Engineering Applications of Artificial Intelligence 82 (2019) 13–29

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

On the regulation of personal data distribution in online advertising platforms☆

José Estrada-Jiménez [a,b], Javier Parra-Arnau [c,*], Ana Rodríguez-Hoyos [a,b], Jordi Forné [b]

[a] Departamento de Electrónica, Telecomunicaciones y Redes de Información, Escuela Politécnica Nacional (EPN), Ladrón de Guevara, E11-253 Quito, Ecuador
[b] Department of Telematics Engineering, Universitat Politècnica de Catalunya (UPC), C. Jordi Girona 1-3, E-08034 Barcelona, Spain
[c] Department of Computer Science and Mathematics, Universitat Rovira i Virgili (URV), E-08034 Tarragona, Spain

ARTICLE INFO

Keywords:
Online advertising
Information privacy
Revenue-invitation rate trade-off

ABSTRACT

Online tracking is the key enabling technology of modern online advertising. In the recently established model of real-time bidding (RTB), the web pages tracked by ad platforms are shared with advertising agencies (also called DSPs), which, in an auction-based system, may bid for user ad impressions. Since tracking data are no longer confined to ad platforms, RTB poses serious risks to privacy, especially with regard to user profiling, a practice that can be conducted at a very low cost by any DSP or related agency, as we reveal here. In this work, we illustrate these privacy risks by examining a data set with the real ad-auctions of a DSP, and show that for at least 55% of the users tracked by this agency, it paid nothing for their browsing data. To mitigate this abuse, we propose a system that regulates the distribution of bid requests (containing user tracking data) to potentially interested bidders, depending on their previous behavior. In our approach, an ad platform restricts the sharing of tracking data by limiting the number of DSPs participating in each auction, thereby leaving unchanged the current RTB architecture and protocols. However, doing so may have an evident impact on the ad platform's revenue. The proposed system is designed accordingly, to ensure the revenue is maximized while the abuse by DSPs is prevented to a large degree. Experimental results seem to suggest that our system is able to correct misbehaving DSPs, and consequently enhance user privacy.

1. Introduction

The growing access of people to information and communication technologies is contributing to reach the so-called "big data era", where the pervasiveness of data is a major input for increasingly personalized and automated online services. One of such services is online advertising, which aims at selecting and directing ads to the right potential customers (personalization) at the right time (real-time), built on multiple parameters, while users browse the Web (Smith, 2014a; Real-time bidding protocol, 0000; Yuan et al., 2012).

This targeted advertising offers crucial benefits to several agents on the Internet. To start, users receive ads tailored to their interests and no longer static ads unrelated to their preferences; consequently, behavioral targeting ensures conversion rates[1] that double those of untargeted ads (Beales, 2010). Furthermore, web sites have access to an entire ecosystem to fund their operation through the money paid by demand side platforms (DSPs), which are advertising agencies acting in representation of advertisers.[2] Also, selling entities are given the opportunity to promote their products over a ubiquitous structure with global reach. The upshot is that most of the content users consume online is supported by ad revenue.

One of the key enabling technologies that makes online advertising so profitable is real-time bidding (RTB), which enables advertisers to compete in real-time auctions to show their ads (Yuan et al., 2013). It is implemented by a management entity called ad exchange. Accordingly, when a user visits a website, her impression is sold to the advertiser (or corresponding DSP) that bids higher, in a matter of milliseconds. Moreover, DSPs are sent bid request messages containing user information (tracking data) to help them tailor ads to the user's preferences and decide the bidding strategy. In this way, the RTB aim is twofold: offering users a personalized experience through targeted ads and, thus, maximizing the profits of the whole advertising ecosystem. Whereas the operation of RTB behind the scenes is pretty opaque and complex for users (Mcdonald et al., 2009), it is quite transparent for

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.engappai.2019.03.013.

* Corresponding author.
E-mail addresses: jose.estrada@epn.edu.ec (J. Estrada-Jiménez), javier.parra@urv.cat (J. Parra-Arnau), ana.rodriguez@epn.edu.ec (A. Rodríguez-Hoyos), jforne@entel.upc.edu (J. Forné).

[1] In online marketing terminology, conversion usually means the act of converting Web site visitors into paying customers.

[2] When referring to the online advertising model in general, the terms 'buyer', DSP and 'advertiser' may be used interchangeably.

https://doi.org/10.1016/j.engappai.2019.03.013
Received 26 June 2018; Received in revised form 4 January 2019; Accepted 19 March 2019
Available online xxxx
0952-1976/

the actors of the advertising ecosystem. For example, ad exchanges provide DSPs with powerful management interfaces that offer very detailed information about the market and even enable buyers to set up their advertising strategies (e.g., by defining a targeting market). Certainly, a lot of benefits arise for the advertising ecosystem from the optimization capability offered by RTB in terms of automation, personalization, profit and transparency.

Yet, despite its proven usefulness, the practices inherent to online advertising and RTB may pose serious privacy risks for users (Estrada-Jiménez et al., 2017). Most of these risks derive from the potential misuse of the user data flowing through the advertising ecosystem. To start, vast user data is mined at very high rates to implement real-time personalization; hence, truly detailed profiles are built about millions of people so fast and uncontrollably (Narayanan and Shmatikov, 2008) that privacy protection is discouraged. Additionally, ad distribution mechanisms based on auctioning user impressions might lead to characterize users as more relevant (or more valuable economically) than others, depending on their profiles (Olejnik et al., 2014); such a differentiation may entail social sorting or discrimination (Speicher et al., 2018), thus an even less private environment. Finally, online advertising builds on interactions among myriads of intermediary ad companies that collect, use and share user data, significantly increasing the risk of data misuse. Ironically, users have no control over how their data is managed in this context.

RTB builds on sharing user data with DSPs to encourage competition and ad personalization, but the unregulated distribution of such data may give rise to concerns. With the aim of helping DSPs decide whether to bid or not for a user impression, an ad exchange distributes to them personal information of the user whose impression is being auctioned (e.g., the URL being visited, the location of the user, or even a label categorizing the user). Thus, not only does the winner DSP receives this input, but also the rest of participating DSPs. This means that there could be agencies maliciously collecting data without even paying for it. We illustrate this risk in Section 3 where we unveil that a given DSP would have paid nothing for at least 55% of the users it tracked in a period of three months. This uncontrolled distribution of user data prompts a non-negligible privacy concern since an increasing number of advertising agencies are relying on RTB to daily reach billions of potential Web customers (Hoelzel, 2015). Although the distribution of personal data among a group of DSPs cannot be entirely stopped without changing the current advertising business model, we report that the potential abuse of these agencies can be tackled with minimum tuning of said data distribution model.

Our proposal builds on regulating the distribution of personal data from the ad exchange to DSPs when a user impression is auctioned. Such regulation consists essentially in limiting the number of DSPs invited to bid, that is lowering the entities to which user data is leaked and, consequently, getting a more private environment. Accordingly, DSPs or similar intermediaries showing a dishonest behavior (e.g., never winning auctions) will be banned from participating in future auctions, which may entail correcting such harmful behaviors. At the same time, our approach strives to maximize the revenue of the ad exchange, looking for a balance with a given privacy level. The upshot is that some privacy can be reached without affecting the business model of the online advertising ecosystem, by slightly modifying the distribution of personal data among intermediary entities such as DSPs. The resulting adjusting effect on the behavior of these entities is relevant since privacy concerns in general do not directly derive from the act of sharing data itself, but from the *inappropriate* sharing of user information (Nissenbaum, 2009).

Unlike our approach, other proposals address this privacy issue through more radical strategies. Research proposals have concentrated on sophisticated mechanisms to anonymize or block the information leaked to third-parties while trying to remain compatible with the current ecosystem, but still requiring important modifications to its architecture and anyhow affecting its economy. On the other hand, commercial solutions have primarily focused on blocking tracking mechanisms at the cost of seriously damaging the Internet business model. However, as concluded in Estrada-Jiménez et al. (2017), it seems very hard to provide more privacy in the online advertising ecosystem without somehow modifying the ad delivery model.

1.1. Related work

In general, the concerns regarding privacy arise from the inappropriate collection, use and sharing of user data (Nissenbaum, 2009). In the context of online advertising, said misuse of user data is potentially present in different moments in time. First, powerful tracking mechanisms are employed by high-level advertising players to "follow" users through the Web (Olejnik and Castelluccia, 0000; Eckersley, 2010; Mayer and Mitchell, 2012). These tracking mechanisms include cookie matching and fingerprinting. When users navigate a website serving ads, third-party interactions from the browser disclose user data to said players, which aggregate and store this information (collection). Then, they process this user data (use) and further distribute it (sharing) to enable personalization for users and to guide the targeting strategy of advertisers.

Within this framework, *external* control over the flow of data could only be enforced before the collection step, i.e., when the web browser leaks data in third-party interactions. Further adjustments require changing online advertising structures. This is why most of the functional solutions to protect privacy in this domain build on managing (essentially detecting and blocking) third-party connections from the user side. These are local approaches, commonly implemented as web browser extensions, that provide users with tracking blocking capabilities. The most popular ad blocker is AdBlock Plus (Adblock Plus, 2015), but other similar tools exist that also provide transparency and user personalization (Parra-Arnau et al., 2016; Sánchez and Viejo, 2018; Achara et al., 2016). In this line, other initiatives propose blocking strategies implemented in brokers (Backes et al., 2012; Guha et al., 2011; Privoxy.org, 2016) that act as local proxies to filter the interactions performed between a group of local users and advertising entities on the Web. Historically, these approaches have detected third-party tracking through static blocking lists that have become extremely long and hard to manage (Easylist, 2016), but recent proposals have improved such detection by using machine-learning techniques (Papadopoulos et al., 2018).

However, ad blockers and anti-trackers suffer from controversial shortcomings. First, its extended use is seriously threatening the business model of the whole Internet. Also, though radical and apparently infallible, ad blocking would have been circumvented by tracking companies by exploiting web sockets (Bashir et al., 2018). Namely, ad blockers might not be as effective as expected.

Looking for more advertising-friendly solutions to preserve privacy, multiple initiatives have emerged from the academic world. Those mostly suggest integrating the active participation of users so they can decide how to manage their data. Some of these works (Backes et al., 2012) propose incorporating trusted third-parties to intermediate the communication between users and advertising players to encrypt or obfuscate user data. Several other approaches present advertising architectures where the exploitation and sharing of data is moved to the user premises, i.e., to the user's browser or a local application (Guha et al., 2011; Toubiana et al., 2010; Fredrikson and Livshits, 2010). This enables users to control how their data is processed and how and when it is shared to third-parties. Two very recent research works (Helsloot et al., 2018, 2017) present protocols to exploit personal data while auctioning user impressions without revealing any personal preferences (in clear text) to advertising parties. As some of the previous approaches, these protocols require that user information be processed locally in the browser, and that a trusted third-party assist in performing operations over encrypted user data.

Other more revolutionary proposals even suggest adapting the advertising model to allow users to be rewarded for ceding their

data (Parra-Arnau, 2017; Brave, 2016) or to enable advertising players to charge users for not tracking them (Mozilla, 0000). Sadly, all this related academic research require modifying the current online advertising model, either in the way user data is exploited or in the mechanism to obtain (economic) value from it. These are important changes that would significantly impact the utility of the user information received by ad platforms, thus negatively affecting their huge revenues. As a consequence, there might not be incentives for the advertising market to adopt them in the short-medium term.

With blocking solutions that are critically tampering with the economy of the Web (Shiller et al., 2017) and academic approaches that are not feasible in the short term, it seems that we need to look beyond to get real privacy. Reaching effective strategies implies starting to disrupt the core of the ecosystem in order to address the moments when user data is processed and shared. Some steps in that line are already taking place thanks to strict privacy regulation recently promulgated (GDPR, 0000) in Europe that is motivating companies to cooperate in favor of the privacy of users. Interestingly, the mere application of transparency initiatives has already allowed to unveil further privacy risks within advertising platforms (Faizullabhoy and Korolova, 2018; Venkatadri et al., 2018).

1.2. Main requirements of the system

In this subsection we include the main requirements around which our proposal revolves, in order to guide the approaches we do next.

- *Simple implementation*. To encourage its implementation in practice, we promote a solution that does not require modifying significantly the architecture of the online advertising ecosystem. This would imply a low adoption cost, unlike other academic approaches that propose rebuilding the current model to protect privacy.
- *Constructive technique*. Looking for an alternative to the arms race started by ad blockers, which is threatening the economy of the Web, we require a mechanism aimed to balance the tradeoff between user privacy and advertising utility (commonly in terms on money). This would limit the level of attainable privacy, but would open the door to further tangible mechanisms to address privacy concerns in this opaque environment.
- *Self-regulatory*. In line with a constructive approach, we uphold a system that allows misbehaving entities to correct their practices against privacy, under the penalty of dynamic punishment. Namely, we require a solution that promotes appropriate practices towards user data to relieve privacy concerns.

Interestingly, the compliance with these three main requirements when designing our system will derive in additional aspects that may go in favor of user privacy.

1.3. Contribution and plan of this paper

In this work, we illustrate the potential misuse of RTB with real data from a publicly available data set. We analyze the data of more than 64 millions of ad-auctions and interactions between a DSP and an ad exchange, to quantify the extent to which a DSP may collect user tracking data without paying for them. To the best of our knowledge, this is the first study reporting quantitative evidences on the misuse of RTB.

Since no preventive mechanism is currently put in place by Google's DoubleClick and AppNexus (the most relevant RTB systems), we hypothesize that such tracking and profiling practices may be rather common. To address this state of affairs, our second contribution is a system that aims to regulate the distribution of user data to third parties during the auctions for ad-impressions, i.e., to whom send the requests for each ad-space bidding.

The proposed solution is designed to strike a *balance* between *the average number of DSPs invited to bid* and the *revenue of the ad exchange* holding the auctions. Limiting the number of DSPs receiving user profiles naturally offer better privacy protection, especially since potential dishonest DSPs will hardly receive user sensitive information under such context. As a consequence, an ad exchange might be motivated to *suppress* the *bid requests* to abusing DSPs, but this would have an impact on its revenue. We formulate the problem of choosing a bid-request distribution as a multi-objective optimization problem that takes into account both aspects, i.e., the number of DSPs invited to bid and RTB profits.

We measure the extent to which user data is disseminated as the average number of DSPs receiving tracking data. Accordingly, for a desired data distribution strategy, our solution recommends, probabilistically and in real time, to which DSPs the ad exchange should send a bid request for any given ad impression, in order to maximize the instant revenue. Evidently, with the aim of preventing abuses and thus supporting privacy, the fewer DSPs receive personal data the better. Experimental results show that our system seems to be able to tackle misbehaving DSPs.

The remainder of this work is organized as follows. Section 2 provides the necessary background in online advertising. Then, Section 3 analyzes the potential abuses and privacy risks we face in this context. Section 4 presents the theoretical analysis of our regulating approach. In Section 5, we evaluate our technique. Section 6 includes a relevant discussion about important topics of our approach and some general incentives to adopt it. Finally, conclusions are drawn in Section 7.

2. Background in online advertising platforms

This section addresses the main concepts involved in the current online advertising ecosystem, in particular with regard to its main players, the interactions among them, and supporting technologies. This knowledge shall provide the reader with the necessary depth to grasp the technical contributions of this work.

2.1. The online advertising landscape

Generally, advertising is conceived as a form of communication aimed at persuading users to buy a product. Even a relative success in such an ambitious (and commercial) objective have led to generate lots of money, so much money that advertising is said to be supporting the existing Internet free access model now (Gayomali, 2014). Due to structure limitations, traditional advertising consisted in massively flooding media with generic ads. Though such massiveness brought interesting revenues, it turned annoying for customers (Rejón-Guardia and Martínez-López, 2014) and caused rejection due to the lack of usability that provoked on web sites. On the other hand, with the rise of the Internet, more granularity became available with regard to user data. Thus, modern (online) advertising has developed a much greater capability of reaching potential customers on an individual basis. For this, recommendation and personalized information systems are being exploited to tailor advertising campaigns to the interests of Web users (Kardan and Hooman, 2013). Then, users are not flooded in their browsers with uninteresting ad content, yet, within ad platforms, a wealth of user information fuels a targeted and optimized strategy.

This optimization is focused on the revenue of the ad distribution system whose core is an auction technology, called real-time bidding (RTB) that allows to assign ad spaces to the highest bidder (Yuan et al., 2013). Along with the use of other technologies, this usually derives in showing ads to the right person and at the right time. Also in this context, more accountability, transparency and effectiveness (Evans, 2009) are provided since ad companies are encouraged to agree on prices that directly match the effort undertaken by the seller with

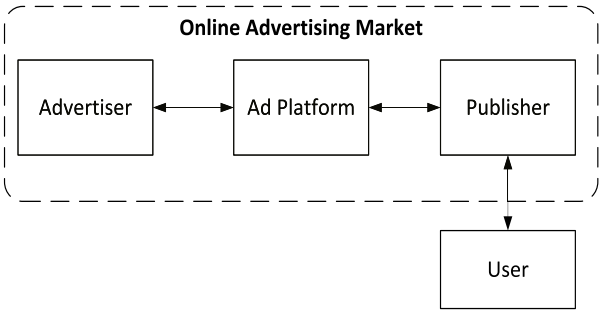

Fig. 1. Main components of the online advertising ecosystem.

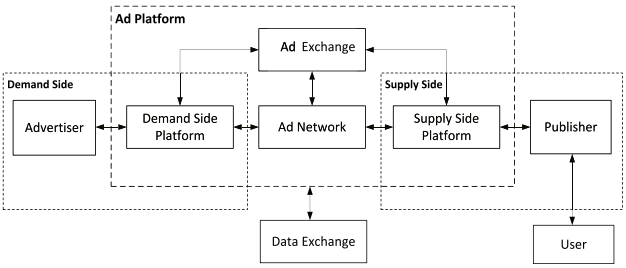

Fig. 2. Disaggregated ad platform scheme and interactions between players.

the benefits received by the buyer (Yuan et al., 2012). Sadly, said accountability and transparency are by no means offered to end users.

Going a little deeper on the structure, the online advertising landscape is triggered by advertisers, who create the demand, and publishers, who generate the supply. Websites have become the publishers by excellence since the content they offer attracts people whose interests can be revealed from intrinsic interactions with the Web. Beyond these fundamental entities in the advertising logic, modern online advertising management has incorporated intermediate entities that help advertisers and publishers navigate the web topology in order to connect them together (Evans, 2009). Such intermediaries are responsible for providing interactive and automatic ad serving that is able to accurately target the intended audience. Said targeting strategy has directly influenced the ad-personalization accuracy, but also the level of transparency (for advertising players) of the process whereby ads are delivered. Of course, outsourcing user data and interactions from publishers and advertisers is required to achieve these goals.

2.2. Online advertising players

As suggested in Section 2.1, the modern online advertising infrastructure builds on three main components. As illustrated in Fig. 1, these components are advertisers, publishers and ad platforms (the set of intermediate entities managing the interactions among the former two). Internet browsing triggers such interactions and user data enables ad targeting; however, users are not given any active role in this context, by default. Yet, a change on their behavior with regard to advertising may significantly impact on the current advertising business model.

Advertisers are entities willing to pay for displaying ads on some spaces of websites (publishers) in order to promote a product to potential customers (Yuan et al., 2012; IAB, 2015). Advertisers and publishers are commonly engaged through intermediate platforms, as shown in Fig. 1, to make their interactions more efficient. This efficiency derives in the capability of advertisers to target ads to their intended audiences.

A **publisher** is an entity, such as CNN or The New York Times, which provides online content (e.g., newspapers, search engines, blogs, etc.), usually through web pages. Since such content draws the attention of users, advertisers pay publishers to be assigned a space in a website, where they can show ads to a given audience.

Ad platforms represent the marketplace where the demand (from advertisers) and the supply (from publishers) of online advertising services are matched (Yuan et al., 2012) (see Fig. 2). They are built of agents (intermediaries) with very specialized roles. On the one hand, they offer interfaces for advertisers and publishers to outsource some of their interactions. On the other hand, they optimize the ad serving process in terms of revenue, flexibility and transparency. Corresponding entities to offer these services have emerged to now give rise to ad platforms.

Ad networks emerged to aggregate ad inventory bought from publishers (ad spaces) in order to resell it to advertisers (OpenX, 2010). By piling ad spaces, ad networks were pioneers in supporting advertisers to reach more selective audiences. Ad networks have evolved into more complex structures, called ad exchanges, though some still operate as they were conceived originally.

Ad exchanges sell their aggregated inventory of ad spaces by means of auctions. They keep consolidating ad spaces from publishers but offer advertisers and publishers more effective and transparent mechanisms to serve ads (Yuan et al., 2012; Mayer and Mitchell, 2012). First, ad exchanges place ads based on automated auctions where advertisers (or those in their representation) "decide" how much to pay for an ad space. The winning bidder is the advertiser that ends up displaying the ad. Secondly, during the auction, ad exchanges share with advertisers "contextual" information about the user who generates the impression they bid for. Such information helps advertisers decide whether to bid for an ad space and how much to bid for it. The auction is held just after a user requests content from a website partnering with the ad exchange. The whole process may take a few tenths of a second. Theoretically, this yields greater efficiency since the ad-delivery process is distributed among the different components of the ad platform (Smith, 2014b). Part of the aggregation strategy of ad exchanges consists in combining multiple ad networks together. This way, advertisers and publishers are relieved from dealing with so many intermediaries. In practice, ad exchanges do not deal directly with advertisers, but with demand side platforms which act in the name of advertisers.

Demand side platforms (DSPs) are entities that work for advertisers, i.e., for the actors generating the demand of ad services. DSPs work on behalf of advertisers, in front of the ad exchange, and help advertisers choose audiences and adequate media to display their ads. By aggregating demand, DSPs are capable of boosting selectiveness and effectiveness for advertisers (Yuan et al., 2012; Mayer and Mitchell, 2012). With the increasing complexity of the advertising ecosystem, advertisers have lost some fine control over ad placement done through DSPs. Consequently, other agents have appeared on the demand side to serve advertisers. We talk, e.g., about trading desks, which give advertisers tools to manage their campaigns more closely and to optimize their strategy according to their needs. Trading desks commonly take advantage of the services provided by various DSPs.

Supply side platforms (SSPs) are entities that work on behalf of publishers, the actors that supply ad spaces to advertisers. SSPs offer publishers an optimized strategy to manage their advertising inventory.

Finally, there are other players in this ecosystem operating on top of demand, supply and ad exchange platforms, which is the case, e.g., of data aggregators and data exchanges. They collect user data to sell it to demand and supply-side platforms to help them make their targeting decisions.

2.3. RTB: the auction technology behind online advertising

When a user visits a Web site with an ad space served through RTB (Yuan et al., 2013), an HTTP request is submitted to the ad exchange, which subsequently sends "bid requests" to potential participants. We note that the number and type of participants involved may vary on a per-auction basis, at the ad exchange's discretion. Within the bid request, the ad exchange generally includes the following data:

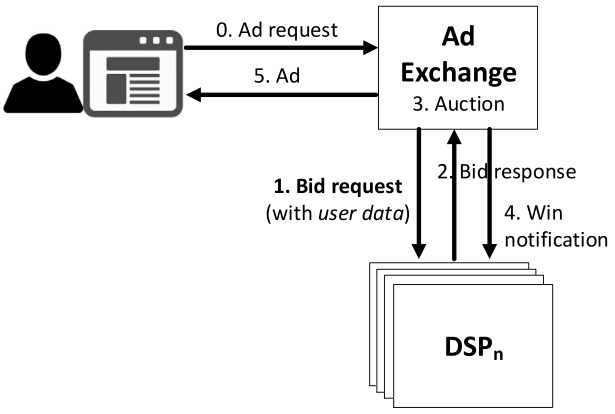

Fig. 3. Interactions among an ad exchange and associated DSPs.

the URL of the page being visited by the user; the topic category of the page; the user's IP address or parts of it; and other information related to their Web browser (Real-time bidding protocol - processing the request, 0000; Yan et al., 2009; Olejnik et al., 2014). Accompanying this information, Google's ad exchange incorporates a bidder specific user ID, which implies that different bidders are given different IDs for a same user. Other RTB-based ad exchanges, alternatively, include their own user's cookies.

Upon receiving the bid request, the bidder may identify the user within its own database through the cookie or identifier. This is provided that the cookie-matching protocol has been executed previously for this user. Thanks to such cookie or identifier, the bidder can track them across those Web pages in which it is invited to bid (Ghosh et al., 2015). From those tracked pages, the bidder can therefore build a profile, maybe complementing tracking and other personal data it may have about the user (Google, Cookie, 2016).

The bid price is then set on the basis of the bidder's targeting objectives, that is, whether it aims to target users visiting certain site categories, browsing from a given location, and/or having some specific profile. To evaluate if the ad-impression meets such objectives, the bidder relies on the aforementioned profile and the information included in the bid request. If interested, the bidder submits a price to the ad exchange, which finally, in a last step, allows the winning bidder to deliver the ad to the user. The winning bidder is evidently the highest bidder, but the price paid is the second-highest bid in the auction (Google, Ad exchange, 2018); these so called second-price auctions look after preventing underbidding and overbidding. It is worth stressing that all this process of gathering user data, ad bidding and delivering is conducted in just tens of milliseconds.

3. Data aggregation-driven privacy risks in online advertising

This section examines in depth the potential abuse and privacy risk object of this paper. We emphasize that these issues derive from the capability of DSPs to track and profile users almost effortlessly and at very low cost. More specifically, user privacy in RTB systems is at risk as a result of: (1) user information is shared with third parties by default; (2) this information is not only delivered to the winner of an auction but also to other entities, and (3) there is an apparent lack of control over the abuse of potential malicious listeners.

Some guidelines are stated by the ad exchange (e.g., Google DoubleClick) regarding the use of auction data. Yet there are not known mechanisms to control such abuse from certain DSPs. Next, to illustrate the aforementioned privacy risk, we analyze a publicly available data set containing bid information of a Chinese DSP.

Table 1
Information, items carried in bid requests, here matched to the potential privacy risks derived from their open distribution and aggregation. A user ID is a string that unambiguously identify a user in a given system, e.g., within the ad exchange's domain. An *IP address* or part of it may be used by DSPs as an identifying parameter and to infer the user's *location*. Device or browser *fingerprints* are also carried within bid requests to serve as powerful identifying parameters built on the specific characteristics of a user entity. Bid requests also transport user behavioral data such as *labels* describing the user's preferences or the category of the site visited. Furthermore, the *URL* or *domain* of this site is communicated. Finally, a minimum *bidding price* is usually included in bid requests as a reference for the auctions.

	Identification	Learning of moving patterns	Cross device tracking	Microtar-geting	Habits tracking	Outlier detection
User ID	×					
IP address		×				
User location		×	×		×	
Device fingerprint	×		×		×	
Web browser fingerprint	×		×		×	
Time stamp		×				
User languages	×			×		
User labels				×		×
URL					×	
Content labels				×		
Minimum bid price			×			×

3.1. Bid requests: the tokens leaking personal data

As explained in Section 2, a bid request not only serves to invite DSPs to participate in the auction of a user's impression. A bid request includes a variety of user data in order to provide DSPs with the necessary feedback to decide whether to bid or not for said impression. Then, the interested DSPs send their bids to the ad exchange in order for an auction to be held. Evidently, the success of personalized advertising tightly depends on the granularity and volume of the information shared with DSPs. Sadly, user privacy decreases to the same extent that personalization improves. As an approach to evidence this privacy risk, we here portray the critical information available about the user and included in bid requests. Fig. 3 depicts the aforementioned interactions among an ad exchange and DSPs. Considering that these interactions are carried out for every single user impression, it illustrates the wealth of personal information flowing to potential participants in the bidding process.

Dozens of fields and subfields carry information concerning the context in which a user impression is held (Google, Real-time, 2017). As described previously, users play a leading role in this context. Thus, much of the information carried to fuel the RTB process characterizes them and, particularly, their behavior. First, a bid request may include a *user's ID* that DSPs may use to individuate them and match previously acquired information with the data included in bid requests. A user ID may be a string that unambiguously identify a user in a given system but not in real life, e.g., within the ad exchange's domain. Furthermore, the user's *IP address* (or part of it) is included in bid requests mostly to help DSPs infer *location information* to execute geographically targeted campaigns. IP addresses can also be used as user identifiers, especially now that IPv6 is providing an almost unlimited addressing space.

Additionally, device and web browser *fingerprint data* may be contained in a bid request, as powerful attributes to better identify users. A fingerprint is a set of attribute values that characterize an entity to the point that could individuate it unequivocally. For example, a fingerprint of a network device might be composed by its operating system's name, its version, the list of applications installed and the list of open ports of the device.

Information about the users' *online behavior* may also be included in the bid requests sent to DSPs, e.g., in the form of a list of (user profile)

Table 2
Parameters describing the iPinYou data set. This includes, e.g., the number of bid log records, unique users involved, or the number of Chinese cities reached.

Bid log records	64.746.749
Logs of won bid requests	19.495.976
Unique users	21.264.865
Data attributes in bid logs	24
Ad exchanges	3
Regions	35
Cities	362
User profile tags	44

tags or categories. These categories reveal the preferences of the user whose impression is auctioned, thus are crucial for DSPs when deciding whether to bid or not. Similar tags depicting the content of the website visited by the user might also be delivered to DSPs along with its *URL* or *domain*. Finally, a *time stamp* indicating the date and time of the user's visit, and a reference *bidding price* to inform the minimum value to bid may be provided by bid requests.

Several privacy risks may derive from this personal information, especially when distributed among several intermediate entities such as DSPs, in a position to aggregate and process said information. To start, although user IDs do not identify a user in real life, a combination of the remaining attributes may inequivocally individuate a user (a few demographic attributes have such an identification power Sweeney, 2002). Users' location information could lead an attacker to learn moving patterns of users to then reveal even further details about their daily activities (Mathai et al., 2015). Device and web browser fingerprints may complement this attack by enabling cross device tracking (Brookman et al., 2017). Not only could users and their activities be geographically tracked using data in bid requests, but also their preferences are learned and may reveal sensitive information (Hill, 2012). In fact, pricing information is already a critical aspect that directly discloses the relative importance of a user. Table 1 maps some of these information items to the potential privacy risks derived from their open distribution and aggregation.

3.2. The iPinYou data set

To illustrate the potential misuse of RTB systems and its real impact on user privacy, we analyze a data set that includes bid information released by a well-known Chinese DSP called iPinYou.

The iPinYou data set (Zhang et al., 2014) contains logs of the ad auctions where this DSP has participated. These logs basically carry three types of information for each auction: (1) user data sent by the ad exchange to the DSP in order for the latter to prepare a bid response, (2) the price paid when it wins the auction, and (3) information on whether the user made a click or a conversion as a response from the ad displayed. User data include some of the parameters described in Section 2.1, e.g., an ID of the user that generated the auction, a timestamp, their browser fingerprint (user-agent), their IP address (its first 3 bytes), their location (region and city), the domain and URL visited, and some user tags representing the categories of interest of the user. Additional information involves the ad exchange that held an auction and the price paid by a DSP (not necessarily iPinYou) to won it. The values of some of these attributes, e.g., IP address, URL and domain visited, are anonymized to preserve the privacy of users. It is important to note that this data set contains information related to the bids in which iPinYou participated, excluding the auctions where iPinYou decided not to bid.

This data set was released in 2013 for an open contest on RTB ad pricing. For the purpose of our analysis, we use the version processed by Zhang et al. in Zhang et al. (2014). We aggregate the data from seasons 2 and 3 of the competition (data from season 1 has different fields than the rest) and we examine the data of almost 65 million bid requests sent to this DSP. We find that these bid requests belong to about 21 million unique users. In Table 2 we summarize the most relevant figures of the data set at hand. In order to facilitate the processing of this data, we used a sample of bid requests corresponding to the users having 70 or more log records in the whole data set, yielding almost 6 thousand users with more than 8 thousand log records.

3.3. Privacy risks and abusing context

User privacy risk starts from the capability of an ad exchange to *identify* the user whose impression is being auctioned. The user ID attribute included in bid requests and thus sent to DSPs inequivocally identifies a user within that context. In fact, if this identifier is already known by a DSP, they could match even more information about the user. In addition, recall, e.g., that a few combined demographic attributes may be very identifiable. Consequently, other attributes such as the user's IP address and the device fingerprint (Eckersley, 2010) might make this risk worse. Namely, although not real-life identifiers, user IDs, when combined with other bid request fields of information, might significantly facilitate the work of a privacy adversary in its bid for individuating a victim.

Public IP addresses could by themselves be very identifiable, too. For this reason, only the first three octets are commonly revealed in bid requests, but it is still evident when the address changes. The uniform change of a user's IP address through the day, if a user is tracked across different geographical areas, might unveil movement patterns, which is sensitive information with regard to user privacy (del Prado Cortez and Frignal, 2014). With respect to this, within the iPinYou data set, we found that a great portion of users (about thirty percent) were associated with three or more different IP addresses.

In addition to IP addresses, other attributes with rare attribute values may help adversaries single out users in real life, even more when analyzed in combination with other attributes. For example, people using Linux operating systems and non standard web browsers (e.g., Opera) could excel so much to become easily identifiable outliers. To have an idea of this, in the iPinYou data set, we found only 206 bid requests (out of millions) including user information coupled with the combination of Linux operating system and the Opera browser.

This process of associating a user's unique identity with their interactions enables *tracking*. Then, working in real time, tracking allows advertising entities to "recognize" users during their impressions and ultimately display a personalized ad. However, tracking also enables these entities to join personal information to build individual user profiles (*profiling*). As in other personalization contexts, such tracking and profiling capabilities are supported by the processing of personal information. Nevertheless, within advertising platforms, personal information flows freely, constantly, and abundantly from the ad exchange to DSPs. Thus, a sort of *oversending of personal data* to third parties might be supporting misuse and worsening privacy risks.

The last statement implies that DSPs essentially do not pay for the user information they receive in bid requests, but for the auctions they win on behalf of advertisers. In practice, upon the reception of bid requests (invitations to participate in auctions), a DSP pays just for the auctions it wins, while it receives user data in the rest of bid requests "for free". Clearly, DSPs may take advantage of the ad exchange's tracking resources at a very low cost. This fact is evidenced in Fig. 4, where we depict the amount of users whose information has been paid by the iPinYou DSP. To illustrate the amount of information potentially collected for free, we can see in this figure that, *for about 55% of the users, this DSP has not paid for any of their bids*. From this we can infer the potential abuse of a third party and the exacerbated risk for the privacy of users if multiple DSPs exhibit a behavior not oriented to participate in auctions, but to take advantage of the large amount of user data distributed by an ad exchange. We would like to stress, however, that this percentage of users tracked for free might be just a lower bound: the released data set does not include the auctions where iPinYou did not bid, but from which it could have received numerous user data costing nothing.

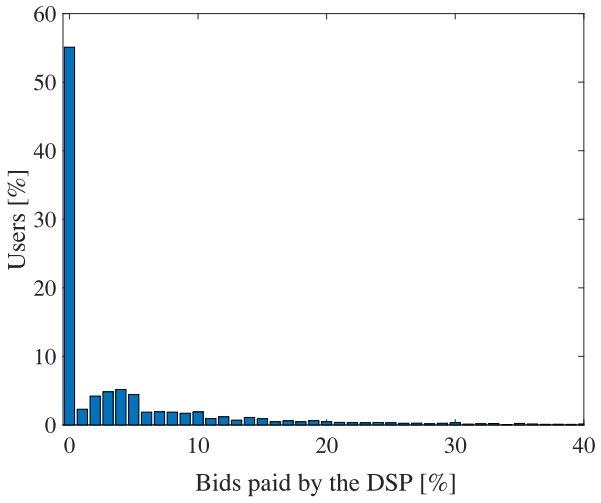

Fig. 4. Percentage of users whose information has been paid by the iPinYou DSP. For about 55% of the users, none of the bid requests triggered from their impressions were paid by the DSP, i.e., the DSP did not pay anything for the auctions held for 55% of the users.

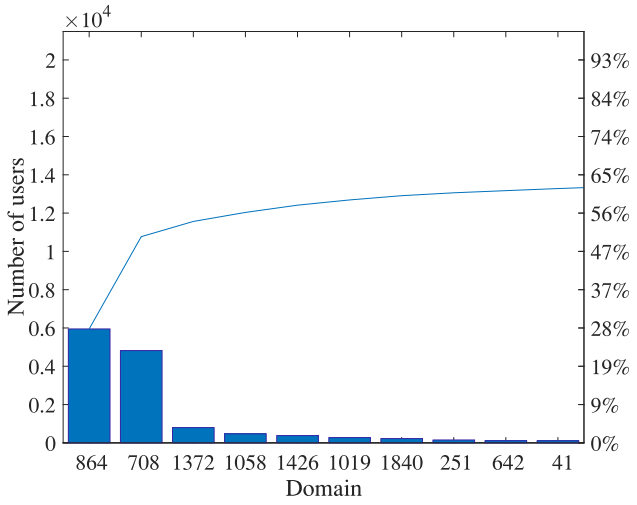

Fig. 5. Amount of users tracked by the most popular domains in the iPinYou data set.

In an attempt to prevent this abuse, ad exchanges clearly prohibit DSPs to use the information in bid requests when a corresponding auction has not been won (Google, Google, 2017). It is also not allowed to use this information for applications other than the ones related to online advertising. However, enforcing such use is hard when the information has already been distributed to third parties; and when, due to an increasingly complex advertising ecosystem, more and more entities are included to outsource specific functions in the demand side (e.g., trading desks).

Data aggregation performed by intermediate entities brings another privacy jeopardy in online platforms. Not only ad exchanges, the core of ad distribution, but also DSPs and even publishers are in a position to *concentrate user data* (Estrada-Jiménez et al., 2017). As expected, in the iPinYou data set, user tracking is concentrated in Google's DoubleClick ad exchange. Furthermore, we found that more than fifty percent of the users in the iPinYou data set would be tracked by only two publishers, probably related to the most popular websites in China (Fig. 5). In other words, having recognized at least 2063 publishers in this data set, less than 0.1% of them concentrates the tracking of more than 50% of the involved users. Powerful *tracking* capabilities are then held in very few hands.

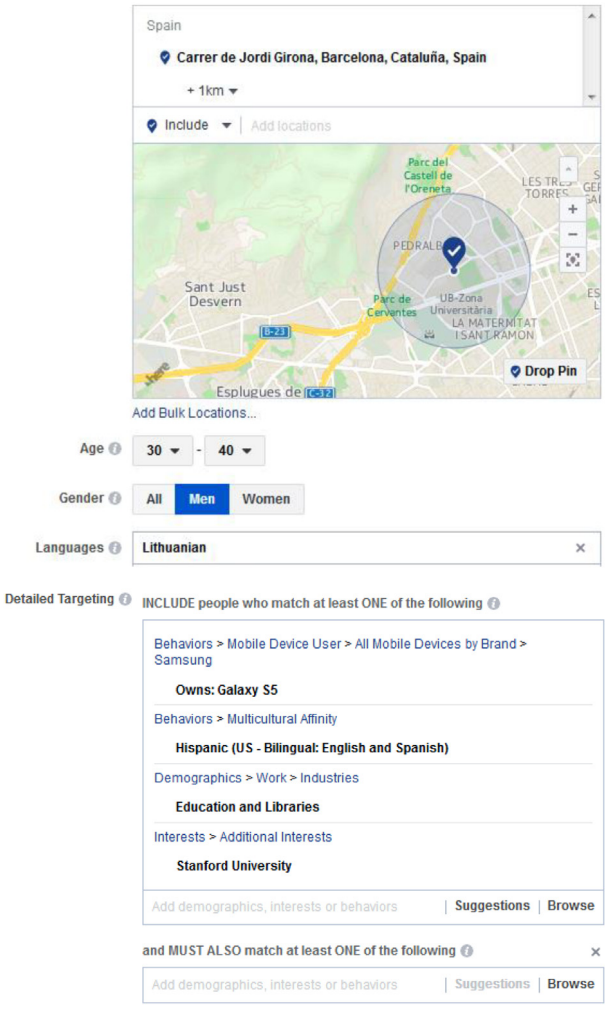

Fig. 6. Interface for advertisers to select an audience for a campaign in Facebook. Its very granular options allow a great power to microtarget users.

Not only the easiness and openness of data collection is threatening the privacy of Internet users, but also the level of detail of the data. The *granularity* of the user data held by these entities has given rise to powerful capabilities of microtargeting. These capabilities have derived in tools to select audiences that may enable even advertisers to target groups of users with great precision (Korolova, 2010). In Fig. 6, we show an interface offered by a social network and a DSP to choose an audience for better ad targeting.

Finally, due to the pervasiveness of online advertising, it is not hard to comprehend its wide reach in the population. The idea that the advertising ecosystem might be collecting information related to large masses of people is reflected in the iPinYou data set. More precisely, based on the user ID and region attributes of the records from this data set, we observed that large portions of the population of important Chinese regions would have been tracked. For example, this DSP (iPinYou) would have information of more that 5% of the population in regions such as Beijing, Guangdong, and Shanghai (see Fig. 7). Considering that, in this case, most of the user data must be "ceded" as input to DSPs for their bidding decisions, gathering such bulks of information seems a very good deal for them. However, this is not good news for the privacy of users, who are probably being observed en masse.

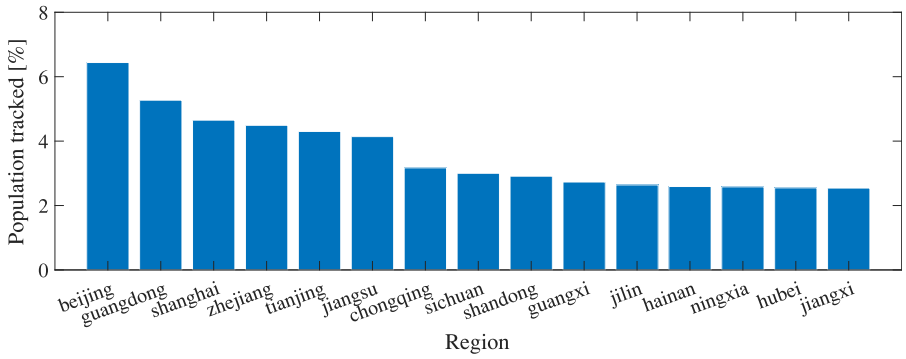

Fig. 7. Population tracked by online advertising entities along different regions of China, as observed in the iPinYou data set.

4. A system for the controlled distribution of bid requests (in RTB)

We propose a system that aims to reduce the oversending of personal data to DSPs, thus ultimately providing some privacy to users in RTB systems. This is done by regulating the distribution of bid requests among intermediary entities such as DSPs or trading desks. Conceptually, this objective could be reached, to a certain extent, by reducing the number of DSPs to which bid requests are sent, thus lowering the instances where user data is aggregated. Naturally, from a practical perspective, our solution is conceived to be implemented within the ad exchange infrastructure since it is the entity in charge of sending bid requests to DSPs when a potential ad impression arises. The proposed system determines, in real time and adaptively, the specific participants of a given ad-space auction, at the cost of some processing overhead at the ad exchange and a potential reduction in revenue incurred by a smaller number of participants in auctions. Being revenue the *raison d'être* of ad exchanges, a trade-off will arise with data distribution control, but with an adequate balance, we shall show that reasonable guarantees can be provided while keeping relatively high profits.

Unlike many of the privacy techniques proposed in the literature for online advertising, a change in the distribution model of bid requests does not entail an important modification of the advertising ecosystem.

4.1. Adversary model

Our technique builds upon the principle of a selective distribution of bid-request information (containing user sensitive data) among potentially interested DSPs. Consistently with this principle, we assume an adversary model in which the bid requests sent by an ad exchange are passively observed and maliciously aggregated by a group of intermediary entities.

We must stress that this adversary model assumes that privacy risk comes from the exploitation of user profiles built from the *aggregation* of user data. Namely, the user data in a single bid request would not entail a significant privacy risk since by itself reveals only a snapshot of the preferences, behavior and demographics of a user at a certain point in time. However, the more user data is aggregated the richer are the resulting profiles, and the higher is the corresponding privacy risk.

As argued in Section 3, RTB-based ecosystems still provide fertile ground for privacy abuse. One of the reasons is the relative ease with which user data can be collected by intermediate and authorized entities such as DSPs and other smaller subsidiary entities (e.g. trading desks). Especially the latter, sometimes being really small companies, are becoming capable of tracking users at a very low cost (or none) and without deploying an important infrastructure. Thus, a privacy gap arises when they are given easy access to an ever-growing universe of aggregated personal data. We propose a system to bridge this gap by penalizing said kind of tracking when it violates the norms established by ad exchanges.

Table 3
Behaviors of DSPs, with regard to their participation in bid auctions, that may go against ad exchanges policies and also in favor of the violation of users' privacy.

Behavior	Description
Silent	A DSP not participating in auctions, and thus not answering bid requests, may be misusing the RTB infrastructure by collecting and exploiting the user data carried in these messages. Ad exchanges recognize said risk when forbidding DSPs using the data for which they have not paid in their policies (Google, Google, 2017). Although this also gives rise to privacy concerns, no further control is made.
Loser	A DSP that looses too many auctions is also suspicious of abusing the RTB infrastructure against privacy. Bidding to lose is possible since bid requests sent to DSPs include information about the minimum price of the user impression auctioned. Just by bidding below the minimum would enable DSPs or related entities to receive user data for free.
Stingy	A DSP bidding too low might be looking for receiving user data when no pricing information is received in bid requests, thus trying to inappropriately exploit the RTB logic in detriment of privacy.

As a note, abusing of such tracking is against the terms of use of the main ad exchanges, which forbid DSPs taking advantage of data for which they have not paid. For example, according to Google DoubleClick Ad Exchange (AdX) Buyer Program Guidelines (Google, Google, 2017), some of the policies that buyers must adhere to are listed next:

- Buyers and any third party to which they provide access to the ad exchange must adhere to the policies.
- The inventory purchased cannot be sold to another sales channel.
- Bid data cannot be used for purposes different from buying on the ad exchange.
- Unless a buyer wins a given impression, it must not use bid data for that impression to create user lists or profile users.

In brief, neither DSPs nor outsourced entities such as trading desks are allowed to exploit bid data coming from an ad exchange, unless they have paid for such data after winning a given auction. Some of the behaviors that might go against ad exchange's policies are described in Table 3.

4.2. Bid request distribution model

As noted in Section 2, the visit of a user to a website that holds an ad space generates a so-called ad impression. Then, an ad exchange auctions said impression among all the available DSPs. To support the bidding decision of DSPs, the ad exchange distributes among them bid requests containing some user data.

We propose reducing the number of DSPs to which a bid request is sent, in order to penalize misbehaving DSPs and to promote privacy.

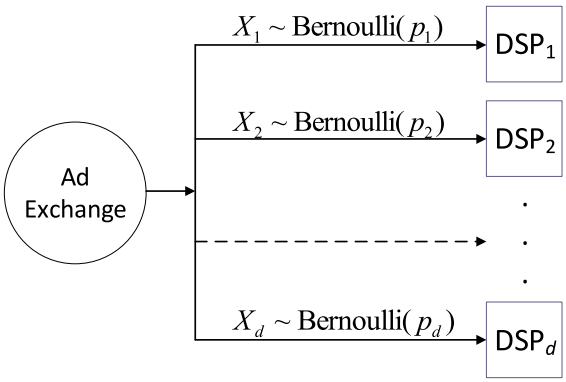

Fig. 8. A depiction of the bid request distribution model we propose for the ad exchange. We model said distribution as the random draw of d Bernoulli trials (represented with d Bernoulli r.v.'s: $X_1, \ldots, X_d$), being d the number of DSPs available. Each r.v. characterizes an experiment with a boolean-valued outcome and a success probability p_i.

To model the distribution of bid requests, we rely on the Bernoulli distribution that characterizes a discrete probability distribution of a random variable whose value is *true* with probability p and *false* with probability $1 - p$. This is the same behavior of the outcome of sending bid requests to DSPs; they will receive requests if behaving well or will not receive bid requests (penalized) if being dishonest. Accordingly, being d the number of DSPs available in a given moment, we model the distribution of bid requests among them as the execution of d Bernoulli trials (or experiments).

These trials can be represented as d independent, identically distributed Bernoulli random variables (r.v.'s) $X_1, \ldots, X_d$, each of which characterizes an experiment with a boolean-valued outcome and a success probability p_i, with $0 \leq p_i \leq 1$. Therefore, when auctioning a user impression, the ad exchange shall send a bid request to DSP_i with probability p_i and shall not do it with probability $1 - p_i$. A given ad exchange's distribution strategy will be defined as the tuple $p = (p_1, \ldots, p_d)$ of the probabilities of sending a bid request to each of the d available DSPs. In Fig. 8, we depict this distribution model for a given user impression.

As introduced previously, to control misbehaving DSPs, we propose bounding the number of DSPs that receive a bid request from the ad exchange. Intuitively, the less the number of receiving DSPs, the higher the level of user privacy. To do it, we introduce a *data distribution control parameter* defined as the average number of DSPs that will receive a bid request, α, with $0 \leq \alpha \leq d$. Namely, in our system, the number of recipient DSPs is bounded to the value of α. Clearly, the number of invited DSPs, being a sum of independent Bernoulli trials, follows a Poisson binomial distribution with mean $\sum_i p_i$. Consequently, our measure of privacy, the average number of participating DSPs, can be computed straightforwardly as $\alpha = \sum_i p_i$.

4.3. A system to balance the number of DSPs invited and ad revenue

Section 4.1 described the adversary model we tackle in this work. In particular, we mentioned that DSPs might go against the policies of the corresponding ad exchange by exploiting a uncontrolled distribution of bid requests. Nevertheless, implementing these policies is by no means a simple task because ad exchanges have no control over the internal dynamics of buyers' data infrastructures. In any case, they do have the capability to regulate how bid requests are distributed to buyers. Then, it is by shaping such distribution of user data, according to the behavior of DSPs, that we propose to bound the amount of information (bid requests) sent to DSPs, with the ultimate aim of enhancing privacy.

Intuitively, a distribution strategy that restricts the recipients of bid requests will reduce the revenue of an ad exchange. Accordingly, we

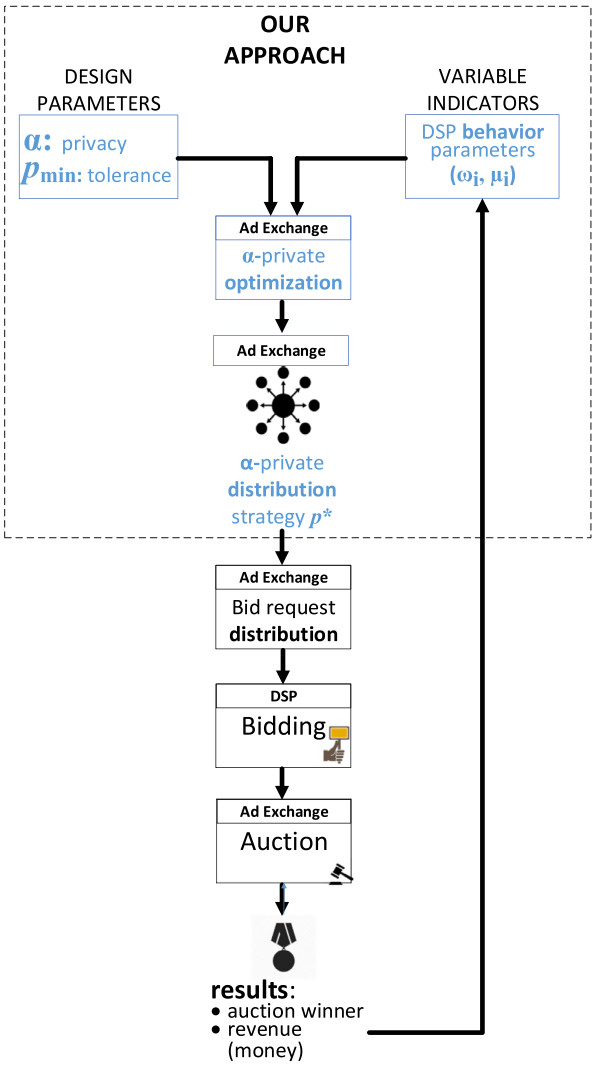

Fig. 9. Methodology implemented to assess our bid request distribution strategy. This flowchart also illustrates how our system integrates to the ad exchange's auctioning system as described in Section 4.3 (the blocks in blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

define a metric of said revenue, in a given auction, as the product of three variables ω_i, μ_i and p_i, for $i = 1, \ldots, d$. Note that maximizing this measure of revenue would imply maximizing the real revenue, according to the distribution model proposed in this work. Both ω_i and μ_i are system parameters taking values in $\mathbb{R}$ and could be interpreted as reputation metrics of a DSP i. A DSP that behaves according to the ad exchange's rules will generate a reasonable revenue and thus will have a better reputation than other DSPs that break the rules. For each DSP i, we define the *winning rate* ω_i as the rate of won bids with respect to the number of bid requests received up to a given instant. Weighting by winning rate enables our model to discourage a potential misuse of the bid request distribution model in online advertising. A DSP that almost always loses is probably just "listening" for user data to tamper with their privacy, thus deceitfully exploiting the online advertising ecosystem. In our proposal, the economical contribution of a DSP winning only a few auctions, even bidding high, will be weighted by its poor wining rate in order to counteract its behavior against privacy.

In addition, we define μ_i as the average money spent by a DSP up to a given instant, that is, the amount of money paid for the won bids divided by the number of bid requests received (we call it *average money*

spent). Next, for the sake of simplicity, we shall refer to the product of ω_i and μ_i as r_i. For the sake of clarity, please refer to Table 4 to find the notation used in this analysis.

We shall denote by p the strategy of distributing bid requests where p_i, already defined in Section 4.2, could be seen as the percentage of traffic sent to DSP i. Evidently, the higher the winning rate ω_i and the average money spent of a DSP$_i$, μ_i, the more likely it is to win an auction (thus having a higher "reputation"). Naturally interested in reaching the maximum possible revenue, an ad exchange will try to send a bid request to the DSPs with the highest product r_i. However, for DSPs with low r_i (i.e., showing bad behavior), in order not to completely eliminate their opportunity to participate in auctions, we will impose a *tolerance* parameter, i.e., a lower bound on p_i, denoted by $p_{\min} > 0$. Thus, with $p_{\min} \leq p_i$, we try to guarantee, for said DSPs, the chance to improve their behavior (reputation) in the future.

According to the justifications in Section 4.2, in our approach we use the parameter α to bound the number of DSPs invited to bid (invitation rate) and that will receive information from the ad exchange. Put another way, α could also be interpreted as a measure of the suppression of bid requests to DSPs. Consistently with this bound, we define a revenue-invitation rate function

$$\mathcal{R}(\alpha) = \max_{\substack{p \\ p_{\min} \leq p_i \leq 1 \\ \sum_{i=1}^{d} p_i = \alpha}} \sum p_i \omega_i \mu_i, \tag{1}$$

which characterizes the optimal trade-off between *revenue* $\mathcal{R}$ and the number of DSPs invited to bid α. From this expression, we aim at finding an optimal strategy of bid request distribution p^*, that satisfies an average participating DSPs α while maximizing the resulting ad exchange's revenue $\mathcal{R}$. Note that this expression establishes a strict restriction (it must be fulfilled) regarding the limit of DSPs that will receive invitations by the ad exchange (α), while its revenue is maximized in a best-effort sense. We would like to stress that the priority in our definition is given to meet this bound, i.e., to prevent abuse and mitigate the privacy risk.

Although we propose modulating (or restricting) the distribution of bid requests to DSPs such that more privacy is provided to users, this does not necessarily imply that ad exchanges lose control over user data. In fact, our approach would leave unchanged the internal logic within ad exchanges for the sake of simplicity and applicability; this includes how user data is collected and processed by ad exchanges. Our proposal focuses rather on the flow of user information from ad exchanges to DSPs, since unnecessary interactions threatening privacy may arise in such data sharing context.

Having presented the main parameters and indicators of our system, we summarize in the next list of steps the actions that the ad exchange must perform to integrate our approach to the auctioning system. Also Fig. 9 illustrates this integration and later on is used to describe the evaluation methodology of our bid request distribution strategy.

Step 1: Set the design parameters of the system: a bid request distribution (privacy) parameter α and a tolerance parameter $p_{\min}$.
Step 2: For each d DSPs, compute and update their variable behavior (reputation) indicators based on their win rate and money spent (ω_i, μ_i).
Step 3: Find an optimal distribution strategy of bid requests $p^* = (p_1^*, \ldots, p_d^*)$ that balance a measure of privacy with revenue.
Step 4: Send bid requests (invitations) only to the α DSPs showing the best behavior indicators.
Step 5: Receive bid responses and auction the user impression.

Table 4
Description of the main variables used in our notation.

Symbol	Description
d	Number of DSPs available in our scenario
p	Tuple representing the ad exchange's distribution strategy. Its elements are the probabilities of sending a bid request to each DSP
α	Average number of DSPs that will receive a bid request, i.e., the average number of DSPs to be invited to the auctions
ω_i	Rate of won bids with respect to the number of bid requests received by a DSP i up to a given instant
μ_i	Average money spent by the ith DSP up to a given instant
r_i	The product of $\omega_i \mu_i$
$p_{\min}$	Lower bound on p_i that guarantees an opportunity to participate in auctions for all DSPs
$\mathcal{R}(\alpha)$	Function modeling the revenue of an ad exchange as a function of the privacy parameter α

4.4. Optimal strategy for the distribution of bid requests

In this section, we shall analyze the revenue-invitation rate function (1) defined in Section 4.3, and present a closed-form solution, albeit piecewise, to the maximization problem. We shall suppose, without loss of generality, that

$$r_1 \geq \cdots \geq r_d. \tag{2}$$

Also, for $k = 1, \ldots, d$, we define a sequence of thresholds α_k as $k(1 - p_{\min}) - d\, p_{\min}$.

Lemma 1. *For any $k = 1, \ldots, d$ and any $\alpha \in [\alpha_{k-1}, \alpha_k]$, the solution to (1) is the distribution strategy*

$$p_j^* = \begin{cases} 1 & , \ j = 1, \ldots, k-1 \\ \alpha - p_{min}(d-k) - (k-1) & , \ j = k \\ p_{min} & , \ j = k+1, \ldots, d \end{cases} \tag{3}$$

and the corresponding maximum revenue yields

$$\mathcal{R}^*(\alpha) = r_k \alpha - r_k p_{min}(d-k) - r_k(k-1) + \sum_{j=1}^{k-1} r_j + p_{min} \sum_{j=k+1}^{d} r_j. \tag{4}$$

Proof. The existence and uniqueness of the solution is a consequence of the fact that we maximize a continuous function over a compact set.

From the assumption (2), it follows intuitively that for an $\alpha < 1$ the solution consists in sending a bid request to the first DSP, i.e., to the DSP having the maximum product $\omega_i \mu_i$. However, the condition $p \geqslant p_{\min}$ ensures the resource, α, must be distributed across all other DSPs, so that all participants can have a chance to receive a bid. The amount of α to be distributed is clearly $d\, p_{\min}$ and hence the remainder $\alpha' = \alpha - d\, p_{\min}$ is the resource to be assigned among the d DSPs. Therefore, $p_{\min} \leqslant \frac{\alpha}{d} \leqslant 1$.

Following the same intuitive principle described above, we proceed to examine the distribution strategy of the remaining α'. Note that, below, all the expressions in terms of α' can be recast in terms of α' on account of $\alpha = \alpha' + p_{\min}$. For notational convenience, define $p' = p - p_{\min}$.

Case 1. $0 \leq \alpha' \leq 1 - p_{\min}$.

Observe that, in this case, any feasible $p' = (p'_1, \ldots, p'_d)$ satisfies

$$p'_1 r_1 + \cdots + p'_d r_d \leq (p'_1 + \cdots + p'_d) r_1 = \alpha' r_1,$$

which implies that the optimal distribution strategy consists in assigning the whole α' to the first DSP, that is $p_1^* = \alpha'$ and $p_i^* = 0$ for $i \neq 1$. More compactly,

$$p^* = (\alpha - (d-1)p_{\min}, p_{\min}, \ldots, p_{\min}),$$

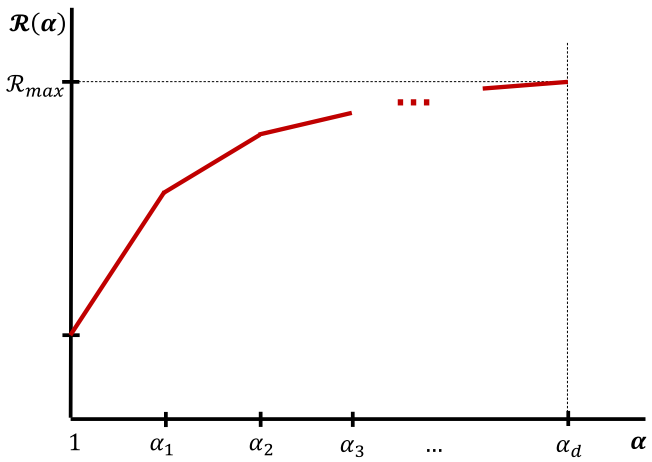

Fig. 10. Conceptual plot of the revenue-invitation rate function. $\mathcal{R}(\alpha)$ is a nondecreasing function defined piecewise. From the labeling assumption (2), the slopes of $\mathcal{R}(\alpha)$ ($\omega_k \mu_k$) decrease as α grows.

by virtue of $\sum p_i^* = \alpha = d\,p_{\min} + \alpha'$.

Case 2. $1 - p_{\min} < \alpha' \leq 2(1 - p_{\min})$.

This case follows in an exactly analogous manner as the previous case and leads to the optimal strategy

$$p^* = (1, \alpha - (d-2)p_{\min}, p_{\min}, \ldots, p_{\min}).$$

Case k. $(k-1)(1-p_{\min}) < \alpha' \leq k(1-p_{\min})$.

By generalizing our analysis for the kth case, written in terms of α as $(k-1)(1-p_{\min}) + d\,p_{\min} < \alpha \leq k(1-p_{\min}) + d\,p_{\min}$, it is straightforward to check that the optimal distribution strategy is

$$p^* = (1, 1, \ldots, \alpha - (d-k)p_{\min}, p_{\min}, \ldots, p_{\min}).$$

Simple algebraic manipulation leads to expression given in the lemma. The derivation of the maximum revenue follows immediately from the optimal strategy as $\mathcal{R}^*(\alpha) = \sum_{j=1}^{d} p_j^* r_j$. □

The optimal bid request distribution strategy in Lemma 1 is interpreted as follows. Given the first condition of our problem (1), $\sum_{i=1}^{d} p_i = \alpha$, the average number of DSPs α to which requests will be sent, has to be distributed among the d available DSPs. According to (3) in Lemma 1, the first $k-1$ DSPs (the ones bidding more and winning more auctions) are by default sent a bid request; the probability of sending them the request is 1. The last $d-k$ DSPs (the ones bidding less and winning less auctions), however, are sent a bid request with a minimum probability $p_{\min}$ according to the first condition of our revenue-invitation rate function (1). Finally, the kth DSP is sent a bid request with the remaining probability $\alpha - p_{\min}(d-k) - (k-1)$. This strategy can be easily explained as a resource allocation problem where α (the "resource to be distributed") is shared among DSPs according to their good behavior, with the aim of satisfying a given bound α.

Next, we proceed to analyze very briefly some properties of the revenue-invitation rate function (4). It is immediate to check the function is piece-wise linear with slopes $\omega_k \mu_k$. Given that this product will never be negative, neither will be the slope of $\mathcal{R}(\alpha)$ and, consequently, it is easy to see that $\mathcal{R}(\alpha)$ is nondecreasing. We cannot characterize $\mathcal{R}(\alpha)$ as increasing because there is the possibility that $\omega_k \mu_k$ is zero. Under the same reasoning, it is immediate to check the monotonicity of this function. Also, from Lemma 1, it is routine to verify the continuity of $\mathcal{R}$ on the interval $\alpha \in [1, d]$. To show the convexity of $\mathcal{R}(\alpha)$, note again that for each k and $\alpha \in (\alpha_{k-1}, \alpha_k]$, the optimal tradeoff function has slope r_k (or $w_k u_k$). From the labeling assumption (2), it follows immediately that $\mathcal{R}(\alpha)$ is defined by the decreasing sequence of positive slopes $r_1, \ldots, r_d$ (or $w_1 u_1, \ldots, w_d u_d$) and therefore is concave. Fig. 10 conceptually illustrates these properties and the results of Lemma 1.

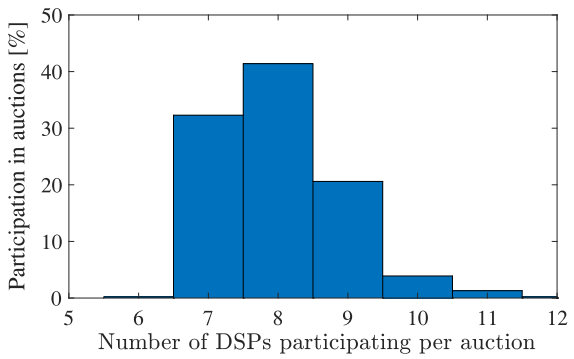

Fig. 11. Number of bid requests (invitations to DSPs) sent per auction for our experiment, with $\alpha = 8$.

From the plot in Fig. 10 we can observe that the behavior of the DSPs (graphically depicted through the slopes r_k) determines that the losses in revenue of the ad exchange could be rather low. Namely, the higher the slope r_k (i.e., the better the behavior) of the first DSPs, the faster $\mathcal{R}(\alpha)$ approaches the ideal revenue $\mathcal{R}_{\max}$. This would entail a more controlled and potentially private bid request distribution (since fewer DSPs would be involved) while not significantly impacting the revenues of the advertising ecosystem. The notation used in this work is summarized in Table 4.

5. Experimental evaluation

Next, we empirically evaluate the solution proposed in Section 4. We describe our experimental methodology and outline the scenario simulated to reproduce the bidding process performed by an ad exchange and a set of DSPs. This allows us to investigate the effect of modifying the bid request distribution model with the aim to enhance user privacy. Our analysis also contemplates measuring said impact in the revenue of the ad exchange.

Since our proposal is to reduce the potential ad buyers to which user data flows, an impact is expected on the revenue obtained by the online advertising ecosystem from these bidders. In particular, given the importance of the advertising business model for the operation of the Internet, we need to show that our proposal does not significantly affect said business model. Accordingly, when applying our strategy, we expect a reduction on the revenue for the ad exchange. However, supported by the optimization approach described in Section 4.3, we also need to verify that this loss in income is acceptable in light of the benefits of a more privacy-respectful system.[3] Furthermore, we shall also verify if, as a result of our multicast solution, the misuse of RTB systems by some DSPs can be effectively addressed.

5.1. Experimental methodology

The proposed solution affects the interaction among an ad exchange and associated DSPs. Recall from Section 2 that DSPs act on behalf of advertisers and thus as bidders (buyers) in the auctions organized by an ad exchange. In order to invite DSPs to participate in a given auction and to provide them with the necessary feedback, the ad exchange distributes bid requests among them, including detailed data about the user whose impression (and corresponding ad space) shall be auctioned. This distribution of user data is adjusted in our approach with the

[3] As commented in Section 1 a great portion of users use ad blockers for privacy reasons. If the proposed system meets the requirements of these privacy-aware users for some α, the loss in revenue due to our multicast strategy (instead of the current broadcast approach) may be more than compensated by the gains of these non-blocking users.

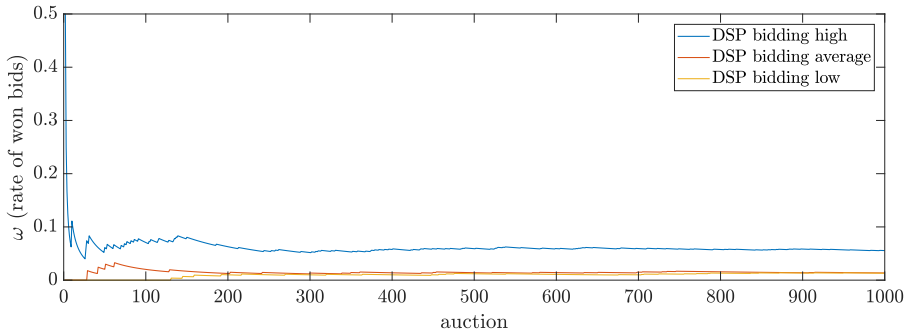

Fig. 12. Rate of won bids for different DSPs behaviors in our experiment. We use $\alpha = 8$ and $\lambda = 0.05$.

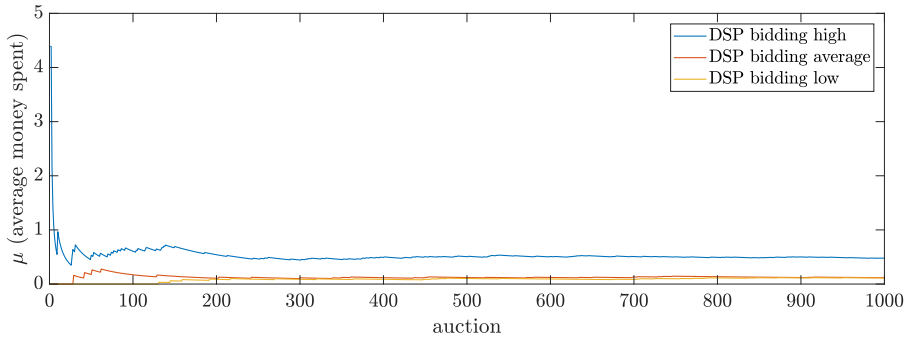

Fig. 13. Average money spent according to different DSPs behaviors in our experiment. We use $\alpha = 8$ and $\lambda = 0.05$.

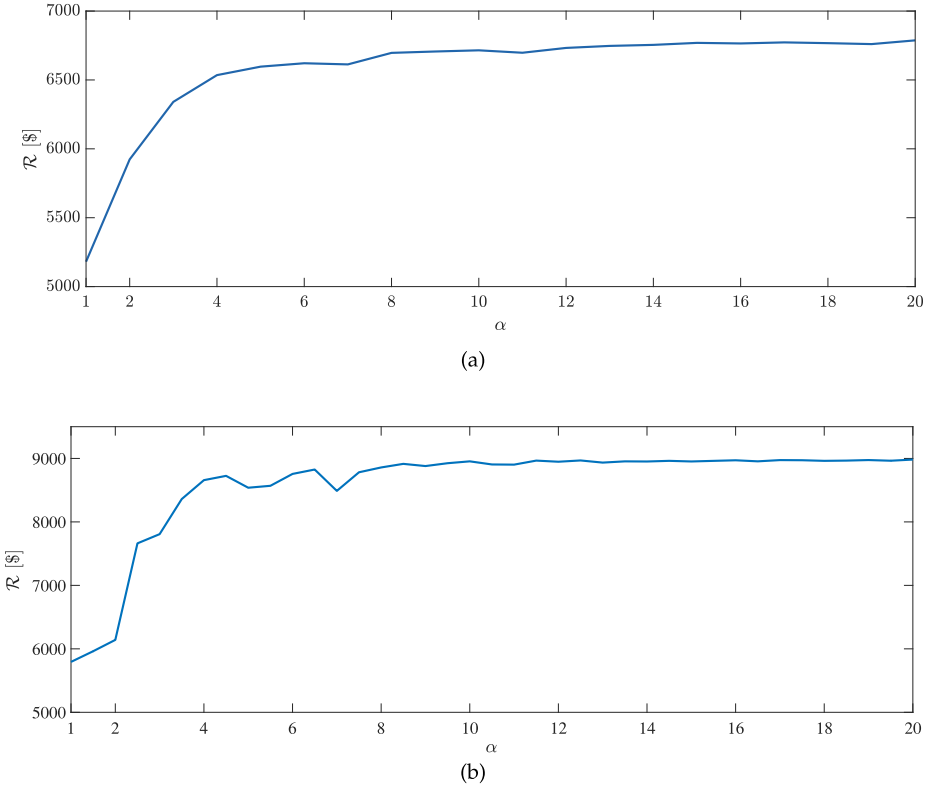

Fig. 14. Revenue obtained by the system for different values of α. For these experiments we use $\lambda = 0.05$ and a set of 20 DSPs so we make α vary from 1 to 20. Experiments are made following two different random distributions when generating the bid requests: (a) uniform and (b) Gaussian.

objective of preventing dishonest behaviors of data collection and thus trying to preserve privacy.

To validate our mechanism, we configure an auctioning scenario that reproduces this behavior, through a Matlab simulation. In this scenario, considering a distribution control parameter α, an ad exchange enables a number of DSPs to participate in each auction, while optimizing its revenue. The main elements of this setup are depicted below.

In our experimental methodology, we simulate real-time auctions in which a variety of DSP types may bid. In order to deploy a more realistic setup, we consider three types of DSPs according to the more likely value of their bids: DSPs bidding high, low, and average. For each auction, the bids from every DSP are randomly sampled from a range of values reflecting these behaviors. For our experiments, such bids are generated following both uniform and Gaussian distributions.

After bids are generated probabilistically at every time instant, an ad exchange instance holds an auction and determines the winner DSP (the one with the highest bid). In line with our privacy proposal, for every auction, not all available DSPs are "invited" (i.e., not all DSPs are sent bid requests), but a number of them, according to the parameter α. Thus, the corresponding activation of DSPs to participate in every auction is enabled by the optimized distribution strategy defined in (3). The strategy depends on two parameters specific to the historical operation (behavior) of each DSP (winning bid rate, and average money spent) and on the privacy parameter α defined by design. Consistently, said parameters of each DSP are calculated before an auction to be used as a kind of reputation metric that fuels the private bid request distribution strategy. Fig. 9 depicts this methodology implemented through a simulation using Matlab R2017a.

After simulating one thousand auctions, we compute the total revenue of the ad exchange by summing all the money effectively spent by the bidders that won at every time instant.

To evaluate if our approach is feasible, we need to examine to which extent it may impact the ad exchange's revenue, which turns to reflect the revenue of the whole advertising ecosystem. Recall that online advertising is said to be supporting the current Internet free business model. Thus, at least for now, this kind of solutions should not significantly tamper with the current ad distribution model since it could negatively affect the economy of online advertising platforms.

5.2. Results

We set up a scenario with twenty DSPs: seven bidding high, seven bidding low, and six bidding between high and low (an average value). Then we simulate an ad exchange instance holding a thousand auctions. Our distribution control strategy is enforced with $p_{\min} = 0.05$ and with $\alpha = 8$. That is to say, to prevent abuses and preserve privacy, bid requests are distributed among eight DSPs in average (those with better behavior) and not among the twenty available. Furthermore, a minimum of 5% of bid requests are distributed among these eight DSPs in order to guarantee all them will participate at some point.

In Fig. 11 we represent the number of DSPs, out of the 20 available, that participate in each auction of our experiments. As expected, this histogram confirms that the number of DSPs participating per auction is 8 in average (the value of α).

Then, we also use Figs. 12 and 13 to characterize the participating DSPs in terms of the rate of won auctions (ω) and the average money spent (μ), respectively. These parameters are measured at every auction, with respect to all the previous auctions. We depict the values to describe the behavior of three DSPs, one from each category. Evidently, these figures show how DSPs with a more desired behavior (bidding higher or spending more) present better indicators ω and μ.

Additionally, we assess the effects of our mechanism on the revenue of the ad exchange. For this, we perform a set of experiments using different values for the parameter α, from 1 to 20 (i.e., we simulated a round of 1000 auctions for each value of α). As α represents the average number of DSPs to which bid requests are sent from the ad exchange, the results from our experiments reveal the impact of the value of this parameter on the total revenue obtained. This impact is illustrated in Fig. 14, for the two different strategies for generating bid requests (uniform and Gaussian distributions). First, note how the revenue increases with the value of α, consistently with the tradeoff commented in Section 4 and depicted in the conceptual plot in Fig. 10. In addition, when $\alpha = 20$, the maximum revenue is reached because, in practice, no control mechanism is applied when all the available DSPs are activated to receive bid requests. Remarkably enough, the revenue when $\alpha = 8$ and onwards is pretty close to the revenue when $\alpha = 20$. Actually, in those cases, revenue is less than 1% lower than the maximum obtained when our strategy is not applied. The importance of this result lies in that the bid request distribution control enables certain privacy guarantees that can be enforced while having a very small impact on the ad exchange's economic benefit. The results observed in these experiments, however, are certainly tied to the specific behaviors assumed for the DSPs. As a matter of fact, our theoretical analysis of the trade-off between revenue and data distribution control found that $\mathcal{R}(\alpha)$ depends on the sequence of slopes $\omega_i\ \mu_i$. The higher the slopes of the first DSPs (sorted according to (2)), the fewer the average number of DSPs needed to obtain revenues close to $\mathcal{R}_{\max}$. On the extreme, the case $\omega_i\mu_i = \omega_j\mu_j$ for all $i, j = 1, \ldots, d$ yields a straight line, which represents the worst scenario in terms of ad revenue.

Finally, we are interested in seeing how our parameter α is capable of regulating the behavior of DSPs. For this, we conduct an experiment with a setup of 3 DSPs, each behaving differently (bidding high, low and average). We simulate a thousand auctions and apply our privacy mechanisms for different values of α, from 1 to 3. We measure the rate r of won auctions with respect to the requests (invitations) received by each DSP from the ad exchange. This rate could be interpreted as a measure of the goodness of the behavior of DSPs as stated in Section 4. A DSP bidding higher will win more bids and spend more money. Accordingly, the higher the rate of won bids, the more desirable is the behavior of a DSP. Conversely, r could also be seen as a measure of the abuse committed by a DSP against the privacy of a user, since a low rate of won auctions (low r) would entail a DSP receiving user data information without paying for it.

The results of this experiments are illustrated in Figs. 15 and 16, where we depict the evolution of the rate of won auctions of different types of DSPs. Respectively, we plot the results obtained from using two different strategies to generate bid requests (uniform and Gaussian) for each type of DSP. First note that, in this context, $\alpha = 20$ represents the case where the ad exchange sends bid requests to all available DSPs, so there is no control strategy applied. In this case, we see that DSPs bidding low have low rates of won auctions, which would imply that they are taking advantage of the advertising system. However, if we analyze the value of this rate as α decreases, we observe that the rate r increases for each type of DSP, which suggests a successful adjustment of the behavior of DSPs. In general, thus, it makes sense to maximize the benefits of the ad exchange subject to a restriction by distribution control (privacy) since the rates of won bids shall improve for small α.

6. Discussion

The big picture of privacy in the online advertising ecosystem

The "hyperconnection" of people to the Internet is making them widely traceable by the providers of third-party applications that enable the collection of personal data. Among such providers we find online advertising platforms, which might be building a mass surveillance structure due to several reasons. First, the presence of advertising online is so *massive* that both the tracking of users and the collection of their data are real-time and ongoing processes. Namely, personal data is continuously leaked to the advertising infrastructure as users browse the Web.

Besides, in this same direction, we have verified that the *reach* of advertising entities is pretty wide. It is a fact that advertisements "follow us" wherever we surf the Internet. As a consequence, user information is regularly processed in bulk and indiscriminately, always in the name of greater personalization. Furthermore, the user information ceded to third parties during ad auctions is extremely *granular*. This facilitates identification of users, e.g., by using identifying attributes (such as IP addresses) or combining them to build a fingerprint. Due to granularity,

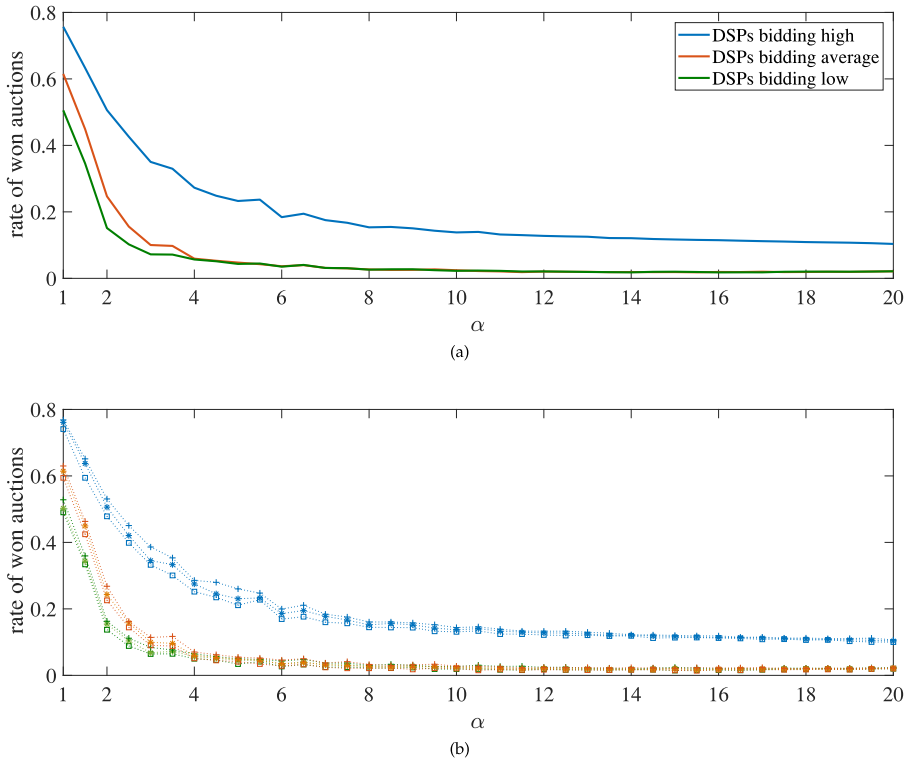

Fig. 15. Evolution of the rate of won auctions for different values of α (from 1 to 20 in steps of 0.5) and $\lambda = 0.05$. We consider 20 DSPs with different behaviors (bidding high, average and low). For each value of α, we repeat the experiment 20 times. The results are depicted averaged in (a) for each type of DSP. In (b), we illustrate results of percentiles 3 (using '+'), 50 (using '*') and 97 (using '□'). For these experiments, bid requests for the different types of DSPs are generated using a uniform distribution.

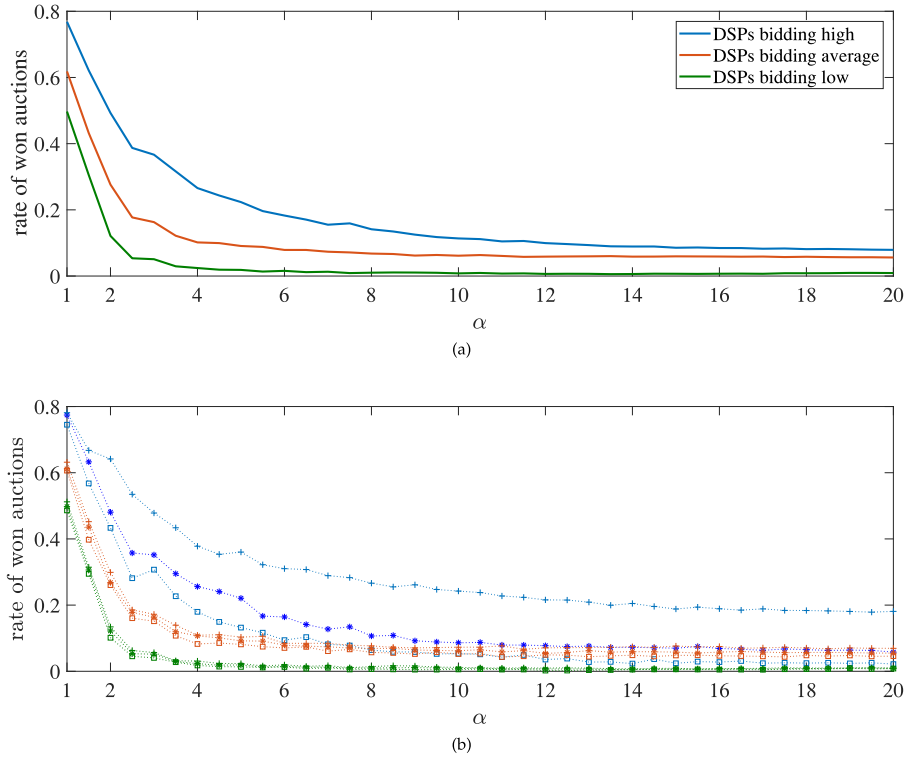

Fig. 16. Evolution of the rate of won auctions for different values of α (from 1 to 20 in steps of 0.5) and $\lambda = 0.05$. We consider 20 DSPs with different behaviors (bidding high, average and low). For each value of α, we repeat the experiment 20 times. The results are depicted averaged in (a) for each type of DSP. In (b), we illustrate results of percentiles 3 (using '+'), 50 (using '*') and 97 (using '□'). For these experiments, bid requests for the different types of DSPs are generated using a Gaussian distribution.

not only identification is feasible, but also other privacy attacks derived from the type of information released. For instance, variations in location data along with IP address changes could unveil user movement patterns. Also, information about sites visited may reveal the interests and behavior of users. Finally, these practices of ubiquitous tracking and aggregation of granular user information is largely concentrated in entities over which little control is enforced. Sadly, this concentration of the power of surveillance is not only affecting the privacy of many users but it is turning advertising entities into dangerous means of massive manipulation, as exemplified in Collins (2017).

This scenario, which is less and less encouraging for privacy, is made worse by the lack of transparency in the sharing of user information among the participating entities. Hence, users are unaware of the complex dynamics behind the advertising ecosystem and the particular privacy risks they are facing online. And so forth, partly motivated by some creepy perceptions regarding online behavioral advertising (Ur et al., 2012), people are increasingly using ad blockers. Whilst emerged to undermine abusive tracking from advertisers, ad blockers bear an interesting concept in giving users a more active role in the advertising ecosystem. This role might consist in providing users with more transparency and radical control over ads. However, very little can be really done if abusive behaviors that exacerbate privacy risks are ignored within the core of advertising platforms.

The privacy risk derived from user data sharing

One of the abusive behaviors that threat privacy in the online advertising ecosystem involves the malicious collection of bid requests by DSPs and related third parties who, violating the terms of service defined by ad exchanges, may be participating in auctions without any interest in winning. This is possible due to the oversending (broadcasting) of bid requests including personal data to DSPs, which is motivated by the need for an ad exchange to maximize profit.

Regulating bid request distribution as a mechanism to preserve privacy

Our contribution to address this issue consists in enabling some control over a crucial part of the advertising ecosystem: the bid request distribution model to DSPs and similar intermediaries. In this line, our experimental results show that reducing the number of DSPs recipients in online advertising through regulating such distribution may virtually cause no losses in revenue for the ad exchange. Particularly, the higher the value of the parameter r_i (product of the rate of won bids and average money spent) for the DSPs with better behavior, the less DSPs need to be contacted to reach the maximum revenue; thus, personal information would be shared among less third parties. Consequently, based on this control strategy, DSPs are encouraged to correct their behavior since, otherwise, their chances of participating and winning in ad auctions fall significantly. Despite the small losses in revenue, we state that win/win outcomes are reached for the interests of ad exchanges (revenue) and end users (privacy) since *actively regulating the behavior of third parties regarding user privacy could significantly discourage harmful attitudes of users towards online advertising (e.g., massively using ad blockers)*.

Unlike other approaches to preserve privacy in online advertising, which contemplate significant modifications in the ecosystem, a great added value is provided by ours since it entails a *minimum change in the bid request distribution strategy*, while leaving the main online advertising infrastructure untouched, albeit personal information is still ceded to third parties. This should be a great incentive for ad exchanges to adopt this kind of mechanisms in order to regulate the behavior of associated agencies and to take additional steps to protect the privacy of end users.

Interestingly, our approach could be extended to complement transparency and control enabled in the user side through an interface, e.g., the one offered by an ad blocker plugin such as Adblock Plus. First, an ad exchange implementing our bid request distribution model might provide users (through this user interface) with the value of α, i.e., the number of entities with which their personal data has been shared (transparency). Furthermore, as a privacy mechanism, the browser plugin could enable users to configure a maximum number of entities with which to share their data (informed decision). Accordingly, if the user data results to be shared with more entities in average, the plugin would block the corresponding ads (control).

In the same line of reducing the potential adversaries to protect privacy, an improvement of our approach could be *targeted auctioning*. This would consist in partitioning our optimization problem to be solved on a per-market basis, i.e., auctioning a user impression only among the DSPs subscribed to a given targeting market. Said otherwise, the specific targets of a DSP at a moment in time could serve as another reputation parameter when the ad exchange auctions a user impression.

Appealing to a change in the bid request distribution model, in the core of the advertising ecosystem, entails a big step towards enforcing privacy in this context; more if the impact of such strategy can be minimized. As depicted in the previous paragraph, although bringing some controlled loss in revenue, our proposal may suggest a paradigm shift with a multiplicative effect for the benefit of user privacy. Not only is activated a technology for ad exchanges to support privacy regulation, but part of the control can be given to users. And further, this approach could serve to alleviate the harmful tensions between advertising systems and users provoked by serious concerns regarding privacy.

Privacy protection with our system

The extent to which our mechanism could protect privacy may also be subject to discussion. Whereas the level of privacy provided by some mechanisms could be quantitatively measured under certain assumptions, whether the given protection is sufficient or not is pretty relative. This is because, in general, the level of privacy provided by any protection mechanism depends on the context, and in our case, it is defined by the requirements set in Section 4.1, the adversary model from Section 1.2, and the strategy proposed in Section 4.2. In this specific framework, our solution could provide great privacy by enabling an ad exchange to strongly support privacy without significantly affecting the revenue of the system. However, the ultimate level of privacy provided would depend on the particular strategy adopted by the ad exchange (either, e.g., aggressive, capping a lot the participation of DSPs; or moderate, not restricting it significantly).

Beyond this limited scenario, other players might still disclose user information, e.g., first parties (publishers), ISPs, data brokers, etc. However, the scope of action of ad exchanges is by far greater. Since DSPs may illegitimately benefit from such capabilities, extremely reducing the amount of DSPs participating in auctions, e.g., to a dozen, would improve privacy to a similar extent.

In any case, ours is a first approach to dealing with privacy issues in this particular context where user data may be inappropriately shared to dishonest DSPs. Interestingly enough, future work might concentrate on giving users further control capabilities focused on modulating the privacy parameter introduced in this work. Thus, if provided with some background information, users themselves would be able to choose the privacy level they feel comfortable with.

Incentives to adopt regulated data distribution

Unlike most proposals to protect privacy in the online advertising ecosystem, ours aims to encourage advertising players (mainly ad exchanges) to adopt it for the benefit of users and ad platforms. In this subsection, we describe the main incentives that these entities would have to implement the mechanism presented in this paper.

- Privacy regulations require more and more control from data controllers and processors over user data collection, use, and distribution (GDPR, 0000). Furthermore, the heavy fines for neglecting user privacy are pushing these entities to adopt protection mechanisms, especially when they manage user data at a massive scale (Solon, 2018). Since the contribution of our work aims at the more private distribution of user data at a relatively low cost, we think that ad exchanges would be strongly motivated to adopt it.
- Not only regulation is urging the advertising ecosystem to endorse privacy protection initiatives, the current ad blocker arms race (Iqbal et al., 2017) is empowering users to protect themselves through radical mechanisms that might be affecting the economic health of the advertising players. In such a conflictive environment, it seems reasonable for these actors, especially those in the core network, to start to give in to the users' legitimate expectations of privacy. Otherwise, the war would grow fiercer, seriously affecting not only the online advertising business model but that of the entire Internet.
- The implementation of the distribution mechanism we propose would not involve significant changes to the online advertising ecosystem. The ad exchange would only have to incorporate a module to distribute bid requests among the "best behaved" bidders for a given context (privacy parameter, targeting market, etc.). The rest of entities would remain unchanged.
- Even though our approach does not tamper with the current online advertising model, it might generate a regulation effect over DSPs. That is, in order to avoid penalization, DSPs would not adhere to dishonest practices when participating in auctions. Interestingly, this value-added service could further improve the system's revenues.
- As it has already happened with other privacy-enhancing initiatives (e.g., Facebook's ad preferences), ours opens the door to the implementation of further transparency and control mechanisms for users. Our mechanism would encourage those ad exchanges committed to respect user privacy to create interfaces for users to examine or even modulate the privacy parameter we are introducing in our work.
- Beyond its technical implications, ad exchanges would be highly encouraged to implement our proposal in order to compensate users for the opacity behind which the exploitation of their data has been hiding. The scandals surrounding the abuse of user information undermine daily the reputation of the advertising ecosystem and hence the trust of end users, the ultimate owners of data (the main input of the online advertising ecosystem). Upon realizing that specific controls are being implemented to protect their privacy, their concerns could be alleviated, since their main concern is not the sharing of information itself, but the inappropriate sharing of their information (Nissenbaum, 2009). Consequently, users themselves could even decide not to block the tracking of privacy-compliant ad exchanges, which is a further incentive for the latter to adopt our mechanism.

7. Conclusions

Undoubtedly, the main privacy concerns regarding online advertising come from the great capability of third parties to aggregate user data. Due to the inherent opacity of this ecosystem, the most known approaches to face such concerns build on radical ad blocking solutions. By entirely blocking ads and partly stopping the leakage of data from the user side, these radical approaches are threatening the current economic model of the Web. On the other hand, with the aim of balancing the trade-off between revenue and the number of invited DSPs (looking for more user privacy), we propose to modify part of the ad delivery model. Our technique arises as a strategy of bid request suppression where interactions carrying user data can be reduced, by design, to offer more privacy, while slightly affecting the revenue of the system. More specifically, we come up with a controlled distribution of bid requests among DSPs in order to reduce the amount of user data shared with said third parties. Nevertheless, our approach comes at the expense of revenue loss incurred by lowering the number of participants within ad auctions. Since this technique would be applied directly in the core of ad platforms, more overwhelming and less harmful results could be obtained.

Part of our contribution lies on an analysis of the privacy risks involved in the massive aggregation of data performed by some online advertising entities. In this line, we strive to characterize the personal information leaked in bidding interactions and some of the derived critical jeopardies. We concentrate on bid request messages that are used to invite DSPs to participate in ad auctions and that carry very granular information about the user online behavior. Thus, using a publicly available data set belonging to a famous Chinese DSP, we unveil the potential capability of advertising intermediaries to do massive surveillance even at a very low cost. Accordingly, we also highlight the power given to advertisers to microtarget users with a very fine precision.

Our main contribution is a mathematical approach to tackle the aforementioned problem of distributing bid requests to less DSPs, while minimally affecting the revenue of the system. We formulate and solve an optimization problem that seeks to maximize the revenue while bounding the participation of DSPs. Thus privacy is enforced through balancing this revenue-invitation rate trade-off.

As a result of our theoretical analysis, we present a close-form solution for the bid request distribution strategy and a revenue-invitation rate function characterizing the optimal trade-off curve. From this analysis, we find an interesting opportunity to cap the number of DSPs that receive bid requests while maintaining a reasonable revenue. From simulations performed over an auctioning scenario, we confirm that the revenue of the system indeed increases with the number of DSPs participating in each auction. However, we find that even when drastically reducing this number (thus, increasing privacy of users) an important portion of revenue may still be preserved. Also, it turns out useful to maximize revenue subject to a restriction that supports privacy when handing out bid requests, because it leads DSPs to behave better (e.g., increasing their rate of won bids), driven by a penalization on abusing the system (e.g., when bidding too low).

Certainly, much remains to be done with regard to privacy in the context of online advertising, especially while balancing said privacy protection efforts with the economic model that holds the free access to Internet today. For this, the ad delivery model itself must be rethinked because its components can implement privacy more effectively, in particular, concerning powerful privacy techniques such as data minimization and transparency for users.

Acknowledgments

This work was partially supported by the Spanish Ministry of Economy and Competitiveness (MINECO), Spain through the projects "INRISCO", ref. TEC2014-54335-C4-1-R, and "Sec-MCloud", ref. TIN2016-80250-R; as well as by the European Commission through the H2020 project "CLARUS". J. Parra-Arnau is the recipient of a Juan de la Cierva postdoctoral fellowship, FJCI-2014-19703, from the MINECO, Spain. The authors gratefully acknowledge the support provided by the Escuela Politécnica Nacional, Ecuador to José Estrada-Jiménez and Ana Rodríguez-Hoyos.

References

Achara, J.P., Parra-Arnau, J., Castelluccia, C., 2016. Mytrackingchoices: Pacifying the ad-block war by enforcing user privacy preferences, arXiv preprint arXiv:1604.04495.

Adblock Plus - surf the Web without annoying ads! Nov. 2015. accessed on 2015-11-15. [Online]. Available: https://adblockplus.org.

Backes, M., Kate, A., Maffei, M., Pecina, K., 2012. Obliviad: Provably secure and practical online behavioral advertising. In: 2012 IEEE Symposium on Security and Privacy. IEEE, pp. 257–271.

Bashir, M.A., Arshad, S., Kirda, E., Robertson, W., Wilson, C., 2018. How tracking companies circumvented ad blockers using websockets. In: Proceedings of the Internet Measurement Conference 2018. ACM, pp. 471–477.

Beales, H., 2010. The value of behavioral targeting, Netw. Advertising Initiative, Tech. Rep. Mar. accessed on 2016-01-15. [Online]. Available: http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf.

Brave, Brave software, Mar. 2016. accessed on 2016-05-30. [Online]. Available: https://www.brave.com.

Brookman, J., Rouge, P., Alva, A., Yeung, C., 2017. Cross-device tracking: Measurement and disclosures. Proc. Priv. Enhancing Technol. 2017 (2), 133–148.

Collins, B., 2017. Russia's facebook fake news could have reached 70 million americans, Aug. [Online]. Available: http://www.thedailybeast.com/russias-facebook-fake-news-could-have-reached-70-million-americans.

Easylist - overview, Mar. 2016. accessed on 2016-05-30. [Online]. Available: https://easylist.github.io.

Eckersley, P., 2010. How unique is your web browser? In: Privacy Enhancing Technologies, Vol. 6205. Springer, pp. 1–18.

Estrada-Jiménez, J., Parra-Arnau, J., Rodríguez-Hoyos, A., Forné, J., 2017. Online advertising: Analysis of privacy threats and protection approaches. Comput. Commun. 100, 32–51.

Evans, D.S., 2009. The online advertising industry: Economics, evolution, and privacy. J. Econ. Perspect. 23 (3), 37–60.

Faizullabhoy, I., Korolova, A., 2018. Facebook's advertising platform: New attack vectors and the need for interventions, arXiv preprint arXiv:1803.10099.

Fredrikson, M., Livshits, B., 2010. Repriv: Re-envisioning in-browser privacy. In: Proc. IEEE Symp. Security, Privacy (SP)(2011).

Gayomali, C., 2014. It would cost each user $232 a year for an ad-free internet, study finds, Aug. accessed on 2016-02-27. [Online]. Available: http://www.fastcompany.com/3034670/fast-feed/it-would-cost-each-user-232-a-year-for-an-ad-free-internet-study-finds.

GDPR, E., 0000. Home page of eu gdpr, línea]. Disponible en: https://www.eugdpr.org/. [Accedido: 19-feb-2018].

Ghosh, A., Mahdian, M., McAfee, R.P., Vassilvitskii, S., 2015. To match or not to match: Economics of cookie matching in online advertising. ACM Trans. Econ. Comput. 3 (2), 12.

Google, Ad exchange auction model, Oct. 2018. [Online]. Available: https://support.google.com/authorizedbuyers/answer/6077702?hl=en.

Google, Cookie matching, Mar. 2016. accessed on 2016-03-07. [Online]. Available: https://developers.google.com/ad-exchange/rtb/cookie-guide#examples.

Google, Google doubleclick ad exchange (adx) buyer program guidelines, Jun. 2017. [Online]. Available: https://www.google.com/doubleclick/adxbuyer/guidelines.html.

Google, Real-time bidding protocol, 2017. accessed on 2017-06-02. [Online]. Available: https://developers.google.com/ad-exchange/rtb/downloads/realtime-bidding-proto.txt.

Guha, S., Cheng, B., Francis, P., 2011. Privad: practical privacy in online advertising. In: USENIX Conference on Networked Systems Design and Implementation. pp. 169–182.

Helsloot, L.J., Tillem, G., Erkin, Z., 2017. Ahead: Privacy-preserving online behavioural advertising using homomorphic encryption. In: Information Forensics and Security (WIFS), 2017 IEEE Workshop on. IEEE, pp. 1–6.

Helsloot, L.J., Tillem, G., Erkin, Z., 2018. Badass: Preserving privacy in behavioural advertising with applied secret sharing. In: International Conference on Provable Security. Springer, pp. 397–405.

Hill, K., 2012. How target figured out a teen girl was pregnant before her father did, Feb. [Online]. Available: https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#764bfe476668.

Hoelzel, M., 2015. The programatic-advertising report: Mobile, video, and real-time bidding drive growth in programmatic, Apr. accessed on 2017-06-30. [Online]. Available: http://www.businessinsider.com/buyers-and-sellers-have-overwhelmingly-adopted-programmatic-with-mobile-leading-growth-2015-4.

IAB, 2015. Digital ad revenues surge 19 percent, climbing to $27.5 billion in first half of 2015, Oct. accessed on 2016-03-06. [Online]. Available: http://www.iab.com/news/digital-ad-revenues-surge-19-climbing-to-27-5-billion-in-first-half-of-2015-according-to-iab-internet-advertising-revenue-report/.

Iqbal, U., Shafiq, Z., Qian, Z., 2017. The ad wars: retrospective measurement and analysis of anti-adblock filter lists. In: Proceedings of the 2017 Internet Measurement Conference. ACM, pp. 171–183.

Kardan, A.A., Hooman, M., 2013. Targeted advertisement in social networks using recommender systems. In: e-Commerce in Developing Countries: With Focus on e-Security (ECDC), 2013 7th Intenational Conference on. IEEE, pp. 1–13.

Korolova, A., 2010. Privacy violations using microtargeted ads: A case study. In: Data Mining Workshops (ICDMW), 2010 IEEE International Conference on. IEEE, pp. 474–482.

Mathai, J., Ramasamy, G., Purusothaman, S., Ezra, K., 2015. Location based mobile advertising framework for commuters. In: Computing and Network Communications (CoCoNet), 2015 International Conference on. IEEE, pp. 928–935.

Mayer, J.R., Mitchell, J.C., 2012. Third-party web tracking: Policy and technology. In: 2012 IEEE Symposium on Security and Privacy. IEEE, pp. 413–427.

Mcdonald, A.M., Reeder, R.W., Kelley, P.G., Cranor, L.F., 2009. A comparative study of online privacy policies and formats. In: Proc. Int. Symp. Priv. Enhanc. Technol. (PETS). Springer-Verlag, Seattle, WA, pp. 37–55.

Mozilla, 0000. Subscribe2web, accessed on 2015-11-21. [Online]. Available: https://air.mozilla.org/subscribe2web/.

Narayanan, A., Shmatikov, V., 2008. Robust de-anonymization of large sparse datasets. In: Proc. IEEE Symp. Secur. Priv. (SP). IEEE Comput. Soc., pp. 111–125.

Nissenbaum, H., 2009. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press.

Olejnik, L., Castelluccia, C., 0000. To bid or not to bid? measuring the value of privacy in RTB, accessed on 2016-05-21. [Online]. Available: http://lukaszolejnik.com/rtb2.pdf.

Olejnik, L., Minh-Dung, T., Castelluccia, C., 2014. Selling off privacy at auction. In: Proc. IEEE Symp. Netw. Distrib. Syst. Secur. (SNDSS).

OpenX, 2010. Ad networks vs. ad exchanges: How they stack up, Jul. accessed on 2016-03-06. [Online]. Available: http://openx.com/blog/openx-releases-new-whitepaper-ad-networks-vs-ad-exchanges/.

Papadopoulos, P., Kourtellis, N., Markatos, E.P., 2018. Cookie synchronization: Everything you always wanted to know but were afraid to ask, arXiv preprint arXiv:1805.10505.

Parra-Arnau, J., 2017. Pay-per-tracking: A collaborative masking model for web browsing. Inform. Sci. 385, 96–124.

Parra-Arnau, J., Achara, J.P., Castelluccia, C., 2016. Myadchoices: Bringing transparency and control to online advertising, arXiv preprint arXiv:1602.02046.

del Prado Cortez, M.N., Frignal, J., 2014. Geo-location inference attacks: From modelling to privacy risk assessment (short paper). In: Dependable Computing Conference (EDCC), 2014 Tenth European. IEEE, pp. 222–225.

Privoxy.org, "Privoxy", Mar. 2016. accessed on 2016-03-17. [Online]. Available: http://www.privoxy.org.

Real-time bidding protocol - cookie matching, 0000. accessed on 2015-10-07. [Online]. Available: https://developers.google.com/ad-exchange/rtb/cookie-guide.

Real-time bidding protocol - processing the request, accessed on 2017-04-07. [Online]. Available: https://developers.google.com/ad-exchange/rtb/request-guide.

Rejón-Guardia, F., Martínez-López, F.J., 2014. Online advertising intrusiveness and consumers' avoidance behaviors. In: Handbook of Strategic e-Business Management. Springer, pp. 565–586.

Sánchez, D., Viejo, A., 2018. Privacy-preserving and advertising-friendly web surfing. Comput. Commun. 130, 113–123.

Shiller, B., Waldfogel, J., Ryan, J., 2017. Will ad blocking break the internet? National Bureau of Economic Research, Tech. Rep..

Smith, M., 2014a. Targeted: How Technology Is Revolutionizing Advertising and the Way Companies Reach Consumers, first ed. AMACOM, New York.

Smith, M., 2014b. Targeted: How Technology is Revolutionizing Advertising and the Way Companies Reach Consumers. AMACOM Div American Mgmt Assn.

Solon, O., 2018. Facebook faces $1.6bn fine and formal investigation over massive data breach, Oct. accessed on 2018-11-10. [Online]. Available: https://www.theguardian.com/technology/2018/oct/03/facebook-data-breach-latest-fine-investigation.

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F.N., Arvanitakis, G., Benevenuto, F., Gummadi, K.P., Loiseau, P., Mislove, A., 2018. Potential for discrimination in online targeted advertising. In: Conference on Fairness, Accountability and Transparency. pp. 5–19.

Sweeney, L., 2002. k-Anonymity: A model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 10 (5), 557–570.

Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H., Barocas, S., 2010. Adnostic: Privacy preserving targeted advertising. In: Proc. IEEE Symp. Netw. Distrib. Syst. Secur. (SNDSS). pp. 1–21.

Ur, B., Leon, P.G., Cranor, L.F., Shay, R., Wang, Y., 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In: Proceedings of the Eighth Symposium on Usable Privacy and Security. ACM, p. 4.

Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K.P., Loiseau, P., Goga, O., 2018. Privacy risks with facebook's pii-based targeting: Auditing a data broker's advertising interface. In: IEEE Symposium on Security and Privacy (SP). pp. 221–239.

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., Chen, Z., 2009. How much can behavioral targeting help online advertising? In: Proceedings of the 18th International Conference on World Wide Web. ACM, pp. 261–270.

Yuan, S., Abidin, A.Z., Sloan, M., Wang, J., 2012. Internet advertising: An interplay among advertisers, online publishers, ad exchanges and web users, arXiv:1206.1754, arXiv preprint.

Yuan, S., Wang, J., Zhao, X., 2013. Real-time bidding for online advertising: measurement and analysis. In: Proceedings of the Seventh International Workshop on Data Mining for Online Advertising. ACM, p. 3.

Zhang, W., Yuan, S., Wang, J., Shen, X., 2014. Real-time bidding benchmarking with ipinyou dataset, arXiv preprint arXiv:1407.7073.