Scientific Research Publishing

Journal of Software Engineering and Applications, 2024, 17, 378-395
https://www.scirp.org/journal/jsea
ISSN Online: 1945-3124
ISSN Print: 1945-3116

Ad Blockers & Online Privacy: A Comparative Analysis of Privacy Enhancing Technologies (PET)

Siddharth M. Madikeri, Vijay K. Madisetti

School of Cybersecurity and Privacy, Georgia Institute of Technology, Atlanta, USA
Email: smadikeri3@gatech.edu, vkm@gatech.edu

How to cite this paper: Madikeri, S.M. and Madisetti, V.K. (2024) Ad Blockers & Online Privacy: A Comparative Analysis of Privacy Enhancing Technologies (PET). *Journal of Software Engineering and Applications*, **17**, 378-395.
https://doi.org/10.4236/jsea.2024.175021

Received: April 28, 2024
Accepted: May 27, 2024
Published: May 30, 2024

Abstract

Online tracking mechanisms employed by internet companies for user profiling and targeted advertising raise major privacy concerns. Despite efforts to defend against these mechanisms, they continue to evolve, rendering many existing defences ineffective. This study performs a large-scale measurement of online tracking mechanisms across a large pool of websites using the OpenWPM (Open Web Privacy Measurement) platform. It systematically evaluates the effectiveness of several ad blockers and underlying Privacy Enhancing Technologies (PET) that are primarily used to mitigate different tracking techniques. By quantifying the strengths and limitations of these tools against modern tracking methods, the findings highlight gaps in existing privacy protections. Actionable recommendations are provided to enhance user privacy defences, guide tool developers and inform policymakers on addressing invasive online tracking practices.

Keywords

Privacy Enhancing Technology (PET), Personally Identifiable Information (PII), OpenWPM, Web Privacy

1. Introduction

When a user visits a website, they are presented by a visual interface that often displays a number of content types intended at giving information or enabling interaction. This material may comprise text, photographs, videos, interactive components such as buttons and forms, and other multimedia elements. Users may interpret this engagement as a simple process of surfing or engaging with material, not aware of the complex procedures taking place behind the scenes.

Any website is mainly used to display some kind of information, such as news, some product's details, social media content, videos, and so on. However, as you read or view that content, the website collects information about you and your surfing habits. What kind of information? Information such as your device's IP address, which can identify your broad geographic location, your browser kind and settings, previous websites you've visited, and even your mouse movements or keystrokes on specific pages. The main website you are visiting collects this data, as do third-party organizations that embed their own monitoring methods in the pages.

Why do these companies want your information? Mainly for advertising purposes: to profile you and your interests so that they can offer you more relevant advertisements. However, the data can also be utilized for analytics to investigate surfing trends, and in some cases, for malicious objectives such as fraud detection. The main worry is that most of this data collection takes place without you knowing about it.

This brings up the distinction between online privacy and security. Privacy refers to the control you have over how organizations acquire, utilize, and distribute your personal data, whereas the latter's main goal is to secure your data from hackers and unauthorized access [1] [2]. While users may accept certain data gathering methods for security reasons, they may be uncomfortable with extensive tracking for targeted advertising purposes if they are not sufficiently explained to about them.

Thus, while websites provide great material and services, the underlying technologies constantly collect information about users as they explore. This raises serious privacy concerns about openness and user control, which this study seeks to investigate further.

2. Background & Related Work

2.1. Web Browsers

Web browsers function as the primary gateway for users to access and interact with content on the internet. Despite their fundamental role of rendering and presenting website information, different browsers approach user privacy and data collection in different ways. These operational differences can significantly impact the extent to which user data is collected and shared during browsing sessions, resulting in divergent user experiences across different browser platforms.

The difference lies in the telemetry and data collection practices employed by browser vendors [3]. Google's Chrome browser is known for its extensive telemetry, capturing data related to user browsing habits, preferences, and system settings. This data is leveraged by Google to improve its services, personalize user experiences, and facilitate targeted advertising within its ecosystem. In contrast, browsers like Mozilla's Firefox and the Tor Browser prioritize user privacy by minimizing data collection and offering robust privacy-focused configurations by default [4].

Moreover, browsers differ in their approaches to third-party content blocking and tracking protection mechanisms. Some browsers, such as Brave and DuckDuckGo, employ aggressive built-in ad-blocking and anti-tracking measures, significantly reducing the ability of websites and third-party entities to collect user data. Other browsers, like Chrome and Safari, offer more limited tracking protection features, relying heavily on user-enabled extensions or manual configuration to enhance privacy [5].

Another critical distinction is the level of transparency and control provided to users regarding data collection practices. Browsers that prioritize privacy, such as Firefox and Tor, emphasize user education, offering comprehensive details about data collection practices and affording users precise control over privacy settings. In contrast, browsers like Chrome may offer fewer user-facing controls, potentially obscuring the extent of their data collection and sharing methods.

These operational differences among web browsers can have profound implications for user privacy and data collection during browsing sessions. As such, understanding these distinctions is crucial for researchers and privacy conscious users alike when evaluating the effectiveness of privacy-enhancing technologies (PETs) and the overall privacy landscape of the web ecosystem.

2.2. Third-Party Online Tracking

Most, if not all, internet users are frequently subjected to various types of online tracking devices that monitor their behavior for a variety of goals, including advertising, user experience enhancement, and third-party data sale. A prevalent approach to online tracking involves the usage of embedded advertisements that include tracking vectors and scripts, that facilitate the collection of user data to enable personalized ad delivery. Similarly, website administrators employ scripts from analytics firms to obtain insights into user interactions with their platforms while also collecting data for their own analytical or commercial goals. Another typical approach of internet tracking is to include social media platform functionality, which allows users to communicate with one another. Social media platforms can track which websites their users visit by requesting embedded resources with each webpage visit [6]. Three prominent vectors for third-party tracking are cookies, HTTP requests, and JavaScript API calls.

1) Third-Party Cookies: The most common technique used to track users across websites. Cookies are small data files stored by websites in the user's browser, primarily used for maintaining session information and personalization. However, cookies from third parties have become a prevalent tracking mechanism. The tracking technique involves the inclusion of a script originating from a third-party tracker on a wide range of websites, such as those that display advertisements. The script then triggers a request from the user's browser to the server of the tracking entity. The tracker then determines whether this request contains a cookie. If a cookie is identified, the tracker associates the request with the related user profile. Otherwise, it builds a new profile and returns a *Set-Cookie* header with a newly formed cookie. The user's browser will associate the received cookie with the domain of the third-party tracker and subsequently includes it in all upcoming requests directed to that domain. This enables the tracker to follow the user across all websites that include a script that initiates a request to the tracker.

2) Third-Party HTTP Requests: This approach takes advantage of websites' intrinsic ability to include resources and material from other domains, accidentally allowing third-party entities to monitor and track user activities across various sites. To retrieve and display these resources, the browser sends individual HTTP requests to the appropriate third-party servers. While these requests are necessary for the integration of various content and functionality, they can (and do) reveal user browsing data to third-party entities. Each request includes the user's IP address, browser and system information, as well as the URL of the referring website. Furthermore, third-party HTTP requests can be used to facilitate user PII and tracking data via methods such as browser fingerprinting. Most users are unaware of the extent to which their browser activities are watched by various third-party domains, limiting their capacity to control their own private data [7].

3) Third-Party JavaScript API Calls: JavaScript APIs present on web pages like Canvas, used for rendering graphics, and WebRTC, intended for real-time communication, can be misused for fingerprinting users based on rendering variations or exposing IP addresses. Other APIs exposing device sensors, battery status, and hardware details enable constructing unique fingerprints for persistent user identification across sites. While these APIs are used for the functionality of websites, the widespread use of third-party scripts accessing these APIs raises privacy concerns by enabling covert tracking without user consent.

2.3. Why Do We Need Ad Blockers?

Ad blockers are PETs that identify and block the loading of ads on web pages. Their essential job is to filter out requests for ad content, scripts, and tracking codes, thus preventing their delivery to the user's browser. Ad blockers not only improve the surfing experience by decreasing clutter and improving page load speeds, but they also provide an important line of protection against invasive tracking technologies used by advertising businesses.

One of the fundamental operational techniques used by ad blockers is to store and continuously update lists of known advertising servers, tracking sites, and related behaviors. These lists, also known as filter lists or blocklists, are analyzed and compiled by ad blocker developers or community-led initiatives. When a user's browser attempts to load a web page, the ad blocker intercepts and inspects the request, checking it against the filter lists. If a request matches one of the entries in the list, the ad blocker prevents it from being loaded, essentially blocking the corresponding advertisement or tracking code. Several recognized ad blockers have acquired broad popularity among privacy-conscious users, with each offering distinct features and capabilities. The ad blockers utilized in this study are explored in greater detail in the following sections.

Ad blockers are excellent tools for improving user privacy and surfing experience, but their effectiveness is contingent on their ability to keep up with the ever-changing landscape of online advertising and tracking techniques. As a result, continual study and evaluation of ad blocker performance against developing threats is critical to guaranteeing their sustained effectiveness in preserving user privacy on the web.

2.4. Exemplary Ad Blockers

This study focuses on four widely-used and representative ad blockers: Ghostery, Privacy Badger, uBlock Origin and AdLock. These tools represent diverse approaches to ad blocking and anti-tracking protection, offering a comprehensive assessment of the current state of privacy defense mechanisms in general.

1) Ghostery: This ad blocker is a complex privacy solution that combines ad blocking features with tracker detection and blocking functionality. In addition to blocking out adverts, Ghostery keeps a comprehensive database of known trackers and provides users with extensive information about the tracking technologies used on visited websites. Such transparency enables consumers to make informed decisions about their online privacy and selectively block or allow specific trackers. Ghostery, the company behind the PET, has expanded its privacy protection features by introducing a privacy suite. This contains a private browser, a secret search engine, and ad-blocking software. Ghostery offers itself as a comprehensive solution for customers looking to improve their online privacy across different touchpoints by providing an integrated ecosystem of privacy-focused solutions [8].

2) Privacy Badger: Developed by the Electronic Frontier Foundation (EFF), Privacy Badger is a browser extension that employs a unique approach to tracking protection. Rather than using a pre-defined blocklist, this PET dynamically learns and blocks non-consensual trackers through heuristic analysis of third-party domains and their tracking behavior [9]. This adaptive strategy seeks to provide strong protection against emerging tracking and monitoring techniques while reducing the possibility of accidentally censoring genuine content.

3) uBlock Origin: Built as a lightweight and efficient ad blocker, uBlock Origin is renowned for its comprehensive filter lists and low resource consumption. Unlike other ad blockers, which primarily target visual ad elements, uBlock Origin takes a broader approach, blocking network-level requests to known advertising and tracking domains. This proactive strategy not only improves page load time but also reduces the privacy threats associated with ad distribution and tracking methods [10].

4) AdLock: This PET combats ads through a multi-layered approach. It uses

DNS filtering to intercept ad server requests, utilizes the hosts file to block known ad domains, and creates a local VPN to analyze and block ad content within app and browser traffic. Like any other ad blocker, it also employs filter lists to identify and block a wide range of ad formats. This combination of techniques disrupts the ad delivery process, offering users an ad-free browsing experience.

2.5. Prior Work

Several important studies have set the foundations for better understanding the mechanics and consequences of web monitoring techniques. However, the dynamic nature of the web ecosystem needs ongoing re-evaluation and study to stay up with the ever-changing landscape of tracking tactics and privacy-enhancing remedies.

Englehardt, S., and Narayanan, A. conducted a study in 2016 titled "*Online tracking*: *A* 1-*million-site measurement and analysis*," [11] which remains one of the most prominent studies in this sector. This comprehensive study conducted a large-scale measurement and analysis of online tracking technologies across 1 million websites. The authors created the OpenWPM platform, which will also be used in this study, and identified several tracking mechanisms, such as the usage of third-party cookies and canvas fingerprinting strategies.

Another study, "*Third-Party Web Tracking*: *Policy and Technology*" [12] by Mayer, J., and Mitchell, J. (2012), investigated the policy and technology elements of third-party web monitoring, stressing the complex ecosystem of data transfers between websites and third-party entities. The authors showed that the absence of transparency and user control over these data flows poses serious privacy concerns.

Krishnamurthy B. and Wills C. contributed to the understanding of web privacy with their paper "*Privacy Diffusion on the Web*: *A Longitudinal Perspective*" [13] in 2009. This study looked at the evolution of privacy-compromising technology across website categories over time, giving light on the widespread use of tracking methods and the implications for user privacy.

While the above studies have produced useful insights and acted as basic works in the field of web privacy, they are somewhat outdated in this age of AI and rapid online adoption of ad technologies. Our study aims to address the following gaps and limitations present in prior work:

1) Outdated Measurements: The above studies were conducted several years ago. Since then, the online tracking ecosystem has evolved significantly, with the introduction of new tracking vectors and the ongoing modification of old techniques. This study presents an updated measurement and analysis of the current state of online tracking, incorporating the most recent innovations in the field.

2) Lack of Comparative Analysis: The above-mentioned research has mostly concentrated on finding and measuring tracking mechanisms, without extensively comparing the performance of various PETs in reducing these concerns.

This research fills this gap by methodically and comprehensively evaluating the efficacy of common ad blockers to third-party tracking.

3) Usage of a Small/Generic Dataset: Prior work has examined tracking mechanisms across only a small pool of websites or a single category of websites. This study explicitly explores the impact of different website categories—Entertainment, Gaming, E-commerce, Sports, and the Tranco list, on the performance of ad blockers. This provides valuable insights into the nuances of web privacy across diverse domains.

By addressing the above limitations and gaps, this study aims to contribute to the ongoing discussion on web privacy by providing an up-to-date and comprehensive analysis of the online tracking landscape, quantifying the effectiveness of commonly used ad blockers. This research endeavor is crucial for informing users, developers, and policymakers about the evolving privacy risks and guiding the development of more robust privacy protection mechanisms.

3. Data Source of Tested Websites

Our dataset contains five categories of websites:

1) Entertainment: Netflix, Spotify, Prime Video, etc.,

2) Gaming: Twitch, Discord, GameSpot, etc.,

3) E-commerce & Shopping: Amazon, Ebay, Walmart, etc.,

4) Sports: ESPN, Bet365, Marca, etc.,

5) Tranco: A list of websites that sources data from multiple providers—Alexa, Cisco Umbrella, Quantcast and Majestic, and averages out the rankings over a thirty-day period.

Each category consists of the top 300 most visited websites worldwide collected up until January 2024. These websites were extracted from https://www.semrush.com and https://www.similarweb.com. Even though a few websites overlap across the above dataset, the inclusion of a range of categories enables a close-to-accurate measurement of the effectiveness of each ad blocker.

Influence of Website Categories on Ad Blocker Performance

While ad blockers are useful tools for protecting user privacy and decreasing exposure to unwanted advertising, their effectiveness and efficacy can vary greatly across different types of websites. The fundamental cause of this mismatch is the unique advertising and tracking tactics used by websites across various domains and organizations.

For instance, most entertainment and gaming websites rely heavily on advertising revenue as their primary monetization approach. As a result, these websites may use more aggressive and intricate advertisement distribution strategies, including obfuscation techniques and lesser-known ad blocking technologies. This can lead to situations in which more popular ad blockers struggle to maintain comprehensive and up-to-date filter lists, resulting in reduced efficacy on these types of websites.

In contrast, e-commerce and shopping websites may prioritize user experience and trust over invasive third-party ads. These websites rely on direct sales and transactions as their primary revenue, incentivizing them to minimize disruptive advertisements that could hinder user shopping experience. As a result, when compared to entertainment and gaming websites, ad blockers may encounter fewer challenges in blocking the relatively straightforward ads employed by e-commerce platforms.

Sports websites, on the hand, present a unique challenge due to their frequent integration of real-time updates and live content. Ad blockers may struggle to keep up with the dynamic nature of these websites. Failure to promptly and proactively update blocklists could result in ad blockers inadvertently allowing certain ads or tracking scripts.

Furthermore, the effectiveness of ad blockers can be influenced by the prevalence of third-party trackers and analytics tools integrated into different website categories. For example, social media platforms and content portals may employ a wide array of third-party services for user engagement analysis, personalization, and targeted advertising. To effectively prevent these many tracking systems, ad blockers must maintain extensive filter lists and apply advanced detection techniques, which can be difficult given the ever-changing nature of these technologies [14].

The adjustments in the advertisement and tracking systems from one category of websites to another make it necessary for ad blockers to be competent in flexible and adaptable practices. This requires constant scrutiny, evaluation, and modifications of filter lists that are tailored to suit various web domains to maintain an uninterrupted ad blocking service of high efficiency which can contribute towards safeguarding consumer privacy and ensuring smooth browsing across the entire web ecosystem.

4. Methodology

To conduct a comprehensive analysis of online tracking mechanisms and evaluate the effectiveness of ad blockers, this study leverages the OpenWPM platform, a powerful web privacy measurement tool, developed by researchers from Princeton University.

4.1. OpenWPM Platform

OpenWPM (**Figure 1**) is an open-source tool designed to study web privacy and online tracking. It is built on the Firefox browser, allowing for automated browsing sessions that collect large datasets on tracking methods.

The tool's main strength is its ability to extract and record a variety of tracking vectors, such as third-party cookies, HTTP requests, JavaScript API calls, and other browser-side instrumentation data. This feature is enabled by integrating custom measurement modules that link to various browser APIs and instrumentation points.

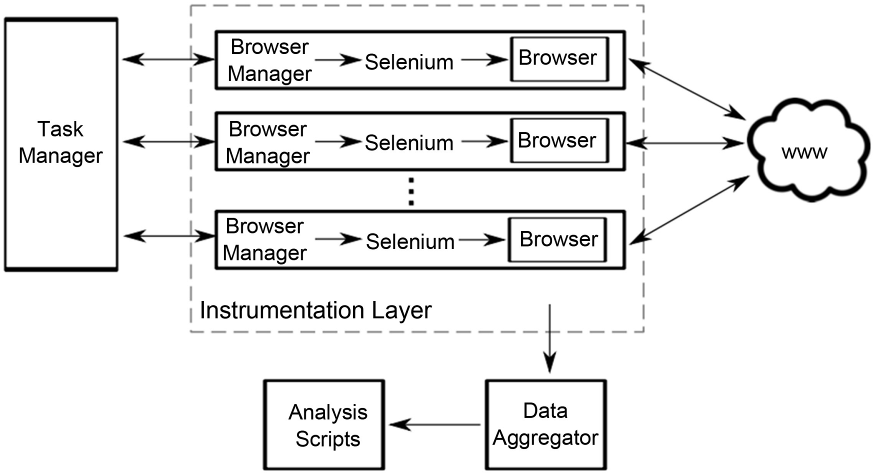

Figure 1. High-level overview of OpenWPM.

We make use of the browser automation, task manager and data collection infrastructure of the tool to crawl across websites and save data. Browser automation is done using Selenium, which by default, occurs using a headless instance of the Firefox web browser. We then adjusted this setting to "native" by creating custom commands with unique profiles for each ad blocker. Native browsers render online pages in a way that simulates the real user experience, allowing them to simulate user interactions more accurately with a website. As a result, data is collected and analysed more accurately.

The tool interacts with Firefox using a Python-based interface, which automates the browsing process. The task manager is set up to manage multiple browsers at once, allowing for methodical examination of certain browser properties and extensions. This arrangement, adapted to researchers' needs, provides a high-level foundation that can be expanded as needed.

The crawl results are recorded in SQLite databases, from which further analysis is performed.

4.2. Implementation

To conduct a comprehensive evaluation of online tracking mechanisms and the effectiveness of the ad blockers across different website categories, a robust data collection process is implemented. For each website category, that includes Entertainment, Gaming, E-commerce, Sports and Tranco, the study involves two distinct crawling modes: a "vanilla" mode without any ad blockers present, and multiple "non-vanilla" or "ad blocker" modes, each configured with a specific ad blocker.

The study makes use of separate browser profiles for each crawling mode, ensuring isolation and preventing potential interference between configurations. The setup process for each website category involves initializing the respective browser profiles and crawling the target websites. Upon completion of the crawling process, the resulting data is stored in category-specific folders, with each folder containing separate SQLite databases corresponding to the vanilla mode and individual ad blocker modes.

The SQLite databases created by the crawling process contain an extensive collection of browsing data organised into several tables. The tables relating to HTTP requests, JavaScript executions, and cookie management are particularly interesting to this study as they provide critical information on the identified tracking vectors: third-party HTTP requests, JavaScript API calls, and cookies.

We created custom Python scripts to extract and analyze pertinent tracking data. These scripts process SQLite databases by extracting information on the third-party vectors established by the visited websites.

Identifying third-party entities is a critical aspect of this process. The scripts leverage the *tldextract* [15] Python package to remove public suffixes (e.g., ".com", ".co.uk") from URLs and compare the resulting domain names. URLs with different subdomains are considered as belonging to the same parent third-party entity, enabling a comprehensive analysis of tracking activities across multiple subdomains.

By extracting and quantifying these tracking vectors across various website categories and crawling modes (vanilla and non-vanilla), this study hopes to shed light on the prevalence of online tracking mechanisms and the relative effectiveness of various ad blockers in mitigating these threats.

5. Results & Observations

The study quantifies the relative effectiveness of each ad blocker in reducing user exposure to third-party tracking entities by comparing the distinct third-party domain counts observed in the "vanilla" crawling mode (without any privacy protections) to those observed in the various "ad blocker" modes (with specific ad blockers enabled). The number of distinct third-party domains identified in ad blocker mode is significantly lower than in vanilla mode, indicating a stronger level of protection from online tracking.

The next sections provide a thorough examination of the three third-party tracking vector counts observed across website categories in both modes.

5.1. Analyzing Online Tracking Vectors

5.1.1. Third Party Cookies

Third-party cookies are the most common tracking vector evaluated in this study. As explained in section 2.2, third-party cookies are set by domains other than the one the user is currently visiting, allowing tracking entities to create persistent user profiles across several websites.

Figure 2 presents a comparative analysis of third-party cookie counts observed during crawls of the top Tranco websites in the "vanilla" mode and "ad block" mode with uBlock Origin as the ad blocker in place. Across most websites examined, the deployment of the ad blocker resulted in a significant reduction in the number of third-party cookies encountered, with cookie counts sometimes

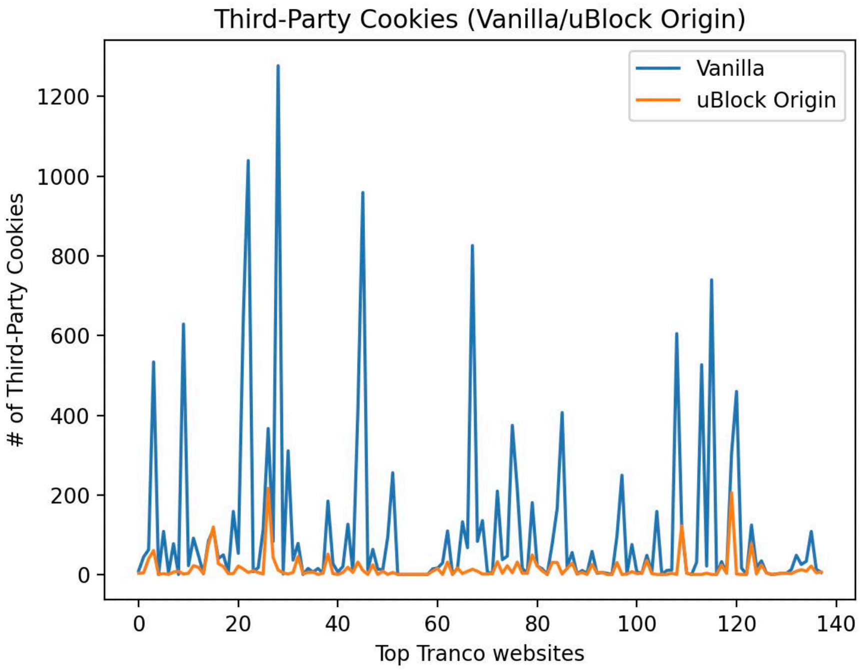

Figure 2. Third-party cookies in vanilla and uBlock Origin ad blocker mode for the Tranco category.

even dropping to zero.

It is worth mentioning that the crawls undertaken in this study do not involve user authentication or any kind of login activity, which could explain the lower reported third-party cookie counts for websites that need users to have accounts. However, this study emphasizes the complex tracking methods used by many website categories, which reflect their unique operating models and data collection tactics.

Overall, the results obtained from the third-party cookie analysis demonstrate the need of PETs in mitigating widespread tracking vectors across a diverse range of websites. The substantial reduction in third-party cookie counts when ad blockers are enabled shows the potential privacy benefits these tools can provide to users seeking to regain control over their online privacy and limit the dissemination of their browsing data to third-party entities.

5.1.2. Third-Party HTTP Requests

HTTP requests initiated by the user's browser to retrieve resources from domains other than the primary website, can expose user data and facilitate tracking by third-party entities.

Figure 3 gives a graphical view analysis of the number of third-party HTTP request counts observed after crawling top sports websites in "vanilla" mode and AdLock "ad block" mode. The graph clearly illustrates a substantial reduction in the number of third-party HTTP requests when the AdLock extension is enabled, highlighting its purpose in blocking this tracking vector.

The graph (**Figure 3**) shows that "vanilla" mode crawls generate a substantial

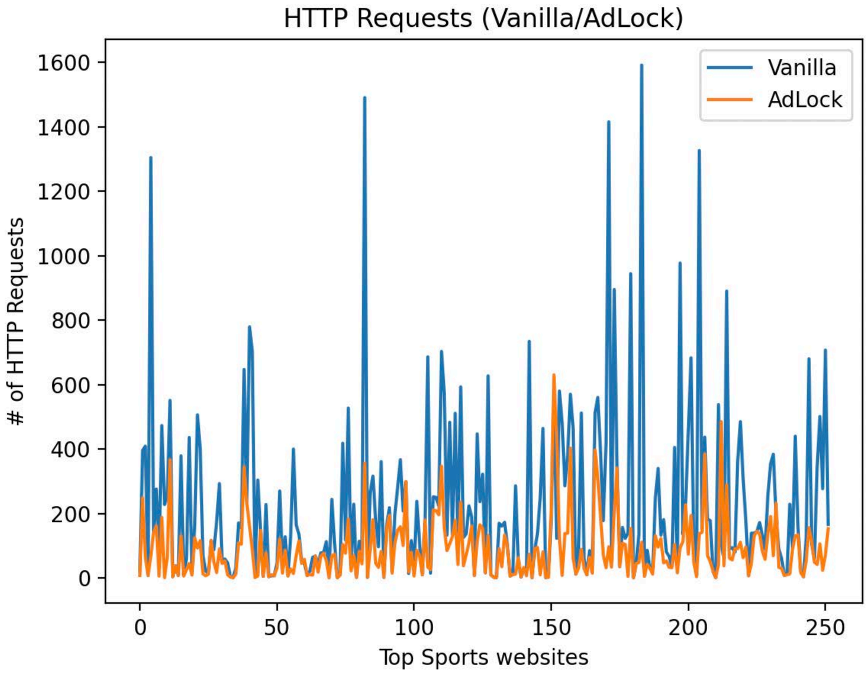

Figure 3. Third-party HTTP requests in vanilla and AdLock ad blocker mode for the Sports category.

number of third-party HTTP requests, often exceeding 1000 for certain domains. The implementation of AdLock results in a significant drop in these requests. This pattern indicates that a significant amount of third-party HTTP requests come from tracking companies and ad networks, which are effectively prevented by the PET.

The effectiveness of these ad blockers in restricting third-party HTTP request tracking is clear across all website categories evaluated, as illustrated by one of the graphs given. Regardless of category, using these ad blockers drastically minimises the number of third-party HTTP requests encountered while browsing any website.

Ad blockers significantly reduce the quantity of third-party HTTP queries, but they do not completely eliminate them. This is due to the legitimate integration of third-party resources, such as Content Delivery Networks (CDNs) or performance optimisation scripts, which may continue to create HTTP requests even when privacy measures are activated.

5.1.3. Third-Party JavaScript API Calls

Certain JavaScript APIs exposed by modern web browsers, while intended for legal purposes such as graphic loading or real-time communication, can be used by third-party entities to monitor and fingerprint users.

The graph in Figure 4 shows a comparative examination of third-party JavaScript API call counts found during crawls of the top e-commerce and shopping websites in both modes. The graph shows a significant reduction in the number of third-party JavaScript API calls when the Ghostery addon is enabled, albeit

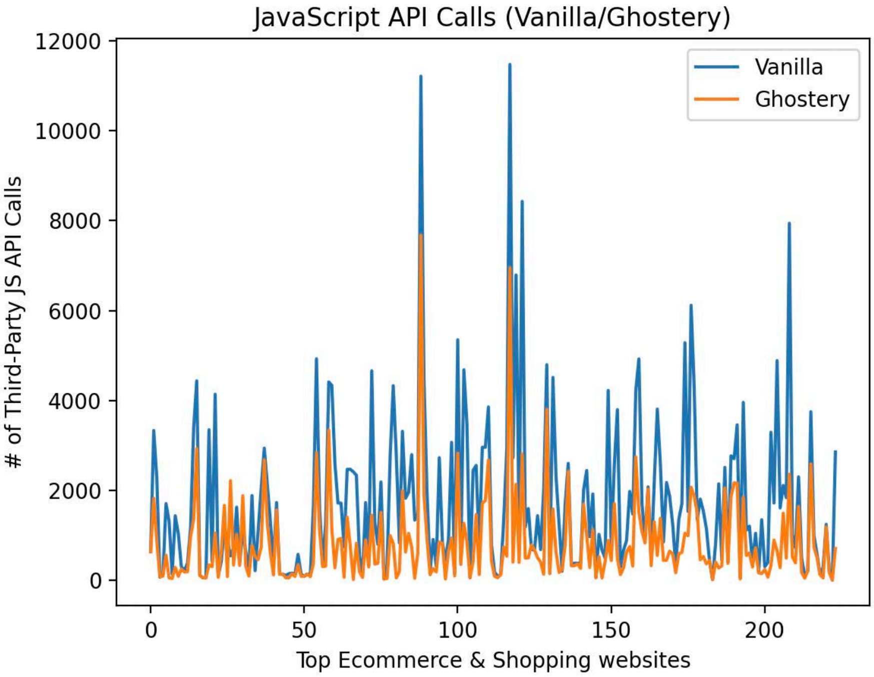

Figure 4. Third-party JavaScript API Calls in vanilla and Ghostery ad blocker mode for the E-commerce category.

with a less obvious effect than third-party cookie and HTTP request mitigation.

While "vanilla" mode crawls contain a high number of third-party JavaScript API calls, potentially exceeding 10,000 for some websites, the Ghostery ad blocker causes a moderate decrease in these call counts. This shows that certain third-party JavaScript API calls may come from tracking entities or adverts that are partially blocked by the browser extension.

As previously stated, not all JavaScript API requests are used for tracking or fingerprinting. Many legitimate third-party scripts and components embedded in websites may leverage JavaScript APIs for core functionality or user experience enhancements. As a result, the reduction in third-party JavaScript API call counts recorded with ad blockers enabled may be less significant than the reduction observed for third-party cookies or HTTP requests, which are primarily utilised for tracking. This discovery shows the complex dynamics of this tracking vector, as well as the issues related with API usage and prospective tracking attempts.

5.2. Analysis across Website Categories

For the **Entertainment** category, as shown in Table 1, the vanilla mode makes 57,946 HTTP requests, 40,716 cookies, and 393,902 JavaScript API calls on average. The use of uBlock Origin reduces these numbers to 15,865 HTTP requests (72.6% reduction), 3133 cookies (92.3% reduction), and 48,761 JavaScript API calls (87.6% reduction). Privacy Badger achieves reductions of 68.9%, 86.1%, and 72.9%, respectively, while Ghostery reduces tracking by 71.1%, 87.2%, and

Table 1. Browser measurements across top Entertainment websites in vanilla and ad block modes.

Entertainment websites	Vanilla	uBlock Origin	Privacy Badger	Ghostery	AdLock
HTTP Requests	57,946	15,865	18,046	16,774	20,999
Cookies	40,716	3133	5666	5227	9423
JS API Calls	393,902	48,761	106,972	75,170	83,895

80.9%. AdLock demonstrated a reduction of 63.8%, 76.9%, and 78.7% across the three vectors.

Coming to the results in the **Sports** category (See Table 2), the vanilla mode counts 57,454 HTTP requests, 33,367 cookies, and 513,339 JavaScript API calls. uBlock Origin reduces these percentages by 68.9%, 92.9%, and 85.1%, respectively. Privacy Badger achieves reductions of 65.5%, 91.4%, and 74.2%, whereas Ghostery decreases tracking by 66.6%, 92.3%, and 80.7%. AdLock reduces 61.1%, 76.7%, and 74.9% across all three tracking technologies.

Coming to the results in the **Gaming** category displayed in Table 3, when in vanilla mode, there are 57,055 HTTP requests, 35,590 cookies, and 405,379 JavaScript API calls involved on average. uBlock Origin reduces these percentages by 78.0%, 97.5%, and 89.9%, respectively. Privacy Badger achieves reductions of 73.6%, 96.1%, and 76.8%, whereas Ghostery decreases tracking by 75.8%, 97.0%, and 83.8%. AdLock reduces 66.1%, 73.8%, and 83.1% over all three vectors.

In the category of **E-commerce and Shopping** (See Table 4), the vanilla mode makes 52,341 HTTP requests, 28,302 cookies, and 382,988 JavaScript API calls. uBlock Origin decreases these figures by 58.3%, 71.9%, and 69.4%, respectively. Privacy Badger achieves reductions of 44.0%, 65.8%, and 48.5%, while Ghostery reduces tracking by 51.4%, 69.1%, and 53.0%. AdLock demonstrates a reduction of 45.6%, 56.1%, and 56.5%.

For the **Tranco** top sites (See Table 5), the vanilla mode makes 32,089 HTTP requests, 15,759 cookies, and 173,707 JavaScript API calls. uBlock Origin reduces these numbers by 58.4%, 86.4%, and 68.9%, respectively. Privacy Badger reduces them by 46.1%, 79.6%, and 51.4%, while Ghostery reduces tracking by 53.5%, 84.5%, and 59.1%. AdLock demonstrates a reduction of 42.9%, 62.7%, and 54.3%. Table 5 shows this breakdown in detail.

5.3. Analyzing the Ad Blockers

According to our testing results and data, uBlock Origin is the most effective PET, continuously achieving the best mitigation rates for third-party HTTP requests, cookies, and JavaScript API calls across all five website categories. This higher performance is due to uBlock Origin's proactive approach to updating filter lists, which allows it to efficiently block a wide range of known tracking sites, scripts, and network requests. Its strong content filtering policies and

Table 2. Browser measurements across top Sports websites in vanilla and ad block modes.

Website Sports	Vanilla	uBlock Origin	Privacy Badger	Ghostery	AdLock
HTTP Requests	57,454	17,880	19,811	19,192	22,382
Cookies	33,367	2355	2858	2563	7769
JS API Calls	513,339	76,273	132,295	98,865	129,006

Table 3. Browser measurements across top Gaming websites in vanilla and ad block modes.

Gaming websites	Vanilla	uBlock Origin	Privacy Badger	Ghostery	AdLock
HTTP Requests	57,055	12,553	15,086	13,833	29,367
Cookies	35,590	875	1381	1063	9308
JS API Calls	405,379	40,851	93,879	65,652	68,693

Table 4. Browser measurements across top E-commerce websites in vanilla and ad block modes.

Commerce-E websites	Vanilla	uBlock Origin	Privacy Badger	Ghostery	AdLock
HTTP Requests	52,341	21,826	29,283	25,430	28,440
Cookies	28,302	7940	9662	8742	12,409
JS API Calls	382,988	117,185	197,158	179,971	166,499

Table 5. Browser measurements across top Tranco websites in vanilla and ad block modes.

Tranco websites	Vanilla	uBlock Origin	Privacy Badger	Ghostery	AdLock
HTTP Requests	32,089	13,344	17,306	14,926	18,301
Cookies	15,759	2148	3213	2437	5881
JS API Calls	173,707	54,051	84,510	71,084	79,352

ability to prevent in-line script injection make it effective at minimising JavaScript-based tracking approaches.

On the other end of the numbers, AdLock consistently exhibits the lowest mitigation rates across two third-party tracking vectors—HTTP requests and cookies. This shows that AdLock may be less effective at completely eliminating online tracking that uses network requests and cookie-based identifiers. However, it is worth noting that AdLock's primary goal is ad-blocking rather than comprehensive anti-tracking, which may explain its inferior effectiveness in these specific vectors.

One interesting observation is that Privacy Badger consistently allows the highest number of third-party JavaScript API calls across all five categories, despite its effectiveness in mitigating the other two tracking vectors. This finding can be due to Privacy Badger's unique approach in blocking online tracking, which relies on heuristics and machine learning to detect and block non-consensual tracking over time. While this method is effective for many tracking mechanisms, it may initially allow certain JavaScript-based tracking to occur until its heuristics can accurately identify and block the offending scripts or domains.

5.4. Top Third-Party Domains that Track

Across all categories, some domains were more frequently blocked by the ad blockers, i.e., these domains were exclusively used as trackers as they were blocked by all the ad blockers. Table 6 displays the prominent tracking domains. *doubleclick.net* and *pubmatic.com* are renowned internet advertising domains recognized for their extensive tracking methodologies. Google uses many tracking domains, two of which are prominently included on the list of commonly blacklisted websites [16]. This suggests a strong centralization of user tracking, which might have significant consequences for user privacy and data security.

6. Conclusions

Our detailed study demonstrates that PETs are effective in protecting user privacy. The analysis reveals significant reductions in the number of third-party tracking mechanisms when browsing with ad blockers enabled. This shows that PETs, particularly ad blockers, can substantially limit online tracking employed by advertisement companies and help protect user privacy.

While PETs give users more control over their data, it raises concerns for advertising companies and content providers who monetize their offerings through ads. Thus, the usage of anti-tracking technologies like ad blockers could potentially disrupt the ad companies' business model, which can lead to conflicts between user privacy and the sustainability of online content displayed by these companies. Thus, evaluating the abilities and results of PETs entails examining both stakeholders' needs to find optimal solutions that balance privacy with revenue generation.

Table 6. Browser measurements across top Tranco websites in vanilla and ad block modes.

Third-Party Domain
https://doubleclick.net/
https://googlesyndication.com/
http://googletagservices.com/
https://pubmatic.com/
https://amazon-adsystem.com/

Overall, our study contributes to the ongoing discussion of web privacy and the usage of PETs like ad blockers as a solution. With growing awareness among internet users regarding online tracking techniques' implications for their privacy and the availability of more anti-tracking solutions, we must continue assessing their effectiveness and potential impact on content providers and users.

7. Current & Future Work

While this study sheds light on the efficacy of PETs in minimising tracking vectors, several areas require further investigation. To begin, comparing PET performance to upcoming tracking techniques such as sophisticated fingerprinting technologies and browser-based mining is critical to ensuring the solutions' long-term usefulness. As the tracking landscape evolves, it is critical to monitor PETs' ability to combat fresh threats.

Our current research examines the potential impact of PETs on website functionality and user experience. This study quantified the mitigation of tracking vectors but did not investigate the trade-offs between privacy protection and potential website breakage or degradation of user experience. Understanding these trade-offs is essential for developing well-balanced PET solutions that preserve both privacy and usability.

Furthermore, current studies are examining the design and implementation of novel PET solutions that deal with the limitations of current technologies while providing more comprehensive and user-friendly privacy protection. Innovative ways and technology may be required to stay up with the ever-changing tracking landscape and fulfil the various needs and preferences of users.

Acknowledgements

We thank the reviewers for their comments that improved the paper.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] GitHub (2021) OpenWPM. https://github.com/openwpm/OpenWPM

[2] Barth, S., Ionita, D. and Hartel, P. (2023) Understanding Online Privacy—A Systematic Review of Privacy Visualizations and Privacy by Design Guidelines. *ACM Computing Surveys*, **55**, 1-37. https://doi.org/10.1145/3502288

[3] Pau, K.N., Lee, V.W.Q., Ooi, S.Y. and Pang, Y.H. (2023) The Development of a Data Collection and Browser Fingerprinting System. *Sensors*, **23**, 3087. https://doi.org/10.3390/s23063087

[4] Telemetry Collection and Deletion. Firefox Help. https://support.mozilla.org/en-US/kb/telemetry-clientid

[5] Apple (2019) Privacy—Features. https://www.apple.com/privacy/features/

[6] Bouke, M.A., Abdullah, A., Alshatebi, S.H., Zaid, S.A. and Atigh, H.E. (2023) The

intersection of targeted advertising and security: Unraveling the mystery of overheard conversations. *Telematics and Informatics Reports*, **11**, 100092. https://doi.org/10.1016/j.teler.2023.100092

[7] Libert, T. (2015) Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites. https://arxiv.org/abs/1511.00619

[8] Ghostery (2019) Ghostery Makes the Web Cleaner, Faster and Safer! https://www.ghostery.com/

[9] Walsh, R. (2019) Privacy Badger Review. Improve Your Privacy—Stop Tracking. https://proprivacy.com/adblocker/review/privacybadger

[10] DeVaney, S. (2021) uBlock Origin—Everything You Need to Know about the Ad Blocker. Firefox Add-Ons Blog. https://addons.mozilla.org/blog/ublock-origin-everything-you-need-to-know-about-the-ad-blocker/

[11] Englehardt, S. and Narayanan, A. (2016) Online Tracking. *Proceedings of the* 2016 *ACM SIGSAC Conference on Computer and Communications Security*, Vienna October 24-28 2016, 1388-1401. https://doi.org/10.1145/2976749.2978313

[12] Mayer, J.R. and Mitchell, J.C. (2012) Third-Party Web Tracking: Policy and Technology. 2012 *IEEE Symposium on Security and Privacy*, San Francisco, 20-23 May 2012, 413-427. https://doi.org/10.1109/SP.2012.47

[13] Krishnamurthy, B. and Wills, C. (2009) Privacy Diffusion on the Web. *Proceedings of the* 18*th International Conference on World Wide Web—WWW'*09, 20-24 April 2009, 541-550. https://doi.org/10.1145/1526709.1526782

[14] Gunnarsson, P., Jakobsson, A. and Carlsson, N. (2022) On the Impact of Internal Webpage Selection When Evaluating Ad Blocker Performance. https://www.ida.liu.se/~nikca89/papers/mascots22.pdf

[15] Kurkowski, J. (2024) John-Kurkowski/Tldextract. GitHub. https://github.com/john-kurkowski/tldextract

[16] Frost, P. (2017) How to Track Google Analytics Conversions across Domains | MSROI. Main Street ROI. https://www.mainstreetroi.com/case-study-how-to-track-google-analytics-conversions-across-domains/