LF-005

遅延時間情報に基づく適応的経路制御の収束性能評価

Evaluation for An Adaptive Network Routing by Using Latency Information

柏崎 礼生 † Hiroki Kashiwazaki
高井 昌彰 ‡ Yoshiaki Takai

1. はじめに

爆発的なインターネットの普及とトラフィックの増大に伴い，局所的な輻輳や障害によるパケットの伝送遅延がネットワーク上で頻発している．一方，現状で用いられている経路制御アルゴリズムはこのようなネットワークのトラフィック状況変化に対して適応的に対処することは困難である．例えば OSPF では，リンクのコスト値を最小とする最短経路を一意に選択するため，輻輳を発生をさせやすいという問題がある．また，MPLS (Multi-protocol Label Switching) 技術を用いて TE(Traffic Engineering) を実現する手法も提案されているが [1]，予め迂回経路を明示的に設定する必要があるため，ネットワークが大規模化するにつれて運用も複雑化する事が考えられる．

本稿では，適応的な経路制御の実現のため，各ノードが自律分散型エージェントとして働く，始点制御でない経路制御アルゴリズム REI(Routing for Environmental Intelligence) を提案する [2]．本研究のアプローチは，人工生命からのアプローチ [3] を参考にしているが，ACO(Ant Colony Optimization) を用いた AntNet のように静的条件での最短経路探索ではなく，動的に変化するトラフィック環境に適応して通信遅延時間を最小に近づける点に特徴がある．各経路制御ノードは，往来するパケットから遅延時間情報を抽出・蓄積し，独立に最善の次ノードを確率的に決定するボトムアップ型の経路制御方法である．ネットワークシミュレーションにより，既存手法との比較実験を行い，提案手法の有効性を示す．

2. 適応的分散経路制御

2.1 前提条件と基本理念

一般に，ある経路において輻輳や障害が生じた場合，そこを経由して来たパケットの伝送遅延時間は増大するため，この遅延情報を各経路制御ノードが観測することで，トラフィック的に「悪い」方向を判断することができる．その結果，悪い方向と判断された場合には，手前のノードで迂回することが可能である (図 1)．

そこで REI では，データパケットの付加的な情報として，そのパケットが経由してきた全ノードの識別子と隣接ノード間の遅延時間の列が与えられるものとする．この遅延時間は固定的に定められるものではなく，各パケットの通過に要した所要時間の測定に基づくものである．遅延時間測定の実装については本稿では言及しない．

また，トラフィックが過負荷に近い状況では，パケットの遅延時間はリンクの通信速度そのものよりも，過負荷によるノード自身のパケット処理能力に強く支配されるため，隣接ノードを結ぶ経路の通信遅延時間（ルーティングおよびフォワーディング処理を含む）については，上りと下りに差がないものと仮定する．これにより，目的ノードからの下りパケットの情報を経路制御に利用することが出来る．

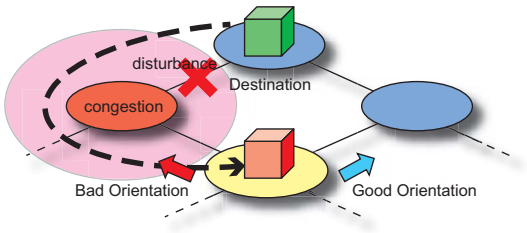

図 1　障害の回避

2.2 パケットの付加情報と経路制御表

各ノードは全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）を有している．ある目的ノード N_d に向けた次ノード候補 N_a には，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．このスコアはノードに到着したパケットに付随する遅延時間情報から得られる．ノード N_p にパケットが到着した場合，そのパケットの付加情報は結果的に $\langle N_1, d_1, N_2, d_2, \cdots, N_{p-1}, d_{p-1}, N_p \rangle$ となっている．ここで N_1 は始点ノードであり，d_m は $N_m \sim N_{m+1}$ 間の遅延時間である．この付加情報に含まれる始点ノードと全ての経由ノード $\{N_1, N_2, \cdots, N_{p-1}\}$ をそれぞれ目的ノードとする経路制御表において，次ノード候補 N_{p-1} に対するスコアの先頭 $S^1_{N_mN_{p-1}}, (1 \le m \le p-1)$ に，各経由ノードから現在パケットが到着したノード N_p までの総遅延時間 $\sum_{i=m}^{p-1} d_i$ を書き込む (図 2)．

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．

2.3 次ノードの選択規則

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．次ノードとして全ての隣接ノードが候補に挙げられるが，既に経由したノードおよび目的ノードへの到達が不可能なノードは除外する．候補に残った隣接ノードを $N_{a_i}, (1 \le i \le r,$ 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

† 北海道大学大学院工学研究科
‡ 北海道大学情報基盤センター

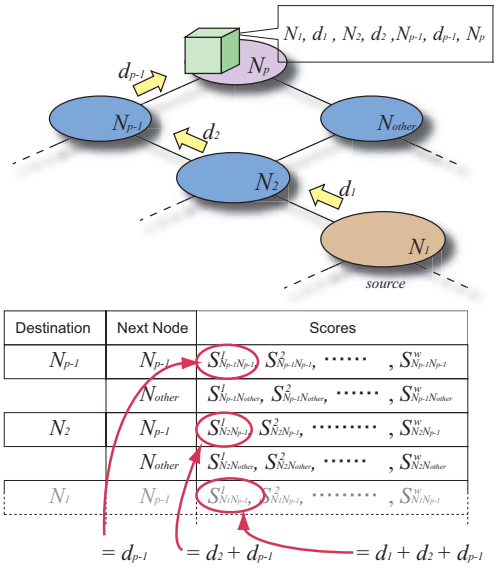

図 2　スコア付き経路制御表

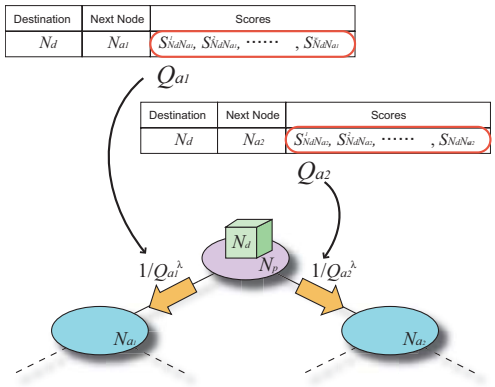

図 3　送付先の確率的選択

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．

次ノードは，全ての候補の代表値 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 (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．

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

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．以下本稿では，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 REIsx と呼ぶ．

3.　比較対象と実験条件

3.1　比較対象

REIsx の特性を実験的に評価するために，比較対象として OSPF を考える．OSPF は広く用いられた動的経路制御手法だが適応性に優れたものとは言い難いため，OSPF をベースに簡単な機能拡張を行う．本稿ではこのような拡張された OSPF を OSPF+, OSPF++ と呼ぶ．

OSPF では，Hello パケットを用いた隣接ノードの生存確認が行われるため，このパケットを頻繁に発生させることによって，適応的な経路制御が可能になるものと考えられる．そこで OSPF+ では，全ての経路制御ノードが 100ss (ss=simulation steps) ごとに Hello パケットを送出し，隣接ノードの生存確認を行うものとする．この際，OSPF の仕様に従い Hello 間隔の 4 倍の時間応答がない場合に隣接ノードがダウンしていると見なし，Link State Update パケットを送出し，経路制御表を更新する．

また OSPF++ では，全ての経路制御ノードが Hello パケット生成間隔 100ss で隣接ノードの生存確認を行うとともに，返信された応答パケットから隣接ノード間の遅延時間を測定するものとする．この際，遅延時間がデフォルトの 10 倍以上となった場合，その値に応じて当該リンクのコストを変更すると共に，Link State Update パケットを送出し，経路制御表を更新する．

3.2　ネットワークモデル

シミュレーション実験のネットワークモデルは 64 個の経路制御ノードから構成され，各隣接ノード間の遅延時間を示すコスト値は全リンクの平均値が 4 となるように，$\{3, 4, 5\}$ の何れかの値を取る．トポロジとして，ランダムに生成されたメッシュ状ネットワークとスケールフリーネットワーク [4] の 2 種類を考える．スケールフリーネットワークの生成には BA モデル [4] を用いて作成した．

また，各ノードには 1 シミュレーションステップ当りに処理可能なパケットの限界量を設ける．この処理限界を越えたパケットがノードの待ち行列に格納された場合，超過分は次ステップに順延され，滞留時間が増加する．本稿では全ての評価実験において処理限界量を 100 packets/ss とした．シミュレーションの 1 ステップを実時間の 1ms と見なし，平均パケットサイズを 400bytes とすると，100 packets/ss のパケット処理はスイッチング能力 100kpps, スループット 320Mbps に相当する．

4.　シミュレーション結果

4.1　一様なトラフィック状態

各ノードはステップごとに一定数のパケットをランダムな送付先に向けて発生させ，一様なトラフィック分布を形成する．この際，単位時間の発生パケット数 (packets/ss) を変化させ，ある 2 ノード間の平均総遅延時間を観測する．ここではメッシュ状ネットワークにおいて最小到達ホップ数が最も大きい 2 ノード間に注目する．これらのノードは互いに 4 packets/ss でパケッ

トを送信し合う．その他のノードは，背景トラフィックとして，ランダムな送付先に対して 4 packets/ss または 8 packets/ss でパケットを発生する．図 4 に REIsx ($\lambda = 4$) および OSPF, OSPF+, OSPF++ それぞれにおける平均総遅延時間の推移を示す．

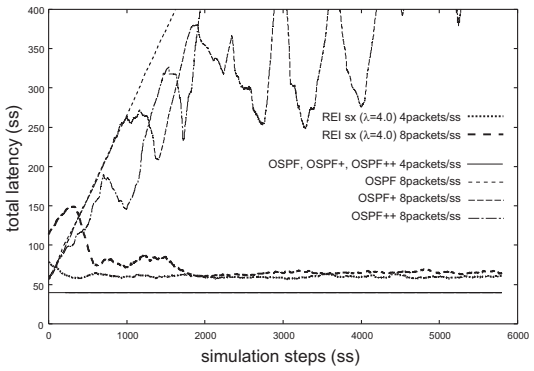

図 4　一様なトラフィック状態における平均総遅延時間の推移

背景トラフィックが 4 packets/ss という状況下では，どのノードもパケット処理限界に達することがない．そのため OSPF を用いた経路制御では，発見された最短経路に沿って遅延時間の安定したパケット配送が行われている．REIsx では，最短遅延経路以外の経路も確率的に選択されるため，結果として平均総遅延時間は OSPF が導く最適値よりも 30% 多くなっている．一方，背景トラフィックが 8 packets/ss となり，多くのパケットがネットワークに流通しはじめると，OSPF では一つの経路にパケットが集中し待ち行列の滞留を生じるため，平均総遅延時間は単調に増加する．OSPF+ や OSPF++ においては経路変更が適宜行われるが，その度に新たなパケット集中を生じるため，遅延時間の増加は避けられない．

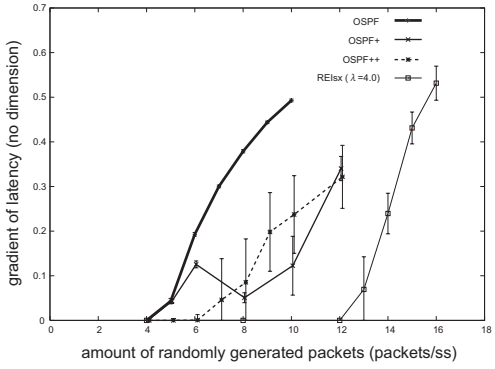

図 5　スケールフリーネットワークにおける平均総遅延時間の増加率

これに対し REIsx では遅延時間の大きな経路を避けて確率的なパケットの分配が実現されるため，平均総遅延時間が発散する現象は見られない．2000ss 経過後の平均総遅延時間は背景トラフィックが 4 packets/ss の場合と同程度までに収束する．

図 5 は REIsx における背景トラフィック量（単位時間あたりの発生パケット数）と平均総遅延時間の増加率の関係を示したものである．ここで，増加率とは，シミュレーションステップの区間 [0, 6000] における平均総遅延時間を最小二乗法で線形近似した時の回帰係数を意味する．ただし平均総遅延時間が収束する場合は平均増加率を 0 とする．OSPF では 6 packets/ss の背景トラフィックで平均総遅延時間の増加が始まるのに対し，REIsx では，$\lambda = 4$ において 9 packets/ss まで許容することが可能である．すなわち，ネットワーク全体で許容可能な総トラフィックの 50% 増加を見込むことが出来る．

4.2　ネットワーク資源の有効利用度

REIsx の経路制御によりネットワーク資源が有効に利用されることを，各ノードに滞留するパケット数の推移から考察する．図 6 は，メッシュ状ネットワークで集中的なトラフィック状態を発生させた状況において，REIsx を用いて経路制御を行った際に各ノードが保有するパケット数の時間変化を示したものである．各ノードが保持するパケット数を 100ss ごとに集計し，1ss あたりの平均滞留パケット数を 25 packets 区切りのヒストグラムで表わしている．各ノードはランダムな宛先に対して 5 packets/ss でパケットを発生させ，最小到達ホップ数が最も大きい 2 ノードが交信するパケット数は 32 packets/ss である．この条件において，OSPF，OSPF+, OSPF++ では平均総遅延時間が発散するが，REIsx の平均総遅延時間は一定値に収束する．

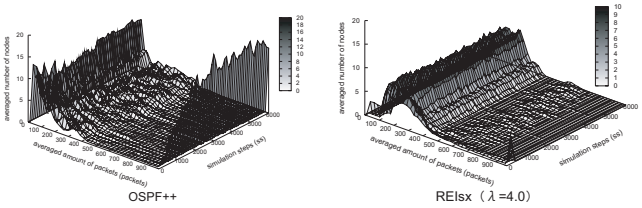

図 6　OSPF++ と REIsx における滞留パケット数の分布変化

リンクの平均コストは 4 であり，ノードのパケット処理限界量は 100 packets/ss であるため，各ノードの滞留パケットが 400packets 以下であれば理論上遅延時間は発散しない．REIsx では，立ち上がりの 400ss までの領域を除いて，1000packets 以上のパケットを保持するノードは存在していない．極端に少ないパケット数のノードも少なく，滞留パケット数が 150〜250packets のノードが全体の 87% を占めている．このように REIsx では遅延時間に基づいて確率的に送付先を決定する事により，一部のノードに過大な負荷を与える事なく，処理能力に余裕のあるノードにパケットを分散させることが可能であり，このことがネットワークのトラフィックの高い許容量の要因である．

4.3 障害発生の状態

トラフィックの適度な集中状態において，ある 2 ノードを結ぶ代表的な経路のコスト値を急速に増加させ，一定時間後に初期状態に回復させる．この際，当該 2 ノード間の平均総遅延時間を観測する．実験に用いるメッシュ状ネットワークの最小到達ホップ数が最も大きい 2 ノードを結ぶ経路は必ずリンク A, B, そして C のいずれかを通り，初期状態では最短遅延経路はリンク A を含む経路となっている．ここではリンク A のコスト値を，2000ss の時点で初期値 4 から 300 に増加させ，4000ss の時点で初期値に戻した．背景トラフィックは 4 packets/ss，当該 2 ノード間の交信パケット数は 16 packets/ss である．図 7 は平均総遅延時間の変化を示している．

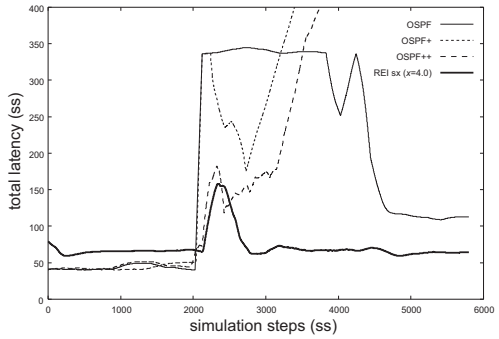

図 7　障害発生時の平均総遅延時間の推移

OSPF+ や OSPF++ において，リンク A のコスト値が増大すると，リンク B またはリンク C を通る経路にパケットが集中する．この集中は背景トラフィックについても同様である．そのためこれらのリンクを通るパケット数は障害発生以前にリンク A を通ったパケットの分だけ上乗せされ，結果として輻輳が生じる．

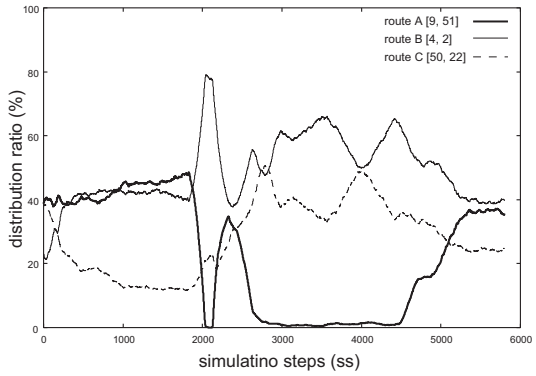

図 8　REIsx における障害時の経路配分確率の推移

これに対し，REIsx では障害による変化に適応し，リンク B とリンク C の双方を経由する経路を選択するため，トラフィックの過度な集中は生じない．REIsx におけるリンク A, B, および C を経由する配分確率の変化を図 8 に示す．障害発生後，リンク A を通る経路の選択確率は 1.0% 以下に低減されている．

REIsx では障害発生直後も，遅延時間の増大した経路を選択し続けてしまうため，一度は遅延の増大が発生するが，その遅延時間増大のピークは平均して 30ss に抑えられる．このピークは次ノード選択規則における各候補ノード代表値の計算頻度によって変化し，64ss ごとに計算を行い代表値を更新した場合に最も低いピークで収束する事が分かっている．障害回復直後に候補ノード代表値情報の変化に伴い，遅延時間の増大が発生してしまうが，これは 15ss 以下に抑えられる．これらの振動は確率分散規則を用いる事により必然的に発生してしまう現象であるが，十分に小さい値に抑えられている事が示されている (図 9)．

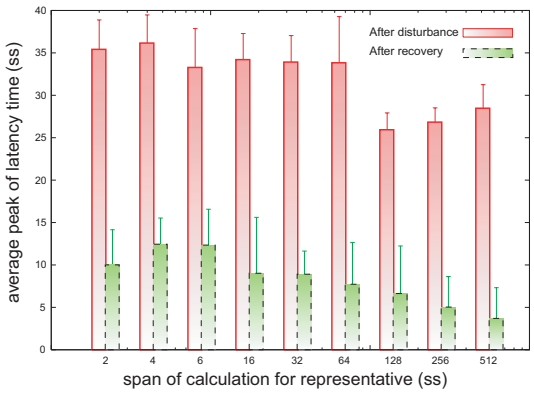

図 9　REIsx における障害発生・障害回復後の平均揺り返し

5. まとめ

トラフィック環境の動的変化に適応する確率的な経路制御手法 REI を提案した．メッシュ状ネットワークとスケールフリーネットワークにおいて，シミュレーション実験を行い，既存の OSPF や機能拡張された OSPF+/OSPF++ よりも優れた適応性と負荷分散性を有することが示された．コネクションを考慮した経路選択と経路制御表の集約化，およびパラメータの自動チューニングは今後の課題である．

参考文献

[1] 山田仁, 高島研也, 仲道耕二, 宗宮利夫, 中後明 : トラフィックエンジニアリングシステムの実機評価, 信号技報, NS2001-8, pp.45-50, April (2001).

[2] H. Kashiwazaki, Y. Takai : REI: An Autonomous Distributed Routing Algorithm, Proc. of IASTED PDSC2004, pp.76-81 (2004).

[3] M. Dorigo, T. Stutzle : Ant Colony Optimization, THE MIT Press (2004).

[4] A.L. Barabási, R. Albert, Emergence of Scaling in Random Networks, Science, 286 (1999) 509.