

IP トラフィックエンジニアリングにおける実証実験の評価

小倉孝夫[†] 鈴木純司[†] 鈴木忠道^{†, ††} 青山友紀^{†, †††}

[†]通信・放送機構 幕張ギガビットリサーチセンター

〒261-0021 千葉県千葉市美浜区ひび野 1-6 パルプラザ幕張 E 棟

Tel: (043)-350-0711/0712 E-mail: {ogura, junji.suzuki}@makuhari.tao.go.jp

^{††}(株)富士通研究所

^{†††}東京大学

〒211-8588 川崎市中原区上小田中 4-1-1

〒113-8656 東京都文京区本郷 7-3-1

TEL:044-754-2618 E-mail: tasuzuki@flab.fujitsu.co.jp

TEL:03-5841-6655 E-mail:aoyama@mlab.u-tokyo.ac.jp

あらまし

インターネットの急速な普及に伴い、ある特定経路にトラフィックが集中するというような問題が生じており、それを解決するためにネットワークの転送能力を最大限に引き出すことを目的としたトラフィックエンジニアリング(TE)技術が注目されている。我々は、JGN(Japan Gigabit Network)上に TE 機能を実装した MPLS(Multi Protocol Label Switching)ルータを 4 拠点に配置し、MPLS ネットワークを構築し、TE による動的負荷分散機能に関する特性評価を行った。具体的には制御パラメータ値の変更による負荷分散制御の安定性に関する傾向を把握し、実際の高度アプリケーションとしてデジタルシネマを対象に適切な制御パラメータ値を設定し、基本動作およびスループットの向上の効果を確認した。

キーワード IP トラフィックエンジニアリング、MPLS、ロードバランシング、実証実験

Evaluation on IP Traffic Engineering Experimentation

Takao OGURA[†] Junji SUZUKI[†] Tadamichi SUZUKI^{†, ††} Tomonori AOYAMA^{†, †††}

[†]TAO Makuhari Gigabit Research Center

E, Pal Plaza Makuhari, 1-6 Hibino, Mihama-ku, Chiba-shi, Chiba, 261-0021, JAPAN

Tel: +81-43-350-0711/0712 E-mail: {ogura, junji.suzuki}@makuhari.tao.go.jp

^{††}FUJITSU LABORATORIES LTD.

^{†††}The University of Tokyo

4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki-shi 211-8588 JAPAN

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, JAPAN

Tel: +81-44-754-2618 E-mail: tasuzuki@flab.fujitsu.co.jp

TEL: +81-3-5841-6655 E-mail: aoyama@mlab.t.u-tokyo.ac.jp

Abstract

Traffic Engineering (TE) is needed in order to optimize IP network resource utilization. Especially, load balancing with TE can avoid traffic concentration on single path between ingress and egress routers. We have built MPLS(Multi Protocol Label Switching) network with TE capability in JGN(Japan Gigabit Network). We evaluated load balancing behavior from the view point of control stability and validated the throughput effect of it.

Key words traffic engineering, MPLS, load balancing, experimentation

1.はじめに

インターネットの急速な普及に伴う IP トラフィックの急激な増加により、高速かつ大規模 IP ネットワークの構築が望まれている。現状、IP ネットワークでは OSPF(Open Shortest Path First)などの IGP(Interior Gateway Protocol)ルーティングプロトコルによって経路制御がなされている。しかし、これらのプロトコルでは、Dijkstra の最短パス探索アルゴリズム^[4]により経路探索を行っているため、最短経路である特定経路にトラフィックが集中するというような問題が生じている。すなわち、現状の IP ネットワークではトラフィック特性、使用可能な帯域情報などがルーティングに考慮されていないため、ある特定のリンクにトラフィックが集中したり、一部のリンクでは利用率が低くなるなど、ネットワークリソースの利用が一部に偏る傾向がある。これらを解決するための一つの技術として、MPLS(Multi Protocol Label Switching)^{[2]-[5]}によるトラフィックエンジニアリング技術^[6]が注目されている。トラフィックエンジニアリング技術は、ネットワーク性能の最適化をねらいとし、具体的にはスループットの向上、パケット廃棄率の低下、転送遅延、ジッタの最小化を図ることを目的としている。

今後、超高速大規模ネットワークのインフラが整備されると共に、広帯域な高度アプリケーション(例えばデジタルシネマの転送速度は数十 Mbps～数百 Mbps である)の出現によって、より一層、ネットワークリソースの有効利用は重要な課題となってくる。最近では、上記 MPLS ネットワークを前提として、トラフィックエンジニアリング技術、特にある特定のリンクに集中するトラフィックを使用率の低いリンクへ自律的に分散させるロードバランシング(動的負荷分散)機能の開発、及び本機能を実装した MPLS ルータ(LSR: Label Switching Router)の実用化が進められている^[7]。すでにこの動的負荷分散機能に関しては、シミュレーション^[8]、および PC ベースの LSR を用いた実験ネットワーク^[9]等で機能検証や定量的評価が行われているが、実際のネットワークにおける有効性評価は行われていない。そこで我々は、まず動的負荷分散機能を

実装した LSR を 4 拠点(幕張ギガビットリサーチセンター、高知トラフィック通信リサーチセンター、北九州ギガラボ、東大)に設置し、これら 4 拠点を超高速大規模ネットワークである JGN (Japan Gigabit Network)で接続した MPLS ネットワークを構築した^[10]。そして、このような実ネットワークを用いて、動的負荷分散機能の有効性評価、及び実データに則した最適なパラメータ値を抽出した。

本章以降の構成を示す。まず、第 2 章で今回、LSR に実装されている動的負荷分散機能の概要について述べる。3 章では構築したネットワークの環境を説明し、4 章で実証実験による負荷分散特性評価、高度アプリケーションによるスループットの向上効果について述べる。最後に 5 章で全体のまとめを行い、今後の課題を記す。

2.動的負荷分散機能の概要

2.1 機能概要

動的負荷分散制御はネットワークのトラフィック状況に応じて自律的に複数の LSP(Label Switched Path)に IP フローを分散させ、トラフィックの輻輳を回避させる。これを実現するためにトラフィック収集通知機能、代替経路検索・設定機能、IP フロー振り分け機能の 3 つの機能から構成される。また、LSR はエッジ LSR とコア LSR から構成され、コア LSR はトラフィック収集通知機能のみを行い、エッジ LSR は 3 つの機能を行う。つまり、エッジ LSR 間で代替経路(エンジニアリングルートの LSP)を設定し、IP フローを振り分け、トラフィックを分散させる。各々の機能を下記に示す。

(1)トラフィック統計情報収集通知機能

各 LSR は、出力側のリンクの使用率(スループット)を一定間隔毎に観測する。その収集した統計情報を OSPF の Opaque LSA(Link State Advertisement)に書き込み、定期的な flooding によって Ingress 側のエッジ LSR に通知する。エッジ LSR は、すべての LSR の統計情報を収集し、データベースを構築することにより、LSP に沿った各リンクのトラフィック状況を知ることができる。

(2) 代替経路検索・設定機能

エッジ LSR は、ある特定経路の一部のリンクでトラフィック量が輻輳閾値を超えているかどうかを判断する。輻輳閾値を超えている場合、代替経路となるエンジニアリングルートを検索する。本機能では Dijkstra の最短パス検索アルゴリズムを改良しており、空き帯域の大きい LSP のルートが検索される。その後、シグナリングプロトコルである RSVP-Tunneling Protocol (RSVP-TE) により明示的に LSP を設定する。

(3) IP フロー振り分け機能

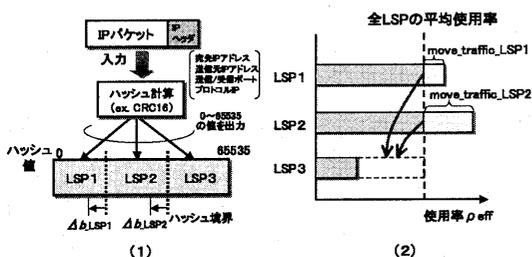


図 1. フロー振り分け処理

設定されたエンジニアリングルートとデフォルトルート（一番、最初に張られる LSP）間でトラフィックを均等に分散させるために、Ingress 側のエッジ LSR では、IP フローの振り分けを行う。図 1(1)に示すように、受信した IP パケットの宛先 IP アドレス等をキーとして、ハッシュ演算（ここでは CRC16 を使用しているため、その範囲は 0~65535 となる）により、LSP を選択する。このハッシュ演算結果の取り得る値の範囲を、同一の Egress 側の LSR に対し既に設定されている LSP の本数で分割しており、この区切をハッシュ境界値と呼んでいる。このハッシュ境界値を LSP の使用率に応じて動かすことで、LSP 間での動的なロードバランシングが可能となる。具体的には、図 1(2)に示すように、各 LSP の実効的な使用率 (ρ_{eff}) が、LSP 群全体の平均負荷に近づくようにハッシュ境界を移動させることにより負荷を調整する。これは、ある LSP に収容されている IP のフローを別の LSP に移動させることに相当する。これにより、トラフィックはバ

ランス良く分散できる。

2.2 制御パラメータ

負荷分散制御を行う上で制御パラメータは、トラフィックの振る舞いに大きく影響を及ぼす。下記に重要な制御パラメータを示す。

統計情報通知間隔: LSR 間で定期的に統計情報を交換する間隔 (OSPF 間隔) を示す。アクセス系、バックボーン系のようなネットワーク種類、または、メッシュ、リング等のネットワークトポロジーによってもトラフィック特性が異なり、また、この収集間隔を小さくすればするほど、LSR の処理が重くなるため LSR の性能に関係する。

Move Granularity: IP のフローを別の LSP に移動させるトラフィック移動量の割合を示し、上記で示したハッシュ境界値の移動量を決定する。LSP_i に関するハッシュ境界の移動量 Δb_{LSPi} は以下の式を用いて計算する。

$$\Delta b_{LSPi} = \frac{move_traffic_LSPi}{\sum \rho_{eff} - LSPi} \times 65535 \times move_granularity$$

ここで $move_traffic_LSPi$ は、LSP_i に関する LSP 全体の平均使用率との差分を表す。また整数値 65535 はハッシュ演算がとりうる値の範囲である。

なお $move_granularity$ は 1.0 以下の値をとり、1 回のフローの移動量を抑制するためのパラメータである。例えば $move_granularity=0.5$ であれば、計算されたハッシュ移動量の 50% (以下、Move Granularity: 50%と記す) だけハッシュ境界を移動する。

Smoothing Factor(SF): パーストラフィック等の急激な変化のトラフィックによる輻輳検出を防ぐため、トラフィックの平滑化 (スムージング) を行う。ここでは過去 n 回 (n: 整数) の統計情報データの平均を示す。

輻輳閾値: 統計収集結果により、輻輳の検出を判断するための閾値を示す。輻輳が生じてから負荷分散を起こすことも可能であるが、閾値を低く設定することにより輻輳前に負荷分散を起こし、事前に輻輳を回避させることができる。そのため、このパラメ

ータ値の設定はネットワーク管理者の運用ポリシーに依存すると考えられる。

3. 実験環境

3.1 ネットワーク構成

図2で示すように、幕張ギガビットリサーチセンター、高知トラフィック通信リサーチセンター、北九州ギガラボ、東大の4拠点に上記、動的負荷分散機能を実装したMPLSルータを設置し、MPLSネットワークを構築した。MPLSルータにはFast EtherインタフェースとATM 155Mbpsインタフェースを配備しており、ATMインタフェースでJGNと接続しており、ATM CBR 120Mbpsのメッシュ構成で4拠点を結んでいる。また、トラフィックの発生は測定器(smartbit, Chariot)を用いて、一定レートのIPトラフィックを発生させ、実験を行った。尚、このような広帯域かつ大規模ネットワークは、将来のバックボーン系ネットワークを想定した実験モデルとして考えられる。

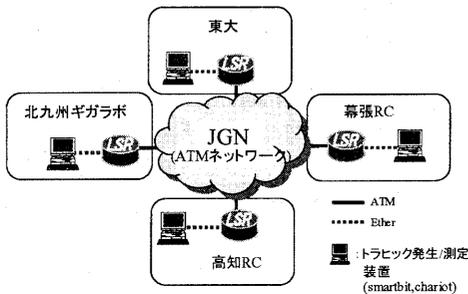


図2. 実験システム概要

3.2 実験モデル

負荷分散機能の特性評価を行うために、下記の2つのモデルで実験を行った(図3)。本実験において負荷分散機能による基本動作の検証を目的とし、トラフィック特性の解析を容易に行うため、北九州のLSRのみ負荷分散機能を動作させ、他の拠点では負荷分散機能を停止した。また、トラフィックは片側方向のみで負荷分散機能の動作を検証した。

モデル1：北九州から高知経由で幕張へトラフィック(54Mbps)を流しておき、その後、高知から幕張へ別

のトラフィック(20Mbps)を流す。この場合、高知・幕張間のトラフィックが輻輳閾値(40%)を超えたことを検出した北九州のLSRは、エンジニアリングルート(LSP)を北九州から東大経由で幕張まで確立する。
モデル2：北九州から高知経由で幕張へトラフィック(54Mbps)を流しておき、その後、北九州から高知へ、及び東大から幕張への2箇所同時にトラフィック(両方、20Mbps)を流す。この場合、北九州・高知間のトラフィックで輻輳閾値(40%)を超えたことを検出した北九州のLSRは、北九州・高知・幕張のトラフィックに対しては北九州・東大・幕張のエンジニアリングルートを確認する。また、北九州・高知のトラフィックに対しては北九州・東大・高知のエンジニアリングルートを確認する。

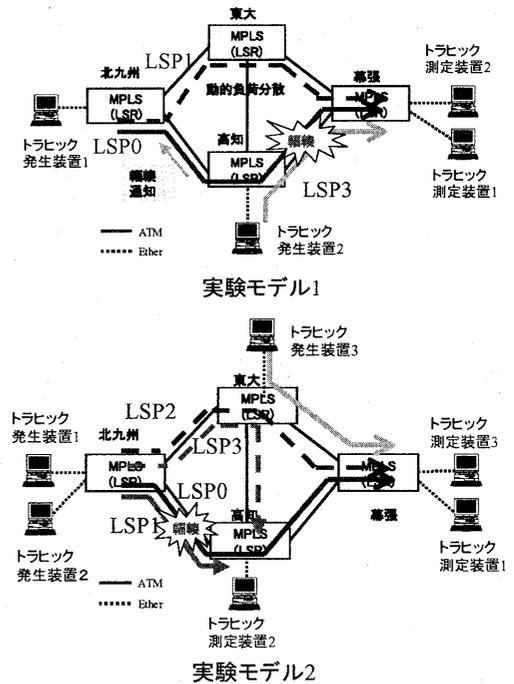


図3. 実験系モデル

このモデルでのトラフィック量は、測定器から発生するIPレベルのトラフィック量を示す。本実験では、輻輳によるデータ転送のロスを防ぐため、輻輳閾値を低く設定することにより、事前に輻輳回避を行う負荷分散機能の検証を行った。また、幕張RCから

各拠点の LSR および測定器をリモートコントロールし、測定を行った。

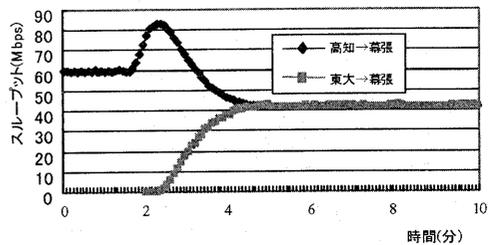
4. 実験結果および考察

4.1 負荷分散特性

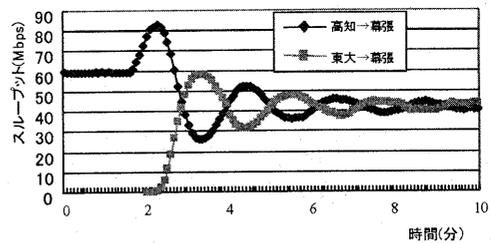
図4は、実験モデル1において Move Granularity 値 (10%,30%,50%) を変更した場合 (Smoothing Factor : 5、統計情報通知間隔 : 5 秒) の経過時間に対するトラヒック量 (スループット) を示している。図4の Move Granularity:10%の場合で示すように、高知・幕張間のトラヒックは減少し、東大・幕張間のトラヒックが急激に増加していることにより、北九州・東大・幕張間でエンジニアリングルートが張られ、これらの2つのルートでトラヒックが均等に分散されていることがわかる。Move Granularity:30%の場合は10%の場合と比較するとトラヒック量は振動しながら均等化される。さらに、50%の場合ではトラヒックは大きく振動したまま均等化されない。これは、各コア LSR での統計収集結果が Ingress 側のエッジ LSR のフロー振り分け処理へ反映されるまでにタイムラグがあることからエッジ LSR のフロー振り分けが過制御になるためである。このようにトラヒックが振動する場合、あるフローにおいてはデフォルトルートとエンジニアリングルート間を行き来して、不安定なデータ転送になると考える。

図5は、実験モデル2において Smoothing Factor 値(1,5,10)を変更した場合(Move Granularity : 30%、統計情報通知間隔 : 5 秒)の経過時間に対するトラヒック量 (スループット) を示している。Smoothing Factor 値が1の場合を見ると、北九州・高知間のトラヒックの増加により北九州・高知・幕張のトラヒックは北九州・東大・幕張のエンジニアリングルートへ、北九州・高知間のトラヒックは北九州・東大・高知のエンジニアリングルートへトラヒックを分散した。そのため、北九州→東大、東大→幕張、東大→高知の各々のスループットが急激に増加した。ここで、北九州・高知間の 20Mbps のトラヒック (LSP1)は空き帯域の大きい北九州・東大・高知間 (LSP3)のルートへすべて移動したことがわかる。

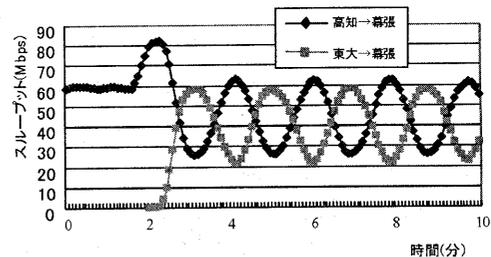
Smoothing Factor 値が 5 の場合は、Move Granularity 値を変更した場合と同様に振動しながらトラヒックが安定する。さらに Smoothing Factor 値が 10 の場合は、トラヒックが振動したまま安定しない。これは統計情報のスムージング処理により得られた値は、実際のトラヒック量に比べ変化量が少なく、実際のトラヒック量まで変化するのに時間がかかるため、過制御になるものと考えられる。



Move Granularity:10%の場合



Move Granularity:30%の場合



Move Granularity:50%の場合

図4. 経過時間におけるスループット特性 (実験モデル1の場合)

このように制御パラメータ (Move Granularity, Smoothing Factor) 値によってトラヒックが不安定な振舞い (トラヒックの振動現象) を起こす場合があり、入力フローに影響を及ぼす。また、負荷分散

機能が動作していても、トラフィックが不安定であると、スループットの向上は図れず効果的でないと考えられる。そのため、できるだけ早くトラフィックが均等に分散できるかがキーとなり、負荷分散制御を実現する適切なパラメータ値の抽出が重要である。そこで、次の章でネットワークの安定性（トラフィックが均等に分散できること）について考察した結果を示す。

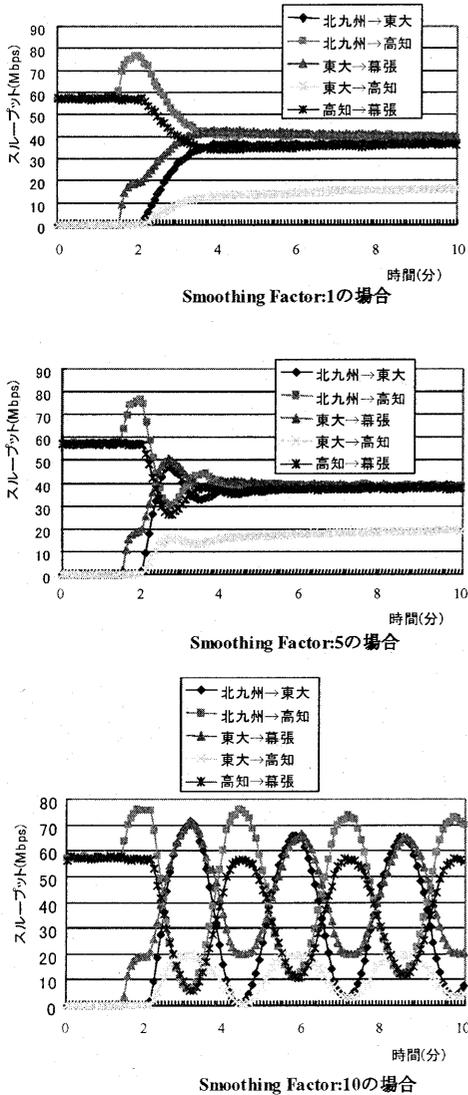


図5. 経過時間におけるスループット特性 (実験モデル2の場合)

4.2 ネットワークの安定性

負荷分散制御を行ってトラフィックが均等に分散されるまでの時間を収束時間（エッジ LSR が輻輳を検出し、エンジニアリングルートの設定要求時からデフォルトルートとエンジニアリングルートを通過する最大リンク使用率の差が0.01%以下、かつ3回連続検出した場合）と定義し、Move Granularity、及び Smoothing Factor との関係を求めるため、収束時間の測定を行った。

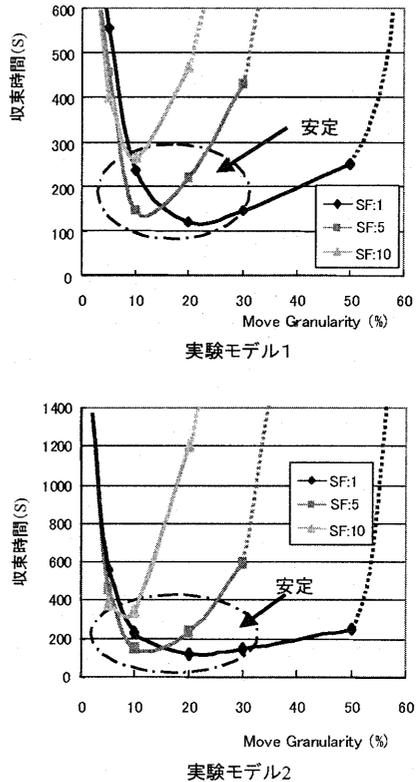
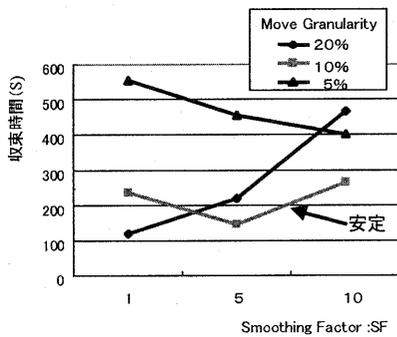


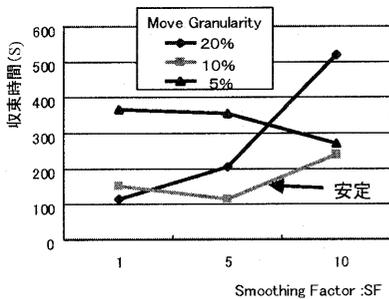
図6. Move Granularity と収束時間との関係

図6は Move Granularity と収束時間との関係を示している。まず、実験モデル1の結果をみるとグラフの形はUの形を示す。Move Granularity 値が小さい場合は、トラフィックの移動量が少しいため、徐々にトラフィック量が均等化され、Move Granularity 値が大きいと振動しやすくなり収束時間もかかり、点線部分より Move Granularity 値が

大きい場合では振動したまま収束しない。この図から安定する領域は、5%から30%のところであることがわかる。また、Smoothing Factor 値によっても傾向は同じであるが、Smoothing Factor 値が小さい方が安定する領域が大きく、Smoothing Factor 値が大きいと安定する領域は小さく振動しやすくなる。これは先程も述べたように統計情報のスムージング処理によりトラヒックの変化量が少なく、実際のトラヒックまで変化するのに時間がかかるため、過制御になるものと考えられる。バックボーン系ネットワークのような短時間での急激なトラヒック変動が少ないような場合を考慮すると Smoothing Factor 値が小さくても良いと考えられる。このように実験モデル1、2の2つの結果で同様な傾向が得られた。



実験モデル1の場合



実験モデル2の場合

図7. Smoothing Factor と収束時間との関係

図7では Smoothing Factor における収束時間を示している。実験モデル1を見ると Move

Granularity 値が 10%の場合、Smoothing Factor の値に関係なく収束時間はほぼ一定であるため、トラヒックが均等に分散される最適なパラメータ値として考えられる。また、実験モデル2においても Move Granularity 値が 10%の場合、Smoothing Factor 値に関係なくトラヒックが均等に分散される最適なパラメータ値として考えられる。

以上の結果より、Move Granularity 値を 10%に設定すれば Smoothing Factor の値に関係なく、300秒以内にトラヒックが分散されることがわかる。バックボーン系ネットワークをターゲットとすれば急激なトラヒック変動は少ないと考えられるため、この収束時間でも、十分効果的な負荷分散が可能であると考えられる。

4.3 負荷分散によるスループット効果

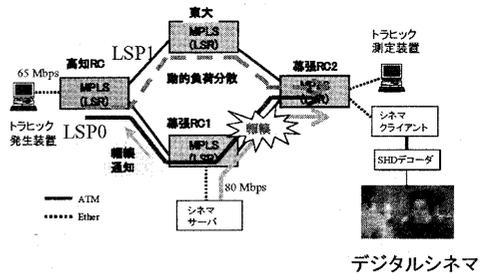


図8. 負荷分散によるスループット向上効果例

上記で抽出した適切なパラメータ値を設定し、実際にデジタルシネマ・アプリケーションを使用して、負荷分散によりスループットが向上した一例を以下で示す。図8は幕張 RC1 から幕張 RC2 へデジタルシネマ・アプリケーションを 80Mbps の IP 転送で行った。また、高知 RC から 65Mbps の IP トラヒックを流し、幕張 RC1・幕張 RC2 間で輻輳を発生させる。幕張 RC1・幕張 RC2 間では ATM レートに換算すると約 170Mbps のトラヒック量になるため、負荷分散機能がない場合、約 50Mbps のセル廃棄 (拠点間は ATM CBR 120Mbps の VP で構成されているため) が起こり、デジタルシネマのアプリケーションはフレーム落ちが生じ、アプリケーション

ョンに大きく影響を及ぼした。一方、負荷分散機能により、高知 RC からのトラヒックを高知 RC・東大・幕張 RC2 のエンジニアリングルートへ分散させ、デジタルシネマの転送フレームの廃棄がなくなることを実際に確認した。ここでは、スループットの向上として、2つのルートに分散することにより170Mbps の転送を可能にしているため、単純に計算すると約 1.4 倍のスループットの向上を図っている。これは、一例であるがスループットの向上率はネットワークの形態（リング、メッシュ等）、トラヒックタイプに依存すると考えられる。

本ネットワーク形態（ただし、幕張 RC2 をメッセ会場とする）で CEATEC JAPAN 2001 に出展し、動的負荷分散機能の有効性に関し、デモンストレーションを行った。

5. まとめ

本稿では、JGN を使用して 4 拠点に MPLS ネットワークを構築し、トラヒックエンジニアリング技術、特に動的負荷分散機能に関する特性評価を行った。その結果、制御パラメータ値の変更による負荷分散制御の安定性がわかり、大規模な実ネットワークにおいて適切な制御パラメータ値を設定することにより複数の経路にトラヒックが均等に分散されることが確認できた。さらに、高度アプリケーション（デジタルシネマ）を使用した形態でも、上記の有効なパラメータ値を設定し、動作確認およびスループットの向上が図れることを確認した。

今後はさらにアプリケーションに注目して、特に TCP/IP フローの負荷分散特性の実証実験を引き続き行い、検討を進める予定である。

参考文献

- [1] R.Gallager et al., "Data Networks", Prentice-Hall, Inc 1987.
- [2] E. Rosen et al., " Multi-Protocol Label Switching Architecture ", RFC3031, January 2001.
- [3] L. Andersson et al., "LDP Specification", RFC3036, January 2001.
- [4] B.Jamoussi et al., "Constraint-Based LSP Setup using

LDP", draft-ietf-mpls-cr-ldp-05.txt, February 2001.

[5] D.Awduche et al., "Extensions to RSVP for LSP Tunnels", draft-ietf-mpls-rsvp-lsp-tunnel-08.txt, February 2001.

[6] D.Awduche et al., "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[7] 宗宮、仲道、高島、小林,"IP トラヒックエンジニアリングにおけるロードバランシング方式の検討", 電子情報通信学会交換機研究会, SSE2000-53, June 2000.

[8] 仲道、高島、宗宮,"IP トラヒックエンジニアリングシステムの安定性検討", 電子情報通信学会交換機研究会, SSE2000-269, March 2001.

[9] 山田、高島、仲道、宗宮、中後,"トラヒックエンジニアリングシステムの実機評価", 電子情報通信学会ネットワークシステム研究会, NS2001-8, April 2001

[10] 鈴木忠道、鈴木純司、榎本、青山,"超高速ネットワークの動的制御に関する一考察", 電子情報通信学会交換機研究会, SSE2000-102, September 2000