1. A Systematic Assessment on Android Third-party Library Detection Tools (*Open Access*)

Zhan, Xian (1); Liu, Tianming (2); Liu, Yepang (3); Liu, Yang (4); Li, Li (5); Wang, Haoyu (6); Luo, Xiapu (7)

Source: *IEEE Transactions on Software Engineering*, 2021; **ISSN:** 00985589, **E-ISSN:** 19393520; **DOI:** 10.1109/TSE.2021.3115506; **Publisher:** Institute of Electrical and Electronics Engineers Inc., Article in Press

Author affiliation: (1) Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China, Hong Kong, (e-mail: csxzhan@comp.polyu.edu.hk) (2) Faculty of Information Technology, Monash University, 2541 Clayton, Victoria, Australia, (e-mail: tianming.liu@monash.edu) (3) Computer Science and Engineering, Southern University of Science and Technology, 255310 Shenzhen, Guangdong, China, 518052 (e-mail: liuyp1@sustech.edu.cn) (4) School of Computer Engineering, Nanyang Technological University, Singapore, Singapore, Singapore, 639798 (e-mail: yangliu@ntu.edu.sg) (5) Faculty of Information Technology, Monash University, 2541 Clayton, Victoria, Australia, 3168 (e-mail: li.li@monash.edu) (6) School of Computer Science, Beijing University of Posts and Telecommunications, 12472 Beijing, Beijing, China, (e-mail: haoyuwang@bupt.edu.cn) (7) Department of Computing, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Hong Kong, (e-mail: luoxiapu@gmail.com)

Abstract: Third-party libraries (TPLs) have become a significant part of the Android ecosystem. Developers can employ various TPLs to facilitate their app development. Unfortunately, the popularity of TPLs also brings new security issues. For example, TPLs may carry malicious or vulnerable code, which can infect popular apps to pose threats to mobile users. Furthermore, TPL detection is essential for downstream tasks, such as vulnerabilities and malware detection. Thus, various tools have been developed to identify TPLs. However, no existing work has studied these TPL detection tools in detail, and different tools focus on different applications and techniques with performance differences. A comprehensive understanding of these tools will help us make better use of them. To this end, we conduct a comprehensive empirical study to fill the gap by evaluating and comparing all publicly available TPL detection tools based on six criteria: accuracy of TPL construction, effectiveness, efficiency, accuracy of version identification, resiliency to code obfuscation, and ease of use. Besides, we enhance these open-source tools by fixing their limitations, to improve their detection ability. Finally, we build an extensible framework that integrates all existing available TPL detection tools, providing an online service for the research community. We release the evaluation dataset and enhanced tools. According to our study, we also present the essential findings and discuss promising implications to the community; e.g., 1) Most existing TPL detection techniques more or less depend on package structure to construct in-app TPL candidates. However, using package structure as the module decoupling feature is error-prone. We hence suggest future researchers using the class dependency to substitute package structure. 2) Extracted features include richer semantic information (e.g., class dependencies) can achieve better resiliency to code obfuscation. 3) Existing tools usually have a low recall. Most existing tools cannot effectively find partial import TPLs, obfuscated TPLs, which directly limit their capability. 4) Existing tools are complementary to each other; we can build a better tool via combining the advantages of each tool. We believe our work provides a clear picture of existing TPL detection techniques and also gives a road-map for future research. IEEE

Main heading: Malware
Controlled terms: Android (operating system) - Codes (symbols) - Feature extraction - Libraries - Mobile security - Open source software - Semantics
Uncontrolled terms: Android - Code - Detection tools - Empirical studies - Features extraction - Library detection - Resilience - Third parties - Third-party library - Tool comparisons
Classification Code: 723 Computer Software, Data Handling and Applications - 723.2 Data Processing and Image Processing - 903.4.1 Libraries
Open Access type(s): All Open Access, Green
Database: Compendex
Data Provider: Engineering Village