Measuring and Analyzing Third-Party Mobile Game App Stores in China

Tingting Wang, Di Wu, *Member, IEEE*, Jiaming Zhang, Min Chen, *Senior Member, IEEE*, and Yipeng Zhou

***Abstract*—In the era of mobile Internet, mobile game apps (i.e., applications) enable users to play games on mobile devices anywhere at any time. Such a change has brought a dramatic revolution to the traditional gaming industry. In this paper, we aim at having a comprehensive understanding of the ecosystem of mobile game apps. To this purpose, we conduct a large-scale measurement study over all game apps hosted by four leading app stores in China, which cover both Android and iOS platforms. We collect information of over 75 000 mobile game apps in a period of three months (October 2014–January 2015). With obtained datasets, we study the scale, evolution, and overlap of game apps in different app stores from a macroscopic level. We find that none of major app stores can provide a complete set of all game apps. We also investigate download patterns of mobile game apps and the impacts brought by user comments and ratings. We observe clear Pareto effect and power-law effect for game app downloads, and there is no strong positive correlation between app score and the number of its downloads. Last, we characterize the features of popular and unpopular game apps and confirm the negative impacts of embedded ads and paid items. We believe our measurement results can provide useful insights and advice for users, developers, and app store operators.**

***Index Terms*—Measurement study, mobile Internet, mobile game apps, popularity analysis.**

I. Introduction

THE PAST decade has witnessed the rapid growth of mobile Internet, and mobile penetration has reached a very high level in recent years [1]. Nowadays, almost everyone has a mobile device (e.g., smartphone, iPad). According to the report in [2], the worldwide shipment of smartphones has reached 1432.9 million in 2016. Driven by the prevalence of mobile devices, a large number of mobile applications have been developed and widely installed by mobile users. As reported in [3], 150 billion mobile apps have been downloaded in 2015, and the number is predicted to reach 200 billion until 2018.

Typically, users are downloading mobile apps from a mobile app store (e.g., Apple's App Store [4], Google Play [5], etc.), which contains a large set of diverse mobile apps. App developers (either individuals or companies) can upload their apps to one or more app stores for fast distribution among mobile users [6]. The mode of app store provides a profitable opportunity for app developers to benefit from app development. In addition to the most famous Apple's App Store and Google Play, many other companies, such as smartphone vendors, telecom operators, software companies and so on, also run their own app stores. The major reason behind their initiatives lies in that, app stores have been becoming a very important portal of mobile Internet, which is similar to the role of search engine for the traditional Internet.

An overview of app markets in China is presented as below to reflect the importance of game apps. In China, among various types of mobile apps, game apps account for the largest portion in the market due to their popularity among users [7]. As reported in [8], mobile game apps take up 24% of the total revenue of the gaming market. Moreover, 53.3% of mobile users who have paid for app downloads are actually paying for game apps [9]. Most of existing research studies focused on analyzing user behaviors [10], app usage [11], malware detection [12], [13] and trustworthiness [14] of mobile apps. However, there is very limited work to specifically study mobile game apps as a whole. Differently, our work focuses our attention on characterizing and understanding the ecosystem of mobile game apps, which has not been fully revealed and well studied.

Our objective is to comprehensively understand the ecosystem of mobile game apps from different angles and identify key factors driving the popularity of game apps. To this purpose, we conduct a large-scale measurement study by retrieving all necessary information over three months from four leading app stores in China, namely, SnapPea [15], MobileAsst [16], AppChina [17] and PushSync [18]. A dedicated platform is developed and deployed to collect and process data. By conducting in-depth analysis, we obtain quite a few interesting and insightful results, which we believe can greatly benefit users, app developers and app store operators. Since user behaviors are similar, we believe that our findings

Manuscript received December 15, 2015; revised May 10, 2016 and July 22, 2016; accepted August 12, 2016. Date of publication August 18, 2016; date of current version December 8, 2016. This work was supported in part by the National Science Foundation of China under Grant 61272397, Grant 61572538, Grant 61572220, and in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant S20120011187. The associate editor coordinating the review of this paper and approving it for publication was B. Stiller. *(Corresponding author: Di Wu.)*

T. Wang, D. Wu, and J. Zhang are with the Department of Computer Science, Sun Yat-sen University, Guangzhou 510006, China, also with the Guangdong Province Key Laboratory of Big Data Analysis and Processing, Sun Yat-sen University, Guangzhou 510006, China, and also with the Collaborative Innovation Center of High Performance Computing, National University of Defense Technology, Changsha 410073, China (e-mail: wangtt22@mail2.sysu.edu.cn; wudi27@mail.sysu.edu.cn; zhangjm27@mail2.sysu.edu.cn).

M. Chen is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: minchen@ieee.org).

Y. Zhou is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: ypzhou@szu.edu.cn).

Digital Object Identifier 10.1109/TNSM.2016.2601329

will also have reference meanings for the operation of other app platforms. In summary, our contributions in this paper can be listed as follows:

- To have a big picture of the ecosystem, a distributed high-performance crawler is developed to collect the information over 75,000 mobile game apps from four leading app stores in China. Through analyzing the data, we find that none of the app stores can provide a complete set of game apps. The comparative study among the four stores enables us to identify all game genres of each store, which are rather diverse due to different user appetites associated with each app store.
- The download patterns are investigated from multiple aspects. We verify that the download pattern of mobile game apps follows *Pareto Effect* and *Power-Law Effect* with truncations on both ends, which was firstly reported by Petsas *et al.* [19]. A new metric based on *entropy theory* is proposed to quantify the stability of app daily downloads. The features of user comments and ratings are also examined.
- We analyze typical characteristics of popular game apps from multiple angles (e.g., downloads, entropy, ratings, comments and categories), in order to identify major factors that determine the popularity of game apps. We observe that popular game apps achieve more stable daily downloads. Moreover, the popularity of game apps does not correlate well with app size. Most popular game apps are in the category *Leisure & Puzzle*. The results are helpful for game app developers to enhance their designs.

The rest of this paper is organized as follows. Section II describes our measurement methodology and obtained datasets. Section III provides a macroscopic analysis of the whole ecosystem of mobile game apps. Section IV studies the download patterns of mobile game apps from multiple aspects. Section V analyzes typical characteristics of popular game apps. Section VI discusses the limitations of our work and possible extension for future study. Section VII surveys related work in the field. Finally, we conclude this paper in Section VIII.

II. Measurement Methodology

To examine the characteristics of mobile game apps, we conduct a large-scale measurement study and collect abundant information from mobile app stores. Next, we will describe our methodologies for data collection and data analysis.

A. Data Collection

To facilitate our data collection, we build a dedicated data collection platform as illustrated in Fig. 1.

The platform consists of a set of hosts (e.g., crawling hosts, game information parser, comment parser) to enable parallel crawling and information processing. Crawling hosts are responsible for crawling Web pages from mobile app stores and then pass them to both game information parser and comment parser. The game information parser and comment parser are used to extract useful information and comments related to a mobile game app from Web pages. All the parsed results

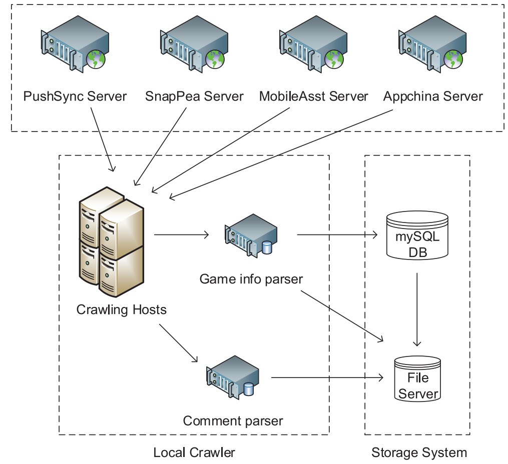


Fig. 1. Architecture of our data collection platform.

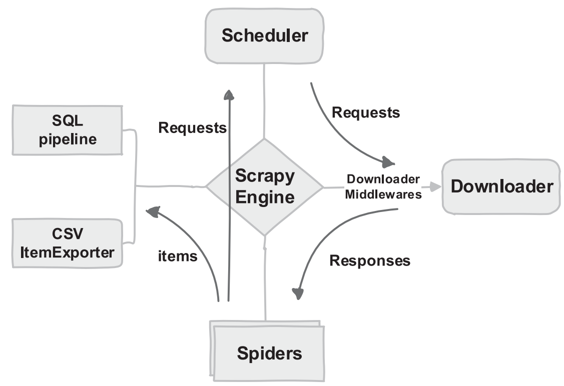


Fig. 2. Major components of our Scrapy-based Web page crawler.

are finally stored into a MySQL database. In addition, we also replicate all the Web pages on a file server to ease further analysis. Our crawler is designed to run at around 22:00 each day, and it is implemented to crawl its known apps before moving to the index pages of app stores to search new apps. The running time is influenced by multiple factors such as network conditions, the total number of game apps maintained by the store and crawling depth. It normally takes several hours (e.g., 1.5 hours for AppChina, 7 hours for PushSync) to complete the crawling task on one app store.

The core module of our data collection platform is a high-performance crawler, which is implemented based on Scrapy [20], an open-source collaborative framework for Web page crawling. Scrapy provides an easy approach to customize a Web spider according to user requirements. Fig. 2 depicts main components of our crawler implementation.

Generally, our crawler contains six major components: *Scrapy Engine* is responsible for controlling data flows among all components of the system; *Scheduler* is responsible for dispatching crawler tasks to a scheduler queue; *Downloader* fetches the next crawler task from the queue, downloads the Web page and then passes it to the spiders; *Spiders* are customized components which are used to parse the Web pages, extract useful items, store additional URLs to the scheduler for further crawling; *SQL Pipeline* is responsible for processing items once they have been extracted by the spiders, and

TABLE I
AVAILABILITY OF INFORMATION ITEMS ON EACH MOBILE APP STORE

Info. Item	SnapPea	MobileAsst	AppChina	PushSync
Title	√	√	√	√
Size	√	√	√	√
Category	√	√	√	√
Version	√	√	√	√
UpdateDate	√	√	√	√
Author	√	√	×	√
Num. of downloads	√	√	√	√
Num. of comments	√	√	√	√
Num. of likings	√	×	√	×
Avg. score	×	√	×	√
Num. of scores	×	√	×	√
Comment details	×	√	×	√
Tags	√	√	×	×
Description	√	√	√	√
Sys. requirements	√	√	√	√
HaveAds	√	√	√	×
isSafe	√	√	√	×
isFree	×	√	×	√

our customized pipeline can feed them to the database; *CSV ItemExporter* helps to export all captured items into a CSV file for storage.

B. Dataset Description

It is generally known that mobile OS vendors and device manufacturers will pre-install a number of apps such as news app, calendar or game center on devices before the devices are purchased by users. Telecom operators will also promote their apps by their distribution channels. To reduce the bias introduced by this kind of mobile app stores, we choose to obtain information of mobile game apps from independent third-party mobile app stores. In our measurements, we select four popular third-party mobile app stores in China as our targets, which include:

- MobileAsst [16] is the largest Android-based app store in China that occupies 19% of market share [21] and owns 120 million users by Dec. 2015 [22]. It is claimed to provide safe and fast distribution of mobile apps.
- SnapPea [15] is known as a simple, elegant and reliable app store among Chinese Android users. Its service was first launched in 2010.
- AppChina [17] is quite popular among experienced mobile game players because it shares abundant strategy tips on their Web forum to assist game players win the games. AppChina requires less steps for app developers to release their apps.
- PushSync [18] is a free iOS-based app store for iPhone/iPad users, especially for those who have their iPhone jailbroken. It provides accurate number of downloads for each iOS-based app. On the contrary, Apple app store does not provide such important information.

According to the survey report in [9], over 62.1% of mobile users in China choose third-party app stores as their default venue to download apps. The above app stores provide abundant information on mobile game apps. By parsing and extracting Web pages, we can create a detailed profile for each game app.

TABLE II
STATISTICS OF OUR DATASETS

AppStore	OS	# of Game Apps	Period
SnapPea	Android	19,863	2014/10/14 - 2015/1/14
MobileAsst	Android	27,845	2014/10/14 - 2015/1/14
Appchina	Android	10,493	2014/10/14 - 2015/1/14
PushSync	iOS	17,644	2014/11/10 - 2015/1/14

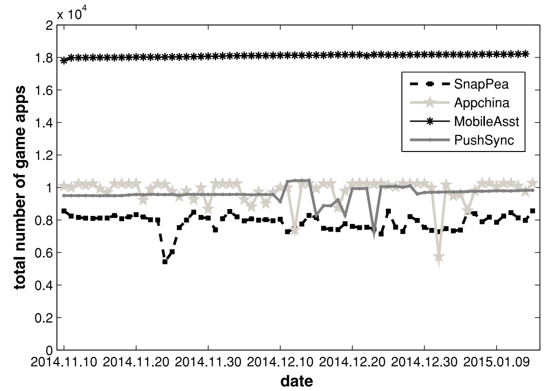


Fig. 3. Total number of mobile game apps in each app store.

Table I shows the availability of the above information items on each mobile app store.

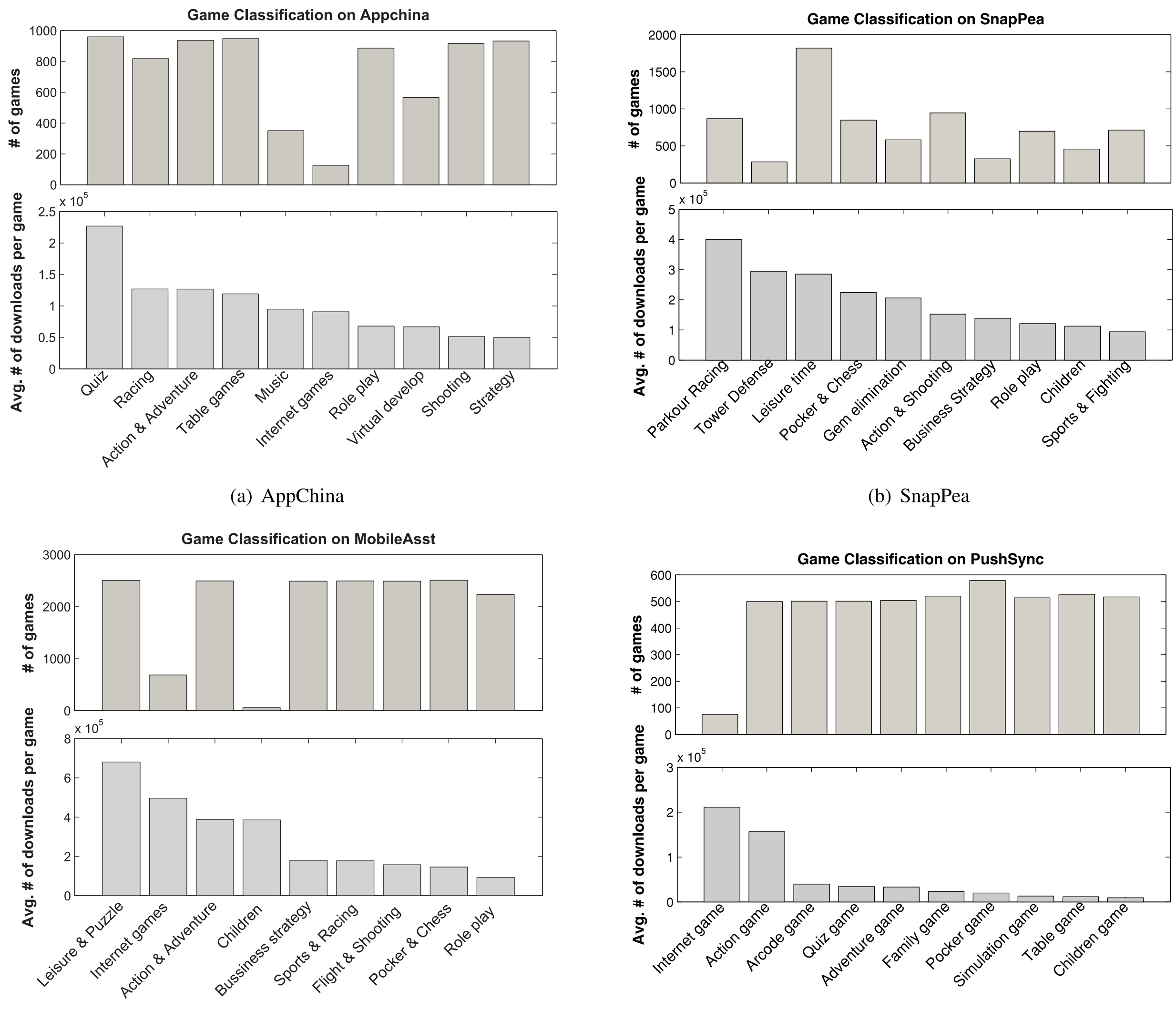
We performed data collection over a period of three months from Oct. 2014 to Jan. 2015, and retrieved information of over ten thousand mobile game apps from each store. The statistics of our datasets is shown in Table II. In the following sections, we will perform detailed data analysis over obtained datasets.

III. MACROSCOPIC VIEW OF THE ECOSYSTEM OF MOBILE GAME APPS IN CHINA

In this section, we perform a macroscopic analysis on the ecosystem of mobile game apps, which enables us to have a high-level picture on the scale and main features of mobile game apps in China.

We first plot the evolution of the total number of unique mobile game apps in each app store in Fig. 3. It is observed that the number of mobile game apps distributed by each app store is pretty stable. Among four app stores, MobileAsst distributes the highest number of mobile game apps, which is around 18 thousand per day. There also exist some small variations in terms of the total number of mobile game apps during the whole measurement period. We conjecture that it is caused by the removal of less popular games and the addition of new games, which are performed by app store operators periodically.

Different app stores classify mobile game apps in different ways. Fig. 4 shows the total number of game apps for each category and the average number of downloads per game in each category. For app store operators, such kind of information is useful to help determine the resource allocated to a particular app category and improve app distribution efficiency. There is high download correlation for game apps within the same category. We found that users show different appetites on different platforms. For AppChina, *Quiz* ranks the first among all categories in terms of the total number of downloads and the



(a) AppChina (b) SnapPea

(c) MobileAssistant (d) PushSync

Fig. 4. Total number of game apps for each category and the average number of downloads per game in each category.

average number of downloads per game. For SnapPea, in spite that *Leisure Time* has the largest number of game apps and the highest total number of downloads, its average number of downloads per game is smaller than *Parkour Racing* and *Tower Defense*. In MobileAsst, *Leisure & Puzzle* is the most highly downloaded category. For PushSync, *Action* ranks the first in terms of the total number of downloads, while *Internet Games* ranks the first in terms of the average number of downloads per game.

Next, we investigate the overlap of mobile game apps distributed by different mobile app stores. We aggregate all the mobile game apps collected from each app store and check their overlaps.[1] Fig. 5 shows the pairwise-intersection of mobile game apps for the four mobile app stores. The values for the top-100 popular mobile game apps are shown in parentheses. We observe that there exists significant overlap between SnapPea and MobileAsst. These two app stores share 38% common game apps among top-100 game apps. However, the overlap among other mobile app stores is commonly less than 16% of their hosted game apps. An interesting observation is that, in spite that PushSync is an iOS-based mobile app store, it still shares quite a few common game apps with other Android-based mobile app stores (e.g., SnapPea, MobileAsst, AppChina). The figure also indicates that none

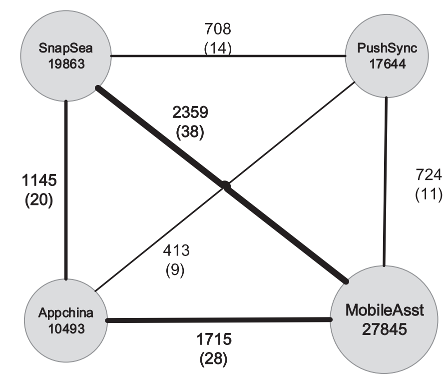


Fig. 5. Pairwise-intersection of all and top-100 mobile game apps for the four mobile app stores (top-100 shown in parentheses).

[1]If two game apps have the same title and version number, we consider them as the same game app.

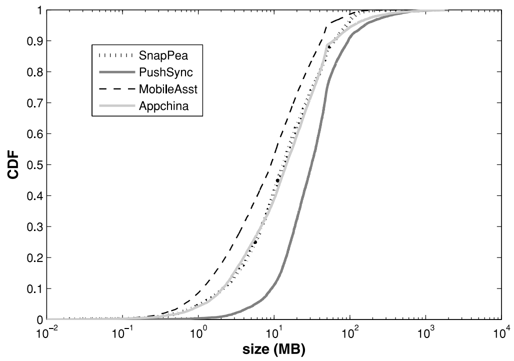


Fig. 6. Size distribution of mobile game apps.

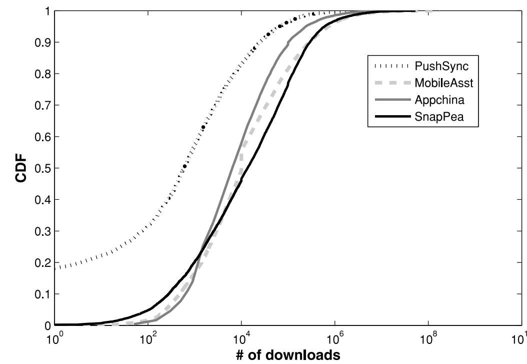


Fig. 7. CDF of the number of downloads for all game apps.

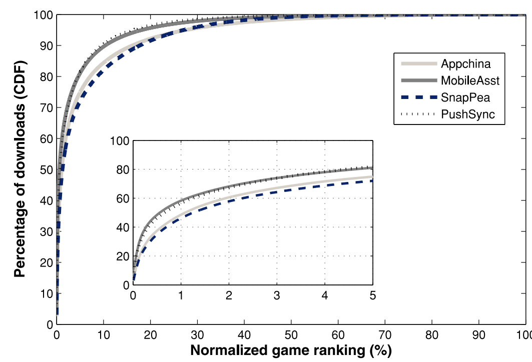


Fig. 8. CDF of game app downloads.

of the major mobile app stores on its own can provide a complete set of mobile game apps. It also points out that mobile game app developers may not prefer to release their apps on all the third-party app stores because it is both time-consuming and costly to provide apps on too many app stores. Uploading an app to more app stores will increase the operational and promotion cost. In addition, app stores have different application procedures, review requirements, and so on, implying the investment of more human resources. Moreover, if a game app becomes popular, the app store will promote the app by putting the app on the front page or allocating more recommendation resources, and that is why we can see a higher overlap rate on the top-100 popular apps.

Fig. 6 illustrates the size distribution of mobile game apps in different app stores. Due to resource constraints (e.g., battery, storage, bandwidth) on mobile devices, it is not surprising to find that most of mobile game apps are smaller than 100 MB. Over 90% of mobile game apps have a size in the range of 1 MB to 100 MB. One interesting finding is that iOS-based mobile game apps (e.g., apps on PushSync) are generally larger than Android-based mobile game apps. It may be caused by specific program structure of iOS apps. For the same game app, its iOS version is commonly larger than its Android version. For example, the Android built-in Unity module automatically compresses the textures in game scenes, while iOS will not perform compression in default.

IV. Understanding User Behaviors in Mobile Game App Stores

In this section, we investigate download patterns of mobile game apps and examine how the downloads are impacted by different factors.

A. Download Patterns of Game Apps

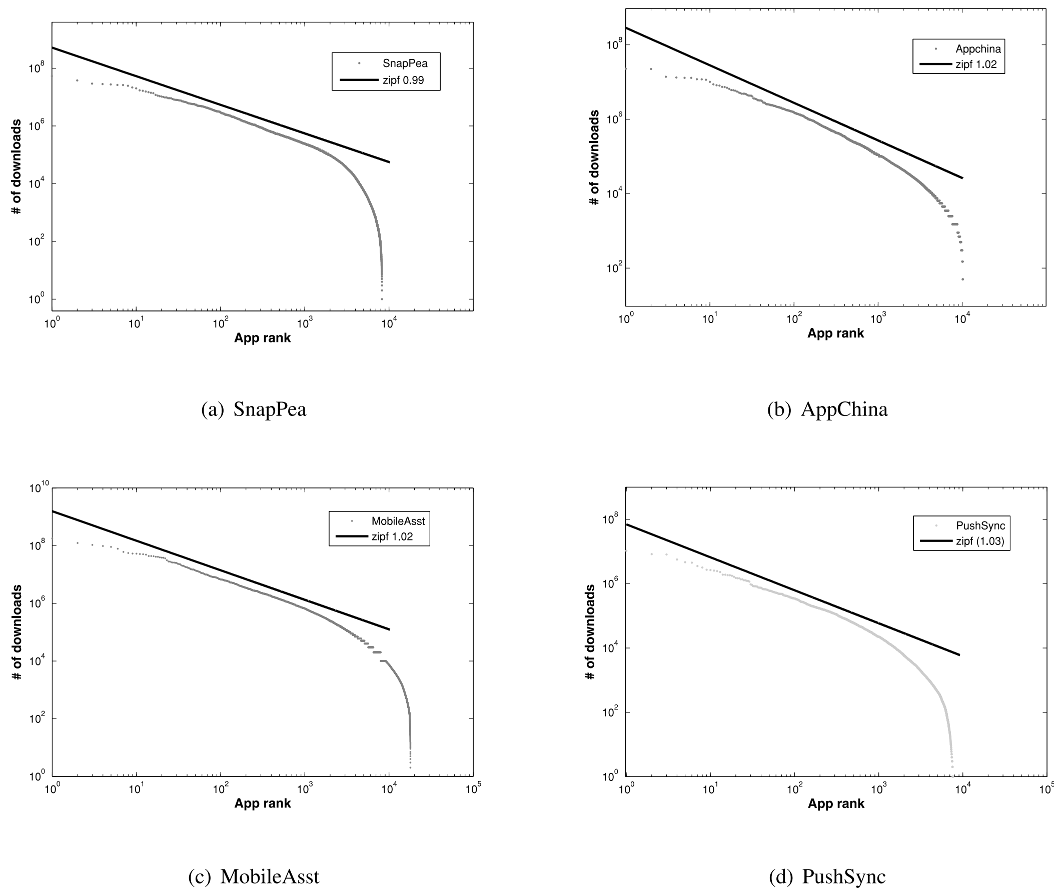
We first plot the cumulative distribution of the number of downloads for all game apps in Fig. 7. For most of Android-based game apps, the number of downloads ranges from 100 to 1,000,000. However, the number of app downloads on PushSync is much smaller than that on other Android-based app stores. Nearly 20% of iOS game apps on PushSync received no downloads. We conjecture that the main reason is that PushSync has a much smaller user population and quite a few unpopular game apps haven't been removed from the app store timely.

Fig. 8 further plots the CDF of the percentage of downloads as a function of game app ranks. The normalized game ranking is calculated from the game ranking. We first sorted all the apps by their total number of downloads in the descending order. r_i represents the ranking of the app and d_i represents the total number of downloads of the app. The normalization equations are presented as follows:

$$R_i = \frac{r_i}{N},$$
$$D_i = \frac{\sum_{j=1}^{j=i} d_j}{\sum_{j=1}^{j=N} d_j}.$$

We observe that 20% of mobile game apps contribute to around 90% of the downloads. Moreover, the top 2% of game apps account for over 60% of all the downloads. The observations clearly point out the download pattern of mobile game apps follows a typical Pareto principle [23], which has also been confirmed in [19]. Such Pareto effect of app downloads can facilitate the distribution of mobile apps by performing effective caching [24] or P2P app sharing. In fact, we have observed the Matthew effect for app downloads. The Matthew effect can be explained intuitively as follows. A top-ranked app has a better chance to be discovered and downloaded by mobile users because popular apps have more opportunities to be recommended by app stores.

To have a better understanding, we plot the number of downloads for game apps ranked by their popularity in Fig. 9.



(a) SnapPea (b) AppChina

(c) MobileAsst (d) PushSync

Fig. 9. The number of downloads for game apps ranked by their popularity.

The distribution of app downloads follows a similar pattern across four app stores. It is obvious to observe that there exists a Zipf-like distribution [25] in the center body and sharp truncations at both ends. The truncation for higher-ranked apps is largely due to the *fetch-at-most-once* principle, which has been first observed in peer-to-peer file-sharing system [26]. Due to the *fetch-at-most-once* principle, users normally download an app only once, which results in that the number of downloads for popular games is not as high as predicted by the Zipf's law. The truncation for lower-ranked apps can be explained by a phenomenon called *clustering effect* [19], which unveils that users tend to download the next app from the same category as their previously-downloaded app. It causes that less popular game apps gain more downloads than predicted by the Zipf's law.

B. Entropy of Daily Downloads

In order to study the fluctuation of daily downloads of game apps, we calculate the measure of *entropy of daily downloads* for each game app. Denote the entropy of the j-th game app over a period of T days as $H_j(T)$. The definition of $H_j(T)$ is given by:

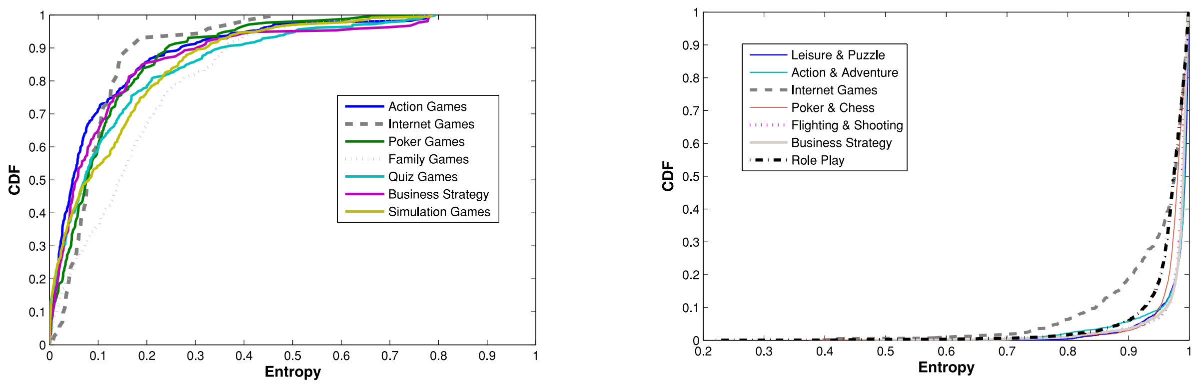
$$H_j(T) = -\frac{1}{\ln T} \sum_{i=1}^{T} \frac{d_{ji}}{\sum_{i=1}^{T} d_{ji}} \ln \frac{d_{ji}}{\sum_{i=1}^{T} d_{ji}},$$

where d_{ji} is the number of downloads of game app j on day i since it has been uploaded to the app store. A higher entropy value implies that the number of daily downloads of a game app is more stable. After normalization, the value of entropy is in the range of [0, 1].

Fig. 10 illustrates the CDF of 60-day normalized entropy of game apps in different categories. Obviously, there exists significant difference between PushSync and MobileAsst in terms of CDF curves. The figure shows that the entropy values of most mobile game apps on MobileAsst are close to 1. The high entropy of MobileAsst may be caused by the rounding operation of the number of total downloads. For a given app on MobileAsst, if its number of total downloads is less than 10,000, the number displayed on the app store is accurate. However, in case that the number of total downloads is higher than 10,000, the last four digits will be rounded to all zeros. For example, 87,650 will be displayed as 80 thousands on the app store. Therefore, the variance of daily downloads will be reduced by such rounding operation, resulting in a high entropy value. The figure also points out the category of a game app has almost little influence on its entropy.

C. User Comments and Ratings

Fig. 11 depicts the cumulative distribution of the number of comments related to each game app. About 30% of the game apps in the Android platform receive no more than one



(a) PushSync (b) MobileAsst

Fig. 10. CDF of 60-day normalized entropy of game apps in different categories.

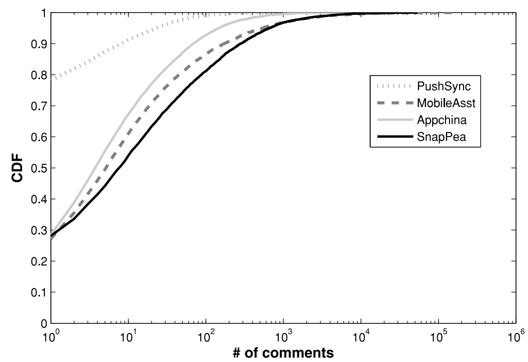


Fig. 11. Cumulative distribution of the number of comments on mobile game apps.

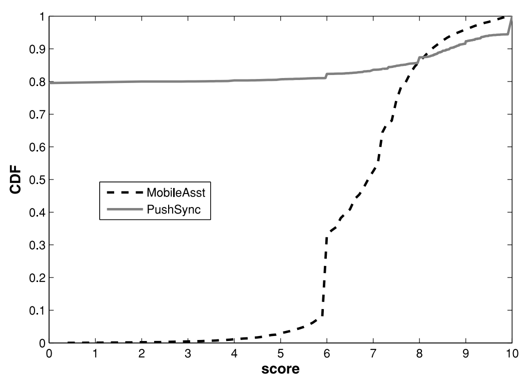


Fig. 12. Cumulative distribution of scores for each mobile game app.

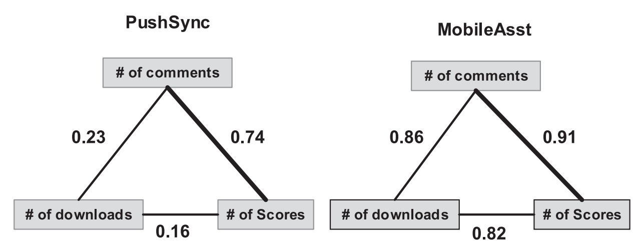


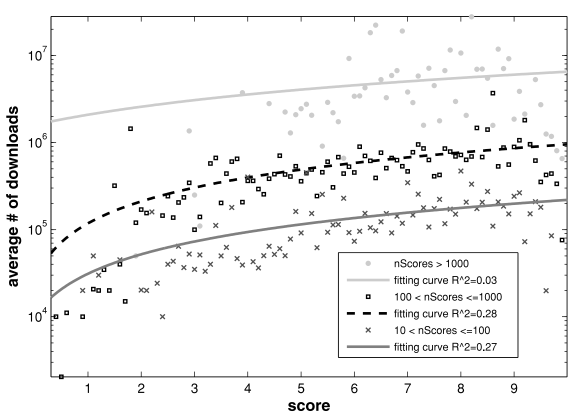
Fig. 13. Spearman correlation coefficient between different fields.

comments, while the percentage is even much larger in the iOS platform which is close to 80%. We further observe that the number of comments of most games on PushSync is less than 100, however the number of comments on SnapPea can reach more than 10,000. We conjecture that it is caused by the fewer number of users on PushSync and its privacy protection mechanism. It is reported in [27] that iOS users take more concerns on security for such third-party app stores, and they are unwilling to register accounts to post comments because of the identification and tracking-related privacy threats.

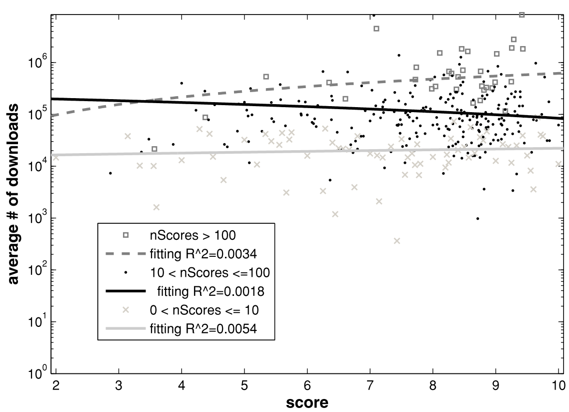
Fig. 12 plots the cumulative distribution of scores for each game app. In this figure, we can observe that 80% of iOS game apps do not receive a score on PushSync. It is because that low-ranked game apps do not have a score on PushSync. However, the condition is quite different on MobileAsst. Most of the game apps obtain a score higher than 6 (note that, the highest score is 10, and the initial score of each game is set to 6).

We have done a deeper investigation about game app downloads, comments and scores, and found the positive correlation in terms of Spearman correlation between the numbers of downloads, comments and app scores on MobileAsst and PushSync. According to Fig. 13, the number of comments has a relatively stronger positive relationship with the number of scores. It indicates that the one who posts a comment will prefer to give a score for the app at the same time. The low interdependency between the number of downloads and the number of comments on PushSync is because that PushSync allows users to make comments without checking their downloads. However, MobileAsst will do the check. From our measurement results, the Spearman correlation coefficient between the number of downloads and the number of comments on the four sites is higher than that of Pearson. This indicates that the rank based on the number of comments can reflect the rank based on the number of downloads.

We usually judge whether a mobile game app is popular by its number of downloads. However, *is a large number of downloads a reliable indicator for a popular game with high score?* We further conduct more investigation to show the relationship between app score and the number of downloads in Fig. 14. However, it is easy to find that there is no strong positive relationship between them, which means that a game app with a large number of downloads may not have a high score.



(a) MobileAsst



(b) PushSync

Fig. 14. Correlation between score and the number of app downloads.

This phenomenon is possibly caused by fake downloading introduced by third parties (e.g., game developers, advertisement companies).

V. Popularity Analysis of Mobile Game Apps

In this section, we analyze typical characteristics of popular game apps. Our main purpose is to identify major factors that impact the popularity of a game app. The results can help game app developers to design more popular game apps. To ease our presentation, MobileAsst and PushSync are selected as the representative for Android and iOS app stores respectively.

A. Characteristics of Popular Game Apps

In our measurement, *Popular Game Apps* are defined as game apps whose total downloads rank the top 1% among all the game apps in a given app store, and the rest 99% game apps are defined as *Unpopular Game Apps*. Since our platform cannot obtain the accurate number of daily downloads for each game app, we can only define *Popularity* as the number of total downloads. It is possible that an unpopular game app has a low and stable daily download rate, while a popular game app only has a few sharp spikes in terms of the number of daily downloads.

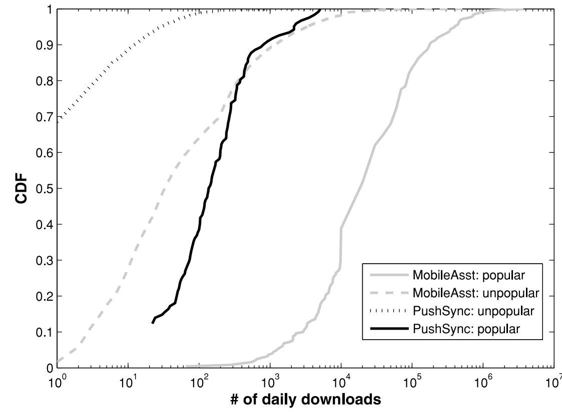


Fig. 15. Daily downloads of popular and unpopular game apps.

Fig. 15 depicts the CDF of the number of daily downloads for popular and unpopular game apps. The number of daily downloads for popular game apps is much higher than that of unpopular games. The daily downloads of most popular games on MobileAsst reach 10,000, which is also higher than that on PushSync.

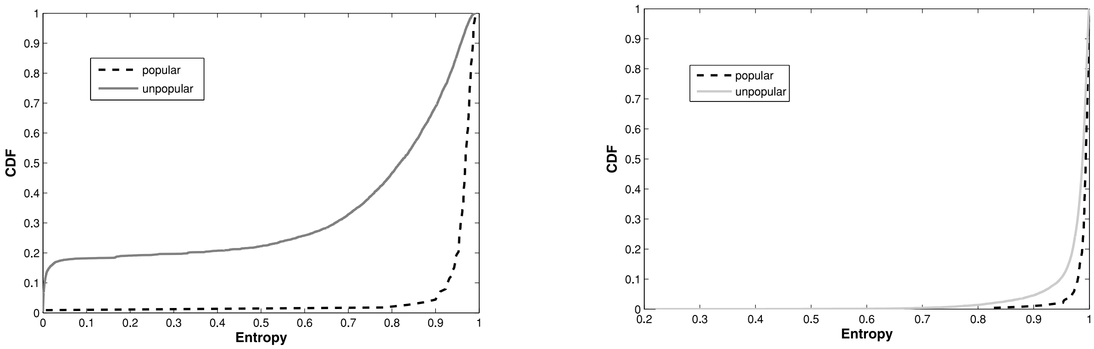
Fig. 16 further plots the CDF of the entropy of both popular and unpopular game apps. We observe that the entropy of popular game apps is generally higher than that of unpopular games. It indicates that popular game apps can attract much more stable daily downloads than unpopular game apps.

We then examine whether the app size has impacts on the popularity of a game app. Fig. 17 provides a scatter plot to depict the relationship between the app size and the number of app downloads in MobileAsst and PushSync. The points above the threshold line correspond to popular games. The figure shows that the sizes of most popular games range from 0.1 MB to 100 MB and there is no obvious causal relationship between the app size and the number of app downloads. It confirms that the app size doesn't play a key role to determine the popularity of a game app.

B. Impacts of Ratings, Comments and Categories

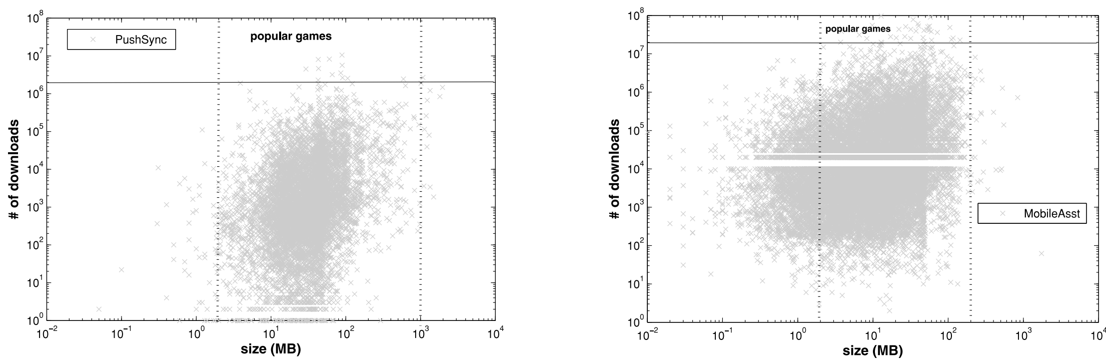
In a third-party app store, users can rate and comment on apps on their features and usability. These scores and comments are helpful for other downloaders to determine whether to download an app or not.

Fig. 18 shows the distribution of average scores of all game apps, namely, the sum of scores divided by the number of scores received by a game app. As the number of scores received by an app also reflects its popularity, we observe that popular game apps normally have a higher score than unpopular game apps on the Android-based MobileAsst. However, the phenomenon on PushSync is not as pronounced as that on MobileAsst. It is a bit counter-intuitive to see that a few unpopular game apps have a higher score than popular game apps. We conjecture that it is caused by the variance introduced by the small number of scores received by unpopular game apps. For example, an unpopular game may have a high score when only a few users download it and rate it with a good score.



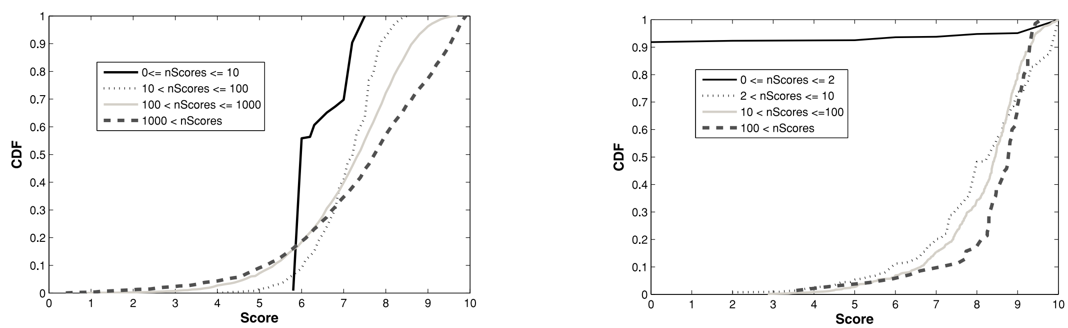
(a) PushSync (b) MobileAsst

Fig. 16. Entropy of popular and unpopular game apps.



(a) PushSync. (b) MobileAsst.

Fig. 17. Relationship between the app size and the number of app downloads.



(a) MobileAsst (b) PushSync

Fig. 18. Cumulative distribution of average scores of all game apps. (Note: *nScores* is the number of scores received by an app.)

Fig. 19 illustrates the CDF of the number of comments for popular and unpopular game apps. It is not surprising to see that popular game apps receive more comments than unpopular game apps. Most popular game apps on MobileAsst receive over 100 comments. However, it is a bit strange that 18% of popular game apps on PushSync receive no comments completely.

We further plot the number of daily comments for the top-5 game apps on MobileAsst and PushSync in Fig. 20(b). The number of daily comments and the number of downloads are mostly stable. The peak of downloads occurs every 7 days, and usually occurs on Sunday. Some special peaks are possibly the results of promotion activities [28], system recommendation, version updates, and so on [29].

We also investigate the impact of game category. From Fig. 21, we can find that the category *Leisure & Puzzle* accounts for 34% of popular games, which is significantly higher than other categories. Such kind of game apps are suitable for killing fragmented leisure time. Users can quickly start the game app, play for a while and then quit the game. We also find that a large fraction of *Poker & Chess* games are unpopular games, and *Children* category has very few games in both popular and unpopular games. We conjecture that the unpopularity of Poker & Chess game apps is caused by the homogenity among such kind of games. In addition, there are already many popular online poker and chess Web sites on the Internet (e.g., Tencent Games, Shanda). It also indicates the selection of game category when

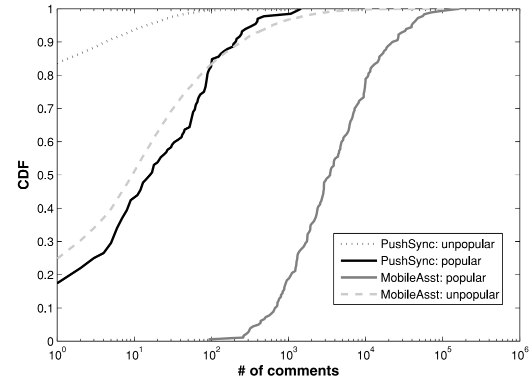
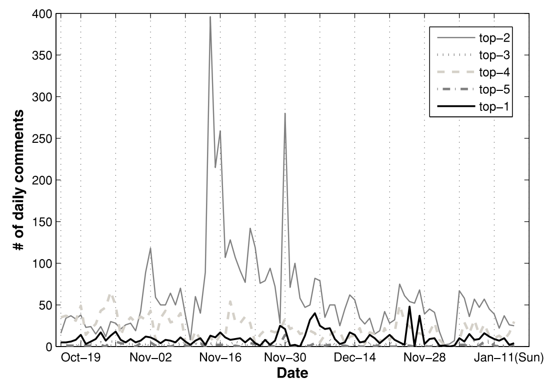
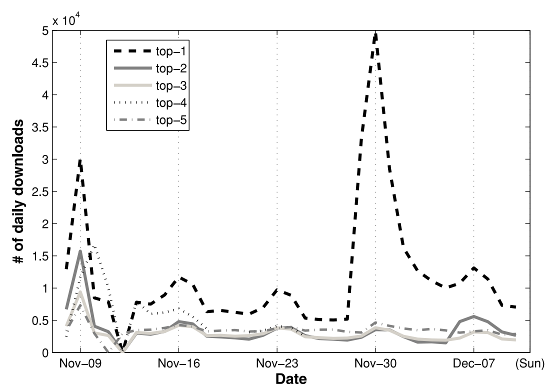


Fig. 19. CDF of the number of comments for popular and unpopular game apps.



(a) MobileAsst



(b) PushSync

Fig. 20. The number of daily comments for top-5 game apps.

developing a game app is critical for a game app to be popular.

C. Impacts of In-App Ads and Purchases

Game apps can be generally classified into *paid apps* and *free apps* based on their difference on the revenue model. For paid game apps, a user needs to pay for one-time charge in order to download the app. For free game apps, most apps contain either advertisements or in-app purchases instead [30]. In-app purchases include all items that a user should pay in a game app, such as game props that enable a user to have super capacity, game stages that a user should pay to play, VIP privilege for advertisement removal, etc.

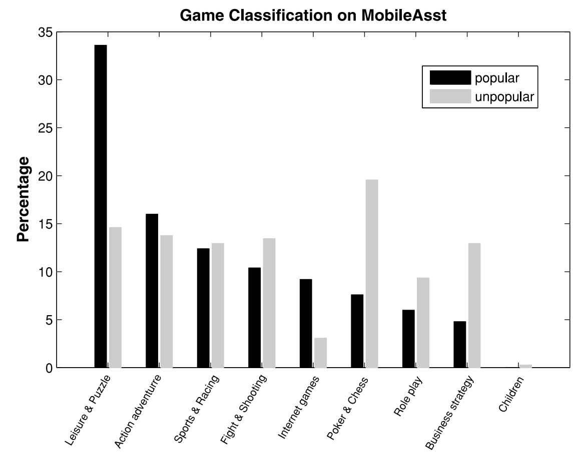


Fig. 21. Classification on MobileAssistant.

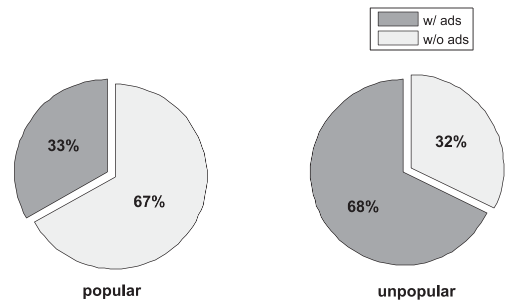


Fig. 22. Percentage of game apps embedded with Ads on MobileAsst.

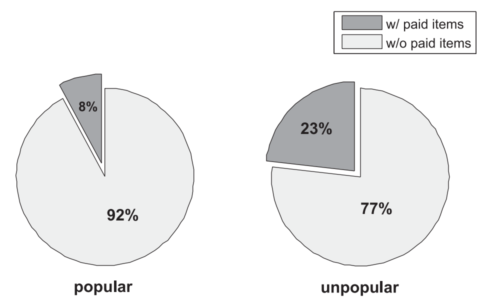


Fig. 23. Percentage of game apps embedded with paid items on MobileAsst.

In our measurement, all the four third-party app stores only provide free game apps. Therefore, we only focus on the study of free game apps in this paper. Next, we will investigate the impacts of embedded items (e.g., in-app ads, in-app purchases) on the popularity of game apps.

Fig. 22 shows that 67% of popular game apps do not contain any advertisements, while 68% of less popular game apps contain advertisements. Based on the above observation, we infer that more in-app advertisements will deteriorate user experience, and thus prevent the prevalence of an app.

In Fig. 23, we show the percentage of game apps embedded with in-app purchases (i.e., paid items) on MobileAsst. The figure points out that, 92% of popular games contain no paid items, while 23% of less popular games contain paid items. It indicates that users are uncomfortable with game apps embedded with paid items.

VI. Limitations and Future Work

In spite that we have made significant efforts, there still exist quite a few technical and non-technical obstacles that limit the scope and depth of our study. In this section, we would like to briefly discuss some limitations of our work and possible extension for further investigation.

First, although it is possible that duplicate apps may be maintained by the same app store, the detection of duplicate apps is not easy, which highly depends on the definition of duplicate apps. If we strictly define duplicated apps as files without any difference, then there are almost no duplicated apps in the same app store. However, we do find game apps with the same title but different version numbers, which can be defined as duplicated apps in some sense because these apps are highly similar. For this study, we only consider two game apps with the same title and version number as the same game app. In the future study, we can relax the definition of duplicate apps (e.g., game apps with the same title but different version numbers, or game apps with similar game contents) for the detection. The attrition rate of addition/deletion of game apps on MobileAsst ranges from -0.1% to 0.1%, which is the lowest, and the rate on Appchina ranges from -10% to 11%, which is the biggest. The rates on PushSync and SnapPea are $[-0.6\%, 0.9\%]$ and $[-6\%, 9\%]$ respectively.

Second, the categorization of game apps on four app stores is quite different. It is not easy to come up with a common categorization across the four app stores. Thus, it is possible that overlapping apps are categorized into different categories on separate app stores. Actually, we observe that around 95% of overlapping apps are marked into different categories in different app stores, in spite that some categories have similar names. In the next step, we also plan to investigate how to design a common categorization framework based on attributes and labels of game apps.

Third, there are be both positive and negative comments on a game app, which have different impacts on its popularity. In this study, we mainly focus on the correlation between the number of comments and the popularity of a game app, without studying the contents in the comments. In our future work, we need to perform text analysis to classify positive and negative comments and investigate their impacts in more details.

Fourth, the number of daily downloads on a few app stores will be rounded when the number is large, which will impact the accuracy of measurement statistics. For example, on MobileAsst, when the number of total downloads of a game app is higher than 10,000, the last four digits will be rounded to all zeros. In this case, the variance of daily downloads will be reduced by such rounding operation. In addition, game apps studied in this work are mostly free apps. It is because that almost all apps in the third-party app stores in China are free, and users are accustomed to downloading for free especially for those Android apps. In the next step, we also plan to extend our measurement scope to cover more game app stores, including both paid and free game apps.

VII. Related Work

Due to the popularity of mobile Internet, there have been plenty of research works to study the characteristics and user behaviors in the mobile app ecosystem from different perspectives.

Xu *et al*. [10] identified the diversity of usage behaviors of smartphone apps from spatial, temporal and user perspectives based on anonymized network measurements. Böhmer *et al*. [11] analyzed the behaviors of 4,100 users by collecting data from virtual sensors on the Android mobile devices. Tongaonkar *et al*. [31] studied the usage patterns of mobile apps by examining the mobile advertisement traffic in network traces. Li *et al*. [32] also studied user behaviors of millions of Android users and gained a few important insights. Petsas *et al*. [19] measured the app popularity distribution, modelled user download behaviors and explored potential pricing strategies for mobile apps. Zhong and Michahelles [33] examined the long tail phenomenon in the mobile app market, and found Google Play is dominated by the top 10% popular apps. Harman *et al*. [34] studied 32,108 non-zero priced apps in the Blackberry app store and found strong correlation between customer rating and the app rank. Liu *et al*. [35] conducted a measurement work to analyze the application popularity in both Android and iOS App Stores.

The measurements on user behavior and app popularity can help on the recommendation of apps in a mobile app market. In this field, there also exist quite a few research studies. Zhu *et al*. [36] developed a mobile app recommendation system that considers both app popularity and user security preferences. Lin *et al*. [37] exploited latent user models derived from Twitter datasets to address the cold-start problem in app recommendation. The security issues of mobile apps also receive significant attention. Yerima *et al*. [12] proposed a new malware detection approach based on Bayesian classification for Android-based mobile apps. Zhou and Jiang [13] collected 1,200 mobile malware samples and systematically characterized their features from multiple aspects. Ng *et al*. [14] investigated the trustworthiness of Android-based mobile apps for App Stores in China.

Different from previous work, we conduct a large-scale measurement study over four leading mobile app stores in China, including both Android and iOS platforms, and focus our attention solely on the ecosystem of mobile game apps. In addition to providing a macroscopic view of mobile game apps in China, we also analyze user behaviors and app popularity in a much more comprehensive manner. To our knowledge, we are the first to adopt the measure of entropy to evaluate the stability of daily downloads of mobile game apps and unveil its relationship with app popularity.

VIII. Conclusion

In this paper, we conducted a large-scale measurement study on four representative mobile game app stores in China. We performed a macroscopic analysis of the whole ecosystem of mobile game app markets in China. We compared the four app stores from multiple aspects to reveal the characteristics of each store. We also characterized the features of popular game apps. This work is helpful for app developers or app store operators to learn more details about Chinese game app markets. As more and more game apps are distributed via both Chinese app markets and global app markets, our findings on Chinese app markets can also help to predict the trend on the global app market. The comparison of Chinese app stores will also be useful to improve the operation efficiency of app store services. The characterization of popular game apps can help app stores to capture user interests more accurately. The comparison between Android game apps and iOS game apps can also provide advice for researchers to investigate the difference between Android users and iOS users. In the future work, we will extend the scope of our study and compare the difference between eastern and western app stores.

ACKNOWLEDGMENT

The authors would like to express special thanks to Hao Tian and Gaoxiang Li for their kindly help, which greatly improves the quality of this paper.

REFERENCES

[1] Ericsson. (Jun. 2015). *Ericsson Mobility Report*. [Online]. Available: http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf

[2] *Worldwide Smartphone Shipments*. (Jan. 2016). [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS40980416

[3] J. Jackson, "IDC: Customer reach and relations in the next-gen mobile app era," Opening Keynote, Mobile Research Summit: Data & Insights, New York, NY, USA, Jun. 2015. [Online]. Available: http://www.mobilemarketer.com/cms/lib/25700.pdf

[4] *Apple Store*. Accessed on Dec. 15, 2015. [Online]. Available: http://www.apple.com/cn/itunes/charts/

[5] *Google Play*. Accessed on Dec. 15, 2015. [Online]. Available: https://play.google.com/store

[6] F. Zhou, J. Liu, G. Simon, and R. Boutaba, "Joint optimization for the delivery of multiple video channels in Telco-CDNs," *IEEE Trans. Netw. Service Manag.*, vol. 12, no. 1, pp. 87–100, Mar. 2015.

[7] iiMedia, "China third-party mobile app store research," iiMedia Res. Group, Hong Kong, Tech. Rep. No.2012Q2, Sep. 2012. [Online]. Available: http://report.iresearch.cn/report/201208/1767.shtml

[8] Y. Zhang, N. Zheng, and X. Wang, "China gaming industry report," Game Publishing Committee, Internet Data Center, CNG, Tech. Rep. 201412, Beijing, China, Dec. 2014. [Online]. Available: http://www.cgigc.com.cn/d/game_report_2014_brief.pdf

[9] iiMedia, "China mobile app store market research report," iiMedia Res. Group, Hong Kong, Tech. Rep. No.2014Q3, Dec. 2014. [Online]. Available: http://www.iimedia.cn/14194788578794634n.pdf

[10] Q. Xu *et al.*, "Identifying diverse usage behaviors of smartphone apps," in *Proc. ACM SIGCOMM Conf. Internet Meas. Conf. (IMC)*, Berlin, Germany, 2011, pp. 329–344.

[11] 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 *Proc. 13th Int. Conf. Human Comput. Interact. Mobile Devices Services (MobileHCI)*, Stockholm, Sweden, 2011, pp. 47–56.

[12] S. Y. Yerima, S. Sezer, G. McWilliams, and I. Muttik, "A new android malware detection approach using Bayesian classification," in *Proc. IEEE 27th Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Barcelona, Spain, 2013, pp. 121–128.

[13] Y. Zhou and X. Jiang, "Dissecting android malware: Characterization and evolution," in *Proc. IEEE Symp. Security Privacy (SP)*, San Francisco, CA, USA, 2012, pp. 95–109.

[14] Y. Y. Ng, H. Zhou, Z. Ji, H. Luo, and Y. Dong, "Which android app store can be trusted in China?" in *Proc. IEEE 38th Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, Västerås, Sweden, 2014, pp. 509–518.

[15] *SnapPea*. Accessed on Jan. 14, 2015. [Online]. Available: https://www.wandoujia.com/

[16] *360 Mobile Assistant*. Accessed on Jan. 14, 2015. [Online]. Available: http://zhushou.360.cn/

[17] *AppChina*. Accessed on Jan. 14, 2015. [Online]. Available: http://www.appchina.com/

[18] *PushSync*. Accessed on Jan. 14, 2015. [Online]. Available: http://tui.tongbu.com/

[19] T. Petsas, A. Papadogiannakis, M. Polychronakis, E. P. Markatos, and T. Karagiannis, "Rise of the planet of the apps: A systematic study of the mobile app ecosystem," in *Proc. Conf. Internet Meas. Conf. (IMC)*, Barcelona, Spain, 2013, pp. 277–290.

[20] *Scrapy*. Accessed on Dec. 15, 2015. [Online]. Available: http://scrapy.org/

[21] NewZoo. (Mar 2016). *Top 10 Android App Stores in China*. [Online]. Available: https://newzoo.com/insights/rankings/top-10-android-app-stores-china/

[22] QuestMobile. (Dec. 2015). *Mobile Internet Industry Review*. [Online]. Available: http://doc.questmobile.com.cn/Mobile_Internet_Industry_Review.2015.pdf

[23] *Pareto Principle*. (2016). [Online]. Available: https://en.wikipedia.org/wiki/Pareto_principle

[24] Y. Li, H. Xie, Y. Wen, C.-Y. Chow, and Z.-L. Zhang, "How much to coordinate? Optimizing in-network caching in content-centric networks," *IEEE Trans. Netw. Service Manag.*, vol. 12, no. 3, pp. 420–434, Sep. 2015.

[25] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, YouTube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas. (IMC)*, San Diego, CA, USA, 2007, pp. 1–14.

[26] K. P. Gummadi *et al.*, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," *ACM SIGOPS Oper. Syst. Rev.*, vol. 37, no. 5, pp. 314–329, 2003.

[27] F. Shirazi and M. Volkamer, "What deters Jane from preventing identification and tracking on the Web?" in *Proc. 13th Workshop Privacy Electron. Soc. (WPES)*, Scottsdale, AZ, USA, 2014, pp. 107–116.

[28] *WeChat Dash Promotion Activities*. Accessed on Nov. 28, 2014. [Online]. Available: http://pao.qq.com/webplat/info/news_version3/6019/6021/6024/m4397/201411/292763.shtml

[29] Z. Li and G. Simon, "In a telco-CDN, pushing content makes sense," *IEEE Trans. Netw. Service Manag.*, vol. 10, no. 3, pp. 300–311, Sep. 2013.

[30] T. M.-H. Tveten, "How to make money on mobile applications," M.S. thesis, Dept. Telematics, Norwegian Univ. Sci. Technol., Trondheim, Norway, 2014.

[31] A. Tongaonkar, S. Dai, A. Nucci, and D. Song, "Understanding mobile app usage patterns using in-app advertisements," in *Proc. 14th Int. Conf. Passive Act. Meas. (PAM)*, Hong Kong, 2013, pp. 63–72.

[32] H. Li *et al.*, "Characterizing smartphone usage patterns from millions of android users," in *Proc. ACM Conf. Internet Meas. Conf. (IMC)*, Tokyo, Japan, 2015, pp. 459–472.

[33] N. Zhong and F. Michahelles, "Google play is not a long tail market: An empirical analysis of app adoption on the Google play app market," in *Proc. 28th Annu. ACM Symp. Appl. Comput. (SAC)*, Coimbra, Portugal, 2013, pp. 499–504.

[34] M. Harman, Y. Jia, and Y. Zhang, "App store mining and analysis: MSR for app stores," in *Proc. 9th IEEE Work. Conf. Min. Softw. Repositories (MSR)*, Zürich, Switzerland, Jun. 2012, pp. 108–111.

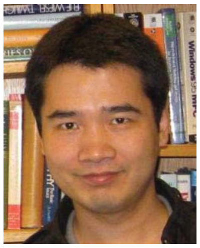
[35] W. Liu, G. Zhang, J. Chen, Y. Zou, and W. Ding, "A measurement-based study on application popularity in android and iOS app stores," in *Proc. Workshop Mobile Big Data (Mobidata)*, Hangzhou, China, 2015, pp. 13–18.

[36] H. Zhu, H. Xiong, Y. Ge, and E. Chen, "Mobile app recommendations with security and privacy awareness," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD)*, 2014, pp. 951–960.

[37] J. Lin, K. Sugiyama, M.-Y. Kan, and T.-S. Chua, "Addressing cold-start in app recommendation: Latent user models constructed from Twitter followers," in *Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval (SIGIR)*, Dublin, Ireland, 2013, pp. 283–292.

Tingting Wang received the B.S. degree in information security from Sun Yat-sen University, Guangzhou, China, in 2014, where she is currently pursuing the master's degree with the Department of Computer Science, under the supervision of Prof. D. Wu.

Her research interests include mobile social networks and big data analysis.



Di Wu (M'06) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 2000, the M.S. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2003, and the Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong, Hong Kong, in 2007. He was a Post-Doctoral Researcher with the Department of Computer Science and Engineering, Polytechnic Institute of New York University, Brooklyn, NY, USA, from 2007 to 2009, advised by Prof. K. W. Ross.

Dr. Wu is currently a Professor and the Assistant Dean of the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include cloud computing, multimedia communication, Internet measurement, and network security. He was a co-recipient of the IEEE INFOCOM 2009 Best Paper Award. He has served as an Editor for the *Journal of Telecommunication Systems* (Springer), the *Journal of Communications and Networks*, *Peer-to-Peer Networking and Applications* (Springer), *Security and Communication Networks* (Wiley), and the *KSII Transactions on Internet and Information Systems*, as well as a Guest Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He has also served as the MSIG Chair for the Multimedia Communications Technical Committee in the IEEE Communications Society from 2014 to 2016. He served as the TPC Co-Chair for the IEEE Global Communications Conference—Cloud Computing Systems, and Networks, and Applications in 2014, the Chair of the CCF Young Computer Scientists and Engineers Forum—Guangzhou from 2014 to 2015, and as a member of the Council of China Computer Federation.

Jiaming Zhang received the B.S. degree in information engineering from the South China University of Technology, Guangzhou, China, in 2015. He is currently pursuing the graduation degree with the Department of Computer Science, Sun Yat-sen University, Guangzhou.

His research interests mainly include cloud gaming, multimedia networking, and distributed system.

Min Chen (M'08–SM'09) received the B.S. degree in electronic and communication engineering from the South China University of Technology in 1999, and the M.S. and Ph.D. degrees in electronic and information technology from the South China University of Technology in 2001 and 2004, respectively.

He is a Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST). He is the Chair of the IEEE Computer Society Special Technical Communities on Big Data. He was an Assistant Professor with the School of Computer Science and Engineering, Seoul National University (SNU) from 2009 to 2012. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of British Columbia, for three years. He was a Post-Doctoral Fellow with SNU for one and a half years. He has over 280 paper publications, including over 150 SCI papers, over 60 IEEE transactions/journal papers, 8 ISI highly cited papers, and 1 hot paper. He has authored two books entitled *OPNET IoT Simulation* and *Big Data Inspiration* (HUST Press, 2015), and a book on big data entitled *Big Data Related Technologies* (Springer Series in Computer Science, 2014). He has over 6820 Google Scholars citations with an H-index of 41. His top paper was cited over 770 times, while his top book was cited 480 times as of 2016. His research focuses on cyber physical systems, IoT sensing, 5G networks, mobile cloud computing, SDN, healthcare big data, medica cloud privacy and security, body area networks, emotion communications and robotics. He was a recipient of the Best Paper Award from the IEEE ICC 2012 and the Best Paper Runner-up Award from QShine 2008. He serves as an Editor or an Associate Editor for *Information Sciences*, *Wireless Communications and Mobile Computing*, *IET Communications*, *IET Networks*, *Wiley I. J. of Security and Communication Networks*, the *Journal of Internet Technology*, *KSII Transaction on Internet and Information Systems*, and the *International Journal of Sensor Networks*. He is a Managing Editor for IJAACS and IJART. He is a Guest Editor for the IEEE NETWORKS, the *IEEE Wireless Communications Magazine*. He was the Co-Chair of the IEEE ICC 2012-Communications Theory Symposium and the IEEE ICC 2013-Wireless Networks Symposium. He was the General Co-Chair for the 12th IEEE International Conference on Computer and Information Technology (IEEE CIT-2012) and Mobimedia 2015. He was the General Vice Chair of Tridentcom 2014. He was a Keynote Speaker for CyberC 2012, Mobiquitous 2012, and Cloudcomp 2015.

Yipeng Zhou received the B.S. degree from the Department of Computer Science, University of Science and Technology of China, and the M.Phil. and Ph.D. degrees from the Department of Information Engineering, Chinese University of Hong Kong (CUHK). He is an Assistant Professor with the College of Computer Science and Software Engineering, Shenzhen University. From 2012 to 2013, he was a Post-Doctoral Fellow with the Institute of Network Coding, CUHK. His main research interests lie in modeling and analysis of large scaled networking systems, analysis of user behaviors of online video systems, and crowdsourcing-based content distribution.