ボットネットの C&C サーバ特定手法の最新データを用いた評価

岡安 翔太†　佐々木 良一†

†東京電機大学
〒120-8551　東京都足立区千住旭町5
okayasu@isl.im.dendai.ac.jp, sasaki@im.dendai.ac.jp

あらまし マルウェアに感染した複数の PC 群から構成されるボットネットによる被害は年々増加している．感染 PC に対策を行ったとしても，ボットネットを操作する攻撃者を特定しない限り，再度感染等の被害が発生し根本的な解決にはならない．そこで著者らは，ボットネットを根源まで追跡する多段追跡システムの第二段追跡方式として，数量化理論２類を用いて C&C サーバを検知する手法の開発を行ってきた．本稿では，マルウェアのドメインリストを公開している DNS-BH のデータから 2014 年度に登録された最新のドメインを用いて，先の特定手法における最適なパラメータ値の設定を行い，検証を通して手法の有効性を確認出来たので報告を行う．

Evaluation of Method for Detecting C&C Server of Botnet using the Latest Data

SHOTA OKAYASU†　RYOICHI SASAKI†

†Tokyo Denki University
5, Senjuasahi-cho, Adachi-ku, Toukyo, 120-8551 JAPAN

Abstract The damage caused by botnet, which consists of multiple PCs infected with malware is increasing year by year. Even if the infected PCs could be found and recovered, it is not a fundamental measure, because the attackers can infect other PCs again easily. Therefore, the authors have developed a method to detect the C & C server using the mathematical quantification theory class II for track the source to botnet as second stage of the multistage tracking system developed by the authors. This paper describes the obtained optimal parameter values to detect the C & C server using the method revised from the method proposed previously, and the data registered in 2014 year part of DNS-BH which shows domain list of malware. The experiment to identify the C & C server using the parameter values confirmed the effectiveness of the revised method.

１　はじめに

近年ボットネットによる被害が増加しており問題となっている．ボットネットとは悪意を持った攻撃者の命令に基づき動作するプログラムに感染した PC(以下，ボット PC)及び攻撃者の命令を送信する指令サーバ(以下，C&C サーバ)からなるネットワークであり，中には数万規模の PC などからなるボットネットもあると言われている[1]．攻撃者が C&C サーバに命令を送ることで，ボットネットに接続されたボット PC はフィッシング目的などの SPAM メールの大量送信や，特定サイトへの DDoS(Distributed Denial of Service)などに利用され，非常に大きな脅威となりうる[2]．これらのボット PC を用いた攻撃の，攻撃元の特定手法として IP トレースバックなどを用いることで，攻撃元を偽装した場合でも検出可能である．しかし，対策が不十分であれば PC に容易に感染するおそれがあるため，根本的な解決とはならない．

このような問題に対して本研究室では，ネットワーク管理者が情報共有を行い，ボット PC や C&C サーバ，攻撃者の特定を目

的とする，多段追跡システムを構成した[3]．本稿は，このうち第二段において C&C サーバ・ダウンローダ(以下，第二追跡対象)を検知する方式に関するものである．

本方式は数量化理論を用いて違法ドメインを識別することにより，ブラックリストに載っていないような第二追跡対象であっても，検出できるというという特長を持っている．しかし，この方式は，研究の結果，時間の経過とともに検知率が落ちていくことが分かっており，経年変化の調査と，それぞれの時点における最適な検知方式の提案をおこなってきた[3]．本報告は，2014 年度を対象に調査を行ったものである．

なお，2008~2011 年度と 2013 年度に関するデータは，従来マルウェア対策研究育成ワークショップ[4]より提供されてきた．しかし，今回 2014 年度データの提供が行われなかったため，ドメインのブラックリストをから有効なデータを抽出することにより，独自に入手できるようにした．

３　関連研究

ボットネットにおける，第二追跡対象の特定を目的とした研究は，特定方法により次の 2 つに分類される．

(1) 第二追跡対象との制御通信に着目した検知方式

C&C サーバと感染 PC 間で行われる通信に着目し，制御通信のペイロードに含まれる文字列などの特徴を分析することで検出を行う手法[5][6]がある．

これら手法は，トランスポート層のポート番号や独自プロトコルといった仕様変更に伴い，対応出来なくなる従来の検出手法と異なり，宛先アドレスや発信元アドレスなどヘッダ情報を除いたデータ本文を検証する為，十分な検証により高い検出精度を出す．しかし，ゼロデイ攻撃等の未検証検体への対応に不十分な点が有る．

(2) 第二追跡対象のドメイン情報に着目した検知方式

感染 PC に潜伏するボットウイルスは，DNS サーバに対して第二段追跡対象の名前解決を行う事がある．

Meng-Han Tsai ら[7]は，第二段追跡対象のドメインに着目し，設定されているドメイン情報や外部リポジトリから取得した情報を併用して，RIPPER と呼ばれるデータマイニング手法を用いて検証をしている．これらは，活動中の C&C サーバに関して高い精度で判別を行う．一方，検出漏れを発生させている．

これに対し，著者らが提案する方式は数量化理論 2 類を用いる方式である．(1)の検知方式では実際の通信を用いる必要があることに対し，提案方式では第二追跡対象の，ドメインの登録期間や逆引きの結果といった，ドメインの特徴を用いる．そのため，解析に必要なデータ取得が容易であり，解析の安全性も高いといえる．

３　第二段トレースバックシステム

第二追跡対象の特定には，数量化理論 2 類を用いた検知方式を行う．本章では数量化理論を用いた検知方式の説明をする．併せて，検知で用いるデータ，特徴要素の抽出手法について説明する．

3.1 数量化理論を用いる検知方式

数量化理論は，元統計数理学研究所所長の林知己夫教授らにより開発されたデータ分析手法である[8]．この内，数量化理論 2 類ではダミー変数の導入による質的データの数量化を行うことで，判別分析に相当する処理を可能にする．

例えば，分析対象データの集合に 1 群と 2 群が混在するとき数量化理論 2 類を用いて 1 群と 2 群に判別するケースを考える．初めに，1 群と 2 群，それぞれを特徴付けるデータをパラメータとして設定した判別式と 1 群と 2 群に判別するための基準となる境界値の 2 種を設定する．この判別式に集合の各要素を入力し出力される判別値を，先に求めた境界値と比較する．この際に，判別値が境界値より高い値ならば 1 群，低い値ならば 2 群と判別を行う．また，使用した集合外の要素についても同様に境界値を比較することで，1 群・2 群どちらに属する可能性が高いかを推定することが可能である．

数量化理論 2 類を用いる検知方式(以下，数量化検知方式)で第二追跡対象の検知を行

うにあたり，「ドメインの特徴量を，数値化処理を行い判別式と判別境界値を求める」パラメータ設定実験と，「求められたパラメータ設定を，異なるデータを用いて検証する」検証実験の二段階に分かれる．

3.2 先行研究

本研究は三原ら[3]が 2009 年に手法を確立してから現在まで継続研究が行われている．その理由はボットネットの特徴が時間経過により変動する傾向にあるためである．その結果，数量化検知方式で求めた従来のパラメータ設定値では対応できなくなってしまう．そのため，最新のデータを用いパラメータの設定値を更新する必要がある．

2011 年度に中村[9]が CCC DATAset[10]を基に，数量化検知方式で用いる特徴量の追加や修正を施した結果，検出精度の改善に成功した．2013 年度には本論文の著者[11]が PRACTICE Dataset 2013[12]を基に 2011 年度に設定されたパラメータの有用性の検証を行いドメインの特徴変動を確認した．そのため，特徴変動で低下した検出精度を改善するため，2013 年度データを用いてパラメータの再設定を行った．結果，検知精度の改善に一定の成果が得られた．しかし，2013 年度パラメータでの検知精度は 80%程度と決して高いとはいえない値となった．

以上のことからドメインの特徴変動に対応するため最新のデータを用いパラメータ設定値を最適化することに加えて，従来使用してきた特徴量について，追加や修正を施すことで，検知精度の改善を行う必要があると言える．

3.3 使用データ

既存方法に倣い，数量化検知方式を用い第二追跡対象の検知を行うため最新のデータが必要である．そこで，ボットネットに関連のあるドメイン(以下，B ドメイン)と，比較対象としてボットネットに関係の無いドメイン(以下，N ドメイン)の 2 種類のデータを用いる．データの詳細を以下に，データの取得数を表 1 に示す．

B ドメイン

(1) 2014 年度 B ドメイン

N ドメイン

(2) 大規模サイト(以下，big ドメイン)

(3) 中規模サイト(以下，mid ドメイン)

(4) 小規模サイト(以下，small ドメイン)

表 1． 使用ドメイン内訳

	ドメイン数(個)
B ドメイン	104
big ドメイン	500
mid ドメイン	152
small ドメイン	32
総数	788

B ドメインには 2014 年度に取得した最新のドメインを用いる．先行研究では B ドメインに攻撃通信が含まれた研究用データセットである CCC DATAset や PRACTICE Dataset から必要ドメインの抽出を行い用いていた．しかし，2014 年度に上記のような研究用データセットの提供が行われなかったため，DNS-BH[13]からボット PC が接続する第二追跡対象のドメインリストを用いる．このリストから 2014 年度に追加されたドメインを新たに発見されたものとみなし，最新の B ドメインと定義し，抽出を行った．なお，(1)は各ボット PC が接続する第二追跡対象のドメインであり，ボットネット自身のドメインではない．

N ドメインはサイトの規模によりドメインの特徴が異なる傾向にある．そのため，サイトの規模を，アクセス数や，メール配信可能規模等を参考に大，中，小，の三段階に分類した．(2)は世界のアクセスランキングトップ 500 を掲載している" The top 500 sites on the web"[14]から取得した．(3)は IR サイトランキング[15]と FORTUNE[16]から中規模企業のドメインを取得した．(4)も同様に FORTUNE から取得した．

3.4 データ解析

数量化検知方式では，3.3 節で述べたデータのドメイン情報を用いた解析を行う．数量化検知方式に利用する特徴量として，調査する項目を表 1 に示す．今回，検知精度の向上を図るため，2011 年度では 8 種類であった特徴量に，新たな項目を追加した．
追加した項目は特徴量番号 8 の TXT レコー

ドである．

各項目の調査には，ドメイン情報を持つDNS サーバに対して特徴量番号 1~8 では dig コマンドを用いて調査を行う．特徴量番号 9 は WHOIS サービスを用いて調査を行う．

表２．　特徴量 9 種類

番号	特徴量
1	逆引き
2	TTL
3	minimum
4	A レコード
5	MX レコード
6	NS レコード
7	CNAME レコード
8	TXT レコード
9	登録期間

各項目の説明として，特徴量番号 1 は，DNS サーバに対して IP アドレスからドメイン名の問い合わせの可否を行う調査．特徴量番号 2，3 は，DNS サーバから取得したドメインの設定情報が記載されている SOA レコードから，設定値を調査．特徴量番号 4~8 は DNS で定義されるドメインについての情報であり，各項目の個数や有無を調査する．特徴量番号 9 は，各レジストリ組織が管理している，ドメインの登録情報から，ドメインの登録日時と利用期限の調査を行う．その差を登録期間とする．本稿では上記 9 つを DNS のドメイン情報の特徴量とし，判別と評価を行う．

４　実験による検証と評価

前３章で示したデータを基に，数量化検知方式での判別と評価を行う．実験には株式会社エスミ社のソフトウェア Excel 数量化理論 Ver3.0[8]を使用する．下記 3 種の実験を通し，ドメインの特徴変動に対応した有用性の高いパラメータ設定を行う．

1. 2013 年度設定パラメータの有効性の検証
2. 2014 年度データでのパラメータ設定
3. 2014 年度設定パラメータの有効性の検証

4.1 実験概要

4.1.1 ドメインの特徴量設定値

数量化検知方式を用いるためにドメインの特徴量を数値化処理する必要がある．特徴量設定値を表３に示す．

表３．　特徴量設定値

逆引き	値
返答なし	1
不正	2
部分一致	3
完全一致	4

TTL	値
1-1000	1
1001-10000	2
10001-	3

minimum	値
1-1000	1
1001-10000	2
10001-	3

MXレコード	値
有り	1
無し	2

Aレコード	値
有り	1
無し	2

NSレコード	値
有り	1
無し	2

CNAMEレコード	値
有り	1
無し	2

TXTレコード	値
有り	1
無し	2

登録期間　（日）	値
1-2500	1
2501-5000	2
5001-	3
N/A	4

表３の設定値を基に，数量化検知方式でパラメータ設定実験と検証実験を行う．そして，最も検知精度の高い特徴量の組み合わせ，判別式，判別境界値を最適パラメータとする．特徴量の最適な組み合わせの項目数の決定方法は次項の赤池情報量基準を用いることで求めることができる．

4.1.2 最適パラメータ数の選定

パラメータ設定実験で最適な特徴量の数を調べるために赤池情報量基準(以下，AIC）[17]を用いる．AIC は元統計数理研究所所長の赤池弘次によって考案された，統計モデルの良さを評価するための指標である．AIC を用いる事で，モデルの複雑さとデータとの適合度のバランスを取る事が可能となる．データを統計的に説明する数式では，用いるパラメータの数を増やせば，測定データとの適合度が高くなる．しかし，無意味なノイズの影響を多く受ける事に繋がり，信頼性が低下してしまう．このような問題に対して，AIC は式(1)によって求められる

最小 AIC 時のパラメータ数を選択する事で，多くの場合，最適なモデルを選択する事が可能となる．

$AIC = -2\ln L + 2k$　　式(1)

L は最大尤度，k は自由パラメータである．今回の場合，k が各要素数に相当し，L が各パラメータ数での数量化理論２類を用いて求めた判別結果と，正答との乖離の最小 2 乗和に相当する．

また，より正確な結果を得るために AIC の比較対象としてベイズ情報基準(以下，BIC)を最適パラメータ数の選定に併用する．BIC は AIC 同様，統計における情報基準の一つである．BIC は AIC と同様に，式(2)によって求められる最小 BIC 時のパラメータ数が，多くの場合，最適なモデルとなる．

$BIC = -2 \cdot \ln(L) + k\ln(n)$　　式(2)

L は最大尤度，k は自由パラメータ，n が観測データの数である．

AIC と BIC の二つを比較することで，より最適なパラメータ数の選定を行う．

4.1.3 実験における検知率

本実験におけるドメインの識別判定方式を表４に示す．識別結果に対する評価指標として，正しく B ドメインと判別されることを True Positive，その割合を True Positive Rate(以下，TPR)．正しく N ドメインと判別されることを True Negative，その割合を True Negative Rate(以下，TNR)．この 2 つを合わせたものを検知率とする．また，B ドメインが N ドメインと判別されることを False Negative，その割合を False Negative Rate(以下，FNR)．N ドメインが B ドメインと判別されることを False Positive，その割合を False Positive Rate(以下，FPR)．この 2 つを合わせたものを誤検知率とする．

表４． 検知判定組み合わせ

	検知結果が真	検知結果が偽
B ドメイン	True Positive	False Negative
N ドメイン	False Positive	True Negative

第二段トレースバックシステムは第二追跡対象の検出を目的としている．そのため B ドメインの検出漏れを防ぐため，False Negative の値が低いことが望ましい．

4.2 2013 年度設定パラメータの有効性の検証

B ドメインの特徴量は時間経過とともに変動することが先行研究から判明した．本節の実験は，2013 年度に数量化検知方式で導き出した最適パラメータが 2014 年度でどの程度有効であるかの検証を行う．先行研究より 2013 年度の最適パラメータ数は AIC より 4 個，特徴量の組み合わせは「逆引き」，「A レコード」，「CNAME レコード」，「登録期間」である．2013 年度検証データを適用したところ検知率は 80%であった．このデータを基に 3.3 節で示した 2014 年度のドメインで検証実験を行う．検証用データには B ドメイン 45 個と N ドメイン 45 個を用いる．N ドメインの内訳は，big ドメイン 15 個，mid ドメイン 15 個，small ドメイン 15 個である．以上の組み合わせデータを 2014 年度検証データとする．検証データを用い 2013 年度との検知率の比較を行う．表５に実験結果を示す．

表５． 検証比較結果

	TPR	TNR	検知率
2013 年度検証データ	88.9%	72.7%	80.8%
2014 年度検証データ	88.4%	77.8%	83.3%

2014 年度検証データの検知率は 83.3%と 2013 年度の検証と大きな差は見受けられなかった．しかし，検知率が 80%前後であることは，決して高い値であるとはいえない．そこで検知率を向上させるため，最新のパラメータに更新する必要があると言える．

4.3 パラメータ設定実験

本節では 2014 年度のドメインから，数量化検知方式により最適なパラメータの設定を行う．設定の教師用データに，2014 年度の B ドメインを 30 個と big ドメインを 30 個使用した．今年度のパラメータ設定実験では 3.3 節で述べたように TXT レコードを追加したデータを用いる．

パラメータ設定実験の結果，最適パラメータ数は AIC，BIC から，ともに 5 個となった．2013 年度では 4 個が最適なパラメータ数であったが，用いたデータ量を増やした

ことで最適なパラメータ数が変化した．今回の実験により高い識別率の最適パラメータが数多く見つかった．そのため，抜粋した最適パラメータ候補に①～④までの通し番号をふり，その通し番号に対応するように識別精度を表６，７に示す．また，教師用データを big ドメイン以外の mid ドメイン，small ドメインやそれらの組み合わせで実験を行ったが，big ドメインを用いる設定が最も識別精度が高かった．教師用データに大規模サイトのドメインを用いることで他の規模まで内包した結果が出たと言える．

表６． 最適パラメータ候補

番号	特徴量組み合わせ				
①	逆引き	TTL	登録期間	TXT	NS
②	minimum	MX	A	登録期間	TXT
③	逆引き	minimum	MX	登録期間	TXT
④	TTL	minimum	登録期間	TXT	NS

※レコード省略

表７． 設定データ識別率

	TPR	TNR	識別率
①	100%	96.8%	98.3%
②	96.7%	96.7%	96.7%
③	96.7%	96.7%	96.7%
④	96.6%	93.5%	95.0%

パラメータ設定実験の結果，①が最も識別率が高い組み合わせとなった．更に，TPR が 100%であり B ドメインを正しく分類することが出来たと言える．表６の最適パラメータ候補が，異なるデータではどの程度有効であるかを次節で検証する．

4.4 検証実験

4.3 節で設定されたパラメータの有用性を検証するため，パラメータ設定用データとは異なるデータで検証実験を行う．そこで 3.2 節で規模ごとに分類した N ドメインを用いる．big ドメイン，mid ドメイン，small ドメインの 3 種類の N ドメインと B ドメインを組み合わせたものを検証用データとする．以下に検証用データの組み合わせと，表８に検証結果を示す．

Big　：B ドメイン 30 個＋big ドメイン 30 個
Mid　：B ドメイン 30 個＋mid ドメイン 30 個
Small：B ドメイン 14 個＋small ドメイン 16 個

表８． 検証結果

	Big		Mid		Small		検知率
	TPR	TNR	TPR	TNR	TPR	TNR	
①	100	96.7	96.7	66.7	100	100	92
②	96.7	100	100	86.7	100	87.5	95.3
③	96.7	96.7	100	70	100	100	92.7
④	96.7	100	100	86.7	100	100	96.7

単位[%]

4.3 節，表 6 の最適パラメータに対し，Big，Mid，Small の検証用データで検証実験を行った．表 7 からパラメータ設定実験では①が最も識別率が高かったが，検証結果では mid ドメインの TNR が低く，N ドメインの誤検知が目立った．総合的に最も検知率が高かったのは④のパラメータ設定である．TPR，TNR が他の組み合わせよりも軒並み高く，検知率は 96.7%である．④の検知漏れに関しては全 150 ドメイン中，B ドメインが 1 件，N ドメインが 4 件であった．

以上により 4.3 節で設定したパラメータ④である「TTL」，「minimum」，「NS レコード」，「TXT レコード」，「登録期間」の 5 つの特徴量の組み合わせが最も有効であると検証することができた．また，2013 年度の最適パラメータよりも高い検知精度となり，B ドメインの検知について一定の成果が得られた．

4.5 最適パラメータにおける特徴量

今回の検証で最適パラメータとされる④のパラメータ設定値である，カテゴリースコアを表 9 に示す．

表 9 を使用し，各特徴量のカテゴリースコアを加算することで，サンプルスコアを求めることができる．例えば，未知のドメイン α が B ドメインであるか N ドメインであるか予想する場合を考える．まず，ドメイン α に対し，各特徴量の調査を行い分類されるカテゴリーとそのスコアを求める．そして，各スコアを加算することでドメイン α のサンプルスコアを求める．この時，求めたサンプルスコアの値が判別的中点よりも高ければ B ドメイン，低ければ N ドメインと予想される．

表９． 最適パラメータのカテゴリースコア

項目名	カテゴリー名	分類数	スコア
TTL	1-1000	38	0.009
	1001-10000	9	-0.041
	10001-	13	0.001
minimum	1-1000	25	-0.177
	1001-10000	13	-0.072
	10001-	22	0.244
NSレコード	有り	38	-0.031
	無し	22	0.053
TXTレコード	有り	28	-0.425
	無し	32	0.372
登録期間	1-2500	28	0.504
	2501-5000	10	-0.395
	5001-	17	-0.670
	N/A	5	0.243
	判別的中点		0.125

以上より，表９のスコアが高ければ高いほど B ドメインの特徴を強く表していると言える．特に TXT レコードと登録期間のスコはドメインの識別に大きな影響を及ぼしている．次項でこの ２ つの特徴量の傾向調査結果を示す．

4.5.1 TXT レコード

表１０． ドメイン別 TXT レコード数

	Bドメイン	bigドメイン	midドメイン	smallドメイン
有り	3	393	96	23
無し	101	107	56	9

単位[個]

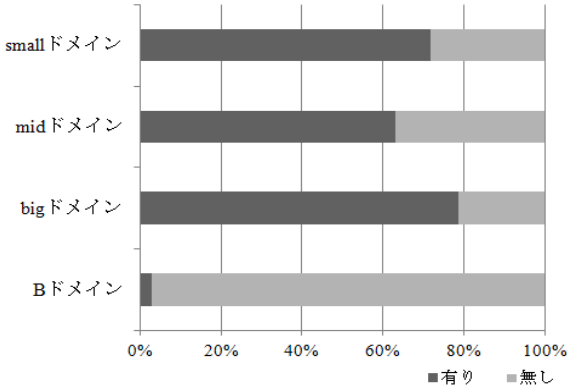

図１． ドメイン別 TXT レコードの割合

TXT レコードは，DNS で定義される情報の一種で，ドメイン外部のソースにテキスト情報を提供する 役割を持つ．ドメインの所有者の確認やメールのセキュリティ対策の実装等に用いられる．B ドメインと N ドメインの TXT レコードの設定状況を比較するため，3.3節の表１のドメインから TXT レコードの有無を調査した．結果を表１０に，その割合を図１に示す．

B ドメインは TXT レコードを設定しない場合がほとんどである．N ドメインはサイト規模により，設定状況が少し異なるが 6 割以上は設定している．

この設定状況の違いが TXT レコードのスコアに顕著に表れている．表９のスコアより，有りは-0.425，無しは 0.372 である．スコアの幅が他の特徴量よりも大きく，ドメインの識別に大きな影響を及ぼしている．

4.5.2 登録期間

前項と同様にドメインの登録期間の調査を行い，その割合を表したものを図２に示す．横軸の単位は日である．

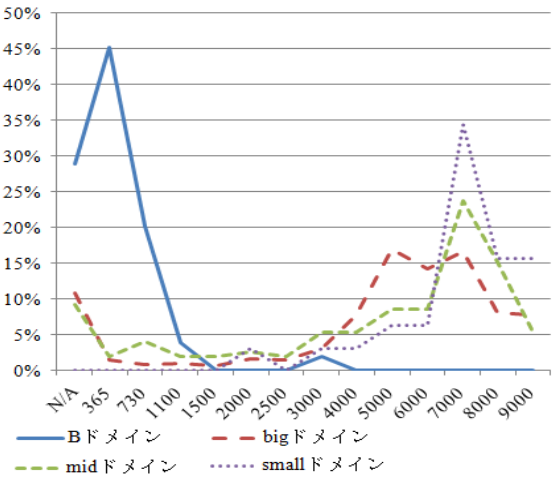

図２． ドメイン別登録期間の割合

図２の B ドメインの多くは比較的，登録期間が短い．さらに，管理情報の取得が出来ないようになっており，登録期間が見つからないドメインも多数存在する．対して N ドメインの登録期間は長い傾向にある．理由として，N ドメインは正規サイトという性質上，長期の運用を目的としているからだ．

このような登録期間の特徴は表９のスコアに大きく影響している．登録期間が 2500 日以下と短い場合のスコアは 0.504 である．この数値は B ドメインを特徴付けるものとして用いられており，全スコアの中で最も

多きな値となっている．対して，登録期間が 5000 日より長い場合スコアは-0.67 である．N ドメインと特徴付けるための，最も影響力の大きな値である．このことから，登録期間は B ドメインと N ドメインを識別するにあたり，大きな指標となっている．

５　おわりに

多段追跡システムのうち，先に提案された第二段トレースバックの解析手法について 2014 年度に取得した新たなデータによる検知方式の検証を行った．先行研究より，時間経過によるドメインの特徴量の変化がもたらす，ドメインの識別率への影響が懸念されていた．今回の検証では 2013 年データと 2014 年データで大きな差は見られなかった．しかし，より精度の高いドメインの識別をおこなうため，新たな特徴量として「TXT レコード」を追加し，最適なパラメータの選定を行い，有効性の検証を行った．結果，大幅な検出精度の改善へとつながった．

しかし，今後も時間経過により，パラメータに用いる特徴量は変動することが予想される．そのため，継続的な特徴量の観測と，その変動に対応した最適なパラメータを自動算出できる仕組み検討する．加えて，プロキシサーバを用いることで要求のあった URL に対し，ドメインの判定を行い対処するフィルタ機能を実装することで，ドメインの特徴変動に動的に対応可能な検知システムの開発を検討する．

また，今回のパラメータの決定には数量化理論を用いたが，機械学習による検証方法も存在する．数量化理論と機械学習では学習方法が異なるため，検知精度や最適パラメータに差がでる可能性がある．そのため，今後は機械学習での検証を試みることで，数量化理論との検知精度の比較を行う．

６　参考文献

[1]警視庁 情報セキュリティ広場，http://www.keishicho.metro.tokyo.jp/haiteku/haiteku/haiteku409.htm

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

[3]三原元，佐々木良一，”数量化理論と CCCDATAsets2009 を利用したボットネットの C&C サーバ特定手法の提案と評価”，*情報処理学会論文誌 VOL.51*，No.9，pp1579-1590

[4]マルウェア対策研究人材育成ワークショップ，http://www.iwsec.org/mws/2014/

[5] D. I. Jang, M. Kim, H. C. Jung, and B. N. Noh, "Analysis of HTTP2P Botnet: Case Study Waledac," 2009 Ieee 9th Malaysia International Conference on Communications (Micc), pp. 409-412, 2009.

[6] Wei. Lu, M. Tavallaee, Ali. A. Ghorbani, “Automatic Discovery of Botnet Communities on Large-Scale Communication Networks，”, ASIACCS '09 Proceedings of the 4th International Symposium on Information , Computer, and Communications Security , 2009

[7] M. H. Tsai, K. C. Chang, C. C. Lin, C. H. Mao, H. M. Lee, and Ieee, "C&C Tracer: Botnet Command and Control Behavior Tracing," in IEEE International Conference on Systems, Man and Cybernetics (SMC), Anchorage, AK, 2011, pp. 1859-1864.

[8]株式会社エスミ，http://www.esumi.co.jp/

[9]中村暢宏，佐々木良一，”累積データを用いたボットネットの C&C サーバ特定手法の評価”，*コンピュータ・セキュリティシンポジウム 2011 論文集,* No.3，pp456-461

[10] 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 ～MWS 2011 Datasets～，MWS2011（2011 年 10 月）

[11]岡安翔太，佐々木良一，”ボットネットの C&C サーバ特定手法の経年変化データを用いた評価”，*第76回全国情報処理学会論文集*

[12] 大村優，畑田充弘：PRACTICE Dataset，MWS2013（2013 年 6 月），http://www.iwsec.org/mws/2013/about.html

[13] DNS-BH – Malware Domain Blocklist，http://www.malwaredomains.com/

[14]The top 500 sites on the web , http://www.alexa.com/topsites/global

[15]Gome，http://www.gomez.co.jp/

[16]FORTUNE，http://archive.fortune.com/

[17] 赤池弘次，甘利俊一，北川源四郎，樺島祥介，下平英俊：赤池情報量基準 AIC（2007）