Challenge title: Consistency between the privacy policies of third-party SDKs disclosed in large real-world apps and their actual behaviors

Context and background: One of the major challenges faced by mobile apps is to ensure the consistency between the privacy policies of third-party SDKs disclosed in large real-world mobile apps and their actual behaviors. Relevant laws and regulations (e.g., CCPA/TC260) require vendors to disclose personal information shared with third-party SDKs in their own privacy policies, and even provide detailed information on the scope of SDK usage and personal information processing. This requires vendors of large real-world apps to accurately and completely disclose in their privacy policies the personal information accessed by or shared to the third-party SDKs.

Third-party SDKs (e.g., AppsFlyer) offer extensive functionality, including user data analysis, which can expedite app development. However, these SDKs may also collect personal information (e.g., IMEI), and transmit it to their servers. Privacy policies [3] are legal documents designed to prevent the misuse of personal information. To protect the user's rights from any unauthorized encroachment, providers of third-party SDKs have a responsibility to accurately disclose the personal information they collect in their privacy policies. Unfortunately, existing research [1] shows that more than 23% of third-party SDKs lack privacy policies and 37% fail to disclose personal information in their privacy policies, let alone do accurately and completely. Relevant laws and regulations (e.g., CCPA/TC260) require that the privacy policy of the app should accurately and completely disclose all user personal information collected by third-party SDKs in a separate section or even a separate document. If the personal information collected by third-party SDKs is inconsistent with the disclosure scope of the app, especially if it exceeds the disclosure scope of the app, the app will face the risk of regulatory penalty or even being removed from the app store.

Previous work [2],[3] has proposed methods to analyze privacy issues of apps, such as analyzing the consistency between apps' privacy policy and behaviors. Zhao et al. [1] has conducted research on the consistency between privacy policies and behaviors of third-party SDKs in apps, such as analyzing 458 SDKs and conducting small-scale consistency studies using static analysis techniques. However, due to the imprecision of static analysis and the small number of SDKs considered, it is challenging for existing solutions to accurately and completely analyze the consistency between the privacy policies of third-party SDKs disclosed in large real-world apps and their actual behaviors. To meet the requirement of vendors of large apps, we need an accurate analysis approach (e.g., a dynamic analysis which can provide witness disclosure) that can precisely and completely obtain the personal information collection of third-party SDKs. Furthermore, it can cover all third-party SDKs used in large apps and all personal information.

Due to some regulations of the enterprise, the dataset of collected third-party SDKs is not provided in this abstract.

Relevant key technical challenges:

We are faced with the following concrete problems and challenges, try to solve them as thoroughly as possible:

1. *Collect as many SDKs as possible to maximize the coverage of the commonly used SDKs in mainstream apps.*
2. *Third-party SDKs often use Java reflection to obtain personal information. How to detect third-party SDKs that use the reflection mechanism to collect user information?*
3. *How to address the efficiency and scalability problem of static analysis for large real-world apps?*
4. *How to extract personal information from the privacy policies of third-party SDKs in apps?*
5. *How to verify whether the personal information is really collected by the third-party SDKs at runtime?*

Expected technical requirements: Ensure that the automated analysis is applicable to the mainstream large apps, with high precision and recall, and is efficient enough (e.g., less than 30min per app for consistency analysis).

REFERENCES

[1] K. Zhao, X. Zhan, L. Yu, S. Zhou, H. Zhou, X. Luo, H. Wang, Y. Liu, "Demystifying Privacy Policy of Third-Party Libraries in Mobile Apps" (ICSE'23), https://arxiv.org/pdf/2301.12348.pdf

[2] X. Zhang, X. Wang, R. Slavin, T. Breaux, and J. Niu, "How does misconfiguration of analytic services compromise mobile privacy?" in 2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE). IEEE, 2020, pp. 1572–1583

[3] L. Yu, X. Luo, J. Chen, H. Zhou, T. Zhang, H. Chang, and H. K. N. Leung, "PPChecker: Towards accessing the trustworthiness of android apps' privacy policies," IEEE Transactions on Software Engineering, 2021, 47(2): 221-242.