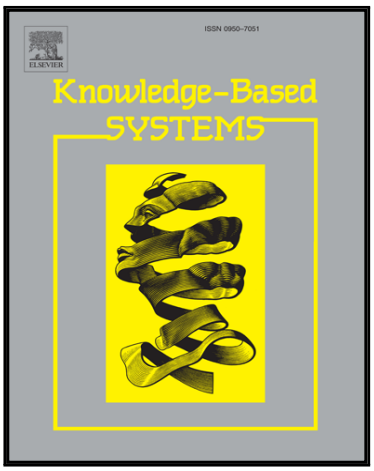

A novel pattern recognition system for detecting Android malware by analyzing suspicious boot sequences

Jorge Maestre Vidal, Marco Antonio Sotelo Monge,
Luis Javier García Villalba

PII: S0950-7051(18)30142-4
DOI: 10.1016/j.knosys.2018.03.018
Reference: KNOSYS 4267

To appear in: *Knowledge-Based Systems*

Received date: 29 November 2017
Revised date: 6 February 2018
Accepted date: 10 March 2018

Please cite this article as: Jorge Maestre Vidal, Marco Antonio Sotelo Monge, Luis Javier García Villalba, A novel pattern recognition system for detecting Android malware by analyzing suspicious boot sequences, *Knowledge-Based Systems* (2018), doi: 10.1016/j.knosys.2018.03.018

A novel pattern recognition system for detecting Android malware by analyzing suspicious boot sequences

Jorge Maestre Vidal[a,*], Marco Antonio Sotelo Monge[a,*], Luis Javier García Villalba[a,*]

[a] *Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), School of Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain*

Abstract

This paper introduces a malware detection system for smartphones based on studying the dynamic behavior of suspicious applications. The main goal is to prevent the installation of the malicious software on the victim systems. The approach focuses on identifying malware addressed against the Android platform. For that purpose, only the system calls performed during the boot process of the recently installed applications are studied. Thereby the amount of information to be considered is reduced, since only activities related with their initialization are taken into account. The proposal defines a pattern recognition system with three processing layers: monitoring, analysis and decision-making. First, in order to extract the sequences of system calls, the potentially compromised applications are executed on a safe and isolated environment. Then the analysis step generates the metrics required for decision-making. This level combines sequence alignment algorithms with bagging, which allow scoring the similarity between the extracted sequences considering their regions of greatest resemblance. At the decision-making stage, the Wilcoxon signed-rank test is implemented, which determines if the new software is labeled as legitimate or malicious. The proposal has been tested in different experiments that include an in-depth study of a particular use case, and the evaluation of its effectiveness when analyzing samples of well-known public datasets. Promising experimental results have been shown, hence demonstrating that the approach is a good complement to the strategies of the bibliography.

Keywords: anomalies, malware, mobile devices, intrusion detection, pattern recognition, sequence alignment

1. Introduction

Over recent years a significant growth in the popularity of mobile devices was observed, which was empowered by their large capacity of connectivity, accessibility, and versatility. Consequently, users in-

*Tel. +34 91 394 76 38, Fax: +34 91 394 75 47

Email addresses: jmaestre@ucm.es (Jorge Maestre Vidal), masotelo@ucm.es (Marco Antonio Sotelo Monge), javiergv@fdi.ucm.es (Luis Javier García Villalba)