International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 04 Apr. 2024, pp: 312-318 www.ijaem.net ISSN: 2395-5252

Analyzing the Efficiency of Web Tracking Systems in Ensuring Eprivacy Compliance: A Comparative Study

[1]Michael Taiwo, [2]Wilson Sakpere, [3]Emmanuel Adediran
[1,2&3]*Lead City University, Ibadan, Nigeria*

Date of Submission: 05-04-2024 Date of Acceptance: 14-04-2024

ABSTRACT: Every time we open our computer systems, laptops, or mobile phones to browse the web. We visit different web sites and open diverse hyperlinks or look for gadgets, what to shop or classify an advert. After a while, a separate website gives us a picture of the equal element we have been seeking out. That means we are being tracked and delivered tailored classified advertisement depending on our previous pursuits and location based totally on cookies content material. What makes the situation complex is that they may not be accredited or permitted to do that. The work aimed to study how are visitors of popular Nigerian websites tracked and how their privacy is affected. For that, all the cookies were identified by category and type. We determine if popular Nigerian webpages comply with the ePrivacy Directive to understand if visitors of popular Nigerian websites were tracked without consent. Finally, we calculate which are effective defense methods against third-party tracking. This study has been based on 22 popular Nigerian websites ranked by Amazon Alexa.com. And for the crawl a python code that generate Web crawling activities was created and executed for crawling purpose. The results showed that 64% of the popular websites use third-party cookies, and most of these websites track visitors without their consent.

Keywords: Web tracking, cookies, ePrivacy Directive, third to third-party tracking, web beacon, Ghostery, Do Not Track, private browsing mode.

I. INTRODUCTION

Most internet users think their surfing information serves responsibly, many assume their identities have none to hide, and a sizable portion of them think they can't utilize certain features of certain web services if they don't think about and agree to certain terms. They are unaware of how their data is being collected, processed, and used, or how using a website service may affect their privacy. It is possible to identify frequently frequented websites by calculating the range of daily visits and landings on an ongoing basis after every trip to a well-known website[1]. The goal of this research is to understand how users of well-known websites are tracked and how this affects their privacy. An analysis of cookies by kind and category is provided in this paper. It is possible to identify the extent of tracking and the compliance of prominent websites with the ePrivacy Directive[2]. Additionally, the outcome displays all cookie-related information for each website category and provides information on whether prominent websites track their visitors without their knowledge. This may be done by checking if websites adhere to the ePrivacy Directive. All of the data are examined and listed in the last section of this study. The 50 well-known Nigerian websites ranked by Amazon Alexa.com served as the basis for this study. The five categories of these websites were news, e-commerce, online services, social media, and services. online crawling was[1] created in the Python programming language to carry out this investigation[3]. Figure1 shows how data and information of web users are grabbed by cookies and are converted into financial means for the organization by selling cookies which are mostly their data, information and users web experience with other organizations which are mostly third-party that are advertisement company and it also aid them in reaching to more audience for their products propaganda, propagation and could also fall into the wrong hands[4].

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 04 Apr. 2024, pp: 312-318 www.ijaem.net ISSN: 2395-5252

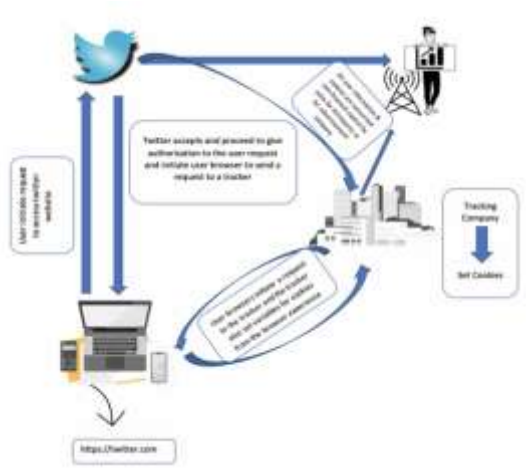

Figure 1: Web Tracking Illustration with www.twitter.com (Authors)

A. Problem Statement

Current web tracking research indicates that many web users have unintentionally lost important or incidental data and information online. Due to the issue of data loss in cyberspace, users frequently lose their personal information carelessly, which leads to fraud, impersonation, web stalking, access to the victim's bank account, and loan applications using the victim's personal information. Most of the time, internet visitors (also known as online users) are unaware of any programmed tracking device that has been integrated via cookies into a website's consent page, resulting in a backend information drop about users' interactions with the website. The most crucial consent page that appears on the screen after visiting a website and gives us the option to cooperate with it by agreeing to the terms and then continuing to communicate with the legitimate site, or to ignore it and lose the option to cooperate with the sites, is typically where most web users lose information. They could also lose a lot of information through phishing links that are frequently used.

B. Aim and Objectives

The paper developed a web crawler program that can track data and information of 22 popular Nigerian website and their cookies and to also find out how visitors of popular websites are tracked and how their privacy is affected. The specific objectives of the study are to:

i. Develop a web scraper program for tracking of cookies.
ii. Collection of different cookies
iii. Calculate and analyze the first- and third-party cookies[3].
iv. Distinguish the cookies by type and their attribute.

II. LITERATURE REVIEW

Several research have revealed that web tracking is a study that needs attention without doubt, as it helps in revealing how to curb the insecurities and irregularities on the web space that might be potent to whisk away with users' data and information without their consent. Research shows that frequent diagnosis and evaluation of this study will lead to further analytical revelation of more measures to curb the means at which data are lost in the web space[6]. But what is web tracking? "Tracking means it can be an act or the process of following something or someone". Websites mostly gather technical data: IP address, screen determination, or the browser used at the website visit. If cookies are used, there are two main assignments: either make browsing experience better or collect browsing info and share with other counterparts[7]. Already in the year 2001, some conclusions were made about cookies and tracking cookies, that are actual 21 years later:

i. Some web users do not know about cookies and how websites administrator might use it against them.
ii. Some web users know how cookies can track them and are unconcerned about it.
iii. Some web users do not know which cookies they will accept and accept all of them.
iv. Some web users want to assume that they are protected from bad usage of cookies and believes that web regulations will help them.

This study will enlighten users on how to understand when being tracked, and how to avoid being monitored. Whenever a web user visits a cookie induced website, the first indication on the website is a cookie banner or cookie notice that will signal the user to interact with the cookie toggle with the displayed option in other to further interacts with the services the website provides[8,9].

The opposite situation occurs when the banner is not used and tracking still occurred (by looking at installed cookies). There can be a wrong understanding that all websites are tracking website visitors[10]. Web track was initially developed to facilitate better marketing, and it is the same with most websites. There are several mechanisms and implications of how tracking can happen. As cybersecurity is about defense, there are several defense methods to minimize the impact of mechanisms and implications. There are several options, how to write about web tracking and there are several organizations that are interested in using

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 04 Apr. 2024, pp: 312-318 www.ijaem.net ISSN: 2395-5252

data and information of web users for different purpose, either for advertisement and marketing, impersonation, fraudulent activities which are illicit and might be complicated for the data owner[11]. For whom the collected tracking data is necessary: i ii Advertisement companies collect information to influence and produce tailored ads to users. The main aim is to find out users' interests in different categories[12]. Law enforcement and intelligence agencies collect information to perform the tasks assigned to them[13]. iii Website owners are interested in their website analytics. It also helps to produce a tailored website advertisement[14].

A related result is that because advertising businesses obtain tracking information and are not concerned about data leaks, they value user data collection more than user privacy. Tracking is more akin to a connection between customers and internet advertising corporations than it is usually a bad thing[10]. Most of it goes to advertising businesses, but if customers are dissatisfied, they can always erase the cookies. Some businesses provide "cross-device" tracking rather than third-party cookies. It implies that a single individual may be recognized via search across several devices. Among the information to collect are IP addresses and the operating system ID[11]. If a website visitor has just searched for bread on Google and is currently surfing Twitter, they may see a bread advertisement there. A high-level review and conception of online tracking has been done, with a focus on the commercial, technological, and privacy aspects. The news category is another illustration[12]. Top users of tracking cookies are news websites. Cookies have been used by news websites since they first appeared. It has been noted several times that news websites have the greatest cookie coefficient. For instance, the most well-known Nigerian websites with trackers are news websites, which also contain many trackers[13].

A. Web Tracking Privacy Implications

"The definition of online privacy is the level of privacy protection an individual has while connected to the Internet. It covers the amount of online security available for personal and financial data, communications, and preferences"[14]. Regarding privacy concerns and data gathering abuses, Timothy Libert published a piece in the New York Times. The issue with news websites is particularly difficult. According to the initial investigation, over 50 different businesses follow individual data to gather as much personal data as they can for commercial use[15]. Data on readers of news websites was shared with third parties, according to an analysis of their privacy rules. Targeted adverts on these news websites and other websites are the consequence of the usage of third-party cookies on these websites to gather information about user activity[16]. The privacy of individuals may be at risk by web tracking[17]. Firms have access to visitor information and can compile thorough internet profiles. Even when visiting websites you trust, visitor data may still be captured and sold to a rival[18]. Mostly private details, but what exactly is the confidential data in which various parties are interested? Name, location, age, gender, and a special computer identity are all examples of personalization in the context of online monitoring. The user's personally identifying information, such as age or gender, can be discovered by examining the HTTP headers[19]. The visitor receives a cookie with a unique ID when they first visit the website, but as many other websites utilize the same third-party cookie, it is possible for the visitor's unique ID and user data to be exposed to third parties. Most websites allow third-party cookies without the user's knowledge on average, and huge firms with US backgrounds (like Google) are typically the ones that do it. Few websites wait for user approval before installing cookies and the vast majority do not offer cookie banners[20]. The two primary privacy problems in online tracking that might result in damaging web tracking situations are private data leaking and identifiability. The most popular technique of identifying utilizes a visitor's IP address to follow them across websites they have visited or uses cookies to track them across pages they have called[21]. To assess if there is a chance of user privacy being compromised, the writers made the decision to examine the third-party tracking. They opt to focus more on the websites that employ a specific word tied to users' privacy (such cookies, cards, or passwords) rather than prohibiting all access tracking. The tracking rate decreased as a result, going from 71% to 24%[22]. They concluded that the usage of third-party trackers raises privacy and online morality issues, and that the more third-party trackers there are connected to one another, the more users might potentially be tracked[23].

B. Cookies

How do cookies work? A stateful browser-server interaction in a stateless protocol is made feasible through cookies. Cookies, or HTTP (Hypertext Transfer Protocol) cookies, are produced and updated by the server, saved by the browser, and sent back and forth between the browser and the server. Web developers typically create cookies. There are several types of cookies.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 04 Apr. 2024, pp: 312-318 www.ijaem.net ISSN: 2395-5252

First-party cookies are created by the domain that hosts the website, and no information is shared between them[24]. The converse is true with third-party cookies, which indicate that information is sent from one website to another, and that the first website has granted permission to gather data. The third party cookies may thus violate your privacy[25]. By categories, we can classify cookies. The most common categories that were used in the context of this work are:

i. Strictly Necessary cookies: Cookies are essential for the provision of the website service.
ii. Performance cookies: Provide statistical information on website usage.
iii. Functionality cookies: Provide enhanced functionality for all website functions. and its
iv. Targeting/Advertising cookies: Create profiles or personalized content. Set mainly by third parties and with the highest privacy risks to visitors.

Although generally safe, cookies cannot ensure privacy. If there is an instance of unsecured information, cybercriminals and fraudsters can utilize cookies to monitor user online activity. Cookies have a lifespan division. The average cookie has a life span of only approximately six months. The browser should expire the cookies at this point. A user can distinguish between a temporary cookie and a tracking cookie using the cookie expiration time. There are two categories of cookies:

i. Non-persistent cookies: mostly session cookies that are deleted after the session. Persistent cookies. Stored in browser memory with expiration time26.

Research titled "Towards a global perspective on web tracking" highlighted Nigerian websites. Nigerian websites are among the TOP 6 AU nations with the greatest quantity of cookies with the longest expiration dates27. Many third-party cookies on websites have a lengthy validity term, which is against Australian legislation. Tracking cookies are the most frequent cookies with a long expiration date, and 80% of third-party cookies have an expiration date of one month or longer.

Figure 2. Cookie syncing between websites[31]

The process of third-party cookie synchronization haven explained implies that a visitor uses three separate IP addresses to view three different websites. The third website will know that this is the same person if it utilizes the same cookies as the other two. The top 20 third-party monitoring firms and their third-party domains are shown below[31]. According to a US-based study on third party cookies, news websites depend on third parties more than other types of websites do. 95% of news websites use third-party material, and they do so in such large quantities that the number exceeds trackers of the 500 most popular websites in the same nation, according to an analysis of news websites in various countries. The US-based internet corporations Google, Facebook, Amazon, and Twitter have a vested interest in gathering information on every news website. The first technological business established in the EU only had cookies on 7% of websites. With at least one third-party tracker, 46% of prominent websites on Alexa.com were monitored. 29% of the websites were monitored by at least five outside trackers32. The fact that Google, a third party, gathers cookie data from 25% of the most popular websites according to Alexa.com.

III. RESEARCH APPROACH

This chapter explains the data gathering methodology, the analysis required for the tracking, monitoring, and evaluation process, as well as how to fulfill both functional and nonfunctional needs. The strategies for reaching the goals outlined in this thesis will be discussed in more detail below, along with a draft and an explanation of the analysis. The purpose of the study and the data collecting, and analytic techniques used to accomplish it are described in this thesis'. The analysis of the most popular or topmost frequently visited website was sorted out from Alexa.com TOP50 visited Nigerian websites (Appendix 1). First crawl showed that many popular websites are back-end websites (no cookies). From there data

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 04 Apr. 2024, pp: 312-318 www.ijaem.net ISSN: 2395-5252

connected with Nigeria websites were sorted out. They were categorized by their content and used IP address was contacted by the web-based crawling program which was designed for the achievement of this thesis. Category was chosen by the web content, and it was divided into five categories: news, web services, e-commerce, social media and services (Table 1). Web Services and services were differentiated by selling aspect. If there is possibility of buying something, paying or carting goods online, then it belongs to services category. Alexa.com was chosen, as in previous research it has been pointed out, that tracking occurs more on the higher ranked websites. In the Alexa.com, some popular websites on the list generated a general error code and in manual check that website does not exist. In general, 22 websites were added for analysis.

IV. RESULTS

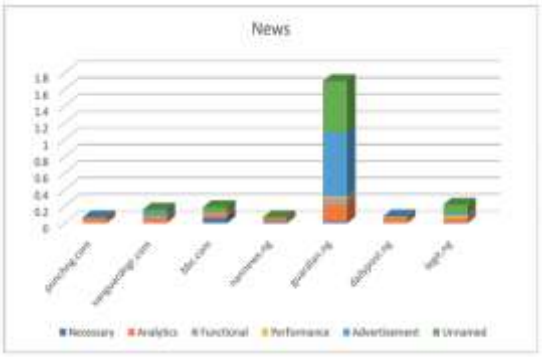

Figure 3 Representation of News Categories Cookies

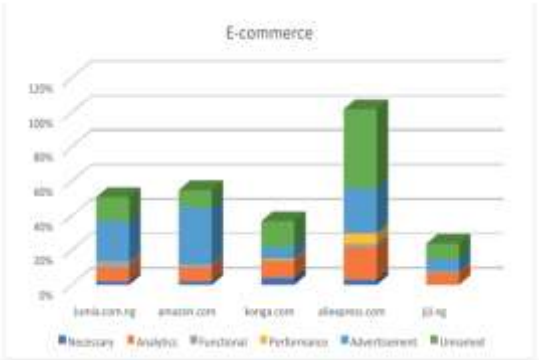

Figure 4 Representation of Ecommerce Categories Cookies (Researcher Taiwo, M., 2023)

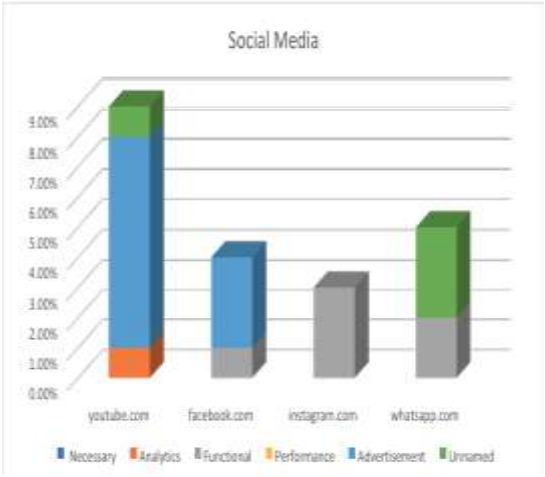

Figure 5 Representation of Social Media Categories Cookies (Researcher M, Taiwo, 2022)

V. CONCLUSION

All in all, this proposal is created fully intent on assisting with making mindfulness on how a site or site executive can have a lot of data about a client, by utilizing client experience or conduct on their site which may be at times without client assent or information. A web crawler python program was produced for the gathering and investigation of information. The framework is a viable, proficient, basic and quick instrument for web treat crawler. The framework will assist with forestalling loss of information and data on the web space and make more consciousness of how outsider functions and how clients can be checked and conveyed a custom fitted commercial. The fundamental strategic and estimation constraint is that the created treat crawler wouldn't collaborate with destinations in manners a genuine client may, and signing into sites doesn't do activities, for example, looking over or clicking joins, and the framework runs just the python Inactive. Web creep of this work showed that few treats had copied values, which had a similar host, name, esteem, and the time stamp. The web treat crawler framework was made to simplify it for protection assessment, outsider treat investigation and assent input. This requires some investment, work, and paper to physically do. Also, you are allowed to give your input with next to no hesitance.

REFERENCES

[1]. Samarasinghe N. & Mannan M. Towards a global perspective on web tracking, Concordia Institute for Information Systems Engineering Concordia University, Montreal, Canada. **Computer & Security** 2019.

[2]. Lutkevich B., Technical Writer, TechTarget, University of Massachusetts Amherst, 2020 https://www.techtarget.com/whatis/definition/third-party-cookie. third-party cookie.
[3]. Sivan-Sevilla I., Parham P., Toward (greater) Consumer Surveillance in a 'cookie-less' World: A Comparative Analysis of Current and Future Web Tracking Mechanisms, **SocArXiv**. 2022. doi:10.31235/osf.io/rauwj.
[4]. Melicher W., Sharif M., Tan J., Bauer L., Christodorescu M., & Leon P., "Do Not) Track Me Sometimes: Users' Contextual Preferences for Web Tracking", In: Proceedings on Privacy Enhancing Technologies; (2): 2016 135–154
[5]. Kim I., Wang W., Kwon Y., Zheng Y., Aafer Y., Meng W., & Zhang X.,"Adbudgetkiller: Online advertising budget draining attack". In:Proceedings of the 2018 World Wide Web Conference. International World Wide Web Conferences Steering Committee, 2018. 297–307
[6]. Dao H. & Fukuda K., "Alternative to third-party cookies: investigating persistent PII leakage-based web tracking, CoNEXT '21"In: Proceedings of the 17th International Conference on emerging Networking Experiments and Technologies, December 2021, pp. 223–229 https://doi.org/10.1145/3485983.3494860
[7]. Chen Q., Panagiotis Ilia, Polychronakis M. & Kapravelos A., "Cookie Swap Party: Abusing First-Party Cookies for Web Tracking, WWW '21" In: Proceedings of the Web Conference 2021, April 2021, pp. 2117–2129 https://doi.org/10.1145/3442381.3449837
[8]. Degeling M., Utz C., Lentzsch C., Hosseini H., Schaub F. & Holz T., "We Value Your Privacy ... Now Take Some Cookies" In: Measuring the GDPR's Impact on Web Privacy, 2019
[9]. Strycharz J., Smith E., Helberger N. & van Noort G., No to cookies: Empowering impact of technical and legal knowledge on rejecting trackingcookies, 2021
[10]. Aliyeva A. &Egele M., "Oversharing Is Not Caring: How CNAME Cloaking Can Expose Your Session Cookies, ASIA CCS '21" In: Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security May 2021 pp. 123–134 https://doi.org/10.1145/3433210.3437524
[11]. Sanchez-Rola I., Balzarotti D. &Santos I., Cookies from the Past: Timing Server-side Request Processing Code for History Sniffing, **Digital Threats: Research and Practice** Volume 1, Issue 4, December 2020 Article No.: 24 pp. 1–24 ttps://doi.org/10.1145/3419473
[12]. Acar G., Eubank C., Englehardt S., Juarez M., Narayanan A. &Diaz C., "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild, CCS '14", In:Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security November 2014 pp. 674–689 https://doi.org/10.1145/2660267.2660347
[13]. MillettL. I., Friedman B. &FeltenE., "Cookies and Web browser design: toward realizing informed consent online, CHI '01", in:Proceedings of the SIGCHI Conference on Human Factors in Computing Systems March, 2001 pp. 46–52 https://doi.org/10.1145/365024.365034
[14]. RoesnerF., Kohno T. &WetherallD., University of Washington; Detecting and Defending Against Third-Party Tracking on the Web, 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 2012), 2012
[15]. HuX., SastryN., MondalM., CCCC: Corralling Cookies into Categories with Cookie Monster, WebSci '21: 13th ACM Web Science Conference 2021, June 2021 Pages 234– 242 https://doi.org/10.1145/3447535.3462509
[16]. PengW., CisnaJ., HTTP cookies – a promising technology, April 1, 2000 J. R. Mayer & J. C. Mitchell, Third-Party Web Tracking: Policy and Technology, IEEE Symposium on Security and Privacy, University of Stanford, Stanford, CA, USA. 2012
[17]. GomerR., RodriguesE. M., Milic-Frayling N. &SchraefelM. C., Network Analysis of Third-Party Tracking: User Exposure to Tracking Cookies through Search, 2013
[18]. AyensonM. D., WambachD. J., SoltaniA., Good N. &HoofnagleC. J., Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning, 30 Jul, 2011
[19]. SchmuckerN., Web Tracking SNET2 Seminar Paper - Summer Term 2011 S. Englehardt & A. Narayanan, Online Tracking: A 1-million-site Measurement

and Analysis, CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security https://doi.org/10.1145/2976749.2978313 October 2016, pp. 1388–1401

[20]. McCarthy L. &YatesD., The use of cookies in Federal agency web sites: Privacy and record keeping issues, July 2010

[21]. FouadI., SantosC., Legout A. & Nataliia Bielova, Did I delete my cookies? Cookies respawning with browser fingerprinting. arXiv:2105.04381 [cs]. 2021 [online] doi: https://arxiv.org/abs/2105.04381.

[22]. GonzalezR., JiangL., AhmedM., MarcielM., CuevasR., Metwalley H. &NiccoliniS., The cookie recipe: Untangling the use of cookies in the wild, 2017 Network Traffic Measurement and Analysis doi:https://doi.org/10.23919/tma.2017.8002896.

[23]. Conference (TMA), 2017 N. Bielova, A. Legout & N. Sarafijanovic-Djukic, 2020. Missed by filter lists: Detecting unknown third-party trackers with invisible pixels, in: Proceedings on Privacy Enhancing Technologies, 2(2020), 2020 pp. 499–518

[24]. BindraC., Building a privacy-first Future for Web Advertising. [online] Google., 2021 doi:https://blog.google/products/ads-commerce/2021-01-privacy-sandbox/.

[25]. AwaleN., PandeyM., DulalA., &TimsinaB., Plagiarism Detection in Programming Assignments Using Machine Learning. in: J Artif Intell Capsule Netw, 2(3), 2020, pp. 177-184.

[26]. AgarwalP., JoglekarS., PapadopoulosP., SastryN., &KourtellisN., Stop tracking me bro! differential tracking of user demographics on hyper-partisan websites. In Proceedings of The Web Conference 2020. pp. 1479–1490.

[27]. CozzaF., GuarinoA., IserniaF., MalandrinoD., RapuanoA., SchiavoneR., &ZaccagninoR., Hybrid and lightweight detection of third party tracking: Design, implementation, and evaluation. Computer Networks, 167(2020), 2020, pp. 106993.

[28]. Estrada-JiménezJ., Rodríguez-HoyosA., Parra-ArnauJ., & FornéJ., Measuring Online Tracking and Privacy Risks on Ecuadorian Websites. In 2019 IEEE Fourth Ecuador Technical Chapters Meeting (ETCM). IEEE, 2019, pp. 1–6

[29]. HuX., Suarez de TangilG., &SastryN., Multi-country Study of Third Party Trackers from Real Browser Histories, 2020 IEEE European Symposium on Security and Privacy (EuroS&P). IEEE, 2020, pp. 70–86.