uc3m | Universidad Carlos III de Madrid

PRITIA-CLOUD

Auditing solutions for ensuring privacy and transparency in cloud services within the Cloud-Edge continuum

Type	National Research Project
Project Title	Next-generation intelligent services in the Cloud-Edge continuum. PRITIA-CLOUD: Privacy, Transparency, and AI-based Optimisation of Efficient Services for the Next Generation of the Cloud-Edge Continuum
Project Acronym	PRITIA-CLOUD

This work has been funded by the PRITIA-CLOUD project (TSI-063100-2022-0015), supported by the Ministry of Economic Affairs and Digital Transformation and the European Union – NextGenerationEU, as part of the Recovery, Transformation and Resilience Plan (PRTR) funds.

Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA ESPAÑA PUEDE

Fingerprinting from ads?: concept, viability and solutions

Miguel Bermejo
Universidad Carlos III de Madrid

Patricia Callejo
Universidad Carlos III de Madrid

Rubén Cuevas
Universidad Carlos III de Madrid

Ángel Cuevas
Universidad Carlos III de Madrid

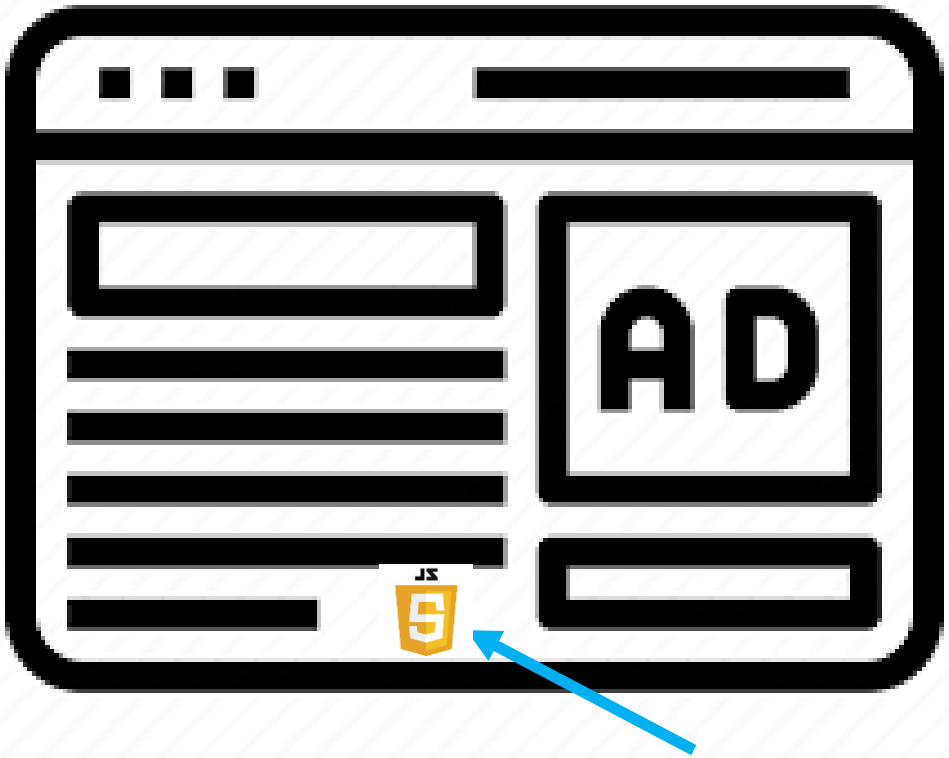
Background: web fingerprinting

Script to collect data

A script inserted in a website HTLM code allows to collect dozens of attributes from the device:

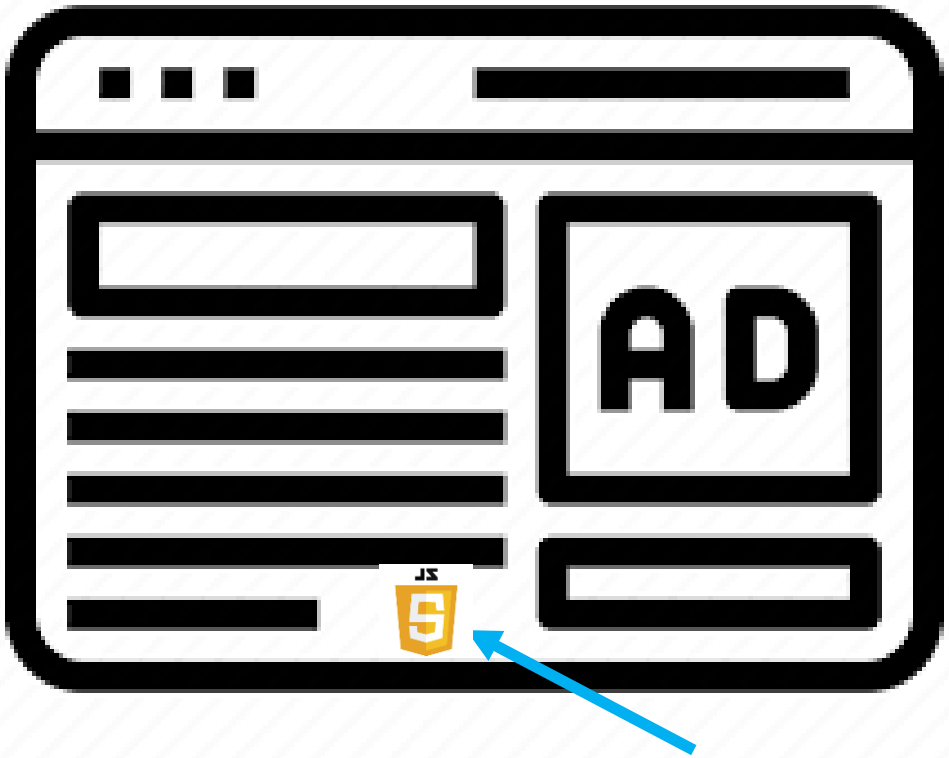
- **Browser version, OS, device**
- **Fonts, screen size, installed add-ons**
- **Graphic and Sound card parameters**
- **etc**

Background: web fingerprinting



Fingerprint = hash(attribute1+attribute2+…+attributeN)

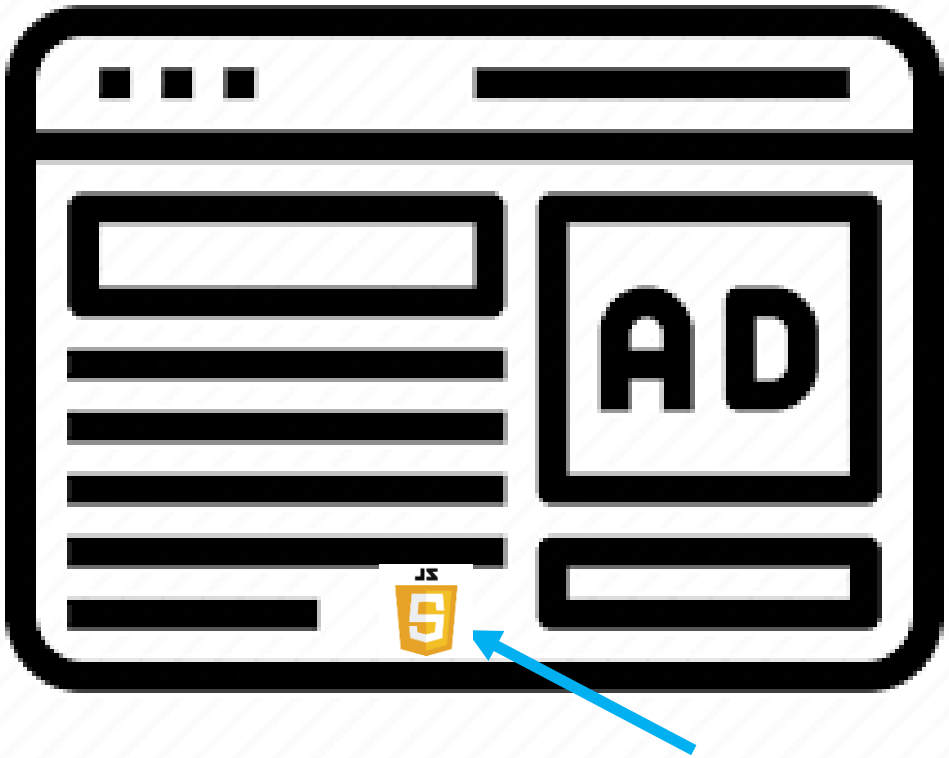
Background: web fingerprinting



Fingerprint = hash(attribute1+attribute2+…+attributeN)

If a sufficient number of parameters with enough discrimination power are collected the fingerprint might be unique in a large pool of devices (100Ks, 1Ms)

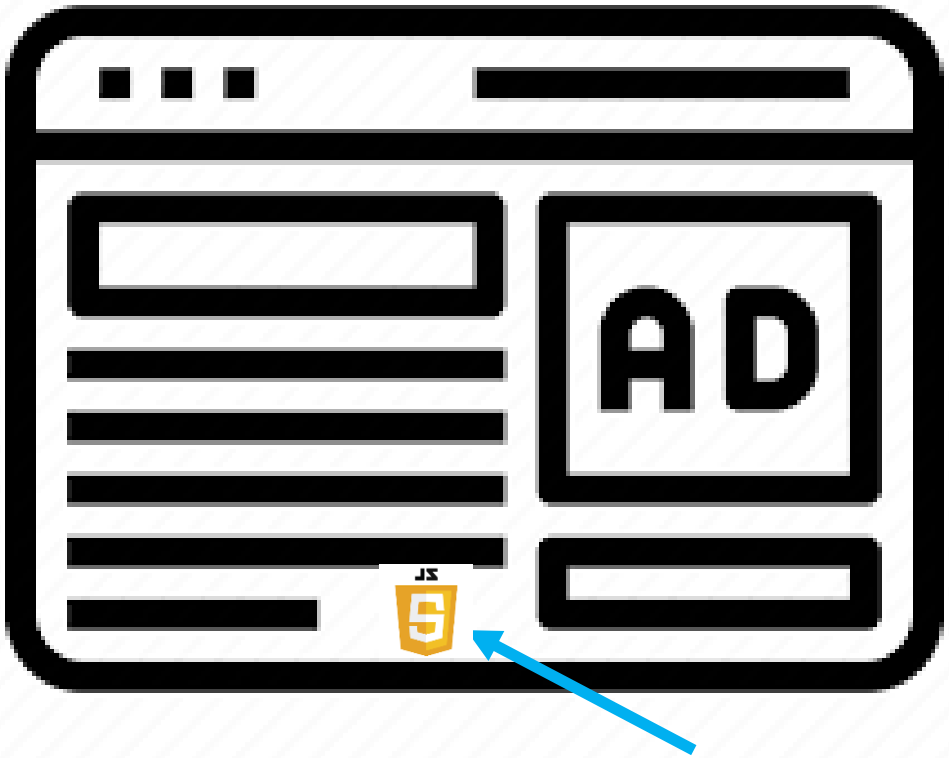
Background: web fingerprinting



Used for different purposes:

- **Tracking of users**
- **Retargeting**
- **etc**

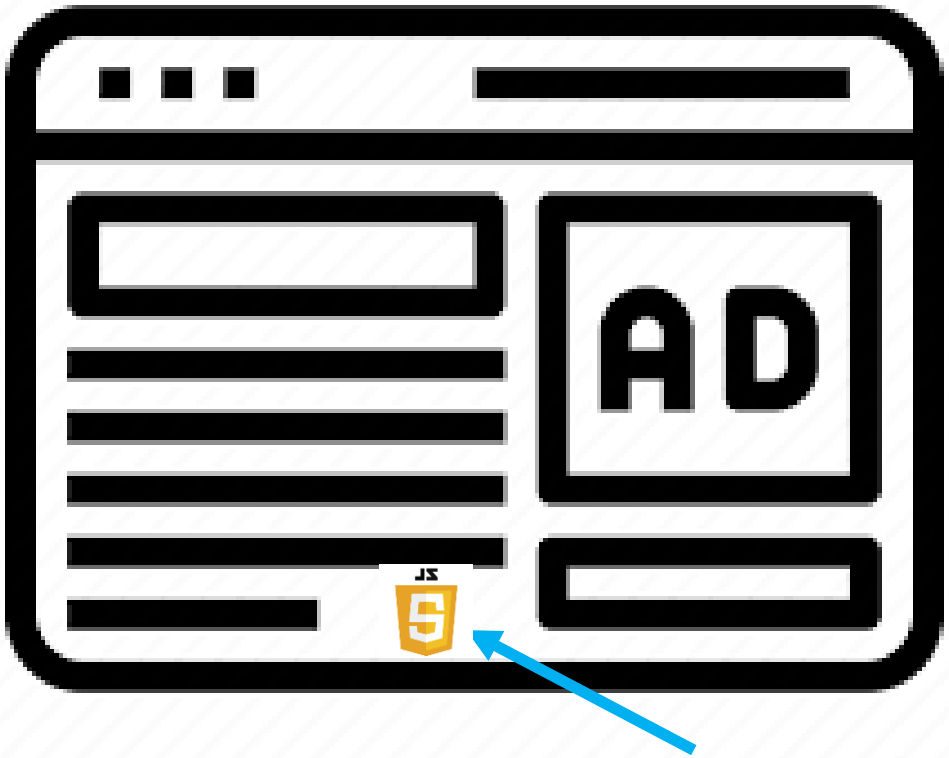
Background: web fingerprinting



It represents a serious privacy thread:

- **Users tracked without consent**
- **Creation of unique identifiers**
- **No compliant with the GDPR**

Background: web fingerprinting



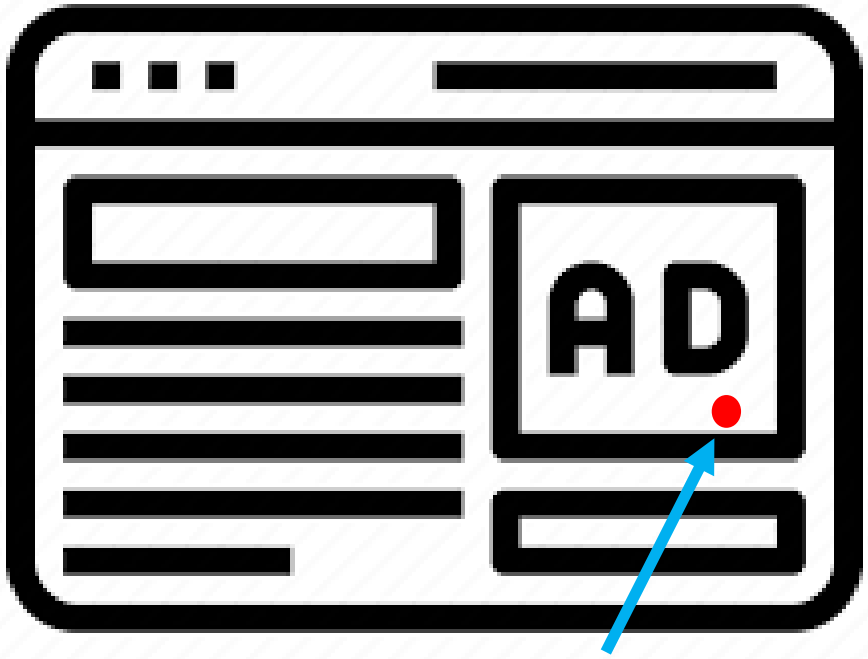
It represents a serious privacy thread:

- **Users tracked without consent**
- **Creation of unique identifiers**
- **No compliant with the GDPR**

Although there are some legit uses:

- **Avoid credential thefts**
- **Fraud identification**

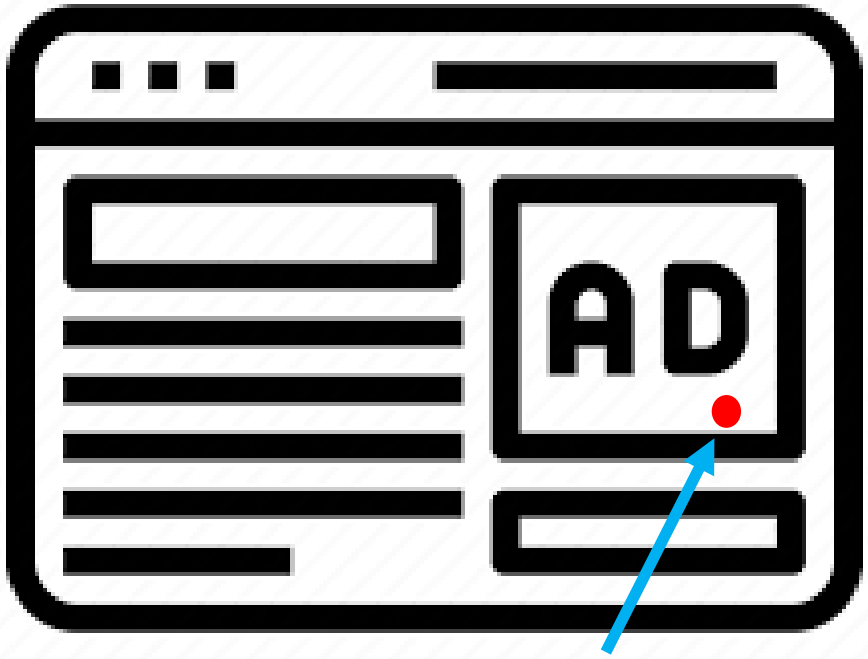
Ad-fingerprinting: Concept



JS code embeeded in an ad

The code to collect attributes is embeeded within an ad

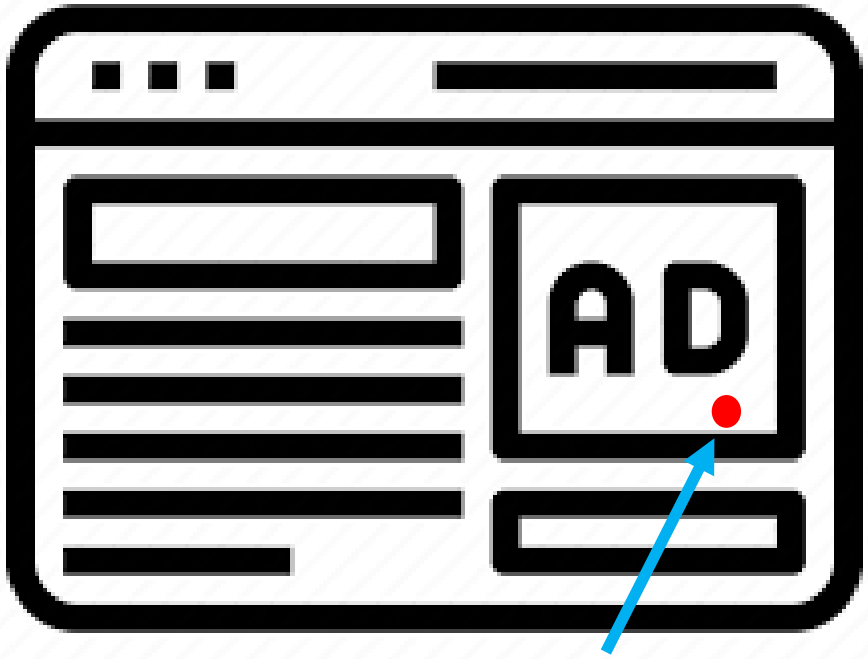
Ad-fingerprinting: Concept



JS code embeeded in an ad

It presents some limitations compared to web fingerprinting due to security policies, e.g., Cross Domain Policy

Ad-fingerprinting: Concept



JS code embeeded in an ad

Complementary uses to web fingerprinting of high value for advertisers:

ATTRIBUTION MODELS

(especially once cookies are fully deprecated)

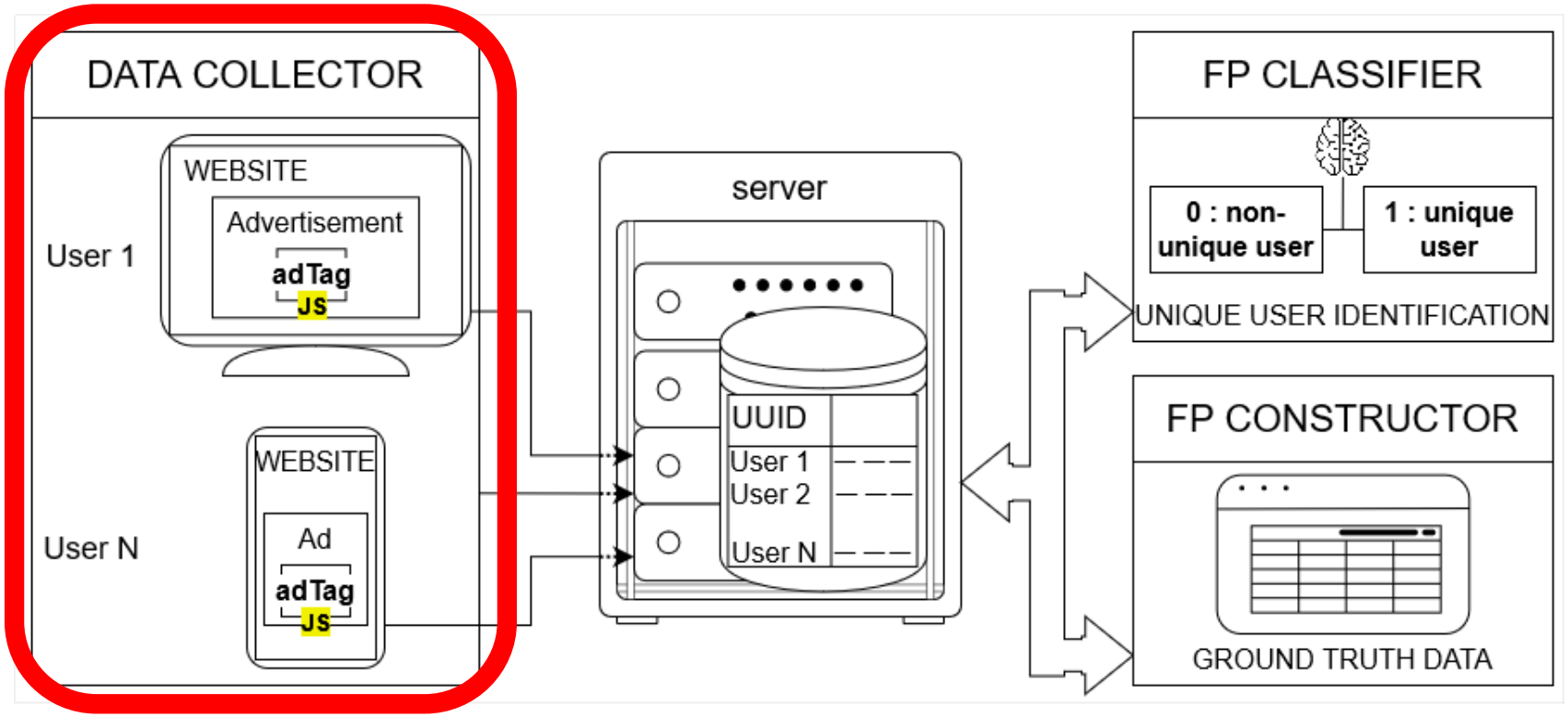
Research questions

1. Is ad fingerprinting viable?

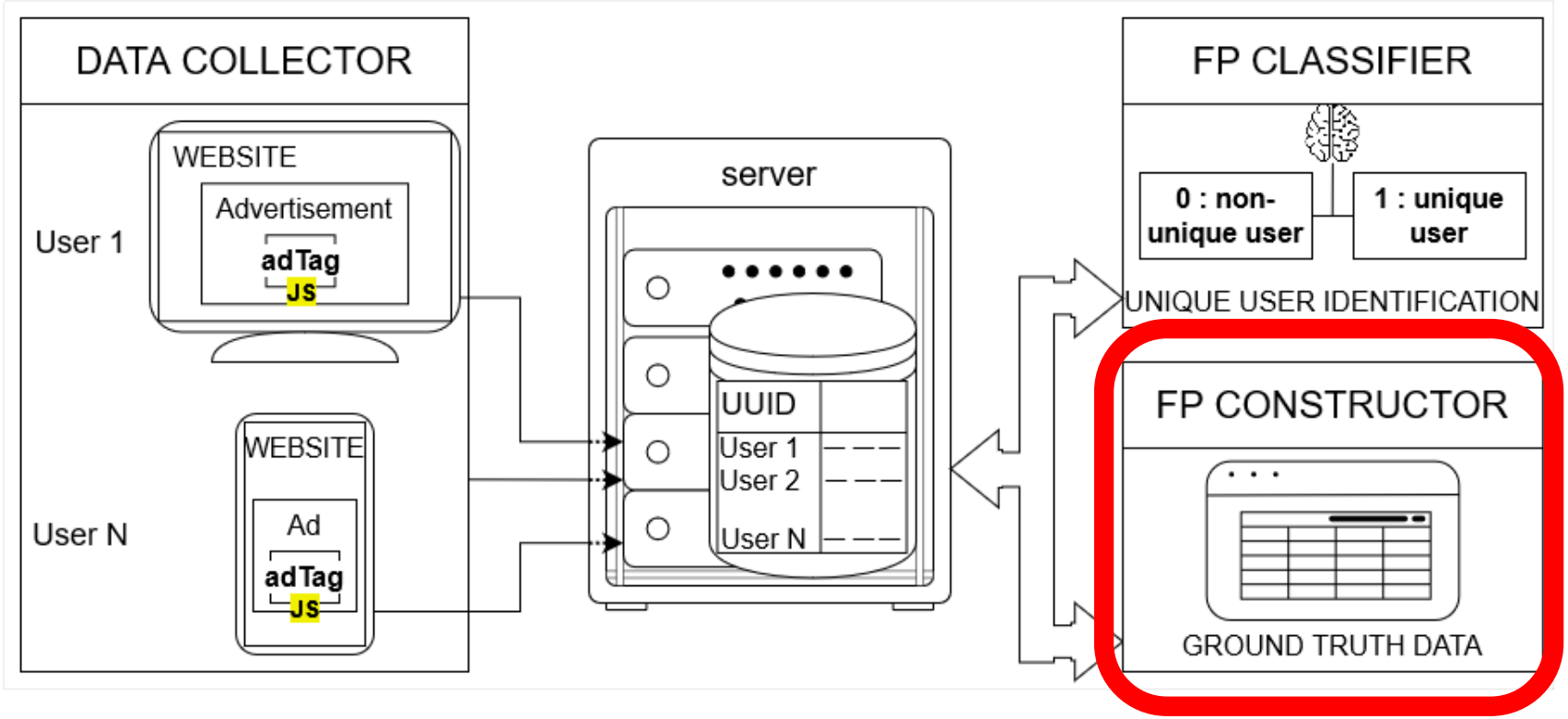
2. How vulnerable are current devices to ad fingerprinting?

3. Is it possible to counter ad fingerprinting with easily adoptable solutions?

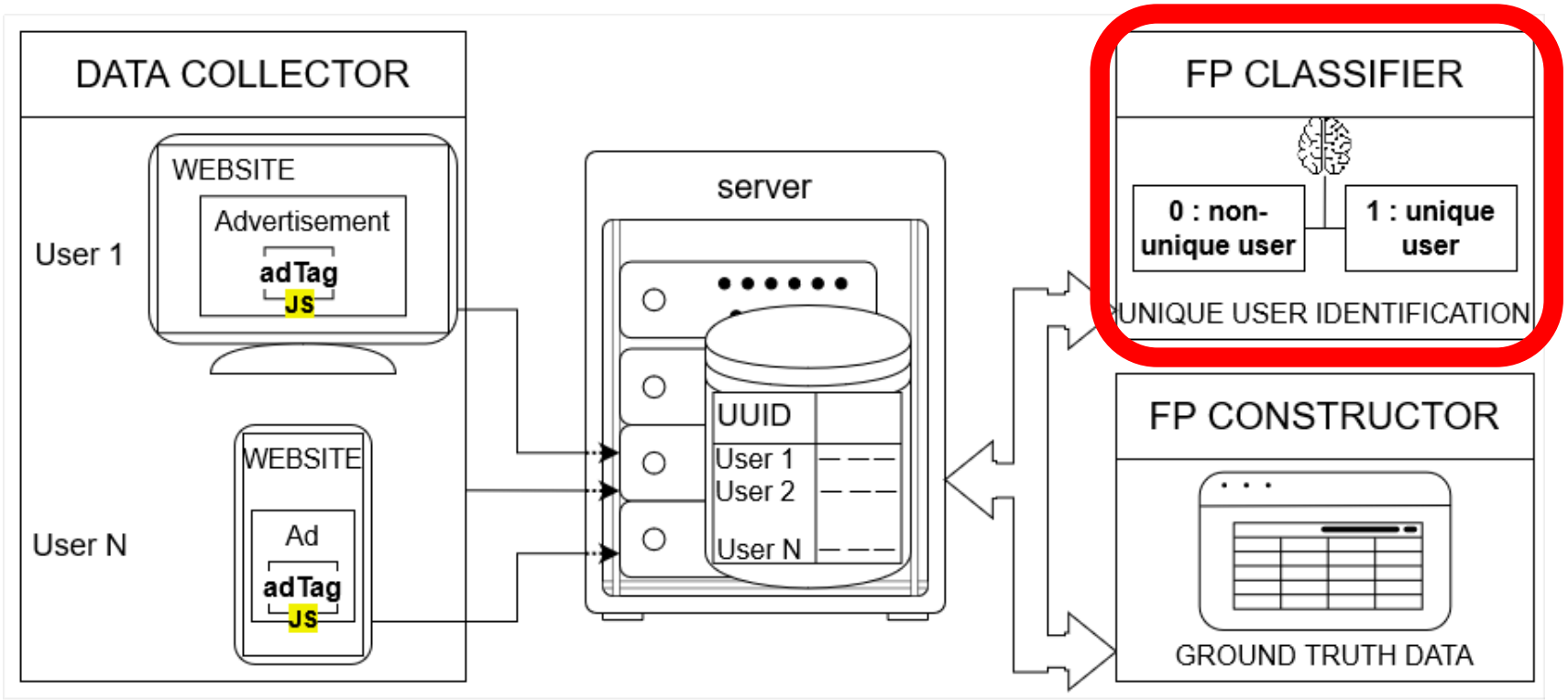
adF system


DATA COLLECTOR
WEBSITE
Advertisement
adTag
JS
User 1
WEBSITE
Ad
adTag
JS
User N
server
UUID
User 1
User 2
User N
FP CLASSIFIER
0 : non-unique user
1 : unique user
UNIQUE USER IDENTIFICATION
FP CONSTRUCTOR
GROUND TRUTH DATA

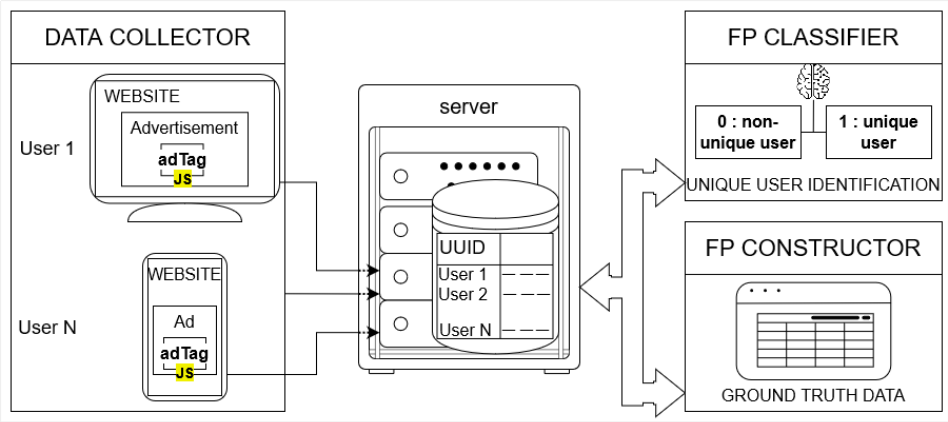
adF system

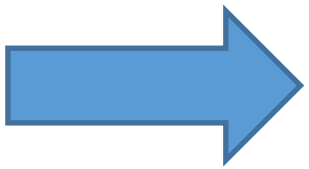

DATA COLLECTOR
User 1
WEBSITE
Advertisement
adTag
JS
User N
WEBSITE
Ad
adTag
JS
server
UUID
User 1
User 2
User N
FP CLASSIFIER
0 : non-unique user
1 : unique user
UNIQUE USER IDENTIFICATION
FP CONSTRUCTOR
GROUND TRUTH DATA

adF system


DATA COLLECTOR
User 1
WEBSITE
Advertisement
adTag
JS
User N
WEBSITE
Ad
adTag
JS
server
UUID
User 1
User 2
User N
FP CLASSIFIER
0 : non-unique user
1 : unique user
UNIQUE USER IDENTIFICATION
FP CONSTRUCTOR
GROUND TRUTH DATA

adF system


DATA COLLECTOR
WEBSITE
Advertisement
adTag
JS
User 1
WEBSITE
Ad
adTag
JS
User N
server
UUID
User 1
User 2
User N
FP CLASSIFIER
0 : non-unique user
1 : unique user
UNIQUE USER IDENTIFICATION
FP CONSTRUCTOR
GROUND TRUTH DATA




unique
Fingerprint
Not unique
fingerprint

Experiments: adF in real ad campaigns

Campaign ID	Start date	End date	Ad Source	Type of Device	OS	CPM	delivered impressions	collected samples in our server	samples with Advertising ID
CampAdFP-010	11 Feb	13 Feb	webs; apps	Mobile; Desktop	All	0.02€	520007	381880	157293
CampAdFP-020	22 May	23 May	webs; apps	Mobile; Desktop	iOS; macOS	0.03€	350550	302539	86471
CampAdFP-021	22 May	23 May	webs; apps	Mobile; Desktop	Linux; Ubuntu	0.04€	9359		
CampAdFP-030	27 May	29 May	webs; apps	Desktop	macOS	0.20€	40048	53111	343044
CampAdFP-031	27 May	29 May	webs; apps	Desktop	Linux; Ubuntu	0.50€	25018		
CampAdFP-040	31 May	2 June	webs; apps	Desktop	macOS	0.04€	547302	590832	343044
CampAdFP-041	31 May	2 June	webs; apps	Desktop	Linux; Ubuntu	0.20€	243790		
CampAdFP-050 *	28 June	30 June	webs; apps	Desktop	macOS	0.36€	12428	23766	22241
CampAdFP-051 *	28 June	30 June	webs; apps	Desktop	Linux; Ubuntu	0.40€	13464		
CampAdFP-APP_1	30 May	30 May	apps	Mobile	Android; iOS	0.10€	40280	594967	292382
CampAdFP-APP_2	06 June	10 June	apps	Mobile	Android; iOS	0.03€	849055		

Datasets

RAW DATASET

Browsers: 1.76M fingeprint samples

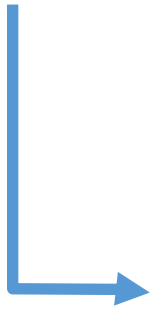
Mobile apps: 870K fingeprint samples

Datasets

RAW DATASET

Browsers: 1.76M fingeprint samples

Mobile apps: 870K fingeprint samples



We remove samples:

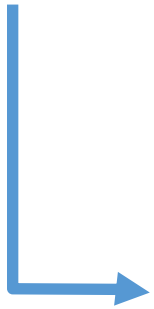
- **Has not an associated advertising id**
- **Its fingeprint appears just once**

Datasets

RAW DATASET

Browsers: 1.76M fingeprint samples

Mobile apps: 870K fingeprint samples



We remove samples:

- **Has not an associated advertising id**
- **Its fingeprint appears just once**

FINGERPRINT DATASET

Browsers: 70k fingeprint samples

Mobile apps: 19,2K fingeprint samples

Fingerprint dataset

FIELD NAME	DEFINITION
DEVICE FINGERPRINT	Device's fingerprint obtained from the Fingerpint Constructor
DEVICE ADVERTISING ID	Device's advertising id obtained from the DSP
TYPE OF DEVICE	Mobile or Desktop
DEVICE OS	Operating System of the device
DEVICE BROWSER	Browser used by the device (*) This field does not apply in the case of mobile apps' samples
ATTRIBUTES	List of attributes collected by Data Collector (65 desktop devices , 34 mobile devices)
GROUND TRUTH UNIQUENESS	Binary variable indicating if the Advertising ID is unique in our dataset. It provides ground-truth information regarding the uniqueness of the device
MESURABLE UNIQUENESS	Binary variable indicating if the adF system identifies the fingerprint as unique or not. **This variable reports the result of the model used by the Fingeprint Classifier.** (*)The classifier is trained using the ground truth uniqueness variable

Vulnerability Metrics

THEORETICAL VULNERABILITY

$$\mathcal{TV} = \frac{N_{tf}}{N_f}$$

MEASURABLE VULNERABILITY

$$\mathcal{MV} = \frac{N_{mf}}{N_f}$$

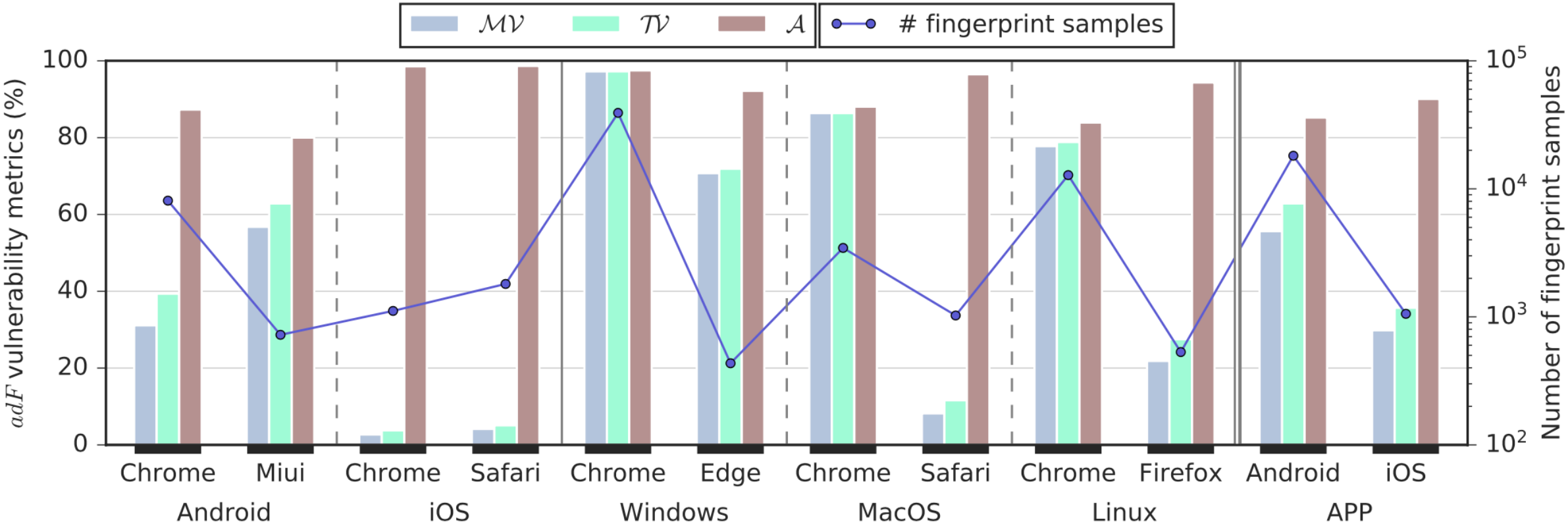
N_f : represents the total number of fingerprints

N_{tf} : represents the total number of fingerprints with ground-truth uniqueness equal to 1

N_{mf} : represents the total number of fingerprints with measurable uniqueness equal to 1

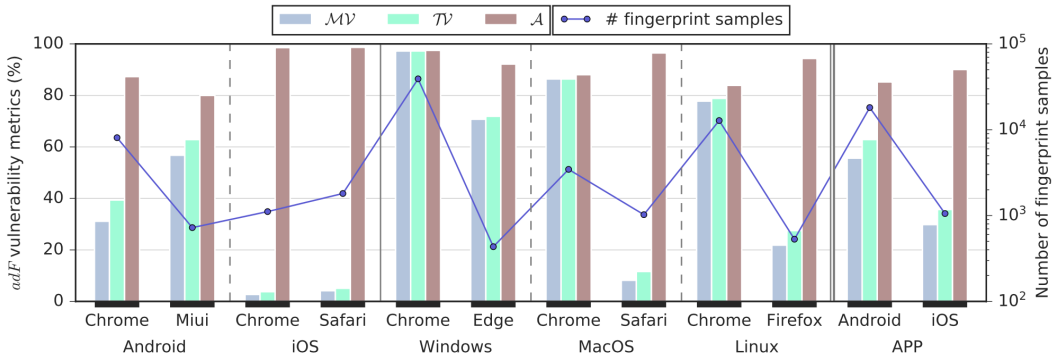
We also measure the accuracy of the adF system, $\mathcal{A}$

Vulnerability of devices to ad fingeprinting



(*) we only consider device configurations with at least 400 samples in our dataset

Vulnerability of devices to ad fingeprinting



If we extrapolate these results to the current market share of the considered device configurations, we conclude that:

1. **At least 70% of desktop devices are vulnerable to ad fingerprinting**

2. **At least 48% of mobile devices are vulnerable through mobile apps, just 20% through mobile browsers**

Reasons behind vulnerability

Reasons behind vulnerability

1. Number of reported attributes

Reasons behind vulnerability

1. Number of reported attributes

2. Cardinality of the attribute, $|S|$

Reasons behind vulnerability

1. Number of reported attributes

2. Cardinality of the attribute, $|S|$

3. Frequency of appearance of different values of an attribute, $|H_n|$

Reasons behind vulnerability

1. Number of reported attributes

Discrimination power

2. Cardinality of the attribute, $|S|$

3. Frequency of appearance of different values of an attribute, $|H_n|$

Countering ad fingerprinting

Current browsers already implement solutions to fight fingerprinting, which basically consist in stopping the reporting of certain attributes

Countering ad fingerprinting: our proposal

Block the reporting of attributes which:

1. Present high discrimination power ($|S| > 25$, $|H_n| > 0{,}1$)
   12 attributes
2. Won't affect the user experience

Countering ad fingerprinting: our proposal

Block the reporting of attributes which:

1. Present high discrimination power ($|S| > 25$, $|H_n| > 0{,}1$)
   12 attributes
2. Won't affect the user experience

Why?:

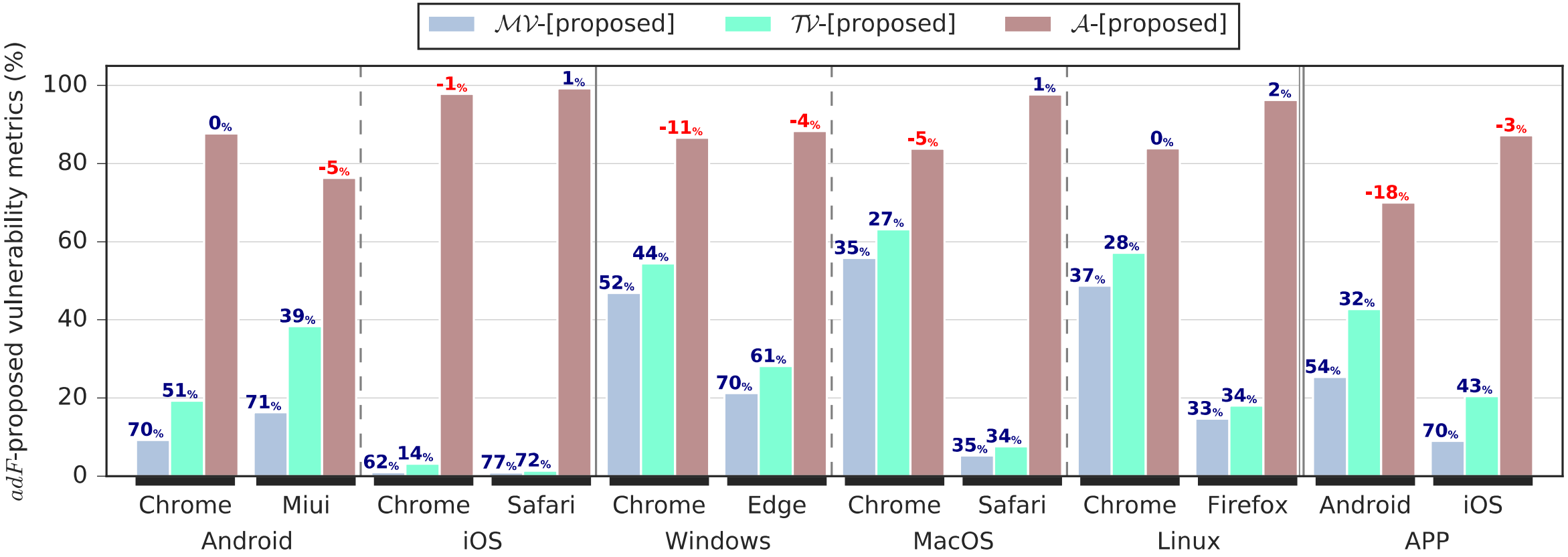
1. It improves state-of-the-art (informed decision)
2. It is easily adoptable by browsers and mobile apps developers

Countering ad fingerprinting: our proposal

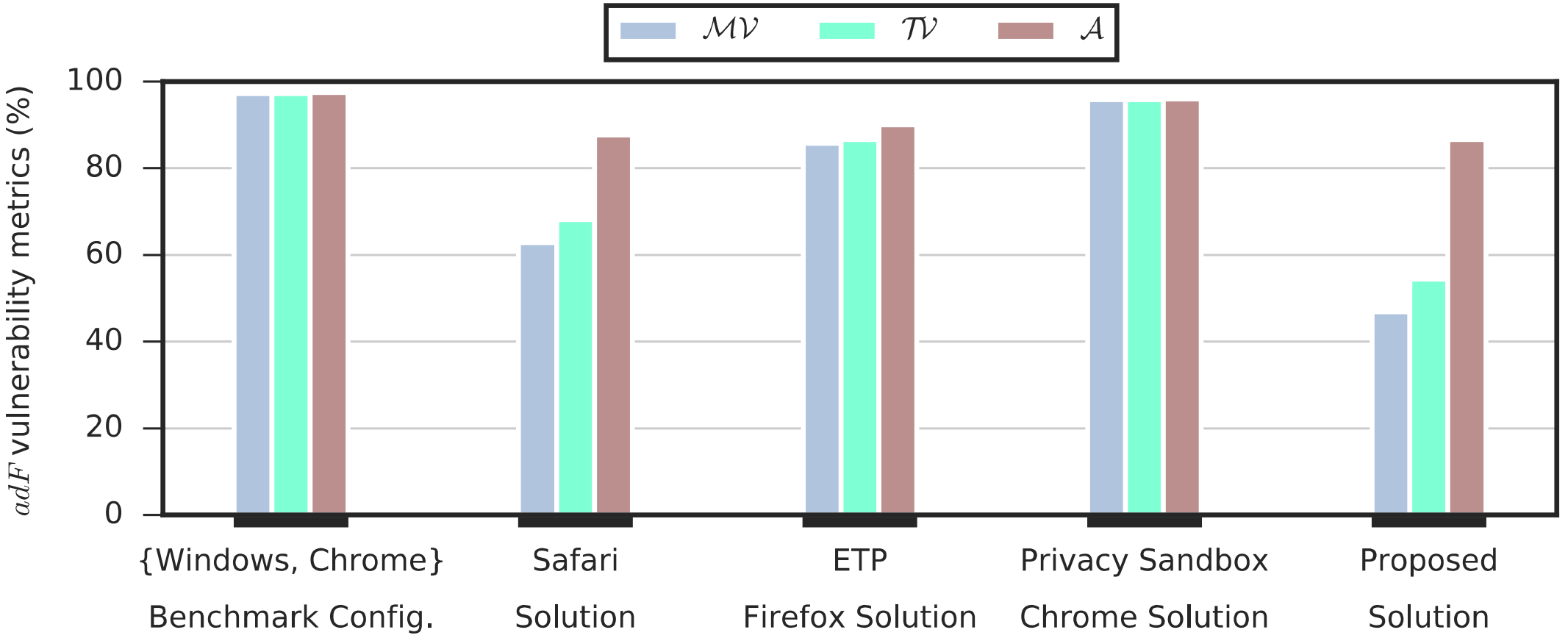
Attributes	MOBILE								DESKTOP												APP			
	Android				iOS				Windows				macOS				GNU/Linux				Android		iOS	
	Chrome		MIUI		Chrome		Safari		Chrome		Edge		Chrome		Safari		Chrome		Firefox					
	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn	|S|	Hn
UserAgent	1521	0,69	40	0,28	136	0,48	83	0,45	309	0,25	26	0,27	89	0,19	51	0,44	160	0,14	57	0,37	5914	0,81	68	0,47
CPU cores	4	0,01	2	0,01	3	0,08	3	0,08	23	0,14	7	0,23	10	0,15	3	0,16	21	0,13	6	0,23	6	0,03	2	0,06
Device memory	6	0,10	3	0,11	1	0,00	3	0,00	6	0,09	4	0,10	4	0,03	2	0,00	6	0,11	2	0,00	5	0,13	1	0,00
Screen: color depth	1	0,00	1	0,00	1	0,00	2	0,00	4	0,00	3	0,01	2	0,08	2	0,10	3	0,00	2	0,02	–	–	–	–
Screen: pixel left	1	0,00	1	0,00	1	0,00	2	0,00	1387	0,16	24	0,09	364	0,23	140	0,22	515	0,18	83	0,37	–	–	–	–
Screen: orientation angle	4	0,04	3	0,03	1	0,00	2	0,00	4	0,00	2	0,00	2	0,00	2	0,00	4	0,06	1	0,00	4	0,01	1	0,00
Screen: orientation type	4	0,04	3	0,03	1	0,00	2	0,00	4	0,00	2	0,00	3	0,00	2	0,00	4	0,07	2	0,00	–	–	–	–
Battery status: charging	2	0,03	2	0,05	1	0,00	2	0,00	2	0,02	2	0,05	3	0,08	2	0,00	3	0,08	2	0,00	–	–	–	–
Simultaneous touch points	3	0,01	1	0,00	2	0,00	4	0,01	18	0,02	5	0,05	1	0,00	3	0,11	8	0,09	2	0,00	4	0,01	2	0,01
Media devices	27	0,20	5	0,01	3	0,10	8	0,10	31	0,13	7	0,19	7	0,05	5	0,12	38	0,21	21	0,23	20	0,13	3	0,11
Languages	471	0,25	16	0,06	20	0,11	24	0,14	926	0,16	24	0,13	370	0,35	23	0,19	960	0,36	32	0,22	767	0,24	28	0,18
PDF viewer enabled	2	0,02	2	0,05	1	0,00	2	0,00	3	0,05	3	0,01	3	0,01	2	0,00	3	0,09	3	0,09	–	–	–	–
User Permissions state	71	0,20	1	0,00	3	0,15	7	0,14	80	0,10	10	0,10	37	0,18	3	0,12	131	0,22	10	0,19	–	–	–	–
Window: available height	155	0,33	36	0,32	27	0,39	25	0,29	480	0,26	52	0,44	418	0,59	174	0,56	500	0,44	146	0,63	243	0,35	12	0,29
Window: available left	1	0,00	1	0,00	1	0,00	2	0,00	161	0,04	8	0,03	91	0,06	14	0,02	143	0,14	54	0,31	–	–	–	–
Window: available top	1	0,00	1	0,00	1	0,00	2	0,00	387	0,03	4	0,01	69	0,19	9	0,15	132	0,12	31	0,27	–	–	–	–
Window: available width	157	0,23	19	0,15	25	0,35	26	0,23	424	0,20	37	0,35	92	0,26	50	0,41	444	0,38	122	0,54	131	0,21	8	0,22
Window: full screen enabled	2	0,08	2	0,10	1	0,00	2	0,00	3	0,03	2	0,08	2	0,08	3	0,01	3	0,07	3	0,11	2	0,03	1	0,00
Storage: quota	902	0,65	169	0,69	1	0,00	4	0,00	17620	0,81	344	0,94	336	0,44	2	0,00	3398	0,73	51	0,18	1880	0,66	1	0,00
navigator porpierties	29	0,07	3	0,06	15	0,17	22	0,21	56	0,09	8	0,08	22	0,06	23	0,32	78	0,17	16	0,21	27	0,11	23	0,32
Plugins	1	0,00	1	0,00	1	0,00	2	0,00	34	0,06	3	0,01	5	0,02	13	0,18	35	0,09	4	0,10	–	–	–	–
Cookie enabled	2	0,01	2	0,01	2	0,10	2	0,09	2	0,00	2	0,01	2	0,00	2	0,10	2	0,00	2	0,07	–	–	–	–
MIME type	1	0,00	1	0,00	1	0,00	2	0,00	12	0,06	3	0,01	6	0,02	7	0,17	11	0,09	6	0,10	–	–	–	–
Canvas	306	0,37	55	0,45	55	0,42	209	0,49	384	0,29	36	0,41	130	0,41	148	0,58	714	0,47	96	0,46	268	0,30	64	0,45
Fonts	3	0,00	2	0,07	9	0,14	10	0,15	6033	0,53	191	0,74	880	0,57	52	0,33	1054	0,36	210	0,70	6	0,00	5	0,14
Bluetooth availability	3	0,08	2	0,05	1	0,00	2	0,00	3	0,07	3	0,09	3	0,09	1	0,00	3	0,10	1	0,00	–	–	–	–
WebGL Extensions	39	0,17	11	0,15	8	0,16	12	0,21	37	0,06	8	0,08	12	0,05	21	0,36	109	0,31	24	0,25	44	0,18	7	0,18
Audio formats: AACP	1	0,00	1	0,00	2	0,00	2	0,01	2	0,00	1	0,00	1	0,00	2	0,01	1	0,00	3	0,01	–	–	–	–
Audio formats: ACC	1	0,00	1	0,00	2	0,00	3	0,01	2	0,00	1	0,00	1	0,00	3	0,01	2	0,00	3	0,01	1	0,00	2	0,01
Audio cxt: base latency	55	0,24	8	0,23	6	0,09	8	0,08	14	0,02	3	0,03	7	0,09	8	0,15	57	0,20	3	0,05	75	0,25	6	0,10
Audio cxt: max channel count	1	0,00	1	0,00	4	0,03	4	0,06	5	0,02	3	0,01	9	0,01	5	0,07	3	0,00	4	0,01	–	–	–	–
Audio cxt: sample rate	3	0,01	2	0,00	5	0,08	5	0,04	10	0,03	3	0,03	6	0,09	7	0,11	5	0,07	2	0,10	4	0,01	5	0,07
Audio cxt: state	2	0,08	2	0,04	1	0,00	2	0,00	2	0,06	2	0,06	2	0,08	1	0,00	2	0,07	1	0,00	–	–	–	–
WebGL (Rend - Param)	182	0,48	64	0,55	8	0,11	11	0,15	1748	0,47	231	0,81	238	0,53	44	0,36	1950	0,65	106	0,55	237	0,42	6	0,12

(*) In bold the attributes we block

Countering ad fingerprinting: results



Countering ad fingerprinting: results



Conclusions

1. Is ad fingerprinting viable?

YES

Conclusions

2. How vulnerable are current devices to ad fingerprinting?

70% of desktop devices are vulnerable
50% of mobile devices are vulnerable

Conclusions

3. Is it possible to counter ad fingerprinting with easily adoptable solutions ?

YES

a simple informed selection of attributes to be blocked reduces the vulnerability up to 72%

Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA ESPAÑA PUEDE

Fingerprinting from ads?: concept, viability and solutions

Miguel Bermejo
Universidad Carlos III de Madrid

Patricia Callejo
Universidad Carlos III de Madrid

Rubén Cuevas
Universidad Carlos III de Madrid

Ángel Cuevas
Universidad Carlos III de Madrid