An Explorative Study of the Mobile App Ecosystem from App Developers’ Perspective

Haoyu Wang1,2, Zhe Liu3,4, Yao Guo1,4, Xiangwen Chen3,4, Miao Zhang2, Guoai Xu2, Jason Hong5
1 Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia
2 Beijing University of Posts and Telecommunications, Beijing, China
3 Key Laboratory of High-Confidence Software Technologies (Ministry of Education)
4 School of Electronics Engineering and Computer Science, Peking University, Beijing, China
5 Carnegie Mellon University, Pittsburgh, US
haoyuwang@bupt.edu.cn, {liuzhe.cs, yaoguo, cherry}@pku.edu.cn
{zhangmiao, xga}@bupt.edu.cn, jasonh@cs.cmu.edu

ABSTRACT

With the prevalence of smartphones, app markets such as Apple App Store and Google Play has become the center stage in the mobile app ecosystem, with millions of apps developed by tens of thousands of app developers in each major market. This paper presents a study of the mobile app ecosystem from the perspective of app developers. Based on over one million Android apps and 320,000 developers from Google Play, we analyzed the Android app ecosystem from different aspects. Our analysis shows that while over half of the developers have released only one app in the market, many of them have released hundreds of apps. We classified developers into different groups based on the number of apps they have released, and compared their characteristics. Specifically, we have analyzed the group of aggressive developers who have released more than 50 apps, trying to understand how and why they create so many apps. We also investigated the privacy behaviors of app developers, showing that some developers have a habit of producing apps with low privacy ratings. Our study shows that understanding the behavior of mobile developers can be helpful to not only other app developers, but also to app markets and mobile users.

Keywords
App developers; Mobile apps; Android; App ecosystem; Google Play; App clone; Mobile privacy

1. INTRODUCTION

With the prevalence of smartphones, mobile apps have seen widespread adoption. The number of apps in both Apple App Store and Google Play has surpassed the two million mark in 2016 and billions of downloads [6, 20], which makes mobile apps a big industry. Recent studies [4, 44] reported that the global mobile app revenues amounted to 41.1 billion US dollars in 2015 and the app economy could double in size to more than 100 billion dollars by 2020. At the same time, the latest estimates [32] indicate that there are 12 million mobile app developers worldwide, representing more than half of the total community, and almost half of app developers focus their attention on the Android platform, which makes mobile app market a very competitive environment.

A large mount of research work have focused on the mobile app ecosystem. One line of research focused on analyzing mobile apps and mobile app markets, including exploring the characteristics and the evolution of mobile apps and app markets [39, 22, 37, 36, 31], security and privacy analysis using either static analysis [34, 17, 28, 42, 40, 30] or dynamic analysis [16, 45, 49, 41], app repackaging detection [47, 40, 43], and mining useful information from app reviews [25, 49, 11], etc. The other line of studies mainly focused on mobile users, including mining and prediction of user behavior and demographics [38, 7, 48], investigating users’ mobile privacy concerns and preferences [28, 29, 12], and mobile app recommendation [24, 51, 50], etc.

App developers are the cornerstone of the mobile app ecosystem. Besides large corporations such as Google and Facebook, individual developers and small companies also play important roles in the app development field. However, few studies have focused on app developers, and very little is known about this part of the mobile app market ecosystem. What is the distribution of app developers in the current app ecosystem? What is the difference in the practices of large organizations, small companies and individual developers? How many of them could survive and make profit in the current mobile app ecosystem?

In this paper, we focus on analyzing the mobile app ecosystem from app developers’ perspective. We crawled the Google Play store and analyzed over 320,000 app developers1 and over 1.5 million Android apps they developed.

We first provided a characterization of all the apps and app developers from Google Play, including the difference between each app category and the distribution of app developers. Then we analyzed the number of apps released

1Here we count developers based on the developer accounts that have released apps on Google Play. Note that while each account may represent many developers from one organization, each individual/company can also apply for multiple accounts.
by each developer, which showed a wide range of distribution from 1 to over 1,000. We classified all developers into four groups based on the number of apps they released, and analyzed the characteristics for different developer groups from the aspects of app quality, developing behaviors, and privacy-related behaviors. Among many interesting results and observations, the following are most prominent:

- The top 1.32% of the apps account for more than 80% of the total number of app installs, while the top 1% of the app developers account for more than 80% of the total number of app installs.

- The developer distribution shows wide diversity for different categories. The categories with the least proportion of popular apps are Medical, Business, and Education, while the categories with the most proportion of popular apps are Photography, Games and Weather.

- Although over half of the app developers have released only one app in the Google Play market, many of them have released more than a hundred apps, or even over a thousand apps. However, a majority of these aggressive developers have released mostly cloned, low-quality apps with very few app installs.

- More than 70% of apps with severe privacy risks are created by 1% of the app developers, which indicates that some developers have the habit to create apps with low privacy ratings.

Based on these results and observations, we discuss the implications of this study, especially on how developers characteristics may affect the practice of not only app developers, but also the impact on app markets and mobile users as well. This paper makes the following key contributions:

- We provide a characterization of more than 320k developers and over 1.5 million apps from Google Play. To the best of our knowledge, this is the first work that analyzed the mobile app market ecosystem from the developer’s perspective in large scale.

- We present a categorization of app developers into 4 groups, and analyzed the difference between them. We also analyzed the developer group that released the most number of apps, exposing the “spamming” phenomenon in the mobile app market.

- We show that understanding the app developers can help mobile app markets and app users to identify low-quality apps and detect misuses, which can be beneficial to the whole mobile app ecosystem.

### Table 1: The distribution of free apps and paid apps in Google Play.

<table>
<thead>
<tr>
<th></th>
<th>#Apps</th>
<th>%Apps</th>
<th>#Installs</th>
<th>%Installs</th>
<th>#Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free</td>
<td>1,277,982</td>
<td>85.11%</td>
<td>91.3B</td>
<td>99.75%</td>
<td>321,877</td>
</tr>
<tr>
<td>Paid</td>
<td>223,573</td>
<td>14.89%</td>
<td>62,547</td>
<td>0.248%</td>
<td>62,547</td>
</tr>
<tr>
<td>Total</td>
<td>1,501,555</td>
<td>-</td>
<td>91.5B</td>
<td>-</td>
<td>338,090</td>
</tr>
</tbody>
</table>

### 2. DATA COLLECTION AND STATISTICS

#### 2.1 Data Collection

We crawled the Google Play webpages in March 2015, and created an index of more than 1.5 million Android apps, among which 1,277,982 of them are free apps. We crawled the metadata of these apps, including the app names, developer names, app ratings, the number of installs, etc. The list represents all the apps that can be crawled from Google Play at that time. The total number of apps is in line with the estimated number from other sources [5, 33].

We also downloaded all the apk files of free apps through the Google Play API. Note that we did not download the apk files of paid apps, because Google has strict restrictions on app purchase frequency and limits the number of apps that can be purchased with one credit card. The distribution of free apps and paid apps is shown in Table 1. Although about 15% of all the apps are paid apps, they only account for 0.2% of the total number of app installs. Because the number of paid apps and the accumulated installs are far less than free apps, we focus on analyzing the developers of free apps in this paper.

> About 15% of all Android apps are paid apps, however they only account for 0.2% of the total app installs.

#### 2.2 App Statistics

**App Distribution** For all the free apps, we show the distribution of their categories (used by Google Play) in Table 2. We can see that about 240,000 apps are game apps\(^2\), and roughly one million apps belong to other app categories.

**App Installs and App Ratings** Figure 1 shows the distribution of app installs and app ratings. Note that the number of app installs from Google Play is presented in ranges

\(^2\) Note that there are currently 19 game sub-categories on Google Play. In this paper, we put them all under the “Game” category due to space limitation.
such as “50,000 - 100,000”, which is not an accurate number. In this paper, we choose the lower bound as the number of app installs for each app.

More than 68% of the apps (over 0.8 million) have a very low number of installs (less than 1000), while only 10% of apps have more than 10k installs. There are more than 16,000 apps have over 1 million installs, while only 10 apps have over one billion installs, which are very popular apps such as WhatsApp, Twitter and Facebook. The average number of app installs for all the apps is about 71,000.

More than 22% of apps have a very low rating score (<3), while more than 35% of apps have a rating between 3 and 4. More than 40% of the apps achieve a high rating (>4), including roughly 10% of them with full mark (score 5).

**Popular Apps** Then we analyzed the distribution of app installs for all apps ordered by the number of installs. The result is shown in Figure 1(c). We regarded the apps with over one million installs as popular apps, which represent the top 1.3% of the most downloaded apps. There are 16,798 popular apps in our dataset. These popular apps account for more than 80% of the total installs.

The distribution of popular apps for each app category is shown in Table 2. Categories “Tools” and “Entertainment” have the most number (>1,000) of popular apps besides “Games”. In contrast, categories “Medical” and “Libraries and Demo” contain less than 50 popular apps each.

The top 1.3% of popular apps account for over 80% of the total installs.

### 3. APP DEVELOPER STATISTICS

#### 3.1 Developer Distribution

There are a total of 321,877 (free) app developers in our dataset, with each developer releasing roughly 4 free apps on average. Note that some large companies have used several developer names, for example Samsung used developer names including “Samsung Electronics Ltd.” and “Samsung Electronics Mobile Biz.”, thus the apps they released have different developer signatures. Because this scenario is rare and difficult to distinguish, we treated them as different developers in this paper.

The distribution of app developers for each app category is shown in Table 2. Besides “Games” category, the categories Entertainment and Tools attracted the most number of app developers, with more than 45,000 developers each. In contrast, only 2,000 developers have released apps in the category Libraries and Demo.

#### 3.2 App Installs Distribution for Developers

For each developer, we first analyzed the accumulated installs and average installs for all the apps they released. As shown in Figure 2, roughly 90% of developers have accumulated installs less than 100k. Only about 2.6% of the developers have accumulated installs higher than 1 million. When we look at the average installs per app for each developer, more than 96% of them are below 100k.

Table 3 shows the developers with top accumulated installs. Google is the obvious No. 1 with over 13 billion accumulated app installs, which is almost 9 times of its closest

---

Table 2: The distribution of all free apps and their developers on Google Play (March 2015).

<table>
<thead>
<tr>
<th>Category</th>
<th>#Apps</th>
<th>#Pop Apps</th>
<th>%Pop Apps</th>
<th>#Dev</th>
<th>%Pop Dev</th>
<th>%Pop Apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>books ...</td>
<td>66,845</td>
<td>247</td>
<td>0.37%</td>
<td>16,763</td>
<td>58</td>
<td>0.35%</td>
</tr>
<tr>
<td>business</td>
<td>73,729</td>
<td>123</td>
<td>0.17%</td>
<td>30,913</td>
<td>29</td>
<td>0.094%</td>
</tr>
<tr>
<td>comics</td>
<td>5,795</td>
<td>54</td>
<td>0.93%</td>
<td>2,497</td>
<td>11</td>
<td>0.44%</td>
</tr>
<tr>
<td>community</td>
<td>28,633</td>
<td>524</td>
<td>1.83%</td>
<td>15,549</td>
<td>125</td>
<td>0.80%</td>
</tr>
<tr>
<td>education</td>
<td>90,323</td>
<td>268</td>
<td>0.30%</td>
<td>35,974</td>
<td>52</td>
<td>0.14%</td>
</tr>
<tr>
<td>entertainment</td>
<td>121,300</td>
<td>1,032</td>
<td>0.85%</td>
<td>45,900</td>
<td>194</td>
<td>0.42%</td>
</tr>
<tr>
<td>finance</td>
<td>27,065</td>
<td>200</td>
<td>0.74%</td>
<td>16,506</td>
<td>29</td>
<td>0.18%</td>
</tr>
<tr>
<td>health</td>
<td>34,275</td>
<td>196</td>
<td>0.57%</td>
<td>17,703</td>
<td>46</td>
<td>0.26%</td>
</tr>
<tr>
<td>libraries</td>
<td>4,393</td>
<td>36</td>
<td>0.82%</td>
<td>2,273</td>
<td>11</td>
<td>0.48%</td>
</tr>
<tr>
<td>lifestyle</td>
<td>92,576</td>
<td>494</td>
<td>0.53%</td>
<td>37,231</td>
<td>97</td>
<td>0.26%</td>
</tr>
<tr>
<td>video player</td>
<td>24,056</td>
<td>391</td>
<td>1.63%</td>
<td>11,633</td>
<td>99</td>
<td>0.85%</td>
</tr>
<tr>
<td>medical</td>
<td>13,617</td>
<td>25</td>
<td>0.18%</td>
<td>7,218</td>
<td>1</td>
<td>0.04%</td>
</tr>
<tr>
<td>music/audio</td>
<td>53,821</td>
<td>596</td>
<td>1.11%</td>
<td>17,249</td>
<td>155</td>
<td>0.90%</td>
</tr>
<tr>
<td>news/magaz</td>
<td>41,502</td>
<td>170</td>
<td>0.41%</td>
<td>18,266</td>
<td>30</td>
<td>0.16%</td>
</tr>
<tr>
<td>personalization</td>
<td>71,777</td>
<td>871</td>
<td>1.21%</td>
<td>12,693</td>
<td>186</td>
<td>1.47%</td>
</tr>
<tr>
<td>photography</td>
<td>18,373</td>
<td>600</td>
<td>3.76%</td>
<td>8,310</td>
<td>157</td>
<td>1.89%</td>
</tr>
<tr>
<td>productivity</td>
<td>31,371</td>
<td>515</td>
<td>1.64%</td>
<td>20,464</td>
<td>113</td>
<td>0.55%</td>
</tr>
<tr>
<td>shopping</td>
<td>22,486</td>
<td>317</td>
<td>1.41%</td>
<td>12,876</td>
<td>84</td>
<td>0.65%</td>
</tr>
<tr>
<td>social</td>
<td>27,840</td>
<td>346</td>
<td>1.24%</td>
<td>17,628</td>
<td>107</td>
<td>0.61%</td>
</tr>
<tr>
<td>sports</td>
<td>32,756</td>
<td>142</td>
<td>0.43%</td>
<td>15,355</td>
<td>25</td>
<td>0.16%</td>
</tr>
<tr>
<td>tools</td>
<td>82,063</td>
<td>1,333</td>
<td>1.62%</td>
<td>45,452</td>
<td>288</td>
<td>0.63%</td>
</tr>
<tr>
<td>transport...</td>
<td>16,126</td>
<td>119</td>
<td>0.74%</td>
<td>10,356</td>
<td>22</td>
<td>0.21%</td>
</tr>
<tr>
<td>travel...</td>
<td>51,806</td>
<td>247</td>
<td>0.48%</td>
<td>22,966</td>
<td>53</td>
<td>0.23%</td>
</tr>
<tr>
<td>weather</td>
<td>5,377</td>
<td>120</td>
<td>2.32%</td>
<td>2,903</td>
<td>35</td>
<td>1.20%</td>
</tr>
<tr>
<td>Games-all</td>
<td>240,003</td>
<td>7,742</td>
<td>3.23%</td>
<td>72,349</td>
<td>1,998</td>
<td>1.79%</td>
</tr>
<tr>
<td>Total</td>
<td>1,277,908</td>
<td>16,798</td>
<td>1.31%</td>
<td>321,877</td>
<td>3,248</td>
<td>1.01%</td>
</tr>
</tbody>
</table>

Table 3: Top 10 app developers with the most number of accumulated installs.

<table>
<thead>
<tr>
<th>Developer</th>
<th>#Apps</th>
<th>Total Installs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Inc.</td>
<td>128</td>
<td>13.2 billion</td>
</tr>
<tr>
<td>Facebook</td>
<td>11</td>
<td>1.5 billion</td>
</tr>
<tr>
<td>Samsung Elec. Co.</td>
<td>20</td>
<td>1.2 billion</td>
</tr>
<tr>
<td>WhatsApp Inc.</td>
<td>2</td>
<td>1.1 billion</td>
</tr>
<tr>
<td>Outfit7</td>
<td>25</td>
<td>817 million</td>
</tr>
<tr>
<td>Rovio Ent. Ltd.</td>
<td>21</td>
<td>591 million</td>
</tr>
<tr>
<td>Gameloft</td>
<td>79</td>
<td>570 million</td>
</tr>
<tr>
<td>Instagram</td>
<td>4</td>
<td>560 million</td>
</tr>
<tr>
<td>Skype</td>
<td>5</td>
<td>511 million</td>
</tr>
<tr>
<td>Twitter, Inc.</td>
<td>3</td>
<td>510 million</td>
</tr>
</tbody>
</table>

Table 4: App developers with top average installs.

<table>
<thead>
<tr>
<th>Developer</th>
<th>#Apps</th>
<th>Average Installs</th>
</tr>
</thead>
<tbody>
<tr>
<td>WhatsApp Inc.</td>
<td>2</td>
<td>525 million</td>
</tr>
<tr>
<td>Flipboard</td>
<td>2</td>
<td>250 million</td>
</tr>
<tr>
<td>Twitter, Inc.</td>
<td>3</td>
<td>170 million</td>
</tr>
<tr>
<td>Instagram</td>
<td>4</td>
<td>140 million</td>
</tr>
<tr>
<td>Facebook</td>
<td>11</td>
<td>138 million</td>
</tr>
<tr>
<td>Google Inc.</td>
<td>128</td>
<td>103 million</td>
</tr>
<tr>
<td>Skype</td>
<td>5</td>
<td>102 million</td>
</tr>
<tr>
<td>DU APPS STUDIO</td>
<td>2</td>
<td>100 million</td>
</tr>
<tr>
<td>Pinterest</td>
<td>1</td>
<td>100 million</td>
</tr>
<tr>
<td>Venticake Inc.</td>
<td>1</td>
<td>100 million</td>
</tr>
</tbody>
</table>

Note that one developer could belong to several categories.
3.3 Which Category is the Most Competitive?

The distribution of developers is fairly diverse for the different app categories. As such, we investigated the difference between each app category and analyzed which app category is most competitive for app developers.

We first analyzed the distribution of popular apps and popular app developers, as shown in Table 2. We found that some app categories are obviously more competitive than other categories. For 27 out of 43\(^4\) app categories, less than 1\% of the app developers have created at least one popular app. App categories such as Business, Entertainment, and Tools attracted plenty of app developers, however, very few of them have created popular apps. Although there are roughly 31k developers work on Business category, less than 0.1\% of them own popular apps, which makes this category a very competitive environment. In contrast, apps and developers in category Photography are easier to achieve popularity, with 3.76\% of the apps and 1.89\% of the developers in this category are popular. The Game app category is generally more popular than other app categories on average. For example, more than 10\% of the apps in sub-category Role Playing (Game) are popular apps (not shown in the Table).

The app categories with the least percentage of popular apps are Medical, Business, and Education. The top app categories with the most percentage of popular apps are Photography, Games and Weather. This result could offer insights to new app developers when they choose which types of apps they choose to work on.

3.4 The Number of Apps Released

Figure 4 shows the number of apps created by each developer. Although over half of the developers have released only one app in Google Play and roughly 90\% of them have created less than 10 apps, many of them (more than 500) have released more than a hundred apps, while some developers have released even over a thousand apps. For example, the developer that released the most number of apps is “Subsplash Consulting”, who has created 1,107 apps.

4. DEVELOPER CHARACTERISTICS

In order to understand the difference between these developers, we first categorized the developers based on the number of apps they have created, and compared their characteristics from various aspects. We then focused on analyzing the aggressive developers who have created at least 50 apps, to understand the reasons behind their behaviors.

4.1 Developer Classification

In order to understand the diversity between different developers, we first classified all the app developers into 4 groups based on the number of apps they released:

- Aggressive Developers. Developers who created more than 50 apps. As shown in Table 5, although only roughly 0.6\% of developers belong to this category, they accounted for 17\% of the total number of apps.

\(^4\)including 19 sub-categories of GAME apps.
and 24.3% of the total number of app installs. However, if we took out Google, which had more than 13 billion total app installs, all the other aggressive developers took up only 9.84% of the total installs.

- **Active Developers.** Developers who created 10 to 50 apps. This group is consist of roughly 6.3% of the total developers, but they occupied 35% of the apps and roughly 30% of the total app installs.
- **Moderate Developers.** Developers who created 2 to 10 apps, which take up about 40.8% of the developers. They accounted for 37% of the apps and 36% of the total app installs.
- **Conservative Developers.** Developers who created only one app. This group, with more than 50% of the developers, accounted for only 13% of the total apps and 10% of the total app installs.

### 4.2 App Popularity vs. Developer Groups

We examined the relationship between app popularity and developer groups. As shown in Table 5, the more apps they created, the higher chance they could be popular. More than 13% of the **Aggressive Developers** are popular developers, while only 0.28% of the **Conservative Developers** are popular developers.

We also compared the number of popular apps released by different developer groups. As shown in Table 5, **active developers and moderate developers** account for most of the popular apps. It is interesting that although the percent of popular developers in the aggressive developer group is much higher than other categories, the percent of popular apps is lowest in these four groups. We will investigate the aggressive developers who have created many apps in detail later.

Figure 5 and Figure 6 show the distribution of accumulated installs and average installs for different developer groups. Developers who created more apps are likely to have more accumulated installs. As to the average app installs, **active developers perform the best. More than 10% of active developers have average app installs higher than 1 million.**

### 4.3 App Ratings vs. Developer Groups

We also compared the average ratings for all apps released by each developer, as shown in Figure 7. The average rating distribution of developers for the 4 groups are similar, which means that there is no obvious relationship between developer groups and average developer ratings. The average developer ratings of **conservative developers** are slightly better than other groups. The reason is that the average installs of conservative developers are fewer than other groups on average, thus the apps they created have fewer user reviews, which makes the app rating easier to control (e.g., add votes themselves).

### 5. DEMYSTIFYING AGGRESSIVE DEVELOPERS

In this section, we focus on aggressive developers that created at least 50 apps. We all understand that large corporations (e.g., Google) have wide business and they have the money and incentive to develop many mobile apps. However, there are more than 2,000 aggressive developers, most of them are very different from companies like Google.

We attempt to answer the following questions:
- Which kinds of organizations/companies/individuals have released so many apps?
- How do these developers create these apps?
- How many of these developers are popular and thus potentially profitable?
- What can we learn from these aggressive developers?

#### 5.1 Overall Distribution

Among the 2,067 aggressive developers, more than 70% of them (1,481 developers) have released 50 to 100 apps. Over 400 developers created 100 to 200 apps, and more than 140 developers created more than 200 apps.

As shown in Figure 5 and Figure 6, although some aggressive developers have large accumulated/average app installs, 40% of them have accumulated installs less than 100k, and more than 70% of them have average installs less than 10k.

---

**Table 5: The distribution of different developer groups.**

<table>
<thead>
<tr>
<th>Developer Category</th>
<th>#Developers</th>
<th># Popular Developers</th>
<th>% Popular Developers</th>
<th># Apps</th>
<th># Popular Apps</th>
<th>% Popular Apps</th>
<th>% Total Installs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressive (&gt; 50 apps)</td>
<td>2,067</td>
<td>277</td>
<td>13.40%</td>
<td>211,908</td>
<td>2,387</td>
<td>1.13%</td>
<td>24.30%</td>
</tr>
<tr>
<td>Active (10–49 apps)</td>
<td>20,299</td>
<td>1,085</td>
<td>5.35%</td>
<td>421,272</td>
<td>5,797</td>
<td>1.38%</td>
<td>29.76%</td>
</tr>
<tr>
<td>Moderate (2–9 apps)</td>
<td>131,225</td>
<td>1,422</td>
<td>1.08%</td>
<td>476,442</td>
<td>6,639</td>
<td>1.39%</td>
<td>35.59%</td>
</tr>
<tr>
<td>Conservative (1 app)</td>
<td>168,286</td>
<td>464</td>
<td>0.28%</td>
<td>168,286</td>
<td>1,975</td>
<td>1.17%</td>
<td>10.34%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>321,877</td>
<td>3248</td>
<td>1.01%</td>
<td>1,277,908</td>
<td>16,798</td>
<td>1.31%</td>
<td>100%</td>
</tr>
</tbody>
</table>

---
5.2 Who are These Developers?

We first analyzed the top 20 aggressive developers with the most number of created apps, as shown in Table 6. It is amazing that although all of them released more than 500 apps, only three developers have popular apps. Most of these developers have average app installs less than 10k. It means that these developers go after quantity alone, rather than app quality.

We also analyzed the app categories these developers focused on. 85% of them only focused on one or two categories. Most of them work on “music and audio”, and “personalization”. These app developers are small companies or individual developers that focus on specific kinds of apps such as theme/wallpaper, books and music/audio, or app service companies who create specialized apps for other institutes (such as church and store) or events.

From another point of view, we further analyzed the top 20 aggressive developers in this group with the most number of installs, and the result is shown in Table 7. These developers go after both app quantity and app quality. All of them have at least 100 million app installs. Note that there is only one developer appeared in both Table 6 and Table 7, namely “Zt.art”. Besides large organizations such as Gameloft and LINE, there are many other companies that specialized on mobile app development, such as Go Dev Team. Almost half of these developers work on games.

5.3 How do They Create Their Apps?

We are interested in why and how these developers created so many apps, so we further investigated the apps they created. We used a combination of several approaches to analyze these apps, including using app clone detection tool [40] and third-party library detection tool [31] to identify shared code of these apps, manually inspecting app resources, and comparing the layout and UI by installing and testing sample apps on smartphones. In the end, we classified the ways that these aggressive developers releasing apps to four types.
For the first type, most of the apps they created are cloned apps, which means that these apps share almost the same code, with only app resources replaced. For example, developer “Kultida Anekboon” has created more than 750 music apps, such as “com.bsw.haney.APercentCircle” and “com.bsw.haney.ARocketToTheMoon”. These apps shared exactly the same code and only resources such as music files and lyrics are replaced. Developer “ZT.art” has created several app templates for different kinds of theme apps, and the same kinds of theme apps shared the same code with only differences in resources. Aggressive developers that focused on wallpaper/theme/books/magazines/weather categories mostly use this way to create new apps.

For the second type of aggressive developers, the apps they created have the same app structures. For example, developer “Subsplash Consulting” has released more than 1000 church apps, and these apps have similar UIs, and they share more than 60% of the code. App service/development companies that create specialized apps for other institutes (such as church and store) mostly use this way to release apps.

For the third type of aggressive developers, they are specialized on certain kinds of apps, and the apps they created share the same development frameworks and they only reuse a small portion of main code. For example, the developer who released many GAME apps usually uses the same GAME engine to create apps, but the different GAME apps do not share the main code.

For the last type of aggressive developers, they created various types of apps, and these apps have almost different functionalities. Large organizations such as Google and Samsung belong to this category.

5.4 Do They Create Popular Apps?

The app installs for aggressive developers are so polarizing. Large companies (e.g., Google) have billions of app installs, while some developers such as “By India Taps” have average app installs less than 100. As shown in Table 5, more than 13% of them are popular developers, most of which are large organization or companies that specially focused on app development. However, for over 211,000 apps they created, only 1.13% of them are popular apps.

5.5 Identifying Spammy Developers

Based on the observed characteristics of developers from Table 6, we regarded the aggressive developers with no popular apps and with an average install number lower than 10,000 as “spamming” developers.

We used this criteria to examine all the 2,067 aggressive developers along with the apps they developed. As a result, 1,597 of them (more than 77%) are identified as “spamming” developers. This result explains our observation in Table 5 that why the percentage of popular apps created by aggressive developers is the lowest.

Figure 8 shows the distribution of the main app categories of these spammy developers. Note that if more than half of the apps one developer created belong to one category, we regarded this category as the main category of the developer. One third spammy developers focused on “Books and References”, “Entertainment” and “Lifestyle” categories.

Next, we checked the apps these developers created on Google Play in October 2016 by searching the app package names, and we found that more than 60% of these apps have already been removed. The reason is that Google Play regularly remove “spammy” apps. Our study has the potential to help app markets to monitor and improve the app market as the number of apps keeps increasing.

Besides a few big companies, aggressive developers are mainly low quality developers who clone and release many similar apps.

6. PRIVACY RATINGS OF DEVELOPERS

Mobile privacy is a hot topic recent years. Many previous studies have either analyzed mobile apps to detect sensitive behaviors or proposed privacy protection approaches for mobile system. In this section, we further investigated the privacy behaviors of app developers.

We used PrivacyGrade [34] to analyze the sensitive behaviors of Android apps and assign privacy grades to each app. PrivacyGrade is based on previous research [28] that used crowdsourcing and machine-learning techniques to analyze the privacy-related behaviors of mobile apps. They categorized the purposes of several hundred third-party libraries (advertising, analytics, social network, etc), used crowdsourcing to ascertain people’s level of concern for data use (e.g. location for advertising versus location for social networking), and clustered and trained a model to analyze apps based on their similarity. The details of PrivacyGrade are omitted in this paper.

We used PrivacyGrade to assign each app a privacy score, ranging from A+ (most privacy sensitive) to D (least privacy sensitive). The distribution of privacy score for the 1.2 million free apps is shown in Figure 9. More than 87% of apps have grade A or A+, and only less than 10% of apps have low privacy score C or D, which means that these apps have potentially privacy risks.
We further analyzed the privacy behaviors of app developers, i.e., the relevance of privacy sensitivity among all apps created by each developer. Thus we first normalized the app grade to a number between 1 to 5. Then we analyzed the average privacy score for each developer, as shown in Figure 10. Roughly 7.6% of developers have low scores (less than 3), and more than 80% of developers have high scores (larger than 4), which means that most of the developers could respect user’s privacy.

We also investigated the developers with lowest privacy scores. For example, the developer “Mobile Roadie” released 499 apps, while 464 of them have the lowest privacy ratings (Score D). As shown in Figure 11, roughly 1% of developers created more than 70% of apps with potential privacy risks. This means that these developers have a tendency to release apps with high privacy risks.

More than 70% of apps with severe privacy risks are created by 1% developers.

7. DISCUSSIONS

7.1 Implications

Although our study has focused on exploring the mobile app ecosystem from the developers’ perspective, it is potentially beneficial to not only app developers, but to app markets and mobile users as well.

For Developers Our study has reveals many characteristics in mobile apps and developers, especially on which app categories are popular and competitive. These information can help developers to understand the current status of mobile app market, and realize which app categories are most competitive, and how hard it is to achieve their goals. One lesson to new developers: releasing hundreds of cloned apps to the market does not necessarily brings you popular apps and earns you any real money!

For App Markets We showed that many developers release hundreds of low-quality, least popular and high privacy risk apps, which is an obvious misuse and may cause disruption to the mobile app ecosystem. This information can also help app markets to find the abnormal developers, pay attention to aggressive developers and low privacy rating developers. We have seen that Google Play has since removed many of these developers and their apps, and we believe our study can help them further with monitoring and improving the app market as the number of apps keeps increasing.

For Mobile Users For mobile users, learning the characteristics of developers is also beneficial. As many developers have the habit to release bad and risky apps, it is better to avoid new apps released by these developers. App developers should be taken into consideration for app recommendation.

7.2 Free Apps vs. Paid Apps

In this work, we only studied the app developers who released free apps in Google Play, because paid apps only account for 0.2% of the total installs and it is hard to download paid apps for us. However, paid app is still a very important role in the app ecosystem, accounting for about 15% of the total number of apps. It will be interesting to investigate the difference between paid apps and free apps, including the difference between their developers, and further analyze their difference in how to gain profit from their apps.
8. RELATED WORK

8.1 App Market Analysis

PlayDrone [39] performed a large-scale characterization of Android apps in Google Play. They downloaded more than 1.1 million Android apps and analyzed the distribution of app developers in Google Play. Then we classified over 320k developers into four groups based on the number of apps they released, and analyzed the characteristics for different developer groups from the aspects of app quality, developing behaviors, and privacy behaviors. The results revealed the wide diversity between app developers.

One interesting group of developers is aggressive developers, who have created more than 50 apps. We conducted a detailed study of these aggressive developers to understand how and why they created so many apps. We found that a majority of these developers tend to release low-quality, less popular and high privacy risk apps. Detecting these developers can help monitor and improve the mobile app ecosystem.

8.2 Mobile App and Mobile User Analysis

One line of work focused on mobile app analysis. Plenty of research have analyzed security and privacy issues in mobile apps, using static analysis [34, 17, 18, 21, 42, 30] or dynamic analysis [16, 45, 49, 41, 27]. Some studies have analyzed mobile apps to detect repackaged app [47, 40, 43], or mining useful information from app reviews [25, 19, 11], etc.

The other line of studies mainly focus on mobile users, including mining and prediction of user behavior and demographics [38, 7, 48], investigating users’ mobile privacy concerns and preferences [28, 29, 12], investigating how users manage mobile apps to understand the preference of mobile user [26], and mobile app recommendation [24, 51, 50], etc.

8.3 Research on Mobile App Developers

Cravens et al. [14] explored the demographic and business model of app developer based on a web-based survey of 352 developers. Balebako et al. [8] have explored how app developers make decisions about privacy and security. Their findings suggested that smaller companies are less likely to demonstrate positive privacy and security behaviors. Bello-Ogunu et al. [9] developed plugins for the Eclipse IDE to guide developers on the set of required permissions when creating Android applications. Websites such as App Brain [5] have published basic statistics of app developers. However, no previous work has detailed analyzed the distribution of app developers, the difference in the practices of various kinds of developers and the privacy behaviors of app developers in large scale.

9. CONCLUSION

In this paper, we present a study of the mobile app ecosystem from the perspective of app developers based on over 1.2 million Android apps and 320,000 developers from Google Play. We first provided a characterization of mobile apps and app developers at scale, analyzed the distribution of app developers in Google Play. Then we classified over 320k developers into four groups based on the number of apps they released, and analyzed the characteristics for different developer groups from the aspects of app quality, developing behaviors, and privacy behaviors. The results revealed the wide diversity between app developers.

This work was partly supported by the National Key Research and Development Program (No. 2016YFB1000105), the National High Technology Research and Development Program of China (No. 2015AA017202), the National Natural Science Foundation of China (No. 61421001), and the Beijing Training Project for the Leading Talents in S&T (No. ljr2015002). Jason Hong’s work was supported in part by the Air Force Research Laboratory under agreement number FA8750-15-2-0281. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

10. REFERENCES
