CprE 491 - sdmay19-38

Android App Forensic Evidence Database

Weekly Report #2

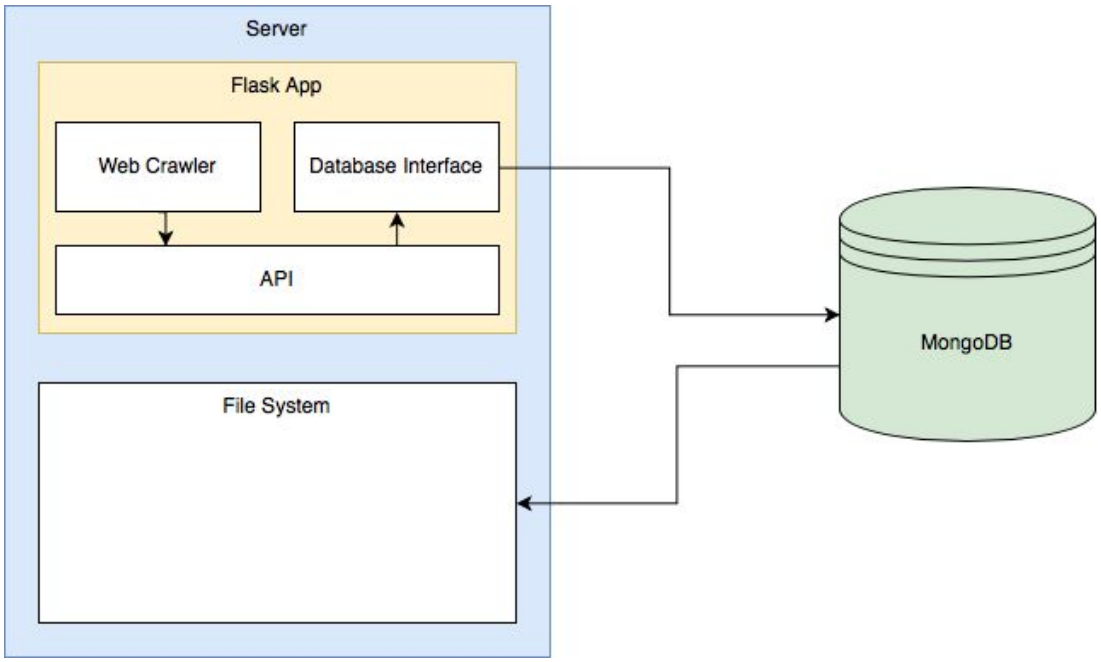
09/13/2018 - 09/20/2018
Client: Neil Gong and Yong Guan
Faculty Advisors: Neil Gong and Yong Guan

Team Members:

Mitchell Kerr -- Tech Lead
Connor Kocolowski -- Report Lead
Emmett Kozlowski -- Meeting Scribe
Jacob Stair -- Testing Lead
Matthew Lawlor -- Meeting Facilitator

Past Week Accomplishments

- Acquired a VM
    - We received a VM from the ecpe department to use as limited storage for containing the information and apks we collect from applications
- Apkpure crawler proof of concept
    - We were able to implement a crawler that processes 20 apps per category from the ApkPure store
        - We limited the number of apps to show that the crawler functions correctly and allows us to prototype quicker
- BeautifulSoup - Apkpure
    - Using the python library, BeautifulSoup, we are able to crawl the categories in the app section. There are 49 categories total and each category holds approximately 400 apps each. We are able to format the name of the app and its meta-data in a text file, but wish to soon add it to a database.
- System Design
    - We can up with an initial architecture for how the system will be built. The design for the system is displayed below.



We are going to have a flask app running the web crawlers and the API to interact with the database. We have chosen to go with a NoSQL database as we are going to be dealing with massive amounts of data and need to scale while maintaining performance. The database will also store the location of the apk file in the file system. The apk files that we download from the app stores will be stored in the file system.

Pending Issues

- Scrapy vs BeautifulSoup
    - We are narrowing down the specifics of this issue, but are still debating which library would be easier for us to use.
- System Design
    - Storing apk files
        - Right now we are going to store the apk files in the filesystem on the server but this might not be a good long term solution.
- ApkPure html inconsistency
    - ApkPure has several different formats for the presentation of app information
        - This usually depends on if the application has multiple downloadable versions

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Mitchell Kerr	Worked on Scrapy implementation of APKPure web crawler	6	9
Connor Kocolowski	Worked on Beautiful Soup implementation of ApkPure web crawler	8	11
Emmett Kozlowski	Researched System Design	6	10
Jacob Stair	Worked on Scrapy library implementation of ApkPure web crawler	7	10
Matthew Lawlor	Continued to work on BeautifulSoup implementation of Apkpure crawler. Acquired a server for us to use.	9	12

Plans for Coming Week

- Decide on Scrapy vs BeautifulSoup
    - We will compare the teams' implementation of each for ease of use and speed of processing.
    - Once we select one of the libraries, we will begin to flesh out the entirety of the ApkPure crawler
- Review System Design
    - We will review the design for the system and make changes according to feedback received. Once the feedback has been incorporated into the design we will start setting up the bare bones architecture.
- Finish crawler for Apkpure
    - We would like to have a finished crawler to build upon for additional app store crawlers.
    - Sort out HTML inconsistencies on APKPure, decide where best to drop in crawler for maximum reusability.
    - Decide how to store APK files for different versions and files which have the same hash across multiple app stores
        - Apps that may be the same APK file on multiple stores may also have different reviews on each store