Why are Android Apps Removed From Google Play? A Large-scale Empirical Study

Haoyu Wang[1,2], Hao Li[3], Li Li[4], Yao Guo[5,6], Guoai Xu[1]
[1] Beijing University of Posts and Telecommunications, China
[2] Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, China
[3] OrangeApk, China [4] Monash University, Australia [5] Peking University, China
[6] Key Laboratory of High-Confidence Software Technologies (Ministry of Education), China
{haoyuwang,xga}@bupt.edu.cn,lihao_0823@yeah.net,li.li@monash.edu,yaoguo@pku.edu.cn

ABSTRACT

To ensure the quality and trustworthiness of the apps within its app market (i.e., Google Play), Google has released a series of policies to regulate app developers. As a result, policy-violating apps (e.g., malware, low-quality apps, etc.) have been removed by Google Play periodically. In reality, we have found that the number of removed apps are actually much more than what we have expected, as almost half of all the apps have been removed or replaced from Google Play during a two year period from 2015 to 2017. However, despite the significant number of removed apps, there are almost no study on the characterization of these removed apps. To this end, this paper takes the first step to understand why Android apps are removed from Google Play, aiming at observing promising insights for both market maintainers and app developers towards building a better app ecosystem. By leveraging two app sets crawled from Google Play in 2015 (over 1.5 million) and 2017 (over 2.1 million), we have identified a set of over 790K removed apps, which are then thoroughly investigated in various aspects. The experimental results have revealed various interesting findings, as well as insights for future research directions.

KEYWORDS

App mining; app store; malware; Android

ACM Reference Format:
Haoyu Wang[1,2], Hao Li[3], Li Li[4], Yao Guo[5,6], Guoai Xu[1] . 2018. Why are Android Apps Removed From Google Play? A Large-scale Empirical Study. In *MSR '18: 15th International Conference on Mining Software Repositories, May 28–29, 2018, Gothenburg, Sweden.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3196398.3196412

1 INTRODUCTION

Since the first Android version was released in 2008, Android has become the most popular platform for mobile devices such as smartphones and tablets. One reason that contributes to the success of Android could be the continuous emergence of new Android apps. As the official app market for Android apps, Google Play has included more than 3.5 million Android apps [8].

Despite the continuous increase of Android apps on Google Play, many Android apps have been removed from Google Play at the same time. In fact, Android apps can be removed for various reasons. For example, as of March 15, 2017, apps without specifically providing privacy policies could risk having Google "limit the visibility of the app" or even removed from the Play store [15]. As another example, apps targeting children that violate the COPPA policy will be also removed from Google Play [11]. COPPA [10], standing for Children's Online Privacy Protection Act, is a US federal law designed for protecting the online privacy of children. In order to build the most trusted store for Android apps, Google Play has explicitly defined various developer program policies, including the developer distribution agreement, which developers should not violate so as to ensure that their apps will not be removed from Google Play. In total, Google has defined 10 types of policies such as *privacy, security and deception* and *spam and minimum functionality* [12].

To the best of our knowledge, no existing studies have explored this direction yet. We, as a community, do not have an overall understanding on the status of removed Google Play apps, neither do we understand the practical reasons behind those removals. To this end, we perform a large-scale empirical study of removed Google Play apps aiming at observing the insights that could benefit both market maintainers and app developers. The best practices observed from the official Google Play market could also shed light on the maintenance of other popular third-party markets. The lessons learned from the removal cases could also be leveraged by app developers to avoid the unfortunate violations, e.g., apps targeting children should never violate the COPPA policy.

In this work, we conduct a large-scale empirical study on removed apps from Google Play based on two app sets collected from 2015 and 2017, which contain roughly 1.5 million and 2.1 million apps, respectively. Each app set contains an arguably complete collection of Android apps from Google Play crawled within a specified period. Because the apps are crawled using the same method, if a previously found app still exists on Google Play, it should also be found at a later time. Thus if an app is available in the 2015 set but is no longer available in the 2017 set (based on the app's unique package name), we consider this app as a removed case.

Based on the above method, we first try to identify the list of apps removed from Google Play during the 2 year span, and then perform a measurement study to understand the distribution of these apps based on their categories and developers. The results give

MSR '18, May 28–29, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5716-6/18/05…$15.00
https://doi.org/10.1145/3196398.3196412

a first impression on the landscape of the removed apps, revealing some unexpected and interesting observations:

- We found that 791,138 apps from the 2015 app set are removed from Google Play. The number is surprisingly high, as we did not expect that *more than half of the apps (out of 1.5 million) were removed from Google Play within two years.*
- Although the user ratings of the removed apps are significantly worse than the apps in general, the popularity (the number of downloads) of these removed apps are almost the same. As a surprising result, more than 500 popular apps with downloads over 1 million have also been removed.
- When we examine the distribution of developers of these removed apps, we found that although more than half of the app developers have at least one app being removed, those developers releasing the most number of apps have been affected most. Some developers with hundreds apps have 100% of their released apps removed from Google Play, partly because their apps are low-quality anyway.

To understand why these apps are removed, we then focus on the top reasons based on our manual inspection, and explore the list of removed apps from six research questions, including malicious apps, privacy-risk apps, fake apps, spamming apps, ad-blocking apps and COPPA-violated kid's apps. Our study have resulted in many interesting observations, which include: (1) apps with high privacy risks are frequently removed from Google Play, as we can see that over 77% of apps with low privacy grade ratings have been removed. (2) We are able to find many fake apps, spamming apps, malware and ad-blocking apps from the list removed apps, which confirms our speculation on the reasons of app removal in Google Play. (3) Only 16.3% of the removed kids' apps have declared privacy policies, and very few of them have disclosed the use of sensitive permissions and third-party services in the app descriptions.

We make our dataset available, including the list of removed apps, along with their metadata and the experimental results of this study, to facilitate further study along this direction. The data can be found at the following website.

https://github.com/HaoLi0823/GooglePlay2015RmvData

2 DATASET

We first need to harvest a set of Google Play removed apps. Unfortunately, to the best of our knowledge, we do not find a single resource that maintains the list of removed Google Play apps. Therefore, we seek to build such a list from scratch. Our idea is to first collect two snapshots of all the available Google Play apps at two different time points and then identify the apps belonging to the first snapshot but somehow are no longer exist in the second snapshot. The retained apps can then be safely considered as removed apps.

We use our previous dataset that are collected between February and March 2015 as the first snapshot of Google Play. This dataset contains over 1.5 million apps, where over 80% of them are free apps. Our crawling strategy starts from a small set of Google Play apps (considered as seeds) that are manually prepared. Then, we use a breadth-first-search (BFS) approach to crawl (1) related apps shown on the web pages recommended by Google Play and (2) other apps released by the same developer. The list we created represents almost all the apps that can be crawled from Google Play at that time. Note that we also downloaded all the apk files of free apps through the Google Play API [14]. Furthermore, We have also taken efforts to crawl the metadata of those apps, including app names, app descriptions, app icons, app version names, developer names, user ratings, the number of installs, the privacy policy address, etc.

For the second snapshot, by the time of this study (in September 2017), we repeated the same process as detailed before to crawl Google Play apps, except that *we take the previous 1.5 million crawled Android apps as our searching seeds*. It could ensure that we check all the 1.5 million apps in 2017[1]. Overall, we are able to collect more than 2.1 million Android apps in the new snapshot.

Table 2 shows the details of our collected datasets. It presents the results that we crawled in 2015 and 2017 separately. In addition to the total number of collected apps, Table 2 also depicts, from the third column to the last column, the number of free apps, paid apps, accumulated installs[2] and the total number of developers, respectively. It is interesting to note that except for paid apps, all the other values obtained in 2017 are higher than that of 2015. Through a manual investigation, we observe that many paid apps available in 2015 now becomes free apps, demonstrating that free apps are a trend in the Android ecosystem. Indeed, as of September 2017, over 93% of our collected Google Play apps are free ones.

3 STATISTICS OF REMOVED APPS

We now present the details of the removed apps based on the previously mentioned two datasets.

3.1 Identifying Removed Apps

Our strategy to recognize removed apps is straightforward: *given an app $a \in GooglePlay2015$ and its package name p, if we cannot find an app $a' \in GooglePlay2017$ that has the same package name p, we consider a is a removed app.*

Overall, we have identified 795,374 removed apps using this strategy, including 684,835 free ones and 110,539 paid ones, which in total have received 14.7 billion downloads and are from 186,595 developers. The first row in Table 3 illustrates these details.

Note that Google Play apps could be removed either by their developers for personal reasons (e.g., shut down of the business), or compulsively by the maintainers of Google Play. Ideally, since our objective is to understand why apps are removed by the maintainers of Google Play, we should not consider such apps that are removed by their developers. However, it is relatively difficult (nearly impossible) to check if a given app is removed by its developers. Through a manual investigation into the previously collected removed apps, we find that some removed apps actually have replacements available on Google Play. Those replacements share the same app names (although the unique package name is changed) and are published by the same developers. For example, the package name of "Opera Mini Web Browser" is "com.opera.mini.android" in our 2015 dataset, while the package name is replaced as "com.opera.mini.native" in the 2017 dataset. It is unlikely that Google Play forces the app

[1]Note that the apps shown on Google Play may differ based on the regions, thus our crawlers run on three different AliCloud servers (locating in China, Japan and USA) to make sure we can actually check whether each app is available or not.
[2]Note that the number of app installs crawled from Google Play is presented in ranges such as "5,000 - 10,000", thus in this study we choose the lower bound as the number of app installs/downloads.

Table 1: Distribution of removed apps based on their released categories on Google Play.

App Category	# Apps	# Removed	% Removed	# Free	# Removed Free	% Removed Free	# Paid	# Removed Paid	% Removed Paid
LIBRARIES AND DEMO	4,749	2,533	53%	4,394	2,360	54%	355	173	49%
WEATHER	6,142	2,654	43%	5,376	2,341	44%	766	313	41%
COMICS	7,493	5,557	74%	5,800	4,342	75%	1,693	1,215	72%
TRANSPORTATION	17,954	7,154	40%	16,128	6,531	40%	1,826	623	34%
MEDICAL	17,931	7,709	43%	13,618	5,734	42%	4,313	1,975	46%
PHOTOGRAPHY	21,270	12,487	59%	18,375	10,988	60%	2,895	1,499	52%
SHOPPING	23,244	12,724	55%	22,484	12,353	55%	760	371	49%
FINANCE	29,685	12,718	43%	27,066	11,637	43%	2,619	1,081	41%
PRODUCTIVITY	38,449	15,133	39%	31,369	12,490	40%	7,080	2,643	37%
COMMUNICATION	32,199	16,705	52%	28,632	15,046	53%	3,567	1,659	47%
SOCIAL	29,804	17,324	58%	27,840	16,251	58%	1,964	1,073	55%
MEDIA AND VIDEO	26,745	18,316	68%	24,122	16,741	69%	2,623	1,575	60%
HEALTH AND FITNESS	40,801	21,403	52%	34,281	18,558	54%	6,520	2,845	44%
SPORTS	38,629	22,061	57%	32,761	19,150	58%	5,868	2,911	50%
NEWS AND MAGAZINES	42,635	22,710	53%	41,512	22,055	53%	1,123	655	58%
TRAVEL AND LOCAL	64,161	29,393	46%	51,815	22,750	44%	12,346	6,643	54%
MUSIC AND AUDIO	60,417	37,320	62%	53,823	33,533	62%	6,594	3,787	57%
BUSINESS	77,285	39,264	51%	73,753	37,724	51%	3,532	1,540	44%
TOOLS	97,667	41,800	43%	82,069	35,435	43%	15,598	6,365	41%
EDUCATION	114,067	50,183	44%	90,323	40,606	45%	23,744	9,577	40%
BOOKS AND REFERENCE	86,559	51,322	59%	66,847	40,169	60%	19,712	11,153	57%
LIFESTYLE	102,849	57,463	56%	92,583	51,936	56%	10,266	5,527	54%
PERSONALIZATION	109,236	63,117	58%	71,776	43,133	60%	37,460	19,984	53%
ENTERTAINMENT	136,838	90,773	66%	121,318	81,100	67%	15,520	9,673	62%
GAME (ALL)	275,371	133,315	48%	240,012	118,278	49%	35,359	15,037	43%
Total	1,502,180	791,138	53%	1,278,077	681,241	53%	224,103	109,897	49%

Table 2: Details of our collected datasets.

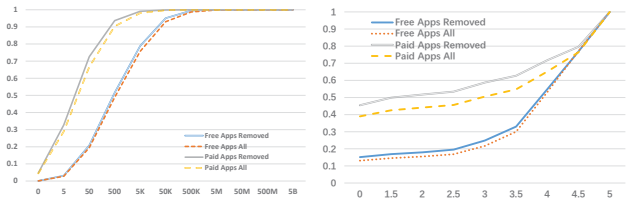
	# Apps	# Free	#Paid	Installs	# Developers
Google Play 2015	1,502,180	1,278,078	224,103	89.9B	338,670
Google Play 2017	2,144,733	2012,893	131,840	193.5B	541,105

Table 3: Details of all the removed apps. Removed apps are recognized purely based on app package names in Step 1, while some removed apps are additionally discarded in Step 2 because they could be potentially removed by their developers rather than the maintainers of Google Play.

	# Apps	# Free	#Paid	Installs	# Developers
Removed Apps (Step1)	795,374	684,835	110,539	14.7B	186,595
Removed Apps (Step2)	791,138	681,241	109,897	14.2B	184,852

developers to update the package names of their apps (resulting in removal of the original apps). In this work, we consider such replacements (i.e., only package name is changed) as developer behaviors and thereby exclude the relevant removed apps from the set of removed apps. As a result, the number of removed apps reduces from 795,374 to 791,138 (cf. Step2 in Table 3).

Overall, about 52.7% of all the apps from the 2015 dataset have been removed from Google Play, among them 53.3% of the free apps and 49.0% of the paid apps from 2015 are removed. Although we expect that many apps will be removed from the app store one way or another, *it is astonishing to see that more than half of the apps have been removed from Google Play with a little over two years*, which reveals the volatile nature of the current Android ecosystem.



(a) App downloads. (b) App ratings.

Figure 1: Cumulative distribution of removed apps based on their downloads and user ratings.

3.2 Distribution of Removed Apps

We now analyze the statistics related to the set of removed apps including app categories, app downloads and ratings.

App Category. Table 1 shows the distribution of the set of removed apps based on their releasing categories on Google Play. Note that we merge all the Game sub-categories due to space limitation. Similar to the total removal rate, most categories have a removal rate around 50%, ranging from 40% to 74%, with ***16 out of 25 categories have at least half of their apps be removed***. Some categories have more than two thirds of their apps, including the *COMICS*, *MEDIA AND VIDEO* and *ENTERTAINMENT* categories, with the *COMICS* at the highest at 74%. This results shows that some categories are more volatile than others. It is also somewhat expected as *COMICS* are time-sensitive.

Table 4: Top 10 developers with the most number of removed apps.

Developer Name	# Removed Apps	# Released Apps	% Removed Apps
-	1037	1037	100%
KoolAppz	955	955	100%
PLACE STARS, Inc.	847	847	100%
Kultida Anekboon	752	752	100%
ZT.art	739	1045	71%
Book21	721	721	100%
yama	709	709	100%
Securenet Systems Inc.	700	777	90%
City Navigator Maps	675	675	100%
Libro Movil	674	682	99%

App Downloads. Figure 1(a) further plots the distribution of the number of downloads for the removed apps, which surprising, has no big difference compared to the download distribution of all the 1.5 million apps in Google Play 2015 dataset. Although we expect that most of the removed apps are due to low-quality, so their downloads should fall into the low range. However, about 5% of the removed paid apps have the number of downloads larger than 500, while roughly 20% of the removed free apps have the number of downloads larger than 5,000.

This result suggests that ***besides low-quality apps that have few downloads, many popular apps were also removed by Google Play during the past 2 years***. For example, out data shows that 503 Android apps with downloads more than 1 million has been disappeared from Google Play, which is worth exploring in details.

App Ratings. Figure 1(b) illustrates the distribution of app ratings for all removed apps. ***Both removed free and paid apps receive relatively poor ratings compared with the overall ratings for all the apps.*** About 60% of the removed paid apps have user rating less than score 3, and 55% of the removed free apps have user rating less than score 4.

3.3 Developers of Removed Apps

Since developers play a key role in the mobile ecosystem for removed apps, and more than half of the developers have at least one app being removed from Google Play as shown in Table 3, we now analyze the characteristics of the developers of removed apps.

Top Removed Developers. Table 4 shows the top 10 developers with the most number of removed apps. Surprisingly, some developers such as *KoolAppz* have released hundreds of apps and almost all of their apps released in 2015 were removed. We manually examine the top 10 developers and find that most of their apps are cloned and repetitive apps that share almost the same code (usually only app resources are replaced). Figure 2 further summarizes the cumulative distribution of developers based on their number of removed apps. The distribution shows that ***more than 20% of the removed apps are released by only 1% of the developers, and more than 60% of the removed apps are released by roughly 10% of app developers***. This result suggests that some app developers have the tendency to release policy-violating (i.e., low-quality) apps.

Table 5: Developer categorization.

Category	# Developer	# Removed Apps	# Avg Removed Apps
Aggressive (>=50)	1,622	173,224	107
Active (10-49)	13,970	278,962	20
Moderate (2-9)	69,664	239,356	3
Conservative (1)	99,596	99,596	1
Total	184,852	791,138	4

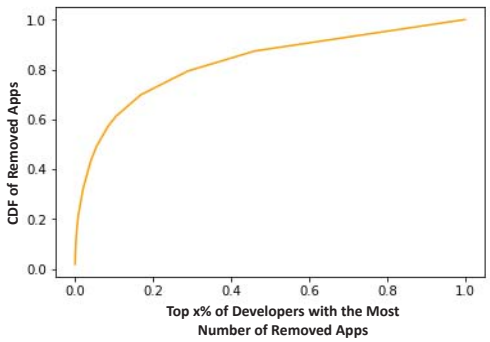


Figure 2: Top app developers with the most number of removed apps.

Developer Categorization. Previous work [83] has proposed to categorize app developers into different groups based on the number of released apps, including *aggressive developers* (released more then 50 apps), *active developers* (released 10 to 49 apps), *moderate developers* (released 2 to 9 apps), and *conservative developers* (released only 1 app). In this study, we use the same categorization to investigate how do developer categories differ in the number of removed apps. As shown in Table 5, there are 1,622 developers belonging to aggressive developers in our removed app dataset, where each developer has 107 apps removed on average. Active developers account for the most number of removed apps, with roughly 14K developers belonging to this group and each of them has 20 apps removed on average.

The results show that those developers releasing hundreds of apps are more likely to be removed, thus their apps are subject to more rigorous inspection by app store maintainers.

Spamming Developers. Previous study [83] has also proposed to identify spamming developers, where they consider aggressive developers with no popular apps (over one million downloads) and with an average install number lower than 10,000 as "spamming" developers. In this study, we use the same criteria to analyze the 338,670 developers in Google Play 2015 dataset. As a result, we are able to identify 2,122 spamming developers. These developers have released a total of 230,771 apps, where 158,647 of them (roughly 70%) are further removed by Google Play. After that, we analyze to what extent apps created by these spamming developers are removed from Google Play. As shown in Figure 3, for 55% of the spamming developers, more than 90% of the apps they released are removed. For 906 spamming developers, all of their apps released in 2015 have been removed. This result suggests that ***spamming developers are the main creators of removed apps***, thus it is

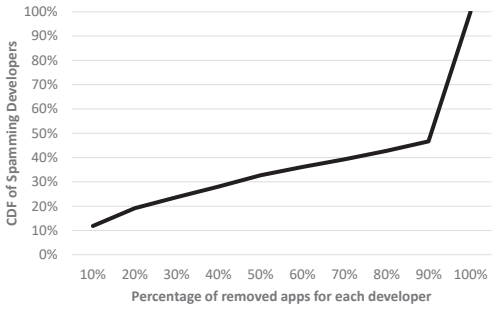


Figure 3: Cumulative distribution of spamming developers vs. percentage of removed apps for each developer.

important for market maintainers to identify these developers and remove the low-quality apps they created.

4 UNDERSTANDING APP REMOVAL

To better understand why so many apps are removed from Google Play, we now present an empirical study on the set of removed apps. In order to understand the reasons, we first create a taxonomy of removed apps based on our manual observation on potential reasons resulted from Google search. We then investigate each of these categories in detail.

4.1 A Taxonomy of App Removal Reasons

We conducted a manual investigation to understand the potential reasons why Android apps are removed from Google Play. To this end, we first thoroughly parse the Google Play policies for app developers [12] and summarize the notable reasons that in principle if violated could result in apps being removed.

Google Play has defined such possible reasons into 10 categories, for which we enumerate them in Table 6. Each category stands for a type of violation that may be associated with various instances. For example, the *Spam and minimum functionality* category contains cases where fake apps as well as spamming apps could be considered as violations that are subject to removal. To help readers better interpret those categories, we have presented in the second column of Table 6 various keywords related to different violations. The keywords are selected by the first three authors of this paper. Each author selects one to three keywords for each category, until they reach a consensus.

In addition to identify the aforementioned reasons that could happen in principle, we are also interested in such reasons that have happened in practice. To this end, we resort to Google search for relevant resources. We first crawled the top three pages (30 webpages) returned by Google with keywords "Google Play removed apps", then we asked *three independent participants* to highlight those relevant pages and cluster each of them into one of the 10 categories as a practical example. Note that a participant's selection of a reason is subjective and it is totally based upon their judgment. If a participant could not conclusively determine the reasons behind a removal, the reason will be labeled as unknown. Among the 30 pages visited, we consider 19 of them are relevant (at least two authors reach a consensus), i.e., the page is released as news or technical reports by popular websites such as Times, CNN, etc.

The clustering results are enumerated in the third column of Table 6. Obviously, the primary reason that leads to app removal from Google Play is *Privacy, security and deception* (with 11 reports). This result is somewhat expected by us as it is indeed very important in our community, considering that the Android platform now has involved over billions of end-users. The second reported reasons fall into the following categories: *Monetization and Ads*, *Spam and minimum functionality*, and *Enforcement*, where each category has exactly two reports. These three categories mainly target the functionality of Android apps, e.g., they cannot be fake apps or they cannot violate certain policies, showing that the app's implementation is also important in order to avoid potential removal. Except for the aforementioned reasons, we also find one report that goes with category *Families and COPPA*, showing that it is also important for app developers to provide a satisfactory environment for child app users.

4.2 Research Questions

The previous subsection has revealed various reasons (both in principle and in practice) that a given app could be removed from Google Play. Since it is hard to go through all the possible reasons, and some reasons are not easy to pinpoint automatically or are difficult to identify statically, in this work, we decide to focus our investigation on six research questions, which are recognized based on the five categories with practical examples.

For category ***"Privacy, security and deception"***, we mainly focus on two research questions. ***RQ1: How many malicious apps were removed from Google Play and how many users were affected by the malware?*** Despite Google Play has adopted some vetting process [2, 13] to keep malware from entering the market, malicious apps are recurrently found in Google Play [18–20]. Therefore, it is interesting to know how many of the removed apps are malware and how many users were affected. ***RQ2: To what extent are Google Play apps removed due to privacy risks?*** Privacy leak has been a long-focused issue in the Android ecosystem that needs to be avoided. Many apps require the access to sensitive data of mobile users. Google Play requires that such accesses should make sense to users, and apps should provide accurate disclosure of their functionality and should perform reasonable behaviors expected by the user [12]. As of now, lots of research studies have been proposed to detect privacy leaks in Android apps. Nevertheless, it is still unknown whether these research outputs have transferred to the removal of privacy-leaking apps.

For category ***"Spam and minimum functionality"***, we focus on two research questions. ***RQ3: Will fake apps be able to enter Google Play? If so, to what extent will they be removed eventually?*** Google has not done enough to prevent devious developers from distributing fake apps to unsuspicious users. These fake apps may disguise themselves as popular apps by simply copying the same or similar app names, icons, and other artifacts from the popular ones. It was reported that a fake *WhatsApp* app has fooled million Android users on Google Play [21]. Thus, it is interesting to explore how many of the removed apps are fake ones. ***RQ4: How many spamming apps are removed from Google Play?*** App developers and app store optimization websites may use some spamming techniques to manipulate the rank of apps or the searching results

Table 6: A taxonomy of potential reasons that Android apps could be removed from Google Play. The examples are summarized based on a thorough examination of all the news (within the top three pages) returned by Google.

Category	Keywords	Examples
Privacy, security and deception	Privacy Policy, Data Leak, Malicious Behavior	[44],[77],[65],[72],[51],[73],[64],[75],[76],[86],[63],[53]
Spam and minimum functionality	Low-quality Apps, Cloned Apps, Fake Apps, Spam	[90], [52]
Monetization and Ads	Payments, Ad-bocking	[39],[50]
Enforcement	Managing and Reporting Policy Violations	[47],[43]
Families and COPPA	Designed for Families, COPPA Compliance	[45]
Restricted Content	Sexuality, Violence, Bullying, Gambling	-
Impersonation and Intellectual Property	Impersonation, Intellectual property	-
Store Listing and Promotion	App Promotion, User Ratings	-
Other Programs	Android Instant Apps	-
Updates and Other Resources	Updates, Other Resources	-

of popular apps. Google Play does not allow apps with misleading, irrelevant, excessive or inappropriate metadata, especially with misleading references to other apps or products.

For category ***"Monetization and Ads"***, we focus on ***RQ5: how many ad blocking apps are removed from Google Play?*** More than 80% of apps in Google Play are free apps and mobile advertisement is commonly used by app developers to monetize their apps. According to the Google Play developer distribution agreements, developers should not develop or distribute apps that "interfere with, disrupt, damage, or access, in an unauthorized manner, the devices, servers, networks, or other properties or services of any third party" [12]. Such apps that do interfere with the normal way other apps operate are recognized as ad-blocking apps, for which Google Play has started to remove them [5, 6].

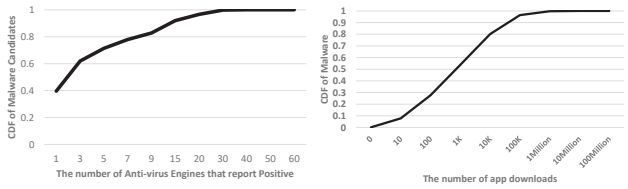
For category ***"Enforcement" and "Families and COPPA"***, we focus on ***RQ6: To what extent Android apps violate the COPPA policy? Do they behave as stated in the app description and privacy policy?*** COPPA regulates the behaviors of operators of online services (including mobile apps) that target at children under age 13 [10]. The task of enforcing COPPA falls mainly to the Federal Trade Commission (FTC) [9]. COPPA requires developers to only collect necessary information from children if it offers a clear description of what private information will be collected and for what purpose. Google Play regulates that kids' apps should be both appropriate for children and compliant with COPPA [12].

4.3 Understanding App Removal

4.3.1 ***RQ1: Malicious Apps.*** To investigate how many of the removed apps are malware, we uploaded all the removed free apps to the VirusTotal Service [35] to check whether they are malicious or not[3]. Among the 681,551 removed free apps, 126,879 of them (roughly 18.6%) are flagged by at least one anti-virus engines. The detailed distribution of the 126,879 apps is shown in Figure 4(a). More than 6% of the removed free apps (41,857 apps) are labeled as malicious by at least five anti-virus engines, and some apps are even flagged by more than 40 anti-virus engines.

Previous work [36, 85] have suggested that some anti-virus engines on VirusTotal may not always report reliable results. Hence, we empirically choose the threshold as 10 to label malware (as what have done by previous studies [36, 87]), meaning that a given app is flagged as malware if it is labeled as malicious by at least ten

[3]Note that the Public API of VirusTotal is limited to 4 requests per minute, thus we have applied 100 public APIs and distributed them to 10 servers to send requests.



(a) Anti-Virus Flags. (b) Downloads.

Figure 4: Distribution of removed malicious apps.

Table 7: Top 10 removed apps reported by the most number of Anti-virus engines from VirusTotal.

Package Name	# Downloads	# Engines
com.tmmspersonal.testeicar.free	100	51
com.ANTISPY.TESTFILE	1000	45
org.starsoftcandy.jewels	500	44
uk.co.extorian.EICARAntiVirusTest	100000	42
com.gp.solitaire	5000	42
com.exyuplus.tv	5000	41
com.storm.phonegap.miaohong	50	41
com.licravins	100000	40
com.gbbcompany.tuvi2014	1000	37
com.arkhamdev.darkarea2lite	50000	37

anti-virus engines. As a result, we are able to identify 21,833 malware samples, which account for roughly 3.2% of the removed free apps. Table 7 shows the top 10 removed apps ranked based on the number of involved anti-virus engines. We believe that ***these malware samples removed by Google could be used to supplement existing malware datasets [1, 36, 88] to support the mobile security research community for future researches.***

We then measure the infected mobile devices. Figure 4(b) shows the distribution of app downloads of these malware samples. Although roughly 96% of them have been downloaded less than 100K times, the total number of app installs for these 21,844 apps achieve 1.4 Billion. The result suggests that unsuspicious users may indeed be exposed to the threats introduced by these malware.

4.3.2 ***RQ2: Privacy-risk Apps.*** To identify privacy-risk apps, we propose to explore the gap between app behaviors and the expectation of mobile users. In this work, we take advantage of PrivacyGrade [32, 33], an open-source project with online service to analyze the sensitive behaviors of Android apps. PrivacyGrade

Table 8: PrivacyGrade assignments to the removed Google Play apps.

Privacy Ratings	# Apps	# Removed Apps	% Removed Apps
C	45,556	34,725	76.2%
D	40,108	31,386	78.3%
Total	85,664	66,111	77.2%

Table 9: Top 10 deleted privacy-risk apps with their number of downloads and privacy ratings.

Package Name	# Downloads	#Rating
com.myxer.android [26]	10 Million	D
com.fotolr.photoshake [24]	10 Million	D
com.sds.android.ttpod [29]	10 Million	D
air.com.elextech.happyfarm [22]	10 Million	D
com.galapagossoft.trialdemo [25]	10 Million	D
com.zdworks.android.toolbox [31]	10 Million	D
com.fingersoft.cartooncamera [23]	10 Million	D
com.outfit7.tomsmessengerfree [27]	10 Million	D
com.unityconceptapps.tcr.kiat [30]	10 Million	D
com.ScnStudios.PoliceCarDriver3D [28]	10 Million	D

is based on previous research [59, 60] that used crowdsourcing and machine-learning techniques to analyze the privacy-related behaviors of mobile apps. The rationale behind PrivacyGrade is that, whether the sensitive permissions should be granted is based on the purpose of permission use in the app and the expectation of mobile users [81, 82]. Based on a large amount of crowd-sourcing data, PrivacyGrade ascertains users' level of concern for data usage (e.g. location for advertising versus location for social networking) and train a machine learning model to predict the privacy risk.

We use PrivacyGrade to assign privacy grades to all the 1,278,078 free apps in our 2015 dataset. The grades are in the range of A+ (most privacy sensitive) to D (least privacy sensitive). As shown in Table 8, ***around 85,664 privacy-risk apps (with low privacy grades of C or D) are presented in the 2015 dataset, while more than 77% of them are removed by Google Play.*** Table 9 shows the top 10 removed privacy-risk apps with their number of downloads. This result suggests that Google Play continues to remove privacy-risk apps, even if they are popular ones (e.g., with 10 million downloads).

4.3.3 ***RQ3: Fake Apps***. We propose a clustering-based approach to identify fake apps in a fast manner. First, we cluster the apps based on their names that we have crawled from Google Play. Among the 1.5 million apps crawled in 2015, 1,329,508 of them are distinct from each other, leading to 11.5% (172,672 apps) of Google Play apps that share at least one name with others.

However, sharing the same app name with others does not directly mean that the app is a fake app. There may have legitimate reasons that a cluster of apps shares the same app name. Indeed, by manually examining some clusters, we find that different apps may name their apps with some common words. For example, there are 395 different apps with the name "Flashlight", and there are 255 different apps with the name "Tic Tac Toe" in our dataset. By performing a further in-depth study, we find that fake apps are usually within small clusters (e.g., size <= 5) that contain generally unpopular apps with a small number of downloads (less than one

Figure 5: Examples of Fake Apps.

order of magnitude) compared with the mimicked (original) apps. Furthermore, the clusters that contain fake apps usually have only one app (the original one) remaining in the 2017 dataset, which indicates that the fake apps are removed by Google Play.

Overall, we have identified 12,803 fake apps (roughly 1.6% of total removed apps). Figure 5 demonstrates two examples we have found in our dataset. The two apps named "eBay" in our dataset are quite easy to distinguish, where the fake app is "com.nadeveloper.ebay" and it was removed by Google Play. The other two fake apps named "Fast Racing 3D" share the same app name with the original popular app, but they have a relatively lower number of downloads. Additionally, since the package name in Google Play is case-sensitive, we find that ***some apps pretend to be the popular ones by declaring similar package names which only differ in letter cases.*** For example, there is a fake app called "com.sampleApp", which copies the same name and icon from the original app that has a package name called "com.sampleapp", which could be very deceptive for Google Play users. Therefore, the app market should pay more attention to these cases.

4.3.4 ***RQ4: Spamming Apps***. Google Play does not allow apps to use misleading, irrelevant, excessive or inappropriate metadata, especially with misleading references to other apps or products. However, spamming developers still insert irrelevant keywords (e.g., the names of popular apps) in their descriptions, so that their apps would appear popular (highly ranked) in the search results, which is a common spamming technique and even used by some app store optimization (ASO) providers.

In order to identify spamming apps with irrelevant descriptions, we resort to a specific strategy, which is to insert names of popular apps in the app's descriptions, proposed by Seneviratne *et al.* [70] that spamming developer might be interested in. To this end, we first collect the top-40 popular app names from each category (25 categories as listed in Table 1, 1,000 popular app names in total). Then, we count the number of popular app names mentioned in the descriptions of removed apps. Since popular app names could be common words, such as "Path" (package name: com.path), "Weather" (package name: com.macropinch.swan), "Circle" (package name: com.ketchapp.circle) and "Music" (package

Table 10: Top 20 popular app names mentioned in the description of removed apps.

App Name	Count	App Name	Count
Google	105,675	Apex Launcher	1,480
Facebook	72,529	Wechat	1,421
Twitter	49,599	Solitaire	1,391
Youtube	28,027	Uber	1,295
Gmail	25,286	Angry Birds	1,125
Instagram	5,390	Adobe AIR	1,120
QQ	4,887	LinkedIn	1,086
Monents	4,350	Pinterest	1,033
Wikipedia	3,918	Tumblr	893
Dropbox	1,769	ESPN	856

name: com.sonyericsson.music), which could be legitimately used by normal apps, we exclude such common words from this study.

Figure 6 shows the distribution of the number of times popular app names appeared in the app description for each app. It is interesting to observe that more than 50% of the removed apps have inserted at least one popular app name in their descriptions, and more than 5% of the removed apps have mentioned at least three popular app names in their descriptions.

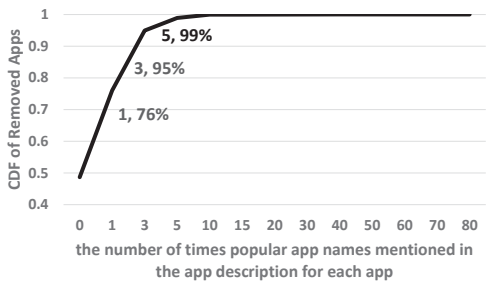


Figure 6: The distribution of the number of times popular app names mentioned in the app description for each app.

We now go one step further to investigate to what extent popular app names are mentioned in the descriptions of removed apps. The top 20 appeared names are enumerated in Table 10. Since Android apps may communicate with Google, Facebook, Twitter to provide social networking functionalities, having those keywords in the app descriptions does not directly mean that these apps are spamming apps. Thus we manually examine some descriptions of removed apps aiming at understanding how popular app names are used in practice. Our examination reveals that apps mentioning over 5 times of other popular apps are highly suspected to be spamming apps, which account for roughly 1% (8487 apps) of the removed apps. Table 11 lists the top 10 spamming apps, among which we can observe that some apps have even mentioned 80 popular app names in their descriptions, resulting in very aggressive behaviors that strongly violate the developer policy of Google Play.

As a case study, Figure 7 presents a real example of an aggressive description belonging to the app "SSNG Racer Lite". The description has defined many popular app names as *keywords*, which are however irrelevant to the actual content of the app. It is also interesting to observe that, although these spamming apps have embedded popular app names into their descriptions aiming to rank higher in the search results in order to attract more users, the number of downloads of those apps are not high, and even none of them is popular app. ***This evidence suggests that it is not feasible to attract users via adding spamming message to the app descriptions and it may face the risk of being removed from the market.***

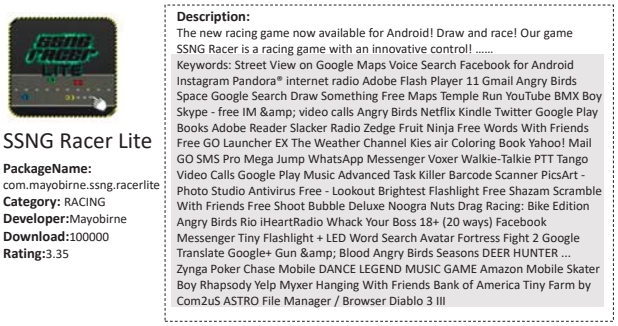


Figure 7: Example of a removed app (SSNG Racer Lite) and its irrelevant description.

Table 11: Top 10 Spamming apps that have listed the most number of irrelevant popular app names in their descriptions.

Package Name	# Downloads	# Pop Names
com.mayobirne.ssng.racerlite	100,000	80
com.mayobirne.ssng.racer	100	80
com.creaple.digdig	10,000	50
com.free.aertsd.game2014	5000	44
akb.studios.guessthelanguages	500	43
com.vvodtopmaket	10,000	34
com.ruleapterol.vvo	10,000	33
bsc.shakeking	1,000	32
com.pumpup	1,000	26
jp.sfproject.adw1	100	26

4.3.5 ***RQ5: Ad-blocking Apps.*** To investigate how many ad blocking apps were removed in the 2015 dataset, we first manually summarize a list of keywords (e.g., ad block, adblock, adblocker, adguard, ads blocker, ads free, etc.) that ad-blocking apps usually used in their app names. Then, we apply these keywords to the 1.5 million apps and find that 35 apps contain at least one of these keywords in their app names. Note that not all of them are ad-blocking apps, because we found some apps usually embed words like "ad free" in their names to indicate no ads were contained in them. By manually reviewing and installing those apps, we confirm that 26 of them are ad-blocking apps, among which 20 of them have already been removed by Google Play. Table 12 enumerates the top 10 ad-blocking apps that are removed. It is interesting to see that most of these removed apps are used to block ads in free apps. Unlike those 20 apps, the remaining six apps are focused on browsers (such as ad-blocking browsers), which however are allowed with the policy of Google Play.

Table 12: Top 10 removed ad-blocking apps.

Package Name	# Downloads
yaya.gugu.trial.adblock	100,000
org.tint.adblock	100,000
com.notification.blocker	100,000
com.av1rus.adblockremover	50,000
fr.flavi1.adblockfree	10,000
com.overlay.adblockbrowser	10,000
com.gmail.calvinloveland.igab	10,000
com.holtoleegames.adblockshooter	5,000
sujeewa.ad3	1,000
sujeewa.ad1	1,000

4.3.6 ***RQ6: COPPA-violated Apps.*** Liu *et al.* [61] have proposed a machine learning classifier to predict whether an app is designed for children based on various features (e.g., app category, content rating, app description, color distribution and usage of the icon and screenshots, etc.). Our study is built upon their work: we also train a similar machine learning classifier and apply it to the 791,138 removed apps. For the 791,138 removed apps, the classifier has identified a total number of 28,319 apps targeting kids.

Privacy Policy. We first analyze how many of the 28,319 apps have declared privacy policies, as COPPA requires app developers to offer a clear "Privacy Notice" of what private information will be collected and for what purpose [10]. ***For the 28,319 apps that targeting children, 23,700 of them (around 83.7%) have no privacy policies declared.***

App Description. COPPA requires app store promotion pages provide individual developers' data collection and sharing practices. It was reported that the FTC staffs have manually reviewed the app descriptions to examine whether apps have provided information about the apps' data practices [3, 4, 7]. Previous studies [66, 68] have been proposed to use natural language processing (NLP) techniques to infer the app's expected behaviors from app descriptions by comparing with the actual behavior extracted from the requested permissions. We therefore leverage WHYPer [66] in this work to check whether these apps disclosure the usage of sensitive information in app description on their promotion pages. WHYPer focuses on three sensitive permissions: "READ_CONTACTS", "READ_CALENDAR" and "READ_AUDIO". Because it takes a relatively long time to analyze all the 28,319 removed kids' apps, we choose only the top 200 apps for each permission (in total 600 apps). ***Surprisingly, only 19% of apps that use "READ_CONTACTS" permission have mentioned it in the app descriptions, the percentages of mentioning "READ_CALENDAR" and "READ_AUDIO" permissions are also quite low, being 27% and 35%, respectively.***

Third-party Services. The FTC staffs have also manually examined the app promotion pages to identify features that may be used for data collection [3, 4, 7], e.g., the ability to make purchases within the app, connect with social media, and serve targeted advertising. These features are often provided by various third party libraries, who may gain access to children's sensitive data as a result. In our study, we first use LibRadar [16, 17, 62], an open source and obfuscate-resilient tool to identify apps that use third-party libraries. we only focus on such libraries that could access to sensitive information by invoking the permission-related APIs [34, 38]. After that, we use heuristics to check whether these apps mentioned the usage of third-party services in their descriptions. We use two kinds of heuristics: (1) we search for the library name (e.g., Admob) in the description; (2) we search for the types or the functionalities of the libraries (e.g., advertisement, advertising, etc). Table 13 shows the result of our heuristic search. It is obvious that a ***large portion of apps use third-party libraries and share the sensitive information with them. However, very few of them have explicitly disclosed these features in their app descriptions.***

Table 13: Disclosure the features of third-party services in app descriptions.

Category	Advertisement	Social Networking	In-app Purchase
% of Apps	53%	19.6%	22.5%
% of Description	5.76%	11.5%	7.05%

5 DISCUSSIONS

We discuss the exploratory implication that our community could observe based on this study and potential limitations of this paper.

5.1 Removal Prediction

In this work, we conduct an analysis of already removed apps aiming at understanding why they are removed by the maintainers of Google Play, which has revealed various findings in different aspects. We believe that these findings can be leveraged to form a symptom-based predictor or even a machine learning-based predictor for predicting the to-be-removed apps before they are really removed. This implication can on one hand keep problematic apps from entering Google Play in the first place, while on the other hand be leveraged by the maintainers of Google Play to provide a channel for app developers to fix highlighted problems.

5.2 Developer Policies of Alternative Markets

We attempt to examine the developer policies of 10 popular third-party app markets in the Chinse market (because Google Play is unavailable), including Baidu Market, Tencent Myapp, 360 Market, Huawei Market, Xiaomi Market, Wandoujia, Anzhi Market, AppChina, HiApk and OPPO Market. Surprisingly, only two app markets (Tencent Myapp, Huawei and Wandoujia) offer explicit developer behavioral policies. Thus it is unclear to us how do these app markets regulate the behavior of app developers, and whether they detect and remove spamming apps or not. As previous work [80, 89] suggested that malware and repackaged apps were found in many third-party markets, the best practices learnt from Google Play could help the third-party market maintainers identify and remove low-quality, malicious, spamming or annoying apps.

5.3 Towards a Better Mobile App Ecosystem

Despite much efforts have been put forward by app analysts for identifying problematic Android apps in the literature, it is still unknown how these approaches may impact the final decision of Google Play, i.e., whether a given app needs to be removed from the market. One reason behind this situation could be the fact that there is no technical support at the moment for facilitating the adoption of existing techniques for removing problematic apps.

We believe that it is not only the responsibility of market maintainers towards removing problematic apps from their markets, but also the responsibility of app developers as well as app analysts. Therefore, all the involved parties, including market maintainers, app developers, and app analysts need to work together so as to elegantly and authentically resolve the problem of removing problematic apps from markets and build a better mobile app ecosystem.

5.4 Threats to Validity

First, the reliability of our empirical results depends on the dataset we have collected. Since Google Play apps could be removed/unpublished by their developers for personal reasons, although we attempt to mitigate this problem by excluding such removed apps that have replacements (i.e., same developer, same app name) on Google Play, our dataset of removed apps may still contain irrelevant apps that may bias our investigation into the reasons why apps are removed by the maintainers of Google Play. Nevertheless, it is non-trivial to fully exclude all the apps that are removed by developers themselves. At the same time, we believe there is no obvious reason why a developer might want to remove their own apps if not for updating or re-releasing it as a new app.

Second, for the sake of simplicity, some of our empirical investigations are conducted with straightforward methods, where the results may not be fully reliable. In our future work, we plan to supplement these investigations with more comprehensive approaches.

Third, these 791,138 apps may be removed by Google Play at any time during the 2.5 years (i.e., the interval between our two datasets). The metadata (e.g., app version, description, downloads, ratings, privacy policy, etc.) and the apks may change during that time. Thus, the removed apps studied in this paper may not be fully representative to the situation when they were removed.

6 RELATED WORK

To the best of our knowledge, our work is the first one that performs a large-scale empirical study of removed Google Play apps. Nevertheless, several research studies, in one way or another, have stepped into this direction. We now discuss the representative ones.

Most notably, many researchers have focused on the security and privacy issues of Android apps in the literature [37, 54, 56, 89]. For example, Zhou et al. [89] present a systematic study for identifying malicious apps on Android markets aim at improving the health of the markets by removing malicious apps. In their empirical experiments, 32 malware are revealed from the Google Play store, which nevertheless is much better than the malware revealed from alternative markets. Through static taint analysis, Li et al. [54] also find similar trends for Google Play apps in terms of privacy leak, where they have identified 337 apps out of 15,000 randomly selected ones that leak private data outside of the device.

The second popular reason causing the removal of Google Play apps could be app spam. Seneviratne et al. [70, 71] propose a method to manually label 1,500 removed apps and found 35% of them are likely to be spam apps, which may provide unrelated app descriptions, not provide a specific functionality (i.e., fake apps), or publish similar apps several times and across diverse categories. Dong et al. [46] have experimentally confirmed that some Android apps do violate the behavioral policy of ad libraries. Indeed, the last point, also known as clone apps or repackaged apps, have been thoroughly investigated by our community [42, 49, 58, 80]. For example, Wukong is proposed by Wang et al. [80] to detect clone apps in the Android ecosystem. It first uses LibRadar [62, 79] to filter out library code and subsequently to select equivalent apps by comparing their fingerprints. Except for spam, clone apps are likely to be injected with malicious payloads [57], which further present security threats to end-users.

For COPPA, Liu et al. [61] have contributed their first step towards providing privacy analysis on mobile apps for children. They have presented a machine learning model for predicting children-focused apps. On step further, Reyes et al. [69] propose to automatically evaluate apps' COPPA compliance based on dynamic execution, network traffic analysis and human-analyst feedback.

Mobile app ecosystem analysis has been widely explored [40, 41, 48, 55, 67, 74, 78, 83, 84]. For example, PlayDrone [78] performed a large-scale characterization of Android apps based on 1.1 million Android apps crawled from Google Play in 2014 and they explored issues such as app evolution, library usage and authentication scheme in Android apps. Bogdan et al. [41] have analyzed 160,000 Google play apps daily for a period of six month aiming to summarize the temporal patterns. Wang et al. [83] have analyzed the Google Play app ecosystem based on over 1.2 million Android apps and 320,000 developers. Taylor et al. [74] analyzed the Google Play over a two-year period to understand how permission usage by apps has changed. None of these studies focused on the characterization of removed apps, and no previous work revealed the fact that a large portion of apps were removed from Google Play.

Overall, we hope our large-scale empirical study on the reasons of dropping apps from Google Play can shed lights on state-of-the-art Android research by helping them realize more advanced approaches to detect to-be-removed apps and simultaneously to keep problematic apps from entering app markets in the first place.

7 CONCLUSIONS

In this work, we present a large-scale empirical study of removed Google Play apps. By crawling and comparing two snapshots of Google Play apps (one at 2015 and another at 2017), we have eventually identified a set of 791,138 removed apps, which indicates that more than half of the apps in 2015 have been disappeared in 2017. We first tried to characterize the set of removed apps based on the categories and developers, then identify potential reasons for app removal and thoroughly explored these removed apps based on 6 research questions. Our experimental results have revealed various interesting findings. The insights we observed in this paper could benefit both app markets and app developers.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No.61702045 and No.61772042), the National Key Research and Development Program of China (No.2017YFB0801903), the Frontier and Key Technology Innovation Project of Guangdong Province Science and Technology Department (No.2016B010110002), the BUPT Youth Research and Innovation Program (No.2017RC40), and Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia under Grant No.ITSM200601.

REFERENCES

[1] 2012. Android Malware Genome Project. http://www.malgenomeproject.org/. (2012). Accessed: 2018-01-15.

[2] 2012. Google 'Bouncer' Now Scanning Android Market for Malware. https://www.pcmag.com/article2/0,2817,2399778,00.asp. (2012). Accessed: 2018-01-15.

[3] 2012. Mobile Apps for Kids: current privacy disclosures are disappointing. https://www.ftc.gov/sites/default/files/documents/reports/mobile-apps-kids-current-privacy-disclosures-are-disappointing/120216mobile_apps_kids.pdf. (2012). Accessed: 2018-01-15.

[4] 2012. Mobile Apps for Kids: Disclosures Still Not Making the Grade. https://www.ftc.gov/reports/mobile-apps-kids-disclosures-still-not-making-grade. (2012). Accessed: 2018-01-15.

[5] 2013. Adblock Plus for Android removed from Google Play store. https://adblockplus.org/blog/adblock-plus-for-android-removed-from-google-play-store. (2013). Accessed: 2018-01-15.

[6] 2013. Google Has Started Removing Ad Blockers from the Play Store. https://lifehacker.com/5990448/google-has-started-removing-ad-blockers-from-the-play-store. (2013). Accessed: 2018-01-15.

[7] 2015. KidsâĂŹ Apps Disclosures Revisited. https://www.ftc.gov/news-events/blogs/business-blog/2015/09/kids-apps-disclosures-revisited. (2015). Accessed: 2018-01-15.

[8] 2017. AppBrain. https://www.appbrain.com. (2017). Accessed: 2018-01-15.

[9] 2017. Children's Privacy. https://www.ftc.gov/tips-advice/business-center/privacy-and-security/children%27s-privacy. (2017). Accessed: 2018-01-15.

[10] 2017. COPPA - Children's Online Privacy Protection Act. http://coppa.org/. (2017). Accessed: 2018-01-15.

[11] 2017. COPPA Violation. https://play.google.com/intl/en-GB/about/families/. (2017). Accessed: 2018-01-15.

[12] 2017. Developer Policy Center". https://play.google.com/intl/en-GB/about/developer-content-policy/. (2017). Accessed: 2018-01-15.

[13] 2017. Google is using machine learning to sort good apps from bad on the Play Store. https://www.theverge.com/2017/7/12/15958372/google-machine-learning-ai-app-store-malware-security. (2017). Accessed: 2018-01-15.

[14] 2017. Google Play API. https://github.com/egirault/googleplay-api. (2017). Accessed: 2018-01-15.

[15] 2017. Google will soon delete apps with no privacy policies from play store. https://www.techrepublic.com/article/google-will-soon-delete-apps-with-no-privacy-policies-from-play-store/. (2017). Accessed: 2018-01-15.

[16] 2017. LibRadar. https://github.com/pkumza/LibRadar. (2017). Accessed: 2018-01-15.

[17] 2017. LibRadar 2.1.0: A tool for Android library detection. https://pypi.python.org/pypi/LibRadar. (2017). Accessed: 2018-01-15.

[18] 2017. Massive Android Malware Outbreak Invades Google Play Store. http://fortune.com/2017/09/14/google-play-android-malware/. (2017). Accessed: 2018-01-15.

[19] 2017. Multi-stage malware sneaks into Google Play. https://www.welivesecurity.com/2017/11/15/multi-stage-malware-sneaks-google-play/. (2017). Accessed: 2018-01-15.

[20] 2017. New Android Trojan malware discovered in Google Play. https://blog.malwarebytes.com/cybercrime/2017/11/new-trojan-malware-discovered-google-play/. (2017). Accessed: 2018-01-15.

[21] 2017. Over a million Android users fooled by fake WhatsApp app in official Google Play Store. https://www.theregister.co.uk/2017/11/03/fake_whatsapp_app/. (2017). Accessed: 2018-01-15.

[22] 2017. Privacy Grade result of app "air.com.elextech.happyfarm". http://privacygrade.org/apps/air.com.elextech.happyfarm.html. (2017). Accessed: 2018-01-15.

[23] 2017. Privacy Grade result of app "com.fingersoft.cartooncamera". http://privacygrade.org/apps/com.fingersoft.cartooncamera.html. (2017). Accessed: 2018-01-15.

[24] 2017. Privacy Grade result of app "com.fotolr.photoshake". http://privacygrade.org/apps/com.fotolr.photoshake.html. (2017). Accessed: 2018-01-15.

[25] 2017. Privacy Grade result of app "com.galapagossoft.trialdemo". http://privacygrade.org/apps/com.galapagossoft.trialdemo.html. (2017). Accessed: 2018-01-15.

[26] 2017. Privacy Grade result of app "com.myxer.android". http://privacygrade.org/apps/com.myxer.android.html. (2017). Accessed: 2018-01-15.

[27] 2017. Privacy Grade result of app "com.outfit7.tomsmessengerfree". http://privacygrade.org/apps/com.outfit7.tomsmessengerfree.html. (2017). Accessed: 2018-01-15.

[28] 2017. Privacy Grade result of app "com.ScnStudios.PoliceCarDriver3D". http://privacygrade.org/apps/com.ScnStudios.PoliceCarDriver3D.html. (2017). Accessed: 2018-01-15.

[29] 2017. Privacy Grade result of app "com.sds.android.ttpod". http://privacygrade.org/apps/com.sds.android.ttpod.html. (2017). Accessed: 2018-01-15.

[30] 2017. Privacy Grade result of app "com.unityconceptapps.tcr.kiat". http://privacygrade.org/apps/com.unityconceptapps.tcr.kiat.html. (2017). Accessed: 2018-01-15.

[31] 2017. Privacy Grade result of app "com.zdworks.android.toolbox". http://privacygrade.org/apps/com.zdworks.android.toolbox.html. (2017). Accessed: 2018-01-15.

[32] 2017. Privacy Rating for Android Apps. https://github.com/CMUChimpsLab/privacyRating. (2017). Accessed: 2018-01-15.

[33] 2017. PrivacyGrade. http://privacygrade.org/. (2017). Accessed: 2018-01-15.

[34] 2017. PScout: Analyzing the Android Permission Specification. http://pscout.csl.toronto.edu/. (2017). Accessed: 2018-01-15.

[35] 2018. VirusTotal. https://www.virustotal.com/. (2018). Accessed: 2018-01-15.

[36] Daniel Arp, Michael Spreitzenbarth, Malte Hubner, Hugo Gascon, Konrad Rieck, and CERT Siemens. 2014. DREBIN: Effective and Explainable Detection of Android Malware in Your Pocket.. In *Ndss*, Vol. 14. 23–26.

[37] Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Yves Le Traon, Damien Octeau, and Patrick McDaniel. 2014. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. *Acm Sigplan Notices* 49, 6 (2014), 259–269.

[38] Kathy Wain Yee Au, Yi Fan Zhou, Zhen Huang, and David Lie. 2012. Pscout: analyzing the android permission specification. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 217–228.

[39] Bogdan Beleapril. 2013. 60,000 apps were removed from the Play Store in February alone. https://www.androidauthority.com/play-store-apps-removed-187094/. (2013). Accessed: 2018-01-15.

[40] Paolo Calciati and Alessandra Gorla. 2017. How do apps evolve in their permission requests?: a preliminary study. In *Proceedings of the 14th International Conference on Mining Software Repositories*. IEEE Press, 37–41.

[41] Bogdan Carbunar and Rahul Potharaju. 2015. A longitudinal study of the Google app market. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 242–249.

[42] Kai Chen, Peng Liu, and Yingjun Zhang. 2014. Achieving accuracy and scalability simultaneously in detecting application clones on android markets. In *Proceedings of the 36th International Conference on Software Engineering*. ACM, 175–186.

[43] Cohen Coberly. 2017. Hundreds of Android apps might be removed from the Google Play Store for misusing accessibility features. https://www.techspot.com/news/71852-hundreds-android-apps-might-removed-google-play-store.html. (2017). Accessed: 2018-01-15.

[44] Shane Curtis. 2017. Google removes 300 Android apps following DDoS attack. https://www.welivesecurity.com/2017/08/31/google-removes-malicious-apps/. (2017). Accessed: 2018-01-15.

[45] Patrick Devaney. 2018. 60 infected apps removed from Google Play Store for displaying pornographic ads to children. https://en.softonic.com/articles/60-infected-apps-removed-from-google-play-store-for-displaying- pornographic-ads-to-children. (2018).

[46] Feng Dong, Haoyu Wang, Li Li, Yao Guo, Guoai Xu, and Shaodong Zhang. 2018. How do Mobile Apps Violate the Behavioral Policy of Advertisement Libraries?. In *Proceedings of the 19th International Workshop on Mobile Computing Systems and Applications (HotMobile)*.

[47] Justin Duino. 2017. Google set to remove apps from Play Store that use Accessibility Services not designed for users with disabilities. https://9to5google.com/2017/11/12/apps-android-accessibility-services-removed-google-play-store/. (2017). Accessed: 2018-01-15.

[48] Jun Gao, Li Li, Pingfan Kong, Tegawendé F Bissyandé, and Jacques Klein. 2018. On Vulnerability Evolution in Android Apps. In *The 40th International Conference on Software Engineering, Poster Track (ICSE 2018)*.

[49] Leonid Glanz, Sven Amann, Michael Eichberg, Michael Reif, Ben Hermann, Johannes Lerch, and Mira Mezini. 2017. CodeMatch: obfuscation won't conceal your repackaged app. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*. ACM, 638–648.

[50] Whitson Gordon. 2013. Google Has Started Removing Ad Blockers from the Play Store. https://lifehacker.com/5990448/google-has-started-removing-ad-blockers-from-the-play-store. (2013). Accessed: 2018-01-15.

[51] IANS. 2017. Google has removed over 500 apps from its Play Store, read why. https://www.gadgetsnow.com/tech-news/google-has-removed-over-500-apps-from-its-play-store-read-why/articleshow/60191102.cms. (2017). Accessed: 2018-01-15.

[52] Sean Michael Kerner. 2017. Google Removes Three Fake Bitcoin Wallet Apps From Google Play. http://www.eweek.com/security/google-removes-three-fake-bitcoin-wallet-apps-from-google-play. (2017). Accessed: 2018-01-15.

[53] Swati Khandelwal. 2017. Over 500 Android Apps On Google Play Store Found Spying On 100 Million Users. https://thehackernews.com/2017/08/android-spyware-malware.html. (2017). Accessed: 2018-01-15.

[54] Li Li, Alexandre Bartel, Tegawendé F Bissyandé, Jacques Klein, Yves Le Traon, Steven Arzt, Siegfried Rasthofer, Eric Bodden, Damien Octeau, and Patrick Mcdaniel. 2015. IccTA: Detecting Inter-Component Privacy Leaks in Android Apps. In *Proceedings of the 37th International Conference on Software Engineering (ICSE 2015)*.

[55] Li Li, Tegawendé F Bissyandé, Jacques Klein, and Yves Le Traon. 2016. An Investigation into the Use of Common Libraries in Android Apps. In *The 23rd IEEE International Conference on Software Analysis, Evolution, and Reengineering (SANER 2016)*.

[56] Li Li, Tegawendé F Bissyandé, Mike Papadakis, Siegfried Rasthofer, Alexandre Bartel, Damien Octeau, Jacques Klein, and Yves Le Traon. 2017. Static Analysis of Android Apps: A Systematic Literature Review. *Information and Software Technology* (2017).

[57] Li Li, Daoyuan Li, Tegawendé F Bissyandé, Jacques Klein, Haipeng Cai, David Lo, and Yves Le Traon. 2017. On Locating Malicious Code in Piggybacked Android Apps. *Journal of Computer Science and Technology* (2017).

[58] Li Li, Daoyuan Li, Tegawendé F Bissyandé, Jacques Klein, Yves Le Traon, David Lo, and Lorenzo Cavallaro. 2017. Understanding Android App Piggybacking: A Systematic Study of Malicious Code Grafting. *IEEE Transactions on Information Forensics & Security (TIFS)* (2017).

[59] Jialiu Lin, Shahriyar Amini, Jason I. Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. [n. d.]. Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy Through Crowdsourcing. In *Proceedings of UbiComp '12*. 501–510.

[60] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I. Hong. [n. d.]. Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In *Proceedings of SOUPS '14*. 199–212.

[61] Minxing Liu, Haoyu Wang, Yao Guo, and Jason Hong. 2016. Identifying and analyzing the privacy of apps for kids. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications*. ACM, 105–110.

[62] Ziang Ma, Haoyu Wang, Yao Guo, and Xiangqun Chen. 2016. Libradar: Fast and accurate detection of third-party libraries in android apps. In *Proceedings of the 38th International Conference on Software Engineering Companion*. ACM, 653–656.

[63] Francisco Memoria. 2017. Google Kicks Malicious Bitcoin Exchange Apps from Play Store. https://www.ccn.com/two-malicious-poloniex-exchange-apps-removed-from-google-play-store/. (2017). Accessed: 2018-01-15.

[64] Alfred Ng. 2017. Google purges malicious Android apps with millions of downloads. https://www.cnet.com/news/google-removes-android-malware-downloaded-up-to-5-9m-times/. (2017). Accessed: 2018-01-15.

[65] Thuy Ong. 2017. Google removes 300 Android apps that secretly hijacked phones for DDoS attacks. https://www.theverge.com/2017/8/29/16219426/google-removes-apps-play-store-hijack-phones-ddos-attacks. (2017). Accessed: 2018-01-15.

[66] Rahul Pandita, Xusheng Xiao, Wei Yang, William Enck, and Tao Xie. 2013. WHYPER: Towards Automating Risk Assessment of Mobile Applications.. In *USENIX Security Symposium*, Vol. 2013.

[67] Thanasis Petsas, Antonis Papadogiannakis, Michalis Polychronakis, Evangelos P Markatos, and Thomas Karagiannis. 2017. Measurement, modeling, and analysis of the mobile app ecosystem. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS)* 2, 2 (2017), 7.

[68] Zhengyang Qu, Vaibhav Rastogi, Xinyi Zhang, Yan Chen, Tiantian Zhu, and Zhong Chen. 2014. Autocog: Measuring the description-to-permission fidelity in android applications. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1354–1365.

[69] Irwin Reyes, Primal Wiesekera, Abbas Razaghpanah, Joel Reardon, Narseo Vallina-Rodriguez, Serge Egelman, and Christian Kreibich. 2017. " Is Our Children's Apps Learning?" Automatically Detecting COPPA Violations. (2017).

[70] Suranga Seneviratne, Aruna Seneviratne, Mohamed Ali Kaafar, Anirban Mahanti, and Prasant Mohapatra. 2015. Early detection of spam mobile apps. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 949–959.

[71] Suranga Seneviratne, Aruna Seneviratne, Mohamed Ali Kaafar, Anirban Mahanti, and Prasant Mohapatra. 2017. Spam mobile apps: Characteristics, detection, and in the wild analysis. *ACM Transactions on the Web (TWEB)* 11, 1 (2017), 4.

[72] Jagmeet Singh. 2017. UC Browser Removed From Google Play Store, Here's the Reason. https://gadgets.ndtv.com/apps/news/uc-browser-disappears-from-google-play-store-ucweb-1775697. (2017). Accessed: 2018-01-15.

[73] Tom Spring. 2017. Apps Infected With SonicSpy Spyware Removed From Google Play. https://threatpost.com/apps-infected-with-sonicspy-spyware-removed-from-google-play/127406/. (2017). Accessed: 2018-01-15.

[74] Vincent F Taylor and Ivan Martinovic. 2017. To update or not to update: Insights from a two-year study of Android app evolution. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*. ACM, 45–57.

[75] time. 2017. Google Gets Rid of Infected Android Apps After Millions Download Malware. http://time.com/4941968/google-infected-android-apps-malware/. (2017). Accessed: 2018-01-15.

[76] Liam Tung. 2017. Chat apps laced with spyware removed from Google Play. https://www.cso.com.au/article/625996/chat-apps-laced-spyware-removed-from-google-play/. (2017). Accessed: 2018-01-15.

[77] Jonathan Vanian. 2017. Google Kicks 500 Apps Off Online Store Over Spyware Concerns. http://fortune.com/2017/08/22/google-spyware-play-online/. (2017). Accessed: 2018-01-15.

[78] Nicolas Viennot, Edward Garcia, and Jason Nieh. 2014. A measurement study of google play. In *ACM SIGMETRICS Performance Evaluation Review*, Vol. 42. ACM, 221–233.

[79] Haoyu Wang and Yao Guo. 2017. Understanding Third-party Libraries in Mobile App Analysis. In *Proceedings of the 39th International Conference on Software Engineering Companion (ICSE-C '17)*. 515–516.

[80] Haoyu Wang, Yao Guo, Ziang Ma, and Xiangqun Chen. 2015. Wukong: A scalable and accurate two-phase approach to android app clone detection. In *Proceedings of the 2015 International Symposium on Software Testing and Analysis*. ACM, 71–82.

[81] Haoyu Wang, Jason I. Hong, and Yao Guo. 2015. Using Text Mining to Infer the Purpose of Permission Use in Mobile Apps. In *The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2015)*. 1107–1118.

[82] Haoyu Wang, Yuanchun Li, Yao Guo, Yuvraj Agarwal, and Jason I. Hong. 2017. Understanding the Purpose of Permission Use in Mobile Apps. *ACM Trans. Inf. Syst.* 35, 4, Article 43 (July 2017), 40 pages.

[83] Haoyu Wang, Zhe Liu, Yao Guo, Xiangqun Chen, Miao Zhang, Guoai Xu, and Jason Hong. 2017. An Explorative Study of the Mobile App Ecosystem from App Developers' Perspective. In *Proceedings of the 26th International Conference on World Wide Web*. 163–172.

[84] Tingting Wang, Di Wu, Jiaming Zhang, Min Chen, and Yipeng Zhou. 2016. Measuring and Analyzing Third-Party Mobile Game App Stores in China. *IEEE Transactions on Network and Service Management* 13, 4 (2016), 793–805.

[85] Fengguo Wei, Yuping Li, Sankardas Roy, Xinming Ou, and Wu Zhou. 2017. Deep Ground Truth Analysis of Current Android Malware. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. 252–276.

[86] Minda Zetlin. 2017. 300 Google Play Store Apps Are Discovered to Contain Malware. https://www.inc.com/minda-zetlin/300-google-play-store-apps-are-discovered-to-conta.html. (2017). Accessed: 2018-01-15.

[87] Min Zheng, Patrick PC Lee, and John CS Lui. 2012. ADAM: an automatic and extensible platform to stress test android anti-virus systems. In *International conference on detection of intrusions and malware, and vulnerability assessment*. 82–101.

[88] Yajin Zhou and Xuxian Jiang. 2012. Dissecting android malware: Characterization and evolution. In *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 95–109.

[89] Yajin Zhou, Zhi Wang, Wu Zhou, and Xuxian Jiang. 2012. Hey, you, get off of my market: detecting malicious apps in official and alternative android markets.. In *NDSS*, Vol. 25. 50–52.

[90] Zeljka Zorz. 2018. 36 fake security apps removed from Google Play. https://www.helpnetsecurity.com/2018/01/04/fake-security-apps-google-play/. (2018). Accessed: 2018-01-15.