Pervasive and Mobile Computing 24 (2015) 101–116

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Service-oriented mobile malware detection system based on mining strategies

Baojiang Cui [a,b,*], Haifeng Jin [a,b], Giuliana Carullo [c], Zheli Liu [d,e,1]

[a] School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China
[b] National Engineering Laboratory for Mobile Network Security, China
[c] Department of Computer Science, University of Salerno, Salerno, Italy
[d] Department of Computer and Information Security, College of Information Technical Science, Nankai University, Tianjin, China
[e] Key Laboratory of Network Security and Cryptology, Fujian Normal University, Fuzhou, China

ARTICLE INFO

Article history:
Available online 16 June 2015

Keywords:
Malware detection
Data mining
Mobile internet
Contraction clustering
SMMDS

ABSTRACT

The large number of mobile internet users has highlighted the importance of privacy protection. Traditional malware detection systems that run within mobile devices have numerous disadvantages, such as overconsumption of processing resources, delayed updating, and difficulty in intersection. This study proposed a novel detection system based on cloud computing and packet analysis. The system detects the malicious behavior of the mobile malwares through their packets with the use of data mining methods. This approach completely avoids the defects of traditional methods. The system is service-oriented and can be deployed by mobile operators to send alarms to users who have malwares on their devices. To improve system performance, a new clustering strategy called contraction clustering was created. This strategy uses prior knowledge to reduce dataset size. Moreover, a multi-module detection scheme was introduced to enhance system accuracy. The results of this scheme are produced by integrating the detection results of several algorithms, including Naive Bayes and Decision Tree.

1. Introduction

Mobile malwares are pervasive in mobile devices. Such malwares cause various problems, such as theft, violation of privacy, and unwanted financial charges. Traditional malware detection methods cannot effectively solve these problems because of several reasons. First, given the computing capacity and battery life limitations of a mobile device, effective detection is difficult to achieve without affecting the normal usage of a user. Second, the malware database often cannot be promptly updated and is thus vulnerable to new malwares [1]. Therefore, a new detection method for mobile malwares that runs outside the device is necessary.

Recent mobile malwares with malicious behavior, such as privilege escalation, remote control, financial charges, and personal information stealing [2], differ from traditional PC malwares which mainly aim at damage the normal function of computers. The final goal of most of the malicious behaviors of mobile malwares is stealing information. Therefore, they must function through the mobile internet to communicate with their servers [3]. From this perspective, if the packets of

* Corresponding author at: School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China.
E-mail address: cuibj@bupt.edu.cn (B. Cui).

[1] Z. Liu and B. Cui contribute to this work equally.

http://dx.doi.org/10.1016/j.pmcj.2015.06.006
1574-1192/

such malwares can be distinguished from others, these malwares could be efficiently detected, and their negative effects can be significantly reduced [4].

A large number of mobile malwares are challenging the competence of mobile operators, who are obliged to protect user privacy and enhance the network service quality [5]. Building a service-oriented mobile malware detection system (SMMDS) based on mining strategies is necessary. Such system can not only ensure the security of users privacy but can also enhance the competitiveness of a mobile operator. The system can provide services to which users can subscribe. By analyzing the packets of subscribers with the use of mining strategies, the malwares installed on mobile devices can be detected [6].

Compared with traditional malware detection systems, SMMDS can enhance the overall security level of the mobile internet while reducing the risks to mobile users. With the use of this approach, the malware packets, once discovered, can be blocked by installing a special device at the internet gateway. Every packet has to go through this device. In addition, the system improves the quality of service and competitiveness of mobile operators. Hadoop and several other cloud computing platforms can generally provide sufficient computation power support to build such a system, which can process large-scale data within a short time.

SMMDS is different from similar products for PCs in several ways. First, most of the similar products for PCs are aiming at intrusion detection by monitoring the traffic. SMMDS aims at identifying the malicious behaviors of the mobile applications. Second, those softwares for detecting PC virus from the internet side often need to work with local softwares which deploy within the PCs to function. SMMDS do not need any local applications in order to function. Moreover, since the difference in behaviors of PC and mobile malwares we mentioned above, the method for identifying these behaviors can be very different.

To meet the demands of users and mobile operators, an SMMDS based on the mining and classification of mobile internet packets is introduced. Moreover, to enhance the performance of the mining process, a novel clustering algorithm called contraction clustering is introduced. The major contributions of the system are as follows:

- This system is the first to move the process of detecting mobile malwares from the devices to the internet without any operation performed on the device. A system that is deployed at the internet gateway is established to analyze all the packets for malware detection.
- Sharing the features of malwares among all users in a unified knowledge database expands the range of detection and enhances the service quality of mobile operators.
- An improved clustering algorithm is introduced. This algorithm processes large-scale data at a high speed with theoretical and experimental proof.

The remainder of this paper is organized as follows: Section 2 summarizes the related works, including the latest detection methods and systems. Section 3 describes the architecture and working process of SMMDS. Section 4 discusses the improved clustering algorithm with its performance proof. Section 5 presents the implementation of the system and the experimental results. Section 6 concludes and presents the directions of future works.

2. Related work

In this section, related works on malware detection systems and methods are reviewed. The systems include those developed for use on PC platforms and mobile devices. These approaches include signature-based and machine learning methods.

2.1. Malware detection systems

Various malware detection systems have been created on PC-based platforms, and some of which are quite typical and exemplary [7–9]. Specially, in 2013, [10] presented a malware protection system which is built into the browser and determines the reputation of most downloads either locally or relying on the server-side data. Inspired by the latter two systems that based their detection on the data of the networks, we move the detection system out of the devices to save their resources.

Recently, malware detection systems of mobile devices are receiving increasing attention from scholars [11]. With the fast development of mobile malwares, protecting the privacies of users of mobile devices is very urgent and necessary. Meanwhile, a few related works were published, which can be divided into three categories:

Local detection. In 2012, [12] proposed a proactive scheme to spot zero-day Android malware by scalably analyzing whether a particular app exhibits dangerous behavior (e.g., launching a root exploit or sending background SMS messages). In 2013, [13] checked temporal properties of the interaction between an application and the Android event system using a graph constructed with static analysis. The two works both used static or dynamic methods to detect mobile malwares locally.

Distributed detection. This kind of detection system solves the problem of over consuming the resources of mobile devices to some extent. In 2010, [14] proposed a semi-distributed malware detection scheme which utilizes the social community structure and reflects a stable and controllable granularity of security to deal with the malware based on the infection history of the community. In 2011, [15] investigated the security versus energy tradeoffs and pointed that to enhance the accuracy of the detection, taking more computing and storage resources would be inevitable. In 2014, [16] proposed a distributed system for mobile malware detection with special optimization for limited storage of mobile devices.