

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 24 日現在

機関番号：17104

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330108

研究課題名(和文) 広域多点間データ転送・共有のためのSDNによる空間・時間・符号化連携制御手法

研究課題名(英文) SDN-based space-time scheduling with information coding on data transfer/sharing over distributed sites

研究代表者

鶴 正人 (Tsuru, Masato)

九州工業大学・大学院情報工学研究院・教授

研究者番号：40231443

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：広域多地点間データ共有のための大容量データ転送における空間・時間・符号化の3つの制御の効果的連携・統合を目的として以下の研究開発を実施した：(a)多地点間大容量データ共有向け中継転送制御、(b)グループ内共有の目的に応じた性能指標および複数グループ間の公平性のモデル化・制御、(c) (a)の一部である複数経路マルチキャストを用いる1対多および多対多ファイル転送やその関連手法のOpenFlow・Linuxシステム上での試作と国内広域テストベッド(JGN-X/RISE)上での実験的評価。

研究成果の概要(英文)：We aim at an effective integration and collaboration of control in space, time, and information coding on large data transfer for distributed data sharing over multiple sites. The outcome includes (a) routing control for massive data sharing over multiple sites; (b) performance metrics for intra-group resource sharing and a fairness modeling and control on inter-group resource sharing; and (c) a proto-type implementation of the multipath-multicast-based one-to-many and many-to-many file transfer schemes by OpenFlow/Linux system as a part of (a) with experimental evaluations on a nation-wide testbed (JGN-X/RISE).

研究分野：情報通信ネットワーク

キーワード：ファイル転送 多対多通信 複数経路転送 マルチキャスト転送 OpenFlow ネットワーク計測 ネットワーク符号化 DTN

## 1. 研究開始当初の背景

(1) 近年、広域ネットワーク（インターネット等）上に分散したエンドノード（データセンタ等のサーバ）の間で相互に大容量データを転送して共有する形態が急増している。同じデータを複数地点に置く目的は、分散計算・検索、キャッシュによるアクセス性能向上（遅延時間低減とトラフィックピークシフト）、バックアップによる耐障害性向上等、様々であるが、クラウドサービスの普及、IoT (Internet of Things) やビッグデータ活用の拡大、さらにインターネットの将来形態の1つとしての情報コンテンツ指向のネットワークの導入等に伴い、広域多地点間大容量データ転送・共有の更なる増大が予想され、その効率的な実現が急務である。これを実現するためには、多地点間大容量データ共有におけるネットワーク資源利用を全体最適化する高度な中継転送制御機能をネットワークが提供する必要がある。

(2) 高度な中継転送制御は、主に空間領域、時間領域、情報符号化領域に分けられる。例えば、送信ノードから複数受信ノードに同じデータを一斉配布する場合、一般に、各受信者は接続トポロジや帯域に関して不均一なため、それらの受信完了時間は様々であり、その状況で各受信者にとっては自身の完了時間が短いことが望ましい。この時、多数の受信者への同時並列ユニキャスト転送は送信者負荷を最大化するだけでなく、受信完了時間の面でも有効ではない。全受信者への単一木によるマルチキャスト転送（本研究ではネットワーク層（以下）の仕組みを指す）は送信者負荷を最小化するが、全受信者の受信完了時間が等しくなり、送信者との間のネットワークが良好で本来は早く受信が完了する受信者にとっては望ましくない。一方、受信者が1人の場合には全資源の独占利用、つまり、送信者からの受信者への複数経路による最大フロー転送が転送完了時間を最小化する。しかし受信者が複数人居る場合は、送信者からの受信者毎の最大フロー転送の同時並列実行は、最大フロー経路間に重なりがあるために必ずしも有効ではない。さらに、送信者が1人の1対多転送だけでなく、一般にデータ共有グループの各メンバからグループ内の他メンバへの多対多転送が発生し、同時実行のために複数の1対多転送を効率的に共存させる経路の空間・時間制御が重要になる。一方、多段階（Multi-round やパイプライン）のノード間転送によるグループ内データ共有を *gossiping* と呼ぶ。このような意味の中継転送では複雑な時間制御（スケジューリング）が必要になる。最後に、実効帯域幅増加や前方消失訂正のための送信者符号化や中継符号化 (Network Coding)、あるいは圧縮を符号化制御と呼ぶ。これらはマルチキャストや複数経路通信においても様々な適用が研究されている。なお、ノード間でのデータ共有の形態に関して、ネットワーク内

部の中継ノードも含めた共有も研究されているが、本研究課題では、エンドノード群がグループを形成し、それらのグループに属するノードがデータを共有するために中継ノードが転送や蓄積等の機能を持つ形態を想定する。

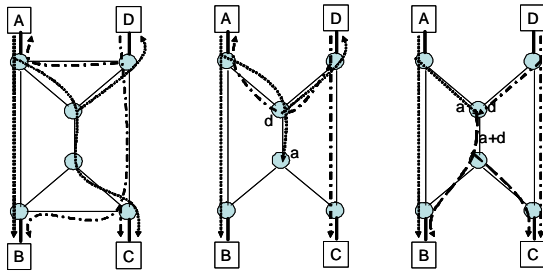
(3) ここで「最適化」は、目的達成に必要な性能を、制約されたネットワーク資源で実現することであるが、性能として、単なるファイル転送完了時間やスループットだけでなく、むしろグループ内のデータ共有達成度が重要である。この際、共有の目的・条件に応じて時間スケールや不完全な共有状態への許容度が異なり、性能指標は共有の達成度とその利得との関係に依存する。さらに、複数グループが共存する場合、仮に見かけは別々の仮想ネットワークを利用していても物理的ネットワーク資源を共有するので、性能指標にも依存した「公平性」が重要となる。

## 2. 研究の目的

(1) そこで本研究課題では、広域に分散したエンドノード（グループメンバ）から構成される通信グループ内での大容量データ共有の効率化のために、空間領域、時間領域、情報符号化領域での中継転送制御を効果的に連携・統合することを目的とする。具体的には、以下の副技術課題：(i) 多地点間大容量データ共有向けの各種中継転送制御とその連携・統合手法、(ii) グループ内共有の目的・条件に応じた性能指標および公平性のモデル化・制御、さらに、複数グループが同一ネットワーク資源を競合利用する場合のグループ間の公平性のモデル化・分析手法、に関する研究開発を目指す。一方、(iii) 手法の一部を実機上に試作し、広域のテストベッド上で実験し、その実用性を分析・評価する。

(2) グループ通信の効率化のための空間制御（相互マルチキャストを含む経路制御）、時間制御（多段階の蓄積・転送による中継）、符号化制御（中継符号化や中継時圧縮）には様々な既存研究が存在するが、前提条件や性能指標が各々で異なるため、それらの効果的な組み合わせや公平な資源割当ての制御は自明ではない。(i)の概念的な例を示す。相互マルチキャスト木の経路、2段階の蓄積・転送、ボトルネックリンク手前での中継符号化の例である。A、B、C、Dの4つのエンドノードは同一長のファイルを1つずつ持ち（各々、a、b、c、d）、6つの中継ノードを経由して接続される。リンクは全二重で、中継ノード間はすべて同一帯域幅、エンドノードとの間の帯域幅はそれに比べて十分大きく、背景トラフィックはない。各エンドノードは自分の持つファイルを他の3ノードに転送する。次図の3つの例は、AおよびDから他ノードへのファイル転送での最初の時間区切り（ファイル長÷中継リンク帯域幅）までの転送経路を示す。BおよびCから他ノードへの経路は明記していないが対称性より

確定する。

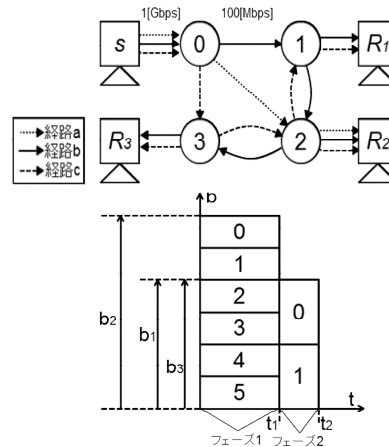


左例は、各送信者から適切な4つの木経路に沿ってマルチキャスト通信だけで4ファイルが競合なく転送できる。しかし、中・右例は中継リンクが2個少ないので、もしこのトポロジ上で単純に相互マルチキャスト通信を行うと、ボトルネックリンクで競合し、3ファイルの内の2つは2倍の時間が掛かって転送完了する。そこで、中例では蓄積・転送スケジュールを適用する。ファイル a は、最初の時間区切りのマルチキャスト通信により B と D に直接転送され、また中央下の中継ノードにも蓄積され、次の時間区切りでそこから C へ転送される。一方、右例では中継符号化も併用する。ファイル a は、最初のマルチキャスト通信により B にはそのまま、C に向けては、途中で a+d となり、C の手前で a+d と d (D から来る) より a が復号される。ただし、中央上の中継ノードでは a と d の構成単位 (パケット等) の排他的論理和による合成が B と C に配送される。次に中央上の中継ノードに蓄積された a が D に転送される。中・右例を受信者の立場から比較すると、共に3ファイルの内の1つだけが2倍の時間掛かるが、受信完了が遅れるファイルは各例で異なる。

### 3. 研究の方法

(1) 上記の (i) および (ii) の2つの副技術課題に関するこれまでの研究成果を拡張し、シミュレーションによる分析・評価を用いながら研究開発を進める。(i) に関して、まず固定・広帯域ネットワーク上での1対多および多対多の大容量ファイル転送を想定し、ネットワーク帯域を最大限利用し、一対多ファイル転送における各受信者の完了時間を短縮するために、ファイルを適切な大きさの均等なブロックに分割し、異なるブロックを複数経路で同時転送するとともにマルチキャストにより無駄な重複転送を減らす手法を検討する。1つのファイルを複数受信者へ複数フェーズに分けて転送し、トポロジや帯域に応じて、ファイルのブロックへの分割、各フェーズの複数受信者への転送経路、各経路で転送するブロックの割当、という3種類の時間・空間割当てを決定する。前提として、この手法では、完全な集中制御によってトポロジを把握し、転送を始める前にすべての受信が完了するまでの全体割当てを計算し、送信者へ転送スケジュールを指示し、ネットワーク上のス

イッチへ経路設定を指示する。次図の単純例を説明する。

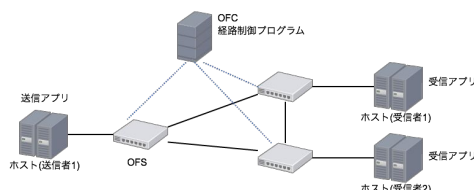


ホスト (S, R1, R2, R3) とスイッチ (0, 1, 2, 3) 間が 1000 [Mbps]、スイッチ間が 100 [Mbps] で接続されており、送信者 S が受信者 R1, R2, R3 に対して同一のファイルを転送する。S から各受信者への最大フローは各々、 $b_1=300$ ,  $b_2=200$ ,  $b_3=200$  [Mbps]、ファイルのブロック分割数は6になり、ブロックを転送するためのスケジュールは図下のようになる。フェーズ1では、優先受信者 R1 は全ブロックを受信完了し、R2, R3 はブロック 0, 1 が未受信となる。フェーズ2では、優先受信者 R2、非優先受信者 R3 とともに全ブロックを受信完了する。各受信者の受信完了時間は各々の下限値と等しく、完璧スケジュールになっている。また、送信者が送信する総データ量のファイル長に対する比 (ロスによる再送は考えない場合) は、各受信者への単一経路ユニキャストの同時並列の場合は同一データを3回送るので3、全受信者への単一木マルチキャストの場合は同一データを1回しか送らないので1であるのに対し、本方式では  $4/3$  となり、受信完了時間を短縮しながら、重複転送も少ないことがわかる。

(2) (i) および (ii) に関連する他の要素技術として、1対多および多対多ファイル転送における複数指標の扱い (最長完了時間、平均完了時間) や公平性分析、ネットワーク符号化の組み込み、ネットワーク符号化付き TCP における再送手法の改良、多段ホップ無線ネットワークでの伝送スケジュールとネットワーク符号化の組み込み、車の移動等の交通網を利用する DTN (蓄積運搬形転送) の枠組みでの大容量データ中継転送の最適化、異種要求を持つグループからなる全体の公平性のための資源割当て理論、経路を自由に設定できるネットワークにおけるリンク状態の実時間把握のための全双方向リンク計測手法や、大規模ネットワークにおける多地点間フロー流量統計 (Origin Destination Traffic Matrix) の推定手法等の検討を行う。

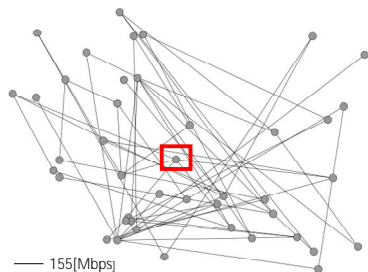
(3) 一方、(iii) に関して、ネットワークにプログラマブルで高度な中継機能を持たせる枠組みとして SDN (Software Defined

Network) が提唱され、その一部を実用化した OpenFlow が注目を集め、データセンタの内部 LAN だけでなく、データセンタ間の世界規模ネットワークにも利用され始めた。OpenFlow は、オーバレイと異なり、ネットワーク内部機能として高速な転送性能と柔軟な経路制御の両立を可能にする。さらに OpenFlow スイッチに PC 等 ( 計算機 ) を連結することで、データ蓄積・加工可能な中継ノードを実現できる。そこで、(i) の一部である複数経路マルチキャストを用いる 1 対多および多対多ファイル転送手法の OpenFlow・Linux システム上での試作を行い、それを用いて NICT が運用する国内広域テストベッド ( JGN-X/RISE ) 上での実験的評価を行う。



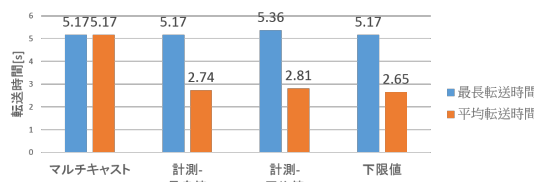
#### 4. 研究成果

(1) (i) の基本部分として、複数経路マルチキャストを用いる 1 対多および多対多ファイル転送手法を提案し、モデル上のシミュレーションによって大規模ネットワークに対しても有効であることを確認した ( 雑誌論文 [1, 4, 5, 13] )。これは、トポロジや帯域に応じて、ファイルをブロックへ分割し、複数の時間フェーズにおける複数受信者へブロック転送経路を決め、各経路に転送するブロックを割当てする。さらに複数ブロックを符号化して 1 つのブロックに合成する場合もある。現実の不均一で複雑なトポロジやデータ共有条件に対して、このような空間・時間・符号化制御を組合せた最適解を求めることは、静的な条件下でも困難である。そこで、ここでは基本的・汎用的な全体手順を設計し、そこから実際のメンバ間転送やその中継の制御を決定・実現するための入力パラメタを「解」として扱い、良い解を見つける反復的なアルゴリズムを設計し、その中で何らかの尺度でのパレート境界を探索した。



図は RENATER と呼ばれるトポロジで、ファイルサイズ 100 [MByte]、送信者 1 人 ( 赤棒 )、受信者 42 人 ( スイッチとの接続帯域は無制限 )、スイッチ間帯域は全て 155 [Mbps] の例で、この場合、ファイル分割数は 12 になる。42

人の受信者中の最長転送完了時間と平均転送完了時間で性能を比較したグラフを示す。「下限値」は、各受信者が自分だけがネットワークを利用して最大フローを用いて受信できると仮定した場合の各転送完了時間の最長と平均である。グラフにはない「同時並列ユニキャスト」はあちこちにボトルネックリンクができ、最長 110.97 [s]、平均 56.87 [s] となり全く使えない。グラフより、最長では提案手法もマルチキャストも下限値を達成するが、平均では提案手法はマルチキャストの半分の時間で済み、下限値に近い。



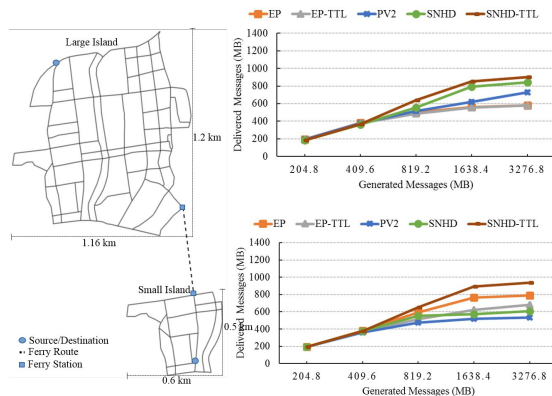
同じネットワークで提案する多対多転送を行うと、上の 1 対多転送の逐次実行に比べても更に転送時間の短縮が図れることが次表から判る。

	最長完了時間	平均完了時間
一対多(理論値)	1223.2	1179.4
提案手法(理論値)	1001.3	882.6
下限値	539.4	466.9

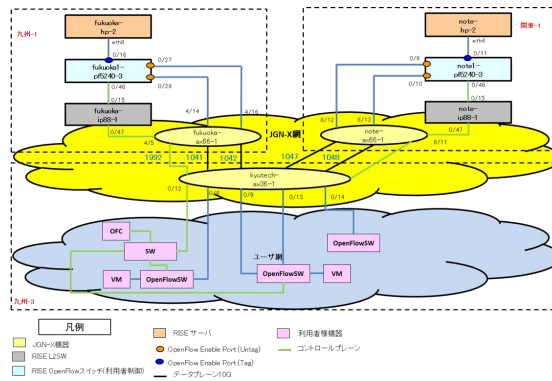
(2) (i) および (ii) の関連技術として、複数グループ競合のモデル化や性能指標に基づく公平性の定義のために、優先度付き公平性を持つ資源割り当ておよび異種の要求を持つグループからなる全体の公平性のための資源割り当ての理論を構築し、シミュレーションによって有効性を示した ( 雑誌論文 [9, 11] )。さらに 1 対多、多対多ファイル転送のスケジュール探索手法にその知見を活かし、また多様な性能指標や公平性に対応できるように、パレート境界を検索する枠組みを組み込んだ。また、必ずしも高速広帯域な通信ネットワーク基盤が整備されていない環境や地域での大容量データ共有を想定し、野外環境センシングを想定した送電鉄塔等の点在固定設備を利用した無線マルチホップ通信での伝送スケジュール最適化やネットワーク符号化の組み込み、離島間での大容量ファイル配布を想定した、島内の車や島間のフェリーの移動を利用した蓄積運搬形無線マルチホップ通信での中継転送最適化、それらの環境におけるパケットロスが多いノード間での確実なデータ転送のためのネットワーク符号化付き TCP における再送手法の改良に関する手法の提案と効果の確認を行った ( 雑誌論文 [2, 3, 7, 8, 10] , 学会発表 [1, 2, 4, 5] )。さらに、経路を自由に設定できるネットワークにおけるリンク状態の実時間把握のための全双方向リンク計測手法の改良や、大規模ネットワークにおける多地点間フロー流量統計 ( Origin Destination トラヒック行列 ) の推定手法の改良を行い、OpenFlow 上に試作し、有効性を実験的に評価した ( 雑誌論文 [6, 12] ,

学会発表[3])

次図は、別の島にある2点間で複数の車やフェリーを使ってデータを運ぶ手法をシミュレーション評価した例である。親島から子島(L2S: 図上)とその逆(S2L: 図下)の方向で送信データ量とそのうち制限時間以内に受信者に届いたデータ量をグラフにしている。送信データ量が少なく輻輳が起きていない(400MB まで)場合は手法間に差はなく送信データをほぼ受信できるが、送信データが多くなると提案手法(SNHD-TTL)が他の既存手法より優ることがわかる。



(3) (iii)として、広域テストベッド JGN-X/RISE 上の福岡および東京の OpenFlow ネットワークと大学内の OpenFlow ネットワークを接続した環境において(次図)、(1)の複数経路マルチキャストを用いる1対多ファイル転送や(2)の全双方向リンク計測の試作システムを実験的に検証した。



(4) これらの研究開発を通して、広域多地点間データ共有のための大容量データ転送における空間・時間・符号化の制御の効果的連携・統合を検討し、OpenFlow等を活用した複数経路マルチキャスト転送の基本手法や各種機能を統合する枠組みを確立し、試作システムの実験により有効性を示し、課題抽出を行なった。現時点での手法は、まだ様々な制約条件・前提条件を必要とするものであるが、理論的・汎用的な枠組みをより拡張することは可能であり、また実機による試作は実用可能性を示しており、今後の発展が期待できる。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

(雑誌論文)(計13件)

- [1] 鶴 正人, 岩本 健志, "ネットワーク制御における資源共有と公平性," 信学論 J99-B(10), 11 pages, 査読有, 2016年10月発行予定.
- [2] Agussalim, Masato Tsuru, "Spray Router with Node Location-Dependent Remaining TTL-Based Message Scheduling in DTNs," IPSJ JIP 24(4), 13 pages, 査読有, to appear in August 2016.
- [3] Agussalim, Masato Tsuru, "Message Transmission Scheduling on Tandem Multi-hop Lossy Wireless Links," Proc. WWIC'16, LNCS 9674, pp.28--39, Springer, 査読有, April 2016. (10.1007/978-3-319-33936-8\_3)
- [4] 岡本洋平, 岩本健志, 小川恭平, 鶴 正人, "複数経路マルチキャストを利用した多対多ファイル転送システムの試作," 信学技報 CQ2016-11, pp.61-66, 査読無, 2016年4月.
- [5] 岩本 健志, 岡本洋平, 鶴 正人, "OpenFlow による複数経路マルチキャストを利用した一対多ファイル転送の分析と改良," 信学技報 NS2015-167, pp.121-126, 査読無, 2016年1月.
- [6] 月岡祐太, 鶴 正人, "OpenFlow ネットワークでの全リンクパケットロス率計測の効率化," 信学技報 IN2015-84, pp.77-82, 査読無, 2015年12月.
- [7] Nguyen Viet Ha, Kazumi Kumazoe, Masato Tsuru, "TCP Network Coding with Forward Retransmission," Proc. IEEE APWiMob'15, pp.136-141, 査読有, Aug 2015. (10.1109/APWiMob.2015.7374970)
- [8] Agussalim, Masato Tsuru, "Node Location Dependent Remaining-TTL Message Scheduling in DTNs," Proc. IEEE APWiMob'15, pp.108-113, 査読有, Aug 2015. (10.1109/APWiMob.2015.7374945)
- [9] Mario Koeppen, Masato Tsuru, "Zonal Fairness in Wireless Infrastructure," Proc. IFIP NTMS'15, 5 pages, 査読有, July 2015. (10.1109/NTMS.2015.7266534)
- [10] Fitriyani Umar, Agussalim, Masato Tsuru, "Message Scheduling for Tandem Multi-hop Wireless Networks," Proc. IEEE APWiMob'14, pp.201-207, 査読有, Aug. 2014. (10.1109/APWiMob.2014.6920287)
- [11] Mario Koeppen, Kei Ohnishi, Masato Tsuru, "Maxmin Fairness under

Priority for Network Resource Allocation Tasks," Proc. IEEE ADMNET'14, pp.49-54, 査読有, July 2014. (10.1109/COMPSACW.2014.12)

- [12] 鶴 正人, "ネットワーク計測の楽しみ ~部分から全体へ、検査・診断から予防へ~ [招待講演]," 信学技法 CQ2013-59, pp. 45-48, 査読無, 2013年11月.
- [13] Akira Nagata, Yoshiaki Tsukiji, Masato Tsuru, "Delivering A File by Multipath-Multicast on OpenFlow networks," Proc. IEEE INCoS'13, pp.835-840, 査読有, Sept. 2013. (10.1109/INCoS.2013.160)

〔学会発表〕(計5件)

- [1] Agussalim, Masato Tsuru, "Static Slot Assignment with Transmission Scheduling on Tandem Multi-hop Lossy Wireless Links," 信学会ソサエティ大会 BS-6-33. (2015年9月10日, 仙台市)
- [2] Nguyen Viet Ha, Kazumi Kumazoe, Masato Tsuru, "Forward Retransmission for Network Coding with TCP in burst loss channel," 信学会ソサエティ大会 BS-6-25. (2015年9月10日, 仙台市)
- [3] Shota Yamashita, Masato Uchida, Masato Tsuru, "Inferring Origin Destination Traffic Matrix with Partial Observation on OpenFlow Networks," 信学会 総合大会 BS-3-38. (2015年3月12日, 滋賀県草津市)
- [4] 鶴 正人, "DTN は何処へ向かうのか?," 情報処理学会 第13回情報科学技術フォーラム(FIT2014), DTN 最前線 ~時空間を超えてデバイスを紡ぐ新しい情報基盤へ~. (2014年9月3日, つくば市)
- [5] 鶴 正人, "DTN の進化と浸透 [依頼講演]," 信学会 総合大会 D ソサエティ・シンポジウムセッション. (2014年3月20日, 新潟市)

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

<http://www.ndrc.kyutech.ac.jp/> (ネットワークデザイン研究センター・ホームページ)

## 6. 研究組織

### (1) 研究代表者

鶴 正人 (TSURU MASATO)

九州工業大学・大学院情報工学研究院・教授

研究者番号: 40231443