KNOWPRIVACY

Project Members: Joshua Gomez, Travis Pinnick, Ashkan Soltani

Project Advisor: Brian Carver

Project Assistants: Sona Makker, Mark Mccans

Report Date: May 14, 2009

EXECUTIVE SUMMARY

Online privacy and behavioral profiling are of growing concern among both consumers and government officials. The Federal Trade Commission (FTC) has made several statements indicating the creation of new regulations if the industry does not improve practices via self-regulation. The goal of this project was to examine both the data handling practices of popular websites and the concerns of consumers in an effort to identify practices which may be deceptive or potentially harmful to users' privacy and, based on our findings, offer potential solutions that policymakers should consider when discussing any new Internet privacy regulations or that website operators could implement to potentially avert or soften such regulation.

The data for this project was pulled from six domains, three regarding user expectations and three regarding website practices. To understand user concerns and knowledge of data collection we looked at surveys and polls conducted by previous privacy researchers. In an attempt to discover what makes users so uneasy as to file a complaint, we looked at records of complaints and inquiries filed to privacy watchdog organizations such as the Privacy Rights Clearinghouse (PRC), the FTC, The California Office of Privacy Protection (COPP), and TRUSTe. Finally, to gain some insight into what aspects of data collection users are being made aware of, we looked at news articles from three major newspapers for topics related to Internet privacy. We analyzed the policies of the 50 most visited websites to get a sense of standard practices regarding the types of data collected about users, how that information is used, and with whom it is shared. We also looked at specific practices such as sharing information with affiliates and third party tracking.

From the previous surveys and polls we found several consistent conclusions: 1) users are concerned about websites collecting information about them and using it behavioral advertising, 2) users desire control over the collection and use of information about them, and 3) users lack knowledge and understanding about data collection practices and policies.

Nearly 40% of the complaints in the data set from the FTC were concerned with opt-out control. Data from the FTC, PRC and COPP all contained a significant number of complaints about Zabasearch and other data broker portals. Qualitative analysis of the complaints suggested that users were mostly concerned with a lack of control over data collection and public display of personal information. Data from TRUSTe showed growing concern with unauthorized use of data. All four data sets supported the theme of users' desire for control.

Analysis of news articles suggests that users were made aware of behavioral profiling and other data collection topics in general. However, discussion of some practices was non-existent. Almost no mention was made of the practice of sharing data with affiliates or of the use of web beacons.

Our survey of privacy policies revealed that most of the top 50 websites collect information about users and use it for customized advertising. Most contained unclear statements about data retention, purchase of data about users from other sources, or notice in the event of company acquisition. Most policies stated that data about users was not shared with third parties while also stating that third party tracking was allowed and that data was shared with corporate affiliates.

Our analysis of web beacons revealed that beacons are ubiquitous on the web. All of the top 50 contained at least one beacon at some point in a one month time period. Some had as many as 100. Of greater concern was the depth of coverage that some tracking companies have. Several of the tracking companies had a beacon on the majority of the top 100 sites. Google in particular had extensive coverage. It had a beacon on 92 of the top 100 sites, and 88% of the total domains reported in the data set.

We requested lists of affiliates with whom data is shared from website operators, but received none. To get a general impression of the number of companies each website could potentially share data with under its current privacy policy, we looked up the parent company of each website and the list of subsidiaries those companies own. The average number of subsidiaries was 297.

From the data, we observed a continuous theme of user concern about data collection and lack of control. We also determine that users only complain when they can see the misuse of data and are made aware of a forum appropriate for their complaint. We also noted that services like Zabasearch have similarities with behavioral profiling.

Despite users concerns about data collection and control, website operators continue to collect and share information about users without sufficient means of opt-out. Self regulation has failed, but users cannot avoid the Internet entirely as it has become engrained in daily life as well as civic and cultural discourse.

Website policies are deceptive when they claim third party sharing does not occur while also claiming that third party trackers are allowed and data about users is shared with affiliates. Most users would consider any entity they have no relationship with to be a third party, regardless of its affiliation to visited website, and sharing space on a webpage in a manner that allows collection of information about a user is essentially the same as sharing data about the user.

Based on our findings we recommend that website operators disclose to users all the information that has been collected about them and with whom it has been shared.

We recommend that users be given the ability to opt out of affiliate sharing and third party tracking.

We recommend that all privacy policies include a link to the FTC's online complaint form.

We recommend that website privacy policies be written in a more readable form without contradictory statements about third party sharing. Any policy that does not meet a minimum standard should be labeled "data collection policy" rather than "privacy policy."