Vol.2009-IOT-7 No.11
2009/10/9

DNS トラフィックデータを利用したボット感染者検出方法

佐藤一道[†1]　石橋圭介[†1]
豊野剛[†1]　三宅延久[†1]

迷惑メール送信や DDoS 攻撃を行うボットネットの脅威が高まっている．ボットウィルスに感染している端末を検出する方法として，悪性ドメインリストを用いて DNS クエリを監視する手法があるが，悪性ドメインを網羅したリストは存在しないため，全てのボット感染ユーザを発見することはできない．そこで，本稿では既知の悪性ドメインリストを利用し，DNS トラフィックデータから新たに悪性ドメインを抽出する手法を提案する．新たに発見された悪性ドメインの名前解決を行うユーザを抽出することで，既知の悪性ドメインリストだけでは捕えることができないボット感染ユーザを発見することができる．提案手法を試験的に評価したところ，既知の悪性ドメインリストには存在しない悪性ドメインを発見でき，新たにボット感染ユーザを検出できたことを確認した．

Analyze DNS Queries for Detecting Anomalous End-hosts and Domain

Kazumichi Sato ,[†1] Keisuke Ishibashi ,[†1]
Tsuyoshi Toyono [†1] and Nobuhisa Miyake[†1]

Botnet activities like sending spam email and DDoS attack have been a serious problem. An approach for detecting infected hosts by using malicious domain names list is proposed. However it may not detect all of infected hosts in that there is no complete malicious domain names list. In this paper, We propose a new method based on co-occurrence frequency between two domain names to detect bot. We apply proposed method to DNS traffic data and malicious domain names list from honeypots. The result show that our approach detect new malicious domain name and bot.

1. はじめに

ボットウィルスと呼ばれる悪意のあるプログラムに感染した端末が第三者に操作され，迷惑メール送信や DDoS 攻撃，他の端末への感染活動などの迷惑行為を行う**ボットネット**の脅威が高まっている．このような脅威から，ネットワーク運用者が管理している網内のボット感染ユーザを検出する要求が高まっている．

ボット感染ユーザを検出する方法として，DNS クエリを監視する手法が挙げられる[6),8)]．ボットウィルスはボット感染ユーザに攻撃指令を与えるサーバや悪意あるプログラムの配布サイトなど[⋆1]と通信するために DNS を用いることが知られている[7)]．この手法ではボットウィルスを解析して得られる指令サーバのドメインリストを用い，その悪性ドメインの名前解決を行うユーザをボット感染ユーザとして抽出するものである．

しかしながら，全ての悪性ドメインを網羅したリストは存在しないため，網内に存在する全てのボット感染ユーザを検出することはできない．そこで，本稿では網内のボットウィルスの感染状況の全容を把握することを目的とする．

この目的を達成するために，本稿では既存の悪性ドメインリストと DNS トラフィックデータを用いて，新たに悪性ドメインを発見する手法を提案する．具体的には，指令サーバなどの冗長化のため，1 つのボットウィルスが名前解決を行う悪性ドメインは 1 つだけではないという性質から導かれる以下の仮説に基づき，ボットウィルスが名前解決を行う悪性ドメインの共起ドメインを抽出することで新たな悪性ドメインの発見を試みる．ここで，共起ドメインとはあるユーザが問い合わせる異なる 2 つのドメイン d_1, d_2 の組と定義する．

仮説　悪性ドメインと共起するドメインの中には，既知の悪性ドメインリストに存在しない悪性ドメインが含まれる可能性がある

本提案手法で新たに発見された悪性ドメインを問い合わせるユーザを抽出することで未知のボット感染ユーザを抽出することができる．

提案手法を DNS トラフィックデータに適用したところ，実際に新たに悪性ドメインを発見することができ，既知の悪性ドメインリストでは捕えられないボット感染ユーザを発見す

†1 NTT 情報流通プラットフォーム研究所
NTT Information Sharing Platform Laboratories

⋆1 以降，本稿ではこれらの通信相手を悪性ドメインと呼び，悪性ドメイン以外のドメインを正規ドメインと呼ぶ．また，ドメインとは FQDN (Fully Qualified Domain Name) のことを差す

ることができた．

以下，本稿では 2 節で本稿の関連研究を紹介し，3 節でボット感染ユーザ検出のステップ，4 節で提案手法の詳細を述べる．5 節では提案手法の評価を述べ，6 節で提案手法の課題，7 節で本稿をまとめる．

2. 関 連 研 究

DNS サーバへの問い合わせドメインに分析し，ボットウィルスに感染しているユーザを特定する手法として文献[8] がある．文献[8] ではある網内に脆弱性を模擬した端末を設置し，その端末へ感染活動を行うボットウィルスの収集および分析を行い，ボットウィルスが名前解決を行う悪性ドメインのリストを生成する．得られた悪性ドメインの名前解決を行うユーザを抽出することでボットウィルスに感染しているユーザを特定する．しかし，脆弱性を模擬した端末に感染活動を行わないボットウィルスも存在するため，網羅的な悪性ドメインリストを生成することはできず，網内の全てのボット感染ユーザを発見することはできない．本稿では上記のようにして得られた悪性ドメインリストを基にし，DNS トラフィックデータから既知の悪性ドメインリストに含まれていない悪性ドメインの抽出を行う．

また，著者らは以前，DNS トラフィックデータのうち，DNS サーバへの問い合わせドメインに注目してワームに感染しているユーザを抽出する方法[4]，悪性ドメインリストの精度を向上させる手法[3] を提案した．

文献[4] ではユーザが名前解決を行うドメインを，ベイジアンスパムフィルタ[2] の手法を応用してスコア付けを行う手法を提案している．しかし，この手法では正規ドメインの名前解決を大量に行った場合，感染していないと判定されてしまう可能性がある．そこで，本稿では悪性ドメインの名前解決を行っているかで感染の有無を判定し，ボットウィルスが正規ドメインの名前解決を大量に行った場合でも正しくボットウィルス感染の有無を判定する．

文献[3] では既知の悪性ドメインリストに含まれる正規ドメイン[⋆1]を取り除き，悪性ドメインリストの正確性を向上させる手法を提案している．正規ドメインを取り除くために，ユーザとそのユーザが名前解決を行うドメインでグラフを生成し，ノード間の類似度からドメインの悪性度を計算する．本稿では既知の悪性ドメインリストを用いてそのリストに含まれていない悪性ドメインを新たに発見する手法を提案しており，既知の悪性ドメインリストには正規ドメインが含まれていないことが前提となっている．この技術と組み合わせることで本提案手法の検出精度の向上が期待できる．

⋆1 ボットウィルスは外部との接続性を確認するために，正規ドメインの名前解決を行うことがある

3. ボット感染ユーザ検出のステップ

本稿で提案するボット感染ユーザの検出方法は，1)DNS 利用ユーザからボット感染ユーザを抽出，2) ボット感染ユーザが名前解決を行ったドメインから新たに悪性ドメインを抽出，3) 発見した悪性ドメインの名前解決を行うユーザの抽出の 3 ステップで構成される．以下，各ステップの詳細を述べる．

ステップ 1 まず，DNS クエリデータから既知の悪性ドメインリストに含まれるドメインの名前解決を行っているユーザを抽出し，DNS 利用ユーザをボット感染ユーザとそれ以外のユーザに分類する (図 1)．ここで，後者のユーザを**みなし非感染ユーザ**[⋆2]と呼ぶ．

ステップ 2 次に，分類したボット感染ユーザが名前解決を行ったドメインのうち，既知の悪性ドメイン以外のものを抽出し，その中から多くのボット感染ユーザが名前解決を行っているドメインを抽出する．これらのドメインは既知の悪性ドメインと頻繁に共起しているドメインであることから，新たな悪性ドメインと判定する（図 2)．

ステップ 3 最後に，ステップ 2 で新たに発見した悪性ドメインを用いて，そのドメインを名前解決を行うユーザを DNS クエリデータから抽出し，そのユーザをボット感染ユーザとする．ここで発見されるみなし非感染ユーザのうち，既知の悪性ドメインリストでは発見できなかったユーザが本提案手法により新たに発見できたボット感染ユーザである (図 3)．

4. 問い合わせドメインの共起関係を用いた悪性ドメインの抽出方法

本節では 3 節で述べたステップ 2 におけるドメインの悪性判定の詳細について述べる．

4.1 ドメインの悪性度

3 節のステップ 2 で行うドメインの悪性判定のために，異なる 2 つのドメインの共起関係を強さを表す**共起度** C を導入する．DNS 利用ユーザの集合を U とし，あるユーザ $u \in U$ が名前解決を行ったドメインの集合を D_u とし，異なる 2 つのドメイン d_i, d_j の共起度 $C(d_i, d_j)$ を以下のように定める．

$$C(d_i, d_j) = \frac{\left|\{u \mid d_i \in D_u \wedge d_j \in D_u\}\right|}{\left|\{u \mid d_i \in D_u \vee d_j \in D_u\}\right|} \quad (1)$$

⋆2 既知の悪性ドメインを問い合わせていないユーザの中にも，ボットウィルスに感染しているユーザが存在していることを前提としているためみなしとしている

Vol.2009-IOT-7 No.11
2009/10/9

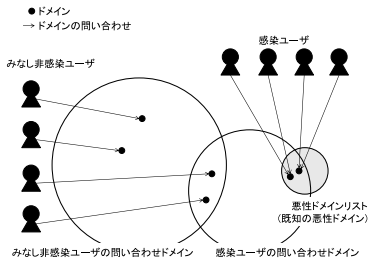

図 1 ステップ 1

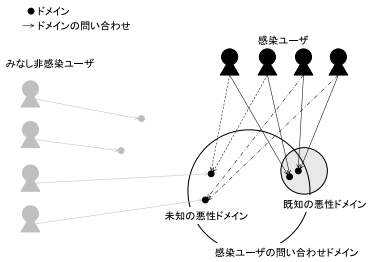

図 2 ステップ 2

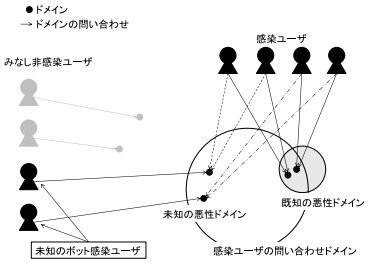

図 3 ステップ 3

式 1 はドメイン d_i または d_j を問い合わせるユーザのうち，両方のドメインを問い合わせるユーザの比を表わすものであり，値が大きいほど 2 つのドメインの関連性が強いと言える．

次に，式 1 を用いてあるドメイン d を悪性かどうか判定するための指標 $S(d)$ を導入する．これを，ドメイン d の**悪性度**と呼ぶ．既知の悪性ドメインの集合を D_M とし，悪性度 $S(d)$ を以下のように定める．

$$S(d) = \sum_{d_m \in D_M} C(d_m, d) \tag{2}$$

式 2 は既知の悪性ドメインと共起する全てのドメインとの共起度を足し合わせたもので，多くの悪性ドメインと共起したドメインほど悪性度が高くなる．

4.2 共起度を用いたボット感染ユーザ発見手法の問題点と解決策

ボット感染ユーザが名前解決を行うドメインには，悪性ドメインの他に実ユーザが WEB ブラウジングやメール送信などを行う際に名前解決を行う**正規ドメイン**が含まれる．ドメインの悪性判定をする上で，正規ドメインを悪性と判定する**誤判定**に注意しなければならない．

式 2 を用いて実験を行ったところ，誤判定を引き起こすドメインやユーザが存在することが分かった．ここではそれらの問題と，その解決策を述べる．

4.2.1 問 題 点

人気ドメインの問題 多くのユーザが名前解決を行うものとして，google や yahoo といった**人気ドメイン**がある．人気ドメインはボットウィルスに感染している端末を使用するユーザの多くが名前解決を行うドメインであるため，既知の悪性ドメインと共起するドメインとしての出現頻度が高くなる．これによって人気ドメインの悪性度が高くなり，悪性ドメインと誤判定されてしまう．

ヘビーユーザの問題 DNS を利用するユーザの中には，1 秒間に数百ドメインも問い合わせを行うようなユーザが存在する．このようなユーザを**ヘビーユーザ**と呼ぶ．ボットウィルスに感染しているヘビーユーザの存在を考えると，そのヘビーユーザが名前解決を行う大量のドメインが既知の悪性ドメインと共起することになる．大量のドメインの中に正規ドメインかつそのヘビーユーザしか問い合わせないものが存在すると，共起度は式 1 の性質から大きくなってしまう．これによってヘビーユーザが名前解決を行うドメインの悪性度が大きくなり，そのドメインが悪性ドメインと誤判定をされてしまう．

このような問題はボット感染ユーザの検知精度を低下させてしまうため，人気ドメインやヘビーユーザの問い合わせドメインの悪性度を下げ，誤判定を少なくする必要がある．

4.2.2 人気ドメインの影響除去

人気ドメインが悪性ドメインと判定されることを防ぐために，ドメイン d の悪性度 $S(d)$ をドメインの人気度合に応じて下げる必要がある．人気ドメインはボット感染ユーザ，みなし非感染ユーザの多くが名前解決を行うドメインである．一方で，悪性ドメインはボット感染ユーザの多くが名前解決を行うドメインであるが，みなし非感染ユーザが名前解決を行うことは少ない．すなわち，あるドメインの名前解決を行うみなし非感染ユーザ数によってドメイン d の人気度合を決定することができ，それを $W(d)$ として以下のように定める．ここで，ボット感染ユーザの集合を U_M，みなし非感染ユーザの集合を U_N である．

$$W(d) = \frac{\left|\{u \mid u \in U_M \wedge d \in D_u\}\right|}{\left|\{u \mid u \in U_M \wedge d \in D_u\}\right| + \left|\{u \mid u \in U_N \wedge d \in D_u\}\right|} \tag{3}$$

式 3 はドメインの人気度合に反比例する重みである．なお，これまでの調査結果[1] による

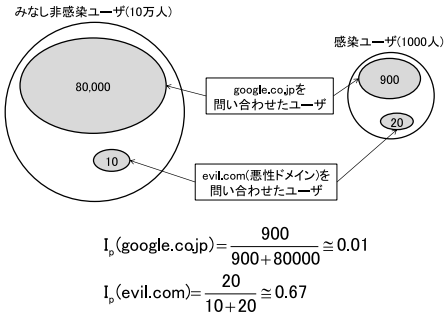

図 4 人気ドメインの除去

と，ボット感染ユーザは全体の 1% 未満である．このことから人気ドメインの名前解決を行うボット感染ユーザは，人気ドメインを問い合わせるみなし非感染ユーザと比較して非常に少ないので，人気ドメインの W の値は非常に小さくなる (図 4)．

式 3 を用いて新たな悪性度を以下のように定める．

$$S_w(d) = S(d) \times W(d) \tag{4}$$

これによってドメインの人気度合に応じた悪性度を算出することができる．

4.2.3 ヘビーユーザの影響除去

ヘビーユーザの問題は，既知の悪性ドメインと共起するドメインが，ボットウィルスが名前解決を行ったドメインなのか，実ユーザが大量に名前解決を行ったドイメンの中に偶発的に表れたものなのかを考慮していないことに起因する．この問題を解決するために，本稿では共起度をユーザが名前解決を行ったドメイン数で重み付けする手法を提案する．具体的にはドメイン d_i, d_j の新たな共起度 $C'(d_i, d_j)$ を以下のように定める．

$$C'(d_i, d_j) = \frac{\sum_{\{u \mid d_i \in D_u \wedge d_j \in D_u\}} 1/\left|D_u\right|}{\left|\{u \mid d_i \in D_u \vee d_j \in D_u\}\right|} \tag{5}$$

式 1 は共起度を 2 つのドメインの名前解決を行ったユーザ数，すなわち共起回数によって得ているため，ユーザが名前解決したドメイン数による共起のしやすさを考慮していなかったが，式 5 は共起回数をそのユーザが名前解決したドメイン数で割ることによって共起のしやすさを考慮したものとなっている．これは，共起度計算の際に 2 つのドメインの名前解決を行ったユーザが得点を持っており，ヘビーユーザほど持っている得点が少なくするよう重み付けしたものと考えることができる．

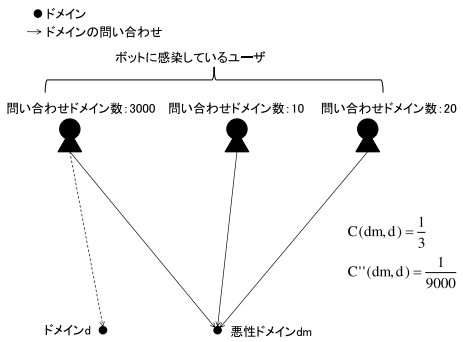

図 5 ヘビーユーザの影響除去

式 2 に式 5 を適用することによってヘビーユーザを考慮した悪性度を得ることができる．さらに，人気ドメインの重み付け $W(d)$ を用いることによって誤判定を引き起す要因を除去した悪性度の計算方法が得られる．この式を S' として，以下のように定める．

$$S'(d) = \sum_{d_m \in D_M} C'(d_m, d) \tag{6}$$

最終的に式 4 に式 3 で定義した $W(d)$ を掛けることで人気ドメインとヘビーユーザの問題を考慮した悪性度の計算手法が得られるこの式を S'_w として以下のように定める．

$$S'_w(d) = \left(\sum_{d_m \in D_M} C'(d_m, d) \right) \times W(d) \tag{7}$$

5. 評　　価

本稿では提案手法を実際の DNS トラフィックデータに適用して評価を行った．評価は 1 節で述べた仮説の検証と，4 節で提案した手法の有効性の検証を行った．

5.1 準　　備

2 節で述べたように，ボットウィルスの解析によって得られる悪性ドメインリストには，正規ドメインが含まれていることがある．それらの正規ドメインは実験結果に大きな影響を

表 1 悪性度上位 1% のドメインで検証用データに合致したドメインの割合

ヘビーユーザ ＼ 人気ドメイン	除去なし	除去あり
重み付けなし	23.0%	27.4%
重み付けあり	65.2%	91.2%

及ぼすため，本稿では手動で正規ドメインを取り除いている．

5.2 仮説の検証

1 節の仮説の正しさを確認するために，*n-fold Cross-Validation*[5] を用いた．これは既知の悪性ドメインを n 分割し，そのうちの k 個のデータを訓練用データ，$n-k$ 個を検証用データとするものである．

本稿では $n=10$，$k=1$ とし，10 通りの検証用データを用いて評価を行った．評価は訓練用データを適用して得られた悪性度上位 n% のドメインを抽出し，その中に検証データがどの程度含まれているかを調査した．検証結果を表 1 および図 6 に記す．図 6 の横軸は抽出した悪性度上位ドメインの割合を表し，縦軸は抽出したドメインの中に，どの程度検証データが含まれているかを表す．値は 10 回の実験の平均値である．特に，上位 1% を抽出したときの結果を表 1 に記す．

表 1 より，単純な共起度を用いた悪性度の計算手法 (式 2) では検証用データの 23% 程度しか発見することができなかったが，人気ドメインやヘビーユーザの問題を考慮した式 7 を用いると，検証用データの 91.2% を発見できていることが分かる．また，図 6 を見ると，提案手法によって検証用データに含まれるドメインの悪性度が高くなり，上位に表われていることが確認できる．

これらの結果により，ボットウィルスが名前解決を行う悪性ドメインの共起ドメインの中には既知の悪性ドメインリストに存在しない悪性ドメインが含まれており，人気ドメインやヘビーユーザが名前解決を行ったドメインを適切に除去することにより，それらが効率良く抽出できることが確認できた．

5.3 提案手法の有効性の検証

提案手法の有効性を評価するために，既知の悪性ドメイン全てを利用して得られた悪性度の高いドメインが真に悪性であるか，新たに発見した悪性ドメインを用いて未知のボット感染ユーザがどの程度発見できたかを評価した．

悪性度の高いドメインが，真に悪性ドメインであるかを判断するために，アンチウィルスソフトウェアベンダのウィルス情報を用いた．悪性度上位 100 個のドメインを評価したところ，真に悪性なドメインが 51%，正規ドメインが 25%，悪性か正規かが不明なものが 13% となった．このうち，悪性度上位 20 個のドメインの評価結果を表 2 に結果を記す．

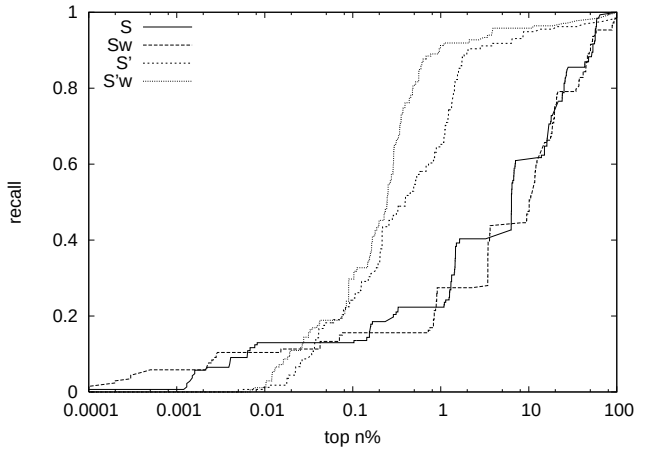

図 6 悪性度上位 n% のドメインで検証用データに合致したドメインの割合

この結果によって，本稿の提案手法によって既知の悪性ドメインリストに含まれる以外の悪性ドメインを発見できることが確認できた．

未知のボット感染者がどの程度発見できたかの評価として，上記の悪性度上位 100 個のドメインのうち真に悪性と認められたドメインと既知の悪性ドメインの名前解決を行うユーザを抽出し，既知の悪性ドメインリストのみを利用した場合と比較してボット感染ユーザがどの程度増加したかを調査した．

評価の結果，提案手法で新たに発見された悪性ドメインを用いると，既知の悪性ドメインリストのみを用いた場合と比較してボット感染ユーザ数が約 3% 増加し，本提案手法が未知のボット感染ユーザを発見できることが確認できた．

6. 今後の課題

前節の評価で本提案手法が未知の悪性ドメインおよびボット感染ユーザを発見できることを確認したが，新たに発見された悪性ドメインと比較してボット感染ユーザ数が少ないことが確認された．これは，新たに発見された悪性ドメインのなかに既知のボット感染ユーザの問い合わせドメインから抽出したものが存在し，未知のボット感染ユーザの発見に繋がらな

表 2 悪性度上位 20 個の評価

悪性度	ドメイン	評価
0.571	spy.nerashti.com	悪性
0.571	bla.bihsecurity.com	悪性
0.571	aaaaaaaaaaaaa.locop.net	悪性
0.500	icq-msg.com	不明
0.319	mail.tiktikz.com	悪性
0.300	x.zwned.com	悪性
0.300	evolutiontmz.sytes.net	不明
0.300	dcom.anxau.com	悪性
0.292	usa.lookin.at	不明
0.292	rewt.buyacaddi.com	悪性
0.250	unkn0wn	存在しないドメイン
0.250	google-analitucs.com/loader/	悪性
0.222	netspace.err0r.info	存在しないドメイン
0.203	win32.kernelupdate.info	悪性
0.203	free.systemupdates.biz	悪性
0.200	zjjdtc.3322.org	不明
0.200	ykln.3322.org	悪性
0.200	dr27.mcboo.com	悪性
0.189	china.alwaysproxy.info	悪性
0.167	home.najd.us	悪性

かったからである．また，本稿では悪性ドメイン判定のための悪性度上位 100 個のドメインを評価したが，最適な悪性判定のための閾値を自動的に生成するには至っていない．

今後の課題として，未知の悪性ドメインおよびボット感染ユーザの検知精度の向上がある．本稿では DNS トラフィックデータのうちユーザから DNS サーバへの問い合わせドメインに注目したが，DNS サーバからユーザへの応答，時系列分析など，他のパラメータに注目した手法を検討したい．また，悪性ドメイン判定の際に検知精度を高く，正規ドメインを悪性と判定する誤検知を小さくするような最適な閾値の決定手法が必要である．

7. ま と め

本稿では DNS トラフィックデータと既知の悪性ドメインリストを利用し，未知の悪性ドメインおよび未知のボット感染ユーザを検出する手法を提案した．

未知の悪性ドメインの発見には，ユーザが名前解決を行うドメインの共起関係に基づきスコリングをすることによって行う．

提案手法を評価した結果，既知の悪性ドメインリストに含まれない悪性ドメインを発見することができ，また既知の悪性ドメインリストのみでは発見できないボット感染ユーザを発見することができた．

今後，本提案手法をボット感染状況を把握するための監視ツールとして応用し，ボット感染ユーザの早期発見などネットワーク管理の一助としたい．

参 考 文 献

1) Cyber Clean Center: 総務省・経済産業省連携 ボット対策プロジェクト. http://www.ccc.go.jp/index.html.
2) Graham, P.: A plan for spam (2002). http://www.paulgraham.com/spam.html.
3) Ishibashi, K., Toyono, T. and Iwamura, M.: Improving accuracy of black domain list by using DNS query graph, *Proceedings of 1st Internet Workshop on Information Network Design* (2008).
4) Ishibashi, K., Toyono, T., Toyama, K., Ishino, M., Ohshima, H. and Mizukoshi, I.: Detecting Mass-Mailing Worm Infected Hosts by Mining DNS Traffic Data, *Proceedings of 1st Internet Workshop on Information Network Design* (2008).
5) Kohavi, R.: A Study of Corss-Validation and Bootstrap for Accuracy Estimation and Model Selection, *Proceedings of 40th International Joint Conference on Artificial Intelligence*, pp.1137–1143 (1995).
6) Kristoff, J.: Botnets, detection and mitigation: DNS-based techniques (2005). Information Security Day.
7) 竹森敬祐，藤長昌彦，西垣正勝：DNS ログに注目した詐称 IP 探索，情報処理学会マルチメディア通信と分散処理 (2008).
8) 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，情報処理学会 コンピュータセキュリティ研究会 (2006).