ストレージ連携情報ネットワークにおける経路制御方式

水 谷 昌 彦† 池 田 博 樹††

大容量コンテンツの共有、配信を効率的に行うためのネットワークアーキテクチャ、および経路制御方法を提案する。ネットワークを流れる情報は、動画・電話などリアルタイム性の要求されるコンテンツと、情報共有を主目的とする非リアルタイムでの利用を想定したコンテンツに分類され、ネットワーク上での情報配信を効率化するためには、これら情報の特性を利用し、ネットワークリソースを有効に活用することが必要である。本稿では、非リアルタイム系データの配信について、ネットワーク内に設置されたストレージの記憶領域情報に基づく経路制御によって、トラフィック集中を緩和し、輻輳の発生を抑制する方法を提案する。これにより、多様かつ大容量なコンテンツが同時に利用されるネットワークでの情報配信効率を向上させることを目的とする。簡単なモデルを用いた性能評価を行った結果、本方式では経路上でのデータパケット廃棄率が高くなる場合についてトラフィック処理効率が改善されることを確認した。

Forwarding Control in Storage Embedded Network

Masahiko Mizutani† and Hiroki Ikeda††

This paper presents network architecture and a method for routing control, which realizes effective handling of large contents. Information transferred in the network can be classified into two categories; one is the data for realtime use such as movie, and the other is the data for storing and/or sharing such as archives and backups. In order to use network resources effectively, it is necessary to take the characteristics of contents into account. In this paper we propose a method of packet-forwarding control for non-realtime large contents. The purpose of this proposal is to minimize occurrence of congestion by distributing traffic load. Evaluation with a simple model shows that our proposal improves the throughtput in congested networks.

1. はじめに

大容量データの転送効率は、ネットワーク技術の研究において、最も重要な研究課題である。近年ネットワークの高速化が進み、物理回線については 10GB Ethernet が製品として登場しており、また 100GB Ethernet をはじめとする回線高速化技術の研究が各研究機関において進められている。一方で ADSL に代表されるブロードバンド環境が一般家庭や企業網に普及し始め、それとほぼ平行して Web を利用した情報配信サービスにも画像や動画などのマルチメディアデータの利用が広がっている[1)]。ネットワークの高速化技術の進展と相まって、ネットワーク上で利用される情報量は回線速度の進展を上回るペースで年々急速に増加しており、配信の効率化は益々重要な課題となっている。

一方、サービス提供のためのネットワーク及びサーバの制御が近年急激に高度化し、個々のユーザニーズに対応するコンテンツ管理と配信品質の管理が重要視されている。例えば、パーソナライズされた高負荷価値サービスの提供を目的として、プレゼンス管理の技術が多く研究されている。ユーザ端末とプレゼンス管理サーバ、さらにはサービス提供サーバの間でプレゼンス情報の登録、更新、その他データ管理のための制御が必要であり、こうした制御のための機能をネットワーク上でサポートすることが、サービスの効率化には必要である。

ネットワーク上でのサービス提供において最も重要な要素は、ユーザのリクエストに対するレスポンス時間の短縮である。データ量の増加に伴い、情報の配信に要する時間が拡大するため、この時の配信遅延の改善が重要な課題となる。遅延は、従来の TCP による転送制御が、現在提供されるコンテンツ配信サービスに見られるような大容量のデータを送信するための機能を持っていないことに起因する。複数の拠点間で発生する、バースト性を持つ大容量トラフィックをネットワーク上で効率的に処理するためには、従来の技術に不足する機能を補うことが必要である。情報通信を効率化するアプローチとして、サーバの処理性能を向上させる方法と、ネットワーク内の中継装置の処理能力を高める方法がとられる。

サーバの処理性能を高める方法には、コンテンツ配信網に利用されるキャッシュサーバの導入が挙げられる。キャッシュサーバからユーザへデータを配信することでサーバの負荷を低減し、またユーザと情報との距離を短縮することで配信遅延を低減することができる。さらに複数のサーバを用意しネットワーク内部でのパケット転送機能とを連携することにより、ネットワークユーザリクエストの特定サー

† (株) 日立製作所 中央研究所
Hitachi, Ltd., Central Research Laboratory
†† (株) 日立製作所 研究開発本部
Hitachi, Ltd., Research and Development Group

バへの集中を回避する方法も用いられる。キャッシュやサーバに分散配置されたデータを連携して管理することにより、ユーザへのデータ配信を高速化する。

遠距離セッションでは経路上のさまざまな要因により制御信号の遅延や、データパケットの喪失及び再送が生じる。これらは転送される情報量に応じて増加するため、転送される情報量の拡大は、ネットワーク上での大きな転送遅延となって現れる。従って、これを防ぐには、データ転送距離を短縮することが有効である。キャッシュサーバを利用することにより、TCP の転送制御に起因する、長距離を転送する際に生じる大きな遅延を低減することが可能である。大容量バッファを持つルータを用い、送達確認区間を短縮することによりデータの転送効率が向上する[2]。

これらはコンテンツ提供者からユーザへの一方向のみの配信形態で有効な方法といえる。しかしながら、頻繁にデータがネットワーク内を移動する場合、キャッシュやサーバによってデータ配信を効率化することは難しい。今後、デジタル化されたコンテンツの利用が広がり、ネットワーク上で利用されるデータも、プロバイダによる提供だけではなく、個人データの逐次利用や遠隔地からの重要データ管理、及び特定コミュニティ内でのデータ共有が進むことが予想される。したがって、データの移動距離を短縮しようとするアプローチでは、今後のネットワークサービスに対応することが困難である。

ネットワーク内でのパケット処理効率を向上させる方法として、特定経路にトラフィックが集中することを避け、ネットワーク上での負荷分散を行うことで個々の中継装置の負荷を低減し、パケット処理速度を向上させる方法がとられる。ルータの負荷に応じて経路を切替えるトラフィックエンジニアリング (TE)[3] やポリシーベースルーティング (PBR)[4] などの方法が提案されている。また装置の負荷を低減する別の方法として、ルータやサーバにおける処理性能向上を目的とするパケット抑制処理がある。RED (Random Early Detection)[5], ECN (Explicit Congestioin Notification)[6] が提案され、現在一般にルータで利用されている。これらの方式における基本的な考え方は、データパケットをルータ若しくはサーバの手前で積極的に廃棄することで、個々の装置に処理が集中する状況を回避しようとするものである。

将来予測されるような、恒常的に負荷が高く、バースト性の強いトラフィックが頻発する状況を想定すると、TE は IP ルーティングの経路制御に基づいて動作するため、変動の激しいトラフィック分布に対し十分に装置負荷を抑制することが難しくなるおそれがある。またパケットを積極的に廃棄するアプローチは、ネットワーク上のトラフィック量をさらに増大させる結果につながることから、次世代のネットワークにおいては適切な制御方法とは考えにくい。従来のネットワークにおいては、データ転送量が小さかったため、データパケットの廃棄及びそれに伴う再送時間はあまり問題にならず、それに伴うトラフィック量の増加が重要視されることがなかった。しかし情報の大容量化が進み、さらなる転送の高速性、確実性が求められる将来においては、トラフィックの増加は最も重要な課題の一つになると考えられる。

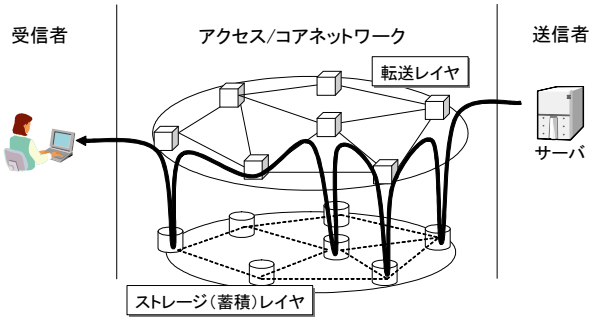

図 1 ストレージ連携情報ネットワーク

本稿では、拡大するトラフィックをネットワーク上で高効率に処理するアーキテクチャを検討する。ネットワーク装置の持つデータパケット転送機能とストレージ装置の持つ情報蓄積機能とを連携する転送制御方法を提案し、その効果を検証した。以下、想定するネットワーク構成を「ストレージ連携情報ネットワーク」と称し、ネットワークモデルと基本的な動作を説明する。

2. ストレージ連携情報ネットワーク

2.1 概　　要

トラフィックの集中がもたらす輻輳とそれに伴う転送遅延を回避するには、トラフィックを調整し、ネットワーク上で同時に転送するデータ量を抑えることが有効である。そこで、ストレージ機能をデータ転送に利用するストレージ連携情報ネットワークを提案する。コンテンツ配信サービスで多く見られるマルチメディアデータに関しては、一定の帯域を要求するリアルタイム系データとは異なり、データ送信時のバースト性が非常に高く、データの即時性は低いという特徴がある。これに対し、回線容量を超過するバーストトラフィックを一時的にネットワークの外におき、転送効率の低下を防ぐことを目的として、ネットワーク上で輻輳が生じた場合に、ノードの転送性能を超える量のパケットを廃棄せずに、一時的にノード内に保管する。ストレージ連携情報ネットワークの概要を図 1 に、またノードの機能構成を図 2 に示す。

2.2 ストレージ機能の活用

ストレージ機能の導入により、ファイルシステムやストレージ領域のブロックアドレス情報をネットワーク内で利用することが可能になる。これらの識別情報は、ネットワーク経路制御に活用することができる。また、パケットを一時蓄積する際に、ストレージ領域の利用状況を管理することでネットワークの混み具合を知ることができ、従来のルータにおける統計情報と連携して柔軟な経路設定が可能となる。具体的には、ストレージ領域の利用状況に基づいて転送経路を決定することで、効率的な転送制御の実現を図る。そのため、従来の IP ネットワークにおける経路制御とは独立に、ストレージ蓄積機能と連携した転送経路テーブルを構築する必要がある。効率的にネットワークトラフィッ

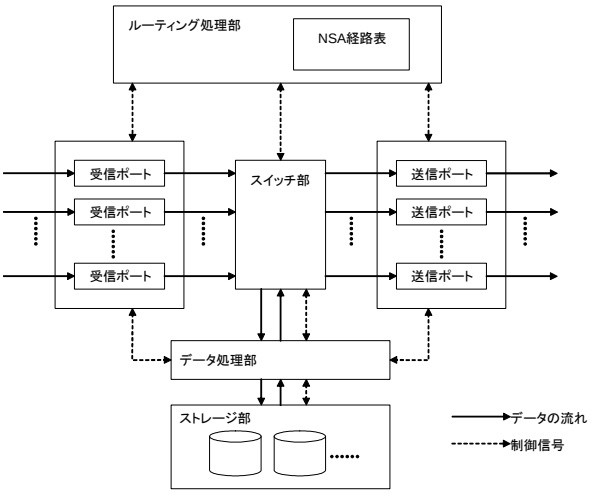

図 2 ストレージ連携ネットワークノードの機能構成図

クを低減させるには、帯域とストレージ領域双方の利用状況を考慮しなければならない。以下、提案する転送制御について説明する。

3. ストレージ領域情報を用いた経路制御方法

3.1 ネットワークストレージアドレス (NSA) の導入

ストレージ領域を識別するためのアドレスとして、「ネットワークストレージアドレス (Network Storage Address: NSA)」を定義する。ストレージ連携情報ネットワーク内での経路決定は、この NSA に基づいて行う。従来のルーティングに用いられる宛先及び送信元の IP アドレスによるパケット振り分けでは、トラフィック集中時にネットワークの輻輳を回避するなど、状況の変化に柔軟に対応することが難しい。そこで、NSA によるアプリケーション層での経路指定を行う新たな制御機構を導入する。

NSA の導入により、ファイルの属性やアプリケーションにより異なる通信条件を考慮し、配信を効率化する柔軟な経路制御を行うことができる。アプリケーション層においてデータのパケット化を行い、このアプリケーション層でのノード識別情報として NSA による経路情報を作成し、パケットの送信先を決定する。図 3 に、アプリケーション層でのアドレス付けを用いた経路制御の概念を示す。転送能力の高い回線もしくは、経路上のホップ数が少ない IP 網での最適経路に対してトラフィックが集中する場合に、それ以外の経路へトラフィックを振り分けることにより、ネットワーク利用の効率化を図る。

NSA を用いた経路制御は、IP 網では通常利用されることの少ない経路上のノードに対しても、ストレージ領域情報に基づいてデータパケットを振り分け、転送処理を行うパケット数を抑制する。そのため、ネットワーク上でのトラフィック処理の偏りを低減し、ネットワーク内でのパケット処理効率が向上する。また、アプリケーション層でのスケジューリング機能と連携してコンテンツの種類に応じた送信方法を選択することが可能になる本機能を導入した場合の効果を以下に挙げる。

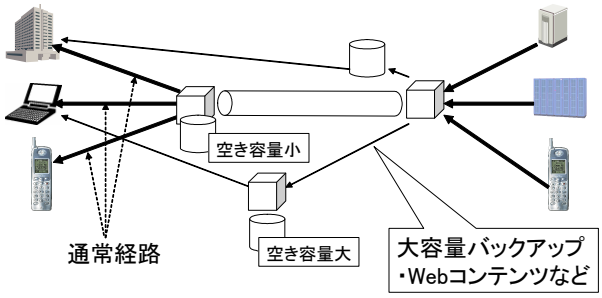

図 3 NSA を利用した経路制御の概念図

- 送信先として自ノードのストレージ領域を指定することによる自ノードの輻輳緩和
- 蓄積パケットのスケジューリング（上位層でのスケジューリング）によるパケット転送効率の向上

3.2 NSA を用いた経路制御方法

ノードに備えられたストレージ領域に対して NSA を設定し、ノード間でストレージ領域の利用状況（データ転送状況）を相互に通知することにより、NSA で記述されるネットワークトポロジに基づいた転送経路を決定し、自ノードから他のノードに至る経路表を作成する。各ノードに保持されるこの NSA で記述される経路表を、「ストレージルーティングテーブル (Storage Routing Table: SRT)」と呼ぶ。各ノードでは、SRT に基づいてパケットの振り分け先を決定し、該当する NSA を持つノードへデータパケットを転送する。

ノード間における経路情報の相互通知では、ストレージ領域情報として、ストレージ領域の分布によって構成されるネットワークトポロジと、各ノードに備えられたストレージ領域の利用状況を取得する。経路の選択は、従来方式で用いられてきた物理回線の状態、ストレージ領域の利用可能な領域情報、更には転送されるデータに要求される処理の優先度に基づいて行う。

図 4 に経路振り分けを行う方法を示す。データは、アプリケーション層で分割され、送信先と送信元の NSA 情報を含むヘッダを付加されて転送される。SRT に基づく経路制御は、IP ルーティングよりも優先される。トラフィック量が少なく、輻輳のない場合には従来の IP 経路と同様の経路を利用しての転送を行う。ノード内ストレージの利用率が高まることは、その経路を流れるトラフィックの増加を意味することから、ストレージ領域の利用状況に基づいて経路上の輻輳状態を判断し、経路上に輻輳が生じている場合は SRT に基づいて、より使用頻度の低い経路へパケットを振り分ける。

SRT を用いた経路選択手順を図 5 に示す。パケット単位でノードへ格納する場合、ストレージ領域の使用率がある閾値を超える場合に、第二候補の経路へデータを振り分ける。また、トラフィックをフローごとに識別することを考えた場合、必要なファイルサイズをストレージ領域に確保できない場合に、次候補の経路へフローを振り分ける。ストレージ領域の利用状況は随時変化し、また転送する情報量もまちまちである。そのためデータフローを受信する度

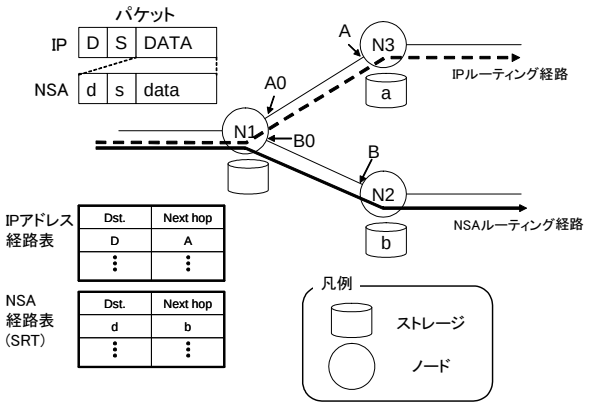

図 4 IP アドレスと NSA とを利用する経路決定方法

に必要な領域を確保できる経路だけを選択するよりも、実際にはアプリケーション層パケット単位でノードへの蓄積及び送達確認を行い、逐次データを送信するように実装することが実用的と考えられる。

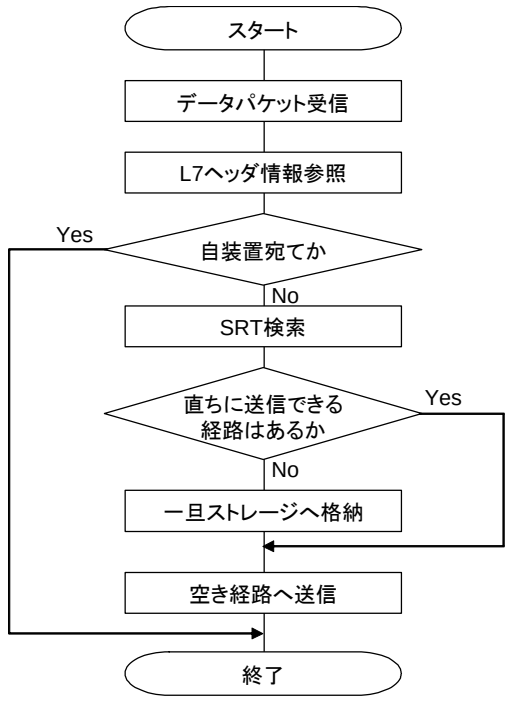

図 5 SRT による経路振り分け処理の手順

3.3 アプリケーションレイヤ経路制御の効果

従来の IP 網ではルーティングプロトコルによって回線コストに応じて経路が選択され、その経路に対してトラフィックが集中する傾向があった。本方式の導入により、従来のコスト値に加えて、ストレージの空き容量に基づく確実な経路設定が可能になる。これにより、転送時の遅延が許容でき、非リアルタイムで利用される情報に関して、ネットワーク上の回線利用効率を向上させ、高効率な配信を可能にする。本方式は特に、確実性がリアルタイム性よりも優先される場合を想定している。そのため転送時に空き領域を持つノードの経路を選択することが特徴となっており、ネットワーク全体のトラフィック処理効率を向上することができる。

提案方式では、利用可能な出力回線全てを利用するため、従来よりも送信バッファの実効的な容量が増加することになる。ノードのトラフィック集中率とパケット処理効率はネットワーク構成により変化するが、一つの経路のみを用いてデータを転送していた従来の方法と比較して、複数の経路に対して送信することでパケット廃棄数が抑制されることになる。個々のノードにおいて複数の送信経路を利用することで、平均的なパケット処理の効率が上がるため、ネットワーク全体のトラフィック処理の効率が向上する。

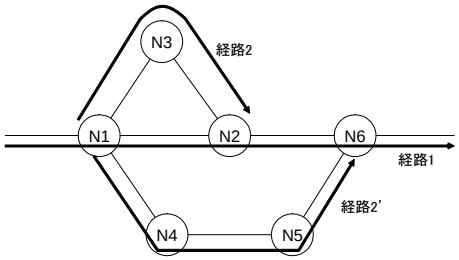

図 6 経路上で迂回が生じる場合のネットワーク構成図

以下、トラフィックが集中するネットワークにおいて、迂回経路を利用する場合のトラフィックの処理効率を考察する。提案方式の適用として、図 6 に示すネットワーク構成を考える。最適経路である経路 1 にトラフィックが集中する場合を仮定し、迂回経路 2 もしくは、経路 2′ により、ある区間をバイパスしてデータパケットを転送する状況を想定する。

一般的に TCP の転送効率は、対象となる TCP セッションでの Round Trip Time (RTT) と経路上のパケット廃棄率が増加するに伴って低下する[2)7)]。ノード間での送達確認に TCP を利用することを仮定すると、その定常状態におけるスループット T は、RTT R とパケット廃棄率 p を用いて、

$$T = \frac{1}{R}\sqrt{\frac{3}{2p}} \tag{1}$$

と表される[7)]。提案方式ではストレージ連携ノード間での送達確認を行う。複数のノードを含む経路でのスループットは、区間上のパケット廃棄率と各ノード間の RTT を一定と仮定すると、式 1 の $1/n$(但し、n はホップ数) となる。経路 1 を用いる場合の区間 b でのスループット、ホップ数、RTT、パケット廃棄率を、それぞれ T_1、n_1、R_1、p_1 とし、経路 2（迂回経路）の場合について、同様に T_2、n_2、R_2、p_2 とすると、双方の経路を利用する場合のスループットの比は、

$$\frac{T_1}{T_2} = \frac{n_2 R_2 \sqrt{p_2}}{n_1 R_1 \sqrt{p_1}} \tag{2}$$

となる。この比が 1 以下になる場合に、迂回経路の利用が効率的であることを示す。図 6 のように、主要な経路付近で迂回経路を確保する場合、n_2 は n_1 とほぼ同じホップ数

を持つと考えられる。ここで、迂回によって経路が 1 ホップ増加する場合 ($n_2 = n_1 + 1$) を考えると、式 2 から、迂回経路により転送効率が向上する条件として、

$$1 + \frac{1}{n_1} \leq \frac{R_1\sqrt{p_1}}{R_2\sqrt{p_2}}. \quad (3)$$

が得られる。経路の RTT は、トラフィックが集中するほど大きくなり、経路間の $RTT \times \sqrt{p}$ の差はより拡大する。$RTT \times \sqrt{p}$ はその区間へのトラフィック集中度を表すパラメータと考えることができる。輻輳の生じている区間 n_1 が大きい程、式 3 で示される経路間のトラフィック集中度の差が小さくなり、迂回経路を用いる効果が期待できることが分かる。従って、特に主要な転送経路周辺の迂回経路を用い、ホップ数がほぼ変化しない場合に、より大きな効果が得られる。

4. 性 能 評 価

4.1 評価モデル

提案方式の効果を検証するため、基本的な経路モデルを仮定したシミュレーションを行った。シミュレーションには、VINT (Virtual InterNetwork Testbed) プロジェクト[8] の ns (Network Simulator) ver. 2 (2.1b9a)[9] を用いた。図 7 にシミュレーションに際して想定するネットワーク構成を示す。ここでは、簡単のためネットワークを 3 台のノードで構成した。通常の IP 経路制御では、図中の区間 b を通過するパケットは、経路 1 を利用することになる。この経路に対するトラフィックが集中する場合の転送効率を調べるため、迂回経路 2 を用意し、区間 b の転送効率の変化を検証する。

ネットワーク構成は、全てのノード間リンクの転送速度を 10[Mbit/s]、遅延時間を 2[ms] とした。各ノードは、アプリケーション層でのデータ転送制御と、TCP によるノード間の送達確認を行う。各ノードのキュー長は 25 パケットとした。ここで、ネットワーク全体での転送効率を検証するため、測定するデータ量は 1MB とし、経路 1 をとる場合と、経路 2 をとる場合の区間 b でのデータ転送に必要な時間を測定した。測定に当たっては、パケット廃棄率の転送効率への影響を調べるため、送信側から受信側へ TCP のストリームを流し、バックグラウンドのトラフィックとして利用した。また、経路 1 を利用する TCP フローの本数を、0–2 本の範囲で変化させ、測定を行った。基礎実験として 10[Mbit/s] の帯域と 1MB のデータを仮定したが、転送効率を 1~10 [Gbit/s]、データサイズを 1~5GB(DVD 相当) とした場合も、トラフィック集中度の上昇に伴い、同様の振る舞いを示すことは明らかである。

4.2 結　　果

シミュレーションの結果を図 8 に示す。パケット転送効率を示すため、経路 1 と経路 2 について、それぞれを利用した場合のデータ転送に要する時間を比較した。グラフは、従来の転送方式である経路 1(実線) のみによる転送に比べ、提案方式の経路 2(破線) を平行して利用する方法によって転送効率が向上することを示している。パケット廃棄率の増加に伴って、従来方式では大幅に転送効率が落ちるのに対し、迂回経路を利用すると効率の低下が抑えられていることが分かる。

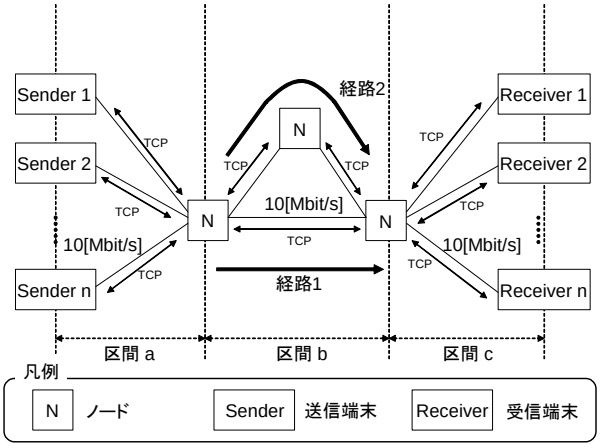

図 7　シミュレーションで想定するネットワーク構成

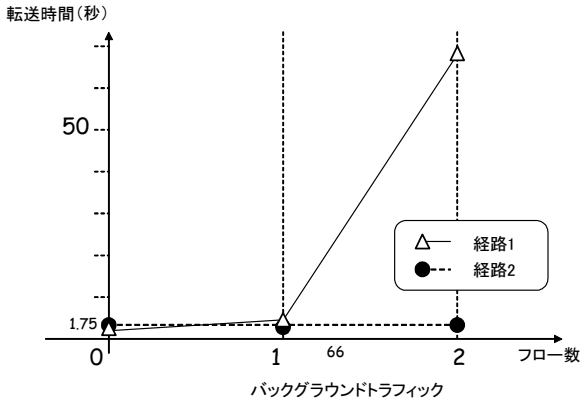

図 8　転送時間の比較

5. 考　　察

簡単なモデルの計算とシミュレーションの結果から、主要な転送経路の近辺に迂回経路を用意することにより、転送効率の向上が見込まれることが確認できた。転送効率の向上は、廃棄されるパケット数の増加と共に顕著になり、大容量コンテンツが頻繁に行き交うことが予想される次世代ネットワークにおいては、有効な手段であると考えられる。

ストレージ機能を利用する経路制御においては、経路上で廃棄されるパケット数は大幅に低減される。ノード間の送達確認と組み合わせることにより、大容量データの転送時など経路上での輻輳が多発する状況での転送効率は大幅に改善する[2]。シミュレーションでも示されているように、送達確認が必要となるデータフローの本数が増えると、その区間内での時間当たり処理可能なパケット数は大きく減少する。これは、ストレージ装置をノード内バッファとして導入するだけでは十分な効果が得られない場合があることを示唆する結果である。再送パケット数を抑制するには、

輻輳状態にあるノード数を減らすことが必要であることが分かる。

本稿で検討した経路制御方式は、ストレージの蓄積機能を利用し、IP ネットワークで偏りがちなネットワーク内のトラフィック分布を、ストレージ領域情報を指標として平均化することで、ノードの輻輳状態を緩和しようとするものである。ノードにおいて主に輻輳の生じる箇所は、出力キューである。従って、輻輳が生じた場合に、送信側で利用できる他のキューを転送に活用することで、各ノード内で時間当たり処理するパケット数は増加し、ネットワーク上で平均化した場合に、リソースを有効に活用することが可能である。

但し、一概にノードの出力側のキューにパケットを振り分けても、有効な経路にパケットを送らない限り転送時間を短縮することにはつながらない。有効な経路数はネットワークトポロジにより変化し、またデータパケットの宛先に近づくにつれて減少する。本稿では、簡単なネットワークモデルに基づくシミュレーションを行ったが、今後、具体的に各種ネットワークを仮定した構成でのシミュレーションにより検証を行う必要がある。

6. ま と め

増大するトラフィックを効率的に転送するための転送制御方式として、ネットワーク内に分散配置されたストレージ領域を識別するアドレスの導入、及び上位層での経路振り分け方式を提案した。本方式は、従来の TCP による送達確認機能を活かすと同時に、レイヤ間の連携による、IP 層のみでの経路処理では不可能な柔軟な経路設定を可能にする。

本方式を、基本的なネットワークモデルに対して適用した結果、ネットワーク上を複数の大容量トラフィックが移動する場合に、従来のネットワーク制御方式と比較して、トラフィックが集中する場合にデータ転送の効率が、改善されることを確認した。本方式は、ストレージ連携情報ネットワーク（図 1）において経路上のトラフィックを転送レイヤから退避すること、及び経路振り分け時の条件を新たに加える際に、ネットワーク利用状況を反映するストレージ領域情報を用いたことを特徴としている。アプリケーション層でネットワークを構築することで、従来技術との連携が可能であり、高度な付加価値を提供するサービスへの適用も考えられる。ストレージ領域情報を用いた経路制御プロトコル及びノード機能の詳細に関し、今後検討を進める予定である。

参 考 文 献

1) 平成 15 年度 情報通信白書 (www.johotsusintokei.soumu.go.jp/whitepaper/ja/h15/index.html)
2) 水谷昌彦, 池田博樹, 宮城盛仁: ”ストレージネットワークにおける情報ルーティング方式”, 電子情報通信学会研究報告, pp.1-6 (2002).
3) Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., Swallow, G., ”RSVP-TE: Extensions to RSVP for LSP Tunnels”, IETF RFC 3209 (2001).
4) Chan, k., Sahita, R., Hahn, S., McCloghrie, K., ”Differentiated Services Quality of Service Policy Information Base”, IETF RFC 3317 (2003).
5) Floyd, S., Jacobson, V., ”Random Early Detection gateways for Congestion Avoidance”, IEEE/ACM Transactions on Networking, V1, N4, pp. 397–413 (1993).
6) Ramakrishnan, K., Floyd, S., ”A Proposal to add Explicit Congestion Notification (ECN) to IP”, IETF RFC 2481 (1999).
7) Padhye, J., Firoiu, V., Towsley, D., Kurose, J., ”Modeling TCP Reno Performance: A Simple Model and Its Empirical Validation”, IEEE/ACM Transactions on Networking, vol.8, no.2, pp. 133–145 (2000).
8) VINT Project (www.isi.edu/nsnam/vint/index.html).
9) The Network Simulator - ns-2 (www.isi.edu/nsnam/ns/).