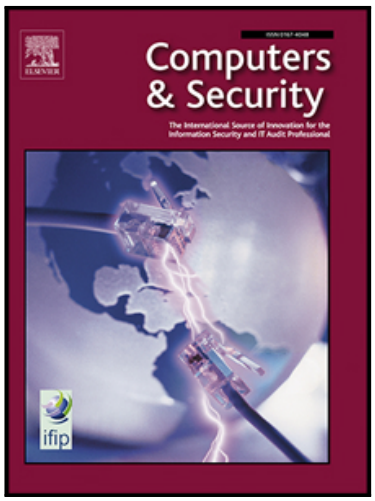

Leveraging Ontologies and Machine-learning Techniques for Malware Analysis into Android Permissions Ecosystems

Luiz C. Navarro, Alexandre K.W. Navarro, André Grégio, Anderson Rocha, Ricardo Dahab

PII: S0167-4048(18)30231-1
DOI: https://doi.org/10.1016/j.cose.2018.07.013
Reference: COSE 1375

To appear in: *Computers & Security*

Received date: 19 March 2018
Revised date: 18 June 2018
Accepted date: 31 July 2018

Please cite this article as: Luiz C. Navarro, Alexandre K.W. Navarro, André Grégio, Anderson Rocha, Ricardo Dahab, Leveraging Ontologies and Machine-learning Techniques for Malware Analysis into Android Permissions Ecosystems, *Computers & Security* (2018), doi: https://doi.org/10.1016/j.cose.2018.07.013

Leveraging Ontologies and Machine-learning Techniques for Malware Analysis into Android Permissions Ecosystems

Luiz C. Navarro[a], Alexandre K. W. Navarro[b], André Grégio[c], Anderson Rocha[a], Ricardo Dahab[a]

[a] *Institute of Computing - University of Campinas (Unicamp), Campinas, SP, Brazil*
[b] *Engineering Department - University of Cambridge - Cambridge, UK*
[c] *Department of Informatics - Federal University of Paraná (UFPR), Curitiba, PR, Brazil*

Abstract

Smartphones form a complex application ecosystem with a myriad of components, properties, and interfaces that produce an intricate relationship network. Given the intrinsic complexity of this system, we hereby propose two main contributions. First, we devise a methodology to systematically determine and analyze the complex relationship network among components, properties, and interfaces associated with the permission mechanism in Android ecosystems. Second, we investigate whether it is possible to identify characteristics shared by malware samples at this high level of abstraction that could be leveraged to unveil their presence. We propose an ontology-based framework to model the relationships between application and system elements, together with a machine-learning approach to analyze the complex network that arises therefrom. We represent the ontological model for the considered Android ecosystem with 4,570 apps through a graph with some 55,000 nodes and 120,000 edges. Experiments have shown that a classifier operating on top of this complex representation can achieve an accuracy of 88% and precision of 91% and is capable of identifying and determining 24 features that correspond to 70 important graph nodes related to malware activity, which is a remarkable feat for security.

Keywords: Malware, Android Permissions, Ontology, Bags of Graphs, Machine Learning, Discriminant Features.

1. Introduction

Smartphones have become ubiquitous computing devices worldwide. A recent Ericsson Mobility Report [1] indicated that smartphones currently represent 55% of all mobile subscriptions globally. The report further projects the number of unique mobile subscribers to reach 6.1 billion by 2022, covering roughly 75% of the world's population. Despite the multitude of different device models and the availability of several different operating systems for smartphones, the Android operating system currently holds 88% of market share [2].

Mobile devices are increasingly being used for activities that directly impact social, work, and financial environments; as such, they have become a primary target for cyber-criminals. A study published by Deloitte [3] concluded that, in the United Kingdom, the top ten usages for smartphones include social networking, emailing, banking, and shopping with similar patterns across other developed countries. To the eyes of a cyber-criminal, social

Email addresses: luiz.navarro@students.ic.unicamp.br (Luiz C. Navarro), akwn2@cam.ac.uk (Alexandre K. W. Navarro), gregio@inf.ufpr.br (André Grégio), anderson.rocha@ic.unicamp.br (Anderson Rocha), rdahab@ic.unicamp.br (Ricardo Dahab)