Computer Security Symposium 2015
21 - 23 October 2015

感染挙動の時系列情報のクラスタリングに基づくマルウェア検知手法

鮫島 礼佳 †　　畑田 充弘 ‡　　吉浦 裕 †　　市野 将嗣 †

† 電気通信大学大学院情報理工学研究科
182-8585 東京都調布市 調布ケ丘 1 丁目 5-1
a.samejima@uec.ac.jp, yoshiura@hc.uec.ac.jp, ichino@inf.uec.ac.jp

‡NTT コミュニケーションズ株式会社
108-8118 東京都港区芝浦 3-4-1 グランパークタワー 16F
m.hatada@ntt.com

あらまし 近年インターネットの普及に伴い, マルウェアが日々開発され, 感染による被害が拡大している. そこで感染後の端末を解析することで, 感染を早期に検知し拡大を防ぐ, 感染検知という手法が注目されている. 本研究ではマルウェア感染時のトラヒックデータをクラスタリングし, 時間変化を考慮した上で検体毎の相関を求め、クラスタリングすることで分析を行っている. その分析結果から, マルウェアの挙動による分類の有効性や, 未知マルウェアの検知手法を提案し, 評価する.

Method for detecting Malware based on clustering of time series information of infection behavior

Ayaka Samejima†　　Mitsuhiro Hatada‡　　Hiroshi Yoshiura†　　Masatsugu Ichino†

†The University of Electro-Communications.
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, JAPAN
a.samejima@uec.ac.jp, yoshiura@hc.uec.ac.jp, ichino@inf.uec.ac.jp

‡NTT Communications Corporation
Gran Park Tower 16F, 3-4-1 Shibaura, Minato-ku, Tokyo, 108-8118, JAPAN
m.harada@ntt.com

Abstract In recent years, increase in damage by malicious software commonly called“ malware ”become a serious problem. So new detection method which detect malware after infection is paid attention. Terminals infected with malware must have peculiar communications. We can detect malware with these peculiar communications from traffic data. We clustering traffic data gotten from terminal infected with malware, and we analyzed each cluster, and read behavior of malware from traffic data. As a result, we could prove that if I can make models of cluster each common behavior of malware, we may be able to detect unknown malware.

1　はじめに

近年, 仕事や生活等の様々な場面でインターネットが必要不可欠な存在となっている. しかし, インターネットが普及し, ユーザの利便性が向上する一方で, マルウェアによる被害が年々増加し問題となっている [1].

従来より行われているマルウェア感染への主な対策として, マルウェア侵入時に検知を行う侵入検知がある. 侵入検知の一般的な手法であるシグネチャ型検知は, 既に解析済みのマルウェアから得た特徴的なシ

グネチャとパターンマッチングを行う事で検知を行う手法である [2]. シグネチャ型の検知手法は, 解析済みのマルウェアの検知率が高い一方, シグネチャの事前登録が必要なため, 未解析のマルウェア, つまり未知のマルウェアの検知は困難という問題がある. 特に, 未知のマルウェアの発生数は年々増加しており, この問題はより深刻となっている. そこで, 未知のマルウェアへの対策として, 感染することを前提とし, 感染後の被害を最小限に抑えるために, 端末が感染していることを検知する感染検知が必要とされている.

一般的にマルウェアに感染した端末は, 感染した端末のネットワーク環境の調査等を行うため, マルウェア感染時特有の通信を行う. そこで本研究では, トラヒックデータを分析し, この感染時特有の通信の挙動を利用することで, 未知のマルウェアの感染検知を行うことを目標とする. そのため, まずは, トラヒックデータから感染時の挙動を表す特徴を抽出する必要がある. そこで本研究では, トラヒックデータの時間的変化を考慮しクラスタリングを行う事で, 感染検知を行う手法を提案する.

以下, 2 章ではトラヒックデータを用いたマルウェアの感染検知を行っている既存研究とその問題点を挙げ, 3 章でその先行研究における問題点に対して本研究で提案する改善手法を説明する. 4 章, 5 章では実際に行った実験の手順とその結果を述べる. そして最後に 6 章で結果に基づいて得られた知見とその考察を行う.

2 先行研究

本章ではトラヒックデータを用いたマルウェアの感染検知を行っている研究事例を紹介する.

宮本らは, ヘッダ情報の種類毎のパケット数を特徴量とし, SVM を利用した異常検知手法の提案を行った [3] . ヘッダ情報の種類毎にそれぞれ 1 分区間のパケット数を算出し SVM を利用して異常検知を行った.

川元らは, ヘッダ情報から得られた 36 種類の特徴量について, 感染検知における有効性の評価を行った [4]. 検知方法は, まず感染と正常のトラヒックデータをそれぞれクラスタリングし、正常コードブックと感染コードブックを作成した. その後テストデータを与え, 最も距離の近いコードブックが正常か感染かによって識別を行った.

市田らは, ヘッダ情報から得た 3 種類の特徴量を元に, 特徴量の時間的状態遷移を考慮した検知手法を提案した [5]. その結果, 時間的状態遷移を考慮した検知手法は, 考慮しなかった検知手法と比較して, 高精度な検知率を示す, といったことが確認された.

大月らは, ペイロード情報から, 既存研究で用いられていた 261 種類の特徴量に対して, 感染検知における有効性の評価を行った [6]. その際, セキュリティベンダーの定義したマルウェアの種類 (ワーム, トロイの木馬, ファイル感染型ウイルス) 毎に評価を行った. その結果, ワームに対しては 4 種類, ファイル感染型ウイルスに対しては 5 種類, トロイの木馬に対しては 15 種類の有効な特徴量を発見した.

これらの先行研究によって, トラヒックデータには正常と感染が分離できる特徴量が含まれており, 更に市田らの研究により, 時間的変化は検知に有効である事が示されている. ここで, トラヒックデータとは, 通信の性質を保有するデータであり, 通信の目的次第で保有する情報の性質は異なる. そのため, ひとつの特徴や性質としてまとめることが難しい. しかし, これらの先行研究では感染後のトラヒックデータを分類せず, 全て一つのクラスタとして扱っている. 大月らは感染後のトラヒックデータを分類してはいるが, セキュリティベンダーの定義に基づいた分類を行っている. セキュリティベンダーの定義はトラヒックデータの挙動に基づいている訳では無く, 感染する OS の種類や, 世間に流布した名称等を用いている場合もあり, さらに同じ検体でもベンダーによって命名が異なる場合がある [1]. そのため, まだ挙動が判明せず, ベンダーに定義されていない未知のマルウェアを検知する際には最適とはいえない. よって感染後の挙動による通信に基づいたトラヒックデータの分類手法と, それを用いた感染検知手法の提案が必要である.

3 提案手法

本研究では, 2 章で述べた先行研究の問題点に対する改善策として, 感染挙動による通信のクラスタリングに基づくマルウェア感染検知手法を提案する. 本章では感染後のトラヒックデータのクラスタリングに

ついて説明する.

3.1 基本方針

本研究では以下の方針を提案する.

(i) 通信のパターンを表すために, 有効とされる特徴量をトラヒックデータから抽出しクラスタリングする.

(ii) 時間変化をクラスタ間の移動で表す. (i) で作成したクラスタ間の各移動パターンの回数 (例えば, クラスタ 1 からクラスタ 2 へ移動した回数) を一定時間毎にカウントし, その回数を特徴として用いる. こうすることで感染挙動による細かな変化を含んだ形で表現することが可能となる. 長期間の変化をみる際には, その一定時間毎のカウントを繰り返し行い, 一定時間毎のカウント数の変化を見る事で, 感染挙動全体の傾向を表現することが出来る.

(iii) 通信パターンの似た検体同士をまとめるため, 時間変化に伴う検体毎のトラヒックデータの類似性を求める. そのために, まず (ii) で, 検体毎にカウントした各移動パターンの回数を用いて, 相関分析を適用し, 検体同士の相関係数を求める. この際, 一般的に感染挙動は検体によって異なるため, 検体によってトラヒックデータの長さが異なる. 相関係数は比較する 2 検体の長さが同じでないと算出できないため, DP マッチングを用いて長さを揃えてから相関係数を求める.

(iv) 感染のトラヒックデータを通信パターンの類似性の高いもの同士で同じクラスタに分類するために, (iii) で算出した相関係数を用いて階層的クラスタリングを行う. その結果, 各クラスタの代表となる検体を通信の典型パターンとする.

3.2 提案手法の概要

(i) 通信パターンの表現方法

まず教師データとなる全てのトラヒックデータに対して, トラヒックデータを 1 秒区間で区切り特徴量を抽出したタイムスロットを用いて, ベクトル量子化を行う. これにより, 正常も感染も含む全ての教師データの通信のパターンをクラスタで表現することが可能となる.

(ii) 時間変化の表現方法

次に時間の変化に伴ってタイムスロットが所属するクラスタの変化を, クラスタ間の移動として調査する. これによって, 通信のパターンの時間変化を表すことが可能となる. しかし, タイムスロットである 1 秒毎の移動ではクラスタ間の移動パターンにばらつきが出来, 移動パターンの特徴を抽出するのが困難といえる. そこで, 30 秒区間の移動, つまりタイムスロット 30 個分の移動の中で, 各移動パターンが表れる回数をカウントし, 大まかな時間変化を含む移動パターンとして表す. これをヒストグラムを用いて表すと図 1 のようになる. 図 1 の横軸はクラスタ番号を 0〜3 とした際の移動パターン (0-0, 0-1, ..., 3-2, 3-3) を, 縦軸は 30 秒間に生じた各移動パターンの回数の割合を表している. そして更に長期間の移動パターンを分析するため, 移動パターン毎に, トラヒックデータが続いている限り, 30 秒区間毎の調査を繰り返し行う. つまり, 図 1 の様なヒストグラムに時間軸が追加され, 30 秒毎に図 1 の様なヒストグラムを追加して行くことで表せる. これを図 2 に示す. 30 秒間の変化を 1 枚のヒストグラムで表現し, さらに長い間の変化を図 2 に示す様にヒストグラム間の変化で表現する.

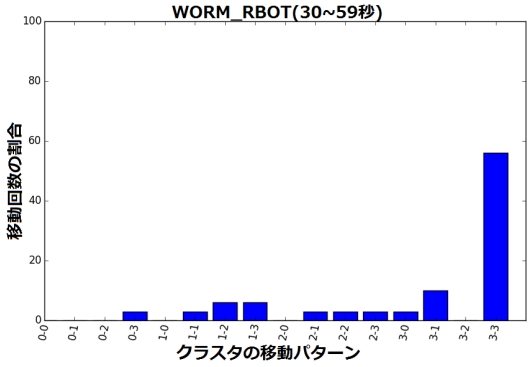

図 1, 30 秒区間の各移動パターンの回数

(iii) 相関係数の算出方法

次に各検体の類似性を分析するため, 検体毎に (ii) を行う. その結果算出される移動パターンの回数を用いて, 相関係数を求める. 相関係数を求めるためには, 類似性を求めたい 2 つの変数の長さが等しくなえればならない. しかしこの提案手

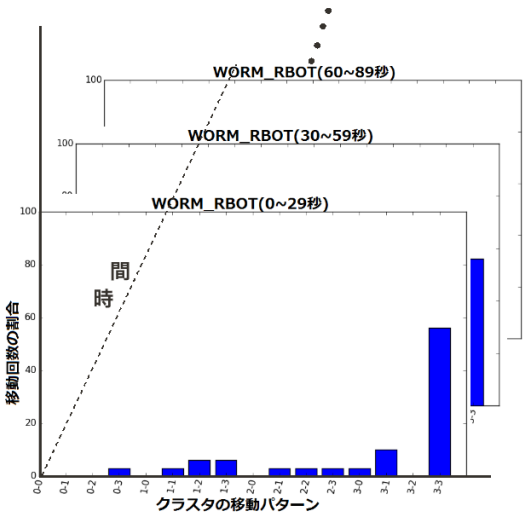

図 2, 長期間の移動パターンの遷移

法では, トラヒックデータの長さが異なる場合, 移動パターンを表す 30 秒区間の区間数も異なってしまうため, このままでは相関係数を算出する事が出来ない. そこで DP マッチングを用いて, 類似性を分析したい 2 検体のデータ数を合わせてから, 相関係数を算出する. DP マッチングを用いる理由は, DP マッチングは時系列情報を考慮したまま長さを合わせることが可能なためである.

相関係数とは, 2 つの変数の間の類似性を示す指標であり, 以下の式を用いて算出できる.

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

DP マッチングは図 3 の様にペナルティを設置することでそのペナルティが最小となる経路を辿り, データを適宜複製して行くことで, 全体としての時系列情報を保持したまま長さを揃える手法である.

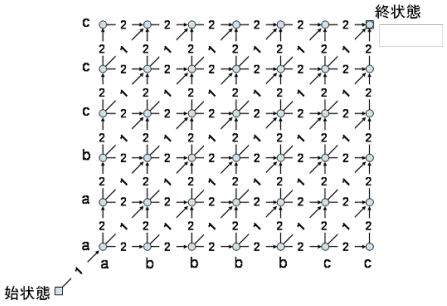

図 3, DP マッチングの例

(iv) 典型パターンの作成方法

全ての検体に関して, 2 検体ずつの組み合わせを総当たりで (iii) を行うことで算出した相関係数を, 階層的クラスタリングを用いてクラスタリングする. 感染検体同士の相関係数をクラスタリングすることで, 感染時の典型的な通信パターンを表すことができる. この典型パターンは各クラスタの代表となる検体の挙動で表す. これにより, 先行研究の様に感染を一つのクラスタとして識別を行った場合に比べ, 識別精度が向上することが期待できる.

ここで, 階層的クラスタリングを用いる理由として, 通信パターンは最終的にいくつのクラスタに分類できるか不明である. しかし, 階層的クラスタリングでは結果となるクラスタ数が不明のままクラスタリングを行うことが可能なため, この手法を用いる.

4 実験

提案手法の有効性を示すため、以下の実験を行った. まずは感染データ全体を使用した場合の有効性を示し, 次に, 感染データの始めの 30 秒のみを使用した早期検知の場合の有効性を示す.

4.1 実験使用データ

本研究では, マルウェア感染時の通信の挙動として MWS Datasets として提供されている, CCC DATAset 2009, 2010, 2011 と, D3M 2012, 2013, 2014, 2015, PRACTICE Dataset 2013, BOS 2014 を利用した [7]. 正常時の通信はあるイントラネットで取得した 2011 年から 2015 年のデータを用いた.

4.2 典型パターンの作成手順

本節では教師データを用いて感染挙動の典型パターンを作成した際に行った実験の手順について, 3.2 で述べた提案手法の手順に沿って説明する. 教師データは 4.1 で述べたデータのうち, 感染は MWS Datasets の CCC2009, 2010, 2011 のハニーポッド 1 台分, D3M2012, BOS2014, PRACTICE2013 の計 320 検体, 正常は 2011 年 2015 年のデータを用いた.

(i) 通信パターンの表現方法

手順 1. 感染データをマルウェアの検体, 年度, 取得環境毎に分割した.

手順 2. トラヒックデータを 1 秒毎に区切り, 先行研究より, 有効とされる 4 種類の特徴量を抽出したタイムスロットを作成した [8]. この特徴量とは, パケットサイズの最小, SYN パケット数, TCP パケットに対する SYN パケットの割合, ACK パケット数の 4 種類である.

手順 3. タイムスロットに含まれる各特徴量の値を Min-Max 法を用いて正規化し、4 種類の特徴量を四次元で LBG-Splitting アルゴリズムを用いてベクトル量子化を行った. この際, クラスタ数は 4 とし, 各クラスタの通信状態を表すコードブックを作成した.

(ii) 時間変化の表現方法

手順 4. 手順 1 で分割した各検体と, 手順 3 で作成したクラスタから, 3.2 の (ii) の手法を用いて時間変化の情報を表した. タイムスロット 30 個分, つまり 30 回の移動毎に, 4 つのクラスタの移動, 計 16 通りの移動パターンが行われる回数をカウントし, 割合として算出した. その結果を図 1 の様なヒストグラムを用いて表した.

手順 5. 手順 4 で算出した 30 秒間で行われる各移動パターンの回数の割合を, 30 秒毎の時系列順に追った. この結果をクラスタ間移動の 16 通り分, 図 2 の様に表した.

(iii) 相関係数の算出方法

手順 6. ここでは, 感染データ全体から DP マッチングを用いて相関係数を求める場合と, 早期発見のため感染データの始めの 30 秒のみを用いて相関係数を求める場合の 2 通りの実験を行った.

(1) 手順 5 で求めた 30 秒毎の移動パターンの回数を全て用いて, 各検体を 2 種類ずつの組み合わせ総当たりで DP マッチングを行った後に相関係数を算出した.

(2) 同様に, 手順 5 で求めた 30 秒毎の移動パターンの回数のうち, 各検体のトラヒックデータの始めの 30 秒のみの移動パターンの回数を用いて, 相関係数を算出した.

(iv) 典型パターンの作成方法

手順 7. 手順 6 で算出した相関係数を, 階層的クラスタリングを用いてクラスタリングした. 相関係数は一般的に 0.2 以下が相関が低く, 0.7 以上が相関が高いといわれている. そこで本実験では, 階層的クラスタリングを行う際の代表として, 全検体との相関係数のうち, 相関係数が 0.7 以上である検体数が最も多い検体を代表と定めた. また, クラスタリングの際に, 同じクラスタに所属する検体数の 70% 以上の検体と相関係数が 0.2 以上でなければ, そのクラスタには分類しない様にした.

教師データのうち, 正常データに関しても, 上記の手順 2〜5 までを同様に行った. 尚, 正常データの 16 通りの移動パターンは, 手順 5 で算出した移動パターンの回数が, 全ての正常データの平均に最も近い時間帯のデータ 300 秒分とした.

4.3 評価実験の手順

本節では, 4.1 で述べた学習の結果をを基に, 本研究の提案手法である感染挙動のクラスタリングの有効性を評価するための実験手順について述べる. テストデータは CCC2011 のハニーポッド 1 台分と D3M2013, 2014, 2015 の計 200 検体を用いた.

手順 8. 4.2 の手順 3 で作成したクラスタのコードブックを基に, テストデータを最もユークリッド距離の近いコードブックを持つクラスタに分類した.

手順 9. 4.2 の手順 4, 5 と同様に 30 秒毎にクラスタ間移動 16 通りの回数をカウントした.

手順 10. (1) 4.2 の手順 7 で作成した各クラスタの典型パターンと, 正常データと DP マッチングを用いて累積最小距離を求めた. その結果, 最も距離の近いグループが感染か正常かによってテストデータの識別を行った.

(2) 更に、グルーピングを行う事の有効性を評

価すべく, 4.2 の手順 7 を行わなかった場合, つまり感染と正常それぞれ 1 グループずつの場合も手順 11 と同様に累積最小距離を求め, テストデータの識別を行った.

5 実験結果

5.1 相関分析の結果

4.2 で述べた 320 検体を 2 検体ずつの組み合わせ, つまり

$$_{320}C_2$$

で 51040 通りの組み合わせで算出した相関係数を用いて相関分析を行った. 以下の図 4 は縦軸, 横軸共に 320 検体をそれぞれ表し, その相関係数をマトリックス形式で表したものである.

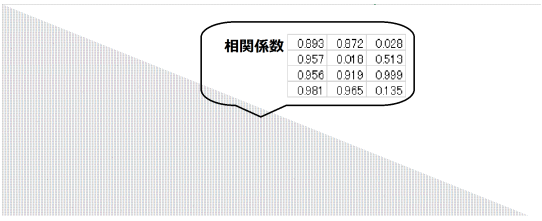

図 4, 相関分析の結果

5.2 評価実験の結果

5.2.1 教師データをグルーピングした場合

DP マッチングを用いて階層的クラスタリングによるグルーピングを行った結果, 感染は 13 グループに分割できた. 教師データ 320 検体の内訳は 210, 62, 20, 14, 4, 2, 2, 1, 1, 1, 1, 1, 1 検体ずつとなった. また, DP マッチングを用いずトラヒックの始めの 30 秒間のみを用いて同様にグルーピングを行った結果, 感染は 6 グループに分割できた. 教師データの内訳は 277, 28, 12, 1, 1, 1 検体ずつとなった. これら各グループに正常 1 グループを追加して評価実験を行った結果を以下に示す. ここで, 識別率とは, 感染検体 200 検体に対して, 正しく感染であると識別された検体数の割合を表す.

表 1 感染データをグルーピングした場合の識別結果

使用した教師データ	全てのデータ 13+1 グループ	始めの 30 秒 6+1 グループ
識別率	98.5%	98%
Malware	197	196
Normal	3	4

5.2.2 感染データをグルーピングしなかった場合

階層的クラスタリングを用いず, 感染 1 グループ, 正常 1 グループとして評価実験を行った結果を以下に示す.

表 2 グルーピングを行わなかった場合の識別結果

使用した教師データ	全てのデータ 1+1 グループ	始めの 30 秒 1+1 グループ
識別率	15%	19.5%
Malware	30	39
Normal	170	161

6 考察

本章では 5 章で得られた結果から, 各クラスタの通信の挙動の性質や, 時間変化の捉え方を考慮しつつ, 感染後のトラヒックデータのクラスタリングに関する考察を行う.

6.1 クラスタリングの有無に関する考察

本節では, 感染データ 320 検体をクラスタリングしたことの有用性について考察を行う.

表 1, 表 2 より, 感染データ全てを用いた場合, 階層的クラスタリングを行ったことによって識別率は 15% から 98.5% まで向上した. また, 感染データの始めの 30 秒のみを用いた場合も同様に, 19.5% から 98% まで向上した. 以上の結果より, 感染データをクラスタリング場合の方がしなかった場合に比べて識別率が圧倒的に高くなることがわかった. ここで, 各感染クラスタの代表データと正常の代表データとの累積最小距離を算出し, クラスタリングを行った場合と行わなかった場合を比較してみた. まず, 感染の全データを用いた場合について考察する. クラスタリ

ングを行って作成した 13 クラスタのそれぞれの代表と, 正常との累積最小距離をそれぞれ算出し, 平均を求めたところ, 約 41.5 となった. クラスタリングを行わずに感染と正常の各代表の距離を求めた所, 約 26.0 となった. 次に, 感染データの始めの 30 秒のみを用いた場合について考察する. クラスタリングを行って作成した 6 クラスタのそれぞれの代表と, 正常との累積最小距離をそれぞれ算出し, 平均を求めたところ, 約 26.8 となった. クラスタリングを行わずに感染と正常の各代表の距離を求めた所, 約 3.8 となった.

以上の結果より, クラスタリングを行わない場合の方が, 行った場合に比べ, 正常クラスタと感染クラスタの累積最小距離は近くなっていることがわかった. 距離が近いということは, 類似性が高いということを示す. つまり, 評価実験を行った際に誤検知をしやすいといえる. これが, クラスタリングを行った場合の識別率が, 行わなかった場合に比べて圧倒的に高くなっていることの理由であると考えられる.

6.2 教師データの長さに関する考察

本節では, 各検体の感染データを全てのトラヒックデータを用いて相関係数を算出した場合と, 通信の始めの 30 秒のみを用いて相関係数を算出した場合の識別率について考察を行う.

6.1 の考察で精度が良いと判断した, 感染データのクラスタリングを行った場合に関して, 表 1 より, 始めの 30 秒のみを用いた場合の識別率が 98% であるのに対して, 全てのデータを用いた場合は 98.5% と向上している. これより, 初めの 30 秒でも高い精度で識別できており, 感染後の長いトラヒックデータを用いることでさらに高い精度で識別できるとわかった.o この結果の理由について考察を行う. まず, 感染データの始めの 30 秒のみで識別を行う事の利点は, 感染後出来る限り短い時間で識別する事が可能となれば, 今後, 感染後の早期検知を行える可能性があるためである. しかし, 今回の結果より, 感染データは感染直後のトラヒックにのみ感染の特徴が現れるとは限らないということがいえる. 感染直後だけでなく, 感染終了時や, 時間的変化に伴う感染全体の挙動から抽出できる感染特有の挙動が存在するため, 感染後の全てのデータを用いた方が識別率が高くなったのだと考えられる.

6.3 クラスタの性質に関する考察

本節では, 感染データをクラスタリングした際, 各クラスタの代表となっている検体のクラスタ間移動のヒストグラムを比較, どこに偏っているか考察. また正常ともヒストグラムレベルでの比較を行う.

ここで以下の表で挙げている各クラスタの代表の検体名はセキュリティベンダーが定義した検体名に基づいている.

表 3　感染データをクラスタリングした場合の正常データとの相関係数

クラスタ番号	所属検体数	代表の検体
0	210	TROJ_FAM
1	62	WORM_DOWNAD
2	20	WORM_RBOT
3	14	WORM_DOWNAD
4	4	WORM_DOWNAD
5	2	WORM_SDBOT
6	2	TROJ_KRYPTIK
7	1	WORM_DOWNAD
8	1	WORM_DOWNAD
9	1	BKDR_SMALL
10	1	WORM_ALLAPLE
11	1	TSPY_ZBOT
12	1	TROJ_MAILBOT

表 4　感染データをクラスタリングしなかった場合の正常データとの相関係数

クラスタ番号	所属検体数	代表の検体
0	277	PE_VIRUT
1	28	WORM_RBOT
2	12	WORM_DOWNAD
3	1	TROJ_FAM28c
4	1	TROJ_FAKEAV
5	1	BKDR_SMALL

上記結果のうち, 同じクラスタ内に, セキュリティベンダーの定義では異なった検体名に定義されているものが多々含まれている. さらにテストデータも, 定義名は異なっていても, 類似した挙動を行う検体が

代表となっているクラスタに分類されているパターンが確認できた. このことから, ベンダーの定義に限らず, 感染挙動によってまとまるクラスタが作成できていることがいえる.

7 まとめ

本論文ではトラヒックデータを用いた, 感染挙動による通信のクラスタリングに基づくマルウェア感染検知手法を提案し, その提案手法に対して識別率を評価した. その結果, 3 章で予測した通り, 感染挙動の通信をクラスタリングした場合, クラスタリングを行わなかった場合に比べて識別率が向上することがわかった. この結果により, 感染挙動の通信をクラスタリングすることの有効性が示せた.

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] 御池鮎樹, マルウェア-情報化社会の破壊者-, 工学社, 2009/09/15, ISBN:978-4-7775-1468-7

[2] 佐々木良一, ネットワークセキュリティ, 電子情報通信学会編, オーム社, 2014, ISBN : 978-4-274-21517-9

[3] 宮本貴朗ら, SVM を用いたネットワークトラヒックからの異常検出, 電子情報通信学会論文誌, Vol. B, 通信 J87-B4, pp593-598, 2004

[4] 川元研治ら, マルウェア感染検知のための経年変化を考慮した特徴量評価に関する一考察, コンピュータセキュリティシンポジウム CSS2011/10, pp277-282

[5] 市田達也ら, 特徴量の時間的な状態遷移を考慮したマルウェア感染検知手法に関する一検討, 暗号と情報セキュリティシンポジウム SCIS2012/01

[6] 大月優輔ら, マルウェア感染検知のためのトラヒックデータにおけるペイロード情報の特徴量評価, マルウェア対策研究人材育成ワークショップ MWS2012/10

[7] 神薗雅紀ら, マルウェア対策のための研究用データセット MWS Datasets 2015 , 情報処 理学会研究報告 Vol.2015-CSEC-70 No.6, Vol.2015-SPT-14 No.6, 2015/07/02

[8] Masatsugu Ichino. et al, Evaluating header information features for malware infection detection, Journal of Information Processing, Vol.23, No.5, , 2015/09(掲載決定).