Chapter 50
Intelligent Malware Detection Using Deep Dilated Residual Networks for Cyber Security

S. Abijah Roseline
VIT Chennai, India

S. Geetha
VIT, Chennai, India

ABSTRACT

Malware is the most serious security threat, which possibly targets billions of devices like personal computers, smartphones, etc. across the world. Malware classification and detection is a challenging task due to the targeted, zero-day, and stealthy nature of advanced and new malwares. The traditional signature detection methods like antivirus software were effective for detecting known malwares. At present, there are various solutions for detection of such unknown malwares employing feature-based machine learning algorithms. Machine learning techniques detect known malwares effectively but are not optimal and show a low accuracy rate for unknown malwares. This chapter explores a novel deep learning model called deep dilated residual network model for malware image classification. The proposed model showed a higher accuracy of 98.50% and 99.14% on Kaggle Malimg and BIG 2015 datasets, respectively. The new malwares can be handled in real-time with minimal human interaction using the proposed deep residual model.

INTRODUCTION

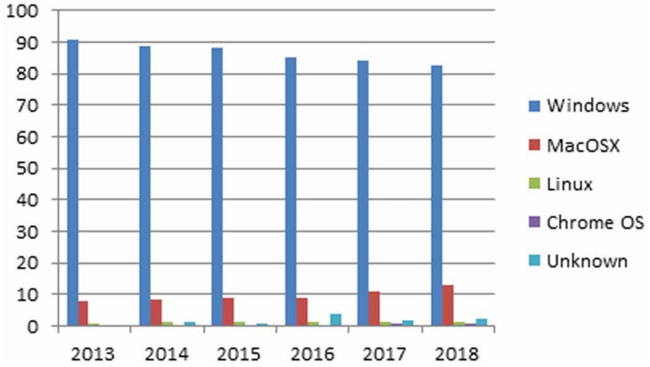
Microsoft windows are the first desktop operating systems with a market share of 82.7% (Statista portal). MacOSX, Linux, Chrome OS, and other unknown operating systems show very less market share as shown in figure 1. The attackers target the widely used Windows OS for achieving their goals. The wider use of computer systems and internet raises the number of security threats such as malware day by day. Cybersecurity is one of the significant areas in this information world with its useful strengths

DOI: 10.4018/978-1-7998-7705-9.ch050

in the everyday aspect of human activities at various levels. Cyber-attacks are uncommonly growing, resulting in greater amounts of data loss and financial loss to individuals or large organizations. Malware is one among the cyber-attacks which are currently sophisticated, stealthy and unknown to users. Security researchers take serious efforts to develop robust detection systems to identify known, as well as unknown malware. The cyber world happens to contain an excessive amount of data which are handled by machine learning applications.

Malware detection and identification of new malware are some of the cybersecurity challenges. Malware with different intents shows different behaviors. The advent of malware detection systems led to the development of detection avoidance mechanisms by the attackers. Although malware authors develop new malware rarely, most of the current malware are variants of existing malware. The previously written malware is slightly changed in any part of the code using any of the obfuscation techniques such as semantic nop insertion, code reordering, etc. Since new malware are similar in some characteristic to previous malware, they can be categorized into different families. But, they did not fulfill the aim of dealing with new zero-day and obfuscated malware with no false positives. Hence, it is necessary to classify malware into various classes or families for robust and intelligent detection of new malware.

Figure 1. The market share of the desktop OS between the years 2013-2018 at the global level



With the spread of new and unseen malware, traditional methods are not sufficient to cope with. Such traditional methods like signature-based methods are sufficient for previously known malware. But, they are not feasible solutions for advanced malware threats. To deal with such advanced threats, advanced machine learning techniques are devised. Particularly, deep learning techniques are more effective than conventional machine learning techniques for pattern recognition applications. Deep learning methods mimic human nervous systems by learning data through abstract and complex representation. The aim of the work is to train the deep learning model to effectively classify and detect the samples in the test dataset file into one of 9 categories (malware families).

The malware classification problem definition and detection solutions are described in the first section. A literature survey describing a review of the various works done for malware classification and detection is given in section 2. The variant deep learning methodologies were discussed in section 3. The proposed method is explained in detail in section 4. The malware dataset details are discussed in section 5. The comparison of various machine learning techniques is done and the effectiveness of the