JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 1 No. 1 (2020)
https://doi.org/10.35877/454RI.jinav174

Identifying Actual Users in a Web Surfing Session using Tracing and Tracking

Umar Danjuma Maiwada[a]*, Aminu Aminu Muazu[a], Izaddeen Kabir Yakasai[b], Rufai Yusif Zakari[c]

[a]*Department of Mathematics & Computer Science, Ummaru Musa Yar'adua University Katsina, Nigeria*
[b]*Department of Computer Science, Middlesex University London, United Kingdom*
[c]*Department of Information & communication Technology, Skyline University, Nigeria*

Abstract

Times New Roman 9pt Nearest neighbor analysis is an analytical method that can be used to determine a dispersion pattern, whether uniform, random, or cluster. Nearest neighbor analysis takes into account the distance, the number of points of dispersion location, the area width, the final result of this analysis is the nearest neighbor index (T) whose value ranges from 0 to 2.15. To determine the factors that influence the pattern of spreading occupancy, in this study using correlation analysis (r). The results of this study based on the calculation of nearest neighbor analysis of the pattern of spreading occupancy in Kendari City as a whole has a pattern of dispersion that tends to random pattern means the distance between the spreading points of occupancy does not have the same distance, and the results of correlation calculations obtained factors that affect patterns of spreading occupancy that tend to be randomized. The most influential factor with positive relation is population growth factor, it can be seen on the correlation result where population growth factor has the highest correlation value is 0.618 and the most influential factor with negative relation is the factor of education facility it can be seen on correlation result where factor Education facility has the highest negative correlation value is -0.526.

Keywords: Type your keywords here, between 3 and 6, separated by semicolons.

1. Introduction

The likelihood of comprehending what a client is doing over the Internet or web has been an issue subsequent to the start of the Internet. Advertising sites and even corporate organisations needed to realise what their customers or clients are dependent upon over the web. Advertising organisations that have sites on the web needed to know how many individuals visit their website or the measure of time a specific client spends on it. Furthermore, corporate organisations needed to know whether their representatives are not doing what they are not supposed to do, and also to screen any sort of suspicious activities happening. (Acar, et.al, 2014)

During that time in 1994 to say, Netscape created an innovation, which was called Magic Cookies that we all know now as cookies.(Olejnik, et.al., 2014) This permitted the server tracking either the measure of time a client spends on the site or which client went to the site. This was an achievement for many people while others see it as a break of security and will prompt conceivable malignant attacks.

* Corresponding author.
E-mail address: umar.danjuma@umyu.edu.ng (Umar Danjuma Maiwada)

Additionally, proxy servers now are utilised to track users. They have a cache in them that aid in keeping a proxy log of visitors. Since it is straightforward to the user, they can log all the URL's a user visited with time and date as well as the IP address of the user (Perlman, et.al, 2016; Acar, et.al, 2014). All these utilisation of cookies and Proxy servers can get you regularly the user's IP address at most after all the tracking. It is however not worthy that; a user's IP address is not ensuring you the actual user as anyone can be using the PC.

2. Problem With Tracking Web Users

At the point when a user starts to surf the web, a session is made between the computer being used and the server through the method of HTTP authentication, which includes a three-way handshake. At the point when a browser sends a request to a server, the data is then sent from the server as a response, a history of this is kept in the history folder of the browser. Additionally, this history can be erased from the folder where it is put away. For the case of an attacker or a criminal erasing the history can be seen as if no website was ever gone by the system (Herrmann, et al, 2010; Ibrahim, et al, 2019).

Tracking is generally used by advertising networks to build up detailed profiles for pinpoint ad-targeting. If you've ever visited a business' website and seen ads for that business on other websites later, you've seen it in action.

This can be overlooked if the security engineer isn't as talented as the attacker or criminal, but when looking at cookies, things are a little bit different. To security specialists, cookies are seen as a more elevated method of tracking exercises on the web than utilising history documents; when a client requests for something from the server and the server responds, the server produces information gotten from the client and stores this on the client's computer, this data put away is the cookie. What is placed inside the cookie are the insights with respect to the client which are the name of the used computer, IP address used, operating system used and other information with regards to the per using history; the IP address of the server, which is all put away on the user end. This likewise can be erased, and this raises the task of tracing, tracking, and identifying somebody who used the computer around then (Millett, et.al, 2001; Olejnik, et.al, 2014).

Like a magic trick, where the magician pulls a rabbit out of a hat that was once empty, we also have an enormous task of providing a way in which we can identify an actual individual who surfed the web at a given time by the use of server logs and cookies and being regardless of it being a single machine or multiple machines (Unger, et. al, 2013).

3. Proposed Solution

At the point when messages are sent through the browser to the server, a protocol is utilised to administer how the messages are sent, and the protocol is Hyper-Text Transfer Protocol (HTTP). The issue concerning this protocol is that the protocol is stateless; by using HTTP, it doesn't recollect the movement of any activity been finished. In place, to defeat this issue, cookies were made to keep up different conditions of a session, authenticate a client, furthermore customise the activities of the session. Cookies on their own, cannot fully track an actual individual of a web session because it wasn't designed for that purpose, but it can be used to track individuals as a result of the information it contains on user. (Millett, et. al, 2001; Tiwana, 1999).

A possible solution to this problem will be to use the cookies, which are stored on the client's machine, sometimes referred to as HTTP cookies because the server generates and store them on the client's machine. After these cookies are stored on the client machine, they can be looked at, accessed and interpreted in order to increase the chances of tracking an individual. Cookies are especially fundamental instruments, which are not, just used to track along on a web session however for different purposes and if cookies happen to get into the wrong hands it can be a huge issue for honest to goodness clients of the web. By getting hold of these cookies, you can get points of interest of distinctive clients' data, for example, IP address of the PC, URL name of pages viewed where the user visited, and this can give anybody the client's surfing patterns (Acar, et. al, 2014; Millett, et.al, 2001).

At whatever point a user or client surfs different websites, and fills data needed by the site, this data are put away in the database, and a portion of the data is likewise put away in the cookies. An illustration, on a site where you are requested to put username and password, and you give the obliged data. In some cases your browser inquires as to whether you will wish to save the password, the data is sent to the server and affirmed while that is done, a treat is produced and put away on the client which contains your username and password. That is the reason why you do not need to give your username and password because the browser has officially mapped that username and password to that site. This can likewise be done when a client fills a demographic structure for an e-commerce webpage when performing Internet shopping; shoes size, trouser length. This are put away in the cookie, and it can be hazardous if in wrong hands as it contains subtle elements of a client without having their assent, and most shopping destinations escape with it in light of the fact that most customers neglect to per use and comprehend the terms and conditions set up. “Information is expensive and can be sold”. (Perlman, et.al, 2002; Tiwana, 1999).

Users who are exceptionally security cognisant can have a tendency to parody their IP address to imitate another person's IP address, so they need to stay unknown when surfing the web, and may have the learning so as not to be tracked. This makes it hard to track such users. Yet using IP address to track a user is that sensible because distinctive systems can be utilised by the same individual or generally IP address changes. Once in a while occurrences happen where numerous clients, for example, in a college get to the same PC now and again. An individual can be tracked in this setting where every individual needs to log in with his or her own separate username and password. IP addresses can now and again be hard to get if by any chance an attacker or criminal plants a sniffer on a client's machine with a specific end goal to listen stealthily on the network (VPN) which utilises a tunnelling methodology to correspond with the focal network server from a remote location. Basically, using IPSec protocol, the IPSec needs to modes; transport mode and tunnel mode. At the point when in tunnel mode, it is more troublesome in acquiring the IP address of the system used to surf the web.

Glancing back at cookies, a machine can't do without cookies however don't be excessively terrified that they are bad because they are proficient apparatuses and also secure, this is because it is only the machine that generated them that can be used to open it or you can get the data found in a cookie on a server log, but this can only be access by an authorised user.

Whenever a client requests information from a server, the information regarding that request is stored also on the server logs. Some of the details stored on a server log are; the IP address, host name, MAC address of the machine used. That’s why the FBI can come knocking on your door first thing in the morning when they have all the details of the information you requested. Don’t do anything naughty though.

4. Implementation

In this section, we will be looking at how server logs and cookies are implemented so as to track and identify an individual. Before going into how this is done, Kurose identified that cookies have four components; (i) a cookie header line is placed in a HTTP response message from the server; (ii) there is also a cookie header line in the second and other HTTP request messages except the initial HTTP request; (iii) a cookie is generated by the server, stored on the client’s machine, and managed by the client’s browser; (iv) a back end database of the web site is created which is related to the information on the cookie. (Tiwana A, 1999) Say a certain user named Mardiyya joins Amazon using the Mozilla Firefox browser. When the browser requests information from the Amazon web server, an identification number is created and assigned to her account, this also creates an entry into the back-end database of the web site which is indexed by the same identification number. The web server responses the request and places this identification number in the HTTP response header (Set-cookie: number) for this scenario we will use an identification number of 4576.

Your browser also sends a user agent every time you connect to a website. This tells websites your browser and operating system, providing another piece of data that can be stored and used to target ads.

Fig. 1. Browser Fingerprinting

Browsers are actually pretty unique. Websites can determine your operating system, browser version, installed plug-ins and their versions, your operating system's screen resolution, your installed fonts, your time zone, and other information. If you've disabled cookies entirely, that's another piece of data that makes your browser unique.

When the browser gets the response and reads the HTTP response message it sees the Set-cookie command and it stores this as a cookie on the client machine, and it manages the cookie from then on. By storing this cookie it is provided with more details such as the server's hostname and the identification number. So whenever Mardiyya visits Amazon and requests a page as she surfs, the browser consults the cookie and gets the identification number and creates a cookie line in the HTTP request. Here is a syntax of the cookie line: - *Cookie: <name>=<value> [; <name>=<value>]*. This syntax contains the name of the cookie, and the value is the identification number of the user in this case Mardiyya.

So each time Mardiyya requests a page from the Amazon web server, it puts this cookie line in the HTTP request header. By doing this, the server using the identification number 4576 keeps track of her activities as she surfs the site. By using cookies, the server can keep records of the items purchased, requested, and items ready to be paid for in the cart or basket so as for her to pay at the end of her shopping session. All this information is stored on the client side in the cookie every time a HTTP response is sent back to the client by also putting a cookie line in the HTTP header; this information is also logged on the server for security and referencing purposes.

After a couple of weeks, Mardiyya requests a page from the Amazon web server and the server are still able to identify her identification number. Why is this possible? That is because the cookie is still stored on the client's machine, even if the cookie is deleted, when it sends the initial HTTP request, it contains information which can be mapped back to the 4576 identification number by using the hostname and IP address details sent by the client. If Diaz decides to provide more information regarding her account, the information provided is added and mapped to the identification number and sent back to the client machine and is stored (Tiwana A, 1999).

Cookies are small pieces of information websites can store in your browser. They have plenty of legitimate uses – for example, when you sign into your online-banking website; a cookie remembers your login information. When you change a setting on a website, a cookie stores that setting so it can persist across page loads and sessions.

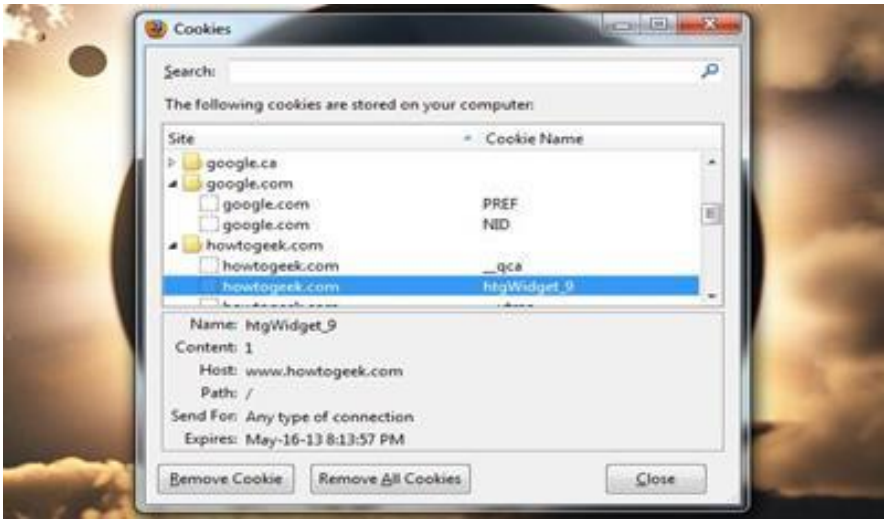

Fig. 2. Cookies & Tracking Scripts

Cookies can also identify you and track your browsing activity across a website. This isn't necessarily a big problem – a website might want to know what pages users visit so it can tweak the user experience. What are really pernicious are third-party cookies.

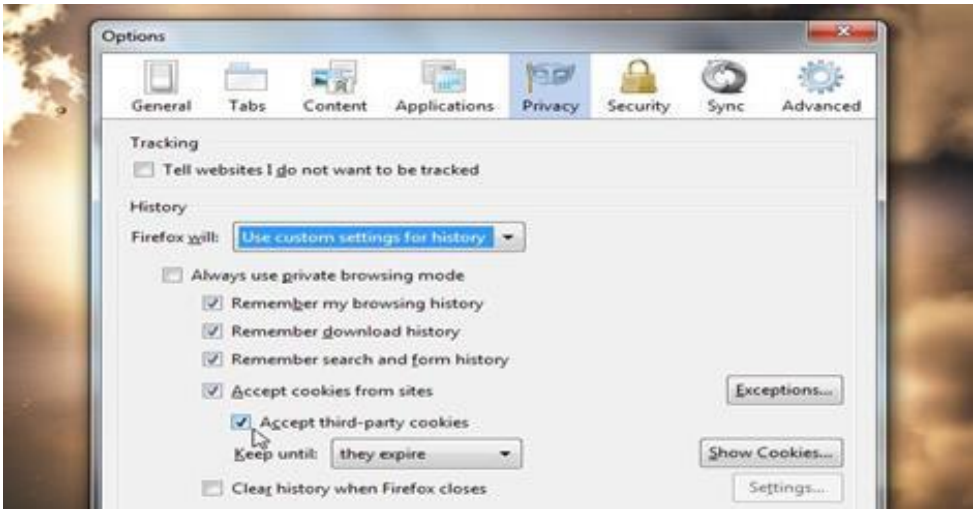

Fig. 3. Third-party cookies

Using this scenario, we can conclude that it is possible to use cookies to track an individual by gathering more cookies of the sites an individual has visited. It is been debated whether there is a better way of keeping people's details online private from both authorized and unauthorized users. Web site operators also use the information gotten from the cookies to evaluate on what users prefer to do on their site, what get their customers interest and by doing so they can build a profile of a user.

4.1. Super Cookies

You can clear your browser's cookies — in fact, we've got a guide to clearing your browser's cookies. However, clearing your cookies isn't necessarily a solution – "super cookies" are increasingly common. One such super cookie is ever cookie. Super cookie solutions like ever cookie store cookie data in multiple places – for example, in Flash cookies, Silverlight storage, your browsing history, and HTML5 local storage. One particularly clever tracking method is assigning a unique color value to a few pixels every time a new user visits a website. The different colors are stored in each user's browser cache and can be loaded back – the color value of the pixels is a unique identifier that identifies the user.

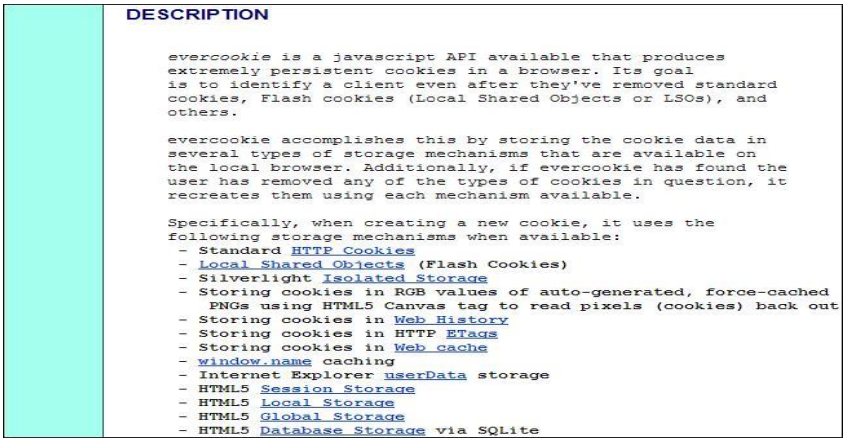

Fig. 4. User Agent

When tracking an individual, an experienced security administrator will look at the server or proxy logs with regards to the identification number of the user or even go as far to eaves drop on the network. The logs as mentioned before are stored for security and reference purposes and they keep record of all request and response sent between the client and the server. Sometimes the request-response isn't done by the server but the proxy because it handles some HTTP request and provides a HTTP response that has been cached on it that rather going all the way to the main server to get the information; don't be scared because all the event performed are stored on the logs.

On the off chance that a client is utilising an extremely secure network to get to the server, it comes hard to eaves drop on the network and not all the data needed can be gotten. So as to conquer this test, here and there in movies we see where a spy or mystery specialists introduces an application on the associates machine without their assent, with information in order to get more dirt on them. That is something that may be done so as to get the data, how that could be possible is by different means; system overhaul is a decent method for doing that. Sorry if you get caught you could wind up in jail, that is if you did it unlawfully. This application keeps record like a cookie or log and the data gotten is sent to the individual who introduced the application on appeal or occasionally. Anyway how about we not overlook that as there are softwares like that, there are likewise softwares that help counteract such things yet keeping a system from putting away cookies, furthermore data of the machine in entirety. Indeed, even browsers erase cookies over a certain time.

At long last, with respect to of the benefits and faults of cookies, we can say we can track a single person through the use of cookies together with learning of conceivable websites the client may have gone by surveying the history folder of the browser and server logs.

4.2. HTTP Referrer

When you click a link, your browser loads the web page you clicked and tells the website where you came from. For example, if you clicked a link to an outside website on How-To Geek, the outside website would see the address of the How-To Geek article you came from. This information is contained in the HTTP referrer header.
The HTTP referrer is also sent when loading content on a web page. For example, if a web page includes an ad or tracking script, your browsers tells the advertiser or tracking network what page you're viewing.
"Web bugs," which are tiny, one-by-one pixel, invisible images, take advantage of the HTTP referrer to track you without appearing on a web page. They're also used to track emails you open, assuming your email client loads images.

5. Conclusion

In outline, we set out to check whether it was conceivable to track and recognize a real person who surfed the web at a certain time, and we can say we did accomplish our objective. By using cookies, server logs, eaves dropping or not withstanding using softwares, we perceived how a client can be tracked by basically appointing an exceptionally unique identification number to a HTTP request or HTTP response. Then again, tracking a specific individual could just be conceivable if the individual gave secret subtle elements by filling forms containing; name, address or email.

This ought not throw anybody off shopping online or doing essential things; how might the things get to your location if that you don't give your location, or be kept insider savvy on the where abouts on your thing without your email. As per Eugene H. Spafford, the most ideal method for security is to be totally offline, detached from the web, and have it flushed.

Acknowledgements

This work is partly supported by a research grant (TETFund) from Umaru Musa Yar'adua University Katsina, Nigeria.

References

Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., & Diaz, C. (2014, November). The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (pp. 674-689).

Banse, C., Herrmann, D., & Federrath, H. (2012, June). Tracking users on the internet with behavioral patterns: Evaluation of its practical feasibility. In *IFIP International Information Security Conference* (pp. 235-248). Springer, Berlin, Heidelberg.

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

Herrmann, D., Gerber, C., Banse, C., & Federrath, H. (2010, October). Analyzing characteristic host access patterns for re-identification of web user sessions. In *Nordic Conference on Secure IT Systems* (pp. 136-154). Springer, Berlin, Heidelberg.

Ibrahim, S. K., & Jebur, Z. T. (2019). Impact of Information Communication Technology on Business Firms. *International Journal of Science and Engineering Applications*, *8*(2), 53-56.

Olejnik, L., Minh-Dung, T., & Castelluccia, C. (2014). Selling Off Privacy at Auction. In *Annual Network and Distributed System Security Symposium* (NDSS).

Millett, L. I., Friedman, B., & Felten, E. (2001, March). Cookies and web browser design: Toward realizing informed consent online. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 46-52).

Perlman, R., Kaufman, C., & Speciner, M. (2016). *Network security: private communication in a public world*. Pearson Education India.

Tiwana, A. (1999). *Web Security*. Digital Press.

Unger, T., Mulazzani, M., Frühwirt, D., Huber, M., Schrittwieser, S., & Weippl, E. (2013, September). Shpf: Enhancing http (s) session security with browser fingerprinting. In *2013 International Conference on Availability, Reliability and Security* (pp. 255-261). IEEE.