Universität des Saarlandes

Understanding Impact of Banner Interaction on User Tracking via Web Cookies

Dissertation zur Erlangerung des Grades des
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Fakultät für Mathematik und Informatik der
Universität des Saarlandes

vorgelegt von

Seyedali Rasaii

Saarbrücken
2025

Dissertation Defense Details

Dean of the Faculty MI:	Prof. Dr. Roland Speicher
Date of the defense:	18th March, 2026

Members of the committee:

Chairperson:	Prof. Dr. Ingmar Weber
Reviewers:	Prof. Anja Feldmann, PhD
	Dr. Devashish Gosain
	Dr. Oliver Gasser
	Dr. Wouter Lueks
Academic assistant:	Dr. Tiago Heinrich

Abstract

The Internet has become a cornerstone of modern society, revolutionizing how people communicate, conduct business, and access information. Its connectivity capabilities have enabled the emergence of new types of services in the form of online platforms such as messaging and shopping. In parallel, advertising has become a dominant revenue model for these platforms. Consequently, advertising companies utilize tracking technologies, such as cookies, to collect large amounts of data, including Personally Identifiable Information (PII), enabling them to deliver personalized content and targeted advertisements. Although these practices enhance user experience and platform monetization, the extensive collection, storage, and analysis of personal data at scale raises concerns regarding user autonomy and privacy.

Recognizing these concerns, governments and regulatory bodies have introduced frameworks, including the General Data Protection Regulation (GDPR), to safeguard user privacy. Under the GDPR, any website or service that collects or processes personal data[1] of individuals located in the EU must obtain users' consent prior to data collection. In response to publishers' efforts to comply with these requirements, consent mechanisms—such as cookie banners—have emerged to inform users about data collection practices and obtain explicit consent for processing their personal data.

In this study, we assess the prevalence of cookie banners and the impact of user interactions with them on the deployment and transmission of web cookies under various configuration setups. To enable this analysis in a fully automated and large-scale manner, we introduce *BannerClick*, a tool capable of interacting with cookie banners with nearly 99% accuracy. Using *BannerClick*, we first examine the web cookie landscape through multiple lenses, including geographic location, device type, and banner interaction modes, providing a comprehensive understanding of how these factors influence the deployment of tracking cookies.

Furthermore, we investigate the emerging trend of cookie paywalls, which allow users to choose between accepting tracking or paying for an ad-free experience, assessing their implications for privacy and accessibility. Lastly, we explore the structural misalignments between existing consent mechanisms and regulatory objectives, demonstrating how seemingly opposing interactions with banners can influence the behavior of subsequently visited websites, thereby extending unwanted tracking via cookies. Ultimately, our findings highlight significant gaps between current consent practices—primarily in the form of cookie banners—and the objectives of privacy laws, emphasizing the need for more practical and user-centric solutions.

[1] GDPR Article 4(1) defines personal data as any information relating to an identified or identifiable natural person (data subject).

Zusammenfassung

Das Internet ist zu einem Eckpfeiler der modernen Gesellschaft geworden und hat revolutioniert, wie Menschen kommunizieren, Geschäfte tätigen und Informationen abrufen. Seine weitreichenden Vernetzungsmöglichkeiten haben zur Entstehung neuer Dienstleistungen in Form von Online-Plattformen wie Messaging und Online-Shopping geführt. Parallel dazu ist Werbung zu einem dominierenden Einnahmemodell für diese Plattformen geworden. Werbeunternehmen setzen dementsprechend Tracking-Technologien wie Cookies ein, um groSSe Mengen an Daten, einschlieSSlich personenbezogener Informationen (Personally Identifiable Information, PII), zu sammeln und dadurch personalisierte Inhalte sowie gezielte Werbung bereitzustellen. Obwohl diese Praktiken die Benutzererfahrung und Monetarisierung verbessern, wirft die umfassende Sammlung und Analyse persönlicher Daten erhebliche Bedenken hinsichtlich der Autonomie und Privatsphäre der Nutzer auf.

Angesichts dieser Bedenken haben Regierungen und Regulierungsbehörden Vorschriften wie die Datenschutz-Grundverordnung (DSGVO) eingeführt, um die Privatsphäre der Nutzer zu schützen. GemäSS der DSGVO muss jede Website oder jeder Dienst, der personenbezogene Daten[2] von Personen innerhalb der EU verarbeitet, vor der Datenerhebung eine ausdrückliche Einwilligung einholen. Als Reaktion darauf haben sich Einwilligungsmechanismen—wie Cookie-Banner—entwickelt, um Nutzer über Datenerfassungspraktiken zu informieren und deren Zustimmung einzuholen.

In dieser Studie analysieren wir die Verbreitung von Cookie-Bannern sowie den Einfluss von Nutzerinteraktionen auf die Platzierung und Übertragung von Web-Cookies unter verschiedenen Konfigurationsbedingungen. Für diese groSSangelegte, automatisierte Analyse stellen wir *BannerClick* vor—ein Tool, das mit einer Genauigkeit von nahezu 99 % mit Cookie-Bannern interagiert. Mithilfe von *BannerClick* untersuchen wir die Cookie-Landschaft des Webs aus verschiedenen Perspektiven, darunter geografische Lage, Gerätetyp und Interaktionsmodi, um zu verstehen, wie diese Faktoren die Platzierung von Tracking-Cookies beeinflussen.

Darüber hinaus analysieren wir den aufkommenden Trend der Cookie-Paywalls, die Nutzern die Wahl lassen, entweder Tracking zu akzeptieren oder für ein werbefreies Erlebnis zu bezahlen, und bewerten deren Auswirkungen auf Privatsphäre und Zugänglichkeit. SchlieSSlich untersuchen wir strukturelle Diskrepanzen zwischen bestehenden Einwilligungsmechanismen und regulatorischen Zielen, um aufzuzeigen, wie verschiedene Interaktionen mit Bannern das Verhalten nachfolgend besuchter Websites beeinflussen und so unerwünschtes Tracking verlängern können. Unsere Ergebnisse verdeutlichen erhebliche Lücken zwischen aktuellen Einwilligungspraktiken—hauptsächlich in Form von Cookie-Bannern—und den Zielen der Datenschutzgesetze und unterstreichen die Notwendigkeit praktischerer, nutzerzentrierter Lösungen.

[2] GemäSS DSGVO Artikel 4 Absatz 1 sind personenbezogene Daten jegliche Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person („betroffene Person“) beziehen.

Dedicated to my family for their presence and supporting energy
throughout my academic journey

Publications

The foundation of this thesis is built upon three following peer-reviewed papers published or submitted to international conferences. All my collaborators are acknowledged as co-authors.

- Ali Rasaii, Shivani Singh, Devashish Gosain, and Oliver Gasser. "Exploring the Cookieverse: A Multi-Perspective Analysis of Web Cookies," In: *Proceedings of the 2023 Passive and Active Measurement Conference (PAM '23)* . Lecture Notes in Computer Science, vol 13882. Springer, Cham. `https://doi.org/10.1007/978-3-031-28486-1_26` (Mar 2023).

- Ali Rasaii, Devashish Gosain, and Oliver Gasser. "Thou Shalt not Reject: Analyzing Accept-or-pay Cookie Banners on the Web," In: *Proceedings of the 2023 ACM on Internet Measurement Conference*, p. 154-161. IMC '23, ACM, New York, NY, USA. `https://doi.org/10.1145/3618257.3624846` (October 2023).

- Ali Rasaii, Ha Dao, Anja Feldmann, Mohammadmahdi Javid, Oliver Gasser, and Devashish Gosain. "Intractable Cookie Crumbs: Unveiling the Nexus of Stateful Banner Interaction and Tracking Cookies," *Privacy Enhancing Technologies 2025 (PET '25)* under review.

Table of Contents

1

Introduction

The Internet has significantly transformed human interaction, reshaping communication, commerce, and access to information on an unprecedented scale [58, 87, 94]. It has become an indispensable part of daily life, enabling seamless global connectivity and a wide range of digital services. However, as the Internet continues to evolve, it introduces new challenges, especially concerning user privacy [2, 38, 133, 152].

The widespread adoption of online platforms—such as shopping sites, news websites, and social media—has created an ideal medium for businesses to deliver advertisements. Today, the web functions as the backbone of an ecosystem that provides free digital applications and services by leveraging user data for targeted advertising. Personalized advertising has thus become the primary revenue model for numerous online platforms, enabling companies to tailor ads according to individual preferences and browsing behavior [114]. While this model supports the financial sustainability of online content and services, it heavily relies on extensive data collection and behavioral tracking—practices that are often carried out without users' awareness [115].

Leading technology corporations, such as Alphabet (Google), Meta (Facebook), and Amazon, have built vast digital ecosystems centered around data-driven advertising. For example, statistics [12, 140] show that Google and Facebook are generating 77% and 98.4% of their revenue from ads. Consequently, these companies continuously refine their ability to analyze, predict, and influence user behavior by collecting massive amounts of personal data [161]. Though these efforts are often justified as necessary for improving user experiences and optimizing services, they also serve commercial interests, allowing advertisers to reach audiences with higher accuracy [80, 114].

To further enhance user targeting, these platforms may leverage online tracking techniques—such as cookies and fingerprinting—that monitor users' activities across websites and devices. As a result, amid growing privacy concerns [11, 83, 134] and the ethical implications of large-scale data collection and processing [142], many regulatory bodies have introduced privacy laws and guidelines to protect user rights and establish boundaries for data collection and use [19, 39]. While these regulations are a promising step forward, their practical effectiveness remains an ongoing challenge.

1.1 Online Tracking

Online tracking[1] encompasses a range of techniques [1,16,35,36,66,84] used to monitor and analyze user behavior across the web. The most common and widely recognized of these techniques are web cookies—small text files stored on a user's device upon visiting websites. Cookies facilitate various functions, such as maintaining user logins or remembering preferences. However, they also serve as powerful tracking tools, especially third-party cookies, which are set by domains other than the one visited by the user. These cookies enable advertisers and analytics providers to follow users across multiple websites, building detailed behavioral profiles[2].

Beyond web cookies, there are many other techniques that enable trackers to bypass existing mitigation strategies, such as blocking third-party cookies. Browser fingerprinting [81], for example, collects unique attributes of a user's browser—such as screen resolution, installed fonts, or browser extensions—to generate an identifier that remains persistent across sessions. Unlike cookies, fingerprinting does not require storing any data on the user's device, making it significantly harder to detect and block. Device fingerprinting [36] extends these capabilities further by leveraging hardware attributes, allowing trackers to recognize users even when they switch between browsers or clear cookies. Similarly, tracking pixels [106]—invisible images embedded in web pages or emails—enable third parties to monitor user interactions without their knowledge.

The covert nature of these tracking techniques and the extensive collection of user data have raised privacy concerns among users and privacy advocates, leading to a growing demand for transparency and control over personal information [11, 83, 134]. As a result, legislative bodies have responded to the increasing awareness of privacy issues, which has been manifested in recent data protection laws such as the General Data Protection Regulation (GDPR) in the European Union [39] and the California Consumer Privacy Act (CCPA) in California [19]. These laws aim to empower individuals by granting them greater control over their personal data and online privacy. For instance, under the GDPR, websites are required to inform users and obtain their consent prior to any data collection or processing.

Consequently, websites now frequently display consent banners to comply with data protection regulations. These banners typically offer users options to accept, reject, or customize tracking preferences. In addition, there has been a notable improvement in the use of privacy-related language and the visibility of websites' privacy terms, such as links to privacy policies [4, 78]. Nevertheless, questions remain regarding the effectiveness of these measures in properly informing users about the privacy implications of their online behavior, as well as in ensuring that websites respect users' autonomy over their personal data and how it is collected and used.

[1] In the context of this work, we use the terms *online tracking* and *web tracking* interchangeably, as *online* here refers to the connection of users to online platforms via web infrastructure, primarily in the form of websites.

[2] GDPR Article 4(4) defines profiling as any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person.

In this work, we investigate cookie banners from multiple perspectives to analyze their prevalence and assess their impact on preventing user tracking via web cookies, based on interactions under different configuration settings. It is important to note that web tracking can also occur through the active collection of user data on online services themselves. For example, when using social media platforms such as Instagram or Twitter, users openly share personal data in ways that can (and often do) facilitate tracking [28, 49, 156, 162]. However, this is not the focus of this study, as we specifically examine tracking mechanisms that operate without the user's explicit active engagement, particularly those related to user consent obtained via banners.

1.2 The Evolution of Privacy

Privacy, once a passive byproduct of localized human interactions, has become a pressing societal necessity in the digital age. Throughout history, humans have actively sought to safeguard sensitive information—a pursuit that predates modern technology. Accordingly, the scope and generic definition of privacy have been evolving as notions of "private" and "sensitive" shift with the collective awareness of each era. Early societies, in their pursuit of security and control, developed strategies to manage access to vital knowledge. For instance, medieval guilds carefully guarded trade secrets, and early postal systems introduced measures to protect private correspondence from interception. In contrast, in today's digital society, privacy is no longer merely a privileged virtue or a protective measure against adversaries; it is increasingly recognized as a fundamental human right and a core societal value.

Looking back, the Industrial Revolution marked a turning point. Before technologies like photography, privacy was mostly intact, as interactions were fleeting—conversations faded with time, and observations stayed within the immediate audience. The invention of the camera disrupted this dynamic by transforming fleeting moments into permanent records. Photography and videography intensified the need to actively protect privacy, as individuals' actions can now be documented, replicated, and disseminated beyond their control on an unprecedented scale. This shift foreshadowed how technological advancements can reshape the scope of privacy and erode the boundaries between public and private life.

With the rise of the Internet, this erosion has intensified. Social media and widespread data collection networks now track, analyze, and monetize nearly every human action, often without individuals' knowledge or explicit consent. Unlike earlier eras, when relocating or altering social contexts can *reset* one's privacy, digital footprints are now persistent and nearly inescapable. Every online interaction—whether a statement, a purchase, or a casual remark—can have long-term implications beyond the control of the individual. As a result, the psychological freedom of "mental purification," a cornerstone of inner well-being and healthy social interactions, has diminished in an era dominated by permanent digital records.

The emergence of generative AI has further complicated this dynamic. Synthetic images, videos, and audio undermine the reliability and trustworthiness of digital

records, challenging traditional notions of authenticity and accountability. In addition, the development of AI models relies on vast amounts of data, incentivizing large-scale data collection to remain competitive in the market. This, in turn, increases the risk of exposing private or sensitive information to fuel model training. However, paradoxically, this same technological shift creates new opportunities for privacy. As digital records become less inherently reliable, individuals may gain plausible deniability regarding digital records. In a (virtual) world where reality itself can be altered, the ability to verify truth weakens, potentially pushing society back to a time when something was generally accepted as fact only if perceived firsthand. Nevertheless, plausible deniability alone does not provide a comprehensive foundation for preserving privacy. First, personal data goes beyond auditory and visual records to include online transactions, interactions, and any form of digital footprint as far as online privacy is concerned. More importantly, privacy extends beyond the mere control of data sharing; it encompasses individual autonomy and the fundamental right to determine how one's data is collected and used, regardless of how it may be interpreted in the future.

Ultimately, privacy transcends individual autonomy. It safeguards creativity, self-expression, and authentic human connection by preserving spaces free from intrusion and manipulation. As collective awareness grows, societies that uphold privacy will foster resilience—empowering individuals to innovate, collaborate, and hold institutions accountable, forming a foundation for long-term prosperity. Conversely, neglecting privacy risks normalizing exploitation, eroding trust, and stifling progress, potentially leading to societal decline. This means privacy is no longer merely a personal right but a societal imperative. Addressing this challenge requires a dual approach. The first priority is education—individuals must be equipped with a deep understanding of privacy, particularly its technological aspects, and be aware of how their choices and actions can protect or jeopardize it. The second priority is innovation—privacy-centric design must be embedded into digital platforms, algorithms, and governance structures. Achieving this demands conscious collaboration and a concerted effort from policymakers, technologists, and institutions.

In this light, privacy should be seen as more than just a legal requirement. It is a key part of fair and responsible technological progress. By making privacy a core value in society and ensuring its integration into new technologies, innovation and human dignity can evolve together rather than in conflict.

1.3 Thesis Goal and Contributions

The primary objective of this thesis is to analyze the current web consent mechanism, namely cookie banners, with respect to regulatory requirements. In particular, through extensive crawl-based measurements on popular websites, this study examines the prevalence of consent banners and investigates the impact of user interactions with them—focusing on how these interactions influence the deployment and transmission of web cookies under various configuration setups.

An initial challenge, however, is conducting large-scale web measurements in a manner that is fast, reproducible, and robust, while minimizing bias. To address this, we introduce *BannerClick*—a tool built on top of OpenWPM, a widely used platform for web privacy measurement. With approximately 99% accuracy in detecting and interacting with banners (*i.e.,* acceptance and rejection), *BannerClick* enables large-scale exploration of the cookie banner landscape across a diverse set of websites and regions with varying configurations (for more details see Chapter 3). This capability allows us to systematically investigate several questions regarding the prevalence and functionality of cookie banners. These questions, along with the methodologies and key findings, are outlined below:

1. **How do geographical locations and legal frameworks affect the behavior of websites regarding consent banners and tracking practices?**
   To address this question, we conduct measurements from eight geographically diverse vantage points, covering regions regulated by GDPR, CCPA, and non-regulated areas in October 2022. Using *BannerClick*, we detect and interact with banners on the Tranco Top 10k domains while continuously collecting cookies (Chapter 4). Our findings reveal that websites display cookie banners on **47%** of sites when accessed from GDPR-regulated regions, compared to less than **30%** in non-EU regions. Additionally, websites in non-GDPR regions set many more tracking cookies than those in GDPR-regulated areas. Before user interaction, **80%** of websites in GDPR regions set at most **4 tracking cookies**, whereas in non-GDPR regions, the number rises to **40**—a tenfold increase.

2. **Do consent banners provide users with easily accessible opt-out options?**
   We observe that among all websites with detected banners, only around **50%** offer easily rejectable banners (*i.e.,* either through a direct reject button on the main view or within the banner's settings), whereas more than **80%** display an accept option. This indicates a bias toward presenting acceptance options more prominently than rejection options. This observation remains consistent across different vantage points.

3. **Do consent banners respect user choices regarding setting cookies?**
   By interacting with banners in a stateless manner, we find that, on average (median), websites set zero tracking cookies when users either reject the banner or take no action. Whereas, this number increases to **3** after accepting the banner. Nevertheless, considering the third quartile, websites set **2** tracking cookies even in rejection and no-interaction modes. While the former highlights the potential impact of banner interaction on cookie deployment, the latter indicates that non-consensual tracking remains a persistent issue.

4. **What are the characteristics and distribution of cookie paywalls, and how different are they compared to regular banners?**
   We conduct an extensive measurement using the top 10k CrUX country-wise top list in March 2023 to investigate the prevalence of cookie paywalls (Chapter 5). Our analysis of 45k websites reveals that **0.6%** implement cookie paywalls,

with the highest prevalence in GDPR-regulated regions. In Germany, **8.5%** of top 1k websites display such paywalls. Subscription costs associated with these paywalls range up to 9 euro per month, with the majority priced at **3 euro**. Additionally, websites employing cookie paywalls send, on average, **42 times more tracking cookies** compared to those using regular banners.

5. **How effective is the current consent mechanism? Does it align with privacy regulatory requirements?**
   By conducting extensive measurements on the Tranco Top 50k domains in a **stateful** manner in July 2024, we demonstrate how seemingly opposing decisions on one website can influence subsequent visits, leading to the transmission of cookies to trackers and ultimately undermining the overall effectiveness of banners (Chapter 6). In particular, we show that **intractable cookies**—tracking cookies set upon banner acceptance—are **sent by around 50% of websites before any interaction with their rejectable banners** during subsequent user visits.

Finally, we examine how the misalignment between current technical infrastructure and privacy regulations has led to the widespread adoption of poorly implemented, omnipresent cookie banners as the dominant consent mechanism. Furthermore, we explore how a more structured approach can enhance both privacy-preserving and user-informing objectives while also addressing the needs of publishers, who rely on advertising as a key revenue source.

1.4 Thesis Structure

The structure of this thesis is designed to address the aforementioned questions systematically. The following is an overview of the thesis structure and the key contributions of each chapter:

- **Chapter 2:** This chapter introduces the necessary terminology and technical **background** for understanding web tracking and privacy regulations. It begins by describing various types of web cookies, including first-party, third-party, and tracking cookies, followed by an overview of online tracking mechanisms. The chapter then introduces relevant privacy laws, such as GDPR and CCPA. Additionally, it provides an overview of cookie consent mechanisms, including Consent Management Platforms (CMPs) and cookie paywalls.
- **Chapter 3:** This chapter introduces ***BannerClick***, a tool built on top of OpenWPM that accurately detects and interacts with cookie banners. It discusses the technical implementation, evaluates its performance, and compares it with existing tools such as *Priv-Accept*. Additionally, the chapter provides documentation on how to use and configure *BannerClick*.
- **Chapter 4:** This chapter utilizes *BannerClick* to conduct a **multi-perspective analysis of web cookies** through an extensive measurement study on the

Tranco Top 10k domains. The results section explores the influence of geographical factors, compliance with user choices, the role of user agents, and the impact of landing pages versus inner pages on cookie deployment in a stateless manner as well as the consistency analysis. The findings are part of a paper published at the Passive and Active Measurement Conference 2023 [120].

- **Chapter 5:** This chapter examines the usability and implications of **cookie paywalls**, a novel type of consent banner that forces users to either accept tracking or pay for a tracking-free experience. It investigates the distribution of cookie paywalls, their relationship with subscription pricing, and potential methods to bypass them. Results from this chapter are part of a paper published at the Internet Measurement Conference 2023 [119].
- **Chapter 6:** This chapter investigates the persistence of tracking cookies during stateful interactions with banners. It introduces **intractable cookies**—cookies set in accepted domains and transmitted before user consent on subsequent websites with rejectable banners—and examines their prevalence and behavior based on factors such as website rank and banner type. Additionally, it explores their manageability and potential solutions to mitigate their occurrence.
- **Chapter 7:** This chapter **concludes** the thesis by summarizing the main findings and highlighting its contributions to web privacy research. It then discusses **final thoughts**, synthesizing insights from the preceding chapters, including the effectiveness of current consent mechanisms, misalignments between technical practices and regulatory goals, and potential future directions for web consent mechanisms and privacy-preserving advertising infrastructure.

2
Background

In this chapter, we provide an overview of web tracking techniques, including cookies and browser fingerprinting, their role in user tracking, and regulatory frameworks such as GDPR and CCPA designed to mitigate privacy risks. Additionally, we examine the adoption of consent banners and Consent Management Platforms (CMPs), highlighting challenges such as deceptive patterns and cookie paywalls.

2.1 Web Tracking

Web tracking refers to techniques used by websites and third parties to monitor, collect, and analyze users' online activities. Various methods can be employed to track users and build a profile of their behavior, including web cookies, browser fingerprinting, URL tracking, tracking pixels, and supercookies [1, 16, 35, 36, 66, 124]. While web tracking is commonly used to personalize content, deliver targeted advertising, and enhance user experiences, it also raises significant privacy concerns [84, 93].

2.1.1 Web Cookies

Web cookies are small pieces of data stored on a user's device by websites to enhance user experience and support essential website functionality [8, 77]. Introduced in the 1990s, cookies were designed to enable websites to maintain session state in stateless HTTP environments. They can be created or modified using JavaScript functions or via the `Set-Cookie` attribute in the HTTP response header. Once stored, cookies are automatically sent back to the server with subsequent requests through the `Cookie` request header, allowing the server to remember the user's actions and preferences across sessions. Cookies are generally classified into two types: **first-party cookies** and **third-party cookies**. First-party cookies are set directly by the website the user is visiting, while third-party cookies are set by external domains, often as a result of loading third-party resources embedded on the website.

Though cookies were originally designed to enhance user experiences by enabling features such as session management and personalization, they have also become a core technology for tracking users on the web. Third-party cookies, in particular, allow advertisers and analytics providers to monitor users across multiple websites, build detailed behavioral profiles, and support targeted advertising. Those specifically used

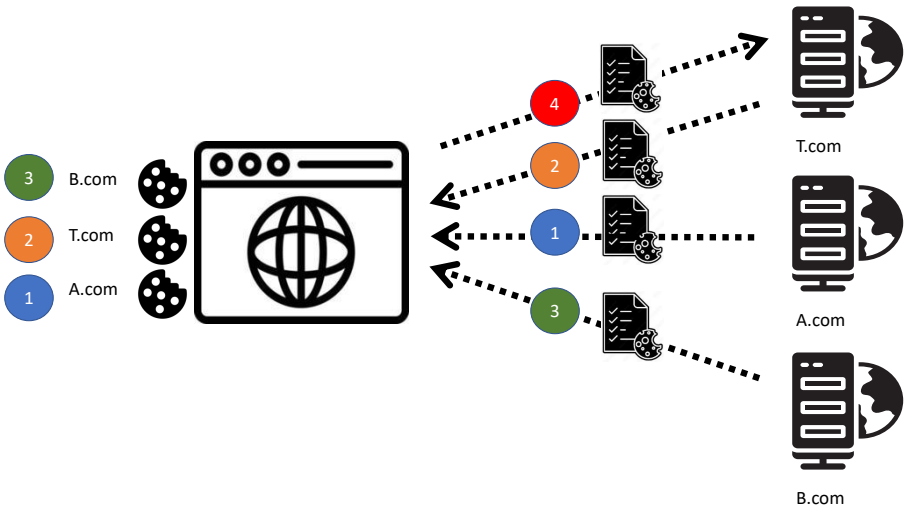


Figure 2.1: An illustration of different types of cookies along with how third-party cookies can be used to track users across different websites.

by tracker domains to monitor user behavior are referred to as **tracking cookies**. These cookies store information that can identify users and associate their activity with the pages they visit and the ads they click. Common data stored in tracking cookies include Unique User Identifiers (UUIDs), Page IDs, and Ad IDs.

Figure 2.1 illustrates different types of cookies and how third-party cookies can be used to track users' online activities. Initially, when a user visits `A.com`, the website sets a first-party cookie via an HTTP response. Along with its own resources, `A.com` includes a third-party resource from `T.com`, which results in a third-party cookie being set by `T.com` on the user's device (step *2*). Later, when the user navigates to `B.com`, the website sets a first-party cookie related to its domain (step *3*). Similar to `A.com`, `B.com` also includes resources from `T.com`. Accordingly, the browser sends a request to `T.com`, which includes the third-party cookie previously set by `T.com` (step *4*). This mechanism enables `T.com` to identify the user as the same individual who visited `A.com`. By collecting and aggregating such information at its backend, `T.com` can build a detailed profile of the user's browsing history across websites like `A.com`, `B.com`, and any other site that loads its resources.

2.1.2 Fingerprinting

Besides cookies, other techniques are also used to track users on the web. **Browser fingerprinting** [81] is one of the most prominent tracking methods, relying on the collection of unique characteristics of a users browser and device. This technique gathers technical details such as screen resolution, installed fonts, browser type, operating system, and plugins. In the case of **device fingerprinting** [36], these capabilities are extended further by incorporating hardware-level attributes, such as the GPU and

CPU. By combining these attributes, a unique "fingerprint" is generated, allowing websites to identify and track users across sessions and websites.

Unlike cookies, fingerprinting does not store any data on the user's device. Instead, it passively collects information through JavaScript APIs during normal browsing activities. Therefore, mitigating browser fingerprinting is more challenging, as it relies on preventing the collection of unique attributes rather than blocking stored data. Privacy-focused browsers, such as Brave and Firefox with Enhanced Tracking Protection (ETP), employ measures like fingerprint randomization and uniform browser configurations to reduce fingerprint uniqueness. Additionally, privacy tools can limit the exposure of APIs commonly used for fingerprinting.

2.2 Privacy and Data Protection Laws

As digital technologies become increasingly integrated into everyday life, privacy has emerged as a fundamental concern in modern society. The widespread collection and use of personal data by online services have heightened concerns about user autonomy and data security. However, the lack of clear communication about data practices often leaves users unaware of how their information is collected, processed, and shared, underscoring the urgent need for robust privacy safeguards.

To address these concerns, governments have introduced regulatory frameworks aimed at protecting user privacy and ensuring transparency in data handling. Among the most notable are the General Data Protection Regulation (GDPR) [39] in the European Union and the California Consumer Privacy Act (CCPA) [19] in California. These regulations mandate that organizations disclose their data practices, empowering users with the knowledge to make informed decisions about their privacy.

A central pillar of these regulations is the emphasis on user consent. Online services must obtain explicit consent before collecting sensitive data or engaging in tracking activities. This requirement has led to the widespread adoption of mechanisms such as cookie banners and privacy notices, which enable users to accept, reject, or customize data collection preferences. By enforcing these practices, privacy regulations aim to align data collection with user expectations and ethical standards.

Despite these efforts, implementing privacy regulations remains an ongoing challenge. Many provisions and requirements are not fully compatible with the current structure of web technologies, creating gaps in enforcement and compliance. Furthermore, the complex ecosystem of online tracking—where multiple entities contribute to data collection and processing—complicates efforts to ensure transparency and enable proper user control. As digital ecosystems continue to evolve, bridging the gap between regulatory intentions and technological realities will be crucial to fostering a truly privacy-conscious digital landscape.

2.2.1 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) [39], implemented by the European Union in May 2018, is one of the most comprehensive legal frameworks designed to protect user privacy and ensure transparency in data handling. It grants users greater control over their personal data while imposing strict obligations on organizations that collect, process, and store such data. The strict requirement of GDPR has significantly influenced how companies worldwide approach data privacy.

GDPR applies not only to residents of EU countries but also to any organization worldwide that processes data originating from the EU. This extraterritorial scope means that organizations handling such data must comply with GDPR requirements, regardless of their physical location, to uphold the privacy and security of EU residents' personal information (Articles 1, 2, and 3).

A core principle of GDPR is the requirement for organizations to obtain explicit and informed user consent before collecting or processing personal data. In the context of online tracking, this has direct implications for technologies such as cookies and browser fingerprinting. GDPR mandates the following:

- *Consent must be freely given* (Article 4(11), Article 7(4)[1]). Users must have the choice to accept or reject tracking without coercion or negative consequences. For example, access to website content cannot be made conditional on accepting all cookies unless strictly necessary for functionality.
- *Consent must be informed* (Articles 12, 13, 14). Websites must provide clear and specific information about the data being collected, its purpose, and the entities involved in processing it. This is typically communicated through cookie banners or privacy notices.
- *Consent must be specific* (Article 7(2)). Users must be able to grant consent for certain types of cookies (e.g., analytics or advertising) while rejecting others. Blanket consents, such as "accept all cookies," are discouraged unless accompanied by more granular options.
- *Consent must be revocable* (Article 7(3)). Users must have the ability to withdraw consent at any time, and organizations are required to ensure this process is as simple as granting consent.

Beyond consent, GDPR also imposes restrictions on tracking technologies by enforcing the following principles:

- *Data minimization* (Article 5(1)). Only the data strictly necessary for the stated purpose should be collected and processed.
- *Purpose limitation* (Article 5(1)). Data collected for one purpose cannot be repurposed without obtaining additional user consent.

[1] Clicking on each Article or Recital of the GDPR mentioned in the text redirects to the corresponding GDPR Provision in the Appendix A.

- *Accountability and documentation* (Articles 5(2), 24, 30). Organizations must maintain records of user consents and demonstrate compliance with GDPR requirements in case of audits.

The implementation of GDPR has led to the widespread adoption of cookie banners and Consent Management Platforms (CMPs), designed to inform users about the use of cookies and collect their preferences regarding tracking. In addition, previous studies suggest that GDPR has been a pivotal step toward creating a more privacy-preserving online environment [30, 31, 76, 137]. By promoting transparency and granting users control over their personal data, it has reshaped the landscape of online tracking. However, despite its stringent requirements, GDPR faces challenges in enforcement and interpretation. Many organizations deploy cookie banners that technically comply with the regulation but use deceptive designs to nudge users toward accepting tracking. Furthermore, the absence of a universal standard for banners results in inconsistent design and disparities in compliance levels. We discuss specific GDPR provisions in detail in Chapter 7.

2.2.2 California Consumer Privacy Act (CCPA)

The California Consumer Privacy Act (CCPA) [19], enacted in June 2018 and effective from January 2020, is a privacy law designed to protect the personal data of California residents. It grants consumers enhanced rights over their personal information and imposes obligations on businesses operating in California or handling the data of California residents.

CCPA applies to organizations that meet specific thresholds, such as having annual gross revenues exceeding $25 million, collecting data on 50,000 or more consumers, households, or devices, or generating 50% or more of their annual revenue from selling consumer data. Unlike GDPR, which focuses on explicit user consent, CCPA emphasizes transparency and grants consumers the right to opt out of the sale of their personal data. In fact, in contrast to GDPR's opt-in model, CCPA allows tracking of users by default unless they explicitly opt out, following an opt-out model instead.

Under CCPA, businesses are required to adhere to the following key principles:

- *Right to Know* (Section 1798.110). Consumers have the right to request information about the categories and specific pieces of personal data collected, the purpose of collection, and any third parties with whom the data is shared.
- *Right to Delete* (Section 1798.105). Consumers can request the deletion of their personal information collected by a business, subject to certain exceptions (e.g., data necessary to complete a transaction or comply with legal obligations).
- *Right to Opt-Out* (Section 1798.120). Consumers have the right to opt-out of the sale of their personal data by businesses. To facilitate this, businesses must provide a conspicuous "Do Not Sell My Personal Information" link on their websites.

- *Right to Non-Discrimination* (Section 1798.125). Businesses cannot discriminate against consumers who exercise their CCPA rights, such as by denying services, charging higher prices, or offering lower-quality goods.

In the context of online tracking, the CCPA primarily addresses the sale and sharing of personal data through mechanisms such as cookies and trackers. However, its enforcement presents several challenges, including ensuring compliance among small and medium-sized businesses and clarifying ambiguities in the definition of "sale" of data. Additionally, some businesses implement opt-out mechanisms that are difficult to locate or understand, undermining the regulations intent.

Despite these challenges, the CCPA represents a clear step toward strengthening privacy protections for California residents. It complements global privacy frameworks such as GDPR by prioritizing consumer rights and promoting transparency, even though its approach to privacy differs in several key aspects.

2.2.3 Other Privacy Regulations

Beyond GDPR and CCPA, various other privacy regulations have emerged globally to address the growing concerns about data privacy and online tracking. These regulations share common goals of enhancing transparency and user control over personal data but differ in their scope and enforcement mechanisms. Some notable examples include:

- *Brazil's General Data Protection Law (LGPD)* [74]. Enacted in 2020, LGPD is Brazils counterpart to GDPR, applying to data processing activities involving Brazilian residents. While similar to GDPR in many aspects, LGPD includes additional provisions emphasizing the role of a data protection officer (DPO) and imposes less stringent penalties for non-compliance.

- *Australia's Privacy Act 1988* [7]. Australias privacy law focuses on protecting personal information through principles like transparency and data minimization. However, unlike GDPR, it applies primarily to organizations with an annual turnover exceeding AUD 3 million, exempting smaller entities.

- *China's Personal Information Protection Law (PIPL)* [144]. Introduced in 2021, PIPL draws inspiration from GDPR but includes unique provisions tailored to Chinas legal and cultural context. It emphasizes data localization, requiring that certain types of personal data collected in China be stored within the country.

- *Japan's Act on the Protection of Personal Information (APPI)* [55]. Revised in 2022, APPI aligns closely with GDPR but differs in its approach to cross-border data transfers, relying heavily on mutual agreements between countries and sectors.

There are key differences between these regulations, often stemming from their cultural and economic contexts. For instance, GDPR and PIPL emphasize data localization and international data transfers, while CCPA prioritizes transparency in large businesses and markets. Additionally, regulations such as Australia's Privacy Act have narrower scopes of applicability, often excluding smaller organizations or certain data processing activities. Despite these differences, all these frameworks share the overarching goal of protecting individuals' personal data and ensuring accountability in data processing practices.

2.3 Consent Mechanisms

In this thesis, in the context of the web and online tracking, consent mechanisms refer to the entire software-based infrastructure designed to (1) inform users about data collection practices and their rights over their data, (2) provide a user interface that allows them to manage their preferences, and (3) ensure that their choices are correctly adopted and implemented by websites and related third parties.

Beyond users themselves, multiple entities are directly or indirectly involved in the consent mechanism infrastructure. These include website operators, third-party vendors, and advertisers. Additionally, third-party services have emerged to provide ready-to-use advertising and analytics solutions. One example is Data Management Platforms (DMP) [48], which are software systems that facilitate the collection, storage, and organization of data for use in marketing, publishing, and business analytics. The data stored in a DMP may include customer information, demographics, mobile identifiers, and cookie IDs. By analyzing this data, DMPs enable businesses to create targeted advertising segments, refine acquisition strategies, and optimize sales.

In addition, there has been a rise in service providers offering partial or full consent mechanisms as a service. Consent Management Platforms (CMP) and Subscription Management Platforms (SMP) are examples of these services. More details on CMPs are provided in Section 2.3.2, and SMPs are discussed in Chapter 5.

2.3.1 Consent Banners

Consent banners, also referred to as cookie banners, cookie notices, or consent notices, serve as the user interface for the consent mechanism infrastructure. Their primary function is to inform users about data collection practices and obtain their consent for such activities. These banners emerged as a direct response to privacy regulations, particularly the General Data Protection Regulation (GDPR), which mandates obtaining clear and informed user consent before collecting or processing personal data.

Typically, consent banners appear when a website loads, notifying users of potential data collection practices during their interaction with the site. These practices vary across websites and may include data collection for analytics, advertising, or

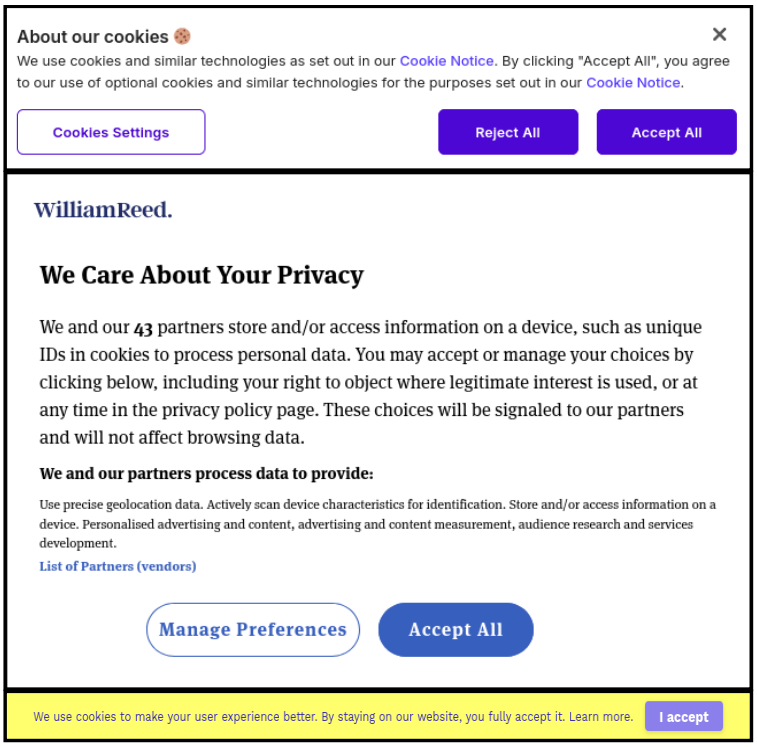


Figure 2.2: Examples of consent banners: from top to bottom, a banner with accept, reject, and settings options; a banner with accept and settings options but no direct reject option; and a banner with only an accept option.

other purposes. The banners also allow users to configure their preferences, such as accepting all cookies, rejecting non-essential cookies, or selecting specific vendors or categories of cookies to enable.

The design and functionality of consent banners vary widely. Some banners provide straightforward options, such as "Accept" and "Reject," while others offer more granular controls, allowing users to manage different types of data processing individually. Additionally, the appearance of banners—including layout, sizing, wording, and interaction flow—differs significantly across websites.

Figure 2.2 illustrates three common consent banner designs, highlighting the variability in user options. The first banner provides explicit "Accept," "Reject," and "Settings" options, offering users a balanced and transparent interface for managing their preferences. The second banner includes "Accept" and "Settings" options but lacks a direct "Reject" button, requiring users to navigate additional steps to decline tracking. The third banner only offers an "Accept" option, significantly limiting user control and potentially coercing consent. Notably, as indicated in the banner's text, the website begins using cookies immediately upon entry, without explicit user permission. These examples demonstrate how banner designs vary across websites, influencing user interactions and the effectiveness of the consent process.

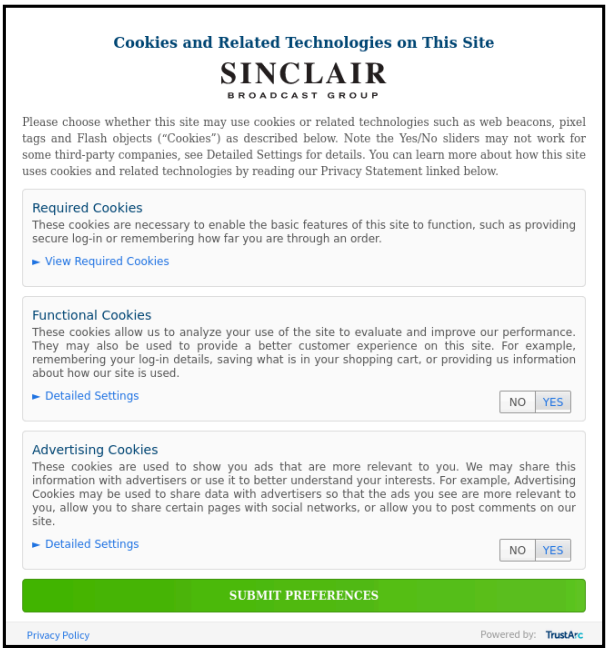


Figure 2.3: An example of a CMP banner with granular consent options, including toggles for different types of cookies and tracking. Users can also selectively accept or reject certain vendors.

Overall, consent banners have become a ubiquitous element of web browsing, serving as a key interface for compliance with privacy regulations. However, their effectiveness in ensuring informed consent and protecting user privacy largely depends on their design and implementation. Poorly designed banners, such as those employing deceptive patterns or obscuring reject options, can undermine the intent of privacy regulations, leading to user frustration and unintentional consent.

2.3.2 Consent Management Platforms (CMPs)

To streamline compliance with privacy regulations, many websites now rely on third-party software solutions for managing user consent. While some still implement custom mechanisms, an increasing number are outsourcing this task to Consent Management Platforms (CMPs).

CMPs help websites comply with regulations such as GDPR and CCPA by providing configurable solutions for displaying consent banners, capturing user preferences, and enforcing compliance within data-handling policies.

Standardization and Functionality of CMPs: CMPs address inconsistencies in consent management by offering a standardized framework for collecting, storing, and sharing user preferences with advertisers, analytics providers, and other third-party

vendors. As shown in Figure 2.3, they often provide granular options for managing consent, enhancing transparency and user control.

Technical Components: CMPs integrate with websites through the following key components:

- *Consent Banners*: Customizable interfaces that allow users to accept, reject, or configure consent preferences for different types of data collection.
- *Consent Strings*: Encoded user choices stored and shared with vendors to ensure compliance.
- *Vendor Management*: A registry of data-processing vendors, ensuring only those with proper user consent can access data.
- *Communication with Vendors*: Standardized APIs relay consent preferences, ensuring vendors comply with regulations.

Transparency and Consent Framework (TCF): [62] The Interactive Advertising Bureau (IAB) Europe developed the TCF to standardize consent management across the digital advertising ecosystem. TCF establishes protocols for CMPs, publishers, and vendors to ensure GDPR compliance. Key components include:

- *Global Vendor List (GVL)*: A registry of vendors declaring data-processing purposes and legal bases.
- *Consent String Encoding*: A standardized format ensuring consistent interpretation of user preferences.
- *Standardized APIs*: Interfaces facilitating interoperability between CMPs and vendors.

Advantages and Challenges of CMPs: CMPs enhance compliance by simplifying consent management and providing a uniform experience for users. However, challenges remain. Many implementations employ deceptive patterns, such as preselected consent options or obscured reject buttons, undermining informed consent. Additionally, discrepancies in vendor interpretations of consent strings create compliance gaps. Moreover, the complexity of granular consent options can lead to consent fatigue, reducing user engagement with privacy settings.

For CMPs to be truly effective, continuous improvements in technology and governance are needed. Refining frameworks like TCF, enforcing stricter guidelines against misleading implementations, and improving user-friendly designs can help ensure that CMPs protect privacy while reducing friction in the consent process.

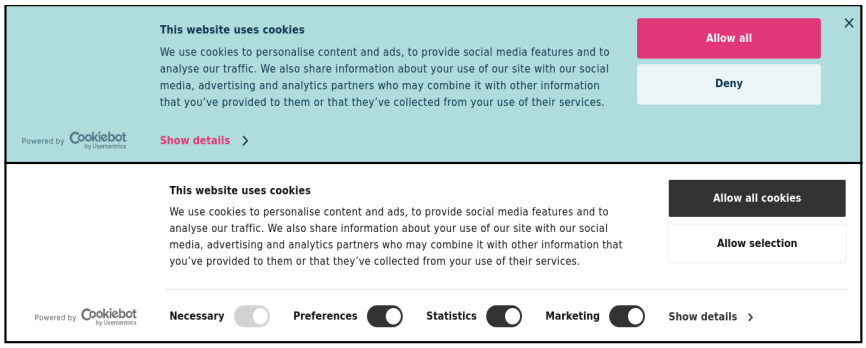


Figure 2.4: Examples of consent banners with deceptive patterns: from top to bottom, a banner with a highlighted "accept" button and a dimmed "deny" option; and a banner with preselected preferences for nonessential cookies.

2.3.3 Deceptive Patterns

As mentioned, the primary goal of tracking technologies like web cookies is monetization, particularly through targeted advertising. These strong incentives may tempt publishers to exploit techniques that maximize user tracking while still appearing to comply with privacy regulations. Some of these techniques, such as *deceptive patterns*, are incorporated at the first stage of the consent mechanism, where users encounter consent banners and decide on their privacy preferences.

Deceptive patterns, a.k.a. "dark patterns," are manipulative user interface designs that influence users to make choices favorable to websites or advertisers, often at the expense of user privacy. These patterns exploit psychological biases or confusion, steering users toward unintended decisions. Two common examples include:

- *Obscuring the "reject" button*, making it harder to decline non-essential cookies or similarly *highlighting the "accept" option* with bold colors or prominent placement.
- *Pre-selected options* to accept all nonessential cookies.

Figure 2.4 shows two consent banners from the same CMP, Cookiebot, highlighting inconsistencies in design (even within a single CMP) and the use of deceptive patterns. The first banner uses a highlighted "accept" button and a dimmed "deny" option, visually nudging users towards accepting all cookies. The second banner takes this further by not only highlighting the "accept" option but also preselecting nonessential cookie preferences by default. This combination biases user decisions, discouraging deliberate and informed choices. These examples illustrate that CMPs, as de facto standards for consent banners, sometimes fail to adhere to good design practices, compromising the spirit of privacy regulations like GDPR.

Overall, deceptive patterns undermine the principle of informed consent by biasing user decisions, making them impulsive or habitual rather than deliberate. Over time, such designs can erode trust in consent banners, causing users to perceive them as intrusive rather than protective of privacy.

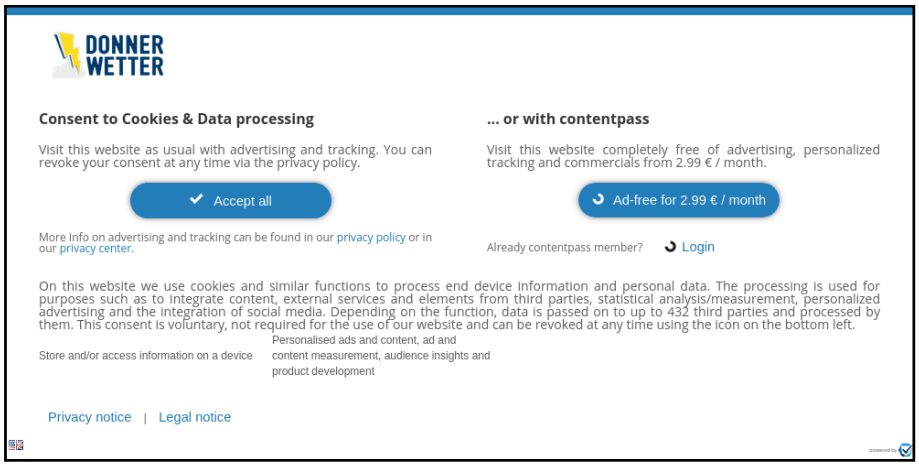


Figure 2.5: An example of a cookie paywall: a type of banner that forces users to either accept tracking or pay for a tracking-free experience.

2.3.4 Cookie Paywalls

Beyond deceptive patterns, cookie paywalls represent another type of banner that can influence user behavior by imposing a more restrictive form of consent [119]. These banners force users to choose between accepting tracking cookies or subscribing to a paid, ad-free experience, without offering a freely available reject option. While they may technically comply with GDPRs consent requirements, cookie paywalls raise concerns about fairness and accessibility. Figure 6.10b illustrates an example of a cookie paywall, where users must either accept tracking cookies or pay 2.99 per month for a tracking-free experience.

Legality Under GDPR: Under GDPR, *consent must be freely given, informed, and unambiguous* (Article 4(8)). While cookie paywalls technically offer a choice, they may violate the freely given principle by coercing users into accepting tracking in exchange for access to content. GDPR Recital 43 warns against conditioning consent on access to services when the data processing is not essential for their functionality—a practice commonly associated with cookie paywalls. This ambiguity raises questions about their legality and compliance with the current provisions of GDPR.

Ethical Concerns: Cookie paywalls create a privacy divide, where only paying users can avoid tracking. This may disproportionately impact low-income users, compelling them to trade their privacy for access to content. However, ethical judgment is to some extent subjective, especially when financial interests are involved.

3

BannerClick

Analyzing the prevalence and behavior of cookie banners—the dominant web consent mechanism—requires a systematic and large-scale measurement approach. This includes examining what options banners provide to users and how websites respond to user interactions, particularly with regard to regulatory compliance.

To conduct such research at scale and effectively analyze the interplay between consent banners, web cookies, and privacy regulations, an appropriate measurement tool needs to meet several key specifications:

- *Automation*: A fully automated system is essential for conducting large-scale measurements. Automation ensures reproducibility for longitudinal analyses, enables consistency checks to validate results, and supports multi-vantage point measurements across diverse geographic regions. Most importantly, it significantly reduces the time required for data collection and analysis.
- *Performance*: The tool must be both fast and accurate. High speed enables broader measurement coverage and, consequently, more comprehensive analysis. Accuracy, on the other hand, prevents biases that may arise from failures—such as incorrectly detecting or interacting with specific banners—thereby ensuring reliable and meaningful results.
- *Data Recording*: To support further analysis and validation, the tool should log all relevant data, including detected banners, user interactions, and the cookies set as a result. This ensures that no critical information is overlooked and enables broader, retrospective analysis.
- *Scalability*: Given the dynamic nature of consent banners, the tool must be adaptable and extensible. It should accommodate new banner types with minimal effort to remain effective in an evolving web environment.

This chapter introduces *BannerClick*, a tool specifically designed to automatically detect and interact with consent banners with 99% accuracy. Built on the Selenium browser automation framework and integrated with OpenWPM, a widely used privacy measurement platform, *BannerClick* enables researchers to conduct large-scale, reproducible, and reliable studies on consent banners, their impact on deployment and transmission of cookies, and compliance with privacy regulations. We detail its design, implementation, and key features, demonstrating how it meets the critical requirements for effective privacy measurement.

The remainder of this chapter is structured as follows: First, we describe how *BannerClick* detects banners on websites. Next, we discuss how *BannerClick* interacts with banners by mimicking user actions such as accepting or rejecting cookies. We then elaborate on its Consent Management Provider (CMP) detection module. Afterward, we evaluate its accuracy and compare its performance with that of an existing tool, Priv-Accept [147]. Lastly, we provide technical details about the implementation of *BannerClick* and include a brief guide on how to configure and use it.

3.1 Banner Detection

The first step *BannerClick* must perform is detecting the presence of a cookie banner on a website. To design this functionality effectively, we first examine the structure of a banner within an HTML page. Typically, as shown in Figure 3.1, banners are separate direct children of the `<body>` tag and contain cookie-related words. *BannerClick* sequentially attempts to locate three specific nodes to pinpoint the banner on a given webpage.

To detect banners, we first construct a corpus of English words that commonly appear in banners by manually inspecting 50 randomly selected websites from the Tranco top-100 domains. This corpus consists of eight unique English words: *cookies, privacy, policy, consent, accept, agree, personalized,* and *legitimate interest.* We then translate these words into 11 different language—German, Swedish, Spanish, Italian, Portuguese, Chinese, Russian, Japanese, French, Turkish, and Persian—expanding the corpus size to 80 words. We later demonstrate that using this corpus enables us to achieve an accuracy of approximately 99% in banner detection.

On a website's HTML page, *BannerClick* first searches for all elements containing a word from our corpus. For example, the `<p>` element (highlighted in blue in Figure 3.1) contains a banner-related word. In addition, if an element contains words from our corpus and meets any of the following conditions, we discard it and proceed to the next element:

1. If the element is set as *invisible*, the banner is not visible to users, and they cannot interact with it.
2. An element with a *negative z-index* is positioned behind other objects on the page. Since a banner must be on top of all elements to remain visible, such elements cannot contain a valid banner.
3. The banner must be within the user's visible area of the webpage. For example, an element located in the footer (i.e., *outside the viewport*) or inside a table is not a valid banner candidate.

Next, it traverses up in the DOM hierarchy towards the HTML element that has either a *positive z-index* or a *fixed position* attribute. Generally, cookie banners are either displayed on top of the webpage content (positive z-index) or maintain the same position on the webpage (fixed position). The element with these properties very

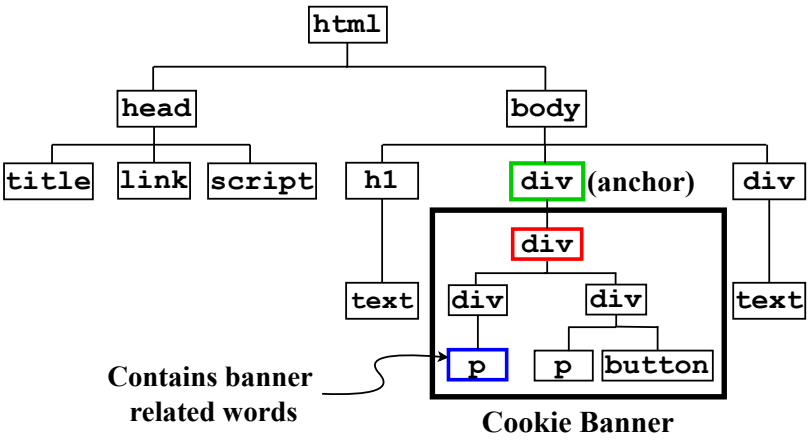


Figure 3.1: An HTML Document Object Model (DOM) containing a banner.

likely contains the banner. We call this the "anchor" element (see the green `<div>` element in Figure 3.1). If *BannerClick* fails to find such an element, it considers the `<body>` element as the div anchor element.

The anchor element contains the banner, but the banner may still be fully contained within some sub-element of the anchor. To find this most-specific element, *BannerClick* traverses down again in the DOM, starting from the anchor element. It uses the following heuristic: the visible elements contained inside the anchor (*e.g.,* banner title, description, and buttons) should also be contained entirely within the more-specific candidate element. Following this heuristic, *BannerClick* continues traversing down the DOM tree until it finds an element that does not completely contain all visible banner elements anymore. This implies that the parent of this element is the most-specific banner-containing element. This is shown as the red `<div>` element in the DOM tree in Figure 3.1.

Some websites might include banners as `iframes`, which are outside the regular website's DOM. In cases where *BannerClick* fails to locate the desired element that contains the banner, it specifically iterates over all visible `iframes`. The above steps are once again repeated inside each `iframe` to detect the banner.

3.2 Banner Interaction

After successfully detecting a banner, *BannerClick* can also interact with it. It can both "accept" and "reject" cookies in an automated manner. To do so, it relies on a corpus of words that are frequently used in cookie banners to indicate acceptance or rejection of cookies. This corpus consists of three categories of words indicating "accept", "reject", and "settings".

To assemble this corpus of words, we access the Tranco top-10K websites and detect the banners on them. We proceeded with those Tranco websites, for which we successfully detect the banner. Next, we identify the language of each of these websites

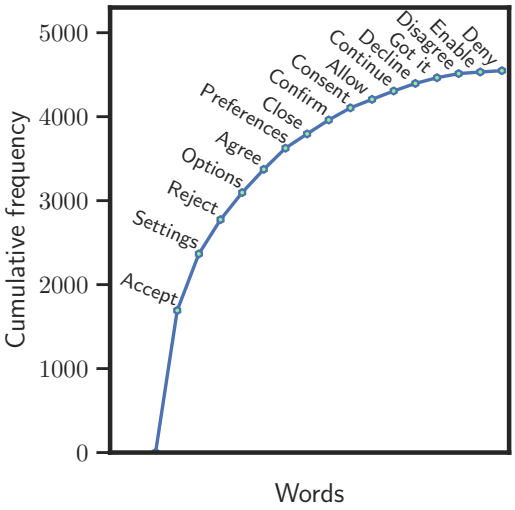


Figure 3.2: English words appearing at differing frequencies inside the buttons of cookie banners.

using Google's `cld3` library [51][1]. We observe that 4215 of these websites are in 12 languages; English alone is the language of more than 77% of those.

To detect commonly used words in a given language, we adopt a simple approach. First, we select all banners in English and identify elements that are likely buttons, such as `<button>` elements or those with button-related classes or tags, such as "btn". Next, we extract the associated words within these elements and count their frequency. We then filter out words that appear in at least 1% of the banners. Figure 3.2 provides examples of such words, including "Accept", "Settings", "Reject", "Options", and "Agree".

For non-English languages (*e.g.,* German), we repeat the same process, but we additionally translate each of these words to English. We then manually check if they are semantically similar to any one of the following three categories: accept, reject, or settings. If the tested word is closer to any of these, we append the word to the appropriate category. We repeat the same process for each of the 11 non-English languages. At the end, we have 172 words in 12 different languages belonging to the three different categories.

After successfully detecting the banner and identifying these words, *BannerClick* automatically clicks the identified button. Throughout all our experiments, we use three modes to interact with websites *i.e.,* "No interaction" (we do not click any button on the banner), "Accept" (we click accept related words), and "Reject" (we click reject related words). If *BannerClick* identifies multiple such elements, it first prioritizes `<button>` elements and then selects the one with the minimum number of

[1] `cld3` at its core uses neural networks to detect the language of any given document. We manually select 20 websites belonging to 10 different languages (*i.e.,* two websites for each language). We identify the language of these websites using `cld3` library and find it to be 100% accurate.

words. For instance, in a banner, a `<p>` element may contain the text, "To accept all cookies, please click the button below", and another `<p>` element simply has the word "accept". Our tool selects the latter, as it is likely the button to provide consent[2].

To provide consent for cookies, *BannerClick* searches for words belonging to the "accept" category in the corpus. When it finds a match, it clicks on the identified element. The process of rejecting cookies is similar: *BannerClick* searches for words inside the banner belonging to the "reject" category in the corpus. However, if no such words are found, *BannerClick* attempts to reject banner policies using the Never-Consent browser extension [123]. Never-Consent searches for specific functions provided by CMPs to reject banner policies (*e.g.,* the OneTrust CMP function `OneTrust.RejectAll()`). It can also interact with banners based on predefined structures associated with different CMPs. For example, some CMPs place the reject button within an element that is consistently marked with a specific tag name.

However, if *BannerClick* still fails to reject the banner, it searches for the third category of words *i.e.,* "settings". This is because, very often, the option to reject cookies is present inside a banner's settings. On a successful match, it clicks on the element containing the "settings" word. If the click is successful, the settings dialogue opens, and *BannerClick* again searches for words belonging to the "reject" category inside this dialogue. Using this approach, *BannerClick* can successfully detect, accept, and reject banners on websites.

We further enhance *BannerClick*'s capabilities for different scenarios through continued exploration in later projects during the course of this study. In particular, we extend its capabilities in detecting cookie paywalls (see Section 5.2). Moreover, we improve its rejection accuracy to better identify banners with pre-selected options in their settings (see Section 6.3).

3.3 CMP Detection

While accessing these websites with *BannerClick*, we also analyze the distribution of Consent Management Providers (CMPs). As introduced in Section 2.3.2 CMPs are platforms that offer cookie consent handling as a service, *i.e.,* websites can include a ready-to-use, yet configurable, banner instead of developing their own cookie banner solution. The IAB Europe Transparency and Consent Framework (TCF) is a GDPR-compliant consent solution that specifies the overall behaviors of CMPs [63]. Based on TCFv2 specifications [62], all CMPs need to implement a `__tcfapi()` function that allows third parties to have access to the users' selected preferences and act accordingly. In *BannerClick* we use this function to record the name of the CMP while crawling a website. During implementations, we observe that—contrary to the specification—not all websites with CMP banners actually implement the `__tcfapi()` function. This specification violation is not limited to a specific CMP.

[2] One can simply detect the `<button>` tags and search for words inside them. However, we observe that banner buttons are not always implemented in this manner. Instead, many websites use other types of tags like `<input>` or `<div>` to implement buttons.

To obtain a better and more comprehensive distribution for CMPs, we additionally incorporate results from the Never-Consent browser add-on [123] into our data. Never-Consent leverages custom APIs, which some CMPs implement in addition or instead of `__tcfapi()`. These custom APIs allow for interaction with CMPs to fetch user-related data or can even trigger a *reject all* event.

If *BannerClick* can call the `__tcfapi()` function on a website, it uses the following command to retrieve the CMP ID:

```
__tcfapi("ping", 2, (pingReturn) =>
    {cmp_id = pingReturn.get("cmpId");});
    return cmp_id;
```

Using this CMP ID, *BannerClick* cross-references the published list of registered CMPs[3] to obtain the CMP name. If the website does not implement the `__tcfapi()` function, *BannerClick* attempts to determine the CMP name using the Never-Consent API. Finally, all records are stored in the "visits" table under the corresponding columns for further analysis. The CMP detection functionalities are implemented in the `cmpdetection.py` module.

3.4 Accuracy Analysis

As mentioned, one of the key factors *BannerClick* must fulfill to ensure reliable results is its accuracy in detecting and interacting with banners. We now evaluate *BannerClick*'s accuracy in both of these functionalities.

Efficacy of banner detection: We test our banner detection approach on the Tranco top-1k websites. We manually inspect and confirm that a total of 518 websites show banners. Using *BannerClick*, we are able to correctly detect banners on 513 websites. Therefore, only 5 websites show a banner, but *BannerClick* fails to detect them. The reasons include the presence of a shadow DOM [98] on the website (`godaddy.com`) and banners having words not present in our corpus (`washington.edu`). Similarly, only 4 websites do not show any banner, but *BannerClick* incorrectly detects a banner. For example, `allaboutcookies.org` has cookie-related words in its DOM, but does not show a banner. Overall, *BannerClick* detects banners with more than 99% accuracy and extremely low FPR (0.008) and FNR (0.009).

Efficacy of banner interaction: As previously mentioned, 518 websites of Tranco top-1K websites show a banner. We manually confirm that 444 of these offer an explicit accept option. The remaining 74 websites do not give the option to explicitly accept (*e.g.,* the banner just has a close button, or there is an implicit accept[4]). *BannerClick* does not click accept button on any of these 74 websites.

[3] `https://cmp-list.consensu.org/v2/cmp-list.json`

[4] Some websites show banners that do not overtly show the "accept" option. For instance banner on `bitly.com`, just states that "By continuing to use this site you are giving us your consent to do this".

In our research, we just consider explicit accept when interacting with banners. This is because, according to GDPR, websites must take users' consent explicitly. Later, for such websites, we quantify the increase in cookies after clicking the accept button. With *BannerClick*, we successfully click accept on 430 out of 444 banners with explicit accept. However, amongst the remaining 14, *BannerClick* clicks the incorrect button on 13 websites. The banners of these websites contain buttons with words that negate the semantic meaning of accept, *e.g.,* "NOT Accept" (which is essentially a reject). Since *BannerClick* does not consider the text's semantics, it incorrectly classifies them as the accept. Lastly, only one website shows a banner with words that we do not have in our corpus. Thus *BannerClick* failed to click the button for that single website. Overall *BannerClick* successfully clicks the button with more than 97% accuracy.

Finally, we calculate *BannerClick*'s reject accuracy by manually checking the screenshots for the Top-1k websites. *BannerClick* successfully reject banners on 377 out of 524 websites and finds that 81 banners do not provide a reject option, resulting in an accuracy of 87.4%. The majority of unsuccessful rejections come from 38 websites that use multi-select mechanisms to reject cookies.

3.5 Comparison With Existing Tools

Jha *et al.* [68] proposed the tool Priv-Accept [147], which automatically attempts to "accept" privacy policies mentioned in a banner. They create a corpus of "accept" related words and compare them with the words present in the DOM of the website. If Priv-Accept finds the accept button, clicks it and compares the website behaviour before and after the click (*e.g.,* page load time).

We compare *BannerClick* with Priv-Accept. First, Priv-Accept is unable to identify and click reject buttons. Second, unlike *BannerClick*, Priv-Accept does not detect the banners but instead inspects the complete DOM for accept-related words and, on a successful match, attempts to click the element containing the word. As a result, it can encounter multiple failures before actually clicking the desired accept button on the banner. On the contrary, *BannerClick* first detects the banner and searches for words contained within the banner. Third, *BannerClick* can click on accept related elements in 12 popular languages, whereas, Priv-Accept only searches for English words. There are other differences, *e.g., BannerClick* looks for banners within the iframes, but Priv-Accept ignores iframes.

We compare both tools on the Tranco top-1k websites. With Priv-Accept, we can click accept on 451 websites, whereas with *BannerClick*, the number is 430. Websites where Priv-Accept can click accept but not *BannerClick* are 66, and vice-versa 59 websites. The vast majority of the former set are websites that do not show an explicit accept option. These are not considered to be explicit accepts by *BannerClick*, however, Priv-Accept considers them. Additionally, Priv-Accept also clicks on the incorrect accept button for 11 websites. The latter group contains websites where Priv-Accept is unable to identify the correct button, *BannerClick* detects banners in iframes, or the website is in a non-English language.

3.6 Documentation

In this section, we document the usage of *BannerClick*. First, we explain how to install the tool and its dependencies. Next, we provide a step-by-step guide on running *BannerClick* for the first time. Finally, we describe how to extend *BannerClick* and adapt it to new use cases.

Installation: BannerClick is implemented as a custom command in OpenWPM. Therefore, OpenWPM must be installed before running BannerClick.

The primary prerequisite for OpenWPM is Conda, an open-source, cross-platform package manager. It can be installed from the official website:

`https://docs.conda.io/en/latest/miniconda.html`

Once Conda is installed, after downloading the *BannerClick* source code from

`https://github.com/bannerclick`

run the `install.sh` script to install all necessary dependencies in a separate Conda environment named `openwpm`.

Note that we have customized `install.sh` to install the Firefox binary from the local `firefox-bin.tar.bz2` file instead of downloading it from the server, as Firefox 95 is no longer available for direct download.

To execute the installation script, use the following command:

```
./install.sh
```

After running the installation script, activate the Conda environment by executing:

```
conda activate openwpm
```

Quick Start: To run *BannerClick* through OpenWPM, use the following command:

```
python demo.py --bannerclick --headless --num-browsers 8 --num-
    repetitions 5 ./bannerclick/input-files/Tranco5Nov.csv
```

Note that BannerClick can also be executed as an independent module outside OpenWPM. To do this, the `init()` method is used to instantiate primary variables such as files and the web driver. By running the `run_all()` method, all core functionalities can be executed for a given set of domains. These domains can be provided as an argument to `bannerdetection.py` using the `-f/-file FILENAME` option or manually by specifying the `file` argument in the `init(file=FILENAME)` function.

Executing the code below follows this process:

```
import bannerclick as bc
bc.init("urlfile.txt")
bc.run_all()
```

1. Extracts URLs from `urlfile.txt` and initializes the web driver.
2. Running `run_all()` triggers banner detection for each URL in the file. If a banner is detected, a screenshot is taken, and its characteristics—such as content, size, language, and more—are stored in the database `crawl-data-js.sqlite`.

Configuration: Aside from the configuration settings for OpenWPM, several parameters in `config.py` can be modified to configure BannerClick. Each parameter is documented directly in the file. Below, we elaborate on a few key parameters:

- `MOBILE_AGENT` Enables simulation of a mobile user agent.
- `GPC_SIGNAL` Enables Global Privacy Control (GPC) signals in the browser.
- `START_POINT` and `STEP_SIZE` Define the range of URLs to be processed from the input file.
- `TIME_OUT` Specifies the timeout threshold for Selenium and, accordingly, OpenWPM for each crawl.
- `SLEEP_TIME` Specifies the wait time between page loads during each crawl.
- `ATTEMPTS` Defines the number of attempts to detect a banner on a page.
- `CHOICE` Determines the interaction mode:
  - 0 No interaction
  - 1 Accept
  - 2 Reject
- `HTTP_INSTRUMENT` Can be set to `True` to enable HTTP logging (same applies for other instrumentation options).

Detection Workflow: Here we elaborate on some key methods used in the banner detection process to detect element with banner-related word, the anchor element, the most-specific element (optimal element).

1. **Finding the element with banner-related word:**
   - `find_els_with_cookie(body_el)` scrapes elements containing cookie-related terms.
   - `pruning_els(els)` eliminates irrelevant elements, such as those inside `<footer>`, `<header>`, `<script>` tags, or invisible elements.
2. **Finding the anchor element:**
   - `find_fixed_ancestors(els_with_cookie)` identifies fixed-position ancestors of the elements with banner-related word. If none are found, the body tag is used.

- `find_by_zindex(els_with_cookie)` searches for ancestors with a positive z-index attribute. If none are found, the body tag is used.
- The mapping `banners_map[first_node] = last_node` groups elements with banner-related word under the same anchor element, indicating they belong to a common banner.

3. **Finding the most-specific element:**
   - `find_optimal(driver, path)` traverses the path from the anchor element to the element with banner-related word to identify the best representative element for the banner.
   - `is_size_equal_to_win(driver, head_el)` checks if the candidate element's size matches the window size. If so, a more compact but inclusive tag is selected.
   - `is_there_major_child(driver, head_el, path)` determines if a child element contains all relevant text and elements but is smaller in size. If such an element exists, it becomes the most-specific element.
   - `is_one_dimension(head_el)` helps detect zero-sized elements to avoid false positives.

Steps 1-3 are repeated for all iframe elements to check for cookie-related frames. If no banner is detected and the webpage is in a non-English language, the page is translated, and the detection process is repeated.

Data Collection Methods Once banners are detected, several methods can be used to collect relevant data:

- `take_current_page_sc()`: Captures a screenshot of the current webpage.
- `take_banners_sc(banners)`: Captures screenshots of detected banners.
- `extract_banners_data(banners)`: Saves banner characteristics into `DataBase.csv`.

4

Multi-Perspective Analysis of Web Cookies

Building on the capabilities of *BannerClick*, introduced in the previous section, we conduct our first study through a systematic exploration of the cookie landscape. By automating cookie banner detection and interaction, *BannerClick* facilitates the analysis of cookies across multiple dimensions with unprecedented precision and scale.

In this chapter, we focus on understanding the impact of cookie banners, geographic locations, user agents, and other factors on cookie deployment. These analyses aim to provide a comprehensive view of the cookie landscape and address critical gaps in existing research. Specifically, we investigate the following key aspects:

- **Interaction with Cookie Banners:** Many GDPR-focused studies overlook the effects of interacting with cookie banners (e.g., accepting or rejecting cookies) [1, 37, 86, 148]. Using *BannerClick*, with detection, acceptance, and rejection accuracies of 99%, 97%, and 87% respectively (see Section 4.2), we detect banners on 47% of Tranco top-10k websites in the EU, compared to less than 30% in non-EU regions (see Section 4.3.1). Interaction with banners increases third-party cookies by 5.5×.

- **Impact of Geographic Locations:** To evaluate GDPR's effectiveness, we compare cookies (especially third-party and tracking cookies) between EU and non-EU vantage points (cf. Section 4.3.2). Without banner interaction, 43% of websites send more tracking cookies from non-EU regions. Even after accepting a banner, 83% of websites send more tracking cookies in non-EU regions, rising to 96% after rejecting banners. These findings demonstrate GDPR's positive impact on reducing third-party and tracking cookies.

- **Consistency of Websites:** Cookie analysis requires consistency in cookie counts across visits. We conduct two statistical tests: First, the coefficient of variation measures consistency when revisiting websites from the same location. Second, the MannWhitney U test [90] measures consistency across visits from different locations. Results show greater consistency within the EU and significant differences between EU and non-EU regions (cf. Section 4.3.4).

- **Cookie Differences Between Landing and Inner Pages:** Landing and inner pages often differ structurally and in content [6]. Cookies exhibit similar disparities (cf. Section 4.3.5). At our US VP, 32% of websites send more third-party cookies on landing pages, while in Germany, 29.7% send more on inner pages. Overall, 27.4% and 15.7% of websites show different third-party and

tracking cookie behavior across vantage points, indicating that focusing only on landing pages may provide an incomplete picture of the cookie landscape.

- **Cookie Differences Between Desktop and Mobile Browsers:** As mobile browsing surpasses desktop usage [44, 139], we examine cookie differences between platforms (cf. Section 4.3.6). At our US East VP, 28% of websites send more third-party cookies on desktop, while in Brazil, 28% send more on mobile. Overall, 14.6% and 9% of websites show different third-party and tracking cookie behavior across vantage points. These findings highlight the need to analyze both desktop and mobile platforms in cookie studies.
- **Impact of Privacy Laws:** We examine the effects of Brazilian and Californian privacy laws [19,131]. Our findings in Section 4.3.7 show CCPA does not directly reduce web cookies. Instead, websites adhering to CCPA tend to send more third-party and tracking cookies, emphasizing the need for further analysis.

Overall, our measurement study highlights that factors such as banner interaction, client location, landing vs. inner pages, and desktop vs. mobile usage impact the cookie landscape. Future research should account for these variables to develop a more comprehensive understanding of web cookies. To encourage reproducibility, we have open-sourced the relevant version (v0.18.0) of *BannerClick* [118] and released our data and analysis scripts [116] at `bannerclick.github.io`.

4.1 Related Work

To regulate the use of cookies, various data protection laws such as the GDPR [39] in the EU or CCPA [19] in California have been enacted in the last years (see Section 2.2 for more details). A large body of previous work attempts to quantify the efficacy of such laws. Dabrowski *et al.* [30] reported less persistent cookie usage for EU users in comparison to US users with Alexa top-100k websites as targets. On the contrary, Sanchez *et al.* [127] claimed that the US appears to approach cookie regulations similar to the EU. We do, however, observe a lower number of TP cookies in the EU when compared to non-EU VPs (see Section 4.3.2).

Furthermore, to check whether website publishers adhere to the EU cookie laws, Trevisan *et al.* [148] developed the tool "CookieCheck" [149]. They reported that half of the websites they tested ($\approx 35k$) from an Italian VP, violate the law *i.e.,* they install profiling cookies[1] before the user's consent.

While studying tracking, Iordanou *et al.* [64] identified the geographic locations of the tracking servers. They found that around 90% of the tracking flows originating in the EU terminate at tracking servers hosted within the EU itself. Additionally, there are multiple measurement studies that highlight how trackers use cookies for user profiling [17,36,40,50,84,85,130]. As an example, Englehardt *et al.* [37] demonstrated

[1] These are cookies that are managed by Web trackers to identify users and are clearly subject to explicit consent according to the GDPR.

that adversaries can reconstruct up to 73% of a user's browsing history using only the collected cookies.

Linden *et al.* [86] took a different direction; they conducted a longitudinal study to assess privacy policies adopted by website publishers before and after GDPR went into effect. They reported that GDPR has a positive impact on privacy policies. Post-GDPR, not only the visual (and textual) representation of policies have improved, but the coverage of important topics *e.g.,* data retention, has also increased. Degeling *et al.* [31] also made similar observations *i.e.,* after GDPR, many websites have added and updated their privacy policies and now show cookie banners to the users. Sørensen *et al.* [137], rather than analyzing the privacy policies, found that after the introduction of GDPR, the number of third parties on EU websites has declined. They noted, however, that it cannot be concluded with certainty that this decline is solely due to GDPR. Kretschmer *et al.* [76] conducted a comprehensive survey of the existing research (> 70 research papers), describing the legal as well as technical aspects of GDPR. They report that the enactment of GDPR has resulted in a decline in third-party tracking and an increase in cookie banners and privacy policies in the EU region.

Santos *et al.* [128] studied cookie banners to analyze how clearly they explain privacy policies. They manually analyzed 400 cookie banners on English language websites that are popular in the EU. They report that 61% of banners used vague language and violated the specificity purpose. Utz *et al.* [154] rather than only focusing on the text of the banners, also studied other factors that can influence user consent decisions (*e.g.,* positioning of the banners on the website). The authors partnered with an e-commerce website in Germany and reported that changing the position of the banner or the text had a significant impact on the users' consent decisions. For instance, if the banner is shown in the lower left part of the screen, users are more likely to interact with it.

More recently, Chen *et al.* [21] conducted a user survey of Californian consumers to analyze how well they understand popular websites' privacy policies. They reported a significant variance in how websites interpret CCPA. Thus, privacy policy disclosures (mandated by CCPA) seem ambiguous to end-users. To this end, Connor *et al.* [105] performed a study to specifically analyze how websites implement "right to opt-out of the sale of users' personal information". They observed that websites implement this mandate in ambiguous ways, which deters the users' motivation to opt-out.

Finally, other research specifically analyzes cookie banners themselves *e.g.,* how clearly they specify privacy policies [128] or the impact of banner location on user consent [154]. Jha *et al.*'s [68] work is closest to our research. Similar to our work, the authors also attempted to interact with the banners in an automated manner to observe differences in cookies. However, as we discuss in Section 3.5, their tool only accepts the privacy policies (of the banner), whereas our tool *BannerClick* has the capability to accept as well as reject a banner's consent.

Table 4.1: Overview of different measurement types.

Measurement Type	Start Date	Duration	Target Websites
Banner Interaction	Jan 20, 2022	20 days	Tranco Top 10k
Consistency Tests	Feb 9, 2022	10 days	Tranco tiered 300
Landing vs. Inner	Mar 8, 2022	4 days	Tranco tiered 300
Desktop vs. Mobile	Feb 27, 2022	10 hours	Tranco tiered 300
Impact of CCPA	Mar 13, 2022	10 hours	Tranco tiered 300

4.2 Methodology

We now present our VP locations, target websites, and our approach to studying the cookie landscape in detail.

4.2.1 Data Collection

We use AWS cloud instances at the following locations as our VPs: Frankfurt (Germany), Stockholm (Sweden), Ashburn (US East), San Francisco (US West), Mumbai (India), São Paulo (Brazil), Cape Town (South Africa), and Sydney (Australia). We select these vantage points to have two VPs inside GDPR countries (Germany and Sweden), two VPs in the US (of which one is in the CCPA state California), one in Brazil (that has LGPD), one in Africa, one in Australia, and one in Asia.

In our measurement study, we use the global Tranco top-10k [113] as target websites for our analysis. The popularity of these websites is measured considering the actual Web traffic of users [129]. Other counterparts like the Cisco Umbrella list [61] and the Majestic Million list [89] are created using indirect sources like DNS queries and URLs embedded in website ads.

Additionally, for experiments that require repeated measurements (*e.g.,* consistency tests), we use a subset of Tranco top-10k websites; we select three sets of websites: Tranco top-100, 1001–1100, and 9901–10k. These sets include websites from the top, middle, and bottom of the Tranco top-10k websites and hence represent different website tiers. We call this subset the "tiered Tranco list". In order to identify a suitable OpenWPM configuration, we perform multiple small-scale test runs. Table 4.1 shows an overview of our final large-scale measurement runs. The longest measurement takes 20 days, in which the Web can change substantially. In order to keep results comparable, we ensure that each website is crawled at a similar time from all vantage points. In the case of failure in one vantage point the website would be excluded from the final result. Moreover, we run OpenWPM in stateless mode and ensure that the browser does not block tracking when accessing websites [107].

4.2.2 Stateless Banner Interaction

As already mentioned, we completely automate our measurement campaign to access the Tranco websites and collecting the cookies using OpenWPM. In particular, in this study, we adopt a stateless approach for collecting cookies, meaning no cookies are stored between visits, and the browser is restarted after each crawl. While this approach may not fully replicate real-world user behavior—where users browse over time and may have previously accepted or rejected cookies—it ensures that cookies collected during each crawl are unbiased and unaffected by prior interactions.

A key advantage of the stateless approach is that it eliminates the influence of prior visits, ensuring consistent and reliable data. This is crucial for evaluating consistency across multiple visits, comparing cookie behavior between landing and inner pages, and analyzing the impact of user agents (e.g., desktop vs. mobile). Additionally, it facilitates systematic analysis of cookie behavior under different banner interaction modes, such as accepting, rejecting, or ignoring the banner.

Overall, the stateless approach is ideal for our goals in this study, offering a robust framework for analyzing cookies across scenarios like consistency, user agent impact, and banner interactions without interference from previously stored cookies.

4.2.3 Cookie Classification

Classifying cookies as first-party or third-party requires identifying the domain of the website as well as the received cookies. In this study, we use the public suffix list [100] to identify the domain of (1) the website and (2) the URL in the *domain attribute* of the cookies. Then for each of the received cookies, we compare its domain with the website's domain. On a successful match, we classify the cookie as first-party; otherwise, we consider it a third-party.

Next, similar to Götze *et al.* [54], we use the justdomains blocklist [70] to identify tracking cookies. This list contains entries from various popular tracking lists *viz.* EasyList, EasyPrivacy, AdGuard, and NoCoin filter lists, only if the *complete domain* is identified as tracking. If the cookie domain matches one of the domains in the justdomains list, we classify it as a tracking cookie. To ensure the correct classification of tracking cookies, we perform a small-scale validation: We identify the top 100 websites sending the most tracking cookies and then we manually inspect the tracking cookie domain. We confirm that well-known tracking domains are indeed sending these cookies (*e.g.,* `doubleclick.net`).

4.2.4 Measurement Setup

We use Amazon EC2 instances in eight different geographic locations. These instances have four CPU cores and are provisioned with 16GB RAM. For our measurements, we use OpenWPM v0.18.0 running Firefox in stateless mode [108] with the following configuration. In each run, we execute 7 browser instances in headless mode, with a

60s Selenium timeout[2]. Empirically, we observe the vast majority of websites to be loaded within these 60s. Moreover, we set the sleep time to 30s, which we experimentally find to be a suitable value. The sleep timer starts when the on-load event is triggered, ensuring that OpenWPM remains on the website for this time period. This is necessary because some cookies are still being set even after the page has finished loading. Furthermore, we set the OpenWPM timeout[3] to 360s (six times larger than the Selenium timeout). *BannerClick* starts detecting the banner (and interacting with it if configured) in three attempts at 0, 10, and 20 seconds after the sleep time has started. We see that more than 94% of banners are detected just on the first try. To aid in manual verification of measurements, *BannerClick* takes a screenshot of the website before interaction, the detected banner, the clicked buttons, and the website after each click.

For the banner interaction measurements from the VP in Germany, which consists of 150,000 separate crawls (10k domains each with 5 repetitions and 3 different modes of interaction), 138,018 are reachable, 946 and 455 exceed Selenium's and OpenWPM's timeout, respectively, for 10,175 the domain is unreachable, 406 trigger exceptions (*e.g.,* due to the lack of a `<body>` tag or page reloading during banner detection). In total, we consider 135,307 successfully completed measurements from all 8 vantage points in our analysis.

4.3 Results

In this section, we present our findings. We begin by analyzing the impact of cookie banners on cookie distribution (Section 4.3.1), followed by the effect of geographical location on cookies (Section 4.3.2). Next, we evaluate the consistency of websites (Section 4.3.4), compare cookies on landing and inner pages (Section 4.3.5), and assess the impact of client platforms (Section 4.3.6). Finally, we examine the effects of CCPA on the deployment of tracking cookies (Section 4.3.7).

4.3.1 Impact of Banner Interaction

We run *BannerClick* on the Tranco top-10k websites [113] to analyze the effect of banner interaction on the deployment of different type of cookies. First, we investigate how many websites we can detect and interact with banners. From the vantage point in Germany, we can successfully detect banners on about 47% out of all accessible websites. *BannerClick* is then able to click on Accept and Reject buttons of the banner for around 40% and 30% of all websites, respectively. Next, Figure 4.1 shows how interacting with banners can substantially impact cookie distribution. After accepting a banner, the number of first-party (FP) cookies increases by more than

[2]Selenium timeout indicates the duration that Selenium waits for a website to be loaded by the browser.

[3]OpenWPM timeout forces the current website crawl to stop upon expiration. That is useful, as Selenium freezes during the loading of some websites (*e.g.,* `bet365.com`).

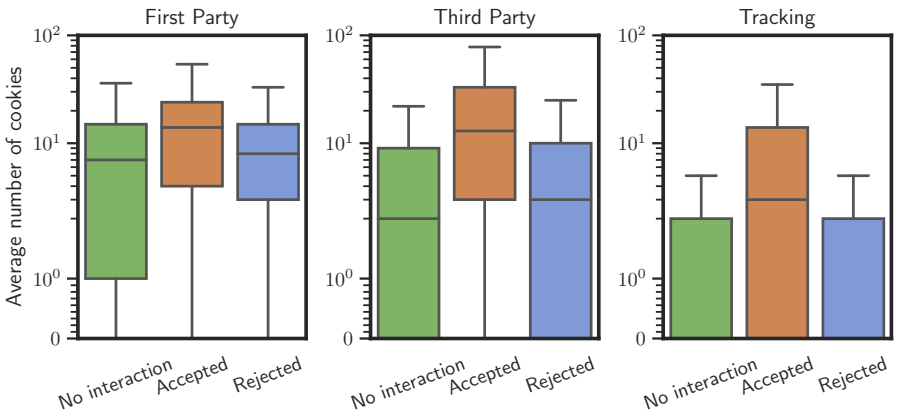


Figure 4.1: Cookie differences between no interaction, accept, and reject from the Germany VP.

1× and the number of third-party (TP) cookies increases by 5.5× on average. As for tracking cookies, the average increase from zero to 7 which shows a significant impact. Also, the minimum number of cookies set by 75% websites (lower quartile) increases from 1 to 4 and 1 to 3 for FP and TP cookies, respectively; for tracking cookies it remains 0. Moreover, we observe a jump in the maximum number of cookies set, which for third-party cookies, and consequently tracking cookies, is quite noticeable.

As for the rejection impact on first-party cookies, we can also see a slight increase in the number of cookies. This might be because of cookies that are being set to keep the state of rejection for future website access. This is further corroborated as we do not see this trend for third-party cookies. Furthermore, we see that the number of tracking cookies is quite low (near zero) when the banner is not accepted, which indicates the effectiveness of GDPR in reducing tracking.

To summarize: *Interacting with consent banners significantly impacts cookie deployment, with accepting banners leading to a substantial increase in first-party, third-party, and tracking cookies, while rejecting them keeps tracking cookies near zero—highlighting both the influence of user choices and the necessity of tools like* BannerClick *to accurately assess web tracking in the wild.*

4.3.2 Impact of Geographical Location

We examine the effect of geographical location on banner interaction and web cookies to observe if websites behave differently (*e.g.,* set a different number of cookies) based on regions. We crawl the Tranco Top-10k websites from eight geographically diverse vantage points (see Section 4.2.1 for more details). While accessing the websites, we interact with the banners in three modes: no interaction, accept, and reject.

Figure 4.2 illustrates the impact of geographic location on the prevalence of detected banners and the availability of different interaction options. In non-EU countries, we

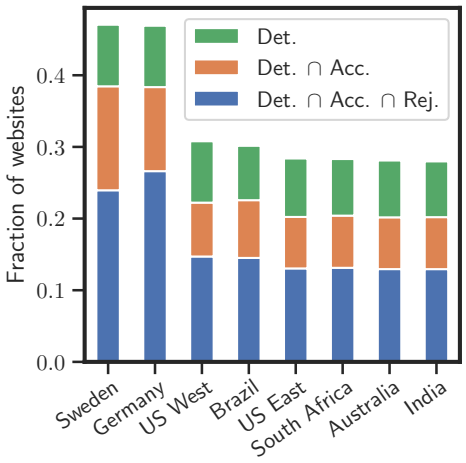


Figure 4.2: Effect of location on banner detection, accept, reject.

detect banners on less than 30% of websites, whereas in EU countries, we observe a higher prevalence (*i.e.,* approximately 47% of the Tranco top-10k). This represents a 56% increase in banner presence in EU compared to non-EU regions. Additionally, across all locations, *BannerClick* is able to accept more banners (blue + orange) than reject them (only blue)[4]. This suggests that banners are designed to favor acceptance options over rejection options.

Furthermore, to analyze the effect of geographical location on cookies, we visit each website five times in each mode and record the number of cookies. If a website is not accessible in five of the iterations at any location, we exclude it from our geographical location analysis. We now report the cookie trends observed at different locations in different modes.

No Interaction Mode: In the no interaction mode, 63% of sites set a different number of TP cookies in at least one location. Of these websites, 56% follow a trend where they set the highest number of TP cookies in either the US East or the US West and the least in Germany and Sweden. We also confirm that in the EU region, about 56% of websites set TP cookies and 30% set tracking cookies even in the no-interaction mode. In non-EU regions, a larger proportion of websites set TP (64%) and tracking cookies (43%). This indicates that GDPR has a positive impact on the reduction of TP and tracking cookies, but still many websites set these cookies without the users' consent. Setting TP (especially tracking) cookies before taking users' approval is a clear violation of GDPR.

Accept Mode: When analyzing the accept mode, we focus on those websites where we can successfully detect and accept banners at all VPs (*i.e.,* 18% of Tranco top-10k). This ensures that banner presence and different banner languages due to varying VPs do not influence our analysis. Amongst them, 21% of websites send precisely the same number of TP cookies at all locations; examples include `truecaller.com`,

[4]The slightly lower number of rejects in Sweden compared to Germany is due to a lack of Swedish reject-related words in our corpus.

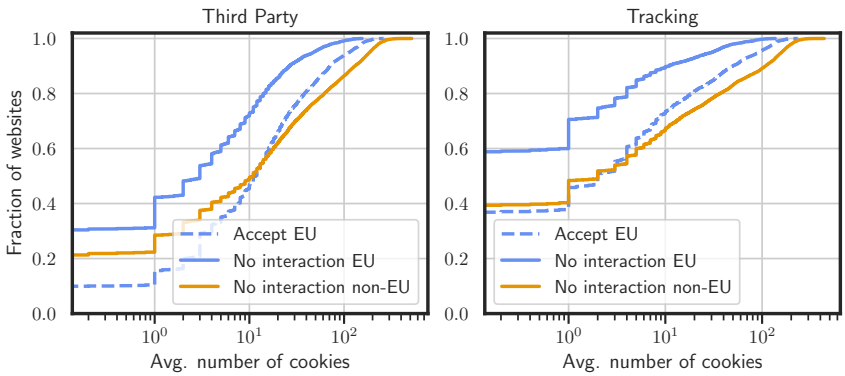


Figure 4.3: ECDF plot with the average number of TP (left) and tracking (right) cookies for websites on which *BannerClick* is able to click accept only in the EU.

`ghostery.com`, and `deepmind.com`. These websites represent an ideal case where users from different regions receive the same number of TP cookies after consenting to the banner. This is noteworthy as even users who reside in regions without strong data protection laws (*e.g.,* India) experience similar privacy standards to those that live in the regions protected by such laws (*e.g.,* EU).

To further assess the impact of GDPR on TP and tracking cookies, we now consider those websites that offer banners *only* in the EU and on which *BannerClick* is able to click the accept button (*i.e.,* 37.6% of the total). For such websites, we observe that the variation in TP cookies is nearly identical for both VPs in the EU. We find a similar trend across the rest of the VPs in non-EU regions. Thus, we aggregate the data points per website for VPs in the EU, and separately for all non-EU ones.

In Figure 4.3 we show an ECDF of the number of TP and tracking cookies for both EU (in blue) and non-EU regions (in orange). It is evident from the figure that, before interaction, about 60% of websites in the EU region set, on average at most 5 TP cookies, and about 80% of websites set, on average at most 4 tracking cookies. On the contrary, in non-EU regions, 60% of the websites set at most 20 TP cookies, and 80% set at most 40 tracking cookies *i.e.,* an increase compared to the EU region by a whole magnitude. Interestingly, 65% and 83% websites set fewer TP and tracking cookies respectively, even after accepting the banner policies in the EU, compared to no interaction at non-EU VPs. This shows that GDPR has a noticeable impact on the number of TP cookies. However, as expected, we find that GDPR does not impact FP cookies: 70% of websites set more or an equal number of cookies after accepting the banner compared to no interaction at the non-EU VPs.

Reject Mode: For the reject mode analysis, we again select websites that again show banners only in the EU, and for which we are able to click the reject button (*i.e.,* 23.7% of the total). We find that 87% and 96% of these, set fewer TP and tracking cookies respectively in the EU after rejecting the banner compared to the no interaction mode at non-EU VPs. We observe a similar trend for FP cookies: 72% of these websites set fewer FP cookies in the same scenario.

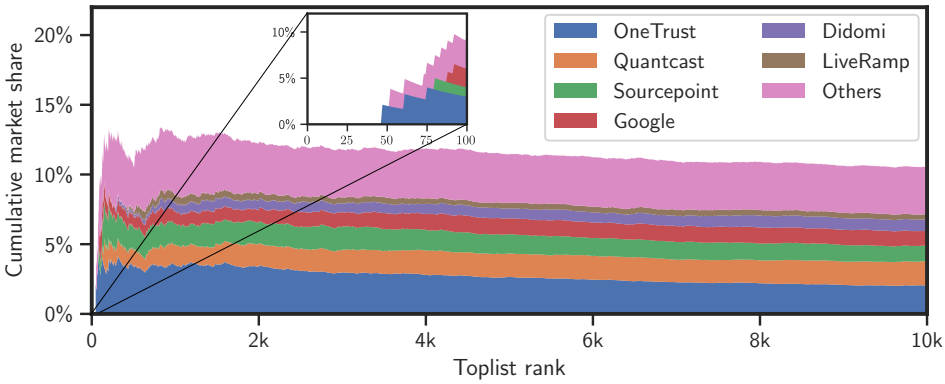


Figure 4.4: CMP distribution depending on the Tranco rank from the Germany VP.

To summarize: *Consent banners are 56% more prevalent and are biased toward offering acceptance options over rejection options when accessing websites within the EU compared to non-EU regions. Additionally, GDPR has a notable impact on reducing third-party and tracking cookies, with websites in the EU setting far fewer cookies than those in non-EU regions—both before and after banner interaction. In contrast, other privacy laws, such as LGPD and CCPA, show no measurable effect on TP and tracking cookies.*

4.3.3 CMP Distribution

As described in Section 3.3, *BannerClick* can detect CMPs using the `__tcfapi` function or the capabilities of the NeverConsent extension. In this study, we analyze the CMP distribution across the Tranco top-10k websites from the Germany VP. This is important as the use of CMPs is increasing, and they are widely regarded as the de facto standard for good practices in banner design and implementation.

In Figure 4.4 we show the cumulative market share of different CMPs for the Tranco top-10k websites. As we can see, in total within the top-1k websites around 13% of websites use CMPs. The CMP deployment remains almost constant with increasing rank, hinting at a consistent CMP deployment between ranks 2k and 10k. The CMP ecosystem is dominated by four companies (OneTrust, Quantcast, Sourcepoint, and Google) which are responsible for more than half of all CMP banners. Interestingly, we can not find a single website in the top 46 websites using a CMP and there is a generally much lower CMP deployment among top-ranked websites (see zoomed-in figure). This can be attributed to the fact that large Internet companies tend to avoid relying on third parties for handling privacy-sensitive data.

Throughout our study, we see a slight increase in CMP usage: From 95 websites out of the top-1k in July 2021 to 107 websites in January 2022. Therefore, it seems that CMPs will continue to play an important role in the cookie ecosystem, which future research should take into account.

As for other VPs, we see fewer CMPs detected on average. This is due to some CMPs not implementing their APIs (*i.e.,* `__tcfapi()` or custom ones), when they do not show a banner, which happens more for non-EU VPs. There is also an increase in the share of CMPs in the category "Others", which underlines that popular CMPs are less likely to provide APIs if no banner is shown.

Finally, we also compare our CMP results to previous work [59]. Their results for CMPs following TCFv1 are similar to our results for the new TCFv2 standard.

To summarize: *CMPs are dominated by four major providers (OneTrust, Quantcast, Sourcepoint, and Google), accounting for over half of all banners. While mid-ranked websites (Tranco 2k10k) show consistent CMP usage, top-ranked sites rarely use them, likely preferring in-house privacy management.*

4.3.4 Consistency Analysis

Next, we analyze the consistency of website cookie behavior, in order to learn how consistently websites send a certain number of cookies. This is important to ensure, that what we measure is not influenced by website randomness, *i.e.,* due to excessively changing third-party content. For statistical consistency analysis, we visit each website of the tiered Tranco top-10k (100 websites each in three different rank tiers) 100 times for each of the three different interactions (no interaction, accept, reject).

Intra-location consistency: To draw meaningful conclusions about cookie characteristics, one must ensure that a website sends a similar number of cookies when accessed multiple times from the same location. *E.g.,* if a website, when accessed for the first time, sends only five cookies, but when accessed the second time, sends hundreds of cookies, it should be classified as inconsistent. For such websites, it is non-trivial to draw meaningful conclusions from the measurements.

From each of the VPs (in eight countries), we measure the intra-location consistency using the coefficient of variation (CoV) as a metric. The CoV is calculated by dividing the standard deviation by the mean. The smaller the CoV, the more consistent the cookie behavior is, when looking at it from each VP separately. We visit each website of the tiered Tranco list from each location and then calculate the CoV based on the number of cookies the website sends. Figure 4.5 (a) shows the ECDF of CoV for third-party cookies. We can clearly see two groups of websites in the plot: EU (Germany and Sweden) on the top and non-EU below that. It seems that when visiting websites from within the EU, they exhibit a more consistent cookie behavior. However, this difference is influenced mainly by the number of websites that send exactly zero third-party cookies, which result in a CoV of zero: When visited from within the EU, more websites send exactly zero third-party cookies compared to when visited from a non-EU VP. This, in turn, leads to the ECDF curves of EU countries starting higher than non-EU countries, exhibiting a shifted but similar curve and later even merging. This is another indicator of the effect of the VP's geographical location in combination with GDPR on cookie behavior, as pointed out in Section 4.3.2. Overall, we find that 75–80% of websites are consistent with a CoV of less than 0.1 (*i.e.,* the

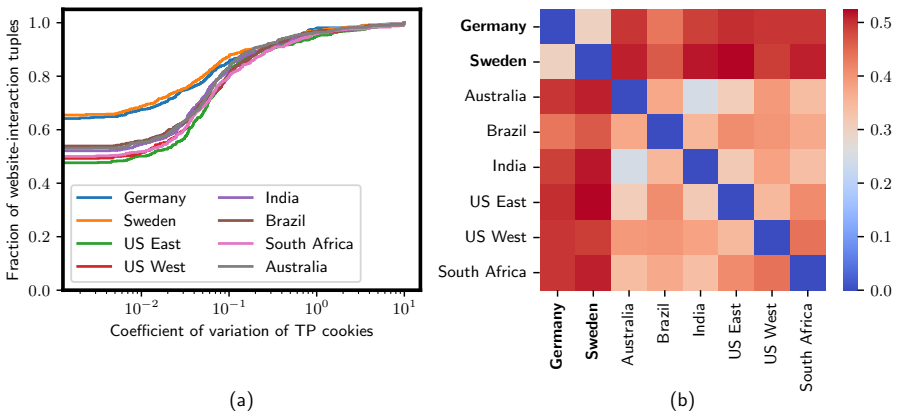


Figure 4.5: (a) Intra-location consistency of third-party cookies. (b) Inter-location statistically significant differences of third-party cookies (EU VPs in bold).

standard deviation is at most 10% of the mean). For first-party cookies (not shown) we see a more similar picture across VPs.

Inter-location consistency: To find statistically significant differences in the number of observed cookies depending on the VP location we use the Mann-Whitney U (MWU) test [90][5]. Again, we crawl websites from the tiered Tranco list 100 times for each interaction (no interaction, accept, reject) from each VP. Then we apply the MWU test with Holm p-value correction [60] and choose a p-value of 0.05 to determine statistical significance. In Figure 4.5(b), we show a heatmap depicting the statistical differences. In the figure, we see two main clusters, *i.e.,* EU vs. non-EU and non-EU vs. non-EU. We find that the majority of differences occur between EU (bold label) and non-EU locations, with more than half of all website-interaction tuples showing a statistically significant difference. On the other hand, if both locations are either in the EU or both outside the EU, we see fewer differences. Moreover, we also confirm that the Tranco rank tier does not affect the differences. An example of such a website is `nytimes.com`, which sends on average, 5 TP cookies when visited from Germany or Sweden, 10 TP cookies from Brazil, and more than 80 TP cookies from other countries. For first-party cookies (not shown), we see a similar picture across VPs, although with fewer differences in total.

To summarize: *Websites exhibit consistent cookie behavior when accessed multiple times from the same location, with 75-80% showing low variability ($CoV < 0.1$). Moreover, statistical analysis (MWU test) confirms that most differences occur between EU and non-EU regions, whereas websites behave more similarly within the same regulatory zones.*

[5]The MWU test is a statistical post hoc test, *i.e.,* it allows to find differences in the cookie distribution between all pairs of VP locations. Our setup fulfills the MWU assumptions, *i.e.,* all test samples from both groups are independent of each other, the samples are ordinal. The distributions of both populations are identical under H_0 and not identical under H_1.

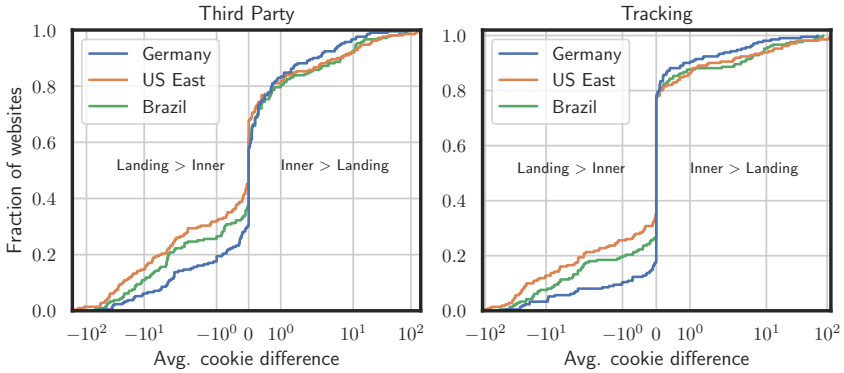


Figure 4.6: Average number of TP cookies comparing landing vs. inner pages.

4.3.5 Impact of Landing vs. Inner pages

When users access a website, they often not only access the website's main landing page but navigate through other inner pages of the website as well. For instance, people visiting the landing page `https://www.bbc.com/` can access the article on the inner page `https://www.bbc.com/sport/football/58920223`. Thus, it is important to study the differences between cookies for landing and inner pages for a given website. We use a simple criterion to classify a link as an inner page. An inner page link must begin with the landing page's fully qualified domain name (FQDN). For instance, **https://www.bbc.com/**sport/football/58920223 is the inner page of **https://www.bbc.com/**.

We intended to use the Hispar list [6] (which contains links to seemingly inner pages) for our analysis. However, we find that many inner pages mentioned in the list either do not begin with the FQDN of the landing page or redirect to completely different domains. For instance, `mail.google.com` is classified as an inner page of `google.com`, which in practice it is not. In general, we observe that more than 50% of inner pages (corresponding to a landing page) in the Hispar list are actually not inner pages. Thus, we use our own automated approach to access a given website's landing and inner pages. For our analysis, we select 10 random inner pages for each landing page as follows.

We first access the landing page of the given website (*e.g.*, `https://www.bbc.com/`). The obtained HTML page contains Web links to inner as well as non-inner pages. Next, we select a link by crawling for `<a>` elements and check whether it is a potential inner page or not. As already mentioned, we simply check that the inner page link must begin with the landing page's FQDN. Using Selenium, we visit this link and extract the final link (that might have changed due to redirection). If the link is an inner page, we append it to the list of inner pages. If it is already in the list, we ignore it and proceed with the remaining ones. Finally, we stop searching for inner pages when either 10 inner pages are found or a total of 50 links (present on the landing page) have been tested. We repeat the same process for all tiered Tranco websites.

In total, we obtain 2273 inner pages corresponding to 300 Tranco websites. We access the set of landing and inner pages from all VPs. Like our other experiments, we visit each webpage (landing and inner) five times in each mode (no interaction, accept, reject) and record the average number of cookies per webpage. Figure 4.6 shows the ECDF of the difference of average TP and tracking cookies from the ten inner pages compared to the corresponding landing page (in the no interaction mode). The negative difference on the x-axis (left part of the figure) corresponds to the fraction of websites where we observe more cookies on a landing page than on inner pages (shown as Landing > Inner). Zero means the same number of cookies is found for both categories. Positive values (right part of the figure) correspond to the fraction of websites where more cookies are sent on inner pages than the landing page (represented as Inner > Landing). Figure 4.6 depicts this difference for three VPs *i.e.,* US East, Brazil, and Germany. We show only these three VPs because we observe nearly the same trend for US East and US West; observations in Brazil are quite similar to India, South Africa, and Australia; the trend in EU countries is almost the same.

At all of our VPs, we find that 12.7% and 8% of websites set more TP and tracking cookies, respectively, on the landing page than on the inner page (*e.g.,* `amazon.com`, `vk.com`, and `youtube.com`). Looking at VPs separately, the proportion of such websites is the highest in US East (32% TP and 24% tracking) and the lowest in Sweden (21% TP) and Germany (12.3% tracking). Moreover, our analysis reveals that 87% of these websites set at least 10 more TP cookies on average on the landing page at all locations. One possible explanation for this trend can be that many websites show more content on the landing page, include more third-party content, and thus set more TP cookies.

Similarly, we observe that 14.7% and 7.7% of websites set more TP and tracking cookies respectively on inner pages across all VPs (*e.g.,* `cnn.com`, `bbc.com` and `reddit.com`). When investigating each VP separately, the proportion of such websites is the highest in Germany (29.7% TP) and South Africa (19.3 tracking), and the lowest in US East (22% TP) and Brazil (15.3% tracking). It is interesting to note that, although GDPR discourages the use of third parties without consent, a substantial fraction of websites prioritize setting TP cookies on inner pages. This can also facilitate user profiling [1] as third-party services can better characterize users' viewing habits and choice of content at a more fine-grained granularity. Overall, our results indicate that studying *only* the landing page provides a partial picture of the TP cookies a user gets. In total, 49.3% and 27.3% of websites set a different number of TP and tracking cookies, respectively, on landing and inner pages at all VPs.

Banners on inner pages: We check for banner presence as a potential contributing factor. Although we find a small number of websites with different banner behavior (*e.g.,* `www.colorado.edu/map`), we generally see a similar number of banners on landing and inner pages. Overall, using *BannerClick*, we detect banners on 22% (US East), 51% (Germany), and 30% (Brazil) of the landing pages of the tiered Tranco list. Correspondingly, we detect banners on 25% (US East), 50% (Germany), and 31% (Brazil) of the inner pages.

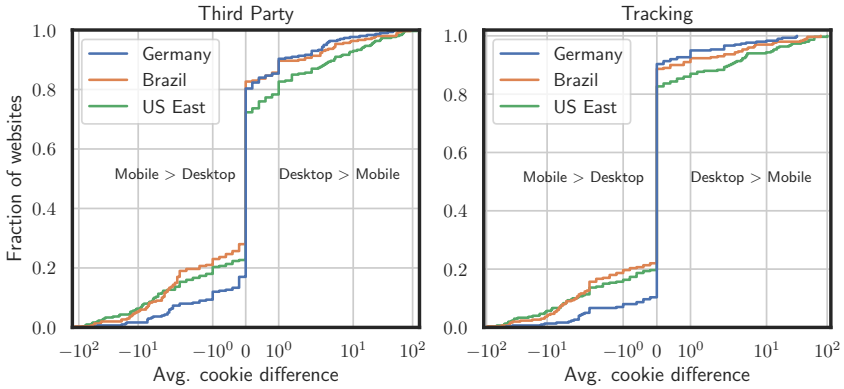


Figure 4.7: Average number of TP cookies comparing mobile vs. desktop.

To summarize: *Focusing solely on landing pages provides an incomplete picture of cookie deployment, as websites exhibit significant differences in third-party and tracking cookies between landing and inner pages.*

4.3.6 Impact of User Agent

We look into the effect of visiting websites from browsers in desktop vs. mobile environments to understand how websites and third parties behave in this context. To visit a website from a mobile browser, we modify the default OpenWPM user agent[6] and the screen size[7]. We manually confirm that modifying these parameters change the appearance of most websites[8] and we see both desktop and mobile versions of the same website. Interestingly, even with these minimal changes, we observe substantial differences between measurements conducted from desktop vs. mobile. We crawl the 300 tiered Tranco websites 5 times in each mode of interaction from all VPs with desktop and mobile configurations.

Figure 4.7 shows the difference between the average number of TP and tracking cookies measured per website when visited from a browser on desktop vs. mobile in the no interaction mode. We subtract the number of cookies observed on mobile from what we observe on desktop. Hence, websites that set more cookies on the desktop yield a positive cookie difference on the x-axis. Vice-versa, if a website sets more cookies on mobile, the cookie difference is negative on the x-axis. We observe that the TP and tracking cookies variation is nearly the same for US East and US West. The data from the VPs in the EU are alike, and the data from the remaining VPs are similar to each other. Hence, we plot the TP and tracking cookies per website for US East, Germany, and Brazil representing their respective classes.

[6]Desktop: "Mozilla/5.0 (X11; Linux x86_64; rv:95.0) Gecko/20100101 Firefox/95.0"; mobile: "Mozilla/5.0 (Android 12; Mobile; rv:68.0) Gecko/68.0 Firefox/93.0".
[7]Desktop: 1366x768; mobile: 340x695.
[8]In some cases this also changes the URL, *e.g.,* by prepending `m.` or `mobile.` to the domain name.

At all VPs, we find that 7.3% and 2.7% of websites set more TP and tracking cookies, respectively, when visited from a desktop (*e.g.,* `bing.com`, `twitch.tv`). On investigating VPs independently, we find that the proportion of such websites is the highest in US East (28% set TP and 17% set tracking cookies) and the lowest in Brazil (17% set TP cookies) and Sweden (9% set tracking cookies). From our analysis, we note that 7% of websites set at least 10 more TP cookies when being visited from a desktop from US East. These facts can be attributed to some websites having more content and, hence more embedded third parties on desktop than on mobile. Many websites, when designed for mobile, decrease the number of advertisements and limit the content to what is visible without scrolling. This reduces data usage and improves the user's viewing experience.

We also observe that 7.3% and 6.3% of websites set more TP and tracking cookies, respectively, when viewed from the mobile environment across all VPs (*e.g.,* `nytimes.com`, `livestream.com`). Distinct VP analysis shows that the proportion of such websites is the highest in Brazil (28% set TP cookies, 22% set tracking cookies) and the lowest in Sweden (15% set TP cookies) and in Germany (10% set tracking cookies). Our analysis shows that 4% websites set at least 10 more cookies when visited from mobile from non-EU VPs. As users are increasingly spending more time on their mobile devices [44], some third parties may prioritize placing more cookies when sites are visited from mobile for better targeting.

Overall, we observe that 14.6% websites set a different number of TP cookies when accessed from desktop and mobile environments at all our VPs. Furthermore, our findings show a higher degree of similarity between desktop and mobile compared to previous work [159], which did not consider banner detection or interaction at all.

Banners on websites browsed from mobile: We check for banner presence as a potential contributing factor in this experiment as well. Using *BannerClick* we detect a similar number of banners on websites when visited from desktop and mobile ($\approx$ 21% US East, 46% Germany, and 26% Brazil).

To summarize: *It becomes imperative that measurements from mobile environments also be considered for a real-world analysis of cookies as websites exhibit differences in third-party and tracking cookie deployment between desktop and mobile environments, with 14.6% of websites setting a different number of TP cookies across all VPs.*

4.3.7 Impact of CCPA

The California Consumer Privacy Act (CCPA) came into effect in January 2020. In the context of CCPA, selling personal information in the form of TP cookies has been a widely debated topic [9]. Thus, we take the first step to studying how CCPA-compliant websites deal with third-party cookies. To analyze the cookie landscape of such websites, we first need to find which websites are overtly complying with CCPA. For this, we use a straightforward approach. Websites covered by CCPA must include a conspicuous hyperlink on their homepage with the text "Do Not Sell My Personal

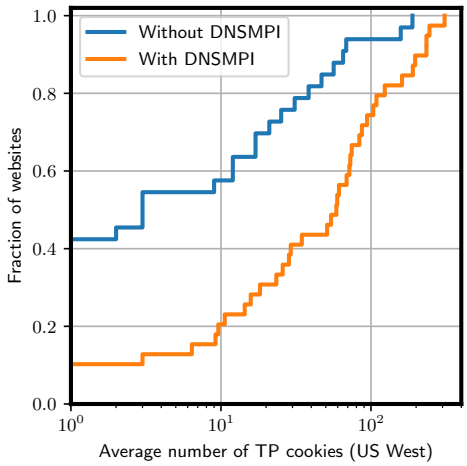


Figure 4.8: Effect of CCPA on cookies: Websites with DNSMPI links send more TP cookies.

Information" (DNSMPI) [103]. We crawl the tiered Tranco list and identify websites that contain this hyperlink[9] .

Out of 300 tiered Tranco websites, we identify that 39 websites contain DNSMPI links from our US West vantage point, 29 websites from US East, and 21 from Germany. This indicates that a user's location impacts whether or not the DNSMPI link is shown. Interestingly, this applies to different locations within the US as well, *i.e.,* we see 11 websites that only show the DNSMPI link to clients from California but not when visiting the website from the US East.

To observe the impact of CCPA on TP cookies, we compare the TP cookies of websites containing DNSMPI links with websites that do not include said links. We select our US West (*i.e.,* California) VP for this analysis. First, we classify the 39 websites with DNSMPI links into three sets belonging to Tranco top-100, 1001–1100, and 9901–10k, respectively. For instance, we obtain 12 websites that belong to the first set. Thus, to have a fair comparison, we randomly select the same number of websites without a DNSMPI link from the Tranco top-100 websites only. We repeat the same process for the other two sets as well. In the end, we compare websites in the same Tranco rank tier. In total, we compare 39 websites with DNSMPI links with the same number of websites without DNSMPI links. This approach ensures that differences in TP cookies are not due to differences in Tranco rank.

Similar to previous experiments, we crawl each website five times and record the number of TP cookies. Figure 4.8 illustrates the variation in average TP cookies for DNSMPI and non-DNSMPI websites (without cookie banner interaction). We can see that websites without DNSMPI (blue line) set a lower number of TP cookies than the websites with DNSMPI (orange line). For example, 42% of non-DNSMPI websites set on average just two or fewer TP cookies, whereas the same fraction of

[9]We use 8 different phrases for searching DNSMPI hyperlinks (*e.g.,* "do not sell my info") as suggested by Van Nortwick *et al.* [103].

DNSMPI websites send 30 or fewer cookies. For tracking cookies, the trend is the same as TP cookies.

We further extend our analysis to Tranco top-10k websites, where we identify a total of 1373 websites with DNSMPI links from the US West. We observe a similar trend as we see with the tiered Tranco list. This shows that CCPA does not have a positive impact on TP cookies by default. On the contrary, websites overtly adhering to CCPA send, on average more TP cookies than non-DNSMPI websites. Furthermore, users need to manually look for the often well-hidden (*e.g.,* in website footers) DNSMPI links and click them to get any real benefit. When it comes to reducing the number of cookies, CCPA seems much less effective than GDPR or similar legislation.

We check if banner presence may be contributing to the TP cookie differences for DNSMPI and non-DNSMPI websites. To our surprise, we find that DNSMPI websites are twice as likely to show a banner compared to non-DNSMPI websites. As a result, DNSMPI websites show a banner more often but still send more TP cookies.

To summarize: *Unline GDPR, CCPA does not inherently reduce TP cookies. Websites with DNSMPI links send more TP cookies on average than those without, suggesting that overt CCPA compliance does not equate to stricter cookie limitations.*

4.4 Discussion

Cookie banner automation: Since GDPR [39] and similar privacy legislation came into effect, cookie banners have become more and more prevalent on the Web. Moreover, during our measurements, we also see a wide variety of different banners. This not only makes automated detection and interaction more challenging for research purposes, but it also hinders browser and extension developers to effectively interact with banners in an automated fashion. These often rely on manually curated rules, do not have the option to reject cookie consent [73], or are no longer maintained [123]. Efforts to offer a general easy-to-use mechanism to refuse all tracking cookies such as HTTP's "Do Not Track" header [92], have not been adopted by the advertising industry and were therefore abandoned. The deployment of Consent Management Platforms (CMPs) can be leveraged as a standardized API for application developers to automate banner interaction. Unfortunately, we confirm previous findings [59] that many CMP websites do not properly implement these standardized APIs, which makes it difficult to make use of them. Moreover, CMPs are almost non-existent for very popular websites, which again leads to a lack of standardization potential for websites most visited by users. Additionally, many cookie banners make it purposefully difficult for people to reject all cookies [136]. As a prominent example, Google has been fined 150 million for not providing users a choice to reject all cookies and was consequently forced to update their cookie banner [33]. All these factors hinder effective banner automation and it is unlikely that the situation will improve without a joint push by browser developers, advertising companies, and lawmakers.

Looking ahead: In order to improve user privacy, browser vendors have recently started to block third-party cookies at various degrees. Mozilla introduced "Enhanced Tracking Protection" in 2019 [158] and is now moving towards completely isolated cookie stores per website [99]. Apple has introduced by-default TP cookie blocking in 2020 [46, 141]. Google has long touted its desire to get rid of TP cookies and proposed a myriad of different possible replacements [23, 47, 132, 143, 155]. Getting rid of TP cookies is likely not the end of user tracking, as different techniques such as Local Storage, IndexedDB, Web SQL, or browser fingerprinting [81] can easily replace TP cookie functionalities [27]. Finally, privacy regulations such as GDPR are not specifically limited to cookies, but require informed consent for any shared user data, irrespective of the used technology. Cookie banners will therefore likely remain a prominent sight in the future, even if the underlying technology might change.

Limitations: Even though we cover a wide range of factors in our work, there are natural limitations to our approach. First, since our banner detection approach leverages words from 12 languages, we might not be able to detect banners on websites using other languages. Second, we use OpenWPM which uses the Firefox browser to access websites. Websites can exhibit different cookie behavior when being accessed from a different browser, such as Chrome or Safari. Third, we solely focus on HTTPS when accessing websites. Since many browsers use an HTTPS-first approach and most websites do support HTTPS [41], we think this focus is warranted. Websites can also be accessed via QUIC, which is not yet widely deployed [160], and we thus do not consider it in our study. Fourth, to classify third-party cookies as tracking cookies, we rely on tracking cookie lists. In order to limit false positive tracking classifications, we use the conservative approach by Götze et al. [54]. Therefore, our identified tracking cookies serve as a *lower bound*. Fifth, to obtain the mobile version of the websites, we modify the OpenWPM user agent and screen size (see Section 4.3.6). Although for most websites, we see the mobile version, for some websites these simple changes are not enough to load the mobile version [159].

4.5 Summary

In this chapter, we leveraged *BannerClick* to conduct a multi-perspective analysis of Web cookies. Running measurements from eight geographic locations across five continents, we identified significant regional differences in cookie behavior. Websites displayed 56% more banners when accessed from an EU vantage point.

We also quantified the impact of banner interactions, finding that websites sent 5.5× more third-party cookies on average after clicking "accept," with a similar increase in tracking cookies. Additionally, our analysis revealed cookie variations based on page type (inner vs. landing) and client platform (desktop vs. mobile), highlighting the complexity of cookie deployment across different contexts.

5

Analyzing Cookie Paywalls on the Web

As demonstrated in the previous chapter, user interaction with cookie banners influences the subsequent deployment of cookies, and rejection can prevent further tracking. However, as shown in Figure 4.2, not all banners offer users a straightforward option to reject tracking. Furthermore, a new form of cookie banner, known as a "cookie paywall," has recently emerged on the web, which is even more restrictive than standard banners in terms of offering a reject option. Unlike regular cookie banners, cookie paywalls further limit user choice. Instead of providing granular consent options, cookie paywalls present users with a binary decision: either consent to tracking or purchase a subscription for an ad-free, tracking-free experience (see Section 2.3.4 for more details).

The GDPR explicitly states that consent must be given freely and unconditionally [26]. As a result, the legality of cookie paywalls remains contentious, with data protection authorities across EU countries holding differing views on the matter [97]. Moreover, the increasing prevalence of cookie paywalls may negatively influence users long-term decision-making and shape their perceptions of cookie banners.

In this chapter, we present the first fully automated, large-scale analysis of the cookie paywall ecosystem. Specifically:

- **Large-scale automated measurement study:** We perform a large-scale automated measurement study to detect cookie paywalls from eight vantage points on 45k websites. We develop a tool to automatically detect cookie paywalls with a precision of 98.2%, which we incorporate and release in the new version of *BannerClick* (v0.21.0) [118] together with our analysis code and data [117] at bannerclick.github.io (see Section 5.2).
- **Characterization of cookie paywall landscape:** We find cookie paywalls on a total of 280 websites (0.6%), with some countries such as Germany seeing a 5 times higher prevalence with 2.9% of reachable websites (see Section 5.3.1). We analyze different cookie paywall pricing schemes, finding that around 80% of websites charge 3 Euro per month or less (see Section 5.3.2). We investigate if buying a cookie paywall subscription indeed protects from tracking and show that subscribers see no tracking cookies compared to an average of 16 tracking cookies seen by non-subscribers (see Section 5.3.3). We uncover that the majority of found cookie paywall websites use Subscription Management Platforms to facilitate the deployment of cookie paywalls (see Section 5.3.4). We highlight

that common ad-blocking solutions are able to block 70% of cookie paywalls using manually curated filter lists (see Section 5.3.5).

- **Discussion of cookie paywall impact:** We discuss the impact of the advent of cookie paywalls, reflect on deceptive patterns to compel users to accept tracking, and reason about the possible future of cookie paywalls (see Section 5.4).

5.1 Related Work

Closest to our research are the works by Papadopoulos *et al.* [110] and by Morel *et al.* [97]. In the former paper, the authors investigate paywalls on websites and classify them into soft (limited number of articles can be read before the paywall is shown) and hard paywalls (a subscription is required to access content on a website). They identify 1.5k websites with some form of paywall-related JavaScript libraries on them. Contrary to our work, they do not look for cookie paywalls on websites. In the latter paper, the authors manually annotate and classify cookie paywalls on websites. They find 13 out of 2.8k websites (0.66%) showing a cookie paywall to the user. In comparison to their work, we completely automate the task of detecting cookie paywalls on websites, characterize the prevalent use of Subscription Management Platforms among cookie paywalls for the first time, and conduct our study on a much larger set of target websites (more than 45k compared to 2.8k).

5.2 Methodology

In this section, we describe the vantage points and target domains for our measurements, detail our cookie paywall detection approach, report on the accuracy of our technique, and discuss the limitations of our approach.

5.2.1 Data Collection

Similar to Chapter 4, we use AWS cloud instances at the following locations as our vantage points (VPs): Frankfurt (Germany), Stockholm (Sweden), Ashburn (US East), San Francisco (US West), Mumbai (India), São Paulo (Brazil), Cape Town (South Africa), and Sydney (Australia). We select these VPs as they include regions with different privacy regulations: GDPR in EU countries (Germany and Sweden), CCPA in California, and LGPD in Brazil. The remaining globally distributed VPs are in countries that have either no or less strict privacy regulations.

We use Google's Chrome User Experience Report (CrUX) [22] for target selection, as it has been shown to be a more realistic toplist [126] compared to Alexa [3] and provides a country-wise list, unlike Tranco [113]. We take the union of the country-wise Google CrUX top 10k domains for each VP country, resulting in 45 222 unique domains reachable across all VPs.

To conduct a large-scale studies across different VPs, we extend BannerClick to automatically detect cookie paywalls.

5.2.2 Cookie Paywall Detection

To measure the prevalence of cookie paywalls, we heavily modified *BannerClick* to be able to detect cookie paywalls in a completely automated way. We enhance BannerClick by adding support for HTML shadow DOMs [98] and implement a tailored technique to detect cookie paywalls on websites.

In our tests, we find that multiple websites with cookie paywalls use shadow DOM environments, which can not be directly modified or inspected by browsers or even Selenium [82] (*e.g.,* it is not possible to look up elements inside shadow DOMs using XPath or CSS selectors). We work around this limitation by looking for possible elements within the main HTML DOM with the `shadow_root` property. Then, we clone and append all child elements within a shadow DOM to the body element of the main document DOM. Thereafter, we find the desired button in the cloned DOM and then run the interaction function on the corresponding element in the shadow DOM. This allows BannerClick to also detect and interact with banners within open and closed shadow DOMs [121].

Before detecting cookie paywalls, we first run BannerClick to detect all types of cookie banners. We then leverage BeautifulSoup [122] to search for cookie paywall-specific words and classify banners as cookie paywalls. As cookie paywalls provide a tracking-free website by paying a subscription fee, we assemble a corpus of cookie paywall-specific words consisting of (1) words related to subscriptions (*i.e.,* abo, abonnent, abbonamento, abonne, abonné, ad-free and subscribe) and (2) currency words and symbols[1]. For each currency word or symbol, we check for a possible payment-related combination, *e.g., $3.99*, *3.99$*, *3.99 $*, or *3.99 $*. If these combinations of currency words or cookie paywall-related words appear in the text of a banner, we classify that banner as a cookie paywall. In total, we find that out of 280 correctly detected cookie paywall websites, 76 make use of a shadow DOM, 132 are embedded in iFrames, and 72 use the main HTML DOM to embed cookie paywalls.

Detection Accuracy: To measure the accuracy of our cookie paywall detection approach, we randomly select 1000 domains from our target list and manually check their screenshots to find the possible existence of cookie paywalls on the website. We find that we correctly detect all 6 present cookie paywall websites. The remaining 994 websites indeed do not show a cookie paywall. Therefore, for these 1000 random websites we have a precision and recall of 100%.

Furthermore, we manually check all 285 websites where we detected a cookie paywall to gain confidence in our detection approach. We find that 280 websites have indeed a cookie paywall, whereas 5 detections are classified as false positives. This results in a detection precision of 98.2%.

[1] We use the top 10 global currencies as well as the official currency of our measurement vantage points: EUR, USD, CHF, AUD, GBP, Rs, BRL, CNY, and ZAR.

VP	Cookie paywalls	Toplist	ccTLD	Language
US East	197	0	0	9
US West	199	0	0	9
Brazil	196	0	0	0
Germany	280	259	233	252
Sweden	276	15	0	0
South Africa	199	0	0	0
India	192	0	0	10
Australia	190	5	0	10

Table 5.1: Number of detected cookie paywalls depending on the country of the vantage point, country-specific toplist, TLD associated with that country, and the most commonly spoken language in that country.

5.2.3 Limitations

Our study provides valuable insights into the prevalence and characteristics of cookie paywalls. However, it is important to consider certain limitations when interpreting the results: First, we use an automated approach with a modified version of the BannerClick tool, achieving a 98.2% precision rate in detecting cookie paywalls. However, false negatives are still possible, and manual verification may not guarantee complete accuracy for all websites. Second, some websites identify web crawlers as bots [79]. Thus when they detect a crawler, they may behave differently—*e.g.,* altering the number of cookies or displaying cookie paywalls differently from a regular user. Although OpenWPM has mechanisms to mitigate bot detection, it is not feasible to completely circumvent bot detection. Hence, our study may not fully represent the actual website behavior experienced by regular users. Third, while our VPs are located in eight different geographical regions across six continents, more VPs in different countries can be added to the study. Thus, future studies can further increase the number of VPs across countries to obtain an even better understanding of cookie paywalls.

Finally, our study primarily examines the technical aspects and deployment of cookie paywalls, not user perceptions or behaviors. Understanding user perspectives requires additional research, such as user surveys or studies.

5.3 Results

In this section, we present results from our cookie paywall measurements, including cookie paywall prevalence across multiple characteristics (Section 5.3.1), subscription pricing (Section 5.3.2), third-party and tracking cookie analyses (Section 5.3.3), a case-study of Subscription Management Platforms (Section 5.3.4), and results from experiments to bypass cookie paywalls (Section 5.3.5).

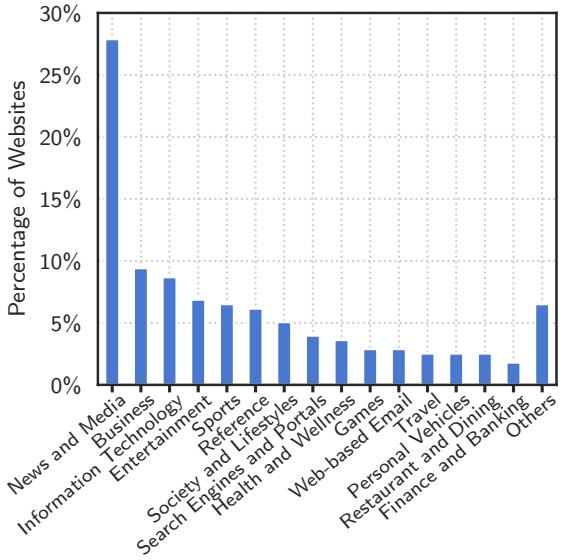


Figure 5.1: Categories of websites showing cookie paywalls.

5.3.1 Cookie Paywall Landscape

We use our modified version of BannerClick to run cookie paywall measurements from eight vantage points targeting 45 222 websites. In Table 5.1 we show different characteristics of our measurements and the detected cookie paywall websites. In total, we find cookie paywalls on 280 unique websites, resulting in an overall cookie paywall rate of 0.6%, a similar rate as found by previous work on a smaller set of target websites [97]. Our vantage points (VPs) in the EU (Germany and Sweden) see around 280 websites with cookie paywalls compared to around 200 for non-EU VPs. This finding is consistent with the generally higher prevalence of cookie banners in the EU [120].

Next, we analyze different characteristics—*i.e.,* country-specific toplists, top-level domains, and language—for *each vantage point* separately. We find that the Germany-specific CrUX toplist (see Section 5.2) contains by far the most detected cookie paywall websites (259, 2.9% of reachable top 10k websites), followed by Sweden (15) and Australia (5). We also find cases where websites on a country-specific toplist show a cookie paywall only when visited from a particular VP[2]. This shows that cookie paywalls are affecting users differently based on the list of popular websites within their country.

To better understand websites showing cookie paywalls to their visitors, we analyze the website top-level domain (TLD), the website's language, as well as the category the website can be attributed to. We find that again the vast majority of cookie

[2]For example, the website pt.climate-data.org is on the Brazilian country-specific toplist, but only shows a cookie paywall when visited from Germany or Sweden. This particular website is in fact operated by a German person, but provides specific subdomains for different languages, *e.g.,* `pt.` for Portuguese.

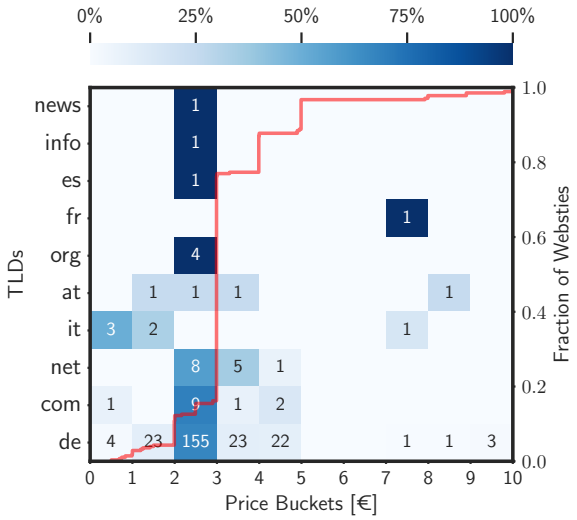


Figure 5.2: Distribution of monthly subscription price for cookie paywall websites.

paywall websites are hosted on Germany's `.de` country-code TLD (ccTLD), followed by generic TLDs (14 on `.com`, 14 on `.net`, 4 on `.org`), and non-VP ccTLDs (6 on `.it`, 4 on `.at`, and 2 on `.fr`).

Next, we inspect the language of the cookie paywall websites using CLD3 [51] to characterize the main target audience. Unsurprisingly, the largest part of these websites are in German[3], followed by English (US, Australia, India), Italian, and Swedish. To characterize the content of the website, we use FortiGuard's Web filter database [43] to assign each website to a category. As shown in Figure 5.1, more than one-fourth of all cookie paywall websites are categorized as news and media, 9% fall into the business category, and 7% are IT-related websites. This highlights that cookie paywalls—although they are most prominent on news websites—go beyond just news websites and are deployed on a large variety of different website categories.

Additionally, we find that cookie paywalls are more prevalent on popular websites, *i.e.,* 1.7% of country-wise top 1k domains show cookie paywalls compared to 0.6% for top 10k domains[4]. Interestingly, if we just consider the top 1k reachable websites for Germany, we detect cookie paywalls on more than 8.5% of websites, almost double the 4.7% in 2022 [97].

To summarize: *Cookie paywalls are most prominent on websites which are popular among users from Germany, where we see them on 2.9% of top 10k websites and 8.5% of top 1k websites. Moreover, cookie paywalls are visible on a wide variety of website categories, with news and media websites making up more than one fourth. In addition, more popular websites are more likely to show cookie paywalls.*

[3]Note that this might also include websites targeted at readers outside Germany, *e.g.,* Austria, Switzerland, or other German-speaking audiences.

[4]Note that the Google CrUX toplist does not contain detailed rank information per website. It rather groups websites into rank buckets, *e.g.,* top 1k or top 10k.

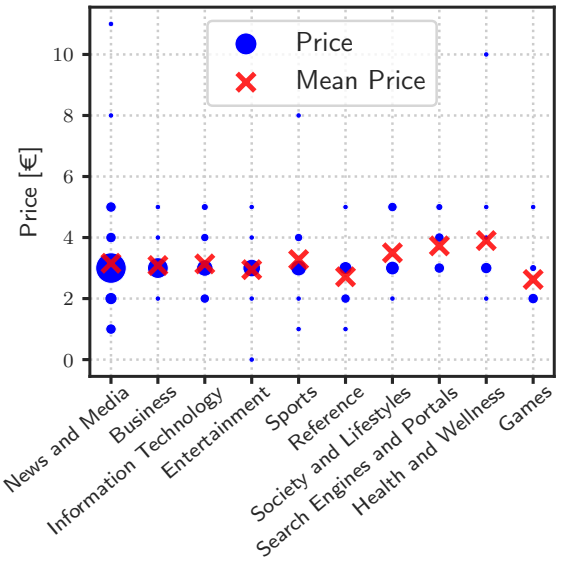


Figure 5.3: Correlation between the category of the websites and price of cookie paywall website subscriptions.

5.3.2 Subscription Pricing

In this section, we analyze the price of cookie paywall subscriptions of all detected 280 websites. We manually inspect each website to determine the price of a subscription. Then, we normalize the subscription price by month and convert it to Euro to make different websites comparable.

In Figure 5.2 we show the distribution of the monthly subscription price for cookie paywall websites. The red line shows an ECDF for the prices of cookie paywalls for all TLDs. We find that around 90% of cookie paywall websites ask for 4 Euro (approx. 4.33 USD) or less per month, and by far the largest fraction of websites charges 3 Euro (3.25 USD), with the majority of these websites being attributed to a Subscription Management Platform in which subscribers just need to pay once to access all partnered websites (see Section 5.3.4). On the other end, a handful of websites ask for 9 Euro (9.74 USD) or more per month. The heatmap in Figure 5.2 shows the occurrence of each price bucket for each TLD separately. We find that TLDs of websites do not have a substantial impact on the prices, as most websites in different TLDs charge between 2 to 3 Euro per month, except for `.it` which are on average cheaper. Furthermore, we explore potential correlations between website categories and subscription prices. In Figure 5.3 the size of the blue data points represents the number of websites falling within each price range, with the red cross showing the mean price per category. We find no obvious relationship between subscription price and website category.

To summarize: *We find that 90% of cookie paywall websites charge at most 4*

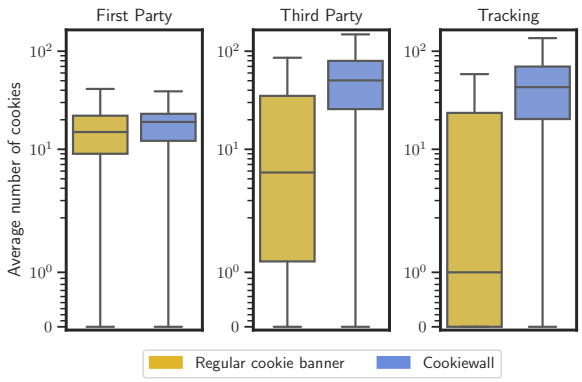


Figure 5.4: Average number of cookies comparing websites with regular cookie banners to cookie paywall websites.

Euro, with some outliers charging upwards of 9 Euro per month. Moreover, we find the prices to be generally similar for different TLDs and website categories.

5.3.3 Impact of Cookie Paywalls on Tracking Cookies

To assess the effect of cookie paywalls on user privacy, we now compare cookies sent by websites with cookie paywalls to websites with "regular" banners. Therefore, we run additional measurements targeting 280 cookie paywall websites and 280 randomly selected websites with regular cookie banners with an accept button. To account for variations in advertisements and consequently sent cookies, we repeat each measurement five times per website and calculate the average number of cookies per website. We then compare the number of first-party, third-party, and tracking cookies after accepting cookie paywalls and regular cookie banners. Similar to previous work [54,120], we use the justdomains blocklist [70] to classify cookies as tracking cookies. If the cookie domain matches one of the domains in the justdomains list, we classify it as a tracking cookie. Note that there exist other techniques to track users that we do not consider in this research, *e.g.,* browser fingerprinting [1], tracking using first-party cookies [20, 32, 101], and the use of invisible pixels and click IDs [10], as we specifically focus on studying the emergence of cookie paywalls. Thus, in the future, a more nuanced analysis focusing on other tracking techniques can be conducted.

Figure 5.4 compares the average number of cookies set by websites with regular cookie banners and cookie paywalls. In the figure, we see a similar number of first-party cookies among both website sets, with a median of 15 and 19 for regular cookie banner and cookie paywall websites, respectively. In contrast, third-party cookies exhibit a stark difference between both website sets. We find many more third-party cookies on cookie paywall websites with a median of 50.4, compared to just 6.8 for cookie banner websites. An even more pronounced discrepancy can be seen for tracking cookies, with cookie paywall websites sending on average 42 times more tracking cookies compared to cookie banner websites (median: 43 vs. 1). This seems to indicate, that websites with cookie paywalls try to monetize their users more

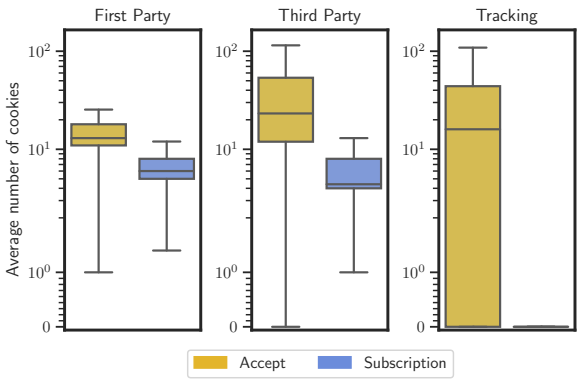


Figure 5.5: Average number of cookies set by websites with contentpass cookie paywall after accepting or accessing with a subscription.

aggressively compared to other websites, either through subscription fees or excessive tracking and advertising.

To summarize: *Cookie paywall websites send 6.4 times more third-party and 42 times more tracking cookies compared to "regular" cookie banner websites. This highlights the focus on monetization efforts of cookie paywall websites.*

5.3.4 Case Study: Subscription Management Platforms

Similar to Consent Management Platforms (CMPs) for regular cookie banners [59, 145], we find two different Subscription Management Platforms (SMPs) for cookie paywalls: contentpass [25] and freechoice [146], which claim to host cookie paywalls for 219 and 167 websites, respectively[5]. These two SMPs provide ad-free access to all partner websites for a monthly fee of 2.99 Euro. Note that only 76 contentpass and 62 freechoice partner websites are in our merged top 10k target list of previous measurements. We also find evidence of interoperability between CMPs and SMPs, with the CMP consentmanager providing integration support for the SMP contentpass [24].

In order to contrast the experience of subscribed users and users accepting tracking on SMP websites, we run an additional measurement for all 219 contentpass partner websites. Thus we create a contentpass account and buy a one-month subscription. We automate the login behavior on each of these websites and compare the sent cookies to clicking "accept". We again run five repetitions per website and average the number of cookies, in order to take website and advertisements variations into account.

Figure 5.5 shows the distribution of the average number of first-party, third-party, and tracking cookies across all 219 contentpass websites. We find a lower number of first-party (FP) and third-party (TP) cookies when accessing these websites with a subscription, with a median of 13 vs. 6 FP and 23.2 vs. 4.4 TP cookies for accept

[5] We observe an increase in these numbers between May and September 2023 to 270 for contentpass and 184 for freechoice.

and subscription, respectively. The most apparent difference can be seen for tracking cookies, where we see no tracking cookies with a subscription compared to a median of 16 when accepting the cookie paywall. Some websites send more than 100 tracking cookies when accessing these websites without a subscription. This underlines that cookie paywall websites are in fact aggressively tracking users, likely to maximize their income from non-subscribing users via ads and to push them towards buying a subscription.

To summarize: *Subscription Management Platforms provide an easy way for website operators to monetize their users by offering them a subscription instead of being tracked and served with ads. While subscribed users see no tracking cookies, users accepting cookie paywalls of the contentpass SMP see a median of 16 tracking cookies with some extreme cases sending more than one hundred tracking cookies.*

5.3.5 Bypassing Cookie Paywalls

This section delves into the feasibility, implications, and tools available for bypassing cookie paywalls on websites. The forcible accept-or-pay scheme of cookie paywalls might in the eyes of some users justify the act of bypassing it without being concerned about ethical considerations. One commonly employed method for bypassing cookie paywalls is the use of ad-blocker browser extensions. Notable examples include "I don't care about cookies" [73] "Ninja Cookie" [102] and "uBlock Origin" [151]. In this section, we focus on investigating the effectiveness of uBlock Origin, one of the most popular ad-blocker extensions.

To evaluate its effectiveness, we conduct a measurement on our 280 detected cookie paywall websites. We enable the uBlock Origin extension[6] and access each of the websites five times. We find that 196 (70%) websites no longer display cookie paywalls across all iterations, while the remaining websites still exhibit the cookie paywall prompt. Note that while browser extensions like uBlock Origin can effectively block resources with domains[7] listed in block lists (such as Easylist), they may not perfectly eliminate all types of cookie paywalls. Some cookie paywalls may be served locally or use lesser-known third-party domains, which may evade the blocking measures. Additionally, we manually inspect these 196 websites and find that all of them except two[8] work normally and do not show any ads.

To summarize: *Browser extensions like uBlock Origin can effectively block 70% of cookie paywalls in our measurements.*

[6] We enable the by default disabled Annoyances filter lists to block cookie paywalls.

[7] Example of patterns in the block lists which prevent further communication with CMPs to show the banners: `*cdn.opencmp.net/*`, `*consentmanager.net/*`, `*usercentrics.eu/*`.

[8] hausbau-forum.de detects uBlock and asks the user for deactivation. promipool.de is clickable but not scrollable.

5.4 Discussion

We now discuss the implications of our findings and present future research directions.

Paywalls vs. Cookie paywalls: Existing research [110] reports on the rise of two types of Internet paywalls—hard and soft. With hard paywalls, users cannot access the website without first buying a subscription. With soft paywalls, users can freely view a certain number of articles before they need to buy a paid subscription. In this paper, we highlight the use of cookie paywalls where users (1) have to pay to *not* opt-in to tracking or, (2) accept using a service with tracking, or (3) cannot access the website's content at all. From a monetary perspective, cookie paywalls are similar to hard paywalls, but overall they adversely impact the clients' privacy. Due to this new "pay or get tracked" model, users may be conditioned to accept tracking cookies rather than paying for their privacy. This can result in privacy laws like GDPR being less effective. Moreover, in the future, websites may charge unreasonably high prices that can further compel users to accept tracking as their default choice. Although previous researchers [56, 145] also highlight the deployment of manipulative and non-compliant consent pop-ups by different CMPs, they do not consider cookie paywalls. For instance, Toth *et al.* [145] report that CMPs like Quantcast provide configuration interfaces to set up cookie banners and restricted website access, *i.e.,* limited (or no access) to website content before interacting with the banners.

Cookie paywalls, Website Content, and Tracking Cookies: Websites that show cookie paywalls may offer important content to their clients. We find that many websites showing cookie paywalls are in top 1k of domains. Thus, users will either provide consent to tracking or pay to avoid it, as they do not want to cease access to the website content. Cookie paywalls have the potential to create two classes of Web users: those that can afford to not being tracked, and those who need to pay for services with their data. In the future, user studies can be conducted to estimate the "monetary value of the content" on cookie paywall websites.

Tracking cookies themselves are used to facilitate ad serving, thus bringing monetary value to the website. To see if there is a correlation between the number of tracking cookies a website sets for "accepting" users and the subscription price, we run an additional small experiment. As shown in Figure 5.6, we observe no meaningful linear correlation between the number of tracking cookies set when accepting tracking and the subscription price.

Circumventing Banners and Cookie paywalls: Interaction with cookie banners may be seen as a nuisance to some users. Thus browsers such as Firefox are working on automatically clicking the reject button (if available) on banners [15]. This approach as well as our tool, can lay the groundwork to also automatically interact with cookie paywalls in the future.

Presently, we can use ad-block extensions and filter lists to evade most cookie paywalls. There is, however, a risk that these extensions may also block necessary scripts, potentially disabling useful functionalities, or introducing security threats. Moreover,

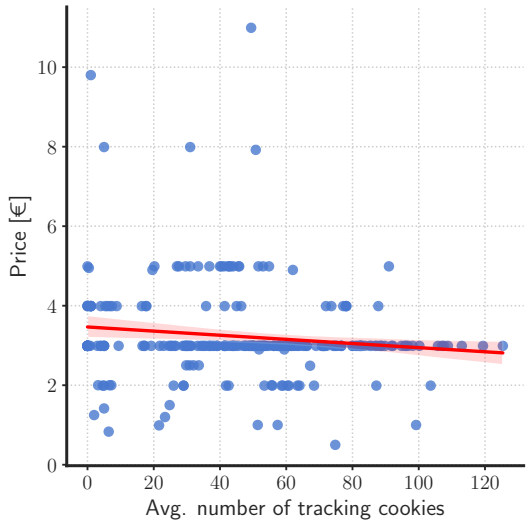


Figure 5.6: Correlation between the number of tracking cookies and price of cookie paywall subscriptions.

since ad-blockers run as a script on the client side, they can themselves be a source of privacy leaks.

Revoking Cookie paywall Acceptance: We find that it is not trivial to switch from cookie paywall acceptance to subscription. If a user has already consented to "accept" on some website's cookie paywall, they must delete their cookies and local storage (specific to the website). After deletion, they will see the cookie paywall on a subsequent visit and can change their choice. Since users will likely not be aware of these necessary additional steps, they might continue to be tracked even though they have subscribed *e.g.,* on a different device.

5.5 Summary

In this chapter, we performed the first automated analysis of the cookie paywall landscape to date. We modified the *BannerClick* to automatically detect cookie paywalls with a precision of 98.2%. Using this tool we crawled 45k websites and found cookie paywalls on 280 of them. We investigated different cookie paywall deployment characteristics and uncovered that they are especially deployed among popular websites in Germany (8.5%). Moreover, we compared cookie paywalls to regular cookie banners and found websites to be sending 42 times more tracking cookies to cookie paywall website visitors. Additionally, we assessed the prevalence of two large Subscription Management Platforms, which provide website operators with easily deployable cookie paywall solutions.

6

Stateful Banner Interaction and Web Cookies

The significant variation among consent banners (Section 2.3.1), the use of deceptive design patterns (Section 2.3.3), and the rise of cookie paywalls (Section 2.3.4) can lead users to unwillingly accept some cookie banners. Previous studies have shown that manipulative tactics affecting user decisions often result in uninformed or coerced consent [75, 88, 136]. For instance, users may accept cookie banners when there is no easy refusal option, resulting in the setting of tracking cookies. These unintentionally accepted cookies may increase the likelihood of being tracked during future web browsing. In fact, contrary to user expectations of seemingly isolated consent, a decision made on one website may influence tracking behavior on others due to shared third-party services and technical loopholes.

In this chapter, we investigate the technical and behavioral mechanisms underlying stateful banner interactions, as well as the subsequent workflows triggered by these interactions. Specifically, we disclose an unrelenting tracking mechanism exploiting web cookies. We find that numerous tracking cookies, *initially set upon banner acceptance on one website, continue to be transmitted to tracker domains even before users interact with the rejectable banners* on other websites. Unlike previous studies that focused on the *stateless deployment* of cookies on websites, we conduct a *stateful* interaction with banners across different websites, demonstrating how seemingly opposing decisions on one website (with accepted banner) can influence tracking behavior on subsequent websites (with rejected banner), ultimately resulting in the *transmission* of cookies to trackers. This not only violates privacy regulations such as the GDPR by undermining the overall effectiveness of banners as the de facto consent mechanism but also creates a false sense of privacy for users when they reject the banners. We refer to these cookies as **intractable cookies** (see Section 6.2).

To substantiate our findings, we conduct two measurement campaigns using *BannerClick* (see Chapter 3). To align its functionality with this study, we enhanced its performance, particularly improving rejection accuracy from 87% to 99%, as well as extending OpenWPM's capabilities in recording transmitted cookies (see Section 6.3). In the first campaign, we crawl the top 20,000 websites from Tranco, accepting cookie banners in the first half and measuring the number of intractable cookies on successfully rejected domains in the second half. For the second campaign, we randomly select and shuffle 20k sites from Tranco's top 50k, following a similar approach. Additionally, we perform both runs in reverse order to gain further insights. Overall, our main findings can be summarized as follows:

- We find that nearly 50% of websites send at least one intractable cookie to third-party tracking domains before obtaining explicit user consent (see Section 6.4.1).
- Regarding the effect of banner interaction, we see no immediate changes on intractable cookies after rejecting banners. However, on average, 25% of intractable cookies are not sent after reloading the webpage with the rejected banner (see Section 6.4.2).
- Furthermore, our measurements show that enabling the Global Privacy Control (GPC) signal in the browser can initially reduce intractable cookies by an average of 30%. An additional 32% reduction is achievable on subsequent visits by also rejecting cookie banners (see Section 6.4.3).
- We observe a notable trend where more popular Tranco websites send fewer intractable cookies compared to less popular ones. Specifically, the top 50 websites send, on average, zero intractable cookies, while the top 10k websites send 25 intractable cookies on average (see Section 6.4.4).
- We note that websites using CMP banners send more than 6.9 times as many intractable cookies compared to those using native banners. Moreover, out of 5,915 accepted domains, 90 websites with cookie paywalls are responsible for setting more than 35% of intractable cookies (see Section 6.4.5).
- We also observe $\approx 60\%$ of intractable cookies have an expiration time of at least 10 days, highlighting their persistence. In addition, around 90% of intractable cookies are set (or refreshed) by at most 1% of accepted websites (see Section 6.4.6).
- Our analysis shows that, on average, each domain has 3.42 different trackers, with each tracker receiving an average of 7.3 intractable cookies. We also verified that the top 20 trackers are indeed well-known tracking companies (see Section 6.4.7).
- Additionally, to assess the effectiveness of partitioned cookies in mitigating intractable cookies by restricting their transmission to *setter website*, we conduct a separate measurement using Chrome. The results show that only 1.3% of all unique tracking cookies are partitioned, with more than half accompanied by non-partitioned cookies from the same tracker domain (see Section 6.4.8).

Finally, our findings offer deeper insight into the complexities of web tracking and the gap between privacy regulations and technical implementations. This understanding lays the groundwork for our discussion in Section 6.5, where we examine GDPR provisions, their limitations in enforcing effective consent mechanisms, and our proposed browser-integrated solution.

6.1 Related Work

In addition to the previously mentioned studies [11, 30, 31, 76, 83, 134, 137] that investigate the prevalence of online tracking and the effects of privacy regulations, some

References	Automated	Reject Coverage	Stateful	Sent Cookie
Trevisan et al. [148]	✓	CMP	✗	✗
Matte et al. [91]	✗	CMP	✗	✗
Jha et al. [68]	✓	N/A	✗	✗
Smith et al. [135]	✓	CMP	✗	✗
Rasaii et al. [120]	✓	87%	✗	✗
Our work	✓	99%	✓	✓

Table 6.1: Overview of the closest previous studies on the misbehavior of tracking cookies.

studies have evaluated the influence of consent banner design on user behavior, specifically their acceptance or denial of consent [13,75,88,136]. Santos et al. [128] analyzed the clarity of cookie banners and found that 61% of them employed vague language, failing to specify privacy policies adequately. Utz et al. [154] explored additional factors influencing user consent, such as banner placement, and reported significant impacts on consent decisions based on these elements. Notably, Nouwens et al. [104] show that merely removing the opt-out button from the first view of banners increases consent rates by approximately 23%. Overall, these studies consistently identify interface interference as a key factor that significantly influences how users interact with banners.

Moreover, many studies specifically analyze the impact of banner interaction on cookie installation. Table 6.1 compares some of the most relevant studies [68, 91, 120,135,148] with our work based on key distinguishing factors, such as whether they were conducted manually or through automation, their rejection coverage, whether they employ stateful or stateless measurements, and whether they focus on the setting of cookies on browsers or sending them to tracker domains. Trevisan et al. [148] developed CookieCheck, a tool that visits websites as a new user and analyzes installed cookies. It focuses solely on profiling cookies set by CMPs that violate the ePrivacy Directive before any user consent is given. Matte et al. [91] used semi-automatic crawl campaigns to detect suspected GDPR and ePD violations in IAB Europe's Transparency and Consent Framework banners. Jha et al. [68] attempted to interact with cookie banners in an automated manner to observe differences in the cookies set. However, their work focused solely on cookie acceptance based on privacy policies. Smith et al. [135] specifically investigated the placement of tracking cookies under the guise of legitimate interest by CMPs, as well as their compliance with properly transmitting users' choices through TC strings. Finally, Rasaii et al. [120] presented in Chapter 4, have conducted a comprehensive measurement study on Tranco's Top 10K sites, analyzing cookies deployment in a stateless manner.

In this research, we aim to investigate how previously accepted cookies on one website may contribute to user tracking on subsequent visits before any explicit consent is given. We use an improved version of BannerClick to detect and interact with banners, exploring privacy violations in a stateful manner. To the best of our knowledge,

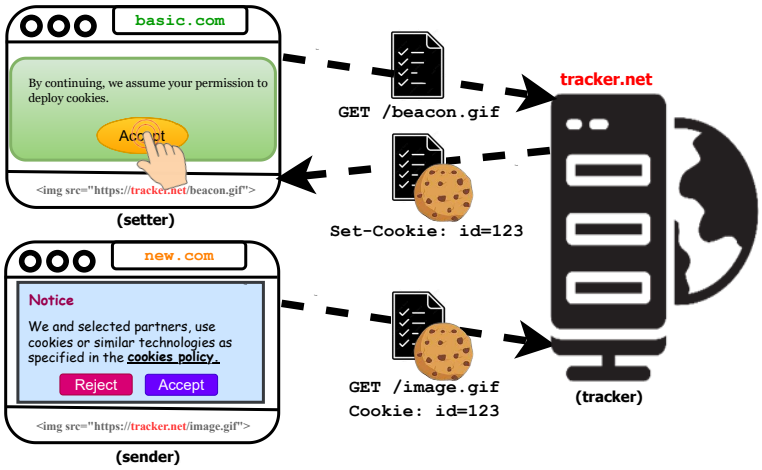


Figure 6.1: The entities involved in intractable cookies transmission.

this specific form of user tracking via web cookies has not been previously explored in the literature.

6.2 Intractable Cookies

In this section, we detail the nature of intractable cookies and the entities involved, as illustrated in Figure 6.1.

In the depicted scenario, a user initially visits `basic.com` and "accepts" the cookie banner. This might result in setting new cookies on the browser. For example, the browser starts sending HTTP requests to load a third-party resource, `beacon.gif`, from `tracker.net`, which sets a cookie (*id*=123). We refer to `basic.com` as the *setter website*, as it establishes the context for the initial cookie-setting in the user's browser, and to `tracker.net` as the *tracker domain*, as it tracks the user. Later, the user accesses `new.com`, which shows a banner with the "reject" option[1]. In this scenario, even though the user has not yet interacted with the banner to explicitly consent to the cookies, during the rendering of the webpage, the browser might send an HTTP request to load third-party resources. For example, `new.com` might embed a resource (`image.gif`) from `tracker.net`, resulting in sending an HTTP request along with the previously stored cookie (*id*=123) back to the *tracker domain*. In this context, `new.com` is referred to as the *sender website*, as it leads to sending the cookie to the *tracker domain*. We call these cookies that are previously set by an accepted *setter website* and then sent by the *sender website* to the *tracker domain* before explicit acceptance of the cookie banner as **intractable cookies**.

[1] We focus on websites with rejectable banners because the concept of banners and unwanted tracking behavior is directly influenced by privacy regulations such as GDPR. A key aspect of these regulations is the user's ability to reject cookies. Therefore, if a banner lacks a reject option, the website is excluded from the study.

It should be noted that existing research primarily assesses the setting of tracking cookies prior to explicit consent in a stateless manner. However, in our study, we introduce intractable cookies to address two key factors overlooked by previous studies. First, our measurements are conducted in a stateful manner. This approach is crucial because internet browsing—and consequently tracking—occurs over sessions and time, with websites often interacting with one another. Second, we focus on the transmission of cookies rather than solely their deployment in browsers, as setting cookies alone does not necessarily violate privacy laws. In addition, as demonstrated in Section 6.4.9, the setting of intractable cookies by *sender website*[2] is relatively uncommon, while many cookies are transmitted to trackers without being set in browsers. Therefore, treating a website as a single, isolated entity and analyzing the deployment of tracking cookies in a stateless manner may fail to capture the full extent of unwanted tracking practices in the wild and subsequently hinder regulators from crafting the most effective regulatory measures.

We elaborate more on the relation between the intractable cookies and privacy regulations as well as possible mitigation approaches like partitioned cookies [52] and their challenges in Section 6.5.

6.3 Methodology

We now describe our data collection approach, enhancements to the OpenWPM platform, improvements to the BannerClick tool for more accurate banner interaction, and the cookie classification method, followed by a discussion of measurement limitations and ethical considerations.

6.3.1 Data Collection

We utilize the popularity index of Tranco as our target list to overcome the shortcomings of current popular site lists, which include instability, unreachable domains, and domains that can be easily manipulated by adversaries [113].

Similar to previous chapters, we use *BannerClick* [120] to conduct large-scale automated crawls and collect data for further analysis. We modify and improve its accuracy in detecting and interacting with banners. This version (v0.26.0) is available in the *BannerClick* git repository [118].

Figure 6.2 depicts the overall scheme of our methodology setup. Each measurement consists of two phases: stateful and stateless. In the stateful phase, BannerClick crawls the first half of the target list. Upon successful *acceptance* of a banner, We aggregate the corresponding cookies, along with their *setter website*, into a database called *Cookie Jar*. At the end of this phase, we store the browser profile to use as the base profile in the next phase.

[2]For instance, visiting `new.com` might also set the cookie (*id*=123), which can be considered more detrimental than merely sending it.

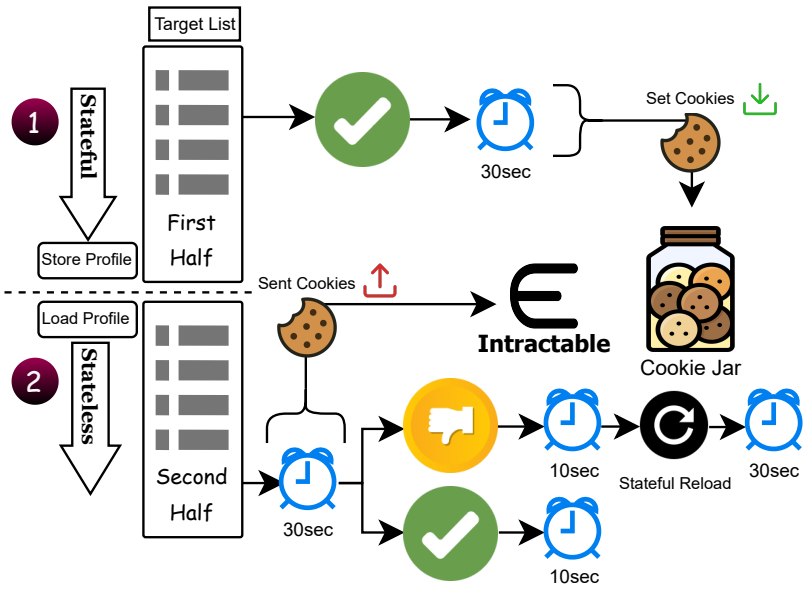


Figure 6.2: Overview of our methodology for measuring *intractable cookies*.

Next, BannerClick crawls the second half of the target list in a stateless manner. For each new iteration, it loads the final profile from the first phase, making each iteration stateful with respect to the accepted domains. After accessing the domain, it waits for 30 seconds and collects all cookies sent through HTTP requests. These cookies are referred to as *Sent Cookies* and are later compared with those in the *Cookie Jar* to detect intractable cookies. In this step, BannerClick interacts with the banner in two separate iterations: one for rejecting and one for accepting. Following each interaction, it waits 10 seconds to capture any immediate changes. Additionally, to assess whether rejection reduces the number of intractable cookies on subsequent visits, we reload the webpage and wait another 30 seconds. The reload event is conducted in a stateful manner, with browser caching disabled to ensure any inconsistencies are related to banner rejection.

We conduct two measurement campaigns from a server hosted in the EU. Table 6.2 details these measurement campaigns:

1. **Regular runs**: In this measurement campaign, we methodically follow Tranco's top 20k websites, first attempting to "accept" banners from the top 10k websites

Run	#Sites	#Accepted	#Rejected	Duration
Regular	20k	3,034	2,379	15 days
Regular-Reverse	20k	3,060	2,424	15 days
Random	20k	2,933	2,578	16 days
Random-Reverse	20k	2,983	2,543	16 days

Table 6.2: Overview of different measurement types.

and then detecting intractable cookies on the bottom 10k. This process is reversed and conducted to explore the impact of website rankings. We refer to these runs as *Regular* and *Regular-Reverse*.

2. **Random runs**: To further assess the intractable cookies on a randomized configuration, we sample and mix up 20k domains from Tranco's top 50k domains list. This shuffled list is then split and examined in two sets, as in the regular runs. We call these runs *Random* and *Random-Reverse*.

6.3.2 Modification to OpenWPM for Stateful Runs

We use OpenWPM to conduct a combination of stateful and stateless crawls aiming to observe the interplay of the *setter* and the *sender website* as mentioned earlier. OpenWPM triggers a new event whenever cookies are set, altered, or deleted and stores the corresponding data in the database. We have adapted its functionality to ensure that every event involving the addition or updating of a cookie results in overwriting the existing cookie in the browser with a new expiration time (Saturday, 01 Jan 2028, 12:12:12 GMT). This modification is crucial because many cookies might otherwise expire before the completion of our measurements (see section 6.4.6). Moreover, a cookie can be deleted by overwriting its expiry time before the current time. We recognize such cases and allow them to proceed without extending the expiration time. In addition, we frequently store the browser profile to resume from the last saved profile upon unexpected crashes.

Furthermore, we have integrated a request parser within OpenWPM that processes HTTP requests by parsing the headers and storing all associated cookies along with the corresponding *sender website*. This enhancement facilitates the comparison of cookies retrieved from HTTP requests on domains where banners are rejected (*i.e., Sent Cookies*) with those stored in *Cookie Jar*.

6.3.3 Improving BannerClick Rejection Coverage

To cover a broader range of banners in our study, we enhance BannerClick's accuracy in detecting and interacting with banners, particularly newer types of banners that it was previously unable to handle. Notably, we observe that many banners now provide buttons allowing users to accept selected preferences through the settings view. Since the preselected options typically consist only of essential cookies, clicking these buttons is expected to reject tracking cookies. Figure 6.3 shows an example of a such banner after clicking the "Settings" button. In this example, all non-essential cookies are disabled by default. To handle these cases, we modify BannerClick so that if it cannot find the "reject" button in the banner's main view, it attempts to click the "settings" button. If successful, it searches for HTML elements related to the settings view, then tries to click "Reject All" or similar buttons, followed by options like "Confirm," "Save," "Accept Selected," or equivalent.

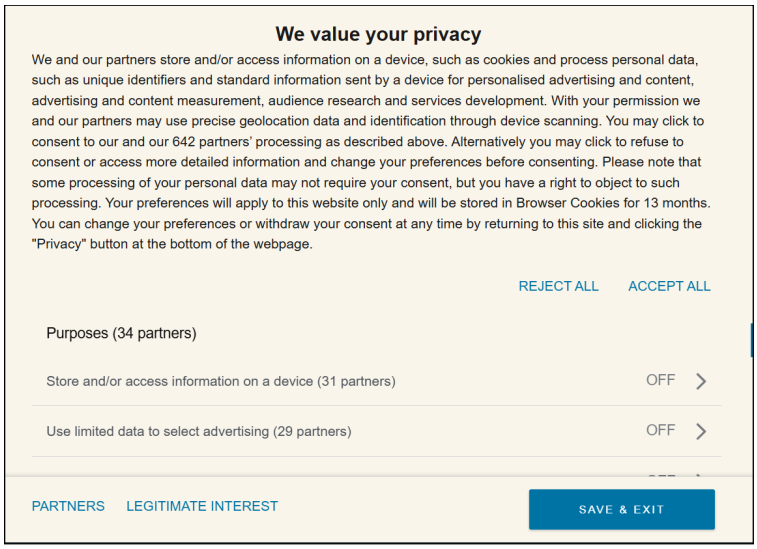


Figure 6.3: Example of a banner with preselected options after clicking the settings button.

To evaluate the accuracy of banner rejection after the changes, we randomly select 1k domains from the top 20k in the Tranco list and attempt to reject their banners. Out of 351 websites with detected banners, BannerClick successfully rejects 263 of them, of which 39 are rejected by clicking on the settings option (20 with preselected options and 19 with a "reject all" option), while the remaining 224 are directly rejected. We manually reviewed the screenshots and found that only one is a false positive and four are false negatives[3], resulting in improving the rejection accuracy from 87% [120] to around 99%. Note that among the rejected banners, there were no cases where the preselected preferences included non-essential cookies. We plan to publish our modified versions of OpenWPM and BannerClick.

6.3.4 Cookie Classification

To categorize cookies as first-party or third-party, we utilize the latest Public Suffix List [100] to determine the effective top-level domains (eTLD+1) of both the visited websites and the cookies' `host` attribute. We then compare the domain of each cookie to the domain of its *setter website*. If they match, the cookie is classified as first-party; if not, it is deemed third-party.

Furthermore, similar to previous works [54, 120], we utilize the *justdomains* blocklist [70] to identify *tracker domains*. This list aggregates entries from various popular tracking lists, including EasyList, EasyPrivacy, AdGuard, and NoCoin filter lists. To obtain the most recent version (*i.e.,* February 2025), we convert the blocklists into the equivalent justdomain lists using the JustDomain converter script [69]. Next, we

[3]The false positive occurs when clicking a "Não, ajustar" button that opens the settings view (*i.e.,* `kinghost.com.br`). False negatives are mainly due to language limitations and uncommon cookie banner designs, such as slide-out panels triggered by the settings button (*e.g.,* `edg.io`).

compare the domains in the lists with the host attribute of the cookies. If there is an exact match or if the host ends with a domain from the list, preceded by a period ('.'), we classify the cookie as a tracking cookie[4].

Subsequently, we categorize a tracking cookie as intractable if it is set on an accepted website in the first phase of the run (*i.e.,* stateful phase) and later sent to the tracker from a website in the second phase of the run (*i.e.,* stateless phase) before the banner is successfully rejected.

6.3.5 Measurement Limitations

Despite our best efforts to eliminate bias from our measurements, we acknowledge that our study does not fully capture the variety of real-world scenarios users encounter while browsing. First, the individual browsing behaviors are far more complex than the direct crawling of a set of ordered domains. Second, website responses may vary based on several factors, including user agents, user activities such as scrolling, and variations in browser settings and capabilities. For example, a study shows that further user interaction and deeper crawling can lead to a 36% rise in the use of tracking technologies like cookies [153].

Moreover, classifying cookies as trackers also presents inherent challenges due to the lack of definitive references on their actual usage. Among the many available approaches, each with its own limitations and advantages [20, 57, 101], we identify tracking domains using the just-domain block filter list due to its widespread use and comparability with related studies. However, these lists are crowdsourced, meaning they are continuously updated and maintained by volunteers. As a result, they may overlook certain cases or contain misconfigurations in their exception procedures. For instance, during manual inspection, we identified several third-party cookies classified as intractable cookies, owned by well-known trackers such as `doubleclick.net` and `demdex.net`, that were not included in the final just-domain lists. In particular, we found that `doubleclick.net` used a cookie named `IDE` as an intractable cookie on 2,300 rejected websites. According to Google documentation [53], this cookie is used to record users' interactions with websites' front-end for advertising analysis.

6.4 Results

Throughout our analysis, to eliminate the potential impact of the number of accepted domains on the cookies stored in *Cookie Jar*, we normalize the number of accepted domains when comparing two runs. Specifically, for the run with a larger number of accepted domains, we compiled a new version of *Cookie Jar* where the cookie *setters* are randomly selected from a number of accepted websites equivalent to that in the run with fewer accepted domains.

[4] We iterate over all entries in the filter lists and compare them with the host of cookies using the following condition:

```
if host == entry or host.endswith('.' + entry) then True
```

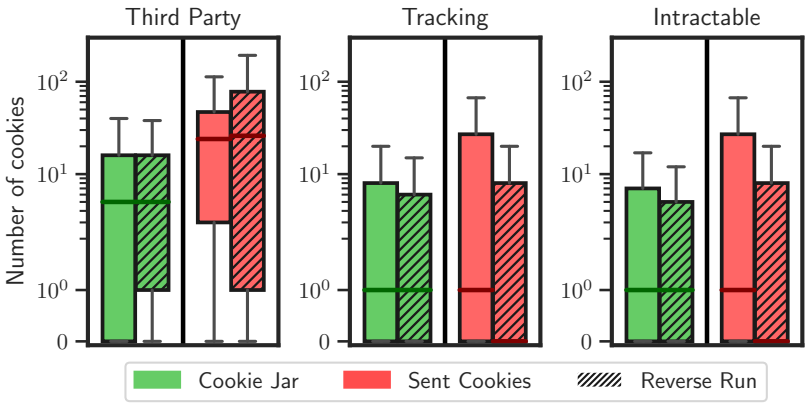


Figure 6.4: Cookie distribution for Regular and Regular-Reverse runs.

6.4.1 Cookie Distribution

Figure 6.4 and Figure 6.5 show the overall distribution of third-party, tracking, and intractable cookies across regular and random runs respectively. Cookies set in accepted domains in the stateful phase (*i.e., Cookie Jar*) and those sent via HTTP headers before the successful rejection of banners in the stateless phase (*i.e., Sent Cookies*) are depicted using green and red boxplots, respectively. Note that, for measuring intractable cookies in *Cookie Jar* (*i.e.,* intractable category green boxplots), we check if a cookie with the same `name` and `host` (the *setter website* might be different) is later sent as intractable cookies. Therefore, when we refer to a cookie in *Cookie Jar* as intractable, it does not imply that it is inherently an intractable cookie. In other words, intractable cookies are essentially third-party tracking cookies that are set with user permission on one website but later sent and propagated covertly without user permission to other websites across stateful browsing sessions, ultimately subverting the core purpose of cookie banners.

In Figure 6.4, for tracking cookies, we see nearly identical *Cookie Jar* boxplots (with median one) across both *Regular* (green) and *Regular-Reverse* (shaded-green). However, more intractable cookies are sent in the *Regular* run compared to *Regular-Reverse* run on average (*i.e.,* the median in the red boxplot is one while it is zero for the shaded-red boxplot). This might indicate that either more popular websites set or the less popular ones sent more intractable cookies. We further investigate the impact of the website ranking in Section 6.4.4. For the random runs (Figure 6.5), the medians of intractable cookies are zero for both, indicating relatively consistent behavior. In total, it is evident that intractable cookies are common across websites. In all runs, the majority of tracking cookies set in the *Cookie Jar* are identified as intractable, *e.g.,* out of 3,583 unique tracking cookies in the *Regular* run, 2,131 cookies are later classified as intractable (see Section 6.4.9 for details).

To further analyze the intractable cookie distribution, we plot the Empirical Cumulative Distribution Function (ECDF) of intractable cookies for regular and random runs (see Figure 6.6). The graphs indicate that around 45% to 55% of websites

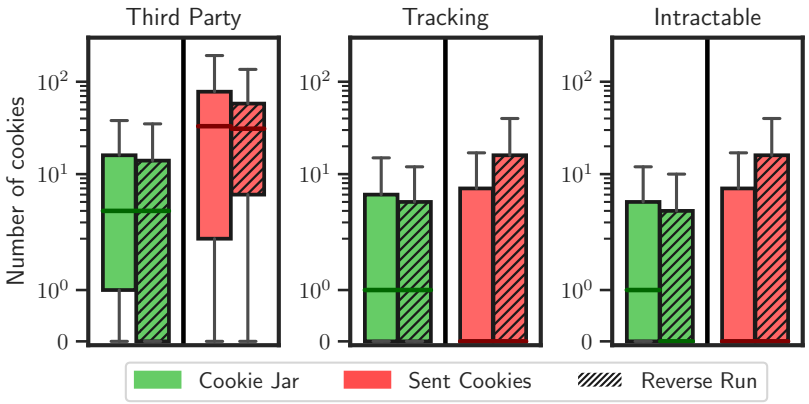


Figure 6.5: Cookie distribution for Random and Random-Reverse runs.

send at least one intractable cookie across all runs. Moreover, we observe on average $\approx 40\%$ more intractable cookies in the *Regular* run compared to the *Regular-Reverse* run, while this difference is relatively smaller for random runs, again indicating the possible impact of website ranking on intractable cookies (see Section 6.4.4).

For a more comprehensive analysis of the impact of banner interaction, website rank, and cookie banner types on intractable cookies in subsequent sections, we use a combined version of the random runs, collectively referred to as the *RandComb* run.

To summarize: *Nearly 50% of websites with rejectable banners still send at least one intractable cookie to tracking domains.*

6.4.2 Impact of Banner Interaction

By definition, intractable cookies are sent to tracker domains prior to any interaction with the banner. We now investigate the possible effect of banner interaction on

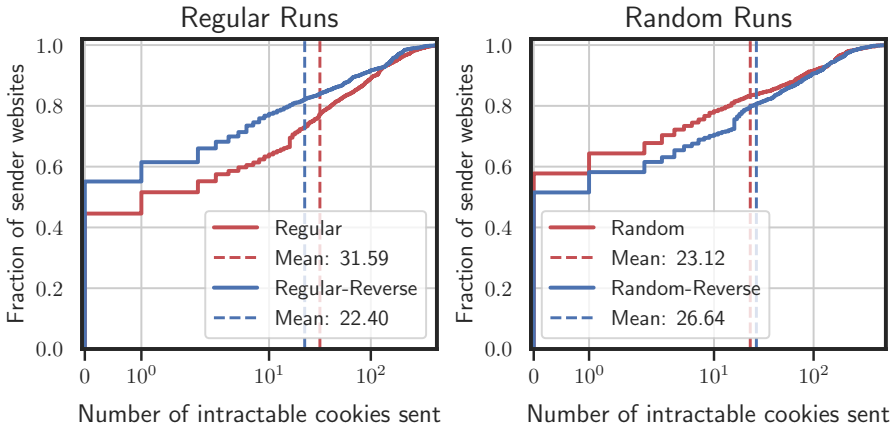
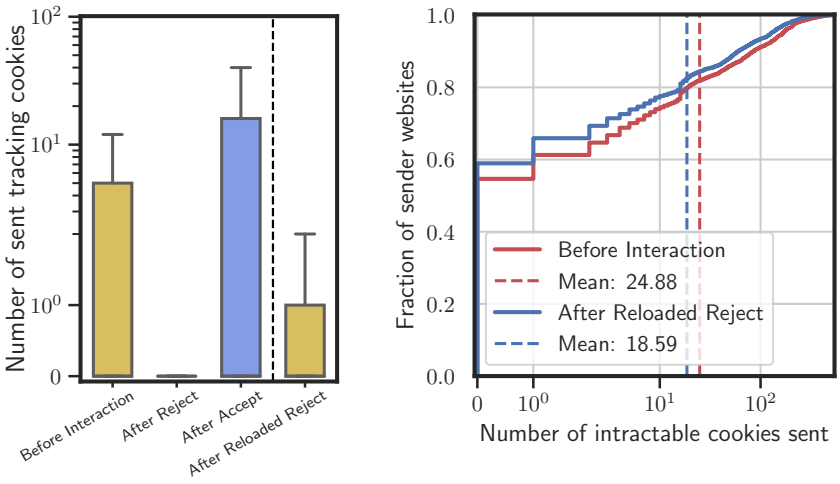


Figure 6.6: ECDF plot for Intractable cookie distribution over websites.



(a) Impact of Banner Interaction

(b) Impact of Reload

Figure 6.7: Impact of banner interaction and reloading on the number of tracking cookies in Cookie Jar sent.

the modification or transmission of intractable cookies. As mentioned in Section 6.3, for each domain in the stateless phase, BannerClick performs two separate iterations: one for rejecting the banner and then reloading the webpage, and one for accepting.

The first three boxplots in Figure 6.7a show the number of tracking cookies sent upon visiting a webpage that were previously stored in the *Cookie Jar*, corresponding to three stages: before interaction (*i.e.,* intractable cookies), after rejecting, and after accepting the cookie banner for *RandComb* run. The flat line for the "After Reject" box indicates that rejecting the banner does not trigger sending additional tracking cookies previously set. We also find that just a few cookies are explicitly deleted after rejecting the banners. In contrast, accepting the banner triggers the transmission of a new set of cookies (*i.e.,* the blue box), which can be considered a valid action since it reflects user consent to the banner.

Additionally, the right-most box in Figure 6.7a shows the number of intractable cookies continuing to be sent after revisiting the rejected website (*i.e.,* a subset of intractable cookies). Compared with the "Before Interaction" box, we see the third quartile dropping from 5 to 1. To further assess the impact of the reload, we plot the ECDF graph of intractable cookies in Figure 6.7b, comparing the stages before interaction and after reloading the rejected webpage. We observe, on average, websites stop sending $\approx 25\%$ of intractable cookies after reloading the rejected webpages.

Overall, considering the "After Accept" stage, we observe that initiating tracking after user interaction with banners is technically feasible. However, as the prevalence of intractable cookies suggests, the current orchestration of web entities often prioritizes tracking before obtaining consent (*i.e.,* the "Before Interaction" stage).

A Case Study: Through our manual inspection, we observed that interacting with a banner can modify a website's front-end configuration. For instance, rejecting a banner may exclude third-party tracking resources from the HTML source. Conversely,

accepting a banner may prompt websites to inject additional third-party resources into the HTML, leading to the transmission of new cookies.

Notably, we found that one of the major entities that govern the website's behavior regarding loading third-party resources is the Consent Management Platform (CMP) (see Section 6.4.5 for more details). For example, `ritzcarlton.com` uses `OneTrust`, one of the most popular CMP [59, 120], to manage user preferences and interactions with the banner. Upon visiting `ritzcarlton.com` for the first time, it renders an iFrame from `demdex.net`, a domain associated with Adobe Audience Manager[5]. In this case, the script inside the iFrame creates new cookies and sends them via XMLHttpRequest API calls to different tracker domains. By default, the browser also sends all previously set cookies along with these requests. Tracker domains can potentially link these cookies together and create a user profile using cookie synchronization techniques [109]. After the user rejects the banner, the CMP stores the user's preference in a cookie, stopping the iFrame from loading and preventing further requests to trackers on subsequent visits to the webpage.

In total, out of the 127,645 intractable cookies detected in the *RandComb* run, around 73% are sent as a result of HTTP requests made by the browser to fetch third-party resources (*e.g.,* img, script, beacon, etc.). The remaining 27% cookies are directly sent via XMLHttpRequest API calls (*e.g., fetch()* method) from script codes.

To summarize: *Rejection of banners does not have any immediate effect on the number of intractable cookies. However, it prevents sending around 25% of intractable cookies in subsequent visits.*

6.4.3 Impact of Global Privacy Control

In addition to interacting with cookie banners, other official, standardized mechanisms have also been developed. One of the most notable is Global Privacy Control (GPC) [45], which has recently gained more attention and is now supported by many browsers and extensions[6]. GPC is a browser setting that signals a user's preference not to be tracked. When enabled, the browser informs websites that users do not want their data to be sold or shared. In 2021, the California Attorney General confirmed that businesses must honor the GPC signal as a valid request to opt out of data sales under CCPA [18]. While GDPR does not explicitly mandate GPC, the signal can be interpreted as a withdrawal of consent for tracking, which websites operating under GDPR should honor.

To measure its impact on the intractable cookies, we perform another iteration in the stateless phase with the GPC signal enabled. Figure 6.8 shows the difference in intractable cookies between the *RandComb* run with and without GPC enabled. We observe that, on average, intractable cookies are reduced by $\approx 30\%$ when GPC is enabled. Notably, around 68% of the remaining intractable cookies overlap with

[5] Adobe Audience Manager is a Data Management Platform (DMP) that collects, manages, and segments user data for personalized advertising and audience targeting.

[6] https://globalprivacycontrol.org/orgs

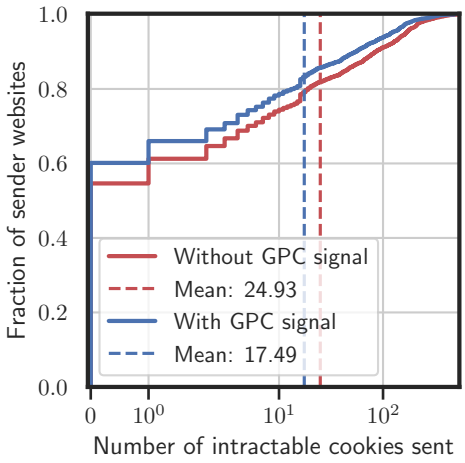


Figure 6.8: Impact of enabling GPC on the number of intractable cookies.

those from the "After Reloaded Reject" shown in Figure 6.7. This indicates that enabling the GPC signal reduces intractable cookies by 30%, and further rejection of the banner can lead to an additional 32% reduction on subsequent visits, resulting in an average of approximately 11.9 intractable cookies per domain.

To summarize: *Enabling the GPC signal reduces the number of intractable cookies by 30% on average initially, with a further 32% reduction possible on subsequent visits by rejecting the banners.*

6.4.4 Impact of Website Rank

Figure 6.9 depicts the average number of intractable cookies set and sent by Tranco websites as per their rank tier. For this analysis, we use the *RandComb* run, which contains 5,915 accepted domains in the stateful phase and 5,121 rejected domains in the stateless phase. The red line shows the average number of intractable cookies sent based on the top list rank of the rejected domains. Accordingly, the blue line shows the average number of set cookies over the top list rank of the accepted domains.

For the red line, we observe an ascending trend where popular websites, on average, send fewer intractable cookies than less popular ones. For instance, we observe that the top 50 websites send, on average, zero intractable cookies[7] . Whereas the top 10k websites send, on average, about 25 intractable cookies. This trend may be explained by the tendency of more popular websites to utilize their own resources, potentially resulting in fewer HTTP requests to third parties and, consequently, fewer intractable cookies being sent. Conversely, the blue line remains constant, with websites setting around 5 intractable cookies on average, regardless of their relative ranking.

[7] This observation is supported by our analysis of regular runs, where none of the 6 rejected domains in the top 50 sent intractable cookies.

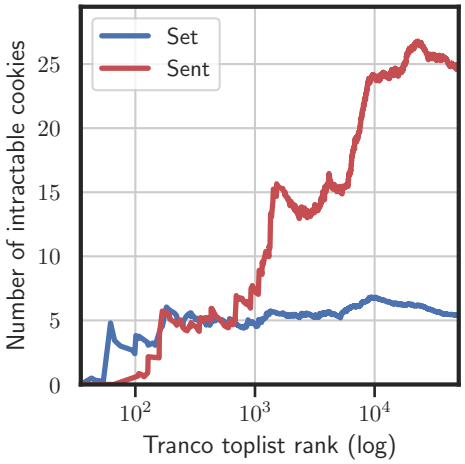


Figure 6.9: Average number of set and sent intractable cookies over Tranco's toplist ranks.

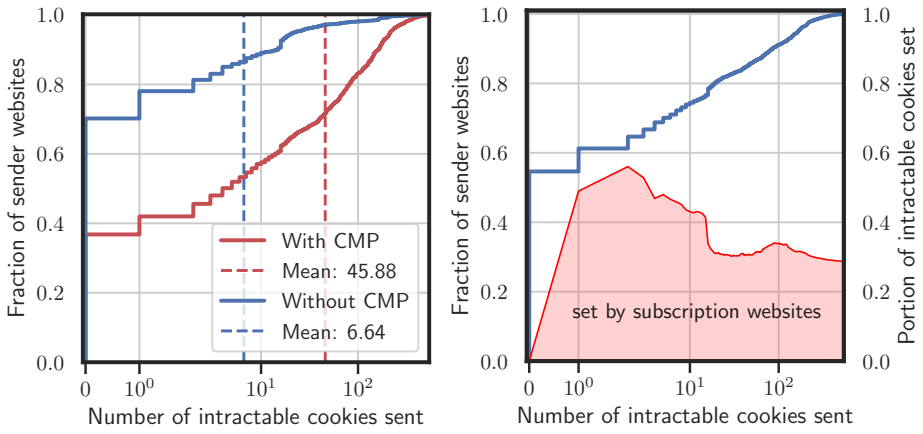
To summarize: *It can be concluded that the rank of websites directly influences the number of intractable cookies, with less popular sites typically sending more.*

6.4.5 Impact of Banner Type

CMP Banners: As mentioned in Section 6.4.2, Consent Management Providers (CMPs) are another entity involved in the context of intractable cookies. Out of 5,121 rejected banners in the *RandComb* run, BannerClick classifies 2,386 as CMP banners. Figure 6.10a illustrates the difference in the number of intractable cookies between websites with and without CMPs. On average, websites embedding CMP banners set 6.91 times more intractable cookies than those using native banners, highlighting a significant discrepancy in cookie deployment. We also find that CMP banners are relatively harder to reject, as $\approx 40\%$ of them require exploring of settings menu involving more than one click to be fully rejected. Whereas over 90% of native banners have a direct reject button and can be rejected with a single click. Overall, this suggests that CMPs may not fully achieve their intended purpose of facilitating user consent and compliance with privacy regulations, as they are generally harder to reject and tend to transmit more intractable cookies[8].

Cookie Paywalls: Recently, a new form of cookie banner called "cookie paywalls" requires users to either opt-in for banner policies (mostly tracking) or pay the cost to reject the cookie paywall through subscription [119]. Figure 6.10b displays the ECDF graph for the *RandComb* run, highlighting the portion of intractable cookies set by websites that display cookie paywall-based banners (totaling 90 detected by *BannerClick* out of 5,915 accepted websites). Note that, following the blue line, 60%

[8]These differences may be due to the likelihood that websites with complex ad interdependencies are more inclined to use CMPs and communicate with tracker domains.



(a) Impact of CMP

(b) Impact of Cookie Paywall

Figure 6.10: Impact of type of banners on the number of intractable cookies for RandComb run.

of websites send a maximum of 2 intractable cookies, of which around 55% of the cookies are set or reset by websites with cookie paywall banners (peak of the red area in the graph). Interestingly, this proportion drops to around 30% and remains relatively constant for websites that send more than 20 intractable cookies. Overall, this shows that even if users somehow manage to reject all other banners, they still get substantial intractable cookies from a few cookie paywalls.

To summarize: *Websites with CMP banners sent around 6.9 times more intractable cookies. In addition, websites with cookie paywalls are responsible for setting* $\approx 30\%$ *of intractable cookies.*

6.4.6 Expiration and Duplication Analysis

As mentioned in Section 6.3, we standardize the expiration times of all retrieved cookies to a uniform date far in the future (Saturday, 01 January 2028, 12:12:12 UTC) to increase the consistency of our measurements and enhance the reproducibility of the findings. We recognize that this approach of artificially extending cookie expiration times may introduce bias into our analysis. Additionally, the diversity and the number of cookies users encounter can vary based on their individual browsing behaviors and the websites they visit. Thus, we examine the actual expiration times of intractable cookies in the *Cookie Jar* and their frequency across accepted websites to better assess the validity of our results.

Figure 6.11 illustrates the distribution of unique[9] intractable cookies based on their expiry time durations and the number of websites setting them for regular runs. On the x-axis, expiry times are segmented into intervals by 1 day, 10 days, 3 months, 1 year, and 'Session' as a special category for session cookies. These are the common

[9]The cookies are grouped by the **name** and **host** attributes.

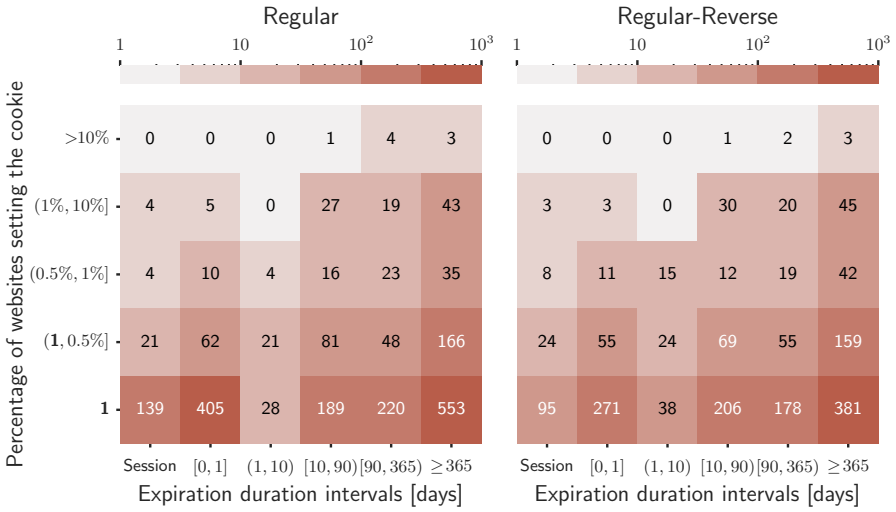


Figure 6.11: Distribution of intractable cookies based on the number of websites setting them and expiration time.

expiry durations for cookies. For example, nearly 25% of intractable cookies have an expiry time set to 1 year. The y-axis shows the buckets for the percentage of accepted websites setting intractable cookies, including a unique category for cookies set once. The numbers displayed within the heatmap cells represent the count of intractable cookies. For instance, the bottom-right cell of the heatmap for the *Regular* run shows that 553 unique intractable cookies have an expiry time exceeding one year and were set by a single website during the entire crawl. Both heatmaps show that about 90% of cookies are set by no more than 1% of websites. Furthermore, our analysis reveals that approximately 40% of these cookies have expiration dates of at least one year suggesting a common practice among individual websites to track visitors over extended periods. A similar pattern is observed in the random runs.

Although cookies are distributed quite sparsely across websites, over 65% of them are set with an expiry time of at least 10 days. This corresponds to the duration required to complete the stateful phase and fill the *Cookie Jar* in all of our four runs. Thus, for over 65% cookies, we can state that our adjustment of the expiration times does not overestimate the prevalence of intractable cookies. However, we cannot conclusively conclude about session cookies and those with smaller expiration times. Nevertheless, we apply these adjustments primarily to ensure reproducibility for further studies and consistent behavior across different runs, while real-world user browsing behavior remains far more complex and influenced by numerous additional variables.

To summarize: *Most intractable cookies persist in the browser for at least 10 days and are set only once throughout the entire crawl.*

	Tracker Domain	#Cookies	#UniqCookies	#Senders
1	amazon-adsystem.com	1,526	2	763
2	adsrvr.org	1,344	2	672
3	criteo.com	2,623	7	649
4	pubmatic.com	19,905	40	621
5	adnxs.com	3,020	5	604
6	casalemedia.com	2,128	4	532
7	id5-sync.com	1,054	7	517
8	openx.net	1,203	3	401
9	smartadserver.com	2,074	7	375
10	sharethrough.com	2,357	12	366
11	3lift.com	1,240	5	365
12	lijit.com	1,084	13	343
13	bidswitch.net	2,723	11	323
14	a-mo.net	5,458	18	311
15	taboola.com	6,118	42	293
16	nr-data.net	282	1	282
17	bidr.io	414	2	282
18	tapad.com	771	3	257
19	liadm.com	444	2	243
20	quantserve.com	454	2	227

Table 6.3: Top 20 *tracker domains* sorted by the total number of intractable cookies associated with them along with the number of unique intractable cookies and the *sender websites* sending them.

6.4.7 Domain Analysis

Having examined the prevalence of intractable cookies (see Section 6.4.1), we now analyze the prevalence of *sender website* and *tracker domain* in their transmission.

The ECDF plots in Figure 6.12 compare the number of intractable cookies and the number of *tracker domains* (responsible for eliciting them) corresponding to the *sender websites* for the *RandComb* run. On average, each *sender website* is associated with 3.42 different *tracker domains*, each with at least one intractable cookie. Furthermore, based on the means, a single *sender website* has an average of 7.3 intractable cookies per *tracker domain*.

In Table 6.3, we present the top 20 *tracker domains* along with the total and unique number of intractable cookies associated with them, sorted by the number of *sender websites* responsible for dispatching them. We observe that most top trackers use only a handful of unique cookies to track users across hundreds of the 5,121 rejected websites. We manually verified that the majority of the top 50 domains owning intractable cookies belong to recognized ad tech companies specializing in programmatic advertising solutions (*e.g.,* `pubmatic.com`), while others provide analytic tools and feedback through session recordings and surveys (*e.g.,* `hotjar.com`).

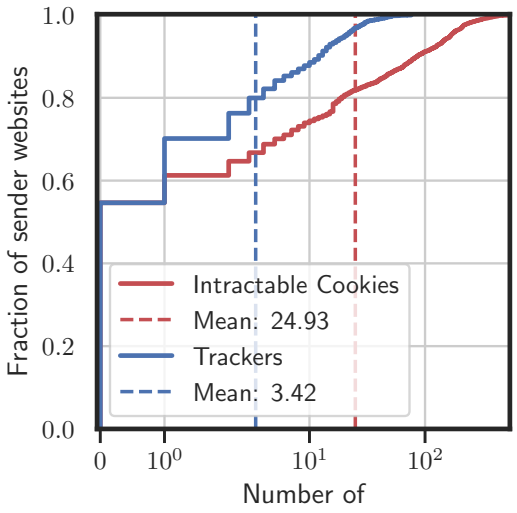


Figure 6.12: A comparison between the number of intractable cookies and the number of associated trackers per *sender website*.

Cookie Type	Total Unique	Partitioned	Along with NP
All Cookies	79,898	521	-
Tracking Cookies	3,177	40	26

Table 6.4: Summary of unique partitioned cookies and those accompanying the non-partitioned (NP) cookies from the same tracker domain.

To summarize: *On average, there are 3.42 different tracker domains per sender website, with each tracker domain receiving an average of 7.3 intractable cookies from the given sender website. We also verified that the top trackers are recognized tracking companies.*

6.4.8 Partitioned Cookies Analysis

Partitioned cookies, a newly proposed feature in Chrome [52] also adopted by other browsers like Firefox, can mitigate intractable cookies by isolating them per *setter website*. In the scenario described in Section 6.4.2, when `basic.com` sets a cookie from `tracker.net`, the cookie is stored in a partitioned context specific to `basic.com`. Later, when the user visits `new.com` (*sender website*), even if it loads a resource from `tracker.net`, the browser does not send the previously set cookie ($id = 123$) because partitioned storage treats `new.com` as a separate context. As a result, `tracker.net` cannot correlate the user's activity across `basic.com` and `new.com`.

Since OpenWPM only supports Firefox and, unlike Chrome, Firefox does not explicitly expose the `partitioned_key` field in cookie storage, we modified *BannerClick* to use Chrome in its default mode. This approach allows us to collect cookies independently of OpenWPM. To measure the prevalence of partitioned cookies, similar to

the stateful phase of the other runs, we conducted a separate run on accepted domains from the *RandComb* run, collecting all cookies set after accepting the banners.

The results show a similar distribution of cookie types compared to previous runs conducted using Firefox. Out of 79,898 unique stored cookies, 521 are set as partitioned cookies, as shown in Table 6.4. However, among 3,177 unique tracking cookies, only 40 (1.3%) are partitioned, of which 26 are accompanied by non-partitioned tracking cookies from the same tracker, with 9 having the same value[10]. Interestingly, we also observe that Chrome does not overwrite existing cookies when their partitioned attribute differs. In other words, the same cookie can be set twice—with and without the partitioned attribute—both of which are sent in subsequent HTTP requests.

Overall, despite nearly three years since partitioned cookies were introduced by Google through the Cookies Having Independent Partitioned State (CHIPS) initiative with the release of Chrome 100 in March 2022, their adoption remains limited and gradual. We discuss the effectiveness of partitioned cookies further in Section 6.5. Nevertheless, a longitudinal study is needed to assess whether partitioned cookies will achieve widespread adoption and be effectively applied in real-world scenarios, ensuring they fulfill their intended purpose of balancing user privacy with functional third-party integrations.

To summarize: *Only 1.3% of all unique tracking cookies are set as partitioned, and more than half of them are accompanied by non-partitioned cookies from the same tracker domain.*

		Total	First Party	Third Party	Tracking	Intractable
Regular	Aggregate	95,131	57,918	37,213	21,920	19,296
	Unique	67,083	57,782	9,301	3,583	2,131
	Avg	31.35	19.09	12.27	7.22	6.36
Regular-Reverse	Aggregate	93,769	55,154	38,615	20,612	18,571
	Unique	67,195	55,115	12,080	3,108	1,769
	Avg	30.64	18.02	12.62	6.74	6.07
Random	Aggregate	84,266	49,315	34,951	17,610	16,041
	Unique	61,349	49,217	12,132	2,667	1,725
	Avg	28.73	16.81	11.92	6.00	5.47
Random-Reverse	Aggregate	83,718	51,180	32,538	17,338	15,455
	Unique	61,421	51,150	10,271	2,823	1,665
	Avg	28.07	17.16	10.91	5.81	5.18

Table 6.5: Cookie distribution across different measurements in *Cookie Jar* database.

		Total	Intractable	Reset
Regular	Aggregate	225,520	75,164	5,952
	Unique	31,239	2,131	530
	Avg	94.80	31.59	2.50
Regular-Reverse	Aggregate	187,827	54,918	5,533
	Unique	28,935	1,769	530
	Avg	77.49	22.66	2.28
Random	Aggregate	214,779	59,605	6,591
	Unique	28,618	1,725	530
	Avg	83.31	23.12	2.56
Random-Reverse	Aggregate	201,862	68,040	6,820
	Unique	27,876	1,665	536
	Avg	79.38	26.76	2.68

Table 6.6: Cookie distribution across different measurements in *Sent Cookies* database.

6.4.9 Cookie Jar and Sent Cookies

In this section, we detail the overall distribution of cookies collected during our measurement campaign across both regular and random runs, as quantified in the numbers presented in Table 6.6.

Overall, the table classifies cookies under two main categories (databases): *Cookie Jar* and *Sent Cookies*. *Cookie Jar* refers to cookies that are set by websites in the stateful phase when their consent banners are accepted by the BannerClick tool [120]. *Sent Cookies* includes cookies extracted from the HTTP requests of websites before the rejection of consent banners. Subsequently, the table's sub-columns categorize the cookies into different types, viz., 'Total,' 'First Party,' 'Third Party,' 'Tracking,' 'Intractable,' and 'Reset.' Specifically, 'Total' denotes the number of all cookies in the databases. 'Reset' refers to intractable cookies overwritten within the domain sending it, *e.g.,* in Figure 6.1, `new.com` may also reset the intractable cookie (*id*=123).

For each run, the first row presents the cumulative count of cookies collected, reflecting the aggregated number of cookies set or sent across the domains. For instance, if a cookie is set twice by two different *setter websites*, it is counted as two. On the other hand, the second row shows the count of unique cookies grouped by their `name` and `host` attributes. Finally, the third row, labeled 'Avg,' depicts the mean number of cookies per domain.

Note that the table presents raw statistics without any adjustments. Across different runs, the number of accepted websites varies, influencing the number of cookies stored in the *Cookie Jar*. Runs with a greater number of accepted domains are likely to have a larger number of unique cookies in *Cookie Jar*, potentially resulting in a higher count of cookies sent per rejected domain. Consequently, a fair comparison cannot be

[10]In most of these cases, the partitioned cookie has a similar `name`, usually appended by 'p' or '_p'.

made across the average numbers presented in the *Sent Cookies* columns of different runs. For a fair comparison, see Section 6.4.

In comparing *Cookie Jar* of regular runs, we observe that the average number of cookies is similar across all categories. In the case of random runs, the behavior across all categories appears relatively consistent. Additionally, in all runs, the counts of 'Unique' intractable cookies are identical between *Cookie Jar* and its corresponding *Sent Cookies*. This consistency is expected, as we classify a cookie in *Cookie Jar* as intractable only if it is subsequently sent from domains where consent is rejected (*i.e.,* intractable cookies in *Sent Cookies*).

Furthermore, the average number of 'Reset' cookies is relatively low compared to the number of intractable cookies sent. This highlights the cohort nature of intractable cookies, as they tend to be sent via HTTP requests without being reset, making them difficult to track by simply observing the current state of cookies in the browser. It also shows that merely considering the deployment of tracking cookies overlooks a large portion of real-world tracking practices.

6.5 Discussion

In the following, we examine ongoing technical solutions proposed by various browsers to mitigate third-party tracking via web cookies (*e.g.,* intractable cookies), evaluating their advantages and limitations.

As discussed in Section 6.4.2, from a technical perspective, the primary cause of intractable cookies may be the *sender websites*' inability to determine the existence of cookies set by previously visited websites (*i.e., setter websites*). To address this issue, several solutions can help mitigate or eliminate the intractable cookies phenomenon. One approach is to *prevent the loading of third-party resources* unless the user explicitly accepts the banner. However, given the current structure of consent mechanisms—where websites handle user preferences via banners and browsers control request transmission—this solution is not feasible.

Alternatively *blocking third-party cookies* is the most straightforward approach to mitigating the privacy-intrusive nature of tracking cookies, including intractable cookies. This approach has already been implemented by browsers like Safari as a default setting [42]. As for Chrome—which accounts for over 65% of the browser market share across both desktop and mobile platforms [138]—privacy-conscious users have the option to customize their settings and block third-party cookies. However, studies [95] show that most users are unaware of these controls or the tracking technologies behind them. More importantly, the debate over tracking extends beyond individual user preferences, as it involves conflicting interests among users, publishers, and advertisers. These competing priorities complicate the feasibility of outright blocking. For instance, Google's July 2024 reversal [112] of its 2020 pledge [27, 150] to phase out third-party cookies in Chrome underscores the tension between privacy advocacy and economic interests.

Nonetheless, if we focus solely on users' interests and assume that blocking third-party cookies enhances their web experience by improving user privacy, the reality is more complex. Entities dependent on advertising revenue will likely shift to alternative tracking methods, such as fingerprinting [5], or adjust their pricing strategies to compensate for the loss of targeted ads [96]. Consequently, from a broader perspective, eliminating third-party cookies entirely may not provide the anticipated benefits for users unless an alternative monetization model is introduced.

Another possible solution to mitigate intractable cookies is *partitioned cookies*. However, based on our observations and given its limited implementation (see Section 6.4.8), partitioned cookies also have several other limitations:

- *Developer Reliance:* Approaches that depend on widespread developer adoption often fail to achieve meaningful deployment. For instance, studies on Content Security Policy (CSP) [72, 125] reveal that less than 2% of websites correctly implement it, with most deployments being ineffective or poorly configured. Although implementing partitioned cookies is less complex than CSP, it still requires modifying attributes like `SameSite` and appending the `_Host` prefix to handle subdomains.
- *Lack of Incentives:* The incentive for adopting partitioned cookies remains unclear (particularly when considering the *tracker domain* as the owner of the cookies) unless explicitly enforced by regulations. In comparison, CSP adoption is driven by its direct relation to the website's own security.
- *Ambiguity of Cookie Purpose:* Cookies serve a variety of purposes, such as saving sessions, managing inter-domain states, tracking, and enhancing security. Simply classifying cookies as partitioned or non-partitioned is insufficient, as owners can justify the use of non-partitioned cookies for various reasons. This perpetuates the issue of an "unclear" cookie purpose.
- *Incompatibility with Consent-Based Tracking:* Most importantly, partitioned cookies lack the technical capability to enable inter-domain tracking, even upon explicit users' consent. This limitation undermines consent-based tracking, as it prevents users from selectively allowing or blocking tracking based on their preferences via cookie banners. For instance, some users may wish to receive personalized advertising while blocking tracking from specific websites, such as those handling sensitive content. However, with partitioned cookies, such granular control is no longer possible, rendering cookie banners ineffective for managing tracking preferences.

6.6 Summary

In this chapter, we reveal the prevalence of *intractable* cookies—tracking cookies that are set by websites where users accept their banners, persistent in the user browser, and sent to tracking domains before the user's explicit consent on subsequent

websites. Through extensive measurements involving 20,000 domains from the Tranco top list, we demonstrated that around 50% of the websites sent at least one intractable cookie. Furthermore, we assessed how banner interaction, enabling GPC signal, can contribute to preventing intractable cookies. We then explored the impact of the website rank and type of the banner on the prevalence of these cookies. Moreover, we analyzed the expiration and duplication characteristics of intractable cookies, along with their domain distribution. Finally, we examined current technical solutions, such as partitioned cookies, that aim to mitigate intractable cookies and discussed their limitations.

7

Summary and Final Thoughts

As a result of privacy regulations such as GDPR, websites are increasingly using cookie banners to inform users and obtain their consent for data collection and processing. While this appears to be a positive step toward a more privacy-conscious web environment, questions remain about the actual effectiveness and impact of cookie banners in protecting user privacy.

In this thesis, we initially introduce *BannerClick*, a tool developed to automatically detect and interact with cookie banners with 99% accuracy. Using *BannerClick*, we then conduct a series of measurements to systematically examine the state of consent mechanisms and cookie-based tracking—with a particular focus on how websites respond to user interactions with banners and whether they comply with privacy regulations such as the GDPR.

In the first study, we examine the role of geolocation, user interactions, and device types in cookie deployment. Our findings show that websites accessed from GDPR-regulated regions display cookie banners more frequently and set fewer tracking cookies compared to those in non-regulated areas. Additionally, website behavior varies significantly based on whether a user is on a landing page or an inner page and whether they are browsing from a desktop or a mobile device.

Next, we analyze the emergence of cookie paywall, which present users with a binary choice: accept tracking or pay for a subscription to access the website. Our findings reveal that websites with cookie paywalls are more prevalent in GDPR-regulated countries and that these websites set significantly more third-party cookies compared to those with standard banners. Additionally, we examine the pricing models of cookie paywalls and explore possible methods to bypass them.

Finally, we introduce *intractable cookies*—cookies that are transmitted to trackers due to prior acceptance on previously visited websites, even before users interact with banners on later websites. We further investigate the relationship between website ranking, banner design, and the prevalence of intractable cookies, highlighting the challenges in enforcing meaningful user consent.

In this chapter, we further discuss the implications of our findings in the broader context of web privacy. We begin by outlining the essential features a proper consent mechanism should include. To support this, we elaborate on the relevant provisions of GDPR and explore their limitations and potential refinements. Next, after reviewing current ongoing solutions, we introduce our proposal for a browser-integrated consent

mechanism. Throughout this discussion, we strive to maintain a balanced perspective and avoid over-speculation. Nevertheless, further scientific research and investigation are required to validate and reinforce these insights.

7.1 Effectiveness of Data Protection Regulations

When it comes to the web environment, the primary objective of privacy regulations such as GDPR and CCPA is to protect users' privacy by informing them and providing them control over their data. However, achieving these objectives requires a broad and comprehensive approach that correctly directs efforts toward transparent, usable, and effective solutions.

In the context of web consent mechanisms, we define transparency, usability, and efficacy as follows:

- **Transparency:** The system must clearly communicate data collection practices to users while emphasizing the importance of their informed decisions in safeguarding their privacy. To achieve this, information should be presented in clear, inclusive language, free from technical jargon, ensuring accessibility for individuals of all backgrounds and technical proficiencies.
- **Usability:** A well-designed consent interface should empower users to make informed privacy decisions without overwhelming them. The design should prioritize simplicity, guiding users through straightforward options and minimizing cognitive strain.
- **Efficacy:** More importantly, the final outcome should be entirely predictable based on the user's decision, ensuring a reliable consent process. For example, if a user accepts consent on one website, this acceptance should not be automatically propagated to other websites.

Based on these criteria, we first evaluate the current state of the GDPR and its limitations in enforcing effective consent mechanisms. We then explore potential improvements and future directions for web privacy regulations that can lead to more robust and user-centric consent solutions.

7.1.1 GDPR and Consent Mechanisms

As we demonstrate throughout this study, the existence of cookie paywalls and intractable cookies indicates that while the advent of privacy regulations such as GDPR appears promising, there is still a notable gap between regulatory intent and actual practices on the web. In the following, we examine the current deployment of cookie banners by addressing two key limitations in the GDPR: the absence of concrete technical guidelines for consent mechanisms and the ambiguity surrounding the entity responsible for their implementation.

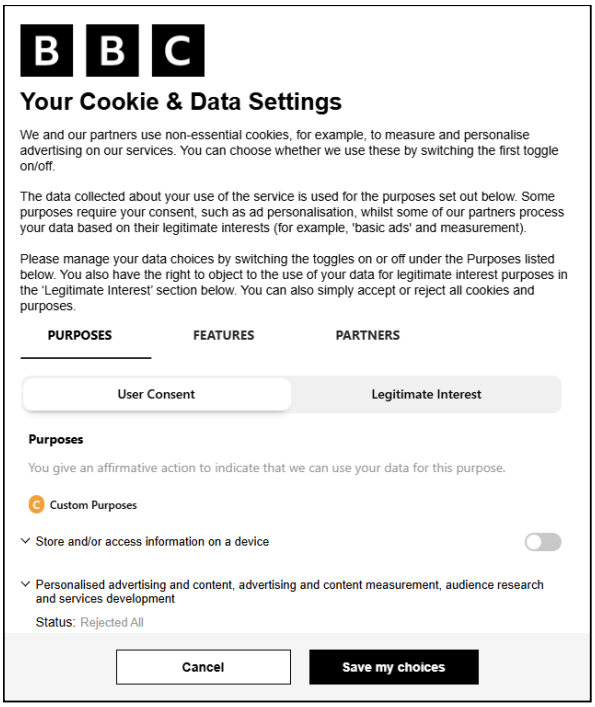


Figure 7.1: An example of a complex cookie banner that requires users to navigate through the settings view, which contains multiple tabs to manage their preferences.

Absence of Concrete Technical Guidelines: While the GDPR defines core criteria for valid consent—such as being freely given, specific, informed, and unambiguous—it lacks concrete design guidelines for translating these principles into practical solutions. Recitals 32 and 43, for instance, suggest that a well-designed banner should offer users multiple options for different purposes. Although most banners provide these options in their settings view along with a "Reject All" button, the absence of an explicit GDPR requirement for such a button on the initial banner screen often forces users to navigate into settings, where they encounter these granular and potentially confusing choices. Figure 7.1 illustrates a typical example of a complex cookie banner. These banners present users with extensive details, often highly technical and difficult for the average user to comprehend. Such designs overwhelm users, particularly those without technical expertise, ultimately diminishing the effectiveness of cookie banners in clearly conveying data collection practices [67, 95, 154, 157].

Moreover, in practice, users do not seek such a high level of control [88, 154]. They typically look for a straightforward way to either opt in or opt out, and it is rare for them to manually select or deselect specific purposes or vendors. Even when users desire such options, they can be presented in a more user-friendly manner—designed to remain accessible yet unobtrusive to the majority of non-curious users. However, this aspect is also not explicitly addressed by the GDPR.

Additionally, the terms "freely given" and "unambiguous" in the GDPR lack precise

definitions, leaving room for varying interpretations[1]. Previous studies show that cookie banners often use deceptive patterns that prompt users to act impulsively or habitually rather than making informed choices [14, 29, 104, 136]. The rise of cookie paywalls [119] further pressures users into accepting tracking, often against their preferences. Consequently, many users reluctantly consent to tracking on certain sites. As we have shown in Chapter 6, these seemingly domain-specific acceptances propagate across later visits via intractable cookies, enabling inter-domain tracking even before users provide explicit consent on a visited site.

Ambiguity of Accountability: Intractable cookies constitute a clear violation of privacy regulations. Under the GDPR, websites must obtain user consent before collecting or processing any data. However, the GDPR's accountability framework remains ambiguous. It designates the "data controller"[2] as the entity responsible for implementing technical and organizational measures to ensure compliance with the Regulation (Article 24). Further, Article 26 introduces the concept of a "joint controller" and mandates that these parties establish an *arrangement* to transparently define their respective responsibilities. A straightforward interpretation of these provisions suggests that all involved entities must collectively clarify and inform users about their data practices.

However, the complex interactions between various entities (e.g., third-party trackers, sender websites, advertisers, DMPs, CMPs) make such an *arrangement* impractical or, at best, difficult to enforce within the current web infrastructure. Consequently, this ambiguity complicates which party is liable, even for privacy lawyers. Without clear accountability, enforcement remains inconsistent, allowing entities to evade responsibility. As a result, the GDPR may struggle to establish an effective framework for consent mechanisms.

In a nutshell, the current deployment of cookie banners appears to be an ad hoc response by publishers to comply with legal requirements, potentially fostering a *false sense of privacy*. While privacy regulations such as GDPR promote transparency and accountability, their effective implementation requires a deeper understanding of user perceptions of consent mechanisms and the roles of different entities involved. Without this understanding and systematic collaboration between regulators, developers, and interdisciplinary experts, these regulations risk backfiring. Poorly designed or implemented policies may fail to achieve their intended purpose, complicating enforcement and ultimately undermining privacy rather than protecting it.

[1] The definition of these terms in Recital 32 relies on other undefined terms, which may lead to subjective interpretations influenced by biased interests.

[2] The GDPR Article 4(7) defines the *data controller* as "the natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data".

7.2 Future Directions

Looking ahead, a more practical and user-centric design is anticipated, possibly integrated at the browser level, allowing users to customize their privacy settings once for the entire browsing session rather than on a per-domain basis.

One promising approach is a browser-integrated consent management portal, where banners are directly connected to the browser, and interactions with them are managed by the browser itself. This will enable the browser to determine when and with which third parties to communicate based on user preferences. While this approach requires substantial changes in the current interactions between web interfaces and browsers, it is both rational and streamlined from an engineering perspective, as it delegates data handling to the core software—browsers.

Alternatively, given that tracking cookies are primarily used for targeted advertising—which is not inherently detrimental and serves as a financial incentive for publishers—a privacy-preserving advertising system that minimizes the leakage of personally identifiable data seems like a better option, similar to Interoperable Private Attribution [65]. Although such systems require significant resources and technical expertise for implementation, privacy-by-design solutions offer a more reliable approach. They effectively reduce the need to disclose data collection activities, thereby minimizing or even eliminating reliance on banners as a means of informing and educating users. This is particularly important considering the trade-offs between time, user experience, usability, effectiveness, and other factors.

In this section, we introduce our proposal for a browser-integrated consent mechanism and discuss how it can address the criteria outlined in Section 7.1 for a desirable consent mechanism.

7.2.1 Browser-Integrated Consent Mechanism

As discussed in Section 7.1.1, one of the major drawbacks of the GDPR is its ambiguity in defining the "data controller" as the entity responsible for collecting and handling user consent. Moreover, obtaining user consent is distinct from data collection and processing and can instead be managed by a separate entity within the data flow. Accordingly, given the current structure of the web ecosystem—where the browser serves as the central element orchestrating communication between various entities—it seems more reasonable and practical to integrate the consent mechanism directly into the browser. This approach streamlines the consent process, ensuring a more consistent, efficient, and scalable implementation that can be easily maintained and updated.

In this model, browsers serve as the intermediary entity responsible for collecting user consent and preferences and applying them accordingly. This approach reduces the burden on both users and websites (*i.e.,* developers). Users can configure their preferences through a well-structured consent management portal within the browser, ensuring their choices are consistently applied across all websites they visit. Likewise,

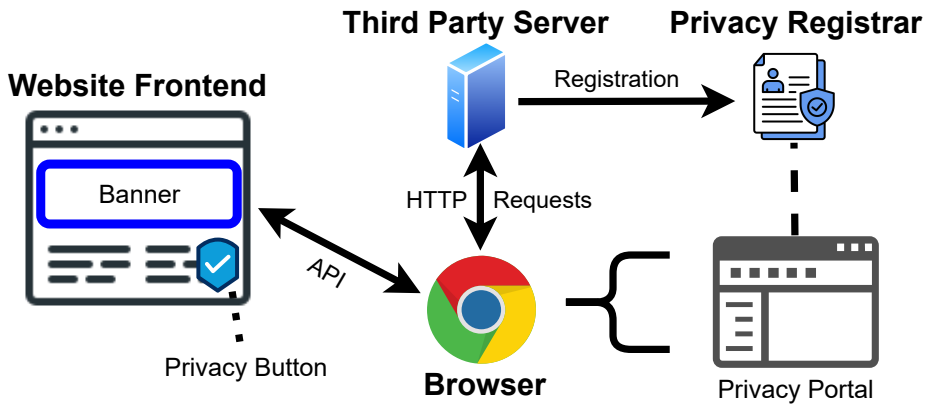


Figure 7.2: Overview of the proposed Browser-Integrated Consent Mechanism and its four major components: browser, website, third party, and registrar.

websites will no longer need to implement their own consent banners, resulting in a more uniform and user-friendly experience.

Figure 7.2 illustrates an overview of our proposal for a browser-integrated consent mechanism. This scheme consists of four major components: the browser, the website, third parties (*e.g.,* tracker or advertiser companies), and the registrar. Initially, each third party that intends to use cross-site tracking technologies, such as cookies, must register with a designated registration agency and declare its purpose, *e.g.,* similar to how vendors register with the IAB Europe Global Vendor List in the context of CMPs [62]. The browser retrieves this information from the registrar and uses it to determine whether to allow or block requests associated with a particular third party based on user preferences.

On the other hand, users can access a *privacy portal*, a browser interface that allows them to set preferences for each website and registered third parties if desired. For instance, they can opt out of certain trackers or specific websites, *e.g.,* based on category. Alternatively, users can modify these preferences upon visiting a website through banners operated by the browser. To facilitate this, the browser provides an easily accessible *privacy button* that enables users to open the banner at any time to revoke or adjust their preferences for the current website and its embedded third parties. While websites can configure the theme and overall appearance of these banners through an API provided by the browser, their fundamental structure, including text and available options, remains consistent across all websites. The banners prominently display easy-to-find "Accept All" and "Reject All" buttons, along with settings for more granular options.

This high-level prototype can be further developed and refined to strengthen user privacy while maintaining essential advertising capabilities. One potential approach is for browsers to treat all third-party cookies as partitioned by default and transmit them cross-site only if explicitly permitted by user preferences. Furthermore, the portal can incorporate a subscription-based model, allowing websites to monetize their content directly as an alternative to ad-supported tracking, similar to existing

cookie paywalls. In this model, websites can adapt their behavior and render the front end based on user choices or subscription status. Although this model may create a divide between paying and non-paying users, its alignment with free market principles can, in the long run, lead to improved services and a more sustainable online environment.

7.3 Ethical Considerations

In conducting all of our measurements, we abide by the ethical guidelines proposed by Partridge and Allman [111] and Kenneally and Dittrich [71], and follow the best measurement practices as described by Durumeric et al. [34]. Our methodology involves running OpenWPM in a manner that mimics the behavior of a typical user browsing with a standard web browser. We utilize dedicated measurement machines, configure informative reverse DNS (rDNS) names, and maintain a website that explains our research activities. Moreover, we offer stakeholders the option to opt-out and be excluded from our measurements. Throughout our measurement period, we did not receive any complaints.

7.4 Conclusion

Privacy regulations form the foundation of the web privacy ecosystem. Their emergence marks a significant step forward, demonstrating how societies and governments adapt to evolving technological landscapes. However, the presence of various types of banners—such as cookie paywalls and banners employing deceptive patterns—along with the persistence of intractable cookies, suggests that there is still much progress to be made in enhancing the effectiveness and practicality of these regulations.

Although technical advancements like partitioned cookies and CMPs have been introduced, the current state of consent mechanisms remains far from ideal. Achieving a consent mechanism that is transparent, user-friendly, and effective requires intentional collaboration among all stakeholders—including developers, regulators, publishers, and advertisers—while balancing the interests of all parties.

Overall, despite notable progress and persistent challenges in user privacy on the web, aligning data protection with data-driven technologies—like many technological advancements—is an ongoing process without a definitive endpoint. As the Internet and web technologies continue to evolve, privacy regulations and consent mechanisms must likewise adapt, requiring sustained attention and continuous refinement.

A

Appendix

A.1 GDPR Provisions

In this section, we summarize the General Data Protection Regulation (GDPR) provisions and recitals cited in this thesis.

Article 1 - Subject-matter and Objectives

- **Subject-matter:** This Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.
- **Fundamental Rights:** This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data.
- **Free Movement of Data:** The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data.

Article 2 - Material Scope

- **Scope of Application:** This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.
- **Exemptions:** This Regulation does not apply to the processing of personal data:
  - in the course of an activity which falls outside the scope of Union law;
  - by the Member States when carrying out activities under Chapter 2 of Title V of the Treaty on European Union;
  - by a natural person for purely personal or household activities;
  - by competent authorities for law enforcement purposes.

Article 3 - Territorial Scope

- **Applicability Within the Union:** This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or processor in the Union, regardless of whether the processing takes place in the Union or not.
- **Applicability Outside the Union:** This Regulation applies to controllers or processors not established in the Union where processing activities are related to:
  - the offering of goods or services to data subjects in the Union, whether or not a payment is required;
  - the monitoring of their behavior within the Union.
- **Application by International Law:** This Regulation applies where Member State law applies by virtue of public international law.

Article 4 - Definitions

- **Personal Data:** Any information relating to an identified or identifiable natural person (data subject); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;
- **Processing:** Any operation or set of operations performed on personal data, such as collection, recording, organization, storage, adaptation, retrieval, use, disclosure, restriction, erasure, or destruction.
- **Profiling:** Any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural persons performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.
- **Controller:** A person or organization determining the purposes and means of processing personal data.
- **Processor:** A person or organization processing personal data on behalf of the controller.
- **Third Party:** A natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data.

- **Consent:** Any freely given, specific, informed, and unambiguous indication of the data subjects wishes regarding data processing.

Article 5 - Principles Relating to Processing of Personal Data

- **Data Minimization:** Personal data shall be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.
- **Purpose Limitation:** Personal data shall be collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes. Further processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes shall not be considered incompatible with the initial purposes, in accordance with Article 89(1).
- **Accountability:** The controller shall be responsible for, and be able to demonstrate compliance with, the principles relating to processing of personal data.

Article 7 - Conditions for Consent

- **Demonstrating Consent:** Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to the processing of their personal data.
- **Withdrawal of Consent:** The data subject shall have the right to withdraw their consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.
- **Consent Requirements:** When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

Article 12 - Transparent Information, Communication and Modalities for the Exercise of the Rights of the Data Subject

- **Transparent Communication:** The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child.

Article 13 - Information to be Provided Where Personal Data are Collected from the Data Subject

- **Information at Collection:** Where personal data relating to a data subject are collected from the data subject, the controller shall, at the time when personal data are obtained, provide the data subject with all of the following information: (a) the identity and contact details of the controller and, where applicable, of the controllers representative; (b) the contact details of the data protection officer, where applicable; (c) the purposes of the processing for which the personal data are intended as well as the legal basis for the processing.

Article 14 - Information to be Provided Where Personal Data Have Not Been Obtained from the Data Subject

- **Information to Data Subjects:** Where personal data have not been obtained from the data subject, the controller shall provide the data subject with the following information: (a) the identity and contact details of the controller and, where applicable, of the controllers representative; (b) the contact details of the data protection officer, where applicable; (c) the purposes of the processing for which the personal data are intended as well as the legal basis for the processing.

Article 24 - Responsibility of the Controller

- **Obligations of the Controller:** Taking into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons, the controller shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation. Those measures shall be reviewed and updated where necessary.

Article 26 - Joint Controllers

- **Joint Responsibility:** Where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers. They shall determine in a transparent manner their respective responsibilities for compliance with the obligations under this Regulation, in particular as regards the exercising of data subjects rights and their respective duties to provide the information referred to in Articles 13 and 14.
- **Arrangements Between Joint Controllers:** The arrangement between joint controllers shall duly reflect their respective roles and relationships vis-à-vis the data subjects. The essence of the arrangement shall be made available

to the data subject. Regardless of the terms of the arrangement, the data subject may exercise their rights under this Regulation in respect of and against each of the controllers.

Article 30 - Records of Processing Activities

- **Record-Keeping Obligation:** Each controller and, where applicable, the controllers representative shall maintain a record of processing activities under their responsibility. This record shall include at least the following information: (a) the name and contact details of the controller and, where applicable, the joint controller, the controllers representative, and the data protection officer; (b) the purposes of the processing; (c) a description of the categories of data subjects and the categories of personal data.

Recital 32

- **Elements of Consent:** Consent should be given by a clear affirmative act establishing a freely given, specific, informed, and unambiguous indication of the data subjects agreement to the processing of personal data relating to them, such as by a written statement, including by electronic means, or an oral statement.

Recital 43

- **Freely Given Consent:** Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment. To ensure that consent is freely given, consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation.

Bibliography

[1] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., Diaz, C.: The web never forgets: Persistent tracking mechanisms in the wild. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 674–689. ACM. https://doi.org/10.1145/2660267.2660347

[2] Acquisti, A., Brandimarte, L., Loewenstein, G.: Privacy and human behavior in the age of information. Science 2015 **347**(6221), 509–514. https://doi.org/10.1126/science.aaa1465

[3] Alexa Internet, Inc.: Alexa top sites. `https://www.alexa.com/topsites`, accessed: 2025-04-04

[4] Amos, R., Acar, G., Lucherini, E., Kshirsagar, M., Narayanan, A., Mayer, J.: Privacy policies over time: Curation and analysis of a million-document dataset. In: Proceedings of the Web Conference 2021. pp. 2161–2173. WWW '21, ACM. https://doi.org/10.1145/3442381.3450048

[5] Angove-Plumb, A.: Browser fingerprinting and the death of third-party cookies. Choice `https://www.choice.com.au/consumers-and-data/data-collection-and-use/who-has-your-data/articles/browser-fingerprinting-and-death-of-third-party-cookies`, accessed: 2025-04-08

[6] Aqeel, W., Chandrasekaran, B., Feldmann, A., Maggs, B.M.: On landing and internal web pages: The strange case of jekyll and hyde in web performance measurement. In: Proceedings of the 2020 ACM Internet Measurement Conference. pp. 132–145. ACM. https://doi.org/10.1145/3419394.3423620

[7] Australian Government: Privacy act 1988 (CTH), `https://www.legislation.gov.au/Details/C2023C00233`, amended and maintained by the Office of the Australian Information Commissioner (OAIC)

[8] Barford, P., Sitaraman, R., Hossfeld, T.: Understanding and managing web cookies: Fundamentals, measurement, and implications. ACM Computing Surveys 2019 **52**(3), 1–36. https://doi.org/10.1145/3342552

[9] Bateman, R.: CCPA: Does using third-party cookies count as selling personal information? `https://www.termsfeed.com/blog/ccpa-third-party-cookies-selling-personal-information/`, accessed: 2025-04-04

[10] Bekos, P., Papadopoulos, P., Markatos, E.P., Kourtellis, N.: The hitchhiker's guide to facebook web tracking with invisible pixels and click ids. In: Proceedings of the ACM Web Conference 2023. pp. 2132–2143. https://doi.org/10.1145/3543507.3583311

[11] Belanger, F., Hiller, J.S., Smith, W.J.: Trustworthiness in electronic commerce: The role of privacy, security, and site attributes. The Journal of Strategic Information Systems 2002 **11**(3-4), 245–270. https://doi.org/10.1016/S0963-8687(02)00018-5

[12] Bianchi, T.: Google: Advertising revenue 2024. `https://www.statista.com/statistics/266249/advertising-revenue-of-google/`, accessed: 2025-04-04

[13] Bielova, N., Litvine, L., Nguyen, A., Chammat, M., Toubiana, V., Hary, E.: The effect of design patterns on (present and future) cookie consent decisions. In: The 33rd USENIX Security Symposium (USENIX Security 2024). pp. 2813–2830. USENIX Association, `https://www.usenix.org/conference/usenixsecurity24/presentation/bielova`

[14] Borberg, I., Hougaard, R., Rafnsson, W., Kulyk, O.: So i sold my soul: Effects of dark patterns in cookie notices on end-user behavior and perceptions. In: Usable Security and Privacy (USEC) Symposium 2022. pp. 1–11. The Internet Society. https://doi.org/10.14722/usec.2022.23026

[15] Brinkmann, M.: Firefox may soon reject cookie prompts automatically. `https://www.ghacks.net/2023/04/17/firefox-may-interact-with-cookie-prompts-automatically-soon/`, accessed: 2025-04-04

[16] Bujlow, T., Carela-Español, V., Solé-Pareta, J., Barlet-Ros, P.: A survey on web tracking: Mechanisms, implications, and defenses. In: Proceedings of the IEEE Communications Surveys & Tutorials 2017. vol. 18, pp. 1352–1380. IEEE. https://doi.org/10.1109/COMST.2015.2513762

[17] Cahn, A., Alfeld, S., Barford, P., Muthukrishnan, S.: An empirical study of web cookies. In: The World Wide Web Conference 2016. pp. 891–901. WWW '16, ACM. https://doi.org/10.1145/2872427.2882991

[18] California Office of the Attorney General: California consumer privacy act (CCPA) - global privacy control, `https://oag.ca.gov/privacy/ccpa/gpc`, accessed: 2025-02-28

[19] Chau, E., Hertzberg, R.: California consumer privacy act. `https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375`, accessed: 2025-04-04

[20] Chen, Q., Ilia, P., Polychronakis, M., Kapravelos, A.: Cookie swap party: Abusing first-party cookies for web tracking. In: Proceedings of the Web Conference 2021. pp. 2117–2129. WWW '21, ACM. https://doi.org/10.1145/3442381.3450071

[21] Chen, R., Fang, F., Norton, T., McDonald, A.M., Sadeh, N.: Fighting the fog: Evaluating the clarity of privacy disclosures in the age of CCPA. In: Proceedings of the 20th Workshop on Privacy in the Electronic Society (WPES) 2021. pp. 157–170. ACM. https://doi.org/10.1145/3463676.3485616

[22] Chrome User Experience Report contributors: Chrome user experience report. https://developer.chrome.com/docs/crux/, accessed: 2025-04-04

[23] Chromium Blog: Potential uses for the privacy sandbox. https://blog.chromium.org/2019/08/potential-uses-for-privacy-sandbox.html, accessed: 2025-04-04

[24] Consentmanager AB: Working with contentpass integration. https://help.consentmanager.net/books/cmp/page/working-with-contentpass-integration, accessed: 2025-04-04

[25] Content Pass GmbH: Contentpass website. https://www.contentpass.net/, accessed: 2025-04-04

[26] Cookie Banner Taskforce: Report of the work undertaken by the cookie banner taskforce. https://edpb.europa.eu/system/files/2023-01/edpb_20230118_report_cookie_banner_taskforce_en.pdf, accessed: 2025-04-04

[27] Cookiebot: Google ending third-party cookies in chrome. https://www.cookiebot.com/en/google-third-party-cookies/, accessed: 2025-04-04

[28] Cotter, K.: Shadow banning is not a thing: Social media algorithms, the attention economy, and the myth of suppression. New Media & Society 2021 **23**(9), 2521–2539. https://doi.org/10.1177/1461444820929328

[29] Coventry, L.M., Jeske, D., Blythe, J.M., Turland, J., Briggs, P.: Personality and social framing in privacy decision-making: A study on cookie acceptance. Frontiers in Psychology 2016 **7**, 1341. https://doi.org/10.3389/fpsyg.2016.01341

[30] Dabrowski, A., Merzdovnik, G., Ullrich, J., Sendera, G., Weippl, E.: Measuring cookies and web privacy in a post-GDPR world. In: Proceedings of the 2019 Passive and Active Measurement Conference. pp. 258–270. Springer. https://doi.org/10.1007/978-3-030-15986-3_17

[31] Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., Holz, T.: We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. In: Network and Distributed Systems Security (NDSS) Symposium 2019. Internet Society. https://doi.org/10.14722/ndss.2019.23378

[32] Demir, N., Theis, D., Urban, T., Pohlmann, N.: Towards understanding first-party cookie tracking in the field. Arxiv Preprint Arxiv:2202.01498

[33] Dillet, R.: Google to update cookie consent banner in europe following fine. https://techcrunch.com/2022/04/21/google-to-update-cookie-consent-banner-in-europe-following-fine/, accessed: 2025-04-04

[34] Durumeric, Z., Wustrow, E., Halderman, J.A.: Zmap: Fast Internet-wide scanning and its security applications. In: The 22nd USENIX Security Symposium (USENIX Security 2013). pp. 605–620. USENIX Association, https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/durumeric

[35] Eckersley, P.: How unique is your web browser? Proceedings on Privacy Enhancing Technologies 2010 **2010**, 1–18. https://doi.org/10.1007/978-3-642-14527-8_1

[36] Englehardt, S., Narayanan, A.: Online tracking: A 1-million-site measurement and analysis. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 1388–1401. ACM. https://-doi.org/10.1145/2976749.2978313

[37] Englehardt, S., Reisman, D., Eubank, C., Zimmerman, P., Mayer, J., Narayanan, A., Felten, E.W.: Cookies that give you away: The surveillance implications of web tracking. In: The World Wide Web Conference 2015. pp. 289–299. WWW '15, ACM. https://doi.org/10.1145/2736277.2741679

[38] Estrada-Jiménez, J., Parra-Arnau, J., Cavallaro, A., Forné, J.: Online advertising: Analysis of privacy threats and protection approaches. Computer Communications 2017 **100**, 32–51. https://doi.org/10.1016/j.comcom.2016.12.016

[39] European Commission: The general data protection regulation (GDPR) in EU. `https://ec.europa.eu/info/law/law-topic/data-protection/`, accessed: 2025-04-04

[40] Falahrastegar, M., Haddadi, H., Uhlig, S., Mortier, R.: The rise of panopticons: Examining region-specific third-party web tracking. In: Traffic Monitoring and Analysis 2014. pp. 104–114. Springer. https://doi.org/10.1007/978-3-642-54999-1_9

[41] Felt, A.P., Barnes, R., King, A., Palmer, C., Bentzel, C., Tabriz, P.: Measuring https adoption on the web. In: The 26th USENIX Security Symposium (USENIX Security 2017). pp. 1323–1338. USENIX Association, `https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/felt`

[42] Fingas, J.: Safari now blocks all third-party cookies by default. Engadget `https://www.engadget.com/2020-03-24-safari-blocks-all-third-party-cookies-by-default.html`, accessed: 2025-04-04

[43] FortiGuard contributors: Web filter lookup | fortiguard. `https://www.fortiguard.com/webfilter`, accessed: 2025-04-04

[44] Gibbs, S.: Mobile web browsing overtakes desktop for the first time. `https://www.theguardian.com/technology/2016/nov/02/mobile-web-browsing-desktop-smartphones-tablets`, accessed: 2025-04-04

[45] Global Privacy Control Group: Global privacy control (gpc). `https://privacycg.github.io/gpc-spec/`, accessed: 2025-02-28

[46] GlobalData Thematic Research: Apple block on third party cookies will change digital media forever. `https://www.verdict.co.uk/apple-halts-third-party-cookies/`, accessed: 2025-04-04

[47] Goel, V.: Get to know the new topics api for privacy sandbox. https://blog.google/products/chrome/get-know-new-topics-api-privacy-sandbox/, accessed: 2025-04-04

[48] Golovan, E.: What is data management platform, how it works and why you really need it in your business, https://igorizraylevych.medium.com/what-is-data-management-platform-how-it-works-and-why-you-really-need-it-in-your-business-ab2e905545f3, accessed: 2025-04-08

[49] Gonzalez, R.: Hacking the citizenry? personality profiling, 'big data' and the election of donald trump. Anthropology Today 2020 **33**(3), 9–12. https://-doi.org/10.1111/1467-8322.12387

[50] Gonzalez, R., Jiang, L., Ahmed, M., Marciel, M., Cuevas, R., Metwalley, H., Niccolini, S.: The cookie recipe: Untangling the use of cookies in the wild. In: Proceedings of the 1st Network Traffic Measurement and Analysis Conference 2017. pp. 1–9. IEEE. https://doi.org/10.23919/TMA.2017.8002896

[51] Google: CLD3 on github. https://github.com/google/cld3, accessed: 2025-04-04

[52] Google: Cookies having independent partitioned state (CHIPS) - privacy sandbox. https://developers.google.com/privacy-sandbox/3pcd/chips, accessed on 05/15/2024

[53] Google: How google uses cookies for advertising, https://business.safety.google/adscookies/, accessed: 2025-02-27

[54] Götze, M., Matic, S., Iordanou, C., Smaragdakis, G., Laoutaris, N.: Measuring web cookies in governmental websites. In: Proceedings of the 14th ACM Web Science Conference 2022. pp. 44–54. ACM. https://-doi.org/10.1145/3501247.3531545

[55] Government of Japan: Act on the protection of personal information (APPI), https://www.ppc.go.jp/en/legal/, accessed: 2025-04-08

[56] Gray, C.M., Santos, C., Bielova, N., Toth, M., Clifford, D.: Dark patterns and the legal requirements of consent banners: An interaction criticism perspective. In: Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems. pp. Article 35, 18 pages. ACM. https://-doi.org/10.1145/3411764.3445779

[57] Gundelach, R., Herrmann, D.: Cookiescanner: An automated tool for detecting and evaluating GDPR consent notices on websites. In: Proceedings of the 18th International Conference on Availability, Reliability and Security 2023. ARES '23, ACM. https://doi.org/10.1145/3600160.3605000

[58] Hamidi, H., Jahanshahifard, M.: The role of the internet of things in the improvement and expansion of business. Journal of Organizational and End User Computing 2018 **30**(3), 24–44. https://doi.org/10.4018/JOEUC.2018070102

[59] Hils, M., Woods, D.W., Böhme, R.: Measuring the emergence of consent management on the web. In: Proceedings of the 2020 ACM Internet Measurement Conference. pp. 317–332. ACM. https://doi.org/10.1145/3419394.3423647

[60] Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 1979 **6**(2), 65–70

[61] Hubbard, D.: Cisco umbrella 1 million. `https://umbrella.cisco.com/blog/2016/12/14/cisco-umbrella-1-million/`, accessed: 2025-04-04

[62] IAB Europe: What is TCF v2.0? `https://iabeurope.eu/tcf-2-0/`, accessed: 2025-04-04

[63] IAB Europe: What is the transparency & consent framework (TCF)? `https://iabeurope.eu/transparency-consent-framework/`, accessed: 2025-04-04

[64] Iordanou, C., Smaragdakis, G., Poese, I., Laoutaris, N.: Tracing cross border web tracking. In: Proceedings of the 2018 ACM Internet Measurement Conference. pp. 329–342. ACM. https://doi.org/10.1145/3278532.3278557

[65] IPA contributors: Patcg-individual-drafts/ipa: Interoperable private attribution (IPA) - a private measurement proposal. `https://github.com/patcg-individual-drafts/ipa/blob/main/IPA-End-to-End.md`, accessed: 2024-05-13

[66] Iqbal, U., Englehardt, S., Shafiq, Z.: Fingerprinting the fingerprinters: Learning to detect browser fingerprinting behaviors. In: IEEE Symposium on Security and Privacy (SP) 2021. vol. 2021-May, pp. 1143–1161. IEEE. https://-doi.org/10.1109/SP40001.2021.00017

[67] Jha, N., Trevisan, M., Mellia, M., Irarrazaval, R., Fernandez, D.: I refuse if you let me: Studying user behavior with privacy banners at scale. In: Proceedings of the 7th Network Traffic Measurement and Analysis Conference 2023. pp. 1–9. IEEE. https://doi.org/10.23919/TMA.2023.00009

[68] Jha, N., Trevisan, M., Vassio, L., Mellia, M.: The Internet with privacy policies: Measuring the web upon consent. ACM Transactions on the Web 2021 **16**(3), 1–24. https://doi.org/10.1145/3555352

[69] JustDomains: Automated scripts to support converting filter lists to "domain-only" lists, `https://github.com/justdomains/ci`, accessed: 2025-02-27

[70] justdomains: Domain-only filter lists. `https://github.com/justdomains/blocklists`, accessed: 2025-04-04

[71] Kenneally, E., Dittrich, D.: The menlo report: Ethical principles guiding information and communication technology research. Ssrn Electronic Journal 2012 . https://doi.org/10.2139/ssrn.2445102

[72] Kerschbaumer, C., Stamm, S., Brunthaler, S.: Injecting CSP for fun and security. In: Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP 2016). pp. 15–25. SciTePress. https://doi.org/10.5220/0005650100150025

[73] Kladnik, D.: I don't care about cookies. `https://www.i-dont-care-about-cookies.eu/`, accessed: 2025-04-04

[74] Koch, R.: What is the LGPD? brazil's version of the GDPR. `https://gdpr.eu/gdpr-vs-lgpd/`, accessed: 2025-04-04

[75] Kosta, E.: Peeking into the cookie jar: The european approach towards the regulation of cookies. International Journal of Law and Information Technology 2013 **21**(4), 380–406. https://doi.org/10.1093/ijlit/eat011

[76] Kretschmer, M., Pennekamp, J., Wehrle, K.: Cookie banners and privacy policies: Measuring the impact of the GDPR on the web. ACM Transactions on the Web (TWEB) 2021 **15**(4), 20:1–20:42. https://doi.org/10.1145/3466722

[77] Kristol, D., Montulli, L.: Http state management mechanism. RFC 2965, Internet Engineering Task Force (IETF) 2001 `https://www.rfc-editor.org/rfc/rfc2965`, accessed: 2025-02-28

[78] Krumay, B., Klar, J.: Readability of privacy policies. In: Data and Applications Security and Privacy XXXIV 2020. pp. 388–399. Springer. https://doi.org/10.1007/978-3-030-49669-2_22

[79] Krumnow, B., Jonker, H., Karsch, S.: How gullible are web measurement tools? a case study analysing and strengthening openwpm's reliability. In: Proceedings of 18th International Conference on Emerging Networking Experiments and Technologies (CoNEXT) 2022. p. 16. ACM. https://doi.org/10.1145/3555050.3569131

[80] Lam, M.S., Pandit, A., Kalicki, C.H., Gupta, R., Sahoo, P., Metaxa, D.: Sociotechnical audits: Broadening the algorithm auditing lens to investigate targeted advertising. Proceedings of the ACM on Human-Computer Interaction 2023 **7**(CSCW2), 360:1–360:37. https://doi.org/10.1145/3610209

[81] Laperdrix, P., Bielova, N., Baudry, B., Avoine, G.: Browser fingerprinting: A survey. ACM Transactions on the Web (TWEB) 2020 **14**(2), 8:1–8:33. https://doi.org/10.1145/3386040

[82] Lavanya: How can we find the xpath for shadow element. `https://www.numpyninja.com/post/how-can-we-find-the-xpath-for-shadow-element`, accessed: 2025-04-04

[83] Lee, C.H., Cranage, D.A.: Personalisationprivacy paradox: The effects of personalisation and privacy assurance on customer responses to travel web sites. Tourism Management 2011 **32**(5), 987–994. https://doi.org/10.1016/j.tourman.2010.08.011

[84] Lerner, A., Simpson, A.K., Kohno, T., Roesner, F.: Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In: The 25th USENIX Security Symposium (USENIX Security 2016). pp. 997–1013. USENIX Association, https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner

[85] Li, T.C., Hang, H., Faloutsos, M., Efstathopoulos, P.: Trackadvisor: Taking back browsing privacy from third-party trackers. In: Proceedings of the 2015 Passive and Active Measurement Conference. pp. 277–289. Springer. https://doi.org/10.1007/978-3-319-15509-8_21

[86] Linden, T., Khandelwal, R., Harkous, H., Fawaz, K.: The privacy policy landscape after the GDPR. Proceedings on Privacy Enhancing Technologies 2020 **2020**(1), 47–64. https://doi.org/10.2478/popets-2020-0004

[87] Loane, S.: The role of the Internet in the internationalisation of small and medium sized companies. Journal of International Entrepreneurship 2005 **3**(4), 263–277. https://doi.org/10.1007/s10843-005-1129-2

[88] Machuletz, D., Böhme, R.: Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. Proceedings on Privacy Enhancing Technologies 2020 **2020**(2), 481–498. https://doi.org/10.2478/popets-2020-0037

[89] Majestic: The majestic million. https://majestic.com/reports/majestic-million, accessed: 2025-04-04

[90] Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics 1947 **18**(1), 50–60. https://doi.org/10.1214/aoms/1177730491

[91] Matte, C., Bielova, N., Santos, C.: Do cookie banners respect my choice? measuring legal compliance of banners from IAB europe's transparency and consent framework. In: IEEE Symposium on Security and Privacy (SP) 2020. pp. 791–809. IEEE. https://doi.org/10.1109/SP40000.2020.00076

[92] Mayer, J., Narayanan, A., Stamm, S.: Do not track: A universal third-party web tracking opt out. https://datatracker.ietf.org/doc/html/draft-mayer-do-not-track-00, accessed: 2025-04-04

[93] Mayer, J.R., Mitchell, J.C.: Third-party web tracking: Policy and technology. IEEE Symposium on Security and Privacy 2012 pp. 413–427. https://doi.org/10.1109/SP.2012.47

[94] McMillan, S.J., Morrison, M.: Coming of age with the Internet: A qualitative exploration of how the Internet has become an integral part of young people's lives. New Media & Society 2016 **8**(1), 73–95. https://doi.org/10.1177/1461444806059871

[95] Mehrnezhad, M., Coopamootoo, K., Toreini, E.: How can and would people protect from online tracking? Proceedings on Privacy Enhancing Technologies 2021 **2021**(1), 105–125. https://doi.org/10.2478/popets-2021-0007

[96] Miller, K.M., Skiera, B.: Economic consequences of online tracking restrictions: Evidence from cookies. International Journal of Research in Marketing 2023 **40**(2), 241–264. https://doi.org/10.1016/j.ijresmar.2023.06.001

[97] Morel, V., Santos, C., Lintao, Y., Human, S.: Your consent is worth 75 euros a year: Measurement and lawfulness of cookie paywalls. In: Proceedings of the 21st Workshop on Privacy in the Electronic Society (WPES) 2022. pp. 213–218. ACM. https://doi.org/10.1145/3559613.3563205

[98] Mozilla: Mdn: Using shadow dom. `https://developer.mozilla.org/en-US/docs/Web/Web_Components/Using_shadow_DOM`, accessed: 2025-04-04

[99] Mozilla: New year, new privacy protection for firefox focus on android. `https://blog.mozilla.org/en/mozilla/new-privacy-protection-for-firefox-focus-on-android/`, accessed: 2025-04-04

[100] Mozilla: Public suffix list. `https://publicsuffix.org/`, accessed: 2025-04-04

[101] Munir, S., Siby, S., Iqbal, U., Englehardt, S., Shafiq, Z., Troncoso, C.: Cookiegraph: Understanding and detecting first-party tracking cookies. In: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 3490–3504. CCS '23, ACM. https://doi.org/10.1145/3576915.3616586

[102] NinjaCookie contributors: Ninja cookie | opt out of non-essential cookies and automatically remove cookie popups. `https://ninja-cookie.com/`, accessed: 2025-04-04

[103] Nortwick, M.V., Wilson, C.: Setting the bar low: Are websites complying with the minimum requirements of the CCPA? Proceedings on Privacy Enhancing Technologies 2022 **2022**(1), 608–628. https://doi.org/10.2478/popets-2022-0030

[104] Nouwens, M., Liccardi, I., Veale, M., Karger, D., Kagal, L.: Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In: Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems. pp. 1–13. CHI '20, ACM. https://doi.org/10.1145/3313831.3376321

[105] O'Connor, S., Nurwono, R., Siebel, A., Birrell, E.: (un)clear and (in)conspicuous: The right to opt-out of sale under CCPA. In: Proceedings of the 2021 Workshop on Privacy in the Electronic Society (WPES). pp. 59–72. ACM. https://doi.org/10.1145/3463676.3485615

[106] Olsen, S.: Nearly undetectable tracking device raises concern. `https://web.archive.org/web/20141107101823/http://news.cnet.com/2100-1017-243077.html`, accessed: 2025-04-09

[107] OpenWPM: Openwpm not using tracking blocking. `https://github.com/openwpm/OpenWPM/issues/101`, accessed: 2025-04-04

[108] OpenWPM: Openwpm stateful vs stateless crawls. `https://github.com/openwpm/OpenWPM/blob/master/docs/Configuration.md#stateful-vs-stateless-crawls`, accessed: 2025-04-04

[109] Papadopoulos, P., Kourtellis, N., Markatos, E.: Cookie synchronization: Everything you always wanted to know but were afraid to ask. In: The World Wide Web Conference 2019. pp. 1432–1442. WWW '19, ACM. https://doi.org/10.1145/3308558.3313542

[110] Papadopoulos, P., Snyder, P., Athanasakis, D., Livshits, B.: Keeping out the masses: Understanding the popularity and implications of Internet paywalls. In: Proceedings of the Web Conference 2020. pp. 1433–1444. WWW '20, ACM. https://doi.org/10.1145/3366423.3380217

[111] Partridge, C., Allman, M.: Ethical considerations in network measurement papers. Communications of the ACM 2016 **59**(10), 58–64. https://doi.org/10.1145/2988445

[112] Picchi, A.: Google reneges on plan to remove third-party cookies in chrome. CBS News `https://www.cbsnews.com/news/google-third-party-cookies-chrome/`, accessed: 2025-04-08

[113] Pochat, V.L., Goethem, T.V., Tajalizadehkhoob, S., Korczyński, M., Joosen, W.: Tranco: A research-oriented top sites ranking hardened against manipulation. In: Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS) 2019. pp. 1–15. Internet Society. https://doi.org/10.14722/ndss.2019.23386

[114] Puglisi, S., Rebollo-Monedero, D., Forné, J.: On web user tracking: How third-party http requests track users' browsing patterns for personalised advertising. In: Mediterranean Ad Hoc Networking Workshop (med-hoc-net) 2016. pp. 1–6. https://doi.org/10.1109/MedHocNet.2016.7528432

[115] Rader, E., Dourish, P.: Awareness of behavioral tracking and information privacy concern in internet users. In: Proceedings of the 10th Symposium on Usable Privacy and Security 2014. pp. 189–204. USENIX Association, `https://www.usenix.org/conference/soups2014/proceedings/presentation/rader`

[116] Rasaii, A.: Analysis scripts and raw data for bannerclick web measurements. `https://doi.org/10.17617/3.1MUYFX`, accessed: 2025-04-04

[117] Rasaii, A.: Analysis scripts and raw data for cookiewall measurements. `https://doi.org/10.17617/3.TREBZR`, accessed: 2025-04-04

[118] Rasaii, A.: Bannerclick on github. `https://github.com/bannerclick/bannerclick`, accessed: 2025-04-04

[119] Rasaii, A., Gosain, D., Gasser, O.: Thou shalt not reject: Analyzing accept-or-pay cookie banners on the web. In: Proceedings of the 2023 ACM Internet Measurement Conference. pp. 154–161. IMC '23, ACM. https://doi.org/10.1145/3618257.3624846

[120] Rasaii, A., Singh, S., Gosain, D., Gasser, O.: Exploring the cookieverse: A multi-perspective analysis of web cookies. In: Proceedings of the 2023 Passive and Active Measurement Conference. https://doi.org/10.1007/978-3-031-28486-1_26

[121] Revill, L.: Open vs. closed shadow dom. `https://blog.revillweb.com/open-vs-closed-shadow-dom-9f3d7427d1af`, accessed: 2025-04-04

[122] Richardson, L.: Beautiful soup documentation. `https://www.crummy.com/software/BeautifulSoup/`, accessed: 2025-04-04

[123] Robin, M.K.: Never-consent on github. `https://github.com/MathRobin/Never-Consent/`, accessed: 2025-04-04

[124] Roesner, F., Kohno, T., Wetherall, D.: Detecting and defending against third-party tracking on the web. In: The 9th USENIX Security Symposium (USENIX Security 2012). pp. 155–170. USENIX Association, `https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/roesner`

[125] Roth, S., Barron, T., Calzavara, S., Nikiforakis, N., Stock, B.: Complex security policy? a longitudinal analysis of deployed content security policies. In: Proceedings of the 27th Network and Distributed System Security Symposium (NDSS 2020). Internet Society. https://doi.org/10.14722/ndss.2020.24053

[126] Ruth, K., Kumar, D., Wang, B., Valenta, L., Durumeric, Z.: Toppling top lists: Evaluating the accuracy of popular website lists. In: Barakat, C., Pelsser, C., Benson, T.A., Choffnes, D.R. (eds.) Proceedings of the 22nd ACM Internet Measurement Conference 2022. pp. 374–387. ACM. https://doi.org/10.1145/3517745.3561444

[127] Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.A., Santos, I.: Can i opt out yet? GDPR and the global illusion of cookie control. In: Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security. pp. 340–351. ACM. https://doi.org/10.1145/3321705.3329806

[128] Santos, C., Rossi, A., Chamorro, L.S., Bongard-Blanchy, K., Abu-Salma, R.: Cookie banners, what's the purpose? analyzing cookie banner text through a legal lens. In: Proceedings of the 2021 ACM Workshop on Privacy in the Electronic Society. pp. 15–28. ACM. https://doi.org/10.1145/3463676.3485611

[129] Scheitle, Q., Hohlfeld, O., Gamba, J., Jelten, J., Zimmermann, T., Strowes, S.D., Vallina-Rodriguez, N.: A long way to the top: Significance, structure, and stability of Internet top lists. In: Proceedings of the 2018 ACM Internet Measurement Conference. pp. 478–493. ACM. https://doi.org/10.1145/3278532.3278574

[130] Schelter, S., Kunegis, J.: Tracking the trackers: A large-scale analysis of embedded web trackers. In: Proceedings of the 10th International Aaai Conference on Web and Social Media 2016. pp. 679–682. AAAI Press

[131] Schreiber, A.: Right to privacy and personal data protection in brazilian law. In: Vicente, D.M., de Vasconcelos Casimiro, S. (eds.) Data Protection in the Internet 2020, Ius Comparatum - Global Studies in Comparative Law, vol. 38, pp. 45–54. Springer. https://doi.org/10.1007/978-3-030-28049-9_2

[132] Schuh, J.: Building a more private web. `https://www.blog.google/products/chrome/building-a-more-private-web/`, accessed: 2025-04-04

[133] Smith, H.J., Dinev, T., Xu, H.: Information privacy research: An interdisciplinary review. Mis Quarterly 2011 **35**(4), 989–1016. https://doi.org/10.2307/41409971

[134] Smith, H.J., Milberg, S.J., Burke, S.J.: State of the information privacy literature: Where are we now and where should we go? Mis Quarterly 2011 **20**(2), 167–196. https://doi.org/10.2307/249477

[135] Smith, M., Torres-Agüero, A., Grossman, R., Sen, P., Chen, Y., Borcea, C.: A study of GDPR compliance under the transparency and consent framework. In: Proceedings of the ACM Web Conference (WWW) 2024. pp. 1227–1236. ACM. https://doi.org/10.1145/3589334.3645618

[136] Soe, T.H., Nordberg, O.E., Guribye, F., Slavkovik, M.: Circumvention by design: Dark patterns in cookie consent for online news outlets. In: Proceedings of the 11th Nordic Conference on Human-computer Interaction: Shaping Experiences, Shaping Society 2020. pp. 1–12. ACM. https://doi.org/10.1145/3419249.3420132

[137] Sørensen, J., Kosta, S.: Before and after GDPR: The changes in third party presence at public and private european websites. In: The World Wide Web Conference 2019. pp. 1590–1600. WWW '19, ACM. https://doi.org/10.1145/3308558.3313524

[138] StatCounter: Browser market share worldwide. `https://gs.statcounter.com/browser-market-share`, accessed: 2025-04-04

[139] Statista: Percentage of mobile device website traffic worldwide from 2015 to 2021. `https://www.statista.com/statistics/277125/share-of-website-traffic-coming-from-mobile-devices/`, accessed: 2025-04-04

[140] Statista Research Department: Global meta advertising revenue 2024, `https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/`, accessed: 2025-04-04

[141] Statt, N.: Apple updates safari's anti-tracking tech with full third-party cookie blocking. `https://www.theverge.com/2020/3/24/21192830/apple-safari-intelligent-tracking-privacy-full-third-party-cookie-blocking`, accessed: 2025-04-04

[142] Su, Y., Jin, L.: The impact of online platforms' revenue model on consumers' ethical inferences. Journal of Business Ethics 2021 **178**(2), 555–569. https://doi.org/10.1007/s10551-021-04798-0

[143] Temkin, D.: Charting a course towards a more privacy-first web. https://blog.google/products/ads-commerce/a-more-privacy-first-web/, accessed: 2025-04-04

[144] The National People Congress of China: Personal information protection law of the people's republic of china (PIPL). http://www.npc.gov.cn/englishnpc/c23934/202108/d7cd17fb36c84672b96d20839856f4c6.shtml, accessed: 2025-04-04

[145] Toth, M., Bielova, N., Roca, V.: On dark patterns and manipulation of website publishers by CMPs. Proceedings on Privacy Enhancing Technologies 2022 **2022**(3), 478–497. https://doi.org/10.56553/popets-2022-0082

[146] Traffective GmbH: Freechoice website. https://freechoice.club/, accessed: 2025-04-04

[147] Trevisan, M.: Priv-accept on github. https://github.com/marty90/priv-accept, accessed: 2025-04-04

[148] Trevisan, M., Traverso, S., Bassi, E., Mellia, M.: 4 years of EU cookie law: Results and lessons learned. Proceedings on Privacy Enhancing Technologies 2019 **2019**(2), 126–145. https://doi.org/10.2478/popets-2019-0023

[149] Trevisan, M., Traverso, S., Bassi, E., Mellia, M.: Cookiecheck tool on github. https://github.com/CookieChecker/CookieCheckSourceCode, accessed: 2025-04-04

[150] Tweh, B., Patel, S.: Google chrome to phase out third-party cookies in effort to boost privacy. The Wall Street Journal https://www.wsj.com/articles/google-chrome-to-phase-out-third-party-cookies-in-effort-to-boost-privacy-11579076089, accessed: 2025-04-08

[151] uBlock Origin contributors: Ublock origin - free, open-source ad content blocker. https://ublockorigin.com/, accessed: 2025-04-04

[152] Ullah, I., Safavi-Naini, R., Babar, M.A.: A comprehensive survey on privacy in online advertising. Arxiv Preprint 2020 https://arxiv.org/abs/2009.06861, accessed: 2025-04-08

[153] Urban, T., Degeling, M., Holz, T., Pohlmann, N.: Beyond the front page: Measuring third party dynamics in the field. In: Proceedings of the Web Conference 2020. pp. 1275–1286. WWW '20, ACM. https://doi.org/10.1145/3366423.3380203

[154] Utz, C., Degeling, M., Fahl, S., Schaub, F., Holz, T.: (un)informed consent: Studying GDPR consent notices in the field. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 973–990. ACM. https://doi.org/10.1145/3319535.3354212

[155] Vale, M.: Privacy, sustainability and the importance of 'and'. https://blog.google/products/chrome/privacy-sustainability-and-the-importance-of-and/, accessed: 2025-04-04

[156] Vos, J.D.: Social media, sociality, and the construction of the 'data subject'. Big Data & Society 2021 **8**(1), 1–12. https://doi.org/10.1177/20539517211003120

[157] Wehner, N., Seufert, M., Schatz, R., Hoßfeld, T.: Do you agree? contrasting google's core web vitals and the impact of cookie consent banners with actual web qoe. Quality and User Experience 2023 **8**(1), 5. https://doi.org/10.1007/s41233-023-00058-3

[158] Wood, M.: Firefox blocks third-party tracking cookies and cryptomining by default. `https://blog.mozilla.org/en/products/firefox/todays-firefox-blocks-third-party-tracking-cookies-and-cryptomining-by-default/`, accessed: 2025-04-04

[159] Yang, Z., Yue, C.: A comparative measurement study of web tracking on mobile and desktop environments. Proceedings on Privacy Enhancing Technologies 2020 **2020**(2), 4–24. https://doi.org/10.2478/popets-2020-0016

[160] Zirngibl, J., Buschmann, P., Sattler, P., Jaeger, B., Aulbach, J., Carle, G.: It's over 9000: Analyzing early quic deployments with the standardization on the horizon. In: Proceedings of the 2021 ACM Internet Measurement Conference. pp. 535–549. ACM. https://doi.org/10.1145/3487552.3487826

[161] Zuboff, S.: Big other: Surveillance capitalism and the prospects of an information civilization. Journal of Information Technology **30**(1), 75–89. https://doi.org/10.1057/jit.2015.5, `https://link.springer.com/article/10.1057/jit.2015.5`, accessed: 2025-04-08

[162] Zuboff, S.: The age of surveillance capitalism: The fight for a human future at the new frontier of power p. 704 (2019). https://doi.org/10.1080/23753234.2022.2086891

List of Figures

List of Tables