Privacy and Security in Online Advertising

Bolun Liu

Introduction

When it comes to user and client concerns with privacy and security in online data collection, there's a lot to talk about. Hence, what the article aims to do here, is discuss some of the most critical privacy concerns and regulations proposals relevant to online advertising.

The current advertising model builds upon an intricate infrastructure composed of a variety of intermediary entities and technologies whose main aim is to deliver personalized ads. For this purpose, a wealth of user data is collected, aggregated, processed and traded behind the scenes at an unprecedented rate. Despite the enormous value of online advertising, however, the intrusiveness and ubiquity of these practices prompt serious privacy concerns. This article surveys the online advertising infrastructure and its supporting technologies and presents a brief overview of the underlying privacy risks and the solutions that may mitigate them. In particular, this article examines the main components of the advertising infrastructure in terms of tracking capabilities, data collection, and privacy risk, and overviews the tracking and data-sharing technologies employed by these components. Then, the article covers a comprehensive survey of the most relevant proposals and Approaches to prevent privacy risk and data abuse, classifies and compares them on the basis of their privacy guarantees and impact on the Web.

Online Advertising Ecosystem

Selecting and directing information are crucial in every aspect of our modern lives, including areas as diverse as health, leisure and research. In the past, these processes were largely manual, but due to the exponential improvements in computation and sophistication of software, they are becoming increasingly automated. The industry of online advertising, lavishly illustrated by Google DoubleClick and real-time bidding (RTB), is an example of the ever-growing automation of these processes, and another crucial aspect of our society — to a large extent, the success of most competitive economic activities is dependent on advertising, particularly on the ability to effectively select and direct information to the right potential customers. Undoubtedly, the advent of the Internet and the Web has created a myriad of new opportunities for advertisers to target billions of people almost effortlessly. However, online advertising is not only ubiquitous. In the early days of the Web, ads were served directly by the publisher as known as the page's owner following a one-size-fits-all approach. But due to the ease with which Web users can be tracked across their page visits, online advertising has also become increasingly personalized. An

example of the sophistication of ad personalization is RTB, which enables advertisers to direct ads to the right user and at the right time, by competing in real-time auctions for the impression of their ads.

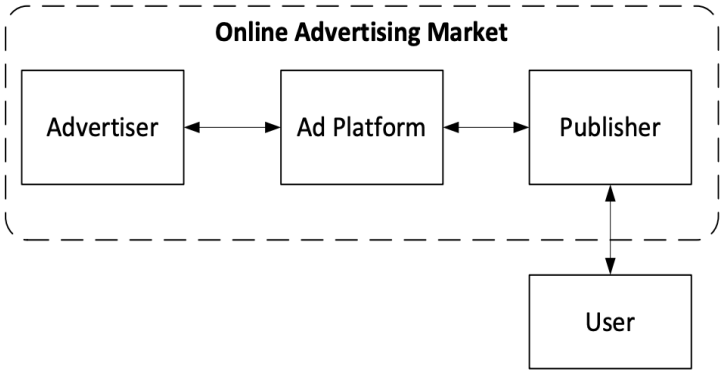

Figure 1. Main components of the online advertising ecosystem

Evidently, personalized advertising is the most effective, and hence the most profitable form of advertising. Those ads relying on a user's browsing interests ensure conversion rates that double those of untargeted ads [1]. On the other hand, from the publishers' perspective, online advertising is the pillar that sustains the Internet's "free" content and services. Nevertheless, advertisers and publishers are not the only entities taking part in this business. In fact, there exists an entire infrastructure at the service of both of them, supported by companies like Google, Facebook, and Twitter. Enabled by these and hundreds of other ad companies, targeting mechanisms take charge of selecting and directing ads to billions of users every day, depending on a number of factors such as the page they are visiting; their browsing history; their IP address, or parts of it; their operating system; the plug-ins installed and other information related to their Web browser [2]; and obviously the objectives and budgets of all advertisers for displaying their ads. User information is therefore an asset fundamental to the efficient and effective delivery of advertising, which is not only handed over to the highest bidder but to many other third parties that are involved in the ad-delivery process.

Security and Privacy Concerns

Unfortunately, evident security risks exist for users when personal, sensitive data about their habits are traded in the name of personalized advertising by an infrastructure that operates in the shadows with virtually no oversight [3]. These security risks can be explained in terms of privacy hazards, social sorting, discrimination, malware distribution, fraud, and others [4][5]. Regarding privacy, serious concerns have been raised by the intrusiveness of practices and the increasing invasiveness of digital advertising. According to recent surveys, two out of three Internet users are worried about the fact that their online behavior is scrutinized without their knowledge and consent. Numerous studies in this same line reflect the growing level of ubiquity and abuse of

advertising, which is perceived by users as a significant degradation of their browsing experience [6][7]. In an attempt to mitigate these privacy and security risks, several approaches have been proposed by a heterogeneous group of actors. 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. On the other hand, commercial solutions have primarily focused on blocking tracking mechanisms at the cost of seriously damaging the Internet business model.

The pervasive dissemination of online advertising on the Internet and the prevailing need for ad platforms and other intermediary entities to collect a wealth of data about Web users prompt serious concerns regarding user privacy [8][9]. In fact, much of the concern regarding privacy and thus regarding privacy threats in online advertising are derived from the risks of misuse of this huge amount of user data, which is held by advertising platforms. Said misuse of user information might include common privacy issues such as data leakage, unauthorized collection of data, and sharing with a third party. Interestingly, as surveyed, the structure of ad platforms and the abilities of their players reflect behaviors strictly coincidental with such privacy issues.

The privacy risks posed by the tracking and profiling practices of the online advertising industry have motivated a variety of privacy-protecting approaches from academia. These research initiatives mostly rely on mechanisms that may support or complement the current economic model of the Web, while others suggest moderate blocking of third-party tracking to protect user privacy. Other plug-and-play proposals are also available to users and are supported commercially. In essence, such approaches provide users with transparency and control functionalities over their browsing data, yet putting at risk the Web economic model, currently built on the revenues of online advertising, through radical blocking mechanisms.

Risks within Online Advertising Interactions

The growing access of people to information and communication technologies is contributing to reaching the so-called “big data era”, where the pervasiveness of data is a major input for increasingly personalized and automated online services. Among such services, online advertising aims at selecting and directing ads to the right potential customers a.k.a personalization at the right time, built on multiple parameters, while users browse the Web [10][11]. 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 are generating greater revenues than those of untargeted ads [12]. Furthermore, as previously described, websites (publishers) 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. 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: real-time bidding, which enables advertisers to compete in real-time auctions to show their ads [13]. 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, RTB's 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 [14], it is quite transparent for 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.

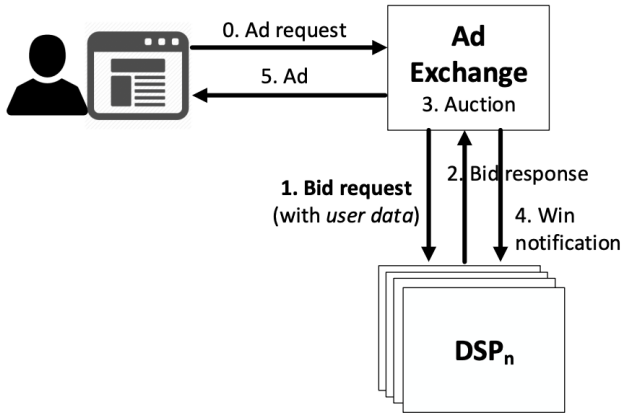

Figure 2. Interactions among an ad exchange and associated DSPs.

Yet, despite its proven usefulness, the practices inherent to online advertising and RTB may pose serious privacy risks for users [15]. 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 [16] that privacy protection is discouraged. Additionally, ad distribution mechanisms based on auctioning user impressions might lead to characterizing users as more valuable economically than others, depending on their profiles [17]; such a differentiation may entail social sorting or discrimination [18], 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.

Table 1. Information, items carried in bid requests, here matched to the potential privacy risks derived from their open distribution and aggregation.

	Identification	Learning of moving patterns	Cross device tracking	Micro targeting	Habits tracking	Outlier detection
user ID	x					
IP address		x				
user location		x	x		x	
device fingerprint	x		x		x	
web browser fingerprint	x		x		x	
time stamp		x				
user languages	x			x		
user labels				x		x
URL					x	
content labels				x		
minimum bid price			x			x

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 such as 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.

In practice, upon the reception of bid requests, 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. 3, where researchers 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.

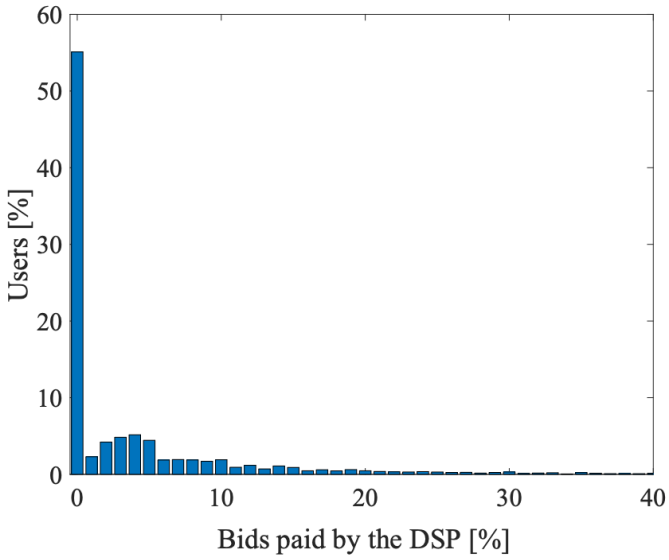

Figure 3. Percentage of users whose information has been paid by the iPinYou DSP

Strategies and Approach to Address Privacy Issue

In an attempt to prevent abuse, ad exchanges clearly prohibit DSPs to use the information in bid requests when a corresponding auction has not been won [19]. 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 on the demand side like trading desks.

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 [20]. Although the distribution of personal data among a group of DSPs cannot be entirely stopped without changing the current advertising business model, it is reported that the potential abuse of these agencies can be tackled by 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 such as never winning auctions will be banned from participating in future auctions, which may entail correcting such harmful behaviors. 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 [21].

Other proposals have addressed 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 [11], it seems very hard to provide more privacy in the online advertising ecosystem without somehow modifying the ad delivery model.

Conclusion

Online advertising is based on tracking technologies to “follow” and monitor users wherever they browse on the Web. Over time, during such tracking, a lot of metadata is collected about a user, which can be employed to build detailed profiles. Moreover, due to the pervasiveness of online advertising, billions of users get involved in this process. Such range and the powerful personalization technologies considered have made online advertising a millionaire business whose revenues are said to be financing the current free Internet access model. Since this successful business revolves around the massive exploitation of user data, there are multiple privacy concerns. This is compounded by the complexity and opacity of the internal structure of the online advertising ecosystem, which further complicates the study of privacy risks, and limits the implementation of protection approaches. As a consequence, privacy protection initiatives neglect important parameters, such as the sustainability of the entire system, making them impractical in the long term. To address this issue, academic proposals for privacy enhancement in online advertising have been brought out, but there is still a long way to go to make a trade-off between commercial profits and user privacy. After all, adjustments to this sophisticated commercial ecosystem involve redistribution, protection, or restriction of multiple interests.

Reference

[1] H. Beales, “The value of behavioral targeting,” Netw. Advertising Initiative, Tech. Rep., Mar.2010, accessed on 2016-01-15.

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

[3] A. M. McDonald and L. F. Cranor, “Americans’ attitudes about internet behavioral advertising practices,” in Proceedings of the 9th annual ACM workshop on Privacy in the electronic society. ACM, 2010, pp. 63–72.

[4] D. Lyon, Ed., Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination. Routledge, Dec. 2002.

[5] U. Senate, “Online advertising and hidden hazards to consumer security,” Tech. Rep., 2014.

[6] “The cost of ad blocking,” PageFair, Res. Rep., Aug. 2015

[7] “The state of online advertising,” Adobe, Tech. Rep., 2012, accessed on 2015- 09-11.

[8] A. Goldfarb and C. E. Tucker, “Privacy regulation and online advertising,” Management Science, vol. 57, no. 1, pp. 57–71, 2011.

[9] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang, “Smart, useful, scary, creepy: perceptions of online behavioral advertising,” in proceedings of the eighth symposium on usable privacy and security. ACM, 2012, p. 4

[10] M. Smith, Targeted: How Technology Is Revolutionizing Advertising and the Way Companies Reach Consumers, 1st ed. New York: AMACOM, Nov. 2014.

[11] “Real-time bidding protocol - cookie matching,” accessed on 2015- 10-07.

[12] H. Beales, “The value of behavioral targeting,” Netw. Advertising Initiative, Tech. Rep., Mar. 2010, accessed on 2016-01-15.

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

[14] M. Hoelzel, “The programatic-advertising report: Mobile, video, and real-time bidding drive growth in programmatic,” Apr. 2015, accessed on 2017-06-30.

[15] J. Estrada-Jim´enez, J. Parra-Arnau, A. Rodr´ıguez-Hoyos, and J. Forn´e, “Online advertising: Analysis of privacy threats and protection approaches,” Computer Communications, vol. 100, pp. 32–51, 2017.

[16] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in Proc. IEEE Symp. Secur., Priv. (SP). IEEE Comput. Soc., May 2008, pp. 111–125.

[17] L. Olejnik, T. Minh-Dung, and C. Castelluccia, “Selling off privacy at auction,” in Proc. Symp. Netw. Distrib. Syst. Secur. (SNDSS), Feb. 2014.

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

[19] Google, “Google doubleclick ad exchange (adx) buyer program guidelines,” Jun. 2017.

[20] M. Hoelzel, “The programatic-advertising report: Mobile, video, and real-time bidding drive growth in programmatic,” Apr. 2015, accessed on 2017-06-30.

[21] H. Nissenbaum, Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press, 2009.