30. FEATURES OF DATA COLLECTION BY USING APPLICATIONS

Matalyga E.A.

Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus

Sinkevich L.E. – Senior Lecturer

This article provides information about what user data is and what is meant by its processing. The main technologies used and the reasons for collecting information are described.

Personal data is any information related to a particular subject. These can be names, emails, messages, geolocation, the information about the use of the application, the date and place of birth, etc. Personal data processing is any action or a set of actions performed with personal data, including collection, systematization, storage, modification, use, depersonalization, blocking and distribution. In most cases, we voluntarily consent to the processing of our personal data. It happens when people register in an online store or create an account in the application.

Absolutely all applications collect information about users. It may be seen by noticing the similarity between your query in one application and the ads appearing after a while in another application. User data is implemented in analytics, marketing, administration, etc.

The most commonly used technologies in collecting information are Software development kit (SDK) and Application programming interface (API). API is a set of protocols and tools that provide data exchange between different components of information systems. APIs can be found in operating systems, programming languages and web services. Companies design their functions and create a set of rules based on which the API can be used in other applications. The developers make use of already existing functions inside their programs which increases the speed of application development.

The SDK is a collection of tools for developing software that is compatible with a particular platform or service. The SDK may include several APIs, pieces of code, and extensive documentation. Companies create the SDK so that developers can work with individual parts of the program without researching each part of it. Data security and fault tolerance of calls to individual services are implemented through the SDK. Application developers use SDKs creating profiles of their users. SDKs themselves are not tracking programs but they help to carry out this process. Often, the information received gets not only into the application but also to third parties who can sell it to other companies.

The results of the pCloud study (Figure 1) show that more than half (52 %) of applications transfer user data to a third party. Instagram (79 %), Facebook (57 %), and LinkedIn (50 %) share the most data with third parties. Instagram shares 79 % of your data including browsing history and personal information with others online [1].

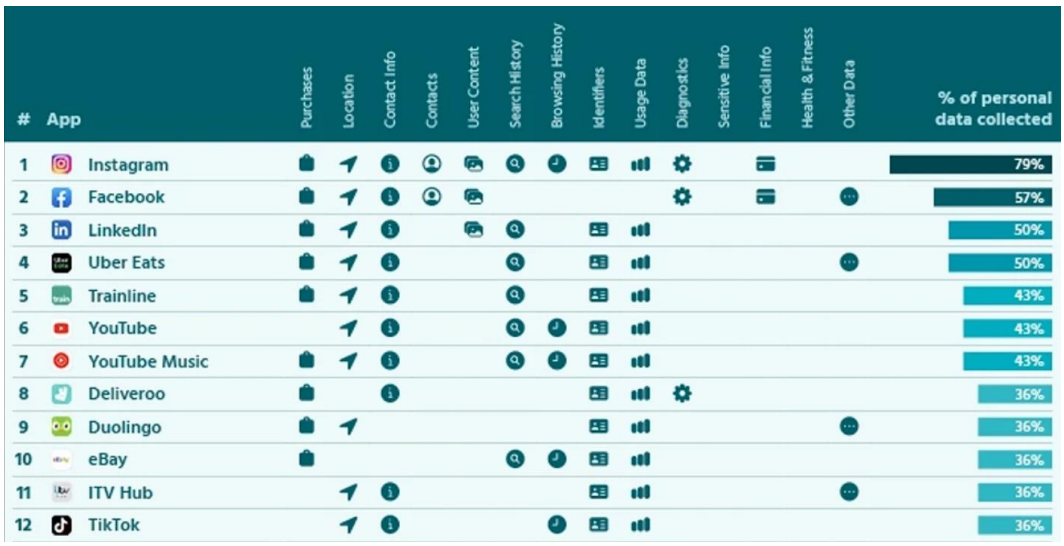

Figure 1 – The results of the pCloud study.

There are several reasons why the user information is collected. Firstly, applications collect user data as the feedback. Since most users have access to the Internet, opening any tab or clicking on a link is immediately recorded and sent to analytics. If a failure occurs while using the application, then all data of the user who has had a technical failure is analyzed in an individual order. Thus, the cause of the problem is found and eliminated.

Secondly, saving user data can speed up some applications. Such applications usually collect data location for automatic date and time settings, account data for automatic authorization, and the most typical usage scenario is to generate server usage search hints or tooltips.

Applications can be used as an advertising platform. Companies are interested in getting the greatest income from advertisements placed in their applications. To achieve it, it is necessary to make sure that the advertisements correspond to the interests of users. The data collected about users is analyzed and the most relevant advertising is provided to them. There is another advertisement technology that should be mentioned. Each user has an advertising ID and it is the same for each application. It is the use of an advertising identifier that explains the existence of similar advertising in different apps. Application companies can promote both their own products and those of other companies. The study conducted by

pCloud revealed that 80 % of apps use the collected data to market their own products in the app and beyond [1].

The collected information about users allows to improve the interface of the application. The user expects the most convenient arrangement of functions, buttons and tabs. The analysis of actions sequences supports the optimization and development of the application.

Basically companies collect information about users to improve the application and, as a result, to get more profit. However, there is some data that causes concern among users. During the coronavirus pandemic, Tectonix released an animated map illustrating the spread of the infection. It showed tourists leaving Florida beaches and returning to their homes across the country as orange dots [2]. Also, according to a TechShielder's study, 60 % of the most frequently used mobile apps in the world store information from users' personal conversations, while 80 % of them collect data about messages you send or receive, and all apps accumulate basic user information such as phone numbers and email addresses. Thus, applications can collect much more information than users think [3].

Users want to trust applications, which they share some data with. However, today the level of trust in companies is decreasing. In 2019, Apple admitted to eavesdropping on its users. A couple of months before that, Facebook was fined $5 billion for the use of confidential user data. The website of the Cambridge Analytics firm was blocked for collecting and transferring information about Facebook users to third parties. Data protection companies understand the possibility of excessive data collection and develop applications that can limit it. As an example, on April 26, 2022, the app Google Play introduced the Data safety section that provides users with information about how apps use their data [4]. Now users can see what part of the data applications is transferred to third parties, whether their data is encrypted, and what purpose applications collect data for. In order for users to have such an opportunity, application developers must fill out a questionnaire in the Data safety section. The app developer is now required to fill in the form with the information about what user data the app collects and shares with third parties.

Perhaps the only company that provides users with access to the data that it collects about them is Google. This information is available in the Activity Tracking section. It stores the history of applications and web searches, location history, and YouTube history. You can also disable advertising personalization in this section. The history of applications and web searches contains information about all the activities that you have ever performed in Google accounts. The location history section stores information about your activities, even when you do not use Google services. YouTube History stores information about what you watch and search on YouTube. These features are used to provide the most up-to-date information for the user, but they can be deactivated if the user wants to. Some users try to limit access to their personal data individually. It is only partly possible. Apple and Android provide device owners with tools to limit tracking, but these tools cannot provide complete privacy.

There are the two most likely solutions to the problem of collecting user data. It is easier for users not to use an app instead of trying to stop tracking attempts. Since any application is based on its users, companies can make a choice towards customers and limit data collection. The second option is based on advertising. According to the studies [5], the cost of placing a specialized advertisement is not profitable enough. Due to the loss of trust from users, specialized advertising for which data is mainly collected, will not be used so often.

References:

1. The most invasive apps: which apps are sharing your personal data? [Electronic resource]. ─ Mode of access: The web's most invasive apps | pCloud. ─ Date of access: 24.03.2023.

2. Want to see the true potential impact of ignoring social distancing [Electronic resource]. ─ Mode of access: https://twitter.com/tectonixgeo/status/1242628347034767361?lang=en. ─ Date of access: 24.03.2023.

3. Hacker Hotspots: The Apps Most Vulnerable to Cybercrime [Electronic resource]. ─ Mode of access: https://techshielder.com/hacker-hotspots-most-vulnerable-apps. ─ Date of access: 24.03.2023.

4. Get more information about your apps in Google Play [Electronic resource]. ─ Mode of access: https://blog.google/products/google-play/data-safety/. ─ Date of access 24.03.2022.

5. Behavioral Ad Targeting Not Paying Off for Publishers, Study Suggests [Electronic resource]. ─. Mode of access: https://www.wsj.com/articles/behavioral-ad-targeting-not-paying-off-for-publishers-study-suggests-11559167195. ─ Date of access: 24.03.2023.