C39 Android™ App Forensic Evidence Database (AndroidAED)

Chen Shi, MS, Iowa State University, Ames, IA 50011; Chao-Chun Cheng, Iowa State University, Ames, IA 50011; Connor Kocolowski, Iowa State University, Ames, IA ; Emmett Kozlowski, BS, Iowa State University, Ames, IA 50014; Justin Kuennen, BS, Iowa State University, Ames, IA 50014; Matthew Lawlor, BS, Iowa State University, Ames, IA 50014; Mitchell Kerr, BS, Iowa State University, Ames, IA 50014; Jacob Stair, BS, Iowa State University, Ames, IA 50014; Zhonghao Liao, PhD, Iowa State University, Ames, IA 50011-1153; Zhenqiang Gong, PhD, Iowa State University, Ames, IA 50011; Yong Guan, PhD, Iowa State University, Ames, IA 50011*

Learning Overview: After attending this presentation, attendees will better understand how AndroidAED will be beneficial for academic researchers whose studies relate to mobile applications that grant them the ability to search through many of the available applications across various third-party app stores.

Impact on the Forensic Science Community: AndroidAED, per this study's research, is the first Android™ app forensic evidence database with the highest precision and the most comprehensive coverage in discovering evidence generated from Android™ apps. This presentation will impact the forensic science community by illustrating how digital forensic investigators can significantly improve the process of investigation of evidence from mobile devices.

With more than 2.7 billion smart phone users across the world, it is no surprise that the mobile app industry is thriving.[1] Android's™ smart phone share increased to 86.7% in 2019 and has remained at more than 75% in the past ten years. A recent study of globally available apps by AndroZoo shows that the number of apps has exceeded eight million and is still rapidly growing.[2] Current digital forensic practices for finding data on these apps is limited, time consuming, and error prone. As an example, Cellebrite® UFED, a commercial mobile device forensic tool, only supports the profiles of 6,000 apps.[3] If mobile devices contain apps that are not included in the 6,000 apps, it cannot work properly. Therefore, an up-to-date, real-world evidence database of Android™ apps from multiple app stores—AndroidAED—is being built employing two analyzing tools developed by this study, namely static and dynamic EviHunter, respectively.[5,6] Web crawlers are created to traverse app stores to collect metadata and download application files. Since each site has a different design and functions, a crawler tailored to each store to collect data was created. The crawlers utilize Beautiful Soup to scrape each webpage and the Selenium driver for webpage interactions. Currently, seven crawlers have been completed and 40 more are under development. The currently supported app markets include: Google® Play Store, APKPure, Uptodown, APKMirror, Aptoide, and F-Droid.[4,7-11] After collecting the file, it will be processed using the forensic analysis tools, collecting where the application is storing the evidentiary information. The information about the application that is collected from each store webpage gets stored into a MongoDB instance. The actual APK files of the applications are stored in a separate file system, and the path to the file is linked to the app entry in MongoDB. In this way, digital forensic investigators can simply query the database to find all the possible evidence data (e.g., locations, photos, call logs, time, etc.) generated by the app and corresponding evidentiary file path. Moreover, considering that the apps installed on the suspect's device can vary on the app's version and source of installation (app store), AndroidAED hosts all the available versions of apps that have been collected from various app markets.

The main contributions of AndroidAED are summarized as follows: (1) AndroidAED, per research, is the first Android™ app forensic evidence database with the highest precision and the most comprehensive coverage in discovering evidence generated from Android™ apps; (2) AndroidAED will be accessible for digital forensic investigators and significantly improves the investigation of evidence from mobile devices; (3) AndroidAED will be available for academic researchers whose studies relate to mobile applications that grant them the ability to search through many of the available applications across various third-party app stores; and (4) AndroidAED will keep updating to provide the most up-to-date evidentiary data for real-world apps, ranging from very popular to very unpopular, available from app stores around the world.

Reference(s):

1. *Statcounter*, 2019. [Online]. http://gs.statcounter.com/os-market-share/mobile/worldwide.
2. K. Allix, T.F. Bissyandé, J. Klein, and Y. Le Traon., Androzoo: Collecting millions of android apps for the research community. *ACM MSR*, 2016.
3. *Cellebrite UFED ultimate*, 2019. [Online]. https://www.cellebrite.com/en/products/ufed-ultimate/.
4. Google Play Store, 2019. [Online]. https://play.google.com/store.
5. C. C.-C. Cheng, C. Shi, N. Z. Gong, and Y. Guan. Evihunter: Identifying digital evidence in the permanent storage of android devices via static analysis. In *ACM CCS*, 2018.
6. Z. Xu, C. Shi, C. C.-C. Cheng, N. Z. Gong, and Y. Guan. A dynamic taint analysis tool for android app forensics. In *SADFE*, 2018.
7. Apkpure. 2019. [Online]. Available: https://apkpure.com/.
8. Uptodown. 2019. [Online]. Available: https://en.uptodown.com/.
9. Apkmirror. 2019. [Online]. Available: https://www.apkmirror.com/.
10. Aptoide. 2019. [Online]. Available: https://en.aptoide.com/.
11. F-droid. 2019. [Online]. Available: https://f-droid.org/en/packages/.

Android™, Database, Evidence