Vol.2013-CSEC-61 No.7
Vol.2013-IOT-21 No.7
2013/5/9

マルウェア感染ホストの特定を目的とした DNS 通信の可視化

牧田大佑† 吉岡克成† 松本勉†

我々は，DNS キャッシュサーバに集まる通信に着目して，ネットワーク内のマルウェア感染ホストを特定する手法について検討している．DNS 通信を解析する場合，観測される通信から多数のホストの名前解決動作の同期性や名前解決されるドメイン名，対応する IP アドレスなどを把握することは難しい．そこで，ホスト群の特徴的な DNS 通信を把握するための可視化手法を提案し，実トラフィックを用いた実験により提案手法の有効性を検証する．

How to Visualize DNS Traffic to Detect Malware-Infected Hosts

DAISUKE MAKITA† KATSUNARI YOSHIOKA†
TSUTOMU MATSUMOTO†

We are developing a method for detecting malware-infected hosts from the traffic of the DNS cache server. When analyzing large DNS traffic, it is difficult to quickly comprehend its contents, such as synchronizations of name resolutions, resolved domain names and corresponding IP addresses. Therefore, we propose a method for visualizing the DNS traffic, and evaluate its effectiveness by experiments using the real DNS traffic.

1. はじめに

情報漏洩やスパムメール送信，DoS（Denial of Service）攻撃を行うマルウェアがインターネット上の大きな脅威となっている．その中でも，ボットと呼ばれる種類のマルウェアは，C&C（Command & Control）サーバから攻撃者の命令を受け取り，それを実行することによって大規模なサイバー攻撃を引き起こすことが知られている．これらの脅威に対抗するため，インターネット上に設置された各種センサから得られるデータを分析することにより，インターネット上のマルウェア感染ホストやその規模を推定する研究が進められている．

我々は，DNS キャッシュサーバに集まる通信に着目してネットワーク内のマルウェア感染ホストを特定する手法について検討している．インターネット上で通信を行うマルウェアの多くは，DNS（Domain Name System）を用いて接続先のサーバとの通信を試みる．そのため，DNS キャッシュサーバの通信を分析することは，ネットワーク内のマルウェア感染ホストを特定する上で有用な情報となる．

DNS キャッシュサーバに集まる通信からマルウェア感染ホストを特定する手法として，マルウェアが使用するドメイン名をブラックリストに登録し，これを用いて検知する手法（ブラックリスト方式）が検討されている．しかし，近年，解析対象となるマルウェアの種類が急増していることに加え，攻撃者はブラックリストの回避を目的として新しいドメイン名を利用する傾向にあり，ブラックリスト方式による検知は困難になりつつある．そこで，マルウェアの動的解析や DNS サーバで得られる DNS 通信を解析することにより，マルウェアに特徴的な名前解決動作を抽出し，その挙動からマルウェア感染ホストやマルウェアが悪用するドメイン（以下，悪性ドメイン）を特定する手法が検討されている[1, 2, 3, 4, 5]. しかし, DNS 通信を解析する場合，観測される通信のテキスト情報から多数のホストの名前解決動作の同期性や名前解決されるドメイン名，対応する IP アドレスなどを把握することは難しい．

本稿では, ホスト群の特徴的な DNS 通信を把握するための可視化手法を提案する．提案手法では，DNS 通信の応答パケットに含まれる，(1)ユーザの IP アドレス，(2)問い合わされたドメイン名，(3)問い合わせに対する応答の 3 つのパラメータを，Query-Response ビュー（QR ビュー）と Time-Series ビュー（TS ビュー）の 2 つのビューを用いて可視化する．QR ビューでは，ある期間の(1)〜(3)の値をそれぞれ平行軸上にプロットし，それらの対応を線で結ぶことにより，その期間の DNS 通信を描画する．一方，TS ビューでは，横軸を時刻，(1)〜(3)を縦軸とする 3 つのグラフを用意し, それぞれに(1)〜(3)に関する期間ごとの統計値を色合いで書き込むことにより, 時系列順に DNS 通信を描画する．そして，この 2 つのビューを時系列順に更新することにより，DNS 通信を動画として表現し，ホスト群の特徴的な DNS 通信の把握を支援する．

本稿の構成は次の通りである．2 章でマルウェアに特徴的な DNS 通信と，DNS 通信の可視化に関する先行研究を取り上げる．3 章で DNS 通信の可視化手法を概説し，4 章で実運用中の DNS キャッシュサーバのトラフィックを用いた可視化とその解析例を示す．そして，5 章でまとめと今後の課題とする．

2. 関連研究

本章では，マルウェアに特徴的な DNS 通信と，DNS 通信の可視化に関する関連研究を取り上げる．

†横浜国立大学
Yokohama National University

2.1 マルウェアに特徴的な DNS 通信

文献[1, 2, 3]では, DNS サーバに集まるトラフィックから悪性ドメインを判定する手法が提案されている．これらの手法は，数週間から数ヶ月のトラフィックを分析することにより，正規のドメインと悪性ドメインで異なる特徴を抽出し，その特徴量からドメイン名の悪性を評価している．

文献[4]では，DNS サーバに集まるトラフィックから，DGA（Domain Generation Algorithm）によって生成されたと推測されるドメイン名を検知する手法が提案されている．DGA とは，接続先のドメイン名を生成する仕組みであり，一部のマルウェアはブラックリストによる検知回避を目的として DGA を使用していると考えられている．文献[4]では，DGA で生成されるドメイン名には応答が存在しない（NXDomain 応答となる）場合が多いという特徴や，生成されるドメイン名の文字列に類似性が存在することに着目している．

我々の先行研究[5]においても，DGA 等の仕組みを用いて多量のドメイン名を名前解決するマルウェア感染ホストを，ホストが名前解決するドメイン集合の類似性に着目して検知する手法を提案した．また文献[6]では，マルウェアを長期間にわたって動的解析したり，同一マルウェアを短期間に複数回動的解析したりすることによって，そのマルウェアに特徴的な DNS 通信の抽出を試みている．

2.2 DNS 通信の可視化

DNS 通信の可視化を試みた先行研究としては，文献[7, 8, 9]がある．

文献[7]では，DNSamp や DNS キャッシュポイズニングといった DNS サーバのセキュリティ問題への対応を目的としたネットワーク管理者向けの DNS 通信の可視化システムが提案されている．このシステムでは，管理者が異常を検出しやすいように，独自の可視化手法に加え，Stacking Graphs や Two Tone Pseudo Color などの複数の表現を利用することにより，DNS キャッシュサーバのログの可視化を試みている．

文献[8, 9]では，同じくネットワーク管理者向けの DNS 通信の可視化システムが提案されている. このシステムは，ネットワーク内のボットの検出を目的としており，ユーザの IP アドレス，問い合わされたドメイン名，問い合わせに対する応答の 3 つのパラメータ（文献[9]では TTL（Time To Live）値が追加されているため 4 つ）を平行軸上に可視化する．彼らのシステムは，問い合わせの頻度や周期性，異常性などから DNS クエリを評価し, 管理者が可視化結果からボットを把握しやすいような工夫を試みている．

3. DNS 通信の可視化

本章では, 我々が提案する DNS 通信の可視化手法について概説する．我々の提案手法の目的は，DNS キャッシュサーバに集まる通信からホスト群の特徴的な名前解決動作を把握することであり，2 つのビューを用いた動画として DNS 通信を表現することにより解析者を支援する（図 1）.

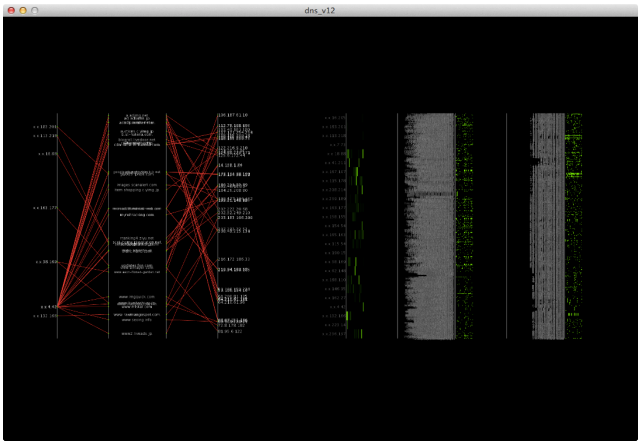


図 1　提案手法による DNS 通信の可視化例

3.1 可視化するパラメータ

DNS はインターネット上でドメイン名と IP アドレスを対応付ける役割を果たしている．ドメイン名を用いてインターネット上のホストと通信を行う場合，まず接続先のドメイン名をネットワーク内の DNS キャッシュサーバに問い合わせ，対応する IP アドレスを取得する．このように，ドメイン名から対応する IP アドレスを取得することを名前解決という．

提案手法では，ドメイン名から IP アドレスを取得するときに使用される A レコード（Address Record）の応答パケットを可視化の対象とし，その応答の時刻情報と次の(1)〜(3)のパラメータを用いて DNS 通信の可視化を試みる．

(1) ユーザの IP アドレス

ドメイン名を問い合わせたユーザの IP アドレス. この IP アドレスは, 応答パケットの IP ヘッダに含まれている宛先アドレス（Destination Address）[10]を使用する.

(2) 問い合わされたドメイン名

ユーザから DNS キャッシュサーバに問い合わされたドメイン名．このドメイン名は，応答パケットに含まれている問い合わせ部（Question Section）に指定されている QNAME の値を使用する[11].

(3) 問い合わせに対する応答

問い合わされたドメイン名に対する応答．通常は，ドメイン名に対応する IP アドレスである. この IP アドレスは，応答パケットに含まれている回答部（Answer Section）に記述されているものを使用する [11]．対応する IP アドレスが複数存在した場合は全ての IP アドレスを抽出し, 名前解決においてエラーが発生した場合はそのエラーを表す文字列（応答が存在しない場合は NXDomain，サーバ障害が発生した場合は ServFail など）を抽出する.

DNS による名前解決と可視化するパラメータ(1)〜(3)の例を図 2 に示す．この図では，ユーザ（クライアント）がドメイン名 www.ynu.ac.jp.を DNS キャッシュサーバに問い合わせ（①），対応する IP アドレス 133.34.27.181 を受け取り（②），Web サーバと通信を行う（③）例と，可視化に使用するパラメータの例を示している．

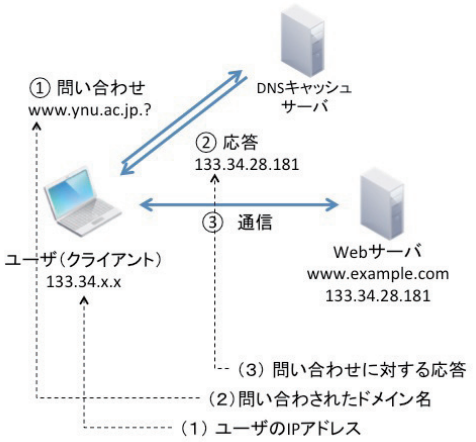


図 2　名前解決と可視化するパラメータの例

3.2 可視化手法

提案手法では，3.1 節で説明したパラメータに関する情報を，次の 2 つのビューに描画する．

- Query-Response ビュー（QR ビュー）：ある期間の問い合わせや応答を描画する
- Time-Series ビュー（TS ビュー）：時系列順に問い合わせや応答に関する統計値を描画する

この 2 つのビューを時系列順に更新していくことにより，動画として DNS 通信を可視化する（図 3）．なお，期間 t は秒，分，時間単位のいずれかに変更を可能にする．

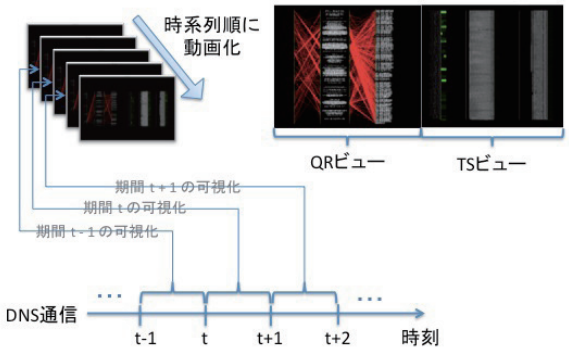


図 3　提案手法の概要

以下，QR ビューと TS ビューの描画方法を概説する．

3.2.1 Query-Response ビュー（QR ビュー）

Query-Response ビュー（QR ビュー）では，ある期間 t の DNS 通信を 4 本の平行軸によって描画する．具体的には，一番左の軸に(1)のユーザの IP アドレスを，中央の 2 本の軸に(2)の問い合わされたドメイン名を，一番右の軸に(3)の問い合わせに対する応答をそれぞれプロットし，それらの対応を線で結ぶことによって，その期間の DNS 通信を描画する．ただし，中央の 2 軸の間には余白を設け，問い合わされたドメイン名をラベルとして描画できるようにした．

QR ビューの描画例を図 4 に示す．

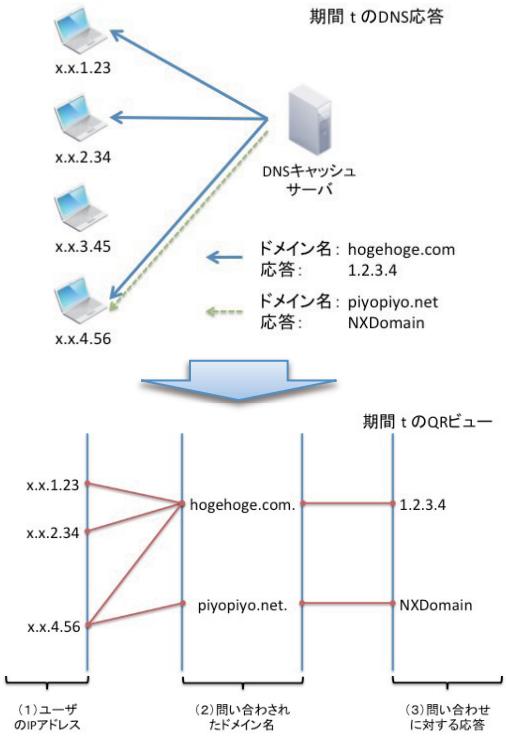


図 4　QR ビューの描画例

各値の配置は，通信における出現順とアルファベット順の二通りで描画できるようにし，各値をプロットする座標は予め一意に決定しておく．実装において，各値の座標は各軸をそれぞれの要素で等分するように調節した．

3.2.2 Time-Series ビュー（TS ビュー）

Time-Series ビュー（TS ビュー）では，期間 t までの DNS 通信に関する統計値をグラフ上に描画する．具体的には，横軸を時刻，縦軸を可視化するパラメータの(1)〜(3)とする 3 つの独立したグラフを用意し，期間 t までのそれぞれのパラメータに関する統計値を色合いで表現する．

グラフ上に色合いで表現される統計値を以下に示す．

- 各ホストの問い合わせ数（左）：各期間における各ホストの問い合わせ数を色合いで表現する．その期間にそのホストが多くの問い合わせをしているほど，対応するグラフ上に明るい色で描画する
- 各ドメイン名を問い合わせたホスト数（中）：各期間において各ドメイン名を問い合わせたホスト数を色合いで表現する．その期間にそのドメインが多くのホストから問い合わされているほど，対応するグラフ上に明るい色で描画する
- 各応答を取得したホスト数（右）：各期間において各応答を取得したホスト数を色合いで表現する．その

期間にその応答を取得するホスト数が多いほど，グラフ上に明るい色で描画する

TS ビューの描画例を図 5 に示す．

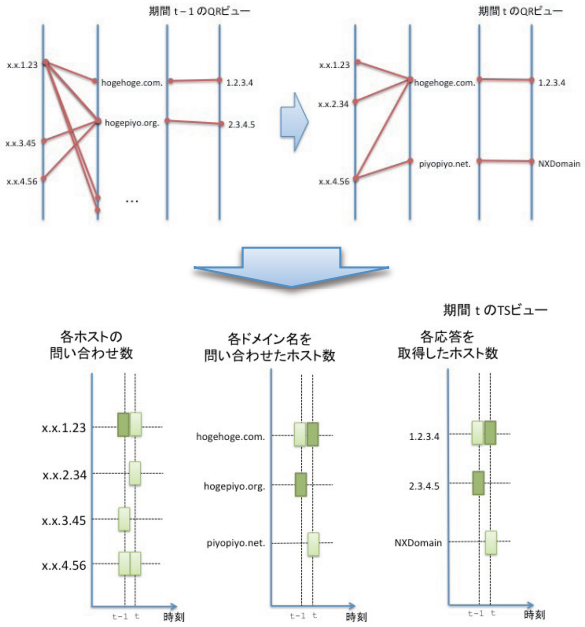


図 5　Time-Series ビューの描画例

各値の配置は，QR ビューと同様，通信における出現順とアルファベット順の二通りで描画できるようにし，各値をプロットする座標も予め一意に決定する．また，TS ビューのグラフは，QR ビューに同期して書き込まれるようにし，色合いは矩形領域としてグラフに描画される．実装において，各値の座標は QR ビューと同様に各要素で各軸を等分するように調節し，矩形領域の幅は固定長（今回の実装では 2pixel），高さは縦軸を全要素で等分するような値とした．

3.3 可視化の前処理

DNS キャッシュサーバにはネットワーク内のホストの問い合わせが集まるため，そのデータは非常に大容量になる．実際に，3.2 節の提案手法で DNS 通信を描画する場合，1 ピクセルに 1 つのデータを書き込んだとしても，各軸で高々数百程度のデータしか描画することができない．そこで，予め分析する対象を選定し，可視化するデータを抽出した上で可視化を行う．

具体的な前処理としては，

- 既知の悪性ドメインに関する DNS 通信を抽出する
- 既知の C&C サーバのドメイン名を問い合わせていたホスト群の DNS 通信を抽出する

などが考えられる．

3.4 可視化システムの実装

提案手法の実装にあたり，pcap 形式の通信データから可視化するパラメータを抽出するプログラムをスクリプト言語の Python[12]，及び，そのライブラリである dpkt[13]を用いて実装した．また，3.3 節の前処理には Unix 系 OS に標準のコマンドを利用し，3.2 節で概説した可視化システムは Casey Reas，Benjamin Fry らを中心に開発されたプログラミング言語 Processing[14]を用いて実装した．

4. 提案手法を用いた可視化とその解析例

本章では，3 章で概説した提案手法を用いて，マルウェアに関連する DNS 通信を解析した例を示す．なお，本章に記載する可視化結果は文献[18]にて公開している．

4.1 解析対象の DNS トラフィック

本稿では，実運用中の DNS キャッシュサーバのトラフィックを解析の対象とした．解析した DNS トラフィックの概要を表 1 に示す．なお，本研究では解析段階において，個々のユーザを特定できる情報は解析していない．

表 1　解析対象の DNS トラフィック

日時	2012 年 6 月平日夜（65 分間）
ホスト数	100 万～200 万
問い合わせ数	1 億～2 億
ドメイン数	300 万～400 万

4.2 解析対象のマルウェア

本稿では，前節の DNS 通信から実際のマルウェアに関係する DNS 通信を抽出し，提案手法を用いて解析を行った．表 2 に本稿で解析したマルウェアの種類と，前処理で用いた DNS 通信の抽出条件の概要を示す．

表 2　解析したマルウェアの種類と DNS 通信の抽出条件

マルウェアの種類	DNS 通信の抽出条件
Virut.g	proxim.ntkrlpha.info.を問い合わせたホスト群の DNS 通信[16]
ZeuS 系①	mail7.digitalwaves.co.nz.と mxs.mail.ru.の両方を問い合わせたホスト群の DNS 通信[6]
ZeuS 系②	ZeuS Tracker[15]に公開されているドメイン名を問い合わせたホスト群の DNS 通信
Conficker.B	[5, 17]で検出されたホスト群の DNS 通信

4.3 DNS 通信の可視化と解析結果

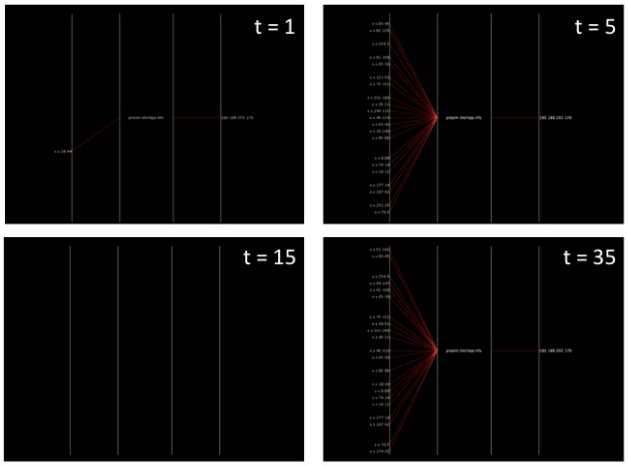
本節では，今回の解析において確認されたマルウェアに特徴的と推測される名前解決動作を 3 つ取り上げる．なお，以降で可視化例として示す静止画像は，図を簡潔にするため，4.2 節に記載した条件から更に対象を絞って可視化している．

4.3.1 名前解決の同期性・周期性

McAfee のセキュリティレポート[16]によると，Win32/Virut.g は proxim.ntkrnlpa.info.にある IRC サーバへの接続を試みる．そこで，このドメイン名を問い合わせているホスト群の DNS 通信を抽出し，提案手法で可視化した．その結果の一部を図 6 に示す．この図より，Virut.g のホスト群は，ある期間（t＝5，t＝35）に同期してそのドメイン名を名前解決していることがわかる．同様の名前解決の同期性は，文献[6]に記載されている mail7.digitalwaves.co.nz.

や mxs.mail.ru.（ZeuS 系①）でも確認され（図 7），mail7.digitalwaves.co.nz.に関しては，周期的に名前解決されている様子も確認することができた．

更に，ZeuS Tracker[15]で公開されているブラックリストで検出されたホスト群（ZeuS 系②）の名前解決動作にも同期性を確認することができた．しかし，先述した Virut.g と ZeuS 系①は分単位の同期性を示したのに対し，ZeuS 系②は秒単位で同期して iqservice.ir.を問い合わせていた（図 8）．



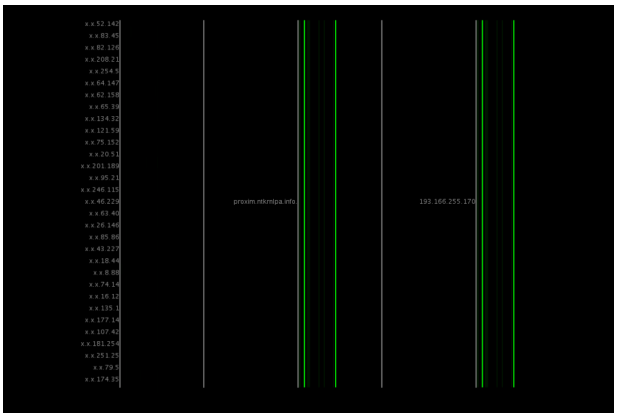


図 6　Virut.g, proxim.ntkrnlpa.info.の可視化例（QR ビュー，TS ビュー）：多数のホストが，ある期間（t=5, t=15）に同期して proxim.ntkrnlpa.info.を問い合わせている．期間は分単位．

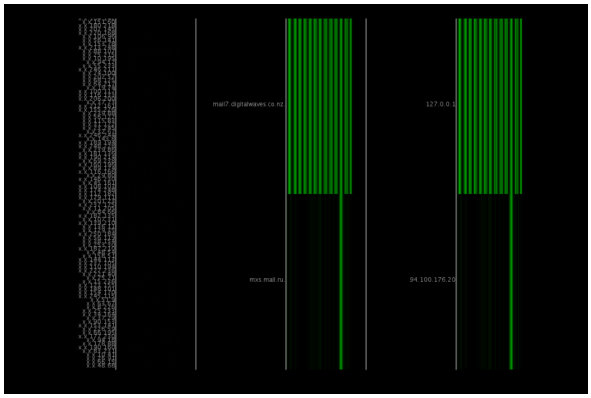


図 7　ZeuS 系①の可視化例（TS ビュー）：mail7.digitalwaves.co.nz.（中央のグラフ上方）は周期的に多数のホストに名前解決されており，mxs.mail.ru.（中央のグラフ下方）はある期間に多数のホストから同期して名前解決されている．期間は分単位．

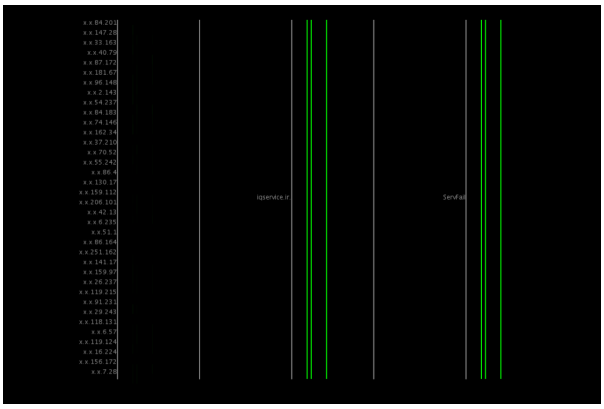


図 8　ZeuS 系②，iqservice.ir.の可視化例（TS ビュー）：多数のホストが，ある期間に同期して iqservice.ir.を問い合わせている．期間は秒単位．

4.3.2 アルファベット順の名前解決

ZeuS 系②のホスト群を解析した結果を図 9 に示す．QR ビューの中央には，ドメイン名がアルファベット順に配置されているため，このホストはアルファベット順に大量のドメイン名を名前解決していたことわかる．このような挙動を示したホストは複数存在していた．これはマルウェアがアルファベット順に並べられたドメイン名のリストを内部に保持し，そのリストに従って名前解決していたためと推測される．

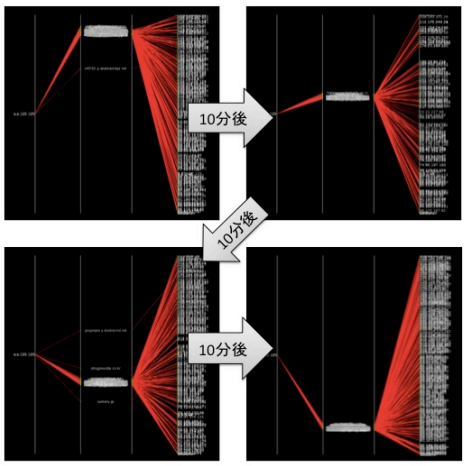


図 9　ZeuS 系②の可視化例（QR ビュー）：このホストは，アルファベット順にドメイン名を名前解決している．

4.3.3 短期間の大量の名前解決

Conficker.B への感染が推測されるホスト群の DNS 通信を可視化した結果の一部を図 10 に示す．この図の TS ビューの問い合わせ数を表すグラフから，各ホストが短期間に大量の名前解決を行っていたことが確認できる．

その期間に名前解決されたドメイン名から，それらは DGA で生成されたものであることが推測されたため，その

ドメイン名に関する DNS 通信を再度抽出し，提案手法で可視化した（図 11）．その結果，DGA で生成されたと推測されるドメイン名は多数存在したが，対応している IP アドレスは 5 つに限定されていた．なお，IP アドレスを逆引きした結果，これらはボットネットのシンクホールとして利用されているものと推測される．

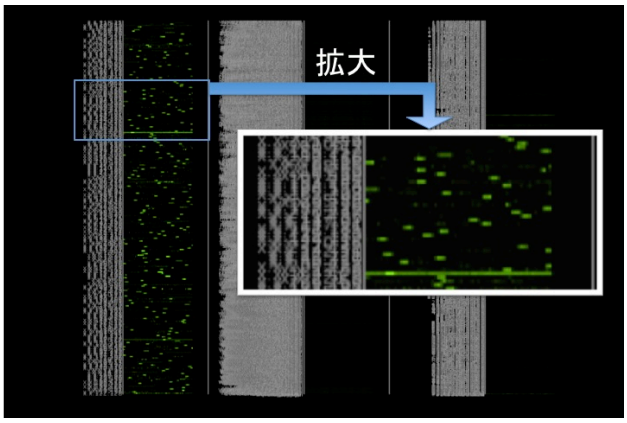


図 10　Conficker.B の可視化例（TS ビュー）：各ホストの問い合わせ数を表すグラフより，これらのホスト群は短期間に大量の名前解決を行っていることが確認できる．期間は分単位．

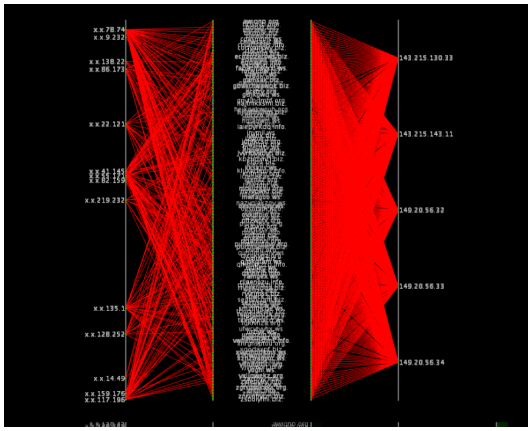


図 11　Conficker.B，DGA で生成されたドメイン名の可視化例（QR ビュー）：多数のドメイン名が名前解決されているが，対応する IP アドレスは 5 つのみであった．期間は分単位．

5. まとめと今後の課題

本稿では，ホスト群の特徴的な DNS 通信を把握するための可視化手法を提案し，その手法を用いて，実運用中の DNS キャッシュサーバの通信に含まれていたマルウェアに関係する DNS 通信を解析した例を示した．その結果，マルウェアによる名前解決の同期性や周期性，アルファベット順の名前解決や短期間の大量の名前解決などのマルウェアに特徴的な名前解決動作を確認することができた．

今後は，解析対象のマルウェアを増やし，マルウェアに特徴的な名前解決動作を多く把握していくとともに，それらの挙動を DNS キャッシュサーバに集まる通信から検知する手法について検討する予定である．また，本提案手法を改良するとともに，DNS 通信を解析するためのユーザインタフェースについても検討していきたい．

謝辞　本研究の一部は，総務省情報通信分野における研究開発委託／国際連携によるサイバー攻撃の予知技術の研究開発／サイバー攻撃情報とマルウェア実体の突合分析技術／類似判定に関する研究開発により行われた．

参考文献

1) Leyla Bilge, Engin Kirda, Christopher Kruegel and Marco Balduzzi,“EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis,” NDSS, 2011.
2) Manos Antonakaks, Roberto Perdisci, David Dagon, Wenke Lee, and Nick Feamster,“Building a Dynamic Reputation System for DNS,” USENIX Security Symposium 2010.
3) Manos Antonakakis, Roberto Perdisci, Wenke Lee, Nikolaos Vasiloglou and David Dagon,“Detecting Malware Domains at the Upper DNS Hierarchy,” USENIX Security Symposium 2011.
4) Manos Antonakakis, Roberto Perdiscit, Yacin Nadji, Nikolaos Vasiloglou, Saeed Abu-Nimeh, Wenke Lee and David Dagon, “From Throw-Away Traffic to Bots: Detecting the Rise of DGA-Based Malware,” USENIX Security Symposium 2012.
5) 牧田大佑，Yin Minn PaPa，吉岡克成，松本勉，“名前解決動作の類似性の基づくマルウェア感染ホストの特定，” SCIS2013
6) 田辺瑠偉，鉄穎，水戸慎，牧田大佑，神薗雅紀，星澤裕二，吉岡克成，松本勉，“長期動的解析によるマルウェアの特徴的な DNS 通信の抽出，” Computer Security Symposium 2012.
7) Pin Ren, John Kristoff, Bruce Gooch, “Visualizing DNS Traffic,” VizSEC’06, 2006.
8) Inhwan Kim, Hyunsang Choi, Heejo Lee, “Botnet Visualization using DNS Traffic,” WISA 2008.
9) Inhwan Kim, Hyunsang Choi, Heejo Lee, “BotXrayer : Exposing Botnets by Visualizing DNS Traffic,” ICONI 2009.
10) J.Postel, “INTERNET PROTOCOL,” http://www.ietf.org/rfc/rfc791.txt,
11) P. Mockapetris, “DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION,” http://www.ietf.org/rfc/rfc1035.txt
12) Python, http://www.python.org/, last visited 2013/03/31
13) dpkt – python packet creation / parsing library, https://code.google.com/p/dpkt/, last visited 2013/03/31
14) Processing, http://processing.org/, last visited 2013/03/31
15) Zeus Tracker, https://zeustracker.abuse.ch/monitor.php, last visited 2013/04/02
16) W32/Virut.g | ウイルス情報 | マカフィー, http://www.mcafee.com/japan/security/virV2008.asp?v=W32/Virut.g, last visited 2013/04/02
17) ThreatExpert, http://www.threatexpert.com/report.aspx?md5=6b6ba315c9f8d83bdb08f2d16dddc1ce, last visited 2012/12/07
18) 横浜国立大学　情報・物理セキュリティ研究拠点, http://ipsr.ynu.ac.jp/dnsviz/index.html