Digital Investigation 13 (2015) 22–37

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

A review on feature selection in mobile malware detection

Ali Feizollah*, Nor Badrul Anuar, Rosli Salleh, Ainuddin Wahid Abdul Wahab

Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 11 October 2014
Received in revised form 2 February 2015
Accepted 4 February 2015
Available online 13 March 2015

Keywords:
Mobile malware
Android
Feature selection
Review paper
Mobile operating system

ABSTRACT

The widespread use of mobile devices in comparison to personal computers has led to a new era of information exchange. The purchase trends of personal computers have started decreasing whereas the shipment of mobile devices is increasing. In addition, the increasing power of mobile devices along with portability characteristics has attracted the attention of users. Not only are such devices popular among users, but they are favorite targets of attackers. The number of mobile malware is rapidly on the rise with malicious activities, such as stealing users data, sending premium messages and making phone call to premium numbers that users have no knowledge. Numerous studies have developed methods to thwart such attacks. In order to develop an effective detection system, we have to select a subset of features from hundreds of available features. In this paper, we studied 100 research works published between 2010 and 2014 with the perspective of feature selection in mobile malware detection. We categorize available features into four groups, namely, static features, dynamic features, hybrid features and applications metadata. Additionally, we discuss datasets used in the recent research studies as well as analyzing evaluation measures utilized.

Introduction

The ubiquity of mobile devices is undeniable because they have brought new possibilities to every days life. Contemporary mobile devices are more powerful when compared to Personal Computers (PCs) ten years ago. Unlike PCs, portability of mobile devices makes them attractive to users. In addition, their small sizes as compared to personal computers play an important role in increasing their popularity. Furthermore, users interests are increasing towards the Rich Mobile Applications (RMA), such as Google Maps that deliver rich user experience along with high interaction (Knoernschild, 2010). However, such popularity has serious security and privacy threats and various other malicious activities. The malicious activities are hidden from the user and are committed in the background or at midnight when the user is asleep (Eslahi et al., 2012). Based on such characteristics, we assess the research works done to detect these malware.

The aim of this paper is to scrutinize various features available in Android malware, since feature selection has considerable effects on results of experiments. We discuss such effect in the following sections. Suarez-Tangil et al. (2013) discuss malware for smart devices in general. However, the paper discusses various types of features very briefly and the authors did not cover all types of available features. Similarly, La Polla et al. (2013) investigates various types of mobile devices, available malware, their effect on the devices and different detection methods. Nevertheless, they did not mention what features they used in detection, considering that features have significant impact on detection. Mohite and Sonar (2014) survey different analysis techniques in mobile malware detection. The paper

* Corresponding author.
E-mail addresses: ali.feizollah@siswa.um.edu.my (A. Feizollah), badrul@um.edu.my (N.B. Anuar), rosli_salleh@um.edu.my (R. Salleh), ainuddin@um.edu.my (A.W.A. Wahab).

http://dx.doi.org/10.1016/j.diin.2015.02.001
1742-2876/

mention examples of detection methods along their description. The paper does not include datasets and evaluation measures. In addition, it does not cover all the recent works comprehensively. Peng et al. (2014) examines evolution of mobile malware, their damages and their propagation model. They included various operating system in the paper, which makes it difficult to examine all available aspects thoroughly. However, we focus on Android operating system and the results are more accurate and comprehensive. Additionally, to the best of our knowledge, surveying Android features is unprecedented in research works.

The rest of this paper is organized as follows. Section 2 gives background information needed for the rest of the paper. Section 3 examines four types of features in mobile malware detection including the static, dynamic, hybrid and applications metadata. We comprehensively analyze each type of features. Section 4 presents discussions regarding the datasets used in the recent research works and their description. Additionally, we discuss the evaluation measures of malware detection in this section. Finally, Section 5 concludes the paper by highlighting important points.

Background

In this section, we present background information. First, we investigate the supremacy of mobile devices and see where mobile devices stand against PCs. Next, we examine how widespread mobile malware are; we then explain different types of malware, ranging from simple one to the most dangerous and sophisticated one. It is beneficial to know popularity of mobile devices, as well as, the seriousness of mobile malware. We scrutinize the importance of feature selection in malware detection in the next sub-section in order to establish the necessity of this research work. Finally, we take a closer look at Android files and their components, since we refer to various components throughout this work.

Supremacy of mobile devices

Popularity of mobile devices is on the rise (Chen and Bilton, 2014). Gartner, an American information technology research and advisory firm, reported that total shipment of mobile devices increased in 2013 by 5.9% and reached 2.35 billion devices compared to the previous year and it is estimated that the growth continues to 2.5 billion devices in 2014 (Gartner, 2013). On the other hand, the shipment of PCs has declined to 305 million units in 2013 and it is expected to decrease below 300 million units in 2014 (Gartner, 2013). Table 1 shows the number of devices shipments in 2012, 2013 and 2014.

Table 1
Worldwide devices shipments (thousands of units) (Gartner, 2013).

	2012	2013	2014
PC (Desktop and Notebook)	341,273	305,178	289,239
Ultramobile	9787	20,301	39,824
Tablet	120,203	201,825	276,178
Mobile Phone	1,746,177	1,821,193	1,901,188
Total	2,217,440	2,348,497	2,506,429

The comparison between PC and mobile devices, ultramobile, tablets, and mobile phones, reveals that the number of PCs is decreasing while the shipment of mobile devices is increasing. In terms of usage of mobile devices, Walker Sands published a report that indicated the Internet traffic pertaining to mobile devices increased. Based on the report, the Internet traffic of mobile devices represents 67% increase in the third quarter of 2013 compared to the same period in 2012 (Sands, 2013).

The rise of android malware

There are numerous mobile operating systems in the market namely, Android (Google, 2014a), iOS (Apple, 2014), Windows Phone (Microsoft, 2014), and BlackBerry (R. in Motion, 2014). Android has dominated the mobile devices industry. Based on a report, a total of 261.1 million devices were shipped in the third quarter of 2013 and 81.3% of the shipped devices were running Android operating system (CNET, 2013). Fig. 1 depicts the dominance of Android among other mobile operating systems.

The number of attacks is steadily going up for Android. Based on the report from F-Secure, Android incorporated 79% of all malware in 2012 compared to 66.7% in 2011 and just 11.25% in 2010 (Techcrunch, 2013). Similarly, Symantec said that number of Android malware increased almost four times between June 2012 and June 2013 (Symantec, 2013). In addition, the period of April 2013 to June 2013 witnessed a massive growth of almost 200% in Android malware. Fortinet (2014), a world leader in high performance network security, announced that within the period of January 1, 2013 until December 31, 2013, they discovered over 1800 new distinct families of malware and the majority of which were Android malware. Malware growth not only degrades performance of the devices, but also has posed serious concerns towards the privacy and security of data (Fortinet, 2014). In February 2014, Symantec stated that an average of 272 new malware and five new malware families are discovered every month targeting Android operating system (Symantec, 2014).

The reason of such enormous increase in Android malware lies in the fact that Android is an open source operating system (Teufl et al., 2013) and the application market of Android, known as Google Play, is not monitored meticulously in terms of security (Feizollah et al., 2013).

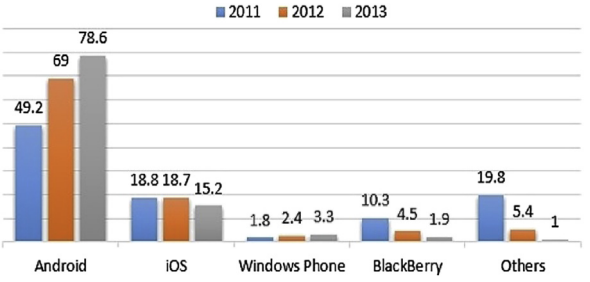

Fig. 1. Mobile operating systems market share (% of global unit shipments) (Motley, 2014).