2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) | 979-8-3503-2720-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/EuroSPW59978.2023.00022

Comparing Privacy Label Disclosures of Apps Published in both the App Store and Google Play Stores

David Rodriguez*, Akshath Jain†, Jose M. del Alamo*, Norman Sadeh†
*ETSI Telecomunicación, Universidad Politécnica de Madrid
{david.rtorrado, jm.delalamo}@upm.es
†School of Computer Science, Carnegie Mellon University
arjain@andrew.cmu.edu, sadeh@cs.cmu.edu

***Abstract*—Apple and Android introduced privacy labels in 2020 and 2022 respectively as a way of providing consumers with succinct summaries of mobile apps' more salient data practices. A number of apps are published in both stores, offering us the opportunity to compare their privacy label disclosures in the two app stores. This paper compares the data practices privacy labels are intended to capture in each store. It then proceeds to analyze the disclosures of 822 apps published in both app stores, focusing on possible discrepancies. This analysis reveals that privacy label disclosures of what is ostensibly the same mobile app can be quite different. We discuss the different possible reasons behind these differences, including the possibility that these discrepancies might be indicative of potential privacy compliance issues. In particular, focusing on data collection disclosures of five different data types (location, contact info, sensitive info, identifiers, and health & fitness) we find discrepancies between iOS and Google Play privacy label disclosures in 66.5% of the mobile apps we analyze.**

1. Introduction

The internet's and digital technologies' explosive growth in recent years has greatly influenced how people interact, consume media, and transact business. Yet, when personal data is involved, the rising use of digital technology also poses grave concerns to people's privacy. In fact, a KPMG study [14] reports that 86% of Americans claim that data privacy is a recent growing concern for them.

In response to these worries, big tech companies (i.e., Google & Apple) have developed privacy labels, which provide users with information about the data collected and shared by apps and the way the data is protected [13]. These labels are designed with the aim of helping users make informed decisions before installing applications based on their privacy practices.

Apple, one of the leading participants in the mobile operating system industry, has a privacy label section built into the App Store. The iOS privacy labels, which went into effect in December 2020, ask app developers to provide explicit breakdowns into the various types of data they gather, including contact information, location data, and browser history. In a similar manner, Android introduced its own privacy labels into the data safety section on the Google Play Store in April 2022.

Privacy labels in both ecosystems must be declared by app developers, which invite users to rely on the veracity of their statements. Nevertheless, app developers may intentionally or unintentionally be omitting information. Our study attempts to shed light on this issue by presenting a comparison between Android and iOS privacy labels. The contributions of this work are as follows:

- A Mapping between iOS and Android labels defining the practices and data types disclosed that could be directly compared between each platform.
- The design of a method for reliably finding iOS applications on Google Play Store (and vice versa).
- A comprehensive comparison between the mapped privacy labels for 822 identical Android and iOS applications.
- A static analysis of 560 Android applications' source code looking for precise and coarse location collection.

To the best of our knowledge, this is the first research work comparing Android and iOS privacy labels usage by app developers. The proposed mapping along with a reliable method for finding the same applications in both marketplaces will enable new studies to be carried out for the benefit of all mobile device users.

The outline of this document is as follows. Section 2 describes iOS and Android privacy labels, presents the mapping identified between these labels, and documents the related work. In section 3, the method followed for the privacy labels comparison is presented. An analysis at scale is conducted in Section 4, highlighting the differences identified. Those differences are discussed in section 5 with the relevant findings. Potential threats to validity are exposed in Section 6 and the paper's final conclusions are reported in Section 7.

2. Background & Related Work

Privacy labels have emerged as a result of the readability and comprehension problems of privacy policies [5], [25], [26]. Their scope is to encourage developers to disclose their applications' privacy practices following a template that allows a better understanding by users. This section will define the particularities of iOS and Android privacy labels. Then, we propose a mapping between the correspondent practices and data types in both ecosystems. Finally, the closest related work is presented highlighting the differences with our contributions.

DOI 10.1109/EuroSPW59978.2023.00022

2.1. iOS Privacy Labels

Since the addition of Apple's privacy labels in December 2020 [20], developers have been asked to describe four privacy aspects of their apps (see Figure 1, bottom): data item, data type, data purpose, and data practice. The data item is the specific data to be collected by the app, which belongs to a higher data type (i.e., category), for example, the "name" data item belongs to the "contact info" data type. The purpose is the app's reason for accessing these data e.g. analytics or app functionality. Data practice describes to what extent the piece of data will be linked to the user identity i.e Data not Linked to You, Data Linked to You, and Data used to Track You. Data not Linked to You refers to de-identified or anonymized data. Data Linked to You refers to data linked to the user's identity e.g via their account, device, or other details. Data used to Track You refers to further linking the data with new third-party data for advertising purposes, or sharing it with a data broker. It's worth noting that in Apple parlance, data collection implies "*sending the data off-device in a way that allows developers or third-parties to access it for a period longer than what is necessary to service the transmitted request in real time*" [4].

iOS app developers are encouraged to declare the type of data to which each practice alludes, as well as the category in which the data may fall and the purpose for accessing it. Data items, data categories, and purposes are declared in the privacy labels from among those provided by apple [4].

2.2. Android Privacy Labels

Android privacy labels went into effect in April 2022, following a similar overall format for data item and data types, but with some remarkable differences in purposes and practices, when compared to iOS labels. Android practices distinguish between data collection and data sharing. Android data collection refers to the same concept as iOS but, interestingly, according to Android terms, a piece of data does not need to be disclosed as "collected" when it is sent off-device over an encrypted connection.

Data sharing refers to a broader concept than iOS Data Used to Track, where "Sharing" refers to transferring user data to a third party. This practice might not be necessarily disclosed in the labels if the data is previously anonymized. As can be seen in Figure 1 (left), data categories, data types, and purposes follow a similar disclosing format compared to iOS [6].

2.3. Privacy Labels mapping

The mapping between iOS and Android labels is not straightforward. By comparing data practices, data items, and purposes we could observe many-to-many relationships between iOS and Android labels. An example of the intricacies of the relationship between labels is shown in Figure 2, where iOS Developer's Advertising or Marketing purpose is mapped to Android Advertising or marketing and Developer communications purposes, while iOS Third-Party Advertising is also mapped to Android Advertising or marketing.

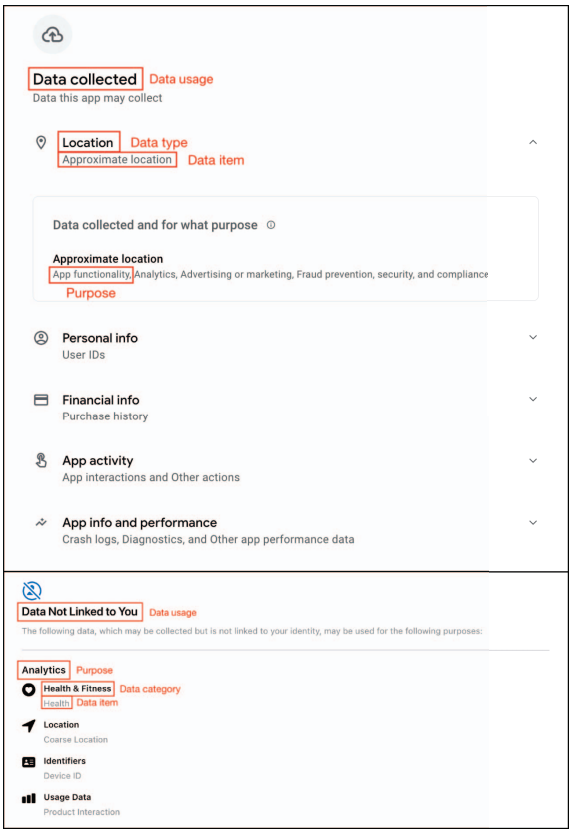

Figure 1. Android (top) and iOS (bottom) privacy labels example

Figure 2. Example of mapping between iOS (left) and Android (right) purpose labels

The many-to-many relationship between the platforms' labels hinders making a straightforward comparison between all iOS and Android data labels. We have therefore limited our comparison to exclusively identical practices and data (i.e., one-to-one relationships). Data Used to Track in iOS alludes to the specific purpose of tracking, which is a subset of possible uses given to data shared in Android. The great difference between these two labels made us limit our comparison to data collection practices and the data items shown in table 1. Purposes were excluded from our analysis.

2.4. Related Work

The notion of privacy labels is an adaptation of nutrition labels introduced by Kelley et al. [7]. Their work aimed to create an information design (i.e., privacy labels) that could improve the comprehensiveness and understanding of privacy policies. In addition, their privacy labels were intended to disclose the collection, use, and sharing of personal data by organizations in an easy-to-understand format. However, the use of privacy labels for

TABLE 1. MAPPING BETWEEN IOS AND ANDROID PRIVACY LABELS

	iOS Label	Android Label
Data practice	Data Linked to You	Data collected
	Data Not Linked to You	
Data items	Sensitive Info	Race and ethnicity
		Political or religious beliefs
		Sexual orientation
	Precise Location	Precise location
	Coarse Location	Approximate location
	Name	Name
	Email Address	Email address
	Phone Number	Phone number
	Physical Address	Address
	Other User Contact Info	Other info
	User ID	User IDs
	Device ID	Device or other IDs
	Health	Health info
	Fitness	Fitness info

the mobile ecosystem was lately proposed by Kelley et al. [8].

Since the adoption of privacy labels by Apple, their various uses and advantages are under discussion. Zhang et al. [25] checked the effectiveness of iOS privacy labels by comparing their readability, comprehensibility, salience, and relevance with those of privacy policies. They conducted an in-depth interview study with 24 iPhone users to investigate their experiences, understanding, and perceptions of Apple's privacy labels. The research concluded that "*Apple's privacy labels still do not fully support users' understanding of disclosed application privacy practices*".

Other related works have analyzed the content of iOS privacy labels [11], [13], [18], and assessed their trustworthiness [9], [24]. In particular, Xiao et al. [24] conducted the first comparison between apps' privacy labels and their actual behavior by conducting a dynamic analysis on 5,102 iOS apps, reporting inconsistencies in 3,423 privacy labels. A study of a similar nature was conducted by Koch et al. [9], analyzing 1,687 iOS apps for privacy labels' correctness. During their analysis, they could observe that "*At least 276 [...] apps violate their privacy label by transmitting data without declaration, showing that the privacy labels' correctness was not validated during the app approval process*".

But even fewer studies [3], [19] have addressed Android privacy labels. Closer to our work, Mozilla has conducted a study [19] on the Android top 20 free apps and top 20 paid apps comparing the privacy policy with the privacy labels for each app. The study concludes by finding discrepancies between privacy policies and privacy labels for nearly 80% of the apps reviewed.

Likewise, a reduced number of studies have addressed the comparison between iOS and Android from a privacy perspective [2], [10], [12]. For example, Kollnig et al. [10] used static and dynamic analysis techniques to assess iOS and Android applications identifying personal data leaks. During this comprehensive work, they also analyzed the recipient's identity and location to uncover compliance issues.

The related works described above have either focused on analyzing privacy labels in a single ecosystem or have compared apps' privacy behavior in both domains. To the best of our knowledge, no prior work has addressed a comparison between iOS and Android privacy labels.

3. Method

In this Section, we describe our analysis methodology. We begin by detailing the iOS and Android apps' selection process in Section 3.1. Afterward, in Section 3.2, we provide details on how we collected the iOS and Android privacy labels. Finally, in Section 3.3, we describe the method to perform a static analysis on the apps and check whether the privacy labels match with actual apps' code.

3.1. iOS and Android apps selection

Conducting a comparison between the privacy labels of App Store and Google Play Store applications requires a dataset containing matching applications from both markets. To create a dataset of this kind we pursued the following steps.

iOS apps selection. We scraped the App Store website to collect the name and other details (e.g., privacy policy URL) of the whole list of available applications. From the list, we randomly selected a subset to conduct this study.

Finding matches in Google Play Store. We created an automated method to find the Android app matching each iOS app in our dataset. The method follows a two-phase pipeline: 1) looking for the iOS app name in the Play Store apps' search bar, and 2) comparing both apps' information to determine if they are actually the same app.

The first phase was a straightforward process where we selected the first result (i.e., potentially the most similar according to Play Store). In order to perform the second phase, we first evaluated several approaches on which we based our comparison: application name, developer name, website URL, privacy policy URL, and app's logo. Our evaluation determined that the logo comparison is the most reliable approach to determining if two applications are indeed the same on both platforms.

The logo comparison method downloads both logos and then performs a comparison based on the Structural Similarity Index Measure (SSIM) [23] between them. Our initial validation in a random sample of 30 applications outputted a 1.0 precision score when using an SSIM threshold of 0.9 in the range of [0-1]. However, we found a few false positive cases when validating a larger dataset.

Apps filtering. Since we needed to ensure that the dataset contained only matching applications for the labels comparison, we conducted an additional step to discard potentially incorrectly tagged apps. This filtering consisted in comparing the privacy policies of both apps available on their corresponding platforms.

Before comparing the privacy policies, we required collecting and ensuring they are indeed privacy policies.

We employed Selenium to retrieve those policies loaded through dynamic code. Likewise, it was necessary to discard those URLs leading to non-privacy policies' websites (e.g., landing pages). To do so, we used a machine learning-based classifier that allows us to differentiate between privacy policies and other texts. This classifier is based on the Support Vector Machines (SVM) algorithm and was trained with 195 manually classified texts, achieving 98.76% precision, 97.56% recall, and 98.15% F1 score when evaluated against 100 unseen English texts. After discarding non-privacy policies texts, we compared them by computing the cosine similarity [21] to dismiss the applications that did not perfectly match (i.e., cosine similarity of 1.0) based on the similarity between privacy policies.

3.2. Privacy labels collection and comparison

iOS labels collection. We iterated the process of scraping the privacy labels for each iOS application. The iOS privacy labels are dynamically loaded in a pop-up window after clicking on a "See details" button. This mandatory interaction requires the use of Selenium to trigger the button and load the HTML. Afterward, the BeautifulSoup python library [16] parses the HTML code and we iterate the collection of practices, types of personal data, and purposes.

Android labels collection. Google Play Store does not reuse the same web resource for the privacy labels disclosure as iOS does. Instead, it serves a different resource where the privacy practices of the app are disclosed (i.e., the safety section). This allows us to scrap this info and load it with BeautifulSoup to parse the HTML. Thus, we can collect the practices, types of personal data, and purposes in a shorter period of time than for iOS apps.

Labels comparison. As we explained in Section 2, we have two types of practices in iOS and Android privacy labels. The Android "data collected" practice could be directly compared with the "Data Linked to You" and "Data Not Linked to You" in iOS, while the data shared in Android cannot be compared with the "Data Used to Track You" due to the different meaning explained in Section 2. Therefore, we limit the comparison of privacy labels to data collected usage. Along the process two different topics will be compared: 1) the aggregated personal data types in each ecosystem and 2) the data collected by the same apps (i.e., Android app and iOS app).

3.3. Comparing Android privacy labels to actual app code

The popularity of third-party libraries has grown to the point that it has surpassed the amount of developers' source code [22]. In some cases, developers can be unaware of the whole behavior of these libraries, which could lead them to incorrectly select their app's privacy labels. Sometimes developers simply do not know or understand in detail how their application behaves in terms of privacy. We also aim to look at the code of Android apps and compare our analysis of the code with disclosures provided in privacy labels.

To do so, we relied on the following pipeline to perform static app analysis. First, we use an unofficial API [1] to download the Android applications. Afterward, we decompile and re-build the java code of the apk file with jadx [17], obtaining the Manifest file along with the application's smali and java code.

Once the application's source code is built, we check the permissions in the manifest file to see whether the app is requesting access to personal data. Additionally, we automatically inspect all java files looking for the API calls that access the personal data. We use this information to perform the comparison with the app privacy labels looking for dissimilarities.

4. Evaluation in the wild

This section describes the evaluation conducted to compare Android and iOS privacy labels following the method described in Section 3.

To collect the dataset of apps and privacy labels we started by looking for details on 35k randomly selected apps in the App Store. While scraping the website to get the name and information for each app, we simultaneously searched for the matching Android application. After this process, we found almost 11k matching candidate pairs of iOS and Android apps. We further analyzed each pair by applying the methods described in Section 3.1, successfully identifying 1,423 exact matches.

Out of these 1,423 exact matches, 1,106 of the iOS apps have privacy labels, while only 911 Android apps do. The intersection set yields 822 apps, namely apps that have published labels in both app stores. Our analysis focuses on these 822 apps.

Figure 3 shows the overall number of Android (green) and iOS applications (blue) claiming to collect each data type. As can be seen, there are mismatches in the privacy labels for some data types. For example, precise and approximate location have a difference of 54% in favor of iOS and 36% in favor of Android respectively for each of these data types. This suggests that when collecting location data, iOS apps tend to favor precise location and Android apps favor approximate location.

Another significant mismatch can be seen between user and device identifiers. 43% more iOS apps disclose collecting user identifiers, while 30% more Android apps claim to collect device identifiers. However, the most alarming gap is between the number of apps claiming to collect sensitive information. According to Apple, sensitive information refers to racial or ethnic data, sexual orientation, disability, and religious or philosophical beliefs, among others. Although we can observe a low number of apps reporting the collection of this data type, we would expect to see exactly the same number of apps on both platforms due to its sensitivity. Interestingly too, almost twice as many iOS apps declare to collect user health data, considered by privacy regulations (i.e., GDPR) as sensitive data.

Nevertheless, the most compelling comparison that can be done between Android and iOS in relation to privacy labels is to determine whether the same types of personal data are claimed to be collected for the same application (in both ecosystems). This is shown in Figure 4, where we report the number of Android apps that disclose the collection of a given data type 1) only in Android (green), 2) only in iOS (blue), 3) in both (yellow). The

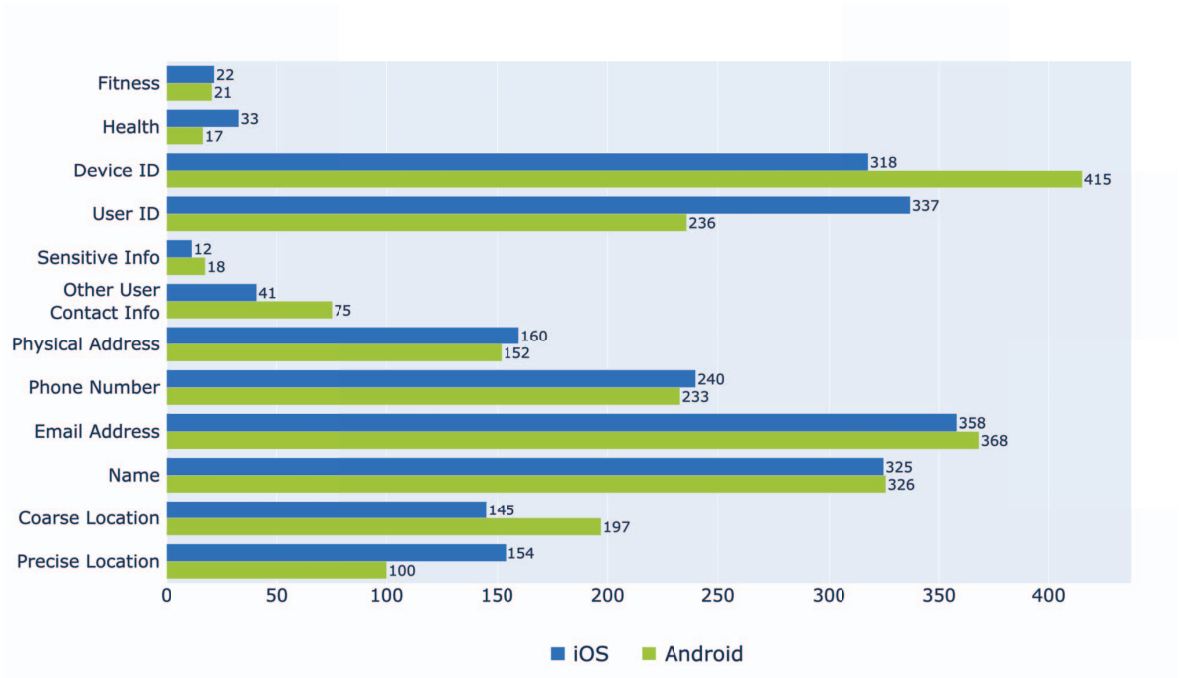

Figure 3. Comparison between the number of Android (green) and iOS (blue) apps declaring the collection of each data type

remaining applications up to the 822 analyzed correspond to those that do not declare the collection of the data in any marketplace. Again, a remarkable gap between Precise location, Coarse location, User ID, and Device ID can be observed Surprisingly, we can observe that only seven applications match reporting the collection of sensitive information, while 16 differ.

The comparison of privacy labels is limited to what developers report their applications to do. However, apps may intentionally or unintentionally be accessing non-reported personal data. To assess their statements, we managed to successfully download and perform static analysis on 560 out of the 822 Android apps.

First, we checked whether applications have access to two types of personal data: coarse location and precise location. Figure 5 shows a comparison between the apps that request access to these data and whether their collection is declared in the privacy labels. As can be seen, 36.6% and 38.2% of the apps that do not declare collecting coarse and precise location respectively, request permissions to access these data.

We further reconstructed the java source code out of the smali code for the 560 applications. As described in Section 3.3, we have looked for the API calls that retrieve the precise and coarse location, observing that for the 54 apps found accessing these two data types, none of them disclose their collection in the privacy labels. Although not conclusive due to static analysis limitations [15] our findings suggest a mismatch between what labels disclose and what apps may actually be doing.

5. Discussion

A **substantial 66.5%** of the 822 applications analyzed **show potential discrepancies between iOS and Android privacy labels**. Moreover, out of the 503 that claim to collect personal data in both marketplaces, **only 16 (3.2%) agree on all the data types mapped** in Section 2. These results suggest notable differences in data practice disclosure in the iOS and Google Play app stores for apps that one would otherwise have expected to have identical or nearly identical data practice disclosures.

There are two major possible explanations for our observations: 1) apps indeed behave differently in each ecosystem and the privacy labels are correct and consistent with that, or 2) the difference suggests apparent inconsistencies of the data practice disclosures of the privacy labels. If it is the first case, we would have applications with exactly the same privacy policy while carrying out different privacy practices. In contrast, the second case is supported by evidence that Android privacy labels show apparent inconsistencies with apps' code. As noted in Section 4, 248 apps (44.3%) request permission to collect coarse or precise location even though none of them are disclosing it in the privacy labels. Moreover, we found 54 of those applications accessing these data in their source code, neither reported on the labels.

Of course, accessing personal data that has not been reported in the data collected section of the privacy labels does not imply non-compliance. As stated in Section 2, in privacy labels parlance "data collected" refers to sending the data off-device, and therefore it might be the case that data is retrieved but never sent out by apps. Nevertheless, it is remarkable to notice that none of the apps found accessing the data did report the collection.

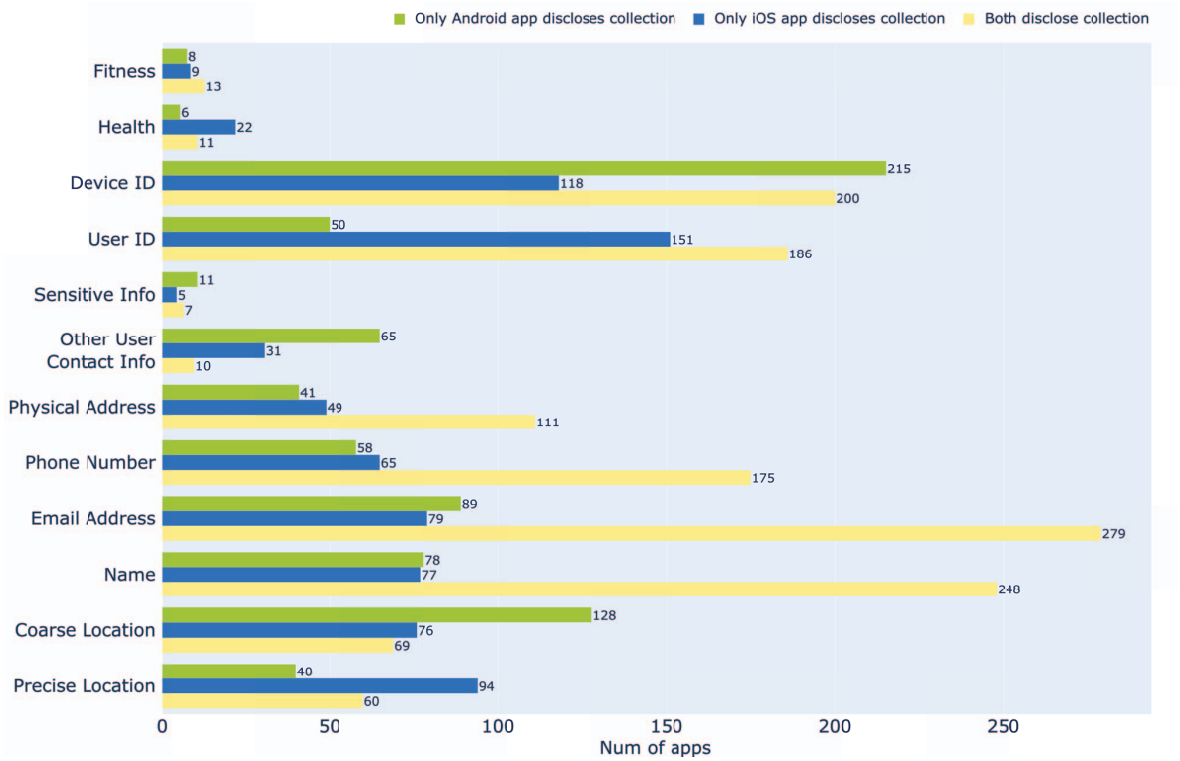

Figure 4. Number of apps disclosing a data item collection in the privacy labels

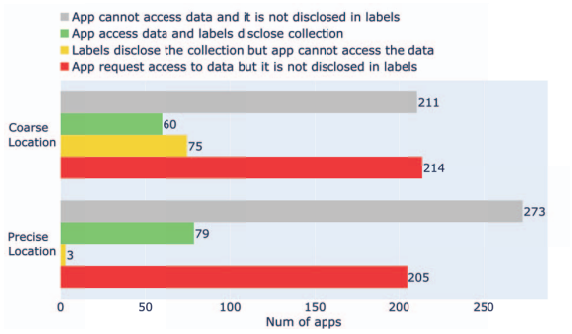

Figure 5. Comparison between Android apps' permissions requested and data collection disclosed in the privacy labels

Differences between the number of applications disclosing collection of precise location and coarse (approximate) location in Android and iOS. In figures 3 and 4 we could see that iOS apps mostly collect precise location while Android does the same with coarse location. Apple considers as precise location the latitude and longitude coordinates with three or more decimal places, equivalent to a dispersion of 110 meters. Google does not define what it considers as precise location but determines that the approximate location is the one capable of locating the user in an area of 3 square kilometers. On the other hand, Apple considers approximate location to be the latitude and longitude coordinates with two decimal places or less, which is equivalent to 1.1 km of dispersion. Thus, the divergences between the definitions of these two data items highlight the disparity between iOS and Android privacy labels, but do not justify the noticeable differences seen in our results.

Apparent inconsistencies found about users' health data being collected in the same applications. We manually inspected apps in two situations: 1) an iOS app that claims to be collecting health data while the same Android app does not; and, 2) the exact opposite situation. Matching with the first situation (1), the Forever GoFit app claims to collect users' health data in iOS but not in Android (with over 50k downloads in Google Play Store). Nevertheless, in the Google Play Store app's description, they state their app "*counts your steps, calories, active time, distance, and record and analyze your sleep and heart rate*". Interestingly, it was last updated during the last month and these functionalities also appear in the Apple Store app's description, along with the same apps' screenshots in both marketplaces. We found the opposite (2) in the VitalFlo Health app, which describes exactly the same functionalities and app screenshots in both marketplaces, while only Android discloses health collection. The main app's purpose is to "*record and track your lung function and symptoms, and automatically sync with your doctor*", where evident health data collection is occurring.

6. Threats to validity

Construct validity. Not all data types can be compared among platforms, only those for which we have found a one-on-one relationship. This may make our comparison not generalizable to other data types. The

number of apps we have analyzed is large and the proposed method for finding iOS apps on Android has a high accuracy, which has been increased by performing an in-depth comparison between privacy policies. This makes negligible the possible error of incorrectly matching one app to another. However, obtaining such a high accuracy along with the fact that not all applications disclose the privacy labels, involved a considerable reduction in the size of the evaluation dataset. The jadx tool decompiles and builds Java files for all Android applications, even if the construction of the source code is not properly achieved. This may lead to an increase in false negative cases of applications retrieving location data type which nevertheless does not introduce false positive results.

External validity. App Store unlike Google Play Store asks developers for a monthly fee to maintain apps on the market. This divergence could lead to greater attention on the app's details provided, and a lower number of unmaintained apps when compared to Google Play Store. This may also be a bias in favor of the iOS applications when comparing the privacy labels.

7. Conclusions

Increasing privacy awareness by users and regulatory pressure by supervisory authorities has led large technology companies (i.e., Google & Apple) to focus their efforts on creating privacy labels for their apps marketplace. In this article, we have compared these labels to check if they are consistent, or if they involve considerable differences for the same applications in the different ecosystems.

Following the proposed method, we collected and compared the privacy labels of 822 apps. Through the comparison, we observed that only 3.2% of the apps disclosing the collection of data coincide in both ecosystems, while a remarkable 44.3% of the analyzed apps are requesting permissions to retrieve data they have not disclosed in their labels. The divergences between iOS and Android privacy labels for applications with the same privacy policy confirm the existence of apparent privacy inconsistencies. We hope that these findings serve as a call to action for regulators. In future work, we aim to conduct a dynamic analysis of the apps to further analyze the actual apps' behavior and compare it with the privacy labels.

8. Acknowledgements

This research has been partially supported by the project TED2021-130455A-I00 funded by MCIN/AEI/10.13039/501100011033 and the Europea Union "NextGenerationEU"/PRTR. Grants from the National Science Foundation Secure and Trustworthy Computing program (CNS-1801316, CNS-1914486) and an unrestricted Privacy Faculty Award from Google helped fund this study.

References

[1] Rehmat Alam. Command line google play apk downloader. github repository. https://github.com/rehmatworks/gplaydl, 2021. Accessed: 2023-03-29.

[2] Ying Chen, Heng Xu, Yilu Zhou, Sencun Zhu, and George Washington. Is this app safe for children?: a comparison study of maturity ratings on android and ios applications. pages 201–212. Association for Computing Machinery (ACM), 5 2013.

[3] Lorrie Faith Cranor. Mobile-app privacy nutrition labels missing key ingredients for success. *Communications of the ACM*, 65:26–28, 11 2022.

[4] Apple Developer. App privacy details. https://developer.apple.com/app-store/app-privacy-details/. Accessed: 2023-03-29.

[5] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. Helping mobile application developers create accurate privacy labels. *Proceedings - 7th IEEE European Symposium on Security and Privacy Workshops, Euro S and PW 2022*, pages 212–230, 2022.

[6] Play Console Help. Provide information for google play's data safety section. https://support.google.com/googleplay/android-developer/answer/10787469?hl=en, 2022. Accessed: 2023-03-29.

[7] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, New York, NY, USA, 2009. Association for Computing Machinery.

[8] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. *Conference on Human Factors in Computing Systems - Proceedings*, pages 3393–3402, 2013.

[9] Simon Koch, Malte Wessels, Benjamin Altpeter, Madita Olvermann, Martin Johns, and : Datenanfragen. Keeping privacy labels honest. In *Privacy Enhancing Technologies Symposium*, pages 486–506, 2022.

[10] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. Are iphones really better for privacy? comparative study of ios and android apps. *Proceedings on Privacy Enhancing Technologies*, 2022:6–24, 9 2021.

[11] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. Goodbye tracking? impact of ios app tracking transparency and privacy labels. *ACM International Conference Proceeding Series*, 22:508–520, 6 2022.

[12] Lydia Kraus, I. Wechsung, and S. Möller. A comparison of privacy and security knowledge and privacy concern as influencing factors for mobile protection behavior. In *Workshop on Privacy Personas and Segmentation (PPS)*, 2014.

[13] Yucheng Li, Deyuan Chen, Tianshi Li, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. Understanding ios privacy nutrition labels: An exploratory large-scale analysis of app store data. *Conference on Human Factors in Computing Systems - Proceedings*, 4 2022.

[14] Orson Lucas, Martin Sokalski, and Rob Fisher. Corporate data responsibility: Bridging the consumer trust gap. https://advisory.kpmg.us/articles/2021/bridging-the-trust-chasm.html?utm_source=vanity&%3Butm_medium=referral&%3Bmid=m-00005652&%3Butm_campaign=c-00107353&%3Bcid=c-00107353. Accessed: 2023-03-29.

[15] Dimitri Prestat, Naouel Moha, and Roger Villemaire. An empirical study of android behavioural code smells detection. *Empirical Software Engineering*, 27:1–34, 12 2022.

[16] PyPI. beautifulsoup4. https://pypi.org/project/beautifulsoup4/. Accessed: 2023-03-29.

[17] GitHub repository. Dex to java decompiler. https://github.com/skylot/jadx. Accessed: 2023-03-29.

[18] Gian Luca Scoccia, Marco Autili, Giovanni Stilo, and Paola Inverardi. An empirical study of privacy labels on the apple ios mobile app store. *Proceedings - 9th IEEE/ACM International Conference on Mobile Software Engineering and Systems, MOBILESoft 2022*, pages 114–124, 2022.

[19] Anne Stopper and Jen Caltrider. See no evil: Loopholes in google's data safety labels keep companies in the clear and consumers in the dark. mozilla foundation. https://foundation.mozilla.org/en/campaigns/googles-data-safety-labels/, 2021. Accessed: 2023-03-29.

[20] Apple Support. About privacy information on the app store and the choices you have to control your data. https://support.apple.com/en-us/HT211970, 2021. Accessed: 2023-03-29.

[21] Vikas Thada and Vivek Jaglan. Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*, 2, 2013.

[22] Haoyu Wang, Yao Guo, Ziang Ma, and Xiangqun Chen. Wukong: A scalable and accurate two-phase approach to android app clone detection. pages 71–82. Association for Computing Machinery, Inc, 7 2015.

[23] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 4 2004.

[24] Yue Xiao, Zhengyi Li, Yue Qin, Xiaolong Bai, Jiale Guan, Xiaojing Liao, and Luyi Xing. Lalaine: Measuring and characterizing non-compliance of apple privacy labels at scale. *arXiv*, June 2022.

[25] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How usable are ios app privacy labels? In *Privacy Enhancing Technologies Symposium*, pages 204–228, 2022.

[26] Shikun Zhang and Norman Sadeh. Do privacy labels answer users' privacy questions? In *Network and Distributed System Security Symposium*, 2023.