

片方向遅延を用いたトラフィックエンジニアリングアルゴリズムの適応性評価

An evaluation for adaptive traffic engineering algorithm with using one way delay.

柏崎 礼生[†] 小林 悟史[§] 河合 修吾[§] 大石 憲且[§] 高井 昌彰[‡]

Hiroki Kashiwazaki Satoshi Kobayashi Shugo Kawai Norikatsu Ohishi Yoshiaki Takai

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一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トラフィックはさらに増大し、その結果、輻輳はより悪化する傾向が見られ、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] による対処も行われている。TE に関する研究をまとめると、より効率的な 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. より多くのトラフィック要求を系全体で許容できること

の四点に集約することが可能である。筆者らはこの四点に対応するために遅延時間を用いた適応的経路制御アルゴリズムを提案した [2]。本稿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。本手法は常時、計測パケットを全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。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、IP ネットワーク上で評価実験を行い、静的ルーティングに対する提案手法の有効性を示す。

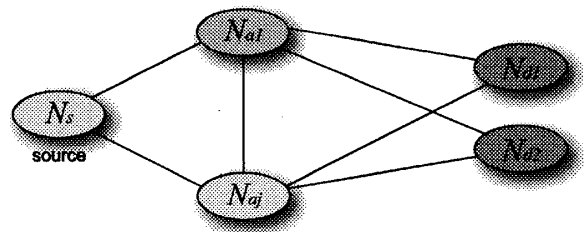
2. 提案アルゴリズム：NREI

本稿では 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) を有しており、任意の目的ノード N_d に向けた次ノード候補 N_a に長さ w のスコア列 $L_{N_d N_a} = \langle S_{N_d N_a}^1, S_{N_d N_a}^2, \dots, S_{N_d N_a}^w \rangle$ が与えられる。このスコア $S_{N_d N_a}^j$ 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}	$S_{N_{d1} N_{a1}}^1, S_{N_{d1} N_{a1}}^2, \dots, S_{N_{d1} N_{a1}}^w, S_{N_{d1} N_{a1}}^w$
	N_{a2}	$S_{N_{d1} N_{a2}}^1, S_{N_{d1} N_{a2}}^2, \dots, S_{N_{d1} N_{a2}}^w, S_{N_{d1} N_{a2}}^w$
N_{d2}	N_{a1}	$S_{N_{d2} N_{a1}}^1, S_{N_{d2} N_{a1}}^2, \dots, S_{N_{d2} N_{a1}}^w, S_{N_{d2} N_{a1}}^w$
	N_{a2}	$S_{N_{d2} N_{a2}}^1, S_{N_{d2} N_{a2}}^2, \dots, S_{N_{d2} N_{a2}}^w, S_{N_{d2} N_{a2}}^w$
N_{d1}	N_{d1}	$S_{N_{d1} N_{d1}}^1, S_{N_{d1} N_{d1}}^2, \dots, S_{N_{d1} N_{d1}}^w, S_{N_{d1} N_{d1}}^w$

図1 スコア付き経路制御表

各ノードは一定時間ごとに任意の目的ノード N_{dst} に対して、計測パケットを全ての次ノード候補 $N_{a1} \dots N_{a_j}$ (j は次ノード候補総数) に対して送信し、自ノードから任意の目的ノードまで、各次ノード候補を経由した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} に対応するスコア列の先頭 $S_{N_{dst} N_{a_n}}^1$ 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、列長を越えたスコアを破棄する。すなわち、スコア列には常に最新の 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

[†] 北海道大学大学院情報科学研究科

[‡] 北海道大学情報基盤センター

[§] 株式会社ネクステック

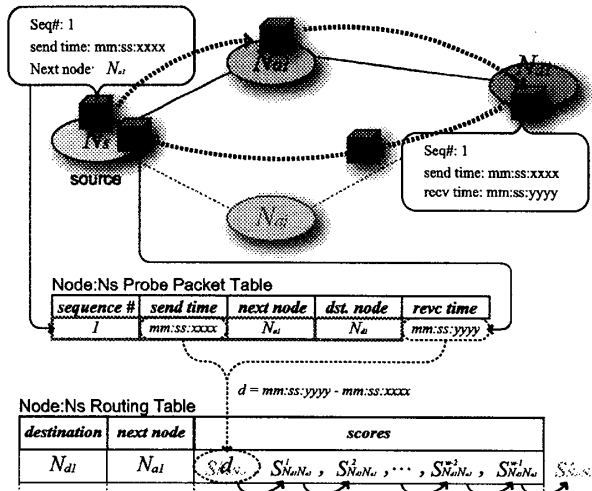


図2 計測パケットの動作と経路制御表の更新手順

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路した次ノード、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路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。

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} , ($1 \leq i \leq r$, 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、 $Q_{a_i} = \sum_{m=1}^w C_m S_{N_d N_{a_i}}^m$ 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 $\lambda (> 0)$ 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 (1) 式で表される。

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

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れたノードが選択される確率を強める働きがある。明らかに $\lambda \rightarrow \infty$ 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 (1) 式を用いる方式を NREI (Network adaptive Routing algorithm for Environmental Intelligence) と呼ぶ。

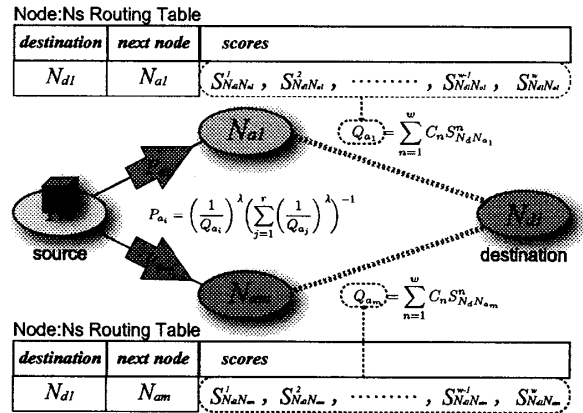


図3 送付先の確率的選択

2.3 per-packet 方式

トラフィック要求に対して動的な負荷分散を行う方法として、与えられた配分割合での負荷分散を実現するためにパケット単位で制御を行う per-packet 方式と、コネクション単位で制御を行う per-flow 方式がある。per-packet 方式は与えられた配分割合を正確に実現する事が可能であり高速な処理を実現する事ができるが、アプリケーションによってはパケットの到着順により通信品質の低下を引き起こす事も考えられる。per-flow 方式ではパケットの到着順が保証されるために通信品質は保たれるが、コネクションの粒度によっては配分割合を正確に実現する事が出来ず、また処理速度が per-packet 方式より劣るためボトルネックとなりやすい。本研究ではバックボーンネットワークがあらゆるアプリケーションの通信品質を保持しようとはせず、よりエンドユーザに近い側 (アプリケーションあるいはアプリケーションサーバやアプリケーションゲートウェイ) に通信品質の保持を任せる立場を取り、per-packet 方式の負荷分散処理を採用する。

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、以下の三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) 及び (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もつ。計測パケットが意図した経路を通るように、パケットには SSRR (Strict Source Routing and Record) オプションが指定される。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 に返送する機能をもつ。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 (1) 式のパラメータ 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。 $\lambda \rightarrow \infty$ 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。

計測パケットの送信頻度は 100ms であり、経路制御表のスコア列のサイズ w は 1000 で、スコア列には 100 sec 分の片方向遅延時間情報が格納されることになる。配分割合の計算頻度は 10 sec であり、経路制御の処理能力を優先するために配分割合は 10% 単位で丸めている。パケット損失が発生した際のペナルティ遅延値 d_p は可変だが、本論文では全ての実験で $d_p = 1000ms$ としている。

3.2 実験ネットワーク

本評価実験では、図 4 に示す四拠点に AR (AR1: 札幌, AR2: 富山, AR3: 山梨, AR4: 高知) を設置した。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) [4] 上に構築された RIBB-II (地域間相互接続プロジェクト, Regional Internet Backbone II) ネットワーク [5] 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 同士を IPv4 over IPv4 によるトンネルで接続し、オーバーレイネットワー

クを構成している。このオーバーネットワークでは、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)

の中継地が異なる二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 10Mbps 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) を設置している (図 4)。

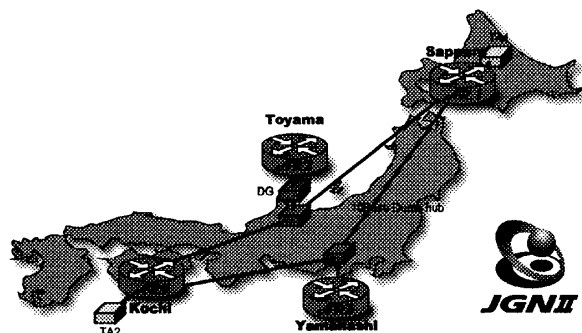


図 4 評価実験ネットワークの構成

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越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、既存の静的ルーティングでの結果を比較する。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 iperf 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 1Mbps から 10 Mbps まで

1Mbps 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。図 5 にその平均値と標準偏差を示す。

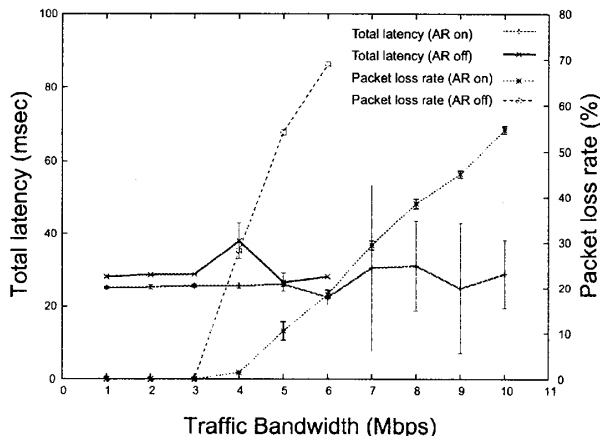


図5 トラフィック量増大に対する片方向遅延時間とパケット損失率の変化

AR2 と AR3 とは半二重の 10Mbps 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越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過ぎ、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ー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 7Mbps 以上のトラフィック要求でジッター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 [6]。そこでパケット損失 20% を一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。

4. 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。その結果、トラフィック量の変化に応じて配分割合を変化させて、より大容量のトラフィックを許容できるネットワークを実現できることを示した。また、片方の経路品質の悪化に対してパケット損失を行わない範囲で配分割合を連続的に変化させ、自律的に迂回路により多くのトラフィックを流す事を実現した。

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負荷分散を行う。そのため、総遅延時間の大きな経路にも常にトラフィックが流れるため、少ないトラフィック要求に対しては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常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、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。また、より迂回路が多く存在するトポロジにおいて提案するアルゴリズムは高い優位性を示す事が評価実験から分かっている。

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事が考えられる。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率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事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] 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, pp3308-3318 (2006).
- [3] 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).
- [4] Japan Gigabit Network II, Advanced Testbed Network for R&D: <http://www.jgn.nict.go.jp/>
- [5] 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).
- [6] 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).