Accepted Manuscript

ANDRODET: An adaptive android obfuscation detector

O. Mirzaei, J.M. de Fuentes, J. Tapiador, L. Gonzalez-Manzano

PII: S0167-739X(18)30931-2
DOI: https://doi.org/10.1016/j.future.2018.07.066
Reference: FUTURE 4384

To appear in: *Future Generation Computer Systems*

Received date : 17 April 2018
Revised date : 21 June 2018
Accepted date : 28 July 2018

Please cite this article as:, ANDRODET: An adaptive android obfuscation detector, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.07.066

AndrODet: An Adaptive Android Obfuscation Detector

O. Mirzaei*, J. M. de Fuentes, J. Tapiador, L. Gonzalez-Manzano

Computer Security Lab (COSEC). Universidad Carlos III de Madrid.
Av. Universidad, 30. ES-28911 Leganes (Spain)
omid.mirzaei@uc3m.es, {jfuentes, jestevez, lgmanzan}@inf.uc3m.es
* *Corresponding author*

Abstract

Obfuscation techniques modify an app's source (or machine) code in order to make it more difficult to analyze. This is typically applied to protect intellectual property in benign apps, or to hinder the process of extracting actionable information in the case malware. Since malware analysis often requires considerable resource investment, detecting the particular obfuscation technique used may contribute to apply the right analysis tools, thus leading to some savings.

In this paper, we propose AndrODet, a mechanism to detect three popular types of obfuscation in Android applications, namely identifier renaming, string encryption, and control flow obfuscation. AndrODet leverages online learning techniques, thus being suitable for resource-limited environments that need to operate in a continuous manner. We compare our results with a batch learning algorithm using a dataset of 34,962 apps from both malware and benign apps. Experimental results show that online learning approaches are not only able to compete with batch learning methods in terms of accuracy, but they also save significant amount of time and computational resources. Particularly, AndrODet achieves an accuracy of 92.02% for identifier renaming detection, 81.41% for string encryption detection, and 68.32% for control flow obfuscation detection, on average. Also, the overall accuracy of the system when apps might be obfuscated with more than one technique is around 80.66%.

Keywords: Obfuscation Detection, Android, Machine Learning, Malware