Journal of Ambient Intelligence and Humanized Computing (2021) 12:1731–1755
https://doi.org/10.1007/s12652-020-02243-0

ORIGINAL RESEARCH

TriDroid: a triage and classification framework for fast detection of mobile threats in android markets

Abdelouahab Amira[1,2] · Abdelouahid Derhab[3] · ElMouatez Billah Karbab[4] · Omar Nouali[1] · Farrukh Aslam Khan[3]

Received: 27 February 2020 / Accepted: 17 June 2020 / Published online: 29 June 2020

Abstract

The Android platform is highly targeted by malware developers, which aim to infect the maximum number of mobile devices by uploading their malicious applications to different app markets. In order to keep a healthy Android ecosystem, app-markets check the maliciousness of newly submitted apps. These markets need to (a) correctly detect malicious app, and (b) speed up the detection process of the most likely dangerous applications among an overwhelming flow of submitted apps, to quickly mitigate their potential damages. To address these challenges, we propose TriDroid, a market-scale triage and classification system for Android apps. TriDroid prioritizes apps analysis according to their risk likelihood. To this end, we categorize the submitted apps as: botnet, general malware, and benign. TriDroid starts by performing a (1) Triage process, which applies a fast coarse-grained and less-accurate analysis on a continuous stream of the submitted apps to identify their corresponding queue in a three-class priority queuing system. Then, (2) the Classification process extracts fine-grained static features from the apps in the priority queue, and applies three-class machine learning classifiers to confirm with high accuracy the classification decisions of the triage process. In addition to the priority queuing model, we also propose a multi-server queuing model where the classification of each app category is run on a different server. Experiments on a dataset with more than 24K malicious and 3K benign applications show that the priority model offers a trade-off between waiting time and processing overhead, as it requires only one server compared to the multi-server model. Also it successfully prioritizes malicious apps analysis, which allows a short waiting time for dangerous applications compared to the FIFO policy.

Keywords Android security · App triage · Malware detection · Data mining · Machine learning

1 Introduction

The Android operating system is largely popular among mobile device users. According to a recent report (Statcounter 2020), Android dominated the smartphone market with a share of 74% in December 2019. Android apps are hosted in markets such as Google Play Store, the official Android app-market. Also, the number of Android apps has noticeably increased, from 1 million in July 2013 to 3.3 million in June 2018. Malware developers highly target Android OS to infect a large number of mobile devices. According to G-DATA (Gdata 2018), the number of discovered malware samples in 2018-Q3 increased over 40% compared to the same period in 2017. It discovered almost 3.2 million new Android malware samples, which represents approximately an average of 7 malware samples every second.

In order to keep a healthy Android ecosystem, app-markets deploy vetting systems to check the maliciousness of newly submitted apps. For example, Google Play (Google

✉ Abdelouahab Amira
amira@cerist.dz

Abdelouahid Derhab
abderhab@ksu.edu.sa

ElMouatez Billah Karbab
e_karbab@encs.concordia.ca

Omar Nouali
onouali@mail.cerist.dz

Farrukh Aslam Khan
fakhan@ksu.edu.sa

1 Research Center for Scientific and Technical Information (CERIST), 16000 Algiers, Algeria

2 Faculty of Exact Sciences, Université de Bejaia, 06000 Bejaia, Algeria

3 Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh, Saudi Arabia

4 Concordia University, Montreal, Canada