Browser fingerprinting

Nataliia Bielova
@nataliabielova

September 28th, 2018
Carnegie Mellon University
I. Internet in 2017

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

<table>
<thead>
<tr>
<th>1995</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browser: Netscape</td>
<td>Browser: Chrome v53</td>
</tr>
<tr>
<td>Language: Fr</td>
<td>OS: Linux</td>
</tr>
<tr>
<td>Browser: Chrome v53</td>
<td>Screen: 1920x1080</td>
</tr>
<tr>
<td>Language: Fr</td>
<td>Language: Fr</td>
</tr>
<tr>
<td>Browser: Chrome v53</td>
<td>Timezone: GMT+1</td>
</tr>
<tr>
<td>OS: Linux</td>
<td>Graphic card: GTX 1080Ti</td>
</tr>
<tr>
<td>Screen: 1920x1080</td>
<td>...</td>
</tr>
<tr>
<td>Language: Fr</td>
<td>...</td>
</tr>
<tr>
<td>Timezone: GMT+1</td>
<td>Graphic card: GTX 1080Ti</td>
</tr>
<tr>
<td>Graphic card: GTX 1080Ti</td>
<td>...</td>
</tr>
</tbody>
</table>

A bigger and richer web

- Audio
- Video
- 3D rendering
- Real-time communications
- Web payments
- Virtual reality
...
### Example of a browser fingerprint

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

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)
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.

- 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)
1. Example of values collected on AmlUnique

Cwm fjordbank glyphs vext quiz, 😊

Cwm fjordbank glyphs vext quiz, 😊
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
• User-agent on Android vs iOS devices

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

Fig. 5. Comparison of anonymity set sizes on the user-agent between Android and iOS devices
What if I disable JavaScript?


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

- How effective is fingerprinting at large scale?

2M fingerprints
33% are unique

Is tracking still possible?

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.

Why the results are so different? Bias in the previous datasets?

Table 1: OS market share distribution.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows</td>
<td>93.5%</td>
<td>63.7%</td>
<td>84%</td>
</tr>
<tr>
<td>MacOS</td>
<td>5.5%</td>
<td>14.9%</td>
<td>11%</td>
</tr>
<tr>
<td>Linux</td>
<td>0.9%</td>
<td>16.9%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Android</td>
<td>72%</td>
<td>55.6%</td>
<td>70%</td>
</tr>
<tr>
<td>iOS</td>
<td>18.8%</td>
<td>42.3%</td>
<td>22%</td>
</tr>
<tr>
<td>Windows Phone</td>
<td>7.6%</td>
<td>&lt;1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

But uniqueness is not enough for tracking!

Research challenges
1. Evaluate fingerprint stability
2. Evaluate the effectiveness of browser fingerprint tracking

Dataset
- 2 browser extensions during 2 years (Jul 2015 – Aug 2017)
- 98,598 fingerprints gathered from 1,905 distinct browsers (data cleaned)

### Fingerprint stability

<table>
<thead>
<tr>
<th>Attribute</th>
<th>50th</th>
<th>90th</th>
<th>95th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>Never</td>
<td>3.1</td>
<td>1.8</td>
</tr>
<tr>
<td>User agent</td>
<td>39.7</td>
<td>13.0</td>
<td>8.4</td>
</tr>
<tr>
<td>Canvas</td>
<td>290.0</td>
<td>35.3</td>
<td>17.2</td>
</tr>
<tr>
<td>Language</td>
<td>Never</td>
<td>215.1</td>
<td>56.7</td>
</tr>
<tr>
<td>Accept</td>
<td>Never</td>
<td>163.8</td>
<td>109.5</td>
</tr>
<tr>
<td>Cookies</td>
<td>Never</td>
<td>Never</td>
<td>Never</td>
</tr>
</tbody>
</table>

Goal: recognise the same browser by his fingerprint even if a fingerprint has changed over time

• How often should a fingerprint be collected so that it remains stable?

• What is the relation between frequency of collection and recognition of the browser?
Evaluation of effectiveness

Test set: 59,159 fingerprints from 1,395 browsers

<table>
<thead>
<tr>
<th>Browser A</th>
<th>fpA1</th>
<th>fpA2</th>
<th>fpA3</th>
<th>fpA4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browser B</td>
<td>fpB1</td>
<td>fpB2</td>
<td>fpB3</td>
<td></td>
</tr>
<tr>
<td>Browser C</td>
<td>fpC1</td>
<td>fpC2</td>
<td>fpC3</td>
<td>fpC4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>fpC5</td>
</tr>
</tbody>
</table>
Simulate the fingerprinting frequency (1 day, 2 days, ..., 20 days)

\[\text{fpA1} \quad \text{fpA2} \quad \text{fpB1} \quad \ldots \quad \text{fpA4} \quad \text{fpC5}\]

**Goal:** compare tracking effectiveness at different collect frequencies

Based on slide of Antoine Vastel
Apply linking algorithms

Link each fingerprint in the generated test set (chronologically)

Chain 1: fpA1, fpA2, fpB1, fpA3, fpA4
Chain 2: fpB2, fpB3
Chain 3: fpC1, fpC2, fpC3
Chain 4: fpC4, fpC5

Based on slide of Antoine Vastel
Evaluate ownership

Ratio of a chain owned by the majoritarian browser

Example: $\text{ownership}(\text{Chain 1}) = \frac{4}{4+1} = 0.8$

Based on slide of Antoine Vastel
Average ownership

Based on slide of Antoine Vastel
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!

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

Slides courtesy of Pierre Laperdrix (Stony Brook University)
How unique is your browser?

https://extensions.inrialpes.fr

~13,000 Chrome extensions

58 websites
Browser extension detection

• via **Web Accessible Resources**

![Chrome Extension](chrome-extension://gpdjojdkbbmdfjfahjcgigfpmkopogic/img/icon_48.png)

unique extension ID

---

**Discovering Browser Extensions via Web Accessible Resources**

Alexander Sjösten  
Chalmers University of Technology  
Gothenburg, Sweden  
sjosten@chalmers.se

Steven Van Acker  
Chalmers University of Technology  
Gothenburg, Sweden  
acker@chalmers.se

Andrei Sabelfeld  
Chalmers University of Technology  
Gothenburg, Sweden  
andrei@chalmers.se

**ABSTRACT**

Browser extensions provide a powerful platform to enrich browsing experience. At the same time, they raise important security questions. From the point of view of a website, some browser extensions are invasive, removing intended features and adding unintended ones, e.g., extensions that hijack Facebook likes. Conversely, from the point of view of extensions, some websites are invasive, e.g., websites that bypass ad-blockers. Motivated by security goals at clash, this

The first and second scenarios present an exclusive point of view of websites, concerned with malicious extensions. The third scenario presents an exclusive view of extensions, concerned with malicious websites. The fourth scenario illustrates legitimate synergies between websites and extensions. Finally, the fifth scenario illustrates the security goals of websites and extensions at outright clash.

**Bank scenario** Bank webpages manipulate sensitive information whose unauthorized access may lead to financial
How unique is your browser?
https://extensions.inrialpes.fr

Browser Extension and Login-Leak Experiment

When you browse the web, small beacons (trackers) are spying on your online activities. Even though such trackers are invisible, they collect information about you such as which pages you visit, which buttons clicked, and what text you typed. This information is often used to show you targeted advertisements and may require you to pay a higher price during online shopping depending on the collected information.

Did you know websites can track you by your browser extensions and web logins?

Recent studies show that you can be tracked based on your web browser properties. In this experiment, we demonstrate that you can also be tracked by

- your browser extensions (such as AdBlock, Pinterest, or Ghostery), and
- the websites you have logged in (such as Facebook, Gmail, or Twitter).

You can learn more here about how these detection techniques work.

In the experiment, we will collect your browser fingerprint, together with the browser extensions installed and a list of websites you have logged in. We only collect anonymous data during the experiment (see our Privacy Policy), we will securely store the data on an Inria server, use it only for research purpose and not share it with anyone outside of Inria.

You can also read the frequently asked questions here.

21 000 users have already tested!
How unique is your browser?

https://extensions.inrialpes.fr
User dataset w.r.t previous studies

Table 2: Previous studies on measuring uniqueness based on browser extensions and our estimation of uniqueness.

<table>
<thead>
<tr>
<th>Study</th>
<th>Fingerprints collected in a study</th>
<th>Extensions targeted in a study</th>
<th>Unique fingerprints in a study</th>
<th>Unique fingerprints in our dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing leaks [54]</td>
<td>204</td>
<td>2,000</td>
<td>56.86%</td>
<td>55.64%</td>
</tr>
<tr>
<td>XHOUND [58]</td>
<td>854</td>
<td>1,656</td>
<td>14.10%</td>
<td>49.60%</td>
</tr>
<tr>
<td>Ours</td>
<td>7,643</td>
<td>13k</td>
<td>39.29%</td>
<td>39.29%</td>
</tr>
</tbody>
</table>

Uniqueness grows as the dataset grows!

How to get a meaningful dataset?

How to define when we have enough users?

Figure 13: Uniqueness of Chrome users based on their extensions only vs. number of users - 204 is the number of users used in [54] and 854 the number of users considered in [58]

How many extensions our users have?

7,643 users of Google Chrome browser

Am I really unique if I use a few extensions?

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

Less protected (15 cookies accepted)
Harder to track (49.7%)

More protected (3 cookies accepted)
Easier to track (54.8%)

*4,000 pages crawled

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** and **stable** over time.

• Test your device on

  [https://amiunique.org](https://amiunique.org) and [https://extensions.inrialpes.fr](https://extensions.inrialpes.fr)