John Cook*, Rishab Nithyanand, and Zubair Shafiq

Inferring Tracker-Advertiser Relationships in the Online Advertising Ecosystem using Header Bidding

Abstract: Online advertising relies on trackers and data brokers to show targeted ads to users. To improve targeting, different entities in the intricately interwoven online advertising and tracking ecosystems are incentivized to share information with each other through client-side or server-side mechanisms. Inferring data sharing between entities, especially when it happens at the server-side, is an important and challenging research problem. In this paper, we introduce KASHF: a novel method to infer data sharing relationships between advertisers and trackers by studying how an advertiser's bidding behavior changes as we manipulate the presence of trackers. We operationalize this insight by training an interpretable machine learning model that uses the presence of trackers as features to predict the bidding behavior of an advertiser. By analyzing the machine learning model, we can infer relationships between advertisers and trackers irrespective of whether data sharing occurs at the client-side or the server-side. We are able to identify several server-side data sharing relationships that are validated externally but are not detected by client-side cookie syncing.

Keywords: cookie syncing, header bidding, online advertising and tracking

DOI 10.2478/popets-2020-0005
Received 2019-05-31; revised 2019-09-15; accepted 2019-09-16.

1 Introduction

Online vs. offline advertising. Online advertising is set to surpass offline advertising (*e.g.,* newspapers, yellow pages, radio, TV) this year. In fact, online advertising revenues in the US are expected to exceed two-thirds of total advertising spending by 2023 [30]. There are several reasons driving this shift from offline advertising to online advertising. First, consumers are increasingly spending more time online. This makes the web a more attractive platform for advertisers. Specifically, consumers in the US now spend about 24 hours a week online, which exceeds the time spent watching TV [47, 65]. Second, online advertising primarily relies on highly automated technologies that enable advertisers to programmatically launch advertising campaigns, measure their effectiveness, and quickly adjust them based on their performance. Programmatic advertising already accounts for 86% of all online display advertising in the US [48]. Third, online advertising allows targeting of advertising campaigns to specific audiences based on their demographics, location, or intents. Personalized online advertising campaigns are reported to be much more effective as compared to their non-personalized counterparts [4].

The online advertising ecosystem includes middle-men. Unlike offline advertising where there is typically a direct relationship between advertisers and publishers, the online advertising ecosystem comprises of several specialized entities that mediate interactions between advertisers and publishers. This is necessitated by the need for technical expertise to participate in protocols, such as real-time bidding (RTB), which require publishers and advertisers to identify, offer, and respond to ad impression opportunities in near real-time. The entities that fill this gap include *supply-side platforms* (SSPs) that put up ad inventory of publishers for sale at *ad exchanges* (AdXes), which are marketplaces that run real-time auctions for individual ad slots. Advertisers bid on individual ad slots auctioned off at AdXes through *demand-side platforms* (DSPs), which use sophisticated models to determine how much to bid for an ad slot based on the user information retrieved from *data management platforms* (DMPs). DMPs gather user information (*e.g.,* browsing history) through a variety of online tracking techniques such as cookies.

Entities engage in data sharing. Intuitively, an ad slot's value as assessed by a DSP, is highly dependent

***Corresponding Author: John Cook:** The University of Iowa, E-mail: john-cook@uiowa.edu
Rishab Nithyanand: The University of Iowa, E-mail: rishab-nithyanand@uiowa.edu
Zubair Shafiq: The University of Iowa, E-mail: zubair-shafiq@uiowa.edu

on the quality of information received from the DMP(s). Consequently, DMPs strive to enhance the quality of information by improving their ability to observe user behavior on different websites and platforms. This can be done by either (1) increasing their presence, as trackers, on the web or (2) developing data sharing relationships with other tracking services. Prior research has shown that only a few organizations (Google, Facebook, Twitter, Amazon, AdNexus, and Oracle) are able to track users on more than 10% of the top 1-million sites [45]. Thus, DMPs often choose to develop data sharing relationships rather than trying to arduously increase their presence on the web. In fact, the RTB protocol has built-in mechanisms to facilitate data sharing between advertisers and trackers. Cookie syncing (a.k.a. cookie matching) in RTB allows two different entities in RTB to exchange their cookies while bypassing the same-origin policy [32, 67]. Cookie syncing essentially allows two entities to map their cookies to each other and get a more complete view of a user's browsing history [52]. A recent study showed that cookie syncing increases the number of entities that track users by almost 7X [69]. Another recent study showed that, despite using privacy-enhancing technologies such as Ghostery and Disconnect, trackers are still able to observe anywhere from 40-80% of a user's browsing history due to cookie syncing in RTB [36].

Transparency of data sharing relationships is important. Privacy researchers and regulators are increasingly interested in studying data sharing relationships between different entities in the intricately interwoven online advertising and tracking ecosystems for several reasons. First, a complete understanding of such relationships can help detect whether a domain is a tracker and, in turn, improve the effectiveness of tracker blocking tools [58]. Blocking tools are presently the most effective protection users can employ against trackers. Second, it is important to uncover data sharing relationships between different organizations for regulatory compliance verification purposes. Both General Data Protection Regulation (GDPR) [3] in Europe and the California Consumer Privacy Act (CCPA) [13] in the US give people the right to know what personal information is being collected and whether (and with whom) it is being shared. Methods that can detect data sharing between different tracking/advertising organizations can help uncover unauthorized or undisclosed data sharing relationships.

Measuring client-facilitated data sharing is insufficient. Analysis of client-facilitated mechanisms in RTB, such as cookie syncing, to detect data sharing between entities is limited due to two reasons. First, prior research relies on different heuristics to detect cookie syncing at the client-side [32, 67, 69]. Unfortunately, these heuristics are brittle to changes in non-standardized implementations of cookie syncing, especially when obfuscation is employed [35]. Second, and more importantly, analysis of client-side mechanisms such as cookie syncing *cannot detect server-side data sharing between entities* [27]. Server-side tracking (*e.g.,* postback tracking [12]) is expected to grow in popularity as mainstream browsers, notably Safari and Firefox [66, 78], have started to implement stringent third-party cookie policies [14]. Thus, it is important to develop methods that can infer both client-side and server-side data sharing between different entities in the online advertising ecosystem.

Inferring server-side data sharing is challenging. It is particularly challenging to infer server-side data sharing because it is not directly observable from purely client-side measurements. To overcome this challenge, prior research has attempted to exploit artifacts that reflect semantics of how online ads are served, rather than relying on specific mechanisms such as cookie syncing. In a seminal work, Bashir et al. [35] exploited retargeting to infer data sharing even if it occurs on the server-side. Their key insight is that retargeting takes place only when data sharing occurs between AdXes on different sites. To operationalize this insight, the authors trained personas to trigger retargeting, which is detected using crowdsourcing, and then analyze inclusion chains to determine whether information is shared at client-side or server-side. Using retargeting to infer server-side data sharing is limited because retargeting represents only a subset of scenarios in which server-side data sharing occurs. More specifically, server-side information exchange is a necessary but not a sufficient condition to trigger retargeting. Furthermore, detecting retargeting is a challenging task that requires significant manual effort that is not only difficult to scale but also susceptible to human errors.

Inferring tracker-advertiser data sharing using header bidding. To address our inability to directly measure server-side data sharing, like previous work [35], we also leverage client-side observable artifacts of the online advertising ecosystem. However, instead of relying on retargeting, we rely on being able to observe the bids placed by DSPs or bidders (on behalf of advertisers) that participate in *header bidding* (HB) – a new programmatic advertising mechanism aimed at increasing publisher advertising revenue as compared to traditional RTB. In contrast to traditional RTB that only exposes the winning bid at the client-side, HB exposes *all* bids made by different advertisers at the client-side. The ability to precisely observe the bids placed by a given advertiser in HB[1] enables us to observe how advertiser bids vary as a persona's browsing history and tracker presence are modified. At a high-level, our approach (named Kashf) to inferring tracker-advertiser data sharing relies on the following insight: *advertisers with knowledge of a user's browsing history will bid differently (potentially higher) than an advertiser having no knowledge of a user's browsing history.* In order to operationalize this insight, we first selectively expose a persona's browsing history to different sets of trackers and record the bids made by an advertiser in HB. We then train interpretable machine learning models, using tracker presence as input features and bid values as the target variable, to accurately predict the bids made by an advertiser. We finally leverage the interpretability of the trained machine learning models to determine which features (*i.e.,* trackers) were most influential in predicting the values of bids placed by an advertiser. This enables us to make inferences about the presence of data sharing relationship (client-side or server-side) between advertisers and trackers.

Key contributions. This paper makes the following two key contributions.

- *Measuring bidding behavior of advertisers* (§3). We are able to draw several novel insights into how different advertisers value users. More specifically, we leverage HB, which exposes *all* bids made by different advertisers, to study how an advertiser's bidding preferences vary for different personas. We find that with the exception of users with a *Health* persona who are universally preferred, advertisers have very different persona preferences. We also find that advertisers often have a strong preference for certain personas and these rarely overlap. Furthermore, we are able to shed light on the practice of underbidding. We find that underbidding is very common with zero bids making up 22% of all bids.
- *Uncovering tracker-advertiser relationships* (§4). Our approach, Kashf, allows us to infer data sharing relationships between advertisers and trackers irrespective of whether they occur at the client-side or the server-side. To this end, we train machine learning models that use tracker information as features to predict the bidding behavior of advertisers with 75-83% accuracy. By analyzing the interpretable machine learning model, we are able to identify data sharing relationships between advertisers and popular trackers, most of which we are able to externally validate. We also demonstrate that many of these inferred server-side data sharing relationships are not detected by client-side cookie syncing.

2 Background

In this section, we provide an overview of online advertising and tracking ecosystems and highlight how they are intertwined.

2.1 Online Advertising Ecosystem

The contemporary online advertising ecosystem relies on programmatic processes to trade ad impressions in near real-time (*i.e.,* typically less than 100ms).

Real-Time Bidding (RTB). RTB is the most widely used programmatic process in online advertising. The typical RTB workflow illustrated in Figure 1 involves interactions between several entities in the advertising ecosystem. These include publishers, publisher ad servers, supply side platforms (SSPs), ad exchanges (AdX), demand side platforms (DSPs), and data management platforms (DMPs). The RTB process has three distinct phases: ad request, bid collection, and ad placement.

- **Ad request.** The workflow is initiated when the browser sends a request to fetch the publisher's web page ①. The publisher's web server responds with the HTML document that contains page content as well as the ad tag ②. While the rest of the page is loaded, the ad tag generates an ad request to an RTB-enabled SSP along with information about the ad slot (e.g., dimension, media type) ③.
- **Bid collection.** The SSP's role is to manage the publisher's ad inventory and put it up for auction at an AdX ④. The AdX notifies the DSP of the available ad inventory by sending a bid request ⑤, which is composed of information from the ad slot as well as any identifiers from the browser (*e.g.,* via cookie

[1] It is noteworthy that in RTB only exposes the winning bid and the corresponding winner in the auction. Moreover, even when a given advertiser wins the auction, RTB does not expose the highest bid due to its use of the second-price auction. Thus, RTB does not allow us to observe the bid placed by a given advertiser.

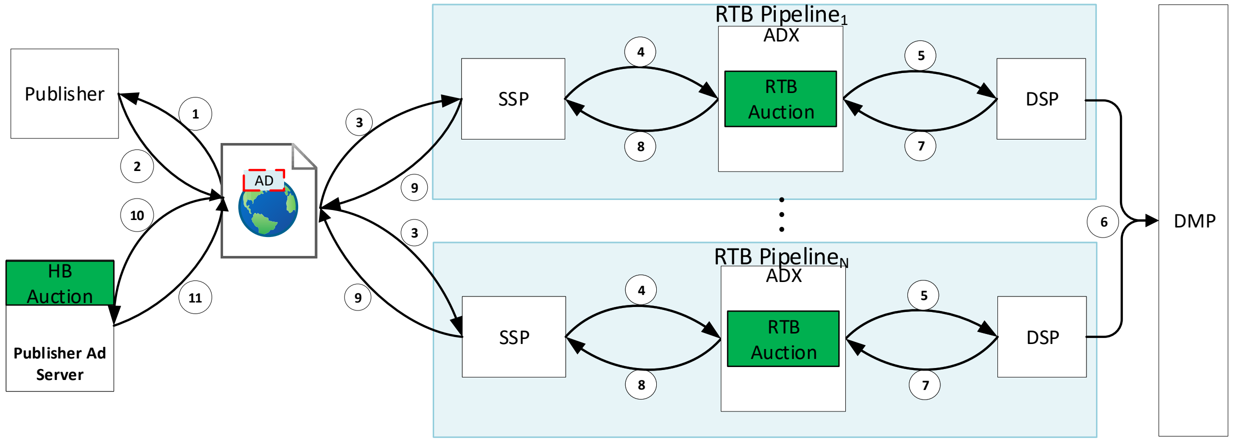

Fig. 1. Online advertising workflow using Real-Time Bidding (RTB) and Header Bidding (HB).

syncing [18, 67]). The DSP evaluates the bid request using the information sent by the AdX and by synchronizing any identifiers with one or more DMPs ⑥. The DSP then acts on behalf of an advertiser by generating and sending the AdX a bid ⑦. DSPs typically implement sophisticated bidding strategies that leverage campaign information from the advertiser, tracking information from the DMP(s), and ad slot information in the bid request.

- **Ad placement.** The AdX collects bid responses from multiple DSPs and uses an auction mechanism (typically a second price auction) to determine the winning bid. If the winning bid value surpasses the impression's minimum sale price set by the publisher, the winning bid and the associated ad is forwarded to the SSP ⑧, which places the ad on a browser page ⑨. If the winning bid value does not surpass the impression's minimum sale price, the impression is presented to the next preferred AdX (as determined by the SSP) and the bidding process is repeated. Auctions occurring at lower levels of the "waterfall" have a residual effect on bidder perception, resulting in progressively lower bids.

Header Bidding (HB). HB is an emerging programmatic process for online advertising that is rapidly gaining popularity due to its promise to increase yield for publishers as compared to traditional RTB [43, 68]. According to a recent survey [49], more than half of the top one thousand websites that offer programmatic advertising already use HB. In contrast to RTB, where the ad inventory is offered to different ad exchanges (and consequently bidders) in a sequential (or waterfall) manner, HB offers the ad inventory to multiple bidders simultaneously. More specifically, in the waterfall model used by RTB, the ad inventory is first offered to higher tier ad exchanges and any leftover inventory is offered to lower tier exchanges. This sequential process results in less competition for bids and subsequently reduces publisher yield from advertising. HB essentially flattens the waterfall, forcing increased competition among different bidders for ad impressions and increasing the yield for publishers. While there is some overlap between RTB and HB, we explain the workflow of the HB model by discussing differences that occur in the ad request (steps ② and ③) and ad placement (⑨ – ⑪) phases. The bid collection process (④ – ⑧) remains unchanged in HB.

- **Ad request.** In HB, the publisher's web server responds with the HTML document that contains page content as well as the ad tag with a HB *wrapper*.[2] The HB wrapper pauses a page's ad tag from being executed and sets a predetermined timeout. While the ad tag is paused, the wrapper simultaneously contacts different demand partners (mainly SSPs) by sending them bid requests ③. While the HB wrapper is awaiting bid responses, parallel auctions are occurring in multiple RTB pipelines as shown in Figure 1.
- **Ad placement.** Each SSP asynchronously sends bid responses to the HB wrapper ⑨. Once the HB timeout expires, all bids are then forwarded to the publisher's ad server ⑩ where a unified HB auction mechanism is used to determine the winning bid and

[2] There are two common implementations of HB. We are discussing client-side HB as opposed to server-to-server HB [7].

price. The browser is notified of the winning bid and the corresponding ad is placed on the page (11).

2.2 Online Tracking Ecosystem

The online tracking ecosystem is composed of a large number of organizations engaging in tracking user behavior across the web. This is accomplished by a variety of techniques including tracking cookies, pixel tags, beacons, and other sophisticated mechanisms. Below we provide a high-level overview of how online tracking works and some aspects of the interplay between online tracking and online advertising.

How online tracking works. In order to deterministically identify users across the web, trackers need to assign unique identifiers to individual users. This is often accomplished using cookies. Cookies are stored at the client-side and are typically structured as key-value pairs that contain identifiers that uniquely identify a user. A tracker present, as a third-party, on multiple domains across the web can read their own cookies linked to a user as they traverse domains. This enables individual trackers to recreate subsets of a user's browsing history.

There are two key limitations of cookies from the perspective of online tracking. First, due to the same-origin policy enforced by browsers, access to a cookie is restricted to the tracker that sets it. This means that two trackers, each with a partial view of a user's browsing history, cannot enhance their knowledge by directly sharing their own cookies with each other. To circumvent this limitation, trackers typically rely on *cookie syncing* in order to map each other's identifiers of a user [67]. Second, since they are stored at the client-side (browser), cookies can be deleted by users. While trackers can always set new cookies, there is still no sound way of linking deleted cookies with new cookies. To circumvent this limitation, trackers now also rely on *cookie respawning* [32] and other stateless (probabilistic) techniques such as *browser fingerprinting* [45].

Tracker relationships. In order to generate a more complete view of a user's browsing history and interests, trackers may collaborate with one another. This is accomplished by using client-side or server-side mechanisms to share information. Client-side mechanisms rely on the user's browser to facilitate an information sharing channel between the collaborating trackers. As a result, these data sharing relationships are directly observable at the client-side. Cookie syncing is a popular client-side mechanism that is used by trackers to facilitate cookie sharing between trackers even in the presence of the same-origin policy. Other client-side mechanisms involve the sharing of other (non-cookie) unique identifiers such as email addresses and unique device identifiers (*e.g.,* IMEI, Android ID, *etc.*) [50, 76]. Server-side mechanisms may rely on an out-of-band information sharing channel between collaborating trackers. Since the user's browser is not involved in the mechanism, these data sharing relationships are not directly observable at the client. Instead, more complex controlled experiments are needed to infer such relationships [35, 36].

Synergy between online advertising and tracking. Two common strategies used by online advertisers for targeting users are contextual targeting and behavioral targeting. In contextual targeting, ads shown to a user are only dependent on the content of the page (or website) being viewed. In contrast, behavioral advertising shows ads that are based on the interests and behavior demonstrated by the user. In recent years, advertisers have started to increasingly rely on behavioral advertising. In fact, just between 2008 and 2018, the ad spend on behavioral ads in the United States increased from $775M to $47B [49]. While many in the online advertising industry claim that behavioral advertising is always more effective that contextual advertising, its effectiveness relies on the quantity and quality of data available about the targeted individual. As a direct consequence, user data obtained from online trackers is vital to the success of the advertising industry. Put another way, data obtained by online tracking is deemed critical to improving the click-through rate (CTR) and return on investment (ROI) in online advertising campaigns [8]. Data management platforms (DMPs), shown in Figure 1, are responsible for feeding user data obtained by trackers into the advertising ecosystem bidding process.

3 Quantifying the Value of Users

In this section, we seek to understand how much advertisers are *willing to pay* to reach different users. Prior work has attempted to understand how much advertisers *pay* to reach different users [53, 67, 71]. There is a subtle but important difference between our and prior work. Prior work is limited to studying the price *actually paid* by only the winning bidder in RTB. Specifically, prior work leveraged the winning price notifications in RTB, which only exposes the winning bid at the client-side. First, note that the winning price is actually *not* the bid value of the winner but rather the bid value of

Question	Results
How much does a user's persona impact bid values?	§3.2.1
How much does user intent matter to advertisers?	§3.2.2
How does bidding behavior vary across advertisers?	§3.2.3
How much do advertisers pay to reach users?	§3.2.4
How common is underbidding?	§3.2.5

Table 1. Questions answered by our study. Questions in **bold** have not been answered by previous work.

the second-highest bidder plus a predetermined amount (typically one cent) because RTB uses the second-price auction. Second, note that RTB's winning price notification does not include any information about the bidders (or their bids) that did not win the auction. Third, note that RTB's winning price notifications often encrypt the winning bid, which prior work [67] assumed (incorrectly [71]) to be the same as plaintext bids. In this section, we leverage HB to empirically understand how much advertisers are *willing to pay* to reach different users while avoiding the limitations of prior work based on RTB. The HB process, as explained in §2.1, typically requires that *all* bids made by different bidders are forwarded to the publisher ad server via the client in plaintext. This client-side access to the details of every bid made by advertisers for different user personas, not just winning bids, allows us to improve and extend the analytical insights drawn by previous work [53, 67, 71]. Table 1 illustrates the contributions of our work towards quantifying the value of users to advertisers.

3.1 Measurement Method

To answer the questions listed in Table 1, we conducted controlled experiments using the following method. At a high-level, our measurement method is explained by: (1) how we crawl web pages, (2) how we create web personas which can signal user intent to complete a transaction, (3) how we gather the bids placed by HB participants on HB-enabled websites.

Web crawling. Our measurements were conducted using a lightly modified version of OpenWPM [45]. OpenWPM was used to automatically load selected webpages. The timeout for each page load was set to 60 seconds. In order to more accurately simulate real user behavior, the bot-mitigation features of OpenWPM were enabled and additional scrolling on the webpage was performed 5 seconds after the browser on-load event fired. Randomized delays of 2-7 seconds were implemented between each page scroll.

Personas created	Adult, Art, Business, Computers, Games, Health, Home, Kids, News, Recreation, Reference, Regional, Science, Shopping, Society, Sports.
Intent sites	`hotels.com`, `zales.com`, `jamesedition.com`, and `luxuryrealestate.com`.

Table 2. List of web personas created and sites used to convey transaction intent by our study. Each persona reflects a user browsing the most popular sites in the corresponding Alexa category. A product page was browsed on each intent site to convey transaction intent.

Creating web personas. In order to gather information about how advertisers value different users, we need to construct *personas* mimicking different users. We constructed personas based on each of the 16 categories found on Alexa's top sites by category [16]. Starting with a clean slate (clean client state), we crawled the top 50 sites in each of these categories using the OpenWPM configuration described above, saving associated browser cookies after each site visit. Each web persona is an accumulation of cookies for a single category and is fully constructed when the crawl is complete. No crawling is performed to construct the control persona – its persona is the absence of cookies. Note that our approach to constructing web personas aligns with previous work within this space [35, 67]. Table 2 lists the 16 different personas used in our study.

Signaling intent. Prior work [67] showed that advertiser bids were generally higher for personas which had previously demonstrated intent to make a transaction (*e.g.,* by navigating to a specific product page on a website). We follow the methodology in [67] and select a small number of sites to signal intent. Table 2 lists the 4 sites on which specific products were chosen to demonstrate transaction intent. In order to create personas which signal transaction intent, we repeated the persona construction method detailed above followed by the intent signaling mechanism described here. We constructed *intent* and *no intent* versions of each of our 16 personas.

Gathering advertiser bids. After constructing intent/non-intent personas, we crawled HB-enabled websites to gather advertiser bids for each of our personas. In order to identify HB-enabled websites, we crawled the Alexa top 10K websites and shortlisted the 25 most popular domains (e.g., espn.com, accuweather.com, cnn.com) that support the most well-known open-source implementation of HB called `prebid.js` [15, 31]. This was done by checking the `prebid.js version` attribute in the `prebid.js` client-side API.[3] Domains returning valid responses were marked as HB-enabled. During each visit to these 25 websites, we made a call to the `prebid.js` API's `getBidResponses` method. All bid responses returned by this method reflect the bids placed by advertisers for the persona/intent being tested. The bid responses were saved and formed the basis of our analysis. We repeated this process 10 times for each intent/no-intent persona.

3.2 Results

We now analyze the bids placed by advertisers for the following 33 personas: 16 personas signaling no intent to complete a transaction, 16 personas signaling intent to complete a transaction, and one control persona with no prior browsing history. Next, we use these bids to answer the questions listed in Table 1.

3.2.1 How much does a user's persona impact bid values?

Table 3 shows the median bid values by the five most prevalent bidders for each of our 16 personas and the control persona. The bid values are expressed in *cost-per-mille* (CPM) which reflects the price paid for 1,000 ad impressions. Focusing on the average column, we are able to draw several conclusions about the impact of personas on average bid values. First, we note that the *Control* persona (without any browsing history) attracts lower bids than most of the trained personas. Second, we note that the *Health* persona attracts significantly higher bids – 1.6x the average bid value across all categories. Similarly, bids are significantly above the average across all categories for *Computers*, *Science*, and *Shopping* personas. Third, we note the *Sports* persona attracts the lowest average bids – 0.6x the average bid value across all categories. Similarly, bidders bid significantly below the average across all categories for *Games*, and *Home* personas. Finally, we note significantly high variation in bid values across bidders for the *Health* and *Computer* personas, which also receive significantly higher bids than other personas. Overall, variability in average bid values allows us to conclude that *a user's persona impacts bids placed by an advertiser.*

3.2.2 How much does user intent matter to advertisers?

Table 4 shows the ratio of the median bids placed by the five most common bidders for personas showing intent to those showing no intent. A ratio near 1 means that bid prices remain similar between *Intent* and *No-Intent* personas. Focusing again on the Avg. column, we are able to draw several conclusions about the impact of showing intent on average bid values. First, we note that the *Control* persona (without any browsing history) attracts significantly lower bid ratios than most of the trained personas. It is also notable that bid ratios from the *Control* persona were near 1 – bidder express significantly lower interest regardless of intent. Second, we note that the *Health* persona attracts the highest average bid ratios – 2.4x higher than the *No-Intent Health* persona. A notable point is that the *Health* persona attracts the highest bid values (Table 3) and bid ratios (Table 4) – on average bidders are willing to pay significantly higher prices for *No-Intent Health* personas and even higher prices for *Intent Health* personas. Similarly, bid ratios were significantly above the average across all personas for *Health*, *Kids* and *Sports*. While *No-Intent Sports* personas attract significantly lower than average bids, the *Intent Sports* persona drew significantly higher bids than the average – bidders are willing to pay higher prices for *Intent Sports* personas. Third, we note that the *Intent Science* persona attracts the lowest average bids and nearly the same as the *No-Intent Science* persona. Similarly, bid ratios were below the average across all categories for the *Science*, *Recreation* and *Control* personas. Finally, we note significantly high variation in bid ratios across bidders for the *Health* and *Kids* personas, which also receive significantly higher bid ratios than other personas. Overall, variability in average bid ratios allows us to conclude that both *user intent and persona impact bids placed by an advertiser.*

3.2.3 How does bidding behavior vary across advertisers?

We now turn our focus to uncovering the differences in the behavior of different bidders. Table 3 and Table 4 illustrate the impact of personas and intent on the bidding behaviors of the five most frequently observed bidders – AppNexus, Rubicon, IX, OpenX, and PubMatic. We breakdown our analysis based on bidders responses to modified personas and intent.

Bidder response to different personas (Table 3). First, Rubicon generally bids more per impression than

[3] Our measurements are dependent on client-side implementation of `prebid.js`, which would not work for server-side implementation [7, 68].

	App.	Rub.	IX	OpX	Pub.	Avg.	Std.
Adult	$0.21^{\downarrow}$	$0.43^{\uparrow}$	0.25	0.34	0.33	0.30	0.08
Arts	$0.34^{\uparrow}$	$0.45^{\uparrow}$	$0.29^{\downarrow}$	0.37	0.36	0.36	0.05
Business	0.28	0.45	0.28	0.30	$0.51^{\uparrow}$	0.36	0.10
Computers	0.20	$0.75^{\uparrow}$	0.21	0.55	$0.73^{\uparrow\uparrow}$	$0.45^{\uparrow}$	$0.25^{\uparrow}$
Games	0.21	$0.33^{\uparrow}$	0.21	$0.34^{\uparrow}$	$0.25^{\downarrow}$	$0.26^{\downarrow}$	0.06
Health	0.21	$1.16^{\uparrow\uparrow}$	0.28	$0.94^{\uparrow}$	0.54	$0.59^{\uparrow}$	$0.39^{\uparrow}$
Home	$0.20^{\downarrow}$	$0.39^{\uparrow}$	0.24	$0.28^{\downarrow}$	$0.28^{\downarrow}$	$0.27^{\downarrow}$	0.07
Kids	0.20	0.41	$0.18^{\downarrow\downarrow}$	$0.47^{\uparrow}$	0.30	0.30	0.11
News	0.19	$0.61^{\uparrow}$	0.24	0.41	0.33	0.34	0.16
Recreation	$0.31^{\uparrow}$	$0.67^{\uparrow}$	$0.23^{\downarrow}$	0.36	0.44	0.41	0.16
Reference	$0.18^{\downarrow}$	0.53	0.20	$0.58^{\uparrow\uparrow}$	0.52	0.37	0.18
Regional	0.23	0.43	0.33	$0.73^{\uparrow\uparrow}$	0.35	0.39	0.17
Science	0.30	$0.70^{\uparrow}$	$0.28^{\downarrow}$	0.44	$0.58^{\uparrow}$	$0.46^{\uparrow}$	0.17
Shopping	$0.40^{\downarrow\uparrow}$	0.56	$0.45^{\uparrow}$	$0.60^{\uparrow\uparrow}$	0.47	$0.49^{\uparrow}$	0.07
Society	$0.22^{\downarrow}$	0.41	0.27	$0.45^{\uparrow}$	0.37	0.32	0.09
Sports	0.19	$0.35^{\uparrow}$	$0.13^{\downarrow\downarrow}$	$0.23^{\downarrow}$	0.30	$0.23^{\downarrow}$	0.07
Control	$0.20^{\downarrow}$	$0.26^{\downarrow}$	0.28	$0.44^{\uparrow}$	$0.37^{\uparrow}$	0.29	0.08
Avg.	$0.24^{\downarrow}$	$0.54^{\uparrow}$	0.26	0.45	0.44	**0.36**	**0.39**
Std.	$0.06^{\downarrow}$	$0.21^{\uparrow}$	$0.07^{\downarrow}$	$0.17^{\uparrow}$	$0.14^{\uparrow}$	**0.08**	**0.13**

Table 3. Impact of user personas. HB median CPMs (USD) across our 16 personas and control persona for the top five bidders (AppNexus, Rubicon, IX, OpenX, and PubMatic) and associated weighted Avg. and Std. among categories and bidders. Bid prices exceeding $\pm\sigma$ among categories are denoted with $^{\uparrow}$ or $^{\downarrow}$. Bid prices exceeding $\pm\sigma$ among bidders are denoted with $^{\uparrow}$ or $^{\downarrow}$. Avg. and Std. among *persona* weighted averages are in **bold red**. Avg. and Std. among *bidder* weighted averages are in **bold black**.

	App.	Rub.	IX	OpX	Pub.	Avg.	Std.
Adult	0.97	0.97	$2.10^{\uparrow}$	0.85	0.95	1.13	0.44
Arts	$1.04^{\downarrow}$	$1.48^{\uparrow}$	1.45	$0.97^{\downarrow}$	1.32	1.26	0.21
Business	1.02	1.09	$2.66^{\uparrow}$	1.01	0.84	1.32	0.66
Computers	1.06	1.19	$2.38^{\uparrow}$	1.18	$0.71^{\downarrow\downarrow}$	1.30	0.55
Games	1.20	1.83	1.81	$1.06^{\downarrow}$	$1.80^{\uparrow}$	1.53	0.34
Health	$1.85^{\uparrow}$	1.24	1.34	$5.96^{\uparrow\uparrow}$	1.21	$2.42^{\uparrow}$	$1.92^{\uparrow}$
Home	1.31	$0.92^{\downarrow}$	$1.50^{\uparrow}$	1.12	1.21	1.19	0.19
Kids	1.31	1.51	$6.00^{\uparrow\uparrow}$	0.76	$1.49^{\uparrow}$	$2.34^{\uparrow}$	$1.99^{\uparrow}$
News	1.05	1.14	$3.57^{\uparrow}$	1.05	0.95	1.62	1.06
Recreation	$1.76^{\uparrow\uparrow}$	1.09	1.04	1.08	0.86	1.15	0.29
Reference	1.01	1.06	$2.80^{\uparrow}$	0.70	$0.60^{\downarrow}$	1.26	0.82
Regional	1.04	$2.24^{\uparrow\uparrow}$	1.46	0.96	0.83	1.35	0.54
Science	1.11	1.02	$0.81^{\downarrow\downarrow}$	$1.12^{\uparrow}$	0.92	$0.99^{\downarrow}$	$0.12^{\downarrow}$
Shopping	1.18	1.42	1.55	1.52	$1.00^{\downarrow}$	1.35	0.21
Society	1.30	$2.15^{\uparrow}$	$2.52^{\uparrow}$	$0.76^{\downarrow}$	0.92	1.51	0.69
Sports	$1.13^{\downarrow}$	$3.00^{\uparrow}$	$3.69^{\uparrow\uparrow}$	$2.85^{\uparrow}$	$1.57^{\uparrow}$	$2.43^{\uparrow}$	0.94
Control	$0.87^{\downarrow}$	$1.32^{\uparrow}$	$1.33^{\uparrow}$	$0.60^{\downarrow}$	0.92	$1.01^{\downarrow}$	0.28
Avg.	1.19	1.45	$2.33^{\uparrow}$	1.07	1.40	**1.48**	**1.49**
Std.	$0.25^{\downarrow}$	$0.54^{\downarrow}$	$1.35^{\uparrow}$	$0.31^{\downarrow}$	$1.26^{\uparrow}$	**0.45**	**0.74**

Table 4. Impact of showing intent. Cells indicate the ratio of median bid values for personas showing intent vs. personas showing no intent for the top five bidders (AppNexus, Rubicon, IX, OpenX, and PubMatic) and associated weighted Avg. and Std. Ratios exceeding $\pm\sigma$ among categories are denoted with $^{\uparrow}$ or $^{\downarrow}$. Ratios exceeding $\pm\sigma$ among bidders are denoted with $^{\uparrow}$ or $^{\downarrow}$. Avg. and Std. among *persona* weighted averages are in **bold red**. Avg. and Std. among *bidder* weighted averages are in **bold black**.

any other advertiser (0.54 USD CPM 1.4x above the average bidder) – regardless of persona. In fact, Rubicon bids the highest average values for 9 of the 16 personas. Second, the *Health*, *Shopping* and *Computers* categories generally attract the most interest from all the bidders. We also find that some bidders show an aversion towards certain personas (*e.g., IX - Sports*, *Kids*) bidding even less than they did for the control persona which had no history attached to it. Finally, we see that OpenX bids significantly more per impression than any other advertiser to place ads in front of our control persona (0.44 USD CPM). At a high-level, our results allow us to conclude that *different bidders have preferences and aversions for different personas and only a few personas are universally preferred.*

Bidder response to demonstrated intent (Table 4). First, we see that while all of our bidders generally had positive responses to intent. The bid ratio for IX is significantly more than other bidders (1.6x more than the average). In fact, IX had the highest intent to no-intent ratio for 10 of our 16 personas. Conversely, PubMatic was found to be the least reactive to intent with their average bid value increasing only by 1.07x. Second, some bidders had increases of nearly 6X in bid values when certain personas demonstrated intent. In particular, OpenX showed a 5.96x increase in their bid values when confronted with an intent *Health* persona. Similarly, IX showed a 6.00x increase in their bid values when intent was demonstrated by the *Kids* persona. Finally, looking at responses to our intent control persona, we see bid increases only for Rubicon and IX (1.32x and

	App.	Rub.	IX	OpX	Pub.	Avg.	Std.
Adult	2.28↑	0.86	2.87	0.72	5.94↑	2.04	1.51
Arts	1.19	0.86	1.08↓	0.94	10.8↑↑	1.77	2.61↑
Business	2.14↑↑	0.86↓	1.76	1.84	-	1.48	0.54↓
Computers	1.06↓	2.81	2.72	0.34↓	1.50↓	2.20	0.86
Games	0.53↓	1.92	1.26	4.48↑	2.58↑↓	1.62	0.95
Health	2.65↑	3.83↑	3.00↑	1.47↓	9.76↑↑	3.99↑	2.43↑
Home	0.62↓	2.50	2.79	-	-	2.19	0.86
Kids	2.84↑	0.86	0.30↓↓	2.02	5.94↑	1.74	1.33
News	0.50↓	2.92↑↑	1.17	0.77	4.86↑	1.64	1.18
Recreation	0.45↓↓	0.86	2.71	4.02↑	-	1.64	1.15
Reference	1.09	0.86	1.39	2.73↑	3.56↑	1.34	0.72
Regional	0.91	1.56	0.81↓	7.73↑↑	-	2.32	2.72↑
Science	2.08↑	3.56↑↑	3.09↑	1.60↓	-	2.69↑	0.86
Shopping	0.45↓↓	1.58	4.87↑↑	0.88	5.94↑	2.56	2.02
Society	1.73	4.00↑↑	1.81	0.94↓↓	3.29↑	2.25	0.96
Sports	1.03	0.86	2.29↑	0.28↓	-	1.12↓	0.57↓
Control	0.62	0.86	2.75↑	0.72	-	1.38	0.94
Avg.	1.27	1.78	2.01	2.25	6.22↑	**2.00**	**2.71**
Std.	0.76↓	1.08↓	0.91↓	2.23↑	3.06↑	**0.66**	**1.61**

Table 5. Impact of user personas on winning bids. HB median CPMs (USD) across our 16 personas and control persona for the top five bidders (AppNexus, Rubicon, IX, OpenX, and PubMatic) and associated weighted Avg. and Std. among categories and bidders. Bid prices exceeding $\pm\sigma$ among categories are denoted with ↑ or ↓. Bid prices exceeding $\pm\sigma$ among bidders are denoted with ↑ or ↓. Avg. and Std. among *persona* weighted averages are in **bold red**. Avg. and Std. among *bidder* weighted averages are in **bold black**.

1.33x increase) while the average showed only a 1.01x increase. This shows that, in general, for many bidders, the knowledge of user personas dominates the decision to increase bid values and intent is only used to decide the magnitude of this increase. At a high-level, our results allows us to conclude that *bidders rarely have similar responses to demonstrated user intent.*

3.2.4 How much do advertisers pay to reach users?

We now turn our attention to understanding *the price advertisers actually pay to reach users.* To answer this question, we first examine the subset of winning bids (*i.e.,* highest bid value in the first-price HB auction) by the five most common bidders for *No Intent* personas as shown in Table 5. First, we note that on average bidders pay $2.00 USD CPM across all personas in order to serve ads – 5.5x the average of the corresponding bid price in Table 5. We see this trend across the board for different personas and bidders. We conclude that bidders have to pay substantially higher prices than their average bids to win the auctions. Second, we note that the average winning bid in HB is 3.4x the average winning bid of RTB (Table XI - Only category column in [67]). There are two main explanations for this difference: (1) auction type (HB typically uses first-price auction and RTB typically uses second-price auction); and (2) bidding structure (HB uses a flattened model to issue bid requests and RTB uses a tiered/waterfall model where bid requests received at lower tiers are interpreted by bidders as bid “left-overs”). Third, we observe some similarities and differences in winning bid trends across personas for HB and RTB [67]. For the *Health* persona, we observe above average winning bids in both RTB and HB. For other personas such as *Games* and *Sports*, we observe a shift from higher than average bids in RTB to lower than average bids in HB. This shift could be due to differences in bidder affinity caused by changing preferences among advertising partners [24, 74] or time/location of measurements [67, 71].

3.2.5 How common is underbidding?

During our bid collection process, we observed many bidders making *zero* bids – *i.e.,* a bid of $0 USD CPM for the impression. There are several reasons for these bids. First, incorrect configurations of HB can lead to zero bids by advertisers. For example, price granularity is a setting made available to publishers which in essence can enforce a minimum bid value. Any bid received below this value is rounded down to zero. Advertisers making bids without correctly accounting for this parameter will generate zero bids. Second, and more interestingly, zero bidding is a form of underbidding – *i.e.,* purposely making low bids with the motivation to gain access to user data (e.g., synced cookie [46, 69]) associated with the impression rather than to win the auction. Although most exchanges which facilitate RTB auctions typically ban such behaviour and enforce mandatory minimum bidding participation, we find no such enforcement in HB auctions. Bids gathered from our experiments allow us to measure the frequency of zero bids, yet they do not let us convincingly distinguish whether they are due to misconfiguration or nefarious intent.

Table 6 shows the fraction of zero bids received for our intent and no-intent persona for each of the top 20

Bidder	Percentage of zero bids		
	No-intent personas	Intent personas	Total
PubMatic*	68.75	66.37	67.70
AppNexus	0.32	0.26	0.29
IX*	19.14	6.19	13.53
Rubicon*	3.54	2.27	2.94
OpenX*	1.04	0.23	0.66
Criteo*	5.83	1.60	3.75
Aol	8.71	9.17	8.90
Sovrn*	10.43	3.87	7.10
Districtm	0.00	0.00	0.00
Conversant	0.00	0.00	0.00
Total (all bidders)	24.21	19.44	22.07

Table 6. The percentage of zero bids made, for intent and no-intent personas, by our 10 most commonly observed bidders. Bidders are sorted by total number of bids placed. Bidders appended with'*' indicate that zero bidding behavior for a bidder was significantly impacted by *Intent* personas.

bidders observed. First, we note that zero bidding is a common occurrence and they make up over 22% of all bids. It is noteworthy that for two of the bidders – PubMatic and Innity – most of their bids are zero bids and together account for a vast majority of all zero bids observed in our measurements. Frequent underbidding observed for these two bidders is indicative of the absence of minimum bidding performance requirements in HB through contract guarantees, which are often enforced in RTB auctions [46]. Second, we assess whether a bidder's likelihood of placing zero bids is *significantly* impacted by personas demonstrating intent to make a transaction. To this end, we apply the chi-square proportions test [38] to compare percentages of zero bids placed by a bidder for *No-Intent* and *Intent* personas. We see a statistically significant (4.77%) decrease in the percentage of zero bids when a persona shows intent to make a transaction. This suggests that bidders are more motivated to make positive bids when the user communicates transaction intent. Finally, we can hypothesize a bidder's motivation to place zero bids by comparing percentage of zero bids between *No-Intent* and *Intent* personas. We suspect that intentional underbidding will lead to a statistically significant difference in the prevalence of zero bids between *Intent* and *No-Intent* personas. We observe a significant decrease in zero bids for PubMatic, IX, Rubicon, OpenX, Criteo, Sovrn, which leads us to suspect that zero bids are unlikely to be caused as a result of configuration errors.

4 Inferring Tracker-Advertiser Relationships

We showed that an advertiser's assessment of the value of a user (*i.e.,* bids) is highly dependent on the available information (*i.e.,* browsing history). To get relevant user information, advertisers (or DSPs bidding on behalf of advertisers) gather information from different trackers (or DMPs in general) through client-side or server-side mechanisms. Inferring such data sharing relationships between different entities in the online advertising ecosystem is an important problem. It is challenging to infer such data sharing relationships, particularly at the server-side because they are not directly observable at the client-side. Next, we present our approach to infer tracker-advertiser data sharing relationships at the client-side or the server-side.

4.1 Proposed Approach

Our approach, named Kashf, leverages the information provided by HB to infer data sharing relationships between trackers and advertisers. Our key insight is that an advertiser's bids for a persona will change when it has an information flow originating from *some* tracker which has seen the persona before. This insight allows us to use bids, which are observable at client-side in HB, as a proxy for the existence of information flows (*i.e.,* data sharing relationship) between a tracker and a bidder. Thus, through careful manipulation of tracker exposure while constructing personas, we can analyze an advertiser's bids to make inferences about its data sharing relationships with the exposed trackers.

We illustrate Kashf with a simplified model showing information flows among key entities in the HB and the associated tracking ecosystem in Figure 2. We start with edge ①, from the client to a tracker. This edge denotes data being gathered from clients by trackers partnering with publishers. Notice that these edges can be observed and manipulated at the client – *i.e.,* trackers can be identified and blocked at the client and thus these edges can be deleted. Edge ② denotes the flow of data from trackers to advertisers. It is the presence of these flows that impacts the advertiser's bids for a user. Unfortunately, these edges, which are crucial for verifying regulatory compliance and building effective privacy-enhancing tools, are not observable or manip-

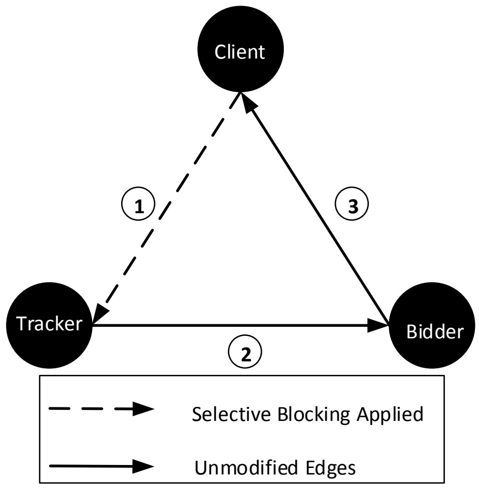

Fig. 2. A simplified model of the information flows between trackers and bidders.

ulated by the client. Finally, edge ③ denotes the bid sent by an advertiser to a client in HB. Notice that these edges are observable by the client in HB. *Our goal is to infer the existence of edge ② given the ability to observe and manipulate edges ① and ③*. We do this as follows:

- We gather a large number of bids from different advertisers (edge ③) while exposing client personas to a select set of trackers (by selectively deleting edge ①).
- We then train a machine learning model to predict the bid values placed by each advertiser based on tracker presence/absence as features. We argue that a model that is able to accurately predict bid values also uncovers data sharing relationships between advertisers and trackers. An accurate model will evaluate tracker presence to gauge their impact on bid values. Put another way, if a machine learning model given edge ① as features is able to predict the values of edge ③, then it must have automatically inferred the presence or absence of edge ②.
- Next, we analyze our interpretable machine learning models to identify the features (*i.e.,* edge ①) which had the most impact on our trained model's bid predictions. The information gain of these features establishes the likelihood of a relationship between the tracker and advertiser (*i.e.,* edge ②). Put another way, trackers that have a high impact on our model for a particular advertiser are more likely to have a data sharing relationship with the advertiser than those having no impact on our model. *After all, under the assumption of one tracker per page, if deleting a tracker edge consistently has no impact on the bid values from the advertiser, it must be true that there is no relationship between the tracker and advertiser.* As we discuss next, our method can generalize to multiple trackers.

Question	Results
Is tracker presence a predictor for advertiser bids?	§4.3.1
Which trackers influence the behaviors of which advertisers?	§4.3.2

Table 7. Questions answered by our study. Questions in **bold** have not been answered by previous work.

4.2 Measurement Method

Table 7 illustrates the contributions of this work towards inferring advertiser-tracker data sharing relationships. To answer the questions listed in Table 7, we conducted controlled measurements as follows. At a high-level, our method is explained by: (1) how we selectively expose trackers to information about our personas (how we manipulate edge ①); (2) how we measure the bids made by advertisers for each of our personas (how we observe edge ③); (3) how we predict bids; and (4) how we identify and validate influence of a tracker on an advertiser.

Exposing user personas to trackers (manipulating edge ①). We constructed 10,000 user personas, each exposing selective characteristics to some subset of trackers, using the following approach.

- *Tracker exposure.* We used EasyList [10] and EasyPrivacy [11] in combination with the outcome of the most recent Alexa top 1-million site crawl [45] to obtain the top 20 most frequently observed tracking organizations and the tracking domains owned by them.[4] We then randomly selected one organization and blocked all their trackers when building a user persona. Specifically, trackers from the selected organization were blocked during the crawling of persona and intent sites described below.
- *Persona selection.* We first randomly selected a persona which for the user persona to mimic. We used the same approach described in §3.1, with the caveat that user persona was only constructed from a random subset of 1-10 of the Alexa top-50 sites

[4] These 20 organizations are: Adobe, Alibaba, Alphabet, AppNexus, Automattic, Baidu, Comscore, Criteo, DoubleVerify, ExoClick, Facebook, Integral Ad Science, Microsoft, Oracle, PubMatic, Quantcast, Sovrn, Twitter, Verizon, and Yandex.

within a persona category (rather than all of the top 50 sites). This reduction was necessary to scale of our experiments to build 10,000 user personas.

- *Intent selection.* Finally, we randomly assigned some of our personas to demonstrate intent to complete an online transaction. The method used was identical to the intent signaling mechanism described in §3.1.

After building each user persona, we waited at least 90 minutes before moving into the bid collection phase.

Measuring advertiser bids (recording edge ③). In order to measure the impact of selectively blocking edge ①, we needed to measure the values obtained through edge ③. This was accomplished by visiting *one* HB-enabled site using each trained persona and gathering the bids made by advertisers. We limited bid gathering to only one site since visiting multiple sites could result in tracker flows from the first site influencing the bids measured on subsequent sites. Visiting a single site allows us to ensure that any tracker-advertiser information flow originates during the persona building phase and not during the bid collection.

Predicting bids. Since our recorded bid values are continuous, we need a method to discretize them. Our bid values were discretized, in a similar manner as previous work seeking to predict encrypted bid values [71], by dividing bid values into classes. Specifically, we divided our dataset of bids values into three classes with the following bid ranges: $[-\infty, \mu - \sigma)$ *low*, $[\mu - \sigma, \mu + \sigma]$ *medium*, and $(\mu + \sigma, +\infty]$ *high*, where μ is the mean bid value and σ is the standard deviation of bid values. Next, we trained a separate Random Forest classifier for each advertiser with the goal of predicting bid classes given the presence of trackers as features. The Random Forest classifier was explicitly chosen due to its interpretable decision tree classification model. We applied 10-fold cross-validation to validate the accuracy of our constructed models. An accurate model for an advertiser demonstrates that tracker presence is a good predictor for bid class.

Validating tracker influence on an advertiser. We want to rank trackers based on their influence on advertiser generated bids. The decision trees produced by the Random Forest classifier rank features based on their importance. The most influential feature, with the highest information gain, is the root node of the tree. The subsequent nodes at lower levels have decreasing information gain on the partitioned data. Thus, given a reasonably accurate model of advertiser's bidding behavior, we analyzed decision trees to obtain a list of trackers ranked by their influence on each advertiser. We then validated this list by comparing the observed relationships with the following sources of known tracker-advertiser relationships:

- *External databases.* We manually searched a variety of sources (*e.g.,* Crunchbase, public company websites, ad-tech blogs, *etc.*) to obtain publicly disclosed advertiser-tracker relationships.
- *Client-side cookie syncing.* The entities in online advertising ecosystem use the *cookie syncing* mechanism to share user identifiers at the client-side while circumventing the browser's same-origin policy. Using the heuristic presented in [69], we detect client-side cookie syncing by looking for identifiers in the URL and the referrer field during our measurements.

4.3 Results

4.3.1 Can tracker presence be used to predict bids?

We now evaluate whether tracker presence can be used as predictors of advertiser bids. Table 8 presents the classification performance of trained machine learning models for the top-5 bidders in our dataset. We note that trained machine learning models can predict bids by different bidders with reasonable accuracy. Specifically, the accuracy ranges from 75% for AppNexus to 83% for IX. It is noteworthy that our trained machine learning models provide comparable accuracy to prior work on predicting encrypted bid values in RTB (82%) [71]. Thus, we conclude that our trained machine learning models can leverage *tracker presence to accurately predict bids by different advertisers.*

Bidders	Accuracy
AppNexus	75%
IX	83%
Openx	81%
Rubicon	82%
PubMatic	78%
Avg.	80%

Table 8. Bid prediction accuracy of machine learning models for top-5 bidders in our dataset.

4.3.2 Tracker Influence On Bidder Behavior

To understand a tracker's influence on an advertiser's bidding behavior, we use the decision tree model generated by our machine learning classifier. As discussed earlier, trackers at the top of the decision tree are more influential than those at the bottom. Table 9 lists the top-3 trackers for each of the top-5 bidders in our dataset. We note that different trackers are the most influential across different bidders. For example, our model ranks DoubleVerify as the most influential tracker for AppNexus while Alphabet as the most influential tracker for PubMatic. We observe that 11 of 15 advertiser-tracker relationships inferred by our model are validated by external databases (10 of 15) or client-side cookie syncing (4 of 15). 3 of these advertiser-tracker relationships are validated by both external databases and client-side cookie syncing. We note 11 potential server-side advertiser-tracker relationships that are not validated using client-side cookie syncing. Of these 11, we are able to validate 7 such server-side relationships using external databases. The remaining 4 may be attributed to previously unknown server-side data sharing relationships, imperfect heuristics to detect cookie syncing, or erroneous inferences by KASHF.

	Tracker 1	Tracker 2	Tracker 3
AppNexus	DoubleVerify [2]	Automattic [19]	Comscore CS,[1]
IX	Sovrn [2]	PubMatic [28]	DoubleVerify [28]
OpenX	Microsoft [25]	AppNexus CS,[17]	Criteo [5]
Rubicon	Verizon CS,[6]	DoubleVerify	Facebook
PubMatic	Alphabet CS	Twitter	Microsoft

Table 9. Tracker influence is ranked in the descending order of information gain for each of the top-5 bidders in our data set. The bidder-tracker relationships that we are able to validate using manual search are marked with a citation. Cookie syncing detected using client-side analysis are marked with ^{CS}.

4.3.3 Implications

It is noteworthy that KASHF is able to uncover several server-side advertiser-tracker relationships that are not observable at the client-side. Our findings seem to indicate that online advertising and tracking ecosystems may be shifting from the client-side to the server-side. We argue that there are several motivations for such a shift from client-side to server-side. First, and perhaps most importantly, advertisers and trackers are shifting to server-side data sharing to circumvent client-side blocking tools. Specifically, a significant fraction of users have installed browser extensions (*e.g.,* uBlock Origin [56], Adblock Plus [29], Ghostery [22], Privacy Badger [26], *etc.*) to block ads and trackers at the client-side. Moreover, mainstream browsers such as Safari and Firefox have enabled anti-tracking protections by default [66, 78]. We believe that advertisers and trackers are likely shifting to server-side data sharing to circumvent client-side blocking mechanisms. Second, advertisers and trackers also prefer server-side implementations due to performance reasons. Specifically, client-side implementation of resource-heavy advertising and tracking logic significantly degrades page load performance [7, 51]. Moreover, client-side implementations are also susceptible to slow response times resulting in auction timeouts (bids arriving after an auction timeout occurs are ignored) [7]. To conclude, our findings highlight the shift from client-side to server-side data sharing in the online advertising ecosystem. As server-side data sharing—which can be inferred by KASHF—becomes more prevalent, it is unclear whether the current generation of client-side blocking tools would continue to remain effective.

4.4 Limitations

Completeness issues. Our study makes two simplifying assumptions that may impact the completeness of our results. First, we restrict our inferences to only include bidder relationships with the top-20 tracking organizations. As a result, we are unable to draw inferences about bidder relationships with smaller tracking services (rank $>$ 20). We argue that, given the extreme skew in tracker coverage across the web [45], our approach would capture the overwhelming majority of data sharing occurring in the advertising and tracking ecosystem. Second, the data sharing relationships in the online advertising ecosystem may be indirect. More specifically, trackers may share data with many non-bidder entities (e.g., SSP, AdX) and bidders may gather data from different data sources (DMP). Our approach is unable to determine whether a tracker-bidder relationship is direct or indirect (*i.e.,* involves other intermediaries). However, as long as the presence of a tracker impacts the bids, our approach is able to infer that there is a direct or indirect tracker-bidder relationship.

Correctness issues. Our study also makes several simplifying assumptions that may impact the correctness of our results. First, the accuracy of our machine learning approach is not perfect. It is possible that some of the tracker-bidder relationship inferences based on our trained machine learning models are incorrect. To overcome this limitation, we take a conservative approach by limiting ourselves to top-3 trackers identified by our machine learning models. As part of our future work, we plan to investigate automated methods to determine the optimal cutoff point given a certain error tolerance. Second, our approach may fail in the presence of tracker-tracker data sharing relationships. Consider an example where the following data sharing relationships are observed: (T_1, T_2), (T_2, A) where T_1 and T_2 are trackers and A is an advertiser. Our technique might conclude that there is no relationship between T_2 and A as a consequence of not observing a change in bidding behavior from A when blocking T_2. However, this conclusion might be incorrect if the reason for no change in A's behavior is the flow of information from T_1 to A via T_2. We mitigate this problem in our work by analyzing trackers at the organizational level. In other words, we assume that all domains within an organization (*e.g.,* doubleclick.net and google-analytics.com belong to Alphabet) *will* share data with each other.

5 Related Work

Prior work related to our research can be categorized into two types: (1) user value quantification and (2) characterization of entities and their relationships in the online advertising and tracking ecosystem.

5.1 Quantifying the Value of a User

As more advertisers rely on online advertising [55] and more Internet users have the expectation of free services [44, 62, 77], online behavioral advertising, facilitated by tracker gathered user data, has become the dominant monetization model on the web. Much of prior work has sought to uncover the value of different types of users (and their data) to different entities in the online advertising and tracking ecosystem. Along these lines, there has been a great deal of interest in understanding how much value users place in the data that they trade for free access to online services. These studies have generally borrowed techniques from psychology and economics to design experiments to implicitly uncover the value that users place on their data. Findings have shown that context dictates privacy valuations of data [33, 54, 57], trustworthiness and intention of the buyer plays a role in privacy valuations [41, 42], and there is a mismatch in the actual and perceived value of user data [39, 53, 70]. From another perspective, there have been efforts [67, 70, 71] to quantify how much user data is worth to advertisers. Such efforts are generally more challenging due to the opacity of the advertising ecosystem – *i.e.,* it is difficult to uncover exactly how much advertisers are paying (bidding) to place ads in front of specific users. These works leveraged the visibility afforded to the user's browser in the RTB auction to uncover the winning bids. More specifically, these works leveraged the fact that the winning bid notification in an RTB auction (including information about the winner and the winning bid value) is relayed to the browser in step ⑨ of the RTB workflow shown in Figure 1.

In a seminal work, Olejnik *et al.*[67] analyzed RTB winning bid notifications to understand variation across different user personas based on their location, time, and browsing history. The authors reported that advertisers pay as little as $0.0005 per site visit. Further, they showed that the prices that advertisers pay vary based on browsing histories reflecting different generic interests (*e.g.,* , games, news, shopping) and specific intents (*e.g.,* , hotel booking, jewelry, electronics). Our work differs and builds upon this work in the following ways.

- First, we are able to shed light on the bidding behavior of different advertisers as we are able to capture bids from each advertiser, not just the winning bid.
- Second, they encountered and ignored encrypted bids, which were (incorrectly [71]) assumed to be comparable to plain text bids. In contrast, we do not encounter encrypted bids in our HB measurements.
- Third, because we can observe bids from different advertisers in HB, we are able to show that advertisers bid differently (by as much as 5.5x for *Health* category in Table 4) for the same user personas.
- Further, through controlled experiments in the second phase of our study, we are able to show that bid variations across different advertisers are, in part, due to differences in advertiser-tracker relationships.

In a follow-up work, Papadopoulos *et al.*[71] addressed the limitation placed by encrypted bid values (encountered by [67]) by developing a machine learning approach to infer values of encrypted winning bids with 82% accuracy. The authors showed that encrypted bids are are 1.7X higher than bids sent in the clear. We build on this work by seeking to understand tracker-advertiser relationships using a similar machine learning approach for modeling bid values. However, unlike their work, we are not interested in predicting encrypted bid values because we do not encounter encrypted bids in HB. Instead, we leverage a machine learning model that can accurately predict an advertiser's bids based on information about presence/absence of trackers to infer tracker-advertiser relationships.

5.2 Characterization of Advertising and Tracking Entities and Relationships

There have been many studies which have sought to measure the prevalence of different entities in the advertising and tracking ecosystem. These include large-scale and longitudinal crawls measuring the prevalence of trackers and different tracking techniques on the web [45, 60, 63, 64, 69], mobile [34, 40, 59, 72, 73, 75], and across multiple platforms [37, 79].

Notably, Englehardt and Narayanan [45] studied the prevalence of different stateful and stateless techniques on the Alexa top 1-million websites. They reported that a few third-parties including Google, Facebook, Twitter,Amazon, and AdNexus cover at least 10% of the top 1M websites. They also showed that client-side cookie syncing is prevalent among third-parties: 45 of the top 50 third-parties sync cookies with at least one other party and the most popular third-party (`doubleclick.net`, an Alphabet-owned AdX) syncs cookies with 118 different third-parties. This highlighted that trackers, even when owned by different organizations, often exchanged data to improve participation in online advertising. In addition to information flows among trackers that are observable at the client-side (*i.e.,* cookie syncing), researchers have also investigated methods to detect server-to-server (S2S) information flows among trackers. This is much more challenging since they are not observable at the client-side (browser).

To address this challenge, Bashir *et al.*conducted controlled experiments and inferred a small subset of S2S information flows by investigating the process of ad retargeting [35]. Since retargeting necessitates data exchange between two AdXes, the authors conducted controlled experiments to trigger and detect ad retargeting and infer S2S information flows. The underlying intuition was that if a retargeted ad was served by an entity that did not observe the original visit which triggered the retargeted ad, then it must have got information about this visit through an S2S information flow. We leverage the same underlying intuition as in Bashir *et al.* [35, 36] to infer tracker-advertiser relationships. However, instead of relying on retargeting as the "signal" for data exchange, we operationalize this intuition using a machine learning model that is trained to predict an advertiser's bid values using presence/absence information of popular trackers.

6 Concluding Remarks

In this paper, we leveraged header bidding (HB) to gain insights into the bidding behavior of advertisers and presented a machine learning approach to infer data sharing relationships between advertisers and trackers. Our work advances the field along two main avenues. First, we are able to provide more nuanced insights into the bidding behavior of online advertisers. While prior research [67, 71] was limited to analyzing only the winning bids in RTB, we are able to observe all bids made by different advertisers in HB. Second, we are able to infer data sharing relationships between advertisers and trackers irrespective of whether it is happening at the client-side or the server-side. While prior work could only detect client-side data sharing [69] or infer server-side data sharing relationships when retargeting occurs [35], our approach is able to infer client-side and server-side data sharing for any advertiser placing bids without relying on specific triggers such as retargeting.

Our work can help existing privacy-enhancing tools in presenting empirically derived inferences about (1) how the data is shared between entities in the online advertising and tracking ecosystems; and (2) what is the perceived value of users. Along the first direction, privacy enhancing tools such as Mozilla Lightbeam [20], uBlock Origin [56], and Ghostery [22] provide users transparency and control over online tracking. Our work can be used to address known limitations [36] of these tools by identifying server-side data sharing practices of online trackers. Along the second direction, our HB measurements can be used to improve existing user valuation tools such as *RTBAnalyzer* [67] and *YourAD-Value* [71] by capturing a more complete picture of *all* advertisers' bidding behaviors. These measurements can further inform micropayment-based alternate web monetization models [61] (*e.g.,* Flattr [21], Contributor [23], BAT [9]) by suggesting how much users should pay a publisher in exchange for blocking ads.

Acknowledgement

This research is supported in part by a Google Faculty Research Award, a grant from the Data Transparency Lab, and the National Science Foundation under grant numbers 1715152 and 1815131.

References

[1] AppNexus Joins Comscore Industry Trust, Comscore. https://www.comscore.com/ita/Public-Relations/Blog/AppNexus-Joins-comScore-Industry-Trust, 2015.

[2] DoubleVerify Launches with the New AppNexus Spend Protection Program. https://www.doubleverify.com/newsroom/doubleverify-launches-with-the-new-appnexus-spend-protection-program, 2015.

[3] General Data Protection Regulation (GDPR). https://gdpr-info.eu, 2016.

[4] Personalization Delivers 3X Consumer Engagement With Digital Advertising, Jivox. https://www.jivox.com/press/personalization-delivers-3x-consumer-engagement-with-digital-advertising/, 2016.

[5] OpenX Strengthens Product and Technology Teams with Key Hires, OpenX. https://www.openx.com/company/press-releases/openx-strengthens-product-technology-teams-key-hires, 2017.

[6] Rubicon Project Partners with Kiip to Automate Mobile In-App Rewarded Inventory, Rubicon Project. http://investor.rubiconproject.com/news-releases/news-release-details/rubicon-project-partners-kiip-automate-mobile-app-rewarded, 2017.

[7] Server-to-Server Header Bidding: The Pros and Cons, The AppNexus Team. https://www.appnexus.com/blog/server-server-header-bidding-pros-and-cons, 2017.

[8] The Economic Contribution of Digital Advertising in Europe, IHS Markit. https://datadrivenadvertising.eu/wp-content/uploads/2017/09/DigitalAdvertisingEconomicContribution_FINAL-1.pdf, 2017.

[9] Basic Attention Token (BAT): Blockchain Based Digital Advertising. https://basicattentiontoken.org/BasicAttentionTokenWhitePaper-4.pdf, 2018.

[10] EasyList. https://easylist.to/easylist/easylist.txt, 2018.

[11] Easyprivacy. https://easylist.to/easylist/easyprivacy.txt, 2018.

[12] Server Postback Tracking Explained. https://help.tune.com/hasoffers/server-postback-tracking-explained, 2018.

[13] The California Consumer Privacy Act of 2018. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375, 2018.

[14] Tracking, Cookies, and ITP 2.0. https://support.partnerstack.com/hc/en-us/articles/360011902273-Tracking-Cookies-and-ITP-2-0, 2018.

[15] Ad Tech Insights - Header Bidding Industry Index, Adzerk. https://adzerk.com/assets/reports/AdTechInsights_Feb2019.pdf, 2019.

[16] Alexa - Top Sites by Category: The top 500 sites on the web, Alexa - An Amazon.com company. https://www.alexa.com/topsites/category, 2019.

[17] AppNexus Third Party Providers, OpenX. https://www.appnexus.com/third-party-providers, 2019.

[18] Cookie Matching|Authorized Buyers, Google. https://developers.google.com/authorized-buyers/rtb/cookie-guide, 2019.

[19] Cookie Policy, Automattic. https://automattic.com/cookies, 2019.

[20] Firefox Lightbeam by Mozilla. https://addons.mozilla.org/en-US/firefox/addon/lightbeam/, 2019.

[21] Flattr - Contributors. https://flattr.com/contributors, 2019.

[22] Ghostery makes the Web Cleaner Safer and Faster! https://www.ghostery.com/, 2019.

[23] Google Contributor. https://contributor.google.com/v/beta, 2019.

[24] Is your pregnancy app sharing your intimate data with your boss? https://www.washingtonpost.com/technology/2019/04/10/tracking-your-pregnancy-an-app-may-be-more-public-than-you-think/, 2019.

[25] OpenX and Microsoft Announce Advertising Partnership, OpenX. https://www.openx.com/company/press-releases/openx-and-microsoft-announce-advertising-partnership/, 2019.

[26] Privacy Badger, Electronic Frontier Foundation. https://www.eff.org/privacybadger, 2019.

[27] Server-side Tracking: General discussion and Common issues in Server-side Tracking, Woopra. https://docs.woopra.com/docs/serverside-tracking, 2019.

[28] Strategic Alliances - Index Exchange. https://www.indexexchange.com/alliances, 2019.

[29] Surf The Web With No Annoying Ads. https://adblockplus.org, 2019.

[30] US Digital Ad Spending Will Surpass Traditional in 2019, eMarketer. https://www.emarketer.com/content/us-digital-ad-spending-will-surpass-traditional-in-2019, 2019.

[31] Why 2018 Was The Year Header Bidding Realized Its Potential, AdExchanger. https://adexchanger.com/ad-exchange-news/why-2018-was-the-year-header-bidding-realized-its-potential, 2019.

[32] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *ACM Conference on computer and Communications Security (CCS)*, 2014.

[33] A. Acquisti, L. K. John, and G. Loewenstein. What is privacy worth? *The Journal of Legal Studies*, 42(2):249–274, 2013.

[34] M. Backes, S. Bugiel, and E. Derr. Reliable Third-Party Library Detection in Android and its Security Applications. In *ACM Conference on computer and Communications Security (CCS)*, 2016.

[35] M. A. Bashir, S. Arshad, W. Robertson, and C. Wilson. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *Proceedings of the 25th USENIX Security Symposium*, 2016.

[36] M. A. Bashir and C. Wilson. Diffusion of User Tracking Data in the Online Advertising Ecosystem. In *Proceedings on Privacy Enhancing Technologies (PETS)*, 2018.

[37] J. Brookman, P. Rouge, A. Alva, and C. Yeung. Cross-device tracking: Measurement and disclosures. *Privacy Enhancing Technologies Symposium (PETS)*, 2017.

[38] I. Campbell. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26(19):3661–3675, 2007.

[39] J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira. Your Browsing Behavior for a Big Mac: Economics of Personal Information Online. In *22nd International Conference on World Wide Web (WWW)*, 2013.

[40] T. Chen, I. Ullah, M. A. Kaafar, and R. Boreli. Information Leakage through Mobile Analytics Services. In *ACM Workshop on Mobile Computing Systems and Applications*

(HotMobile), 2014.

[41] D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis. A study on the value of location privacy. In *ACM Workshop on Privacy in Electronic Society (WPES)*, pages 109–118. ACM, 2006.

[42] G. Danezis, S. Lewis, and R. J. Anderson. How much is location privacy worth? In *Workshop on the Economics of Information Security (WEIS)*, 2005.

[43] A. Dey. Header Bidding vs RTB: Understanding the Differences. https://blognife.com/2018/09/08/header-bidding-vs-rtb-understanding-the-differences/, 2018.

[44] W. Dou. Will Internet Users Pay for Online Content? *Journal of Advertising Research*, 44, 02 2005.

[45] S. Englehardt and A. Narayanan. Online Tracking: A 1-million-site Measurement and Analysis. In *ACM Conference on Computer and Communications Security (CCS)*, 2016.

[46] J. Estrada-Jiménez, J. Parra-Arnau, A. Rodríguez-Hoyos, and J. Forné. On the regulation of personal data distribution in online advertising platforms. *Engineering Applications of Artificial Intelligence*, 82:13–29, 2019.

[47] L. Fisher. Surveying The Digital Future, The 2017 Digital Future Report, Center for the Digital Future at USC Annenberg. http://www.digitalcenter.org/wp-content/uploads/2013/10/2017-Digital-Future-Report.pdf, 2017.

[48] L. Fisher. US Programmatic Ad Spending Forecast Update 2018, eMarketer. https://www.emarketer.com/content/us-programmatic-ad-spending-forecast-update-2018, 2018.

[49] L. Fisher. Header Bidding Update 2018. What's the Outlook for Web, Mobile App and Video? https://www.emarketer.com/content/header-bidding-update-2018, 2019.

[50] I. fouad, N. Bielova, A. Legout, and N. Sarafijanovic-Djuki. Tracking the Pixels: Detecting Unknown Web Trackers via Analysing Invisible Pixels. In *arXiv:1812.01514v2*, 2019.

[51] G. A. Fowler. It's the middle of the night. Do you know who your iPhone is talking to? https://www.washingtonpost.com/technology/2019/05/28/its-middle-night-do-you-know-who-your-iphone-is-talking, 2019.

[52] A. Ghosh, M. Mahdian, R. P. McAfee, and S. Vassilvitskii. To match or not to match: Economics of cookie matching in online advertising. *ACM Transactions on Economics and Computation*, 3, 2015.

[53] J. González Cabañas, A. Cuevas, and R. Cuevas. FDVT: Data Valuation Tool for Facebook Users. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2017.

[54] J. Grossklags and A. Acquisti. When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. In *Workshop on the Economics of Information Security (WEIS)*, 2007.

[55] L. Handley. Half of all advertising dollars will be spent online by 2020, equaling all combined 'offline' ad spend globally. https://www.cnbc.com/2017/12/04/global-advertising-spend-2020-online-and-offline-ad-spend-to-be-equal.html, 2016.

[56] R. Hill. An efficient blocker for Chromium and Firefox. Fast and lean, uBlock Origin. https://github.com/gorhill/uBlock#ublock-origin, 2019.

[57] B. A. Huberman, E. Adar, and L. R. Fine. Valuating Privacy. *IEEE Security & Privacy*, 3(5):22–25, 2005.

[58] V. Kalavri, J. Blackburn, M. Varvello, and K. Papagiannaki. Like a Pack of Wolves: Community Structure of Web Trackers. In *International Conference on Passive and Active Network Measurement (PAM)*, 2016.

[59] A. Le, J. Varmarken, S. Langhoff, A. Shuba, M. Gjoka, and A. Markopoulou. AntMonitor: A system for monitoring from mobile devices. In *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*, pages 15–20. ACM, 2015.

[60] A. Lerner, A. K. Simpson, T. Kohno, and F. Roesner. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *USENIX Security Symposium*, 2016.

[61] M. Lesk. Micropayments: An idea whose time has passed twice? *IEEE Security Privacy*, 2(1):61–63, 2004.

[62] T.-C. Lin, J. S.-C. Hsu, and H.-C. Chen. CUSTOMER WILLINGNESS TO PAY FOR ONLINE MUSIC: THE ROLE OF FREE MENTALITY. *Journal of Electronic Commerce Research*, 14(4), 2013.

[63] J. R. Mayer and J. C. Mitchell. Third-Party Web Tracking: Policy and Technology. In *IEEE Symposium on Security and Privacy*, 2012.

[64] G. Merzdovnik, M. Huber, D. Buhov, N. Nikiforakis, S. Neuner, M. Schmiedecker, and E. Weippl. Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools. In *IEEE European Symposium on Security and Privacy*, 2017.

[65] R. Molla. Next year, people will spend more time online than they will watching TV. That's a first. https://www.recode.net/2018/6/8/17441288/internet-time-spent-tv-zenith-data-media, 2018.

[66] N. Nguyen. Latest Firefox Rolls Out Enhanced Tracking Protection. https://blog.mozilla.org/blog/2018/10/23/latest-firefox-rolls-out-enhanced-tracking-protection/, 2018.

[67] L. Olejnik, M.-D. Tran, and C. Castelluccia. Selling Off Privacy at Auction. In *Proceedings of the 2014 Network and Distributed System Security Symposium*. Internet Society, 11 2014.

[68] M. Pachilakis, P. Papadopoulos, E. P. Markatos, and N. Kourtellis. No More Chasing Waterfalls: A Measurement Study of the Header Bidding Ad-Ecosystem. *arXiv:1907.12649*, 2019.

[69] P. Papadopoulos, N. Kourtellis, and E. Markatos. Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask. In *The Web Conference (WWW)*, 2019.

[70] P. Papadopoulos, N. Kourtellis, and E. P. Markatos. The Cost of Digital Advertisement: Comparing User and Advertiser Views. In *World Wide Web Conference (WWW)*, 2018.

[71] P. Papadopoulos, N. Kourtellis, P. R. Rodriguez, and N. Laoutaris. If You Are Not Paying for It, You Are the Product: How Much Do Advertisers Pay to Reach You? In *ACM Internet Measurement Conference (IMC)*, 2017.

[72] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, S. Sundaresan, M. Allman, and C. K. P. Gill. Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem. In *Network and Distributed System Security Symposium (NDSS)*, 2018.

[73] I. Reyes, P. Wijesekera, A. Razaghpanah, J. Reardon, N. Vallina-Rodriguez, S. Egelman, and C. Kreibich. "Is Our Children's Apps Learning?" Automatically Detecting COPPA Violations. In *IEEE Workshop on Technology and Consumer Protection (ConPro)*, 2017.

[74] A. Senior. John Hancock Leaves Traditional Life Insurance Model Behind to Incentivize Longer, Healthier Lives. https://www.johnhancock.com/content/johnhancock/news/insurance/2018/09/john-hancock-leaves-traditional-life-insurance-model-behind-to-incentivize-longer--healthier-lives.html, 2018.

[75] N. Vallina-Rodriguez, J. Shah, A. Finamore, Y. Grunenberger, K. Papagiannaki, H. Haddadi, and J. Crowcroft. Breaking for commercials: characterizing mobile advertising. In *ACM Internet Measurement Conference (IMC)*, 2012.

[76] N. Vallina-Rodriguez, S. Sundaresan, A. Razaghpanah, R. Nithyanand, M. Allman, C. Kreibich, and P. Gill. Tracking the Trackers: Towards Understanding the Mobile Advertising and Tracking Ecosystem. In *Workshop on Data and Algorithmic Transparency (DAT)*, 2016.

[77] C. L. Wang, Y. Zhang, L. R. Ye, and D.-D. Nguyen. Subscription to fee-based online services: What makes consumer pay for online content? *Journal of Electronic Commerce Research*, 6(4):304, 2005.

[78] J. Wilander. Intelligent Tracking Prevention 2.0. https://webkit.org/blog/8311/intelligent-tracking-prevention-2-0/, 2018.

[79] S. Zimmeck, J. S. Li, H. Kim, S. M. Bellovin, and T. Jebara. A Privacy Analysis of Cross-device Tracking. In *USENIX Security Symposium*, 2017.