Week 3: Empirical Study of Google Playstore

Nate Bachmeier

TIM-7140:Software Engineering

March 14, 2021

Northcentral University

# Empirical Study of Android Malware

## Problem Statement

Few inventions are more personal than the mobile phone. Some studies estimate that the average consumer touches their phone over 2600 times per day across 73.8 hours per month (Reizer, 2019; Winnick & Zolna, 2016). Those users interact with dozens of specialized apps that broker their lives with unfettered access to their privacy. Anyone with a valid email address and twenty-five-dollars can publish content into Google’s Playstore (Microsoft, 2018). This

low entry barrier for both developers and consumers fuels its growth to 86.2% of the market (MobileApps, 2021). However, it also raises questions regarding Android apps’ quality and accessibility.

## Purpose Statement

Krutz et al. (2015) collected and analyzed metadata across 4416 versions of 1179 F-Droid projects. F-Droid is an installable catalog of FOSS (Free and Open Source) apps for the Android platform (F-Droid, 2021). Numerous researchers are leveraging this information to publish papers on trends across the mobile community. While these efforts shine light into the problem space, they are not complete. Even thoroughly analyzed, F-Droid represents 0.1% of the size of Google Play Store’s three million apps (Statista, 2021).

Instead, a modern version of this study needs to occur, which leverages APK Mirror (APKM). APKM exposes a RESTful interface for downloading versioned content from Google Playstore. It also provides a searchable interface cross-referencing multiple publications from the same development team. After collecting a representative sample, automated tooling will extract metadata about each program’s instance. A feature assessment will determine normal behavior for predicting the quality and accessibility of Android mobile applications.

# Data Collection Process

APKM curates a feed of apps with a proven track record of being relevant and unique. After archiving new app versions, they raise a notification through Push Bullet. This approach enables researchers to collect information in real-time. Additionally, their web portal allows throttled queries to prevent massive downloading of their collection.

## Desired Features

|  |  |
| --- | --- |
| Property | Description |
| App Metadata | The name, category, and description |
| System Requirements | Android OS Version and hardware requirements |
| Author | Publisher information |
| Revision | Revision description, publication date, and version |
| Rating | Review count and number of stars |
| Manifest | Permissions, services, and intents |
| Quality Metrics | Results of SonarQube |

## Backfilling Strategy

## Forward Monitoring