An Analytical Study of Web Tracking: In a Nutshell

AKM Bahalul Haque, Farhat Tasnim Progga, Md. Amdadul Bari

Abstract— Privacy over the internet has become a significant concern nowadays as users are ubiquitously tracked on the web. While using the internet, each of our footprints can be seen using the proper tracking mechanism. The users can be tracked on the web based on the interpretation and measurement of user's data while browsing. These user data include session identifiers, cookies, HTML local storage. The web tracking can be used in both right and wrong ways. For medical issues, security reason and financial reasons web tracking can be beneficial one while on the other hand tracking someone intentionally or accidentally reeks the privacy of the user which can be used for advertising or identity theft. In this paper, various web tracking techniques are discussed and analyzed to show how the users are vulnerable on the web while accessing it.

Index Terms— Web tracking, Privacy, Cookies, Fingerprinting, Security, Session, Anonymity

1 INTRODUCTION

WEB tracking is one of the most underrated words in this world as only a handful of people related to information technology especially the security and privacy sectors know how it works. The first thing that comes to our mind is what web tracking is actually and what it depicts. Well, in a simple sense we can say that web tracking is a method of collecting user information and statistics. In general, this information is collected for improving the user experience. That is while a system is being operated, the usage information and user information are sometimes collected to observe which types of response the system provides to its users or if there are any issues. Moreover, different service providers, content providers, and third-party companies collect user data too. This data is provided to other organizations and to their organizations for improving their business.

The rest of the paper is organized as follows: section II presents related work, section III presents the advantages of web tracking, IV presents the disadvantages of web tracking, section V presents the mechanism of web tracking, section VI discusses about how one can prevent themselves from being web tracked and finally section VII draws the conclusion of the presented study.

2 RELATED WORK

The current techniques of web tracking have been built through many evaluations. There are several tracking mechanisms [1] which are involved like; cookies, fingerprinting, HTML Local storage, the session identifier, etc. There are several types of fingerprinting techniques involved, e.g., operating system instance fingerprinting, canvas fingerprinting, network and location fingerprinting, device fingerprinting, browser fingerprinting, etc. in the browser fingerprinting HTML and CSS fingerprinting techniques [2]. CSS fingerprinting works while comparing with different elements of CSS in the browser. As we know that different web browser implements CSS differently, so from those attributes of different CSS families it is possible to identify the browser version and instances. On the other hand, HTML5 fingerprinting involves how the browser implements HTML standards. During the first stage online tracking, there was only way to track users, and that is through IP address and session ids sent through getting and POST method [3]. Cookies are another reason for web tracking. The authors in [4], expressed about this issue. There are many cookie leaks [5]. For example, some session cookies are stored only for that session means only for a specific browsing session. Several other types of cookies used for web tracking such as super cookies, ever cookies, etc.

3 BENEFIT OF WEB TRACKING/WHY IT IS DONE

3.1 Medical Issues

Different medical research institutes do this tracking for better research activities like tracking the patients while they are under any treatment. Their health conditions, behavioral analysis helps the authority for better treatment towards the patient. Moreover, it helps the scientists to further research on the specific disease. Nowadays a lot of smartphones have tracker options. Also, there are devices which solely used for tracking the fitness of the user. These data are gathered and calculated. Later the user is given specific suggestions for improving their daily routine. So, in this case, it can be said that to serve the human health issues better, different companies integrate these features with the regular usage devices. If there is vast research on patients about various diseases, there is a greater chance of making prevention as the first step than to cure. As the statistical data can be analyzed to recognize what are the pre-symptoms of being sick in case of specific diseases. For example, there is a program by WHO (World Health organization) about public health surveillance which aims at providing an early warning system for damage controlee like public health emergencies.

- *AKM Bahalul Haque is currently working as a Lecturer at North South University, Bangladesh, PH-+8801643481347. E-mail:bahalul.haque@northsouth.edu*
- *Farhat Tasnim Progga is currently pursuing bachelor program in computer science and engineering in North South University, Bangladesh, PH-+8801630471990. E-mail: farhat.progga@northsouth.edu*
- *Md. Amdadul Bari is currently pursuing bachelor program in computer science and engineering in North South University,Bangladesh,PH-+8801843105424.E-mail:amdadul.bari@northsouth.edu*

2.2 Financial Institutions

Financial institutions like banks and other state organizations track the online and offline (seldom) financial activities of the people for detecting fraudulent activities, e.g., money laundering and against any emerging crisis. This kind of tracking also helps the government to facilitate the support towards the citizens who might be falling into crisis.tables and figures will be processed as images..

2.3 Security Reasons

Various government organizations like police and intelligence agencies track online activities of suspected people for better knowing about any future threats against the country. Using this, different terrorist activities can be stopped around the world.

2.4 Others

There are other reasons or motivations for tracking user activities online. For example, nowadays autonomous cars have become available. The companies monitor each user activities while driving through their website and a built-in tracker in the car. This helps to locate any vehicle in any emergencies. If someone needs to find any lost vehicle, they can easily find it online through the website.

3 DrawBacks of webtracking

The first and foremost disadvantage is the privacy violation. In each case, if a user is tracked online intentionally or accidentally that person's privacy is broken. Personal data is stored online which is used for various activities later. E-commerce websites use the data to advertise online, use the address to send newsletter even if the user does not want it. They use this information for the publicity stunt and rating improvement. The information provided on one website are speeded online to increase the site visiting more frequent and

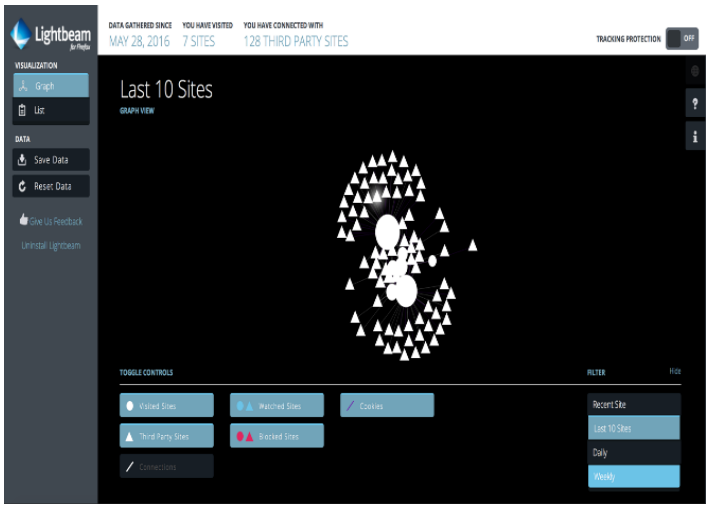

Fig. 1. Websites linking to third-party sites using cookie transfer (leak)

send in invitations for their purpose.

It is possible to show a map of statistics about the visited site, the third-party sites that are connected using a Firefox plugin. A demo is shown in the above figure and this it can be easily noticed that how many sites have been visited (only 7) and how many third-party sites have been added as a cookie in my browser. That means, these sites also have some of the information given in those seven websites. If only seven locations can do this, it can be readily understood that how all the sites that people visit every week or every month can do. If someone knows that he can be tracked online in any way that person will get anxious to share his idea or thoughts. Moreover, even if he might not say anything intentionally, his activity can be under suspicion in any circumstances though that person will not know about it.

4 Mechanisms of web tracking

There are several tracking mechanisms [1] from time to time which is involved in web tracking. Though it is not one of the most popular topics for people, it is still a matter worth discussing.

4.1 Session Identifiers

The user information is transferred to the website through the URL. Though it is an old method of collecting user data from the web, it was only available for once just, that is for the one session only. Web form authentication is another method of tracking internet user. When a user is using the internet sometimes, it is needed to register on the web that is to provide the user information to that website. After that, the user does not need to provide the information again as they are stored in cookies, and the website knows that he or she is already registered.

4.2 Cookies

It is one of the most popular and effective ways of tracking a user's online activity. A cookie is a small file, approximately limited to 4KB, stored on the user's computer. For each cookie, there are some unique identifiers. The identifiers are generated at random. However, creating the cookies and storing them does not require the user identification, so the user sometimes does not have any idea. On the other hand, the user must accept to store cookies occasionally. Later if the user does not clean the cookies, it is stored in the user's storage device. All the cookies are not of the same type, and we also have seen that the accuracy of this cookies depends on the user's acceptance and response. Moreover, there are tracking and non-tracking cookies. As explained before, the user sometimes must register to a website while using its service and facilities. This user gives the user-name and registration information only for once. Later user does not have to provide that information. That is because the information is stored in a browser in the user's computer for the time. After that, when the user tries to visit the same website again, the browser interacts with cookies and gets his credentials for using the service until the next time. In this case, the website is constantly getting the user information like the time zone he is in, the location browser history, etc. These cookies are usually transferred from one place to another. For example, Microsoft, Google track their users through their other services. The services exchange their cookies for better performance. The user can use

their service very efficiently like logging into one service and use all of them. Another example of cookie usage can be found at Wizz Air website [6]. This website is using cookies for improving the performance of the website. The information which is supplied by cookies is anonymous as well as it helped them to understand how their visitor's castoff Wizz Air websites for better presenting of their content to the users. The cookies that are being used in Wizz Air are strictly necessary cookies, Functionality cookies, Performance cookies, online behavioral advertising cookies.

4.3 HTML Local Storage

As HTML5 global storage was not implemented due to some policy issues. For this reason, the local storage option came into play. It facilitated some following features:

• No plug-in required.

• Permanent storage of object till the process is obstructed by the user or by the websites. That is if the object is not removed by either the website itself or by the user, it still stays as a permanent object.

• The storage can be up to 5 MB. The local storage for this is cleared while the cookies are cleared, so it also has a flexible removal option.

• Another type of storage is the HTML5 session. A user can also be tracked using this option, but only for a limited amount of time as like the HTML5 local storage it also uses the local storage option but only for a browser session. Once the session is closed the object is also deleted with it.

4.4 Fingerprinting

It is a more advanced way of tracking the user on the web. It takes more parameter than the other tracing mechanism making it robust. It has some following features. It does not require any cookies and even doesn't matter if the browser is cookie enabled or not. Works without any trace leaving to the user. No way of knowing by the average users if they are being watched over the internet. There are different types of fingerprinting each having different properties. Some of them are given below with a brief description.

Network and location fingerprinting

This type of fingerprinting is based on the HTTP. The incoming requests are the IP address which also reveals the geographical areas, where they are coming from. By using proper tools like network tools, the ISP (internet service provider) address can be known. It is also possible to tell if the user is using any proxy address. If it is, then the flash cookies come into the advantage for bypassing this. The two-renowned technical fingerprinting companies, Innovation & Threat Matrix does the job of getting the real IP address. The users not using the proxy, or any anonymity tools can be easily tracked by some free tools available on the internet. It is possible to get the exact location of them from the IP address they provide to us. In the case of mobile internet users, it is hard to track the users based on the IP address as it shows different locations in the tracker.

Device fingerprinting

Another method of tracking by fingerprinting. This method does not depend on the browser specifics. It is based on JavaScript and has so far proven very useful in some sectors. Its functionalities can be described as follows: The IP address is collected. From the first two octets, operating system specifications, browser specifications, and font information are collected. The font specific information of the targeted user's browser is collected with the help of JavaScript. It's been noted that the Turkish Ministry of National Education [7] is known to be the only government site using gov. domain using device fingerprinting on a large scale. When the user visits the website, they are tagged with a flash object which repeatedly sends all kinds of information about the user's device.

Operating System Instance Fingerprinting

It is described earlier that JavaScript can obtain operating system information. JavaScript can also get not only the operating system but also the other specifications. Through the fingerprinting methods, the color depth of the monitor and the display dimensions can also be identified. Fingerprinting companies are mainly the clients of getting these things done correctly. Moreover, flash objects used in the cookies like flash cookies also reveals other amazing information like audio capability, webcam capability, printing capability, etc. [1,8]. So, it is seen that almost anything can be revealed by web tracking methods which are improving day by day.

Browser Fingerprinting

It is most commonly known as Cookie Monsters or HTML5 and CSS fingerprinting [1, 2]. This fingerprinting works through several types of tags. There are a lot of new tags that have been introduced recently. There are almost 242 tags [9] has been added which might be suitable for the HTML5 fingerprinting. A brief walk to how fingerprinting is collected [10]-

• When a connection is made the user agent and accept header are automatically delivered to the site.

• Access to the plugins is given by the JavaScript mainly.

• There is a flash plugin in the system and its API provide specific attributes regarding the system e.g. operating system version, font lists, versions etc.

• Canvas fingerprinting also works in the same way.

For example, in the website https://amiunique.org the following information is collected:

• the User-agent header
• the Accept header
• the Connection header
• the Encoding header
• the Language header
• the list of plugins
• the platform
• the cookies preferences (allowed or not)
• the Do Not Track preferences
• the screen resolution and its color depth
• the use of local storage
• the use of session storage
• a picture rendered with the HTML Canvas element
• a picture rendered with WebGL
• the presence of AdBlock etc.

A cross browser fingerprint test [10] has been shown above. In both the cases the fingerprinting test has been able to collect all the necessary information from the system including the browser information, font history, resolution, and operating system version. Fingerprinting mechanism is an effective

mode of tracking a user across the web and gathering information about the user needed for necessary purpose.

Canvas Fingerprinting:

Canvas fingerprinting [1] is one of the newest modes of collecting information from the web. It is like drawing an element on the screen, more like a canvas drawing in the display. Canvas is a property of HTML5 which is used to draw animations via JavaScript virtually. The fact has shown that the same graphic can be rendered in different devices differently because of the difference of the system property. The image processing powers and engines are different. The processor speed, graphics, operating system versions are different. So, the rendering must be different based on these properties. The test can be done online (www.browserleaks.com).

Ever Cookie (Super Cookie)

We have already seen different types of cookies which help to track the user. Some cookies use local shared storage and stay for a long period of time. These cookies can be deleted from the browser. If someone knows the way to clean the cookies it can easily be deleted. Here a brief introduction of Ever cookie is given. From its name, it is clearly understood that this type of cookie intends to stay in the user's computer permanently. It is, in fact, a JavaScript API [11]. It was discovered by Samy Kamkar. He thought of making such a cookie which could last forever even the user has deleted his cookies. In fact, it does so. It uses different storage mechanisms to store the data. Later when the cookie is deleted from one storage it tries to retrieve the same from different storage locations available. So, it is seen that the cookie is in fact recreated after every deletion. The storage mechanisms that are used while the recreation is below [11]:

- Standard HTTP Cookies
- HTTP Strict Transport Security (HSTS) Pinning
- Local Shared Objects (Flash Cookies)
- Silverlight Isolated Storage
- Storing cookies in RGB values of auto-generated, force cached
- PNGs using HTML5 Canvas tag to read pixels (cookies) back out • Storing cookies in Web History
- Storing cookies in HTTP ETags
- Storing cookies in Web cache
- window.name caching
- Internet Explorer user Data storage
- HTML5 Session Storage
- HTML5 Local Storage
- HTML5 Global Storage
- HTML5 Database Storage via SQLite
- HTML5 Indexed DB
- Java JNLP Persistence Service
- Java CVE-2013-0422 exploits (applet sandbox escaping)

Some characteristics of ever cookie are

- Cross-browser accessibility. If there are flash locally shared object cookie, Silverlight isolated storage etc. in play.
- The client does not need to install anything, as a result, it can run silently under the hood. On the other hand, the server at least needs to have access to the JavaScript cookies.

5 How not to be tracked

Cookie clearing would be the undoubted advice from many and of course towards many users. As average users do not care for the cookies stored, instead they feel happy getting the system behaves as they want to be while behind this, they are ignorant that their data might be used to recognize the patterns of their activities online. Private browsing mode is one of the most straightforward solutions to avoid caching or cookie storage. Safari private browsing is one of the most secure as it also prevents cookies ever to create and track the users online [11]. To remain anonymous online Tor browser can be a more delicate choice as it has been long praised for protecting anonymity online. It connects through various relays over the world to connect to the destination site. The onion routing makes almost impossible to track the online user activities. Though there has been some improvement in tracking users online even if they are using Tor, yet it is one of the best means to remain anonymous. VPN (Virtual Private Network) is another way of not being tracked or hiding the real information as it is masked under a different network. So, this can also be a useful sword against web tracking. Though it is virtually impossible to hide one online, it worth a try to use each possible means so that a user cannot be tracked down by some third-party organizations for their benefit, not ours and giving our information to use it on their purpose.

6 Conclusion

As nowadays web tracking has become a threat to everyone and most people are not aware of its consequences. So, in this paper we precisely discussed about the throwback and tools of web tracking. Then, we come to an end by putting an elucidation for how everyone can preserve themselves from being web tracked.

Acknowledgment

I would like to thank my co-authors and my fellow colleagues for their support in this paper.

References

[1] Bujlow, T., Carela-Español, V., Solé-Pareta, J., & Barlet-Ros, P. (2015). "Web tracking: Mechanisms, implications, and defenses." arXiv preprint arXiv:1507.07872

[2] T. Unger, M. Mulazzani, D. Fruhwirt, M. Huber, S. Schrittwieser, and E. Weippl, "Shpf: Enhancing http (s) session security with browser fingerprinting, in Availability, Reliability and Security (ARES)", 2013 Eighth International Conference on. IEEE, 2013, pp. 255261.

[3] K. McKinley, Cleaning Up After Cookies, iSEC PARTNERS, Tech. Rep., December 2008, accessible: https://www.nccgroup.trust/us/our-research/cleaning-up-after-cookies/

[4] F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web", in Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation, ser. NSDI12. Berkeley, CA, USA: USENIX Association, 2012, pp. 1212. [Online]. Available: http://dl.acm.org/ citation.cfm?id=2228298.2228315

[5] M. Nasir, "Tracking and Identifying Individual Users in a Web Surfing Session", Computer and Network Security, Middlesex University, London, Tech. Rep., January 2014, accessible: https://www.academia.edu/4214725/Tracking and Identifying Individual Users in a Web Surfing Session

[6] https://wizzair.com/static/docs/default-source/downloadable-documents/cookies_new_address_en_6996a2a8.pdf

[7] Ministry of National Education TC MILL EGITIM BAKANLIG I, Available: http://meb.gov.tr

[8] Vasilyev, V. (2018). Anonymous browser fingerprint. Contribute to Valve/fingerprintjs development by creating an account on GitHub. JavaScript. Retrieved from. https://github.com/Valve/fingerprintjs (Original work published 2012)

[9] Am I unique? (n.d.). Retrieved November 1, 2018, from https://amiunique.org/faq

[10] Cross-browser fingerprinting test 2.0. (n.d.). Retrieved October 1, 2018, from https://fingerprint.pet-portal.eu

[11] Samy Kamkar - evercookie - virtually irrevocable persistent cookies. (n.d.). Retrieved September 15, 2018, from https://samy.pl/evercookie/