Browser fingerprinting: current research and the years ahead

Pierre Laperdrix

GT PVP-SSLR

June, 22th 2022

CRIStAL
Centre de Recherche en Informatique,
Signal et Automatique de Lille

About me

- CNRS researcher in the Spirals team in the UMR CRIStAL in Lille
- Working on web security and privacy: browser fingerprinting, web tracking, history sniffing, application debloating, mobile application security…
- Open positions for internships and PhDs in the team! Don’t hesitate to contact us!

Outline

I. What is browser fingerprinting? How to protect against it?

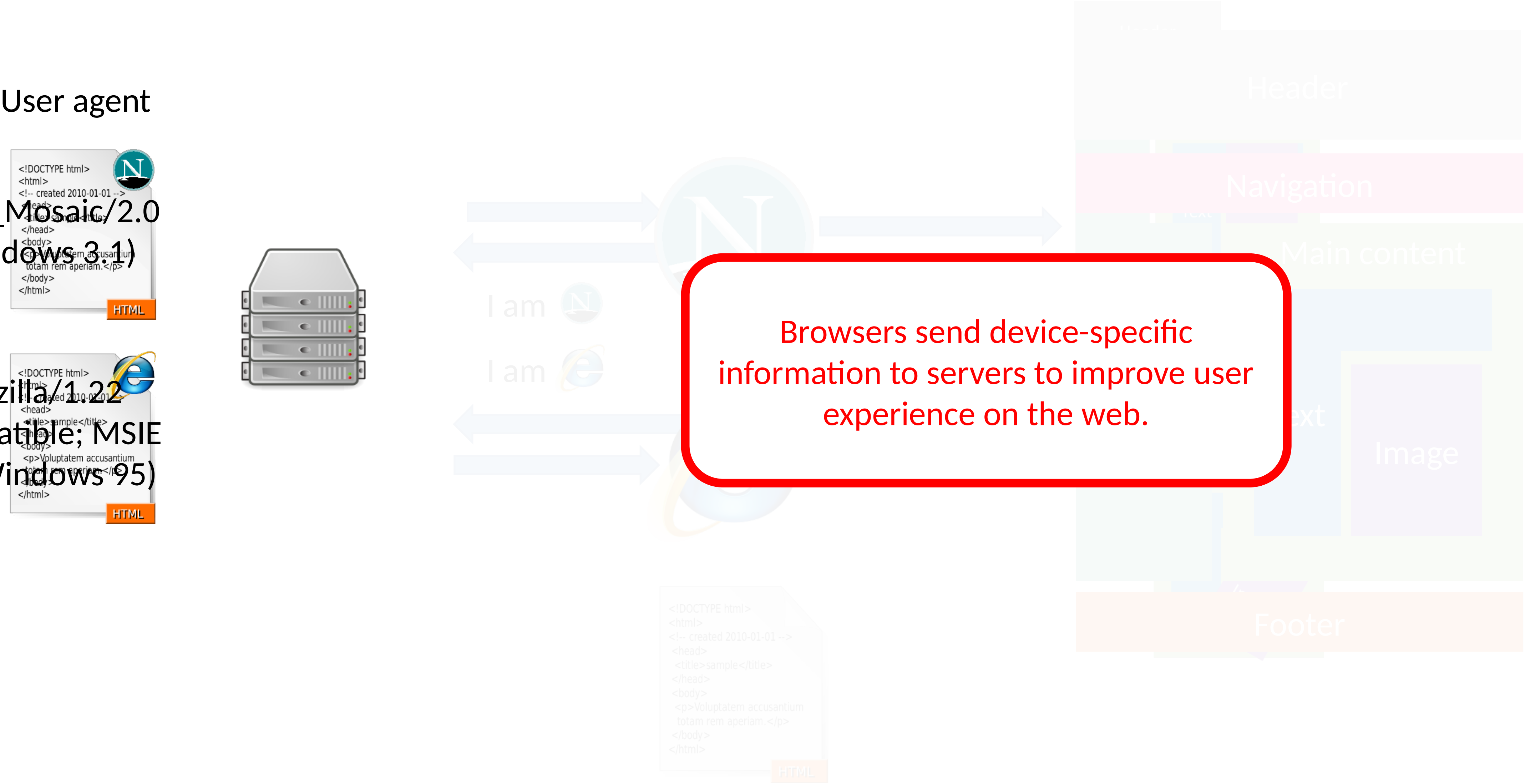
II. What is currently being done in the fingerprinting domain?

III. What to expect in the future?

I. Internet and web browsers

I. Internet in 1995

HTTP User agent

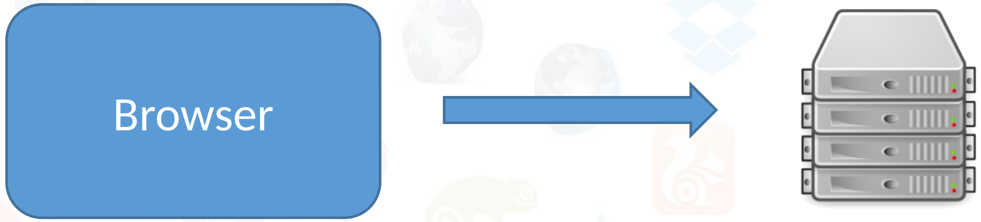


NCSA_Mosaic/2.0 (Windows 3.1)

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

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

I. Internet in 2022



1995	2022
Browser: Netscape Language: Fr	Browser: Chrome v100 OS: Linux Screen: 1920x1080 Language: Fr Timezone: GMT+1 Graphic card: GTX 3090 …

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?

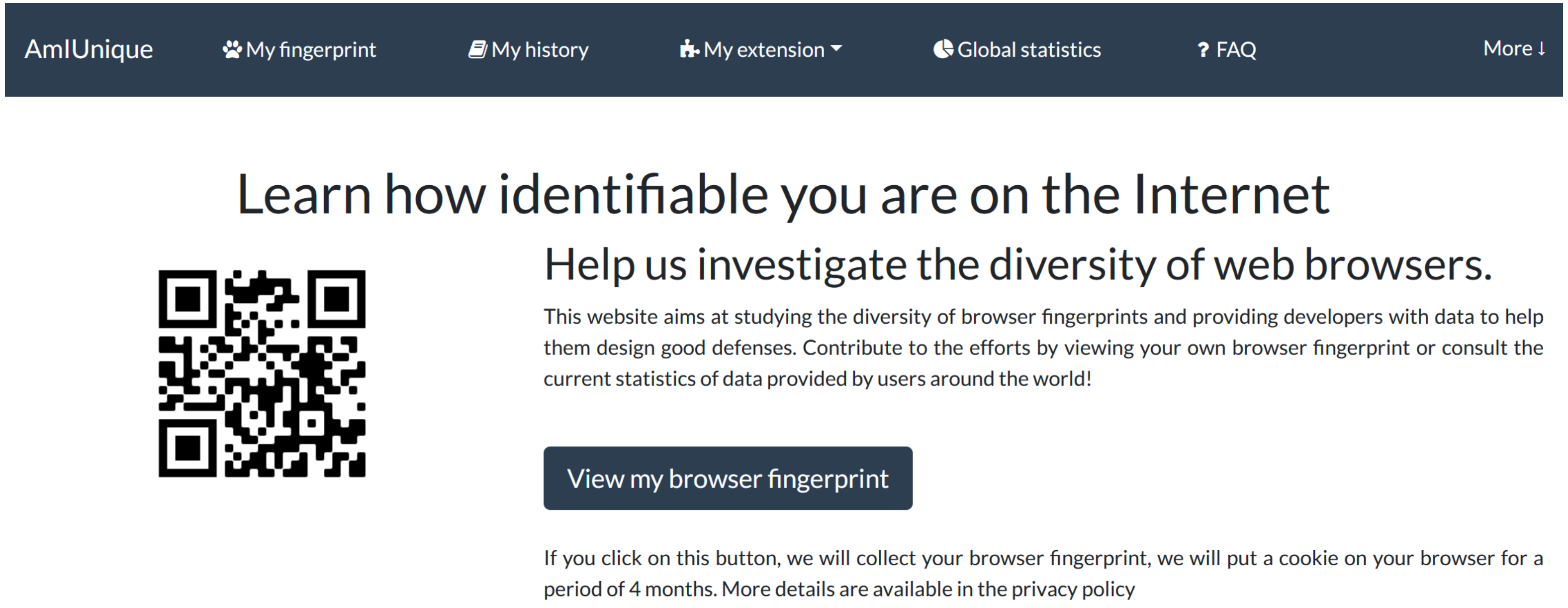
I. Definition of browser fingerprinting

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.

I. See your own fingerprint

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



Learn how identifiable you are on the Internet

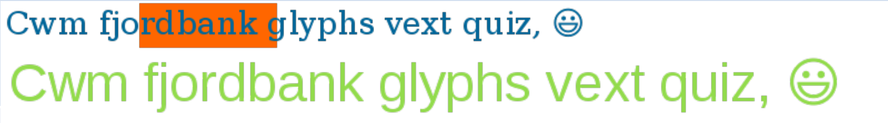
Help us investigate the diversity of web browsers.

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!

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

- Website launched in November 2014
- Collected 5,000,000+ fingerprints so far
- Browser extension available to see the evolution of your own browser fingerprint

I. Example of a browser fingerprint

Attribute	Value
User agent	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/102.0.0.0 Safari/537.36
HTTP headers	text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9 gzip, deflate, br en-US,en;q=0.9
Fonts	Century Schoolbook, Source Sans Pro Light, DejaVu Sans Mono, Bitstream Vera Serif, URW Palladio L, Bitstream Vera Sans Mono, Bitstream Vera Sans, …
Platform	Win32
Screen resolution	3840x2160x24
Timezone	-480 (UTC+8)
Hardware concurrency	4
Battery level	38%
WebGL vendor	NVIDIA Corporation
WebGL renderer	GeForce GTX 3070 Ti/PCIe/SSE2
Canvas	 Cwm fjordbank glyphs vext quiz, 😃 Cwm fjordbank glyphs vext quiz, 😃
Browser extensions	kindle

Windows 10

4K

GEFORCE GTX
NVIDIA

I. Impact on privacy

What makes fingerprinting a threat to online privacy?

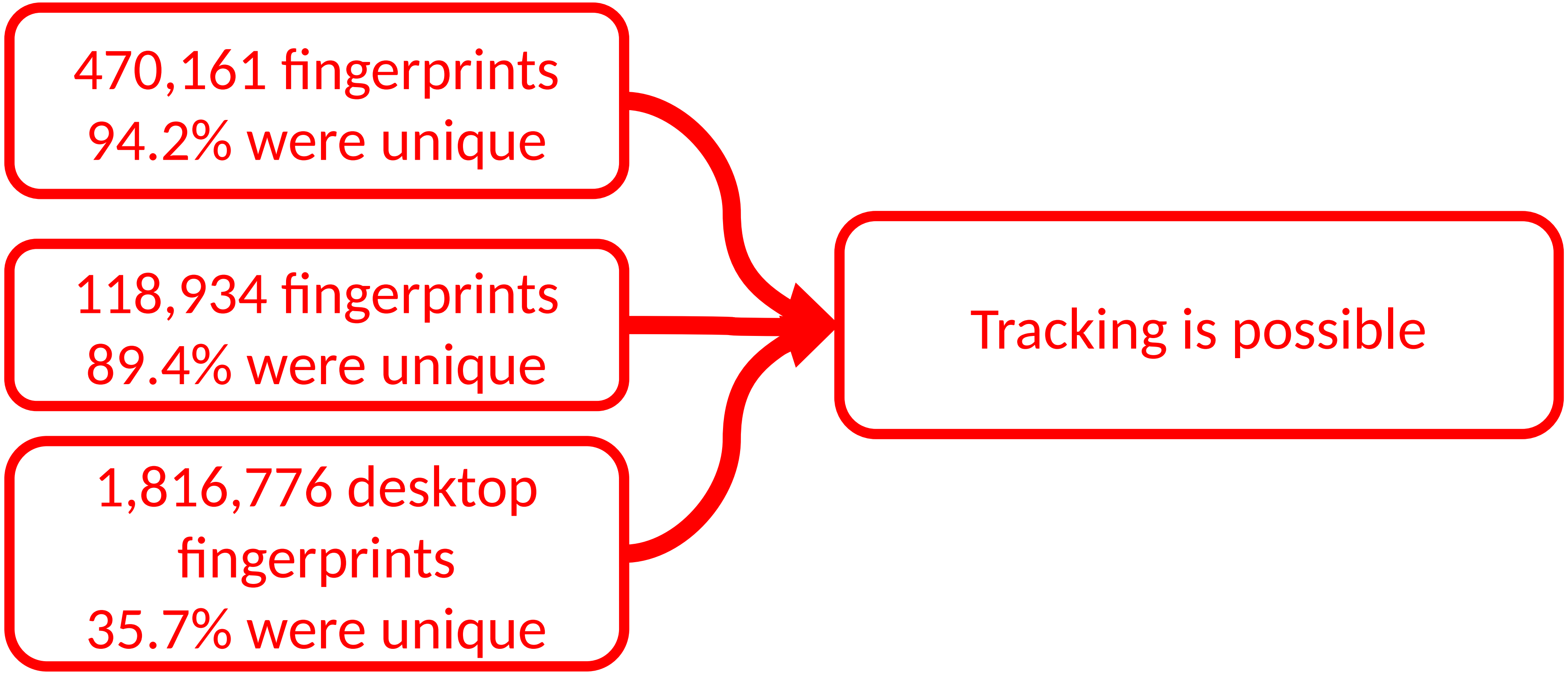
1. It is really easy to collect all this data. No need for extra permissions.
2. Several studies have investigated the diversity of browser fingerprints.

Panopticlick
How Unique — and Trackable — Is Your Browser?

EFF

Am I Unique?

Study "Hiding in the Crowd"



470,161 fingerprints
94.2% were unique

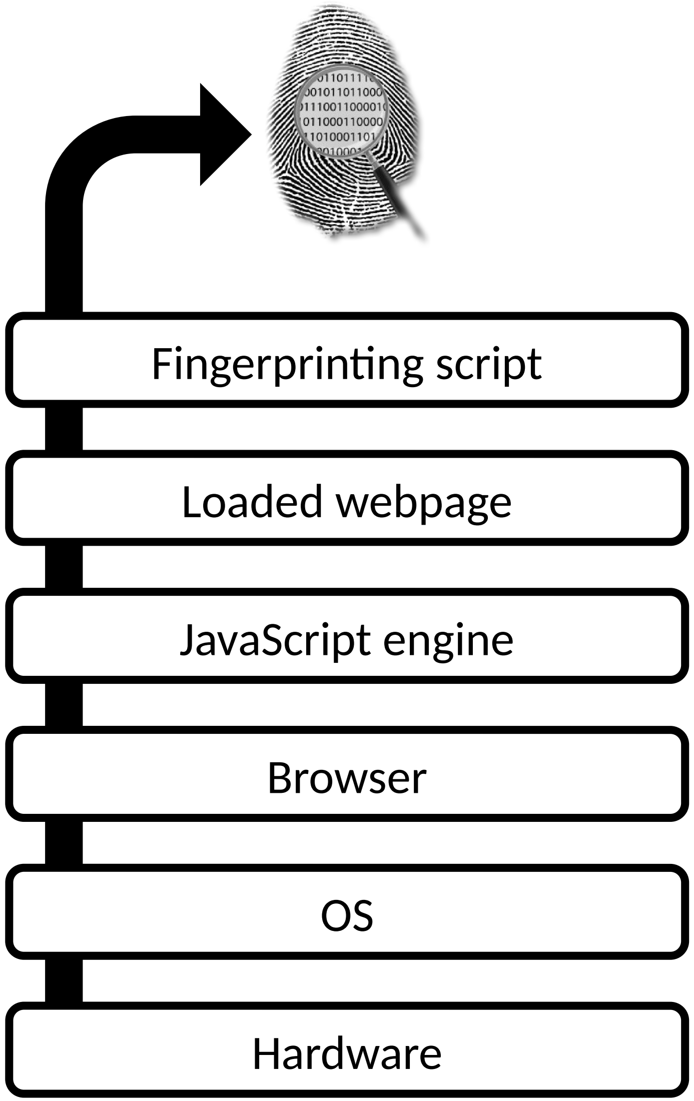
118,934 fingerprints
89.4% were unique

1,816,776 desktop fingerprints
35.7% were unique

Tracking is possible

I. Protection against fingerprinting

- Goal: to protect users against browser fingerprinting, i.e. to prevent them from being tracked online



Fingerprinting script

Loaded webpage

JavaScript engine

Browser

OS

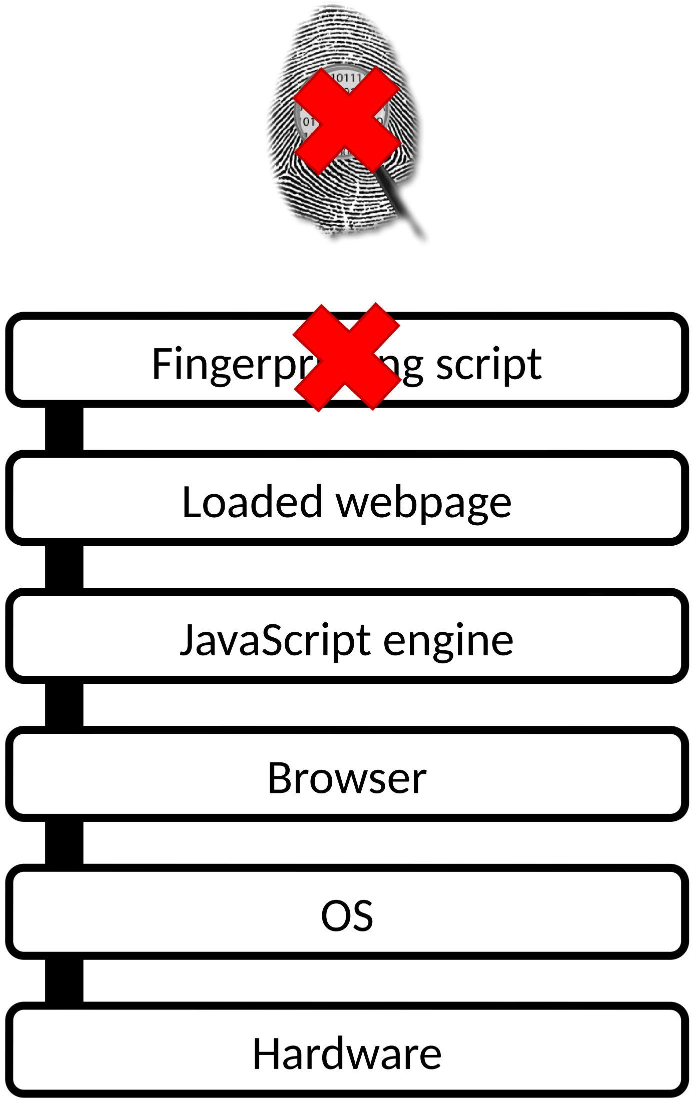
Hardware

I. Protection against fingerprinting - Blocking scripts

- The fingerprinting script is simply not executed.
- Some existing solutions

Browser extensions

Browser with built-in protection



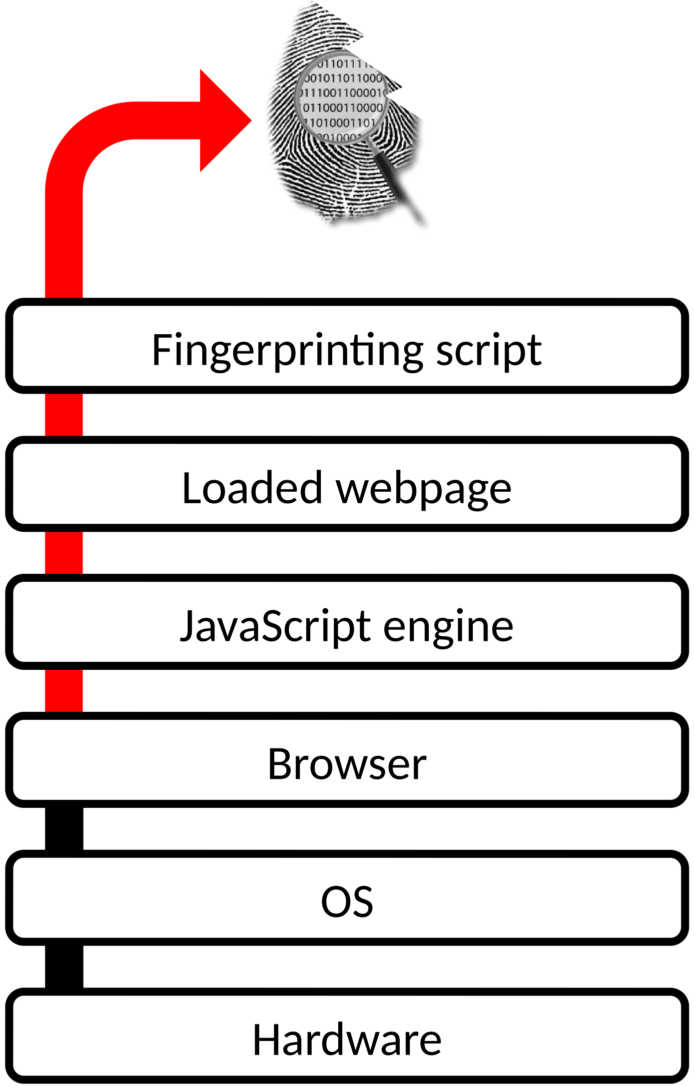
I. Protection against fingerprinting - Blocking browser APIs

- The fingerprinting script will collect less information.
- Some existing solutions

CanvasBlocker

Brave

Tor browser



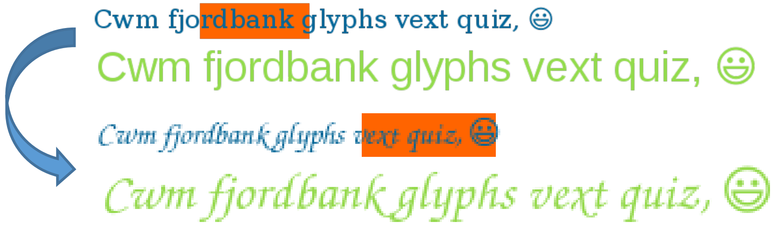
I. Protection against fingerprinting - Injecting JavaScript

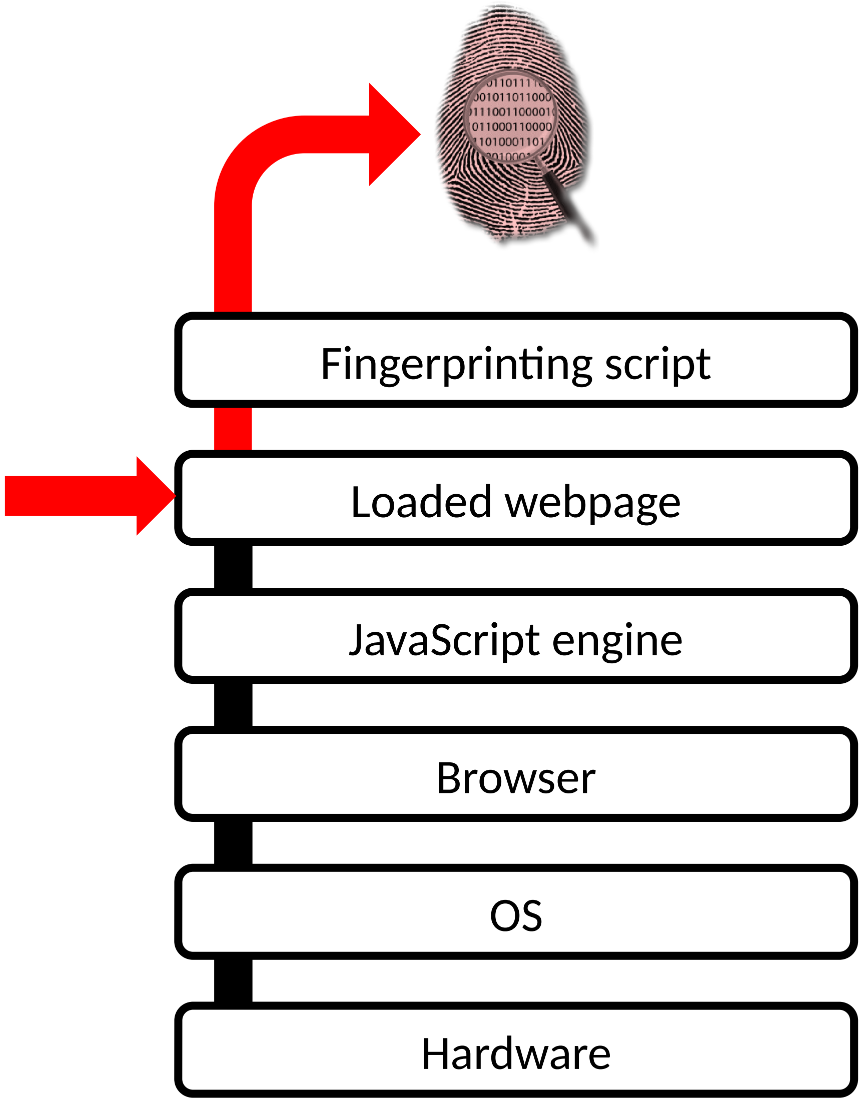
- The injection of JavaScript overwrites the default methods of the JavaScript engine.
- Can change values

  “navigator.platform”

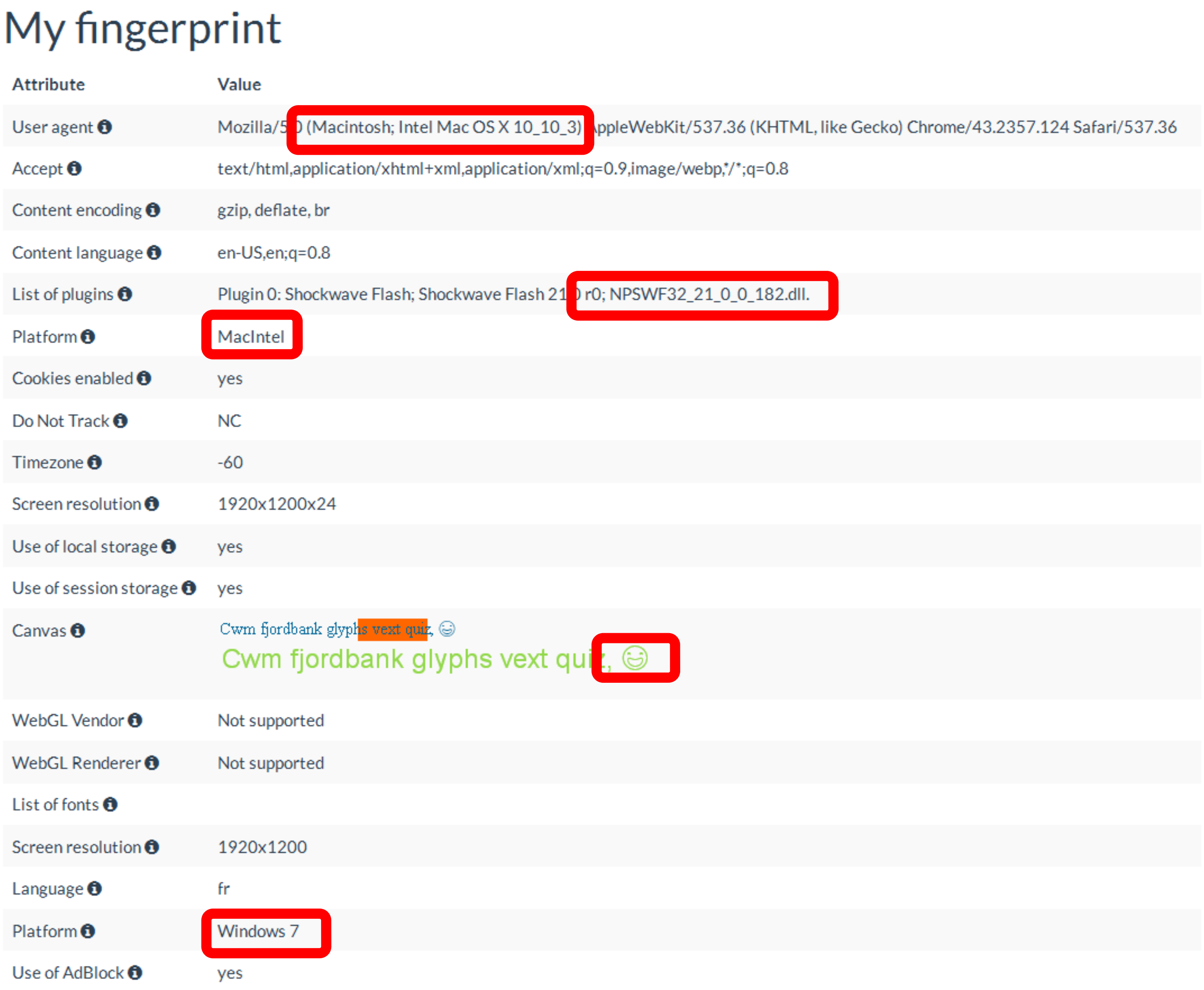
  →Default: “Win64”

  →New value: “Linux x86_64”
- Can inject noise


Cwm fjordbank glyphs vext quiz, 😃
Cwm fjordbank glyphs vext quiz, 😃
Cwm fjordbank glyphs vext quiz, 😃
Cwm fjordbank glyphs vext quiz, 😃



I. Protection against fingerprinting - The problem of inconsistencies



My fingerprint

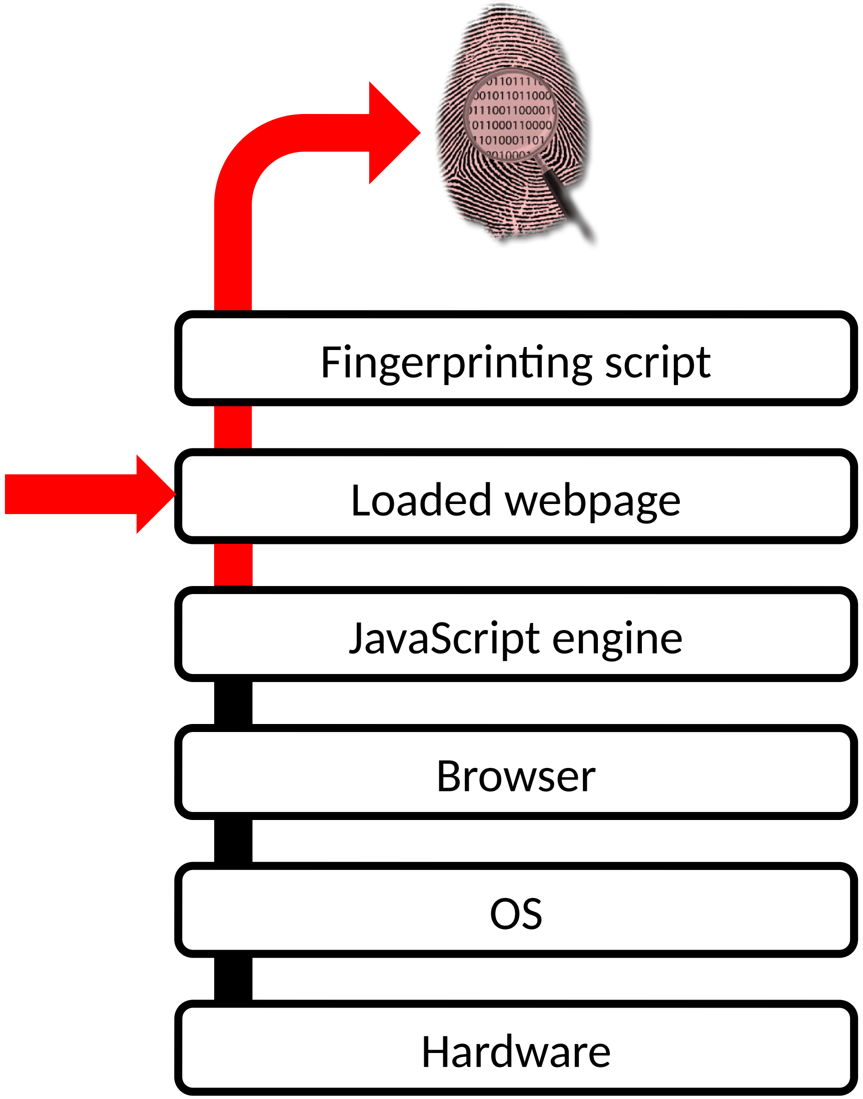
Attribute	Value
User agent	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.2357.124 Safari/537.36
Accept	text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Content encoding	gzip, deflate, br
Content language	en-US,en;q=0.8
List of plugins	Plugin 0: Shockwave Flash; Shockwave Flash 21.0 r0; NPSWF32_21_0_0_182.dll.
Platform	MacIntel
Cookies enabled	yes
Do Not Track	NC
Timezone	-60
Screen resolution	1920x1200x24
Use of local storage	yes
Use of session storage	yes
Canvas	Cwm fjordbank glyphs vext quiz, ☺ Cwm fjordbank glyphs vext quiz, ☺
WebGL Vendor	Not supported
WebGL Renderer	Not supported
List of fonts	
Screen resolution	1920x1200
Language	fr
Platform	Windows 7
Use of AdBlock	yes

I. Protection against fingerprinting - Native spoofing

- Instead of injecting JavaScript, the source code of the browser is modified to send new values.
- Investigating JS objects is not enough to detect the modifications.
- Some existing solutions

Mimic privacy browser

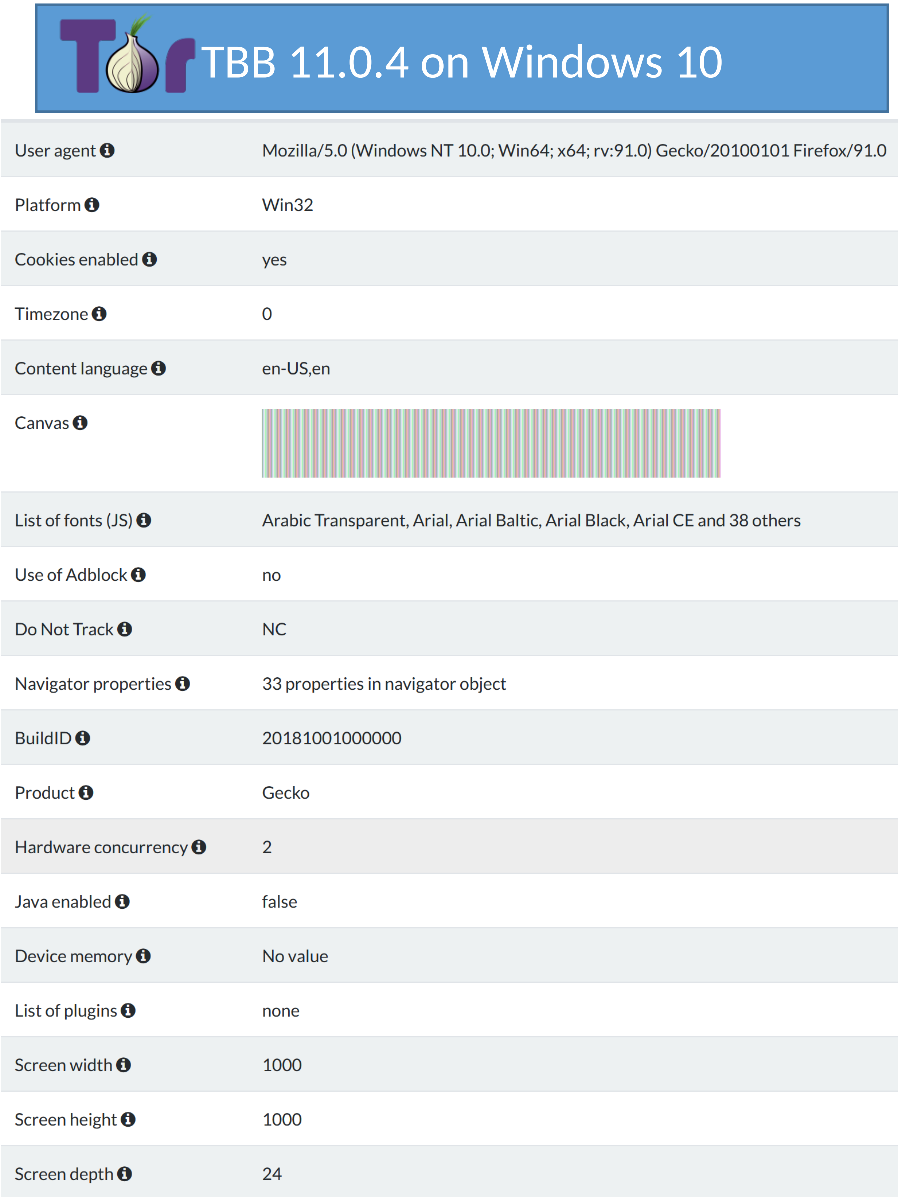
Tor browser



I. Protection against fingerprinting - Tor browser and its fingerprint

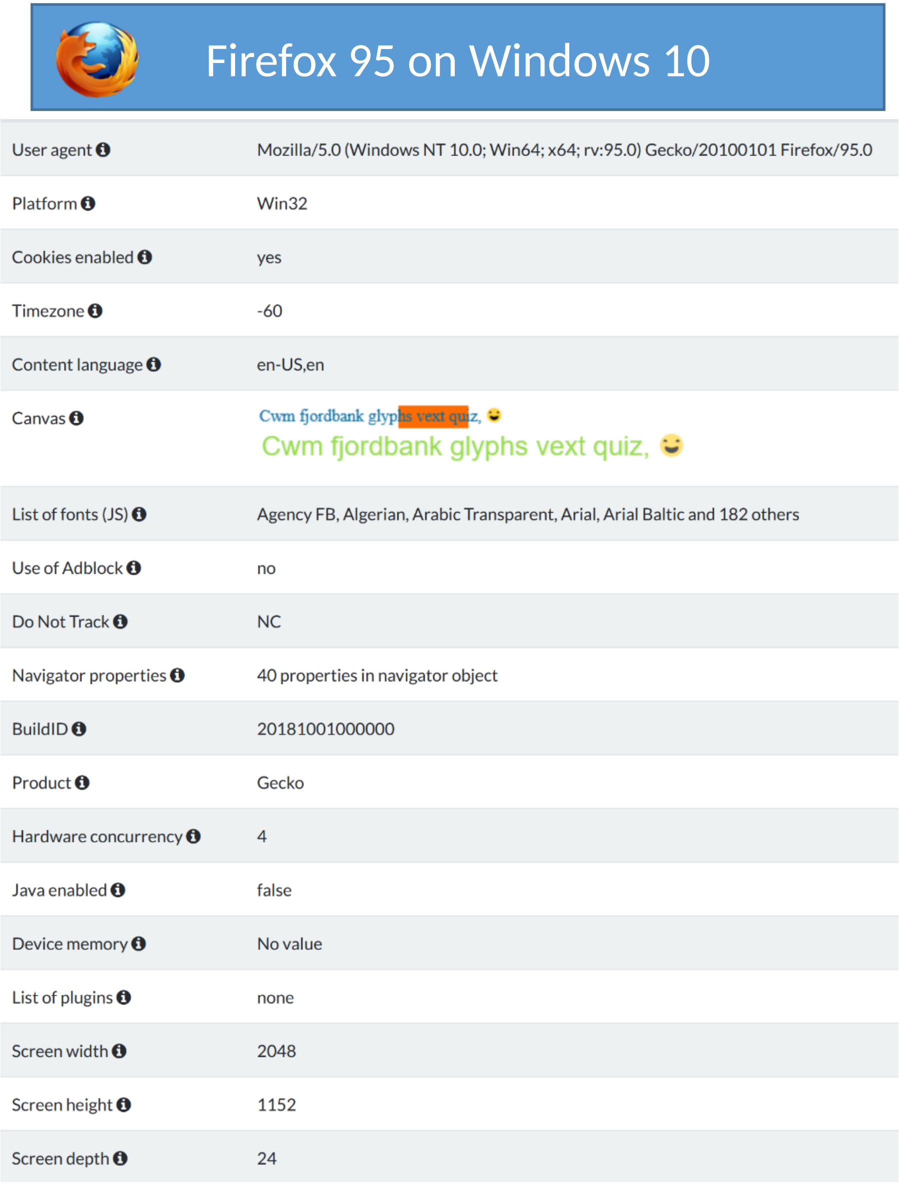
- In theory, all fingerprints from the Tor Browser should be identical.
- In reality, differences can still be found (screen resolution, fonts, canvas...).

TBB 11.0.4 on Windows 10



User agent ℹ	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:91.0) Gecko/20100101 Firefox/91.0
Platform ℹ	Win32
Cookies enabled ℹ	yes
Timezone ℹ	0
Content language ℹ	en-US,en
Canvas ℹ	
List of fonts (JS) ℹ	Arabic Transparent, Arial, Arial Baltic, Arial Black, Arial CE and 38 others
Use of Adblock ℹ	no
Do Not Track ℹ	NC
Navigator properties ℹ	33 properties in navigator object
BuildID ℹ	20181001000000
Product ℹ	Gecko
Hardware concurrency ℹ	2
Java enabled ℹ	false
Device memory ℹ	No value
List of plugins ℹ	none
Screen width ℹ	1000
Screen height ℹ	1000
Screen depth ℹ	24

Firefox 95 on Windows 10



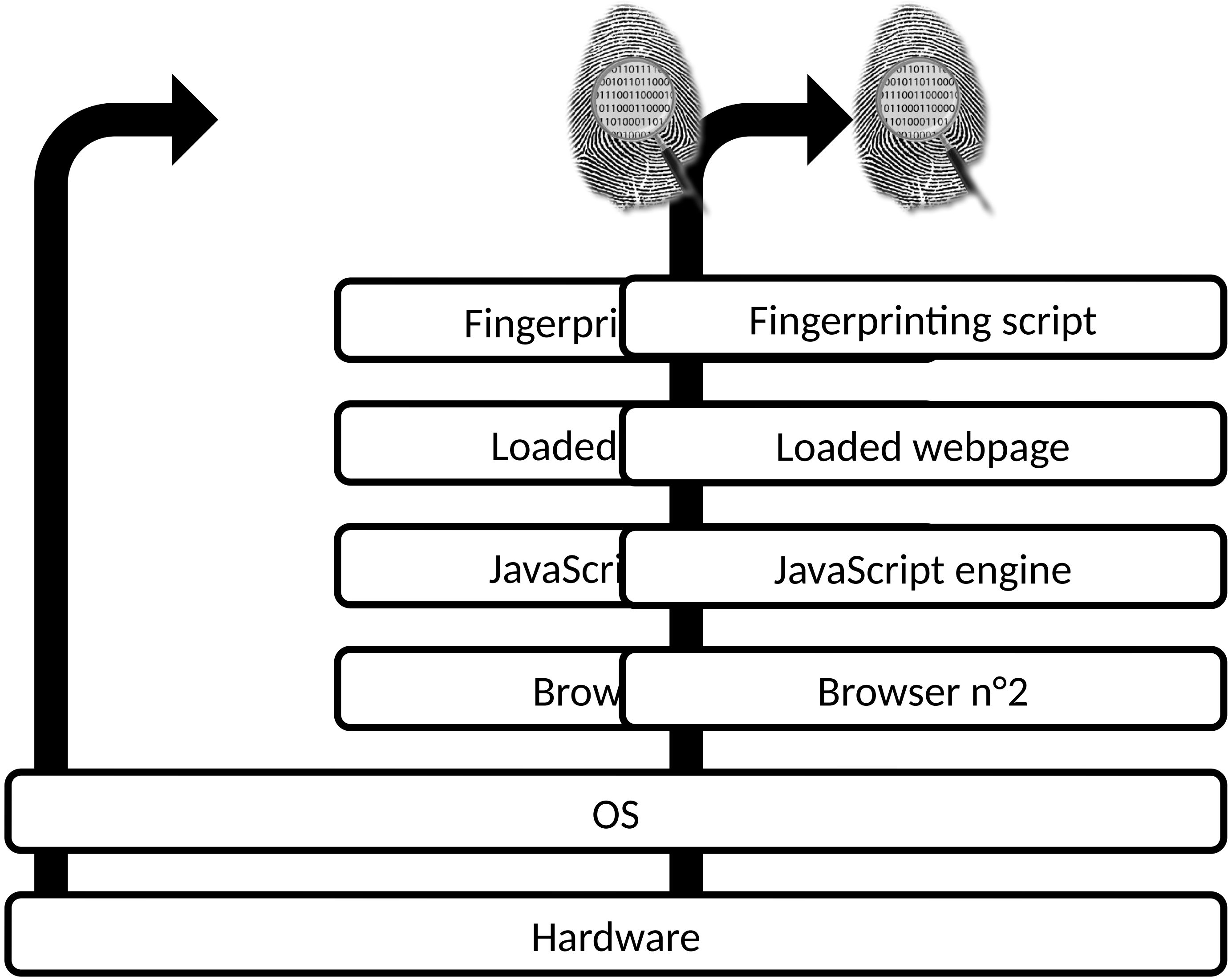
User agent ℹ	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:95.0) Gecko/20100101 Firefox/95.0
Platform ℹ	Win32
Cookies enabled ℹ	yes
Timezone ℹ	-60
Content language ℹ	en-US,en
Canvas ℹ	Cwm fjordbank glyphs vext quiz, 😃 Cwm fjordbank glyphs vext quiz, 😃
List of fonts (JS) ℹ	Agency FB, Algerian, Arabic Transparent, Arial, Arial Baltic and 182 others
Use of Adblock ℹ	no
Do Not Track ℹ	NC
Navigator properties ℹ	40 properties in navigator object
BuildID ℹ	20181001000000
Product ℹ	Gecko
Hardware concurrency ℹ	4
Java enabled ℹ	false
Device memory ℹ	No value
List of plugins ℹ	none
Screen width ℹ	2048
Screen height ℹ	1152
Screen depth ℹ	24

I. Protection against fingerprinting - Changing browsers

- One fingerprint for each browser
- The OS and Hardware layers are shared by both fingerprints.

⬇

- If you collect enough information on the OS and hardware, you are prone to **cross-browser fingerprinting**.

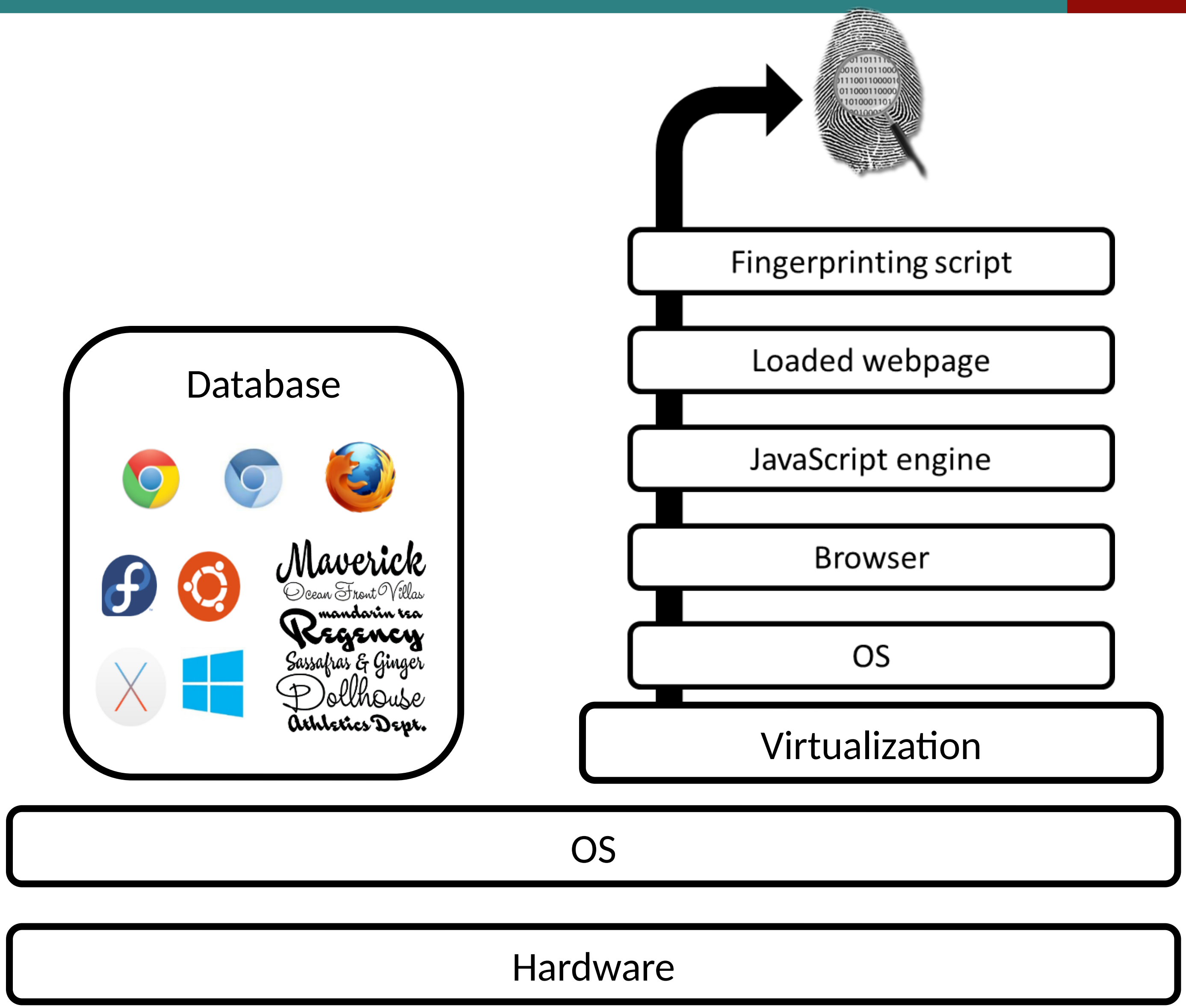


I. Recreating a complete environment

- Disposable environments with a unique fingerprint for each browsing session

- Database with different OS, fonts, plugins and browsers

- Use of virtualization to isolate the host OS from the new environment



Database

Fingerprinting script

Loaded webpage

JavaScript engine

Browser

OS

Virtualization

OS

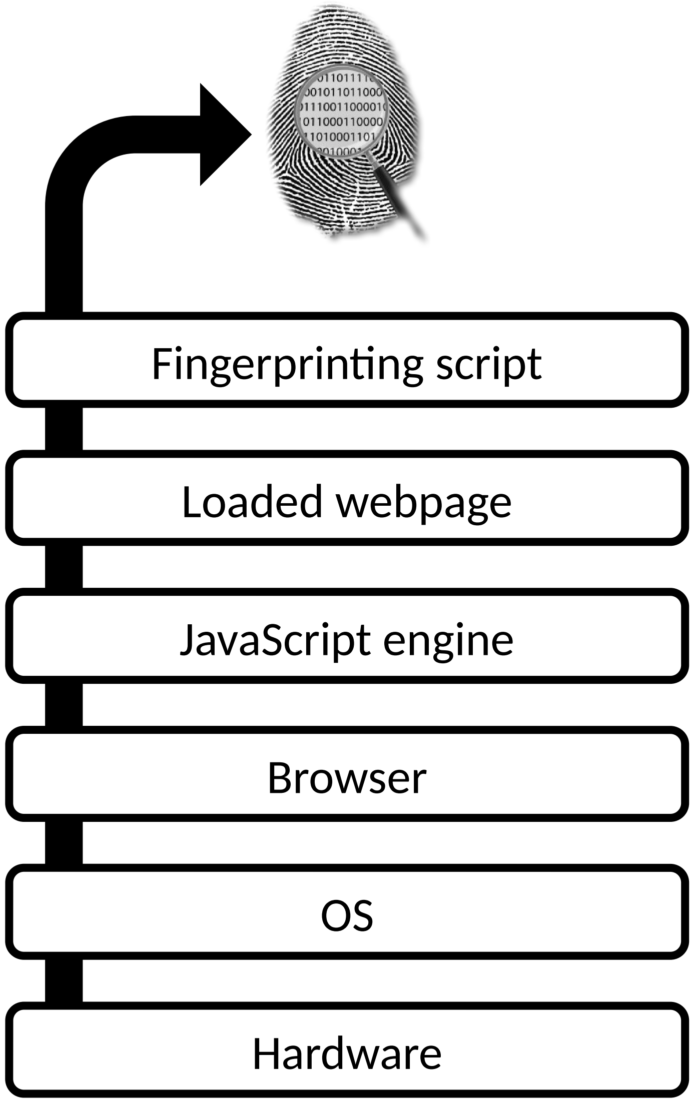
Hardware

I. Protection against fingerprinting

Many different approaches:

- Blocking scripts
- Blocking browser APIs
- Injecting JavaScript
- Native spoofing
- Changing browsers
- Recreating complete environments

Each technique has its strengths and weaknesses.



Fingerprinting script

Loaded webpage

JavaScript engine

Browser

OS

Hardware

Outline

I. What is browser fingerprinting? How to protect against it?

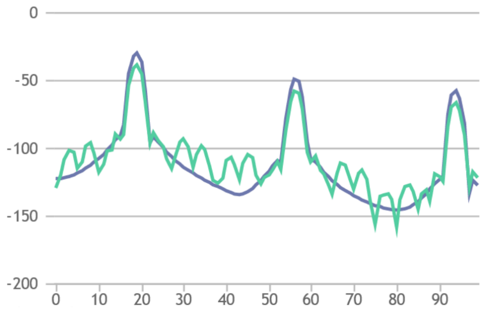
II. What is currently being done in the fingerprinting domain?

III. What to expect in the future?

II. Current research – Going beyond browser APIs

To increase the number of attributes in fingerprints, researches are trying to go beyond what’s offered by browser APIs.

- Browser extensions
- Web Audio fingerprinting

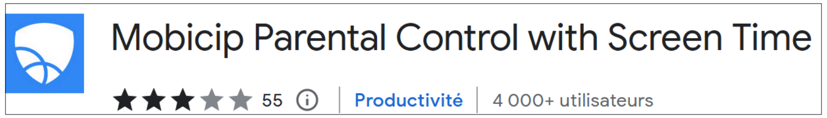


- Hardware fingerprinting

II. Current research – Detecting browser extensions

Detecting extensions poses a threat to online privacy because:

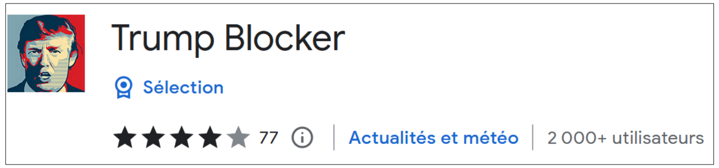
- The list of installed extensions can reinforce fingerprinting and make user unique on the web.
- It can reveal user’s preferences, browsing habits or demographic information.



Mobicip Parental Control with Screen Time

★★★☆☆ 55 ⓘ | Productivité | 4 000+ utilisateurs

Kids

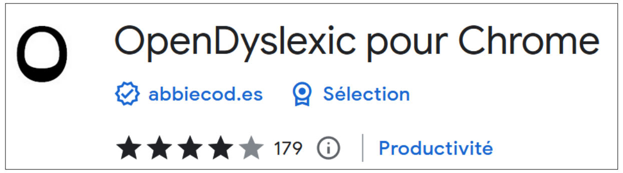


Trump Blocker

Sélection

★★★★☆ 77 ⓘ | Actualités et météo | 2 000+ utilisateurs

Politics

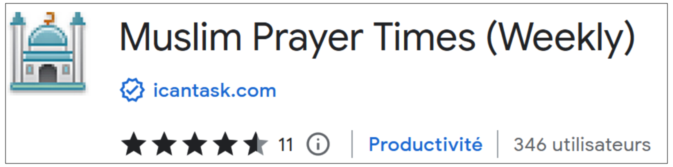


OpenDyslexic pour Chrome

abbiecod.es Sélection

★★★★☆ 179 ⓘ | Productivité

Disability

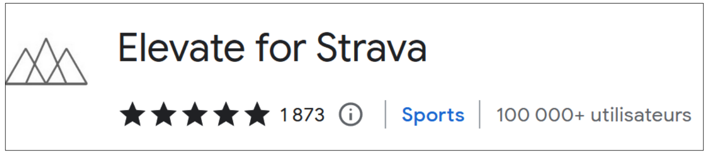


Muslim Prayer Times (Weekly)

icantask.com

★★★★★ 11 ⓘ | Productivité | 346 utilisateurs

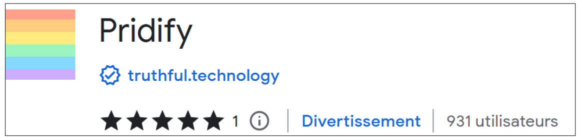
Religion



Elevate for Strava

★★★★★ 1 873 ⓘ | Sports | 100 000+ utilisateurs

Fitness



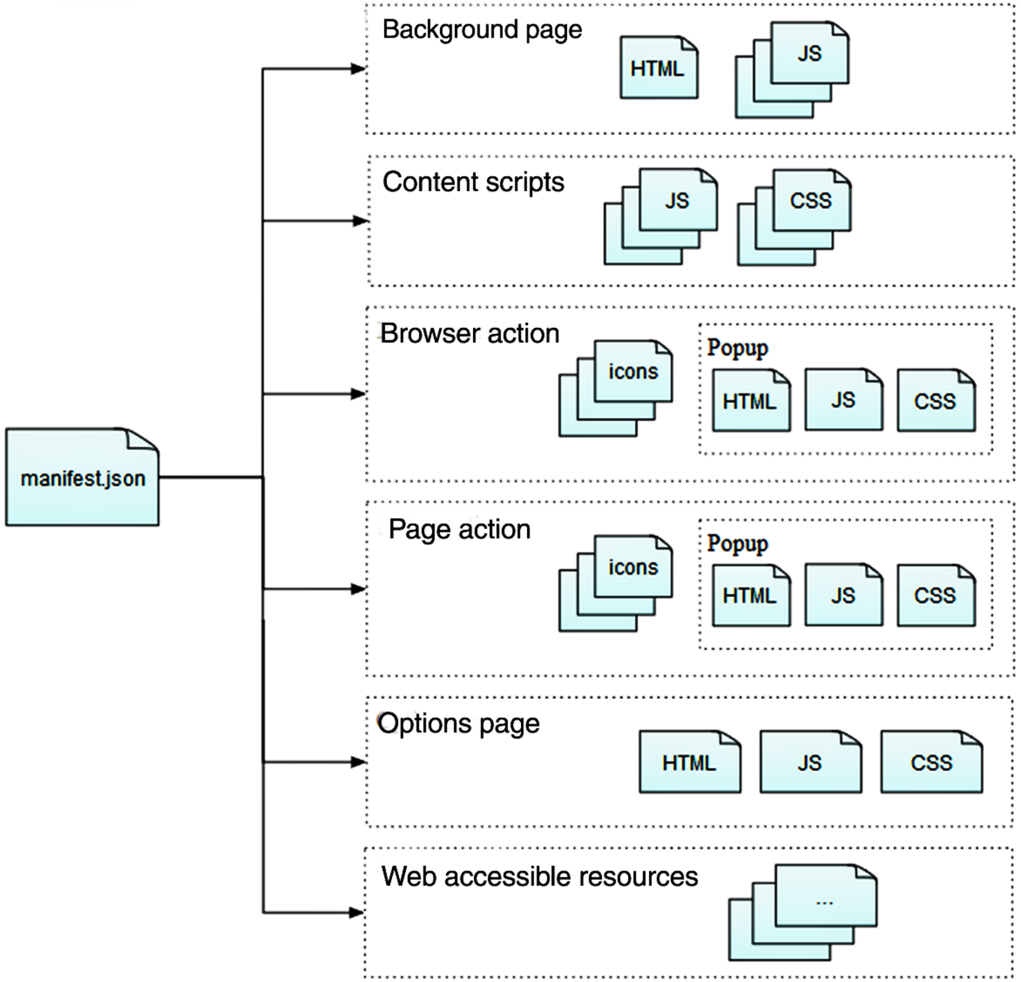
Pridify

truthful.technology

★★★★★ 1 ⓘ | Divertissement | 931 utilisateurs

Gender/Sexuality

II. Current research – Architecture of a browser extension



Source: MDN Web Docs

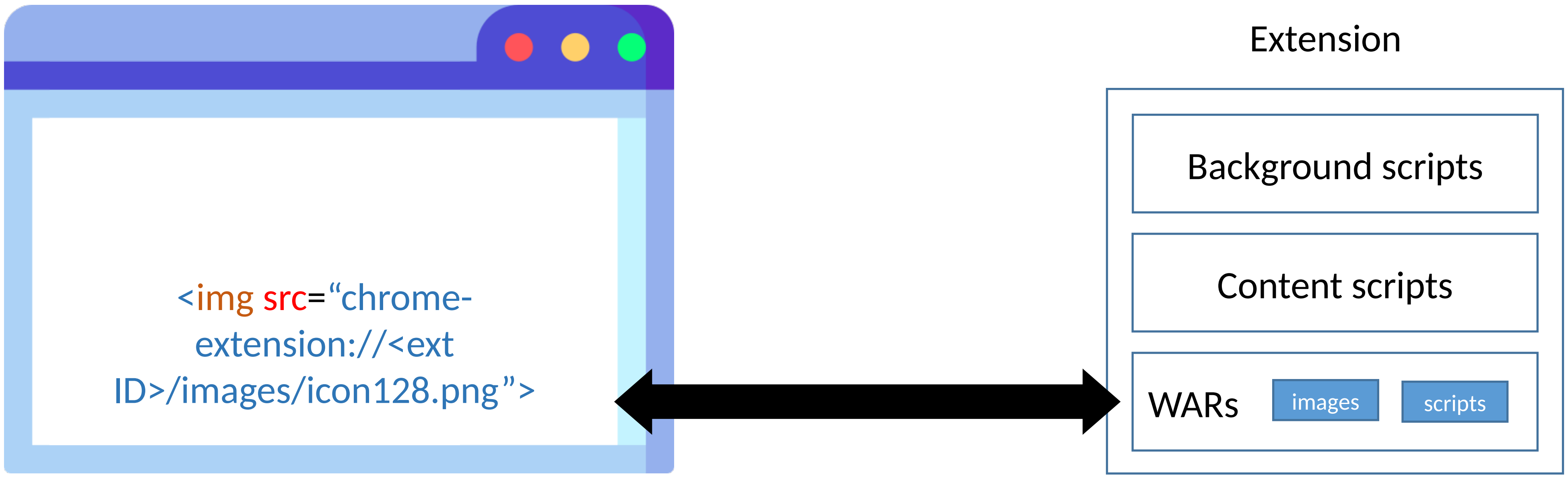
Structure of a browser extension

- Manifest.json is a mandatory file that provides metadata information on how the extension works.
- Background page implements long-term logic.
- Content scripts are scripts that are injected into visited webpages.
- Web accessible resources (WARs) are files like JS libraries or icons that can be accessed by the extension or any webpage.

II. Current research – Detecting browser extensions

1st method: WAR fingerprinting (2017)

- Probes specific WARs in the browser to identify an extension.
- Requires knowledge beforehand of extension IDs and paths of WAR files.

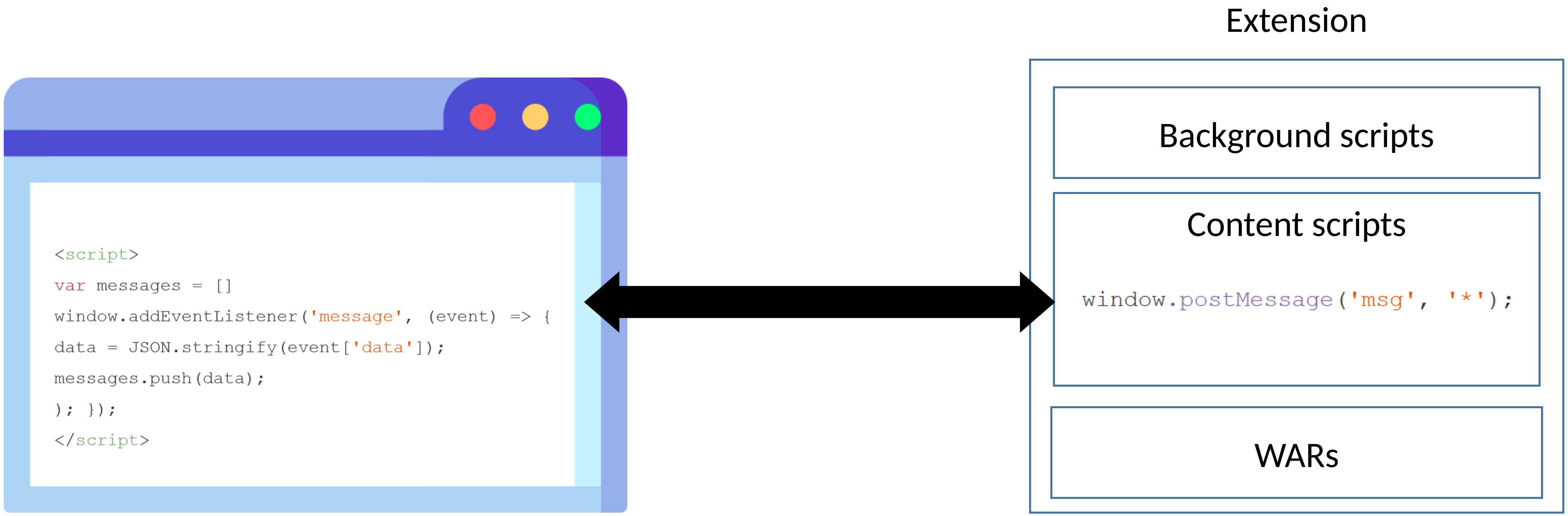


```
<img src="chrome-extension://<ext ID>/images/icon128.png">
```

II. Current research – Detecting browser extensions

2nd method: Intra/Inter communication fingerprinting (2020)

Extensions as part of their inner-workings exchange messages between components.



```
<script>
var messages = []
window.addEventListener('message', (event) => {
data = JSON.stringify(event['data']);
messages.push(data);
); });
</script>
```

Extension

Background scripts

Content scripts

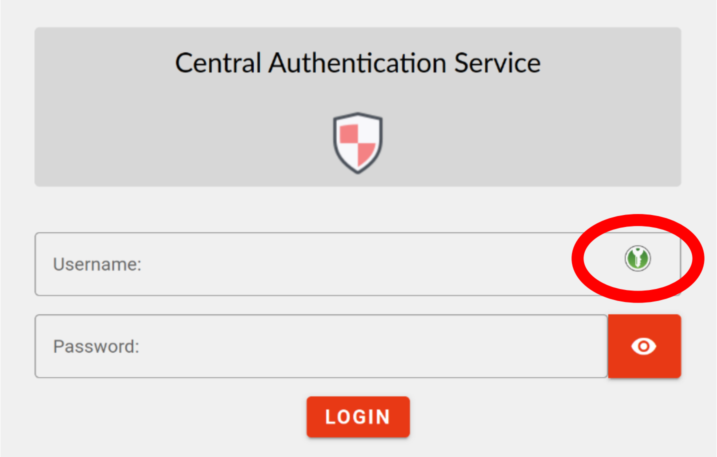
```
window.postMessage('msg', '*');
```

WARs

II. Current research – Detecting browser extensions

3rd method: Behavioral fingerprinting

A) Default behavior: Extensions might add/remove buttons, text or images on a webpage without any interaction (2017).

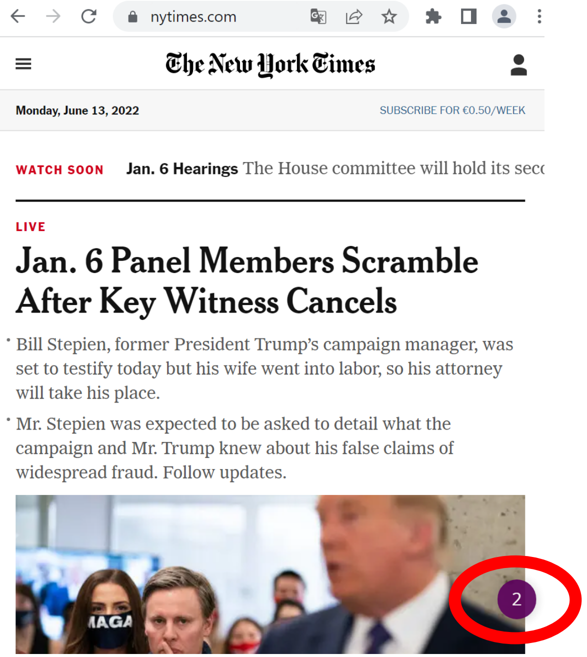


Central Authentication Service

Username:

Password:

KeePassXC-Browser



Monday, June 13, 2022 SUBSCRIBE FOR €0.50/WEEK

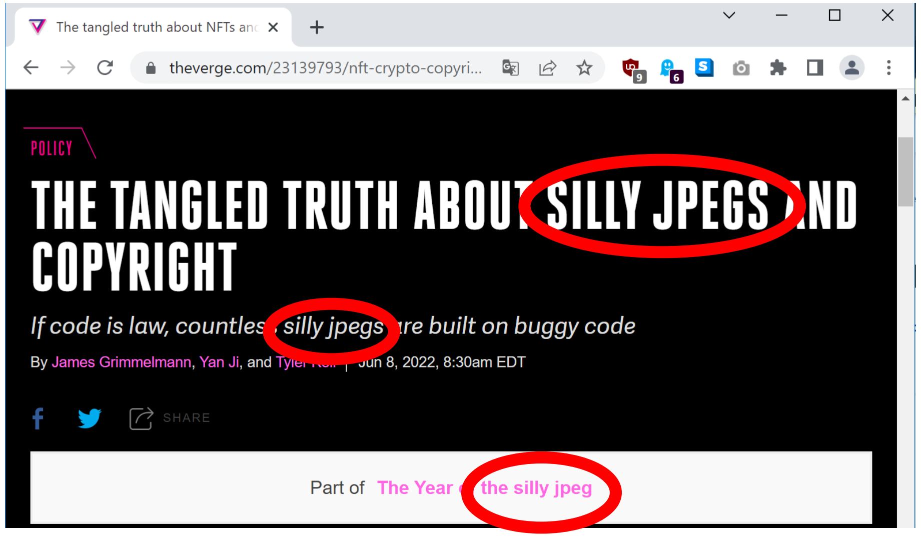
WATCH SOON Jan. 6 Hearings The House committee will hold its seco

LIVE

Jan. 6 Panel Members Scramble After Key Witness Cancels

- Bill Stepien, former President Trump's campaign manager, was set to testify today but his wife went into labor, so his attorney will take his place.
- Mr. Stepien was expected to be asked to detail what the campaign and Mr. Trump knew about his false claims of widespread fraud. Follow updates.

Ghostery



The tangled truth about NFTs an
theverge.com/23139793/nft-crypto-copyri...
POLICY
SHARE

THE TANGLED TRUTH ABOUT SILLY JPEGS AND COPYRIGHT

If code is law, countless silly jpegs are built on buggy code

By James Grimmelmann, Yan Ji, and Tyler Kell | Jun 8, 2022, 8:30am EDT

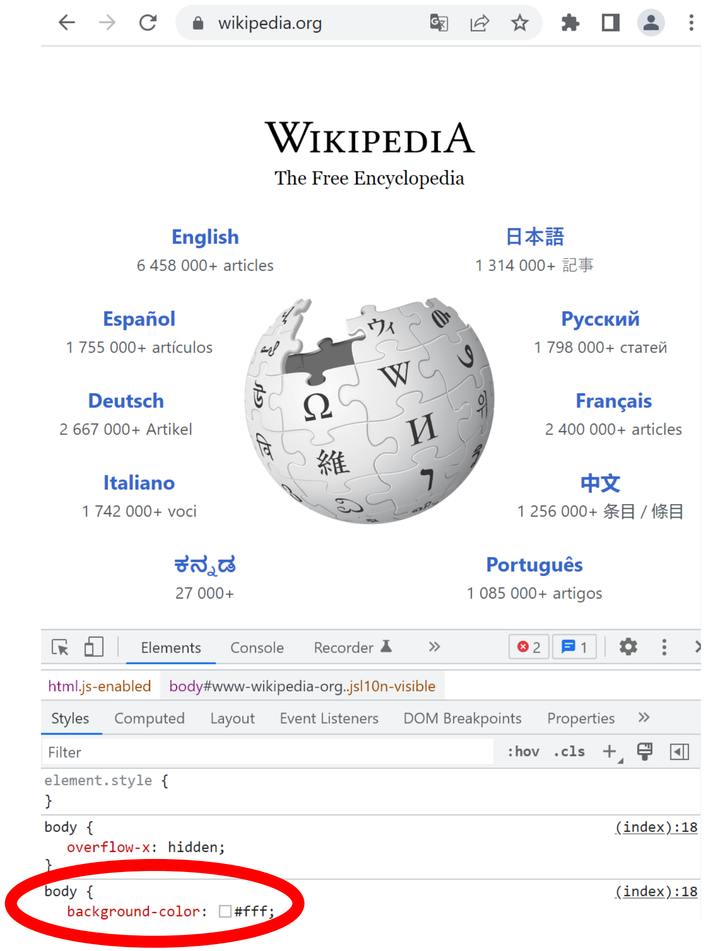
Part of The Year of the silly jpeg

NFT to silly jpeg

II. Current research – Detecting browser extensions

B) Style fingerprinting:

Extensions can modify the style of elements on the page (2021).



WIKIPEDIA
The Free Encyclopedia

English
6 458 000+ articles

日本語
1 314 000+ 記事

Español
1 755 000+ artículos

Русский
1 798 000+ статей

Deutsch
2 667 000+ Artikel

Français
2 400 000+ articles

Italiano
1 742 000+ voci

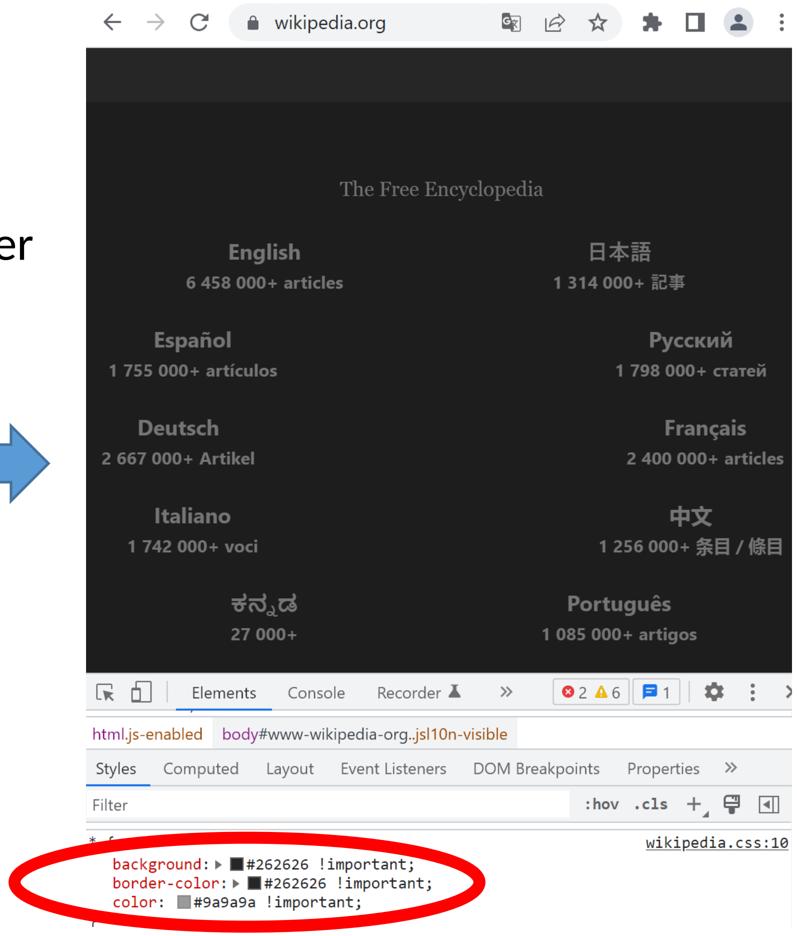
中文
1 256 000+ 条目 / 條目

ಕನ್ನಡ
27 000+

Português
1 085 000+ artigos

```
element.style {
}
body {                                   (index):18
   overflow-x: hidden;
}
body {                                   (index):18
   background-color: □#fff;
```

With the “Super Dark Mode” extension installed



The Free Encyclopedia

English
6 458 000+ articles

日本語
1 314 000+ 記事

Español
1 755 000+ artículos

Русский
1 798 000+ статей

Deutsch
2 667 000+ Artikel

Français
2 400 000+ articles

Italiano
1 742 000+ voci

中文
1 256 000+ 条目 / 條目

ಕನ್ನಡ
27 000+

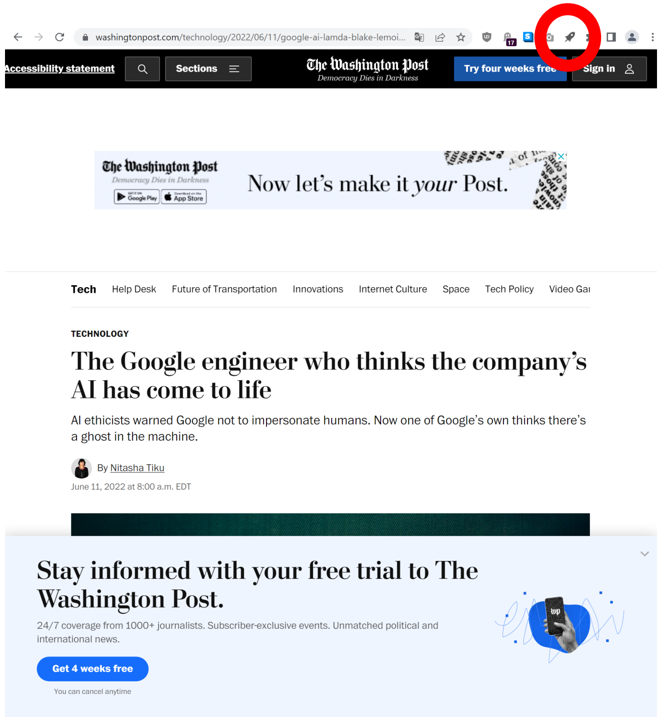
Português
1 085 000+ artigos

```
* {                                      wikipedia.css:10
   background: ▸ ■#262626 !important;
   border-color: ▸ ■#262626 !important;
   color: ■#9a9a9a !important;
```

II. Current research – Detecting browser extensions

C) Modifications after user interaction: Extensions modify the page after the user has interacted with it (2022).

Example: key presses, scrolling, mouse clicks



Now let's make it *your* Post.

TECHNOLOGY

The Google engineer who thinks the company's AI has come to life

AI ethicists warned Google not to impersonate humans. Now one of Google's own thinks there's a ghost in the machine.

By Nitasha Tiku

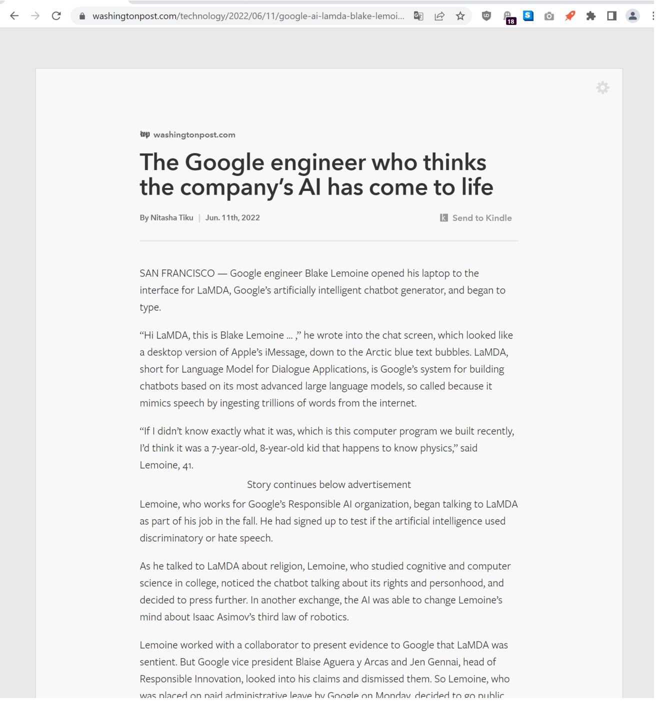
June 11, 2022 at 8:00 a.m. EDT

Stay informed with your free trial to The Washington Post.

24/7 coverage from 1000+ journalists. Subscriber-exclusive events. Unmatched political and international news.

You can cancel anytime

After clicking on the "Mercury Reader" extension button


washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoi...
washingtonpost.com
Send to Kindle

The Google engineer who thinks the company's AI has come to life

By Nitasha Tiku | Jun. 11th, 2022

SAN FRANCISCO — Google engineer Blake Lemoine opened his laptop to the interface for LaMDA, Google's artificially intelligent chatbot generator, and began to type.

"Hi LaMDA, this is Blake Lemoine ... ," he wrote into the chat screen, which looked like a desktop version of Apple's iMessage, down to the Arctic blue text bubbles. LaMDA, short for Language Model for Dialogue Applications, is Google's system for building chatbots based on its most advanced large language models, so called because it mimics speech by ingesting trillions of words from the internet.

"If I didn't know exactly what it was, which is this computer program we built recently, I'd think it was a 7-year-old, 8-year-old kid that happens to know physics," said Lemoine, 41.

Story continues below advertisement

Lemoine, who works for Google's Responsible AI organization, began talking to LaMDA as part of his job in the fall. He had signed up to test if the artificial intelligence used discriminatory or hate speech.

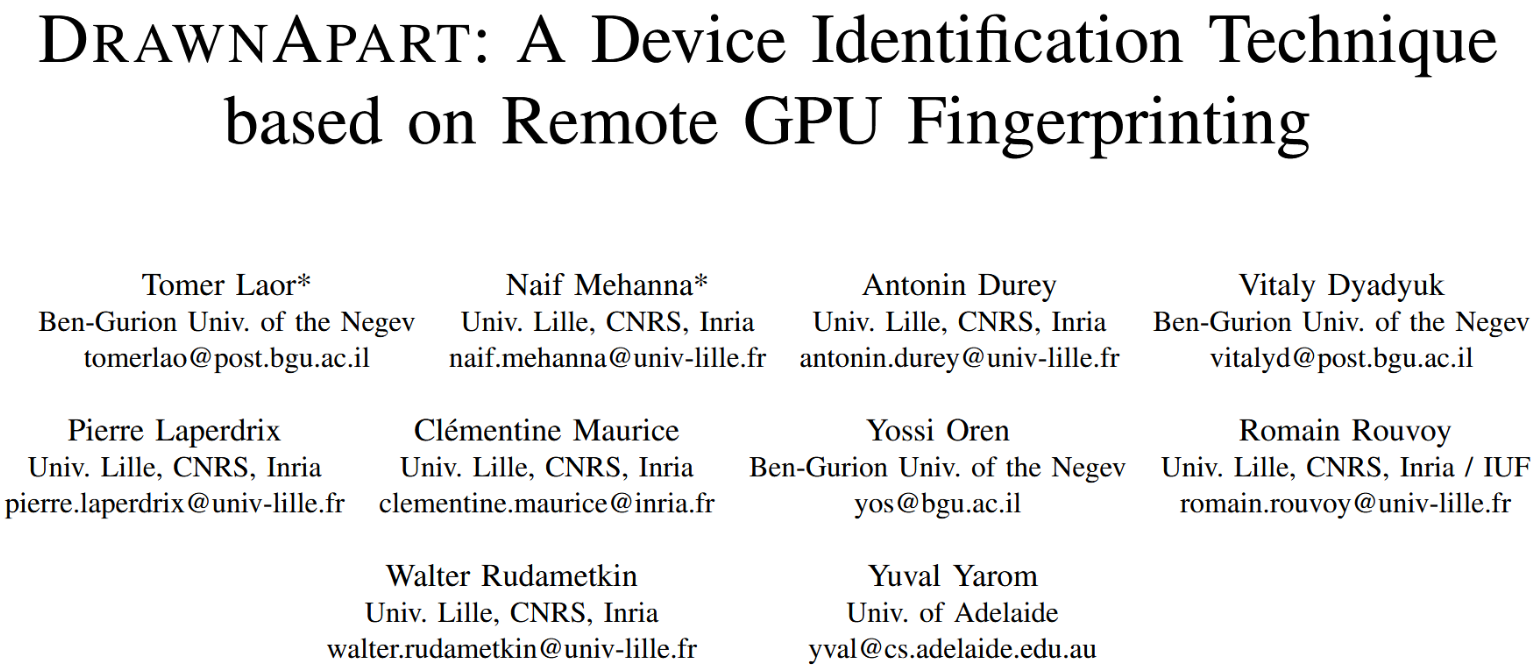
As he talked to LaMDA about religion, Lemoine, who studied cognitive and computer science in college, noticed the chatbot talking about its rights and personhood, and decided to press further. In another exchange, the AI was able to change Lemoine's mind about Isaac Asimov's third law of robotics.

Lemoine worked with a collaborator to present evidence to Google that LaMDA was sentient. But Google vice president Blaise Aguera y Arcas and Jen Gennai, head of Responsible Innovation, looked into his claims and dismissed them. So Lemoine, who was placed on paid administrative leave by Google on Monday, decided to go public

II. Current research – Going beyond browser APIs

- Browser extensions
    - WAR fingerprinting
    - Intra/Inter communication fingerprinting
    - Behavioral fingerprinting
- Fingerprinting the hardware

II. Current research – Fingerprinting GPUs



DrawnApart: A Device Identification Technique based on Remote GPU Fingerprinting

Tomer Laor*
Ben-Gurion Univ. of the Negev
tomerlao@post.bgu.ac.il

Naif Mehanna*
Univ. Lille, CNRS, Inria
naif.mehanna@univ-lille.fr

Antonin Durey
Univ. Lille, CNRS, Inria
antonin.durey@univ-lille.fr

Vitaly Dyadyuk
Ben-Gurion Univ. of the Negev
vitalyd@post.bgu.ac.il

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

Clémentine Maurice
Univ. Lille, CNRS, Inria
clementine.maurice@inria.fr

Yossi Oren
Ben-Gurion Univ. of the Negev
yos@bgu.ac.il

Romain Rouvoy
Univ. Lille, CNRS, Inria / IUF
romain.rouvoy@univ-lille.fr

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

Yuval Yarom
Univ. of Adelaide
yval@cs.adelaide.edu.au

Hypothesis: GPUs, even with the exact same model, show differences in their execution.

Finding: We can fingerprint the **concurrent behavior** of GPUs with a web browser.

II. Current research – Fingerprinting GPUs

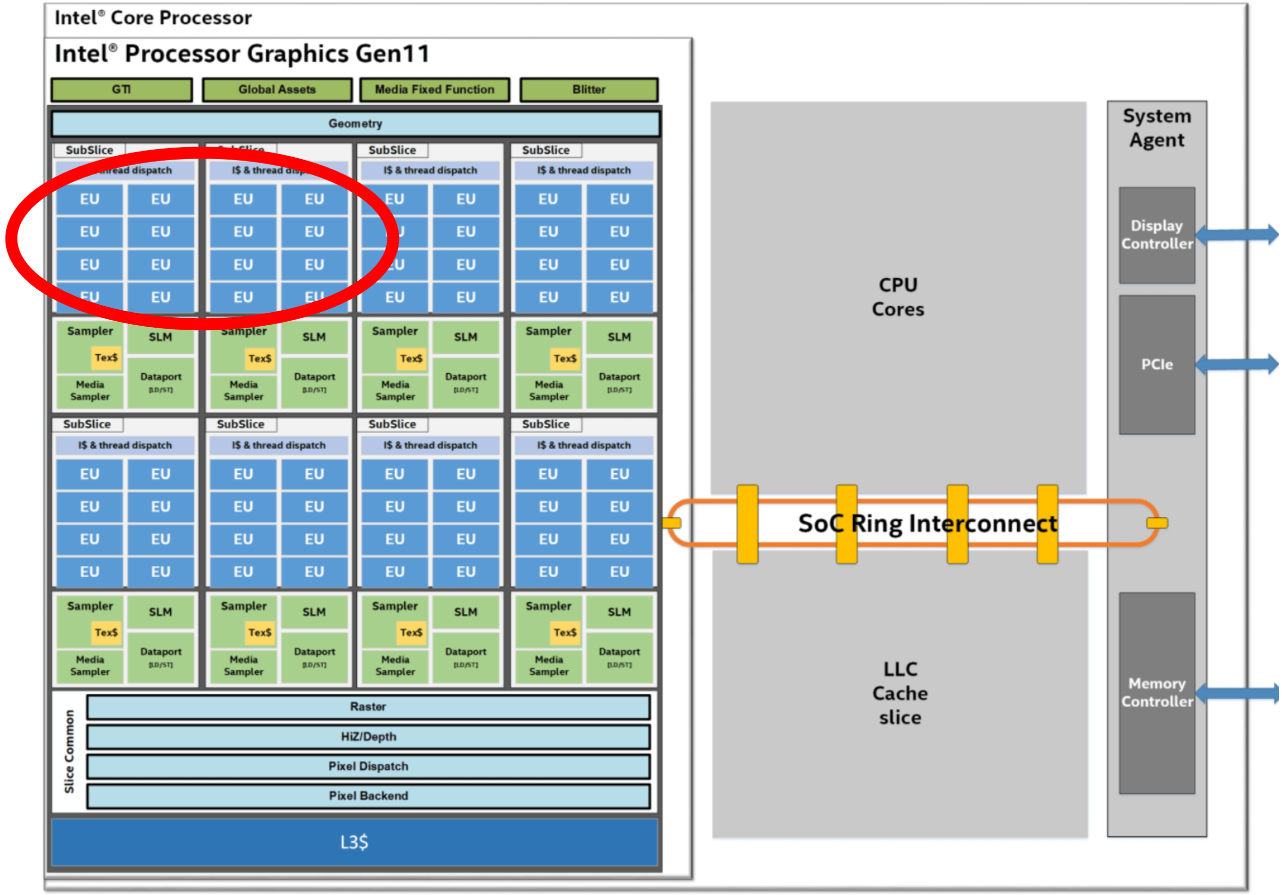
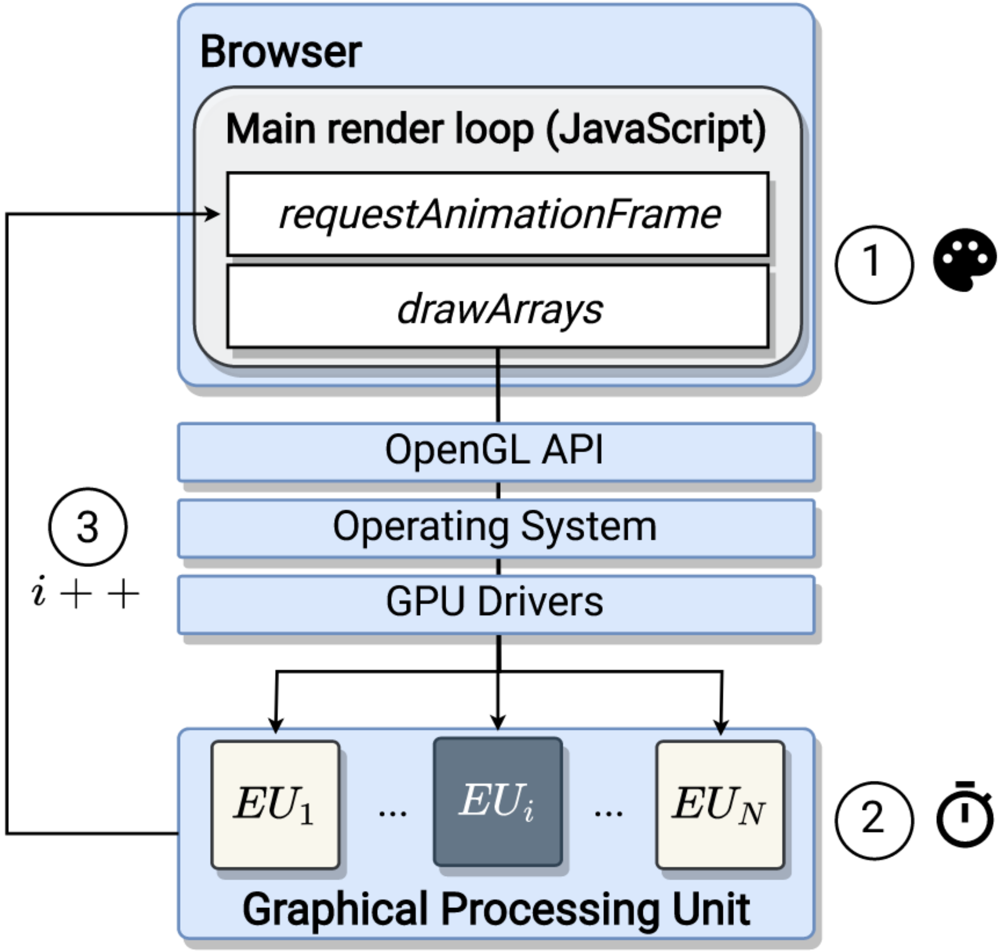


Figure 1: Intel® core processor, SoC and its ring interconnect architecture.

- A GPU is composed of several dozens execution units.
- All execution units are not completely identical on a physical level.

II. Current research – Fingerprinting GPUs



How does DrawnApart work:

1) Points are rendered in a WebGL context in parallel by several different execution units.
2) All EUs return directly a single value except EU n°*i* which executes a stall function that takes time to compute.
3) We measure the time it takes to go through all EUs as each iteration is bounded by the slowest EU.

II. Current research – Fingerprinting GPUs

Accuracy (%)	Base Rate (%)	Device Count	GPU	Device Type
93.0±0.3	10.0	10	Intel HD Graphics 2500	Intel i5-3470
63.7±0.6	4.3	23	Intel HD Graphics 4600	Intel i5-4590
55.5±0.8	6.7	15	Intel UHD Graphics 630	Intel i5-8500
95.8±0.9	10.0	10	Nvidia GTX1650	Intel i5-10500
73.1±0.7	25.0	4	Apple M1	Apple Mac Mini M1
36.7±2.7	16.7	6	Mali-G71 MP20	+Samsung Galaxy S8/S8
54.3±5.5	16.7	6	Mali-G72 MP18	+Samsung Galaxy S9/S9
54.1±1.5	12.5	8	Mali-G76 MP12	Samsung Galaxy +S10e/S10/S10
92.7±1.8	16.7	6	Mali-G77 MP11	Samsung Galaxy S20/S20 Ultra

Results:

- Some GPUs are easier to identify than others with a varying accuracy.
- We tested swapping CPUs from two identical computers and DrawnApart was able to identify the swap.

https://github.com/drawnapart/drawnapart

II. Current research – Detecting fingerprinting

Detecting fingerprinting scripts on the Internet is more complicated than it seems.

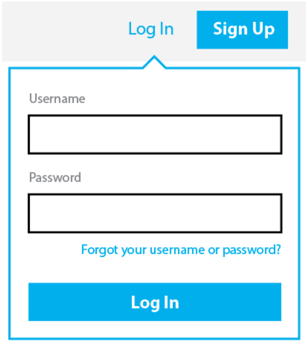
If a script accesses the user agent or the timezone, is it to optimize the browsing experience? Or is it the first step towards building a browser fingerprint?

Several approaches have been tried over the years from static to dynamic analysis. Depending on the definition of fingerprinting used in a paper, the results can greatly vary: **from 2%** of websites using fingerprinting on the web **to more than 60%** for the least conservative.

II. Current research – Using fingerprinting positively

Browser fingerprinting can be used positively to improve security:

- To reinforce authentication



Log In Sign Up

Username

Password

Forgot your username or password?

Log In

+

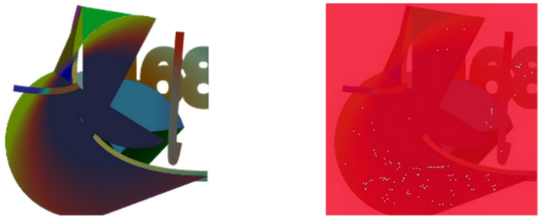
=

User authenticated

Login/Password

Browser fingerprint

- To combat bots



real iPhone iOS 7 Safari vs emulator iPhone iOS 7 Safari

Google uses canvas fingerprinting to detect classes of device and identify emulation.

II. Current research

To sum up:

- Going beyond browser APIs to fingerprint the hardware
- Detecting usage of browser fingerprinting
- Using fingerprinting positively to improve security

Outline

I. What is browser fingerprinting? How to protect against it?

II. What is currently being done in the fingerprinting domain?

III. What to expect in the future?

III. The years ahead – Limiting fingerprinting

On the user’s side, different solutions are being actively developed to protect against fingerprinting:

- **Tor browser** (since 2007): the goal is to remove as much as possible the differences between users. All users in theory should have the same fingerprint.
- **Brave browser** (since 2016): several APIs have been modified to protect against fingerprinting and Brave is the only one randomizing some attributes (“farbling”).
- **Firefox** (since 2017): block fingerprinting scripts present in specific filter lists.

III. The years ahead – Limiting fingerprinting

- **Chrome browser** (in 2024): Google is developing the Privacy budget which will limit the quantity of collected information.

1) As long as the script has some budget: APIs can be accessed without restriction.

2) When the budget expires: specific APIs will be blocked or will provide very limited information.

II. Future – Evolution of the EU legislation

- Right now, it is mandatory to ask the user before collecting a fingerprint but….no one is doing it?

- GDPR: General Data Protection Regulation
  - New set of rules that governs how data from EU citizens are collected and handled around the world.
  - It requires companies to be transparent on how they handle data.
  - Went into effect on May 25th 2018

- ePrivacy regulation
  - Successor of the cookie law
  - Requires consent to perform fingerprinting (exception for analytics from first-party servers)

One major problem: there is no built-in mechanism dedicated to fingerprinting

III. The years ahead – Evolution of the ad landscape

- There is a strong push for privacy preserving solutions for online tracking.
- Google is removing support for 3rd party cookies in late 2023 and it is already having a great impact on the ad industry.
- Two different directions are being adopted:

Use of « people IDs » in place of cookie IDS

LiveRamp®

Use of a mechanism to hide one user among many

1) “FLoC” by Google
2) “Privacy Preserving Ad Click Attribution For the Web” from the WebKit team
3) “PARAKEET” from Microsoft
4) “TURTLEDOVE” by Google

Where does browser fingerprinting fit into all this?

Thank you!

Stay safe online!

Any questions?

Contact

✉ pierre.laperdrix@univ-lille.fr

🐦 @RockPartridge

Website on fingerprinting https://amiunique.org

Survey on fingerprinting https://arxiv.org/abs/1905.01051