

OBIDROID

“These aren’t the Droids you’re looking for”

Team Members & Roles

Team Members	Team Roles	Contact
Luis Aguilar	Coding & Development	luis@berkeley.edu
Morgan Wallace	Coding & Development	morgan.m.wallace@gmail.com
Shreyas	Coding & Development	shreyas@ischool.berkeley.edu

Background

Google’s Android platform is currently one of the most popular smartphone platforms in the world, with over 81% market share ^[1] and over 10 billion app downloads. Due in part to platform’s popularity, Android malware is a pervasive and increasingly troublesome problem.

At its most benign, Android malware consists of adware. More malicious Android malware can potentially steal user information, transmit GPS data, insert banking Trojans, send premium texts, or transform the phone into a bot to be used in a denial of service attack or spam campaign. Also, there could be times where popular, fair apps might be indulging in practices of unfairness ^[2] & deception ^[3], which most users may not even be aware of.

The FTC tries to protect the users against such apps by enforcing their policies of unfairness and deception. But this process is still passive. Because of the scale of the app store, the FTC relies on user complaints to examine an application, and hence it is a passive activity not an active pursuit of **unfair** applications.

Project Details

We aim to examine an app based on its description and user reviews to sift through a large collection of applications and flag applications that are indulging in unfair practices. We aim to flag the applications based on user grievances.

Some of the indicators could be:

- App description
- User reviews

We aim to examine the above largely through sentiment analysis, aided by information extraction about features that are being talked about in the conversation and cross-checking it with description of the app, if that is an indicator of an unfair practice. For eg, a *torch* app for android may be a popular app, and the sentiment around it could largely positive, but if a few negative comments raise concern about its contact list hijack practice, it should be flagged. But we do wish to account for the fact that *all negative* comments do not amount to a grievance.

Project Goals

Our main aim is develop a **filter** system for flagging *unfair* apps, via customer reviews and app descriptions. We do not aim to *predict* if an app is unfair or not using the reviews/description. Instead we aim to help scale down the problem of policing every app periodically on the app store by building a good indicator/flagging system.

Assessment of Project Goals

Acceptable Ground Rules

- From our point of view, we would like to cast a wider net overall so as to flag most of the potentially unfair apps. Hence, **False Positives** for good apps that may be flagged for closer view is an acceptable scenario.
- Also, if an unfair app doesn’t have sufficient user reviews, it would be hard to ascertain their unfairness.

Assessment Scenarios

- a known malware with sufficient user reviews indicating user grievances should not be missed.
- the tool should be *sustainable* as the user reviews are updated often, and it should be able to apply the assessment metrics periodically
- The assessment of our tool should be examined in overall general category of apps as well as in individual app categories like *games* etc

Deliverables

- Since there is no public API for accessing data about Android App Store apps, we have to build a
 - *crawler* : to be able to crawl and compile a list of apps in overall and general categories periodically from the app store
 - *scraper* : to be able to take a list of Android apps and scrape their significant features (app description & user reviews)
 - *analyzer* : analyze each app on a NLP based metrics to assess if assess grievances of users.
 - *metric/rubric* : a set of metrics that are being employed to assess the unfairness of apps.

Work Breakdown

Task	Description	Duration (week(s))
Project Scaffolding	Crawler & Scraper Scaffolding, DataBase Design	1 week
Labeling & Feature Extraction	Create Training Set with manually attached labels, extract features	1 week
Develop Feature Models	Develop models for supervised classification	2 week
Finalization	Assessment & finalization	1 week

Declaration

This project is being jointly pursued by Luis Aguilar, Kristine A Yoshihara & Shreyas in Info 219 Privacy, Security & Cryptography under the guidance of Prof Doug Tygar.

1. Android takes record smartphone share at expense of iPhone and BlackBerry ↩
2. **FTC Unfairness:** What is Deception? Representation, omission or practice that is likely to mislead the consumer; **Perspective of ‘Reasonableness’** - if the practice affects a certain group, then evaluate reasonableness from the perspective of that group; **‘Material’** - does it affect customer decision with regard to a product or service ↩
3. **FTC Deception** practices that cause or are likely to cause substantial injury to consumers that are not outweighed by the benefits to consumers or competition and are not reasonably avoided by consumers. ↩