Third-Party Data Leaks on the 500 Most Popular Websites

Robin Carlsson
University of Turku
Turku, Finland
crcarl@utu.fi

Henna Lohi
University of Turku
Turku, Finland
hmlohi@utu.fi

Sammani Rajapaksha
University of Turku
Turku, Finland
syraja@utu.fi

Panu Puhtila
University of Turku
Turku, Finland
papuht@utu.fi

Timi Heino
University of Turku
Turku, Finland
tdhein@utu.fi

Sampsa Rauti
University of Turku
Turku, Finland
sjprau@utu.fi

Abstract

With the advent of digitalization, the use of online services for everyday tasks has greatly increased. In this study, we examine privacy of the 500 most popular websites in the world to determine to what extent these websites leak their visitors' personal data to third parties. We analyze the network traffic of the websites to see whether they leak the URLs visited by the user and search terms. Our findings show that 58.0% of the studied websites leak such personal data to outside actors. These results demonstrate that the user's privacy is in danger on a large portion of the studied websites. The sheer number of found third-party services on many websites also introduces several challenges concerning sustainability and fair data processing practices.

CCS Concepts

• **Security and privacy → Web application security**.

Keywords

Websites, online privacy, data leaks, third-party services

ACM Reference Format:
Robin Carlsson, Henna Lohi, Sammani Rajapaksha, Panu Puhtila, Timi Heino, and Sampsa Rauti. 2024. Third-Party Data Leaks on the 500 Most Popular Websites. In *2024 9th International Conference on Information Systems Engineering (ICISE 2024), December 14–16, 2024, Chiang Mai, Thailand.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3711954.3711960

1 Introduction

In present time, daily tasks are carried out more and more on the web. Shopping, banking, communication and enjoying entertainment often take place on web platforms. Modern websites also regularly use third-party services, which is often a great threat for visitors' privacy when personal data is collected and sent to remote servers, often without users' consent and knowledge. Third-party web analytics, for example, give website maintainers insights on meeting business goals and achieving good usability on the website. At the same time, however, the large technology giants and other third parties providing these services reap the benefits by collecting users' personal data. Users often do not understand the meaning and nature of this collected personal data [12].

For instance, URLs and search terms often leak to private third-party corporations on modern websites [11]. The URLs and searches may contain sensitive data such as the user's health status, political views or sexual orientation. These kinds of sensitive data items are considered special category data subject to greater protection in the General Data Protection Regulation (GDPR). It is noteworthy that also the companies outside the EU that process the personal data of EU citizens must comply with the GDPR. From this viewpoint, collecting personal data without consent and proper informing of users is a privacy problem.

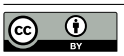
The contributions of this study are as follows. It presents an extensive network traffic analysis of the 500 most popular websites and their third-party services, illustrating the fact that even without the consent, one website can have dozens of third-party services collecting the user's personal data. Additionally, we make an argument about excess third-party services posing a notable challenge to sustainability of software development. We also give some recommendations to mitigate the adverse effects the use of third parties has on privacy and energy consumption.

The rest of the paper is structured as follows. Section 2 reviews the related work on third-party data leaks. Section 3 covers the study setting and methods of our research. Section 4 presents the results of our network traffic analysis. Section 5 discusses the implications of our findings for user privacy and sustainable software development. Finally, Section 6 concludes the paper.

2 Related work

There is a large body of research on third-parties in web-based services, covering many different topic areas. In this survey, we have included papers which are very similar to ours, i.e. studies on the data leaks to third parties, but also research done on other aspects of privacy, such as the use of dark patterns and privacy policies in the most popular websites in EU area, since these studies are both thematically and methodologically related to our own research.

According to the study by Naryshkin et al. [13] both direct and indirect leaks of user privacy occurred on 75% of the 120 websites that were analyzed through material created by third parties. The majority of the leaks were collected by a small group of third parties, according to the researchers, who categorized the leaks into different categories, such as those via cookies or URLs.



ICISE 2024, Chiang Mai, Thailand

ACM ISBN 979-8-4007-1736-9/24/12
https://doi.org/10.1145/3711954.3711960

A paper by Kontaxis et al. [8] addresses the privacy risks associated with social login platforms like Facebook, which are increasingly used by millions of websites. The authors suggest a framework that enables users to authenticate with third-party websites using a reduced Facebook session, disclosing only the essential personal information, in order to reduce these dangers. The system, known as SudoWeb, is a Chrome extension that protects users' privacy by default and allows them to increase their session's permissions when needed.

Research by Libert [11] exemplifies that 91% of health-related websites make third-party requests, with 70% of websites leaking sensitive condition and treatment information, risking privacy.

Degeling et al. [6] assessed the effectiveness of GDPR by investigating whether the top 500 websites in each of the EU countries had adopted the use of privacy policies, or if the policies existing prior to GDPR were updated to conform to it. Their findings indicated that the mandates of GDPR had increased the use of privacy policies and cookie consent banners by roughly 16% each across EU.

The study by Sivan et al. [18] addresses the key problem of location data leakage from Android devices, in which user locations are exposed through Internet traffic without their knowledge. With the help of 71 participants' network traffic over a time period of 37 days in empirical evaluation, the researchers discovered that more than 85% of devices leaked location data. This allowed inference into user POIs with an accuracy of 61%. The findings underline how universal and acute the problem of location data exposure is, raising severe privacy concerns and pointing out the risks of potential misuse for the inference of sensitive information on users.

According to a report by cookie bot [5], despite GDPR obligations ad tech trackers were found on 89 percent of EU government websites, with only Spain, Germany, and the Netherlands avoiding commercial trackers and Finland being the seventh highest.

A study by Zheutlin et al. [21] indicated that several health-related websites leak the user's data to ad trackers and third-party cookies. This happens on most commercial and non-profit websites that are creating full health profiles of the user, which are required for targeted ads. This practice is quite common, but, at the same time, most users are unaware of how their data is being used, or even the extent of its specificity.

Krisam et al. [9] studied the use of dark patterns in cookie consent banners of top 500 German websites. They concluded that 85% of the websites used dark patterns to manipulate the user to accept all cookies, and only 21.5% offered the option of rejecting all cookies with one click.

The study by Yu et al. [20] indicates that more than half of the 19,483 global hospital websites analyzed leverage tracking scripts and some even replay sensitive information back to external servers. Meanwhile, even though some hospital privacy policies asserted that no data sharing with third parties would occur, some violations have been found.

In another study by Zheutlin et al. [22] more than two-thirds of 85 accredited digital pharmacies they inspected used multiple data-tracking methods, including ad trackers, third-party cookies, and sharing data with Facebook and Google Analytics. Despite the growing use of digital pharmacies, especially during the COVID-19 pandemic, there is little regulation or transparency regarding how personal health data is handled.

A study by Rauti et al. [15] reveals significant data privacy issues in Finnish online pharmacies, with prescription information linked to personal data being leaked in 57 cases out of 163. Most of these pharmacies (145) have integrated third-party services. The results show a greater need for control and data protection by design in online pharmacies.

Similarly, the study by Samarasinghe et al. [16] reveals that there is a significant number of worldwide government websites and Android apps that are harbouring commercial trackers, with 17% of websites and 37% of apps hosting Google trackers. Additionally, 27% of government apps leak sensitive user information whereas some government sites and apps are flagged as malicious.

Samarasinghe et al. [17] found in a separate study that 32 percent of religious websites and 78 percent of religious Android applications have Google trackers, with some sending sensitive information in clear text. Moreover, many religious sites and applications have security vulnerabilities, with some applications flagged by VirusTotal as malicious.

Compared to previous studies, our study presents an extensive study on data leaks on most popular websites in the world. The current study also specifically explores the setting in which no consent for data collection is given by the user. Our study is conducted from the viewpoint of European privacy regulation. Unlike most of the previous work, we concentrate on URL and search term leaks, which are popular sensitive data. Moreover, we discuss the issue of excessive third parties from a sustainability point of view, which has largely been overlooked in previous studies.

3 Study setting and methodology

We used the "Moz Top 500 Websites" list[1] of the most popular 500 websites in the world as a dataset for this study. The list is based on a link-based search engine ranking metric called "Domain Authority"[2]. In order to analyze the 500 websites on the list, we used an automatic tool we have specifically built to browse websites and analyze their network traffic to third parties. The tool is written in Python and uses Selenium WebDriver library for browser automation. It detects URL addresses and search terms that leak to external actors in a setting where the user has not consented to data collection. We have presented the technical implementation of our tool in more detail in [4].

The network traffic analysis was performed manually for a small number of websites on which the tool failed[3]. For the manual analysis, we used Google Chrome's Developer Tools. In both automatic and manual analysis, the recorded network traffic was analyzed, and the detected third parties and the sensitive personal data items (mainly the URLs and search terms) were listed.

We specifically focused on 3 different types of leakages:

- *Website URL:* This leakage type was chosen because it directly reveals that the specific website has been accessed. While it alone does not contribute to identification of the website user, combined with device information it can be used for

[1] https://moz.com/top500

[2] https://moz.com/learn/seo/domain-authority

[3] Because our tool is a software robot that aims to mimic user behavior and browses the websites, the failures were mostly cases in which the analyzed websites had pop-up elements such as cookie consent banners positioned on top of the search functionality, which prevented our tool from using the search

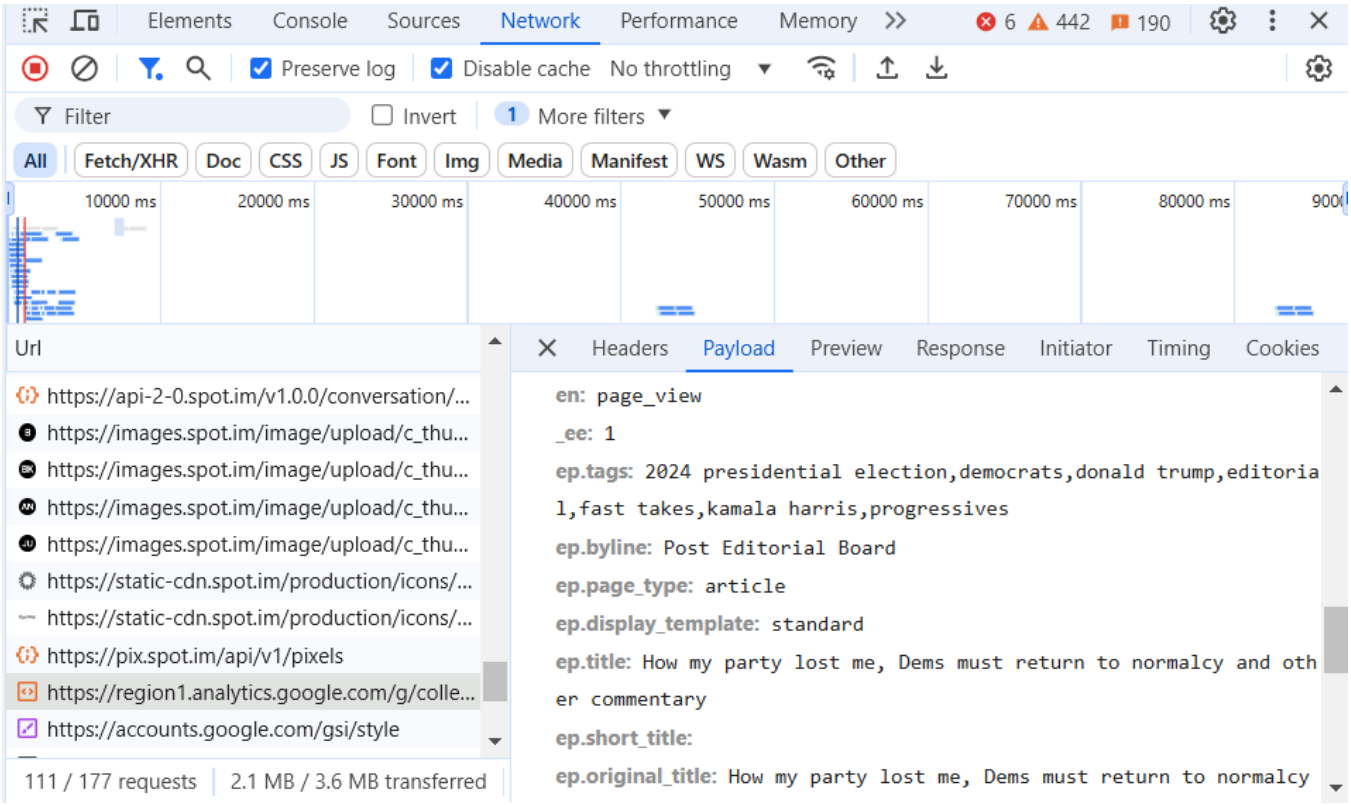


Figure 1: Contextual data about reading a political article leaking to Google.

constructing a digital fingerprint of the user in question, with several records of this type indicating a pattern of what websites the specific user is interested in.

- *Search term:* Leaks of the used search term can be used to gauge interested in to specific topic. Depending on the theme and purpose of the website, these kinds of leaks can reveal potentially sensitive information about the website user.
- *Device information:* Device information, such as device IDs and IP addresses, can be used in identifying the specific person using the website. Combining data items such as operating system and browser information, screen size etc. can also help in uniquely identifying the user. Some web analytics tools, such as Google Analytics, use unique identifiers that can be used to definitely identify specific unique browser-device pairings[4].

Of the studied 500 websites, 26 were not functional in the sense that users could navigate them. This was either due to IP being blocked due to sanctions, for example the website of Russia Today, `rt.com`, or the fact that the domain was not an actual website but used to host and serve content such as `googleusercontent.com`. These nonfunctional websites have been omitted from the calculations in the results section.

We also manually inspected the cookie consent banners of these websites. The focus in this analysis was to determine:

- Is it possible for the user to deny cookies?
- Is the user informed of the data collection?
- In addition to these, we also noted other aspects of interest related to the cookie consent process.

In the current study, the term *personal data* is important, and we define it briefly here. We use the definition laid out by the GDPR and also used by the Finnish Office of the Data Protection Ombudsman. According to this definition, personal data is "all data related to an identified or identifiable person"[5]. Following this definition, technical details such as IP addresses, device identifier, and accurate location data count as personal data. It is also worth noting that while many technical details alone cannot be used to identify someone, identification may become possible when several details such as screen size and operating system version are combined. Therefore, such data items are also included in the definition of personal data.

The real challenge in personal data leaks is the fact that the identifying personal data such as the IP address is combined with potentially sensitive contextual data, such as the URLS of the specific pages the user has visited. For example, the third parties may learn the health related URLs a specific user visits and the search terms they use [11]. All this often happens without the user's knowledge. Figure 1 shows an example of how information about reading

[4]https://www.owox.com/blog/use-cases/google-analytics-client-id/

[5]See https://gdpr.eu/eu-gdpr-personal-data/ and https://tietosuoja.fi/en/what-is-personal-data

a political article leaks to Google without the user's consent. The URL address of the accessed article, along with the title and the tags related to the article, are all sent to Google's servers.

4 Results

Out of the 474 functional websites, 271 (57.2%) leaked the URLs of visited subpages. A search function was available on 333 websites and search terms leaked to at least one third party in 178 cases (53.5%). That is, both the visited URLs and inputted search terms leak more often than not, even without the user's consent or properly informing them. Lastly, 275 websites (58.0%) leaked either visited pages, search terms, or both.

Table 1 shows the websites with most unique third parties. The website of Asahi Shimbun, one of Japan's largest newspapers, comes on top with 43 third parties. New York Post and The Sydney Morning Herald take the second and third places with 30 and 29 third parties, respectively. Generally, it seems many large media websites also include an excessive number of third party services. These numbers are very high when we keep in mind that no consent for data collection has been given.

The average number of third parties on the functional 474 websites was 3.28 while the median was 1.00. On such popular websites, using over 3 third-party services without consent can be considered a high number. When we look at the average numbers of third parties in separate groups of 100 websites (again excluding the non-functional websites), 100 largest websites on top have the average of 0.92 third parties, the second set of 100 websites 2,91, the third 3.29, the fourth 3.83, and finally the fifth 5.48. This constantly increasing number may be partly due to better data protection practices on the very top websites, but another reason is likely the fact that many of the top websites (several Google's domains, Facebook, LinkedIn, Adobe, Amazon, Apple, Microsoft, Tiktok etc.) are notable data collectors themselves and do not include any competing third-party services or analytics.

Table 1: Websites with the highest number of unique third parties. Websites with 20 or more third parties are included.

Websites	Third parties
Asahi.com	43
Nypost.com	30
Smh.com.au	29
Theverge.com	24
Nbcnews.com	24
Gsmarena.com	24
Tabelog.com	22
Thestar.com	21
Nicovideo.jp	20
Politico.com	20
Adweek.com	20

Table 2 shows most frequent third parties receiving personal data. It is not surprising that Google takes the first place. Google Analytics is the most popular web analytics service in the world. It tracks and reports web traffic and user behavior. Along with Google Analytics, the DoubleClick online advertising service owned by Google is also popular. Twitter (X Corp) is the second popular third party found in the current study, in most cases used to track users on various websites. ScorecardResearch is a market research service of Full Circle Studies. It collects data on users' browsing behavior, building profiles and reporting on Internet trends. Other top companies include Amazon (mostly the Amazon Ads service) and Meta (mostly the Meta Pixel tracking service).

Table 2: Most frequent third parties receiving personal data. Each leak type to a specific third party has been counted once per website.

3rd Party	URL leaks	Search term leaks
Google	211	144
Twitter (X Corp)	68	50
Scorecard Research	51	44
Amazon	41	35
Meta (Facebook)	59	34
New Relic	38	28
Chartbeat	30	26
Rubicon Project (Magnite)	28	22
AppNexus	32	21
Pubmatic	19	13

Table 2 also shows the frequency for URL leaks and search terms leaks – both leak types counted once per each website. Google is clearly the most popular third party; 42.2% (211) of the studied websites leak the visited URLs to Google. Respectively, search terms leak to Google on 28.8% (144) of the websites. These are notably high numbers, especially when we take into account the fact that consent for cookies and data collection has not been given.

Table 3 shows the 15 most popular third-party domains that received personal data on analyzed websites. Each domain has been counted only once per website. We can see that among these 15 top domains, there are 7 different domains related to Google's services (Google Analytics and DoubleClick).

Figure 2 roughly shows the initial destinations the leaked personal data is sent to and the number of leaks going to a specific area. The fact this study is carried out in Finland affects the results in the sense that many third-party connections go to Europe. However, it is evident that the majority (57.4%) of personal data seems to end up outside EU and EEA, while the minority (42.6%) stays inside Europe. This may not be surprising for websites hosted outside of Europe, but providing services for EU citizens and collecting their personal data without consent still violates the GDPR. In this sense, the most popular websites in the world do not seem that concerned with GDPR compliance.

As shown in Figure 3, consent practices and informing the users about data collection varied. In total, 268 websites allowed the user to reject cookies and data collection. On the other hand, 46 websites informed the user of cookies or data collection but did not give the possibility to reject them. Curiously, 6 websites allowed the user to decline cookies and data collection only with a paid subscription! 180 websites did not have an option to reject cookies and data collection and also did not inform users about them with

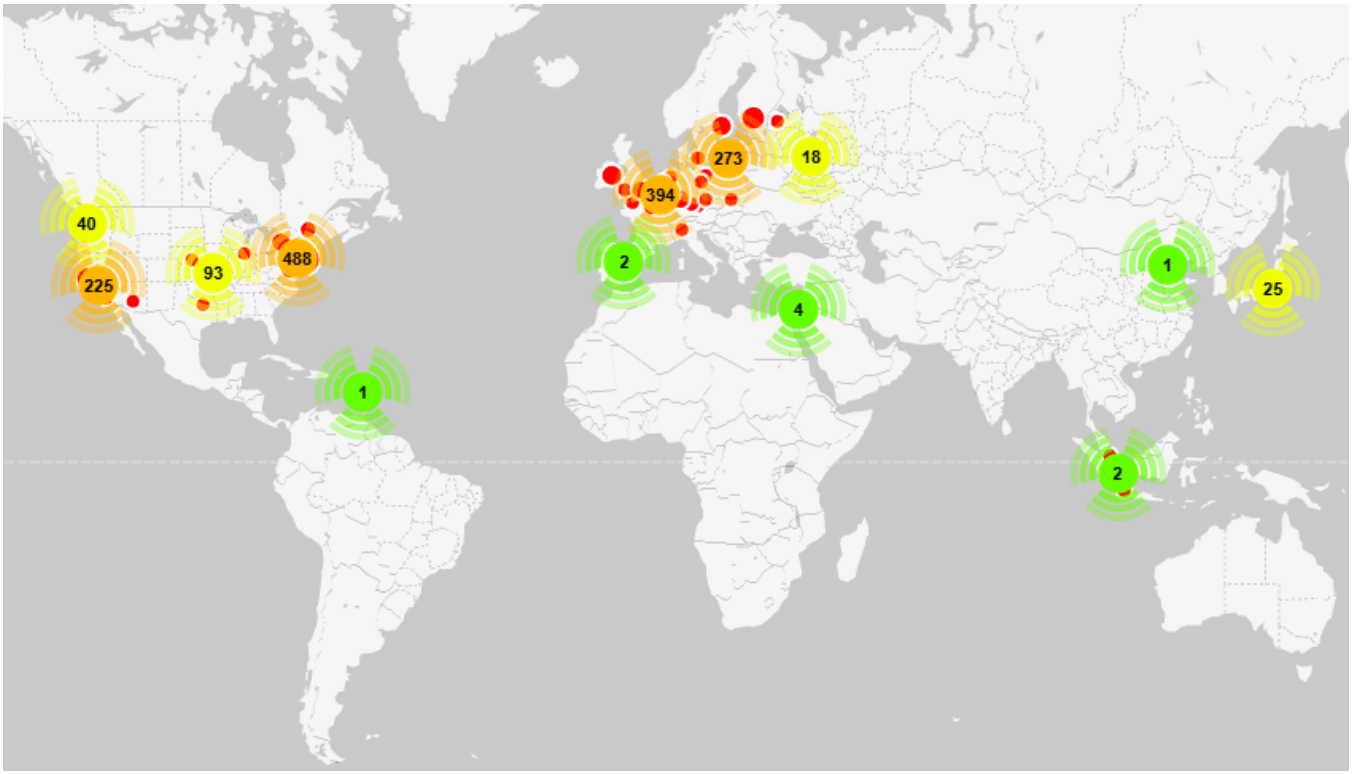


Figure 2: The destinations of the detected personal data leaks.

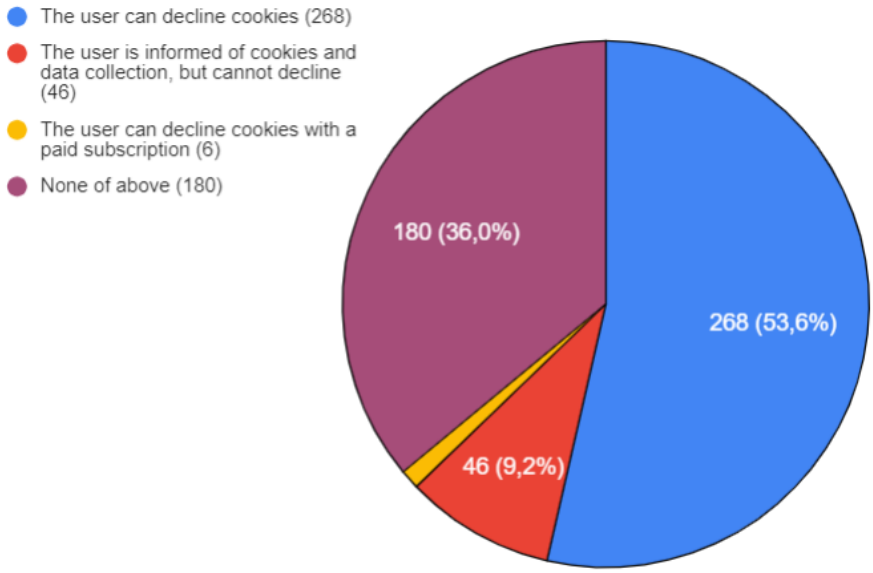


Figure 3: Asking for consent and informing the user on the studied websites.

a popup or other clear notification upon arrival. It is worth noting, however, that out of 500 studied websites, 217 did not have any network traffic to third-party services in our experiments. Many of these websites may not contain any third parties or cookies and therefore do not need to ask consent or inform users. Naturally, privacy regulations and laws are different around the world and informed consent is not required everywhere. However, the GDPR should be followed whenever EU citizens' personal data is being

Table 3: Most frequent domains. Each domain has been counted once per website.

Domain	Occurrences
www.google-analytics.com	129
region1.google-analytics.com	99
www.facebook.com	62
region1.analytics.google.com	53
sb.scorecardresearch.com	52
googleads.g.doubleclick.net	47
www.google.com	36
www.google.fi	35
securepubads.g.doubleclick.net	34
bam.nr-data.net	32
aax.amazon-adsystem.com	31
ping.chartbeat.net	27
fastlane.rubiconproject.com	26
ib.adnxs.com	21
px.ads.linkedin.com	19

processed [2]. It can also be argued that obtaining informed consent would be a recommendable practice universally.

5 Discussion

5.1 Key findings

The key findings of the current study can be summarized in the following points:

- Well over half (57.2%) of the studied pages leaked the URL addressed of the visited pages to third parties without obtaining consent. Depending on the visited pages and the web service in question, this may reveal sensitive details about the user's, especially over a long period of time.
- In a similar vein, over half (53.5%) of the websites with a search function leaked search terms inputted by the user to third parties without the user's consent.
- The numbers of third parties on the analyzed websites are excessive, with the highest number of third parties being 43 and over 9% having more than 10 third parties. Especially media websites were found to regularly contain a large number of third-party services.
- Over half of the found data leaks on the analyzed websites went to third parties located outside the EU or EEA without the user's consent.

5.2 Implications for software developers

Our research reveals significant issues with third-party data leaks and consent practices across the studied websites. Of particular concern is the high number of third parties – even up to 43 per website – in a setting where the user has not given their consent to cookies or data collection. The analyzed websites also often operated without asking for users' consent or properly informing them about the third parties. Even if one argues some third-party tracking on the sites studied websites is justified, there is still a major problem with how many of these services are used. Having an average of 3.28 different third-party services on a single website collecting personal data is too much.

This calls for more attention for the selection of third parties. It is important to carefully consider which third parties are used in a specific web service. The use of each third-party service should be well justified. It is also important to understand what personal data is collected by the services employed on the website and where it is going. Network traffic analysis approaches like the one we have used in this study can be applied to study outgoing data.

It is clear that many of the studied websites are not transparent enough about the data collection and processing activities that take place when a user visits the website. Currently, the users do not have a real possibility to decide what personal data about them is processed and by which parties. This goes against the very definition of privacy as the user's right to decide how their personal data is accessed, used, and disclosed. Even if the studied websites do not necessarily contain highly sensitive personal information, users should have the ability to control how their data is collected. This is why websites should be designed with privacy in mind from the very beginning.

The user´s should be informed about the data collected through consent banners or privacy policies. Additionally, the developers should ensure the accept and reject option in the cookie banners are available and provide the services according to the permissions given. Furthermore, developers can implement encryption protocols (such as HTTPS) and use of privacy-preserving techniques (like hashing, pseudonymization or tokenization) to prevent leaking of search terms and other sensitive data to third parties.

In practice, GDPR compliance often has a large gap between promise and delivery, even in Europe [7]. With the 500 most popular websites, the problem is even more pronounced. As the most companies behind the websites will not have to face serious consequences from violating the GDPR, they are not as likely to follow this legislation as their European counterparts. This undermines the legitimacy of the GDPR's applicability outside the EU. While European legislation may not motivate web developers around the world to fix the privacy issues we have presented in the current study, the goals of transparency, privacy and sustainability definitely should do so.

5.3 Implications for sustainability

Energy consumption and emissions caused by the Internet use have also increased, which presents a real sustainability challenge and calls for the greening of the Internet [3, 19]. Consequently, the number of third parties used on website is also important thing to consider from a sustainability point of view. Websites running excessive numbers of third-party scripts leads to the following sustainability issues:

- *Increased use of computing power*: The CPU usage is increased as results of scripts third-party services run on client side. It has formerly been demonstrated that adding just a couple third-party analytics services may have a significant effect on a website's energy consumption [14]. As this paper clearly demonstrates, often the users are forced to waste computational power and energy by running third-party service scripts without their consent.

- *Higher bandwidth usage*: Employing multiple third-party services also increases HTTP requests and bandwidth consumption, leading to higher energy usage. With each page load, several third-party services trigger HTTP request, and many of them also make request periodically or when the user performs a certain action on a website, further contributing to energy consumption.
- *Notable disk space usage*: Multiple third-party services collect the same or similar data that they may replicate on multiple servers. Disk space is therefore used unnecessarily in several locations, creating data waste. Maintaining this data storage requires energy.

Consuming energy causes carbon emissions and has significant environmental effects [1], at least if the third parties do not make use of green data centers with renewable energy. Minimizing data collection and the number of third-party services is therefore not only a question of privacy, but also an important consideration in green software development.

5.4 Implications for users

Although GDPR as a law is only applicable to EU citizens, international websites also have to take it into account when providing services for Europeans. The privacy of personal data, of course, is even wider, universal concern with several potential implications for users. This study has addressed two possible data leak scenarios in particular. Firstly, the visibility of user preferences (such as URLs of sub pages visited), can expose personal browsing behavior to third parties potentially compromising users' anonymity. Secondly, the leaking of search terms containing sensitive personal data poses significant privacy risks for the user. The search terms in particular can be sensitive, revealing information such as health conditions, religious beliefs, political affiliations, sexual orientations etc. could be visible to third parties without the user's knowledge. In terms of GDPR, this is very problematic because not only are these data categories especially sensitive (special category data), but users are often neither informed nor given opportunity to consent to websites' data-sharing practices as our findings clearly demonstrate. This violates users' right to control the use of their personal data. Additionally, few websites did not offer users the option to reject cookies, unless they subscribed to paid services, which creates an unfair paywall around privacy, which should be a free, fundamental right to.

All these privacy issues can lead to feelings of invasion of privacy and loss of control over one's personal information. Furthermore, even attempts to influence an individual's preferences and emotions could be possible. These kinds of ethically problematic actions violate the individual's autonomy and cause emotional distress or develop negative feelings towards the hosting entity [10]. In the long run and in extreme cases, even the user's mental health and safety can be in danger because of the data leaks. While the third parties found in our study may not abuse collected personal data themselves, there is also possibility that the data ends to some more malevolent "fourth party" that can then cause the above mentioned problems to users.

6 Conclusion

We have presented a study on third-party data leaks on the 500 most visited websites in the world. We have seen that personal data is extensively collected without consideration for the user's consent. This is not only a problem for privacy and compliance with the GDPR, but also a concern for sustainable web development. Avoiding unnecessary and excessive third-party services and informing users properly about data collection are necessary in order to build more transparent, private and green web services. This especially goes for high-traffic websites, where adverse effects on users and environment are multiplied.

Acknowledgments

This research has been funded by Academy of Finland project 327397, IDA – Intimacy in Data-Driven Culture.

References

[1] Alex O Acheampong. 2018. Economic growth, CO2 emissions and energy consumption: what causes what and where? *Energy Economics* 74 (2018), 677–692.
[2] Adèle Azzi. 2018. The challenges faced by the extraterritorial scope of the General Data Protection Regulation. *J. Intell. Prop. Info. Tech. & Elec. Com. L.* 9 (2018), 126–137.
[3] Aruna Prem Bianzino, Claude Chaudet, Dario Rossi, and Jean-Louis Rougier. 2010. A survey of green networking research. *IEEE Communications Surveys & Tutorials* 14, 1 (2010), 3–20.
[4] Robin Carlsson, Panu Puhtila, and Sampsa Rauti. 2023. Towards an automatic tool for detecting third-party data leaks on websites. *Proceedings http://ceur-ws.org ISSN* 1613 (2023), 0073.
[5] Cookiebot. 2019. Ad tech surveillance on the public sector web. https://www.cookiebot.com/media/ 1121/cookiebot-report-2019-medium-size.pdf.
[6] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. 2018. We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. *arXiv preprint arXiv:1808.05096* (2018).
[7] Timi Heino, Sampsa Rauti, Robin Carlsson, and Ville Leppänen. 2023. Third-party services as a privacy threat on university websites. In *Proceedings of the 24th International Conference on Computer Systems and Technologies*. 134–138.
[8] Georgios Kontaxis, Michalis Polychronakis, and Evangelos P Markatos. 2012. Minimizing information disclosure to third parties in social login platforms. *International Journal of Information Security* 11 (2012), 321–332.
[9] Chiara Krisam, Heike Dietmann, Melanie Volkamer, and Oksana Kulyk. 2021. Dark Patterns in the Wild: Review of Cookie Disclaimer Designs on Top 500 German Websites. In *Proceedings of the 2021 European Symposium on Usable Security* (Karlsruhe, Germany) *(EuroUSEC '21)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3481357.3481516
[10] Lauren I Labrecque, Ereni Markos, Kunal Swani, and Priscilla Peña. 2021. When data security goes wrong: Examining the impact of stress, social contract violation, and data type on consumer coping responses following a data breach. *Journal of Business Research* 135 (2021), 559–571.
[11] Timothy Libert. 2015. Privacy implications of health information seeking on the web. *Commun. ACM* 58, 3 (2015), 68–77.
[12] Yabing Liu, Han Hee Song, Ignacio Bermudez, Alan Mislove, Mario Baldi, and Alok Tongaonkar. 2015. Identifying personal information in internet traffic. In *Proceedings of the 2015 ACM on Conference on Online Social Networks*. 59–70.
[13] Konstantin Naryshkin. 2010. *Study on the Leakage of Private User Information Via a Range of Popular Websites*. Ph. D. Dissertation. Worcester Polytechnic Institute.
[14] Panu Puhtila, Lauri Kivimäki, Timi Heino, Jari-Matti Mäkelä, Sampsa Rauti, and Tuomas Mäkilä. 2024. The Effect of Analytical Tools on Energy Consumption in Websites. To be published, accepted to ICT 4 Sustainability (ICT4S 2024)..
[15] Sampsa Rauti, Robin Carlsson, Sini Mickelsson, Tuomas Mäkilä, Timi Heino, Elina Pirjatanniemi, and Ville Leppänen. 2024. Analyzing third-party data leaks on online pharmacy websites. *Health and Technology* 14, 2 (2024), 375–392.
[16] Nayanamana Samarasinghe, Aashish Adhikari, Mohammad Mannan, and Amr Youssef. 2022. Et tu, brute? Privacy analysis of government websites and mobile apps. In *Proceedings of the ACM Web Conference 2022*. 564–575.
[17] Nayanamana Samarasinghe, Pranay Kapoor, Mohammad Mannan, and Amr Youssef. 2022. No salvation from trackers: Privacy analysis of religious websites and mobile apps. In *International Workshop on Data Privacy Management*. Springer, 151–166.

[18] Nir Sivan, Ron Bitton, and Asaf Shabtai. 2019. Analysis of location data leakage in the Internet traffic of Android-based mobile devices. In *22nd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2019)*. 243–260.
[19] Yuhwa Suh, Kiyoung Kim, Aran Kim, and Yongtae Shin. 2015. A study on impact of wired access networks for green Internet. *Journal of Network and Computer Applications* 57 (2015), 156–168.
[20] Xiufen Yu, Nayanamana Samarasinghe, Mohammad Mannan, and Amr Youssef. 2022. Got sick and tracked: Privacy analysis of hospital websites. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 278–286.
[21] Alexander R Zheutlin, Joshua D Niforatos, and Jeremy B Sussman. 2021. Data-tracking on government, non-profit, and commercial health-related websites. *Journal of general internal medicine* (2021), 1–3.
[22] Alexander R Zheutlin, Joshua D Niforatos, and Jeremy B Sussman. 2022. Data-tracking among digital pharmacies. *Annals of Pharmacotherapy* 56, 8 (2022), 958–962.