

# 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月31日現在

機関番号：12102

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22560359

研究課題名（和文） BGP・TCPとの連携制御によるP2Pコンテンツ配信の最適化方式

研究課題名（英文） P2P File-Download Control Method Associated With Router Network Parameters

研究代表者

川原崎 雅敏（KAWARASAKI MASATOSHI）

筑波大学・図書館情報メディア系・教授

研究者番号：70375517

研究成果の概要（和文）：

P2P トラフィックが急増して ISP 網設備を圧迫しているが、P2P は物理網（ISP 網）の状態を意識せずに動作するため、実際には ISP 網リソースを不必要に占有していると考えられる。本研究では、ピア間で自律的に RTT や TCP ウィンドーサイズ等のネットワーク状態を測定し、物理網的に近くにあり可用帯域の広い経路を持つピアを選択してダウンロードする制御により、ISP 網設備の効率的利用と P2P ファイル取得時間の短縮を図る方式を提案し、その有効性をシミュレーションにより明らかにした。

研究成果の概要（英文）：P2P applications impose an enormous traffic load on the Internet. Actually, however, topological difference between P2P network and underlying physical network (Internet network) is creating inefficient use of the physical network resources. To this problem, we proposed P2P download control methods using router network parameters (e.g., RTT, TCP window size and BGP path information) that can achieve effective use of router network resources, thus relieve ISP traffic burden and improve P2P application performance. Simulation results revealed the effectiveness of the proposed methods.

交付決定額

（金額単位：円）

|        | 直接経費      | 間接経費      | 合 計       |
|--------|-----------|-----------|-----------|
| 2010年度 | 1,500,000 | 450,000   | 1,950,000 |
| 2011年度 | 1,000,000 | 300,000   | 1,300,000 |
| 2012年度 | 900,000   | 270,000   | 1,170,000 |
| 年度     |           |           |           |
| 年度     |           |           |           |
| 総 計    | 3,400,000 | 1,020,000 | 4,420,000 |

研究分野：工学

科研費の分科・細目：電気電子工学、通信・ネットワーク工学

キーワード：P2P、物理網、連携制御、TCP、性能解析、BGP

## 1. 研究開始当初の背景

国内のインターネットトラフィックは急速に増大し続けているが、その8割程度をP2Pトラフィックが占めているとの報告がある。その結果、ISPは設備増設を迫られているが、

P2Pアプリケーションは物理網（ISP 網）の状態を意識せずに動作するため、オーバーレイ網である P2P ネットワーク上で近傍のユーザからファイルを取得したとしても、物理網では必ずしも近傍から取得したことには

ならない。事実、P2P トラフィックは不必要にプロバイダ境界をよぎっているとの報告があり、ISP 設備を有効利用できていないばかりでなく、プロバイダ間リンクがボトルネックとなって品質が劣化する恐れがある。

このような状況を改善すべく、世界的に研究や実験が始まっている。中でも、SIGCOMM に発表された P4P[1]が注目され、IETF においても対応する Working Group [2]が設立され、Verizon やエール大学等による実証実験も行われている。P4P とは、物理網の状況を意識して P2P の性能を向上させる試みである。プロバイダが P4P ディスタンスと呼ばれる「仮想的なピア間の距離」を計算して i-Tracker と呼ばれるサーバに提供し、ディスタンスが近いピアからファイルを取得することでパフォーマンスを最適化できることを示している。文献[3]も同様の問題意識に基づき、極力 AS(Autonomous System) 内からファイルを取得する方法を提案している。

しかしながら、P4P には幾つか問題がある。第一は、前提とする P2P のファイル共有方式である。i-Tracker という名前からも推測されるように、P4P のアーキテクチャは BitTorrent 型の P2P ファイル共有が暗黙の前提になっている。世界的には BitTorrent が主流であるが、日本では Winny 等が主流であるため P4P 方式はそぐわないと考えられる。

第二は、P4P 方式の運用上の課題である。IETF の議論においても明らかになってきたように、ISP は自らのネットポロジを教えず、P2P ユーザも自らの IP アドレスを教えない。その結果、i-Tracker に物理網の状況を集約することは実際には困難であり、i-Tracker の情報に基づいて P2P ファイル取得を最適化するというロジックは成り立たなくなる。

以上の背景から、国内における P2P トラフィックの急増に対処するためには P4P に代わる有効な方式の開発が望まれる。

## 2. 研究の目的

本研究は、i-Tracker といった集中サーバを利用せず、RTT・TCP 輻輳制御・BGP ルーティングなど既存インターネットの測定・制御機能を P2P アプリケーションに連携させることにより、物理網の状況を意識した P2P ファイル取得を実現することで、この問題を解決しようとするものである。前提とする P2P ファイル共有は BitTorrent ではなく Gnutella 型を想定し、サーバを用いない自律分散制御により、物理網に不必要な負荷をかけることなく効率的なファイル取得を実現できることを目指す。

【文献】 [1] H. Xie, Y.R. Yang, A.Krishnamurthy, Y.G. Liu, A. Silberschatz, "P4P: Provider Portal for Applications," ACM SIGCOMM Computer Communication Review, Vol.38, No.4, pp.351-362, Aug. 2008

[2] IETF WG: "Application-Layer Traffic Optimization (alto)", Low Extra Delay Background Transport (ledbat)

[3] V. Aggarwal, A. Feldmann, C. Scheidele, "Can ISPs and P2P users cooperate for improved performance?", ACM SIGCOMM, Vol.37, No.3, July 2007

## 3. 研究の方法

### (1) 課題のモデル化

Gnutella 型ファイル共有アプリケーションでは図 1(a)に示すように、ファイルを要求するピア A は P2P 網上に Query を送信し、欲しいファイルを保持しているピア B1-B5 はピア A に QueryHit を返信する。ピア A は P2P 網上で最も自らに近いピア、すなわち最初に応答してきたピア B3 からファイルをダウンロードする。しかし、物理網上は図 1(b)に示すように B3 よりも B1 や B5 の方が A に近い (ルータホップ数が少ない)。したがって、B3 よりも B1 か B5 からダウンロードした方がネットワークリソースの消費量は少ない。一方、ファイルをダウンロードするためには経路上にファイル転送に足る十分な空き帯域が必要である。したがって最適な相手ピアは、B1-B5 の中で、物理網上で A に近く、A への経路上に十分な空き帯域があるピアであると言え、課題は最適ピアの選択問題に帰着される。

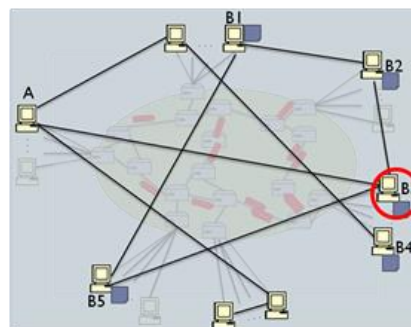


図 1(a) P2P 網での動作

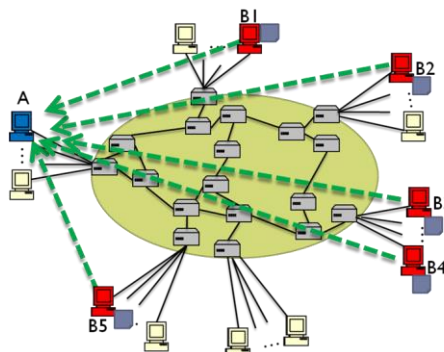


図 1(b) 物理網での動作

## (2)最適ピアの選択方法

ファイル要求ピアが複数のファイル保持ピアの中から最適なピアを選択する指標として、①物理網的に近いピアの選択には **RTT** とルータホップ数、②経路の帯域が広いピアの選択には **TCP** ウィンドーサイズとパケットペア方式による可用帯域推定を用いる方法を検討した。さらに、**ASP** 境界を極力よぎらないピアの選択に **BGP** パス情報を用いる方法を検討した。

## (3) ネットワークモデル

上記①、②による制御効果の比較評価には計算機シミュレーションを用いた。この際、ネットワークのモデル化が課題となる。**P2P** 網と物理網 (**ISP** 網) の2階層モデルとし、**RTT** やウィンドーサイズに有意差が出る程度の大規模網モデルとする必要がある。ルータ数は100とし、代表的な**ISP** 網ではルータ数とアクセスリンク数の比が平均して1対3.39であることから、ピア数は339とした。

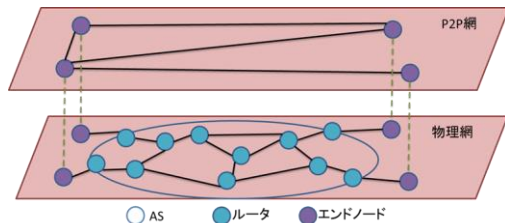


図2 2階層のネットワークモデル

ネットポロジータについては、**P2P** 網はノード次数がべき乗則に従うモデルで良く近似できることが知られているので、代表的な **BA** (**Barabási-Albert**)モデル[4]で近似した。物理網をべき乗則で近似した研究例もあるが、これではルータ間距離やリンク容量をモデル化できない。そこで物理網は **FKP** (**Fabrikant, Koutsoupias, Papadimitriou**)モデル[5]を用いることとした。

**FKP** モデルは、式(1)を用いて完全グラフにノードとリンクを付加していくことで、ノードの位置関係を記述可能なネットワークポロジータを生成する。

$$\min_j \alpha \cdot d_{ij} + h_j \quad (1)$$

式(1)は追加ノード  $i$  と接続する既存ノード  $j$  を決める際に用いられ、 $i-j$  間の物理距離  $d_{ij}$  と、ノード  $j$  の論理距離  $h_j$  の重み付け和 (重み:  $\alpha$ ) が最小になる  $j$  を選択する。ノードはユークリッド平面上にランダムに配置され、その座標から物理距離が決まる。論理距離はノードホップ数に相当する。重み  $\alpha$  を大きくすると物理距離が重視されてノードは近くのノード同士でつながろうとする。逆に  $\alpha$  を小さくすると論理距離が重視され、追加ノードはより少ないホップ数であらゆるノードにたどり着けるようにノ

ード次数の高いノードと接続するようになり、ハブノードが形成される。 $\alpha=45$  とした。なお、**FKP** モデルはノード次数1のノードが多く生成される性質があるが、代替経路を確保するために最低ノード次数を2とした。また、リンク容量は一律に **24Mbps** とした。**FKP** モデルで生成した物理網トポロジータを図3に示す。

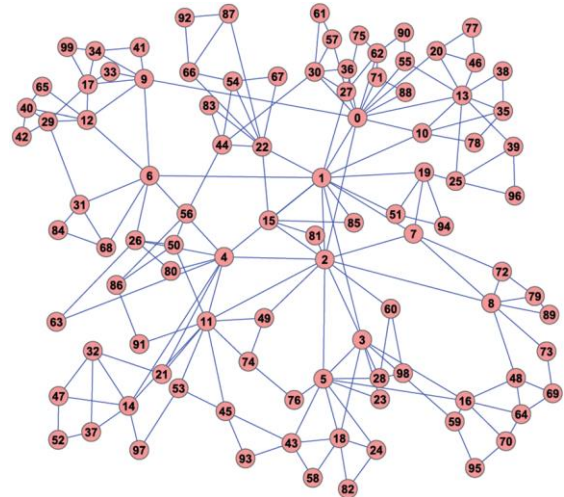


図3 FKPモデルで生成した物理網（ルータ網）

【文献】[4] 今野紀雄, 井出勇介著, 『複雑ネットワーク入門』, 第1版, (株)講談社, 2008, 161p

[5] Lun Li, et.al., “A First Principles Approach to Understanding the Internet’s Router-level Topology”, **SIGCOMM’04**, pp3-14, 2004

## (4) シミュレータの作成

物理網は上述した **FKP** モデルを用い、**NS2** (**Network Simulator-2**)を利用してシミュレータを作成した。要件は以下の3つである。

- 1) 現実的な物理網の要素(伝搬遅延、ノード次数など)を有すること
- 2) **P2P** アプリケーションのピアとしてエンドノードを有すること
- 3) リクエストピア  $A$  と候補ピア  $B1-B5$  間の経路において、伝搬遅延、可用帯域といった物理網状態に有意な差が出ること。

図3のモデルでは、ノード間経路上のリンクの総伝搬遅延は最長で **84.6ms**、最短で **0.0057ms** となり、経路によって伝搬遅延に差が生じている。さらにリンクの可用帯域に差を作るため、本シミュレーションではルータ網にバックグラウンドトラフィックを流した。

一方、**P2P** 網はピア数を339個(物理網のエンドノードに相当)、最低ノード次数を2として、トポロジータジェネレータ **BRITE** により作成した。なお、ダウンロードファイルサイズは **10MB** とした。

## (5) シミュレーションシナリオ

シミュレーションの流れを以下に示す。



- ① ファイル保持ピア F をエンドノードからランダムに 50 個選出する。
- ② リクエスト A が発生する。A は F に選ばれなかったエンドノードからランダムに選出される。発生タイミングはポアソン生起による。バックグラウンドトラヒックと同程度のリンク使用率になるようにパラメータを調節した。リクエスト A は 1 秒～40 秒の間に 258 回発生する。
- ③ A はファイル検索を行い、F の中からダウンロード元候補ピア B1~B5 を探す。P2P 網上でホップ数が小さい F を順番に選択していく。ここで、同じホップ数の F が複数あった場合は、その中からランダムで選ぶ。
- ④ B1~B5 に対してダウンロード制御を行う。すなわち、5 つの候補 B1~B5 から最適ピア選択方式（提案方式）にしたがってダウンロード元を決め、ダウンロードを開始する。
- ⑤ シミュレーション期間中、②～④を繰り返す。A および B1~B5 は、シミュレーション中に複数回選出されうる。

以上の流れでシミュレーションを行った。提案方式と比較するため従来方式である Gnutella 方式についてもシミュレーションを行った。Gnutella は最初に QueryHit を受信したノードからダウンロードを開始するので、常に B1 からダウンロードすることとした。

#### (5) 制御パラメータの計測方法

提案方式では、最適ピアを選択する指標として RTT や TCP ウィンドーサイズを利用する。これらはプローブパケットを送信して測定するが、プローブトラヒックも ISP 網リソースを消費するため、方式評価には測定負荷も含めて行う必要がある。

##### (a) RTT 方式

候補ピアからリクエストピアに向けて Ping プロブを 4 回送信して RTT を計測し、その平均値が小さい候補ピアからダウンロードする。RTT 方式は、物理網上で近傍にあり、混雑が少ない候補ピアを選択する。

##### (b) TCP ウィンドーサイズ方式

候補ピアからリクエストピアに向けて TCP コネクションを張り、ペイロードサイズ 1Byte のプローブパケットを 600 個転送して 0.01 秒間隔でウィンドーサイズ(W\_size)を計測し、その平均値が最も大きい候補ピアからダウンロードする。W\_size は可用帯域が広いほど大きくなり、開き方も早くなる。W\_size 方式ではより可用帯域が広い候補ピアを選択する。

##### (c) パケットペア方式

2 個のプローブパケットを連続的に送信し、ボトルネックリンク通過前後のパケット間

ギャップ( $\Delta_{in}$ 、 $\Delta_{out}$ )の変化から、(2)式により経路の可用帯域 A を推定する[6]。推定値がより大きい候補ピアを選択する。C はボトルネックリンクの容量、すなわち経路のリンク容量である。

$$A = (1 - (\Delta_{out} - \Delta_{in}) / \Delta_{in}) \times C \quad (2)$$

C も未知であるので、同じくパケットペア方式である CapProbe 法[7]により推定する。

#### 【文献】

[6] J.Strauss, et.al., “A measurement study of available bandwidth estimation tools,” in IMC, 2003

[7] R.Kapoor, et.al., “CapProbe: a simple and accurate capacity estimation technique,” Proc. ACM SIGCOMM '04, pp.67-78, 2004

## 4. 研究成果

本研究の目的は、①P2P トラヒックが物理網にかけける無用なトラヒック負荷の削減、② P2P アプリケーションのパフォーマンス向上、である。そこで、提案した各ダウンロード制御方式（RTT 方式、ウィンドーサイズ方式、パケットペア方式、および無制御の Gnutella 方式）を、物理網にかけける負荷量とダウンロード効率の観点から以下を指標として比較評価した。

#### 【物理網にかけける負荷の評価指標】

- 平均ネットワークリソース使用量(NRU)
- 同時ダウンロード接続数（同時 DL 数）

#### 【パフォーマンス向上の評価指標】

- ダウンロード完了数（DL 完了数）
- 平均ダウンロード完了時間（DL 時間）

### (1) 実験結果

#### (a) 平均 NRU

図 4 は各方式の平均 NRU である。RTT 方式は物理網的に近傍なピアを選択するため平均 NRU が小さい。一方、W\_size 方式と Pkt\_Pair 方式は迂回してでも可用帯域の広い候補を選択するため平均 NRU が大きくなっている。

Gnutella の平均 NRU を 100 とすると、RTT 方式は 84、W\_size 方式は 109、Pkt\_Pair 方式は 113（内、プローブ負荷が 14）となった。

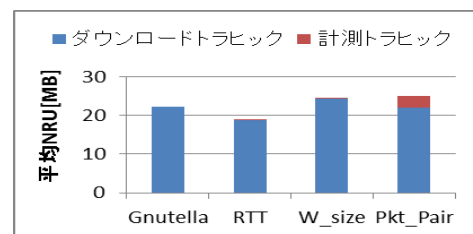


図 4 平均 NRU

#### (b) 同時ダウンロード接続数

各ルータ間リンクの同時 DL 数の時間推移を図 5 に示す。W\_size 方式と Pkt\_Pair 方式は可用帯域の広いピアからダウンロードする

ため同時 DL 数が小さくなっている。

Gnutella 方式は網の混雑状態を考慮せずに DL を試みるため、多くのリンクで同時 DL 数が大きくなり、一度上がった同時 DL 数はほとんど下がらない。即ち多くの DL が互いの経路を圧迫し合って完了しない状態に陥っている。一方、RTT 方式は同時 DL 数が大きいリンクこそ少ないが、最大同時 DL 数は Gnutella 方式より高い。特に、経路が短いピアを選ぶためにハブノードを通る経路で同時 DL 数が大きくなり、負荷が特定リンクに集中して DL 効率が低下している。

W\_size 方式はほぼ全リンクにわたって同時 DL 数が低く、DL トラヒックの負荷分散に大きな効果があり、DL 効率も高い。これは可用帯域の広い候補ピアから DL するため、他の DL が存在している経路を避けるためである。パケットペア方式も W\_size 方式に次ぐ負荷分散効果がある。

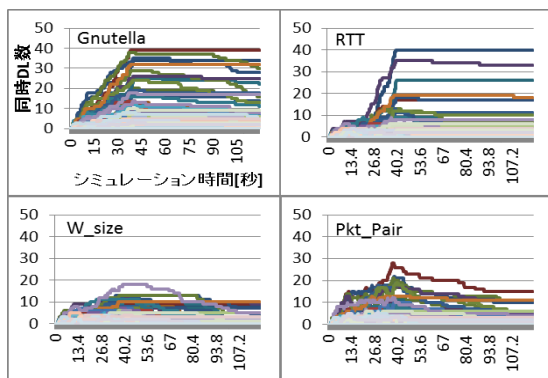
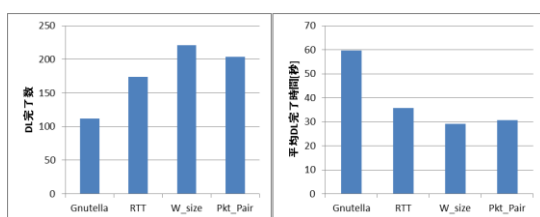


図 5 同時 DL 数

### (c)ダウンロード性能

図 6 は各方式の DL 完了数(a)と平均 DL 時間(b)である。W\_size 方式は最も多くのダウンロードを最も少ない時間で実現しており、ダウンロード性能が高いことがわかる。



(a)DL 完了数

(b)平均 DL 時間

図 6 ダウンロード性能

### (2)考察

RTT 方式は、地理的に最も近い候補ピアを予測してファイルダウンロードを行うことで Gnutella 方式に比べてネットワークリソース使用量を 16%削減した。同時に、混雑した経路もある程度推測して避けるので、ダウンロード性能も向上する。ただし、タイミングによっては混雑した経路でも RTT を取得できてしまうので、混雑状況を見逃して地理的

な近さのみで候補を判断し、結果的にハブノード周辺にダウンロードトラヒックを大量に流してしまう。これはダウンロードトラヒックの一極集中を発生させ、ダウンロード性能を低下させている。

ウィンドーサイズ方式は、可用帯域が広い候補ピアからファイルをダウンロードする。可用帯域は広いが遠方にいる候補ピアを選ぶことがあるため、ネットワークリソース使用量は従来方式よりも 9%程度悪化する。しかし、そうしてダウンロードトラヒックを分散させることで、トラヒック負荷を平滑化させるとともに、ダウンロード性能を従来方式の 2 倍以上に向上させることが判明した。

パケットペア方式は、ウィンドーサイズ方式と同様に負荷の平滑化やダウンロード性能向上の効果を示したが、推定値の誤差が大きく、制御効果は十分ではない。また、プローブパケットのパケットサイズが他の提案方式よりも大きいので、計測負荷も膨大である。

以上の考察から提案方式の中では、負荷分散とダウンロード性能の向上に高い効果を示すウィンドーサイズ方式が P2P のダウンロード制御に最も有効な方式であると言える。

### (3) 結論

本研究は、P2P トラヒックが ISP 網を圧迫している現状において、サーバや ISP 事業者の助けを借りず、ピア間の自律分散的な制御により物理網状態を予測し、最適なピアからファイルをダウンロードすることにより、物理網の利用効率とファイルのダウンロード効率を同時に最大化する方式について提案と評価を行った。

#### (a) 物理網パラメータを利用する方法

RTT で経路の物理網的な距離と混雑の度合いを、ウィンドーサイズおよびパケットペアで経路の可用帯域を予測し、物理網的に近く、可用帯域の広い経路を持つピアからダウンロードする方式を提案し、その有効性を大規模網シミュレーションにより検証・評価した。その結果、RTT 方式は無駄なトラヒックを減らすのが P2P トラヒック負荷を特定リンクに集中させてしまうことが判明した。パケットペア方式は負荷分散とダウンロード性能向上に一定の効果を示したが、可用帯域の推定誤差とプローブ負荷の大きさが課題となった。これらに対し、ウィンドーサイズ方式は、若干のネットワークリソース使用量増と引き換えに、効果的な負荷分散とダウンロード性能向上をもたらすことが判明した。提案方式の中では最も有効な制御方式であると言える。

#### (b) BGP パス情報を利用する方法

BGP パス情報の利用については実網で検証を行った。筑波大学キャンパスネット

ワークの関門ルータに BGP パス情報をホスト端末に渡す機能を追加し、同一の Web コンテンツを保持する海外の複数サイトにアクセスして調べたところ、経路上の一部の ISP が BGP パス情報を隠ぺいしてしまっており、エンド・エンドのパス情報が得られるケースはほとんどなかった。BGP パス情報をピア選択に用いる方式は現実的とは言えない。

#### (c) 実網実験

提案方式の実網実験についても検討を進め、研究室環境で実機に方式を実装して正常動作を確認した。これをグローバルな P2P 実験網である PlanetLab で実験するため、米国の PlanetLab 事務局に筑波大学学術情報メディアセンターから正式に参加申請を行った。しかし 1 年半以上経過しても進捗がなく、参加は実現していない。参加でき次第、実験開始する予定である。

#### 5. 主な発表論文等

〔雑誌論文〕(計 8 件)

1. M. Kawarasaki, K.Suzuki, G. Itoh; "P2P File-Download Control Method Associated With Router Network Parameters", Proc. of European Modelling Symposium 2010 (EMS2010), pp500-506, 2010 (査読有)
2. M. Kawarasaki, G. Itoh: "Network Aware Peer Selection Method for P2P File-Downloading Using Packet-Pair Measurement", Proc. of ICUMT2011, 7pages, 2011 (査読有)
3. Y. Yoshida and M. Kawarasaki: "Relay Node Based Proactive Load Balancing Method in MPLS Network with Service Differentiation", Proc. of IEEE Workshop on User-Centric Networking (U-Net'12), pp7050-7054, 2012 (査読有)
4. Y. Shomura, K. Yoshida, A. Sato, S. Matsumoto, K. Itano: "A Traffic Analysis Using Cardinalities and Header Information", Proc. of ICNC2010 (International Conference on Networking and Computing), pp55-62, 2010 (査読有)
5. S. Takada, A. Sato, et.al.: "A P2P Approach to Scalable Network-Bootstrapping", Proc. of ICNC2012, pp201-207, 2012 (査読有)
6. S. Takada, A. Sato, et.al." A Simple Collaborative Method in Web Proxy Access Control for Supporting Complex Authentication Mechanisms", Proc. of Networking, Applications and Worksharing (CollaborateCom2012), 10pages, 2012 (査読有)
7. K. Fujiwara, A. Sato, et.al., "DNS traffic analysis - Issues of IPv6 and CDN -", Proc.

of SAINT2012, pp129-137, 2012 (査読有)

8. K. Fujiwara, A. Sato and K. Yoshida, "DNS traffic analysis - CDN and the World IPv6 Launch -", Journal of Information Processing, Vol.21, No.3, 8pages, 2013.1 (査読有)

〔学会発表〕(計 7 件)

- ① S. Takada, A. Sato, et.al.: "A P2P Approach to Scalable Network-Bootstrapping", ICNC2012, pp201-207, 2012.12.5, Okinawa Gender Equality Center (Japan)
- ② S. Takada, A. Sato, et.al." A Simple Collaborative Method in Web Proxy Access Control for Supporting Complex Authentication Mechanisms", Collaborate-Com2012, 2012.10.16, Renaissance Pittsburgh Hotel (USA)
- ③ K. Fujiwara, A. Sato, et.al., "DNS traffic analysis - Issues of IPv6 and CDN -", SAINT2012, 2012.6.16, Swiss Hotel Izmir (Turkey)
- ④ Y. Yoshida and M. Kawarasaki: "Relay Node Based Proactive Load Balancing Method in MPLS Network with Service Differentiation", IEEE U-Net'12, 2012.6.11, Ottawa Convention Centre (Canada)
- ⑤ M. Kawarasaki, G. Itoh: "Network Aware Peer Selection Method for P2P File-Downloading Using Packet-Pair Measurement", ICUMT2011, 2011.10.5, Danubius Hotel Arena (Hungary)
- ⑥ M. Kawarasaki, K.Suzuki, G. Itoh; "P2P File-Download Control Method Associated With Router Network Parameters", EMS2010, 2010.11.17, Grand Hotel Duomo (Italy)
- ⑦ Y. Shomura, K. Yoshida, A. Sato, S. Matsumoto, K. Itano: "A Traffic Analysis Using Cardinalities and Header Information", ICNC2010, 2010.11.17, Hiroshima University (Japan)

#### 6. 研究組織

##### (1) 研究代表者

川原崎 雅敏 (KAWARASAKI MASATOSHI)  
筑波大学・図書館情報メディア系・教授  
研究者番号：70375517

##### (2) 研究分担者

佐藤 聡 (SATO AKIRA)  
筑波大学・システム情報系・講師  
研究者番号：90285429