

Beyond Google Play: A Large-Scale Comparative Study of Chinese Android App Markets

Haoyu Wang¹, Zhe Liu², Jingyue Liang², Narseo Vallina-Rodriguez^{3,4}
Yao Guo^{2*}, Li Li⁵, Juan Tapiador⁶, Jingcun Cao⁷, Guoai Xu¹

¹ Beijing University of Posts and Telecommunications,

² Key Laboratory of High-Confidence Software Technologies (MOE), Peking University, ³ IMDEA Networks, ⁴ ICSI,

⁵ Monash University, ⁶ Universidad Carlos III de Madrid, ⁷ Indiana University Bloomington

ABSTRACT

China is one of the largest Android markets in the world. As Chinese users cannot access Google Play to buy and install Android apps, a number of independent app stores have emerged and compete in the Chinese app market. Some of the Chinese app stores are pre-installed vendor-specific app markets (e.g., Huawei, Xiaomi and OPPO), whereas others are maintained by large tech companies (e.g., Baidu, Qihoo 360 and Tencent). The nature of these app stores and the content available through them vary greatly, including their trustworthiness and security guarantees.

As of today, the research community has not studied the Chinese Android ecosystem in depth. To fill this gap, we present the first large-scale comparative study that covers more than 6 million Android apps downloaded from 16 Chinese app markets and Google Play. We focus our study on catalog similarity across app stores, their features, publishing dynamics, and the prevalence of various forms of misbehavior (including the presence of fake, cloned and malicious apps). Our findings also suggest heterogeneous developer behavior across app stores, in terms of code maintenance, use of third-party services, and so forth. Overall, Chinese app markets perform substantially worse when taking active measures to protect mobile users and legit developers from deceptive and abusive actors, showing a significantly higher prevalence of malware, fake, and cloned apps than Google Play.

CCS CONCEPTS

• Information systems → Web mining; • Security and privacy → Mobile and wireless security; • Networks → Mobile networks;

KEYWORDS

App ecosystem, Android market, malware, cloned app, third-party library, permission, Google Play

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IMC '18, October 31–November 2, 2018, Boston, MA, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5619-0/18/10...\$15.00

<https://doi.org/10.1145/3278532.3278558>

ACM Reference Format:

Haoyu Wang¹, Zhe Liu², Jingyue Liang², Narseo Vallina-Rodriguez^{3,4}, Yao Guo^{2*}, Li Li⁵, Juan Tapiador⁶, Jingcun Cao⁷, Guoai Xu¹. 2018. Beyond Google Play: A Large-Scale Comparative Study of Chinese Android App Markets. In *2018 Internet Measurement Conference (IMC '18)*, October 31–November 2, 2018, Boston, MA, USA. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3278532.3278558>

1 INTRODUCTION

According to recent reports, there are more than 700 million Android users in China [84]. Due to the restriction of Google's services in China since late 2010 – and by extension of Google Play [46, 73] – hundreds of millions of Chinese Android users resort to alternative markets to buy and install Android apps. This restriction over Google services in China has been seen as a business opportunity by many Chinese Internet companies (e.g., Tencent and Baidu) and smartphone manufacturers (e.g., Huawei and Xiaomi). Despite the fact that these app markets target mainly the Chinese Android users, they are also available to users from all over the world.

The diversity and large number of third-party markets in China have made it difficult for both mobile users and app developers to choose the most appropriate one(s) to discover or distribute their apps. This state of affairs has also opened new opportunities for malicious actors: previous work has suggested that repackaged apps, including malware, are widely distributed in Google Play, but especially through third-party markets [52, 53, 57, 58, 60, 89, 102].

To the best of our knowledge, no previous work has performed a systematic and comparative study across different app markets, including the Chinese ones. To fill this research gap, we perform a multi-dimensional and large-scale study covering more than 6.2 million apps to identify the differences between Google Play and 16 popular Chinese Android app markets. We begin our study by offering a high-level characterization of these app stores, discussing features such as their copyright checks, app auditing processes, their strategies to attract app developers, and their transparency efforts (Section 2). Second, after presenting our dataset and app collection method (Section 3), we compare their download distributions, user rating distributions, and presence of third-party tracking and advertising libraries (Section 4). Third, we study their catalog similarities and their publication dynamics, with emphasis on detecting the distribution of the presence of a given developer and app across stores (Section 5). We then provide an in-depth analysis of malicious and deceptive behaviors across app markets, discussing the presence of fake and cloned apps, over-privileged apps, and malware (Section 6). We conclude our paper with a discussion around the state of affairs in the Chinese Android ecosystem, and its implications to users and developers alike (Section 7).

Our main research contributions are as follows:

- We conduct a comparative study of various intra- and inter-market features. Our results reveal a long tail distribution of app popularity, with the top 1% of apps usually accounting for over 80% of total downloads across the 17 studied markets. Further, we observe the presence of heterogeneous behaviors across markets (e.g., in terms of code maintenance and metadata consistency).
- We find that the set of third-party libraries (e.g., SDKs provided by advertising and tracking services) embedded in Android apps are different for those published in Chinese stores than in Google Play. This observation could be explained by the inability to access Google services such as Google Analytics and AdMob from China, and Chinese developers' need to monetize their apps through services specialized in the Chinese market.
- Popular apps are more likely to be simultaneously published in multiple markets compared to unpopular ones. However, there is a strong market bias across developers: 1) 57% of Google Play developers do not publish their apps on any of the Chinese markets, and 2) almost half of the Chinese-specific developers do not publish apps in Google Play.
- We analyze the prevalence of various types of malicious behaviors in our dataset, specifically the presence of fake and cloned apps, over-privileged apps, and malware. Google Play clearly outperforms Chinese markets in all dimensions of our study thanks to their positive efforts to eradicate these behaviors. Our results reveal that the presence of malicious and repackaged apps in the majority of Chinese app stores is significant and prevalent over time (10%, on average, in the case of malware), in some cases reaching almost 1 in 4 apps in the market.
- In order to estimate the extent to which app markets implement security checks on submissions, we performed a second crawl 8 months after the first snapshot. Our exploration suggests that over 84% of the potentially malicious apps found in Google Play were removed. This differs considerably in the case of Chinese markets, with malware removal ratios varying from 0.01% to 34.51% in the best case.

To the best of our knowledge, this is the first comparative study between Google Play and alternative Chinese app stores at scale, longitudinally and across various dimensions. Our results motivate the need for more research efforts to illuminate the widely unexplored Chinese mobile and web ecosystem. We believe that our efforts can positively contribute to bring user and developer awareness, attract the focus of the research community and regulators, and promote best operational practices across app store operators. We have released our dataset, along with the experiment results, to the research community at: <http://market.orangeapk.com/>

2 CHINESE ANDROID APP MARKETS

Due to the access restrictions of Google Play in China, Chinese Android users resort to a large ecosystem of alternative third-party Android app markets, which could be classified into three categories according to their nature:

- **Vendor-specific app markets.** China has a vast and powerful smartphone manufacturing industry with well-known vendors such as Huawei, Xiaomi, and Lenovo. Almost every Chinese

smartphone vendor maintains its own app market, which also comes pre-installed on their devices.

- **Web companies.** Chinese Internet giants such as 360, Baidu, and Tencent also compete in the Chinese Android market with their own app stores. These companies usually provide support to some smartphone vendors behind the scenes. For example, the Sony app store in China is powered by Baidu App Market, and the Smartisan app store is supported by Tencent Myapp Market.
- **Specialized markets.** A number of relatively small Chinese companies are specialized in Android app services. They usually make profit through app promotion and other business-oriented partnerships with app developers/companies. For example, 25PP is an Android app market powered by the PP smartphone assistant, which is a popular management system app in China. Similarly, Wandoujia is an app store provided by a company focused on app recommendation, especially for mobile games.

In this study, we first resort to several independent industry reports about app market ranking in China [3–5, 14, 29, 30]. We cover all the top 10 Android markets in China, excluding the Vivo market (ranks 6 to 10 in China), because the Vivo market does not provide a web-based app download interface, which makes it difficult for us to crawl the apps. Our list covers the app stores for the top five smartphone vendors in China [27], three top Chinese web companies, and eight popular specialized Android app markets. The app markets in this list cover more than 98% of active users in China [3, 29]. The majority of these markets target Chinese Android users, but some of them operate at a global scale, particularly those run by Android handset vendors. For example, Huawei's app market is also popular in Europe, Latin America, and the Middle East [20].

2.1 Features of Chinese App Markets

In this section, we study some critical aspects and features across app stores, including their openness to developers, their publication and app auditing process, and their transparency, as shown in Table 1. For that, we first registered a developer account for each market and then manually examined their *developer policies*, *terms of service* and other documents released by these markets [6, 8, 12, 13, 15–17, 21–25, 28, 33].

- (1) **Openness:** Most Chinese app markets allow third-party developers to publish their apps for free². However, a small number of app stores enforce some limitations. For instance, Lenovo's MM market only allows registered companies to release apps [22], whereas OPPO market only allows publishing apps falling in specific categories, such as "wallpaper" and "theme" apps [25]. Vendor markets such as OPPO and Xiaomi force developers to release apps that are fully compatible with their own devices [25, 33]. Finally, App China explicitly limits an APK size to 50 MB [13].
- (2) **Copyright checks:** In order to limit the publication of fake and cloned apps, all the Chinese markets but HiApk and PC Online perform copyright ownership checks. Developers should submit a "Software Copyright Certificate" indicating that they are the original authors of the released apps.
- (3) **Publishing incentives:** Chinese app stores provide a number of incentive mechanisms for encouraging app developers to

²Google Play's registration fee costs \$25 [19].

Table 1: Dataset size and market features for Google Play and the 16 Chinese markets studied in this paper.

Market	Type	Size (#Apps)	Aggregated Downloads	#Developers	% Unique Developers	Openness	Copyright Check	App Vetting	Security Check	Vetting Time	Quality Rating	Incentive#1	Incentive#2	Incentive#3	Privacy Policy Advertisement	In-app Purchase
Google Play	Official	2,031,946	193 B	538,283	57.04	✓	✓	✓	✓	Hours	✓	✓	✓	✓	✓	✓
Tencent Myapp	Web Co.	636,265	82 B	294,950	10.61	✓	✓	✓	✓	1 day	✓	✓	✓	✓	✓	✓
Baidu Market	Web Co.	227,454	94 B	107,698	15.10	✓	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
360 Market	Web Co.	163,121	50 B	90,226	6.80	✓	✓	✓	✓	1 day	✓	✓	✓	✓	✓	✓
OPPO Market	HW Vendor	426,419	57 B	209,197	14.37	Partial ¹	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
Xiaomi Market	HW Vendor	91,190	-	55,669	5.78	✓	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
MeiZu Market	HW Vendor	80,573	19 B	50,451	0.58	✓	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
Huawei Market	HW Vendor	51,303	83 B	32,927	5.66	✓	✓	✓	✓	3-5 days	✓	✓	✓	✓	✓	✓
Lenovo MM	HW Vendor	37,716	24 B	24,565	0.79	✓	✓	✓	✓	2 days	✓	✓	✓	✓	✓	✓
25PP	Specialized	1,013,208	56 B	470,073	19.06	✓	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
Wandoujia	Specialized	554,138	38 B	291,114	0.97	✓	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
HiApk	Specialized	246,023	17 B	115,191	3.65	✓	N/A	N/A	N/A	N/A	✓	✓	✓	✓	✓	✓
AnZhi	Specialized	223,043	12 B	74,145	21.93	✓	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
LIQU	Specialized	179,147	26 B	101,336	6.10	✓	✓	✓	✓	N/A	✓	✓	✓	✓	✓	✓
PC Online	Specialized	134,863	0.2 B	65,225	2.58	✓	N/A	N/A	N/A	N/A	✓	✓	✓	✓	✓	✓
Sougou	Specialized	128,403	3 B	66,759	4.04	✓	✓	✓	✓	1 day	✓	✓	✓	✓	✓	✓
App China	Specialized	42,435	-	23,699	3.22	✓	✓	✓	✓	1-3 days	✓	✓	✓	✓	✓	✓
Total		6,267,247	754 B	1,035,992												

publish their apps. These models could be classified into three categories. The first one is “**The Starting App and Exclusive App Free Promotion**”, a common mechanism across markets which gives stores publication exclusivity for a period of time in exchange for actively taking measures to promote the app, typically during 24 hours. The second category is “**High Quality App Free Promotion**”. Some markets have a qualification of high-quality apps. Apps that meet the criteria to obtain such a qualification could request the markets to promote them for free. The third category is “**Editors’ Choice**”, in which the store recommends apps based on personal opinions.

- (4) **Auditing process:** All app markets but HiApk and PC Online indicate that apps are published after an inspection and vetting process. Moreover, eight markets (Google Play, Tencent, OPPO, Xiaomi, Meizu, Huawei, Anzhi and AppChina) claim to incorporate *human inspections* attempting to complement the automated auditing process. The general approach is to use automated security analysis tools first to identify possible threats³, and then manually check the most suspicious submissions. For example, a majority of the top apps in the Huawei market are labeled with a sign indicating that they went through manual inspection before being made publicly available, and it is reported that Huawei has a large human inspection team [1]. Excluding HiApk and PC Online, Chinese alternative app stores also explicitly check and report security issues on the apps (e.g., malware and aggressive adware). The inspection time varies across markets, from several hours (Google Play) to roughly 5 days (Huawei market). 360 market requires all the developers

to use their packaging tool 360 Jiagubao [34] to obfuscate apps before entering the market.

- (5) **App quality ratings:** Only Tencent Myapp market and 360 market explicitly report that they rate the quality of published apps based on downloads, user comments, developer level and other metrics. For high quality apps, they could provide more market resources (e.g., advertise them on the starting page) for app promotion to attract high quality developers.
- (6) **Transparency:** As opposed to Google Play, none of the Chinese app markets require developers to publish their *privacy policies* whenever they obtain and use sensitive user data. However, nine markets (Google Play, Tencent Myapp, Baidu, 360, OPPO, Huawei, 25PP, Sougou and AppChina) explicitly inform users whether the apps contain *advertisements*. Only Google Play and 360 market report the presence of *in-app purchases* in the apps.

3 APK COLLECTION

We implemented a crawler to harvest APKs from Google Play and the 16 alternative Chinese Android app stores listed in Table 1 in August 2017. For each app, we also collect publicly available metadata as provided by the app stores, including, among others, the app name, version name, app category, description, downloads, ratings and release/update date.

We follow different strategies to crawl each market. In the case of Google Play, we use a list of 1.5 million package names provided by PrivacyGrade [78] as the *searching seeds*. We use a breadth-first-search (BFS) approach to crawl (1) additional related apps recommended for each one of our seeds by Google Play, and (2) other apps released by the same developer. In order to avoid potential regional bias, we instrumented our crawler to support both English and Chinese languages. Chinese app markets index apps in different

³Some web companies have released their own security analysis tools for Android apps, e.g., 360 Mobile Security [7] and Baidu PhoneGuard [68, 69].

ways. Consequently, we adapt our crawler to the indexing behavior of each Chinese app market. For instance, as of this writing, Baidu’s app market indexes apps incrementally⁴.

We launched several crawlers in parallel via 50 Aliyun Cloud Servers [37] between August 15 and August 30, 2017. However, published Android apps can be updated by the developers at any time, potentially affecting our analysis. To overcome this challenge, we implement a “parallel search” strategy in our crawler. As long as we identify a new app (based on its package name) in one of the 17 markets, we immediately search this app (using either the app name or its package name, according to different markets) in all the remaining markets to crawl it simultaneously if found. Note that we will crawl all the listed searching results and add them to the searching seeds. After roughly 8 months, we launched a second, one-week crawling campaign in April 30, 2018 for analyzing whether any of the studied malicious apps has been removed from each individual stores (Section 7).

3.1 Dataset

Table 1 reports the number of harvested APKs per store. We crawled metadata for 6,267,247 different apps across all app stores, and 4,522,411 APK files. To the best of our knowledge, our dataset is the largest cross-store APK collection obtained by the research community. The mismatch between app metadata and APKs is due to Google Play’s rate limiting mechanism, which limited our APK collection efforts to a random sample of 287,110 of them. We resorted to AndroZoo [11, 67] to obtain offline the APK files for 1,553,382 of the missing Google Play apps, using the package name and version name as primary key. Note that, although this dataset does not cover all the available apps in these markets due to the limitation of our BFS app crawling method⁵, we believe that our dataset has covered the most popular apps in both Google Play and the Chinese markets. Further, due to the parallel search strategy, the apps studied across markets will not bias the results.

4 GENERAL OVERVIEW

We now study high-level characteristics of Google Play and the 16 Chinese app stores. Using Google Play as a reference, we briefly discuss differences along various dimensions such as catalog diversity, user downloads, Android API support, third-party libraries and user ratings.

4.1 App Categories

App stores give app developers the freedom to publish their apps in specific app categories. However, each Android market implements a different taxonomy of apps. While Google Play defines 33 app categories (excluding game app subcategories), Huawei Market only has 18 categories. In order to perform a fair comparison across markets, we manually develop a consolidated taxonomy containing 22 app categories, as shown in Figure 1. Due to the lack of enforcement and lax supervision over the metadata provided by app developers,

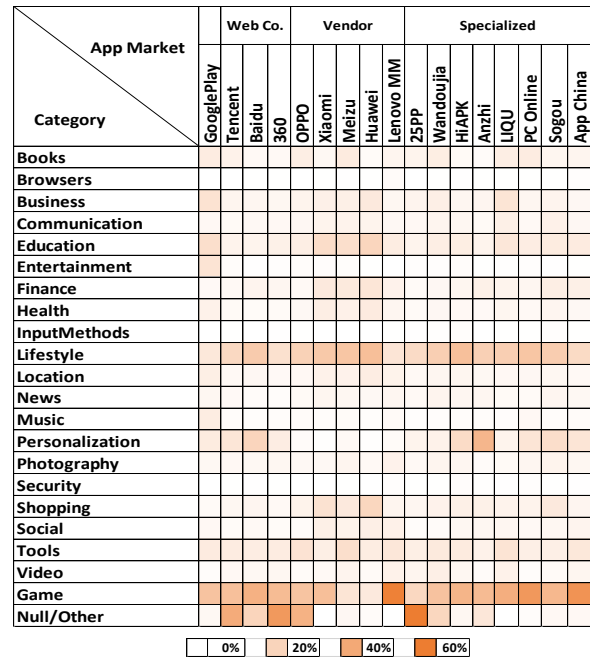


Figure 1: Distribution of app categories.

in Tencent, 360, OPPO, and 25PP markets, we classify 40% of the apps from these stores as “Other” category⁶.

It is noticeable that Games account for roughly 50% of all apps across markets, while other popular categories include Lifestyle and Personalization. The least popular categories are Browsers, InputMethods and Security tools. Note, also, how the distribution of published apps across categories for the majority of Chinese app stores follows closely Google Play’s distribution. A number of app stores, especially vendor ones such as Meizu, Huawei and Lenovo’s, present a different distribution of categories.

4.2 User Downloads

The majority of app stores report the actual number of user installs per app while Google Play bins them into installation ranges (e.g., “50,000 - 100,000”). However, this metadata may not be consistent across stores. Xiaomi and AppChina do not report this information at all. Further, we suspect that some of the app stores might be reporting the number of user downloads, likely higher than the number of user installs, instead of user installs. For comparison purposes and minimize bias, we normalize the number of user installed apps for each app store (excluding Xiaomi and AppChina) to Google Play’s ranges⁷.

As Table 1 reports, the apps in Google Play have 193 Billion aggregated downloads⁸. No Chinese app store gets closer individually to this volume despite the size of the Chinese market in terms of user-base. However, the number of aggregated downloads across all the 16 studied markets is three times higher than that of Google

⁴ We use the following syntax: <http://shouji.baidu.com/software/INTEGER.html>.

⁵ As of this writing, the number of apps in Google Play is 2,893,556 [39], while we only crawled roughly 70% of them.

⁶ Apps published in these markets can report NULL or non-descriptive categories (e.g., “Unclassified”, “102229”).

⁷ e.g., 75,123 after normalization becomes [50,000, 100,000].

⁸ Estimated by considering the lower bound limit of Google Play’s install range.

	0-10	10-100	100-1K	1K-10K	10K-100K	100K-1M	>1M
Google Play	4.05%	17.90%	30.52%	25.38%	15.15%	5.62%	1.21%
Tencent Myapp	55.87%	12.37%	15.50%	10.38%	4.21%	1.21%	0.35%
Baidu Market	0.00%	34.98%	25.91%	23.21%	7.65%	5.40%	2.26%
360 Market	16.54%	16.08%	19.25%	25.79%	12.78%	7.24%	1.97%
Huawei Market	0.10%	0.00%	38.05%	27.33%	17.64%	11.73%	4.16%
Xiaomi Market	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Wandoujia	1.96%	4.74%	43.66%	35.24%	12.17%	1.77%	0.38%
HiApk	0.00%	0.00%	78.24%	13.15%	5.93%	2.05%	0.53%
Anzhi Market	0.10%	1.35%	49.72%	42.83%	4.86%	0.84%	0.23%
OPPO Market	0.00%	0.00%	84.31%	10.47%	3.16%	1.55%	0.43%
25PP	0.27%	4.63%	68.02%	20.34%	4.82%	1.49%	0.37%
Sougou	0.77%	17.83%	55.13%	22.27%	2.51%	1.15%	0.31%
MeiZu Market	7.63%	13.50%	45.37%	19.54%	7.97%	4.28%	1.42%
LIQU	0.01%	0.03%	0.01%	71.83%	22.32%	5.14%	0.61%
App China	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Lenovo MM	0.04%	14.70%	0.00%	53.54%	16.78%	11.02%	3.19%
PC Online	13.07%	74.19%	8.62%	2.98%	0.91%	0.21%	0.02%

Figure 2: Distribution of downloads across markets.

Play. This figure shows the importance of Chinese Android markets when aggregated.

The distribution of app downloads follows a power-law distribution, regardless of the app market, as shown in Figure 2. In general, 85% of the analyzed apps have less than 10K installs. However, subtle differences arise when looking in detail on a per-store basis after ranking the apps by their number of installs. On average, the top 0.1% of the apps account for more than 50% of the total downloads, regardless of the app store. However, the top 0.1% of apps published in Tencent MyApp account for more than 80% of the total downloads while more than 55% of its published apps have almost no downloads (< 10). On the other hand, 15% of the apps published in vendor app stores like Huawei’s and Lenovo’s have more than 100K installs. This suggests that there are significant differences in the popularity and quality of the apps published in certain app stores, as we will investigate later.

4.3 Minimum API Level

Android app developers can declare in the app manifest the minimum Android API level supported by their apps. This information could offer insights about whether app developers are trying to maximize app customers, or whether they try to target top-end users. Figure 3 shows the distribution of minimum API level declared for each app in each market. The result suggests that API levels 7-9 (i.e., Android versions 2.1.x to 2.3.2) are the most widely supported minimum API levels by the analyzed apps. However, the percentage of apps in alternative Chinese markets supporting low API levels is 3x higher than that of Google Play in general: roughly 63% of apps in Chinese third-party markets support API levels lower than 9, as opposed to 22% in the case of Google Play.

We further analyzed the release or update time of these apps across markets. This is also a metric used for estimating whether developers actively maintain their apps, a strong signal for code quality [59, 66]. Figure 4 shows the distribution of the release/update time of the apps in our dataset, as reported by the markets. As we can see, roughly 90% of apps in Chinese alternative markets were released/updated before 2017, while the number in Google

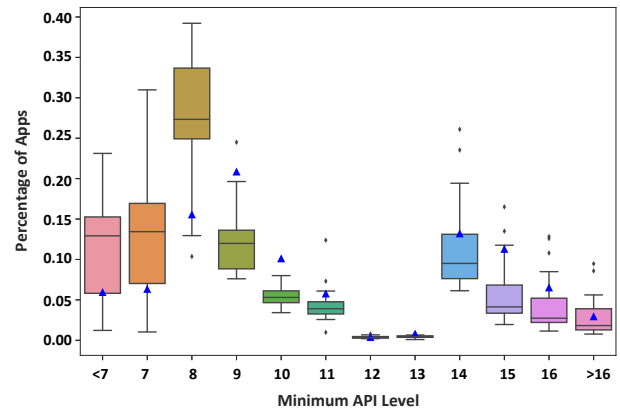


Figure 3: Distribution of minimum API level declared for each app for the analyzed markets. The triangle symbol represents the value for Google Play, while the box-plots represent the values across the 16 Chinese alternative stores.

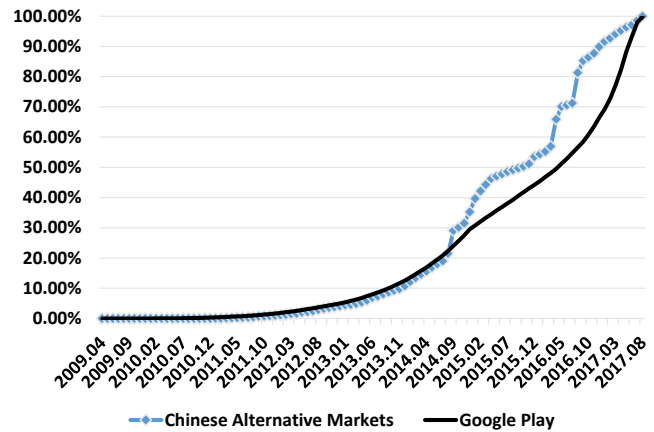


Figure 4: Distribution of app release/update dates.

Play is 66%. Further, only 5% of apps published in Chinese stores were updated/released within 6 months before launching our crawling campaign, while more than 23% of Google Play apps were released during the same time frame. This finding suggests that most of the apps published in Chinese markets likely support low API-level as they were released years ago. These apps do not get timely updates, hence likely exposing their user-base to various security risks [59], and do not take advantage of features introduced in newest Android versions.

4.4 Third-party Libraries

Third-party services form an integral part of the mobile ecosystem: they ease app development and enable features such as analytics, social network integration, and app monetization through advertisements [79, 86, 88]. However, aided by the general opacity of mobile systems, such services are also largely invisible to users,

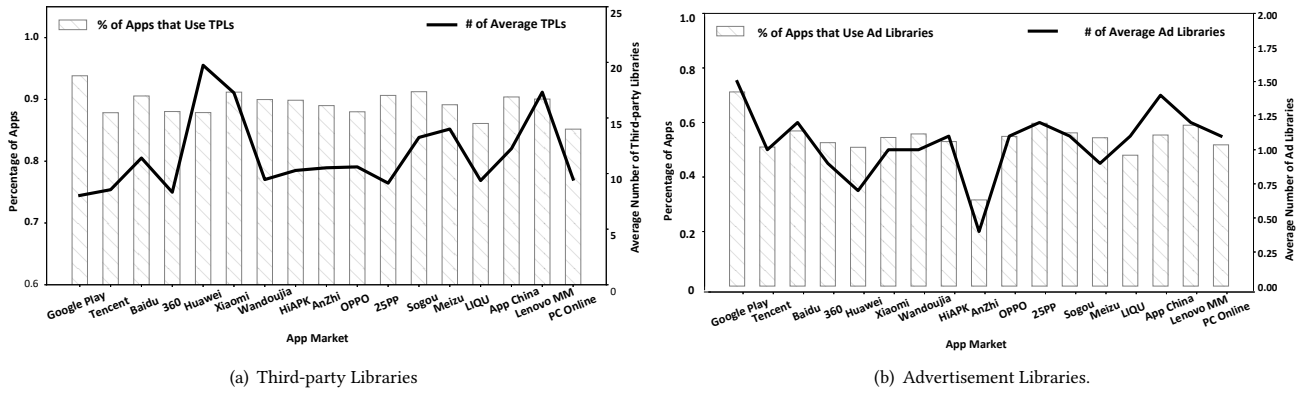


Figure 5: Presence of third-party libraries across app stores.

Table 2: Top 10 third-party libraries for Google Play and Chinese markets apps. Chinese market specific libraries are highlighted in the table.

Google Play		
Package Name	Type	Usage (%)
com.google.android.gms	Development	66.1
com.google.ads	Advertisement	62.1
com.facebook	Social Networking	21.5
org.apache	Development	20.5
com.squareup	Payment	13.8
com.google.gson	Development	12.9
com.android.vending	Payment	12.5
com.unity3d	Game Engine	11.8
org.fmod	Game Engine	9.6
com.google.firebase	Development	9.0
Chinese Markets		
com.google.ads	Advertisement	25.7
org.apache	Development	24.1
com.google.android.gms	Development	20.5
com.tencent.mm	Social Networking	17.3
com.baidu	Development, Map	16.9
com.umeng	Analytics, Advertisement	16.5
com.google.gson	Development	16.3
com.alipay	Payment	11.0
com.facebook	Social Networking	10.7
com.nostra13	Development	10.6

hence causing potential privacy risks [43, 61, 72, 79, 81, 82, 88]. This is aggravated by the lack of transparency enforcement across alternative app stores (Section 2): no alternative Chinese store requires developers to publish a privacy policy, and only a handful of them actively report the presence of ad services or in-app purchases in published apps.

Although existing studies have created several tools or datasets for third-party library detection [43, 65, 74, 80], they are either too old or incomplete to fulfill our research purpose in this paper. For example, LibRadar [70, 74] is a widely used and obfuscation-resilient tool to identify third-party libraries used in Android apps. However, it was created in 2016, and it relies on a feature dataset of libraries extracted from Google Play apps. Considering that our apps are

crawled in August 2017, and most of our apps are from Chinese markets, it may fail to report new libraries (or new versions) as well as libraries specific to the Chinese market.

To this end, we have applied the clustering-based approach introduced in LibRadar [74] to the 6 million apps we collected in this paper, and build a new and complete feature dataset of third-party libraries covering apps in both Google Play and the Chinese markets. At a result, we have created a dataset containing 5,102 libraries with 672,052 different versions. We then manually examined the top 2,000 libraries and labeled them into different categories⁹. In order to identify the company behind each one of them as well as the purpose of the library, we search the unobfuscated package name in Google, and refer to several sources, including AppBrain library classification [39], PrivacyGrade classification [78], and Common Library classification [65]. We group them in 5 different categories by their purpose or offered service: ad network, analytics, social networking, development tools, and payment.

As shown in Figure 5(a), the presence of third-party services varies from app store to app store, yet it remains high: Google Play has the highest presence of embedded third-party libraries in their published apps (roughly 94% of published apps have a third-party library) whereas PC Online presents the lowest penetration (85% of published apps). Differences also appear in terms of the total number of libraries per app when inspecting specific stores, specially in vendor-provided app stores. While the average app embeds more than 10 third-party libraries, those published in 360 market have 20 third-party libraries embedded on average. This number contrasts significantly with the 8 libraries found in average for Google Play apps.

Most Popular Third-party Libraries. Table 2 lists the top 10 third-party libraries found for apps published in Google Play and all Chinese markets, respectively. Google-related libraries used for advertisement and analytics services dominate in Google Play: they can be found in more than 60% of published apps. It is interesting to see that, although Google services are blocked in China, Google-related libraries can be also found in Chinese markets, with more than a quarter of apps in Chinese markets embedding Google-related advertising libraries [79, 86]. We further explored these

⁹Note that for libraries with multiple versions, we only need to label one of them.

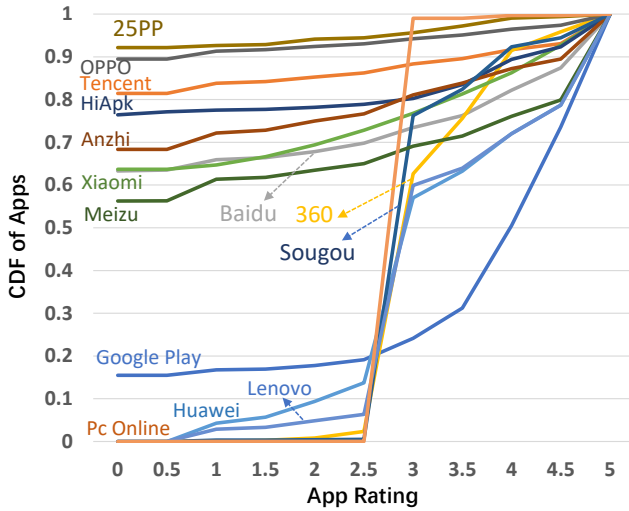


Figure 6: Distribution of app ratings across markets.

apps and identified two leading reasons for this. The first reason is that most of these apps do not release Chinese-specific versions. This implies that the subset of applications relying on Google Services found on alternative Chinese app stores are identical to those present in Google Play. The second reason is that some markets crawl apps in Google Play to enlarge their application catalog: more than 30,000 apps published in Baidu market are explicitly labeled as crawled from Google Play in the developer name field. Nevertheless, we found many instances of third-party libraries specific to the Chinese market across app stores. For instance, instead of Facebook’s GRAPH API [18], more than 17.3% of the apps published in Chinese markets embed Tencent Wechat library [32], a popular Chinese social networking SDK. Alipay (a payment SDK) and Baidu (a library for development also offering map support) are also used by more than 10% of the apps published in Chinese markets, hence replacing Google vending and Google Maps, respectively.

Advertising libraries. Identifying advertising libraries is a non-trivial task, as suggested by previous studies [53, 71, 75]. We leverage AppBrain and Common Library classification [65] to identify and classify third-party ad libraries. We have manually labeled 282 advertisement-specific libraries (with 56,011 versions) in total. As shown in Figure 5(b), around 70% of the apps published in Google Play use any of the labeled ad libraries, while 53.2% of apps in Chinese markets use at least one ad library. It is worth mentioning that Google AdMob dominates Google Play with roughly 90% of the advertisement market share, while the Chinese mobile ad ecosystem is more decentralized. Google AdMob and Umeng are the two most popular ad libraries, accounting for 80% of the mobile ad market share in China, while more than 200 ad libraries compete for the remaining 20% of the market.

4.5 App Ratings

We conclude our app store comparison with a brief analysis of how users rate published apps. The rating scores are crawled from app markets. Note that if an app does not receive any rating score, we set it as 0 by default. Figure 6 shows the CDF of app ratings for all

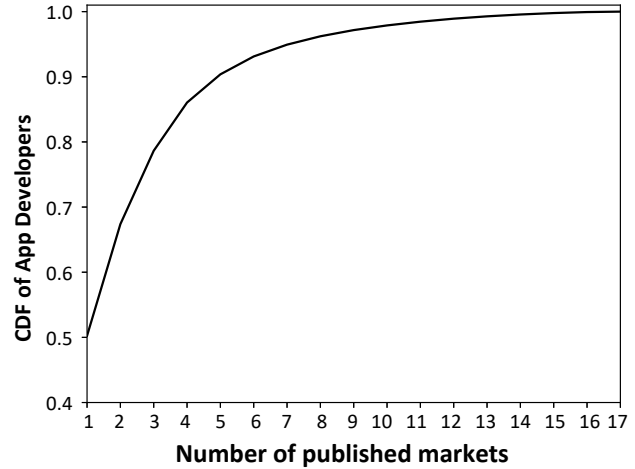


Figure 7: CDF of developer published markets.

the considered markets. The distribution shows that app ratings vary greatly across Chinese markets but it is possible to identify two clear patterns:

- **Pattern #1:** More than 80% of apps in Chinese third-party app markets do not receive user ratings at all, around 90% of these apps have less than 1,000 downloads. This pattern can be found in 25PP, OPPO and Tencent Myapp markets. This trend indicates that most of the apps published in these markets are low-quality and unpopular Android apps – a trend in-line with the app download distribution shown in Figure 2.
- **Pattern #2:** Finally, we notice that the distribution for several markets (e.g., PC Online in the bottom) contains many apps with ratings between 2.5 and 3 out of 5. We tried to upload some testing apps to PC Online and found that they use a default rating of 3, instead of a default rating 0, which could be the reason leading to this distribution.

Google Play, instead, presents a pattern completely different to that of any Chinese app market: only 9.3% of Google Play apps have not been rated by users, while more than 50% of them have received ratings higher than 4.

5 PUBLISHING DYNAMICS

In this section, we investigate the publishing dynamics of app developers. We focus on analyzing the publishing distribution for each developer and app across each store¹⁰. We define “**single-store**” released apps as those available only in a single market of our dataset; otherwise, we label them as “**multi-store**” apps. Note that it is possible that the “single-store” app would appear in other markets that are not covered in this paper. This, however, does not affect our comparative study.

5.1 App Developers

Android mobile apps must be signed with a developer key before being released. We used the tool ApkSigner [9] to extract the app developer signature present on each APK. This metadata, embedded

¹⁰We identify unique apps across markets based on their package names (or app ID).

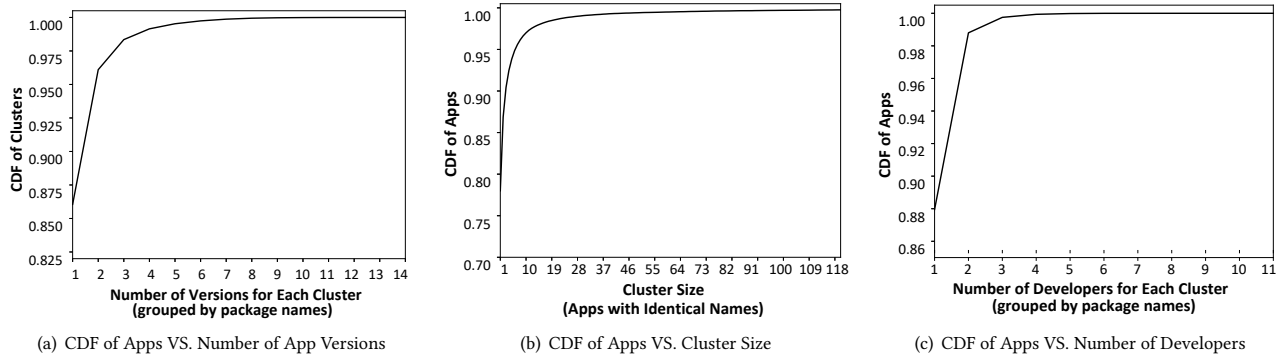


Figure 8: CDF of apps vs. (a) number of different versions (b) cluster size, and (c) number of developers.

on each executable, cannot be spoofed or modified by malicious actors¹¹. We identified slightly over 1 million app developers – all of them with different signatures – in our dataset, as summarized in Table 1.

Our analysis reveals that app developers follow different publishing strategies by targeting app stores and users in different ways. More than half of the developers release their apps in Google Play, and around 48% of them focus solely on Chinese alternative markets. Out of these developers found on Google Play, 57% of them do not release their apps in Chinese markets, possibly due to language barriers or a lack of understanding of the fragmented Chinese ecosystem.

Figure 7 shows the CDF of the number of app markets targeted by each app developer. Around 20% of the app developers publish their apps in more than 3 app stores simultaneously, but only a few of them (just 696) roll out their apps in the 17 markets simultaneously. It is interesting to note that over 10% of the developers target exclusively one single Chinese store. This trend is more prevalent for those markets with a larger app catalog (e.g., Tencent and 25PP), which also offer incentives to app developers for the exclusive publishing rights of their software.

5.2 Single- and Multi-store Apps

Single-store Apps. More than 77% of the apps published in Google Play are single-store ones. This result is expected, as Google Play has a global presence and its catalog has far more apps than any other market individually. On average, 11% of the apps published in alternative Chinese app stores are single-store, though this figure varies across stores. For example, while AnZhi, OPPO and 25PP have over 20% of single-store apps, both Wandoujia and Meizu markets have less than 1% of single-released apps. A manual inspection of the apps exclusively published in Meizu reveals that they are popular apps explicitly developed for Meizu-branded handsets (e.g., `com.meizu.flyme.wallet` and `com.meizu.media.reader`).

Multi-store Apps. Between 20% and 30% of the apps published in Chinese alternative markets are also present in Google Play. The analysis also indicates that many Chinese markets share a significant fraction of their app catalogs: for instance, 80% of the apps published in 25PP are also released in Huawei, Wandoujia,

Meizu and Lenovo markets. This trend is also present among the top 1% most popular apps (by downloads) for each market: over 80% of the top 1% most popular apps are shared across all Chinese markets. Catalog similarities between top apps in Chinese stores and Google Play are, instead, low. This finding confirms that many developers target exclusively Chinese app stores.

5.3 IDE and App Store Introduced Biases

The previous method offers an upper-bound estimation of catalog overlaps between stores. However, an important remaining question is: *are two apps with the same package name and app version identical?* An alternative and stricter method to identify whether two apps are identical is comparing the hash (e.g., MD5) of their APK contents. This method allowed us to find a total number of 546,703 apps in our dataset with identical package names, version code and developer but different MD5. For instance, we have 14 different hashes for the app `com.kugou.android.v8.7.0`. After manually inspecting their DEX files (i.e., main function code), we conclude that those apps are identical: the only difference between them is their `META-INF/kgchannel` file¹². This confirms that relying on the package name, version number and developer signature are sufficient to accurately identify similar apps despite these subtle differences. Finally, we also identify instances of app store-introduced differences resulting from stores forcing app developers to follow certain requirements prior to publication. A notable case is 360 market, which requires developers to obfuscate their apps with 360 Jiagubao before uploading it to the app store [34].

5.4 Outdated Apps

Another reason potentially preventing us from identifying multi-store released apps are unsynchronized roll-outs of new app versions across stores. We now relax the condition to identify two identical apps: we only consider the app package name and developer signature, excluding the app version¹³. As shown in Figure 8(a), roughly 14% of apps have simultaneously published multiple versions in different stores, up to 14 different versions in extreme cases.

¹²The `META-INF/kgchannel` files/directories are created, recognized and interpreted by the Java 2 Platform to configure apps, extensions, class loaders and services. The main purpose of them is to differentiate the source of app users (i.e., from which market the app is installed).

¹³We assume that app version numbers are assigned incrementally regardless of app stores.

¹¹We found that one developer (with the same signature) may correspond to multiple names across markets with slight variations, e.g., Chinese name vs. English name.

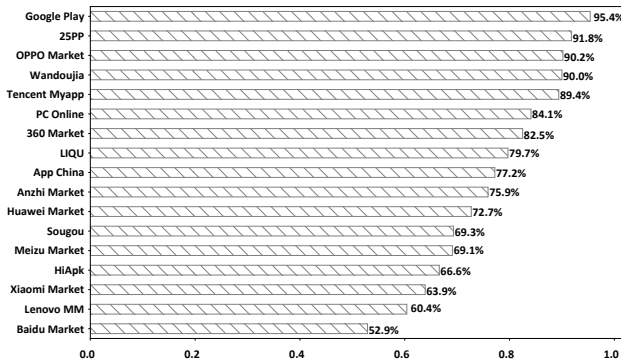


Figure 9: A comparison of app updates across markets.

Because we use a “parallel search” strategy in our crawler (Section 3), the elapsed time between all crawls for a given app across markets is in the order of a few minutes, so those are intentional actions or poor software maintenance practices of the developers. This behavior is not limited to poorly maintained unpopular apps.

Figure 9 details the overall distribution of outdated apps across app stores. Note that for this analysis we exclude single-store apps which are always updated by definition. Besides unfixed bugs and potential vulnerabilities, publicly available outdated apps also hinder users from enjoying newly added features. This can decrease the perceived quality of the apps, and overall hurts the brand equity of the market. This observation suggests that developers may prioritize roll-outs in specific app stores. Google Play has the highest version number across all app stores: 95.4% of the apps published there have the highest app version number. This is not the case for stores like Lenovo MM and Baidu markets, where more than 39% of their apps might be outdated according to their version number.

6 DEVELOPER MISBEHAVIORS

In this section, we study the prevalence of various types of malicious behaviors across markets. Specifically, we study the presence of *fake apps*, *cloned apps*, *over-privileged apps*, and *malware*. The differences between fake and cloned apps are subtle but substantial. Malicious developers can release fake apps that masquerade as the legitimate one but stealthily perform malicious actions on the user’s device. We define those as “fake apps” [103]. We consider “cloned apps” as those that are the results of repackaging legitimate ones [89].

6.1 Fake Apps

We exploit the fact that fake apps usually try to emulate the app name of a legitimate one, but are published with different package names [38, 63]. We applied a clustering-based method to efficiently identify fake apps at scale. First, we build a cluster enforcing a strict matching of app names. As shown in Figure 8(b), around 22% of the apps in our dataset share the same name with at least another app, all of them with different package names, either in the same or in a different store. Not all the identified apps are necessarily fake, as developers may have legitimate reasons for releasing different apps (package names) with the same app name. This is the case of: 1) apps sharing common names like *Flashlight*, *Calculator*, or *Wallpaper*;

Table 3: Fake and cloned apps across stores. SB and CB stand for Signature-Based and Code-Based clones, respectively.

Market	Fake (%)	Clones	
		SB (%)	CB (%)
Google Play	0.03	4.01	17.82
Tencent Myapp	0.53	8.24	22.73
Baidu Market	0.48	10.98	17.38
360 Market	0.50	5.43	23.26
Huawei Market	0.33	11.54	18.76
Xiaomi Market	0.0	8.00	20.11
Wandoujia	0.39	5.98	21.23
HiApk	0.64	7.51	20.08
AnZhi Market	0.57	4.92	20.71
OPPO Market	0.38	5.85	20.94
25PP	0.35	7.16	24.08
Sougou	1.83	4.86	18.28
MeiZu Market	1.14	6.65	18.42
LIQU	0.40	5.32	16.68
App China	0.0	10.17	13.23
Lenovo MM	0.67	7.81	16.37
PC Online	1.89	8.60	23.34
Average	0.60	7.24	19.61

and 2) apps released by the same developer with different package names for different platforms¹⁴.

To this end, we applied a heuristic rule to remove legitimate clusters. Generally, the apps in a cluster include different developer signatures. By manually analyzing 100 randomly selected clusters of different size, we found out that 83% of fake apps form small clusters (size < 5 with uncommon names) of unpopular ones (*i.e.*, downloads ≤ 1,000) and a popular one with more than 1 million installs (the official app). Table 3 summarizes the percentage of fake apps identified in each market using this heuristic. The result suggests that fake apps are present in all app stores, including Google Play. Nevertheless, Meizu, PC Online and Sougou stores have a percentage of fake apps above the average. Note that our heuristic is straightforward yet very effective in identifying apps that use similar names to camouflage as the official apps.

The largest number of fake apps in absolute terms correspond to 25PP and Tencent Myapp, with 3,591 and 3,347 apps, respectively. Relative to the market size, PC Online (with 1.89%) and SouGou (1.83%) lead the ranking of markets with higher presence of fake apps. Overall, our results suggest that many app markets do not take enough efforts to identify and remove fake apps, despite all of them—but PC Online and HiApk—requesting copyright checks and performing app auditing before publication (Section 2). While we did not identify any fake app in Xiaomi and App China, Google Play presents a marginal number of fake apps (572 in total).

6.2 Cloned Apps

Cloned apps often share a large portion of the metadata with the original app, but they are obviously signed by different developers. We explored the prevalence of cloned apps using two separate strategies: a signature-based approach (which aims at identifying apps with the same package name but different developer signatures) and a code-based approach (*i.e.*, apps with high code similarity

¹⁴*e.g.*, com.sogou.map.android.maps and com.sogou.map.android.maps.pad are two different versions of Sogou Map.

but different package names). However, we are also interested in identifying the source market in which the original app has been published. As it is non-trivial to identify the original app given a pair of cloned apps [56, 101], we resort to a heuristic approach to solve this: the app with more downloads is regarded as the original one. Unfortunately, this may generate false positives as it may be possible for the cloned app to have more installs than its original version. Unfortunately, to the best of our knowledge the research community has not developed a more accurate method to solve this problem [51, 89, 101].

Signature-based clones. As in the previous section, we first cluster all the apps by their package name and then compare the app developer signatures for each cluster. We consider that two apps are clones if they share the same package name but do not have a common developer signature. Since package names are supposed to distinctively identify an Android app, it is expected that they should be unique across different Android markets and that they are signed with the same developer key.

Figure 8(c) shows the distribution of apps with respect to the number of developer signatures obtained in a cluster. Overall, 12% of apps have at least 2 clones released by different developers. For example, the app `com.dino.dinosuperapp` has been published in 15 different markets by 11 different developers. To better understand the nature of these clones, we manually examined 100 randomly selected pairs of signature-based clones. In all cases, we observed that clones are actually repackaged apps, *i.e.*, apps created by disassembling the original app, making modifications, and finally reassembling the resulting code into a different app. Even if we cannot cover all cases manually, our analysis suggests that there are no legitimate reasons behind these identified clones.

Code-based clones. Since cloned apps can also modify the package name, we implemented a different approach based on analyzing code similarity to identify cloned apps. Previous work has proposed different approaches for app repackaging detection [51, 52, 58, 99, 101, 102]. In this paper, our implementation is based on WuKong [89], which proved to be an accurate and scalable two-phase approach for app clone detection. We first extracted Android API calls, Intents, and Content Providers for each app and created a feature vector per app with more than 45K dimensions. We then used a variant of the Manhattan distance to measure the similarity between each pair of vectors. Specifically, for n -dimensional feature vectors A and B , their distance is given by

$$distance(A, B) = \frac{\sum_{i=0}^n |A_i - B_i|}{\sum_{i=0}^n (A_i + B_i)}$$

If the resulting distance between the computed vectors for a pair of apps exceeds a certain threshold – we experimentally selected a conservative threshold of 0.05, which corresponds to a 95% similarity – and they are signed with different signatures, we consider these two apps as potential clones. For those apps flagged as potential app clones, we performed a second code-level comparison to refine the results as introduced by WuKong. In this second step, we consider two apps to be clones when they share more than 85% of the code segments. Due to space limitations, we omit the implementation details here.

Previous work has suggested that, on average, more than 60% of an app’s code come from third-party libraries [89]. This figure

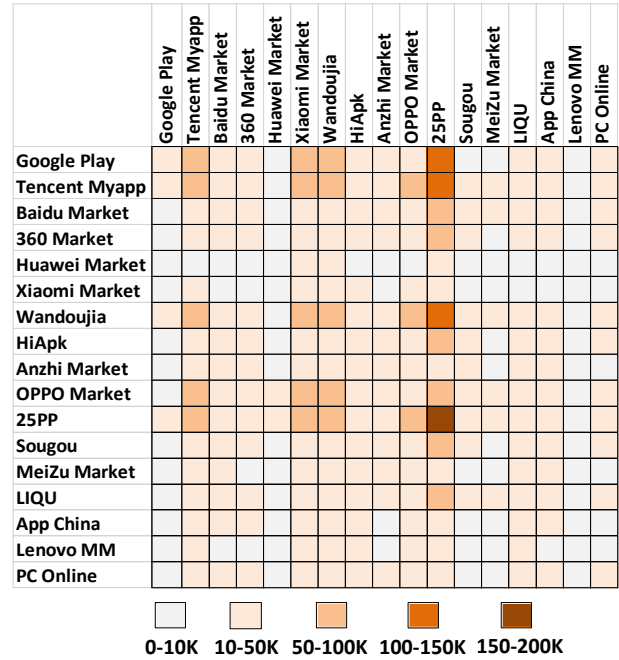


Figure 10: Intra- and inter-market app clones.

is relevant for our analysis since libraries may cause both false positives and false negatives when detecting code clones [65]. To overcome this limitation, we leveraged LibRadar [70, 74] to identify and eliminate the impact of third-party libraries on our code-based app clone study.

Results. Table 3 summarizes the distribution of signature- and code-based clones for each market. Code-based clones (roughly 20%) are generally more common than signature-based clones (roughly 10%). This result is in line with figures reported in previous work [89, 102] and suggests that attackers are more interested in advanced cloning methods that go beyond changing app package names and manipulating the code. We further illustrate the source market of cloned cases in the heatmap rendered in Figure 10. Both intra-market and inter-market clones are considered¹⁵. For each cell (row X , column Y), the color represents the number of cloned apps in market Y that were originally published in market X . Google Play is the premier source for cloning apps: it presents the large number of apps being cloned into Chinese markets. We can also detect interesting trends when looking at the destination of these apps. Market 25PP has the largest number of cloned apps, which are mainly copied from Google Play, followed by Tencent Myapp and Wandoujia. Surprisingly, intra-market clones are also quite common in addition to inter-market clones. As shown in Figure 10, more than 181,677 apps in 25PP market have similar apps to those originally from the same market.

6.3 Over-privileged Apps

Previous studies [41, 54] have analyzed the gap between requested permissions and those actually used by Android apps. An app is said to be “over-privileged” when it requests more permissions

¹⁵Only code-based clones are presented as signature-based clones do not involve any intra-market clones.

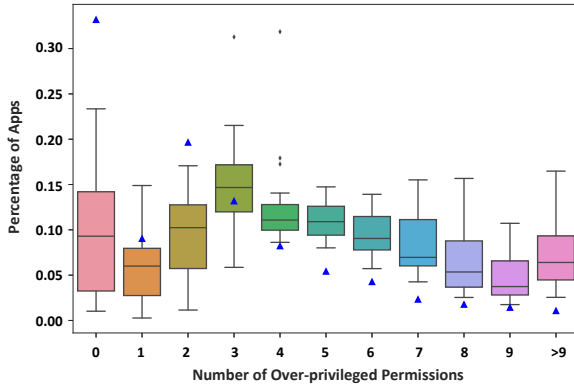


Figure 11: Distribution of over-privileged apps in Google Play and Chinese markets. The triangle symbol represents the value for Google Play, while the box-plots represent the values across the 16 Chinese alternative stores.

(listed in the `AndroidManifest.xml`) than those actually used in their functionalities. Previous work [97] has suggested that more than 85% of Android apps published in vendor-customized phones suffer from this issue. Since permissions constitute an explicit declaration of what sensitive resources an app will use [91, 93], over-privileging an app is undesirable because: (i) it is a violation of the principle of least privilege [2]; (ii) it exposes users to unnecessary permission warnings; and (iii) it increases the attack surface [44] and the impact of the presence of a bug or vulnerability [54].

Intuitively, this gap can be identified first by building a permission map that identifies what sensitive permissions are needed for each API call/Intent/Content Provider, and using static analysis to determine what permission-related invocations an app makes. Then, we can compare the actually used permissions by the app with the requested permissions listed in `AndroidManifest.xml`. To do this, we leveraged data provided by PScout [26, 41], specifically a list of 32,445 permission-related APIs, 97 permission-related Intents, 78 Content Providers URI Strings, and 996 Content Provider URI fields¹⁶.

In general terms, apps published in Chinese markets tend to request more sensitive permissions, *i.e.*, those labeled as dangerous by Google [10], than Google Play apps. Figure 11 shows the distribution of over-privileged apps across markets grouped by how many permissions in excess each app has. Note that, in general, Chinese markets contain more over-privileged apps than Google Play. Approximately 65% of the apps in Google Play are over-privileged, while the percentage in Chinese markets is roughly 82%. In two particular cases (25PP and App China), more than 95% of apps requested at least one unused permission. Apps often request no more than 10 unused permissions, 3 being the most common value. The most common over-privileged sensitive permissions are

¹⁶Note that we use the API-Permission mapping for Android 5.1.1, which may not reflect the new sensitive APIs introduced in subsequent system versions. However, more than 90% of apps in our dataset target API levels less than 5.1. In addition, a well-known limitation of static over-privilege app analysis is its inability to handle Java reflection and dynamic code loading [90].

Table 4: Percentage of apps labeled as malware in each market by AV-rank.

Market	AV-rank (% apps)		
	≥ 1	≥ 10	≥ 20
Google Play	17.03	2.09	0.32
Tencent Myapp	34.15	11.16	3.45
Baidu Market	42.77	12.24	3.30
360 Market	41.40	12.35	3.10
OPPO Market	42.97	16.43	6.00
Xiaomi Market	55.11	9.12	1.82
MeiZu Market	51.40	10.70	3.14
Huawei Market	57.48	4.71	0.57
Lenovo MM	54.20	7.53	1.52
25PP	32.36	8.26	2.06
Wandoujia	31.99	7.98	2.19
HiApk	41.89	11.12	2.72
AnZhi Market	55.32	11.37	2.41
LIQU	45.91	13.00	4.27
PC Online	55.93	24.01	8.37
Sougou	52.41	16.53	4.59
App China	48.55	14.13	4.27
Average	36.49	12.30	3.69

READ_PHONE_STATE (52.38%), ACCESS_COARSE_LOCATION (36.28%), ACCESS_FINE_LOCATION (33.83%), and CAMERA (19.98%).

6.4 Malware Prevalence

In order to investigate the presence of malicious and undesirable apps in our dataset, we uploaded all the apps to VirusTotal [31], an online analysis service that aggregates more than 60 anti-virus engines, which is widely adopted by the research community. Previous studies [40, 96] have suggested that some anti-virus engines may not always report reliable results. In order to deal with such potential false positives, we analyzed the results grouped by how many engines (**AV-rank**) flag a sample as malware. Previous work have argued that a threshold of 10 engines is a robust choice [40, 61, 100].

Overall Result. Table 4 shows the overall detection results. Remarkably, roughly 50% of the apps in Chinese markets are flagged at least by one anti-virus engine, while the percentage for Google Play is considerably lower (17.03%). According to the threshold of “ $AV-rank \geq 10$ ”, around 2% of the apps in Google Play are labeled as malware, while the percentage in Chinese markets is much higher. In fact, for 11 out of the 16 Chinese markets the percentage of malware exceeds 10%. A particularly remarkable case is the PC Online market, with more than 24% of its apps labeled as potentially malicious. In absolute terms, Tencent and 25PP markets host the largest number of malicious apps (70,988 and 83,655, respectively). On the opposite side, we find Huawei’s market, with a figure (4.71%) comparable in magnitude to that of Google Play (2.09%).

Top Malware. Table 5 lists the top 10 malicious apps according to their AV-Rank. Note that two of them (`com.trustport.mobilesecurity_eicar_test_file` and `com.zoner.android.eicar`) correspond to the AV benchmarking apps developed by the European Institute for Computer Antivirus Research (EICAR). The remaining apps—and others that were manually inspected by us—clearly show potentially malicious behaviors. For example, `com.ypt.merchant`, published in 5 markets, poses itself as a legitimate mobile point-of-sale (mPOS) for merchants and individuals.

Table 5: Top 10 malicious apps by their AV-rank.

Package Name (malware family)	AV-Rank	Markets
com.trustport.mobilesecurity_eicar_test_file (eicar)	48	Wandoujia, 25PP
games.hexalab.home (mofin)	47	LIQU
com.wb.gc.ljfk.baidu (ramnit)	47	Baidu, HiAPK
com.ypt.merchant (ramnit)	46	Tencent, Wandoujia, OPPO, 25PP, LIQU
com.wsljtwinmobi (ramnit)	46	Tencent, 25PP
com.wb.gc.ljfk.tx (ramnit)	45	Tencent
com.wgljd (ramnit)	45	Tencent, 360
com.zoner.android.eicar (eicar)	44	Google Play, Wandoujia, 25PP
com.zhiyun.cnhyb.activity (ramnit)	44	Baidu
com.fai.shuiligongcheng (ramnit)	44	25PP

Repackaged Malware. The Android Genome Project [103] suggested that app repackaging is the main way for malware distribution, and 86% of the 1,260 samples are repackaged malware. However, this dataset is outdated (collected in 2011) and the number of samples is relatively small so it may no longer provide a representative picture of the current Android malware landscape. Thus, we further analyzed how many malware samples in our dataset are repackaged apps. To this end, we merged the malware results with the app clone detection results as shown in Section 6.2, and observed that only 38.3% of these malware samples are repackaged apps. This result suggests that app repackaging is no longer the main way for malware spreading. We believe this is an interesting observation for our community, and we leave to future work analyzing the newest trends in malware spreading strategies.

Malware Family. We further analyzed the distribution of malware families across Google Play and Chinese markets. To do this, we used AVClass [83] to obtain the family name (label) of each identified malware. Figure 12 shows the distribution top 20 malware families. An interesting finding is that the distribution of malware families differs greatly between Google Play and Chinese markets. The most popular malware family in Chinese markets is **kuguo** (12.69%), while it only corresponds to 0.6% of malware in Google Play. Roughly, 45% of the malware present in Google Play belong to the family **airpush** (29.04%) and **revmob** (15.09%). We further enlarged our threshold to “ $AV - rank \geq 20$ ” and found that it shows generally similar malware family distribution.

7 POST-ANALYSIS

All markets have strict policies to conduct copyright and security checks (Section 2). Yet our results reveal that they still host a significant amount of fake and cloned apps, as well as malware samples. As introduced in Section 3, we performed a second crawl for each app store about 8 months after the first one in order to quantify whether the stores made any effort to remove those samples from their catalogs¹⁷. As shown in the first column of Table 6, over 84% of the potential malicious apps found in Google Play have been removed. However, the percentages of malware removal in Chinese alternative markets vary from 0.01% (PC Online) to 34.51%

¹⁷We exclude HiAPk from this analysis as it has discontinued its services by the end of 2017. In addition, OPPO can only be accessed now using their market app.

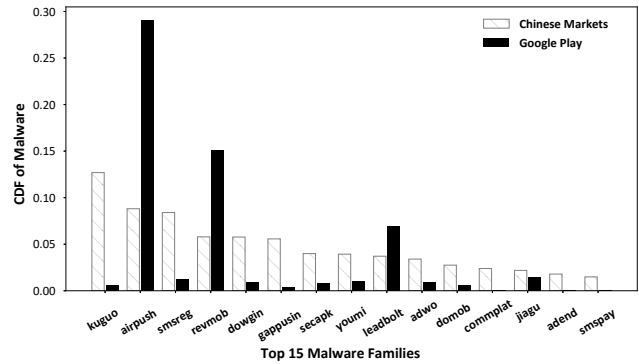


Figure 12: Distribution of top 15 malware families in Google Play and Chinese markets.

Table 6: Percentage of removed malware across markets. The third column indicates the number of apps also published and removed in Google Play (GPRM).

Market	%Malware Removed	#Overlapped with GPRM	%Removed
Google Play	84%	-	-
Tencent Myapp	8.75%	7,157	3.1%
Baidu Market	23.99%	1,422	34.53%
360	43%	1,198	34.22%
Xiaomi	32.50%	636	31.13%
Meizu	29.18%	668	26.20%
Huawei	26.92%	169	23.08%
Lenovo MM	22.75%	263	16.35%
25PP	19.63%	7,804	17.31%
Wandoujia	34.51%	5,289	44.74%
AnZhi	27.61%	632	25.78%
LIQU	14.08%	1,878	11.18%
PC Online	0.01%	1,117	0.00%
Sougou	24.24%	1,082	22.00%
App China	20.51%	546	30.24%

(Wandoujia). We extracted and inspected in detail those apps with an AV-rank ≥ 10 removed from Google Play (GPRM) between our crawls. 11,623 of them were also found in at least one Chinese app store, and over 70% of them are still hosted by at least one Chinese market by the end of April 2018, as shown in the fourth column of Table 6. Tencent and PC Online are clearly the Chinese stores in which those potentially malicious apps still survive.

8 DISCUSSION

Our results reveal that potentially malicious and deceptive activities are more common in Chinese markets than in Google Play. Figure 13 presents a multi-dimensional comparison of four representative Chinese app markets and Google Play. Tencent Myapp, one of the largest Chinese app stores by their aggregate number of downloads, hosts a significant amount of mobile malware. This store seems to be more indulgent with malicious developers, including those publishing malware as well as fake and cloned apps. Although Tencent Store claims to perform manual inspection for all the submitted apps, our empirical observations seem to contradict it. We find a similar behavior for PC Online. However, in this case

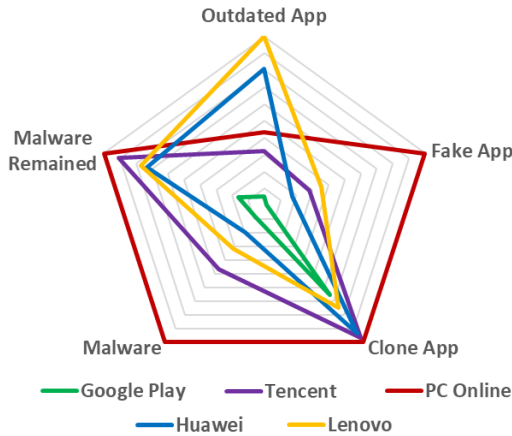


Figure 13: Multi-dimensional comparison of Google Play, Tencent, PC Online, Huawei and Lenovo markets. For each metric, we normalize it to the scale [0, 100], and the center represents 0.

we could not find any developer policy describing security checks on apps prior to publication.

Huawei and Lenovo markets show a clearly different behavior. These stores publish popular apps and present similar app ratings and download distributions. They also seem to have strict mechanisms to prevent malware distribution: only 4.71% and 7.53% of their apps, respectively were labeled as malware, figures comparable in magnitude to that of Google Play. The purpose of the stores and their market segment can also influence in the presence of malware, possibly due to their need to protect their brand reputation. Lenovo’s MM market does not allow individual developers to publish apps, a practice that could help them mitigate the spread of malware and low-quality apps. However, Huawei and Lenovo markets still have a significant number of outdated apps, which could hinder users from enjoying newly added features and other app improvements (e.g., bug fixing). This practice could contribute to decrease the perceived quality of the apps, hurting as a result the brand equity of the app store.

9 RELATED WORK

Previous research efforts have performed large-scale mobile app analysis [35, 45, 47, 50, 55, 85]. However, alternative Android markets have not been well studied by the research community yet. To the best of our knowledge, our work is the largest and most exhaustive comparative analysis made between the official Google Play store and Chinese alternative markets.

Large-scale App Repositories. AndroZoo [67] is an academic effort focused on compiling a large-scale dataset of APKs. This research effort has enabled a number of studies focusing on malicious practices and privacy risks of Android apps [42, 48, 64, 98]. AndroZoo uses purpose-built crawlers to harvest more than 5M APKs from 12 app stores and 5 Chinese markets with roughly 1.5 Million apps. The work of Ishii *et al.* [62] is the closest to ours. They investigated 4.7M Android apps covering 27 app markets, mainly obtained from AndroZoo [11], to understand the security management of global third-party markets.

Measurement of Google Play. Many research efforts have been focused on Google Play. PlayDrone [87] also performed a large-scale characterization of 1.1 million apps published in Google Play. They explored various issues such as app evolution and authentication schemes. Bogdan *et al.* [49] analyzed 160,000 Google Play apps daily for a period of 6 months, aiming to summarize the temporal patterns. Ali *et al.* [36] quantitatively compared app market attributes (e.g., ratings and prices) of Apple store and Google Play based on 80,000 app pairs. Wang *et al.* [92] presented a large-scale study of 791,138 removed Google Play apps to identify potential reasons for app removal. Wang *et al.* [94] analyzed the mobile app ecosystem from the perspective of app developers based on over 1.2 million apps and 320,000 developers.

Measurement of Alternative Markets. For third-party markets, Petsas *et al.* [77] analyzed 4 alternative markets to understand the downloading patterns and popularity trends. Ng *et al.* [76] assessed the trustworthy level of 20 Chinese app markets, but they only studied roughly 500 apps. Wang *et al.* [95] have studied gaming apps across 4 Chinese markets to understand their scale and evolution. WuKong [89] was proposed to identify repackaged apps in five Chinese app markets.

10 CONCLUSION

In this work, we have conducted a large-scale mobile app analysis to understand various features of several Chinese Android app stores and how they compare to Google Play. Specifically, our analysis covers over 6 million Android apps obtained from 16 Chinese app markets and Google Play. Overall, our results suggest that there are substantial differences between the Chinese app ecosystem and Google Play, though some minor commonalities are also found. We have identified a significant number of developers and third-party services specialized in the Chinese market. We have also found a higher prevalence of fake, cloned, and malicious apps in Chinese stores than in Google Play, possibly due to market operators indulgently overlooking copyright and security checks over the apps. We believe that our research efforts can positively contribute to bring user and developer awareness, attract the focus of the research community and regulators, and promote best operational practices across app store operators.

ACKNOWLEDGMENTS

We sincerely thank our shepherd Prof. Zhenhua Li (Tsinghua University), and all the anonymous reviewers for their valuable suggestions and comments to improve this paper. This work is supported by the National Key Research and Development Program of China (grant No.2018YFB0803603), the National Natural Science Foundation of China (grants No.61702045, and No.61772042); the BUPT Youth Research and Innovation Program (No.2017RC40); Spain’s Ministry of Economy and Competitiveness (grant TIN2016-79095-C2-2-R); the Madrid Region’s Technologies 2014 Research Program (grant S2013/ICE3095); the US National Science Foundation (grant CNS-1564329); and the European Union’s Horizon 2020 Innovation Action programme (grant Agreement No. 786741, SMOOTH Project).

REFERENCES

- [1] Human Inspection Team in Huawei, 2016. <http://android.tgbus.com/news/bd/201609/552547.shtml>.
- [2] Principle of least privilege - Wikipedia, 2017. https://en.wikipedia.org/wiki/Principle_of_least_privilege.
- [3] The top 10 Android app stores in China in 2017, 2017. <https://technode.com/2017/06/02/top-10-android-app-stores-china-2017/>.
- [4] 2017-2018 App Market Ranking in China-iiMedia Research, 2018. <http://www.iimedia.cn/60947.html>.
- [5] 2018 Top 10 App Markets in China, 2018. https://www.sohu.com/a/233552979_427024.
- [6] 360 Market - App Vetting, 2018. <http://dev.360.cn/wiki/index/id/18>.
- [7] 360 Security - Free Antivirus, Booster, Cleaner, 2018. https://play.google.com/store/apps/details?id=com.qihoo.security&hl=en_US.
- [8] Ali Platform - App Vetting, 2018. <http://aliapp.open.uc.cn/wiki/?p=140>.
- [9] Android Developer - APK Signer, 2018. <https://developer.android.com/studio/command-line/apksigner>.
- [10] Android Developers - Permissions Overview, 2018. <https://developer.android.com/guide/topics/permissions/overview>.
- [11] AndroZoo, 2018. <https://androzoo.uni.lu/>.
- [12] Anzhi Platform - App Vetting, 2018. <http://dev.anzhi.com/help.php?type=help&id=67>.
- [13] App China Platform - App Vetting, 2018. <http://dev.appchina.com/dev/help?choose=check>.
- [14] App Market Ranking in China, 2018. <http://chanye.07073.com/guonei/1756627.html>.
- [15] Baidu Market - App Vetting, 2018. <http://app.baidu.com/docs?id=18&frompos=401006>.
- [16] Developer Policy Center - Google Play, 2018. <https://play.google.com/about/developer-content-policy/>.
- [17] Developer Policy Center - Tencent Myapp, 2018. <http://wiki.open.qq.com/wiki/%E5%BA%94%E7%94%A8%E4%B8%A%E6%9E%B6%E8%A7%84%E5%88%99>.
- [18] Facebook Graph API, 2018. <https://developers.facebook.com/docs/graph-api>.
- [19] How to use the Play Console, 2018. <https://support.google.com/googleplay/android-developer/answer/6112435?hl=en>.
- [20] Huawei has surpassed Apple as the world's second largest smartphone brand, 2018. <https://www.theverge.com/2017/9/6/16259810/huawei-apple-global-smartphone-sales>.
- [21] Huawei Market - App Vetting, 2018. <http://developer.huawei.com/consumer/cn/devservice/doc/50104>.
- [22] Lenovo Market - App Vetting, 2018. http://open.lenovo.com/developer/adp/helpData/database_detail.jsp?url=http://open.lenovo.com/sdk/yysh/.
- [23] LIQU Platform - App Vetting, 2018. http://dev.liqucn.com/User/show_article/6.
- [24] Meizu Market - App Vetting, 2018. <http://open-wiki.flyme.cn/index.php?title=%E9%AD%85%E6%97%8F%E5%BA%94%E7%94%A8%E5%95%86%E5%BA%97%E5%BA%94%E7%94%A8%E5%AE%A1%E6%A0%B8%E8%A7%84%E8%8C%83>.
- [25] OPPO Market - App Vetting, 2018. <https://open.oppomobile.com/wiki/index#id=73486>.
- [26] PScout: Analyzing the Android Permission Specification, 2018. <http://pscout.csl.toronto.edu>.
- [27] Smartphone Market in China, 2018. <https://www.statista.com/topics/1416/smartphone-market-in-china/>.
- [28] SOGOU Platform - App Vetting, 2018. <http://zhushou.sogou.com/open/question-14.html>.
- [29] Top 10 Android App Stores in China, 2018. <https://newzoo.com/insights/rankings/top-10-android-app-stores-china/>.
- [30] Top 10 Chinese App Markets, 2018. http://www.sohu.com/a/223230295_100075221.
- [31] VirusTotal, 2018. <https://www.virustotal.com/>.
- [32] WeChat SDK, 2018. <https://open.weixin.qq.com/>.
- [33] Xiaomi Market - App Vetting, 2018. <https://dev.mi.com/console/doc/detail?pid=879>.
- [34] 360 Jiagu, 2017. <http://jiagu.360.cn>.
- [35] V. M. Afonso, P. L. de Geus, A. Bianchi, Y. Fratantonio, C. Kruegel, G. Vigna, A. Doupe, and M. Polino. Going native: Using a large-scale analysis of android apps to create a practical native-code sandboxing policy. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2016.
- [36] M. Ali, M. E. Joorabchi, and A. Mesbah. Same app, different app stores: A comparative study. In *Proceedings of the International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, 2017.
- [37] Aliyun ECS, 2017. <https://ecs-buy.aliyun.com>.
- [38] B. Andow, A. Nadkarni, B. Bassett, W. Enck, and T. Xie. A study of grayware on google play. In *Proceedings of the IEEE Security and Privacy Workshops*, 2016.
- [39] Monetize, advertise and analyze Android apps, 2017. www.appbrain.com/.
- [40] D. Arp, M. Spreitzenbarth, H. Gascon, K. Rieck, and C. Siemens. Drebin: Effective and explainable detection of android malware in your pocket. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2014.
- [41] K. W. Y. Au, Y. F. Zhou, Z. Huang, and D. Lie. Pscout: analyzing the android permission specification. In *Proceedings of the ACM SIGSAC conference on Computer and communications security (CCS)*, 2012.
- [42] V. Avdiienko, K. Kuznetsov, I. Rommelfanger, A. Rau, A. Gorla, and A. Zeller. Detecting behavior anomalies in graphical user interfaces. In *Proceedings of the International Conference on Software Engineering Companion (ICSE-C)*, 2017.
- [43] M. Backes, S. Bugiel, and E. Derr. Reliable third-party library detection in android and its security applications. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2016.
- [44] A. Bartel, J. Klein, Y. Le Traon, and M. Monperrus. Automatically securing permission-based software by reducing the attack surface: An application to android. In *Proceedings of the IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2012.
- [45] M. Bierma, E. Gustafson, J. Erickson, D. Fritz, and Y. R. Choe. Andlantis: Large-scale android dynamic analysis. *arXiv preprint arXiv:1410.7751*, 2014.
- [46] How to Access Google Play Store in China?, 2017. <https://www.bestvpnprovider.com/how-to-access-google-play-store-china/>.
- [47] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In *Proceedings of the International conference on Human computer interaction with mobile devices and services*, 2011.
- [48] P. Calciati and A. Gorla. How do apps evolve in their permission requests?: a preliminary study. In *Proceedings of the International Conference on Mining Software Repositories (MSR)*, 2017.
- [49] B. Carburnar and R. Potharaju. A longitudinal study of the google app market. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2015.
- [50] P. H. Chia, Y. Yamamoto, and N. Asokan. Is this app safe?: a large scale study on application permissions and risk signals. In *Proceedings of the International conference on World Wide Web (WWW)*, 2012.
- [51] J. Crussell, C. Gibler, and H. Chen. Attack of the clones: detecting cloned applications on Android markets. In *Proceedings of the European Symposium on Research in Computer Security (ESORICS)*, 2012.
- [52] J. Crussell, C. Gibler, and H. Chen. Scalable semantics-based detection of similar Android applications. In *Proceedings of the European Symposium on Research in Computer Security (ESORICS)*, 2013.
- [53] F. Dong, H. Wang, L. Li, Y. Guo, T. F. BissyandÃ, T. Liu, G. Xu, and J. Klein. Frauddroid: Automated ad fraud detection for android apps. In *Proceedings of the ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 2018.
- [54] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner. Android permissions demystified. In *Proceedings of the ACM conference on Computer and communications security (CCS)*, 2011.
- [55] C. Gibler, J. Crussell, J. Erickson, and H. Chen. Androidleaks: Automatically detecting potential privacy leaks in android applications on a large scale. In *Proceedings of the 5th international conference on Trust and Trustworthy Computing (TRUST)*, 2012.
- [56] C. Gibler, R. Stevens, J. Crussell, H. Chen, H. Zang, and H. Choi. AdRob: examining the landscape and impact of Android application plagiarism. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2013.
- [57] M. Grace, Y. Zhou, Q. Zhang, S. Zou, and X. Jiang. Riskranker: Scalable and accurate zero-day android malware detection. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2012.
- [58] S. Hanna, L. Huang, E. Wu, S. Li, C. Chen, and D. Song. Juxtapp: a scalable system for detecting code reuse among Android applications. In *Proceedings of the International Conference on Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA)*, 2012.
- [59] S. Hassan, W. Shang, and A. E. Hassan. An empirical study of emergency updates for top android mobile apps. *Empirical Software Engineering*, 22(1):505–546, 2017.
- [60] Y. Hu, H. Wang, Y. Zhou, Y. Guo, L. Li, B. Luo, and F. Xu. Dating with scambots: Understanding the ecosystem of fraudulent dating applications. *arXiv preprint arXiv:1807.04901*, 2018.
- [61] M. Ikram, N. Vallina-Rodriguez, S. Seneviratne, M. A. Kaafar, and V. Paxson. An analysis of the privacy and security risks of android vpn permission-enabled apps. In *Proceedings of the Internet Measurement Conference (IMC)*, 2016.
- [62] Y. Ishii, T. Watanabe, F. Kanei, Y. Takata, E. Shioji, M. Akiyama, T. Yagi, B. Sun, and T. Mori. Understanding the security management of global third-party android marketplaces. In *Proceedings of the ACM SIGSOFT International Workshop on App Market Analytics*, 2017.
- [63] S. M. Kywe, Y. Li, R. H. Deng, and J. Hong. Detecting camouflaged applications on mobile application markets. In *Proceedings of the International Conference on Information Security and Cryptology*, 2014.
- [64] L. Li, A. Bartel, T. F. BissyandÃ, J. Klein, Y. Le Traon, S. Arzt, S. Rasthofer, E. Bodden, D. Ocateau, and P. Mcdaniel. IccTA: Detecting Inter-Component Privacy Leaks in Android Apps. In *Proceedings of the International Conference on Software Engineering (ICSE)*, 2015.

- [65] L. Li, T. F. Bissyandé, J. Klein, and Y. Le Traon. An investigation into the use of common libraries in android apps. In *Proceedings of the IEEE International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 2016.
- [66] L. Li, T. F. Bissyandé, H. Wang, and J. Klein. Cid: automating the detection of api-related compatibility issues in android apps. In *Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)*, 2018.
- [67] L. Li, J. Gao, M. Hurier, P. Kong, T. F. Bissyandé, A. Bartel, J. Klein, and Y. Le Traon. Androozoo++: Collecting millions of android apps and their metadata for the research community. *arXiv preprint 1709.05281*, 2017.
- [68] Z. Li, W. Wang, C. Wilson, J. Chen, C. Qian, T. Jung, L. Zhang, K. Liu, X. Li, and Y. Liu. Fbs-radar: Uncovering fake base stations at scale in the wild. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2017.
- [69] Z. Li, W. Wang, T. Xu, X. Zhong, X.-Y. Li, Y. Liu, C. Wilson, and B. Y. Zhao. Exploring cross-application cellular traffic optimization with baidu trafficguard. In *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2016.
- [70] LibRadar - A detecting tool for 3rd-party libraries in Android apps, 2017. <https://github.com/pkumza/LibRadar>.
- [71] B. Liu, B. Liu, H. Jin, and R. Govindan. Efficient privilege de-escalation for ad libraries in mobile apps. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2015.
- [72] M. Liu, H. Wang, Y. Guo, and J. Hong. Identifying and analyzing the privacy of apps for kids. In *Proceedings of the International Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2016.
- [73] Z. Lu, Z. Li, J. Yang, T. Xu, E. Zhai, Y. Liu, and C. Wilson. Accessing google scholar under extreme internet censorship: A legal avenue. In *Proceedings of the ACM/IFIP/USENIX Middleware Conference: Industrial Track (Middleware)*, 2017.
- [74] 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 International Conference on Software Engineering Companion (ICSE-C)*, 2016.
- [75] A. Narayanan, L. Chen, and C. K. Chan. Addetect: Automated detection of android ad libraries using semantic analysis. In *Proceedings of the IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2014.
- [76] Y. Y. Ng, H. Zhou, Z. Ji, H. Luo, and Y. Dong. Which android app store can be trusted in china? In *Proceedings of the IEEE Computer Society International Conference on Computers, Software and Applications (COMPSAC)*, 2014.
- [77] T. Petsas, A. Papadogiannakis, M. Polychronakis, E. P. Markatos, and T. Karagiannis. Measurement, modeling, and analysis of the mobile app ecosystem. *ACM Trans. Model. Perform. Eval. Comput. Syst.*, 2(2):7:1–7:33, Mar. 2017.
- [78] Privacy Grade, 2017. <http://privacygrade.org/>.
- [79] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, S. Sundaresan, M. Allman, C. Kreibich, and P. Gill. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2018.
- [80] A. Razaghpanah, N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, P. Gill, M. Allman, and V. Paxson. Haystack: In situ mobile traffic analysis in user space. *ArXiv e-prints*, 2015.
- [81] J. Ren, M. Lindorfer, D. J. Dubois, A. Rao, D. Choffnes, and N. Vallina-Rodriguez. Bug fixes, improvements,... and privacy leaks. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2018.
- [82] J. Ren, A. Rao, M. Lindorfer, A. Legout, and D. Choffnes. Recon: Revealing and controlling pii leaks in mobile network traffic. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2016.
- [83] M. Sebastián, R. Rivera, P. Kotzias, and J. Caballero. Avclass: A tool for massive malware labeling. In *Proceedings of the International Symposium on Research in Attacks, Intrusions, and Defenses (RAID)*, pages 230–253. Springer, 2016.
- [84] List of countries by smartphone penetration, 2017.
- [85] D. Sounthiraraj, J. Sahs, G. Greenwood, Z. Lin, and L. Khan. Smv-hunter: Large scale, automated detection of ssl/tls man-in-the-middle vulnerabilities in android apps. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2014.
- [86] N. Vallina-Rodriguez, J. Shah, A. Finamore, Y. Grunenberger, K. Papagiannaki, H. Haddadi, and J. Crowcroft. Breaking for Commercials: Characterizing Mobile Advertising. In *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2012.
- [87] N. Viennot, E. Garcia, and J. Nieh. A measurement study of google play. In *Proceedings of the International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, 2014.
- [88] H. Wang and Y. Guo. Understanding third-party libraries in mobile app analysis. In *Proceedings of the IEEE/ACM International Conference on Software Engineering Companion (ICSE-C)*, 2017.
- [89] 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 the International Symposium on Software Testing and Analysis (ISSTA)*, 2015.
- [90] H. Wang, Y. Guo, Z. Tang, G. Bai, and X. Chen. Reevaluating android permission gaps with static and dynamic analysis. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, 2015.
- [91] H. Wang, J. Hong, and Y. Guo. Using text mining to infer the purpose of permission use in mobile apps. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2015.
- [92] H. Wang, H. Li, L. Li, Y. Guo, and G. Xu. Why are Android Apps Removed From Google Play? A Large-scale Empirical Study. In *Proceedings of the International Conference on Mining Software Repositories (MSR)*, 2018.
- [93] H. Wang, Y. Li, Y. Guo, Y. Agarwal, and J. I. Hong. Understanding the purpose of permission use in mobile apps. *ACM Transactions on Information Systems (TOIS)*, 35(4):43, 2017.
- [94] H. Wang, Z. Liu, Y. Guo, X. Chen, M. Zhang, G. Xu, and J. Hong. An explorative study of the mobile app ecosystem from app developers' perspective. In *Proceedings of the International Conference on World Wide Web (WWW)*, 2017.
- [95] T. Wang, D. Wu, J. Zhang, M. Chen, and Y. Zhou. Measuring and analyzing third-party mobile game app stores in china. *IEEE Transactions on Network and Service Management*, 13(4):793–805, 2016.
- [96] F. Wei, Y. Li, S. Roy, X. Ou, and W. Zhou. Deep ground truth analysis of current android malware. In *Proceedings of the International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*, 2017.
- [97] L. Wu, M. Grace, Y. Zhou, C. Wu, and X. Jiang. The impact of vendor customizations on android security. In *Proceedings of the ACM SIGSAC Conference on Computer and communications security (CCS)*, 2013.
- [98] X. Yang, D. Lo, L. Li, X. Xia, T. F. Bissyandé, and J. Klein. Characterizing malicious android apps by mining topic-specific data flow signatures. *Information and Software Technology*, 2017.
- [99] F. Zhang, H. Huang, S. Zhu, D. Wu, and P. Liu. ViewDroid: towards obfuscation-resilient mobile application repackaging detection. In *Proceedings of the ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*, 2014.
- [100] M. Zheng, P. P. Lee, and J. C. Lui. Adam: an automatic and extensible platform to stress test android anti-virus systems. In *Proceedings of the International conference on Detection of Intrusions and Malware, and Vulnerability assessment (DIMVA)*, 2012.
- [101] W. Zhou, Y. Zhou, M. Grace, X. Jiang, and S. Zou. Fast, scalable detection of "piggybacked" mobile applications. In *Proceedings of the ACM Conference on Data and Application Security and Privacy (CODASPY)*, 2013.
- [102] W. Zhou, Y. Zhou, X. Jiang, and P. Ning. Detecting repackaged smartphone applications in third-party Android marketplaces. In *Proceedings of the ACM Conference on Data and Application Security and Privacy (CODASPY)*, 2012.
- [103] Y. Zhou and X. Jiang. Dissecting android malware: Characterization and evolution. In *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*, 2012.