Hindawi
Scientific Programming
Volume 2022, Article ID 2508690, 15 pages
https://doi.org/10.1155/2022/2508690

Research Article

Evaluating the Privacy Policy of Android Apps: A Privacy Policy Compliance Study for Popular Apps in China and Europe

Kaijun Liu,[1,2] **Guoai Xu**,[1,2] **Xiaomei Zhang**,[3] **Guosheng Xu**,[1,2] **and Zhangjie Zhao**[4]

[1]*School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*National Engineering Research Center of Mobile Network Security, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[3]*China Cybersecurity Review Technology and Certification Center, Beijing 100020, China*
[4]*Beijing Big Data Center, Beijing 100101, China*

Correspondence should be addressed to Kaijun Liu; liukaijun@bupt.edu.cn

Received 20 February 2022; Revised 26 June 2022; Accepted 26 July 2022; Published 23 August 2022

Academic Editor: Zhihan Liu

Recently, with the increase in the market share of the Android system and the sharp increase in the number of Android mobile apps, many countries and regions have successively launched laws and regulations related to data security. The EU's GDPR and China's Information Security Technology-Personal Information Security Specification are two of the most important bills, affecting vast areas and large populations. Both regulations impose requirements on privacy policy specifications for Android apps. With these requirements, however, apps' privacy policies have become larger. Researchers have conducted studies on whether the actual privacy behavior of apps conforms to their privacy policy description but have not focused on compliance with the privacy policy itself. In this paper, we propose evaluation metrics for privacy policy compliance and evaluate popular apps by analyzing privacy policies and apps. We applied our method to 1,000 apps from the Google Play Store in Europe and 1,000 apps from the Tencent Appstore in China. We detected a number of app privacy policy noncompliance issues and discovered a number of privacy issues with third-party services and third-party libraries.

1. Introduction

According to statistics [1], in October 2021, the Android [2] system accounted for 71% of the global smartphone market. Since mobile apps built for Android can access sensitive personal data such as user location, network information, and unique device identification, this substantial market share increases the risk of users' personal data breaches. Acts on personal data protection have been introduced around the world in recent years. The General Data Protection Regulation (GDPR) [3] is a data protection act introduced by the European Union in 2016. On May 25, 2018, the GDPR came into force. The law has a huge impact worldwide because any company that stores or processes personal information about EU citizens in EU countries must comply with the GDPR even if they have no business presence in the EU. In 2017, China issued the *Information Security Technology-Personal Information Security Specification* (hereinafter referred to as "Specification") [4] and then updated the standard in 2020 and released the English version of the standard simultaneously [5].

The above two regulations affect a large number of people, have a wide range of influence in the world, and are the basis for much data supervision work. Both of these laws or rules lay out their obligations to data controllers and clarify the rights of data subjects. In the case of Android apps, the data controller is the organization represented by the app developer, and the data subject is the user of the app. One of the obligations of the data controller is to provide the data subject with information on data acquisition, collection, sharing, and storage and inform the data subject of their rights to the data. This part of the information is often referred to as the privacy policy [6]. Both the GDPR and the Specification include the corresponding requirements for privacy policies, guiding developers to write qualified privacy policies.

In previous studies, researchers conducted research on the app privacy policy and the actual privacy behavior of the app, and analyzed whether the privacy policy was consistent with the app's behavior [7]. Researchers also focused on the issue of consent notices displayed by apps after the relevant laws took effect [8]. But the community lacks research on the compliance of the app's privacy policy itself. Figure 1 shows an excerpt of the privacy policy of Super Space Cleaner [9], an app with over 5 million downloads on the Google Play Store. It is not difficult to find that such a disclosure of privacy behavior is not an effective way to show users what kind of data the app needs and how the data will be used. Even if the user agrees to the policy and uses the app, the user has no idea what personal data might be sent or stored. This raises a question: is the privacy policy of the popular apps GDPR-compliant? Does a noncompliant privacy policy mean more noncompliant data breaches?

To investigate app privacy policy compliance issues, we conducted a study on the top 1,000 most popular apps each from Google Play Store in Europe and Tencent Appstore in China [10]. We used popular apps from Google Play Store in Europe to test their privacy policies for compliance with GDPR and popular apps from the Tencent Appstore in China to test their privacy policies for compliance with the Specification. We first identified the specification of privacy policy writing in accordance with GDPR and the Specification and then built an automated and scalable pipeline for analyzing the app's privacy policy specification and the app's possible collection of private data with actual privacy behavior and applying it to our dataset. The related code and sample app list are available at https://github.com/xingyueren-qinmu/PPE_code.

Our research makes the following contributions:

(i) A privacy policy evaluation scheme is proposed. By analyzing the privacy policy text and app, the readability, completeness, and accuracy of the policy are obtained.

(ii) We tested 1,000 app samples each from the Google Play Store and Tencent Appstore to understand the privacy policy compliance of current popular apps.

(iii) We analyzed the test results and discussed the presentation format of the apps' privacy policies, the completeness and accuracy of privacy policies, user consent violations, and the privacy behaviors of third-party services and libraries.

This article is organized as follows: Section 2 introduces the relevant work and research objectives of our research; Section 3 introduces the privacy policy requirements in GDPR and the Specification, as well as our perspective on evaluating privacy policies; Section 4 introduces our privacy policy text analysis method; Section 5 presents our app analysis method; Section 6 is an indepth analysis of our data results, Section 7 is the discussion; and Section 8 is the conclusion.

> This information includes your device brand, device model, network status, etc. The device information we collect does not contain any user-sensitive information, such as device ID or any information that can be used to permanently identify the user or device.

Figure 1: An excerpt of the privacy policy of Super Space Cleaner.

2. Context of Our Work and Research Questions

Our work focused on whether the privacy policies of popular Android apps in the two major markets are complete under the requirements of GDPR and the Specification and whether a noncompliant privacy policy means that the app has more privacy issues. Is there a strong correlation between the app's privacy policy and the app's own privacy practices? In this section, we present the related work of the study and our research goals.

2.1. Context of Our Work. With the popularity of the Android operating system and the widespread use of Android apps, many researchers are concerned about the veracity of the privacy policy of the app and whether the actual privacy behavior of the app is consistent with the privacy policy.

In 2016, there was still lack of definition for the privacy policy writing specification. Yu et al. [11] used NLP to analyze the privacy policy of Android apps sentence by sentence and analyzed the app's calls to privacy APIs and data sending behaviors to determine whether the remaining privacy policies matched. Slavin et al. [7] focused on the language structure of privacy policies and proposed a framework based on a privacy policy phrase ontology and a set of mappings from API methods to policy phrases, which was used to detect violations. Wang et al. [12] focused on discovering guidance for private data input from a GUI and proposed a static analysis method that can be used to analyze dynamic GUI interfaces for large-scale analysis of app interfaces and identify possible data breaches. Kununka [13] et al. conducted a detailed study on a small number of apps and found that the privacy behavior and privacy policies of the target samples were inconsistent. However, due to the small number of analysis samples, there may be data discrimination problems. Yu et al. [14] designed a tool to comprehensively analyze privacy policies, app bytecodes, app advertisement descriptions, and app permissions to discover app privacy violations. During their research, they proposed nine semantic patterns in privacy policies, covering most of the privacy policy topics. In 2018, Ferrara and Spoto [15] performed the automated static analysis of apps to customize reports for the four major players in the GDPR compliance process. Momen et al. [16] compared whether the app privacy problem had improved before and after the introduction of GDPR through a detailed analysis of app permissions. Their research shows that the GDPR has a certain normative effect on app privacy behavior. Fan et al.

[17] proposed an automated system to detect GDPR violations of apps and privacy policies by identifying data practices declared in the app's privacy policy and data-related behaviors in the app code. Guamán et al. [18, 19] focused on whether the app violated the relevant provisions of the GDPR when transmitting data across borders. In 2021, Nguyen et al. [8] examined the illegal behavior of apps collecting privacy data before the user had agreed to the privacy statement. They paid special attention to the data collection behavior of the advertising domain and informed the developers of this behavior.

2.2. Research Questions. In this work, we examined the legal compliance with privacy policies for popular Android apps. We first proposed three metrics for analyzing app privacy policy compliance and developed an automated pipeline to implement the analysis method. Then, we performed the tests on apps in China and Europe. Finally, we conducted an indepth analysis of the detection data. Specifically, our study aims to answer the following research questions:

RQ1: what are the requirements of laws and regulations for the app's privacy policy? We summarized and compared the GDPR and the Specification in the privacy policy-related regulations and sorted a total of 10 requirements.

RQ2: How can we evaluate a privacy policy's compliance? We proposed three evaluation perspectives for privacy policies, which are the integrity of app privacy policy; privacy data collection before user consent; and accuracy of the app privacy policy. Then, we designed the corresponding automated detection method and applied the method to 1,000 popular apps each from Google Play Store in Europe and Tencent Appstore in China.

RQ3: what is the privacy policy compliance of Tencent Appstore (popular apps in China) and Google Play Store in Europe (popular apps in Europe)? We compare the privacy policy compliance of popular apps from the Tencent Appstore and Google Play Store in Europe from multiple perspectives and conduct indepth discussion on privacy violations by third-party services.

2.3. Data Collection.

Regulation Text. We downloaded the GDPR text file from the GDPR official website [3] (updated to 23.05.2018). We downloaded the full text of the specification from the China National Standard Full-Text Open System [20] (updated to 06.03.2020). The research on the content of regulations in the following sections will be based on these two texts.

App Samples. To obtain the popular apps for Google Play Store in Europe, we crawled the popular app list on Diandian Data [21] from October 2021 for each EU member state and got the most popular 500 free apps and 500 paid apps sorted by the number of downloads. Then, we downloaded these apps using a web crawler. We examined the top 100 most downloaded apps in each of 10 categories in October 2021 from the Tencent Appstore, for which reason we cannot get the total ranking of apps for the Tencent Appstore but can only collect the rankings of classified apps in the Tencent Appstore. See the app list in our GitHub repository for details. The app study described in the following sections is based on these 2,000 apps.

Privacy Policy Samples. The above 2,000 apps have links to their privacy policies on the corresponding pages of the app store. We crawled the privacy policies of these apps. The follow-up studies are conducted based on these samples.

3. Privacy Policy Specification

3.1. Legal Background of Privacy Policy. In GDPR Section 3, Rights of the Data Subjects, clause 13, "Information to be provided where personal data is collected from the data subject," and clause 14, "Information to be provided where personal data has not been obtained from the data subject," specify that data controllers need to provide information to data subjects. Articles 15–22 enumerate the rights of the data subject. From this, the GDPR requirements for privacy policy writing can be distilled. In the Specification, Article 5.5 "Personal Information Protection Policy" proposes the requirements of the personal information protection policy for personal information controllers, and a template example of the personal information protection policy is shown in Appendix of Specification.

Using the requirements and templates of GDPR and the Specification, we summarize the specifications for the writing of privacy policies. The main topics of privacy policies are as follows:

Summary of the Policy. This part includes the scope of products or services to which the personal information protection the policy applies, applicable personal information subject type, effective and updated time, and the policy directory.

How the App Collects and Uses Personal Data? This part needs to list the types of personal information required and the method of collection (such as obtaining permission through the API and obtaining it through cookies) based on different business functions or interest.

How Personal Data Are Shared, Transferred, and Publicly Disclosed? This section needs to explain the reasons for sharing and transferring personal information; the types of personal information that need to be shared and transferred; and the types and identities of data recipients. This section also needs to describe the types and reasons for the public disclosure of personal data, where appropriate, when the consent of the data subject is not needed (e.g., such as information required by a legal authority).

How the Data Controller Protects the User's Personal Data? This part of the Specification requires the data

controller to explain the security protection measures for the user's personal data, the personal information security protocol currently followed, and the certification and security obtained. There are no clear requirements in this part of the GDPR.

How the Data Controller Stores the User's Personal Data? This section explains where and for how long the data controller stores the user's personal data.

Data Subject Rights. This section needs to express to the users their rights as a data subject, including data access rights, rectification rights, deletion rights, restriction of processing rights, data portability rights, right to object, and automation of personal decision-making.

How the Data Subject Can Lodge a Complaint or Appeal with the Supervisory Authority? This section requires the data controller to publish the complaint, the reporting method, and the response time to the complaint for the user.

How to Handle Children's Personal Information? This section needs to explain that the collection of children's information requires the consent of the guardian.

In addition to the above requirements, the GDPR requires that the privacy policy contains a corresponding legal basis, that is, the policy needs to state that the basis for writing is the GDPR. In the Specification, complaints and responses are mandatory in privacy policies.

3.2. Explicit Privacy Policy to Users. According to our analysis, the average length of the English privacy policy is more than 3,000 words, and the average length of the Chinese privacy policy is more than 8,000 words, which makes it difficult for users to read before using the app. As a result, various laws and regulations require apps to explicitly ask users for consent to ensure that users are aware of the app's privacy acquisition behavior. GDPR requires consent in Section 2.2 to be freely given, specific, informed, and unambiguous. In recent years, China has contacted and issued a number of testing bases and testing specifications, requiring that, after the app is installed and before the basic business starts, the personal data processing rules should be clearly displayed to users through the interactive interface. Therefore, whether it is an app applicable to the GDPR or the Specification, when the app starts to obtain and collect the user's private data, it must seek the user's consent through a visual interface.

Based on the above two regulations' requirements on privacy policy, our evaluation of the privacy policy was divided into three parts:

(1) *Evaluation of the Integrity of the Privacy Policy*. Pay attention to whether the content of the app privacy policy is complete and whether it contains the topics required by law. Section 4 presents our specific approach.

(2) *Evaluation of Asking for Consent for Collecting Privacy Data*. Pay attention to whether the privacy regulations are clearly stated in the app and whether data collection is started after the user agrees. Section 5 introduces this part in detail.

(3) *Evaluation of the Accuracy of the Privacy Policy*. This part focuses on whether the content of the privacy policy is consistent with the actual privacy behavior of the app. The specific analysis method is introduced in Section 5.

4. Privacy Policy Integrity

In this section, we introduce the privacy policy integrity analysis method. The text of the privacy policy is expressed in various ways, which brings difficulties to the analysis. Considering the analysis efficiency, combined with the characteristics of the privacy text with paragraph subtitles, we chose to analyze the theme of the privacy policy. We considered a privacy policy to be complete if it addressed all the topics covered by the relevant regulations.

4.1. Privacy Policy Text Preprocessing. We crawled the privacy policy from the app stores, and due to the diverse structure of the privacy policy page, we used the URL2io service [22] to extract the body of the web page. URL2io filtered out the node tree that may be the text in the webpage source code and used the evaluation algorithm based on machine learning to predict the text node. Of the 2,000 privacy policies, we successfully obtained 1,834 privacy policy texts using URL2io and 62 privacy policies manually; the rest were either linked incorrectly or did not exist in text form.

We chose to classify the title content in the text of the app to judge the integrity of the subject of the privacy policy. The reason we did not extract the subject from the body paragraphs is that developers formulate privacy policies differently. In the absence of standard datasets for training classification models, it is difficult to accurately extract the content of paragraphs. In contrast, analyzing headlines that consist of only one sentence is much easier. We first used regular expression matching to obtain the content of the text title, such as matching #, ∗, and other format markers, along with paragraph numbers, and built a text tree according to the title number and title level and output to the privacy policy integrity analysis module.

4.2. Method of Privacy Policy Integrity Analysis. Figure 2 shows the workflow of integrity analysis. First, we manually extracted some of the titles in the privacy policy to construct a topic-candidate title (T-CT) table. Then, we used Bert [23] to calculate the features of the candidate titles (CTs) and the privacy policy titles (PPTs) to be tested. Next, we used the cosine similarity algorithm to calculate the similarity between PPT and CT. Finally, we determined the threshold of similarity through several experiments to realize the privacy policy integrity analysis work.

Topic-Candidate Title Table Building: Both English and Chinese privacy policies are diverse in the way they

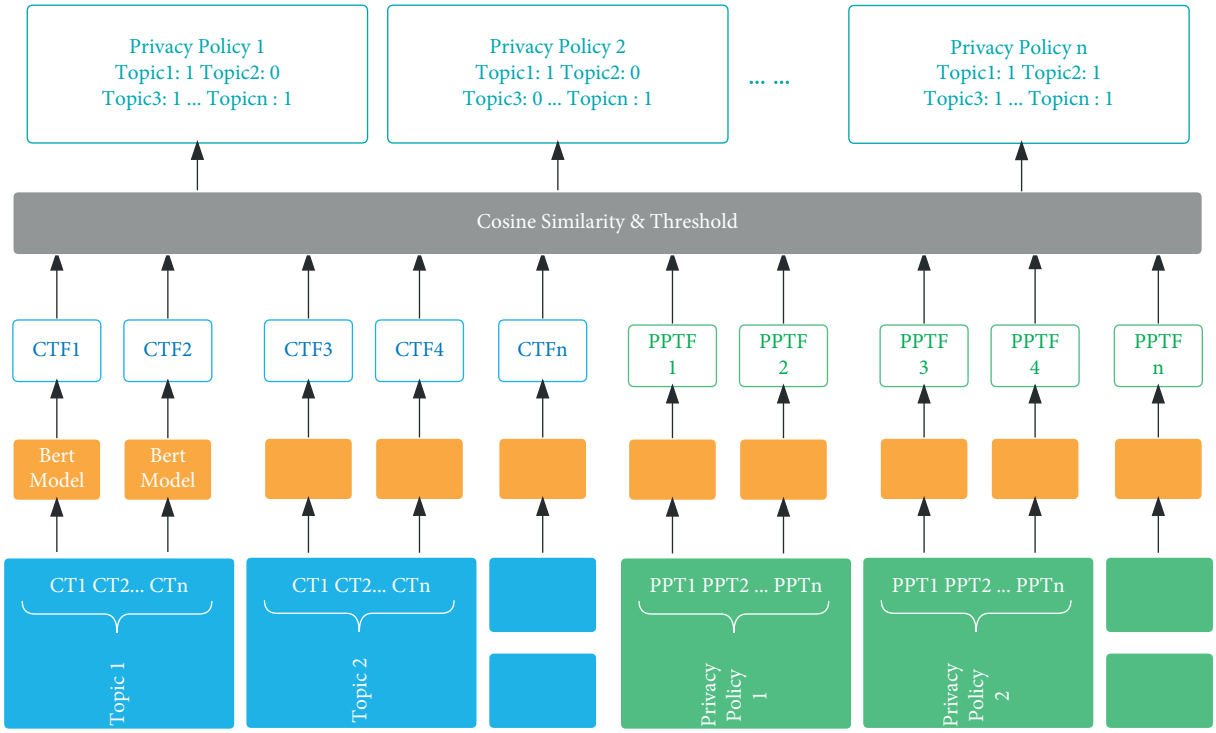

Figure 2: Workflow of privacy policy subject integrity analysis.

express and describe the same topic matter. Therefore, if only one expression of the topic is used to calculate similarity with the title in the privacy policy text, it may lead to false negatives. That is, the privacy policy actually contains the topic, but the similarity between the privacy policy text and the topic text is too low because of the different expressions. To avoid this, we need to create a table of topic-candidate titles. We manually sampled 40 policies from each of the two stores. The criteria we followed when sampling was that the sample met the subject requirements of GDPR and the Specification. To avoid discrimination, multiple samples from the same developer were not used, and samples were evenly distributed according to the number of downloads (we automatically extracted the downloads of the app from the store page when obtaining the samples). We extracted the description of each subject that met the requirements from the sample as a CT. Figure 3 shows possible expressions for the topic "how to collect users' personal data", all of which make up our T-CT table.

"Information Collection": [
"How Do We Collect Your Information",
"Information We Collect About You",
"What information Do We Receive",
"The information we collect",
"Collection of your personal infromation",
"What we collect"
]

Figure 3: Candidate titles of the topic "how to collect users' personal data?"

Bert-Based Sentence Feature Calculation (Figure 4). We chose the Bert model for our experiments. Bert is a very popular pretrained language model that understands language well and performs downstream tasks well with a small amount of data fine-tuning. Bert used Transformer's bi-directional attention mechanism to design the network and designed two semi-supervised training methods, Mask Language Model and Next Sentence Prediction, in the pre-training process. The first part is the embedding layer, the second part is the encoder layer, and the third part is the pooler layer. The embedding layer is divided into three parts: word embedding, segmentation embedding, and position embedding. The encoder layer uses the bi-directional attention mechanism of the transformer, and there are 12 encoder layers consisting of the attention mechanism network and the forward fully connected network. The pooler layer is responsible for downscaling the result vector from the encoder layer to the final result vector. In our study, for the Chinese privacy policy, we use the BERT-Large Chinese model proposed by Cui et al. [24], which is a Chinese Bert model with a word list size of 21128, a number of nodes in the hidden layer of 1024, a total of 24 layers in the encoder layer, and a number of parameters of 330 M. For the English privacy policy, we use the BERT-Large English model provided by Turc et al. [25]. The word table size is 30,522, the number of nodes in the hidden layer is 1024, the encoder layer has 24 layers, and the number of parameters is 340 M.

The specific feature process is as follows. First, we use the Bert model to compute the feature vectors for each CT in the T-CT table in turn. Specifically, the average of the

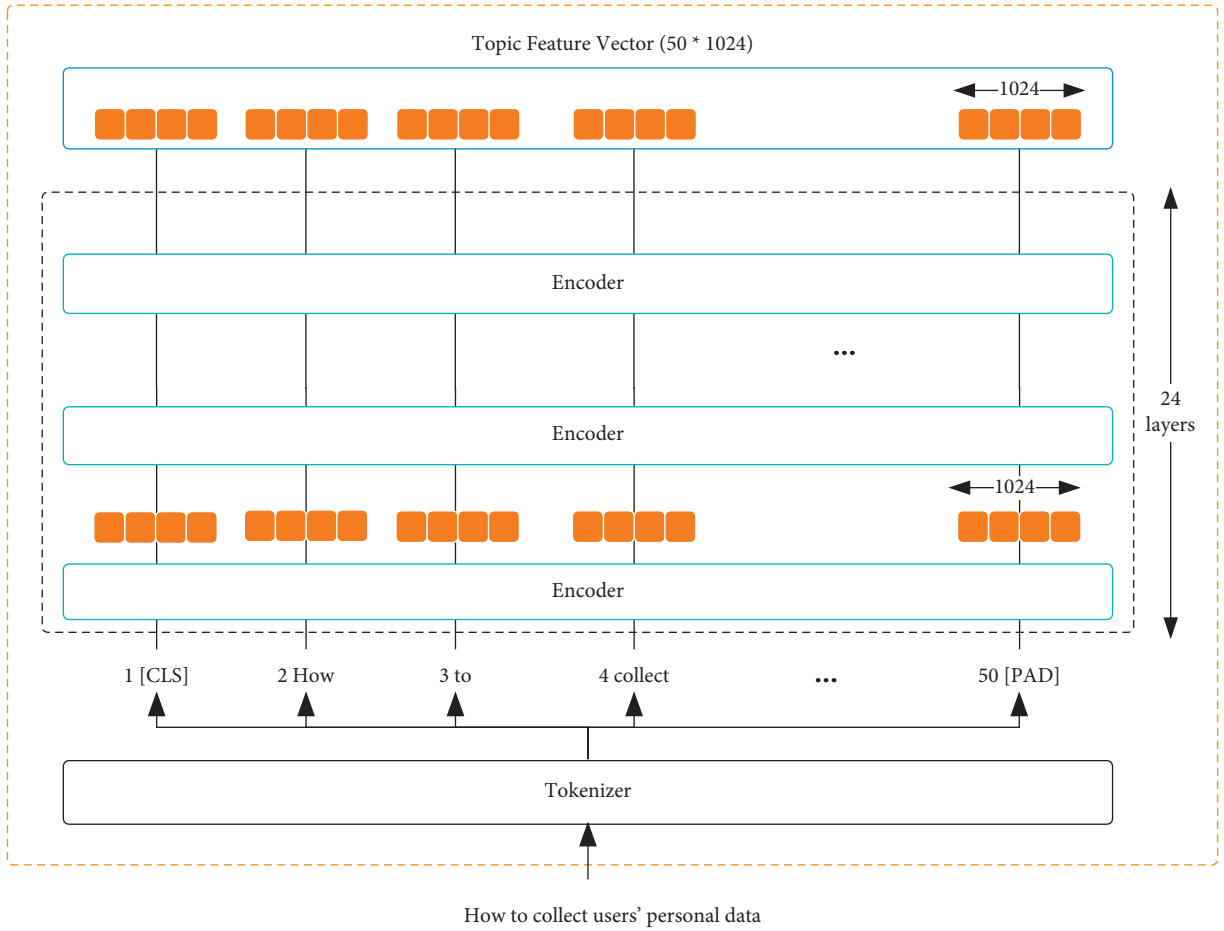

Figure 4: Bert model we use.

embedding output for each word in the headline is calculated. Then, for each privacy policy to be tested, we extract the titles in the text as PPTs and compute their feature vectors using the same method. The feature vectors of all our CTs and PPTs are recorded in the library and used for subsequent similarity calculations:

Similarity Calculation. We computed the semantic features of all CTs and PPTs using the Bert model. Here, we chose the cosine similarity algorithm to calculate the similarity between the two feature values. The cosine similarity algorithm measures the similarity between two vectors by measuring the cosine of the angle between them. The sensitivity of this algorithm to the direction of the vectors makes it widely used in classification work. For the output of the Bert model, we traverse the eigenvalues of each PPT and use cosine similarity to calculate its similarity to each CT eigenvalue. Formula (1) shows the calculation of the similarity of semantic features of two titles, where CT_i denotes the ith candidate title, PPT_j denotes the jth privacy policy title, and $W_k(x)$ denotes taking the kth dimensional component of the eigenvector$x$. Once the similarity between two titles exceeds a threshold, we assume that the privacy policy to which the PPT belongs contains the topic to which the CT belongs. After several rounds of debugging, we finally set the similarity threshold to 0.914 for Chinese and 0.922 for English.

$$\mathrm{sim}\left(CT_i, \mathrm{PPT}_j\right) = \frac{\sum_{k=1}^{1024} W_k(CT_i) \times W_k\left(\mathrm{PPT}_j\right)}{\sqrt{\sum_{k=1}^{1024} \left(W_k(CT_i)\right)^2} \times \sqrt{\sum_{k=1}^{1024} \left(W_k\left(\mathrm{PPT}_j\right)\right)^2}}. \tag{1}$$

5. App Analysis

In this section, we discuss two analyses we applied to apps in the Google Play Store and Tencent Appstore. Through dynamic analysis, we detect data collection without consent, and we detect the correctness of privacy policies through static analysis on the app. We describe our testing methods and processes in Sections 5.1 and 5.2, respectively.

5.1. Data Collection without Consent. Both the GDPR and the Specification require apps to ask users for consent, and that consent must be freely given, specific, informed, and unambiguous. That is to say, the acquisition and transmission of personal data must only take place after

the user's active consent (e.g., "click to accept"); otherwise, the app's privacy policy may be deemed to be in violation.

To detect whether the app's privacy policy is clearly showed and contains potential violations, we designed an automatic detection scheme to dynamically analyze the app. We chose to run the inspected app on a device (Pixel 1, Android 7.1.2) equipped with the Xposed framework [26] and our designed Hook module to analyze the app's private data acquisition behavior. To identify app data sending issues, we monitored the app's traffic. To intercept the TLS traffic, devices used our own root certificate for detection, and we used mitmproxy [27] to do this. We installed our own root certificate as a system certificate. Our detection ran in the environment described above. Figure 5 shows our analysis process. We first installed the app and then executed it automatically. During the execution process, we monitored API calls and network traffic. The specific scheme is as follows.

App Installation. Nguyen et al. [8] agreed to all the permission requests of the apps when installing them. Their purpose was to make the app have sufficient ability to obtain all possible data and identify the behavior of sending private data. In this research, for the same app, we used two installation methods: granting all permissions and retaining all permissions, that is, pipeline A simulated a normal usage scenario for the user, and no permissions were granted during installation; in pipeline B, all permissions were granted to the app during installation.

Automated Execution. We automate the execution of each app after it is installed. When the app runs for the first time, it may display a welcome screen that is not related to the analysis. For this situation, we customized the automatic running rules to let the app enter its real starting interface and then stop interfering with the running of the app. At this point, the app may have three forms of expression: the permissions for the app; displaying the privacy consent interface; and entering the main interface. We used the Android system tool UI Automator [28] to obtain the structural information of the interface for subsequent analysis. Note that apps do not ask us to pay and create an account during this process, and we are not required to do so.

API Monitoring and Traffic Monitoring. In the process of automated execution, we used the Xposed framework to monitor 41 APIs related to privacy data and cryptography (i.e., javax.crypto.Cipher and Base64 API). Part of the classes and the related privacy data are shown in Table 1. The API call parameters, return values, and call stack information were recorded in the log. We used mitmproxy to monitor the content of the request traffic and record the host, parameters, and request raw data in the log. API logs and traffic logs were used in subsequent analysis.

API and Traffic Correlation Analysis. We first searched the traffic for private data that can be obtained through the API, and we focused on the data shown in Table 2. These data are from Appendix A of the Specification, but some of them were not found during the analysis. So, the final test results are based on the display in Section 6.3. We obtained the relevant values from the test device via the Android debug bridge. We traversed the traffic log, performed string matching on the parameters, and requested the raw data from each request. The data used for string matching contains the private data content and the category to which it belongs. Taking "latitude 32.899" for instance, we use the various precisions of 32.899 and various spellings of the word "latitude" for string matching. Once there is a hit, the app has a privacy violation. Many apps send out encrypted traffic content, and simple string matching is ineffective with this kind of traffic. Our monitoring of cryptographic APIs gave us the ability to analyze the content of encrypted traffic. Consider the following example:

$$y = \mathrm{AES}(x), z = \mathrm{base64}(y). \tag{2}$$

when we monitor the traffic whose request raw is z (or contains z), by traversing the input and output of the cryptography API, we can get the input y from the Base64 call and get the input x from the AES call, so as to decrypt the cypher traffic. To deal with the possibility of multiple layers of encryption (such as AES (AES (. . .AES (x). . .))), we keep traversing the cryptographic API until the associated input and output content cannot be obtained. In the previous research, we found that, among the 48 lines of traffic, only one line was encrypted twice, so this design has little impact on the analysis efficiency of the system. We tested the above scheme and the scheme based on cryptography decryption in [29] on the "No. 1 Community" app and analyzed 12 encrypted API Hook results and 8 traffic monitoring results. Our method took 0.434 s, and the method in [29] took 1.568 s (the test environment was MACOS/2.6 GHz 6core Core-i7/16 GB RAM). The advantages were significant.

Identification of Privacy Leakage Behaviors. We divided privacy behaviors into three types: data leakage, potential data leakage, and data acquisition. The act of obtaining and sending private data before asking for the user's consent violates the GDPR and the Specification, and we marked it as a kind of data leakage; the use of a private network protocol and delayed sending (saving the data in a file and transmitting it through the Internet at some point in the future) was considered obtaining private data and encrypting it, so it was marked as a potential data leak. For behaviors that only obtain private data without sending or encrypting it, we marked it as data acquisition because

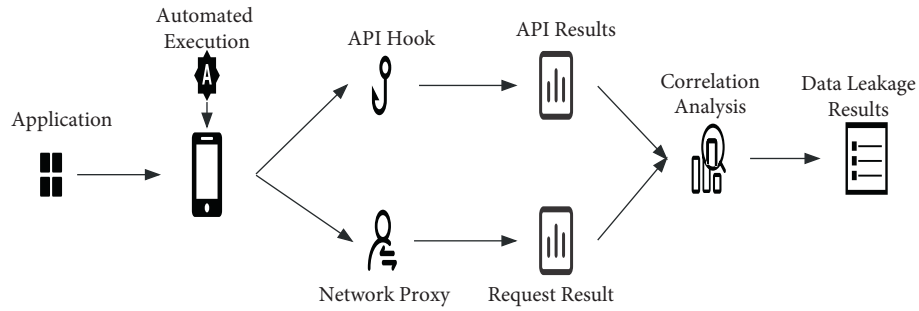

Figure 5: User consent analysis.

Table 1: Classes and related privacy data.

Class	Number of methods	Related privacy data
ContentResolver	7	Contacts, videos, pictures, SMS, recents, bookmarks, calendar
SmsManager	3	SMS
TelephonyManager	14	Phone number, location, IMSI, carrier
WifiManager	3	BSSID, SSID

Table 2: Device-related data.

Data type	Description
Android_ID	Android system ID
MAC	MAC address of WiFi interface
SN	Serial number
IMEI	Device ID
BSSID	MAC address of WIFI hotspot device
SSID	WIFI hotspot name
IMSI	International mobile subscriber identity
Phone number	Test device phone number
Location	Test device location
E-mail	Android account
Contact	Name and number of contacts
SMS	Short message service information
Recents	Call logs

it is unreasonable to obtain private data without the user's consent.

Using the above pipeline analysis, we obtained the private data acquisition and sending behavior of an app before obtaining the user's consent (regardless of whether the app asks the user for privacy consent) and defined the relevant behavior as data leakage, potential data leakage, or data acquisition. The analysis of relevant results is given in Section 6.

5.2. Privacy Policy Accuracy Analysis. We considered an app's privacy policy to be accurate if all kinds of privacy that may exist in an app are made clear to the user in the privacy policy. We divided privacy into two categories. One was the privacy that requires the app to apply for permission and obtain it through the API, such as geographic location, contacts, and photos. The other was to request the user to enter personal privacy information in the GUI, such as the user's name, age, and home address. For permission privacy, we extracted the permissions applied inside the app and searched the privacy policy by string matching according to the permission-privacy dictionary to determine whether the privacy policy is accurate. For private data obtained through the GUI interface, we focused on the interface context because the text in the interface guides the user to enter the private data. Our detection strategy was as follows: we first decompressed the APK file and then obtained all layout files in all resource directories and string.xml in the values directory. Then, we traversed and scanned the layout file, extracted the text attributes of all components, and recorded if the text attribute was a plaintext string; if the text was a resource address, we accessed the address recursively (because the content pointed to by the address may also be an address) and obtained string records. When dealing with Chinese characters, we paid attention to the encoding of the text, most of the apps have UTF-8 encoding, and a few are GBK. For analysis, we unified the character encoding to UTF-8. Finally, we performed a string matching analysis on the recorded string and 5 categories of personal data. As shown in Table 3, since the types of personal data of users are not specified in the GDPR, the types of privacy we focus here all come from the types of personal data specified in the Specification.

6. Indepth Analysis

In this section, we present our empirical findings on privacy policy violations for apps on the Google Play Store and Tencent Appstore. We first describe how the app's privacy policy is presented; next provide an in-depth analysis of the completeness and accuracy of the app's privacy policy based on the test results; then analyze the privacy policy's solicitation of user consent;besides discuss the compliance between privacy policy and the third-party services. Finally, we compare the compliance with the privacy policy of the two stores.

6.1. How the App Privacy Policy Is Presented? Among the 2,000 app samples in the Tencent Appstore and Google Play

Table 3: 5 categories of personal data.

Data category	Data
Personal basic infomation	Name, birthday, gender, racial, country, company...
Account information	E-mail account, bank account, transaction record
Identity information	ID, passport, military ID, social security card...
Heath information	Height, weight, medical history, allergy history...
Device information	SMS, contacts, app list, phone number, location, call record...

Store, we successfully obtained the privacy policy texts for 1,886 apps. There are three main reasons for the failure to obtain the privacy policy text: (a) the app does not list the privacy policy link in the store; (b) the app's privacy policy link has a problem (the link cannot be accessed, or the privacy policy link does not point to its privacy policy text page); (c) the privacy policy of the app is displayed in the PDF format or image format. Figure 6 shows the results of our analysis. Interestingly, missing links to privacy policies and link problems were mostly found in the Google Play Store app. We also noted that the Google Play Store requires that apps' privacy policies not be displayed in the PDF format [30], and 24 apps violated this rule.

Based on the above findings, we randomly sampled the three presentations of privacy policies. We manually installed and viewed the privacy policies of 30 apps, 10 of which were presented in text, 10 with pictures, and 10 presented in the PDF form. We found that only five apps opened the privacy policy page through the default browser, and the privacy policy presentations of these five apps were all text. Among the privacy policies presented as pictures and PDFs, eight apps had a zoom function on their viewing pages; three apps in PDF format and two apps in image format could not zoom in on a 5-inch $1920 * 1080$ screen, making it hard to see. Interestingly, we found that two apps' privacy policy linkspoint to a PDF file in the app store, while point to a web page within the app. Since extracting privacy policies from apps is a difficult task, we did not compare all the apps'privacy policy links in the app store and within the app.

In general, the privacy policy can be best displayed in text. Because whether the developer chooses to use Android WebView or the built-in browser of the mobile phone to display the privacy policy, the user can display a clear privacy policy by zooming in and out of the page. The text format is convenient for typesetting. Multiple privacy policies made by the same developer are unified via the format, thereby improving the user's reading experience. In contrast, pictures and PDF formats may not only increase the reading difficulty for users but also cannot guarantee that the display page of the privacy policy will not be tampered with (through traffic hijacking) and can only increase the cost of page tampering to a limited extent.

6.2. App Privacy Policy Completeness and Accuracy. As described in Section 4.1, we performed text integrity analysis on the privacy policy texts of 1,886 successfully obtained apps. The overall conclusion is that the privacy policies of the Tencent Appstore app are more normative than those of the

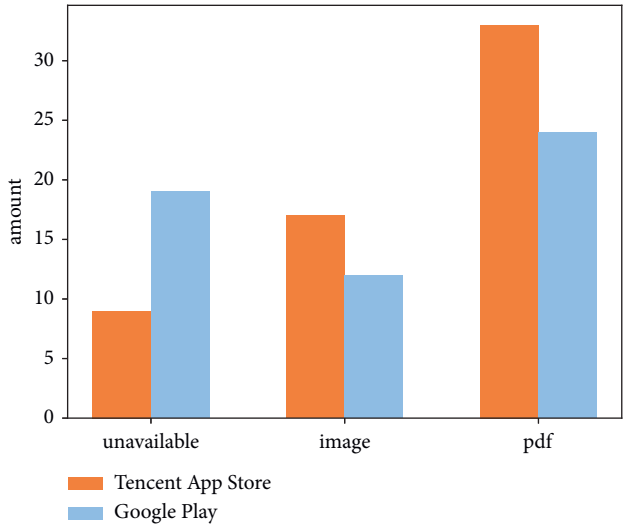

Figure 6: Reasons for the failure to obtain the privacy policy text.

Google Play Store. Figure 7 shows a comparison of the integrity of the app privacy policies of the two stores. Note that "Complaints and Responses" is a subject covered by the Specification requirements, and "Legal Basis" is a subject covered by the GDPR requirements. It can be found that "how personal information is collected and used" and "the sharing, transfer, and public disclosure of personal information" are topics covered in 91% of apps' privacy policies for both stores. Since the template of the privacy policy is given in the Specification, the privacy policy of the Tencent Appstore is relatively complete. The topic with the lowest inclusion rate is "Complaints and Responses," with an inclusion rate of 77%, and the inclusion rates of other topics are all higher than 88%. Compared with the Tencent Appstore, in the Google Play Store samples, except for "data subjects' rights" and "update of the privacy policy," the topic inclusion rate is lower than that of the Tencent Appstore samples. The inclusion rates of the topics "information storage," "protection for minors," and "legal basis" are all below 60%. Note that, for topics with an inclusion rate of less than 70%, we searched the full text through keyword extraction and string matching to prevent app developers from not presenting this content as separate topics.

Ninety-two percent of the privacy policies we successfully analyzed included the topics "how data is collected and used" and "how data is shared." Among the privacy policy topics listed in Section 3.1, these two topics are often directly related to privacy data types. Does the existence of the subject mean that the content of the privacy policy is

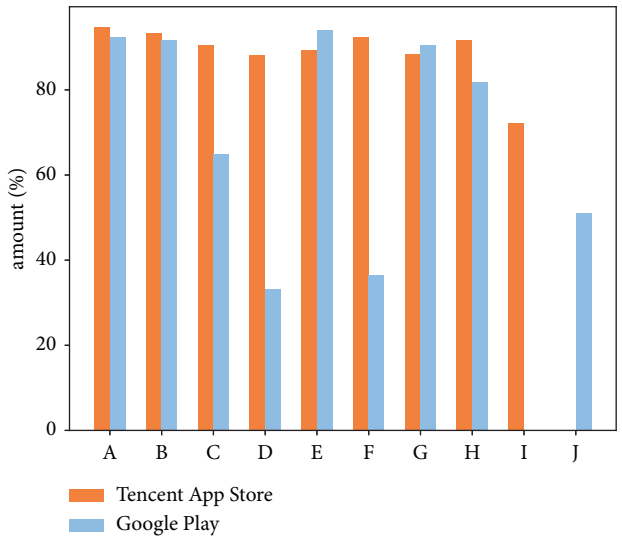

Figure 7: Privacy policy completeness. A: collection and usage of information; B: information sharing, transfer, and public disclosure; C: security of information; D: information storage; E: rights of the user; F: protection of children; G: policy updates; H: contact us; I: complaints and responses; J: legal basis.

accurate? Both the Specification and the GDPR require that the app privacy policy describe the specific types of personal data collected, used, and shared. Accordingly, we performed the interface text analysis in Section 5.2 on 2,000 apps and extracted the privacy texts we cared about from the app interface layout and resource files. After analysis, we found that the personal data most frequently requested by apps in Google Play Store and Tencent Appstore are mobile phone numbers, email addresses, usernames, geographic locations, and text messages. Mobile phone numbers and email addresses ranked first in the Google Play Store and Tencent Appstore, respectively. It can be seen that the strategies for app registration and login in China and in other countries are different. We used text matching to search for relevant personal data in the text of the app's privacy policy and counted the possible personal data collection behaviors that were not marked in the app's privacy policy. In the Google Play Store samples, at least six items of privacy data were not marked in the privacy policies of 52 apps, and we marked these apps as "inaccurate privacy policies".Compared to the samples from the Google Play Store, the Tencent Appstore performed better in this regard, only 25 apps had at least six items of private data that were not marked.

Interestingly, most of these inaccurate privacy policies covered at least six privacy policy topics. Among the 77 apps (with at least six items of privacy data not marked), only three apps had fewer than six privacy policy topics. Since the installation packages of these apps are each at least 70 MB, we did not manually run the apps and determine whether these personal data guides that do not appear in the privacy policy appear in the app interface, because traversing almost all of the interfaces of these apps is not possible.

From our evaluation of the integrity and accuracy of privacy policies, we can see that, under the requirements of the GDPR and the Specification, the privacy policies of popular apps have a high level of integrity and accuracy. After data analysis, it was found that the overall compliance of the sample privacy policy of the Tencent Appstore is better than that of the Google Play Store. It can be seen that the privacy policy template suggested in regulation can play a good guiding role in the normative privacy policy.

6.3. Privacy Behavior of the App before Consent Is Obtained.

Personal Data in Our Analysis. GDPR Article 4 [31] states that "personal data" means any information relating to an identified or identifiable natural person. The definition applies broadly. Under this definition, data tied to user devices (e.g., Android serial number), user network data (e.g., IP address), and user account data (e.g., e-mail) should be included in the scope of personal data. This is not friendly for security analysis research because there is no definite scope. Examples of specific personal information data are given in Appendix A of the Specification. Among them, personal communication data, personal Internet data, personal device data, and personal location data are the data that can be detected during dynamic analysis. Accordingly, our experiments focus on the data in the first column of Tables 4 and 5.

Data Leakage That Occurs before User Consent. As described in Section 5.1, we utilized two installation strategies for the app: granting no permissions upon app installation and granting all permissions upon app installation. The installation without granting any permission simulates the real scenario when the app is in use. Since our automated runtime script agrees to permission requests when permission pop-ups are granted, the data shown in Table 4 is the data leakage detected when partial permissions are granted. Also, this is the real data leakage problem that users may encounter when using the app. Granting all permissions to the app at installation allows the app code to access and send data to the maximum extent possible. The data in Table 5 shows the data leakage of the app with this installation method.

Granting required permissions, we found that 73 apps had DL (data leakage) before user consent, 82 apps had PDL (potential data leakage), and 716 apps had DA (data acquisition). With all permissions granted, the number of three privacy behaviors increased significantly, with 309 apps undergoing DL, 91 apps having PDL, and 970 apps having DA. We compared the types of private data involved in the two control groups, as shown in Tables 4 and 5. Based on the data detected above, we first analyzed whether the app explicitly asked for the user's consent and then analyzed the destination domain of private data leakage.

Request for user consent: when detecting the privacy behavior that occurs before the app obtains the user's consent, we saved the start-up interface information

Table 4: Privacy behaviors with required permissions.

Data type	DL		PDL		DA	
	Google	Tencent	Google	Tencent	Google	Tencent
IMEI	87	68	24	22	259	377
OS ID	—	—	—	—	377	259
SN	29	3	14	17	2	13
SSID	14	10	—	5	55	126
BSSID	2	10	—	—	41	117
App list	23	—	2	1	21	41
MAC	13	—	—	22	24	49
Location	11	12	—	1	7	26
IMSI	—	—	—	—	10	79
Number	—	1	—	—	3	6
E-mail	—	—	—	—	32	17
IP	—	—	—	—	106	123
Carrier	—	—	—	—	106	92
GSM ID	—	—	—	—	—	2
Sensor	—	—	—	—	190	92

Table 5: Privacy behaviors with all permissions.

Data type	DL		PDL		DA	
	Google	Tencent	Google	Tencent	Google	Tencent
IMEI	17	75	7	15	63	216
OS ID	132	87	20	14	527	279
SN	5	26	15	10	2	27
SSID	25	6	—	8	81	124
BSSID	24	10	—	—	66	124
App list	33	8	6	6	31	23
MAC	16	21	—	5	31	57
Location	68	9	1	12	110	130
IMSI	12	53	—	3	36	131
Number	1	2	—	—	12	17
IP	—	—	—	—	134	111
Carrier	—	—	—	—	172	127
E-mail	—	—	—	—	49	22
GSM ID	—	—	—	—	24	51
Sensor	—	—	—	—	298	93
Contact	—	—	—	—	11	6
Media	—	—	—	—	29	71
SMS	—	—	—	—	3	—
Call log	—	—	—	—	1	—
Calendar	—	—	—	—	1	3

after the app was installed for the first time. According to the requirements of GDPR 4–11 and the specification, when the app needs to obtain or process the user's private data, it needs to ask the user for consent visually. A screenshot of the user consent interface in the app My Spectrum [32] is shown in Figure 8. Based on the consent requirements in the two regulations, we determined that, after the first installation, apps with any of the behaviors of DL, PDL, and DA needed to present a consent solicitation page similar to that displayed by My Spectrum. This page needs to include a privacy statement, and a link to the privacy policy and clearly display the "Agree" and "Decline" buttons. After analysis, we found that 246/404 apps in the Tencent

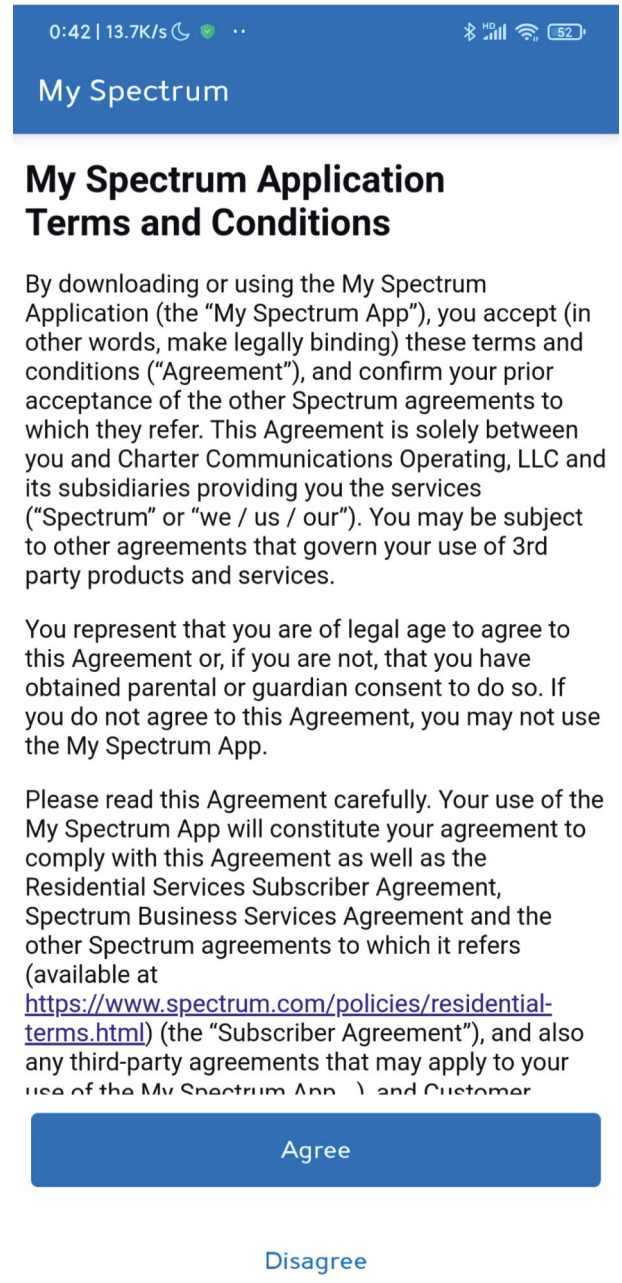

Figure 8: User consent interface in My Spectrum.

Appstore sample explicitly displayed a consent solicitation page, while only 6/569 apps in the Google Play Store did. The GDPR clearly requires that apps need to seek the user's privacy consent when opening the screen, so the fact that the user's consent has not been obtained when the collection or processing of private data has already begun is very serious.

Privacy Data Destination Domains. Among the privacy breaches we detected, we identified 208 domain names that were used as private data sending destination domains, and we named these domains private data domains (PDD). Since there are multiple subdomains

under the registrable domain names, we used the public domain name suffix list [33] to further resolve these domain names into registrable domains. In the end, 158 apps in the Tencent Appstore contacted 81 PDDs, and in the Google Play Store, 151 apps contacted 75 PDDs. The results show there are situations where multiple apps access the same PDD. Here, we speculated that third-party libraries and third-party services are causing this phenomenon. We assumed that if a PDD appears in at least five apps and these five apps are not from the same developer, this PDD was a domain name accessed by a third-party service or a third-party library. Through the above hypothesis screening, we found that 24 PDDs were the target domains of 426 DLs, accounting for 68% of the times private data was sent without user consent. This result shows that only a very small number of first parties collect private data. In contrast, most of the private data is leaked or may be leaked to third parties because developers rely heavily on third-party services for various needs, such as personalized push, analytics services, and social networking. Figure 9 shows the top five third-party service domains for two stores. Among them, qq.com and branch.io are the domains with the highest number of data breaches in the Tencent Appstore and Google Play Store in Europe, respectively, and the companies they belong to are both well-known third-party service providers.

6.4. About Third-Party Service. In Section 6.3, we listed the top five third-party service domains for each of the two stores. In Section 3.1, we mentioned that the Specification and GDPR require apps to clearly state in their privacy policies which third parties they share information with, to explain the types and ways in which information is shared and to list the privacy policies of third-party services. Our integrity analysis indicated that 88% of the apps that used these services had a privacy policy that said so. Still, data breaches are widespread. For this, we ask the following two questions and try to find the answer.

How can users assert their rights when third-party services are involved? When reading and analyzing the developer documents of third-party services, we noticed that these privacy policies all mention how users manage their own information and claim the rights of information subjects. However, we found that, for different third-party services, the difficulty for users to assert the rights of information subjects was different. Thanks to having their own clients, users of QQ open SDK [34] and Paypal [35] can manage the information collected by other apps through these third-party services by logging into the corresponding clients. For example, the user can log in to the QQ [36] app, disable the QQ service connected to other apps in the settings, and delete the data. The provider branch gives users a variety of ways to delete data and exit the branch service in the privacy policy and also provides corresponding functions to developers through the SDK, but it cannot control whether the app developer makes corresponding settings in the app. mParticle [37], to which mparticle.com belongs, and recommends that users implement the rights of data subjects through e-mail contact. Umeng [38], to which umeng.com and umsns.com belong, recommends users log in to its official website to process personal data, but we did not find a suitable entry. It is not difficult to see that, at the current stage, it is difficult for users to protect their rights against data violations involving third-party services.

Is solicitation of user consent for third-party services solely the responsibility of the third-party developer? We analyzed the privacy policies of the top three third-party services for two stores (listed in Figure 9), which are QQ, Umeng, JPush [39], branch [40], mParticle, and Apptentive [41]. It was found that these third-party services all required developers to list the use of the service in the app's privacy policy, and the user's consent to the app's privacy policy was deemed to be consent to the third-party service's privacy policy. Therefore, once the user agrees to the privacy policy of the app, the third-party service can start collecting or using the user's private data. Among the six third-party services we analyzed, only Umeng required developers to perform a delayed initialization configuration in its own compliance document [42] to ensure that users agreed to the app's privacy policy before the Umeng SDK was initialized by the app. Therefore, we concluded that it is the responsibility of both the app developer and the third-party service that the third-party service starts to collect the user's private information before the user's consent. The app developers did not adequately delay initialization, and the third-party services did not clearly indicate the relevant compliance requirements to the app developers in documents similar to the "Service Access Guide" (the above Umeng's reminder was not in the service access documentation).

6.5. Summary and Recommendations. From the above analysis of the results, it is not difficult to find that popular apps in the Tencent Appstore perform better than those in the Google Play Store in terms of privacy policy writing and user consent. Especially in terms of the user consent solicitation, only 84 apps out of 1,000 samples in the Google Play Store prompt apps to review and agree to their privacy policy summary and privacy policy text before starting their regular services (including registration and login services). There are probably two reasons for this. The first is that Specification was introduced later than the GDPR, and several studies [43, 44] have indicated that China has taken the GDPR into account in writing its data security regulations and has refined many of the elements of the GDPR to suit its national context. This makes the Specification a better reference for both application developers and data security regulators. The second is that, in recent years, China has continued its massive regulation of data security for apps.

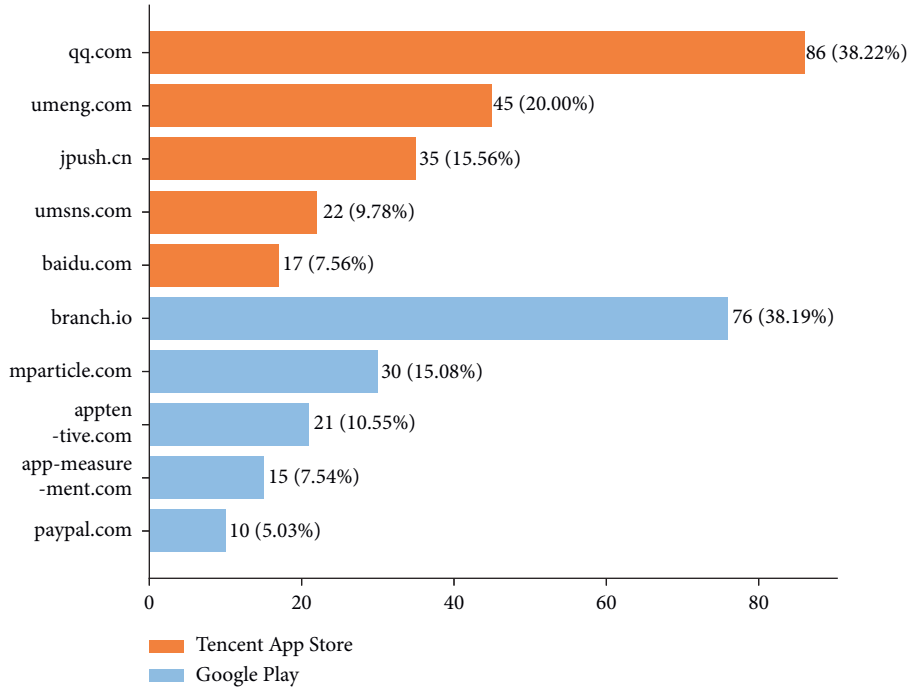

Figure 9: Third-party service domains for each of the two stores.

Multiple meetings and briefings [45] have prompted app developers to write more compliant app privacy policies and write apps with fewer data security concerns.

However, although the app privacy policy notification of the Tencent Appstore is better, the collection of private data before the user's consent has not been reduced because of this. As described in Section 6.3, this is mainly caused by third-party libraries and third-party services. In the statistics of third-party libraries and third-party services, the privacy behavior of third-party service providers in China accounts for a very high proportion. At a time when the Android ecosystem is highly globalized, it is difficult for most apps to avoid using various third-party services due to functional requirements or performance requirements. Privacy violations of third-party services may cause apps to violate their applicable data protection laws.

Through the detection of privacy policies and app samples, we evaluated the privacy policy compliance of popular apps in two app stores based on their applicable regulations. Overall, the privacy policy compliance of the Tencent Appstore app is better than that of the Google Play Store.

Based on the above analysis, we make the following recommendations to app developers:

(1) Present the privacy policy in text form. Presenting it in image or PDF format may cause reading difficulties for users.

(2) If the applicable regulations or specification documents provide a privacy policy template, the privacy policy should be written according to the template; if there is no template, one can try to learn the privacy policy of the top app from among the popular apps or the app publishing platform itself.

(3) When using a third-party service, carefully read the privacy policy or compliance document of the service to understand the privacy requirements of the third-party service for app developers and avoid potential privacy issues. At the same time, when it comes to third-party services, we hope that developers will try their best to provide users with a convenient way to claim data rights.

We recommend that the Google Play Store conduct more rigorous testing of the apps' privacy policies, including detecting the presentation of the privacy policies and the normative nature of the apps to solicit user consent. An automated evaluation process should be established in order to analyze whether app privacy policies comply with the legal requirements of the region. The store can then ask developers to improve their privacy policies based on the analysis results. In this way, privacy violations can be reduced.

7. Discussion

7.1. Poor Presentation of Privacy Policy. We mentioned in Section 3.2 that the average length of the privacy policy of the app is more than 2,000 words (more than 8,000 Chinese characters). In the first half of 2021, US consumers used an average of 46 apps per month [46]; if consumers needed to fully understand the privacy policies of these apps, they may need to read more than 100,000 words. Our analysis also found that each app uses at least two third-party libraries, and as required by law, the app's privacy policy must include a link to the privacy policies of these third-party services. The privacy policies of some large third-party services also contain links to the privacy policies of the third-party services they use. Some

service providers, such as DingTalk [47], do not write special privacy policies for their own SDKs or third-party services. As a consequence, the developer can only quote the privacy policy of the service provider, and the third-party privacy policy links used in it may be as many as 20. Take Douyin [48], the most downloaded app in the Tencent Appstore, as an example. To fully understand the privacy policy of the app and its use and indirect use of third-party services, users need to read at least 42 privacy policies, and the number of words is as high as 579,000 (Chinese characters), which is unbearable for consumers. It can be seen that developers of various apps and third-party services should also think about how to serve their customers more reasonably when writing a complete privacy policy in accordance with the law.

7.2. Responsibilities of Third-Party Services. Third-party libraries and third-party services often choose to hide technical details in a black-box manner to protect their code, which makes it difficult for developers to understand how the app works and to discover privacy behaviors before users agree. In the current globalization of the Android ecosystem, an excellent third-party service may be used by app developers from all over the world. How to avoid violating the privacy protection laws of multiple regions is a test for third-party service developers. Third-party service developers urgently need to make changes to provide user consent interfaces to app developers or directly display consent dialog windows to users. For those third parties who play the role of data processors, it is the responsibility of the developer team and enterprises to faithfully provide data processing services without "crossing the border" of the identity of the data controller.

7.3. Limitations in Our Method. Our approach naturally suffers from certain limitations. First, we do not have access to information beyond what is stated in the privacy policy. This is for two reasons. One is that it is difficult to fully trigger all processes of the app and automate registration and login, which is an inherent challenge for dynamic analysis. The other is that, as described in Section 7.1, the text of the privacy policy contains complex links that create multiple layers of references and nesting, making it difficult to obtain the full privacy policy and therefore to conduct privacy data cross-analysis.

Second, we use the Xposed framework to dynamically analyze the apps, which cannot resist the antidetection methods of the apps very well. The privacy behavior based on the native code is also difficult to identify through the Xposed framework. Also, since the Xposed framework only works well on Android 7.1.2, our detection may have missed some private data operations that apps only trigger on higher Android versions, which may have led to some false negatives.

Furthermore, we will have difficulty accessing some restricted domains during the test because of the policy restrictions in our region. Therefore, some common domains such as facebook.com, Google.com, etc. may be missing in the test results. Although we chose to download the sample apps directly from the Google Play Store on the European servers, this problem was also difficult to solve.

8. Conclusions

In this paper, we proposed three evaluation metrics for app privacy policy compliance, which are integrity of app privacy policy, privacy data collection before user consent, and accuracy of app privacy policy. Then, we analyzed 2,000 popular apps on the Google Play Store and Tencent Appstore to understand app privacy policy compliance with the GDPR and the Information Security Technology-Personal Information Security Specification. Compared with the previous study, this paper discusses the integrity of the privacy policy text in depth and increases the variety of privacy data in terms of analyzing consent issues and privacy policy accuracy, and also increases the scope of text scanning in general.

In our privacy policy completeness study, we found that the way regulations are written has an impact on the completeness of privacy policies. Our research points out that the Specification does a better job than the GDPR in guiding developers to specify a qualified privacy policy. Our study also found that there were 55 and 219 apps, respectively, in the Google Play Store and Tencent Appstore with poor privacy policies, 117 and 151 apps, respectively, that started to collect users' privacy data before obtaining the user's consent to the privacy policy, and 215 and 916 apps, respectively, that did not fulfil the obligation of the consent solicitation. The results reveal how the two stores' popular apps violate their respective laws. Based on our indepth analysis of the data, we found that incomplete app privacy policies are not directly linked to actual privacy violations, as a large number of privacy behavior violations are caused by popular third-party services. Finally, we put forward some suggestions for app stores and third-party service providers, respectively.

Data Availability

The program code, sample app list, and a list of candidate titles for privacy policies used to support the results of this study are hosted in our GitHub repository at the address included in the article. The intermediate data supporting the findings of this study, including the intermediate results of the privacy policy integrity analysis, and the intermediate results of the app samples dynamic analysis, are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant no. 61873069).

References

[1] https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/.
[2] https://www.android.com/.
[3] https://gdpr-info.eu/.
[4] http://www.ahstu.edu.cn/wlzx/info/1011/1478.htm.
[5] https://www.tc260.org.cn/front/postDetail.html?id=20200918200432.
[6] https://en.wikipedia.org/w/index.php?title=Privacy_policy&oldid=1066039196.
[7] R. Slavin, X. Wang, M. B. Hosseini et al., "Toward a framework for detecting privacy policy violations in android application code," in *Proceedings of the 38th International Conference on Software Engineering*, ACM, Austin Texas, May 2016.
[8] T. T. Nguyen, M. Backes, N. Marnau, and B. Stock, "Share first, ask later (or never?)-Studying Violations of GDPR's Explicit Consent in Android Apps," in *Proceedings of the USENIX Security Symposium*, 2021.
[9] https://firebasestorage.googleapis.com/v0/b/spacecleaner-16568.appspot.com/o/privacy_policy.html?alt=media.
[10] https://sj.qq.com/.
[11] L. Yu, X. Luo, X. Liu, and T. Zhang, "Can we trust the privacy policies of android apps?" in *Proceedings of the 2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, IEEE, Toulouse, France, June 2016.
[12] X. Wang, X. Qin, M. B. Hosseini, R. Slavin, T. D. Breaux, and J. Niu, "GUILeak: tracing privacy policy claims on user input data for Android applications," in *Proceedings of the 40th International Conference on Software Engineering*, ACM, Gothenburg Sweden, May 2018.
[13] S. Kununka, N. Mehandjiev, and P. Sampaio, "A Comparative Study of Android and IOS mobile Applications' Data Handling Practices versus Compliance to Privacy policy," in *proceedings of the IFIP International Summer School on Privacy and Identity Management*, pp. 301–313, Springer, Italy, June 2018.
[14] L. Yu, X. Luo, C. Qian, and S. Wang, "Revisiting the Description-To-Behavior Fidelity in Android applications," in *Proceedings of the 2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, pp. 415–426, IEEE, Osaka, Japan, March 2016.
[15] P. Ferrara and F. Spoto, "Static analysis for GDPR compliance,".
[16] N. Momen, M. Hatamian, and L. Fritsch, "Did app privacy improve after the GDPR?" *IEEE Security & Privacy*, vol. 17, no. 6, pp. 10–20, 2019.
[17] M. Fan, L. Yu, S. Chen et al., "An Empirical Evaluation of GDPR Compliance Violations in Android mHealth apps," in *Proceedings of the 2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*, pp. 253–264, IEEE, Coimbra, Portugal, October 2020.
[18] D. S. Guamán, J. M. Del Alamo, and J. C. Caiza, "GDPR compliance assessment for cross-border personal data transfers in android apps," *IEEE Access*, vol. 9, pp. 15961–15982, 2021.
[19] D. S. Guamán, X. Ferrer, J. M. Del Alamo, and J. Such, "Automating the GDPR Compliance Assessment for Cross-Border Personal Data Transfers in Android Applications," 2021, https://arxiv.org/abs/2103.07297.
[20] http://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=4568F276E0F8346EB0FBA097AA0CE05E.
[21] https://app.diandian.com/rank/googleplay/11-4-0-2-0?time=1655136000000&timetype=custom.
[22] http://url2io.applinzi.com/.
[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," 2018, https://arxiv.org/abs/1810.04805.
[24] Y. Cui, W. Che, T. Liu et al., "Pre-training with Whole Word Masking for Chinese bert," 2019, https://arxiv.org/abs/1906.08101?amp=1.
[25] I. Turc, M.-W. Chang, K. Lee, and K. Toutanova, "Well-Read Students Learn Better: On the Importance of Pre-training Compact Models," https://arxiv.org/abs/1908.08962.
[26] https://repo.xposed.info/module/de.robv.android.xposed.installer.
[27] https://mitmproxy.org/.
[28] https://developer.android.com/training/testing/ui-automator.
[29] X. Wang, A. Continella, Y. Yang, Y. He, and S. Zhu, "Leakdoctor: toward automatically diagnosing privacy leaks in mobile applications," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, pp. 1–25, 2019.
[30] https://support.google.com/googleplay/android-developer/answer/10144311.
[31] https://gdpr.eu/article-4-definitions/.
[32] https://play.google.com/store/apps/details?id=com.brighthouse.mybhn&hl=zh_CN&gl=US.
[33] https://Ko-Zu. publicsuffixlist[M]. [2022-01-27].
[34] https://wikinew.open.qq.com/%23/home.
[35] https://developer.paypal.com/home.
[36] https://im.qq.com/mobileqq/.
[37] https://www.mparticle.com/.
[38] https://www.umeng.com/.
[39] https://docs.jiguang.cn/compliance_guide/app_compliance_guide/app_compliance_guide1.
[40] https://branch.io/zh/.
[41] https://www.apptentive.com/.
[42] https://developer.umeng.com/docs/147377/detail/214876.
[43] https://www.allbrightlaw.com/CN/10475/5e82fe8147bccd41.aspx.
[44] https://www.secrss.com/article/11900.
[45] https://www.miit.gov.cn/jgsj/xgj/fwjd/index.html.
[46] U. S. Consumers, *Used an Average of 46 Apps Each Month in the First Half of 2021*, USA, https://sensortower.com/blog/apps-used-per-us-smartphone.
[47] https://terms.alicdn.com/legal-agreement/terms/suit_bu1_dingtalk/suit_bu1_dingtalk202010070946_49604.html.
[48] https://sf3-cdn-tos.douyinstatic.com/obj/ies-hotsoon-draft/douyin_agreement/privacy.html.