



PRIVACY PANDA

Final Report

05.12.2021


About Us

Privacy Panda was founded by five graduate students at Carnegie Mellon University with degrees in security engineering, privacy engineering, and business.

Introduction

The Children's Online Privacy Protection Act (COPPA) aims to protect children's privacy online. It enforces regulations on websites that are "directed to children under 13" (Federal Trade Commission, n.d.). These regulations focus on privacy policies, data collection, and parental consent to access the websites. COPPA forms the foundation for what our team chose to prioritize with this project.

One popular application that aims at monitoring children's web browsing is Net Nanny. This is an application that has been in production since 1993. Net Nanny allows a guardian to set time constraints, block applications, web filtering, and more. Unfortunately, it is



expensive depending on the number of devices under its umbrella. Also, no free option is available. Net Nanny also has the ability to track where a child is located. The admin of the app can set specific locations and be notified when the user leaves that area, such as a house or school. This feature has resulted in some privacy concerns.

Over the last three months, our group has worked extensively to develop products to address concerns we had of the application mentioned above. An extremely important feature of our products is that we do not store any user information. Our extension is a better version of browsing history. Regardless of what your child visits, we are not storing any personal data on our servers. This is an important distinction from the application mentioned above. While we give some power to the parent to see what has been viewed, we also consider the privacy of the child.

Motivations and Project Goals

The following project goals were our number one priority in designing our application, and informed most of our decisions.

To provide COPPA compliance data to parents

We built our application using the Children's Online Privacy Protection Act (COPPA) to credentialize and support each decision that we made. We are tracking COPPA clauses in the websites that our customers' children are visiting because these clauses are a good foundation for understanding privacy in children's websites.

To educate parents on privacy issues

Another main focus was to ensure that parents are not just blindly using our application to track or control their children's internet usage. Additionally, we have focused this application to be geared towards children 13 years old and younger (in accordance with COPPA, which is directed towards children of this age range). Our educational tools work to provide parents with the information necessary to control their privacy online to their own standards.

Focus on edge processing

An important part of our design for both the website and extension was to not store any user information. Everything the user views is stored locally, just like browser history, except we have added more functionality to allow the parent to see more detail. This design strategy allows us to be more secure and private as a company.

Motivations of Implementation Decisions

I. Local Storage

A major aim of Privacy Panda was to not store any user information in a server, and to have all information stored locally, with the user. This was especially important if we were to implement child account functionality; we felt that it was best that a product built with the aim of protecting the privacy of children online did not collect the information of these children. We chose Google Chrome due to its widespread use and due to the widespread documentation of extension development for the browser. Using Chrome's inbuilt `chrome.storage` API followed from this decision, as it was the most straightforward way to store user data locally.

II. Authentication

We ideally wanted to do client-side authentication to align with our vision of providing a product that does not store any user information in a Privacy Panda server. We ultimately chose to use OAuth2 to have users authenticate using their Google credentials; this way, we do not store user information, and the dashboard is protected from the view of children. Currently anyone with a Google Account can view the dashboard, which is something we would aim to limit if given more time in this project. However, the current implementation still fulfills the major aim of preventing children from accessing the dashboard, at least until they are old enough to make a Google Account.

III. Dashboard

Only domains are shown in the browsing history section instead of full URLs to preserve some privacy for the child (e.g. we display `www.google.com` instead of `www.google.com/baby-pandas`). Only the first visit to a domain is shown, as it is more important to provide the COPPA information to parents than to clutter the dashboard with a history of every non-unique website visit. The COPPA Clauses shown in the dashboard are the sentence before, sentence containing, and the sentence after the mention of "COPPA" in a privacy policy. A method more complex than this would have been difficult to implement, so this is the method that was chosen. This privacy policy information is not stored or saved locally as it would be too large to fit in local storage. Instead, we use the [Fetch API](#) to make web requests from a pre-filtered json file with privacy policy information that is stored in an AWS S3 Bucket. This allows the information to be quickly pulled for display in the dashboard whenever a user refreshes the dashboard.

Alerts are based on the alert words that appear in the title of a page, and not based on the URL. Basing alerts on the title of a page is useful as it captures search queries as well as titles of web pages. In an alert, we display the alert word and a timestamp, and also display the page title to give context to the alert. For example, an alert is thrown with the word “hash”, but if a child is seeking to learn about hash functions, it is probably not a cause for concern. Finally, parents and guardians are unable to choose alert words due to the potential for over-monitoring as a result.

IV. Incognito Functionality

In order for our implemented functionality in incognito to work, the user would need to enable incognito functionality within the Chrome extensions page first. There are three options for Incognito Functionality with Chrome extensions: Spanning, Split, and Not Allowed. The default is Spanning Mode, because it interfaces well with servers. In Split Mode, the data collected on incognito is stored separately from that collected on a normal tab. Furthermore, the processes cannot communicate with each other so the user could open the dashboard on an incognito tab. Not Allowed Mode prevents the extension from being enabled in incognito at all. We chose to implement split mode because our extension does not rely on a server, we want the user to be able to open a dashboard in incognito, and it is more private to store data separately.


Privacy by Design Strategies

Data-Oriented Strategies

I. Minimize

With Privacy Panda, our aim was to not store any user data within a server. We accomplished this by having user browsing data stored locally (within the browser), which could then be displayed in the extension dashboard. Our implementation of authentication serves less as true authentication and more as a barrier to entry for children, but this was also done without any new user credentials being stored with Privacy Panda. Additionally, instead of scraping the information of all websites as a user browsed, we instead pulled COPPA clause information from a file hosted in an AWS S3 Bucket.

II. Separate



User data is stored locally, so there is no central location for data to be breached from. User data can only be found in the browser where Privacy Panda is installed.

III. Abstract

On the dashboard, instead of displaying full URLs visited, we only return the domains (so instead of reddit.com/r/wallstreetbets, we only would show reddit.com). Dates and times of these visits are shown, but only the first visit to a domain is shown.

IV. Hide

Access to browsing history and alert history is only available through the dashboard. To log into the dashboard, one must log in with a Google Account. This will ensure that younger children will not have access to the dashboard.

Process-Oriented Strategies

I. Inform

Within the extension and on the website, there is a section named “About Privacy Panda”. We include the following information in this section: “Using the Privacy Panda extension, browsing data is stored locally on your computer. This means that it is only accessible from your browser. IP Addresses and some user device information is automatically stored by the web server we use to host our website, but this information is only accessible to the hosting server administrators”.

II. Control

At present, there are no controls regarding the display of browsing information available within the dashboard.

III. Enforce

Privacy Panda’s Privacy Policy: “Privacy Panda is committed to your privacy. We do not store any data on users of our extension package, and as such, there is no data for us to process or store. If our data handling practices change, all users are to be notified through updates to the extension software as well as notices here on our splash page.”

IV. Demonstrate

Privacy Panda does not store user data in its servers, and as such does not log any data. There is no data to be reported on, so we are unable to retrieve data for Subject Access Requests.

Biggest Implementation Challenges

I. Authentication

Authentication is one of the most important parts of Privacy Panda. We spent a lot of time finding a way. And at last, we decided to use OAuth2 from Google to authenticate our users instead of creating a server ourselves to do it. OAuth2 is the industry-standard protocol for authorization which provides a mechanism for users to grant web and desktop applications access to private information without sharing their username, password and other private credentials (OAuth2, 2018). As none of our team had experience in this, it took great effort and time. We first generated a key from Chrome Extension and used that to access Google API and Services. Then we went to Google Cloud Platform to create our project in order to access Google OpenID Connect. After this, we could do the coding with JavaScript and HTML. One of the biggest challenges here was that it is difficult to fit the code into background.js as our extension was half-completed at that time without the Authentication things. Both parts of the code were tested separately and both worked. Solving the code conflict took several days. We communicated in the group and solved the problem together. We got a lot of help from [“An Object Is A”](#).

A new problem here is that we have authorized the users using OAuth2 from Google, but anyone who has a legal google account can access the dashboard of our extension theoretically if he can access the user's computer. At this time, we cannot fix the problem. Our rough idea is that we could relate the browsing history to the current log-in account and other accounts cannot access that.

II. Interfacing with Chrome storage

Chrome storage works through specific APIs on Javascript. In order to display a user's browser history, we had to store information from every site visited. The information was accessed only by “chrome listeners” that acted on specific browser actions. The listeners were accessed with Javascript files, but we had to display them using HTML files for them to appear on the dashboard. As none of our team has worked with Javascript, HTML, or Chrome storage before, interfacing between them was one of our big challenges. This was a challenge that lasted the majority of the semester because there were different processes for different pieces of

information. In the end, we overcame this challenge by using documentation and tutorials when we could, and trial and error when we couldn't.

III. Configuration file for website server

Hosting for the website was necessary for the completion of our project. We used an AWS EC2 server to host our website. On this server, we still needed NGINX to run our web application. Upon pulling our source code to the server, we needed to configure NGINX to read from the files, run our website through different ports and names, as well as reading the static files for styling and javascript. Although pretty straightforward in text, this took some time to get up and running. Using various blogs and YouTube videos, we were finally able to have our website up and running with no issues.

IV. Honorable mention: Regex

Parsing the Privacy Policy file was done with a Python script. The logic used to filter the script was not difficult to write, but understanding how to write Regular Expressions that would return the correct format for the COPPA Clauses took some time.


Privacy Impact Assessment

I. Description of Stakeholders

During the development process, our team built a user persona. Gus Andrews, in his post "User Personas for Privacy and Security" outlines what personas are used for, and how they can be helpful in the development phase of any application (Andrews, 2017). Our persona is named Susan, and she is a stay-at-home mom of three who wants to ensure that her kids are being safe on the internet. She does not know much about privacy, but is willing to learn. She doesn't always know where to look for the resources. See Appendix A for the details of her persona.

This is just one example of our diverse potential user group, to include any parent that wants to track privacy policies of the websites that their children visit. This application is flexible enough to work with parents of one child or many.

Our team also felt it was important to take into account the needs of the children. While our application is solely geared towards keeping children safe online, we understand that they might not always view it as helpful. Children need a certain level of independence to properly grow and make mistakes. We work to give some independence back to the child by way of educating parents on how to talk to



children about privacy at different ages and stages of their life. Our goal is to be flexible enough to reach any parent-child dynamic, understanding that all relationships and individuals are different.

II. Stakeholder Consultation Plan

Our number one priority is ensuring that our customers are always informed of our decisions and any updates that may be pertinent to their use of the application. For that reason we are committed to keeping an updated privacy policy, which our stakeholders can find located on our splash page at all times under the tab “Privacy Policy”.

Finally, we have a robust set of resources on our splash page which parents can use to educate themselves about privacy policies in general. We hope that this will provide some context for parents to make their own educated decisions about the privacy policies of websites that their children frequent.

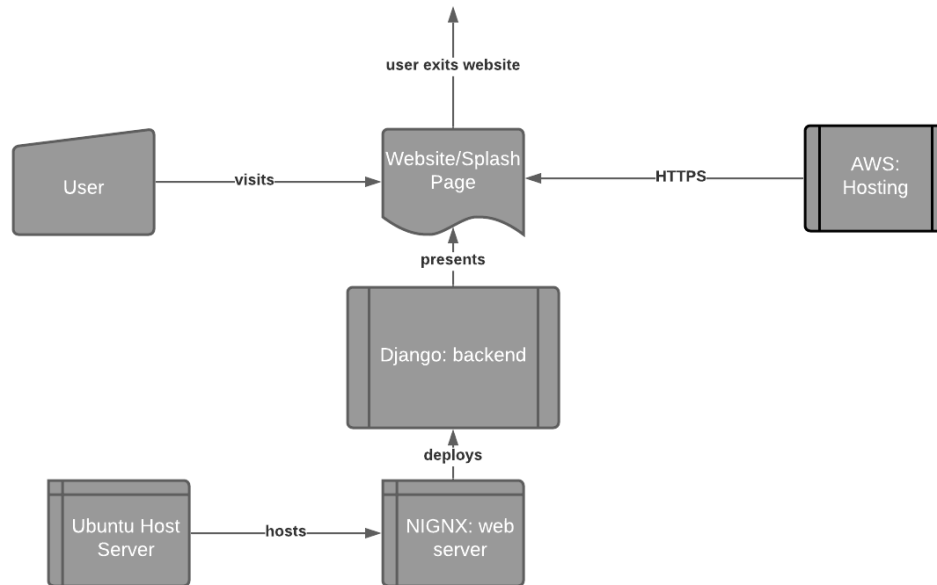
III. Information Flows with Diagrams

The following diagrams outline information flows within both our splash page website and extension application. As discussed in Motivations and Project Goals, we focused on processing data in a way that ensured complete privacy for our users. Because of this priority, our information flow diagrams do not contain any extraneous information.

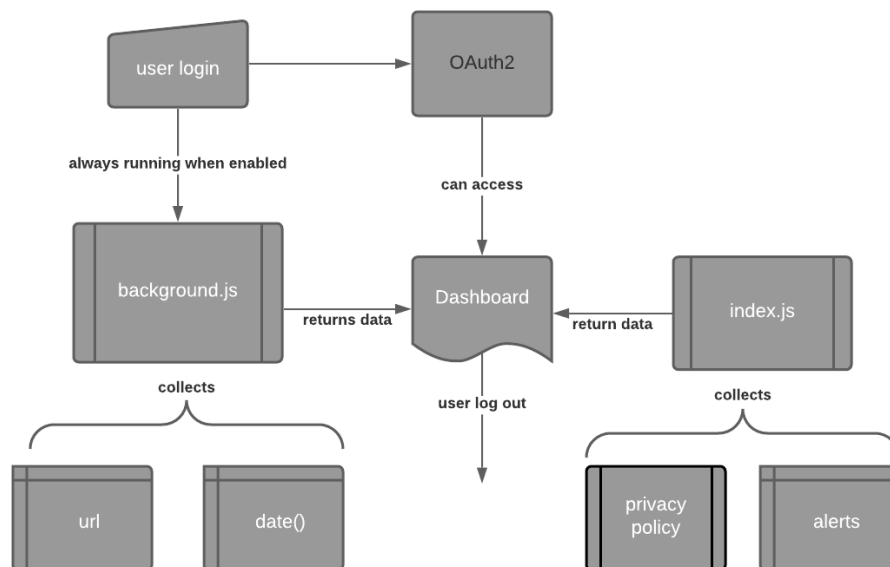
As discussed, we used Google’s OAuth2 for authentication. Using OAuth allows to avoid handling authentication credentials.

Additionally, all data collected about browsing history is stored locally on the user’s computer. This way, our users can have faith that their browsing history is private from external attackers and need not worry about our handling of their data.

Unified Modeling Language for Splash Page




Unified Modeling Language for Extension



IV. Privacy Risks and Solutions

We identified various privacy risks throughout the semester and addressed them during presentations. There are two main categories of risk: a parent violating the privacy of their child, and the extension being used to track someone who is not a



child and does not consent. Even though this extension is intended to give a parent more control over their child's privacy online, we did not want to create a tool that can be used to micromanage a child. There could be sites that are not problematic, but that the child does not want the parent to know about. Additionally, there could be topics that a child wants more information on, but does not want to ask their parents. So long as these topics are not harmful to the child, we did not want to violate their privacy. The second category could manifest in cases of intimate partner violence where an abuser uses our extension to monitor a spouse's online activity, for example.

The biggest decision we made was to not deploy this extension for public use. Though we believe it is very helpful when used correctly, the possibility of the risks mentioned above outweigh the benefits of deploying it. We do not currently have the experience to adequately protect our users from these risks, but we did make various implementation decisions to address them. We addressed these risks first by limiting the information that we provided to the user. We focused on COPPA mentions that appeared in a website's privacy policy because that is the main purpose of the extension. Additionally, we added granularity to the website url when we displayed it on the dashboard so that the specifics of a child's activity can still be private. For example, if a child was reading through a subreddit page, our browsing history would only display reddit.com, not the subreddit. Lastly, we created a whole website for educational purposes to help the parents without harming the children.

V. Trade-offs Between Usability and Privacy

The decision to create an extension is itself a tradeoff between usability and privacy. An extension can give personalized, real time information to parents about how to protect their children's privacy. However, it can be difficult to set up and use. In order to address this concern we made a website that has onboarding instructions, but there are still various steps. In this case, we leaned towards privacy and made the extension anyway, but in other cases we leaned towards usability. Originally we had 7 columns of information on a user's dashboard. They contained details about the website their children visited and details about how to access more information. We reduced this to 4 columns containing the most crucial pieces of information. This usability decision was taken to not overwhelm the user with too much information.

References

- Andrews, G. (2017, June 29). User Personas for Privacy and Security. Medium.
<https://medium.com/@gusandrews/user-personas-for-privacy-and-security-a8b35ae5a63b>
- Federal Trade Commission. (n.d.). Children's Online Privacy Protection Rule ("COPPA").
<https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/Childrens-online-privacy-protection-rule>
- Fetch API—Web APIs | MDN. (2021, April 17).
https://developer.mozilla.org/en-US/docs/Web/API/Fetch_API
- Manifest—Incognito. (2015, September 25). Chrome Developers.
<https://developer.chrome.com/docs/extensions/mv3/manifest/incognito/>
- OAuth2: Authenticate users with Google - Chrome Developers. Chrome Developers. (2018). Retrieved 11 May 2021, from
https://developer.chrome.com/docs/extensions/mv3/tut_oauth/

Appendix A: Susan's Persona

Overview:

- Soccer mom
- Anxious over her kids' online safety
- PTA president
- Stay-at-home mom
- 3 children
 - 4th grader
 - 6th grader
 - 9th grader

Technology Expertise Level:

- Watched The Social Dilemma & wants to protect her kids online privacy, but doesn't know how
- LOVES Facebook
- Limits her kids screen time
- Uses default browsers and applications that come with her phone/computer

Technology Use:

- Likes to read blogs about cupcake recipes
- Kids have a family iPad
- Has a cell and a home computer
- Active on the PTA Facebook group

Access Locations:

- Home
- Café down the street

Threats from Technology Use:

- Accidentally hit "reply all" on an email with sensitive information

- Worried that her kids are distracted with other websites during zoom school
- Heard from her social group that zoom is collecting personal data but doesn't know how to stop it
- Worried that her kids social medias and games are selling information to other sites that can damage her kids reputations for college admissions

Physical Threats:

- Cyber bullying
- Information leaked through unregulated websites (like home address or phone numbers)
- Worried that her kids might obsess over social media standing. That could lead to anxiety
- Worried that her kids are dangerously influenced by models online and conventional body norms

Needs:

- Feel confident that her kids are safe online
- An easy way to monitor which websites her kids are accessing
- Feel part of her kids' lives (school, sports)
- Be social