PrivacyTracker: A User-Centered Browser Plugin to Understand Online Tracking

Ashwin Srinivasan

Pittsburgh Allderdice High School

Project No.: SCM116

Category: Computer Science/Math

Table of Contents

Abstract

Online privacy is a growing concern to users. Most popular websites partner with third parties to track users' activity across the Internet in order to target advertisements. The majority of users are not comfortable with these practices. Existing privacy tools have been ineffective in helping users understand online tracking and make privacy decisions because the information they show is unintelligible to users. The design hypothesis of this project was that longitudinal, personalized data about online tracking would be more helpful to users than only the names of tracking companies. A Google Chrome browser plugin was developed in this project to present personal data-driven information, which is not available in existing privacy tools. The design of the plugin was informed by a survey conducted to assess users' knowledge of web privacy and responses to possible representations of tracking. The plugin was able to display accurate, personalized information about how third parties track users and was more effective than existing commercial software.

Introduction

Online privacy is a growing concern to users. The majority of websites on the Internet use online tracking and behavioral targeting, which involve utilizing user browsing activity to target tailored advertisements and content to individuals [1]. However, studies have shown that users are not comfortable with third-party data collection practices [2] such as selling consumer data and analyzing user interests. If given a choice, most users would block online tracking companies [3].

Currently, there are several tools that allow consumers to control online data collection, such as opt-out cookies, AdChoices, Do Not Track, and privacy browser extensions [1]. However, these utilities have been ineffective due to confusing interfaces and lack of actionable of information about the tracking ecosystem. Their model of tracking is fundamentally incorrect; the

data provided is focused on tracking occurring at one instant, rather than longitudinally. Users desire an intuitive privacy tool that provides personalized information about tracking companies as well as mechanisms to control online tracking [2].

Goal

The goal of this project was to develop a Google Chrome privacy plugin to assist users in understanding online tracking by providing user-centered information about tracking companies, which is not present in existing privacy tools. The longitudinal data presented by the plugin about precisely when a user has been tracked and what companies may have inferred about them allows users to make informed decisions about their online privacy. In order to determine the most useful types of information about online tracking, a survey was conducted to assess users' responses to possible representations of tracking. The results of the survey guided the information presented by the plugin and provided empirical evidence to support the plugin's design.

Background Research

Online Tracking

Online tracking is the practice of collecting user data from first-party websites through authorized third parties [1]. It is used by most of the top websites on the Internet [3]. An example of online tracking is when an advertiser tracks a user's visit to a webpage about "travel destinations" and displays an advertisement for travel destinations when the user visits an unrelated webpage, targeting the same user.

There are three types of tracking companies: advertising services, social networks and analytics companies. Advertising services target advertisements to users based on the estimated audience of a publisher, the content of a webpage, or user browsing activity [1]. Contextual advertising, where an advertisement is based on the content of a webpage, is the simplest example

of a tailored advertisement. A more complex method is online behavioral advertising (OBA), which profiles users based on online activities and uses this data to determine the advertisements that interest them. This technique can significantly increase the click-through rate of advertisements [3].

The second of the three types of tracking companies are social networks, which offer various forms of integration to websites that can be embedded in webpages; the data collected from these tools is often sold for advertisement targeting.

Analytics companies allow websites to understand their visitors by providing data about demographics, web browser user agents, and user interactions with various forms of content. Some analytics companies monetize their service by collecting payment from website clients, but the majority of analytics companies offer free services that are monetized by selling the data collected to advertising networks. Content providers have the same business model, where a free content hosting service is monetized by selling the data they collect from users.

Behavioral Targeting

Behavioral targeting involves following the activities of an individual user over time and using the data to target tailored advertisements and content [4]. Websites can purchase data about their audience to target appealing advertisements and content, and advertising companies themselves buy data to develop a complete profile of users [5]. Although tracking services cannot buy personally identifiable information about consumers, they analyze user activity to discover consumer interests.

A variety of technologies are used in behavioral targeting, enabling companies to collect data about users and target advertisements, provide content, or sell data. Stateless tracking, or "fingerprinting", generates a nearly unique identifier from the user's web browser. It is estimated that 83.6% of web browsers have been identified with fingerprinting techniques. However, stateful

tracking, also known as "supercookies,", where a unique pseudonymous identifier is placed on a user's computer [1], is the primary behavioral targeting method to track users over time. Stateful tracking can be implemented through Adobe Flash cookies and HTML5 local storage.

Critics of behavioral targeting believe it is wrong to collect data about users without their consent. Privacy advocates have called for an opt-in model where users consent to tracking before they become subject to tailored advertisements or content. This approach has been dismissed as being unrealistic by advertising firms, who have defended behavioral targeting by claiming that Americans actually want tailored advertisements, which can only be provided by profiling users. Websites often claim that they must sell user data in order to buy content [4].

Past research has shown that approximately 50% of users find advertisements useful [3]. On the other hand, when informed of online behavioral advertising techniques, 60-80% of participants in one study were opposed to it [4]. Users have described behavioral targeting as "creepy" and "scary" [3].

Existing Privacy Tools

Currently, there are several tools that enable consumers to control data collection, including privacy browser extensions, opt-out cookies, and Do Not Track [1]. However, these utilities have proven ineffective due to confusing interfaces and lack of information about trackers [2].

Privacy browser extensions such as Ghostery and Disconnect function by identifying HTTP requests to trackers listed in a tracking protection list (TPL) [1] and allow users to block specific trackers (Figures 1-2). For example, Ghostery, which has more than 300,000 Mozilla Firefox users [5], merely shows a list of "tracker source URLs" for trackers connected to each webpage. Disconnect suffers from the same flaws; it only shows the number of requests to each tracking company, which the majority of users cannot understand. Users do not find this information useful or intelligible, and might wonder why they are being tracked and what the

tracker knows about them. In addition, current privacy browser extensions overwhelm users with configuration options and misleading interface elements [2].

Figure 1: Information for the advertising company "DoubleClick" shown in Ghostery after visiting *http://www.msn.com*. Rather than actionable information, the data shown is merely a meaningless list of tracker URLs.

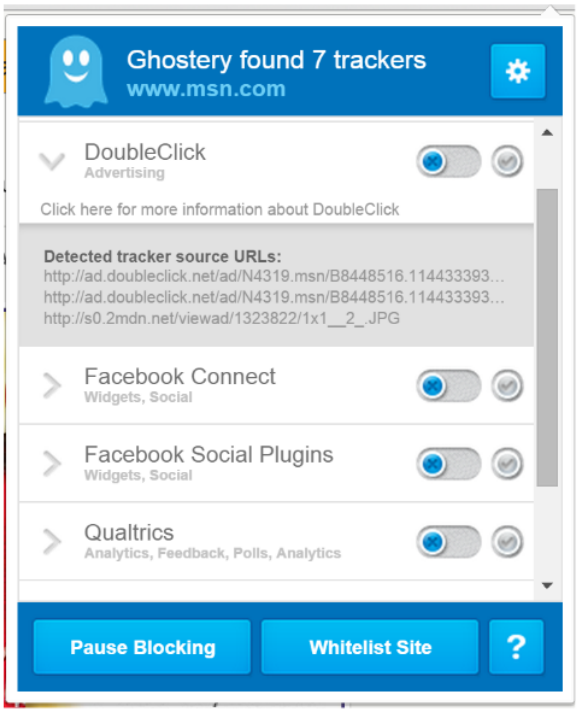
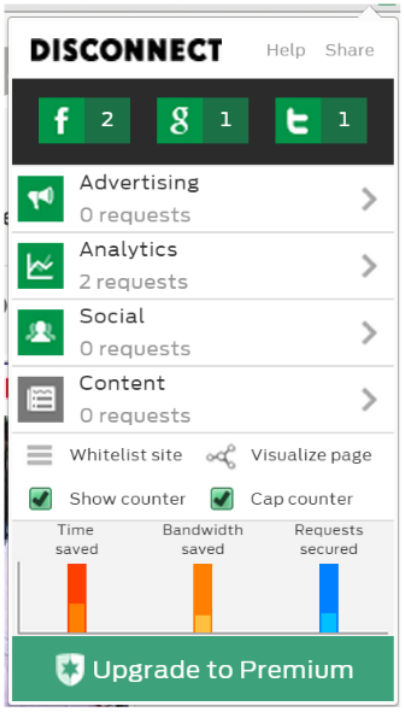


Figure 2: Main window of Disconnect.me plugin after visiting *http://www.msn.com*. Very little intelligible and actionable tracking information is given to the user.



Opt-out cookies disable tracking from a particular analytics or advertising company [1]. Tools that set universal opt-out cookies such as PrivacyMark have confusing interfaces involving jargon unfamiliar to average users [2]. Do Not Track (DNT) is a voluntary standard developed by the World Wide Web Consortium (W3C) that utilizes a HyperText Transfer Protocol (HTTP) header containing online tracking preferences [1]. The standard has not been widely adopted by advertisers and fails to limit behavioral targeting [5].

Overall, users desire an intuitive privacy tool that provides personalized information about trackers and mechanisms to control behavioral targeting [2].

Plugin Design

Design Hypothesis

The design hypothesis of this project was that data-driven, longitudinal information about online tracking companies is more relevant to users than information about tracking occurring at one moment that is shown by existing privacy software. Existing privacy plugins show only a list of tracker names connected to a webpage, which is meaningless to users. To the average user, this does not give enough information to decide whether to block each company. On the other hand, the plugin developed in this project shows a list of webpages that each tracker has used to collect data about the user as well as user interests inferred by each tracker. This longitudinal data was based on user browsing activity and mirrors the data collected by third parties.

User Survey Methodology

In order to inform the design of the privacy plugin, a survey was conducted to assess users' knowledge of online tracking and perceptions of various representations of data collection hypothesized to be helpful to users. Participants were recruited through Amazon Mechanical Turk.

Mechanical Turk is a web-based platform for recruiting subjects, known as "workers" [6], to complete short human intelligence tasks (HITs) [7]. Workers are compensated for each task they complete [8]. Mechanical Turk was selected because it provides a diverse group of survey participants [9], and inherently requires participants to be Internet users [7]. Previous research has shown that participants recruited through Mechanical Turk are more representative of the general population than subjects from traditional recruitment methods [6] and are more diverse than American undergraduate student samples typically used in research [8].

The privacy survey conducted in this project took 10-15 minutes to complete and consisted of 43 questions, hosted on the SurveyGizmo website. The 50 participants were paid $0.80. Duplicate responses were prevented by ensuring that the IP address of each participant was unique [10]. The word "privacy" was not mentioned in the survey to avoid participant bias [11].

The survey contained three sections, and each section addressed a different aspect of online tracking. The first section asked participants to rank the accuracy of 11 statements concerning behavioral targeting from "Definitely False" to "Definitely True." Questions from this component were based on validated questionnaires from previous studies [12] and some questions were repeated to verify participants' attention.

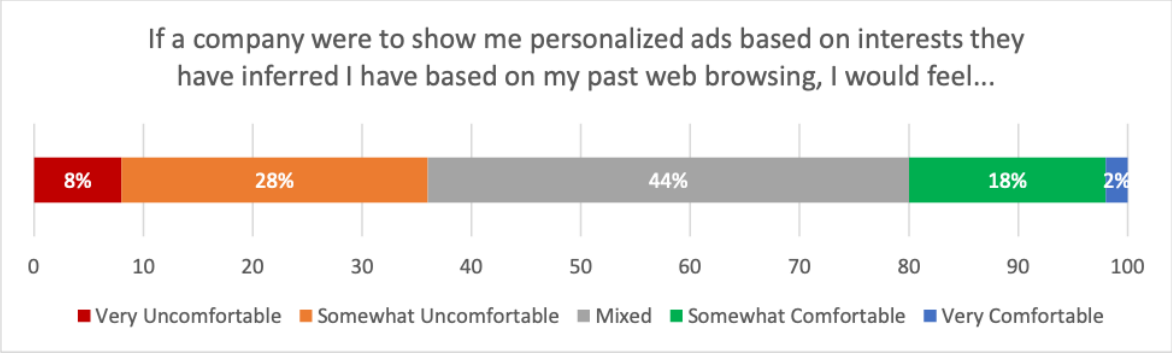
The second part of the survey was designed to determine participants' perceptions of behavioral targeting. Five questions asked participants to rate their comfort with examples of online tracking from "Very Uncomfortable" to "Very Comfortable." For three other questions, participants rated their desire for tailored advertisements, content, and discounts on a scale from "I Definitely Don't Want This" to "I Definitely Want This."

The final component of the survey was directly related to the design of the plugin. Participants were shown possible representations of online tracking in a screenshot of a hypothetical privacy tool that consisted of a header and information about tracking. Participants were shown 12 screenshots of how a privacy plugin might appear in randomized order, each with a different type of information about tracking that was hypothesized to helpful to users. To ensure that the personal interests of participants did not interfere with their responses, participants were asked to assume that they were interested in "Clothing, Food, and Travel." For each sample screen, participants rated the usefulness of the screenshot from "Very Useless" to "Very Useful" as well as their likelihood to block the tracking company from "Very Unlikely" to "Very Likely."

User Survey Results

The results of the survey showed that the majority of participants were aware of online tracking and behavioral targeting practices. A total of 98% of participants correctly believed that analytics and advertising companies track them when they visit a website. In addition, 96% of participants understood that advertising companies "probably" or "definitely" use their interaction with content to tailor advertisements. However, 80% of participants were uncomfortable with or had mixed feelings about online behavioral advertising (Figure 3). Approximately 50% of participants were "somewhat" uncomfortable with third-party tracking and none were comfortable with selling user data. In addition, 74% of participants did not want or were unsure about tailored advertisements.

Figure 3: Participants' attitudes towards online behavioral advertising.



In participants' opinions, the least useful screenshot of a hypothetical privacy plugin was the control group (Figure 4). Notably, this control group was most similar to existing privacy tools focusing on tracking occurring instantaneously, rather than longitudinally. On the other hand, participants found other types of tracking information to be more useful. For example, 66% of participants thought that a list of webpages the tracker knows the user visited was useful. The list of user interests inferred by the tracker used to target the user was also useful. The most useful information was the possible future action of a tracking company (Figures 5-7). In addition,

providing the reason a tracker is connected to a first-party webpage was influential in whether the participant would block the tracker.

Figure 4: Usefulness of control group screenshot which was similar to existing privacy tools that focus on tracking occurring at one moment, rather than longitudinally. According to survey participants, this was the least useful information about online tracking.

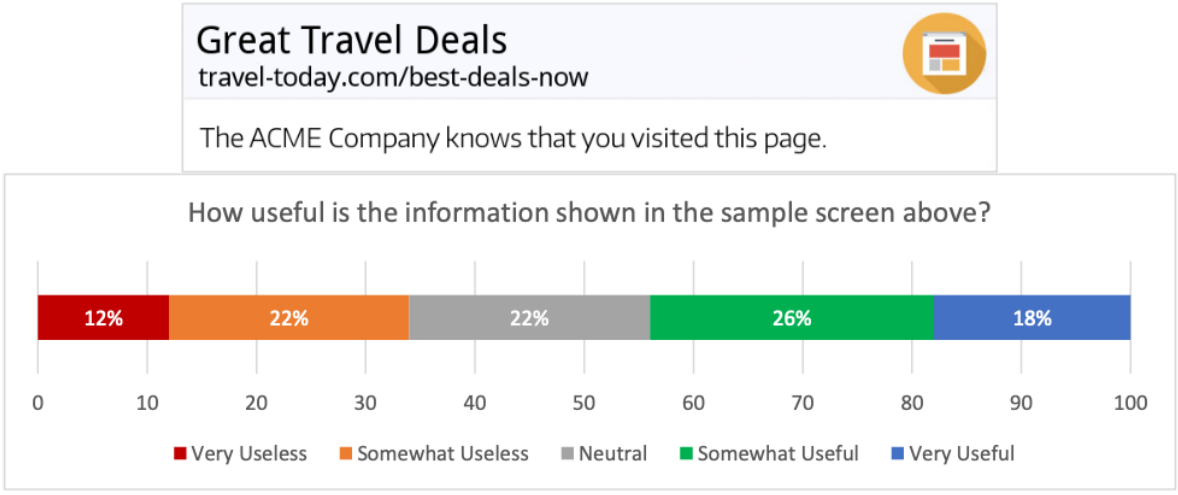


Figure 5: Participants' responses regarding usefulness of screenshot showing webpages that a tracker knows the user visited.

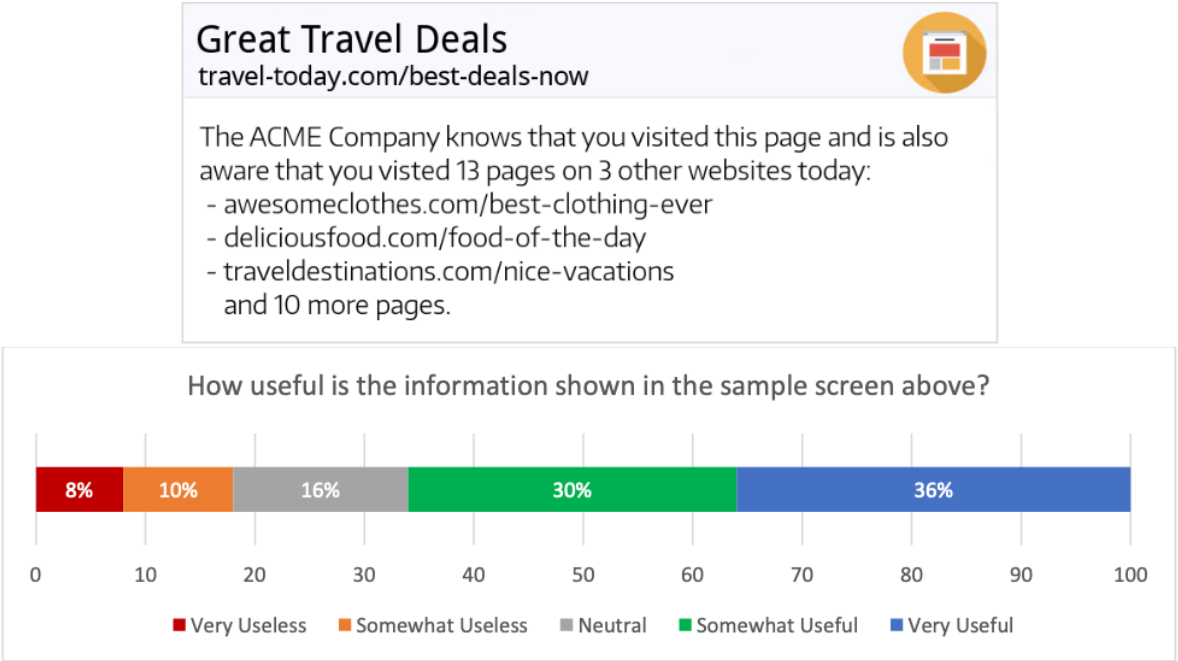


Figure 6: Participants' responses regarding usefulness of screenshot showing user interests inferred by a tracker.

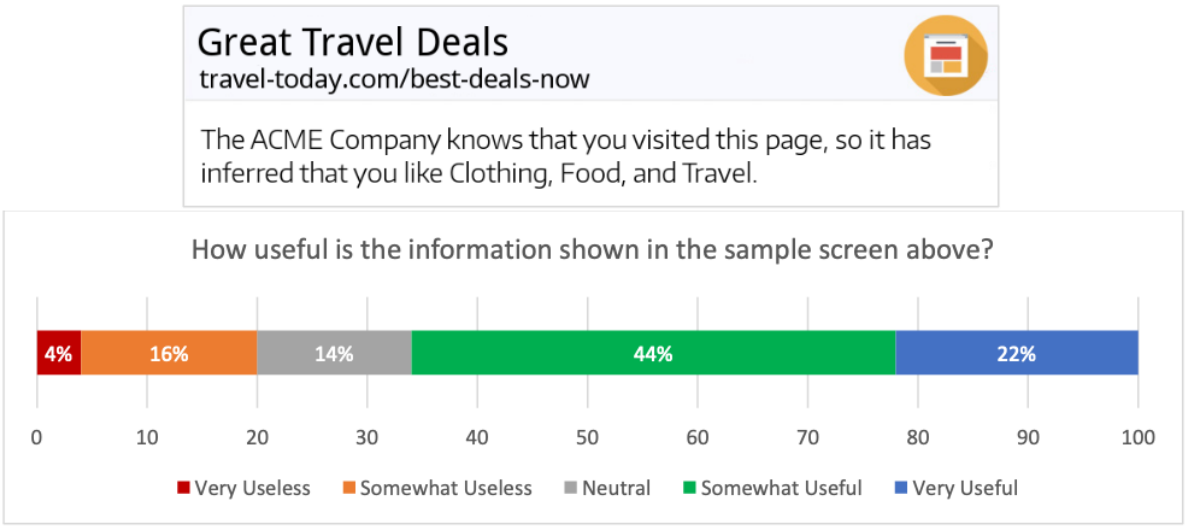
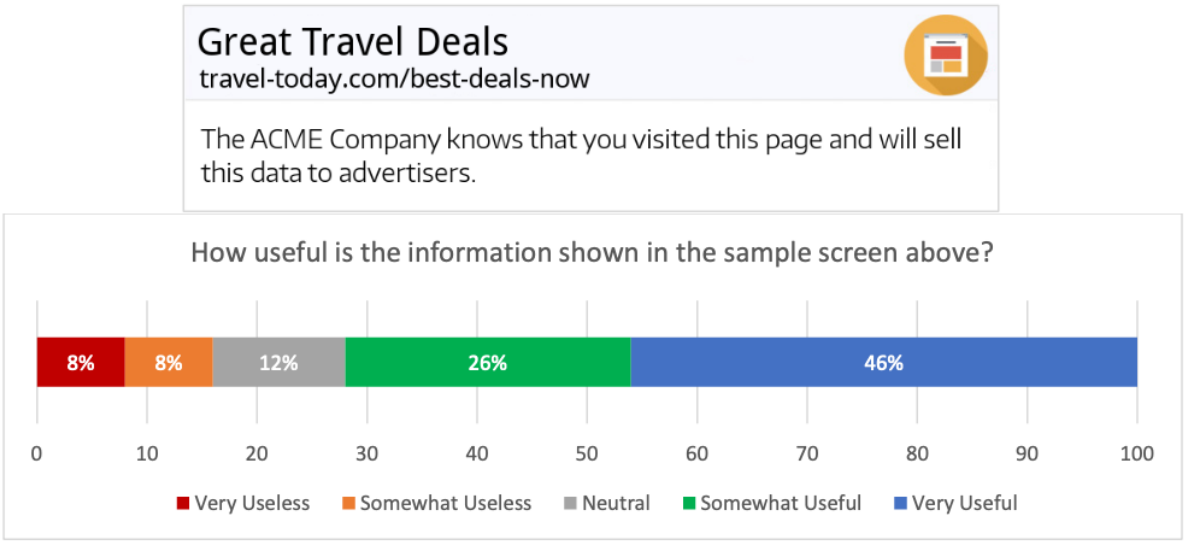


Figure 7: Participants' responses regarding usefulness of screenshot explaining that a tracker will sell user data to advertisers.



Design Criteria

A privacy plugin was developed that builds on the survey results indicating that participants considered longitudinal, personal data-driven more useful than the tracking data displayed by current privacy tools. The plugin developed in this project, called "PrivacyTracker," offers

personalized information about how a user's data has been collected by tracking companies, alongside explanations of the tracking ecosystem that are designed to be as straightforward as possible.

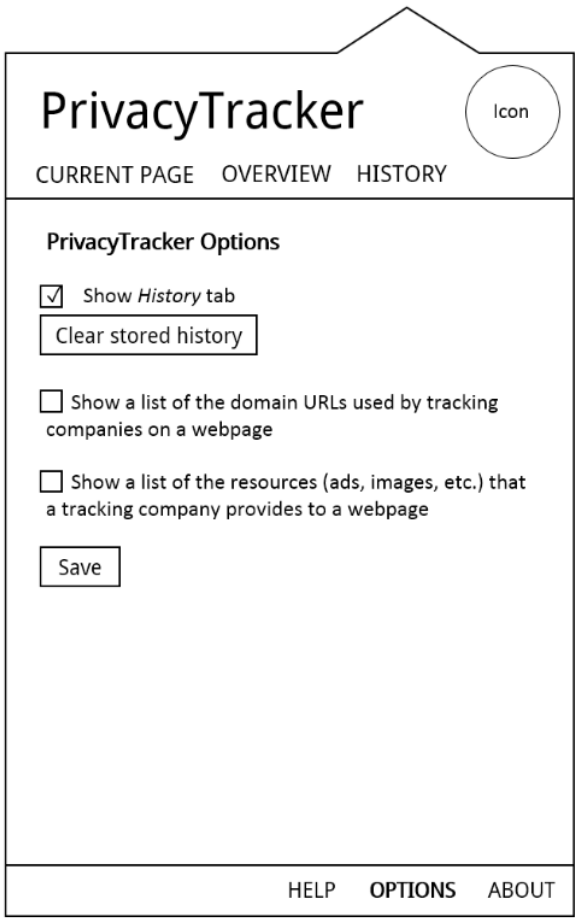
According to the majority of survey participants, the most useful information about online tracking included webpages a tracker knows the participant visited and interests inferred about the participant. In addition, since the reason a third-party tracker is connected to a first-party website was an influential factor in whether survey participants would block the tracking company, the plugin grouped trackers into four categories for tracking companies: analytics services, advertising companies, social networks, and content providers. To avoid privacy jargon, the title of the category was written as a description in non-technical, easily understandable language.

Before implementing the plugin, design wireframes were created to show the placement of the user interface elements and text in the plugin window (Figure 8). The plugin consists of three tabs: “Current Page,” “Overview,” and “History.” When a user opens the plugin, the Current Page tab shows the trackers connected to the webpage that the user is currently visiting. Users can click on a tracker to display its tracking profile, which includes information about the number of HTTP requests made to the tracking company, the list of websites that the tracker also knows the user visited, and the tracking company’s inferred user interests. Based on this data, the user can make an informed decision regarding whether to block or enable the tracker. The Overview tab shows all tracking companies that are profiling the user, and the History tab provides a list of all webpages the user has visited. The user can click on a history entry to load that webpage and view tracking information about it. At the bottom of the window, plugin help, customization options, and a description of the plugin are available to the user (Figure 9). These wireframe designs, as well as

plugin performance and accuracy, were used as the design criteria to test the plugin in ten trials after implementation.

Figure 8: Design wireframe of tracker information shown in plugin.

Figure 9: Design wireframe of options mode in plugin window.



Plugin Implementation

The privacy plugin in this project was developed for Google Chrome, the most popular web browser globally [13], in order to make it available to the largest number of web users.

The plugin was implemented using the JavaScript programming language as a cross-platform Chrome extension. The user interface was written in HyperText Markup Language (HTML) and Cascading Style Sheets (CSS). Chrome Application Programming Interfaces (APIs)

were used to communicate with the browser for webpage and tab data. The tracking protection list from the open-source Disconnect utility was included in the plugin to categorize trackers [14]. The plugin contains 2560 lines of code. All code was written only by the author of this paper.

The plugin creates databases for history, trackers, and options that are stored in Chrome's SQLite settings database. When the user loads a webpage, the plugin's background script records the webpage in the history database and logs all HTTP requests. If a user has blocked a tracking company that is detected in an HTTP request, the script uses Chrome's webRequest API to cancel the request. A unique identification number is given to each database entry associated with a particular webpage so that history entries can be linked to tracker information. The event listener that is called for webpage loading is an asynchronous function, so individual browser tabs can be logged and processed simultaneously.

The plugin's popup script is responsible for loading the main window when the user clicks the plugin icon in the Chrome toolbar. When the popup script is called, it checks if the user is connected to the Internet. If this condition is satisfied, the script uses Chrome's storage API to load history and tracker databases. The tracking protection list is loaded from a JavaScript Object Notation (JSON) file to into the script and the plugin checks to ensure that all data has been loaded correctly. However, since the storage API runs asynchronously, synchronous data verification would always fail because it would execute before the data is loaded. To prevent this, the data verification is nested in the callback for the storage API call, allowing the plugin to load smoothly.

After all data has been loaded, the plugin categorizes requests to trackers (Algorithm 1). This process involves matching HTTP request addresses to entries in the tracking protection list, creating a JavaScript object for each tracker. Each category that contains at least one tracker is displayed in the plugin window, colored red with color intensity based on the number of trackers.

Algorithm 1: Algorithm used to collect and categorize HTTP requests to tracking companies. Trackers were categorized into analytics, advertising, social network, or content provider.

```
for each HTTP request URL in all HTTP requests for this page, do
    for each tracker name in tracking protection list, do
        if tracker URL equals HTTP request URL, then
            if tracker name not found in tracker category, then
                Create entry tracker name in tracker category
            Add tracker URL to tracker name in tracker category
    end
end
```

When a category is clicked, the trackers in the category are shown with their icons and the number of webpages the tracking company knows the user visited. Trackers that use unsecure requests are shown in red, and disabled trackers appear in gray. Clicking on a tracker shows its tracking profile, including the other pages the tracker knows the user visited (called "tracked websites") which are determined using an algorithm that collects the identification numbers of entries in the history database that have an HTTP request with the same tracker (Algorithm 2). This is the longitudinal information displayed in the plugin that is not found in existing software.

Algorithm 2: Algorithm used to determine other webpages that a tracking company is aware that the user visited in the past.

```
for each HTTP request URL in all HTTP requests for all pages, do
    if current tracker URL equals HTTP request URL, then
        Add HTTP request ID to tracked websites
end
```

For each tracked webpage, the plugin uses topic data from Amazon's Alexa service to determine the interests extracted from that page, creating a list of interests used by the tracker to target the user. Alexa is also used to retrieve the country of origin and popularity ranking for each tracker. After reviewing this information, the user can elect to block the tracker.

Plugin Testing

The fully implemented plugin is shown in Figures 10-11. Figure 10 shows the profile for "DoubleClick," an advertising company, after visiting three popular webpages. Figure 11 shows all companies tracking the user, allowing them to understand how their data is being collected.

Figure 10: Tracking profile for advertising company DoubleClick in plugin after visiting *mlb.com*, *flipboard.com*, and *msn.com* shows what DoubleClick knows about the user, including two webpages visited by the user and four user interests.

Figure 11: Overview tab of plugin after visiting *mlb.com*, *flipboard.com*, and *msn.com* shows list of companies tracking the user, allowing users to understand what third parties know about them overall.

PrivacyTracker clearly describes how advertising companies target tailored advertisements to the user and provides personalized information about tracking. In Figure 10, the plugin explains that DoubleClick is aware that the user visited MSN and MLB websites and has inferred that the

user is interested in Recreation and Sports, Baseball, and has two other interests. This actionable, longitudinal information is not found in any privacy tool on the market today.

The plugin was tested in ten trials where the trackers shown in the plugin were compared to trackers shown in Ghostery and Disconnect after visiting three of the top 500 websites on the Internet [15]. In each of the ten trials, the plugin detected the same 16 trackers as existing privacy tools after visiting the three webpages. The plugin was also able to successfully block trackers.

Discussion

A privacy plugin was developed in this project to display longitudinal information based on the user's own web browsing being tracked by third parties, which is a feature not found in existing privacy tools. The design of the plugin incorporated the results of a survey that assessed users' knowledge of data collection as well as perceptions of various representations of tracking.

As shown in previous studies, survey participants in this project also possessed a basic knowledge of online data collection practices, were uncomfortable with behavioral targeting, and wanted user-centered information about online tracking [2, 3, 4].

The plugin allowed users to understand online tracking on a webpage, empowering them to make informed privacy decisions and take control of their private information on the Internet. The fully implemented plugin was effective in displaying a list of trackers for popular websites and providing detailed information about each company. When a user blocked a tracker, the plugin was able to cancel all HTTP requests to that tracker. As a result, webpage elements dependent on the blocked tracker such as advertisements or images were not shown.

The interface of the plugin is more intuitive than existing privacy tools, with less technical vocabulary to explain online tracking. After the user has visited multiple pages that have a certain tracking company, the plugin displayed additional information about that tracker including tracked

webpages and interests, supporting the design hypothesis. On the other hand, existing privacy tools including Ghostery and Disconnect merely display a list of tracker addresses.

The plugin could be further improved by enhancing the loading time of the plugin by rewriting the popup script to use multiple processor cores on the user's computer. Additional improvements include a tutorial for first-time users and encryption of data stored locally. To increase the accuracy of survey results, a larger group of participants could be surveyed with questions about more possible representations of tracking. In the future, a user study can be conducted to evaluate the usability of the plugin and verify that users find the plugin intuitive and useful in their daily use. A future goal for the plugin is to release it in the Chrome Web Store.

The application of the privacy plugin developed in this project is for the majority of users who are not comfortable with online tracking. The plugin satisfies users' desires for privacy tools that are not found in existing software. Protecting user data from third-party tracking can ultimately prevent cybersecurity threats and risks such as identity theft and phishing attacks [1, 16].

Conclusion

In conclusion, the privacy plugin developed in this project presented personal data-driven information about tracking companies in an intuitive interface, allowing users to make informed decisions about whether to block trackers. The results of the survey used to inform the design of the plugin showed that users are aware of behavioral targeting practices and a significant number of users are not comfortable with online tracking. The implemented plugin supported the project design hypothesis and provided features not found in existing privacy software. Future studies to further validate the design hypothesis can be conducted to test the usability of the plugin, which can be improved with performance optimizations. Overall, the plugin empowers users to take control of their personal data and can be used by consumers who are opposed to online tracking.

References

1. Mayer, J. R., & Mitchell, J. C. (2012). Third-Party Web Tracking: Policy and Technology. *Proceedings of the IEEE Symposium on Security and Privacy*, 413-27.
2. Leon, P. G., Ur, B., Cranor, L. F., Shay, R., & Wang, Y. (2012). Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 589-598.
3. Ur, B., Leon, P. G., Cranor, L. F, Shay, R., & Wang, Y. (2012). Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS)*.
4. Turow, J., King, J., Hoofnagle, C., Bleakley, A., & Hennessy, M. (2009). Americans Reject Tailored Advertising and Three Activities that Enable It. *Social Science Research Network.*
5. Balebako, R., Leon, P. G., Shay, R., Ur, B., Wang, Y., & Cranor, L. F. (2012). Measuring the Effectiveness of Privacy Tools for Limiting Behavioral Advertising. *Proceedings of the IEEE Symposium on Security and Privacy*.
6. Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. Political Analysis, 20(3), 351-368.
7. Mechanical Turk Overview. (n.d.). Retrieved December 27, 2014, from https://www.mturk.com/mturk/help?helpPage=overview
8. Buhrmester, M. D., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5.
9. Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411-419.
10. About SurveyGizmo. (n.d.). Retrieved December 9, 2014, from http://www.surveygizmo.com/company/about/
11. Braunstein, A., Granka, L., & Staddon, J. (2011). Indirect Content Privacy Surveys: Measuring Privacy without Asking about It. *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS)*.
12. McDonald, A. M., & Cranor, L. F. (2010). Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising. *Research Conference on Communications, Information and Internet Policy (TPRC)*.
13. Nelson, K. (2014, June 5). Chrome Is Now More Popular Than Internet Explorer. Retrieved June 12, 2014, from http://mashable.com/2014/06/05/chrome-popular-internet-explorer/
14. Tracking Protection. (n.d.). Retrieved December 28, 2014, from https://disconnect.me/moreprivate
15. Top 500 Global Sites. (n.d.). Retrieved December 28, 2014, from http://www.alexa.com/topsites
16. Cybersecurity Privacy Practical Implications. (n.d.). Retrieved February 15, 2015, from https://epic.org/privacy/cybersecurity/