未知マルウェア検知に向けたマルウェア通信の実態調査

畑田 充弘 †‡　　森 達哉 †

† 早稲田大学 基幹理工学研究科
169-8555 東京都新宿区大久保 3-4-1
{m.hatada, mori}@nsl.cs.wasead.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 膨大な数のマルウェアが多種多様な攻撃に利用され，効率的な解析のために動的解析システムが利用されている．その中でも危険性の高い新種のマルウェアへの対応を早期に行うために，既知マルウェアの自動分類や未知マルウェアの識別といった研究が行われている．本稿では，マルウェア通信に着目し，未知マルウェアの検知に向けた実態調査として，2014 年 12 月からの 6ヶ月間において待受型及び巡回型ハニーポットで収集したマルウェア数万ファイルの動的解析時の通信データの解析結果を報告する．

Characterizing Network Behavior of Malware: Toward Detecting New Malware Families with Network Monitoring

Mitsuhiro Hatada†‡　　Tatsuya Mori†

†Graduate School of Fundamental Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku, Tokyo 169-8555, JAPAN
{m.hatada, mori}@nsl.cs.wasead.ac.jp

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

Abstract Dynamic analysis system are widely used to efficiently analyze massive malware. In order to handle new threat on a preferential basis, researches on automated classification of known malware and detection of unknown malware have been conducted respectively. Several tens of thousands of malware were collected by server type honeypot and web crawler type honeypot from December 2014 to May 2015. In this paper, we report the statistics and distinctive trend of recent malware communication captured in our dynamic analysis environment.

1　はじめに

膨大な数のマルウェアが多種多様な攻撃に利用され，関連するインシデント対応も 2013 年に比べ 2014 年は 9%増加し、43%から 52%へと増加しているという調査結果もある [1]．効率的なマルウェアの解析のために，単独のアンチウイルスソフトによる検知や複数のアンチウイルスソフトによる検知 [2] の他，動的解析システムあるいはサンドボックスが多く利用されるようになっている [3], [4], [5], [6]．アンチウイルスソフトの検知結果には，マルウェアのタイプ (ワーム，トロイの木馬，バックドア等) ，ファミリー，亜種を区別する検知名を付与したり，検知手法 (ジェネリック手法やヒューリスティック手法) よって検知名を付与したりしている．一方で，マルウェアの動的解析結果に加えて静的情報を加味して

マルウェアの自動分類を行う研究 [7] や，API コールに着目し一定の分類や類似検体の抽出が可能であることを示した研究 [8] もあり，アンチウイルスソフトの検知結果との一致を評価している．

しかしながら，アンチウイルスソフトでは検知できないマルウェアが多数あることや，検知名にも間違いや製品差異があり，また，動的解析を行う環境 (OS，ネットワーク等) によって動作が変化するマルウェアも多数あるため，マルウェアの挙動を正しく把握できているかどうか定かではない．膨大なマルウェアの中でも，インシデント対応の優先度を判断する上で，より危険性の高い未知のマルウェアを検知し，対応を早期に行う必要がある．我々は，多くのマルウェアが感染後に何らかのネットワーク挙動を行うことに着目し，マルウェアによる通信に基づいて従来とは異なる分類の定義や未知マルウェアの検知を行うことを検討している．

そのためにも，マルウェアの通信データに基づく検知・分類に有用な特徴や知見を得る必要がある．本論文では，2014 年 12 月から 2015 年 5 月の 6ヶ月間において，待受型及び Web 巡回型ハニーポットで収集した 78,376 ファイルの動的解析によって得られた通信データの分析結果を報告するとともに，いくつかの特徴的な通信挙動を報告する．本論文の主な貢献を以下に示す．

- 動的解析システムの判定結果に基づいて，78,376 ファイルの通信の基本的な傾向を分析し，悪性ファイルによる通信は TCP のセッション数や送受信サイズが大きく，HTTP レスポンスのステータス・コードには 410 (Gone) が少ない，といった知見を得た．
- 各プロトコルにおける悪性ファイルの特徴的な通信の事例 (例えば TCP での名前解決，UPnP での外部通信，DNS シンクホールされている Web サーバの HTTP レスポンス等) を示し，上記知見とともに特徴量を詳細に分解・定義・組み合わせることで，通信データに基づく新たな分類の定義や未知マルウェアの検知の検討の第一歩を進めた．
- 動的解析環境，解析対象データ，インターネットにおける対策など本研究テーマを取り巻く課題や制約の議論は，新たな研究テーマをも示唆している．

以降．2 章では関連研究を述べ，3 章で分析対象データの取得環境とデータセットの概要を示し，4 章で主要なプロトコルにおける統計データと特徴的なマルウェア通信を報告する．5 章でデータセット及び分析に関わる課題を議論し，5 章でまとめを行う．

2 関連研究

マルウェアをアンパック・逆アセンブルし，プログラムコードの類似性に基づいた自動分類 [9] や，制御フローグラフの類似性で分類を行う [10] などバイナリを対象とした研究は従来から多く存在する．マルウェアの通信に着目して，パケットのヘッダやペイロードから未知マルウェアの検知に有効な特徴量を評価した研究 [11]，[12] や，LPC ケプストラム分析によるマルウェア感染の検知精度を評価した研究 [13] があるが，実験データ数が少ないという課題もある．大規模なネットワーク環境におけるマルウェアの伝搬を HTTP 通信に着目して検知する研究 [14] もあり，動的解析システムで解析した数千件のマルウェアの通信の概要が紹介されている．URL ブラックリストを生成するためにマルウェアの動的解析時の通信を調査した研究 [15] もあるが，HTTP に特化した調査に留まる．HTTP と DNS 通信に着目し，対象ホストからの外向き通信の因果関係を分析するアプローチで，発見しにくいマルウェアの検知を試みる研究 [16] も行われている．

文献 [17] では，独立成分分析によりマルウェア通信と正常通信を分離し，Random Forests による分類から 2 種類のマルウェアファミリを判定しており，文献 [18] では，ネットワークフロー間の遷移をグラフ化し，グラフサイズや枝の数の平均等の特徴量から決定木を用いて 12 種類のマルウェアファミリを分類しているが，マルウェアの通信挙動を広く調査・報告しているものではない．そこで我々は，大量のファイルの動的解析時の通信データを様々な観点で分析し，今後の検知・分類に有用な知見を得ることを目的として，マルウェア通信の実態調査を行う．

3 実態調査の環境とデータセット

3.1 環境

本論文で対象とする 78,376 ファイルの動的解析時の通信データ (pcap ファイル) は図 1 のような環境で収集・蓄積したものである．待受型ハニーポットは Windows XP ベースの高対話型ハニーポット数十台で構成されており，複数拠点でそれぞれインター

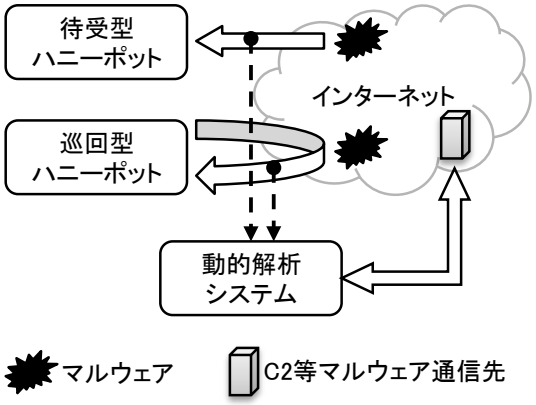

図 1: データ収集・分析環境

表 1: データセットの基本情報

収集期間	2014/12/01 - 2015/05/31
pcap ファイル数	78,376
データサイズ	
IP	30.9 GB
TCP	30.7 GB
UDP	174.9 MB
ICMP	2.7 MB
Media Types 別対象ファイル数	
x-pe-app-32bit	76,265
x-dosexec	1,250
zip	389
x-pe-dll-32bit	249
x-dosexec-dll	134
msoffice-doc	29
x-pe-app-64bit	26
msoffice-xls	13
x-pe-dll-64bit	9
msoffice-xlsx	5
msoffice-ppt	3
x-pe-driver-64bit	3
msoffice-pptx	2
判定結果別対象ファイル数	
悪性	38,588
不審	34,546
良性	5,242

ネットに接続している．また，巡回型ハニーポットは Windows XP 及び 7 ベースで Adobe Reader 等の脆弱な複数のプラグインをインストールした Internet Explorer による高対話型ハニーポット [19] 数台が，複数拠点からインターネットに接続している．Web 空間を巡回するために元となる URL のリストは，独自に収集している URL ブラックリストやその候補のリストである．動的解析システムは，待受型ハニーポットと巡回型ハニーポットの通信からファイルを復元し，Windows XP 及び 7 ベースの仮装環境でファイルを一定時間実行する．複数の拠点のインターネット接続を利用し，C2 等マルウェアが通信する先とのパケットを記録している．通信から復元されるファイルには，脆弱性を悪用した攻撃によってダウンロードさせられるマルウェアの他，特に巡回型ハニーポットでは結果的にマルウェアではないプログラムファイルやドキュメントファイル等も含んでいることが前提であり，何らかの通信が発生したファイルを対象にしている．なお，動的解析システムの出力は pcap ファイルの他，静的解析を含むエキスパートのナレッジに基づく解析レポートがあり，ファイルやレジストリ操作，ネットワーク挙動等の履歴からマルウェアらしい挙動のリスト (悪性挙動) やマルウェアらしさのスコア (悪性スコア) が記載されている．

3.2 データセット

今回実態調査を行うデータセットの基本情報として，収集期間，総ファイル数，イーサフレームにおける IP パケット及びその内訳 (TCP，UDP，ICMP) のデータサイズ，Media Types 別の解析対象ファイル数，悪性スコアに基づいた判定結果 (悪性，不審，良性) 別の解析対象ファイル数を表 1 に示す．通信データでは，TCP が 99%を占め UDP と ICMP は非常に少ない．ファイルの種類では，DLL を含む 32bit の PE ファイルが 97%を占める一方で，64bit の PE ファイルや Office ファイルは非常に少ない．判定結果は悪性 49%，不審 44%，良性 7%という割合であり，大半が何らか疑わしいファイルであるということになる．

4 分析

前章で述べたデータセットについて，悪性判定結果別に各プロトコルの分析結果を報告する．3 種類の悪性判定結果以外にも動的解析レポートに基づく挙動のパターン等をもとにしたより詳細な分類も今後検討する必要はあるが，本論文では大きな粒度の判定結果に基づいて分析を行う．また，各プロトコルにおいて特徴的な通信の事例を示す．

表 2: TCP セッション数，送受信サイズ

判定	項目	セッション数	受信サイズ	送信サイズ
悪性	Avg	5.55	0.53 MB	0.02 MB
	Max	6427	11.37 MB	30.99 MB
	SD	43.24	799256.16	171189.39
不審	Avg	3.32	0.25 MB	0.01 MB
	Max	1378	16.71 MB	0.37 MB
	SD	9.82	1197230.45	30201.80
良性	Avg	1.87	0.23 MB	0.01 MB
	Max	215	14.98 MB	0.38 MB
	SD	6.44	1319407.03	30915.24

4.1 TCP

解析対象ファイルによる TCP 通信のセッション数，受信サイズ，送信サイズについて，平均 (Avg)，最大 (Max)，標準偏差 (SD) を表 2 に示す．標準偏差は大きくなるものの，悪性ファイルはセッション数の平均，送受信サイズの平均ともに大きくなる傾向がある．悪性ファイルによる最大送信サイズ (30.99 MB) は，Sality あるいは Padobot と呼ばれるマルウェアであり．445/tcp で 6,000 を超える内外のホストの共有ドライブに感染コードを送ろうとしたものであった．

HTTP

TCP の中でも主要なプロトコルといえる HTTP について分析する．解析対象ファイルによるリクエスト・メソッドのパターン，レスポンスのステータス・コード，レスポンスのコンテンツ・タイプについて，頻度上位 5 種類をそれぞれ表 3，4，5 に示す．

リクエスト・メソッドのパターンは，解析対象ファイルが HTTP リクエストを行った順にそのメソッドを表記したものである．例えば”G”であれば GET を一度だけであり，”G,P”であれば GET の後に POST でリクエストを行ったということを表している．不審/良性ファイルともに上位 2 種類がリクエストなし (N/A)，GET を 2 回であるのに対して，悪性ファイルは 30%が 1 度だけ GET，25%が GET の後に POST を 2 回，GET を 3 回というパターンで半数以上を占めている．このようなパターンの一部を目視で確認した範囲では，ファイルをダウンロードするための GET リクエストや，インターネットへの接続性を確認するための特定サイトへの GET リクエストの後，何らかのデータを POST で送信するといった挙動が見られた．また，悪性/不審/良性に共通するリクエストなし (N/A) は，解析環境の Windows OS が時刻同期を行うための DNS/NTP 通信や，インターネット接続確認を行う NCSI (Network Connectivity Status Indicator) の DNS 通信は発生しているものの HTTP 通信には至っていない場合が多く見られる．ここでは，頻度上位 5 種類のみを記載しているが，全パターンは悪性/不審/良性それぞれで 960/665/167 種類となっていることからも，より悪性ファイルの方が多様なパターンで HTTP リクエストを行うといえる．

HTTP レスポンスのステータス・コードは悪性/不審/良性で大半が 200 (OK) となっている．しかしながら，Conficker あるいは Downadup と呼ばれるマルウェアの接続先は，DGA (Domain Generation Algorithm) によって生成されるドメイン名または特定のドメイン名となり，多くが CERT 等により DNS シンクホールされている．そして，HTTP の接続先からは，”シンクホールされているドメインへの接続”を示す HTML コンテンツをステータス・

表 3: HTTP リクエスト・メソッドのパターン

悪性	[%]	不審	[%]	良性	[%]
G	30	N/A	28	N/A	57
G,P,P,G,G,G	25	G,G	17	G,G	11
N/A	10	G,P	8	G	10
G,P,G,G,G	9	P,G,G,G,G	8	G,P	9
P,G,G,G,G	7	G	7	P,G,G,G,G	1

G:GET，P:POST，N/A:リクエストなし

表 4: HTTP レスポンスのステータス・コード

悪性	[%]	不審	[%]	良性	[%]
200	94.4	200	87.1	200	76.3
404	1.9	410	5.1	410	8.7
403	1.4	404	2.3	302	4.4
302	0.7	304	1.4	304	3.9
301	0.4	302	1.0	404	2.3

200:OK，301:Moved Permanently，302:Found，304:Not Modified，403:Forbidden，404:Not Found，410:Gone

表 5: HTTP レスポンスのコンテンツ・タイプ

判定	コンテンツタイプ	[%]
悪性	image/png	26
	application/json; charset=UTF-8	19
	text/html	18
	application/octet-stream	17
	text/plain	6
不審	image/png	22
	text/html	14
	image/jpeg	10
	text/html; charset=utf-8	9
	text/plain	5
良性	image/png	18
	text/html	12
	text/html; charset=utf-8	11
	text/html; charset=UTF-8	9
	image/gif	6

表 6: SSL セッション数、送受信サイズ

判定	項目	セッション数	受信サイズ	送信サイズ
悪性	Avg	0.19	0.03 MB	0.96 KB
	Max	45	9.16 MB	0.51 MB
	SD	1.42	273496.86	8075.51
不審	Avg	0.18	0.01 MB	0.38 KB
	Max	31	5.19 MB	0.12 MB
	SD	1.51	107243.18	3571.80
良性	Avg	0.21	0.01 MB	0.40 KB
	Max	32	9.31 MB	0.21 MB
	SD	1.12	147947.44	3886.35

表 7: UDP セッション数，送受信サイズ

判定	項目	セッション数	受信サイズ	送信サイズ
悪性	Avg	2.78	1.7 KB	0.60 KB
	Max	1347	9.17 MB	0.71 MB
	SD	13.43	73377.28	5708.12
不審	Avg	3.16	0.75 KB	1.74 KB
	Max	1105	5.14 MB	46.81 MB
	SD	9.06	29026.35	263902.69
良性	Avg	3.01	0.46 KB	0.31 KB
	Max	202	0.05 MB	0.04 MB
	SD	5.38	1545.10	1010.24

コード 200 で受信するため，このような HTTP レスポンスも含んでいる点は注意が必要である．不審/良性の頻度 2 位となっている 410 (Gone) は Web サイトの管理社が意図的に設定したものと考えられるため，悪性ファイルの通信先となるような攻撃者が準備した Web サイトでは，わざわざそのような設定をしないことが考えられる．

HTTP レスポンスのコンテンツ・タイプでは，悪性ファイルの頻度 2 位と 4 位にある”application/json”と”application/octet-stream”が不審/良性と比べて上位になっている．これらは，MultiPlug といったアドウェアあるいは PUA (Potentially Unwanted Applications) が，他の実行ファイルをダウンロードしたり，HTTP POST リクエストに対して json 形式でレスポンスを受け取ったりといったものが多く確認できた．

SSL

TCP の中でも主要なプロトコルである SSL は，HTTP (全フレームサイズ:0.34 GB) の約 4.5 倍 (1.54 GB) であった．解析対象ファイルによる SSL 通信のセッション数，受信サイズ，送信サイズについて，平均 (Avg)，最大 (Max)，標準偏差 (SD) を表 6 に示す．平均セッション数は悪性ファイルより良性ファイルの方が大きく，分布に大きな偏りがあるものの，悪性ファイルは不審/良性ファイルに比べて平均受信サイズ・送信サイズともに大きくなっている．悪性ファイルに限らず，SSL 通信先の頻度上位には，クラウドストレージサービスや CDN 事業者の IP が多く見られた．

4.2 UDP

解析対象ファイルによる UDP 通信のセッション数，受信サイズ，送信サイズについて，平均 (Avg)，最大 (Max)，標準偏差 (SD) を表 10 に示す．TCP (表 2) と異なり，UDP セッション数の平均は悪性ファイルが不審/良性よりも小さくなっている．不審ァイルによる最大送信サイズ (46.81 MB) は，c0f06 (マルウェア検体の sha1 ハッシュ値先頭 5 文字のみ記載) であり，Trojan.Win32.Staser (Kaspersky 社検知名，以降同社検知名を用いる) の亜種によるものであった．このマルウェア (今回の分析期間後に悪性ファイルとして判定されるようになっている) は，あるホストに”k=15 バイトの英数小文字”を HTTP で POST したレスポンスとして，”-udp 203.0.113.142:80 -timeout 1 -thread 22”といったリストを 3 種類受信し，そのうちのある Web ホスティングサイトが管理する IP アドレスの 80/udp へ何らかのデータを送信しようとする挙動が見られた．また，特徴的な通信として，悪性ファイルの中に，UPnP でブロードキャストアドレス (239.255.255.250) を介して中国の複数の IP アドレスと通信を行っていた Backdoor.Win32.FirstInj の亜種も見られた．

DNS

UDP の中でも主要なプロトコルといえる DNS について，クエリーとレスポンスのレコードタイプの頻度上位 5 種類をそれぞれ表 8，9 に示す．悪性/不審/良性に関わらず 94%以上が A レコードのクエリーであり，悪性ファイルに特徴的なクエリーとしては，MX レコードのクエリーがごく僅かに確認できている．これは主にスパムメール送信を目的とする Trojan.Win32.Cutwail の亜種により，主要なフリーメールのドメイン名の名前解決を行っていたものによる．レスポンスのレコードタイプでは，Conficker による DGA に基づくドメイン名の名前解決ができず，結果として SOA レコードのみの応答が悪性ファイルでは大きくなったものと考えられる．

悪性ファイルには，Google DNS で名前解決を行う

表 8: DNS クエリーのレコード・タイプ

悪性	[%]	不審	[%]	良性	[%]
0x0001	98.42	0x0001	94.36	0x0001	95.76
0x001c	1.03	0x001c	5.58	0x001c	4.21
0x000f	0.51	0x0010	0.05	0x000c	0.02
0x0002	0.01	0x0021	0.003	0x0021	0.01
0x0010	0.01	0x000c	0.003	-	-

0x0001: A, 0x0002: NS, 0x000c: PTR, 0x000f: MX, 0x0010: TXT, 0x001c: AAAA, 0x0021: SRV

表 9: DNS レスポンスのレコード・タイプ

悪性	[%]	不審	[%]	良性	[%]
0x0001	58.8	0x0001	54.3	0x0001	53.5
0x0002	28.5	0x0002	26.6	0x0002	29.0
0x0006	6.1	0x0005	16.9	0x0005	15.7
0x0005	4.9	0x001c	1.6	0x001c	1.4
0x001c	1.1	N/A	0.4	0x0006	0.5

0x0001: A, 0x0002: NS, 0x0005: CNAME, 0x0006: SOA, 0x001c: AAAA, N/A: レスポンスなし

マルウェア (HEUR:Trojan.Win32.Generic) も確認された．また，不審ファイルには 53/tcp で名前解決を行うマルウェア (Trojan-Downloader.Win32.Agent) の亜種も確認された．

その他

本データセットの中で，MWS Datasets [20] の D3M 2015[21] と共通するファイルとして，b76a9 (マルウェア検体の sha1 ハッシュ値先頭 5 文字のみ記載) の動的解析時の pcap ファイルがあった．これは HEUR:Trojan.Win32.Generic として検知されるマルウェアであり，2014 年 5 月 2 日に動的解析された pcap では，複数のホストに対して HTTP の GET リクエストを送信し，301(Moved Permanently) や 404 (Not Found) のレスポンスを受け，約 60 秒経過後に 354 ホストに対して UDP のハイポートに接続を試み，解析時間中には 33 ホストからの応答を受信しており，P2P 通信を行っている様子がわかる．本論文で使用しているデータセットと通信の種類 (HTTP，UDP による P2P) や HTTP リクエストの User-Agent が”Opera/9.50 (Windows NT 6.0; U; en)”という点では同様であったが．UDP 通信と HTTP 通信のパケットが時系列で混在してる点は異なった．

表 10: ICMP 宛先数，送信回数，タイプ

判定	項目	宛先数	回数	Type0	Type3	Type8
悪性	Avg	1.63	3.25	0	0.01	3.24
	Max	8271	16467	0	7	16467
	SD	115.00	228.96	0	0.20	228.96
不審	Avg	0.02	0.05	0	0.04	0.004
	Max	3	633	0	633	60
	SD	0.14	3.44	0	3.42	0.40
良性	Avg	0.02	0.05	0.00	0.02	0.03
	Max	3	90	16	7	90
	SD	0.14	1.4	0.22	0.20	1.39

Type 0: Echo Reply, Type 3: Destination Unreachable, Type 8: Echo Request

4.3 ICMP

解析対象ファイルによる ICMP 通信の宛先数，回数，ICMP タイプについて，平均 (Avg)，最大 (Max)，標準偏差 (SD) を表**??**に示す．悪性ファイルによる ICMP 通信の宛先数/回数の平均が大きくなっているのは，2 つのファイルが要因となっている．最大宛先数 (8,271) に計 16,467 回の Echo Request を送信している悪性ファイルと，7,109 個の宛先に計 14,147 回の Echo Request を送信している悪性ファイルは，ともに Virut あるいは Allaple と呼ばれるマルウェアの亜種で，それぞれある /16 のネットワークをスキャンしている挙動が見られる．悪性ファイルの中には，内部ホストへの Echo Request に加えて，外部 (ポルトガル) のホストへも Echo Request を行うマルウェア (Virus.Win32.Qvod) も確認された．不審ファイルによる ICMP 送信回数の最大値 633 と ICMP Type3 (Destination Unreachable) はどういうファイルによるものであるが，相当数の Echo Request は観測されていない．これは 4.2 で通信挙動を解説した c0f06 検体があるホストの 80/udp へ送信した UDP パケットが，宛先ポートへ届かず ICMP Type3 として帰ってきたものである．

5 議論

動的解析環境　一般に知られる動的解析環境の課題として，マルウェアによる解析環境検知の結果，十分な解析結果が得られない場合がある．一例としては実際の解析時間以上の時間経過後や再起動後のみに動作を開始するようなマルウェアの場合，通信データが取得できない可能性がある．個々のマルウェアの初動に応じて，柔軟に解析リソースを制御する等の高度化が必要となる．また，解析環境そのものに

よる通信 (時刻同期や NCSI) をデータセットには含んでいるため，適切な前処理によって除外する必要もある．

解析対象データ　待受型及び巡回型ハニーポットで収集しているファイルを対象データとしているが，データに偏りがある可能性は十分にある．そもそも通信をともなわないマルウェアも存在するが，マルウェア全体の中でどの程度のマルウェアが通信を伴うかといった評価も必要である．本論文では，悪性/不審/良性といった大分類をもとに分析を行ったが，分類間の差異について概観したものの統計的検定を伴ったものではないため，今後の検討を進めるにあたって検証する必要がある．また，未知マルウェアの検知に向けては，有効な特徴量に応じたより詳細な分類や，現状ではグレーな位置付けともいえるアドウェアあるいは PUA をどのように位置づけるかという考慮も必要となる．

インターネットにおける対策　DNS シンクホールや DPI (Deep Packet Inspection) によるフィルタリングなど，各国の CERT 機関や ISP 等によってインターネットにおいて有効な対策が進んでいる．その結果，解析環境からのマルウェアによる通信もその対策による影響があることは避けられない．実社会の利益に対して研究が足かせとならないよう注意する必要がある．

6　まとめ

様々な攻撃に利用され日々新たに出現するマルウェアの中から，インシデント対応の優先度を判断する上で，より危険性の高い新種のマルウェアへの対応を早期に行うことが望ましい．近年，マルウェアの静的解析結果に基づく自動分類や，未知マルウェアの検知の研究とともに，マルウェアの動的解析結果として得られた通信データに基づいて，マルウェアの自動分類を行う研究が進んでいる．しかしながら，関連研究ではマルウェアの通信データから複数の特徴量を定義して，一定の精度で複数種類のマルウェアの分類を行っているものの，マルウェアの通信挙動を広く調査・報告しているものはない．そこで我々はまず，マルウェアの通信データに基づく検知・分類に有用な知見を得ることを目的として，2014 年 12 月から 2015 年 5 月の 6ヶ月間に，待受型及び Web 巡回型ハニーポットで収集した 78,376 ファイルの動的解析で得られた通信データの分析を行った．具体的には，動的解析の判定結果として悪性/不審/良性と分類したファイルの通信データをもとに，TCP，UDP，ICMP，HTTP，SSL，DNS の基本的な統計情報の分析と，特徴的な通信挙動の事例報告を行った．その結果，悪性/不審/良性ファイルの通信データそれぞれにおいて異なる傾向があることを明らかにした．また，さらに分析を通じて得た課題や対策について議論した．今後，報告した特徴的な事例以外も調査を進め，個々の特徴量を更に詳細化し，組み合わせていくことによって，多様な特徴量をもとに検知・分類の方式を検討していく予定である．

参考文献

[1] “2015 global threat intelligence report.” `https://nttgroupsecurity.com/`.

[2] “Virustotal.” `https://www.virustotal.com/`.

[3] K. Aoki, T. Yagi, M. Iwamura, and M. Itoh, “Controlling malware HTTP communications in dynamic analysis system using search engine,” in *Proceedings of Third International Workshop on Cyberspace Safety and Security*, pp. 1–6, Sep. 2011.

[4] “Cuckoo sandbox.” `http://www.cuckoosandbox.org/`.

[5] “Malwr.” `https://malwr.com/`.

[6] “Anubis.” `https://anubis.iseclab.org/`.

[7] 畑上英毅, 橋本正樹, 堀合啓一, and 田中英彦, “マルウエア動的解析に於ける自動分類手法の研究,” IPSJ SIG Technical Reports 2011-CSEC-52 51, Mar. 2011.

[8] A. Fujino, J. Murakami, and T. Mori, “Discovering Similar Malware Samples Using API Call Topics,” in *Proceedings of the 2015 IEEE Consumer Communications and Networking Conference (CCNC 2015)*, pp. 146–153, Jan. 2015.

[9] 岩村誠, 伊藤光恭, and 村岡洋一, “機械語命令列の類似性に基づく自動マルウェア分類システム,” **情報処理学会論文誌**, vol. 51, pp. 1622–1632, Sep. 2010.

[10] S. Cesare and Y. Xiang, “Classification of malware using structured control flow,” in *Proceedings of the Eighth Australasian Symposium on Parallel and Distributed Computing - Volume 107*, pp. 61–70, Jan. 2010.

[11] Y. Otsuki, M. Ichino, S. Kimura, M. Hatada, and H. Yoshiura, “Evaluating payload features for malware infection detection,” *Journal of Information Processing*, vol. 22, pp. 376–387, Apr. 2014.

[12] M. Ichino, K. Kawamoto, T. Iwano, M. Hatada, and H. Yoshiura, “Evaluating header information features for malware infection detection,” *Journal of Information Processing*, vol. 23, Sep. 2015. (to appear).

[13] 岩野透, 吉浦裕, 畑田充弘, and 市野将嗣, “LPC ケプストラム分析を利用したマルウェアの感染検知,” **情報処理学会論文誌**, vol. 56, Sep. 2015. (to appear).

[14] Luca Invernizzi and Stanislav Miskovic and Ruben Torres and Sabyaschi Saha and Sung-Ju Lee and Christopher Kruegel and Giovanni Vigna, “Nazca: Detecting malware distribution in large-scale networks,” in *Proceedings of the ISOC Network and Distributed System Security Symposium (NDSS '14)*, pp. 1–16, Feb. 2014.

[15] 畑田充弘, 稲積孝紀, 有川隼, and 田中恭之, “サンドボックス解析結果に基づく url ブラックリスト生成方式に関する事例調査,” IPSJ SIG Technical Reports 2015-CSEC-66 47, Jun. 2015.

[16] H. Zhang, D. D. Yao, and N. Ramakrishnan, “Detection of stealthy malware activities with traffic causality and scalable triggering relation discovery,” in *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*, pp. 39–50, Jun. 2014.

[17] H. Mekky, A. Mohaisen, and Z.-L. Zhang, “POSTER: Blind Separation of Benign and Malicious Events to Enable Accurate Malware Family Classification,” in *Proceedings of 2014 ACM CCS*, pp. 1478–1480, Nov. 2014.

[18] S. Nari and A. A. Ghorbani, “Automated Malware Classification based on Network Behavior,” in *Proceedings of ICNC 2013*, pp. 642–647, Jan. 2013.

[19] M. Akiyama, T. Yagi, Y. Kadobayashi, and T. Hariu, “Client Honeypot Multiplication with High Performance and Precise Detection,” *IEICE Trans. on Info. and Sys.*, vol. E98-D, pp. 775–787, Apr. 2015.

[20] M. Hatada, M. Akiyama, T. Matsuki, and T. Kasama, “Empowering anti-malware research in Japan by sharing the MWS Datasets,” *Journal of Information Processing*, vol. 23, Sep. 2015. (to appear).

[21] 神薗雅紀, 秋山満昭, 笠間貴弘, 村上純一, 畑田充弘, and 寺田真敏, “マルウェア対策のための研究用データセット ～mws datasets 2015～,” IPSJ SIG Technical Reports 2015-CSEC-70 6, Jun. 2015.