高品質インターネット　3－3
(2002. 6. 26)

# MPLSを用いた次世代IXにおける転送能力評価

楠田 友彦[i]　石原 丈二[ii]　西内 一馬[iii]　永見 健一[ii]
中川 郁夫[i]　江崎 浩[iv]　菊池 豊[iii]

i インテック・ウェブ・アンド・ゲノム・インフォマティクス株式会社
ii 株式会社東芝
iii 高知工科大学
iv 東京大学

E-mail: i kusuda@isl.intec.co.jp, ikuo@intec.co.jp, ii {takeshi3.ishihara, ken.nagami}@toshiba.co.jp
iii(kazuma, yu)@kikuken.org, iv hiroshi@wide.ad.jp

**あらまし**　ISP(Internet Service Provider)を相互に接続する技術として IX(Internet eXchange)がある。我々はMPLS(Multi Protocol Label Switching)を用いた次世代IXとしてMPLS-IX技術を提案し、JGN(Japan Gigabit Network)を用いたテストベットを構築して運用を行っている。本稿では、ルータのパケット転送能力比較を中心に、MPLS-IX と既存IX技術との比較を行う。

**キーワード**　IX, MPLS

# Evaluation of packet forwarding performance on MPLS-IX

Tomohiko KUSUDA[i]　Takeshi ISHIHARA[ii]　Kazuma NISHIUCHI[iii]　Kenichi NAGAMI[ii]
Ikuo NAKAGAWA[i]　Hiroshi ESAKI[iv]　and Yutaka KIKUCHI[iii]

i INTEC Web and Genome Informatics Corporation
ii Toshiba Corporation
iii Kochi University of Technology
iv University of Tokyo

E-mail: i kusuda@isl.intec.co.jp, ikuo@intec.co.jp, ii {takeshi3.ishihara, ken.nagami}@toshiba.co.jp
iii(kazuma, yu)@kikuken.org, iv hiroshi@wide.ad.jp

**Abstract**　We propose a new IX(Internet eXchange) architecture using MPLS(MultiProtocol Label Switching), called MPLS-IX. This paper describes the evaluation of packet forwarding performance on MPLS-IX.

**Keyword**　IX, MPLS

## 1．はじめに

近年、ADSL、FTTH等の高速なアクセス回線が一般家庭に浸透し始めており、ISP(Internet Service Provider)間におけるトラフィック交換を従来よりも高速、広帯域、かつ、効率的に行いたいという要求が強くなってきている。

一般に、ISP間のトラフィックを効率的に交換する技術として、IX(Internet eXchange)が利用されている。

しかし、現在最も利用されているEthernetスイッチによるLANベースのIXでは、こうしたトラフィック増加に対して転送速度の面から対応が困難になっていくことが予想される。

我々は、こうした状況を解決するために、高速なトラフィック交換を可能にする次世代IX技術に関する研究を目的とする「次世代IX研究会」を立ち上げた。

本研究会では、次世代のIXとして、MPLS(Multi Protocol Label Switching)をベースとしたIXのアーキテクチャを提案している。このアーキテクチャをMPLS-IXと呼ぶ。

我々は、これまで、JGN(Japan Gigabit Network)を利用したMPLS-IXのテストベット構築やマルチベンダによるMPLSルータの相互接続実験を行い、MPLS-IXの機能検証を行ってきた。

本稿では、相互接続実験の一環として行っているルータの転送能力測定試験についての報告を行い、その結果からMPLS-IXと既存IX技術との比較を行う。

## 2．MPLS-IX

次世代IX研究会では、新しいIXのアーキテクチャとして、MPLS-IXの提案を行っている。以下にその概要と特徴を示す。

### 2．1　MPLX-IXのアーキテクチャ

MPLSとは、従来のIPによるパケット転送とは異なり、「ラベル」と呼ばれる固定長の識別子によってパケットを転送する技術である。

MPLS-IXは1つ、あるいは、複数台のMPLSに対応したCoreルータにより構成される。

Coreルータ間はOSPFによりIPの経路制御が行われている。ISP等のIX接続組織は他組織との経路交換を行うためのルータをCoreルータに接続する。このルータをEdgeルータと呼ぶ。

Edgeルータでは、FEC(Forwarding Equivalence Class)と呼ばれる同一のルール(ここでは、経路交換を行う対向組織のEdgeルータへの宛先IPアドレス)に基づいて転送されるパケットの集合に対して、ラベルを割り当て、パケットの中に格納して転送する。

MPLS-IX網を構成するCoreルータでは、パケットに格納されたラベルの情報からパケットのネクストホップのルータを認識して転送を行う。Coreルータでは、パケットの転送に関してIPヘッダの情報をいっさい参照しない。

ラベルを用いることにより、経路交換を行うEdgeルータ間に仮想的なパスが形成される。このパスをLSP(Label Switching Path)と呼ぶ。

IXユーザはこのLSP上でBGP4により経路交換を行う。Coreルータは経路交換にいっさい関与しない。

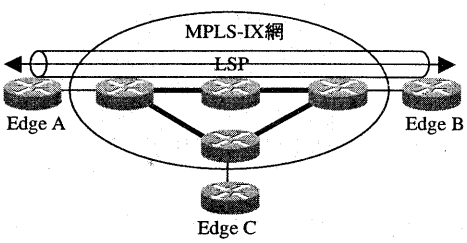


図1　MPLS-IXの概要

### 2．2　MPLS-IXの特徴

MPLS-IXは以下の特徴を持つ。

・　データリンクメディア非依存

MPLS技術を用いることで、Ethernet、ATM、POSといったデータリンクメディアの差異を抽象化することができる。そのため、特定のデータリンクメディアに依存せず、柔軟で高速なトラフィック交換が可能になる。今後、更なる高速インターフェースが登場した場合でも、基本的にMPLSに対応させることが可能であり、拡張性に優れている。

・　広域分散

MPLS-IXは、前述のように、POS等の広域利用可能なデータリンクメディアにも対応している。Coreルータ間の接続にこうしたデータリンクメディアを利用することにより、IXへの接続点を広域に分散配置することが可能である。これにより、広域分散環境における組織間の相互接続が可能になる。

また、データリンクメディアの選択次第でIXに接続するEdgeルータをIXの接続点が存在する施設に置く必要がなくなり、従来よりもメンテナンスが容易になる。

・LANベースIXの問題点の解決

従来のLANベースのIXでは、接続ルータがサブネットを共有するため、ブロードキャストトラフィックの影響を受ける。また、拡張の際には、アドレスのリナンバリングが必要になる。

MPLS-IXでは、ユーザ間に仮想的なパスを設定することにより、第3者のトラフィックの影響を遮断することが可能になる。

## 2. 3 実証実験

次世代IX研究会では、MPLS-IXのアーキテクチャの機能検証を目的として、いくつかの実証実験を実施している。

・ テストベットの構築

広域分散環境におけるIX機能の実証実験を行うため、JGN(Japan Gigabit Network)上にMPLS-IXのテストベットを構築した。東京、大阪、富山、高知の4県を拠点とし、6箇所の施設にCoreルータを設置した。現状では、20程度の組織が接続を行い、MPLS-IX上で経路交換を行っている。(2002年5月現在)

・ ルータ相互接続実験

MPLSは比較的新しい技術であるため、現状では、ルータベンダによっては我々が提案するMPLS-IXを実現するための機能が十分に実装されていない部分が見られる。そこで、こうした機能の改善要求とマルチベンダ環境での相互接続性の検証を行うためにルータ相互接続実験を実施している。

2002年5月現在、3回の相互接続実験を行い、10社を超えるルータベンダが実験に参加している。その実験の一環として、ルータの転送能力測定試験を行っている。

# 3. 転送能力測定試験

本章では、ルータの転送能力測定試験について述べる。

## 3. 1 転送能力測定試験の目的

転送能力測定試験の主な目的として以下のものが挙げられる。

・ ラベル転送における転送速度の検証

IPによる転送とMPLSによる転送とでは、まったく異なるアルゴリズムによって転送が行われる。

IPによるパケット転送は、宛先IPアドレスをもとに行われる。

一方、MPLSによる転送は、パケットに格納されたラベルをもとに行なわれる。ラベルは固定長の4バイトの識別子であり、そのうちの20ビットのラベル値が転送テーブルからネクストホップを選択するインデックスとして使用される。

このように、IPとMPLSでは、転送アルゴリズムが大きく異なっているが、MPLSを用いてIXを構築することを考えた場合、少なくとも既存のLANベースIXと同程度の転送速度が得られることが望ましい。

・ ラベル転送における遅延の検証

今後、インターネットにおいて、ストリーミングやVoIPといったリアルタイムアプリケーションの利用増加が考えられる。こうしたリアルタイムアプリケーションにとって、遅延は品質の維持において重要な問題となる。

転送速度と同様、遅延に関しても従来のLANベースIXと同程度の性能が必要とされる。

## 3. 2 転送能力測定試験の構成

転送能力測定試験では、MPLS-IXを構成するルータがラベルによってパケットを転送する際の転送能力を測定する。

また、比較のため、従来のLANベースのIXに接続することを想定したIPによる転送能力測定も行う。

・ IPの転送能力測定

LANベースのIXにおける転送能力測定は、図2のようにISP A

の顧客 User A から ISP B の顧客 User B に向けてトラフィックが流れている状況を想定して行う。

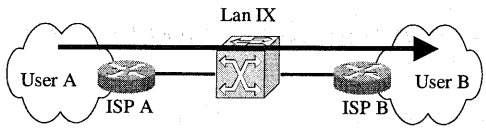


図2　LAN ベースの IX の構成

図2の場合、ISP A のルータは、User A から User B へのパケットを通常どおり、IP ヘッダの情報に基づいて転送している。この場合の ISP A のルータのパケット転送能力を測定する。

・MPLS の転送能力測定

MPLS-IX における転送能力測定は、LAN ベースの IX における転送能力試験と同様に、ISP A の顧客 User A から ISP B の顧客 User B に向けてトラフィックが流れている状況を想定して行う。

図3に構成を示す。

MPLS-IX を構成するルータは、パケットに対する処理の違いによって3つに分類される。転送能力測定はそれぞれの処理ごとに行う。

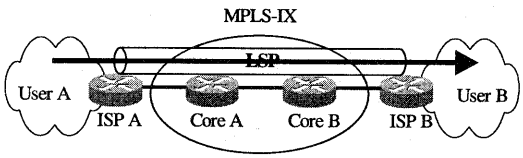


図3　MPLS-IX の構成

・ PUSH

パケットのラベルスタックに対してラベルを追加する処理。

ISP A のルータで PUSH が行われる。User A から転送されてきた IP パケットにラベルを追加する。経路交換を行う対向ルータごとに異なるラベルが追加される。転送はラベルに基づいて行われる。

・ SWAP

ラベル付きで転送されてきたパケットのラベルを、新しいラベルに付け替える処理。

Core ルータで SWAP が行われる。Core A では、ISP A からラベル付きで転送されてきたパケットのラベルを参照し、ISP B へ向かうラベルに付け替える。ラベルを付け替えられたパケットはネクストホップである Core B に転送される。この際の転送はすべてラベル情報に基づいて行われる。

・ POP

ラベル付きで転送されてきたパケットのラベルを取り除く処理。

ISP B のルータで POP が行われる。ISP B のルータは、Core B からラベル付きで転送されてきたパケットからラベルを取り除く。そして、ラベルを取り除いたパケットを、IP ヘッダ情報をもとにネクストホップである User B へ転送する。

### 3. 3　転送能力測定試験の方法

今回の測定試験では、アジレントテクノロジー社のルータテスタを測定器として使用した。ルータテスタは、さまざまなパラメータを持つパケットの生成機能や経路情報の注入を行う BGP4、MPLS においてラベルの配布を行う LDP、RSVP-TE といった各種プロトコルのエミュレーション機能を備えている。このエミュレーション機能を利用することにより、2. 2章で示した図2, 3のような仮想的ネットワークを1台のルータテスタで構築することができる。

図4にルータテスタを使用する場合の物理構成を示す。

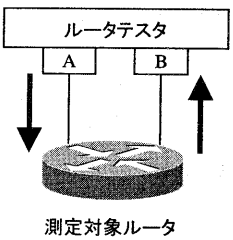


図4　ルータテスタによる実験構成

ルータテスタは、インターフェースの種類ごとに複数のポートを持っており、例えば、ポート A を図2における User A、ポート B を ISP B のルータと仮定した設定を行うことにより、測定対象ル

ータをISP A のルータと見立てた実験が可能になる。

・ IP 試験

図5 に IP 転送試験の論理的な構成を示す。測定器と測定対象ルータの接続インターフェースは GigabitEthernet を用いる。図 5 では、測定器を 2 台として記述しているが、物理的には 1 台で構わない。

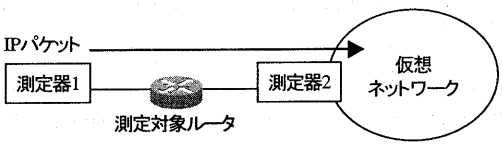


図5 IP 試験

IP による転送試験は以下のように行う。

まず、測定器 2 から測定対象ルータに BGP4 でネットマスク長 24 の 50000 経路を注入する。その経路へのネクストホップは測定器 2 とする。測定器 1 から測定対象ルータに注入した経路すべてに向けた IP パケットを送信し、測定器 2 で測定対象ルータが 1 秒間に転送したパケット数(PPS)、パケット転送による 1 秒間の平均遅延を測定する。パケットを送信する時間は 1 分間とする。

測定器 1 から転送するパケット数は、測定対象ルータにおいてパケットロスが発生せず、平均遅延が一定の値で安定して推移する状態を維持可能な限界値に設定する。ただし、転送するパケット数が少ない場合の遅延と比較して、オーダが変化するほど大幅に異なる値で安定している場合は、測定の対象から除外する。例えば、ある測定対象ルータに対して 1000pps でパケットを送信した場合に、遅延が 10$\mu$s で安定しているとする。同じ測定対象ルータを対象に転送パケット数を 2000pps に増加させて遅延を測定した時に、1000$\mu$s 前後で一定に安定したとしても、測定の対象外とし、10$\mu$s 前後で安定するまで転送パケット数を制限する。

転送パケット数の限界値を設定する場合、pps 単位ではなく、1Mbps 単位でパケット数を調整する。

送信する IP パケットのサイズは、46、66、130、258、514、1026、1498 バイトとし、それぞれの場合について測定器 2 での受信パケット数、平均遅延を測定する。パケットのサイズには、IP ヘッダの 20 バイトも含む。

・ PUSH 試験

まず、測定対象ルータと測定器 2 の間に LSP を設定する。これにより、測定対象ルータでは、測定器 2 へ転送するパケットに対し、ラベルを PUSH する処理を行う。測定器 2 では、ラベルが埋め込まれたパケットが受信される。ラベル配布プロトコルは、LDP、RSVP-TE を問わない。いずれかの方法で LSP が確立できればよい。確立する LSP は 1 本とする。

LSP の設定を行った上で、IP 転送試験と同様、測定対象ルータに 50000 経路を注入し、測定器 1 からその経路に対して IP パケットを送信する。測定器 2 で受信したパケット数、遅延を測定する。パケット送信時間、転送パケットサイズも IP 試験と同様である。

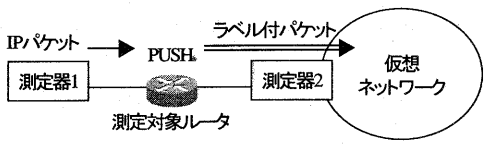


図6 PUSH 試験

・ SWAP 試験

SWAP 試験では、測定器 1 と測定器 2 の間に LSP を設定する。ラベル配布プロトコルは LDP、RSVP-TE を問わない。

測定器 1 から測定器 1 と測定器 2 の間に確立された LSP にパケットを送信し、測定器 2 での受信パケット数、遅延を測定する。

送信パケット数の決定方法、パケット送信時間、転送パケットサイズは IP 試験と同様である。

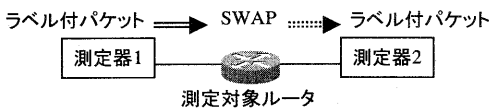


図7 SWAP 試験

・ POP 試験

POP 試験では、測定器 1 と測定対象ルータの間に LSP を設定する。ラベル配布プロトコルは LDP、RSVP-TE を問わない。LSP を確立できれば良い。

測定器 1 から測定対象ルータとの間に確立された LSP にパケッ

トを送信し、測定器2で受信されるパケット数、遅延を測定する。POPにより、測定器2ではIPパケットを受信する。

送信パケット数の決定方法、パケット送信時間、転送パケットサイズはIP試験と同様である。

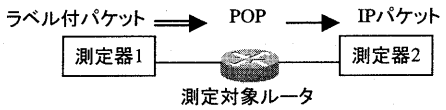


図8　POP試験

## 4. 結果

測定は5つのベンダのルータを対象に行った。

MPLSの場合、4バイトのラベルをEthernetフレームに格納するため、IPの場合よりもパケット1つ当りのサイズが大きくなる。そのため、MPLSとIPとでは、GigabitEthernetにおいて転送可能である最大PPSが異なるため、転送したパケットの個数では、単純にIPとMPLSの転送能力の比較はできない。

そこで、GigabitEthernetで転送可能であるIPパケットの最大PPSを図9の式により算出し、MPLSパケットの最大PPSを図10の式により算出する。これらの式で計算した結果、それぞれのIPパケットサイズにおける最大PPSは表1のようになる。

この値を用いて実測値をパーセンテージで表し、転送能力の比較を行う。

各ベンダの測定結果を図12～16で示す。パケットサイズごとにIP、PUSH、SWAP、POP試験それぞれの場合のPPS、平均遅延を表している。

$$
\begin{aligned}
(\max) &= \frac{1G}{(IPG + preamble + DA + SA + TYPE + IP + FCS) \times 8} \\
&= \frac{10^9}{(12 + 8 + 6 + 6 + 2 + IP + 4) \times 8} \\
&= \frac{125000000}{38 + IP}
\end{aligned}
$$

図9　GigabitEthernetにおけるIPの最大PPSの計算式

$$
\begin{aligned}
(\max) &= \frac{1G}{(IPG + preamble + DA + SA + TYPE + Label + IP + FCS) \times 8} \\
&= \frac{10^9}{(12 + 8 + 6 + 6 + 2 + 4 + IP + 4) \times 8} \\
&= \frac{125000000}{42 + IP}
\end{aligned}
$$

図10　GigabitEthernetにおけるMPLSの最大PPSの計算式

表1　GigabitEthernetにおける最大PPS

| パケットサイズ | 46 | 66 | 130 | 258 | 514 | 1026 | 1498 |
|---|---|---|---|---|---|---|---|
| IP | 1488095 | 1201923 | 744048 | 422297 | 226449 | 117481 | 81380 |
| MPLS | 1420464 | 1157414 | 726744 | 416669 | 224822 | 117042 | 81170 |

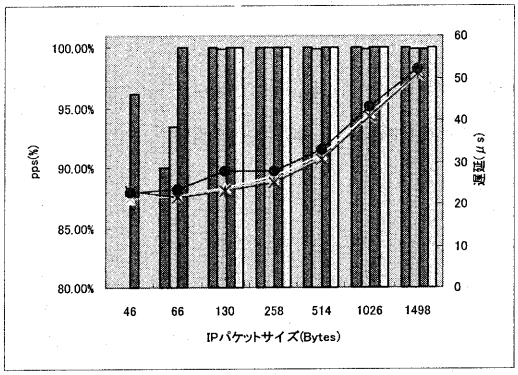


図11　A社の結果

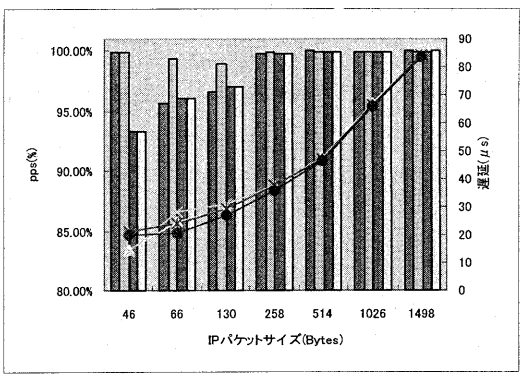


図12　B社の結果

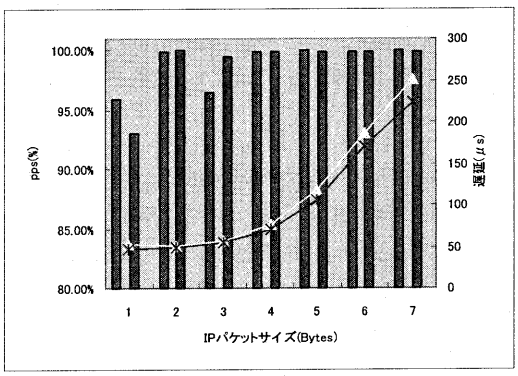


図13　C社の結果

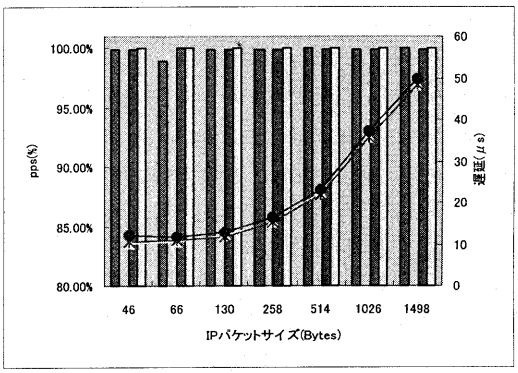


図14　D社の結果

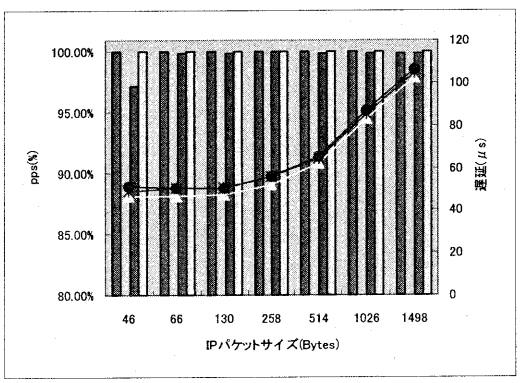


図15　E社の結果

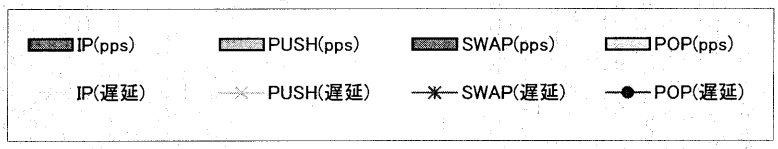


図16　凡例

## 5. 考察

各ベンダともIPによる転送とMPLSによる転送では、PPSの割合において大きな差はない。IP、MPLSともにほぼワイヤレートによる転送が可能であった。

MPLSの比較では、ほとんどのベンダで、PUSH、SWAPに比べてPOPがわずかながらPPSが高いという結果が得られた。

パケットサイズが130バイト以下の場合において、転送能率が低下するベンダがあった。パケットサイズが小さい分、転送すべきパケット数が多くなるため、処理可能なパケット数の限界を超え、処理しきれなくなったと考えられる。

また、パケットサイズが小さい場合には、IPとMPLSでの転送能力に違いが見られる傾向があった。しかし、ベンダによって、MPLSによる転送の方が高速であったり、IPによる転送の方が高速であったりと結果がまちまちであり、実装に依存した傾向であると考えられる。

遅延もPPSと同様に、IPとMPLSの間に大きな差は見られない。ただし、ほとんどのベンダにおいてIP、PUSH、SWAPの遅延よりもPOPの遅延が1～2$\mu$s程度高い傾向が見られた。

## 6. 終わりに

本稿では、MPLSによる転送能力測定試験についての結果から、IPによる転送能力とMPLSによる転送能力には差がないことを示した。

今回は、あるISPがIXを利用して1つのISPと接続することを想定した実験を行った。しかし、より現実的には、複数のISPとの接続を想定しなければならない。次回以降、そうした環境下における実験を行う必要があると考えられる。

次世代IX研究会では、今後、MPLS-IX上でのIPv6によるトラフィック交換や通信の品質保証などの新しい機能の追及を行い、導入していく予定である。

## <参考文献>

[1]次世代IX研究会, http://www.distix.net/

[2]中川 郁夫, 林 英輔, 高橋 徹, 江崎 浩, "次世代インターネットエクスチェンジの技術動向," 情報処理(IPSJ Magazine), vol.42 no.7, Jul. 2001.