

機関番号：12605

研究種目：基盤研究(B)

研究期間：2008～2010

課題番号：20300025

研究課題名(和文) P2Pファイル共有ネットワークの測定と制御に関する研究

研究課題名(英文) Studies on Measurement and Control for P2P File Sharing Networks

研究代表者

川島 幸之助 (KAWASHIMA KONOSUKE)

東京農工大学・大学院工学研究院・教授

研究者番号：90345330

研究成果の概要(和文)：

Winny ネットワークを測定することにより、ファイルを持っているピア・持っていないピアの挙動の違いを明らかにし、参加・離脱のプロセスのモデル化を行った。Winny ネットワーク内のファイル流通制御として、検索リンクを制御することにより、ファイルを見つけにくくする手法と、ファイル検索のインデックス情報を改変することにより、ファイルを見つけにくくする手法を確立した。

研究成果の概要(英文)：

We measured Winny network to model behavior of peers with/without a file, and modeled join/leave process of peers. To control file distribution in Winny network, we developed two file control methods and confirmed the effectiveness. The one is controlling the search network to control file search query. Another one is the index poisoning for the file search.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	7,500,000	2,250,000	9,750,000
2009年度	4,200,000	1,260,000	5,460,000
2010年度	3,100,000	930,000	4,030,000
総計	14,800,000	4,440,000	19,240,000

研究分野：総合領域

科研費の分科・細目：情報学・ 計算機システム・ネットワーク

キーワード：ネットワーク計測，通信トラヒック，トラヒック制御，トラヒック分析，ピアツーピア

1. 研究開始当初の背景

インターネットの利用形態は日々変化し、トラヒック量はこの10年間で1000倍に達している。その主たる原因として、高速のインターネットアクセス回線、高性能PCの家庭での普及を背景とした、P2P(ピアツーピア)ファイル共有アプリケーションの利用者が増加したことがあげられる。さらに、P2Pフ

ァイル共有アプリケーションを用いた、著作権のある動画、音楽、ソフトの共有や、個人情報漏洩の問題が大きな社会問題となっている。現在のP2Pファイル共有アプリケーションソフトの抱える問題として下記があげられる。

問題1 巨大なトラヒックの生成による、他のアプリケーションへの影響。

- 問題 2 著作権のあるファイルの違法な共有。
 問題 3 個人情報の漏洩。

これら 3 つの問題が発生した原因として、ピア P2P ネットワークが管理者の存在しない自律分散型のネットワークのために、制御が不能に陥ってしまったことが挙げられる。これに対して、P2P ネットワークに対して能動的に制御を行うことにより P2P ネットワーク上を流れるファイルを制御できることを明らかにしてきた。具体的には制御ピアが P2P ネットワークに参加することにより、流出ファイルを検索できなくしたり、流通するファイルを破壊したりすることが可能であることを確認している。

Winny ネットワークに参加するユーザ数は、30 万人程度であることが測定されているが、本研究者らの詳細な解析により、Winny ネットワークは通常 8 万～11 万程度のピアで構成され、そのうちファイルをアップロードしているピアは 5～6 万程度であることを明らかにした。すなわち、P2P ネットワークの規模はかなり小さいことを確認している。

これらの事実から、プロトコルが公開されてしまえば、管理が難しいと言われてきた自律分散制御が行われているピア P2P ネットワークでさえ、現在の研究室で用意できるコンピュータシステムの能力で制御が可能であることが予想される。

2. 研究の目的

前述の 3 つの問題を解決するために、次の 3 つの研究課題に取り組む。

(1) ピア数、ピアの挙動、ファイル保持の状況の測定解析

よりよいネットワーク設計、制御を実現するためには、まず、最初に現状を把握する必要がある。近年の測定技術の向上により、大学の研究室レベルにおいても、高度なネットワーク測定が可能となってきた。P2P 技術はこれから発展が期待されている技術であるが、ピア（ユーザ PC）がサーバとクライアントの両方の役目を果たすことから、ユーザの“挙動”，すなわち、P2P ファイル共有システムにおいては、ユーザの参加と離脱、ファイルのアップロード/ダウンロード状況、検索クエリの出し方等により、システム全体としてのパフォーマンスに大きく影響する。本研究課題では、Winny を対象としてピアの挙動を測定により解析し、数式でモデル化をする。解析結果を用いて P2P ネットワーク制御方式を検討する。

(2) ファイル検索リンクの制御によるトラヒック、流通ファイル制御

制御ピアを P2P ファイル共有ネットワー

クに参加させ、ピアに対して検索リンクを張ることにより、実質、ファイル検索リンクを機能させなくすることが可能である。ファイルが見つかりにくくなることで、ファイルの拡散を抑えることが可能である。

複雑ネットワークの分野で明らかになってきている“ネットワークのもつ性質”等を利用することにより、すべてのピアに対してではなく、一部のピアに対して制御を行うことで有効な制御を行えることが考えられる。Winny を対象として、どのピアに対してどのような制御をすることで、ファイルの拡散に影響があるのかを確認し、有効な制御方式を明らかにする。

(3) ポイズニングによるトラヒック、流通ファイル制御

一度 Winny ネットワークに流出したファイルを消すことは不可能であった。しかし、オーバーレイネットワークではユーザの PC がネットワーク上の“ルータ”として振舞う。つまり、ユーザの PC はセキュリティ上、脆弱である。本研究課題では、P2P ファイル共有ネットワーク内に偽の情報を流すことで、実ネットワークでのファイル流通制御方式を確立する。

3. 研究の方法

(1) ピア数、ピアの挙動、ファイル保持の状況の測定解析

測定ピアを P2P ネットワークに参加させ、P2P アプリケーションのプロトコルを用いて内部の情報を得る。これまでの測定方法では、測定ピアの P2P ネットワークへの接続位置を変えることで測定を行っているが、接続したピア（図 1 のピア B）から得られる情報のみを採取していたため、測定速度が遅いという問題があった。測定方式を改良し 15 分ごとのネットワークのスナップショットを測定できるようにする。さらに、ファイル検索ネットワークのスナップショットからピアの参加・離脱の特性を解析する。特にファイルを持っているピアと持っていないピア（フリーローダ）の違いに着目する。

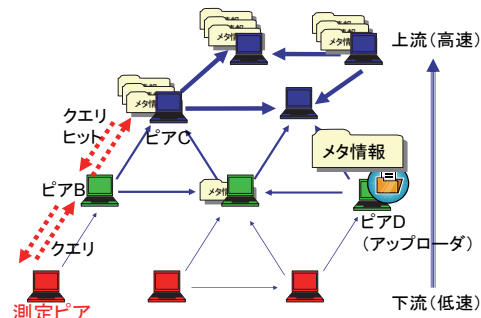


図 1 Winny の検索ネットワーク

(2) ファイル検索リンクの制御によるトラフィック，流通ファイル制御

これまで，流出ファイルを持つピアに対してファイル検索リンクを集中させることにより，対象ファイルの拡散を防ぐ方式の開発を行ってきた。Winny ネットワークでは，検索クエリを上流のピアに転送する。上流に制御ピアを配置し，クエリを制御し，ネットワーク上で見つかるファイルを減少させる方式を開発する (図 2)。これにより，ファイルの共有回数が減少し，混雑時のトラフィックの削減効果も期待できる。

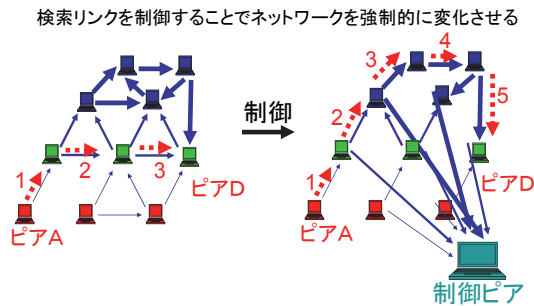


図 2 検索リンクの制御

(3) ポイズニングによるトラフィック，流通ファイル制御

Winny ネットワークにファイルが流出すると，削除機能がないため大きな社会問題となっている。ファイル検索のため，ピアがメタ情報である“キー”を持つ。制御対象となるファイルのキーを偽キーで書き換えることにより，ファイルの発見を困難にし，ファイルの検索効率を低下させることができる (図 3)。ダウンロード側ではファイルとしては存在するが，他のピアから検索されずに実質ネットワークから消失する。このような，Winny ネットワークに対するインデックスポイズニング手法を明らかにする。

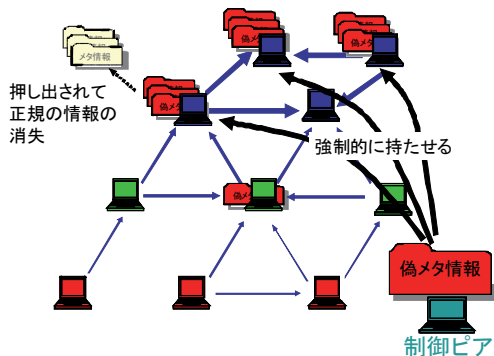


図 3 インデックスポイズニング

4. 研究成果

(1) ピア数，ピアの挙動，ファイル保持の状況の測定解析

測定ピアを P2P ネットワークに参加させ，P2P アプリケーションのプロトコルを用いて内部の情報を測定し，分析を行った。特にファイル検索ネットワークピアの参加・離脱の特性を解析した。

ピア数は，昼は，夜と比較し，30%少ない。ネットワークに参加している間にファイルを削除したり，アップロードを開始したりすることが分かった。特に朝になると，ファイルを持ったピアが多くネットワークから離脱することが分かった (図 4)。

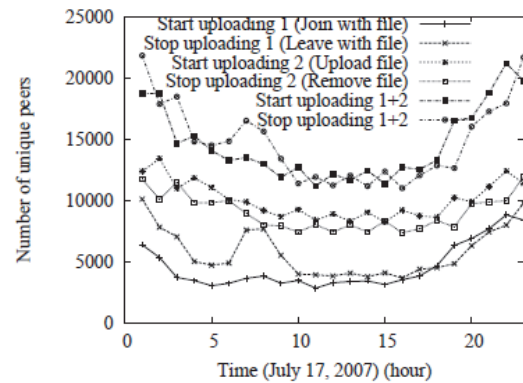


図 4 ピアの挙動 (アップロード開始・中止，参加・離脱)

ピアのネットワークへの参加過程が，ファイルを持っているピア，いないピアともに負の二項分布に従うことを明らかにした (図 5)。負の二項分布は，分散が指数分布と比較して大きいため，性能評価の際の入力として用いると大きな違いが生じることが考えられる。

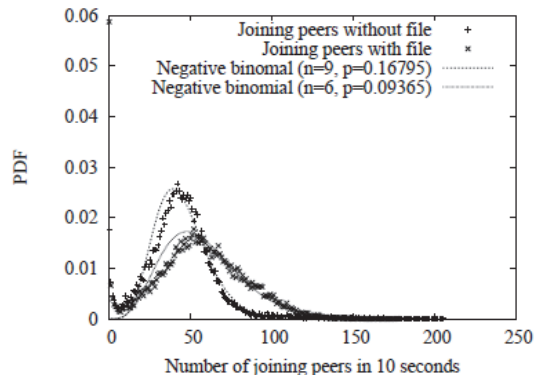


図 5 参加の過程 (ファイルあり・なし)

(2) ファイル検索リンクの制御によるトラフィック，流通ファイル制御

Winny ではファイルの位置情報であるファイルキーを発見しないとファイルをダウ

ダウンロードすることができない。そこで、ファイルダウンロードの前段階であるファイル検索に注目した。ダウンロードを行いたいピアは検索ネットワークに対し、検索クエリを発行しファイルキーを探す。提案方式では、この検索クエリを制御（ブロック）し、検索結果を得られない状態にすることでファイルダウンロードを不可能にする。検索クエリをブロックするためには一般ピアに対し検索リンクを張り、検索クエリを受け取る必要がある。提案方式は、一般ピアに対し検索リンクを張りにいく制御ピア群と、制御ピア群を管理する管理サーバから構成される。制御ピアは一般ピアに上流側から接続する。これは、Winny では下流側から上流側に検索クエリを発行するからである。また、Winny では検索クエリが6ホップ上流のピアまで転送されるという特徴がある。そこで、制御ピアは検索クエリが集中しやすい箇所に接続し、効率よく検索クエリを制御する。

そこで、提案方式では、制御ピアが接続しているピアから様々な指標の回数を計測し、これらの各指標の回数が大きければ大きいほど重要なピアであると判断する。管理サーバは制御ピアから各ピアにおける指標の数を集計し、ピアの重要度を順位付けする機構を持つ。制御に用いる指標は一般ユーザの行動特性（以下ピア特性と呼ぶ）とネットワーク構造に基づく特性に大別される。制御に用いる指標を以下に示す。

- (a) クエリ発行回数
- (b) ファイル所持数
- (c) 申告速度
- (d) クエリ転送回数

(a), (b) はピア特性である。Winny では、下流から上流側にクエリが転送される。そして、クエリには転送されてきた経路情報が記載されている。制御ピアではこのクエリ経路情報を分析することでクエリを発行したピア、転送したピアを特定する。(a)では各ピアがクエリを発行した回数を計測する。(b)ではファイルキーを分析し、キーに記載されているファイルの位置情報から各ピアのファイル所持数を特定する。ここでは、ファイル所持数を計測する。ファイルを多く持っているピアは多くのファイルキーを拡散する可能性があり重要である。(c), (d) はネットワーク構造特性である。申告速度が大きいピアは検索ネットワークの上流に存在する確率が大きくなる。つまり、検索クエリが集まる確率が高いピアといえる。(c)では、各ピアの申告速度を計測する。(d)では、クエリの経路から各ピアがクエリを何回転送しているかを特定する。クエリ転送回数が多いピアは他ピアのクエリを中継する回数が多

く、多くのピアのクエリをブロックできる可能性がある。ここでは、クエリの転送回数を計測する。

提案方式を実際に稼働する P2P ファイル共有ネットワーク Winny に適用し、その有効性と制御が P2P ファイル共有ネットワークへ与える影響を評価するために、Winny プロトコルに基づく各システムの実装を行った。検索ネットワーク制御システムは 7 台の制御ピアと 1 台の管理サーバによって構成される。制御ピアはマルチスレッドプログラムで構築され、制御ピア 1 台あたりのスレッド数は 2,000 本で、制御ピア 7 台のスレッド数の合計は 14,000 本である。なお、このスレッド本数は制御用マシンの性能による。制御ピアは一般ピアに検索リンクを張り検索を制御する一方、転送されるクエリやキーを解析し、前述した各指標により、制御対象ピアを順位付けする。

測定評価の結果、提案方式の適用により、クエリ転送回数が多いピアを上位から制御することで、検索成功率を 87%減少させることができた。Winny では、クエリが上流側のピアに転送される際、また帰還する際にファイルキーを経路上に複製する働きをする。つまり、検索クエリによってファイルキーが拡散することを意味する。提案方式の適用により、検索クエリ帰還率が減少すると、検索ネットワーク上にファイルキーが複製される確率は減少する。したがって、ファイルキーがヒットする確率も減少することになる。

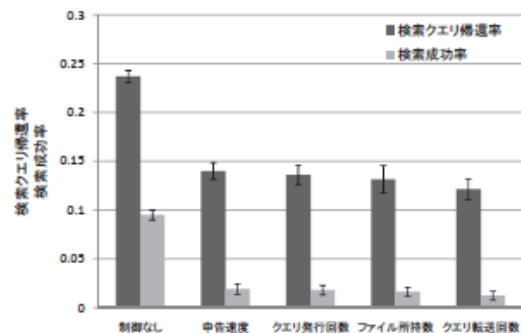


図6 クエリ帰還率、検索成功率

提案手法では検索ネットワークを制御することによって検索成功率を低下させ、ファイル発見までの時間を遅延させることができると考えられる。ここで、一般ユーザの検索をどの程度遅延させればよいのかを考察する。図6より、検索ネットワーク制御によって検索成功率を 87%削減できたので、ファイルキー発見までの時間を 7 倍程度にすることが期待できる。ファイル発見までの時間の実測値を 7 倍にしたものを図7に示す。ファイル発見までの時間が 7 倍になると、同時検索台数が 8 台以下の場合には 48 時間以内にファイルキーを発見できない可能性が高

くなることがわかる。

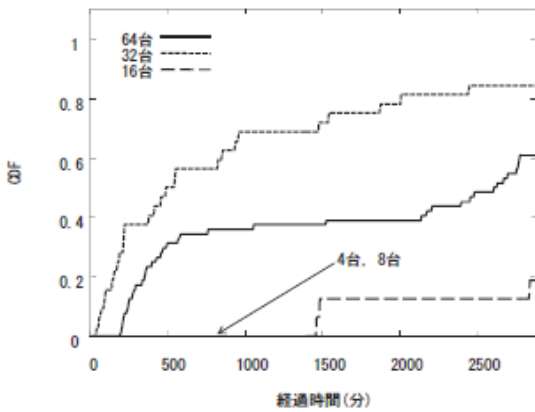


図7 ファイル発見までの時間(累積分布)

今後の課題として、本研究ではファイルID検索に対する制御の評価を行ったが、検索ヒットの条件が緩くなるキーワード検索に対しても評価を行う必要がある。また、Winy ネットワークの特性を解析したが、他のP2Pネットワークで共通した性質や現象を明らかにし、P2Pファイル共有ネットワークに対するファイル流通制御方式に活かすことが必要である。

(3) ポイズニングによるトラヒック、流通ファイル制御

提案手法では、制御対象ファイルの流通に必要な本物のキーを、Winy ネットワーク上から完全に「消去」する。本物のキーがWiny ネットワーク上から完全に消去された場合は、ファイル検索ピアが発行する検索クエリに本物のキーがまったくクエリヒットしなくなる。また、提案手法で拡散するダミーキーが、Winy ネットワークに与える影響を最小限にとどめるために、制御に用いたダミーキーをWiny ネットワーク上からすぐに消去する。一方、従来手法は、ネットワーク上に大量のダミーキーを拡散させ、ファイル検索ピアにダミーキーばかりをクエリヒットさせ、本物のキーをほとんどクエリヒットさせないことで制御している。従来手法は制御中に、本物のキーとダミーキーが両方ともネットワーク上に残り続けるが、提案手法では制御中に、本物のキーとダミーキーが両方ともネットワーク上に残らないという点で異なる。

Winy ポイズニングシステムを用いてネットワーク上にダミーキーを拡散し、制御対象ファイルの流通制御を行う。Winy ポイズニングシステムは、5台のWiny ポイズニングピアと1台のシステム管理サーバによって構成される。Winy ポイズニングピアには、

Winy ネットワーク上のアクティブピアのアドレスを収集する機能、アクティブピアに対してダミーキーを拡散する機能の2つの機能を実装している。Winy ポイズニングピアはマルチスレッドプログラムで構築され、5台のスレッド数の合計は3,000スレッドとなる。Winy ポイズニングシステムは、Winy ネットワークを構成する10万台以上のすべてのアクティブピアに対して、15分以内に接続を試行することができる。システム管理サーバはWiny ポイズニングピアが収集したユニークなアクティブピアのアドレスを管理し、それらを重複がないように5台のWiny ポイズニングピアに割り当てることで、システム全体が協調動作するように管理する。

評価実験の結果、提案手法をWiny ネットワークに適用した場合は、Winy ピアが特定のファイルのキーを入手できる確率が、適用しない場合の平均0.005%まで低下することを確認した(図8)。

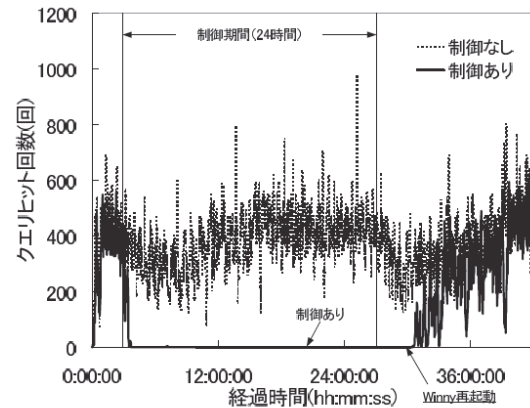


図8 クエリヒット回数

また、提案手法を用いた制御は、1,000個のファイルを同時に制御した場合であってもインデックスの汚染率は、5%程度であり、既存の方式と比べ、Winy ネットワークへの負荷が少ないことを示した(図9)。

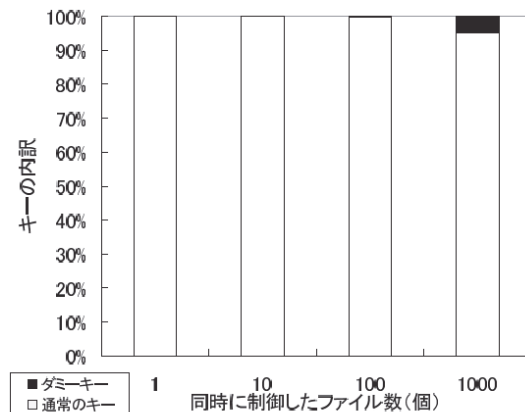


図9 インデックスの汚染率

今後は、インデックスポイズニングを実運用するための大規模分散制御用システムへの拡張，ならびに，多数のファイルを効率良く制御するための提案手法の改良を行う必要がある。

5. 主な発表論文等

〔雑誌論文〕(計 2 件)

① Satoshi Ohzahata, Konosuke Kawashima, "An Experimental Study of Peer Behavior in a Pure P2P Network," Journal of Systems and Software, 査読有, vol. 84, no. 1, pp. 21-28, 2011.

② 吉田雅裕, 大坐畠智, 中尾彰宏, 川島幸之助, "Winny ネットワークに対するインデックスポイズニングを用いたファイル流通制御方式," 情報処理学会論文誌, 査読有, vol. 50, no. 8, pp. 2008-2022, 2009.

〔学会発表〕(計 15 件)

① 中澤大暁, "ピア特性とネットワーク構造に着目した P2P 検索ネットワーク制御方式," 電子情報通信学会情報ネットワーク研究会, 20110120, 大阪大学.

② 大坐畠 智, "ネットワークへの情報流出対策技術の動向," 電子情報通信学会 2010 年ソサイエティ大会, 20100916, 大阪府立大学.

③ Hiroaki Nakazawa, "A Method for Controlling Search in a Pure P2P File Sharing Network," The 8th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT), 査読有, 20100617, マレーシア.

④ 渡部 友也, "An Analysis of Peer and Query Behavior in Winny Network by Passive Measurement," 2010 IEICE General Conference, 20100318, 東北大学.

⑤ 平野 誠, "P2P ファイル共有ネットワークのグラフ的特徴の解析および評価," 第 72 回 情報処理学会 全国大会, 20100309, 東京大学.

⑥ 中澤大暁, "Winny ネットワークにおける検索ネットワーク制御方式," 電子情報通信学会情報ネットワーク研究会, 20101115, 広島市立大学.

⑦ 上野真弘, "Winny ネットワークにおける高次数をもつ非協力ピアを用いたトラヒック制御," 電子情報通信学会ネットワークシステム研究会, 20091016, 熊本県立大学.

⑧ 渡部友也, "Winny ネットワークにおける申告速度に注目したピアとクエリの挙動の定量的分析," 電子情報通信学会ネットワークシステム研究会, 20091016, 熊本県立大学.

⑨ Satoshi Ohzahata, "A Measurement

Study on Peer Behaviors for a Pure P2P Network," 2009 International Conference on Advanced Information Networking and Applications (AINA 2009), 査読有, 20090526, イギリス.

⑩ プトゥラプラタマ, "ファイル検索クエリを用いた Winny ネットワークのトポロジ測定方式," 電子情報通信学会ネットワークシステム研究会, 20090416, 芝浦工業大学(東京).

⑪ 中澤大暁, "ピュア P2P アプリケーションにおけるファイル検索制御法の提案とその測定分析," 電子情報通信学会ネットワークシステム研究会, 20090416, 芝浦工業大学(東京).

⑫ Pratama Putra, "A Measurement Method for Capturing Winny Network Topology," 2009 IEICE General Conference, 20090319, 愛媛大学.

⑬ Masahiro Yoshida, "Controlling File Distribution in Winny Network Through Index Poisoning," 2009 International Conference on Information Networking (ICOIN 2009), 査読有, 20090123, タイ.

⑭ 吉田 雅裕, "Winny ネットワークにおけるインデックスポイズニングの適用と評価," 電子情報通信学会ネットワークシステム研究会, 20080912, 東北大学.

⑮ 上野 真弘, "高次数をもつ非協力ピアの参加が非構造化 P2P ファイル共有ネットワークへ及ぼす影響の評価," 電子情報通信学会ネットワークシステム研究会, 20080911, 東北大学.

〔その他〕

ホームページ等

<http://www.tuat.ac.jp/~kkawa/>

6. 研究組織

(1) 研究代表者

川島 幸之助 (KAWASHIMA KONOSUKE)
東京農工大学・大学院工学研究院・教授
研究者番号: 90345330

(2) 研究分担者

大坐畠 智 (OHZAHATA SATOSHI)
電気通信大学・大学院情報システム学研究科・准教授

研究者番号: 30361744

(H20→H21: 連携研究者)

(3) 連携研究者

中尾 彰宏 (NAKAO AKIHIRO)
東京大学・大学院情報学環・准教授
研究者番号: 60401238