Toward (Greater) Consumer Surveillance in a 'Cookie-less' World:
A Comparative Analysis of Current and Future Web Tracking Mechanisms

Ido Sivan-Sevilla & Patrick T. Parham (UMD)

ABSTRACT

The anticipated shift of the advertising industry away from third-party cookies has been marketed as 'privacy friendly.' New cookie-less tracking technologies are being proposed, but the consumer privacy implications of those technologies are far from clear. To what extent ad-networks are going to change their practices, and no longer rely on cross-site consumer surveillance or historically rich consumer profiles for advertising purposes? Can the new tracking technologies become GDPR-compliant, given the significant compliance pushback against cookie-based advertising mechanisms?

Our study seeks to evaluate the potential privacy harms of cookie-less advertising ID solutions by (1) building novel typology of tracking specifications to assess the privacy impacts of tracking technologies; (2) deductively analyzing cookie-based tracking mechanisms by collecting novel data on persistent user identification across 50 popular websites that work with all four main supply side platforms (SSPs) - Pubmatic, OpenX, AppNexus, and Rubicon; and (3) contrasting those findings with deductive analysis of data collected from technical industry documentation on the three main cookie-less ID architectures - The Trade Desk Unified ID 2.0, LiveRamp ID, and Secure Web Addressability Network (SWAN).

We find how the new tracking architectures can make consumer surveillance in the Web dynamically wider, more persistent over time, and extra vulnerable to integration of first- and third-party data by advertisers. This could lead to the circumvention of consumer targeting restrictions posed by the main advertising platforms, in case advertisers choose to bid on ads based on sensitive profile categories. In contrast to existing criticism on cookie-less tracking solutions that mostly focuses on deficiencies in consent mechanisms or the lack of a governing body for the new solutions, our study underscores the structural impact of ad networks on the potential privacy harms caused by the proposed tracking mechanisms. Structures and processes of ad networks might lead to potentially longer, wider, and richer consumer surveillance, compared to cookie-based tracking mechanisms. Our findings also question the ability of suggested tracking architectures to become GDPR compliant. Despite civil society pushback against the current cookie-based tracking ecosystem and the marketing of new tracking solutions as 'privacy-preserving-advertisement,' we show how these technologies enable granular tracking and targeting of consumers. 'Singling out' individuals might become easier for advertisers, given the historically rich consumer profiles that advertisers are now incentivized to build. In contrast to existing debates about Universal ID solutions and GDPR compliance, that mostly focus on the lack of data controller and data processor roles in the proposed architectures, we are stressing the extent of profiling and targeting that the new solutions allow and the way they violate key GDPR principles such as article 9 and recital 26.

Our empirical findings not only challenge the assumption that excluding third-party cookies would decrease consumer surveillance, but also show the potential for consumer surveillance to expand, seriously questioning the ability of those solutions to comply with the GDPR. A shift in one tracking instrument, as central as that instrument might be, cannot bridge the inherent gaps in consumer privacy caused by the structures & incentives of the online advertising market.

Toward (Greater) Consumer Surveillance in a 'Cookie-less' World:
A Comparative Analysis of Current and Future Web Tracking Mechanisms

Ido Sivan-Sevilla & Patrick T. Parham (UMD)

1 - Introduction

Ad-based monetization mechanisms of Web content are about to move away from their main tracking instrument – third-party cookies (Binns, 2022; Choi et al., 2020; O'Reilly, 2020). Arguably a major win for privacy advocates, this would theoretically disable cross-site surveillance, enhancing consumers' privacy online. New 'cookie-less ID Solutions' are extensively discussed by ad industry stakeholders, who frame them as 'Privacy Preserving Advertisement' (PPA) initiatives (Thomson & Rescorla, 2021). The privacy and compliance implications of those prospective ID solutions, however, are still unclear. Are these new 'Universal ID Solutions' really 'privacy-friendly' as marketed by the industry? To what extent Ad Networks are going to change their practices and no longer rely on cross-site consumer surveillance for marketing purposes? Does it really prevent individuals from being 'singled-out' by future online ad solutions as clearly stressed by the GDPR?

To better understand the future of Web tracking and privacy compliance, we compare the three main cookie-less tracking architectures - The Trade Desk Unified ID 2.0, LiveRamp RampID, and Secure Web Addressability Network (SWAN) - to current cookie-based tracking mechanisms in 50 popular websites working with all four main supply-side-platform (SSP) networks - Pubmatic, OpenX, AppNexus, and Rubicon. Our comparison is based on deductive tracking analyses according to our suggested typology of tracking specifications for assessing the privacy implications of tracking technologies.

We first introduce a novel typology of tracking specifications, based on five different categories, to assess the privacy harms associated with tracking technologies. Through this typology we aim to systematically measure and reveal to what extent cookie-less tracking mechanisms differ from one another and from current cookie-based tracking with regards to privacy harms. Second, to capture current cookie-based tracking practices, we collect data from 50 popular websites that work with the four main SSPs in the industry. We investigate the extent to which those actors persistently identify users across sites based on their unique ID cookie. Those SSPs are in a position to link user identities across browsing experiences and conduct cross-site consumer surveillance. We scrape the ads.txt files in the root domains of top million popular websites (based on the tranco rank) to construct publisher networks grouped around the same SSP. To reduce false positives, we inspect the HTMLs of 50 popular websites in each SSP network to verify that those websites practically work with that SSP. Then, we trace the usage of ID cookies within and across the four SSP networks. We investigate the construction of user identities by SSPs and trace persistent identification of individuals across websites. Through various web crawlers, we collect data about tracking cookies installed via HTTP Headers, URL Parameters, and browsers' local storages. We couple that with technical documentation from the ad industry to further

understand the privacy implications of cookie-based tracking. Third, we contrast our analysis of cookie-based tracking with the persistent user identification enabled by suggested cookie-less solutions. We rely on journalistic reports and technical industry documents, considering the three most dominant cookie-less tracking solutions discussed by the industry, and deductively evaluating them based on our tracking typology.

Our findings reveal how the new tracking architectures can make consumer surveillance in the Web (1) Dynamically wider, linking user identities across, not only within SSP networks; (2) More persistent over time by relying on deterministic identification data such as PIIs; and (3) Vulnerable to the integration of first- and third-party data for consumer profiling by Incentivizing advertisers to link their user data with purchased data. This could allow advertisers to secretly bid on sensitive user categories, potentially circumventing targeting restrictions of the main advertising platforms.

In contrast to existing criticism on cookie-less tracking solutions that mostly focus on deficiencies in consent mechanisms or the lack of a governing body, our study underscores the structural impact of ad networks on the potential privacy harms by the proposed tracking mechanisms. Structures and processes of ad networks lead to potentially longer, wider, and richer consumer surveillance, compared to cookie-based tracking. Our findings also question the ability of suggested tracking architectures to become GDPR compliant. Despite civil society pushback against the current cookie-based tracking ecosystem (Lomas, 2021) and the marketing of new tracking solutions as 'privacy-preserving-advertisement' (Thomson & Rescorla, 2021), we show how these technologies enable granular tracking and targeting of consumers. 'Singling out' individuals might become easier for advertisers, given the historically rich consumer profiles that advertisers are incentivized to build. In contrast to existing debates about Universal ID solutions and GDPR compliance, that mostly focus on the lack of data controller and data processor roles in the proposed architectures (Schiff, 2021e; Schiff; 2022a), or the role of cookies in the suggested solutions (Asim, 2021; Asim, 2022), we are stressing the extent of profiling and targeting that the new solutions allow and the way they violate key GDPR principles such as article 9 and recital 26.

Those findings highlight how online tracking has become a deeply entrenched norm, with a shift in one tracking instrument – third-party cookies – unable to fundamentally change the behavior and user-profiling appetite of ad industry actors. Websites will still heavily rely on third-party request networks for ad delivery (Gopal et al., 2022), and the structure of ad networks is expected to stay the same, with all third-parties involved having an interest in building their own database about individuals to augment their ad bidding and delivery decisions (Martin, 2022). We ultimately argue that progress toward limiting persistent consumer identification in the digital advertising industry has not been made. Cross-site surveillance is here to stay, with the structure of ad networks and the extent of persistent user identification determining the levels of users' online privacy and the ability of the ad industry to meaningfully comply with the law.

To evaluate the privacy implications of future tracking technologies, the next section highlights existing gaps in analyzing future tracking solutions and the importance of considering the privacy impacts of the structural components and processes of ad networks. Then, we develop five tracking specifications for evaluating the privacy harms of tracking technologies. In Section 3.1

we present our data collection & analysis for current cookie-based tracking. In Section 3.2, we analyze the privacy impacts of cookie-less tracking solutions. Section 3.3 summarizes the comparison of cookie-based and cookie-less tracking technologies. Section 4 discusses the implications of our findings and Section 5 concludes, detailing study limitations and future questions.

2 - Ad Networks (lack of) Privacy Compliance

Consumer privacy is gaining momentum. Revelations on how platforms use consumers' data (Mac & Kang, 2021), the unprecedented wave of federal and state privacy bills (Lively, 2022), and the on-going questioning on GDPR compliance by the advertising industry (Lomas, 2021) created pressure on browser owners to declare their end of support in the muchly criticized third-party cookies (Shein, 2021). The anticipated shift in the main tracking instrument for targeted Ads, a market projected to grow to $525B by 2024 (Edelman, 2020), led to a burst of alternative user ID solutions (Asim, 2021), marketed by the industry as 'privacy friendly.'

Existing criticism on cookie-less tracking solutions mostly focus on deficiencies in consent mechanisms (Kaye, 2021a) or the lack of a governing body (UnifiedID2, 2022). A recent report from Mozilla surfaces more specific privacy concerns, identifying how new tracking solutions provide no mechanism to prevent access to users' data (Thomson & Rescorla, 2021). Still, discussion of these cookie-less tracking proposals and their privacy implications have not placed enough emphasis on the intermediary role that advertising technology networks will play in continuing to stitch together persistent consumer identification (Asim, 2021; Asim, 2022). In contrast, we argue that the structural impact of ad networks on the potential privacy harms of new tracking technologies should be carefully analyzed.

Discussion over the ability of suggested tracking technologies to become GDPR compliant is also limited, mostly focusing on the lack of data controller and data processor roles in the proposed architectures (Schiff, 2021e; Schiff; 2022a), or the role of cookies in the suggested solutions (Asim, 2021; Asim, 2022). But at the center of those solutions, we argue, are the profiling and targeting capabilities enabled for ad-network actors. Those capabilities should be assessed in light of GDPR's principles regarding data subjects' control over their data, the processing of sensitive data, and the ability of data controllers to 'single out' individuals.

Thus, the structural components and processes of ad network actors should be at the center of consumer tracking and targeting evaluations. Surprisingly, those ad-networks, which facilitate almost half of Web monetization mechanisms (Choi et al., 2020), have received very little empirical attention from the Information Systems (IS) literature. Most works on privacy, including works related to online advertising, are very specific to users' privacy concerns and consumer choices (Aguirre et al., 2015; Dinev, 2014; Dinev et al., 2013; Dinev & Hart, 2006; Goldfarb & Tucker, 2011, 2015; Hui et al., 2007; T. Li & Unger, 2012; Y. Li, 2011, 2012; Malhotra et al., 2004; Xu et al., 2011). Less is known about the privacy threats that consumers experience online and how they challenge industry compliance. Simply put, the 'so what' story about Web privacy remains empirically under-discussed by the IS literature and this study aims to fill some of the gap.

Specifically, the shift in the use of third-party cookies as an evasive tracking instrument by ad-based Web monetization mechanisms (Jones, 2020; Lavin, 2006) was not critically analyzed. We aim to shed light on potential user tracking and targeting in a cookie-less world, questioning the ability of post-cookie solutions to comply with privacy law. We intend to highlight the operations of Ad Networks, which represent a chain of third-party actors that monitor consumers' behavior online and follow consumers across multiple websites for marketing purposes (D'Annunzio & Russo, 2020). While publishers collect information about their own visitors, it is the ad network that collects the most information to track and target consumers across the Web (Bashir et al., 2016). Those Ad Networks affect market outcomes, but research remains sparse regarding the implications of network actors' decisions (Choi et al., 2020).

To fill those gaps and understand the structural impacts of ad networks on consumers' privacy, we deductively analyze cookie-based and cookie-less tracking mechanisms based on a novel tracking typology below, that enables us to understand the privacy impacts of tracking technologies.

3 - Tracing Web Tracking

Following Binns (2022), we use a working definition of tracking provided by the World Wide Web Consortium (W3C)'s tracking protection group according to which 'tracking is the collection of data regarding a particular user's activity across multiple distinct contexts, and the retention, use, or sharing of data derived from that activity outside the context in which it occurred' (W3C Working Group, 2019). According to this definition, not every data collection is considered 'tracking.' As long as data collected within one context stays in the same spatial and temporal context, we should not consider such data collection as 'tracking.' However, when a party somehow collates different data points about the individual from different data sources and/or time stamps, this should count as tracking (Binns, 2022).

The specifications of tracking vary considerably. Tracking can take place via various user identification instruments - cookies (Jones, 2020), fingerprinting (Englehardt & Narayanan, 2016), favicons (Solomos et al., 2021), personally identifiable information (PII), and etc. Tracking can be conducted by different parties (public or private) who can gain visibility on data subjects across different contexts, in different time stamps, and for different purposes - marketing, law enforcement, national security, user engagement, public health, etc. There are ample opportunities and vectors for consumers to be tracked, especially when tracking actors own various services with which consumers directly engage - from search engines and platforms to mobile software, hardware, websites, and for our purposes in this paper, ad network functions.

We focus specifically on tracking as an opaque commercial practice for online consumers on the Web, emerged through symbiotic relationships between websites and third parties (Gopal et al., 2022), and facilitated through automated data capture, making 'passive' commercial surveillance almost inevitable for individuals on the Web.

To evaluate and assess the extent of tracking, we develop a framework for tracking specifications that details what we view as the most important criteria for measuring the impact of tracking technologies on consumers' privacy. We suggest the following five tracking specifications:

[1] ***User Identification Instrument*:** How can individuals be identified? What is the tracking instrument through which trackers assign user IDs and potentially follow users and collate data points for profiling purposes (i.e. cookies, tokens, fingerprints, favicons).

[2] ***Cross-context User Visibility:*** Which companies can persistently identify consumers across sites? We are interested to understand who are the ad network actors that can identify and 'enjoy' visibility over consumers across the Web (i.e., SSPs, DSPs, advertisers, publishers, Ad Exchanges).

[3] ***Longitudinal Tracking:*** Does the tracking technology enable consumers to be tracked over time?

[4] ***Circumvention of Targeting Restrictions:*** Does the tracking mechanism enable/incentivize advertisers to build rich first-party consumer profiles and hiddenly escape targeting restrictions by advertising platforms?

[5] ***User Data Sources:*** Can we limit participating data actors for profiling purposes? What are the possible sources for data collation? Which data points about the user can be gathered and used for targeting purposes? (i.e. current browsing behavior, past browsing behavior, offline data from digital footprints, first party data, third party data).

Through a deductive analysis based on the tracking specifications above, sections 3.1 & 3.2 evaluate cookie-based and cookie-less tracking mechanisms. Section 3.3 compares the mechanisms, revealing how cookie-less tracking alternatives are likely to enable greater consumer surveillance than current, cookie-based, tracking practices.

3.1. Cookie-based Tracking

In order to contrast future cookie-less tracking mechanisms with current cookie-based tracking by ad networks, we first aim to reach a comprehensive understanding of current tracking practices. To do so, we rely on browser-side observations and study the ability of ad networks to persistently identify and potentially track users across websites.

To assess the usage of persistent identifiers within and across SSP networks, we assembled a list of publisher sites that work with all four main SSP networks in the advertising industry. The four SSPs selected were based on the authors' familiarity with the networks' prominent position in the digital advertising industry. We developed two different selection criteria to assemble a final list of publisher sites. For the first selection criteria, inclusion of a publisher was based on the publisher site listing the SSP in their 'ads.txt' file. With this criterion met, we then ranked sites based on the 'Tranco' popularity index (Tranco, n.d.). Our initial assembled list of the top 100 popular websites that list all four main SSPs in their ads.txt file was crawled by visiting their landing pages (and not inner-site pages) per site. Analyzing the initial results, we saw that

certain publishers did not register cookies from all four SSPs. While ads.txt indicate which partners are eligible to sell a publishers ad inventory, based on the instances observed in crawling 100 sites we assume that just because a partner is listed does not mean that the publisher is currently working with the partner. In the second selection criteria, inclusion of a publisher was based on manually checking in the Chrome browser Developer Tools - Cookies Storage table, that a cookie from each of 4 SSPs registered in at least 1 of 10 refreshes of the landing page. We were able to develop a list of 50 sites from checking sites in the 'Tranco' popularity index that list all four SSPs in their ads.txt file (see appendix #1). From this manual check that indicated that a cookie does not always register when a landing page is loaded, we decided to crawl the 50 sites 10 separate times and average the results. In the 10 separate crawls, we again visited only the publisher landing pages.

To evaluate the usage of persistent user identifiers by trackers among the four SSP networks we traced stateful tracking via HTTP cookies, as they are still the most dominant technique to identify online users across websites (Roesner et al., 2012; Fouad et al., 2020). To collect tracking information (i.e HTTP cookies, JavaScript Operations, and HTTP headers) of each publisher site we used an open source-based automated web crawler – OpenWPM - that simulates real users' activity and records website responses, metadata, cookies used, and scripts executed (Englehardt and Narayanan, 2016). We performed 10 individual stateful crawls and set the crawl to use only one browser instance. We did not set the "Do Not Track" setting in the configuration to allow for persistent identification and configured the simulated browser to accept all 3rd-party cookies. We also used the "bot detection mitigation" to scroll randomly up and down visited pages. We set sleep time between publisher sites to five seconds, and timeout between websites to 100 seconds. The 10 crawls were run on a local machine on September 20th & 21st, 2022, and data were recorded in a SQLite database.

Within the SQLite database file created by the crawl, the 'http_requests', 'javascript', and 'javascript_cookies' tables were analyzed. SSPs were recognized based on the 'host' field and filtered based on rows containing the SSP network name. Publisher site was established based on the 'top_level_url' field. The persistent identifier was recognized based on a concatenation of the cookie name and cookie value fields. For the crawl of publisher sites, stateful tracking was enabled and we traced the cookie storage of our browser to analyze how the same cookie IDs were used across websites within and across SSP networks. We refer to the SSPs that use the same cookie ID value in two or more publisher sites as 'persistent identifiers,' as they are identically identifying the user across the sites they work with. We identified SSP ID cookies based on the syntax included in each company's privacy policy (See Table 1 below) (Magnite, 2021a; OpenX, 2022; Pubmatic, 2020; Xandr, 2022).

SSP Network	ID Cookie Name
Pubmatic	**KADUSERCOOKIE**
OpenX	**i**
AppNexus	**uuid2**
Rubicon	**khaos**

Table 1: ID Cookie Names Used for Tracing Persistent Identification of Users Per SSP Network

In our data analysis, we wanted to first observe whether certain SSPs persistently identify users across publisher sites within their networks. We observed that SSPs utilized a persistent identifier in 77.6-90% of sites (See Table 2 below). At the same time, we acknowledge that the persistent identification of users is happening on the server-side as well (e.g., Acar et al., 2014), in ways that are more challenging for researchers to detect. Hence, we expect our results to be considered as a lower bound on the amount of persistence identification of users by SSPs across the crawled websites. A full visualization of persistent identification patterns within SSP networks, by site, can be found in appendix #2.

SSP Network	Average percentage of websites across which our browser was persistently identified
Pubmatic	86.4%
OpenX	77.6%
AppNexus	90%
Rubicon	90%

Table 2: Amount of Persistent Identification Across Websites Per Crawled SSP Network

Second, we wanted to see the overlap of persistent identifiers across SSPs. Can we spot the same ID cookie value by two or more SSPs, hinting that those SSPs are dynamically identifying individuals in the same way to potentially link user data across the sites they work with? Interestingly, from our browser-side observations, we saw how in all cases, despite one instance recurring across all 10 individual stateless crawls we could not fully explain (see appendix #3), persistent identification across SSP networks was non-existent. We could not trace the same cookie ID or any cookie value being dropped on our browser by two different SSPs. We acknowledge that cookie-syncing between SSPs might happen on the server-side, away from our crawler, but competition considerations between those actors make it unlikely. Advertisers choose to work with multiple SSPs to bid on users and base the selection of SSP partners primarily on ability to reach different audience sizes by publisher unique monthly visitors and delivery of accurate inventory performance (Sluis, 2018; Vargas, 2022a). These considerations related to

expanding the addressable audience and differentiation in services offered by SSPs demonstrate clear competition between SSPs that would disincentivize cookie sharing.

In summary, going back to our tracking specifications, we observed how ***users can be identifie***d via third-party cookies placed by supply-side networks (SSPs) on publisher sites. This provides SSPs with ***cross-context visibility*** when using ID cookies to persistently identify and potentially track individuals across the websites they are embedded in. Still, our data show that SSP-based identification of users remains within the network of publisher sites per SSP, and not crossing SSP networks. The observed tracking method also enables ***tracking over time***, as long as consumers can be identified or linked to the same cookie.

Going beyond our data collection and relying on our review of the industry's publications and technical documents, third-party cookies also enable advertisers to learn about users' browsing history and past behavior for bidding on publishers' ad inventory across sites. Importantly, cookies can be linked to other data sources, allowing advertisers to pair their data with cookies, beyond what passive surveillance of browsing behavior can provide, ***potentially circumventing targeting restrictions*** by advertising platforms. Advertisers are able to match third-party cookie identifiers to data purchased from data brokers to target users based on potentially sensitive assembled categories (Experian, n.d.; Sherman, 2021). In terms of ***user data sources***, minimal limitations are currently in place. Even though we have not directly measured potential participants in user tracking, previous studies showed how a range of ad networks actors, online and offline, can contribute to profiling consumers for marketing purposes (Choi et al., 2020; Wei et al., 2020).

3.2. Cookie-less tracking

Following Google's announcement in 2020 that the company would no longer support third-party cookies within its digital advertising products, first-party data collection from publisher sites began to arise as the digital advertising consensus to preserve some of the functionality provided by third-party cookies to tracking and targeting in the programmatic bidding process (Schuh, 2020; Southern, 2020). Publishers began to better organize and intensify user identification in first-party held data through means such as subscriptions, non-paying subscriber registration, and newsletters sign-ups (Asim, 2021). While first-party user data segments have used targeting categories constructed from demographic and on-site behavioral data to provide users with relevant advertisements directly on individual publisher websites, advertisers are still interested in capturing information about user behavior across the web that resembles tracking from cookie-based persistent identification patterns. As the preservation of precision in targeting relevant audiences that occurs from being able to follow behavior across sites is still a top priority for advertisers, SSPs were initially thought to potentially be a coordinating body that could aggregate first-party data across publisher sites so that advertisers could still have access to relevant audiences across publisher sites (Joseph, 2021). However, alternative solutions, labeled under the umbrella terms 'Universal IDs' and 'Alternative IDs,' have started to proliferate that work across multiple SSPs networks, enabling the identification of individual users more persistently than current, SSP-partitioned identification.

In order to compare proposals to existing cookie-based solutions, we have selected what we consider to be the three primary alternative identifier solutions - 'The Trade Desk Unified ID 2.0 (UID 2.0),' 'LiveRamp RampID,' and 'Secure Web Addressability Network (SWAN).' Our selection was informed by our reading of trade publications covering the digital advertising industry and the frequency of reporting on specific solutions. We have incorporated the reporting into the data we have collected along with messaging and technical documentation that the authors of the different solutions have made public. Hence, based on our suggested tracking specifications in section 3, we analyze each of the three main cookie-less tracking solutions discussed by the industry.

In terms of the ***instrument of user identification***, the three solutions vary. For SWAN, an individual is identified when they first visit a publisher site that has adopted the solution (Asim, 2022; Schiff, 2021d; Thomson & Rescorla, 2021). Upon loading the publisher page, the user is presented with a pop-up asking for consent to show personalized advertising on the current site and other sites that have adopted the solution. As part of the pop-up, the user also has the option to share their email address which can serve as an identifier. Regardless if the user accepts the option to receive personalized advertising with or without sharing their email address, a first-party cookie is placed by the initially visited publisher site within the SWAN network that creates a pseudonymous identifier that is stored on the user's browser. For UID 2.0, the solution authored by a top demand-side platform (DSP), user identity is established by logging via emails into publisher sites. Publishers store the email address in a first-party cookie placed on the page (Asim, 2022; Thomson & Rescorla, 2021; UnifiedID2, 2022). The email is matched to a UID2.0 through a corresponding token is also created and is used for encryption of the UID2.0 and can be decrypted only by partners that receive a decryption key through agreeing to the Unified ID 2.0's terms of service. The creation of the pseudonymous UID2 is managed by the UID2.0 service. The third solution under analysis, The LiveRamp RampID, is distinct from the other previous two, in that it is interoperable with other ID solutions, including UID2.0 (Asim, 2022). The RampID uses user email addresses that are matched with email addresses that are shared with publishers in exchange for content (Asim, 2022; LiveRamp, 2022c). Instead of placing a first-party cookie, LiveRamp matches IDs in the ecosystem through its proprietary authenticated traffic solution (ATS) (Asim, 2022; LiveRamp, 2022d). The ability of this solution to further identify users in the ecosystem is also attached to other offline PII (phone number, address history), based on information that advertisers can match with assembled first, second and third-party data.

Regarding ***cross-context user visibility,*** we see that access to SSPs and their publisher partners still gives advertisers the opportunity to reach users across sites. The orientation of an ID solution recreates that of the existing SSP to publisher tracking. While partnering with primary SSPs to coordinate the identifier programmatic bidding is transacting on, these solutions are also recreating an identifier to recognize users across SSPs (Asim, 2022). In Section 3.1, we showed how current cookie-based identification of users seems partitioned by individual SSPs. For the analyzed cookie-less tracking solutions, however, we argue that not only are these networks replicated across publisher sites in SSP partnerships, but also the integration of ID solutions by industry leading SSPs could potentially lead to identification of users across SSPs, not only within SSPs. As a result, greater identification of users across publisher sites will be enabled. All three analyzed solutions have been supported and committed adoption from primary SSPs, and The

Trade Desk Unified ID 2.0 has announced partnerships with all four SSPs we analyzed as part of our data crawl (Asim, 2022; Schiff, 2020a; Schiff, 2020b; Schiff, 2020c; Schiff, 2021a; Schiff, 2021d). This delegates a large amount of responsibility to those that will be assigned with governing these solutions, as the view of consumer behavior across the web will be expanded.

Longitudinal tracking is clearly enabled by all three cookie-less tracking solutions. The persistence derives from deterministic data serving as the basis for each solution (Asim, 2022; Kaye, 2021b). Each offers users the option to opt out of personalized advertising universally from all participating partners (Asim, 2022; LiveRamp, 2022b; SWAN-community, 2021; UnifiedID2, 2022). An argument can also be made that persistence can last for a greater duration than third-party cookies, which users frequently delete, as the choice to opt out could mean the user losing access to publisher content. While advertisers are able to bid on users that are classified to a persistent identifier, these specific solutions have not yet answered whether they will support the most persistent targeting method, retargeting or remarking - the following of a user after an initial action that classifies them as potentially more likely to carry out a desired action attributed to a digital advertisement.

Interestingly, all three solutions further encourage ***circumvention of targeting restrictions by advertising platforms.*** Advertisers can persistently identify users based on their first-party data, easily overriding targeting policies by main advertising platforms. We have previously alluded to the responsibility of those in charge of governing the new tracking solutions as they provide for greater persistent identification across the web. But what we highlight here is an under-discussed gap in cookie-less proposals. Advertisers can now enjoy an increased role in persistent identification through not just relying on publisher first-party data but also on the ability of the advertisers themselves to upload data to be encoded for targeting. Within The Trade Desk Unified ID 2.0, for instance, the company mentions ‘First-Party Relationships’ capabilities, where an advertiser is able to upload first-party data to be encoded to the UID2 for activation across publisher sites (UnifiedID2, 2022). Similarly, LiveRamp offers advertisers the opportunity to ‘onboard’ their data where PII can be uploaded in order to be converted into RampIDs and organized by segment so that they can be activated in more than 500 different partner platforms (LiveRamp, 2022a). SWAN does not provide a great deal of information related to these capabilities, but states on its homepage that it is ‘Complementary to CRM data’ (Secure Web Addressability Network (SWAN), 2021). This capability, enabled by all three solutions, was found in other alternatives to create the potential to further obfuscate advertiser targeting practices.

Such capability is concerning based on the work that can be done by advertisers before PII is encoded by identity solutions for targeting. As stated in Section 3.1, advertisers also choose to target users beyond the signals from cookies and purchase third-party data to expand profiling of users beyond what passive surveillance of browser behavior can provide. The emphasis on first-party data in cookie-less solutions has encouraged advertisers to begin leveraging their existing customer base to find other users that resemble the traits of current customers through modeling practices to create ‘look-alike’ segments (LiveRamp, 2020a). Compared to advertisers’ efforts in the current cookie-based ecosystem, and with the level of granularity enabled by targetable information IDs still unclear, advertisers are now investing more in technology that can match first-

party subscriber data to other datasets (Vargas, 2022b). This work is carried out on an 'identity graph' that allows advertisers to manage individual-level data and encode the data through an ID method of choice, providing a centralized system to merge online and offline identifiers into a consolidated profile to pair with purchased third-party data and activate selected audiences with multiple partners (LiveRamp, 2020c). An identity graph poses a threat to individual privacy as it allows advertisers to develop rich profiles of existing customers through pairing data purchased from data brokers but also non-customers across the open web (Vargas, 2022b). The Unified ID 2.0 and LiverRamp RampID now enable alarming ID linkage capabilities in this technology (Schiff, 2021b; Schiff, 2021c; Schiff, 2022b; The Trade Desk, 2021c). When an advertiser is able to upload a list of IDs by segment, there is the potential that advertisers are not only uploading explicit lists of existing customer PII but also emails that have been grouped according to specific audience segmentation categories. As these segments are constructed using data purchased from data brokers, there is a chance the construction of segments and profiles could involve sensitive categorical information. These IDs aim to be interoperable across major platforms, and the inability to verify or specify the segmentation practices of PII being converted to transact in major digital advertising platforms creates the opportunity for advertisers to violate targeting practices specific to these platforms (Google, 2022; Meta 2021; Twitter, 2022). There is no mechanism to verify how advertisers have segmented first-party data before importing into these systems, and new ID solutions makes sensitive population segmentation and bidding very attractive for advertisers.

For example, a credit card company might want to target African-American men specifically, but is unable to do so though the restricted targeting categories on certain primary digital advertising platforms. Instead, the credit card company could assemble a list of prospective customers based on data broker acquired data that can tie email address to race. If the credit card company chose to encode their data with the LiveRamp ID, for instance, and activate on any of the 500 partner platform destinations the company claims to partner with, the company just provides a link to certain companies' targeting restrictions without a thorough review (LiveRamp, 2022a, LiveRamp, 2022b). Previous research has identified instances where advertiser-uploaded lists have violated platform policies when targeting users, using categories such as race, religion, politics, sex life, or health (Wei et al., 2020). We argue that new tracking ID solutions place a significant amount of importance on leveraging first-party data that encourages advertisers to look for ways to link online and offline data through ID solutions, allowing for targeting practices with little to no oversight by primary platforms. This capability encourages advertisers to profile users before even interacting with primary platforms, circumventing existing platform restrictions.

Regarding ***user data sources used for tracking***, we found that similar to cookie-based tracking, the limitation of participants allowed to utilize and transact on these IDs remains unclear. The basic structure of third-parties responsible for identifying individuals remains almost identical. A different party now provides the persistent identifier and advertisers still pass IDs to bid on individuals across sites through demand-side platforms. Providing general terms detailing how participants are expected to adhere to certain principles, promises of outlining a code of conduct, or keeping proprietary governance privileged, it is unclear what other members of the current ecosystem can participate (Asim, 2022; LiveRamp, 2022c; Thomson & Rescorla, 2021; SWAN-

community, 2021; UnifiedID2, 2022). This issue has not been resolved, even though cookie-less tracking solutions claim to provide great individual privacy through limiting the sharing of persistent individual-level information to various parties. We found that parties responsible for governing the solutions are not clear nor are the terms that participants are expected to follow. The SWAN solution will be governed by the SWAN Network itself which has outlined a set of 'Model Terms' that details how participants are expected to adhere to information sharing practices (Thomson & Rescorla, 2021; SWAN-community, 2021). The Trade Desk claims that the company plans to turn over control to an 'Administrator,' a role that has not yet been filled (Asim, 2022). The Trade Desk documentation further mentions a 'code of conduct' that participants must follow, but is not currently available (UnifiedID2, 2022). The role of the administrator was originally supposed to be maintained by the Interactive Advertising Bureau, but the organization has chosen to no longer pursue supporting the solution in this capacity (Katsur, 2022; Mitchell, 2021). Prebid has also declined to serve as the administrator (Shields, 2022). Both SWAN and the Unified ID 2.0 are open source, but LiveRamp ID is a proprietary solution that is managed by the company itself (Asim, 2022; LiveRamp, 2022c). Based on the lack of transparency, it is not clear how these solutions will enforce standards to ensure participants do not violate terms of operation that have not been fully defined or made public.

3.3. Comparing cookie-based and cookie-less tracking mechanisms

Table 3 below summarizes our comparison between current and future Web tracking solutions by the advertising industry. The table shows how consumer surveillance is expected to expand in terms of (1) wider persistent identification patterns, (2) potentially spanning tracking across larger time periods, and (3) incentivizing advertisers to circumvent targeting restrictions and bid on consumers based on sensitive and richer first-party data profiles.

Our findings show how surveillance on the web is in the process of increasing from its current cookie-based organization to one transacting on ID solutions. First, persistent identification has increased as identification is not limited to individual SSP networks. Second, the source of data responsible for determining identity is derived from PII and consent mechanisms, making it difficult for individuals to opt out of an exchange for content that enables tracking. Third, the capability to upload data has encouraged advertisers to find ways to segment audiences before interacting with primary platforms. This has resulted in the purchase of offline data from data brokers that can be paired with existing customer data or used to derive potential customers based on similarity to those current customers or other preferred traits. This practice allows for the selection of traits that can violate the policies of primary platforms that prohibit certain sensitive categories to be used for targeting of users. The violation of platform policy is a result of not having a method to thoroughly verify how individuals are encoding generated segments to ID solutions before activating in platforms.

	Cookie-based Tracking	Cookie-less Tracking		
	SSP Cookie IDs	**UID 2.0**	**SWAN**	**LiveRamp**
User Identification Instrument	Passive placing of third-party ID cookies	First-party cookie based on consent obtained on publisher site & sharing of email	First-party cookie based on consent obtained on publisher site & sharing of email	Proprietary authenticated traffic solution based on consent on publisher site & sharing of email merged with offline name, address, and phone number
Cross-site User Visibility	Within all four SSPs, across 77%-90% of sites in a network, but not across SSPs	Across, not only within all four SSP networks	Across, not only within all four SSP networks	Across, not only within all four SSP networks
Longitudinal Tracking	Yes	Yes - with reliance on more deterministic data	Yes - with reliance on more deterministic data	Yes - with reliance on more deterministic data
Circumvention of Targeting Restrictions	Yes	Yes - and encourage greater profiling by pairing advertiser 1st-party data to purchased 3rd-party data	Yes - Alludes to the ability to pair CRM data	Yes - and encourage greater profiling by pairing advertiser 1st-party data to purchased 3rd-party data
User Data Sources	No limitations are in place	Limitations on participants and governance mechanisms still unclear	Limitations on participants and governance mechanisms still unclear	Limitations on participants and governance mechanisms still unclear

Table 3 - Tracking Specifications for marketing purposes via cookies and cookie-less solutions

4 - Discussion

Based on data collection from the four main SSP actors on cookie-based tracking, and analysis of technical documentation from the advertising industry on cookie-less tracking solutions, our study highlights three tracking specifications through which the main cookie-less tracking solutions are expected to increase consumer surveillance on the Web. First, persistent identification of consumers across sites is likely to cross SSP networks, creating potentially greater real-time visibility on consumers. Second, identification mechanisms are expected to rely on PIIs, making them more persistent and likely to better track users over time. Third, with the pivoting of the industry toward 'identity graphs' and first-party data, advertisers are now incentivized to target consumers based on rich first- and third-party data profiles, potentially overriding existing targeting restrictions in case advertisers choose to target consumers based on sensitive and forbidden categories. Hence, consumer tracking on the Web is likely to become dynamically wider, involve richer consumer histories, and rely on a greater variety of data sources.

In contrast to existing criticism on cookie-less tracking solutions, our study shows the structural impact of ad networks on the potential privacy harms by the proposed tracking mechanisms. We systematically contrast future solutions with current, operational, tracking mechanisms based on our developed tracking specifications. We look beyond deficiencies in consent mechanisms (Kaye, 2021a; Thomson & Rescorla, 2021) or lack of a clear governing body (UnifiedID2, 2022) to show how structures and processes of ad networks lead to potentially longer, wider, and richer consumer surveillance, compared to current tracking mechanisms.

Our findings pose serious questions on the ability of the suggested tracking mechanisms to become GDPR compliant. Despite civil society pushback against the current cookie-based tracking ecosystem (Lomas, 2021) and the marketing of new tracking solutions as 'privacy-friendly' (The Trade Desk, 2021a; Thomson & Rescorla, 2021), we show how the new mechanisms can practically enable granular tracking and targeting of individuals based on sensitive categories. Moreover, 'singling out' individuals might become easier by advertisers, given the historically rich consumer profiles that advertisers are incentivized to build (GDPR's recital 26). In contrast to existing debates about Universal ID solutions and GDPR compliance, that mostly focus on the lack of data controller and data processor roles in the proposed architecture (Schiff, 2021e; Schiff; 2022a), or the role of cookies in the suggested solutions (Asim, 2021; Asim, 2022), we would like to stress the extent of profiling and targeting that the new solutions allow, and their possible violations of key GDPR principles.

First, the potential richness of personal information involved in the projected profiling tactics may go against or beyond consumers' reasonable expectations and might infringe applicable data protection principles and rules. When an advertising company, for instance, joins its own first-party data with third-party data sources, as described in Section 3.2, this may result in personal data being used beyond their initial purpose and in ways the individual could not reasonably anticipate. The profiles built by the advertiser might involve an inference of interests or characteristics which individuals had not actively disclosed, undermining the ability of individuals to exercise control over their personal data (EDPS, 2018), or creating an opportunity for advertisers to 'single out' individuals from their data (GDPR's recital 26). Moreover, GDPR's

transparency requirements might be violated as the role of different parties in this process is probably unclear to the user given the potential circumvention of platforms' targeting restrictions (see Section 3.2 for an example).

Second, the suggested tracking mechanisms might encourage discrimination and exclusion. Targeting by advertisers can involve criteria that, directly or indirectly, have discriminatory effects relating to an individual's racial or ethnic origin, health status or sexual orientation, or other protected qualities of the individual concerned. The potential for discrimination in targeting arises from the ability for advertisers to leverage the extensive quantity and variety of personal data given the pivoting of new tracking technologies around deterministic identification. As explained in Section 3.2, the ability of advertisers to link their own segmentation of users with ID solutions enables granular and possibly discriminatory targeting that overrides existing, non-enforceable, targeting restrictions, and thus, violates requirements in GDPR's Article 9 about processing special categories of personal data.

5 - Conclusion

The privacy implications of ad networks structures and processes lead to a privacy-concerning ad landscape even without third-party cookies. The implementation of cookie-less tracking solutions by the AdTech complex potentially enables greater dynamic visibility on consumers, longer consumer tracking, and the assembling of more sensitive consumer profiles.

Looking ahead, even though there is not a clear primary tracking solution in the group of offerings, and with industry leaders not believing that one solution will be responsible for continued persistent identification alone, the Trade Desk Unified ID 2.0 has emerged as a leading solution. Despite uncertainty about whether the Trade Desk Unified ID 2.0 and other solutions will be compliant with evolving privacy regulation due to the similarities between the proposal and existing tracking methods, support continues to amass for the Trade Desk based on marketing language that emphasizes privacy but also promises to deliver on the same capabilities provided by third-party cookies (Asim, 2021; The Trade Desk, 2021a). The criticism of the Trade Desk UID could be viewed as outweighed by not only industry support from agencies such as IPG, Omnicom, and Publicis Groupe (Bürgi, 2021a; Bürgi, 2021b; The Trade Desk, 2021b; The Trade Desk, 2021c) but also by Google. While Google initially announced in 2021 that it would not support email or alternative third-party identifiers in its ecosystem due to their belief that the solution would not be sustainable in the evolving regulatory environment (Temkin, 2021), the company has reserved course through the introduction of encrypted signals from publishers (ESP) product that allows publishers to share encrypted first-party data and alternative identifiers via Google's Ad Manager (Schiff, 2022c; Google Ad Manager Help, 2022). With this backing, the Trade Desk now has support from the parties with high interests in maintaining budgets and practices in programmatic advertising.

Given our findings in this study, this is an alarming trend. The implementation of UID 2.0 by the advertising industry might make consumer surveillance even worse. As long as the structure of ad networks will not fundamentally change, consumers' profiling will persist as a highly valued commodity.

Our study has a few limitations. First, for our evaluation of cookie-based mechanisms, we crawled 50 landing pages of popular websites and not their inner pages, where tracking is known to be more pervasive. Second, we compare cookie values based on identical strings, even though some actors might encrypt or hash their cookie values, making us miss some persistent identification trends. Third, we cannot trace information being shared on the server side and fully capture the amount of detail ad network actors have about the individual user. The three limitations lead us to assume that our findings on cookie-based tracking suggest a lower bound for actual persistent identification patterns. Previous studies support our assumption, acknowledging that advertising actors often match first-party user data with the observed cookie IDs to achieve more granular targeting capacities (Trusov et al., 2016).

Future follow-up research projects could invest in understanding how compliance to privacy initiatives is being shifted to the individual advertiser level, as demonstrated in Section 3.2. Advertising platforms create policies to restrict sensitive categories from targeting, but at the same time still giving advertisers the capability to target users with precision. Those loosely-enforceable mechanisms are part of the privacy problem of ad-networks and a structural change is much needed.

References

Acar Gunes, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. (2014). “The Web Never Forgets: Persistent Tracking Mechanisms in the Wild.” *Proceedings of the ACM Conference on Computer and Communications Security (CCS).*

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness. *Journal of Retailing*, *91*(1), 34–49. https://doi.org/10.1016/j.jretai.2014.09.005

Asim, A. (2021, November 17). *Digiday Media Research:A comprehensive guide to third-party cookie alternatives.* Digiday https://digiday.com/media/a-digiday-media-guide-to-third-party-cookie-alternatives/

Asim, A. (2022, June 9). *Digiday+ Research: A guide to the top 10 ID alternatives for publishers.* Digiday. https://digiday.com/media/digiday-research-a-guide-to-the-top-10-id-alternatives-for-publishers/

Bashir, M. A., Arshad, S., Wilson, C., & Robertson, W. (2016). Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. *Proceedings of the 25th USENIX Security Symposium August 10–12, 2016 • Austin, TX*, 17.

Binns, R. (2022). Tracking on the Web, Mobile and the Internet of Things. *Foundations and Trends® in Web Science*, *8*(1–2), 1–113. https://doi.org/10.1561/1800000029

Bürgi, M. (2021a, August 10). *Unified ID 2.0 quietly amasses more support from the agency world, but publishers aren’t as convinced.* Digiday. https://digiday.com/media/unified-id-2-0-quietly-amasses-more-support-from-the-agency-world-but-publishers-arent-as-convinced/

Bürgi, M. (2021b, November 17). *Omnicom Media Group formally endorses UID 2.0 in a bid to move the post-cookie future forward.* Digiday. https://digiday.com/marketing/omnicom-media-group-formally-endorses-uid-2-0-in-a-bid-to-move-the-post-cookie-future-forward/

Choi, H., Mela, C. F., Balseiro, S. R., & Levy, A. (2020). Online Display Advertising Markets: A Literature Review and Future Directions | Information Systems Research. *Information Systems Research*, *2*(31), 556–575. https://doi.org/10.1287/isre.2019.0902

D’Annunzio, A., & Russo, A. (2020). Ad Networks and Consumer Tracking. *Management Science*, *66*(11), 5040–5058. https://doi.org/10.1287/mnsc.2019.3481

Dinev, T. (2014). Why would we care about privacy? *European Journal of Information Systems*, *23*(2), 97–102. https://doi.org/10.1057/ejis.2014.1

Dinev, T., & Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts. *European Journal of Information Systems*, *22*(3), 295–316. https://doi.org/10.1057/ejis.2012.23

Edelman, G. (2020, October 5). Ad Tech Could Be the Next Internet Bubble. *Wired*. https://www.wired.com/story/ad-tech-could-be-the-next-internet-bubble/

EDPS. (2018). *EDPS Opinion on online manipulation and personal data*. https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf

Englehardt S. and A. Narayanan. (2016). “Online Tracking: A 1-million-site Measurement and Analysis.” *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1388-1401.

Experian. (n.d.). *ConsumerView: Data by the Numbers*. https://www.experian.com/content/dam/marketing/na/assets/ems/marketing-services/documents/infographics/consumerview.pdf

Fouad I., N. Bielova, A. Legout, N. Sarafijanovic-Djukic. (2020). “Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels.” Available in: https://arxiv.org/abs/1812.01514

Goel, Vinay. (2021, June 24). An updated timeline for privacy sandbox milestones. *The Keyword*. https://blog.google/products/chrome/updated-timeline-privacy-sandbox-milestones/
Google Ad Manager Help. (2022, March). Share encrypted signals with bidders (Beta). G*oogle Ad Manager Help*. https://support.google.com/admanager/answer/10488752

Goldfarb, A., & Tucker, C. (2011). Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science*, *30*(3), 389–404. https://doi.org/10.1287/mksc.1100.0583

Goldfarb, A., & Tucker, C. E. (2015). Standardization and the Effectiveness of Online Advertising. *Management Science*, *61*(11), 2707–2719. https://doi.org/10.1287/mnsc.2014.2016

Google. (2022). *Personalized advertising - Advertising Policies Help*. https://support.google.com/adspolicy/answer/143465?hl=en

Google Ad Manager Help. (2022, March). Share encrypted signals with bidders (Beta). G*oogle Ad Manager Help*. https://support.google.com/admanager/answer/10488752

Gopal, R. D., Hojati, A., & Patterson, R. A. (2022). Analysis of third-party request structures to detect fraudulent websites. *Decision Support Systems*, *154*, 113698. https://doi.org/10.1016/j.dss.2021.113698

Hui, K.-L., Teo, H. H., & Lee, S.-Y. T. (2007). The Value of Privacy Assurance: An Exploratory Field Experiment. *MIS Quarterly*, *31*(1), 19–33. https://doi.org/10.2307/25148779

Jones, M. L. (2020). Cookies: A legacy of controversy. *Internet Histories*, *4*(1), 87–104. https://doi.org/10.1080/24701475.2020.1725852

Joseph, S. (2021, November 17). *'They will need to use multiple routes': Shifts appear in the publisher-SSP union, as alternative identifiers proliferate.* Digiday. https://digiday.com/media/they-will-need-to-use-multiple-routes-shifts-appear-in-the-publisher-ssp-union-as-alternative-identifiers-proliferate/

Kaye, K. (2021a, March 24). *Third-party cookie replacements fall short of consent and transparency promises*. Digiday. https://digiday.com/media/third-party-cookie-replacements-fall-short-of-consent-and-transparency-promises/

Kaye, K. (2021b, April 1). *WTF is the difference between deterministic and probabilistic identity data?.* Digiday. https://digiday.com/media/wtf-is-the-difference-between-deterministic-and-probabilistic-identity-data/

Katsur, A. (2022, February 28). Tech Lab Update on UID2.0. *IAB Tech Lab*. https://iabtechlab.com/blog/tech-lab-update-on-uid2-0/

Lavin, M. (2006). Cookies: What do consumers know and what can they learn? *Journal of Targeting, Measurement and Analysis for Marketing*, *14*(4), 279–288. https://doi.org/10.1057/palgrave.jt.5740188

Li, T., & Unger, T. (2012). Willing to pay for quality personalization? Trade-off between quality and privacy. *European Journal of Information Systems*, *21*(6), 621–642. https://doi.org/10.1057/ejis.2012.13

Li, Y. (2011). Empirical Studies on Online Information Privacy Concerns: Literature Review and an Integrative Framework. *Communications of the Association for Information Systems*, *28*. https://doi.org/10.17705/1CAIS.02828

Li, Y. (2012). Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, *54*(1), 471–481. https://doi.org/10.1016/j.dss.2012.06.010

Lively, T. K. (2022, July 7). *US State Privacy Legislation Tracker*. https://iapp.org/resources/article/us-state-privacy-legislation-tracker/

LiveRamp. (2020a, March 31). *Look-alike Modeling: The What, Why, and How.* https://liveramp.com/blog/look-alike-modeling-the-what-why-and-how/

LiveRamp. (2020b, August 11). *Platform-Specific Distribution Information*. https://docs.liveramp.com/connect/en/platform-specific-distribution-information.html

LiveRamp. (2020c, August 28). *What's the Difference between a DMP and an Identity Graph?.* https://liveramp.com/blog/difference-between-dmp-data-management-platform-identity-graph/

LiveRamp. (2022a, March 8). *Onboarding Your Data*. https://docs.liveramp.com/connect/en/onboarding-your-data.html

LiveRamp. (2022b, May 6). *Consumer Requests for Opt-Outs, Data Access, or Data Deletions*. https://docs.liveramp.com/connect/en/consumer-requests-for-opt-outs,-data-access,-or-data-deletions.html

LiveRamp. (2022c, June 30). *RampID Methodology*. https://docs.liveramp.com/connect/en/rampid-methodology.html

LiveRamp. (2022d, July 21). *Authenticated Traffic Solution*. https://docs.liveramp.com/privacy-manager/en/authenticated-traffic-solution.html

Lomas, N. (2021, June 16). Adtech 'data breach' GDPR complaint is headed to court in EU. *TechCrunch*. https://social.techcrunch.com/2021/06/16/adtech-data-breach-gdpr-complaint-is-headed-to-court-in-eu/

Mac, R., & Kang, C. (2021, October 3). Whistle-Blower Says Facebook 'Chooses Profits Over Safety.' *The New York Times*. https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html

Magnite (2021a, August 27). *Data Subject Rights Policy*. Magnite. https://www.magnite.com/legal/data-subject-rights-policy/

Magnite (2021b, August 27). *Platform Cookies statement*. Magnite. https://www.magnite.com/legal/platform-cookie-statement/

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, *15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Marotta, V., Wu, Y., Zhang, K., & Acquisti, A. (2022). The Welfare Impact of Targeted Advertising Technologies. *Information Systems Research*, *33*(1), 131–151. https://doi.org/10.1287/isre.2021.1024

Martin, K. (2022). Finding Consumers, No Matter Where They Hide: Ad Targeting and Location Data. In *Ethics of Data and Analytics: Concepts and Cases* (pp. 99–111). Auerbach Publications. https://doi.org/10.1201/9781003278290-16

Meta. (2021, November 9). *Removing Certain Ad Targeting Options and Expanding Our Ad Controls*. https://www.facebook.com/business/news/removing-certain-ad-targeting-options-and-expanding-our-ad-controls

Mitchell, J. (2021, January 21). Reviewing Unified ID 2.0 for Long-Term Industry Value – IAB Tech Lab. *IAB Tech Lab*. https://iabtechlab.com/blog/reviewing-uid2-for-long-term-industry-value/

OpenX (2022, March 3). *OpenX Ad Exchange Privacy Policy*. OpenX. https://www.openx.com/privacy-center/ad-exchange-privacy-policy/

O'Reilly, L. (2020, January 14). Google plans to kill off third-party cookies in Chrome "within 2 years." *Digiday*. https://digiday.com/media/google-plans-kill-off-third-party-cookies-chrome-within-2-years/

Parkin, R. (2021, June 27). *What the Delay to the End of Third-Party Cookies Means for Advertisers*. AdExchanger. https://www.adexchanger.com/data-driven-thinking/what-the-delay-to-the-end-of-third-party-cookies-means-for-advertisers/

Pubmatic (2020, July 1). *Platform Cookie & Other Similar Technologies Policy*. Pubmatic. https://pubmatic.com/legal/platform-cookie-policy/

Roesner Franziska, Tadayoshi Kohno, and David Wetherall. (2012). "Detecting and defending against third-party tracking on the web." In *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, NSDI*, pages 155–168.

Schiff, A. (2020a, November 17). *Magnite Hops Aboard The Unified ID 2.0 Train*. AdExchanger. https://www.adexchanger.com/online-advertising/magnite-hops-aboard-the-unified-id-2-0-train/

Schiff, A. (2020b, November 19). *PubMatic Is The Latest Ad Tech Company To Join Unified ID 2.0*. AdExchanger. https://www.adexchanger.com/publishers/pubmatic-is-the-latest-ad-tech-company-to-join-unified-id-2-0/

Schiff, A. (2020c, December 16). *OpenX Is Latest SSP To Join Unified ID 2.0*. AdExchanger. https://www.adexchanger.com/online-advertising/openx-is-latest-ssp-to-join-unified-id-2-0/

Schiff, A. (2021a, March 3). *Xandr Integrates With Unified ID 2.0 And Outlines Its Identity Roadmap*. AdExchanger. https://www.adexchanger.com/online-advertising/xandr-integrates-with-unified-id-2-0-and-outlines-its-identity-roadmap/

Schiff, A. (2021b, April 21). *ID Graph Provider Infutor Joins The Club With Support Unified ID 2.0.* AdExchanger. https://www.adexchanger.com/online-advertising/id-graph-provider-infutor-joins-the-club-with-support-unified-id-2-0/

Schiff, A. (2021c, May 26). *LiveRamp Launches Identity Resolution For First-Party Data. AdExchanger*. https://www.adexchanger.com/data-exchanges/liveramp-launches-identity-resolution-for-first-party-data/

Schiff, A. (2021d, May 27). *SWAN Vs. SWAN: The Differences Between The Two 3P Cookie Alternative Proposals*. AdExchanger. https://www.adexchanger.com/online-advertising/swan-vs-

swan-the-differences-between-the-two-3p-cookie-alternative-proposals/

Schiff, A. (2021e, November 16). *Unified ID 2.0 Faces Roadblocks In Europe As A Result Of GDPR*. AdExchanger. https://www.adexchanger.com/privacy/unified-id-2-0-faces-roadblocks-in-europe-as-a-result-of-gdpr/

Schiff, A. (2022a, March 3). *IAB Tech Lab Declines To Be The Admin For UID2 (But Hope Springs Eternal)*. AdExchanger. https://www.adexchanger.com/online-advertising/iab-tech-lab-declines-to-be-the-admin-for-uid2-but-hope-springs-eternal/

Schiff, A. (2022b, March 24). *Epsilon Is Making Its Identity Platform Interoperable With Unified ID 2.0*. AdExchanger. https://www.adexchanger.com/data-exchanges/epsilon-is-making-its-identity-platform-interoperable-with-unified-id-2-0/

Schiff, A. (2022c, March 24). *Google's Encrypted Signals Program Just Entered Open Beta, And Here's What You Need To Know About It*. AdExchanger. https://www.adexchanger.com/ad-exchange-news/googles-encrypted-signals-program-just-entered-open-beta-and-heres-what-you-need-to-know-about-it/

Schuh, J. (2020, January 14). Building a more private web: A path towards making third party cookies obsolete. *Chromium Blog*. https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html

Shein, E. (2021, September 14). *Third-party cookies are going away: What advertisers, marketers and consumers should know*. TechRepublic. https://www.techrepublic.com/article/third-party-cookies-are-going-away-what-advertisers-marketers-and-consumers-should-know/

Sherman, J. (2021). *Data Brokers and Sensitive Data on U.S. Individuals*. Duke University Sanford School of Public Policy. https://sites.sanford.duke.edu/techpolicy/report-data-brokers-and-sensitive-data-on-u-s-individuals/

Shields, R. (2022, January 7). *With Prebid at the helm of UID 2.0, indie ad tech marches to a unified beat but not all voices are in harmony*. Digiday. https://digiday.com/media/with-prebid-at-the-helm-of-uid-2-0-indie-ad-tech-marches-to-a-unified-beat-but-not-all-voices-are-in-harmony/

Sluis, S. (2018, November 20). *Advertiser Perceptions: How SSPs Can Win Market Share From Google*. AdExchanger. https://www.adexchanger.com/platforms/advertiser-perceptions-how-ssps-can-win-market-shar-from-google/

Solomos, K., Kristoff, J., Kanich, C., & Polakis, J. (2021). Tales of Favicons and Caches: Persistent Tracking in Modern Browsers. *Proceedings 2021 Network and Distributed System Security Symposium*. https://doi.org/10.14722/ndss.2021.24202

Southern, L. (2020, January 29). *Beyond ad targeting, the demise of the third-party cookie will hit key digital media functions*. Digiday. https://digiday.com/media/cookie-collateral-damage/

Secure Web Addressability Network (SWAN). (2021). *Secure Web Addressability Network (SWAN)*. https://swan.community/

SWAN-community (2021). *Secure Web Addressability Network (SWAN) - Model Terms Explainer*. https://github.com/SWAN-community/swan/blob/main/model-terms-explainer.md

Temkin, D. (2021, March 3). Charting a course towards a more privacy-first web. *Google*. https://blog.google/products/ads-commerce/a-more-privacy-first-web/

The Trade Desk. (2021a, February 21). *What the Tech is Unified ID 2.0?*. The Trade Desk. https://www.thetradedesk.com/us/news/what-the-tech-is-unified-id-2-0

The Trade Desk. (2021b, April 8). *Publicis Groupe all-in on first-party identity solution*. The Trade Desk. https://www.thetradedesk.com/us/news/publicis-groupe-all-in-on-first-party-identity-solution

The Trade Desk. (2021c, July 28). I*PG Leans into Unified ID 2.0 as a Closed Operator*. The Trade Desk. https://www.thetradedesk.com/cn/news/ipg-leans-into-unified-id-2-0-as-a-closed-operator

The Trade Desk. (2022). *Unified ID 2.0 Partners*. https://www.thetradedesk.com/us/about-us/industry-initiatives/unified-id-solution-2-0/unified-id-2-partners

Thomson, M., & Rescorla, E. (2021, August). *Comments on SWAN and Unified ID 2.0*. Mozilla. https://mozilla.github.io/ppa-docs/swan_uid2_report.pdf

Tranco. (n.d.). *A research-oriented top sites ranking hardened against manipulation—Tranco*. *Retrieved July 26, 2022, from* https://tranco-list.eu/

Twitter. (2022). *Targeting of Sensitive Categories*. https://business.twitter.com/en/help/ads-policies/campaign-considerations/targeting-of-sensitive-categories.html

UnifiedID2 (2022). *uid2docs*. https://github.com/UnifiedID2/uid2docs

Vargas, A. (2022a, February 4). *Google Reigns Supreme In Latest Advertiser Perceptions SSP Report, But Competition Is Tight Among Everyone Else*. AdExchanger. https://www.adexchanger.com/online-advertising/google-reigns-supreme-in-latest-advertiser-peceptions-ssp-report-but-competition-is-tight-among-everyone-else/

Vargas, A. (2022b, April 12). *Goodway Group Stitches Together Identity Graph To Complement Brands' First-Party Data*. AdExchanger. https://www.adexchanger.com/agencies/goodway-group-stitches-together-identity-graph-to-complement-brands-first-party-data/

W3C Working Group. (2019, January 22). Tracking Compliance and Scope. Tracking Compliance and Scope. https://perma.cc/3HXA-F47U

Wei, M., Stamos, M., Veys, S., Retinger, N., Goodman, J., Herman, M., Filipczuk, D., Weinshel, B., Mazurek, M., & Ur, B. (2020). What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In *29th USENIX Security Symposium (USENIX Security 20)* (pp. 145-162).

Xandr (2022, April 29). *Digital Platform Cookie Policy*. Xandr. https://www.xandr.com/privacy/cookie-policy/

Xu, H., Dinev, T., Smith, J., & Hart, P. (2011). Information Privacy Concerns: Linking Individual Perceptions with Institutional Privacy Assurances. *Journal of the Association for Information Systems*, *12*(12). https://doi.org/10.17705/1jais.00281

Appendix #1 – 50 Sites Crawled to Evaluate Cookie-Based Tracking on SSP Networks

https://www.yahoo.com,
https://www.msn.com,
https://www.tumblr.com,
https://www.nytimes.com,
https://www.sohu.com,
https://www.imdb.com,
https://www.ebay.com,
https://www.forbes.com,
https://www.washingtonpost.com,
https://www.dailymail.co.uk,
https://www.tinyurl.com,
https://www.nature.com,
https://www.weather.com,
https://www.usatoday.com,
https://www.cnet.com,
https://www.tribunnews.com,
https://www.aol.com,
https://www.goodreads.com,
https://www.time.com,
https://www.foxnews.com,
https://www.ted.com,
https://www.merdeka.com,
https://www.wired.com,
https://www.independent.co.uk,
https://www.latimes.com,
https://www.huffingtonpost.com,
https://www.kompas.com,
https://www.theverge.com,
https://www.speedtest.net,
https://www.detik.com,
https://www.huffpost.com,
https://www.nicovideo.jp,
https://www.britannica.com,
https://www.buzzfeed.com,
https://www.usnews.com,
https://www.nypost.com,
https://www.merriam-webster.com,
https://www.9gag.com,
https://www.sciencedaily.com,
https://www.apnews.com,
https://www.youm7.com,
https://www.mirror.co.uk,
https://www.timeanddate.com,
https://www.gsmarena.com,
https://www.politico.com,
https://www.ndtv.com,
https://www.dictionary.com,
https://www.chron.com,
https://www.vnexpress.net,
https://www.thesaurus.com

Appendix #2 - Persistent Identification Within the Four SSP Networks

Site	Pubmatic	OpenX	AppNexus	Rubicon	Total
https://www.yahoo.com	5	9	10	3	27
https://www.msn.com	3	1	6	1	11
https://www.tumblr.com	8	10	10	10	38
https://www.nytimes.com	10	10	10	10	40
https://www.sohu.com	4	4	4	4	16
https://www.imdb.com	10	1	10	10	31
https://www.ebay.com	10	10	10	10	40
https://www.forbes.com	10	10	10	9	39
https://www.washingtonpost.com	10	9	9	9	37
https://www.dailymail.co.uk	10	10	10	10	40
https://www.tinyurl.com	7	4	10	10	31
https://www.nature.com	10	0	10	10	30
https://www.weather.com	8	8	7	10	33
https://www.usatoday.com	10	10	10	10	40
https://www.cnet.com	9	6	9	9	33
https://www.tribunnews.com	10	10	10	10	40
https://www.aol.com	4	3	8	6	21
https://www.goodreads.com	5	3	5	6	19
https://www.time.com	10	10	10	10	40
https://www.foxnews.com	10	10	10	10	40
https://www.ted.com	4	4	1	5	14
https://www.merdeka.com	10	10	1	10	31
https://www.wired.com	10	10	10	10	40
https://www.independent.co.uk	10	7	10	9	36
https://www.latimes.com	10	10	10	10	40
https://www.huffingtonpost.com	10	10	10	10	40
https://www.kompas.com	10	10	10	10	40
https://www.theverge.com	10	0	10	10	30
https://www.speedtest.net	10	10	10	10	40
https://www.detik.com	10	9	9	8	36
https://www.huffpost.com	9	10	8	9	36
https://www.nicovideo.jp	10	10	10	10	40
https://www.britannica.com	10	10	10	10	40
https://www.buzzfeed.com	7	10	10	10	37
https://www.usnews.com	10	10	10	10	40
https://www.nypost.com	9	5	10	9	33
https://www.merriam-webster.com	10	10	10	10	40
https://www.9gag.com	10	10	10	10	40
https://www.sciencedaily.com	3	10	10	10	33
https://www.apnews.com	10	10	10	10	40
https://www.youm7.com	5	10	10	10	35
https://www.mirror.co.uk	2	1	3	4	10
https://www.timeanddate.com	10	4	10	10	34
https://www.gsmarena.com	10	10	10	10	40
https://www.politico.com	10	10	10	10	40
https://www.ndtv.com	10	10	10	10	40
https://www.dictionary.com	10	10	10	10	40
https://www.chron.com	10	10	10	10	40
https://www.vnexpress.net	10	0	10	9	29
https://www.thesaurus.com	10	10	10	10	40
Average Site Persistence Per SSP	**43.2**	**38.8**	**45.0**	**45.0**	
Average Site Peristence % Per SSP	**86.4%**	**77.6%**	**90.0%**	**90.0%**	

Appendix #3 - observed sharing of information between Pubmatic and Rubicon

In our data collection on cookie-based tracking in SSP networks we did see that in two unique cookies' hostnames associated with Pubmatic, Rubicon was named in stateless crawl 1 of 10 (See Table 4 below). We did see other occurrences with different persistent identifiers and similar host name syntax in each of the 10 stateless crawls.

Looking further into Rubicon's (now known as 'Magnite') privacy policy, the company lists Pubmatic as an 'Advertising Cookies – Third Party Technology Providers' partner with 'DMP, Onboarders, Data Providers' activities associated with this classification.

We are still trying to figure out what this potential cookie syncing between those actors means for consumer tracking.

Persistent Identifier	Host Name
KADUSERCOOKIE26F46775-238D-433A-B321-8B4D7D2349A9	https://ads.pubmatic.com/AdServer/js/user_sync.html?gdpr=&gdpr_consent=&us_privacy=1---&predirect=https%3A%2F%2Fprebid-server.**rubiconproject.com**%2Fsetuid%3Fbidder%3Dpubmatic%26gdpr%3D%26gdpr_consent%3D%26us_privacy%3D1---%26account%3D%7B%7Baccount%7D%7D%26f%3Db%26uid%3Dnull
	https://ads.pubmatic.com/AdServer/js/user_sync.html?gdpr=0&gdpr_consent=&us_privacy=1NNN&predirect=https%3A%2F%2Fprebid-server.**rubiconproject.com**%2Fsetuid%3Fbidder%3Dpubmatic%26gdpr%3D0%26gdpr_consent%3D%26us_privacy%3D1NNN%26account%3D%7B%7Baccount%7D%7D%26f%3Db%26uid%3Dnull

Table 4: Possible ID cookie sharing between Pubmatic and Rubicon in stateless crawl 1 of 10