Understanding What They Do with What They Know (Short Paper)

Craig E. Wills
Worcester Polytechnic Institute
Worcester, MA USA
cew@cs.wpi.edu

Can Tatar
Worcester Polytechnic Institute
Worcester, MA USA
can@cs.wpi.edu

ABSTRACT

This work seeks to understand what “they” (Web advertisers) actually do with the information available to them. We analyze the ads shown to users during controlled browsing as well as examine the inferred demographics and interests shown in Ad Preference Managers provided by advertisers.

In an initial study of ad networks and a focused study of the Google ad network, we found many expected contextual, behavioral and location-based ads along with combinations of these types of ads. We also observed profile-based ads. Most behavioral ads were shown as categories in the Ad Preference Manager (APM) of the ad network, but we found unexpected cases where the interests were not visible in the APM. We also found unexpected behavior for the Google ad network in that non-contextual ads were shown related to induced sensitive topics regarding sexual orientation, health and financial matters.

In a smaller study of Facebook, we did not find clear evidence that a user’s browsing behavior on non-Facebook sites influences the ads shown to the user on Facebook, but we did observe such influence when the Facebook Like button is used to express interest in content. We did observe Facebook ads appearing to target users for sensitive interests with some ads even asserting such sensitive information, which appears to be a violation of Facebook’s stated policy.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

Keywords

Online Advertising, Privacy

1. INTRODUCTION

There has been much work measuring the extent to which third-party advertisers are in a position to observe and correlate user behavior across a broad range of first-party Web sites [14, 1, 12]. Other work has shown that these advertisers are not only in a position to infer information about users, but directly obtain known information about users via social networking and other types of sites [11, 13, 10]. However an important question that has gone largely unanswered, which we seek to investigate in this work, is understanding what “they” (the advertisers) actually do with this information available to them.

We are aware of two previous studies that are relevant to this question. The first gathered information on text ads shown to automated users visiting different first-party sites [9]. The researchers were able to determine that different ads were shown based on different sites visited, but they did not examine the nature of these differences. The second performed a controlled study to measure behavioral targeting for four training topics on text ads as a basis for evaluating the effectiveness of privacy tools designed to limit this targeting [2].

Our work makes a number of contributions. First, we not only examine how advertisers use behavioral information in serving ads, but take a more comprehensive approach to see if and how this information is combined with location and personal characteristics of a user. Second, we introduce a variety of sensitive topics in our testing to understand how they are handled by advertisers. Third, we examine not just text ads, but also those shown as images or Flash.

In addition to what ad networks reveal about what they know via the ads they display, some have provided transparency for users to see what inferred demographics and interests are being associated with their browsing behavior. These Ad Preference Managers (APMs) allow users to view and edit their preferences. Another contribution of our work is to examine the behavior of these APMs as they provide additional insight on what advertisers are doing with what information they receive.

We use a modification of our methodology in a smaller study of Facebook to study how it translates user information and a user’s behavior on non-Facebook sites into ads served on the social networking site. Recent work such as [15], has demonstrated that Facebook is in a position to track user behavior on non-Facebook sites.

The remainder of the paper begins with a classification for how information is used to serve ads in Section 2. Section 3 describes our testing methodology and Section 4 presents results for the Google Ad Network. It is followed by a smaller study of Facebook in Section 5 and then concluding remarks. A more detailed version of the paper with illustrative examples captured during testing is also available [16].

WPES’12, October 15, 2012, Raleigh, North Carolina, USA.

2. AD CLASSIFICATION

As a means to help understand what third parties do with the information they obtain we have developed a two-dimensional classification with one dimension for known information and one dimension for inferred information. This classification is shown in Figure 1 and focuses on the ads served by an ad network.

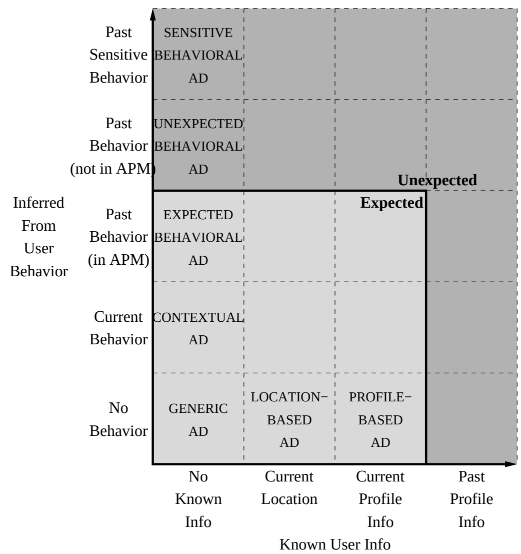

Figure 1: Classification of Ads Shown

The horizontal dimension shows information that is known about a user where a user's current location, obtained via a precise latitude/longitude location or a geoIP mapped location, is one type of known information about a user. We note that location could be shown as a third dimension, but add it to the known information dimension for simplicity of display. The resulting horizontal dimension indicates ads may be displayed using no known information about a user, using the user's current location, using information from a profile available on the current page (e.g. using information from a user's Pandora profile to show ads while the user visits Pandora) and using information obtained from a previous profile to show an ad on a subsequent page (e.g. using information from a user's Pandora profile for displaying an ad on a subsequent page).

The vertical dimension shows information inferred about a user's browsing behavior including no behavior, context of the current Web page and past browsing behavior. As shown in the figure, we separate past browsing behavior information into three types: behavior captured in the Ad Preference Manager for the user, behavior not captured in the APM and behavior that deals with sensitive topics.

The boxes within the figure show potential combinations of information used for an ad where we adopt the notion of a "generic ad" for an ad that has known no context on either dimension. While the CAPITALIZED TEXT in the figure labels ads based on only known or inferred information we expect combinations also exist—such as ads combining context of the current page along with a user's current location.

The shading of the figure groups combinations into *expected* and *unexpected* based on what we believe would be the perception of most users. Thus on the horizontal dimension we expect to see location and profile information for the current page being used while on the vertical dimension we expect to see contextual advertising as well as behavioral advertising that is captured in the APM of the associated ad network. We do not expect to see profile information being used on subsequent pages nor do we expect to see behavioral ads for topics not captured in an APM or such ads dealing with sensitive topics.

3. METHODOLOGY

Our approach is to study specific ad networks. An ad network is largely a black box in terms of how it works. However, an ad network does have a number of inputs that can be controlled and outputs that can be observed as shown in Figure 2. Specifically the sites that are visited and the textual input (search terms) provided to an ad network can be controlled. Similarly, current location and profile information can be controlled. Manipulating these inputs in a controlled manner allows us to detect if they come through as visible output—in the APM of the ad network, in the displayed ads themselves or in both.

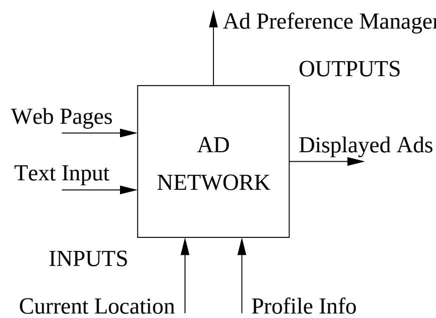

Figure 2: Inputs and Outputs of an Ad Network

Our first step in studying a particular third-party ad network is to identify a set of first-party sites for which the advertiser has a presence when the first-party site is loaded in a browser. We primarily found these first-party sites by using results from an ongoing longitudinal Web crawl using the methodology described in [12]. We also include sites dealing with sensitive topics (e.g. health and sexual orientation) as well as sites where a user creates an online profile of information. We identify about 15 first-party sites for each ad network.

Each experiment for an ad network consists of daily sessions for a ten-day period. A session comprises visiting the first-party Web sites successively and performing representative actions on these sites.

The same controlled browser is used for an entire experiment. At the beginning of the experiment we delete all cookies and history in the test browser. We do not explicitly delete Flash LSO cookies, but did not observe any such cookies for the ad networks we tested. For sites requiring a login, we do so in the first session of the experiment and then do not log out during the remainder of the experiment. All tests are run from Worcester, Massachusetts, which is part of the Boston metropolitan area.

During an experiment, we observe the ads served by the relevant ad server and check the APM contents during the course of the sessions to see how it changes based upon the visited sites and inputs. We do not click on any ads during a session thus providing no additional information about interests. We also record all HTTP traffic and object content using the Fiddler proxy [5] running on the same computer as the test browser. This traffic and content is saved for later analysis.

For this work, all tests are performed manually with a session typically taking 20-30min. We plan to automate the tests moving forward, but the manual approach not only allows us to drive the test, but to observe and understand the rendered ads whether they be text, image or Flash. These observations helped us in automating the analysis of the collected content. We are also able to stop and capture examples of different types of ads based upon our classification in Figure 1. Because the tests are performed sequentially, the results are impacted by the churn of ads over time. Because of this churn we are cautious in attaching significance to small differences when reporting results from separate experiments.

4. RESULTS

We applied our methodology to ad networks that provide an APM. From this set of ad networks, we initially studied four of the larger ones: AOL, BlueKai (actually a data exchange), Google and Yahoo!.

Based upon initial results, we choose to focus on more systematic analysis of the Google ad network for a number of reasons: it is the largest ad network with a presence on approximately 60% of popular Web sites based on our own recent data collection; it was the only ad network for which we saw evidence of unexpected results from our initial work; and on March 1, 2012 it modified its privacy policy [8] in part to show "more relevant ads" affording an opportunity to study how activities on Google-owned sites affects ads on non-Google sites.

Using a set of 20 sites (see [16] for list), we performed six experiments with a total of 55 sessions (including one session of 8 days and one of 7) in May/June 2012 where in each experiment we induced different sets of interests. Details on the interests and how they were induced are described in Section 4.2. An Internet Explorer browser with a default configuration (all cookies accepted, JavaScript execution enabled, no ad blocking) was used for all experiments.

The saved content from each session was analyzed in an automated manner where we only considered content that was served directly or indirectly by the Google ad network itself. Indirection was detected via the string "google" or "doubleclick" being in the `Referer` header or as part of JavaScript code that was serving the ad. We used a keyword-based analysis for each induced interest where we searched the saved content for one of a set of keywords related to that interest. Automated analysis of image and Flash content was done by matching keywords in URLs. Matches were subsequently verified to ensure correctness and that the ad was not contextual based upon the contents of the page.

In analyzing results, we start with the four combinations in the lower left corner of Figure 1. It is not surprising that we observed numerous ads each session related to the contents of the current page (contextual advertising) or having no clear relationship to the current or previous behavior (generic advertising). Location-based advertising is also common. Analyzing the data for the keywords "worcester," "boston," "massachusetts" and "new england," we verified that all 55 experimental sessions exhibited at least one instance of location-based advertising. It is also not unusual to see the combination of contextual and location-based advertising.

4.1 Profile-Based Ads

We next looked for evidence of profile-based ads (two right-most columns in Figure 1) using the sites linkedin.com and pandora.com in which we established an account with the profile information such as age, gender, location and job information. Although linkedin.com transmits much profile information to DoubleClick, we did not find evidence that this information was being used by DoubleClick in serving ads on linkedin.com.

However, we did find evidence that profile information on Pandora is being used when DoubleClick serves ads on the site. We found two types of ads where the ads match information in the Pandora profile. The first matches the profile location of New York while the second shows an ad from match.com with a default age range that corresponds to the age in the profile. We would not expect the profile information to be used for advertising on subsequent first-party sites, which is this last column in Figure 1. Evidence for such subsequent use is not clear. We found no evidence that the age in the profile is being used as the Google APM consistently converged on demographics of a 65+ age male in all of our experiments. The profile age did not influence this inferred characteristic nor did we observe the use of the profile age in ads except those for match.com on Pandora.

Evidence for the subsequent use of the profile location is less clear. Using our automated analysis with keywords "nyc" and "new york," we examined sessions where the profile location of New York was passed to Google via LinkedIn and Pandora. We found that 59% of such sessions contained at least one ad (not on LinkedIn or Pandora) for this location. Unfortunately, we did not perform any experiments without LinkedIn and Pandora so we do not have a baseline for comparison.

4.2 Behavioral Ads

We next test for the presence of behavioral advertising (rows 3 and 4 in Figure 1) by inducing a number of interests in each of our experiments. The complete set of interests, along with how they were induced and the keywords used when we analyzed the collected data are shown in Table 1. Within an experiment, more than one, but fewer than all interests were induced.

Table 1: Induced Behavioral Interests

Induced Interest	How Induced?	Match Keyword(s)
cars	Ford, Toyota sites	ford, toyota, cars, autos, mazda, honda, jeep
dogs	search term	dog, k-9, pets, veterinarian, puppies
golf	search term	golf
investment	Bloomberg site	finance, invest, stock, market, trusts
miami	location selection	miami, south beach
tennis	search term	tennis, racquet

Before examining results on how these induced interests are mapped into behavioral ads, we first examine how these interests influence the evolution of the APM—the other output of the ad network—over time. Figure 3 shows the evolution of the APM over eight sessions (days) of one experiment where cars, golf, investments and miami were the induced interests from Table 1.

Figure 3 shows the number of interest and demographic categories displayed by the Google Ads Preferences Man-

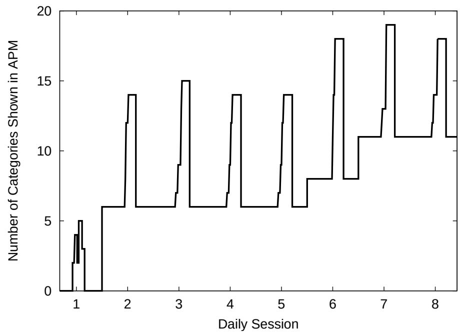

Figure 3: Evolution of APM Over Time

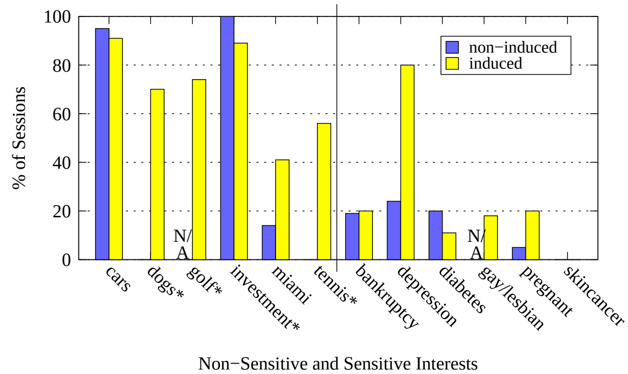

Figure 4: Percentage of Sessions Displaying Non-Contextual Ad Matching Induced Interest (* indicates interest was shown in APM when induced.)

ager [6] at the beginning of, during and after each session. The horizontal gaps between each session represent elapsed time in which no Web site is visited with our test browser. At no point during the experiment were categories manually added or removed; what is shown reflects only inferences made based upon browsing behavior.

While the process used by Google for adding and removing APM categories is not known, the behavior in Figure 3 appears to reflect a two-stage process for mapping browsing behavior to the set of categories. We see short-term behavior where interest (not demographic) categories are added in response to input text and the content of particular pages. Observation shows that visiting pages of nytimes.com and to a lesser extent bloomberg.com results in immediate addition of categories related to the content of these pages. In contrast, we do not observe this behavior when other news sites such as cnn.com or cbsnews.com are visited. However these short-term additions are fleeting as we observe each of them being dropped from the APM within an hour of when the session is completed.

The addition of categories that persist long-term occurs between sessions. For example in Figure 3, we see that the number of categories in the APM returns to zero shortly after the completion of the first session, but by the time the second session is started the next day, six (four interest, two demographic) categories have been added and these categories persist across the short-term comings and goings of categories in subsequent sessions.

With a better understanding of how the APM works, we now examine the other visible output of the Google Ad Network—the ads themselves. In analyzing the results we use the keywords (and appropriate variants) shown in Table 1 to match ads in each session that were served by Google and are not contextual—for example ads for financial investments on bloomberg.com are contextual. Behavioral ads are non-contextual ads corresponding to an induced interest. In our testing we analyze the number of non-contextual ads for an interest both when it is induced and not induced.

The left portion of Figure 4 shows a summary of results for the six interests given in Table 1. In each case, the results show the percentage of sessions in which at least one non-contextual ad match occurs for the given interest. Results are provided for sessions in which the interest is not induced and sessions for which it is. Note the interest golf was induced in all sessions so non-induced results are not-applicable in the figure. All results are based on at least 10 sessions with most results based on more than 20 sessions.

Figure 4 shows a number of interesting results. Non-contextual ads for dogs and tennis were only found when the interest was induced and in each case corresponding categories populate the APM. In contrast, ads for cars and financial investment always or almost always are found whether or not the interest is explicitly induced. What is unexpected about this result is how the Google APM is handled in each case. For investment, the Google category "Finance - Investing" appears in all relevant experiments while for cars, we never saw a category for automotives in general or Ford or Toyota in particular appear in the APM during an experiment. We observed similar behavior for the interest Miami where non-contextual ads were observed, but in most experiments the interest was not reflected in the APM.

It can be argued that the absence of an expected interest in the APM is not an issue. What is shown in the APM is just a heuristic and a user can add the interest themselves. However if the APM is accurate for some interests and not for others then it is inconsistent and not complete in representing what is known about the user, which was part of the rationale in introducing APMs in the first place.

4.3 Induced Sensitive Interests

We next repeated our methodology for a range of sensitive interests to test for evidence of such ads in light of the policy that "Google will not show ads based on sensitive information or interest categories, such as those based on race, religion, sexual orientation, health, or sensitive financial categories" [7]. Consistent with this statement, we observed that induced sensitive interests had no effect on the APM categories. The interests we tested included sexual orientation as well as ones on sensitive health and financial matters. These interests are shown in Table 2.

Table 2: Induced Sensitive Interests

Induced Interest	How Induced?	Match Keyword(s)
bankruptcy	search term	bankrupt, chapter 7, debt, tax relief, foreclosure
depression	health search term	depression
diabetes	health search term	diabetes
gay/lesbian	gaylife, thenewgay sites	lgbt, lesbian, gay, mat_boy
pregnancy	health search term	pregnant, ob/gyn, infant, baby, birth
skin cancer	health search term	skin cancer, melanoma

Despite Google's statement, we do see non-contextual ads for induced sensitive interests in our collected data. One example is where an ad about depression appears on nytimes.com. Similarly, we found ads that match the induced interests of bankruptcy and pregnancy. An ad for bankruptcy is also contextual as it appeared on the accuweather.com page for Miami weather, which was being induced as an interest. We also observed an ad for a New England center for "Getting Pregnant After 30" on macmillandictionary.com combining a sensitive induced topic along with location. Similarly we observed instances of match.com ads with photos of men on non-gay sites as well as ads advocating "LGBT for Obama" on thefreedictionary.com.

These examples clearly show that the Google Ad Network is serving non-contextual ads related to induced sensitive topics. One obvious question is what behavior is observed when the topics are not induced. These results are shown in the right portion of Figure 4.

The overall results reflect that non-contextual ads for sensitive interests are shown less frequently than those for non-sensitive interests. The results also reflect the frequency that such ads are shown does not differ significantly except for depression where 8 out of the 10 (80%) sessions in the one experiment in which it was induced contained an ad on nytimes.com. We initially suspected these would actually be an example of contextual advertising, but found no evidence of "depression" in the contents of the nytimes.com page and have no basis to conclude they are contextual.

In summary, Google may not be serving ads based on sensitive interest categories, but it is serving non-contextual ads related to sensitive interests—whether induced or not.

4.4 Induced Interests from Google Sites

In the last portion of our study, we examined some of the impact of Google's modified March 1, 2012 approach to sharing information across Google properties [8]. We used this opportunity to study how activities on Google-owned sites affects ads on non-Google sites. For this part of our study, we changed the source from which interests were induced. Specifically, in some experiments the interests of golf, bankruptcy and depression were induced based on queries to the Google search engine. Similarly, YouTube was used to search for and view videos based on the interests of dogs and pregnancy.

We observed that interests induced via YouTube were reflected in ads and in APM categories. For example, we observed two non-contextual ads for dogs and pregnancy that were served on thefreedictionary.com within a session in which these topics were induced on YouTube. The inducement of dogs caused this category to be included in the Google APM in a similar manner as if it was induced from a non-Google site. In contrast, we did not observe that the inducement of golf as a Google search term caused this topic to be included as an APM category. The results show that induced interests using Google search show similar frequency as when the interests are not induced at all while the two YouTube-induced interests correlate better in frequency with induced interests on non-Google sites. See [16] for more details.

5. FACEBOOK RESULTS

We also used the ads served by Facebook to its users as a means to understand how the social networking company makes use of the information received from its users. This work was partially motivated by studies such as [15] and our own work showing Facebook as a growing *third-party* with a presence on nearly 40% of popular first-party sites.

This portion of our study was less extensive than our previous study and examined two specific research questions. First, in comparison with information provided by a user on Facebook itself how does a user's browsing behavior on non-Facebook sites influence the ads this user sees on Facebook both when the content is and is not Facebook-liked? Second, how is sensitive browsing behavior and information about a user handled in ads by Facebook?

5.1 Non-Facebook Browsing Behavior

To answer these questions we used a similar methodology to that described in Section 3 in that we constructed a set of first-party sites having the Facebook Like button that was being served by Facebook. The first-party sites induced both non-sensitive and sensitive interests.

Using these sites and interests, we conducted four experiments, each of 15-20 sessions, where at the beginning of each experiment we logged into Facebook and remained logged in for the remainder of the experiment. At the end of each session we visited Facebook then saved the results of our browsing session for later analysis. In the first experiment, we did not induce any interests, but simply recorded the displayed Facebook ads in each session. Second, for each session we visited the Web sites in our test set, but did not click on the Facebook button. This experiment examines whether non-Facebook behavior is influencing Facebook ads. Third, we repeated the previous experiment, but in this experiment we not only visited sites, but Facebook-liked pages we browsed on these sites. Fourth, we did not visit any non-Facebook sites, but instead induced interests by indicating them as Facebook interests at the beginning of the experiment. We also indicated that our male user was interested in men.

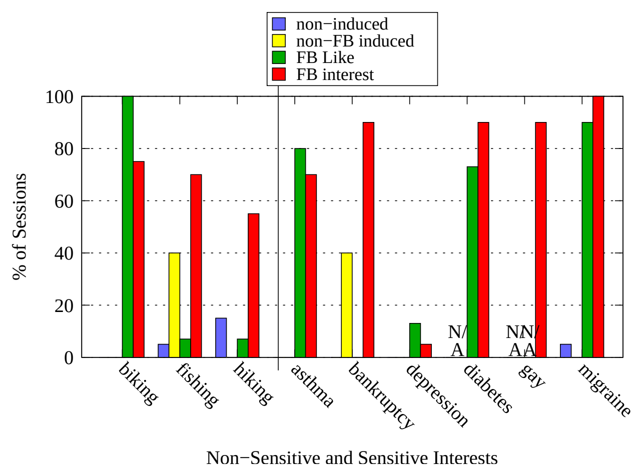

Figure 5: Percentage of Sessions Displaying Facebook Ad Matching Induced Interest

Results for non-sensitive and sensitive interests from these four experiments are shown in Figure 5. As shown, Facebook ads rarely match non-induced or non-Facebook-behavior-induced interests. The only instances for the latter case are ads for a Facebook fishing game and ones about college debt, which may relate to bankruptcy. In contrast, Figure 5 shows generally a high percentage of sessions included ads for interests induced either by Facebook-liking a page or explicitly including the topic as a Facebook interest. Collectively, these results do not provide clear evidence that a user's browsing behavior on non-Facebook sites influences the ads shown on Facebook itself unless a user Facebook-likes the site content.

5.2 Facebook Ads for Sensitive Topics

We also looked more closely at how Facebook handles ads for sensitive topics. This question is relevant as the Facebook policy for ads [3] says "Ad text may not assert or imply, directly or indirectly, within the ad content or by targeting, a user's personal characteristics within the following categories: race or ethnic origin; religion or philosophical belief; age; sexual orientation or sexual life; gender identity; disability or medical condition (including physical or mental health); financial status or information ..."

The results in Figure 5 show a high percentage of sessions where ads matching sensitive interests are shown. Despite targeting sensitive topics, these ads are apparently acceptable according to guidelines provided by Facebook for creating ads [4] because they make statements describing the product or service and not necessarily the characteristics of the user.

However other ads that we observed do appear to violate Facebook's own guidelines. One ad showed the user's age of 32. Another encouraged recipients to join gay men in creating their own roommate listing. Yet another asks recipients "do you have diabetes?" These ads assert sensitive information directly or indirectly through language such as asking if a user has a sensitive condition or encouraging them to join others with a sensitive condition implying that the recipient has this characteristic. We found ads asserting a sensitive characteristic in *each* of the experimental sessions where the interest was induced as a Facebook interest, primarily through ads for diabetes, migraines and sexual orientation.

6. SUMMARY AND FUTURE WORK

In summary, our initial study of a few ad networks and our focused study of the Google ad network found many expected contextual, behavioral and location-based ads along with combinations of these types of ads. We also observed some profile-based ads. We generally found that behavioral ads based upon induced interests were shown as categories in the Ad Preference Manager of the ad network, but found a couple unexpected cases where the interests were not visible in the APM. We also found unexpected behavior for the Google ad network in that non-contextual ads were shown related to induced sensitive topics regarding sexual orientation, health and financial matters. However, we also found such ads displayed when the sensitive topics were not induced meaning that Google may not be showing behavioral ads for these topics, but users with such sensitive topics may be unable to discern the difference.

In a smaller study of Facebook, we did not find clear evidence that a user's browsing behavior on non-Facebook sites influences the ads shown to the user on Facebook, but we did observe such influence when the Facebook Like button is used to express interest in content. We did observe Facebook ads appearing to target users for sensitive interests with some ads even asserting such sensitive information, which appears to be a violation of Facebook's stated policy.

Our work has a number of directions for future work. Automating the data collection will allow for more experiments and allow them to be done in parallel to reduce the impact of ad churn. We plan a longitudinal study as the results reported here may change over time. We also plan to evaluate the effectiveness of various privacy protection measures.

Addendum

We disclosed the results of our study to Google and Facebook as possibly being inconsistent with their stated policies. We did not receive any responses to our disclosures.

7. REFERENCES

[1] Julia Angwin. The web's new gold mine: Your secrets. *Wall Street Journal*, July 30 2010. http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html.

[2] Rebecca Balebako, Pedro Leon, Richard Shay, Blase Ur, and Lorrie Faith Cranor. Measuring the effectiveness of privacy tools for limiting behavioral advertising. In *Web 2.0 Workshop on Security and Privacy*, May 2012.

[3] Facebook advertising guidelines, Last Revised: March 20, 2012. http://www.facebook.com/ad_guidelines.php.

[4] Facebook help center: Ad and sponsored stories copy, image, targeting, and destination. http://www.facebook.com/help/?page=245316378826196.

[5] Fiddler web debugging proxy. http://www.fiddler2.com/fiddler2/.

[6] Google Ads Preferences. http://www.google.com/ads/preferences/.

[7] Interest-based advertising: How it works - Google ads preferences, 2010. http://www.google.com/ads/preferences/html/intl/en/about.html.

[8] Google privacy policy, Last Modified: March 1, 2012. http://www.google.com/intl/en/policies/privacy/.

[9] Saikat Guha, Bin Cheng, and Paul Francis. Challenges in measuring online advertising systems. In *Proceedings of IMC*, November 2010.

[10] Balachander Krishnamurthy, Konstantin Naryshkin, and Craig E. Wills. Privacy leakage vs. protection measures: The growing disconnect. In *Proceedings of the Web 2.0 Security and Privacy Workshop*, pages 1–10, Oakland, CA USA, May 2011.

[11] Balachander Krishnamurthy and Craig E. Wills. On the leakage of personally identifiable information via online social networks. In *WOSN*, 2009.

[12] Balachander Krishnamurthy and Craig E. Wills. Privacy diffusion on the web: A longitudinal perspective. In *Proceedings of the World Wide Web Conference*, 2009.

[13] Jonathan Mayer. Tracking the trackers: Where everybody knows your username, October 2011. http://cyberlaw.stanford.edu/node/6740.

[14] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. Detecting and defending against third-party tracking on the web. In *Symposium on Networked Systems Design and Implementation*, April 2012.

[15] Arnold Roosendaal. Facebook tracks and traces everyone: Like this! Technical report, Tilburg Law School, November 2010. Legal Studies Research Paper Series No. 03/2011. http://ssrn.com/abstract=1717563.

[16] Craig E. Wills and Can Tatar. Understanding what they do with what they know. Technical Report WPI-CS-TR-12-03, Computer Science Department, Worcester Polytechnic Institute, August 2012. http://www.cs.wpi.edu/~cew/papers/tr12-03.pdf.