CprE 491 - sdmay19-38

Android App Forensic Evidence Database

Weekly Report 4

10/04/2018 - 10/11/2018
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

- Database Model
    - Our initial schema is shown below. It will consist of three main json documents. One for the app, the app store and the comments/reviews. The app document will point to all the app stores that we have collected that application from. The app store document will contain all metadata gathered and the information for all the versions on the store. The comments document will contain all comments and reviews that are on that app store for that application.

      Application Document

      ```
      {
          "_id": ObjectId("someString"),
          "package_name": "string",
          "app_name": "string",
          "store_ids": [
              ObjectId("someString"),
              ObjectId("someString")
          ]
      }
      ```

App Store Document

```
{
        "_id": ObjectId("someString"),
        "store_name": "string",
        "package_name": "string",
        "app_name": "string",
        rest of metadata collected
        "comments": ObjectID("someString")
        "versions": [{
                        "version_number": "number",
                        "sha": "string",
                        "signature": "string",
                        "date_published": ISODate(“date”),
                        "path_to_apk": "path"
                },
                {
                        "version_number": "number",
                        "sha": "string",
                        "signature": "string",
                        "date_published": "date",
                        "path_to_apk" : "path"
                }
        ]
}
```

Comment document

```
{
        "_id": ObjectId("someString"),
        "versions": [
         {
                "user_name" : "name",
                "date" : ISODate(“date”),
                "text" : "text"
         },
         {
                "user_name" : "name",
                "date" : ISODate(“date”),
                "text" : "text"
         }
```

```
        ]
    }
```

Pending Issues

- Crawlers are very individualized and need to be modified into more general structure to allow for crawlers to be called as services
- CyBox only allows 7 files to be uploaded at a time
    - This is problematic as we will be using thread to upload each individual APK. We will need to limit the threads so that CyBox does not throw errors.
- UpToDown's unique page structure is making some metadata difficult to obtain

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Mitchell Kerr	APKMirror can now reach every application and find metadata	6	22
Connor Kocolowski	Fixed apk download process (apk link tag was changed on apkPure site).	5	22
Emmett Kozlowski	Worked on designing the database model	5	19
Jacob Stair	UpToDown crawler now obtains basic metadata	5	22
Matthew Lawlor	Looked into the Box SDK and was able to upload a file to CyBox with Python.	3	21

Plans for Coming Week

- Add functionality to each crawler to collect all the application information on the store
- Continue work on APKMirror and uptodown app crawlers
    - Reach majority of applications and find all relevant metadata
    - Look for similarities across crawlers to add abstraction
- Begin work on additional crawlers
- Research possibility of creating python library for crawling app stores
- Set up vm to begin running apkpure app crawler and to correctly upload data into our database

- Upload multiple APK files to CyBox
- Flask App to communicate with database