Computers
& Security
The International Source of Innovation for the Information Security and IT Audit Professional
ELSEVIER
ifip

Early-Stage Malware Prediction Using Recurrent Neural Networks

Matilda Rhode, Pete Burnap, Kevin Jones

PII: S0167-4048(18)30554-6
DOI: 10.1016/j.cose.2018.05.010
Reference: COSE 1347

To appear in: *Computers & Security*

Received date: 22 December 2017
Revised date: 8 April 2018
Accepted date: 15 May 2018

Please cite this article as: Matilda Rhode, Pete Burnap, Kevin Jones, Early-Stage Malware Prediction Using Recurrent Neural Networks, *Computers & Security* (2018), doi: 10.1016/j.cose.2018.05.010

Early-Stage Malware Prediction Using Recurrent Neural Networks

Matilda Rhode[a,*], Pete Burnap[a], Kevin Jones[b]

[a] *School of Computer Science and Informatics, Cardiff University*
[b] *Airbus Group*

Abstract

Static malware analysis is well-suited to endpoint anti-virus systems as it can be conducted quickly by examining the features of an executable piece of code and matching it to previously observed malicious code. However, static code analysis can be vulnerable to code obfuscation techniques. Behavioural data collected during file execution is more difficult to obfuscate, but takes a relatively long time to capture - typically up to 5 minutes, meaning the malicious payload has likely already been delivered by the time it is detected.

In this paper we investigate the possibility of predicting whether or not an executable is malicious based on a short snapshot of behavioural data. We find that an ensemble of recurrent neural networks are able to predict whether an executable is malicious or benign within the first 5 seconds of execution with 94% accuracy. This is the first time general types of malicious file have been predicted to be malicious during execution rather than using a complete activity log file post-execution, and enables cyber security endpoint protection to be advanced to use behavioural data for blocking malicious payloads rather than detecting them post-execution and having to repair the damage.

Keywords: malware detection, intrusion detection, recurrent neural networks, machine learning, deep learning

1. Introduction

Automatic malware detection is necessary to process the rapidly rising rate and volume of new malware being generated. Virus Total, a free tool which can be used to evaluate whether files are malicious, regularly approaches one million new, distinct files for analysis each day[1][1].

Commonly, automatic malware detection used in anti-virus systems compares (features extracted from) the code of an incoming file to a known list of malware signatures. However, this form of filtering using static data is unsuited to detecting completely new ("zero-day")

[*] Corresponding author

Email addresses: rhodem@cardiff.ac.uk (Matilda Rhode), burnapp@cardiff.ac.uk (Pete Burnap), kevin.jones@airbus.com (Kevin Jones)

[1] 0.935 million on 2nd December 2017