A Research Study on User Concerns Regarding Online Advertising

Lalit Agarwal
DA-IICT, Gandhinagar
200901159@daiict.ac.in
Supervisor
Prof. Anish Mathuria
Off-Campus Supervisor
Dr. Saurabh Panjwani, Dr. Sharad Jaiswal
Bell Labs, Alcatel-Lucent
Manyata Embassy Tech Park Silver Oak (Wing A),
Outer Ring Road, Nagavara,
Bangalore - 560045

***Abstract*— Recent studies have highlighted user concerns with respect to tracking of users' online activity to deliver targeted ads and the need for better consumer choice mechanisms to address this phenomena. We re-investigated this issue and found that although concerns for tracking of web-browsing behavior remain strong, other aspects of online advertising like the possibility of being shown ads with embarrassing and sensitive content also upset users as much and are often voiced as greater concerns than the concern of being tracked. Current-day blocking tools are insufficient to redress the situation: users demand selective filtering of ad content (as opposed to, say, blocking out all ads). In order to address this situation, we developed a modified version of ad-block plus tool which blocks only embarrassing ads in contrast to blocking all ads.**

***Index Terms*— online advertising, third-party tracking, privacy, embarrassment, ad-blocking tools**

I. INTRODUCTION

Advertising companies today on the web use a technique called online behavioral advertising (OBA) through which they show ads which match user's interests. It is conducted by using third-party tracking which means that browsing pattern of a user is collected using a third-party file called a cookie stored on the user's machine which is used to show relevant ads to the users. The delivery of tailored ad content from the advertisers on the publishers' sites has allowed greater interactivity between the consumers and the advertisers. However, recent studies have highlighted user concerns with respect to third-party tracking and online behavioral advertising (OBA) and the need for better consumer choice mechanisms to address these phenomena. Users are also concerned about ads containing inappropriate contents being shown to them. Current-day blocking tools are insufficient to redress the situation: users demand selective filtering of ad content (as opposed to, say, blocking out all ads) and are not satisfied with mechanisms that only control third-party tracking and OBA. As part of the study, we try to re-examine the privacy concerns related to third-party tracking and online advertising especially concerning embarrassing ads among the users and propose a design which suits their needs.

II. PROJECT OBJECTIVES

- Understanding the concept of online behavioral advertising (OBA) and how cookies are used by ad networks to provide customized ads to users.
- Developing an ad-extraction tool to identify and extract online ads from a given list of web-pages.
- Studying the question of effectiveness of OBA vs. other forms of advertising on the web and re-investigate through a user study the question of user attitudes towards third-party tracking and online advertising.
- Building a modified version of the existing Adblock Plus tool to block embarrassing advertisements only in contrast to blocking all the advertisements and therefore still ensuring some of the benefits of OBA.

III. WHAT IS ONLINE BEHAVIORAL ADVERTISING?

Online Behavioral advertising or OBA is a modern technique in online advertising which is used by advertisers to tailor ad content to users based on their past browsing habits. A smart way of identifying the target pool for a product is to use information that a customer appeared to be trying to find information about that product elsewhere. This allows advertising companies to deliver targeted ads to the users thereby increasing their revenue. As a result, the advertisers pay for few impressions of the ads and enjoy a higher conversion rate. According to the results by Search Engine Watch, only eight percent of all online advertising is behaviorally targeted.

IV. THIRD-PARTY TRACKING

Third-party web tracking refers to the practice by which a third-party entity i.e. an entity other than the website which the user has visited, keeps a track of the user's visit to various websites. Third-party tracking is highly prevalent on the web

today. It provides desirable functionality for the website owners by providing personalization, site analytic, and targeted advertising. The trackers are able to gather a large browsing profile about a user and thus are able to provide better service to its customers (the embedding websites) and to the user himself (e.g., in the form of personalization). Third-parties generally include advertising networks (Doubleclick), analytic companies (Google Analytic), social networks (Facebook).

A. How does it work?

Third-parties track users by storing a small file called cookie on their machine. A cookie contains a unique identifier which uniquely identifies the user's browser. Cookies are set either by HTTP responses having a "Set-Cookie" header or by scripts running in the page. When a page contains content from a third-party, a third-party cookie is usually stored by the third party while delivering the content. These cookies contain a unique identifier which uniquely identifies the user's browser and keeps track of the websites visited by him.

B. What information do they collect?

The data collected by the third-parties to be used for behavioral advertising is not related to the personal information of the users. Third-parties claim that they don't collect user's personal information like email address, contact number etc. They only keep a track of the websites visited by the user using the cookie stored on the user's machine. However, a user's browsing pattern can reveal a lot of sensitive information about them like their marital status, sexual orientation, medical conditions etc.

C. Current Policies on tracking

In 2010, United States Federal Trade Commission (FTC) proposed a "Do Not Track" mechanism which requests web-application to disable its tracking of an individual user. The online advertising industry- Network Advertising Initiative (NAI), Digital Advertising Alliance (DAA) and the Interactive Advertising Bureau Europe, IAB allow users to opt out of behavioral advertising by displaying an opt-out icon alongside an ad. However, collection of user data remains unaffected.

V. Third-party tracking and OBA

The popularity of OBA is growing and it is being increasingly discussed in both industrial and academic circles because of its privacy implications. OBA is normally implemented by web third parties who partner with different websites, track information about individual visits to these sites and use this information to create browsing profiles of the visitors. The more websites a third party partners with, the more information it gathers about the sites' visitors and the better profiles it can create of individual users. This information is subsequently used by third parties to channel advertisements to users based on the profiles it created for them. The most popular third party which does this kind of "cross-site" tracking today is DoubleClick, a fully-owned subsidiary of Google.

VI. User concerns related to online behavioral advertising

As web browsing history is inextricably linked to personal information, third-party tracking has become a topic of increased public debate. OBA provides both benefits and downsides to users. If the user's interests have been accurately profiled, he/she will receive more relevant advertising. However, collecting data about users' online activities can potentially infringe their privacy. According to a 2009 study by Bleakley et. al [1], if given a choice, 68% of Americans definitely would not allow advertisers to track them online even if their online activities would remain anonymous. A 2012 study by Ur et al. [2] participants had strong concerns about data collection, and the majority of participants believed that advertisers collect personally identifiable information.

The continued long-term display of such annoying/ embarrassing ads may exert negative effects on the publisher, the user, and the advertiser. Firstly, annoying ads can exert negative effects on publishers by creating its bad reputation among the users. Apart from the user abandonment effects, annoying ads might signal that the website on which the ad is placed on lacks stability, reputability, or safety. Secondly, annoying ads can exert a negative impact on users by getting in the way of the user consuming the publisher's content, undermining the very reason that brought them to the site.

VII. User Study

We investigated the question of user attitudes towards third-party tracking and online advertising, studying it in the context of 53 web users in India. We used one-on-one, in-depth interviews to understand user attitudes towards OBA and the perceptions towards OBA by situating it in the context of how online ads, in general, are perceived by users. This enabled us to evaluate concerns for OBA relative to other user concerns in the realm of online advertisements. We find that embarrassing ads tend to cause greater concern to the users than third-party tracking, and furthermore, knowledge of the latter and its effects can heighten user concerns for embarrassment.

A. Methodology

The majority of our interaction with users was via a one-on-one, semi-structured interview centered on online advertising and OBA. We used a power-point slide deck to give the users a brief idea of third-party tracking and OBA through examples before asking them their views about it. We also included a brief (2-minute) tutorial on how to control behavioral advertising and told them about ad-block plus tool. All interviews were audio-recorded and later transcribed. Each participant provided written, informed consent.

B. Participant Sample

Our sample was gender-balanced (26 F, 27 M) and we took care to sample a mix of people from both technical (bachelor's degree in engineering) and non-technical back-grounds (25 technical, 28 non-technical). The age range was 22 to 42, with a mean age of 30.7. They were all active web users, reporting

to be spending between 1 to 8 hours (mean 3.4 hours) browsing on a personal PC every day. A majority of our participants (89% of all females, 55% of all males) reported to be viewing ads regularly, a size-able number reported having clicked on them intentionally (67% F, 37% M) and a noticeable fraction reported having converted a click into a purchase (17% F, 11% M).

VIII. Findings

A. Concerns about third-party tracking

We find that users in our study were quite concerned about third-party tracking and OBA but their concerns were largely centered on a fear of personally-identifiable information and financial data being lost to third parties. Outside of this fear, users were indifferent to tracking and often, report positively towards it. Participants had a neutral attitude towards third-party tracking with only 25% of participants opposing the idea. The majority of the participants voiced resistance to being tracked on email and banking websites, financial investments and adult content websites.

B. Perception of OBA ads

Participants' attitudes towards being shown OBA ads were, in general, more positive than those towards third-party tracking. More than 75% of them said that they would like to be shown such ads. The ad-categories on which many users wanted ads were travel and tourism(66%), apparel (51%) and Arts and entertainment (49%). One concern that was voiced several times by many participants (70%) was with respect to the repetitiveness of OBA ads. Participants reported being annoyed by OBA ads shown to them repeatedly, sometimes even long after they had made a purchase through the ad.

C. Concerns about embarrassing ads

More than the issue of third-party tracking, users were deeply concerned and sensitive about the content they get exposed to in online ads. A majority of the users in our study reported past experiences of being shown ads with embarrassing and suggestive content which had upset them. We were surprised to find that a majority of our participants (39/53, i.e. 74%) had experienced situations in which they were shown online ads which they perceived as carrying embarrassing content.

While some stated having been embarrassed by ads when browsing in private, the majority reported instances in which the embarrassment was caused by being in the vicinity of other people. When we probed participants about the nature of websites on which they have observed sensitive ads, six of the participants (less than 15%) reported to be seeing such ads only on websites with pirated content or otherwise interpreted by them as "suspicious" sites. In contrast, more than 60% claimed to have seen embarrassing ads even on "normal" websites which they categorized under email, news or video-streaming sites. Users in our study reported to have encountered embarrassing ads even on certain "good" websites (e.g., email sites) and claimed that they continue to visit such sites for their information needs. Those in our study were fairly consistent in defining embarrassing ads as graphic ads that either contain sexually explicit content or information on online dating etc.

Among the participants who had experienced embarrassing ads, their concern towards being shown such ads was markedly greater than that towards being shown OBA ads, or even towards third-party tracking. Nearly all participants in this category stated to be more worried about being shown embarrassing ads than about tracking or OBA. We asked participants to provide us a qualitative concern rating on a 5-point Likert scale to the three possibilities- being tracked by third parties, being shown OBA ads and being shown embarrassing ads. The results are depicted in Fig. 1.

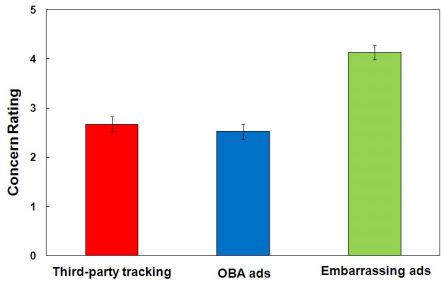

Fig. 1. Mean participant concern ratings for being tracked by third parties, being shown OBA ads and being shown embarrassing ads.

D. Towards an end-user tool

While much emphasis has been given to the issue of privacy in behavioral advertising in prior work, our study suggests that this may not be the issue that web users are most worried about today (within the realm of online advertising). A large number of users in our study reported being more concerned about seeing embarrassing advertisements online than about their browsing history being tracked by third parties, which means that at least in some geographies, embarrassment from online ads is a matter of significant user concern.

We envision an end-user browser plugin tool which enables users to achieve two objectives:

- Selective tracking control i.e., disabling tracking on embarrassing websites which may be matrimony, dating, adult websites.
- Selective ad blocking i.e., blocking out embarrassing advertisements which contain inappropriate/sensitive content which may make the user uncomfortable.

IX. Ads Extraction Tool

The next step was to develop an ad-extraction tool which would identify and extract the contents of both image and textual advertisements present in a given web-page. The main objective of this tool was to identify the ads being shown to the users and analyze the content of the advertisements for the study. During the user study, we found that the users were quite concerned about being shown advertisements with

embarrassing content which included adult sites ad, dating and matrimony ads. We wanted to know that how many such advertisements were actually being shown to the users. Also we wanted to identify the common features among these embarrassing ads to build a filter list which could be used to block them.

A. Tools Used

In order to perform the extraction of ads, the following tools and algorithms were used-

- **Selenium Web driver**- It is a browser automation tool which automates web application for testing purposes and allows running automated scripts to perform various tasks. The Firefox web driver was used which automates the Firefox browser. Using the web driver, the web-page was loaded and was parsed to retrieve the advertisements present in that web-page. The advertisements which were inside iframes loaded using JavaScript which took some time to load and therefore slowed down the execution. (`https://code.google.com/p/selenium/wiki/FirefoxDriver`)
- **Jsoup Parser**- In order to parse the HTML content fetched by the Selenium web driver, Jsoup parser was used. It is a java library which parses the HTML content and allows extracting and manipulating data using DOM, jquery-like methods. (`http://jsoup.org/`)
- **Aho-Corasick** string matching algorithm was used for comparing the URLs with a list of ad-keywords to check if a given link was an advertisement or not.
- Also the **Easylist's** list of advertisement filters which is also used by many Ad-block plus users was used to identify links which were ads from all the links. Two lists were used for the same.
  - General advertising keyword list- It contains a list of general advertising keywords. It was used to see if a given link is an ad or not.
  - Third-party network list- It contains a list of third-parties which are involved in delivering advertisements across the web. This list was used to check if the advertisement was from a third-party network or not.

B. Extracting the advertisements

As there could be many links inside a HTML page, in order to identify the links which are ads, the URLs are compared with a list of ad-keywords to identify if a particular link is an ad or not. We used the list of filters from the Easylist subscription which is also used by most of the Adblock Plus users. In order to compare all the links with the list of filters, Aho-Corasick string matching algorithm is used. These are the steps involved:-

1) Using Selenium's Firefox web driver, the web-page from which the ads are to be extracted is fetched. The Selenium web driver automates the Firefox browser i.e. creates an instance of the browser and waits for the page to load and allows JavaScript to execute if required.
2) Once the page loads completely, the HTML content of the page is fetched from the browser and is parsed using Jsoup to look for all the links present in the webpage.
3) **For ads outside iframes**, once the page completely loads, we parse the HTML content of the page using Jsoup to look for anchor tags. For each anchor tag, it compares the href link in that anchor tag with the list of advertisement filters to see if that link is an ad or not. If it is an ad, it stores the link in a file.
4) **For ads inside iframes**, once the page loads the content of iframes, the web driver identifies all the iframes in the HTML page and compares the source links of the iframes to the advertisement filter list using the Aho-Corasick string matching algorithm. If the string match occurs, the iframe content is parsed to look for anchor tags and then the links in these anchor tags are compared with a list of ad filters. If the string match occurs again, it is considered to be an ad and the link is stored in a file.
5) In order to get the content of these ads, we process HTTP get requests on the links collected and extract the meta content, title of the landing web-pages from the HTTP response to identify the content of these advertisements.

C. Methodology

We ran the ad-extraction for around 5,000 web pages and looked for advertisements in these web-pages. The data used for this part included the web browser history of the participants which were collected during the user study. We were able to get around 50,000 URLs from 28 users. We randomly selected 3 sets of URLs from these 50,000 URLs each containing 500, 2500 and 5,000 URLs respectively and ran our test on these URLs to look for embarrassing ads. Multi-threading was used to run five Selenium Firefox driver instances simultaneously for faster execution. The tests resulted in an excel file containing the following information for each ad-

- Ad-Title- The ad title displayed for text ads
- Ad-Content- The ad content displayed for text ads
- Ad-Display URL- The ad URL displayed for text ads
- Ad-Src URL- Source URL of the ad, extracted from the HTML code of the ad
- LPTitle- The title of the landing page of the ad
- LPURL- The URL of the landing page of the ad
- URL- The URL of the web-page on which this ad was present
- ImgSrc- Source address of the image, if it was an image ad

Apart from the excel file we also downloaded the HTML pages and images of these ads for doing further analysis. The content of these excel files were scanned to identify embarrassing textual ads and a separate list of such ads was created using the excel files. For image ads, we individually browsed through the images which we downloaded to see if they could be embarrassing. This resulted in a list of embarrassing ads

present in those pages which we used to analyze and build a filter for blocking embarrassing and sensitive ads.

D. Results

We identified embarrassing/sensitive ads from the ad data we collected by running the ads-extractor tool. We considered an ad sensitive if it belonged to any of the following categories- dating, matrimony, nightwear and adult sites. We were able to extract 4014 text and 2321 image ads out of which, we found 73 (2%) embarrassing text ads and 51 (2%) embarrassing image ads. The majority of the embarrassing ads were either dating ads or matrimony ads. We also looked at their source URLs, landing page content to build a list of filters to block such ads.

E. Limitations of the tool

Limited tools are available on the internet which automates a browser and allows iframes to load using JavaScript. We tested some headless browsers tools but they failed to give appropriate results. Due to the absence of any another alternative, we had to use Selenium's Firefox web driver. Since each instance of the web driver opened a new Firefox window, it slowed down the process when working with multiple URLs. Also since targeted ads are usually loaded using JavaScript, the Firefox driver waited for the entire page to load and then only parsed the HTML page which added to the delay. Also the tool failed to extract flash ads because since they are embedded objects, it is not easy to identify if they contain an ad or not.

X. Ad-blocking tools

There are several tools which are freely available which gives users an option to control both third-party tracking and block ads being shown to them. One such tool is Ad-block Plus which is one of the most popular ad blocking and content-filtering extension available for Firefox, Chrome and other major browsers. It blocks all ads including textual, image, flash ads using a list of filters. According to ABP's official website, only 25% users reported that they use it because they want to view zero ads. Many users have reported to be using it with the motivation for blocking annoying and embarrassing ads.

A. How Ad-block Plus works?

Ad-block Plus acts like a proxy between the browser and the web server. It monitors the HTTP request being sent to the web server and blocks HTTP requests based on the source addresses of the request. A Javascript object called content policies gets called by Ad-block plus whenever there is something to be loaded in the browser. It checks the address and matches it with a list of ad-blocking filters to decide whether it should be allowed to load.

The file "contentPolicy.js" contains an implementation of the nsIContentPolicy XPCOM interface for controlling loading of various types of content. This allows the HTTP requests being sent to be filtered if they match the filter list.

Whenever there is a HTTP request being sent, the shouldLoad() function of the nsIContentPolicy interface calls the processNode() function. The processNode() function checks if the address of the request belongs to a whitelist filter. If it belongs to the whitelist filter it doesn't block the requests. Otherwise it calls the matchesAny() function present in "matcher.js" which checks if the address matches with any of the items in the filter list. If a match occurs, the processNode() functions returns "false" to the calling function shouldLoad() which then simply blocks the requests from being sent.

B. Limitations of Ad-Block Plus

While these tools provide a solution to the privacy concerns raised by users related to OBA, if most of the users start using them, there could be serious consequences for the advertising companies, publishers and also impact the users indirectly. It could lead to economic loss for these advertising companies. Also websites, blogs who get their revenues from these advertising networks may eventually terminate their services. At the same time, users who would like to keep themselves updated with the new products of their choice would be at loss if they block all the ads.

Though, considering the rising concerns related to OBA, these tools play an important role but none of them offers a topic based control which is the need of the hour. There is a need for a tool which provides selective blocking of ads and selective control of tracking instead of blocking all the ads.

C. Perception of Ad-blocking tools

By and large, participants responded favorably to the idea of Ad-block Plus but a few were also displeased by the sheerness with which it restricted ad consumption. Only three participants in our sample reported to have used the tool prior to the study. Among the remaining, the majority (62%) expressed an interest in using the tool. Two participants were explicit in stating that their main motivation to use the tool was to eliminate embarrassing ads.

XI. Ad-Filter- A Modified version of Ad-Block Plus

More than half of the participants in our study who were disinterested in Ad-block Plus said that they lacked interest not because they did not want to see ads blocked, but because they wanted to stop irrelevant and embarrassing ads only, something that Ad-block Plus still does not have good controls for. We modified the Ad-block Plus tool to block embarrassing/sensitive advertisements only in contrast to blocking all the advertisements. For this we had to create our own filter lists which only blocked inappropriate advertisements and also had to do modify the existing Ad-block Plus code.

A. Modifications done to Ad-Block Plus

- *Creating a new filter list:* We created a new filter list consisting of only embarrassing website URLs in contrast to including all the advertisement URLs. Our findings from the user study reveal that participants were fairly

consistent in defining embarrassing ads as graphic ads that either contain sexually explicit content or information on online dating. The list included URLs of websites belonging to the following categories- dating, matrimony, nightwear, adult sites. We also referred to alexa.com, ranker.com and many other websites on the internet to create a comprehensive lists of such websites. We were able to create a list with around 200 unique websites for each category.

- *Allowing iframes and javascripts to load:* Since Adblock plus blocked HTTP requests, the requests for loading any iframes and javascripts were also blocked. In order to allow these elements to load, we checked if the request was being sent to load an iframe or a javascipt and if it was true we allowed it to load. This allowed the content of the iframes and Java-scripts to be loaded in the browser. Another modification done was to get the source URLs for the image ads. Usually for image ads, a request is sent to the server to fetch the images, the address on which the HTTP requests are sent is the image address and not the ad source URL. We checked if the request was being sent for an image element and if it was so we got the ad-source URL by accessing the parent node's href attribute.
- *Sending XMLHttpRequest to get the final page of the ad:* In most of the cases the landing page URL is present inside the source URL in a parameter called as adurl. In such cases, we are able to filter ads from the source URL itself. For other cases, where the landing page URL is not present inside the source URL, we send a HTTP get request on the ad-source URL to get the landing page URL and then do a matching on the landing page URL. This allows us to fetch the final URL of the ad and check if it belongs to any of the sensitive categories.
  We used the XMLHttpRequest (XHR) API for the same which sends HTTP requests directly to a web server and fetches the server response data directly back into the script. The "Location" field in the response header contains the landing page URL.

B. Limitations of the tool

The tool contains a predefined list of URLs belonging to four categories- dating, matrimony, nightwear, adult sites. Though we have tried to build a comprehensive list, still there could be some sensitive sites which may not belong to this list and hence will not be blocked. In such a case, the user will have to add that filter to the list using the tool. Also the tool only blocks embarrassing image ads. It does not work for flash and text ads.

C. Future Work

An important area of future work is to give topical preferences to the users where they can choose the categories from a list for which they want ads to be blocked. We also plan to extend the category list from these four topics to a wider range of topics which include categories which users might want to block like politics, religion etc. We also plan to add a module allowing third-party tracking to be controlled on selective websites. Due to time constraint it was not possible to implement these features and they are a part of our future plan.

XII. Conclusion

While much emphasis has been given to the issue of privacy in behavioral advertising in prior work, our study suggests that this may not be the issue that web users are most worried about today. A large number of users in our study reported being more concerned about seeing embarrassing advertisements online than about their browsing history being tracked by third parties, which means that at least in some geographies, embarrassment from online ads is a matter of significant user concern. Tools like Ad-Block Plus have a singular focus on ad blocking and control third-party tracking as a secondary objective. We believe that there is a need of a tool like Ad-Filter which selectively block ads.

References

[1] J. Turow, J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy. Americans reject tailored advertising and three activities that enable it. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214

[2] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In Proc. SOUPS, 2012.

[3] http://adblockplus.org/en/faq_internal

[4] F. Roesner, T. Kohno, and D. Wetherall. Detecting and Defending Against Third-Party Tracking on the Web. In Proc. of NSDI, 2012

[5] J. Mayer and J. Mitchell. Third-party web tracking: Policy and technology. In Proc. of IEEE Symposium on Security and Privacy, 2012.

[6] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In Proc. SOUPS, 2012.