

Title: Open Android - An Android Feature Database for Data Scientists

Authors: Robert Schmicker

Major: Computer Science

Minor: Mathematics

Department: Tagliatela College of Engineering

Advisors: Frank Breitinger, Ibrahim Baggili

Abstract:

With a market share of around 80%, it is superfluous to mention that the Android operating system is and will be a popular target for malware developers. In order to build reliable tools, researchers need samples from which they usually extract features (e.g., strings) which then are analyzed. Given that process, there are three major problems: (1) it is hard to get malware samples, (2) data scientists are not necessarily familiar with feature extraction and (3) feature extraction can be resource intensive (e.g., time consuming). Therefore, we developed *Open Android* which provides a public data set of Android application features for data scientists to use in testing their algorithms. It allows to gather necessary information without spending the time to program feature extraction tools. Data scientists also receive access to the malware samples' information, without having to spread the malicious code itself. Lastly, there is now a central data set for data scientists to effectively compare algorithms against.