

A Study on Google Play Store

Sunny Arora¹, Jatinder Singh Bal²

^{1,2}Guru Kashi University, Talwandi Sabo

Abstract

Many Android users download and use third-party apps from the Google Play store, but very little is known about these apps as a whole. In order to index and analyse more than 1,100,000 Android apps every day, we built PlayDrone, the first scalable crawler for the Google Play store. As a result of PlayDrone's hacking techniques, over 880,000 free Android apps can be accessed online, including their source code, for the first time.

The difficulty of large-scale monitoring has an impact on our knowledge of app market dynamics. When it comes to things like app updates, the effect of developer actions on app popularity, and coveted spots in the best-of lists, this is especially true. Keeping categories of apps intact is a constant challenge in the rapidly evolving app market ecosystem. During the registration process, app developers must choose the category they believe best fits their app. Approach leaves room for misuse and gaming by the registrant, in addition to the ambiguity in selecting the right category. Every now and then, the app store will update its list of categories and possibly re-categorize apps.

Keywords: Google play Store, Android applications, Google play overview.

1. Introduction

Users can access a wide range of third-party applications via the Google Play store. Millions of Android phone and tablet owners sign up for Google and third-party services to download and use these applications. Millions of people download content from the Google Play store, despite the fact that the content is largely unchecked by the store's administrators.

PlayDrone, the first scalable Google Play store crawler and application analysis framework, has been developed by our team. four key methods are employed by PlayDrone. There are a number of ways to get around Google's security measures, and PlayDrone is one of them. Dictionary-based attacks are used to find applications, Furthermore, in order to communicate with Google Play's servers, the Android client for Google Play has been recompiled to use insecure communications protocols. One of the advantages of PlayDrone is its use of higher-level languages and frameworks, which allow it to perform highly concurrent, distributed processing with minimal effort. Besides Ruby, PlayDrone uses Sidekiq and Redis key-value stores. PlayDrone's performance can be boosted by adding a cluster of servers, allowing it to efficiently crawl the Google Play store as its content grows. Third, all of PlayDrone's metadata and decompiled sources are kept in a Git repository. PlayDrone analyses the evolution of Google Play store content to use this simple versioning system to track and manage multiple versions of each app. Elasticsearch is used as a distributed real-time search and analytics engine to make it simple to analyse and explore the Google Play store metadata and content using an indexing schema derived from the Google Play store API.

As mobile device technology has advanced, users have been able to transform from mere consumers of technology to creators of new mobile experiences. New mechanisms for software distribution, such as app markets like Google Play, collect developer-written software and make it available to smartphone users. In contrast to the desktop paradigm, where software is distributed directly from developers, this centralised approach to software distribution is implemented.

The impact of market interactions on future technology is influenced by both developers and users. As a result, both developers and consumers are affected by a lack of knowledge about popular app markets. Even in the case of app stores, there is no guidance for users when it comes to making a choice between apps that claim to be similar in functionality, for example.

Manufacturers of mobile devices and platforms have set up app markets where developers can sell their apps to potential customers. Over a million apps are available in over 15 different categories on the two most popular app stores, the APPLE App Store and the GOOGLE Play Store. Even more companies, such as mobile phone service providers, are now creating their own app stores of their own. App stores are becoming more difficult to navigate and distinguishing between different types of apps is becoming more difficult as they grow in popularity. Thus, app developers find it difficult to get their products noticed. There is still a long way to go before "app search optimization" technology catches up to web search.

2. Google Play Overview

App Distribution Channel: Google Play, the company's app store, can be found here. Each app submitted by a developer receives its own webpage in addition to the Google Play home page and the search interface. It contains meta-information about the application that keeps track of relevant information about it (e.g., name, category, version, size, prices). In addition, Google Play categorises apps from "Arcade & Action" to "Weather.". Apps of interest can be downloaded and installed by users, who can then review them. A review can be given a score of 1 to 5 stars. The average of all user ratings is used to assign a star rating to each app. The app's website also includes information on how many people have used it (e.g., rating, number of installs, user reviews). When deciding whether or not to install a new application, this information is used by users.

App Development: In order to submit an app to Google Play, an Android developer must first pay \$25 for a publisher account. It's a win-win situation, as the fee encourages higher-quality products while discouraging spam. Google Play allows developers to submit as many apps as they want. Automated malware scanning is provided by the Bouncer [3] service on Google Play to reduce spam. Developers have the option of charging a fee for their apps or distributing them free of charge.

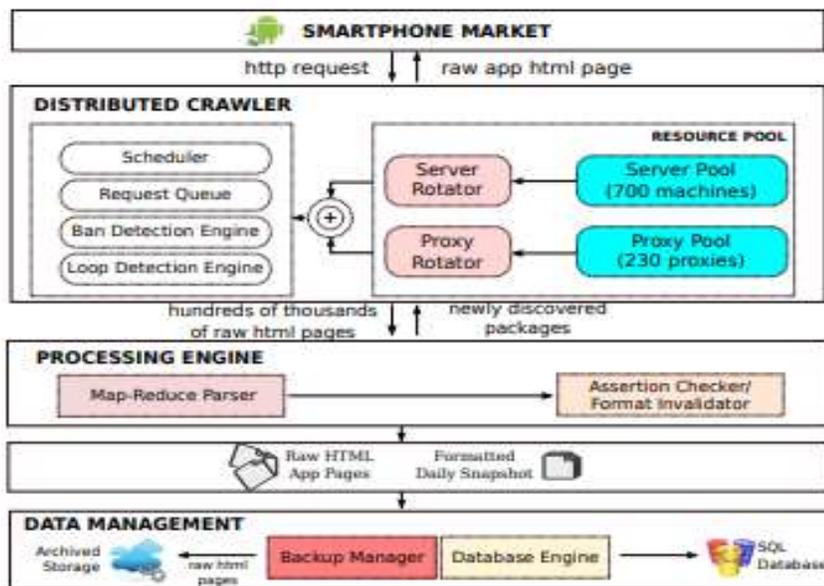


Figure 1. Architecture of iMarket, the developed GooglePlay crawler. It consists of a distributed crawler, processing engine and data management components.

Permission Model: This model of security is called Capability-based. A manifest file called `AndroidManifest.xml` is required for each app to declare the capabilities (permissions) it requires. When an app is downloaded from the Google Play website, the user is presented with a screen that displays the permissions requested by the app. A "all-or-none" approach to installing the application means granting it all of the requested permissions.

Interfacing with Google Play

As a result, PlayDrone requires an account with a Google account in order to crawl the Google Play store effectively. This could have resulted in the accounts being suspended by Google, so we decided to harvest a large number of accounts. We had to deal with two issues in order to accomplish this quickly and effectively. In order to create a Google account, you must first solve a CAPTCHA. If an IP attempts to register more than five Google accounts on the same day, Google requires phone verification.

By utilising a crowdsourcing Internet marketplace service, we were able to address both of these issues at a very low cost. Any such service, it's possible to make use of CAPTCHA solver services like Death by Captcha. It was created to allow users to submit their Google account information back to us via Amazon's Mechanical Turk service, <http://playdrone.io>. It's a service that compensates registered users for doing menial manual labour, such as checking email. We put out a call for help on Mechanical Turk, and the instructions were as follows: (1) Open a new tab in an incognito or guest browser window. You can sign up for a Google account at <https://accounts.google.com/SignUp>. (2) Except for "Mobile phone" and "current email address," all of the requested information should be completed except for those two. (3) Sign in to your new account at <http://playdrone.io/accounts/new> with your email address and password. (4) Reply with the verification code that was returned. By providing a return confirmation code, the user can be reimbursed for their time and effort. Google account information submitted to playdrone.io is validated to ensure that it is not duplicative and can authenticate with Google services before the code is given out. Dedicated users can only complete one task at a time because Mechanical Turk prevents them from doing the same thing twice.

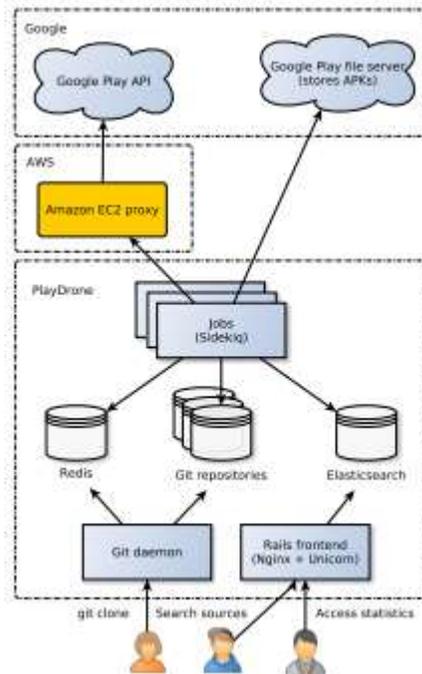


Figure 2. PlayDrone crawler architecture.

3. Google Play Characteristics

PlayDrone uses a comprehensive index of Google Play application sources and metadata to present aggregated characteristics of Android apps. Table 2 lists all of the 887,220 Google Play store apps that PlayDrone has identified as either free or paid, as well as the categories under which they fall. Since they are displayed in a separate top-level directory in Google Play, game applications are listed and classified separately. Each application can only be listed in one category in Google Play, and the categories are arranged from most to least popular. Over 3.5 times as many free apps as paid apps are available today on the App Store. As shown in Table 1, paid applications make up only 0.05 percent of total downloads from the store.

Number of Applications

	June 22, 2013	November 30, 2013
Free apps	691,518	994,216 (+28%)
Paid apps	195,702	223,258 (+14%)
All Apps	887,219	1,107,475 (+25%)

Cumulative download counts (min-max)

	June 22, 2013	November 30, 2013
Free apps	22G-85G	31G-116G (+37%)
Paid apps	111M-428M	126M-488M (+14%)
All Apps	23G-85G	31G-117G (+37%)

Table 1: Number of applications and cumulative download counts on June 22, 2013 and November 30, 2013.

4. Conclusion

Google Play, a major app store, is the subject of this study, which examines trends over time. Using a variety of hacking techniques, we've created PlayDrone, a system that can successfully search through Google Play. Adding more servers is all it takes for PlayDrone to grow, and it's fast enough to crawl Google Play every day, downloading and decompiling over 1.1 million Android apps. With PlayDrone, we can characterise Android apps in Google Play on a massive scale and show how the content of apps changes over time, despite Google Play having thousands of apps, only a small percentage of those are free, despite the fact that free apps account for the vast majority of downloads. Customers at risk were identified and alerted by working with service providers such as Amazon, Facebook and Google. Use PlayDrone to better understand Android apps and improve the quality of content in Google Play, as demonstrated by these results. According to the results of this research, 'family' and 'games' apps in the Google Play store are the most downloaded.

5. References

1. B.-G. Chun, S. Ihm, P. Maniatis, M. Naik, and A. Patti. *CloneCloud: Elastic Execution Between Mobile Device and Cloud*. In *Proceedings of the 6th European Conference on Computer systems (EuroSys 2011)*, Apr. 2011.
2. J. Crussell, C. Gibler, and H. Chen. *Attack of the Clones: Detecting Cloned Applications on Android Markets*. In *Proceedings of 17th European Symposium on Research in Computer Security (ESORICS 2012)*, Sept. 2012.
3. J. Crussell, C. Gibler, and H. Chen. *AnDarwin: Scalable Detection of Semantically Similar Android Applications*. In *Proceedings of 18th European Symposium on Research in Computer Security (ESORICS 2013)*, Sept. 2013.
4. N. d'Heureuse, F. Huici, M. Arumathurai, M. Ahmed, K. Papagiannaki, and S. Niccolini. *What's App?: A Wide-Scale Measurement Study of Smart Phone Markets*. *Mobile Computing and Communications Review*, 16(2):16–27, Apr. 2012.
5. A. P. Felt, M. Finifter, E. Chin, S. Hanna, and D. Wagner. *A Survey of Mobile Malware in the Wild*. In *Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices (SPSM 2011)*, July 2011.
6. Abu Farha, I., & Magdy, W. (2021). *A comparative study of effective approaches for Arabic sentiment analysis*. *Information Processing & Management*, 58(2), 102438. doi:10.1016/j.ipm.2020.102438
7. Zhang, L. and Liu, B., "Sentiment analysis and opinion Mining", In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning and Data Mining*. Springer, Boston, MA, 1152–1161, (2017). doi:10.1007/978-1-4899-7687-1_907

8. Rizk, N.M., Ebada, A. and Nasr, E.S., "Investigating mobile applications' requirements evolution through sentiment analysis of users' reviews". *11th International Computer Engineering Conference (ICENCO)*, 123-130, 2015.
9. Aldabbas H., Bajahzar A., Alruily M., Ali Adil Q., Rana M. Amir Latif, Farhan M., "Google Play Content Scraping and Knowledge Engineering using Natural Language Processing Techniques with the Analysis of User Reviews", *Journal of Intelligent Systems*, Volume 30: Issue 1, 2020.
10. E. Brynjolfsson and M. Smith. *Frictionless commerce? a comparison of internet and conventional retailers*. *Management Science*, 2000.
11. B. Carbunar and R. Potharaju. *A longitudinal study of the google app market*. In *Proceedings of the IEEE/ACM ASONAM*, 2015.
12. W. Enck, D. Ocate, P. McDaniel, and S. Chaudhuri. *A study of android application security*. In *Proceedings of the 20th USENIX Conference on Security*, 2011.
13. A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner. *Android permissions demystified*. In *Proceedings of the ACM CCS*, 2011.
14. H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*. ACM, 2012, pp. 1617–1621.
15. H. Zhu, E. Chen, H. Xiong, H. Cao, and J. Tian, "Mobile app classification with enriched contextual information," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1550–1563, 2014.