Evaluating User Understanding and Effectiveness of Data Safety Labels *In-Situ*

Assoc. Prof. Adam J. Aviv, The George Washington University, CV
Dr. Sai Teja Peddinti, Google

Problem Statement: Privacy nutrition labels offer a compact and more readable description of the functionality of apps and services and have been recently adopted for describing mobile applications behaviors by both Apple (as *privacy labels*) and Google's Android (as *data safety labels*). While there is much promise in clearly synthesizing app functionality into labels, it remains unclear how users both understand these labels and how they will effectively use labels in managing the privacy hygiene of their mobile devices and the apps installed therein. In this proposal, we outline three experiments using an *in-situ* methodology of participants reviewing data safety labels of their own apps on their own phone to determine: (a) how users understand data safety labels and what functionality they expect an app to have based on the labels; (b) do data safety labels influence the choice of app for installation; and, (c) how do users review labels for previously installed apps, including cases when the app updates its labels after installation. The expected outcomes of this research will inform how and where data safety labels should be presented to users to help them make the most informed privacy decisions that can be integrated into the Play Store. This research will also support the design of new tools that can be integrated into the Android environment, where prompts and nudging could be used to assist users, similar to how runtime permissions are currently used for reporting on unused permissions by apps.

Researchers' Background: Associate Professor Adam J. Aviv has a strong record of investigating the user privacy perceptions and actions of Google products. This includes exploring usage of Google's third-party apps (Balash 2022a) and Google's My Activity dashboard (Farke 2021). He also has active research on Apple's privacy labels, including a longitudinal analysis of privacy label adoption in the Apple App Store (Balash 2022b) that is ongoing with over a year of data currently collected (the paper is in submission). Prof. Aviv is a prior recipient of the Google Faculty Research Award in 2019 to study in-app authentication mechanisms on Android apps. This has partially supported a number of publications on authentication for mobile devices (Munyendo 2021, Munyendo 2022) and apps (Bailey 2021), and is the primary funding source for ongoing work in large scale measurement of authentication methods of Android apps as collected from APK Mirror sites. Prof. Aviv will work closely with Dr. Peddinti (Sr. Research Scientist in the Privacy Research group at Google) on this new research direction into data safety labels, and the two have been actively meeting on the topic and are excited to pursue this research direction in the coming years.

Approach: We propose three primary and independent studies, each of which will use an *in-situ* methodology, whereby participants will respond to questions and prompts on their own device in relation to their own previously installed or soon to be installed apps. To make such studies possible, we will develop a custom instrument app, to be installed on the participants device, that has permission to monitor new app installations and review currently installed apps. The app will communicate with a server, managed by the researchers, to distribute study material based on actions of the user with respect to app installation or an update to previously installed apps' data safety labels. Critically, the development of the survey app will enable in-context and on-phone inquiry into the users' understanding and behavior regarding data safety labels. Such an app can be developed from scratch, or built on prior work of Paco[1],

[1] https://www.pacoapp.com/

which served a similar purpose in prior research[2], but may be currently out of date with latest Android releases and also the needs of this study.

Study 1: User understanding of data safety labels: A key research question we seek to answer in this study is, simply, how do users understand the data safety labels with respect to the functionality of apps they currently use? We propose to use the custom survey app to review the installed apps on a participant's device, review them for data safety labels and app genera and category, and then present the participant with a set of closed-item Likert questions and open response items with respect to the apps functionality and associated labels. This may include, for example, displaying a weather app, like the WeatherUnderground, with its data safety labels. The participant will then answer Likert questions for concern and benefits for each label, and then provide an open-text response describing the functionality of the app, particularly primed for data privacy features. By varying which apps are shown, with respect to the coverage of the data safety labels and app genre, we can assess how users' match data safety labels to app functionality in a meaningful way.

Study 2: Post-installation review of data safety labels: The custom survey app can detect newly installed apps by listening for the ACTION_PACKAGE_ADDED intent broadcast by the Android system's package installer. When this occurs during the study period, participants can be prompted about how they selected that given app for installation. Anecdotally, most participants will likely *not* review data safety labels and will have other motivations for installing the app, such as the app is required for a purpose without alternatives, or choose the app based on the rating or the reviews. But this provides an opportunity to prompt the participant about the data safety labels to determine if better highlighting the labels would lead to a different choice of app with fewer concerning labels (which could be determined by looking at related apps in the Play Store). Additionally, survey prompts could inform us to which set of data safety labels are most important to display and if this varies across participants (due to personal privacy sensitivities); the result could lead to updates in the Play Store interface.

Study 3: Data safety labels updates to apps: Data safety labels are currently an install time mechanism, similar to that of install-time permissions. The opportunity for reviewing and decision-making is ephemeral, and once an app is installed, it is difficult and very unlikely that a user will return to view the labels, for example, to choose to uninstall the app and replace it with a less permissive one. At the same time, even if the labels were reviewed at install time, they can change when a developer updates the app, and that decision point is no longer valid. In this study, we seek to determine how to help users in reviewing the data safety labels of previously installed apps. The study will consist of two main tasks: first, participants will be shown a general overview of their apps installed and the labels for those apps, and second, they will be prompted to answer survey questions whenever an installed app changes the labels, reporting on the differences. In both cases, we also query about the participants desire to uninstall or select a different app as a result of the labels, or if they feel that is impossible or unnecessary.

Study Recruitment: We envision running each of the studies independently of each other and recruiting using survey platforms like Prolific[3]. As we are not performing interventional, experimental research, rather exploratory studies that seek to expose possible theories for how users interact with data safety labels, we hope to keep our recruitment size relatively small and survey timeline relatively short. For

[2] https://research.google/pubs/pub46261/

[3] https://www.prolific.co/

Study 1, we hope to recruit at least n=200 participants, which should provide board coverage to the kinds of apps installed and data safety labels. It will also be more than sufficient for qualitative thematic analysis of the open text responses. Study 2 and 3 are more longitudinal in nature and thus will require more coordination. For each of these studies, we seek to recruit two independent samples of n=80 participants, with the expectation that 50% may drop out before the end of the study period (~3-months). This sample should still be able to provide meaningful insights into both how users consider data safety labels during installation and post-installation. We've included a $6,000 budget for human subject payments and $2,400 (40%) for recruitment costs.

Human Subject Research Ethics: The proposed research involves human subject research. Dr. Aviv has significant experience in conducting human subject research ethically. All studies will be approved by GW's institutional review board (IRB), use best practices in data collection, and meet community ethics standards for HCI research.

Expected Outcomes and Connection to Google's Product and Services: From this research we anticipate two key outcomes, both of which can be applied to improving products and services at Google:

1. *Presentation and Discussion:* As an outcome of Study 1, we will provide deeper understanding of the users' expectations of data safety labels as they connect to expected app functionality. As Google may seek to perform automated checking of data safety labels, such research could directly inform how to present discrepancies to users so that they can make informed decisions about their privacy.
2. *Tools and Methods*: The second outcome of this research comes from Study 2 and 3, which focus on how users decide to install an app or review previously installed apps based on the data safety labels. The outcome of this research, particularly the app that reviews new or previously installed apps, could be transformed directly into a new tool (or product) that helps users better review the data safety labels associated with the apps they install, including offering alternatives apps and mitigating strategies.

Proposed Budget: The following is a budget for 100% of a GRA student for 1 year of funding, including tuition, fringe benefits (e.g., healthcare), and human subject research payments and recruitment costs (e.g., overhead for using MTurk or Prolific). **Total Budget: $83,892**

- Student Salary: $36,667 (12-mon.)
- Fringe: $2,915 (7.95%)
- Tuition: $35,910 (18 cred.)
- Human Subj. Research: $6,000
- Human Subj. Recruit.: $2,400 (40%)
- **Total: $83,892**

References:

- (Balash 2022a) David G. Balash et. al. Security and Privacy Perceptions of Third-Party Application Access for Google Accounts. *31st USENIX Security Symposium* (USENIX Security 22). 2022.
- (Balash 2022b) David G. Balash et. al. Longitudinal Analysis of Privacy Labels in the Apple App Store. Arxiv. 2022.
- (Munyendo 2022) Collins W. Munyendo et. al. "The Same PIN, Just Longer": On the (In)Security of Upgrading PINs from 4 to 6 Digits. *31st USENIX Security Symposium (USENIX Security 22)*. Aug. 2022
- (Farke 2021) Florian Farke et. al. Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity. *30th USENIX Security Symposium* (USENIX Security 21). 2021.
- (Baily 2021) Daniel V. Bailey, et. al. ``I have no idea what they're trying to accomplish:'' Enthusiastic and Casual Signal Users' Understanding of Signal PINs. *17th Symposium on Usable Security and Privacy*. (SOUPS'21). 2021
- (Munyendo 2021) Collins W. Munyendo et. al Using a Blocklist to Improve the Security of User Selection of Android Patterns. *17th Symposium on Usable Security and Privacy*. (SOUPS'21). 2021