

# Privacy Violations Using Microtargeted Ads: A Case Study

Aleksandra Korolova  
Department of Computer Science  
Stanford University  
korolova@cs.stanford.edu

**Abstract**—In this paper we propose a new class of attacks that exploit advertising systems offering microtargeting capabilities in order to breach user privacy.

We study the advertising system offered by the world’s largest online social network, Facebook, and the risks that the design of the system poses to the privacy of its users. We propose, describe and provide experimental evidence of several novel approaches to exploiting the advertising system in order to obtain private user information.

We communicated our findings to Facebook on July 13, 2010, and received a very prompt response. On July 20, 2010, Facebook launched a change to their advertising system that made the kind of attacks we describe much more difficult but not impossible to implement.

## I. INTRODUCTION

As the number of social network users grows and the social networks expand their feature set to offer users a greater range of the type of data they can share, the concerns related to privacy of this data increase as well. One of the big concerns users have when they share personal information on social networking sites is that the service does not “sell” their personal information to advertisers [1], [2].

Although leading social networks such as Facebook have indeed refrained from selling the information to advertisers, they have established advertising systems that enable personalized social microtargeted advertising. To reconcile the conflicting goals of microtargeted advertising and protecting privacy of users’ personal information, Facebook has implemented an advertising system that provides a layer between individual user data and advertisers. More specifically, the advertising system collects from advertisers the ads they want to display as well as targeting criteria that they’d like the users to satisfy, and then delivers the ads to people who fit those criteria [3]. However, it turns out that building an advertising system that provides an intermediary layer is not sufficient to provide the guarantee of “deliver the ad ... without revealing any personal information to the advertiser” [3], [4], as many of the details of the advertising system’s design influence the privacy guarantees the system can provide, and an advertising system without privacy protections built in by design remains vulnerable to determined and sophisticated attackers.

The solution of an intermediary layer is a common one. As observed by Harper [5] “most websites and ad networks do not “sell” information about their users. In targeted online

advertising, the business model is to sell space to advertisers - giving them access to people (“eyeballs”) based on their demographics and interests. If an ad network sold personal and contact info, it would undercut its advertising business and its own profitability.”

This work proposes and gives experimental examples of several new types of attacks on user private data based on exploiting the microtargeting capabilities of Facebook’s advertising system. These examples contribute to understanding of the ease of implementing such attacks and raise awareness of the many ways that information leakage can happen in microtargeted advertising systems.

**Paper Organization.** In Sections II and III we describe the Facebook experience from user and advertiser perspectives, then introduce the causes of privacy leaks and our experimental evidence for them in Sections IV and V. We discuss our results, their implications, and related work in Sections VI-VIII and conclude in Section IX.

## II. BACKGROUND FROM FACEBOOK USERS’ PERSPECTIVE

### A. Types of user information

When users sign up on Facebook, they are required to provide their first and last name, their email, gender, and date of birth<sup>1</sup>. They are also immediately encouraged to upload a picture and fill out a more detailed set of information about themselves, such as *Basic Information*, consisting of current city, hometown, interested in (women or men), looking for (friendship, dating, a relationship, networking), political and religious views; *Relationships*, consisting of a relationship status (single, in a relationship, engaged, married, it’s complicated, in an open relationship, widowed); *Education and Work* information; *Contact Information*, including address, mobile phone, IM screen name(s), and emails; as well as *Likes and Interests*. The *Likes and Interests* profile section can include things such as favorite activities, music, books, movies, TV, as well as *Pages* corresponding to brands, such as Starbucks or Coca Cola, events such as 2010 Winter Olympics, websites such as TED.com, and diseases, such as AIDS. Any user can Like any Page about any topic.

<sup>1</sup>It is against Facebook’s Statement of Rights and Responsibilities to provide false personal information <http://www.facebook.com/terms.php>

Typically, Facebook users fill out and keep updated daily a variety of information about themselves, thus seamlessly sharing their lives with their friends.

### B. User privacy expectations on Facebook

Facebook enables users to choose their privacy settings which determine who can see which information in their profile. One can distinguish five significant levels of privacy settings specifying the visibility of a particular type of information: *Everyone*, *Friends of Friends*, *Friends Only*, *Hide from specific people* and *Only me*. A very natural set of privacy settings, and one for which there is evidence<sup>2</sup> most people would strive for if they had the ability and patience to navigate Facebook’s ever-changing privacy interface, is to restrict the majority of information to be visible to “Friends only”, with some basic information such as name, location, a profile picture, and a school (or employer) visible to “Everyone” to enable search and distinguishability from people with the same name. In certain circumstances, one might want to hide particular pieces of one’s information from a subset of one’s friends (e.g. sexual orientation information from co-workers, relationship status from parents), as well as keep some of the information visible to “Only me” (e.g. because filling out the information is required by Facebook or to enable receipt of Page’s updates in one’s Newsfeed, without revealing one’s interest in that Page to anyone).

Facebook users have shown time [6] and again [7] that they expect Facebook to not expose their private information without their control [8]. This vocal view of users, privacy advocates, and legislators on Facebook’s privacy changes has recently been acknowledged by Facebook’s CEO [4], resulting in a revamping of Facebook’s privacy setting interface and re-introduction of the options to restrict the visibility of all information, including that of *Likes and Interests*. Users are deeply concerned about controlling their privacy according to a Pew Internet and American Life Project study [9], which shows that more than 65% of social network users say they have changed the privacy settings for their profile to limit what they share with others. Facebook users have been especially concerned with the privacy of their data as it relates to sharing of it with advertisers [2], [1].

## III. BACKGROUND FROM ADVERTISERS’ PERSPECTIVE

### A. Ad creation process and Targeting options

An ad created using Facebook’s self-serve advertising system consists of specifying the ad’s destination URL, Title, Body Text and an optional image.

The unique and valuable proposition [10] that Facebook offers to its advertisers are the **targeting criteria** they are allowed to specify for their ads. As illustrated in Figure 1, the advertiser can specify such targeting parameters as Location

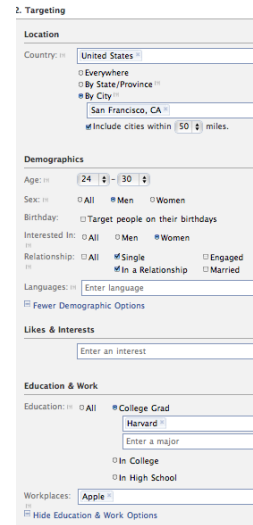


Figure 1. Campaign targeting interface

(including a city), Sex, Age or Age range (including a “Target people on their birthdays” option), Interested In (all, men or women), Relationship status (e.g. single or married), Languages, Likes & Interests, Education (including specifying a particular college, high school, and graduation years), and Workplaces. The targeting criteria can be flexibly combined, e.g. targeting men who live within 50 miles of San Francisco, are male, 24-30 years old, single, interested in women, Like Skiing, have graduated from Harvard and work at Apple. If one chooses multiple options for a single criteria, e.g. both “Single” and “In a Relationship” in Relationship status, then the campaign will target people who are “single **or** in a relationship”. Likewise, specifying multiple interests, e.g. “Skiing”, “Snowboarding”, targets people who like “skiing **or** snowboarding”. Otherwise, unrelated targeting criteria such as age and education are combined using a conjunction, e.g. “exactly between the ages of 24 and 30 inclusive, who graduated from Harvard”. During the process of ad creation, Facebook provides a real-time “Estimated Reach” box, that estimates the number of users who fit the currently entered targeting criteria. The diversity of targeting criteria that enable audience microtargeting down to the slightest detail is an advertisers’ (and, as we will see, a malicious attacker’s) paradise.

The advertiser can also specify the time during which to run the ad, daily budget, and max bid for Pay for Impressions (CPM) or Pay for Clicks (CPC) campaigns.

### B. Matching ads to people

After the ad campaign is created, and every time it is modified, the ad is submitted for approval that aims to verify its adherence to Facebook’s advertising guidelines<sup>3</sup>. From our experience, the approval is occasionally performed

<sup>2</sup>as evidenced by 100,000 people using an open-source privacy scanner *Reclaim Privacy* <http://www.reclaimprivacy.org>

<sup>3</sup>[http://www.facebook.com/ad\\_guidelines.php](http://www.facebook.com/ad_guidelines.php)

manually and other times - automatically, and focuses on checking adherence to guidelines of the ad image and text.

For each user browsing Facebook, the advertising system determines all the ads whose targeting criteria the user matches, and chooses the ads to show based on their bids and relevance.

Facebook provides detailed performance reports specifying the total number of impressions and clicks the ad has received, as well as the number of unique impressions and clicks, broken up by day; as well as rudimentary responder demographics. The performance report data is reported close to real time.

#### IV. ATTACK TYPE CHARACTERIZATION

In this work we illustrate that the promise by several Facebook executives [4], [3], [11], [1] that Facebook "[doesn't] share your personal information with services you don't want", and in particular, "[doesn't] give advertisers access to your personal information" [4], "don't provide the advertiser any ... personal information about the Facebook users who view or even click on the ads" [11] is something that the advertising system strives to achieve but does not fully accomplish. In other words, we show that despite Facebook's advertising system serving as an intermediary layer between user data and advertisers, the design of the system, the matching algorithm, and the user data used to determine the fit to campaign's targeting criteria, combined with the detailed campaign performance reports, contribute to a system that leaks private user information.

We experimentally investigate the workings of the Facebook's advertising system and establish that:

- Facebook uses private and "Friends Only" user information to determine whether the user matches an advertising campaign's targeting criteria
- The default privacy settings lead to most users having a publicly visible unique set of features
- The variety of permissible targeting criteria allows microtargeting an ad to an arbitrary person
- The ad campaign performance reports contain a detailed breakdown of information, including number of unique clicks, respondents' demographic characteristics, and breakdown by time,

which we show leads to an attacker being able to design and successfully run advertising campaigns that:

- A. Infer information that people post on the site in "Only me", "Friends Only", and "Hide from these people" visibility mode
- B. Infer private information not posted on Facebook through ad content and user clicks
- C. Display intrusive and "creepy" ads to individuals

#### V. ATTACK EXPERIMENTS

We now describe in detail our approach and experiments successfully implementing examples of inference attacks

listed in the previous Section<sup>4</sup>.

A. *Infer information that people post on the site in "Only me", "Friends Only", and "Hide from these people" visibility mode.*

Consider the following blueprint for an attack (Algorithm 1 **Inference from Impressions**) aimed at inferring information that a user entered on Facebook but has put into an "Only me" or "Friends Only" visibility mode. According to the privacy settings, it should not be available for observation to anyone except the user herself, or to anyone except the user's friends, respectively. The attack will bypass this restriction by running several advertising campaigns that take advantage of the design and implementation of the Facebook's advertising system.

For ease of notation, we represent each advertising campaign as a mixture of conjunctions and disjunctions of boolean predicates, where campaign  $A = a_1 \wedge (a_2 \vee a_3)$  targets people who satisfy criteria  $a_1$  (e.g. went to Harvard) and criteria  $a_2$  (e.g. "Like skiing") or  $a_3$  (e.g. "Like snowboarding").

---

#### Algorithm 1 Inference from Impressions

---

- 1: **Input:** A user  $U$  and a feature  $F$  whose value from the possible set of values  $\{f_1, \dots, f_k\}$  we'd like to determine, if it is entered by  $U$  on Facebook.
  - 2: Observe the profile information of  $U$  visible to the advertiser that can be used for targeting.
  - 3: Construct an ad campaign with targeting criteria  $A$  combining background knowledge about  $U$  and information visible in  $U$ 's profile, so that one reasonably believes that only  $U$  matches the campaign criteria of  $A$ .
  - 4: Run  $k$  ad campaigns,  $A_1, \dots, A_k$ , such that  $A_i = A \wedge f_i$ . Use identical and innocuous content in the title and text of all the ads. Specify a very high CPC (or CPM) bid, so as to be reasonably sure the ads would win an auction among other ads for which  $U$  is a match.
  - 5: Observe the impressions received by the campaigns over a reasonable time period. If only one of the campaigns, say  $A_j$ , receives impressions, from a unique user, conclude that  $U$  satisfies  $f_j$ . Otherwise, refine campaign targeting criteria, bid, or ad content.
- 

Note that Algorithm 1 **Inference from Impressions** does not require a user  $U$  to pay attention or click on the ad in order for the attack to succeed in inferring the user's private information. The necessary and sufficient conditions for attack's success are: ability to choose  $A$  that identifies the user  $U$  uniquely<sup>5</sup>; Facebook's matching algorithm using

<sup>4</sup>For ethical reasons, all experiments conducted were either: 1) performed with consent of the people we were attacking or aimed at fake accounts; 2) aimed at Facebook employees involved with the advertising system; 3) aimed at inferring information that we do not plan to store, disclose, or use.

<sup>5</sup>We discuss the feasibility of this in Section VI-A.

the information of whether  $U$  satisfies  $f_i$  when determining campaign match; the user  $U$  using Facebook sufficiently often so that the ads have a chance to be displayed to  $U$  at least once over the observation time period, if  $U$  matches the targeting criteria; the advertising system treating campaigns  $A_1, \dots, A_k$  equally.

We run several experiments following this blueprint, and experimentally establish that the advertising system satisfies the above conditions. In particular, we establish that Facebook uses "Friends Only" and "Only me" visible user data when determining whether a user matches an advertising campaign, thereby enabling a malicious attacker posing as an advertiser to infer information that was meant by the user to be kept private or "Friends only", violating user privacy expectations and the company's privacy promises [4], [1], [3], [11].

1) *Inferring a friend's age:* The first experiment shows that using Facebook's advertising system it is possible to infer the age of a particular person, who has set the visibility of that information to be restricted to only her.

We attack a friend of the author, who has entered her birthday on Facebook (because Facebook requires every user to do so) but has specified that she wants it to be private by selecting "Don't show my birthday in my profile" option in the Info section of her profile and by selecting "Make this visible to Only Me" in the Birthday Privacy Settings. Accordingly, she expects that no one should be able to learn her age; however, our experiments demonstrate that it is not the case.

We know the college she went to and where she works, which happens to be a place small enough that she is the only one at her workplace from that college. Following the blueprint of **Inference from Impressions** we created several identical ad campaigns targeting a female at the friend's place of work who went to the friend's college, with the ads differing only in the age of the person being targeted - 33, 34, 35, 36, or 37.

From observing the daily stats of the ad campaigns' performance, we successfully (and correctly) inferred the friend's age - 35, as only the ad targeted to a 35-year-old received impressions. The cost of finding out the private information was a few cents. The background knowledge we utilized related to the friend's education and workplace, is also available in her profile and visible to "Friends Only". Based on prior knowledge, we pruned our exploration to the 33-37 age range, but could have similarly succeeded by running more campaigns, or by first narrowing down the age range by running campaigns aimed at "under 30" and "over 30", then "under 40" and "over 40", then "under 34" and "over 34", etc.

2) *Inferring a non-friend's sexual orientation:* Similarly, following the same blueprint, we succeeded in (correctly) inferring sexual orientation of a non-friend who has posted that she is "Interested in women" in a "Friends Only"

visibility mode. We achieved Step 3 of the blueprint by targeting the campaign to her gender, age, location, and a fairly obscure interest publicly visible to everyone, and used "interested in women" and "interested in men" as the varying values of  $F$ .

3) *Inferring information other than age and sexual orientation:* The private information one can infer using techniques that take advantage of Facebook's fine-grained targeting capabilities and the campaign performance reports is not limited to user age or sexual orientation. Similarly, an attacker posing as an advertiser can also infer a user's relationship status, political and religious affiliation, presence or absence of a particular interest, as well as exact birthday using the "Target people on their birthdays" targeting criteria. This information might be entered into Facebook but set to private, "Friends Only", or visible to friends "Except these people" - the privacy settings which we established that an attacker is able to circumvent. Although using information obtained in such a way is against Facebook's Terms of Service, a determined malicious attacker would not hesitate to disregard it. Moreover, an innocent user would find such an attack surprising, since even if they did not actively click on an ad, by merely adding this information to their profile and logging in to their account they enabled the advertiser to infer private data.

#### *B. Infer private information not posted on Facebook through ad content and user clicks*

While at the core of the just described privacy breach enabling one to infer private data is Facebook's use of private data to determine campaign matches, two other potential privacy breaches are possible due to the microtargeting enabled by Facebook's advertising system.

Suppose one wants to find out whether a colleague is having marital problems, a celebrity is struggling with drug abuse, or whether an employment candidate enjoys gambling or is trying to get pregnant. The blueprint for such an attack (see Algorithm 2 **Inference from Clicks**) would be to target the campaign at the individual of interest, and use the ad content<sup>6</sup> and observation of clicks in order to establish whether the individual is concerned with certain issues.

Any user who clicks on an ad devised according to **Inference from Clicks 2** blueprint reveals that the ad's topic is of interest to him. However, the user is completely unaware what targeting criteria led to this ad being displayed to him, and whether every single user on Facebook or only one or two people are seeing the ad, and thus, the user does not suspect that by clicking the ad, he possibly reveals sensitive information about himself in a way tied to his identity. Finally, when the ad is clicked, the attacker posing as an advertiser can set a cookie with full knowledge of person's identity and interests.

<sup>6</sup>e.g. "Having marital difficulties? Our office offers confidential counseling."

---

**Algorithm 2 Inference from Clicks**

---

- 1: **Input:** A user  $U$  and topic of interest  $T$
  - 2: Observe the profile information of  $U$  visible to the advertiser that can be used for targeting.
  - 3: Construct an ad campaign with targeting criteria  $A$  combining background knowledge about  $U$  and information visible in  $U$ 's profile, so that one reasonably believes that only  $U$  matches the campaign criteria of  $A$ .
  - 4: Run the ad campaign with campaign targeting criteria  $A$  and ad content, picture, and text inquiring about  $T$ , linking to a landing page controlled by an attacker.
  - 5: Observe whether one receives impressions on the ad to establish that the ad is being shown to  $U$ . Make conclusions about  $U$ 's interest in topic  $T$  based on whether there are clicks on the ad.
- 

For ethical reasons, the experiments we successfully ran to confirm the feasibility of such attacks contained ads of more innocuous content: inquiring whether a particular individual is hiring for his team and asking whether a person would like to attend a certain thematic event.

### C. Display intrusive and "creepy" ads to individuals

One can also take advantage of microtargeting capabilities in order to display funny, intrusive, or creepy ads. For example, an ad targeting a particular user  $U$ , could use the user's name in its content, along with phrases ranging from funny, e.g. "Our son is the cutest baby in the world" to disturbing, e.g. "You looked awful at Prom yesterday". For these types of attacks to have the desired effect, one does not need to guarantee Step 3 of Algorithm 2: an intrusive ad may be displayed to a wider audience, but if it uses a user's name, it will only have the desired effect on that user, since others will simply deem it irrelevant after a brief glance.

### D. Other potential inferences

Finally, the information one can infer by using Facebook's advertising system is not limited to the private profile information and information related to the contents of the ads the users click.

For example, one can estimate the frequency of a particular person's Facebook usage, determine whether they have logged in to the site on a particular day, or infer the times of day during which a user tends to browse Facebook. To get a sense of how private this information may be, consider that according to American Academy of Matrimonial Lawyers, 81% of its members have used or faced evidence from Facebook or other social networks in the last five years [12], with 66% citing Facebook as the primary source, including a case when a father sought custody of kids based on evidence that the mother was on Facebook at the time when she was supposed to attend events with her kids [13].

More broadly, going beyond individual user privacy, one can run campaigns in order to infer the age or gender distribution of employees of particular companies, estimate the amount of time employees of particular companies spend on Facebook, the fraction of them who are interested in job opportunities elsewhere, etc. The combination of microtargeting and detailed reporting can give insights into interests and behavioral patterns of certain groups that are interesting from a social science perspective, as well as insights that can be used to manipulate companies and people.

## VI. DISCUSSION OF RESULTS AND THEIR REPLICABILITY

### A. Targeting individuals

The first natural question that arises with regards to the attack blueprints and experiments presented is whether creating an advertising campaign with targeting criteria that are satisfied only by a particular user is feasible for a large fraction of Facebook's users. There is overwhelming experimental and theoretical evidence that it is indeed the case.

As pointed out by [14], 87% of all Americans (or 63% in follow-up work by [15]) can be uniquely identified using zip code, birth date, and gender. Moreover, it is easy to establish [16], [17] that 33 bits of entropy are sufficient in order to identify someone uniquely from the entire world's population. Recent work [18] successfully applies this observation to uniquely identify browsers, based on characteristics such as user agent and timezone that browsers make available to websites. Given the breadth of permissible Facebook ad targeting criteria, it is likely feasible to collect sufficient background knowledge on anyone to identify them uniquely.

The task of selecting targeting criteria matching a person uniquely is in practice further simplified by the default Facebook privacy settings, that make profile information such as gender, hometown, interests and Pages available to everyone. An obscure interest shared by few other people (and there are lots of them), combined with one's location is likely to yield a unique identification, and although the step of selecting these targeting criteria requires some thinking and experimentation, common sense, combined with easily available information on the popularity of each interest or Page on Facebook, easily yields a desired campaign. For users who have changed their default privacy settings to be more restrictive, one can narrow the targeting criteria by investigating their education and work information through other sources. An attacker, such as a stalker, malicious employer, insurance company, journalist or lawyer, is likely to have the resources to obtain the additional background knowledge on their person of interest or may have this information provided to them by the person himself through a resume or application. Friends of a user are particularly powerful in their ability to infer private information about the user, as any information the user posts on Facebook,

facilitates their ability to refine targeting and create campaigns aimed at inferring information kept in the "Only me" visibility mode or inferring private information not posted using **Inference from Clicks**. .

#### B. *Danger of Friends of Friends, Page and Event Admins*

Additional power to successfully design targeting criteria matching particular individuals comes from the following two design choices of Facebook's privacy settings and ad campaign creation interface:

- All profile information except email addresses, IM, phone numbers and exact physical address is by default available to "Friends of Friends".
- The campaign design interface offers options of targeting according to one's *Connections on Facebook*, e.g. targeting users who are/aren't connected to the advertiser's Page, Event, Group, or Application, or targeting users whose friends are connected to a Page, Event, Group, or Application.

While these design choices are aimed at enabling users to share at various levels of granularity and enabling advertisers to take full advantage of social connections and the popularity of their Page(s) and Event(s), they also offer an unprecedented opportunity for breaching privacy through advertising. For example, an attacker may entice a user to Like a Page or RSVP to an event they organize, through prizes and discounts. What a user most likely does not realize when Liking a Page or RSVPing to an event is that he immediately makes himself vulnerable to the attacks of Section V. Furthermore, since the Connections targeting also allows to target friends of users who are connected to a Page, the fact that one's friend Likes a Page, immediately makes one also vulnerable to attacks from the owner of that Page, leading to a potential privacy breach of one's data without any action on one's part.

#### C. *Mitigating uncertainty*

A critic can argue that there is an inherent uncertainty both on the side of Facebook's system design (in the way that Facebook matches ads to people, chooses which ads to display based on bids, does campaign performance reporting) and on the side of user usage of Facebook (e.g. which information and how people choose to enter in their Facebook profile, how often they log in, etc.) that would hinder an attacker's ability to breach user privacy. We offer the following counter-arguments:

**Uncertainty in matching algorithm.** The attacker has the ability to create multiple advertising campaigns as well as to create fake user profiles (see Section VI-D) matching the targeting criteria of those campaigns, in order to reverse-engineer the core aspects of how ads are being matched to users, in what positions they are being displayed, how campaign performance reporting is done, which of the

targeting criteria are the most reliable, etc. In fact, in the course of our experiments, we identified that targeting by city location does not work as expected, and were able to tweak the campaigns to rely on state location information. For our experiments and in order to learn the system, we created and ran more than 30 advertising campaigns at the total cost of less than \$10, without arousing suspicion.

**Uncertainty in user information.** Most users log in to Facebook almost every day, thus enabling a fairly quick feedback loop: if, with a high enough bid, the attacker's campaign is not receiving impressions, this suggests that the targeting criteria require further exploration and tweaking. Hence, although a user might have misspelled or omitted entering information that is known to the attacker through other channels, some amount of experimentation, supplemented with the almost real-time campaign performance reporting, including the number of total and unique impressions and clicks received, is likely to yield a desired campaign.

**Uncertainty in conclusion.** Although attacks may not yield conclusions with absolute certainty, they may provide reasonable evidence. A plausibly sounding headline saying that a particular person is having marital problems or is addicted to pain killers, can cause both embarrassment and harm. The detailed campaign performance reports, including the number of unique clicks and impressions, the ability to run the campaigns over long periods of time, the almost real-time reporting tools, the incredibly low cost of running campaigns, and the lax ad review process, enables a determined attacker to boost his confidence in any of the conclusions.

#### D. *Fake accounts*

As the ability to create fake user accounts on Facebook may be crucial for learning the workings of the advertising system and for more sophisticated attacks (see Section IX), we comment on the ease with which one can create these accounts.

The creation of fake user accounts (although against ToS) that look real on Facebook is not a difficult task, based on our experiments, anecdotal evidence<sup>7</sup>, [19] and others' research [20]. The task can easily be outsourced to Mechanical Turk, as creation of an account merely requires picking a name, email, and filling out a CAPTCHA. By adding a profile picture, some interests and some friends, the fake account quickly becomes hard to distinguish from a real account. What makes the situation even more favorable for an advertising focused attacker, is that typically fake accounts are created with a purpose of sending spam containing links to other users, an observation Facebook relies upon to mark an account as suspicious [21]; whereas the fake accounts created for the purpose of facilitating attacks of Section V would not exhibit such behavior, and would thus be much harder to distinguish from a regular user.

<sup>7</sup><http://rickb.wordpress.com/2010/07/22/why-i-dont-believe-facebooks-500m-users/>

## VII. VIEWS ON MICROTARGETING: UTILITY VS PRIVACY

From the advertisers' perspective, the ability to microtarget users using a diverse set of powerful targeting criteria offers a tremendous new opportunity for audience reach. Specifically on Facebook, over the past year, the biggest advertisers have increased their spending on Facebook advertising more than 10-fold [22] and the "precise enough" audience targeting is what encourages leading brand marketers to spend their advertising budget on Facebook [10]. Furthermore, Facebook itself recommends targeting ads to smaller groups of users<sup>8</sup>, as such ads are "more likely to perform better".

In a broader context, there is evidence that narrowly targeted ads are clicked as much as 670% more than ordinary ones [23], [24] and that very targeted *audience buying* ads, e.g. directed at "women between 18 and 35 who like basketball"<sup>9</sup> are valuable in a search engine setting as well.

The user attitude to microtargeted personalized ads is much more mixed. A user survey by [25] shows that 54% of users don't mind the Facebook ads, while 40% dislike them, with ads linking to other websites and dating sites gathering the least favorable response. Often, users seem perplexed about the reason behind a particular ad being displayed to them, e.g. a woman seeing an ad for Plan B contraceptive may wonder what in her Facebook profile led to Facebook matching her with such an ad and feel that the social network calls her sexual behavior into question [26].

Even more broadly, recent work has identified a gap in privacy boundary expectations between consumers and marketers [27]. According to a Wall Street Journal poll<sup>10</sup>, 72% of respondents feel negatively about targeted advertising based on their web activity and other personal data. A recent study [28] shows that 66% of Americans do not want marketers to tailor advertisements to their interests, and 52% of respondents of another survey claim they would turn off behavioral advertising [23].

Most people understand that in order to receive more personalization they need to give up some of their data [29]. However, they rely on the promises such as those of [3], [4] that personalization is done by the entity they've entrusted their data with, and that only aggregate anonymized information is shared with external entities. However, as our experiments in this work demonstrate, this is not the case, and information that has been explicitly marked by users as private or information that they have not posted on the site but is inferable from the content of the ads they click, leaks in a way tied to their identity through the current design of the most powerful microtargeted advertising system. If peo-

ple were aware of the true privacy cost of ad microtargeting, their views towards it would likely be much more negative.

## VIII. RELATED WORK

The work most closely related to ours is that of Wills and Krishnamurthy [30] and Edelman [31] who have shown that clicking on a Facebook ad, in some cases, revealed to the advertiser the user ID of the person clicking, due to Facebook's failure to properly anonymize the HTTP Referer header. Their work has resulted in much publicity and Facebook has since fixed this vulnerability [32].

The work of [33] proposes a system that enables behavioral advertising without compromising user privacy through a browser extension that runs the targeting algorithm on the user's side. The widespread adoption of their system in the near term seems unlikely.

Several pranks have used Facebook's self-serve advertising system to show an innocuous or funny ad to one's girlfriend<sup>11</sup> or wife<sup>12</sup>. However, they do not perform a systematic study or suggest that the advertising system can be exploited in order to infer private information.

## IX. CONCLUSION AND CONTRIBUTIONS

In this work, we have studied the privacy implications of the world's currently most powerful microtargeted advertising system. We have identified and successfully exploited several design choices of the system that can give rise to private information leakage through running advertising campaigns on the system.

We have notified Facebook of the vulnerabilities in their advertising system on July 13, 2010. Facebook has been very receptive to our feedback and has implemented fixes on July 20, 2010, which will make the kinds of attacks we describe much harder to implement, and thus mitigate the risk to users.

### A. Why Facebook's fix is insufficient

As far as we can tell externally, the approach Facebook took following the disclosure of our work to them is to introduce an additional check in the advertising system, which at campaign creation stage looks at the "Estimated Reach" of the ad created, and suggests to the advertiser to target a larger audience if the "Estimated Reach" does not exceed a soft threshold of about 20 people. While we applaud Facebook's prompt response and efforts in preventing the execution of attacks proposed in this work, their fix is insufficient to fully guarantee user privacy.

To bypass the additional restriction while implementing Algorithm 1 **Inference from Impressions**, it suffices for the attacker to create more than 20 fake accounts (Section VI-D) that match the user being targeted in the known attributes.

<sup>8</sup><http://www.facebook.com/help/?faq=14719>

<sup>9</sup><http://blogs.wsj.com/digits/2010/07/15/live-blogging-google-on-its-earnings>

<sup>10</sup><http://online.wsj.com/community/groups/question-day-229/topics/how-do-you-feel-about-targeted>

<sup>11</sup><http://www.clickz.com/3640069>

<sup>12</sup><http://www.gabrielweinberg.com/blog/2010/05/a-fb-ad-targeted-at-one-person-my-wife.html>

To bypass the restriction while implementing Algorithm 2 **Inference from Clicks**, one can either take a similar approach of creating more than 20 fake accounts, or target the ad to a slightly broader audience than the individual, but further personalize the ad to make it particularly appealing to the individual of interest (e.g. by including the individual's name or location in the ad's text).

Hence, although the minimum campaign reach restriction introduces additional complexity into executing attacks, we believe that the difficulty does not increase significantly enough so as to make the attack infeasible for a determined and resourceful adversary.

### B. Better solutions for Facebook's system

An advertising system that uses only profile information designated as visible to "Everyone" by the user when determining whether the user matches a campaign's targeting criteria would protect users from private data inferences using attacks of type **Inference from Impressions**. Indeed, if private and "Friends Only" information is not used when making the campaign match decisions, then the fact that a user matches a campaign provides no additional knowledge about this user to an attacker.

If Facebook were to commit to using only fully public information in their advertising system, it would likely degrade the quality of the audience microtargeting that they are able to offer advertisers, as well as create an incentive for Facebook to encourage users to share their information more widely in order to improve their advertising (something that Facebook has been accused of but vehemently denies [1]). However, we believe that currently this is the best solution to guarantee users the privacy protections that Facebook promises [4], [3], [11], [1].

We do not know of a solution that would be fully foolproof against **Inference from Clicks** attacks. The *Power Eye* technology [34], [35], that allows consumers to mouse over the ad to get a view of the data that was used to target the ad, offers some hope in providing the user with the understanding of the information they might be revealing when clicking on a particular ad. However, the hassle and understanding of privacy issues required to evaluate the breadth of the targeting and the risk that it poses is beyond the ability of a typical consumer, and thus, the best solution from the perspective of protecting one's privacy is to avoid clicking any of the ads.

An imaginative adversary can come up with other attacks than the ones we described. For example, one can exploit an Auto-complete browser vulnerability [36] in combination with a Facebook advertising campaign focused on a certain topic, in order to cheaply and at scale infer sensitive information about thousands of users. Until a comprehensive solution to protecting privacy in microtargeted advertising systems can be developed, Facebook could mitigate the risk to users through thoughtful design choices regarding:

- Using only public information for ad match determination.
- Setting the default privacy settings for all data as "Friends Only".
- Less detailed campaign performance reports avoiding inclusion of private information even if it is presented in aggregate form.
- Increased financial and logistical barrier for creating ad campaigns.
- Re-thinking of targeting based on Connections to people and Pages (see Section VI-B).
- Evaluation of a campaign as a whole, and not only the content of the ad, during the campaign review process.

It is an open question how to protect the privacy in its broader sense as described in Section V-D, applied in the context of entities rather than individuals.

### C. The broader picture: Microtargeted Advertising and Privacy

The challenges we have investigated in this work, of designing microtargeted advertising systems offering the benefits of fine-grained audience targeting while aiming to preserve user privacy, using the example of Facebook's advertising system, are applicable to a variety of other companies entrusted with user data and administering their own advertising systems, e.g. Google [37]. We have demonstrated that using an intermediary layer that handles the matching between users and ads is not sufficient for being able to provide the privacy guarantees users and companies aspire for, and that a variety of design decisions play a crucial role in the ease of breaching user privacy using the proposed novel class of attacks utilizing advertising campaigns.

We believe that the design of microtargeted ad systems balancing the needs of users, advertisers and web service providers is an important direction for future research, and plan to present our approach towards the private-by-design advertising systems that satisfy rigorous privacy guarantees in follow-up work.

#### ACKNOWLEDGMENTS

The author gratefully acknowledges support by Cisco Systems Stanford Graduate Fellowship and Award IIS-0904325, and thanks Ashish Goel for thoughtful feedback and constructive suggestions.

#### REFERENCES

- [1] B. Schnitt, "Responding to your feedback," The Facebook Blog, <http://blog.facebook.com/blog.php?post=379388037130>, April 5, 2010.
- [2] B. Szoka, "Privacy MythBusters: No, Facebook doesn't give advertisers your data!" <http://techliberation.com/2010/07/06/privacy-mythbusters-no-facebook-doesnt-give-advertisers-your-data>, July 6, 2010 (accessed August 7, 2010).



- [3] S. Sandberg, "The role of advertising on Facebook," *The Facebook blog*, July 6, 2010. [Online]. Available: <http://blog.facebook.com/blog.php?post=403570307130>
- [4] M. Zuckerberg, "From Facebook, answering privacy concerns with new settings," *The Washington Post*, May 24, 2010.
- [5] J. Harper, "It's modern trade: Web users get as much as they give," *The Wall Street Journal*, August 7, 2010.
- [6] D. Kravets, "Judge approves \$9.5 million facebook beacon accord," *The New York Times*, March 17, 2010.
- [7] J. Kincaid, "Senators call out Facebook on instant personalization, other privacy issues," *TechCrunch*, April 27, 2010.
- [8] —, "Live blog: Facebook unveils new privacy controls," *TechCrunch*, May 26, 2010.
- [9] M. Madden and A. Smith, "Reputation management and social media," *Pew Internet and American Life Project*, May 26, 2010.
- [10] N. O'Neill, "Barry Diller: we spend every nickel we can on Facebook," Interview to CNN Money <http://www.allfacebook.com/2010/07/barry-diller-we-spend-every-nickel-we-can-on-facebook>, July 23, 2010 (accessed August 7, 2010).
- [11] E. Schrage, "Facebook executive answers reader questions," *The New York Times* <http://bits.blogs.nytimes.com/2010/05/11/facebook-executive-answers-reader-questions/>, May 11, 2010.
- [12] AAML, "Big surge in social networking evidence says survey of nation's top divorce lawyers," February 10, 2010. [Online]. Available: <http://www.aaml.org/?LinkServID=2F399AE0-E507-CDC7-A1065EE2EE6C4218>
- [13] L. Italie, "Divorce lawyers: Facebook tops in online evidence in court," Associated Press, [http://www.usatoday.com/tech/news/2010-06-29-facebook-divorce\\_N.htm](http://www.usatoday.com/tech/news/2010-06-29-facebook-divorce_N.htm), June 29, 2010.
- [14] L. Sweeney, "Uniqueness of simple demographics in the u.s. population," in *Carnegie Mellon University, School of Computer Science, Data Privacy Lab White Paper Series LIDAP-WP4*, 2000.
- [15] P. Golle, "Revisiting the uniqueness of simple demographics in the us population," in *WPEIS: Proceedings of the 5th ACM workshop on Privacy in electronic society*, 2006, pp. 77–80.
- [16] A. Narayanan, "About 33 bits," <http://33bits.org/about/>, 2008.
- [17] P. Eckersley, "A primer on information theory and privacy," *Electronic Frontier Foundation*, January 27, 2010.
- [18] —, "How unique is your web browser?" in *Privacy Enhancing Technologies*, 2010, pp. 1–18.
- [19] M. Arrington, "Being Eric Schmidt (on Facebook)," *TechCrunch*, October 10, 2010.
- [20] T. Ryan and G. Mauch, "Getting in bed with Robin Sage," *Black Hat USA*, July 28, 2010.
- [21] C. Ghioffi, "Explaining Facebook's spam prevention systems," *The Facebook Blog*, June 29, 2010.
- [22] B. Womack, "Facebook advertisers boost spending 10-fold, COO says," *Bloomberg*, August 3, 2010.
- [23] J. Mullock, S. Groom, and P. Lee, "International online behavioural advertising survey 2010," Osborne Clarke, May 20, 2010.
- [24] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen, "How much can behavioral targeting help online advertising?" in *WWW*, 2009, pp. 261–270.
- [25] S. Su, "User survey results: Which ads do Facebook users like most (and least)?" <http://www.insidefacebook.com/2010/06/15/facebook-users-survey-results-ads>, June 15, 2010.
- [26] B. Stone, "Ads posted on Facebook strike some as off-key," *The New York Times*, March 3, 2010.
- [27] G. R. Milne and S. Bahl, "Are there differences between consumers' and marketers' privacy expectations? a segment-and technology-level analysis," *Journal of Public Policy and Marketing*, vol. 29, no. 1, pp. 138–149, 2010.
- [28] J. Turow, J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy, "Americans reject tailored advertising and three activities that enable it," *Social Science Research Network*, September 29, 2009. [Online]. Available: <http://ssrn.com/abstract=1478214>
- [29] N. Carr, "Tracking is an assault on liberty, with real dangers," *The Wall Street Journal*, August 6, 2010.
- [30] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," in *WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks*, 2009, pp. 7–12.
- [31] B. Edelman, "Facebook leaks usernames, user ids, and personal details to advertisers," <http://www.benedelman.org/news/052010-1.html>, May 20, 2010.
- [32] M. Jones, "Protecting privacy with referrers," Facebook Engineering's Notes <http://www.facebook.com/notes/facebook-engineering/protecting-privacy-with-referrers/392382738919>, May 24, 2010 (accessed August 7, 2010).
- [33] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas, "Adnostic: Privacy preserving targeted advertising," in *17th Annual Network and Distributed System Security Symposium, NDSS*, 2010.
- [34] M. Learmonth, "'Power Eye' lets consumers know why that web ad was sent to them," *Advertising Age* [http://adage.com/digital/article?article\\_id=144557](http://adage.com/digital/article?article_id=144557), June 21, 2010.
- [35] T. Vega, "Ad group unveils plan to improve web privacy," *The New York Times*, October 4, 2010.
- [36] J. Grossman, "Breaking browsers: Hacking auto-complete," <http://jeremiahgrossman.blogspot.com/2010/08/breaking-browsers-hacking-auto-complete.html>, August 2, 2010 (accessed August 7, 2010).
- [37] J. Vascellaro, "Google agonizes on privacy as ad world vaults ahead," *The Wall Street Journal*, August 10, 2010.