Vol.2016-CSEC-75 No.6
2016/12/1

エントロピーを用いた 初期侵入段階におけるRATの通信検知手法の考察

宇野真純[1,a)]　石井将大[2]　猪俣敦夫[1,3]　新井イスマイル[4]　藤川和利[4]

概要：Remote Access Trojan/Tool(RAT) とは標的型攻撃の初期侵入段階においてまず用いられる遠隔操作を可能にするツールである．標的型攻撃の検知においては，情報探索から端末制御段階までに RAT の通信を検知することが有用とされている．先行研究 [1] では，抽出された特徴が短期間の通信パケットであることから正常なアプリケーションとの区別が困難であることや，特定の通信プロトコルを使うことのみ想定した場合など環境に依存することがあるため，検知の条件を回避するための偽装が容易であること等の問題が存在する．本研究では，ある特定の通信プロトコルを用いるなどの制約された環境に依存せず，初期の侵入段階における RAT 通信の検知を目的とする．具体的には，先行研究で用いられた In/Out bound 通信のパケット数やバイト数などの複数の特徴に加え，通信パケットから新たにエントロピーを計算して特徴とした検知手法を提案する．エントロピーを用いることにより，限定された環境に依存しないなどの理由から限定した条件の回避による偽装が困難となる．さらに，本研究では RAT が確立した C&C サーバとの通信トラフィックのパケットの特徴より，攻撃者が行動を開始するまでの間の RAT 通信のエントロピーは小さくなると仮定し，検知においてエントロピーが示す情報が有用であることを示す．

キーワード：ネットワークセキュリティ, 侵入検出・検知, RAT, 標的型攻撃, エントロピー

A RAT detection method by using packet entropy on early intrusion stage

Masumi Uno[1,a)]　Masahiro Ishii[2]　Atsuo Inomata[1,3]　Ismail Arai[4]　Kazutoshi Fujikawa[4]

Abstract: Confidential information leaked accidentally by targetted attacks causes a serious social issue. In targeted attacks, Remote Access Trojan/tool (RAT) is mainly used. It is important to detect the RAT activity on intrusion stage to minimize damage by the attack. The detection of the RAT is getting more and more difficult with technological advance. Previous studies can not detect RAT which uses various kinds of protocols and they cannot detect advanced RAT. In this study, we aim to detect an early intrusion stage of RAT communication. This study uses packet entropy of the communication.

Keywords: Network Security, Intrusion Detection, RAT, APT, entropy

1. はじめに

特定の組織を狙った標的型攻撃による被害が深刻化し，社会問題となっている．標的型攻撃とは特定の組織を対象とし，機密情報や個人情報を狙った攻撃の総称である．特

[1] 奈良先端科学技術大学院大学情報科学研究科
Graduate School of information ,Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara
[2] 東京工業大学情報理工学院
School of Computing, Tokyo Institute of Technology , 2-12-1 Ookayama, Meguro-ku, Tokyo
[3] 東京電機大学
Tokyo Denki University, 5 Senju Asahi-cho, Adachi-ku, Tokyo
[4] 奈良先端科学技術大学院大学総合情報基盤センター
Information Initiative Center, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara
[a)] uno.masumi.uh1@is.naist.jp

徴として，通信が偽装されるため発覚されにくいこと，計画的であり長期的に潜伏されること，日々その攻撃手法が変化し続けていることが挙げられる．また，標的型攻撃は以下の複数の段階を踏んで行われる．

(1) 事前準備: 標的の情報収集．不正プログラムの準備
(2) 初期侵入: 標的型メール送信，不正プログラムの実行
(3) 端末制御: C&C サーバ間通信の確立，感染環境の確認
(4) 情報探索: 内部活動ツール送出，LAN 内情の探索
(5) 情報集約: 有益な情報の収集
(6) データ送出: 収集情報の入手

標的型攻撃の初期侵入段階において，Remorte Accsess Trojan/Tool （RAT）と呼ばれる遠隔操作ツールが用いられる．トレンドマイクロ社の報告では，トレンドマイクロ社のネットワーク監視対応においておよそ 4 社に 1 社の割合で RAT の通信が観測されている [2].

また，標的型攻撃の検知において，初期侵入から端末制御段階である侵入時活動までに RAT の通信を検知することが有用とされている [3]．寺田ら [4] の動的活動観測報告によると，RAT が被害者端末に侵入し実行されてから攻撃者が遠隔操作を開始するまでの時間は最短で 7 分，最長で 38 時間であった．

本研究では，ある特定の通信プロトコルを用いるなどの制約された環境に依存せず，初期の侵入段階における RAT 通信の検知を目的とし，Jang ら [1] で用いられた In/Out bound 通信のパケット数やバイト数などの複数の特徴に加え，通信パケットから新たにエントロピーを計算して特徴とする検知手法を提案する．特徴としてエントロピーを用いることにより，侵入される標的対象の環境に依存しないことや，検知条件を回避するための攻撃者による通信の偽装が困難となる．さらに，本研究では RAT が確立した C&C サーバとの通信トラフィックのパケットは送信間隔に偏りがあることから，攻撃者が行動を開始するまでの間の RAT 通信のエントロピーは小さくなると仮定し，実験を行い，RAT の検知においてエントロピーを特徴として用いることが有用であることを示した．さらにそのエントロピーを新たに特徴として用いて RAT の通信を検知する手法について考察を行う．

本研究は以下の通りに構成される．2 章では関連研究と本研究の位置づけについて説明し，3 章ではエントロピーを用いた特徴の抽出とその定義に関して述べ，比較実験を行い，考察を行った．また，4 章でエントロピーを用いた検知手法の考察を行い，5 章において今後の課題とまとめについて述べる．

2. 関連研究

2.1 RAT の検知手法に関して

RAT の検知方式は主としてネットワークベースのものとホストベースのものに分けられる．ネットワークベースの検知手法では RAT と正常なアプリケーションの通信特徴の違いに着眼し，ネットワークから得られるパケットなどの通信の情報を用いて検知を行う．それに対し，ホストベースの検知手法では，バイナリ解析や API コールの監視など，主にホスト上で取得されるプロセス情報を用いて検知を行う．それぞれの特徴としてホストベースのものは利用できる情報が多いという利点を持つ．しかし，その一方で情報を収集するために時間を要するという欠点が存在する．ネットワークベースのものは，利用できる情報は限られているが，リアルタイム性に優れ，早い段階で検知が可能である．

表 1 先行研究に用いられた特徴の比較

	packet num	packet byte	port num	その他
JAN ら [1]	◎	◎		in/out rate
山田ら [5]	◎	-	-	通信と操作を紐付
山内ら [6]	◎	◎	-	アクセス時間間隔
Zeng ら [7]	○	○	○	packefloe num
Li ら [9]	◎	◎	○	in/out packet rate

本研究では侵入時活動までに RAT の通信を検知することを目的としているため，時間を要するホストベースの検知手法は不向きだと考えられる．よってネットワークベースを採用する．

2.2 ネットワークベースの検知

ネットワークベースの検知手法において，既存研究は以下の通りに挙げられる．また**表 1** は既存研究に用いられた特徴をまとめたものである．

山田ら [5] の研究では，SMB の操作と通信の開始を結びつけ，特徴として検知に用いている．この手法では，攻撃の通信の特徴として，特定の プロトコル（SMB）が利用されることを前提と考え，そのプロトコルの特徴に強く依存しているため，異なるタイプの通信が用いられた場合に検知するのは困難である．また，検知に RAT 通信と内部攻撃通信の相関を用いているが，2 つの通信の間にダミーの操作や通信が行われると，相関が取れず検知が出来ない．

山内ら [6] は，In/Out bound 通信のパケット数の合計とデータサイズの合計，セッション時間，アクセス回数，アクセス時間間隔を特徴とし，SVM，ナイーブベイズ，ロジスティック回帰などを用いて HTTP 通信を行う botnet の C&C トラフィックの分類を行い評価している．この研究は HTTP 通信を行う場合のみに有効であり．他のプロトコルを用いた通信に対応が出来ない．

Zeng ら [7] の研究も同様にボットネットの検知である．ホストで得られる情報とネットワークから得られる情報より検知を行っており，他の検知手法と同様に，ネットワークベースの特徴として C&C トラフィックより得られるパケット数やパケットのバイト数を用いている．

また，エントロピーを用いた異常検知の先行研究として溝口ら [8] は人間と機械のデータ送信間隔に違いがあるこ

表 2 early stage 内の特徴の比較

	Pack Num	In Byte	Out Byte	In Pack	O/I Pac	OB/OP
adwind RAT	241	63	66	79	2.05	0.4
dropbox	9	57	78	4	2.035	11.4

とに着眼してボットの検知を行っている．対象としているボットは人間の挙動を模倣しない単純な挙動を取るボットであり，エントロピー用いて通信にスコアリングすることで，ボットの挙動が一定的であり通信にぶれがないためにエントロピーが低くなると仮定している．しかし RAT はボットと違い人間が操作する上挙動が一定的ではないため，この手法では RAT の検知は出来ない．

Li ら [9] の研究では RAT の検知のための MANTO というシステムを作成している．MANTO は学習フェーズと検知フェーズから成る．まず学習フェーズにおいてトレーニングセットを用いて悪性な通信の検知モデルを生成し，リアルなトラフィックのデータを検知フェーズで入力として特徴ベクトルを生成する．学習フェーズで作成した検知モデルとの類似度を K-平均法で測り，その通信が悪性かを判断している．MANTO は特徴ベクトルを生成するために In bound 通信と Out bound 通信の Byte 数の比率を用いているが，P2P 通信を行う RAT が HTTP や HTTPS 通信を行う RAT よりもその比率が異なるために検知できず，それが全体の検知率を下げている．

Jang ら [1] は検知に用いる情報の取得期間として新たに early stage の定義を行い，検知を行っている．early stage とは感染端末から SYN パケットが送信されてから次の通信パケット到着までにインターバルタイムが T 秒以上間が開くまでの間であり，T=1.0 を最適としている．しかし用いる特徴は early stage 内で観測されたパケットの情報のみであるため，検知できるものに限界があり誤検知を生む可能性が非常に高い．例として先行研究に用いられていない RAT と通常のアプリケーションの特徴比較を**表 2** に示す．先行研究で用いられた特徴だけでは正常なアプリケーションと RAT の通信パケットの差は少なく，互いを明確に判別するための特徴としては不十分である．

以上より，標的型攻撃における RAT 通信の検知では，以下の二点が求められる．

- 情報探索段階までに得られるデータで検知が可能であること
- 特定の挙動やプロトコルなどに依存せず RAT の亜種にも対応が可能であること

表 1 より，RAT の通信の検知において先行研究で用いられた共通の特徴である In/Out bound Packet の数，In/Out bound Packet のバイト数は検知に用いる特徴として有用であると推測できる．本研究では，それらの特徴に加えて，RAT の通信間隔の偏りをエントロピーを用いて数値化し，新たに特徴として用いる．

3. エントロピーによるパケット到着間隔の偏りの数値化

3.1 エントロピー

本研究では，パケットの到着時間の偏りを特徴と考え，値として表現するためにエントロピーを用いる．エントロピーとは情報理論の概念において情報の不確実さを表す尺度である．

事象の発生確率が高ければ高いほどその情報の持つ確実性は高いと言えるため，エントロピーは小さくなる．また逆に発生確率が低ければ多くの情報を含んでいるとされ，エントロピーは大きくなる．

計算にはシャノンエントロピーを用い，以下の式で定義される．

$$H(X) = \sum_{i=1}^{M} -p_i log_2 p_i \tag{1}$$

本研究において RAT による通信パケットの送信時間間隔の偏りをエントロピーを用いて表し，正常なアプリケーションと RAT の通信の差を明確にする．p_i を以下のように定義する．

S は 3 ウェイハンドシェイクの SYN パケットが到着してから s 秒間の間に得られたパケット数であり，T は区間 t 秒間内の Inbound パケットの個数である．

パケットキャプチャを開始した s 秒内の時間を t 秒間ずつ分割し，その中に含まれているパケットの個数を表している．つまり，p_i が大きいとは，到着したパケットが偏っているということであり，p_i が大きい場合，式より求められるエントロピーの計算結果 $H(X)$ は小さくなる．また，M は p_i の個数である．

$$p_i = \frac{T}{S} \tag{2}$$

T=区間幅 t 秒間内の Inbound Packet の個数
S=s 秒間に得られた Inbound Packet の個数

また In bound 通信の偏りを正しく抽出するためには適切な区間 t と適切な全体の幅 s を求める必要がある．

エントロピーを求めるために適した区間幅 t と全体の秒数 s を選択するため以下の実験を行った．RAT は Jan ら [1] の研究で用いられていない Darkcommet RAT，用いられていないかつ先行研究で使用される特徴では正常なアプリケーションと判別が困難な adwind RAT を用いる．正常なアプリケーションは従来の手法で Darkcommet RAT とは判別困難な DropBox，adwind RAT と判別困難な TeamViewer など複数個選択し，エントロピーの値の比較を行った．

3.2 実験

3.2.1 区間幅 t に関するエントロピー

エントロピーの算出に用いるための t に関して，適切な値を求める．切り取る区間幅 t を変更し，T の値がエントロピーにどのような変化を与えるか実験を行った．

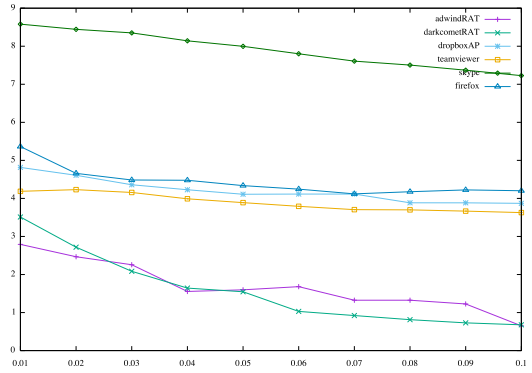

図 1 区間幅 t を変更した際のエントロピーの変化

表 3 区間幅 t を変更した計算結果 $H(X)$ の比較

	t=0.01	t=0.02	t=0.04	t=0.06	t=0.08	t=0.1
adwind RAT	2.791726	2.46702455	1.5549918	1.6811611	1.3240546	0.6509344
DarkCommet RAT	3.511217	2.7178983	1.6432400	1.0297571	0.81291087	0.67744
DropBox Apl	4.8148198	4.6076511	4.2282167	4.11427573	3.8852005	3.8702589
TeamViewer	4.1866836	4.2305954	3.9896522	3.7921732	3.69944	3.6277628
Skype	8.58060	8.44404182	8.142642	7.801937	7.50486135	7.2272798
Firefox(Browser)	5.36339	4.6586920	4.47690665	4.244126	4.175846	4.201838

表 3 は s=30(sec) であり，区間幅 t を変化させたときのエントロピーの比較表である．区間幅 t の値が大きくなればなるほど全体的にエントロピーが小さくなる傾向が確認できる．また，は??はその値を線グラフで表現している．t の値が増加した場合, エントロピーの値は減少していくが，RAT は正常なアプリケーションよりも値が低くエントロピーの減少率が高いことが理解できる．

これは本手法では区間幅 t に存在するパケットが 0 となる場合，エントロピーの計算において定義上 0 は計算されないためであり，また RAT の t の区間幅を狭めると，大きな偏りを分割してしまい，突出した値が取れず通信間隔はほぼ等間隔とみなされるためである．ゆえに Keep-alive 通信までのインターバルタイムがエントロピーの計算結果に与える影響も少なくなり，より偏りを拾えない．

3.2.2 全体の期間 s に関するエントロピー

エントロピーの算出に用いる s の適切な値を求める．全体の秒数 s に関して，S の値の変化がエントロピーにどのような変化を与えるか実験を行った．

表 4 は s(sec) の値を変化させた時，エントロピーの値にどのような変化を与えるかをまとめたものである．また，は図 2 より，t=0.05 として s の値を小さくした時，特に感染端末に潜入した RAT が攻撃者の C&C サーバへ 3 ウェイハンドシェイクの SYN パケット送信後から攻撃者との

表 4 全体のパケット数 S を変更した計算結果 $H(X)$ の比較

	s=10	s=20	s=30	s=40
adwind RAT	2.64519759	2.79172633	2.7917263	3.2134908
Dark commet RAT	0.0640815368	0.22090394	3.51121736	3.62061607
DropBox Apl	4.49958529	4.78246853	4.81481982	5.52345308
TeamViewer	3.61966927	3.79710468	4.14683605	4.01576184
Skype	6.67910306	7.77506885	8.58060862	9.17483767
Firefox(Browser)	4.3856628	4.66839673	5.36339555	4.89764141

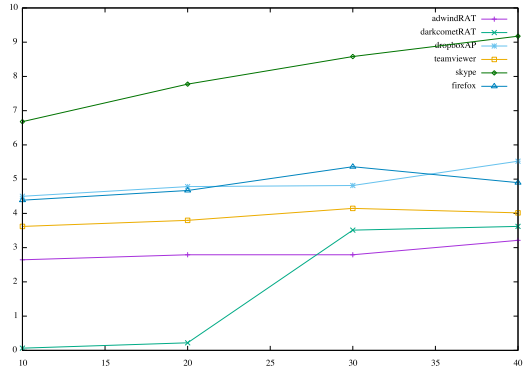

図 2 パケット採取時間 S(sec) を変化した際のエントロピーの変化

ネットワーク確立までの通信パケット数が少ない RAT は傾りが大きく，値が極端に小さくなる傾向がある．これは Keep-alive 通信による定期的な通信のパケットがエントロピーの計算に含まれず，データの偏りが突出した結果だと考察できる．逆に，S の値が大きくなった場合 RAT と正常なアプリケーションの値の差が少なくなっている．これは s を長く取るほど Keep-alive 通信が影響を与え，値の偏りを小さくしているからである．

よって s は Keep-alive 通信が行われるまでに設定しなければ，検知に適したエントロピーの値を取れない．また，本研究は初期潜入から端末制御段階までの検知を目的としているため，s は長期間であってはいけない．

3.3 結果考察

3.21 の実験結果より適切な T の値は t=0.1 となった．区間幅が狭い場合では，偏った通信を分割してしまうため，偏りを上手くエントロピーで表現できない．一方で 3.22 の実験結果より S の適切な値は s=10 となった．s=40 の時の結果から，RAT による Keep-alive 通信のパケットを含むパケットを s 秒内で拾ってしまった場合，値がその影響を大きく受け，パケットの偏りが取れなくなったと考察できる．そのため，s の値は Keep-alive 通信が起こるまでの間が有用である．

RAT はパケットの到着時間間隔に特徴を持つ．例として Skype と adwindRAT のパケットの総出力グラフをは図 34 に示す．RAT は短時間で C&C サーバとの通信を確立し，インターバルタイムを置いてから攻撃者の司令を待つ間に Keep-alive 通信を行う．攻撃者の C&C サーバとのネットワーク確立部分のパケットの少なさとインターバルタイム，Keep-alive 通信におけるそれぞれから為る inbound 通

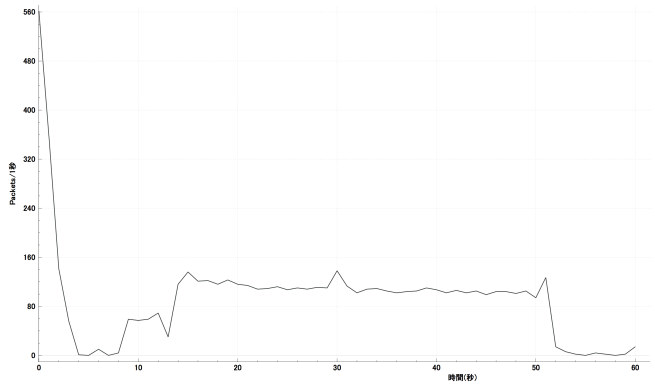

図 3 Skype のパケット出力

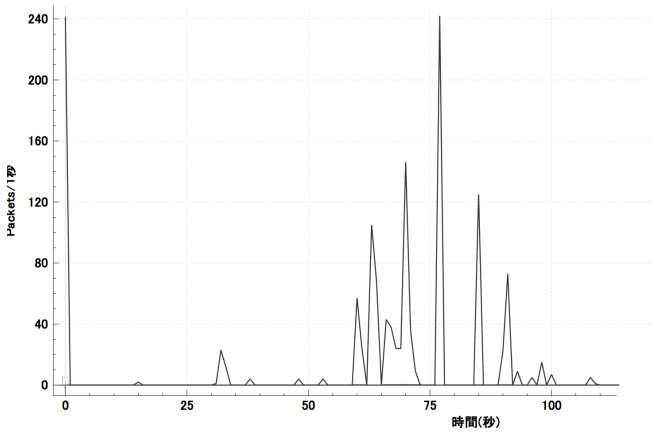

図 4 adwind RAT のパケット出力

信の偏りが RAT のエントロピーが小さくなる原因だと考察できる．

従来の検知方式では，adwind RAT のような C&C サーバとの通信を確立までの間に通信を行うパケット数の多い RAT の通信と通常のアプリケーションとの判断は困難であった．しかし，エントロピーを用いて通信の偏りを算出することにより，その違いを明確にすることが可能となった．

以上より，適切な区間幅 t と適切な全体の秒数 s を用いて，ネットワーク確立開始から Keep-alive 通信までを含めた値でエントロピーを用いることが有用であるといえる．また，適切な区間を取ることが出来なくても t の値を増加させることにより，エントロピーの減少率を RAT の特徴とすることが可能である．

4. エントロピーを特徴として用いた検知方式の考察

Jan ら [1] の研究に習い，入力を通信パケットとし，3 ウェイハンドシェイク開始時の SYN パケットが到着してからインターバルタイムが 1 秒以下になるまでの間に RAT の通信より特徴を抽出する．またそれとは別に同じ入力データから s 秒以内のパケットよりエントロピーを計算する．計算した値を教師データとし，教師あり機械学習アルゴリズムを用いて，通信を RAT のものであるかそうではないかを判別する．

本手法に用いる特徴とは，先行研究で述べた In/Out bound 通信のパケット数，バイト数や，Out bound パケットの数で byte 数を割ったものなどの特徴に加えて，エントロピーの値とその変化率を用いる．それにより，先行研究よりもエントロピーを算出するための時間は必要だが，侵入時活動段階以内で，検知ができなかったネットワーク確立までにパケット数が多い RAT の通信にも対応し，またその反対にパケット数が少ない通常のアプリケーションとも相互に RAT との通信の判断が可能となる．また，新たにエントロピーによる特徴を利用することによって，新たに誤検知を生む可能性もあるため，更なる考察が必要である．

5. まとめ

標的型攻撃における RAT の通信の検知については，情報探索から端末制御が行われる侵入時活動までに検知を行うことが重要である．しかし従来の検知方式では，特定のプロトコルや動作に依存していること，特徴抽出に用いる情報が少ないために誤検知率が高くなることや，また，情報は多いが侵入時活動段階までに必要なデータが揃えられないことなどの問題を指摘した．この課題に対し，通信パケット到着時間間隔の偏りをエントロピーを用いて数値化することにより，先行研究で検知が出来ない RAT と正常なアプリケーションとの判断が可能となった．そしてエントロピーを特徴のひとつとして用いることが有用であることを示した．

今後の課題として検体の増加や，機械学習アルゴリズムに関する考察，エントロピーを特徴として用いた RAT の通信検知手法について実際にシステムを構築して検証を行うこと，システムの評価などが挙げられる．

参考文献

[1] Dan Jan and Kazunari Omote: *A RAT Detection Method Based on Network Behavior of the Communication' s Early stage*, "IEICE TRANS Vol.E99-A NO1 January 2016 p.145-153", 2016.
[2] 特定非営利活動法人 日本セキュリティ監査協会：APT による攻撃対策と情報セキュリティ監査研究会 APT 対策入門: 新型サイバー攻撃の検知と対応 (2012).
[3] トレンドマイクロ株式会社：国内標的型サイバー攻撃分析レポート 2016 年版 状況と目的に応じて攻撃を変化させる攻撃者 (2016).
[4] 寺田真敏, 堀健太郎, 成島佳孝, 吉野龍平, 萩原健太: 研究用データセット「動的活動観測 2015」, Computer Security Symposium 2015.
[5] 山田正弘, 森永 正信, 海野 由紀, 鳥居 悟, 武仲 正彦:. 組織

内ネットワークにおける標的型攻撃の諜報活動検知方式, The 31st Symposium on Cryptography and Information Security, 2014.
[6] 山内一将, 川本淳平, 堀良彰, 櫻井幸一. 機械学習を用いたセッション分類による C&C トラフィック抽出, The 31st Symposium on Cryptography and Information Security, 2014.
[7] Yuanyuan Zeng and Xin Hu and Kang G. Shin: *Detection of Botnets Using Combined Host-and Network-Level Information*, IEEEIIFIP International Conference on Dependable Systems & Networks(DNS), 2010.
[8] 溝口 誠一郎, 堀良彰, 櫻井幸一: エントロピーを用いた機械的特徴のスコアリングとボット検知への応用, 平成 23 年度 電気関係学会九州支部連合大会, p.654-655, 2012.
[9] Shicong Li and Xiaochun Yun: *A General Framework of Trojan Communication Detection Based on Network Traces*" *Architecture, and Storage*, IEEE Seventh International Conference on Networking, 2014.