マルウェアの可視化とその応用に関する研究

松重 雄大 †　浦辻 和也 ‡　甲斐 博 ‡　森井 昌克 †

† 神戸大学大学院工学研究科
657-8501 兵庫県神戸市灘区六甲台町 1-1
matsushige@stu.kobe-u.ac.jp
mmorii@kobe-u.ac.jp

‡ 愛媛大学大学院理工学研究科
790-8577 愛媛県松山市文京町 3
uratsuji@hpc.cs.ehime-u.ac.jp
kai@cs.ehime-u.ac.jp

あらまし 近年，マルウェアの増加に伴い解析結果の数も膨大になり，マルウェアの特徴や類似性を把握することが困難となっている．本論文ではマルウェア同士の相違点を明確にすることを目的とし，マルウェアの解析結果より一意に定まる可視化 3D モデルを提案する．マルウェアの動作内容を機能として分類し，マルウェアを複数の機能の集合体としてモデル化することにより，マルウェアの特徴を反映した可視化が可能になる．3D モデルを用いた応用としてマルウェアの流行状況の可視化システムを提案する．時間軸上の動的な可視化により，日本国内における 1 日単位でのマルウェアの遷移を把握することが可能である．

A Study on Visualization of Malware and Its Applications

Takahiro Matsushige†　Kazuya Uratsuji‡　Hiroshi Kai‡　Masakatu Morii†

†Graduate School of Engineering, Kobe University
1-1, Rokkodai-cho, Nada-ku, Kobe-shi, Hyogo 657-8501, JAPAN
matsushige@stu.kobe-u.ac.jp
mmorii@kobe-u.ac.jp

‡Graduate School of Science and Engineering, Ehime University
3 Bunkyo-cho, Matsuyama-shi Ehime 790-8577, JAPAN
uratsuji@hpc.cs.ehime-u.ac.jp
kai@cs.ehime-u.ac.jp

Abstract In recent years, the number of analysis results has becomes enormous with an increase of malware. It is difficult to determine the similarity and features of malware. In this paper, the purpose is to clarify the differences between malware. We propose a visualization 3D model uniquely determined from the analysis of malware. Visualization by classifying the behavioral description of the malware as functions and by modeling malware as a collection of multiple functions reflecting the characteristics of malware. We propose the epidemic situation visualization system of malware as an application using the 3D models. It is possible to grasp the translation of malware on a daily basis in Japan by a dynamic visualization on the time axis.

1　はじめに

マルウェアへの対策を行うために静的解析手法や動的解析手法など解析手法の研究が盛んに行われている．大手のセキュリティソフトベンダでは自動解析システムを用いた網羅的な解析を行っており，解析結果がレポートとして公開されている．しかしマルウェアは複雑な動作を行うことが多く，解析結果を読み解くためには高度な専門知識と多くの時間を要する．またマルウェアの増加に伴い解析結果の数も膨大になり，マルウェアの特徴や類似性を把握することが困難となっている．

マルウェアの中には既存のマルウェアの一部のみを変更したものが存在することから，同じ特徴を持つマルウェアを亜種として分類することで解析結果を整理できる．マルウェアを一定の評価基準の下で分類することで，マルウェア間の差異や類似性の検証以降の研究を効率的に行うことができると考えられる．しかしマルウェアの明確な分類基準はなく，注目する基準によって分類結果が異なる．マルウェアの分類に関する研究として，マルウェアの感染時の動作に着目した分類 [1] や静的解析結果と API の組み合わせによる分類 [2]，マルウェアの挙動情報を指標として利用する分類 [3] 等がある．またマルウェアの挙動情報と API ログを組み合わせて指標とする分類 [4] がある．

本論文ではマルウェアの動作内容を機能として分類し，マルウェアを複数の機能の集合体として可視化する方法を提案する．ファイルの参照や書き換え等がそれぞれ個別の機能となる．マルウェアの解析結果として ESET 社の解析結果 [5] を利用する．機能分類の結果はトロイの木馬やワーム等の種類を元に作成する 3D モデル上に各機能に対応する 3D モデルを配置する．類似した機能を持ったマルウェア同士の 3D モデルは類似することになる．

また可視化モデルを用いたマルウェアの流行状況の可視化システムを提案する．類似した特徴を持ったマルウェアが流行しているか直観的に把握可能になると考えられる．時間軸上での可視化のため，日本国内の 1 日単位でのマルウェアの遷移情報を得ることが可能である．

2　マルウェアの分類

2.1　既存の分類

マルウェアはコンピュータへの侵入方法や動作の特徴から，ワームやトロイの木馬等の種類に大別される．しかし種類分類以降の詳細な分類について明確な定義は存在しない．詳細な分類を行うことで詳細な動作内容を把握できるだけでなく，亜種間の違いを明確にすることが可能になる．

マルウェアの分類に関する一つの指標としてセキュリティソフトベンダによるマルウェアの命名がある．しかし命名方法に厳密な定義はなく命名の際に注目する点が異なることから，セキュリティソフトベンダごとに名前が異なるマルウェアは多い．命名後に適切な名前に変更する場合もあるが，適切な分類であるかを判断する有効な評価基準が存在しないため，名前のばらつきは解消されていない．

2.2　機能への対応

一般にマルウェアは感染活動，ネットワークへのアクセス等特定の動作を行うものが多い．感染活動として OS やソフトウェア，ディレクトリ構造などを調査し，データの書き換えや追加を行う．本論文ではマルウェアの持つ特徴を機能群とそれの属する機能に分類する．マルウェアの解析結果からマルウェアの特徴情報を取得し各機能に分類することにより，特定の機能の組み合わせによる詳細な分類が可能であると考えられる．マルウェアの機能群を以下の 4 種類に分類した．

情報収集　マルウェアは感染や個人情報の取得等を行う際に対象のコンピュータの情報を取得する必要がある．ファイル，ソフトウェアのバージョン，パスワード等を収集するための機能をこの機能群に分類する．「特定ファイルの調査」「ユーザ情報の収集」「パスワードの収集」を機能として定義する．

感染活動　マルウェアの流通，ボットネットの作成や維持を行う機能をこの機能群に分類す

る.「ファイル，レジストリ，DLL の作成」「ファイル，レジストリ，DLL の書き換え」「ファイル，レジストリ，DLL の削除」を機能として定義する.

破壊活動 データやシステムの破壊や外部からの侵入経路を構築するバックドアの作成の機能をこの機能群に分類する.「バックドアの作成」「システムの破壊」を機能として定義する.

外部への操作 特定サイトへのアクセスやネットワーク上の他のコンピュータへのアクセス，メールの配信等を行う機能をこの機能群に分類する.「サイトへのアクセス」「メールの送信」「サーバへのログイン」「ダウンロード」「ネットワーク共有」を機能として定義する.

3 マルウェアの機能分類

マルウェアの解析結果からマルウェアの特徴を抽出し，機能群に分類する．マルウェアの特徴を調査するため ESET 社のマルウェア解析結果 [5] を用いる．マルウェアの種類，動作内容（機能），機能の対象数を取得する.

マルウェアの種類は解析結果上で Category に示されている文字列を用いる．マルウェアはコンピュータへの侵入方法においてワーム型，トロイの木馬型に 2 分可能であることから，本論文では上記 2 種類を対象とする.

マルウェアの動作内容は複数行の文章で記述されている．またマルウェアの機能に続けてファイル名等の機能の対象が列挙されている．文章から機能を抽出するために形態素解析プログラム TreeTagger[6] を用いて単語に分解し，各機能に対応する単語が存在するか確認する．機能の分類と対応する単語を表 1 に示す.

機能及び機能対象は以下のステップで抽出する.

ステップ 1 TreeTagger を用いて文を単語に分解し，各単語を原型に変換する.

表 1: 機能の分類と対応単語

機能群	機能	対応単語
情報収集	特定ファイルの調査	search
	ユーザ情報の収集	collect
	パスワードの収集	password
感染行動	ファイル，レジストリ，DLL の作成	create copy
	ファイル，レジストリ，DLL の書き換え	modify add
	ファイル，レジストリ，DLL の削除	delete
破壊活動	バックドアの作成	backdoor
	システムの破壊	end
外部への動作	サイトへのアクセス	connect
	メールの送信	mail
	サーバへのログイン	server
	ダウンロード	download
	ネットワーク共有	network

ステップ 2 文を表 1 に対応する単語数が最も多い機能へ分類する．対応する単語がない場合はステップ 3 に移動する.

ステップ 3 文を機能対象へ分類する．ステップ 1 からステップ 3 を繰り返すことにより，機能と機能対象のみが抽出される.

文中に表 1 に対応する単語がない場合，機能対象へ分類されることから，機能に対応する単語を増やすことで機能分類の精度向上が期待できると考えられる.

4 マルウェアの 3D モデル化

解析結果から抽出した種類と機能に 3D モデルを与え，マルウェアを 3D モデルで表現する．種類から与えられる 3D モデル上に機能から与えられる 3D モデルを配置する．種類と 3D モデルの対応はトロイの木馬型をクラゲ型，ワーム型をナマコ型とする．機能と 3D モデルの対応を表 2 に示す.

同じ機能群に属する機能の形状は同一であり，異なる色で表現される．機能の 3D モデルの作成数は機能数 1 と機能の対象数の和で決定される.（ある機能において機能対象が存在しない場合，作成されるモデル数は 1 個である.）

表 2: 機能と 3D モデルの対応

機能	形状	色
特定ファイルの調査	楕円形	赤
ユーザ情報の収集		緑
パスワードの収集		青
ファイル，レジストリ，DLL の作成	円錐	赤
ファイル，レジストリ，DLL の書き換え		緑
ファイル，レジストリ，DLL の削除		青
バックドアの作成	お椀型	赤
システムの破壊		緑
サイトへのアクセス	つの型	赤
メールの送信		緑
サーバへのログイン		青
ダウンロード		黄
ネットワーク共有		シアン

マルウェアの 3D モデル化の例を図 1 に示す．

例として示したマルウェアはトロイの木馬であり，クラゲ型で作成されている．機能分類の結果を表 3 に示す．表 2 より「ダウンロード」は「黄色のつの型」,「ユーザ情報の収集」は「緑色の楕円形」,「サイトへのアクセス」は「赤色のつの型」であり，上面から螺旋状に配置されていることが確認できる．

5　マルウェア流行状況の可視化

5.1　流行状況把握の課題

マルウェアの流行状況を把握するためには多くの情報が必要となる．

感染率　マルウェア流行状況の指標となる．

時間　時間推移を確認する上で重要な情報である．

マルウェア間の類似点　特徴の類似したマルウェアが流行しているか調べる際には，特徴を直観的に把握できる必要がある．

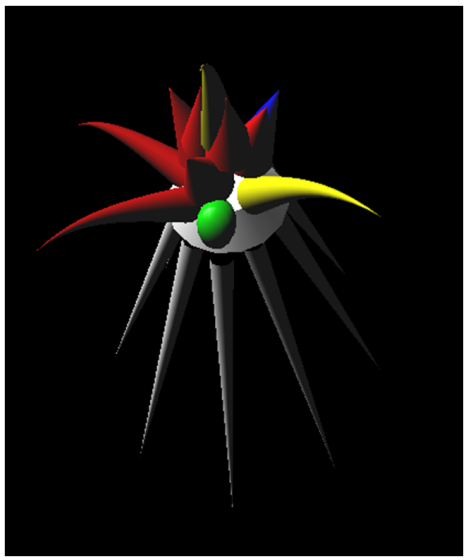

図 1: 3D モデル化の例

感染率及び時間は二次元グラフで表現可能であるが，個々のマルウェアの特徴を考慮することは困難である．前節で述べたマルウェアの 3D モデルを用いることで，マルウェアの特徴の類似点を把握し易い流行状況の可視化が可能になると考えられる．

5.2　流行状況データ

マルウェア流行状況のデータとして ESET 社の統計情報を利用する．1 日ごとの統計情報が日々更新されていることから，統計情報を蓄積することで流行状況の時間推移を 1 日単位で捉えることが可能になる．統計情報には国別の感染率 (Infection ratio) とマルウェア別の感染率がある．本論文では日本国内に着目した可視化を提案する．国別の感染率を元に可視化するモデル数を，マルウェア別の感染率を元に可視化するマルウェアの種類及び割合を決定する．日本国内の感染率が増加すれば表示されるモデル数も増加することで，直観的に把握することが可能になる．

表 3: 機能分類結果

機能	機能モデル数
ダウンロード	1
ファイル，レジストリ，DLL の作成	2
ファイル，レジストリ，DLL の作成	2
ファイル，レジストリ，DLL の作成	1
ファイル，レジストリ，DLL の削除	1
ダウンロード	1
ユーザ情報の収集	1
サイトへのアクセス	2

5.3 提案システム

提案可視化システムは図 2 に示す通り，取得部と可視化部から構成される．

取得部では定期的に ESET 社の統計情報にアクセスし，HTML 形式のデータを取得する．取得したデータから国別の感染率及びマルウェア別の感染率を抽出し，データベースに保存する．またデータに含まれる各マルウェアの特徴情報にアクセスし，取得データに対して機能分類を行う．機能分類の結果も同様に保存する．感染率の高いマルウェアの中には特徴情報が存在しない場合がある．詳細な解析が今後行われ特徴情報が公開された際には，保存している情報を更新する必要がある．

可視化部ではデータベースに保存されたデータを取得し，各マルウェアの 3D モデルを作成する．3D モデルは前節と同様に，種類のモデルと機能のモデルで構成する．国別の感染率（R）から全体のモデル数 (M) を，マルウェア別の感染率 (r_i) から各マルウェアのモデル数 (m_i) を以下の式により設定する．ただし M，m_i はモデル数のため，小数点以下切り捨てとする．

$$M = 10R \tag{1}$$

$$m_i = \frac{r_i}{\sum_{i=1}^{n} r_i} M \tag{2}$$

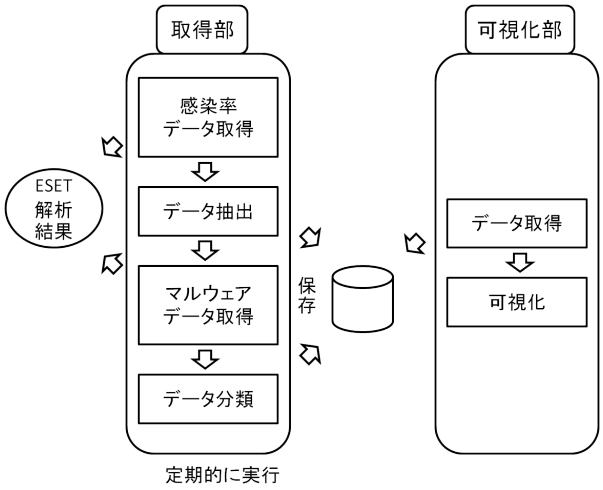

図 2: 可視化システムの流れ

例えば日本国内の感染率が 1.95%の場合，19 個のモデルを作成する．各マルウェアのモデル数の比は各マルウェアの感染率の比とほぼ同一になる．

一定時間おきに可視化対象となるデータを更新し動的な可視化を行うことで，流行状況の時間推移を確認することができる．日本以外の国に対しても同様に流行状況のデータを取得することで，各国のマルウェアの遷移を比較することも可能になると考えられる．

6 まとめ

本論文ではマルウェア解析結果に対して機能分類を行い，3D モデルで可視化する手法を提案した．文を単語に分割し，機能に対応する単語が最多の機能に文を分類していくことで，マルウェアを機能の集合体と捉えた可視化を行った．機能ごとのモデルの形状及び色は一意であることから，機能の類似したマルウェア同士は 3D モデルが類似することになる．

また 3D モデルを利用した流行状況の可視化システムを提案した．マルウェアの特徴を反映した 3D モデルを用いることで，機能の類似性を考慮した流行状況の把握が可能になると考えられる．また時間軸上の動的な可視化であることから，1 日単位でのマルウェアの遷移を把握し易いと考えられる．

今後はマルウェア流行状況の可視化システムを実装し，時間推移の中でマルウェアの特徴を直観的に把握可能であるか評価を行う．また機能分類の精度向上のため，マルウェアの機能分類に用いる単語の再定義を行う予定である．

参考文献

[1] 名坂康平，酒井崇裕，山本匠，竹森敬祐，西垣正勝，“自動実行登録に基づくマルウェアの分類に関する検討，”情報処理学会研究報告，2010-CSEC-50-40，pp.1-5，2010.

[2] 岩本一樹，和崎克己，“静的解析によるマルウェアの分類と結果の検討，”マルチメディア，分散，強調とモバイル (DICOM2010) シンポジウム，pp.477-491，2010.

[3] 堀合啓一，今泉隆文，田中英彦，“マルウェア亜種の動的挙動を利用した自動分類手法の提案と実装，”情報処理学会論文誌　50(4)，pp.1321-1333，2009.

[4] 正力達也，伊沢亮一，森井昌克，“マルウェアの挙動情報を用いたマルウェア分類システムんの提案と実装，”2011 年情報セキュリティシンポジウム (SCIS2011)，2A2-3，2011.

[5] ESET Virusradar
http://www.virusradar.com/en/home/world

[6] TreeTagger
http://www.cis.uni-muenchen.de/ schmid/tools/TreeTagger/

[7] 浦辻和也，松重雄大，甲斐博，森井昌克，”Malware visualization based on the behavior and its classification”，第 13 回情報科学技術フォーラム (FIT2014)，2014.