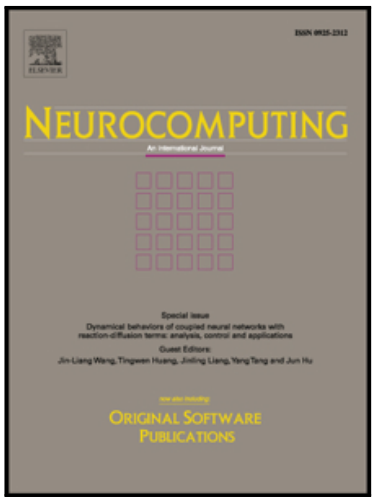

Android Malware Detection with Unbiased Confidence Guarantees

Harris Papadopoulos, Nestoras Georgiou, Charalambos Eliades, Andreas Konstantinidis

PII: S0925-2312(17)31767-8
DOI: 10.1016/j.neucom.2017.08.072
Reference: NEUCOM 19073

To appear in: *Neurocomputing*

Received date: 25 April 2017
Revised date: 17 July 2017
Accepted date: 16 August 2017

Please cite this article as: Harris Papadopoulos, Nestoras Georgiou, Charalambos Eliades, Andreas Konstantinidis, Android Malware Detection with Unbiased Confidence Guarantees, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.08.072

Android Malware Detection with Unbiased Confidence Guarantees

Harris Papadopoulos*, Nestoras Georgiou, Charalambos Eliades, Andreas Konstantinidis

Department of Computer Science and Engineering, Frederick University, Cyprus

Abstract

The impressive growth of smartphone devices in combination with the rising ubiquity of using mobile platforms for sensitive applications such as Internet banking, have triggered a rapid increase in mobile malware. In recent literature, many studies examine Machine Learning techniques, as the most promising approach for mobile malware detection, without however quantifying the uncertainty involved in their detections. In this paper, we address this problem by proposing a machine learning dynamic analysis approach that provides provably valid confidence guarantees in each malware detection. Moreover the particular guarantees hold for both the malicious and benign classes independently and are unaffected by any bias in the data. The proposed approach is based on a novel machine learning framework, called Conformal Prediction, combined with a random forests classifier. We examine its performance on a large-scale dataset collected by installing 1866 malicious and 4816 benign applications on a real android device. We make this collection of dynamic analysis data available to the research community. The obtained experimental results demonstrate the empirical validity, usefulness and unbiased nature of the outputs produced by the proposed approach.

Keywords: Malware Detection, Android, Security, Conformal Prediction, Class Imbalance, Unbiased Predictions, Confidence Measures, Confidence Guarantees, Random Forests

1. Introduction

The evolution of ubiquitous smartphone devices has given rise to great opportunities with respect to the development of applications and services spanning from simple messaging and calling applications to more sensitive financial transactions and Internet banking services. As a result, a great deal of sensitive

*Corresponding author
Email address: h.papadopoulos@frederick.ac.cy (Harris Papadopoulos)