Towards Privacy-Preserving Mobile Apps: A Balancing Act

Dengfeng Li Wing Lam Wei Yang Zhengkai Wu Xusheng Xiao Tao Xie
University of Illinois Urbana-Champaign, Case Western Reserve University
{dli46,winglam2,weiyang3,zw3,taoxie}@illinois.edu,xusheng.xiao@case.edu

ACM Reference format:
Dengfeng Li Wing Lam Wei Yang Zhengkai Wu Xusheng Xiao Tao Xie. 2017. Towards Privacy-Preserving Mobile Apps: A Balancing Act. In *Proceedings of Symposium and Bootcamp on the Science of Security (HotSoS), Hanover, Maryland USA, April 2017 (HoTSoS'17),* 1 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

As mobile apps are increasingly becoming data-driven, these apps tend to collect much app usage data to carry out their promised utilities and enhance user experiences. Unfortunately, some highly sensitive information in the data provides little or no benefit towards delivering the apps' utilities. For instance, for an app whose purpose is to show video game trailers, it is unnecessary to request and send its users' phone number and contact list to a remote server [1]. There is a strong need for a framework to help protect users' app usage data while retaining the app's utility efficacy (e.g., the number of enabled features).

There are three main challenges in realizing such framework. First, it is difficult to correctly identify security-sensitive information in the app usage data. For instance, user input text (such as "My password is 12345") can contain sensitive information, and such framework needs to understand the semantic meaning of such text in order to know whether sensitive information is present or not. Second, because utilities of apps vary dramatically, there is a need for generically applicable program analysis to measure the impact of information anonymization on the level of utility efficacy. Third, balancing privacy preservation and utility efficacy requires fine-grained analysis on privacy specification (such as a privacy policy declared by the app's developers) and the app.

To address these challenges, we propose a privacy framework that enables a mobile app's developers to determine what sensitive information can be anonymized while maintaining a desirable level of utility efficacy.

2 PROPOSED PRIVACY FRAMEWORK

We next describe the overview of our proposed privacy framework (shown in Figure 1) and its key components.

First, we detect sensitive information collected by the given app. By adapting our previous work [2], our technique leverages UI rendering, geometrical layout analysis, and natural language processing to identify input fields that may accept sensitive information. Our detection also leverages static data flow analysis to detect sensitive information (such as a GPS location) obtained from the system.

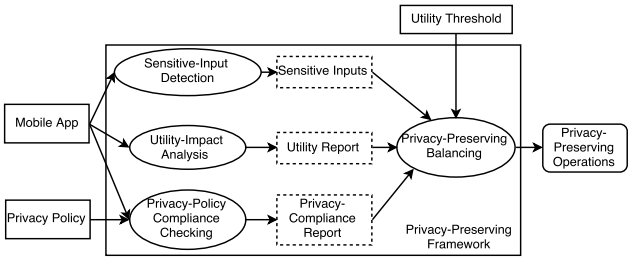


Figure 1: Proposed privacy framework

Second, to understand the impact of anonymizing a specific user input on the level of utility efficacy, we propose to anonymize each input, dynamically measure its impact on the level of utility efficacy using F-measure (the harmonic mean of precision and recall) as our metric, and produce a utility report.

Third, we aim to check whether the sensitive information collected by an app is privacy preserving, with respect to the app's declared privacy policy. We propose to conduct static data flow analysis on the app and its backend server to verify their behavior against the declared privacy policy. For instance, Sen et al. [3] proposed a data flow analysis for Map-Reduce-like big data systems to track how user data flow in the system and check compliance with its declared privacy policy. Such analysis gives users confidence that the sharing of sensitive information is legitimate and avoids privacy over-preservation that degrades the level of utility efficacy.

Finally, leveraging results from the preceding three components, we anonymize various sensitive information while assuring that the level of utility efficacy is above a user-predefined threshold. Our goal is to maximize the utility efficacy and privacy preservation by combining different levels of anonymization.

Acknowledgments. This material is based upon work supported by the Maryland Procurement Office under Contract No. H98230-14-C-0141. This work was supported in part by National Science Foundation under grants no. CCF-1409423, CNS-1434582, CNS-1513939, CNS-1564274, and a Google Faculty Research Award.

REFERENCES

[1] Carlos Castillo. 2012. Android Malware Promises Video While Stealing Contacts. (2012). https://securingtomorrow.mcafee.com/mcafee-labs/android-malware-promises-video-while-stealing-contacts/
[2] Jianjun Huang, Zhichun Li, Xusheng Xiao, Zhenyu Wu, Kangjie Lu, Xiangyu Zhang, and Guofei Jiang. 2015. SUPOR: Precise and Scalable Sensitive User Input Detection for Android Apps. In *Proc. 24th USENIX Conference on Security Symposium (USENIX Security'15)*. 977–992.
[3] Shayak Sen, Saikat Guha, Anupam Datta, Sriram K. Rajamani, Janice Tsai, and Jeannette M. Wing. 2014. Bootstrapping Privacy Compliance in Big Data Systems. In *Proc. 2014 IEEE Symposium on Security and Privacy (SP '14)*. 327–342.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
HoTSoS'17, Hanover, Maryland USA
© 2016 Copyright held by the owner/author(s). 978-x-xxxx-xxxx-x/YY/MM…$xx.xx
DOI: 10.1145/nnnnnnn.nnnnnnn