

離散イベント型シミュレーションを用いた トラフィックエンジニアリング手法の提案

柏崎 礼生^{†1} 高井 昌彰^{†2}

ハイビジョン機器の普及により安価に高品質な映像のストリーミングが可能となった。最近では4K映像を撮影・放映することができる研究拠点も増え、コンテンツのリッチ化がさらに進んでいる。こういった巨大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り、映像の品質が劣化する事が問題となる。本稿では小～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なトラフィックエンジニアリング手法を提案する。

A Proposal of Traffic Engineering Method by Using Discrete Event Simulation.

HIROKI KASHIWAZAKI^{†1} and YOSHIAKI TAKAI^{†2}

By widely spreading of high definition cameras and broadband networks, now we can inexpensively stream over-HD quality movies onto the Internet. The number of research facilities which have 4K resolution (its horizontal resolution is approximately 4,000 pixels) cameras and projectors have been increasing. Multimedia contents have been getting richer and richer. 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.

1. 背景

ネットワーク回線を用いた映像のストリーミングはインターネットの黎明期から行われており、技術の発展と並行して今なお進化し続けている。かつて一対一の片方向で行われていたストリーミングは、デバイスの高性能化、計算能力の向上、そしてネットワークの広帯域化により、高解像度の映像を多対多でリアルタイムに送受信する形に進化している。特に高解像度化の進化は著しく、4K映像を撮影・放映する事が可能な拠点も増加しており、また単に撮影された映像を放映するだけでなく、クラウドコンピュータサービスにより手軽に安価に利用できるようになった高性能コンピュータを用いて視覚化されたデータをリアルタイムに重畳表示する事で現実を強化する表現も可能となっている。利用分野も工業・医療分野における遠隔操作や遠隔治療といった双方向レイダシステムに始まり、科学技術イベント・地域系イベントのリアルタイム中継、複数台のHDカメラを用いて空間の全周映像

をネットワーク経路で遠隔地で再現する芸術分野でのリアルタイムインタラクションなど多様な分野に及んでおり、今後もさらなる応用と新たな利用分野の開拓が期待される。ネットワーク回線についても拠点間を結ぶ回線が学術網や研究網、キャリア網など複数の回線と接続する事が可能となり、マルチホーム環境を安価に整備・用意することができる。

しかしながら実際には、マルチホーム環境を用意することが可能であっても、そのうちの1つの回線しか使われていない事が少なくない。拠点に接続された各回線が実際の程度のトラフィックを許容する事ができるか、他拠点がどの程度のトラフィック要求を発生させているかを知る事が困難であり、アプリケーションごとの帯域制限といった特有の制約が各回線それぞれに異なったポリシーで課せられていることもある。それらの制約に配慮した上での効率的な利用を行うことが難しい。このような条件下では、極めて大きなトラフィックが一つの回線に重畳した結果、パケット損失を生じさせてしまうクロストラフィックが発生する。しかし各拠点の局所的な情報だけではこの損失を回避することが困難である。そのため従来の映像配送網では、大きなトラフィック要求の発生に対して事前に人手で配送経路のアレンジを行う必要があり、その運用

^{†1} 東京藝術大学 芸術情報センター
Art Media Center, Tokyo University of the Arts
^{†2} 北海道大学 情報基盤センター
Hokkaido University Information Initiative Center

コストの大きさが積極的に映像配信を行うことの阻害要因ともなっている。

このような状況の下、各拠点が配送経路のアレンジを意識することなく複数の回線を利用することができ、トラフィックを適応的に配分しネットワーク資源を有効活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。計算機の能力が著しく向上した現在における TE 手法は、単にトラフィックを分散させるだけでなく、リアルタイムに計測されたネットワーク情報に応じて自律的な経路組み替えを実現し、伝送品質に影響を及ぼす総遅延時間やパケット損失率を低減することで、肥大化・複雑化したトラフィック要求に対してより細やかな制御を実現する事が求められている。

2. 従来研究および本研究の目的

従来の TE 手法は大きく分けるとオンライン方式とオフライン方式とに分類することができる²⁾。オンライン方式の TE は実トラフィックの計測から経路を即座に迂回路への切り替えを実現する手法で、この手法の代表例として MPLS を用いた MPLS-TE が挙げられる³⁾。MPLS-TE は同一組織内バックボーンなどにおいては有用であるが、比較的高価で運用の難しい MPLS ルータを必要とする点や、異なる機種 MPLS ルータで相互接続性に不安定さが懸念される事から⁴⁾、拠点によって運用スタッフの技術レベルの差が大きい共同研究拠点などで気軽に導入して運用することが難しい。特に MPLS を用いたオンライン方式の TE では複数の LSP (Label Switched Path) をあらかじめ設定することを前提としている⁵⁾。本研究が対象とするマルチホーム環境の相互映像配信網においては多様な粒度のトラフィック要求が各々の拠点間で発生するため、LSP を予め用意する方法では取り得る全ての経路の LSP を張らなければ細やかな経路制御を実現する事が困難であり、運用が現実的とは言えない。一方で IP カプセリングを利用したオーバーレイルーティング手法を用いれば膨大な数の LSP を事前に用意することなく柔軟な経路制御を実現することが可能である⁶⁾。

MPLS を用いないオンライン方式の TE においてはパケット単位でトラフィックを振り分ける手法があるが、確率的なトラフィック配分方法では経路振動が発生することが知られており、リオーダーリングを発生によりジッターの増大やパケット損失率の増大につながり、伝送品質の安定性に影響を及ぼす。また、トラフィックを分散させることで TE を実現する手法においては、一般的に経路の遅延差により生じるリオーダーリングの低減やループの回避が必要であるとされる⁷⁾。リオーダーリングの発生やパケットの損失に対して 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⁸⁾、拠点数が増えるにつれて、また各拠点からのトラフィック要求量が増大するに伴い復号に要する計算量も増大する点において、本稿が想定する多対多の高品質映像送受信網での問題解決が難しい。トラフィック要求の送信元と目的地に対して経路を一意に定める決定論的なトラフィック配分手法を用いれば、リオーダーリングの低減やループの回避を実現しながら細かな TE を実現することが可能である。

以上のオンライン方式の TE に対してオフライン方式の TE とは、トラフィック要求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・設定する手法である。ネットワーク環境の変化に対して実時間で対応性するためにはネットワークを構成する全ての拠点間でネットワーク情報の計測を行い、計測されたデータを元に運用において現実的な時間で解を導くことが求められる。得られたネットワーク情報をもとに最適解を導く方法として、グラフ理論を用いた解法がある。ネットワークのある 2 点間の流量を最大化する ford-fulkerson 法は他の 2 点間の流量を考慮しないためクロストラフィックに対処する事が難しい⁹⁾。Applegate らは線形計画モデルを用いてトラフィック要求とその近接部分を含む周辺に対して効率的な経路制御を計算したが¹⁰⁾、この手法は特定のトラフィック要求を前提としており、実際のネットワークでは刻々とネットワーク情報は変化するため現実的な対応は難しい。最小遅延経路制御を実現するために複数経路を計算する MDVA¹¹⁾ などの手法も提案されているが、帯域幅を考慮に入れていないため過大なトラフィック要求によりパケット損失が発生し得る。

実際のネットワークでは他の拠点間で送受信されるトラフィック要求が他のトラフィック要求に対して影響を与え、一方で帯域制限といったネットワーク運用管理ポリシーによって回線品質も変化する。このように時々刻々と環境の変化するネットワークにおいてはトラフィック要求量を計測するだけでなく、ある時点で発生するトラフィック要求量に対する同時刻のパケット損失率も計測し、その回線のトラフィック要求量に対する損失特性を得た上で最適解を導かなければならない。こういった細かなパラメータが与えられた問題に対して、離散イベント型シミュレータを用いれば、トラフィック要求量、損失特性、そして伝送遅延をパラメータとして与えて、拠点間を結ぶネットワークで発生し得るパケット損失量やトラフィックの伝送遅延時間を評価することができる。離散イベント型シミュレータには最適解を探索する能力がないので、設定された問題に対する最適解を求める際には経路組み合わせを総当たりで評価する必要があり、全ての経路組み合わせについて評価を行った場合、計算量は膨大なものになる。

以上をふまえて本稿では、オーバーレイルーティン

グを対象とした決定論的な TE 手法を提案する。本稿で提案する TE 手法を実現するため、各拠点に適応的経路制御 (Adaptive Routing; 以下 AR) ノードを各拠点に設置し、隣接する拠点との片方向遅延時間とパケット損失率、および他拠点に対するトラフィック要求量を常時計測する。ネットワーク状況の変化を検出すると、あらかじめ設定された代表ノードは各拠点が計測したネットワーク情報を収集し、評価シミュレーションのパラメータを設定する。代表ノードは、有効そうな経路組み合わせを優先的に評価する総当たり計算を行い、現実的な時間内でトラフィック損失率と伝送総遅延時間の増大を抑制し得る準最適解を探索する。この評価により得られた経路組み合わせをオーバーレイネットワーク上で経路制御に適用することで、適応的なトラフィックエンジニアリングを実現する。本手法では送信元 (src) と目的地 (dst) の組み合わせに対してループのない経路を一意に定めるため、経路組み替え時以外ではパケットのリオーダーリングやループが発生しない。計測のために用いられるプローブパケットは拠点間で送受信される総トラフィック要求量に比べて十分に小さく、伝送品質に与える影響は無視できる。

3. 提案アルゴリズム

本稿で提案する TE 手法は、複数の拠点が、1 つ以上のお互いに独立した回線で接続されたネットワークを対象とする。各拠点間を結ぶ回線の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、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¹²⁾、これらのトポロジ情報を重畳させて各拠点間を結ぶ全ての回線のトポロジを有向グラフで表す (図 1)。

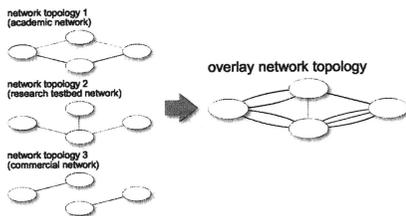


図 1 拠点間ネットワークトポロジの作成
Fig. 1 Description of Network Topology

各拠点には AR 用ノード群を拠点のゲートウェイに設置し、用意された各回線をこの AR 用ノード群に接続する。AR 用ノード群は拠点間ネットワーク間でのトラフィック要求に対してのみ IP カプセリングを施しオーバーレイルーティングを行う。AR 用ノード群

は以下の 4 つのノードにより構成される。

- 計測ノード
- AR 管理ノード
- 計算ノード
- オーバーレイルーティングノード

計測ノードは他拠点とのネットワーク情報を計測し収集する。計測から障害を検知し、経路組み替えを行うまでの流れを以下に示す (図 2)。

- (1) 計測ノードが一定時間の継続的なパケット損失を検知すると、これを障害とみなす。
- (2) 障害を検出した計測ノードは代表拠点の AR 管理ノードに報告し、経路の再計算アナウンスを要求する。
- (3) 要求を受けた代表拠点の AR 管理ノードは全ての拠点の計測ノードに最新のネットワーク情報を提供するようにアナウンスを行う。
- (4) アナウンスを受けた各拠点の計測ノードはネットワーク情報を代表拠点の AR 管理ノードに送信する。
- (5) 代表拠点の AR 管理ノードは受信したネットワーク情報から、経路組み合わせを評価するためのシミュレーションのパラメータ設定を定める。また、現行の経路組み合わせにおけるシミュレーション評価を行い、得られた評価値を目標値として設定する。経路組み合わせを並べ替え、各拠点ごとに異なる計算範囲を適切な範囲で切り出す。各拠点の計算ノードにシミュレーションのパラメータ、目標値、計算範囲を送信する。
- (6) 各計算ノードは受けたシミュレーション情報から、指定された範囲の経路組み合わせの評価シミュレーションを行い、目標値よりも良い評価値の経路組み合わせが見つかる代表拠点の AR 管理ノードに即座に通知を行う。指定された範囲の経路組み合わせ評価を全て終了した時は代表拠点の AR 管理ノードに要求を出し、新たな経路組み合わせ評価シミュレーションの範囲情報を得て計算を繰り返す。
- (7) 代表拠点の AR 管理ノードは得られた最良の経路組み合わせを採用し、一定時間ごとに全ての拠点のオーバーレイルーティングノードに対して経路組み替えのアナウンスを行う。頻繁な経路組み替えによる伝送品質の劣化を避けるため、前回切り替え時の評価値との差が、指定された組み替え閾値に満たない場合は経路組み替えを行わない。

以下で、AR 用ノード群に含まれる計測ノード、AR 管理ノード、計算ノードの働きについて詳しく説明する。

3.1 計測ノード

計測ノードは以下の 3 つのネットワーク情報を収集する。

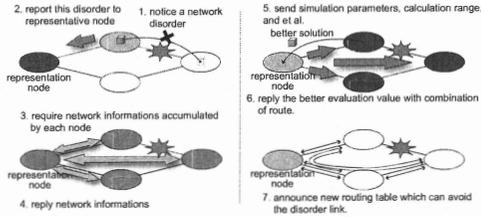


図2 障害検知から経路組み替えまでの流れ
Fig. 2 Procedure of reroute.

- 自拠点から他の全ての拠点に対して発生するトラフィック要求量 τ
- ネットワーク的に隣接した (1 ホップで到達可能な) 拠点との間の片方向遅延時間 δ
- ネットワーク的に隣接した拠点との間のパケット損失率 λ

他拠点に対するトラフィック要求量を計測するために計測ノードには sFlow のような高速に各拠点へのトラフィック要求量を計測することが可能な機能を実装する。各拠点を識別するために n 個の拠点を N_1, \dots, N_n と表す時、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} として、拠点 N_i の計測ノードは $\tau_{i1}, \tau_{i2}, \dots, \tau_{in} (n \neq i)$ を計測する。

片方向遅延の計測には各拠点の時刻が同期している事が求められる。遅延時間を計測するため、100 μ s オーダー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¹³⁾、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。拠点 N_i の計測ノードは隣接する拠点 N_j の計測ノード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} とパケット損失率 λ_{ij} を計測する。計測パケットは隣接拠点にのみ送られ、オーバーレイルーティングは行われず下位レイヤーのルーティングアルゴリズム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、計測されたこれらのネットワーク情報に対して各々の長さ w の配列を用意し、新たに得られたネットワーク情報をこの配列の先頭に格納する。新しい情報が書き込まれる前に、古い配列の要素を1つずつ後退させ、列長 w を越えたスコアを破棄する。すなわち配列には常に最新の w 個のネットワーク情報が保存されることになる (図3)。

パケット損失率と同時刻に計測されたトラフィック要求量とを組にして、パケット損失率の線形近似直線を求める。配列に格納されたパケット損失率が0である時点のトラフィック要求量の中で最大のトラフィック要求量、およびパケット損失率が0より大きいトラフィック要求量に対して最小自乗法による線形近似を行う。隣接するノード N_i とノード N_j のリンクに対して $a_{ij}\tau + b_{ij}\lambda + c_{ij} = 0$ となる係数 a_{ij}, b_{ij}, c_{ij} をこのリンクのトラフィック損失率を表す特徴量として、

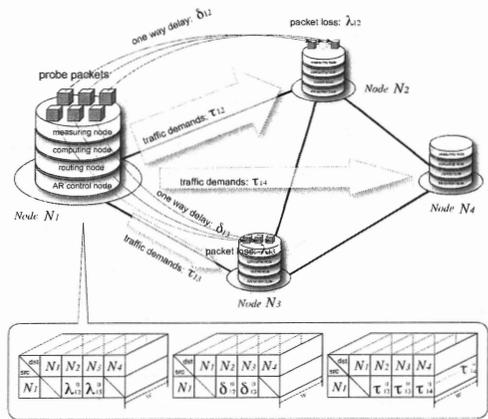


図3 ネットワーク情報の収集と格納
Fig. 3 Measuring Network Information and Accumulation

新しい計測情報が取得されるたびに算出する (図4)。

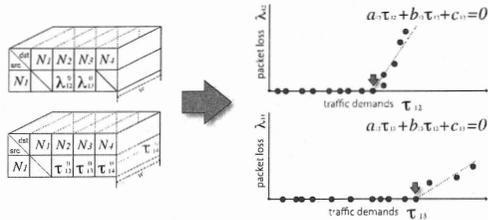


図4 トラフィック要求量-パケットロス特性と特徴係数
Fig. 4 Characteristics coefficients of packet loss conducted by traffic demands and packet loss

3.2 AR 管理ノード

障害の検出が報告されると、代表拠点の AR 管理ノードは管理全拠点にネットワーク情報の提供をアナウンスする。全拠点のネットワーク情報を得ると、AR 管理ノードは経路組み合わせの評価を行うための問題設定を行う。

図5に示す4ノードのリング状トポロジを例として、問題設定手法を示す。 N_1 から N_2 に到るループのない経路は1ホップで目的ノードに到着する経路 $\langle N_0, N_1 \rangle$ と3ホップで到着する経路 $\langle N_0, N_2, N_3, N_1 \rangle$ の2つが存在する。このトポロジにおいて上り経路と下り経路は同じ経路を通するものとする、任意の2ノード N_i と $N_j (j \neq i)$ の組み合わせ6つにおいて全て2つの経路が存在する。

ここで、全ての2ノード N_s, N_d の組み合わせに対して、ループのない経路を全て書き出し、収集された片方向遅延時間情報 δ を用いて各々の経路の総遅延時間を算出する。この総遅延時間をキー値とし、 N_s, N_d を結ぶ各経路を総遅延時間が短い順でソートする。図5の例では全ての2ノードの組み合わせが6つあり、各

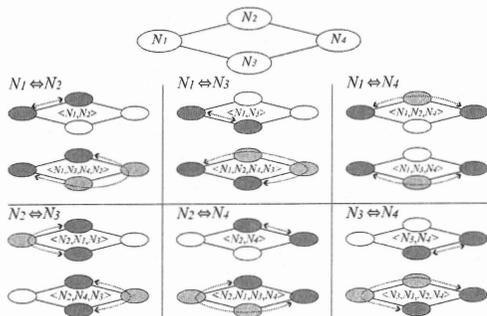


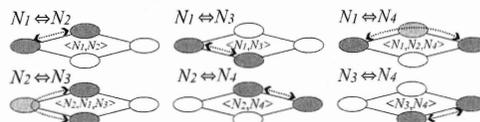
図 5 任意の 2 ノードに対する経路の組み合わせ
Fig. 5 Path Combination for any two nodes combination

2 ノードを結ぶ経路は 2 通りずつあるので、経路組み合わせは全部で $2^6 = 64$ 通りあり、各経路組み合わせに対して離散イベント型シミュレータを用いて評価計算を行う。この計算は各拠点の計算ノードに分担させて計算させる。各拠点には全ての 2 ノードの組み合わせに対して、ソートされたどの順位の経路を組み合わせで評価するかを指定したメッセージを送る。順位の指定は (n_1, n_2) のように範囲指定することができる。図 5 では 2 ノード組み合わせが 6 つあるので、6 つの要素を持つ配列 $\{[N_1, N_2], [N_1, N_3], [N_1, N_4], [N_2, N_3], [N_2, N_4], [N_3, N_4]\}$ (各要素は順位を指定する 2 ノードの組み合わせを表す) をメッセージとして送信する。例えば N_1 の計算ノードに $\{1, 1, (1, 2), (1, 2), (1, 2), (1, 2)\}$ というメッセージを送ると、 N_1 の計算ノードは指定された範囲の $1 \times 1 \times 2 \times 2 \times 2 \times 2 = 16$ 個の経路組み合わせを評価する。

3.3 計算ノード

経路組み合わせの計算要求を AR 管理ノードから受け取ると、計算ノードはその経路組み合わせから各拠点ごとに固有の経路制御表を作成する。受け取った計算要求には拠点を結ぶネットワークトポロジにおける全ての 2 ノードの組み合わせに対してループのない経路が 1 つ定められている。この 2 ノードの組み合わせは送信元 (src) と目的地 (dst) であり、この src と dst の組み合わせを持つパケットが先ほど計算要求に記された経路を通る。例えば図 5 において任意の 2 ノード N_1, N_2 を src と dst とする経路が $\langle N_1, N_3, N_4, N_2 \rangle$ のように決定されたならば、src: N_1 , dst: N_2 のパケットが N_3 に届けられた時の次送付先 (next hop) は N_4 になり、逆に src: N_2 , dst: N_1 のパケットが N_3 に到着した時の送付先ノードは N_1 となる。このようにして、任意の src: N_s , 任意の dst: N_d のパケットが任意のノード N_i に届いた時の next hop を一意に定める src-dst 型経路制御表を作ることができる。src: N_s , dst: N_d が辿る経路に N_i が含まれない場合、 N_i が持つ経路制御表の $(src, dst) = (N_s, N_d)$ および $(src, dst) = (N_d, N_s)$ に対応する欄は空欄となり、 N_i

に src: N_s , dst: N_d あるいは src: N_d , dst: N_s のパケットが到着した時は破棄される (図 6)。このようにして作られた経路制御表を用いると、任意の 2 拠点を結ぶ経路が一意に定まり、特別なラベリングをすることなくネットワーク上を流れるトラフィックをループのない経路に分散させることができる。



node N_1 routing table	node N_2 routing table	node N_3 routing table	node N_4 routing table	
src \ dst	N_1	N_2	N_3	N_4
N_1	N_2	N_3	N_4	reach
N_2	reach	N_1	reach	N_4
N_3	reach	N_2	N_1	N_4
N_4	reach	reach	reach	N_1

図 6 各ノードの経路制御表
Fig. 6 Routing Table on Each Node

この経路制御表に従ってパケットの配送が行われる離散イベント型シミュレーションを行い、各経路組み合わせの評価値を求める。

3.4 探索優先順位の変更と経路の間引き

本提案手法では、障害が発生している状態で計測された系全体のネットワーク情報およびその時点での経路組み合わせをもとに離散イベント型シミュレーションで評価を行い、この評価値よりもトラフィック損失率の低い経路組み合わせを探索する。伝送損失率の非常に小さな差で、極めて大きな伝送総遅延時間をもたらす経路を選択される事のないように、伝送損失率には許容幅を持たせ、その許容枠の中で最短の伝送総遅延時間を準最適値として選択する。2 拠点を結ぶ複数の経路は、計測された片方向遅延時間から導かれる経路の総遅延時間をキーとしてソートされている。これにより探索の初期段階でより伝送総遅延時間の短い経路から順に組み合わせられ探索が行われる。またソートされた各経路が、最短遅延経路を基準として n 倍以上 $n + 1$ 倍未満の遅延時間となる経路をクラス n に分類する。すなわちクラス 1 には最短遅延経路を含む最短遅延経路の伝送総遅延時間の 2 倍までとなる経路が含まれる。全ての 2 拠点の組み合わせに対する経路の組み合わせを決定する時、全ての経路の組み合わせを総当たりで探索するのではなく、このクラス番号の小さい順から探索候補に入れることで、長大で現実的でない経路を探索から除外する。

4. 離散イベント型シミュレータ

与えられた経路制御表を用いれば、計測されたトラフィック要求量の発生がその経路組み合わせで配送された時に、どの程度のパケット損失と総伝送遅延時間

をもたらしかを離散イベント型シミュレータで求めることが可能である。既存の離散イベント型シミュレータを経路評価のために用いる場合、評価実験のパラメータを読み込ませる際のオーバーヘッドが問題となり、膨大な回数の評価実験を現実的な運用の時間内に行わせる事が難しい。また本稿で提案するような src と dst の組み合わせに対して経路が一意に定まるアルゴリズムで、映像伝送のような固定ビットレートのトラフィック要求が発生するという条件においては、パケット単位でのシミュレーションではなく、まとまったパケット群で粒度の荒いシミュレーションを行う事で精度を下げずに高速な評価を行う事が可能である。そこで ns-2 の設計を参考に¹⁴⁾、固定ビットレートのトラフィックを発生させた時のトラフィック損失率と総伝送遅延時間のみに特化した簡易高速版シミュレータを実装し、評価に用いる。

4.1 設 計

簡易高速型シミュレーションは単位時間 1 ss (simulation step) の時間粒度で進行する。各拠点に他の各拠点に対してトラフィック要求を 1 ss ごとに計測された量だけ発生させる。発生したトラフィック要求を与えられた経路制御表と、計測された片方向遅延に従って、次のイベント駆動時間に移動させる。このようにシミュレーションはトラフィック発生とトラフィックの移動の 2 つ処理が交互に行わる。経路組み合わせの評価に用いるのは、目的ノードに到達したトラフィックの総量 E_s と到達できずに途中で損失されたトラフィックの総量 E_λ である。これらの初期値を 0 にする。1 ss は 1 ms を模するものとする。本シミュレータでは、送信元ノードと目的ノードを同じくする 1 ss 内に発生したパケットの群れを離散化されたトラフィックの最小単位として扱い、この離散トラフィックがキューに投入され、ルーティングされ、目的ノードに到着するとキューから除外される。1 つの離散トラフィックは、送信元ノード *src*、目的ノード *dst*、その離散トラフィックが滞留しているノード (current node) *cn*、送付先ノード (next hop) *nh*、*cn* での残滞留時間 *R*、その離散トラフィックの大きさ *T*、そしてその離散トラフィックが発生してから総時間 *D* の 7 つの情報によって表現される。このシミュレーションは任意の経路組み合わせを評価するものであるから、あらかじめ各ノードごとの経路制御表は作られており、ノード N_i において送信元ノード *src* と目的ノード *dst* に対する送信先ノード *nh* は $NH(N_i, src, dst)$ により与えられるものとする。

4.2 トラフィック要求の発生

シミュレータのキューに全ての送信元ノード N_s と全ての目的ノード N_d の組み合わせに対する離散トラフィックを投入する。離散トラフィックは発生したばかりの状態なので、滞留しているノード *cn* は N_s であり、送付先ノード *nh* は $NH(N_s, N_s, N_d)$ となる。

ノード N_i からノード N_j に対して計測された片方向遅延時間を $\delta_{i,j}$ とすると、残滞留時間 *R* は $\delta_{N_s, nh}$ となる。同様にノード N_i において計測された他ノード N_j への 1 ms あたりのトラフィック量の平均値を $\tau_{i,j}$ とすると、離散トラフィック量の大きさ *T* は $\tau_{N_s, nh}$ として与えられる。離散トラフィックが発生してから総時間 *D* は 0 である。

4.3 トラフィック要求の移動

シミュレータのキューにある全ての離散トラフィックの残滞留時間 *R* から 1 を減じ、同様に離散トラフィックが発生してから総時間 *D* に 1 を加える (ただし直前のトラフィック発生ステージで発生したものを除く)。残滞留時間 *R* が 0 になった離散トラフィックは送付先 *nh* に指定されたノードに移動する。

この時、移動処理を行った全ての離散トラフィックについて、滞留しているノード *cn* と送付先ノード *nh* ごとの累計値を求める。この累計値は *cn* と *nh* により定まるリンクに単位時間あたり流れたトラフィック総量に等しい。計測ノードで算出された、リンクのパケット損失率に関する特徴量を用いて、そのリンクにおけるトラフィック総量に対応するトラフィック損失率 $\lambda_{cn, nh}$ が 0 より大きい時は、*cn* から *nh* へと移動処理が行われる全ての離散トラフィックに対して、離散トラフィックの大きさから $\lambda_{cn, nh}$ を減じ、減じられた分を損失トラフィック総量 E_λ に加える。

滞留しているノード *cn* を *nh* に書き換え、書き換えられた *cn* が目的 *dst* と一致する場合は後述する到着処理に回す。送付先 *nh* は書き換えられた *cn* を用いて $NH(cn, N_s, N_d)$ に書き換える。残滞留時間 *R* は、書き換えられた *nh* を用いて $\delta_{cn, nh}$ となる。到着処理に回された離散トラフィックは、その離散トラフィックの大きさ *T* と、離散トラフィックが発生してから総時間 *D* とを掛け合わせた値を到着トラフィック総量 E_s に加え、キューから削除する。

5. 実 験

5.1 予備評価実験

実装された簡易高速版シミュレータの精度を評価するために、3 種類のネットワークポロジを BRITE¹⁵⁾ を用いてノード数 8 の仮想ネットワークを生成した。各ネットワークの設定を図 7 に示す。グラフの生成モデルは Barabási-Albert と Waxman を利用した。Waxman モデルではノード配置方法によりグラフ構造が変化するため、Heavy Tail と Random の 2 つのノード配置方法を用いた。

帯域幅は全て 100 Mbps で一定とし、各ノードを結ぶリンクの片方向遅延時間は仮想ネットワークでの配置された距離を参考に、全片方向遅延時間の平均が 4 ss となるような整数値で与えた。各ノード間のトラフィック要求量は固定ビットレートで 1 Mbps から 10

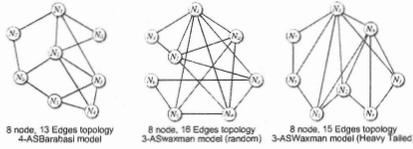


図 7 予備評価実験のためのネットワークのグラフ構造

Fig. 7 Illustration of the graph structure for preliminary evaluation

Mbps の範囲で一様に分布させ、トラフィック要求量分布を 100 パターン用意する。シミュレーション開始からトラフィック要求量分布に従って各ノードが他ノードにトラフィックを送信する評価実験を簡易高速版シミュレータと ns-2 のそれぞれで行い、100 ss でシミュレータを停止させて総トラフィック損失率と総遅延時間を比較した。ns-2 で得られた総トラフィック損失量を $E_{\delta ns-2}$ 、総伝送遅延時間を $E_{\lambda ns-2}$ とし、簡易高速版シミュレータで得られた総トラフィック損失量を $E_{\delta ths}$ 、総伝送遅延時間を $E_{\lambda ths}$ とする。

3つのネットワークトポロジーと100個のトラフィック要求量分布の組み合わせ、総計300パターンの実験において、 $\frac{E_{\delta ths}}{E_{\delta ns-2}} \times 100$ の値の平均値は99.2%、標準偏差は 2.05×10^{-2} 、 $\frac{E_{\lambda ths}}{E_{\lambda ns-2}} \times 100$ の値の平均値は98.8%、標準偏差は 6.22×10^{-3} であった。簡易高速版シミュレータは固定ビットレートでの評価計算について十分に高い精度のシミュレーションを行うことが出来る事が示された。

また、各トラフィック要求量分布に対する簡易高速版シミュレータの評価計算を全ての経路組み合わせに対して行う。100 ss で打ち切る評価計算について1秒間で 1.0×10^6 の経路組み合わせを評価することが出来るものとし、1秒間で得られた最適解と全探索で得られた最適解を比較する。この処理能力は Intel Xeon X5570 2.93 GHz を2基、メインメモリを16 GB 搭載したマシン1台で前述の評価計算を行った時の平均処理能力の10倍の値である。経路評価計算は仮想ネットワークの各ノードで用意されている計算ノード、合計10台に分散して計算されるものとし、10台で10倍の経路組み合わせを評価することが出来るものとする。計算を分散させる際のオーバーヘッドは無視できるものとする。

先ほど同様、総計300パターンの実験において、274パターン(91.3%)が全探索で得られた最適解と同じ解を1秒以内に発見する事ができた。他18パターン(6.0%)は1秒以内に最適解とトラフィック損失率が0.1%異なる範囲の解を発見しており、残り8パターン(2.7%)については10秒以内に最適解と伝送損失率が0.1%異なる範囲の解を発見している事がわかった。

5.2 評価実験

8ノード程度の小規模で迂回路が複数存在するトポ

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状況を作り出し、この問題に対して提案手法が実運用において現実的な時間内に迂回路を探索できることを ns-2 によるシミュレーション実験により示す。

図8に示すトポロジーのネットワークを対象としてネットワークシミュレータ ns-2 を用いた評価実験を行った。各拠点ノードを結ぶリンクは全二重、各ノード間リンクのバックボーン帯域幅は100 Mbps、転送遅延は図に示す通りとした。各ノードは Drop tail 方式でキューの管理を行うものとする。各ノードに CBR (UDP) のソースと UDP null を置く。代表ノードはノード4であり、隣接するノードとのリンクにおいてパケット損失が0.1%以上変化した時に経路組み合わせの再計算を行うものとする。経路組み替えアナウンスは1 sec ごとに行う。1つの経路評価計算は100 ss で打ち切り、その時点での E_{δ} と E_{λ} で評価を行う。

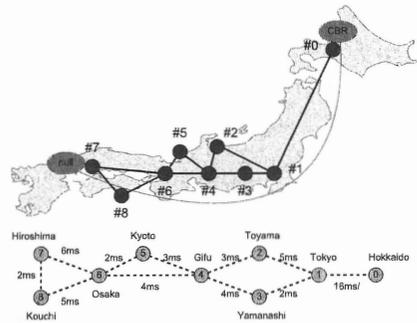


図 8 評価実験のネットワーク

Fig. 8 Network Construction in Evaluation Experiment

上記ネットワークにおいて以下のようなイベントを時系列に沿って発生させる。

- (1) 0 sec において全てのノードは全ての他ノードに対して 2^n kbps の CBR トラフィック要求を発生させる。 n は $8 \leq n \leq 12$ の範囲で一様分布で定める。
- (2) 50 sec から 5 sec ごとにノード0からノード7への CBR トラフィック要求を1 Mbps ずつ増加させ90 Mbps まで増大させる。
- (3) 500 sec においてノード0からノード7への CBR トラフィック要求を元に戻す。

トラフィック要求量の増大に対するトラフィック損失について、ノード3からノード4のリンクにおけるトラフィック総量とトラフィック損失率の変化を図9に示す。パケット損失率はトラフィック要求量を増加させた後、4 sec 後に計測された値である。TE を有効にした場合は、87 Mbps でパケット損失率が0.1%を越えて経路組み替えが発生するが、4 sec 以内に準最適解が適用されて損失を抑制することに成功している事が分かる。経路組み替え後は98 Mbps まで損

失が発生せず、迂回経路へのトラフィック分散が適切に行われていることが分かる。

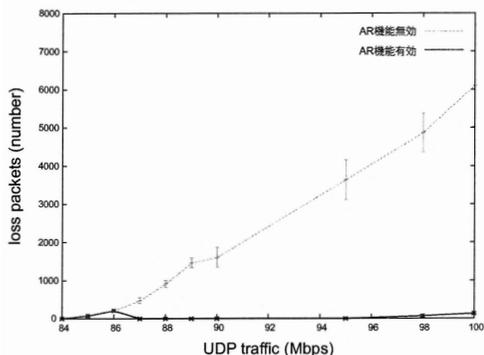


図9 区間 [3,4] におけるトラフィック要求量の増大に対するトラフィック損失量の変化

Fig.9 Change of Packet Loss: between node 3 and 4

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路の品質変化に対して適応する決定論的な経路制御手法を提案した。各拠点到に設置された計測ノードは独立して自律的にネットワーク情報を計測、取得、保持し、ネットワーク品質の劣化を判断すると代表ノードに経路組み合わせの再計算の依頼を行い、代表ノードは経路組み合わせの評価を各拠点の計算ノードで分散計算させることにより、障害に適応することが可能なTE手法を提案した。ネットワークシミュレータ ns-2 を用いた評価実験を行い仮想的なトラフィックのフロー計算を行うことで、提案手法がネットワークの変化に対して動的に経路を変更させる事を観測した。

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) Kandula, S., Katabi, D., Davie, B. and Charny, A.: Walking the tightrope: responsive yet stable traffic engineering, SIGCOMM Comput. Commun. Rev., Vol.35, No.4, pp.253-264 (2005).
- 6) 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).
- 7) Srikanth Kandula, Dina Katabi, Shantanu Sinha, Arthur Berger: Dynamic Load Balancing Without Packet Reordering, ACM SIGCOMM Computer Communication Review, Vol.38, No.2, pp.53-62 (2007)
- 8) 近堂 徹, 西村浩二, 相原玲二, 前田香織, 大塚玉記: 高品質動画伝送における FEC の性能評価, 情報処理学会論文誌 Vol.45, No.1, pp.84-92 (2004).
- 9) 滝根哲哉, 伊藤大雄, 西尾章治郎: 岩波講座 インターネット (5) ネットワーク設計理論, 岩波書店 (2001).
- 10) Applegate, D. and Cohen, E.: Making intradomain routing robust to changing and uncertain traffic demands: understanding fundamental tradeoffs, SIGCOMM '03: Proc. 2003 conference on Application, technologies, architectures, and protocols for computer communications, New York, NY, USA, pp.313-324, ACM Press (2003).
- 11) Vutukury, S. and Garcia-Luna-Aceves, J.J.: MDVA: A Distance-Vector Multipath Routing Protocol, INFOCOM, pp.557-564 (2001).
- 12) 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, 2007(72), 103-108 (2007).
- 13) 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).
- 14) Teerawat Issariyakui, Ekram Hossain: Introduction to Network Simulator NS2, Springer, ISBN-13: 978-1441944122 (2009)
- 15) Medina, A., Lakhina, A., Matta, L., and Byers, J.: BRITE: An Approach to Universal Topology Generation, International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems- MAS-COTS '01 (2001).