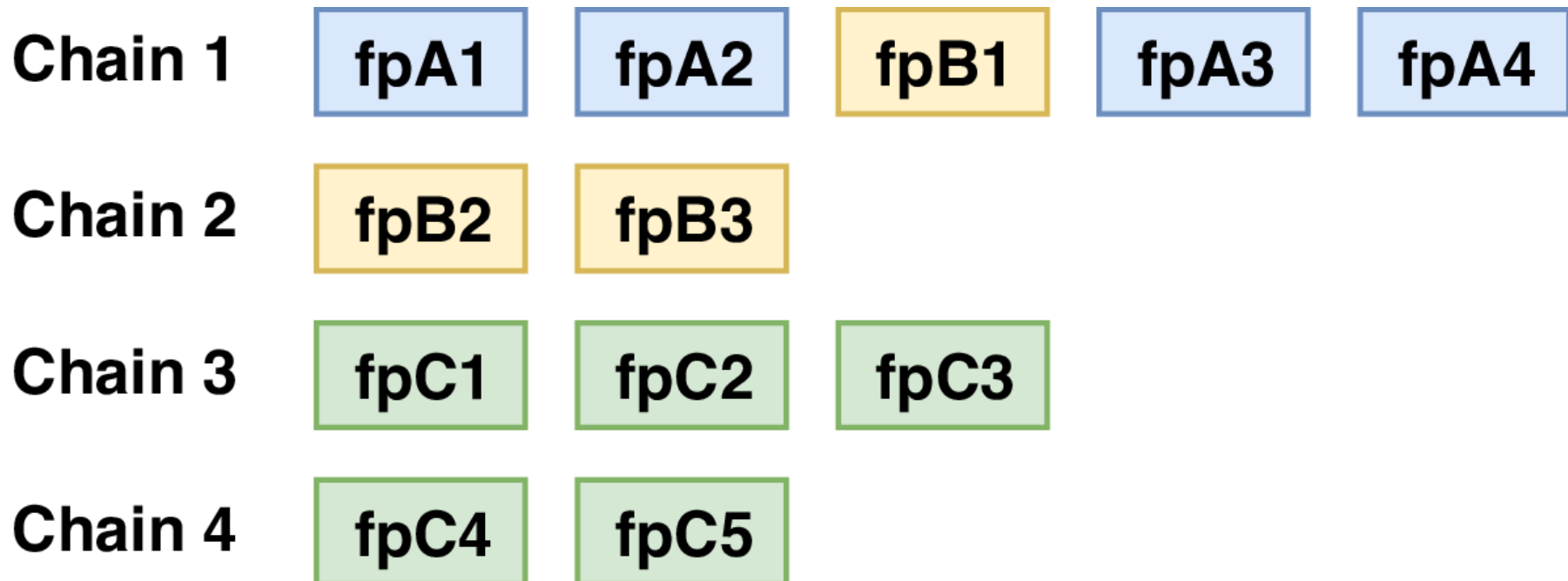
- ~20% of browsers are “*untrackable*”
 - Fingerprints too similar
 - Unpredictable evolutions
- ~26% of browsers are highly trackable
 - Unusual configurations (e.g. multiple languages)
 - Specialized hardware or software

Linking fingerprints [IEEE S&P'18]

Fingerprints (chronological collection order)

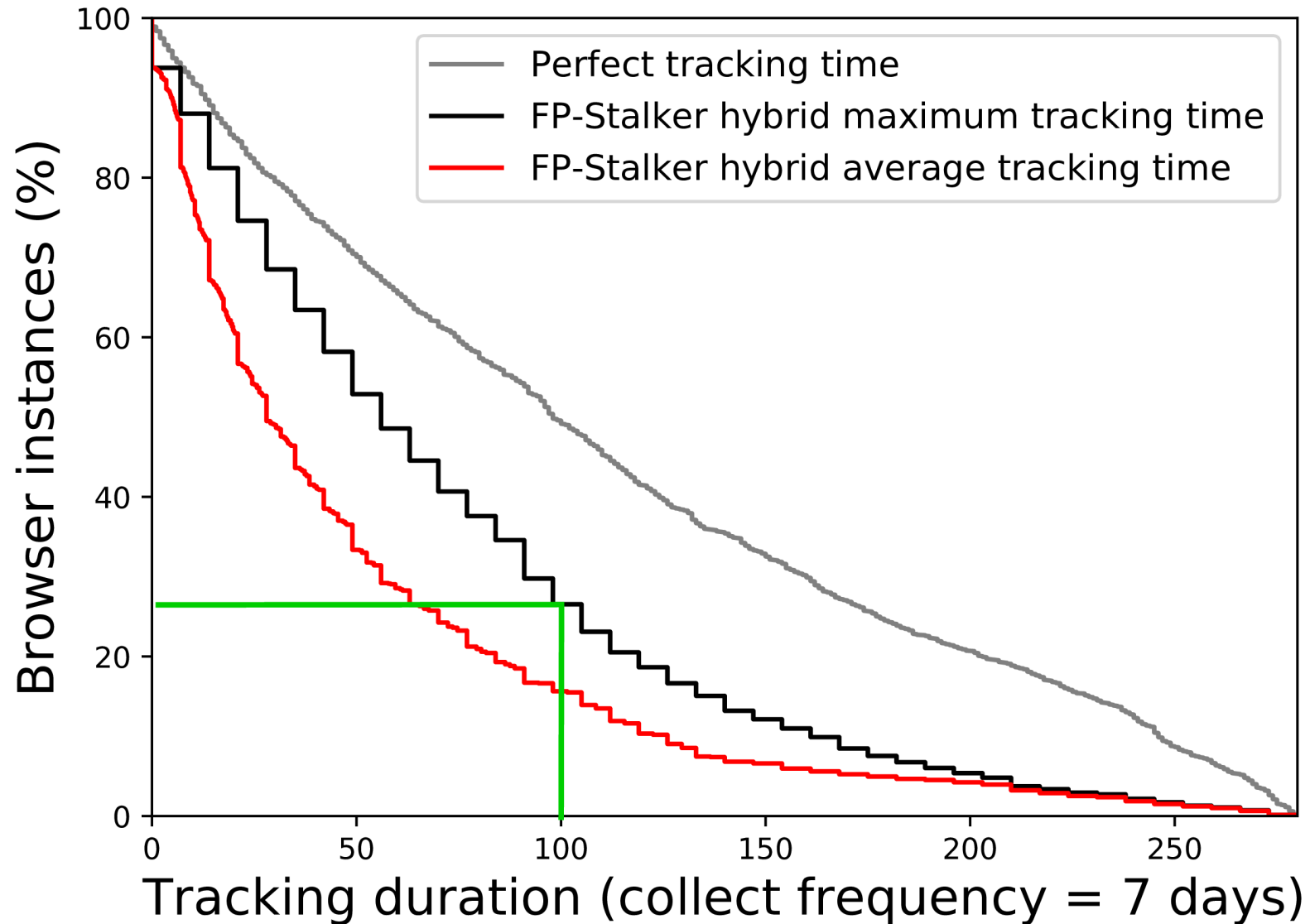


Result of linking algorithm (mistakes shown)

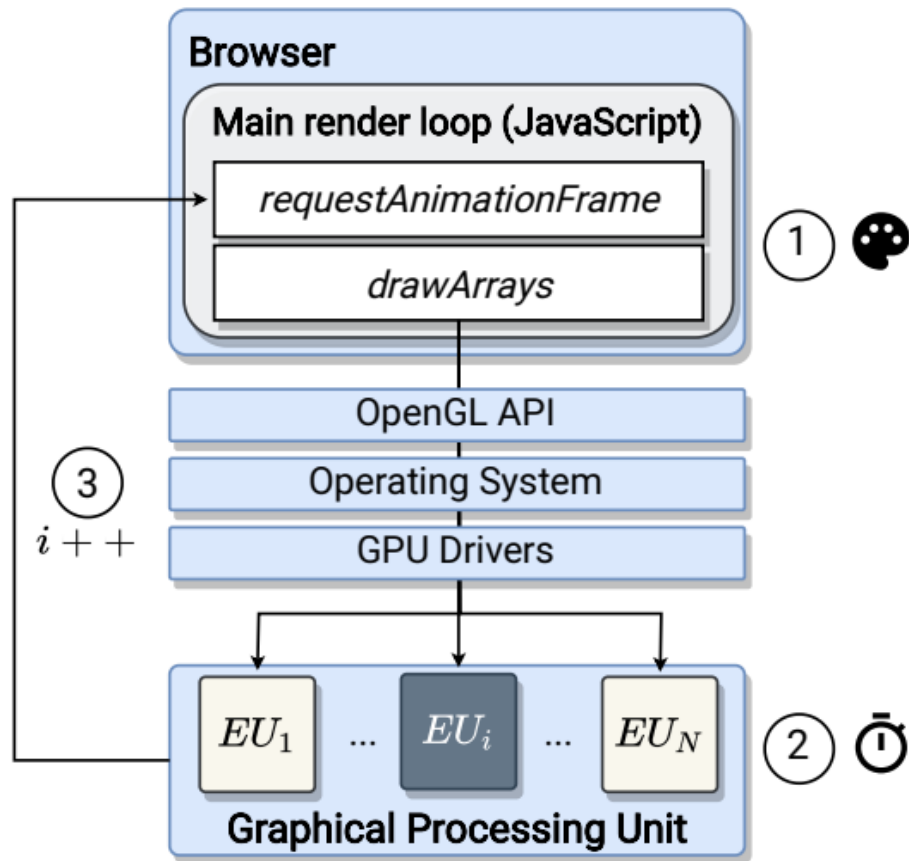


Fingerprint tracking duration

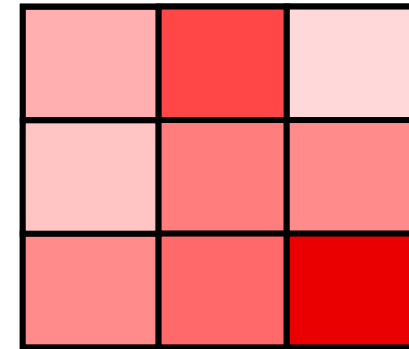
[IEEE S&P'18]



DrawnApart : GPU Fingerprinting [NDSS'22]



0.4	0.9	0.1	0.2	...	1.5
-----	-----	-----	-----	-----	-----

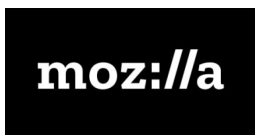


66% re-identification increment over SOA

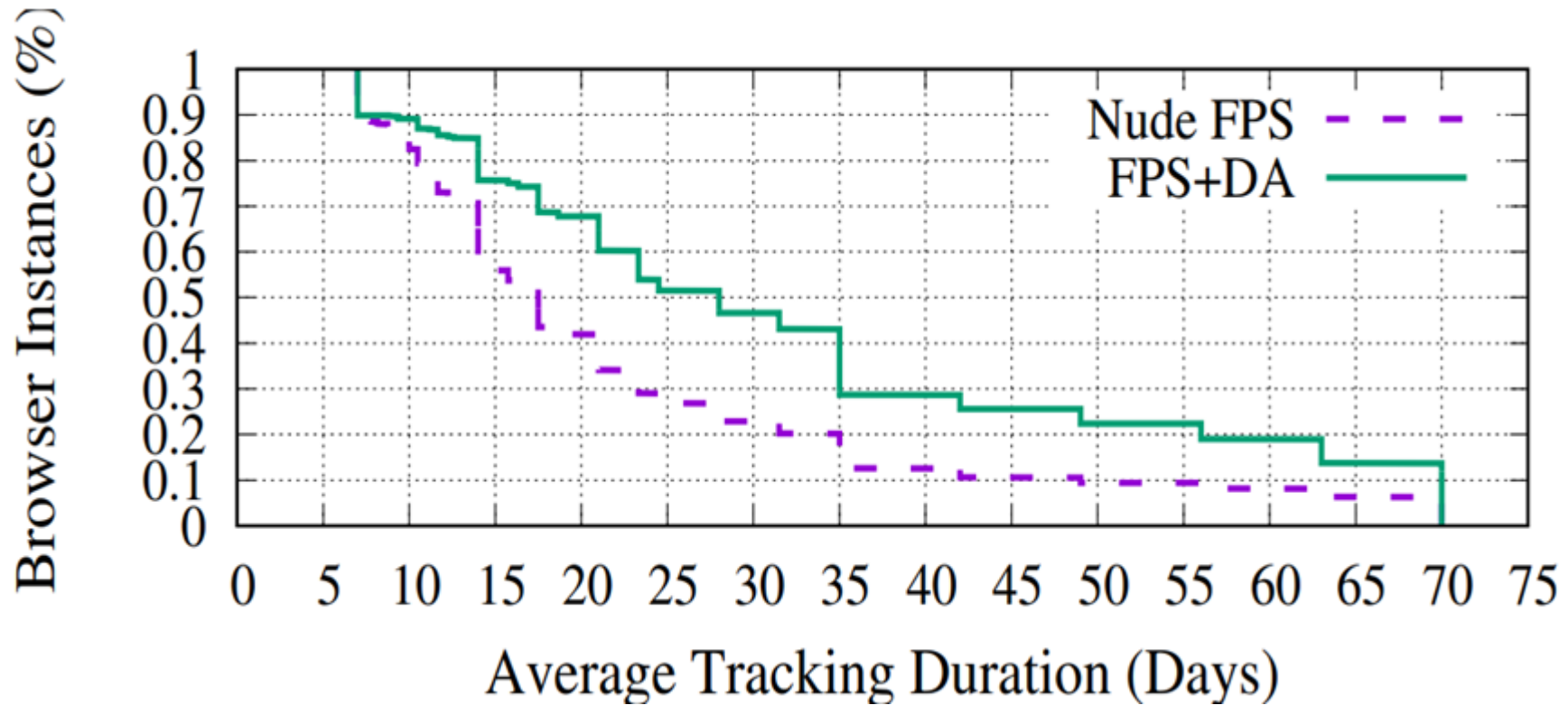
Bug Bounty

Spec update

1st Place – Applied Research



DrawnApart : GPU Fingerprinting [NDSS'22]



66% improvement over FP-Stalker (FPS: 17 days - FPS+DA 27.5 days)

What about fingerprinting consistency?

FP-Scanner : Detecting incoherencies in countermeasures
[USENIX Security'18]

Cwm fjordbank glyphs vext quiz, 😊

Cwm fjordbank glyphs vext quiz, 😊

Cwm fjordbank glyphs vext quiz, 😊

Cwm fjordbank glyphs vext quiz, 😊

Key takeaways

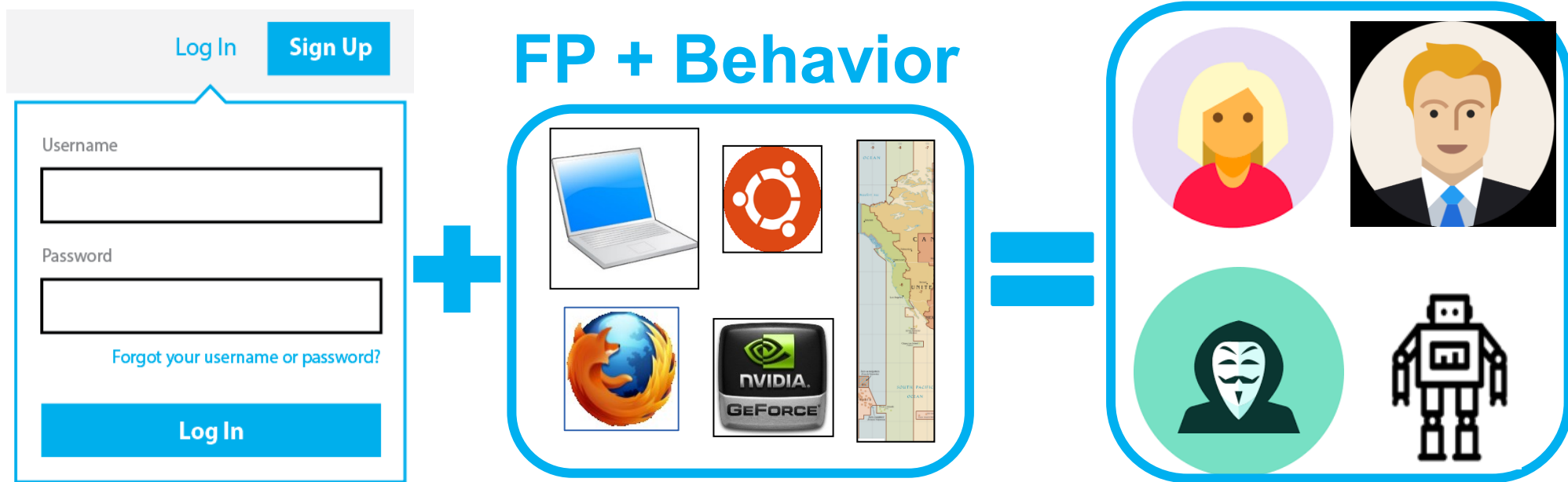
Fingerprinting is a threat to privacy [S&P'16]

- But not the *ultimate tracking technology* [S&P'18]
 - Complements cookies (e.g., cookie respawning)
- Few “good” defenses [UsenixSec'18]
 - But better ones can be made [IWIP'18, SEAMS'15]
- Many more risks to explore [WWW'20, NDSS'22]

→ **Fingerprinting can be used for security**

Bot detection [MADWeb'20] MFA [DIMVA'21]

FP-Locker : Multi-Factor Authentication Systems with Browser Fingerprinting

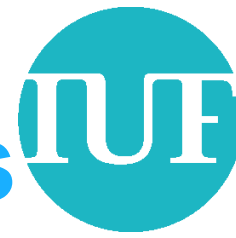


Design of a novel authentication system

CAS Prototype

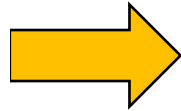
Technology Readiness Level 6 (*Demo in Relevant Environment*)₄₃

Identify new privacy risks

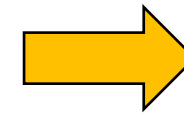


institut
universitaire
de France

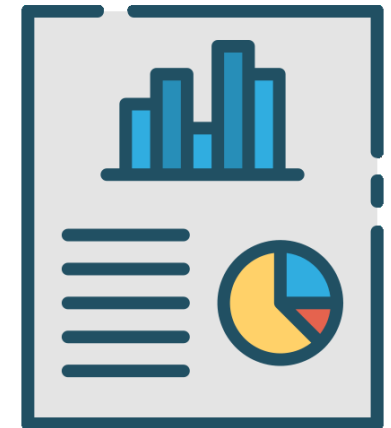
Battery of Tests



Web technologies



Models / Reports



How do we consistently identify new risks ?

From new features, changes to browser code, extensions

Systematically exploring configurations



Fingerprint to identify privacy bugs



Browser Runtime Configuration

Environment Configuration

Operating System Configuration

Firmware Configuration

Hardware Configuration

Thousands of configuration options



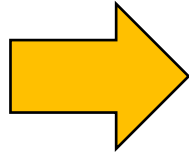
ARM

intel

Systematically exploring configurations



Compile options



Identify privacy bugs



Browser Runtime Configuration

Environment Configuration

Operating System Configuration

Firmware Configuration

Hardware Configuration

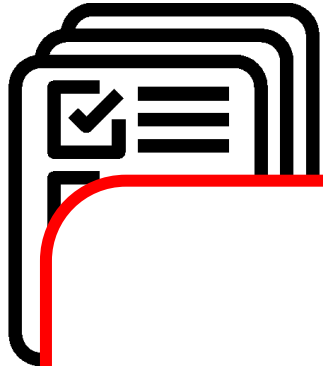
Thousands of configuration options



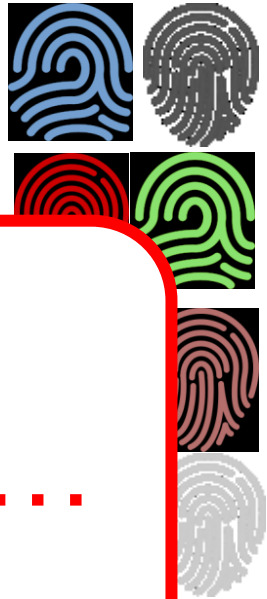
ARM

intel

Systematically exploring configurations



Identify privacy bugs



And don't forget the 200k extensions in the Chrome Store... and Android apps



Environment Configuration



Operating System Configuration

Firmware Configuration

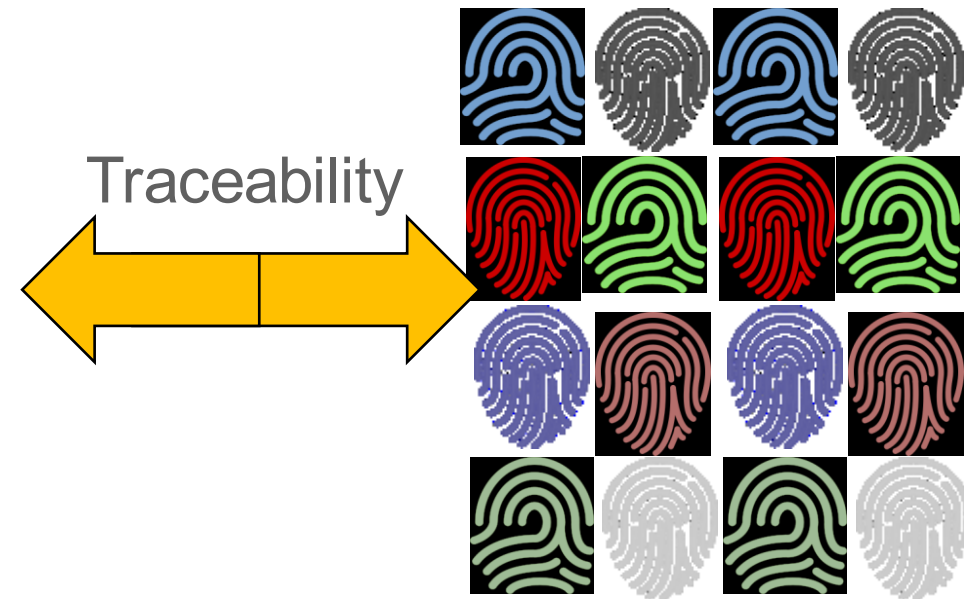
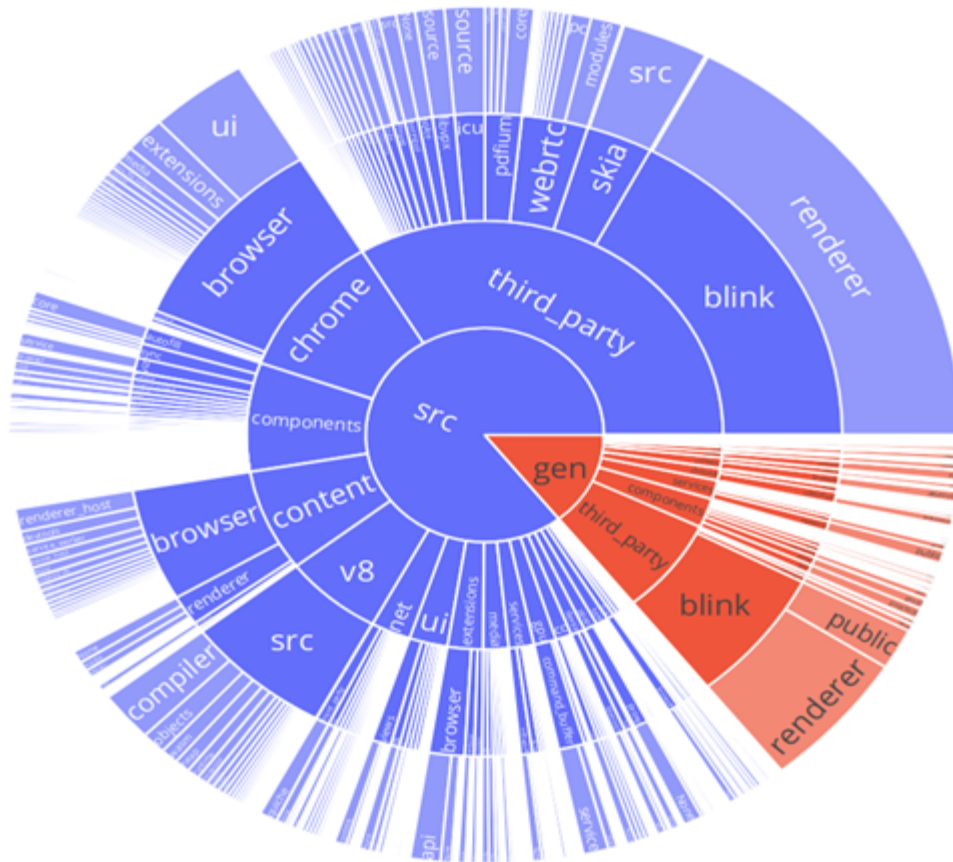
ARM

intel

Hardware Configuration

Thousands of configuration options

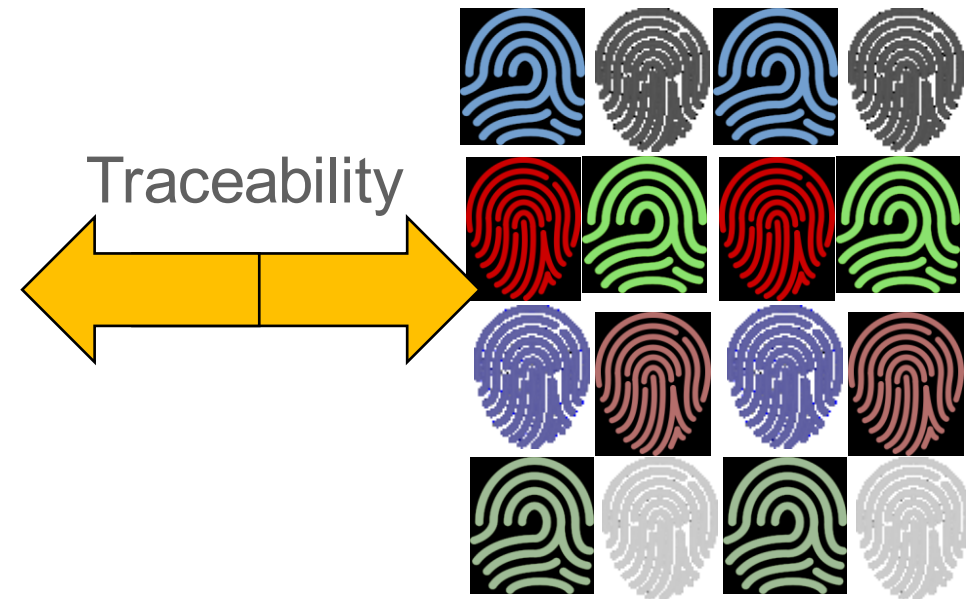
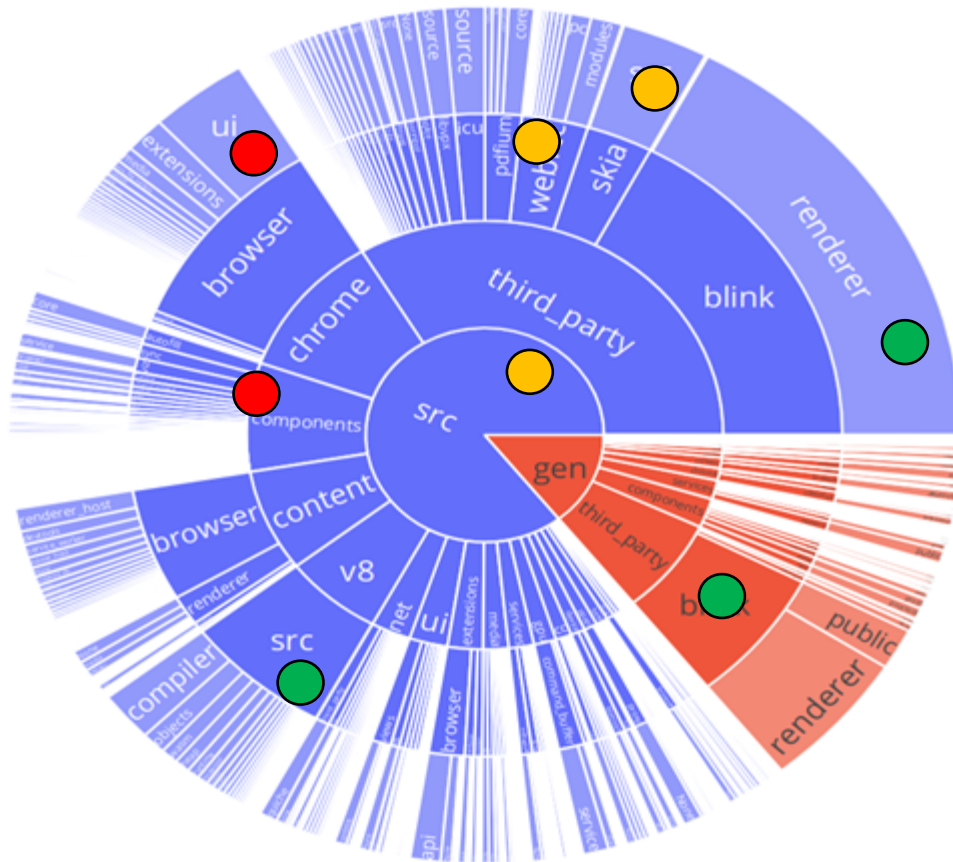
It's all in the source code !



How can we identify fingerprint attributes in source code ?

From commit history, component, functions, ...

It's all in the source code !



How can we identify fingerprint attributes in source code ?

From commit history, software component, functions, variables, ...

Paper under review

FP-Rainbow : Fingerprint-based Browser Configuration Identification

Maxime Huyghe¹, Clément Quinton¹ and Walter Rudametkin²

Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

`{maxime.huyghe,clement.quinton}@univ-lille.fr`

² University of Rennes, IRISA, IUF

`walter.rudametkin@irisa.fr`

FP-Rainbow : Pipeline

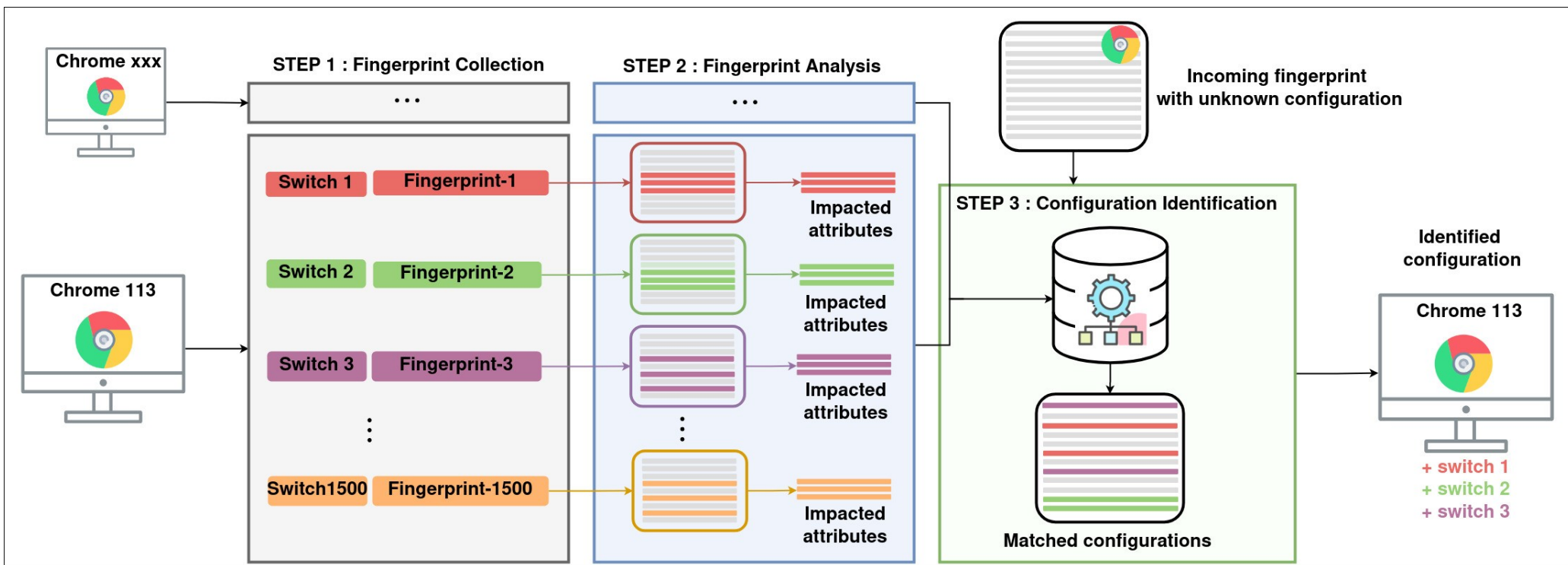
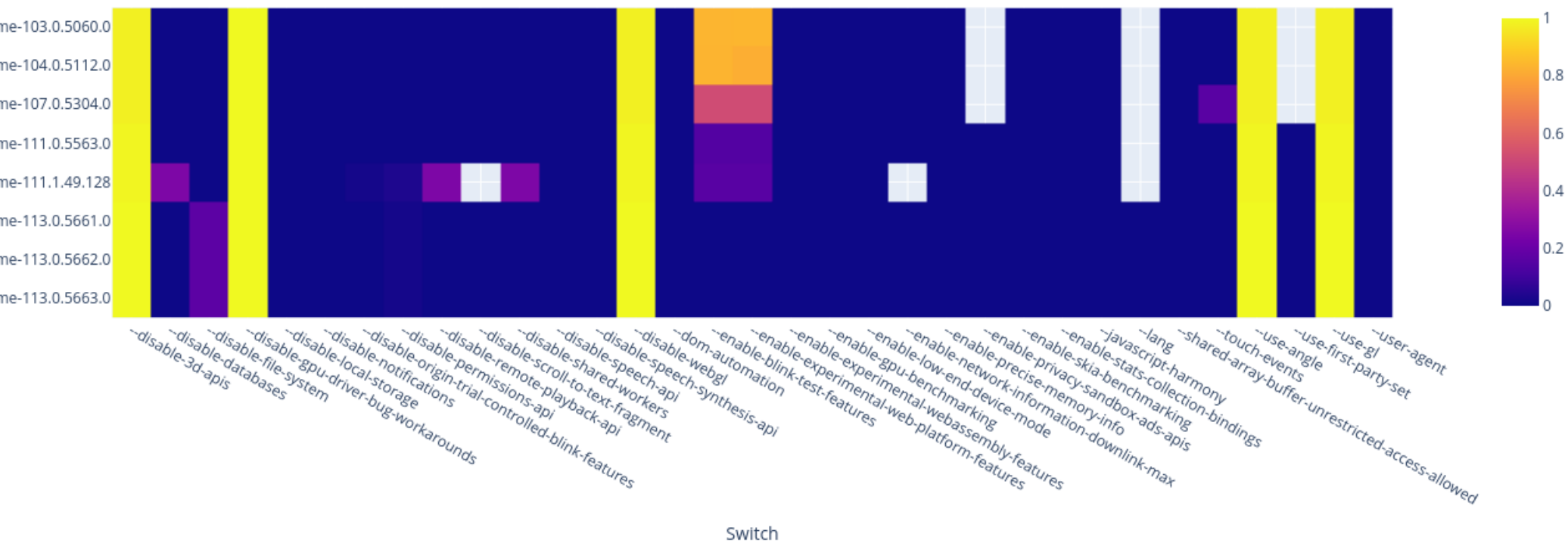


Table 1. Analysis of the effects of switches on the BOM and the FingerprintJS, including the generation of browser fingerprints, the frequency of timeouts, and the attribute density for specific browser versions

Browser Version	Number of switches that impact the BOM	Number of switches that impact FingerprintJS	Fingerprint Generated	Timeout during the test	Attributes per Fingerprint	Attributes on reference Fingerprint
109.0.5413.0	48	17	1477	4	12875-15750	14924
110.0.5476.0	49	18	1478	5	12871-15829	14963
111.0.5563.0	47	17	1478	4	12895-15903	15015
112.0.5607.0	47	17	1478	4	12895-15870	15015
113.0.5669.0	47	17	1478	4	12894-15888	15014
114.0.5731.0	47	17	1478	4	12910-15938	15059
115.0.5788.0	47	17	1478	4	13014-16018	15163
Headless109.0.5413.0	36	13	1491	4	12315-15242	14403
Headless110.0.5476.0	38	14	1491	5	12311-15321	14364
Headless111.0.5563.0	37	13	1492	4	12335-15395	14364
Headless112.0.5607.0	37	13	1492	4	12335-15362	14455
Headless113.0.5669.0	38	14	1492	4	12334-15380	14545
Headless114.0.5731.0	38	14	1492	4	12396-15430	14454
Headless115.0.5788.0	38	14	1492	4	12500-15510	14924

FP-Rainbow : Detect configurations

Use cases ?



What about Android ?

PhD Sihem BOUHENNICHE (2024-2027)

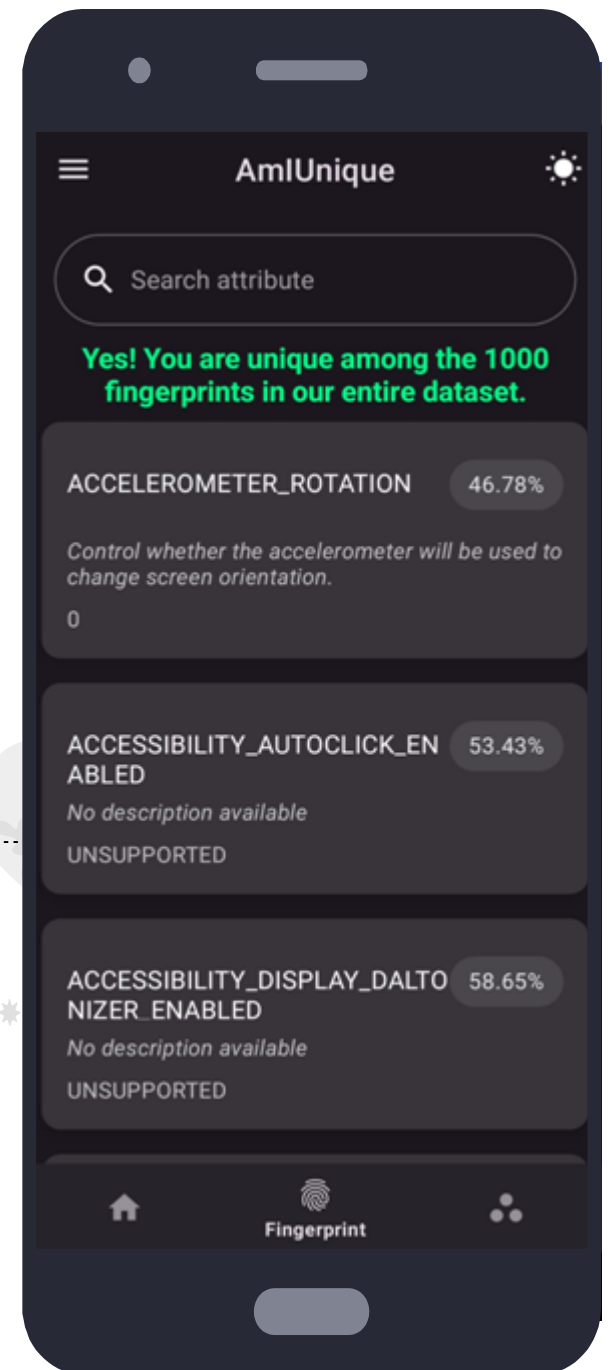
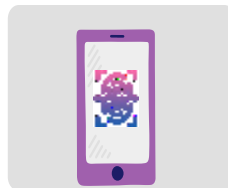
Semi-exhaustive method
to retrieve zero-permission
attributes



Querying Android API
using different techniques
like Java Reflection



~280 native attributes
retrievable by an Android
application



More alternatives to tracking cookies?

A Survey on Web Tracking: Mechanisms, Implications, and Defenses

Tomasz Bujlow, *Member, IEEE*, Valentín Carela-Español, Josep Solé-Pareta, and Pere Barlet-Ros

II	Session-only tracking mechanisms	4
II-A	Session identifiers stored in hidden fields	4
II-B	Explicit web-form authentication	
II-C	window.name Document Object Model	
III	Storage-based tracking	11
III-A	HTTP cookies	
III-A1	Explicit web-form authentication	
III-A2	Cookie leaks / syncing	
III-A3	Advertising networks	
III-B	Flash cookies and Java	
III-C	Flash LocalConnection	
III-D	Silverlight Isolated Storage	
III-E	HTML5 Global, Local, and	
III-F	Web SQL Database and IndexedDB	
III-G	Internet Explorer userData	
IV	Cache-based tracking	11
IV-A	Web cache	
IV-A1	Embedding identifiers in cache	
IV-A2	Loading performance test	
IV-A3	ETags and Last-Modified	
IV-B	DNS cache	
IV-C	Operational caches	
IV-C1	HTTP 301 redirect cache	
IV-C2	HTTP authentication cache	
IV-C3	HTTP Strict Transport Security	
IV-C4	HTTP Strict Transport Security	
V	Fingerprinting	11
V-A	Network and location fingerprinting	11
V-B	Device fingerprinting	
V-C	Operating System instance fingerprinting	
V-D	Browser version fingerprinting	
V-E	Browser instance fingerprinting	
V-F	Browser instance fingerprinting	
V-G	Other browser instance fingerprinting	
V-G1	Panoptlick project	
VI	Other tracking mechanisms	11
VI-A	Headers attached to outgoing requests	
VI-B	Using telephone metadata	
VI-C	Timing attacks	
VI-D	Using unconscious collaboration	
VI-E	Clickjacking	
VI-F	Evercookies (supercookies)	
VII	Identification of the tracked	11
VII-A	Legitimate first-party services	
VII-B	Leaking information to third parties	
VII-C	Selling information to third parties	
VII-D	Using web hacks	
VII-E	Intended deanonymization	
VIII	Purposes and implications of tracking	17
VIII-A	User-oriented search	17
VIII-B	Online advertising	18
VIII-C	Web analytics and usability tests	18
VIII-D	Assessing financial credibility	18
VIII-E	Price discrimination	18
VIII-F	Determining the insurance coverage	19
VIII-G	Impact on the job market	19
VIII-H	Government surveillance	19
VIII-I	Identity theft	19
VIII-J	Third-party tracking	20
IX	Methods and tools for avoiding and auditing tracking	20
IX-A	Blocking advertisement services	20
IX-B	Hiding the IP address	22
IX-C	Modification of data sent over the network	23
IX-D	Opt-out cookies	23
IX-E	Do Not Track	23
IX-F	Using privacy-focused search engines	24
IX-G	Private browsing mode	24
IX-H	Clearing the browser cache and history	24
IX-I	Execution blocking	24
IX-J	E-mail aliases	24
IX-K	Self-destroying file systems	24
IX-L	Discovering how tracking works	24
IX-M	Combined tools	26

IPV4 ADDRESS



Eric Lawrence 🎻

@ericlaw

Follow



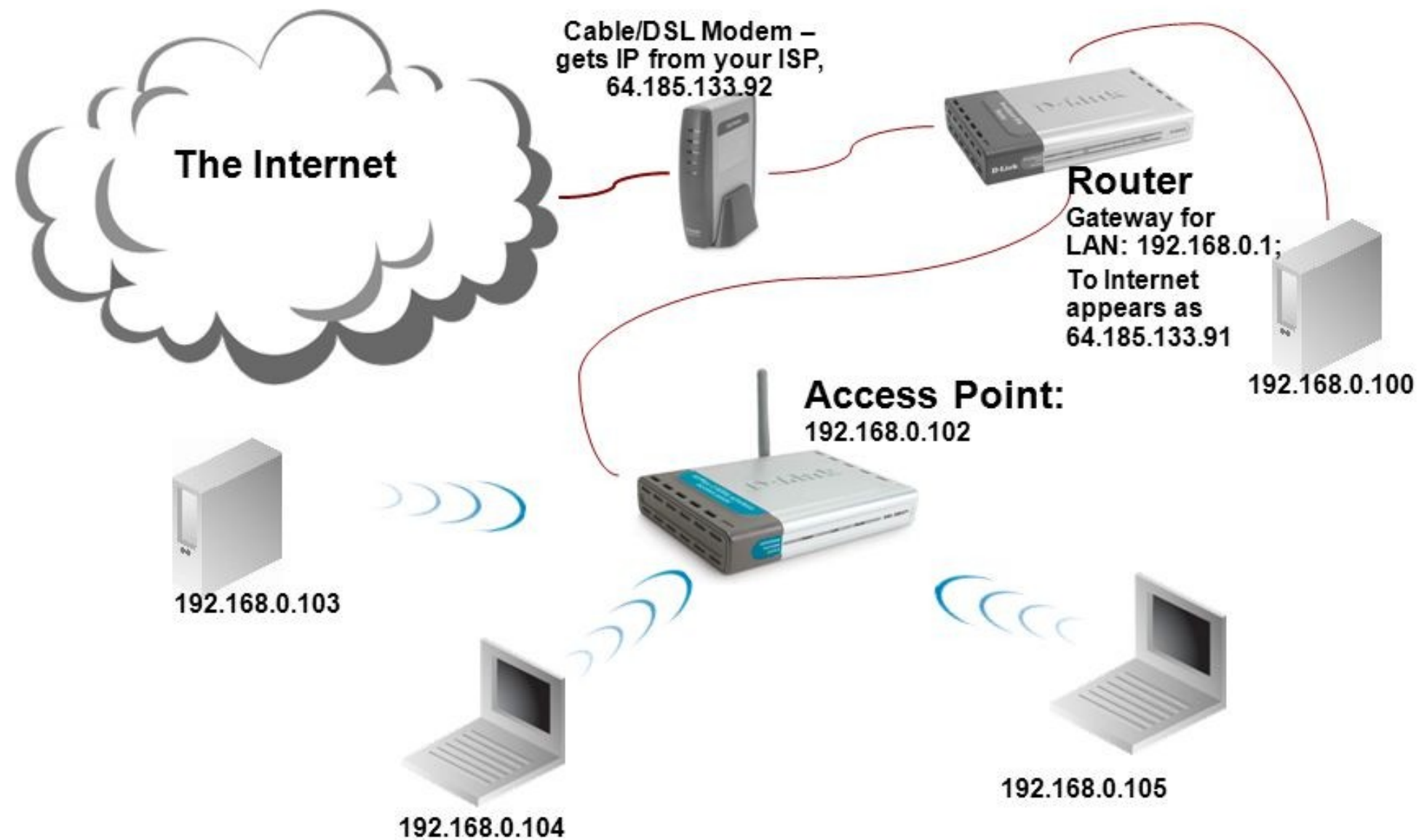
If your browser leaks your IP address, the rest of its "Privacy" features are of limited value.

9:20 AM - 16 Jun 2019

IPv4 Addresses

- IPv4 (32 bit)
 - 193.51.236.124
 - 4 billion, distributed across regions
- Types
 - Static Address
 - Dynamic Address
- DHCP (Dynamic Host Configuration Protocol)
 - DHCP leases
 - Duration of 'lease' negotiated, then address released back to server

Home Network- IP Addresses



DHCP Lease Cycle

Renewal T1
50% of Lease time



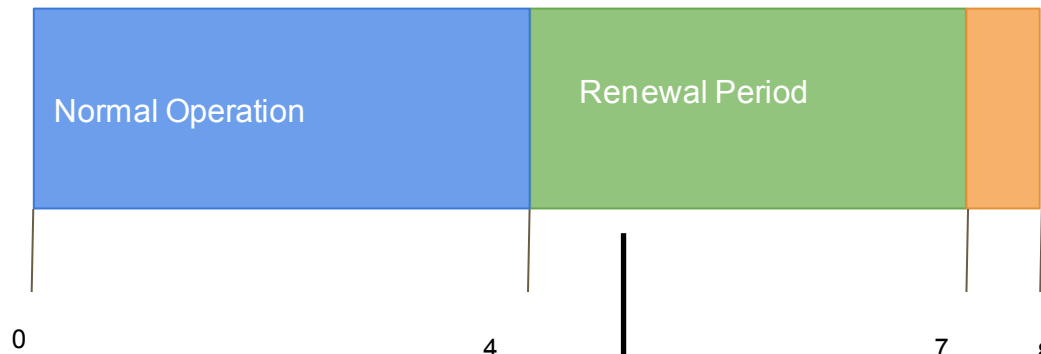
4 days



7 days

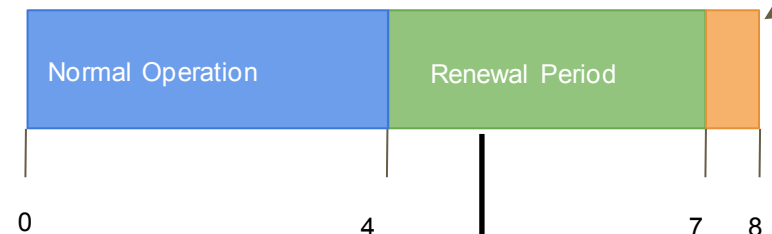
Rebinding Period

193.51.236.124



Renewal
No IP change

193.51.236.124



Renewal
No IP change

193.51.236.124

IP transitions: Antoine

193.51.236.160 -> 77.128.127.191, 2018-05-24 08:00:00
77.128.127.191 -> 193.51.236.160, 2018-06-01 13:00:00
193.51.236.160 -> 77.128.127.191, 2018-06-10 17:00:00
77.128.127.191 -> 193.51.236.160, 2018-06-15 09:00:00
193.51.236.160 -> 77.128.127.191, 2018-07-05 15:00:00
77.128.127.191 -> 193.51.236.160, 2018-07-10 13:00:00

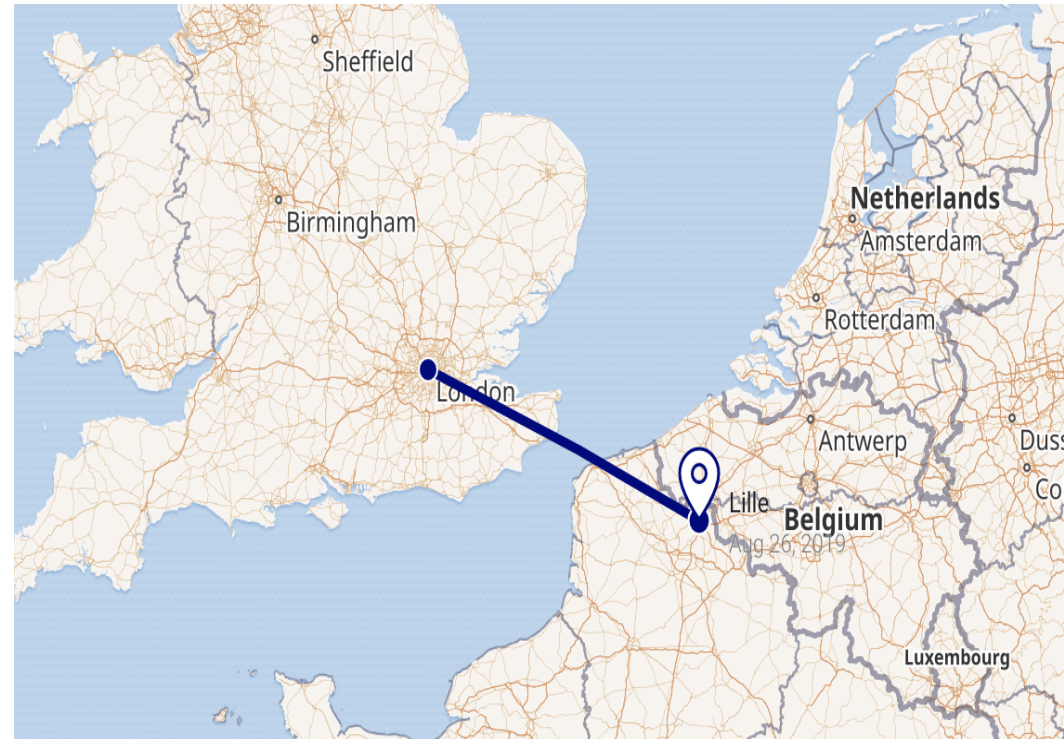
Lille

79.173.173.4 -> 109.145.107.134, 2018-07-12 20:00:00
109.145.107.134 -> 79.173.173.4, 2018-07-16 11:00:00
79.173.173.4 -> 109.145.107.134, 2018-07-30 18:00:00
109.145.107.134 -> 79.173.173.4, 2018-07-31 09:00:00
79.173.173.4 -> 86.190.33.234, 2018-08-11 16:00:00
86.190.33.234 -> 79.173.173.4, 2018-08-13 12:00:00

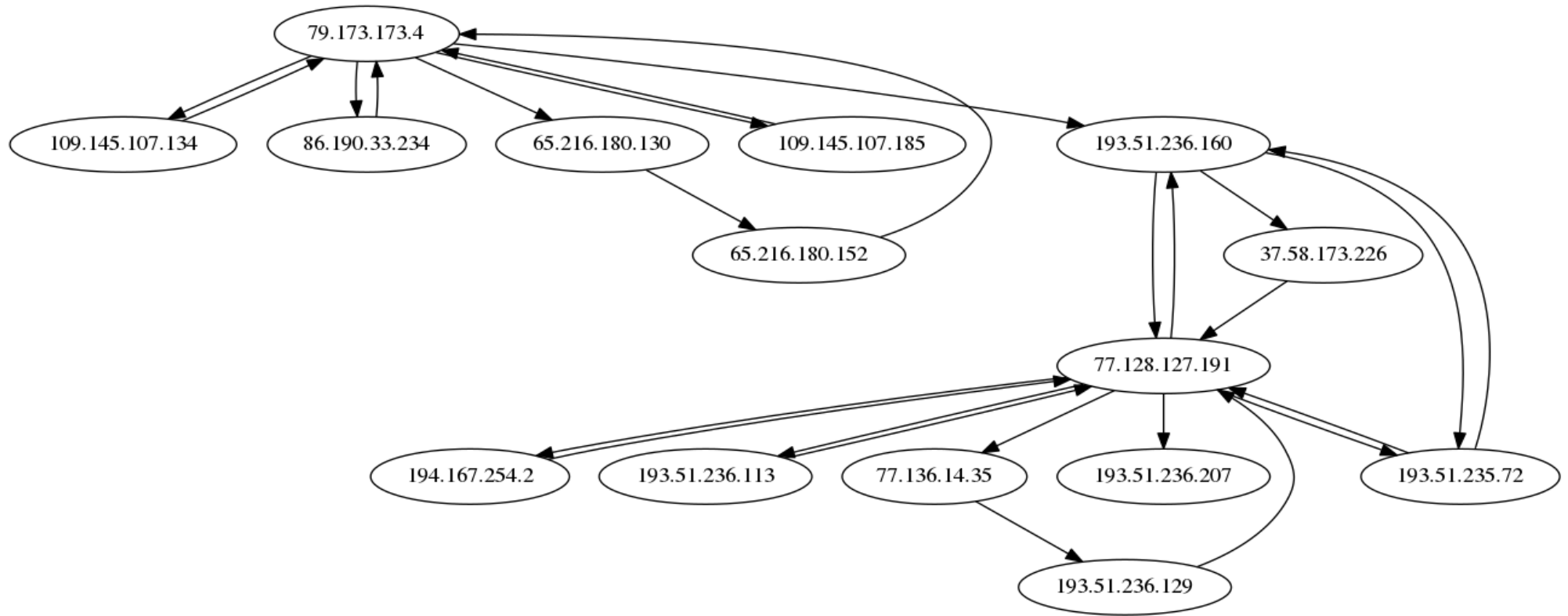
London

193.51.236.160 -> 77.128.127.191, 2018-09-06 17:00:00
77.128.127.191 -> 193.51.236.160, 2018-09-07 13:00:00
193.51.236.160 -> 77.128.127.191, 2018-09-07 17:00:00
77.128.127.191 -> 194.167.254.2, 2018-09-12 13:00:00
194.167.254.2 -> 77.128.127.191, 2018-09-12 18:00:00
77.128.127.191 -> 193.51.236.160, 2018-09-13 13:00:00

Lille



Antoine's IP transitions [WWW'2018]



Some repeated addresses

Some long-lived addresses

Cookie reconstruction ?

Don't Count Me Out: On the Relevance of IP Address in the Tracking Ecosystem

Vikas Mishra
Inria / Univ. Lille
vikas.mishra@inria.fr

Pierre Laperdrix
CNRS / Univ. Lille / Inria
pierre.laperdrix@univ-lille.fr

Antoine Vastel
Univ. Lille / Inria
antoine.vastel@inria.fr

Walter Rudametkin
Univ. Lille / Inria
walter.rudametkin@univ-lille.fr

Romain Rouvov
Univ. Lille / Inria / IUF
romain.rouvov@univ-lille.fr

Martin Lopatka
Mozilla
mlopatka@mozilla.com

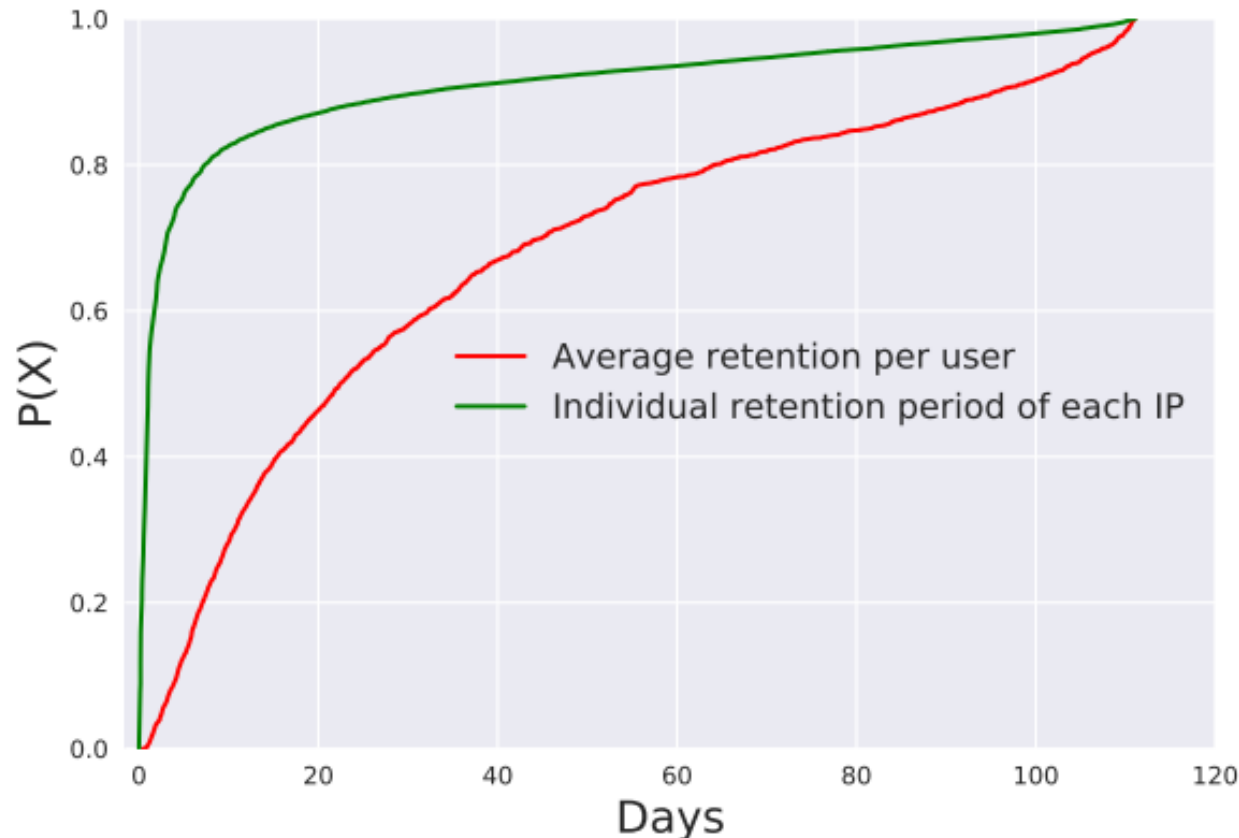
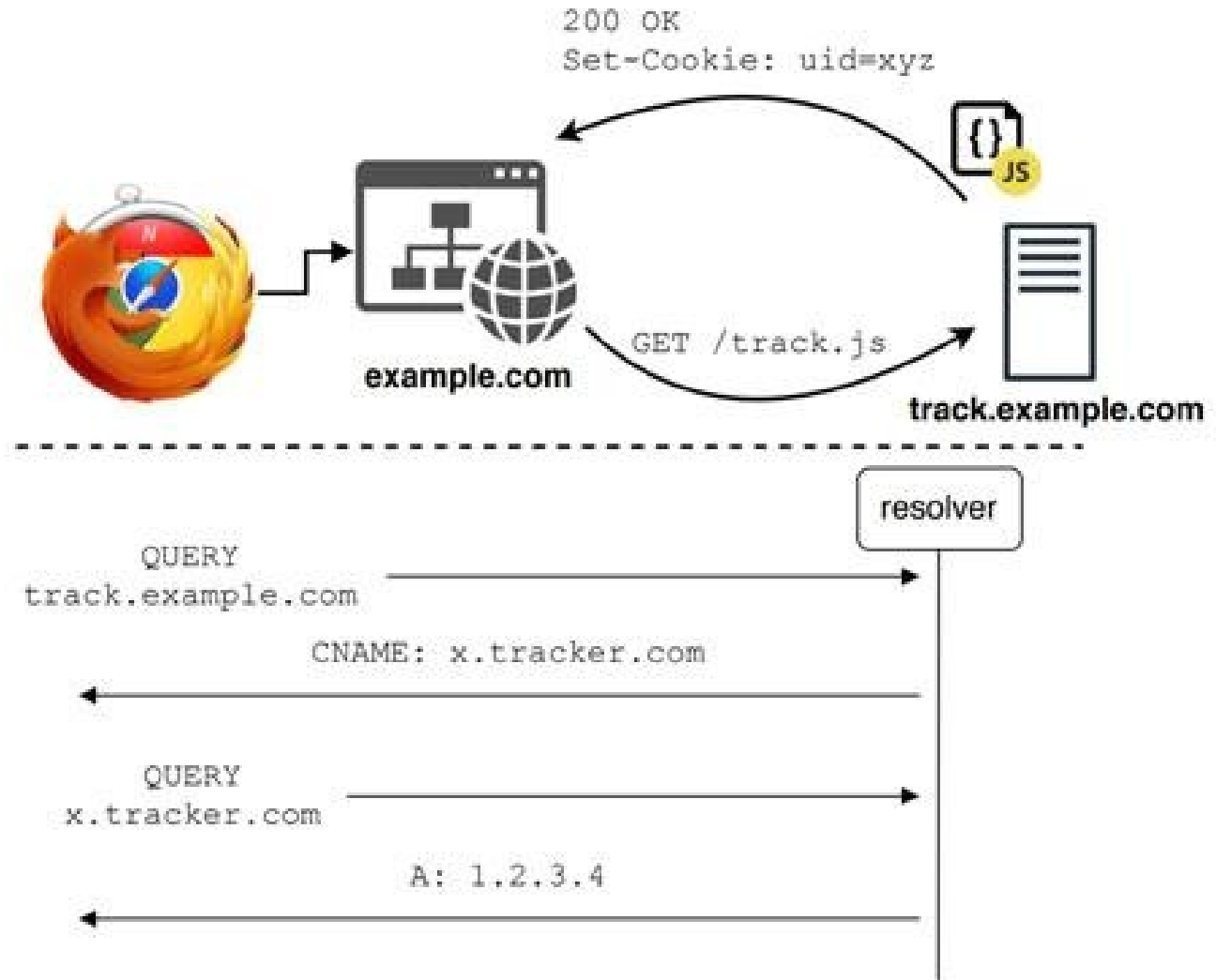


Figure 4: Retention period of IP addresses. The green line

CNAME Cloaking

DNS-based Tracking Evasion



But... Same Origin Policy ?

Device Linking: Personal Identifiable Information

Did you leak anything ?

Login

Email

Telephone

Name

Location

Home address

ZIP Code

Date of birth

IP Address

...



Client-side vs Server-side tracking

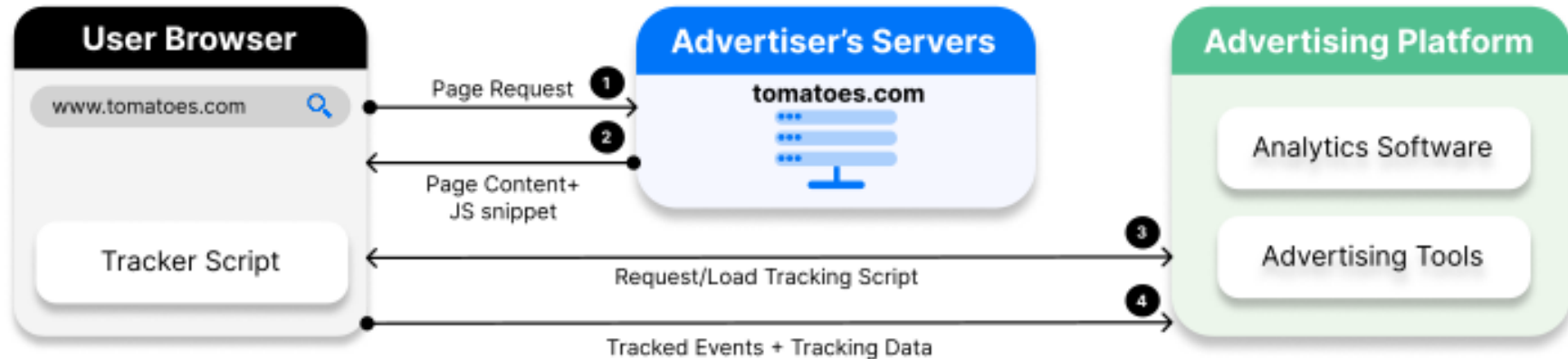


Figure 1: Client-side tracking.

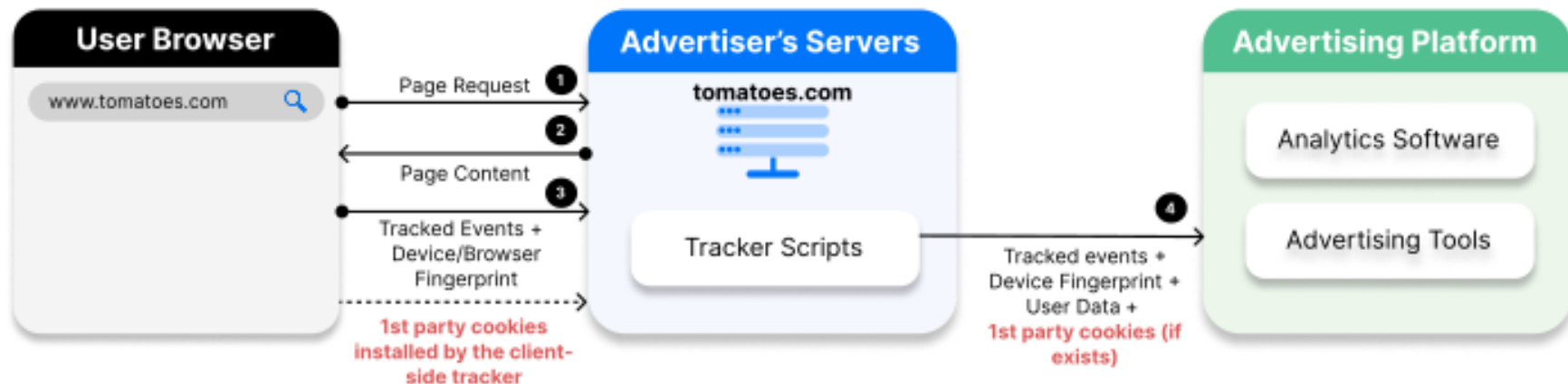


Figure 2: Server-side tracking.

Server-side tracking

- Two main SST technologies
 - Google Tag Manager (GTM)
 - Meta Conversions API (CAPI)
 - Often combined with Meta Pixel (client-side)
 - Uses Personal Identifiable Information to retarget you !!!

How effective is Meta at tracking?

[PETS 2024]

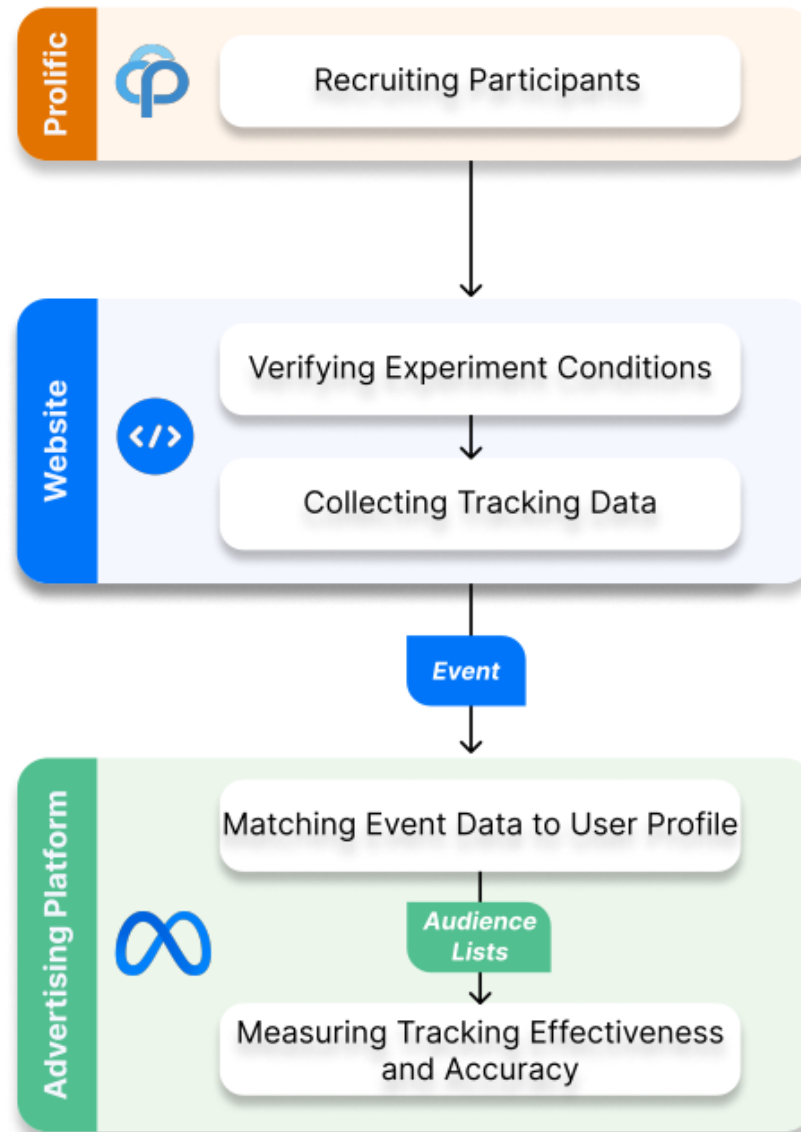


Figure 3: General workflow of experiments.

How effective is Meta at tracking?

[PETS 2024]

Table 7: Examples of cookies installed by Meta’s client-side trackers used for advertising.

Cookies	Domain	Content
c_user	Facebook	The Facebook account ID of the user currently connected on the browser. Lifespan of 1 year.
fr	Facebook	Stores Facebook account details. Used for ad delivery and improving ads relevancy. Lifespan of 90 days.
sb	Facebook	Used to store browser details. Lifespan of 373 days.
_fbp	First-party	Facebook Browser ID. A unique ID saved under the website domain when the user first visits the website. Used for advertising and site analytics. Lifespan of 90 days.
_fbclid	First-party	Facebook Click ID. When users click on ads on Facebook, the link includes a "fbclid" query parameter. When users land on the target website, the Meta Pixel automatically adds this query parameter to the _fbclid cookie. It is used to report actions, such as purchases, generated through Facebook ads. Lifespan of 90 days.
usid	Facebook	A session cookie that collects a combination of the user’s browser and unique identifiers. Used for tailored advertising.
oo	Facebook	Used for Facebook advertisement and behavioral targeting. Lifespan of 5 years.

How effective is Meta at tracking?

[PETS 2024]

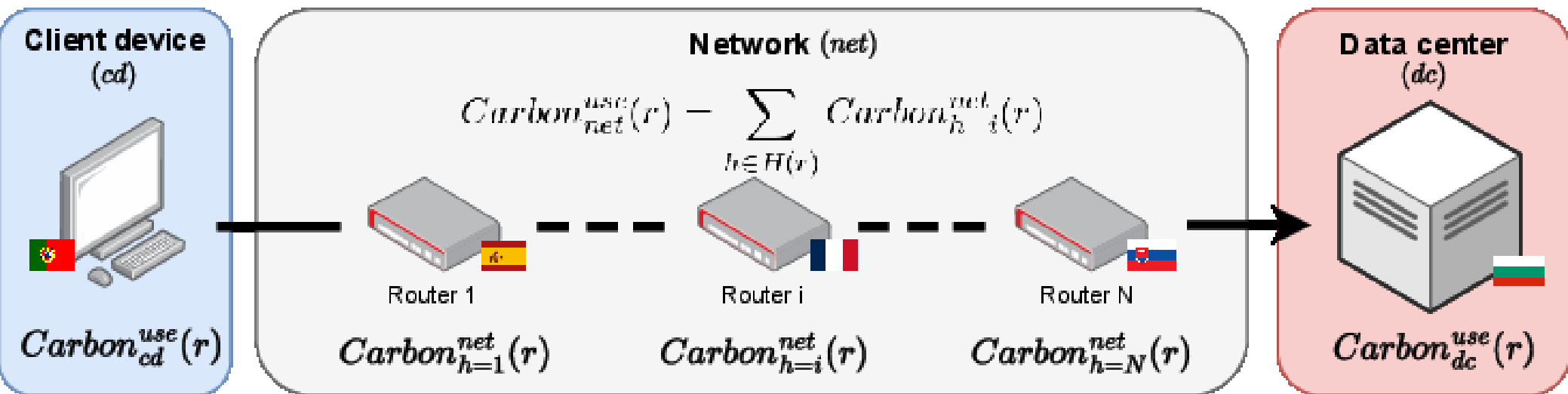
Experiment	#	fbp	Users	Pixel	CAPI	Overlap
Desktop: Chrome	1	No	400	46%	51%	43%
	2	No	1000	43%	44%	48%
	3	Yes	500	44%	45%	73%
	4	Yes	500	42%	45%	74%
Mobile: Chrome	1	No	300	61%	34%	50%
	2	No	250	N/A	51%	N/A
	3	No	175	N/A	50%	N/A

- N/A indicates that the tracker was not implemented on the website.

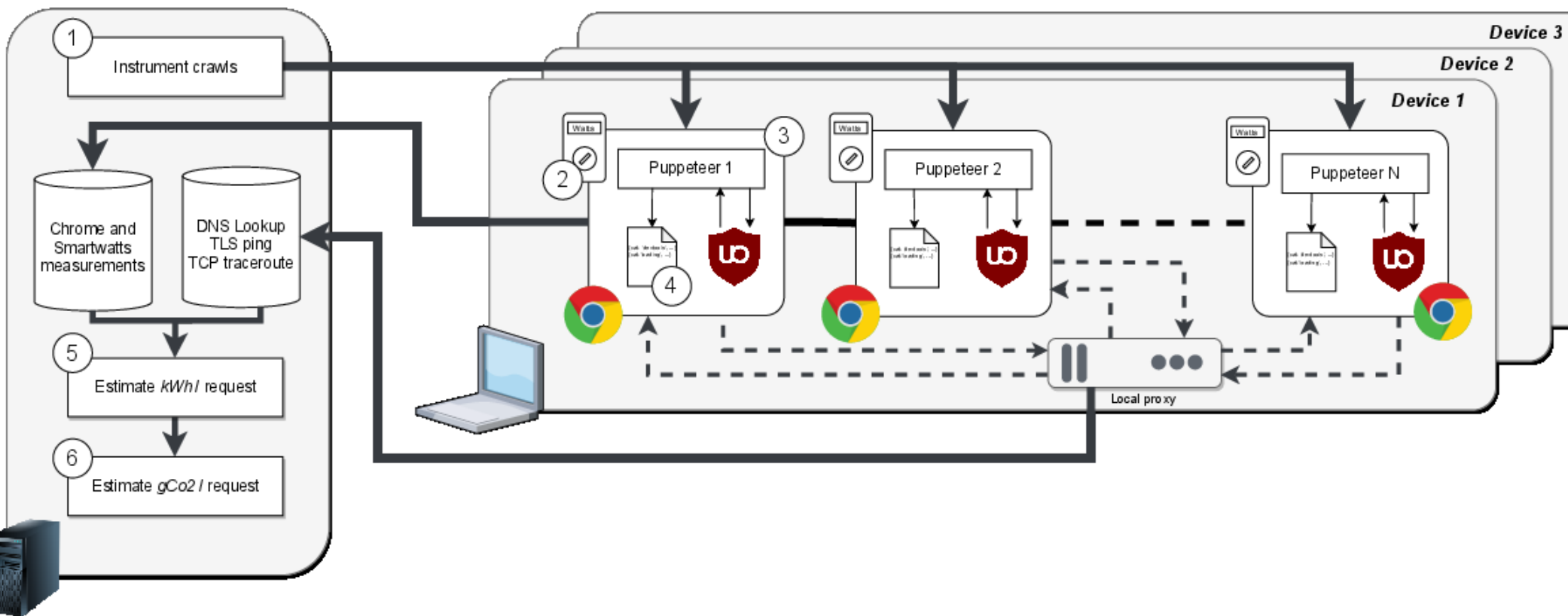
Implementation	Events	Reach	Accuracy	Overlap
Pixel 1	1722	61.09%	81.45%	52%
CAPI 1	1833	19.74%	64.80%	
Pixel 2 (isolated)	812	63.30%	100%	58%
CAPI 3 (isolated)	958	16.32%	59.6%	

MORE FUN STUFF !!!

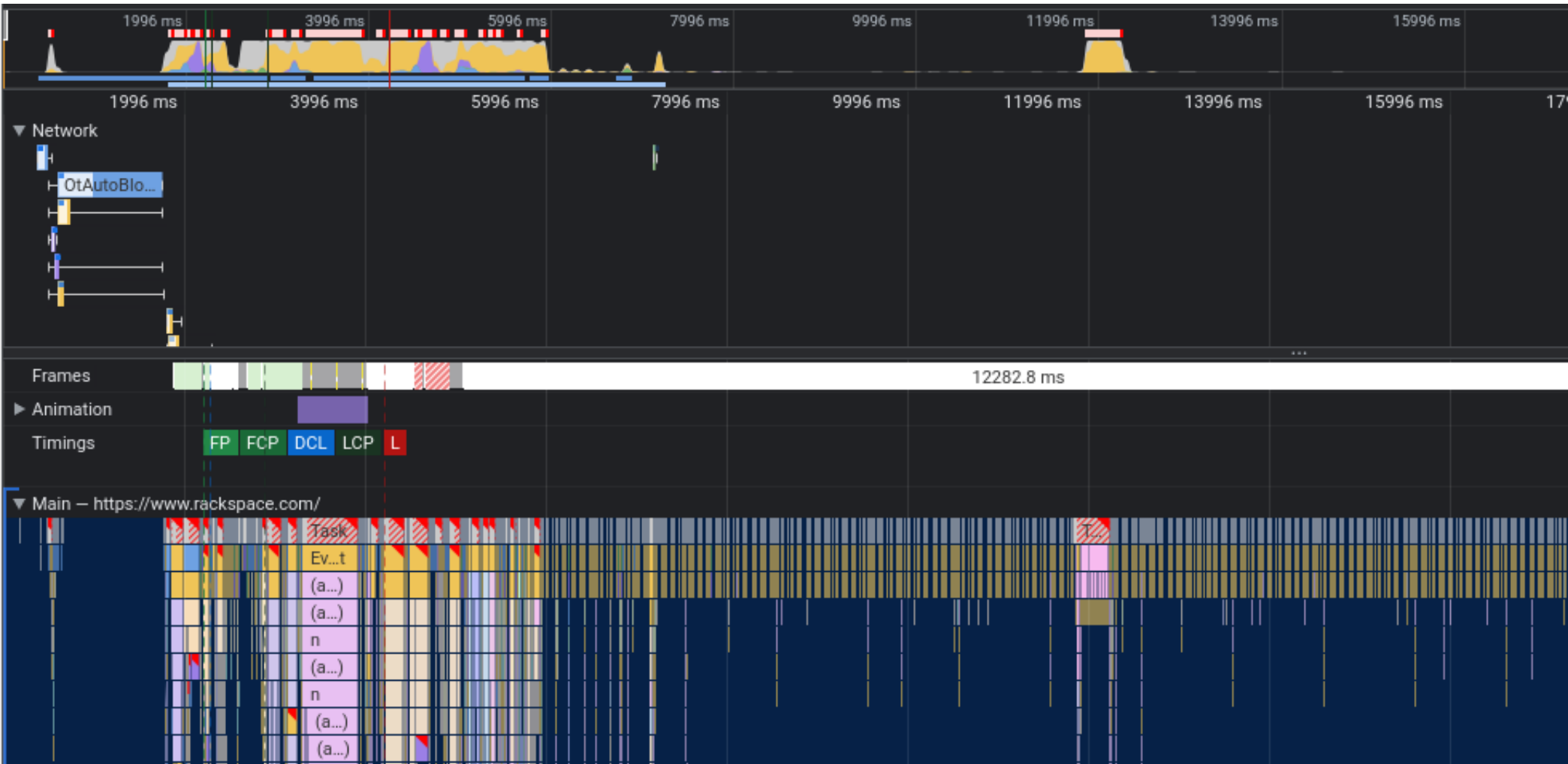
Ad-Carbon : End-to-End analysis of the Carbon Footprint of Advertising



Ad-Carbon : End-to-End analysis of the Carbon Footprint of Advertising



Ad-Carbon : End-to-End analysis of the Carbon Footprint of Advertising



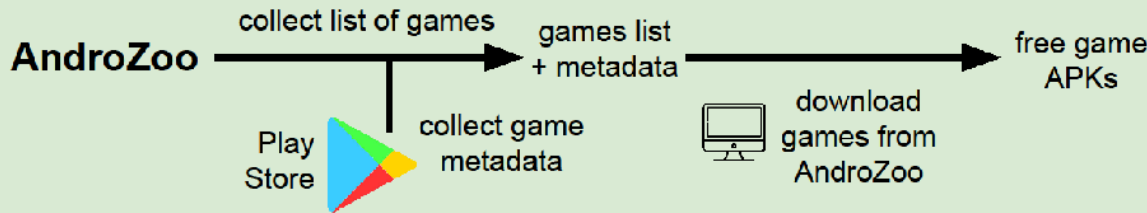
Ad-Carbon : End-to-End analysis of the Carbon Footprint of Advertising

We crawled 10,562 domains

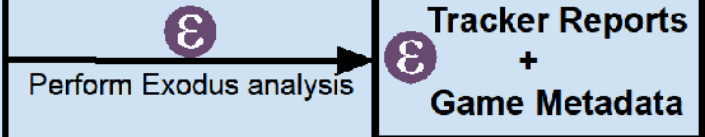
- 31,394 pages
- Advertising increases carbon emissions by **144%** for our crawls
- Accepting cookie banners
 - 90% carbon footprint increase
 - 38.8% more advertising requests

Do free Android games track you more ?

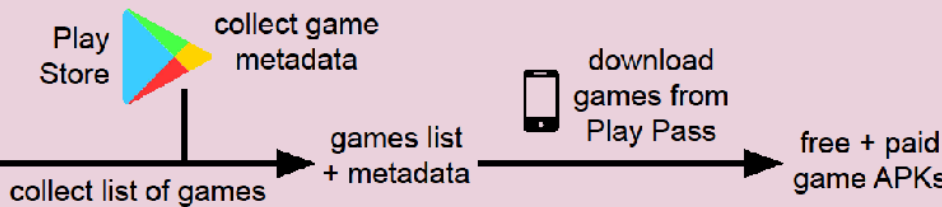
① AndroZoo games collection



③ Exodus analysis



② Play Pass games collection



The Price to Play: a Privacy Analysis of Free and Paid Games in the Android Ecosystem

Pierre Laperdrix, Naif Mehanna, Antonin Durey, Walter Rudametkin

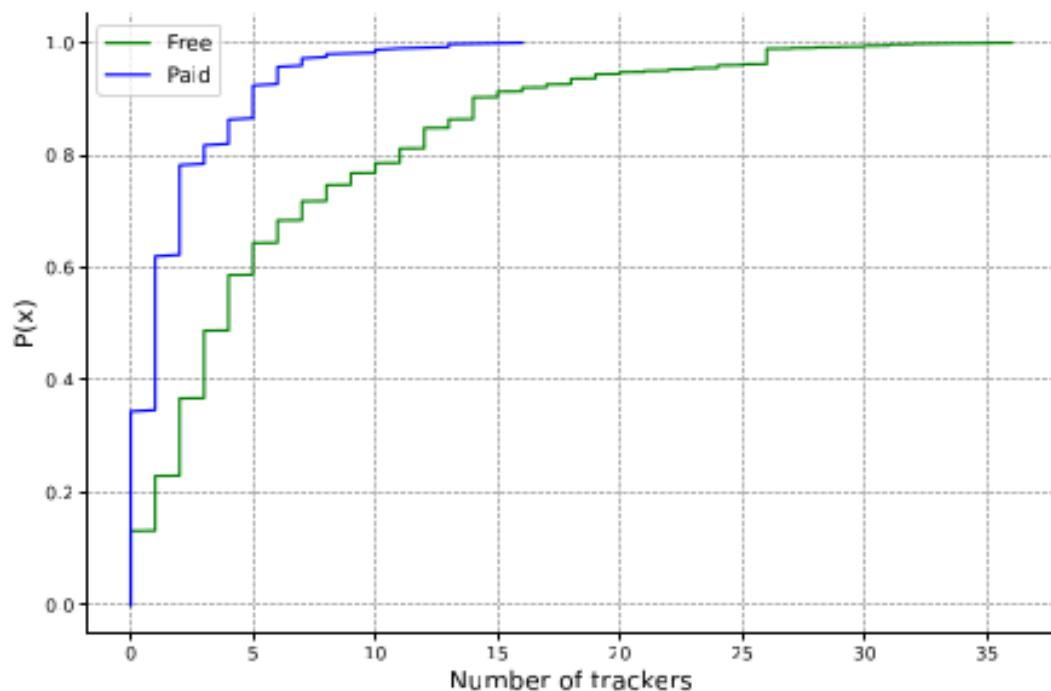


Table 1: Source of the games present in our dataset.

	Free games	Paid games	Total
AndroZoo	6,035	0	6,035
Play Pass	320	396	716
Total	6,355	396	6,751

Table 2: Overview of the presence of trackers in the games of our dataset

	Percentage of games with trackers	Average number of trackers per game	Standard deviation
Free	86.79%	6.11	6.65
Paid	65.31%	1.80	2.38

Figure 2: Distribution of trackers across free and paid games.

**CONCLUDING REMARKS
(ONLY 5 MORE SLIDES)**

Surveillance Capitalism



The Age of Surveillance Capitalism. S. Zuboff⁷⁸

Surveillance Capitalism



Adtech 2024

~\$800B USD



Areas of academic interest

Large scale monitoring

Identifying and repairing new privacy risks

Recommender systems

- **AI Bias, Influence, Manipulation, Targeting**

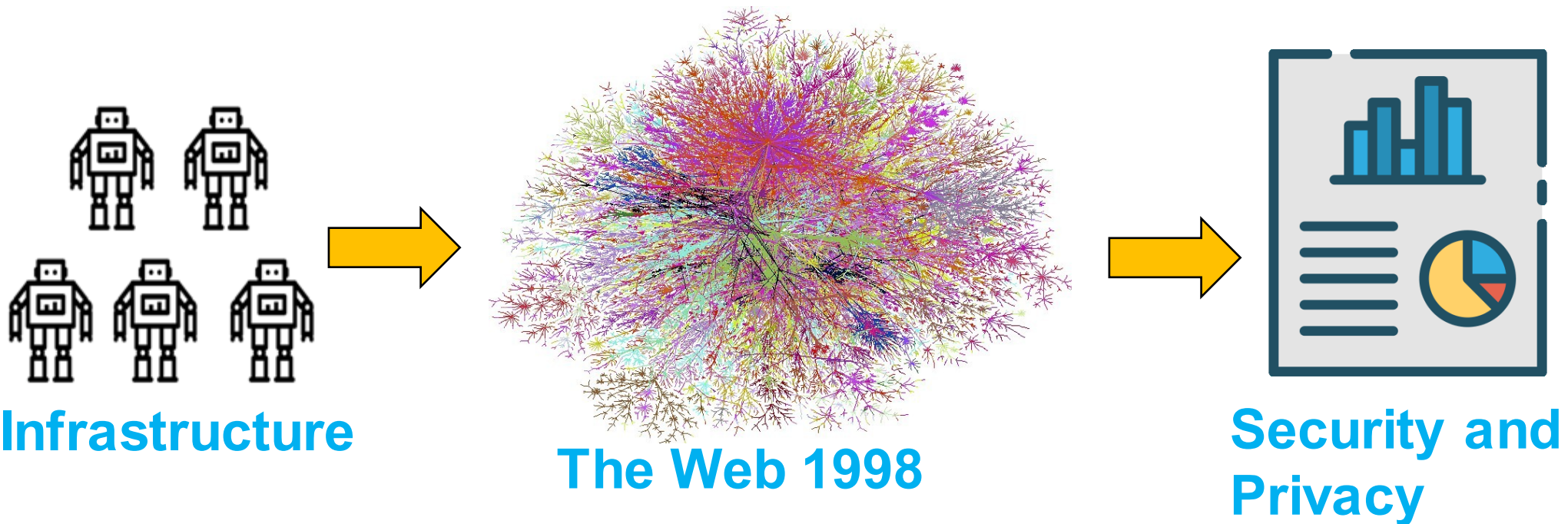
Legislation to protect privacy

- **GDPR, ePrivacy, Digital Markets Act**

Societal impact of omnipresent surveillance

- **Social cooling**

Large scale Web monitoring



How do we study complex systems ?

[DIMVA 21]

[MADWeb 20]

[WWW 20a]

[WWW 20b]

moz://a

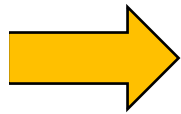


eye/o

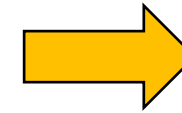
DATA DOME

New privacy risks

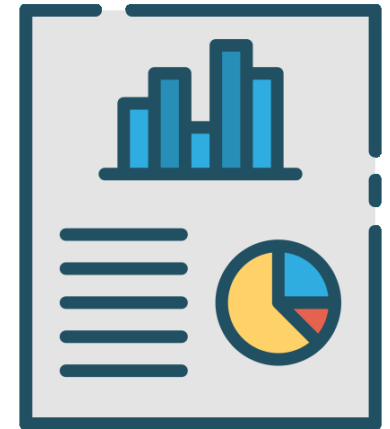
Battery of Tests



Web technologies



Models / Reports



How do we consistently identify new risks ?

[NDSS 22]
[Bug Bounty Mozilla]
[UsenixSec'18]
[PETS'21]

The end

Backup slides

Next Steps

Release the App
to collect real
world fingerprints



1

Run experiments
in controlled
environments on
different
emulators



The impact of
permissions on
fingerprinting



2

Longitudinal
study: the impact
of the Android
API evolution on
fingerprinting



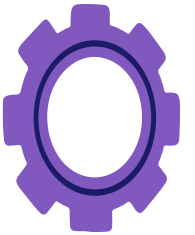
3

Other Directions

Device fingerprinting in the iOS ecosystem



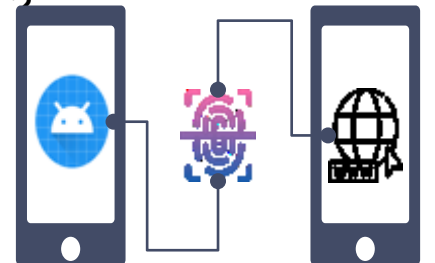
What is the impact of the device's settings on the fingerprint?



Do Android applications use device fingerprinting?

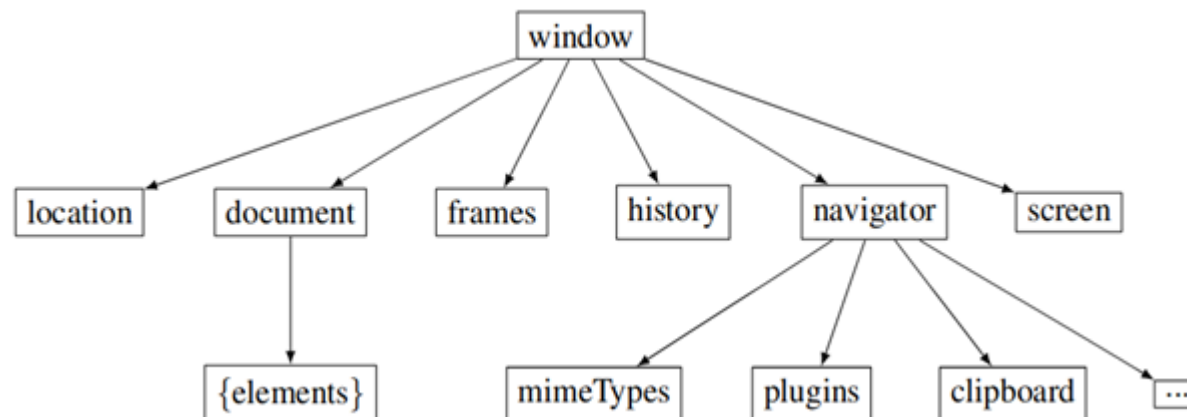
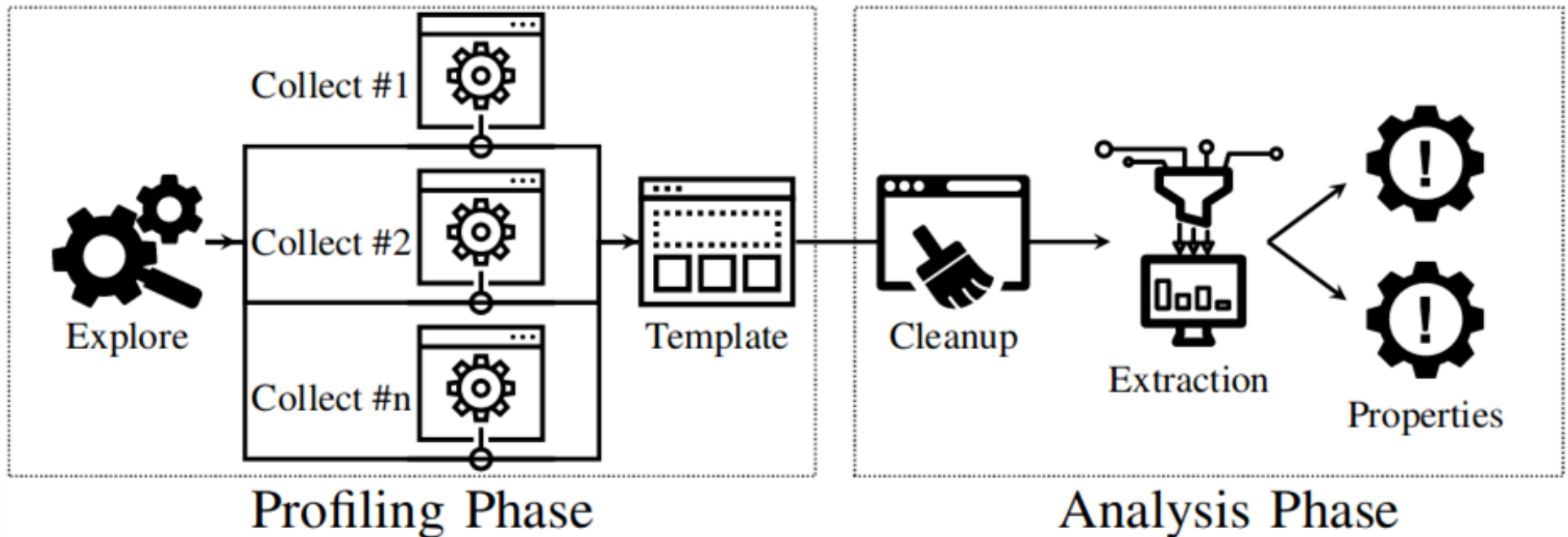


Can we link browser fingerprints and device fingerprints to the same device?



JavaScript Template Attacks [NDSS19] :

Systematic fingerprint attributes extraction (no function calls)



JavaScript Template Attacks [NDSS19] :

Results

Documented
vs. discovered
properties

Browser	MDN	JavaScript Template
Firefox	2247	15 709
Chrome	2698	13 570
Edge	1806	9666
Firefox Android	2104	15 612
Chrome Android	2676	13 119
Tor browser	2247 [†]	15 639

Some properties
are random or
duplicates

Browser	Exploration	Without duplicates	Usable
Firefox	18 443	16 450	15 709
Chrome	15 585	13 604	13 570
Edge	13 752	11 850	9666
Firefox Android	18 214	16 296	15 612
Chrome Android	15 556	13 608	13 119
Tor browser	17 217	15 645	15 639

“It will be very hard for people to watch or consume something that has not in some sense been tailored for them.”

Eric Schmidt, Google

DON'T WORRY, IT'S ONLY
MARKETERS COLLECTING
OUR PERSONAL DATA
SO THEY CAN CREATE
MORE RELEVANT
ADVERTISING FOR US.



TOM
FISH
BURNIE

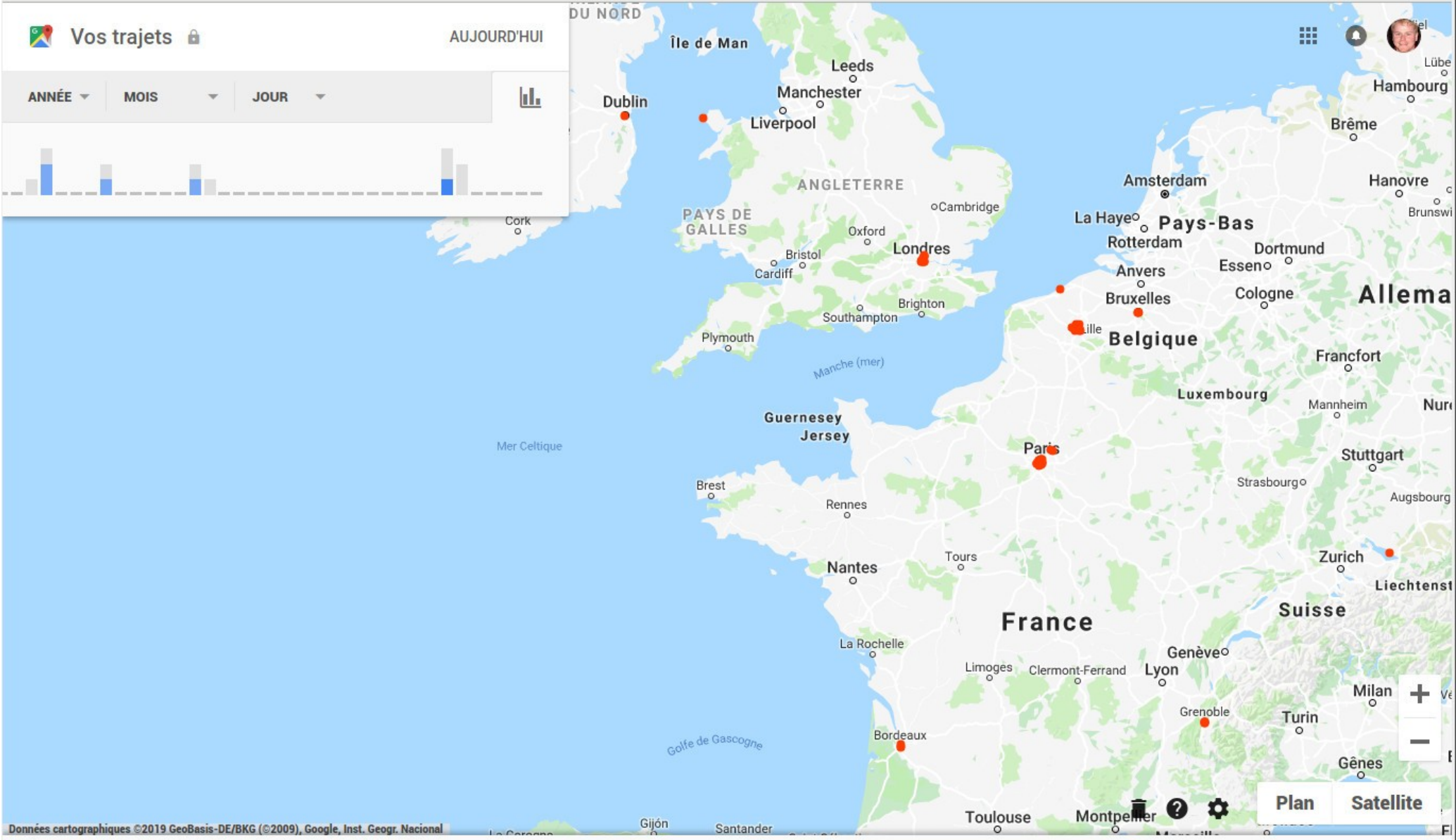
Exercise 1: Your location

- If activated, Google constantly stores your location and creates a timeline
- <https://www.google.com/maps/timeline?pb>
- Show it to your colleague!

Vos trajets

AUJOURD'HUI

ANNÉE MOIS JOUR



Données cartographiques ©2019 GeoBasis-DE/BKG (©2009), Google, Inst. Geogr. Nacional

111 lieux
Consultez les lieux que vous fréquentez le plus souvent et tous ceux dans lesquels vous vous êtes rendu grâce à l'historique des positions.

Bordeaux
26-29 juin 2018
AUTRES TRAJETS

L'historique des positions est activé
Votre position est mise à jour par votre appareil mobile, et vous seul pouvez y accéder.
GÉRER L'HISTORIQUE DES POSITIONS

Domicile et lieu de travail
Ajouter l'adresse de votre lieu de trav

Vos trajets

AUJOURD'HUI

2018 mai 25



Après-midi dans le lieu suivant :
Electronic Frontier Foundation
 Vendredi 25 Mai 2018

10,4 km
 2 h 46 min

Ajouter un lieu

À pied - 2,4 km

24 min

Lower Nob Hill

11:15 - 11:49

San Francisco, Californie, États-Unis

À pied - 700 m

6 min

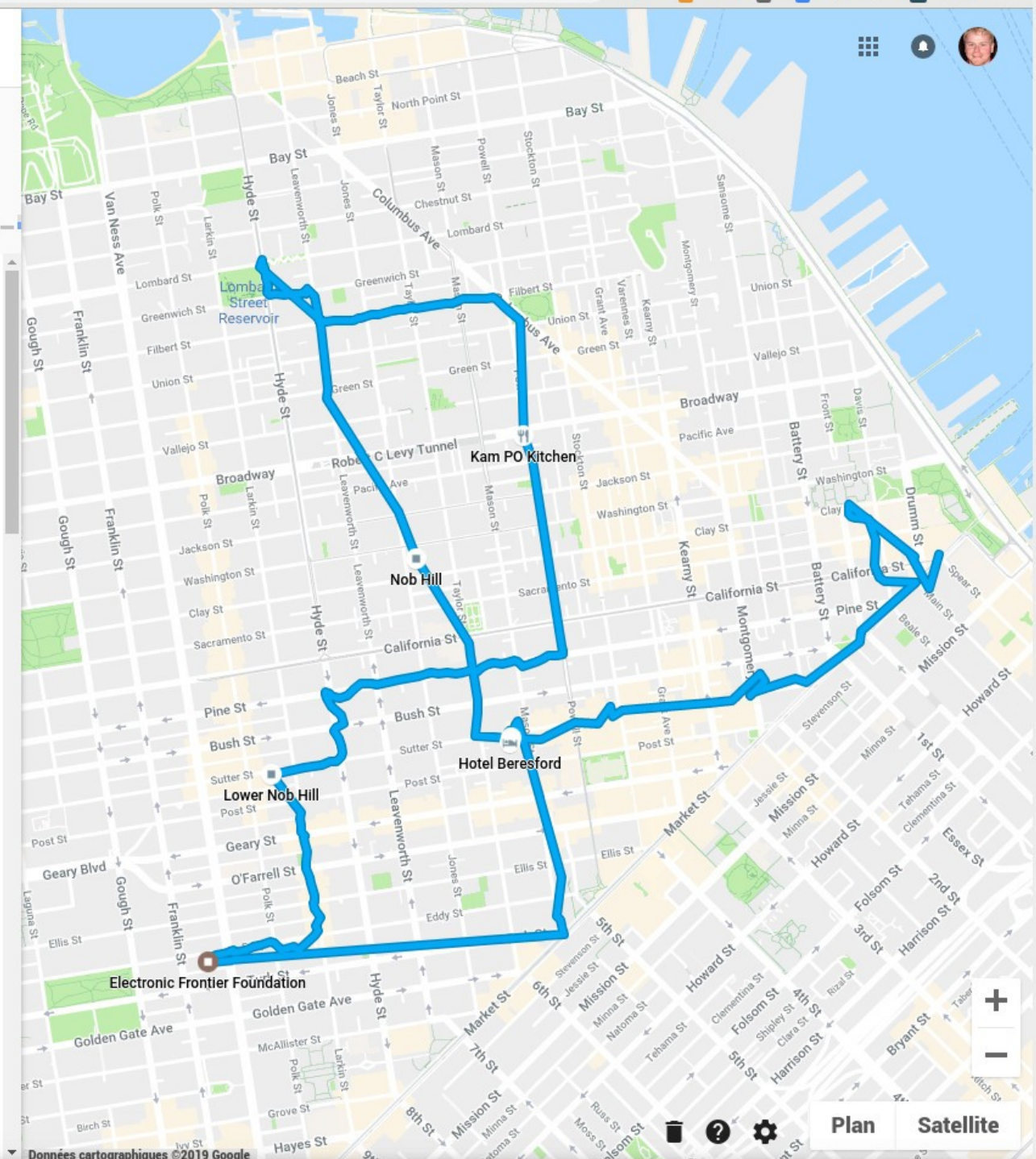
Nob Hill

11:55 - 12:12

San Francisco, Californie, États-Unis



+ 2 AUTRES



Plan Satellite

IP Address Dataset

- AmlUnique browser extensions for Chrome and Firefox
- June 6, 2019 - September 25, 2019
- 41,566 unique IP addresses from 5,443 browser instances
- Frequency 4 hours
- Desktop Devices
- IPv4 addresses

- IP sharing behind NAT
 - 1,046 users share at-least 1 IP with another user

- However, a set of IP addresses retained by a user over a long duration are highly discriminating
 - pair-wise comparison
 - 93% users had a unique set of long-lived IP Addresses

Uniqueness of IP Address

- IP sharing behind NAT
 - 1,046 users share at-least 1 IP with another user
- However, a set of IP addresses retained by a user over a long duration are highly discriminating
 - pair-wise comparison
 - Jaccard Similarity (J) = 0 for 99.97% of pairs of users.
 - 93% users had a unique set of long-lived IP Addresses