Understanding the Privacy Risks of Popular Search Engine Advertising Systems

Salim Chouaki, Oana Goga, Hamed Haddadi, Peter Snyder

Understanding the Privacy Risks of Popular Search Engine Advertising Systems

Salim Chouaki
LIX, CNRS, Inria, Ecole Polytechnique, Institut Polytechnique de Paris
salim.chouaki@inria.fr

Oana Goga
LIX, CNRS, Inria, Ecole Polytechnique, Institut Polytechnique de Paris
oana.goga@cnrs.fr

Hamed Haddadi
Imperial College London, Brave Software
h.haddadi@imperial.ac.uk

Peter Snyder
Brave Software
pes@brave.com

ABSTRACT

We present the first extensive measurement of the privacy properties of the advertising systems used by privacy-focused search engines. We propose an automated methodology to study the impact of clicking on search ads on three popular *private* search engines which have advertising-based business models: StartPage, Qwant, and DuckDuckGo, and we compare them to two dominant data-harvesting ones: Google and Bing. We investigate the possibility of third parties tracking users when clicking on ads by analyzing first-party storage, redirection domain paths, and requests sent before, when, and after the clicks.

Our results show that privacy-focused search engines fail to protect users' privacy when clicking ads. Users' requests are sent through redirectors on 4% of ad clicks on Bing, 86% of ad clicks on Qwant, and 100% of ad clicks on Google, DuckDuckGo, and StartPage. Even worse, advertising systems collude with advertisers across all search engines by passing unique IDs to advertisers in most ad clicks. These IDs allow redirectors to aggregate users' activity on ads' destination websites in addition to the activity they record when users are redirected through them. Overall, we observe that both privacy-focused and traditional search engines engage in privacy-harming behaviors allowing cross-site tracking, even in privacy-enhanced browsers.

CCS CONCEPTS

• **Security and privacy → Privacy protections**; **Privacy protections**; • **Networks → Network measurement**; **Network measurement**.

KEYWORDS

Search engines, advertising systems, cross-site tracking, privacy, measurement.

IMC '23, October 24–26, 2023, Montreal, QC, Canada

ACM ISBN 979-8-4007-0382-9/23/10…$15.00
https://doi.org/10.1145/3618257.3624823

ACM Reference Format:
Salim Chouaki, Oana Goga, Hamed Haddadi, and Peter Snyder. 2023. Understanding the Privacy Risks of Popular Search Engine Advertising Systems. In *Proceedings of the 2023 ACM Internet Measurement Conference (IMC '23), October 24–26, 2023, Montreal, QC, Canada.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3618257.3624823

1 INTRODUCTION

Privacy-focused search engines such as DuckDuckGo, StartPage, and Qwant [3, 9, 10] promote a strategy of respecting users' privacy and promise not to track users' search and browsing behavior, all while delivering relevant search results. However, private search engines rely on advertising for revenue, and use traditional advertising platforms to deliver ads: DuckDuckGo and Qwant use Microsoft's advertising system, while StartPage uses Google's advertising system. These search engines are often ambiguous on the privacy properties of the ads that appear on their search page, and their consequent privacy properties remain unexplored to the best of our knowledge.

In this work, we aim to fill this gap by conducting the first study of the privacy properties of the advertising systems of three major privacy-focused search engines: DuckDuckGo, StartPage, and Qwant, and how they compare to more popular search engines: Bing and Google. We investigate the privacy properties of these search engines when they: (i) present search ads to users, (ii) when a user clicks on an ad, and (iii) when the user lands on the advertiser's page.

We implement an automated measurement methodology to measure if and how users can be re-identified (hence, their privacy is compromised) when clicking on search ads on each search engine (see Section 3). We build an open-source implementation of this methodology in the form of a Puppeteer-based pipeline that simulates search queries and ad clicks. We apply this crawling methodology to the five search engines, providing a full dataset with visited websites, cookies created, locally stored values, and web requests to search engines' servers and/or other third parties when clicking ads. We use filter rules from several major open-source lists to detect web requests to online trackers, and we propose a methodology to differentiate user identifiers from non-tracking values in query parameters and cookies values.

We then present in Section 4 a systematic analysis of our dataset to investigate privacy harms before clicking an ad, during clicking an ad, and after clicking an ad and reaching the advertiser's website.

We find that users' privacy is not harmed *until* users click on an ad. Privacy-focused search engines do not appear to attempt to re-identify users across visits or queries and do not include resources from, or make network requests to known trackers. However, we find that users' privacy is compromised by **all** studied search engines in various ways once users click on an ad.

Disappointingly, we find that all search engines record additional information about the user and/or the users' clicks after the user has clicked on an ad. Private search engines capture data related to the clicked ad, including the ad provider, destination URL, and the ad's position within the search results page, along with the user's browsing data, such as the search query, device type, and browser language. Private search engines do not store user identifiers upon ad clicks, in contrast to traditional search engines that record user identifying values. Furthermore, we find that all search engines in our study engage in navigation-based tracking. Navigation-based tracking refers to tracking techniques that are redirecting users through one or more redirectors when navigating from one website to another in order to share user information across sites [33]. Navigation-based tracking does not require third-party cookies and can be used to circumvent browsers' privacy protections from cross-site tracking using partitioned cookies storage. Alarmingly, we observe that privacy-focused search engines engage in more navigation-based tracking than non-privacy-focused ones: We observe navigational tracking on 4% ad clicks on Bing, on 100% ad clicks on Google, on 100% ad clicks on DuckDuckGo, on 86% ad clicks on Qwant, and on 100% ad clicks on StartPage.

On the destination page, we check whether the search engine requires advertisers to abide by privacy-respecting practices by measuring whether advertisers include trackers or other known privacy-harming resources. We found that 93% of ads destination pages (across all five search engines) included tracker and privacy-harming resources. Finally, we check whether search engines or redirectors aid advertisers in profiling visitors by measuring the data they receive in the form of user-describing query params. We find that advertisers receive user identifiers as query parameters in 68%, 92%, and 53% of cases for DuckDuckGo, StartPage, and Qwant, respectively. This practice, known as UID smuggling, enables redirectors to aggregate more user behavior data if they have scripts on the ads' destination websites and they store the user-identifying parameters they receive. Notably, in the case of private search engines, the user-identifying parameters are not set by the search engine but by the redirectors encountered between the search engine's and the advertiser's sites.

Our results indicate that privacy-focused search engines' privacy protections do not sufficiently cover their advertising systems. Although these search engines refrain from identifying and tracking users and their ad clicks, the presence of ads from Google or Microsoft subjects users to the privacy-invasive practices performed by these two advertising platforms. When users click on ads on private search engines, they are often identified and tracked either by Google, Microsoft, or other third parties, through bounce tracking and UID smuggling techniques. Particularly, advertisers receive unique user identifiers through query parameters in most ad clicks, which can enable cross-site tracking even in privacy-enhanced browsers that block third-party cookie tracking.

2 BACKGROUND

This section briefly discusses the policies and approaches of the main search engines alongside popular tracking approaches.

2.1 Private search engines

We study the two dominant search engines that rely on user tracking for personalized search results and advertisements, namely Google and Bing, and three of the most popular privacy-branded search engines that provide users with non-personalized results and ads: DuckDuckGo, StartPage, and Qwant [11, 29]. Private search engines can either build their own independent search indexes or use big tech search engines like Bing, Google, or Yahoo to provide search results. Both types of private search engines claim not to store users' search histories and not to collect nor share tracking and personal data. We now describe the advertising systems employed by the different private search engines and present a summary of their data-sharing policies outlined in their respective *About* pages.

DuckDuckGo is a standalone search engine that maintains and uses its own search index alongside other indexes, such as Bing's, to provide search results [31]. DuckDuckGo relies on Microsoft's advertising system but only serves ads based on the search results and not the behavioral profiles of users [30]:

> *"search ads on DuckDuckGo are based on the search results page you're viewing instead of being based on you as a person"*

When clicking an ad on DuckDuckGo, the user is redirected to the ad's landing page through Microsoft Advertising's platform. DuckDuckGo claims Microsoft does not store ad-click behaviors from DuckDuckGo for purposes other than accounting and does not associate ad-clicks with users' profiles [18]:

> *"When you click on a Microsoft-provided ad that appears on DuckDuckGo, Microsoft Advertising does not associate your ad-click behavior with a user profile. It also does not store or share that information other than for accounting purposes."*

This implies that Microsoft can, though currently chooses not to, link the ad-click data to an existing Microsoft user profile. The privacy policy is signed by both DuckDuckGo and Microsoft.

Qwant is a standalone EU-based search engine that allows users to access online resources without being tracked nor profiled [32]. Qwant relies on Microsoft's advertising system to deliver ads in their search results pages. Although Qwant reports transmitting *some information* concerning search queries to Microsoft to enable the latter to present pertinent advertisements, it remains unclear which specific information is shared. In addition, to detect fraud, Qwant uses a specialized service offered by Microsoft, which has access to the user's IP address and the browser "User-Agent". Qwant assures that this service does not have access to the search query, which is sent to another service that does not know the IP address of the user [32].

Unlike DuckDuckGo, which also uses Microsoft advertising, we did not find any mention to ad-click behavior information on Qwant's privacy policy. They do not mention whether Microsoft stores this data and for what purposes they use it.

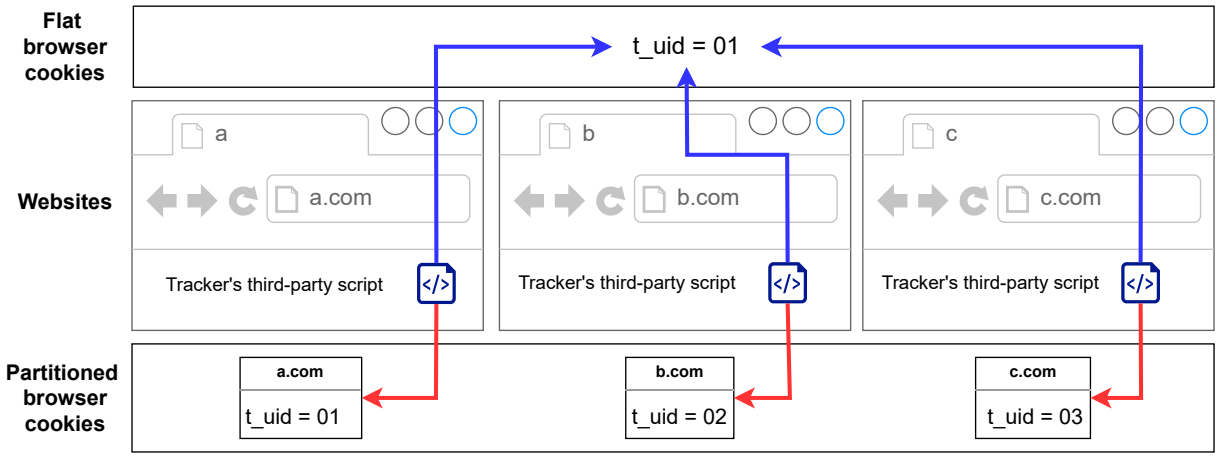

Figure 1: Cookie tracking in flat vs. partitioned cookies storage.

StartPage is a meta-search engine that allows users to obtain non-personalized search results from Google's search index while protecting their privacy. StartPage relies on Google AdSense to show ads to users on their results page. According to StartPage's privacy policy, the search engine serves strictly non-personalized ads since it does not share any identifiable information with Google. Therefore, ads displayed on the search results page are solely based on the user's search query [38].

Regarding ad-click behavior data, the privacy policy does not make any reference to whether Google tracks or profiles users based on this information. Nevertheless, StartPage emphasizes that by clicking on an ad, users leave the protection of StartPage's privacy policies and become subject to the data collection practices of the website they are redirected to [37].

> *"By clicking on an ad, like any other external website you click on after performing a StartPage search, you leave the privacy protection of StartPage and are subject to those websites' data collection policies."*

2.2 Cross-site tracking

Cross-site tracking refers to the practice of following a user across multiple first-party websites and associate their browsing activities to a unique identifier. Web tracking practices require first-party websites (e.g. the content providers) to share data about a user's activity with third parties (the trackers). Online tracking has been traditionally implemented through browser cookies. However, due to increasing adoption of cookie-blocking browsers and extensions, and the push on adopting partitioned cookies storage on web browsers, more and more trackers started to rely on navigational tracking techniques. We next discuss the details on how these two techniques work.

2.2.1 Cookie tracking. To enable cross-site cookie tracking, whenever a user visits a first-party website, the website makes a request to the third-party website (the tracker). This allows the tracker to set a cookie, which will identify the user and will be associated with the browsing activity of the user. For example, when the user visits a website *A* that makes a request to the tracker *T*, the tracker associates the cookie identifier of the user with the fact that the user visited website *A* (see Figure 1). Later, when the user visits website *B*, which also makes a request to the tracker *T*, the tracker will be able to associate the cookie identifier of the user with the fact that the user visited website *B*. Hence, the tracker will be able to know that the user visited both websites *A* and *B*.

This was initially possible because browsers had a common cookie storage containing all cookies, and trackers could read their corresponding cookies regardless of which first-party website allowed the tracker cookie to be set (see Figure 1). However, several browsers, such as Safari, Firefox, and Brave, have implemented partitioned storage to prevent using cookies for cross-site tracking [33]. These browsers use a partitioned cookies storage with a hierarchical namespace where a tracker accesses a different storage area on each website that loads it, preventing trackers from matching or assigning the same identifiers to users across multiple websites. Hence, cross-site tracking based on cookies can no longer be performed on these browsers. Chrome -the most used web browser- is in the process of testing partitioned cookies storage but does not use it by default [27, 28].

2.2.2 Navigational tracking. Navigational tracking refers to tracking techniques that use one or more URL navigations to share user information across sites. Navigational tracking does not require third-party cookies and can be used to circumvent browsers' privacy protections from cross-site tracking using partitioned cookies storage.

Bounce tracking is a navigational tracking technique that refers to redirecting users through one or more redirectors when navigating from one website to another. To allow this, a website *A* containing links to another website *B* does not directly link to the target *B* but instead links to an intermediary *redirector* (R)–the tracker (see Figure 2). When users click on a link on website *A*, they are taken to the redirector first, which then redirects them to the intended destination (website *B*) or other intermediary redirectors. The website *A* can directly change the actual link of the destination (b.com) to a redirection link (r.com), or a redirector's third-party

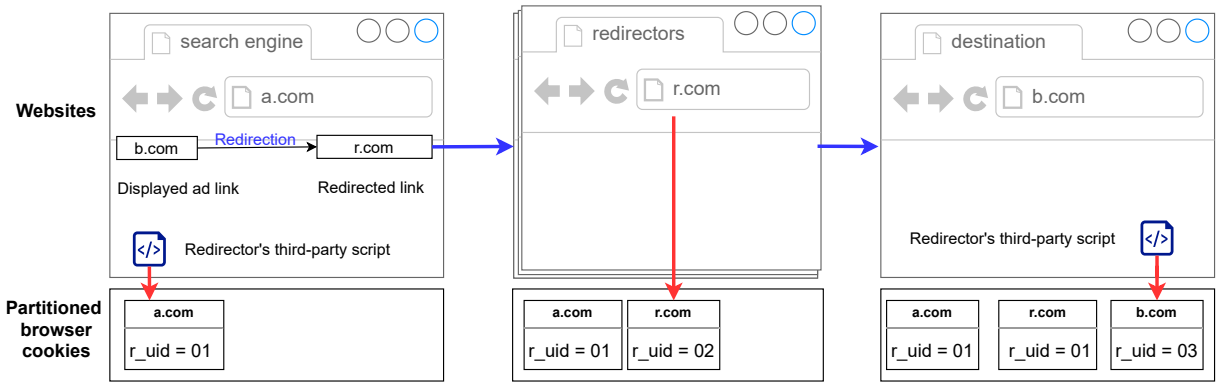

Figure 2: Bounce tracking.

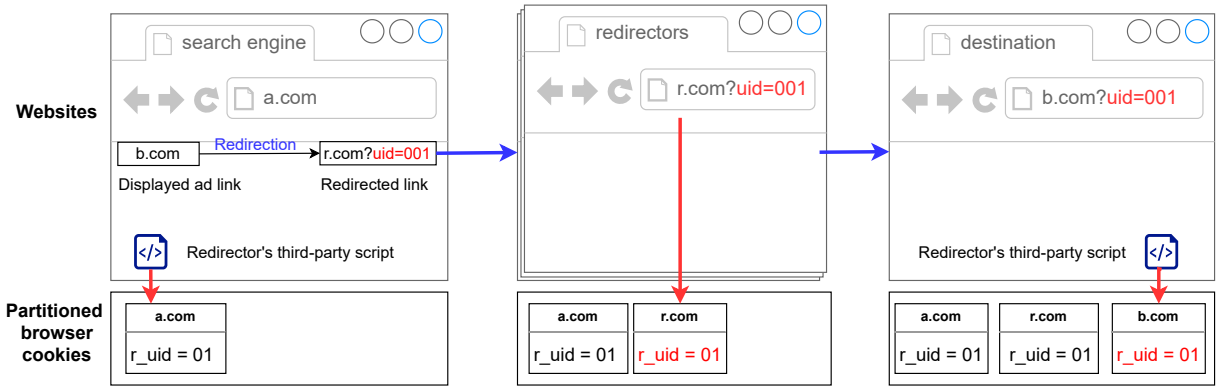

Figure 3: UID smuggling

script can do it. On its turn, the redirector can change the destination link again and send it further to other redirectors. Hence, from the link in the ad on the website *A*, one cannot know all the different redirectors the users will pass through when they click on an ad. We call the *redirection path* all the websites a user navigates through to arrive from *A* to *B*. Since, from a browser perspective, the redirector is the first-party domain, it can read or set cookies in its own partition [26]. In the following, we describe what data redirectors can infer according to the redirector's behavior.

(1) If the redirector does not set a first-party cookie, it will only know that a user went from website *A* to website *B* and will not be able to link this to other user browsing activities.

(2) If the redirector sets a first-party cookie, it will be able to aggregate all the activity of the user that is redirected through it (either from website A or other websites that use it as a redirector), hence, it will allow cross-site tracking.

(3) If the redirector also sets third-party cookies on websites *A* and *B*, it will not be able to link the activity of the user on website *A* with the activity of the user on website *B*, and with the activity of the user that goes through its own site (through redirects) since they do not share the same user ID [33]. Hence, while bounce tracking allows to a certain degree, cross-site tracking, it does not have the same coverage as the traditional (and soon-to-be obsolete) third-party cookie tracking.

UID smuggling is a navigational tracking technique that modifies users' navigation requests by adding information to the navigation URLs in the form of query parameters. In addition, similar to bounce tracking, UID smuggling may redirect the user to one or more third-party trackers before redirecting the user to the intended destination. Figure 3 describes this process. When a user clicks on a link on a website *A*, the originator page itself or a tracker on the page–through a script–decorates the URL by adding the originator's user identifier (UID) as a query parameter. The user then passes through zero or more redirectors which are invisible to him. Each of these redirectors can get the UID from the query parameter and has permission to store it in a first-party cookie under the redirector's domain. Finally, the user is sent to the destination website B, and the redirector can forward or not to website *B* the UID it received from *A*. All the trackers on website B will be able to read the UID from the query parameter and know that it was the UID sent by the originator (through request headers).

UID smuggling is more powerful than bounce tracking. Trackers using UID smuggling regain the ability to share UIDs across websites with different domains and can circumvent restrictions from partitioned cookie storage spaces [33]. For example, they can link the user's visits to the website *A* with the user's visits to website *B* and the user's activity that goes through its site (through redirects) since they can all be linked to the same user ID. In addition, UID

smuggling can help other trackers on website *B* (and website *A*) to link users' browsing activity across all the websites that received the UID as a query parameter.

Hence, bounce tracking and UID smuggling are two powerful techniques to circumvent browser's solutions for forbidding third-party cookie tracking.

3 MEASUREMENT METHODOLOGY

We develop a measurement methodology to capture network flows when clicking on an ad from a search engine results page. Using multiple crawlers, we simulate a large number of search engine queries in order to collect a sample of information flows per search engine. For each request, we collect the cookies created, the locally stored values, and the web request sent by the browser. In addition, we rely on several open-source datasets to detect web requests to online trackers. We consider five main search engines: Google[1], Bing[2], DuckDuckGo[3], StartPage[4], and Qwant[5]. We use Google and Bing as baselines to compare with the other three, which claim to have higher privacy standards and protective measures in place.

3.1 Crawling system

Each crawling iteration begins at a search engine's main page, where our system will type a query and access the search engine results page. Next, it chooses one of the displayed ads to click on to access its destination website. Then, the navigation path passes through zero or more redirectors before landing on the ad's destination website. The redirectors are invisible to the user but can be identified through an analysis of network requests initiated by the browser. Each of these redirectors can read the query parameters added by the search engine or other intermediaries and store them locally or send them to other third parties. The system records all first-party and third-party cookies, local storage values, and web requests at each step. We run each iteration in a new browser instance to ensure no stale data is cached from previous iterations.

Depending on the search engine, ads are either part of the main page or are loaded through an iframe. We use scrapping techniques to detect them and rely on several HTML elements' attributes. For instance, all ads on StartPage are inside an HTML element titled "Sponsored Links". Moreover, we use hyperlink values to detect Google ads since they all link to "www.googleadservices.com/*".

Our system prioritizes ads with landing domains it has not visited yet, aiming to maximize the number of different destination websites. Each time a crawler clicks on an ad, our system adds the domain of its landing URL to the list of visited websites. In the subsequent iterations, the crawler first extracts the landing domains of all the displayed ads. The landing domains are included within the HTML objects of the advertisements on all search engines. The crawler then gives preference to click on ads leading to domains that have not been encountered in the list of visited websites.

We reproduced these steps for 500 search queries on the five search engines. We randomly choose them from Google Trends [21] and movie titles from MovieLens [6]. All iterations were performed in "accept" cookies mode. Table 1 represents the number of different search queries we typed, the number of different destination pages we landed on, and the number of different domain paths we collected for each search engine.

Table 1: Number of search queries, destination websites, and redirection paths.

	# Queries	# Different destination websites	# Different redirection paths
Bing	500	98	131
Google	500	102	134
DuckDuckGo	500	56	94
StartPage	500	60	107
Qwant	500	60	88

We implemented our system using Puppeteer [7] to automate visiting search engines' websites, typing search queries, detecting and clicking on one of the displayed ads, and waiting for 15 seconds on the ad's destination website. We reproduce these steps multiple times from the same IP address for each search engine. To reduce the chance of being identified as bots, we use puppeteer-extra-plugin-stealth [8]. This plugin applies various techniques to make the detection of headless Puppeteer crawlers by websites harder.

Puppeteer allows us to record cookies and local storage for each request. However, it does not guarantee that it can attach request handlers to a web page before it sends any requests [33]. Hence, detecting and collecting web requests only using Puppeteer might cause losing some of them. We use a Chrome extension alongside Puppeteer crawlers to record web requests during all the crawling time. We do not observe a significant difference between web requests recorded by crawlers and web requests recorded by the extension. In median, the crawlers recorded 97% of the requests recorded by the extension. The code of the crawling system and the dataset are available at https://github.com/CHOUAKIsalim/Search_Engines_Privacy.

3.2 Detection techniques

Detection of trackers: We use URL filtering to detect web requests to online trackers. We use filter rules from two open-source lists: EasyList [4] and EasyPrivacy [5]. EasyList is the most popular list to detect and remove adverts from webpages and forms the basis of many combination and supplementary filter lists [19]. EasyPrivacy is a supplementary filter list that detects and removes all forms of tracking from the internet, including tracking scripts and information collectors [19]. These filter lists are used by extensions that aim to remove unwanted content from the internet, like AdBlock and uBlock. We combined and parsed these lists using adblock-rs [1] and obtained 86 488 filtering rules.

In addition, we use the Disconnect Entity List [2] to get the entities of online tracker domains. It is a dictionary where keys represent entities such as Google, Microsoft, and Facebook, and values represent the web domains that belong to each entity. Hence, to get the entity of a tracker, we iterate over all values and search to what entity is the tracker domain associated with. This list contains 1 449 entities and 3 371 related web domains.

[1] https://www.google.com/
[2] https://www.bing.com/
[3] https://duckduckgo.com/
[4] https://www.startpage.com
[5] https://www.qwant.com/

Detection of bounce tracking: We classify an instance as bounce tracking when an advertisement's destination link is altered to pass through one or more redirectors. To construct the redirection sequence, we trace the series of URLs the browser navigates through after clicking an ad and prior to reaching the advertisement's intended landing page. We further validate the redirection sequence by examining the HTTP response headers, precisely the 'Location' and 'status code' headers. These headers divulge the redirection process, as the 'Location' header contains the new redirection URL, and status codes such as '301 Moved Permanently,' '302 Found,' '307 Temporary Redirect,' and '308 Permanent Redirect' indicate the occurrence of redirection [17].

Detection of UID smuggling and user identifiers: To detect UID smuggling, we need to differentiate between query parameters that represent user identifiers and non-tracking query parameters such as session identifiers, dates, and timestamps. We consider all query parameters, localStorage, and cookie values. We call them tokens. There are 6 971 unique tokens in our dataset. We perform the following filtering, which is similar to the one performed by Randall et al. [33]:

(i) Each iteration is executed in a new browser instance; hence, user identifiers should not be shared across browser instances. We discard tokens that are the same across all or a subset of browser instances.
(ii) For each browser instance and search query, we analyze the tokens resulting from the URLs of all ads that appear on the results page (which are usually in the form of googleadservices.com/.... /aclk?..cid=CAESbeD2ZWCwqFv3e-2k_....). We discard tokens with different values for the different ad URLs as they likely correspond to ad identifiers.
(iii) To detect session identifiers, we store the profile of each iteration in a separate directory and execute an extra iteration per browser instance one day later to see which values of cookies/parameters change. We discard tokens with different values in the two iterations as they are more likely session identifiers.
(iv) Similar to [33], we use programmatic heuristics to discard particular values. We discard tokens that appear to be timestamps (values between June and December 2022 in seconds and milliseconds), tokens that appear to be URLs, tokens that constitute one or more English words ([20]), and tokens that are seven characters long or less.

After using these filters, we are left with 1 942 tokens. We manually investigated them and observed a non-negligible number of false positives. Hence, we manually filtered the remaining tokens and removed those composed of any combination of natural language words, coordinates, or acronyms. In the end, we are left with 1 258 user-identifying tokens, which we consider to be user identifiers.

4 RESULTS

This section presents the results of applying the presented methodology to the five selected search engines. We measure how users' privacy is affected before, during, and after clicking on a search ad. We find that the advertising systems on all evaluated search engines result in privacy harm, even for search engines that market themselves as privacy-respecting. We find that how, and to what degree, user privacy is harmed varies across each evaluated system.

The rest of this section proceeds as follows. Section 4.1 begins by presenting measurements of how user privacy is impacted *before* users click on an ad (i.e., after the user has received answers to their search query, but before the user clicks on an advertisement contained among or alongside the search results). Section 4.2 presents measurements of how user privacy is effected *during* clicking on an advertisement (i.e., after the user has clicked on an advertisement, but before the user arrives at the advertisement's destination). Finally, Section 4.3 gives measurements of how user privacy is affected *after* clicking on an advertisement (i.e., after the user has arrived at the final destination of the advertisement link, and scripts are executed on the advertiser's website).

4.1 Before clicking on an ad

We first present measurements of how the advertising systems used by popular search engines affect user privacy before a person has clicked on any advertisement. At this point in the process, the user has submitted a query to the search engine and received a results page. The returned results include at least two types of links: "organic results" (i.e., websites that contain content the search engine thinks relates to the query) and "paid results" (i.e., advertisements that the search engine has been paid to show to users).

This subsection presents measurements of how user privacy is impacted before the user has selected, clicked on, or otherwise "engaged" with a search advertisement. Since a user will only click on a fraction of the advertisements they are presented with, users will be effected by these "before" privacy harms more frequently than the privacy harms presented in later subsections.

4.1.1 First-party reidentification. We first measure whether search engines track or reidentify users across queries and visits. We find that the non-privacy-focused search engines (i.e., Bing and Google) track users across visits and are able to link different search queries to the same user who made those queries. The privacy-focused search engines, on the other hand, do not appear to attempt to reidentify users across visits or queries, aligning with the claims made in their privacy policies (see Section 2.1). We measured whether search engines are able to reidentify users across queries and visits by looking for whether search engines stored unique user identifiers in the browser's first-party storage (e.g., cookies, localStorage). Specifically, we inspected the DOM storage area for each site and looked for stored values that appeared to be unique identifiers, using the heuristics described in Section 3.2. We observed that Google and Bing did store such user identifiers; the other search engines did not.

We note that some privacy-focused search engines *did* store other values in first-party storage, but that they were used for purposes other than user identification (e.g., client-side storage of user preferences).

4.1.2 Requests to trackers. We also measured whether search engines harmed user privacy by communicating with trackers when presenting advertisements. We did not observe any search engine including resources from, or making network requests to, known trackers.

We checked for communication with known trackers by i. recording the URLs of all the network requests made by the browser when rendering the search results, and ii. checking those URLs against popular filter lists (as described in Section 3.2). These URLs comprise both the sub-resources (e.g., scripts, images, videos) loaded by the results page and the third-party requests made using the Web networking APIs (e.g., XMLHttpRequest, `fetch()`, web sockets).

We note that we were only able to measure the client-side network behavior of each search engine, and could only observe whether the search engine pages themselves were sharing information with known trackers. We were not able to measure how or if each search engine communicates with trackers on the server-side.

4.2 When clicking on an ad

Next, we measure how user privacy is affected after the user clicks on an ad, but before the user has arrived at the ad's destination (usually, a page controlled by the party placing the advertisement). This step of the process involves systems run by both the search engine itself and the advertising platform paying for the ad.

During this stage, the advertising system may try and accomplish several goals, including fraud detection (i.e., attempting to detect if the "click" was the result of an automated system, intending to increase how much the advertiser pays the search engine) and user profiling (i.e., recording information about the user clicking the ad to combine with existing user profiles). Simultaneously, the search engine may use this step in the processs to try and achieve other goals, including quality of service measurements (i.e., ensuring that advertisements render correctly) or additional user profiling (i.e., recording which ad the user clicked to "enrich" whatever information the search engine may have about the user).

We find that the measured search engines vary widely in how they treat user privacy when the user clicks on an ad. However, we also find that the advertising systems engage in privacy-harming behaviors and share user identifying information with third parties across all measured search engines, despite the privacy-focused branding adopted by some search engines.

4.2.1 Search engine page behaviors. First, we measured what behaviors the search engine's page engages in *after* the user clicks on an ad but *before* the browser begins navigating away from the search engine's page (and towards the advertisement's destination page). These behaviors might be things like recording which advertisement the user clicked on or how long the user waited before clicking, and are implemented with browser APIs like "onclick" handlers and "ping" attributes [16].

We measured each search engine's post-click behaviors by recording what network requests happened on the page after each advertisement was clicked on. We find that all search engines record additional information about the user and/or the user's click, after the user has clicked on an ad.

Bing. Clicking on an advertisement on Bing results in additional first-party (i.e., within Bing) network requests. In all iterations, clicking caused a request to be sent to https://bing.com/fd/ls/GLinkPingPost.aspx. These requests included several query parameters, including the clicked ads' destination websites. Furthermore, these requests include user identifiers, for instance, communicated in the `MUID` cookie –A cookie identifying unique web browsers visiting Microsoft sites- [6].

Google. Clicking on ads on Google results in additional first-party web requests. In all cases, the browser sends POST web requests to https://google.com/gen_20?. These requests include user identifier values communicated in cookies such as `NID` and `AEC` [7].

DuckDuckGo. Clicking on an advertisement on DuckDuckGo results in additional first-party network connections to https://improving.duckduckgo.com. These requests include several query parameters, such as the search query, the ad provider (Bing in all cases), and the destination URL of the clicked ad. Next, the browser sends an additional network request that fetches a JavaScript file served from https://duckduckgo.com/y.js. This request includes several query parameters containing information about the ad and the link to which the user should be redirected (link to Bing servers). We note that none of the query parameters nor the cookies sent with these web requests matched our user heuristics for user identifiers.

Qwant. When clicking on an advertisement on Qwant, a first request is sent to https://qwant.com/action/click_serp, including information about the user's browser, such as the type of the device and the browser language, along with the search query. Furthermore, this request contains information on the clicked ad (e.g., its position on the results page and the destination website). Then, another request is sent to https://api.qwant.com/v3/redirect/, including the URL to direct the user to. These two connections do not include user identifiers as query parameters nor as cookies values.

StartPage. Clicking on an advertisement on StartPage results in an additional first-party request to https://startpage.com/sp/cl. This request includes information about the position of the clicked ad on the results page, but does not include the ad's destination URL. Similar to DuckDuckGo and Qwant, requests to StartPage servers do not include user identifiers.

In summary, we find that all search engines, traditional and privacy-focused alike, record information about users' ad clicks. They all collect data about the clicked ad, such as its position on the results page or destination URL. However, only traditional search engines (Google and Bing) include user identifiers with web requests to their servers.

4.2.2 Navigation Tracking. Next, we measure whether the advertising systems in search engines engage in navigation-based tracking, a technique for tracking users that circumvents browser privacy protections by directing a user through otherwise unrelated sites. Section 2.2.2 provides a high-level summary of how navigating tracking works and why it is an effective method of circumventing tracking protections in many browsers. We find that most of the search engines in our data set engage in navigation-based tracking at least some of the time. Further, we find that the *privacy-focused search engines engage in navigation-based tracking for the majority of placed ads.*

We measure the navigation tracking we observed on the selected search engines in three dimensions: i. the distribution of how many sites the user is "bounced" through when they click on an ad on

[6] https://learn.microsoft.com/en-us/clarity/cookie-list
[7] https://policies.google.com/technologies/cookies

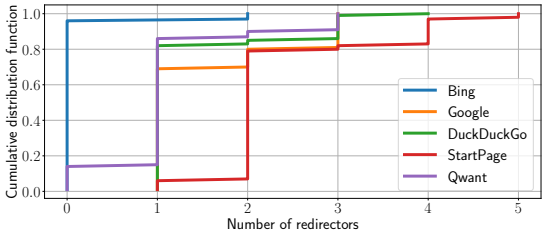

Figure 4: CDF of the number of different redirectors for Bing, DuckDuckGo, Google, and StartPage.

Figure 5: CDF of the number of different redirectors that store UID cookies for Bing, DuckDuckGo, Google, and StartPage.

each search engine, ii. how many different organizations a user is exposed to during navigation tracking episodes (distinct from the number of pages or domains), and iii. the distribution of the number of sites in the redirection path that store user-identifying cookies.

Number of sites visited. Figure 4 presents the distribution of the number of different sites (i.e., $eTLD + 1$) each search engine directs the user through when clicking on an ad. We observe that clicking on an ad on Bing generally results in being redirected through the fewest number of sites (96% of ad clicks on Bing result in no other site being visited except for Bing and the final destination site). Clicking on sites on DuckDuckGo, Google, and Qwant typically results in visiting one other site (respectively, 82%, 69%, and 72% of clicks result in an intermediate navigation to a site different than the search engine and the ad's destination). Clicking on ads on StartPage resulted in (on average) visiting the largest number of different sites (93% of clicks resulted in visiting at least two sites other than StartPage and the ad's destination).

Number of organizations visited. However, we note that all redirections are not equal in their privacy impact; the marginal privacy harm is generally much lower if a site redirects the user between two sites the company owns, versus the user being redirected between two sites owned by unrelated companies. More concretely, there is little-to-no additional privacy harm if Google bounces a user—and passes information about the user—from google.com to googleadservices.com, while there *is* privacy harm if Google bounces a user–and the user's information—from google.com to facebook.com (i.e., Facebook learns new information they otherwise would not learn).

Understanding the privacy harm of navigation tracking requires considering *which* sites the user is being "bounced" between. Table 2 presents the five most common redirection paths for each search engine, and Table 7 in the appendix presents the most common sites in the redirection paths. Moreover, we group redirectors' domains by the organization to which they belong using the Disconnect Entity List [2]. Table 3 presents the fraction of navigation paths that include a website from each organization across all search engines.

We observe that the impact of navigation tracking differs widely between search engines. On one hand, the navigation tracking that occurs from clicking on ads on Google results in little additional privacy harm; the most commonly immediately visited sites are also operated by Google (i.e., googleadservices.com and ad.doubleclick.com). On the other hand, we find that navigation tracking significantly harms user privacy on privacy-branded search engines. In all three cases, users are either usually directed to Bing sites (100% and 76% of the time for DuckDuckGo and Qwant, respectively) or Google sites (100% of the time for StartPage).

While these results are alarming—since these are search engines advertising that they are privacy-preserving —they are not inexplicable. DuckDuckGo and Qwant rely on Bing to provide search ads, and StartPage relies on Google.

Number of sites that identify users. The extent of privacy harm resulting from bounce tracking depends on two key factors: the behavior of the redirector (i.e., whether the redirector stores user-identifying cookies) and the type of cookie storage used by the browser (flat or partitioned). The lowest level of privacy harm occurs when the redirector does not store any user-identifying cookies. In this case, the redirector can infer the source and destination of the navigation event (i.e., the search engine and the ad's website). However, if the user navigates through the same redirector multiple times, the redirector cannot aggregate the tracking data from different visits to the same user.

In contrast, if the redirector sets UID cookies on users' browsers, it can combine tracking data each time the user bounces through it. Specifically, if a user clicks on multiple ads on the same search engine and is redirected through the same redirector each time, the redirector can aggregate all the websites the user has visited. Moreover, if the user's browser has flat cookie storage, the redirector can potentially aggregate the user's activity and match it to the same user instance on every website where the redirector has a script.

Figure 5 presents the distribution of the number of different redirectors in the navigation paths that store UID cookies for each search engine, and the Table 4 lists these redirectors that store UID cookies on users' browsers. Our observations indicate that the level of privacy harm resulting from bounce tracking varies considerably across different search engines. While the navigation tracking that occurs when users click on ads on traditional search engines appears to cause little privacy harm, as users are identified by sites operated by third-party entities in only 4% and 8% of navigation paths for Bing and Google, respectively. In contrast, for the three privacy-branded search engines, users are identified by sites operated by third-party entities in most cases. Precisely, more than 95% of users clicking on ads on DuckDuckGo, StartPage, or Qwant are identified by Bing, and Google identifies users clicking on ads on StartPage in 100% of cases. As a result, Google and Bing might associate the

Table 2: Top five most common navigation domain paths when clicking an ad for each search engine.

Search engine	Domain paths	Frequency
Bing	bing.com - destination	96%
	bing.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	3%
	bing.com - t23.intelliad.de - 1045.netrk.net - destination	1%
Google	google.com - googleadservices.com - destination	69%
	google.com - googleadservices.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	17%
	google.com - googleadservices.com - pixel.everesttech.net - ad.doubleclick.net - destination	4%
	google.com - googleadservices.com - monitor.clickcease.com - destination	4%
	google.com - googleadservices.com - monitor.ppcprotect.com - destination	2%
DuckDuckGo	duckduckgo.com - bing.com - destination	82%
	duckduckgo.com - bing.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	14%
	duckduckgo.com - bing.com - 6102.xg4ken.com - destination	2%
	duckduckgo.com - bing.com - clickserve.dartsearch.net - ad.doubleclick.net - tpt.mediaplex.com - destination	1%
	duckduckgo.com - bing.com - pixel.everesttech.net - destination	1%
StartPage	startpage.com - google.com - googleadservices.com - destination	73%
	startpage.com - google.com - googleadservices.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	17%
	startpage.com - google.com - destination	6%
	startpage.com - google.com - googleadservices.com - 6008.xg4ken.com - destination	1%
	startpage.com - google.com - googleadservices.com - clickserve.dartsearch.net - ad.doubleclick.net - monitor.ppcprotect.com - destination	1%
Qwant	qwant.com - bing.com - destination	66%
	qwant.com - destination	14%
	qwant.com - bing.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	10%
	qwant.com - track.effiliation.com - destination	3%
	qwant.com - click.linksynergy.com - destination	3%

Table 3: Fraction of navigation paths that include a website from each organization across all search engines.

	Bing	Google	DuckDuckGo	StartPage	Qwant
Adobe	0%	**4%**	**1%**	0%	**1%**
Conversant Media	0%	0%	**1%**	0%	0%
DuckDuck Go	0%	0%	100%	0%	0%
Facebook	0%	0%	0%	**1%**	0%
Google	**3%**	100%	**15%**	100%	**11%**
Kenshoo	0%	**2%**	**2%**	**1%**	0%
Microsoft	100%	0%	100%	0%	79%
Nielsen	0%	0%	0%	**1%**	0%
PPCProtect	0%	**2%**	0%	**1%**	**1%**
Qwant	0%	0%	0%	0%	100%
Rakuten	0%	0%	0%	0%	**3%**
StartPage	0%	0%	0%	100%	0%
Unknown	**4%**	**23%**	**15%**	**19%**	**16%**

destination website visited by the user through the advertisement to the user profile, especially if the user's browser has flat-cookie storage.

4.3 After clicking on an ad

Finally, we measure how user privacy is impacted once the user has "finished" clicking on a search ad and has arrived at the advertiser's page. We measure how the search engine/advertiser relationship effects user privacy in two ways: first, by measuring whether advertisers include trackers or other known-privacy-harming resources, and two, by measuring if and what kinds of information the search engine's advertising system provides to the advertiser (in the form of user-describing query params). This first measure relates to whether the search engine requires advertisers to abide by privacy-respecting practices; the latter measure relates to whether search engines' advertising systems collude with advertisers to aid advertisers in profiling visitors.

Redirectors in navigation paths can aggregate more data about the user's behavior if they have scripts on the ads' destination websites. For this, they need to match users using either third-party cookies if they are enabled by the browser or UID smuggling. In this section, We investigate whether redirectors can aggregate users' activity on ads destination websites by analyzing online trackers, whether they receive UID as query parameters, and whether they store them. We recorded these requests by keeping the crawlers on the ads' destination pages for 15 seconds for all iterations.

4.3.1 Requests to online trackers. We first measure whether search engines protect their users by requiring advertisers to be privacy-protecting. We measure this by loading the website each clicked

Table 4: Redirectors that store UID cookies.

Bing	Google	DuckDuckGo	StartPage	Qwant
ad.doubleclick.net (3%)	googleadservices.com (98%)	bing.com (95%)	google.com (100%)	bing.com (78%)
t23.intelliad.de (1%)	ad.doubleclick.net (21%)	ad.doubleclick.net (14%)	googleadservices.com (94%)	ad.doubleclick.net (11%)
1045.netrk.net (1%)	pixel.everesttech.net (4%)	6102.xg4ken.com (2%)	ad.doubleclick.net (18%)	click.linksynergy.com (3%)
	monitor.ppcprotect.com (2%)	pixel.everesttech.net (1%)	6008.xg4ken.com (1%)	pixel.everesttech.net (1%)
	3825.xg4ken.com (2%)			monitor.ppcprotect.com (1%)
				tracking.deepsearch.adlucent.com (1%)

search advertisement leads to, recording the URLs of all sub-resources and network requests made when loading and executing the page, and comparing those URLs against EasyList and EasyPrivacy.

We find that the web pages users are taken to when they click on ads search engines contain many trackers and privacy-harming resources. Further, we observe this for both "standard" and "privacy-focused" search engines alike. Specifically, we find that 93% of ads destination pages (across all five search engines) include tracker and privacy-harming resources. Broken down by search engine, we observed 277, 218, 326, 437, and 260 different tracker third parties over all iterations, and a median of 9, 11, 6, 8, and 6 different online trackers per iteration for Bing, Google, DuckDuckGo, StartPage, and Qwant, respectively.

In order to understand which companies track users on ad destination pages, we group the domains that observed tracking resources are served from by "entity" using the Disconnect Entity List [2] For example, using the entity list, we group tracker resources served from the domains google.com and doubleclick.com to the same entity (i.e., Google). Table 5 presents the top entities of trackers we observed on ad destination pages. For instance, we see that Google is the top entity for online trackers on destination pages for StartPage (36%), and we saw that all StartPage redirection paths go through Google servers. Hence, if the browser implements a flat cookies storage, Google can match the StartPage user on the ads destination website and aggregate data about his activity on it in 36% of the cases. We make the same observation for Microsoft trackers on Qwant (4.3%).

4.3.2 User identifiers. Finally, we measure if the advertising systems of the search engines aid advertisers in tracking users across sites by transmitting unique identifiers (or other personal or otherwise individual values) across site boundaries through query parameters.

As discussed in Section 3.2, this technique is sometimes called UID smuggling and is an increasingly common technique trackers and sites use to circumvent browser privacy protections (such as blocking third-party cookies or partitioning browser storage). For example, if an advertiser places an ad for https://site.example, the advertising system might collude with the advertiser to allow the advertiser to profile the user by appending unique identifiers to the destination URL. The search engine's advertising system might, for example, append information the advertising system knows about the user to the advertiser's destination URL (creating a URL like https://site.example?user_id=<id>, so that the advertiser can learn more about the user, harming the user's privacy.

We measure whether search engines' advertising systems collude with advertisers to track users across sites by examining the query parameters the search engine (or other intermediate party in a navigation chain) includes in the URL of the advertiser's destination page. We collect all of the query parameters in the destination ad URLs and extracted values that appeared to be unique identifiers using the heuristics described in Section 3.2.

We find that advertising systems collude with advertisers most of the time across all search engines, *even private ones*. Clicking ads on all five search engines resulted in user identifiers being passed to advertisers. We found user identifiers in query parameters in 80%, 94%, 68%, 92%, and 53% for Bing, Google, DuckDuckGo, StartPage, and Qwant, respectively. Most of these parameters are MSCLKID (Microsoft Click Identifier) or GCLID (Google Click Identifier), two *unique identifiers* used for ad-click tracking. MSCLKID is added to the landing page URL by Microsoft Advertising and GCLID is added by Google Ads when users click on their respective ads. Advertisers use these click IDs to identify and track ad clicks; advertisers might store click-tracking first-party cookies to track actions taken after the ad click [14, 24, 25]. Table 6 represents the fraction of iteration where the web request to the ad's destination page included MSCLKID, GCLID, or other parameters. We can see that in search engines that use Microsoft advertising (DuckDuckGo and Bing), we find both MSCLKID and GCLID. However, in ones that use Google advertising (Google, StartPage, and Qwant), we do not find MSCLKID.

Moreover, we investigate whether ads destination pages persist the UID query parameters they receive. We cross-reference values obtained from destination pages' first-party storage (e.g., cookies and localStorage) with the query parameters these pages receive. We find that MSCLKID values are persisted in 15%, 17%, and 1% of cases for Bing, DuckDuckGo, and Qwant, respectively. As for GCLID, we find that a cookie is created in 5%, 10%, and 13% of cases for Bing, Google, and StartPage.

Table 5: Top entities of online trackers reached by crawlers on each search engine.

Bing	Google	DuckDuckGo	StartPage	Qwant
unknown (32.0%)	unknown (34.8%)	unknown (29.5%)	Google (36.0%)	Google (26.3%)
Google (24.4%)	Google (28.7%)	Google (21.8%)	unknown (28.1%)	Amazon (23.4%)
Microsoft (13.8%)	Microsoft (10.5%)	Amazon (16.3%)	Microsoft (4.3%)	unknown (22.4%)
Facebook (3.8%)	Amazon (3.1%)	Facebook (3.4%)	Facebook (3.2%)	Microsoft (4.2%)
Criteo (2.4%)	Criteo (2.5%)	Criteo (2.2%)	Criteo (3.0%)	Criteo (3.8%)

Table 6: Fraction of iteration where the ad's destination page received MSCLKID, GCLID and other UID attributes as query parameters.

	MSCLKID	GCLID	other UID parameters
Bing	79%	12%	3%
Google	0%	92%	8%
DuckDuckGo	66%	12%	6%
StartPage	0%	92%	12%
Qwant	51%	8%	7%

5 LIMITATIONS

Our measurement methodology has some limitations. First, we only look for user identifiers transferred in query parameters and do not detect them when they are transferred in other methods. For instance, previous work [33, 39] found that trackers sometimes decorate their own URL in the document.referrer header with user identifiers and reads them on the destination page. Second, we run all our crawling iterations from the same IP address. Consequently, if some query parameters are user IP address based, they will have the same value across all iterations, and thus we would not consider them as user identifiers. Finally, our results, particularly those observed after clicking on ads, are subject to variation based on the ads we selected and the search queries we used. Different search queries could potentially trigger distinct ads and lead to diverse advertisers, potentially exhibiting different behaviors. Nonetheless, our primary objective is to demonstrate the potential for third-party tracking when interacting with ads on private search engines.

6 RELATED WORK

Search engines and online tracking have received a lot of attention from the research community in the past decades. We review next only studies closest to our work.

Search engines. A first line of work has measured to which extent we can observe personalization in search engine results [23, 34] and ads [22]. For instance, Hannak et al. [23] have developed a methodology for measuring personalization in search results, applied it to Bing, Google, and DuckDuckGo, and found that Bing results are more personalized than Google ones while they did not find any noticeable personalization for DuckDuckGo.

A second line of work has focused on solutions to protect users' privacy from search engines and prevent web profiling. Castellà-Roca et al. [12] presented a computationally efficient protocol that provides a distorted user profile to the search engine to preserve users' privacy.

Finally, several studies have proposed privacy-preserving search-personalizing solutions for search engines. For instance, Shen et al. [36] analyze various software architectures for personalized search and envision possible strategies with a client-sided personalization. Xu et al. [40] suggest helping users choose the content and degree of detail of the profile information built by search engines.

To the best of our knowledge, there is no study investigating the privacy properties of the advertising systems used on private search engines.

Online tracking. Several works analyzed the usage of cross-site tracking techniques in the wild [15]. Chen et al. [13] propose a data flow tracking system to measure user tracking performed through first-party cookies that third-party JavaScript sets. They found that more than 97% of the websites they have crawled have first-party cookies set by third-party javascript and that on 57% of them, there is at least one cookie containing a unique user identifier diffused to multiple third parties. Roesner et al. [35] measured how user tracking occurs in the wild. They found that multiple parties track most commercial pages and estimate that several trackers can each capture more than 20% of a user's browsing behavior.

In response to browsers implementing partitioned cookies storage to protect users from cross-site tracking through browser cookies, several trackers are adopting other tracking strategies such as bounce tracking or UID smuggling. Koop et al. [26] analyzed a dataset of redirection chains in the wild and found that 11% of websites redirect to the same 100 top redirectors. Moreover, they demonstrate that these top redirectors could identify users on the most visited websites. Randall et al. [33] measured the frequency of UID smuggling in the wild and found that it is performed on more than 8% of all navigations in their dataset. We use a similar method to identify user identifiers among all cookie values and query parameters by implementing automatic filtering followed by a manual inspection.

All these studies were conducted in the wild, and to the best of our knowledge, no study focuses on navigational tracking techniques performed on search engines.

7 CONCLUSION

In this paper, we presented the first systematic study of the privacy properties of the advertising systems of five popular search engines: Two traditional ones, Google and Bing, and three private ones, DuckDuckGo, StartPage, and Qwant. We investigated whether, and to which extent, search engines through their advertising systems, engage in privacy-harming behaviors that allow cross-site tracking.

Despite the privacy intentions and promises of private search engines, our findings reveal the failure of privacy-focused search

engines to fully protect users' privacy during ad interactions. Users on all measured search engines, including the privacy-focused ones, are subject to navigation-based tracking by third parties. We find that all search engines engage in bounce tracking when clicking on ads, where users are sent through several redirectors before reaching the ads' destination websites. Surprisingly, our results indicate that privacy-focused search engines engage in more bounce tracking than non-privacy-focused ones. While private search engines themselves do not engage in user tracking, their reliance on traditional advertising systems (Microsoft or Google) renders users susceptible to tracking by those systems. *Although we cannot directly attribute this tracking to the search engines themselves, it is evident that they are enabling it through their reliance on Microsoft and Google's advertising systems.*

Inspecting the privacy policies of the search engines in light of our findings reveals interesting disparities. While our results demonstrate that Microsoft is capable of tracking DuckDuckGo users when they click on ads, DuckDuckGo asserts that Microsoft does not associate ad-click data with user profiles. On the other hand, Qwant, which also relies on Microsoft advertising for a significant fraction of its ads, do not document the utilization of ad-click data by Microsoft and whether it is used to enhance user profiles. Similarly, StartPage explicitly states that clicking on ads subjects users to the data collection policies of other websites.

Our study highlights the need for increased attention to privacy protection within the advertising systems of search engines. One potential solution to protect users' privacy for private search engines would be to reduce their reliance on third-party advertising systems. Developing their own advertising platform could provide greater control over privacy practices, although the feasibility and complexity of such an approach remain uncertain. Alternatively, private search engines could collaborate with advertising systems such as Microsoft and Google, forging partnerships that proactively tackle privacy concerns. For instance, private search engines could negotiate agreements with the ad provider that prevent redirecting users who click on ads placed within private search engines to additional third parties. This approach would minimize the extent of third-party tracking, limiting it to the ad provider only. Moreover, search engines like StartPage and Qwant could follow the lead of DuckDuckGo by seeking agreements with advertising systems to prevent the use of ad-click identifiers for user profile enrichment. These proactive steps would enhance user privacy while maintaining advertising partnerships with larger platforms.

ACKNOWLEDGMENTS

This research was supported in part by the French National Research Agency (ANR) through the ANR-17-CE23-0014, ANR-21-CE23-0031-02, and MIAI@Grenoble Alpes ANR-19-P3IA-0003 grants and by the EU through the 101041223, 101021377, and 952215 grants.

REFERENCES

[1] Last accessed September 11, 2023. Ad Block engine in Rust. https://www.npmjs.com/package/adblock-rs
[2] Last accessed September 11, 2023. Disconnect Entity List. https://github.com/mozilla-services/shavar-prod-lists/blob/master/disconnect-entitylist.json
[3] Last accessed September 11, 2023. DuckDuckGo seach engine. https://duckduckgo.com/
[4] Last accessed September 11, 2023. EasyList. https://easylist.to/easylist/easylist.txt
[5] Last accessed September 11, 2023. EasyPrivacy. https://easylist.to/easylist/easyprivacy.txt
[6] Last accessed September 11, 2023. MovieLens. https://movielens.org/
[7] Last accessed September 11, 2023. Puppeteer. https://www.npmjs.com/package/puppeteer
[8] Last accessed September 11, 2023. Puppeteer Extra Plugin Stealth. https://www.npmjs.com/package/puppeteer-extra-plugin-stealth
[9] Last accessed September 11, 2023. Qwant seach engine. https://www.qwant.com/
[10] Last accessed September 11, 2023. StartPage seach engine. https://www.startpage.com/
[11] Brave. Last accessed September 11, 2023. What are the best private search engines? https://brave.com/learn/no-tracking-search-engine/
[12] Jordi Castellà-Roca, Alexandre Viejo, and Jordi Herrera-Joancomartí. 2009. Preserving user's privacy in web search engines. *Computer Communications* 32, 13 (2009), 1541–1551. https://doi.org/10.1016/j.comcom.2009.05.009
[13] Quan Chen, Panagiotis Ilia, Michalis Polychronakis, and Alexandros Kapravelos. 2021. Cookie Swap Party: Abusing First-Party Cookies for Web Tracking. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 2117–2129. https://doi.org/10.1145/3442381.3449837
[14] Google Click Identifier (GCLID): Definition. Last accessed September 11, 2023. Google Click Identifier (GCLID): Definition. https://support.google.com/google-ads/answer/9744275
[15] Nurullah Demir, Daniel Theis, Tobias Urban, and Norbert Pohlmann. 2022. Towards Understanding First-Party Cookie Tracking in the Field.
[16] MDN Web Docs. Last accessed September 11, 2023. The Anchor element - ping attribute. https://developer.mozilla.org/en-US/docs/Web/HTML/Element/a#ping
[17] MDN Web Docs. Last accessed September 11, 2023. Redirections in HTTP. https://developer.mozilla.org/en-US/docs/Web/HTTP/Redirections
[18] DuckDuckGo and Microsoft. Last accessed September 11, 2023. DuckDuckGo Help Pages - Company Ads by Microsoft on DuckDuckGo Private Search. https://help.duckduckgo.com/duckduckgo-help-pages/company/ads-by-microsoft-on-duckduckgo-private-search/
[19] EasyList. Last accessed September 11, 2023. Overview. https://easylist.to/
[20] Github. Last accessed September 11, 2023. PyEnchant. https://pyenchant.github.io/pyenchant/
[21] Google. Last accessed: September 11, 2023. Stats and Analysis. https://trends.google.com/trends
[22] Saikat Guha, Bin Cheng, and Paul Francis. 2010. Challenges in Measuring Online Advertising Systems. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (Melbourne, Australia) *(IMC '10)*. Association for Computing Machinery, New York, NY, USA, 81–87. https://doi.org/10.1145/1879141.1879152
[23] Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring Personalization of Web Search. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 527–538. https://doi.org/10.1145/2488388.2488435
[24] Google Analytics Help. Last accessed September 11, 2023. Common questions about Google Ads Clicks and Analytics Sessions. https://support.google.com/analytics/answer/4588454?hl=en
[25] Microsoft Help. Last accessed September 11, 2023. Auto-tagging of Microsoft Click ID. https://help.ads.microsoft.com/apex/index/3/en/60000
[26] Martin Koop, Erik Tews, and Stefan Katzenbeisser. 2020. In-Depth Evaluation of Redirect Tracking and Link Usage. *Proceedings on Privacy Enhancing Technologies* 2020 (10 2020), 394–413. https://doi.org/10.2478/popets-2020-0079
[27] Milica Mihajlija. Last accessed September 11, 2023. Cookies Having Independent Partitioned State (CHIPS). https://developer.chrome.com/docs/privacy-sandbox/chips/
[28] Milica Mihajlija. Last accessed September 11, 2023. Cookies Having Independent Partitioned State (CHIPS) origin trial. https://developer.chrome.com/blog/chips-origin-trial/
[29] NordVPN. Last accessed September 11, 2023. The best private search engines for secure browsing. https://nordvpn.com/blog/private-search-engines/
[30] DuckDuckGo Help Pages. Last accessed September 11, 2023. Company - Advertising and Affiliates. https://help.duckduckgo.com/duckduckgo-help-pages/company/advertising-and-affiliates/
[31] DuckDuckGo Help Pages. Last accessed September 11, 2023. Privacy - Anonymous Localized Results. https://help.duckduckgo.com/privacy/anonymous-localized-results/
[32] Qwant. Last accessed September 11, 2023. Legal information. https://about.qwant.com/en/legal/confidentialite
[33] Audrey Randall, Peter Snyder, Alisha Ukani, Alex C Snoeren, Geoffrey M Voelker, Stefan Savage, and Aaron Schulman. 2022. Measuring UID smuggling in the wild. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 230–243.
[34] Shamma Rashed, Tasnim Said, Amal Abdulrahman, Arsiema Yohannes, and Monther Aldwairi. 2022. Evaluating Web Search Engines Results for Personalization and User Tracking. (2022). https://doi.org/10.48550/ARXIV.2211.11518

[35] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and Defending against Third-Party Tracking on the Web. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation* (San Jose, CA) *(NSDI'12)*. USENIX Association, USA, 12.

[36] Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2007. Privacy Protection in Personalized Search. *SIGIR Forum* 41, 1 (jun 2007), 4–17. https://doi.org/10.1145/1273221.1273222

[37] StartPage. Last accessed September 11, 2023. Can I advertise on StartPage? https://support.startpage.com/hc/en-us/articles/5076181310612-Can-I-advertise-on-Startpage-

[38] StartPage. Last accessed September 11, 2023. Privacy Policy. https://www.startpage.com/en/privacy-policy

[39] WebKit. Last accessed September 11, 2023. Tracking Prevention Policy. https://webkit.org/tracking-prevention-policy/

[40] Yabo Xu, Ke Wang, Benyu Zhang, and Zheng Chen. 2007. Privacy-Enhancing Personalized Web Search *(WWW '07)*. Association for Computing Machinery, New York, NY, USA, 591–600. https://doi.org/10.1145/1242572.1242652

8 APPENDIX

Ethics

Our experiments were conducted in a completely automated manner, without any human involvement or use of user data. Furthermore, the measurements we performed imposed minimal overhead on the well-resourced ad networks

Table 7: Most common redirectors (and their fractions) in domain navigation paths when clicking an ad on search engines.

Bing	Google	DuckDuckGo	StartPage	Qwant
clickserve.dartsearch.net (38%)	googleadservices.com (65%)	bing.com (74%)	google.com (42%)	bing.com (71%)
ad.doubleclick.net (37%)	ad.doubleclick.net (14%)	clickserve.dartsearch.net (11%)	googleadservices.com (39%)	ad.doubleclick.net (10%)
t23.intelliad.de (13%)	clickserve.dartsearch.net (13%)	ad.doubleclick.net (11%)	clickserve.dartsearch.net (7%)	clickserve.dartsearch.net (9%)
1045.netrk.net (12%)	pixel.everesttech.net (3%)	6102.xg4ken.com (2%)	ad.doubleclick.net (7%)	track.effiliation.com (3%)
	monitor.clickcease.com (3%)	tpt.mediaplex.com (1%)	6008.xg4ken.com (1%)	click.linksynergy.com (2%)
	monitor.ppcprotect.com (1%)	pixel.everesttech.net (1%)	monitor.ppcprotect.com (1%)	pixel.everesttech.net (1%)
	3825.xg4ken.com (1%)		t.myvisualiq.net (1%)	awin1.com (1%)
			monitor.clickcease.com (1%)	zenaps.com (1%)
			ad.atdmt.com (1%)	deepsearch.adlucent.com (1%)
				monitor.ppcprotect.com (1%)