Targeted, Not Tracked: Client-side Solutions for Privacy-Friendly Behavioral Advertising

Mikhail Bilenko, Matthew Richardson, and Janice Y. Tsai

Microsoft Corporation
1 Microsoft Way, Redmond, WA 98052
{mbilenko,mattri,janice.tsai}@microsoft.com

Abstract. The current discussion of potential *Do Not Track* regulation for online advertising is worrisome for the advertising industry, as it may significantly limit the capability for targeted advertising, a key revenue source for online content. The present discourse conflates the behavior tracking and ad targeting processes, leading to the presumption that providing privacy must come at the cost of eliminating advertisers' targeting capability. This paper focuses on a family of methods that facilitate behavioral targeting while providing consumer privacy protections. This is achieved by differentiating between client-side and server-side tracking. Client-side solutions provide for mechanisms and policies that address the privacy concerns over lack of user control over data while providing advertising platforms with the ability to target users. We compare and contrast several client-side methods along several dimensions of user privacy, adoption effort, and trust. A novel client-side profiling method is proposed that differs from prior work in not requiring installation of additional software by the user and providing compatibility with existing ad serving infrastructure. Empirical evaluation of the method on large-scale real-world datasets demonstrates the potential for high targeting performance of client-side techniques. We hope that by considering such middle-ground approaches, the present debate will converge towards solutions that satisfy both advertisers' desire for targeting and users' desire for privacy.

1 Introduction

Privacy concerns related to online advertising have grown over the past several years among web users, as reflected in coverage of the issues surrounding "behavioral tracking" in the popular press (e.g. the Wall Street Journal's "What They Know" series [15]). These concerns have attracted government attention, resulting in regulatory proposals [2], workshops [1], as well as Congressional hearings and legislation. The combination of popular sentiment, media attention and government involvement is likely to result in action to protect consumer privacy in online advertising in the near future. A series of discussions that has taken place recently involving legislators, advertising and technology industries, privacy advocates, and regulatory bodies has resulted in the *Do Not Track* (DNT) solution gaining the greatest momentum.

DNT relies on incorporation of privacy-protecting features in web browsers that either discourage or block communication with third parties that perform behavioral advertising. Different DNT implementations proposed by browser manufacturers can be grouped into three categories: domain blocking, opt-out cookies, and HTTP headers. Domain blocking allows the user to specify domains which the browser should never contact. In contrast, with opt-out cookies and HTTP headers the browser contacts the target domain but informs it that the user wishes not to be tracked. These latter two solutions require the user to trust that the target domain will comply. Evaluating these DNT variants in the context of the regulatory framework put forth by the Federal Trade Commission (FTC) [2] demonstrates that none of them currently meet all criteria in their formal interpretation: each of the three strikes a different balance between ease-of-use, universality and enforceability. Crucially, DNT has so far also failed to win the endorsement of the online advertising industry, which continues to advocate for self-regulation being sufficient.

In this position paper, we argue that the current discussions of *Do Not Track* are hampered by the lack of clear distinction between *tracking* (collection and aggregation user behavior data) and *targeting* (use of this data during ad selection). Demarcation of the two processes is crucial, as they are increasingly being performed by multiple parties, whose interactions are increasingly non-trivial both technically and financially. Additionally, because any party performing either tracking or targeting can also be the content publisher (first party), policies that do not distinguish these differences are inherently ambiguous and hence ineffective.

Differentiating between tracking and targeting has significant implications for protecting both consumer and industry interests. To advertisers and ad platforms, tracking is only a means to an end of increasing advertising effectiveness, which is achieved by targeting. For users, there appears to be a gap in attitudes towards data collection and targeted advertising. Survey results reported by Hallerman [9] indicate that 55% of respondents are very or somewhat comfortable with ad targeting, while another survey by McDonald and Cranor [11] found that over two-thirds of respondents have agreed or strongly agreed that “someone keeping track of my activities online is invasive”. While these results should be viewed in the context of the fact that most users lack fundamental understanding of how tracking and targeting work [10], they nonetheless indicate that the two processes are perceived differently.

There *is* a family of solutions that differentiates between tracking and targeting. In contrast to the existing DNT discussions and browser-based solutions, these *client-side tracking* proposals protect the privacy of users while still enabling advertisers to target ads to them. This makes such solutions attractive to both users and advertisers, and are an important category of solution that should be considered in any future discussions on DNT and behavioral targeting. Client-side tracking solutions store behavioral data on the user’s machine, giving users complete control over their data, ensuring that their behavioral information can be edited or deleted permanently as needed. In addition to providing contrastive analysis of previously proposed client-side tracking mechanisms, this position paper describes a new approach, *Client-only Profiles(CoP)*, demonstrating how it can be implemented using a recently proposed machine learning method for constructing compact profiles [4]. Unlike previously proposed solutions, CoP does not require a user-installed browser plugin, and relies on very minor modifications to the existing advertising infrastructure. We argue that client-side profiling strikes a balance between user privacy, advertiser revenue, and user and advertiser adoption, and that future exploration in this direction can yield effective alternatives to the binary policies currently being discussed.

2 Targeting vs. Tracking

To consumers, businesses, and advertising content providers, the term **tracking** may have different connotations. For this paper, we adopt the definition put forth by the Center on Democracy and Technology in the context of online behavioral advertising [3]: "Tracking is the collection and correlation of data about the Internet activities of a particular user, computer, or device, over time and across non-commonly branded websites for any purpose other than fraud prevention or compliance with law enforcement requests." Tracking can be performed by the first party or a third party, where the first party is the functional entity with which a user reasonably expects to exchange data, while the third party is any other other functional entity.

It is crucial to distinguish between **tracking** and **targeting**: the former refers to the process of data *collection* and *processing*, while the latter focuses on the *use* of processed data for personalization in the context of a specific task, such as advertising. This distinction is effectively disregarded in the ongoing public debates. To some degree, the confusion is a reflection of the overall poor level of understanding of tracking technologies, behavioral advertising, and the risks of related information exchange [11]. However, the distinction is critical, because it represents the fact that there are numerous parties involved in data collection, processing, and its use for advertisement selection. Both the tracking and targeting steps can be performed by several entities, and any policy or technology proposal must take the complexity of ad delivery pipelines into account to be effective and unambiguous.

The moniker *Do Not Track* masks this complexity and is ambiguous on multiple levels. In addition to conflating tracking and targeting, the public policy focus has been on "Do Not Track for the Purpose of Behavioral Advertising," which neglects tracking performed for non-advertising purposes (e.g., data collected could be used for differential pricing on retail websites). While targeting is always performed by one or more parties involved in advertisement selection as explained in the next section, tracking may be performed by any of the entities involved. In addition to parties involved in targeting, users may be tracked by specialized data aggregation firms (data exchanges), which subsequently sell the data data to interested parties, including advertising platforms. In addition, tracking may also be performed by the first party, which then provides aggregated information along with the ad request.

2.1 Tracking Mechanics

The architectural details of the tracking infrastructure are not widely publicized, and have mostly been revealed via reverse-engineering studies performed by researchers and journalists [5, 13]. The lack of transparency is not surprising given the combination of intellectual property value of the underlying technology and the sensitivity of the related privacy issues.

Third-party tracking is typically performed via an element of the viewed page that sends an HTTP request to the tracking server, passing the properties of the current context and client-side identification data. The sending element is typically either a tracking pixel (small image hosted on the tracking server), or JavaScript code loaded with the page that sends the request. Identification of the user can be performed via an ID stored in a cookie, as well as via indirect methods such as relying on the first-party user ID encoded in the URL of viewed page, which is passed as the Referrer header. Additionally, it has been shown that combining standard request headers such as the IP address and UserAgent string can lead to high-precision identification [6].

User and context properties passed to the tracking server may include attributes of the viewed page. These attributes may include its category in some publisher-defined taxonomy, or a search query if the page was visited via a link from a search engine. In addition, the publisher may provide any additional user data, such as user-submitted demographic or location data, or data summarizing prior behavior of the user. Tracking servers also receive user data that it stores client-side, e.g., in its cookie, or in specialized plugin cookies (known as local shared objects), or in HTML5 browser local storage. Tracking data may also include demographic and location data (self-reported or inferred), attributes derived from user browsing activity (e.g., inferred interest categories), or specialized features (e.g., identifiers encoding specific products which the user has viewed on a retail site). The tracking platform can either store all user-related information on the client-side, which requires updating the information regularly, or utilize server-side storage.

2.2 Targeting Mechanics

The effectiveness of advertising is directly affected by the availability of data used to estimate the user's potential responsiveness to the advertisement creative or the underlying product. Such information may come in various forms: demographic and location information, aggregated information about user's past behavior, and raw behavioral information can all be factors that affect user's desirability to the advertiser. Access to additional information describing the user beyond the context in which the advertisement is served allows advertisers to modulate the prices they are willing to pay to optimize their return-on-investment (ROI).

Advertising platforms that perform their own tracking utilize the tracking information in ad selection. In recent years, an increasing volume of display advertisements is allocated via ad exchanges, also known as real-time bidding (RTB) platforms [12]. A content publisher sends an ad request to the ad exchange, which then forwards it to multiple advertising networks. User information collected via tracking may be provided by the first party (the publisher), or by any of the involved third parties: the ad exchange, ad networks submitting the bids, or dedicated data exchanges (tracking-only companies), which may partner with any of the former entities. As a result, actual ad selection may involve not one, but multiple user profiles accumulated by a number of agents, where any of them can be stored either server-side or client-side. Matching of user IDs is an additional complication, which is resolved by a establishing the mappings across the different parties and caching them, typically performed by the exchange.

3 Targeted, But Not Tracked

This section discusses how the distinction between targeting and tracking can be exploited by technological approaches which reduce or prevent user tracking, while allowing advertising networks to retain all or most of the revenue gains achieved from targeting. These solutions make use of *client-side* aggregation of personal data, which allows the user to be targeted while leaving user in posession of their data. It is know that major concerns over behavioral advertising include users' lack of control over the data describing their past behavior, as well as the insufficient transparency of the data collection and retention, as evidenced by users' poor understanding of these processes and policies [10]. Client-side user profiling solutions respect users' desire for privacy while maintaining the advantages (both to users and service providers) that comes with behavioral advertising, and hence provide a strong alternative to clear-cut, binary solutions like *Do Not Track*.

First, we give an overview of three recently proposed solutions in this space, which require the user to install a custom plugin in their browser that performs behavioral tracking and advertisement targeting. Then, we present a novel approach, Client-only Profiles (CoP), which naturally fits into the existing advertising ecosystem. CoP does not require any installation or actions by users or advertisers, and requires only minimal changes from advertising networks. We also summarize relevant results of an empirical study for one possible algorithmic implementation on which CoP can be based: a machine learning method titled *Predictive Compact Profiles* [4]. These results show that the CoP approach can potentially retain nearly all of the revenue gains obtained from behavioral ad targeting, while preserving users' right to privacy and control of their data.

3.1 Plugin-Based Client-Side Profiling

Plugin-based client-side solutions make use of a browser extension installed on the user's machine to incorporate user preference in advertisement selection. The plugin maintains a collection of the user's browsing and behavioral data on the user's machine, and uses it to facilitate targeting during ad selection. The three primary approaches that fall into this category are Privad [8], Adnostic [14], and RePRIV [7].

Privad [8] exemplifies an approach whose goal is complete user privacy. User behavior is monitored by a plugin installed on the client machine that maintains a user profile built from it. Ad platform's server provides the plugin with all (or a large subset) of the potential ads that may be displayed, from which the plugin selects the actual ad to be displayed utilizing the profile to achieve targeting. Ad impressions and clicks are encrypted and passed through a third-party dealer which is not able to view the information, but anonymizes its source before passing it to the ad network (i.e., hides the user's IP address from the ad network). Thus, the ad network does not know which ads were shown to or clicked by which users, but obtains aggregate statistics.

Adnostic [14] takes a similar approach to Privad: a browser plugin selects the ad to display with the aid of a locally constructed profile. In contrast to Privad, Adnostic takes the view that ad impressions should be kept hidden from the service provided, but not ad clicks. This makes the ad platform less vulnerable to click fraud, but also reveals the targeting attributes of a user when that user clicks on an ad. When a user visits a webpage, the ad network sends 10 to 20 ads which can be displayed, and the client plugin selects one of these based on local targeting attributes. The information about which ad is displayed is encrypted and provided to the ad network in a form that prevents the network from knowing which ad was shown. Occasionally (monthly, for example), aggregated encrypted data is provided to a trusted third party that decrypts it and informs the network how many times each ad was viewed.

Both Privad and Adnostic make fraud detection difficult for ad platforms (though less so in Adnostic), which is a significant concern from the perspective of ad platforms. Both approaches also increase network traffic and page load times since they move a significant portion of the ad selection step to the client along which requires transferring the ad inventory. Another concern is advertiser budget constraints: with Adnostic, an advertising platform must estimate when an ad's budget will expire in advance before sending it to the client, which may lead to ads being shown too many or too few times depending on the quality of the prediction. Furthermore, these two approaches take a significant portion of control over tracking and targeting out of the hands of the advertising network, reducing its ability to innovate and experiment with new targeting methods.

RePriv [7] takes a slightly different approach to targeting without server-side tracking. It constructs user profiles from the raw browsing data on the client machine, and sends the profiles up to ad platform's server to facilitate targeting server-side. In contrast to the previous two approaches, this allows the ad network to view user data and perform whatever personalization it desires at the time the ad is requested. The ad network can also provide custom miner modules to the client that extract data from the user's raw behavior, allowing the network to develop more complex targeting mechanisms. The user has the option to review the data that will be sent to the ad network, and either approve or disapprove of its release. Since the ad network is directly selecting ads and recording clicks, this approach solves many of the difficulties that Privad and Adnostic have with regard to fraud, budgets, and innovation.

3.2 Native Client-Side Profiling

This section introduces a novel approach, Client-only Profiles (CoP), that also stores behavioral information on the client but, unlike the methods discussed in the previous section, does not require the user to download and install a custom browser plugin. Most importantly, CoP gives users control over their data while allowing platforms to target advertisements without making significant structural changes to the current delivery or pricing mechanisms.

In CoP, user behavior is maintained in aggregated form along with a cache of raw recent behavior in the browser cookie associated with the ad network. The ad network receives the cookie with the ad delivery request from the user's current context (the web page the user is loading), and returns targeted ads with an updated cookie containing a refreshed profile. As with RePriv and Adnostic, while the ad network does receive the raw user browsing behavior, it is bound by policy to discard the information once it has returned the targeted ads and the cookie to the user; it also must not store the user ID with any logs of ad impressions and ad clicks. Thus, the only record of user behavior is maintained on the client in the cookie, leaving the user with the option of deleting their profile at any time, knowing that there are no records associated with them remaining on the server. Because this method relies on policy compliance by ad networks, it is not enforceable. However, this is the same assumption that the leading DNT solutions (Opt-out cookies and HTTP header) and RePriv make. Given that policy violations are detectable (by manipulations of client-side data), and bear significant legal and public relations ramifications, compliance assumptions are reasonable. We also note that CoP can be implemented to incorporate server-side encryption of profiles, preventing their interception in non-secure HTTP traffic.

CoP requires ad networks to construct incrementally-updated user profiles to facilitate targeting. In our initial implementation of CoP for search advertising, we model the gains from ad targeting by considering *bid increments* that advertisers can specify along with their bids. Increments are triggered when the user has shown a past and has expected future interest in the ad's topic. Bid increments are commonplace in display advertising platforms, however they are based only on explicitly known demographic attributes, or broad, loosely defined segments. In search advertising, advertisers have an analogous interest in adjusting their keyword bids for users known to have had a past interest in the keyword's topic.

To maintain a profile of the user's predicted future interests, different profile constructions can be employed. Here, we summarize a machine learning based approach recently proposed in [4]. The approach utilizes features the encode recency and frequency of past behavior associated with a keyword and its neighbors (related keywords), as well as context-independent keyword and user properties. For each candidate keyword considered for inclusion in the profile, a scoring function trained via a machine learning approach predicts, for that user, the likelihood that the user will click on an advertisement associated with the keyword in the future. Based on predicted probability, top-k keywords are selected to comprise the user's profile, which is then used to trigger bid increments during future selection. Targeting is thus performed under the restriction that the only available user information is that which is stored in the client-side cookie on the user's machine, yet is able to closely match the predictive accuracy of targeting that relies on the user's complete behavior history. In the next section, we summarize an empirical study fully described in [4] that examines whether revenue is lost in moving from traditional server-side tracking to client-side profiling based on a particular algorithmic implementation of profile construction in CoP.

3.3 Revenue Impact Analysis

Client-side vs. server-side keyword profiles were evaluated using two months of search and advertising behavior logs for 2.4 million users of the Bing search engine, sampled randomly from the overall, larger pool of bot-filtered US-based active users, where active users were defined as those users who had used the search engine (issued at least one query) on at least 30 of the 60 days in the time period. The first six weeks of data were used for training the machine-learning based predictor used for keyword selection for profiles. Training is performed by simulating the profile construction process and utilizing the subsequent behavior to obtain a training label (ad click or lack thereof) for every keyword that was a candidate for inclusion in the profile. With the utility predictor trained on the first six weeks, the efficacy of online profiling was evaluated by simulating profile construction over the seventh week, using behavior following the construction period to estimate utility.

Client-side profiles contain two portions: the profile, which matches keywords, and an additional keyword cache used to enhance the profile construction candidate pool. Both components were constrained to be no greater than 100 keywords, which ensures that profiles fit the 4KB cookie size limit. Profile construction uses the utility model that corresponds to matching the keywords for which future ad clicks are observed, which is equivalent to bid increments. Cache construction is more straightforward: it is based on least-recently-used caching, a standard approach that typically has good performance and is efficient to compute.

Utility is reported as the fraction of ad clicks in post-profile-construction behavior for which the profile matched the bidded keyword, and hence would have triggered the bid increment for advertisers who specify it. This metric, percentage of incremented clicks, can be viewed as the percentage of ad revenue that would be increased via increments.

Figure 1 illustrates this relative performance of client-side profiles (with their limited knowledge of user history) with respect to server-side profiles for different profile sizes, which correspond to server-side tracking. The figure demonstrates that maintaining a modest cache size alongside the profiles allows achieving targeting performance comparable to that of server-side profiling, but without the need to track user behavior server-side. For example, if profiles are limited to 20 keywords, utilizing a cache of the 50 most recent queries allows capturing 97% of the revenue gain achieved by server-side tracking while providing users

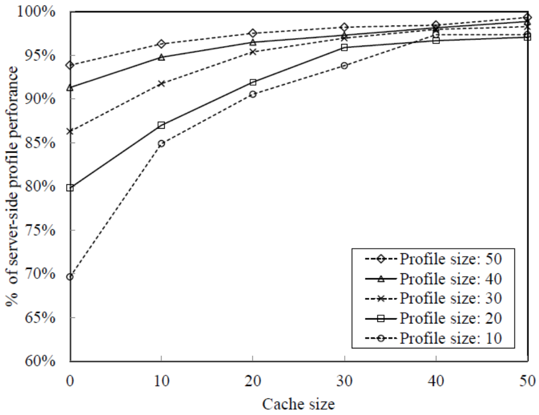

Fig. 1. Relative client-side profile utility (utility as a proportion of server-side utility for profiles of the same size)

full control of their data. These results demonstrate the practicality of allowing users to opt-out from server-side tracking with minimal revenue or performance cost. Complete description of the methodology and extended results are available in [4].

4 Discussion

These four approaches make different trade-offs in a space with dimensions of efficiency, changes from the existing infrastructure, guaranteed privacy, flexibility of targeting mechanism, and more. In Table 1, we compare the four methods, along with the existing methodology along these dimensions.

The dimensions are:

- **Profile Construction**: Is the profile constructed on the client (**C**) or server (**S**)? The user profile consists of the behavioral data stored for the user. RePriv constructs the profile on the client, but using code that can be sent from the server down to the client, indicated by C(*).
- **Profile Storage**: Is the profile stored on the client or server? Profile storage determines what party has control over the profile and the ability to manipulate or delete it.
- **Ad Targeting**: Is ad targeting performed by the client or server?
- **User Control over Profile**: Does the user maintain full control over the content of their profile?
- **Can target based on (non-contracted) third party behaviors**: Does the system allow ad networks to utilize user behaviors on web sites with which the ad network has no relationship?
- **User behaviors revealed to platform:** Which of the following user behaviors are revealed to the ad platform, tied to a particular user:

  - **Page Visits**: User visits to pages on which the platform serves ads.
  - **Profiles**: Profile information derived from raw user behavior.Derived user information, such as aggregate counts of the number of times the user has visited a page of a particular topic.
  - **Ad Impressions**: The ads viewed by the user. *agg* indicates that this information is provided in aggregate.
  - **Ad Clicks**: The ads clicked by the user. *agg* indicates this information is provided in aggregate
- **Required Changes To**: What changes would be needed in order to implement the proposed solution, by:
  - **Client**: The user's machine. "Plugin" means a Web browser plugin would have to be installed by the user.
  - **Advertiser**: The advertiser. "Slow stats" means that the advertiser will not receive real-time statistics about their advertisement budget and performance.
  - **Platform**: The advertising platform (network). Indicates a general notion of the quantity of work required for the platform to implement the system and change their targeting to work within the proposed system.
  - **Extra Parties**: Are there any additional new parties (besides the client, advertiser, and platform) necessary? "Online dlr" refers to the *dealer* required by Privad, which must be online with high availability. "Offline ttp" refers to the *trusted third party* by Adnostic, which needs only process data occasionally (they suggest once a month).
- **Requires unified topic taxonomy**: Does the system require ad networks to agree on a single unified behavioral targeting topic taxonomy?
- **Requires trust in platform**: Does the user need to trust that the company running the advertising platform is trustworthy? This is simply a reflection of the breakdown of which data is received by the advertising platform shown in the middle portion of the table.
- **Increased traffic from ads**: What (if any) is the increase in the network traffic between the user and the platform due to communicating more ads?
- **Increased traffic from profile**: What (if any) is the increase in the network traffic between the user and the platform due to communicating user profile information?

One of the primary divisions between methods is whether the ad personalization is done by the client or the server. In Privad and Adnostic, personalization is done on the client, whereas for RePriv and CoP, it is done on the server. This design decision has important implications. Methods which personalize on the client require multiple ads to be downloaded per page view, causing increased network traffic and/or slower page load times. Methods which personalize on the server necessarily must reveal to the server which ads were shown to the user, and also must pass up to the server any profile information that is used for user targeting. These latter methods rely more heavily either on providing the user controls to edit what is sent to the servers, or make assumptions about server

Table 1. Comparison of various proposals for respecting user privacy while still enabling ad targeting.

Concept	Privad	Adnostic	RePriv	CoP	Today
Profile Construction	C	C	C(*)	S	S
Profile Storage	C	C	C	C	S
Ad Targeting	C	C	S	S	S
User Control over Profile	Y	Y	Y	Y	N
Can target based on (non-contracted) third party behaviors	Y	Y	Y	N	N
User behaviors revealed to platform:					
Page Visits		Y	Y	Y	Y
Profiles			Y	Y	Y
Ad Impressions	agg.	agg.	Y	Y	Y
Ad Clicks	agg.	Y	Y	Y	Y
Required changes to:					
Client	plugin	plugin	plugin		
Advertiser		slow stats			
Platform	lots	lots	minor	very minor	
Extra Parties	online dlr	offline ttp			
Requires unified topic taxonomy	yes	yes	no	no	no
Requires trust in platform	no	some	some	yes	yes
Increased traffic from ads	download all ads	download 10x ads			
Increased traffic from profile			profile	cookie	

policies regarding retention and use of the user data in the future (in the case of CoP, the policy is that the server is not allowed to remember the profile after it has been used to personalize the ad). A primary advantage of personalization on the server is that it enables up a wider variety of personalization methods. In the case of ads, not only can the server select which ad to show, but also vary the ranking of a set of ads, as well as charge differentially for them. Further, servers may choose to change the ad copy (title, text, URL) or appearance of the ad as well. By assuming that ad personalization is equivalent to simply selecting one of N ads, Privad and Adnostic are provide a much more limited set of personalization options to the ad network.

A second significant difference between approaches is whether the ad network has the ability to develop targeting techniques. Much of the literature assumes that targeting should be done based on categorical membership. In our experience, other targeting methods can be as or more effective, such as fine-grained keyword-based targeting. Other platforms may find other attributes more beneficial, such as estimating user demographics, or geographics. Techniques which perform personalization on the client must also prescribe how that personalization is done. This means all ad platforms will be required to conform to a single personalization technique, a goal that is undesirable and untenable in our opinion. For example, Adnostic uses a single categorization system built into the client based on natural language processing heuristics. Ad networks are unlikely to want to give up control over their category schema or the ability to develop more and more advanced techniques for categorizing users into their schema.

One advantage of the plugin-based approaches is that they provide ad networks with the opportunity to target ads based on a user's entire browsing history, as opposed to just the portion of the history for which the ad network was able to observe the user. On the other hand, plugins must be downloaded and installed by end-users, increasing the difficulty of method adoption. Further, plugins solutions are less flexible, since they constitute executing modules that are shared across multiple ad networks, requiring cross-network agreement for modification and potentially slowing the rate of innovation and progress in tracking and targeting methods. RePriv circumvents this difficulty by providing a mechanism for servers to send routines to the plugin in a secure and verified manner, but this openness comes at a cost – RePriv asks users for permission when new routines are installed, and, more importantly, each time user profile information is sent to the ad platform.

Taken as a whole, we believe the CoP approach, with its flexibility, efficiency, and similarity to the existing ad serving infrastructure, strikes the most effective balance between users' privacy, ad networks desire to personalize, and feasibility of implementation.

5 Conclusion

In this paper, we focus on the current conflation of tracking and targeting in policy discussion, which leads to the false assumption that any method that allows users to opt out of tracking must also neccessarily prevent ad targeting. This leads to ambiguity because different parties may be performing tracking and targeting, resulting in a popular misunderstanding that it is only possible to either have both tracking and targeting operating, or neither.

Distinguishing between tracking and targeting leads to a family of middle-ground methods which are required to store behavioral profiles locally on the user's machine, while allowing the ad platform to target ads. Three previously proposed approaches in this family all assume installation of client-side plugins and a topic-based representation of user profiles. The paper presented a novel approach, CoP, which also relies on client-side profile storage, but departs from prior work in not requiring additional software or a singular profile representation, making it directly compatible with existing advertising platform infrastructure. A possible implementation of CoP using a recently proposed machine learning algorithm for compact profile construction [4] is discussed. Empirical evalution on a large-scale, real-world dataset comparing CoP with traditional server-side tracking demonstrates that client-side approaches have the potential to give users control of their data without significant losses of revenue for the advertising industry.

References

1. Exploring privacy: A roundtable series. *Federal Trade Commission*, 2010.
2. Protecting consumer privacy in an era of rapid change: A proposed framework for businesses and policymakers. *Federal Trade Commission*, December 2010.
3. What does "Do Not Track" mean? a scoping proposal by the Center for Democracy & Technology. *Center for Democracy and Technology*, January 31 2011.
4. M. Bilenko and M. Richardson. Predictive client-side profiles for personalized advertising. In *Proceedings of 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-11)*, August 2011.
5. P. Eckersley. How online tracking companies know most of what you do online. *Electronic Frontier Foundation*, September 2009.

6. P. Eckersley. How Unique Is Your Web Browser? In *Privacy Enhancing Technologies Symposium*, pages 1–18. Springer, 2010.
7. M. Fredrikson and B. Livshits. RePriv: Re-Envisioning In-Browser Privacy. In *Proceedings of the IEEE Symposium on Security and Privacy*, May 2011.
8. S. Guha, A. Reznichenko, K. Tang, H. Haddadi, and P. Francis. Serving ads from localhost for performance, privacy, and profit. In *Proceedings of the 8th Workshop on Hot Topics in Networks (HotNets-09)*, 2009.
9. D. Hallerman. Behavioral targeting attitudes. *eMarketer*, June 2008.
10. A. McDonald and L. Cranor. Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising. In *38th Research Conference on Communication, Information and Internet Policy*, 2010.
11. A. M. McDonald and L. F. Cranor. Americans' attitudes about internet behavioral advertising practices. In E. Al-Shaer and K. B. Frikken, editors, *WPES*, pages 63–72. ACM, 2010.
12. S. Muthukrishnan. Ad exchanges: Research issues. In *Proceedings of 5th Workshop on Internet and Economics(WINE)*, 2009.
13. S. Schoen. New cookie technologies: Harder to see and remove, widely used to track you. *Electronic Frontier Foundation*, September 2009.
14. V. Toubiana, H. Nissenbaum, A. Narayanan, S. Barocas, and D. Boneh. Adnostic: Privacy preserving targeted advertising. In *Proceedings of the 2010 Network and Distributed System Security Symposium*, 2010.
15. J. Valentino-DeVries. What they know about you. *The Wall Street Journal*, July 31 2010.