



Comparing Privacy Labels of Applications in Android and iOS

Rishabh Khandelwal
University of Wisconsin–Madison
Madison, WI, USA
rkhandelwal3@wisc.edu

Paul Chung
University of Wisconsin–Madison
Madison, WI, USA
paul.chung@wisc.edu

Asmit Nayak
University of Wisconsin–Madison
Madison, WI, USA
anayak6@wisc.edu

Kassem Fawaz
University of Wisconsin–Madison
Madison, WI, USA
kfawaz@wisc.edu

ABSTRACT

The increasing concern for privacy protection in mobile apps has prompted the development of tools such as privacy labels to assist users in understanding the privacy practices of applications. Both Google and Apple have mandated developers to use privacy labels to increase transparency in data collection and sharing practices. These privacy labels provide detailed information about apps' data practices, including the types of data collected and the purposes associated with each data type. This offers a unique opportunity to understand apps' data practices at scale. In this study, we conduct a large-scale measurement study of privacy labels using apps from the Android Play Store ($n=2.4M$) and the Apple App Store ($n=1.38M$). We establish a common mapping between iOS and Android labels, enabling a direct comparison of disclosed practices and data types between the two platforms. By studying over 100K apps, we identify discrepancies and inconsistencies in self-reported privacy practices across platforms. Our findings reveal that at least 60% of all apps have different practices on the two platforms. Additionally, we explore factors contributing to these discrepancies and provide valuable insights for developers, users, and policymakers. Our analysis suggests that while privacy labels have the potential to provide useful information concisely, in their current state, it is not clear whether the information provided is accurate. Without robust consistency checks by the distribution platforms, privacy labels may not be as effective and can even create a false sense of security for users. Our study highlights the need for further research and improved mechanisms to ensure the accuracy and consistency of privacy labels.

CCS CONCEPTS

• Security and privacy → Usability in security and privacy; • Human-centered computing → HCI design and evaluation methods.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WPES, November 26, 2023, Copenhagen, Denmark

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0235-8/23/11...\$15.00
<https://doi.org/10.1145/3603216.3624967>

KEYWORDS

privacy nutrition labels, google data safety section, apple privacy label, consistency, cross-platform analysis

ACM Reference Format:

Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. 2023. Comparing Privacy Labels of Applications in Android and iOS. In *Proceedings of the 22nd Workshop on Privacy in the Electronic Society (WPES '23)*, November 26, 2023, Copenhagen, Denmark. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3603216.3624967>

1 INTRODUCTION

Privacy practices have become crucial in the digital era as users increasingly seek transparency and control over their personal data. Traditional privacy policies, however, have shown limitations in effectively conveying these practices to users. Studies have revealed issues such as complexity and user avoidance, resulting in privacy policies often being ineffective [9, 16]. In response to these challenges, privacy labels emerged as a potential solution. Introduced by Kelley et al. [18], privacy nutrition labels aim to summarize privacy practices clearly and concisely, enhancing users' visual comprehension.

Major tech companies have embraced privacy labels to enhance user awareness. Apple introduced Apple Privacy Labels (APL) in the App Store in 2020, and Google followed suit with Data Safety Sections (DSS) in the Google Play Store in 2022 (examples shown in Figure 1). However, despite their increasing adoption, questions persist regarding the accuracy and consistency of the privacy practices reported in these labels. For example, prior studies have shown that while privacy labels benefit users by making privacy practices more accessible [33], they can be inaccurate due to developers' knowledge gaps or resource limitations [25]. Such inaccuracies can confuse and harm users by providing a false sense of security, thereby increasing their privacy risks.

Previous research [2, 32] has explored the consistency of privacy labels by examining their alignment with dataflows [32] and privacy policies [2]. Achieving consistency with dataflows through dynamic code analysis [32] can be challenging to scale, while ensuring complete consistency with privacy policies is complicated due to their coverage of practices for multiple services, including websites and apps. Moreover, privacy policies may lack the same level of granularity as privacy labels. In this study, we focus on an under-explored aspect of consistency, namely the alignment of privacy labels across different platforms. Our investigation aims to better understand developers' data practices and their alignment

with disclosed privacy labels as we uncover potential discrepancies across platforms. To achieve this, we address two key research questions:

- **RQ1:** What practices are developers reporting in privacy labels? How do these practices vary with the metadata of the apps, such as popularity, age rating, and app cost?
- **RQ2:** Do apps have different practices across platforms?

By studying these questions, we explore privacy labels’ transparency and reliability aspects. Understanding what practices developers disclose in privacy labels provides valuable insights into the level of transparency and user communication these labels offer. Additionally, exploring cross-platform inconsistencies in privacy practices can help identify potential gaps in privacy disclosures. Such gaps can lead to user confusion, false security, and enhanced privacy risks.

To address the research questions, we comprehensively analyze privacy labels for apps listed on both the Google Play Store and Apple App Store. We develop a scraper to collect metadata for over 2.4M apps from the Play Store and 1.38M apps from the App Store. Subsequently, we examine the self-reported practices of developers on both platforms, analyzing how privacy practices vary with app metadata, such as popularity, age rating, and cost.

We also compare the privacy practices of apps cross-listed on both platforms. We curate a dataset comprising apps cross-listed on both platforms to perform this analysis. We create a common mapping to compare the privacy practices from the Data Safety Section (DSS) and Apple Privacy Label (APL). Our findings reveal that developers often disclose different practices for the same app across the two platforms. To gain deeper insights, we conduct case studies to understand the nature of these discrepancies and explore potential reasons for the inconsistency. Furthermore, we delve into the possible factors that can explain the observed inconsistency, aiming to identify underlying reasons for the discrepancies in privacy practices across the two platforms.

In this work, we make the following contributions:

- We perform large-scale measurements of privacy practices reported in privacy labels across two major platforms - App Store (n=1.38M) and Google Play Store (n=2.4M). We filter out apps with less than 1000 downloads for Google Play Store. This limits the number of apps on the Google Play Store to 1.14M. We find that only 50.2% of the apps provide privacy labels on the Google Play Store, whereas on the App Store, only 69.2% of the apps contain privacy labels.
- We also identify 165K apps cross-listed on both platforms, with 100K apps having privacy labels on both, and compare these privacy labels. Surprisingly, we find that privacy labels for 51.5% of the apps are not consistent across the different platforms.
- We create a mapping between iOS and Android labels, enabling a direct comparison of the disclosed practices and data types between the two platforms.
- We provide the first large-scale datasets for privacy labels for Android (n=1.14M) and iOS (n=1.38M). Further, we curate a dataset with apps cross-listed on both platforms.

To the best of our knowledge, this is the first work to comprehensively compare Android and iOS privacy labels reported by

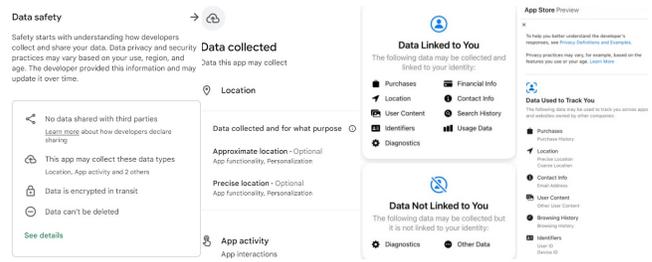


Figure 1: Illustrative Example of privacy nutrition labels.

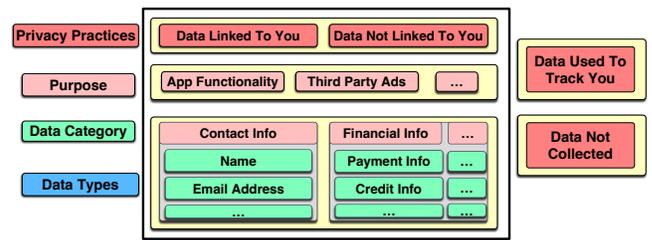


Figure 2: The Hierarchy of Apple Privacy Labels

app developers at this scale. By establishing a common mapping between the practices and data types disclosed on both platforms, we enable a rigorous and direct examination of privacy disclosures across iOS and Android apps. We hope this common mapping serves as a foundational framework for assessing the consistency and transparency of privacy labels, paving the way for data-driven analysis and evidence-based recommendations to enhance data disclosures.

2 BACKGROUND AND RELATED WORKS

Privacy Nutrition Labels. Originally introduced by Kelley et al. [18, 19], privacy nutrition labels aim to summarize the privacy practices of websites in a nutrition label format for better visual comprehension. They later designed the “Privacy Facts” display to allow the users to consider privacy while installing apps [20]. More recently, researchers proposed an Internet of Things (IoT) security and privacy label [12, 13] to surface privacy and security information related to IoT devices to the users. Researchers have also studied the design and evaluation of privacy notices and labels [7, 10, 11, 14, 18–21, 28, 29].

In December 2020, Apple implemented privacy nutrition labels for the app store and mandated app developers to provide privacy information for their apps through the Apple Privacy Label (APL). Recently, Google also required app developers to include a Data Safety Section (DSS) on the Google Play Store. An example of the Data Safety Section and Apple Privacy Label can be seen in Figure 1.

Apple Privacy Label. The Apple Privacy Label (APL) is a four-level hierarchy (as shown in Figure 2). The top level consists of four high-level privacy practices, known as *Privacy Types*. The second level of the label discusses the purpose for data usage, while the

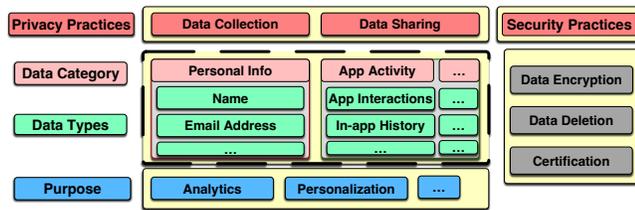


Figure 3: Google Data Safety Section

third and fourth level describes high-level *Data Categories* and fine-grained *Datatypes*, respectively. At the top level, *No Data Collected* denotes that the app does not collect any user data.

Among the other three categories, *Data used to Track you* covers the practices when user data is linked with third-party data for targeted advertising, Ad measurement, or sharing with a data broker. Notably, tracking does not apply when the data is never sent off the device in a way to identify the user or device, or if the data is used for fraud detection. *Data linked to you* covers the personal information and data linked to the user’s identity as opposed to *Data not linked to you*.

The next level describes the purposes for which data collected in *Data linked to you* and *Data not linked to you* may be used. Apple defines five main purposes: *Third party advertising and marketing*, *Developers’ advertising and marketing*, *Analytics*, *Product Personalization*, *App Functionality* and *Other Purposes*. It is important to note that *Data Used to Track you* does not get a purpose level as its purpose is to track the users. In the *Data Categories* level, Apple defines 14 categories of data such as *Contact Info* (consisting of personal information), *Health and Fitness*, *Financial Info* etc. *Data Categories* consists of the final level - *DataTypes* which consists of 32 fine-grained datatypes that the developers can use, such as *App Interactions*, *Precise Location*, *Contacts*, *Phone* etc. An illustrative example of APL is shown in Fig. 1.

Google Data Safety Section The Data Safety Section (DSS) also consists of four levels, where the first is high level *Privacy Practices*. The second and third levels consist of *Data Categories* and *Data Types*, and the fourth level consists of *Purpose*.

The first level includes three practices: *Data Collection*, which covers the details about the data that is collected and its intended use; *Data Sharing*, where the developers disclose what data is shared with third parties; and *Security Practices* that covers the data practices related to user choice and data security. *Security Practices* include three tags: *Encrypted in Transit*, *Data Deletion Option*, and *Review against Global Security Standards*.

In the second level, *Data Categories* includes 14 categories such as App Info and Performance and App Activity. Each *Data Category* can also have *Data Types*, which provide fine-grained information about the data used by the app. For example, *App Activity* includes *App Interactions and Installed App*, as shown in Fig. 3. The final level of the Data Safety Section consists of *Purposes* that describe the reasons for collecting or sharing the data.

We note that even though the two privacy labels (APL and DSS) have some overlap at the lowest level, they cover different high-level practices. For instance, APL focuses on surfacing tracking practices and the linkability of the data. DSS focuses on data-centric practices,

including collection, sharing, encryption, and deletion. In the rest of the paper, we will use APL and DSS to denote privacy labels for iOS and Android apps, respectively. Further, we use the term *Privacy Labels* to refer to both APL and DSS collectively.

Usability of Privacy Labels. Researchers have studied the usability of APLs from both users’ [33] and developers’ [25] perspectives. Zhang et al. [33] studied 24 iPhone users to understand their experiences and perceptions of privacy labels on the app store. They uncovered that users find the labels confusing with unfamiliar terms. From the developers’ perspective, Li et al. [25] interviewed 12 iOS developers and reported that the sources of errors by developers in privacy labels included both under-reporting and over-reporting data collection. They further concluded that the label design is generally confusing for the developers either due to known factors (lack of resources, improper documentation) or unknown factors (preconceptions, knowledge gaps). Researchers recently built and evaluated a tool [15] that helps iOS developers generate privacy labels by identifying data flows through code analysis. While these works focus on the usability evaluation of APL, our work compares the privacy practices in APL with those present in DSS.

Studies on Privacy Labels. Similar to our work, Xiao et al. [32] characterize non-compliance of Apple privacy labels by studying data flow to label consistency of 5K iOS apps. They also provide insights for improving label design. This work is complementary to ours as we measure the consistency of privacy labels with the data practices mentioned across the platforms.

The works most similar to ours perform longitudinal measurement of privacy labels to understand the adoption and evolution of Apple privacy labels over time [5, 25, 30]. In particular, Scoccia et al. [30] conducted an empirical study of 17K apps to characterize how sensitive data is collected and shared for iOS apps. They found that free apps collect more sensitive data for tracking purposes. Li et al. [25] and Balash et al. [5] collected weekly snapshots of Apple privacy labels and characterized the privacy practices mentioned in privacy labels for 573k apps. Balash et al. [5] also perform additional correlation analysis with app meta-data like user rating, content rating, and app size.

Our work is different in two ways. First, we provide complimentary analysis by analyzing privacy labels from Apple and Google to provide a comprehensive understanding of practices mentioned in APL and DSS. In doing so, we verify prior works’ findings on how sensitive data is being collected and used in the Apple Privacy Label. Second, we create a dataset with cross-listed apps on both platforms to understand how developers disclose their practices on different platforms. To the best of our knowledge, ours is the first work performing comparative analysis across the two platforms.

3 DATA COLLECTION PIPELINE

We show an overview of the data collection pipeline in Fig. 4. We begin by scraping the metadata and privacy labels for the apps from Google Play and Apple App Store (Section 3.1). Finally, we identify cross-listed apps between Google Play and Apple App Store (Section 3.2).

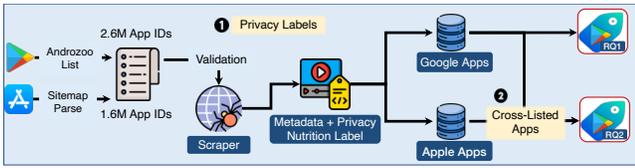


Figure 4: Overview of the data collection pipeline. RQs here refer to the Research Questions introduced in Section 1

3.1 Privacy Labels

First, we describe the collection method for our privacy labels (both DSS and APL) datasets.

Google Data Safety Section. We collected a snapshot of the Data Safety Sections (DSS) for 2.49M apps present on the Play Store on November 25, 2022. Google required app developers to complete the data safety section by July 20, 2022. To collect the data safety section, we start with the APK list provided by Androzo [3]. This daily updated list consists of up-to-date Android app ids from various sources, including those from the Google Play store. Using the app ids and a customized version of publicly available Google Play Store scraper library `google-play-scraper` [1], we capture the metadata of each app, including its data safety sections. We used four local machines to perform the scraping. The total time to retrieve data for 2.49M apps from Google Play is 24 to 48 hours. We note that this set also includes apps with very low download counts. To ensure that apps with low download counts do not skew our analysis, we filter out apps with fewer than 1000 downloads, resulting in a total of 1.14M apps with 573k having privacy labels. We refer to this dataset as **DSS Dataset**.

Apple Privacy Labels. We collected a snapshot of the Apple Privacy Labels (APLs) on November 13, 2022. We begin by parsing the XML site map for the app store to curate the dataset.¹ Using the URLs from the sitemap, we use the Apple Store Catalogue API to extract the metadata for each app, including the privacy nutrition label. We crawled using 11 instances of Google Cloud functions to scrape 1.6M apps in 15 hours.

We extracted information for 1.38M apps from the 1.6M apps available, filtering out those with non-English content. As a result, we obtained 955K (69.2%) with APLs. In comparison, Balash et al. [5] in March 2022 found that 60.5% of apps had Apple Privacy Labels. The higher percentage in our study suggests that new APLs are still being added to apps. We refer to this dataset as **APL Dataset**.

3.2 Identifying Cross-Listed Apps

We outline the process of identifying cross-listed apps across both platforms in Fig. 5, which forms the basis for comparing privacy labels in our analysis (Section 5). The absence of unique identifiers in cross-platform apps makes this task particularly challenging [17].

We address this challenge by employing a heuristic based on combinations of pseudo-identifiers, including the app name, developer name, privacy policy, and developer website. We start by considering apps with the same name on both platforms (n=220K). We note that this set can include false positives as different apps

¹We used the `ultimate-sitemap-parser` library.

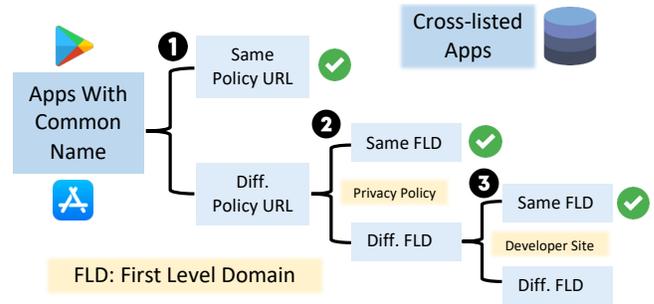


Figure 5: Pipeline for the generation of cross-listed app dataset. The rationale for using both full privacy policy URL and its FLD is to capture the cases where the apps may have different policies for iOS and Android.

can have the same names. To eliminate false positives, we rely on the following selection criteria: If the privacy policy URL matches the apps, we consider them as a unique match (n=85K). In certain cases, like the *N TLC Catalog* app on the Play Store and App Store, platform-specific identifiers may be included in the URLs for privacy policies, resulting in a different URL. If this is the case, we match the first-level domain of the privacy policy URLs and identify them as unique matches (n=54K). Additionally, some apps may not contain a link to the privacy policy despite it being encouraged on both platforms. If we cannot find the privacy policy URL in the app stores metadata, we broaden our coverage by also matching the first-level domain of the developer website, which is available on both platforms. This approach yields a total of 25K matches, resulting in 165K apps in both the Apple Play Store and Google Play Store.

Manual Verification: To assess whether our heuristic results in false positive matches, two of the authors manually verified a total of 450 app pairs, 150 app pairs randomly sampled for each of the three heuristics and found that no app from Google Play Store was matched to an incorrect app from App Store. For our analysis, having an accurately mapped set is more important than capturing all instances of cross-listed apps, as false positives can skew the consistency results.

Rationale for Starting with Apps with the Same Name: In our methodology for identifying cross-listed apps, we consider apps with identical names. While this approach may overlook apps with slightly different names across platforms, we believe it is justified for our analysis’s specific focus on checking the consistency of reported practices across platforms. Ensuring precise mapping is crucial to avoid false positives that could potentially distort the results.

Cross-listed Apps Dataset Using the method described above, we find a total of 165K cross-listed apps. Among these apps, we find that 5% have privacy nutrition labels only on the Google Play Store, 20.2% have the label only on the Apple App Store, 60.8% have labels on both the platforms and 13.9% do not have a privacy nutrition label on either platform. The higher rate of privacy labels for the App Store can be understood as Apple enforced nutrition labels on

their platform earlier than Google, giving more time for developers to add the details in the APL.

3.3 Ethical Considerations

We collected data only from publicly available web pages and APIs. While our data collection scripts might load Google and Apple’s servers, we were careful not to abuse these resources. In particular, we added back-off strategies in case of errors and waited for sufficient time before retrying for the failed cases. Furthermore, for privacy policy extraction, we were respectful of robots.txt and only extracted HTML when the website allowed us to.

3.4 Takeaways

Our measurement pipeline yielded two significant observations. Firstly, app developers have shown a sluggish response in adding privacy labels to their apps, even after the hard deadlines have passed. As of November 2022, privacy labels are only present for 69% of the apps on the Apple app store and 50.2% of the apps on the Google Play store. Secondly, our data collection pipeline resulted in the curation of three extensive datasets: one for Apple Privacy Labels (n=955K), another for Google Data Safety Section (n=573K), and a third dataset containing apps cross-listed across both platforms (n=165K).

4 PRACTICES IN PRIVACY LABELS [RQ1]

4.1 Google Data Safety Section

In this section, we analyze the DSS dataset (Section 3.1) comprising 573K apps. We first discuss the practices present in DSS and then examine how these practices vary with age rating, price, and popularity.

Data Collection and Sharing: Among the apps having DSS, we saw 42.3% collecting at least one type of data, and 35.8% sharing at least one data type (purple bars in the top plot for Fig. 6). This suggests that most apps on the Play Store report do not collect or share data. This is in contrast with the findings from prior work [31] that found that the majority of the apps use at least one third-party application, which has been shown to collect sensitive information [8, 26]. One possible explanation is that developers struggle to understand third-party libraries’ collection and sharing practices. This is also supported by prior research [6, 25].

We also note that among the apps not collecting any data, around 23% report sharing data. This is because *Data Collection* is defined as the instance when the developers retrieve the data from the device using the app [4], whereas *Data Sharing* is defined as when the data is transferred from the device to a third party. Thus, per definitions, the developers can share data without collecting it if the application uses third-party libraries, which send data directly to third-party servers.

Security Practices: We find that 23% of the apps do not provide any details of their security practices. 65% of the apps encrypt data they collect or share while it’s in transit, and 42% allow the users to request that their data be deleted or automatically anonymize data within 90 days. Notably, we find that 40% of the apps state that they do not collect or share data but encrypt the data in transit. As apps need network permissions to transmit data, we cross-verified

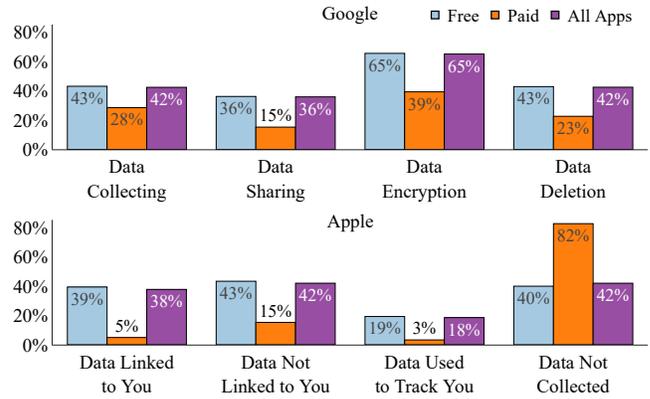


Figure 6: Distribution of privacy types in Google Data Safety Sections and Apple Privacy Labels. The normalization is done by the total number of apps with privacy labels.

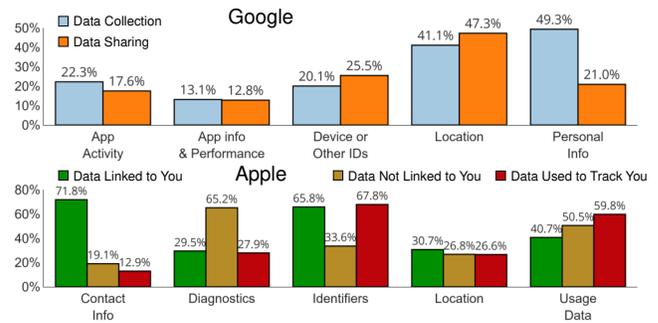


Figure 7: Distribution of Top-5 data categories for high-level practices for apps in Play Store (top) and App Store (bottom). The normalization is done by the total number of apps with privacy labels. For plots with data categories, see Fig. 14 in the Appendix.

encryption practices with apps’ network permission requests and find that 10.5% apps do request network permission but do not encrypt data, potentially exposing user data in plain text. Additionally, 12.6K of apps (with 1.8K apps with more than 100K downloads) do not request network permissions, yet state that they encrypt data in transit, suggesting that some developers might be over-reporting their practices, consistent with prior research [22, 25].

Category and Purpose Level Practices: In Fig. 7, we present the top-5 data categories for *Data Collection* and *Data Sharing* by apps in the Play Store. The full plot includes all data categories in Fig. 14 (Appendix). Our findings indicate that the data categories *Personal Information* and *App activity* are among the most frequently collected and are primarily used for *App functionality* and *Analytics*. However, *Location* and *Device Ids* are more commonly shared for the purpose of *Advertising or Marketing*. We emphasize that this flow poses serious privacy risks and allows for third-party tracking. We also observe that sensitive data types such as *Audio*, *Files and Docs*, and *Health and Fitness* are collected less frequently, with the most common purpose being *App functionality*.

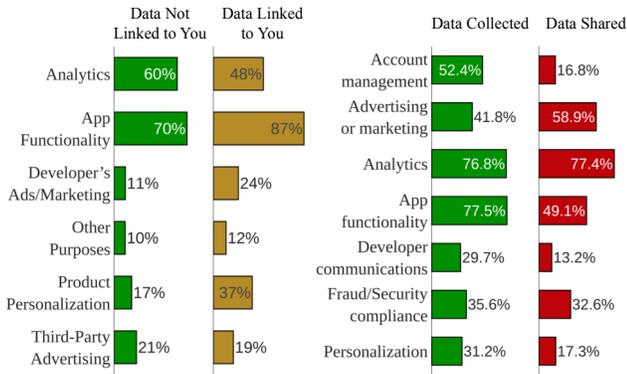


Figure 8: Distribution of purpose for high-level privacy practices in APL

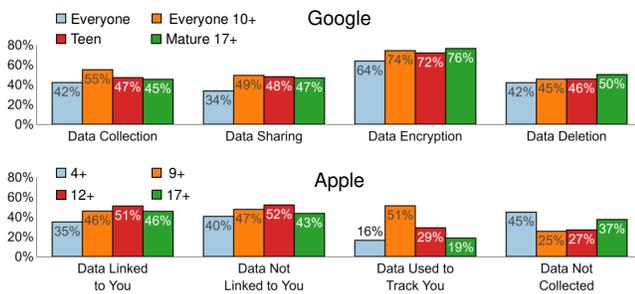


Figure 9: Distribution of privacy types based on age rating for DSS and APLs. The normalization is done by the total number of apps with privacy labels.

Figure 8 shows the distribution of purpose for data collection for apps on the Play Store. We find that *Analytics* and *App functionality* are among the most popular purposes for which the apps request data. Furthermore, out of the 7 possible purposes for collecting data, over 4K apps list 6 or more purposes for the data they collect, which may indicate that app developers list all purposes out of convenience. For example, *Workplace from Meta* with over 15M+ downloads, lists the same 6 purposes for all the data they collect like access to *Installed Apps, SMS or MMS, Music Files*. This is consistent with the findings of Li et al. [25], who suggest that developers may over-report in cases of ambiguity.

Variation of Practices with Popularity: We first investigate the relationship between privacy practices and app popularity. We classify apps into three categories based on their number of downloads: extremely popular (greater than 1M download, n=56K), semi-popular (more than 10K downloads, n=524K), and low-popular (less than 10K downloads, n=621K). Our findings reveal that 1) the fraction of apps displaying Data Safety Sections (DSS) increases with the popularity of the apps (42% for low-popular, 51% for semi-popular, and 76% for extremely popular) and 2) the fraction of apps collecting and sharing data is less for popular apps (41% for low-popular, 46% for semi-popular and 12% for extremely popular). These results suggest that developers from popular apps tend to report more privacy-friendly practices.

Variation of Practices with Age Rating: Next, we examine how the practices of apps differ based on their age rating as determined by the Google Play Store. The Play Store assigns five different age ratings: Everyone, Teen, Mature 17+, and Everyone 10+². We acknowledge this distinction’s importance, as apps accessible to children and teens (falling in the Everyone and Teen categories) are expected to have higher transparency and collect less data. However, our analysis of the dataset reveals that 59% of apps with the *Mature 17+* rating have a Data Safety Section (DSS). In contrast, the fraction of apps with a DSS in the other age ratings ranges from 47% (Everyone) to 55% (Everyone 10+). The data practices for different age ratings are shown in Fig. 9. We find that the fraction of apps having *Data Collection* and *Data Sharing* is lowest for apps rated for *Everyone*, whereas apps targeting *Mature 17+* have the highest encryption rate.

Variation of Practices with Price: Finally, we study the difference in practices based on whether the app is available for free, free with in-app purchases, or paid. We find that 68% of the paid apps have DSS, whereas, for free apps, only 46% have DSS. Fig. 6 shows the distribution of high-level practices with free and paid apps. We note that for paid apps, a fraction of apps collecting and sharing data is lower. Furthermore, apps with *Data Encryption* and *Data Deletion* are lower because the apps are collecting and sharing fewer data. This suggests that paid apps tend to have better data practices.

Internal Consistency: Analyzing the app permissions with the data collected/shared, we observe the developers report practices inconsistent with other declared practices or with the app permissions. For example, we find that 40% of the apps state that they do not collect or share data, but encrypt the data in transit. We delved deeper into this observation by cross-verifying security practices with apps’ network permission requests. 59% of apps do not request network permissions yet state that they encrypt data in transit. It is unclear why apps would need to encrypt transit data if they are not collecting or sharing data.

We also cross-verified the collected and shared data types from the DSS to the app permissions. Several apps report collecting or sharing data types without asking for the corresponding permissions. For example, 11.5% of the apps report collecting or sharing precise location data without obtaining location permissions. Another example is 23.7% of the apps report collecting or sharing files and documents without the “Photos/Media/Files” permissions.

The inconsistencies in app developers’ self-reported privacy practices and permissions significantly affect user trust and data security. When apps claim not to collect or share data but mention encrypting data in transit, it raises concerns about the accuracy and transparency of their practices. Similarly, when apps collect or share specific data types without requesting the corresponding permissions, it creates doubts about how the data is being accessed and utilized. These inconsistencies undermine the reliability of privacy labels and may lead users to make ill-informed decisions, potentially exposing them to privacy risks and data breaches. Addressing these discrepancies is crucial for building user confidence and ensuring that privacy practices align with what is communicated to users through app permissions and privacy labels.

²Google also has Adults 18+ rating, but we found less than 200 apps in this category and decided to filter it out for this analysis

4.2 Apple Privacy Labels

Next, we examine the Apple Privacy Label (APL) dataset (Section 3.1) consisting of privacy labels from 955K apps. We first discuss the practices present in APL and then dive into variations of practices with an age rating and price. Finally, we conclude by comparing the low-level practices mentioned in APL and DSS

High-Level Practices: In our dataset, 42% of apps collected data from users that were not linked back to the user (Data Not Linked to You), whereas 37% of apps did collect data that is linked to the user (Fig. 6). Note that apps could collect multiple types of data, some of which may be linked to the users while others may not. Furthermore, around 18% of the apps reported collecting data used to track the users. Note that this reflects the status of the APLs after the *Apple Tracking Transparency* policy was implemented, which requires developers to obtain consent from users before tracking but does not entirely prevent an app from monitoring a users' activity. We also find that 42% of apps report not collecting data from users. Recent works [23, 24] analyzing iOS apps have found that at least 80% of the apps still use tracking libraries in the apps. Further, these libraries have been shown to collect user data [8, 26]. Similar to the case of Android developers, this discrepancy can be explained by the third-party libraries' lack of transparency of privacy practices, which is confusing for the developers.

For *Data Used to Track You*, we find that *Usage Data* and *Identifiers* are most commonly used. We note here that Apple defines *Tracking* as when data collected is linked with third-party data for targeted advertising and when the data is shared with a data broker. Additionally, we observe that 25% of the apps collecting *Location* information also use it for tracking. This poses severe privacy risks to the users as entities can track the physical location of the users, which can reveal sensitive details about users' habits and routines.

Data Category and Purpose Level Practices: In Fig. 7 (bottom), we show the top-6 data categories mentioned in the high-level practices in the APL dataset. We find that for *Data Linked to You*, *Contact Information* and *Identifiers* are collected most frequently, whereas for *Data Not Linked to You*, *Diagnostics* and *Usage Data* are collected most frequently. Apple defines *Contact Information* as name, email, phone number, and physical address, whereas *Usage Data* refers to product interactions and advertising data such as information about the ads that the user has viewed. Analyzing purposes for these data categories (Figure 8), we find that nearly 60% of the apps use these data categories for *App functionality* and *Analytics*. It is also worth noting that *Contact Information* is used for *Advertisements* in only 8% of the apps that collect this information, indicating that apps generally do not use personal information for advertisements. We also note that *Identifiers*, commonly used for tracking users for targeted advertising is used for *Advertisement or Marketing* in more than 20% of the apps that collect *Identifiers*. Interestingly, *Location*, under *Data Linked to You* is also used for *Advertisement or Marketing* by 20% of the apps that collect *Location*.

Variation of Practices with Age Rating Next, we investigate the correlation between the privacy practices described in the Android Permission List (APL) and the age rating and price of apps. The App Store assigns four different age ratings: 4+, 9+, 12+, and 17+ (which roughly align with the rating system used by the Google Play Store). Our analysis reveals that the fraction of apps with an age rating of

17+ is highest at 76%. However, we note that the high-level data practices, shown in Figure 1, are consistently more privacy-friendly for apps with lower age ratings. For instance, only 13% of the apps with an age rating of 4+ track users. Similarly, data collection for these apps is also consistently lower than that of other categories.

Variation of Practices with Price: Finally, we categorize the dataset into free and paid apps and examine the differences in privacy labels. Recall that for the Play Store, we observed that paid apps contained more DSS than free apps. We find the reverse trend for APL, with 70% of the free apps having APL as compared to 52% paid apps. On the other hand, the high-level practices are decidedly better for the paid apps, as shown in Fig. 6 (bottom chart). For instance, 82% of the paid apps reported not collecting any data, while only 3% of paid apps mentioned using data to track the user. This indicates that the paid apps on iOS platforms are more friendly than the free apps.

Comparison Between DSS and APL: As discussed in Section 2, DSS and APL provide different information to the users and cannot be directly compared based on high-level practices. However, since the underlying data collected is the same, we can compare the practices shown in Fig. 7. We observe that the fraction of apps requesting similar datatypes is much smaller for apps on the Play Store than that of the App Store (with a notable exception of *Location*). This can be attributed to the fact that developers have had a longer to work with the APL framework, while the DSS framework is still relatively new. In our communication with app developers, one developer mentioned that they tried different answers on the data safety form. We also received communication indicating that some developers updated their DSS based on the questions that we had asked. This indicates that the developers are unclear on the process involved in the data safety forms, which might result in some inaccuracies in the DSSs. This is also supported by the study conducted by Li et al. [26], where they find that app developers find it challenging to fill out the privacy labels, especially because the frameworks for Apple and Google are starkly different and can create confusion.

4.3 Takeaways

The analysis presented here results in three main takeaways: 1) Privacy practices reported in the privacy nutrition labels differ from the privacy practices derived using app analysis by prior works [31]. Specifically, previous works have shown that third-party libraries are used in the majority of apps and that these libraries collect sensitive information from the users. This is inconsistent with what we find in the privacy labels. This inconsistency can be explained by the fact that privacy practices of third-party libraries are often vague and create confusion among the developers (consistent with findings from literature [25]). 2) We show that paid apps and apps open to all age groups, including children, are more privacy-friendly. As shown in Fig. 6 and Fig. 9, these apps are less likely to engage in tracking, data collection, and data sharing. 3) Fig. 7 also shows that location data is often used for advertising, marketing, and tracking. This poses severe privacy risks, as location data can reveal sensitive information about an individual's habits and routines. This suggests that further attention should be paid to the use of location data in mobile apps and the potential risks it poses to user privacy.

| DSS Purposes | → | APL Purposes |
|--|---|--------------------------|
| Advertising or Marketing | → | Advertising or Marketing |
| Analytics | → | Analytics |
| App Functionality | → | App Functionality |
| Fraud prevention, Security, and Compliance | → | App Functionality |
| Personalization | → | Personalization |
| Account Management | → | N/A |
| Developer Communication | → | N/A |

Table 1: Table showing the common mapping from Data Safety Card to Apple Privacy Label

5 PRACTICES ACROSS PLATFORMS [RQ2]

The privacy labels on Android and iOS platforms cover distinct aspects of data practices. For instance, DSS focuses on security practices, whereas APL lacks such coverage. Despite their differences in high-level practices, a significant overlap exists in the lower-level attributes, specifically the datatype and purpose. As a result, we leverage these lower-level practices to compare the disclosed practices by app developers.

To facilitate this comparison, we first identify the common datatypes and purposes present in both labels. Subsequently, we thoroughly examine two key factors: 1) the datatypes and 2) the combinations of datatypes and purposes. Our primary aim is to assess the consistency of privacy labels across both platforms. We hypothesize that, for a given app, data practices should demonstrate similarities between the two platforms. To validate this hypothesis, we analyze privacy labels for a substantial sample of 100K apps that disclose their practices on Android and iOS platforms.

5.1 Mapping DSS categories to APL categories

To compare privacy practices for apps in APL and DSS, we establish a common mapping between the datatypes and purposes used in the two labels. As previously mentioned, the two labels reveal different high-level practices. APL emphasizes tracking and linkability of collected data without distinguishing between data collected and shared. Conversely, DSS focuses on security practices and whether data is collected or shared with third parties. It is important to note that the datatype and purpose tags employed in both labels may denote varying concepts. For instance, in APL, App functionality encompasses fraud prevention and security measures, while in DSS, they are represented as distinct tags, with separate ones for app functionality, fraud prevention, and security measures.

Mapping Purposes. APL enumerates four distinct purposes for data collection, while DSS lists seven purposes, as illustrated in Table 1. We observe that *Fraud Prevention, Security, and Compliance* constitutes a separate tag in DSS, while in APL, it is included within *App Functionality*. Additionally, since there are no equivalent tags for *Account Management* and *Developer Communication* in APL,

| Google DataType | → | Apple DataType |
|-------------------------------|---|-------------------------|
| Approximate Location | → | Coarse Location |
| Address | → | Physical Address |
| Political Or Religious Belief | → | Sensitive Info |
| Sexual Orientation | → | Sensitive Info |
| Emails | → | Emails Or Text Messages |
| Sms Or Mms | → | Emails Or Text Messages |
| Files And Docs | → | N/A |
| Calendar | → | N/A |
| Crash Logs | → | Crash Data |
| Diagnostics | → | Performance Data |
| Device Or Other Ids | → | Device Id |

Table 2: Table showing the mapping of datatypes from Data Safety Card to Apple Privacy Label

we exclude them from DSS during this comparison. The complete mapping for purposes is available in Table 1.

Mapping DataTypes. In our examination of the definitions for the datatypes collected/shared in APL and DSS, we observe that DSS provides finer granularity. For instance, DSS features separate tags for *Race and Ethnicity*, *Sexual Orientation*, and *Political or Religious beliefs*, whereas APL groups these categories under *Sensitive Information*. Moreover, DSS introduces separate categories for *Calendar*, *Files and Docs*, and *Music Files*, which APL does not include. Notably, we also find identical datatypes represented with different tag names. For example, *App Interactions* in DSS corresponds to *Product Interaction* in APL, and *Diagnostics* data in DSS aligns with *Performance* data. The comprehensive mapping is provided in Table 3 in Appendix A.

5.2 Findings

We conduct a comparative analysis of the self-reported privacy practices among 100K apps listed on both Android and iOS platforms, each of which provides privacy labels. Our investigation revolves around two primary questions: a) How do the high-level practices of data collection compare between the two labels? and b) Is there consistency in the stated purposes for using various datatypes between the two platforms?

5.2.1 Comparison of Data Collection. To determine the disparity in data collection practices, we rely on the *Data Not Collected* tag for the iOS platform and the *Data Shared* and *Data Collected* tags for the Android platform. Our findings indicate that a total of 22K (22%) apps report different data collection practices across the two platforms. Among these apps, 42% declare data collection on Android, while 58% do so on the iOS platform. Upon further examination, we discovered that 18% of these apps have amassed more than 100k

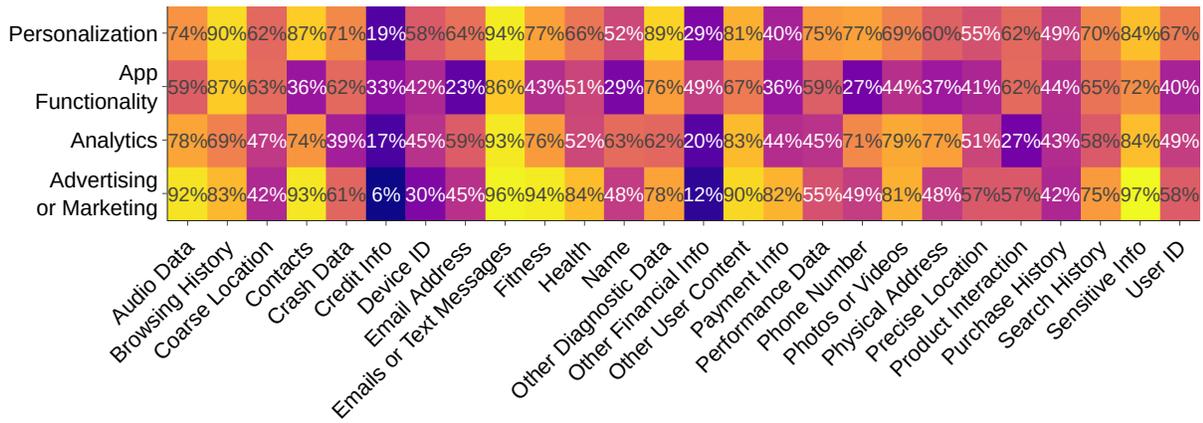


Figure 10: Normalized Heatmap showing the inconsistencies in the datatype-purpose pair. Normalization is done for each cell block in the heatmap, i.e., for each datatype-purpose pair, we normalize with the total number of apps that have that datatype-purpose pair.

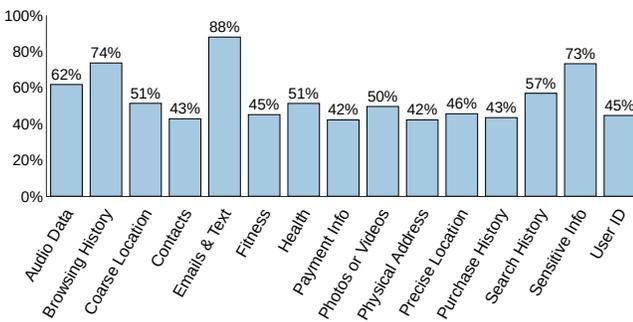


Figure 11: Distribution of inconsistent apps with datatypes. Each datatype is normalized with the number of apps using that particular datatype on either platform. Note that we have omitted some datatypes here for brevity. The full distribution can be found in the Fig. 14 in the Appendix

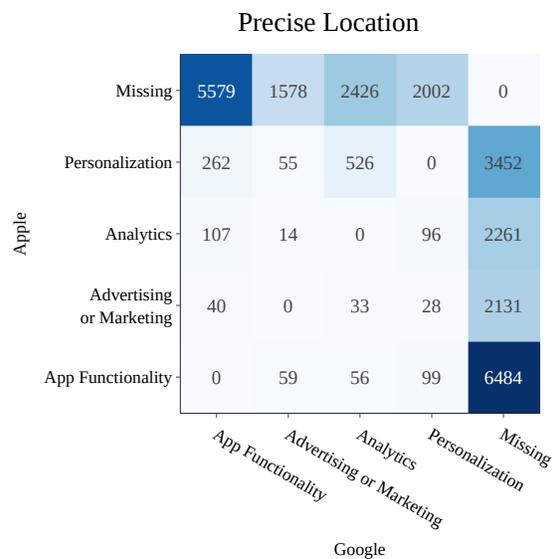


Figure 12: Confusion metric for Precise Location. Each row shows what the purpose in Apple for precise location was mislabeled in Google, whereas each column shows what the purpose in Google was mislabeled in Apple. Takeaway: When the developers mentioned app functionality in Google, a lot of times, it was confused with advertising or analytics or personalization, indicating that the frameworks are not working as they are supposed to, i.e. the users might not be getting accurate information about privacy practices of the apps.

downloads, with 5% boasting over 1M downloads, signifying that even popular apps exhibit this inconsistency.

For instance, the app *KineMaster - Video Editor*, a video editing application with over 400M+ downloads on Google Play Store, claims not to collect any data in the Play Store, but on the App Store, it asserts the collection of sensitive data such as *Location* and *Identifiers*.

The presence of such discrepancies in self-reported data collection practices undermines the credibility of the Privacy Label framework. This poses a significant concern for users, as they may base their decisions on inaccurate information, thereby exposing themselves to increased privacy risks.

5.2.2 Comparison of Fine-grained Practice. We perform the fine-grained analysis along two dimensions: 1) DataType: We examine whether the privacy labels report the collection or sharing of the

same datatypes. 2) DataType-Purpose pairs: We compare the common datatype-purpose pairs in both labels. Any app that lacks at least one datatype-purpose pair present in both sets is considered

inconsistent. Additionally, instances, where a datatype-purpose pair is present in one label but missing in the other, are also tagged as inconsistent. For instance, if an app’s DSS states (*Location - Personalization*), while APL states (*Location - App Functionality*), we consider the app inconsistent. Similarly, if an app’s DSS has (*Location - Personalization*), APL has (*Location - App Functionality*), and also (*Location - Personalization*), we still treat it as inconsistent, as the tag (*Location - App Functionality*) is not common in both labels.

Among the cross-listed apps with privacy labels, we find that at least 60% exhibit at least one inconsistency. For example, in the app *Tiktok*, DSS indicates data collection of users’ contact lists for ‘Advertising and Marketing’ purposes, while APL states that the app does not collect a contact list.

Fig. 11 illustrates the inconsistency in datatypes across the two platforms. Notably, *Sensitive Information*, *Browsing History*, and *Emails or Text Messages* exhibit the highest inconsistencies between the platforms. From Fig. 11, we observe that DeviceID and Product Interactions are the data categories with the most significant inconsistencies. Additionally, *Precise Location* and *Coarse Location* demonstrate inconsistency with *Advertising*, implying that at least in one of the labels, location data is used for advertising, thereby raising privacy concerns for users.

To explore datatype-purpose inconsistencies, we depict the normalized heatmap of inconsistent apps in Fig. 10. Each cell block in the heatmap is normalized by the total number of apps that have that specific datatype-purpose pair. Our findings reveal that *Fitness* and *Sensitive Information*, when used for *Advertising or Marketing*, frequently demonstrate inconsistencies. The plot also indicates that while *Sensitive Information* and *Fitness* data are not commonly collected (Fig. 11), when they are collected, they often exhibit inconsistencies in privacy labels across both platforms. On a positive note, *Credit Information* and *Financial Information* show the least number of inconsistencies, which is reassuring considering the sensitive nature of this data.

Case Studies to Understand the Inconsistencies. As shown earlier (Figure 10 and Figure 11), app developers often report different practices on different platforms. However, this analysis primarily focuses on the distribution of apps exhibiting inconsistencies across data types and purposes without providing insights into the specific areas of inconsistency. To better understand the inconsistencies in self-reported purposes, we conduct in-depth case studies with two data types: *Precise Location* and *Purchase History*. For each app that requests these data types, we perform a detailed comparison of the purposes stated in DSS and APL to identify the specific areas of mismatch.

Figure 12 illustrates the incorrectly matched purposes for the data type *Precise Location*. The figure shows that the highest mismatch rate occurs when a purpose is absent from one platform but present on the other. For instance, in the case of the app *Snapchat*, it collects *Precise Location* for *App Functionality* and *Advertising or Marketing* purposes on Google Play Store. However, on the Apple App Store, it additionally collects *Precise Location* for *Personalization* and *Analytics* purposes.

Similarly, in Figure 13, we present the inconsistency of apps for the *Purchase History* data type. We observe a similar pattern in this data type, with most inconsistencies arising from missing

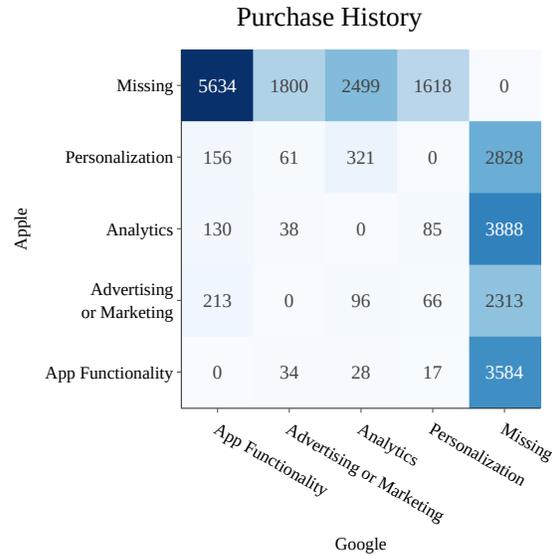


Figure 13: Confusion metric for Purchase History. Each row shows what the purpose in Apple for purchase history was mislabeled in Google, whereas each column shows what the purpose in Google was mislabeled in Apple.

labels in the complementary platform. Additionally, we identify inconsistencies where developers report collecting data for entirely different purposes. For instance, in the case of *Twitch TV*, it collects *Purchase History* for *App Functionality* on Google Play Store. In contrast, on the Apple App Store, it contains the same data type for *Analytics* and *Personalization* purposes.

5.3 Takeaway

In this section, we analyzed the consistency of privacy labels for the same apps across the two platforms. We find that 60% of the cross-listed apps had at least one inconsistency between APL and DSS. We further find that inconsistencies are highest for *Sensitive Information*, *Browsing History*, and *Emails or Text Messages* datatypes. Through a detailed analysis of datatype-purpose inconsistencies, we find that *Emails and Text Messages* when used for Advertising results in inconsistencies 96% of the time, indicating a concerning problem with disclosure of practices in privacy labels.

6 DISCUSSION

In this paper, we investigated the consistency of privacy labels with privacy policies and labels on other platforms. Our findings suggest a significant degree of inconsistency in privacy labels. Overall, there is a need for greater consistency in how privacy practices are disclosed to users within and between platforms. In this section, we discuss the implications of our findings and suggest potential solutions for improving the transparency and consistency of privacy practices. We also discuss the limitations of our study.

Comparison between the two labels. We analyzed the Data Safety Sections and the Apple Privacy Labels and found that the two labels cover different aspects of data practices. While both labels

provide information about the types of data that apps collect, Apple’s privacy label does not distinguish between data collection and sharing. Apple’s privacy label is more explicit about data practices like linkability, third-party advertising, and tracking. In contrast, data safety sections lack these details but does inform the users about the safety of their data (*Data Encryption*) and the choices that they have with developers (*Data Deletion Option*). These practices may be of particular interest to the users in light of the GDPR [27], which requires companies to provide a clear and explicit purpose for collecting and using personal data. Regulations like the GDPR and the CCPA also provide the right to delete the data to the users, which is covered in Data safety forms but not in the Apple privacy labels.

The comparison between the two labels highlights the importance of considering multiple sources of information when evaluating the data practices of apps. By combining the information provided by both labels, users can make more informed decisions about their privacy and the apps they choose to use.

Inconsistencies in disclosed practices across platforms. Our findings indicate inconsistencies between the privacy labels in the Apple Privacy Labels and the Google Data Safety Sections for the same apps. One possible reason for these inconsistencies is the confusing framework for privacy labels. While previous research [25] has shown that privacy labels are useful for developers and users, it also highlighted that filling privacy labels is perceived as challenging extra work. On top of that, developers are also unclear about definitions, which can result in confusion and inaccurate privacy labels. This confusion can be compounded by the fact that different platforms may use different terminology to describe similar practices. For example, in Apple’s privacy label, the term *tracking* is used when data collected is linked with third-party data for advertising purposes or when data is shared with a third party, which can be confusing to the developers, even when they are asked to pay close attention [25].

Another possible reason for the inconsistencies we observed is the casual attitude of some developers toward disclosing their data practices. Some developers may not fully understand the data practices of their apps or may not prioritize accurately disclosing this information to users. Finally, the platforms lack consistency checks to ensure accurate information in the privacy labels. Without these checks, developers can provide misleading or incomplete information about their data practices to meet the requirements.

We note that these inconsistencies can have serious consequences for users, as they may be confused about the privacy practices of their apps. If the practices disclosed in the privacy labels are inaccurate, it can reduce the efficacy of these labels as a tool for helping users make informed decisions about their privacy. Even worse, it could induce a false sense of security in users, who may assume that their data is being handled in a certain way when it is not.

Usability of Privacy Labels. Even though our analysis finds inconsistencies between privacy labels and privacy practices, evidence suggests that privacy labels generally carry more specific information about the practices. They include information about the types of data an app collects, how it is used, and whether it is shared with third parties. This information can be beneficial for users concerned

about their privacy and want to ensure they only use apps that respect their personal data.

However, the accuracy of privacy labels is not guaranteed. While developers are required to disclose their data practices to obtain a privacy label, there is no guarantee that the information they provide is accurate or complete. As such, platforms need to recognize that developers may not always be honest about their data practices. Therefore, it is necessary to have systems in place to verify the accuracy of privacy labels and to hold developers accountable for any discrepancies. This is particularly important because the false labels can create a false sense of security among the users.

One potential model for regulating privacy labels is a system similar to the one used for food nutrition labels, which are regulated by the Food and Drug Administration (FDA). A regulatory body could be established to oversee privacy labels and ensure they are accurate and consistent. This could help to build trust among users and encourage developers to be more transparent about their data practices.

Limitations. We used multiple heuristics to identify these apps for our cross-listed app analysis. However, these heuristics were likely unable to cover all the cross-listed apps. Moreover, for app-to-app comparison, we assume that both the Android and the corresponding Apple apps are similar enough to have similar privacy practices; however, these apps could be using different third-party libraries, which require different permission, leading to the inconsistency we observe in our analysis. However, the observed inconsistency (over 60% of the apps) strongly suggests that different third party libraries may not be the only factor contributing to the differences.

7 DATASET RELEASE

We plan to release the datasets curated in the work to the research community after publication. By sharing these datasets, we aim to contribute to the advancement of research in the domain of privacy practices and data disclosure on mobile app platforms. The availability of these comprehensive datasets will enable researchers to conduct further investigations, validate findings, and explore various aspects of privacy labels.

8 CONCLUSION

In conclusion, our large-scale measurements of Privacy Labels have provided valuable insights into the privacy practices of apps. By analyzing Data Safety Sections for 2.5M apps and Apple Privacy Labels for 1.38M apps, we provided a comprehensive picture of the privacy practices of the applications. On the one hand, privacy labels provide users with more specific information about the data practices of apps than traditional privacy policies. However, our comparison of Privacy Labels for cross-listed apps in the Play Store and Apple Store showed differences in the practices disclosed, indicating that developers are not consistently disclosing the same information on different platforms. This can confuse users and make it difficult to make informed decisions about which apps to use based on their privacy concerns. Overall, these findings highlight the importance of carefully reviewing Privacy Labels and other sources of information when evaluating the privacy practices of apps. They also suggest that there is a need for improved transparency and accountability in the app industry, as developers may

not always be accurately disclosing their data collection and use practices. A more transparent system will allow the consumers to be aware of the data collection and use practices of the apps and make informed decisions about their privacy.

ACKNOWLEDGMENTS

This work was supported by the NSF through awards: CNS-1942014 and CNS-2003129, and by gifts from Google, NVIDIA and Meta. Finally, we thank the reviewers for their fruitful discussions and recommendations.

REFERENCES

- [1] 2022. *JoMingyu/google-play-scraper: Google play scraper for Python*. <https://github.com/JoMingyu/google-play-scraper>
- [2] Mir Masood Ali, David G Balash, Chris Kanich, and Adam J Aviv. 2023. Honesty is the Best Policy: On the Accuracy of Apple Privacy Labels Compared to Apps' Privacy Policies. *arXiv preprint arXiv:2306.17063* (2023).
- [3] Kevin Allix, Tegawendé F. Bissyandé, Jacques Klein, and Yves Le Traon. 2016. AndroZoo: Collecting Millions of Android Apps for the Research Community. In *Proceedings of the 13th International Conference on Mining Software Repositories (Austin, Texas) (MSR '16)*. ACM, New York, NY, USA, 468–471. <https://doi.org/10.1145/2901739.2903508>
- [4] Understand app privacy & security practices with Google Play's Data safety section Computer Google Play Help. 2022. https://support.google.com/googleplay/answer/11416267?hl=en&visit_id=638094609270086018-2285502702&p=data-safety&rd=1#zippy=%2Csecurity-practices
- [5] David G Balash, Mir Masood Ali, Xiaoyuan Wu, Chris Kanich, and Adam J Aviv. 2022. Longitudinal Analysis of Privacy Labels in the Apple App Store. *arXiv preprint arXiv:2206.02658* (2022).
- [6] Rebecca Balebako, Abigail Marsh, Jialiu Lin, Jason I Hong, and Lorrie Faith Cranor. 2014. The privacy and security behaviors of smartphone app developers. (2014).
- [7] Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. 2015. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. 63–74.
- [8] Theodore Book, Adam Pridgen, and Dan S Wallach. 2013. Longitudinal analysis of android ad library permissions. *arXiv preprint arXiv:1303.0857* (2013).
- [9] Fred H Gate. 2010. The limits of notice and choice. *IEEE Security & Privacy* 8, 2 (2010), 59–62.
- [10] Lorrie Faith Cranor. 2012. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.* 10 (2012), 273.
- [11] Lorrie Faith Cranor. 2022. Mobile-app privacy nutrition labels missing key ingredients for success. *Commun. ACM* 65, 11 (2022), 26–28.
- [12] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. 2020. Ask the experts: What should be on an IoT privacy and security label?. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 447–464.
- [13] Pardis Emami-Naeini, Janarth Dheendrayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. 2021. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase IoT devices?. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 519–536.
- [14] Grace Fox, Colin Tonge, Theo Lynn, and John Mooney. 2018. Communicating compliance: developing a GDPR privacy label. (2018).
- [15] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. 2022. Helping mobile application developers create accurate privacy labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 212–230.
- [16] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj Agarwal. 2016. How short is too short? implications of length and framing on the effectiveness of privacy notices. In *Twelfth symposium on usable privacy and security (SOUPS 2016)*. 321–340.
- [17] Ashish Hooda, Matthew Wallace, Kushal Jhunjhunwalla, Earlene Fernandes, and Kassem Fawaz. 2022. SkillFence: A Systems Approach to Practically Mitigating Voice-Based Confusion Attacks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–26.
- [18] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "Nutrition Label" for Privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (Mountain View, California, USA) (SOUPS '09). Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. <https://doi.org/10.1145/1572532.1572538>
- [19] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. 2010. Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (CHI '10). Association for Computing Machinery, New York, NY, USA, 1573–1582. <https://doi.org/10.1145/1753326.1753561>
- [20] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3393–3402.
- [21] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as Part of the App Decision-Making Process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 3393–3402. <https://doi.org/10.1145/2470654.2466466>
- [22] Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. 2023. Unpacking Privacy Labels: A Measurement and Developer Perspective on Google's Data Safety Section. *arXiv preprint arXiv:2306.08111* (2023).
- [23] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. 2021. Are iPhones Really Better for Privacy? Comparative Study of iOS and Android Apps. *arXiv preprint arXiv:2109.13722* (2021).
- [24] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. 2022. Goodbye tracking? Impact of iOS app tracking transparency and privacy labels. *arXiv preprint arXiv:2204.03556* (2022).
- [25] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I Hong. 2022. Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels. In *CHI Conference on Human Factors in Computing Systems*. 1–24.
- [26] Jialiu Lin. 2013. *Understanding and capturing people's mobile app privacy preferences*. Ph.D. Dissertation. Carnegie Mellon University.
- [27] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. 2018. The privacy policy landscape after the GDPR. *arXiv preprint arXiv:1809.08396* (2018).
- [28] Aleecia M McDonald, Robert W Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. 2009. A comparative study of online privacy policies and formats. In *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 37–55.
- [29] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. 2015. A design space for effective privacy notices. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*. 1–17.
- [30] Gian Luca Scoccia, Marco Autili, Giovanni Stilo, and Paola Inverardi. 2022. An empirical study of privacy labels on the Apple iOS mobile app store. (2022).
- [31] Haoyu Wang, Yao Guo, Ziang Ma, and Xiangqun Chen. 2015. Wukong: A scalable and accurate two-phase approach to android app clone detection. In *Proceedings of the 2015 International Symposium on Software Testing and Analysis*. 71–82.
- [32] Yue Xiao, Zhengyi Li, Yue Qin, Jiale Guan, Xiaolong Bai, Xiaojing Liao, and Luyi Xing. 2022. Lalaine: Measuring and Characterizing Non-Compliance of Apple Privacy Labels at Scale. *arXiv preprint arXiv:2206.06274* (2022).
- [33] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. 2022. How Usable Are iOS App Privacy Labels? *UMBC Faculty Collection* (2022).

A MAPPING DATA TYPES FROM GOOGLE DATA SAFETY SECTIONS TO APPLE PRIVACY LABELS

In Table 3, we show how we map the data types in DSS to APL. The mapping is done based on the description provided in the documentation of the privacy labels.

This table is a continuation of the table in Section 5, Table 2. We observe that most datatypes have a one-to-one mapping while only a few, like *Music Files*, *Other In-App Messages*, *Other Info*, do not have a direct mapping to APL.

B DISTRIBUTION OF HIGH-LEVEL CATGEORIES

In Figure 14, we show the complete distribution of the inconsistent apps with regard to their datatypes across APL and DSS.

In Figure 15, we show the distribution of the high-level datatypes of apps found in the App Store. We observe that the most used datatypes are: *Diagnostics*, *Identifiers*, *Location*, and *Usage Data*

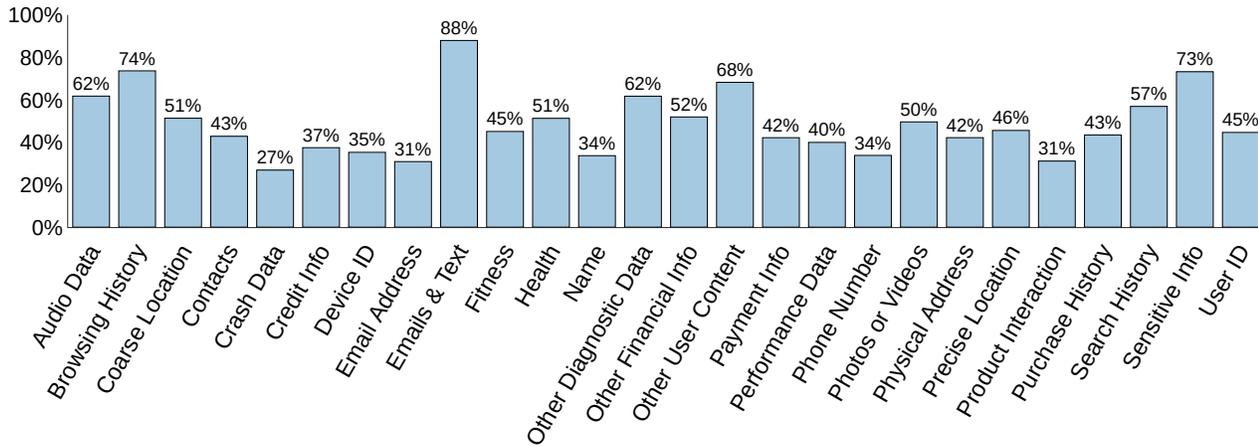


Figure 14: The complete distribution of inconsistent apps with datatypes. Each datatype is normalized with the number of apps using that particular datatype on either platform.

| Google DataType | → | Apple DataType |
|------------------------------|---|-----------------------|
| Precise Location | → | Precise Location |
| Name | → | Name |
| Email Address | → | Email Address |
| Phone Number | → | Phone Number |
| Race And Ethnicity | → | Sensitive Info |
| User Ids | → | User Id |
| User Payment Info | → | Payment Info |
| Credit Score | → | Credit Info |
| Other Financial Info | → | Other Financial Info |
| Purchase History | → | Purchase History |
| Health Info | → | Health |
| Fitness Info | → | Fitness |
| Other In-App Messages | → | N/A |
| Photos | → | Photos Or Videos |
| Videos | → | Photos Or Videos |
| Voice Or Sound Recordings | → | Audio Data |
| Music Files | → | N/A |
| Other Audio Files | → | N/A |
| Contacts | → | Contacts |
| App Interactions | → | Product Interaction |
| Other User-Generated Content | → | Other User Content |
| In-App Search History | → | Search History |
| Other Actions | → | N/A |
| Web Browsing History | → | Browsing History |
| Other App Performance Data | → | Other Diagnostic Data |
| Other Info | → | N/A |

Table 3: Table showing the remaining mapping of datatypes from Data Safety Card to Apple Privacy Label

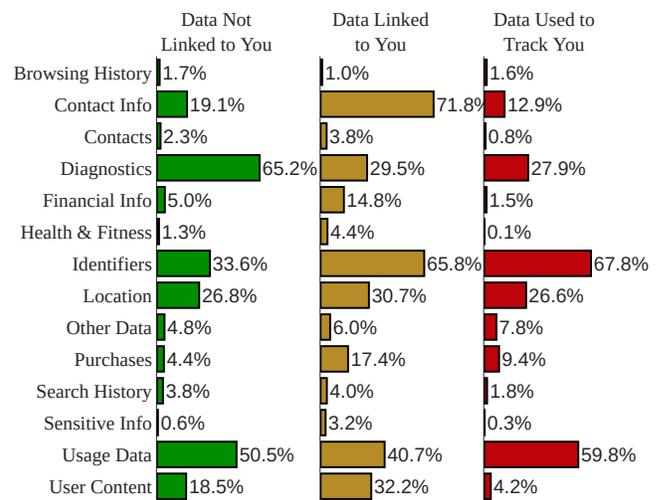


Figure 15: Distribution of datatypes with privacy types for Apple Privacy Labels.