To Be or Not to Be (in the EU): Measurement of Discrepancies Presented in Cookie Paywalls

Andreas Stenwreth[1,*,a], Simon Täng[1,*,b] and Victor Morel[2,c]
[1] *Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden*
[2] *Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden*
{*andstenw, simonta, morelv*}*@chalmers.se*

Keywords: Cookie Banners, Cookie Paywalls, Pay-or-Okay, GDPR, Web Measurements.

Abstract: Cookie paywalls allow visitors to access the content of a website only after making a choice between paying a fee (paying option) or accepting tracking (cookie option). The practice has been studied in previous research in regard to its prevalence and legal standing, but the effects of the clients' device and geographic location remain unexplored. To address these questions, this study explores the effects of three factors: 1) the clients' browser, 2) the device type (desktop or mobile), and 3) the geographic location on the presence and behavior of cookie paywalls and the handling of users' data. Using an automatic crawler on our dataset composed of 804 websites that present a cookie paywall, we observed that the presence of a cookie paywall was most affected by the geographic location of the user. We further showed that both the behavior of a cookie paywall and the processing of user data are impacted by all three factors, but no patterns of significance could be found. Finally, an additional type of paywall was discovered on $\sim 11\%$ of the studied websites, coined the "double paywall", which consists of a cookie paywall complemented by another paywall once tracking is accepted.

1 INTRODUCTION

Cookie paywalls, sometimes denoted "pay-or-okay banner", allow visitors to access the content of a website only after making a choice between paying a fee or accepting tracking. This practice has recently triggered interest from regulators, resulting in Meta being charged for breaching the Digital Markets Act (European Commission, 2024) after introducing a pay-or-okay model on Facebook and Instagram. As this model is put under review, data is needed to give a thorough basis for regulators and legislative decision-making. However, previous studies have been limited in scope in regard to the plurality of browsers and environments (e.g. desktop or mobile) (Morel et al., 2023; Rasaii et al., 2023) despite observations that the browser may affect the presence of a paywall on a website (Morel et al., 2022). This paper therefore addresses the following research questions by studying the effect of the factors *browser*, *device type* and *geographic location*: 1) How is the prevalence of cookie paywalls affected by the studied factors? 2) Do the factors affect how much of a website can be accessed before a cookie paywall appears? 3) Do the factors affect how and by whom user data is processed?

When studying the pay-or-okay model, we discovered a new type of paywall, the "double paywall". The double paywall initially presents a cookie paywall, which is later complemented by an additional paywall after tracking (often performed via the use of cookies) has been accepted. We contacted a legal scholar, who argued for the need for quantitative insights regarding the adoption of this practice to provide a valuable basis for further legal examination. As a result, this paper also studies the question: 4) How widespread is the occurrence of double paywalls?

To answer these questions, we develop a crawler to extract data from 804 websites using 26 combinations of browser, operating system and geographic location. The main contributions of this paper can be summarized as follows: (1) We provide the largest dataset of cookie paywalls to date. (2) We found that the geographic location of the user has an impact on whether a cookie paywall is presented. (3) We introduce the first dataset of double paywalls (93 websites). This data has already been presented during a hearing to the European Commission in July 2024.

[a] https://orcid.org/0009-0000-4466-8561
[b] https://orcid.org/0009-0007-4340-4370
[c] https://orcid.org/0000-0001-9482-8906
[*] Authors with equal contribution and importance.

Stenwreth, A., Täng, S. and Morel, V.
To Be or Not to Be (in the EU): Measurement of Discrepancies Presented in Cookie Paywalls.
DOI: 10.5220/0013107800003899

In *Proceedings of the 11th International Conference on Information Systems Security and Privacy (ICISSP 2025) - Volume 1*, pages 183-190
ISBN: 978-989-758-735-1; ISSN: 2184-4356

2 BACKGROUND

This section describes the relevant legal notions in EU law, the actors and the technical standard central to our analysis, and the related work.

2.1 Legal Landscape

In the EU, two main data protection laws, namely the General Data Privacy Regulation (GDPR) (European Parliament and Council of the European Union, 2016) and the ePrivacy Directive (ePD) (European Parliament and Council of the European Union, 2002), regulate the processing, storing, and management of personal data (for the GDPR), and of cookies more precisely (for the ePD). In this context, personal data is “any information relating to an identified or identifiable natural person”, and because a significant amount of data on the web is considered personal data (typically, data stored in web cookies), organizations collecting such data (i.e. controllers) must abide by the requirements laid out in these texts. They notably require that controllers choose a **legal basis** (i.e. a legally valid high-level reason), such as the *consent* of the user or the *legitimate interest* of the organization (GDPR Recital 36), and a specific purpose (EDPB, 2020, Parag. 121) for the processing of personal data. When using cookies for the intention of tracking or targeted advertising, the GDPR and ePD state that user *consent is the only applicable legal basis* (European Parliament and Council of the European Union, 2002, Art. 5(3)) and users must be notified of any tracking and be able to act, opt out, or leave before any tracking is initiated (Article 6).

2.2 Transparency and Consent Framework

The Transparency and Consent Framework (TCF), created by the Interactive Advertising Bureau (IAB), is an industry standard and tool designed to create a standardized experience when making privacy choices on websites (IAB Europe, 2023b). The standard specifies eleven purposes for data processing that can rely on one or both legal bases of consent and legitimate interest (depending on whether the purpose relates to targeted advertising or not). The elicitation of consent and the communication of the purposes and legal bases for data processing is implemented through consent notices, such as cookie banners, commonly provided by Consent Management Platforms (CMPs).

CMPs provide functions including presenting a consent banner to ask for user consent and logging the provided response of a user (IAB Europe, 2024). After consent is given,[1] the CMP regulates the activation of cookies and other technologies based on the given consent in line with the EU data protection regulations. Additionally, CMPs package user preferences in a standardized payload called the Transparency and Consent String (TC String).

A TC String is an encoded HTTP-transferable string that enables communication of transparency and consent information (IAB Tech Lab, 2023). The TC String contains information such as the number of vendors to which data is conveyed based on user consent or legitimate interest, as well as the purposes for which consent and legitimate interest are used as a basis for data processing. The information is passed to all relevant parties including the data subject, the publisher, and the vendors. A TC String is stored in the user’s web browser as either a persistent cookie or as an entry in the browser’s local storage.

Together with presenting a consent banner, many CMPs also provide integration with cross-site subscription-based models – provided by companies called Subscription Management Platforms (SMPs) such as Content Pass GmbH (Content Pass GmbH, 2024a) and Traffective GmbH (Traffective GmbH, 2024a) –, as a way of offering additional revenue streams for websites. These SMPs provide the option of paying a monthly fee to gain access to all partnered websites without personalized advertisement (Pfau, 2023).

2.3 Related Work

The first study to investigate the privacy impacts of paywalls was conducted by (Papadopoulos et al., 2020) in which paywalls were categorized based on the restrictions put on the user. However, *cookie* paywalls were first studied by (Morel et al., 2022) who manually identified and found cookie paywalls on 13 out of the 2800 websites they studied. The detection process was later automated by (Morel et al., 2023) and (Rasaii et al., 2023) by employing a web crawler using the Mozilla Firefox web browser. Neither of the studies considered the use of different browsers nor examined the role of mobile devices. (Morel et al., 2023) found 431 websites using cookie paywalls, all of which used the TCF, and (Rasaii et al., 2023) found 280 websites. Both studies used language-based approaches to identify whether a website uses a cookie paywall, leveraging paywall-related keywords.

[1] Note that for legitimate interest, no affirmative action is expected, only informing users is required.

3 METHODOLOGY

We present here how we designed the web crawler used to collect data from 804 websites in June 2024. The crawler is built using Selenium (Software Freedom Conservancy, 2024) and Appium (OpenJS Foundation, 2023) to allow the use of desktop and mobile browsers to search websites for cookie and double paywalls and collect and analyze TC Strings. It leverages natural text processing and HTML IDs and class names commonly used by cookie paywalls to detect cookie and double paywalls. Moreover, the crawler is configurable using three factors, the geographic location, operating system and browser, to allow for the comparison of website behavior between different device setups.

3.1 Factors and Dataset

We use two geographic locations, Gothenburg in Sweden and New York in the USA, covering both a location covered by the GDPR and ePD (Gothenburg) and one that is not (New York). The Swedish location was accessed directly and the USA location was set up using a VPN service. Moreover, we use five operating systems: Windows, macOS, Linux, Android, and iOS, as well as the top four most used web browsers: Google Chrome, Safari, Microsoft Edge, and Mozilla Firefox (StatCounter, 2023). These combine into 26 different configurations of the crawler.

The dataset consists of the 431 websites found to be using cookie paywalls by (Morel et al., 2023), the 280 websites found by (Rasaii et al., 2023), plus an additional 441 and 215 websites listed on the webpages of the only two SMPs known to the authors at the time of the study, *contentpass.net* (Content Pass GmbH, 2024b) and *freechoice.club* (Traffective GmbH, 2024b). After cleaning the dataset of duplicates and websites that did not present cookie paywalls, these combine into a final dataset of 804 websites[2]

3.2 Exploratory Phase

We conducted pilot crawls throughout the start of the study to create a set of HTML IDs and class names used for the detection of cookie paywalls as well as a corpus of syntagms[3] related to cookie paywalls. During these crawls, we also found different implementations of cookie paywalls where a user has to interact with the website (move the mouse or scroll) for a paywall to appear, or two-tiered cookie paywalls that present themselves as normal cookie banners until the user declines cookies. These findings were later used in the design of the automated detection approach of the full-scale crawler.

3.3 Cookie Paywall Detection Approach

To accommodate for the different types of implementations of cookie paywalls, the detection algorithm is run several times, with one action performed between each run. A website gets flagged as not using a cookie paywall only if moving the mouse, scrolling, and trying to click a reject button did not result in any cookie paywall being detected.

To find web elements related to cookie paywalls, the detection algorithm searches for HTML IDs and class names associated with cookie paywalls. It does this by parsing the Document Object Model (DOM) representation of a website.

Elements found to relate to cookie paywalls are first searched for text confirming the presence of a cookie paywall using the corpus of syntagms. Two syntagms are, in general, considered sufficient to identify a cookie paywall due to the two-choice nature of a cookie paywall, containing a payment option and a cookie option. If any of the syntagm combinations are found, the website is flagged as using a cookie paywall. If no cookie paywall is found in the searched elements, the text in the entire DOM is searched using the same approach.

Due to the infeasibility of a manual verification on the entire dataset, we measured the accuracy of the cookie paywall detection by manually inspecting a random subset of 10% of the collected data. We found that the crawler produced false negatives for six cookie paywalls (three websites in two configurations). Thus, the crawler has an accuracy and recall of 99.7% and precision of 100%.

3.4 TC String Collection and Analysis

When a cookie paywall is found, the crawler retrieves the TC String at two separate instances: before and after all cookies are accepted. The TC string is collected primarily by leveraging the TCF API or alternatively, if the TCF API is unavailable, by searching the browser's cookie and local storage.

If found, the TC String is collected and decoded using the website *iabtcf.com* (IAB Tech Lab, 2024). Using this decoder, the program extracts the purposes for data processing and the number of vendors to which data is conveyed based on both consent and legitimate interest.

[2] https://github.com/Trooja/cookie-paywall-discrepancies/blob/main/cookie-paywalls.csv

[3] Sets of words with a sequential relationship.

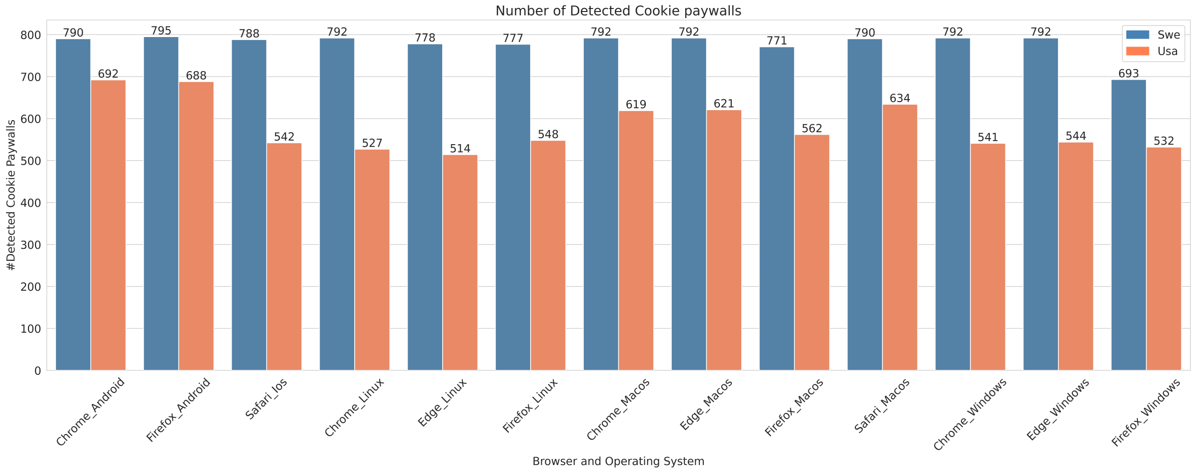

Figure 1: Number of cookie paywalls detected for each combination of the studied factors.

3.5 Double Paywall Detection Approach

After accepting cookies, the website can be interacted with and traversed in order to search for additional paywalls. When exploring the dataset, these paywalls were only found on “premium” parts of websites, such as subscription-only articles, often indicated by some “premium icon” with a descriptive class name, such as *premium-icon*. If no elements with premium indicators are found, the crawler searches for any article, and as a last resort, any link that does not lead off the website.

The detection of additional paywalls is done heuristically in an ad hoc fashion by searching the webpage for visible elements that either have a class name or ID indicating the presence of a paywall, for example *paywall* or *paid-barrier*. If such an element is found, the crawler takes a screenshot of the element and saves the URL for manual inspection. If no potential paywalls are found on a webpage, the crawler returns to the previous page and searches again up to three times.

To verify the accuracy of the double paywall detection approach, the screenshots taken by the crawler and the collected URLs were manually inspected. The crawler identified 104 websites as presenting a double paywall, 93 of which had an actual double paywall, resulting in an accuracy of 88%.

4 RESULTS

This chapter presents the data collected from the websites in the dataset, see Section 3.1. The data was collected using the 26 possible configurations of the crawler, that is, combinations of browser, operating system and geographic location. Further details can be found in the full version of the paper.[4]

4.1 Prevalence of Cookie Paywalls

Figure 1 presents the number of cookie paywalls detected in each of the configurations. The highest number of cookie paywalls found in any configuration was 795, using Firefox on Android from the Swedish vantage point, and the lowest was 514 when using Edge on Linux from the USA. On average, 780 cookie paywalls were detected for each configuration in Sweden, and 581 for the configurations in the USA. Additionally, the number of detected cookie paywalls in Sweden was higher than that of the USA across all configurations of the crawler.

When considering the CMPs used by the websites, it was found that the CMP *Traffective GmbH* was overrepresented among the websites presenting a cookie paywall in Sweden but not in the USA. This CMP was used by approximately 70% of these websites for all browser and operating system combinations, except when using Android. When using an Android device, the CMP *Traffective GmbH* was the second most prevalent CMP at 21% after *Sourcepoint Technologies Inc* at 23%.

4.2 Actions Required to Display a Cookie Paywall

Cookie paywalls were detected by the crawler on 800 websites and produced complete data for 380 websites, that is, a cookie paywall was found in every combination of the studied factors. The vast major-

[4] https://arxiv.org/abs/2410.06920

ity of these 380 websites (approximately 99.3%) presented a cookie paywall immediately after entering the website, regardless of the browser, operating system and geographic location. Very few websites required the user to move the mouse (0.64%) and even fewer (0.06%) required the user to scroll. Across all combinations, only 17 websites showed any kind of discrepancy in what action was required between different configurations of the crawler, and each of these websites displayed the cookie paywall immediately in at least one configuration. No website required the user to scroll or move the mouse in all configurations.

4.3 Processing of User Data Conveyed Through the TC String

The crawler produced complete data (a valid TC String was collected in every configuration) for 238 websites before accepting cookies and 335 websites after accepting cookies. The following results are derived from data collected on these 238 and 335 websites.

4.3.1 Number of Vendors

The number of websites that showed any discrepancies in the number of registered vendors after accepting cookies were 124 and 23 for consent and legitimate interest, respectively. Before accepting cookies, no websites showed any discrepancies in the number of vendors using consent, and 16 websites showed discrepancies for vendors using legitimate interest.

The mean number of vendors for consent after accepting cookies for all websites was 280.23 when aggregating over all combinations of factors. When comparing this value with the aggregated number of vendors for each studied factor, the largest deviation from the total mean was no more than 0.20%, that is, 0.57 vendors. For legitimate interest, the largest deviation from its total mean of 37.55 vendors was 0.34%, i.e. 0.13 vendors.

The difference was also small on a per-website basis where the max difference between two combinations for a given website was 26 vendors[5] for consent after acceptance and the average max difference was 1.21 with a standard deviation of 2.53. More detailed information can be found in Table 1.

4.3.2 Purposes for Data Processing

No website registered any purposes based on consent before cookies were accepted. After cookies were accepted, only one website registered different purposes based on consent depending on the configuration used on the crawler. On this website, fewer purposes were registered for all combinations involving Firefox. For legitimate interest, 16 and 37 websites differed between configurations before and after accepting cookies, respectively. In all cases where some discrepancy was found, the difference consisted of a varying number of purposes registered. When accessing these websites from Sweden, the average number of purposes registered was slightly lower than when accessing from the USA both before and after consenting to cookies (0.860 and 0.873 before, 2.945 and 3.144 after).

It was further found that 20 websites used legitimate interest as a basis to select or create a profile for personalized ads or content when accessed from the USA, whereas no websites did this from the Swedish vantage point. When including the websites where complete data could not be obtained, this number increases to 39. All of these websites used *consentmanager.net* as their CMP.

4.4 Prevalence of Double Paywalls

We found 93 websites[6] that presented a double paywall. Out of these, 73 websites were classified as news websites according to Cyren's URL Category Checker (Cyren, 2024), followed by the categories *Computers & Technology* and *Business*. Furthermore, 68 websites were hosted in Germany and all double paywalls were hosted in EU countries.

Table 1: Statistics over the max difference in the number of vendors on a per-website basis.

Cookies Accepted	Legal Basis	Maximal Max Diff	Mean Max Diff	Standard Deviation
No	Consent	0	0	0
	Legitimate Interest	3	0.12	0.43
Yes	Consent	26	1.21	2.53
	Legitimate Interest	6	0.30	0.88

[5] Found on the website *as.com* with 806 vendors on Firefox on Windows from the USA and 780 on Chrome on Linux from Sweden.

[6] https://github.com/Trooja/cookie-paywall-discrepancies/blob/main/double-paywalls.csv

5 DISCUSSION

This section discusses the results presented in Section 4 followed by limitations of the study and suggestions for future work.

5.1 Prevalence of Cookie Paywalls

Out of the studied factors, the geographic location has the largest impact on whether a website presents a cookie paywall or not. For all combinations of browser and operating system, fewer cookie paywalls were detected when accessing websites from the USA than when accessing them from Sweden. It was found that, in all but two combinations, the majority of websites showing this type of behavior used the CMP *Traffective GmbH*. Thus, the CMP presents a pattern of almost systematically choosing whether to present a cookie paywall based on the geographic location of the user. However, six websites using this CMP did not exhibit this behavior in any of the combinations, indicating the presence of other, possibly website-related, factors. Additionally, the CMP was only used by 20% of the websites displaying this behavior on the two combinations using Android, which would suggest that the choice of browser and operating system is also of some importance.

The type of device (mobile or desktop) accessing a website does not appear to be directly correlated to the prevalence of cookie paywalls. There were larger differences in the number of detected cookie paywalls between the two mobile devices than the difference between the iOS device and any of the desktop operating systems. However, using Android as the operating system, especially in the USA, resulted in more cookie paywalls being encountered. Thus, the role of the operating system seems to be of greater importance than the type of device used. Furthermore, the operating system seems to play a bigger role when accessing websites from the USA than when accessing them from Sweden.

In general, the choice of web browser does not seem to impact the prevalence of cookie paywalls on its own. In the majority of cases, fewer cookie paywalls were encountered when using Firefox compared to the other browsers, but the average number of cookie paywalls did not differ by more than 25 websites. **However, the combination Firefox on Windows from Sweden produced an outlier.** This combination presented 78 fewer cookie paywalls than the second lowest encounter rate in Sweden, and only one more cookie paywall than the highest encounter rate in the USA. The reason for this outlier is unknown but may partly be a result of the CMP *Sourcepoint Technologies*, as this CMP was used by approximately 67% of the websites with no cookie paywall in this configuration.

To summarize: *The prevalence of cookie paywalls was most affected by the geographic location used to access them, with more cookie paywalls being displayed from the Swedish vantage point than the USA. The browser had little effect on the number of cookie paywalls encountered, one combination of all three factors produced an outlier. The device type did not seem to directly correlate to the prevalence of cookie paywalls, but the operating system of the device did.*

5.2 Actions Required to Display a Cookie Paywall

The vast majority of cookie paywalls appeared immediately, and few websites required that the crawler moved the mouse before the cookie paywall was displayed. Only 17 websites, approximately 4% of the websites with complete data, displayed differences in the required action when accessed using different configurations of the crawler. This small number of websites showing any kind of discrepancy makes it infeasible to determine if these differences depend on the studied factors or inconsistencies in the behavior of the crawler, network latency or some other factor. Consequently, it is plausible that these discrepancies are website specific rather than a result of the different configurations used.

To summarize: *99.3% of displayed cookie paywalls appeared immediately, and only 17 websites required any type of action for the cookie paywall to appear. Due to the low number of websites displaying discrepancies, the individual effect of the studied factors remains unclear and the possibility that this was a website-specific behavior cannot be ruled out.*

5.3 Processing of User Data Conveyed Through the TC String

The analysis of the number of vendors and purposes for the different legal bases was solely conducted on websites that provided this information for every combination. This choice enabled the creation of comparable results between configurations, but may also have led to potential patterns existing in the entire dataset being missed.

Discrepancies in the number of vendors between different combinations of browser, operating system and geographic location were found, but the differences were too small to make any

concrete connection to a specific factor. On a per-website basis, some differences were discovered between the combinations of the studied factors, but these differences were, in general, small. Furthermore, no combination of factors deviated significantly from the total mean, with the largest deviation being 0.34%. These deviations were a result of a small subset of websites that on average differed by one vendor.

The number of registered purposes varied between different combinations of the studied factors, but too few websites showed any discrepancies to distinguish a pattern. Less than 12% (at most 37 websites) showed some kind of discrepancy between different configurations for user consent and legitimate interest, both before and after accepting cookies. More purposes were registered when accessed from the USA than from Sweden, but due to the low number of websites presenting any type of discrepancy, this cannot be attributed to a greater trend.

When accessed from the USA, 39 websites used legitimate interest as a basis to select or create a profile for personalized ads and content, but no websites did so when accessed from the Swedish vantage point. The use of this legal basis for these purposes is prohibited in the current version of the TCF framework (TCF v2.2) after a decision by the Belgian Data Protection Authority in 2022 (IAB Europe, 2023a). This indicates that websites using *consentmanager.net* as its CMP may, based on the geographic location, disregard the rules of the IAB Europe TCF and change how the collected user data is processed.

To summarize: *Discrepancies were found in the number of vendors user data was shared with, and the purposes for which these vendors processed data, but the differences were small and were only found on a subset of the websites. On these websites, a slight tendency for more purposes was displayed when accessed from the USA.*

5.4 Prevalence of Double Paywalls

Approximately 11.6% of websites used a double paywall, of which 73% were hosted in Germany. The majority of websites using a double paywall (78%) were categorized as news websites, including several large news sites such as *zeit.de*, *lemonde.fr* and *abc.es*. This is significantly higher than the proportion of news sites among websites using cookie paywalls found in previous research (Morel et al., 2023). A possible explanation could be that this practice lends itself well to news websites, as it may allow a website to entice users to subscribe to the newspaper through providing a subset of its content, whilst still being able to monetize this audience through targeted advertising. However, the predominance of news websites may instead partially be an effect of a bias in the methodology. The detection of double paywalls was automated, using a limited number of indicators that an additional paywall was used. Thus, extending the set of indicators could provide a less biased search and potentially a wider variety of website categories.

To summarize: *The use of double paywalls is fairly widespread, with 93 websites using one. The majority of these were classified as news websites and found in Germany.*

5.5 Limitations

Our study provides insight into what discrepancies can be found in cookie paywalls between different browsers, operating systems and geographic locations. However, it is important to consider certain limitations: first, an automated approach was used to collect data for the analysis of the study. A subset of this data was manually examined to provide a degree of confidence for the conclusions, but this manual verification might not guarantee complete accuracy for all the collected data. Second, websites have previously been proven to be able to detect Selenium-driven web browsers and consequently alter their behavior (Cassel et al., 2022). It has also been shown that websites may deliver different content to detected web bots than to normal users (Jonker et al., 2019). Because of this, the crawler may not fully represent the regular website behavior of an actual user. Finally, the data for this study was collected over several weeks. Thus, there exists a possibility that some websites would have changed their behavior during the time of the data collection.

5.6 Future Work

Future studies may want to use a larger set of geographic locations and a wider range of browsers on common operating systems to cover more combinations of factors and get a more fine-grained overview. Another continuation of this study would be to explore not only the presence of cookie paywalls, but also what, if anything, replaces a cookie paywall on a website if it appears in some configurations but not in others. Additionally, a comparison of the tracking conducted by such websites could provide an understanding of how the collection of consent affects how user data is collected and used.

6 CONCLUSION

We presented in this paper the first study on the effects of the web browser, device type and geographic location on the presence and behavior of cookie paywalls, and their handling of users' data. We built an automated crawler to collect data from 804 websites with confirmed cookie paywalls. Using this data, we showed that all factors affected cookie paywalls to some degree – the location being the most impactful – , and that changing the combination of the factors predominantly affected the presence of the cookie paywall. Finally, we produced the first dataset of a new type of paywall coined *double paywall*.

ACKNOWLEDGMENTS

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

REFERENCES

Cassel, D., Lin, S.-C., Buraggina, A., Wang, W., Zhang, A., Bauer, L., Hsiao, H.-C., Jia, L., and Libert, T. (2022). OmniCrawl: Comprehensive measurement of web tracking with real desktop and mobile browsers. In *POPETS '22*.

Content Pass GmbH (2024a). contentpass. https://www.contentpass.net/. (Accessed on 2024/05/07).

Content Pass GmbH (2024b). Partner websites. https://www.contentpass.net/en/publications. (Accessed on 2024/03/21).

Cyren (2024). Website URL category check. https://data443.com/cyren-url-category-check-gate/. (Accessed on 2024/06/18).

EDPB (2020). Guidelines 05/2020 on consent under regulation 2016/679. Accessed on 2024/09/23.

European Commission (2024). Commission sends preliminary findings to meta over its "pay or consent" model for breach of the digital markets act. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3582. (Accessed on 2024/09/23).

European Parliament and Council of the European Union (2002). Directive 2002/58/EC of the european parliament and of the council of 12 july 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (directive on privacy and electronic communications).

European Parliament and Council of the European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

IAB Europe (2023a). FAQ: APD DECISION ON IAB EUROPE AND TCF - updated february 2023. https://iabeurope.eu/wp-content/uploads/FAQ_-APD-DECISION-ON-IAB-EUROPE-AND-TCF-Updated-Febraury-2023.docx.pdf. (Accessed on 2024/03/13).

IAB Europe (2023b). The transparency & consent framework (TCF) v2.2. https://iabeurope.eu/transparency-consent-framework/. (Accessed on 2024/03/13).

IAB Europe (2024). TCF for CMPs. https://iabeurope.eu/tcf-for-cmps/. (Accessed on 2024/09/25).

IAB Tech Lab (2023). Transparency and consent string with global vendor & CMP list formats. https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework/blob/master/TCFv2/IAB%20Tech%20Lab%20-%20Consent%20string%20and%20vendor%20list%20formats%20v2.md. (Accessed on 2024/03/13).

IAB Tech Lab (2024). @iabtcf. https://iabtcf.com/?/decode#/. (Accessed on 2024/05/16).

Jonker, H., Krumnow, B., and Vlot, G. (2019). Fingerprint surface-based detection of web bot detectors. In *ESORICS '19*.

Morel, V., Santos, C., Fredholm, V., and Thunberg, A. (2023). Legitimate interest is the new consent - large-scale measurement and legal compliance of IAB europe TCF paywalls. In *CCS '23*.

Morel, V., Santos, C., Lintao, Y., and Human, S. (2022). Your consent is worth 75 euros a year - measurement and lawfulness of cookie paywalls. In *WPES'22*.

OpenJS Foundation (2023). Welcome. Appium Documentation. https://appium.io/docs/en/latest. (Accessed on 2024/09/18).

Papadopoulos, P., Snyder, P., Athanasakis, D., and Livshits, B. (2020). Keeping out the masses: Understanding the popularity and implications of internet paywalls. In *WWW '20*.

Pfau, D. (2023). PUR models status quo on the european market. https://iabeurope.eu/wp-content/uploads/PUR-Modelle-bvdw_20231004-en.pdf.

Rasaii, A., Gosain, D., and Gasser, O. (2023). Thou shalt not reject: Analyzing accept-or-pay cookie banners on the web. In *IMC '23*.

Software Freedom Conservancy (2024). The selenium browser automation project. https://www.selenium.dev/documentation/. (Accessed on 2024/09/18).

StatCounter (2023). Browser market share worldwide. https://gs.statcounter.com/browser-market-share. (Accessed on 2024/03/13).

Traffective GmbH (2024a). Freechoice | Deine Wahl! https://freechoice.club/. (Accessed on 2024/05/07).

Traffective GmbH (2024b). The freechoice universe. https://freechoice.club/en/partner/. (Accessed on 2024/03/21).