The Price of Free: Privacy Leakage in Personalized Mobile In-App Ads

Wei Meng, Ren Ding, Simon P. Chung, Steven Han, and Wenke Lee
College of Computing
Georgia Institute of Technology
{wei, rding, pchung34, s.han}@gatech.edu, wenke@cc.gatech.edu

***Abstract*—In-app advertising is an essential part of the ecosystem of free mobile applications. On the surface, this creates a win-win situation where app developers can profit from their work without charging the users. However, as in the case of web advertising, ad-networks behind in-app advertising employ personalization to improve the effectiveness/profitability of their ad-placement. This need for serving personalized advertisements in turn motivates ad-networks to collect data about users and profile them. As such, "free" apps are only free in monetary terms; they come with the price of potential privacy concerns. The question is, how much data are users giving away to pay for "free apps"?**

In this paper, we study how much of the user's interest and demographic information is known to these major ad networks on the mobile platform. We also study whether personalized ads can be used by the hosting apps to reconstruct some of the user information collected by the ad network. By collecting more than two hundred real user profiles through surveys, as well as the ads seen by the surveyed users, we found that mobile ads delivered by a major ad network, Google, are personalized based on both users' demographic and interest profiles. In particular, we showed that there is a statistically significant correlation between observed ads and the user's profile. Since users of different demographics tend to get ads of different contents, we also demonstrated the likelihood of learning users' sensitive demographic information such as gender (75% accuracy) and parental status (66% accuracy) through personalized ads. These findings illustrate that in-app advertising can leak potentially sensitive user information to any app that hosts personalized ads, and ad networks' current protection mechanisms are not sufficient for safe-guarding users' sensitive personal information.

I. Introduction

In-app advertising allows mobile application developers to generate revenue despite publishing their work for free. As in traditional web-based advertising, personalization improves the effectiveness of in-app advertising (and thus increases the revenue earned by app developers). It is well understood that such personalization is only possible if certain user information (e.g. interests, demographic information) is available to the party that serves advertisements, and thus privacy leakage is always a concern. While ad personalization has been well studied for web, relatively little research explores mobile ad personalization in terms of what user information is being collected. We believe research focused on mobile ad personalization is a significant pursuit for the following reasons: 1) Mobile devices are a lot more intimate to users; they are carried around at all times and are being used more and more for sensitive operations like personal communications, dating, banking, etc. Therefore, privacy concerns regarding what information is collected for ad personalization are more serious. 2) Unlike in-browser advertising, where the advertisement content is strictly isolated from the rest of the displayed page by the well-known "same origin policy", in-app advertising operates in a new and less understood environment. Thus, in this paper we try to answer the following two questions which will be of great interest not only to privacy-conscious users of mobile applications, but also to advertisers who try to target specific audience groups:

1) To what degree are in-app advertisements personalized to target different attributes of a user (e.g. interest, demographics)?
2) How much can an app learn about a user by observing personalized advertisements?

To achieve our goals, we collected ground truth demographic data from more than two hundred real users and tested the correlation between the demographic data with advertisements observed by each volunteer user. This correlation allows us to establish that certain advertisements are statistically more likely to be shown to users of one demographic group than another. We also used the data collected to train models to predict a user's interest/demographic information based on advertisements he/she receives. The accuracy of the generated model indicates how much an ad-hosting app can learn about the user by merely observing the personalized advertisements received.

The work presented in this paper marks significant improvement in the methodology for studying mobile ad personalization as well as an extension in scope for such studies. Specifically, our work is the first that is based on ground truth data collected from real users, while prior work [25], [37] mostly study advertisements received by synthetic users that are expected to have certain interests or belong to certain demographic groups. We argue that results obtained using synthesized users can never be conclusive if we do not know how the studied ad-network builds user profiles. For example, while one may try to make a user appear like a middle aged white male to the advertiser/ad network by downloading

NDSS '16, 21-24 February 2016, San Diego, CA, USA

http://dx.doi.org/10.14722/ndss.2016.23353

and running apps that are predominately used by the target population, one cannot know for sure that this is actually a signal that the advertiser/ad network is listening for or whether this signal is strong enough for the advertiser to conclude the user is a middle-aged white male.

In contrast to prior work that mainly studied how a user's interests are used in ad personalization, our work also studied how one's demographic information is used. We believe this is an important extension in scope, because the collection and utilization of demographic information (like income) should raise more concerns than personalization based on a user's interests alone. Despite previous studies suggesting that demographic based targeting had not been widely observed in mobile in-app ads [37], our findings (discussed next) turned out to be different, which may have resulted from the different approach we applied in our study (see Section IV-D for a discussion on why our approach is more successful in identifying ad personalization based on the user's demographics).

Previous studies have shown the possibility of privacy leakage through web advertising [26], [38]. We studied whether the environment of in-app personalized mobile ads presents new privacy threats. Specifically, we investigated whether users' demographic profile can be reconstructed based on the ad contents delivered to mobile apps. Ideally, as the case of in-browser advertising, personalized ads are delivered directly to users in the iframe of ad networks, and thus only users can know the ad content personalized on their information. However, unlike web advertising, mobile in-app advertising allows app developers to access users' personalized ads. These ads might reflect users' real interests and other demographics, because the ads run in the same process space of the app.

To maintain a reasonable scope for our work, we focused on the Android platform and Google's mobile ad network - AdMob. Given the major market shares of Android (76.6%) in current mobile shipments and that of AdMob (35%) on Android devices [14], [24], we believe they are very good representations for studying mobile advertisements. We note that even though we only studied one ad network, the same methodology can be applied to study other ad networks and to determine whether/how much other user data (e.g. sexual orientation) is effectively used in ad personalization and might be leaked to apps hosting those ads. The amount of leaked personal data depends on the degree of personalization of the studied ad-network, *e.g.*, adversaries are expected to learn less or even no personal information from a less sophisticated ad network than from Google's ad network.

Our analysis on data collected from real users shows that mobile ads delivered by Google are heavily personalized based on both user's interests and demographic information. More than 57% of ad impressions for 41% of the users match user's real interests. More than 73% of ad impressions for 92% of users are correlated with user's demographic information. We note that our findings are, to the best of our knowledge, the first to suggest a user's demographics information plays a significant role in determining which ads he/she receives.

We further demonstrated that in-app personalized advertising opens a new channel for leaking personal information collected by ad networks to any party that has access to the ads. Specifically, we showed that an adversary with access to user's personalized ads is able to predict a user's gender and parental status with an accuracy of 75% and 66%, respectively. In addition, the adversary is able to predict a user's age with an accuracy of 54%, which is significantly higher than that by random guess (33%). A surprising finding with a more profound privacy implication is that by using the same method, adversaries can also predict other information about that user, such as income, political affiliation, and marital status, with significantly higher accuracy than random guesses, even though none of this demographic information is known to be used for personalization. In fact, some demographic identifiers (e.g. ethnicity, political affiliation and religious beliefs) are deemed so sensitive that Google explicitly stated that they will not be used in ad personalization [7], [10]. This finding shows that in in-app advertisement settings, a guarantee from Google is no longer enough for protecting the user's privacy, since user information that Google uses for personalization can be inadvertently leaked to any third party that host Google ads, and Google has no control over how such leaked information can be used to derive more sensitive information about the user. As pointed out in [27], even the most benign looking user information can be combined to infer more sensitive details about the user.

The contributions of our work are as followed:

1) We examined personalization in mobile in-app ads to determine which demographic information may have been used for advertisement personalization.
2) We were the first to seek ground truth regarding user's demographics and personal interests and to develop a new methodology for leveraging real users in evaluating the quality of mobile ads personalization. By using the new methodology, we were able to answer whether certain ads were correlated with user's demographics information, and to determine how much ad networks know about their users.
3) We studied the possibility that app developers can extract user's demographic information merely by observing personalized advertisements delivered to the user by an ad network (which we assume has already "learned" such data, but is not explicitly sharing it).
4) We demonstrated that sensitive information - including that which is not explicitly used in serving tailored ads - could be leaked through personalized mobile in-app ads.

The rest of this paper is structured as follows. We present background about mobile in-app advertising in Section II and describe our methodology in Section III. In Section IV we present our data and characterize mobile in-app ad personalization. We demonstrate the possibility of privacy leakage in mobile in-app advertising in Section V and discuss limitations of our study and countermeasures in Section VI. Finally, we discuss related work in Section VII and conclude in Section VIII.

II. Background

In this section, we briefly describe the ecosystem of mobile advertising and its related targeting mechanisms. We also discuss the differences between web advertising and mobile

advertising that cause potential privacy leakage on mobile platforms.

A. Ecosystem of Mobile Advertising

Publishers, advertisers, and ad networks are the three main components of both web and mobile advertising. The only difference between web and mobile advertising is that in web advertising, the only kind of publishers are owners of websites, while in the mobile case, publishers can also be developers of apps who might spare some of the screen real-estate for in-app ads (e.g. banner). Advertisers, on the other hand, set up ad campaigns to show their ads to specific users in apps if requested by the publishers. In return, the advertisers pay the publishers for serving their ads, which might potentially generate more transactions later from users if they are interested. In order for the publishers to connect with the advertisers, ad networks are formed. By partnering with millions of publishers, it is possible for the ad networks to integrate user information contributed by participating apps, generate profiles to predict various attributes of the user (e.g. age, gender, income level) and use these profiles to push targeted ad campaigns from advertisers to certain group of users. Such data collection and profile building is of paramount importance to all three parties, since accurate targeting is crucial for both effectiveness of ad delivery [29] and increasing publisher revenue [19].

B. Targeting in Mobile Advertising

Ad networks can monitor app activities, app lists, device models, etc. on mobile devices to automatically collect and infer the users' demographic and interest profiles. Information like demographics, geo-locations, etc. can also be provided from app developers through ad control APIs [15] for better quality targeting in order to maintain a higher click-through rate of ads, resulting in higher revenue. On major platforms like Android, since most users would login to their Google account before starting to use the devices, more personal information can be gathered from these accounts. With all the potential paths for information collection, an ad network is able to use these personal features to create/update user profiles, and push personalized in-app ads to targeted users.

We have studied the interface provided by major ad networks (*e.g.*, Google) for advertisers to specify their target population, and concluded ad networks generally provide the following three types of targeting: topic targeting, interest targeting and demographic targeting. Such offerings to the advertisers suggest that the ad networks have at least some estimate for each user regarding the attributes that can be used for targeting. In Section IV, we will try to confirm our hypothesis and also look for evidence to suggest that the ad network (Google) may have collected other information about a user that is not currently made targetable for advertisers. Notice that each of the targeting mechanisms can be employed in combination with other options for more accurate ad personalization [6].

Topic Targeting. Topic targeting lets advertisers place their mobile ads in apps that are related to the ad content. Simply by selecting one or more ad topics through an ad network interface, advertisers can have the ad network deliver to apps that are relevant. For example, by targeting the "Autos & Vehicles" topic, advertisers can ensure that auto-related ads are pushed to apps that include content about cars or other automotive themes. More precise subtopics, such as "Truck & SUVs", are also included in the general topic of "Autos & Vehicles" to achieve more effective topic targeting [17].

Interest Targeting. Interest targeting involves reaching to users interested in products and services similar to those advertisers offer, even when they are using apps that are not directly related to the products or services that are advertised. The interest profiles of users can be pre-built by the ad network, based on users' usage patterns on mobile devices, ad categories that they have clicked on before, and so on. Cross-platform correlation for interest profile might also be necessary for locating the same user across PC and mobiles. By having advertisers choose the interest categories, the ad network can advertise to those who have shown interests in the same categories before in their profiles [11].

Demographic Targeting. Advertisers use demographic targeting to deliver ads to users who are within a chosen demographic group. For example, if the advertised business caters to a specific set of users within a particular age range (e.g. younger people like sport cars better), then targeted ads to that group of people are more effective than others [12].

C. (Lack of) Isolation for In-App Advertising

In web advertising, the in-browser ads are usually delivered directly in iframe from ad networks to users [35]. These ads on websites are isolated from publishers of websites in terms of its content and code due to the Same Origin Policy (SOP). Thus, usually only the users will be able to view the ad content that is personalized by ad networks based on their collected personal information. In-app advertising however, has targeted ads running in the same processes as the apps themselves with the same permission level. Therefore, all app developers are able to access to users' personalized ads in their own apps, which can be reverse-engineered to show users' real interests and demographics. In fact, Shekhar et al. have also mentioned the necessity of separating processes between an application and its advertising for security purposes [43]. In the current study, we are examining the same argument from the privacy perspective.

Even though a recent report shows that Google has considered utilizing HTTPS protocol to encrypt ad-related traffic [3], we argue that the protection is not useful regarding current privacy concerns. Since encryption can only protect at the level of communication channel between apps and ad networks, ad content is in plaintext at the time received ads are being displayed to users. Hence, app developers can still access the targeted ads delivered to their apps in the decrypted form.

III. Methodology

In this section, we first describe our research problem, then discuss the challenges and outline our approaches.

A. Goals of Study

We seek to answer two key questions regarding user privacy in personalized mobile advertising:

1) What personal information about real end users can a dominant mobile ad provider such as Google know and use in personalized mobile advertising? Specifically, we want to understand how much mobile ad providers know about real users and how that knowledge regarding real users is exploited for providing personalized ads.
2) Could personalized mobile in-app ads be served as a channel of private user information leakage? More specifically, could an adversary (i.e. mobile app developer) with access to personalized mobile ads gain any information about real users?

To answer the above questions, we need to clearly define private user information on mobile devices with respect to personalized advertising. We study two classes of personal information in our work:

Interest Profile. A user interest profile models a user's behavior on the web or/and on the phone and is built by online trackers and ad providers. It consists of labels of tens or even hundreds of interest categories. Interest targeting in mobile advertising targets users who match the combination of interest categories specified by the advertiser. For instance, Bob is a fan of video games and he spends 2 hours playing games on his smart phone each day. Bob also reads many articles about sports news through specific applications on his phone. A interest profile like $\{Games, Sports\}$ well represents Bob's interest. A developer of a new basketball game may ask ad providers like Google to target users that have interest profiles similar to Bob's. Bob may see and click an ad ofr the basketball game and then become a user of this game.

Demographics. In recent years, ad providers have started to provide a more sophisticated targeting option - demographic targeting - for advertisers. For example, advertisers can target users by gender, age and parental status on Google AdWords [1]. This indicates that ad providers are actively tracking and modeling private personal information other than interests. Google has confidently shown its knowledge of user's gender, age and parental status in its personalized service. This raises the question of what other personal information online trackers are trying to learn from their users, which concerns both consumers and policy makers. In this study, we examine the following demographic categories: *Age, Gender, Education, Income, Ethnicity, Political Affiliation, Religion, Marital Status, and Parental Status*.

B. Challenges and Our Proposed Approaches

In the process of designing our experiment to determine which information is used by Google for personalizing advertisements, we have identified two challenges that any similar experiment will need to overcome.

Triggering personalization based on target attributes

To determine whether certain user information is collected and used for advertisement personalization, we need to devise a method to "provide" the ad network with that piece of user information. For example, if we want to determine whether the user's gender is used for advertisement personalization, we need to make sure that *if gender is indeed used, the ad network should have high confidence in its estimation of the user's gender when it is serving ads to a user under observation.* A previous approach to answering this question was to build artificial user profiles by performing certain actions that are believed to be observed by the ad network (e.g. installing certain apps that are predominately installed by one gender but not the other).

We find this approach circular in nature. In particular, if we are trying to determine *what* user information is used in personalization, we must assume we do not know *how* the ad network deduces the personal information by observing the user's behavior. In fact, even the set of user behaviors observed/used by the ad network to form the user's profile is generally unknown to us. As such, there is no reliable way for us to say, for example, "if the ad network is providing gender based personalization, it must have concluded that the user under observation is a male after we have performed these operations". In other words, if our experiments based on synthesized user behavior come back negative, we cannot tell if that is because the studied user attribute is not used for personalization or if it is because of flaws in the profile synthesis process.

In this work, we overcame the above problem by recruiting real users and collecting their demographic information as well as the personalized advertisements observed by these users. By having ground truth from real users, negative results that show no difference in advertisements observed by people of different demographics can be concluded as "the ad network failed to provide advertisement personalization based on that piece of demographic information".

Isolate personalization from non-target attributes

A related problem we face in trying to determine if certain user information is used in as personalization is controlling for the other factors that are known to be used in ad personalization. For example, in in-browser advertising, we know that the user's geo-location is an important factor in determining which ads he/she sees. Similarly, if we collect advertisements seen by different users on apps they installed, the difference in the ads they receive may not be based on demographic differences; rather, they may be caused by the categories of ads requested by different apps using ad control API. From our experience in designing similar experiments, we are certain that such noise must be eliminated if we are to draw any statistically significant conclusion confirming the existence of personalization based on any user attributes that are not previously known to be used for personalization. To this end, we chose to collect personalized advertisements seen by different users with our own tailored app that does not employ any ad control API, and always send requests for advertisements from our own IP address.

C. Experiment Design

In this subsection, we present the details of our experiments, which was approved by our Institute Review Board (IRB).

Subject recruitment. We recruited Android users located in the United States from Amazon Mechanical Turk [4] as subjects to complete surveys regarding the subject's demographics

and interests. Each subject was also required to install our Android app for ad traffic collection. Using the surveys, we are able to gather ground truth about end users for evaluating mobile ad personalization. This eliminates the artifact effect of building synthesized user profiles. To ensure users pay attention to the survey, we inserted some trick multiple choice questions in random order. They are considered simple to solve and require no more than basic skills (e.g. 1+1=?). Subjects' answers to the survey would be rejected if they failed to complete those trick questions correctly. The survey questions and multiple choice questions were also randomized in order, for the purpose of removing potential biases from subjects' responses. Using survey responses along with the ads collected in our app, we were able to analyze the relationship between personal user information and mobile ads.

Ad collection. To isolate the impact of application-based targeting, we designed a blank Android app dedicated for collecting mobile ads. The app initiates 100 ad requests to ad networks without setting any targeting attributes. We selected Google AdMob as our target for ad data collection, due to Google's dominance in the mobile advertising industry. Since location targeting is prevalent in practice, we established a secure VPN connection to our server from the user's device to isolate impact of the location. In particular, all Internet traffic through all the apps installed on subject's phone was tunneled through the VPN service provided by our App, but we only collected advertising traffic generated by our App during the data collection phase. To avoid collecting ads intended for applications other than ours, we instructed our Mechanical Turk subjects to keep our app running without operating other apps until the app finished data collection and turns off the VPN tunnel. The entire data collection process took about 2-15 minutes, depending on the network condition of the user's device.

We note that the VPN tunneling employed in the ad collection process makes all ad requests from our blank app originate from our IP addresses. As a result, ads collected by our app are personalized for users at our location, instead of the individual subject's real location. While this can be considered a kind of noise we inject into our data, we argue that this is also an advantage of our experimental design: we can eliminate the influence of geo-location on ad profiling, and better study how the other aspects of a user's profile (demographics, interest) affect the ads she/he receives. Furthermore, if geo-location were included in the study, we would need a significantly higher number of volunteers to cancel out its influence.

We believe that our experiment had limited impact on the subject's ad profile for two reasons: 1) we believe a subject's profile is built upon long term observation of how he/she interacts with his/her phone/apps, and thus any influence we introduce over a short interval with 100 ad requests will be insignificant, and 2) "blank" ad requests (i.e. requests that do not specify any information regarding the intended audience) from a blank app, with no clicks on the received ads, should present the ad network with no useful signals for updating the subject's profile.

Landing URL extraction. We use the landing URL (the destination URL that a user agent will be redirected to after clicking/touching an ad) as the representation of an ad in our analysis. Specifically, we tried to extract the landing URL directly from the meta data contained in the HTML source of an ad. Through our analysis on the HTML sources representing ads, we identified several attributes and keywords that helped us extract the landing URL of an ad. The attributes and keywords we used for landing URL extraction were *buildRhTextAd, adurl=, final_destination_url, destination_url, destinationUrl, click_url*, and *go.href*. For ads that we were not able to extract a valid landing URL, we replayed the ads on our server by reopening the HTML source in a desktop web browser and clicking through the ad to reach the final landing page. We also replayed some ads that the extracted landing URL directs to some known ad networks, which usually further relay the user's visit to the final destination or other ad networks. In the end, we had to replay only 2,372 out of 39,671 (5.98%) ads in our dataset to get the final landing URL; thus, our approach allowed us to minimize the impact of our study on the mobile ad ecosystem.

Landing URL post processing. The landing URLs extracted using the above approach usually contain a long list of tail attributes, which are used to identify the source of the visit, *e.g.* creative_id, campaign_id , mobile app name, and ad network. We cluster the landing URLs by removing those tail attributes and grouping them into ad campaigns. Ad campaigns are further merged if they share the same domain and prefix but different resource names, except for Google Play Store apps which all start with https://play.google.com/store/apps/details?.

Ad categorization. Each ad (landing URL) is categorized into one of the 24 root interest categories that Google provides for as targeting options. If an ad is of a Play Store app (identified by the above URL pattern), we use a script to directly extract its corresponding interest category from its Play Store webpage. For the other ads, we rely on Google Ad Preference [5] for labelling. In particular, we built a Google Ad Preference crawler using a headless browser - PhantomJS [9]. We crawled the interests that were generated by Google right after visiting each ad landing URL for a consecutive of 10 times starting with a blank browser profile. Such interests were used to label the ads correspondingly. Yahoo Content Analysis API [18] is also used in cases where we do not get any assigned categories on Google Ad Preference page from the previous method. All the Yahoo categories are mapped to their closest representation in the 24 Google root interest categories. For those ads that cannot be automatically categorized by the above steps, we manually assign the category that best matches the contents of the ad landing page. We had three persons categorizing the mobile ads independently. If the decisions on an ad from the three human labellers conflict, a final label is selected based on mutual agreement.

Ad-containing packages detection. Ad networks employ information like the list of installed apps and apps a user has used to infer his/her personal information [12]. Specifically, all interaction with one ad network will be captured through apps that contain the code of the ad network. Ad networks' libraries are typically called in the same manner by different developers. The UI elements that are used for renderring ads have the same name or identifier across different applications. This enables one to learn the applications that include the same ad library on a user's mobile device. We detected all packages

that include Google's AdMob library through the Android PackageManager the user has installed. Such information is helpful in understanding the profiling mechanism of black-box ad networks. We used the list of ad-containing packages as one class of features in our privacy evaluation in Section V.

IV. Characterization of Mobile Ad Personalization

In this section, we first present details of our dataset collected from 217 real users, then study demographics-based and interest-profile-based mobile ad personalization.

A. Dataset

We created a Human Intelligence Task (HIT) on Amazon Mechanical Turk for workers located in the United States. Each worker was asked to install our app, which will collect 100 ads when executed. After the completion of the data collection process, the worker was asked to fill out a survey regarding his or her interests and demographic information (as defined in Section III-A). We had run our HIT for 12 days on Amazon Mechanical Turk. In the first 3 days of our experiment, each worker was compensated for $1.00. To attract more workers to work on our HIT, we increased the reward amount to $1.25 on the 4th day.

In total, we were able to collect survey responses from 284 users and successfully collected 100 ads for each of 217 users. The other 67 users quit before the ad collection was complete, and we discarded the partial data collected. Table I shows the distribution of the demographics of the 217 users in our dataset. Out of 39,671 ad impressions we captured, 33,135 ad impressions for 695 unique ads were issued from the 217 users. We observed that some users had run our data collection application multiple times, and for these users, we only used the first 100 ads collected. Figure 1 displays the distribution of the 695 unique ads in terms of ad impressions. Surprisingly, two ads (*Zoosk - #1 Dating App*[1]: 3,461, *Samsung for Business*[2]: 2,602) accounted for 28% of total ad impressions in our dataset. In the 695 unique ads, 500 (72%) are of applications on the Google Play Store, which generate 9,124 impressions (42%). The remaining 195 (28%) ads contribute to 12,576 impressions (58%). Figure 2 shows the distribution of the 695 unique ads in terms of number of users. Five ads were delivered to more than 50% of the 217 users. Figure 3 gives the number of unique ads displayed to each user, Figure 4 breaks down the ad impressions into interest categories, and Figure 5 presents the number of users that have received ads in each category.

TABLE I: Demographics distribution of subjects.

Gender		Political Affiliation		
Female	Male	Independent	Democrat	Republican
95	122	108	80	29
43.78%	56.22%	49.77%	36.87%	13.36%

Parental Status		Income		
Not a parent	Parent	< $30K	$30K-$60K	> $60K
128	89	107	67	43
58.99%	41.01%	49.31%	30.87%	19.82%

Religion		
Atheist	Non-Christian	Christian
82	47	88
37.79%	21.66%	40.55%

Age				
18-24	25-34	35-44	45-54	55+
45	106	47	14	5
20.74%	48.85%	21.66%	6.45%	2.30%

Marital Status		
Single	Married	Separated, divorced or widowed
124	73	20
57.14%	33.64%	9.22%

Ethnicity				
Other	Hispanic	Asian	African American	Caucasian
8	12	12	23	162
3.69%	5.53%	5.53%	10.60%	74.65%

Education			
High school or less	Associates	Bachelor	Master or higher
78	50	71	18
35.94%	23.04%	32.72%	8.30%

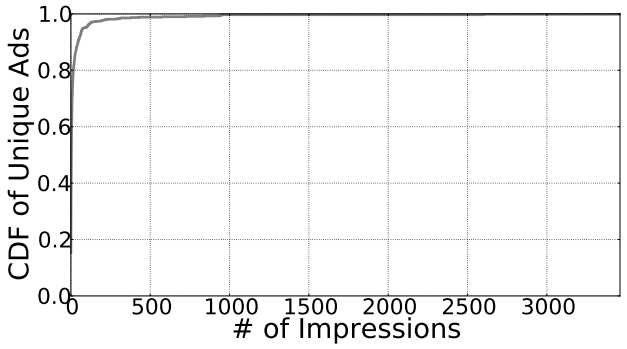

Fig. 1: Impression distribution of unique ads.

B. Interest Profile Based Personalization

In this subsection, we study interest profile based personalization. Specifically, we try to understand how well ad networks learn about the real users' interests.

We use $P_{user,i}$ as the real user interest profile derived from the survey response from user i. The ad interest profile $P_{ad,i}$ is defined as the set of interest categories of all ads delivered to user i. We use the following three metrics to evaluate the similarity between real user interest profile and ad interest profile.

1) Size of an interest profile, which is number of categories in each interest profile.
2) Precision, which is defined as $|P_{user,i} \cap P_{ad,i}|/|P_{ad,i}|$. Precision represents the fraction of categories in an ad interest profile that match the user's real interest profile. It measures how precisely ad networks know user's real interests.
3) Recall, which is defined as $|P_{user,i} \cap P_{ad,i}|/|P_{user,i}|$. Recall is the fraction of categories in the real user interest profile that are presented in the ad interest profile. It represents the ad network's coverage of the users' real interests.

[1]https://play.google.com/store/apps/details?id=com.zoosk.zoosk
[2]http://www.samsung.com/us/business/samsung-for-enterprise/

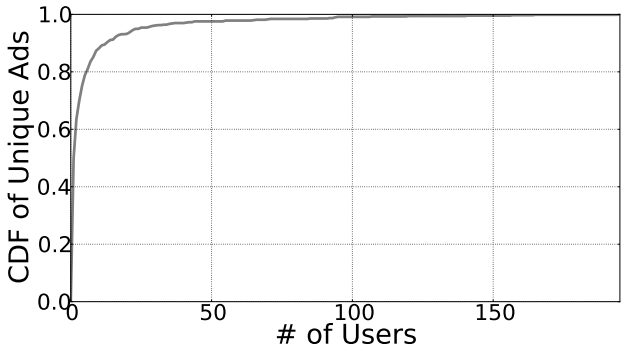

Fig. 2: User distribution of unique ads.

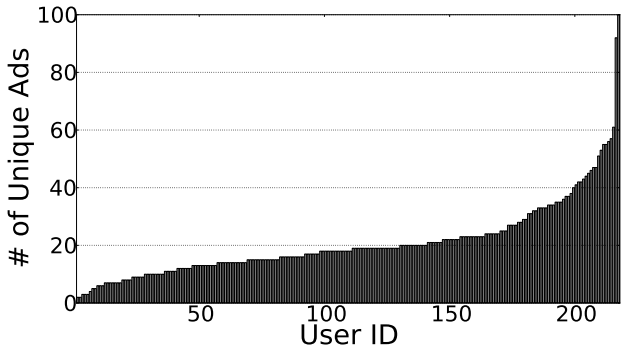

Fig. 3: Number of unique ads of each user.

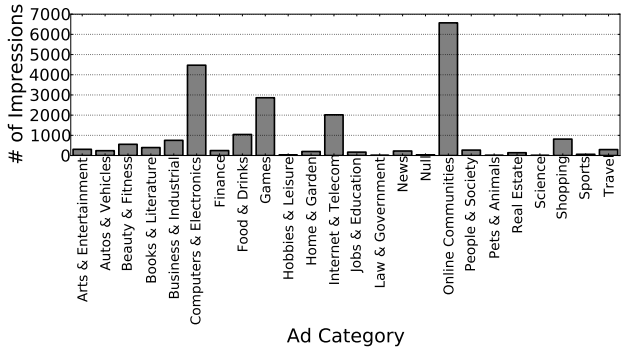

Fig. 4: Number of ad impressions in interest categories.

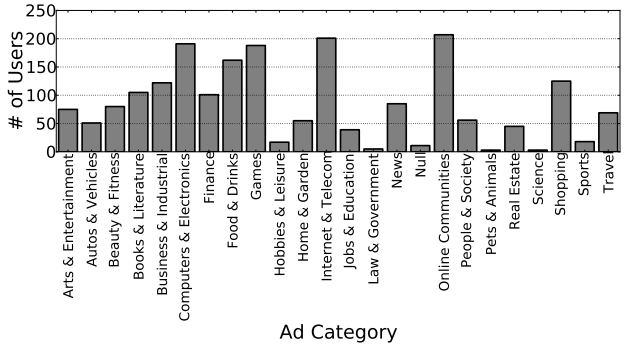

Fig. 5: Number of users in interest categories.

We first show the sizes of real user interest profile and ad interest profile for each user in Figure 6. The sizes of the two interest profiles vary significantly across users, indicating a diverse distribution of user interests. As we can see in Figure 6, there is no clear correlation between the sizes of the two profiles, suggesting size of interest profile is not a good metric for evaluating the similarity between the two profiles.

The distributions of Precision and Recall are shown in Figure 7 and Figure 8, respectively. For over 79% of the users, at least 21% of the categories in the ad interest profiles are correct. For 11% of the users, at least 83% of the categories in the ad interest profiles are correct. In terms of Recall, Google could cover at least half of real user interests for 60% of the users. The results demonstrate that ad networks like Google can build accurate interest profiles of mobile users and use the profiles built for personalizing mobile in-app advertisements.

To understand how many ads are personalized based on real user interests, we further looked for ads that match real user interests, which we refer to as *precise ads*. Figure 9 displays the number of precise ad impressions of the users. The result suggests that Google is actively personalizing a large fraction of its ad deliveries. For 41% of users, more than 57% of their ad impressions match their real interests.

Summary. Our analysis shows that mobile ads are highly personalized based on user interests. The ad interest profiles derived from observation of 100 ad impressions are quite close to users' real interest profiles with good precision and recall. More than 83% of the categories in the ad interest profiles are correct for 11% of users, and more than 50% of real user interest categories are covered in ad interest profiles for 60% of users. We further find a large fraction of mobile ads match with real user interests. More than 57% of ad impressions for 41% of users fit users' real interests.

C. Demographics Based Personalization

As discussed in Section II, since ad networks allow advertisers to target their ads towards specific demographic groups, we strongly believe the ad networks already have profiles that capture various demographic information of the users. In this subsection, we seek to quantify to what extent real users' demographics may have been used for ad personalization. Note that as shown in [1], [12], gender, age and parental status are the only 3 demographic categories that Google explicitly allows advertisers to use for targeting purpose. Thus our observation of strong correlations between ads and other categories of demographic information may not be the results of explicit ad targeting. However, for the sake of brevity, in the following discussion we will attribute strong correlation between ads and demographic information to (possibly unintended) personalization.

We grouped the 217 users into different demographic sets for each demographic category. We employed statistical tests to determine whether one ad is correlated with a given demographic category. Specifically, we counted the number of times an ad is shown to users in different demographic groups. Then the ad was tested for independence with the demographic category by using Pearson's chi-squared test. We excluded ads

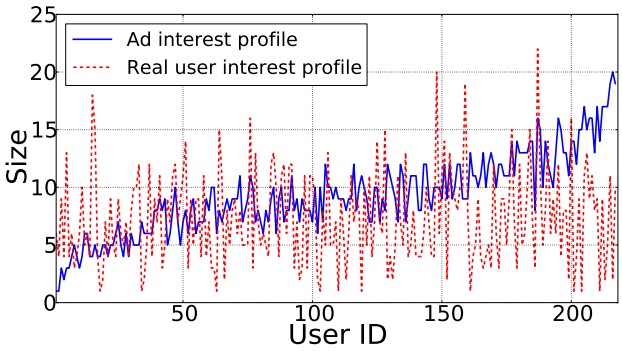

Fig. 6: Number of interest categories in real user interest profile and ad interest profile.

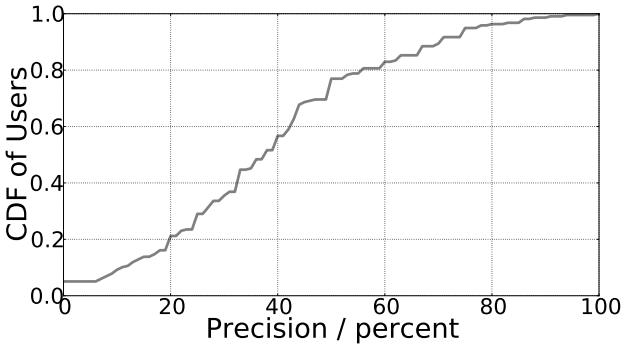

Fig. 7: Precision distribution of user profiles.

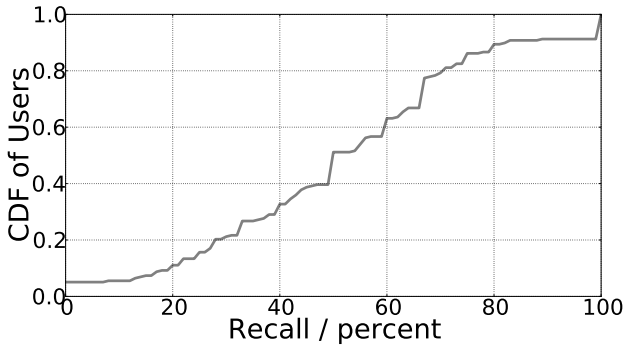

Fig. 8: Recall distribution of user profiles.

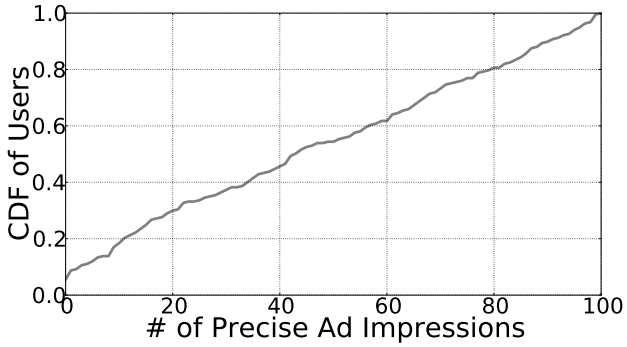

Fig. 9: Number of precise ad impressions of users.

with an expected number of impressions fewer than 5 for any demographic group, which is a common practice when applying Pearson' chi-squared test. The null hypothesis is that the ad being tested is independent of the demographic category. We set the significance level for our tests to be 0.005. If the p-value of one ad is less than the significance level, we reject the null hypothesis and label the ad as personalized based on (correlated with) the demographic category under test.

The number of unique personalized ads in each demographic category is presented in Figure 10. It is not surprising that many ads are targeting users by gender. For example, the ad for the game *Game of War - Fire Age*[3] is shown to 66 males (70%) for 614 impressions (64%), while only 28 females (30%) received the remaining 342 impressions (36%). On the other hand, the ad impressions of the game *Cookie Jam*[4] are dominated by female users. 182 impressions (96%) were shown to 13 females, while 4 males share the remaining 7 impressions (4%). There are even 33 ads that are exclusive for one gender class.

Parental status is the second most personalized demographic category in our result. Ads of social applications and websites like *Zoosk - #1 Dating App* and *Facebook*[5] are leaning toward users that are not a parent. 2,312 ad impressions (67%) of Zoosk and 403 ad impressions (70%) of Facebook are shown to non-parent users. Interestingly for Marital Status, we found social applications and websites also show preference to non-married groups.

We are surprised that there are many ads that are dependent on users' income level. In our dataset we found ads of many games (*e.g.*, *FINAL FANTASY Record Keeper*[6] : 71%, *Cookie Jam* : 69%, *League of Angels -Fire Raiders*[7] : 67%, *World Series of Poker - WSOP*[8] : 67%) are shown more toward users in our low income group (with annual gross income below $30,000). We could not find any income targeting option on Google AdWords and Google AdMob for advertisers. We do not think advertisers on Google currently are explicitly targeting users by income. The result suggests that Google may tailor ad deliveries based on users' income as a result of its personalization algorithms to further increase click-through-rate or conversion rate, which we cannot prove for sure. However, the practice of ad syndication makes it possible that those impressions may be purchased through other ad networks that offer income targeting and other demographics targeting [16]. It is also possible that the correlation with income we observed is a result of income's correlation with age, parental status, or other demographic information. For example, in our dataset older people who have children generally have higher incomes.

Our statistical test suggests 58 unique ads are correlated with users' religion. However, we found only one ad, *Peace With God*[9], which is clearly related to religion. 57 impressions

[3]https://play.google.com/store/apps/details?id=com.machinezone.gow
[4]https://play.google.com/store/apps/details?id=air.com.sgn.cookiejam.gp
[5]https://www.facebook.com/r.php
[6]https://play.google.com/store/apps/details?id=com.dena.west.FFRK
[7]https://play.google.com/store/apps/details?id=com.gtarcade.loa.ph
[8]https://play.google.com/store/apps/details?id=com.playtika.wsop.gp
[9]http://peacewithgod.net/mobile/

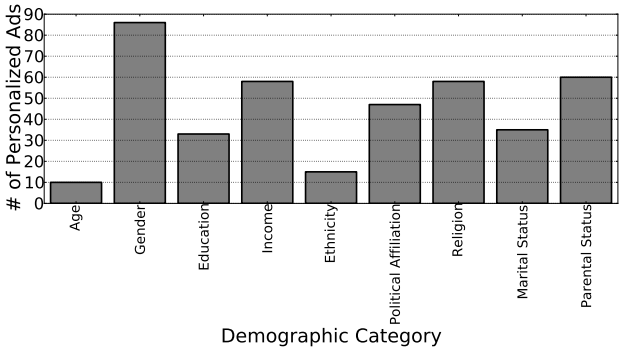

Fig. 10: Number of unique ads that are personalized based on demographics.

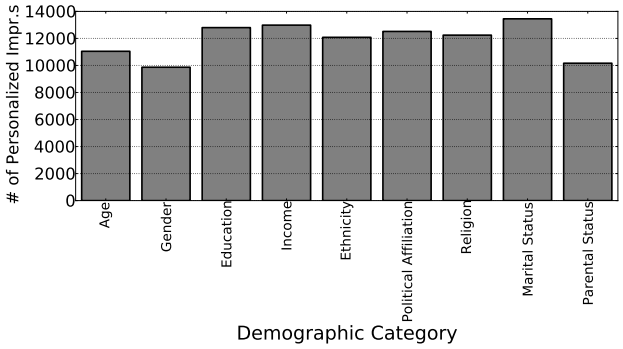

Fig. 11: Number of ad impressions that are personalized based on demographics.

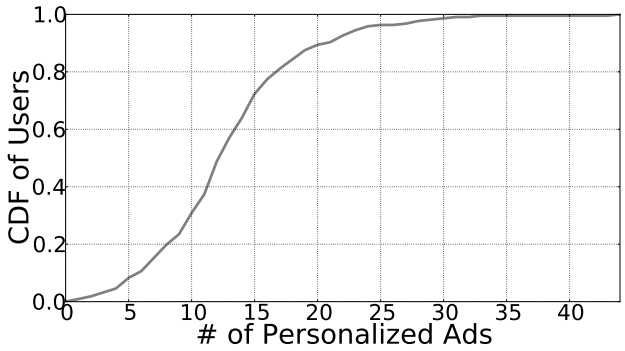

Fig. 12: Number of unique ads that are personalized based on demographics across users.

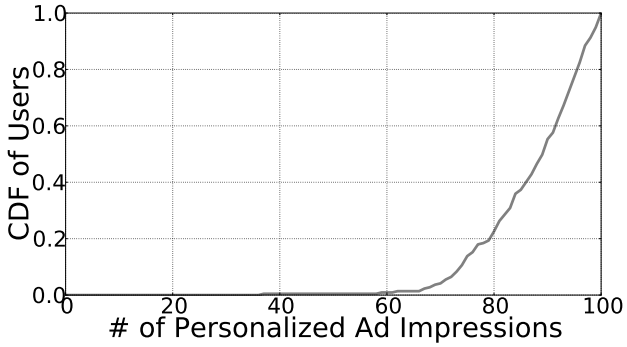

Fig. 13: Number of ad impressions that are personalized based on demographics across users.

(57%) of this ad are shown to Christians. We conjecture carefully that the dependencies of the other 57 ads may be a result of religion's correlation with other demographic categories. Similarly, we could not find any explicit evidence of targeting by Political Affiliation. We cannot explain the correlations of ads with users' Political Affiliation due to a lack of insight into ad networks' secret personalization algorithms.

For age, education and ethnicity, we observed personalization in lower degrees in terms of number of correlated ads. However, Figure 11 gives a different view of demographics based personalization. We present the number of impressions of the personalized ads in each demographic category in Figure 11. Except for gender, all demographic categories have a number of personalized ad impressions greater than 10,000 (46%). Although there are fewer personalized ads in some demographic categories, the effects of personalized ads in these categories are not significantly lower than those of personalized ads in other categories.

We further quantify the impact of demographics based personalization on individual users. Figure 12 and Figure 13 show the distribution of users in terms of number of demographically personalized ads and number of demographically personalized ad impressions they receive, respectively. 76% of the users have received at least 10 ads that are personalized based on demographics. Surprisingly, at least 73 out of the 100 impressions are of ads that are personalized based on demographics for 92% of the users. Note that the values derived here include personalized ads and impressions displayed to demographic groups that are not the primary targets of the ads.

Summary. Using statistical tests, we found ads that are personalized based on (correlated with) demographics. Gender is the demographic category that we observed the highest number of unique personalized ads. Personalization may be an explicit targeting option expressed by advertisers (age, gender and parental status), or it may be result of an ad network's proprietary personalization algorithms (income, religion, etc.). Our results in Section V will shed more light on our observation about demographics based personalization.

We also found that demographics based personalization in mobile advertising is prevalent in practice. 76% of our users have received at least 10 demographically personalized ads, and more than 73% of ad impressions of 92% of users are demographically personalized. Ads that are delivered exclusively to certain demographic groups are highly indicative of a real user's personal information. Together with non-exclusive ads that are correlated with some demographic categories, those ads may be a good representation of real user's demographic profile. This raises a great opportunity for adversaries to learn the private personal information of real users. As we will demonstrate in Section V, personalized ads serve as a new channel for privacy leakage on mobile devices.

D. Comparison with Previous Studies

Previous studies showed that demographic information of users was not commonly used for mobile in-app ad personalization (only found on Google's ad network), and the user's geo-location played a significantly more important role than his/her interests/demographics in determining what in-app advertisements he/she is receiving [37]. In contrast, our results illustrate that both demographic and interest profiles of users have a statistically significant impact on how in-app ads are selected for the same ad network studied (We did not study geo-location as discussed in Section III). Some explanations for the discrepancy between the results across studies are as followed.

First of all, what [37] measured to show the significance of demographics information in ad personalization is very different from what we are measuring. In particular, most of the results in [37] used keywords/ad-control attributes in ad requests generated by applications to measure *how often demographics based ad personalization is requested by apps*. However, in this paper, we studied *how often one's demographics information strongly correlates with the ads he/she receives* and use it as an indication of successful demographics-based ad personalization. In other words, [37] measured personalization requested by the app developers, while we looked for evidence of successful personalization performed by the ad network, which may have a significantly better ability to profile a user than app developers.

Secondly, instead of using synthesized user profiles to harvest mobile in-app ads as in the previous studies, we collected ads using real user profiles, which led us to observe more ads that are correlated with user's interest and demographics. We believe that authentic user profiles could generate the right signals to trigger ad personalization, while synthesized user profiles might not. In addition, we studied more complete demographic profiles from users than previous ones, which enabled us to discover new types of demographics based targeting that were not observable in previous studies. We believe that our findings compliment those from previous studies and helps the research community better understand mobile in-app ad personalization.

V. Privacy Leakage through Personalized Mobile Ads

As shown in Section IV, mobile ads are highly personalized based on user interests and demographics. As a result, a set of ads collected from a user's device can be seen as an accurate representation of that user's real personal information, potentially including sensitive data. We have already shown in our real user study that the ad interest profiles inferred from 100 ad impressions closely match with real user interest profiles. It is also possible to infer user's demographics from personalized mobile ads, as we will demonstrate in the following subsections. The rest of this section is organized as followed: we discuss the feasibility of privacy leakage through personalized mobile ads in Section V-A; our experiment setup and definition of evaluation metrics will be presented in Section V-B; finally, in Section V-C, we present the evaluation results. Section V-D discusses the privacy implication from our experiment.

A. Technical Feasibility

Previous studies have shown the possibility of reconstructing user interest profiles in web advertising under the threat model that the adversary can eavesdrop on a victim's unencrypted ad traffic [26]. To provide stronger security in serving online advertisements, the online advertising industry is taking steps towards enhancing the security of ad transmission through HTTPS protocol [2], [3]. The adoption of HTTPS in ad serving could certainly defeat the above attack in web advertising. However, we argue that in the scenario of personalized mobile advertising, the threat to user privacy is much greater even with the protection that HTTPS provides.

In contrast with web advertising where the personalized ad contents are protected from publishers and other third-parties by the Same Origin Policy, there is no isolation of personalized ad contents from the application developers on a mobile platform such as Android. As such, an adversary does not need to sniff the ad traffic of a victim mobile device user. Even when HTTPS is enforced, any host application can still read the ad contents displayed within it regardless of whether encryption is enabled during data transmission. The ability of observing personalized ads on mobile devices opens a new attack vector for private personal information leakage, which we will demonstrate next.

B. Demographics Learning from Personalized Mobile Ads

In Section IV, we showed that one could recover more than half of the real user interests for 60% of users by categorizing ads into interest categories. In this section, we will try to determine if the same can be done for the user's demographics information, which is potentially more sensitive than the user's interests. In particular, we applied machine learning algorithms to build models for predicting a user's demographic information (for all studied categories) based on the ads he/she has seen, and evaluated the accuracy of the generated models.

We use the combination of the number of impressions of ads that are correlated with each demographic category and the list of installed apps that contain the Google AdMob library as features. Each sample is labeled with one class in the corresponding demographic category according to the survey response. We tested a set of basic classification algorithms (Decision Tree, Logistic Regression, Multinomial Naive Bayes, K-Nearest Neighbors, Random Forest, SVM) to estimate our ability to learn users' demographic info. A dummy classifier that predicts by randomly guessing was used as the baseline classifier for comparison.

The classifiers are implemented using the *scikit-learn* package [13] of Python. All classifiers are evaluated with 5-fold cross validation to avoid overfitting. Specifically, the 217 samples are randomly divided into 5 different sets (folds), and for each fold, the other 4 folds of samples are used as training set to train the model. The resulting model is then validated using the remaining fold as test set. For the sake of fairness, all classification algorithms are cross-validated using the same division of 5 folds.

We used the accuracy of the prediction as the metric for measuring the severity of privacy leakage through personalized

mobile ads. We define the accuracy of a classification model as the number of accurate predictions divided by number of all predictions. Note that one prediction is accurate only when the predicted class is exactly the same as the label. Thus for binary classification problems (*e.g.*, gender and parental status) the dummy classifier has accuracy of 50%. For multi-class classification problems, which are harder than binary classification problems, the accuracy of dummy classifier is 1 divided by number of possible classes. We report the cross validated accuracy (the mean of accuracies of the 5 validations) as the accuracy of one classifier.

A point worth emphasizing is that in a perfectly safe/privacy-preserving system, the adversary should have no advantage in knowing victims' personal information, *i.e.* the adversary cannot have better accuracy than that obtained from tossing coins. Thus if the accuracy of an adversary's model is significantly higher than that of the dummy classifier, it suggests that the adversary has significant advantage in learning victims' personal information. Our goal is to understand the possibility of privacy leakage in personalized mobile ads, thus any result that is above the baseline accuracy (the accuracy of the dummy classifier) should be considered as a potential source of privacy leakage. We present our findings in next part of this section.

C. Evaluation

Table II lists the accuracy of all the classifiers and the accuracy of the dummy classifier for each demographic category. The cell in bold represents the highest accuracy score in each column. Overall, SVM performs the best in predicting all demographic categories. For all demographic categories, we could find at least one classifier that performs much better than the dummy classifier. This is particularly true for gender, that three classifiers (Logistic Regression, Multinomial Naive Bayes, and SVM) are able to predict accurately for over 70% of the users. Such a result was expected, as we have shown in Section IV that gender is the most heavily used in ad personalization. The same three classifiers also perform well for parental status, with accuracies above 65%. Surprisingly, four classifiers are able to accurately predict the ethnicity of more than 70% of users, in contrast with the 20% accuracy of the dummy classifier, but this can be attributed to the distribution of our sample. By comparing with the result in Table I, we find that the distribution of our subjects is highly biased in terms of ethnicity. 74.7% of our subjects are Caucasians, and according to the United States Census Bureau, 72.4% of the U.S. population was Caucasian in 2010 [8]. With this knowledge, anyone is able to predict with an accuracy of around 72.4% on data set that is randomly sampled. Thus, we do not claim that the high prediction accuracy of our classifier for ethnicity comes from the advantage of having access to real user's personalized ads. High bias in sample distribution (*e.g.*, ethnicity) is known to result in biased classifiers. Unfortunately, we find the distributions of our subjects are biased for age, education, income, political affiliation, religion, and marital status as well, due to the small size of our dataset.

To remedy the aforementioned limitation of our data set, we reorganized our data to make it more evenly distributed between different classes (i.e. different values in each studied category). To this end, we merged some of the less popular classes in age, marital status, political affiliation, income and education; the distributions of demographics in the merged classes are as shown in Table III. As a result, some of the previous multi-class classification problems are reduced to binary classification problems. We augmented the dummy classifier by using a majority selection strategy, *i.e.* it always outputs the most popular label in the training set. Table IV lists the accuracies of all the classification models on the new classification problems. Since the number of classes in the 5 demographic categories was reduced, we observed improvement on the performance of classifiers. The accuracies of our classifiers for age, income, marital status and political affiliation are better than random guess and the case with prior knowledge of population distribution. For instance, we could accurately learn the income, or marital status for more than 60% of users. The information derived from personalized ads indeed helps one predict a users' personal information with better accuracy. However, the price of the performance improvement is the coarser granularity of the prediction. For example, the new classifier can not differentiate people whose annual income is higher than \$60K with people who earn less than \$60K but higher than \$30K per year. None of the classifiers performs significantly better than augmented dummy classifier for education and religion, which suggests the adversary has little advantage for the two categories.

In addition to studying the possibility to predict individual aspects of the user's demographic, we also try to determine for each user how many of the 9 studied demographic categories the adversary may learn by monitoring his or her personalized ads. We record the accurate demographic predictions of each user in the 5-fold cross validation. Figure 14 presents the distribution of the number of correct predictions for demographic categories across users. For 91% of the users, at least 4 demographic categories of the users could be accurately predicted[10]. There are even 2 users that the predictions for all the 9 demographic categories are correct. Figure 15 shows the breakdown of the correct predictions for each user.

Summary of Results. We demonstrated the possibility of leaking user's sensitive personal information through personalized mobile ads to third-party app developers. With data from about 200 users, we were able to build classifiers that predict gender with over 70% accuracy and parental status with over 65% accuracy. By balancing the subject distribution in age, income, political affiliation, and marital status, we could predict a user's corresponding demographic class with significantly better accuracy than random guess and the case with prior knowledge of population distribution. We are not able to build classification models that have significant advantage over an augmented dummy classifier for education, ethnicity and religion. We discuss the privacy implication of our results next.

D. Interpreting our Results

Our results presented in Table II and Table IV indicate that our ability to predict a user's gender, age and parental status is significantly higher than that of predicting other types of demographics information. On the surface, this is neither surprising nor alarming in a privacy perspective. It

[10]This can be different set of 4 categories for different users

TABLE II: Accuracy of classifiers of demographic categories.

	Age	Education	Ethnicity	Gender	Income	Marital Status	Parental Status	Political Affiliation	Religion
Decision Tree	**0.51**	0.30	**0.76**	0.64	**0.47**	**0.62**	0.62	**0.50**	0.35
Logistic Regression	0.38	0.37	0.72	0.73	0.45	**0.62**	0.65	**0.50**	0.39
Multinomial NB	0.37	0.35	0.61	0.73	0.36	0.49	0.65	0.41	0.43
K-Nearest Neighbors	0.39	0.34	0.75	0.65	0.45	0.47	0.59	0.45	0.40
Random Forest	0.39	0.36	0.68	0.67	0.43	0.59	0.58	0.46	0.41
SVM	0.49	**0.40**	0.75	**0.74**	**0.47**	0.59	**0.66**	0.49	**0.46**
Dummy	0.20	0.25	0.20	0.50	0.33	0.33	0.50	0.33	0.33

TABLE III: Reorganized distribution of demographics of subjects.

Age			Marital Status	
18-27	28-33	34+	Single	Not single
71	71	75	124	93
32.72%	32.72%	34.56%	57.14%	42.86%

Political Affiliation		Income	
Independent	Non-Independent	< $30K	> $30K
108	109	107	110
49.77%	50.23%	49.31%	50.69%

Education		
High school or less	Associates	Bachelor or higher
78	50	89
35.94%	23.04%	41.02%

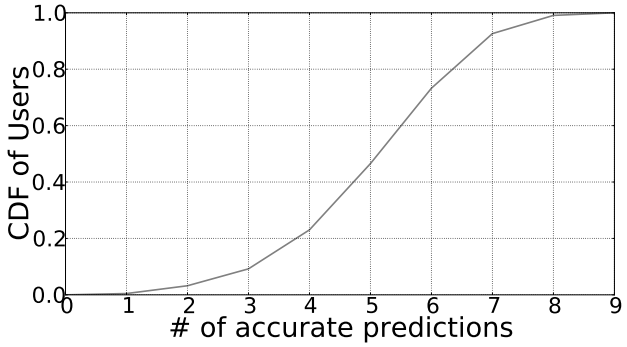

Fig. 14: Number of accurate predictions for demographic categories across users.

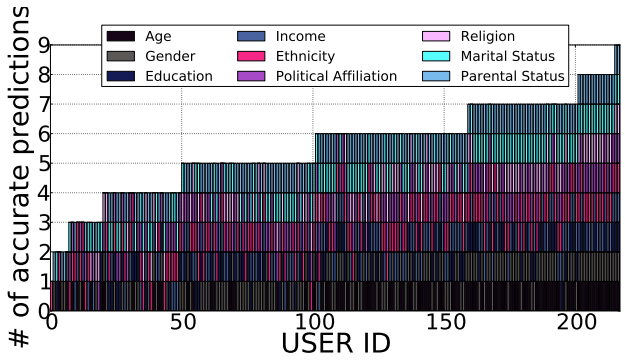

Fig. 15: Breakdown of accurate predictions for demographic categories of users.

is unsurprising because as seen in [1], [12]; gender, age and parental status are the three targeting options offered in Google's current ad product. Therfore, one can deduce user's demographic information in these categories based on what ads they see. Simply put, *our results confirmed that Google can deliver on what it is offering advertisers; it can correctly deliver ads to the specified demographic groups.* One can also argue that the privacy concern caused by leaking one's gender, age and parental status is very minimal.

However, the real surprise lies in the adversary's non-trivial gain in his/her ability to predict aspects of the user's demographic other than age, gender and parental status. Some of the other demographic information (e.g. political beliefs) is deemed so sensitive that Google explicitly stated [7], [10] that they will not even be collected as part of a user's profile; for the rest, there is no known documentation that suggests Google is using them for personalization. We believe our success in gaining some knowledge of these other aspects of the user's demographic can be explained by the Federal Trade Commission's study on data brokers [27], which shows that general non-sensitive information collectively could be used to infer more sensitive information. In particular, it is very possible that there are very strong correlations between one's age, gender and parental status with his/her other demographic information. For example, parents are much more likely to be married, and older people generally have higher incomes. This highlights the more profound privacy implication of our work.

Privacy Implication In the traditional web settings, Google can protect its users by enforcing published policies that prohibit the collection of the more sensitive user demographic information. This is an effective protection since none of Google's profiling information is leaked to the websites that host advertisements. However, such protection is no longer effective in mobile settings, where any ad-hosting app can observe the personalized advertisement being shown to the user. In this new ad-hosting environment, Google is, to a very large extent, sharing its profile of the user with the app developers and potentially other ad-networks, and Google cannot dictate how this shared/leaked information is used. As shown in our results and [27], even the sharing of the most benign user demographic information can have the adverse effect of allowing third parties to gain some sensitive demographics information about the users, and the threat from the privacy leak through this new channel is only going to increase when Google starts using other "benign" user demographic information for ad personalization.

Due to the lack of separation between in-app advertisements and the rest of the host app logic, the host app can observe all personalized advertisements without needing

TABLE IV: Accuracy of classifiers of reorganized demographic categories.

	Age	Education	Ethnicity	Gender	Income	Marital Status	Parental Status	Political Affiliation	Religion
Decision Tree	0.52	0.38	0.75	0.64	0.54	0.60	0.63	0.54	0.38
Logistic Regression	**0.54**	0.41	0.73	0.72	0.57	0.59	0.64	**0.59**	0.42
Multinomial NB	0.45	0.44	0.73	**0.75**	**0.62**	0.61	0.64	**0.59**	0.41
K-Nearest Neighbors	0.41	0.37	0.74	0.60	0.52	0.62	0.56	0.54	**0.43**
Random Forest	0.42	0.40	0.71	0.70	0.51	0.60	0.60	0.58	0.38
SVM	0.44	**0.45**	**0.75**	0.73	0.54	**0.63**	**0.66**	**0.59**	**0.43**
Dummy	0.33	0.33	0.20	0.50	0.50	0.50	0.50	0.50	0.33
Augmented Dummy	0.35	0.41	0.75	0.56	0.51	0.57	0.59	0.50	0.41

any extra permission. Since ad personalization is inherently performed based on the personal information of the user, by revealing to the host app what ad is being displayed to the user, the ad network may be inadvertently leaking some of its collected user information to the app developer. Our study shows that such leaked information can be used to accurately derive some of the user's demographic information. This is especially true for information like gender, age, and parental status, which are known to be used in ad targeting. In addition, some information that ad networks might not be explicitly collecting or using could also be leaked to the app developer. Our results indicate that one can predict significantly better a user's income, political affiliation, and marital status over random guess by observing the personalized ads that are served to a user. The information thus inferred may then be used to request ads from other higher-paying ad networks. The leaked user information may also be used for price discrimination. For example, the same good could be sold at different prices to users in different income groups. Furthermore, the private information can also be sold or transferred to other parties.

Our results highlight the need for protecting private user data (personalized ads) from unauthorized parties (app developers) on Android. The equivalence of Same Origin Policy should be provided on Android to isolate personalized ads from application context to protect user's privacy. Proposals like AdSplit [43] and AdDroid [39] are a good starting point for separating ads from applications on Android. Furthermore, before such isolation between in-app advertisement and the host app is established, ad networks should balance the gain in revenue and the risk of the user's privacy when they decide to personalize ads using more detailed or sensitive user demographic information.

VI. Discussion

A. Limitations

The main limitation of our study is the small sample size and the uneven demographic distribution of our data set. We argue that such limitations do not invalidate our results. The ads that are correlated with demographics are selected by applying statistical tests. And since we are using a significance level of 0.005 in our statistical test, we are 99.5% confident that the correlations are not observed by chance. Similarly, by using 5-fold cross validation for evaluating our ability to learn user's demographic information based on the ads he/she receives, our results in Section V confirms that the threat of leaking sensitive user information through personalized ad is real. Furthermore, we argue that aggressive/malicious app developers or ad-networks can achieve significantly better accuracy than what we've shown in Section V for two reasons: 1) they can invest more resources to obtain better ground truth data, and 2) they can observe ads received by users for a longer period of time (and thus have more highly personalized ads in their data set). In future work, we plan to apply our technique to other ad-networks and to attempt to collect ad data for a longer period of time. We will also try to improve our results in Section V by experimenting with techniques to better clean our data set (e.g. remove users who appear to receive a lot of non-personalized ads) as well as techniques like multi-task learning to better leverage our advantage in predicting the user's gender, age and parental status.

B. Countermeasures

The root cause of the studied privacy leak from personalized ads to the hosting application is the lack of isolation between the ads and the app. Thus, adopting HTTPS to protect the ad traffic will not stop the problem. While previous work [39], [43] highlighted the need for isolating ad libraries largely from the perspective of separating permissions of ad-related code from code of the hosting app, our work in Section V shows that there is also a need to prevent the hosting app from reading the ad library's data when that data is derived from the ad-network's private information.

As the essential core of the mobile advertising ecosystem, ad networks are responsible for protecting users' privacy. Since the above ad isolation techniques have not been widely adopted, ad providers should build defense mechanisms into their products to protect users' privacy. One possible defense could be adding noise or randomness into the personalized results. For example, ad networks could make a larger fraction of their ad deliveries to be non-personalized or contextual ads instead of maximizing personalization of every ad impression. A similar technique has been proposed for the scope of privacy protection in online searches (*e.g.* adding noise into user's search history) [22], and could make it more difficult for an adversary to learn user's personal information and mitigate part (if not all) of the privacy threat we identified in Section V.

Besides adding noise to personalized ads, ad networks may also provide coarser grained targeting options for advertisers. For example, instead of enabling advertisers to precisely target users that are 26 years old, ad networks may only provide a range (*e.g.* 25-34) for targeting. Such approaches may result in coarser granularity of adversary-accessible personal information and decrease the severity of privacy leakage. Google AdMob is already offering ad targeting only for coarse-grained age groups; we encourage other ad networks to adopt a similar model in their targeting offerings.

The idea behind both of the proposed countermeasures is to trade off the quality of ad personalization to limit the degree of privacy leak through such ads. We cannot expect all ad networks to adopt such an approach because less personalized ads may contribute to a loss in ad revenue. We will leave it as an open problem to identify a strategy that can avoid such tradeoff and still work in the current ad-hosting environment (where there is no isolation between the logic/data of the ad-library and the main app).

VII. Related Work

Privacy in Online Advertising. The privacy issues related to online advertising have been the focus of quite a number of studies. For example, Roesner *et al.* [42] showed the prevalence of third-party web tracking and designed a browser extention for defending against social widget tracking. Acar *et al.* [20] studied three advanced web tracking mechanisms - canvas fingerprinting, evercookies and the use of "cookie syncing" in conjunction with evercookies - and suggested that even sophisticated users can face great difficulties in evading tracking. XRay [34] tried to identify how various tracking information is being utilized by targeted ads. Korolova [33] presented attacks that exploit Facebook's advertising system to infer private user information. Barford *et al.* [23] found widespread use of ad targeting mechanisms on the web and showed significant correlation between user interest profile and in-browser ads. Olejnik *et al.* [38] examined the leakage of users' browsing histories through Cookie Matching and Real-Time Bidding. Datta *et al.* [28] explored how user behaviors, Google's ads and Ad Setting interact in the web settings. Finally, Castelluccia *et al.* [26] demonstrated an attack very similar to what we presented in Section V, where the adversary tries to reconstruct user's interest profile from unencrypted personalized in-browser ad traffic of synthesized users.

These works focus mainly on advertising in web pages. In contrast, our work focused on similar issues in an in-app advertising setting. As we have mentioned in the introduction, not only is the personalization of in-app advertisements less understood, it also has the potential to raise more serious privacy concerns, due to the intimate nature of mobile phones. Also, as compared to [26] which requires the adversary to have the capability to intercept the victim's network traffic, the attack we presented in Section V can be carried out by any app on Android.

Privacy-Preserving Advertising. A number of systems have been proposed for privacy-preserving personalization. Privad [31], [40], Adnostic [44], and RePriv [30] achieved this goal by using generalized user profiles and moving ad personalization to the client side. ObliviAd [21] leveraged secure hardware to provide privacy guarantees. Mor *et al.* [36] designed Bloom cookies for encoding a user's profile in a privacy-preserving manner. Hardt *et al.* [32] proposed a differentially private distributed protocol that simultaneously achieves reasonable level of privacy, efficiency and quality in personalization on smart phones. While these proposals protect users' private information or identifiers from being leaked to ad networks, they cannot stop the attack in Section V, which only requires observation of the end results of personalization.

Mobile Ad Personalization. Nath [37] presented MAdScope for harvesting in-app ads and characterizing in-app targeted ads. By studying the keywords/ad-control attributes included in ad requests from different apps, the author found that only one of the top ten in-app ad networks is using behavioral targeting, and demographic information is not commonly used in in-app ads. Book *et al.* [24], [25] surveyed how app developers used ad control APIs to show ads targeting their presumed user population, and studied mobile ad targeting using simulated user profiles and found targeting based on users. In contrast to these two studies, we focused on personalization in the absence of any input from the app developers, and instead of using synthesized user profiles, we harvested ads and demographic information from real users. Our results suggested that a large fraction of ad impressions are correlated with demographic information.

Ad Isolation on Android. Recent studies on isolating advertising from application could provide solutions to the privacy leakage problem we studied in Section V. AdSplit [43] is an Android extension that allows an app and the ad library to run as separate processes with different permissions. AdDroid [39] also separates privileged advertising functionality from host applications on Android. Roesner *et al.* [41] designed LayerCake to support cross-principal embedded interfaces on Android.

VIII. Conclusion

We have studied how user information is utilized by major ad providers for in-app ad personalization on mobile devices and to what extent ad networks know about the user's interest and demographic information. We have also investigated if in-app advertisements can be a channel for leaking user information collected by ad networks to apps hosting these advertisements.

By collecting both the profile and observed mobile ad traffic from 217 real users in a survey, we found that mobile ads delivered by a major ad network (i.e. Google AdMob) are highly personalized based on both users' demographic and interest profiles. Specifically, we showed that more than 57% of ad impressions delivered to 41% of users matched the users' interests. Also, more than 73% of ad impressions for 92% of users were found to be personalized based on users' demographics.

We also demonstrated that personalized in-app advertising can leak potentially sensitive personal information to any app that hosts ads. Specifically, we achieved high accuracies in demographic categories that are explicitly used as targeting options, and showed that information that is not used in serving tailored ads could also be leaked to app developers. These findings illustrate that more protection is needed to defend against privacy leakage in personalized mobile in-app ads.

Acknowledgment

The authors would like to thank the anonymous reviewers and our shepherd, Manuel Egele, for their help and feedback,

as well as our operations staff for their proofreading efforts. This research was supported by the NSF award CNS-1017265, CNS-0831300, CNS-1149051 and DGE-1500084, by the ONR under grant N000140911042 and N000141512162, by the DHS under contract N66001-12-C-0133, by the United States Air Force under contract FA8650-10-C-7025, by the DARPA Transparent Computing program under contract DARPA-15-15-TC-FP-006. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF, ONR, DHS, United States Air Force or DARPA.

REFERENCES

[1] Add demographic targeting to an ad group. https://support.google.com/adwords/answer/2580282?hl=en.

[2] Adopting encryption: The need for https. http://www.iab.net/iablog/2015/03/adopting-encryption-the-need-for-https.html.

[3] Ads take a step towards "https everywhere". http://adwords.blogspot.com/2015/04/ads-take-step-towards-https-everywhere.html.

[4] Amazon mechanical turk. https://www.mturk.com/mturk/welcome.

[5] Google Ads Settings. http://www.google.com/ads/preferences.

[6] Google Ads Targeting Tools. http://www.google.com/ads/displaynetwork/manage-your-ads/targeting-tools.html.

[7] Key terms- privacy & terms - google. http://www.google.com/policies/privacy/key-terms/#toc-terms-sensitive-categories.

[8] Overview of race and hispanic origin: 2010. http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf.

[9] Phantomjs. http://phantomjs.org/.

[10] Privacy policy - privacy & terms - google. http://www.google.com/policies/privacy/.

[11] Reach people interested in your products or services. https://support.google.com/adwords/answer/2497941?hl=en.

[12] Reach people of specific demographics. https://support.google.com/adwords/answer/2580383?hl=en.

[13] scikit-learn: Machine learning in python. http://scikit-learn.org/stable/.

[14] Smartphone os market share, q4 2014. http://www.idc.com/prodserv/smartphone-os-market-share.jsp.

[15] Targeting - admob android guides - google developers. https://developers.google.com/mobile-ads-sdk/docs/admob/android/targeting.

[16] Targeting options - advertisers - mobgold. http://www.mobgold.com/web/advertiser/target-option.

[17] Topic Targeting. https://support.google.com/partners/answer/2769377?hl=en.

[18] Yahoo content analysis api. https://developer.yahoo.com/contentanalysis/.

[19] The Value of Behavioral Targeting. http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf, 2009.

[20] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 674–689. ACM, 2014.

[21] M. Backes, A. Kate, M. Maffei, and K. Pecina. Obliviad: Provably secure and practical online behavioral advertising. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy*, pages 257–271. IEEE Computer Society, 2012.

[22] E. Balsa, C. Troncoso, and C. Diaz. Ob-pws: Obfuscation-based private web search. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy*, pages 491–505. IEEE Computer Society, 2012.

[23] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan. Adscape: Harvesting and analyzing online display ads. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 597–608. ACM, 2014.

[24] T. Book and D. S. Wallach. A case of collusion: A study of the interface between ad libraries and their apps. In *Proceedings of the Third ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, pages 79–86. ACM, 2013.

[25] T. Book and D. S. Wallach. An empirical study of mobile ad targeting. *arXiv:1502.06577*, 2015.

[26] C. Castelluccia, M.-A. Kaafar, and M.-D. Tran. Betrayed by your ads!: Reconstructing user profiles from targeted ads. In *Proceedings of the 12th International Conference on Privacy Enhancing Technologies*, pages 1–17. Springer-Verlag, 2012.

[27] F. T. Commission. Data brokers: A call for transparency and accountability. 2014.

[28] A. Datta, M. C. Tschantz, and A. Datta. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 1(1):92–112, 2015.

[29] A. Farahat and M. C. Bailey. How effective is targeted advertising? In *Proceedings of the 21st International Conference on World Wide Web*, pages 111–120. ACM, 2012.

[30] M. Fredrikson and B. Livshits. Repriv: Re-envisioning in-browser privacy. In *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, pages 131–146. IEEE Computer Society, 2011.

[31] S. Guha, B. Cheng, and P. Francis. Privad: Practical privacy in online advertising. In *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*, pages 169–182. USENIX Association, 2011.

[32] M. Hardt and S. Nath. Privacy-aware personalization for mobile advertising. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, pages 662–673. ACM, 2012.

[33] A. Korolova. Privacy violations using microtargeted ads: A case study. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops*, pages 474–482. IEEE Computer Society, 2010.

[34] M. Lécuyer, G. Ducoffe, F. Lan, A. Papancea, T. Petsios, R. Spahn, A. Chaintreau, and R. Geambasu. Xray: Enhancing the web's transparency with differential correlation. In *Proceedings of the 23rd USENIX Conference on Security Symposium*, pages 49–64. USENIX Association, 2014.

[35] B. Liu, A. Sheth, U. Weinsberg, J. Chandrashekar, and R. Govindan. Adreveal: Improving transparency into online targeted advertising. In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks*, pages 12:1–12:7. ACM, 2013.

[36] N. Mor, O. Riva, S. Nath, and J. Kubiatowicz. Bloom cookies: Web search personalization without user tracking. In *Proceedings of the 2015 Network and Distributed System Security Symposium*. ISOC, 2015.

[37] S. Nath. Madscope: Characterizing mobile in-app targeted ads. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 59–73. ACM, 2015.

[38] L. Olejnik, T. Minh-Dung, C. Castelluccia, et al. Selling off privacy at auction. In *Proceedings of the 2014 Network and Distributed System Security Symposium*. ISOC, 2014.

[39] P. Pearce, A. P. Felt, G. Nunez, and D. Wagner. Addroid: Privilege separation for applications and advertisers in android. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, pages 71–72. ACM, 2012.

[40] A. Reznichenko and P. Francis. Private-by-design advertising meets the real world. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 116–128. ACM, 2014.

[41] F. Roesner and T. Kohno. Securing embedded user interfaces: Android and beyond. In *Proceedings of the 22nd USENIX Conference on Security Symposium*, pages 97–112. USENIX Association, 2013.

[42] F. Roesner, T. Kohno, and D. Wetherall. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, pages 155–168. USENIX Association, 2012.

[43] S. Shekhar, M. Dietz, and D. S. Wallach. Adsplit: Separating smartphone advertising from applications. In *Proceedings of the 21st USENIX Conference on Security Symposium*, pages 553–567. USENIX Association, 2012.

[44] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas. Adnostic: Privacy preserving targeted advertising. In *Proceedings of the 2010 Network and Distributed System Security Symposium*. ISOC, 2010.