Understanding How Third-Party Libraries in Mobile Apps Affect Responses to Subject Access Requests

Nikita Samarin* and Primal Wijesekera*†
*University of California, Berkeley and †International Computer Science Institute

I. Introduction and Background

Mobile app developers rely on Application Programming Interfaces (APIs)—a set of functionalities made available in the form of *third-party libraries* (TPLs)—to deliver essential and non-essential functionality [9]. Unfortunately, the decreased development costs associated with TPLs often come at the expense of the privacy and security of consumers [4], whose personal information can be inconspicuously accessed and transmitted by TPLs. Comprehensive privacy laws, such as the GDPR or CPRA, impose certain obligations on software developers who collect personal information directly or indirectly, including through the usage of TPLs.

Prior research has demonstrated that app developers often lack an understanding of the TPLs they integrate and how to configure them [1]. The lack of visibility and adequate control of privacy-threatening behaviors of TPLs often leads to privacy harm to users. It could expose developers to increased risks of non-compliance with data protection regulations, which require companies to disclose their data collection and sharing practices and respond to consumers' requests to access the personal information held by the company.

Given the difficulties of keeping track of privacy-relevant behaviors of TPLs, we hypothesized that **an app developer is less likely to disclose the personal information collected by TPLs in response to a consumer's subject access request, as opposed to the personal information that the developer collects themselves**. In this proposal, we present our preliminary results that offer support for our hypothesis.

II. Methodology

We selected 160 top-ranked Android mobile apps to analyze, which we downloaded together with their privacy policies in November 2021. To determine which privacy policies contained instructions for submitting a verifiable consumer request (VCR) under the California Consumer Privacy Act (CCPA), two researchers from our team independently labeled the text of each policy. Our analysis indicated that out of the selected 160 apps, 109 (68%) included CCPA-specific disclosures in their privacy policies (with Krippendorff's alpha = 0.81, indicating an acceptable level of inter-rater agreement [6]).

App Testing. We manually tested the selected 109 apps using phones with our instrumented version of Android 9 that monitored resource accesses and all network traffic, regardless of the use of TLS. (Prior work has applied the same tool [1], [2], [5], [8], [10].) We set up each test phone—to be used by an individual tester in California—to use its pseudonymous identifiers, such as the phone number, email address, usernames, device identifiers, and other types of information. We categorized each observed destination domain as either first- or third-party using the same approach as in [11].

Verifiable Consumer Requests. The CCPA recognizes that a first party can collect personal information directly or *indirectly* and requires the developer (i.e., the first party) to disclose any personal information it has collected about the user, including through or by a service provider or contractor. [3]. For each tested app, a California resident submitted a VCR to the app developer, requesting specific pieces of personal information accessed, collected, and shared by the app. Out of the 109 apps, we received our data in 80 cases.

Ethics. The IRB at our institution determined that this study does not meet the legal definition of human subjects research, as it involves the examination of institutional processes and does not collect data *about* individuals [7].

III. Preliminary Results

To test our hypotheses, we used a two-sample z-test. We compared the proportions of personal information disclosed in response to our VCRs that was: 1) collected directly by the developer and 2) collected indirectly by a third-party library.

In the 582 flows, we observed transmissions only to third-party domains in 178 cases, of which only 20 (11%) were disclosed to us in response to our VCRs. We observed transmissions to first- and third-party domains in the remaining 404 cases, of which 266 (66%) were disclosed. The difference between these proportions was statistically highly significant ($p < 0.001$) supporting our stated hypothesis.

IV. Discussion and Future Work

Our initial results support our hypothesis that app developers are less likely to disclose the personal information collected by TPLs in response to a VCR, than the personal information that the developer collects themselves. Given the growing list of privacy regimes focusing on data governance practices and making developers accountable for all the resource access during an app execution, third-party code execution poses a grave threat to consumers and app developers.

In our future work, we want to correctly attribute data flows that originate in the code of the TPL as opposed to the developer's codebase. We also want to understand how TPL vendors can offer greater transparency of their information practices and enable developers to comply with subject access rights. These technical and regulatory questions must be answered to make the mobile app ecosystem more privacy-conscious.

Acknowledgements

This work was supported by the U.S. National Science Foundation (under grant CNS-1817248), the National Security Agency (under contract H98230-18-D-0006), the Center for Long-Term Cybersecurity (CLTC) at U.C. Berkeley, and by CITRIS and the Banatao Institute at the University of California.

References

[1] Noura Alomar and Serge Egelman. Developers say the darnedest things: Privacy compliance processes followed by developers of child-directed apps. *Proceedings on Privacy Enhancing Technologies*, 2022.

[2] Benjamin Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Serge Egelman. Actions speak louder than words:Entity-Sensitive privacy policy and data flow analysis with PoliCheck. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 985–1002, 2020.

[3] California Consumer Privacy Act. Cal. Civ. Code §1798.130(a)(3)(A). https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5, 2018. Accessed: 2022-01-28.

[4] Álvaro Feal, Julien Gamba, Juan Tapiador, Primal Wijesekera, Joel Reardon, Serge Egelman, and Narseo Vallina-Rodriguez. Don't accept candy from strangers: An analysis of third-party mobile sdks. *Data Protection and Privacy*, 13(13):1, 2021.

[5] Catherine Han, Irwin Reyes, Álvaro Feal, Joel Reardon, Primal Wijesekera, Narseo Vallina-Rodriguez, Amit Elazar, Kenneth A Bamberger, and Serge Egelman. The price is (not) right: Comparing privacy in free and paid apps. *Proceedings on Privacy Enhancing Technologies*, 2020(3), 2020.

[6] Klaus Krippendorff. Computing krippendorff's alpha-reliability. 2011.

[7] Office of Human Research Protections. What is human subjects research?, 2022. https://www.hhs.gov/ohrp/sites/default/files/OHRP-HHS-Learning-Module-Lesson2.pdf.

[8] Joel Reardon, Álvaro Feal, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, and Serge Egelman. 50 ways to leak your data: An exploration of apps' circumvention of the android permissions system. In *28th USENIX security symposium (USENIX security 19)*, pages 603–620, 2019.

[9] Pasquale Salza, Fabio Palomba, Dario Di Nucci, Andrea De Lucia, and Filomena Ferrucci. Third-party libraries in mobile apps. *Empirical Software Engineering*, 25(3):2341–2377, 2020.

[10] Madelyn R. Sanfilippo, Yan Shvartzshnaider, Irwin Reyes, Helen Nissenbaum, and Serge Egelman. Disaster Privacy/Privacy Disaster. *Journal of the Association for Information Science and Technology*, 71(9):1002–1014, 2020.

[11] Rahmadi Trimananda, Hieu Le, Hao Cui, Janice Tran Ho, Anastasia Shuba, and Athina Markopoulou. Ovrseen: Auditing network traffic and privacy policies in oculus vr. *arXiv preprint arXiv:2106.05407*, 2021.