DNS クエリログのクエリ数に着目した異常ホストの検出

渡辺 拳竜[1]　池部 実[2]　吉田 和幸[3]

概要：大分大学のキャッシュ DNS サーバのクエリログを約 9 ヶ月間分析した結果，一般ホストとは異なる挙動をする 4 つの異常ホストを観測した．クエリログから送信元 IP アドレスと問合せた FQDN をキーにして集計した．その結果，一般ホストに関しては 1 つの FQDN あたりのクエリ数が少なく，さまざまな FQDN についてクエリを送信していた．一方，異常ホストは他のホストが対象としない FQDN に大量のクエリを送信している．そのため異常ホストの場合，1 つの FQDN あたりのクエリ数は多くなる傾向にある．本論文では，上記の特徴から，1 日のクエリログから学内に存在する異常ホストを検出する手法について提案する．本提案手法を 1 日ごとのクエリログに適用し，分析した結果について報告する．

A Detection method for anomalous host based on the DNS query counts

Watanabe Kenryu[1]　Ikebe Minoru[2]　Yoshida Kazuyuki[3]

1. はじめに

インターネットの発達と普及に伴い，我々が日頃から利用している Web ページの閲覧や電子メールなどのようなサービスにはインターネットが不可欠な存在になっている．ユーザがインターネットを利用するためには IP アドレスが必要である．IP アドレスは単なる数字列であるため，ユーザにとってわかりにくい．そこで，IP アドレスとユーザにわかりやすいドメイン名との対応関係を管理する DNS(Domain Name System) は，インターネットの根幹を支える重要なプロトコルのひとつである．ユーザは，キャッシュ DNS サーバに対してドメイン名を問合せ，IP アドレスを取得することでインターネットを利用する．

キャッシュ DNS サーバを問合せるホストの中にはボットなどのマルウェアに感染したホストが存在する可能性がある．2013 年 6 月に学内においてボット感染ホストが 1 件報告された．このボット感染ホストは攻撃者との通信をするため，キャッシュ DNS サーバへ C & C(Command and Control) サーバの FQDN を問合せる挙動が見られた．ボット感染ホストの問合せと正常なホストの問合せに差異を発見することができれば，キャッシュ DNS サーバのクエリログを調査することで問合せをしたホストが正常かどうかを判別できるのではないかと考え，クエリログを分析した．

本論文では，学内のキャッシュ DNS サーバのクエリログから，DNS クエリ数を集計することにより，異常ホストを検出する手法について提案する．また，本提案手法をクエリログに適用し，分析した結果を報告する．

本論文の構成は，第 2 章にてキャッシュ DNS サーバを利用したマルウェア感染ホストや悪性ドメインの発見に関する関連研究について述べる．第 3 章にて一般ホストと異常ホストの分析結果について述べる．第 4 章にてキャッシュ DNS サーバのクエリログのクエリ数に着目した異常ホストの検出手法とその分析結果について述べる．第 5 章にてまとめと今後の課題について述べる．

2. 関連研究

キャッシュ DNS サーバを利用してネットワーク内の異常なクエリを送信したホストの発見や，ボットをはじめと

[1] 大分大学大学院工学研究科知能情報システム工学専攻
Course of Computer Science and Intelligent Systems, Graduate School of Engineering, Oita University
[2] 大分大学工学部知能情報システム工学科
Department of Computer Science and Intelligent Systems, Faculty of Engineering, Oita University
[3] 大分大学学術情報拠点情報基盤センター
Center for Academic Information and Library Services, Oita University

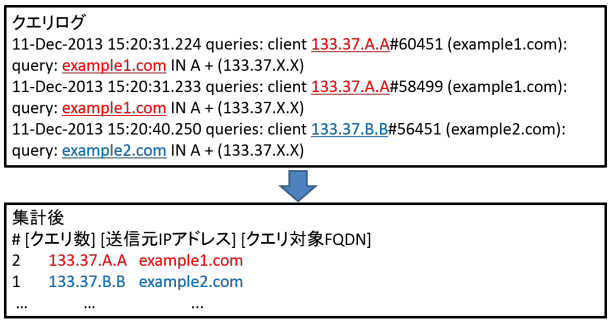

図 1 送信元 IP アドレスと FQDN の集計方法

したマルウェア感染ホストの発見を目的とした研究について述べる．

牧田ら [1] は，キャッシュ DNS サーバに集まるトラフィックを可視化することで，マルウェア感染ホストを特定する手法を提案している．牧田らの手法では，DNS 通信の応答パケットに含まれる，ユーザの IP アドレス，問合せられたドメイン名，問合せに対する応答の 3 つをパラメータとし，Query-Response ビューと Time-Series ビューの 2 つのビューを用いて可視化している．提案手法を用いて，運用中のキャッシュ DNS サーバの通信に含まれていたマルウェアに関する DNS 通信を解析した．その結果，マルウェアによる名前解決の同期性や周期性，アルファベット順での名前解決や短期間での大量な名前解決要求など通常ユーザとは考えにくいマルウェア特有の名前解決動作を確認した．

田中ら [2] は，キャッシュ DNS サーバのクエリログを解析することにより，マルウェア感染ホストが問合せる未だ報告されていない不正 Web サイトのドメインを発見するための手法を提案している．マルウェアは感染したホストをさらに他のマルウェアに感染させるために，悪性 Web サイトへアクセスさせる傾向がある．また，通常のホストは悪性のある Web サイトの FQDN に対するクエリを送信することは考えにくい点から，MDL(Malware Domain List) などのブラックリストにより報告された既知の悪性ドメインと共通するドメインを持つ FQDN をクエリ送信したホストを抽出し，それらのホストが重複してアクセスする Web サイトの中には未知の悪性ドメインが含まれている可能性がある．提案手法を用いた評価実験の結果，ブラックリストの報告にない未知の悪性ドメインを抽出した．

3. 一般ホストと異常ホストの分析結果

大分大学のキャッシュ DNS サーバにおける 2013 年 6 月 1 日から 2014 年 2 月 20 日までの約 9 ヶ月間のクエリログについて，図 1 のように 1 クエリごとの送信元 IP アドレスと FQDN を抽出，集計し，クエリ数に着目して分析した．その結果，通常とは異なる挙動を示す 4 つの送信元ホストを観測した．以下では，まず一般ホストの挙動を説明し，その後，観測した 4 つの異常ホストの挙動を説明する．

表 1 2014 年 2 月 12 日におけるクエリ数の多い FQDN(上位 4 件)

FQDN	クエリ数	クエリ対象とした IP アドレス数
teredo.ipv6.microsoft.com	40,113	734
www.google.com	18,551	1,037
twitter.com	16,289	827
www.facebook.com	15,983	942

表 2 2014 年 2 月 12 日における送信元 IP アドレスと FQDN の集計 (全 455834 組)

クエリ数	組数	割合	累積
10 回未満	413,140	90.63 %	90.63 %
10 から 49 回	38,891	8.53 %	99.17 %
50 から 99 回	2,466	0.54 %	99.71 %
100 回以上	1,337	0.29 %	100.00 %

3.1 一般ホストの挙動

本論文における一般ホストは，異常と判断した 4 つのホストを除いた学内ホスト全体と定義する．一般ホストの挙動は約 9 ヶ月間のクエリログの中から 2014 年 1 月 20 日から 2014 年 2 月 20 日の約 1 ヶ月間を抽出し，分析した結果を示す．

3.1.1 クエリ数の多い FQDN の調査

まず，一般ホストが日常的に利用している FQDN について調査した．調査期間中でクエリ数が最多であった 2014 年 2 月 12 日のクエリログから FQDN を抽出し，集計した．クエリ数上位 4 件の FQDN を表 1 に示す (学内ドメインへのクエリは除外する)．出現回数が最も多かった teredo.ipv6.microsoft.com は IPv4 を IPv6 にトンネリングする際に用いられるサーバの FQDN であり，Windows 側の設定で自動的に有効になっている場合がある．この FQDN をクエリ対象とした送信元 IP アドレスは 734 件であった．www.google.com，twitter.com，www.facebook.com に関しては，Web サイトのアクセス状況を調査している Alexa[3] の集計情報において，上位に位置づけられる人気度の高い FQDN であった．これらの FQDN をクエリ対象とした送信元 IP アドレスは，2 月 12 日のクエリログ中に多数存在していた．

以上のことから，一般ホストが日常的に利用している FQDN は複数の送信元 IP アドレスからの問合せによってクエリ数が増加していると考えられる．

3.1.2 1 つのホストあたりのクエリ数

続いて，1 日ごとのクエリログについて一般ホストがクエリ対象とする FQDN のクエリ数を調査した．2014 年 2 月 12 日について，図 1 の手順により集計した．送信元 IP アドレスと FQDN を 1 つの組とし，何回現れたかを集計し，クエリ数とする．調査した結果を表 2 に示す．

表 2 より，455834 組のうち，クエリ数 10 回未満である

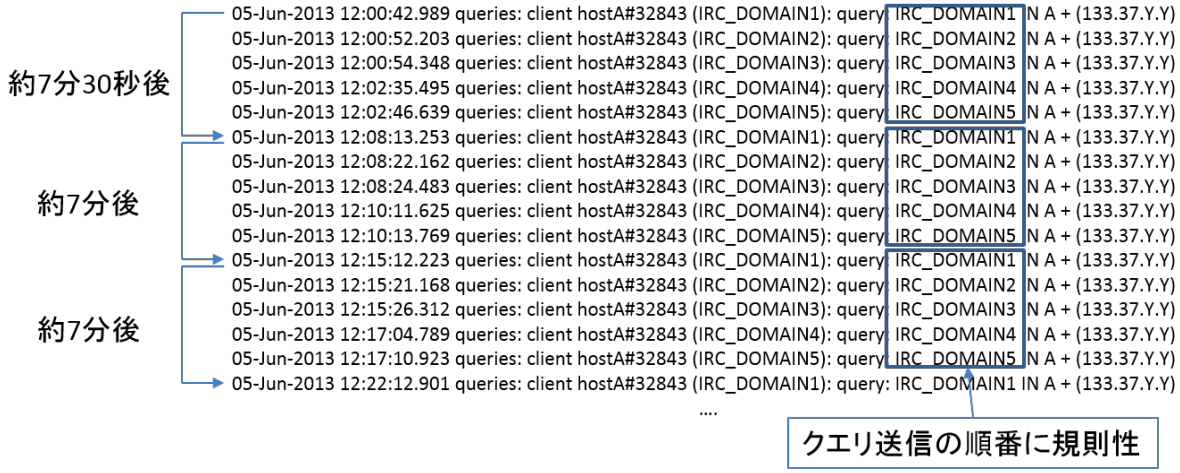

図 2　指令サーバと考えられる 5 つの FQDN のクエリ送信状況の抜粋 (ホスト A)

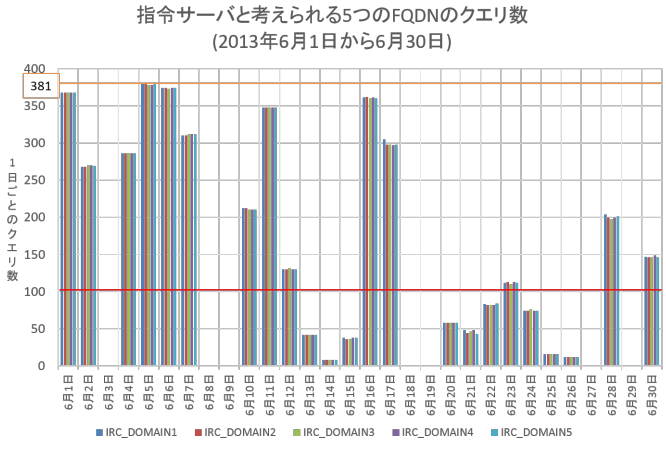

図 3　指令サーバと考えられる 5 つの FQDN のクエリ数

組が全体の約 90 %であった．また，クエリ数 100 回以上である組が全体の約 0.29 %である．分析期間全体を通して調査した結果，いずれの日も表 2 と類似する傾向であった．3.1.1 節で述べた，一般ホストが日常的に利用する FQDN はクエリ数 100 回以上の組に含まれる場合があるが，組数は非常に少ない．このことから，1 日ごとのクエリログにおいて，一般ホストが同一 FQDN を 100 回以上問合せることはほとんど存在しないと推測される．

3.2　異常ホストの挙動

本節では，一般ホストとは異なる挙動を示した 4 つのホスト A，B，C，D の，それぞれの挙動について分析した結果を述べる．

3.2.1　ホスト A の挙動

ホスト A は学内においてボットに感染し，活動していたホストである．ホスト A について観測した期間は 2013 年 6 月 1 日から 2013 年 6 月 30 日である．ホスト A は 5 つの IRC サーバの FQDN を問合せていた．ボットは攻撃者と通信する際に，IRC(Internet Relay Chat)[4] を使用するものが多く，その場合 IRC サーバが指令サーバとなるため，これら 5 つの FQDN は指令サーバの FQDN であると推測される．図 2 は指令サーバの FQDN と推測される 5 つの FQDN に対するクエリを抜粋したものである．図 2 の時刻と FQDN に注目すると，ホスト A は 1 つの FQDN を約 7 分間隔で規則的に，5 つの FQDN を順番にクエリを送信する特徴が観測された．これはボットが指令サーバとの接続のためにプログラムもしくは設定ファイルに記述された FQDN をクエリしていたと推測される．2013 年 6 月 1 日から 6 月 30 日において，1 日ごとにホストが指令サーバの FQDN をクエリした回数を集計した結果を図 3 に示す．

図 3 より，6 月 5 日では 381 回のクエリが観測された．これは観測期間中最大のクエリ数であった．しかし，ホスト A が稼動していても指令サーバの FQDN をクエリしない日が存在していた．図 3 の赤線はクエリ数 100 回を示している．指令サーバの FQDN を問合せている日のうち，14 日間はクエリ数が 100 回を超えており，指令サーバへの接続を試みる際には一般ホストの挙動に対して比較的多いクエリ数が観測された．

これらの FQDN の悪性を評価するため virustotal[5] により検索したところ，2 つの FQDN が悪性であると判断された．また，これらの 5 つの FQDN はホスト A 以外の送信元 IP アドレスがクエリ対象することはなかった．

3.3　ホスト B の挙動

ホスト B はチャットをサービスしているサイトの FQDN を定期的に問合せる挙動が観測された．ホスト B について観測した期間は 2013 年 11 月 29 日から 2013 年 12 月 23 日である．ホスト B がこの FQDN を問合せた回数を集計した結果を図 5 に示す．図 5 の赤線はクエリ数 100 回を示したものである．観測期間において，いずれの日もクエリ数が 100 回以上となっていた．また，他のホストがこの FQDN を問合せた形跡はなかった．

```
5分後
5分後
:
02-Dec-2013 07:45:31.232 queries: client hostB#30099 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 07:50:31.220 queries: client hostB#32664 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 07:55:31.207 queries: client hostB#31196 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 08:00:31.194 queries: client hostB#30023 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 08:05:31.180 queries: client hostB#32983 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 08:10:31.172 queries: client hostB#30924 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 08:15:31.155 queries: client hostB#31127 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 08:20:31.141 queries: client hostB#39057 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 08:25:31.227 queries: client hostB#37170 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
02-Dec-2013 08:30:31.113 queries: client hostB#34661 (CHAT_DOMAIN): query: CHAT_DOMAIN IN A + (133.37.Y.Y)
...
```

図 4　チャットをサービスしているサイトの FQDN のクエリ送信状況の抜粋 (ホスト B)

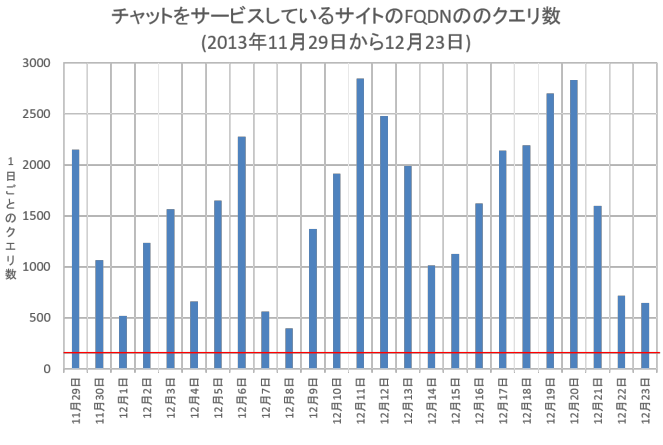

図 5　チャットをサービスしているサイトの FQDN のクエリ数

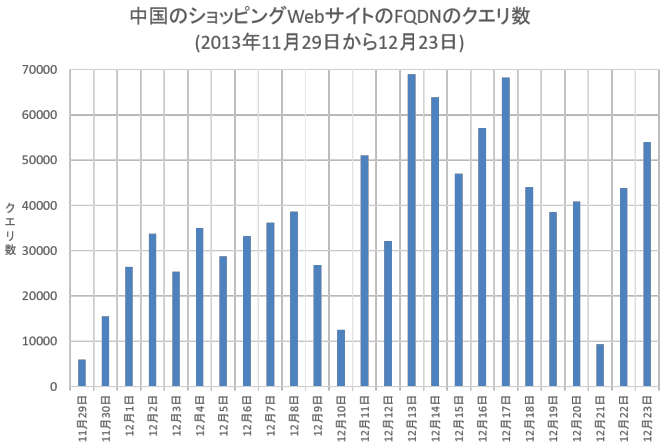

図 6　中国のショッピングサイトの FQDN のクエリ数

次にホストがこの FQDN のクエリを送信している様子について一部抜粋したものを図 4 に示している．

図 4 の時刻に注目すると，約 5 分間隔で規則的にクエリが送信されていた．このような規則的なクエリ送信のため，1 日あたりのクエリ数は大きくなる．観測期間中において最低 396 回のクエリが送信されていた．チャットをサービスしている接続先であり，またその接続先へ規則的にクエリを送信する点が，ホスト A の挙動と類似している．そのためホスト B はボットに感染している疑いがある．

3.4　ホスト C の挙動

ホスト C は中国のショッピングサイトのドメイン名を含んだ FQDN を大量に問合せていた．ホスト C について観測した期間は 2013 年 11 月 29 日から 2013 年 12 月 23 日である．この FQDN のクエリ数を集計した結果を図 6 に示す．観測期間において，12 月 13 日に最大 69,026 回，11 月 29 日に最小 5,897 回のクエリが観測された．この FQDN を問合せる他のホストは観測されなかった．問合せの目的は不明であるが，一般ホストの挙動と比較すると何らかの異常なホストであると推測される．

3.5　ホスト D の挙動

ホスト D は大量の FQDN を問合せていた．ホスト D について観測した期間は 2013 年 6 月 24 日から 2013 年 7 月 18 日である．観測期間のうち，ホスト D によるクエリ数が最も多かったのは 7 月 1 日の 2,614,294 回であった．また，クエリ対象の FQDN は，775,620 種類であった．一般ホストの場合，1 日に問合せる FQDN の種類はたかだか数十から数千種類であるのに対し，ホスト D の問合せは非常に多い．さらに，ホスト D からの大量のクエリが連日観測されていた．その期間キャッシュ DNS サーバは，ホスト D からの大量のクエリにより大きな負荷がかかっていたものと推測される．ホスト D は頭文字が”a” から始まる FQDN を”az” まで問合せると”b” から始まる FQDN を問合せる挙動が観測された．また，ホスト D がクエリ対象とした FQDN は後日同じ FQDN をクエリ送信とするがほとんどなかった．その中でホスト D が活動している日に必ずクエリが送信され，かつ他のホストがクエリ対象としていない FQDN を 3 つ発見した．それらのクエリ数の集計を図 7 に示す．図 7 の赤線はクエリ数 100 回を示したものである．3 つの FQDN のうちいずれか 1 つのクエリ数が 100 回を超える日は観測期間中 23 日であった．これら 3 つの FQDN をキャッシュ DNS サーバ側で名前解決しないように設定したところ，アルファベット順に FQDN を問合せる行為は収束した．

virustotal により 3 つの FQDN の悪性の調査した結果，

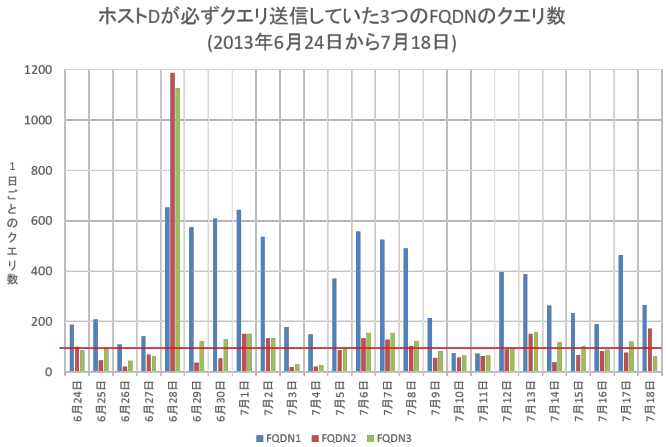

図 7 ホスト D が必ずクエリ送信していた 3 つの FQDN のクエリ数

そのうち 1 つが悪性であると判断された．ホスト D の挙動は，牧田ら [1] が ZeuS 系ボットとして示したアルファベット順に FQDN をクエリする特徴に類似していた．そのためホスト D は ZeuS 系のボットに感染していた可能性がある．

3.6 一般ホストと異常ホストの挙動の特徴

一般ホストと異常ホストの挙動を分析し，以下の 2 つの特徴を得た．

[特徴 1] ：一般ホストは 1 日に同じ FQDN を問合せる回数は少なく，異常ホストは同じ FQDN の問合せる数が多い．

[特徴 2] ：一般ホストが日常的に利用する FQDN は複数の学内ホストが問合せ元として存在するが，異常ホストは他の学内ホストがクエリ対象とすることがない FQDN を問合せる．

この結果に基づき，ホストがクエリ対象する FQDN のクエリ数に着目した異常ホスト検出手法について検討する．

4. クエリ数に着目した異常ホスト検出手法の検討

4.1 異常ホスト検出手法におけるクエリ数の閾値

[特徴 1] より，異常ホストはクエリ数を用いて抽出する．一般ホストの分析結果より，1 日あたり同じ FQDN に対して一般ホストのクエリ数が 100 回以上となるのは全体の 0.29 %と非常に少ない．一方，4 つの異常ホストは特定の FQDN のクエリ数が 100 回を超える場合が多い．そのため，1 日ごとのクエリログについて，1 ホストあたりの同一 FQDN のクエリ数が 100 回以上となる場合を抽出条件とする．

4.2 whitelist による調査対象の絞込み

表 2 より一般ホストがクエリ数 100 回以上となる組は 1,337 組存在しているため，4.1 節の抽出条件のみでは新規の異常ホストのクエリが抽出される．一方，一般ホストからのクエリも多く含まれる．そこで 4.1 節の抽出条件で抽出した FQDN について，[特徴 2] により，複数の送信元 IP アドレスからクエリ対象とされているものは除外する．whitelist は抽出した FQDN から一般ホストのクエリであると推定される FQDN をまとめたものを作成する．学内において同種のボットに感染したホストが存在した場合，同じ指令サーバの FQDN を同期間にクエリする可能性があるため，ここではクエリ対象とした送信元 IP アドレスが 10 件以上存在する FQDN は除外する．

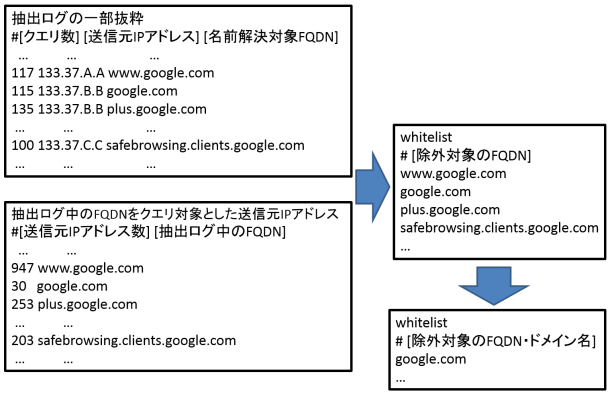

図 8 whitelist の作成手順

4.2.1 whitelist の作成手順

4 月 21 日から 4 月 25 日の 5 日間のクエリログを用いて，[特徴 2] の条件に適合する FQDN を収集する．

まず，4 月 21 日のクエリログについて，クエリ数，送信元 IP アドレス，クエリ対象 FQDN の組を集計し，クエリ数 100 回以上となる組を抽出する．

4 月 21 日のクエリログについて，抽出された FQDN をクエリ送信した送信元 IP アドレス数を集計し，10 件以上の送信元 IP アドレスからクエリ対象となる FQDN ならば，whitelist に追加する (図 8)．

4 月 22 日のクエリログについて，クエリ数，送信元 IP アドレス，クエリ対象 FQDN の組を集計し，クエリ数 100 回以上となる組を抽出する．4 月 21 日分の whitelist と照合し，FQDN を除外する．

4 月 22 日のクエリログについて，残りの FQDN をクエリ送信した送信元 IP アドレス数を集計し，10 件以上の送信元 IP アドレスからクエリ対象となる FQDN ならば，whitelist に追加する．

この手順を 5 日分繰り返し，除外対象の FQDN を収集した．

作成した whitelist は，共通のドメイン名を集約する．

4.3 異常ホスト検出手法の実行手順

1 日分のクエリログに対し，以下の手順に従い抽出する．

(1) クエリ数による抽出 : 1 クエリごとの送信元 IP アドレスとクエリ対象 FQDN を取り出し，クエリ数を集計し

表 3　2014 年 4 月 30 日から 5 月 2 日の抽出ログに含まれていた組数

4 月 30 日	5 月 1 日	5 月 2 日
95 組	84 組	80 組

たのち，クエリ数が 100 回以上となる組を抽出する．

(2) whitelist による FQDN の除外 : (1) で作成した抽出ログ中のすべての FQDN について whitelist と照合し，一致するものは除外する．

ホスト A の挙動 (図 3) ように，異常ホストは悪性の FQDN を必ずしも毎日クエリ送信することはないと推測される．そのため，数日間のクエリログに異常ホスト検出手法を適用し，分析する．

4.4　異常ホスト検出手法を用いた分析結果

2014 年 4 月 30 日から 5 月 2 日のクエリログについて，異常ホスト検出手法を用いて分析し，抽出された組数を表 3 に示す．3 日間の抽出ログについて，これまで発見した 4 つの異常ホストと同一の FQDN が存在するかを調査した．その結果，4 月 30 日の抽出ログ中にホスト B に該当する FQDN をクエリ送信していた学内ホストを検出した．この FQDN のクエリ送信は，図 4 に示したホスト B の挙動と一致しており，同一の異常ホストに分類できる．また 4 月 30 日の抽出ログには，田中ら [2] が未だ報告されていない新種の不正 Web サイトの可能性があると判定した FQDN が存在していた．

4.5　考察

学内ホストが 1 日にクエリ送信する FQDN のクエリ数に着目した異常ホスト検出手法について検討した．抽出条件となるクエリ数の閾値は 100 回とした．2014 年 4 月 21 日から 4 月 25 日における一般ホストのクエリと推定される FQDN を whitelist に追加する．whitelist により FQDN を除外することで，抽出ログの組数は 1000 組から 100 組程度に絞ることができる．2014 年 4 月 30 日から 5 月 2 日の 3 日間のクエリログに異常ホスト検出手法を適用し分析した結果，ホスト B と同一の FQDN をクエリ送信している学内ホストを検出することができた．今回の 3 日間のクエリログの調査では，これまで発見した 4 つの異常ホスト以外の特徴を持つ学内ホストの検出はできなかった．しかし，これから調査期間を増やしていくことで，本検討手法により異常ホストを検出することが可能であると考えられる．異常ホストを検出し，異常ホストの活動を遮断することを目的とする．

5.　おわりに

5.1　まとめ

大分大学のキャッシュ DNS サーバのクエリログを 2013 年 6 月 1 日から 2014 年 2 月 20 日まで分析した結果，一般ホストとは異なる挙動をする 4 つのホストを発見した．一般ホストと異常ホストのクエリ送信に関する挙動を分析した結果，1 つのホストあたりのクエリ数と FQDN をクエリ送信した送信元 IP アドレス数の点において差異を発見した．これらの特徴にもとづいて，キャッシュ DNS サーバのクエリログを用いて異常ホストを検出する手法を検討した．本検討手法を 3 日分のクエリログに適用し，分析した結果，異常ホストであるホスト B と同一の FQDN をクエリ送信する新たな学内ホストを検出した．

5.2　今後の課題

今回の調査期間では，これまで発見した異常ホスト以外の挙動をする学内ホストを検出することはできなかった．しかし，今後調査期間を増やすことで，検出可能であると考察する．また，今回検討した異常ホスト検出手法はクエリ数に着目したため，別の特徴から異常ホストを検出可能かを検討していく．例えば，規則性のあるクエリ送信をしていた学内ホストに着目して分析することで新たな特徴を発見できると考えられる．

参考文献

[1] 牧田大佑，吉岡克成，松本勉，”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”，情報処理学会研究報告，第 61 回 CSEC・第 21 回 IOT 合同研究発表会，Vol.2013-IOT-21，No.7，pp.1-6，2013 年 3 月

[2] 田中晃太郎，長尾篤，森井昌克，”DNS ログからの不正 Web サイト抽出について-解析手法とその匿名化-”，情報処理学会，コンピュータセキュリティシンポジウム 2013 論文集，Vol.2013，No.4，pp.132-138，2013 年 10 月

[3] Alexa，http://www.alexa.com//

[4] C.kalt，” Internet Relay Chat Architecture”，RFC 2810,April 2000，https://tools.ietf.org/html/rfc2810

[5] virustotal，https://www.virustotal.com/