The Art and Craft of Fraudulent App Promotion in Google Play

Mizanur Rahman*
Amazon, USA
mrahm031@fiu.edu

Nestor Hernandez*
FIU, Miami, USA
nestorghh@gmail.com

Ruben Recabarren
FIU, Miami, USA
recabarren@gmail.com

Syed Ishtiaque Ahmed
University of Toronto, Toronto, CA
ishtiaque@cs.toronto.edu

Bogdan Carbunar
FIU, Miami, USA
carbunar@gmail.com

ABSTRACT

Black Hat App Search Optimization (ASO) in the form of fake reviews and sockpuppet accounts, is prevalent in peer-opinion sites, e.g., app stores, with negative implications on the digital and real lives of their users. To detect and filter fraud, a growing body of research has provided insights into various aspects of fraud posting activities, and made assumptions about the working procedures of the fraudsters from online data. However, such assumptions often lack empirical evidence from the actual fraud perpetrators. To address this problem, in this paper, we present results of both a qualitative study with 18 ASO workers we recruited from 5 freelancing sites, concerning activities they performed on Google Play, and a quantitative investigation with fraud-related data collected from other 39 ASO workers.

We reveal findings concerning various aspects of ASO worker capabilities and behaviors, including novel insights into their working patterns, and supporting evidence for several existing assumptions. Further, we found and report participant-revealed techniques to bypass Google-imposed verifications, concrete strategies to avoid detection, and even strategies that leverage fraud detection to enhance fraud efficacy. We report a Google site vulnerability that enabled us to infer the mobile device models used to post more than 198 million reviews in Google Play, including 9,942 fake reviews. We discuss the deeper implications of our findings, including their potential use to develop the next generation fraud detection and prevention systems.

CCS CONCEPTS

• **Security and privacy → Social network security and privacy**; **Social aspects of security and privacy**;

KEYWORDS

Search Rank Fraud; Crowdturfing; Fake Review; Opinion Spam; App Store Optimization

*Both authors contributed equally to the paper

CCS '19, November 11–15, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6747-9/19/11…$15.00
https://doi.org/10.1145/3319535.3345658

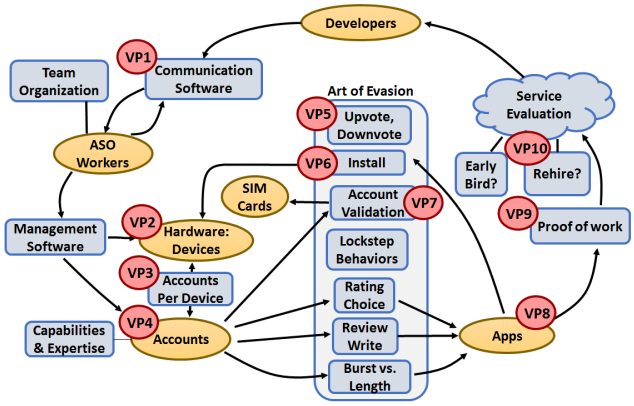


Figure 1: Map of discovered fraud workflow in Google Play. Orange ovals denote tangible participants and assets, blue rectangles denote several investigated capabilities, behaviors or strategies. Small red ovals represent fraud vulnerability points that we identified and discuss in § 6.

ACM Reference Format:
Mizanur Rahman, Nestor Hernandez, Ruben Recabarren, Syed Ishtiaque Ahmed, and Bogdan Carbunar. 2019. The Art and Craft of Fraudulent App Promotion in Google Play. In *2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19), November 11–15, 2019, London, United Kingdom.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3319535.3345658

1 INTRODUCTION

Popular online services that provide millions of users with access to products, news, social relationships and peer-opinions, are besieged by fraudulent behaviors, that skew public opinion and bias product reputation and popularity [25, 35, 45, 54, 75, 80, 91]. To reduce the effects of such behaviors, commercial peer-opinion sites employ proprietary solutions to detect and filter fraud, e.g., [22, 30, 34, 47, 56, 65, 67, 71, 79, 98]. Similarly, a substantial body of academic research has focused on the detection aspect of the fraud problem, and has proposed and used assumptions about the behaviors and capabilities of fraudsters, that are based on intuition, extracted from small datasets of fraud, or revealed by collaborators within commercial sites. While such previous efforts have revealed important insights into the operations of fraudsters, most have not been validated with empirical feedback from the actual perpetrators.

In an effort to address this limitation, we first performed a structured interview study comprised of 118 questions, with 18 Black Hat App Search Optimization (ASO) workers that we recruited

from 5 freelancing sites, concerning fraud that they post on Google Play. Second, we performed a quantitative investigation with data that we collected from 39 other ASO workers recruited from the same sites. The data includes 1,164 Google Play accounts that the 39 ASO workers revealed to control, and 21,767 fake reviews posted from these accounts for 6,362 unique apps. Further, we identified, and report a Google site bug that enabled us to infer the mobile device models used to post 198,466,139 reviews for the 6,362 apps.

Based on the findings of our studies, we present the fraud workflow map of Figure 1, showing newly identified and previously explored fraud capabilities, behaviors and detection avoidance strategies. Specifically, we report multiple, novel insights into the working patterns of ASO workers, including that they (1) pool in physical, brick-and-mortar offices, friends-and-family organizations, and online teams, (2) have either a well-articulated role and are salaried on a regular basis, or are part of unstructured teams and share earnings, (3) have access to many user accounts, of both sockpuppet (fake) and organic (controlled by real users) types, (4) have access to large and diverse stocks of low to high-end, and new to old mobile device models, (5) flexibly outsource work when their number of accounts or device models are insufficient, and (6) implement interactive work verifications, and punish cheaters.

Further, our studies provide evidence that supports several observations and assumptions made by previous fraud detection work, about, e.g., the emergence of organic fraud [50, 52, 101], the timing of fraud [38, 53, 61, 63, 64, 96, 100], the fake review writing process [38, 42, 51, 52, 60, 61, 63–65, 74, 76, 96, 97, 99] and the choice of ratings [24, 51, 52, 63–65, 74, 93, 94].

However, we also report and validate concrete, participant-revealed behaviors that do not fit the mold of assumptions made in previous work, including lockstep behaviors [32, 48, 59, 77, 81, 86, 93, 94, 97, 100] or posting reviews in bursts [26, 28, 32, 36–38, 40, 42, 43, 51, 52, 59–61, 63, 65, 92, 95, 96, 100].

We also found and report participant-claimed techniques to bypass Google-imposed verifications, e.g., user account validations and review-posting sanity checks, and even strategies to leverage Google's fraud detection mechanisms to improve fraud efficacy, e.g., downvoting negative reviews to trigger their removal, or using singleton accounts that exploit detection cold-start problems.

Finally, and importantly, we identify several vulnerability points in the fraud workflow, and propose defenses that exploit them. In summary, we introduce the following contributions:

- **ASO worker studies**. Present empirical data from actual ASO workers, to advance our understanding of their work, through interviews and a quantitative analysis of gold standard fraud data [§ 4].
- **ASO worker capabilities, behaviors and strategies**. Report new findings on the capabilities and behaviors exhibited by ASO workers [§ 5]. Provide evidence that supports several observations and assumptions made by previous detection work. Report and validate concrete strategies to avoid detection, including departures from existing assumptions. Build a first map of the Google Play fraud workflow.
- **Google Play vulnerabilities**. Identify and report a bug that can be exploited to collect device model information from reviews [§ 4.2]. Report Google Play verifications claimed to be ineffective by participants. [§ 5].
- **Impacts**. Identify vulnerability points in the fraud workflow and discuss their potential to advance fraud detection and prevention work [§ 6].

2 RELATED WORK

Social network fraud studies have focused on the identification of fraud and its effects on social network users. For instance, Thomas et al. [83] identified 1.1 million accounts suspended by Twitter and studied the behavior and lifetime of spam accounts, the campaigns they execute, and the wide-spread abuse of legitimate web services. Thomas et al. [85] have further investigated fraudulent Twitter account markets to monitor prices, availability, and fraud perpetrated by 27 merchants over 10 months. Stringhini et al. [82] studied Twitter follower markets by purchasing followers from different merchants, and discovered patterns and detected market-controlled accounts in the wild. Critical operational details of the fraud market have remained however mostly unstudied. Our work seeks to address this, by both documenting and validating operational procedures of ASO workers who target Google Play.

De Cristofaro et al. [36] studied page "likes" from Facebook ads performed by "like" farms using honeypot pages, and analyzed them based on temporal, social and demographic characteristics of the likers. McCoy et al. [62] studied the business model of "online pharma", using ground truth datasets of transaction logs. Further, Springborn et al. [78] purchased fraudulent traffic for honeypot websites and analyzed the underlying pay-per-view networks and their delivery methods. In comparison, we conduct an interview study to directly engage and seek insights from fraud perpetrators, then support them through an analysis of empirical fraud data.

Bursztein et al. [31] observed that manually hijacked Google accounts exhibited activity at a tight daily schedule, homogeneous daily time table and similar tools and utilities used in parallel on different victims from varying IPs. While none of our participants claimed to use hijacked accounts or to have been the victim of such attacks, some participants did claim and exhibit lockstep behaviors in their use of accounts. We further note that our study is more general, as it concerns the entire fraud workflow.

Other similar studies have different goals. To highlight the methods and prevalence of scammers, specific to Nigeria, Park et al. [69] collected three months of data using an automated system which posts honeypot ads on Craigslist, and interacts with scammers. Portnoff et al. [73] used NLP and ML-based methods to determine post type, product and price on cybercriminal market offerings. Further, Wang et al. [88] used empirical crawled data to identify SEO campaigns and documented their impact on promoting search results for several luxury brands. In contrast, in our fraud study, we seek to also identify (1) Google Play vulnerabilities that fraud workers found and exploit, (2) evolutions in fraudulent behaviors to avoid detection, and (3) their intrinsic weaknesses, to be exploited by the next generation fraud detection solutions.

3 BACKGROUND

We consider peer-opinion app markets, e.g., Google Play [9], who host accounts for products, developers and users. Developers use their accounts to upload apps and information about them. User accounts enable users to establish an online identity, search for, install and review apps. A review consists of a star rating (1–5)

and text, and also includes the profile photo and name of the user account from which it was posted.

Search rank fraud and crowdsourcing. The search rank of apps has significant impact on the returns made by their developers. Thus, developers have incentives to maximize their app's visibility. In this paper we focus on developers who attempt to engineer the search rank of their apps by hiring specialized, online Black Hat ASO workers, to perform review and install count manipulation. Developers and fraudsters connect through several sites, that include general-purpose crowdsourcing sites [7, 19, 39], specialized fraud sites [1, 2, 4, 16, 18], and social networks (e.g., Facebook groups).

Fraud detection and defenses. Online systems implement a suite of fraud detection and defense mechanisms [66, 79, 98]. For Google Play, such observable mechanisms include:

- **Account validation**. Request users to prove control of a mobile phone, e.g., by providing its calling number, then retrieving a code sent to it through SMS.
- **Install-then-review**. Users can review an app only if they install it first [20].
- **Filter fake reviews**. Detect and remove reviews suspected of being fake.
- **Close fraudulent accounts**. Identify and close user and developer accounts suspected of behaviors that violate the site's terms of service.

4 METHODS

Our study involves both a qualitative exploration of and a quantitative investigation into various aspects of fraud production. In this section we describe both studies.

4.1 Qualitative Study

The qualitative study of our work is comprised of in-depth interviews with 18 ASO workers. We recruited participants from several Facebook ASO groups, and also Upwork [19], Fiverr [39], Zeerk [21], and Peopleperhour [14], all popular among ASO workers. We identified 560 such workers, and invited them to participate in our study through the 1-on-1 communication services of the corresponding sites. We include the recruitment message in the auxiliary document.

72 of them responded to our invitation. To select participants who are actively involved in ASO jobs, we asked the responders, 3 questions, all for Google Play: (1) "how many accounts do you control?", (2) "for how long have you been actively doing ASO?", and (3) "on how many ASO jobs did you work, approximately?".

We identified 25 participants who control at least 100 accounts on Google Play, have been active for at least 1 year, and have completed at least 100 ASO tasks. Following recruitment, and before starting the interview, we read to these participants the introductory script included in the auxiliary document. 18 of them (all male, 19-29 years old, located in Bangladesh(13), India(4) and New Zealand(1)) agreed to participate.

In the following, we refer to the interview participants as P1, .., P18. With these participants, we conducted a structured interview study that had 46 questions, with additional 72 questions for clarifications, see auxiliary document. The questions range from demographic information to workflow, and from the devices used to the operational methods employed. We conducted the interviews over Skype, between August and October, 2018. Interviews lasted from 33 to 66 minutes (M = 46.38, SD = 12.34). We paid a rate of 5 USD for every 15 minutes a participant spent in our interview. We audio recorded the interviews with the participant permission, then transcribed and anonymized the data.

We analyzed the anonymized data using the Grounded Theory method [33]. We used open coding to identify 169 unique codes, including both abstract and concrete labels. Two members of our team independently coded the data. The inter-coder agreement was 84.61%. In the cases where codes of the two coders did not match, a discussion was held with a third member of our team, to decide the final code. We used axial coding to relate the generated codes, and ended up with 22 categories grounded in the collected data. Some of the categories are: account blending, account creation, devices, early-bird fraud, extreme reviews, strategy, etc. We have then further refined our categories into the codes that form subsection titles in § 5.

4.2 Quantitative Investigation

We performed a quantitative investigation with user accounts collected from 39 ASO workers, different from the qualitative study participants, but recruited using the same methods described in $ 4.1. In the following, we refer to the quantitative study participants as F1, .., F39. Each of the selected workers claimed to control up to 500 Google Play accounts ($M = 211, SD = 166$), and each shared the IDs of at least 15 Google Play accounts that they control. This yielded a total of 1,164 account IDs for analysis.

We then crawled the 6,362 unique apps that the ASO workers reviewed using those IDs, and that were available in Google Play. These apps had received 21,767 reviews from the 1,164 worker-controlled accounts, and a total of 218,167,727 reviews. We used the AppBrain API [3] to collect the category and release date of each app.

Device model data collection. We have collected information provided by Google Play about the devices used to post fraudulent reviews. Google Play's client-side enforced functionality, allows an authenticated user to filter reviews according to the model of her registered devices. We used this functionality to query the reviews posted for an app, for all possible device models, and thus identify the device model used to post any individual review. We used the list of 21,597 Google supported devices [10], that contains the parameters that we needed to identify the device models used to post the above 21,767 reviews, posted from the 1,164 ASO worker-controlled accounts, as perceived by Google's systems. In addition, we collected the device release date and price (in EUR) from GSM Arena [12] and Gadgets360 [8].

4.3 Ethical Considerations

Some ASO work is considered unethical according to several ethical frameworks, and many ASO workers belong to low-paid vulnerable groups. This is why our study took utmost care to follow the best ethical practices for conducting sensitive research with vulnerable populations [29]. Our study had a very clear declaration of the researchers' identity, research objective, and potential impact on the participants' work without following any sort of deception. The whole study procedure was scrutinized and approved by the institutional review board of a major North American university

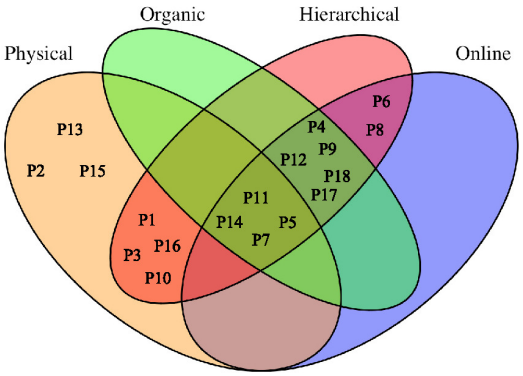


Figure 2: Venn diagram of participant categories, reveals diversity and complexity of fraud organizations. Participants are part of teams that are either (1) physically co-located or online, (2) hierarchical or flat, and (3) sockpuppet account based or organic.

(IRB-18-0077@FIU). We include our recruitment message and introductory script in Appendix A. We include a discussion of the process of our recruitment, the possible reasons for our participants to respond, and other relevant issues, in the auxiliary document.

We used GDPR [70] recommended pseudonymisation for data processing and statistics, and other generally accepted good practices for privacy preservation. After data collection, we have deleted all device-to-identity links and only generated statistics that allowed us to validate our assumptions. We have avoided obtaining additional information about the devices used or the accounts involved. We have contacted Google about our discovered device model identification issue, through Google's vulnerability reward program (VRP) [11] (issue: 119676181). Google has accepted our finding and has invited us to join their hall of fame.

5 FINDINGS

We organize, analyze and report findings from the interview and quantitative studies. Figure 1 provides a map of the topics that we investigated.

5.1 Team, Location, and Organization

All the 18 interview participants claimed to be part of organizations dedicated to posting fraud in Google Play. Our data shows that ASO workers assemble in various organizational structures. While some of them work in a team where each person has a well-articulated role and they are salaried on a regular basis, many of them work in a more unstructured team and the whole team share their earnings. We classify ASO teams into several categories, based on their location, organization type, the type of fraud, and profit sharing structure. Figure 2 shows the Venn diagram of the 18 participants grouped according to 4 of these categories, for readability.

Team size. The first column of Table 1 lists the team sizes claimed by each participant for their organization, including both physically co-located and online team members. 5 participants claimed to work alone. The other 13 participants claimed to have a team with at least 10 members. Notably, P4 claimed to be part of a big company with around 150 people in their team, who organize 15,000 organic ASO workers through virtual (WhatsApp, Facebook) groups.

Figure 3: Photo taken by participant P10 in our study, with the premises and (anonymized) employees of his business. Photo reproduced with permission from the participant.

Physical co-located vs. online teams. Seven participants (Figure 2) claimed to work with a physically co-located team. 5 of them claimed to have brick and mortar offices. Figure 3 shows a photo taken by P10, with the premises and employees of his fraud team. 7 others claimed to have strictly online teams. The remaining 4 claimed to be a part of hybrid organizations that (1) are a physical team, including working alone with their own devices and accounts, and (2) have access to online ASO workers. Notably, P18 said (1) *"I run a mobile repair shop. I use the devices that I get to repair."* and (2) *"I share the link in my group and they review it."* P11 said *"I use two types of accounts, my friends and family, and my own 100 accounts."*

Organization structure: hierarchical vs. flat. 15 participants claimed a hierarchical structure of their organizations (Figure 2). 11 of them described specific roles in their organizations, that include *job managers*, who interface with the developers and manage work from the marketplace, *team admins*, who organize, distribute tasks, and verify the work of *review posters*, and *new account creators*. For instance, P3 said *"I am one of the admins in our team and we have 10–12 admins. Under each admin, we have 15–20 members. All admins work as subcontractors, and some of our other team members work with the developers and manage work from the marketplace."* However, 2 participants claimed to work in teams with a flat organization. For instance, P15 said *"We all work together. There is no hierarchy."*

Organic fraud. 9 participants claimed to organize or be part of online teams of "organic" users, workers who use their personal accounts to post fake reviews (Table 1). P5 said *"I also have my own Facebook group where I have combined 60 real users to write reviews."* P7 did not specify the number of organic accounts that they can access, but stated *"we have 3,000 accounts. If we need more we run CPI/CPA campaign where people get an incentive to install apps."*

Profit sharing. One participant claimed to pay team members a monthly salary, while another one claimed an even split among members. Three of them mentioned preferential cuts for the job manager (10–25%) and team lead (10–50%) and equal split of the rest among the actual review posters. Two participants claimed a flat rate for the review posters ($0.40 per review). The rest of the participants did not respond to this question.

Summary. Our study thus confirms observations made by existing work, that fraud is perpetrated by experts who control either (1) many sockpuppet user accounts, e.g., [28, 37, 55, 59, 60, 63, 64, 77, 93, 95, 99] or (2) *organic fraudsters*, i.e., real account owners

		Accounts		Devices		
P	Members	Organic	Inorganic	Mobile	Laptop	Online
P1	40	0	15,000	300	0	0
P2	12	0	300	40	0	0
P3	195	0	1,500	200	0	0
P4	150	15,000	0	0	0	15,000
P5	12	100	0	0	0	60
P6	1	0	1,500	0	0	500
P7	50	N/A	3,000	1,000	0	0
P8	35	0	150	0	0	100
P9	15	400	0	0	0	450
P10	30	0	450	30	35	0
P11	1	200	100	45	0	200
P12	1	500	0	0	0	500
P13	13	0	80,000	13	13	0
P14	34	5,000	0	0	0	5,000
P15	10	0	300	50	0	0
P16	50	0	500	70	0	0
P17	1	1,000	0	0	0	1,000
P18	1	500	30	30	0	500

Table 1: Number of team members, and of accounts and devices claimed by the 18 interview participants.

recruited online [13, 17]. Our study also provides concrete numbers and extends the existing literature by adding that (1) ASO workers can be hybrid (e.g., both organic and sockpuppet masters) and (2) product developers can hire multiple types of expert ASO workers to promote their products.

Participants claimed to charge between $0.5 and up to $6 per posted review (M = 2.16, SD = 1.86), and to have between 1 and 6 years of experience in ASO jobs (M = 3.03, SD = 1.53). During this time, they claimed to have worked on between 150 and 4,000 apps in total, and between 6 and 50–60 apps in the past month (M = 34.11, SD = 18.37). They also declared a diverse educational background, including 2 masters degrees, 11 completed bachelor degrees, 2 ongoing bachelors, and 4 high school graduates.

5.2 Fraud Capabilities and Expertise

The middle columns of Table 1 list the number of user accounts claimed to be controlled by or accessible to each of the 18 participants. Most participants control a few hundred accounts, however, a few control or have access to several thousands: P13 claimed to be part of a team of 13 workers who control 80,000 accounts.

7 participants, each claiming to control thousands of accounts, also claimed to be able to write an "unlimited" number of reviews for a single app, i.e., more reviews than the developer can ask or afford (as inferred from the participant's past experience). The other 11 participants, with up to 3,000 accounts, claimed to be able to write a number of reviews that was consistent (i.e., smaller or equal) to the number of accounts they previously claimed to control.

To provide perspective on several of these claims, Figure 4 shows the number of accounts revealed, and the number of unique apps reviewed from those accounts, by each of the participants in our quantitative study (§ 4.2). In total, we have crawled information from 1,164 accounts and the 6,362 unique apps that were reviewed from these accounts. Even in this limited gold standard dataset, one participant (F18) was able to reveal 83 accounts that he controls, and F35 has reviewed 927 unique apps from his 42 accounts.

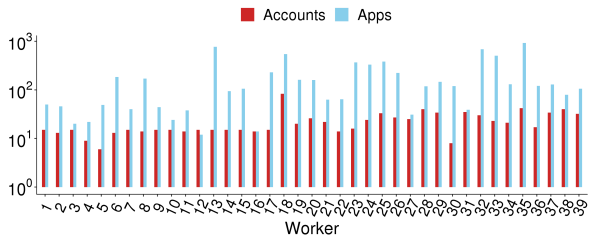


Figure 4: Number of accounts revealed by F1,..,F39 and number of apps reviewed from them. F18 revealed 83 accounts. 14 workers have reviewed at least 150 apps from the revealed accounts. F35 has reviewed 927 apps!

5.3 Hardware: Devices

All the interview participants claimed to own or have access to multiple mobile devices. The last columns of Table 1 list the number of devices, organized by types, claimed to be controlled or accessible by each participant. 9 participants claimed to post fraud from mobile devices; 11 participants claimed this also happens from the mobile devices of organic ASO workers that they control. 2 participants said that they also post from emulators running in laptops, e.g., P13 claims to have 13 laptops and use the BlueStacks emulator [5] to install and review apps, and also 13 smartphones.

P8 and P18 have an almost 1-to-1 account-to-device mapping. Participants such as P2, P3, P7, P10 and P15, have a small but many-to-one mapping, e.g., up to 7 accounts per device. Others, such as P1 and P13, claim to have significantly more accounts than devices (e.g., 15,000:300 and 80,000:30 respectively).

Mobile device models. Several participants claim access to communities of organic users (see Figure 2), thus to a diverse set of devices. 4 participants (P1, P10, P11, P13) claimed to own only low-end, cheap devices. Others (P7, P15, P16) claimed to own a mix of low, medium and high-end devices, dominated by low-end devices. For instance, P7, who claimed to own more than 1,000 devices said that (1) *"we try to choose cheap devices with more features and memory,"* however (2) *"we also have high-end phones like Nokia, Samsung, which we need to review virtual/augmented reality apps"*.

Device source. Most participants claimed to purchase their devices on the regular market. However, P11 said, about his claimed 45 devices, that *"I have bought them from the black market with a very low price."* Further, as mentioned in § 5.1, P18 claimed to run a mobile device repair shop, and use the devices he is supposed to repair, to write reviews.

Device storage. 6 participants claimed to store the devices on a table, easily accessible. P1 claimed to store the devices in a separate room. P7 said that *"the department who handle reviews and installs is on a different floor, and high-end phones are kept in the locker after use for safety."* We also asked P7 about how they manage to charge 1,000 devices. He claimed that they have a dedicated team to manage all the devices, and charge a device every 2–3 days. Further, he claimed that they keep the devices on during office time, and switch them off after 11pm–midnight.

App-device compatibility issues. When asked about what they do when they need to promote an app that is not compatible with their devices, 9 participants (P5, P6, P8, P10, P11, P13, P14, P16, P17) said that it never happened. However, P7 said that he runs

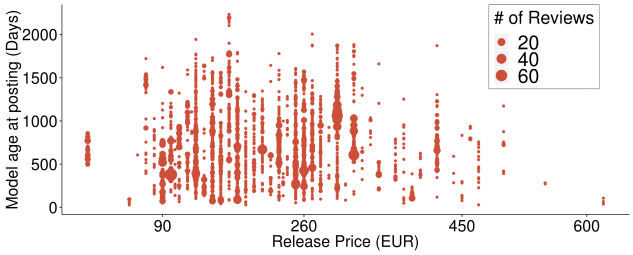


Figure 5: Scatter plot of device release price (EUR) vs. model age (Days) at posting time, for each of 9,942 reviews posted from 344 unique device types. Most devices are old and low-end (45.98%) or mid-end (31.41%), or fresh and low-end (15.31%). High-end and even free devices have been used!

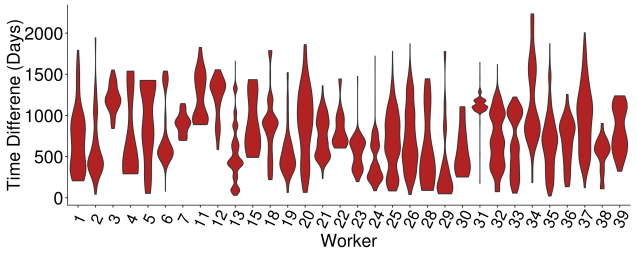


Figure 6: Per-worker distribution (violins) of the "age" of devices used to post reviews, i.e., the time difference in days between the review date and the release date of its posting device. Workers not shown had insufficient known device models. F3, F7, F11, and F31 use old devices. Most others (F1, F2, F13, F20, etc), use both newly released and old devices.

campaigns to recruit ASO workers who own compatible devices, or even purchase such devices. P9 and P15 said that they provide as many reviews as they can from their compatible devices, and contact the developer to explain the problem. P12 skips the job.

Quantitative Investigation. We used the technique described in § 4.2 to find 344 unique device models, used to post 9,942 of the 21,767 reviews written from the accounts controlled by the 39 participants. We found that 12 participants posted reviews from at least 20 different device models; F35 used at least 84 distinct device models. However, participants F9 (215 reviews), F10 (166), F14 (162), F16 (67), F17 (459), and F27 (197) have posted reviews only from devices of unknown models. We confirmed that the "unknown" device category includes reviews posted from Google Play's website interface and certain types of emulators.

Figure 5 shows the relationship between the device release price (in Euros) and the device model age at posting time, for each of 9,942 presumed fake reviews posted from 344 unique device models. We consider that a device is low, mid, or high-end, if its release price is in the range $[0, 260)$, $[260, 450)$, and $[450, \infty)$ respectively [87]. We classify a device model age into Fresh ($<$ 1 year), Middle-aged (12-18 months), and Old ($>$ 18 months). We found that 61.3% of reviews were posted from low-end, 38.2% from mid-end, and 0.5% from high-end devices, while 77.39% are from old and 19.66% from new models. Further, most of these reviews were written from old low-end devices (45.98%), old mid-end (31.41%) and fresh low-end (15.31%) devices.

A notable case is that of tablets given away (price 0EUR, leftmost points in Figure 5) by the Uruguayan government to students as part of an inclusion plan named Plan Ceibal [15]. Participants F25 and F32 used this device model to write 159 reviews for 137 apps. In addition, 3 reviews were posted from Galaxy S9+ devices whose price exceeds 600EUR (rightmost points in Figure 5).

Figure 6 shows the per-worker distribution of the "age" of their devices: the time difference between the review date and the device release date for all the fake reviews posted from known devices. 13 ASO workers have each posted at least 100 reviews from devices that are over 6 months old. Additionally, F13, F24, F25, F32, F33, and F35 have each posted at least 30 reviews from devices that are less than 6 months old. We conclude that different workers rely on stocks of either old devices, new devices or a mix of old and new, to post fraud.

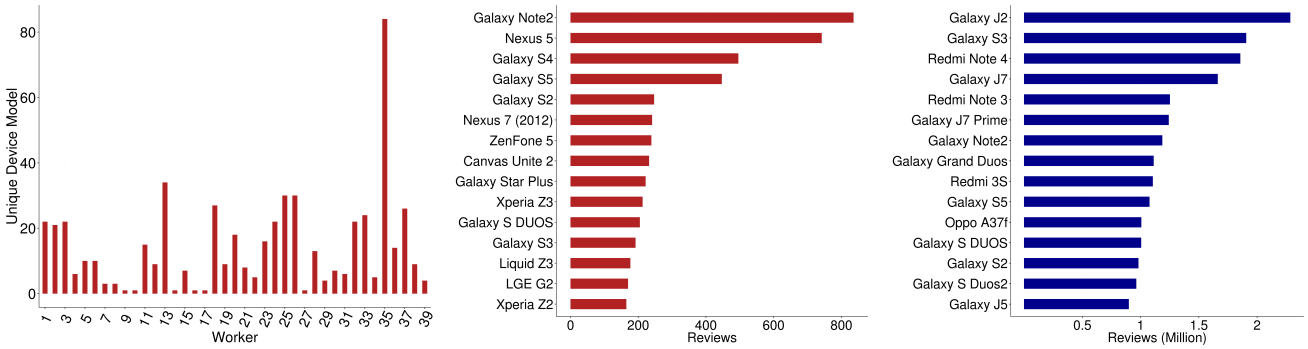
We found that 93.8% of the 9,942 fake reviews were posted from smartphones and 6.2% from tablets. Figure 7(a) displays the number of unique device models used by ASO workers, including the "unknown" category (i.e., not among the 21,597 officially supported device models provided by Google [10]). While F35 has used 85 unique device models, participants F9, F10, F14, F16, F17, and F27 have posted all their reviews from unknown devices. Figure 7(b) shows the popularity of device models used by the 39 participants, over all their 9,942 reviews posted from devices of "known" models. The top 6 most used devices by ASO workers to post these reviews are Galaxy Note 2 (836 reviews), Nexus 5 (742), Galaxy S4 (496), S5 (447), S2 (247) and Nexus 7 (241). Further, Figure 7(c) shows the popularity of the top 15 most popular devices, out of 11,934, that were used to post 198,466,139 reviews in Google Play.

Summary. We found ASO workers who claim to have access to large number of devices, either owned, or accessed through their communities of organic fraud. This claim is partially confirmed through our gold standard fraud data. Both in our interviews and in the quantitative study, we found that ASO workers have a diverse stock of low to high-end and new to old devices. Participants with many devices reported streamlined solutions to manage them, while those with fewer devices reported ways around cost limitations and compatibility issues, e.g., further outsourcing jobs.

5.4 Software

Team formation. 10 interview participants (P3, P5, P6, P8, P9, P11, P12, P14, P17, P18) said that they used Facebook and/or Whatsapp to create online teams. For instance, P6 said that *I have a Facebook group of more than 500 people, from different locations in Bangladesh, collected from various freelance groups in Facebook."* P9 hints at eligibility criteria: *"To build a team, we first post message in Facebook groups. Then we contact those who respond, personally, and talk to them. We then decide if each is eligible, then we include him in our Facebook group."* P17 claimed access to multiple groups, *"We have 20 groups of real users in WhatsApp."*

Team communications. For communications, the above 10 participants claimed to use the corresponding Facebook and Whatsapp messenger app. P6 said *"I post the app link in my Facebook group, and ask them to download and post reviews."* P11 said *"When I get a job, I send them messages in WhatsApp or I reach them personally."*



(a) (b) (c)

Figure 7: (a) Number of distinct devices per ASO worker (F1 .. F39) including unknown category. F9, F10, F14, F16, F17, and F27 have only unknown devices; F35 used at least 84 distinct device models. (b) Device model popularity for top 15 devices used by ASO workers to post reviews. The 39 participants have used 344 distinct device models. (c) Device model popularity for top 15 devices in the wild. 11,934 unique device models were used to post over 198 million reviews in Google Play.

P7 however claimed to use specialized software: *"We have our own system where we push the apps. Users who use our system get the notifications about the new task and once they complete the task they get paid. Due to the privacy policy, I can't disclose the system name."*

Account maintenance. 5 interview participants (P1, P11, P13, P15, P16) said they access their accounts regularly. 12 participants said that they access them manually. However, 3 participants (P13, P15, P16) said they use scripts and automatic login systems to periodically access their accounts, keep them alive, and report if any are inaccessible. For instance, P13 said *"We have built a system in Linux where if we input 100 accounts, the system automatically logs into those accounts, and keeps them alive."* The participants who organize organic users said that organic users access their accounts regularly.

Job automation. P6 said that *"We can post reviews, ratings and installs using bots if the client has no problem. The bot names are like QZ362, YNX32, or something like these."* All of the other participants said that they write their reviews manually, and do not use any script for this purpose.

5.5 Techniques: The Art of Evasion

Awareness of fraud detection. All interview participants are aware of their fake reviews being detected and deleted. All of them have reported that Google deleted some of their reviews. Although most of them have reported deletion as a small or negligible percentage (under 5%) of all the reviews they posted, four of our interview participants have said that 10–20% of their reviews were deleted. P6 said that the review deletion percentage depends on the app and ranges from 2% to 30%. Most participants said however that it is very infrequent for their accounts to be deleted. P2 said that *"Sometimes the email might be disabled; in that case the review will still be shown as written by a Google User."*

Perceived reasons for deletion. Participants reported diverse reasons for deletion:

• *Device re-use.* P5 and P10 blame it on using the same device to write multiple reviews for an app: *"I always track the screenshot that my workers provide as work proof. If I see two or more reviews from one worker have been deleted, I am pretty sure that they have used the same device for those reviews."* Proof of work details in § 5.12.

• *Improper VPN use.* P10 also blamed VPN: *"One safe way is, login from normal IP, then write review from VPN. If you login using VPN, Google will detect this as fraud."*

• *Improper app use.* P12 said that Google deletes reviews if the users *"do not care to use the app and keep it installed for more days."* More details in the app retention part of § 5.5.

• *Extended account use.* P3, P9, P18 report that using the same account to write many reviews in a short time, may trigger redflags.

• *Misfires of Google fraud detection.* P6 blames it on Google: *"Sometimes genuine reviews get deleted and sometimes multiple reviews from same devices don't get deleted."*

User account validation. P2 and P3 said that they prefer to use e-mail to validate user accounts. P3 also said that Google may force them to use phone numbers. Only P16 claimed that *"we use virtual phone numbers and Google accepts them."* All others said that they use real phone numbers to validate accounts.

Real phone numbers require access to SIM cards, which can be expensive. However, participants revealed ingenious solutions to bypass this limitation. For instance, P3, P10, P11 and P17 use friends and family: P3 said that *"We use our friends and family phone numbers. For example, I meet a friend on the road, I ask him to check the message and I use his phone number to verify an account."* P10 said that *"In Bangladesh one person can buy as many as 20 SIM cards using his credentials. [..] For example, for my 450 Gmail accounts I have used at least 200 phone numbers."* P5 mentioned that he borrowed SIM cards from friends. P7 and P15 use phone number verification services. Concretely, P7 said *"we pay other people to get a one-time code from their mobile SMS to verify those accounts."* P13 said that they purchase user accounts that are already validated.

Several participants reported limitations on phone number reuse. For instance, P3 and P8 said that one number could be used for 3–5 accounts but not immediately, while P1 said that *"between two verification using the same number, we have to wait at least 3 months."*

Review without install. When asked, P5, P10, P13 and P18 said that one can review an app without its prior installation from a device on which the account is logged in [20]: *"Click on install then stop installing immediately. The app would not be installed but it will allow us to write reviews."* We have tested this claim and verified that

it works as suggested. This vulnerability breaks Google's intended security design [20] and facilitates the creation of fake reviews by reducing the amount of resources needed from the ASO worker.

App installation and use. 14 participants claimed to wait, open, or even use the app before reviewing it. P5 and P9 wait a few hours before reviewing the installed app. P9 claimed to also use it for 5–10 minutes. P6 and P8 claimed to open the app 1–2 times before reviewing. P7 claimed to use the app as a normal user. P10, P13 and P16 claimed to keep the app open for 3–15 minutes before writing the review. P12, P14, P17 and P18 claimed to recommend to their online and organic teams to open the app for a few minutes and even use it before reviewing. P4 said *"We try to navigate all the pages of the app before writing the reviews."*

All the participants admitted to perform retention installs. P10 said that this is required to prevent filtering: *"Google takes 72 hours to verify the review. If you delete the app in this period, Google will drop the review."* Most participants said that they keep the app for a few days after reviewing it: 1 day (P1, P5 and P15), 2–3 days (P4, P5, P8, P10, P13, P14), 1–2 weeks (P17), and 7 days – 2 months (P2). P4 said that his workers keep the app until they need the space.

Upvote, Downvote. 6 of the 18 participants (P5, P7, P10, P14, P15, P16) said that they upvote reviews written by their team from other accounts. P7 said *"We upvote the reviews put by our team and also other reviews which are positive."*

P10 said that his team downvote negative reviews of the apps they target, in order to trigger Google's filtering mechanism, thus remove those reviews. P7 said *"We provide upvote and downvote services to move positive reviews to the top and negative to bottom."*

Singleton accounts. P1, P2, P7, P10, P13, and P15 said they worked on jobs where they had to create accounts just to post one review and then to abandon them. P1 and P2 said that the cost of such reviews is higher, $8 and $10 respectively. The reason for this is due to the effort to create an account, which will not be amortized over multiple fake review posting activities. The reason given by the participants for being requested to do this is that Google does not filter reviews posted by singleton accounts, since its fraud detection module needs more information to build a reputation for the account.

Account blending. 12 participants claimed to have seen jobs that required only the use of old accounts. However, P1, P2, P7, P10, P11, P13, P15 said that they have worked on jobs where they only used fresh accounts. P10 said that *"We do it because Google always keeps the reviews received from new accounts."* P1, P2, P7, P16, P18 said that they regularly use a mix of old and new accounts. In § 5.13 we report account creation and purchase strategies.

Noisy reviews. P2, P3, P5, P7, P10, P13, P15 and P16 said that they do not review other apps to avoid detection. Of the physically co-located teams, only P1 said that they review products for which they have not been hired, which they pick at random. 7 participants with online and organic team members (P4, P6, P8, P11, P14, P17, P18) said that their online team members do review other apps, which they normally use in their real life. P4 said *"That's why we use real users. We don't need to follow any strategy. The real users' behaviors serve the purpose of authenticity. We always instruct them to use other popular apps from their accounts."*

Device reset. P10 said that before logging in to an account, they flush the virtual device and change its MAC address. After using the account and virtual device pair for a few days to install and review apps, they log out and repeat the process with another account. They then leave the previous account unused for 1–1.5 months: *"after that interval, Google does not check that the new login is from the same MAC address as the previous one."* P13 similarly claim to stay logged in to the account for 3 days, then they reset the device (using cccleaner) before logging in to the next account.

VPN use. P1, P3, P5, P13, P15 admitted to use VPNs, while the other 10 explicitly claimed to not use them. P3 said *"We use VPN or proxy only when it is required in the job specification. For example, if I need to install from USA, we have to use USA proxy server. (sic)"*

Emulator use. P10 and P13 said that their teams use virtual devices running in laptops. The others claimed to use mobile devices or have access to real users equipped with mobile devices.

Summary. Several of our interview participants confirmed several observations proposed in previous work: (1) ASO workers adjust their behaviors to avoid detection [24, 36, 44, 68, 74, 74], including using VPNs [58, 85], and mobile device emulators running on PCs [58, 81, 96]. However, P10 noted that improper use of VPNs can also trigger fraud filtering. (2) ASO workers also write genuine reviews, for products for which they have not been hired [24, 36, 38, 52, 74, 89]. However, this is only supported by participants who claimed to recruit and use organic ASO workers. (3) Some participants claimed to upvote their own reviews [68]. (4) Some participants also report using singleton accounts [63, 74, 76, 92, 100]. We however report a surprising motivation for this, which is not convenience, but rather a fraud detection strategy that exploits cold-start problems of Google's fraud detector.

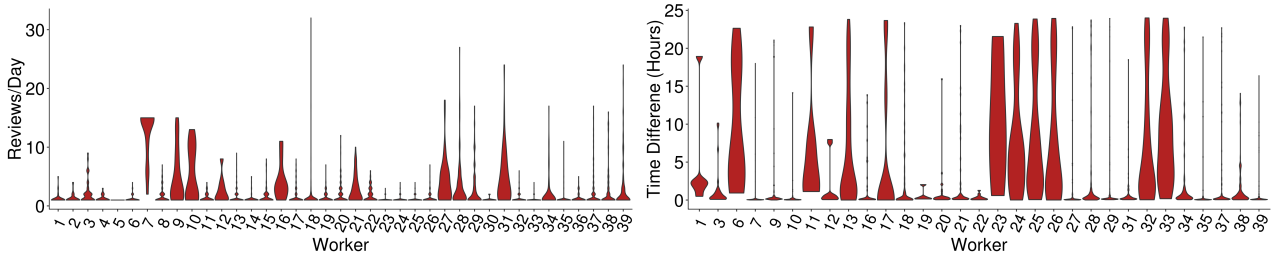
Further, we identified new black hat ASO behaviors, that include downvoting negative reviews to promote their filtering by Google, and the unexpected benefits of using singleton accounts. Participants revealed ingenious solutions to bypass Google-imposed verifications, and validate the user accounts that they control, with real phone numbers. They provide circumstantial support for previous work studying the underlying technical and financial capabilities of social network fraudsters [84].

Several participants reported the ability to bypass Google's check of preventing reviews without prior app installation. However, to avoid filtering, all participants said they use a combination of app interaction, delaying of review posting, and retention installs.

Further, we conjecture that the claimed use of a blend of older with newly created accounts, enables ASO workers to replenish or increase their base of accounts controlled, build the reputation of older accounts, and reduce chances of detection of lockstep behaviors (§ 5.8) and the use of singleton accounts.

5.6 Review Burst vs. Campaign Length

We now present findings on the timing of the review process. 16 interview participants claimed to have seen jobs (1–45 in the past month) that specify how many reviews per day the workers should post. For instance, P5 said that *"Most buyers don't want to get all the reviews in a single day. They want a slow rate, like 2–3 reviews each day. To maintain this rate, they provide the review text on a daily basis."* However, P6 also said that *"some developers with money don't care whether reviews stay or not. They just need the number of reviews, quality doesn't matter. They just want short-time business."*



(a) (b)

Figure 8: (a) Per-worker distribution of the number of reviews per day for each targeted app. (b) Per-worker distribution of time difference in hours between consecutive reviews posted within one day for targeted apps. F7, F9, F10, F16, F27, F28, F31, tend to post more reviews per day, in bursts. F1, F3, F19, F20, F23, F29, F35, F37-39 post few daily reviews, but in bursts. Others like F6, F11, F13, F23-F26, F32, F33 post few daily reviews, but space them through the day (post one every 8-9 hours).

P1, P3 and P5 reported that they suggest to the hiring developers, the rate of posting reviews. P5 said *"If the developer asks for 30 reviews each day, I have to warn him that it's harmful to his app as Google may detect this as fake. Then I'll suggest to him that I will take 10 days to provide 30 reviews."* Most participants suggest 2–3 reviews per day, but some (e.g., P11, P14, P17, P18) recommend higher numbers, up to 30–40 reviews per day (P14).

Several participants suggested that the number of recommended daily reviews is a function of the app's existing review count. Concretely, P6 said, *"for new apps with less installs, it is better not to provide many reviews each day. But for popular apps, 20–50 reviews each day would be acceptable."*

P10 revealed a different strategy: *"We provide a slow rate at the beginning. Like 1 review per day, or 5 reviews in 6–7 days. After 10 reviews we start posting 2 reviews each day. After 150 reviews we can provide 3–4 reviews each day."*

All the participants except P4 said that they have seen ASO jobs that require a duration for the review posting campaign. P6 and P15 said that this is rare, and that developers are more concerned about the total number of reviews. However, P2 said almost all the jobs he has seen in the past month, mention the campaign length. In the past month, 5 participants have seen 3–5 such jobs, 4 have seen 6–10 jobs, and 5 have seen 11–35 such jobs. 12 participants reported longest seen required campaigns of 1–6 months, and 6 participants reported campaigns of 7–18 months.

Quantitative Investigation. Figure 8(a) shows the per-worker, violin-shaped distribution of the number of reviews per day, posted from accounts controlled by the 39 ASO workers, for each targeted app. Figure 8(b) shows the violin plots for the distributions of the inter-review times (only those posted within the same day). We observe several participants, e.g., F7, F9, F10, F16, F27, F28, F31, who tend to post more reviews per day, an do so in bursts. We also see participants, who even though write fewer reviews per day, still tend to post them in bursts (F1, F3, F19, F20, F23, F29, F35, F37-39). However, as also reported by the interview participants, we also found ASO workers who post only a small number of reviews per day and space them well through the day. Notably, F6, F11, F13, F23-F26, F32, F33 have a mean inter-review time of 8-9 hours.

Further, we call a worker's *active interval* for an app, the time span (in days) between the worker's last and first review for the app from accounts that we know he controls. Figure 9 shows the per-worker active interval distribution over the 316 apps that received

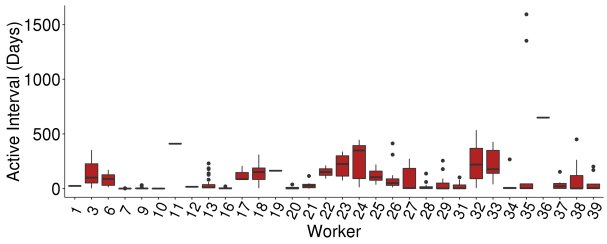


Figure 9: Per-worker distribution of active intervals (in days) over apps targeted. Each point represents the active interval of an ASO worker for an app. We observe workers who have posted reviews for certain apps, for more than 1 year, and up to more than 4 years.

at least 10 reviews from the 39 participants. Some ASO workers were often active for more than 1 year for an app.

Summary. We found ASO workers who post fake reviews in rapid bursts in both our qualitative and quantitative investigations. This is consistent with assumptions made in previous fraud detection work, e.g., [26, 28, 32, 36–38, 40, 42, 43, 51, 52, 59–61, 63, 65, 92, 95, 96, 100]. However, multiple interview participants have revealed both developer and ASO worker assumptions that Google flags review bursts. Some participants also claimed to push back on developers who asks for many daily reviews. Our quantitative analysis reveals ASO workers whose behavior is consistent with these statements. Interview participants further revealed avoidance techniques that include (adaptive) rate control.

Rate control implies longer campaigns, as workers need more time to post their review quota. This is further supported by statements made by several interview participants and by evidence we extract from the quantitative investigation.

5.7 Accounts Per Device Strategies

Participants revealed mixed strategies for the number of accounts used on a device, and the number of reviews that they publish from a single device. P10, P11, P13, P18 said that they only log in to one account at a time, on any device that they control. P18 has 30 devices and 30 accounts, and a 1-to-1 mapping between accounts and devices. P11 said that *"If we provide multiple reviews from one device, Google will keep only one review for that device."* P5, P6, P7, P8, P9, P15 and P16 claimed to log in to multiple accounts (2–5) from a single device and also instruct their remote workers to do the same.

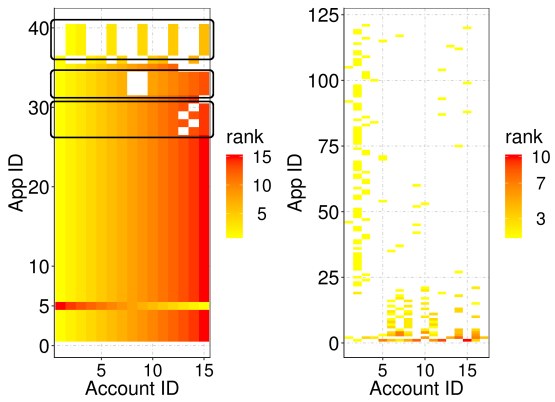


Figure 10: Lockstep matrices for F7 (left) and F32 (right). Rank (color) indicates the order in which an account was used to review an app. F7 exhibits strong lockstep behaviors, having used almost all his revealed 15 accounts to review all the 40 apps (exceptions shown within black rectangles). F32 however exhibits less obvious reviewing patterns.

However, P5 and P9 claimed to only provide one review from one device for an app. P15 and P16 keep track of which accounts they use to log in to any device, and once they log out from one account, they wait 7–10 days before they use it again.

P5 claimed to use a fixed set of 2–3 accounts to log in to one device at a time, then uses those accounts to review multiple apps. However, he also claimed that he only provides one review from one account for an app. P6 said that he instructs his remote workers to log in to at most 2 accounts from any device (at a time), however, they can review the target app from both accounts. P7 mostly use 2–3 accounts from a device for safety. P8 claims to login to 5 accounts on his device, and his Whatsapp group members log in to 3–5 accounts per device. P9 claims that he has logged into 4 accounts in a device, but he does not allow his workers to post more than one review from any device.

Summary. ASO workers generally claim that it is possible to review an app from different accounts using the same device. We have tested this claim and verified that it works as suggested. This vulnerability facilitates the creation of fake reviews by reducing the amount of resources needed from the ASO worker.

5.8 Lockstep Behaviors

Interview participants revealed different strategies to choose which of their accounts and devices to use for a job. Several participants revealed lockstep-indicative behaviors, based on a spreadsheet of accounts and devices that they maintain across all their jobs. P5, P7, P10, P13, P18 select the devices in a sequential, round-robin manner, while P5, P7, P13, P15, P16, P18, select the accounts sequentially. For instance, P15 claimed that *"We have statistics on how many times an account was used previously. From there we try to find accounts that have been used fewer times. We also track which device was used for which account, so next time we use the same device for that account."*

Others however claimed non-lockstep indicative behaviors. 7 of the 18 participants (P6, P8, P9, P11, P12, P14, P17) claimed a random choice of accounts and devices, including made by their remote online employees. P16 claimed to monitor the reviews filtered, and choose accounts based on their filter avoidance success rate.

To investigate lockstep behaviors in the gold standard fraud data (§ 4.2), we used frequent itemset mining [23, 64] to discover sets of apps that are co-reviewed by many accounts in the same or similar order. Intuitively, a set of apps reviewed by the same, many user accounts, is said to be "frequent". More formally, let $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ be a set of apps, and let $\mathcal{U} = \{u_1, u_2, \ldots, u_m\}$ be a set of users in Google Play. We say that a set $A \subseteq \mathcal{A}$ is s–frequent if $\frac{|\{u \in \mathcal{U}; A \subseteq T_u\}|}{|\mathcal{U}|} \geq s$ where $T_u = \{a \in \mathcal{A}; a \text{ is reviewed by } u\}$.

We used the A-priori algorithm [23, 57], to find per-worker *maximal* frequent itemsets: frequent itemsets for which none of their immediate supersets are frequent. 25 of the 39 participants had maximal frequent itemsets with $s = 0.5$. That is, they used at least half of their accounts to review common subsets of apps.

Figure 10 shows *lockstep matrices* for two of the ASO workers. In the lockstep matrix M_{ij} of a worker, columns are user accounts controlled by the worker and rows are apps reviewed from those accounts. $M_{ij} \in [n_w]$ denotes the chronological order of the review posted by account j on app i. n_w is the total number of reviews posted by the worker to app i. ASO worker F7 (left) shows a nearly perfect lockstep behavior with the same set of 15 accounts used for almost all the 40 apps, and in the *same order*. We also see attempts at "variation": F7 uses his accounts in exact reverse order to promote app 5. Further, for several sets of apps (black rectangles in Figure 10), F7 does not use the same set of accounts, and uses all his other accounts in the same order.

However, 14 participants exhibit less pronounced lockstep behaviors, e.g., F36 (Figure 10 right). Out of 121 apps reviewed, in only two apps, F36 used more than 50% of the 17 accounts he revealed.

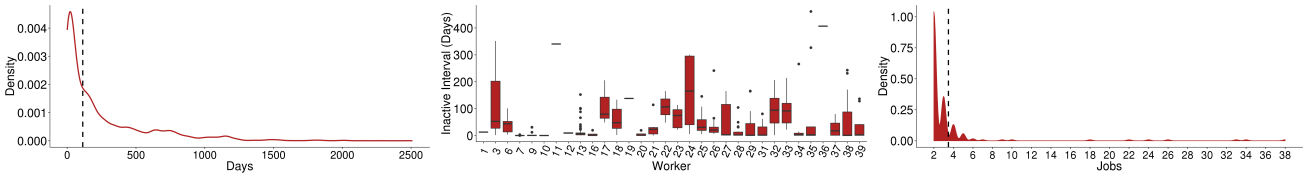
Summary. 6 out of 18 interview participants claimed lockstep-indicative behaviors; 25 of the 39 quantitative study participants exhibit lockstep behaviors, some even using their accounts in the *same order* to review multiple apps. This is consistent with and provides evidence for assumptions made in previous work, e.g., [32, 48, 59, 77, 81, 86, 93, 94, 97, 100].

However, we also report claims (8 of 18 participants) and evidence (14 of 39 participants) of random account and device choice. We conjecture that ASO workers may adopt evasion strategies, e.g., by using different sets of accounts for different jobs, and use organic workers, less likely to be frequently active at the same time.

5.9 Timing: Fraud Event Points

Early bird fraud. 14 participants said that they have worked on recently launched apps, and either the hiring developer mentions that the app was recently launched, or that they infer this information based on the app status when posting their first review. Declared numbers range from 1–2 jobs in the past month (P1, P11) to 20–40 (P9, P10, P13). P7 said that *"We even work on apps which are going to be launched soon. A few of our clients rely on our agency from pre-launch to launch and then post-launch."*

Re-hires. All 18 participants claimed to have been re-hired for apps that they previously promoted (total times M = 186.1, SD = 190.7, Min = 15, Max = 600). P1 said that *"If the app is getting bad reviews, the developer will hire us again to get good reviews. We have seen this case for minimum 30 to 40 apps per year."* P12 said *"I have around 20 regular clients. They hired me for the same app, around 40–50 times."* Further, all of the 18 participants claimed to have regular customers, who hire them to promote multiple apps.



(a) (b) (c)

Figure 11: (a) Relative likelihood for the time difference between launch time and reviews by ASO workers, for 585 apps that received at least 10 fraudulent reviews. Vertical dashed line is the median. (b) Per-worker distribution of the maximum inactive interval measured in days for each targeted app. 8 participants, e.g., F7 and F9 are intensely active, however, F3, F24, F32 and F33 exhibit more evidence of later rehiring. (c) Density function of number of jobs received by ASO workers from the same developer. One worker worked on 38 apps of the same developer. The vertical dashed line corresponds to the median value.

Quantitative Investigation. Figure 11(a) plots the time difference in days, between the app launch time and the posting time of each review from a fraudulent account controlled by any of the 39 participants in the quantitative study (§ 4.2), over the 585 apps that received at least 10 fraudulent reviews in total. The distribution is left-skewed, with 50% of the reviews being posted after less than 3 months after app launch. However, we observe cases where the first reviews from any of the accounts of our 39 participants, are posted long after the app was released: the median and 3rd quartile are 113 and 344 days respectively. Thus, about 25% of the fake reviews were written after one year.

We call the *inactive intervals* of an ASO worker for an app, to be the time differences between consecutive reviews that he posted to that app, from accounts that he controls. Figure 11(b) shows the per-worker distribution of the *maximum inactive interval* computed over each app that the worker reviewed from accounts that he controls. We show only the workers with enough points to compute statistics. 8 workers have very short inactive intervals, thus are more intensively active for the apps that they target. However, ASO workers such as F3, F24, F32 and F33, have longer inactive intervals, suggesting rehiring. For instance, we found 16 cases where the worker was inactive for more than 8 months for an app.

Figure 11(c) plots the density function of the number of apps uploaded by the same developer, and reviewed by the same worker, over the 39 workers of the quantitative investigation. We observe that the mean number of jobs assigned is 3.48, and 7 workers have been hired by the same developer more than 10 times. We found one developer that hired 6 workers to each promote at least 10 apps.

Summary. Our qualitative and quantitative studies provide evidence confirming observations and assumptions made in previous work, that (1) ASO workers tend to be hired early after app launch, or even before launch, to control review sentiment, see e.g., [38, 53, 61, 63, 64, 96, 100] and (2) developers rehire some of these workers at later times, when honest feedback reduces the product rating [53].

5.10 Review Writing

We asked interview participants about, and report findings on the source of review text, plagiarism, and review length:

Review text source. 2 participants (P3, P4) said that they always write their own reviews. The other participants said that they both receive or request the review text from the developer, and they also write their own reviews. P2 said that they receive instructions about the reviews from the developer. P11 reported developers who provide review samples, from which they are supposed to generate variations. 3 participants (P7, P8, P15) said that they either prefer or even ask the developer to provide the review text. P3 and P13 said that they study the app before writing the review. P13 claimed to ask the developer to provide the app's main features, which he uses to fabricate reviews.

Review posting process. The participants revealed a mixed strategy of typing the reviews directly on the device, vs. cut-and-pasting them from a separate source. 11 of the 18 participants said that they type the reviews directly from their devices. For instance, P5 said that they cut-and-paste reviews if provided by the hiring developer, otherwise they type their own (short reviews). However, P7 noted that most devices do not allow cut-and-paste. Several participants organize teams of remote ASO workers, thus stated that they are not aware of their review-typing actions.

Review plagiarism: 8 participants (P1, P3, P5, P12, P13, P14, P15, P18) denied plagiarism and self-plagiarism. P2, P4, P6, P9, P11 and P17 however admitted to some form of plagiarism. P2 blamed it on developers: *"Yes, sometimes we copy, but only if buyers mention the source, for example, apps hosted in other sites."* P4 said that *"we don't copy-paste. But our reviews are short and sometimes similar."* P16 said *"We have a review data set, and we use those reviews for all apps. Sometimes we change a the reviews bit for different apps."* P9 said that *"Not exact copy-paste. But sometime we copy and modify reviews from other apps that are similar."* P12 also complained about some organic users, who are careless and write random comments, e.g., "nice game" for a non-game app.

Review length: 11 participant claimed that their reviews exceed 10 words (10–40). P3 and P4 admitted that their reviews are short (3–5 words). P4 motivated this choice: *"We don't use many words or big sentences because Google may match the pattern. We always use short messages like "Good app", "Awesome", "Fantastic". These are very common but easy to write and Google may not complain."* P6 argued that *"if you write too long reviews, they will certainly look like paid reviews, because real users don't have time to post a paragraph."*

Quantitative Investigation. Figure 12 shows the empirical CDF of the review word count over all the reviews posted by the 39 participants, and also only for F7 and F26, who wrote 542 and 771 reviews respectively, and are the ASO workers with the most distant CDFs from one another: $\mathbb{P}(Length \leq 10|F7) = 0.88 \gg \mathbb{P}(Length \leq 10|F26) = 0.06$. The overall fake review word count CDF is closer to F7, with the overall $\mathbb{P}(Length \leq 10) = 0.63$.

Further, we identified exact review duplicates among the 21,767 reviews posted by the 39 participants (§ 4.2), and sorted them by

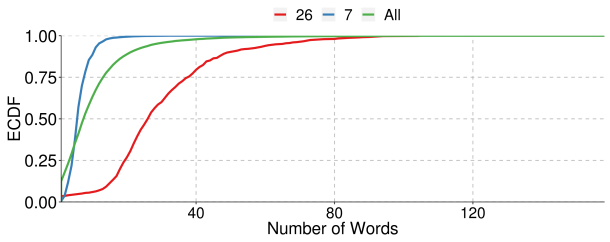


Figure 12: Empirical CDF for two extreme behaviors shown by two participants. All other workers have their corresponding CDF between these two curves and are not displayed for better visualization. We note that $\mathbb{P}(Length \leq 25|F3) = 0.99 \gg \mathbb{P}(Length \leq 25|F26) = 0.46$, and the all-worker ECDF is closer to worker 7 who writes shorter reviews.

the geometric mean between the number of ASO workers who have written the review and its overall frequency. An advantage of the geometric mean is that it gives a balance between two quantities that are in different ranges. 993 reviews were empty (154.37). The next 10 most repeated reviews ordered by geometric mean were "good" (30.51), "Good" (27.42), "nice" (20.63), "Love it" (15.19), "app" (13.71), "Excellent" (13.26), "Awesome" (12.64), "Like it" (11.83), "Nice app" (11.18), "Great app" (10.95). We note that these reviews are short, generic, and app-agnostic. This analysis validates the survey answers by some ASO workers, that short reviews may be preferable since long reviews may trigger Google's defenses and block their content.

Summary. Most interviewed participants said that the text of the reviews is provided by the developers, but also they can write their own reviews. Consistent with previous observations, e.g. [38, 42, 51, 52, 60, 61, 63–65, 74, 76, 96, 97, 99], several participants admitted to reuse common linguistic patterns and copy reviews across similar products. We also confirmed this finding in our quantitative study.

Further, most participants claimed to write short reviews, which is also reflected in our gold standard fraud data. Previous work, e.g., [38, 49, 51, 52, 60, 65], also made this observation, and attributed it to the fraudster lack of experience with the product. However, we also present evidence of ASO workers who post much longer reviews. We conjecture that fraud evasion can also be a factor.

5.11 Ratings

Rating choice strategies. All 18 interview participants admitted writing mostly 4 or 5-star reviews unless they receive special instructions from the developers. 8 participants (P3, P7, P8, P9, P10, P11, P14, P18) said that they receive instructions on the ratio of review ratings from the developers. For instance, P12 said that *"developers request us to write a few 4, 3 and even few 1 star reviews."*

When there are no instructions on the rating distribution, several participants claimed to maintain their own ratio. For instance, P5, P16, P17 claimed to post a 10% vs. 90% ratio of 4 to 5 star reviews, P2, P10, P18 have a 20%-80% ration, P1, P9 have a 30%-70% ratio and P13 has a 40%-60% ratio. 3 participants (P6, P11, P14) said that they do not maintain any specific ratio, while P4 and P12 post only 5-star reviews.

3 participants claimed strategies to also post lower ratings, in order to avoid detection. For instance, P6 said that: *"if the average rating goes up to 4.3 or 4.4, I also write a few 3-star reviews."* P7 said that *"when posting more than 200 reviews, we suggest to the client to have at least 5 to 6 reviews with 3 star ratings."* P15 claimed to post a 10%-30%-70% ratio of 3, 4, 5-star reviews.

Negative campaigns. When asked if they were ever hired to post negative (1-2 star) reviews, and how many such jobs they worked on, only two participants said that they participated in such negative review campaigns. P3 had participated in only one such job, but later morally objected to it, while P4 also admitted to have worked on only a few such jobs (5-7). The other participants said that they never participated in negative campaigns. We did not ask participants how many such campaign jobs they have seen.

The gold standard fraud data we collected from 39 participants confirms that 95.52% of the 21,767 reviews posted from the accounts they control, were either 4 or 5 stars. Only 1.67% were 3-star and 2.81% were 1 or 2 star reviews.

Summary. Both interview participants and gold standard fraud data reveal the prevalence of positive ratings. This confirms observations and assumptions made in previous fraud detection work, e.g., [24, 51, 52, 63–65, 74, 93, 94]. However, we found that negative review campaigns (or negative ratings) are unpopular. Further, several interviewed participants reported rating-level detection evasion strategies, e.g., the sprinkling of neutral and negative ratings, among positive reviews.

5.12 Proof of Work

After ASO workers finish their jobs, it is expected for developers to ask for proof of work. 12 participants said that they use screenshots of their reviews. 5 participants said that they send the usernames of accounts that they used to post reviews. P6 claimed *"I check my reviews for 2–3 days and then send the permalinks that are direct links of the review I post, or names I used to post the reviews."*

Team-level verifications. Work verifications can take place at the team level. For instance, P3 said that *"[..] we ask everyone to post reviews in the team. Then I track how many reviews we provide and they also send me the screenshot. If the buyer requires the screenshots I send him those too."* P6 said that *"If we get a report that any review is being deleted then we check that user's mobile and ask him to provide a screenshot of the app installed immediately. If he fails to provide that, I flag him as a bad user and we consider him less for the next tasks."* P9 also verify that their team members do not post multiple reviews from the same device, by looking at the screenshots sent.

Follow-up. P3 said that *"Sometimes, the developer keeps track of the reviews we post, and gives us 24 hours to show that the reviews are alive. If any review is deleted during this time, we have to re-post the reviews."* P7 claim to provide guarantees of reviews sticking for 5–7 days and refill deleted ones for free.

5.13 Account Creation

13 of the 18 interview participants, mentioned use of fake name generators, e.g., [6], to name their user accounts. Some of them create account names to correspond to specific geographic regions, as sometimes also requested by developers. P2 even claimed to send the chosen names to the employer for feedback. P11 claimed to use random names from Google search and P7 said that they have their own name database. P4, P7 and P14 said that their use of organic ASO workers, ensures that they use real user names.

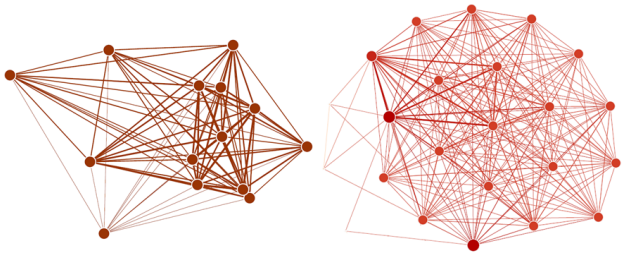


Figure 13: Co-review graphs built over the accounts claimed to be controlled by (left) F13 and (right) F32. Edge width is proportional to the number of apps reviewed in common by the endpoint accounts. 14 accounts revealed by F13 form a clique, and on average, any two accounts reviewed 78 apps in common.

7 participants said that they add profile pictures, which they retrieve from different sources, e.g., Google search, Google Plus, pixabay.com, to make the account look more authentic. P9 said *"After we use fake name generator to create the account name, we search the name in Google Plus and choose a profile, then we choose a random person from the list of followers and use his image for the account profile."* P10 however said that *"We use no picture as picture defines your demographics. Buyers do not want this now."*

Account Creation vs. Purchase. 6 participants (P1, P3, P7, P10, P13, P16) claimed to create new accounts periodically, ranging from once a day (P10) to once a month (P16). P2 and P9 claimed to create new accounts when they don't have enough accounts for a job, especially when the job requests accounts from a specific geographic region. P5 and P18 create new accounts when Google deletes some of their accounts. P15 create new accounts when the job requests more reviews than they can provide. 5 participants (P1, P5, P13, P17) admitted to purchase new accounts. P1 claimed to have purchased more than 10,000 accounts, while P13 claimed to have purchased 47,000 accounts. Two participants (P1 and P3) volunteered the fact that they age their new accounts (1–2 months) before using them to post reviews.

Summary. The claims of fake name generator use, provide evidence toward limited variability in naming patterns for the worker-controlled accounts, as previously assumed [85]. We identified profile photo plagiarism behaviors, but also a claimed developer-driven trend to avoid profile photos.

5.14 Validation and Efficacy of ASO

Validation of quantitative study. Collecting ground truth fraud data attributed to the workers who created it, is a difficult task. We believe that any process to obtain such ground truth data needs to involve the workers. In addition, to gain confidence in the correctness of the accounts claimed to be controlled by the 39 workers, we used *co-review graphs* built over the accounts claimed to be controlled by each worker: nodes are user accounts, and edges have weights that denote the number of apps reviewed in common by the end-point accounts. Figure 13 shows example co-review graphs built over the accounts revealed by F13 and F32.

Figure 14(top) shows the average co-review weight of the accounts claimed to be controlled by each of the 39 ASO workers, i.e., the ratio of the sum of all edge weights to the number of edges.

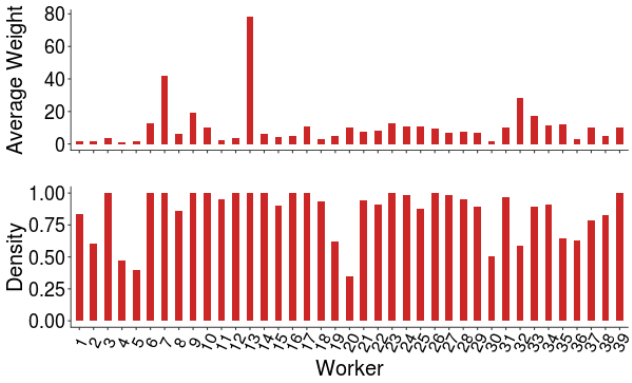


Figure 14: Density and average weight for co-review graphs of 39 ASO workers. 12 workers have complete graphs (density=1). 30 workers have graphs with density at least 0.75.

Figure 15: Active vs. inactive accounts controlled by the 39 quantitative study participants. We observe diverse success in keeping accounts active on the long term.

Figure 14(bottom) shows the edge density of the worker co-review graphs, i.e., the ratio of the number of co-review edges to the maximum number of edges possible in that graph. The co-review graphs of 12 of the workers are cliques, i.e., any two accounts have reviewed at least one app in common. Further, the co-review graphs of 16 workers have an average weight of at least 10, up to 78.61 for F13. This is in contrast to the probability of co-rating two apps in Apple's China App Store, of 0.163% (computed over 0.5 million random accounts) [93].

In addition, we manually investigated the accounts revealed by the 39 workers, and found multiple instances of repeated profile photos, mostly of glamorous people, and simple patterns in the account names.

Efficacy of ASO. To investigate the efficacy of the ASO strategies employed by the 39 workers who participated in our quantitative study, we look at (1) the number of accounts that they control that are still active, and (2) the impact of their ASO campaigns.

Figure 15 shows the number of accounts controlled by each of the 39 workers, that are active and inactive (i.e., Google returns 404 not found error). Of the 1,164 accounts known to be controlled by the 39 workers, 120 were inactive (10.30%) in May 2019. Qualitative study participants stated that they never abandon accounts unless they are closed by Google or Google filters all their reviews. Thus, Figure 15 reveals diverse success among the 39 ASO workers, in terms of being able to keep their accounts active long term: while a majority of the workers have all their accounts still active, including the workers with more than 40 accounts, several workers had a

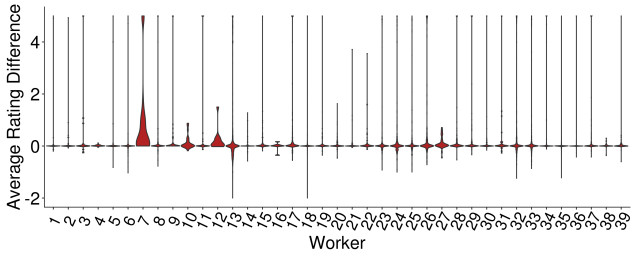


Figure 16: Impact of campaigns conducted by the 39 quantitative study participants, on the average rating of apps for which they campaigned. We observe diverse success in increasing the average rating of targeted apps.

majority of their accounts closed. Notably, 36 out of the 47 accounts controlled by F34 are closed, as are 29 out of 35 accounts of F31.

In addition, we studied the *impact* of a worker on each app on which he has performed an ASO campaign. We denote the impact I_A of an ASO worker W for an app A to be the change in A's rating during W's active interval. Specifically, $I_A = R_f - R_i$, where R_i is A's "initial" average rating, i.e., before the first review posted by the worker for A, from any of his accounts, and R_f is A's "final" average rating, after W's last review posted for A. Figure 16 shows the violin plots of the distribution of impact values, over all the apps campaigned by each of the 39 ASO workers, from all the accounts that each controls. We observe diverse abilities of these workers. For workers like F7, F12, and F21, we observe only positive impact on the average ratings of all the apps that they target. Most workers however have mixed impact, with many of their targeted apps seeing up to 5 star increase in average rating during their active interval, and a few others seeing up to a 2 star drop. We observe however that overall, apps seem to benefit from the campaigns in which these workers have contributed.

We conclude that different strategies have different impact on the ability of ASO workers to avoid detection and impact the ratings of apps that they target.

Our study has several limitations. First, we do not know all the accounts controlled by the 39 ASO workers. Second, we cannot pinpoint the exact strategies that are responsible for the success to maintain accounts active or ensure that reviews are not filtered. Such an analysis would require detailed experiments that explore the impacts of altering a single feature of a fraud detection algorithms that is kept a close secret. Third, the impact that we computed, is oblivious to simultaneous campaigns being conducted by other workers on the same apps. Finally, our computed average rating of an app is imperfect, since (1) we do not have access to ratings posted without reviews, and (2) may not correctly model Google's algorithm, that e.g., may assign weights to ratings based on perceived usefulness, fraudulence or recency [72]. We describe more limitations of our studies, in § 7.

6 DISCUSSION AND RECOMMENDATIONS

The varied capabilities, behaviors and evasion strategies claimed and exhibited by the studied participants, suggest that fraud detection solutions should cast a wider net. While some of our participants seem to fit the mold of assumptions made in previous work, we present claims and evidence of evolution, perhaps fueled by the competitive nature of the market. In this section, we propose disruption strategies for each vulnerability point identified in the fraud workflow of Figure 1, and discuss potential implications of our study's findings, on future fraud detection and prevention solutions. The opaque nature of commercial fraud detection systems prevents us from establishing the costs and scalability of implementing the proposed recommendations, or from determining if they are already implemented. However, manual verification of statements made by ASO workers revealed several weaknesses in Google's defense. Some of the following defenses propose to address them.

VP1: Proactive Fraud Monitoring. Recruiting WhatsApp/Facebook groups need to aggressively accept new collaborators. We verified that these communication channels are easy to infiltrate. Thus, we recommend to proactively detect campaigns at this point, and flag apps likely to receive fraudulent reviews, and suspicious accounts engaged in posting fraud.

VP2: Device Fingerprinting. We observe that device models and their per-country popularity can be used to detect reviews written from accounts claiming to be from a country where the posting device is not popular. However, this vulnerability could also be used by ASO workers to blend in with normal users, by mimicking the distribution of devices observed in Google Play.

Further, this device-model leaking bug can also be used by computer criminals to perform reconnaissance on potential victims. Figure 7(c) in shows the top 15 most popular devices, out of 11,934, that were used to post 198,466,139 reviews in Google Play. An adversary could use this bug to for instance, identify owners of device models known to be vulnerable, e.g., [27, 90]. We notified Google about the dangers of this bug, see § 4.3.

VP3: 1-to-1 Review-To-Device. Our interviews and experiments revealed that a user can download an application once, and review it as many times as the number of accounts she has logged in to the device (up to 5, claimed by, e.g., P8). We suggest enforcing that a device can be used to post only 1 review per downloaded app.

VP4: Organic Fraud Detection. We suggest the use of account activity levels to differentiate organic from inorganic (sockpuppet) accounts. Organic ASO workers are likely to use their devices continuously, like the normal users that they almost are. Sockpuppet accounts are more likely to experience inactive interludes given the dynamic of their workflow (§ 5.5). Account activity includes but is not limited to the number of apps with which the account interacts per time unit, the duration of such interactions, and the number of other Google services (maps, gmail, drive, music, etc) to which it is subscribed. Additionally, our data and experiments reveal that some workers may even be posting only laptop-based reviews as all their reviews were written from devices of *unknown* models. Our study suggests that these workers are more likely to control sockpuppet accounts. This requires however future validation.

VP5: Monitor Review Feedback. An account should be able to upvote or downvote a review only if it has installed the respective app on at least one device. We verified that this is not currently enforced by Google Play. Fraud attribution (see below) can also be used to discount upvotes from accounts known to be controlled by the same ASO worker as the one that posted the review.

VP6: Verify App Install and Retention. We recommend developing protocols to verify that an app has been or is still installed on the device, e.g., before accepting a user review from that device.

While remote attestation inspired solutions (e.g., [46]) will not be secure without device TPMs, defeating such solutions will require significant investment from ASO workers.

VP7: Account Validation and Re-validation. The cellular provider used during account validation can also be used to detect inconsistencies with the claimed profile (e.g., location) of the user account. Further, several ASO workers mentioned using SIM cards of others to validate their accounts. Peer-opinion sites could ask users to re-validate their accounts at random login times (e.g., veiled as "improved authentication security"), especially if their validating SIM cards have also been used for other accounts.

VP8: App Usage. Most ASO workers suggest that they use apps before reviewing them, and keep them installed after review for a while, to mimic genuine behaviors. However, we believe (but have not investigated) that features extracted from per-app waiting times, app interaction modes and times, and post-review behaviors, are different for honest vs. fraudulent accounts, and could be used to pinpoint sockpuppet and organic fraud accounts. For instance, it is suspicious if an app receives a good review soon after it was downloaded, has received little interaction, and is quickly uninstalled. Coupled with VP6, mandating wait times to post reviews will impact the number of apps that an ASO worker device can store, thus the number of apps that a worker can target at a time.

VP9: Mislead ASO Workers Through Fraud Attribution. SIM cards can also help attribute sockpuppet accounts to the ASO workers who control them, see e.g., [41]. Account-to-ASO worker attribution can be used to reduce worker ability to adjust to detection [79]: to mislead ASO workers into believing that their actions are effective, peer-opinion sites could show removed fake positive reviews only to the accounts used to post them, the other accounts suspected of being controlled by the same worker, and the account of the app developer. This would force ASO workers to partition their account set into monitoring-only sets that cannot be used to post fraudulent reviews, and regular fraud-posting accounts.

VP10: Once a Cheater, Always a Cheater. Our qualitative and quantitative studies (§ 5.9) provide evidence that developers rehire ASO workers not only for the same app, but also for other apps that they develop. We recommend to monitor overlapping accounts that review sets of apps by the same developer, and redflag fraud developers early on.

7 LIMITATIONS

Recruitment Bias. We have not performed a complete exploration of the ASO worker universe, and cannot claim that our participants are a representative sample. Our recruitment process is biased, since we selected only candidates who (1) we could reach out to, (2) responded, (3) were English speakers, (4) were willing to participate after approving the consent form, and (5) claimed qualifying capabilities (i.e., control at least 100 accounts, have at least 1 year of ASO expertise and participated in at least 100 ASO jobs, § 4.1).

For instance, out of the 560 contacted workers, 72 replied to our invitation, 25 qualified, and 18 agreed to finally participate. Thus, other workers will likely have both fewer and more capabilities than the participants in our studies. However, from the answers and data that we collected, we reveal previously unknown ASO strategies, provide insights into previously proposed defenses that may be effective against them, and report Google defense vulnerabilities.

We leave for future work an investigation into the ability of deception and more substantial financial incentives, to increase the recruitment success rate and identify novel ASO strategies. We believe that our approach is a best effort in recruiting workers, without the use of deception.

Generalization of Results. We have used crowdsourcing sites such as Upwork, Fiverr, Zeerk, and Peopleperhour for years, and have found them to be reliable sources of ASO activities. In addition, we have also found and used, after being pointed out by multiple ASO worker contacts, large groups in Facebook, that specialize in ASO. However, we do not claim that we were able to contact most of the active ASO workers.

The participants in our studies also claimed expertise in fake reviews and ratings in Google Maps, Apple Store, Amazon, Facebook and Twitter, fake installs in Apple App Store, fake likes and followers in Facebook and Instagram, and influential tweets in Twitter. However, we did not ask participants, questions about their strategies in other platforms. Thus, we do not claim that our findings apply to other sites or other types of ASO work.

Validation of Findings. Due to the sensitivity of the topic surveyed and data collected, we did not perform the quantitative and qualitative studies on the same participants. Our quantitative study is also performed only on a subset of the accounts controlled by 39 participants. We have corroborated multiple survey answers with quantitative measurements, and also manual verification by the authors. In § 5.14 we describe the process we used to validate the data collected in the quantitative study. However, several participant claims are difficult to validate (e.g., team organization, size and location, capabilities, interactions with employers, number of devices controlled, etc). The particular nature of our participants, makes any suspicion on these topics, legitimate.

8 CONCLUSIONS

In this paper we present results from the first structured interview study of 18 ASO workers we recruited from 5 sites, concerning their fraud posting work in Google Play, and also a quantitative investigation with data that we collected from 39 other ASO workers recruited from the same sites. We report Google Play vulnerabilities, and new findings about the capabilities, behaviors and detection avoidance strategies claimed and exhibited by ASO workers.

Taken together, our study is limited by the difficulty to recruit participants and the sensitivity of the data. The presented findings are hence needed to be understood as situated information and not as generalized facts. Since the nature of fraud detection research involves elimination of risks and vulnerabilities, the presented findings, even with all their limitations, provide new suggestions for future research. Further, given the observed ASO worker ability to adapt, we believe that future research should focus on collecting more such information from diverse sources, to extend and ensure the continued relevance of our findings.

9 ACKNOWLEDGMENTS

This research was supported in part by the NSF under grants CNS-1840714 and CNS-1527153, NSERC grants RGPIN-2018-06185 and DGECR-2018-00103, and the Florida International University's Dissertation Year Fellowship.

REFERENCES

[1] [n. d.]. App Reviews. http://www.app-reviews.org.
[2] [n. d.]. App Such. http://www.appsuch.com.
[3] [n. d.]. AppBrain. https://www.appbrain.com/info/about.
[4] [n. d.]. Apps Viral. http://www.appsviral.com/.
[5] [n. d.]. BlueStacks. https://www.bluestacks.com/.
[6] [n. d.]. Fake Name Generator. Your Randomly Generated Identity. https://www.fakenamegenerator.com/.
[7] [n. d.]. Freelancer. http://www.freelancer.com.
[8] [n. d.]. Gadgets 360. https://gadgets.ndtv.com/.
[9] [n. d.]. Google Play. https://play.google.com/store?hl=en.
[10] [n. d.]. Google Play Help – Supported Devices. https://support.google.com/googleplay/answer/1727131?hl=en.
[11] [n. d.]. Google Vulnerability Reward Program. https://www.google.com/about/appsecurity/reward-program/.
[12] [n. d.]. GSMArena. https://www.gsmarena.com/.
[13] [n. d.]. microWorkers. https://microworkers.com/.
[14] [n. d.]. PeoplePerHour. https://www.peopleperhour.com.
[15] [n. d.]. Plan Ceibal. https://www.ceibal.edu.uy/en/institucional.
[16] [n. d.]. Rank Likes. http://www.ranklikes.com/.
[17] [n. d.]. RapidWorkers. https://rapidworkers.com/.
[18] [n. d.]. The Social Marketeers. http://www.thesocialmarketeers.org/.
[19] [n. d.]. Upwork Inc. https://www.upwork.com.
[20] [n. d.]. Write a review on Google Play. Google Play Help, https://tinyurl.com/yc9stfy3.
[21] [n. d.]. Zeerk. https://zeerk.com/.
[22] 2018. How Artificial Intelligence detects fake reviews. Scitech Europa, https://tinyurl.com/ycjwtmfw.
[23] Rakesh Agrawal and Ramakrishnan Srikant. 1994. Fast Algorithms for Mining Association Rules in Large Databases. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 487–499. http://dl.acm.org/citation.cfm?id=645920.672836
[24] Leman Akoglu, Rishi Chandy, and Christos Faloutsos. 2013. Opinion Fraud Detection in Online Reviews by Network Effects. In *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, Massachusetts, USA, July 8-11, 2013*.
[25] Tasneem Akolawala. 2018. Google Play Store Removes Millions of Fake Reviews and Bad Apps With New Anti-Spam System. Gadgets360, https://tinyurl.com/ya6g2v9n.
[26] Prudhvi Ratna Badri Satya, Kyumin Lee, Dongwon Lee, Thanh Tran, and Jason (Jiasheng) Zhang. 2016. Uncovering Fake Likers in Online Social Networks. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16)*. ACM, New York, NY, USA, 2365–2370. https://doi.org/10.1145/2983323.2983695
[27] Brian Barrett. 2018. Millions of Android Devices are Vulnerable Right Out of the Box. Wired, https://www.wired.com/story/android-smartphones-vulnerable-out-of-the-box/.
[28] Alex Beutel, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, and Christos Faloutsos. 2013. CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 119–130. https://doi.org/10.1145/2488388.2488400
[29] Dearbhail Bracken-Roche, Emily Bell, Mary Ellen Macdonald, and Eric Racine. 2017. The concept of vulnerabilityin research ethics: an in-depth analysis of policies and guidelines. *Health research policy and systems* 15, 1 (2017), 8.
[30] Kyle Bradshaw. 2018. Play Store's machine learning based anti-spam system removes millions of reviews per week. 9To5Google, https://tinyurl.com/ya3b6xjg.
[31] Elie Bursztein, Borbala Benko, Daniel Margolis, Tadek Pietraszek, Andy Archer, Allan Aquino, Andreas Pitsillidis, and Stefan Savage. 2014. Handcrafted Fraud and Extortion: Manual Account Hijacking in the Wild. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, New York, NY, USA, 347–358. https://doi.org/10.1145/2663716.2663749
[32] Qiang Cao, Xiaowei Yang, Jieqi Yu, and Christopher Palow. 2014. Uncovering Large Groups of Active Malicious Accounts in Online Social Networks. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. ACM, New York, NY, USA, 477–488. https://doi.org/10.1145/2660267.2660269
[33] Kathy Charmaz and Linda Liska Belgrave. 2007. Grounded Theory. *The Blackwell Encyclopedia of Sociology* (2007).
[34] Jason Cipriani. 2016. Google starts filtering fraudulent app reviews from Play Store. ZDNet, https://tinyurl.com/hklb5tk.
[35] Nicholas Confessore, Gabriel Dance, Richard Harris, and Mark Hansen. 2018. The Follower Factory. *The New York Times* (Jan 2018). https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html
[36] Emiliano De Cristofaro, Arik Friedman, Guillaume Jourjon, Mohamed Ali Kaafar, and M. Zubair Shafiq. 2014. Paying for Likes?: Understanding Facebook Like Fraud Using Honeypots. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, New York, NY, USA, 129–136. https://doi.org/10.1145/2663716.2663729
[37] Amir Fayazi, Kyumin Lee, James Caverlee, and Anna Squicciarini. 2015. Uncovering Crowdsourced Manipulation of Online Reviews. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*. ACM, New York, NY, USA, 233–242. https://doi.org/10.1145/2766462.2767742
[38] G Fei, A Mukherjee, B Liu, M Hsu, M Castellanos, and R Ghosh. 2013. Exploiting burstiness in reviews for review spammer detection. In *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*. 175–184.
[39] Fiverr. [n. d.]. https://www.fiverr.com/.
[40] Stephan Günnemann, Nikou Günnemann, and Christos Faloutsos. 2014. Detecting Anomalies in Dynamic Rating Data: A Robust Probabilistic Model for Rating Evolution. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. ACM, New York, NY, USA, 841–850. https://doi.org/10.1145/2623330.2623721
[41] Nestor Hernandez, Mizanur Rahman, Ruben Recabarren, and Bogdan Carbunar. 2018. Fraud De-Anonymization for Fun and Profit. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. ACM, New York, NY, USA, 115–130. https://doi.org/10.1145/3243734.3243770
[42] Atefeh Heydari, Mohammadali Tavakoli, and Naomie Salim. 2016. Detection of Fake Opinions Using Time Series. *Expert Syst. Appl.* 58, C (Oct. 2016), 83–92. https://doi.org/10.1016/j.eswa.2016.03.020
[43] Bryan Hooi, Neil Shah, Alex Beutel, Stephan Günnemann, Leman Akoglu, Mohit Kumar, Disha Makhija, and Christos Faloutsos. 2016. BIRDNEST: Bayesian Inference for Ratings-Fraud Detection. In *Proceedings of the 2016 SIAM International Conference on Data Mining, Miami, Florida, USA, May 5-7, 2016*. 495–503. https://doi.org/10.1137/1.9781611974348.56
[44] Bryan Hooi, Hyun Ah Song, Alex Beutel, Neil Shah, Kijung Shin, and Christos Faloutsos. 2016. FRAUDAR: Bounding Graph Fraud in the Face of Camouflage. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 895–904. https://doi.org/10.1145/2939672.2939747
[45] Danny Yuxing Huang, Doug Grundman, Kurt Thomas, Abhishek Kumar, Elie Bursztein, Kirill Levchenko, and Alex C. Snoeren. 2017. Pinning Down Abuse on Google Maps. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1471–1479. https://doi.org/10.1145/3038912.3052590
[46] Markus Jakobsson. 2018. Secure Remote Attestation. *IACR Cryptology ePrint Archive* 2018 (2018), 31. http://eprint.iacr.org/2018/031
[47] Mark Jansen. 2018. Here's how the Google Play Store detects fake ratings and reviews. Digital Trends, https://tinyurl.com/yc5hvyq5.
[48] Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. 2014. Inferring Strange Behavior from Connectivity Pattern in Social Networks. In *Advances in Knowledge Discovery and Data Mining*, Vincent S. Tseng, Tu Bao Ho, Zhi-Hua Zhou, Arbee L. P. Chen, and Hung-Yu Kao (Eds.). Springer International Publishing, Cham, 126–138.
[49] Nitin Jindal and Bing Liu. 2007. Review Spam Detection. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 1189–1190. https://doi.org/10.1145/1242572.1242759
[50] Parisa Kaghazgaran, Majid Alfifi, and James Caverlee. 2019. TOmCAT: Target-Oriented Crowd Review ATtacks and Countermeasures. In *International AAAI Conference on Web and Social Media, ICWSM*.
[51] Parisa Kaghazgaran, James Caverlee, and Majid Alfifi. 2017. Behavioral Analysis of Review Fraud: Linking Malicious Crowdsourcing to Amazon and Beyond. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017*. 560–563.
[52] Parisa Kaghazgaran, James Caverlee, and Anna Squicciarini. 2018. Combating Crowdsourced Review Manipulators: A Neighborhood-Based Approach. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18)*. ACM, New York, NY, USA, 306–314. https://doi.org/10.1145/3159652.3159726
[53] Santosh KC and Arjun Mukherjee. 2016. On the Temporal Dynamics of Opinion Spamming: Case Studies on Yelp. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 369–379. https://doi.org/10.1145/2872427.2883087
[54] Helen Knapman. 2019. Fake five-star review farms are flooding Amazon with positive comments, says Which? The Sun, https://tinyurl.com/yafthxdd.
[55] Srijan Kumar, Justin Cheng, Jure Leskovec, and V.S. Subrahmanian. 2017. An Army of Me: Sockpuppets in Online Discussion Communities. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 857–866. https://doi.org/10.1145/3038912.3052677
[56] Srijan Kumar and Neil Shah. 2018. False Information on Web and Social Media: A Survey. *CoRR* abs/1804.08559 (2018). arXiv:1804.08559 http://arxiv.org/abs/

1804.08559
[57] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. 2014. *Mining of Massive Datasets* (2nd ed.). Cambridge University Press, New York, NY, USA.
[58] Huayi Li, Zhiyuan Chen, Arjun Mukherjee, Bing Liu, and Jidong Shao. 2015. Analyzing and Detecting Opinion Spam on a Large-scale Dataset via Temporal and Spatial Patterns. In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*. 634–637.
[59] Huayi Li, Geli Fei, Shuai Wang, Bing Liu, Weixiang Shao, Arjun Mukherjee, and Jidong Shao. 2017. Bimodal Distribution and Co-Bursting in Review Spam Detection. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1063–1072. https://doi.org/10.1145/3038912.3052582
[60] Shanshan Li, James Caverlee, Wei Niu, and Parisa Kaghazgaran. 2017. Crowdsourced App Review Manipulation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1137–1140.
[61] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Lauw. 2010. Detecting Product Review Spammers Using Rating Behaviors. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*. ACM, New York, NY, USA, 939–948. https://doi.org/10.1145/1871437.1871557
[62] Damon McCoy, Andreas Pitsillidis, Jordan Grant, Nicholas Weaver, Christian Kreibich, Brian Krebs, Geoffrey Voelker, Stefan Savage, and Kirill Levchenko. 2012. PharmaLeaks: Understanding the Business of Online Pharmaceutical Affiliate Programs. In *Presented as part of the 21st USENIX Security Symposium (USENIX Security 12)*. USENIX, Bellevue, WA, 1–16. https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/mccoy
[63] Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Spotting Opinion Spammers Using Behavioral Footprints. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*. ACM, New York, NY, USA, 632–640. https://doi.org/10.1145/2487575.2487580
[64] Arjun Mukherjee, Bing Liu, and Natalie Glance. 2012. Spotting Fake Reviewer Groups in Consumer Reviews. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*. ACM, New York, NY, USA, 191–200. https://doi.org/10.1145/2187836.2187863
[65] Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S. Glance. 2013. What Yelp Fake Review Filter Might Be Doing?. In *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, Massachusetts, USA, July 8-11, 2013*.
[66] Kazushi Nagayama and Andrew Ahn. 2016. Keeping the Play Store trusted: fighting fraud and spam installs. Android Developers Blog, https://android-developers.googleblog.com/2016/10/keeping-the-play-store-trusted-fighting-fraud-and-spam-installs.html.
[67] Shirin Nilizadeh, Francois Labrèche, Alireza Sedighian, Ali Zand, José Fernandez, Christopher Kruegel, Gianluca Stringhini, and Giovanni Vigna. 2017. POISED: Spotting Twitter Spam Off the Beaten Paths. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)*. ACM, New York, NY, USA, 1159–1174. https://doi.org/10.1145/3133956.3134055
[68] Shashank Pandit, Duen Horng Chau, Samuel Wang, and Christos Faloutsos. 2007. Netprobe: A Fast and Scalable System for Fraud Detection in Online Auction Networks. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. 201–210.
[69] Youngsam Park, Jackie Jones, Damon McCoy, Elaine Shi, and Markus Jakobsson. 2014. Scambaiter: Understanding Targeted Nigerian Scams on Craigslist. In *21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, California, USA, February 23-26, 2014*.
[70] TE Parliament. 2016. Regulation (eu) 2016/679 of the european parliament and of the council. *Official Journal of the European Union* (2016).
[71] Sarah Perez. 2016. Amazon bans incentivized reviews tied to free or discounted products. Tech Crunch, https://tinyurl.com/zgn9sq3.
[72] Sarah Perez. 2019. Google Play is changing how app ratings work. Tech Crunch https://techcrunch.com/2019/05/08/google-play-is-changing-how-app-ratings-work/.
[73] Rebecca S. Portnoff, Sadia Afroz, Greg Durrett, Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, Damon McCoy, Kirill Levchenko, and Vern Paxson. 2017. Tools for Automated Analysis of Cybercriminal Markets. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 657–666. https://doi.org/10.1145/3038912.3052600
[74] Shebuti Rayana and Leman Akoglu. 2015. Collective Opinion Spam Detection: Bridging Review Networks and Metadata. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. ACM, New York, NY, USA, 985–994. https://doi.org/10.1145/2783258.2783370
[75] Brian Reigh. 2017. Fake reviews on the Play Store reportedly growing and getting smarter. Android Authority, https://tinyurl.com/yc4fo9dk.
[76] Vlad Sandulescu and Martin Ester. 2015. Detecting Singleton Review Spammers Using Semantic Similarity. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)*. ACM, New York, NY, USA, 971–976. https://doi.org/10.1145/2740908.2742570
[77] Jonghyuk Song, Sangho Lee, and Jong Kim. 2015. CrowdTarget: Target-based Detection of Crowdturfing in Online Social Networks. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. ACM, New York, NY, USA, 793–804. https://doi.org/10.1145/2810103.2813661
[78] Kevin Springborn and Paul Barford. 2013. Impression Fraud in On-line Advertising via Pay-Per-View Networks. In *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. USENIX, Washington, D.C., 211–226. https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/springborn
[79] Tao Stein, Erdong Chen, and Karan Mangla. 2011. Facebook Immune System. In *Proceedings of the 4th Workshop on Social Network Systems (SNS '11)*. ACM, New York, NY, USA, Article 8, 8 pages. https://doi.org/10.1145/1989656.1989664
[80] Rebecca Stewart. 2019. Instagram's fake follower purge has had 'little effect' on fraudulent influencers. The Drum, https://tinyurl.com/y7ja52h5.
[81] Gianluca Stringhini, Pierre Mourlanne, Gregoire Jacob, Manuel Egele, Christopher Kruegel, and Giovanni Vigna. 2015. EVILCOHORT: Detecting Communities of Malicious Accounts on Online Services. In *24th USENIX Security Symposium (USENIX Security 15)*. USENIX Association, Washington, D.C., 563–578. https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/stringhini
[82] Gianluca Stringhini, Gang Wang, Manuel Egele, Christopher Kruegel, Giovanni Vigna, Haitao Zheng, and Ben Y. Zhao. 2013. Follow the Green: Growth and Dynamics in Twitter Follower Markets. In *Proceedings of the 2013 Conference on Internet Measurement Conference (IMC '13)*. ACM, New York, NY, USA, 163–176. https://doi.org/10.1145/2504730.2504731
[83] Kurt Thomas, Chris Grier, Dawn Song, and Vern Paxson. 2011. Suspended Accounts in Retrospect: An Analysis of Twitter Spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11)*. ACM, New York, NY, USA, 243–258. https://doi.org/10.1145/2068816.2068840
[84] Kurt Thomas, Dmytro Iatskiv, Elie Bursztein, Tadek Pietraszek, Chris Grier, and Damon McCoy. 2014. Dialing Back Abuse on Phone Verified Accounts. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. ACM, New York, NY, USA, 465–476. https://doi.org/10.1145/2660267.2660321
[85] Kurt Thomas, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. 2013. Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse. In *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. USENIX, Washington, D.C., 195–210. https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/thomas
[86] Tian Tian, Jun Zhu, Fen Xia, Xin Zhuang, and Tong Zhang. 2015. Crowd Fraud Detection in Internet Advertising. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1100–1110. https://doi.org/10.1145/2736277.2741136
[87] Robert Triggs. 2017. Flagship? Mid-range? Budget? Find the best phone for you. AndroidAuthority, https://www.androidauthority.com/flagship-mid-range-budget-best-phone-815330/.
[88] David Y. Wang, Matthew Der, Mohammad Karami, Lawrence Saul, Damon McCoy, Stefan Savage, and Geoffrey M. Voelker. 2014. Search + Seizure: The Effectiveness of Interventions on SEO Campaigns. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, New York, NY, USA, 359–372. https://doi.org/10.1145/2663716.2663738
[89] Guan Wang, Sihong Xie, Bing Liu, and Philip S. Yu. 2011. Review Graph Based Online Store Review Spammer Detection. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining (ICDM '11)*. IEEE Computer Society, Washington, DC, USA, 1242–1247. https://doi.org/10.1109/ICDM.2011.124
[90] Tom Warren. 2017. 41 percent of Android phones are vulnerable to 'devastating' Wi-Fi attack . The Verge, https://www.theverge.com/2017/10/16/16481252/wi-fi-hack-attack-android-wpa-2-details.
[91] Rolfe Winkler and Andrea Fuller. 2019. How Companies Secretly Boost Their Glassdoor Ratings. The Wall Street Journal, https://tinyurl.com/yc7t2nk4.
[92] Sihong Xie, Guan Wang, Shuyang Lin, and Philip S. Yu. 2012. Review Spam Detection via Temporal Pattern Discovery. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*. ACM, New York, NY, USA, 823–831. https://doi.org/10.1145/2339530.2339662
[93] Zhen Xie and Sencun Zhu. 2014. GroupTie: Toward Hidden Collusion Group Discovery in App Stores. In *Proceedings of the 2014 ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '14)*. ACM, New York, NY, USA, 153–164. https://doi.org/10.1145/2627393.2627409
[94] Zhen Xie and Sencun Zhu. 2015. AppWatcher: Unveiling the Underground Market of Trading Mobile App Reviews. In *Proceedings of the 8th ACM Conference*

on Security & Privacy in Wireless and Mobile Networks (WiSec '15). ACM, New York, NY, USA, Article 10, 11 pages. https://doi.org/10.1145/2766498.2766510

[95] Zhen Xie, Sencun Zhu, Qing Li, and Wenjing Wang. 2016. You Can Promote, but You Can'T Hide: Large-scale Abused App Detection in Mobile App Stores. In *Proceedings of the 32nd Annual Conference on Computer Security Applications (ACSAC '16)*. ACM, New York, NY, USA, 374–385. https://doi.org/10.1145/2991079.2991099

[96] Chang Xu. 2013. Detecting Collusive Spammers in Online Review Communities. In *Proceedings of the Sixth Workshop on Ph.D. Students in Information and Knowledge Management (PIKM '13)*. ACM, New York, NY, USA, 33–40. https://doi.org/10.1145/2513166.2513176

[97] Chang Xu and Jie Zhang. 2015. Combating Product Review Spam Campaigns via Multiple Heterogeneous Pairwise Features. In *Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, Canada, April 30 - May 2, 2015*. 172–180.

[98] Fei Ye and Kazushi Nagayama. 2018. In reviews we trust, Making Google Play ratings and reviews more trustworthy. Android Developers Blog, https://tinyurl.com/yb67uy73.

[99] Junting Ye and Leman Akoglu. 2015. Discovering Opinion Spammer Groups by Network Footprints. In *Proceedings of the 2015 ACM on Conference on Online Social Networks (COSN '15)*. ACM, New York, NY, USA, 97–97. https://doi.org/10.1145/2817946.2820606

[100] Junting Ye, Santhosh Kumar, and Leman Akoglu. 2016. Temporal Opinion Spam Detection by Multivariate Indicative Signals. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17-20, 2016*. 743–746.

[101] Haizhong Zheng, Minhui Xue, Hao Lu, Shuang Hao, Haojin Zhu, Xiaohui Liang, and Keith W. Ross. 2018. Smoke Screener or Straight Shooter: Detecting Elite Sybil Attacks in User-Review Social Networks. In *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*.

A RECRUITMENT MATERIAL

We are researchers from FIU, a university in the US, looking for freelancers with provable App Search Optimization (ASO) expertise in Google Play, willing to participate in a survey. We will ask you questions about your experience working as an app search optimization (ASO) freelancer. We are conducting this survey part of an effort to increase our understanding of how the ASO process optimizes mobile apps.
Your participation in this study is confidential. We will never reveal to anyone any information that may be linked to you, including the fact that you participated in our study.
Your participation is completely voluntary and you may choose to withdraw at any time or not answer questions that you do not feel comfortable answering. If you agree to participate, please send me an e-mail at mrahm031@fiu.edu.

Figure 17: Recruitment message sent to each identified ASO worker.

Figure 17 shows the recruitment message that we sent to each ASO worker that we identified. Figure 18 shows the script that we read to each ASO worker who replied to the recruitment message and qualified for our study.

Thank you for agreeing to participate in this study. My name is Mizanur Rahman, and I am a student at FIU.
In this study, I would like to ask you questions about your experience working as an app search optimization, or ASO, freelancer. The questions will explore your perspectives on ASO strategies in Google Play. The study should take up to 1 hour.
If you decide to participate, you will be one of up to 100 people in this study. We will pay you $5 for every 15 minutes of your time, that is, $20 if we talk for 1 hour.

The benefits of your participation include receiving feedback on vulnerabilities that your strategies may have, and also helping us better understand and model the app search optimization process in Google Play.
Please note that some of the questions that we will ask you, may be upsetting. You can skip any questions you don't want to answer, or stop the study entirely, at any time. Your participation in this study is voluntary. You are free to participate in the study or withdraw your consent at any time during the study. Your withdrawal or lack of participation will not affect any benefits to which you are otherwise entitled.
In addition, once we publish our results, other parties, including Google, may use them to try to develop techniques to detect your activities. We note that you already run this risk, even if you do not participate in our study. This is because other developers who hire you, may work for Google, and could use data that they collect from you, to directly impact your activities, e.g., block your accounts or remove the reviews that you write. However, we will never do this.

Please be assured that your participation in this study is confidential. We will keep the records of this study private and protected to the fullest extent provided by law. In any sort of report we might publish, we will not include any information that will make it possible to identify you. We will store records securely, and only the researcher team will have access to the records. However, your records may be reviewed for audit purposes by authorized University or other agents who will be bound by the same provisions of confidentiality.
Now, please read the consent form at the following link, https://fiu.qualtrics.com/jfe/form/SV_8wYphZYyVQ4lTz7, and tell me if you want to participate in the study. If you want to participate, please click on the button at the end of the form, that says "I consent". Before we begin, do you have any questions?

Figure 18: Introduction script read by interviewer to ASO workers who responded to the recruitment message, and qualified for the study, before starting the study.