Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.778123

PRIVACY GUARD: EFFICIENT PRIVACY PRESERVATION METHOD FOR PRIVACY PROTECTION OF WEB USERS

Ram Mohan Rao P[1,4,*], AP Siva Kumar[2], Murali Krishna[3,4]
[1]Department of CSE, JNTU Anantapur, Andhra Pradesh India
[2]Professor, Department of CSE, JNTU Anantapur, Andhra Pradesh India
[3] Professor, Department of CSE, SV College of Engineering and Technology, Tirupati, Andhra Pradesh India
[4]Department of CSE, Sphoorthy Engineering College Hyderabad, Telangana India

Abstract:

Extensive usage of web applications and online services like ecommerce, finance, insurance, social media etc. has become part of our day-to-day life. By virtue of which, lot of personal, sensitive and confidential data is being uploaded to public domains. Privacy of the users is more compromised due to the tracking of web activity by the browsers and the application vendors. User's online activity is being tracked by both the browsers and application vendors using cookies, tracking scripts and third party sharing. Hence the privacy of the user is at stake while using web applications. The major privacy threats include auto profiling, disclosure, discrimination, surveillance etc. In the recent past, there is an increase in user awareness with respect to privacy. Few privacy protection tools were developed in the recent past. In this paper we proposed a novel method for web data protection called privacy guard with improved privacy and compared the same with existing privacy protection tools. We have experimented with 50 plus websites in three different browsers viz. chrome, firefox and edge. Our privacy guard exhibited better performance by blocking all types of cookies, preventing functional java script from execution and detecting third party data sharing programs with reasonable load time to render the page. Our privacy guard can also be used as a browser extension. This paper gives a detailed description of how privacy guard was developed, the privacy threats addressed by privacy guard and its comparison with existing privacy protection tools like privacy badger, NoScript, Ad block plus and Ghostery.

Keywords: Web Privacy, Third party sharing, Tracking scripts, Web application, Privacy guard

1. INTRODUCTION

Privacy in web has become a major concern with more than 79% of websites tracking the users web activity and user data. As part of web tracking, the websites gather considerable amount of data which includes location, IP address, browsing activity, login, gender etc. This information is shared with third parties and also used to build the user profile. The profile of a web user may contain information related to employment status, financial and medical condition, user preferences etc. The profile will be used to offer value added services to the user. However, profiling without user consent is a privacy breach. Tracking is done by web sites by using 1) HTTP request headers, 2) cookies, and 3) functional java scripts. Some of the HTTP headers used for tacking are listed in Table 1. [1]

Table 1 HTTP Headers that can be useful in web tracking

Sno	Name of the header	Purpose
1	Date	The date and time at which the message was originated.
2	Forwarded	Disclose original information of a client connecting to a web server
3	Prefer	Permit client to request certain behaviors be employed by a server while processing a request.
4	Origin	request for cross-origin resource sharing
5	Tk	Tracking status header

The information collected through the Http headers is called as browser finger printing.

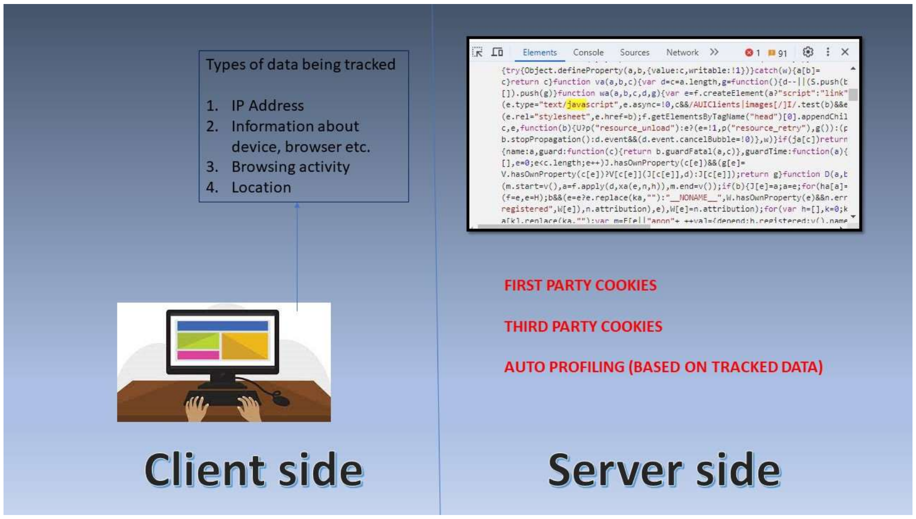
Cookies: cookie is a piece of information stored by web servers in web browsers. cookies are used by web servers to study the user activity on their web site, manage the user session and gather data for third party vendors.

First party cookies: cookies used by the web server to monitor and manage users, are called first party cookies.

Third party cookies: cookies generated and placed by a different web server other than the one user is browsing.

Java script tracking codes: other than HTTP headers and cookies, special java script code can be embedded into the HTML page to track user behavior. 80% of the web traffic is third party cookies.[2][3]. A typical web tracking activity is described in Figure 1.

Figure 1 Typical Web Tracking elements



Types of data being tracked

1. IP Address
2. Information about device, browser etc.
3. Browsing activity
4. Location

```
{try{Object.defineProperty(a,b,{value:c,writable:!1})}catch(w){a[b]=
c}return c}function va(a,b,c){var d=c=a.length,g=function(){d--||(S.push(b
[]).push(g)}function wa(a,b,c,d,g){var e=f.createElement(a?"script":"link"
(e.type="text/javascript",e.async=!0,c&&/AUIClients|images[/]I/.test(b)&&e
(e.rel="stylesheet",e.href=b);f.getElementsByTagName("head")[0].appendChil
c,e,function(b){U?p("resource_unload"):e?(e=!1,p("resource_retry"),g()):(p
b.stopPropagation():d.event&&(d.event.cancelBubble=!0)},w)}if(ja[c])return
{name:a,guard:function(c){return b.guardFatal(a,c)},guardTime:function(a){
[],e=0;e<c.length;e++)J.hasOwnProperty(c[e])&&(g[e]=
V.hasOwnProperty(c[e])?V[c[e]](J[c[e]],d):J[c[e]]);return g}function D(a,b
(m.start=v(),a=f.apply(d,xa(e,n,h)),m.end=v());if(b){J[e]=a;a=e;for(ha[a]=
(f=e,e=H);b&&(e=e?e.replace(ka,""):"__NONAME__",W.hasOwnProperty(e)&&n.err
registered",W[e]),n.attribution),e),W[e]=n.attribution);for(var h=[],k=0;k
a[k].replace(ka,""):var m=F[e||"anon"+ ++va]={depend:h.registered:v().name
```

FIRST PARTY COOKIES

THIRD PARTY COOKIES

AUTO PROFILING (BASED ON TRACKED DATA)

Client side

Server side

Some of the applications of web tracking include recommendation systems, targeted advertising, and auto profiling. The recommendation systems are widely used by ecommerce sites. Most of the blogs, vlogs use targeted advertising and auto profiling is done, to offer value added services to the users. However, all these applications may lead to privacy threats like disclosure of personal data, surveillance, discrimination, personal embarrassment etc. The online activity is increasing exponentially and so are the privacy concerns. There is an increase in the privacy awareness of the users and in order to protect and safe guard user's privacy in the web, many countries have legislated privacy bills with GDPR (General Data Protection Regulation) of European Union being the pioneer. Many privacy protection tools were developed in the recent past which include Ghostery, Ad block plus, No Script, and Privacy Badger. Detailed description of these privacy protection tools is provided in the next section with comparative study followed by our contribution which is a more effective privacy protection tool called Privacy Guard.

2. Related Work

There is a significant improvement in the privacy awareness among the web users with 91% of the users are reluctant to advertisements in the web, treating them to intrusive and harmful [4]. This has resulted in the development of many privacy preservation tools. The most prominent and popular privacy preservation tools are Privacy badger, NoScript, Ad block plus and Ghostery. we have studied these four privacy preservation tools by deploying them in three different browsers viz. Chrome, Firefox and Microsoft Edge.[5]

Experimental Setup:

1.Each privacy preservation tool viz. Privacy Badger, NoScript, Ad block plus and Ghostery are deployed in three browsers Chrome, Firefox and Microsoft Edge.

2. Using web crawling and selenium web driver we accessed top 50 websites in the year 2023[6]. List of

websites crawled is described in Table 2.

3. We have tested each privacy tool with respect to blocking of cookies, tracking scripts, third party data cookie

and other functional java script.

4. There are limitations in every privacy preserving tool with respect to blocking of all threats and performance

i.e., load time of web pages.

5. We have developed a more efficient browser extension called Privacy Guard which offers enhanced privacy

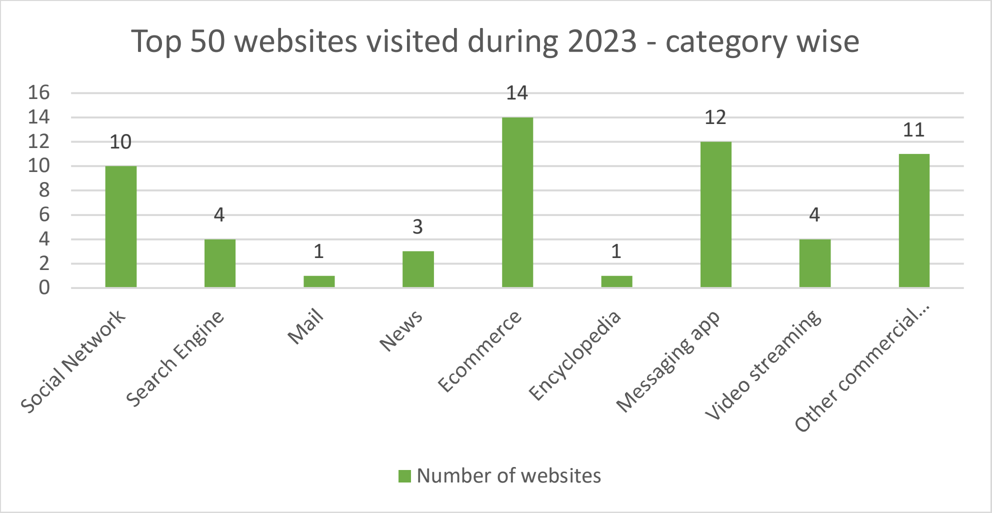
protection than the existing privacy preserving tools.

Table 2. List of top 50 websites (based on visits) crawled using Python web crawler script and web driver of selenium is used to extract information related to web pages.

Sno	Website	Category	Sno	Website	Category
1	Facebook	Social network	26	Spotify	Music site
2	Google	Search engine	27	Telegram	Social network
3	Twitter	Social network	28	Aliexpress	Ecommerce
4	LinkedIn	Social network	29	Flipkart	Ecommerce
5	Reddit	Social network	30	Myntra	Ecommerce
6	Yahoo	Search engine	31	OTTO	Ecommerce
7	GMAIL	Mail service	32	Zalanda	Shoe company
8	Amazon	Ecommerce	33	Naver Korea	Search engine
9	CNN	News	34	Whatsapp	Messaging app
10	Instagram	Social network	35	Xvideos	Video sharing
11	Wikipedia	Encyclopedia	36	Yandex	Search engine
12	Youtube	Video sharing	37	Pornhub	Porn site
13	MSN	News	38	Microsoft Bing	Search engine
14	NY times	News	39	Twitch	Live streaming
15	Pinterest	Social network	40	VK	Social network
16	Baidu	MNC	41	Livedoor	Interset service provider
17	Bing	Search engine	42	B&m Auto pecas	Brazil auto giant
18	eBay	Ecommerce	43	Amazon-India jobs	Ecommerce

19	Microsoft	MNC	44	Amaznon Germany	Ecommerce
20	Netflix	Live streaming	45	Amazon Canada	Ecommerce
21	Tik Tok	Video sharing	46	Amazon UK	Ecommerce
22	Mail.ru	Social network	47	Amazon Mexico	Ecommerce
23	QQ	Messaging service	48	Amazon Italy	Ecommerce
24	Weibo Corporation	Social network	49	Amazon Mexico	Ecommerce
25	OK.RU	Social network	50	Amazon Spain	Ecommerce

Figure 2 Types of websites based on category (top 50 websites visited during 2023)



Privacy Badger: It uses an internal blacklist method to block various types of tracking codes including finger printing and cookies. The privacy badger maintains and internal list of websites that will be blocked. The list will be continuously updated with new websites which are found to be violating privacy laws. It works by sending the Do Not Track header with every page request, and assessing whether the user is still being tracked. It cannot detect any malware and it is not open source. We have deployed privacy badger in all three browsers viz. chrome, firefox and edger respectively and found that privacy badger was able to block cookies of ecommerce websites like amazon etc. but failed to block the cookies of Google Search engine. [7]

NoScript: No Script is based on the white listing technique where user has to authorize to allow any executable content in the website and hence the default behavior is to block all website contents. A white list-based tool allows content that has been explicitly authorized by the user, so the default behavior is to block all website content. However, this may result in usability problems and requires considerable user interaction.[8]

Adblock plus: it is primarily developed to prevent unwanted ads to be displayed during web access. It basically uses a black listing technique with large number of customizable rules. Adblock plus is also good at preventing auto download of any web resources from the black listed advertising agencies and trackers. [9]

Ghostery:

Ghostery can detect and block tracking scripts and cookies from webpages in order to improve privacy and focus on only important content. It also scans the DOM tree for Advertisements, tracking, and other entities stored in a predefined blacklist. It also provides a facility where user can activate or deactivate the privacy protection.[10]

- Privacy: Privacy Badger, NoScript, and Ghostery are designed with a primary focus on user privacy, aiming to prevent tracking and data collection.
- Security: NoScript enhances security by blocking potentially harmful scripts, reducing the risk of malicious activity.
- Ad Blocking: Adblock Plus is specifically tailored for blocking ads, while the others also address tracking and privacy concerns.
- User Control: NoScript provides the most granular control over scripts, but it requires more user interaction.

The effectiveness of these tools may vary based on user preferences, the specific threats they want to mitigate, and the balance between privacy and convenience. Users often choose a combination of these tools to create a comprehensive privacy and security solution. We have developed a more effective, light weight and easy to deploy browser extension called Privacy Guard which combines all the benefits of the various privacy protection tools discussed above.

3. PRIVACY GUARD: EFFECTIVE AND EFFICIENT PRIVACY PROTECTION TOOL

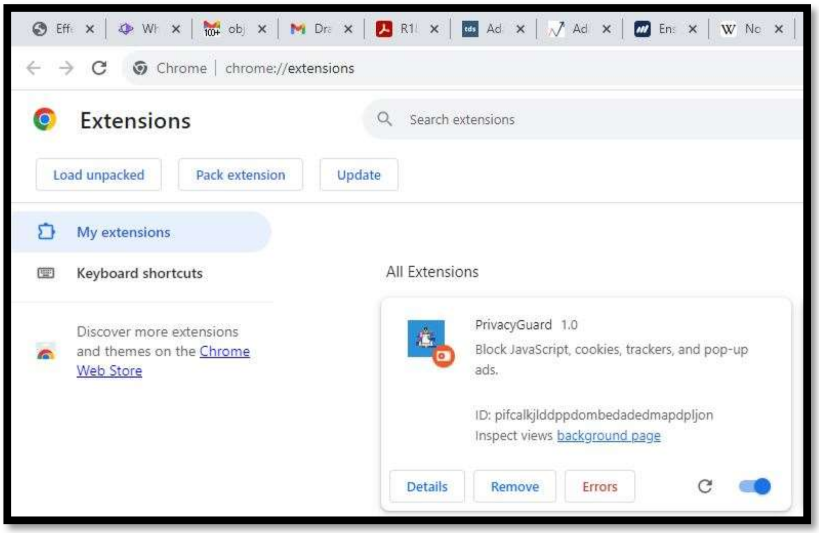
Privacy guard is a novel privacy protection tool which combines features of all existing privacy protection tools into one and offers better performance with respect to load time. Privacy guard was developed using java script and CSS. It is a heuristic approach where the DOM tree of every website is read, identify the functional java script code present if any and block any executable code representing trackers, advertisements etc. our privacy guard also blocks cookies which are used by search engines, ecommerce sites and social media platforms which are the potential sources of privacy breach. The list of trackers which are blocked by our privacy guard are shown in the table 3.

Table 3 Trackers found and blocked by Privacy Guard

Sno	Name of the tracker	Purpose	Used by
1	Google Analytics	Traffic analysis	All websites powered by Google
2	Facebook pixel	To study how Facebook users respond to the advertisements shown to them	All websites that use Facebook for advertisements.
3	Hotjar	Session recorders	Various websites for improving user experience
4	Double click	Advertisements	Google to serve advertisements
5	Crazy Egg	Session tracking	Dell, Yahoo, Twilio, Zendesk etc.
6	KISSmetrics	Study customer behavior and retention to improve marketing strategies	MICROSOFT, Dassault systems, Carvan, Lucid etc.
7	YouTube API	To gain insights on user engagement, demographics, user preference and recommendations	YouTube LLC.

Privacy Guard was designed to block all the above mentioned web trackers and was tested on all the mentioned web sites. Privacy guard was found successful in blocking the trackers, cookies and any other functional java scripts when tested on the top 50 websites listed in Table 2. Privacy guard was deployed as a Chrome browser extension and alert statements are used to show the various types of cookies blocked by the Privacy Guard. Figure 3 and 4 describes the Privacy Guard blocking the cookies used by www.google.com

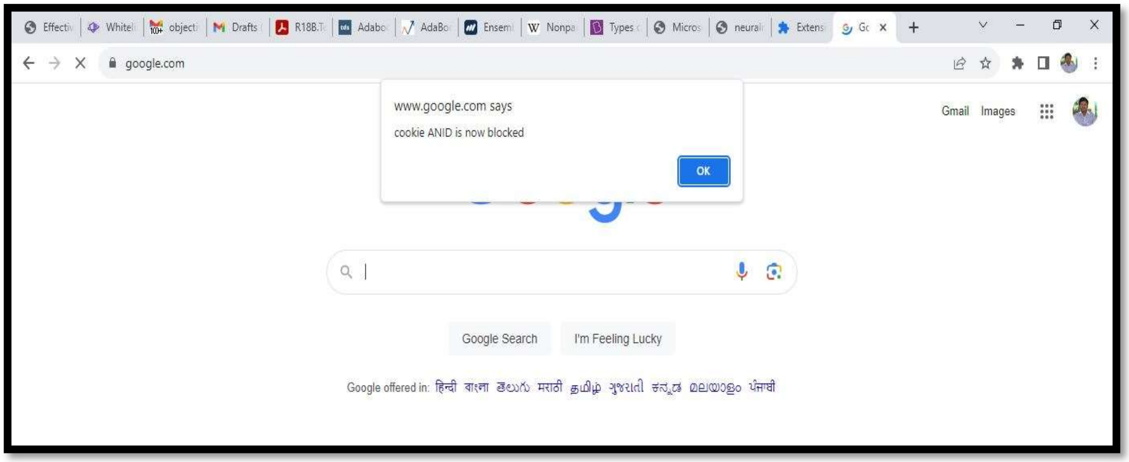
Figure 3 Privacy Guard added to chrome extension



PrivacyGuard 1.0
Block JavaScript, cookies, trackers, and pop-up ads.

ID: pifcalkjlddppdombedadedmapdpljon
Inspect views background page

Figure 4 ANID cookie blocked by Privacy Guard


google.com
Gmail
Images
Google Search
I'm Feeling Lucky
Google offered in: हिन्दी বাংলা తెలుగు मराठी தமிழ் ગુજરાતી ಕನ್ನಡ മലയാളം ਪੰਜਾਬੀ
OK

www.google.com says
cookie ANID is now blocked

Using Privacy Guard, we were successful in blocking cookies, web trackers and advertisements. We have also measured the load time of the web page before and after deployment of the Privacy Guard extension in Google Chrome, Firefox and Edge. The load time without extension was 0.20 seconds for www.google.com and 0.21 seconds after deploying the Privacy Guard extension. Hence the differences in the load time of web pages are also found to be negligible. The table below shows the load times for few popular websites.

Table 4. Load times of 5 websites with and without Privacy Guard measured in seconds

		GOOGLE CHROME BROWSER	MICROSOFT EDGE	FIRE FOX

Sno	Website visited	Before deploying Privacy Guard in seconds	After Deploying Privacy Guard in seconds	Before deploying Privacy Guard in seconds	After Deploying Privacy Guard in seconds	Before deploying Privacy Guard in seconds	After Deploying Privacy Guard in seconds
1	www.google.com	0.20	0.21	0.24	0.24	0.25	0.26
2	www.youtube.com	0.63	0.97	0.76	0.98	0.82	0.98
3	www.myntra.com	0.22	0.22	0.23	0.23	0.25	0.25
4	www.netflix.in	0.58	0.73	0.64	0.65	0.77	0.79
5	www.linkedin.com	0.35	0.36	0.54	0.54	0.57	0.61

It has been observed that the differences in the load times of 5 websites before and after deployment of Privacy Guard is negligible, with Google Chrome offering the better performance than Firefox and Microsoft Edge with respect to load time of the websites. Hence Privacy Guard is proved to be a more effective and efficient solution for blocking harmful java scripts, cookies, trackers and unwanted advertisements.

4. RESULTS AND DISCUSSION

Privacy of a user is a major concern in web because almost all web sites track the user behavior, retrieve personal data from the browser, share the data with others which may to lead to disclosure of personal, sensitive and private information of the users. It is very important to safe guard user's privacy inline with privacy legislations like GDPR. In this regard we developed a novel privacy preservation tool called Privacy Guard which offers a better privacy when compared to existing privacy protection tools like Ghostery, Privacy Badger, NoScript and Adblock plus. Privacy Guard can be used as a browser extension and it is all in one protection from cookies, functional java script, trackers and unwanted advertisements. Table 5 provides a comparative study and analysis of existing privacy protection tools and Privacy Guard. Privacy Guard was proved to be more effective and efficient with respect to privacy protection and load time of websites.

Table 5. Comparison of Privacy Guard with other existing privacy tools

Features / Methods	Cookies	Java script	Third party tracking and ad blocking	User base	Load time (tested on 50 websites) in seconds

PRIVACY GUARD	**Blocks**	**Blocks functional script**	**Block tracking**	**NIL**	**01.357289**
Ghostery	Blocks	Does not detect	Block third party tracking	2000000+	01.789
NoScript	Blocks	Blocks	Blocks third party tracking	600000+	01.45
Ad block plus	Does not block	Block	Prevents malware and tracking	1000000+	01.25
Privacy Badger	Does not block	Does not block	Prevents tracking by using DO NOT TRACK header in every web page.	1000000+	01.11

5. CONCLUSION

In the recent times, there is an increase in the awareness of the users with respect to privacy threats involved in web and the growing concern has led to creation of privacy legislations in many countries with GDPR being the pioneer. The existing privacy preserving tools and our contribution which is more enhanced, improved privacy protection tool (Privacy Guard) will be able to offer privacy protection with respect to cookies, trackers and advertisements but they do not comply to all the laws of the GDPR. There is a need to implement all privacy laws by ensuring the websites to comply all the GDPR laws, create awareness among the users and design privacy protection tools using Artificial Intelligence, such that they can learn from the experience and evolve on their own. Hence there is a huge scope in the field of data privacy with respect to web user activity.

DECLARATIONS

Conflict of Interest: The authors declare that they have no conflict of interests.

Competing Interests: The authors have no competing interests

Funding Information: No funding provided

Author contribution: This work is done as part of academic research by the corresponding author under the guidance of the second and third author.

Data Availability Statement: will be provided up on request

Informed Consent: Not Applicable

REFERENCES:

1. Tirandazi, Peyman, Seyed Mojtaba Hosseini Bamakan, and Aref Toghroljerdi. "A review of studies on
internet of everything as an enabler of neuromarketing methods and techniques." *The Journal of*
Supercomputing 79.7 (2023): 7835-7876.

2. Bubukayr, M.; Frikha, M. Effective Techniques for Protecting the Privacy of Web Users. *Appl. Sci.* **2023**, *13*,
3191. https://doi.org/10.3390/app13053191

3. Gotze, Matthias, et al. "Measuring web cookies in governmental websites." *Proceedings of the 14th ACM*
Web Science Conference 2022. 2022.

4. Urman, Aleksandra, and Mykola Makhortykh. "You are how (and where) you search? Comparative analysis of web search behavior using web tracking data." *Journal of Computational Social Science* (2023): 1-16.

5.Peukert, Christian, et al. "Regulatory spillovers and data governance: Evidence from the GDPR." *Marketing Science* 41.4 (2022): 746-768.

6.Gudavalli, Aneesha, and G. JayaLakshmi. "Implementation of Test Automation with Selenium Webdriver." *Grenze International Journal of Engineering & Technology (GIJET)* 8.1 (2022).

7.Smith, Karen Louise, and Elysia Guzik. "Developing Privacy Extensions: Is it Advocacy through the Web Browser?." *Surveillance & Society* 20.1 (2022): 64-81.

8.Schöni, Lorin, Karel Kubicek, and Verena Zimmermann. "Block Cookies, Not Websites: Analysing Mental Models and Usability of the Privacy-Preserving Browser Extension CookieBlock." *Proceedings on Privacy Enhancing Technologies* 2024.1 (2023): 192-216.

9.Bubukayr, Maryam Abdulaziz Saad, and Mounir Frikha. "Web Tracking Domain and Possible Privacy Defending Tools: A Literature Review." *Journal of Cybersecurity* 4.2 (2022): 79.

10. Dettman, David. "Third-Party Tracking in Online Public Library Environments in the United States and Canada: A Statistical Analysis." *Evidence Based Library and Information Practice* 18.2 (2023): 120-122.