Front. Eng. Manag. 2022, 9(4): 640–652
https://doi.org/10.1007/s42524-022-0228-y

RESEARCH ARTICLE

Zitong LI, Zhuoya FAN, Junxu LIU, Leixia WANG, Xiaofeng MENG

Large-scale App privacy governance

Abstract Recently, the problem of mobile applications (Apps) leaking users' private information has aroused wide concern. As the number of Apps continuously increases, effective large-scale App governance is a major challenge. Currently, the government mainly filters out Apps with potential privacy problems manually. Such approach is inefficient with limited searching scope. In this regard, we propose a quantitative method to filter out problematic Apps on a large scale. We introduce Privacy Level (P-Level) to measure an App's probability of leaking privacy. P-Level is calculated on the basis of Permission-based Privacy Value (P-Privacy) and Usage-based Privacy Value (U-Privacy). The former considers App permission setting, whereas the latter considers App usage. We first illustrate the privacy value model and computation results of both values based on real-world dataset. Subsequently, we introduce the P-Level computing model. We also define the P-Level computed on our dataset as the PL standard. We analyze the distribution of average usage and number of Apps under the levels given in the PL standard, which may provoke insights into the large-scale App governance. Through P-Privacy, U-Privacy, and P-Level, potentially problematic Apps can be filtered out efficiently, thereby making up for the shortcoming of being manual.

Keywords privacy risk, Privacy Level, quantification, large-scale App governance

1 Introduction

In recent years, the privacy infringements of mobile phone applications (Apps) have aroused wide concern (Degirmenci, 2020), and the voice of strengthening App governance has become much stronger. In January 2019, the Office of the Central Cyberspace Affairs Commission, the Ministry of Industry and Information Technology, the Ministry of Public Security, and the State Administration for Market Regulation of China jointly issued the *Announcement on the special governance of illegal use of personal information by mobile phones using Apps*. This announcement decided to perform special acts toward the illegal collection and misuse of user information by Apps from three perspectives: The App privacy policy, the assessment of personal information use by mobile phones, and establishing App personal information security certification system. Since March 2019, the working group on App governance has evaluated more than 1000 of the most popular Apps and Apps with problematic behavior based on reports from netizens. In December 2019, four departments of China jointly published the *Identification of Apps' illegal collection and use of personal information*. App governance has attracted extensive attention. However, at present, it is not effective enough as governments mainly browse and filter Apps manually. Specifically, they first identified a number of Apps with possible problems through user complaints and manual searching, and then checked what user information is collected by the Apps and how it is being used. Finally, they notified the offending Apps for rectification. Such method has two limitations.

First, the sampling method is only applicable to a small scale. Currently, the government mainly identifies potentially problematic Apps by looking into user complaints and focusing on commonly used Apps, which only points to a few Apps. After the final round of filtering, the number of illegal Apps and rectification requirements even decreases. According to incomplete statistics in 2019, nearly 4 million Apps can be found on Internet App stores. However, only 31000 Apps were detected, and 3129 clues of violation behavior were investigated in Cleaning Net 2019 Act, a special campaign launched by the Ministry of Public Security of China. In addition, in the special operation "Special action for telecom and Internet industries to improve network data security" and "Special rectification work for App infringement on users'

Received November 27, 2021; accepted July 27, 2022

Zitong LI, Zhuoya FAN, Junxu LIU, Leixia WANG, Xiaofeng MENG (✉)
School of Information, Renmin University of China, Beijing 100872, China
E-mail: xfmeng@ruc.edu.cn

This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 62172423, 91846204, and 61941121).

rights and interests" carried out by the Ministry of Industry and Information Technology of China, only 236 Apps are forced to rectify themselves (Personal Information Protection Task Force on Apps, 2019). At present, no such method is designed for large-scale App filtering. As a result, only a small part of massive Apps can be inspected.

Second, the efficiency of manual checking is low. When evaluating and analyzing how Apps gather and use user information, the manual way is more accurate, but the cost of time and manpower is also high. Considering that many Apps with potential illegal behavior may still be getting away, the need for large-scale App filtering tool is quite urgent.

Considering the two shortcomings, we propose a quantitative method to filter Apps with high probability of having privacy problems on a large scale. The proposed method is based on the App permission requests and App usage, that is, how many people is using this App. The probabilities of App having privacy problems are quantified into Permission-based Privacy Value (P-Privacy) and Usage-based Privacy Value (U-Privacy). Having calculated these two values, Apps can be divided into different Privacy Levels (P-Levels). Here, we present one kind of level standard, namely, PL standard, based on real-world dataset, and its correctness is also verified. In our result, Apps at higher levels under PL standard are more likely to have potential privacy problems. When inspecting large-scale Apps, P-Level, or in our case, the PL standard, may serve as reference for potentially problematic Apps. This method has two advantages.

First, the searching scope is expanded. Through programmatic calculation, P-Levels can be calculated at one time for a large number of Apps. Considering that the process is automatic, people can go through several Apps at once.

Second, the efficiency is improved, the time cost is greatly reduced, and early manual research and user report are not needed because the computation of P-Level is automated. Authorities can focus on Apps at high P-Levels with more specific targets, which can save time and manpower. Besides, our method is also adaptive, which means when the dataset changes, new P-Privacy, U-Privacy, and P-Level can be computed rapidly.

This paper has three main contributions:

(1) Permission-based Privacy Value and Usage-based Privacy Value are proposed to measure an App's probabilities of having privacy problems.

(2) A kind of P-Level standard (PL standard) that provides a straightforward and practical tool for large-scale filtering over Apps is proposed. Defining a practical standard for P-Level is concerning. Here, we compute on real dataset and explore different methods, to make the standard consistent with reality. To our best knowledge, we are the first to study how to determine the levels of a large-scale App privacy regulation.

(3) The proposed method is verified with large-scale dataset from the real world. We show the correctness of PL standard by using the data released from the government. The feasibility of adopting this method in App governance is discussed, and an analysis for App privacy in Chinese mobile market is also presented (Section 5).

The subsequent structure of this paper is presented as follows. Section 2 introduces the work related to privacy risk assessment and classification methods. Section 3 mainly clarifies how to compute App privacy values and presents our results on large real datasets. Section 4 focuses on P-Level computing model and defines PL standard, and Section 5 analyses the distribution of the number of Apps and App usage under PL standard and App categories, respectively, thereby providing insights into large-scale App governance.

2 Related work

Our work is derived from privacy analysis. Considering that Apps may leak users' private information during use, many works are recently focusing on privacy analysis, which aims at evaluating Apps' potential probability of such leakage. According to Meng et al. (2019), privacy analysis methods for Apps could be mainly grouped into privacy policy analysis, static code analysis, dynamic analysis, and permission analysis.

Static code analysis focuses on App code reviewing to discover privacy problems. For example, Son et al. (2021) proposed a privacy estimation approach that considers how much the personal data usage pattern of a certain App differs from those of Apps with the same functions. They parsed target App code searching for functions or constants used to collect user information. Singh et al. (2019) used graphs to represent data flow among different application programming interfaces (APIs) to find malicious Apps. Zhang et al. (2020) clustered Apps according to their APIs, and the outliers outside the clusters were considered risky. Dynamic analysis refers to observing the information flow while using an App to evaluate the potential risk. Hayes et al. (2020) employed dynamic analysis in his research by considering network connections during an App's usage when the connections might do harm to users' private information.

Permission analysis mainly assesses privacy based on the permissions requested by Apps. After the users' approval of certain requests, an App can obtain access to critical operations, such as short message service (SMS) and getting location, which may leak private information. Peng et al. (2012) detected potentially harmful Apps by using probabilistic tools based on requested permissions, as well as Apps' categories, whereas Felt et al. (2011) identified malicious Apps by inspecting whether the App requires unnecessary permissions. Wu et al. (2021) used deep learning to determine the relationship between Apps' introduction and their requested permissions. Users can avoid installing malicious Apps if they know whether the Apps' introduction is consistent with their actually requested permissions beforehand. In addition, Chia et al. (2012) considered user ratings and external community ratings to assess Apps' privacy risk better. Biswas et al. (2016) presented SDroid to provide the optimal permission management based on an end users' opinion. They assessed the requested permissions, especially the over-claimed ones, and asked users to grant permissions to Apps selectively. In addition, 3P Framework (Biswas et al., 2017) enables the users to have greater control about granting permissions while installing Apps. This framework sits between the Kernel and the Android application packages (APKs), which provides the minimum required permissions for the App to work.

We propose a highly efficient privacy quantification method based on permission analysis, which quantifies the App privacy by checking the permissions requested at runtime. The work most similar to ours is DroidRisk (Wang et al., 2013), which assesses security risk based on permission request patterns from benign Apps and malware. The major difference is that DroidRisk considers a malware's existing probability. However, it is not applicable to large App datasets because of its complexity. Although we have a more detailed division of sensitive permissions, we focus on implementation for large-scale analysis.

Having calculated the privacy values, we explore different measures to decide P-Levels. Reasonable P-Level can not only help people intuitively judge whether the privacy value is high or not, but also provide reference for App filtering. We use data classification methods to determine P-Level boundaries. At present, two types of classification methods can be applied in data privacy and security field, one is access control-oriented, and the other is data range-oriented.

In access-control oriented classification, data are categorized into different levels based on their nature, and only users with specific identities can access the data of certain levels. This approach requires a deep understanding of the data semantics. Judgment toward the value of the data by users, data collectors, and even third parties should also be considered. Related work in this field includes Information Security and Privacy Classification (ISPC), which divides personal information into four categories: High, medium, low, and unclassified. Unclassified information can be seen by public without additional protection. Low-level information is usually only available to employees and approved non-employees. Medium-level information is only accessible to a specific group of employees. High-level information is highly confidential and accessible only to designated individuals. The classification considers the sensitivity of the data itself and controls the visible range of the data. Facebook's privacy level identification (Grauschopf, 2020) is a similar exercise in controlling the visibility of data. It allows users to set the visibility, thereby dividing the data into five categories: Visible to all, indirect friends, direct friends, certain people, and only themselves.

In data range-oriented classification, the data are usually numerical, and thus can be divided into different classes according to the range of the value. There are two common classification methods: Equal division and clustering. Hu (2007) used the equal division to grade the risk value of the quantified information system and classified the normalized risk value into five levels of "very low", "low", "medium", "high", and "very high", which correspond to the risk value range of [0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), and [0.8, 1.0], respectively. Lu et al. (2014) quantified privacy universality and confidentiality in the Internet of Things scenario and proposed Privacy Information Security Classification (PISC) model adopting clustering, with the calculation result divided into four levels, namely, "low", "medium", "high", and "very high". Among the two classification methods, equal division is simple and direct, whereas the clustering considers more about the statistical characteristics of data. Both methods have their own strengths.

In access control-oriented classification, the data types are relatively diverse, and thus profound understanding and description of the data are required. The data range-oriented classification is mainly for numerical data, and the values falling in a certain range are grouped as a level. The rules are simple and easy to understand. In our case, both privacy values are numerical data within a certain range, so we adopted data range-oriented classification.

3 App privacy value

App privacy refers to the probability of user personal information disclosure caused by Apps. The probability can be quantified from two perspectives: The permission setting and the App usage. In this paper, we propose privacy value, which can be calculated given a set of Apps, their requested permissions, and usage information.

We first introduce Permission-based Privacy Value (P-Privacy) and Usage-based Privacy Value (U-Privacy) and then define P-Level based on the two values in Section 4. P-Privacy is calculated from the permission setting of an App. It considers the situation of a single user. That is, once the App is installed, regardless of how many users it has, the new user will face the privacy risk of P-Privacy. U-Privacy combines the usage of the App with permission setting and considers the accumulating privacy of the App on its whole user group. P-Privacy and U-Privacy depict privacy from different perspectives, and connections and differences between them are observed. High P-Privacy means that more permissions related to users' personal information are requested during use, whereas the U-Privacy is not necessarily high because the total usage can be low. Furthermore, an App high in U-Privacy may not actually acquire many important permissions but collects a large amount of user information because of its large user group.

3.1 Permission-based Privacy Value (P-Privacy)

Definition 1. Permission sensitivity (Zhu et al., 2021). This value is used to measure the damage to user privacy caused by obtaining a permission.

This variable reflects the sensitivity of user information contained in a single permission. In Meng et al. (2019), permission sensitivity is divided into four levels (1, 2, 3, 4). Different permissions are assigned corresponding levels according to the problems that may arise from the information obtained by permissions and the difficulty of resolving them. The higher the sensitivity of a permission is, the greater harm it may cause due to its disclosure (Meng et al., 2019).

According to the Android Development Manual, to control access to restricted data (e.g., system status, and contact information) and the execution of restricted operations (e.g., connecting to paired devices, and recording audio), Android sets several permissions to support the protection of user privacy. Android API 30 version currently has 182 permissions, which are classified into normal, dangerous, signature, and special permissions. Each type specifies the restricted data that an App can access and the restricted operations that it can perform after being granted the permissions. Among the 182 permissions, 34 are dangerous permissions and 46 are signature permissions. Considering the category of permissions and the frequency of permissions being obtained, we focus on reading calendar, writing calendar, reading contact list, writing contact list, and 39 other sensitive permissions (Meng et al., 2019) and analyze App privacy on a large scale.

Definition 2. Permission-based Privacy Value (P-Privacy). This value represents the privacy risk inherent in an App's permission settings, regardless of the size of the App's user group.

To calculate P-Privacy, for each sensitive permission i, whose sensitivity is $s_i \in \{1, 2, 3, 4\}$, $p_i \in \{0, 1\}$ indicates whether this permission is requested. For an App, its P-Privacy is

$$P\text{-}Privacy = \textstyle\sum_{i=1}^{l} p_i \times s_i \,, \tag{1}$$

where l is the total number of sensitive permissions. The result will eventually be normalized to the interval [0, 1] using sigmoid function $\frac{1}{1+e^{-\alpha+w}}$, where α is the original P-Privacy level and w is a positive float used to adjust the mapping interval. Here, we set w to the median of P-Privacy levels of all Apps in the dataset.

P-Privacy considers privacy issues in the worst case, assuming that the App will definitely leak user data after obtaining permission. It is calculated on the basis of 39 privacy permissions and pre-set privacy sensitivity.

3.2 Usage-based Privacy Value (U-Privacy)

Definition 3. Usage-based Privacy Value (U-Privacy). This value represents the privacy risk caused by App usage beside permission setting. Except for 39 privacy permissions and pre-set privacy sensitivity, App usage is also included during calculation. Apps with wide usage will have higher value accordingly.

Notably, the maximum usage of all Apps in the dataset is set as $U_{\max}$. For one App, its usage amount is denoted as U, then its U-Privacy is

$$U\text{-}Privacy = \frac{|\ln(U)| \times P\text{-}Privacy}{|\ln(U_{\max})|}. \tag{2}$$

Similarly, the values will be normalized to [0, 1] with sigmoid function.

3.3 Analysis

In this section, we first compute privacy value using real-world dataset according to the method illustrated in Sections 3.1 and 3.2, and then we analyze the generated results.

The dataset in our experiment includes user behavior data obtained from the stratified sampling of different administrative regions in China and App data obtained from major App stores. The user behavior data are provided by cooperative companies, and thus is not publicly available. The App data are crawled from online App stores. To collect App data, we developed a crawler to extract the name, category, and the requested permissions of each App automatically from webpages.

We calculate P-Privacy and U-Privacy of each App in the dataset, showing P-Privacy and U-Privacy distribution and different privacy value distribution on App categories for verification.

3.3.1 Dataset

The dataset used for App privacy computing consists of two parts. One is the user behavior data, which includes the user log of installing or uninstalling the App, and the other is the App data composed of App information crawled from the network.

The sample users are from the stratified sampling data of prefecture-level administrative regions in China. The sampling proportion of population in each prefecture-level city is 2%–3%. The total size of the user behavior dataset is 36722417 or approximately 36.7 million.

The user behavior data are the event log data of the user installing, maintaining, and uninstalling the App, including the device ID, App package name, App installation status (uninstalling, maintaining, and installing), and report time. All data are desensitized. The data format is shown in Table 1.

App data refers to the App information obtained from third-party application stores. Specifically, we crawled App data from Wandoujia, Yingyongbao, and Xiaomi App store in December 2020. The App information includes App name, category, developer, version, and requested permission. The total size of this dataset is 406053 or approximately 400000.

3.3.2 P-Privacy and U-Privacy distribution

Figure 1 shows the distribution of the calculated P-Privacy and U-Privacy. For each point in the figure, the x coordinate represents P-Privacy, the y coordinate represents U-Privacy, and the color brightness of the point represents the number of App that corresponds to P-Privacy and U-Privacy.

From Fig. 1, we can infer that as for all Apps in the dataset, both privacy values of the majority are low. Most Apps are concentrated in the range of [0.3, 0.4] and [0, 0.05], whereas Apps with high privacy values are the minority. The average for P-Privacy and U-Privacy is 0.49 and 0.12, respectively. This value can reflect that App privacy problem in China is not that severe overall.

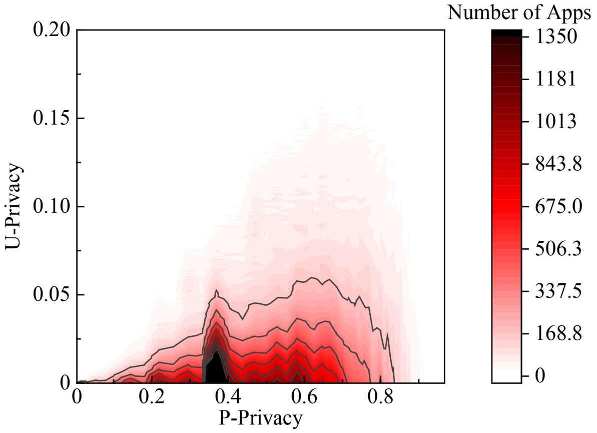

Fig. 1 Density of P-Privacy and U-Privacy distribution.

The Pearson correlation coefficient of P-Privacy and U-Privacy is 0.53, which indicates a correlation between P-Privacy and U-Privacy. However, this correlation is not strong. The proposal of U-Privacy is necessary because P-Privacy only considers permissions requested, whereas U-Privacy considers the permission and the usage. The correlation coefficient shows that the App usage, as another factor independent of the permissions requested, has a great impact when measuring the privacy of different Apps.

3.3.3 Verification

Figures 2 and 3 show the distribution of P-Privacy and U-Privacy in different App categories, respectively, and we can also compare the calculated result with reality for verification. In Figs. 2 and 3, each boxplot shows the App privacy values of the corresponding category. The distribution of the bottom and the top horizontal line means the minimum and maximum of privacy values, respectively. The top and bottom edges of the box figure refer to quartiles, middle horizontal line refers to the median, and the black spot is the average.

According to Figs. 2 and 3, the three categories with the highest average P-Privacy are health, socializing, and shopping. Likewise, three categories with the highest average U-Privacy are security, health, and shopping. Among them, the P-Privacy of security Apps is not high, but the U-Privacy of them is at the top of the list, which is consistent with widespread installation and use of security Apps.

By comparing Figs. 2 and 3, we can see that the P-Privacy of various types of App have a relatively uniform upper limit, which is basically between 0.9 and 1.0, while the lower limit is uneven. U-Privacy has a uniform lower limit, which is close to 0, while the upper limit varies greatly. This finding shows that only from the perspective of obtaining permissions of Apps, all kinds of Apps have obtained more sensitive permissions, even those that do not need too much user personal information for normal use. For U-Privacy, if the number of users is considered, there are Apps with very few users in each category. As

Table 1 User behavior data format

Attribute name	Description	Data type	Data length
Device ID	Device identification number	varchar	200
Package name	App installation package name	varchar	200
Status	App installation status: “0” for “uninstall”, “1” for “maintaining”, “2” for “newly installed”	int	/
install_type	Installation type: “−1” for “unknown type”, “0” for “normal application”, “1” for “system application”, “2” for “upgraded system application”, “3” for “pre-installed application”	int	/
Time	First report time or installation time	varchar	100
last_reported	Latest report time. If Status is zero, then this is uninstallation time accordingly	varchar	100

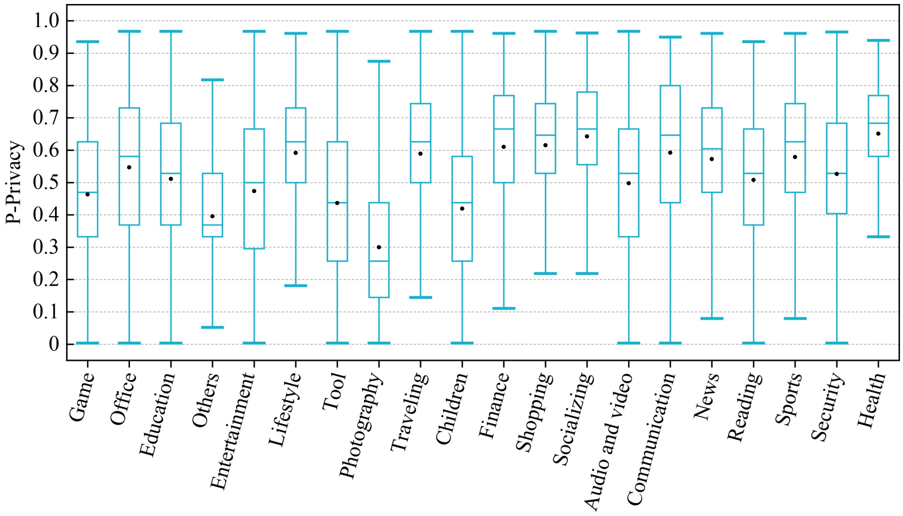

Fig. 2 P-Privacy distribution for different App categories.

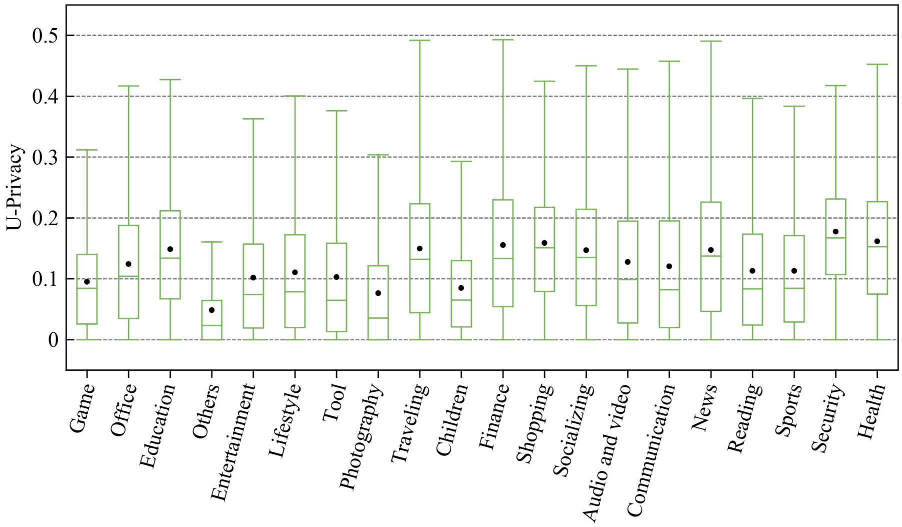

Fig. 3 U-Privacy distribution for different App categories.

the positions of boxes in the boxplot are generally low, it shows that most Apps have few users in each category. These results also correspond to reality.

4 App Privacy Level

Section 3 quantifies App privacy to a certain range. Here, we propose Privacy Level (P-Level) to provide reference for App governance better. We continue by using the dataset in Section 3.3.1 to compute for the P-Level.

4.1 Privacy Level

Definition 4. P-Level is a number within a certain range that can reflect the degree of privacy value. In our settings, P-Level has k possible integer values and ranges from 1 to k.

With the P-Level result, people can focus on Apps at higher levels and act more purposefully. Privacy values can be divided into different levels in many ways. However, it is the key to define the level standard that is consistent with and can be applied to reality. We use the real dataset in Section 3.3.1 to ensure level standard practicality and explore the two methods mentioned in Section 2 to determine the level standard: Equal division and clustering.

(1) **Equal division.** This method adopts the idea of equalization to divide the data interval into k parts. The interval length is 1, so it can be divided into k intervals with length of $1/k$. Except that the k-th interval range is $[1-1/k, 1]$, other i intervals have the range of $[(i-1)\times(1/k), i\times(1/k))$, $i\in\{1, 2, ..., k-1\}$. By observing that an App's P-Privacy and U-Privacy falls into which interval, we can find its corresponding P-Level.

(2) **Clustering.** Clustering is an improvement on equalization. It is statistical analysis technique that divides objects into relatively homogeneous groups. Numerical-oriented cluster analysis methods include K-means and K-mediods clustering (Zhang and Zhou, 2019). Between them, K-means clustering has low spatial and temporal computation complexity, and thus is suitable for processing large-scale and low-dimensional data. Consequently, K-means clustering is applied in this paper.

After clustering the existing privacy values into k clusters, the maximum or minimum of each cluster can be taken as the P-Level boundary, which is the same form as equal division. By comparing an App's P-Privacy or U-Privacy along with the calculated dividing boundaries, we can find the App's corresponding P-Level.

4.2 PL standard

We propose two kinds of P-Level standard by equal division and clustering. The equal division has been clarified in Section 4.1. For clustering, implementing details will be shown below.

We employ K-means to determine the P-Level boundary. To compare with the equal division, k is set to 10 in K-means clustering. Given two dimensions of privacy values, we consider two kinds of clustering: One is to take the (P-Privacy, U-Privacy) of each App as its corresponding coordinates to conduct clustering in two-dimensional space, and the other is to use one-dimensional clustering for P-Privacy and U-Privacy. In the latter case, an App has two P-Levels for P-Privacy and U-Privacy.

Having calculated P-Privacy and U-Privacy in Section 3.3, we first attempted to use the two-dimensional clustering. Initial cluster centers for clustering are randomly selected. After K-means clustering, the (P-Privacy, U-Privacy) distribution of each cluster is shown in Fig. 4. As shown in Fig. 4, in two-dimensional clustering, the boundaries between different clusters are difficult to define. Therefore, this clustering method is less practical.

For the second clustering method, P-Privacy and U-Privacy are divided into 10 clusters, and the minimum value in each cluster is taken as the level boundary. For example, for 10 clusters, given the i-th cluster ranges from $lower_i$ to $upper_i$, then we choose $lower_1$, $lower_2$, ..., $lower_9$ to separate privacy values. The computed P-Level results are shown in Table 2.

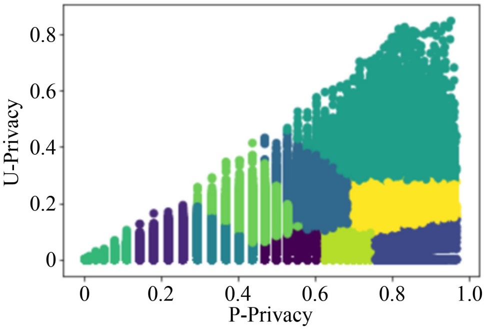

Fig. 4 Two-dimensional clustering results.

Compared with two-dimensional clustering, one-dimensional clustering has clear level boundaries. As such, it is more applicable in real life. In the subsequent part of this paper, we only use one-dimensional clustering P-Level results, which are referred as the PL standard.

Definition 5. PL standard is the P-Level standard given in this paper, which is specified in Table 2. Although there are many other ways to determine P-Level standards, we especially define the standard presented here as PL standard. It is calculated by one-dimensional clustering privacy values on the dataset in Section 3.3.1. The following analysis in Section 5 is also mainly based on this standard.

Compared with the manual approach, our method used to decide P-Level standard is not only effective and more applicable but also adaptive to large-scale scenario. We can easily adapt this method to different datasets by rerun clustering. However, handling different datasets manually requires much more time.

4.3 Verification

The App P-Level can be used to filter problematic Apps more efficiently, thereby making up for the disadvantages of manual work. This subsection tends to verify the correctness of PL standard based on the real-world data released by the government by comparing high-level Apps and Apps noted by the government.

The Ministry of Industry and Information Technology of China has published Problematic App List in 2020, and Cyberspace Administration of China (2021a) once sent out notification as 84 Apps misuse users' information. We summarize the involved Apps and calculate the appearing frequency of Apps at high levels to verify the correctness of our standard.

Table 2 PL standard for P-Privacy and U-Privacy, respectively

P-Level	Range	
	P-Privacy	U-Privacy
1	[0.0000, 0.1450)	[0.0000, 0.0198)
2	[0.1450, 0.2571)	[0.0198, 0.0590)
3	[0.2571, 0.3326)	[0.0590, 0.0994)
4	[0.3326, 0.4377)	[0.0994, 0.1395)
5	[0.4377, 0.5000)	[0.1395, 0.1819)
6	[0.5000, 0.5806)	[0.1819, 0.2303)
7	[0.5806, 0.6467)	[0.2303, 0.2889)
8	[0.6467, 0.7161)	[0.2889, 0.3663)
9	[0.7161, 0.8000)	[0.3663, 0.4902)
10	[0.8000, 1.0000]	[0.4902, 1.0000]

A total of 308 Apps are presented in Section 3.3.1 dataset and in the list released by the Ministry of Industry and Information Technology of China. We calculate the proportion of Apps whose level is larger than 5 to larger than 9 in the list, and the results are shown in Table 3.

Among these Apps, the proportion of Apps at high levels to all Apps in the list is regarded as the recognition rate. If regarding those whose P-Privacy or U-Privacy levels are higher than or equal to 7 as the Apps at high P-Levels, then the recognition rate would reach 80%.

In addition, in May 2021, the Cyberspace Administration of China, in accordance with the law and relevant regulations, inspected the collection and use of personal information of some popular Apps. They inspected Apps, such as security management or online lending, and notified a list of Apps that had illegally collected or used personal information (Cyberspace Administration of China, 2021b). In this list, 54 Apps were included in the dataset in Section 3.3.1. We also calculated the recognition rates for different P-Levels under equal division and clustering, respectively. The results are shown in Table 4.

Table 3 Proportion of Apps with different P-Privacy and U-Privacy levels in Cyberspace Administration of China (2021a)

High P-Level Apps proportion		Equal division	Clustering
P-Privacy level	⩾ 5	93.07%	94.72%
	⩾ 6	88.12%	90.10%
	⩾ 7	73.93%	**80.20%**
	⩾ 8	48.51%	66.67%
	⩾ 9	18.48%	45.87%
U-Privacy level	⩾ 5	46.86%	94.72%
	⩾ 6	19.47%	92.08%
	⩾ 7	7.26%	**83.17%**
	⩾ 8	1.32%	74.92%
	⩾ 9	0.08%	60.73%

Table 4 Proportion of Apps with different P-Privacy and U-Privacy levels in Cyberspace Administration of China (2021b)

High P-Level Apps proportion		Equal division	Clustering
P-Privacy level	⩾ 5	87.04%	87.04%
	⩾ 6	83.33%	83.33%
	⩾ 7	68.52%	**72.22%**
	⩾ 8	55.56%	64.81%
	⩾ 9	33.33%	53.70%
U-Privacy level	⩾ 5	44.44%	77.78%
	⩾ 6	27.78%	72.22%
	⩾ 7	14.81%	**70.37%**
	⩾ 8	5.56%	57.41%
	⩾ 9	0.00%	50.00%

The recognition rate in the clustering can reach 70% if the level higher than 7 is regarded as the high level. According to Tables 3 and 4, the overall recognition rate of P-Privacy is higher than that of U-Privacy in equal division, whereas the rate of U-Privacy is higher than P-Privacy in clustering in Table 3. Comparing the two computing methods, the overall recognition rate of clustering is higher than that of equal division.

According to the two lists released by the government, we calculated the proportion of Apps at high P-Levels for verification. The results demonstrate the correctness of the proposed method. When trying to filter out potential problematic Apps, people can refer to P-Level to expand the scope and improve efficiency, as well as accuracy.

5 Observation on App PL standard

The PL standard proposed in Section 4 can serve as an efficient tool in APP governance. First, it can be calculated efficiently while considering millions of Apps for one time. Second, by looking into features of Apps at high levels, people can put regulations more purposely. This section provides an overall view on large-scale App distribution based on PL standard. We begin from the App statistics of PL standard and then further explore privacy value, App usage, and category distribution on PL standard to provide more insights.

5.1 PL standard statistics

Figures 5 and 6 show the number of Apps in each P-Level under clustering and equal division, respectively. For P-Privacy, the standard deviation of App number in clustering and equal division is 9900.6 and 19783.0, respectively. For U-Privacy, the standard deviation of App number in clustering and equal division is 15690.1 and 42729.1, respectively. Although the equal division is simple and direct, it does not consider the characteristics of data distribution. The clustering considers the statistical characteristics of the data; hence, the results are more uniform. Therefore, in the following analysis, we focus on the clustering results, that is, the PL standard defined in Section 4.

If P-Level is applied to select potentially problematic Apps, then the number of Apps at different levels should be considered. When determining the lower bound to define high P-Level Apps, the lower the bottom level is, the more Apps will be above that level, thereby widening inspecting scope. Thus, when applying P-Level to filter Apps, the number of Apps and the recognition rate should be considered comprehensively.

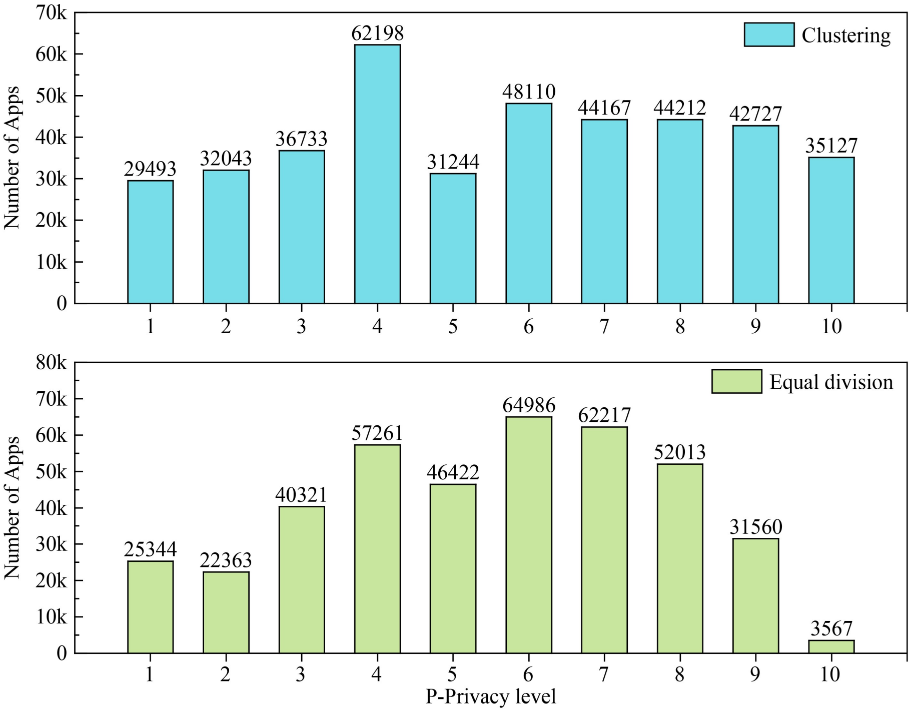

Fig. 5 Distribution of App number based on P-Privacy in different dividing methods.

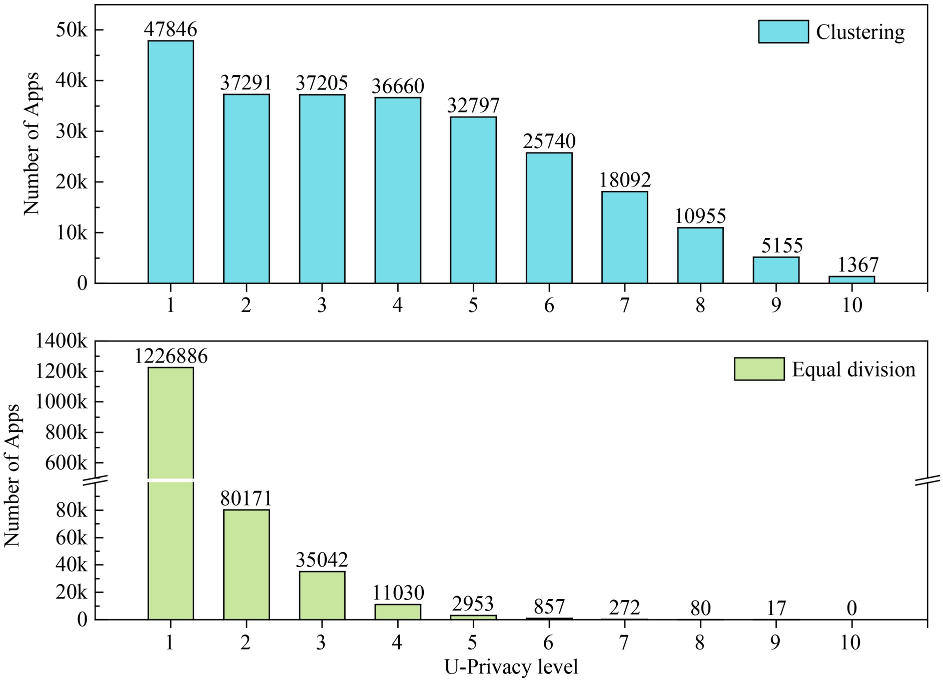

Fig. 6 Distribution of App number based on U-Privacy in different dividing methods.

5.2 Privacy value

We calculate the U-Privacy distribution in different P-Privacy levels according to PL standard, and the results were shown in Fig. 7.

According to Fig. 7, Apps with more sensitive permissions (high P-Privacy) are more likely to have wide usage (high U-Privacy). In addition, for some Apps, the impact of low usage exceeds the impact of sensitive permissions on privacy values. In Fig. 7, for different P-Privacy levels, the lower limits are basically the same, that is, they present a triangle distribution. Besides, with the increase in the P-Privacy level, the maximum of U-Privacy also rises. This observation indicates that the more sensitive permissions an App obtains, the more likely it will have wide usage. However, the lower limit of value is basically consistent, showing that there are also Apps with low U-Privacy among Apps with high P-Privacy. For these Apps, the impact of usage on their privacy values exceeds the impact of permissions.

Usage and sensitive permissions have a reflection on each other, indicating that Apps asking for more user information makes it easier to depict and predict user behavior. They can update products and promote popularization according to user preferences better, which results in increasing usage.

5.3 App usage

Figures 8 and 9 show the average usage of Apps that correspond to P-Privacy and U-Privacy levels, respectively, according to PL standard.

In Figs. 8 and 9, the number of Apps at high P-Levels is small, but the usage is large. However, although many Apps have low privacy values, their usage is less than one tenth to that of those Apps with high privacy values. This conclusion is particularly obvious in the U-Privacy levels. The average usage of Apps and its quantity show a "scissors gap". However, the situation is more moderate in P-Privacy levels. The number of mid-level (P-Privacy level 4) Apps is the largest, and high-level Apps number is significantly larger than low-level ones.

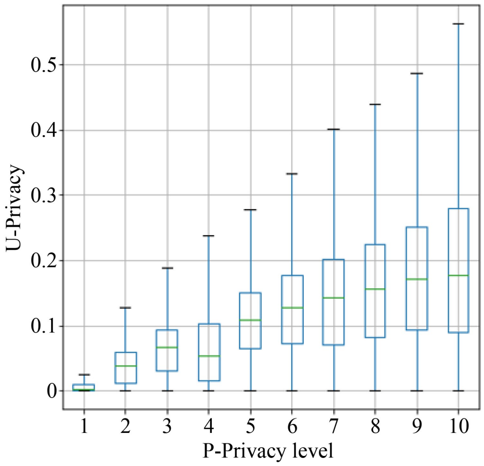

Fig. 7 Distribution of U-Privacy in different P-Privacy levels.

In addition, the level of P-Privacy and U-Privacy have the similar trend, but the difference among U-Privacy levels is more obvious. Specifically, usage increases exponentially as the level goes up. In P-Privacy, the difference between the highest and lowest usage is about 30 times, which can reach 7000 times in U-Privacy. It can be speculated that the usage is included in the U-Privacy calculation, which may be the reason why the average usage of App at high levels is much larger than that in the P-Privacy calculation.

5.4 App categories

Figures 10 and 11 show the distribution of different App categories at various levels in PL standard. In Figs. 10 and 11, the larger the bubble is, the more Apps of a certain category are under the corresponding P-Level.

According to the P-Privacy levels, lifestyle, shopping, socializing, finance, and office Apps are mostly at high levels. Photography Apps are mostly at the low levels, and other Apps are relatively evenly distributed at various levels. Vertically, Apps at low levels are mainly concentrated in photography and tool, whereas those at high levels are mainly in lifestyle and tool. In other words, tool Apps occupy a large proportion in both levels.

In terms of U-Privacy levels, horizontally viewing by category, all kinds of Apps are basically concentrated in the low levels. Lifestyle, game, and tool Apps are most prominent. Vertically, low-level Apps are dominated by lifestyle, game, and tool, whereas high-level Apps are mostly lifestyle, education, and tool Apps. Consequently, in App governance, attention should be paid to common Apps, which are more prone to privacy problems.

6 Conclusions

This paper proposes P-Level to facilitate App governance. To compute P-Level, two privacy values, namely, Permission-based Privacy Value and Usage-based Privacy Value, are presented. Then, we explore two ways of dividing privacy values into P-Levels: Equal division and clustering. According to real-world dataset results, the numbers of Apps are more evenly distributed under P-Level calculated by clustering, and it is defined as the PL standard. Besides, when verifying the correctness of P-Level on the problematic App list released by government, PL standard also shows better recognition ability. The analysis about the PL standard, App usage, and categories have also been given. Overall, the PL standard based on real-world dataset serves as an automated tool for future large-scale App governance with good efficiency, accuracy, and adaptivity.

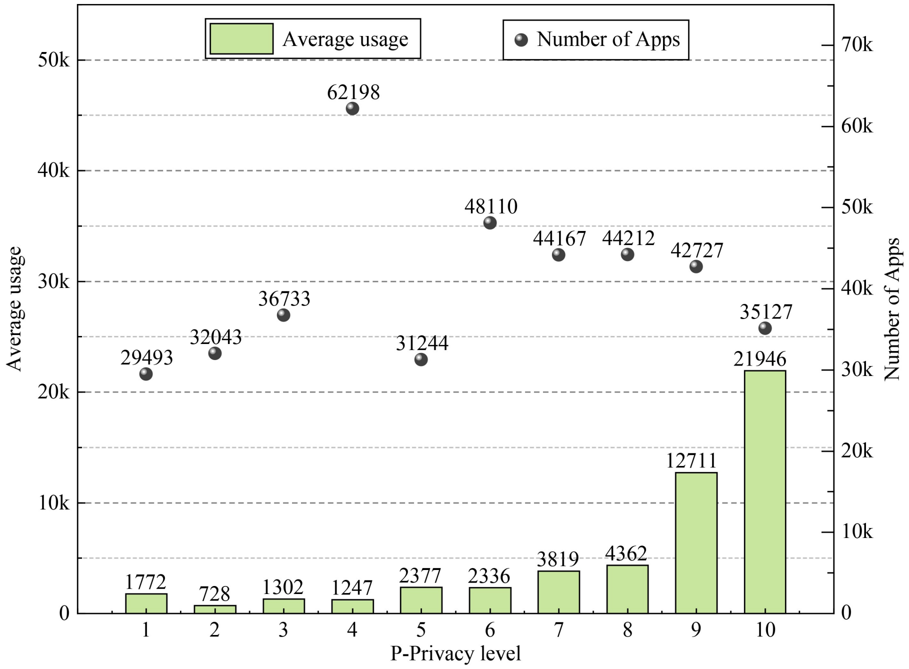

Fig. 8 Average usage of Apps in different P-Privacy levels.

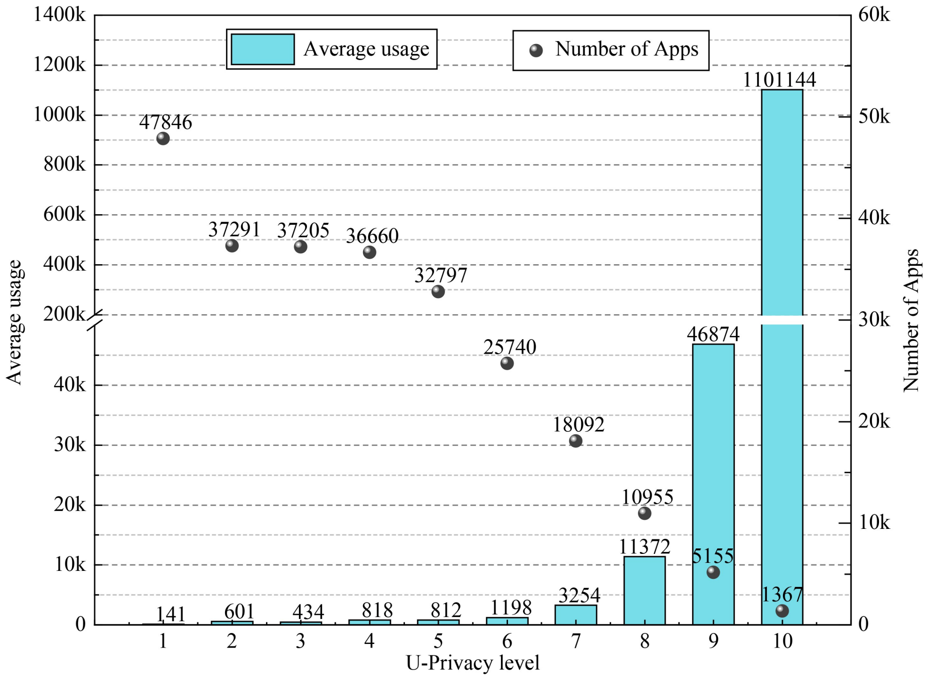

Fig. 9 Average usage of Apps in different U-Privacy levels.

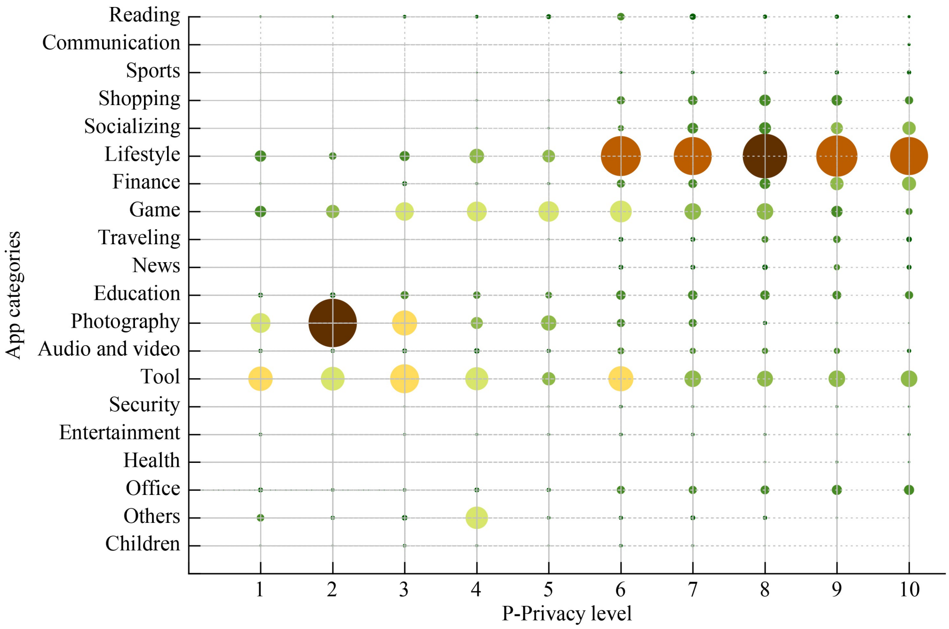

Fig. 10 App distribution in different categories under P-Privacy levels.

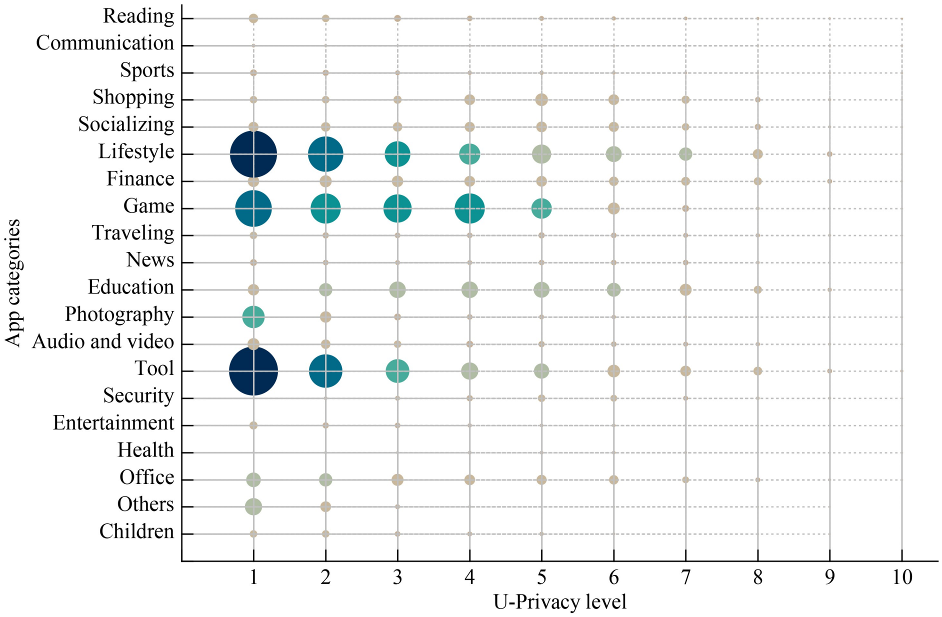

Fig. 11 App distribution in different categories under U-Privacy levels.

References

Biswas S, Wang H, Rashid J (2016). Android permissions management at App installing. International Journal of Security and Its Applications, 10(3): 223–232

Biswas S, Sharif K, Li F, Liu Y (2017). 3P framework: Customizable permission architecture for mobile applications. In: Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications. Guilin: Springer, 445–456

Chia P H, Yamamoto Y, Asokan N (2012). Is this App safe? A large scale study on application permissions and risk signals. In: Proceedings of the 21st International Conference on World Wide Web. Lyon: Association for Computing Machinery, 311–320

Cyberspace Administration of China (2021a). Notice on illegal collection and use of personal information in 84 Apps including Tencent Phone Manager (in Chinese)

Cyberspace Administration of China (2021b). Notice on illegal collection and use of personal information in 105 Apps including Tiktok (in Chinese)

Degirmenci K (2020). Mobile users' information privacy concerns and the role of App permission requests. International Journal of

Information Management, 50: 261–272

Felt A P, Chin E, Hanna S, Song D, Wagner D (2011). Android permissions demystified. In: Proceedings of the 18th ACM Conference on Computer and Communications Security. Chicago, IL: Association for Computing Machinery, 627–638

Grauschopf S (2020). Facebook privacy levels: Understanding Facebook's levels of privacy. Online Paper

Hayes D, Cappa F, Le-Khac N A (2020). An effective approach to mobile device management: Security and privacy issues associated with mobile applications. Digital Business, 1(1): 100001

Hu Y (2007). Research on Risk Assessment Method of Network Information System. Dissertation for the Doctoral Degree. Chengdu: Sichuan University (in Chinese)

Lu X, Li Q, Qu Z, Hui P (2014). Privacy information security classification study in Internet of Things. In: Proceedings of the International Conference on Identification, Information and Knowledge in the Internet of Things. Beijing: IEEE, 162–165

Meng X F, Zhu M J, Liu J X (2019). Quantitative research on privacy risk of large-scale mobile users. Journal of Information Security Research, 5(9): 778–788 (in Chinese)

Peng H, Gates C, Sarma B, Li N H, Qi Y, Potharaju R, Nita-Rotaru C, Molloy I (2012). Using probabilistic generative models for ranking risks of Android Apps. In: Proceedings of the ACM Conference on Computer and Communications Security. Raleigh North, CA: Association for Computing Machinery, 241–252

Personal Information Protection Task Force on Apps (2019). Governance report on Apps' illegal collection and use of personal information (in Chinese)

Singh A K, Jaidhar C D, Kumara M A A (2019). Experimental analysis of Android malware detection based on combinations of permissions and API-calls. Journal of Computer Virology and Hacking Techniques, 15(3): 209–218

Son H X, Carminati B, Ferrari E (2021). A risk assessment mechanism for Android Apps. In: Proceedings of the International Conference on Smart Internet of Things (SmartIoT). Jeju: IEEE, 237–244

Wang Y, Zheng J, Sun C, Mukkamala S (2013). Quantitative security risk assessment of Android permissions and applications. In: Proceedings of the 27th Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy. Newark, NJ: Springer, 226–241

Wu Z, Chen X, Lee S U J (2021). FCDP: Fidelity calculation for description-to-permissions in Android Apps. IEEE Access, 9: 1062–1075

Zhang X H, Zhang Y, Zhong M, Ding D Z, Cao Y Z, Zhang Y K, Zhang M, Yang M (2020). Enhancing state-of-the-art classifiers with API semantics to detect evolved Android malware. In: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. Association for Computing Machinery, 757–770

Zhang Y L, Zhou Y J (2019). Review of clustering algorithms. Journal of Computer Applications, 39(7): 1869–1882 (in Chinese)

Zhu M J, Ye Q Q, Meng X F, Yang X (2021). Privacy risk quantification of mobile application based on requested permissions. Scientia Sinica (Informationis), 51(7): 1100–1115 (in Chinese)