What Phone Applications Want From You

Jonathan Bryan[1] and Josh Herman[2]

I. Objective

Technology over the years has paved the way for many ground-breaking and innovate discoveries that have allowed for incredible progress in society. The integration of technology into society has become an ever-growing and necessary aspect to the lives of many individuals and has resulted in a large amount of users becoming reliant on this powerful entity to make their lives marginally more simple. One primary conduit of technology that has become a focal point in the lives of many, is the presence smartphones. With the ever expansive growth and influence presented by this monolithic addition to society, there has been an increase in concern regarding the perversion of users privacy, and the desire by large companies to obtain valuable data from these users. Through smartphones, it is possible for users to gain access to a plethora of apps to make their daily lives more manageable or to simply entertain themselves, with some examples of apps being *Trello*, *Facebook*, and many others. With the large expanse of options users have at their disposal, it is important for users and companies hosting these apps to understand how companies may be trying to manipulate users and gather personal data about them.

For this study, our aim is to produce a quantitative analysis over a large amount of apps found on the Google Play Store to observe the varying permissions requested by apps. Alongside this, we also plan on applying certain filters to these results such as popularity, genre, and so on, to determine if there is a significant difference in, say, the difference between what permissions are requested from an app that is used by million of users versus an app used by a few thousand. By doing so, we hope to be able to distinguish apps that may be requesting unnecessary permissions for malicious intentions. Similarly, we also hope that by performing this study we can incite the need to improve user awareness to prevent users from forgoing their data to companies who would look to misuse of sell their information.

II. Motivation

Many apps on app stores today require permissions on the user's phone to function properly. These can include location services, microphone/camera, SMS, storage, and even body sensors. While some of these apps legitimately need and use these permissions, there are some apps that do not need these permissions. For instance, it is clear why a messaging app would need access to the phone's SMS system, or a voice communication application would need access to the microphone. However, a simple game application does not need access to the phone (phone number, making calls, etc). While this may seem obvious at first glance, many companies have successfully marketed their applications in such a way that users easily become addicted. This type of addiction can encourage users to accept any and all permission requests simply to continue using the app, with no thought into what exactly they are agreeing to. As such, there is an obvious need to improve user awareness and to observe on a large scale the type of information and access that these businesses want from users.

III. Data

Our data is going to be collected from the Google Play Store and the Amazon App store. We aim to create a bot using *Python* which will go through a large amount of varying apps found throughout the app store and extract valuable information such as the permissions, popularity, genre, and so on.

A. Comparative Analysis - App Genres

By observing the varying permissions found throughout each genre, we hope to determine if there is a specific genre that seems to contain more malicious results than others. For instance, what kind of permissions do games generally require versus productivity apps.

B. Comparative Analysis - App Stores

This project also aims to compare hows apps' permissions differ between app stores in general. Specifically, we would look at the Google Play Store and the Amazon App Store. It will be interesting to see if there's a change in permissions required for apps on the different app stores, especially since one of the app stores is much more popular than the other.

C. Large vs Small Applications

With the possibility of app designers selling data to other companies to, for example, improve targeted advertisements, the topic of which audience sizes are targeted more by companies becomes all the more interesting. For instance, we could compare a game created by a large company that has a large fan base to a game that is developed by a small company and has fewer fans.

IV. Responsibilities

The tasks for this project are going to be split into three sections. First data gathering, second visualization, and third analysis.

A. Data Gathering

This task will be led by Josh Herman. This task will include gathering information from the Google Play Store and the Amazon App Store. This will be accomplished by creating a bot to get information from both stores by means of web scraping.

B. Data Visualization

This task will be led by Jonathan Bryan. It will consist of filtering the data gathered from the bot to visually display the import aspects associated with our research.

C. Data Analysis

This task will be led by both students. In this section, we will analyze the data gathered from the bot, alongside the information we can infer from the data visualization to justify our conclusions.

V. Time-line

A. Bot Creation

To properly analyze multiple app stores on a large scale to obtain a grandiose amount of data, we will require the assistance of a bot. This will be done using Python and will be the primary tool used for data scraping. While we will be tinkering with the bot throughout this project, we hope to have the first version of this bot done by **October 20.**

B. Data Collection

For this task, We will need to collect data from the Google Play Store and from the Amazon Play Store. This task would include collecting the data, and sorting it into their respective collections. We aim to have this section finished by **October 30.**

C. Data Visualization

Data visualization is the task of incorporating visual representations of the data we have collected through graphs, making it easier to present the data. We aim to have this finished by **November 7.**

D. Analysis

Analysis is a task of combing through the data, and coming to the conclusions that are reasonable considering the data we have collected. We expect this process to take a week, and to be finished by **November 14.**

E. Final

This task includes compiling everything into a final report, and presenting it to the class. We expect this to be an easier section since most of the work will already be finished. We aim to have this task finished by **November 21**

VI. Expected Outcome

A. Large vs Small Applications

The hypothesis for this section is that applications made by larger companies will require more permissions than apps made by smaller companies. We believe that larger companies are going to gather more data from users than they really need.

B. Apps on Different Stores

The expected result for this analysis is that apps available across multiple platforms will retain the same permissions across each platform.

C. General Outcome

The final expected outcome is a more general one. We suspect that many apps on the various app stores require more permissions then they really require. We also believe that while an app legitimately requires a permission, it can lead to a security vulnerability.