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

Potentially misleading representations may run afoul of the FTC's prohibitions against deceptive practices and state laws

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

Domain Name	$\text{Free} \wedge \overline{\text{Paid}}$	$\text{Free} \wedge \text{Paid}$
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

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.

vlopee :-anroai-:-cm/studio/(cmm(an-T)-62 line/aapt2(,.)]TJEMC /Span <</MCID 52 >>BDC)-1.921 -11[2] -“Apktoold oold for rivrsee ylosed,d

VIII. Acknowledgments

This work was supported by the U.S. National Security Agency’s Science of Security program (contract H98230-18-D-0006), the of and Security (contract

182D-0906), the National Science ogrents

the 1643296) and 1717248e

vatioe program ogrenm

the the

a T315 gresrpagenca and the for

eaty wuldslike to thnke

W450 ijresikr(a 192 (for)92 (eedback,r)92 (as 20e(wellr)92 (as)92 Refjohürsp)-218 L’)55 ykikoet.

EEENCS

“AAPI[1]2- vlopeers