マルウェアによる感染活動の目的推定のための トラヒックデータとAPIログに基づく分析

鮫島 礼佳[1,a)]　村上 純一[2,b)]　吉浦 裕[1,c)]　市野 将嗣[1,d)]

概要：近年，インターネットを介した攻撃が増加し，被害が深刻なものとなってきている．インターネットを介した攻撃は，一般的にマルウェアが用いられる．攻撃の数が増えると同時に，マルウェアの種類や数も日々増加して行く中，従来の様にマルウェア一つ一つを解析し，それぞれのマルウェア用に対策を講ずるのでは，対策が追いつかない．そこで，攻撃の目的が事前にわかれば，その目的の部分を重点的に防御できるような対策を講ずることができるのではないかと考え，攻撃の目的推定を最終目標と定める．その最終目的に向けて，まず，攻撃に利用されるマルウェアによる感染活動の目的がわかれば，攻撃に使用されるマルウェア次第で攻撃自体の目的も推定できるのではないかと考える．そこで，本研究では，攻撃の目的推定という最終目標に向けての初期検討として，トラヒックデータと API ログを用いたマルウェアの目的分類を行う．その結果，いくつかのマルウェアの目的を読み取ることが出来たので，ここに報告する．

キーワード：マルウェア，目的分類，トラヒックデータ，API ログ

Analyze of Traffic Data and API Log for Estimating Object of Malware Infection Activity

Ayaka Samejima[1,a)]　Junichi Murakami[2,b)]　Hiroshi Yoshiura[1,c)]　Masatsugu Ichino[1,d)]

Abstract: In recenr years, increase in attack by way of Internet become a serious problem. Attack by way of Internet are used by Malware. As number of attack are increasing, number of Malware are increasing. But conventional method, considering measure per Malware, can not catch up with new Malware. We think that if we recognize purpose of attack, we can consider better measure. We think that if we recognize purpose of Malware, we can recognize purpose of attack. So, we clustering Malware by purpose from analyzing traffic-data and API-log. As a result, we could read some purpose of malware.

Keywords: Malware, Object Classification, Traffic Data, API Log

1. はじめに

近年，個人向け組織向けに関わらず，インターネットを介した攻撃が増加し，被害が深刻なものとなってきている．インターネットを介した攻撃は，一般的にマルウェアを用いて行われるため，攻撃の増加に伴いマルウェアの数も年々増加している [1]．現在の攻撃対策として，アンチウイルスソフトの利用や，攻撃に利用されそうな通信を制限するフィルタリングルールの導入，人力による 24 時間の通信監視，等が挙げられる．しかし，攻撃の数が増え，マルウェアの種類や数が日々増加して行く中，従来の様にマルウェア一つ一つを解析し，それぞれのマルウェア用に対策を講ずるのでは，対策が追いつかず，防御が薄くなって

[1] 電気通信大学大学院情報理工学研究科
The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan
[2] 株式会社 FFRI
FFRI Inc.
1-18-18 Ebisu, Shibuya-ku, Tokyo 150-0013, Japan
[a)] a.samejima@uec.ac.jp
[b)] murakami@ffri.jp
[c)] yoshiura@hc.uec.ac.jp
[d)] ichino@inf.uec.ac.jp

しまう．

そこで，本研究では，攻撃の目的が事前にわかれば，その目的の部分を重点的に防御できるような対策を講ずることができるのではないかと考え，攻撃の目的推定を最終目標と定める．攻撃の目的を推定するためには，既に発生している攻撃の目的を知っておく必要がある．インターネットを介した攻撃は，一般的に攻撃の目的達成の手段として．マルウェアが用いられる場合が多い．このことから，あるマルウェアが攻撃においてどのような役割を担っているか，何を成すために作成されたマルウェアなのか，というマルウェアの目的がわかれば，攻撃に使用されるマルウェア次第で攻撃自体の目的も推定できるのではないかと考える．そこで，本研究では，攻撃の目的推定という最終目標に向けての初期検討として，マルウェアによる感染活動から目的を推定するために，トラヒックデータと API ログからマルウェアを目的毎に分類し，データの分析を行った．

以下，2 章ではマルウェアの分類や攻撃推定を行っている既存研究の紹介と，本研究の目的を述べる．3 章ではマルウェアの目的を分析した実験について述べ，その結果を 4 章に示す．結果に関する考察は 5 章で述べ，最後に 6 章でまとめと今後の課題，展望について述べる．

2. 先行研究と研究目的

本章では，本研究の目的であるマルウェアの目的による分類とは異なるが，マルウェアへの感染検知の観点でトラヒックデータと API ログを用いたマルウェアの解析と，分類を行っている既存研究と，攻撃者の意図や能力の推定を行っている既存研究について述べる．まず 2.1 節で既存研究の紹介を行い，2.2 節で既存研究の課題や有効性を挙げ，それを基に本研究での目的を述べる．

2.1 先行研究の紹介

本節では，まず本研究の目的であるマルウェアの目的による分類とは異なるが，マルウェアへの感染検知の観点でマルウェアを解析し，分類を行っている既存研究を紹介する．次に，攻撃の早期推測や，攻撃の意図や能力の推定を行っている既存研究を紹介する．

まず，マルウェアへの感染検知の観点で，マルウェアの解析，分類を行っている既存研究を紹介する．マルウェアの解析，分類を行っている研究として，トラヒックデータを用いた解析を行っている研究と，API ログを用いて解析を行っている研究がある．トラヒックデータを用いてマルウェアの解析を行っている研究として，桑原ら [2] は，トラヒックデータから抽出した通信の特徴を用いて，マルウェアの感染の有無や，感染パターンの分類を行った．千葉ら [3] は，トラヒックデータから抽出した可変的または不変的な通信の特徴を用いて，マルウェア感染時の通信内容の分類とテンプレート化を行い，実際にテストデータを検知することでそのテンプレートの有効性を示した．

API ログを用いてマルウェアの解析を行っている研究として，青木ら [4] は，API の関数名のみを用いて検体の分類を行い，ベンダーの定義による分類と比較することでその有効性を示した．武部ら [5] は，API コール列を特徴ベクトル化し，SVM により機械学習を行うことで亜種マルウェアの推定を行った．川古谷ら [6] は，データの送受信に関わる API と，それによって呼び出される API との依存関係を用いて，送受信されたデータの内容の特定を行った．

次に，攻撃の早期推測や，攻撃の意図や能力の推定を行っている既存研究を紹介する．川北ら [7] は，ソーシャルメディア上の情報を集め分析することで，攻撃の兆候を早期に捉えられることを示した．実際に，早期に兆候を捉えることで，攻撃への対処を迅速に行い，攻撃を受ける期間を短縮できたことから，その有効性が示されている．芦野ら [8] は，インターネット上に存在するノイズを観測し分析することで，そのインターネットノイズがただの意味の無いノイズではなく，何らかの意図を持った攻撃の一種であることを発見した．インターネットノイズの挙動や通信データを解析することで，その通信の意図や攻撃システムを推測できることを示した．渡部ら [9] は，上級セキュリティ技術者の個々のノウハウをデータ化し共有しやすくするために，マルウェアの解析を行った．実際にマルウェア感染時のトラヒックデータ分析する事で，マルウェアの活動，機能，目的をそれぞれ分類し，それらを紐付けることで，攻撃の意図やマルウェアの背景にある攻撃活動の知見をナレッジとしてデータ化して扱えるようになることを示した．

2.2 研究目的

2.1 節で紹介したマルウェアへの感染検知の観点でマルウェアを解析し，分類を行った研究 [2], [3], [4], [5], [6] より，トラヒックデータもしくは API ログを用いることで，マルウェアの分類が出来ることが示されている．しかし，これらのマルウェアの分類は検知の観点で行われており，マルウェアの目的による分類ではない．

また，攻撃者の意図や能力の推定を行った研究 [7], [8], [9] より，攻撃によって発生する情報から，攻撃の意図や目的，能力を推定することで，攻撃を早期に検知または事前に攻撃の発生を推定出来ることが示されている．このことより，攻撃の目的を推定し，攻撃の早期検知または事前の攻撃検知を行うことで，事前に対策を練ることができると考える．そのためには，実際に発生した攻撃から，攻撃側の情報を分析する必要がある．近年では攻撃の手段として一般的にマルウェアが用いられている．マルウェアを目的毎に分類することができれば，そのマルウェアを用いた攻撃の目的推定に繋がると考える．マルウェアの感染活動は通

信だけでなく端末内部にも表れる．そのため，先行研究のように通信データからだけでなく，API ログも用いて端末内部の挙動も分析し，目的を読み取って行く必要があると考える．

そこで，本研究では，攻撃の目的推定による事前対策を目標に，マルウェアによる感染活動から目的を推定するという観点から，感染後のトラヒックデータと API ログの解析を行い，マルウェアの目的による分類を行った．

3. 実験

本章では，解析を行った結果，読み取ることができたマルウェアの目的について述べる．

3.1 使用データ

本研究では．株式会社 FFRI 提供のマルウェア感染時のデータ 931 検体を用いて解析，分類を行った．このデータは，2015 年 12 月から 2016 年 3 月までに取得したマルウェア検体を，Cuckoo Sandbox2.0-dev で実行した際のトラヒックデータと，API ログのデータである．

3.2 実験方法

本節では，本研究で 3.1 節で述べたマルウェア検体の感染時トラヒックデータと API ログを分析した結果，推測できたマルウェアの主な目的とその特徴を述べる．

2 章で述べた通り，マルウェアを目的毎に分類し，その分類の意味，つまりマルウェアの目的を列挙している先行研究は見つけられなかった．そこで，本研究でマルウェア感染時のトラヒックデータと API ログを分析した結果推測できたマルウェアの目的を以下に列挙する．

- 内部設定操作系
  - (1) データの暗号化
  - (2) 通信経路の確保
  - (3)bot 化
- 外部との通信系
  - (4) ファイルのダウンロード
  - (5) 感染拡大
  - (6) 環境調査

上記の目的を分類するにあたって，まずは通信にマルウェアの挙動が表れているか否かで，目的 (大) によって検体を分類した．通信に挙動が表れていないものは，内部設定操作系，として，API ログから目的を推測した．また，通信にマルウェアの挙動が表れているものは，外部との通信系，として，まずはトラヒックデータから目的を推測し，次に API ログからより具体的にマルウェアの目的を推測した．

4. 結果

本章では，3 章で述べた実験の結果を述べる．3.1 節で述べた 931 検体を目的毎に分類した際の内訳を**表 1** に示す．

表 1 マルウェアの目的による分類
Table 1 Object Classification of Malware

目的 (大)	目的 (小)	検体数
内部設定操作系	(1) データの暗号化	112
	(2) 通信経路の確保	108
	(3)bot 化	18
外部との通信系	(4) 感染拡大	61
	(5) ファイルのダウンロード	66
	(6) 環境調査	16

5. 考察

本章では 4 章で示した結果を基に考察を行う．それぞれの目的を推測するにあたって，まずトラヒックデータの解析を行った．はじめに各検体のトラヒックデータから，DNS クエリの発生の有無と，DNS クエリが存在する場合，検体毎にクエリにあるホストの種類数を調べた．次に，API ログから，ファイルやレジストリの書き換え，変更，削除の有無と，そのような API が発生していた場合は発生回数とその対象となるファイルやレジストリのパスを調べた．各目的に分類されたマルウェアのトラヒックデータと API ログから読み取れた特徴を**表 2** に示す．

今回，推測できた目的のうち，特に (1) データの暗号化と (2) 通信経路の確保について，表 2 の特徴に示した部分をグラフで表した．(1) データの暗号化の特徴を**図 1** に，(2) 通信経路の確保の特徴を**図 2** に示す．図 1，図 2 をみると，同じ目的であっても検体によって，ファイル書き換え系 API の出現回数と，DNS クエリの数とに偏りがあることがわかる．この偏りは，同じ目的を達成する為のマルウェアであっても，目的達成のための手法が異なることを表しているといえる．つまり，この偏りが類似しているものは，一つのマルウェアから発生した同じ目的を持つ亜種であると考えられる．

6. まとめ

本研究では，攻撃の目的推定という最終目標に向けた初期検討として，マルウェア感染時のトラヒックデータと API ログを用いたマルウェアの感染活動による目的の分析を行った．そして，今回の分析で推測できたマルウェアの感染活動による目的を，各目的毎に統計的に分析し，考察を行った．その結果，マルウェアの感染活動による目的毎に，発生する API の属性やその回数，DNS クエリ等の外部との通信の有無に，傾向があることがわかった．

本研究は，まずマルウェアの感染活動による目的にどのようなものがあるか知るための分析のため，目的の推測を自動化できていないのが問題点と言える． 今後は，より多くの検体を分析できるように，マルウェアの感染活動による目的推測の自動化アルゴリズムの提案を行っていく．

表 2　マルウェアの目的毎のデータ的特徴
Table 2　Object Classification of Malware

目的 (大)	目的 (小)	特徴
内部設定操作系	(1) データの暗号化	RANSOM_CERBEA 系，CRYPAURA 系，CRYPTESLA 系，CRYPWALL 系 …ファイルの書き換え API 多数 (2500〜6500 回程) http 通信少数 (0〜10 回) RANSOM_LOCKY 系 …ファイルの書き換え API(0 回または 400 回程)，http 通信少ない (0〜10 回) RANSOM_CRILOCK 系 …ファイルの書き換え API 少ない (4 回程)，http 通信多数 (30〜70 回) RANSOM_CRYPCTB 系 …ファイルの書き換え API ほぼ無し (0〜1 回)，http 通信無し (0 回)
	(2) 通信経路の確保	BKDR_BLADABI 系 …ファイル書き換えほぼ無し (0〜5 回)，レジストリキー書き換え多数 (20〜200 回) http 通信ほぼ無し (0-3 回) RAP 系 …システム構成の把握系 API 多数
	(3)bot 化	TROJ_GEN 系 …DNS クエリ多数 (多数のホスト名にクエリを投げている) …レジストリキー書き換え多数 (30〜300 回)
外部との通信系	(4) 感染拡大	TROJ_FORCON 系 …同じネットワークグループの探索多数 ファイルの共有，ユーザー設定書き換え系の API 多数
	(5) ファイルのダウンロード	TROJ_DYER 系，TROJ_FORCON 系 …ネットワークの設定書き換え API，http 通信多数 (GET 多数) TROJ_FRS 系 …COM 書き換え API 多数
	(6) 環境調査	TROJ_FRS 系 …ネットワーク環境調査 API 多数

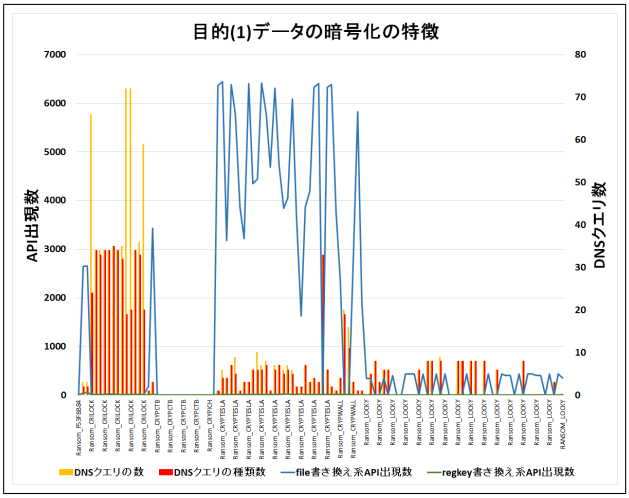

図 1　目的 (1) データの暗号化
Fig. 1　Object(1) Cypher of Data

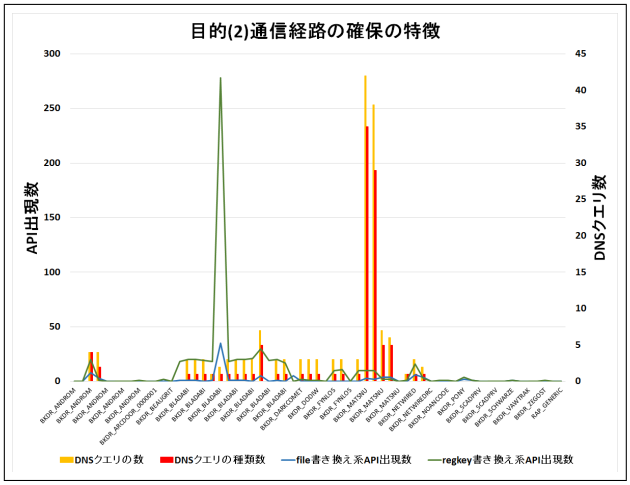

図 2　目的 (2) 通信経路の確保
Fig. 2　Object(2) Establishment of Communication Path

また，最終目標である攻撃の目的推定に向けて，マルウェアの感染活動による目的毎の分類手法の提案や，分類後のマルウェアを用いた攻撃全体の目的推定の手法の提案等を行っていく．

謝辞　本研究は JSPS 科研費 15H01684 の助成を受けたものです．

参考文献

[1] McAfee Labs : McAfee Labs 脅威レポート 2016 年 6 月 (オンライン), http://www.mcafee.com/jp/resources/reports/rp-quarterly-threats-may-2016.pdf.

[2] 桑原和也，菊池浩明，寺田真敏，藤原将志：パケットキャプチャーから感染種類を判定する発見的手法について，コンピュータセキュリティシンポジウム (2009).
[3] 千葉大紀，八木毅，秋山満昭，青木一史，針生剛男：感染後通信検知のための通信プロファイリング技術の設計と評価，コンピュータセキュリティシンポジウム (2014).
[4] 青木一樹，後藤滋樹：API コール情報を利用したマルウェアの階層型分類，電子情報通信学会総合大会 (2016).
[5] 武部嵩礼，後藤滋樹：Paragraph Vector を利用した亜種マルウェア推定法，電子情報通信学会総合大会 (2016).
[6] 川古谷裕平，塩治榮太朗，岩村誠，針生剛男：API コール間のデータ依存関係を利用したマルウェア通信内容の特定，コンピュータセキュリティシンポジウム (2013).
[7] 川北将，島成佳：テクニカル分析によるサイバーインシデント予測手法の検討，電子情報通信学会信学技報 (2016).
[8] 芦野佑樹，島成佳：インターネットノイズに対する偽装応答機能の実装と観測に基づいた意図が不明なリクエストに関する考察，暗号と情報セキュリティシンポジウム (2015).
[9] 渡部正文，島成佳，田辺瑠偉，吉岡克成：マルウェア活動の背景にある攻撃者の意図に関する知見のナレッジ化方式，コンピュータセキュリティシンポジウム (2014).