Privacy concerns and protection measures in online behavioural advertising

Leon J. Helsloot Gamze Tillem Zekeriya Erkin
Delft University of Technology
Department of Intelligent Systems, Cyber Security Group
Delft, The Netherlands
l.j.helsloot@student.tudelft.nl, {g.tillem,z.erkin}@tudelft.nl

Abstract

Online Behavioural Advertising (OBA), the practice of showing advertisements based on a person's web browsing behaviour, has become a vital component of the ad-supported web. The tracking of users' browsing behaviour that is needed for OBA, however, raises privacy concerns. We give an overview of the OBA landscape, and describe which user information is collected, which techniques are used to perform the collection, and how user information is shared between companies. Moreover, we discuss the privacy concerns that are raised by current OBA practices. After identifying privacy concerns, we describe a range of existing techniques to protect user privacy in online advertising. These techniques are compared based on their feasibility in the current advertising ecosystem, including the potential utility they offer advertising companies and how well they can be integrated with current trends in online behavioural advertising. Finally, we identify open problems in the protection of user privacy in online advertising.

1 Introduction

Online advertising is a multi-billion dollar industry, forming one of the driving economic factors behind free web services [5]. With a worldwide spend of $178 billion in 2016 [8], advertising has become a pervasive phenomenon on the internet as we know it today, allowing publishers to offer content to users free of charge. In the past few years, however, an increasing number of people have chosen to prevent advertisements from being shown on web pages they visit. An estimated 615 million devices use ad blocking tools, which amounts to 11% of the global internet population, and the number is expected to grow further [14].

The consequence of the increased use of ad blockers is that publishers suffer from a significant decrease in revenue from the advertising space they offer. The worldwide loss of ad revenue due to ad blocking was estimated to be $41.4 billion in 2016, which is 23% of the total ad spend [15]. In an effort to address the economic impact of ad blocking, some publishers now request users of ad blockers to allow ads on their websites, pay for content that was previously free, or deny access to users of ad blockers altogether [2]. Such practices have sparked an arms race of technologies to block advertisements, and to circumvent blockage. These developments are a threat to the business models of many websites with freely available content. For a sustainable ad-supported internet economy, it is necessary to address the objections users have to advertisements.

In a recent survey among users of a popular ad blocker, 32% of respondents indicated that privacy concerns were among the reasons for using an ad blocker [2]. Similarly, in a 2011 survey about privacy and advertising among United States residents, 94% of respondents considered online privacy an important issue, and 70% of respondents believed that online advertising networks and online advertisers should be responsible for safeguarding an individual's online privacy [20]. The practice of selecting advertisements based on users' browsing behaviour, and the data collection required for such

targeting, is one of the factors leading to privacy concerns [21]. Although advertisements tailored to users' interests are recognized as being useful for both users and publishers, users are reluctant to accept behaviourally targeted advertisements due to a mistrust of advertising companies. Moreover, users experience a lack of control over what information is collected about them [21].

In this paper, we give an overview of the OBA landscape, identify privacy concerns raised by the practice of behavioural advertising, and describe existing tools to protect user privacy in online advertising. Moreover, we identify open problems and future research directions related to privacy protection in OBA. This paper is organized as follows. In Section 2, we give an overview of the current trends in online advertising. Section 3 describes the privacy issues raised by advertising practices, after which we outline various protection measures in Section 4. Finally, we identify open problems related to privacy in online advertising in Section 5

2 Online advertising landscape

The online advertising landscape has become a complex landscape of different entities and mechanisms. In this section, we describe the Real-Time Bidding (RTB) mechanism, the practice of Online Behavioural Advertising (OBA), and techniques used to collect user information by tracking web browsing behaviour.

2.1 Real-time bidding

The current trend in online advertising is the RTB mechanism of buying and selling advertisements [22]. RTB facilitates a real-time auction of advertising space, which allows buyers to decide how much to bid per impression. This gives advertisers fine-grained control over how they spend their advertising budget, as well as on which pages and to whom their advertisements are shown. These auctions take place at an ad exchange, which operates much like a stock exchange for ad impressions. The real-time nature of such auctions, in which bids must be placed in a fraction of a second, requires the whole auction process to be carried out programmatically.

To manage the added complexity of RTB, other types of platforms emerged along with ad exchanges. Such platforms include Demand-Side Platforms (DSPs), which bid on individual impressions from multiple inventories for advertisers, and Supply-Side Platforms (SSPs), which support publishers in optimizing advertising yield. Moreover, the opportunity to target ads at individual users has led to the emergence of Data Management Platforms (DMPs), which provide user data to DSPs and SSPs.

The interplay between the various parties in the online advertising ecosystem is illustrated in Fig. 1. Prior to ad impressions, advertisers set up their campaign, including targeting criteria and campaign characteristics, with the DSP. When a single ad impression is triggered by a user visiting a web page that includes an ad slot, the user's web browser is instructed to contact the SSP to request an advertisement. The SSP forwards the ad request to the ad exchange, along with user data collected by the SSP. The ad exchange then sends a bid request to all connected bidders (DSPs), each of which may contact DMPs to retrieve additional information about the user. Each DSP decides how much to bid on behalf of its advertisers and responds to the ad exchange with the bid value and the associated advertisement. After collecting responses from bidders, the ad exchange chooses the winning bid and sends a win notice back to the winner. The ad exchange then sends the advertisement of the winning bidder to the SSP, which forwards it to the user. Finally, the advertisement is displayed in the user's browser. If the ad is clicked, the user's browser contacts the advertiser, and the advertiser records the ad click.

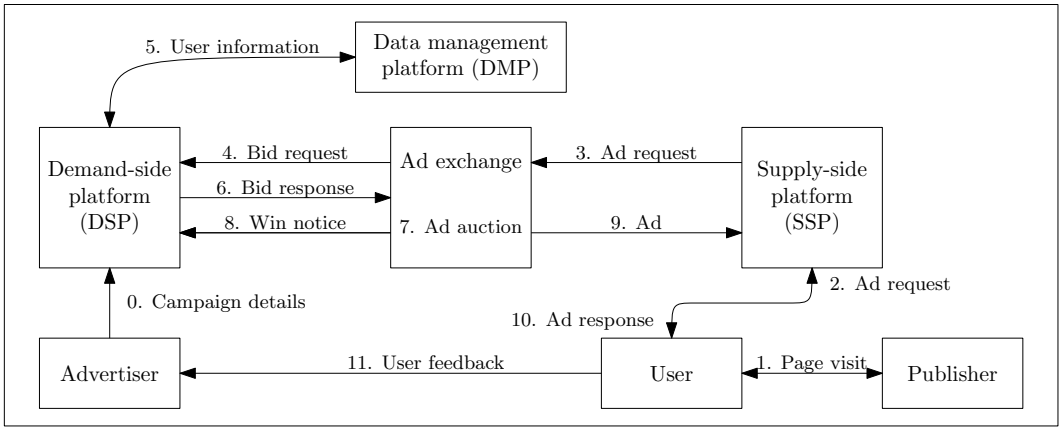

Figure 1: Overview of the online advertising ecosystem.

2.2 Online behavioural advertising

The RTB ecosystem allows advertisers to target advertisements at individual users based on previously exhibited behaviour, such as visited webpages and clicked advertisements. This practice of showing the most relevant advertisement for a viewer based on their browsing behaviour is known as Online Behavioural Advertising (OBA). OBA has been shown to greatly improve click-through rates of advertisements, allowing for highly effective targeting of advertisements at specific consumer groups or even individual consumers [23]. Users see advertisements that are more relevant to their interests, and may help them discover a product or service they need. Advertisers, on the other hand, benefit from accurately targeted advertisements because they are more likely to reach the desired audience, and thus increase the expected return on investment of the advertising budget. Finally, publishers benefit from an increased value of the advertising space they offer, allowing for the continuation of the 'free content' business models that internet users have become accustomed to.

OBA requires information about both the user and the visited webpage to be passed along a number of parties. Typically, a webpage includes a piece of code that causes the user's web browser to contact an SSP. The SSP then identifies the user, and sends information about the user to the ad exchange. The ad exchange, in turn, sends this user information, along with information about the webpage and publisher, to all DSPs participating in the auction. Each of these DSPs may have its own information about the user, and may contact one or more DMPs to obtain additional user information. The gathered information is then used to predict the response of the user to specific advertisements, which in turn is used to place a bid on the ad impression.

2.3 User tracking

To track the web browsing behaviour of users across the web, companies within the advertising ecosystem use a variety of techniques. Most of these techniques require collaboration between the tracker and the publisher, typically in the form of the publisher including some content to be loaded from the tracker's servers on their web pages. By collaborating with many publishers, trackers can gain an increasingly complete view of web page visits across the web.

The most well-known tracking method uses HTTP cookies that store unique user identifiers associated with the domain of a tracker. Typically, trackers instruct publishers to include on their pages a small, 1×1 pixel transparent image that is loaded from the tracker's domain. When such a pixel is loaded, the third-party cookie associated with the tracker's domain is sent with the request, along with an HTTP referrer header

containing the URL of the web page on which the pixel was included. This mechanism allows the tracker to log visits to all pages on which its pixels are included.

Browser cookies can be easily removed by users, however. Therefore, some trackers have adopted resilient tracking techniques to circumvent users' tracking preferences, such as *evercookies* or fingerprinting. Evercookies refer to combinations of different browser storage vectors such as Flash cookies, browser cache, and HTML local browser storage, which are used to recreate deleted HTTP cookies. Browser fingerprinting refers to tracking techniques which do not rely on local browser storage to persist an identifier, but identify users based on unique combinations of browser and device characteristics. These techniques typically rely on inclusion of JavaScript from the tracker, which reports back to the tracker's servers with a user identifier and the visited web page. Both evercookies and fingerprinting are used in practice, and combinations of tracking techniques are used to increase the resilience against countermeasures [1].

3 Privacy violations in behavioural advertising

The nature of the data that is used for OBA, as well as the exchange of such data between a large number of parties, raises privacy concerns. Dozens of third party advertising companies track users across the web to keep track of the web pages visited by users, which products they purchase, users' location, which devices and software they use, and more [6]. Such information about user activity is widely shared between companies, often without users' knowledge or consent [13].

Behavioural targeting primarily relies on a user's web browsing history to estimate the probability of a user clicking on an advertisement. Such browsing histories include visited web pages, but also search queries and previously clicked advertisements. The browsing history is rarely the only input for the estimation, however. Apart from characteristics of the current page, additional information about the user is gathered to improve the accuracy of the estimation. Such information commonly includes the user's location, (partial) IP address, and user agent, which describes the specific browser and operating system used. When possible, this may be further augmented with demographics, such as age and gender, and a history of online purchases.

There are a number of issues with widespread collection and sharing of user data. Firstly, the nature of the data that is collected is highly sensitive. Pages visited by a user can indicate location, interests, purchases, sexual orientation, financial challenges, medical conditions, and more.Moreover, web browsing histories are analysed at large scale to find patterns of activity, which allows even more inferences about users' interests. While these inferences are made to serve advertisements tailored to users' interests, knowledge of people's behaviour and interests, possibly including political, financial, or medical information, leaves opportunity for abuse.

The possibility of abuse is further increased by the shape of the advertising landscape. User information is not securely stored at a single party, but is widely shared amongst parties. Advertising companies normally have access only to partial web browsing histories, as each company relies on the cooperation of publishers to allow tracking on their web pages. Using a technique called *cookie matching*, however, parts of these histories are leaked to other companies. Cookie matching is a widespread practice that allows ad exchanges, DSPs and other parties to synchronize unique user identifiers and share data [13]. Users are given little insight and control over which data is collected about them and which parties have access to this data. Due to information leakage via cookie matching, and the exchange of personal data via DMPs, user data is spread to a large number of third parties.

Browsing histories not only contain sensitive personal information, they are also tightly linked to users' identities. Olejnik et al. [12] found that a vast majority of internet users had unique browsing histories, and that these unique histories could be used as fingerprints of users. With a small number of observed web page visits,

users could be identified based on their browsing fingerprint. Moreover, even indirect observations about user history, such as interest categories inferred by some advertisers, were sufficient to uniquely identify a large number of users [12]. It is therefore possible to link distinct browsing sessions to the same user, implying that some countermeasures against tracking, such as private browsing modes offered by most web browsers, may not be effective at protecting users.

Possible attackers are not limited to companies within the advertising ecosystem; adversaries acting as advertisers or monitoring users can also abuse behavioural advertising to learn sensitive information about users. Korolova [7] shows that precise targeting mechanisms and reporting options offered by Facebook can be abused by an advertiser to infer private attributes of user profiles. Their attack shows that information that should not be visible to advertisers can be leaked using precise targeting, undermining the trust people may have in Facebook to hide private information from third parties.

4 Protection measures

The privacy risks posed by tracking and profiling practices in online advertising have led to the development of a variety of tools to protect the privacy of users [6]. Some tools aim to be compatible with the current economic model of ad-supported websites, whereas others offer methods to block online advertisements altogether. This section gives an overview of existing tools aimed at privacy protection in online advertising, describes the underlying techniques, and evaluates the potential utility the tools offer advertising companies in the current OBA landscape.

Tracker blocking A widely used approach to protect privacy in online advertising is to use client-side tools that block requests to tracking parties [10]. Such blocking tools operate on the user's machine, without requiring any cooperation from tracking parties or website operators. Some blocking tools aim to prevent advertisements from being shown altogether, whereas others focus primarily on blocking trackers.

While tracker blocking gives users some control over which trackers are contacted by their browser and which advertisements they see, tracker blockers generally fail to completely block tracking. Moreover, blocking network requests breaks some websites, particularly when using heuristic detection techniques [10]. Finally, ad blockers and tracker blockers severely limit the advertising revenue for publishers, thus endangering the business model of many websites [6].

Obfuscation and anonymization An approach to hide user interests from trackers without blocking content is to add noise to the collected data, a technique known as obfuscation [4]. Obfuscation has been used to make tracking a user more difficult by randomizing browser attributes, which protects against browser fingerprinting by making sessions unlinkable [11]. Moreover, an attempt has been made to obfuscate user interest profiles generated by Google by inserting dummy traffic. However, the effect of obfuscation was overshadowed by a high variance of user profiles over time [4].

While randomization of web browser characteristics or user interests is likely to have a somewhat higher utility to publishers than completely blocking advertisements, such obfuscation techniques do not allow any behavioural targeting. Papaodyssefs et al. [16] therefore use k-anonymity to ensure that browsing histories are not uniquely identifiable, yet contain useful behavioural information. The anonymization is performed by only passing browsing events to the tracker if they do not make the browsing signature of the user distinguishable from at least $k-1$ other users. If a browsing event threatens to make a user's browsing history too dissimilar from other browsing histories, one of two intervention policies is applied: either the tracking cookie is dropped, or the user's

browsing history is obfuscated with web pages that increase the similarity with other browsing histories. The proposed architecture, however, relies on a proxy that has access to the browsing histories of all users. The privacy problem is thus shifted from one tracking party to another.

These obfuscation-based approaches have been shown to have some effect on user profiles in practice, but neither the effect on user privacy nor the impact on advertising utility is thoroughly analysed. Obfuscation attempts in similar domains have been shown to be unsuccessful at protecting user privacy, e. g. in web search [17]. Furthermore, feedback from users with obfuscated profiles, in the form of e. g. clicks on advertisers, may lead to a decrease in targeting performance for all other users if user profiles contain false interests.

Local profiling Many of the proposed solutions to enhance the privacy of OBA keep user profiles on the user's own machine. Local profiling techniques construct a user profile, typically consisting of interest keywords, from a user's web browsing behaviour. These profiles are then made available to advertising companies in a privacy-preserving way. Advertising companies thus no longer have access to detailed web browsing histories, which reduces the amount of information leaking to third parties.

In Adnostic [19], both profiling and targeting are performed within the user's web browser. When a user visits a web page, an ad network sends back a list of multiple ads based on the page content. Adnostic then selects the most relevant advertisement from the downloaded list by comparing topic tags of the advertisement with the local user profile. A major drawback of Adnostic is that it does not hide the user's web browsing behaviour from the advertising network. Furthermore, only a small set of advertisements is downloaded, and this set is selected without any behavioural targeting.

While local user profiling can be used to alleviate many of the privacy concerns in OBA, its adoption in existing solutions is subject to some pitfalls. All solutions described here assume an online advertising model where a central party is responsible for selecting which advertisement to show to a user. This selection is based on a combination of interest keywords and targeting keywords submitted by advertisers. In the current RTB model, however, advertisement selection involves many different bidders and ad selection models trained on massive amounts of user data. Generalization through local user profiling is thus expected to result in worse personalization performance than current advertisement selection algorithms [6]. Moreover, as described in Section 3, even generic interest keywords are sensitive information and can be used to uniquely identify web users.

Cryptographic approaches Cryptographic constructions have been used to address two privacy problems in online advertising: retrieving ads from a server without disclosing which ad was viewed, and privately reporting aggregate ad click statistics. ObliviAd [3] uses secure hardware to protect user privacy. To request an ad, a user sends a locally generated profile containing interest keywords to an ad broker, who is responsible for selecting and serving ads. The broker uses a Secure Coprocessor (SC) to match profile keywords to ad keywords, which ensures that no profile information can be learned by the broker and additionally allows remote verification of the program running on the SC. The SC retrieves ads from an encrypted storage using an Oblivious RAM (ORAM) scheme proposed by Shi et al. [18], which protects the access pattern of the SC on the ad storage. To privately report ad clicks, ObliviAd uses electronic tokens which are encrypted and signed by the SC. Tokens of different users are mixed on the SC before being published, which hides the link between users and clicks from the broker. However, hardware-based solutions seem infeasible without increasing the cost beyond the point of profitability. Moreover, the current RTB model involves many parties in advertisement selection, and uses models for ad selection that are more complex than keyword matching. Finally, companies may not be willing to disclose which algorithms they use for ad selection due to competition among advertising platforms.

Adnostic [19] uses a voting system based on additively homomorphic encryption to privately report ad clicks. Additively homomorphic encryption schemes allow anyone with access to a public key pk and ciphertexts $\mathcal{E}_{pk}(m_1)$ and $\mathcal{E}_{pk}(m_2)$ to create the ciphertext $\mathcal{E}_{pk}(m_1+m_2)$. Such schemes allow individual encrypted 'votes' or ad clicks to be aggregated by a trusted party before being decrypted to reveal aggregate statistics. Voting schemes as suggested in Adnostic, however, are based on downloading a set of advertisements in advance, making the schemes unsuitable for real-time targeting.

Privacy of selected advertisements While much of the research on privacy in online advertising focuses on sensitive information being exposed to advertising companies, behavioural targeting can reveal interests to other parties as well. To protect against microtargeting attacks as described by Korolova [7] (see Section 3 for a description of the attack), Lindell and Omri [9] propose a differentially private mechanism for releasing campaign statistics. Differential privacy requires that a change of a single database entry changes the distribution of responses given by the database only slightly, which is achieved by adding randomized noise to the output of queries for campaign statistics such as impressions. The added noise ensures that an attacker cannot determine if a specific impression occurred. However, advertisers might be hesitant to accept being charged based on noisy campaign statistics, and many ad selection systems use click feedback to optimize campaigns. Moreover, an attacker can make accurate inferences by averaging a sufficiently large number of campaign statistics [7].

5 Discussion and future work

While several approaches to enhance the privacy of OBA have been suggested, none of the existing solutions can achieve user privacy in the current RTB model without harming the economic benefits of OBA. Previous work limits the available targeting information, underestimates the complexity of the advertising landscape and ad selection models, or precludes learning preference patterns from individual responses. Future research directions are to explore potential privacy violations in online advertising mechanisms, and to create protection tools that harmonize the economic significance of OBA with users' expectation of privacy in the complex advertising landscape.

To alleviate privacy concerns within the current RTB model, numerous challenges need to be overcome. Where most previous work assumes a single entity to perform ad selection, the RTB model requires many different entities to perform part of the ad selection by placing bids. This distributed ad selection requires the dissemination of user information to a multitude of parties, which becomes even more challenging to do privately if parties wish to exchange data. Moreover, bidders with a small set of similar campaigns may expose users to homogeneity attacks. If a bidder with a portfolio consisting entirely of campaigns for loans wins an auction, they can infer that the user may have an interest in loans. Finally, bidders typically require user feedback, such as clicks on ads, to improve their targeting models. In previous work, such feedback is aggregated or mixed to obtain global click-through rates, but modern ad selection models use individual responses as training data. It is thus necessary to privately report such individual responses.

In this paper, we identified privacy concerns raised by the practice of OBA and discussed existing tools to improve user privacy. The ad blocking arms race makes it evident that, in order to maintain the ad-supported web as we know it today, the privacy concerns of users must be addressed while maintaining the profitability of online advertising. While each of the discussed approaches addresses a relevant problem, the challenges associated with the current RTB model cannot be solved with existing tools. In future work, the identified challenges must thus be overcome, such that the vast amount of information available on the internet can remain freely accessible.

References

[1] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild". In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. Nov. 2014, pp. 674–689.

[2] M. An. *Why People Block Ads (And What It Means for Marketers and Advertisers)*. HubSpot Research, July 2016.

[3] M. Backes, A. Kate, M. Maffei, and K. Pecina. "ObliviAd: Provably Secure and Practical Online Behavioral Advertising". In: *2012 IEEE Symposium on Security and Privacy*. May 2012, pp. 257–271.

[4] M. Degeling and T. Herrmann. "Your Interests According to Google - A Profile-Centered Analysis for Obfuscation of Online Tracking Profiles". In: *ArXiv e-prints* (Jan. 2016). arXiv: `1601.06371 [cs.CY]`.

[5] J. Deighton and H. M. Brierley. *Economic Value of the Advertising-Supported Internet Ecosystem*. Interactive Advertising Bureau, Sept. 2012.

[6] J. Estrada-Jiménez, J. Parra-Arnau, A. Rodríguez-Hoyos, and J. Forné. "Online Advertising: Analysis of Privacy Threats and Protection Approaches". In: *Computer Communications* 100 (Mar. 2017), pp. 32–51.

[7] A. Korolova. "Privacy Violations Using Microtargeted Ads: A Case Study". In: *2010 IEEE International Conference on Data Mining Workshops*. Dec. 2010, pp. 474–482.

[8] V. Letang and L. Stillman. *Global Advertising Forecast*. MAGNA, Dec. 2016.

[9] Y. Lindell and E. Omri. "A Practical Application of Differential Privacy to Personalized Online Advertising." In: *IACR Cryptol. EPrint Arch.* (Mar. 2011). Cryptology ePrint: `2011/152`.

[10] G. Merzdovnik, M. Huber, D. Buhov, N. Nikiforakis, S. Neuner, M. Schmiedecker, and E. Weippl. "Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools". In: *2nd IEEE European Symposium on Security and Privacy*. Apr. 2017.

[11] N. Nikiforakis, W. Joosen, and B. Livshits. "PriVaricator: Deceiving Fingerprinters with Little White Lies". In: *Proceedings of the 24th International Conference on World Wide Web*. 2015, pp. 820–830.

[12] L. Olejnik, C. Castelluccia, and A. Janc. "Why Johnny Can't Browse in Peace: On the Uniqueness of Web Browsing History Patterns". In: *5th Workshop on Hot Topics in Privacy Enhancing Technologies*. July 2012.

[13] L. Olejnik, T. Minh-Dung, and C. Castelluccia. "Selling Off Privacy at Auction". In: *Network and Distributed System Security Symposium*. Feb. 2014.

[14] PageFair. *The State of the Blocked Web*. Feb. 2017.

[15] PageFair and Adobe. *The Cost of Ad Blocking*. Aug. 2015.

[16] F. Papaodyssefs, C. Iordanou, J. Blackburn, N. Laoutaris, and K. Papagiannaki. "Web Identity Translator: Behavioral Advertising and Identity Privacy with WIT". In: *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*. Nov. 2015, 3:1–3:7.

[17] S. T. Peddinti and N. Saxena. "On the Privacy of Web Search Based on Query Obfuscation: A Case Study of TrackMeNot". In: *Privacy Enhancing Technologies*. July 2010, pp. 19–37.

[18] E. Shi, T.-H. H. Chan, E. Stefanov, and M. Li. "Oblivious RAM with $O((\log N)^3)$ Worst-Case Cost". In: *Advances in Cryptology – ASIACRYPT 2011*. Dec. 2011, pp. 197–214.

[19] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas. "Adnostic: Privacy Preserving Targeted Advertising." In: *Network and Distributed System Symposium*. Mar. 2010.

[20] TRUSTe. *2011 Consumer Research Results: Privacy and Online Behavioural Advertising*. July 2011.

[21] 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*. July 2012, 4:1–4:15.

[22] J. Wang, W. Zhang, and S. Yuan. "Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting". In: *ArXiv e-prints* (Oct. 2016). arXiv: `1610.03013 [cs.GT]`.

[23] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen. "How Much Can Behavioral Targeting Help Online Advertising?" In: *Proceedings of the 18th International Conference on World Wide Web*. Apr. 2009, pp. 261–270.