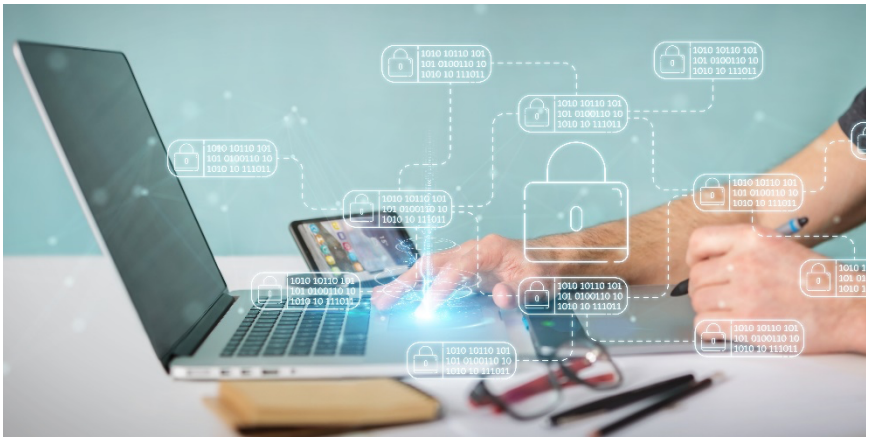
A-SIT PLUS GmbH

Privacy comparison between paid and free apps



Privacy comparison between paid and free apps

Autor:
Emina Ahmetovic
emina.ahmetovic@iaik.tugraz.at

Datum: 25.04.2023

Abstract/Zusammenfassung:
The statistics report that 97% of the Android apps from the official store are free. Nevertheless, a few free apps have their counterpart in the form of apps that are available to purchase. The possibility of choosing between free or paid versions often assumes that paid apps offer better privacy protection. In this project, we aim to evaluate the privacy features between free and paid apps. To do this, we create a dataset of top paid-free pairs. Then, we examine the privacy characteristic, such as permissions and data safety declaration. Lastly, based on our findings, we conclude to what extent the paid apps meet assumed privacy expectations.

Inhaltsverzeichnis

Contents

1. Introduction

Free Android applications are dominant in the Play Store. The distribution of paid and free apps in 2022 goes in favor of free apps by 97 percent[1]. With only 3 percent of applications available to be purchased at the store, this trend is unlikely to change much in the following years. Nevertheless, mobile markets also offer additional pricing models that complement the free and paid version. In particular, freemium and paidmium models are among the popular app pricing models that end users can choose from. While *free* apps are free of charge for users and do not rely on in-app payment, *freemium* apps refer to the apps that are free to download but still offer in-app purchases. Similarly, *paid* apps require users to make a one-time payment upon downloading apps, and they do not offer in-app purchases, while *paidmium* are paid apps that also offer in-app purchases for additional revenue. The variety of pricing models provides users with the opportunity to customize the apps according to their preferences and needs, but they also allow developers to establish a business model and gather profit. In fact, the projected revenue for 2022 for paid apps in the app market is US$5.25bn compared to US$204.90bn for in-app purchase revenue[2].

1.1. Users' understanding of paid apps

To attract users, many applications developer choose to distribute free versions. However, developers usually opt for one revenue-generation model or a combination of multiple to ensure a profit. This brings privacy differences between paid and free apps into question. The monetization model of the apps and their high distribution among apps raise the question if free apps come at the cost of privacy, where privacy is being traded for getting free apps. On the other hand, it also debates if paid apps offer better protection of privacy. Some studies on users' understanding of free and paid apps and their privacy features suggest there is a common belief that paying for an app would mean users' data will not be shared or collected (Han C, 2020 Jan;2020(3)). This belief is strongly associated with paid apps that promote ad-free or no-ads or with similar keywords that users interpret as a version that offers better privacy-preserving techniques. However, prior research that is based on static and dynamic analysis of the apps has shown this is, in practice, misleading.

1.2. Contribution

This study sheds light on the privacy differences between paid and free apps. While paying for the apps is justified by the further functionalities and features they unlock, we focus on a common belief that between the free and paid version of the app, the latter is characterized by better privacy behavior. Our objective was to examine privacy features and to answer the following research question: what are the privacy differences between the free app and its paid counterpart? Hence, we aim to conclude if paying for apps means different privacy practices. As a part of our contribution, we created a dataset with privacy-related data available on the Play Store of the 1620 free and paid app pairs. We use these data to examine and compare privacy-related features, such as data safety and permissions. Finally, conclude our study based on the evaluation results. The results of our analysis show a major overlapping between free and paid apps in terms of data safety. In the following sections, we will explain our results in more detail.

1.3. Outline

The rest of the study is organized as follows. In section 2, we will provide the background of the privacy features in Android apps. In section 3, we will introduce our methodology. In section 4, we will provide the evaluation results. In section 5, we will discuss the related work. In section 6, we will provide a conclusion.

[1] https://www.statista.com/statistics/266211/distribution-of-free-and-paid-android-apps/

[2] https://www.statista.com/outlook/dmo/app/worldwide

2. Background

In this section, we provide a short overview of the permission mechanism in Android, the data safety section, different analyses of applications, as well as an overview of the monetization models.

2.1. Permissions on Android

A great number of mobile apps that provide useful features in everyday life are offered to Android users. Nevertheless, as the apps can invade the privacy of users, the permission model (Developers., 2023) is introduced as a protection mechanism that gives users control over their data. Based on the scope of the restricted actions and access to the restricted resources, Android defines normal and dangerous protection levels.

Normal protection level is assigned to the Install-time permissions (Developers, Install-time permissions, 2023), such as INTERNET, that are characterized by limited access to restricted resources. Since they execute actions with a minimal impact on the other app or system, they are automatically granted by the system at the install time (a time when a user installs the app). The install time permissions are not displayed to the users by default but rather can be found in the Play Store details. They cannot be revoked, and there are no runtime checks. They are also granted to all users on the device.

On the other hand, runtime permissions (Developers, Runtime Permissions, 2023), such as SMS, CONTACTS, CALENDAR, LOCATION, and similar, are assigned to the dangerous protection level since they allow access to the restricted data, and allow app to perform restricted action. They are requested at runtime (right before performing the restricted action) and need explicit approval from the user. Users should be aware that runtime permissions can access private data, which potentially contains sensitive information. In contrast to the install time permissions, users can revoke runtime permissions at any time.

2.2. Analysis of Android Applications

The prior work that aims to detect privacy differences between paid and free apps have performed application analysis that was usually either static, dynamic, or a combination of these two.

Dynamic application analysis is a process of detecting and analyzing the malicious behavior of the application while the application is running. Since the monitoring and tracking of private data happen in real-time, this method is not affected by obfuscation, encryption, or similar factors. Nevertheless, the dynamic analysis is also not able to detect privacy leaks that are not triggered at the runtime. An additional limitation is high resource consumption due to the real-time operation, making these operations less affordable on some mobile devices (Kang, 2021).

In contrast to dynamic analysis, static analysis is done before running an app. Static analysis is a process of extracting application-specific features such as permissions and API calls by analyzing the source code or its binary representation. Since the static analysis works on the source code level, the detection of the malware is affected by some evasion techniques, such as obfuscation of the code or dynamic code loading. An additional limitation of the static analysis is that it can produce a larger number of false positives. A popular open-source tool for static analysis on Android is FlowDroid (Arzt, 2014).

Taint analysis is an information flow analysis technique used to track the flow of sensitive information from defined sources to defined sinks. Sources are defined as resources users want to protect, such as location or phone number, whereas sinks are defined as the points where the resources could leave the device, such as methods related to the transmission of the data. The taint analysis aims to identify the data leakage by observing the information flow between sources and sinks, and if the data from sources will reach a sink (Zhang, 2021) . It can be either static or dynamic, and it is mostly used to detect privacy leaks. A widely used tool for static taint analysis in Android apps is TaintDroid (Enck, 2014).

2.3. Data safety on Android

To increase transparency and help users understand how apps handle their data, developers utilize the data safety section to fill out the necessary information and display it to users. The purpose of data safety on Android is to disclose what data are collected and shared and to provide more information about security and privacy practices in the app. This is a way of providing users with enough information to make an informed decision about the apps they use. The following data are disclosed in the data safety section:

- **Data collection**, as the transmission of data off a device from the app, either done by libraries or SDK, collecting data from a web view that is open inside the app, or data transmitted that has been processed ephemeral, as well as user data that has been collected pseudonymously. Data that is only processed locally and not sent off the device are not included in the data collection, as well as the end-to-end encryption.
- **Data sharing**, as a sharing of collected data from the app with a third party. Data sharing includes the following cases: off-device, on-device, from the app libraries and SDK, as well as from the web view opened from the app. The following cases exclude sharing: transferring fully anonymized data, transmitting data to the service provider, transmitting data for legal purposes, or transmitting data with user consent.
- **Data handling**, where developers can disclose which collected data type is optional and which one is required. In the optional case, users have control over data collection and can decide, in contrast to the required case where collected data is required for the app functionality.
- **Privacy and security practices**, where developers have an opportunity to disclose whether their app uses encryption when transmitting data or whether it provides a deletion mechanism where users can request that their data be deleted.
- **Family policy.** Applications that target children as an audience must follow Google Play's Families policy requirements[3]. Information if the app is compliant with the family policy is displayed within the data safety section.
- **Independent security review.** If the application is validated against a global security standard, developers can add this information to complement the data safety section.

There are different type and categories of data that has been collected or shared by the app. Moreover, developers are also asked to provide the purpose[4] of data handling, such as app functionality, analytics, developer communications, advertising or marketing, fraud prevention, security, and compliance, personalization, and account management. Data types and their categories are outlined in table 1.

Category	Data type
Location	Approximate location
	Precise location
Personal info	Name
	Email address
	User IDs
	Address
	Phone number
	Race and ethnicity

[3] https://support.google.com/googleplay/android-developer/answer/9893335?hl=en

[4] https://support.google.com/googleplay/android-developer/answer/10787469?hl=en#zippy=%2Cdata-collection%2Cpurposes

	Political or religious beliefs
	Sexual orientation
	Other info
Financial info	User payment info
	Purchase history
	Credit score
	Other financial info
Health and fitness	Health info
	Fitness info
Messages	Emails
	SMS or MMS
	Other in-app messages
Photos and videos	Photos
	Videos
Audio files	Voice or sound recordings
	Music files
	Other audio files
Files and docs	Files and docs
Calendar	Calendar events
Contacts	Contacts
App activity	App interactions
	In-app search history
	Installed apps
	Other user-generated content
	Other actions
Web browsing	Web browsing history
App info and performance	Crash logs
	Diagnostics
	Other app performance data
Device or other IDs	Device or other IDs

Table 1. The table shows data types and their categories. Details of data types and their descriptions can be is outline in the office website[5].

2.4. Privacy policy on Android

To better evaluate which apps to download, users can acquire more information about the privacy of the app in their privacy policy. This field contains the link to the policy that can be downloaded by users. This optional information is located in the Google play details section, and it is written by developers who wish to share with users their privacy regulations[6].

2.5. Overview of the monetization models

A monetization model is a business model that is used for generating income from a product, service, creation, intellectual property, or similar[7] . It refers to the process in which the non-revenue item can generate income[8].

[5] https://support.google.com/googleplay/android-developer/answer/10787469?hl=en#zippy=%2Cpurposes%2Cdata-types

[6] https://support.google.com/googleplay/answer/2666094?hl=en

[7] https://www.mightynetworks.com/encyclopedia/monetization-model

[8] https://www.investopedia.com/terms/m/monetize.asp

There are many different monetization models[9] that free apps can use to gather profit. These business opportunities often include:

- **Advertising**, refers to the model where developers show third-party ads in their application to generate revenue. This is also one of the most popular models since it is versatile and can fit into most apps, but at the same time apps can remain free for end users[10].
- **Affiliate marketing** refers to a model similar to advertising; however, in this case, the income in the form of a commission fee is generated every time the users download the app or perform defined operations inside the app.
- **In-app purchases** refer to one of the most popular monetization modes where users can buy additional features from the app or goods. For example, this is often the case in game apps where users can pay for the next level and, in general, where purchases unlock additional features of the app.
- **Paid subscriptions** are a popular model among software as a service apps, where the free trial period is introduced at the beginning and followed by the paid subscription that reaches a higher level of context. For instance, many popular newspapers use the subscriptions model.
- **Freemium** is a model that combines the premium and free apps, meaning that the free version available for users is usually associated with basic functionalities. However, users must pay for additional features, context, or functionalities. For this model, it is important that the core functionality of the app remains free of charge.
- **Sponsorship** is similar to advertising, however, it is based on one advertiser that is in some way connected with the purpose of the app.
- **Crowdfunding** is a model popular among start-ups where funding platforms can support the app by donating money. Some of the popular platforms for crowdfunding are Kickstarter[11], Indiegogo[12], CrowdFunder[13], AppFunder[14].
- **Transaction revenue, and similar**. If some of the functionalities the app provides are money transactions, this can be used to get a fee from the transactions. This is popular among hotel and flight booking apps.

3. Methodology

One of the first steps was to create a dataset. To compare the privacy features of the free and paid apps, we create a dataset that consists of privacy related metadata of the apps. We took an approach where we compare free apps with their corresponding paid pairs. This means that for each free app, we try to find the paid counterpart from the same developer. As we are not purchasing apps, our analysis uses only data available in Play Store. For

[9] https://theappsolutions.com/blog/marketing/free-app-money/

[10] https://www.bornfight.com/blog/how-to-make-money-with-your-mobile-app-top-10-app-monetization-models/

[11] http://www.kickstarter.com/

[12] http://www.indiegogo.com/

[13] https://www.crowdfunder.co.uk/

14 https://appsfunder.guru/

this purpose, we have scraped the most popular appIDs from Google Play Store that have the ranking TOP FREE. Furthermore, we process our search to find the ones available only in English, and we ended with around 50 000 appIDs. As we try to identify if and which apps have paid counterparts, we search for information about app developers. After finding the developerID for each app, we find paid apps from that developer. Not all free apps have their paid versions, but also, as paid and free versions of the same developers are not really associated in Play Store, we have opted for applying algorithms in Python that measure the similarity between two strings. To find which paid appID would be a counterpart of the free appID, we compared a couple of approaches, where we have decided to use FuzzyWuzzy[15], a Python library that uses the Levenshtein Distance[16], to calculate the differences between sequences and patterns with a score of similarity that is greater or equal than 80%. The following code snippet shows a screenshot of the pairs with their corresponding score.


```
[{
        "paidApp": "lysesoft.andftppro",
        "freeApp": "lysesoft.andftp",
        "score": 91
}, {
        "paidApp": "air.LecturaMusicalPracticaPRO",
        "freeApp": "air.LecturaMusicalPractica",
        "score": 95
}, {
        "paidApp": "com.atypicalgames.radiationisland",
        "freeApp": "com.atypicalgames.radiationislandfree",
        "score": 94
}, {
        "paidApp": "com.jumobile.manager.systemapp.pro",
        "freeApp": "com.jumobile.manager.systemapp",
        "score": 94
}, {
        "paidApp": "pro.cryptotab.android",
        "freeApp": "max.cryptotab.android",
        "score": 86
}, {
        "paidApp": "com.shenyaocn.android.usbdualcamerapro",
        "freeApp": "com.shenyaocn.android.usbdualcamera",
        "score": 96
}]
```


After manually inspecting the created dataset, we noticed that, in most cases, pairs had been correctly identified. However, in some cases, such as identifying "paidApp": "pro.cryptotab.android", as a counterpart of the "freeApp": "max.cryptotab.android", we noticed that the process should be further refined. To optimize our matching techniques, we have added an additional condition, where the differences between two strings should only contain the following keywords, such as **'full', 'pro', 'lite', 'free', 'paid', 'ads', 'light',** and similar.

At the end of our process, we collected 1 640 pairs, with the data related to our privacy analysis that includes data safety, permissions, privacy policy, category, and similar. One of the examples for such a pair can be seen in the following code snippet:

[15] https://github.com/seatgeek/fuzzywuzzy/blob/master/fuzzywuzzy/fuzz.py#L236

[16] https://people.cs.pitt.edu/~kirk/cs1501/Pruhs/Spring2006/assignments/editdistance/Levenshtein%20Distance.htm

```
[{
        "paidApp": {
                "genreID": "TOOLS",
                "securityPractices": "[]",
                "score": "3.9",
                "sharedData": "[]",
                "ratings": "1043",
                "permissions": ["Phone", "Device ID & call information"],
                "appID": "lysesoft.andftppro",
                "collectedData": "[]",
                "description": " 'AndFTP Pro unlocks advanced features for AndFTP application. AndFTP is a FTP, FTPS, SCP and SFTP client. It provides commands to rename, delete, set permissions on remote files and folders. It can upload or download files and folders recursively. It supports RSA and DSA keys for SSH. You need AndFTP free installed. Features in Pro version are SCP support, folder synchronization, custom commands and import settings from file.<br><br>Pro version acts as an unlock key, it does not have any icon and you cannot open it. Once installed it unlocks all features of the free application. You can check it by running the free application, then Menu-&gt;Options-&gt;Advanced and you should see &quot;License: Pro&quot;.',\n  ",
                "title": "AndFTPPro",
                "privacy policy url": "http://www.lysesoft.com/about/privacy.html"
        },
    "freeApp": {
                "genreID": "TOOLS",
                "securityPractices": "[]",
                "score": "3.6",
                "sharedData": "[]",
                "ratings": "31196",
                "permissions": ["Photos/Media/Files", "Storage"],
                "appID": "lysesoft.andftp",
                "collectedData": "[]",
                "description": " 'AndFTP is a FTP, FTPS, SCP, SFTP client. It can manage several FTP configurations. It comes with both device and FTP file browser. It provides download, upload, synchronization and share features with resume support. It can open (local/remote), rename, delete, update permissions (chmod), run custom commands and more. SSH RSA/DSA keys support. Share from gallery is available. Intents are available for third party applications. Folder synchronization are available in Pro version only.',\n  ",
                "title": "AndFTP (your FTP client)",
                "privacy policy url": "http://www.lysesoft.com/about/privacy.html"
        }
}]
```


The collected apps belong to the 45 categories, and the distribution of the app according to the categories is depicted on the following table:

SHOPPING	5	AUTO_AND_VEHICLES	33
ENTERTAINMENT	0	NEWS_AND_MAGAZINES	4
SPORTS	20	MEDICAL	23
VIDEO_PLAYERS	49	HOUSE_AND_HOME	3
SOCIAL	7	COMICS	2
PHOTOGRAPHY	55	GAME_PUZZLE	40

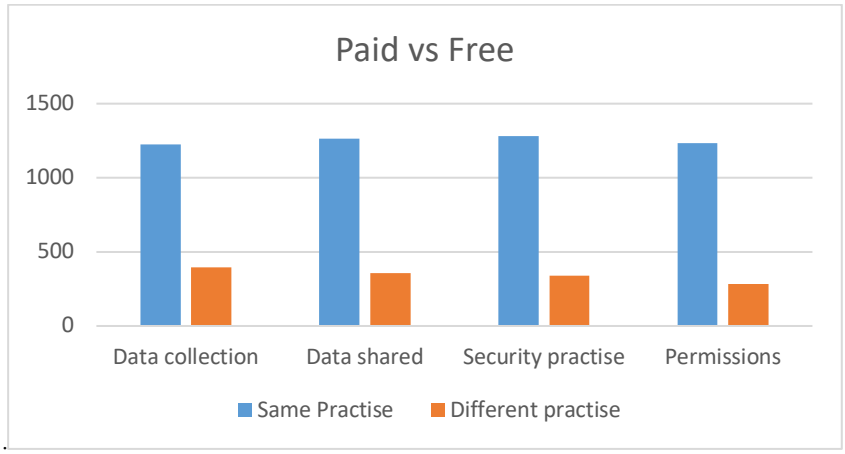
ART_AND_DESIGN	12	GAME_BOARD	11
FINANCE	28	GAME_EDUCATIONAL	38
COMMUNICATION	31	PARENTING	2
BOOKS_AND_REFERENCE	41	GAME_ARCADE	42
MAPS_AND_NAVIGATION	24	GAME_CASUAL	25
HEALTH_AND_FITNESS	54	GAME_SIMULATION	55
TOOLS	246	LIBRARIES_AND_DEMO	4
WEATHER	28	GAME_WORD	11
GAME_CASINO	5	GAME_TRIVIA	5
BUSINESS	25	GAME_RACING	12
EVENTS	1	GAME_ADVENTURE	44
MUSIC_AND_AUDIO	115	GAME_SPORTS	6
LIFESTYLE	32	GAME_ACTION	41
TRAVEL_AND_LOCAL	17	GAME_MUSIC	6
GAME_STRATEGY	52	PERSONALIZATION	46
EDUCATION	114	PRODUCTIVITY	118
GAME_ROLE_PLAYING	34	GAME_CARD	27

Table 2. This table shows the distribution of the apps according to their category.

4. Evaluation

We have used the data from our dataset to make a privacy comparison between paid and free app pairs. Rather than just analyzing privacy practices for all free and paid apps, we took an approach in which we found free-paid pairs of the apps, and we based our analysis on those pairs. This approach gives us more reliable results; however, the process of finding pairs has a drawback where the number of apps is significantly reduced due to the inability to identify pairs correctly or due to the fact that not all free apps have their counterparts.

We separate our evaluation into four parts: Data collected, Data shared, Security and privacy practices, and Permissions. The overall evaluation shows significant overlapping between paid and free app practices for the outlined categories, as seen in the diagram below. In separate sections, we elaborate on our findings in more detail.



Picture 1. The diagram depicts the differences between paid and free apps in four categories: data collection, data sharing, security practices, and permissions.

4.1. Data collected

We compare the data collection section of the free app and its paid counterpart. We analyzed 1620 pairs overall. Our study shows that **76 % (1225)** have the same data collection practice for both the free and paid version. We define the same collection practice as the one where free and paid apps collect the same data type from the same category and for the same purpose. The remaining **24% (395)** have different practices when it comes to collecting data. We investigate further what these differences are. The results we obtained are summarized in the following table 3, where the first column, Data collection type, is the type of data that has been collected. The second column is Paid app, which is the number of occurrences of that data type in paid apps, and the third column, Free app, marks the number of data collection occurrences in free apps. The results clearly point out that free apps collect, in general, more data than paid apps for some types, such as crash logs, device or other IDs, diagnostics, app interactions, approximate location, email address, other app performance data, and user ids. For the other types, the results are quite similar.

Data collection type	Paid app	Free App
Crash logs	150	250
Device or other IDs	143	312
Diagnostics	121	216
App interactions	119	220
Approximate location	63	136
Email address	57	61
Other app performance data	56	97
User IDs	42	57
Other actions	37	63
Precise location	32	34
Name	29	26
Other user-generated content	15	15
Photos	13	14
Purchase history	11	25
Other info	7	5
Other in-app messages	7	6
Installed apps	6	24
In-app search history	5	5
Voice or sound recordings	3	4
Health info	3	3
Files and docs	3	5
User payment info	2	1
Other financial info	2	2
Videos	2	3
Phone number	2	4
Fitness info	2	7
Race and ethnicity	1	1
Sexual orientation	1	1
Calendar events	1	1
Contacts	1	2
SMS or MMS	1	1
Address	1	0
Music files	0	1
Other audio files	0	1
Web browsing history	0	1

Table 3. The table depicts the type of data that has been collected by both paid and free apps and the occurrence of the collected data for each paid and free app.

4.2. Data shared

We compare the data shared section of the free app and its paid counterpart. We analyzed 1620 pairs overall. Our study shows that **78% (1263)** have the same data-sharing practice for both the free and paid version. We define the same shared practice as the one where free and paid apps are sharing the same data type from the same category and for the same purpose. The remaining **22% (357)** have different practices when it comes to sharing data. We investigate further what these differences are. The results we obtained are summarized in the following table 4, where the first column, Data shared type, is the type of data that has been shared. The second column is Paid app, which is the number of occurrences of that data type in paid apps, and the third column, Free app, marks the number of data shared occurrences in free apps.

Data shared type	Paid app	Free App
Crash logs	92	300
Device or other IDs	87	177
Diagnostics	79	174
App interactions	75	169
Approximate location	44	156
Email address	7	5
Other app performance data	23	55
User IDs	12	25
Other actions	23	39
Precise location	15	42
Name	2	0
Other user-generated content	4	2
Photos	5	3
Purchase history	6	16
Other info	0	5
Other in-app messages	1	1
Installed apps	3	8
In-app search history	5	7
Voice or sound recordings	3	3
Health info	0	0
Files and docs	3	2
User payment info	3	1
Other financial info	1	1
Videos	1	1
Phone number	2	0
Fitness info	0	2
Race and ethnicity	1	0
Address	1	0
Web browsing history	0	4

Table 4. The table depicts the type of data that has been collected by both paid and free apps and the occurrence of the collected data for each paid and free app.

Table 4 shows the differences between free and paid app data-sharing practices. Similarly to data collection, they differentiate primarily by a significantly larger number of data-sharing occurrences in free apps for some types, such as crash logs, device or other IDs, diagnostics, app interactions, and approximate location. On the other side, the differences between free and paid apps for less popular data types are negligible.

4.3. Security practices

We compare the security and privacy practices of the free app and its paid counterpart. We analyzed 1620 pairs overall. Our study shows that **79% (1279)** have the same security and privacy practice for both the free and paid version. The remaining **21% (341)** have different practices when it comes to sharing data. We investigate further what these differences are, and we obtain the following results:

Security and privacy practices	Paid apps	Free apps
Data is encrypted in transit	52	256
Data can't be deleted	43	184
You can request that data be deleted	29	102
Data isn't encrypted	20	30
Committed to follow the Play Families Policy	29	18
Independent security review	0	1

Table 5. The table depicts the type of security and privacy practices by both paid and free apps and the occurrence of them for each paid and free app.

Similar results are also obtained for security and privacy practices. However, some features that have higher occurrences in paid apps versus free apps are commitment to follow Play Family Policy.

4.4. Permissions

We compare the permissions of the free app and its paid counterpart. We analyzed 1620 pairs overall. Our study shows that **76% (1235)** have the same permissions for both the free and paid version. The remaining **24% (385)** have different practices when it comes to sharing data. We investigate further what these differences are, and the obtained results, similarly to the prior section, show that free apps declare more permissions in comparison to paid apps.

Permissions	Paid apps	Free apps
Wi-Fi connection information	30	138
Storage	68	96
Photos/Media/Files	68	96
Phone	22	57
Device ID & call information	19	55
Location	17	46
Contacts	19	33
Identity	17	27
Device & app history	6	19
Microphone	5	10
Camera	19	14
Calendar	5	7
Wearable sensors/Activity data	1	3

Table 6. The table depicts the type of permissions that has been associated with both paid and free apps, and the occurrence of those permissions for each paid and free app.

5. Related work

In a study by Han et al. (Han C, 2020 Jan;2020(3)), user expectation towards free and paid apps was examined. The authors surveyed 998 participants, and the results strongly suggest that participants expect that paid apps offer better privacy and security practices than their free counterparts. The study has indicated that most users see the ads as an indicator that the app is tracking and collecting data to a certain degree, but they also assume that the lack of ads means protection for their personal data. In their study, they applied static and dynamic analysis. They revealed that 45% of the paid versions used the same third-party libraries as free versions and that 74% had the same dangerous permissions as the free app. In addition, they reveal that 32% of the paid apps have the same data collection and transmission behaviors as their free counterparts. As our study does not purchase the paid apps and therefore excludes static or dynamic analyses, we form our study based on the metadata that is provided with the app. As the Android applications are distributed centrally, it means that also a variety of metadata, such as app data safety, privacy policy, descriptions, reviews, permissions, and similar are available to users. We focus our study mostly on the difference in data safety of the apps, in contrast to the prior work.

6. Conclusion

From the results obtained comparing privacy features available in Play Store, we can summarize the following:

- **76%** of data collection practices are the same in the free and paid apps
- **78%** of data shared practices are the same for free and its paid counterpart app
- **79%** of security and privacy practices are the same for free and its paid counterpart app
- **76%** of permissions are the same for free and its paid counterpart app

Given these numbers, we can conclude that users might wrongly think that paying for the apps would provide better privacy behavior. As with the other prior work, we argue that our study has shown very few differences in terms of outlined privacy segments between free and paid apps.

This study, however, has its limitations. First, it considers only free and paid models, where other pricing models are out of scope of this study. Second, we rely on the matching process to find pairs, in contrast to other studies where these features have been manually inspected by outsourcing this task. In addition, we were unable to perform static or dynamic analysis as we were not purchasing paid apps. This study shows the potential to further explore differences between paid and free apps, as it revealed some interesting findings that might contradict users' beliefs or expectations. One of the approaches would involve traditional analysis that could include code analysis to get more information about the differences in behavior.

References

Arzt, S. a. (2014). Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. *Acm Sigplan Notices*, 49(6):259–269.

Developers, G. (2023). Install-time permissions. pp. https://developer.android.com/guide/topics/permissions/overview#install-time.

Developers, G. (2023). Runtime Permissions. p. https://developer.android.com/guide/topics/permissions/overview#runtime.

Developers., G. (2023). Permissions on Android. p. https://developer.android.com/guide/topics/permissions/overview.

Enck, W. a.-G. (2014). Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)*, 32(2):1–29.

Han C, R. I.-R. (2020 Jan;2020(3)). The price is (not) right: Comparing privacy in free and paid apps. *Proceedings on Privacy Enhancing Technologies.*

Kang, H. a. (2021). {A modified flowdroid based on chi-square test of permissions. *Entropy*, 23(2):174.

Zhang, J. a. (2021). Analyzing android taint analysis tools: FlowDroid, Amandroid, and DroidSafe. *IEEE Transactions on Software Engineering*.