UNIVERSITÉ CÔTE D'AZUR

Inria INVENTEURS DU MONDE NUMÉRIQUE

Browser fingerprinting

Nataliia Bielova

@nataliabielova

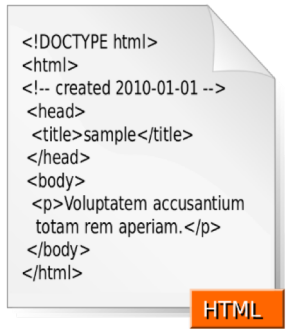
Privacy, Security and ethical aspects of data

Université Cote d'Azur

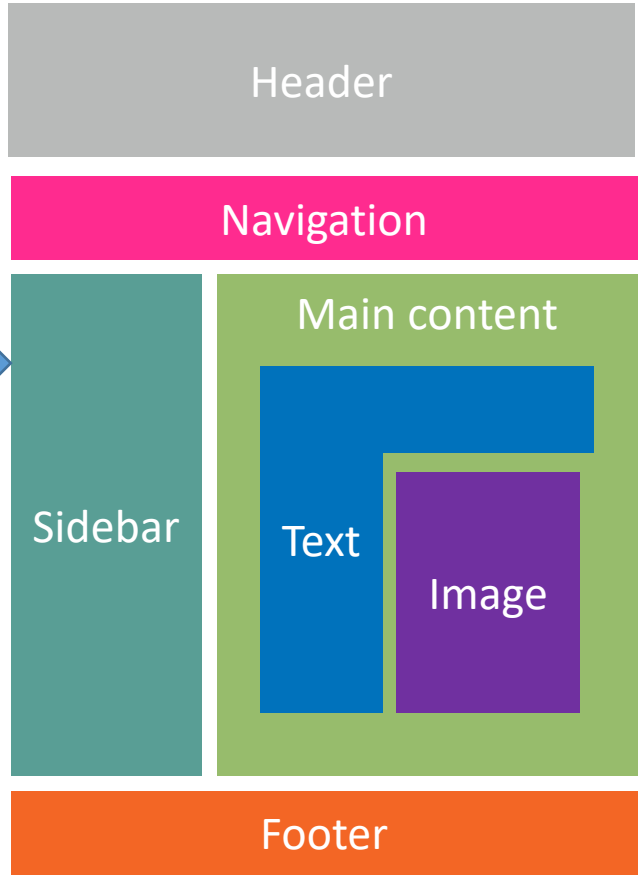
Outline

- A brief history of Web browsers
- What is browser fingerprinting?
- From basic to advanced fingerprinting

I. Internet and web browsers



Browser



Slides courtesy of Pierre Laperdrix (Stony Brook University)

I. Internet in 1995

HTTP User agent

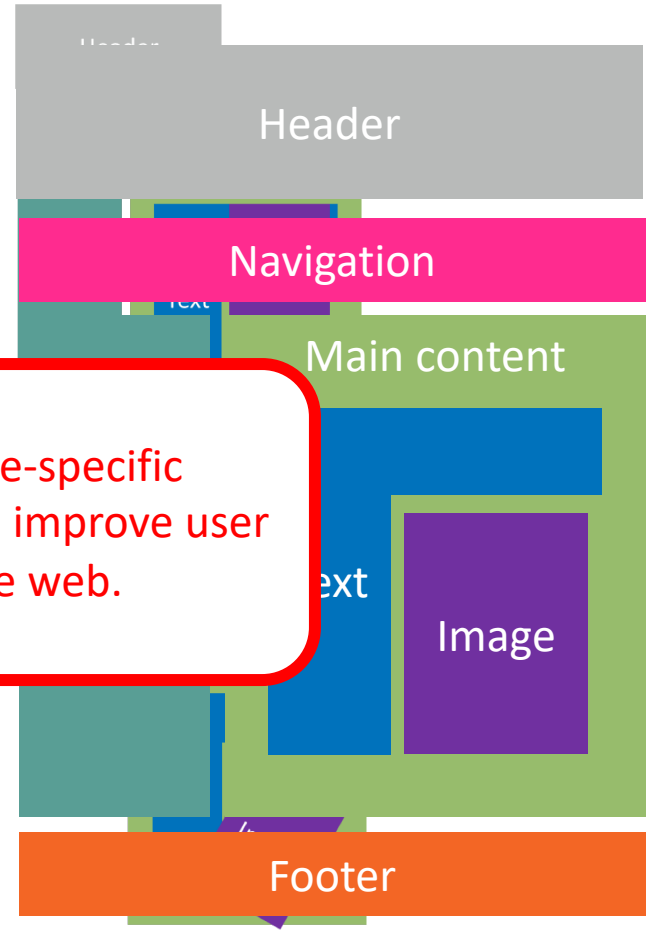
NCSA_Mosaic/2.0 (Windows 3.1)

Mozilla/1.22 (compatible; MSIE 2.0; Windows 95)

I am

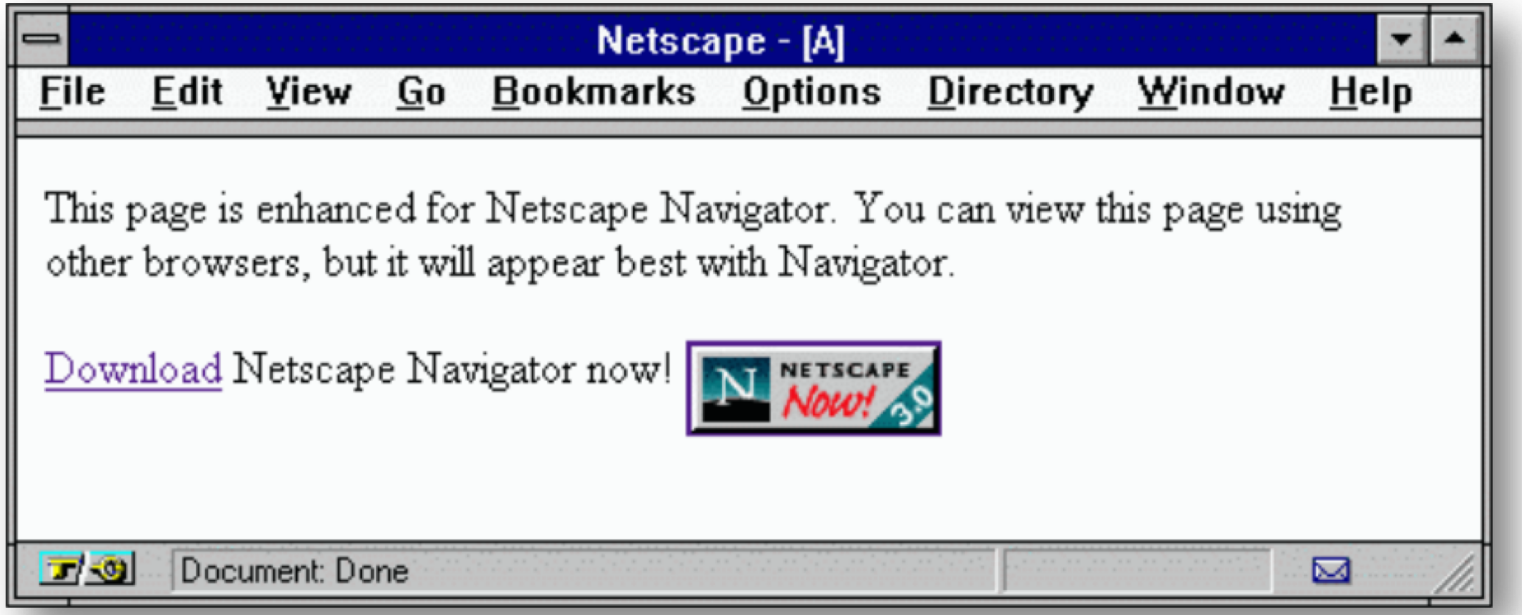
I am

Browsers send device-specific information to servers to improve user experience on the web.



I. Internet in 1995

- Every website announces with **what browser** it is recommended to visit the website



I. Internet in 2017

1995	2017
Browser: Netscape Language: Fr	Browser: Chrome v53 OS: Linux Screen: 1920x1080 Language: Fr Timezone: GMT+1 Graphic card: GTX 1080Ti …

A bigger and richer web

- Audio
- Video
- 3D rendering
- Real-time communications
- Web payments
- Virtual reality

…

What happens when we start collecting all the information available in a web browser?

Example of a browser fingerprint

Attribute	Value
User agent	Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:55.0) Gecko/20100101 Firefox/55.0
HTTP headers	text/html, application/xhtml+xml, application/xml;q=0.9,*/*;q=0.8 gzip, deflate, br en-US,en;q=0.5
Plugins	Plugin 0: QuickTime Plug-in 7.6.6; libtotem-narrowspace-plugin.so; Plugin 1: Shockwave Flash; Shockwave Flash 26.0 r0; libflashplayer.so.
Fonts	Century Schoolbook, Source Sans Pro Light, DejaVu Sans Mono, Bitstream Vera Serif, URW Palladio L, Bitstream Vera Sans Mono, Bitstream Vera Sans, ...
Platform	Linux x86_64
Screen resolution	1920x1080x24
Timezone	-480 (UTC+8)
OS	Linux 3.14.3-200.fc20.x86 32-bit
WebGL vendor	NVIDIA Corporation
WebGL renderer	GeForce GTX 650 Ti/PCIe/SSE2
Canvas	Cwm fjordbank glyphs vext quiz, 😀 Cwm fjordbank glyphs vext quiz, 😀

Slides courtesy of Pierre Laperdrix (Stony Brook University)

I. Definition of browser fingerprinting

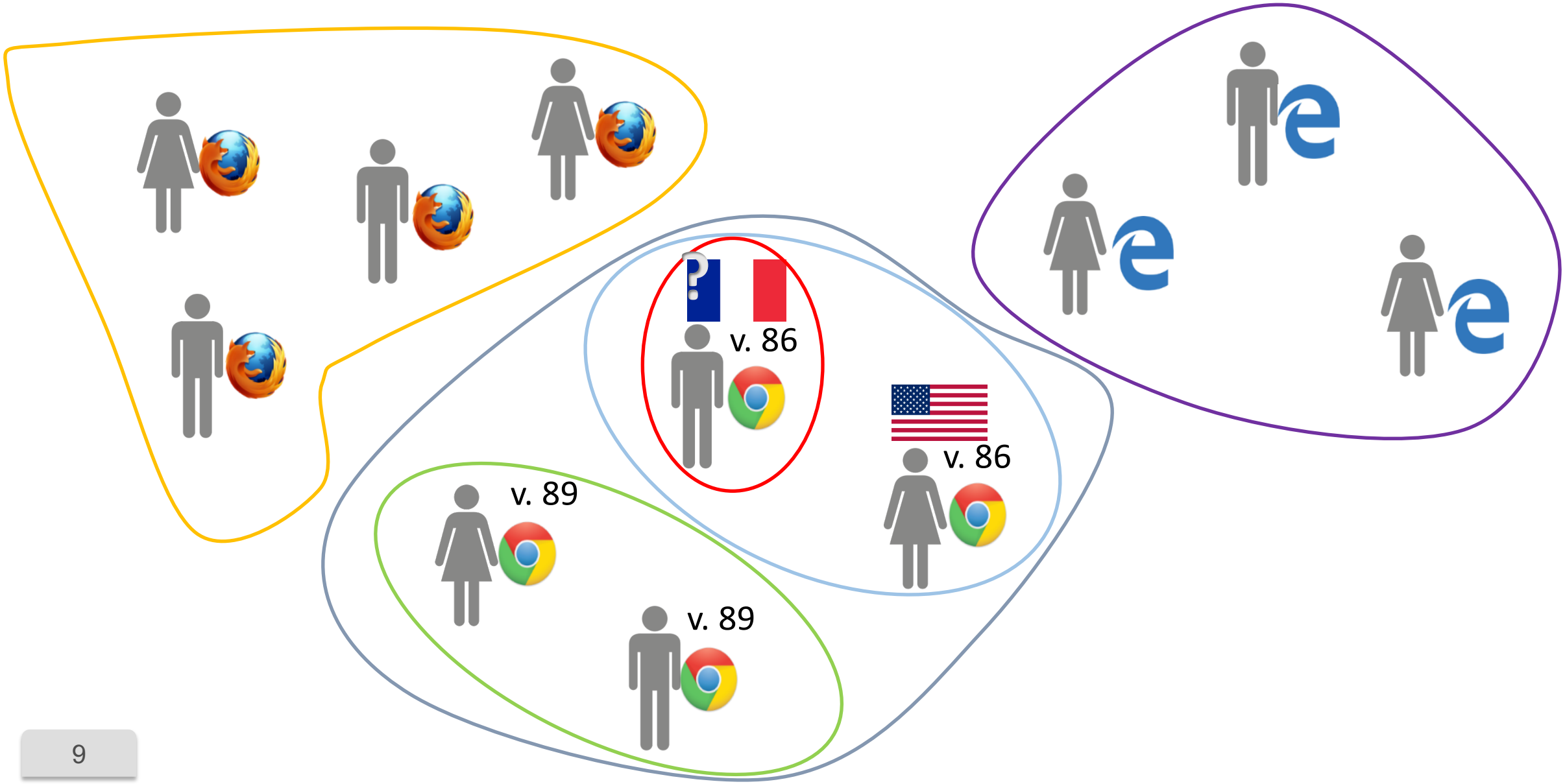
Definitions

- A browser fingerprint is a set of information related to a user's device from the hardware to the operating system to the browser and its configuration.

- Browser fingerprinting refers to the process of collecting information through a web browser to build a fingerprint of a device.

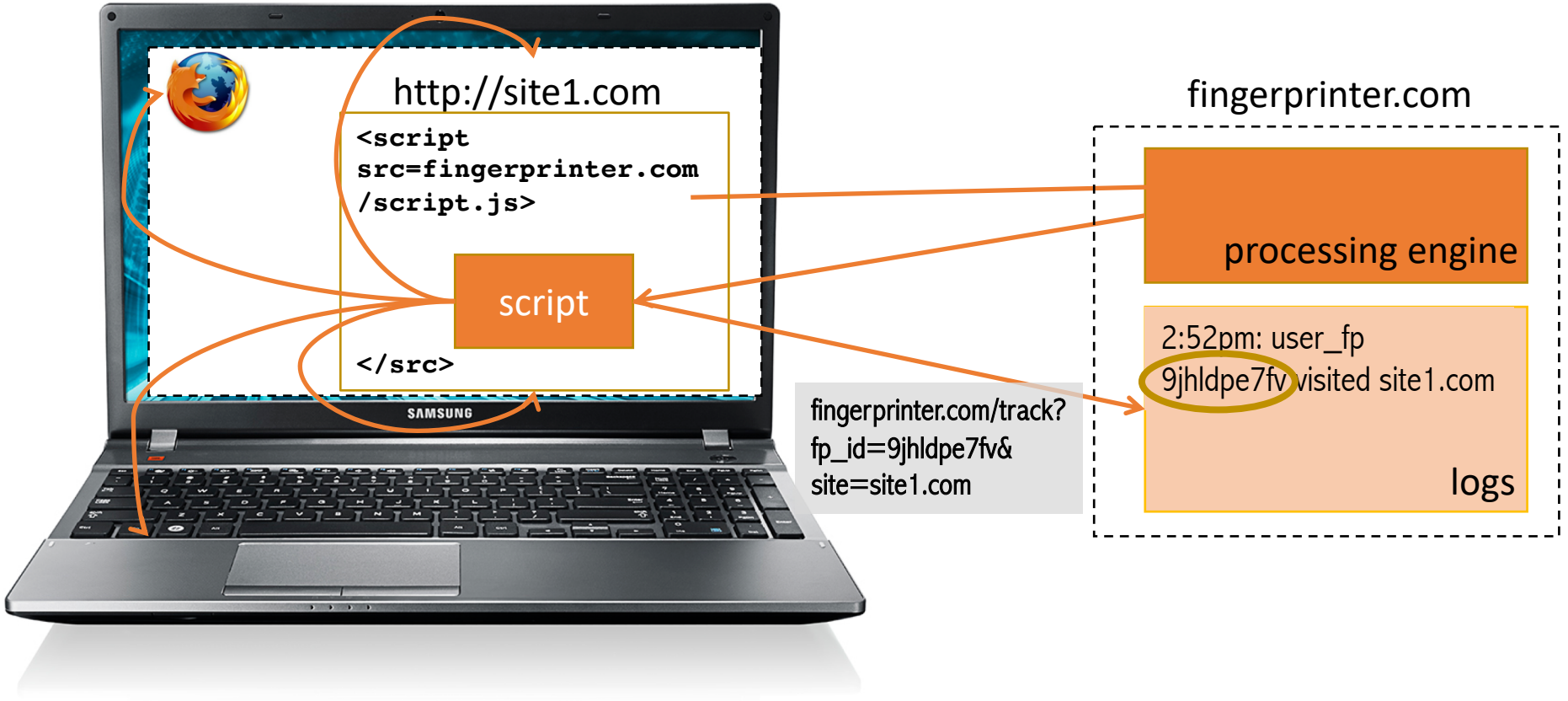
Slides courtesy of Pierre Laperdrix (Stony Brook University)

How can we be identified by a browser fingerprint?



Browser fingerprinting used for tracking

Browser and operating system properties are used to track repeated visits to a site.



Comparison of the emoji on different devices and OSs

(a) Windows 7

(b) Windows 10

(c) Linux

(d) iOS

(e) Firefox OS

(f) Android 4.3 and before

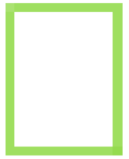
(g) Android 4.4

(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

https://hal.inria.fr/hal-01285470/document

What makes fingerprinting a threat to online privacy?

Two studies have investigated the diversity of browser fingerprints.

Am I Unique?

470,161 fingerprints
94.2% were unique

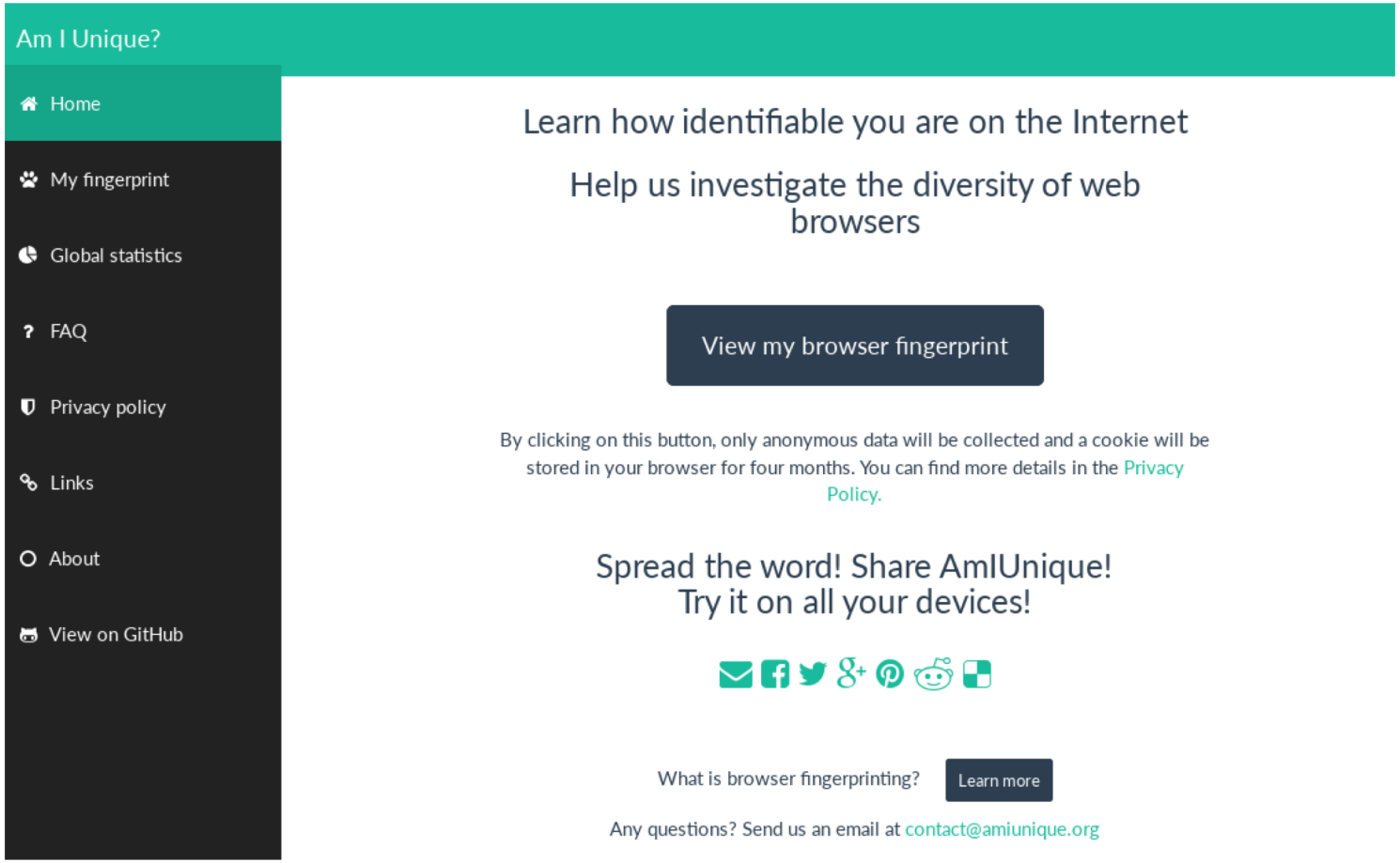
118,934 fingerprints
89.4% were unique

Tracking is possible

Slides courtesy of Pierre Laperdrix (Stony Brook University)

I. See your own fingerprint

https://amiunique.org (Am I Unique)



- Website launched in November 2014
- Collected 660,000+ fingerprints so far
- Browser extension available to see the evolution of your own browser fingerprint

Slides courtesy of Pierre Laperdrix (Stony Brook University)

Fingerprinting

- Panopticlick [Eckersley, PET'2010]

> Your browser fingerprint **appears to be unique** among the 2,419,678 tested so far.
>
> Currently, we estimate that your browser has a fingerprint that conveys **at least 21.21 bits of identifying information.**

- Information needed to **uniquely identify a browser**
  - n – number of connected devices: 5 000 000 000
  - $log_2 n$ – number of bits for a unique id: 33 bits

- **Idea: distinguish user's browsers** by accessing browser features and using their probability distributions

I. Example of values collected on AmIUnique

Slides courtesy of Pierre Laperdrix (Stony Brook University)

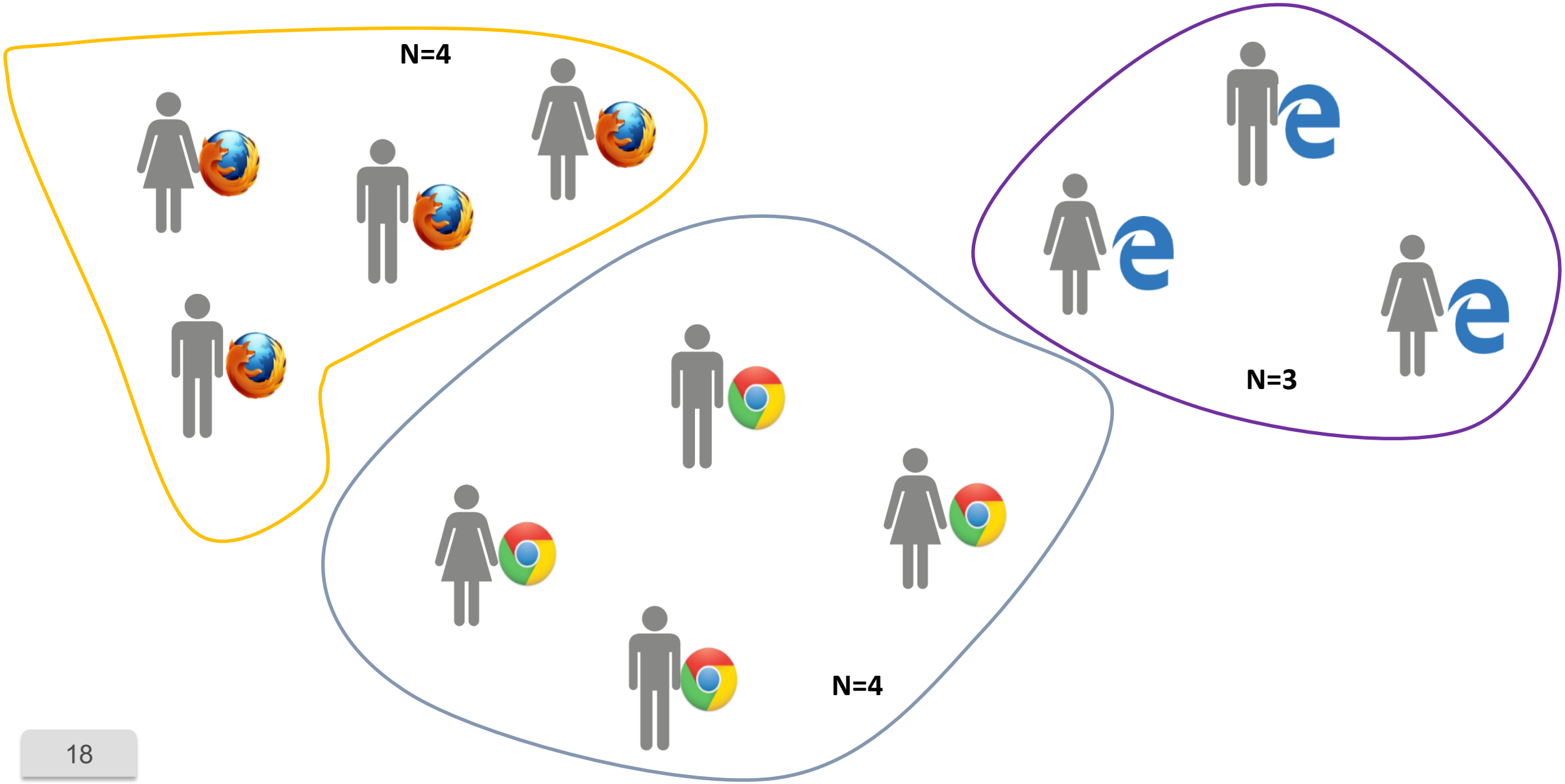
How to evaluate uniqueness in a dataset?

	Browser type	Browser version	Language
1	Firefox	83.0	EN
2	Firefox	83.0	EN
3	Firefox	82.0	EN
4	Firefox	82.0	EN
5	Chrome	86.0	EN
6	Chrome	86.0	FR
7	Chrome	89.0	EN
8	Chrome	89.0	FR
9	Edge	85.0	FR
10	Edge	85.0	FR
11	Edge	87.0	FR

Anonymity sets

	Browser type
1	Firefox
2	Firefox
3	Firefox
4	Firefox
5	Chrome
6	Chrome
7	Chrome
8	Chrome
9	Edge
10	Edge
11	Edge

Anonymity sets



Anonymity sets

	Browser type
1	Firefox
2	Firefox
3	Firefox
4	Firefox
5	Chrome
6	Chrome
7	Chrome
8	Chrome
9	Edge
10	Edge
11	Edge

N=4

N=4

N=3

Anonymity set size N=4 with 8 elements

Anonymity set size N=3 with 3 elements

Anonymity sets

	Browser type	Browser version
1	Firefox	83.0
2	Firefox	83.0
3	Firefox	82.0
4	Firefox	82.0
5	Chrome	86.0
6	Chrome	86.0
7	Chrome	89.0
8	Chrome	89.0
9	Edge	85.0
10	Edge	85.0
11	Edge	87.0

N=2

N=2

N=2

N=2

N=2

N=1

Anonymity set size N=2 with 10 elements

Anonymity set size N=1 with 1 element

Anonymity sets

	Browser type	Browser version	Language
1	Firefox	83.0	EN
2	Firefox	83.0	EN
3	Firefox	82.0	EN
4	Firefox	82.0	EN
5	Chrome	86.0	EN
6	Chrome	86.0	FR
7	Chrome	89.0	EN
8	Chrome	89.0	FR
9	Edge	85.0	FR
10	Edge	85.0	FR
11	Edge	87.0	FR

What are the anonymity sets here?
Compute it yourself!

Example of a browser fingerprint

Attribute	Value
User agent	Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:55.0) Gecko/20100101 Firefox/55.0
HTTP headers	text/html, application/xhtml+xml, application/xml;q=0.9,*/*;q=0.8 gzip, deflate, br en-US,en;q=0.5
Plugins	Plugin 0: QuickTime Plug-in 7.6.6; libtotem-narrowspace-plugin.so; Plugin 1: Shockwave Flash; Shockwave Flash 26.0 r0; libflashplayer.so.
Fonts	Century Schoolbook, Source Sans Pro Light, DejaVu Sans Mono, Bitstream Vera Serif, URW Palladio L, Bitstream Vera Sans Mono, Bitstream Vera Sans, ...
Platform	Linux x86_64
Screen resolution	1920x1080x24
Timezone	-480 (UTC+8)
OS	Linux 3.14.3-200.fc20.x86 32-bit
WebGL vendor	NVIDIA Corporation
WebGL renderer	GeForce GTX 650 Ti/PCIe/SSE2
Canvas	Cwm fjordbank glyphs vext quiz, 😀 Cwm fjordbank glyphs vext quiz, 😀

Slides courtesy of Pierre Laperdrix (Stony Brook University)

I. Example of values collected on AmIUnique

Some user-agents

- Mozilla/5.0 (Windows NT 6.1; WOW64; rv:34.0) Gecko/20100101 Firefox/34.0
- Mozilla/5.0 (iPhone; CPU iPhone OS 8_1_2 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12B440 Safari/600.1.4
- Mozilla/5.0 (Android; Mobile; rv:27.0) Gecko/27.0 Firefox/27.0
- Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_2) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/39.0.2171.95 Safari/537.36
- Mozilla/5.0 (X11; Ubuntu; Linux i686; rv:34.0) Gecko/20100101 Firefox/34.0

Slides courtesy of Pierre Laperdrix (Stony Brook University)

I. Example of values collected on AmIUnique

Other custom user-agents

- godzilla/5.0 (X122; BSD; rv:500.0) Gecko/20100101
- pouet
- “54. When a warlike prince attacks a powerful state, his generalship shows itself in preventing the concentration of the enemy's forces. He overawes his opponents, and their allies are prevented from joining against him.”
- Deepnet Explorer 1.5.3; Smart 2x2; Avant Browser; .NET CLR 2.0.50727; InfoPath.1)
- NSA
- Game Boy Advance
- eat it

Slides courtesy of Pierre Laperdrix (Stony Brook University)

How to compare datasets: Anonymity sets

- User-agent on Desktop vs Mobile devices

More than 20% of mobile users are unique because of their browser's UserAgent!

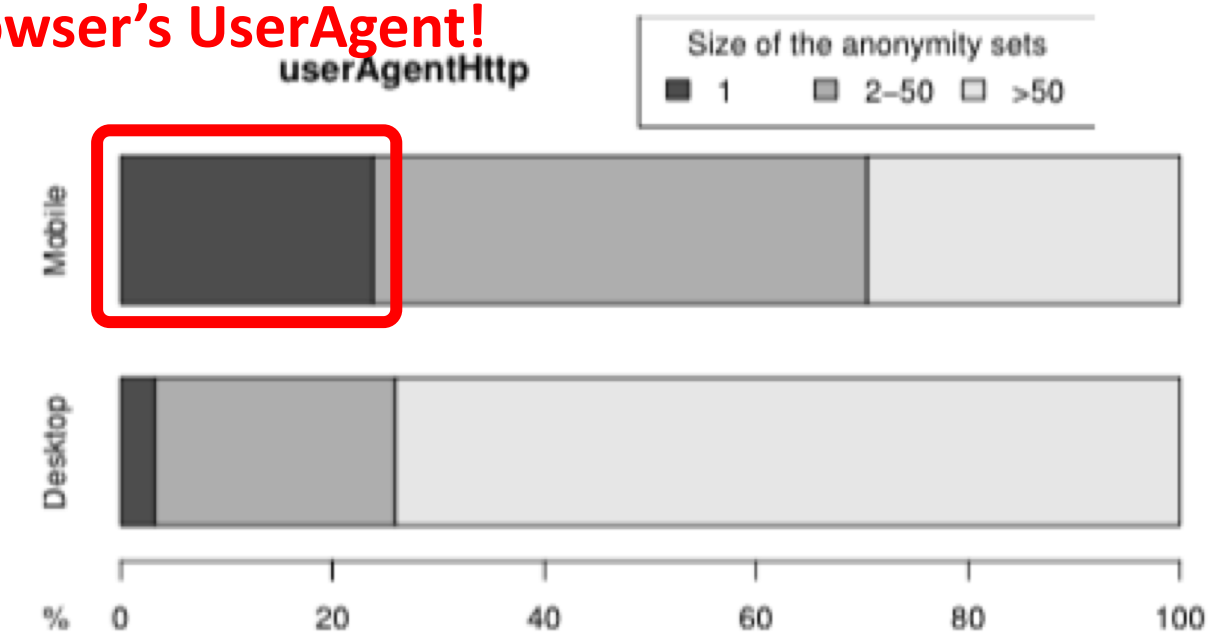


Fig. 4. Comparison of anonymity set sizes on the user-agent between desktop and mobile devices

25

Anonymity sets for mobile devices

- User-agent on Android vs iOS devices

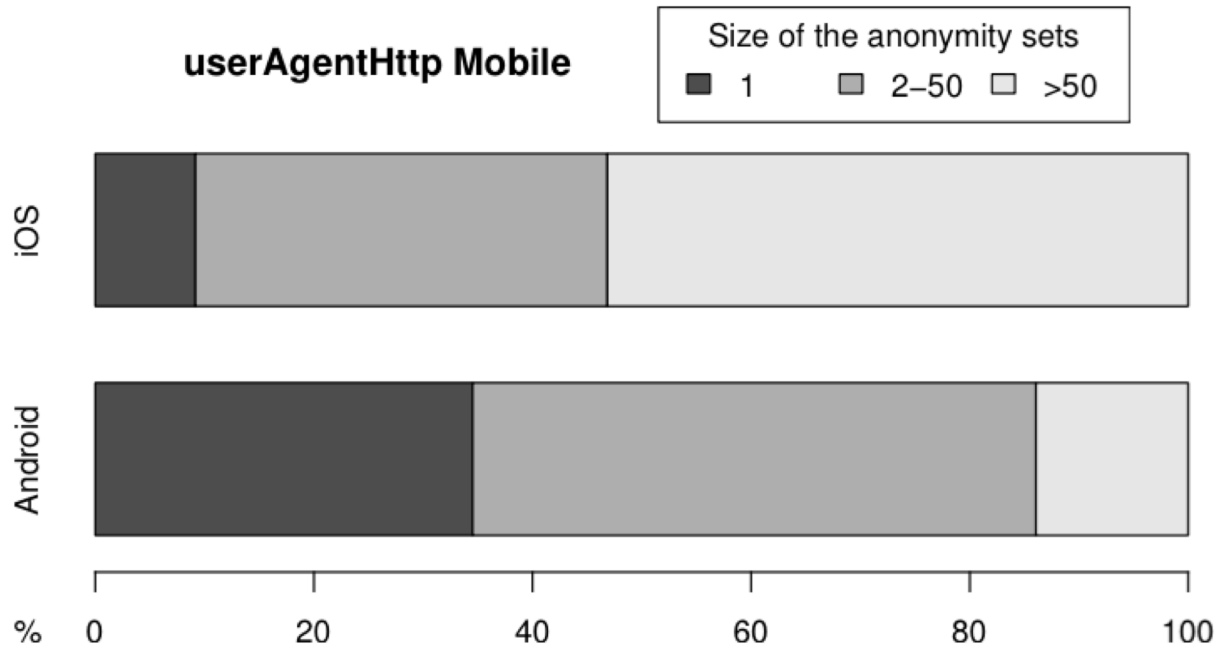


Fig. 5. Comparison of anonymity set sizes on the user-agent between Android and iOS devices

26

What if I disable JavaScript?

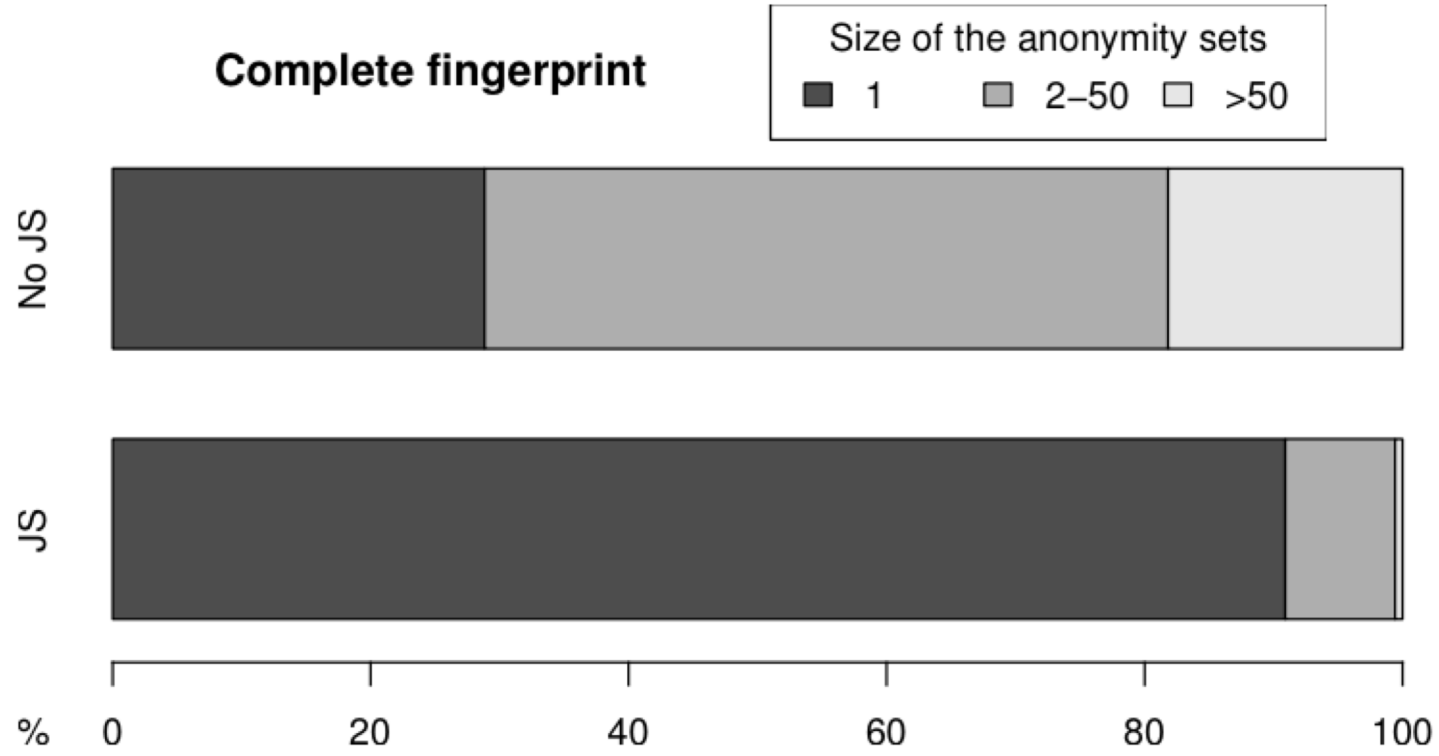


Fig. 9. Comparison of anonymity set sizes on the complete fingerprint between devices with and without JavaScript

I. Summary

- Servers can easily collect information about a device to form what is called a **browser fingerprint**.
- There is so much diversity that users can be **tracked** online if their fingerprint is **unique**.
- Test your device on

  https://amiunique.org

Slides courtesy of Pierre Laperdrix (Stony Brook University)

https://hal.inria.fr/hal-02864872

Browser Fingerprinting: A Survey

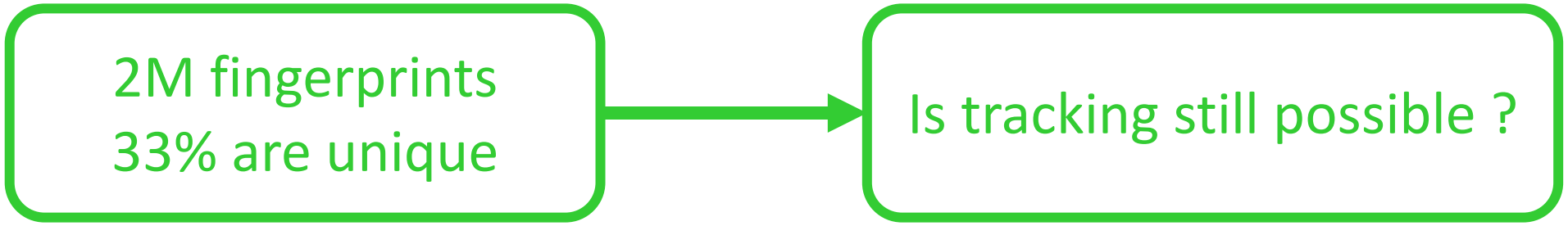
Pierre Laperdrix [1, 2] , Nataliia Bielova [3] , Benoit Baudry [4] , Gildas Avoine [5] Details

1 CRIStAL - Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL) - UMR 9189
2 SPIRALS - Self-adaptation for distributed services and large software systems
Inria Lille - Nord Europe, CRIStAL - Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL) - UMR 9189
3 PRIVATICS - Privacy Models, Architectures and Tools for the Information Society
Inria Grenoble - Rhône-Alpes, CITI - CITI Centre of Innovation in Telecommunications and Integration of services
4 KTH - Royal Institute of Technology [Stockholm]
5 ASCII - Analyse d'interactions stochastiques intelligentes et coopératives
CMAP - Centre de Mathématiques Appliquées - Ecole Polytechnique, Inria Saclay - Ile de France

Abstract : With this article, we survey the research performed in the domain of browser fingerprinting, while providing an accessible entry point to newcomers in the field. We explain how this technique works and where it stems from. We analyze the related work in detail to understand the composition of modern fingerprints and see how this technique is currently used online. We systematize existing defense solutions into different categories and detail the current challenges yet to overcome.

I. Summary

- How effective is fingerprinting at large scale?



Hiding in the Crowd: an Analysis of the Effectiveness of Browser Fingerprinting at Large Scale

Alejandro Gómez-Boix, Pierre Laperdrix, Benoit Baudry

The Web Conference (**WWW 2018**)

Slides courtesy of Pierre Laperdrix (Stony Brook University)

2M users in France (WWW 2018)

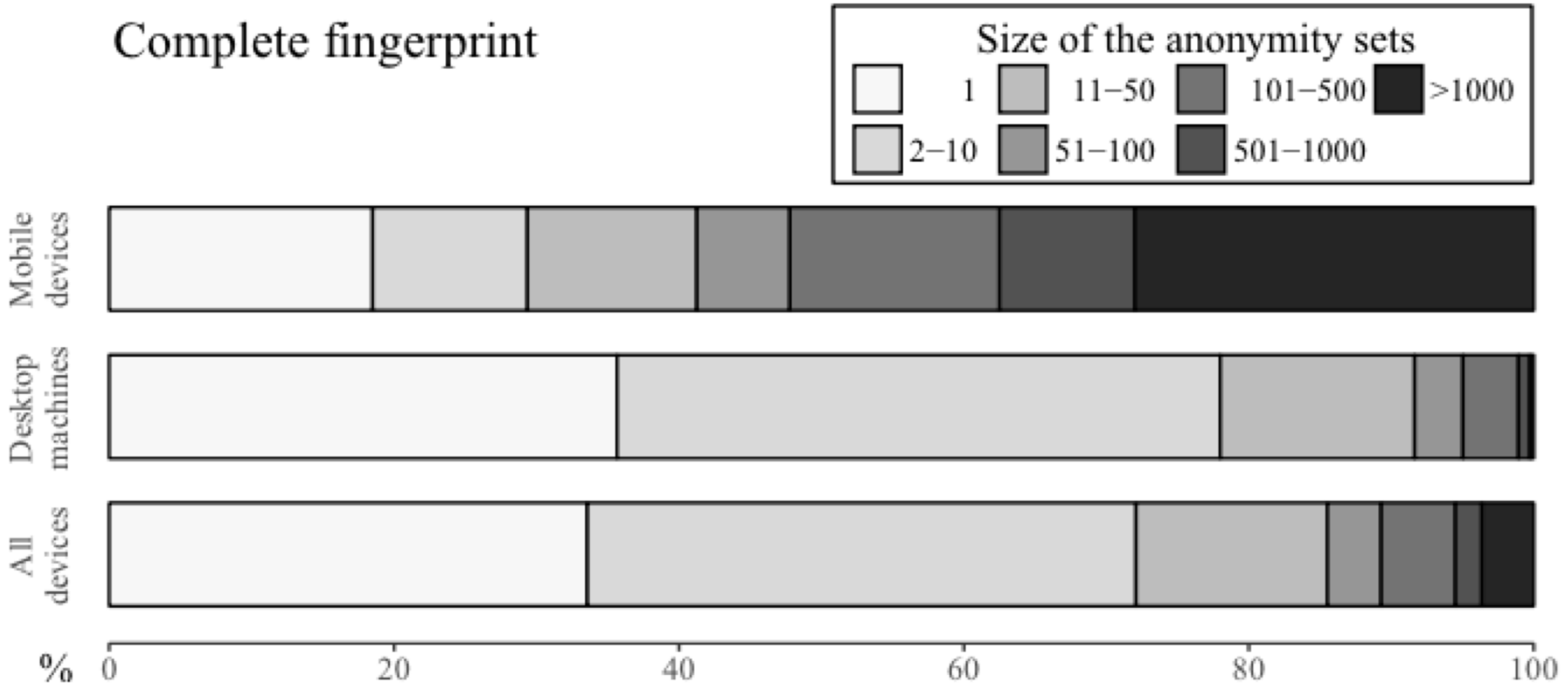
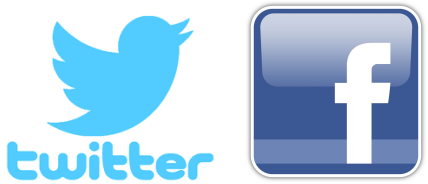


Figure 3: Comparison of anonymity set sizes between mobile devices and desktop/laptop machines.

Slides courtesy of Pierre Laperdrix (Stony Brook University)

New Fingerprinting Methods

- **Privacy Paradox**
  - Users' fingerprints can be enriched by their browser extensions
  - Moreover, we found an attack allows to detect 58 web services where the user is logged in!



G.G. Gulyás, D. F. Some, **N. Bielova** and C. Castelluccia. To Extend or not to Extend: on the Uniqueness of Browser Extensions and Web Logins. *WPES@ACM CCS 2018.*

32

I. Plugins VS Browser extensions

- **Plugins** were created to display content not supported by the browser
  - Flash Java Silverlight

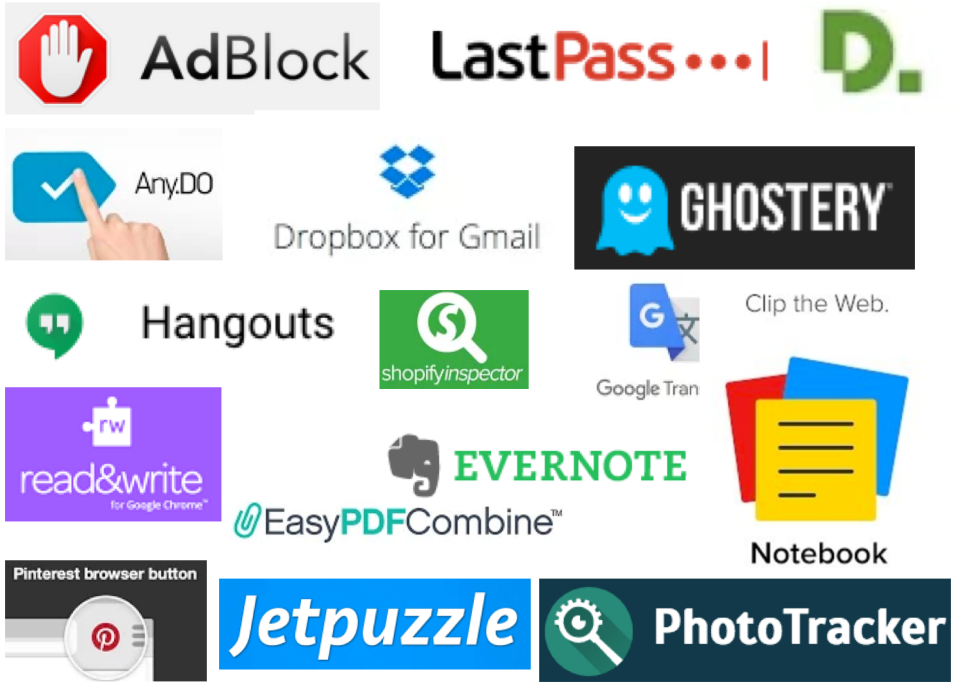
- All installed plugins are accessible via the `navigator.plugins` JavaScript object

- **Extensions** extend or modify default behavior of a browser
  - AdBlockPlus, LastPass, Ghostery, Pinterest

- There is no API that webpages can use to detect all installed extensions

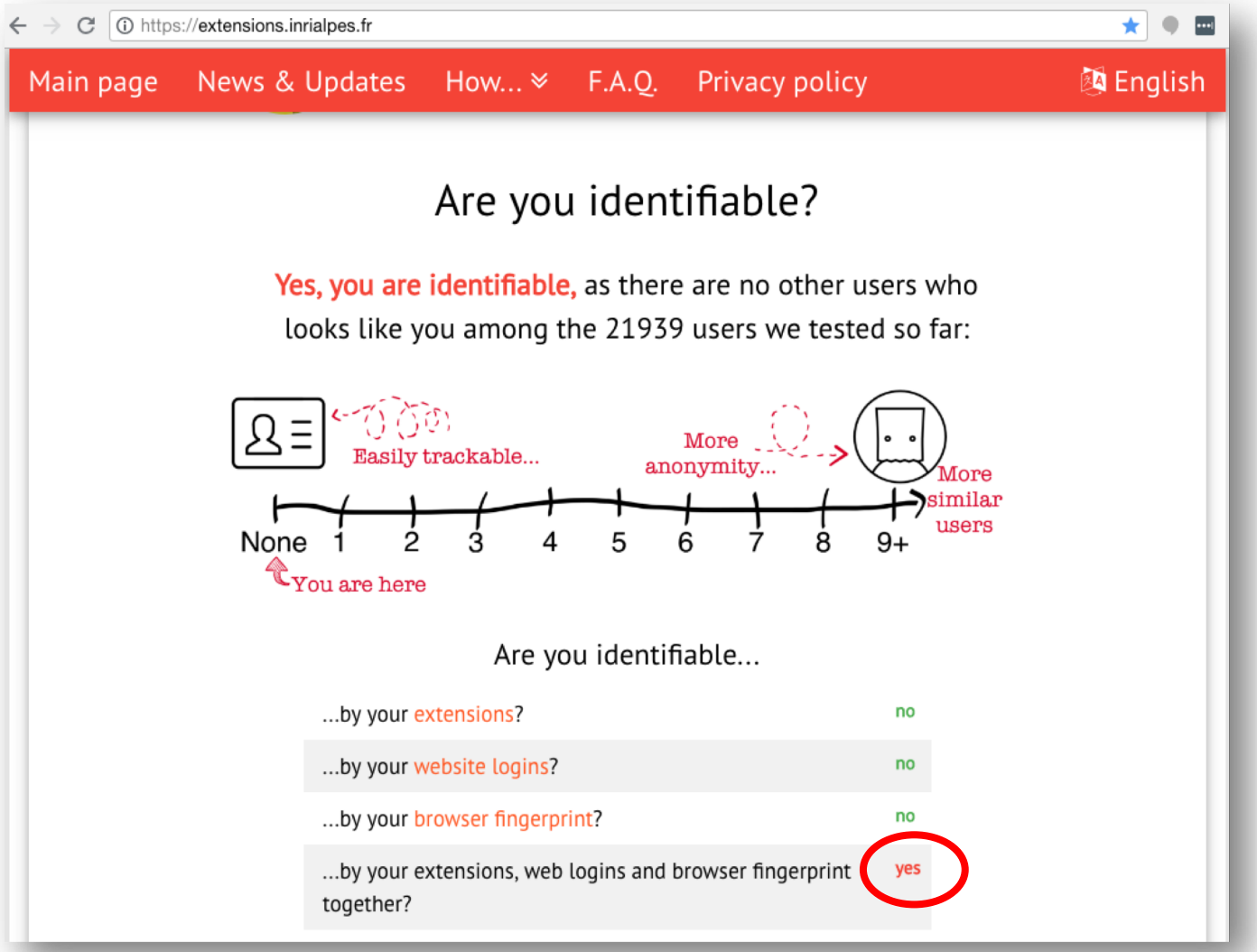
How unique is your browser?

- Browser extension detection
- ~13 000 extensions

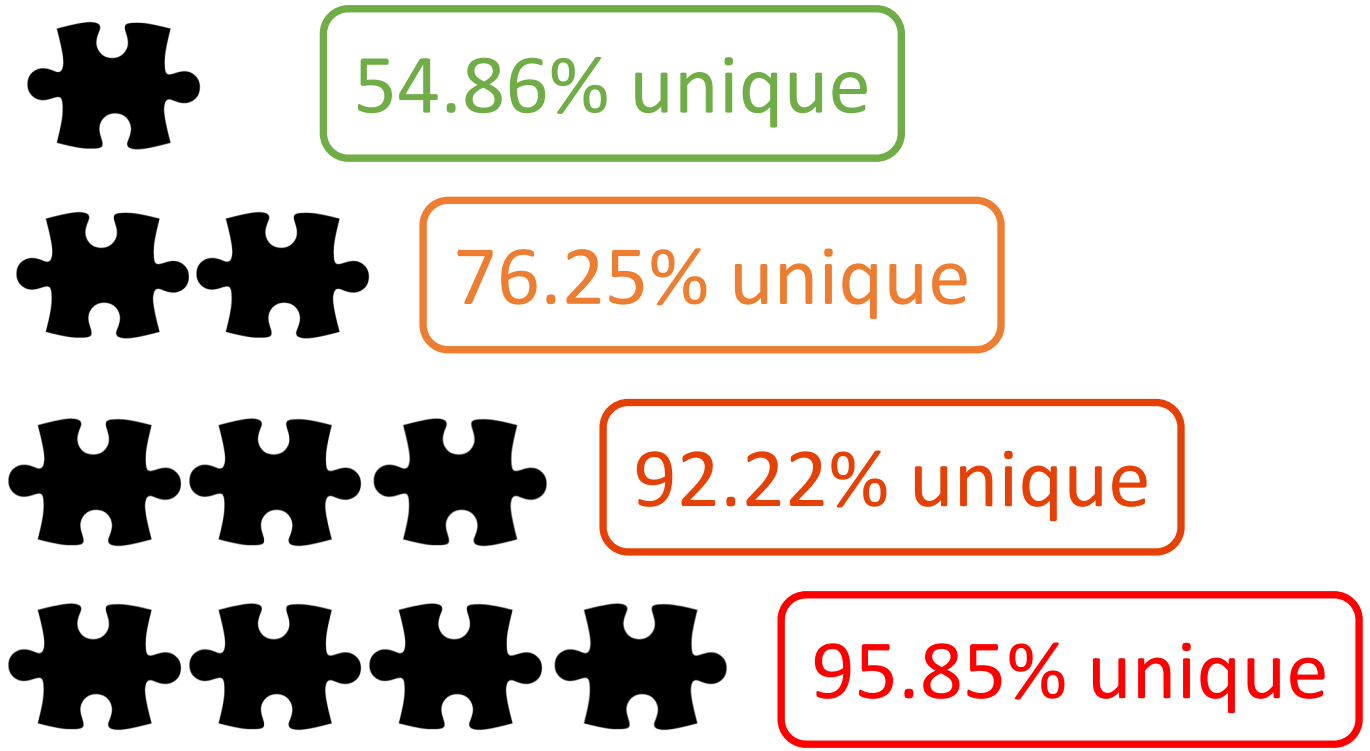


- Websites a user is logged in
- 58 websites

How unique is your browser?



Am I really unique if I use a few extensions?



G.G. Gulyás, D. F. Some, **N. Bielova** and C. Castelluccia. To Extend or not to Extend: on the Uniqueness of Browser Extensions and Web Logins. *WPES@ACM CCS 2018.*

36

The dilemma of privacy extensions

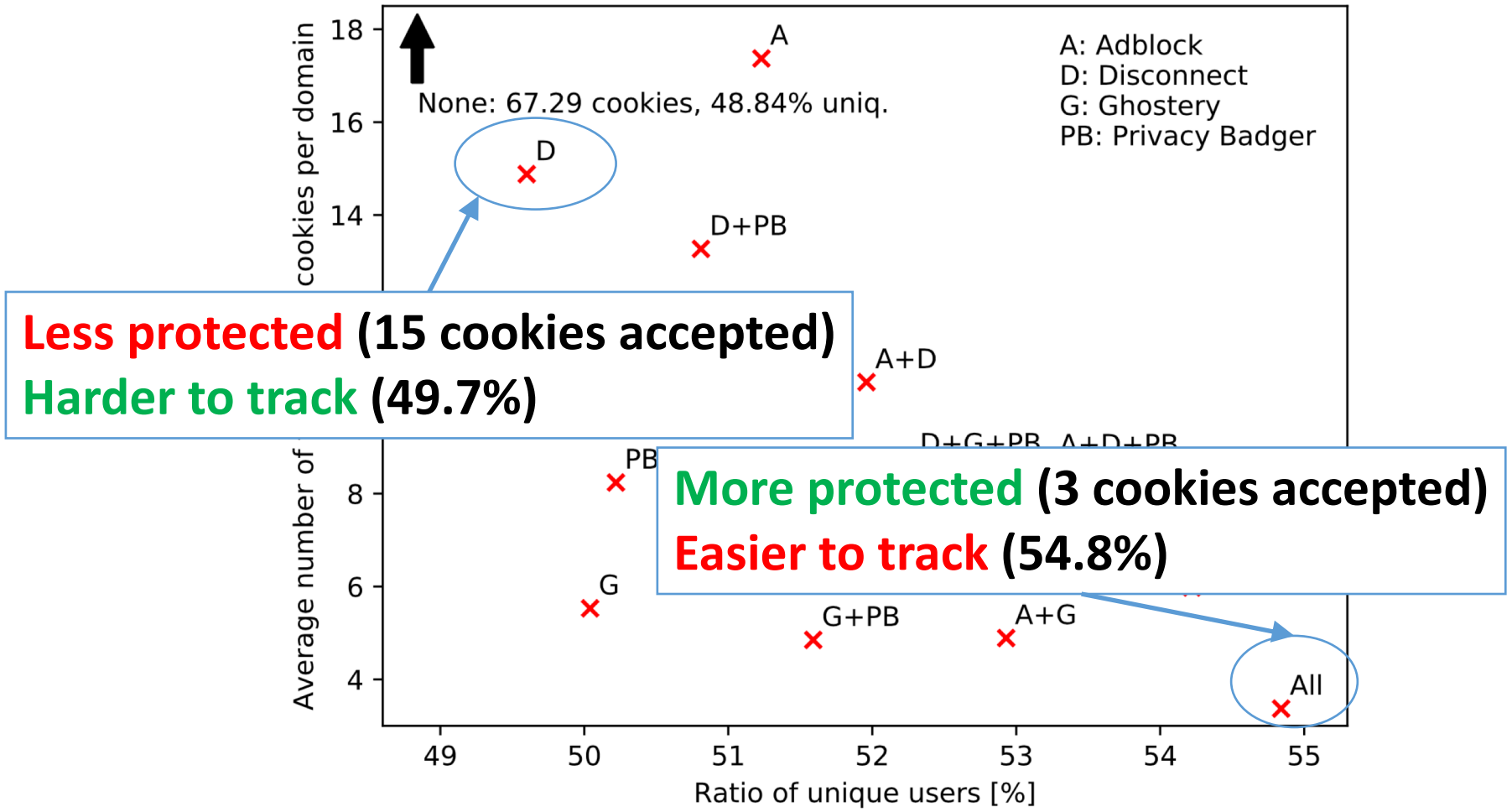
- Privacy extensions **block some trackers**
- Privacy extensions **make a user more unique**

- What is the trade-off between **privacy gain** (some trackers are blocked) and **privacy loss** (user is more unique)?

G.G. Gulyás, D. F. Some, **N. Bielova** and C. Castelluccia. To Extend or not to Extend: on the Uniqueness of Browser Extensions and Web Logins. *WPES@ACM CCS 2018.*

37

Uniqueness of users vs. number of accepted third-party cookies



*4,000 pages crawled

38