マルウェアの転送ログを利用したボットの活動分析

金井 瑛 †　　水谷 正慶 †　　武田 圭史 †　　村井 純 ‡

† 慶應義塾大学大学院 政策・メディア研究科
‡ 慶應義塾大学 環境情報学部
252-8520 神奈川県藤沢市遠藤 5322

{kanai,mizutani,keiji,jun}@sfc.wide.ad.jp

あらまし　インターネット利用者のセキュリティを脅かすボットは，日々大量の亜種が出現している．また，高度な遮蔽技術が用いられることもあり，対策が困難である．ボットネット対策にはボットの特性や活動傾向を知ることが必要であるが，一般的なボットの特性はよく知られていない．本稿では，日本国内に設置したハニーポットのマルウェア転送イベントを用いて，マルウェアを配布しているノードの所属する国に着目した傾向分析を実施した．その結果，日本国内からの転送にはネットワークの利用形態やトラフィックの流量と深い関係があると分かった．また，特殊なマルウェアの配布元ノードを発見し，攻撃者がマルウェアの検知を避けているために常に新しいマルウェアを配布している様子を観測できた．

Analysis of Bot Activities Based on Malware Transfer Records

Akira Kanai †　　Masayoshi Mizutani †　　Keiji Takeda †　　Jun Murai ‡

†Keio University - Graduate School of Media and Governance
‡Keio University - Faculty of Environment and Information
5322, Endo, Fujisawa-shi, Kanagawa 252-8520, Japan

Abstract　Bot threatens Internet users' security with malicious behaviors. The new varieties of botware appear everyday. Since those bots conceal their existences and activities, it is difficult to defend from them. Properties of the general botware are not well known, although knowing their trends is necessary in taking actions against them. In this paper, the event logs of malware activities, on some hosts within Japan, were analyzed. By classifying them in terms of geographic locations where the bots actually came from, relationships were discovered between the infections and Internet trends in Japan, such as network topologies and traffic trends. Also, it was found that the attackers update their botware often enough, to avoid from being detected by anti-malware solutions.

1　はじめに

近年，インターネット上で感染を広げ，感染に成功したノードを攻撃者が集中管理できるようにするマルウェアへの対応が急務とされている．攻撃者はマルウェアに感染したノードの集合であるボットネットを用いて，主に迷惑メールの送信や感染ノード上に保存されているクレジットカード番号など金銭的価値のある情報を取得できる．攻撃者は不正な活動により利益を得られるため，その存在を隠蔽し，規模を拡大させるために高い技術力を導入する十分な動機を持つと言われている [1]．ボットネットの対策には多くの研究が進められているだけではなく，ボット感染予防推進事業が国とともに開始されるなど [2]，ますますボットネットに対する関心は高まっている．ボットネット被害の低減にはボットネットの構成ノード (ボット) とボットネットに関する多くの知識が求められるが，ボットの亜種発生頻度の高さや，高度な隠蔽技術により，対策が困難である．

本稿では国内のハニーポットにおける転送の記録

が含まれている研究用データセット CCC DATAset 2008 の攻撃元データ (以降，研究用データセット) を用いて，マルウェアの配布元ノードの国およびタイムゾーンに着目した分析を実施した．

本稿は 5 節で構成される．まず第 2 節では研究用データセットの概要と，転送方法の種類について説明する．第 3 節では，配布元ノードの国の分布について分析し，国ごとの特徴を示す．第 4 節では，1 日における配布元ノード数の分布について着目し，各国におけるノードの稼働状況と配布元ノード数との関係を調査した．続く第 5 節では，データセットに含まれる特殊な配布元ノードが転送するマルウェアの種類の変化を述べる．最後に第 6 節で本稿の調査で得られた結果をまとめる．

2　調査データ

本稿で用いた研究用データセットには，日本国内 112 台のハニーポットにおいて取得された 2007 年 11 月 1 日から 2008 年 4 月 30 日まで 6ヶ月間にわたるイベントのログが 2,942,221 件含まれている．

ボットの一般的な感染活動は**攻撃**，**転送**と**接続**の 3 段階に分けられ次のように活動する [3].

1. 攻撃:脆弱なノードに対して小さなコード (ダウンローダ) を送り込む.
2. 転送:ダウンローダはインターネット上の配布元ノードからマルウェアをダウンロードする.
3. 接続:マルウェアに感染したノードはボットネットの管理サーバに接続する．以後は管理サーバからの命令に従い活動する.

研究用データセットには感染活動のうち，転送に関する情報のみが含まれており，攻撃と接続に関する情報は持たない.

2.1　データ項目

研究用データセットの各レコードが含む項目は次の通りである.

- 時刻
- 配布元ノード情報
- プロトコル
- ポート番号
- 通信方向
- ハッシュ値
- マルウェアの名称
- ファイル名

時刻はハニーポットが転送を開始した時刻である．**配布元ノード情報**はハニーポットがマルウェアを転送する際に接続する通信先ノードの情報が含まれており，配布元ノードを一意に識別できる．**プロトコル**と**ポート番号**はダウンロードに用いられたトランスポート層のプロトコル情報である．**通信方向**は，転送の通信を開始した方向によってプッシュ型とプル型の 2 つに分類される．**ハッシュ値**は転送されたマルウェアのハッシュ値が含まれており，**マルウェアの名称**はイベントが発生した翌日にトレンドマイクロ社のウィルス検知エンジンにおいて判別されたマルウェア名を示す．ただしマルウェアが判別できなかった場合は UNKNOWN と表記される．**ファイル名**は転送が完了したマルウェアを保存しようとするノード上のディレクトリ名とファイル名である.

2.2　プッシュ型転送とプル型転送

マルウェアの転送にはプッシュ型とプル型がある．以下にそれぞれの転送方法と特徴について述べる．それぞれの転送方法を用いた感染の手順を図 1 に示す．まず，攻撃では多くの場合，攻撃元ノードが脆

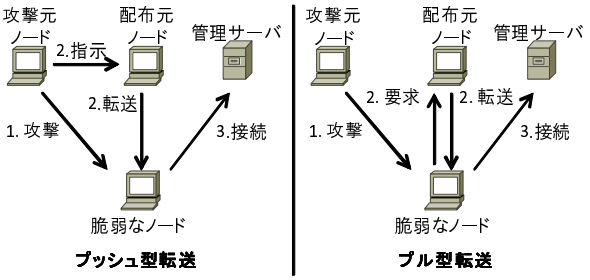

図 1: プッシュ型転送とプル型転送の例

弱なノードのソフトウェアに含まれる脆弱性を利用してダウンローダと呼ばれる小さなソフトウェアを脆弱なノードに送り込み実行させる．攻撃が成功すると，ダウンローダは転送を開始する.

図左のプッシュ型転送ではダウンローダが実行されると脆弱なノードはあらかじめプログラムされたポート番号でインターネットからの通信を受け付ける．攻撃元ノードは配布元ノードに対して脆弱なノードの IP アドレス情報とポート番号を伝える．ただし，攻撃元ノードと配布元ノードは同一のノードである場合もある．配布元ノードは受け取った情報を用いて，脆弱なノードにマルウェアのアップロード

を開始する．プッシュ型転送はインターネット上のノードから通信が開始されるため，NAPT など，外部からの通信が制限された環境においてプッシュ型の転送は動作しない．

図右のプル型転送では脆弱なノードはダウンローダが実行されるとあらかじめプログラムされた IP アドレスおよびポート番号に接続し，マルウェアをダウンロードする．プル型転送は脆弱なノードから通信が開始されるため，外部からの通信が制限された環境においても動作する．

転送が終了し，マルウェアに感染したノードはボットネットの管理サーバに接続する．以後は管理サーバからの命令に従い活動する．

2.3 国情報の取得

本稿では，配布元ノードの情報と国別情報を結びつけた分析を行う．配布元ノードから配布元ノードの国情報を推定する手段として MaxMind 社の提供する GeoIP[4] を用いた．

3 マルウェアの配布元ノードの国別特性

3.1 国ごとの配布元ノード数

第 2 節ではデータセットに含まれるデータはプッシュ型とプル型に分類されると述べた．研究用データセットに含まれる全イベントのうち，122,399 イベント (4.2%) はプッシュ型転送のイベントであり，2,819,822 イベント (95.8%) はプル型転送のイベントである．また，データセットに含まれる配布元ノードの総数は 259,742 ノードであり，42,197 ノード (16.2%) がプッシュ型転送の配布元ノード，217,545 ノード (83.8%) がプル型転送の配布元ノードである．

本節では，配布元ノードが所属する国を調査し，その傾向を示す．まず，データセットからプッシュ型転送を利用するノードのみを抽出し，各配布元ノードの国を調査した．図 2 にノード数が多い上位 20 国の分布を示す．

同様に，プル型転送を利用するノード数の国毎の分布を図 3 に示す．また，ノード数が最も多い JP(日本) を除いた上位 2 位から 20 位までのノード総数を図 4 に示す．

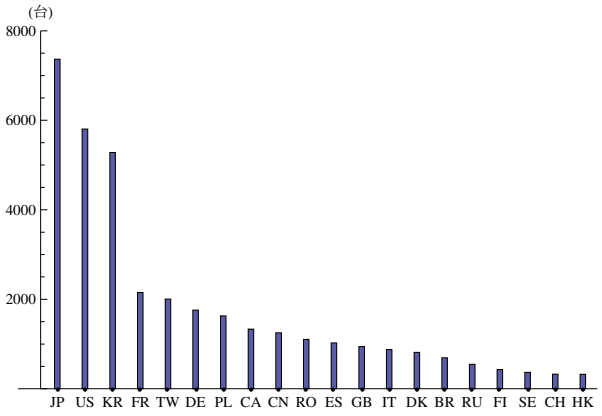

図 2: プッシュ型配布元ノードの国別分布

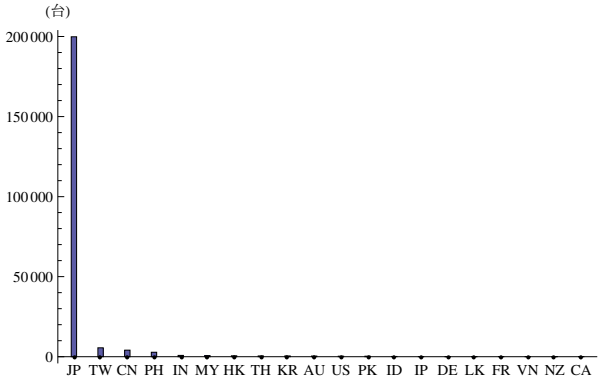

図 3: プル型配布元ノードの国別分布

3.2 日本からの転送と転送方法の関係についての考察

日本の配布元ノードは 207,241 ノード (全体の 79.8%) と全体の大部分を占めた．同一国内からの転送が多い理由にはボットあるいはボットネットの特性が関係すると考えられる．

図 5 にマルウェアの原種の 1 つである [5] Agobot から抜粋したソースコードを示す．ソースコードは Agobot の転送が完了し，管理サーバに接続した後に実行するスキャンの動作を示している．*mFakeMsg* はボットがスキャンに必要とする情報を格納する構造体であり，*m_cScanner* の *HandleCommand* 関数の引数として $mFakeMsg$ 構造体を指定すると，*szLocalIp* の所属する ClassB のアドレス空間に対してスキャンを開始する．*szLocalIp* にはボットの IP アドレスが含まれている．スキャンにより脆弱性を持つノードを発見すると，Agobot は攻撃を開始する．

グローバル IP アドレスは各地域や各国のインターネットアドレス管理組織 (Internet Registry) に管理

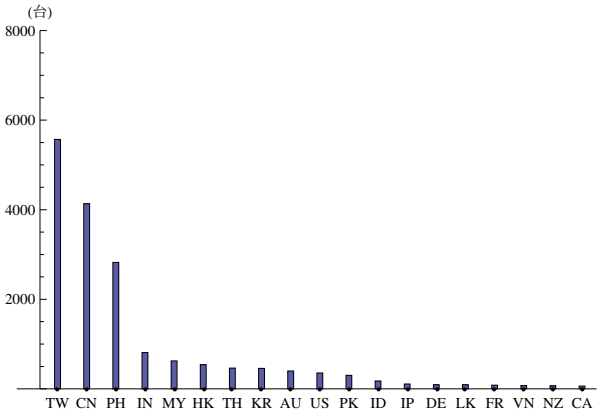

図 4: プル型配布元ノードの国別分布 (2 位以降)

```
//ノードの ClassB ネットワークを攻撃対象とする
//また，スキャン時間は 10000000ms とする
mFakeMsg.sChatString.Format(".scan.dcom    %s/16
10000000", szLocalIp);
//攻撃を開始
g_cMainCtrl.m_cScanner.
HandleCommand(&mFakeMsg);
```

図 5: Agobot における脆弱性スキャンの実装

されており，これらの組織は連続した IP アドレス空間をネットワーク事業者などのアドレス管理者に割り振る．そのため，感染したノードが近隣のアドレス空間に対してスキャンを開始すると，必然的に近隣のアドレス空間を利用している日本国内へのスキャンと攻撃が多くなり，ボットネットは同一国内のノードを多く含む結果となる．そのため，今回のデータセットでは配布元ノードの国として日本が多くなっていると考えれられる．

3.3 国と転送方法の関係についての考察

図 2 と図 3 を比較すると，多くの国において転送方法に傾向がみられる．たとえば，US(アメリカ) では 94.3%,KR(韓国) では 92.0%の配布元ノードがプッシュ型の転送であるのに対し，日本では 96.4%,PH(フィリピン) では 91.1%の配布元ノードがプル型の転送と，転送方法が偏っている．

日本においてプッシュ型転送が少ない理由として国内における家庭のインターネット接続形態が挙げられる．プッシュ型転送はインターネットからの接続を受け付ける必要があるので，ブロードバンドルータなどの NAPT 機能やファイアウォール機能によって外部からの通信が制限されたネットワーク内ノードでは機能しない．日本では多くの家庭において複数台のノードが存在しているため，NAPT 機能などにより外部からの通信が制限されているネットワークが多く，プッシュ型の転送によって配布するダウンローダを用いるマルウェア自体が拡散できないのではないかと考える．

4 時差を考慮したマルウェアの活動状況の分析

4.1 各国の１日における攻撃ノード数の割合

配布元ノードとなるノードはインターネットに接続している時期にしか活動できない．そこで国に着目した際の 1 日の配布元ノード数の分布は国民の活動時間帯と関係すると想定できる．

研究用データセットの期間中おける h 時から $h+1$ 時の平均的なマルウェアの配布元ノードの数を $E_{(h)}$ とする．ここで平均的な 1 日における h 時から $h+1$ 時の一意なマルウェアの配布元ノードの数の割合を $TP_{(h)}$ は，

$$TP_{(h)} = \frac{E_{(h)}}{\sum_{n=0}^{23} E_{(n)}}$$

である．

日本，アメリカと TW(台湾) における時間毎の平均的な攻撃元ノード数の分布を図 6 に示す．

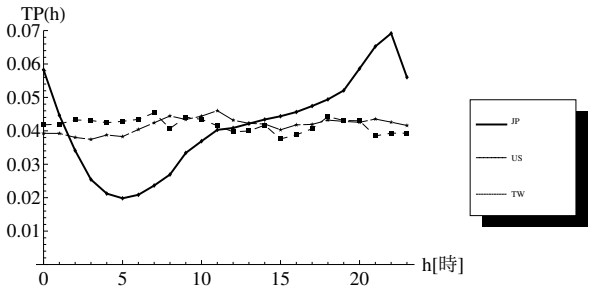

図 6: 日本，アメリカ，台湾からの１日の攻撃ノードの分布

4.2 国内からの転送数とトラフィックの関係についての考察

日本に着目すると，明朝はイベントの割合は少なく，おおよそ 9 時にかけて割合が増加し，しばらく同様の割合が続き，さらに 21 時頃から 23 時頃まで

増加している．この傾向は，日本の商用 ISP の接続点である NSPIXP2 のトラフィック傾向と類似している [6]．これは，第 3.2 節の考察と同じように，転送がボットネットの性質と深く関連しているからと考えられる．NSPIXP2 のトラフィックが多い時間帯はボットを含む日本国内のインターネットに接続しているノード数が多いと考えられ，国内への攻撃が増加する．その結果，国内の転送イベントが増加する．

台湾とアメリカにおいては，仮定が正しいとするのであれば，日本と同様の傾向あるいは時差分だけ X 軸にシフトしたグラフが描かれるが，そのような傾向はみられない．今回は研究用データセットからこの要因を推測できなかった．今後，攻撃のイベントデータなどと関連を調べて調査する必要がある．

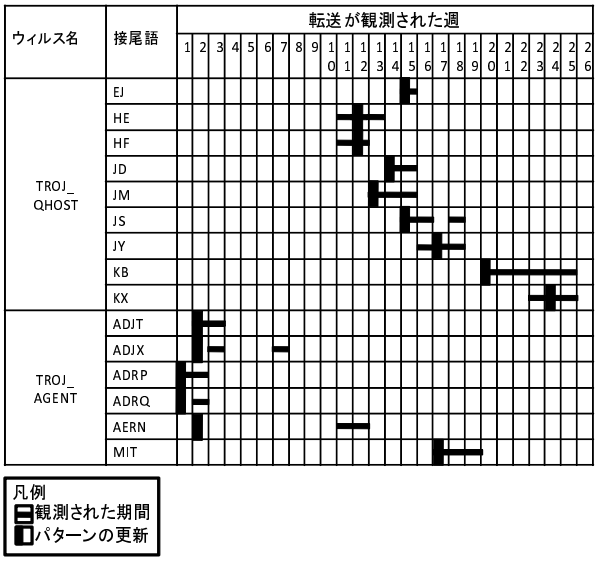

図 7: ノード α から転送が観測されたマルウェア

5 マルウェア配布サイトに関する分析

5.1 マルウェアの発生時期

研究用データセットには 2,942,221 件のレコードが含まれており，28.10%にあたる 826,962 件は特定の ノードα との転送イベントである．この ノードα はカナダに割り振られた IP アドレスを利用している．さらにカナダが配布元ノードとなっているイベント総数 878,084 件のうち 94.1%を ノードα が占めている．ノードα からの転送はすべてプル型転送であり調査期間の毎日，何らかの転送イベントが記録されている．また，期間によって転送されるマルウェアの種別が異なる．本節では，ノードα から配布されるマルウェアとその時期について考察する．

図 7 に ノードα から送信された代表的な 2 種類のマルウェアと，そのマルウェアの転送イベントがどの週に観測されたのかを示す．なお，イベント記録時点でハッシュ値 α を持つマルウェアが UNKNOWN と判定されていても，以降のイベントでハッシュ値 α となるマルウェアはマルウェア α と記録されていれば，ハッシュ値 α のマルウェアすべてをマルウェア α として扱う．

図の各列は取得したマルウェアをトレントマイクロ社のウィルス検知エンジンで検索した結果のウィルス名，接尾語と転送イベントが発生した週を示す．転送が観測された週とは，観測が開始された 2007 年 11 月 1 日から同月 7 日までを第 1 週，同月 8 日から同月 14 日までを第 2 週とした際の時期を示す．各行はマルウェアの名称と，そのマルウェアが観測された期間を示す．また，そのマルウェア亜種についてトレンドマイクロ社のウィルス検知パターンファイル (パターンファイル) が更新された最も近い時期を示す．たとえば，$TROJ\_QHOST.EJ$ は第 15 週にパターンファイルが更新されたが，同週に転送が観測されていることがわかる．

今回のマルウェアの命名ルールはトレンドマイクロ社のウィルス命名ルール [7] に従っている．例えば，TROJ_QHOST.EJ の接頭語 $TROJ$ はこの丸ウェアがトロイの木馬の一種であり，$QHOHT$ はマルウェアの名を示す．接尾語は亜種情報を示す．接尾語アルファベットが若いほど，早くに発見されたものであることを示す．図内では期間中，亜種 EJ がノード α から送信された $TROJ\_QHOST$ のもっとも最初に発見された亜種である．ただし，過去に 1 度発見されたマルウェアが特性の変化によって再度パターンファイルが更新される事例もある．

図よりノード α から転送するマルウェアは頻繁変わっていることが読み取れる．$TROJ\_QHOST$ の亜種に着目すると，第 11 週から亜種 HE, HF の転送が観測された．これらの亜種に対するパターンファイルは翌週更新された．第 14 週から第 15 週にかけて亜種 EJ, JM, JD, JS の転送が観測されている．第 19 週にはこれらの亜種の転送はすべて観測されなくなり，代わって亜種 KB の転送が観測された．そして，第 23 週に亜種 KX がの転送が観測さ

れ，翌週にパターンファイルが更新された．

また，他のマルウェアでも同様の傾向がみられる．$TROJ\_AGENT$ は第 1 週から第 3 週にかけて，亜種 $ADJT, ADJX, ADRP, ADRQ$ と複数の亜種が観測された．第 2 週には亜種 $AERN$ のパターンファイルが更新されたが ノードα からの転送は観測されていない．第 4 週からしばらく $TROJ\_AGENT$ は観測されなかったが，第 7 週には亜種 $ADJX$, 第 11,12 週には亜種 $AERN$ の転送が短期的に確認された．再び，第 17 週には接尾語が大きく異なる亜種 MIT が観測された．

5.2 パターンファイル更新時期とマルウェア発生時期の関係についての考察

これら 2 つのウィルスにおいて，多くの亜種の最初の転送はパターンファイルの更新とほぼ同時期に発生している．また，$TROJ\_QHOST$ の亜種 HE, HF, JY, KX や $TROJ\_QHOST.KX$ などのいくつかのマルウェアでは観測の時期がパターンファイルの更新よりもかなり前に観測されており，パターンファイルの更新が新しいマルウェアの出現に追い付いていない様子がわかる．さらに，大半のマルウェアは最初の転送の観測から 2,3 週間以内に転送が観測されなくなっており，ボットネットの拡散に用いられるマルウェアが古いマルウェアから新しいマルウェアに移り変わっている．これらの傾向から，攻撃者がウィルス検知エンジンによるマルウェアの検知を避けるために常に新しいマルウェアを利用している様子が見られた．

6 まとめ

本稿では，マルウェアが感染する手順の 1 つである転送に着目し，日本国内のハニーポットで取得された転送イベントのデータセットを分析した．

まず，データセットに含まれる配布元ノードの国はほぼ日本である．これは，多くのマルウェアにおける感染手法が感染したノードの近隣のネットワークを対象とする性質と相関があると考えられる．また，転送をプッシュ型とプル型の 2 種類に分けて考えると，国によって，転送の種類に偏りがある．特に日本ではプル型の転送が多く，多くの家庭において NAPT 装置を利用している背景が関係すると考えられる．

次に，1 日における配布元ノードの数を国ごとに調査した．日本では，インターネットのトラフィック量と配布元ノードの数に深い関係がみられ，インターネットに接続しているノード数と転送イベント数に相関があるといえる．ただし，日本以外の国では生活と関係した配布元ノードの数の関係が見られなかった．今後，転送が発生した原因となる攻撃のログと合わせた解析などが必要と考えられる．

最後に，データセットに含まれる特殊な配布元ノードが転送するマルウェアの移り変わりに注目した．この配布元ノードは観測期間中の毎日において何らかのマルウェア転送が観測され，転送されるマルウェアの多くの移り変わりはトレンドマイクロ社のパターンファイル更新の時期と一致している．近年の攻撃は検知を避けるために様々な手段が取られているといわれているが，その 1 つの事例として，転送するマルウェアを常にアップデートしている配布元ノードを発見できた．

本稿の分析を通して，配布元ノードの国はハニーポットを設置した国と関係が深いと考えられ，また，特殊な配布元ノードを観察し続けると新種のマルウェアを観測できると期待できる．今後は，転送だけではなく攻撃の情報を含んだデータセットと比較し，日本以外の 1 日における配布元ノードの数に特徴が見受けられない原因や，攻撃元ノードと配布元ノードの国の関係など新たな視点での分析など，ハニーポットで取得した情報の特性をより深く調査していく．

参考文献

[1] Trend Micro. ボットネットの脅威とソリューション：フィッシング, Oct 2006. http://jp.trendmicro.com/imperia/md/content/jp/threat/specificsolutions%/trendphishingwhitepaper.pdf.

[2] IPA. ボット感染予防推進事業の開始について, Feb 2008. http://www.ipa.go.jp/security/isg/bot.html.

[3] Paul Bacher,Thorsten Holz,Markus Kotter,Georg Wicherski. Know your enemy: Tracking botnets, May 2005. http://www.honeynet.org/papers/bots/.

[4] MaxMind. GeoIP, 2008. http://www.maxmind.com/app/ip-location.

[5] V. Yegneswaran P. Barford. An inside look at botnets. *Special Workshop on Malware Detection, Advances in Information Security, Springer Verlag*, 2006.

[6] WIDE Project. NSPIXP2 Traffic, Sep 2007. http://nspixp.wide.ad.jp/2/.

[7] Trend Micro. ウイルス名の付け方, 2008. http://jp.trendmicro.com/jp/threat/threat_types/name/.