Do You Get What You Pay For? Comparing The Privacy Behaviors of Free vs. Paid Apps

Catherine Han[1], Irwin Reyes[2], Amit Elazari Bar On[1], Joel Reardon[3], Álvaro Feal[4,5], Kenneth A. Bamberger[1], Serge Egelman[1,2], Narseo Vallina-Rodriguez[2,4]

[1]UC Berkeley, [2]International Computer Science Institute, [3]University of Calgary, [4]IMDEA Networks Institute, [5]Universidad Carlos III de Madrid

***Abstract*—It is commonly assumed that the availability of "free" mobile apps comes at the cost of consumer privacy, and that paying for apps could offer consumers protection from behavioral advertising and long-term tracking. This work empirically evaluates the validity of this assumption by investigating the degree to which "free" apps and their paid premium versions differ in their bundled code, their declared permissions, and their data collection behaviors and privacy practices.**

We compare pairs of free and paid apps using a combination of static and dynamic analysis. We also examine the differences in the privacy policies within pairs. We rely on static analysis to determine the requested permissions and third-party SDKs in each app; we use dynamic analysis to detect sensitive data collected by remote services at the network traffic level; and we compare text versions of privacy policies to identify differences in the disclosure of data collection behaviors. In total, we analyzed 1,505 pairs of free Android apps and their paid counterparts, with free apps randomly drawn from the Google Play Store's category-level top charts.

Our results show that over our corpus of free and paid pairs, there is no clear evidence that paying for an app will guarantee protection from extensive data collection. Specifically, 48% of the paid versions reused all of the same third-party libraries as their free versions, while 56% of the paid versions inherited all of the free versions' Android permissions to access sensitive device resources (when considering free apps that include at least one third-party library and request at least one Android permission). Additionally, our dynamic analysis reveals that 38% of the paid apps exhibit all of the same data collection and transmission behaviors as their free counterparts. Our exploration of privacy policies reveals that only 45% of the pairs provide a privacy policy of some sort, and less than 1% of the pairs overall have policies that differ between free and paid versions.

I. Introduction

Mobile app marketplaces offer consumers a large selection of products: as of 2017, the Google Play Store offered approximately 3,500,000 available Android apps [9], while Apple's iOS App Store lists approximately 2,200,000 apps [11]. Many apps are available free of charge, while others require consumers to pay a one-time fee to download them: roughly 7.8% of apps in the Google Play Store require payment [8], as compared with 6% of iOS apps [19].

Common app pricing models include free, paid, "freemium," and "paidmium" [25]. Free apps are available at no up-front cost and do not offer in-app purchases, while paid apps require the user to pay for the initial download. The "freemium" model raises revenue primarily from in-app purchases, while the app itself comes at no up-front cost. Likewise, the "paidmium" model also relies heavily on in-app purchases, though the app itself also comes at a cost.

In aggregate, free apps attract over 10 times the volume of downloads as paid apps [21]. Developers of free apps rely on other ways to generate revenue besides directly collecting money from paying consumers. Partnering with advertising networks to serve ads to users is one such major alternative revenue stream for free apps. Google's AdMob, for instance, is found in more than 1 million apps, and has yielded more than $1 billion collectively to developers [14].

It has become apparent that users often trade their privacy for these "free" apps [10]. The question, however, remains unanswered for paid apps—are consumers of paid apps truly safe from extensive user profiling and tracking? Users paying for apps expect them to be of higher quality compared to free versions [20], and a common selling point to that end is the removal of ads in paid versions. Even media outlets have reinforced these consumer expectations, stating that paid apps have better security and privacy assurances than free apps [7]. The lack of ads, however, might give the false assurance of being free from extensive data collection, a practice often associated with user tracking for the purpose of ad targeting. That is, even if an app does not display ads, it may still perform invasive tracking for the purpose of serving highly-targeted ads in *other* apps.

Regulators have been pushing tech companies for increased transparency about their data collection practices. In a recent landmark ruling against Google, the French data regulator CNIL levied a 50 million Euro fine for a breach of the EU General Data Protection Regulation's (GDPR) transparency and informed consent requirements concerning data collection for personalized ads [12]. Recent privacy legislation passed in California [3] further exemplifies rising regulatory and public concern surrounding consumers' ability to make informed decisions about their digital safety.

Exploring if app behaviors comport with user expectations and if "ad-free" representations may be misleading consumers can inform regulators, policymakers, and consumers alike.

Potentially misleading representations may run afoul of the FTC's prohibitions against deceptive practices and state laws prohibiting unfair business practices, as well as general privacy regulations, such as the GDPR and CCPA. Finally, such inquiry can also inform economic models exploring the viability of "pay for privacy" consumer protection models [6].

To that end, we explore the differences and similarities in the implementation and data collection practices of free Android apps and their paid counterparts offered on the Google Play Store, across 1,505 pairs of apps. On average, at least 10,000 users have installed each pair of apps. We measured their prospective differences along four key aspects: different third-party libraries—which may be used for advertising and tracking—bundled with the apps, the nature of the permissions they access, the types of sensitive data shared with third-party services, and the differences in privacy policies offered by each version of the app.

From our results, we make four key observations concerning pairs whose free versions exhibited at least one of each metric, respectively:

- 48% of paid versions include all the same third-party libraries as the free versions.
- 56% of paid versions inherit all the same permissions from their free versions.
- Most of the app pairs (free and paid versions of the same app) exhibit an all-or-nothing model, in which 38% of paid versions exhibit the same data collection behaviors as the free versions and 45% have none of it.
- In spite of increasing regulatory pressure to improve transparency, only 45% of the pairs in our corpus provide a privacy policy, and less than 1% of corpus pairs have policies that differ between the free and paid apps.

II. Related Work

This section is a summary of relevant prior work on the analysis of mobile app privacy and on the compared analysis of free and paid apps.

A. Analysis of Mobile App Privacy

Previous work has analyzed the collection of personal information through both static and dynamic analysis. Static analysis consists of evaluating software without execution [5], whereas dynamic analysis focuses on tracking the transmission of sensitive information at runtime. Runtime behavior is often paired with the observation of network traffic to identify personal data dissemination. To automate the process of such analysis, researchers have developed several tools to not only simulate user interaction, but also give summaries of network traffic [16], [22], [27].

The research in this paper combines static and dynamic analysis methods introduced in previous work [16], as well as broadening the analysis of mobile apps by loosening the constraints on our corpus in two key ways: (1) including both free and paid apps for direct comparison to one another; and (2) having a broader scope than apps that are specifically designed for children and families.

B. Comparison of Free and Paid Apps

Prior research also sought to examine the relationship between free and paid mobile apps. Researchers have used static analysis to examine the prevalence of tracking libraries and their data collection behaviors in free and paid apps [23]. Other studies have investigated vulnerabilities associated with the maintenance of software and inclusion of third-party libraries in apps with different monetization models [22], [24].

Earlier works center on a broad comparison of a body of free apps with a body of paid apps. Our work offers a novel view on the comparison between free and paid apps by presenting a precise, side-by-side analysis of specifically constructed pairs of apps: a free app and its paid "premium" version. Our approach compares apps that are directly related to one another: the same general app from the same developer, but offered separately as free and paid versions.

C. Paying for Privacy

There has been previous research to determine the value of privacy to consumers and how privacy benefits could be leveraged as a selling point by businesses [26]. Survey data indicates that online consumers place great importance on having insight and control over how companies handle their personal data [18].

With growing concern surrounding online privacy, prior work indicates that some consumers are indeed willing to pay a premium for products that protect privacy [26]. In our future analysis, we expect to conduct a survey that explores a complementary perspective on the price of privacy: observing users' perceptions of privacy behaviors of paid and free app versions, specifically considering whether, when apps are advertised as "ad-free," most consumers believe that this is synonymous with "better privacy." With this, we can identify ways in which the behavior of apps differ from users' expectations, ultimately determining if the "ad-free" representation misleads consumers.

III. Methodology

The goal of this paper is to compare the implementations and data collection practices of free Android apps and their paid counterparts offered on the Google Play Store. In this analysis, we generalize different app monetization models into two overarching categories: we define "free apps" as those that are available for download on the app store at no up-front cost; and we define "paid apps" as apps that require a one-time payment to download. Our focus is on paid apps in which the consumer pays for the app as a single discrete product, not for a continuously renewed service. We acknowledge that apps may employ other monetization strategies, such as the "freemium" or "paidmium" models, in which potentially recurring in-app purchases generate revenue for the developer. Though we are aware that some apps do offer in-app purchases to disable ads, these are beyond the scope of this study.

The Google Play Store does not reliably link free apps to their paid versions, or even indicate if a corresponding paid version exists at all. Therefore, we first developed our own

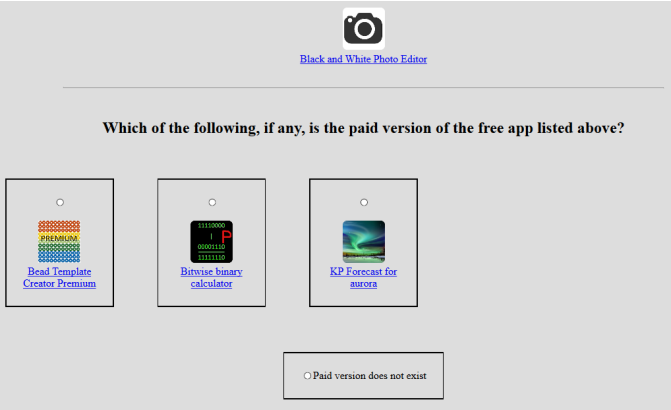

Fig. 1. MTurk task for selecting the paid version of a free app.

method to identify pairs of free and paid versions of the same general app (e.g., "Quick PDF Scanner FREE" and "Quick PDF Scanner PRO"). We evaluate and compare the behavior of these pairs using both static and dynamic analysis techniques.

A. App Corpus

We formed our app corpus by consulting the AppCensus database,[1] which is regularly updated by crawling the "Top Free" charts in each of the Play Store's categories. We then created a labeling task on Amazon Mechanical Turk. We presented workers with a free app and a list of all paid apps from the same developer (Figure 1). In order to increase the likelihood of valid free/paid pairings, we only presented workers free apps whose titles or package names contained the words "free" or "lite," as those keywords would suggest that a "paid" or "full" version exists. If the free app did not have a corresponding paid version, workers were instructed to select the "None" option. We presented each free app to three different workers, then manually adjudicated the responses for agreement and correctness. Workers were paid $0.10 for each match in consensus with the others.

This process yielded 1,583 potential pairs to analyze. To keep costs under control, we discarded 33 pairs whose paid app was priced more than $10. Due to region-locking or device compatibility issues, we were unable to purchase, download, or install apps in 45 additional pairs. In the end, we assembled a corpus of 1,505 pairs of free apps and their paid counterparts.[2] All of them were available on the Google Play Store as of December 2018 and represented 1,159 unique developers. The free apps in this corpus had a median install count of 10,000 (reported as the lower bound of a range in the Play Store, *e.g.,* "10,000+").

B. Evaluating Apps

We looked for similarities across pairs of free and paid apps along four dimensions: (1) the portion of third-party packages found in the free app that are also included in the paid version; (2) the portion of Android permissions declared by the free app also declared by the paid app; (3) the portion of sensitive network transmissions performed by the free app also seen in the paid app; and (4) the differences in the privacy policies within pairs. We believe these four aspects are a good representation of apps' data collection and sharing behaviors. We employed the following methods to evaluate these:

Static Analysis: We used the Android Asset Packaging Tool *(aapt)* [1] to extract the permissions apps request for various device resources. We then identified differences in permissions within pairs of free and paid apps. Additionally, we relied on Apktool [2] to examine apps' file structures for the package names that comprise the app. We identified third-party libraries by eliminating package names that share the same first two levels as the app package (*i.e.,* disregarding code belonging to the core app). This revealed what third-party libraries—possibly used for monetization and data collection—are shared between free apps and their paid counterparts.

Dynamic Analysis: We used dynamic analysis methods derived from earlier work [16] to automatically evaluate apps by executing them in an instrumented environment (deployed on identical Nexus 5X smartphones) that captures apps' network traffic. We relied on the Android SDK's Application Exerciser Monkey tool [15] to automatically explore apps without user intervention. Although there is no guarantee that paired apps have identical user interfaces, we controlled for differences in app execution by providing both apps with the same random input stream at the same time. This increases the likelihood that observed differences in app behavior arose from implementation differences, rather than differences in input.

At the end of each paired execution, we analyzed the captured network data to identify which sensitive data types were sent to which remote services—services that could be for ads, profiling, crash reporting, etc. We focused on detecting the transmission of sensitive data that can be used to uniquely track a user over time and across different services: persistent identifiers, such as the Android Advertising ID (AAID), IMEI, and Wi-Fi MAC address; as well as personally identifiable information (PII), such as geolocation, name, and phone number. In order to detect the transmission of sensitive data, we not only used simple string matching, but also relied on methods from previous work [16] for the decoding of obfuscated network traffic, which uses regular expressions formed from the manual inspection of different data encoding schemes.

Privacy Policies: We obtained privacy policies for apps in our corpus by following the "Privacy Policy" link on each app's Play Store listing. Using a headless Firefox instance controlled by Selenium, we downloaded the HTML content for the linked policies. A BeautifulSoup-powered script subsequently stripped away inline scripts and extracted the body of the policy in plain text. We compared the text versions of the privacy policies linked from free and paid apps to identify differences in their disclosed data collection behaviors.

[1] https://search.appcensus.io/

[2] https://github.com/io-reyes/play-store-purchase/blob/master/data/pairs-conpro.csv

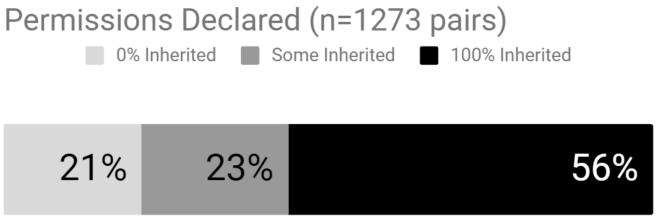

Fig. 2. Frequency of Android permissions inherited between free/paid pairs, where the free app requested at least one Android permission.

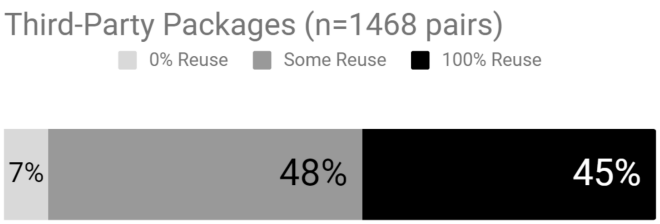

Fig. 3. Frequency of third-party package reuse among free/paid pairs, where the free app had at least one third-party package.

IV. Limitations

We acknowledge that free and paid versions of an app pair may have UI differences, and that those variations may produce differences in execution despite receiving the same input stream. We thus represent our results as a "best-effort" at-scale comparison of pairs of apps.

In constructing our corpus, we limited our paid app selection to those that were no more than $10, as some apps were unreasonably priced (*e.g.,* more than $100) and thus had few actual downloads, if any. App pricing, however, is heavily skewed towards less expensive apps, so omitting these apps should have minimal impact on our results; enforcing this threshold resulted in the exclusion of only 33 pairs (2.1%) out of the initial 1,583 pairs identified.

We applied a simple strict definition for "paid" apps, distinguishing free and paid only by observing the price associated with the installation of the app. We disregarded any in-app purchases, e-commerce, or recurring transactions that may occur once the app is installed. As of 2018, approximately 6% of the Google Play Store consists of paid applications, in which paid is defined as having a price associated with the installation of the app itself [4].

V. Analysis

This work focuses on measurable differences in privacy between free and paid versions, so all presented comparisons are conditioned on the free app having at least one observation for any of the corresponding metrics. In each of the following analyses, we disregard pairs in which the free app had no third-party packages, no permission requests, or no sensitive data shared with a third-party service, respectively.

We note that there are indeed some paid apps that have observations along these dimensions that were not seen in their free counterparts. However, these represent only a small portion of our corpus: 175 paid apps requested permissions not declared by their free versions, and 67 paid apps transmitted data not observed in the free release. We stress that our analysis quantifies the degree to which free apps' behaviors along these three metrics are carried over to their corresponding paid versions.

A. Declared Android Permissions

The Android permission system serves to protect user privacy. Apps must hold appropriate permissions to use various device resources (*e.g.,* Internet access and information about the device) and access sensitive user data (*e.g.,* phone number). A subset of Android's permissions are deemed "dangerous" because they guard sensitive resources that directly affect user security and privacy, such as the contact list or location information [13]. All of the resources categorized as dangerous permissions require user consent at runtime.

Of the 1,505 pairs in our corpus, 1,273 had free versions that declare at least one Android permission (either regular permissions or "dangerous"). In 56% of these pairs, the paid version (Figure 2) declared all of the same permissions held by the free version. That is, paid apps held all the same privileges as free versions in a majority of the time that any permissions are declared. The most common permissions that both the paid and free versions requested were the ones that gave access to network state, Internet, and writing to external storage; of these, the permission that safeguards reading and writing to external storage is deemed as "dangerous" under Android's own categorization.

We note too that 21% of pairs have paid apps that do not request any of the permissions declared by their corresponding free versions. This suggests potential over-permissioning of free apps in these cases, in which free apps hold permissions that may not be necessary for those apps' core functionality. Of these permissions, the most common one exclusive to the free version was for Internet access. For the free/paid pairs that had no permissions overlap at all, 20% of the permissions held exclusively by the free version were "dangerous" Android permissions. Overall, this implies that free apps will likely have access to permissions that the paid app does not, putting users' privacy at higher risk by requesting unnecessary permissions.

B. Bundled Third-Party Packages

The use of third-party code is common practice in software engineering to expedite development. In mobile apps, third-party libraries allow for pre-built functionality like graphics rendering, advertising, and analytics, among others. Third-party code bundled in apps has the same privileges as the host app, and can access all the same device resources and personal data available to the host app.

Of the 1,505 pairs in our corpus, 1,468 had at least one third-party package in the free version. Of these (Figure 3), we observed that 45% of paid apps contained the same third-party libraries as the free versions, while 7% of paid apps showed no third-party libraries carried over from the free version. The remaining 48% of paid apps had varying degrees of third-party

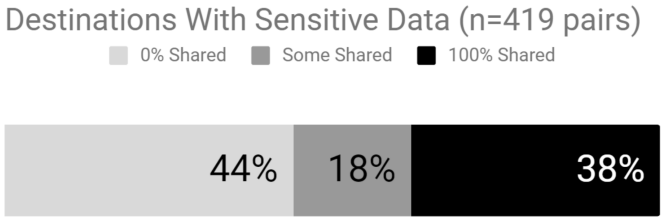

Fig. 4. Frequency of unique domain destinations shared between free/paid pairs, where the free app transmitted sensitive data to at least one domain.

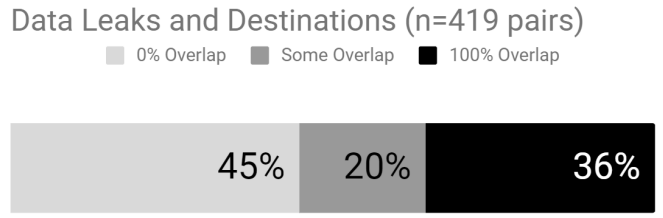

Fig. 5. Frequency of unique sensitive data type-domain pairs shared between free/paid pairs of apps, where the free app transmitted sensitive data to at least one remote domain.

library reuse from the free version to the paid version. This suggests that paid apps are likely to contain most, if not all, of the same third-party libraries as the free versions. Although we acknowledge that our analysis does not account for third-party libraries included but not actually executed (i.e., dead code), these results show that developers of paid apps have little motivation to remove externally-produced code in paid apps. This may leave paying consumers exposed to the same potential for third-party data collection as found in free apps.

Based upon the library categorizations of LibRadar [17], we analyzed the types of third-party libraries present in free and paid versions of apps, focusing our attention on libraries associated with libraries labeled as "Advertising" and "Analytics"; some of the common libraries categorized as Advertising include `vungle.com`, `mopub.com`, `applovin.com`, `ads.google.com`, and `inmobi.com`. Similarly, prevalent instances of Mobile Analytics libraries include `sdk.flurry.com`, `analytics.google.com`, and `android.crashlytics.com`.

Focusing on advertising libraries specifically, LibRadar detected at least one ad library present in either the free or paid release (or both) in 831 pairs. Of those, there were 802 free apps where ad libraries were detected, while only 408 paid apps were found to contain ad libraries. This suggests that ad libraries are most likely present in either free versions of apps exclusively, or to a lesser extent, in both the free and paid versions of an app.

C. Network Transmissions

Third-party services bundled in apps routinely collect various data from users and their devices. For example, crash reporting services might gather hardware specifications and usage telemetry to help developers debug their apps, while advertising networks collect persistent identifiers and personal information to better target users with ads relevant to them. By observing all of the network traffic associated with an app, we can discern the types of sensitive data being transmitted (*e.g.,* Android Advertising ID, user e-mail, geolocation, etc.) and the recipient of that data.

Domain Name	$Free \wedge \overline{Paid}$	$Free \wedge Paid$
unity3d.com	22	63
facebook.com	23	50
flurry.com	13	42
crashlytics.com	4	37
onesignal.com	2	14
applifier.com	7	9

TABLE I
THIRD-PARTY DOMAINS CONTACTED BY AT LEAST TEN APPS AND WHICH TEND TO REMAIN ACTIVE IN THE PAID VERSION.

Among the 1,505 pairs of apps that we examined, there were 419 pairs in which the free version transmitted sensitive data to online services over the Internet. Out of these 419 pairs, we observed that 44% of these pairs' paid versions (Figure 4) did not communicate with any of the domains that the free version did, while 18% shared some destinations with the free version. Conversely, 38% of these pairs' paid versions communicated with all of the same domains as the free version. Overall, there is a near equal distribution among the two ends at "almost no shared" and "almost all shared" domains. When considering sensitive data type and domain pairs instead of only domains, the distribution was nearly identical (Figure 5).

We further investigated the particular *domains* that lie on these two extremes of possible behavior. That is, which domains are more likely to *receive* data from paid versions versus those more likely to be *removed* or deactivated in the paid version. We narrowed the search for those observed transmitting PII and which were contacted by more than ten apps. We compared the set of domains contacted across each app pair. We counted whether the transmission was observed exclusively in the paid version, exclusively in the free version, or appeared in both. Tables I and II illustrate our results, the former listing domains that remain active in the paid version and the latter listing domains that are not. We did not observe any domains used by more than one app pair that were more likely to be active exclusively in the paid version. When counting transmissions, we omit the sub-domain from the domain name; *e.g.,* transmissions to both `api.vungle.com` and `ingest.vungle.com` are counted as `vungle.com`.

Table I presents a subset of third-party domains that received PII from apps in our experiment. This list contains only domains contacted by at least ten different apps and which tend to *remain* in the paid version. This means that more free-paid app pairs contacted these domain in the paid version than exclusively in the free version. Unity is a combination of game engine along with associated advertising and analytics, and we see that it sends the Android ID, the Android

Domain Name	$Free \wedge \overline{Paid}$	$Free \wedge Paid$
chartboost.com	31	10
manage.com	26	0
liftoff.io	24	1
mopub.com	20	1
adcolony.com	19	4
applovin.com	17	8
adjust.com	16	6
amazon-adsystem.com	15	0
appbaqend.com	11	0
startappservice.com	10	0
supersonicads.com	10	2
appsflyer.com	8	6
kochava.com	8	3
vungle.com	8	3
heyzap.com	7	6

TABLE II
THIRD-PARTY DOMAINS CONTACTED BY AT LEAST TEN APPS AND WHICH TEND TO BE DEACTIVATED IN THE PAID VERSION.

Advertising ID, the IMEI, and the Wi-Fi MAC addresses in both versions. Applifier, a Unity-owned entity that began as a cross-promotional network for apps, we observed receiving the Android Advertising ID.

Table II presents the list of third-party domains that received PII, were contacted by at least ten apps, and which tended to be *removed* in the paid version. This presents the opposite extreme as Table I. In contrast to the domains in Table I, Table II contains fewer analytics companies and more explicit advertising companies, *e.g.,* ones that serve advertising impressions to end users. Nonetheless, the data show that many paid apps still transmit personal information to advertisers.

D. Privacy Policies

Google Play allows application developers to provide privacy policies in their apps' Google Play Store listings. We implemented a crawler to fetch the privacy policy for each analyzed app in our dataset. Ultimately, we were only able to download privacy policies for 45% of the corpus. Of the privacy policies that we could not find, the vast majority (87%) were due to broken links (HTTP 404 errors). These results alone illustrate how absurd it is to expect users to make informed decisions about their online privacy.

In the end, we were able to examine 739 app pairs for which we found a privacy policy for both the free and paid versions. We examined a pair of policies by first performing a `diff`, and then manually examining any differences. We discarded the differences caused by Javascript code in the HTML and those that differ only in the title of the page and not the the content. We found nine pairs of apps in which the paid version says that there is no data collection whatsoever and three pairs in which the paid version collects less data than the free version. This analysis shows that the vast majority (98% of 739) of pairs of free and paid apps either do not offer a privacy policy for both versions of the app or the policies are identical. We acknowledge that our methods don't take into account the possibility that a single policy discloses behaviors for both free and paid versions (*i.e.,* in different sections within the same text), so further text similarity analysis is required. However, a majority of pairs in our corpus still lack privacy policies for at least one of the apps.

VI. DISCUSSION

Based on our analysis, in roughly half the cases, free and paid versions of the same app tend to bundle the same third-party code, request the same permissions, and share the same personal information with third parties. This observation runs counter to the general belief that paying for the app protects the consumer from extensive data collection and tracking. It is even more troubling that there is no easy way for consumers to determine whether a given paid app actually affords greater privacy protections than its free counterpart.

In a small minority of free and paid pairs, however, the paid app did request fewer permissions (Figure 2). This suggests that at least some app developers have made an effort to make sure that paid versions are less intrusive and likely to meet consumer expectations.

Of the three metrics we used in our analysis, the nature of the sensitive data being shared with third parties is the most critical. This data shows that many apps, regardless of their monetization model, are sharing sensitive data with third parties, potentially defying consumer expectations. During our analysis we found that 38% of paid apps shared the same sensitive data with the same third parties as their free counterparts.

In this preliminary study, we did not analyze what drove certain developers to these decisions, nor did we examine consumers' expectations. More work is needed to better understand the differences and likely reasons behind why paid versions still collect data. Without properly understanding developers' motives, it is difficult to come up with solutions to increase the transparency of these apps, and to propose future policies and regulations to protect consumers. It is especially important for this information to be made available so that consumers can make informed decisions. This has to be a unified effort among policy makers, developers, and platforms, such as the Google Play Store.

In future work, we expect to further expand our sample size and do more dynamic analysis to find irrefutable evidence of violations and to find themes or patterns among those violations. Such patterns will play a key role in trying to figure out solutions. We also expect to conduct a survey to quantitatively measure the differences in consumer expectations surrounding the privacy protections afforded by paid apps.

An initial analysis of privacy policies across pairs of free and paid apps showed that most developers do not have different legal documents for versions of the apps with different monetization models. In future work we plan to extend our method to be able to automatically detect differences in privacy policies without the need for manual inspection. We also plan to take a deeper look into the issue by performing text similarity analysis on the apps privacy policies to see how different they are, not only across different monetization models but also across completely different apps.

VII. Conclusion

This paper presents a multi-dimensional analysis of the measurable benefits that consumers can expect to receive when paying for an app by employing both static and dynamic analysis, uniquely performing a large-scale, one-to-one comparison between a free version of an app and its paid counterpart.

Our preliminary results show that the privacy benefits of paying for apps are tenuous at best, and are likely to mislead consumers, making it impossible for them to make informed decisions about their privacy.

VIII. Acknowledgments

This work was supported by the U.S. National Security Agency's Science of Security program (contract H98230-18-D-0006), the Department of Homeland Security (contract FA8750-18-2-0096), the National Science Foundation (grants CNS-1817248 and CNS-1564329), the European Union's Horizon 2020 Innovation Action program (grant Agreement No. 786741, SMOOTH Project), the Rose Foundation, the Data Transparency Lab, and the Center for Long-Term Cybersecurity at U.C. Berkeley. The authors would like to thank Primal Wijesekera for feedback, as well as Refjohürs Lykkewe.

References

[1] "AAPT2 | Android Developers," https://developer.android.com/studio/command-line/aapt2.

[2] "Apktool - A tool for reverse engineering 3rd party, closed, binary Android apps." https://ibotpeaches.github.io/Apktool/.

[3] "California Consumer Privacy Act (CCPA)," https://www.oag.ca.gov/privacy/ccpa.

[4] "Distribution of free and paid Android apps in the Google Play Store from 3rd quarter 2017 to 1st quarter 2018," https://www.statista.com/statistics/266211/distribution-of-free-and-paid-android-apps/.

[5] D. Amalfitano, A. R. Fasolino, P. Tramontana, B. D. Ta, and A. M. Memon, "MobiGUITAR: Automated Model-Based Testing of Mobile Apps," *IEEE Software*, 2015.

[6] Amina Wagner, Nora Wessels, Peter Buxmann, Hanna Krasnova, "Putting a Price Tag on Personal Information - A Literature Review," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[7] S. Angeles, "Are Free Apps Safe?" https://www.businessnewsdaily.com/4868-free-app-security-risk.html, archived at https://web.archive.org/web/20181129010454/https://www.businessnewsdaily.com/4868-free-app-security-risk.html. Last Accessed: November 28, 2018.

[8] AppBrain, "Free vs. paid Android apps," https://www.appbrain.com/stats/free-and-paid-android-applications, archived at https://web.archive.org/web/20181129003827/https://www.appbrain.com/stats/free-and-paid-android-applications. Last Accessed: November 28, 2018.

[9] AppBrain, "Number of Android apps on Google Play," https://www.appbrain.com/stats/number-of-android-apps, archived at https://web.archive.org/web/20181129003859/https://www.appbrain.com/stats/number-of-android-apps. Last Accessed: November 28, 2018.

[10] B. X. Chen, "How to Protect Your Privacy as More Apps Harvest Your Data," https://www.nytimes.com/2017/05/03/technology/personaltech/how-to-protect-your-privacy-as-more-apps-harvest-your-data.html, archived at https://web.archive.org/web/20181129005245/https://www.nytimes.com/2017/05/03/technology/personaltech/how-to-protect-your-privacy-as-more-apps-harvest-your-data.html. Last Accessed: November 28, 2018.

[11] A. Dogtiev, "App Download and Usage Statistics (2018)," http://www.businessofapps.com/data/app-statistics/, archived at https://web.archive.org/web/20181130221155/http://www.businessofapps.com/data/app-statistics/. Last Accessed: November 30, 2018.

[12] C. Fox, "Google hit with £44m GDPR fine over ads," https://www.bbc.com/news/technology-46944696, archived at https://web.archive.org/save/https://www.bbc.com/news/technology-46944696. Last Accessed: January 21, 2019.

[13] Google, "Dangerous permissions," https://developer.android.com/guide/topics/permissions/requesting.html#normal-dangerous, accessed: August 17, 2017.

[14] I. Google, "Get paid to show relevant ads from over a million advertisers with Google AdMob," https://developer.android.com/distribute/best-practices/earn/show-ads-admob, archived at https://web.archive.org/web/20181129004421/https://developer.android.com/distribute/best-practices/earn/show-ads-admob. Last Accessed: November 28, 2018.

[15] Google, Inc., "UI/Application Exerciser Monkey," https://developer.android.com/tools/help/monkey.html.

[16] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, Serge Egelman, ""Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale," in *Proceedings on the 2018 Privacy Enhancing Technologies (PET2018)*, 2018, pp. 63–83.

[17] Z. Ma, H. Wang, Y. Guo, and X. Chen, "Libradar: fast and accurate detection of third-party libraries in android apps," in *Proceedings of the 38th international conference on software engineering companion*. ACM, 2016, pp. 653–656.

[18] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model," *Information Systems Research*, vol. 15, no. 4, pp. 336–355, December 2004.

[19] L. Mirani, "The amount most people are willing to pay for an app is $0 - until they've actually downloaded it," https://qz.com/129699/the-amount-most-people-are-willing-to-pay-for-an-app-is-0-until-theyve-actually-downloaded-it/, archived at https://web.archive.org/web/20181114231539/https://qz.com/129699/the-amount-most-people-are-willing-to-pay-for-an-app-is-0-until-theyve-actually-downloaded-it/. Last Accessed: November 14, 2018.

[20] M. Panzarino, "Why you should want to pay for apps," https://thenextweb.com/apps/2011/04/24/why-you-should-want-to-pay-for-apps/, archived at https://web.archive.org/web/20181129005820/https://thenextweb.com/apps/2011/04/24/why-you-should-want-to-pay-for-apps/. Last Accessed: November 28, 2018.

[21] Rajiv Garg and Rahul Telang, "Inferring App Demand from Publicly Available Data," 2013.

[22] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, S. Sundaresan, M. Allman, C. Kreibich, and P. Gill, "Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem," in *Proc. of NDSS Symposium*, 2018.

[23] S. Seneviratne, H. Kolamunna, and A. Seneviratne, "A Measurement Study of Tracking in Paid Mobile Applications," in *Proc. of ACM WiSec*, 2015.

[24] Takuya Watanabe, Mitsuaki Akiyama, Fumihiro Kanei, Eitaro Shioji, Yuta Takata, Bo Sun, Yuta Ishi, Toshiki Shibahara, Takeshi Yagi, Tatsuya Mori, "Understanding the origins of mobile app vulnerabilities: a large-scale measurement study of free and paid apps," in *Proceedings of the 14th International Conference on Mining Software Repositories*, 2017, pp. 14–24.

[25] A. K. Y. Tang, "Mobile App Monetization: App Business Models in the Digital Era," 2016.

[26] J. Tsai, S. Egelman, L. Cranor, and A. Acquisti, "The effect of online privacy information on purchasing behavior: An experimental study," in *Proceedings of the 2007 Workshop on the Economics of Information Security (WEIS'07)*, Pittsburgh, PA, USA, 2007.

[27] Yuta Ishii, Takuya Watanabe, Fumihiro Kanei, Yita Takata, Eitaro Shioji, Mitsuaki Akiyama, Takeshi Yagi, Bo Sun, Tatsuya Mori, "Understanding the security management of global third-party Android marketplaces," in *Proceedings of the 2nd ACM SIGSOFT International Workshop on App Market Analytics*, 2017, pp. 12–18.