© The British Computer Society 2015. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

Advance Access publication on 13 July 2015
doi:10.1093/comjnl/bxv047

Toward a Taxonomy of Malware Behaviors

André Ricardo Abed Grégio[1,2]*, Vitor Monte Afonso[2],
Dario Simões Fernandes Filho[2], Paulo Lício de Geus[2] and Mario Jino[2]

[1] *Center for Information Technology Renato Archer, Campinas, Sao Paulo, Brazil*
[2] *University of Campinas, Campinas, Sao Paulo, Brazil*
**Corresponding author: gregio@lasca.ic.unicamp.br*

Malicious code attacks pose a serious threat to the security of information systems, as malware evolved from innocuous conceptual software to advanced and destructive cyber weapons. However, there is still the lack of a comprehensive and useful taxonomy to classify malware according to their behavior, since commonly used names are obsolete and unable to handle the complex and multi-purpose currently observed samples. In this article, we present a brief survey on available malware taxonomies, discuss about issues on existing naming schemes and introduce an extensible taxonomy consisting of an initial set of behaviors usually exhibited by malware during an infection. The main goal of our proposed taxonomy is to address the menace of potentially malicious programs based on their observed behaviors, thus aiding in incident response procedures. Finally, we present a case study to evaluate our behavior-centric taxonomy, in which we apply identification patterns extracted from the proposed taxonomy to over 12 thousand known malware samples. The leveraged results show that it is possible to screen malicious programs that exhibit suspicious behaviors, even when they remain undetected by antivirus tools.

Keywords: malware taxonomy; malware behavior; malware analysis; incident response

Received 29 July 2014; revised 18 May 2015
Handling editor: Albert Levi

1. INTRODUCTION

A malicious software, or malware, is a set of instructions that run on a system to make it do arbitrary activities on behalf of an attacker [1], or to act in a way (automated or not) that threatens security aspects of the compromised system, its users and associated data. Malware evolved from the harmless concept of self-replicating automata to multipurpose stealthy code capable of destroying physical assets. However, despite all evolution, the malware counter-measures research field suffers from the lack of a general-purpose, behavior-based taxonomy able to address the myriad of new samples. This taxonomy should be informative and practical, to help security professionals in grouping incidents caused by malicious programs based on their reported or observed activities.

Existing malware taxonomies usually map observed features—which may be the descriptive predominant behavior, structural organization (topology) or type of attacks launched—to obsolete naming schemes, thus assigning this feature to a predefined class (e.g. the behavior of appending itself to another file is usually linked to the class of 'viruses'). This kind of taxonomy does not embrace the plethora of distinct and complex malware samples currently observed in the wild. Therefore, there is a need of a new taxonomic scheme to address modern malware by grouping samples of malicious programs according to their observed execution behavior.

One of the most popular defense mechanisms against malware is the antivirus (AV), whose role is to detect the maximum number of malicious programs despite their name or exhibited features. In order to try to organize how several AVs assign names to the files they detect as malicious, the security community put efforts to create standard naming schemes [2], as well as distributed databases about malware samples and their behavior [3–5]. The increasing rise of malware variants and the need of manual procedures to extract and input information about the evaluated sample into those naming schemes do not contribute to their adoption. Moreover, we believe that since an