ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 11, Issue 1, pp: (71-78), Month: January - March 2023, Available at: www.researchpublish.com

Device Fingerprinting and Identification

Omar Khalid Alfaraj

Saudi Aramco Oil Company

Dhahran, Saudi Arabia

DOI: https://doi.org/10.5281/zenodo.7759680

Published Date: 22-March-2023

Abstract: **Browsers are our portal to the web, and if users want a secure, adequate and smart services, websites need to know some information about their browser, operating system and their hardware device. Modern day browsers have evolved a lot; browsers that provide the beauty of the web also provide a dark side, a rich environment of exploitable information that can identify a device. This is called "device fingerprinting," which is the technique of identifying visitors of a website and verifying if they are who they claim to be. The idea of fingerprinting is to gather some data that helps to recognize the user amongst others. The data collected could be categorized into three main types: Firstly, the device-related data such as graphics card type, and CPU information. Secondly, the information related to the software such as operating system and browser. Thirdly, the user behavioral data such as the mouse speed and keyboard pattern. Websites used to use cookies for identification, however, there is a new and better methodology nowadays. This paper will show what can be used for device fingerprinting as an alternative to cookies. I conclude that there is always a trade-off between privacy and a good service on the internet.**

Keywords: **Fingerprints, browsers, uniqueness, verification, personal data, cookies.**

1. INTRODUCTION

Device and browser fingerprints are terms used to refer to hardware details and browsers information that makes it unique among others. Therefore, device fingerprinting is the technique of identifying a user by collecting information related to their hardware device, operating system, browser, and behaviour. The technique is widely used to identify and authenticate the users of the application or website for security and advertisement purposes. That information can be collected through JavaScript, Flash, Java and PHP. When the values of those attributes are combined, they can give a unique fingerprint. Furthermore, advertisements companies used to rely on cookies for identification of users, however, cookies can be easily deleted at any time. Therefore, many companies resort to device fingerprinting data because it was shown to be a better solution, and provide better services for users on the internet, such as correctly rendering the page content.

2. RELATED WORKS

Several studies have been conducted to explain and examine the device and browser fingerprints, however, the most popular ones are the following:

1- "How Unique Is Your Web Browser?" by Peter Eckersley, which is a very relatable research to this paper's topic. In 2010 Eckersley developed the Panopticlick.eff.org website, which was aiming at gathering device specific data through scripts that run on browsers. He collected 10 attributes from visitors, which will detect users' browsers and their environments. Eckersley mentioned that around 83% of visitors had unique fingerprints, and this percentage increased to reach 94% in cases where Java or Flash Plugins were enabled. He also showed that the fonts and plugins lists that were gathered via Flash API and JavaScript API respectively played a major part in giving distinguishing between visitors and providing better results. [1].

2- Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints. A paper done by AMIUnique.org, which is a project that was launched six years after the Panopticlick study, tried to provide a more effective

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 11, Issue 1, pp: (71-78), Month: January - March 2023, Available at: www.researchpublish.com

way of fingerprinting by looking at the parts that have not been addressed in the previous study, such as exposure to mobile phones fingerprinting and comparing the ability of that with the ability of fingerprinting the computer. There were some similar results and they even benefited from the previous study. This study also compared what they have produced with what had already been done [2].

3- "FPDetective: Dusting the Web for Fingerprinters," which was one of the few studies that analyzed the web pages that use the device fingerprint methodology. According to the researchers Nikiforakis et al, they analyzed the fingerprint for three well-known commercial companies and found out that privacy of users was not taken into consideration as it should be, and the fingerprinting scripts were snooping on users to extract as much information about their devices. Moreover, they said about 5% of the first 100,000 sites use canvas fingerprinting scripts on their home pages [3]. This study also focused on several aspects that previous studies did not focus on, such as canvas and fonts fingerprints.

What is unique in this study is the number of attributes mentioned compared to the ordinary ones like: Flash and JavaScript. User behaviours properties such as mouse movement, scrolling, and keyboard speed are very strong features to identify visitors. I focused on the most important attributes that usually narrow down the circle of users' possible results. Attributes should be universal, unique, stable, collectable, resistant to forgery and tampering, and have a high entropy value.

3. THE IMPORTANCE OF DEVICE FINGERPRINTING

Identification is getting more and more important, especially with the high number of cybercriminals targeting online bank and social media accounts nowadays, which is a serious problem that might leads to impersonation, money loss or worse. If users got their details stolen then the privacy of that victim is breached. So, a mechanism that can verify the devices to recognize users is strongly needed. If a user suddenly changed the device that they usually use on this website or application then more verification is needed to distinguish between the real account holder and hackers.

When a user is created in a system that uses device fingerprinting, the information gathered from the user device will be stored in the system database. In future logins, the system will gather the information again and compare it with the already stored user's details in the database. This mechanism eliminates most fraud tactics. If the real account holder happens to change the device usually used to log in then further confirmation will be taken to allow the user login such as Two-Factor Authentication. If the user conformed that he/she is the owner of that account, then the new data of the device should be added for that user by updating the database. This means a new device fingerprint will be mapped to the user. Figure 1 below shows the overall idea in a simple way.

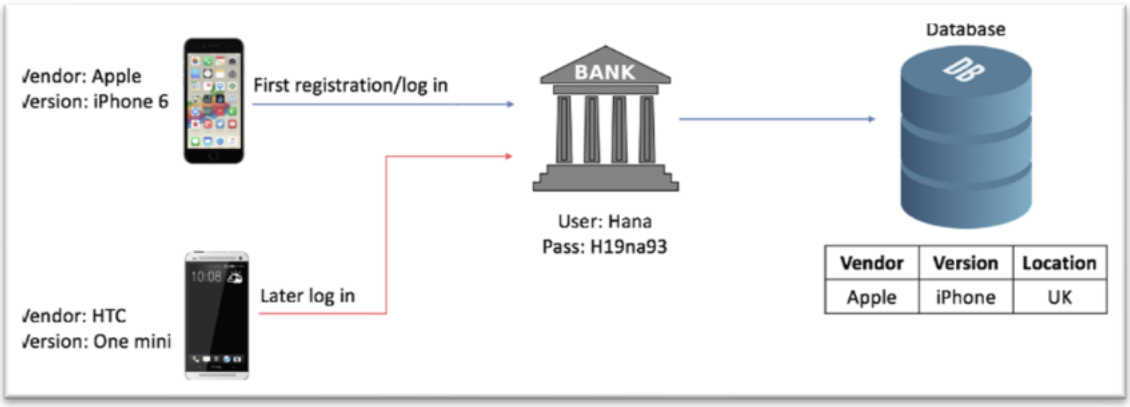

FIGURE 1: Storing user data

Nevertheless, in some cases such as a stolen device and stolen credentials it will be difficult to detect a fraud if systems only depend on device hardware related information. One way to overcome such scenarios is to consider user behavioural data and combine it with all other data to create a special and unique fingerprint. Behavioural data can be:

1) Mouse movements: measuring the time a user takes to move the cursor from a place to another.

2) Scrolling speed: measuring the speed of a cursor movements on the page.

3) Keyboard pattern: measure the typing speed and style of the user.

Those data can be gathered by asking the user to select the right photo between multiple options, measuring the speed of the mouse since accessing the site, and checking how long it takes the user to enter the credentials. Since different people will have different speed and styles, we can exploit those for more accurate identification. See Figure 2 for an example of how to test mouse speed. This technique is called "Cross Fingerprinting" because it can detect the user across different browsers on the same device.

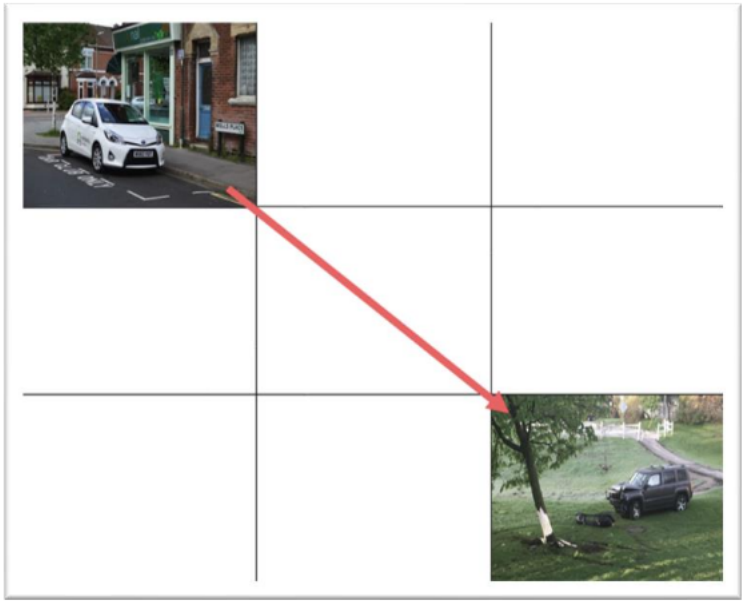

FIGURE 2: Testing mouse movement speed

The fingerprinting mechanism raised some serious concerns because users cannot control what will be fetched from their devices once they visit a web page. However, more people started to learn about these practices that violate their privacy on the internet. As result of that, VPNs, private browsers, and extensions such as AdBlock and Gostery started to become more common. These strengthened the desire of users to be untraceable and unidentified. A lot of users still prefer comfort to privacy as pointed out by the Laperdrix, Rudametkin and Baudry [2]. And this seems to be a bit logical as many web users do not pay attention to these privacy issues and they are not qualified to understand it, so they sacrifice privacy in exchange for better service.

A challenging part is to identify what information is feasible to be collected and how to formulate methods to extract them. The next section of this paper will explain the old and new methods of extracting data from users on the web with more details.

4. THE OLD METHODS FOR IDENTIFICATION

1- Cookies

Identifying users started with the concept of cookies which is "The most common and well-known method for persisting data on the client. As a part of a HTTP response, any web server can issue unique identifiers to first-time visitors and have the browser use the stored values on all future requests to a specific site" [4]. Cookies are used to send the state information related to the users' browser, which can be utilized for authentication. In addition to that, it is used to save users' preferences, keep accounts logged in, and provide appropriate advertisement for users based on the websites they recently visited or the items they have bought. In general, cookies do many things to provide a better surfacing experience of internet. However, cookies pose a serious threat to user privacy, and also might take some space on the user's device. Therefore, some people delete cookies or use the private browsing to avoid the cookies side effect.

There are two types of cookies used on the internet:

1) First-party cookies

2) Third-party cookies

The second cookies type is the most common nowadays. It is set and read by another party that provides a service such as advertising companies [5]. Since information on a website with third-party cookies can be read by external providers, that means other websites might be able to track visitors of that website as well. Therefore, many warnings from experts have been said about the privacy, and they debate whether they should or should not allow cookies while using the internet for protecting the privacy [6].

2- Super Cookies and Session Cookies

The problem with the regular cookies, whether they are first party or third party, is that they are easy to find and delete via the browser settings, which cause a loss to the advertising companies and other organizations who depend on collecting cookies for business and income. This is the reason why a new type of cookies have been introduced that are stronger and more reliable. They work similarly to the regular cookies but they are harder to detect and delete, and are therefore called "Super Cookies." In contrast to the regular cookies, most super cookies are stored in different locations of the computer, which makes deleting more difficult. Many websites have been using Super Cookies instead of regular cookies. Super Cookies can regenerate the users' profile without their knowledge after users delete the normal cookies from the browser [1].

The previous cookies were persistent cookies, however, there is another type of cookie called "Session Cookies," also known as "Transient Cookies." It gets deleted when the user closes the browser automatically. The session cookies are stored in a temporary memory and it is not kept until it gets deleted manually from the memory. Basically, this cookie does not collect data about the user browser or user device, however, they store information in the form of session identification [7].

Figure 3 below summarizes the cookies procedure, which consists of three steps:

1- When the user accesses the web server, a HTTP request will be sent

2- Upon receiving the request, the server will create the cookies and send to the client

3- If the user accesses the page or subpage again, the server will send the same cookies it has stored for this user.

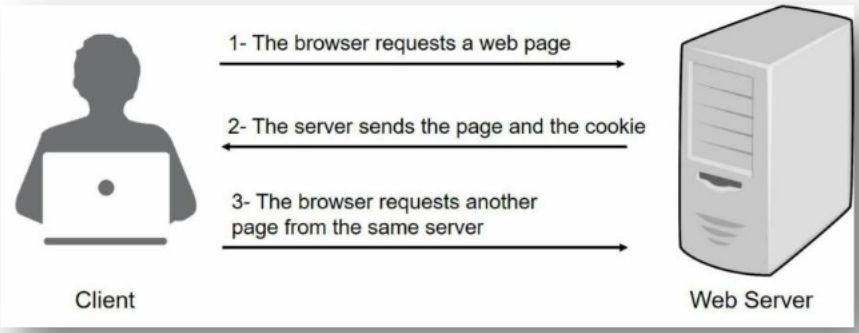

FIGURE 3: Procedure of request between client and server

5. THE NEW METHODS FOR DATA EXTRACTION AND IDENTIFICATION

The knowledge of cookies danger that affects the privacy has become apparent to many. So, some internet users resort to avoid cookies by browsing anonymously, which is known as "private mode" browsing. Thus, fingerprinting came to the surface to make up cookies' limitations. Although the fingerprinting mechanism can use cookies, it does not depend on the existence of cookies to identify a user. When users enter a webpage, the page will be designed with embedded JavaScript codes that help with loading pages, clicking buttons, or filling empty fields. The website will start fetching data that is needed for identification of users. It will collect a list of attributes – that when combined together – can uniquely identify and distinguish between visitors of a website and track them on the web. By tracking, I mean the linking of visits to web pages as made by the same device. These data will be stored in variables, which will later be sent to PHP code file to be pushed to the database server.

1- JavaScript's Objects

JavaScript was developed in 1995 by Bernard Eich. JavaScript is a scripting programming language that is used with HTML and CSS to make a webpage more interactive and to do complex thing such as dynamic updating, animate 2D, 3D which HTML and CSS alone cannot do. JavaScript is a very powerful tool that can be used to fingerprint visitor of a website when exploiting some of its objects such as the navigator and screen objects. Below tables shows what those objects can extract from the user device and browser:

TABLE 1: Navigator object's properties

Property/function	It's job
cookieEnabled	Determines if cookies are enabled in the browser.
Language	Returns the language of the browser.
Platform	Returns for which platform the browser is compiled.
Product	Returns the engine name of the browser.
User Agent	Returns the user-agent header sent by the browser to the server.
Java Enabled	Specifies if the browser has Java enabled.

TABLE Error! No text of specified style in document.**: Screen object's properties**

Property/function	It's job
Height	Returns the total height of the screen.
Width	Returns the total width of the screen.
Color Depth	Returns the bit depth of the color palette for displaying images.

Other properties that can be gathered using JavaScripts are: the user agent, language of the device, IP address, and list of fonts. Those properties help in the process of identification of a browser when they are put together. A study that has been done by Nicholas Zakas in 2010 [8] showed that only 1% of people who use modern browsers have their JavaScript disabled. That means relying on JavaScript is perfectly acceptable since it is enabled in almost all users' browsers. Users nowadays are using the web all the time, and wherever they go, they use it via mobile phones, tablets, and desktops. In 2009, Mayer and Mitchell conducted a fingerprinting experiment on 1,328 clients by hashing the concatenated content that he collects from navigator and screen objects, and what he discovered was that around 96% of them was uniquely identified [9].

2- Java Language Attributes

Java is one of the most popular programming languages that supports object-oriented programming. Java can make the webpage more intractable with users. Due to this reason, programmers tend to code the online games in java language. People might get confused between Java and JavaScript due to the similarity of their names. However, Java is typically compiled into bytecode and run over a java virtual machine while JavaScript is interpreted on the web browser [10].

The need for Java in such a mechanism comes from the limitation of JavaScript where it can only fetch data about the user browser and software that the browser is running on. Meanwhile, Java can be used to gather more hardware information such as the total memory space and details of CPU core. Java provides a wider chance of gathering information than JavaScript. The popular way to use Java on HTML pages is via extending the Applet class, and adding applet tag in the HTML page. This way, when users open a webpage, the java file will be executed. Unfortunately, this way might be useful in some contexts but not in this project as java applet is used for simulation. What is needed is to run java to fetch the user data to be stored in the database.

Another approach for using Java on HTML pages can be done via JavaScript by importing the "JavaPoly.js" library. It is a collection that extends java virtual machine to give java code an appropriate environment to be compiled and run. The beautiful advantage of this collection is that it will work whether users have java environment on their computer or not, and the code will be executed straight away. This is in contrast to java applet, which needs a Java SE Runtime environment to be executed and must be executed in a window [11].

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 11, Issue 1, pp: (71-78), Month: January - March 2023, Available at: www.researchpublish.com

3- WebGL & Canvas details

WebGL is short for "Web Graphics Library," which is an API in JavaScript used in drawing and rendering 2D and 3D graphics based on the OpenGL API. This attribute is now supported by most browsers, and it provides data about the underlying GPU of the user device. Moreover, there is another attribute that can help to enhance the uniqueness of the identification, which is the "Canvas". It has been discovered recently. In the canvas fingerprinting technique, when users visit a website that has canvas fingerprinting code, their browser is instructed to "draw" a hidden line of text or/with 3D graphic and the result will depend on the hardware and software depending on the operating system and the graphical card. As you can see in Figure 4, the open mouth emoji when executed will be different in windows than in mac [2]. Thus, this gives us the power to do test on the operating system, and on the hardware as well.

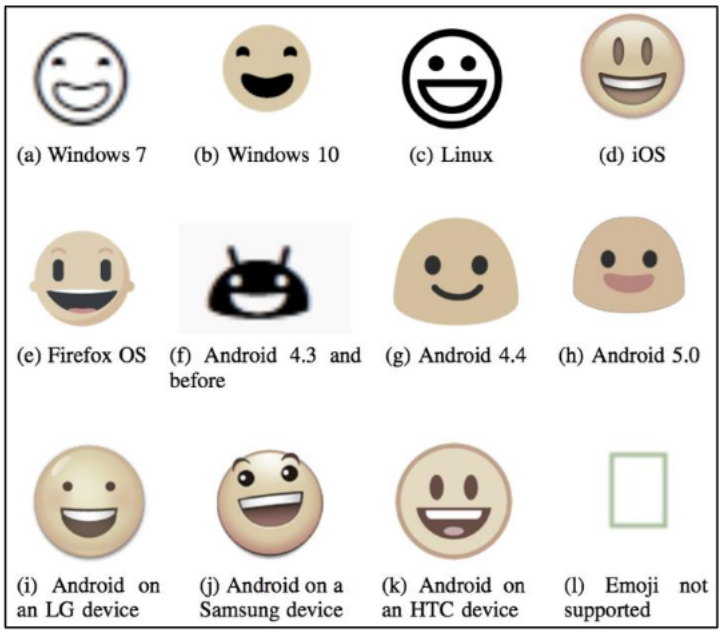

FIGURE 4: Comparison of the "smiling face with open mouth" emoji on different devices and operating systems.[2]

4- HTTP Headers

HTTP headers stands for "HyperText Transfer Protocol." This is used in necessity in fingerprinting because it can increase fingerprint uniqueness by 36%, as stated by Laperdrix et al. When connecting to a server, the browser automatically sends the list of HTTP headers used, such as the user agent, language of the web, and other headers. Browser extensions might change or add more headers, which gives extra information about the user device configuration. These details can be collected using JavaScript through objects that will be discussed later. Since modifying the HTTP header is doable, fetching those data from more than one property is a good thing to do to check for inconsistency according to Laperdrix [2].

5- Flash Plugins

Adobe Shockwave Flash is considered to be the most used plugin in the world for playing animation. There are around half a million users who installed and enabled this plugin in their browsers according to the Adobe website. This plugin is used to provide a better service while surfing the internet mainly to allow users to view interactive web content like video games, business presentation, and advertisement. Thus, it is important that the flash is installed on the browser to provide the best possible service to the users.

Flash Player is another plugin, which has a similar function to the previous plugin. Thus, Flash is also used by many users to display active contents. The difference between the two plugins is that Shockwave shows content that has been created with Adobe Director software whereas Flash Player displays content that has been created with Flash Professional CS5. Both are free web players from Adobe. These Flash plugins are very helpful for fingerprinting users' browsers since it is installed on many of them. These plugins have a different version, which can also be used as an identifying feature.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 11, Issue 1, pp: (71-78), Month: January - March 2023, Available at: www.researchpublish.com

Figure 5 below shows an example of what can be fetched from a user's device when accessing a page with a fingerprinting mechanism.

User agent	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_12_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/65.0.3325.181 Safari/537.36
Platform	MacIntel
Language	en-US
Is Java Enabled?	Java is not enabled.
Is Flash Enable?	Flash id disabled
Height	800
Width	1280
Color Depth	24
Cookies	Cookies are enabled.
Name	Welcome again chrome2
IP address	2.223.55.87
Time and Location	BST
Canvas	CheckUniqueness <canvas> 1.0
Canvas Value	ata:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAASwAAACWCAYAAABkW7XSAAAcn0lEQV
DoNotTrack	DoNotTrack is disabled
WebGLVendor	Intel Inc.~Intel HD Graphics 4000 OpenGL Engine
AdBlock	AdBlock is disabled
Number of CPU core	4
Fonts	"Arial","Arial Black","Arial Narrow","Arial Rounded MT Bold","Arial Unicode MS","Baskerville Old Face","Batang","Bauhaus 93","Bell MT","Bernard MT Condensed","Book Antiqua","Bookman Old Style","Britannic Bold","Brush Script MT","Calibri","Calisto MT","Cambria","Cambria Math","Candara","Century Gothic","Century Schoolbook","Colonna MT","Comic Sans MS","Consolas","Constantia","Cooper Black","Copperplate Gothic Bold","Copperplate Gothic Light","Corbel","Courier","Courier New","Curlz MT","Edwardian Script ITC","Engravers MT","Eurostile","Footlight MT Light","Franklin Gothic Book","Franklin Gothic Medium","Gabriola","Garamond","Geneva","Georgia","Gill Sans MT","Gloucester MT Extra Condensed","Goudy Old Style","Gulim","Haettenschweiler","Harrington","Helvetica","Helvetica Neue","Impact","Imprint MT Shadow","Lucida Bright","Lucida Calligraphy","Lucida Console","Lucida Fax","Lucida Grande","Lucida Handwriting","Lucida Sans","Lucida Sans Typewriter","Lucida Sans Unicode","MS Gothic","MS Mincho","MS PGothic","MS PMincho","MS Reference Sans Serif","Matura MT Script Capitals","Microsoft Himalaya","Microsoft Sans Serif","Microsoft Yi Baiti","MingLiU","MingLiU-ExtB","MingLiU_HKSCS","MingLiU_HKSCS-ExtB","Mistral","Monaco","Mongolian Baiti","Monotype Corsiva","Onyx","PMingLiU","PMingLiU-ExtB","Palatino Linotype","Papyrus","Perpetua","Perpetua Titling MT","Playbill","Rockwell","Rockwell Extra Bold","SimHei","SimSun","SimSun-ExtB","Stencil","Tahoma","Times","Times New Roman","Trebuchet MS","Tw Cen MT","Verdana","Wide Latin"
Mouse	1756
Scroll	44

FIGURE 5: Device fingerprint example

6. CONCLUSION

Fingerprinting is the technique of identifying a user by collecting information about their device, software, and behaviour. Fingerprinting is used due to the increase of visitors, which raises the need to identify them to make sure that users are who they claim they are, or track them for other purposes such as advertisement. Websites tend to use different ways to identify visitors. Some website used to depend on cookies, however, cookies can be deleted, which is problematic for the advertisement companies or websites. Others found a new way, which is fingerprinting. This paper showed more attributes that can be collected compared to the other papers and studies. User behaviours properties such as mouse movement, scrolling, and keyboard speed are very strong features to identify visitors. I focused on the most important attributes that usually narrow down the circle of users' possible results. Attributes should be universal, unique, stable, collectable, resistant to forgery and tampering, and have a high entropy value. I conclude that there is always a trade-off between privacy and a good service on the internet.

REFERENCES

[1] Eckersley, P. (2010). How Unique Is Your Web Browser? Retrieved Nov 13, 2022, from https://panopticlick.eff.org/static/browser-uniqueness.pdf

[2] Laperdrix, P., Rudametkin, W., Baudry, B. (2016) Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints. Retrieved March Dec, 2022, from https://hal.inria.fr/hal-01285470v2/document

[3] Acar, G., Juarez, M., Nikiforakis, N., Diaz, C., Gürses, S., Piessens, F., and Preneel, B. (2013). FPDetective: Dusting the Web for Fingerprinters. Retrieved March 08, 2023, from https://www.esat.kuleuven.be/cosic/publications/article-2334.pdf

[4] Janc, A., & Zalewski, M. (2014). Technical analysis of client identification mechanisms. Retrieved March 08, 2023, from https://www.chromium.org/Home/chromium-security/client-identification-mechanisms

[5] Krishnamurthy, B., Wills, C. (2006). Generating a privacy footprint on the Internet. Retrieved March 08, 2023, http://web.cs.wpi.edu/~cew/papers/imc06.pdf

[6] Duong, V. (2011). A Proposal of a Cross-Browser User Tracking Method with Browser Fingerprint. Retrieved March 08, 2023, from https://www.sfc.wide.ad.jp/thesis/2011/files/nobita-publish-thesis.pdf

[7] Dacosta, I., Chakradeo, S., Admad, M., and Traynor, P. (2011). One-Time Cookies: Preventing Session Hijacking Attacks with Disposable Credentials. Retrieved Feb 28, 2023, from https://smartech.gatech.edu/bitstream/handle/1853/37000/GT-CS-11-04.pdf?sequence=1&isAllowed=y

[8] Nicholas, Z. (2010). How many users have JavaScript disabled?, Retrieved March 08, 2023, from https://web.archive.org/web/20110218221930/http://developer.yahoo.com/blogs/ydn/posts/2010/10/how-many-users-have-javascript-disabled/

[9] Mayer, J., and Mitchell, J. (2012). Third-party web tracking: Policy and technology. In Proceedings of the IEEE Symposium on Security and Privacy, p413–427.

[10] Dincer, K., Fox G. (1997) Using Java and JavaScript in the Virtual Programming Laboratory: A Web-Based Parallel Programming Environment. Retrieved March 09, 2023, from https://pdfs.semanticscholar.org/60dd/af4af7fead3274a446a0b272276a1fe7de75.pdf

[11] JavaPoly. (n.d.). Retrieved March 10, 2023, from https://www.javapoly.com