情報処理学会論文誌 Vol.54 No.1 404–413 (Jan. 2013)

オーバレイルーティング網における広帯域映像配信のための適応的トラフィックエンジニアリング

柏崎 礼生[1,a)] 高井 昌彰[2,b)]

受付日 2012年1月10日, 採録日 2012年10月10日

概要：ハイビジョン機器などの普及により高精細映像のストリーミングが可能となった．大きなトラフィック要求を発生させるコンテンツを複数拠点で相互配送する場合，クロストラフィックの発生によりパケット損失が起こり，映像品質の劣化が問題となる．本論文では小中規模のマルチホーム環境に適用可能な，離散イベント型シミュレータを用いた適応的トラフィックエンジニアリング手法を提案する．提案手法はネットワークの送信元と送付先の組合せからなる経路の組合せに対して，ネットワーク情報の実測値をもとに漸進的な探索処理を行い，実時間で準最適な経路を適用する手法である．クラウドコンピューティング環境の普及により強大な計算機資源を利用することが可能となったが，本提案手法はこれを利用して高速な経路の組合せ評価を実現する．本提案手法の実時間対応性とトラフィック負荷分散をシミュレータを用いて評価し，その有効性を示す．

キーワード：トラフィックエンジニアリング，マルチホーム，ネットワークシミュレーション

An Adaptive Traffic Engineering for Deliverying High Bandwidth Movie on Overlay Routing Network

Hiroki Kashiwazaki[1,a)] Yoshiaki Takai[2,b)]

Received: January 10, 2012, Accepted: October 10, 2012

Abstract: By widely spreading of high definition cameras and broadband networks, now we can stream over-HD quality movies onto the Internet easily. When we distribute the movies each other between some facilities, packet losses caused by cross traffic become a serious problem. This paper proposes an adaptive traffic engineering method by using discrete event simulator. This method searches suboptimal paths combination from the enormous combinations of paths determined from the pairs of all source node and all destination node. We use the cloud computing service to evaluate these combinations rapidly, which we can utilize powerful computing resource by. This method is tested in ns-2 simulation on the 11 nodes network. The results show that proposed traffic engineering method has better adaptability and high-speed performance in cross traffic avoidance.

Keywords: traffic engineering, multi-home network, network simulation

1. 背景

製造業分野におけるロボットの遠隔操作や医療分野における遠隔診断・遠隔手術，さらには芸術分野でのリアルタイムインタラクション表現などにおいて，中継拠点間の高精細映像配送のためのネットワーク利用が拡大している．一般に映像データのネットワーク配送は，日食観測などの宇宙・天文学イベントや地域の祭典などのリアルタイム中継においても古くから行われているものである．しかし高精細映像配送によるトラフィック要求はネットワーク帯域を占有し，他のアプリケーションの通信品質に大きな影響を与えるため，複数のネットワーク回線を有する中継拠点

[1] 大阪大学サイバーメディアセンター
Cyber Media Center, Osaka University, Ibaraki, Osaka 567–0047, Japan
[2] 北海道大学情報基盤センター
Hokkaido University Information Initiative Center, Sapporo, Hokkaido 060–0811, Japan
[a)] reo@cmc.osaka-u.ac.jp
[b)] takai@iic.hokudai.ac.jp

においては異なるトラフィック要求をネットワーク回線ごとに振り分ける配送経路の適切な組合せを設定することが求められる．

中継拠点に接続された各回線は，その他のアプリケーションでも利用されているため，高精細映像配信を行おうとする各拠点からのトラフィック要求量と同時に，実際に許容可能なトラフィック要求量を正確に把握する必要がある．また，各拠点・各回線ごとに帯域制限などのネットワークポリシが課せられていることもある．異なる拠点で発生したトラフィックが 1 つの回線に重畳した場合には，クロストラフィックが発生し，配送される映像の品質に影響を与える．しかし各拠点の局所的なネットワーク情報だけでは，この通信品質の劣化を回避することは困難である．従来の映像データのネットワーク配送では，大きなトラフィック要求の発生に対して輻輳を発生させない配送経路の組合せの導出と経路の組替えを事前にオペレータの人手で行っており，この運用コストの大きさが映像配送におけるネットワーク利用拡大の阻害要因にもなっている．

そのため，各拠点がトラフィック要求に対する配送経路の組替えを意識することなく複数のネットワーク回線を利用することができ，トラフィック要求を適応的に各回線に配分することでネットワーク資源を有効利用するトラフィックエンジニアリング（Traffic Engineering，以下 TE）[1] 手法が研究・開発されている．計算機の処理能力が著しく向上した現在において，高精細映像配送を対象としたリアルタイムの TE 手法は単にトラフィック要求を分散させるだけでなく，計測されたネットワーク情報に応じた自律的な経路組替えを実現することが求められている．映像の通信品質に影響を及ぼすパケット損失率やインタラクションの応答性能に影響を及ぼす総遅延時間を低減することで，多様なトラフィック要求に対して細やかな制御を実現する必要がある．

2. 本研究の目的

従来の TE 手法は，計測されたネットワーク情報の累計からトラフィック要求量の最大値を見積もりトラフィックの配分を行うオフライン方式と，実トラフィックから計測された情報をもとにリアルタイムにトラフィックの配分を行うオンライン方式とに分けられる [2]．映像配送においては拠点間を結ぶネットワーク回線の利用可能帯域をリアルタイムに検出する必要があるため，本研究ではオンライン方式の TE 手法を採用する．

オンライン手法の TE 手法としては同一組織内のバックボーンにおける MPLS-TE が実運用で利用されているが [3]，MPLS ルータの相互接続性における問題点が指摘されており [4]，広域に分散した不均一なネットワークを結ぶ手法として採用することは難しい．計測されたネットワーク情報に基づいて OSPF や BGP のパラメータをリアルタイムで調整する手法も提案されているが [5], [6]，拠点間が複数のネットワーク回線で接続されている環境においては，既存の経路制御アルゴリズムで柔軟なトラフィック配分を実現することは困難である．

トラフィック要求の配分を細かい粒度で実現する方法として，パケット単位での配分を行う手法があげられるが [7]，リオーダリングの低減やループの回避が必要である [8]．FEC を用いてデータを冗長化し，受信側で復号することで対応する手法も提案されているが [9]，拠点数が増えるにつれて符号・復号に要する計算量も増大するため，本研究が想定する多対多の高精細映像配送での問題解決が難しい．

マルチホーム環境における TE において，低レイヤでのルーティングやスイッチングで細やかな制御を実現するためには，拠点どうしのネットワーク運用の協調が必要となり，また手法によっては特定のプロトコルに対応した機材を新たに増設する必要があるため，コストの増大が問題となる．一方でオーバレイネットワークを用いた TE では，運用コストを増大させることなく既存の設備を用いて安価に仮想的なネットワークを構築することができ，要素ネットワークのトポロジを意識せずに TE が可能となるというメリットがある．しかし一般にオーバレイネットワークは，その構成基盤となる要素ネットワークの影響を受けるため，ネットワーク品質や安定性の問題が指摘されている．この問題に対処した先行研究として，ノード間のネットワーク情報である片方向遅延時間を評価値とした適応的経路制御手法 [10] をオーバレイネットワーク上で実装し，地域間相互接続網 [11]（RIBB）での映像伝送における実証実験で評価を行った例がある [12]．

以上をふまえ，本論文ではネットワークポリシの異なる複数拠点間において柔軟でリアルタイム性のある TE を実現するため，広域に分散配置された 10 カ所程度の拠点間で高精細映像をオーバレイルーティングで配送することを対象とした TE 手法を提案する．本提案手法は，要素ネットワークのネットワーク情報を収集して利用することにより要素ネットワークの変動に対応することが可能であり，多様なトラフィック要求を各ネットワーク回線に適応的に配分する経路組合せを探索し，組替えを行うものである．シミュレーション実験を行うことにより，運用上支障のない現実的な時間内でパケット損失や総遅延時間を低減させる経路の組合せの探索および組替えを行うことができることを示す．

3. 基本的フレームワーク

本論文の TE 手法は，複数の拠点が，1 つ以上の互いに独立したネットワーク回線で接続されたオーバレイネットワークを対象とする．トポロジが公知である学術ネットワークや研究用テストベッドにおいてはその情報を用い，

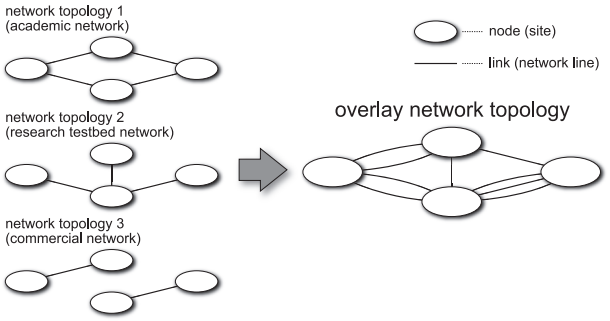

図 1 オーバレイネットワークのトポロジ
Fig. 1 Overlay network topology.

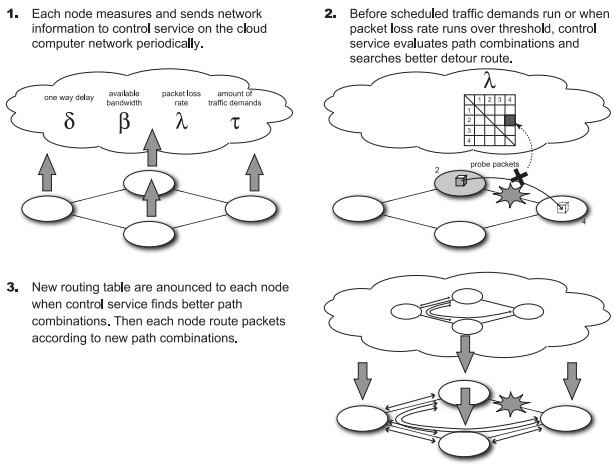

図 2 オーバレイルーティングの経路組替え
Fig. 2 Rerouting for overlay network.

またトポロジが公知でない商用ネットワークにおいては遅延計測によるトポロジの推定を行い [13]，これらのトポロジ情報を重畳させて各拠点間を結ぶすべてのネットワーク回線のトポロジを 1 つの多重有向グラフで表す（**図 1**）．

オーバレイネットワークにおいて適応経路制御（AdaptiveRouting: AR）を実現するため，各拠点のゲートウェイに AR ルータを設置し，各ネットワーク回線を接続する．AR ルータはオーバレイルーティング機能に加え，隣接拠点との片方向遅延時間，パケット損失率，利用可能帯域および他拠点へのトラフィック要求量の計測を行う機能を有している．AR ルータは拠点を結ぶネットワーク間でのトラフィック要求に対してのみオーバレイルーティングを行う．

経路組合せの評価計算や経路の組替えを AR ルータに広告する管理サービスをオーバレイネットワークから分離するために，管理サービスの機能はインターネット上のクラウドコンピューティング環境で実装するものとする．経路組合せの評価計算は，あらかじめスケジュールされたトラフィック要求が発生する一定時間前に行い，輻輳を低減させる経路に組み替える．また，この評価計算はパケット損失率の上昇を検知した際にも行われる．AR ルータによるネットワーク情報の収集から経路の組替えまでの流れを以下に示す（**図 2**）．

(1) 各拠点の AR ノードは計測したネットワーク情報を定期的にクラウドコンピューティング環境上にある管理サービスに送信する．
(2) 予約されたトラフィック要求が発生する一定時間前，あるいは指定した閾値を超えるパケット損失率が計測された時点で，管理サービスは経路組合せの評価計算を開始し，パケット損失率を改善できる経路を探索する．
(3) パケット損失率と総遅延時間を改善できる経路が発見されると，管理サービスはすべての拠点の AR ルータに経路の更新をアナウンスし，AR ルータは新しい経路組合せに従って拠点間のオーバレイルーティングを行う．

4. 経路組合せの評価計算

4.1 ネットワーク情報の収集

各拠点のゲートウェイに設置された AR ルータは以下の 4 つのネットワーク情報を収集する．

- 隣接した拠点との間の片方向遅延時間 δ
- 隣接した拠点との間のパケット損失率 λ
- 隣接した拠点との間の利用可能帯域 β
- 自拠点から他のすべての拠点に向けて発生するトラフィック要求量 τ

n 個の拠点を $N_1, \cdots, N_n$ と表すとき，拠点 N_i の AR ルータは隣接する拠点 N_j の AR ルータに対して計測パケットを送信し，拠点 N_i，N_j 間の片方向遅延時間 δ_{ij} の計測を行う．また計測パケットを用いた利用可能帯域 β_{ij} の推定を行う．N_i の AR ルータから送信された計測パケットには送信時刻 t_{send} が記載され，隣接する拠点 N_j の AR ルータ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 N_j から N_i へ向けて送信される計測パケットの中にカプセリングされて返送され，計測パケットが N_i に到着するとカプセルから遅延時間情報が取りだされる．その後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．計測パケットには連続した番号を記載しておく．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損失が発生したものと見なし，単位時間あたりのパケット損失率を計測する．映像伝送で求められる利用可能帯域変動の時間粒度を考慮し，利用可能帯域推定には pathChirp [14] や流体モデルを用いた測定手法 [15] を用いるものとする．

拠点 N_i，N_j 間に複数のリンクが用意されている場合は，それぞれの回線における片方向遅延時間，利用可能帯域，およびパケット損失率の計測を行い，個々のリンクを区別して扱う．また拠点 N_i で発生した拠点 N_j へのトラフィック要求量を τ_{ij} とするとき，拠点 N_i の AR ルータは自拠点以外のすべての拠点に対するトラフィック要求

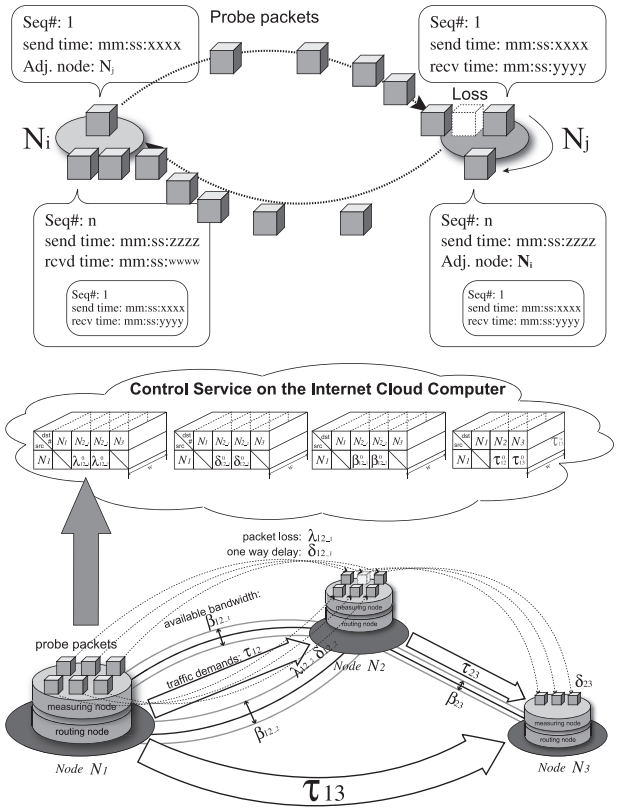

図 3 ネットワーク情報の計測と収集
Fig. 3 Measuring of network information.

量 $\tau_{i1}, \tau_{i2}, \cdots, \tau_{in}$（$n \neq i$）を計測する．計測パケットは隣接拠点にのみ送られ，オーバレイルーティングの要素ネットワークのルーティングに従い隣接拠点に転送される．一定時間ごとに計測パケットは送信され，収集されたネットワーク情報はクラウドコンピューティング環境上で動作する管理サービスに送られる（**図 3**）．

片方向遅延の計測には各拠点の時刻が同期していることが求められる．遅延時間を計測するため，100 μs オーダの精度での時刻同期が必要となるが，NTP の精度向上手法を用いて [16]，輻輳時にも高精度の時刻同期を実現できる対処を施すものとする．またネットワーク情報を 10 msec ごとに計測して 1 sec ごとに管理サービスに送信した場合でも，ネットワーク情報の送信に要する帯域は 10 kbps 程度であり，他のトラフィック要求への影響は無視することができるものとする．

4.2 管理サービス

クラウドコンピューティング環境上に配置された管理サービスは，ネットワーク情報の蓄積，トラフィック要求のスケジュール予約管理，および経路組合せの評価計算を行う．管理サービスは各拠点の AR ルータから送られてくる片方向遅延時間 δ，パケット損失率 λ，利用可能帯域 β およびトラフィック要求量 τ に対して各々の長さ w のバッファを用意し，新たに得られたネットワーク情報をバッファの先頭に格納する．新しい情報が書き込まれる際に，最も古い情報を破棄することで，つねに最新の w 個のネットワーク情報をバッファに保存する．

予約されたトラフィック要求が発生する一定時間前になると，新たなトラフィック要求の発生によってパラメータが変化した条件での経路組合せの評価計算を行う．経路組合せの探索においては，環境の動的な変化に対する応答性を向上させるため，ある水準の準最適解を 1 度に求めるのではなく，漸進的な解空間の探索処理を行い，一定時間ごとに評価の良い解が発見され次第，ただちに経路の組替えを行う．すなわち，管理サービスは各拠点の AR ルータに対して，パケット損失を一定水準以上改善することが期待できる経路を定期的にアナウンスする．また各拠点の AR ルータから収集したネットワーク情報を参照し，閾値を超えたパケット損失率の上昇を検知した場合においても，前述と同様な経路組合せの評価計算を行う．ただし，規模の大きなネットワークにおいてはすべての経路組合せを評価するために膨大な時間を要するため，運用上設定した時間で計算を打ち切るものとする．評価計算の開始時点でのネットワーク情報を用いて評価計算を行うものとし，評価計算中に変動したネットワーク情報に追従して評価計算のパラメータを変更することは行わない．また，暫定的な経路に停留することを避けるため，一定時間ごとに最新のネットワーク情報をもとに経路組合せの評価計算を行い，パケット損失や伝送総遅延時間をあらかじめ設定した値以上改善することができる経路が見つかった場合に経路組替えアナウンスを行う．アナウンスの頻度は各 AR ノードとクラウドコンピューティング環境上の管理サービスとの遅延時間の最大値よりも大きな値とし，経路組替えを行うタイミングを同期できるようにする．

本手法ではループのない経路制御を実現するため，送信元ノード（src）と目的ノード（dst）の組合せに対してループのないすべての経路を探索の候補とし，往路と復路は同じ経路をたどるものとする．すべての src と dst のペアに対して 1 つずつ経路を選択し，経路組合せを定めると，src と dst の各ペアのトラフィック要求量は既知であるため，その経路組合せに従ってオーバレイルーティングを行った場合の総パケット損失量と総遅延時間を離散イベント型シミュレーションで求めることができる．

4.3 経路のスコアリング

すべての src と dst 間の経路にそれぞれ総遅延時間に基づくスコアを付与し，スコアの小さい経路を優先的に選択して全体の経路組合せを構築する．各経路の総遅延時間は計測された片方向遅延時間を用いて算出する．src:N_s，dst:N_d のすべての経路において最小の総遅延時間を M_{sd} とするとき，src:N_s，dst:N_d の各経路の総遅延時間を M_{sd} で除した値をその経路のスコアとする．

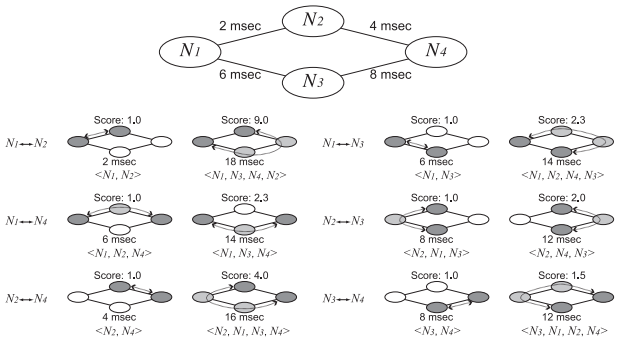

図 4 総遅延時間に基づく経路のスコアリング
Fig. 4 Path scoring based on total latency.

図 4 に示す 4 ノード（4 拠点）のリング状トポロジを例とすると，2 ノードの組合せ $[N_1, N_2]$ において N_1 から N_2 に至るループのない経路は $\langle N_1, N_2\rangle$ と $\langle N_1, N_3, N_4, N_2\rangle$ の 2 つである．経路 $\langle N_1, N_2\rangle$ の総遅延時間は 2 msec，経路 $\langle N_1, N_3, N_4, N_2\rangle$ の総遅延時間は 18 msec であるため，経路 $\langle N_1, N_2\rangle$ のスコアは 1.0，経路 $\langle N_1, N_3, N_4, N_2\rangle$ のスコアは 9.0 となる．他の 2 ノードの組合せ，$[N_1, N_3]$，$[N_1, N_4]$，$[N_2, N_3]$，$[N_2, N_4]$，$[N_3, N_4]$ についても同様にそれぞれスコアを算出する．

スコアの小さい経路を優先的に採用して経路組合せを作り，その経路組合わせに対して評価計算を実行する．すなわち，得られたすべてのスコア値を昇順にソートしたリスト $\langle 1.0, 1.5, 2.0, 2.3, 4.0, 9.0\rangle$ の並び順に従って，小さいスコア値を持つ経路を選択し，可能なすべての経路組合せを作成する．スコアが同値の場合にはホップ数の少ない経路を選択する．最初に作られる経路組合せは，スコア 1.0 の経路を組み合わせたもの，すなわちすべての 2 ノードの組合せにおける最短遅延時間の経路を組み合わせた $[\langle N_1, N_2\rangle, \langle N_1, N_3\rangle, \langle N_1, N_2, N_4\rangle, \langle N_2, N_1, N_3\rangle, \langle N_2, N_4\rangle, \langle N_3, N_4\rangle]$ である．

次いでスコア 1.5 である経路 $\langle N_3, N_1, N_2, N_4\rangle$ を含めたすべての経路組合せを作成する．ただし，すでに作られた組合せは作らず，重複を排除する．そのため新たに作られる経路組合せは $[\langle N_1, N_2\rangle, \langle N_1, N_3\rangle, \langle N_1, N_2, N_4\rangle, \langle N_2, N_1, N_3\rangle, \langle N_2, N_4\rangle, \langle N_3, N_1, N_2, N_4\rangle]$ となる．

次にスコア 2.0 である経路 $\langle N_2, N_4, N_3\rangle$ を含めたすべての経路組合せを作成し，$[\langle N_1, N_2\rangle, \langle N_1, N_3\rangle, \langle N_1, N_2, N_4\rangle, \langle N_2, N_4, N_3\rangle, \langle N_2, N_4\rangle, \langle N_3, N_4\rangle]$ および $[\langle N_1, N_2\rangle, \langle N_1, N_3\rangle, \langle N_1, N_2, N_4\rangle, \langle N_2, N_4, N_3\rangle, \langle N_2, N_4\rangle, \langle N_3, N_1, N_2, N_4\rangle]$ が新たに作られる．このようにして残りの経路 $\langle N_1, N_3, N_4\rangle$，$\langle N_1, N_2, N_4, N_3\rangle$，$\langle N_2, N_1, N_3, N_4\rangle$，$\langle N_1, N_3, N_4, N_2\rangle$ を順に含めてすべての経路組合せを作成する．

スコアの小さな経路を優先的に用いて経路組合せを生成することにより，評価計算において総遅延時間の短い経路組合せが後から発見されることを防ぐことができる．またスコアが同じ場合にはホップ数の少ない経路を選択することで，トラフィックが重複する区間の少ない経路組合せを作ることができ，クロストラフィックの発生可能性の少ない経路組合せから評価することが期待できる．

ここでは遅延時間のみをスコアとして用いたが，利用可能帯域をスコアに加味することもできる．その場合は OSPF [17] を参考にして利用可能帯域の逆数を求め，遅延時間との加重和でスコアを計算する．

4.4 シミュレーションによる評価計算

経路組合せが定まれば，各ノードに src:N_s，dst:N_d のパケットが到着したときの送付先ノード N_n を示す経路制御表を作成し，この経路制御表に従う離散イベント型シミュレーションを行う．パケットを一定時間発生させ，そのすべてが目的ノードに到着するまでをシミュレーションすることにより，総パケット損失量 Λ と総遅延時間 Δ を算出する．離散イベント型シミュレーションで用いられるノード間の遅延時間，帯域，トラフィック要求量のパラメータは，それぞれ計測された片方向遅延時間 δ，利用可能帯域 β，トラフィック要求量 τ により決定される．本研究ではパケットのキューイングおよびパケット損失のモデルとして ns-2 の設計を参考にして [18]，すべての 2 ノードの組合せに対して一意に経路の定まる経路制御に特化した離散イベント型シミュレータを実装した．

ある経路組合せに対するシミュレーションが終了し，得られた総パケット損失量がその時点での準最適解（準最適な経路組合せ）の総パケット損失量 Λ_{min} よりも小さい場合には，この経路組合せを新たな準最適解として採用し，Λ_{min} および総遅延時間 Δ_{min} を更新する．なお総パケット損失量が Λ_{min} と同値の場合には，総遅延時間と Δ_{min} を比較し，総遅延時間が小さい場合にはその経路組合せを準最適解に採用する（**図 5**）．

ある経路組合せに対するシミュレーションの途中で，Λ_{min} よりも大きな総パケット損失量が算出された場合，そのシミュレーションを打ち切り，次の経路組合せの評価に進む．ある経路組合せの評価計算の結果は，他の経路組合せの評価計算に対して，シミュレーションの打ち切りにおいてのみ影響を与え，評価計算そのものに依存関係はない．したがって，ある評価計算の結果を待つことなく，次の評価計算を行うことができるため，評価計算の処理を並列化することができる．

5. 実験

5.1 シミュレータの精度評価

実装した離散イベント型シミュレータの精度を評価するために，ノード数 8 のネットワークを 3 種類，BRITE [19] を用いて生成した．各ネットワークを**図 6** に示す．グラフの生成モデルは Barabási-Albert と Waxman を用いた．Waxman ではノード配置方法によりグラフ構造が変化す

表 1 離散イベント型シミュレータで得られる結果の相対的精度
Table 1 Relative precision of the network simulator to ns-2.

	4-ASBarabasi	3-ASWaxman (random)	3-ASWaxman (Heavy Tailed)
総パケット損失量	99.8% ($\sigma = 2.13 \times 10^{-2}$%)	99.0% ($\sigma = 2.02 \times 10^{-2}$%)	98.8% ($\sigma = 2.00 \times 10^{-2}$%)
総遅延時間	98.6% ($\sigma = 5.88 \times 10^{-3}$%)	98.4% ($\sigma = 7.17 \times 10^{-3}$%)	99.4% ($\sigma = 5.61 \times 10^{-3}$%)

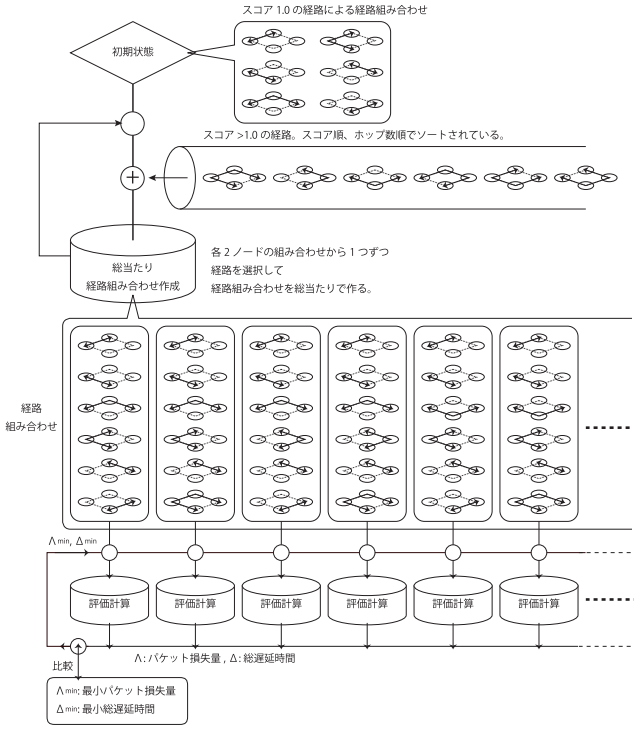

図 5 経路組合せの生成と評価計算
Fig. 5 Making path combinations and evaluation.

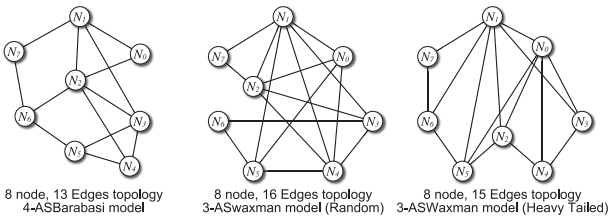

図 6 シミュレータの評価のためのネットワーク
Fig. 6 Networks for the simulator evaluation.

るため，Random と Heavy Tailed のノード配置方法を用いた．

ネットワークの帯域幅はすべて 100 Mbps とし，各リンクの片方向遅延時間は平均値 4 simulation steps（ss）の一様分布により与えた．映像配送網においてすべてのノード間でトラフィック要求が発生することは現実的でないため，過去 10 年間に RIBB で行われた映像配送での配信状況を参考に，各ノードは平均 3 ノードに対してトラフィック要求を発生させるパターンを作成した．

トラフィック要求量は固定ビットレートで 2^n Mbps（n は $0 \leq n \leq 4$ の範囲で一様分布）とし，トラフィック要求量のパターンを 200 個用意する．シミュレーション開始から 100 ss までパケットを発生させ，すべてのパケットが目的ノードに到着するまでの総パケット損失量と総遅延時間を算出する．

ネットワークシミュレータの比較対象として ns-2 を取り上げ，同様の条件ですべてのトラフィック要求量のパターンに対する総パケット損失量と総遅延時間を算出した．**表 1** は，本研究で実装した離散イベント型シミュレータで得られた結果を，ns-2 のシミュレーション結果（最適な経路組合せに従って経路制御が行われるように設定済み）に対する相対値として示したものである．実装した離散イベント型シミュレータが十分な精度で総パケット損失量と総遅延時間を求めることができることが分かる．

5.2 シミュレータの計算速度評価

実装された離散イベント型シミュレータの計算速度を評価するため，前節と同じ条件でトラフィック要求量のパターンに対してすべての経路組合せの評価を行い，最適解を求めた．その際の解探索の進捗状況と，発見された準最適解の変化を**図 7**，**図 8** に示す．横軸はシミュレーションの実行時間を対数で表している．縦軸にはトラフィック要求量の異なるパターン 200 個を並べ，最適解に到達するのに要する時間の順でソートしている．

トラフィック要求量の各パターンにおいて，総パケット損失量は，シミュレーション開始時の経路組合せにおける総パケット損失量と最適解の総パケット損失量で正規化し，色の濃淡変化で示されている．また，正規化された総パケット損失量がゼロとなる時点，すなわちトラフィック要求量の各パターンで最適解に至った時点を実線で結んでいる．加えて，同様の総パケット損失量が 0.2 となる時点を点線で結んでいる．経路組合せのシミュレーションは Intel Xeon 2.93 GHz（12 コア）を 2 基，メインメモリを 2 GB 搭載した計算機 10 台で行った．シミュレータは C++で実装し，FreeBSD 8.2 上で動作させた．

実験に使用したすべてのネットワークトポロジで 11 sec 以内に最適解に達しており，最適解の探索に 1 sec 以上を要しているのはトラフィック要求全パターンの 10%以下である．また，2 sec 以内には正規化された総パケット損失率 0.2 以下の解を得ることができている．

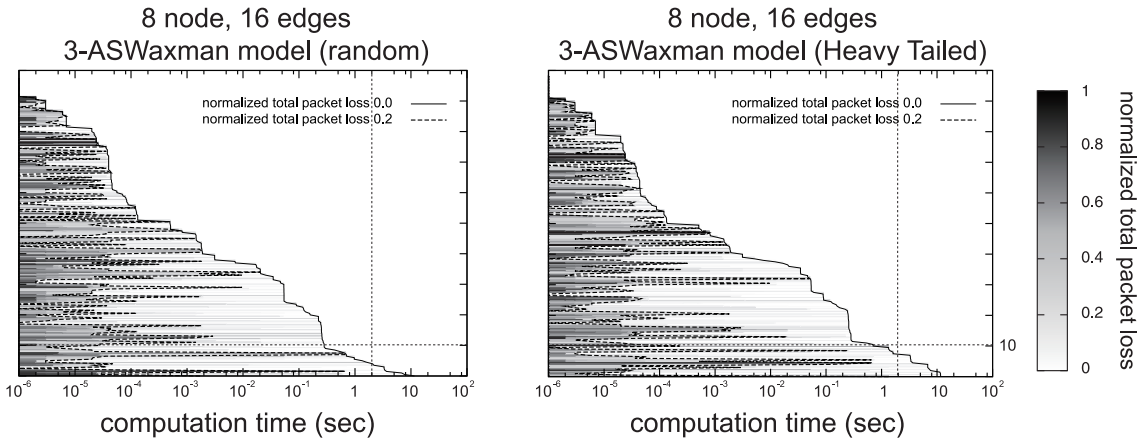

図 8 経路組合せの最適解への収束（Waxman モデル）

Fig. 8 Convergence to optimal solutions in the Waxman model topologies.

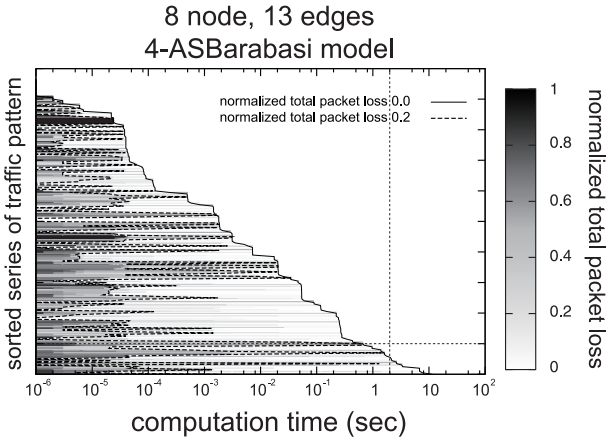

図 7 経路組合せの最適解への収束（Barabási-Albert モデル）

Fig. 7 Convergence to optimal solutions in the Barabási-Albert model topology.

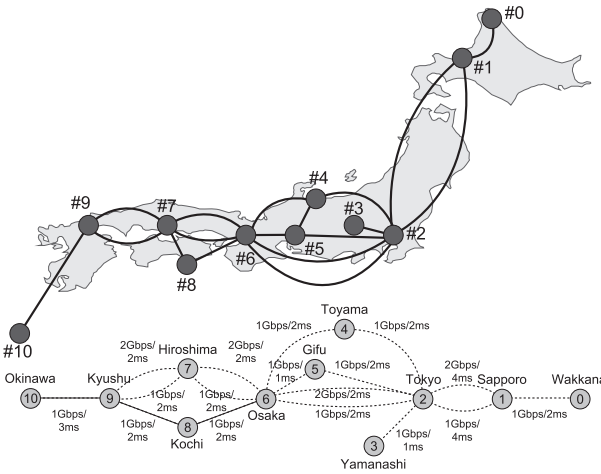

図 9 評価実験のネットワーク

Fig. 9 Network for evaluation experiment.

5.3 評価実験

実際のネットワーク運用における本提案手法の有効性を示すため，RIBB のトポロジを参考に，迂回路が複数存在する 11 ノードのトポロジにおいてクロストラフィックにより輻輳が発生する条件を設定し，評価実験を行った．

図 9 に示すネットワークを対象として ns-2 を用いた評価を行う．隣接するノードを結ぶリンクは全二重で，その帯域幅と遅延は図中に示すとおりである．経路組合せの再計算は，リンクのパケット損失率が 0.1%以上となったときに行われるものとする．ネットワーク情報の計測は 1 msec ごとに行われているものとし，経路組替えは 10 msec ごとに行われるものとする．経路評価計算ではパケット群を一定時間発生させ，発生したパケット群がすべて目的ノードに到着したときの総パケット損失率と総遅延時間を評価する．

また，トラフィック要求の時間変化に対する適応性をみるために，以下のようなイベントを時系列に沿って発生させる．

(1) 各ノードは他のすべてのノードに対して 2^n Mbps の CBR トラフィック要求を発生させる．n は $0 \leq n \leq 8$ の範囲で一様分布で定める．

(2) 時刻 2 sec から 6 sec まで#0 から#10 への CBR トラフィック要求を発生させることを予約する．トラフィック要求量は 512 Mbps とする．

(3) 時刻 6 sec で各ノードは他のすべてのノードに対して 2^n Mbps の CBR トラフィック要求を発生させる．n は $0 \leq n \leq 8$ の範囲で一様分布で定める．

(4) 時刻 8 sec から 12 sec まで#0 から#10 への CBR トラフィック要求を発生させる．このトラフィック要求は予約なしであり，トラフィック要求量は 512 Mbps である．

経路組合せのシミュレーションは 5.2 節と同様に Intel Xeon 2.93 GHz（12 コア）を 2 基搭載した計算機 10 台で行い，シミュレータは FreeBSD 8.2 上で動作させた．パケット損失が発生するリンクにおける総パケット損失量の時間変化の結果を**図 10** に示す．

時刻 2 sec に予約されたトラフィック要求があるため，TE を有効にしている場合は他のトラフィック要求発生の予約のない時間範囲において十分に早い時刻から経路組合

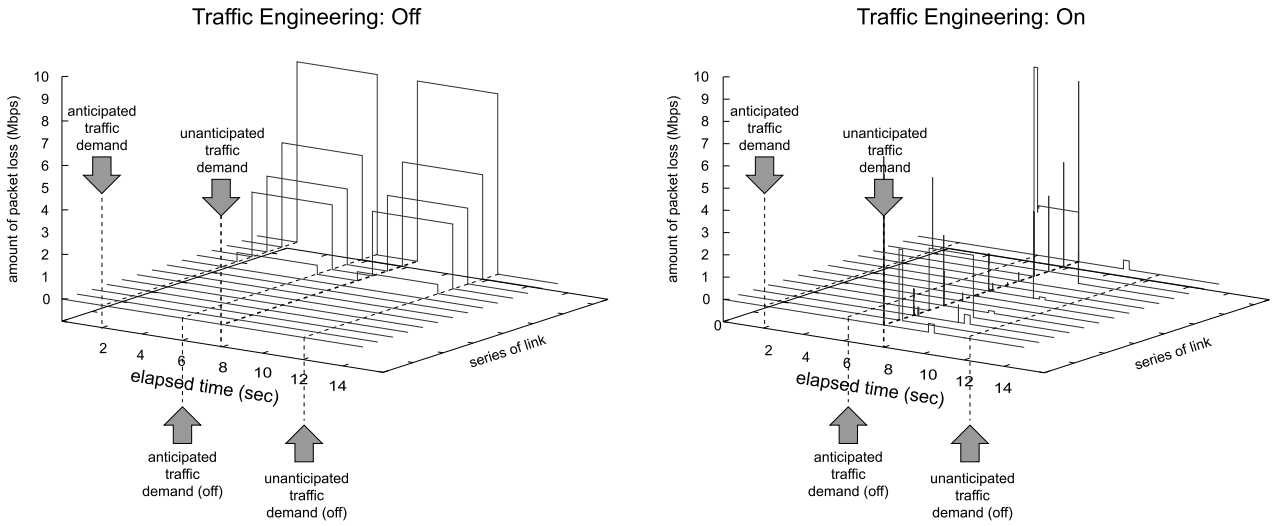

図 10 総パケット損失量の時間変化
Fig. 10 Change of packet loss.

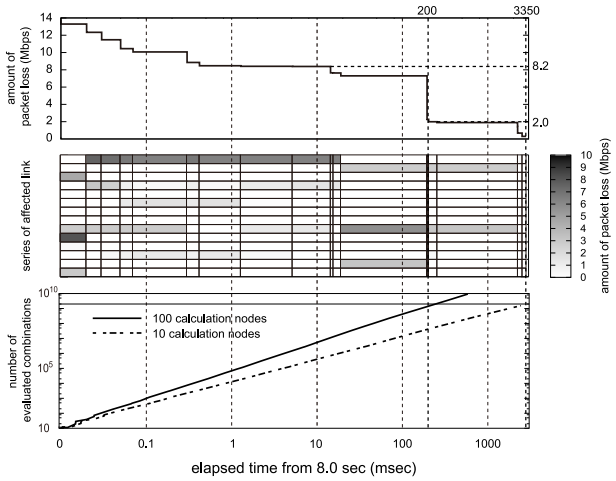

図 11 総パケット損失量と経路組合せ探索の時間変化
Fig. 11 Change of packet loss and path combination evaluation.

せの評価計算を開始する必要がある．ここでは時刻 0 sec の時点で経路組合せの評価計算が開始され，1,680 msec 後にパケット損失のない経路を発見し，その経路が広告されることによって，時刻 2 sec の時点では準最適な経路組合せに変更されている．そのためトラフィック損失量 0 を維持している．一方，時刻 8 sec から始まるトラフィック要求では予約がないため，TE を有効にしている場合でも事前に経路組合せの評価計算を行うことができずトラフィック損失が発生する．しかし，パケット損失増大を検出した AR ノードの管理サービスへの障害報告を起因としてネットワーク情報の収集が始まり，経路組合せの評価計算が行われるため，10 msec ごとにより最適な経路組合せに更新されていく．

時刻 8 sec から始まる予約のないトラフィック要求に対し，パケット損失が発生するリンクにおける総パケット損失量の時間変化の詳細を**図 11** に示す．準最適解の探索はトラフィック要求の発生から 1,000 msec 以内の変化が大きいため，時間軸を対数軸としている．上段のグラフが評価計算によって得られた総パケット損失量，中段のグラフは時刻 8 sec から 12 sec の間でパケット損失が発生する各区間のパケット損失量の変化である．

評価計算における準最適解は，0.1 msec 以下の粒度でより良い解が探索されるが，実際の経路に反映されるのは 10 msec 単位であるため，予約のないトラフィック要求の発生後 10 msec の時点で，総パケット損失量が 8.2 Mbps の経路に切り替わる．その後，数度の経路切替えが行われ，トラフィック要求発生後 200 msec の時点で総パケット損失量が 2.0 Mbps の経路に切り替わり，最終的に 3,350 msec 後にパケット損失のない経路に切り替わっている．

図 11 下段に，評価した経路組合せ数の時間変化を示す．10 台の計算機で評価計算を行った際の時間変化のほか，100 台の計算機で同様の計算を行った場合の推定時間を併記している．このトポロジにおける経路組合せの総数は 1.7×10^{58} 個ある．3,350 msec でトラフィック損失量が 0 となる経路組合せを発見するまで 4.1×10^{9} 個の経路組合せを評価している．

5.4 考察

トラフィック要求量の変動が緩やかな場合では，5.3 節の評価実験のようにネットワーク情報の変化の検知から経路組合せの評価までに十分長い時間をかけることができるが，ネットワーク情報の変化が急峻な場合では，評価された経路組合せを適用しても有効な障害回避とならない可能性がある．後者の場合は，新しい経路組合せを適用した後に計測されるパケット損失率と評価計算で推定されたパケット損失率に明確な差が現れるため，この差があらかじめ設定した閾値を超えるときには，経路組合せの評価計算途中であってもネットワーク情報を更新して新たな経路組

合せの評価を行うことが望ましい．また，オーバレイネットワークを構成する要素ネットワークの変動については予測が困難であるが，クラウドコンピューティング環境における仮想マシンのデプロイと起動は分のオーダで実現可能であるため，要素ネットワークの中長期的な変化に応じて必要な計算ノード数を増減させることも有効である．

先の 5.3 節で示した評価実験はシミュレータ上で行われているため，すべての拠点の時刻が同期している．しかし実際のネットワークにおいては時刻同期に差異があり，そのため経路組替えのアナウンスを行っても，切替えのタイミングの差異によりパケット損失が発生する可能性がある．また経路組替えの前後でパケットの到着順序に不整合が発生し，系が不安定になる可能性もある．これに対処する手法として，アプリケーション側で誤り訂正情報を付加し，バッファリングを行うことが考えられる．また提案手法では設定値以上に改善される経路組合せでなければ経路組替えのアナウンスが行われないため，この設定値をより大きな値にすることにより，頻繁な経路組替えを抑制することができる．

経路切替え時にもトラフィック要求はネットワーク上を伝送されており，このトラフィック要求は経路切替え前の経路を通過することを想定されてルーティングされている．そのため新しい経路の組合せによって作られる経路制御表によっては，送付先ノードが存在せずにパケットが破棄されることになり，ルーティングの安定性を過渡的に損なう可能性がある．この問題はパケットが目的ノードに到着する以前に経路制御表が組み替わることに起因するため，経路制御表の更新の時系列の順に各 AR ノードが経路制御表を 2^n 個保持する手法により対応することが可能となる．各経路制御表には $0 \sim 2^n - 1$ の識別番号を与え，パケットが送信元 AR ノードで発生する際に n ビットの識別子を付記し，経路制御表更新の時系列の中で，どの経路制御表を参照してルーティングされるかを定める．これによりパケットが目的ノードまで転送される途中で経路制御表の組替えが行われた場合においても，そのパケットは目的ノードに到着するまでは到達性が保証された過去の経路制御表でルーティングされることが可能となる．n の値は想定される経路切替えの頻度に依存して定められる．経路組替えが最短で 1 sec ごとに行われ，かつ最大総遅延経路における総遅延時間が 1 sec 未満である場合は，パケットの発生から到着までに 2 回以上の経路組替えは発生しないため $n = 1$ で対応可能となる．

6. まとめ

各拠点が片方向遅延，トラフィック損失率，トラフィック要求量を計測し，これを集約してトラフィックの適切な経路制御に関する準最適解を探索することで，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オンライン方式のトラフィックエンジニアリング手法を提案した．各拠点に設置された AR ルータは独立して自律的にネットワーク情報を計測，取得し，管理サービスに情報を集約して障害を検知する．スケジュールされたトラフィック要求の発生や障害の発見をトリガとして経路組合せの再計算を行い，管理サービスはすみやかに迂回経路を広告する．

評価実験により，本提案手法は 8 ノードのネットワークにおいて十分な精度で経路組合せを高速に評価することができ最適解を現実的な時間で得ることができることを示した．また 11 ノードの現実に即したネットワークにおける評価では，突発的なトラフィック要求の発生に対しても数秒間で準最適解を探索し，適応的に障害対応を実現できることを示した．拠点数の増大にともない計算量は非線形に増大するため，階層化による対処が今後の課題である．

謝辞 本研究は平成 24 年度北海道大学情報基盤センター共同研究「インタークラウドをより拡張するための地域間相互接続の調査検証」，平成 24 年度国立情報学研究所共同研究「“Trans-Japan Inter-Cloud Testbed” の構築に向けたネットワーク基盤に関する検討」，平成 24 年度学際大規模情報基盤共同利用・共同研究拠点公募型共同研究「分散クラウドシステムにおける遠隔連携技術」による支援を受けました．

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] 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).

[4] 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).

[5] Ye, T., Kaur, H.T., Kalyanaraman, S. and Yuksel, M.: Large-Scale Network Parameter Configuration Using an On-line Simulation Framework, *IEEE/ACM Trans. Networking*, Vol.16, No.4, pp.777–790 (2008).

[6] Tamura, H., Okubo, T., Inoue, Y., Kawahara, K. and Oie, Y.: Implementation and Experimental Evaluation of On-Line Simulation Server for OSPF-TE, *Proc. 7th International Conference on Hybrid Intelligent Systems* (*HIS 2007*), pp.259–264 (2007).

[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] 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).
[10] 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).
[11] 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).
[12] 柏崎礼生，小林悟史，河合修吾，大石憲且，高井昌彰：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，情報処理学会論文誌，Vol.49, No.3, pp.1194–1203 (2008).
[13] 菊池 豊，藤井資子，山本正晃，永見健一，中川郁夫：遅延計測による日本のインターネットトポロジーの推定，情報処理学会研究報告，Vol.2007, No.72, pp.103–108 (2007).
[14] Ribeiro, V., Riedi, R., Baraniuk, R., Navratil, J. and Cottrell, L.: pathChirp: Efficient Available Bandwidth Estimation for Network Paths, *Proc. Passive and Active Measurement Workshop* (2003).
[15] 浜 崇之，藤田範人，地引昌弘：トラフィック短期変動の影響を考慮した利用可能帯域測定方式，信学技報，CQ2007-26, pp.67–72 (2007).
[16] 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).
[17] Moy, J.: The OSPF Specification, RFC 3272 (1989).
[18] Issariyakui, T. and Hossain, E.: *Introduction to Network Simulator NS2*, Springer, ISBN-13: 978-1441944122 (2009).
[19] Medina, A., Lakhina, A., Matta, I. and Byers, J.: BRITE: An Approach to Universal Topology Generation, *International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems*, *MASCOTS'01* (2001).

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．同年北海道大学情報科学研究科助手（後に助教）．平成 22 年東京藝術大学芸術情報センター特任助教．平成 24 年大阪大学サイバーメディアセンター助教．適応的ネットワークルーティング，インタークラウドコンピュ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，萌え領域に興味を持つ．電子情報通信学会，人工知能学会，IEEE，ACM 各会員．

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（役員補佐相当職）．平成 23 年同大学情報基盤センター長．北海道大学評議員．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．平成 17 年 NPO 法人北海道地域ネットワーク協議会理事・副会長．平成 23 年情報処理学会北海道支部長．電子情報通信学会，IEEE，国際 CIO 学会各会員．