片方向遅延を用いたネットワークトラフィックの適応的負荷分散手法

柏崎礼生[†1] 小林悟史[†2] 河合修吾[†2]
大石憲且[†2] 高井昌彰[†3]

インターネットが普及し利用が熟成するにつれ，トラフィック要求の恒常的な増加・複雑化が問題となっている．過大なトラフィックが 1 つの経路制御ノードに集中すると輻輳が生じ，伝送遅延の増大，パケット損失が発生するほか，再送により輻輳はより悪化して他の経路制御ノードにまで影響が伝播する．このため，輻輳の発生を回避し回線品質を保つために，ネットワークの帯域，経路制御ノードの処理能力をより有効に利用するための様々なトラフィックエンジニアリング（TE）技術が研究・開発されている．しかし既存のオフライン方式の TE 手法は時間即応性に欠け，トポロジの変化に対応できない．またオンライン方式の TE 手法も単一障害点の問題や複雑なトポロジのネットワークに適用する困難さが指摘されている．本論文では筆者らが提案した遅延時間を用いた適応的経路制御手法を改良し，片方向遅延を用いたアルゴリズム NREI（Network adaptive Routing for Environmental Intelligence）を提案する．IP ネットワークで評価実験を行い，提案アルゴリズムを利用しない静的ルーティングを行った結果と比較し，提案手法はトラフィック要求の増大および経路の遅延の悪化に対して適応的な負荷分散を実現できることが確認できた．

An Adaptive Load Balancing for Network Traffic by Using One-way Delay

Hiroki Kashiwazaki,[†1] Satoshi Kobayashi,[†2] Shugo Kawai,[†2]
Norikatsu Ohishi[†2] and Yoshiaki Takai[†3]

As the Internet becomes increasingly popular, constant increase in demand for network traffic has been an issue. In order to avoid traffic congestion and to maintain link quality, various traffic engineering (TE) technologies, which utilize processing ability of a routing node, are being researched and developed. However, existing “offline-method of TE” lacks reaction sensitivity and adaptability in topology changes. On the other hand, “online-method of TE” contains a SPOF (single point of failure) issue and an adjustability issue in a network of complex topology. This paper proposes NREI (Network adaptive Routing algorithm for Environmental Intelligence) which is based on an one way delay algorithm. NREI is developed from the delay time based adaptive routing algorithm which authors proposed in the past. NREI is tested in an IP network to compare with static routing. The test results show that proposed routing algorithm has better adaptability in congested path avoidance and network load balancing.

1. はじめに

インターネットが広範に普及し，流通するコンテンツのリッチ化が進むことによりネットワークトラフィックの総量は恒常的に増加している．過大なトラフィック要求が 1 つの経路制御ノードに集中すると輻輳が生じ，伝送遅延の増大，パケット損失が発生する．また再送要求の発生によりトラフィックはさらに増大し，その結果，輻輳はより悪化して他の経路制御ノードにまで影響が伝播する．輻輳の発生を回避し回線品質を保つために，ネットワークの回線帯域の増強，経路制御機器の処理速度の向上といったハードウェアによる対処だけでなく，ネットワークの回線利用率を最適化し，ネットワーク全体の通信性能を向上させる目的で通信経路を操作するトラフィックエンジニアリング（Traffic Engineering，以下 TE）[1] による対処も行わ

†1 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University
†2 株式会社ネクステック
Nextech Co., Ltd.
†3 北海道大学情報基盤センター
Hokkaido University Information Initiative Center

れている．

従来の TE 手法は以下の 2 つにまとめられる[2]．トラフィック量をあらかじめ計測しておき，その最大値をもとに最適経路を計算・設定するオフライン方式の TE は，実際に変動する複雑なトラフィック要求への実時間対応が難しくネットワークトポロジの変化への対応が困難であり柔軟性や拡張性に乏しい．またオフライン方式の TE は事前に計測したトラフィックと実トラフィックが一致する場合に有効性を発揮するが，これら 2 つのトラフィック要求が完全に同一となることはない．

もう 1 つの手法は，これらの問題に対処するために研究・提案されているオンライン方式の TE であるが，その多くが MPLS[3] ネットワークを前提にしており[4]，MPLS ルータを必要とする点や，安価な MPLS ルータにおける相互接続性が不安定であることから[5] 適応範囲が狭いという問題点がある．MPLS を用いたオンライン方式の TE[6] では，イングレスエッジが単一障害点となる可能性があり，また複数の LSP（Label Switched Path）があらかじめ設定されていることを前提としている．そのため大規模なネットワークにおいて多様なトラフィックが発生し輻輳がおこるような状況において，その輻輳に対応できる迂回路を予測し，状況に応じて迂回路を選択することは現実的に困難である．MPLS を用いないオンライン方式の TE においては経路振動が発生することが知られており[7]，パケット損失やジッタといった伝送品質および安定性に問題がある．

オフライン方式およびオンライン方式の TE の問題点から，TE に求められる機能は

(1) 実トラフィックへの実時間対応性
(2) 大規模かつ複雑なトポロジに対応可能なスケーラビリティとフレキシビリティ
(3) 単一障害点を持たない自律分散性
(4) 適切なトラフィック分散と経路振動の抑制を両立すること

の 4 点に集約することが可能である．

トラフィック分散の手法として，あらかじめ定めた複数のパスに等しくトラフィックを分散させる決定論的手法があるが，トラフィックの状況に応じて適応的にパスを増減させたり，ダイナミックに迂回路を変更したりすることが難しい．これに対して，存在しうるパスに対してトラフィックを確率的に分散させる確率論的手法を自律分散的に行おうとすると，経路の遅延差によるリオーダリングの低減やループの回避が一般的に求められる[8]．ループの回避手法としてパケットに経由経路を記載する手法が考えられるが，パケットに対する処理が煩雑なためスケーラビリティを損なうという問題がある．しかし，エンドユーザの PC 処理性能の向上にともない，一般的なネットワーク利用においては，アプリケーション側でのリオーダリング処理や，ループによるパケットロスのエラー訂正を行うことで，アプリケーションの品質を保つことが可能である[9]．

経路制御ノードの自律的な計測によって得られる回線品質情報として，回線利用率や他ノードとの遅延時間があげられる．回線利用率は輻輳の評価基準にはなるが，回線利用率をより低くすることを目標値として自律分散で競争を行わせると経路の収束が難しく，パケットの総遅延時間の振動が大きくなる．それに対して遅延時間を評価基準にすると，経由するノードでの転送遅延を 0 と仮定したとき，各ノード間の伝搬遅延時間をコストとする重み付きグラフの最短経路が最適値（最短遅延時間）となるため，経路の収束を実現しやすい．片方向遅延を用いた適応的経路制御手法として，ゲートウェイレベルの経路制御手法が提案されているが，単純な輻輳回避手法であり，効率的なトラフィック分散ができない[10]．

そこで筆者らは遅延時間を用いた自律分散型の適応的経路制御アルゴリズムを提案した[11]．この手法はパケットに経由経路・経路間遅延情報を随時記載し，この情報から各経路制御ノードが受動的かつ自律的に他ノードとの遅延時間を収集し，遅延時間をスコアとする経路制御表を構築するアルゴリズムである．この経路制御表のスコアを用いて次ノード候補の評価値を決定し，重み付き確率分散で経路制御を行う．評価実験によりこの手法がネットワークリソースを有効に利用して，一部のノードに輻輳を生じさせることなく，より多くのトラフィック要求を系全体で分散して処理できることを示し，特に，より迂回路が多く存在するトポロジにおいて優位性が高いことが分かった．しかしすべてのパケットに経路・遅延情報を記載する点や，上り・下り遅延の対称性を前提にしている点が実ネットワークへの適用の障害であった．

本論文ではこのアルゴリズムを改良し，各ノードが独立して能動的に片方向遅延の計測を行い，この遅延情報を評価値とする経路制御表を自律的に構築する経路制御手法 NREI を提案する．常時，計測パケットをすべてのノードに発信し，これにより得られた遅延時間情報を用いて経路の評価を行うことでトラフィック量の変動に対する即時対応を実現した．また，実ネットワークで発生するパケット損失を遅延時間情報に置

き換えることでパケット損失の回避を実現した．本研究は遅延センシティブでないネットワークアプリケーションでの利用を前提としたオーバレイネットワークでの TE 手法であり，リオーダリングおよびループについてはアプリケーション側での処理で回復を行うこととする．遅延センシティブなネットワークアプリケーションについては，別のオーバレイネットワークで対処することを前提とした．計測パケットの総トラフィック量は十分に小さく，伝送品質に与える影響は無視できる．この手法により混雑時には特に明示的な設定を行うことなく迂回路にトラフィックを分散させることができる．この改良アルゴリズムを PC ルータに実装し，アルゴリズムの応用性を重視して IP ネットワーク上で評価実験を行い，決定論的に経路を設定するルーティングに対する提案手法の有効性を示す．

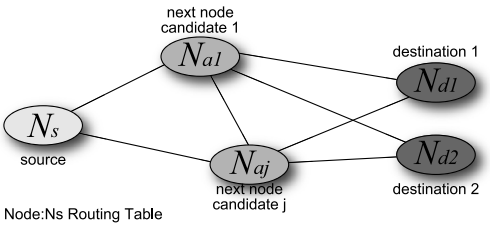
2. 提案アルゴリズム：NREI

ネットワークの遅延尺度として RTT（Round Trip Time）が広く知られている．しかし伝送遅延の上り遅延と下り遅延の非対称性から，往復路遅延を経路制御の評価値として用いることは論理的説得力に乏しい．それに対し，近年 NTP の計測精度を飛躍的に向上させる研究が行われており[12),13)]，これを利用した片方向遅延（One Way Delay: OWD）を評価尺度として帯域や混雑状況を推測する応用手法が期待されている．また，精度の高い計測を実現するためには計測経路をトラフィックが流通するネットワークとは独立させて計測させるべきと考えられるが，一定間隔にブロードキャストされた NTP の同期パケットがキューイング遅延の影響を受けにくいことを利用し，回線利用度が高い状態における高精度の時刻同期を実現する研究も行われている[14)]．広域ネットワークにおける経路制御ノード間の平均遅延時間に注目するならば，NTP を用いた片方向遅延で十分な精度を保つことができるといえる．

そこで本論文では NTP を用いて計測された片方向遅延情報を用いてネットワークトラフィック要求を適応的に分散する．片方向遅延を計測するために NTP の計測パケットを用い，計測された片方向遅延情報をもとに経路制御表を構築する．

2.1 経路制御表と計測パケット

各ノードはすべての目的ノードに関する経路制御表（次ノード候補の集合 = すべての隣接ノード）を有しており，任意の目的ノード N_d に向けたすべての次ノード候補 N_a に長さ w のスコア列 $L_{N_dN_a} = \langle S^1_{N_dN_a}, S^2_{N_dN_a}, \cdots, S^w_{N_dN_a} \rangle$ が与えられる．このスコア $S^j_{N_dN_a}$ はノードが任意の目的ノードに対して送信した計測パケットから得られた片方向遅延情報である（**図 1**）．



Node:Ns Routing Table

destination	*next node*	*scores*
N_{d1}	N_{a1}	$S^1_{N_{d1}N_{a1}}$, $S^2_{N_{d1}N_{a1}}$, ……, $S^{w-1}_{N_{d1}N_{a1}}$, $S^w_{N_{d1}N_{a1}}$
	N_{aj}	$S^1_{N_{d1}N_{aj}}$, $S^2_{N_{d1}N_{aj}}$, ……, $S^{w-1}_{N_{d1}N_{aj}}$, $S^w_{N_{d1}N_{aj}}$
N_{d2}	N_{a1}	$S^1_{N_{d2}N_{a1}}$, $S^2_{N_{d2}N_{a1}}$, ……, $S^{w-1}_{N_{d2}N_{a1}}$, $S^w_{N_{d2}N_{a1}}$
	N_{aj}	$S^1_{N_{d2}N_{aj}}$, $S^2_{N_{d2}N_{aj}}$, ……, $S^{w-1}_{N_{d2}N_{aj}}$, $S^w_{N_{d2}N_{aj}}$
N_{a1}	N_{a1}	$S^1_{N_{a1}N_{a1}}$, $S^2_{N_{a1}N_{a1}}$, ……, $S^{w-1}_{N_{a1}N_{a1}}$, $S^w_{N_{a1}N_{a1}}$

図 1 スコア付き経路制御表
Fig. 1 Routing table with scores.

各ノードは一定時間ごとに任意のノード N_{dst} を目的ノードとして，計測パケットをすべての次ノード候補 $N_{a_1} \cdots N_{a_j}$（j は次ノード候補総数）に対して送信し，自ノード–各次ノード候補–目的ノードという各経路の片方向遅延を計測する．送信された計測パケットには送信時刻 t_{send} が記載され，目的ノードに到着した計測パケットには受信時刻 t_{recv} が追記される．送信時刻と受信時刻が記載されたパケットは目的ノードから送信元ノードへ返送され，そのパケットの目的ノード N_{dst} と経由した次ノード N_{a_n} とに対応するスコア列の先頭 $S^1_{N_{dst}N_{a_n}}$ に，受信時刻と送信時刻の差分 $d = t_{recv} - t_{send}$ が片方向遅延として追記される．経路制御表に新しいスコアが書き込まれる前に，古いスコア列の要素を 1 つずつ後退させ，列長を超えたスコアを破棄する．すなわち，スコア列にはつねに最新の w 個のスコアが保存される（**図 2**）．

計測パケットには連続した番号を記載しておく．計測パケットを発生させるノードは，各計測パケットの目的ノードと経由した次ノード，および送信時刻情報をこの番号と紐付けすることで，片方向遅延を計測して戻って来たパケットとそうでないパケットを識別することができる．パケットを送信してから d_{limit} 以上経過した場合は途中の経路でパケットドロップが発生したものと見なし，そのパケットの目的ノードと経由した次ノードに対応するスコア列の先頭にペナルティ遅延値 d_p を与える．

計測パケットはすべての隣接ノードに対して送付された後，次節で述べる次ノード選択規則に従い目的

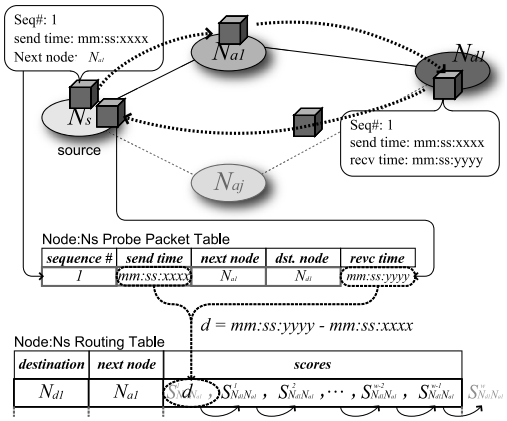


図 2 計測パケットの動作と経路制御表の更新手順
Fig. 2 Protocol for probe packets and updating routing table.

ノードまで運ばれる．また近傍の片方向遅延情報に関して情報の精度を上げるため，自ノードからの最短ホップ数が h_n 以下のノードに対しては，経路を始点で明示的に与えて遅延計測を行う．このとき明示的に与えられる経路は，各隣接ノード N_{a_n} から目的ノード N_{dst} への最短ホップ数の経路 $[N_{a_n}, \cdots, N_{dst}]$ の先頭に自ノード N_s を加えた経路 $[N_s, N_{a_n}, \cdots, N_{dst}]$ である．この経路が任意の隣接ノード N_{a_i} に対して複数あるとき，任意のノード N_{a_i} と目的ノード N_{dst} に対応する経路制御表には複数の経路から得られる遅延時間情報がスコア列に格納される．

2.2 次ノード選択規則

あるノードにパケットが到着した際，そのノードが目的ノードでなければ，目的ノードに対応する経路制御表のスコア列を参照し，次ノードを決定する．次ノードはすべての隣接ノード N_{a_i}（$1 \le i \le r$, r は候補総数）とする．

候補が複数存在する場合，目的ノード N_d，候補ノード N_{a_i} のスコア列に対する代表値 Q_{a_i} を，$Q_{a_i} = \sum_{m=1}^{w} C_m S^m_{N_d N_{a_i}}$ とするスコア列全体に対する加重平均で与える．ここで C_m は負の傾きを持つ線形の加重関数である．

Q_{a_i} は遅延時間であり，すべての候補の中で最小の Q_{a_i} が選ばれることが望ましいが，決定論的に選択すると負荷分散が行われず，系全体でより多くのトラフィック要求を受容することが困難となる．そこで次ノードは，すべての候補の代表値 Q_{a_i} を λ（> 0）乗した値の逆数で加重された確率分散で決定されるものとする（**図 3**）．すなわち候補ノード N_{a_i} が次ノードとして選ばれる確率 P_{a_i} は式 (1) で表される．

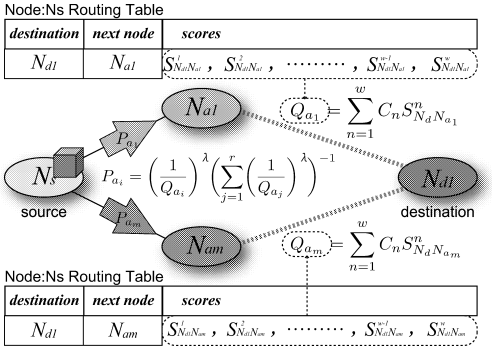


図 3 送付先の確率的選択
Fig. 3 Probabilistic selection of the next-node.

$$P_{a_i} = \left(\frac{1}{Q_{a_i}}\right)^{\lambda} \left(\sum_{j=1}^{r} \left(\frac{1}{Q_{a_j}}\right)^{\lambda}\right)^{-1} \qquad (1)$$

次ノードの選択確率を定めるパラメータ λ は，候補の中から相対的に少しでも優位なノードが選択される確率を強める働きがある．明らかに $\lambda \to \infty$ では，最良の候補が確率 1 で選ばれることになる．以下本論文では，送付先決定に λ をパラメータとして含む式 (1) を用いる方式を NREI（Network adaptive Routing algorithm for Environmental Intelligence）と呼ぶ．

ホップ数が h_n 以上の自ノード N_s と目的ノード N_{dst} の組に対する計測パケットもこの選択確率で目的ノードまでの遅延を計測するため，各隣接ノード N_{a_n} と N_{dst} に対するスコア列には，その隣接ノードからとりうる様々な経路の遅延時間が混在する．その結果，ネットワーク上に経路遅延に偏りのある部分が存在しても，その部分から遠いノードにおいては他経路からの遅延情報との平均によって偏りが平滑化され，遠くからその部分を強く回避する傾向を作らず，近づくにつれて回避傾向が強くなる経路制御が実現される．

3. 評価実験

提案手法の有効性を評価するために，NREI アルゴリズムを実装した遅延適応ルータ（AR）を用意し，実験ネットワークを構成して 2 種類の評価実験を行った．

3.1 実　　装

AR の実装は，Xeon 1.86 GHz，512 MBytes の PC で OS は NetBSD 3.0.2 を用いて行った．AR の機能は，片方向伝送遅延の計測・評価機能，決定された配分割合に従いパケットを配送する機能に分類される．片方向伝送遅延の計測は，NTP により時刻同期した AR 間において，タイムスタンプを記載した UDP パ

ケットを送受信することで行われる．AR では，以下の 3 種のプロセスが動作する．

(1) 計測パケットを送信するプロセス
(2) 計測パケットを受信するプロセス
(3) 計測を評価するプロセス

片方向遅延の計測区間を AR1 → AR2 とした場合，AR1 では (1) および (3) が，AR2 では (2) のプロセスが動作する．

プロセス (1) は，タイムスタンプが埋め込まれた UDP パケットを指定された AR に指定された経路を通るように送信する機能を持つ．片方向遅延時間の計測精度を上げるため，ホップ数 h_n 以下の近傍目的ノード宛ての計測パケットには SSRR（Strict Source Routing and Record）オプションを指定する．プロセス (2) は，計測パケットを受信した時刻を取得し，受信時刻の情報を送信元の AR に返送する機能を持つ．プロセス (3) は，(1) および (2) により記録された送受信時刻の差から伝送遅延を算出し，前章のアルゴリズムに基づいてパケット配分割合を計算する．また，計算された配分割合をパケット配送機能に伝達する．

パケット配送機能は，OS のカーネル内に実装した．同機能は既存の IP ルーティング機能とは別のルーティングテーブルを持つ．このテーブルには，宛先 IP プレフィックスに対する次ホップのリストが配分割合付きで記載される．個々の IP パケットの宛先アドレスに対して，最長一致するプレフィックスが選択され，配分割合に基づき次ホップが選択される．宛先アドレスに対してプレフィックスが選択されない場合には，OS が持つ IP ルーティング機能に基づき次ホップが決定される．

配分割合を決定する式 (1) のパラメータ λ は，次ノード候補の中から相対的に優位なノードが選択される確率を高める働きを持つ．$\lambda \to \infty$ では最良の次ノード候補が確率 100%で選択されることになり，このアルゴリズムによる決定を決定論的に行うか，一様な確率分散で行うかを制御するパラメータである．λ は今回，変更が可能な形で実装した．また，要求トラフィックを式 (1) により求められた配分割合で分散させる方法として，求められた割合を正確に実現することのできる per-packet 方式を用いた．

遅延時間の計測頻度が高すぎると経路制御表に微視的な情報ばかりが集積されることになる．試験的な計測の結果，10 ms オーダでの遅延の大きな変動はないと考え，計測パケットの送信頻度は 100 ms とした．配分割合の計算頻度は，1 sec ごとに行った場合において総遅延時間の変化の観測が難しく，一方で計算頻度が低いとネットワーク品質の変化に対応できないため，計算頻度を 10 sec ごととし，少なくとも 10 sec 間は直近に計算された配分割合に基づきトラフィックが処理されるようにした．また，急激な配分割合の変動を抑制するため，過去数回分の配分割合の変化を次の配分割合の計算に取り入れることができるようにスコア列のサイズ w を 1,000 に定めた．スコア列のサイズに関連し，1 回のペナルティ遅延値 d_p が代表値を過度に増幅させないように本論文ではすべての実験で $d_p = 1{,}000$ ms としている．また経路制御の処理能力を優先するために配分割合は 10%単位で丸め，配分割合の計算頻度を高めた場合でも追従できるようにした．

3.2 実験ネットワーク

本評価実験では，以下の 4 拠点に AR を設置した．

- AR1：札幌（北海道大学情報基盤センター）
- AR2：富山（インテック・ウェブ・アンド・ゲノム・インフォマティクス株式会社）
- AR3：山梨（山梨県立大学）
- AR4：高知（高知工科大学）

各拠点は，JGN II（超高速・高機能研究開発テストベッドネットワーク）[15] 上に構築された RIBB-II（地域間相互接続プロジェクト，Regional Internet Back-Bone II）ネットワーク[16] により相互に接続している．AR によるルーティング空間を構築するため，各拠点の AR どうしを IPv4 over IPv4 によるトンネルで接続し，オーバレイネットワークを構成している．このオーバレイネットワークでは，互いの AR をルーティング時の次ホップとして利用できる．

本実験においては，札幌と高知にユーザネットワークを構築し，トラフィック発生装置（Traffic Agent: TA1，TA2）を設置した．このユーザネットワーク間のトラフィックを，AR により富山または山梨を経由する形でルーティングを行う．すなわち，

(1) AR1（札幌）– AR2（富山）– AR4（高知）
(2) AR1（札幌）– AR3（山梨）– AR4（高知）

の中継地が異る 2 本の伝送路を作成し，遅延計測をそれぞれの伝送路ごとに実施する．AR1 または AR4 は，計測から得られた配分割合に基づき，パケットの転送先として AR2 または AR3 を選択する．なお，中継地となる AR2 および AR3 においては，帯域を制限して輻輳を意図的に発生させるため，AR2，AR3 は JGN II 回線に対して 10 Mbps の半二重で接続している．さらに，伝送遅延を制御するために，AR2 と JGN II 回線の間に遅延を msec 単位で制御する装置（遅延発生装置 Delay Generator: DG）を設置してい

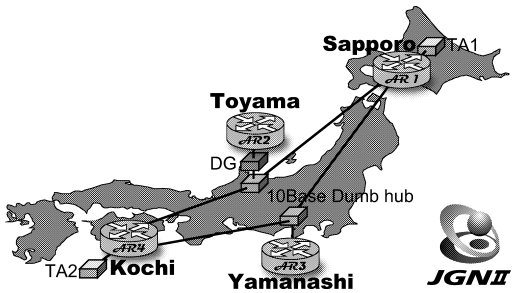


図 4 評価実験ネットワークの構成
Fig. 4 Network construction for evaluation experiments.

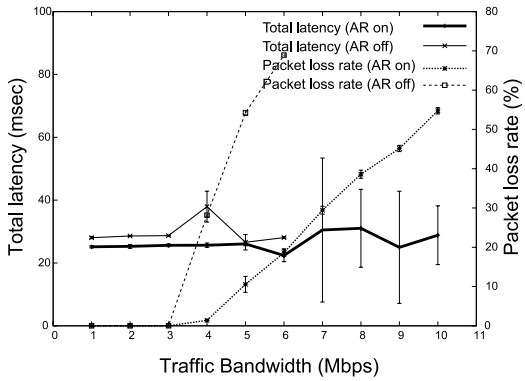


図 5 トラフィック量増大に対する片方向遅延時間とパケット損失率の変化
Fig. 5 Changes of delay and packet loss ratio for increment of network traffic.

る（**図 4**）.

3.3 実験 1. トラフィック要求の増大

提案したアルゴリズムを実装した AR によるネットワークで適応的な負荷分散が実現できることを確認するために，片一方の伝送路のみの使用では帯域を超えるような量のトラフィック要求を発生させる評価実験を行う.

構築した実験ネットワーク上の AR4 に接続されたトラフィック発生装置 TA2 から AR1 に接続された TA1 へ UDP パケットによるトラフィック要求を発生させる．トラフィック量を増加させた際の TA2 から TA1 への片方向遅延時間の変化とパケット損失率を計測し，AR 機能を有効にした結果と，一方路のみを固定的に選択する方式での結果とを比較する．トラフィック要求の発生およびパケット損失率の計測には，ネットワークの帯域やスループットを解析するソフトウェアである iperf[⋆1] を用い，送信パケットのサイズならびに送出間隔を調整することで，トラフィック量（Mbps）を制御する．また，トラフィック発生中の [TA2，TA1] 間の片方向総遅延時間は TA2，TA1 で ntpd を動作させて計測した．次ノード選択確率を定めるパラメータは $\lambda = 4$ に設定し，トラフィック量を 1 Mbps から 10 Mbps まで 1 Mbps ごとに変化させ，平均片方向遅延時間とパケット損失率を 10 回計測した．また，近傍の遅延計測で明示的経路を設定するパラメータは $h_n = 2$ に設定した．**図 5** にその平均値と標準偏差を示す.

AR2 と AR3 とは半二重の 10 Mbps ダムハブを経由して接続されているため，AR 機能が有効でない場合は 3 Mbps のトラフィック量を超えた時点でパケット損失率が飛躍的に増大する．7 Mbps 以上のトラフィック量を発生させるとパケット損失率が高くなりすぎ，パケット損失率の計測が不可能となった．また 4 Mbps のトラフィック量が発生した時点で到着したパケットの総遅延時間とジッタの増大が観測された.

一方で AR 機能を有効にした場合，4 Mbps でのパケット損失率を AR 機能が無効の場合の 5%まで低減させている．11 Mbps 以上はパケット損失の増大により計測が不可能になる．また総遅延時間は 7 Mbps 以上のトラフィック要求でジッタが増大するが平均総遅延時間に大きな変動は見られず，AR 機能により適切な負荷分散が行われている．Reed-Solomon 符号を用いた FEC（Forward Error Correction）で冗長化されたネットワークストリームであればパケット損失が 20%であっても 2.5%の損失まで回復できる[17]．そこでパケット損失 20%を 1 つの指標とした場合，この評価実験のネットワークでは 2 つの経路にトラフィックを自律的に分散させることでおよそ 2 倍のトラフィック要求に応じることができる.

3.4 実験 2. 片側経路の遅延増大

地域間で高品位映像伝送を行うなど広帯域のトラフィック要求が行われるイベントではすべての拠点における中継機器の性能やネットワークの性質が均一とは限らず，そのため伝送途中にある中継路の品質が悪化する事態がしばしば観測される．このような障害が発生した場合には別経路への切替えを手動で行うのではなく，バックボーンが自律的かつ連続的に配分割合を変更することが望ましい．提案手法が遅延時間の増大に対して自律的に適応的な配分割合の変更を行うことを評価するため，以下の実験を行った.

AR2 に接続された遅延発生装置を用いて AR1–AR2–AR4 の経路の総遅延時間に 0 ms から 100 ms までの遅延を段階的に 10 ms 単位で人為的に追加した後，0 ms に戻す．このとき，AR4 に接続された TA2

⋆1 http://dast.nlanr.net/Projects/Iperf/

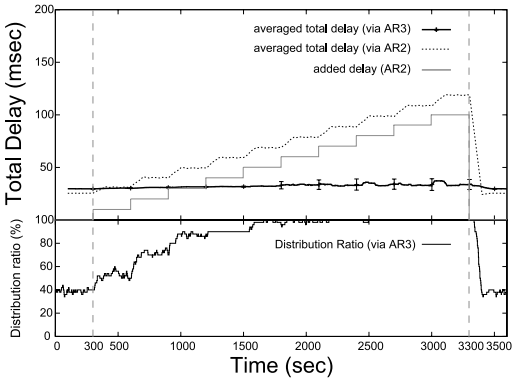


図 6 経路遅延の増大に対するパケット配分割合の変化（トラフィック量：3 Mbps）
Fig. 6 Changes of packet distribution for path delay increment (3 Mbps).

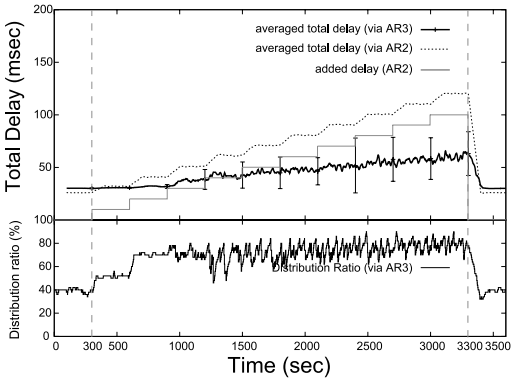


図 7 経路遅延の増大に対するパケット配分割合の変化（トラフィック量：4 Mbps）
Fig. 7 Changes of packet distribution for path delay increment (4 Mbps).

から，AR1 に接続された TA1 に対して UDP パケットによるトラフィックを発生させ，AR4 における経路選択割合と AR1 に到着したパケットの平均総遅延時間を AR2 経由と AR3 経由のそれぞれについて計測した．発生させたトラフィック量は 3 Mbps，4 Mbps，5 Mbps であり，次ノード選択確率を定めるパラメータを $\lambda = 4.0$ に設定した．各トラフィック量における評価実験を 5 回行い，平均総遅延時間および配分割合の変化を**図 6**，**図 7**，**図 8** に示す．AR3 経由の平均総遅延時間については 300 sec ごとの標準偏差を並べて示す．

遅延発生装置による人為的な遅延が発生していない状況 [0, 300 sec] において AR2 経由の平均総遅延時間は 25.5 msec，AR3 経由の平均総遅延時間は 29.7 msec である．この総遅延時間を代表値として式 (1) を用いた配分割合の理論値は AR2：AR3=1：0.54 となり，配分割合の 10%の丸めを含めると計測された配分割合 AR2：AR3=60%：40%に一致する．3 Mbps のトラフィック量を発生させた状態では，すべてのトラフィックが片方の経路に集中しても輻輳は発生しない（図 6）．300 sec 付近から AR2 経由経路の総遅延時間が段階的に増大するのに対して配分確率も段階的に AR3 を経由する側が高くなり，遅延発生装置が 50 ms の遅延を発生させた時点（1,500 sec 近傍）で AR3 経由経路が 100%で選択される．人為的な遅延は 3,300 sec で 0 sec に戻る．配分割合は 95 sec で元の 60%：40%に戻る．

トラフィック量が 4 Mbps になると片方の経路に集中した場合輻輳が発生しパケット損失を生じる（図 7）．配分割合が AR2：AR3=20%：80%を超えると AR3 を経由する経路に 3.2 Mbps 以上のトラフィックが発生し，パケット損失によるペナルティ遅延が経路制御

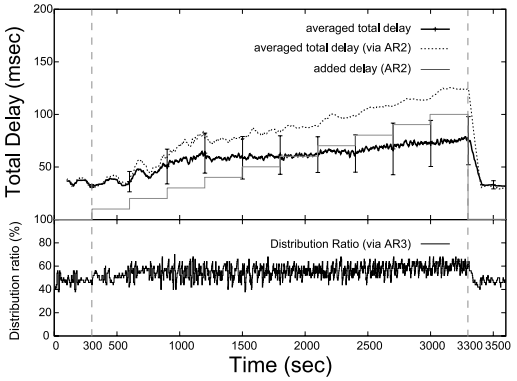


図 8 経路遅延の増大に対するパケット配分割合の変化（トラフィック量：5 Mbps）
Fig. 8 Changes of packet distribution for path delay increment (5 Mbps).

表に与えられる．そのため遅延発生装置が 40 ms の遅延を発生させた時点（1,200 sec 近傍）から配分割合の振動が生じ始めるが，[1,200, 3,300] における配分割合の平均値は 73.9%（2.92 Mbps），標準偏差は 7.15%である．パケット損失が発生する臨界付近までトラフィックを AR3 経由に配分するため，機器の処理能力超過による伝送遅延が発生し，AR3 経由の総遅延時間は緩やかに増大し，2,400 sec の時点で標準偏差は最大で 26 msec となるが，AR2 を経由する経路の 50%程度の総遅延時間の経路を優先的に選択していることが分かる．

同様にトラフィック量が 5 Mbps においても片方の経路だけにトラフィックが集中すると輻輳によりパケット損失が発生する（図 8）．この場合における [1,200, 3,300] における配分割合の平均値は 56.5%（2.82 Mbps），標準偏差は 5.32%であり，4 Mbps に

おける結果同様，AR3 経由でパケット損失が発生する量のトラフィック要求近傍で安定していることが分かる．パケット損失により与えられるペナルティ遅延値が，遅延発生装置により与えられる遅延よりも大きいため，パケット損失の回避が優先されて配分割合が決定される．

4. 考　　察

4.1 トラフィック要求の増大に関する考察

トラフィック要求を増大させた場合における自律的負荷分散性能を評価した実験 1. において，1 つのパスにトラフィック要求を集中させた場合との比較を行った．このような 1 組の始点ノードと目的ノードが 2 つのパスで接続されている明快なトポロジにおいては equal cost multi-path（ECMP）を用いてトラフィックを分散させる手法が一般的である．ECMP におけるトラフィックの分割手法は実装依存だが，per-flow で分割することが推奨されている[18]．高品位映像伝送のようにフローが非常に大きい粒度である場合は複数あるパスを有効に利用できないが，フローの粒度が十分に小さければパス数分のトラフィックを許容することができる．評価実験 1. の実験環境において AR4–AR3–AR1 と AR4–AR2–AR1 を等コストリンクとして ECMP を用いると仮定すると，実験 1. の結果から 6 Mbps（1 パス 3 Mbps×2）までパケットロスが発生せず，高いトラフィック要求が発生した場合においても遅延時間の偏差が小さく抑えられることが分かる．一方，提案手法を用いた経路制御手法においては，片一方の経路でパケットロスが発生するまでトラフィックを割り当ててしまうため，ECMP の方が高い有効性を示す．トラフィック要求の増大に柔軟かつ効率的に対応するためには，片方向遅延情報だけでなく，帯域占有率やパケットロスの発生するトラフィック要求量の計測値を配分割合の決定式に導入することが求められる．

ECMP のような決定論的な手法ではネットワークのトポロジと，それに対するトラフィック要求が既知である場合において効率的なトラフィック要求の配分を行うことができるが，トポロジが複雑になるとパスの設定が困難となり，大規模なネットワークにおいてはトラフィック要求が既知とはなりえない．クロストラフィックの発生により輻輳が発生し，あらかじめ設定したパスのすべての品質が悪化した場合，ECMP の運用では代替となる品質の良いパスを計測して探し，その経路を等コストパスとして設定する必要が生じる．しかしその際，帯域に余裕のあるパスが複数存在しなければ，各パスの空き帯域に応じてトラフィックの配分割合をパスごとに設定することは困難である．提案手法は片方向遅延時間の測定値に応じた配分割合を自律的に決定するため，そのようなパスの探索やネットワークアレンジを必要としない．この点において提案手法の優位性がある．

4.2 片側経路の遅延増大に関する考察

遅延を人為的に変動させ配分割合の偏向を観測した実験 2. において，図 6 と図 7 を比較すると，片方の経路に 3.0 Mbps のトラフィックが流れている状態までは転送遅延が発生せず，3.2 Mbps のトラフィックでパケットロスが発生していることが分かる．これは帯域を制限するために用いた機器がバッファ容量の非常に小さな 10 Base-T のダムハブであるため，バッファリングによる転送遅延が発生するまでもなくパケットロスを発生してしまい，転送遅延そのものよりもパケットロスのペナルティ遅延による配分割合の変動がおこっているものである．一方パケットロスが発生した場合のペナルティ遅延により，輻輳が発生した経路制御ノードを敏感に避けることが可能であり，計測された片方向遅延情報により経路制御表が更新されることによって，輻輳が緩和したノードに対してトラフィックを徐々に回復させることを可能にしている．提案手法は本評価実験のようなネットワークにおいても有効性を示すが，バッファ容量の大きなノードから構成されたネットワークで複雑なトラフィックによる輻輳が発生する状況において，より高い適応性と障害回避を実現すると考えられる．

提案手法は片方向遅延を評価値として配分割合を決定するため，たとえば衛星回線や経路距離の大きな回線など，極端に遅延の大きな代替経路があった場合に，その代替経路の帯域に余裕があってもまったく使用されない結果となる．評価実験における実装で述べたように配分割合は 10%単位で丸めているため，本実装においてたとえば遅延時間の小さな経路と大きな経路の 2 つのパスが存在した場合，配分割合決定パラメータ $\lambda = 4$ では，遅延時間差が 2.11（$= \sqrt[4]{20}$）倍以上になったときに遅延時間の大きな経路の配分割合が 0%となる．λ の変更によりジッタがどの程度の範囲で収まるかを指定することが可能であると考えられる．

5. ま と め

能動的に片方向遅延を計測し，これを利用してトラフィック要求の動的変化や経由する経路の品質変化に対して適応する確率的な経路制御手法 NREI を提案した．各経路制御ノードは独立して自律的に経路制御表

を構築し，片方向遅延時間に応じて重み付けされた確率的な次ノード選択によって平均総遅延時間の小さな経路に配分割合を大きく与える負荷分散を実現する．NREI を実装したルータを JGN II に配置し，迂回経路を持つトポロジのネットワークを構築して評価実験を行った．

プローブパケットにより取得された片方向遅延情報は次ノード選択に反映させ，より短い遅延時間の経路を高い確率で選択することで実トラフィックへの実時間対応性を実現している．また各経路制御ノードに個別の設定を投入することなく，すべて単一のアルゴリズムにより駆動させることで適応的な経路制御を実現している．ネットワークのトポロジが大規模化・複雑化した場合においても，トポロジに応じたネットワークアレンジを施す必要がなく，スケーラビリティとフレキシビリティを確保することができる．またこのことは，系の中で中央集権的な経路制御ノードを持たないことを意味しており，単一障害点を持たない．評価実験で示したように本提案手法は複数の経路に対して自律的にトラフィックを分散させつつ総遅延時間をより小さくする経路を選択することで，決定論的に 1 つのパスにトラフィック要求を集中させるよりも多くのトラフィック要求を系全体で許容することができた．また配分割合の振動を評価実験では 5〜7%に抑えることを示した．

提案手法は片方向遅延情報のみから配分割合を決定するため，パケット損失や輻輳を生じさせない程度のトラフィック要求に対しても確率的に負荷分散を行う．そのため，総遅延時間の大きな経路にもトラフィックを流す可能性があり，少ないトラフィック要求に対しては最適解と比較して平均総遅延時間を増大させてしまう．しかし，提案手法は高品質動画像といった帯域を逼迫するようなトラフィックがつねに発生するネットワークにおいて特に有効性を示す．

選択される経路の確率的な広がりは，次ノード選択規則のパラメータ λ によって定められる．λ が大きければ，総遅延時間の平均値は減少するが，その分散は増大する．適切な λ 値を選ぶことによって，平均値と分散をバランスさせ，総遅延時間の最大値を低く抑えることが可能である．

本研究の実装では per-packet 方式を用いているためにコネクションの品質を損ねることが考えられる．複雑で変動の激しいトラフィック要求に対して自律的に適応的な制御を実現するために，片方向遅延だけでなく回線利用度などの情報を加味してアルゴリズムを改善し，高い処理能力を保ちながら通信品質を高めることが今後の課題である．

参 考 文 献

1) Awduche, D., Chiu, A., Elwalid, A., Widjaja, I. and Xiao, X.: Overview and Principles of Internet Traffic Engineering, RFC 3272 (2002).
2) 小原泰弘，今泉英明，加藤 朗，中村 修，村井 純：広範なトラフィック要求に対応する負荷分散経路計算アルゴリズム，情報処理学会論文誌，Vol.48, No.4, pp.1627–1640 (2007).
3) Rosen, E., Viswanathan, A. and Callon, R.: Multiprotocol Label Switching Architecture, RFC3031, IETF (Jan. 2001).
4) 熊木健二，中川郁夫，永見健一，長谷川輝之，阿野茂浩：キャリアネットワークにおける MPLS TE LSP 確立に関するロードバランス手法の提案と評価，情報処理学会論文誌，Vol.48, No.4, pp.1616–1626 (2007).
5) 菊池 豊，石原丈士，永見健一，楠田友彦，菱岡裕男，西内一馬，羽田友和，水村雅明，正岡 元，池田浩志，中川郁夫，江崎 浩：異機種ルータの相互接続試験活動—新しいネットワークアーキテクチャの導入を促進するために，信学技報，Vol.106, No.15, SS2006-4, pp.19–24 (2006).
6) Awduche, D., Berger, L., Li, T., Srinivasan, V., Swallow, G.: RSVP-TE: Extensions to RSVP for LSP Tunnels, RFC3209 (2001).
7) Anderson, E.J. and Anderson, T.E.: On the Stability of Adaptive Routing in the Presence of Congestion Control, *INFOCOM '03* (2003).
8) Kandula, S., Katabi, D., Sinha, S. and Berger, A.: Dynamic Load Balancing Without Packet Reordering, *ACM SIGCOMM Computer Communication Review*, Vol.38, No.2, pp.53–62 (2007).
9) 岸田崇志，前田香織，河野英太郎：ネットワーク障害物を乗り越えるテレビ会議用ゲートウェイの開発，情報処理学会論文誌，Vol.48, No.4, pp.1552–1561 (2007).
10) Jo, M., Kim, H.D. and Kim, H.: An Adaptive Routing Method for VoIP Gateways Based on Packet Delay Information, *IEICE Trans. Communications*, Vol.E88-B, No.2, pp.766–769 (2005).
11) 柏崎礼生，高井昌彰：遅延時間情報に基づく適応的ネットワークルーティング，情報処理学会論文誌，Vol.47, No.12, pp.3308–3318 (2006).
12) 岩間 司，金子明弘，町澤朗彦，鳥山裕史：高速ネットワークを利用した高精度時刻比較，電子情報通信学会論文誌 D，Vol.J89-D, No.12, pp.2553–2563 (2006).
13) 山田雄介：IP ネットワーク上の時刻同期手法，電子情報通信学会技術研究報告，Vol.105, No.280, pp.1–6 (2005).

14) 町澤朗彦，岩間 司，鳥山裕史：毎正秒パケット到着感覚（PAI）に基づいた時刻同期方式，電子情報通信学会論文誌 B，Vol.J89-B，No.10，pp.1855–1866 (2006).
15) Japan Gigabit Network II, Advanced Testbed Network for R&D. http://www.jgn.nict.go.jp/
16) 菊池 豊，中川郁夫，樋地正浩，八代一浩，林 英輔：ジャパンギガビットネットワーク：4地域間相互接続実験プロジェクト，情報処理，Vol.43，No.11，pp.1171–1177 (2002).
17) 近堂 徹，西村浩二，相原玲二，前田香織，大塚玉記：高品質動画像伝送における FEC の性能評価，情報処理学会論文誌，Vol.45，No.1，pp.84–92 (2004).
18) Thaler, D. and Hopps, C.: Multipath Issues in Unicast and Multicast Next-Hop Selection, RFC 2991 (2000)

(平成 19 年 6 月 12 日受付)
(平成 19 年 12 月 4 日採録)

柏崎 礼生（正会員）

平成 11 年北海道大学工学部システム工学科卒業．平成 15 年同大学大学院修士課程修了．平成 17 年同大学院博士課程中途退学．工学修士．現在，同大学院情報科学研究科コンピュータサイエンス専攻助教．適応的ネットワークルーティングに関する研究に従事．情報ネットワークの可視化，人工生命，アニメーション，フィギュアに興味を持つ．電子情報通信学会，人工知能学会，IEEE，ACM 各会員．

小林 悟史

平成 6 年北海道大学工学部情報工学科卒業．工学士．同年デービーソフト（株）入社．平成 9 年（株）ネクステック設立．現在，同社取締役副社長．AS 間経路制御，MPLS，オーバレイネットワーク，インターネットの品質計測の研究および，IPv6 を中心としたネットワークプログラミングの開発に従事．

河合 修吾

平成 3 年北海道大学文学部行動科学科卒業．文学士．同年デービーソフト（株）入社．平成 8 年（株）コアシステム入社．平成 9 年（株）ネクステック取締役着任．平成 17 年より同取締役副社長，現在に至る．主にネットワークシステムの構築および監視運用に従事．その他，IP のドメイン内および AS 間経路制御の研究および運用，VPN システムの開発および運用に従事．

大石 憲且

平成 3 年北海道大学農学部農芸化学科卒業．農学士．同年デービーソフト（株）入社．平成 7 年任意団体（当時）北海道地域ネットワーク協議会で BGP4 ならびに AS 運用の研究．平成 9 年（株）ネクステック設立．代表取締役社長（現任）．平成 15 年 NPO 法人北海道地域ネットワーク協議会理事就任（現任）．医療情報ネットワーク相互接続研究会，次世代 IX 研究会，北海道広域高速学術ネットワーク検討会等で，広域経路制御，MPLS，インターネット VPN，メトロエッジの開発・実用化に従事．

高井 昌彰（正会員）

昭和 58 年東北大学工学部電子工学科卒業．昭和 63 年同大学大学院工学研究科博士課程修了．工学博士．同年東京大学理学部助手．平成元年北海道大学工学部講師．平成 4 年同助教授．平成 7 年同大学大型計算機センター助教授．平成 15 年同大学情報基盤センター教授．平成 16 年同大学情報基盤センター副センター長．平成 18 年同大学 CIO 補佐官．現在に至る．超並列・分散処理システム，コンピュータグラフィックス，コンピュータネットワークの研究に従事．電子情報通信学会，IEEE，国際 CIO 学会各会員．