# 相互通信に着目した暗号化 P2P トラヒックの弁別手法

鈴木将史[†]　阿部洋丈[††]　岡部正幸[†††]
梅村恭司[†]

P2P ファイル共有アプリケーションによる情報漏洩や著作権侵害，一部のヘビーユーザによるトラヒック占有は大きな問題となっている．このため，通信トラヒックから P2P の通信を弁別することで上記の問題を抑止することが求められている．これをうけて様々な弁別手法が提案されているが，従来の弁別手法では通信が暗号化された場合に P2P トラヒックを弁別することが困難であるといった問題が生じる．そこで本稿では P2P トラックの相互接続性に着目し，暗号化された通信でも P2P トラヒックを弁別できる手法を提案する．また，提案手法に対して暗号化された P2P トラヒックを用いて評価を行う．さらに，Planetlab を用いて広域ネットワークでの測定を行い，提案手法の適用範囲を明らかにする．

# An analytical method for the pure P2P traffic that focus attention on a bidirectional connection for encrypted connection

Masashi Suzuki,[†] Hirotake Abe,[††] Masayuki Okabe[†††]
and Kyoji Umemura [†]

Today, There are problems such as the information leak and the copyright infringement by the peer-to-peer file-sharing application, and the traffic occupation by some heavy users. For these reasons, P2P communication discrimination becomes necessary to suppress the above problems. As a result, various discrimination techniques have been proposed. However, if the communication is encrypted, it is usually difficult to distinguish P2P traffic from ordinal traffic by the existing discrimination technique. In this paper, we propose a technique to detect P2P traffic in encrypted communications, focusing on the inter-connecting behaving of P2P traffic. We have evaluated using the proposed methodology for an encrypted P2P traffic. In addition, by measurement in the wide-area networks using planetlab, we have discussed the limitation of the proposed methodology.

## 1. はじめに

### 1.1 背　景

近年，Winny などの P2P ファイル共有アプリケーションでは情報流出や著作権侵害などの問題が発生している．また，このほかの問題として P2P ファイル共有アプリケーションによって動画や音楽など巨大なサイズのファイル交換が行われるため，ネットワークトラヒックが大幅に増大しており，インフラへの多大な負担となっている[5]．このため，P2P ファイル共有アプリケーションのトラヒックを特定し，上記の問題を抑止することが期待されている．

これをうけて P2P ファイル共有アプリケーションを特定するためにいくつかの研究がなされている．例えばパケットのペイロードを解析することによって P2P ファイル共有アプリケーションを特定する One Point Wall[3] がある．これはペイロード部に含まれる P2P ファイル共有アプリケーション特有のビットパターンを検出する方法である．このほかにもトランスポート層のヘッダ情報から得られる情報を利用して，P2P ファイル共有アプリケーションのトラヒックパターンを弁別する方法もある[4]．しかし，これらの解析手法ではトランスポート層での暗号化が施されてしまった場合，解析するための情報が読み取れなくなってしまう．そのため，上記で挙げた弁別手法では暗号化通信

† 豊橋技術科学大学情報工学系
Information and Computer Science, Toyohashi University of Technology
†† 大阪大学サイバーメディアセンター
Cybermedia Center, Osaka University
††† 豊橋技術科学大学情報メディア基盤センター
Information and Media Center, Toyohashi University of Technolog

という条件下では P2P トラヒックを正しく弁別することができないといった可能性も生じる．

### 1.2 本研究の目的

前節で述べたように既存の解析手法では暗号化された通信から P2P トラヒックを弁別することが難しいことがわかる．そこで，我々は P2P ファイル共有アプリケーションの双方向通信に着目して P2P トラヒック解析を行う．この手法は通信が暗号化されていない条件での解析手法[2] が提案されている．本稿ではこの手法を拡張し，トランスポート層での暗号化が施された条件下でも P2P トラヒックを弁別できる手法を提案する．なお，この手法は[1] をもとにしており，実データの測定をもとに改良したものである．

### 1.3 本稿の構成

本稿は全 5 章で構成される．

第 1 章では，本研究の背景と目的について述べた．

第 2 章では，本研究のベースとなる P2P トラヒック弁別手法と，本研究で提案する P2P トラヒック弁別手法について述べる．

第 3 章では，暗号化を施した通信に対して本研究の手法を用いた場合に懸念される問題点について述べる．そして，これらの問題点に対する解決策を示す．

第 4 章では，様々なトラヒックに対して本研究の手法を適用した結果を示す．これにより，本研究の手法が有効であるかどうかを検討する．

第 5 章では，本研究の結論を述べる．

## 2. 双方向通信に着目した P2P トラヒック弁別手法

### 2.1 ベース研究について

通常のサーバとクライアント間の通信では，クライアント側からサーバ側へ通信路の確立を行う．このような通信モデルはクライアントサーバモデルと呼ばれる．しかし，一部のピュア P2P の通信ではクライアント同士で双方向に通信路を確立する必要があるため，各ノード間で向きの異なる二本の通信路が存在することになる．また P2P 通信では各ノードが自由に参加，離脱できるため，頻繁にノード間で通信路を確立する必要がある．このような通信モデルはピュア P2P モデルと呼ばれる．このように頻繁に双方向に通信路を確立するような通信はサーバとクライアントによる通信では存在せず，ピュア P2P での通信以外にはあまり存在しない．そこでこの双方向通信に着目することで，クライアントサーバモデルとピュア P2P モデルを区別して P2P トラヒックの弁別を行う． しかし，このベース研究では TCP のヘッダ情報を用いて通信路の確立を判別しているため，通信が暗号化されてしまった場合には P2P トラヒックを弁別することができないという問題がある．

### 2.2 暗号化通信における P2P トラヒック弁別の提案手法

通信路の確立を行うためにネゴシエーションを行う．このネゴシエーションの手順はスリーウェイハンドシェイクと呼ばれる．スリーウェイハンドシェイクはデータ通信に先立って SYN パケット，SYN/ACK パケット，ACK パケットの順に 3 回の通信を行う．これらのパケットの送受信後に通信路が確立される．

ベース研究の手法ではこの SYN パケットや SYN/ACK パケットを TCP のヘッダ情報を見ることで確認していた．しかし，暗号化された通信では TCP のヘッダ情報がわからないため， ベース研究のように SYN パケットや SYN/ACK パケットを確認することはできない．そこで本稿では，事前に調査した SYN パケットと SYN/ACK パケットの時間間隔 $\alpha$ と，相互にスリーウェイハンドシェイクが行われる時間間隔 $\beta$，それと SYN パケットと SYN/ACK パケットのパケット長をもとに P2P トラヒックを弁別する．これらの時間間隔を**図 1** に示す．

以上の特徴量を用いて P2P トラヒックを弁別する．まずスリーウェイハンドシェイクの判定条件は SYN パケットと SYN/ACK パケットのパケット長が共に 62byte であることから，62byte のパケットが送信されてから時間間隔 $\alpha$ の間に 62byte のパケットが返信された場合にスリーウェイハンドシェイクであると判定する．その後，時間間隔 $\beta$ 以内に反対側のホストから再度上記の判定方法でスリーウェイハンドシェイクを行っていることが判定されれば，P2P 通信であると判定する．

これにより，たとえ暗号化によって TCP のヘッダ情報を読み取ることができなくとも，パケット長と時間間隔 $\alpha$，$\beta$ を用いることでスリーウェイハンドシェイクを判別できるため，P2P トラヒックを弁別することができると考えられる．また，時間間隔 $\alpha$，$\beta$ は P2P アプリケーションの通信をキャプチャしたトラヒックをもとにして決める (具体的には後述する)

## 3. 暗号化通信について

### 3.1 暗号化について

本稿では通信の暗号化に使用する暗号化プロトコルとして IPsec[6] を想定する．IPsec は複数の暗号化方式を採用することができるので，通信相手との間で通

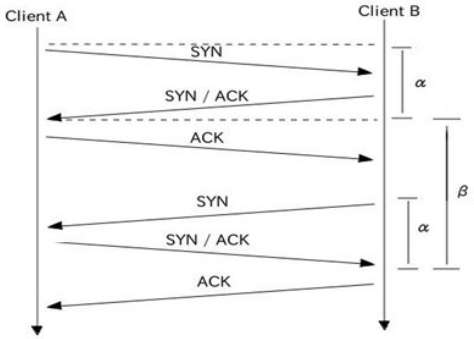


**図 1** P2P トラヒックの特徴量

信の設定を合わせるためにパラメータを共有する必要がある．このとき通信相手と共有するパラメータを SA と呼ぶ．SA には様々なパラメータがあるが，重要なパラメータとなるのがセキュリティプロトコルとモードである．セキュリティプロトコルには ESP と AH がある[7]．ESP はパケットの暗号化機能を提供し，AH は発信元の認証，完全性認証を提供する．モードにはトランスポートモードとトンネルモードの 2 つがある．トランスポートモードはペイロードだけを，トンネルモードは IP パケット全体をカプセル化する．

本稿ではセキュリティプロトコルに ESP，モードにトランスポートモードを選択して議論を進める．セキュリティプロトコルを ESP とした理由としては，AH では通信の暗号化ができず認証しか提供されていないためである．また，トランスポートモードを選択した理由は，今回は IP ヘッダまで暗号化するわけではなく，TCP ヘッダまで暗号化することを想定しているためである．なお，トランスポートモードを選択したからといって，パケット長が短いパケットは短く，長いパケットは長いことに変わりはない．さらに IP ヘッダが暗号化された場合でも，Point-to-Point で P2P ファイル共有アプリケーションが通信を行っていることは判別できる．このため，トランスポートモードを選択したからといって，一般性が失われることはない．

### 3.2 通信の暗号化による問題点と対策

#### 3.2.1 パケット長の変更による影響

IPsec によって暗号化を行う場合，パケットに対して暗号化のためのデータが付加される．そのため，トラヒック弁別で必要な特徴量であるパケット長が変化してしまうといった問題がある．

そこで，Windows XP で利用できる IPsec によって暗号化された通信をキャプチャして，暗号化されていない通信と暗号化された通信を比較してパケット長

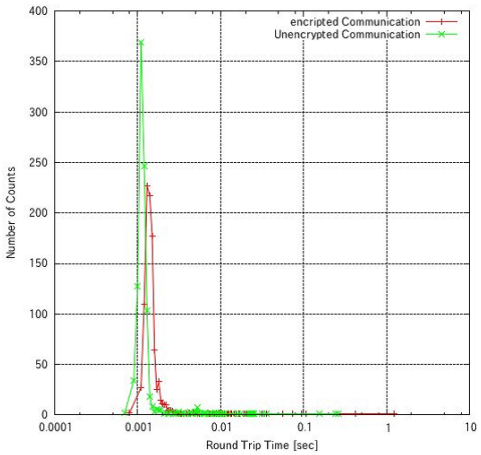


**図 2** 3 ウェイハンドシェイクの時間間隔の比較

の増加を調べた．結果として，すべてのパケット長が 32byte ずつ増加していることがわかった．これが他の環境であったとしても，パケット長の増加はプロトコルに依存するためにそれほど変化はない．これにより，暗号化された通信に対して本稿の手法を用いるには，特徴量として用いるパケット長を 32byte 増加させればいいことがわかった．

#### 3.2.2 暗号化処理による遅延の影響

暗号化通信を行う際には暗号化や復号化処理が必要となる．この処理によってラウンドトリップタイム（RTT）が増加することが想定される．RTT が増加してしまった場合，SYN パケットを送ってから SYN/ACK パケットが返ってくる時間間隔が長くなり，トラヒックの弁別に対して影響がでる可能性がある．

前節と同様に Windows XP を利用した IPsec によって暗号化されたトラヒックと暗号化されていないトラヒックからそれぞれ 1000 件のスリーウェイハンドシェイクを抽出し，それぞれの時間間隔と比較した．結果を**図 2** に示す．図 2 から暗号処理による平均の遅延時間は約 0.001 秒であることがわかった

## 4. 評 価 実 験

### 4.1 暗号化通信に対する弁別手法の評価実験

本節では，よく知られたピュア P2P ファイル共有アプリケーションである Winny[8] を用いて実験を行う．実験では IPsec によって暗号化された通信を用いて本稿の弁別手法が暗号化通信に対して有効性があるかどうか検討する．また，暗号化プロトコルには Windows XP で利用することができる IPsec を用いる．

実験には，Winny による通信のみをキャプチャした２種類のトラヒックと，Winny による通信パケットを含まない１種類のトラヒックを用いる．Winny のみの通信をキャプチャしたトラヒックは，通信が暗号化されたトラヒック（IPsec トラヒック）と，通信が暗号化されていないトラヒック（nonIPsec トラヒック）がある．また，Winny による通信パケットを含まない通信トラヒック（nonP2P トラヒック）は，暗号化されていない通信をキャプチャしたものである．これらのトラヒックの一覧を**表 1** に示す．

前述の各トラヒックに対して弁別手法を適用することで，弁別手法の有効性を検討する．実験の流れを以下に示す．

(1)　Winny のみの通信トラヒックを後述する実験環境から取得する．Winny パケットを含まないトラヒックは研究室のトラヒックから取得する．各トラヒックの測定時間はそれぞれ 3 時間とした

(2)　P2P トラヒックから特徴量 $\alpha$，$\beta$ を決定した．具体的な内容は後述する

(3)　特徴量 $\alpha$，$\beta$ を用いてトラヒックの判定条件を設定した．この条件に従ってそれぞれのトラヒックに対して弁別手法を用いた結果から弁別手法を評価する

#### 4.1.1　P2P トラヒック収集のための実験環境

実験環境にはホスト OS として Linux(Ubuntu9.04) をインストールした 5 台のマシンを用意した．このうち 4 台のホストマシンには仮想環境として VMware Server を導入し，残り一台をパケットキャプチャのためのマシンとした．VMware Server を導入した 4 台のホストマシン上で，それぞれ 2 台ずつの仮想マシンを動作させ，この仮想マシン上で Winny を動作させることによって Winny による通信のみをキャプチャしたトラヒックを取得した．上記の構成を**図 3** に示す．

#### 4.1.2　特徴量である時間間隔 $\alpha$ の決定

実験環境から取得した P2P トラヒックから 150 件のスリーウェイハンドシェイクを抽出して，SYN パケットと SYN/ACK パケットの通信間隔を計測した．計測結果を**図 4** に示す．図 4 から SYN パケットと SYN/ACK パケットの時間間隔は最長で 0.0418 秒であった．これから時間間隔 $\alpha$ を $\alpha$=0.042 とした．このように設定した理由は，スリーウェイハンドシェイクを見逃さないようにするためである．また $\alpha$ を長く設定したとしても時間間隔 $\beta$ によって P2P トラヒック自体の誤検出は防ぐことができる

**表 1**　トラヒック一覧

| | 暗号化された通信 | 暗号化されていない通信 |
|---|---|---|
| winny通信を含む | IPsecトラヒック | nonIPsecトラヒック |
| winny通信を含まない | ――― | nonP2Pトラヒック |

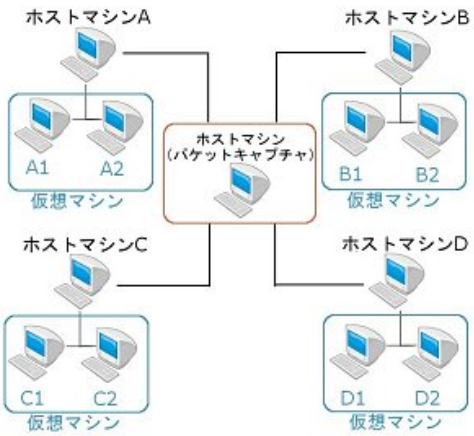


**図 3**　実験環境の構成

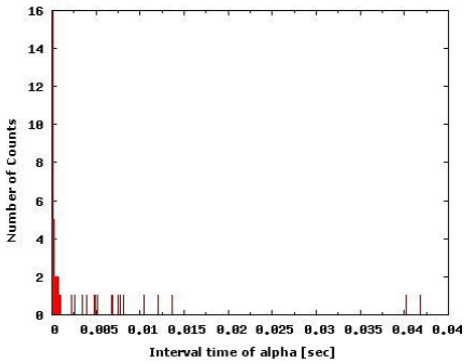


**図 4**　時間間隔 $\alpha$ の測定

#### 4.1.3　特徴量である時間間隔 $\beta$ の決定

相互にスリーウェイハンドシェイクを行う時間間隔 $\beta$ を計測する．実験環境で得た P2P トラヒックから相互にスリーウェイハンドシェイクを行う際の時間間隔を 320 件観測した．観測結果が**図 5** である．図 5 から相互にスリーウェイハンドシェイクを行う時間間隔は最長で 0.307 秒であった．$\alpha$ では最長より長い時間としたが，$\beta$ はこのようにすることは適切ではない．時間間隔 $\beta$ を大きな値に設定する場合，誤検出が大きくなる可能性がある．そのため，ROC カーブを作成することで最適な特徴量 $\beta$ を特定した．パケット長を 62byte，特徴量 $\alpha$ を 0.042 とし，特徴量 $\beta$ を 0.01 間隔で 0 秒から 0.25 秒変化させたときの nonIPsec トラヒックの検出率を y 軸，nonP2P トラヒックの検出率を x 軸としてプロットした図を**図 6** に示す．図 6 か

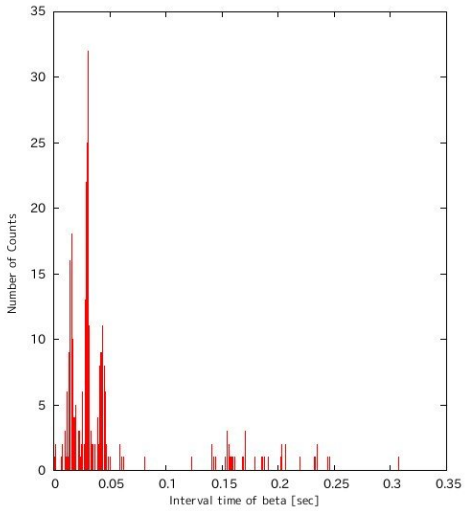


**図 5**　時間間隔 $\beta$ の測定

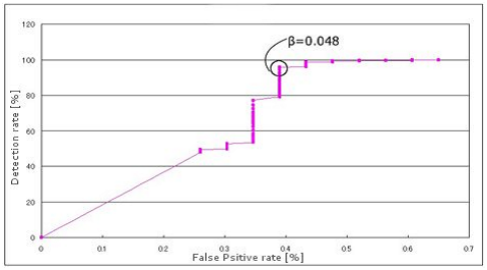


**図 6**　時間間隔 $\beta$ の ROC カーブ

ら検出率の高く，誤検出の少ないときの $\beta$ のパラメータは 0.048 であった

#### 4.1.4　暗号化された通信に対する弁別結果

取得したトラヒックから弁別手法を評価する．nonP2P トラヒックと nonIPsec トラヒックに対しての特徴量を $\alpha$ =0.042，$\beta$ =0.048，パケット長=62 とした．また，IPsec トラヒックに対しての特徴量 $\alpha$，$\beta$ は 3．2 章の結論から遅延時間を考慮して $\alpha$ を 0.001 秒増加させ，$\beta$ を 0.003 秒増加させた．$\beta$ を 0.003 秒増加させた理由として 1 回目のスリーウェイハンドシェイクから 2 回目のスリーウェイハンドシェイクが行われるまでに ACK パケット，SYN パケット，SYN/ACK パケットが送信されるためである．このため IPsec トラヒックに対しての特徴量は $\alpha$ =0.043，$\beta$ =0.051，パケット長=94 とした．各トラヒックの弁別結果を**表 2** に示す．

IPsec トラヒックは通信が暗号化されてしまっているため，スリーウェイハンドシェイクが行われた回数を正確には把握することができないが，以下のように考えて処理した．IPsec トラヒックと nonIPsec トラヒックは同様の環境で同様の時間だけキャプチャしたトラヒックであるため，Winny が通信を行った回数は同程度になると考えた．これを踏まえて IPsec トラヒックと nonIPsec トラヒックの結果を比較すると，nonIPsec トラヒックの検出数に比べて IPsec トラヒックの検出数は約 1/4 まで下がってしまっている．ただ，nonP2P トラヒックの結果からわかるようにフォールスポジティブレート（FPR）は 0.5064%と低いので，IPsec トラヒックの検出数が少なくなったとしても，Winny 通信を行っているトラヒックを発見することができると考えられる．

**表 2**　各トラヒックへの弁別手法の適用

| トラヒック | コネクション数 | | FNR[%] |
|---|---|---|---|
| | 検出数 | 合計数 | |
| nonIPsecトラヒック | 5959 | 6234 | 4.4113 |
| Ipsecトラヒック | 1865 | | |

| トラヒック | コネクション数 | | FPR[%] |
|---|---|---|---|
| | 検出数 | 合計数 | |
| nonP2Pトラヒック | 11 | 2172 | 0.5064 |

### 4.2　遅延時間を考慮した弁別手法の評価

上記の実験ではローカル環境における提案手法の評価を行った．ルーティングが安定している環境で本方式は有効と考えられるが，ルーティングが複雑なインターネットなどの環境では，ローカル環境のように安定で短い遅延時間とはならないことが予想される．そのため，インターネットのテストベッドである Planetlab を利用してインターネットに近い環境でのネットワーク遅延を調べることで本稿の提案する方法の適用の限界を検討する．

### 4.3　Planetlab 上でのネットワーク遅延の影響

実験環境と Planetlab 上でそれぞれ 100 件のスリーウェイハンドシェイクの遅延時間を観測した．Planetlab で使用したノードは pl2．Planetlab.ics.tut.ac.jp (133.15.59.2) と，planetlab-02.naist.jp (163.221.11.72) である．それぞれの環境で観測した遅延時間の平均と分散を**表 3** に示す．表 3 から実験環境と Planetlab の遅延時間の平均を比較すると，実験環境に比べて，Planetlab の遅延時間の平均が大変大きいことがわかる．また分散に関しても実験環境が安定していることに比べて，Planetlab の観測データにはばらつきがあることがわかる．ここで，それぞれの観測結果をプロットしたものを**図 7** に示す．このデータからわかるように，平均値から± 0.02 の範囲では実験環境は 90%の

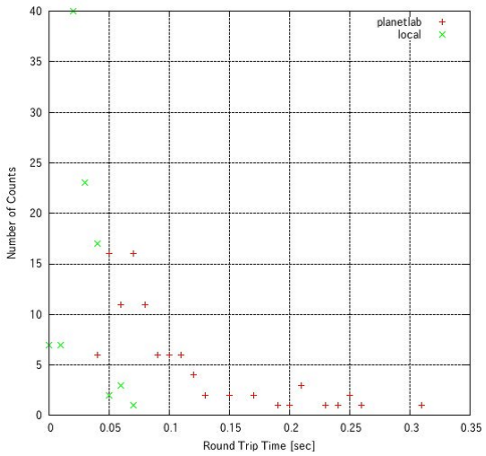


**図 7** Planetlab と実験環境上でのネットワーク遅延の観測結果

観測データが集まっていることに対し，Planetlab では 30%ほどの観測データしかないことがわかる．さらに，場合によって大きな遅延が発生する．そのため広域ネットワークに対しては時間間隔だけでは情報が不足していることがわかった

## 5. 結　　論

本稿では通信が暗号化された P2P トラヒックを弁別する手法を説明し，その手法の有効性を調査した．IPsec によって暗号化された通信トラヒックに対して 2.2 章で述べた弁別手法を適用した．P2P による通信を含まないトラヒックによって誤検出を調べた結果，フォールスポジティブレートが 0.5064%と大変低かった．通信が暗号化された場合では，通信が暗号化されていない場合に比べて P2P アプリケーションによって確立されたコネクションの検出数が約 1/4 まで下がったが，通信が暗号化された状態でも，ローカルネットの環境ならば本手法は P2P トラヒックを検出する手法として妥当であるとわかった．

さらに，Planetlab を用いてインターネットに近い環境での遅延時間を調べることで，提案手法が適用できる範囲を調査した．結果として実験環境での遅延時間の平均値に比べて Planetlab 上での遅延時間の平均値が大きいことがわかった．さらに，Planetlab では実験環境に比べて観測データにばらつきはあることがわかった．以上の結果から，平均値からはずれたデータが多くなる広域ネットワークにおいては，本稿の提案手法をそのまま使用することに問題があることが判明した．

**表 3** 遅延時間の平均と分散

| | Local | Planetlab |
|---|---|---|
| 平均 | 0.03048 | 0.10851 |
| 分散 | 0.00019 | 0.00943 |

## 参　考　文　献

1) 三浦明日香, 梅村恭司, 阿部洋丈, 岡部正幸: SYN パケットの呼応に着目した P2P トラヒックの表示, 情報処理学会全国大会講演論文集， pp.239-240 (2009)
2) 松田崇， 中村文隆， 若原恭， 田中良明: 相互接続における順逆接続間隔を利用した P2P トラヒック分別手法， 信学技報， No.NS2006-237, pp.415-420 (2007)
3) ”One Point Wall” http://www.onepointwall.jp/
4) 松田崇， 中村文隆， 若原恭， 田中良明， 大崎淳， 千田浩一， 加藤圭， 飯塚正: PureP2P ファイル共有トラヒックの特性解析， 信学技報, No.NS2005-2， pp.5-8 (2005)
5) 亀井聡: P2P 技術がネットワークインフラに及ぼす影響と課題， コンピュータソフトウェア, Vol.22， No.3， pp.8-18， 日本ソフトウェア科学会, (2005)
6) ”Security Architecture for the Internet Protocol”, RFC 4301, IETF
7) ”Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH)”, RFC 4305, IETF
8) 金子勇: Winny の技術， アスキー， (2007)