

Privacy Violations Using Microtargeted Ads: A Case Study

Aleksandra Korolova*

Abstract. We propose a new class of attacks that breach user privacy by exploiting advertising systems offering microtargeting capabilities. We study the advertising system of the 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.

The work illustrates how a real-world system designed with an intention to protect privacy but without rigorous privacy guarantees can leak private information, and motivates the need for further research on the design of microtargeted advertising systems with provable privacy guarantees. Furthermore, it shows that user privacy may be breached not only as a result of data publishing using improper anonymization techniques, but also as a result of internal data-mining of that data.

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 to implement in practice, even though, as we discuss, they remain possible in principle. We conclude by discussing the broader challenge of designing privacy-preserving microtargeted advertising systems.

Keywords: Facebook, social networks, targeted advertising, privacy breaches

1 Introduction

As more people rely on online social networks to communicate and share information with each other, the social networks expand their feature set to offer users a greater range of the type of data they can share. As a result, more types of data about people is collected and stored by these online services, which leads to increased concerns related to its privacy and re-purposing. One of the big concerns users have when they share personal information on social networking sites is the possibility that their personal information may be sold to advertisers [39, 44].

Although leading social networks such as Facebook have refrained from selling the information to advertisers, they have created systems that enable advertisers to run highly targeted social advertising campaigns. Not surprisingly, the goals of enabling highly targeted advertising and protecting the privacy of users' personal information entrusted to the company are at odds. To reconcile these conflicting goals, Facebook has

*Department of Computer Science, Stanford University, <mailto:korolova@cs.stanford.edu>

designed an advertising system which provides a separation layer between individual user data and advertisers. Concretely, Facebook collects from advertisers the ad creatives to display and the targeting criteria which the users being shown the ad should satisfy, and delivers the ads to people who fit those criteria [38]. Through experiments, in this paper we demonstrate that an advertising system providing an intermediary layer between users and advertisers is not sufficient to provide the guarantee of “deliver the ad ... without revealing any personal information to the advertiser” [38, 53], 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 is vulnerable to determined and sophisticated attackers.

Building an advertising system that serves as an intermediary layer between user data and advertisers is a common approach to user data monetization. As observed by Harper [18], “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 evidence of feasibility of several new types of attacks for inferring private user information by exploiting the microtargeting capabilities of Facebook’s advertising system. The crux of the attacks consists of crafting advertising campaigns targeted to individuals whose privacy one aims to breach and using the ad campaign performance reports to infer new information about them. The first type of attack, **Inference from Impressions**, enables an attacker posing as an advertiser to infer a variety of private information about a user from the fact that the attacker matched the campaign targeting criteria. The second type of attack, **Inference from Clicks**, enables inferences from the fact that a user takes action, such as a click, based on the content of the ad. This work also contributes to understanding of the ease of implementation of proposed attacks and raises awareness of the many ways that information leakage can happen in microtargeted advertising systems. It provides an example of a real-world system in which internal data mining of users’ private data entrusted to the company can lead to privacy breaches.

Paper Organization. In Sections 2 and 3 we describe the Facebook experience from user and advertiser perspectives. We introduce the underlying reasons for privacy leaks, attack blueprints, and present our experimental evidence of their success in Section 4. We discuss our results, their implications, and related work in Sections 5-7. We conclude in Section 8 with a discussion of Facebook’s response to our research disclosure and a discussion of the challenges of designing provably private microtargeted advertising systems.

2 Facebook from the users’ perspective

In this section we describe the types of information that users can include in their Facebook profiles and the privacy controls available to them.

2.1 User profile information

When users sign up on Facebook, they are required to provide their real first and last name, email, gender, and date of birth.¹ 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 the 2010 Winter Olympics, websites such as TED.com, and diseases such as AIDS. Any user can *Like* any Page about any topic. Since the launch of Facebook's Open Graph API [45], users are able to Like many entities on the web, such as webpages, blog posts, products, and news articles. Users can also post status updates, pictures, and videos; ask questions; and share links through Facebook, potentially enabling Facebook to learn further details about their interests through data mining of these pieces of content.

Many Facebook users complete and actively update [15] this variety of information about themselves, thus seamlessly sharing their interests, current activities, thoughts, and pictures with their friends.

2.2 User privacy

Facebook provides the ability to limit who can see the information a user shares on Facebook through a privacy setting specific to each category of information. 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 many users would strive for if they had the technical sophistication and patience to navigate Facebook's ever-changing privacy interface,² 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., date of birth, which is required by Facebook or interest in a certain Page, in order to receive that Page's updates in one's Newsfeed, without revealing one's interest in that Page to anyone).

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

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

Facebook users have shown time [24] and again [22] that they expect Facebook to not expose their private information without their control [21]. This vocal view of users, privacy advocates, and legislators on Facebook’s privacy changes has recently been acknowledged by Facebook’s CEO [53], resulting in a revamping of Facebook’s privacy setting interface and a 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 [29], 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 the sharing of it with advertisers [44, 39].

3 Facebook from the advertisers’ perspective

In this section, we describe the design of Facebook’s advertising system at the time this research was performed (spring and summer of 2010).

3.1 Ad creation process and Targeting options

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

The unique and valuable proposition [36] 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 (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 advertiser’s (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.

2. Targeting

Location
Country:

- ☐ Everywhere
- ☐ By State/Province
- ☒ By City
 - ☒ Include cities within miles.

Demographics
Age: -
Sex: ☒ All ☐ Men ☐ Women
Birthday: ☐ Target people on their birthdays
Interested In: ☒ All ☐ Men ☐ Women
Relationship: ☐ All ☒ Single ☐ Engaged
☒ In a Relationship ☐ Married
Languages:
☐ Fewer Demographic Options

Likes & Interests

Suggested Likes & Interests

- ☐ Wakeboarding
- ☐ Wake Boarding
- ☐ Jet Skiing
- ☐ Kneeboarding
- ☐ Snowshoeing
- ☐ Snowboard

Education & Work
Education: ☐ All ☒ College Grad

- ☐ In College
- ☐ In High School

Workplaces:
☐ Hide Education & Work Options

Figure 1: Campaign targeting interface

3.2 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.³ Based on our experiments, it seems that the approval is occasionally performed 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 ad campaign performance reports specifying the total number of impressions and clicks the ad has received, 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.

4 The Attacks

We illustrate that the promise by several Facebook executives [53, 38, 40, 39] 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” [53], “don’t provide the advertiser any ... personal information about the Facebook users who view or even click on the ads” [40] is something that the advertising system strives to achieve but does not fully accomplish. 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 the 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 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 many users having a publicly visible uniquely identifying 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,

³http://www.facebook.com/ad_guidelines.php

which we show leads to an attacker posing as an advertiser being able to design and successfully run advertising campaigns that enable them to:

1. Infer information that people post on Facebook in “Only me”, “Friends Only”, and “Hide from these people” visibility mode
2. Infer private information not posted on Facebook through ad content and user response
3. Display intrusive and “creepy” ads to individuals

We now describe in detail two novel attacks that exploit the details of the advertising system’s design in order to infer private information and our experiments implementing them.⁴

4.1 Infer information posted on Facebook with “Only me”, “Friends Only”, and “Hide from these people” privacy settings through ad campaign match

Attack 1: **Inference from Impressions** is aimed at inferring information that a user has entered on Facebook but has restricted to be visible to “Only me” or “Friends Only.” According to the privacy settings chosen by the user, this information should not be available for observation to anyone except the user themselves, or to anyone except the user’s friends, respectively. The proposed attack will bypass this restriction by running several advertising campaigns targeted at the user and differing only in the targeting criteria corresponding to the unknown private information the attacker is trying to infer. The difference in campaign performance reports of these campaigns will enable the attacker to infer desired private information.

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”).

The necessary and sufficient conditions for the attack’s success are: the ability to choose targeting criteria A that identify the user U uniquely;⁵ Facebook’s user-ad matching algorithm showing the ad only to users who match the ad targeting criteria exactly and using 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.

⁴For 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.

⁵We discuss the feasibility of this in Section 5.1.

Attack 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 CPM (or CPC) 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.
-

We run several experiments following the blueprint of Attack 1, 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 [53, 39, 38, 40].

We also remark that a typical user would find Attack 1 **Inference from Impressions** very surprising, as the advertiser is able to gain knowledge about things the user might have listed in their profile even if the user U does not pay attention to or click on the ad.

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 information to only be visible by themselves.

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 Information 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. The ads whose age target does not match the friend’s age will not be displayed, and the ad that matches her age will be, as long as the ad creative is reasonably relevant and the friend uses Facebook during the ad campaign period.

From observing the daily stats of the ad campaigns’ performance, particularly, the number of impressions each of the ads has received, we 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.

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 .

Inferring information other than age and sexual orientation

The private information one can infer using techniques that exploit the microtargeting capabilities of Facebook’s advertising system, its ad-user matching algorithm, and the ad campaign performance reports, is not limited to user age or sexual orientation. 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 criterion.

Although using private user information obtained through ad campaigns is against Facebook’s Terms of Service, a determined malicious attacker would not hesitate to disregard it.

4.2 Infer private information not posted on Facebook through microtargeted ad creative and user response to it

The root cause of privacy breaches possible using Attack 1: **Inference from Impressions** is Facebook’s use of private data to determine whether the users match targeting criteria specified by the ad campaign. We now present a different kind of attack, Attack 2: **Inference from Clicks**, that takes advantage of the microtargeting capabilities of the system and the ability to observe a user’s response to the ad in order to breach

privacy. The goal of this attack is to infer information about users that may not have been posted on Facebook, such as a particular user’s interest in a certain topic. The attack proceeds by creating an ad enticing a user U interested in topic T to click on it, microtargeting the ad to U , and using U ’s reaction to the ad (e.g., a click on it) as an indicator of U ’s interest in the topic.

Suppose an attacker 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. Attack 2: **Inference from Clicks** targets the campaign at the individual of interest, designs the ad creative that would engage an individual interested in the issue (e.g., “Having marital difficulties? Our office offers confidential counseling.”), and observes whether the individual clicks on the ad to infer the individual’s interest in the issue.

Attack 2 Inference from Clicks

- 1: **Input:** A user U and a topic of interest T .
 - 2: Observe the profile information of U visible to the advertiser that can be used for targeting.
 - 3: Construct 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 criteria of A .
 - 4: Run an ad campaign with targeting criteria A and ad content, picture, and text inquiring about T , linking to a landing page controlled by an attacker.
 - 5: Observe whether the ad receives impressions to ensure that it is being shown to U . Make conclusions about U ’s interest in topic T based on whether the ad receives clicks.
-

Any user who clicks on an ad devised according to the blueprint of **Inference from Clicks** reveals that the ad’s topic is likely of interest to him. However, the user does not suspect that by clicking the ad, he possibly reveals sensitive information about himself in a way tied to his identity, as he 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.

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.

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 the

success of Step 3 of Attack 2—an intrusive ad may be displayed to a wider audience, but if it uses a particular user’s name, it will likely only have the desired effect on that user, since others will likely deem it irrelevant after a brief glance.

4.3 Other possible inferences

The information one can infer by using Facebook’s advertising system is not limited to the private profile information and information inferred from the contents of the ads the users click.

Using the microtargeting capability, 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 or become in the future, 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 [1], 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 [19].

More broadly, going beyond individual user privacy, one can imagine running ad campaigns in order to infer organization-wide trends, such as the age or gender distribution of employees of particular companies, the amount of time they spend on Facebook, the fraction of them who are interested in job opportunities elsewhere, etc. For example, a data-mining startup Rapleaf has recently used [50] their database of personal data meticulously collected over several years, to compare shopping habits and family status of Microsoft and Google employees. Exploitation of powerful targeting capabilities and detailed campaign performance reports of Facebook’s advertising system could potentially facilitate a low-cost efficient alternative to traditional investigative analysis. Insights into interests and behavioral patterns of certain groups could be valuable from the social science perspective, but could also have possibly undesired implications, if exploited, for example, by insurance companies negotiating contracts with small companies, stock brokers trying to gauge future company performance, and others trying to exploit additional information obtained through ad campaigns to their advantage.

5 Discussion of Attacks and their Replicability

In this section, we discuss the feasibility of selecting campaign features for targeting particular individuals, the additional privacy risks posed by “Connections targeting” capabilities of the Facebook advertising system, the ways in which an attacker can increase confidence in conclusions obtained through the attacks exploiting the advertising system, and the feasibility of creating fake user accounts.

5.1 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 practically feasible. There is strong experimental and theoretical evidence that it is indeed the case.

As pointed out by [43], 87% of all Americans (or 63% in follow-up work by [14]) can be uniquely identified using zip code, birth date, and gender. Moreover, it is easy to establish [33, 10] that 33 bits of entropy are sufficient in order to identify someone uniquely from the entire world’s population. Recent work [9] successfully applies this observation to uniquely identify browsers based on characteristics such as user agent and timezone information that browsers make available to websites. Although we did not perform a rigorous study, we conjecture that 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 liked visible to everyone. An obscure interest shared by few other people, 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 enables the creation of 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 all information the user posts in “Friends Only” privacy mode facilitates their ability to refine targeting and create campaigns aimed at inferring information kept in the “Only me” mode or inferring private information not posted using **Inference from Clicks**.

5.2 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 adver-

tiser’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 facilitate the opportunity for a breach of 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 is that by Liking a Page or RSVPing to an event he makes himself more vulnerable to the attacks of Section 4. Furthermore, since the Connections targeting also allows to target friends of users who are connected to a Page, if one’s friend Likes a Page, it also makes one 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.

5.3 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, and does campaign performance reporting) and on the side of user usage of Facebook (e.g., which information and how people choose to enter it 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 5.4) 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. For example, in the course of our experiments, we identified that targeting by city location did 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. Half of Facebook’s users log in to Facebook every day [3], 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 plausible 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.

5.4 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, we comment on the ease with which one can create these accounts.

The creation of fake user accounts (although against the Terms of Service) that look real on Facebook is not a difficult task, based on our experiments, anecdotal evidence,⁶ [2] and others' research [37]. The task can 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 to the fake account, it 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 [13]; whereas the fake accounts created for the purpose of facilitating attacks of Section 4 would not exhibit such behavior, and would thus, presumably, be much harder to distinguish from a regular user.

6 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 more than 10-fold [51] and the "precise enough" audience targeting is what encourages leading brand marketers to spend their advertising budget on Facebook [36]. Furthermore, Facebook itself recommends targeting ads to "smaller, more specific" groups of users,⁷ as such ads are "more likely to perform better".

In a broader context, there is evidence that narrowly targeted ads are much more effective than ordinary ones [32, 52] and that very targeted *audience buying* ads, e.g., directed at "women between 18 and 35 who like basketball"⁸ are valuable in a search engine ad setting as well.

The user attitude to microtargeted personalized ads is much more mixed. A user survey by [42] shows that 54% of users don't mind the Facebook ads, while 40% dislike

⁶<http://rickb.wordpress.com/2010/07/22/why-i-dont-believe-facebooks-500m-users/>

⁷<http://www.facebook.com/help/?faq=14719>

⁸<http://blogs.wsj.com/digits/2010/07/15/live-blogging-google-on-its-earnings>

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 a 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 [41].

Even more broadly, recent work has identified a gap in privacy boundary expectations between consumers and marketers [31]. According to a Wall Street Journal poll,⁹ 72% of respondents feel negatively about targeted advertising based on their web activity and other personal data. A recent study [47] 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 [32].

Many people understand that in order to receive more personalization they need to give up some of their data [6]. However, they rely on the promises such as those of [38, 53] 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 people were aware of the true privacy cost of ad microtargeting, their views towards it would possibly be much more negative.

7 Related Work

The work most closely related to ours is that of Wills and Krishnamurthy [25] and Edelman [11] 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 [20].

The work of [16] observes that ads whose ad creative is neutral to sexual preference may be targeted exclusively to gay men, which could create a situation where a user clicking on the ad would reveal to the advertiser his sexual preference.

Several pranks have used Facebook’s self-serve advertising system to show an innocuous or funny ad to one’s girlfriend¹⁰ or wife¹¹. However, they do not perform a systematic study or suggest that the advertising system can be exploited in order to infer private information.

The works of [46] and [17] propose systems that perform profiling, ad selection and

⁹<http://online.wsj.com/community/groups/question-day-229/topics/how-do-you-feel-about-targeted>

¹⁰<http://www.clickz.com/3640069>

¹¹<http://www.gabrielweinberg.com/blog/2010/05/a-fb-ad-targeted-at-one-person-my-wife.html>

targeting on the client’s (user’s) side and use cryptographic techniques to ensure accurate accounting. These proposals require a shift in the paradigm of online advertising, where the ad brokers relinquish the control of the way profiling and matching is performed and rely on a weaker client-side model of the user, which seems unlikely in the near-term.

8 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 enable new kinds of attacks and inferences of user private data through advertising campaigns.

8.1 Facebook’s response and other possible solutions

Following the disclosure of our findings to Facebook on July 13, 2010, Facebook promptly implemented changes to their advertising system that make the kinds of attacks we describe much harder to execute.

Their approach was to introduce an additional check in the advertising system, which at the campaign creation stage looks at the “Estimated Reach” of the ad created, and suggests to the advertiser to target a broader audience if the “Estimated Reach” does not exceed a soft threshold of about 20 people. We applaud Facebook’s prompt response and efforts in preventing the execution of attacks proposed in this work, but believe that their fix does not fully eliminate the possibility of proposed attacks.

Although we did not perform further experiments, it is conceivable that the additional restriction of sufficiently high “Estimated Reach” can be bypassed in principle for both types of attacks proposed. To bypass the restriction while implementing Attack 1: **Inference from Impressions**, it suffices for the attacker to create more than 20 fake accounts (Section 5.4) that match the user being targeted in the known attributes. A sufficient number of accounts matching the targeting criteria in the system would permit running the ad, and attacker’s control over the fake accounts would enable differentiating between the impressions and clicks of targeted individual and fake accounts. To bypass the restriction while implementing Attack 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 introduced additional complexity into reliably executing attacks, the restriction does not seem to make the attacks infeasible for determined and resourceful adversaries.

A better solution to protect users from private data inferences using attacks of type 1: **Inference from Impressions** would be to use only profile information designated

as visible to “Everyone” by the user when determining whether a user matches a campaign’s targeting criteria. 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 beyond what they could infer by simply looking up their public profile on Facebook.

Although using only fully public information in the advertising system would come closest to delivering on the privacy promises made by Facebook to its users [53, 38, 40, 39], it would also introduce a business challenge for Facebook. As much of the information users share is “Friends Only”, using only information shared with “Everyone” would likely degrade the quality of the audience microtargeting that Facebook is able to offer advertisers, and hence create a business incentive to encourage users to share their information more widely in order to monetize better (something that Facebook has been accused of but vehemently denies [39]). Another approach would be to introduce an additional set of privacy controls to indicate which information the users are comfortable sharing with advertisers; however, this would create significant additional cognitive burden on users navigating an already very complex set of privacy controls [12].

We do not know of a solution that would be fully foolproof against **Inference from Clicks** attacks. The *Power Eye* concept [27, 49], providing consumers with a view of the data used to target the ad upon a mouseover, 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 likely 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.

As discussed in Section 5, there are several other aspects of Facebook’s advertising system that make it particularly vulnerable to attacks aiming to infer individual’s private information. Mitigating the risks to users could be accomplished through thoughtful design choices regarding:

- Careful choice of user information used for ad match determination.
- Default privacy settings, perhaps setting them as “Friends Only” for all data.
- Less detailed campaign performance reports, avoiding inclusion of private information even if it is presented in aggregate form.
- Increased financial and logistical barriers for creating ad campaigns.
- Re-thinking of targeting based on Connections to people and Pages (see Section 5.2).
- Evaluation of an ad 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 privacy in its broader sense as described in Section 4.3, applied in the context of entities rather than individuals.

8.2 The Broader Challenge: Enabling Microtargeted Advertising while Preserving 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, will become applicable to a variety of other companies entrusted with user data and administering their own advertising systems (e.g., Google) as they move to enable better targeting [48]. We have demonstrated that merely 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 seemingly minor design decisions play a crucial role in the ease of breaching user privacy using the proposed novel class of attacks.

As microtargeted advertising is becoming increasingly important for the online economy, finding ways to design microtargeted advertising systems that balance the privacy needs of users and the business and utility needs of advertisers and web service providers is of essence. This work, as well as a variety of others [5, 4, 26, 35, 34], have demonstrated that ad-hoc approaches to protecting privacy inevitably fail in the world of creative and sophisticated adversaries in possession of auxiliary information from multiple sources. Therefore, we believe that systems satisfying rigorous guarantees of privacy, such as *differential privacy* (see [7]), would provide the most robust starting point.

The existing work on algorithms for guaranteeing differential privacy does not seem to immediately apply to this context because of the following unique real-world characteristics of the problem domain:

- The ability of an attacker to run multiple campaigns, repeatedly, and over an extended period of time.
- The difficulty of analyzing whether two campaigns are identical or the extent of their overlap.
- The business-need to perform accurate reporting and billing based on the number of impressions and clicks received by an ad.
- The ability of an attacker to create users with characteristics he controls (e.g., through creation of fake online profiles or injection of search queries).
- The business-need to deliver the ad to users exactly matching the campaign criteria.
- The ever-increasing number of features and their combinations available for targeting, and the correlations between them.

- The constantly evolving and very detailed user profiles and interests (based on their social network profiles, or email and search activity).

For example, consider taking a commonly used approach for achieving differential privacy—adding properly calibrated random noise [8], in this case, to the true number of unique impressions and clicks received by a particular ad [28], before reporting these statistics to the advertiser. From the business logic perspective, advertisers might be hesitant to agree to pay according to the noisy, rather than true, utility received from their ad campaigns. From the privacy perspective, they might have other ways of inferring the true number of clicks received, e.g., through an analytics tool on their ad’s landing page. Furthermore, differential privacy is typically achieved by adding symmetrically distributed random noise. If fresh noise is added for every campaign and every time campaign performance is reported, then an attacker can make accurate inferences by running a sufficiently large number of identical campaigns for a sufficiently long time period and averaging the reported statistics. If noise is randomly generated only once per campaign, and reused subsequently, then the web service needs to reliably identify which campaigns are identical and store the noise; furthermore, the statistics will always be either under- or over- reported. The system would also need to ensure private information could not be inferred by running ad campaigns targeting overlapping sets of users or using overlapping sets of features.

Another conceptual approach for achieving privacy would be akin to performing a broad match—showing the ad not only to users who match the campaign targeting criteria perfectly, but also to users who match them less well (e.g., in proportion to some quality of the match score using the exponential mechanism of [30]). The approach of modifying each campaign into a broad match campaign would significantly undermine the utility potential of microtargeted advertising to reach a highly targeted audience for whom the ad is most relevant. It would require advertiser buy-in, an ability to design a meaningful match score for a variety of user profile features and ad campaigns, as well as a payment scheme that charges advertisers differently and fairly depending on the quality of the match without jeopardizing privacy.

Thus, understanding the space of possible solutions for designing microtargeted advertising systems and quantifying the trade-offs in user privacy, and web services’ and advertisers’ utility for each of them, is an important and rich direction for future work.

Acknowledgments

The author gratefully acknowledges support by Cisco Systems Stanford Graduate Fellowship and NSF Award IIS-0904325, and thanks Ashish Goel for thoughtful feedback and constructive suggestions. This article was presented in an earlier version at the 2010 IEEE International Conference on Data Mining Workshops, December 13, 2010, in Sydney, Australia.[23]

References

- [1] AAML (2010). Big surge in social networking evidence says survey of nation's top divorce lawyers. <http://www.aaml.org/?LinkServID=2F399AE0-E507-CDC7-A1065EE2EE6C4218>.
- [2] Arrington, M. (2010). Being Eric Schmidt (on Facebook). *TechCrunch*.
- [3] — (2010). Facebook COO: 175 million people log into Facebook every day. *TechCrunch*.
- [4] Backstrom, L., Dwork, C., and Kleinberg, J. M. (2007). Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. In *WWW*, 181–190.
- [5] Barbaro, M. and Zeller, T., Jr. (2006). A face is exposed for AOL searcher No. 4417749. *The New York Times*.
- [6] Carr, N. (2010). Tracking is an assault on liberty, with real dangers. *The Wall Street Journal*.
- [7] Dwork, C. (2011). A firm foundation for private data analysis. *Commun. ACM*, 54(1): 86–95.
- [8] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *TCC*, 265–284.
- [9] Eckersley, P. (2010). How unique is your web browser? In *Privacy Enhancing Technologies*, 1–18.
- [10] — (2010). A primer on information theory and privacy. *Electronic Frontier Foundation*.
- [11] Edelman, B. (2010). Facebook leaks usernames, user ids, and personal details to advertisers. <http://www.benedelman.org/news/052010-1.html>.
- [12] Gates, G. (2010). Facebook privacy: A bewildering tangle of options. *The New York Times*. <http://www.nytimes.com/interactive/2010/05/12/business/facebook-privacy.html>.
- [13] Ghiossi, C. (2010). Explaining Facebook's spam prevention systems. *The Facebook Blog*.
- [14] Golle, P. (2006). Revisiting the uniqueness of simple demographics in the us population. In *WPES: Proceedings of the 5th ACM workshop on Privacy in electronic society*, 77–80.
- [15] Grove, J. V. (2009). Just married: Groom changes Facebook relationship status at the altar. <http://mashable.com/2009/12/01/groom-facebook-update>.

- [16] Guha, S., Cheng, B., and Francis, P. (2010). Challenges in Measuring Online Advertising Systems. In *Proceedings of the 2010 Internet Measurement Conference (IMC)*. Melbourne, Australia.
- [17] — (2011). Privad: Practical Privacy in Online Advertising. In *Proceedings of the 8th Symposium on Networked Systems Design and Implementation (NSDI)*. Boston, MA.
- [18] Harper, J. (2010). It’s modern trade: Web users get as much as they give. *The Wall Street Journal*.
- [19] Italie, L. (2010). Divorce lawyers: Facebook tops in online evidence in court. Associated Press, http://www.usatoday.com/tech/news/2010-06-29-facebook-divorce_N.htm.
- [20] Jones, M. (2010). Protecting privacy with referrers. Facebook Engineering’s Notes <http://www.facebook.com/notes/facebook-engineering/protecting-privacy-with-referrers/392382738919>.
- [21] Kincaid, J. (2010). Live blog: Facebook unveils new privacy controls. *TechCrunch*.
- [22] — (2010). Senators call out Facebook on instant personalization, other privacy issues. *TechCrunch*.
- [23] Korolova, A. (2010). Privacy violations using microtargeted ads: A case study. In *2010 IEEE International Conference on Data Mining Workshops*, 474–482. Sydney, Australia.
- [24] Kravets, D. (2010). Judge approves \$9.5 million Facebook ‘Beacon’ accord. *The New York Times*.
- [25] Krishnamurthy, B. and Wills, C. E. (2009). On the leakage of personally identifiable information via online social networks. In *WOSN ’09: Proceedings of the 2nd ACM Workshop on Online Social Networks*, 7–12.
- [26] Kumar, R., Novak, J., Pang, B., and Tomkins, A. (2007). On anonymizing query logs via token-based hashing. In *WWW*, 629–638.
- [27] Learmonth, M. (2010). ‘Power Eye’ lets consumers know why that web ad was sent to them. Advertising Age http://adage.com/digital/article?article_id=144557.
- [28] Lindell, Y. and Omri, E. (2011). Personal communication.
- [29] Madden, M. and Smith, A. (2010). Reputation management and social media. *Pew Internet and American Life Project*.
- [30] McSherry, F. and Talwar, K. (2007). Mechanism design via differential privacy. In *FOCS*, 94–103.

- [31] Milne, G. R. and Bahl, S. (2010). Are there differences between consumers' and marketers' privacy expectations' a segment- and technology-level analysis. *Journal of Public Policy and Marketing*, 29(1): 138–149.
- [32] Mullock, J., Groom, S., and Lee, P. (2010). International online behavioural advertising survey 2010. Osborne Clarke.
- [33] Narayanan, A. (2008). About 33 bits. <http://33bits.org/about/>.
- [34] Narayanan, A., Shi, E., and Rubinstein, B. I. P. (2011). Link prediction by de-anonymization: How we won the Kaggle social network challenge. *CoRR*, abs/1102.4374.
- [35] Narayanan, A. and Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy*, 111–125.
- [36] O'Neill, N. (2010). 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>.
- [37] Ryan, T. and Mauch, G. (2010). Getting in bed with Robin Sage. Black Hat USA.
- [38] Sandberg, S. (2010). The role of advertising on Facebook. *The Facebook blog*. URL <http://blog.facebook.com/blog.php?post=403570307130>
- [39] Schnitt, B. (2010). Responding to your feedback. The Facebook Blog, <http://blog.facebook.com/blog.php?post=379388037130>.
- [40] Schrage, E. (2010). Facebook executive answers reader questions. The New York Times <http://bits.blogs.nytimes.com/2010/05/11/facebook-executive-answers-reader-questions/>.
- [41] Stone, B. (2010). Ads posted on Facebook strike some as off-key. *The New York Times*.
- [42] Su, S. (2010). User survey results: Which ads do Facebook users like most (and least)? <http://www.insidefacebook.com/2010/06/15/facebook-users-survey-results-ads>.
- [43] Sweeney, L. (2000). 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*.
- [44] Szoka, B. (2010). 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>.
- [45] Taylor, B. (2010). The next evolution of Facebook platform. <http://developers.facebook.com/blog/post/377>.

- [46] Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H., and Barocas, S. (2010). Adnostic: Privacy preserving targeted advertising. In *17th Annual Network and Distributed System Security Symposium, NDSS*.
- [47] Turow, J., King, J., Hoofnagle, C. J., Bleakley, A., and Hennessy, M. (2009). Americans reject tailored advertising and three activities that enable it. *Social Science Research Network*. <http://ssrn.com/abstract=1478214>.
- [48] Vascellaro, J. (2010). Google agonizes on privacy as ad world vaults ahead. *The Wall Street Journal*.
- [49] Vega, T. (2010). Ad group unveils plan to improve web privacy. *The New York Times*.
- [50] Wauters, R. (2011). Googlers buy more junk food than Microsofties (and why Rampleaf is creepy). *TechCrunch*.
- [51] Womack, B. (2010). Facebook advertisers boost spending 10-fold, COO says. *Bloomberg*.
- [52] Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., and Chen, Z. (2009). How much can behavioral targeting help online advertising? In *WWW*, 261–270.
- [53] Zuckerberg, M. (2010). From Facebook, answering privacy concerns with new settings. *The Washington Post*.

