通信トラフィックの時系列分析によるボット活動の可視化と特徴検出

池田　潤一[1]　　岩村　誠[1,2]　　秋岡　明香[3]　　村岡　洋一[4]

[1]早稲田大学大学院基幹理工学研究科
169-8555 東京都新宿区大久保3-4-1
{junichi.ikeda, iwamura.makoto}@muraoka.info.waseda.ac.jp

[2]NTT 情報流通プラットフォーム研究所
180-8585 東京都武蔵野市緑町3-9-11
iwamura.makoto@lab.ntt.co.jp

[3]電気通信大学情報システム学研究科情報ネットワークシステム学専攻
182-8585 東京都調布市調布ヶ丘1-5-1
akioka@is.uec.ac.jp

[4]早稲田大学理工学術院
169-8555 東京都新宿区大久保3-4-1
muraoka@waseda.jp

あらまし　本研究の目的は, ボットネット対策のために, ボット通信の特性や傾向を把握することである. 本研究では, 研究用データセット CCC DATASet 2009 の攻撃通信データを用い, ボットに感染した端末の通信相手について時系列分析を行った. 実験の結果, ボットの典型的な各挙動について可視化を行い, ネットワークトラフィックから識別可能なことが分かった.

Visualizing and Detecting Bot Activities by Time-series Analysis of Network Traffic

Junichi Ikeda[1]　　Makoto Iwamura[1,2]　　Sayaka Akioka[3]　　Yoichi Muraoka[4]

[1]Graduate School of Fundamental Science and Engineering, Waseda University
3-4-1 Ohkubo, Shinjuku-ku, Tokyo, 169-8555 Japan
{junichi.ikeda, iwamura.makoto}@muraoka.info.waseda.ac.jp

[2]NTT Information Sharing Platform Laboratories
9-11, Midori-Cho 3-Chome Musashino-Shi, Tokyo 180-8585 Japan
iwamura.makoto@lab.ntt.co.jp

[3]Graduate School of Information Systems, The University of Electro-Communications
1-5-1, Chofugaoka, Chofu-shi, Tokyo, 182-8585
akioka@is.uec.ac.jp

[4]Faculty of Science and Engineering, Waseda University
3-4-1 Ohkubo, Shinjuku-ku, Tokyo, 169-8555 Japan
muraoka@waseda.jp

Abstract　The purpose of this study is taking measure of bot net by understanding the characteristic and the tendency to the bot communication. We used the attack communication data of CCC DATASet 2009 on time-series analysis of the other party of the communication of an infected terminal with bot. From the results of the experiments, we understood to identify typical bot activities from the network traffic.

1. はじめに

コンピュータシステムが社会基盤の一部として定着するにつれ, その障害がもたらす社会的損失も大きなものになっている. しかし、近年インターネットに接続した端末からのDDoS, ウィルスなどの妨害攻撃, 情報漏えいやspamなどセキュリティの問題が増加傾向にあり, 被害は拡大している. 特にセキュリティ侵害への対処方法や再発防止への要求が高まってきているが, これらの技術はまだ十分に確立されているとは言いがたく, 対策は急務に迫られており, 検知・防御が求められている.

中でもマルウェアの感染やメールによるウィルス, 悪意のあるファイルにより, 気付かずマルウェアを実行し感染することが多い. マルウェアによる脅威は，高度に複雑化され，脅威となるマルウェアやその配布元サイト，攻撃者の存在が隠蔽されている. マルウェアの機能を特定することは難しい[1]. このようなマルウェアの感染は，Webアプリケーションとクライアントアプリケーションの脆弱性が合わせて悪用されていることが多い.

マルウェアの中でも, ボットネットへの接続による被害が拡大している. ボットネットとは，ボットなどのウィルスによって外部の人間によりコントロールされるようになった複数のコンピュータをつなぐネットワークのことを指す. ボットは従来のワームやウィルスのように自動的に感染を拡大せず，Harderと呼ばれる攻撃者からの指令を受けて活動するため，その実態の把握が難しいといわれている. ボットネットに感染した端末はC&Cサーバからの命令により操作され, 情報の流出, DDoSやspamなど踏み台攻撃に利用される. また, ボットネット経由でC&Cのコントロールにより新たなマルウェアをダウンロード・更新し, 機能や目的を容易に変更しうる. そのため, ボットに感染した端末は時間の経過と共に挙動を変えうる. 加えてボットは攻撃者からの指令で行動をとるため, ユーザは感染に気がつきにくいという問題点がある. ボットによる被害が拡大している背景には, セキュリティ対策が不十分なコンピュータを, ネットワークに接続することなどが挙げられる. これらボットへの感染を防ぐためにも, ボットの挙動の傾向や特徴を知る必要がある.

本研究では, ハニーポットを利用したシステムを構築し, 得られた通信トラフィックのキャプチャ（研究データセットCCC DATASet 2009）に対して時系列分析を行う. 時系列に対し, 挙動を可視化することにより, ボットネットの実態や挙動を観測し, 時間と攻撃元・攻撃先IPアドレスの関連性を評価する. 得られた結果を評価することで, ボットの実態, 動向の観測を行う. また, どのIPにどのような攻撃が来やすいか考察することを目指す.

以下, 2章において, 既存の技術と研究とその課題について説明する. 3章で時系列分析を用いたボットの挙動解析手法について提案する. 4章で実験概要について説明し, 5章でそれら解析結果を述べる. 6章で考察し, 最後に7章でまとめる.

2. 関連研究

侵入検知や状況認識における可視化の研究はいくつかある. ネットワークの状況を表示したり, 時間的属性によりデータを表示するためにアニメーションを使用したり, 色を利用することにより, 利用者にとって状況の認識を支援しているシステムが多い.

警視庁@police[2] において, 定点観測が行われている. 攻撃手法や国別攻撃元などの統計が公開されており, それぞれの情報が可視化して表示されている.これらの情報は毎時15分頃に更新されるのでリアルタイム性にも優れ, 逐次情報を把握することができる.

IronPort Systemsは, 高精度のスパムメール検知機能を備えている. 電子メール管理者が, 自社のネットワークに電子メールおよびWebトラフィックを可視化できるようにした. インターネットコミュニティの各メンバーが、スパムの傾向やウィルスの大規模感染、スパイウェアやその他のWebベー スの脅威を、従来に比べより簡単に追跡できるようにすることを可能にした.

NVisionIP[3] はインタラクティブにネットワークフローを可視化するためのツールである. さまざまなネットワーク機器のインターフェースを通過するトラフィックの詳細なデータであるNetflowデータを用いることで, ネットワーク状態を示すために, 色を用いることでIPアドレスにおける

特定のポート活動を表示している. ただし, NVisionIPは現在の状況を把握し, 不正侵入検知における監視を支援する目的としていると考えられる. ゆえに現在の状況を把握するのには適しているが, 過去の挙動やどのような異常により現在の異常状況になったかなどの情報を得るためには向いていない.

上記のような技術や研究はあるが, ボットの挙動解析の問題として, ボットプログラムの進化, 機能の更新は大変早く頻繁に行われており, そのため, 対策を立てるための傾向や特徴を捉えることが困難になっている. このような問題点からボットの挙動や本来の目的を効率的に解析するために, 時系列分析により網羅的に通信トラフィックのキャプチャを解析するシステムを提案する.

3. 提案手法

3.1 時系列分析

本システムでは, ボットの挙動を解析する目的で, 攻撃通信データを用いて時系列分析を行う. 時系列分析とは, 時間経過ごとに記録されたデータ列や数値列からモデルを作成することである. また, その結果から, 全体の傾向や特徴を導き出す分析手法のことである. 本研究では, 時間軸に IP アドレスを関連付け, ボットの時間軸における外部からの攻撃やボット感染端末による感染活動などの挙動の傾向を導き出す.

3.2 システム概要

研究用データセット CCC DATASet2009 の攻撃通信データの中には, ボット感染端末と C&C サーバや攻撃端末などとの様々な双方向通信データが混ざっている. そのデータの中でも, TCP-syn のパケットと UDP のパケットそれぞれに注目し、それらを攻撃関連パケットと断定して解析を行った.

TCP-syn と UDP のパケットに注目し, これらをそれぞれ分類した. 一つ目はボットに感染した端末から外部への通信であり, 二つ目は外部からハニーポットへの通信である. この二つに分類し, ボットの挙動を解析した. ここでボットの挙動として, C&C サーバとの通信, 感染活動, マルウェアの更新などが挙げられるが, 本研究では, 感染活動のみに着目した.

感染活動の挙動を抽出するため, 横軸に時間, 縦軸に IP アドレスを取り, ボット感染端末からの通信とボット感染端末への通信それぞれの挙動をプロットする. これにより, 双方向の通信におけるボットの攻撃やデータのやりとりなどの挙動を網羅的に抽出することができ, 攻撃の特徴を視覚的に把握することができる.

4. 実験

4.1 実験環境

本研究で利用した実験環境の端末は, 以下の表 1 に示した通りである.

表 1: 実験環境

OS	Windows XP SP3
CPU	Intel Core (TM)2 Extreme 3GHz
メモリ	2GB

4.2 解析対象

本研究で用いた解析対象は, 研究用データセット CCC DATASet2009[4] である. このデータセットはサイバークリーンセンター[5] で収集しているボット観測データであり，マルウェア検体、攻撃通信データ，攻撃元データから構成されたボット観測データ群である．

本研究では, 研究データセットCCC DATASet2009 の攻撃通信データを利用した. CCC DATASet2009の攻撃通信データは，ハニーポット2台 (x.x.21.xとx.x.22.x)への通信を2日分フルキャプチャしたデータである. 識別情報 (IPアドレス) が含まれているが，検体名称が含まれていない. 本研究では, 2009年3月13日と2009年3日14日の2日分のデータを対象にした. これらデータを解析し, データの時刻情報とIPアドレス情報を照合することで，ボットの挙動を抽出した.

対象データを 3.2 節で述べたように分類した場合の, 各分類におけるパケット数を表 2 に示す. この表より, UDP での攻撃が TCP-syn よりも多いことがわかる.

表 2: 2 日分の攻撃通信データの内訳

	パケット内容	パケット数
ボットからの通信	TCP-syn で送信元 IP が x.x.21.x か x.x.22.x のパケット	481336
	UDP で送信元 IP が x.x.21.x か x.x.22.x のパケット	850046
ボットへの通信	TCP-syn で送信元 IP が x.x.21.x と x.x.22.x 以外のパケット	4585
	UDP で送信元 IP が x.x.21.x と x.x.22.x 以外のパケット	409640

5. 実験結果

本章では, 研究用データセット CCC DATASet2009 の攻撃通信データの TCP-syn のパケットと UDP のパケットを時系列分析した結果について述べる.

5.1 TCP-syn に注目したボットからの通信

ここでは, ハニーポットから外部への通信の中で, TCP-syn のパケットに注目した結果を述べる. 解析対象のパケットから TCP-syn で送信元 IP が x.x.21.x か x.x.22.x のパケットを対象とした. 横軸にパケット送信時刻, 縦軸に宛先 IP アドレスをプロットしたものを図 1, 2 に示す.(ここで図 1 は 2009 年 3 月 13 日, 図 2 は 2009 年 3 月 14 日のものである)

ここで図 1, 図 2 では宛先アドレスの範囲が広すぎるため, 詳細な分析をすることができない. そこで, 図 1 の宛先 IP アドレス 118.8.200.0/22 部分を拡大したものを図 3 に示す.

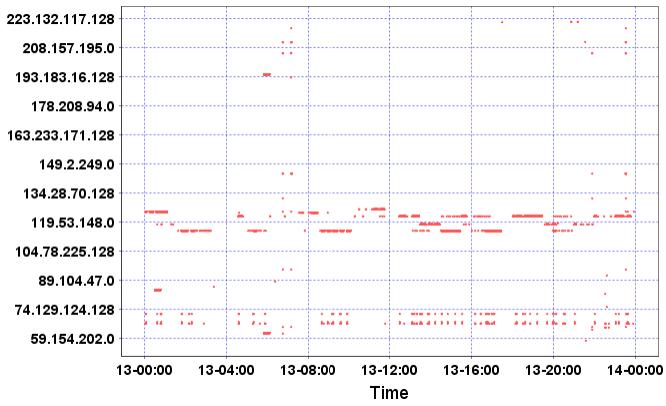

図 1: TCP-syn に注目したボットからの通信 (2009 年 3 月 13 日)

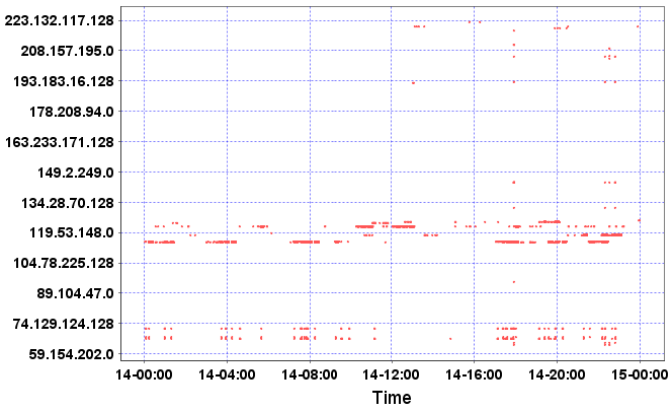

図 2: TCP-syn に注目したボットからの通信 (2009 年 3 月 14 日).

図 3: 118.8.200.0/22 付近の拡大図 (2009 年 3 月 13 日)

図 3 より, 縦方向の筋を確認することができる. これは特定のアドレスに対して網羅的にアクセスしている攻撃であり, アクセススキャン型と呼ぶ. このようなアドレススキャン型の攻撃がいたるところで確認できた.

5.2　TCP-syn に注目したボットへの通信

ここでは, 外部からハニーポットへの通信の中で, TCP-syn のパケットに注目した結果を述べる. 解析対象のパケットから TCP-syn で送信元 IP が x.x.21.x と x.x.22.x 以外のパケットを対象とした. 横軸にパケット送信時刻, 縦軸に送信元 IP アドレスをプロットしたものを図 4, 5 に示す. (ここで図 4 は 2009 年 3 月 13 日, 図 5 は 2009 年 3 月 14 日のものである)

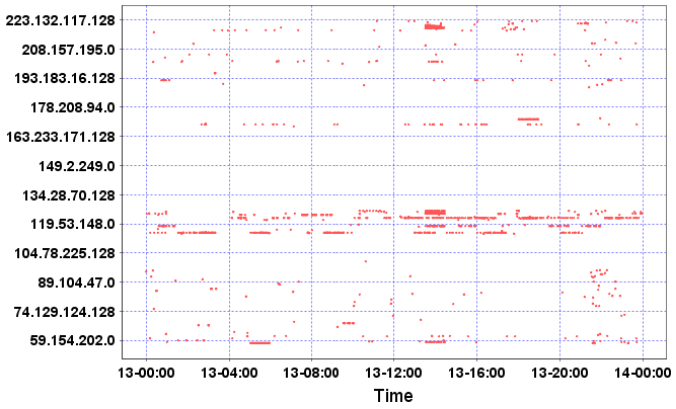

図 4: TCP-syn に注目したボットへの通信
(2009 年 3 月 13 日)

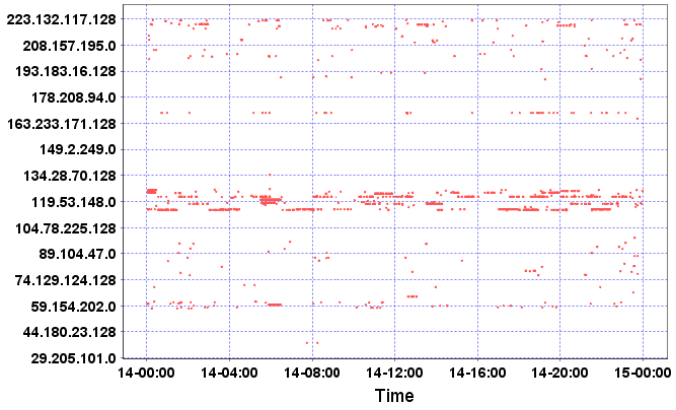

図 5: TCP-syn に注目したボットへの通信
(2009 年 3 月 14 日)

図 4, 5 の中央に見られる帯は, 特定のアドレス付近から継続的なアクセスが行われた様子である. その中でも port135, 139, 445 への攻撃が多く見られた. port135 のペイロードを観察したところ, REMACT といったリモートアクセスを仕掛けるパケットを確認し, Blaster であると推察できた. Port445 への攻撃は SMB であった. また, 13 日 14:00 あたりに port33887 へ集中的に攻撃があった.

5.3　UDP に注目したボットからの通信

ここでは, ハニーポットから外部への通信の中で, UDP のパケットに注目した結果を述べる. 解析対象のパケットから UDP で送信元 IP が x.x.21.x か x.x.22.x のパケットを対象とした. 横軸にパケット送信時刻, 縦軸に宛先 IP アドレスを宛先ポート番号ごとにプロットしたものを図 6 に示す.

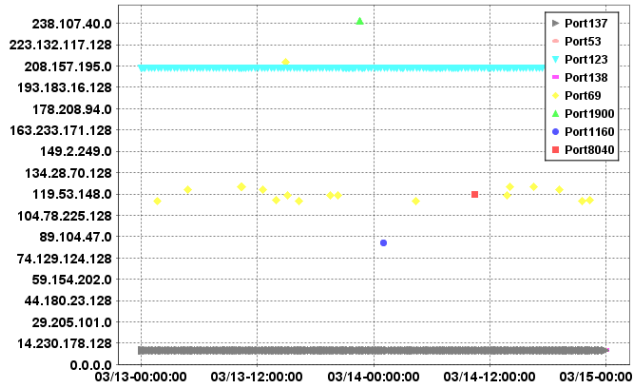

図6: UDPのパケットに注目したボットからの通信(2日分)

図6からわかるように横方向の筋を観測することができる. 図6での下にある横方向の筋はport53,への通信があった. port53はDNSとの通信であり, その詳細を分析すると, 特定のドメインを引いているものがあった. これはそのサイトにアクセスし, ファイルをダウンロードするためのアクセスであると考えられる. もう一つは, AレコードでIPアドレスを引いているものがあった. この意図は不明であるが, そのような特徴のある挙動を観測することができた. port137, 138は, ボット感染端末の周囲のIPアドレスのNetBIOSに対して継続的に通信している. この結果から, ボットは周囲の身近な端末に対して通信する傾向があることがわかる.

また, 図6での上にある横方向の筋は, port123への通信が確認できた. port123はNTPであり, 自分自身がインターネットに繋がっているかを確認するための通信であると考えられる.

5.4　UDP に注目したボットへの通信

ここでは, 外部からハニーポットへの通信の中で, UDP のパケットに注目した結果を述べる. 解析対象のパケットから UDP で送信元 IP が x.x.21.x と x.x.22.x 以外のパケットを対象とした. 横軸にパケット送信時刻, 縦軸に宛先 IP アドレスを宛先ポート番号ごとにプロットしたものを図 7 に示す.

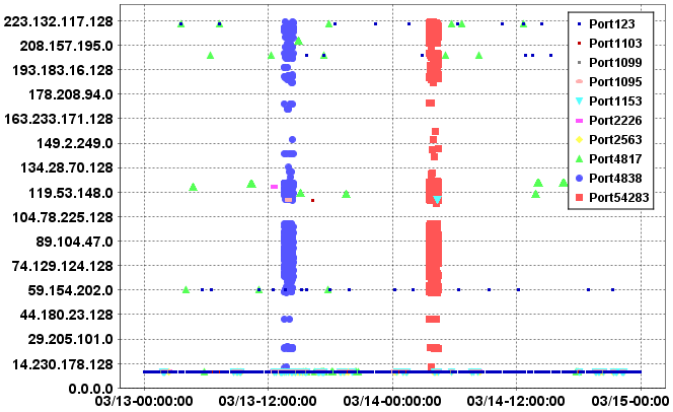

図7: UDPのパケットに注目したボットへの通信(2日分)

図7より, 13日14:00や14日4:00あたりの時間帯に大量のプロットが見られる. これより, port4838やport54283への局所的な時間で広域なIPアドレスから攻撃を観測できた. 特に, 14日4:00あたりに受けている攻撃はPort54283へのSubSevenという攻撃ツールからの攻撃であると考察できる.

6. 考察

攻撃通信データの中でも TCP-syn と UDP のパケットに注目した. TCP-syn のパケットを攻撃パケットと見なすことで, さまざまなパターンに分けて挙動を抽出することができた. ボットからの通信のプロット図では, 縦方向の筋を観測することができた. これはボットから外部端末への継続的な感染活動であると断定できる. ボット感染端末への通信のプロット図では様々な国からアクセスが網羅的にきていることを見ることができた.また, UDP のパケットに注目し, 解析した結果, 横方向の筋を確認できた. これはボット端末の周りに攻撃したり, 特定の IP を継続的に攻撃していることがわかる.

以上より, TCP-syn と UDP のパケットによる解析結果を比較すると, TCP-syn パケットはアドレススキャン型の攻撃を観測し, UDP パケットは特定のアドレスに対する攻撃の傾向が多く観測することができた.

7. まとめ

本研究では, 研究データセット CCC DATASet2009 の攻撃通信データを時系列分析によって解析した. 攻撃通信データの中でも, TCP-syn のパケットと UDP のパケットに注目し, 時間軸と IP 軸に対し, 通信挙動をプロットし可視化することで, ボットに感染した端末の全体の挙動を検出した. これは, ボットの挙動を網羅的に検出するという点で有効なシステムであると考える.

今後の課題としては, このようなボットの特徴的な挙動を検知し, 挙動解析を自動化できるようなシステムの構築が考えられる.

参考文献

[1] 高橋正和，村上純一，須藤年章，平原伸昭，佐々木 良一，”フィールド調査によるボットネットの挙動解析”, 情報処理学会論文誌，vol.47 No.8 p2512-p2523 Aug. 2006.

[2] 警察庁セキュリティポータルサイト@police, http://www.cyberpolice.go.jp/detect/observation.html

[3] Kiran Lakkaraju, William Yurcik, and Adam J Lee, NVisionIP: Netflow Visualizations of System State for Security Situational Awareness, ACM SIGSAC, pp65-72, 2004

[4] 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）

[5] サイバークリーンセンター（CCC), 2009年04月度, サイバークリーンセンター活動実績 https://www.ccc.go.jp/report/200904/0904monthly.html