IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 October 2024, accepted 4 November 2024, date of publication 8 November 2024, date of current version 2 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3494539

RESEARCH ARTICLE

Do Cookie Banners Respect My Browsing Privacy? Measuring the Effectiveness of Cookie Rejection for Limiting Behavioral Advertising

MATEO ORMENO[1], HA DAO[2], VALERIA HERSKOVIC[1], AND KENSUKE FUKUDA[3], (Member, IEEE)

[1]Computer Science Department, Pontificia Universidad Católica de Chile, Santiago 7821093, Chile
[2]Max Planck Institute for Informatics, Saarbrücken, 66123 Saarland, Germany
[3]National Institute of Informatics, Tokyo 1018430, Japan

Corresponding author: Mateo Ormeno (maormeno@uc.cl)

This work was supported in part by the NII Internship Program; and in part by the National Center for Artificial Intelligence CENIA FB210017, Basal ANID.

ABSTRACT Online behavioral advertising (OBA) is a method within digital advertising that exploits web users' interests to tailor ads. Its use has raised privacy concerns among researchers, regulators, and the media, emphasizing the need for a reliable mechanism to measure its prevalence. However, there is a lack of systematic research on how user consent choices affect OBA presence, and no open-source frameworks exist for large-scale automated OBA measurement. To address this, we design and implement *OpenOBA*, a new framework for automated OBA discovery on the web. *OpenOBA* is a general, modular, and scalable framework to support essentially any OBA measurement. With it, we conduct a study to measure the impact of three user consent choices for cookies on OBA, uncovering a complex online privacy landscape. We first confirm the presence of OBA by comparing the increased likelihood of encountering ads from a specific topic, i.e., *Style & Fashion*, when browsing with an artificially induced behavior versus when browsing without any particular behavior. Then, we find that the *Accept All* choice significantly raises the number of OBA ads. For the *Reject All* option, on the other hand, we observe that it reduces the number of unique third-party tracking cookie hosts (tracker domains) by around 70%, yet it still shows ads related to the user's interests. Notably, we also find that OBA ads are only served through Google-related domains across the three banner interaction configurations used, despite the involvement of up to 191 different tracker domains in the *Accept All* configuration. This underscores the dominant role of major players in the OBA ad market. Finally, to foster reproducibility and further research, we open-sourced our framework and released all data and analysis scripts.

INDEX TERMS Web privacy, web privacy measurements, OBA, OBA presence measurements, *OpenOBA*.

I. INTRODUCTION

In the last decade, the Internet has become firmly ingrained in fundamental aspects of daily life [1]. While offering numerous enhancements, it has also increased the visibility of people's digital footprints [2], leaving a significant trail of online behavior and interactions. In 2023, 65.7% of the world's population used the Internet [3]. Its widespread usage, coupled with sophisticated web tracking mechanisms [4], has expanded online data collection without boundaries, placing user data as a valuable commodity. Similarly to oil, this commodity propels one of the largest

The associate editor coordinating the review of this manuscript and approving it for publication was Gang Li.

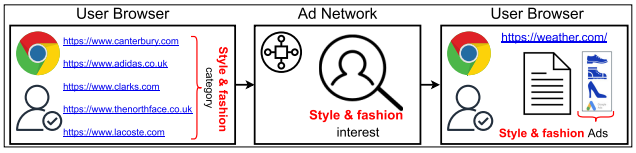

FIGURE 1. Online behavioral advertising explanation.

emerging industries of the Internet: *online advertising*, with a market estimated at USD 257.97 billion in 2024 and projected to reach USD 431.76 billion in 2029 [5].

The online advertising industry thrives on promoting products and services through Internet platforms, leveraging data analytics and strategies to target specific audiences. Central to this is user data —encompassing behavior, preferences, and demographics— which is fundamental in crafting and delivering personalized ads. This is known as Online Behavioral Advertising (OBA), defined as the practice in online advertising wherein information about the interests of web users is incorporated in tailoring ads [6]. OBA involves collecting data on users' browsing habits, search history, and website interactions, enabling advertisers to dynamically create detailed user profiles, which are used to predict their interests and preferences. For example, when a user visits several websites related to the *Style & Fashion* category, this browsing activity is tracked and categorized under the *Style & Fashion* interest by the ad network. Subsequently, when the user navigates to an unrelated site, like *weather.com*, the ad network recognizes their interest in *Style & Fashion* and serves relevant *Style & Fashion* ads in the browser (see Figure 1).

Owing to its dependence on web tracking, OBA has raised privacy issues among internet users. This has sparked discussions about the benefits of personalized ads versus their potential intrusiveness [7], [8], [9] and has been accompanied by privacy concerns from researchers, regulators, and the press [6], [10], [11], highlighting the need for a reliable method to measure OBA prevalence. However, there is currently no open-source framework for conducting experiments involving automatic measurements of OBA presence. This lack of accessible, scalable tools has made it difficult for researchers to conduct consistent and comprehensive studies across different web environments. Furthermore, while many studies have explored the storage and behavior of third-party cookies, particularly during user interactions with cookie banners [12], [13], [14], [15], there remains a significant gap in understanding how these interactions directly influence OBA presence. Specifically, little research has been done to systematically assess how varying consent choices (e.g., *Accept All*, *Reject All*, or *No Action*) impact the volume and targeting of OBA ads.

In this paper, we present a novel framework designed to measure the prevalence of OBA on the web automatically. Our approach leverages automated browsing experiments that capture both third-party tracking behaviors and their relation to OBA presence, offering a comprehensive view of how user interactions, particularly with cookie banners, influence the occurrence of OBA. The main contributions of the paper are as follows:

1) We design and implement *OpenOBA*, a flexible framework for automated measurements of OBA presence, which enables the construction and analysis of experiments simulating web users' experiences with targeted online advertising based on their browsing history. Our framework includes all necessary input components to simulate user behaviors browsing the web while collecting encountered ads, thus enabling the detection of OBA ads as intended (§ III).
2) Using *OpenOBA*, we then conduct a comprehensive measurement study to analyze the presence of OBA across various scenarios involving cookie consent interactions (§ IV-A). We first find a significant increase in both the quantity and targeting precision of ads when users adhere to the *Accept All* cookies option. Conversely, users adhering to *Reject All* cookies or taking *No Action* effectively see their targeted ads reduced, but not completely eliminated, indicating that some level of profiling continues (§ IV-B1). By analyzing third-party tracking cookies, we find that this heightened targeting is underpinned by extensive third-party tracking, as demonstrated by higher numbers of unique tracking cookies associated with the *Accept All* configuration. We then focus on the ad-serving domains, finding that all targeted ads come from *googleadservices.com* and *doubleclick.net*, although there are up to 191 tracker domains that implant their cookies on the user browser (§ IV-B2).
3) We conclude by discussing the implications (§ V-A) of our work and exploring the acknowledged research opportunities that could provide a deeper and further understanding of OBA measurements (§ V-B). We also suggest possible research directions to enhance online privacy transparency on the web. (§ V-C).

II. BACKGROUND AND RELATED WORK

A. THIRD-PARTY WEB TRACKING

Third-party web tracking refers to the practice of an entity, other than the domain directly visited by the user, that identifies and collects information about web users. It is used for online behavioral advertising that targets users with ads based on their profiles or interests, which may be considered threatening to user privacy.

Third-party web tracking is complicated in that it relies on a wide variety of web tracking technologies, ranging from stateful tracking mechanisms that recognize users by retrieving information stored on users' machines to stateless tracking mechanisms that recognize users without storing any information. Undoubtedly, many tracking techniques have been developed to maximize the benefits of tracing user browsing behavior. These have been largely studied to get a better understanding of the risks involved. Some approaches detected the main mechanisms behind specific web tracking

techniques [13], [16], [17], [18], [19], [20], analyzed cookie synchronization that bypasses same-origin policies on the web [21], [22], [23], and explored the unique identifier stored in a cookie or embedded as a parameter in a URL [24].

B. WEB PRIVACY MEASUREMENT FRAMEWORKS

In the field of web privacy, researchers have developed various frameworks and tools to systematically measure and analyze the privacy implications of web tracking mechanisms. These tools (e.g. OpenWPM [13], FPdetective [23], Chameleon [25], and Common Crawl [26]) are essential for understanding and studying how personal data is collected, shared, and used across the web. OpenWPM was developed using Python and utilizing a Firefox browser with the Selenium automation tool for website navigation. This tool is well-regarded for its comprehensive features, speed, and capacity to handle large-scale measurements. Its effectiveness was demonstrated through its widespread use in diverse web measurement studies [19], [27], [28], [29]. Therefore, our research integrates its web crawling capabilities into our framework implementation.

C. PRIVACY REGULATIONS AND COOKIE BANNER MEASUREMENTS

Cookie banners are popups on websites used to inform visitors about the use of cookies and to obtain their consent as required by data privacy regulations around the world, such as the GDPR in [30] in Europe. These regulations mandate transparency and user consent before any personal data collection occurs through cookies.

Prior research on cookie banner interaction showed significant variation in compliance and effectiveness. Some findings suggested that higher acceptance rates are due to the downplaying of rejection options [31], misleading or vague language violating the specificity purpose [32], and the influence of the banner location on user consent [33]. More closely aligned with cookie banner effectiveness and usage measurements, Jha et al. [34] showed that ignoring cookie banners offered biased and partial views of the Web and that web tracking was much more prevalent after accepting, while Rasaii et al. [14] specifically quantified the effects of cookie banner interaction, indicating that accepting cookies sent around five times more third-party cookies than other options, with a similar trend for tracking cookies.

D. OBA PRESENCE MEASUREMENTS

OBA is the practice of tailoring advertising based on the tracking of user's online activities [35]. Tracking providers usually track an individual web usage history across multiple sites to target ads. Past research studied OBA in the context of the online advertising market. Balebako et al. [36] presented a methodology for measuring behavioral targeting based on web history to evaluate the effectiveness of privacy tools to limit OBA. Carrascosa et al. [6] introduced a methodology for measuring and understanding OBA in the online advertising market by relying on training artificial online personas representing behavioral traits. Solomos et al. [37] proposed a methodology for detecting cross-device tracking and measuring the factors affecting its performance by triggering cross-device targeted ads.

Unlike previous studies, our work measures OBA presence with third party tracking in sight, contemplating an ad context given by specific cookie banner interactions. To the best of our knowledge, *OpenOBA* is the first open-source framework that enables large-scale automated measurements of OBA on the web, while also exploring the complexities of web cookies and user tracking, providing a detailed perspective largely overlooked in earlier research.

III. OpenOBA: A FRAMEWORK FOR AUTOMATIC MEASUREMENTS OF THE PRESENCE OF ONLINE BEHAVIORAL ADVERTISING

In this section, we present *OpenOBA*, a framework designed to automate the measurement of the presence of OBA on the web. Note that we are not claiming or attempting to reverse engineer the complex series of real-time bidding; we merely intend to build a framework to detect if there is any correlation between the ad displayed to a user and their past browsing behavior. We outline the design requirements of the framework and its core components.

A. DESIGN REQUIREMENTS

To accomplish our goal of automatically measuring OBA ads, we need a solution that can run experiments involving the simulation of Internet users' behavior that would make them a target of OBA based on the simulated interest user browsing. The framework must be robust enough to handle external uncertainties while maintaining the integrity and reproducibility of the simulations. In addition, it should be easily adaptable and flexible to updates in response to emerging web technologies to remain relevant and accurate in its analyses. We propose the following essential features and characteristics that a model would need to comply with the requirements detected through our research:

1) **Configurable experiment interface**: Provides accessible configurable parameters throughout each experiment stage.
2) **Web crawler**: Provide web crawling as the primary mechanism for simulating user behavior and executing researcher-defined simulations. The web crawler mimics human browsing patterns to interact with websites.
3) **Advertisement scraping**: Capture advertisements from websites in a way such that their content can be analyzed and categorized.
4) **Experiment continuity**: Support conducting experiments over multiple days, allowing for pauses and the resumption of previous sessions, crucial due to the unpredictable nature of OBA.

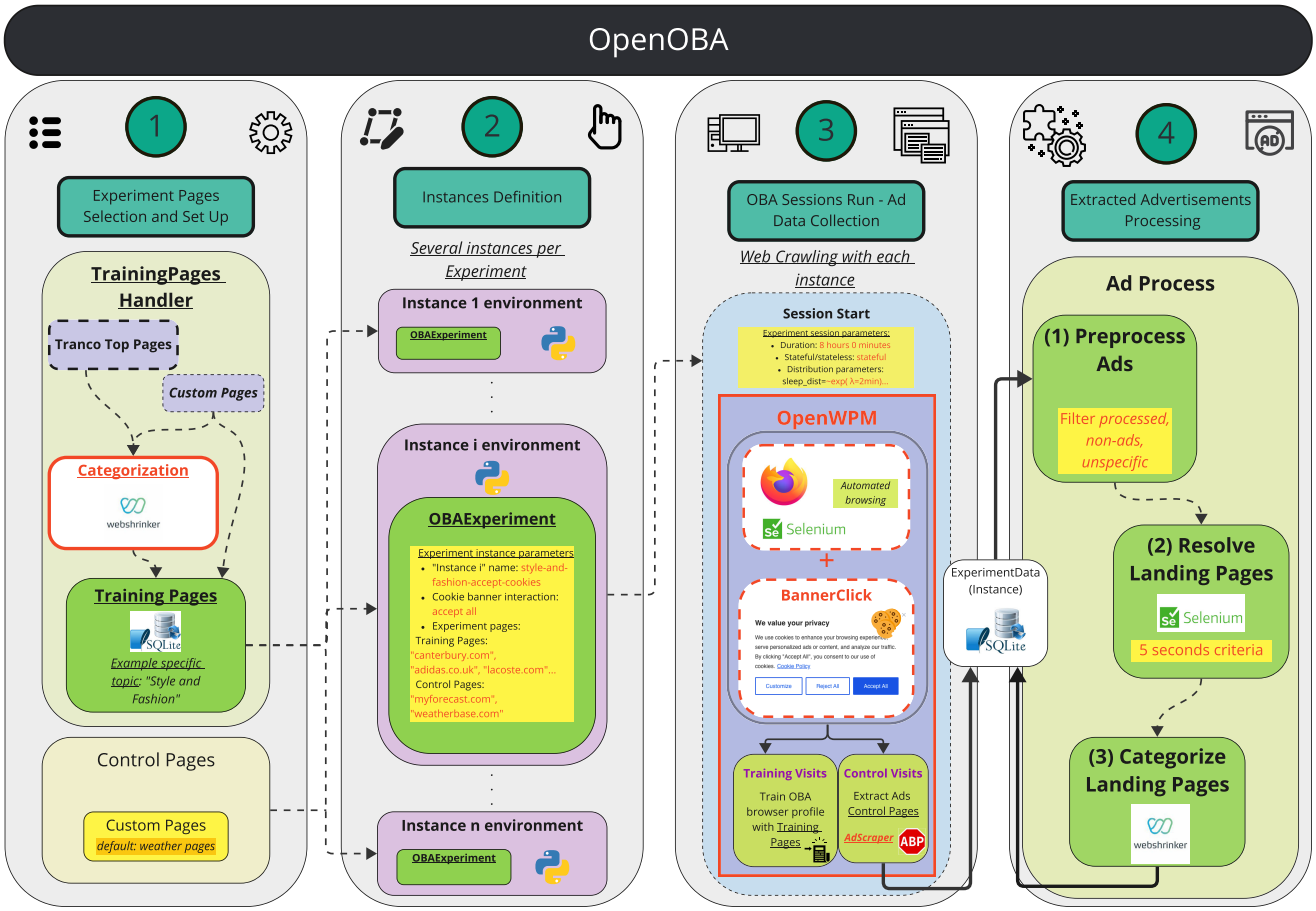

FIGURE 2. High-level overview of OpenOBA.

B. DESIGN AND IMPLEMENTATION

At a high level, the *OpenOBA* framework treats OBA as a black box, checking whether a user profile's browsing activities target OBA. The framework is structured into four stages, from the initial experiment definition to measuring OBA. These stages are: (1) Experiment Pages Selection and Set Up, (2) Instances Definition, (3) Sessions Run - Ad Data Collection, and (4) Ad Data Processing. Each stage, including the main components and interactions involved, is described in detail (see Figure 2).

1) EXPERIMENT PAGES SELECTION AND SET UP

The first step involves setting up the environment and preparing what we refer to as *training pages*. These are the websites that induce specific browsing behavior depending on the nature of the experiment.

To bolster our framework, we integrate the Tranco Top 10k web pages, a research-oriented top site ranking list resistant to manipulation [38]. We store a cached version of this list from April 2023 in an SQLite database and categorize each URL using the WebShrinker service [39]. Webshrinker is a service that can categorize domains according to the marketing-oriented taxonomy of the Interactive Advertising Bureau (IAB) [40]. This taxonomy provides a two-level hierarchy of categories and an indicator of certainty for identified categories, which we also store. Depending on the experiment's needs, these data can be accessed via the **TrainingPagesHandler** component, which can retrieve from the cached database, update the list by repeating the process for a newer version of Tranco, or handle and categorize *custom pages*.

Next, we define the *control pages* set. These are web pages used to gather ads and whose content is used to evaluate OBA for each experiment. These pages should include a sufficient number of display ads and belong to a neutral and clear topic to minimize noise in the browser profiles. Following previous work where similar *control pages* were used [6], [41], we select popular websites from a distinguishable web category, *weather pages*[1], that consistently display a high amount of dynamic ads and meet our requirements.

[1]http://myforecast.com/, http://weatherbase.com/, http://theweathernetwork.com/, http://weather.com/,http://weather2umbrella.com/

2) INSTANCES DEFINITION

The next stage involves defining the various instances of the experiment. In our framework, an experiment comprises multiple instances of the OBAExperiment component. Each instance is essentially a version of the experiment configured with specific settings. By having multiple instances with identical configurations but controlled variations, we can systematically compare how these variations induce observable differences in advertisements and network activity data. An instance configuration depends on the values of two types of parameters: experiment instance parameters and experiment session parameters.

1) **Experiment instance parameters**: The parameters are designed to remain constant across all sessions within a given instance, reflecting the core nature of the experiment. They can also include specific adjustments for each instance, depending on the focus of the study. These are:
   - *Instance name*: Unique experiment instance identifier.
   - *Cookie banner interaction*: Specific interaction to be taken when encountering cookie banners (e.g., *Accept All*, *No Action*, *Reject All*) on visited pages.
   - *Experiment pages*: Predefined sets of *training pages* and *control pages*. Depending on the nature of the study, different pages can be assigned for each instance's sets.
2) **Experiment session parameters**: These parameters are specific to individual sessions and do not alter the overall nature of the experiment:
   - *Duration*: The duration for an experiment session.
   - *Stateful/Stateless crawl*: To either keep or clear the state and browsing history/navigation data in the browser during the session.
   - *Statistical distributions parameters*: Values for the statistical distributions that define browsing behavior, such as wait and sleep times between visits, page loading times, and the ratio of training to control visits to be performed.

Dividing an experiment into multiple executable units with custom configurations (i.e., *instances*) enables broader, more flexible measurements and comparisons. Researchers can customize each instance to focus on specific factors by manipulating and controlling various web browsing variables. The complexity and number of these variables determine the number of instances required for the experiment.

3) OBA SESSIONS RUN - AD DATA COLLECTION

After the instances are defined and set up, the next step is arguably the most important since it is where OBA happens, and the data for its measurement is collected. Each *session run* is performed by each instance's **OBAExperiment**. Simply put, through a web crawler, it trains a profile browser with *training pages* while randomly visiting *control pages* where OBA could be manifested, saving the data.

This feature is built on two maintained tools focused on research: OpenWPM [13] and Bannerclick [14]. This places the implementation on a shared academic ground with other academic lanes related to web privacy while still satisfactorily addressing all of its technical needs. For web crawling capabilities, we use OpenWPM [13], a popular Python open-source tool for web privacy measurements that uses the Selenium web driver [42] for the Firefox browser (see § II-B). Following their structure and SQLite schema, we can track, configure, relate, and analyze the different browsers, website visits, screenshots, page sources, web traffic, browser profile management, and many other functions. This is the core component of our framework. In addition, we use Bannerclick [14] to incorporate automatic handling of cookie banners in the webpages visited for the OBA measurement experiments.

The first step is to create a fresh new browser profile (for every new *experiment instance*) or to load an existing experiment profile with a given *experiment name* to resume an *experiment instance* that has already been created in a previous session. The crawling process is structured into *training visits*, which train the instances browser profiles with the content of the *training pages*, and *control visits*, which expose the browser profiles to advertisements from the *control pages* and extract their content. To minimize profile contamination with the *control pages* content, *control visits* must be fairly spaced out in between *training visits*, which requires manipulating the training/control visits ratio to some degree while still maintaining some randomization to avoid bot detection. This is achieved by *control_visits_rate*, an *experiment instance parameter* (from 0 to 100%) that will determine the probability of choosing a *control visit* each time the crawler needs to choose the upcoming visit, set as 20% by default. Besides, to always ensure a higher number of *training visits*, each time that *training* is chosen, the following 0 to 2 visits will also be randomly locked to *training visits* as well. The specific process followed when choosing each type of visit is described next:

1) ***Training visits***:
   - *Page selection*: Determine the number of web pages to visit from the training set, with outcomes ranging from one to three pages.
   - *Waiting time*: Introduce a random wait time before page access, following an exponential distribution with a mean duration of 3 minutes, as noted in [43].
   - *Page Visit*: Use the OpenWPM crawler to visit a randomly selected training page. Handle the cookie banner using the BannerClick tool upon page load.
   - *Post-visit waiting*: After a successful page load, wait for a uniformly distributed period between 45 and 90 seconds to allow website aggregators to categorize the browser profile based on the training page topic.
   - ** Repeat the last three steps for each page selected in Page Selection*
2) ***Control visits***:
   - *Page selection*: Select a single random website from the designated *control pages*, known for a high density of ads.

- *Waiting time*: Pause for a random duration similar to the wait time in the training visit's second step.
- *Page visit*: Navigate to the control page using the OpenWPM crawler. Upon successful page load, manage the cookie banner via the Bannerclick tool to accept cookies as configured.
- *Advertisement loading*: Wait for a period uniformly distributed between 30 and 60 seconds to ensure ads are fully loaded.
- *Ad scraping*: Identify all possible ads within the page source by matching lists of EasyList selectors with every loaded frame without directly visiting the AdURLs to avoid profile contamination. Capture screenshots of each element matched as an ad and compile its URL.
- *Final screenshot*: Conclude the visit by taking a full-page screenshot after ad extraction using OpenWPM's designated command.

Each experiment instance is run for several sessions (preferably days) with the same browser profile to ensure that the tracking aggregators have enough time to recognize the pattern of the *training pages*' topics and relate the browser profile to it.

Note that the time complexity of our proposed approach primarily depends on the number of browsing scenarios and the volume of data collected during each experiment. Specifically, the framework's complexity scales linearly with the number of simulated sessions and the volume of third-party tracking requests processed per session, ensuring it can handle large-scale experiments efficiently.

4) EXTRACTED ADVERTISEMENTS PROCESSING

The last stage of an OpenOBA experiment aims to identify and categorize the ads collected during the crawling sessions. The data of each extracted ad is stored in two formats: a screenshot image and the URL, namely *AdURL*, which is the URL assigned to the particular ad from the ad-serving provider (e.g., DoubleClick, Taboola, PubMatic).

While screenshots provide a visual record of the advertisement's explicit content and layout, our primary focus is on AdURLs. AdURLs are the links embedded in advertisements that direct users to the landing pages. Landing pages hold the actual value of the advertisement. They are what the ad network or aggregator decides to show to the user. Without these landing pages, fully understanding the advertisement's content, message, target audience, and commercial intent becomes impossible. By resolving *AdURLs* and following the redirects to reach the final landing page, we collect the ads displayed on the *control pages*.

In particular, the process followed during this last step is the following.

(1) ***Preprocess ads***: Gather all extracted ads that have not been processed yet (without an associated categorized landing page) and apply the *non-ad* and the *non-specific* filters. *Non-ad* filter excludes all the ads whose URL leads to its ad provider's settings or help pages, while *non-specific* ads redirect to a search engine index page. In this way, we cover the anticipated but unavoidable false positives that our Ad Scraping tool (see previous stage) incurs.
(2) ***Resolve landing pages***: Resolve all the remaining AdURLs using a Selenium stateless browser. This browser will wait until the window URL value stays the same for 10 seconds uninterrupted, indicating that the actual window URL is the final *landing page* URL. This URL is then updated as the landing page URL of all the ads in the database that share the same AdURL.
(3) ***Categorize landing pages***: Categorize all the uncategorized landing pages using the same WebShrinker API of the first stage.

This process allows us to associate ads with categories separately on each experiment instance, enabling the users to study the occurrence and persistence of OBA across several aspects when comparing different experiments' *training pages*, parameter variations, and even instance parameter variations.

IV. INVESTIGATING ONLINE BEHAVIORAL ADVERTISING IN THE CONTEXT OF COOKIE INTERACTION

In this section, we perform experiments to evaluate the effectiveness of web tracking by exploring the prevalence of OBA in the context of cookie banner interactions, particularly in instances where the use of cookies has been explicitly declined.

Here, we extend a methodology similar to [6], [36], and [37], where we also cover the handling of cookie banners to study the relation between the different options (i.e., *Accept All*, *No Action*, *Reject All*) and the exposure to OBA and web tracking. This approach leverages our OpenOBA tool to set up and run automatic experiments to measure the presence of OBA, which will be described in the following subsections.

A. EXPERIMENT TRAINING, CONTROL PAGES SELECTION AND SET UP

1) TRAINING AND CONTROL PAGES SELECTION

As stated in § III-B1, we first identify the input for our experiment.

Training pages selection: we need a set of *training pages* belonging to the same category to represent a specific web browsing persona. This category has to be easily distinguishable from others but with enough variety of pages so that the behavior is more realistic as a truly interested and knowledgeable user. The pages also require to have an interactive cookie banner to be handled. The *Style & Fashion* category matches our requirements given its relevance in the online advertising environment, representing the second largest e-commerce category [44]. To this end, we use the *TrainingPages* component of the OpenOBA framework and retrieve the 20 most popular websites from the Tranco Top 10K sites that are confidently categorized as *Style and Fashion* by Webshrinker [39], which also have a cookie banner interaction.

TABLE 1. Census measurement configurations.

Experiment name	VM-A1	VM-B1	VM-C1
Cookie banner interaction	Accept All	No Action	Reject All
Training pages	*Style & Fashion* from Tranco top 10K sites*		
Control pages	Default (weather pages)		
Session duration	8 hours /day		

* Nine sites are also categorized as *Shopping*.

Control pages selection: Since the default *OpenOBAcontrol pages* have been carefully selected, we use those pages for our measurement (see § III-B1).

2) EXPERIMENT SETUP

To make sure there is no common identifier belonging to other tracking techniques shared between them [23], [45], we install the OpenOBA framework on three different Google Cloud Platform Compute Engine *n2d-standard-4* instances. Each instance is set with the same parameters for the *OBAExperiment* component of *OpenOBA*, but each with a different *cookie banner choice* option. Thus, the complete experiment, namely *OBA Run*, consists of three instances running *OpenOBA* in parallel. We create the instances in London, United Kingdom, to make sure that the crawlings are run within a GDPR-covered region, which requires websites to ask for user's consent to use cookies, mostly as cookie banners. In this way, every instance has the same *Style & Fashion*-categorized *training pages* being crawled to train a browser profile according to our proposed framework in § III-B3, while automatically handling the *cookie banner choice* on every visit. When a control visit is performed, the ads are extracted to be processed and analyzed later. To maximize exposure to the advertising ecosystem, we operate three instances simultaneously for eight hours a day over six days. The details of each experiment instance configuration are summarized in Table 1.

To validate our results, before starting the *OBA Run*, we create the *Random Run* by making an exact copy of each of the experiment's instances in their corresponding machine with two key differences: (i) no training visits and (ii) stateless browsers with clean browsing histories for every control visit. We enforce the exact same cookie banner action and control visits as their corresponding *OBA Run* instance, with the same *control pages* the same number of times each, in the same order. This way, we isolate the effects of our variables of interest (i.e., manifesting a specific browsing behavior and taking a specific cookie banner action) by contrasting our results to the ones obtained from browsers exposed to the same online ad environment that did not go through a process of simulating a specific user behavior.

Note that, to ensure the reproducibility and applicability of our study, we followed the experimental setup inspired by [46]. The experimental details are outlined in Table 3.

B. RESULTS

Here, we present our findings, offering insights into the effectiveness of web tracking and the prevalence of OBA under various cookie consent scenarios.

1) PRESENCE OF OBA AND IMPACT OF COOKIE BANNER CONFIGURATIONS

Since we intend to study OBA exposure throughout a prolonged period, we use *the number of unique ads by visit* as the unit to measure OBA presence, where two ads are considered *unique by visit* if they are either extracted during a different control page visit or if they have different AdURLs. In other words, we eliminated duplicate ads during each visit. Also, we only keep the ads whose AdURL domain belongs to auctioneers authorized to sell the impression inventory on the *control pages*. This way, since each ad is individually relatable to OBA, the main criterion is the presence/absence of OBA ads per visit. To this end, we consider 789 ads for *Random Run* and 1,376 ads for *OBA Run* in our analysis.

In Figure 3, we present the distribution of ad categories for each configuration in *Random Run*, and in Figure 4 for *OBA Run*. For each category, we calculate the percentage of *unique ads by visit* and rank them from highest to lowest. Note that an ad could belong to one or more categories. Our analysis reveals that *Random Run* exhibits a diverse distribution of ad categories, with notable percentages in *Technology & Computing* and *Hobbies & Interests* across all three banner configurations. Conversely, for *OBA Run*, there is a significant increase in the *Shopping* category, from 14.5% in *Random Run* to 48.04% in *OBA Run*, making it the most common category in both *Accept All* and *Reject All* configurations. Further examination of our *training pages* shows that nine out of 20 are categorized as *Shopping*, establishing it as the second most represented category. Additionally, the *Style & Fashion* category, absent in all *Random Run* instances, appears in nearly 1.96% of all ads in *OBA Run*, with a 2.2% presence in both *Accept All* and *No Action* configurations. This means that specific browsing behaviors increase the likelihood of encountering OBA ads. When focusing on the impact of cookie banner choices on OBA ads, we also observe that the *Accept All* option prominently displays more OBA ads than the *No Action* and *Reject All* options. However, these consentless options still show ads that target user interests.

2) AD-SERVING DOMAINS AND THIRD-PARTY TRACKING COOKIES ANALYSIS

We extend our analysis and include the traffic data collected during the experiment to study third-party tracking cookies in each cookie banner choice interaction and compare it to the OBA differences suggested by the ads data. Here, similar to previous works [14], [47], we define a cookie to be associated with *third-party web tracking* when its *host* appears in any of the blocklists, including EasyList [48], EasyPrivacy [49], and Adservers [50].

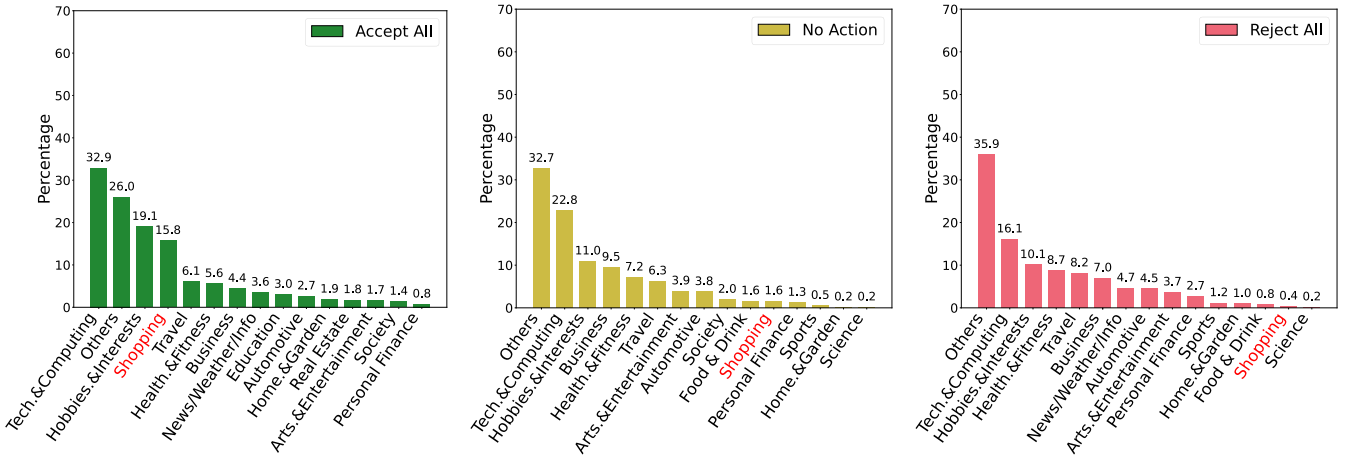

FIGURE 3. Breakdown of ads categories shown on control visits for ***Random Run***.

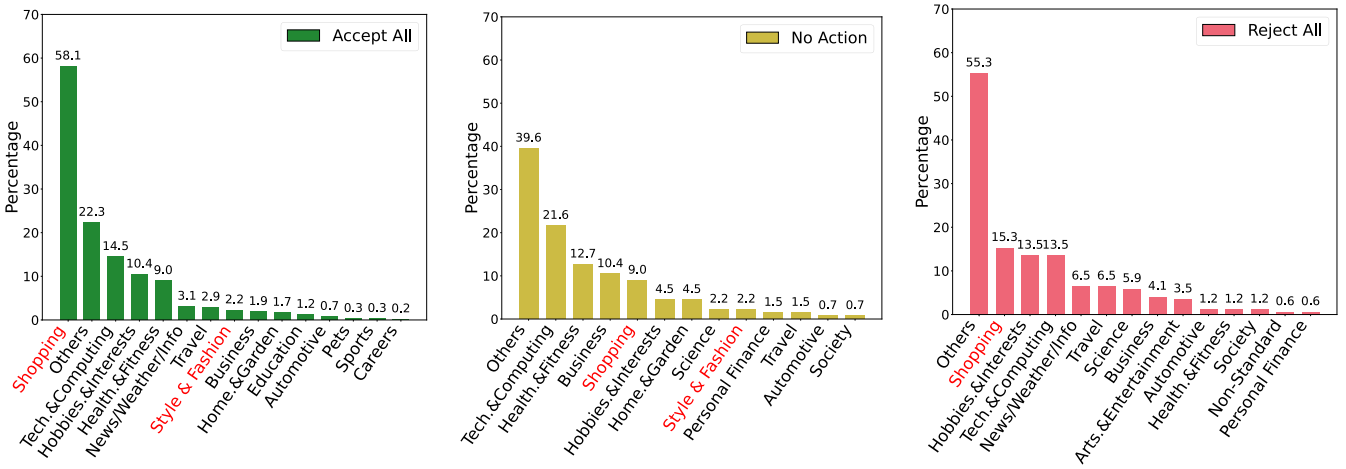

FIGURE 4. Breakdown of ads categories shown on control visits for ***OBA Run***.

We first show the unique tracking cookie hosts for the three *OBA Run* instances among six sessions and their *overlap* in Figure 5. We confirm that *Accept All* cookies results in the highest number of unique third-party tracking cookie hosts, e.g., 191 unique cookies among six sessions, indicating significant exposure to tracking. On the other hand, *No Action* and *Reject All* do not introduce any new third-party tracking cookie hosts; they merely reduce the number present in the *Accept All* configuration. Choosing to take *No Action* or *Reject All* cookies significantly reduces the number of unique third-party tracking cookie hosts encountered, with totals of 52 and 57, respectively. This represents a decrease of approximately four times compared to the *Accept All* configuration, emphasizing the effectiveness of these actions in limiting third-party tracking cookies. We also note that a subset of 49 third-party tracking cookies persist regardless of the user's cookie preferences. Worryingly, the primary objective of many entities within the digital advertising and analytics sectors seems to be to gather as much data as possible, with little regard for user consent or preferences.

We then break down the top 10 tracker domains sorted by the number of third-party tracking cookies for each *OBA Run* instance in Table 2. There are four tracker domains, *rubiconproject.com*, *openx.net*, *3lift.com*, and *adsrvr.org* (green), which show a significant decrease in number and fall from the top 10 tracker domains in both the *No Action* and *Reject All* configurations compared to the *Accept All* configuration. Interestingly, *adsrvr.org* (red), the ad server for The Trade Desk [51], a real-time bidding ad exchange, displays a different pattern. The number of cookies associated with *adsrvr.org* is higher in the *No Action* and *Reject All* configurations compared to the *Accept All* configuration. This unexpected result suggests that cookie banner consent mechanisms result in more persistent tracking efforts by particular domains.

We also present the number of OBA ads by the ad-serving domain in Figure 6. Interestingly, we find that although various tracking domains set tracking cookies, only two domains serve OBA ads: *googleadservices.com* and *doubleclick.net*. Both of these domains are Google-related. This indicates that

TABLE 2. Top 10 tracker domains sorted by the total number of cookies involved in OBA Run.

#	Tracker Domain	# Cookies
1	adnxs.com	5,967
2	casalemedia.com	2,004
3	rubiconproject.com	1,869
4	openx.net	1,524
5	mediavine.com	1,505
6	demdex.net	1,434
7	criteo.com	1,189
8	3lift.com	1,079
9	evergage.com	994
10	adsrvr.org	946

(a) *Accept All*

#	Tracker Domain	# Cookies
1	evergage.com	998
2	mediavine.com	815
3	stickyadstv.com	641
4	criteo.com	619
5	adnxs.com	515
6	demdex.net	471
7	casalemedia.com	387
8	media.net	327
9	bounceexchange.com	274
10	postrelease.com	205

(b) *No Action*

#	Tracker Domain	# Cookies
1	evergage.com	1,034
2	mediavine.com	575
3	stickyadstv.com	495
4	criteo.com	435
5	adnxs.com	419
6	demdex.net	391
7	bounceexchange.com	286
8	casalemedia.com	263
9	media.net	231
10	tribalfusion.com	194

(c) *Reject All*

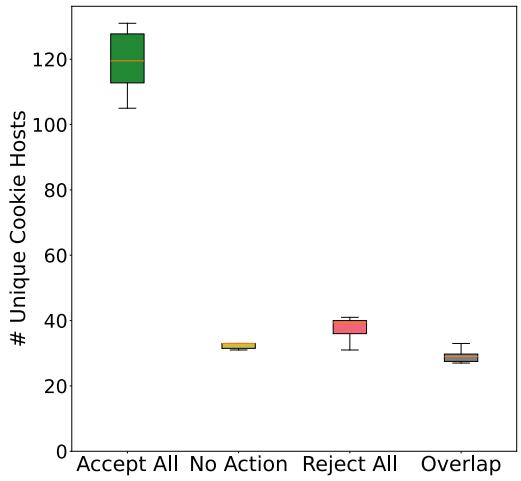

FIGURE 5. Distribution of unique third-party tracking cookie hosts (tracker domains).

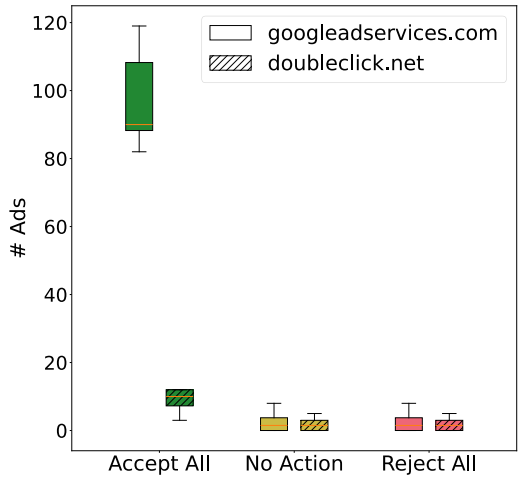

FIGURE 6. Breakdown number of OBA ads by ad-serving domain.

Google has a significant presence and control in the OBA ad market.

V. DISCUSSION

A. IMPLICATIONS

The implications of this work are far-reaching, offering valuable insights into the dynamics of OBA and its interaction with user consent mechanisms. *OpenOBA*, the framework for OBA measurement presented in this paper, helps to enhance our understanding of how targeted advertising operates, revealing the significant influence of cookie consent choices on ad volume and targeting. Notably, the findings demonstrate that even rejecting cookies does not fully eliminate personalized ads, raising concerns about the effectiveness of current consent frameworks and prompting a reevaluation of their privacy protections. Furthermore, the dominance of Google-related domains in ad-serving, despite the involvement of numerous third-party trackers, highlights the concentrated power of major platforms in the OBA ecosystem. These results have important implications for policymakers, suggesting the need for stronger regulatory measures to protect user privacy more effectively. By providing OpenOBA as an open-source framework, the study also facilitates further research and fosters reproducibility, enabling others to explore the complex relation between user behavior, tracking, and targeted advertising. Overall, this work paves the way for the development of more robust privacy tools and a more comprehensive understanding of digital advertising's impact on user privacy.

B. OpenOBA FRAMEWORK APPLICATIONS

Our *OpenOBA* framework presents numerous opportunities for research aimed at gaining deeper insights into OBA measurement and web tracking practices.

Firstly, *OpenOBA* framework can be used to compare OBA and web tracking across different profile interests. This setup would show how OBA and tracking behaviors vary by interest, revealing which interests attract more targeted advertising and tracking activities. The expected outcome is a detailed understanding of category-specific advertising strategies and tracking intensity.

In addition, it can be used to assess the impact of geographic location on OBA and web tracking by using multiple browsers trained with the same list of *training pages* but operating these browsers from different geographic locations. This approach allows for the comparison of OBA and tracking behaviors across different locations, showing how regional differences affect advertising targeting and

TABLE 3. Study reproduction.

		ID	Criterion	Description
Dataset (see § IV-A1)		C1	Analyzed sites	Training pages are selected from the top 10,000 Tranco sites, with control pages from five weather-related websites.
		C2	Analyzed pages	A JSON file listing all analyzed pages was published in [62].
		C3	Site/page selection	Describes the selection process for analyzed sites.
		C4	Multiple measurements	Consistent pages were used across all experiments.
Experiment Design (see § IV-A2)	**Crawler Setup**	C5	Crawling tech used	OpenOBA framework was used in the experiment.
		C6	Adjustments to crawling tech	N/A
		C7	Extensions to crawling tech	N/A
		C8	Bot detection evasion	N/A
		C9	Publicly available crawler	OpenOBA framework is available in [62].
		C10	User interaction mimicry	User behavior was simulated by browsing training and control pages for 8 hours daily over 6 days.
	Experiment Env.	C11	Crawling strategy	A stateful crawling strategy was used.
		C12	Crawl location	Measurements were conducted in London, United Kingdom.
		C13	Browser adjustments	OpenWPM with Firefox version 108 was used.
		C14	Data processing pipeline	The data processing pipeline was detailed in the manuscript.
Evaluation		C15	Openly available results	All data and analysis scripts are available in [62].
		C16	Result/success overview	Results on OBA presence and cookie banner impact are presented in § IV-B.
		C17	Limitations	Limitations are discussed in § V-C.
		C18	Ethical considerations	Ethical considerations are discussed in § V-D.

tracking practices. The expected outcome is a comprehensive understanding of how geographic factors influence OBA.

By focusing on these experimental setups, future research can provide valuable insights into how interest, location differences, and legal frameworks influence online privacy and advertising strategies. This contributes to a better understanding and improvement of online privacy practices.

C. LIMITATIONS AND FUTURE WORK

Firstly, our framework incorporates ad scraping techniques inspired by prior work [52], [53], [54], leveraging EasyList selectors to detect visible ads. Despite its effectiveness, it can sometimes miss ads due to the evolving complexity of ad delivery mechanisms and the limitations in capturing dynamically served content. Hence, we intend to refine our scraping mechanisms to enhance accuracy and comprehensiveness in ad detection. Additionally, exploring alternative strategies for identifying ads in dynamically changing web environments could further bolster the robustness of our framework.

Secondly, our study has not addressed the potential impact of the browser being used by the user on OBA prevalence. Different browsers have varying levels of privacy protection, cookie handling, and built-in tracking prevention mechanisms, which can influence the extent of OBA. For instance, browsers like Firefox and Safari have introduced enhanced tracking protection, whereas Chrome handles cookies and tracking differently. While our study did not focus on browser-specific behavior, we intend to investigate how these differences across browsers affect OBA dynamics and tracking practices. Understanding this could provide further insights into how the choice of browser influences user privacy in the context of targeted advertising. In addition, we acknowledge that our current study is focused on analyzing OBA presence in relation to cookie consent choices, without comparing our results to a broader range of baselines such as privacy-enhancing tools (e.g., ad blockers, tracking prevention measures) or different browser-based tracking protections. While incorporating such comparisons would provide additional insights and add value to the analysis, they are beyond the scope of this study. In future work, we plan to address this limitation by expanding our experiments to include these broader baselines, allowing for a more comprehensive evaluation of OBA practices and the effectiveness of privacy protections across different scenarios.

Finally, we intend to integrate social media platforms into the *OpenOBA* framework to reveal OBA practices within these digital spaces. By examining targeted advertising behaviors and their link to user screen time on social media, researchers can uncover the mechanisms of online tracking and advertising on these platforms. While relevant research practices for social media are not yet part of this study, future work will focus on expanding the sample size, increasing the diversity of experiments, and exploring

a broader range of user behaviors across both traditional websites and social media platforms. This integration is set to enhance our understanding of how users' online behaviors are tracked and monetized, contributing to the wider conversation on online privacy and paving the way for more effective privacy-enhancing technologies and regulatory arrangements.

D. ETHICAL CONSIDERATIONS

This research is conducted following the ethical guidelines established by Partridge and Allman [55] and Kenneally and Dittrich [56], also adhering to the best measurement practices as recommended by Durumeric et al. [57]. Our methodology includes using *OpenOBA* to emulate a standard user navigating with a web browser.

VI. CONCLUSION

The study presented in this paper provided the design and implementation of a comprehensive framework, *OpenOBA*, for measuring OBA presence and its implications in the context of cookie banner interactions. Through our experiments, we demonstrated the framework's reliability by obtaining significant insights into how user consent choices — accepting, ignoring, or rejecting cookie banners — impact the prevalence and nature of targeted ads, all done in an automatic manner. Finally, to foster reproducibility and further research, we open-sourced our framework and released all data and analysis scripts [58].

ACKNOWLEDGMENT

The authors thank to Johan Mazel for his valuable comments at the start of this project.

REFERENCES

[1] (2023). *Most Popular Reasons for Using the Internet Worldwide As of 2nd Quarter 2023*. Accessed: Jan. 12, 2024. [Online]. Available: https://www.statista.com/statistics/1387375/internet-using-global-reasons/

[2] (2021). *Number of User Data Points Collected From Select Ios Social Media Apps Worldwide As of Mar. 2021, By Type*. Accessed: Nov. 1, 2024. [Online]. Available: https://www.statista.com/statistics/1305349/data-points-collected-apps-ios-by-type

[3] (2023). *Number of Internet and Social Media Users Worldwide As of Oct. 2023*. Accessed: Nov. 1, 2024. [Online]. Available: https://www.statista.com/statistics/617136/digital-population-worldwide/

[4] T. Bujlow, V. Carela-Español, J. Solé-Pareta, and P. Barlet-Ros, "A survey on web tracking: Mechanisms, implications, and defenses," *Proc. IEEE*, vol. 105, no. 8, pp. 1476–1510, Aug. 2017.

[5] (2024). *Online Advertising Market—Size, Share, Growth & Report*. Accessed: Mar. 1, 2024. [Online]. Available: https://www.mordorintelligence.com/industry-reports/online-advertising-market

[6] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris, "I always feel like somebody's watching me: Measuring online behavioural advertising," in *Proc. ACM CoNEXT*, 2015, pp. 1–13.

[7] T. Dehling, Y. Zhang, and A. Sunyaev, "Consumer perceptions of online behavioral advertising," in *Proc. IEEE CBI*, 2019, pp. 345–354.

[8] S. Aiolfi, S. Bellini, and D. Pellegrini, "Data-driven digital advertising: Benefits and risks of online behavioral advertising," *Int. J. Retail Distrib. Manage.*, vol. 49, no. 7, pp. 1089–1110, Jul. 2021.

[9] (2023). *View of Targeted Online Advertising Among Americans, By Age Group*. Accessed: Jan. 12, 2024. [Online]. Available: https://www.statista.com/chart/18146/view-of-targeted-online-advertising-among-americans/

[10] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang, "Smart, useful, scary, creepy: Perceptions of online behavioral advertising," in *Proc. SOUPS*, 2012, pp. 1–15.

[11] Y. Wu, S. Bice, W. K. Edwards, and S. Das, "The slow violence of surveillance capitalism: How online behavioral advertising harms people," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Jun. 2023, pp. 1826–1837.

[12] C. Matte, N. Bielova, and C. Santos, "Do cookie banners respect my choice: Measuring legal compliance of banners from IAB Europe's transparency and consent framework," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2020, pp. 791–809.

[13] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 1388–1401.

[14] A. Rasaii, S. Singh, D. Gosain, and O. Gasser, "Exploring the cookieverse: A multi-perspective analysis of web cookies," in *Proc. PAM*, 2023, pp. 623–651.

[15] D. Bollinger, K. Kubicek, C. Cotrini, and D. Basin, "Automating cookie consent and GDPR violation detection," in *Proc. 31st USENIX Secur. Symp.*, Aug. 2022, pp. 2893–2910.

[16] V. Dudykevych and V. Nechypor, "Detecting third-party user trackers with cookie files," in *Proc. 3rd Int. Sci.-Practical Conf. Problems Infocommunications Sci. Technol.*, Oct. 2016, pp. 78–80.

[17] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz, "The web never forgets: Persistent tracking mechanisms in the wild," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2014, pp. 674–689.

[18] I. Fouad, N. Bielova, A. Legout, and N. Sarafijanovic-Djukic, "Missed by filter lists: Detecting unknown third-party trackers with invisible pixels," in *Proc. PETS*, 2020, pp. 499–518.

[19] H. Dao, J. Mazel, and K. Fukuda, "CNAME cloaking-based tracking on the Web: Characterization, detection, and protection," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 3, pp. 3873–3888, Sep. 2021.

[20] H. Dao and K. Fukuda, "Alternative to third-party cookies: Investigating persistent PII leakage-based web tracking," in *Proc. 17th Int. Conf. Emerg. Netw. EXperiments Technol.*, Dec. 2021, pp. 223–229.

[21] P. Papadopoulos, N. Kourtellis, and E. Markatos, "Cookie synchronization: Everything you always wanted to know but were afraid to ask," in *Proc. World Wide Web Conf.*, May 2019, pp. 1432–1442.

[22] T. Urban, D. Tatang, M. Degeling, T. Holz, and N. Pohlmann, "Measuring the impact of the GDPR on data sharing in ad networks," in *Proc. 15th ACM Asia Conf. Comput. Commun. Secur.*, Oct. 2020, pp. 222–235.

[23] G. Acar, M. Juarez, N. Nikiforakis, C. Diaz, S. Gurses, F. Piessens, and B. Preneel, "FPDetective: Dusting the web for fingerprinters," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2013, pp. 1129–1140.

[24] M. Falahrastegar, H. Haddadi, S. Uhlig, and R. Mortier, "Tracking personal identifiers across the web," in *Proc. PAM*, 2016, pp. 30–41.

[25] *Chameleon*. Accessed: May, 3, 2024. [Online]. Available: https://github.com/ghostwords/chameleon

[26] *Common Crawl*. Accessed: May 3, 2024. [Online]. Available: https://commoncrawl.org/

[27] D. Cassel, "Omnicrawl: Comprehensive measurement of web tracking with real desktop and mobile browsers," in *Proc. PETS*, vol. 1, 2022, pp. 227–252.

[28] S. Siby, U. Iqbal, S. Englehardt, Z. Shafiq, and C. Troncoso, "WebGraph: Capturing advertising and tracking information flows for robust blocking," in *Proc. USENIX Secur.*, 2022, pp. 2875–2892.

[29] U. Iqbal, C. Wolfe, C. Nguyen, S. Englehardt, and Z. Shafiq, "Khaleesi: Breaker of advertising and tracking request chains," in *Proc. USENIX Secur.*, 2022, pp. 2911–2928.

[30] (2016). *European Commission: The General Data Protection Regulation (gdpr) in Eu*. Accessed: Jan. 19, 2024. [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/

[31] D. Klein, M. Musch, T. Barber, M. Kopmann, and M. Johns, "Accept all exploits: Exploring the security impact of cookie banners," in *Proc. 38th Annu. Comput. Secur. Appl. Conf.*, Dec. 2022, pp. 911–922.

[32] C. Santos, A. Rossi, L. Sanchez Chamorro, K. Bongard-Blanchy, and R. Abu-Salma, "Cookie banners, What's the purpose: Analyzing cookie banner text through a legal lens," in *Proc. 20th Workshop Workshop Privacy Electron. Soc.*, Nov. 2021, pp. 187–194.

[33] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz, "(Un)informed consent: Studying GDPR consent notices in the field," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2019, pp. 973–990.

[34] N. Jha, M. Trevisan, L. Vassio, and M. Mellia, "The Internet with privacy policies: Measuring the Web upon consent," *ACM Trans. Web*, vol. 16, no. 3, pp. 1–24, Aug. 2022.
[35] *Self-Regulatory Principles for Online Behavioral Advertising: Tracking, Targeting, and Technology*. Accessed: Mar. 5, 2024. [Online]. Available: https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-staff-report-self-regulatory-principles-online-behavioral-advertising/p085400behavadreport.pdf
[36] R. Balebako, P. Leon, R. Shay, B. Ur, Y. Wang, and L. Cranor, "Measuring the effectiveness of privacy tools for limiting behavioral advertising," in *Proc. W2SP'12-SP*, 2012, pp. 1–10.
[37] K. Solomos, P. Ilia, S. Ioannidis, and N. Kourtellis, "Talon: An automated framework for cross-device tracking detection," in *Proc. RAID*, 2019, pp. 227–241.
[38] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczynski, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2019, pp. 1–22.
[39] (2024). *Webshrinker*. Accessed: Dec. 12, 2024. [Online]. Available: https://webshrinker.com/
[40] (2024). *Labtaxonomy*. Accessed: Dec. 2, 2024. [Online]. Available: https://www.iab.com/guidelines/content-taxonomy/
[41] C. Yeung, U. Iqbal, Y. T. O'Neil, T. Kohno, and F. Roesner, "Online advertising in Ukraine and Russia during the 2022 Russian invasion," in *Proc. ACM WWW*, 2022, pp. 2787–2796.
[42] (2024). *Selenium*. Accessed: Dec. 2, 2024. [Online]. Available: https://www.selenium.dev/
[43] R. Kumar and A. Tomkins, "A characterization of online browsing behavior," in *Proc. 19th Int. Conf. World Wide Web*, Apr. 2010, pp. 561–570.
[44] M. Shi, C. Cardie, and S. Belongie, "Fashionpedia-ads: Do your favorite advertisements reveal your fashion taste?" 2023, *arXiv:2305.02360*.
[45] V. Mishra, P. Laperdrix, A. Vastel, W. Rudametkin, R. Rouvoy, and M. Lopatka, "Don't count me out: On the relevance of IP address in the tracking ecosystem," in *Proc. Web Conf.*, Apr. 2020, pp. 808–815.
[46] N. Demir, M. Große-Kampmann, T. Urban, C. Wressnegger, T. Holz, and N. Pohlmann, "Reproducibility and replicability of web measurement studies," in *Proc. ACM Web Conf.*, Apr. 2022, pp. 533–544.
[47] M. Gotze, S. Matic, C. Iordanou, G. Smaragdakis, and N. Laoutaris, "Measuring Web cookies in governmental websites," in *Proc. 14th ACM Web Sci. Conf.*, Jun. 2022, pp. 44–54.
[48] (2024). *Easylist*. Accessed: Dec. 2, 2024. [Online]. Available: https://easylist.to/easylist/easylist.txt
[49] (2006). *Easyprivacy*. Accessed: Mar. 5, 2024. [Online]. Available: https://easylist.to/easylist/easyprivacy.txt
[50] (2024). *Adservers List*. Accessed: May, 13, 2024. [Online]. Available: https://pgl.yoyo.org/adservers/
[51] *The Trade Desk*. Accessed: May 3, 2024. [Online]. Available: https://better.fyi/trackers/adsrvr.org/
[52] E. Zeng, M. Wei, T. Gregersen, T. Kohno, and F. Roesner, "Polls, clickbait, and commemorative 2 bills: Problematic political advertising on news and media websites around the 2020 U.S. elections," in *Proc. 21st ACM Internet Meas. Conf.*, Nov. 2021, pp. 507–525.
[53] E. Zeng, T. Kohno, and F. Roesner, "What makes a 'Bad' ad? User perceptions of problematic online advertising," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–24.
[54] E. Zeng, T. Kohno, and F. Roesner, "Bad news: Clickbait and deceptive ads on news and misinformation websites," in *Proc. ConPro*, 2020, pp. 1–11.
[55] C. Partridge and M. Allman, "Ethical considerations in network measurement papers," *Commun. ACM*, vol. 59, no. 10, pp. 58–64, Sep. 2016.
[56] E. Kenneally and D. Dittrich, "The menlo report: Ethical principles guiding information and communication technology research," *SSRN Electron. J.*, vol. 7, no. 3, pp. 1–22, 2012.
[57] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMAP: Fast Internet-wide scanning and its security applications," in *Proc. 22nd USENIX Conf. Security (SEC)*, 2013, pp. 605–620.
[58] *OpenOBA—Framework for Online Behavioral Advertising Measurement*. Accessed: May 3, 2024. [Online]. Available: https://github.com/fukuda-lab/OpenOBA

MATEO ORMENO is currently pursuing the Master of Science degree in engineering with Pontificia Universidad Católica de Chile, Santiago, Chile. He is doing a graduate research internship at the National Institute of Informatics (NII), Japan. His research interest includes web privacy measurements.

HA DAO received the Ph.D. degree from the National Institute of Informatics (NII), Graduate University for Advanced Studies (SOKENDAI), Japan. She is currently a Postdoctoral Researcher with the Max Planck Institute for Informatics, Germany. Her research interests include online privacy and data protection.

VALERIA HERSKOVIC received the Ph.D. degree from Universidad de Chile. She is currently an Associate Professor with the Department of Computer Science, Pontificia Universidad Católica de Chile. Her research interests include human-centered computing and human-centered artificial intelligence.

KENSUKE FUKUDA (Member, IEEE) received the Ph.D. degree in computer science from Keio University, Kanagawa, Japan, in 1999.

He is currently a Professor with the National Institute of Informatics (NII) and the Graduate University for Advanced Studies (SOKENDAI). His research interests include the measurement and analysis of internet traffic, network management, and security.