Visualizing Trackers in Browsing Data Using PrivacyBadger

Aryan Srivastava
Brown University

Ragna Agerup
Brown University

William Kuenne
Brown University

1. Abstract

It has become increasingly difficult for an average person to understand how their data is being used on the internet. On average, 79% of websites globally are secretly tracking you [1]. In addition to this, trackers are actually invisible to the user on the pages they appear, providing a misperception of what information is actually out there. Ghostery, a browser extension that aimst to tackle the problem of showing and blocking invisible trackers, have reported a total of about 2,000 trackers on the web [2]. We provide a way for users to collect data about the trackers and in addition, we have created a website to visualize the number of trackers found, on what websites they appear, and who they are. Lastly, we have provided a way for users to permanently opt-out of being tracked by the most notorious trackers found on user's profiles, in just one click.

2. Introduction

Privacy online is almost non-existent, transparency in data collecting by companies is equally so. User's online activity is being tracked, and stored as data and information which is then sold to deliver personalized ads. Transparency for the users is lacking, and an evening searching the web leads to hundreds of trackers that are now collecting information about you.

There are currently several tools that help you block trackers on the web, some of these being Privacy Badger, Ghostery and more recently browsers such as Firefox and Chrome. These tools seek to solve some of these problems, such as blocking unnecessary trackers and showing what trackers appear on the different websites. However, they don't provide a mechanism to understand the bigger picture of where these trackers come from and how to opt out of giving them their information.

Privacy Badger allows users to download the data it stores about their browsing history. This can help the user understand what data is actually displayed. However, the data is stored as a json file. A json file is almost impossible to read, and only ever convenient in scenarios when there will be analysis, cleaning and pretty printing of the file. In our case, on the other hand, it has proved to be a good attribute to tackle our problem, and show the users the answers they are really looking for.

We aim to provide a mechanism for users, and potentially researchers, to use a modified version of Privacy Badger to collect data about their browsing habits, and provide a simple and user-friendly website to display visualizations summarizing the presence of trackers in the user's activity online.

3. Background

When a user clicks on a website, let's say nytimes.com, there are a bunch of cookies on that page. To be able to read and access material, you accept the cookies. This is visible to you. What you don't visually accept, is trackers, which are still installed on that website to track your behaviour. These trackers track your behaviour, the time you spend on different pages and certain links that you click on.

Now your information is shared with that company, and trackers are beginning to slowly gather information about your browsing. Many companies even do cross-web-tracking, which means that they are tracking your actions across several websites, to build a more in-depth

picture about what type of user you are. All to be able to find more ads that can target your profile. We aim to break this down for you and provide an easy way to know and understand who they are and how to get out of it.

However, none of these web extensions aim to solve the problem of visualizing to the user who these trackers are, where they are and how they control your searches online. We have taken matters into our own hands to give the user the transparency that they need.

4. Design

4.1 Using the Modified PrivacyBadger Extension to Collect Data

The user first needs to install our unpackaged modified PrivacyBadger extension. There are instructions to do so on the Github that are simple and straightforward. Once a user has installed the modified extension on the browser of their choice (Chrome or Firefox), they need to delete the pre-trained seed data that PrivacyBadger uses from the settings window. They also need to enable learning based on browsing, and disable the sending of "Global Privacy Control"[3] and "Do Not Track"[4] signals. After following these simple steps, PrivacyBadger will start collecting data about the user's browsing and the trackers it witnessed.

When the user is ready to download their data (for example, after they have visited the specific websites they wish to get data for), they can simply export user data from the "Manage Data" tab in the settings of the extension. This will download a .json file including a list mapping every tracker that was found during the user's browsing to a list of websites it was found on.

4.2 Visualization Website

The visualization aims to break down the data in a clear way so that the user has a complete understanding of what websites and trackers do. We created a web page in React.js that allows the user to upload their downloaded Privacy Badger file, which then parses the data and displays it.

We have implemented several features that make it easy to break down the data and show the user how tracking works. Our first component is a bar chart that sorts the websites combined with their trackers and displays the top 10 websites with the most trackers on them. This will give the user a clear picture of what websites track a lot and how many trackers you can find on these pages.

The next component is a list of these top trackers where we break down the details of the tracker. Here we can see the list of trackers for each website, in order, sorted from the highest number of trackers to the lowest number of trackers.

Lastly, we have implemented a bubble chart that maps the number of occurrences of each tracker on the websites visited by the user. This shows what tracker does the most tracking on you. In addition, we have created a database with the most common trackers and their opt out links, and thus, clicking one of the bubbles will direct you to that tracker's opt out page.

5. Implementation

Our implementation consists of two components - the modified Privacy Badger extension that collects the user's data, and the visualization website that breaks down and shows that data.

Our project has been implemented in React.js using JavaScript, HTML and CSS. On the back-end, we created a server where the file uploaded by the user is posted. On the front-end, the file is fetched and displayed to the user through various visualizations. We use express to create the server, and multer to handle the file that is uploaded. We used Nodemon to monitor changes that auto-reloads when changes are made. This makes it so that we don't have to restart the server manually each time.

On the front-end side, we have used react bootstrap for styling and react-router-dom to connect the html pages for a smooth transition. Additionally, for the visualization, we have used d3 which is another JavaScript library that helped us create the different visualization components and made them responsive, such as the bubble chart displayed below in Figure 1.

We wrote data processing functions that obtained useful lists and statistics from the 'snitch_map' stored in the .json user data file collected from the modifier Privacy Badger. Figure 2. Shows an example of an entry in the snitch_map. The key stores the tracker and the value stores a list of all the websites it was found on.

Figure 1. The bubble chart visualization of trackers with opt-out links on our website

Figure 2. An example entry in the snitch_map

5.1 Modifying Privacy Badger

Modifying Privacy Badger to work as we wanted it to required changing only one line of code! Moreover, the change was only changing the value of the "TRACKING_THRESHOLD" variable in the constants.js file from 3 to some high value; we changed it to 9999999.

This worked because Privacy Badger blocks a tracker that has been observed tracking on more than TRACKING_THRESHOLD websites. If a tracker is blocked, any new websites it is found on aren't added to the collected data. A blocked tracker, if it is a dynamic tracker, also cannot load other trackers. Therefore, making sure Privacy Badger never blocks anything by making TRACKING_THERSHOLD a high value fixes both of these issues.

Lastly, we got a lot of help from one of the developers of Privacy Badger through their Github repo's issues page to clarify what we needed to change in order to get our desired result. [5]

6. Evaluation

We collected data from a browsing session on multiple news websites. Visualizing the downloaded data from the modified Privacy Badger on our website yielded interesting and insightful results. We found out the most notorious websites and trackers and were able to opt-out of several of them.

7. Individual Contributions

Aryan - I worked on the data collection method by figuring out how to modify and use Privacy Badger for our purpose. I also worked on the visualization website, mostly writing the data processing functions and some styling.

Ragna - I worked on the visualization and implemented the backend server for posting and loading the data. I also used the data processing functions that Aryan wrote to create components that rendered and displayed the visualizations. Lastly, I worked on the database for the trackers and their opt-out links.

William -

References

[1]
[2] Ghostery®. "Submit a Tracker." *Ghostery*, 4 Dec. 2020, www.ghostery.com/submit-a-tracker/.

REFERENCES

[1]
[2] Ghostery®. “Submit a Tracker.” *Ghostery*, 4 Dec. 2020, www.ghostery.com/submit-a-tracker/.

[3] https://globalprivacycontrol.org/
[4] https://www.eff.org/issues/do-not-track
[5] https://github.com/EFForg/privacybadger/issues/2716