AppWatcher: Unveiling the Underground Market of Trading Mobile App Reviews

Zhen Xie
Dept of Computer Science and Engineering
The Pennsylvania State University
zhenxie@cse.psu.edu

Sencun Zhu
Dept of Computer Science and Engineering &
College of Information Sciences and Technology
The Pennsylvania State University
szhu@cse.psu.edu

ABSTRACT
Driven by huge monetary reward, some mobile application (app) developers turn to the underground market to buy positive reviews instead of doing legal advertisements. These promotion reviews are either directly posted in app stores like iTunes and Google Play, or published on some popular websites that have many app users. Until now, a clear understanding of this app promotion underground market is still lacking. In this work, we focus on unveiling this underground market and statistically analyzing the promotion incentives, characteristics of promoted apps and suspicious reviewers. To collect promoted apps, we built an automatic data collection system, AppWatcher, which monitored 52 paid review service providers for four months and crawled all the app metadata from their corresponding app stores. Finally, AppWatcher exposes 645 apps promoted in app stores and 29,680 apps promoted in some popular websites. The current underground market is then reported from various perspectives (e.g., service price, app volume). We identified some interesting features of both promoted apps and suspicious reviewers, which are significantly different from those of randomly chosen apps. Finally, we built a simple tracer to narrow down the suspect list of promoted apps in the underground market.

Categories and Subject Descriptors
K.6.5 [Management of Computing and Information Systems]: [Security and Protection]; D.2.8 [Software Engineering]: Metrics

Keywords
Underground market, Mobile app reviews, Opinion Mining, App Stores, Fake Reviews

1. INTRODUCTION
The great popularity of smart devices (e.g., tablets and smartphones) has attracted more and more software developers to engage in mobile application (app) development. As of September 2014, both Apple App Store [1] and Google Play [2] offered 1.3 million apps. The entire app market has reached up to 35 billion dollars in 2013 and it could hit 70 billion dollars by 2017 [3]. Meanwhile, the competition to attract customers has become more furious than ever. To promote their apps, some developers choose to buy reviews from companies like bestreviewapp [4], dailyappshow [5]. This review trading market will cause great damage to app market players. On one hand, these paid reviews may mislead honest customers and cause both time and money losses to them. On the other hand, it will demoralize those developers who try their best effort to improve app qualities in order to receive more and better reviews. Therefore, paid reviews are forbidden by app store vendors. Apple App Store warns the developers not to cheat the user reviews with fake or paid reviews in Clause 3.10 of “App Store Review Guidelines” [6]; GooglePlay requires the developers not to manipulate product ratings or reviews with unauthorized means like fake or paid reviews in its “Google Play Developer Program Policies” [7]. Moreover, according to “The FTC’s Revised Endorsement Guides” [8], paid reviews are illegal if endorsers fail to show whether they are paid to write the reviews (currently almost no reviewers state whether their reviews are paid or not).

By the place where a paid review is posted, the app review market can be grouped into two types. The first type is to post reviews directly in app stores (referred to as app store promotion). Unlike traditional product stores, app stores are highly centralized; for example, iTunes is the only store that distributes apps for non-jailbroken devices of iTouch, iPhone, and iPad. Therefore, app reviews posted in these stores could be read by all the app users. However, as one user is only allowed to write one review for an app, non-conforming developers have to find many reviewers to promote their apps. This is often difficult. Thus, some service providers have established a business to charge developers who seek reviews and pay app users who like to make money for writing app reviews. The second promotion type is to post reviews on popular websites (referred to as web promotion), which have a large population of readers and followers in social networks like Facebook and Twitter. They may quickly broadcast app reviews to their readers and followers who are potential app users, and meanwhile, these websites would get paid by app developers. Until now, this app review trading market is still under the ground and unknown to the community.
In this paper, we aim to unveil the underground market of trading mobile app reviews and collect a set of promoted apps as the ground truth to further study their statistical characteristics. We joined their promotion communities as a passive reviewer or follower and collected apps assigned to us. We have built an automatic data collection system, called AppWatcher, to monitor 52 service providers every day. In four months, AppWatcher has gathered 645 apps reviewed by app store promotion and 29,680 apps reviewed by web promotion. It has also collected app meta data like rating, text comment, and reviewer information, and randomly chosen 186,263 apps from Apple App Store for comparison study. We have statistically studied service providers, promotion incentives, promoted apps, and suspicious reviewers. Moreover, we have compared their features with randomly chosen apps and tested the significance of their difference.

Our key contributions can be summarized as follows.

1. **Market Study:** To our best knowledge, we are the first to study the underground market of mobile app review promotion. Our results is very helpful to understand the current situation and growth of this market.

2. **New Discoveries:** We have found some interesting features of both promoted apps and suspicious reviewers. These features are significantly different from those of randomly chosen apps and they are critical to design advanced algorithms for detecting promoted apps or reviewers.

3. **Promoted Apps:** Except service providers, it is very difficult to get the ground truth of whether an app is a promoted app or not. All the promoted apps we have collected by joining as a recruited reviewer are definitely promoted. This app dataset is important to measure the accuracy of promoted app detection algorithms [9][10].

4. **Promoted App Tracer:** We design a simple app tracer, which can be extended by adding heuristics we discovered, to narrow down the suspect list of promoted apps in the underground market.

2. **DATA COLLECTION METHODOLOGY**

2.1 **Service Provider Collection**

Our basic way to find app promotion service providers is searching for their websites with the help of search engines. To look for app store review providers, we search with various combinations of keywords such as “paid”, “buy”, “app”, “review”, “rating”, “app stores”. For website review providers which charge for app reviews, we use the combinations of “paid”, “buy”, “app”, “review”, “evaluation”, “rating”. In addition, we have found one blog [11] that keeps updating the list of such websites.

2.2 **Service Provider’s Meta Data Collection**

To collect a service provider’s meta data information, we read its pages like FAQ, About, Pricing. Specifically, we collected such data as service price, reward to reviewers, service launching time. Service launching time is the time when a service provider started to review apps. Normally, a website shows its creation time (i.e., year) at the bottom of the front page. If there is no such information, we would refer to the whois information of the domain and use the domain creation time instead.

2.3 **Promoted App Collection**

We have built an automated data collection system, named AppWatcher, to gather promoted apps from the service providers. AppWatcher is implemented in JAVA and Python and it has 12,908 lines of source code. As illustrated in Fig. 1, the system includes four major modules: website monitor, assignment collector, app meta data crawler, and database.

![Figure 1: The interactions between AppWatcher and review trading market. The detailed actions between them are as follows. A1: App developers upload their apps to app stores; A2: Some developers buy reviews from service providers; A3: The "website monitor" module monitors target service providers to collect paid reviews. If the service provider is recruiting reviewers online, "assignment collector" module will join as a passive reviewer to collect assigned app information; A4: the module of "app meta data crawler" crawlers paid reviews along with their dates; A5: "app meta data crawler" retrieves these promoted apps’ meta data from their corresponding app stores.](image)
app, so our collector only collects the latest 5,000 reviews or all reviews (when total number is below 5,000).

2.4 Ethical Consideration

Our study on the underground market for review/ranking fraudulence is ultimately related to the real app market. One important rule in our experiments is not to influence the real market. Therefore, we have chosen to be an passive observer. Specifically, when crawling apps from a website, we slowed down the speed of crawling (i.e., twice a day) to avoid traffic jamming. In addition, when joining as a reviewer in a reviewer recruiting website, we only passively received the assignments without posting any review for any app. When crawling app meta data from app stores, we also limited the downloading speed.

3. APP STORE PROMOTION

3.1 Dataset

We have found 26 websites that have been involved in app store promotion. Among them, five websites were closed after being exposed very recently. These websites are used to advertise their services. Ten of them are even used to recruit reviewers online, and two of these ten websites require to review an app or pay for registration fees before joining. To not violate the ethical regulation, we deployed AppWatcher to only monitor the remaining eight websites. We registered an account as a reviewer in each website and recorded apps assigned to us. We logged into these websites twice a day and observed the change of these assignments. In total, we have collected 645 apps, which were mostly promoted from July 1st to October 7th, 2014.

Among these promoted apps, 618 are from the iOS platform while only 27 apps are from the Android platform. Since the population of promoted Android apps is not sufficient for statistical study, we chose to focus on iOS apps. 54 iOS apps, which were promoted by “reviewfordev”, have been removed from Apple App Store, so we collected the meta data of 564 iOS apps with 497,259 text comments and 462,896 reviewers.

For comparative study, we also randomly selected 179,353 apps from Apple App Store, which include 10% apps from each subcategory that has over 100 apps and all apps in each subcategory with less than 100 apps. We collected their meta data, including 9,399,014 text comments, and 6,722,558 reviewers.

3.2 Analysis of Service Providers

For all the service providers engaged in app store based promotion, we mainly study when their services started, how much they charge customers, how much they reward reviewers, and how many apps they have reviewed.

3.2.1 Service launch time

Service launch time is the time when service providers start to review apps. The first service was started in 2010 and then more and more service providers emerged in 2012 and 2013. On their websites, many of them mention “1 million apps in iTunes” as the incentive for using their services. With over a million apps in an app store, it is extremely difficult for a new app to acquire initial reviews, while with no or few reviews an app would be ranked far behind and look much less attractive to app users. Therefore, their review promotion services could help app developers booster up their apps out of this difficult time.

3.2.2 Service price and reward

Service price is the price of writing one text review along with a rating score. In Fig. 2(a), we depict the lowest price for one review from each service provider. The highest price is $12.5 from “buyklout” [12], and the lowest one is around $0.5 from “androidappreviewsource” [13]. Basically, buying more reviews will be given more discounts. For example, “appiness” [14] charges $9 (US dollars) for each review if buying 5 reviews, $8 for 10, $7 for 50, and $6 for 100 reviews.

The reward for writing a review varies from $0.5 to $5. For each qualified review, “appiness” and “reviewfordev” pay $5 and $4, respectively, while most others pay $0.5. Some service providers use other type of rewards. For example, “appredeem” pays 1000 points, which can be cashed out; “appwin” sets up a lottery for app reviewers.

3.2.3 Promoted App volume

Except “reviewfordev” [15], none of the service providers shows the total number of promoted apps on their websites. However, some of them recruit online reviewers, so we can join as a passive reviewer and monitor app review assignments for more than four months. As shown in Fig. 2(b), the most popular service provider is “bestreviewapp” [4], which has accepted 251 apps. Some other providers like “apprebates” [16](84), “getappreviews” [17] (38), “appredeem” [18] (34), “promodispenser” [19] (32), and “giftmeapps” [20](30), have less apps. For the rest three, only 3 or less apps were found. In total, we have found 473 apps promoted within the monitoring period.

Except service providers, probably no one knows the accurate count of paid reviewers (or accounts). However, some of them have offered review packages, which state the number of reviews they are able to deliver for one app. For app stores like iTunes, each user is only allowed to write one review for each app. Therefore, the maximum number of reviews they are advertising to deliver for one app could reflect the number of reviewers they can hire or the number of accounts under their control. From Fig. 3, we can see that “4xn” [21] has the largest number of reviewers (i.e.,1,000). The second largest one is “getappreviews” [17], which can deliver 500 reviews at a time. The others can deliver less than 200 reviews.

---

1 "reviewfordev" displays all the apps it has promoted, and some of these apps were promoted before our monitoring.
3.3 Analysis of Promotion Incentive

3.3.1 Promotion Incentives

To find out when and why a developer made his decision to start app promotions, we analyzed app events like version updates and rating changes that happened before the promotion.

First, we retrieved all the version update history from iTunes and identified the most recent version update before the promotion. Further, we calculated the intervals (in terms of days) between the version update and the promotion. The average interval is 39 days. Fig. 4(a) shows the distribution of all the intervals. Specifically, 47% promotions were launched within one week after a new version update, 62% within two weeks, 80% within 60 days, and 92% within 120 days.

Second, we studied apps’ weekly average ratings for four weeks before the promotion. As shown in Fig. 4(b), nearly half of average ratings are one star, 80% average ratings are below three stars, and 90% below four stars. That is, most apps had received low average ratings before promotion.

Third, we looked into the distribution of weekly reviewer quantity. As illustrated in Fig. 4(c), more than 80% apps have less than 10 reviewers in a week and more than 90% apps have less than 50 reviewers. That is, most promoted apps have few reviewers in weeks before promotion.

Statistically, low average ratings and few reviewers are the most likely motivation for promotion.

3.3.2 Promotion Completion Time

To avoid the notice of app store vendors, most service providers try to post reviews gradually rather than posting bulk reviews in a very short time. Another reason could be that promotion reviews are completed by human, which limits the speed. Nevertheless, most of the reviews are promised to be completed within one week (e.g., appiness, bestreviewapp, appreviews4sale), or two weeks (e.g., appreviewsource). For example, as shown below, appiness mentions its strategy in its FAQ page.

Q: How long does it take to complete the reviews? A: It’s not a precise science because people are involved. Reviewers agree to review an app within 72 hours of receiving the request. Some may do it immediately, some may do it after a day and some may push it to the limit. However, we like this 72 hour window because this ensures that the downloads and reviews looks natural.

Q: I’d like to order 100 or more reviews but I want it to be spread out over time?
A: No problem. First, reviewers can take anywhere from 0 to 72 hours to review the app, so you will have a natural trickling in of reviews.

To confirm this, we studied the number of days needed to complete a review task. The quickest task was completed within one day, while the slowest one took up to 115 days. The average completion time was 21 days. As illustrated in Fig. 5, nearly 77% promotions were completed within 30 days and nearly 90% promotions were completed within 60 days.

3.4 Analysis of Promoted Apps

For all the promoted apps, we collected their meta data and then studied their platforms, average ratings, total number of reviewers, category distributions, and developers. For the purpose of comparison, we also randomly chose the same number of iOS apps that have at least one text comment.

All the promoted apps are either on iOS (96%) or on Android (4%). Even though both app stores have more than 1.3 million apps, iOS app developers were reported to have 85% more revenue than Android apps [23]. This could motivate more developers to promote their iOS apps than Android apps. As promoted Android apps are too few, next we only focus on iOS apps.

The average ratings from reviewers who posted text comments are shown in Fig. 6. Over 90% promoted apps have positive (i.e., >3 stars) average ratings and 80% average ratings are above 4 stars. Compared to randomly chosen apps, promoted apps have much higher average ratings. To test whether their difference is statistically significant, we

---

Google Play does not provide version update events on its website, so we only focus on iTunes.

This website has been closed very recently.
conducted a Welch’s t-test with the null hypothesis that the mean of promoted apps’ average ratings are identical to that of random apps’ average ratings. The test returned a small p-value (p-value < 0.05), which denied the null hypothesis and supported the hypothesis that promoted apps have significantly larger average ratings than random apps.

We also depict the total number of reviewers. From Fig. 7 we can see that promoted apps have more reviewers than random apps. Their difference was also tested to be significant by Welch’s t-test.

To study the category distribution, we count the number of apps in each category. As an iOS app could be assigned multiple categories, the app would be counted in more than one category. As shown in Fig. 8(a), the top first category is “Games” following by “Newsstand”, “Photo & Video”, “Lifestyle”, and “Productivity”. Compared to random apps in Fig. 8(b), the category distribution of promoted apps is nearly the same. However, promoted apps in each category (except “Social Networking”, “Medical”, and “Navigation”) have more paid apps than random apps have.
For promoted iOS apps, we collected all the reviewers who posted text comments. In this section, we study their behaviors in terms of co-rating and average rating. For comparative study, we randomly chose the same number of reviewers from Apple App Store.

Even though we cannot tell which reviewers were recruited, those who have reviewed multiple promoted apps after promotion started are suspicious. Given the set of 564 promoted apps, we first collect the number of reviewers who reviewed them, and then calculate the number of reviewers who have reviewed a certain number of promoted apps. In Fig. 9, we can observe that while the most number of reviewers reviewed only one promoted app, many reviewers have rated over tens of promoted apps. There are even two reviewers who reviewed 246 promoted apps. In contrast, for random (i.e., randomly chosen) apps, none of their reviewers have reviewed more than three apps in the chosen set (we do not show a figure here). Therefore, active reviewers of promoted apps are likely recruited reviewers.

![Figure 9: Number of reviewers (y-axis) who have reviewed certain number of promoted apps (x-axis)।](image)

Further, we collect those reviewers (also called suspicious reviewers) who reviewed at least two promoted apps during monitoring time. In total, we find 1,363 such reviewers. We then count the total numbers of apps in the app store they have reviewed, including apps that are not in our known promotion app set. In Fig. 10, we list all such reviewers in x-axis based on the number of reviewed apps (in the decreasing order). On average, each suspicious reviewer reviewed 9 apps, while the highest number is close to 800. We can also observe that the first 200 suspicious reviewers all have reviewed at least 100 apps in the app store. In contrast, less than 30 random reviewers have reviewed at least 100 apps in the app store. Therefore, overall suspicious reviewers have reviewed significantly more apps than random reviewers.

![Figure 10: Number of apps reviewed by one reviewer.।](image)

Furthermore, we study the average ratings provided by reviewers. For suspicious reviewers, we chose those who had reviewed at least two and at least six promoted apps. While for random reviewers, we randomly chose those who had reviewed at least two and at least six apps in the app store. As illustrated in Fig. 11, for those who reviewed at least two apps, nearly 1,200 (88%) suspicious reviewers, but only 800 (59%) random reviewers, have average ratings higher than 4 stars. In the case of six apps, nearly 400 (67%) suspicious reviewers, but only 200 (33%) random reviewers, have average ratings higher than 4.5. Therefore, average ratings from suspicious reviewers are significantly higher than that from random reviewers in both cases. This observation is also supported by a Welch’s t-test.

![Figure 11: The average ratings to all apps from reviewers rated different number of apps.।](image)

### 3.6 Promoted App Tracer

Joining as a recruited reviewer to monitor review assignments, we have collected many promoted apps from several service providers. However, these apps are probably only a small part of all promoted apps. This is because this kind of promotion services has been launched for years, while AppWatcher only monitored them for four months. That is, we were unable to directly discover apps promoted before our monitoring. Besides, some service providers may recruit their reviewers through other channels like invitation only membership and phone call groups. Nevertheless, the good news is that the entire review history of each reviewer has been recorded and stored in app stores. Next, we present a simple approach to demonstrate how to leverage our knowledge of promoted apps as the ground truth to find unknown promoted apps in app stores.

Intuitively, if a reviewer has reviewed many apps flagged as abused apps, he is probably recruited and his reviews to the unflagged apps also become suspicious. Further, if an app has received many such suspicious reviews, it becomes highly suspicious. Specifically, on one hand, given a set of promoted apps, reviewers who have reviewed more known promoted apps are more likely to be a recruited reviewer. In our previous comparison between reviewers of promoted apps and reviewers of randomly chosen apps, we observed that no reviewers from randomly chosen apps have reviewed
more than three apps from the app set. However, reviewers of promoted apps have reviewed far more apps (up to 246). Those reviewers who have reviewed many promoted apps and given high rating scores are highly likely recruited reviewers. On the other hand, given a set of recruited reviewers, the apps rated by many of them are also highly likely promoted apps. By following down the chain from known promoted apps to suspicious reviewers, and then to other promoted apps, we are able to expand the set of potentially promoted apps and finally get a suspect list. This suspect list can be further examined and verified by app store vendors with other additional evidences such as credit card numbers bound with each review account, IP address of submitting a review, etc.

Based on these observations, we design an iterative algorithm to trace apps promoted by recruited reviewers. As illustrated in Fig. 12, our algorithm starts from a known promoted app set (e.g., App Set I), which are the promoted apps we previously collected. Then, it inspects each reviewer of these promoted apps and labels reviewers who have reviewed at least \( N_a \) promoted apps (e.g., \( N_a = 5 \)) as recruited reviewers (e.g., Attacker Set I). Among all the apps reviewed by these reviewers, those apps reviewed by at least \( N_r \) recruited reviewers (e.g., \( N_r = 25 \)) are then tagged as promoted apps, and they form a new apps set (e.g., App Set II). This process continues iteratively until no suspicious reviewers or apps can be found.

Figure 12: Our promoted app tracer that starts from a set of promoted apps.

We have implemented this algorithm in JAVA and applied it to search for suspicious apps in Apple App Store. By setting \( N_a \) and \( N_r \) to 5 and 25 in our experiments, respectively, we have reported additional 2,410 suspicious apps (not including the previous 546 iOS apps collected by AppWatcher), along with 157,502 potential recruited reviewers.

For the lack of ground truth and additional information (e.g., reviewer account information, IP address) for validation, we are unable to verify the above results (i.e., suspicious apps and reviewers) one by one. Instead, we study the overall distributions of these suspicious apps and reviewers and their differences with randomly chosen apps and reviewers. If they have similar distributions as the known promoted apps (Section 3.4) and recruited reviewers (Section 3.5), they are very likely to be promoted apps and recruited reviewers.

We randomly chose 500 apps reported by our tracer, and studied their average rating distribution. As illustrated in Fig. 13, the average ratings of these suspicious apps are much higher than that of random apps. Their difference is significant, supported by Welch’s t-test (p-value < 0.05). On the other hand, compared to Fig. 6, the average rating distributions of known promoted apps and suspicious apps are nearly identical. That is, these suspicious apps have similar rating characteristics as known promoted apps, which indicates highly likely they are also promoted apps.

Figure 13: Average rating of suspicious apps.

Likewise, as shown in Fig. 14, the average ratings of suspicious reviewers are also much higher than that of random reviewers. Compared to Fig. 11, the average rating distributions of suspicious reviewers are similar in both scenarios. In other words, the suspicious reviewers reported by our tracer have the similar rating characteristics as recruited reviewers. Hence, they are likely recruited reviewers.

Figure 14: Average rating provided by suspicious reviewers.

Note that the purpose of this basic app tracer is to provide a generic framework and demonstrate how to utilize known promoted apps and basic heuristics to detect other unknown promoted apps. The criteria and suggested heuristics are not fixed or perfect, which can be extended to include other evidences like correlation between app’s average ratings and review quantities [10]. We also note that during the traversal process some popular apps might be accidentally added to the app set. As a result, some honest reviewers would be mistaken as suspicious reviewers, which would in turn bring more popular apps into the suspicious app set. Clearly, by excluding popular apps (e.g., those in top-200 rank charts for at least a few weeks) in our algorithm, one may filter out such noise. Also, one may increase the values of \( N_a \) and \( N_r \) when running our algorithm to reduce the chance of normal apps being included. After all, compared to over 1.2 million apps (as of June 2014) in iTunes store, with the current parameter setting, our basic tracer output 2,410 apps for fur-
ther investigation, which only accounts for 0.2% of all apps. With additional information such as reviewer credit card number, IP addresses, purchasing history, origin of country, app store may accurately pinpoint promoted apps among this short list.

4. WEB PROMOTION

Some service providers help developers advertise apps on their websites and charge for writing and publishing reviews. However, such paid reviews are forbidden by app store vendors like Google and Apple. Moreover, according to FTC [8], paid reviews without explicitly stating they are paid are illegal. In this section, we study these paid web promotion service providers, apps they have promoted, the objectiveness of their reviews, and promotion incentives of developers.

4.1 Dataset

We have found 31 service providers which charge developers for app reviews. For each service provider, we manually collected its meta data like service launching time, app volume, service price from its website.

Then, AppWatcher was started to monitor these websites everyday. It has found 29,680 promoted apps by July 15, 2014. For those apps with app ids shown on the websites, it further collected their meta data from the corresponding app stores. The meta data includes such information as total number of raters, distribution of ratings, average rating, category, version changes, developers, etc. It also collected all of the reviews along with review date, reviewer identify, rating score, targeted app version, review content. In total, we collected meta data for 7,077 apps, 5,349 developers, along with 5,290,281 reviews.

4.2 Analysis of Service Providers

For all the 31 websites we have monitored, we analyze when they started, what their service prices are, and what the market revenue is.

![Figure 15: App volume and service price of each service provider](image)

(a) App volume (b) Service price

4.2.1 Service launch time

Service launch time is the time when a service provider started. Most websites were launched in 2008, 2010, and 2011. On July, 2008, Apple opened the app store iTunes [1]. On March 6, 2012, GooglePlay\(^5\) launched android app store. Right after these openings, six service providers started their services. On September, 2010, GooglePlay expanded its service to 29 additional countries [2], and Apple App Store reached 10 billion downloads on January 22, 2011 [1]. Clearly, the quick increase of app market and its extremely huge user group have not only attracted many developers, but also quickly motivated some people to provide mobile app review services.

4.2.2 App volume

The volumes of promoted apps on different websites are displayed in Fig. 15(a). The first four websites have reviewed more than 3,000 apps, while the other websites have reviewed less than 3,000 apps. The number of apps on the first four websites accounts for 55% of the total number of promoted apps.

4.2.3 Service price

Service price is the price of a single app review posted on a service provider website. As illustrated in Fig. 15(b), the highest service price is from “androidheadlines” [24], which is $250 for one app review. Five websites charge customers about $150 and the others charge less than $100. Their prices could be determined by factors like pageviews, social media followers. For example, “androidheadlines” claims to own 350,000+ followers in the social network. As for the second highest one (i.e., smartappsforkids), it has 120,000 followers.

4.2.4 Market revenue and market trend

Based on the minimal service price and app volume in each website, the market revenue for web promotion is estimated to be at least $2,562,021.

The growth trend of the entire market is illustrated in Fig. 16. The monthly app volume has seen a great increase in early 2011. At around May 2012, it suddenly increased to over 900 apps. As for the monthly average rating, it first gradually increased from 3 star to 4 star from August 2008 to December 2010. Then, it varied between 3 star and 4 star for some time. Since 2012, it has been stable at the 4.5 star level. Overall, both the monthly app volume in the web promotion market and the average rating of promoted apps have increased in recent years.

![Figure 16: Monthly changes of app volume and average ratings](image)

4.3 Analysis of Promoted Apps

The promoted apps are mostly iOS apps, which account for 93.4%. The second most promoted apps are Android apps and they account for 6.4%. Other platforms, including Windows Phone, BlackBerry, Kindle Fire, only account for 0.16%. The most likely reason is that iOS apps make the most revenue, leading Android apps by (85%) [23].
The total number of reviewers of an app.

As the total number of apps in Windows, Kindle Fire, BlackBerry, is much smaller, we only analyze promoted apps based on iOS and Android. For all the promoted apps, we count the number of reviewers who provided both ratings and text comments. As illustrated in Fig. 17, over 70% apps have less than 500 raters and reviewers, and over 80% apps have less than 1000 raters and reviewers. Note that for random apps (i.e., randomly selected iOS apps), shown in the same figure, over 95% apps have less than 500 raters. This percentage is much higher than that of promoted apps. Therefore, in general, promoted apps have more reviewers than random apps.

![Figure 17: Total number of reviewers of an app.](image)

The categorial distribution of promoted apps is depicted in Fig. 18. Here, the right y-axis shows the percentage of paid apps among all apps in each category. The top five categories of promoted iOS apps are Games, Newsstand, Books, Productivity, and Lifestyle. In all the categories, more paid apps than free apps have been promoted. While the categories of promoted apps (Fig. 18) are very similar to that of random apps (Fig. 8(b)), promoted apps have higher percentages of paid apps than free apps random apps have in each category. This is not surprising though. As paid apps generate revenue to app developers directly, developers of paid apps are more willing to pay for web promotion.

The number of promoted apps from the same developer is also studied. For all the 5,349 developers, 1,016 developers (19%) have promoted at least two of their apps while the remaining 4,333 developers (81%) promoted one app. The top three developers have promoted 177, 84, and 69 apps, respectively.

### 4.4 Analysis of Review Objectiveness

Nearly all the service providers claim that their reviews will be objective and they cannot guarantee any positive review. Only a few of them mention that they will send negative reviews back to the developers. To understand the objectiveness of their ratings, we collected editor’s ratings and further studied rating distribution.

From Table. 1, we can see that, 92% ratings by website service provides are positive ratings (i.e., 4 or 5 stars⁶). 5% ratings are neutral (i.e., 3 stars) and 3% ratings are negative (i.e., 1 or 2 stars). Compared to ratings of random apps (68% positive and 19% negative), editors’ ratings have far more positive ratings and less negative ones. Moreover, among editors’ ratings, 5-star ratings account for 60% whereas only 9% of random ratings are 5 stars. Overall, more editors’ ratings are positive and less are negative.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Random Apps</th>
<th>Promoted Apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 star</td>
<td>9%</td>
<td>60%</td>
</tr>
<tr>
<td>4 star</td>
<td>58%</td>
<td>32%</td>
</tr>
<tr>
<td>3 star</td>
<td>14%</td>
<td>5%</td>
</tr>
<tr>
<td>2 star</td>
<td>18%</td>
<td>2%</td>
</tr>
<tr>
<td>1 star</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Further, in Fig. 19, we depict the proportion of positive ratings from each website⁷. For the first few websites, nearly all the ratings are positive. Only three websites have positive ratings below 50%.

![Figure 18: App category distribution (iOS)](image)

**Figure 18:** App category distribution (iOS)

**Figure 19:** The proportion of positive ratings in each website

### 4.5 Time for Promotions

For developers, the ultimate goal of using web promotion is to make more revenue, but what is the time for them to decide to use promotion? In this section, by studying the intervals between web promotion and app version update, weekly average rating, and weekly rater quantity before the promotion, we can get the answer and further infer the stimulus for promotion.

From Fig. 20(a), we can see that 50% promotions happened within one month of a new release and over 80% happened within 100 days. Therefore, one to three months after

⁶We have converted different rating scales to five star rating.

⁷Only those websites providing ratings along with text comments are included.
a new version release have been the common time for web promotion.

Further, in Fig. 20(b), we depict the CDF of the average ratings in four weeks just before the web promotion. As we can see, over 14% apps have average rating of 1 star, 52% apps less than 2 stars, and up to 73% less than 3 stars.

We also studied the weekly number of raters in four weeks before the promotion. As illustrated in Fig. 20(c), nearly 85% apps have less than 3 raters in each week, 90% apps less than 7 raters, and 95% apps less than 21 raters.

The above study shows that these promoted apps have experienced low average ratings or few raters in the weeks before the promotion. Our conjecture is that after a developer released a new version of his app, normally he waited for one to three months, hoping his app would become popular and get good reviews. However, when observing both low rating and few reviewers for his app, he decided to seek help from web promotion.

### 4.6 Cross Promotion

As the main channels of promotion on the underground market, app store promotion and web promotion share many common features. For example, they all charge developers and write positive reviews; a few service providers dominate the market and promote a lot more apps than others; the promoted apps usually experienced few reviewers or low average ratings before promotion. Nevertheless, there are also essential differences. First, the sources of app reviewers are different. For app store promotion, service providers either manipulate lots of accounts or recruit reviewers online. But for web promotion, each service provider hires a few editors. Second, after releasing a new app, developers often wait longer before choosing web promotion than developers who choose app store promotion. One possible explanation is that app store promotion is useful to get initial reviews for a new release, while web promotion is for app advertisement. The latter does not necessarily result in more reviews.

Some developers might resort to both ways of app promotion. We cross checked all the apps promoted by app store promotion from July 1st to October 7th, 2014 and those by web promotion till October 7th, 2014. In total, we have found 98 such apps out of 546 apps (18%). As an app often has multiple versions, developers might promote different app versions in a different way. Among these cross promoted apps, 12 apps were promoted for the same version, while the other 77 apps were promoted for different versions. Moreover, 11 developers (owners) have cross promoted more than one app. The top three developers have cross promoted 8, 4, and 3 apps, respectively.

### 5. RELATED WORK

**Underground Market Study** Underground market is the center of connecting illegal merchants with special service providers including CAPTCHA solving systems [25], Twitter account spam [26], fake reviews, etc. These markets usually rely on some popular crowd-sourcing websites like Amazon’s Mechanical Turk, Freelancer. Amazon’s Mechanical Turk has been studied from various perspectives including worker demographics [27] [28], behavior research [29], marketplace [30]. Some other crowd-sourcing websites like Zhubaij [31], Sandaha [32], MicroBlogs [33] were also studied [34] from the perspective of campaigns on these websites. Freelancer was studied [25] from the role of freelance labor in web service abuse like account registration and verification, social network linking. Some forums was also studied in trading twitter spam accounts [26]. Unlike the above underground markets, app review market consists of service providers whose services only focus on writing mobile app reviews. They tried many sophisticated strategies to elude detection algorithms and some of them have built a reward system to recruit reviewers from the wild.

**Opinion Spam Detection** Opinion spam was first studied by Jindal [35] and various approaches have been proposed [36] [37] [38] [39] [40] to detect fake reviewers in traditional market like Amazon. Comment spam of blogs was also studied using spam structure [41], content based algorithm [42], language models [43] etc. Moreover, collusive reviewer groups are also studied [44] [45].

In the mobile app review market, opinion spam is even more serious than ever because of huge marketplace and centralized app distribution systems like iTunes, GooglePlay. However, due to the natural differences between mobile app markets and traditional markets in terms of review content, reviewer volumes, reviewer behavior, etc, previous approaches do not apply directly in mobile app market. So far very few approaches [10] [9] have been proposed to detect either collusive reviewers and fake reviews, respectively. One challenging issue with existing detection approaches is the lack of ground truth to verify their detection results. The outcome from this work can provide a dataset as the ground truth.

### 6. CONCLUSION

As the size of an app market (e.g., Google Play and iTunes) has exceeded a million of apps, it is very hard to make a new app known in an app market. This makes the underground market of trading app reviews increasingly popu-
lar. Review-based promotions in mobile app markets could discourage benign developers who try hard to improve app quality. In this paper, we first built a system called AppWatcher to monitor the app promotion market, collect promoted apps, and analyze facts such as service launch time, service price, app volume, possible promotion incentives. We then revealed the characteristics of promoted apps and suspicious attackers, which can serve as important heuristics to design automated and effective algorithms to identify promoted apps in the entire app store such as iTunes. The dataset collected by AppWatcher can also serve as the ground truth to evaluate the power of any app promotion detection algorithm.

7. ACKNOWLEDGEMENT

This work was supported in part by NSF grant CCF-1320605 and a Google gift. We also thank the reviewers for helpful comments.

8. REFERENCES

[38] G. Wang, S. Xie, B. Liu, and P. S. Yu, “Review graph based online store review spammer detection,” ICDM ’11.