Chapter 10

Similarity Measure for Obfuscated Malware Analysis

P. Vinod
SCMS School of Engineering and Technology, India

P. R. Rakesh
SCMS School of Engineering and Technology, India

G. Alphy
SCMS School of Engineering and Technology, India

ABSTRACT

The threats imposed by metamorphic malware (capable of generating new variants) can easily bypass a detector that uses pattern-matching techniques. Hence, the necessity is to develop a sophisticated signature or non-signature-based scanners that not only detect zero day malware but also actively train themselves to adapt to new malware threats. The authors propose a statistical malware scanner that is effective in discriminating metamorphic malware samples from a large collection of benign executables. Previous research articles pertaining to metamorphic malware demonstrated that Next Generation Virus Kit (NGVCK) exhibited enough code distortion in every new generation to defeat signature-based scanners. It is reported that the NGVCK-generated samples are 10% similar in code structure. In the authors’ proposed methodology, frequencies of opcodes of files are analyzed. The opcodes features are transformed to new feature spaces represented by similarity measures (37 similarity measure). Thus, the aim is also to develop a non-signature-based scanner trained with small feature length to classify unseen malware and benign executables.

INTRODUCTION

Malware is a generic term that refers to software which does undesired malicious activity in computer systems. As the amount of available data multiply, the problem of managing the information turn out to be more difficult. The increased use of internet file sharing has led to wide spread of malware. The consequence of this prevalence is that many computer systems are vulnerable and are infected with malicious programs. Zero day attacks cause great destruction to the computer world. Apart from the new attacks, existing malware threats are transformed into new ones. Malware detectors have not evolved to mitigate sophisticated attacks.
Therefore, there is a need urgent need to develop robust detector which can identify not only the existing malwares but also unseen and obfuscated malwares. Static signature based detection technique has been a dominant within Antiviruses. Many writers make use of virus constructors for developing new malware. These malware kits allow hackers to generate new malware specimens with minimal knowledge. Some of the tools are Next Generation Virus Kits (NGVCK), G2, Mass Code Generation (MPCGEN), Virus Creation Lab (VCL 32) etc. We consider variants of a malware from various tools like NGVCK, G2 etc. Earlier studies have already shown that NGVCK tool provides enough obfuscation in subsequent malware generations. Thus, traditional signatures can prove to be ineffective when dealing with unknown variants. In this work, we propose static analyzing technique to extract relevant features from malware. Based on these relevant features we determine similarity amongst pair of files using 37 similarity measurement indices. In an effort to classify the samples, the objectives are as stated below:

1. To find whether various distance/similarity measures are effective in classification of malware and benign executables.
2. To evaluate the effectiveness of virus generation tool kits like NGVCK, G2 etc in the generation of strong metamorphic variants.
3. To evaluate the effects of feature reduction in classification accuracy.
4. To find which classifier produces better result for the test/train model.
5. To find actegory of features that contribute to the detection scheme.
6. To evaluate response of the proposed detector on imbalanced dataset.

In the remaining part of this chapter we briefly introduce different types of malware. Obfuscation methods adopted by metamorphic engine to generate strong metamorphic malware is also discussed. Subsequently, we discuss malware detection techniques proposed by researchers to identify metamorphic malware. Also, we introduce machine learning methods for the detection of malicious code. Later part of this chapter we discuss our proposed methodology based on similarity measurement indices. Finally experiments, results, inferences are covered. We close the chapter with pointers to future research in the challenging domain of desktop and mobile malware detection.

MALWARE AND DETECTION TECHNIQUES

Malicious programs have a long history and ever since the invention of malicious programs their detection have attracted the anti-malware community. Malicious programs aimed at detecting useful information confined to system and users, they remain dormant and undetected. Anti-detection mechanisms have evolved into complexity. However, sophisticated malicious program also known as polymorphic and metamorphic viruses have evolved in unprecedented rate. Detection of malcode that can obfuscate or morph future instances is a big challenge and needs to be addressed critically. In the following subsection, we discuss different types of malware with their detection techniques.

Malware

Malware or unwanted software’s are the program created to compromise normal functionality of computer, gather sensitive information by gaining super user privilege and bypass access control. Malicious programs prevails itself in numerous forms such as codes, scripts, active contents and other software. Often, malware is confused as defective codes (also known as bug), whereas malware is skillfully implanted to disrupt the functioning of the host program.
Related Content

The Desire for Privacy: Insights into the Views and Nature of the Early Adopters of Privacy Services
www.igi-global.com/chapter/desire-privacy-insights-into-views/23109?camid=4v1a

Avoiding Pitfalls in Policy-Based Privacy Management
www.igi-global.com/chapter/avoiding-pitfalls-policy-based-privacy/21339?camid=4v1a

Network Intrusion Detection Using Multi-Objective Ensemble Classifiers
www.igi-global.com/chapter/network-intrusion-detection-using-multi-objective-ensemble-classifiers/152965?camid=4v1a

Design and Implementation of a Framework for Assured Information Sharing Across Organizational Boundaries
www.igi-global.com/article/design-implementation-framework-assured-information/2493?camid=4v1a