Targeting privacy issues of online behavioural advertising

Bert Tilmans, Jeroen Slobbe, Robbert van den Berg

1 Introduction

The online advertising market topped a record revenue of 26 billion US dollar over the year 2010 [10]. This market is still growing and many advertisers and marketeers make use of it for their campaigns.[10]. Given these figures, it is not surprising that the advertising business is constantly looking for methods to improve ad responses and thus revenue. A popular and successful method is to target ads specifically to each visitor by creating a profile with personal information about this visitor [11]. This is referred to as targeted advertising or Online Behavioural Advertising (OBA). The success of targeted advertising also introduces problems. First of all, OBA relies on personal information, which gives rise to privacy issues. Second, OBA can be very discriminating or manipulative - think of a case in which a specific website visitor receives higher pricings for products than another visitor, only because one visitor has a different 'shopping pattern'.

One often hears that "Behavioural targeting is inherently in conflict with privacy"[16]. While this may or may not be true, effort should be made to decrease the privacy issues to a minimum, i.e. with a Privacy Enhancing Technology (PET). In this paper we will first discuss how OBA works and explain the benefits. After that we will analyse the techniques used for OBA and the privacy issues. Finally we will introduce and apply a framework for evaluating three PETs which we will extensively analyse, testing their privacy enhancing abilities and weaknesses.

2 Targeted Advertising

Targeted advertising was invented because conventional banners and ad systems were shown to be unpopular amongst web users. The repetitive and obtrusive nature of the ads were common reasons for unpopularity. Besides, ads often proved to be completely irrelevant to the user's interest [9].

2.1 Terminology

Before explaining how targeted advertising works, some relevant definitions are presented. The terms *targeted advertising* and *online behavioural advertising* (OBA) are considered equivalent in this paper. OBA can be defined as *the tracking of the web user's online activities - including the conducted searches, visited web pages and viewed content - in order to deliver targeted ads to each user's individual (estimated) interest* [15]. In contrast to OBA, the classic approach for online advertising was based only on contextual advertising. Contextual advertising consist of ads placed within the content of a generic, third-party web page [8].

Furthermore, in this paper, privacy is considered as the right to seclude oneself or information about oneself and to determine whether, when, how, and to whom, one's personal or organizational information is to be revealed [4].

2.2 Process flow

Nowadays, behavioural advertising is a sophisticated process involving many actors. It starts with advertisers trying to get the attention of potential customers by advertising. Obviously they like to spend their advertising budget as effective as possible. Advertisers make use of the ad networks

(e.g. Doubleclick) because of their ability to effectively spread ads. The ad network does this by harvesting information about web users (e.g. age, gender, whereabouts, interests etc). This information allows the ad network to specifically target the users that are most likely be interested in the advertiser's services or products. The ad network then places the ad on a suitable website controlled by a third party, referred to as ad publisher. The publisher in return gets paid for the ad placement [11] [6].

In *Table 3*, the actors which are strongly related to the targeted advertising process are summarised. Also their goals and (business) activities are specified. In the next paragraph, a more detailed description of each actor's benefits (goal/interest) will be indicated.

Table 1: Table of involved actors

Actor	*Activity*	*Goal/Interest*	*Example*
Web users	Clicks on ads	Accurate ads, free services	Alice/Bob
Advertisers	Pays for ad placement	Reach correct audience	Sellers
Ad publishers	Publishes ads on page	Earn money	NY Times
Ad networks	Buy and sell advertisements	Earn money	AdBrite/Doubleclick/Facebook

2.3 Advantages

The OBA process has a number of benefits compared to the classical approach. A clear advantage of behavioural advertising over contextual advertising is the fact that it is more likely to show ads in the viewer's spectrum of interest. Therefore, advertisements which make the user feel annoyed are less likely to pop-up. Another advantage is that publishers offering free content or services (e.g. Gmail or Youtube) can be financially supported by the ad system, allowing them to stay online even without fixed income. Furthermore, free content could be offered to the end-users.

- Ad network
  The ad network makes money by charging the advertisers. Charging could be conducted within different business models, for example *pay per click* or receiving a percentage of a successful buy. In return to the advertisers, using each user's profile, ad networks can guarantee higher probability in placing ads which are that user's interest.

- Advertiser
  The advertiser pays the ad network(s) to gain more confidence in spreading their ads to interested users than to uninterested ones. This of course should result in having more users buying the advertiser's products or services and thus increasing the advertiser's revenue.

- Ad publisher
  The publisher acquires money for every view, click, successful sell or successful registration initiated by a user. For having a user click on an ad, the ad publisher will receive more money than when only an ad is viewed. If the ad network supplies more advertisements which are in the user's interest by applying behavioural advertising, the ad publisher will acquire more money.

- Web user
  OBA enables the user to receive more interesting, and thus less annoying, advertisements. Furthermore, thanks to the financial support to publishers by OBA, the user can enjoy many more free online services.

2.4 Techniques

OBA completely depends on the creation of accurate user profiles by the ad network. Whenever a user profile is incomplete or incorrect, the user will receive ads which he may not like . Therefore,

the advertisement industry uses several techniques to create user profiles which are as complete and correct as possible. Below, several techniques are listed and elaborated.

- Cookie

  A cookie is a piece of text received from the web server - when a user is visiting a web page - and is saved on his computer by the web browser. The cookie can be used to store site preferences, authentication information, session identifiers, a unique user identifier and more. Whenever visiting this specific web page again, the cookie is sent back to the web server which uses this cookie again. For example, Google Adsense stores a cookie with a unique number on your computer so they can track you when visiting ad publisher's web pages (connected to Google Adsense's ad-network) [1].

- Web bug

  A web bug usually is a small, mostly invisible, part of a web page or non-plain text e-mail which allows the ad network to track who has viewed the web page or e-mail. When the web page is viewed, the web bug will be downloaded from the ad network's web server, which can create a list of user specific ip addresses and corresponding web bugs downloaded (and thus visited web pages). Even more user specific information can be stored when client-side Javascripts are used. Using web bugs does not require to have client-side information storage (as with cookies), which makes it harder to protect from.

- Web browser fingerprinting

  A web browser fingerprint contains information about the browser version, operating system it runs on, installed plug-ins and more. All this information together can make a web browser uniquely identifiable. The Electronic Frontier Foundation (EFF) published [2] an online tool which compares your web browser's fingerprint to many other - already scanned - web browser's fingerprints, and tells if your web browser is uniquely identifiable amongst those. Results of this EFF project are presented in Eckersley's paper [2] and says that amongst almost 500,000 (privacy conscious) users, 83,6% of the web browsers fingerprints had an instantaneously unique fingerprint. Since no client side storage of information is needed, protection from web browser fingerprinting is hard.

2.5 Privacy Issues

As explained, the OBA system deals with lots of privacy sensitive information about users. Next, several issues are discussed.

2.5.1 User information leaks

One of the core issues with collection user information is keeping it secret to the public domain. Probably the most famous case is the AOL scandal [5]. In 2006 AOL published an 'anonimised' list of search queries of 650,000 users, not expecting lots of information would be extracted through correlation of the query keywords and personal identifyable information (which were present in many queries). This case was a clear example that seemingly small bits of information can lead to unique identification of a user.

2.5.2 Users expectations are violated

The so called *informational privacy* - the privacy to control ones personal information - is currently being violated because users have no control of their personal information being stored by ad networks and the other actors. Most users do not know who has access to their personal information, where it is stored and for how long. Let alone, having users not even realising the existence of the personal information gathering ad networks, thanks to the seemingly invisible behaviour of profile creation by ad networks [13]. The latter case is a big issue since users can not protect themselves from (privacy violating) risks they don not understand [13].

But even if users delete their visible and controllable identifiers like cookies, then web bugs, spyware, hidden identifiers (browser fingerprints) - which can be collected by XSS-attacks - can still be used to collect privacy concerning information. Then, having (legally mandatory) opt-in systems do not guarantee no data will be collected. Besides, how many people actually read the privacy consents? Distrust in the 'advertising world' makes users think personal information will be collected anyway, even if they pay for certain privacy preserving services [13].

2.5.3 Potential useless/unknown information collection by companies

Often companies blindly store every piece of data they collect, possibly because they hope it may become useful some day. Or even worse, because they don't know it exists anymore. Some of this data is used as test data for experimental systems. Other data are leftovers from legacy or unused systems, e.g. when two companies merged. Another place where outdated information is stored is in older backups. Good back-up policies facilitate in daily, weekly, monthly and one year back-up. This means that some piece of information, which is ordered (by court or individually) to be removed, still remains in the company for at least a year. Most companies store hard drives of older machines in case a system crash occurs. Most of the time everything is deleted. Unfortunately, a default delete command does not erase the file but only the reference to it. This way, recovering the data from the seemingly erased storage device is still possible.

2.6 OBA versus privacy, the core issue

As illustrated in the previous sections, OBA is a large improvement over classical advertising systems, but comes at a large price: the violation of users privacy. Internet users do not have the ability to control if and how much (personal) information they supply or leak to the ad networks and in lesser amount to the advertisers and ad publishers. This shortcoming will be the core issue addressed in this pape. More accurately, we state the core problem as:

Analyse the possibility of an ad system which allows tailored ads while maintaining the user's privacy

In this paper, we will analyse several privacy enhancing technologies (PETs) which would ensure that OBA and user's privacy are no longer mutually exclusive. We will do this by conducting a literature study on several recently published papers which propose a solution for the privacy problems of OBA. The validity, feasibility and usability of the proposed PETs will be checked as well as a comparison between the systems. Furthermore we will add some suggestions and check at what degree our defined core problem has been covered.

3 Solution (PETs)

In order to mitigate or solve the problems as defined in section 2.6, below the PETs *Adnostic*, *PPOAd* and *RSA group model* are evaluated.

3.1 Current inadequate solutions

Currently, an accessible method to improve ones privacy with respect to OBA are browser plug-ins that somehow try to prevent tracking or otherwise try to prevent the creation of an accurate user profile by the ad networks.

- Tor

  Tor is a privacy protecting system which *prevents anyone from learning your location or browsing habits*[3]. It works by tunnelling and encrypting web traffic through a non-fixed network of computers (proxies), thereby hiding identiers like ip-addresses. Unfortunately, a user may still be uniquely identifiable, e.g. by using browser fingerprinting or correlating search queries or browsing patterns.

- Track me not

  An alternative but popular approach is to confuse the tracking systems by obfuscating search and browse behaviour. This is what the Firefox plug-in *track me not* does. Periodically it sends search queries to various providers such as AOL, Yahoo and Google.

Both solutions decrease the ad network's effectiveness through obfuscation. This is inadequate because it still leaks information and leads to inaccurate ads. A PET which takes care of the privacy issues, as stated in chapter 2, and preserves the effectiveness of OBA at the same time is desirable.

3.2 PET rating framework

In order to rate and compare the three evaluated PETs (see paragraph *3.2 - 3.4*) the following framework (*Table 2*) will be used. To make clear what the balance is between the advertising world's interest in OBA and that of the users, several criteria are used. It is necessary to have the criteria because:

- Integration
  Having a PET which will take a huge amount of effor to be operational is less appealing for ad networks and users. Even when it takes small effort for the user only, then still the ad networks are probably not eager to spend huge amounts of money for the sake of the user. Unless, of course, having a PET is obligated by law and no other PETs are more suitable. Therefore, having a PET which requires little adjustments and extensions to the current OBA system is desired.

- Scalability
  Whenever a PET will decrease the user experience in such a way - e.g. by having high latency because of long cryptographic computation times - the users will get annoyed, then they probably will stop using the PET. Same holds for the ad networks, but then the experience relies on the decrease in effectiveness of their systems caused by the *expensive* (cryptographic) computation times.

- TTPs
  Having TTP(s) as part of PET is a discussible weakness. Many systems, using TTP(s), have proven to be secure and globally accepted (e.g. the Public Key Infrastructue for SSL certificates). Though, keeping the tremendeous amount of 26 Billion US dollars of OBA revenue in mind, the probability of having the ad networks colluding with the TTP(s) is non-negligible.

- Attack and Attack impact
  Obviously, having a lower likelyhood of attacks and lower negative impacts on the things PETs try to achieve (e.g. privacy of user profiles) is better than having the opposite case. However, a decision between having a low likelyhood together a large impact and having a high likelyhood together with a small impact, may not be clear. Though, since interesting systems will get hacked anyway at some point, having a lower impact is more important than having a low likelyhood.

3.3 Adnostic

Toubiana et. al.'s proposed and created a PET called *Adnostic*, which is a client-side tool. The main goal is to preserve user privacy without diminishing the widely successful economic model currently used by ad networks. With this goal, several issues like billing, advertising budgets, network latency, targeting efficacy, malicious behaviour and others had to be addressed. In order to fulfil these goals, Toubiana used elliptic cryptographic functions, a Trusted Third Party (TTP) and the widely used bookmark history on delicous.com.

Table 2: PET rating framework

Criteria	*Description*	*Rate (0-5)*
Integration	Does the PET integrate with the current system?	0:= 'no, many adjustments needed' 5:= 'yes, PET is extension-like'
Scalability	Is the PET scalable, i.e. is processing time or bandwidth not drastically increased?	0:= 'yes, processing time grows exponential' 5:= 'no, processing time grows polynomial'
TTPs	Are any TTPs part of the PET? If so, what is the privacy impact when collusion occurs between a TTP and one or more of the other parties?	0:= 'PPTs exist, major collusion impact' 5:= 'no PPTs, minor collusion impact'
Attacks	How likely are certain attacks to occur. est data for experimental syste	0:= 'very likely' 5:= 'not likely at all'
Attack impact	If a certain attack succeeds, what will be the the privacy impact?	0:= 'major impact' 5:= 'minor impact'

3.3.1 Design

Next we will give a short description of the basic elements of Adnostic. In the next paragraph the correctness, strengths and weaknesses of Adnostic will be reflected to our rating framework.

1. Profiling
   Profile creation happens client-side instead of server-side (i.e. for every ad network different servers and different profile creation methods). Adnostic makes use of a general segmentation table for user interests. For creating the profile, *Delicious.com* - a social bookmarking website - is used to derive a corpus of tags which will be used when categorising. Because of the general profile, all ad networks can make use of it, even the smaller ones. The big advantage of Adnostic over current OBA systems is the user's ability to control their profile.

2. Ads selection by server
   At the beginning of the billing period - say the beginning of the month - an encrypted counter will be set to zero: $C_{ID} \leftarrow E(pk, 0)$. By using existing contextual advertising and pricing algorithms, the ad network selects n ads from its ad database. The ad network remembers the IDs of the selected ads. Then the public key pk (that is only know to a Trusted Third Party, for which the use will become clear when we describe the billing step) and the ads will be sent as an ordered set $V = (pk, ad_1, ad_2, ..., ad_n)$.

3. Ad selection by Adnostic
   When Adnostic receives the ordered set V it selects one of the n ads to be displayed. It is not in the scope of this paper to describe the ad selection algorithm. Though, it may be important to note that the algorithm is a trade-off between the essentially arbitrary computations over the user's profile and browsing history on the one hand, and a method for rating ad's relevance to the user on the other. Now, Adnostic will create a binary vector $v = (v_1, v_2, ..., v_n) \in \{0, 1\}^n$ that represents which ad is shown to the user. Then, $W = (E(pk, v_1), E(pk, v_2)..., E(pk, v_n))$ is sent to the ad network. In order to address click-fraud, Adnostic does a zero-knowledge proof to the ad network to proof that for all v_is (with $v_i \in v$, $v_i \in \{0, 1\}$), and that Adnostic selected exactly one ad to be viewed, not more or less.

4. Billing
   When the ad networks received W and the zero-knowledge proof succeeded, a scalar c is applied to each element of W, resulting in the vector $W' = (E(pk, c \cdot v_1), E(pk, c \cdot v_2)..., E(pk, c \cdot v_n))$. Next the additive homomorphic property to add encrypted values to the counter C_{ID} is used. Then at the end of each billing period - say the end of the month - the counter C_{ID} is sent to the TTP, who will decrypt the vector W' and tells the ad network who to pay what (possibly together with a zero-knowledge proof).

3.3.2 Remarks about correctness and implementation

The ad selection algorithm used by Adnostic may be less accurate than the (bigger) ad networks's algorithms. This is a big drawback for the both the user and the ad networks, since possibly less relevant ads will be viewed. One way to solve this is to let the bigger ad networks make their algoritms available. But one can imagine that the ad networks are not very eager to hand out their algorithms. Furthermore, the n ads that are sent to Adnostic are selected using contextual advertising algorithms, which may be less effective than current OBA systems. Though integration of Adnostic will be relatively ease (install add-on, install TTP do small adjustments to current OBA systems), ad networks probably are not very eager to downgrade their ad targeting accuracy.

Selection of n ads is always a balance between performance (less cryptographic computations) and ad selection accuracy (more ads to choose from). Toubiana et. al. performed some tests from which they concluded that $n < 30$ is feasible. Unlike in the current OBA system, the ad selection is performed on a relatively small number of ads. This may make Adnostic's approach less effective. Furthermore, no (estimated) upper bound is given for the value of n. But they suggested *frequency capping* - rescricting the frequency of viewing an ad to the user - as well as using less HTTP-headers for communication with Adnostic, in order to be able to increase the value of n. Nevertheless, Adnostic's scalability will not strongly rely on the exact value of n, and the results by Toubiana et. al. show that many users can use Adnostic at the same time. Furthermore, ad selection may not be as effective as the advertisers and ad networks probably would like to see, but the overall implementation is good. For all users it is as easy as installing an add-on and the ad networks need to create a relatively small extension for their servers. This way Adnostic will quite some points for *implementation* in our rating framework.

3.3.3 Attacks and Flaws

The security of Adnostic's system strongly relies on the trust in several elements. It should have trust:

1. in the TTP not colluding with ad networks, which may be a bit naïf (if collusion occurs, the initial privacy issues remain); rate for *TTP* in the framework will decrease a bit as well as the *attack impact*
2. in the users, not erasing their profile always; rate for *integration* will decrease a bit
3. in the used cryptographic tools, s.t. the ad networks are not able to break the algorithms and so still learn what ads are viewed and still create profiles as they do now; rate for *attacks* and *attack impact* won't decrease when the cryptographic algoritms are chosen carefully
4. in the ad networks, not secretly still using current OBA systems as well as Adnostic at the same time and so still learn what ads are viewed and create profiles as they do now (which means, the initial privacy issues remain), which is very likely to occur; rate for *attacks* as well as for *attack impact* will drastically decrease

3.4 PPOAd

PPOAd is a targeted ad system which aims to be both privacy preserving and secure. It was developed by E. Androulaki and S.M. Bellovin from Colombia University as a protection against profiling by ad entities while also being resistant to click fraud to protect financial incentives.

3.4.1 Requirements

The PPOAd system has two main requirements: *security* and *privacy*. Security is subdivided in the following points:

- Correctness: if all parties are honest, ad networks and publishers will get paid by the advertisers for each of their ads clicked, while preserving privacy.

- Fairness: Parties get paid if they do their duty properly.
- Accountability: Detection and identification of misbehaving parties.
- Unframability: No user can be framed as being responsible for misbehaviour eg click fraud.
- Authentication: Users can only use the system if they are properly authenticated.

Privacy is also subdivided in the following two points:

- Unlinkability: A system entity must not be able to profile a particular honest user.
- Anonymity: A system entity must not be able to trace a particular browsing activity back to a particular user.

These requirements ensure that click fraud cannot occur as well as protecting the privacy of honest users. Additional but less accurately stated requirements are profitibilty and deployability. The former means that the system should at least be as profitable as current OBA systems, and the latter requires that PPOAd systems can be implemented without substantial changes to current systems.

3.4.2 Design

The PPOAd system introduces another actor in the OBA system as discussed in section 2.2. This extension is the User Ad Proxy, or UAP. It is responsible for proxying traffic between the user and each website that is visited by that user. The design consist of three core operations:

1. The user registers by the UAP.

   In the user registration procedure a registration credential *regtick* is issued. This *regtick* authorizes a user U as a member of the PPOAd system multiple times anonymously and unlinkably. Also a wallet with adticks, $W_{adticks}$, is provided. These adticks enables the user to click on ads.

2. The user visits a publishers website.

   The users visits a website via the UAP, sending his ad preferences and demonstrating knowledge of the *regtick* using a zero-knowledge protocol. The UAP contacts the publishers websites and passes on the information from the user, who then gets his preferred ads.

3. The user clicks an ad.

   When the users clicks on an ad, he links one of his adticks to a combination of publisher, ad network and product serial information, which is enough to uniquely identify a particular ad.

3.4.3 Remarks about correctness & implementation

In the registration phase the user needs to provide the UAP his credentials. The user creates a public key pair and provides the UAP with the public part. In the next step both collaborate in a procedure where the *regtick* of the user is created but this should be done in a way the UAP cannot link the *regtick* to the user. However the UAP gains a transcript of the procedure which serve blacklistability purposes. Note that creating the *regtick* blind towards the UAP is not trivial, the paper omits the details on how this is done. In the end the UAP stores the combination of user credentials, public key and regtick transcript in his database.

When a user using the PPOAd system requests as website , he has to authenticate to the UAP. The user can do this by proving possession of his regtick anonymously to the UAP using a zero-knowledge protocol. The UAP learns that the user has a valid regtick but doesn't learn the contents. If the UAP accepts, the user is supplied with a *tick* that allows the user to visit the website requested. This system inherently has a number of drawbacks. If the UAP is down,

the user is unable to surf anonymously with respect to behavioural advertising. If the system becomes popular the load on the UAP becomes huge. The paper does mention it is possible to have a distributed network of UAPs however they would still need to communicate with eachother to exchange which user is authenticated, what users a registered etc.

In the next phase, the user has to set up his ad preferences in the PPOAd software. These are communicated to the UAP which relays them to the ad-networks. These ad-networks are thus unable to link the preferences to the users. The networks in their turn provide the publishers with a list of ads. This is were the targeting takes place. A critical note is that in reality ad networks probably would nog accept this solution. The preferences determined by the user are less accurate and thus less successful as those determined by behavioural advertising. Finally the user supplies the*tick* to the publisher which then shows the web page along with the targeted ads to the user. Note that this connection should happen through an anonymizing network. Otherwise the publisher could match the users ip adress to the supplied ad preferences. This however has severe consequences for usability and deployability requirements. The user will have to set up such an anonymizing network (eg Tor) which generally slows down browsing performance (thus usability decreases). Besides a certain level of technical knowledge is required of the user, although the anonymizing functionality could be included in the PPOAd software (the same that is responsible for gathering users ad preferences).

In case the user clicks an ad, the publisher notifies the advertiser belonging to the ad network that placed the ad. This is also currently done for billing purposes. Together the ad parties construct the ad information, eg. advertisement id, publisher, timestamp and the transcript of the registration of the user by the UAP (both contained in the *tick*). This is then send back to the user to verify. This step ensures *unframability* because it disallows to spoof adclicks that users never did. Upon acceptance, the user will, in collaboration with the ad network, spend an *adtick* (it is bound to the ad information). However if the user has clicked more then a predefined maximal number of times on an (the) ad the ad network will detect this click fraud. In this way the *correctness* and *fairness* security requirements are met. A consequence is that the user can be blacklisted or his anonimity revoked (note that this still doens't mean the identity of the user is made public, just that his public key is revoked from the UAP). In the end, if all is well the user is presented with the advertisers website.

3.5 Attacks and Flaws

3.5.1 Point of attacks

In this scenario, the most likely point of attack is the UAD. In case this PET will be extensively used, a denial of service type of attack against the UAP would effectively disable its working. Likely the load on the UAD is already quite large so this kind of attack would be very feasable. However the gain for an attacker is only possible privacy loss of the users since if they are unable to authenticate the users are also unable to obtain a *tick* and thus visit a publishers site. The solution is to fall back on normal browsing with the known privacy risks.

3.5.2 Effectiveness against behavioural advertising.

Ad brokers can temporarily circumvent the PPOAd system if the user is not careful. Consider a user browsing the web using tor and PPOAd. Now if the user logs in any service that is connected to or controlled by an ad broker (eg google mail) the ad broker could build a profile on that information. Since the same endpoint is used for this session, the ad network is still able to collect browsing information (as long as the session lasts) and combine this with the preference list it gets through the UAP. This however is technically more difficult, the tracking should combine the users login information as well as ip address and unique browser fingerprint.

3.6 The Semi-private PIR Scheme

This scheme is developed by A. Juels in his paper about targeted advertising and privacy. [12]

3.6.1 Design

A threshold PIR scheme allows a consumer c (with $c \in C$ and C is the set of consumers) to request an advertisment ad (with $ad \in AD$ and AD is the set of advertisers) belonging to his profile p (with $p \in P$ and P is the set of profiles) in such a way that the server s ($s \in S$ and S is the set of servers) doesn't learn any information about the request. We assume for simplicity that most of the servers are honest. The El Gamal key pairs (x, y) are appropriately distributed among the servers. Each consumer has an El Gamel key pair (Y_{C_i}, X_{C_i}). Now the scheme can start. The Advertiser chooses a negotiant function $F_{ad} : \mathrm{P} \rightarrow \{1, 2....n\}$. The consumer uses this function on his profile and posts the encrypted result $(E_y[F_{ad}(P(i))], i)$to the bulletin board. When the bulletin board has collected enough encrypted profiles (or some other trigger constraint) the bulletin board applies a mixnetwork to the Set of collected encrypted profiles and sends them to the server. The server replaces the encrypted profile with an encrypted advertisment (For simplicity we assume that an advertisment may be encrypted as a single (El Gamal) ciphertext) and sends it to another mixnetwork. Now the server applies a quorum-controlled asymmetric proxy re-encryption function to obtain $(E_{yc_i}[ad_{F_{ad}(P(i))}], i))$ from $(E_y[F_{ad}(P(i), i)]$. The consumer receives the encrypted pair from the server and can decrypt his advertisment by using his El Gamel key pair.

3.6.2 Remarks about correctness & implementation

The pet relies on a number of assumptions which are not proven. The first is the Decision Diffie-Hellman assumption[7]. A second assumption is that the scheme assumes that most of the servers are honest but doesn't give a measurement when the scheme isn't secure anymore. A thirth assumption is needed for the key-distribution. The paper states that when the pedersen[14] key exchange is used it should be secure. However the paper doesn't proof that the combination of both scheme's or any other scheme is secure. A problem for the implementation would be the use of mixnetworks. Because mixnetworks use computationally intensive (cryptographic) operations. An other problem for the scheme is the bulk encryption. The paper presents a solution for this problem based on mix networks but leave that for the users.

3.6.3 Attack and flaws

A weakness in this scheme is that the Advertiser may construct the negotiant function. For example the advertiser could encode F_{AD} in a way that it decodes the ID of the customer in the output. Then there is a possibility that the Advertiser receives that request from the bulletin board and is able to enemurate information about the consumer.

4 Conclusion

In this paper the dangers posed by behavioural advertising regarding to privacy have been extensively examined. We have shown that the ad networkers and publishers collaborate to profile particular users in order to provide accurate targeted advertising. For this they use advanced tracking techniques which are not easily circumvented by general web users. Also in this paper we discussed several technical solutions that should address the privacy problems of OBA while still retaining pleasant browsing and accurate advertising. This succeeded to a certain level but has several drawbacks concerning usability and scalability. Also, there are some attacks that could endanger the PETs. We conclude with a summary of the discussed PET's and a rating table.

Considering the comments provided in the respective sections, we rated the PETs in the following table:

Table 3: Framework rating where 0 is very bad and 5 is very good

Criteria	*Description*	*Adnostic*	*PPOad*	*RSA-pet*
Integration	Does the PET integrate well within the current system?	**4**	**4**	**3**
Scalability	Is the PET scalable (is processing time and bandwidth acceptable)?	**3**	**2**	**1**
TTPs	Is the PET free of the use of trusted third parties and thus of collusion between ad parties?	**4**	**2**	**0**
Attacks	How likely are attacks to occur?	**1**	**4**	**2**
Attack impact	If attacks succeeds, to what extend is privacy lost?	**0**	**3**	**2**

4.1 Adnostic

Adnostic is an easy to implement PET which takes care of decentralising the user profiles which are currently stored at the ad network servers. It is a brave attempt to keep away all personal information from the ad networks. Unfortunately, it fails in (cryptographically) guaranteeing that the ad networks can not remain using the currently used OBA sytems. Which is, compared to the other flaws in the system (e.g. TTPs and users deleting their profiles), the biggest disadvantage. Therefore, we think Adnostic is a fairly good try, but Toubiana et. al. should also consider colluding parties and those who do not abide the law.

4.2 PPOad

The PPOad PET is a complicated and ingenieus scheme combining a lot of techniques. Beneficial to this is that in a way it is just like the swiss-knife protocol, it is rather complete and satisfies a lot of desirable requirements both on the privacy as on the security (anti click-fraud) level. It does however lend heavily on the usage of a anonymizing network which hinders usability. Also, real clever tracking techniques can bypass the system although this is unlikely and may decrease profiling accuracy.

4.3 The Semi-private PIR Scheme

The PET is currently far away from a working implementation. However, since it enables both the user and the advertiser to reach his goal, this is a nice research area. Combining with research in mixing networks, it could be the solution when the user asks to defend his privacy.

References

[1] Google adsense privacy statement. https://www.google.com/support/adsense/bin/topic.py?hl=nl&topic=146, juni 2011.

[2] A method in determining the browser uniqueness. http://panopticlick.eff.org/browser-uniqueness.pdf, juni 2011.

[3] Tor: Protect your privacy. defend yourself against network surveillance and traffic analysis. https://www.torproject.org/, juni 2011.

[4] The web-based business dictionary. http://www.businessdictionary.com/definition/privacy.html, juni 2011.

[5] Website about the aol scandal in 2002. http://aolscandal.com/, juni 2011.

[6] E. Androulaki and S. M. Bellovin. A secure and privacy-preserving targeted ad-system. 6054 LNCS:123–135, 2010.

[7] Dan Boneh. The decision diffie-hellman problem. *Lecture Notes in Computer Science*, 1423, 1998.

[8] D. Chakrabarti, D. Agarwal, and V. Josifovski. Contextual advertising by combining relevance with click feedback. *Proc. 17th Int'l Conf. on WWW*, pages 417–426, April 2008.

[9] Osborne Clark. International online behavioural advertising survey 2010. 2010.

[10] IAB. Iab internet advertising revenue report. 2011.

[11] J. Sydow M. Jaworska. Behavioural targeting in on-line advertising: An empirical study. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5175 LNCS:62–76, 2008. cited By (since 1996) 1.

[12] A Juels. Targeted advertising ... and privacy too. In *Proceedings of the 2001 Conference on Topics in Cryptology: The Cryptographer's Track at RSA*, CT-RSA 2001, pages 408–424, London, UK, UK, 2001. Springer-Verlag.

[13] A.M. McDonald and L.F Cranor. Beliefs and behaviors:internet users understanding of behavioral advertising. 2010.

[14] T. Pedersen. *Non-interactive and information-theoretic secure verifiable secret sharing.*, volume 91. 1991.

[15] Jessica L. Rich. Ftc staff proposes online behavioral advertising privacy principles. http://www.ftc.gov/opa/2007/12/principles.shtm, December 2007.

[16] Vincent Toubiana, Helen Nissenbaum, Arvind Narayanan, Solon Barocas, and Dan Boneh. Adnostic: Privacy preserving targeted advertising. 2010.