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

動的解析ログを用いたマルウェアの早期目的推定に向けた特徴量の予測手法に関する検討

A Study on Feature Prediction Methods for Early Object Estimation of Malware Using Dynamic Analysis Logs

朝倉 紗斗至 * 中川 恒 † 押場 博光 † 市野 将嗣 *
Satoshi Asakura Ko Nakagawa Hiromitsu Oshiba Masatsugu Ichino

キーワード マルウェア，動的解析，目的推定，早期推定，クラスタリング

あらまし

近年，マルウェアの巧妙化により，その対策が重要となっている．マルウェアの解析において，動的解析は広く行われている手法である．動的解析を行う際にマルウェアがどのような侵害活動を目的としているかを把握しておくことは解析に役立つと考える．例えば，マルウェアの目的が他のマルウェアをダウンロードするものだと推定することができれば解析者はその結果を念頭に，より詳細な解析をスムーズに進められる．また，マルウェアの中には長期間実行が継続するものがあり，解析には長時間記録されたログが必要である．そのため，長時間のログの収集を待たずに，短時間のログを用いて行う「早期の目的推定」ができる必要があると考える．加えて，解析環境によってマルウェアの実行が途中で止まる場合もある．この場合にも対応できるように少ないログのみで行える目的推定は必要だと考える．

本研究ではマルウェアの早期の目的推定のため，感染開始時点から短時間記録されたログを用いて高精度に目的推定を行うことを目指す．そのため，本研究では動的解析ログを用い，ある検体の短時間のログ（ログＡとする）で得られた特徴量から同じ検体の長時間のログ（ログＢとする）で得られた特徴量を予測し，予測した特徴量を目的推定に利用する．このとき，ログＡはログＢの先頭部分のログである．この手法を行った筆者らの既存研究[1]では，特徴量を予測する際に１つの回帰モデルのみを用いていたが，本稿ではより正確な予測のため，マルウェアをクラスタリングした後にクラスタごとに回帰モデルを作成し，予測を行う手法を提案する．また，既存研究[1]のときとは実験方法をいくつか変更した．具体的には，検体数の増加，推定対象ラベルの変更，マルチラベル推定の実施，ログから抽出する挙動情報の追加，などを行った．

実験では，MWS Datasets [2] の一部として提供されている Soliton Dataset 2019, 2020 に含まれるマルウェア 546 検体の動的解析ログを使用し，提案手法を用いた目的推定実験を行った．実験にあたり，マルウェアの目的を 7 種類（ダウンロード，暗号化，拡散，情報窃取，マイニング，遠隔操作，ドロップ）定義し，各検体にラベルとして付与した．このとき，1 検体が複数の目的を持つ場合はマルチラベルとした．実験の結果，提案手法により短時間のログを用いたときの目的推定精度が長時間のログを用いたときの精度と同程度まで向上した．これにより，目的推定に要する時間を短縮することができた．

参考文献

[1] 朝倉紗斗至，他，“動的解析ログを用いた特徴量の予測によるマルウェアの早期機能推定に関する検討”，コンピュータセキュリティシンポジウム 2020 論文集，pp.602-609 (2020)

[2] 寺田真敏，他，“マルウェア対策のための研究用データセット MWS Datasets ～コミュニティへの貢献とその課題～”，情報処理学会，Vol.2020-IFAT-139 No.8 (2020)

* 電気通信大学，〒182-8585 東京都調布市調布ヶ丘 1-5-1
† 株式会社 FFRI セキュリティ，〒100-0005 東京都千代田区丸の内３丁目３番１号 新東京ビル２階