XXXXXXXXX

Creating Better Privacy on the Web

CSC 540 - Fall 2013
October 30, 2013

With increasing concern for Internet privacy, it's no wonder there are many studies on web tracking and the information gathered by websites and advertisement groups. Users should understand that when they visit a website, objects from other sites may be loaded in the background, which the third-party sites use to track users' information as they browse the Internet. By information I mean the data gathered by web sites from user input (as form data or profile data), IP addresses, and system fingerprinting. The information collected can be used to provide targeted advertisements to a user based on the collected information and sometimes inferred information. Other uses of the information gathered are unknown, but it may be compiled into demographic databases which are sold to marketers. Clearing the browser cache and history, and removing cookies can prevent trackers from tracking across browsing instances, but may have a negative impact on a user's browsing experience and do not prevent tracking on sites where the user is logged-in [1]. Ad blocking may prevent some tracking objects from loading, but do not always catch everything [1, 2]. In response to these failures in current privacy protection techniques, researchers have developed tools to prevent a larger percentage of third-party tracking. As current privacy protection methods are not sufficient to block all third-party tracking, and as trackers become more proliferated and are able to bypass current protection methods, more research is needed to test the efficiency of current tools and to create better ones. In order to do that, it is best to understand how trackers and current protection tools work. It should be noted that the attacks and leaks of private information described in these studies are only those that happen as a result of trackers embedded in web pages, not those caused by malware.

An early study by Krishnamurthy and Wills shows the need for better privacy for users. A concern brought up in this study is the issue of re-identification, or "the ability to relate supposedly anonymous data with actual identities" [2]. The study shows that it may be possible for the information gathered by trackers to be used with other profile information to link the two data sets. The connections between sites the user directly visits (visible sites) and those loaded in the background (hidden sites, or trackers) may allow this behavior [2]. They also found that traditional blocking techniques (ad blocking and blacklisting) are only partially effective because of the problem of properly identifying trackers [2]. One of the faults of current ad blocking software is its reliance on blacklists and whitelists for determining allowed and blocked objects. As new tracking websites are created, the ad blocker needs to be updated with the latest information, like a malware scanner is.

Jackson, Bortz, Boneh, and Mitchell discuss and develop tools to prevent privacy leakage as a result of lack of proper implementation of the same-origin policy. The same origin policy limits the interaction between websites of different domains in a user's web browser [3]. Due to the improper implementation of this policy, a number of attacks against privacy are possible [4]. One attack is possible because browsers at the time of this study don't protect cache using same-origin policy, allowing third-party websites to access cache files to determine what sites a user has visited. Another attack involves visited link differentiation which may allow sites to access the browser history database to determine which sites were visited. To prevent these attacks, the researchers developed tools, called "SafeHistory" and "SafeCache" for the Firefox web browser to help protect cache and browser history from these attacks. For example, with these tools the user will only see links in a page as being visited if they were accessed from that page or domain, otherwise it would show as not being a visited link even if that page had been accessed by the user through another domain.

As of this writing the tools suggested in this study are not supported by the latest version of Firefox. Although some updates to Firefox since this study have increased the security with regards to the way the browser handles third-party cookie blocking, it does not support history protection in its default settings. From the researchers' web page at http://www.safehistory.com/, users can access test cases to see how their browsers currently protect from certain kinds of tracking.

This study classifies trackers by how they are imbedded in the page creating four categories: no tracking, non-cooperative, semi-cooperative, and cooperative [4]. A secondary classification differentiates trackers that are single-session or single-site, with those that are multiple-session or multiple-site. This classification expands on the classification in the previous study by describing how the tracker is imbedded in the web page. This classification is good for enforcing the same origin policy on trackers that violate it. However this may also limit the ability of the tools suggested to only block those trackers and not others, which users desire.

Roesner, Kohno, and Wetherall provide "the most complete study of web tracking to date" [1]. After observing how different trackers function, they created a classification system based on the behaviors of the trackers, which allowed them to determine which behaviors could be blocked by certain techniques and develop a tool to block behaviors that could not previously be blocked. However, their findings showed that some trackers use multiple behaviors and could circumvent certain blocking techniques that only catch one of the behaviors of the tracker. A particularly concerning observation that they made was evidence of cookie leaks, where one site can gain access to the information in another site's cookie data.

> *For example, `msn.com` and `bing.com`, both owned by Microsoft, use cookie leaking mechanisms within the browser to share cookies with each other…. This enables Microsoft to track a unique user across both MSN and Bing, as well as across any site that may embed one of the two.* [1].

Unfortunately the only way to prevent this leak is by blocking or deleting cookies at the end of the browser session, meaning the user will have a limited web experience and may not be able to save states across browsing sessions. However this may not be enough to prevent tracking [1, Section 5.1 Clearing client-side state]. Even when the user deletes browsing history and cookies, the information gathered by trackers in further web browsing may allow for re-identification, especially if the tracker can match the gathered data with an existing profile [2]. Using current browsers and tools, there is no way to completely protect from cooperative tracking of this behavior.

Roesner, Kohno, and Wetherall also show that current defenses against tracking are inadequate. They show that third-party cookie blocking can be ineffective because of the way that different browsers implement their third-party blocking rules. "While Firefox blocks third-party cookies from both being set as well as from being sent, most other browsers (including Chrome, Safari, and Internet Explorer) only block the setting of third-party cookies." [1]. Social media sites often set cookies when a user visits their site, but also use trackers on different web sites to allow their clients to express interest in a certain product, article, or page. In these cases, the trackers may be able to use the cookie, which was set when the user visited the social media site's page, when the user visits a page on a different domain containing the social media site's tracker [1]. This behavior only happens in browsers that do not block the sending of third-party cookies when third-party cookie blocking is enabled. They also show that the "Do Not Track" header, Pop-up blocking, and "Private Mode" browsing, which can be enabled in browsers, only prevent trackers using certain behaviors.

The tool these researchers create, called "ShareMeNot", is intended to prevent trackers from social media websites to be used on third-party pages unless the user chooses to express interest in the page. This is done by blocking the "like" or "+1" buttons that social media sites often put on other pages to track users' interest. The researchers decided on this tool because they found that it was the type of tracker with a behavior that could not be blocked by other methods. The fear is that the social media sites may track users' visits to the third-party site regardless of their expressed interest (through clicking the "Like", et al. buttons) and may infer data that might be used in advertising or creating demographics. Another important feature of this add-on is to maximize privacy protection while minimizing the impact on a user's web browsing experience [1]. ShareMeNot allows users to choose to allow a social media button on third-party websites when they choose to. ShareMeNot is currently available for Chrome and Firefox browsers.

Although this tool is useful in blocking the behavior of social media trackers, other add-ons and settings would be needed in order to protect users from different tracking behaviors.

Willis and Tatar provided better information of how advertisers use the information gathered from user web browsing. Their study focuses on the Google ad network (Google Analytics and Doubleclick) and on Facebook advertisements. They provided a classification model of server selected advertisements based on attributes of known user information, such as location and profile information, and observed user behavior, such as user input. The information sets are divided into two groups: expected behavior (showing advertisements that are of interest to the user because of information gathered), and unexpected behavior (showing advertisements that may be of a sensitive nature or may be inferred by tracking from external sites or outdated profile/preference information) [5]. This information can be used in future studies in determining what behavior should be blocked and what should be allowed. It is also useful in showing how advertising networks function.

Their study shows that Facebook does not appear to track users on third-party sites for advertisements on `facebook.com` unless the user expresses interest on the third-party site [5]. However, it was also shown that sometimes the information gathered by trackers was used in selecting advertisements of a sensitive nature, inconsistent with Facebook's stated policy [5].

Studies have shown that traditional privacy methods do not completely protect users as they visit multiple websites. While some tools have been created to help protect users' privacy against trackers that exhibit certain behaviors, there currently does not appear to be one tool that does so completely. Users posting information to the sites, as in the cases of web searches or posting to social media sites, is not covered by privacy, but it is alarming that this information may be linked with unrelated information from the user's web browsing. The behaviors of some trackers can be blocked by denying third-party cookies or by using traditional ad-blocking software. However there are few tools that will block trackers that cooperate with the sites in which they are imbedded, or block forced redirects. As further research is conducted, I hope that tools will be made for a wider variety of browsers that will help protect users' privacy and protect against third-party tracking.

References

[1] Roesner, F., Kohno, T., and Wetherall, D. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation* (NSDI'12). USENIX Association, Berkeley, CA, USA, 12-12.

[2] Krishnamurthy, B. and Wills, C.E. Generating a privacy footprint on the internet. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement* (IMC '06). ACM, New York, NY, USA, 65-70.

[3] Ruderman, J. The same origin policy, 2001. http://www.mozilla.org/projects/security/ components/same-origin.html

[4] Jackson, C., Bortz, A., Boneh, D., and Mitchell, J.C. Protecting browser state from web privacy attacks. In *Proceedings of the 15th international conference on World Wide Web* (WWW '06). ACM, New York, NY, USA, 737-744.

[5] Wills, C.E. and Tatar, C. Understanding what they do with what they know. In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society* (WPES '12). ACM, New York, NY, USA, 13-18.