A Partial Reproduction of Malware Detection with *RevealDroid*

Xiaoqin Fu
Washington State University, Pullman, USA
xiaoqin.fu@wsu.edu

Haipeng Cai
Washington State University, Pullman, USA
haipeng.cai@wsu.edu

ABSTRACT

We summarize our partial reproduction of *RevealDroid*, one of the latest malware classifier for Android. We outline our reproduction methodology and discuss our findings.

ACM Reference Format:
Xiaoqin Fu and Haipeng Cai. 2019. A Partial Reproduction of Malware Detection with *RevealDroid*. In *Proceedings of Recognizing and Rewarding Open Science in Software Engineering (ICSE-ROSE 2019).* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/1122445.1122456

WHO

Reproduced paper: We partially produced the *RevealDroid* paper, titled *Lightweight, Obfuscation-Resilient Detection and Family Identification of Android Malware* and published in *ACM Transactions on Software Engineering and Methodology, 26(3):11, 2018* by Joshua Garcia, Mahmoud Hammad, and Sam Malek.

Reproducer: *RevealDroid* was partially reproduced by Xiaoqin Xu and Haipeng Cai as part of one of their ongoing project that is to be presented as a poster at ICSE 2019. The poster tile is *On the Deterioration of Learning-Based Malware Detectors for Android.*

WHAT

RevealDroid is a malware detection and family categorization technique for Android. Out of a larger exploratory set of features, *RevealDroid* used three classes of features eventually: counts of different categories of Android SDK APIs used, counts of calls to native code, and characterizing numbers on reflection usage.

WHY

Malware detection has been extensively studied in recent years, yet it remains an open problem. In particular, a main reason that new malware keeps emerging although numerous detectors have been developed is that a detector trained on older samples may not be able to classify newer samples accurately. This paper on *RevealDroid* not only addresses the obfuscation-induced challenges to high-accuracy malware detection, but sensibly evaluated the capabilities of the technique for detecting malware that is newer than training samples. This is intriguing and very important for Android security, hence motivating our reproduction study of *RevealDroid*.

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICSE-ROSE 2019, 25–31 May, 2019, Montreal, Canada
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

WHERE

In our reproduction study, we used the three classes of features chosen by *RevealDroid*. However, given the focus of our project on sustainable malware detection, we only reproduced the studies on malware detection, although the original paper also included studies on malware family identification. For the same reason, we only focused on the time-aware setting (i.e., training the classifier on apps whose ages were knowingly different from the ages of apps for testing), although the original study additionally considered time-agonistic setting (selecting training and testing samples without considering their age). Moreover, we extended the time-agnostic setting such that the training and testing span multiple years explicitly. In the original study, the authors split training and testing data as per the first day of a year, without considering a specific length of the spans.

HOW

First, we replicated the original study on malware detection in the time-aware setting successfully—we obtained the same results as reported originally. This established confidence about the correctness of our experimental setup. Then, we applied *RevealDroid* to a different set of 24,780 apps, including varying numbers of benign apps and malware of different years, ranging from 2010 to 2017. We trained *RevealDroid* on benign and malicious samples both from a year x, and then used the trained model to predict the labels of benign and malicious samples both from a year $x + N$, where $x \in [2010,2016]$ and $N \in [1,7]$. As a result, we conducted 28 instances of training and testing. For each instance, we computed the detection precision, recall, and F1 accuracy. Our results have been summarized in our paper, and our replication package can be found here.

DISCUSSION

The artefact of *RevealDroid* was provided by the authors as an open-source package. The package included many handy scripts for separately using different components of the tool. We found these scripts along with the open-source nature very useful as they allowed us to check the correctness of each of our reproduction steps easily. The project documentation also gives details on how to set up the tool, which greatly facilitated our reproduction. We had to change some of the original scripts because of their inclusion of hard-coded paths or environment dependencies, and we added a few new scripts in order to evaluate the technique in the extended time-aware setting. In this new setting, *RevealDroid* did not perform as well as it did in the original study, which is reasonable as it did not aim at *sustainable* detection. Nevertheless, our effort for the reproduction was minor. This process confirmed the merits of being open-source along with clear documentation and handy setup utilities for successful replication and reproduction.