Open Source Intelligence for Malicious Behavior Discovery and Interpretation

Cameron Noakes

February 2022

1 Sentence Overview

The relevant study within this literature focuses its efforts on knowledge collection from the ATTCK framework tactics, techniques and procedures (TTPs), malicious behavior identification using deep learning, and the identification of specific API calls.

An ATTCK based Malicious Behavior Analysis system (MAMBA) for Windows malware is proposed which incorporates manipulated resources and malicious activities in the neural network model which is related to the deep learning outlined. The proposed solution also yields high mapping functionality from the discovered malicious behaviors to relevant ATTCK techniques in the ATTCK matrix.

2 Introduction

Open Source Intelligence (OSINT) provides experience and knowledge from cybersecurity to form a common knowledge base with publicly available information online. The strength of the ATTCK framework (a popular OSINT knowledge base) is its structure and openness in collecting and sharing cyber threat intelligence and potential attacks and the associated vectors that come along with it.

An overall outline of the fast ever evolving attack scenarios in applications such as ATTCK and alike helps cybersecurity analysts handle potential attacks as they are found or executed.

The explained focus of this research and literature is to focus on analyzing the dynamic behavior of malware using the knowledge from ATTCK and with the use of neural networks using tools such as Cuckoo Sandbox, CWSandbox, and APIf to help detect malicious behavior.

The researchers later extract and organize the relations of the TTPs from the ATTCK framework matrix. In this study, the researchers employ the attention mechanism to identify and interpret the outcome from the proposed model, the outcome wont be a Boolean but more of a decision between TTPs, API calls and the associated resources in this study.

3 Background and Motivation

The researchers analysed an example a malware sample classified as part of the JCry malware family. Detailed about its created processes is given and the discovered TTPs.

JCry is an example of ransomware disguised as an adobe flash player update which, in turn, creates malicious files when executed to maintain persistence. The ATTCK framework manages to identify such attacks through its various TTPs for each step of the attack (initial threat and the persistence).

4 Techniques and Execution Trace

The MITRE domain provides textual descriptions of TTPs for which the proposed model MAMBA can extract the needed resources and match them with arguments for when using the API calls. This can help outlined a strong correlation between ATTCK techniques and Windows API calls.

The main task of the MAMBA model is to align a resource annotated with a TTP in the ATTCK framework to a manipulated resource used by any malware.

5 Knowledge Extraction from MITRE ATTCK Framework

To start extracting the knowledge from the ATTCK framework, the literature outlines how the researchers needed to extract a disclosed resource 'r' related to technique 'y' for every technique in the ATTCK matrix and then perform regular expressions (regex) for the 'r' extraction of a token, where a token is a complete path of a resource. After this a full collection of techniques from the ATTCK framework matrix can be gathered.

Several usages of mathematics and algorithms are used to help with the malicious behavior identification using deep learning some examples are weight matrices and probability distribution. Pseudo code is used to explain the MAMBA Neural Network with the Dataset Statistics being shown in a table on page 7.

6 Conclusion

The MAMBA proposed system incorporates ATTCK techniques to be used for the deep learning neural network in relation between resources and API calls for malicious behavior identification.

The proposed solution was then evaluated by big datasets and the techniques in the ATTCK framework matrix. The researchers outline how the model still could need improvements and due to data collection limitations could make the data less complete.

7 What problem does this solve?

The reasoning behind the study and proposed system is due to needing to understand the characteristics of malware, including their activities and manipulated resources on the target machines, there was a need for a proposed model to help aid in the knowledge and mapping to API calls with the use of ATTCK TTPs (most commonly just the techniques).

8 How do they solve it?

The researchers develop a proposed system that is Malicious Behavior Analysis system (MAMBA) for Windows malware through deep learning and neural networks and use ATTCK techniques to map malicious behavior identification to API calls.

9 How do they assess it?

The evaluation is done through comparing the performance of MAMBA neural network and other methods using the ATTCK framework and big datasets to answer the asked Q1 and Q2 within the literature. Comparisons are made for the performance of MAMBA with two rule-based systems (Cuckoo Signatures) and Regular Expressions (RegEx) and five traditional machine learning methods.