7W-06

画像特徴量を用いた未知マルウェアの検知・分類アルゴリズムに関する検討

小寺　建輝†　房安　良和†　泉　隆†

日本大学†

１．　はじめに

近年，IoT への注目に伴い，IoT デバイスを標的としたマルウェアが出現している．例えば，2016 年秋には「Mirai」と呼ばれるマルウェアが出現し，多くの IoT デバイスが感染の被害にあった．さらに，Mirai は作成者によってソースコードが公開されており，それを改変した亜種が大量に作成されている．これに対し，ウイルス定義ファイルを用いてマルウェアを検知する従来のパターンマッチング法では，パターンが定義されていない亜種を検知することは難しい．このような問題を解決するため，機械学習により亜種を検知・分類する研究がある．その中でも，Windows 系マルウェアを画像化し，画像認識によって亜種を該当するファミリに分類する研究[1]では，高い識別精度でマルウェアを分類できたことが報告されている．これは，亜種が元のコードの一部のみを改変して作成されるため，元のマルウェアとその亜種，つまり同一ファミリでは視覚的に類似した画像が得られるためである.そこで本研究では，画像認識により亜種を分類する技術を Mirai のように亜種を多く含む IoT マルウェア(Linux 系マルウェア)の検知・分類へと応用することについて検討を行う．

本稿では，同一ファミリ等の類似したマルウェアの画像同士をグループ化して検知モデルを構築し，各検知モデルを利用して亜種等のマルウェアを検知するアルゴリズムを検討する．

２．　マルウェアの画像化[1]

マルウェアを画像化する手法を以下に示す．また実際にマルウェアをファミリごとに画像化した例を図 1 に示す．

(1) 対象ファイルを 1Byte(8bit)ずつ読み込み 1 次元配列に格納する．
(2) ファイルサイズ(配列の要素数)に応じて幅を決定し，1 次元配列から 2 次元配列へ変換する．
(3) 配列の要素の値は 1Byte であり，0-255 の範囲となるため，その値を画素値として 256 階調のグレースケール画像を生成する．
(4) 画像の幅に合わせて画像を最近傍法により正方形にリサイズする．

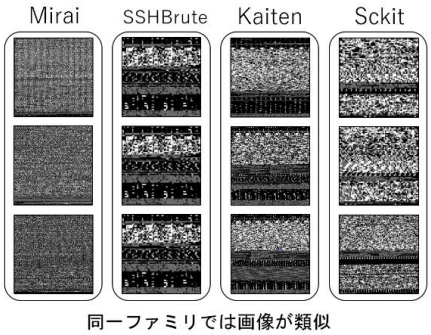

図 1. 各ファミリにおけるマルウェアの画像

３．　マルウェア検知アルゴリズム

検知モデルの学習及びそのモデル利用してマルウェアを検知するアルゴリズムを以下の４つのフェーズに分けて説明する．

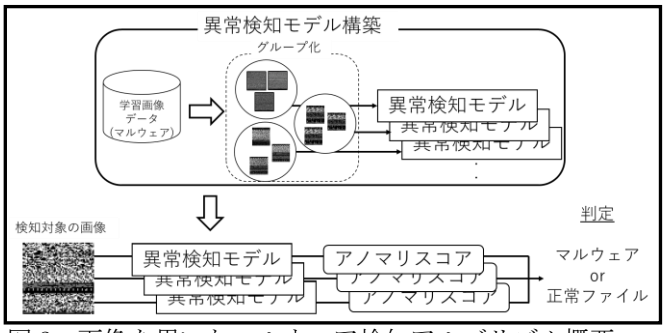

図 2. 画像を用いたマルウェア検知アルゴリズム概要

(1) 類似した画像のグループ化
(2) 異常検知モデルの構築　　学習
(3) アノマリスコアの算出
(4) マルウェアの判定　　検知

(1) 類似した画像のグループ化

学習データであるマルウェアにファミリ名のラベルを割り当て，ファミリごとにグループ化を行う.他にも，クラスタリングを用いることにより，画像特徴量が類似したもの同士をグループ化する手法が考えられる．

(2) 異常検知モデルの構築

検知対象の画像が(1)で作成したグループに属するか否かをアノマリスコア(異常度)により判定する異常検知モデルを各グループで構築する.また，判定のために各モデルでアノマリスコアに閾値を設定する．

“A study on Detection and Classification algorithm of unknown Malware by using Image feature”
† Tateki Kodera ,Yoshikazu Husayasu ,Takashi Izumi・Nihon University

(3) アノマリスコアの算出
　検知対象の画像を各モデルに入力し，アノマリスコアを算出する.
(4) マルウェアの判定
　検知対象の画像のアノマリスコアがあるモデルで閾値未満であった場合，そのモデル(グループ)に該当するマルウェアであると判定する. また，全てのモデルにおいてアノマリスコアが閾値以上であった場合，正常ファイルと判定する.

4． 実験

本実験では, ファミリごとに画像をグループ化し構築した各ファミリの異常検知モデルにおいて，各モデル(グループ)に該当するマルウェア(同一ファミリ＝亜種)を入力した時の検知率と，正常ファイルを入力した際に当該グループのマルウェアと判定する誤検知率ついて検証を行った.

ここで，モデルの構築及び検知率の評価に 27 ファミリ 1399 検体のマルウェアを，誤検知率の評価に 7783 検体の正常ファイルを利用して 10 分割交差検証を行う. また，モデル構築のための画像特徴量に 320 次元の Gist 特徴量[2]，学習アルゴリズム及びアノマリスコアの算出に Isolation Forest[3]を採用した. 各モデルにおけるアノマリスコアの閾値は，任意に決定した閾値候補に対して検知率と誤検知率による F 値を求め，F 値が最大となるときの閾値を採用している.

実験結果を表 1 に，各ファミリのモデルにおけるアノマリスコアの例を図 3 に示す.

表 1. 実験結果

グループ名 (ファミリ名)	検体数	検知率(%)	誤検知率(%)
Hajime	17	100	2.3
Mare	20	95	0.1
Ganiw	246	93.9	1.3
Grip	41	90.2	3.9
Gafgyt	368	88.6	6.5
Ramen	30	86.7	3.7
Chapro	14	85.7	0
SSHBrute	13	84.6	0
Scalper	13	84.6	10.5
Mayday	32	78.1	0
Mrblack	111	77.5	0.9
Race	40	77.5	11.7
SSHscan	12	75	0
Kaiten	109	73.4	19.2
Darlloz	15	73.3	5.4
Brk	22	72.7	0.5
Dcom	20	65	17.1
Sorso	14	64.3	0.3
DNSamp	31	58.1	1.4
Phobi	13	53.8	5.6
Matrics	27	48.1	7.6
Lotoor	39	46.2	24.8
Lion	22	45.5	20.7
Openssl	16	43.8	0.1
Mirai	70	40	17.1
Telf	31	29	8.8
Sckit	13	7.7	3.3

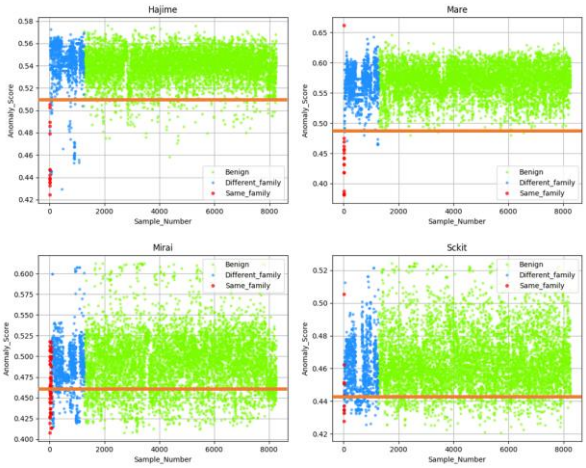

図 3. 各ファミリのモデルにおけるアノマリスコア (赤:同一ファミリのマルウェア，青:異なるファミリのマルウェア. 緑:正常ファイル，橙:閾値)

表 1 より，Hajime のモデルでは，検知率 100%であり，同一ファミリの全ての亜種を検知できた. さらに，約 3 分の 1 のファミリのモデルでは，80%以上の検知率で亜種を検知できた. また，検知率が高いモデルは誤検知率が低い傾向にあった. これは，図 3 の上 2 つ(Hajime や Mare)のモデルのように，同一ファミリと正常ファイルのアノマリスコアの差が明確なためである. このようなモデルでは，アノマリスコアに閾値を設定することで亜種の検知が可能であると考えられる. 一方，一部のファミリのモデルでは検知率が低い結果となった. これは，図 3 の下 2 つ(Mirai や Sckit)のようなモデルのように，同一ファミリと正常ファイルのアノマリスコアに差が現れなかったためである. これについては，学習データのマルウェアに割り当てたファミリ名に誤りがあり，正しいグループに分類されなかったことや Mirai のようにマルウェアごとに動作対象の CPU アーキテクチャが異なる場合，同一ファミリであっても画像の類似度が低くなることが原因と考えられる. つまり，類似した画像のグループ化に問題があると考えられるため，クラスタリングを用いて，画像特徴量の類似性に基づいたグループ化を行い，改善する必要がある.

5． まとめ

本稿では，マルウェアの画像をもとにマルウェアの亜種を検知する異常検知モデルを検討し，一部のモデルにおいて有用性を確認することができた.

今後は，クラスタリングを用いることで，類似した画像のグループ化の改善を行う.

参考文献

[1] L. Nataraj, et al. : ”Malware Images:Visualization and Automatic Classification”, VizSec’11(2011-07)
[2] A. Olivia and A. Torralba : ” Modeling the shape of a scene: a holistic representation of the spatial envelope”, Intl. Journal of Computer Vision, Vol.42, No.3, pp.145–175(2001)
[3] Fei Tony Liu, et al. : “Isolation-Based Anomaly Detection”, ACM Transactions on Knowledge Discovery from Data(TKDD), Vol.6, No.1, pp.1-39(2013-03)