講演番号： 21

動的解析と静的解析を併用したマルウェア検知の研究

A-7　Study of the malware detection that used dynamic analysis and static analysis

武田 理史† 宮保 憲治†

Masafumi TAKEDA Noriharu MIYAHO

† 東京電機大学情報環境学部情報環境学科

† School of Information Environment, Tokyo Denki University

1. はじめに

近年、ハッカーは国際的に組織化され攻撃のテクノロジーは急速に高度化し、新規のマルウェア発生件数は増加の一途をたどっている。同時に、近年のマルウェアは、解析を妨害する処理がされているため、解析情報の取得も困難である。

本稿では、動的解析と静的解析を併用することによって得られた、解析情報から特徴量を抽出し、機械学習に活用することにより、マルウェアの検知を行う手法を提案する。

2. 実験内容

まず、本実験で使用する動的解析、静的解析、機械学習の手法について以下に述べる。

動的解析は、オープンソースの動的解析ツールである CuckooSandbox[1]を使用した。CuckooSandbox は、実行ファイルや文書ファイル等の各種データファイルを CuckooSandbox に投入することで、自動的に仮想環境上で実行 、解析し、その結果をレポート化した HTML、JSON ファイルや通信データを記録した PCAP 形式のファイルを出力するツールである。CuckooSandbox を活用したマルウェアを解析するための仮想環境を Ubuntu 上に、VirtualBoxをインストールすることにより構築した。

静的解析を行う場合は、まず、解析対象のマルウェアに対して文字列の抽出を行う。次に、難読化ツールの特定をし、難読化状態を解除した後に、CuckooSandbox を活用して解析する。CuckooSandbox のレポート内で得られる StaticAnalysis の情報を使用することで、静的解析を行うことができる。

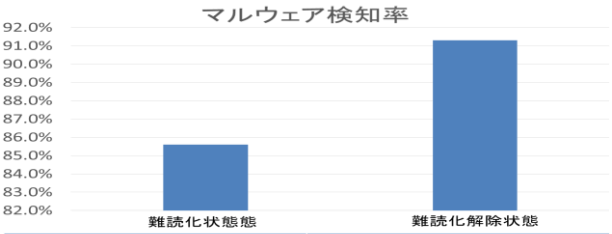
機械学習は、プログラミング言語 Python の機械学習用ライブラリ Scikit-Learn[2]を用いて実装した。特徴量として、解析により得られた Win32API の関数名を使用した。また、アルゴリズムとして、SVM (SupportVectorMachine)を使用した。

次に、実験内容について以下に述べる。

EXE 形式のマルウェア 530 体と、難読化ツールを特定し、難読化を解除した状態のマルウェア 130 体、および Windows7の環境に存在する EXE 形式の非マルウェア 530 体を活用して CuckooSandbox で解析を行った。解析結果から得られた API の関数名を特徴量として、機械学習により、難読化状態のマルウェアと、難読化を解除した状態のマルウェアの検知率を比較した。

3. 実験結果

難読化状態のマルウェアと、難読化を解除した状態のマルウェアの検知率の比較を行った結果、難読化状態のマルウェアの検知率が 85.6%であった。一方、難読化を解除した状態のマルウェアの検知率は 91.3%となり、5.7%の検知率向上を実現できることが分った。



マルウェアの状態	検知率
難読化状態	85.6%
難読化解除状態	91.3%

図1. マルウェア検知率の比較

4. まとめと今後の課題

本実験では、難読化状態を活用した動的解析技術のみ使用した場合と、難読化を解除した状態を加えて、動的解析と静的解析の技術を併用した場合の検知率の向上効果を検証した。

その結果、後者がより高いマルウェアの検知が可能であることを検証した。

参考文献

[1]CuckooSandbox:http://www.cuckoosandbox.org/, 2015/05/10

[2] http://scikit-learn.org/stable/, 2015/11/07