DNS 通信の挙動からみたボット感染検知方式の検討

東角　芳樹、鳥居　悟
株式会社富士通研究所　ソフトウェア＆ソリューション研究所

あらまし：ハニーポットが行う DNS 通信に着目し，研究用データセット CCC DATAset 2008 の攻撃通信データを解析した．ボットに感染し活動にいたる過程で発行される DNS クエリの特徴を整理し，ネットワーク監視によるボット感染検知の実現可能性について考察した．その結果，さまざまなホストの名前や内部のホストの名前の解決を行なっているなど,従来にはない挙動が抽出できた．これらは，感染初期に固有の通信挙動であり，また，実行コードの解析では得られない，という点で有効な解析結果であると考える．
キーワード：**ボット**，**ハニーポット**，**ログ解析**，**感染検知**

Study for Detecting Bot-Infected PC based on behavior of DNS query

Yoshiki Higashikado, Satoru Torii
Software and Solution Laboratories, Fujitsu Laboratories LTD.

Abstract : We analyzed the attack communication data of CCC DATAset 2008, focusing to DNS query which the honeypot sends to. In this paper, we clarify the feature of the DNS query from infection of bot to activity.　The feasibility study of the detection method was also presented. As a result, unknown activity of bot has been extracted, such as solving various hosts' name and an internal hosts' name. We think that those feature are valuable, as they means the communication of early stage of infection and they cannot be got by the code analysis.
Keyword : Bot, honeypot, log analysis, infection detection

1. はじめに

近年のマルウェア特にボットは，環境に応じて動作を変更するなど巧妙になってきている．感染を検出するためにはその挙動を的確に把握することが必要であり，いろいろな環境における動作事例の解析結果を広く共有することが求められる．一方で，公表されている解析結果は一部の組織における観測に基づくものであり，感染 PC が多いといわれている日本国内 ISP 環境における観測結果はほとんど発表されていない．

そこで，共通素材としての研究用データセット CCC DATAset 2008 の攻撃通信データ（以降，CCC2008 攻撃通信データ）を用いて，ボット感染時の DNS クエリの挙動を整理し，これまでに報告された解析結果との比較評価を行なう．

CCC2008 攻撃通信データは日本国内 ISP に設置されたハニーポットの観測結果であり，ボット感染の初期フェーズの通信挙動が観測されている．既存手法の比較評価に活用できると共に，ボット感染を早期に検知するための手法を選定することに役立つものと考えられる．さらには，研究用標準データとして活用できることが期待できるものと考えている．

以降では，解析対象である CCC2008 攻撃通信データの概要を説明し，シーケンシャルマルウェアの動作に関連した特徴的な解

析結果を示す．さらに，DNS クエリに着目した既存研究成果を紹介し，CCC2008 攻撃通信データを対象に同様の傾向の有無を解析した結果を述べる．

2. 研究用データセット CCC DATAset 2008

研究用データセット CCC DATAset 2008 は，サイバークリーンセンター(https://www.ccc.go.jp/)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．

ここでは，その概要と基礎データとなる解析情報について述べる．

2.1. CCC2008 攻撃通信データの概要

本研究で使用した CCC2008 攻撃通信データは，ハニーポット 2 台への通信を 2 日分フルキャプチャしたデータである．この 2 台のハニーポットの OS は，それぞれ Windows XP*と Windows 2000*である．これらハニーポットは，日本国内 ISP のインターネット回線に接続されており，定期的にクリーンな状態にリセットされている．

ハニーポットは単にネットワークに接続されているだけであり，その上ではメールの送受信や Web へのアクセスといった操作をしておらず，そのため受動的攻撃の被害のような傾向は得られていないようである．

2.2. 基本統計

取得されているパケットはすべて IP パケットである．その内訳を表 1 に示す．

表 1：取得されているパケットの内訳

IPパケットの種類	パケット数
ICMP	145,469
UDP	3,038,494
TCP	12,717,980
合計	15,901,943

* Windows は米国 Microsoft Corporation の米国およびその他の国における登録商標です。

ここで，ループバックパケットは除外している．以降の解析結果でも同様にループバックパケットを除外する．

3. セッションごとの通信の挙動

観測対象のハニーポットは，定期的にクリーンな状態にリセットされている．リセットされてから次にリセットされるまでの間をひとつのセッションと定義する．これにより，それぞれのセッション内の通信挙動を解析することで，攻撃からの一連の活動の流れを把握することができると考える．

ここでは，近年話題になっているシーケンシャルマルウェアの概要を述べると共に，CCC2008 攻撃通信データを解析した結果のなかでシーケンシャルマルウェアの動作に関連した特徴的な解析結果を紹介する．

3.1. シーケンシャルマルウェア

インターネット上のサーバからプログラム等をダウンロードする「ダウンローダ」を介した感染事例が報告されている[1][2]．このようなマルウェアはシーケンシャルマルウェアもしくは多段攻撃(multistaged attacks)と呼ばれている．

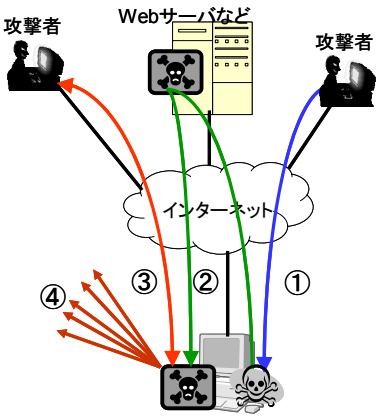


図 1：シーケンシャルマルウェアの動作

この動作を図 1 に示す．今回の CCC2008 攻撃通信データでの観測結果を踏まえると，感染に至る流れは以下の 4 つの手順に分類できる．

① 攻撃/侵入：主にポート番号 135 の RPC (Remote Procedure Call)を使ってダウンローダが送り込まれる．

② 不正プログラムのダウンロード：TFTP やHTTPを使って不正プログラムをダウンロードする．
③ 攻撃者の制御通信：IRC メッセージのやり取りが行なわれ，動作を制御される．使用されるポート番号はIRCの割り当て番号とは異なる場合がある．
④ 踏み台の拡大/攻撃：SPAM 送信や DoS 攻撃，他 PC の探索などが行なわれる．

3.2. 特徴的な挙動を示すセッション

CCC2008 攻撃通信データを解析し，このような攻撃，ダウンロード，活動に至る挙動が存在するかどうかを整理した．すべてのセッションの通信を解析した結果のなかから，特徴的な挙動示すものを図 2 に示す．

ここで，グラフの横軸は時間経過を示し，縦軸は通信量を示す．通信方向が分かるように，上段が外部からの通信(intrusion)を示し，下段が内部のハニーポットからの通信(extrusion)を示している．また，通信内容の区別がつくように色分けされている．

(1) 感染に至らなかった（図 2(1)）

何回かの攻撃通信が到達しているものの，感染には至らなかったケースである．このほか，まったく攻撃通信が観測されなかったセッションも見受けられた．

(2) 本体の入手に至らなかった（図 2(2)）

攻撃が成功し，ダウンローダが稼動している．しかし，ダウンロード先のサイトから応答がなく再送を繰り返している．このように，ダウンロード先サイトからの応答がないケースも見受けられた．

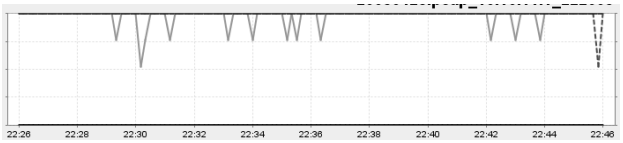
(3) ひと通りの動作が確認された（図 2(3)）

攻撃から活動に至る流れがひと通り観測されたケースである．ここで観測された活動は，近隣サイトへのスキャン行為であった．

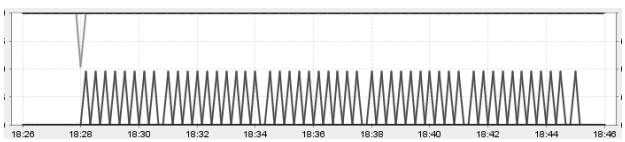
(4) 複数の感染が確認された（図 2(4)）

すでに感染し活動が行なわれている途中で，さらに他のマルウェアに感染し活動がおこなわれたケースである．

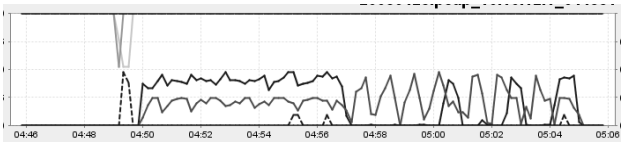
(1) 感染に至らなかったケース



(2) 本体の入手に至らなかったケース



(3) ひと通りの動作が確認されたケース



(4) 複数の感染が確認されたケース

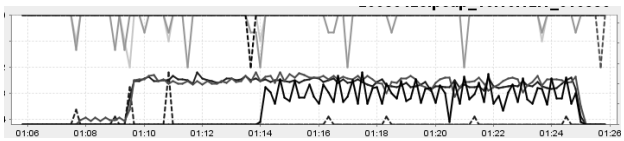


図 2：特徴的な挙動を示すセッション例

4. DNS クエリに着目した既存の研究

シーケンシャルマルウェアのようにダウンローダやボットが活動すると，その本体を入手するサイトや C&C サーバとの通信が発生する．これらの宛先は FQDN で定義されているため，DNS への問合せ（クエリ）が発生する．この問合せに着目した研究がすでにいくつか行なわれている．

ここでは，これら既存の研究成果を整理する．主に，実環境における調査と，マルウェアを解析した調査の二種類に分類できる．

4.1. 実環境における調査

Schonewille ら[3]は，サーバ側とネットワークの両方でモニタしたクエリを解析した結果を報告している．調査対象は，オランダの国立学術研究コンピュータネットワークである SURFnet である．この調査では，以下の 4 つの切り口が有効と結論づけている．

・ブラックリストに載っているドメインの問合せ
・問合せ数ランキングの上位(Top 10)

・リゾルバの逸脱
・普通でない qtype(MX,AXFR, IXFR)

武藏ら[4]は，大学内 DNS サーバへの大量の DNS アクセスを調査した．その結果，ボットに乗っ取られていると思われる PC 端末からの正引きクエリには，IP アドレスがキーワードとして直接記述されていることを見出した．

4.2. マルウェア解析に基づく調査

高橋ら[5]は，ハニーポットで収集したマルウェアを静的解析して入手した FQDN のリストを作成し，各 FQDN を解析した結果を報告している．この調査では，マルウェアが使用する FQDN には，Dynamic DNS や DNS Service を利用しているもの，名前解決ができないもの，正引きと逆引きとの結果が異なるなどの特徴があると述べられている．

朝長ら[6]は，ハニーポットで収集したマルウェアを動的静的に解析して FQDN ブラックリストを作成し，DNS サーバへの正引きと比較する方式を提案している．

5. CCC2008 攻撃通信データにおける DNS クエリ

これら既存の研究成果を踏まえつつ，CCC 2008 攻撃通信データに対しても同様の傾向の有無を解析することは，共通認識を醸成するとともに，技術の有効性や効果を客観的に確認できると期待する．

5.1. 解析にあたって

さまざまな環境におけるボットの挙動を整理することは，適切なボット対策を導出するうえで重要と考える．特に，ボットに感染されている PC は，その大部分が ISP 環境内に接続されているものであるといわれている．すなわち，ISP 環境内における動作を明らかにしておく必要がある．

また，ボットの挙動は，C&C サーバによる制御など，実行コードに記載されていない部分の影響をも受ける．すなわち，外部との通信が行われているハニーポット上での挙動を解析することが必要と考える．

なお，今回の調査では，CCC2008 攻撃通信データから抽出されたホスト名に対しては，直接アクセスするなどの行為は行っておらず，正引きと逆引きの結果の照合や，Dynamic DNS や DNS Service を利用しているかどうかの調査は行っていない．

5.2. DNS クエリの内容

CCC2008 攻撃通信データに含まれていた DNS 通信のうち，ハニーポット 2 台から発行された全リクエストを集計した．その結果を，表 2 に示す．

Windows 2000 のハニーポットでは，武藏らが指摘した，IP アドレスを指定した正引きが多数行われていることが確認できた．一方で，Windows XP のハニーポットではこのような挙動は見られなかった．また，Windows XP のハニーポットでは，Schonewille らが指摘した，リゾルバの逸脱が確認できた．一方で，Windows 2000 のハニーポットではこのような挙動は見られなかった．これらの挙動の違いは，ボットが使用する WindowsAPI の動作が OS により異なることが原因ではないかと類推する．

表 2：DNS の全リクエスト

Windows XP		
	正引き(A)	3,320
	DNSサーバ指定(NS)	75
	逆引き(PTR)	30
	起動時のtime.windows.comの問合せ	172
	ホスト名にIPアドレスを指定(A:正引き)	0
	小計	3,597
Windows 2000		
	正引き(A)	1,425
	DNSサーバ指定(NS)	0
	逆引き(PTR)	504
	起動時のtime.windows.comの問合せ	0
	ホスト名にIPアドレスを指定(A:正引き)	15,018,349
	小計	15,020,278

一方，普通でない qtype(MX, ,AXFR, IXFR)は確認されなかった．このような挙動は，SPAM 送信すなわち大量のメールを送信するため，送信先アドレスを入手するために多数のアドレス問い合わせを行う故の挙動と考えられる．今回のような感染初期の挙動には現れないのであろうと類推する．言い換えると，このような挙動を示す PC は，すでに制御下におかれており感染後長い時間が

経過しているものであるといえる.

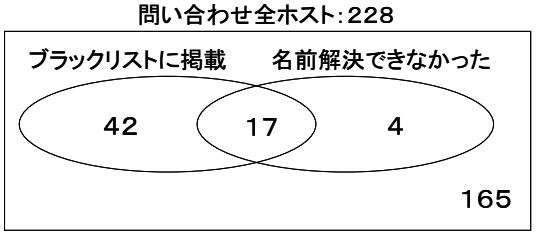
5.3. 正引きに記載されていたホスト名

正引きの引数に指定されていたホスト名を抽出し, それらがブラックリストに掲載されているかどうか, CCC2008 攻撃通信データ中の当該通信の応答が得られていたかどうかを整理した. 結果を図 3 に示す.

ここで使用したブラックリストは, Snort.conf Samples Project [7] の config-hosts(2008/07/14 版)を使用した.

ここから, ハニーポットが名前解決を要求したホスト名の全種類から比較すると, ブラックリストに掲載されたもの, および, 名前解決できなかったものの割合いは, 必ずしも多くないことがわかる. 言い換えると, ハニーポットからはさまざまなホストに対しての名前解決の要求が発行されているといえる.

(1) Windows XP



(2) Windows 2000

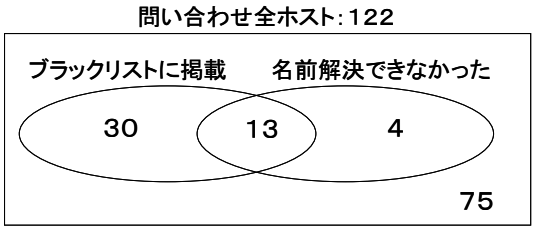


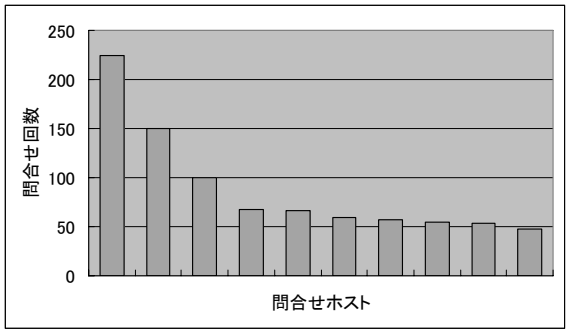
図 3 : 正引きされていたホスト名の分類

5.4. 問い合わせ頻度

正引きの引数に指定されていたホスト名ごとに, それぞれ問い合わせ回数を集計した. 図 4 に, 上位 10 ホストの問い合わせ回数を示す.

ハニーポットが定期的にクリーンな状態にリセットされているとはいえ, 高い頻度での問い合わせが発行されているといえる. ここから, Schonewille らの調査結果である「問合せ数ランキングの上位(Top 10)」の切り口が有効であると考えられる.

(1) Windows XP



(2) Windows 2000

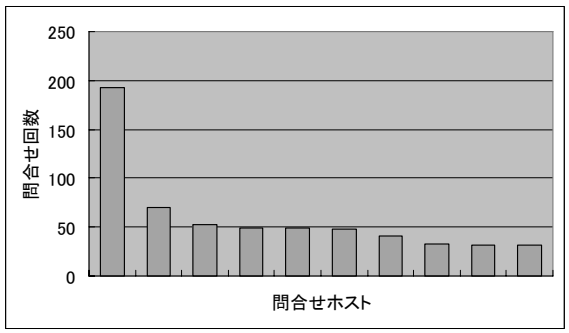


図 4 : 上位 10 ホストの問い合わせ回数

5.5. 問い合わせ頻度の高いホスト名

正引きの引数に指定されていた回数の多いホスト名をそれぞれ上位 50 件（同率を含め, Windows XP では 55 ホスト, Windows 2000 では 52 ホスト)がどのようなものであるかを調査した. 表 3 に集計結果を示す.

表 3 : 問い合わせ頻度の高いホスト

Windows XP	
ブラックリストに掲載	18
ポータルサイト	14
メールサーバ	12
不明	10
その他	1
小計	55

Windows 2000	
ブラックリストに掲載	26
メールサーバ	10
ローカルISP	6
IRC	2
ポータルサイト	1
不明	7
小計	52

問い合わせ数の多いサイトのなかでは, ブラックリストに掲載されているものがトップに位置している. 一方で, 通常の使用においても発生するであろう, 一般的なポータルサイトやメールサーバへの問い合わせの頻度も高い.

また, ハニーポットが設置されている ISP 内のホスト名を名前解決しようとしている傾向が見られる. これはこれまでにない

傾向として，注目に値すると考えられる．

5.6. 検知アラートが挙がるセッション

CCC2008 攻撃通信データの中には，本体の入手に至らず感染の途中で終わっている事象や，一度に複数のマルウェアに感染している事象などが含まれており，これらの中から感染事象だけを適切に抽出することは困難である．

そこで，ハニーポットのリセット間隔ごとのセッションを 1 セットとして，定量的に整理するための母数とする．すなわち，このセッション中にひとつ以上の DNS クエリが発行されたセッションを「感染された」とする．また，セッション中にこれまでに示した特徴に該当する通信がひとつでも見受けられたら「検知した」とする．

セッション単位での特徴の検出状況の結果を表 4 に示す．

表 4：特徴が現れたセッションの数

	Windows XP		Windows 2000	
	検知数	検知率	検知数	検知率
ブラックリスト	90	0.96	85	0.93
リゾルバの逸脱	25	0.27	0	0.00
普通でないqtype	0	0.00	0	0.00
IPアドレスの正引き	0	0.00	78	0.86

ここから，一連の行動が含まれるセッションに対して，ブラックリストを用いることで感染検知のアラートを挙げることが有効と考えられる．また，Windows 2000 を対象とした場合，IP アドレスで正引きを行なう行為を契機とすることが有効である．

ここで，検知率は検知したセッション数を感染されたセッション数で除した値で求められる．但し，感染していないものを感染したと誤って検出する false positive は評価できていない．

6. まとめ

ボットに感染し活動にいたる過程で，ハニーポットが発行する DNS クエリの挙動を整理した．これは，感染初期の通信挙動であり，実行コードの解析だけでは導出できなかったであろう，という点で有効な解析結果であると考える．

特に，日本国内 ISP 環境における挙動の解析結果として，これまでに報告されていた特徴との比較確認ができたとともに，さまざまなホストの名前や内部のホストの名前の解決を行なっているなど，従来にはない挙動が確認できた．この結果は，ボット感染 PC が発行する通信を識別し，ボット感染を早期に検知するための手法の選定に役立つと考える．

いろいろな環境における動作事例の解析結果を広く共有することは，適切なボット対策を行なうために必要なことである．今回の結果がその一部を担うことができれば幸いである．また，今後も継続的に，このような共通的な研究用データセットが提供されることを期待するところである．

参考文献

[1] “近年の標的型攻撃に関する調査研究”, 独立行政法人情報処理推進機構(IPA)

[2] Symantec Internet Security Threat Report Volume XII: September, 2007

[3] Antoine Schonewille, Dirk-Jan van Helmond, “The.Domain Name Service as an IDS. How DNS can be used for detecting and monitoring badware in a network”, 2006, http://staff.science.uva.nl/~delaat/snb-2005-2006/p12/report.pdf

[4] 武藏泰雄，松葉龍一，杉谷賢一，”プロトコル異常検知による A レコード型 DNS パケット分散サービス妨害攻撃の阻止,” 情報処理学会研究報告，IPSJ SIG Notes 2005-DSM-38-(5)

[5] 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，Aug.2006.

[6] 朝長秀誠，田中英彦, “Botnet の命令サーバドメインネームを用いた Bot 感染検出方法に関する研究,” 情報処理学会研究報告, IPSJ SIG Notes 2006-CSEC-35-(3)

[7] Snort.conf Samples Project http://www.bleedingsnort.com