

## 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 24 年 6 月 6 日現在

機関番号：17104

研究種目：基盤研究(B)

研究期間：2009～2011

課題番号：21300024

研究課題名（和文） 大規模広域ネットワーク上での大容量ファイル高速転送のための中継サービス基盤技術

研究課題名（英文） A relayed file transfer framework to reduce the response times of large file transfer on a large-scale wide-area network

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

研究成果の概要（和文）：大規模広域ネットワーク上で、IP パケット転送の上位で高機能中継ノード間の蓄積形ファイル中継転送を行い、多数ユーザの長大ファイルの同時転送の応答時間を短縮する枠組みを検討した。要素技術として、送信・中継・受信ノード間での高速公平なデータ転送プロトコル、ファイル中継の時間・空間的スケジューリング、中継ノードの適切配置、に関する手法を考案し、シミュレーションや、PC と Openflow を用いたプロトタイプ実験によって評価した。

研究成果の概要（英文）：We have discussed a framework to reduce the response times of simultaneous large file transfers by many users in a large-scale wide-area network. In that, on top of IP packet routing, the files are stored and relayed among special advanced relay nodes arranged on the network. A variety of essential techniques in terms of a high-speed fair data transfer between sender/relay/receiver nodes, a temporal and spatial file relay scheduling, and an effective arrangement of the relay nodes were proposed and then evaluated through simulations and/or PC and Openflow-based prototype experiments.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	4,400,000	1,320,000	5,720,000
2010年度	5,000,000	1,500,000	6,500,000
2011年度	2,500,000	750,000	3,250,000
年度	0	0	0
年度	0	0	0
総計	11,900,000	3,570,000	15,470,000

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：トラフィックエンジニアリング、高速 TCP、ファイル転送、スケジューリング、高機能中継ノード、Openflow

## 1. 研究開始当初の背景

(1) インターネットは、アクセス回線の高速化とそれに対応するためのバックボーン回線の増強の繰り返しによって、その転送能力が大幅に向上してきたが、それでも、近年の YouTube などの大容量動画ファイルダウン

ロードの増大は、ネットワークプロバイダやその上でサービスを提供するアプリケーションプロバイダにとって火急の課題になってきている。実際、日本のバックボーン上の通過トラフィックは年々増加し、既に毎秒テラビットに達している。さらに今後、無圧縮ハ

イビジョン伝送や大規模科学計測データの広域共有・交換など、大容量データの高速転送への要求は更に高まると予想され、ネットワーク回線の増強や、アプリケーションレベルでの転送方式や従来の転送経路の工夫では太刀打ちできない。

(2) 今のインターネットはユーザ間（エンドツーエンド）制御によるファイル転送を行っているが、それは、個々の連続転送単位（通信フロー）のレートや継続時間が比較的小さく、互いに独立な膨大な数のフローが共存する状況を想定しているからであり、一部の通信フローが高レートで大容量データを転送し、それらが競合するような場合には、個々のファイル転送完了までの遅延時間（応答時間）が増大し、また、全体の資源利用効率も低下する場合がある。

この問題の解決のためには、ネットワーク側のパケット単位での経路制御と各ユーザ側のエンドツーエンドでの転送制御だけでは限界があり、ファイルまたはそれを構成する転送単位での資源利用・資源競合に関するネットワーク側の能動的な空間的制御（経路選択、マルチパス制御など）および時間的制御（転送単位でのスケジューリング）の総合的な適用が有効と考えられる。

## 2. 研究の目的

(1) 上記の問題を解決するために、本研究課題では、大規模広域ネットワーク内部の要所要所に高機能中継ノード（ファイル中継ノード）を置き、一般の IP パケット転送の上位のファイル中継ノード間で蓄積形中継転送を行い、多数ユーザの長大ファイルの同時転送を効率的に実現する枠組みを検討して、その可能性・有効性を示す。一般に、中間ノードでの中継（store-and-forward）は、中継しないで直接転送する場合（cut-through）に比べて、中間ノードでの蓄積にかかる分の転送時間が増え、それに伴うコストも掛る。しかしながら、ノード間の転送方式の自由度が増し、またネットワーク全体の時間・空間的利用の最適化が図れる。

(2) 想定する形態は以下の通り。大容量ファイルを転送したいユーザは送信ノードから直接受信ノードへファイルを転送せず、入口中継ノード（ファイル中継サービス）へファイルを転送する。入口中継ノードはファイルを受信しながら必要なら適切な大きさの転送単位に分割し、経路やタイミングをスケジュールしながら0個以上の中継ノードを経由して出口中継ノードへ転送する。出口中継ノードは最終受信ノードに対して、1ファイルとして見えるように転送を行う。

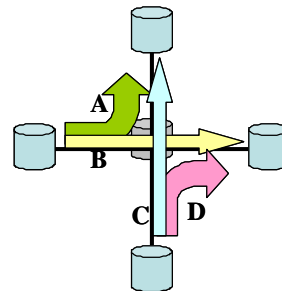
(3) 具体的には、このような形態を用いて多数ユーザの長大ファイルの同時転送を効率よくネットワーク上で共存させ、それらのファイル転送の応答時間を短縮する「ファイル中継サービス」を実現するための3つの要素技術：(i) 送信・中継・受信ノード間での高速公平なデータ転送プロトコル、(ii) ファイル中継の時間・空間的スケジューリング、(iii) 中継ノードの適切配置、に関する手法を相互の連携も踏まえて考案・評価することを目的とする。

## 3. 研究の方法

(1) 上記の3つの要素技術に関するこれまでの研究成果を拡張して研究開発を進める。

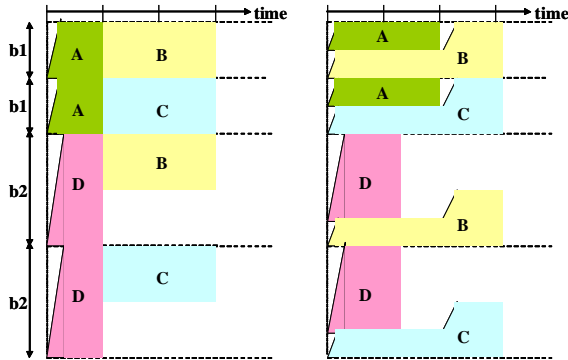
(i) では、長距離広帯域回線上での大容量ファイルの転送、特に複数フローが競合する場合の問題はよく知られており、送受信ノード間の距離などに起因する制御遅延が精密な送信レート制御を困難にし、送信レートが高すぎると輻輳の発生に起因する多量のパケット再送によって無駄に帯域を消費したり、逆に送信レートが低すぎると帯域を十分に活かせなかったりする。それによって、非効率だけでなく不公平性も顕在化する。さらに、無線通信を含む場合などに、高速であるが不安定であったり、往路と復路の特性が異なったりすることで、従来のTCPが有効に機能しなくなる。中継ノードを置くことでノード間距離が短縮され、制御遅延が改善できる場合があるが、それだけでは解決できず、ネットワークの状態の推定に基づくTCPの改良や、TCPとは別の複数の異なる通信路を統合的に利用するトランスポートプロトコルを検討する。

(ii) と (iii) に関しては、最も基本的なアイデアは、蓄積中継によって、競合を避けつつリンク帯域を最大限有効利用することで、平均応答時間を短縮する点である。



単純な例として、上図は4つのノード間でフローA（左→上）、B（左→右）、C（下→上）、D（下→右）を転送することを示しており、Aのフロー長は、B、C、Dの半分、ネットワークの中央と左および中央と上の間の帯域幅(b1)は、右および下との間の帯域幅(b2)の半分、とした場合、適切なスケジュー

リングが下図左で、通常の4フロー同時並行転送が下図右である。前者の方がすべてのフローにおいて転送終了時刻が早い。また、(i)との関係において、フロー転送途中で他の競合フローが終了あるいは開始すると、送信レート制御（特に開始時）の不完全さから転送効率が悪くなる。



これらのスケジューリングは発生するファイル転送要求に応じて動的に行う必要があるが、完全な情報に基づく集中形計算によって最適解を見つけることは大規模ネットワークにおいては困難であり、どのような情報がどの時点で利用できて、それに基づく制御を行うのか、が重要になる。

(2) これらの技術に関して、制約論理プログラミングや整数計画法なども援用した理論的解析・考察に基づいて手法を検討し、計算機シミュレーションで効果を分析する。さらに、proof-of-concept としてのLinux PCやOpenflowを用いたプロトタイプ実験を、商用インターネットや NICT が提供するJGN2plus/JGN-X 高速ネットワークテストベッド、あるいは高速インターネット実験衛星 WINDS などの上で実施し、分析・評価する。

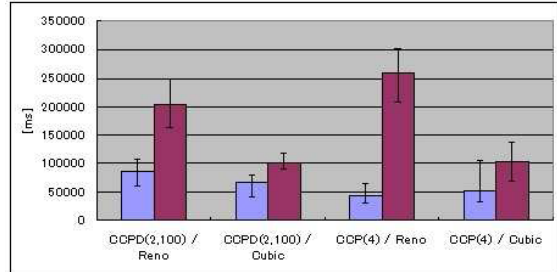
なお、Openflow の利用は本研究課題申請時には想定していなかったが、柔軟なフローベースの経路制御を可能にし、新世代ネットワークのための実験ツールとしても注目され始めたため、NEC（株）の協力も得て (<http://www.nec.co.jp/press/ja/1102/0102.html>)、手法の試作と実験のために導入した。

#### 4. 研究成果

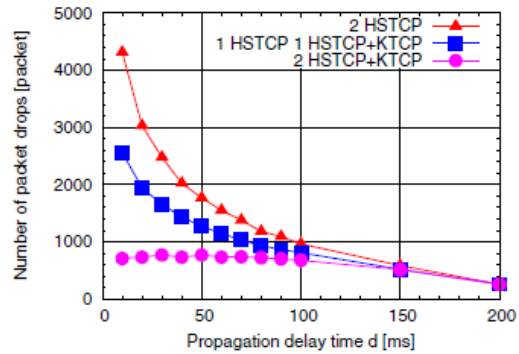
(1) (i) の送信・中継・受信ノード間での高速公平なデータ転送プロトコルに関して、高速転送向き TCP の輻輳制御を改良し、途中のルータのバッファ容量が小さい場合やパケットロスの多い環境での効率を改善する手法や、長期間に渡るフローが存在する場合の競合による非効率性を改善する手法を開発し、Linux 上の試作を用いたネットワークエ

ミュレータと実ネットワーク上での実験や計算機シミュレーションによって評価し、効果を確認した。

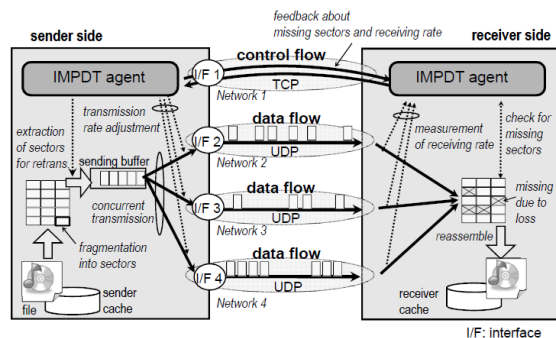
次図（学会発表[4]）は、提案手法 CCP (Capacity and Congestion Proportional) と CCPD (CCP plus Derivative) と TCP の既存手法 (NewReno および Cubic) をインターネット上で競合させた場合の転送完了時間を比較し、提案手法の有効性を示す。



次図（雑誌論文[4]）は、提案手法 KTCP を既存の高速 TCP 手法である HSTCP (High-Speed TCP) に組み込んだ場合に、組み込まなかった場合と比べてパケットロス率が大幅に低減していることを示す。



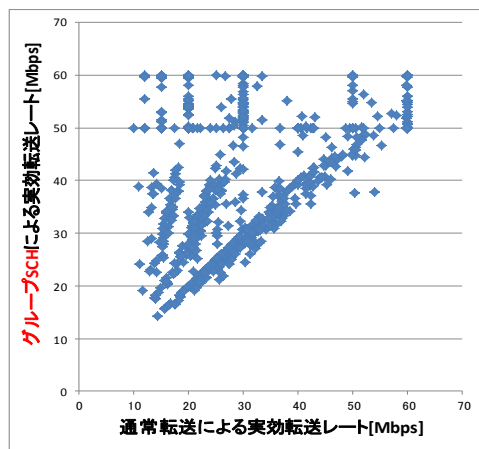
一方、ノード間に複数の異種・不安定な経路がある場合にそれらを統合的に扱うことで TCP よりも転送完了時間を短縮する新しい方式を提案・開発し、計算機シミュレーション、JGN2plus、インターネット、さらに WINDS での実験を通して効果を確認した。次図（雑誌論文[1]）は方式の基本的な仕組みである。



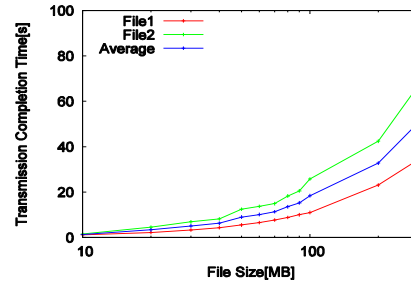
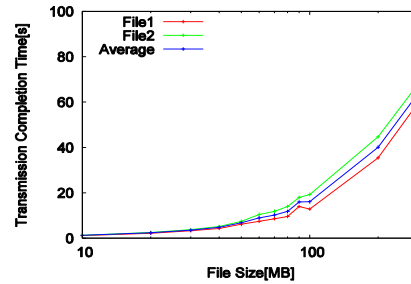
(2) (ii) のファイル中継の時間・空間的ス

ケジューリングと(iii)の中継ノードの適切配置に関して、これらは不可分な部分があったため、一体として取り組んだ。多様なファイル長やボトルネックリンクの混在状況に対応させ、複数のフロー転送の競合を適切な条件の範囲内で回避し各フローの転送完了時間(応答時間)を短縮するデータ中継スケジューリングの基本方式を、蓄積・中継ノードの基本配置方式も含めて検討した。また、高次数の中継ノードへの通過フローの集中を回避する経路制御方式も考案した。さらに、ファイル転送だけでなく転送後の処理も含めた応答時間を短縮するためのスケジューリング手法も考案した(雑誌論文[2])。これらはすべて、計算機シミュレーションによって効果が確認された。

次図(学会発表[1])は、25本のリンクを持つ格子状のネットワークで3箇所のリンクが60[Mbps]で、他はすべて100[Mbps]の帯域幅を持つものを想定し、様々な長さのファイルを同時に8個転送する時の、通提案中継スケジューリングを行う場合と行わない場合(通常転送)の各ファイルの平均スループットのシミュレーション結果である。提案手法により、平均だけでなく、大多数の個々のファイル転送時間を短縮できることを示している。



一方、実際のシステムでは、実装上のオーバヘッドやTCPに起因する非効率性など様々な要因によって期待した効果が出ない恐れもある。そこで、最も基本的な部分を簡易システムとしてPCとOpenflowを用いて試作し、小規模実験ネットワークで評価した。次図(学会発表[2])は、2つのファイル転送を100[Mbps]上で競合させた場合の応答時間を示す。上が従来転送(2フロー同時)、下がスケジュール転送で、前者は平均62秒掛かるが、後者は平均50秒で完了している。



また、OpenFlowを用いた実用化を考えるために各スイッチにおける経路変更の時間粒度の最適化の基礎検討も行った(学会発表[3])。

なお、中継ノードの適切配置に関しては、送信ノードとネットワーク内のボトルネック手前のノードの2段の蓄積が単純かつ有効性が高いことを示唆したが、コストなどの問題で設置できる台数や位置が限定される場合のトレードオフの十分なモデル化や検討ができておらず、今後の課題である。ただし、将来的には多くのルータが高機能中継ノードに置き換えられるとも考えられる。その場合には、静的な最適配置だけではなく、動的な中継・非中継の切り替えが重要になる。そこで、輻輳度合いに応じてTCPの転送を分割して中継に切り替える基本手法についても検討を行い(雑誌論文[3])、シミュレーションによって効果を確認した。

(3) 以上の研究成果を通じて、広域ネットワーク内部の要所要所に置かれたファイル中継ノード間での蓄積形中継転送に基づいた、多数ユーザの長大ファイルの同時転送を効率的に実現する枠組みの基本的な有効性を示し、Openflowなどの最新の技術を利用することで実用化への可能性も示すことができたと言える。しかし、実用化にあたっては、スケーラビリティを確保することが最大の課題であり、全体管理の仕組みや自律分散的な処理について、研究開発を進めたい。また、中継ノードでの待ち時間を利用した圧縮などの併用についても今後の発展課題として残っている。

## 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 6 件)

- [1] A. Nagata, S. Yamamura, M. Tsuru, "Data transfer exploiting multiple heterogeneous challenged networks - implementation and application," Int. J. Space-Based and Situated Computing, 2(2):112-122, 2012. (<http://www.inderscience.com/browse/index.php?journalID=373>), 査読有.
- [2] H. Yamamoto, M. Tsuru, et al., "Parallel Transferable Uniform Multi-Round Algorithm for Minimizing Makespan," IEICE Trans. Comm., E95-B(5):1669-1678, 2012. (10.1587/transcom.E95.B.1669), 査読有.
- [3] M. Shimamura, T. Ikenaga, M. Tsuru, "Splitting TCP connections adaptively inside networks," IEICE Trans. Inf. Syst., E95-D(2):542-545, 2012. (10.1587/transinf.E95.D.542), 査読有.
- [4] S. Yoshimizu, H. Koga, K. Kouyama, M. Shimamura, K. Kumazoe, M. Tsuru, "Kyushu-TCP: Improving Fairness of High-Speed Transport Protocols," IEICE Trans. Commun., E93-B(5):1104-1112, 2010. (10.1587/transcom.E93.B.1104), 査読有.

[学会発表] (計 13 件)

- [1] 佐藤 広和, 鶴 正人, "ファイル転送時間短縮のための蓄積中継型フロースケジューリング手法," 情報処理学会火の国情報シンポジウム 2012, C-4-3, 8 pages, 福岡県飯塚市, 2012 年 3 月 16 日.
- [2] 中山 慶彦, 鶴 正人, "ファイル転送時間短縮のための OpenFlow スイッチを用いたフロースケジューリング," 電子情報通信学会技術報告 111(468), NS2011-255, pp.427-432, 宮崎県宮崎市, 2012 年 3 月 9 日.
- [3] H. Kudou, M. Tsuru, "Effects of routing granularity on communication performance in OpenFlow networks," Proc. IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing, pp.596-601, 24 Aug 2011, Victoria, BC, Canada.
- [4] D. Cavendish, M. Tsuru, "TCP Congestion Avoidance using Proportional plus Derivative Control," Proc. 3rd Int. Conf. on Evolving Internet, pp.20-25, 22 June 2011,

Luxembourg City, Luxembourg.

[その他]

<http://www.ndrc.kyutech.ac.jp/>

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