複数観測データを用いたボットネットの活動分析に関する一考察

畑田 充弘[†]　　寺田 真敏[‡]
[†]NTT コミュニケーションズ株式会社
〒163-1421 東京都新宿区西新宿 3-20-2 東京オペラシティタワー21F
[‡]中央大学大学院　理工学研究科
〒112-8551　東京都文京区春日 1-13-27

概要　近年，様々な経路での感染活動や，大量のスパムメール送信，DDoS 攻撃，フィッシングサイト開設など，ボットネットに起因する不正行為が社会的問題になっており，その豊富な機能や高度な運用によって実態の把握が困難であるといわれている．ボットネットの活動傾向を把握するために，本稿では，不正行為に関わる複数の観測データを用いたボットネットの活動分析結果について報告する．
キーワード　ボットネット，スパムメール，フィッシング，ブラックリスト

A Study of Botnet Activity Analysis using Multiple Monitoring Data

Mitsuhiro HATADA[†]　　Masato Terada[‡]
[†]NTT Communications Corporation
Tokyo Opera City Tower 21F, 3-20-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo, 163-1421 JAPAN
[‡]Graduate School of Science Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku Tokyo 112-8551, Japan

Abstract　Today, botnet activities are major concern such as various infection method, massive spreading E-mail spam, DDoS attacks and construction of phishing site. And it is hard to understand the activities of botnet by its rich functions and high practical operation. This paper shows some results of analysis about botnet activities using multiple monitoring data.
Keywords　Botnet, Spam Mail, Phishing, Black List

1. はじめに

近年，様々な経路での感染活動や，大量のスパムメール送信，DDoS 攻撃，フィッシングサイト開設など，ボットネットに起因する不正行為が社会的問題になっている．ボットネットはボットプログラムに感染した多数のコンピュータと C&C と呼ばれる制御サーバで構成され，Herder と呼ばれる攻撃者の命令により前述のような不正行為やボットプラグラム自身の機能追加等が行われ，活動傾向の把握が困難であるといわれている．一方で，文献[1]で整理されているように，ボットプログラムを含むマルウェアの感染検知や動的／静的解析，広域観測といった視点で数多くの研究が行われている．また，広域ネットワークから様々なレイヤの情報を収集して分析を支援する検討も行われている[2]が，不正行為に関わる複数の異なる観測データを用いたボットネットの活動傾向の分析例は少ない．本稿では，これらの分析例を示し考察を行う．

2. 観測データ

ボットネットの活動傾向を把握するための観測データには，自らハニーポットを設置する等のデータ収集環境を構築してログを収集した収集データと，インターネット上で公開され

ているようなブラックリストとして収集できる公開データに大別できる(図 1). 一般に，収集データは, 詳細なログを得ることができるが, ボットネットの活動傾向を分析する上では, 収集環境の網羅性が課題となる．公開データは, その元となるデータの収集方法や公開方法に起因して, 収集データと同様に網羅性の課題に加えて，信憑性や即時性が課題となる.

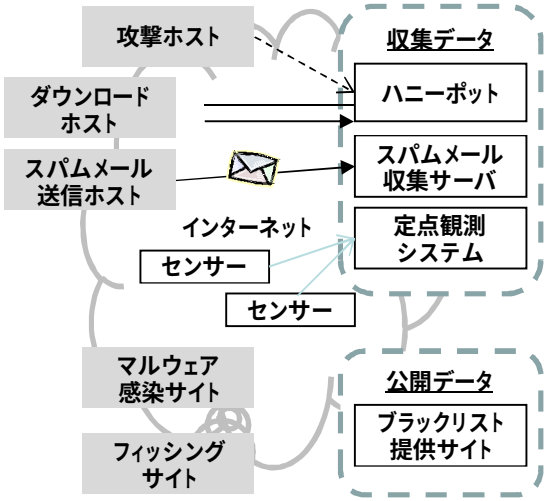

図 1. 観測データ

これらの観測データに関わる課題は今後の検討とし，本稿では，これらの複数の観測データを組み合わせたボットネットの活動傾向の分析結果について報告する. 以下, 本稿で使用した各観測データについて述べる.

2.1. 感染活動ログ

総務省と経済産業省の共同プロジェクトとして, ISP を通じたボット感染者への注意喚起等を行っているサイバークリーンセンター(CCC)[3]から提供された CCC DATAset 2008 の攻撃元データを用いる．攻撃元データには，多数のハニーポットで収集したマルウェアダウンロード時のタイムスタンプやダウンロードホスト IP アドレス，ウイルス名称等が含まれている (**観測データ No.1**).

2.2. スパムメールログ

広告サイトやフィッシングサイトへの誘導を目的として, ボットに感染した後に指令を受けて大量にスパムメールを送信したり, その踏み台となったりすることが知られている. スパムメール収集サーバを構築し, スパムメールの発信元と誘導先の IP アドレスを観測データとして利用する.

(1) 発信元 IP アドレス

収集した各スパムメールのヘッダから Received:フィールドを元に発信元 IP アドレスを抽出し, 観測データとして記録した (**観測データ No.2-1**). ヘッダ改ざんにより偽装された IP アドレス等も含んでいる可能性もあるが, 本稿では偽装がないものとして調査を実施した.

(2) 誘導先 IP アドレス

収集した各スパムメールの本文に含まれる URL を抽出し，直後に本文中の URL に基づいて, DNS クエリにより得た IP アドレスを観測データとして記録した (**観測データ No.2-2**).

2.3. マルウェア感染ブラックリスト

従来からのリモート攻撃型の感染活動とは異なり，Web ブラウザの脆弱性を突いて感染する Web 感染型マルウェアもボットネットの感染活動の一つとして問題になっており実態調査などが行われている[4]. このようなマルウェア感染への対策の一例として, マルウェアに感染させられるサイトの発見者によって, 登録及び検証されたブラックリストが公開されている.

今回はその一例として，1 時間毎に更新される Malware Block List[5]を情報源として，1 日 2 回(日本時間 0 時及び 12 時)取得し，ブラックリストの取得後，都度リスト中の URL に基づいて, DNS クエリにより得た IP アドレスを観測データとして記録した (**観測データ No.3**) .

2.4. フィッシングサイトブラックリスト

インターネット利用者にとって直接的な被

害となる，オンラインバンクの ID／パスワード等の搾取のために，数多くのフィッシングサイトが日々開設されており，発見者によって，登録及び検証されたブラックリストが公開されている．

今回はその一例として，PhishTank[6]を情報源として，1 日 2 回(日本時間 6 時及び 18 時)，VALID かつ ONLINE 状態の最新の 1,000 件を取得し，2.3 同様に都度 IP アドレスを観測データとして記録した (**観測データ No.4**) ．

2.5. 定点観測データ

JPCERT/CC が運用している定点観測システム ISDAS[7]は，インターネット上にセンサーを分散配置し，脆弱点探索のためのスキャン活動等を把握するための情報として有用である．定点観測データには，攻撃検知時のタイムスタンプや発信元及び宛先の IP アドレス及びポート番号が含まれている (**観測データ No.5**) ．

3. 分析結果

各観測データの分析対象期間は 2008 年 4 月 1 日から 2008 年 4 月 30 日の 1 ヶ月間とした．分析対象期間において，観測データ No.1 のイベント数は 406,555 件，No.2-1 及び No.2-2 の元となるスパムメール収集数は 1,534,057 通，ユニークな誘導先 URL 数は 671,823 件，No.3 のユニークな URL 数は 1,713 件，No.4 のユニークな URL 数は 22,772 件，No.5 のイベント数は 264,970 件であった．

分析の方針として， 下記の組み合わせにおいて，全ての観測データの共通項となる IP アドレスと時間軸による関連付けを行う．

- 感染活動ログ(No.1)と各観測データ
- 感染活動ログ(No.1)と複数観測データ

3.1. 感染活動ログと各観測データ

まず，概況として各観測データから抽出したユニークな IP アドレス数を表 1 に示す．また，分析対象期間において観測データ No.1 と各観測データのユニーク IP アドレス一致数とその国別分布を表 2 に示す．

表 1. 観測データのユニーク IP アドレス数

No.	ユニーク IP アドレス数
1	37,914
2-1	412,953
2-2	1,155
3	905
4	9,463
5	52,498

表 2. 観測データ No.1 との一致数及び国別分布

No.	一致数	国（数）
2-1	75	JP(26), US(6), RU(6), RO(5), MX(5), Other(27)
2-2	7	KR(5), CN(1), JP(1)
3	1	IL(1)
4	7	KR(4), CN(1), PH(1), US(1)
5	3,690	JP(3054), CN(153), TW(94), KR(58), US(53), Other(278)

(1) 観測データ No.1 と No.2-1

一致した 75 個の IP アドレス全てについて，日単位での感染数とスパムメール受信数の推移から，増減の傾向に類似性が見える(図 2)．

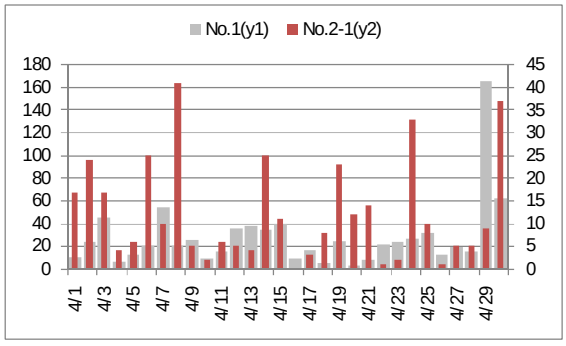

図 2. 感染数とスパムメール受信数

(2) 観測データ No.1 と No.2-2

一致した 7 個の IP アドレスのうち，No.4 と重複しない IP アドレスが 2 個存在した．こ

のうち, 例として 1 個の IP アドレスについて, 日単位での感染数と誘導先 URL が含まれるスパムメール受信数の推移を図 3 に抽出した. 図 3 から感染活動と誘導先が短い期間においてのみ連動して利用されていることがわかる. 誘導先 URL は 2 種類あり, TLD(トップレベルドメイン)が異なるが SLD(セカンドレベルドメイン)は同一のものであった.

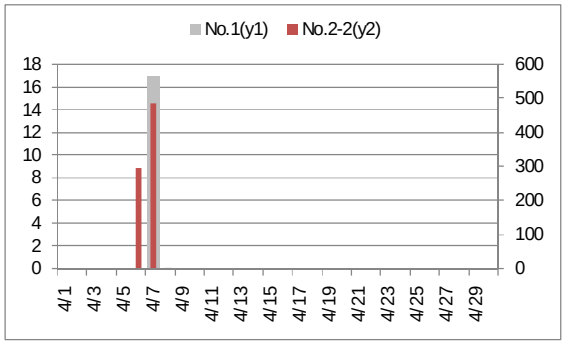

図 3. 感染数とスパムメール受信数

(3) 観測データ No.1 と No.3

一致した 1 個の IP アドレスについて, 日単位での感染数とブラックリストにユニークな URL が登録されていた期間(BL 登録期間)を図 4 に示す.

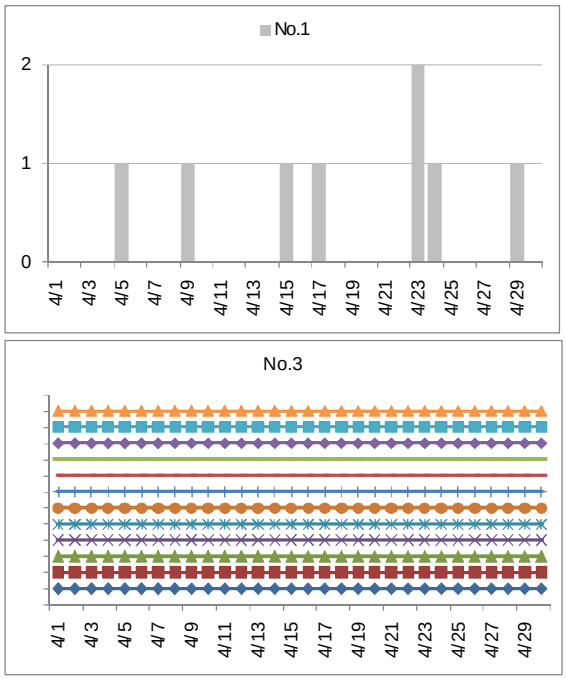

図 4. 感染数と BL 登録期間

感染数は少なく, 12 種類の URL の BL 登録期間に対してまばらであるのは, ブラックリストが公開されているため, 攻撃者側もリストに基づく対策を回避していることが考えられる. なお, 12 種類の URL は TLD, SLD が同一であり, 第 3 レベルドメインあるいはホスト部が異なる URL であった.

(4) 観測データ No.1 と No.4

一致した IP アドレスのうち, No.2-2 とも重複しない IP アドレスは 2 個存在した. このうち, 例として 1 個の IP アドレスについて, 日単位での感染数とブラックリストにユニークな URL が登録されていた期間をそれぞれ図 5 に示す.

フィッシングサイトとしての BL 登録期間は短期間ではあるが, 分析対象期間を通して感染活動が確認できる.

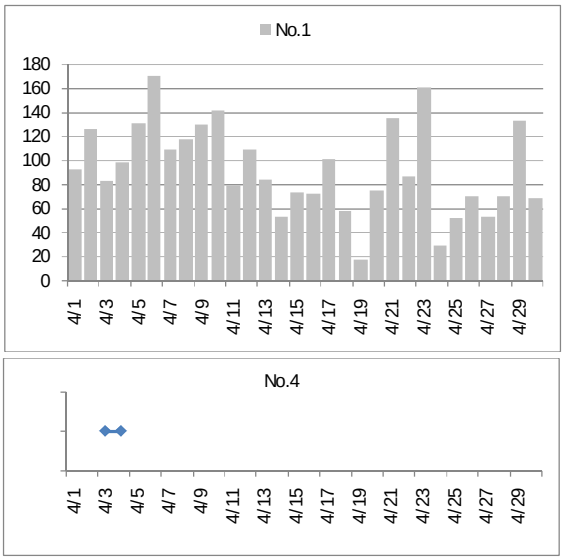

図 5. 感染数と BL 登録期間

(5) 観測データ No.1 と No.5

一致した IP アドレスのうち, 同一日に出現した IP アドレスの第 1 オクテットと第 2 オクテットによる頻度を No.1 と No.5 について算出した結果を図 6 に示す. これより, 12x.10x 付近のホスト出現が共通して多いことが確認できる. この出現頻度傾向は, 観測データ

No.1(図 7)だけの結果とは異なる.

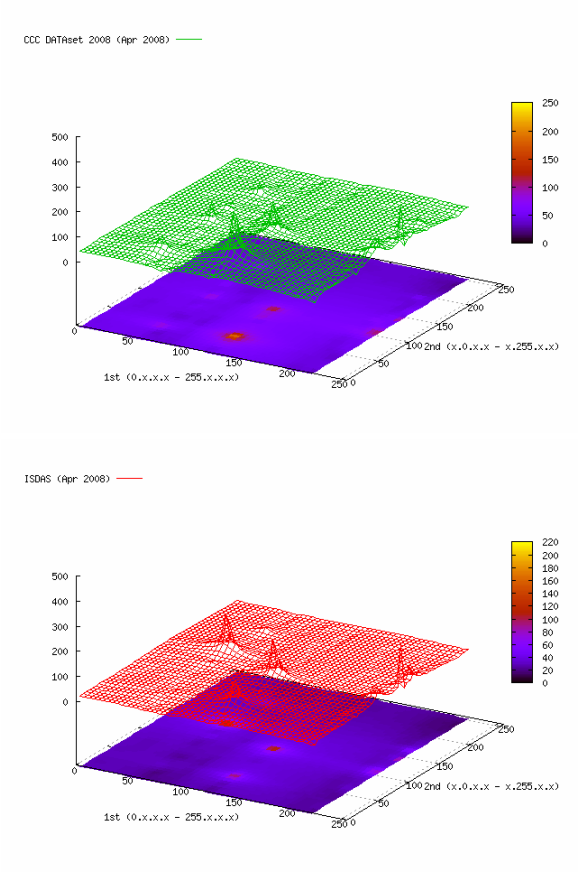

図 6. 同一日での出現頻度
(上段：観測データ No.1 下段：観測データ No.5)

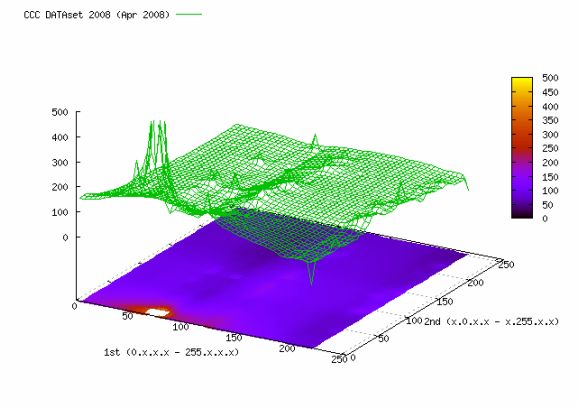

図 7. 観測データ No.1 の出現頻度

3.2. 感染活動ログと複数観測データ

(1) 観測データ No.1，No.2-2，No.4

5 個の IP アドレスが，観測データ No.1，No.2-2，No.4 で一致した．Whois 情報によると 4 個の IP アドレスは KR の同一 netname であり，1 個の IP アドレスは CN であった．また，観測データ No.1 では全て同一ハッシュ値のマルウェアを 80/TCP でダウンロードしており，4 月 7 日時点でトロイの木馬に分類される TROJ_MUTANT.AD として検出されるようになったものである．

KR の 1 個の IP アドレスについて，日単位で感染数と誘導先 URL が含まれるスパムメール受信数の推移，及び URL がフィッシングサイトのブラックリストに登録されていた期間を図 8 に示す．3 種類の観測データはともに 4 月前半に分布しており，誘導先 URL の 31 種類はフィッシングサイトのブラックリストに登録されていた URL とは異なるものであった．誘導先 URL31 種類の内，21 種類は 1 日のみ誘導先 URL として用いられていた．また，KR の 4 個の IP アドレスについての誘導先 URL やフィッシングサイトのブラックリスト URL に登録されていた URL には，共通するドメイン，FQDN が多数存在する．

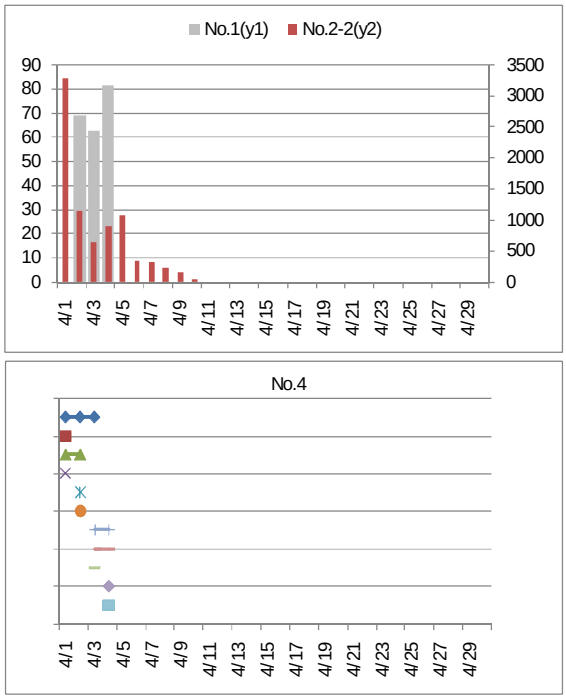

図 8. 感染数，スパム受信数，BL 登録期間

(2) 観測データ No.1，No.2-1，No.5

7 個の IP アドレスが観測データ No.1，No.2-1，No.5 で一致した．そのうちの 1 個の

IP アドレスについて，それぞれの観測データにおける検知状況及び No.5 の各センサーにおける検知状況を図 9，図 10 に示す．スパムメール発信元としての同ホストの利用に比べて，マルウェアのダウンロードや感染に至る攻撃元としての利用頻度が高い．これは前者がスパムメール送信依頼という受動的な活動であり，後者はボットネットの支配下に置くホストの拡大のため，ある程度能動的な活動であることが理由として考えられる．また，No.5 の各センサーにおける検知状況によると，当該 IP アドレスを攻撃元としたイベント検知は sid A 及び sid B の 2 センサーのみであったことから，攻撃活動自体はある程度局所的なものであると推定できる．

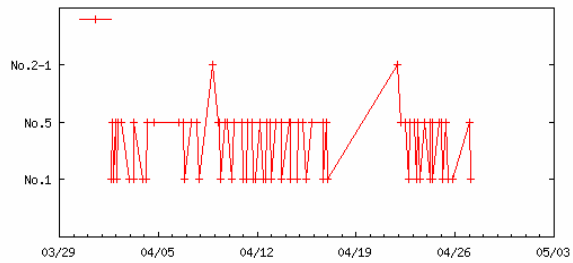

図 9．各観測データにおける検知状況

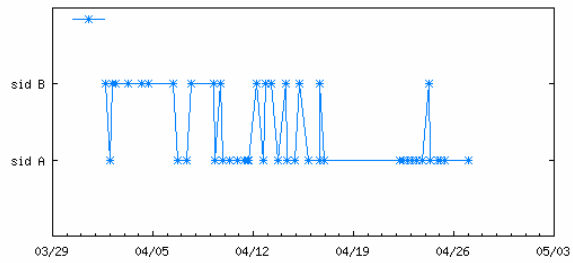

図 10．観測データ No.5 の各センサーにおける検知状況

4. おわりに

本稿では，ボットネットの活動の一端と考えられるインターネット上の不正行為に関わる複数の異なる観測データを用いて，IP アドレスと時間軸を中心に行った分析例を示した．

攻撃者にとっては，対策を困難にすることは望ましいことであり，バーチャルドメインによるホスティングサーバ群や，DNS による負荷分散されたサーバ群などが，ボットネットの運用に適しているため，攻撃の対象になりやすいことが十分に考えられる．加えて，URL 短縮サービスや検索エンジンを利用した Web ページのリダイレクト等を含む URL の多段リンク構造も同じ目的からボットネットの運用に適している．

今後，ますます巧妙な手法によりボットネットの活動傾向の把握，対策が困難になってくることが十分に考えられる．観測データの拡充や網羅性の検証に加えて，URL や IP アドレスに基づくブラックリストを複数の観測データに基づいてリアルタイムに更新管理し，各種フィルタ等の対策に活用する仕組みの検討が必要になると考えている．

謝辞

本研究の遂行にあたって，定点観測データならびに研究用データセットを提供頂いた JPCERT/CC とサイバークリーンセンターに感謝する．

参考文献

[1] 藤原将志，他：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)

[2] 鬼頭哲郎，他：マルチレイヤ型広域モニタリングに関する検討，情報処理学会 CSEC 研究報告 No.16 p279-284(2007 年 3 月)

[3] サイバークリーンセンター，https://www.ccc.go.jp/

[4] 秋山満昭，他：クライアントハニーポットを用いた Web 感染型マルウェアの実態調査，CSS2008(2008 年 10 月)

[5] Malware Block List，http://www.malware.com.br/index.shtml

[6] PhishTank，http://www.phishtank.com/

[7] ISDAS，http://www.jpcert.or.jp/isdas/