SCIS 2022 2022 Symposium on
Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 – 21, 2022
The Institute of Electronics,
Information and Communication Engineers

シンボルが削除されたIoTマルウェアにおける自然言語処理を用いた関数名推定
Function Name Estimation by Natural Language Processing in Symbol-Stripped IoT Malware

イボット　アリジャン * 大山 恵弘 *
Ibot Arijan Yoshihiro Oyama

キーワード IoT マルウェア, 関数名推定, 自然言語処理, ニューラル機械翻訳

あらまし

近年，IoT デバイスの普及に伴いセキュリティが貧弱な IoT デバイスが増加し，それを狙ったマルウェアが増加している．増加を続ける IoT マルウェアの動向を把握するために，マルウェアを正しく解析する手法が必要とされている．しかし IoT マルウェアは，今まで研究されてきた Windows マルウェアと異なる点が存在する．それは多くの IoT マルウェアが静的リンクでありながらシンボルが削除されている点である．シンボルが削除されたファイルでは関数の境界や名前がわからなくなるため，関数からマルウェアの機能を推定することが難しくなる．

これを解決するために Artuso らの研究 [1] では，自然言語処理の翻訳アルゴリズムを用いた関数名推定を行っている．彼らの手法ではアセンブリ命令列と関数名を対訳として扱うことで，翻訳アルゴリズムを関数名推定に応用している．しかし，彼らの研究では機械学習モデルの違いなどによる精度の差は調査しているが，他の関数名推定手法と比較が行われていない．また，マルウェアを使った関数名推定や，推定された関数名がマルウェアの分類や検知にどれほど有効であるかは示されていない．本研究では Artuso らの研究を拡張し，IoT マルウェアを対象とした関数名推定手法を提案する．彼らの研究では Linux/x86-64 を対象としていたが，本研究では IoT マルウェアに多い Linux/Arm を対象としている．

実験では，IoT マルウェアから取得した関数をデータセットとして使い，提案手法の評価を行った．その結果，提案手法は精度と学習時間の面で，既存手法より優れていることがわかった．また，推定された関数名を使ったマルウェア分類の実験も行った．その結果，既存手法と比較してホワイトボックスな分類手法でありながら，既存手法と同等の精度を得ることができた．

本研究では，提案手法に関する詳細な調査も行った．提案手法では基本的な自然言語処理アルゴリズムを使ったが，近年の強力な自然言語処理アルゴリズムでも実験を行った．その結果，精度は少し向上したものの計算コストがかなり大きくなることがわかった．また，転移学習に関しての実験も行った．Ubuntu のリポジトリから取得した大量のファイルを使って学習し，その後少量のマルウェアで学習済みモデルを調整した．その結果，転移学習は提案手法においても有効であるとわかったが，提案手法の精度が十分に高かったため大きな精度向上には至らなかった．

提案手法が推定を間違えた関数を調査したところ，関数内で使用されるメモリアドレスのみ異なる関数が多いことがわかった．この問題を解決するために，関数内で呼び出された関数を選択的にインライン展開する方法などが考えられる．また，提案手法は逆アセンブルを必要とするため，既存手法と比べて前処理にかかる時間が長いという問題がある．これらの問題については今後の研究課題とする予定である．

参考文献

[1] Artuso, F., Di Luna, G., Massarelli, L., Querzoni, L.: In Nomine Function: Naming Functions in Stripped Binaries with Neural Networks, arXiv (2019), https://arxiv.org/abs/1912.07946

* 筑波大学, 茨城県つくば市天王台 1-1-1, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan.