# Web Privacy Tools and Their Effect on Tracking and User Experience on the Internet

Donnya Ajdari*
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA
dajdari3@gatech.edu

Tyler Stocksdale*
Computer Science and Electrical Engineering Department
University of Maryland, Baltimore County
Baltimore, MD
tstocks1@umbc.edu

Chris Hoofnagle
School of Law
University of California, Berkeley
Berkeley, CA
choofnagle@law.berkeley.edu

Nathan Good, Ph.D
Good Research
Berkeley, CA
nathan.good@gmail.com

***Abstract*—Websites often serve data from third parties. This content can have various purposes, one of which is advertising. While third-party advertising supports free services and information on the internet, the desire to target it and measure outcomes contributes to tracking of users. Users sensitive to such tracking can use a number of tools to mitigate this tracking.**

**This paper presents a study on various web privacy tools analyzing the direct effect of each tool on the user. This research was conducted as a part of a larger project[1] that is currently ongoing. We found that different web privacy tools do, in fact, reduce privacy invasions and produce a more pleasant user experience for internet users.**

***Keywords—web privacy tools; third-party tracking; internet user experience; online privacy.***

## I. INTRODUCTION

In this paper, we first provide some background information that will contribute to the motivation of this project and also present our objectives in conducting this research. We then describe some previously done research that has contributed to or inspired our research. Next, we provide a detailed description of our methods in data collection. Included is a thorough analysis of our considerations and design choices before we started collecting data. The data collection system that we developed is described in detail. After this we report our results and then wrap up with a discussion of our results and our conclusions. Finally, we outline work that could be done in the future to extend, modify or improve various aspects of our study.

### A. *Web Privacy Tools*

Web privacy tools exist to aid a user in protecting personal information and to block unwanted trackers and/or advertisements when surfing the web. These web privacy tools are implemented as browser extensions, some of which employ default settings, and others that require some level of configuration by the user in order to let the user ultimately decide what private information is protected. Often included in the configuration of a particular web privacy tool is the ability to block trackers used for advertising, analytics, beacons[2], privacy notices, and widgets. These tools attempt to enable users to make informed decisions [2] and control ones exposure to tracking and advertisement companies.

### B. *Cookies*

When a user (client) requests a website, the client and the website's server communicate by exchanging Hypertext Transfer Protocol (HTTP) messages. HTTP is a stateless protocol meaning that there is no record of previous Web transactions between the client and the server. Cookies were designed to remember certain information about the client and are stored as a string of characters (letters and numbers) on the user's hard drive [3]. This allows servers to maintain client state information on the backend database of a website. Cookies have many purposes and some common uses are to implement online shopping carts, facilitate user authentication, and remember user preferences [3]. Cookies can also be used to enable websites to implement advertising, social network integration, and record and analyze users' browsing activities across other unrelated websites [4]. In other words, they can be used to track a user.

* This work was done when the author was visiting University of California, Berkeley.

[1] Web Privacy Census. Led by Dr. Chris Hoofnagle at UC Berkeley

[2] For definition of beacon, see [1]

### C. Page Load Time

Web page loading speeds are one the most important design aspects of a website [5]. Slow response times cause users to abandon websites. This can result in many things such as a loss of customers on online shopping websites, loss of website revenue from advertisements, and a negative website reputation. It has been shown that a one second increase in page load time results in a fifteen fold increase in users who will abandon the webpage [6]. Therefore, page load time has a direct correlation with user experience on the internet. It is also known that minimizing the number of HTTP requests can dramatically speed up the load time of a web page [7]. Also, excessive use of multimedia data on a web page contributes to long page load times [5]. Advertisements use HTTP requests to load onto the page and they are often multimedia elements consisting of pictures, video, and sound. This leads to the hypothesis that advertisements slow down page load times.

### D. Objectives

The main objective of this paper is to analyze different web privacy tools and how they affect a user. Specifically, we want to ascertain how these tools prevent tracking through the use of cookies, and also if these tools create a more pleasant user experience by reducing page load time. This paper aims to answer the following two questions:

1. How effective are web privacy tools at preventing tracking?
2. Do web privacy tools improve a user's browsing experience?

For the purposes of this project, we correlate a decrease in the number of cookies with a decrease in tracking, and also a decrease in page load time with an increase in a user's browsing experience.

## II. Related Work

### A. Web Privacy Census

The Web Privacy Census is currently an ongoing project at the University of California, Berkeley. Its goal is to be able to define and quantify how consumers are being tracked on the internet and also to be able to make empirical statements about the state of internet tracking and privacy [8]. This project is, as the name implies, meant to be similar to a census in that its methods must be consistent and repeatable over time. This project collects and analyzes data on a site-by-site basis in order to make statements about the level of tracking on the most popular websites [9]. The Web Privacy Census project analyzes thousands of different websites collecting and storing data from each one. Because of this, an automated process is used in order to speed up data collection [9].

Our research is an extension of this project and needed to be designed to have similarly consistent and repeatable methods. We also needed to employ a site-by-site analysis so that the data collected can be easily integrated into the existing data of this project. Lastly, it was highly preferable that our process for data collection was fully automated.

### B. Jonathan Mayer

Jonathan Mayer, a Ph.D. student in computer science at Stanford University, has done research on multiple web privacy tools and their effect on third-party web tracking. In 2011, he found that web privacy tools vary in their effectiveness of preventing third-party tracking, but the top performing tools were ones that blocked third-party advertising [10].

Our research was loosely modeled after Jonathan Mayer's research. We wanted to update his findings with newer versions of select web privacy tools, but also analyze how these tools affect user experience through changes in page load times. The findings of his research gave us some ideas and criteria for selecting the web privacy tools we were going to use.

## III. Methods

As mentioned before, the methods of our data collection were required to adhere to certain guidelines. These guidelines are consistency, repeatability, site-by-site analysis, and easy integration into the existing Web Privacy Census project. In addition to these requirements there were other considerations that needed to be kept in mind:

First, collecting data on a site-by-site basis requires that data collected from one site must not leak over into data collected from another. This requires that after each site visitation, user data must be cleared from the browser, so that upon visiting the next site, the data collected would not be contaminated. This process of clearing user data is not trivial because of the existence of what are called zombie cookies [11].

Second, data collection using different web privacy tools must happen as close to concurrently as possible because of the ever-changing nature of websites. At any given time, a website may serve advertisements from different companies, change its privacy policy, or make other changes that would impact the consistency of the data collected. Since our data is supposed to represent a snapshot in time, each site's data needs to be collected across all web privacy tools simultaneously.

Third, each of the web privacy tools needs to be configured as similarly as possible to provide a consistent level of privacy protection. In addition, each of the tools should be configured to keep the default settings whenever possible because users of these web privacy tools tend not to change settings away from their defaults. [12]

Lastly, it is preferred that data collection occurs as quickly as possible so that multiple samples of data may be collected. To make more accurate conclusions, multiple samples of data should be collected for each website and averaged in order to reduce the effect of outlying data. Assuming each sample takes roughly the same amount of time to collect, it is easy to see how the total data collection time will quickly multiply when additional samples are collected.

### A. *Design Choices*

Since our project uses a site-by-site analysis, our first choice was which websites to collect data from. We decided to use two different lists of websites. The first list is the top 100 US ranking sites according to Quantcast. The second list contains the 100 websites with the most number of cookies according to data collected by the Web Privacy Census. The first list was chosen in order to test each of the web privacy tools on websites that a normal internet user would visit. The second list was chosen in order to conduct a "stress" test on each of the web privacy tools. We surmised that the websites with the most cookies would contain the most third-party tracking and have relatively high page load times.

The second design choice we had to make was which web privacy tools we were going to use and how to configure them. We took into account which web privacy tools were used in Jonathan Mayer's research and also which web privacy tools are popular with today's internet users. The tools we decided on were Adblock Plus (version 2.3), Ghostery (version 2.9.6), and DoNotTrackMe (version 2.2.9.618). Each of these tools had different configuration options. The options we chose for each web privacy tools were selected in an attempt to make the configuration of each tool as similar as possible.

Adblock Plus requires a first time setup to obtain user input on configuration options. We chose to disable tracking and remove social media buttons (widget blocking). We did not choose to enable malware blocking because this is a study of tracking only. After making these choices, Adblock Plus did not ask for any additional configuration input, so all other settings were left to the defaults.

Ghostery uses a wizard in order for the user to specify configuration settings. We first chose to enable auto-updating of Ghostery's library of known trackers. Next, we chose to block all trackers and all cookies known to be used for tracking purposes. Ghostery also has some advanced settings that the user can adjust, but we chose to leave those settings unchanged.

DoNotTrackMe is a tool created by Abine and has no wizard or first time setup. Unlike Adblock Plus and Ghostery, it does not wait for the user to specify configuration options to start blocking trackers. DoNotTrackMe's default settings are to "block all tracking companies everywhere" and to "use Abine suggestions". We did not adjust any settings for DoNotTrackMe since it did not require initial user input on configuration options in order to operate.

Our last design choice was how to measure third-party tracking and user experience on the internet. As stated in the introduction of this paper, cookies are widely used as a method of tracking by online advertisers. Since our chosen web privacy tools were all configured to block tracking, any reduction in the number of cookies on a website can be seen as a reduction in tracking. Also mentioned in the introduction of this paper is the importance of page load time to a user. As a result, we decided to record how long it takes for each website to load and use this as a measure of user experience.

### B. *System Setup and Execution*

Creating a system that would collect the data necessary for this project took careful planning and consideration. Since we needed an easy way to clear user data, and to collect data in parallel across multiple web privacy tools, we decided to make use of virtual machines (VM's). Each VM was set up with a different web privacy tool installed on the Firefox web browser. One VM did not have a web privacy tool installed and was used as a control. (Fig. 1) Each VM can be started and operated simultaneously. This both cuts down on data collection time and provides concurrent data collection. We used VirtualBox to create and manage our VM's locally. VirtualBox provides an option to take a "snapshot" of a VM and the ability to restore the VM to this "snapshot" at a later time. This allowed us to easily erase any user data that was accumulated by restoring the machine to an earlier point in time when no user data existed.

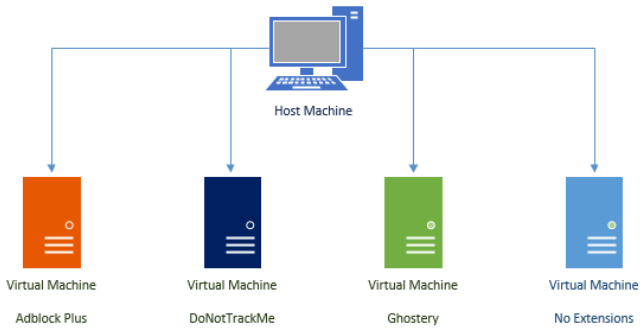


Fig. 1. Virtual machine setup.

VirtualBox also provides a command line API to manage and control the VM's that it was used to create. We created a python script to facilitate the communication between the machine hosting the VM's (host) and the VM's themselves. This python script on the host machine made use of VirtualBox's command line API in order to instruct the VM's to collect data for a certain website. The python script first started each VM, and then simultaneously gave each VM the instruction to fetch data from one URL retrieved from a list of URL's. Once the VM's had completed collecting the data, the python code powered off each machine and restored the snapshot from an earlier point in time. This process was repeated for each URL in the list. After all data was collected, data parsing and analysis took place. (Fig. 2)

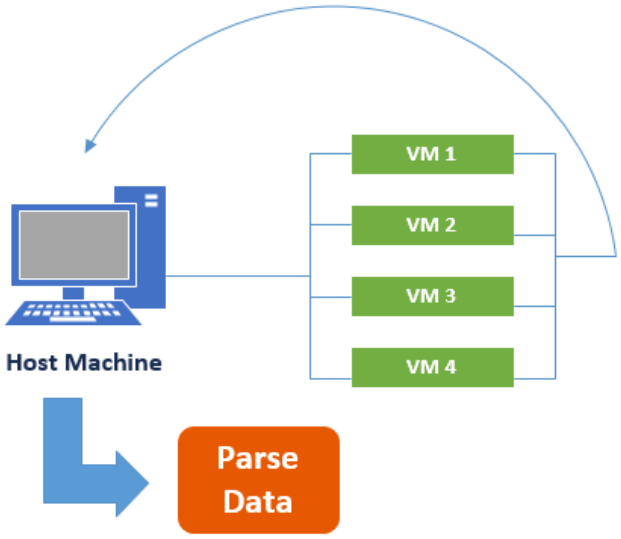


Fig. 2. Data collection process.

On each of the VM's, identical Java code was used to facilitate the data collection. This code made use of the Selenium webdriver. First, the webdriver started an instance of the Firefox web browser and navigated to the URL passed from the host machine. We used the Firebug extension to automatically export the load time of the web page. Once the page was loaded, the Java code extracted the cookies database file that Firefox uses to store persistent cookies. Session cookies were not collected. This information (load time and cookie data) was then sent back to the host machine and stored.

In order for the data that we collected to be integrated with the data collected by Web Privacy Census, a trivial level of data parsing would be required to append this data to the database that contains the Web Privacy Census data.

## IV. Results

For both lists of URL's (top 100 ranked US websites and top 100 websites with the most cookies) we collected data three times and averaged the resulting number of cookies and page load times. The data collected from the VM's with the web privacy tools installed was compared to the control VM that had no web privacy tools installed. We report our findings in a percentage decrease fashion. We measure the effectiveness of a web privacy tool by the percent decrease in tracking (number of cookies) and percent increase in user experience (decrease in page load time). Some of the page load time data we collected were considered outliers because these load times were largely inconsistent with the other load times collected from the same website. These data were omitted from our results because they heavily skew the average to the point that the result is no longer accurate.

We found that for the top 100 ranked US sites, Adblock Plus provided an average of 67.5% decrease in tracking across all websites, Ghostery provided a 76.4% decrease, and DoNotTrackMe provided a 60.8% decrease. Figure 3 is a graphical summary of these results.

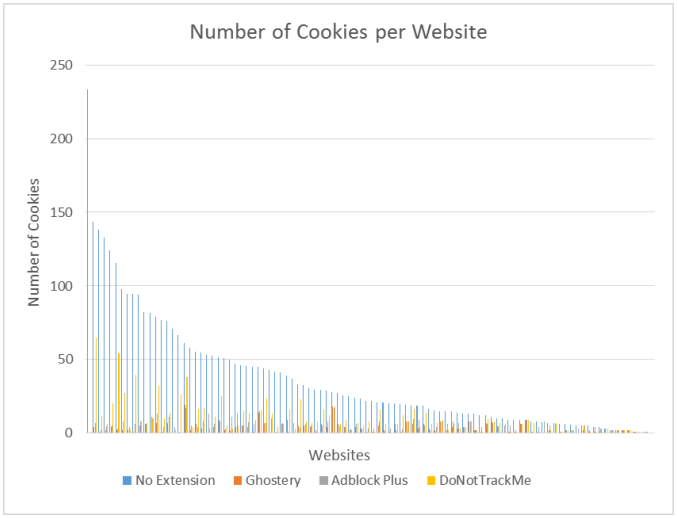


Fig. 3. Tracking data for top 100 ranked US sites.

We found that for the top 100 sites with the most cookies, Adblock Plus provided an average of 84.2% decrease in tracking across all websites, Ghostery provided a 89.0% decrease, and DoNotTrackMe provided a 73.9% decrease. Figure 4 is a graphical summary of these results.

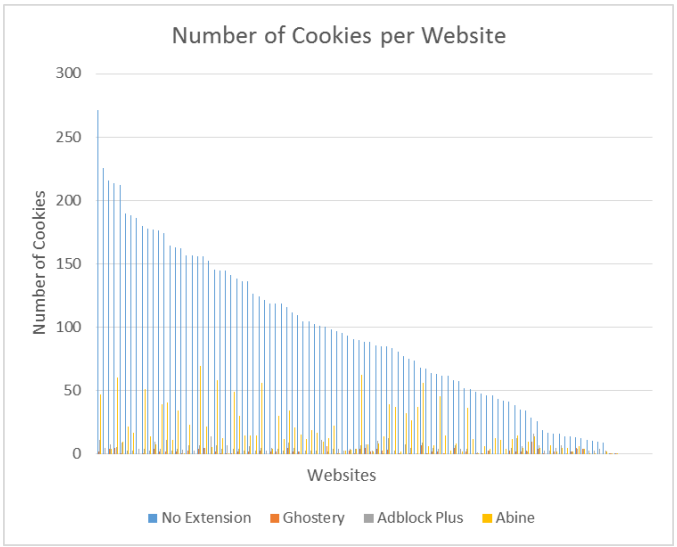


Fig. 4. Tracking data for top 100 sites with the most cookies.

We found that for the top 100 ranked US sites, Adblock Plus provided an average of 16.7% increase in user experience across all websites, Ghostery provided a 24.4% increase. DoNotTrackMe produced data which contained mostly negative page load times. Because of this, we have deemed this data as inconclusive. Figure 5 is a graphical summary of these results.

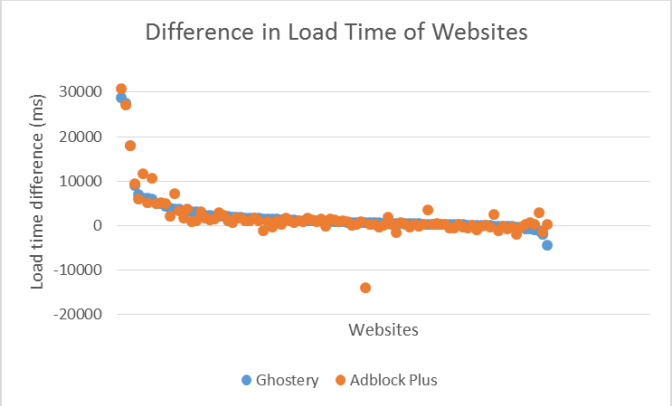


Fig. 5. User experience data for top 100 ranked US sites.

We found that for the top 100 sites with the most cookies, Adblock Plus provided an average of 4.1% increase in user experience across all websites, Ghostery provided a 10.5% increase. DoNotTrackMe produced data which contained mostly negative page load times. Because of this, we have deemed this data as inconclusive. Figure 6 is a graphical summary of these results.

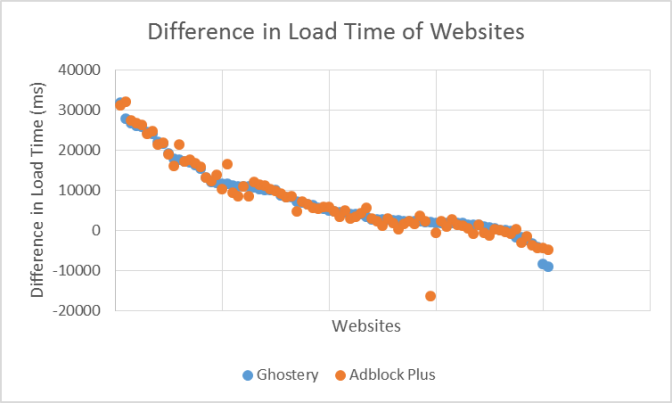


Fig. 6. User experience data for top 100 sites with the most cookies.

## V. Conclusion and Future Work

From the data we collected and our analysis of it, we can undoubtedly confirm Jonathan Mayer's research conclusions that web privacy tools that block third-party advertising are very effective at reducing third-party tracking. These web privacy tools do allow some cookies to exist however. This does not mean that these tools are allowing tracking. Recall that cookies have more uses than only tracking. Blocking cookies that are not involved in tracking could result in significant usability problems. Imagine if the cookies used to facilitate the implementation of an online shopping cart were blocked. If this were to happen, the online shopping cart would no longer be operable. The most efficient configuration of a web privacy tool would reduce tracking, while simultaneously causing little to no effect on the usability of a website. If a tool is configured too strongly to block tracking, usability will suffer. If a tool is configured too strongly to promote usability, there will not be as much of a reduction in tracking. Finding this middle ground is often hard to accomplish and is why we see differences in performance across the web privacy tools we studied. Due to minor configuration differences in each of these tools, each tool will have a slightly different balance of usability and tracking reduction. In our own research, we noticed that the functionality of some minor parts of a website was often affected by our own (mostly default) configuration of these web privacy tools.

We also conclude that these web privacy tools are very effective in contributing to a more pleasant user experience on the internet through faster page load times. It is also noteworthy that these web privacy tools also visually remove advertisements which some internet users consider "nonsensical, uninformative, forgettable, ineffective, and intrusive" [13].

In the future, plenty of improvements could be made on the current state of this project. Due to the time constraints on this project, there is room for improvement in the efficiency of python and Java code used to control the VM's and to collect data. More efficient code would result in faster data collection times and also promote the readability of the code. The code could also be made more robust by employing better error handling and eliminating hard-coding. This would allow the code to run more smoothly and also enable custom configuration as a general web crawler. In addition to this, the process of data collection and data parsing could be combined into one to save time and reduce the manual workload. Extra features could also be added to the code to complement its current support of text message or email alerts when errors are encountered and when the time consuming process of data collection has finished.

This research could also be extended in terms of data collection. One could analyze more than just 100 websites at a time in an effort to obtain a more comprehensive picture of how web privacy tools affect users. Additional data could be collected by reconfiguring our existing web crawl system. This data could be used to answer different questions about the state of the internet. Specifically, session cookies could be collected and analyzed. Lastly, additional web privacy tools could be implemented and their effects analyzed.

In its current state, our virtual machine system is very complicated and sometimes cumbersome to use. Future work could include streamlining this system to promote easy use, maintenance, and modification. This system could also be modified to automate the virtual machine creation process, so that more virtual machines could be added with little effort. Lastly, a very useful tool for this system would be remote access. This system uses large amounts of RAM and processor time and was deployed on a machine with one of the fastest processors on the consumers market and with an extraordinary amount of RAM. Being able to remotely access this machine from anywhere in order to collect data at any time would be a great benefit to an already impressive system.

## Acknowledgment

This work was supported in part by TRUST, Team for Research in Ubiquitous Secure Technology, which receives support from the National Science Foundation (NSF award number CCF-0424422).

Special thanks to Chris Hoofnagle, Nathan Good, Aimee Tabor, the TRUST staff, and other REU interns, without which this project would not have been possible.

## REFERENCES

[1] A.R.A. Bouguettaya and M. Y. Eltoweissy, "Privacy on the Web: facts, challenges, and solutions," Security & Privacy Magazine, IEEE, vol. 1, pp. 40-49, 2003

[2] Ghostery. About Ghostery. [Online]. Available: http://www.ghostery. com/about

[3] Weihong Peng, Jennifer Cisna, "HTTP cookies – a promising technology", Online Information Review, Vol. 24 Iss: 2, pp.150 – 153, 2000

[4] Mayer, Jonathan R., and John C. Mitchell. "Third-party web tracking: Policy and technology." *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 2012

[5] Nah, Fiona Fui-Hoon. "A study on tolerable waiting time: how long are web users willing to wait?." *Behaviour & Information Technology* 23.3 (2004): 153-163.

[6] Li, Wen-Syan, Wang-Pin Hsiung, Dmitri V. Kalashnikov, Radu Sion, Oliver Po, Divyakant Agrawal, and K. Selçuk Candan. "Issues and evaluations of caching solutions for web application acceleration." In *Proceedings of the 28th international conference on Very Large Data Bases*, pp. 1019-1030. VLDB Endowment, 2002.

[7] King, Andrew B. *Speed up your site: web site optimization*. New Riders, 2003.

[8] C. J. Hoofnagle and N. Good. (2012 October) The Web Privacy Census. [Online]. Available: http://law.berkeley.edu/privacycensus.htm

[9] C. J. Hoofnagle & N. Good. (2012 October) The Web Privacy Census Methods. [Online]. Available: http://www.law.berkeley.edu/ 14457.htm

[10] J. Mayer. (2011, September) Tracking the trackers: Selfhelp tools. [Online]. Available:http://cyberlaw.stanford.edu/node/6730

[11] Schmücker, Niklas. "Web Tracking." (2011).

[12] P. G. Leon, B. Ur, R. Balebako, L. F. Cranor, R. Shay, and Y. Wang, "Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising," Carnegie Mellon CyLab, Tech. Rep. 11-017, October 2011.

[13] McCoy, Scott, et al. "The effects of online advertising." *Communications of the ACM* 50.3 (2007): 84-88.