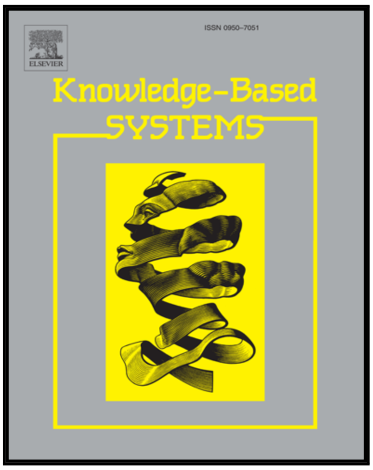

Detection of Malicious Webmail Attachments Based on Propagation Patterns

Yehonatan Cohen, Danny Hendler, Amir Rubin

PII: S0950-7051(17)30533-6
DOI: 10.1016/j.knosys.2017.11.011
Reference: KNOSYS 4106

To appear in: *Knowledge-Based Systems*

Received date: 3 September 2017
Revised date: 8 November 2017
Accepted date: 10 November 2017

Please cite this article as: Yehonatan Cohen, Danny Hendler, Amir Rubin, Detection of Malicious Webmail Attachments Based on Propagation Patterns, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.11.011

Detection of Malicious Webmail Attachments Based on Propagation Patterns

Yehonatan Cohen[a,*], Danny Hendler[a], Amir Rubin[a]

[a]*Department of Computer Science, Ben-Gurion University of The Negev, Be'er Sheva 84105, Israel*

Abstract

Email remains one of the key media used by cybercriminals for distributing malware. Based on a large data set consisting of antivirus telemetry reports, we conduct the first comprehensive study of the properties of malicious webmail attachments. We show that they are distinct among the general web-borne malware population in terms of the malware *reach* (the number of machines to which the malware is downloaded), malware type and family. Furthermore, we show that malicious webmail attachments are unique in the manner in which they propagate through the network.

We leverage these findings for defining novel features of malware propagation patterns. These features are derived from a time-series representation of malware download rates and from the community structure of graphs that model the network paths through which malware propagates. Based on these features, we implement a detector that provides high-quality detection of malicious webmail attachments.

Keywords: Malware; Time Series Analysis; Community Detection; Early Detection; Service Provider

1. Introduction

Traditional antivirus software relies on signatures to uniquely identify malicious files. Once a file is determined to be malicious, its signature is computed and added to a signatures database. Malware writers, on the other hand, have responded by developing obfuscation techniques with the goal of evading signature-based detection. Polymorphic and metamorphic malware utilize dead-code insertion, subroutine reordering, encryption, and additional techniques, in order to alter a file's content and make signature-based detection more difficult (Musale et al., 2015; You and Yim, 2010). A consequence of this arms race is that numerous new malware instances are generated every day, thus limiting the effectiveness of static detection approaches. For effective and timely malware detection, signature-based mechanisms must be augmented with detection approaches that are harder to evade.

[*] Corresponding author. Tel: +972 (0)525282229

Email addresses: `yehonatc@cs.bgu.ac.il` (Yehonatan Cohen), `hendlerd@cs.bgu.ac.il` (Danny Hendler), `t-amirub@microsoft.com` (Amir Rubin)