Writing 2 - Technical Summary
Data Safety Labels
By: Victoria Hennemann

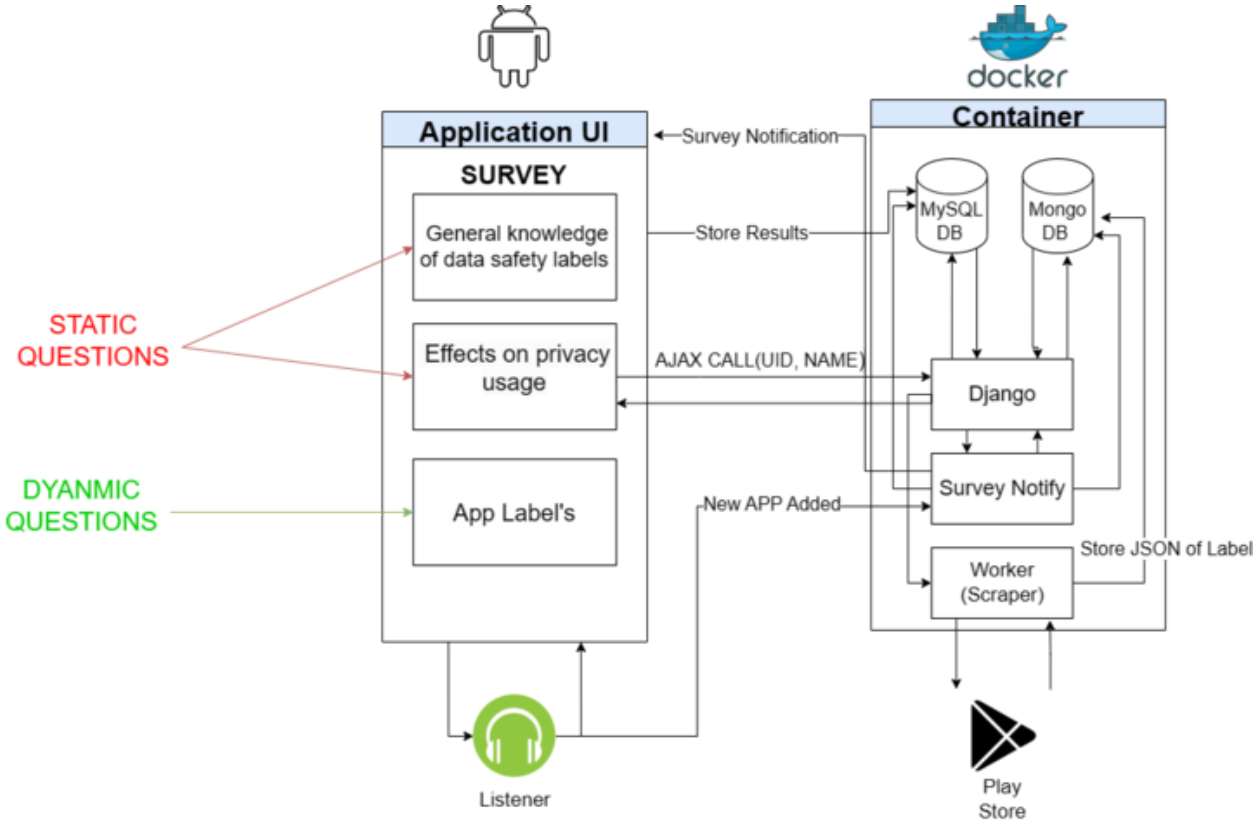
Are people considering data safety labels prior to installing an app? Our innovation aims to tackle this question by providing an Android application that will serve as a research tool to assess user's knowledge about data safety labels. While having a simple frontend component, our backend is what makes our application technically novel. A prior way to examine user's perceptions of data safety labels was to generate a static survey which user's would complete. Our app allows for the dynamic generation of surveys, therefore we assess users when a user downloads an application. This allows researchers to assess users on the most updated labels in the easiest and most convenient way possible.

Our first objective is to develop a user-friendly interface that assesses the user's knowledge of a data safety label in the form of a survey. Second, our backend should allow for the dynamic generation of surveys with the least number of updates to the frontend. Our final product application should consider: *latency* (when should the user be prompted to answer a survey) and *sufficiency* (number of questions asked to users). Within the backend we aim to have some form of logic that will limit the number of questions asked based on factors of the label that are deemed more important.

Our frontend is built in Android studio which allows us to see front end interactions on a google device. We are also modifying a python web scraper to fit our specific goals. Our backend will consist of these main components: NGINX web server, a MYSQL DB, a Mongo database, Django, an updater (listener), and a worker (web scraper). These will all be hosted in Docker containers. Making the backend work in the most efficient way possible with the least amount of code will be technically challenging, however we are planning testing periods between each prototype to make the application more efficient. A diagram of the backend is shown in Figure 1 below.

This research is funded in its entirety by Google. Therefore, any materials needed such as test phones, database credits, or other escalated versions of open source software will be funded by the budget given to the Usable Security and Privacy Lab. We estimate that this project will have a really heavy backend which will include data querying, building a web scraper, and forming points of communication between databases, servers, and the user interface. Therefore, we estimate around 1,000-3,000 lines of code. We expect to put our research application through multiple iterations of testing prior to starting our first study. Our first testing period will begin once our alpha prototype is complete. This prototype should scrap the web store when a user downloads the application, generate questions using Survey JS, send the questions to the user, and obtain user input. This prototype will be demo-ed to a Google Researcher for feedback. Our second milestone is to create the Beta prototype that will have all the capabilities that we outlined above. This Beta prototype will go through formal pilot testing to receive feedback before our final product is used for testing.

Figure 1.


docker
Application UI
SURVEY
General knowledge of data safety labels
Effects on privacy usage
App Label's
Survey Notification
Store Results
AJAX CALL(UID, NAME)
New APP Added
Container
MySQL DB
Mongo DB
Django
Survey Notify
Worker (Scraper)
Store JSON of Label
STATIC QUESTIONS
DYANMIC QUESTIONS
Listener
Play Store