社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2008－CSEC－42（47）
2008／7／25

コネクション解析によるP2P通信端末検知手法

重本 倫宏† 大河内 一弥† 寺田 真敏†

†(株)日立製作所 システム開発研究所

〒212-8567 神奈川県川崎市幸区鹿島田890 日立システムプラザ新川崎

あらまし 近年，Winny，Share等のP2Pファイル交換ソフトによる情報漏洩の多発や，P2P通信によるトラヒックの圧迫が大きな問題となっている．このような問題を解決するためには，ネットワーク上のP2P端末を検知することが重要な課題となる．本稿では，コネクションを解析することにより，P2Pファイル交換ソフトの種別によらず，P2P通信を行っている端末を検知する方式を提案する．さらに，実環境を用いた実験を行い，提案手法が有効に動作する条件について考察する．

キーワード P2P，コネクション解析，検知

P2P Node Detection method based on Connection Analysis

Tomohiro Shigemoto† Kazuya Okochi† and Masato Terada†

† Systems Development Lab., Hitachi Ltd.

Hitachi System Plaza Shin-Kawasaki 890 Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa, 212-8567 Japan

Abstract Recently, increasing number of information leaks by viruses in the P2P network are becoming big problem. And P2P file sharing also accounts for an astonishing volume of current Internet traffic. Therefore, this research aims to detect terminals of P2P file sharing application from Internet traffic. In this paper, we propose a P2P Node Detection method based on two types of Connection Analysis: "Established connection ratio" and "Distributed port ratio". In addition, we also give experiments for finding suitable threshold of the proposed method.

Keyword P2P, Connection Analysis, Detection

1. はじめに

近年，Winny，Share等のP2Pファイル交換ソフトによる情報漏洩が多発し，未だ鎮静化の兆しが見えない状況にある．情報漏洩事故を起こした場合，経営トップの監督責任が問われることも予想され，実際に訴訟を起こされた事例も出始めている．さらに，Winnyなどに代表されるP2Pファイル交換ソフトは，巨大なP2Pネットワークを構成して音楽データや画像データなどサイズの大きなファイルのやり取りに利用されており，上記情報漏洩問題だけでなく，ISPネットワークリソースの圧迫といった問題も引き起こしている．

これらの問題を解決するために，ネットワークを流れるトラヒックから，P2P通信を検知する多くの研究がなされてきた[1-5]．例えば，P2Pファイル交換ソフトの通信を検知し，通信をブロックするOne Point Wall[1]などの製品がいくつか存在している．これらの製品の多くは，パケットの中身を解析し，個々のP2Pファイル交換ソフトの通信に表れる文字列やビットパターンが含まれているかどうかを確認することで検知を行っている．しかし，最近のP2Pファイル交換ソフトにはパケットの一部が暗号化されているものもあり，暗号解読が必要となる場合も多く，新たなP2Pファイル交換ソフトに対応するためには，開発コストや時間を要する．そのため，パケットの中身を解析することなしに，P2P通信の特徴からP2P通信の検知を行う方法が考案されてきた．文献[2]では，クライアント／サーバ関係に着目し，P2P通信を特定する方法について提案している．文献[3]では，ノード間のアクセス関係をグラフ化し，その直径の大きさからP2P通信の特定を行っている．文献[4]では，NetFlowやsFlowのフロー情報を解析し，P2Pファイル交換ソフトのトラヒックを異常トラヒックとして検知する手法を提案している．文献[5]では，トラヒックフローの統計値よりP2Pファイル交換ソフトのトラヒックを特定する方法を提案している．しかし，これらのトラヒックの特徴を用いた検知手法では，P2Pファイル交換ソフトが稼働してい

るホスト上で，他のアプリケーションを利用していた場合に，そのアプリケーションが行う通信によって，P2P 通信の検知精度が低下してしまうという問題も指摘されている．

本稿では，P2P ファイル交換ソフトと他のアプリケーションが同時に利用されている場合にも，パケットの中身を解析することなしに，P2P 通信を行っている端末を特定する手法を提案する．また，実環境を用いた実験を行い，提案手法が有効に作用する閾値について考察する．

2. P2P 通信端末特定手法

本章では，P2P ファイル交換ソフトと同時に他のアプリケーションを併用している場合にも P2P 通信端末を特定する手法を提案する．

2.1. P2P 通信の特徴

P2P 通信を行っている端末を特定するにあたり，P2P 通信のみのトラヒックの解析を行った．以下に P2P 通信の特徴を示す．

(1) コネクション確立成功割合

TCP 通信を行う際には，TCP コネクションの確立を行い，そのコネクション上で実際のデータを送受信する．TCP コネクションを確立するためには，接続要求元から相手先へ確立要求（SYN パケット：SYN フラグを ON にしたパケット）が送信される．確立要求を受け取った相手先は，それに対する応答（SYNACK パケット：SYN フラグと ACK フラグを ON にしたパケット）を返す．さらに，接続要求元がそれに対する応答を送信することで，コネクションが確立される．P2P ファイル交換ソフトは，利用者の端末同士で直接通信を行っている．利用者の端末は，サーバのように常時起動されているわけではなく，電源が落とされている場合も少なくない．このため，確立要求を送信した先の電源が落とされていれば，応答が返ってこず，P2P ファイル交換ソフト利用者の通信は，Web アクセスなどの通信に比べコネクション確立の成功割合が低くなる傾向がある（図 1）．コネクション確立の成功割合は以下の式により求められる．

$$\text{コネクション確立成功割合} = \frac{SYNACK\text{パケット受信数}}{SYN\text{パケット送信数}}$$

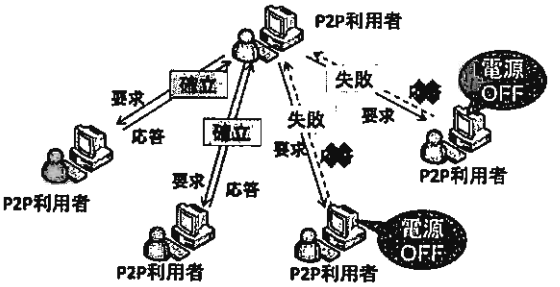

図 1 P2P 利用者の確立成功割合

(2) 利用ポートの分布割合

P2P ファイル交換ソフトは，ISP の帯域制限や FW による通信遮断を迂回するために，起動時にランダムに待ち受けポートの設定を行う．このため，P2P ファイル交換ソフト利用者の通信は，Web アクセスなどの通信に比べ利用ポート（宛先ポート）が広範囲に分布する傾向がある（図 2）．利用ポートの分散割合は以下の式により求められる．

$$\text{ポート分散割合} = \frac{\text{宛先ポートが異なる}SYN\text{パケット送信数}}{SYN\text{パケット送信数}}$$

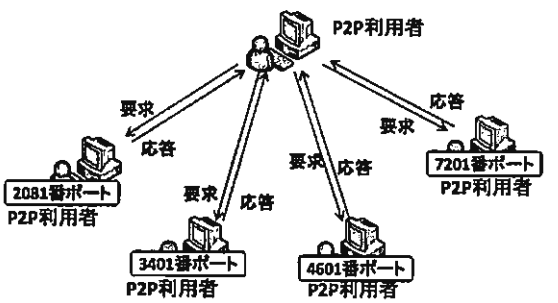

図 2 P2P 利用者のポート分布

2.2. 他のアプリケーションを併用した際の特徴

P2P 通信と同時に他のアプリケーションを利用している場合には，他のアプリケーションが行う通信の影響によって，2.1 節で述べた P2P 通信の特徴が埋もれてしまう場合がある．

実際に，P2P ファイル交換ソフトを用いて P2P 通信を行っているトラヒックと，P2P ファイル交換ソフトと他のアプリケーションを同時に利用したトラヒックを取得し，2.1 節で述べたコネクション確立成功割合と利用ポート分布の特徴について調査を行った．P2P 通信を行う P2P ファイル交換ソフトとして，現在国内において最も利用者が多い Winny を用いた[6]．また，他のアプリケーションとして Web アクセスを用いた．

図 3，4 に Winny のトラヒックと，Winny と Web ア

クセスを併用したトラヒック，及び WEB アクセスのみのトラヒックのコネクション確立成功割合と利用ポート分散割合を示す．観測時間 10 分間のテストデータを 10 セット用意した．

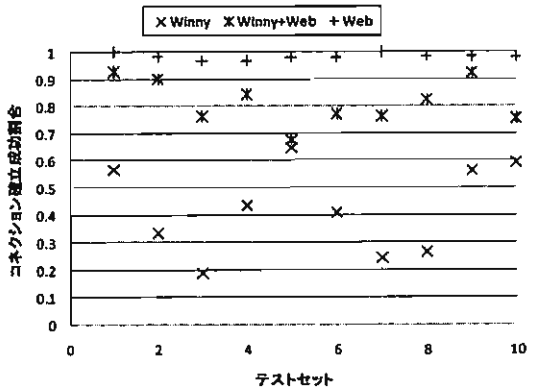

図 3 コネクション確立成功割合

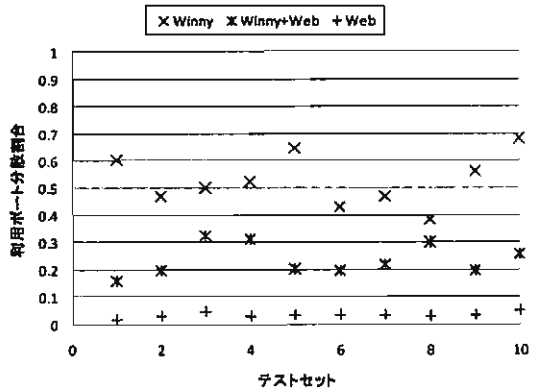

図 4 利用ポート分散割合

図 3，4 より，Winny のみによる通信を行っている場合はコネクション確立成功割合が低く，利用ポート分散割合が高いことが分かる．これに対して，Winny と Web アクセスを同時に行っている場合には，コネクション確立成功割合が高くなり，利用ポート分散割合が低くなる．このように，Web アクセスを併用している場合には，P2P 通信のみの場合とは異なり，Web アクセスのみの場合に似た特徴を示していることが分かる．このことは，確立成功割合の特徴を用いた場合，或いは利用ポート分散割合の特徴を用いた場合でも，他のアプリケーションを併用している P2P 通信端末を特定することは難しいということを意味している．

2.3. 提案手法の概要

2.2 節で述べたように，他のアプリケーションを併用した場合には，P2P 通信のみの場合とは異なる特徴を示すため，確立成功割合や利用ポート分散割合で P2P 通信端末を精度よく特定するのは困難である．

しかし，コネクション確立成功割合と利用ポート分散割合の特徴を組み合わせて判断することによって，他のアプリケーションのみを用いた場合と，他のアプリケーションを併用した場合を区別できる．

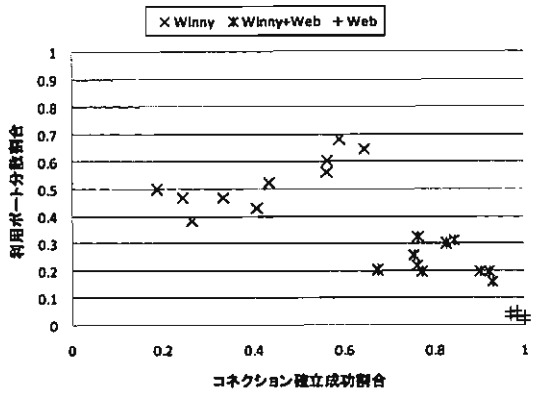

図 5 確立成功割合とポート分散割合

図 5 は Winny による通信のみのトラヒックと，Winny と Web アクセスを併用した際のトラヒック，Web アクセスのみのトラヒックそれぞれの確立成功割合と利用ポート分散割合を示した図である．

図 5 に示す通り，Winny と Web アクセスを併用した場合でも，Web アクセスのみの場合と比べ，コネクション確立成功割合は低く，かつ利用ポート分散割合が高いという特徴があることがわかる．

提案手法は，上記の特徴を利用し，P2P 通信端末を特定する．図 6 に P2P 通信端末特定システムの概要図を示す．

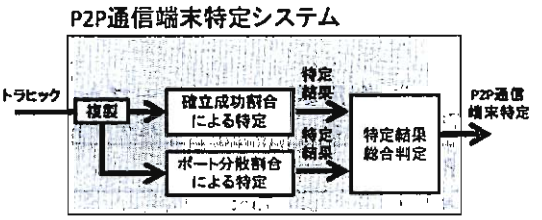

図 6 P2P 通信端末特定システム

2.4. 提案手法の実現方法

提案手法では，コネクション確立成功割合を求めるために，端末ごとの SYN パケット送信数及び，SYNACK パケット受信数を計数し，コネクション確立テーブルに格納する．例えば，図 7 に示すように P2P 利用者 A，B，C が P2P ファイル交換ソフトを利用していたとする．利用者 C がオフラインになったとしても，利用者 C が過去に利用していたという情報から，利用者 A，B は利用者 C への接続を試みる．しかし，利用者 C はオフラインのため確立要求に対する応答を返さない．

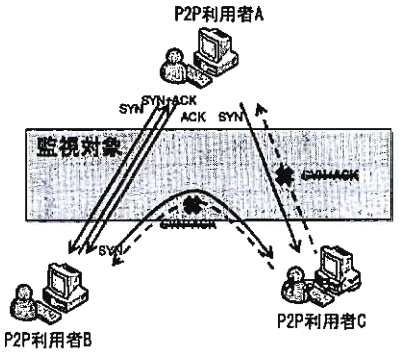

図 7 P2P 利用者の通信

図 7 に示した通信が行われた場合の分析装置におけるコネクション確立テーブルの状態を表 1 に示す．

表 1 コネクション確立テーブル

端末	SYN パケット送信数	SYNACK パケット受信数
P2P 利用者 A	2	1
P2P 利用者 B	1	0

提案手法では，利用ポート分散割合を求めるために，利用ポート分散テーブルも用いる．利用ポート分散テーブルでは，SYN パケットを検知するたびに，端末情報と，利用ポートを格納する．利用ポート分散テーブルの例を表 2 に示す．

表 2 利用ポート分散テーブル

端末	宛先ポート番号	SYN パケット送信数
P2P 利用者 A	80	45
	5431	5
P2P 利用者 B	2375	2

2.5. 検知アルゴリズム

提案手法の検知アルゴリズムを以下に示す．

1. 基幹部のルータを流れるパケットの中から SYN フラグが ON になっているパケットのみを分析装置へとコピーする．
2. パケット中の ACK フラグが ON になっていれば，宛先 IP の SYNACK パケット受信数を 1 増やす．ACK フラグが ON になっていなければ，送信元 IP の SYN パケット送信数を 1 増やし，宛先ポート番号を記録する．
3. 一定時間経過後，コネクション確立成功割合と利用ポート分散割合を求める．例えば，表 1 のコネクション確立テーブルから，利用者 A のコネクション確立成功割合は 0.5，利用者 B のコネクション確立成功割合は 0 となる．また，表 2 の利用ポート分散テーブルから，利用者 A の利用ポート分散割合は 0.04，利用者 B の利用ポート分散割合は 0.5 となる．
4. コネクション確立応答の割合がある閾値以下の場合，或いは，利用ポート分散割合がある閾値以上の場合に，その端末を P2P 通信端末として判定する（閾値については 3 章で検討する）．

3. 閾値の検討

本章では，実環境を用いた実験を行い，提案手法が有効に動作する閾値について検討した結果を述べる．

図 8 に，Winny による通信のみのトラヒックと，Winny と Web アクセスを併用した際のトラヒック，Web アクセスのみのトラヒックの確立成功割合と利用ポート分散割合を示した図である．10 分間の通信データが含まれているテストデータを，それぞれ 20 セット用いた．

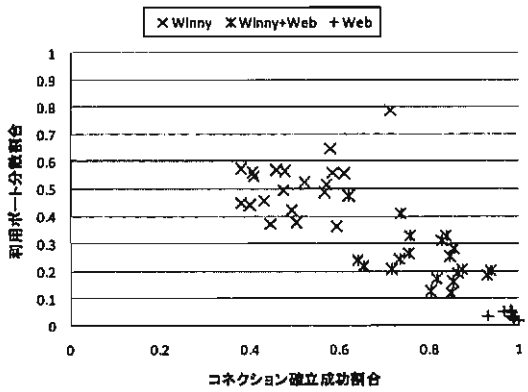

図 8 Winny と Web アクセス

図 8 より，Winny のみのトラヒック，Winny と Web アクセスを用いたトラヒックでは，Web アクセスと比べて，コネクション確立成功割合は低く，利用ポート

分散割合は高くなっていることが分かる．

汎用的な P2P 通信端末の特定を行うため，Share を用いた場合の確立成功割合と利用ポート分散割合を図 9 に示す．

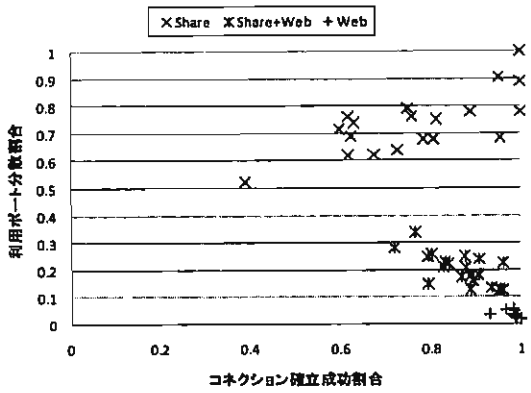

図 9 Share と Web アクセス

Share では，コネクション確立成功割合が高いデータがいくつか存在した．しかし，利用ポート分散割合は Web アクセスと比べて値が高くなり，P2P 通信端末として特定することが可能となる．

図 10 に，Cabos を用いた場合の確立成功割合と利用ポート分散割合を示す．ただし，10 分間の通信データが含まれているテストデータを，それぞれ 10 セット用いた．

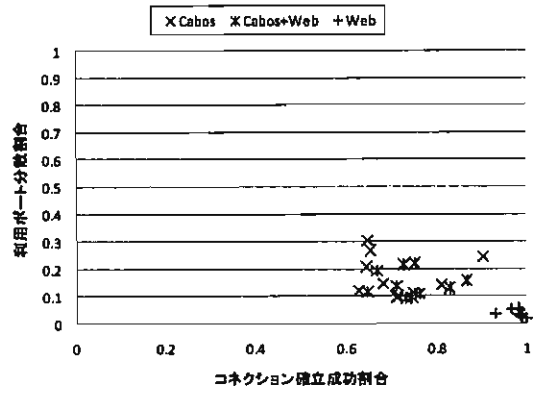

図 10 Cabos と Web アクセス

図 10 より，Cabos については，利用ポート分散が低くなるデータが存在した．しかし，コネクション確立成功割合は Web アクセスと比べて値が低くなり，P2P 通信端末として特定することが可能となる．

以上の評価実験により，本提案手法を用いて，コネクション確立成功割合が 0.9 以下であるか，利用ポート分散割合が 0.1 以上である場合にその通信を行っている端末を P2P 通信端末として判断すれば，本評価実験に用いた P2P ファイル交換ソフトを利用していた端末を全て検知できる（図 11）．

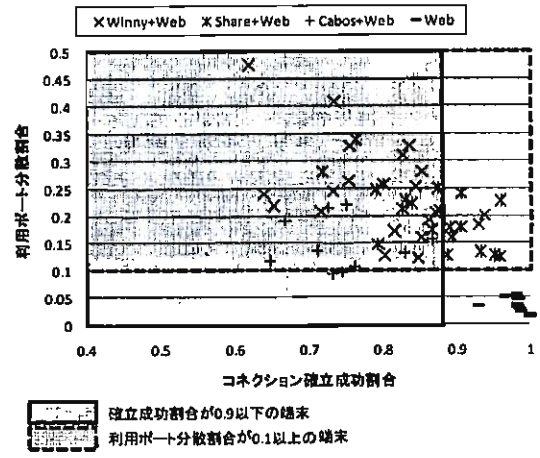

図 11 P2P 通信端末特定の閾値

4. 考察

(1) P2P 通信端末特定の閾値

評価実験において，P2P 通信を行っているかどうかの判断に，コネクション確立成功割合が 0.9 以下であるか，利用ポート分散割合が 0.1 以上である場合を用いた．しかし，P2P ファイル交換ソフトの種類によって閾値を変化させる方法もある．例えば Winny のみの通信の場合には，コネクション確立成功割合が 0.8 以下であり，また，Share のみの通信であれば，利用ポート分散割合が 0.5 以上となっている．適切な閾値を選択することで，P2P ファイル交換ソフトの種類まで判別できると考えられる．

(2) P2P 通信端末判定期間

評価実験においては，10 分間トラヒックを監視したのち P2P 通信端末の判定を行った．しかし，Share については，10 分間に送信された SYN パケット数が少ないテストセットが存在したため，コネクション確立応答割合が高い値となってしまった．P2P 通信端末の判定を一定時間に行うのではなく，一定量 SYN パケット送信数が溜まった場合に判定するなどの手法をとることにより，検知精度が向上すると考えられる．

(3) デフォルト待ち受けポート

Cabos については利用ポート分散割合が他の P2P フ

ファイル交換ソフトに比べて低くなる傾向があった．これは，Cabos のデフォルト（初期値）の待ち受けポートが 6346 に設定されており，待ち受けポートの設定を変えずにファイル交換を行っている利用者が多いためだと考えられる．そのため，デフォルトの待ち受けポートが設定されている P2P ファイル交換ソフトについては，待ち受けポート番号と P2P ファイル交換ソフトの対応表を用いることで，P2P ファイル交換ソフトの種別まで判別できると考えられる．

(4) トラヒック監視場所

提案手法を用いてトラヒックを監視する場所について考察する．図 12 に個人の PC がプロバイダなどを介してインターネットに接続する様子を模式的に表した図を示す．

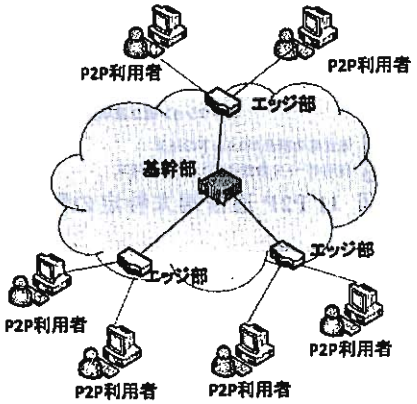

図 12 P2P 通信の監視対象

監視を行う場所については，P2P ファイル交換ソフトを稼働させている個人の端末，個人の端末と ISP の間に置かれるネットワークエッジ部のルータ，IX などのネットワーク基幹部のルータが考えられる．個人の端末から送信されるデータは，エッジ部または基幹部のルータを介して受信先へ届けられる．トラヒックの監視に要する装置の台数は，ネットワークの基幹部に近づくに従い少なくなる．また，P2P ファイル交換ソフトは P2P ネットワーク上の多数の端末と通信を行う特性上，ネットワーク基幹部の通信には多くの P2P 通信端末の通信が含まれている可能性が高い．

以上のことから，ネットワーク基幹部を監視することで効率よく P2P 通信端末の特定を行うことができる．また，提案手法では，SYN フラグが ON になっているパケットの情報のみから，P2P 通信端末の特定を行う．このため，トラヒックが集中するネットワーク基幹部であっても適用可能だと考えられる．

5. おわりに

本稿では，ファイル交換ソフトと他のアプリケーションを同時に利用している際にも，P2P 通信端末を検知する手法を提案した．また，提案手法をいくつかの P2P ファイル交換ソフトに適用した結果，P2P ファイル交換ソフト種別に依存することなく P2P 通信端末を特定することができた．

今回の評価実験では，P2P ファイル交換ソフト以外の通信として Web アクセスを用いた．しかしながら，回線環境，併用するアプリケーションの種類などによって適切な閾値は変化する．今後は，Web アクセスの他にも，メールサーバとの通信や，FTP サーバとの通信など様々なアプリケーションが同時に利用されている状態での評価を行っていく．さらに，他の特徴量を併用することによって検知精度を向上させていくことを検討する．

謝辞

本研究は総務省から受託した「ネットワークを通じた情報流出の検知及び漏出情報の自動流通停止のための技術開発」の成果の一部です．本研究を進めるにあたって有益な助言と協力を頂いた関係者各位に深く感謝いたします．

文　献

[1] One Point Wall, http://www.onepointwall.jp/

[2] 大坐畠智，川島幸之助，“クライアント／サーバ関係に着目したピュア P2P アプリケーショントラヒック特定方式と評価，”情報処理学会論文誌 Vol49，No.2 pp.988-998, Feb.2008.

[3] Constantinou, F. and Mavrommatis, P. “Identifying Known and Unknown Peer-to-Peer Traffic,” Proc. 5th IEEE International Symposium on Network Computing and Applications, pp. 93-102, 2006.

[4] 藤井聖，中村豊，藤川和利，砂原秀樹，“通信先ホスト数の変化に注目した異常トラヒック自動検出手法の提案と評価，”電子情報通信学会論文誌 Vol.J88-B，No.10 pp.1922-1933, 2005.

[5] 金子広幸，大坐畠智，荻原洋一，寺田松昭，川島幸之助，“ピュア P2P 型アプリケーショントラヒック特定法とトラヒック特性解析，”IEICE NS 研究会，NS2004-5，信学技報，Vol. 104，No. 17，pp. 17-20, 2004.

[6] 社団法人コンピュータソフトウェア著作権協会，利用実態のアンケート調査，クローリング調査の結果（2007 年 12 月）
http://www2.accsjp.or.jp/news/release071221.html
