様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月　 ８ 日現在

機関番号：１２１０２
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ１１７４１

研究課題名（和文）マルウェア解析における例外的事象への対策技術

研究課題名（英文）Countermeasure Techniques for Exceptional Events in Malware Analysis

研究代表者
大山　恵弘（Oyama, Yoshihiro）

筑波大学・システム情報系・准教授

研究者番号：１０３６１５３６

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：本研究課題では，多くのマルウェアで処理や目的を解析で把握することが困難になっている問題を解決するための要素技術を開発した．ソフトウェア工学の技術をマルウェア解析に適用するにはどんな手法が必要かという問題や，それらのうち解析の精度と効率を特に向上させるものはどれで向上幅はどの程度かという問題について解明や考察を進め，有益な結果を得た．具体的には，攻撃前に例外を発生させて実行を終了するマルウェアを安定的に実行する技術や，破損していて解析や実行に失敗するマルウェアなどを不完全な情報だけを用いて解析する技術を開発した．また，最新のマルウェアによる解析回避などの処理の実態についても明らかにした．

研究成果の学術的意義や社会的意義
本研究成果の意義はソフトウェア工学の技術をマルウェア解析に適用する方法論を構築したことと，その有効性を検証したことである．ソフトウェア工学の技術のマルウェアへの適用は今まで途上段階にあった．本研究では，そのような技術をマルウェア解析でも有効に利用できることを示した．本研究で取り組んだ「問い」は，(1) ソフトウェア工学の技術をマルウェア解析に適用するにはどんな手法が必要か，(2) それらを適用したときの解析の精度と効率はどうなるか，であった．それらを明らかにする上で一定の成果を得た．本研究の学術的独自性と創造性は，マルウェアの挙動やファイルに関する例外的な事象や特徴を対象とする点にあった．

研究成果の概要（英文）：In this research project, we developed elemental technologies to solve the problem that it is difficult to understand the operations and purpose of many malware programs by analysis. We have obtained useful results by clarifying and discussing issues related to the methods needed to apply software engineering techniques to malware analysis, as well as the methods that specifically improve the accuracy and efficiency of the analysis and the extent of the improvement. Specifically, we developed techniques for stable execution of malware that terminates execution with an exception before the actual attack activity, and techniques for analyzing corrupted malware that the analyst fails to analyze or execute, using only incomplete information. We also clarified the actual state of various operations by the latest malware such as analysis evasion.

研究分野：ソフトウェア

キーワード：マルウェア

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

マルウェアによるセキュリティ被害の度合いと範囲は拡がる一方である．マルウェアの検知と攻撃防止のために必要な処理の 1 つはマルウェア解析である．マルウェア解析とは，マルウェアそのもの（検体）や攻撃の挙動をもとに，マルウェアの処理，構造，意図などを明らかにする作業である．

近年ではマルウェアが洗練されてきているためマルウェア解析が著しく困難化している．一部のマルウェアは自身が解析されることを想定し，解析を回避するための処理を実行する．具体的には，自身が解析されているかどうかを検査し，もし解析されていると判断したら，実行を終了させたり，ダミーの処理を実行し続けたりするなどの解析回避のための対応を行う．この結果，解析回避の機能を持つマルウェアを実行しても，早期に実行を終了したり，攻撃として意味のある挙動を示さなかったりして，マルウェアの攻撃や意図を解析者が把握できなくなる．

この問題に対してはこれまで多くの研究が行われてきた．解析回避の挙動を明らかにする研究もあれば，解析回避を行うマルウェアからもできる限り有用な情報を抽出する手法の研究もあった．しかしそれらの多くはマルウェア実行における例外的事象は中心的に扱っておらず，それらの事象が発生しても効果的，効率的に解析を進めるための技術については研究や開発が不足していた．また，昨今，マルウェア検体そのものが破損しているという例外的状況も散見されるにも関わらず，その状況についての想定が欠けており，大半の技術や研究ではそのような検体を除外するか，通常の状況における手法を適用し続けるものとなっていた．

２．研究の目的

本研究の目的は，マルウェア解析の特に例外的事象の扱いに対してソフトウェア工学の技術を適用する方法論の構築と，その有効性の検証である．ソフトウェア工学の技術は今まで主に通常の（善良な）ソフトウェアに対して開発，適用されており，マルウェアへの適用は途上段階にあった．具体的には，2 つのサブ研究課題に取り組む．1 つ目は，ソフトウェア工学の技術によるマルウェアの安定的実行である．2 つ目は，ソフトウェア工学の技術による破損マルウェアの理解と復元である．核心をなす学術的「問い」は 2 つである．1 つ目は，ソフトウェア工学の技術をマルウェア解析に適用するにはどんな手法が必要か，である．2 つ目は，それらのうち解析の精度と効率を特に向上させるものはどれで向上幅はどの程度か，である．本研究の学術的独自性と創造性は，マルウェアの挙動やファイルにおける頻発する事象や特徴ではなく，例外的な事象や特徴を対象とする点にある．

具体的には，本研究では多くのマルウェアで処理や目的を解析で把握することが困難になっている問題を解決する．攻撃前に例外を発生させて実行を終了するマルウェアや，破損していて解析や実行に失敗するマルウェアを主に対象にする．取り組む具体的なテーマは，第一に，ソフトウェア工学技術によるマルウェアの安定的実行である．研究代表者が開発した例外処理や解析妨害対策の技術と既存技術の組み合わせにより，未知であったマルウェアの特徴をより多く解明することを目指す．第二に，ソフトウェア工学技術による破損マルウェアの理解と復元である．新しい破損マルウェア分類技術と既存技術の組み合わせにより，破損などの異常に強い解析技術の構築を目指す．

３．研究の方法

研究目的を達成するために複数の方法で研究を実施した．

まず，例外を発生させるマルウェアのための解析技術を開発した．マルウェアはしばしば例外の発生により自身の実行を終了させる．一般には，このような例外は特定の条件において発生するが，マルウェア解析においては，潜在的な脅威を理解するために，そのような例外が発生しなかった場合の挙動情報も収集する必要がある．そこで，例外を発生させたマルウェアの実行を終了させることなく強制的に継続させる方法を設計，実装する．その方法においては，マルウェアが発生させた例外を捕捉し，その例外を無効化，すなわち，あたかも最初から発生しなかったかのように偽装する．これは例外発生時のマルウェアのメモリとレジスタを書き換え，例外の種類に応じて柔軟に配送を制御することによって実現する．この方法を用いると，マルウェアは致命的な例外をある意味では乗り越えて実行を継続できる．

次に，フォレンジックにおいて復元されたなどの理由で一部が破損しているマルウェアのための解析技術を開発した．

第一に，マルウェア自身によるファイル消去などによりデータの一部が欠損したマルウェア検体を対象にした，アンチウイルスによるマルウェアの種別同定に関する調査を行った．具体的には，欠損した状態でどこまでマルウェアの種別を同定できるのか，マルウェアのデータのうちどの部分が欠損すると同定に影響が出るのかを調査した．

第二に，欠損マルウェアに対してアンチウイルスではなく機械学習技術を用いたマルウェア種別の同定を試みた．欠損マルウェアのデータは欠損しており，ヘッダ情報や挙動解析情報などの有用な特徴が得られないため，同定には画像特徴量を使用した．

第三に，表層解析結果の特定の部分のみを用いた，機械学習によるマルウェア分類に関して有用な実験結果や知見を示した．表層解析は，マルウェアのファイル名，ハッシュ値，文字列などリソースの情報，セクションの情報，PE ヘッダの情報など，マルウェアを実行したり逆アセンブルしたりせずに得られる表層的な情報を収集する解析手段である．具体的には，これまで注目されることが少なかった，マルウェアの PE ファイル中のドキュメント化されていないヘッダ（Rich ヘッダ）のみを用いることを試みた．Rich ヘッダは Microsoft 社のリンカを使用して作られた PE ファイルに付与されるヘッダであり，その PE ファイルのビルド時の環境情報を含んでいる．また，Rich ヘッダはパックと呼ばれる操作を受けても残存し，攻撃者の妨害を受けにくいことが指摘されている．そこで，この Rich ヘッダのみを特徴量として機械学習によるマルウェア分類を試みた．実験では，既存のマルウェアデータセットから Rich ヘッダを持つ検体を抽出して作ったデータセットを用いた．

第四に，マルウェア画像を入力として受け取るマルウェア検知器を対象とした，機械学習を意識した攻撃の 1 つである敵対的パッチ攻撃の効果に関する実験を行い，様々な攻撃と効果の関係についての有用な実験結果や知見を示した．敵対的パッチ攻撃とは，入力値に特殊な摂動（敵対的パッチ）を加えることで，機械学習モデルの判別を誤らせる攻撃である．具体的には，攻撃者がマルウェアに敵対的パッチを挿入することにより検知を回避することが可能かどうか，ファイル中のどの位置にパッチを挿入するとより回避の成功率が上がるか，を実験で調査した．

他にも，最新のマルウェアの解析回避の処理やパスワード総当たり攻撃の処理を分析する研究も行った．

４．研究成果

マルウェア解析において，より有用な情報を抽出できる形で例外的事象を扱うための複数の要素技術を開発し，実験を通じてそれらの有効性を実証した．

まず，例外を発生させるマルウェアのための解析技術については，著名なサンドボックスシステムである Cuckoo Sandbox を提案方法で拡張し，例外発生時にメモリ読み出しを偽装したりメモリ書き込みを無視したりする処理を実装した．さらに，例外の無効化を実施すると無限ループで実行が進まなくなる問題が顕著になることがわかったので，無限ループからできるだけ賢く脱出するための方法を設計し，そのシステムに実装した．そのシステムを利用して提案方式の有無にしたがってマルウェアの実行がどう変化するかを観測した．その結果，30%以上のマルウェアの実行では提案方法により API 呼び出し回数やアクセスされる資源数が増加した．これは，提案方法を使えばそれまで認識できていなかったマルウェアの挙動情報を抽出できるようになることを示すものである．

一部が破損しているマルウェアのための解析技術については以下の成果を得た．第一に，マルウェア自身によるファイル消去などによりデータの一部が欠損したマルウェア検体を対象にした，機械学習によるマルウェアの種別同定については，欠損の仕方をもとに人工的にマルウェアを欠損させる実験の結果，特にファイルの先頭や末尾の欠損がマルウェアの同定や悪性判定の精度を著しく下げることを明らかにした．第二に，欠損マルウェアに対する機械学習技術を用いたマルウェア種別の同定については，アンチウイルスによる同定に致命的な悪影響を与えるファイル先頭の欠損があったとしても，学習に用いるデータによっては，かなり高い精度で同定が可能であることを明らかにした．第三に，表層解析結果の特定の部分のみを用いた，機械学習によるマルウェア分類については，Triplet Network を採用した深層学習によりマルウェア分類を行い，層化 5 分割交差検証を実施した結果，かなり高い精度で分類が成功した．また，Rich ヘッダを含む実行ファイルをパックしたときに Rich ヘッダが残存するかどうかを，複数のパッカーを用いて検証した．その結果，著名なパッカーの中でも Rich ヘッダを消去せず残存させるパッカーがあることがわかった．第四に，敵対的パッチ攻撃の効果については，敵対的パッチによって検知を回避することは可能であり，特に，ファイル中で最初に登場するセクションの直前に敵対的パッチを配置する場合に最も高い割合で検知を回避できることがわかった．

最新のマルウェアの解析回避の処理の分析については，主に Soliton データセットというマルウェアデータセットのマルウェア解析ログを対象にして調査を行い，解析回避と思われる処理の実行回数や，そのような処理を実行した検体を示した．また，得られた結果について考察し，解析回避処理は頻繁に観測される処理ではないことや，サンドボックスにより検出された解析回避処理には誤検出が多い可能性があることを示した．同時に，解析回避処理である可能性が高い処理が確かに実行されていることも実例とともに示し，注意を喚起するとともに対策を検討する助けになる情報を提供した．

パスワード総当たり攻撃の処理の分析については，主に Soliton データセットと FFRI データセットというマルウェアデータセットを対象に調査や実験を行い，マルウェア解析ログから実際にパスワードリストを抽出して得られた特徴を報告した．実験の結果，ログ内の文字列リストから探索の鍵となるパスワードを検出してその周辺の文字列を調べるという方法により良質なパスワードリストを抽出できたことを示した．抽出したパスワードリストはサイズの面でもパ

スワードの特徴の面でも多様だったが，大半は，これまでのセキュリティ分野での研究などでよく指摘される頻出パスワードリストのパスワード群を多く含んでいた．同時に，頻出パスワードリストには含まれないがマルウェアがパスワード候補として保持すると思われる文字列群も複数把握することができた．分析を通じて得られたこれらの情報は今後パスワードリストによる総当たり攻撃を検出，防止する上で有用であると考えている．

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 小久保 博崇、江田 智尊、大山 恵弘	４．巻 62
２．論文標題 データ欠損を起こしたマルウェアの機械学習による名称同定および悪性判定	５．発行年 2021年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 1879～1892
掲載論文のDOI（デジタルオブジェクト識別子） 10.20729/00214231	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 小久保博崇，大山恵弘	４．巻 61
２．論文標題 マルウェア検体のデータ欠損がアンチウイルスによるマルウェア同定に与える影響の調査	５．発行年 2020年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 1414～1427
掲載論文のDOI（デジタルオブジェクト識別子） 10.20729/00206790	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Yoshihiro Oyama and Hirotaka Kokubo	４．巻 -
２．論文標題 Forced continuation of malware execution beyond exceptions	５．発行年 2022年
３．雑誌名 Journal of Computer Virology and Hacking Techniques	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s11416-022-00457-8	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計8件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 大山恵弘
２．発表標題 Soliton Dataset 2021におけるマルウェアによる解析回避処理の調査
３．学会等名 コンピュータセキュリティシンポジウム 2021
４．発表年 2021年

１．発表者名
鈴木克弥，大山恵弘

２．発表標題
SleepHop: 動的バイナリ計装によるマルウェアのタイミング攻撃の無効化

３．学会等名
コンピュータセキュリティシンポジウム 2021

４．発表年
2021年

１．発表者名
イボット アリジャン，大山恵弘

２．発表標題
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

３．学会等名
2022年暗号と情報セキュリティシンポジウム

４．発表年
2022年

１．発表者名
小久保博崇，大山恵弘

２．発表標題
ドキュメント化されていないヘッダを活用した機械学習によるマルウェア分類

３．学会等名
2022年暗号と情報セキュリティシンポジウム

４．発表年
2022年

１．発表者名
小久保博崇

２．発表標題
データ欠損を起こしたマルウェアの機械学習による同定

３．学会等名
コンピュータセキュリティシンポジウム 2020

４．発表年
2020年

１．発表者名
玉林亜喬

２．発表標題
マルウェアの動的解析回避処理の傾向についてのSoliton Dataset 2020の分析

３．学会等名
コンピュータセキュリティシンポジウム 2020

４．発表年
2020年

１．発表者名
イボットアリジャン

２．発表標題
IoTマルウェアの分類における画像化を用いた手法とシステムコール列を用いた手法の比較

３．学会等名
情報処理学会 第186回マルチメディア通信と分散処理・第92回コンピュータセキュリティ合同研究発表会

４．発表年
2021年

１．発表者名
鈴木克弥

２．発表標題
クロック分解能を用いたRaspberry Pi仮想マシンの検出

３．学会等名
情報処理学会 第186回マルチメディア通信と分散処理・第92回コンピュータセキュリティ合同研究発表会

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
システムセキュリティ研究室 研究成果
https://syssec.cs.tsukuba.ac.jp/wp/index.php/work/

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関