今後脅威となりうるマルウェア配布元ホストの早期発見に関する一考察

大井 俊介 †　　加藤 貴司 †　　Bhed Bahadur Bista†　　高田 豊雄 †

† 岩手県立大学大学院　ソフトウェア情報学研究科
020-0193 岩手県岩手郡滝沢村滝沢字巣子 152-52

g231g009@edu.soft.iwate-pu.ac.jp, {k-kato,bbb,takata}@soft.iwate-pu.ac.jp

あらまし　マルウェアの配布元ホストは，活動期間やマルウェアの転送ログなどにおいて様々な特徴が現れる．マルウェア配布元ホストの中から，ウィルス対策ソフト等で検知されない，未知のマルウェアを積極的に配布するホストを早期に発見することは，マルウェア感染ホストの拡大対策として効果的であると考える．本稿では，研究用データセット CCC DATAset 2008・2009 の攻撃元データ，及び攻撃通信データより，マルウェア配布元ホストの特徴を分析し，今後脅威となりうるホストの早期発見を目的とした分析結果と一考察について述べる．

Early Detection of Hosts which Distribute Malware that will Become the Threat in Future

Shunsuke Ohi†　　Takashi Katoh†　　Bhed Bahadur Bista†　　Toyoo Takata†

†Graduate School of Software and Information Science, Iwate Prefectural University
152-52 Sugo, Takizawa, Iwate 020-0193 Japan

g231g009@edu.soft.iwate-pu.ac.jp, {k-kato,bbb,takata}@soft.iwate-pu.ac.jp

Abstract　Malware distribution hosts has various kind of characteristics such as active period or traffic pattern. Early detection of malware distribution hosts that distribute unknown malwares is important for effective countermeasure of outbreak of them. In this paper, based on CCC DATAset 2008 and 2009, we analyze attack source data and attack traffic data, and show characteristics of malware distribution hosts which is effective for early detection of potential threat.

1　はじめに

インターネットの利用が広く普及し，それに伴い，マルウェアによる被害が増加している．

マルウェアとは悪意のあるソフトウェアの総称であり，その中でもボットと呼ばれるマルウェアの被害が増大している．ボットに感染したホストは，DoS 攻撃やスパムメール送信などの踏み台となり，意図せずコンピュータ犯罪の加害者になりうる危険性を秘めている．また，ボットは非常に多くの亜種が出回り，一般的なシグネチャの照合によるウィルス対策ソフトなどでの対策が難しい．さらに，ボット同士でネットワークを構築するボットネットの規模性の高さにより，攻撃元の特定も困難である [1].

従来のボットの研究は，ボットのバイナリ解析や，感染後のホストのレジストリやファイルの走査，ボット感染後のインターネットトラフィックの分析などが主な研究内容となっており，マルウェア配布元ホストの分析に絞った研究は少ない．

本稿では，サイバークリーンセンター (CCC) によって収集されたインターネットのログから成る，ボットネット観測研究用データセット “CCC DATAset 2008・2009” [3] を用いて，長期間と短期間から確認できるマルウェア配布元ホストの特徴を分析し，今後脅威と成り得るマルウェア配布元ホストの早期発見を目的とした一考察について述べる．

2 今後脅威となりうるマルウェア配布元ホストの検出に向けた分析項目の設定

本稿では，ボットに感染したホストが，マルウェア配布元ホストからマルウェアをダウンロードする行動に着目する．使用するデータは，CCC DATAset 2008・2009 における攻撃元データ，及び攻撃通信データである．

ボットに感染したホストは，インターネット上に配置された，マルウェアを所持する Web サーバや TFTP サーバ（以下，マルウェア配布元ホスト）にアクセスし，マルウェアのダウンロードやアップデートを試みる．このマルウェア配布元ホストは，活動期間やマルウェアの転送ログなどにおいて様々な特徴が現れることが既存研究により示された [2].

本稿では，なるべく早期にマルウェア配布元ホストの活動や傾向を特定し，被害の拡大を防ぐ指標となる情報を示す．そのために，シグネチャの不一致などによりウィルス対策ソフトに検知されないマルウェア (以下，UNKNOWN マルウェア) を配布するホストの早期特定と，UNKNOWN マルウェアを配布するホストになりうるホストの予想分析を行う．

本稿での分析項目を以下に述べる．

- 長期間から見るマルウェア配布元ホストの活動分析 (3.2 節)
- 短期間から見るマルウェア配布元ホストの活動分析 (3.3 節)
- 短期間から見る UNKNOWN マルウェア配布元ホストの活動分析 (3.4 節)
- UNKNOWN マルウェア配布元ホストとなりうるホストの特徴分析 (3.5 節)

3 マルウェア配布元ホストの分析

3.1 分析に用いるデータ

本稿において使用するデータは，CCC DATAset 2008・2009 の攻撃元データ，及び攻撃通信データである．

攻撃元データは，2008 年 5 月 1 日から 2009 年 4 月 30 日までのハニーポット約 100 台のマルウェア検体取得ログであり，攻撃通信データは，ハニーポット 2 台の 2 日間 (2009 年 3 月 13 日 (以下，3 月 13 日)，2009 年 3 月 14 日 (以下，3 月 14 日)) のパケットキャプチャデータである．攻撃通信データにおけるハニーポット 2 台は，いずれも攻撃元データのハニーポットの数に含まれる．

分析対象とするマルウェア配布元ホストは，攻撃通信データにマルウェア配布元ホストとして出現した，55 ホストを基準として分析を行う．

3.2 長期間から見るマルウェア配布元ホストの活動分析

長期的にボットの活動を見た場合の，マルウェア配布元ホストの特徴を示す．分析期間は 2008 年 5 月 1 日から 2009 年 4 月 30 日である．

月単位でマルウェア配布元ホストの集計を行った結果，55 ホスト中 48 ホストが，2009 年 3 月のみにマルウェア配布元ホストとして活動している．つまり，大半のマルウェア配布元ホストは短期間のみに出現している．長期的に活動するホストを図 1，図 2 で示す．

マルウェア配布元ホストを長期間からみた場合，長期的に活動しているマルウェア配布元ホストの生存確認程度の情報は得られたが，ボットの拡大対策には役立て難い結果となった．

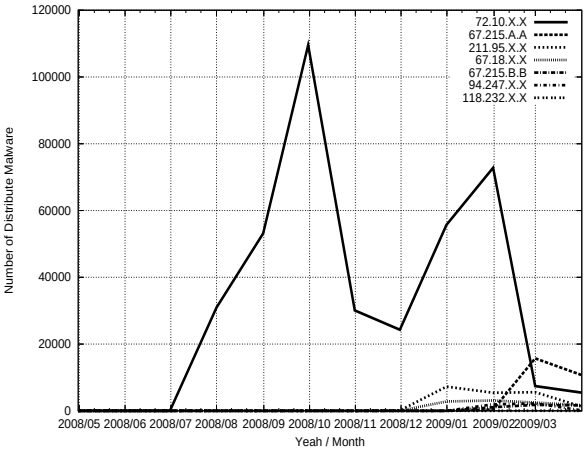

図 1: 長期的に活動するマルウェア配布元ホスト

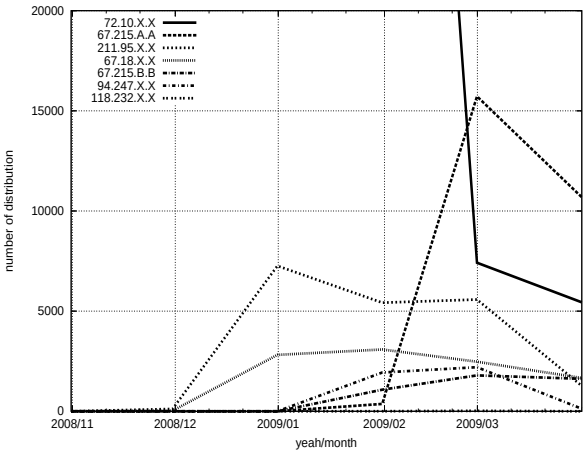

図 2: 長期的に活動するマルウェア配布元ホスト（一部拡大）

3.3 短期間から見るマルウェア配布元ホストの活動分析

長期的なホストの分析では，長期的に活動するマルウェア配布元ホストの傾向を把握できた．しかし，大半のマルウェア配布元ホストは，ある一定期間のみに出現するものであった．よって，マルウェア配布元ホストの情報をボットの被害拡大対策に利用するには，マルウェア配布元ホストを長期間から見た場合の分析だけではなく，短期間から見た場合の分析も必要となる．ダウンロード配布元ホストの早期発見に向け，長期的にマルウェア配布元ホストを見た場合との差異について述べる．分析期間は，3 月 13 日と 3 月 14 日の 2 日間である．

3 月 13 日において，全てのハニーポットにより取得されたマルウェア数は 3388 個であり，その中で攻撃通信データのログに出現した 55 ホストが配布したマルウェアの割合は 49.2%となった．3 月 14 日において，全てのハニーポットにより取得されたマルウェア数は 4241 個であり，その中で 55 ホストが配布した割合は 41.8%となった．両日に共通して現れたホストは 55 ホスト中 9 ホストあり，この 9 ホストは，図 1，図 2 に出現したホストと同一のホストが大部分を占める．9 ホストの内訳を表 1 に示す．

設定した 2 日間において，僅か 2 台のハニーポットを分析することで，全てのハニーポットで取得した約半数のマルウェアの配布元が特定可能であり，残り半数のマルウェアは，新出ホストから配布されるものであることを示した．つまり，長期間における分析と同程度の結果を，2 日間程度のログ情報で得られることを示した．

3.4 短期間から見る UNKNOWN マルウェア配布元ホストの活動分析

既知のマルウェアは，ウィルス対策ソフトの導入などにより対策が可能であるが，UNKNOWN マルウェアの対策に関しては，一般的に難しい問題である．

マルウェア配布元ホストの中から，UNKNOWN マルウェアを積極的に配布するホストを早期に発見し，そのホストとの通信を遮断することは，マルウェア感染ホストの拡大対策として効果的であると考える．分析期間は，3.3 節と同様に 3 月 13 日と 3 月 14 日の 2 日間である．

3 月 13 日において，UNKNOWN マルウェアの取得数は 234 個であり，3 月 14 日においては 378 個であった．UNKNOWN マルウェア配布元のホストは，3.2 節や 3.3 において出現した，長期的に活動するマルウェア配布元ホスト 4 台が 8 割を占めている．その内訳を表 2 に示し，残りのホストの内訳を表 3 に示す．表内の数字は，(1) そのホストが配布した UNKNOWN マルウェアの個数と，そのホストが配布した全てのマルウェアの個数，(2) そのホストが配布した全てのマルウェアにおいて UNKNOWN マル

表 1: UNKNOWN マルウェア配布元ホスト

ホスト / 日時	図 1，図 2 に出現したホスト					
	67.215.A.A	72.10.X.X	211.95.X.X	67.18.X.X	67.215.B.B	94.247.X.X
3/13	376	376	157	100	58	50
3/14	416	416	319	100	50	138

ホスト / 日時	図 1，図 2 に出現しなかったホスト		
	122.18.X.X	114.164.X.X	124.102.X.X
3/13	161	37	11
3/14	14	59	4

表 2: 短期間から見た UNKNOWN マルウェア配布元ホスト (既出ホスト)

	67.18.X.X	67.215.A.A	94.247.X.X	211.95.X.X
3/13	100 ／ 100（100%）	0 ／ 376（0%）	50 ／ 50（100%）	55 ／ 157（35.0%）
3/14	100 ／ 100（100%）	28 ／ 416（6.7%）	138 ／ 138（100%）	16 ／ 319（5.0%）

表 3: 短期間から見た UNKNOWN マルウェア配布元ホスト (新出ホスト)

	61.235.X.X	62.90.X.X	62.140.X.X	80.239.X.X
3/13	0 ／ 22 （0%）	6 ／ 6（75.0%）	none	1 ／ 1（100%）
3/14	30 ／ 56（53.6%）	29 ／ 29（100%）	1 ／ 1（100%）	none

	92.62.X.X	98.126.X.X	122.211.X.X	147.202.X.X
3/13	20 ／ 20（100%）	none	1 ／ 1（100%）	0 ／ 2（0%）
3/14	6 ／ 6（100%）	2 ／ 2（100%）	none	2 ／ 8（25.0%）

	195.47.X.X	208.75.A.A	208.75.B.B	219.167.X.X
3/13	none	none	none	1 ／ 1（100%）
3/14	12 ／ 12（100%）	5 ／ 5（100%）	9 ／ 9（100%）	none

ウェアが占める比率，から構成される．

UNKNOWN マルウェアの個数より考察すると，UNKNOWN マルウェアの配布量はホストの活動期間に比例して伸びると予想できるが，UNKNOWN マルウェアの比率より考察すると，必ずしもそうとは限らない．例えば，ホスト 72.10.X.X は活動期間は最も長いホストであったが，UNKNOWN マルウェアを配布した数は 0 である．また，3 月 14 日において最もマルウェア配布数が多いホスト 67.215.A.A の UNKNOWN マルウェアの配布量が，僅か 6.7% であるのに対し，マルウェア配布量がホスト 67.215.A.A の配布量の 4 分の 1 程度のホスト 67.18.X.X は，全てのマルウェアが UNKNOWN マルウェアであった．つまり，マルウェア配布元ホストの活動期間やマルウェア配布量と，UNKNOWN マルウェア配布量の関係性は薄い．これらの結果を踏まえ，3.5 節にて検証を行う．

3.5 UNKNOWN マルウェア配布元ホストとなりうるホストの特徴分析

UNKNOWN マルウェア配布元ホストについて検証することを以下に述べる．

- UNKNOWN マルウェア配布元ホストの所属国による分類 (3.5.1 節)
- UNKNOWN マルウェア配布元ホストとの通信において現れる特徴 (3.5.2 節)
- UNKNOWN マルウェア配布元ホストの応答速度 (3.5.3 節)

表 4: UNKNOWN マルウェア配布元ホストの所属国による分別

	US	CN	JP	その他
ホスト数	5	2	2	7

3.5.1 UNKNOWN マルウェア配布元ホストの所属国による分類

UNKNOWN 配布元ホストが，ある特定の地域に偏って所属している可能性も考えられるため，表 2，表 3 に出現した 16 ホストの所属国を WHOIS プロトコル [4] を用いて分類した．(表 4 [1]) その結果，所属国によって大きな偏りはなく．所属国の情報から UNKNOWN 配布元ホストの特定や推測は難しい．

3.5.2 UNKNOWN マルウェア配布元ホストとの通信において現れる特徴

攻撃通信データから確認できる，UNKNOWN 配布元ホスト 4 ホスト (表 2) の特徴を示す．

4 ホストに共通する特徴として，全ての UNKNOWN マルウェア転送において tcp/80 番ポートが使用されており，マルウェアの転送要求の際にハニーポットから送信される，SYN パケット，ACK パケット，RST パケット，FIN/ACK パケット，HTTP リクエストが 1 度に 2 回以上送信され，それによる重複パケットが発生する特徴を観測した．また，4 ホスト中 3 ホストとの通信において，ハニーポットが FIN/ACK パケットを用いた正規の接続切断を行った後に，RST パケットを送信する特徴を観測した．

しかし，これらの特徴は，UNKNOWN マルウェアを配布するホストのみならず，既知のマルウェアを配布するホストとの通信においても確認できた．さらに，これらの特徴は，感染したマルウェアの攻撃特性に左右されると考えられ，UNKNOWN マルウェア配布元ホストの早期特定に用いることは難しい．よって，本節で述べた特徴は，マルウェアの感染を検知する際の指標の一つとして用いるのがよいと考えられる．

3.5.3 UNKNOWN マルウェア配布元ホストの応答時間

これまでの分析結果において，UNKNOWN マルウェア配布元ホストの大半は，多数の UNKNOWN マルウェアや既知のマルウェアを積極的に配布していることを示した．つまり，UNKNOWN マルウェアの配布元ホストは，多数のボット感染ホストからの接続要求を受け続けていると考えられる．さらに，UNKNOWN マルウェアを配布するために，何らかの形でマルウェアのアップデートを行っており，既知のマルウェアのみを配布しているホストに比べると，ホストのネットワーク応答時間に何らかの特徴が現れると考えた．そこで，ハニーポットがマルウェア配布元ホストに接続要求を行い，接続が承認されるまでの時間に着目し，検証を行った．

検証の詳細は，攻撃通信データのハニーポット 2 台と通信を行ったマルウェア配布元ホストにおいて，ハニーポットがマルウェア取得の際，マルウェア配布元ホストとの 3 ウェイハンドシェイクを完了するまでの時間 (以下，接続完了時間) の平均時間と，接続完了時間における標準偏差，及び標準誤差を求めた．結果を表 5 に示す．

表 5 の結果より，UNKNOWN 配布元ホストと既知のマルウェア配布元ホストを比較すると，標準偏差，及び標準誤差で，UNKNOWN 配布元ホストの値が高くなる特徴を示した．例として，UNKNOWN マルウェアの配布元ホスト 67.215.A.A と，既知のマルウェアのみの配布元ホスト 72.10.X.X は，3 月 13 日と 3 月 14 日において同数のマルウェアを配布しており (表 1)，表 5 において平均接続完了時間の値が近い．しかし，ホスト 72.10.X.X はホスト 67.215.A.A に比べ，接続完了時間にばらつきが少ないため，標準偏差，及び標準誤差の値が小さい．つまり，ホスト 67.215.A.A は，マルウェアを配布する行動以外の何らかの行動の影響により，ハニーポットとの接続完了時間にばらつきが生じ，標準偏差，及び標準誤差の値が大きくなったと予想する．

[1] US：アメリカ合衆国，CN：中国，JP：日本，その他：イギリス，イスラエル，エストニア，カナダ，デンマーク，ドイツ，ラトビア

表 5: マルウェア配布元ホストの接続完了時間の分析

	UNKNOWN マルウェア配布元ホスト			
	67.18.X.X	67.215.A.A	94.247.X.X	211.95.X.X
平均接続完了時間 (ms)	168.8686667	210.6174583	303.2366667	196.8854444
標準偏差	7.361879938	15.98932249	12.77546462	12.22903586
標準誤差	3.292332797	3.334004209	9.033617665	4.323617091
	既知のマルウェア配布元ホスト			
	72.10.X.X	114.164.X.X	122.17.X.X	122.18.X.X
平均接続完了時間 (ms)	216.3158	16.83654545	9.943666667	9.884166667
標準偏差	4.417365658	3.086028792	1.550649828	1.290350394
標準誤差	1.472455219	0.975887991	1.096475009	0.577062239

接続完了時間の標準偏差，及び標準誤差の閾値の決定により，UNKNOWN マルウェア配布元ホストの指標となる情報を得られる可能性を示した．

4 まとめと今後の課題

本稿では，CCC DATAset 2008・2009 の攻撃元データと攻撃通信データを用いて，長期間と短期間から見た場合のマルウェア配布元ホストの特徴を示し，UNKNOWN マルウェア配布元ホストの早期発見を目的とした分析結果と一考察について述べた．以下に分析結果をまとめる．

長期間におけるマルウェア配布元ホストの分析では，約半数のマルウェア配布元ホストは特定できるが，残り半数のマルウェア配布元ホストを特定するのは困難であることを示した．また，短期間におけるマルウェア配布元ホストの分析では，長期間におけるマルウェア配布元ホストの分析と同程度の情報を得られることを示した．さらに，マルウェア対策ソフトで検知されない，UNKNOWN マルウェアを積極的に配布するホストの応答時間の検証を行い，UNKNOWN マルウェア配布元ホストの早期特定に役立つ指標を示した．

今後の課題として，3.5.3 節において検証した，マルウェア配布元ホストとの接続完了時間の標準偏差，及び標準誤差の閾値の検討と，評価が課題となる．また，本稿において分析できた UNKNOWN 配布元ホストとの通信データのサンプル数が少ないため，より多くの UNKNOWN 配布元ホストとの通信データを分析する必要がある．

今後の展望として，我々が研究・開発を行っている，個人参加型のインターネット観測システム ABLA[5] を用いて，マルウェア配布元ホストの情報を共有し，ネットワーク管理者をターゲットとしたボットの拡大予想システムを構築する予定である．

参考文献

[1] Markus Jakobsson, Zulfikar Ramzan：“Crimeware: Understanding New Attacks and Defenses”，Addison-Wesley（2008）．

[2] 石井宏樹，他：“ダウンロードホストに着目したマルウェアの活動分析”，MWS 2008，Vol.2008，No.8，pp.97–102，(2008)．

[3] 畑田充弘，他：マルウェア対策のための研究用データセットとワークショップを通じた研究成果の共有，MWS2009（2009 年 10 月）．

[4] RFC3912 - WHOIS protocol（2009/9/4）http://www.rfc-editor.org/rfc/rfc3912.txt

[5] 葛野弘樹，他：“モバイルエージェントを用いた分散型インターネット観測システムの提案”，情報処理学会論文誌，Vol.47，No.5，pp.1393–1405（2006）．