Freely Given Consent? Studying Consent Notice of Third-Party Tracking and Its Violations of GDPR in Android Apps

Trung Tin Nguyen
CISPA Helmholtz Center for Information Security
Saarland University
tin.nguyen@cispa.de

Michael Backes
CISPA Helmholtz Center for Information Security
backes@cispa.de

Ben Stock
CISPA Helmholtz Center for Information Security
stock@cispa.de

ABSTRACT

Adopted in May 2018, the European Union's General Data Protection Regulation (GDPR) requires the consent for processing users' personal data to be *freely given, specific, informed, and unambiguous*. While prior work has shown that this often is not given through automated network traffic analysis, no research has systematically studied how consent notices are currently implemented and whether they conform to GDPR in mobile apps.

To close this research gap, we perform the first large-scale study into consent notices for third-party tracking in Android apps to understand the current practices and the current state of GDPR's consent violations. Specifically, we propose a mostly automated and scalable approach to identify the currently implemented consent notices and apply it to a set of 239,381 Android apps. As a result, we recognize four widely implemented mechanisms to interact with the consent user interfaces from 13,082 apps. We then develop a tool that automatically detects users' personal data sent out to the Internet with different consent conditions based on the identified mechanisms. Doing so, we find 30,160 apps do not even attempt to implement consent notices for sharing users' personal data with third-party data controllers, which mandate explicit consent under GDPR. In contrast, out of 13,082 apps implemented consent notices, we identify 2,688 (20.54%) apps violate at least one of the GDPR consent requirements, such as trying to deceive users into accepting all data sharing or even continuously transmitting data when users have explicitly opted out. To allow developers to address the problems, we send emails to notify affected developers and gather insights from their responses. Our study shows the urgent need for more transparent processing of personal data and supporting developers in this endeavor to comply with legislation, ensuring users can make free and informed choices regarding their data.

CCS CONCEPTS

• **General and reference** → **Measurement**; • **Security and privacy** → **Privacy protections**; *Usability in security and privacy*.

KEYWORDS

Android Security; Consent; GDPR; User Privacy

ACM Reference Format:
Trung Tin Nguyen, Michael Backes, and Ben Stock. 2022. Freely Given Consent? Studying Consent Notice of Third-Party Tracking and Its Violations of GDPR in Android Apps. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22), November 7–11, 2022, Los Angeles, CA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3548606.3560564

CCS '22, November 7–11, 2022, Los Angeles, CA, USA

ACM ISBN 978-1-4503-9450-5/22/11…$15.00
https://doi.org/10.1145/3548606.3560564

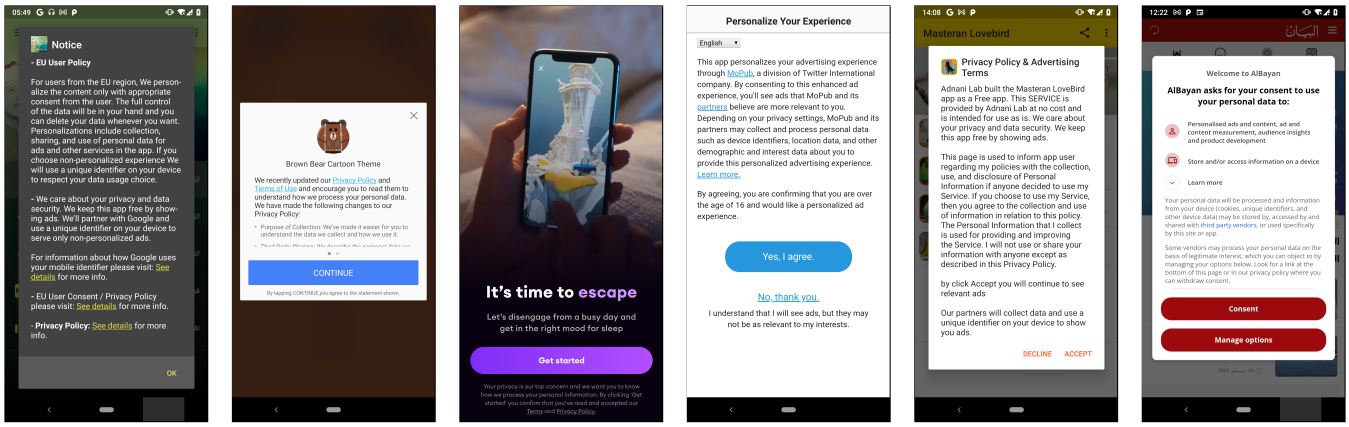
1 INTRODUCTION

Every time we load a page on a commercial website or use a mobile app, information about us and about what we are doing online will be broadcast to large numbers of companies, most notably for advertising purposes [21, 33]. This user tracking happens hundreds of billions of times every day, which is becoming a major problem for individuals' rights regarding their data [21, 41, 43, 44].

To protect user privacy, regulatory bodies around the globe have attempted to address the user tracking problem through regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) — which mandate online services to *transparently* disclose how they handle personal data and grant users crucial data protection rights [11, 55]. As a result of the GDPR legislation developers are, for example, adding or changing privacy policy, having additional consent notice popups, or removing third-party libraries [73].

Under the GDPR, to be legally compliant, an app is required to obtain users' consent before sharing any personal data with third parties if such parties use the data for their own purpose [16, 55]. As such, until now, there have been various ways such consent notices could be obtained in mobile apps in the European Union (see Figure 1). However, if consent is the legal basis for processing data, the GDPR requires the consent must be *freely given, specific, informed, and unambiguous*. Further, the users must have given consent through a statement or by a clear *affirmative* action prior to the data processing [23, 25]. Apps must therefore provide easy ways of giving or refusing consent to collecting and processing users' personal data. Note that if the only option to use the app is to agree to personal data sharing as described in the privacy policies and otherwise, users have to uninstall the app, this cannot be considered *"freely given"* under GDPR.

However, little research has been done to systematically study the violations of GDPR consent requirements in mobile apps. Most researchers focused on analyzing the app privacy policies to identify unexpected behaviors or privacy violations, i.e., comparing an app's actual behavior and the declared data processing in the privacy



(a) (b) (c) (d) (e) (f)

Figure 1: Example of various types of consent notices in Android apps.

policy [2, 57, 59, 77, 78]. Recently, Nguyen et al. [43] performed the first large-scale measurement in Android apps to detect apps sending personal data to third parties by analyzing the network traffic generated by apps *without* prior consent. Although researchers have started to look into GDPR consent violations in the mobile ecosystem and their impact on the users, but the analysis has been done extensively on app network traffic to understand the potential violations. However, no study has systematically analyzed the current practices of implemented consent notices in mobile apps or, more importantly, whether these consent notices can be legally justified under GDPR. Our work is the first to close this research gap by answering the following research questions:

- *Do mobile apps implement any form of consent notices? What are the common properties of such consent notices?*
- *Do these implemented consent notices can be legally justified under GDPR?*
- *Are developers aware of the GDPR consent requirements and the violations of their implementation?*

To answer these research questions, we perform a large-scale study with 239,381 Android apps available through an EEA country Play Store, allowing us to provide a comprehensive overview of the consent notices currently implemented in mobile apps in the wild. Specifically, we propose a mostly automated and scalable approach using image processing and natural language processing techniques to identify the implemented consent notices and their current practices. As a result, we recognize four widely implemented mechanisms to interact with the consent user interfaces from 13,082 apps, such as confirmation-only notices that feature a button with the text "OK" or "I agree" (e.g., (a), (b) and (c) in Figure 1), or notices that provide options to either accept or decline the data sharing (e.g., (d) and (e) in Figure 1). Based on the identified mechanisms, we then extend prior work to develop a tool that automatically detects users' personal data sent out to the Internet with different consent conditions (i.e., based on the choice mechanism to interact with the consent notice).

Doing so, we find 30,160 apps do not even attempt to implement consent notices for sharing users' personal data with third-party data controllers, which mandate explicit consent under GDPR. In contrast, out of 13,082 apps that implement consent notices, we find 2,688 (20.54%) apps violate at least one of the GDPR consent requirements, such as trying to deceive users into accepting all data sharing, sharing before explicitly given consent, or even continuously transmitting data when users have explicitly opted out. Further, to inform app developers about their implementation problems and understand the reasons behind them, we send emails to inform 1,127 affected developers (who have implemented any form of consent notices) and gather insights from their responses. Based on the insights from both developers and our own analysis, we show the urgent need for more transparent processing of users' personal data and further supporting developers in this endeavor to comply with strict law standards, ensuring users can make free and informed choices regarding their data.

In summary, our paper makes the following contributions:

- We systematically study the current practices of consent notices implemented in Android apps in the wild. In particular, we use image processing and natural language processing techniques to analyze the apps' user interfaces and construct a large dataset of 13,082 consent user interfaces, which are categorized into four interaction mechanisms. We believe our results can inform future research on the current practices of consent notices in Android apps and enable them to build tools that help developers comply with legislation such as obtaining valid consent under GDPR.
- We further build a tool that automatically detects users' personal data sent out to the Internet with different consent conditions. We then apply it to perform a first large-scale measurement on the 239,381 Android apps in the wild to understand the current state of the potential violation of GDPR consent requirements[1]. While prior studies primarily

[1]We note that we refer to the violations as potential because we carefully worded not to make legally conclusive statements since this could amount to legal consulting strictly regulated by our national law. Therefore, only a judicial ruling can provide

focus on network traffic analysis, we are the first to analyze the app's consent user interfaces to evidence the potential violations of GDPR.
- To enable developers to address the problem before other parties might take any legal actions, we further send emails to inform affected developers and gather insights from their responses. Finally, based on our results, we make an urgent call for more transparent processing of users' personal data and better tools to support developers.

Organization. The remainder of the paper is structured as follows. Section 2 describes the legal background of GDPR consent requirements and our analysis of potential violations of these requirements in practice. Section 3 presents our approach to identifying consent notices currently implemented in Android apps. Section 4 presents our large-scale analysis of Android apps and demonstrates our approaches to detecting potential GDPR consent violations. Section 5 presents our email notifications to affected developers. Section 6 discusses the scope of our work. Section 7 describes related work, and Section 8 draws conclusions.

2 LEGAL BACKGROUND OF GDPR CONSENT

This section describes the legal background of the GDPR consent requirements. We then briefly outline our legal analysis of the potential GDPR consent violations in Android apps based on many authoritative legal documents in this specific domain. Finally, we note that the GDPR governs all collecting and processing of personal data related to individuals situated in the EU and EEA, which is used for our legal analysis in this work. Besides GDPR, the ePrivacy Directive [49] also applies to how third parties gather consent to accessing information stored on the consumers' device (known as "cookie law" on the Web), but this is beyond our scope.

2.1 Legal Background

On May 25, 2018, the European Union's GDPR mandates a legal justification for the processing of personal data of all Europen residents (**data subjects**) [55]. For example, in mobile apps, developers act as **first-party data controllers** by directly determining the means and purposes for collecting and processing users' personal data, e.g., providing users with the app's functionalities and services. While the third parties (parties external to this app developer) are considered **third-party data controllers** if they receive and use the data for their own purposes and gains (which are not controlled by the first party). As examples, this can be done in order to conduct market analyses, create and monetize user profiles across customers for advertising purposes, or improve their services.

GDPR Article 6 [24] contains the six general justifications for collecting and processing users' personal data. In particular, the processing shall be lawful only if and to the extent that at least one of the following applies: the processing may be based on **consent**, the fulfillment of a contract, compliance with a legal obligation, protection the vital interests of the data subject, the fulfillment of public interest, or the data controller's legitimate interests. In practice, most advertising companies rely on consent or legitimate interests as the legal basis for processing users' personal data for profiling and targeted advertising (i.e., since the legal ground necessary for the performance of a contract does not apply in these circumstances [8, 21]). However, the European Data Protection Board (EDPB) and many legal studies state that it seems unlikely these companies' legitimate interests may claim to outweigh the fundamental rights and freedoms of the data subject and the legitimate interests grounding is not considered to be an appropriate lawful basis for the processing of personal data [16, 21, 22, 43]. Consequently, such companies have to rely on consent as the legal basis for their processing operations. In case the processing is based on consent, the GDPR requires consent to be *freely given, specific, informed, and unambiguous* (GDPR Art. 7 [25]). Further, the data subject (which is the user) must have given consent through a statement or by a clear affirmative action (GDPR Art. 4(11) [23]).

Our research focuses explicitly on these aspects of GDPR consent requirements. In particular, with respect to the regulations mentioned above, transmitting users' personal data to a third-party data controller without *freely given, specific, informed, and unambiguous* consent for the purpose of targeted advertisement is considered violating GDPR.

2.2 Legal Analysis of Potential GDPR Consent Violations

As a result of in-depth legal analysis, we aim to systematically study the following potential legal violations in Android apps specific to GDPR consent requirements. In addition, we cited expert-generated legal sources to argue whether the declared practices violate the aforementioned regulations.

2.2.1 Lack of Consent Notices. Apps transmit users' personal data with third-party data controllers for advertising purposes without implementing any kind of consent notices.

This practice violates the requirement of GDPR consent which mandates mobile apps to obtain users' explicit consent before sharing users' personal data with third-party data controllers [16]. As such, due to a lack of consent notices, these apps do not have a valid legal basis for processing personal data under GDPR. Therefore, collecting and processing users' personal data without implemented consent notices is violated GDPR. The user has to be explicitly asked to consent to personal data processing for advertising purposes through a statement or by a clear affirmative action, and this consent must not be grouped with, e.g., consent to download the app or consent to access certain APIs on the phone, or "consent" packaged in terms and conditions or privacy policies.

Further, GDPR requires consent to be *informed*. In particular, it must fulfill certain conditions: "*The controller shall take appropriate measures to provide any information [...] relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child [GDPR Art.12(1)]*". Notably, Article 29 Working Party [51] states that providing information to data subjects prior to obtaining their consent is important to enable them to make informed decisions, understand what they agree to, and exercise their right to withdraw their consent — if the controller does not provide accessible information, user control becomes illusory, and consent will be an invalid basis for processing.

legal certainty about whether they are actual violations. However, in Section 2 we reason why they should be considered violations by directing to relevant regulations and legal precedents.

Besides, the consent must be *unambiguous*: "*If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language [GDPR Art. 7(2)]*". Therefore, the user's consent must be easily differentiated from other declarations or even consent to other processing activities.

2.2.2 Sharing Data Before Given Consent. Apps implement consent notices to obtain users' consent for sharing personal data, but the apps transmit data before any given explicit consent by users.

This practice violates the requirement of *explicit consent*. Under GDPR, the data controller should obtain verbal or written confirmation about the specific processing [Recital 32]. Article 29 Working Party states that consent can not be based on an opt-out mechanism (e.g., users have to withdraw their by default opted-in consent by turning off personalized ads through their device settings), as the failure to opt-out is not a clear affirmative action [51]. Instead, the user has to actively give their consent, i.e., by clicking "I agree" on a consent form. Merely continuing to use an app or other passive behavior does not constitute explicit consent.

Lastly, this practice further violates the requirement of *prior consent* which requires that the consent has to be obtained prior to any processing activity of personal data to be considered valid [51]. For example, neither of the consent dialogues in Figure 1 is valid consent under GDPR if the data sharing occurs before the user has explicitly given their consent.

2.2.3 No Way to Opt Out. The consent notice does not offer a way to refuse consent. The most common case is a consent notice that simply informs the users about the app's data share.

This practice violates the requirement of *unambiguous* consent, which specifies that the users must have given consent through a statement or by a clear affirmative action (GDPR Art. 4(11) [23] and Art. 7 [25]). According to Article 29 Working Party [51], as a general rule, the GDPR prescribes that if the data subject has no real choice, feels compelled to consent, or will endure negative consequences if they do not consent, then consent will not be valid. Accordingly, consent will not be considered "free" if the data subject is unable to refuse or withdraw their consent without detriment. Furthermore, any element of inappropriate pressure or influence upon the data subject (which may be manifested in many different ways) which prevents a data subject from exercising their free will, shall be invalid consent.

As shown in Figure 1 (a), Figure 1 (b), and Figure 1 (c) none of the apps provide any meaningful ways of giving or refusing consent to the sharing of personal data. The only option to use the app is to agree to share personal data as described in the consent dialogues or packaged in the privacy policies. Since users cannot use the app without consenting to these purposes (has to choose between giving consent or uninstalling the app), the consent cannot be considered as being "*freely given*".

2.2.4 Non-Respect of Choice. Apps transmit users' personal data with third-party data controllers for advertising purposes after users explicitly rejecting/opting-out consent.

The legal analysis of Matte et al. [41] shows that this practice violates the lawfulness principle established in Articles 5(1)(a) and 6(1) of the GDPR: for the processing to be lawful, it must be based on a legal ground. The EDPB further specifies that *"if the individual decided against consenting, any data processing that had already taken place would be unlawful"* due to lacking legal basis for processing. For example, the apps in Figure 1 (d) or (e) will be violated GDPR legislation, if data sharing still occurs after the user has explicitly rejected the data sharing.

3 METHODOLOGY

Our goal is to systematically study the current practices of consent notices implemented in the mobile ecosystem and then examine whether they conform to GDPR. Although identifying cookie consents (i.e., ePrivacy Directive) on the Web has been studied in-depth, little research has been done to systematically study GDPR consent implementations in the Android apps. Recently, Kollnig et al. [36] first attempted to study the absence of consent notices to third-party tracking from a small set of Android apps (i.e., 1,297 apps) by manually inspecting each app's user interfaces. However, such a manual process is insufficient in identifying and studying the status quo of currently implemented consent notices and the potential GDPR consent violations in Android apps at scale. Furthermore, consent notices currently found vary in their appearances (e.g., the consent dialogues display the same notices for sharing data may look very different in many apps) and the underlying functionality in mobile apps (see Figure 1). For example, the consent notices in Figure 1 (a), (b), and (c) only display a notice that informs users about the collecting and sharing of users' data without further information, and there are no ways to reject the data sharing from these consent notice user interfaces; the consent dialog in Figure 1 (f) shows other types of consent mechanisms from third-party providers that offer complex opt-in choices (e.g., consent management platforms).

More specifically, we first propose a mostly automated and scalable solution to identify currently implemented consent notices in Android apps in the wild. We use real Android devices to run each app (without interactions with the app user interface) and take the app screenshots (apps' user interfaces). The underlying assumption is that the app has to show the consent notices before sharing data to be legally compliant, which is the first time users open the app. From the collected app screenshots, we then perform the image processing and natural language processing techniques to identify the privacy-related user interfaces, which potentially are consent notices for sharing users' personal data with third parties (Section 3.1). Then we conduct the clustering analysis on these privacy-related user interfaces to group them by their similarity. Finally, based on the clustering results, we carefully manually verify each group to identify any form of consent notices (Section 3.2).

In the following, we outline how we conduct each of the steps in more detail.

3.1 Collecting Privacy-Related User Interface

We aim to cover all different forms of consent notice user interfaces currently implemented in the Android apps in the wild such as self-implemented consent notices by app developers, consent management platforms (CMP), or consent SDKs from third-party

providers. Therefore, it would be insufficient if we based on any specific consent notice designs or user interfaces to identify the currently implemented consent notices. In fact, most of the consent notices contain a certain keyword that is related to privacy, such as *"privacy", "privacy policy", "data policy", or "gdpr"* [14, 68]. However, not all apps' user interfaces with such keywords are supposed to be consent notices (e.g., app developers simply added the "privacy policy" text that navigates the user to the privacy policy page, which is not a consent notice). Therefore, our first step aims to collect such privacy-related user interfaces, which potentially contain consent notices for sharing the users' personal data with third-party services.

Specifically, to collect privacy-related user interfaces in Android apps, we first install the app in question and then open it but do not interact. We then take the app screenshot (i.e., an app's user interface is everything the user can see and interact with) after waiting for five seconds for the app to be fully initialized. The underlying assumption is that the app should show the consent notices before sharing data, which is the first time users open the app. To do that, we reply on *six* rooted devices (Pixel[2], Pixel 3a, and Pixel 6) running Android 9 or 12 to analyze a given app. Recall that our goal is to systematically study the current practices of consent notices in the wild at scale. Hence, relying on static analysis techniques, which may produce a vast amount of false positives and even could not understand what the actual user interface of such consent notices are, is not an option [10, 38, 70].

Finally, we extract the text from the collected apps' screenshots by using optical character recognition (OCR) [65]. We then perform the string-matching with the privacy-related keyword list from the prior work Degeling et al. [14] (i.e., which contained phrases from all 24 official languages, plus four EU languages) to identify whether the screenshots are privacy-related user interfaces.

3.2 Identifying Consent Notices

After collecting the privacy-related user interfaces, we now want to detect if they are any form of privacy notices or consents for sharing data with third-party services. However, not all privacy-related user interfaces are consent notices (e.g., the app presents only the "privacy policy" text on the screen). Moreover, it is practically impossible to manually analyze all collected privacy-related user interfaces to identify the consent notices. Therefore, we apply a mostly automated and scalable approach using natural language processing techniques to identify the consent notices at scale. Generally, we first perform the clustering technique based on the extracted texts from these privacy-related user interfaces to group them by their visual representation and content. Based on these groups, we then carefully manually inspect each group to identify any form of privacy notices or consents and systematically study the current practices of such consent notices. This approach allows us to verify a large number of apps in manageable ways, and allows us to provide a comprehensive overview of the consent notices currently implemented on mobile apps at scale.

More specifically, we apply several text processing methods to analyze extracted natural language text from collected privacy-related user interfaces. These methods are broadly classified into three primary tasks: text preprocessing, feature extraction, and clustering. In the following, we outline how we conduct each step in more detail.

3.2.1 Text Preprocessing. We first apply the following widely-used text preprocessing techniques [32, 40, 44, 72]: correcting misspelling from ORC errors through *autocorrect* [18] (*e.g., "imagec"* to *"image"*); normalizing and lemmatizing all words, *e.g.*, removing punctuations, converting letters to lowercase and reducing the inflectional forms of a word (*e.g., "sent"*, and *"sending"* to *"send"*); and removing generic stop words such as *"are"* and *"the"*; lastly we remove words that are not generic stop words but are specific to the app such as app name, package name, number, and date time.

3.2.2 Feature Extraction. We then use a bag-of-words model to extract features from the preprocessed texts of privacy-related user interfaces [44, 60]. In particular, let $T = \{t_1, t_2, ..., t_n\}$ be a set of all unique terms in the corpus of privacy-related user interfaces. A text feature of vector of the i^{th} privacy-related user interface is denoted as $t_i = \{t_1, t_2, ..., t_k\}$. For example, the raw text is *"We use device identifiers to personalise content and ads"*, after applying the preprocessing, the generated preliminary text vector is: *t = {"device", "identifier", "use", "ad", "personalise", "content"}*.

Additionally, we employ the hypernym strategy to resolve synonyms and introduce more general concepts for identifying related topics [32] (i.e., added to each term of the feature vectors all subconcepts of the five levels below it based on Wordnet corpus [42]). Finally, we perform word stemming on all terms (*e.g., "personalise"* and *"personalising"* to *"personalis"*). For instance, with the text vector {"*personalise*"}, the final text feature vector will be {"*personalis*", "*person*", "*individu*", "*individualis*"}.

Finally, we convert each term, in the text vector to numeric one by using the term-frequency inverse document-frequency (TF-IDF). The TF-IDF value for each element is calculated as:

$$tfidf(t,d) = tf(t,d) * idf(t) = \frac{1+N}{1+df(d,t)}$$

where t refers to the selected term, d refers to the text vector, tf is the absolute frequency of a term, *i.e.*, $tf(t,d)$ is the number of times a term t occurs in a given d, idf is the term's inverse document frequency, N is the number of text lists in the corpus, and $df(d,t)$ returns the number of text lists that contain the target term t.

3.2.3 Hierarchical Clustering Analysis. Lastly, we use agglomerative hierarchical clustering to identify similar privacy-related user interfaces. Specifically, we use Ward's method [71], and the similarity score or distance between two vectors is calculated by cosine similarity:

$$similarity(\vec{v_i}, \vec{v_j}) = cos(\vec{v_i}, \vec{v_j}) = \frac{\vec{v_i}.\vec{v_j}}{\|\vec{v_i}\|.\|\vec{v_j}\|}$$

3.2.4 Manually Identify Consent Notices. Finally, this leaves us with groups of privacy-related user interfaces for identifying consent notices. We now manually inspect each group and classify any form of information about data practices as a privacy notice and any affirmative user agreement and action to data practices

[2]The first generation of Pixel smartphones.

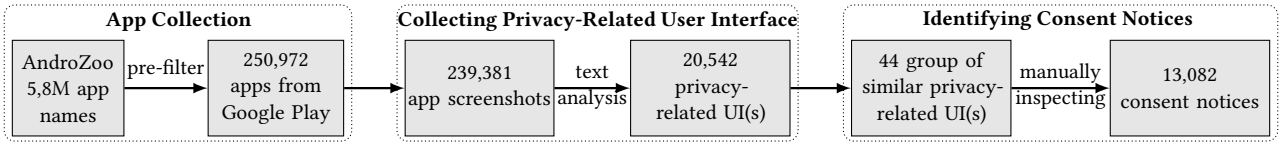


Figure 2: Overview of the methodology to identify consent notices in Android apps and the (intermediate) results.

as consent (i.e., based on the current practice in the field [36]). In fact, the GDPR consent requirements are far more specific, and strict standards must be adhered to. However, this is an intended option to increase the objectivity of our classification to identify the implemented consent notices. Further, we argue that it does not affect the validity of our results regarding the identified potential GDPR consent violations in Section 4.

4 LARGE-SCALE ANALYSIS

This section presents the results of our empirical study of Android apps on Google Play regarding the currently implemented consent notices and whether they are obtained legally under GDPR requirements. More specifically, we first outline how we construct the app dataset for our analysis (Section 4.1) and subsequently present our consent analysis results (Section 4.2). We note that we performed all technical testing and experiment in EEA country where the GDPR applies, i.e., our geolocation is EEA country and the Play store is set to the EEA country variant accordingly. Based on the identified consent notices, to empirically study the potential violations of GDPR consent requirements (outlined in Section 2), we further perform a network traffic analysis to detect apps actually send users' personal data to third-party data controllers for advertising purposes (see Section 4.3). This allows us to ensure the potential violations indeed took place. Finally, we report the observed potential violations in Section 4.4.

4.1 App Dataset Construction

Our analysis aims to assess the state of potential GDPR violations in Android apps in the wild. Therefore, we crawled all free Android apps from October 2021 to March 2022 on the Google Play store (EEA country location-based) based on the list of apps from AndroZoo [1] (which has 5,8M of Android apps' names). As such, we cannot generalize our findings to paid apps by limiting our analysis to free apps. However, this is also in line with previous large-scale Android security research [19, 43, 47, 64]. We further applied the following filter to get the most relevant apps to our study (e.g., filtered out apps that developers do not maintain; focusing on apps that have the capacity to access users' personal data, such as persistent unique identifiers on users' devices). Notably, we consider those apps that meet the following conditions:

- Apps have at least 10,000 downloads (i.e., the popularity of the apps). This factor allows us to regard our findings to represent widespread practices of the potential GDPR consent violations and their effects on millions of users.
- Apps request sensitive permission such as GPS location, contact [28]. These apps have the capacity to access and share highly sensitive information (which are considered personal data under GDPR) to third-party services.
- Apps have the latest update later than May 2018, when GDPR went into effect, i.e., app developers have not maintained the outdated apps, and most of them violated the legislation [43]. Therefore, our study does not include those apps with the latest update before May 2018.

As a result, we obtained 250,972 Android apps. In the next step, we present how to identify the consent notices on these apps.

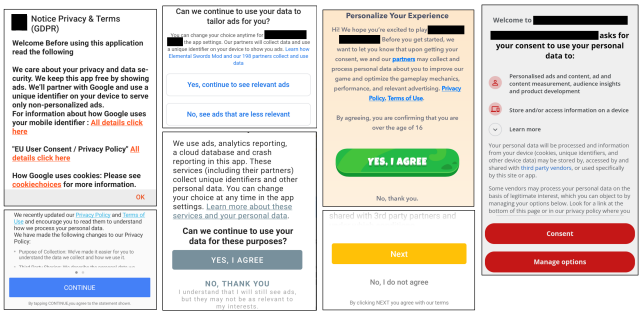
4.2 Identifying Consent Notices In The Wild

Figure 2 shows an overview of our methodology for identifying consent notices in 250,972 Android apps and the intermediate results. More specifically, our technique suffers from certain limitations which keep us from analyzing all apps in the dataset. In particular, we successfully analyzed about 239,381 (95.38% of 250,972) apps by using dynamic analysis (i.e., install the app, open the app, take a screenshot). Unfortunately, the remaining 11,591 either crashed or detected the analysis environment, which potentially affects the completeness of our results.

Out of the 239,381 successfully analyzed apps, we identified 20,542 privacy-related user interfaces using the keyword matching techniques. Furthermore, based on these 20,542 collected privacy-related user interfaces, we identified 44 groups of similar user interfaces (see Appendix A for more details on how to select the best threshold for the clustering analysis and to examine the cluster quality). From 44 groups, we then manually inspected each group to filter out those without consent notices (e.g., UI has only the "privacy policy" text). Finally, for the remaining 36 groups, we manually inspected each UI to identify consent notices (e.g., any form of information about data practices, any affirmative user agreement). As a result, we identified 13,082 privacy-related user interfaces that are any form of consent notices.

Recall that our goal is first to understand the current practices of implemented consent notices in the Android app market (which no study has systematically analyzed before) and then to examine whether they conform to GDPR. Therefore, we not only reported whether an app displayed a consent notice (based on 36 groups), but also analyzed and categorized the types of consent notices based on their interaction options. In particular, by manually inspecting these screenshots of identified consent notices, we identified the four mechanisms for user interaction that are currently widely implemented by Android apps[3] — the classification is based on the definition of prior work on the Web area [14]. Figure 3 shows some example of each mechanism. Based on the choice mechanism to

[3]We note that the clustering of similar privacy-related user interfaces is based on the consent notices' full content, i.e., including the explanation text and the consent choices. As such, when we manually categorized 36 groups of consent notices based on their interaction options, the same consent interaction mechanism may contain multiple groups (e.g., the things that make it different are the text of the choices, such as the choices of the confirmation-only group could be "Start", "Agree", "Next").



(a) Confirmation (b) Personalized ads (c) Binary choices (d) Complex Choices

Figure 3: Example of the four types of consent choices.

interact with the consent notice, we then built an automatic tool to empirically study the potential violations of GDPR (Section 4.3).

- **Confirmation-only**: 5,740 (43,87%) consent notices display a UI element (e.g., a button, checkbox) with an affirmative text such as "OK", "I agree", "Start", clicking on which is interpreted as an expression of user consent.
- **Opt-out personalized ads**: 3,949 (30,19%) consent notices implement a simple choice to refuse consent limit to personalized advertising only, but this does not necessarily prevent the tracking.
- **Binary choices**: 2,871 (21,95%) consent notices have two or more UI elements, but mainly offer two main functionalities which allow users to either accept or decline the data sharing on the app.
- **Complex choices**: 522 (4%) consent notices provide complex opt-in choices such as consent management platform (CMP), consent SDKs from third-party services.

Overall, our classification results provide a comprehensive understanding of the kinds of consent notices in the current Android app market. First, as our data indicate, the confirmation-only consent notices are widely implemented in Android apps (i.e., 43,87%), in which the users have no ways to reject the data sharing. The only option is to agree with data sharing as described in the consent notices, privacy policies, or terms and conditions. If not, users have to uninstall the app, such a practice which cannot be considered *"freely given"* under GDPR. Secondly, 30,19% of apps implemented the opt-out personalized ads consent notices. However, such a choice might make users incorrectly assume that refusing to see personalized ads prevents all tracking [36]. Notably, there are not so many apps that provide users with reject entirely the data sharing (i.e., 21,95% of binary notices, and 4% of the complex choices, which also not easy to reject all of the data sharing at one time).

GDPR Article 6 contains six general justifications for collecting and processing users' personal data. As such, even though those 13,082 apps implemented any form of consent notices, they may be based on other legal grounds for processing users' personal data. Thus they do not legally require consent under GDPR. In order to justify whether those consent notices are legally compliant, only legal experts and authorities can make the decision, and we exclude such discussions from this work. Instead, we specifically focus on those apps that implement consent notices and send users' personal data to third-party advertising data controllers, which mandate users' consent under GDPR. Doing so, we ensure that a potential violation indeed took place and avoid the case where the apps relied on using legitimate interests as a legal basis for processing users' personal data. Hence, in the following, we further perform a network traffic analysis to detect apps that send users' data to third-party data controllers for advertising purposes.

Limitations. We naturally suffer from certain limitations from OCR and dynamic analysis, in which the collected privacy-related user interfaces and the identified consent notices may be incomplete by our approach. First, the text content may be missed when extracting from app screenshot to text by using OCR, or the taken screenshot could be only a small part of the consent notices.

However, relying on static analysis techniques, which are well known for producing unsound results, is not an option [10, 38, 70]. On top of that, with the static analysis, we could not see the actual appearance of such privacy-related user interfaces, which is necessary for our analysis to identify the consent notices and their current practices. To demonstrate the insufficient of static analysis, we first randomly sampled 1,000 apps from those 20,542 apps that have privacy-related user interfaces (see Section 4.2). We then performed a static analysis to extract the text from the *strings.xml* resource of these apps (where the apps store UI text), and then apply the same keyword matching to the extracted text (i.e., identifying privacy-related user interfaces). As a result, the static analysis approach only identified 345 (34.5% of 1,000) apps that have privacy-related text. We found that the static analysis has missed the majority (65.5%) of privacy-related user interfaces.

Therefore, we argue that the static analysis is insufficient to identify privacy-related user interfaces (potentially containing consent notices). However, similar to any other static analysis, our approach also has drawbacks. In particular, we statically analyzed the *strings.xml* resource to identify the privacy-related text (which could potentially be used by the consent notice user interfaces). In practice, the app may dynamically load consent notices from the Web API, consent CMP(s), third-party SDK(s), or dynamic content. Further, developers may add hardcoded text into the UI layout files or the app's code, which could be obfuscated (i.e., require more advanced de-obfuscation tools to analyze [7, 75]). As such, our static analysis could miss such cases, which can only be analyzed by dynamic analysis (i.e., actually running the app).

Further, to estimate the false negatives of our proposed approach, we randomly sampled 100 non-flagged apps (identified as not having any form of consent notices by our approach) and manually checked each app. Among them, we found that four apps crashed, one app had a consent notice, and 95 apps had no consent notices.

Finally, we only consider consent notices written in English, supported by the natural language processing techniques that we used, to ensure that we understand the actions we perform, such as measuring the quality of privacy-related user interface clustering and categorizing the consent notices in Section 3.2. As such, our results have not covered 1,996 (9.72% of 20,542) privacy-related user interfaces that are not in English but instead, e.g., German, Spanish.

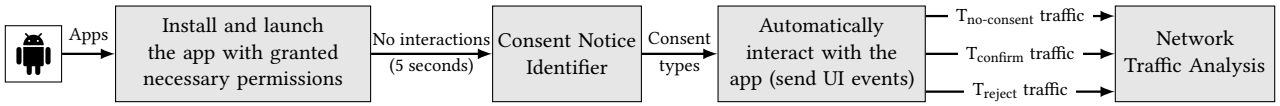


Figure 4: Overview of our methodology to detect data sent out to the Internet with different given consent conditions.

Choices	Text Keyword
Confirm	*"yes", "agree", "ok", "continue", "consent", "confirm", "understand", "start", "okay", "enable", "got it", "go", "join", "next", "understood", "play", "allow"*
Reject	*"no, thank you", "decline", "no, see ads", "no, thanks you", "no, thank", "cancel", "opt out", "disagree", "no", "not", "refuse", "deny", "exit", "cancel", "postpone", "later", "do not", "reject"*

Table 1: Text keywords based on each type of consent choice.

4.3 Automating Potential GDPR Violations Detection

To empirically study the potential violations of GDPR consent requirements in Android apps (outlined in Section 2), we further perform a network traffic analysis to detect apps actually send users' personal data to third-party data controllers for advertising purposes. In particular, we extend prior work to build an automatic dynamic analysis tool that detects apps sending data to the Internet with different given consent conditions. Based on the collected network traffic and the identified consent notices, we then perform a legal analysis of potential violations of GDPR consent requirements (Section 4.4). Generally, we measure the following conditions: (1) we look for apps that send users' personal data to third-party data controllers for advertising purposes to see whether they implement any consent notices; (2) for those that have implemented consent notices, we further investigate whether they send data before any given consent; (3) whether they allow users to refuse the consent; (4) and finally, those apps where we explicitly opt-out from consent and they still share data.

We followed best practices established by prior work [31, 43, 54, 56] to collect network traffic for identifying third-party services and privacy leaks in Android apps. In order to intercept the TLS traffic, our six rooted devices were instrumented with our own root certificate (i.e., by using MitM proxy [12]), and the given app was instrumented to detect and disable SSL Pinning by using *objection* [46]. To interact with the app automatically, we then extended the lightweight test input generator that sends random or scripted input events/UI interactions to the app based on our configurations (i.e., DroidBot [39]). The collected network traffic will be stored in our database for later analysis.

Figure 4 shows an overview of our methodology to detect data sent out to the Internet with different given consent conditions. Generally, we first install the app and then grant all apps' requested permissions listed in the manifest, i.e., install and runtime permissions. Subsequently, we launch the app, run it up to 150 seconds based on different given consent conditions, and record its network traffic. We note that, between each condition, we uninstall the app and clear all of the app stored data to ensure the apps show consent notices each time (i.e., the consent notice may not show again when the users have already made their choices).

In the following, we outline how we analyze the network traffic of 13,082 apps (that implemented consent notices, in Section 4.2) based on each consent condition.

- **No consent**: First, we aim to detect apps' network traffic without users' explicit consent ($T_{no\text{-}consent}$ traffic). To achieve this, we simply open the app but do not interact with it at all. The underlying assumption is that if network traffic occurs when this app is opened without any interactions, we have naturally not consented explicitly to any type of data collection by third parties.
- **Confirm consent**: Second, we detect apps' network traffic after accepting the consent ($T_{confirm}$ traffic). Based on the identified consent notices in Section 4.2, our tool will first automatically identify the current user interface is consent notice or not and then click on the UI element that indicates the acceptance on the consent user interface (see the first row in Table 1) by using the string-matching technique and then afterward randomly navigating the apps.
- **Reject consent**: Third, we also aim to detect apps' network traffic after rejecting the consent (T_{reject} traffic). We now configure our tool to click on the UI element that indicates to reject consent (see the second row in Table 1) and then record its network traffic while running.

Additionally, we want to detect apps that transmit users' personal data with third-party data controllers for advertising purposes without implementing any form of consent notices, potentially violating the outlined *"Lack of Consent Notices"* in Section 2. Therefore, from the 239,381 successfully analyzed apps, we excluded those apps in the set of 13,082 apps, and then automatically analyzed the remaining apps to collect their network traffic.

Finally, we extend the code from Nguyen et al. [43] to search for the users' personal data in the outgoing network streams originating from each device (Table 2 listed types of personal data), and to identify third-party domains operated by ad-related companies (i.e., based on list of 45 advertisement domains of data controllers from their legal analysis).

4.4 Observed Potential Violations

In this section, we present the results of our empirical study of 239,381 Android apps on Google Play regarding the potential violations of GDPR consent requirements.

4.4.1 Overview of Network Traffic Analysis. Out of the 239,381 successfully analyzed apps, we identified 101,484 (42.39% of 239,381) apps that contacted to 20,968 unique fully-qualified domain names by either sending or receiving some data in our experiment. Is it

Data Type	Description
AAID	Android Advertising ID
BSSID	Router MAC addresses of nearby hotspots
Email	Email address of phone owner
GPS	User location
IMEI	Mobile phone equipment ID
IMSI	SIM card ID
MAC	MAC address of WiFi interface
PHONE	Mobile phone's number
SIM_SERIAL	SIM card ID
SERIAL	Phone hardware ID (serial number)
SSID	Router SSIDs of nearby hotspots
GSF ID	Google Services Framework ID

Table 2: Types of personal data we consider in our work.

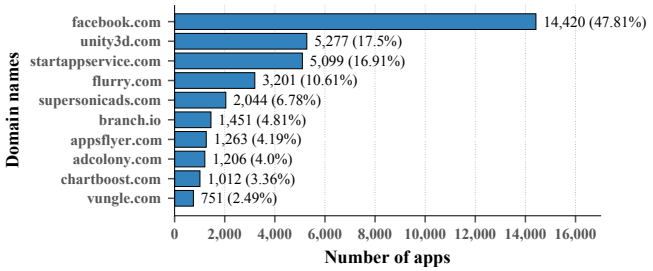


Figure 5: Top 10 ad-domains that received personal data from 30,160 apps that have no consent notices.

known that a single registerable domain may uses many subdomains (e.g., `api2.branch.io`, `api.branch.io`). Therefore, to normalize these hosts to their registerable domain (`branch.io` in the above cases), we rely on the public suffix list [53]. As a result, we identified 14,209 registerable domains (referred to as "domain names") that were contacted by those 101,484 apps. Notably, we identified 41,639 (41,03% of 101,484) apps sent users' personal data to 1,484 domain names.

Those 1,484 domain names that receive personal data may operate with their own privacy policies and further share the data with their partners, which could be broadcast to large numbers of different companies. Therefore, to understand how personal data may be processed, the legal basis for processing, and whether the processing is compliant with GDPR, we have to read the entire privacy policies of all the involved partners of those services. However, it is infeasible to conduct such an in-depth analysis of hundreds of privacy policies. Therefore, in the following potential GDPR consent violations analysis, we primarily focus on the list of 45 domains from [43] that are operated by advertising companies and hence definitely act as data controllers. We leave an automated analysis of privacy policies and assessment of potential data controllers to future work. In that case, our framework would be able to detect more potential violations. By using the conservatively established list from prior work, we are confident to not suffer from any false positive, yet our results naturally only serve as a lower bound of potential violations.

In the following subsections, we will now elaborate on the analysis results of each potential violation outlined in Section 2 in greater detail.

4.4.2 Lack of Consent Notices. We detected 32,341 apps sent users' personal data to 43 of those 45 third-party ad-related domains (data controllers) from [43], which would require explicit consent to receive users' personal data. However, we identified a significant number of 30,160 (93.26% of 32,341) apps have no implemented any form of consent notices. Figure 5 shows the top 10 ad-related domains that received personal data without consent notices in our dataset, counting the number of apps that sent data to them. The potentially violated apps occur across different app categories, such as the top 5 categories that have more potential violating apps than others are GAME (20.51%), ENTERTAINMENT (8.10%), EDUCATION (5.65%), PERSONALIZATION (5.13%), and BOOKS & REFERENCE (4.96%).

Our results show a significant number of apps that potentially violate GDPR consent requirements, specifically due to a lack of legal basis for processing. For this, we provide empirical evidence of a widespread *lack of consent notices* in the Android app market, which is not given through automated network traffic analysis on GDPR violations by prior work [43]. Overall, our data indicate that the vast majority of apps do not even attempt to achieve GDPR compliance. It could well be that developers are unaware of the data sharing and the need to obtain user consent for such data sharing and collection.

4.4.3 Sharing Data Before Given Consent. Out of 13,082 apps that implemented consent notices, we identified 3,007 (23%) apps that sent users' personal data to the Internet before any given consent (in $T_{\text{no-consent}}$). Notably, 2,181 (16,67% of 13,082) apps sent personal data to third-party data controllers for advertising purposes which mandates explicit consent. It could be that the developers bundle the data sharing with consent to use the app (i.e., confirmation-only consents). We found that 42.6% of 2,181 apps implemented the confirmation-only consent type. However, for such cases, in Section 2, we show that personal data transfer must only occur after the user has actively consented (e.g., by clicking accept), such "consent" packaged in terms and conditions or privacy policies are not compliant, which render invalid consent under GDPR.

On the contrary, a large number of 57.4% potential violating apps implemented other consent types (i.e., including 27.92% opt-out of personalized ads, 21.37% binary choices, and 8.12% complex choices). However, even though they provide choices for the users to reject the data sharing, but the apps behave differently compared to the consent notice user interfaces. Weir et al. [73] surveyed app developers and observed that most developers' changes were cosmetic due to the GDPR legislation (e.g., adding dialogues). We confirm and evidence the existence of this widespread problem among app developers and further highlight that such cosmetic changes do not fulfill the legal conditions for collecting valid consent under GDPR. Further, this may be caused by app developers' misconfiguration of third-party libraries. The other likely reason is that third-party consent platforms and services employ the opt-out mechanism rather than opt-in. Their services first transmit data and then ask users to opt-out, or send data along with a flag indicating the opt-out. However, at the very least, there is no technical reason why this

	Sharing Data Before Given Consent (N=2,181)	No Way to Opt-Out (N=1,084)	Non-Respect of Choice (N=134)
Confirmation-only	929 (42.60%)	1,084 (100%)	–
Opt-out of personalized ads	609 (27.92%)	–	48 (35.82%)
Binary choices	466 (21.37%)	–	28 (20.90%)
Complex choices	177 (8.12%)	–	58 (43.28%)

Table 3: Types of consent notices and number of potentially violated apps (percentages relative to each violation type).

would even be necessary, and this opt-out mechanism is also not compliant with the GDPR consent requirements.

4.4.4 No Way to Opt Out. From 5,740 apps that implemented the confirmation-only consent type, we identified 1,387 of them sent personal data to the Internet. Another 1,084 of them sent to third-party advertising data controllers. For such apps, the only option to use the app is to agree to the sharing of personal data as described in the privacy policies. If the data subject has to choose between giving consent to third-party processing for profiling and behavioural advertising purposes, or uninstalling the app, the consent cannot be considered to be "*freely given*".

4.4.5 Non-Respect of Choice. We found 134 apps that still sent users' personal data to third-party advertising data controller after explicitly opting-out the data sharing from the consent user interface. Also, surprisingly, these types of potential violations can even occur in popular apps with millions of installs. For example, we find an app that has more than 5M installs that sent the AAID along with other persistent identifiers (i.e., IMEI, MAC) to the same third-party advertising data controllers. Other apps with more than 100M installs still shared the AAID with the third-party data controller for advertising purposes after explicitly rejecting the consent.

4.4.6 Summary of Observed Potential Violations. In summary, we perform a large-scale analysis of the potential violations of GDPR consent requirements on a set of 239,381 Android apps in the wild. Doing so, we identified 30,160 apps do not even attempt to implement consent notices for sharing users' personal data with third-party data controllers, which mandate explicit consent under GDPR. In contrast, out of 13,082 apps implemented consent notices, we identified 2,688 (20.54%) apps violate at least one of the GDPR consent requirements (i.e., an app could violate more than one GDPR consent requirement). In particular, 2,181 (16,67% of 13,082) apps sent personal data to third-party data controllers before given explicit consent. Among these 2,181 apps, 42.6% of apps are from the confirmation-only group (see the first column of Table 3). On the other hand, 1,084 (8.28% of 13,082) apps sent to third-party data controllers in which their consent notices do not offer a way to refuse consent. Notably, all of them are from the confirmation-only group (see the second column of Table 3). Further, 134 apps that still sent data after explicitly opting out of the data sharing from the consent user interface. 58 out of 134 apps are from the complex choices group (see the third column of Table 3).

Interestingly, a significant number of potential violations are related to Android's Advertising ID (AAID), i.e., nearly 99% of apps at least sharing this personal data (Table 4 shows the type of personal data that we detected). According to Google, an AAID is *"a unique, user-resettable ID for advertising, provided by Google Play services. ... It enables users to reset their identifier or opt-out of personalized ads"* [26]. Furthermore, even Google's brand Admob explicitly lists the AAID as personal data in their documentation for delivering ads [29]. While Google itself remained vague on the characterization of the AAID as personal data, the IAB Europe GDPR Implementation Working Group already established in their 2017 Working Paper on personal data that "Cookies and other device and online identifiers (IP adresses, IDFA, AAID, etc.) are explicitly called out as examples of personal data under the GDPR" [30]. In May 2020 NOYB – European Center for Digital Rights [45], a European not-for-profit privacy advocacy group, lodged a formal complaint about the AAID with Austria's data protection authority. The complaint states that although the AAID is personal data Google does not adhere to the requirements of valid consent. Android users have no option to deactivate or delete the tracking ID, only to reset it to a new one.

More recently, Google has taken a first action regarding this matter. As part of the Google Play services update[4], the advertising ID will be removed when users opt-out of personalization using the advertising ID in Android Settings (i.e., any attempts to access the identifier will receive a string of zeros instead of the identifier) [27]. However, Article 29 Working Party states that consent can not be based on an opt-out mechanism (e.g., users have to withdraw their by default opted-in consent by turning off personalized ads through their device settings), as the failure to opt-out is not a clear affirmative action [51]. This indicates that none of the controllers who claim that data subjects can withdraw their by default opted-in consent by turning off personalized ads through their device settings, have a valid consent to process personal data [21]. In contrast, Apple has recently taken active action for mandatory prior consent for sharing of Advertising Identifiers for its iOS 14 update [3] explaining that even dynamic advertising identifiers are considered personal data.

Not Easy to Withdraw Consent. GDPR Article 7(3) specifies that the controllers must ensure the data subject can withdraw their consent as easily as giving consent and at any given time. Notably, Article 29 Working Party [51] states that data subjects must be able to withdraw consent via the same interface, as switching to another interface for the sole reason of withdrawing the consent would require undue effort. Furthermore, the controller must make the withdrawal of consent possible free of charge or without lowering

[4]The Google Play services' updates will affect Android 12 starting in late 2021 and then will expand to affect apps running on all devices starting April 1, 2022

	Lack of Consent Notices (N= empirical30,160)	Sharing Data Before Given Consent (N=2,181)	No Way to Opt Out (N=1,084)	Non-Respect of Choice (N=134)
AAID	29,952 (99.31%)	2,166 (99.3%)	1,077 (99.4%)	134 (100%)
BSSID	30 (0.1%)	—	—	—
EMAIL	1 (0.0%)	—	—	—
GPS	524 (1.74%)	23 (1.1%)	13 (1.2%)	2 (1.5%)
IMEI	336 (1.11%)	23 (1.1%)	10 (0.9%)	3 (2.2%)
MAC	214 (0.71%)	39 (1.8%)	27 (2.5%)	3 (2.2%)
SERIAL	3 (0.0%)	—	—	—
SSID	31 (0.1%)	—	—	—

Table 4: Types of data and number of apps sending this data to ad-related domains (percentages relative to each violation type).

service levels [50]. To investigate whether developers provide the easy options to withdraw consent, we randomly sampled 100 apps that potentially violated at least one of the consent requirements and checked the app functionality. Specifically, we checked for options to allow withdrawal consent in the app settings. Among these 100 apps, we found only 16 apps present any options to withdraw consent notices. Overall, this indicates that most apps may not provide meaningful ways to withdraw consent, which is also necessary for valid consent under GDPR. However, further studies need to be conducted to confirm this problem with larger samples. Finally, we note that finding privacy-related app settings is challenging due to the difficulty in locating them from an app's user interface. The more challenging is to automatically detect those related to consent. We leave this challenge for future work.

5 DEVELOPER NOTIFICATION

To enable developers to address the incorrect consent implementations, we notified affected developers, focusing mainly on the potentially violated apps that have implemented consent notices. On the one hand, this enables them to address the issues before other parties might take any legal actions (e.g., being fined for breaching data protection law [13, 62]). Second, we wanted to gain insights into the underlying reasons that caused the observed phenomena in the first place. In addition, we informed all notified developers about the study purpose, our methodology, and contact information (i.e., email address, phone number) to contact in case they had questions or concerns. We note that our institution's ethics guidelines do not mandate approval for such a study.

Based on the developers' detailed contact information in the Play Store, we extracted the publicly available email addresses to send the notifications[5]. Similar to the work of [43], to access how many developers received our reports, rather than including the technical details in the email, we further sent developers a link to our Web interface. Specifically, in our Web report, we briefly explained our testing methodology, showed the developers information about potential violations, accompanied by the corresponding legal references (i.e., GDPR consent requirements), and detailed which hosts received which type of data. Further, to gain some insights into the underlying reasons that caused the identified problems, we asked participants if they had been aware of the potential violations of their apps, their general understanding of the personal data that their app sent to third-party services, and their understanding of GDPR consent requirements as well as proposals for tool support. We decided to have this rather than a full-fledged study, as we wanted to keep the overhead for respondents as low as possible to prompt more responses.

[5]Email template: https://raw.githubusercontent.com/cispa/gdpr-consent/main/email-template.md

5.1 Notification and Accessed Reports

The potentially violated apps may have been updated (i.e., changed the problematic code, removed from the Play store) between our download and notification date. Thus, we further checked their availability and last update time before sending our notifications. Doing so, out of the 2,688 apps that implemented consent notices and potentially violated at least one of the GDPR consent requirements, we find 829 apps had been removed or updated with a newer version by the time we conducted our notification on April 04, 2022. We took this step to ensure that we would not notify developers who had removed the problematic code between our dataset download and notification date. Also, a single developer may have responsibility for more than one app in the store. Therefore, to ensure we do not send multiple emails to developers, we grouped emails to developers to receive only one email with multiple report links. We followed currently best practices established by existing work [15, 43, 61] allowing developers to opt-out of our study.

In total, we notified 1,127 developers responsible for the remaining 1,859 potentially violated apps. Of those developers, only one asked to be removed from our experiment and do not wish to be involved in further study. Until April 26, 2022, we saw 505 accessed reports for 225 apps. Notably, considering that a single developer may have multiple apps affected by the same issue, we count the overall number of apps for which their developer accessed *some* report; totaling 266 (14.31% of 1,859) apps for which we reached their developer.

5.2 Developer Responses

In addition to the accessed reports and the updated apps, we also analyzed the responses we received from developers to understand the underlying reasons that caused the problems. In total, this amounted to 43 distinct developers for which we classified emails. Note that not all respondents answered the stated questions from

our email notification[6]. The response rate for our questionnaires was low, as might be expected. However, it is in line with prior work [43] which received 448 responses from 11,914 notifications.

Of the 43 respondents, 25 acknowledged receipt of our email and wanted to take it under advisement. Five stated that they required further investigation within their respective companies. Two have mentioned they updated the apps and further inquired with us to recheck their apps. Notably, two respondents argued that the EU was not their primary market and inquired us about potential solutions to the problems. We faithfully answered all of the emails while stating that we cannot provide conclusive individual legal assessments. On the other side, three respondents disagreed with our assessment.

When asked about the data collection, 9/14 respondents said they were not aware of the types of data being collected, and 5/14 said they knew GDPR protected this data. Of the 13 respondents who answered our question regarding being aware of GDPR consent requirements, 9 said yes and passed all the flags to the SDK, and four are not aware. Three explained their app was outdated, and four said this was a bug.

Regarding our final question about developer support, we received seven answers. Of those, six wanted to have an automated tool like our to analyze their apps for compliance, while two asked for better documentation around how to implement GDPR compliance. Finally, three respondents argued that third-party tools should be compliant by default, e.g., *"a "one-stop solution" as Unity3D plugin that ensures to fully cover all current (and future) GDPR requirements"*.

5.3 Updates to Notified Apps

To assess our notification's impact on the affected apps, we downloaded new versions of all apps that had looked at our reports at least one by April 26, 2022. Then, we re-ran our pipeline for each app with an updated version to assess if the changes were related to the reported GDPR infringement. For the 266 apps for which we reached a developer, 8 apps were removed from Google Play, 111 apps were no longer available to download from EEA country, and 147 apps have been updated. Of those 147 apps, 84 still potentially violated at least one of the GDPR consent requirements, leaving the remaining 63 apps which fixed the problems.

We note that the overall number of apps that addressed the issue is low, as might be expected from sending unsolicited emails to prospective participants. In fact, app developers are more likely to take action when they receive such notifications from their service providers [52], e.g., *"THE only person who can claim anything is my service provider-Play Store"*. Besides, we believe that the seemingly minor change in overall numbers can be attributed to a lack of time to address the issue properly.

6 DISCUSSION

Our results thus far have shown that many apps do not even attempt to implement consent notices for sharing users' personal data with third-party data controllers for advertising purposes, despite the GDPR requirements. Although little research has been working on this, we further show that even from the apps implementing any form of consent notices, there are still many, namely 20.54%, containing at least one potential violation of GDPR. Given these insights, we now discuss further the problems.

6.1 Widespread Violation of GDPR Consent

On the legal side, EU regulators have already been active. Recently, the Norwegian Data Protection Authority (DPA) imposed a fine of $7.17M on Grindr [20] for obtaining invalid consent under GDPR (i.e., users had to agree to the entire privacy policy but not to a specific processing operation, and do not have the choice not to consent). The France DPA (CNIL) fined Google $170M and Facebook $68M for breaching French and GDPR laws (i.e., these tech giants were using manipulative dark patterns to try to force consent) [62]. In late October 2018, the CNIL also made a ruling decision on an advertising company, which suggests that bundling consent to partner processing in a contract is not valid consent under the GDPR [63]. The CNIL stated that controllers have to implement a compliant consent mechanism (i.e., it must be *freely given, specific, informed, and unambiguous*) and ensure that any personal information is collected and processed lawfully. This means that when receiving personal data from a partner company, the receiving party must demonstrate that the transmitting party also relied on legally compliant consent mechanisms [21]. However, our results show a significant skew toward apps sending out personal data to advertisement companies without valid consent under GDPR, i.e., 30,160 apps do not even attempt to implement consent notices, 2,688 apps that implemented a form of consent notice but potentially violate at least one of the GDPR consent requirements. Those apps did not present the user with legally compliant consent mechanisms under GDPR, which has consequences for the validity of consent for any third parties acting as controllers, e.g., those advertising companies that received personal data.

In practice, many third-party services claim that they operate based on consent passed on through contractual terms with their customers, which would be the app developers in this case (e.g., Facebook required developers to obtain appropriate legal basis consent before sending data via their SDK [17], while it by default automatically collects data). However, those third-party data controllers can neither demonstrate valid legal consent nor a legitimate interest that overrides the consumer's fundamental right to privacy for behavioral profiling and targeted advertising [21].

6.2 Transparency of Processing Users' Data

We find that the majority of apps have not asked for informed consent from the users for tracking and profiling third-party services (i.e., 30,160 apps do not even attempt to implement consent notices). Thus, it is practically impossible for users to know which third parties receive and process their data. Moreover, even if users had the time and knowledge to read and understand privacy policies, these documents are excessively complicated and obtuse, and the majority (71%) of apps lack privacy policies even though they are obligated to have one [21, 78]. Therefore, we show the urgent need for more transparent processing of users' personal data and further allowing them to exercise their fundamental rights and freedoms. In mobile apps, users can take only a few actions to limit or prevent

[6] All notified developers were informed that their responses are pseudonymous and could be used in our paper.

tracking and data sharing, while there are various tools to prevent tracking in the Web browsers [9, 41].

6.3 Lack of Support for Developers

Is it known that developers are currently in a disadvantaged position, where third parties make it cumbersome for developers to comply with GDPR [43]. However, as first-party data controllers, developers are legally responsible for ensuring that the users' data is lawfully processed. Based on the received responses, there is a clear need for better information and documentation from third-party services and assurance tools that help developers comply with strict law standards, e.g., *"I would love there to be a standardised implementation requirement for this. It seems every 3rd party SDK we have uses different ways of implementing consent. The documentation can be quite unclear, so having a tool to analyse app traffic would be incredibly useful."*. Therefore, we strongly call on providing developers with clear requirements or guidance for how GDPR consent has to be legally obtained.

7 RELATED WORK

Researchers are actively and continuously studying the legislation violations of online services after GDPR went into effect in May 2018. Among others, existing works have extensively studied the cookie consent compliance on the Web.

In particular, Kampanos et al. [34] shows that the majority of websites in the UK and Greece lack cookie consent notices, i.e., only roughly 45% have a cookie notice. Many studies further have shown that a lot of websites do potentially violate the GDPR consent requirements, such as do not allow users to refuse data collection, installing tracking and profiling cookies before the user gives explicit consent [14, 37, 58, 66, 67, 69]. Regarding the cookie consent interface, Matte et al. [41] perform the first study to compare the interface of the cookie notices shown to the users to the website behaviors. On the other hand, Utz et al. [68] inspected how the design of consent popups from websites nudge users into uninformed consent by conducting a study with real website visitors.

However, little research has been done to measure the GDPR violations of consent on mobile apps. Recently, Nguyen et al. [43] performed the first large-scale measurement on Android apps to detect apps sending personal data to third parties without prior consent. On the other hand, Kollnig et al. [36] tried to study the absence of consent notices to third-party tracking from a small set of Android apps (i.e., 1,297 apps) by manually inspecting each app — which is insufficient in identifying and studying the status quo of currently implemented consent notices in Android apps. Until now, there have been various ways that consent could be obtained in mobile apps. However, while existing works [36, 43] mainly focus on analyzing apps' network traffic to detect GDPR violations, little or no research has systematically studied how the consent notices are currently implemented in mobile apps and whether they are legally obtained under GDPR.

Different from prior work, we perform the first large-scale study into consent notices of third-party tracking in Android apps in the wild to understand the current practices and the current state of GDPR's consent violations. While prior studies primarily focused on network traffic analysis or manually studied a small set of samples, we semi-automatically analyzed a large scale of apps' consent user interfaces to investigate whether they are *freely given, specific, informed, and unambiguous* with respect to GDPR consent requirements. Additionally, another line of work aims to analyze the app privacy policies or privacy labels to identify potential GDPR violations, i.e., determining whether an actual app's behavior is consistent with what is declared in the app privacy policy or privacy label [2, 35, 57, 59, 77, 78]. Researchers have developed different techniques to detect privacy violations in mobile apps and to identify third-party advertising and tracking services using static analysis [4, 5, 44, 48] or dynamic analysis [6, 57, 74, 76].

8 CONCLUSION

In this paper, we performed a systematic study into consent notices of third-party tracking in 239,381 Android apps in the wild to understand the current practices and the current state of GDPR's consent violations. As a result, we first recognized four widely implemented mechanisms to interact with the consent user interfaces from 13,082 apps. We found 30,160 apps do not even attempt to implement consent notices for sharing users' personal data with third-party data controllers, which mandate explicit consent under GDPR. In contrast, out of 13,082 apps implemented consent notices, we identified 2,688 (20.54%) apps potentially violate at least one of the GDPR consent requirements, such as trying to deceive users into accepting all data sharing or even continuously transmitting data when users have explicitly opted out. We sent notification emails to inform affected developers and gather insights from their responses. Our study showed the urgent need for more transparent processing of personal data and supporting developers in this endeavor to comply with legislation, ensuring users can make free and informed choices regarding their data.

ACKNOWLEDGMENTS

We thank the anonymous shepherd and reviewers for their insights and constructive feedback.

REFERENCES

[1] Kevin Allix, Tegawendé F. Bissyandé, Jacques Klein, and Yves Le Traon. 2016. AndroZoo: Collecting Millions of Android Apps for the Research Community. In *MSR*.
[2] Benjamin Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Serge Egelman. 2020. Actions Speak Louder than Words: Entity-Sensitive Privacy Policy and Data Flow Analysis with POLICHECK. In *USENIX Security*.
[3] Apple. 2022. User Privacy and Data Use. https://developer.apple.com/app-store/user-privacy-and-data-use/. 2022/04/29.
[4] Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Yves Le Traon, Damien Octeau, and Patrick McDaniel. 2014. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. *ACM Sigplan Notices*.
[5] Kathy Wain Yee Au, Yi Fan Zhou, Zhen Huang, and David Lie. 2012. Pscout: analyzing the android permission specification. In *CCS*.
[6] Ravi Bhoraskar, Seungyeop Han, Jinseong Jeon, Tanzirul Azim, Shuo Chen, Jaeyeon Jung, Suman Nath, Rui Wang, and David Wetherall. 2014. Brahmastra: Driving apps to test the security of third-party components. In *USENIX Security*.
[7] Benjamin Bichsel, Veselin Raychev, Petar Tsankov, and Martin Vechev. 2016. Statistical deobfuscation of android applications. In *CCS*.
[8] European Data Protection Board. 2019. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects". https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf. 2019/02.

[9] Dino Bollinger, Karel Kubicek, Carlos Cotrini, and David Basin. 2022. Automating Cookie Consent and GDPR Violation Detection. In *USENIX Security*.

[10] Richard Bonett, Kaushal Kafle, Kevin Moran, Adwait Nadkarni, and Denys Poshyvanyk. 2018. Discovering flaws in security-focused static analysis tools for android using systematic mutation. In *USENIX Security*.

[11] CCPA. 2022. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa. 2022/04/29.

[12] Aldo Cortesi, Maximilian Hils, Thomas Kriechbaumer, and contributors. 2010–. mitmproxy: A free and open source interactive HTTPS proxy. https://mitmproxy.org/ [Version 6.0].

[13] Datatilsynet. 2022. Intention to issue € 10 million fine to Grindr LLC. https://www.datatilsynet.no/en/news/2021/intention-to-issue--10-million-fine-to-grindr-llc2/ 2022/04/29.

[14] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. 2019. We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. In *NDSS*.

[15] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide scanning and its security applications. In *USENIX Security*.

[16] europa.eu. 2022/04/29. "Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC (Article 29 Working Party). https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[17] Facebook. 2022. FB SDK Best Practices for GDPR Compliance. https://developers.facebook.com/docs/app-events/gdpr-compliance/. 2022/04/28.

[18] Filyp. 2022. autocorrect. https://github.com/filyp/autocorrect 2022/04/28.

[19] Felix Fischer, Konstantin Böttinger, Huang Xiao, Christian Stransky, Yasemin Acar, Michael Backes, and Sascha Fahl. 2017. Stack overflow considered harmful? the impact of copy&paste on android application security. In *SP*.

[20] NOYB – European Center for Digital Rights. 2022. NCC & noyb GDPR complaint: Grindr fined € 6.3 Mio over illegal data sharing. https://noyb.eu/en/ncc-noyb-gdpr-complaint-grindr-fined-eu-63-mio-over-illegal-data-sharing 2022/04/28.

[21] forbrukerradet.no. 2022/04/29. OUT OF CONTROL. https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf.

[22] GDPR. 2013. Opinion 03/2013 on purpose limitation (WP 203), adopted on 2 April 2013. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf 2022/04/21.

[23] GDPR. 2021. Art. 4 Definitions. https://gdpr.eu/article-4-definitions/ 2021/02/01.

[24] GDPR. 2021. Art. 6 Lawfulness of processing. https://gdpr.eu/article-6-how-to-process-personal-data-legally/ 2021/02/01.

[25] GDPR. 2021. Art. 7 Conditions for consent. https://gdpr.eu/article-7-how-to-get-consent-to-collect-personal-data/ 2021/02/01.

[26] Google. 2021/02/02. Advertising ID. https://support.google.com/googleplay/android-developer/answer/6048248?hl=en.

[27] Google. 2022. Advertising ID. https://support.google.com/googleplay/android-developer/answer/6048248?hl=en 2022/04/28.

[28] Google. 2022. Android Permission. https://developer.android.com/reference/android/Manifest.permission 2022/04/28.

[29] Google. 2022. Obtaining Consent with the User Messaging Platform. https://developers.google.com/admob/ump/android/quick-start 2022/04/24.

[30] IAB Europe GDPR Implementation Group. 2017. The definition of Personal Data - Working Paper 02/2017. https://iabeurope.eu/wp-content/uploads/2019/08/20170719-IABEU-GIG-Working-Paper02_Personal-Data.pdf.

[31] Peter Hornyack, Seungyeop Han, Jaeyeon Jung, Stuart Schechter, and David Wetherall. 2011. These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In *CCS*.

[32] Andreas Hotho, Steffen Staab, and Gerd Stumme. 2003. Ontologies improve text document clustering. In *ICDM*.

[33] ICCL. 2022/03/08. Landmark litigation. https://www.iccl.ie/rtb-june-2021/.

[34] Georgios Kampanos, Siamak F Shahandashti, and Name Name. 2021. Accept All: The Landscape of Cookie Banners in Greece and the UK. In *IFIP SEC*.

[35] Simon Koch, Malte Wessels, Benjamin Altpeter, Madita Olvermann, and Martin Johns. 2022. Keeping Privacy Labels Honest. *PoPETs* (2022).

[36] Konrad Kollnig, Pierre Dewitte, Max Van Kleek, Ge Wang, Daniel Omeiza, Helena Webb, and Nigel Shadbolt. 2021. A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps. In *SOUPS*.

[37] Ronald Leenes and Eleni Kosta. 2015. Taming the cookie monster with dutch law–a tale of regulatory failure. *Computer Law & Security Review* (2015).

[38] Li Li, Tegawendé F Bissyandé, Mike Papadakis, Siegfried Rasthofer, Alexandre Bartel, Damien Octeau, Jacques Klein, and Le Traon. 2017. Static analysis of android apps: A systematic literature review. *Information and Software Technology* (2017).

[39] Yuanchun Li, Ziyue Yang, Yao Guo, and Xiangqun Chen. 2017. Droidbot: a lightweight ui-guided test input generator for android. In *ICSE-C*.

[40] Rachel Tsz-Wai Lo, Ben He, and Iadh Ounis. 2005. Automatically building a stopword list for an information retrieval system. In *DIR*.

[41] Célestin Matte, Nataliia Bielova, and Cristiana Santos. 2020. Do Cookie Banners Respect my Choice?: Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. In *SP*.

[42] GA Miller. 1995. WordNet: a lexical database for English. *COMMUN ACM* (1995).

[43] Trung Tin Nguyen, Michael Backes, Ninja Marnau, and Ben Stock. 2021. Share First, Ask Later (or Never?) Studying Violations of {GDPR's} Explicit Consent in Android Apps. In *USENIX Security*.

[44] Trung Tin Nguyen, Duc Cuong Nguyen, Michael Schilling, Gang Wang, and Michael Backes. 2020. Measuring User Perception for Detecting Unexpected Access to Sensitive Resource in Mobile Apps. In *ASIA CCS*.

[45] NOYB – European Center for Digital Rights. 2021. Google: If you don't want us to track your phone – just get another tracking ID! https://noyb.eu/en/complaint-filed-against-google-tracking-id. 2021/01/17.

[46] objection. 2021. Runtime Mobile Exploration. https://github.com/sensepost/objection. 2021/01/17.

[47] Marten Oltrogge, Nicolas Huaman, Sabrina Amft, Yasemin Acar, Michael Backes, and Sascha Fahl. 2021. Why Eve and Mallory Still Love Android: Revisiting {TLS}({In) Security} in Android Applications. In *USENIX Security*.

[48] Xiang Pan, Yinzhi Cao, Xuechao Du, Boyuan He, Gan Fang, Rui Shao, and Yan Chen. 2018. Flowcog: context-aware semantics extraction and analysis of information flow leaks in android apps. In *USENIX Security*.

[49] The European Parliament and the Council of the European Union. 2002. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). (2002).

[50] Data Protection Working Party. 2010. Opinion 4/2010 on the European code of conduct of FEDMA for the use of personal data in direct marketing. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp174_en.pdf. 2022/04/28.

[51] Data Protection Working Party. 2016. Guidelines on Consent under Regulation 2016/679 (wp259rev.01). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051. 2020/09/04.

[52] Sai Teja Peddinti, Igor Bilogrevic, Nina Taft, Martin Pelikan, Ulfar Erlingsson, Pauline Anthonysamy, and Giles Hogben. 2019. Reducing Permission Requests in Mobile Apps. In *Proceedings of ACM Internet Measurement Conference (IMC)*.

[53] publicsuffixlist. 2021. publicsuffixlist. https://github.com/ko-zu/psl. 2021/05.

[54] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, and Phillipa Gill. 2018. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In *NDSS*.

[55] General Data Protection Regulation. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. *OJEU* (2016).

[56] Jingjing Ren, Ashwin Rao, Martina Lindorfer, Arnaud Legout, and David Choffnes. 2016. Recon: Revealing and controlling pii leaks in mobile network traffic. In *MobiSys*.

[57] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't somebody think of the children?" examining COPPA compliance at scale. *PETS* (2018).

[58] Iskander Sanchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. 2019. Can i opt out yet? gdpr and the global illusion of cookie control. In *Asia CCS*.

[59] Rocky Slavin, Xiaoyin Wang, Mitra Bokaei Hosseini, James Hester, Ram Krishnan, Jaspreet Bhatia, Travis D Breaux, and Jianwei Niu. 2016. Toward a framework for detecting privacy policy violations in android application code. In *ICSE*.

[60] Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. 2010. Short text classification in twitter to improve information filtering. In *ACM SIGIR*.

[61] Ben Stock, Giancarlo Pellegrino, Christian Rossow, Martin Johns, and Michael Backes. 2016. Hey, you have a problem: On the feasibility of large-scale web vulnerability notification. In *USENIX Security*.

[62] Techcrunch. 2022. France spanks Google $170M, Facebook $68M over cookie consent dark patterns. https://techcrunch.com/2022/01/06/cnil-facebook-google-cookie-consent-eprivacy-breaches/ 2022/04/28.

[63] Techcrunch. 2022. How a small French privacy ruling could remake adtech for good. https://techcrunch.com/2018/11/20/how-a-small-french-privacy-ruling-could-remake-adtech-for-good/ 2022/04/28.

[64] Vasant Tendulkar and William Enck. 2014. An application package configuration approach to mitigating android ssl vulnerabilities. *MoST* (2014).

[65] Tesseract. 2022/03/08. Tesseract-OCR. https://github.com/tesseract-ocr/tesseract.

[66] Stefano Traverso, Martino Trevisan, Leonardo Giannantoni, Marco Mellia, and Hassan Metwalley. 2017. Benchmark and comparison of tracker-blockers: Should you trust them?. In *TMA*.

[67] Martino Trevisan, Stefano Traverso, Eleonora Bassi, and Marco Mellia. 2019. 4 years of EU cookie law: Results and lessons learned. *PETS* (2019).

[68] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un) informed Consent: Studying GDPR Consent Notices in the Field. In *CCS*.

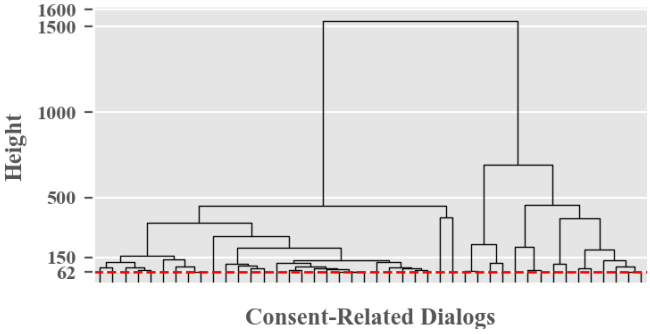


Figure 6: The dendrogram of our hierarchical clustering.

[69] Pelayo Vallina, Álvaro Feal, Julien Gamba, Narseo Vallina-Rodriguez, and Antonio Fernández Anta. 2019. Tales from the porn: A comprehensive privacy analysis of the web porn ecosystem. In *IMC*.
[70] Yan Wang, Hailong Zhang, and Atanas Rountev. 2016. On the unsoundness of static analysis for Android GUIs. In *PLDI*.
[71] Joe H Ward Jr. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* (1963).
[72] Takuya Watanabe, Mitsuaki Akiyama, Tetsuya Sakai, and Tatsuya Mori. 2015. Understanding the Inconsistencies between Text Descriptions and the Use of Privacy-sensitive Resources of Mobile Apps. In *SOUPS*.
[73] Charles Weir, Ben Hermann, and Sascha Fahl. 2020. From Needs to Actions to Secure Apps? The Effect of Requirements and Developer Practices on App Security. In *USENIX Security*.
[74] Primal Wijesekera, Arjun Baokar, Ashkan Hosseini, Serge Egelman, David Wagner, and Konstantin Beznosov. 2015. Android permissions remystified: A field study on contextual integrity. In *USENIX Security*.
[75] Lei Xue, Hao Zhou, Xiapu Luo, Le Yu, Dinghao Wu, Yajin Zhou, and Xiaobo Ma. 2020. Packergrind: An adaptive unpacking system for android apps. *IEEE Trans. Softw. Eng* (2020).
[76] Zhemin Yang, Min Yang, Yuan Zhang, Guofei Gu, Peng Ning, and X Sean Wang. 2013. Appintent: Analyzing sensitive data transmission in android for privacy leakage detection. In *CCS*.
[77] Le Yu, Xiapu Luo, Xule Liu, and Tao Zhang. 2016. Can we trust the privacy policies of android apps?. In *DSN*.
[78] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman M Sadeh, Steven M Bellovin, and Joel R Reidenberg. 2017. Automated Analysis of Privacy Requirements for Mobile Apps.. In *NDSS*.

A HIERARCHICAL CLUSTERING ANALYSIS

To select the best threshold for identifying similar privacy-related user interfaces, we considered the tradeoff between the quality of the clustering against the number of clusters. By manually inspecting the cluster quality and testing with different distances ranging from 50 to 150 based on the dendrogram of our hierarchical clustering in Figure 6, we have the best quality result at the height 62. The best quality cluster means that the majority of privacy-related user interfaces in the same group would be more similar to each other than in another group. Besides interpreting the dendrogram, we used the silhouette score to evaluate the quality of clusters. By inspecting the silhouette plot, we found that the first highest was at 9 clusters, and the second highest was at 44 clusters. Then the number of clusters started to increase significantly in correlation with the value of the silhouette score.