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

Haoyu Wang[1,2], Zhe Liu[3,4], Yao Guo[3,4]*, Xiangqun Chen[3,4], Miao Zhang[2], Guoai Xu[2], Jason Hong[5]
[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. Specially, 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, 18, 21, 42, 30] or dynamic analysis [16, 45, 49, 41], app repackaging detection [47, 40, 43], and mining useful information from app reviews [25, 19, 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?*

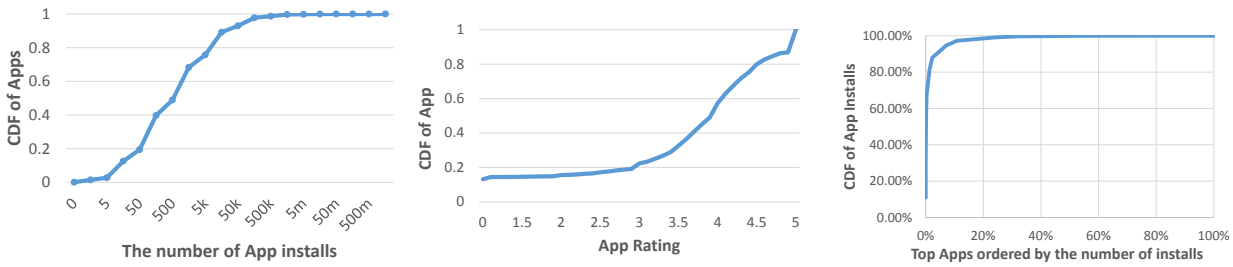
In this paper, we focus on analyzing the mobile app ecosystem from app developers' perspective. We crawled the Google Play store and analyzed over 320,000 app developers[1] and over 1.5 million Android apps they developed.

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

*corresponding author

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052712

[1]Here 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.



(a) (b) (c)

Figure 1: Distribution of app installs and app rating.

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.

	#Apps	%Apps	#Installs	%Installs	#Dev
Free	1,277,982	85.11%	91.3B	99.75%	321,877
Paid	223,573	14.89%	226M	0.248%	62,547
Total	1,501,555	-	91.5B	-	338,690

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.

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

Category	#Apps	#Pop Apps	%Pop Apps	#Dev	#Pop Dev	%Pop Dev
books ...	66,845	247	0.37%	16,763	58	0.35%
business	73,729	123	0.17%	30,913	29	0.094%
comics	5,795	54	0.93%	2,497	11	0.44%
communi...	28,633	524	1.83%	15,549	125	0.80%
education	90,323	268	0.30%	35,974	52	0.14%
enterta...	121,300	1,032	0.85%	45,900	194	0.42%
finance	27,065	200	0.74%	16,506	29	0.18%
health ...	34,275	196	0.57%	17,703	46	0.26%
libraries ...	4,393	36	0.82%	2,273	11	0.48%
lifestyle	92,576	494	0.53%	37,231	97	0.26%
video player	24,056	391	1.63%	11,633	99	0.85%
medical	13,617	25	0.18%	7,218	1	0.014%
music/audio	53,821	596	1.11%	17,249	155	0.90%
news/magaz	41,502	170	0.41%	18,266	30	0.16%
personaliz...	71,777	871	1.21%	12,693	186	1.47%
photography	18,373	690	3.76%	8,310	157	1.89%
productivity	31,371	515	1.64%	20,464	113	0.55%
shopping	22,486	317	1.41%	12,876	84	0.65%
social	27,840	346	1.24%	17,628	107	0.61%
sports	32,756	142	0.43%	15,355	25	0.16%
tools	82,063	1,333	1.62%	45,452	288	0.63%
transport...	16,126	119	0.74%	10356	22	0.21%
travel ...	51,806	247	0.48%	22,966	53	0.23%
weather	5,377	120	2.23%	2,903	35	1.20%
Games-all	240,003	7,742	3.23%	72,349	1,298	1.79%
Total	1,277,908	16,798	1.31%	321,877	3,248	1.01%

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.

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

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

Table 4: App developers with top average installs.

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

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[3]. 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

[3]Note that one developer could belong to several categories.

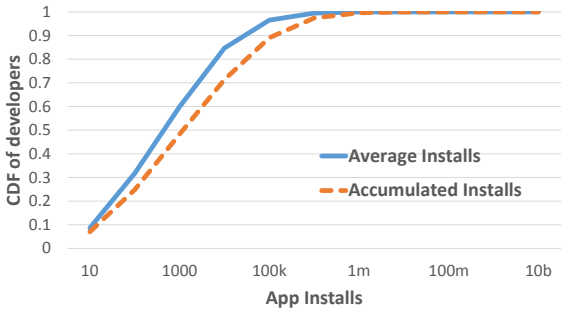


Figure 2: Distribution of *Accumulated Installs* and *Average Installs* for all app developers.

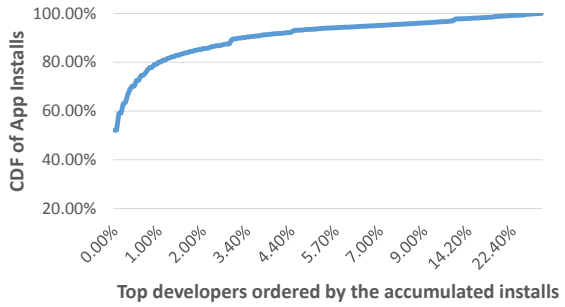


Figure 3: The proportion of the total number of app installs for top developers ordered by the Accumulated Installs.

competitor (Facebook). For the top 10 developers, only Outfit7 (the creator of app *Talking Tom and Friends*), Rovio Entertainment (the maker of app *Angry Birds*) and Gameloft (the creator of app *Dungeon Hunter*) focus on mobile app development, while the remaining developers are top IT corporations that have wide business.

Table 4 shows the top app developers with the most number of *average installs*. *WhatsApp* is in the first place, with an average installs of more than 500 million. *Flipboard* is a company focusing on magazine apps, which has an average installs of over 250 million. *Venticake* and *DU APPS STUDIO* are mobile app development companies, while other top developers are large IT corporations.

Popular Developers We then analyzed the distribution of the total number of app installs for top developers, which is shown in Figure 3. We regarded the top 1% of developers as *popular developers*, with each popular developer achieving accumulated installs higher than 3 million. In total, these popular developers occupied 80% of the total installs.

> *The top 1% of the developers with the most app installs account for 80% of the total installs.*

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

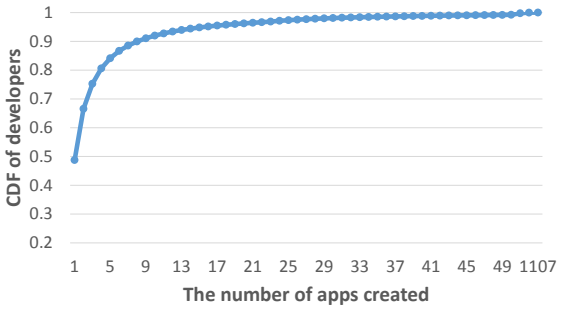


Figure 4: Distribution of the number of apps each developer created.

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.

Table 5: The distribution of different developer groups.

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

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 33% 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).

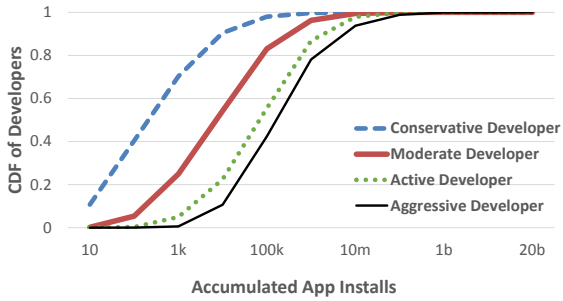


Figure 5: A comparison of the distribution of accumulated app installs for different developer groups.

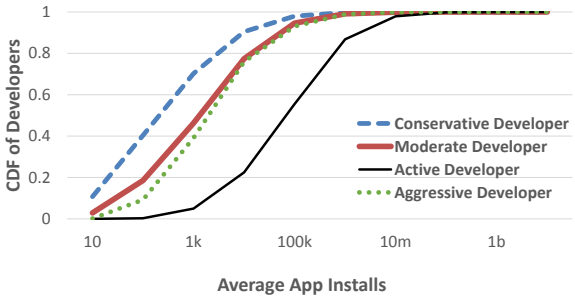


Figure 6: A comparison of the distribution of average app installs for different developer groups.

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 6: Top 20 aggressive developers with the most number of created apps.

Developer	#Apps	#Pop Apps	# Total Installs	# Avg Installs	App Categories	What is going on?
Subsplash Consulting	1107	0	1,594,013	1440	education lifestyle	Most of them are similar church apps
SkyOrange	1037	0	270,843	261	various	Most are specific apps for shop, school and other institutes
CrowdCompass	912	0	146,994	161	business	Most are specific apps for conferences or events
ZT.art	900	83	187,131,300	207,923	personalization	Most are cloned theme apps
Brainpub for Theme	806	0	5,411,160	6714	personalization	Most are cloned theme apps
Securenet Systems Inc.	777	0	753,555	970	music and audio	Most are cloned music radio apps
Tobit.Software	776	0	468,256	603	various	Most are similar news, photo and events apps
Kultida Anekboon	752	0	783,241	1042	music and audio	Most are cloned music apps
Shopgate GmbH	723	0	651,625	901	shopping	Most are similar shopping apps for specific stores
Lisbon Labs	702	1	3,490,923	4972	books	Most are cloned books and dictionary apps
Artem Chop	668	0	4,383,210	6562	personalization	Most are cloned theme apps
Magzter Inc.	640	0	685,656	1071	news&magazines	Most are cloned digital magazine apps
CTS cBroadcasting	615	0	293,410	477	media and video	Most are cloned video apps
MagazineCloner	613	0	684,620	1117	news&magazines	Most are cloned digital magazine apps
IDJ Group	604	0	375,218	621	various	Most are cloned video apps
iConnect	601	0	8,607,030	14,321	personalization	Most are cloned theme apps
Freedom Design	566	8	34,698,100	61,304	personalization	Most are cloned theme apps
Nobex Technologies	549	0	264,028	481	music & audio	Most are cloned radio apps
ReverbNation Artists	544	0	352,300	648	music & audio	Most are cloned theme apps
By India Taps	520	0	24,454	47	business	Most are cloned Yellow Pages apps
Total	14,412	92	251,069,954	17,420	-	-

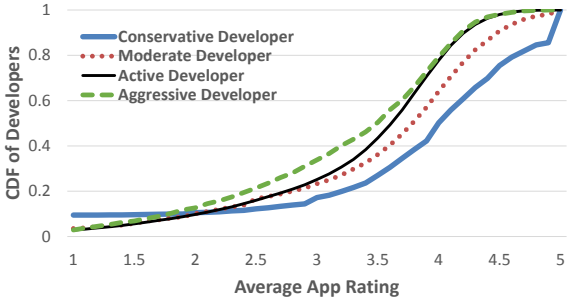


Figure 7: A comparison of the distribution of average developer rating for different developer groups.

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 *Google* and *LINE*, there are many other companies that specialized on mobile app development, such as *Gameloft* and *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.bswhaney.APerfectCircle*" and "*com.bswhaney.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 who 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 Spamming 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

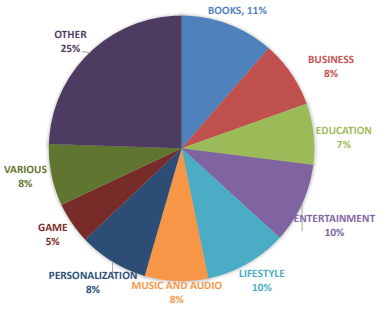


Figure 8: The distribution of main app categories of spammy developers.

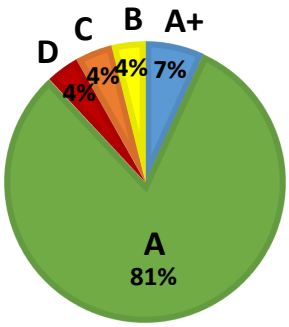


Figure 9: The distribution of privacy scores of 1.2 million free Android apps from Google Play.

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.

Table 7: Top 20 developers with the most number of app installs in the *Aggressive Developers* group.

Developer	#Apps	#Pop Apps	# Total Installs	# Avg Installs	Category
Google Inc.	128	90	13,211M	103,214,656	various
Gameloft	79	65	570M	7,215,696	game
LINE	130	76	409M	3,153,354	various
Sony	120	45	393M	3,281,400	various
6677g.com	145	99	368M	2,544,586	game
GO Dev	128	47	297M	2,322,895	various
Glu	66	46	252M	3,831,970	game
Disney	66	35	222M	3,375,759	various
ZT.art	900	83	187M	207,923	personal...
Motorola	52	22	180M	3,462,829	various
Magma Mobile	103	64	179M	1,738,942	game
GO Launcher	210	33	174M	828,729	various
TabTale	223	85	155M	695,883	game
Verizon	62	19	140M	2,266,758	various
Mobi Systems	175	10	129M	742,776	books&refs
Com2uS USA	61	34	123M	2,032,295	game
Vasco Games	343	50	121M	353,821	game
HTC	67	19	119M	1,777,448	various
Dexati	482	45	109M	227,075	game
Yahoo Japan	101	33	107M	1,062,242	various

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 (large 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

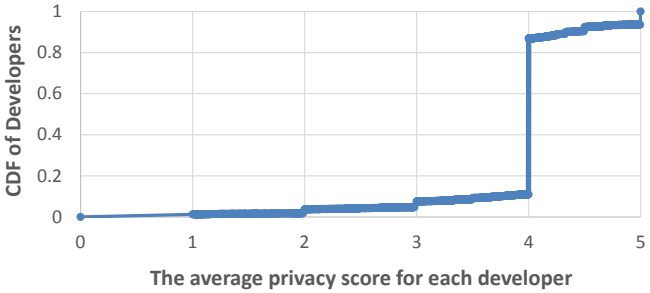


Figure 10: The distribution of average privacy score.

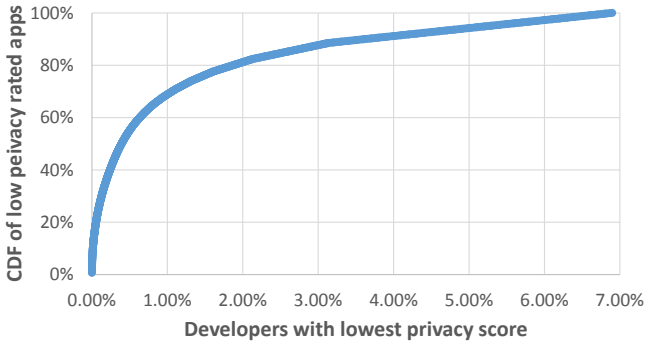


Figure 11: Developers with lowest privacy scores.

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 and explored issues such as app evolution, library usage, duplicative app content and authentication scheme in Android apps. Heureuse et al. [15] performed temporal measurement analysis of app markets including topics such as market growth, app pricing and top vendors, etc. Zhong et al. [46] examined the long tail of Google Play, which suggested that Google Play is more of a market strongly dominated by popular hit products than a "long-tail" market where unpopular niche apps aggregately contribute to a substantial portion of popularity. Holzer et al. [23] analyzed the trend of mobile app market from aspects such as portal (centralized/decentralized), devices etc, and showed implications for app developers. However, they did not analyze real app data or developer data. Some research work focused on market-level detection of mobile malware [10, 35] or grayware [1]. Besides, several companies and websites such as AppBrain [5], Android Rank [3], PrivacyGrade [34], LibRadar [13] and AndroidLib [2] offer app analysis services and regularly publish statistics about market analysis. Compared with this paper, these research/websites have not detailed further analyzed app developers yet.

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.

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.

Acknowledgment

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. 61421091), and the Beijing Training Project for the Leading Talents in S&T (No. ljrc 201502). 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

[1] B. Andow, A. Nadkarni, B. Bassett, W. Enck, and T. Xie. A study of grayware on Google Play. In *Proceedings of Workshop on Mobile Security Technologies*, 2016.

[2] AndroidLib, 2016. `http://www.androlib.com/`.

[3] Android Rank, 2016. `http://www.androidrank.org/`.

[4] App Annie Forecast, 2016. `https://www.appannie.com/cn/landing/forecast/`.

[5] AppBrain, 2016. `http://www.appbrain.com/`.

[6] Wikipedia *App Store(iOS)*, 2016. `https://en.wikipedia.org/wiki/App\_Store\_(iOS)`.

[7] R. Baeza-Yates, D. Jiang, F. Silvestri, and B. Harrison. Predicting the next app that you are going to use. In *Proceedings of WSDM '15*, pages 285–294, 2015.

[8] R. Balebako, A. Marsh, J. Lin, J. Hong, and L. Cranor. The privacy and security behaviors of smartphone app developers. In *Proceedings of the 2014 Workshop on Usable Security (USEC)*, 2014.

[9] E. Bello-Ogunu and M. Shehab. Permitme: integrating Android permissioning support in the IDE. In *Proceedings of the 2014 Workshop on Eclipse Technology eXchange*, pages 15–20. ACM, 2014.

[10] S. Chakradeo, B. Reaves, P. Traynor, and W. Enck. Mast: Triage for market-scale mobile malware analysis. In *Proceedings of WiSec'13*, pages 13–24.

[11] N. Chen, J. Lin, S. C. H. Hoi, X. Xiao, and B. Zhang. Ar-miner: Mining informative reviews for developers from mobile app marketplace. In *Proceedings of ICSE '14*, pages 767–778, 2014.

[12] E. Chin, A. P. Felt, V. Sekar, and D. Wagner. Measuring user confidence in smartphone security and privacy. In *Proceedings of SOUPS '12*.

[13] Detecting libraries in Android apps, 2016. http://radar.pkuos.org/.

[14] A demographic and business model analysis of today's app developer, 2016. http://www.doyoubuzz.com/var/f/xh/gJ/xhgJP6i_3R7alpfo5mCrKD4wLVByQzFvtIONbUGMcuX2AEWTZ-.pdf.

[15] N. d'Heureuse, F. Huici, M. Arumaithurai, M. Ahmed, K. Papagiannaki, and S. Niccolini. What's app?: A wide-scale measurement study of smart phone markets. *SIGMOBILE Mob. Comput. Commun. Rev.*, 16(2), 2012.

[16] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In *Proceedings of OSDI '10*.

[17] W. Enck, D. Octeau, P. McDaniel, and S. Chaudhuri. A study of Android application security. In *Proceedings of the 20th USENIX Conference on Security*, 2011.

[18] A. P. Felt, K. Greenwood, and D. Wagner. The effectiveness of application permissions. In *Proceedings of WebApps'11*.

[19] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh. Why people hate your app: Making sense of user feedback in a mobile app store. In *Proceedings of KDD '13*, pages 1276–1284, 2013.

[20] Wikipedia *Google Play*, 2016. http://en.wikipedia.org/wiki/Google_Play.

[21] M. Grace, Y. Zhou, Q. Zhang, S. Zou, and X. Jiang. Riskranker: Scalable and accurate zero-day Android malware detection. In *Proceedings of MobiSys'12*, pages 281–294, 2012.

[22] M. Harman, A. Al-Subaihin, Y. Jia, W. Martin, F. Sarro, and Y. Zhang. Mobile app and app store analysis, testing and optimisation. In *Proceedings of MOBILESoft'16*, pages 243–244, 2016.

[23] A. Holzer and J. Ondrus. Mobile application market: A developer's perspective. *Telematics and Informatics*, 28(1):22 – 31, 2011.

[24] A. Karatzoglou, L. Baltrunas, K. Church, and M. Böhmer. Climbing the app wall: Enabling mobile app discovery through context-aware recommendations. In *Proceedings of CIKM '12*, pages 2527–2530.

[25] H. Khalid, E. Shihab, M. Nagappan, and A. E. Hassan. What do mobile app users complain about? *IEEE Software*, 32(3):70–77, 2015.

[26] H. Li, W. Ai, X. Liu, J. Tang, G. Huang, F. Feng, and Q. Mei. Voting with their feet: Inferring user preferences from app management activities. In *Proceedings of WWW '16*, pages 1351–1362.

[27] Y. Li, Y. Guo, and X. Chen. Peruim: Understanding mobile application privacy with permission-ui mapping. In *Proceedings of UbiComp '16*, pages 682–693.

[28] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of UbiComp '12*, pages 501–510.

[29] J. Lin, B. Liu, N. Sadeh, and J. I. Hong. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Proceedings of SOUPS '14*, pages 199–212.

[30] M. Liu, H. Wang, Y. Guo, and J. Hong. Identifying and analyzing the privacy of apps for kids. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications (HotMobile '16)*, pages 105–110, 2016.

[31] Z. Ma, H. Wang, Y. Guo, and X. Chen. Libradar: Fast and accurate detection of third-party libraries in Android apps. In *Proceedings of the 38th International Conference on Software Engineering Companion (ICSE '16)*, pages 653–656, 2016.

[32] There are 12 million mobile developers worldwide, 2016. http://www.businessofapps.com/12-million-mobile-developers-worldwide-nearly-half-develop-android-first/.

[33] Number of available apps in Google Play, 2016. https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/.

[34] Privacy Grade, 2016. http://privacygrade.org/.

[35] V. Rastogi, Y. Chen, and W. Enck. Appsplayground: Automatic security analysis of smartphone applications. In *Proceedings of CODASPY '13*, pages 209–220, 2013.

[36] F. Sarro, A. A. Al-Subaihin, M. Harman, Y. Jia, W. Martin, and Y. Zhang. Feature lifecycles as they spread, migrate, remain, and die in app stores. In *Proceedings of RE'15*, pages 76–85.

[37] S. Seneviratne, H. Kolamunna, and A. Seneviratne. A measurement study of tracking in paid mobile applications. In *Proceedings of WiSec'15*, pages 7:1–7:6, 2015.

[38] S. Seneviratne, A. Seneviratne, P. Mohapatra, and A. Mahanti. Your installed apps reveal your gender and more! *SIGMOBILE Mob. Comput. Commun. Rev.*, 18(3):55–61, 2015.

[39] N. Viennot, E. Garcia, and J. Nieh. A measurement study of Google Play. In *Proceedings of SIGMETRICS'14*, pages 221–233, 2014.

[40] H. Wang, Y. Guo, Z. Ma, and X. Chen. Wukong: A scalable and accurate two-phase approach to Android app clone detection. In *Proceedings of ISSTA '15*, pages 71–82.

[41] H. Wang, Y. Guo, Z. Tang, G. Bai, and X. Chen. Reevaluating Android permission gaps with static and dynamic analysis. In *Proceedings of GLOBECOM '15*.

[42] H. Wang, J. I. Hong, and Y. Guo. Using text mining to infer the purpose of permission use in mobile apps. In *Proceedings of UbiComp '15*, pages 1107–1118.

[43] H. Wang, Z. Wang, Y. Guo, and X. Chen. Detecting repackaged Android applications based on code clone detection technique. In *SCIENCE CHINA Information Sciences*, volume 44(1), pages 142–157, 2014.

[44] Worldwide mobile app revenues in 2015, 2015 and 2020 (in billion U.S. dollars), 2016. https://www.statista.com/statistics/269025/worldwide-mobile-app-revenue-forecast/.

[45] Y. Zhang, M. Yang, B. Xu, Z. Yang, G. Gu, P. Ning, X. S. Wang, and B. Zang. Vetting undesirable behaviors in Android apps with permission use analysis. In *Proceedings of CCS'13*, pages 611–622.

[46] N. Zhong and F. Michahelles. Google Play is not a long tail market: An empirical analysis of app adoption on the Google Play app market. In *Proceedings of SAC '13*, pages 499–504.

[47] W. Zhou, Y. Zhou, X. Jiang, and P. Ning. Detecting repackaged smartphone applications in third-party Android marketplaces. In *Proceedings of CODASPY '12*.

[48] X. Zhou, S. Demetriou, D. He, M. Naveed, X. Pan, X. Wang, C. A. Gunter, and K. Nahrstedt. Identity, location, disease and more: Inferring your secrets from Android public resources. In *Proceedings of CCS '13*, pages 1017–1028.

[49] Y. Zhou, X. Zhang, X. Jiang, and V. W. Freeh. Taming information-stealing smartphone applications (on Android). In *Proceedings of TRUST '11*, pages 93–107.

[50] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Mobile app recommendations with security and privacy awareness. In *Proceedings of KDD'14*, pages 951–960.

[51] K. Zhu, X. He, B. Xiang, L. Zhang, and A. Pattavina. How dangerous are your smartphones? app usage recommendation with privacy preserving. *Mobile Information Systems*, 2016:6804379:1–6804379:10, 2016.