DroidDetector: Android Malware Characterization and Detection Using Deep Learning

Fangfang Li

May 31, 2018

Abstract

Smartphones and mobile tablets are rapidly becoming indispensable in daily life. Android has been the most popular mobile operating system since 2012. However, owing to the open nature of Android, countless malwares are hidden in a large number of benign apps in Android markets that seriously threaten Android security. Deep learning is a new area of machine learning research that has gained increasing attention in artificial intelligence. In this study, we propose to associate the features from the static analysis with features from dynamic analysis of Android apps and characterize malware using deep learning techniques. We implement an online deep-learning-based Android malware detection engine (DroidDetector) that can automatically detect whether an app is a malware or not. With thousands of Android apps, we thoroughly test DroidDetector and perform an in-depth analysis on the features that deep learning essentially exploits to characterize malware. The results show that deep learning is suitable for characterizing Android malware and especially effective with the availability of more training data. DroidDetector can achieve detection accuracy, which outperforms traditional machine learning techniques. An evaluation of ten popular anti-virus softwares demonstrates the urgency of advancing our capabilities in Android malware detection.

1. Introduction

Android dramatically surpassed a billion shipments of its devices in 2014 and has remained the No.1 mobile operating system since 2013, according to a recent report from Gartner. Android markets, such as theoogle Play Store and other third-party markets, play an important role in the popularity of Android devices. However, the openness of Android makes these markets hot targets for malware attacks and causes countless instances of malware being hidden behind a large number of benign apps that seriously threatens users security and privacy. Moreover, a report from McAfee Labs reveals that 3.73 million pieces of mobile malware were identified in 2013, increasing an astounding from the end of 2012. Consequently, an urgent need arises to develop powerful solutions for Android malware detection [1]. Unfortunately, the Android market currently has no such solution.

Today, the main countermeasure to defense against malware on Android platforms is a risk communication mechanism that warns users about the permissions required before installing each app.

2. Feature Extraction

To systematically characterize Android apps (i.e., both malware and benign apps), we conduct static and dynamic analyses to extract features from each app, as shown in Fig. 1.

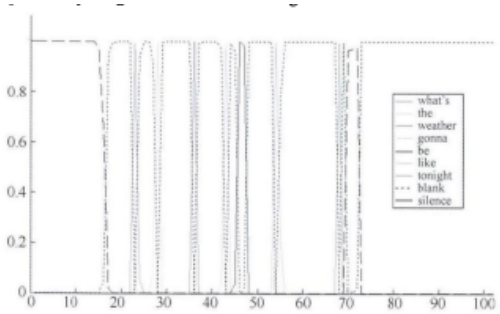

Figure 1. An example of word CTC.

All the features fall under one of three types: required permissions, sensitive APIs, and dynamic behaviors. Among them, required permissions and sensitive APIs are extracted through the static analysis, whereas dynamic behaviors are extracted through dynamic analysis. Specifically, all we need is the installation file of each Android app [2]. Therefore, the goal of people's hard work is to achieve a visual assisted driving system that has the ability to track roads on highways and avoid collisions with vehicles ahead. One point to be pointed out here is that the computer plays the role of replacing the human brain in the computer vision

system, but it does not mean that the computer must complete the visual information processing according to the human visual method. Computer Vision can and should process visual information based on the characteristics of the computer system. However, the human visual system is by far the most powerful and complete visual system known to people.

References

[1] K. Katyal, A. Alpher, and P. Burlina. Leveraging deep reinforcement learning for reaching robotic tasks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1

[2] G. Y, W. S, D. M, and X. J. Raptorx-angle: real-value prediction of protein backbone dihedral angles through a hybrid method of clustering and deep learning. *Prentice Hall Professional Technical Reference*, 19(4):100, 2018. 1