An Experiment on Facebook Ads for Testing Sensitive Information Flow

Jagannathan Chakravarthy and Hsin Miao

{*jchakrav,hsinm*}*@andrew.cmu.edu*

***Abstract*—We explored the advertisements served on Facebook by extending the AdFisher tool to collect advertisements from Facebook and performed statistical tests on the data obtained to draw conclusion on whether Facebook ads are in compliance to their privacy policy. We investigated this by conducting three experiments in which we tried to gather ads from newly created profiles, collected ads from existing Facebook profiles and collected ads from profiles after exposing them to a particular treatment. We enumerated the results and also elaborated on the problems encountered along the path which hindered scaling up the experiments to obtain statistically significant results. We also presented future steps that can be taken to mitigate some of the problems encountered.**

I. Introduction

THE effectiveness of ad campaigns have inflated due to targeted marketing, which has been made possible through websites which track user data and thereby predict a user's interests. This raises a lot of concerns for the user, who would like to know which set of personal information they provide is used for targeted advertising. For this reason, the websites that acquire or collect the personal information from users tell them which set of this information will be used for targeting and which set is not used through their privacy policy. However, we cannot directly determine whether these policies are complied. They can be ascertained by conducting indirect experiments such as analyzing the ads that the user encounters on the internet.

Google and Facebook are two of the largest ad providing platforms. Our work is motivated by Datta et al. [1] in which experiments were conducted on Google ads. They found some correlations between the user provided data, the websites the user visited and the advertisements they see. The result implies possible violations of Google's privacy policy with respect to use of sensitive information for targeted advertising. In our project, we adopted a similar approach given in [1], and our goal is to ascertain if Facebook ads are in compliance to their privacy policy/guidelines. We extended the tool *AdFisher* developed by Datta et al. [1] so that it can automatically login using Facebook accounts, and collect the ads shown on Facebook News Feed.

In the following sections, we will first describe the extension of AdFisher, and then several experiments will be stated. Finally, we will discuss the results and possible future works in the project.

II. Related Work

Analysis on Google Ads has been extensively studied by many researchers. Our work was inspired by [1] in which a tool named AdFisher is used to collect Google ads from different websites by exposing the user to different treatments and analyzing how the treatments correspond to the Ads shown for the users. They conclude that Google Ad-settings is opaque about some features and transparent about other features and these variances lead to seemingly discriminatory ads. Their main results were that when a user visits websites related to substance abuse, he gets significantly different ads as compared to a user who does not visit these websites although the Google Ad-settings page does not show any changes reflecting to the user's interest. Also, when male and female users are exposed to the same treatment, they found that male users get significantly more ads related to high-paying jobs than their female counterparts.

III. Extension on AdFisher

In this section, we will describe how to extend AdFisher to Facebook. There are two modules: account authentication and advertisements collection. By using these functions, we can easily analyze the ads on Facebook by AdFisher.

A. Account Authentication

The module will first visit Facebook (www.facebook.com), and the login by typing the Facebook account and password. We obtained the goal by using Selenium [2], the python API used in AdFisher.

First, we locate the text area that is for account, and then type the account by the using this function:

```
email=driver.find_element_by_name("email")
email.send_keys("ACCOUNT")
```

Second, we use the same way to insert the password into the password area. In order to really login into Facebook, the "Log In" button on the webpage or the key *return* on the keyboard should be clicked. We use this code `Keys.RETURN` in Selenium that simulates a "click" action of the keyboard to login.

B. Advertisements Collection

In the section, we will describe how we collect the ads from Facebook. Figure 1 is an example of a Facebook Ad. The ad is located in a container where the class name of the container object is *"ego_unit"*. The class name of the title of an ad is *"_5vwh"*; also, the class name of the hyperlink to the ad is *"_4xvg"*; moreover, the class name of the ad content is *"_5vwk"*. In the example ad, the ad title is "SPANX.com Official Site", the hyperlink is "spanx.com", and the ad content is "Shop the ... Free Shipping!"

By analyzing the HTML attributes, we can use Selenium to collect the ads by using the procedure below:

```
find_elements_by_class_name("ego_unit")
find_elements_by_class_name("_5vwh")
find_elements_by_class_name("_4xvg")
find_elements_by_class_name("_5vwk")
```

After collecting the information including title, hyperlink, and the content of an ad, we store it into the log the file by using the format that AdFisher uses for further analysis.

IV. Issues on new Facebook accounts

We will discuss the relationship between a newly created Facebook account and the ads on its News Feed in the section. In the experimental design in our proposal, we would like to create eight accounts and divided them into two groups: control group and experimental group. After different treatments on these two groups, we would like to determine whether there is a statistical difference between the two sets of ads collected by these two groups.

Fig. 1. An example of Facebook ad.

However, we found that there is no ads in a new Facebook account. In contrast, there are two ads in our personal Facebook accounts. If we analyze the ads on our existing Facebook accounts, there must be many confounding factors that affect the results of the experiments. To be more specific, we cannot conclude that the treatments such as visiting shopping websites in our experiments really affect the collected ads because of those confounding factors. In order to ensure the causality of our experiments, we first investigated how to make a new Facebook account show an ad.

The first experiment is to investigate the relationship between the ads and the followed fan pages. Our hypothesis is that the number of followed online-shopping-fan-pages will contribute to online shopping ads on News Feed for a new Facebook account. Because Facebook does not provide the top list of its fan pages, we visited *alexa.com* [3] and followed the fan pages of top 50 online clothing shopping websites.

In the experiment, the account followed a fan page from the list obtained from *alexa.com*, and then collected the ads for twenty reloads. We sequently followed the fan pages by the rank on the list. However, we did not obtain significant results in the experiment. In fact, we did not obtain any ads after following 50 online clothing shopping fan pages. The result shows that following fan pages merely is not enough to show the ads. It implies that there must be other factors that affect the ads on the News Feed for a new Facebook account.

TABLE I
THE P-VALUES FOR DIFFERENT SPLITFRACT VALUES. DIFFERENCE OF ADS ARE MEASURED FROM DIFFERENT FACEBOOK ACCOUNTS.

splitfrac	p-value
0.1	0.250191
0.2	0.062656
0.3	0.015665
0.4	0.003931
0.5	0.00196

V. COMPARISON OF TWO ACCOUNTS

In this experiment, the Facebook accounts of Jagannathan and Hsin were used to collect ads. This experiment was done to know if the ads served to the two users were inherently different. For treatment 1, Jagannathan's profile was used and for treatment 2, Hsin's profile was used. In both the treatments, the users were logged in to Facebook and the ads were collected immediately. The experiment was run for 20 blocks with 10 reloads in each block. The total number of ads collected for each user was 392 and 400 for Hsin and Jagannathan respectively. These ads were subjected to analysis and p-value for this experiment was found for different splitfrac values. Splitfrac value determines what percentages of the samples should be used for training and testing. The results are shown in Table I. The p-value less than 0.05 is considered a significant value. These results hint at the possibility that the two users were indeed served different ads.

VI. THE EFFECT ON ONLINE SHOPPING WEBSITES

In this experiment, the Facebook account of one user was used to collect ads after exposing that user to online shopping (clothing) websites. This was done to determine whether there is a correlation between the ads that are served to the user and the user's online activity. The experiment was set up such that the ads were collected after the user has been exposed to websites related to online shopping in one treatment and in another treatment, the user remained idle and then the ads were collected from Facebook. The experiment was run for 15 blocks with 10 reloads in each block. The total number of ads that were collected for the two treatments were 303 and 296 respectively . These ads were subjected to analysis and p-value for this experiment was found for different splitfrac values. The results are shown in Table II.

TABLE II
THE P-VALUE FOR DIFFERENT SPLITFRACT VALUES. DIFFERENCE OF ADS ARE MEASURED FROM TWO TREATMENTS.

splitfrac	p-value
0.1	0.499342
0.2	0.249891
0.3	0.06223

This indicates that the online browsing activity might not have any immediate effect on the ads served to the user on Facebook. The reasons for this might be that the number of websites to which the user was exposed to might not be big enough to create a significant impact on the ads served to the user, or that the time needed for the browsing activity to affect the ads served might not be immediate.

VII. DISCUSSIONS

Several problems were encountered while collecting the data for the first part of the experiment in which the advertisements from Facebook were analyzed to ascertain if Facebook ads are in compliance to their privacy policy/guidelines. The list of problems is given below:

First, in the experiments done based on Google ads, no user creations were required. For our experiment, ads could not be collected without creating profiles on Facebook. This was a major restriction since many profiles had to be created to run a single experiment.

Second, several new accounts were created but none of them were provided with any advertisements. To improve the activity of these profiles, several Fan Pages were "Liked" and "Followed" and friends were added. Also several websites were visited while the user was logged into Facebook. None of these activities resulted in producing advertisements on the new users' Facebook page. This highly restricted the scalability of the experiment since the experiments had to be performed on existing users who had advertisements on their Facebook page.

Third, unlike Google, Facebook does not provide the option of viewing what knowledge Facebook has

acquired about the user. Therefore, it is impossible to know what information Facebook has gathered based on users' activities. Also, unlike Google, there is no option to opt-in or opt-out of interest based advertisements and no option to view or edit the interests of a particular user. This restricted the use of same profiles for different experiments.

To get significant results and to scale up the experiments, there are some possible ways to show ads on new Facebook accounts. For example, the visit of online shopping websites might lead to online shopping ads in new Facebook accounts. After investigating the ways to show the ads on News Feed for new accounts, further experiments can be conducted to examine whether Facebook complies its Advertising Guidelines. In addition, the next step would be to scale up the experiment for the comparison of different accounts in which ten browsers will be used to collect advertising data over 100 blocks so that a significant number of ads are gathered. The experiment would be set up in such a way that five browsers will login as one user and the other five as another user. Moreover, the next step for the experiment that finds the relationship between online shopping websites and ads would be to increase the online activities of the users by increasing the number of websites to be visited. Also, the measurement can be scaled up so that more ads can be collected to determine if there is any significant difference in the ads served based on the online activity of the user.

VIII. Conclusion

In this project we extended AdFisher on Facebook advertisements. By our implementation, advertisements on Facebook can be easily collected by AdFisher. In the experiments we found that Facebook provides targeted advertisements to different users. However, we were not able to scale up the experiments because Facebook does not provide advertisements for newly created users. This limitation restricted us from obtaining significant results. To get more insight on the problem, further works are required to examine whether Facebook complies with their privacy policies.

References

[1] A. Datta, M. C. Tschantz, and A. Datta, "Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination," *arXiv:1408.6491*.

[2] SeleniumHQ, "http://www.seleniumhq.org/,"

[3] Alexa, "Top Sites in Shopping, Clothing. [online]. available: http://www.alexa.com/topsites/category/Top/Shopping/Clothing,"

[4] Facebook, "Facebook Advertising Guidelines [online]. available: https://www.facebook.com/ad_guidelines.php,"