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

川元　研治 †　　市田　達也 †　　市野　将嗣 ‡　　畑田　充弘 †‡　　小松　尚久 †

† 早稲田大学理工学術院基幹理工学研究科
169-8555　東京都新宿区大久保 3-4-1
{kawamoto, ichida, komatsu}@kom.comm.waseda.ac.jp

‡ 電気通信大学大学院情報理工学研究科
182-8585　東京都調布調布ヶ丘 1-5-1
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 17F
m.hatada@ntt.com

あらまし 本研究では，マルウェア感染検知の既存研究でよく用いられている特徴量に対して，マルウェアに感染している感染トラヒックとマルウェアに感染していない正常トラヒックの識別実験により特徴量評価を行った．その際，特徴量毎にベクトル量子化で作成した正常時，感染時のコードブックとテストデータとの特徴空間上での距離を用いて識別を行った．本稿では，感染トラヒックデータとして CCCDATAset，正常トラヒックデータとして同じデータ収集日におけるあるイントラネットのトラヒックデータを使用して，年によらずマルウェア感染検知において有効である特徴量について考察した結果を報告する．

A study of feature evaluation considering effects of year for malware detection

Kenji Kawamoto†　Tatsuya Ichida†　Masatsugu Ichino‡　Mitsuhiro Hatada†‡
Naohisa Komatsu†

†Graduate School of Fundamental Science and Engineering, Waseda UNIversity
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, JAPAN
{kawamoto, ichida, komatsu}@kom.comm.waseda.ac.jp

‡Graduate School of Informatics and Engineering, The University of Electoro-Communications
1-5-1 Chofugaoka, Chofu-si, Tokyo, 182-8585, JAPAN
ichino@inf.uec.ac.jp

†‡NTT Communications Corporation
Gran Park Tower 17F, 3-4-1 Shibaura, Minato-ku, Tokyo, 108-8118 Japan
m.hatada@ntt.com

Abstract In this paper, we evaluated features used in existing researches based on the experiment that showed how each features could discriminate anomaly traffic that was infected with malware from normal traffic that was not infected with malware. In this evaluation, we made discriminations using the distance between the normal or anomaly codebook made by each features using vector quantization and test data. In this paper, we used CCCDATAset as anomaly traffic data, some traffic data on an intranet which was the same date as anomaly traffic data as normal traffic. Then we report our consideration which features are effective for malware detection with no relation to year effects.

1　本研究の背景と目的

昨今のインターネットの普及により，マルウェアの脅威が広がっている．マルウェアとは悪意のあるソフトウェア (Malicious Software) の略称であり，その被害は個人情報の流出やパソコンの乗っ取りというように我々の生活を脅かす存在となっている．文献 [1] によると 2011 年度上半期の日本国内での被害報告数は約 4 千件にものぼっており，活動が表面化しないボットネットによる被害や Web からの感染の増加，加えて日に日に新種のマルウェアが発生しているという現状で，早急に対策を講じる必要がある．

これまでの対策研究としては，文献 [2] で整理されているが，既知のマルウェアについての対策が中心であり，未知のマルウェアについての対策が不十分という問題がある．そこで本研究では，マルウェアに感染していない状態とマルウェアに感染している状態そのものに違いがあると考え，トラヒックデータを用いた未知のマルウェア感染検知に着目する．さらに，トラヒックデータには時間的な変化があり，時間的な変化に着目することで感染検知の性能が向上する可能性がある．例えばバイオメトリクスでは，発話時の唇動作個人認証において複数のアルゴリズムが提案されているが，時系列を使用しないアルゴリズムと使用するものを比較した際，後者の方が高い精度での認証が可能であることが示されている [3]．

本研究では，ネットワークユーザのプライバシーの問題を考慮し，インターネットとユーザ間のトラヒックデータのパケットペイロードは参照せずに，ヘッダ情報から様々な特徴量を抽出し，それらを識別器に入力することで感染の有無を判定する．しかし，ヘッダ情報から抽出する特徴量について，既存の研究では十分な評価が行われていない．そこで本稿では，感染トラヒックとして CCCDATAset2009, 2010, 2011[4](以下 CCC2009, CCC2010, CCC2011)，正常トラヒックとして同じデータ収集日におけるあるイントラネットのトラヒックデータを用いて，各特徴量の感染トラヒックと正常トラヒックを識別する能力を評価し，マルウェア感染検知に有効と思われる特徴量について検討する．さらに，3 年間の経年変化も考慮することで，年々移り変わるマルウェアや新種のアプリケーションによるトラヒックの変化があっても，大きな影響を受けずに，識別することができる特徴量について考察する．

以下では，既存研究で良く用いられている特徴量を紹介し，本稿で評価する特徴量について述べる．そして，各特徴量の識別能力を評価する実験を行い，識別率の高い特徴量についての考察を述べる．

2　既存の特徴量

既存のマルウェア感染検知やネットワーク異常検知に関する研究で用いられている特徴量を紹介する．

文献 [5] では，タイムスロット型の検出モジュールとフローカウント型の検出モジュールを組み合わせてネットワーク異常検知を行っている．タイムスロット型とは，一定の時間間隔 (タイムスロット) でのトラヒック流量をカウントし，各特徴量を抽出する方式である．タイムスロット型で用いられている特徴量には，TCP の各フラグの出現回数 5 種類や TCP，UDP，ICMP パケット数等がある．一方，フローカウント型とは，フロー毎に特徴を抽出する方式である．フローとは，プロトコル，送信元 IP アドレスとその送信元ポート番号，宛先 IP アドレスとその宛先ポート番号が同じパケット群である．フローカウント型で用いられている特徴量には，パケット数，フラグメントされたパケット数，同一ポート番号のフロー出現回数の逆数等がある．本研究ではタイムスロット型のネットワーク観測方式を採用する．なぜなら，実際の感染検知では，迅速なマルウェア検知が求められるため，同一フローのパケットを全て収集してからでないと各特徴量を抽出できないフローカウント型だと，リアルタイム性に欠けるためである．

文献 [6] では，ネットワークトラヒックから複数の通常状態を定義するクラスタリング手法を提案している．その際に，特徴量として ICMP，SYN パケット数，FIN パケット数，SYN,FIN 以外の TCP パケット数，UDP パケット数を 60 秒毎にカウントし，それらを正規化したものを用いている．

文献 [7] では，ボットと人間の通信挙動には差異があると考え，特定のホスト間におけるデータ送信時間間隔を特徴量として見ることにより，正常なクライアントとボットクライアントの差異を確認している．

既存研究より，特定のパケット数，到着間隔，TCP フラグ，ポート番号に関する特徴量がよく用いられていることがわかる．

表 1: 特徴量 36 種類

番号	特徴量 [単位]
1	パケット数
2	パケットサイズの総数 [byte]
3	パケットサイズの平均 [byte]
4	パケットサイズの最小 [byte]
5	パケットサイズの最大 [byte]
6	パケットサイズの標準偏差 [byte]
7	到着間隔の平均 [秒]
8	到着間隔の最小 [秒]
9	到着間隔の最大 [秒]
10	到着間隔の標準偏差 [秒]
11	SYN パケット数
12	FIN パケット数
13	PSH パケット数
14	ACK パケット数
15	RST パケット数
16	URG パケット数
17	SYN/ACK パケット数
18	FIN/ACK パケット数
19	PSH/ACK パケット数
20	RST/ACK パケット数
21	TCP パケット中の SYN パケット割合
22	TCP パケット中の FIN パケット割合
23	TCP パケット中の PSH パケット割合
24	TCP パケット中の ACK パケット割合
25	TCP パケット中の RST パケット割合
26	TCP パケット中の URG パケット割合
27	TCP パケット中の SYN/ACK パケット割合
28	TCP パケット中の FIN/ACK パケット割合
29	TCP パケット中の PSH/ACK パケット割合
30	TCP パケット中の RST/ACK パケット割合
31	ICMP 到達不能メッセージ数
32	UDP パケット数
33	送信元ポート番号が 69/UDP のパケット数
34	送信元ポート番号が 80/TCP のパケット数
35	送信元ポート番号が 110/TCP のパケット数
36	送信元ポート番号が 443/TCP のパケット数

3 識別実験

3.1 用いる特徴量

本研究では既存研究をもとに，パケットのヘッダ情報から取得できる情報およびその統計値を特徴量として用いる．本研究で検討対象とする特徴量 36 種類を表 1 に示す．

3.2 実験諸元

3.2.1 評価方法

本研究での感染トラヒックと正常トラヒックの識別方法について説明する．はじめに，ベクトル量子化を用いて，感染トラヒックのみを用いて学習を行った感染コードブックと，正常トラヒックのみを用いて学習を行った正常コードブックを予め作成する．今回は，各特徴量を個別に評価することが目的であるので 1 次元コードブックを作成した．トラヒックデータから特徴量を抽出する際のタイムスロット幅は，0.1 秒，1 秒，10 秒，100 秒の 4 種類とし，ベクトル量子化のアルゴリズムには，LBG+Splitting を用い，そのレベル数は 2，4，8，16，32 の 5 種類とした．そして，予め感染トラヒックか正常トラヒックかのラベル付けされた各特徴量毎の 1 次元テストデータを与え，テストデータと感染，正常コードブックとの特徴空間上でのユークリッド距離を計算し，感染コードブックとの距離の方が小さければ感染，正常コードブックとの距離の方が小さければ正常と識別している．

識別結果に対する評価指標として True Positive Rate(以下 TPR) と True Negative Rate(以下 TNR) を用いる．TPR は感染トラヒックを感染トラヒックと正しく識別できた割合である．TNR は正常トラヒックを正常トラヒックと正しく識別できた割合である．各特徴量について，2009 年，2010 年，2011 年のトラヒックデータを用いて，各タイムスロット毎に感染か正常か識別し，TPR，TNR をそれぞれ算出した．

3.2.2 使用したデータについて

本研究では，感染コードブック作成のための学習データに CCC2009，正常コードブック作成のための学習データに 3 月 13 日から 3 月 15 日の 2009 年のトラヒックデータを用いた．テストデータは，感染トラヒックに CCC2009, CCC2010，CCC2011，正常トラヒックに感染トラヒックと同じデータ収集日のトラヒックデータを用いた．

本研究では，感染トラヒックとして CCC2009, 2010, 2011 の攻撃通信データを使用した．しかし，これらの攻撃通信データにはマルウェアに感染するまでのトラヒックが含まれている．今回の特徴量評価実験では感染時のみのデータを用いる必要がある．そこで本研究では，攻撃通信データから感染以降のトラヒックのみを切り出した．その手順としてまずは，取得環境独自の制御パケットをフィルタリングで除外した．次にハニーポットの OS のリセット間隔で切り出す．そして，切り出したファイルを攻撃元データのログファイルの時刻と照らし合わせて，感染を確認し，実際の感染攻撃の開始パケットを探し，それ以降を感染トラヒックとして抽出した．

4　経年変化を考慮した特徴量評価

特徴量全体の傾向としては，年を追うごとに識別率が低下する特徴量が多かった．これは，年々移り変わるマルウェアや新種のアプリケーションの発生にともなう，トラヒックの複雑化，多様化が原因だと思われる．今後も新種のマルウェアやアプリケーションは増加すると思われるので，マルウェア感染検知では，トラヒックの変化の影響をあまり受けずに正しく識別できる特徴量を用いることが必要とされる．その点で経年変化を確認することは重要であると考えられる．

感染検知に有効な特徴量について検討するため，経年変化が小さく TPR,TNR が共に高い特徴量と，経年変化が小さく TPR のみ高い特徴量を抜き出す．なお，経年変化が小さく識別率が高い特徴量とは，あるタイムスロット幅，ベクトル量子化レベル数において，3 年間の TPR, または TNR が 90%以上であった特徴量を指す．マルウェア感染検知の要件は，感染と正常を正しく識別できる特徴量を用いること，加えて，感染のみを正しく識別できる特徴量も合わせて使用することである．

4.1　TPR,TNR 共に高い特徴量

今回の実験では，パケットサイズの最小値が唯一 3 年間の経年変化が小さく，TPR,TNR が共に 90%を超えていた特徴量であった．特徴量としてパケットサイズの最小を用いた際の，経年変化が小さくなるタイムスロット幅と量子化レベル数の組み合わせ，および TPR,TNR をそれぞれ以下の表 2，表 3 に示す．

表 2: パケットサイズの最小の TPR

タイム スロット幅	量子化 レベル数	2009 TPR	2010 TPR	2011 TPR
1 秒	8	99.3%	99.8%	91.5%
1 秒	16	98.9%	100%	91.5%
10 秒	16	99.1%	100%	94.6%
100 秒	8	98.4%	98.7%	93.8%
100 秒	16	98.7%	98.9%	95.5%

次に，タイムスロット幅 1 秒における，パケットサイズの最小値についてのテストデータの平均と標準偏差を以下の表 4 に示す．

表 4 より，感染トラヒックの方が正常トラヒックよりパケットサイズの最小値と，最小値の変動が大きくなることがわかる．すなわち，正常トラヒックでは常にパケットサイズの最小値が 60byte に近い値を取るが，感染トラヒックではパケットサイズの最小値が様々な値を取り得るという性質の違いがある．この性質の違いは，感染検知に有効であると思われる．

表 3: パケットサイズの最小の TNR

タイム スロット幅	量子化 レベル数	2009 TNR	2010 TNR	2011 TNR
1 秒	8	99.3%	100%	98.4%
1 秒	16	99.3%	100%	98.4%
10 秒	16	100%	100%	100%
100 秒	8	100%	99.8%	100%
100 秒	16	100%	99.8%	99.8%

表 4: テストデータの統計値 (パケットサイズの最小)

	2009	2010	2011
平均 (感染)	70.3[byte]	63.8[byte]	97.0[byte]
平均 (正常)	60.0[byte]	60[byte]	60.7[byte]
標準偏差 (感染)	36.0[byte]	2.2[byte]	50[byte]
標準偏差 (正常)	0.2[byte]	0[byte]	4.9[byte]

4.2　TPR のみ高い特徴量

経年変化が小さく，TPR が 3 年間を通して 90%以上の特徴量について，3 年間の TPR の平均が最も高くなるときの TPR の平均値，およびタイムスロット幅とベクトル量子化レベル数を以下の表 5 にまとめる．特徴量番号は表 1 の番号に従う．

表 5 の特徴量を大別すると，パケット数，パケット割合，ポート番号に関するパケット数に分けることができる．

4.2.1　パケット数

表 5 より，タイムスロット幅 0.1 秒，量子化レベル数 2，特徴量として RST/ACK パケット数を用いると 2009 年,2010 年,2011 年の TPR が 100%となっている．しかし，タイムスロット幅を 0.1 秒としたときのテストデータを見ると，感染トラヒック，正常トラヒック共に全てのタイムスロットで RST/ACK パケットが 0 個か 1 個しかないことがわかる．しかし，感染コードブックは正常コードブックより小さい値をとっているため，テストデータとコードブックとの距離を計算すると，感染コードブックとの距

表 5: TPR が高い特徴量の最大 TPR とその条件

特徴量番号	最大 TPR	タイムスロット幅	量子化レベル数
4	98.0%	100 秒	32
14	99.4%	1 秒	8
15	99.8%	10 秒	8
17	99.9%	0.1 秒	2
18	99.9%	1 秒	4
19	96.1%	100 秒	2,8
20	100%	0.1 秒	2
21	98.7%	0.1 秒	16
24	97.9%	0.1 秒	8
25	99.8%	0.1 秒	4
27	99.3%	0.1 秒	8
28	98.1%	1 秒	32
29	98.5%	0.1 秒	8
30	99.4%	1 秒	16
31	99.7%	1 秒	32
32	99.9%	0.1 秒	4
34	100%	0.1 秒	16
35	100%	0.1,1,10,100 秒	2,4,8,16,32
36	100%	0.1,1,100 秒	2,4,8,16,32

離の方が小さくなってしまう．よって，RST/ACK パケット数が 0 か 1 しかないタイムスロットを全て感染トラヒックと判断してしまったため，TPR が 100%になったと考えられる．このように，今回の識別方法では，ほとんどのタイムスロットで特徴が似ているが，特徴の違いが現れる頻度が少ない特徴量を正しく識別できていないことがわかった．

同様の問題が SYN パケット数にも当てはまる．SYN パケット数は既存研究でよく用いられている特徴量である．ポートスキャンや SYN フラッド攻撃の際には，SYN パケットが大量に流れるが，それ以外のときでは感染トラヒック，正常トラヒック共にほとんど流れていない．しかし，学習にポートスキャンや SYN フラッド攻撃などを含んだトラヒックを用いると，感染コードブックの方が正常コードブックより大きな値をとってしまう．そして，SYN パケット数が少ないタイムスロットを全て正常トラヒックと判断してしまう．このような特徴量は，今回の識別方法では感染トラヒックと正常トラヒックを正しく識別できていないが，他の特徴量と組み合わせて用いることで，感染検知に有効な特徴量になり得ると思われる．

さらに，今回の実験においてある特定のパケットの数を特徴量とすることには問題がある．感染トラヒックと正常トラヒックには，1 タイムスロット中のパケット数に大きく差があるという問題が存在するからである．一例として，2011 年のテストデータにおける感染トラヒックと正常トラヒックのパケット数の平均を以下の表 6 に示す．

表 6: 1 タイムスロットにおけるパケット数の平均

タイムスロット幅	感染	正常
0.1 秒	2.4	53.5
1 秒	6.2	252.8
10 秒	21.8	3051.6
100 秒	549.1	8724.6

今回の実験におけるテストデータの 1 つのサンプルは，1 つのタイムスロットと対応している．そのため，1 つのタイムスロットにおいて，感染トラヒックと正常トラヒックにパケット数の差があり過ぎるとそれが特徴となり，感染トラヒックと正常トラヒックに違いが現れてしまい，特徴量を正しく評価できない場合がある．そこで，パケットの数自体に関する特徴量の有効性を補完するための手段として割合を考える．

4.2.2 パケット割合

例えば，ACK パケット数に関して，タイムスロット幅を 10 秒とすると，経年変化が小さく TPR が高かった．次に，TCP パケット中の ACK パケット割合について見てみる．タイムスロット幅を 10 秒としたときの，2011 年の感染トラヒックと正常トラヒックの TCP パケット中の ACK パケット割合のヒストグラムを以下図 1, 図 2 に示す．

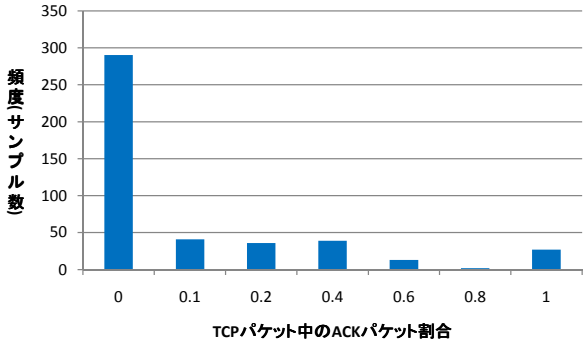

図 1: TCP パケット中の ACK パケット割合 (感染)

図 1, 図 2 より，感染トラヒックでは ACK パケット割合が低く，正常トラヒックでは ACK パケット割合が高いことが確認できた．正常トラヒックではサイズの大きいデータのやり取りが多く，データ通

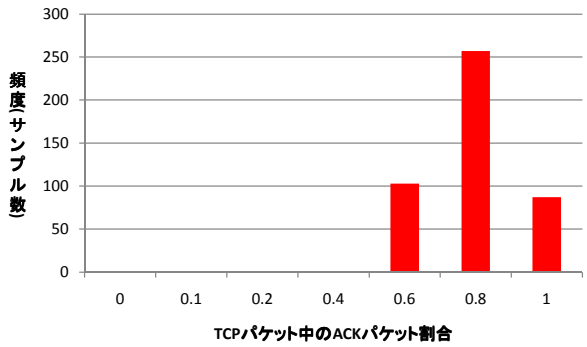

図 2: TCP パケット中の ACK パケット割合 (正常)

信の際の制御パケットが多く流れているため, ACK パケット割合が高くなっていると考えられる.

このように，特徴量としてある特定のパケット数を用いる場合，パケット数だけではなく，その割合も確認することで，感染トラヒックと正常トラヒックの明らかな違いを確認することができた．そして，ACK パケット数，および TCP パケット中の ACK パケット割合は，感染検知に有効な特徴量であることがわかった．しかし，その他の TCP パケット割合は，感染トラヒック，正常トラヒック共に割合が低く，割合がほとんどのタイムスロットで 0.1 より低かった．すなわち，感染トラヒックと正常トラヒックの全パケット数が同じになった場合，パケット数では感染トラヒックと正常トラヒックに違いが現れない可能性があると思われる．

4.2.3　ポート番号に関するパケット数

ポート番号に関するパケット数を特徴量とした場合，TPR が 100%となる場合が多い．しかし，送信元ポート番号が 80/TCP のパケット数は，4.2.1 節で挙げた感染トラヒックと正常トラヒックのパケット数の差の影響を大きく受け，正しく評価できていない可能性がある．さらに，送信元ポート番号が 110/TCP,443/TCP のパケット数にも，4.2.1 節で挙げた RST/ACK パケット数の評価の際の誤識別の問題がある．このような感染トラヒックと正常トラヒックで特徴の違いが現れる頻度の少ない特徴量を正しく評価するためには，別の識別方法を用いる必要があると考えられる．

5　まとめ

本稿ではマルウェア感染検知で用いる特徴量について，その識別能力を調査し，マルウェア感染検知に有効な特徴量について検討した．今回の識別実験から，特徴量全体として，識別率が年を追うごとに低下しているという傾向があり，さらに，今回の識別方法と使用したデータでは，識別率を正しく評価できない特徴量があることもわかった．しかし，特徴量として特定のパケット数を用いる場合，その割合も同時に使用することで感染トラヒックと正常トラヒックの違いが明らかになることを確認した．そして，パケットサイズの最小と ACK パケット数および TCP パケット中の ACK パケット割合が，感染検知に有効である特徴量として使用できる可能性があることを示した．

今後は，識別率の妥当性を検証できなかった特徴量に対して，マルウェア感染検知に対して有効であるか調査し，特徴量を組み合わせたマルウェア感染検知手法について検討していく．

参考文献

[1] インターネット脅威マンスリーレポート　2011 年 5 月度
http://jp.trendmicro.com/jp/threat/security_ news/monthlyreport/article/20110602082147.html

[2] 藤原将志，寺田真敏，安部哲哉，菊池浩明，”マルウェアの感染方式に基づく分類に関する検討”，情報処理学会 CSEC 研究報告，No.21，p177-182，2008 年 3 月

[3] 市野将嗣，坂野鋭，小松尚久，”核非線形相互部分空間法による話者認識”，”信学論 (D-)”,vol.J88,no.8,pp.1331-1338,2005.

[4] 畑田充弘，中津留勇，秋山満昭，”マルウェア対策のための研究用データセット〜MWS 2011 Datasets〜，”マルウェア対策研究人材育成ワークショップ 2011(MWS2011),October 2011.

[5] 佐藤陽平，和泉勇治，根元義章，”複数の検出モジュールの組み合わせによるネットワーク異常検出の高精度化”，電子情報通信学会，信学技報，2004 年

[6] 平松尚利，和泉勇治，角田裕，根元義章，”複数の通常状態を用いたネットワーク異常検出”，電子情報通信学会，信学技報，2006 年

[7] 釘崎祐司，笠原義晃，堀良彰，櫻井幸一，”データ送信間隔に着目した挙動の観測に基づくボット検知手法”，SCIS2009