第Ⅴ部

ラベルスイッチ技術によるインターネットの構築実験

第5部
ラベルスイッチ技術によるインターネットの構築実験

第1章　はじめに

LAST (LAbel Switch Technology) WGでは、ラベルスイッチ技術を用いた研究開発を行っている。ラベルスイッチ技術は、IETF MPLS (Multi Protocol Label Switching) WGで標準化されているため、MPLSと呼ばれることもある。

本報告書では、MPLSを用いた分散IXのアーキテクチャ提案とそれに必要な機能の説明とあやめプロジェクトによるMPLSのBSD上の実装報告および実証実験報告を行う。

第2章　MPLSを用いた分散IXアーキテクチャ

ISP(Internet Service Provier) 間の相互接続を効率的に行う技術としてIX(Internet eXchange) が使われている。従来のIX(Internet eXchange) 技術の典型的な例はイーサネットなどで構成されるLAN型IXとATMを用いたATM型IXである。これら2つは、いくつかの問題点がある。例えば、ATMではOC-48(2.4Gbps) が限界といわれており、また、イーサネットでは、ギガビットイーサネットが現在の主流であるため、これ以上の高速なデータリンクでIXを用いてISP間を接続することはできない。しかし、OC-192など高速なデータリンクが存在して来ているため、これに見合う高速なデータリンクでISP間を接続する技術が必要がある。本章では、様々なデータリンクメディアを利用できるMPLSの利点を生かしたIXアーキテクチャを提案し、このMPLS IXに必要なルータの機能および設定項目を整理する。

2.1 背景

インターネットは、多数のISP (Internet Service Provider) が存在し、そのISP間を相互接続することによって成り立っている。ISP間の相互接続を効率的に行う技術としてIX (Internet eXchange) が使われている。現在では、世界に数百ものIXが存在し、IXはISP間の膨大なトラフィック交換を支える重要な通信基盤の役割を果たしている。

従来のIX (Internet eXchange) 技術の典型的な例はイーサネットなどで構成されるLAN型IXとATMを用いたATM型IXである。これら2つは、いくつかの問題点がある。1つの例として、ATMではOC-48 (2.4 Gbps) が限界といわれており、また、イーサネットでは、ギガビットイーサネットが現在の主流であるため、これ以上の高速なデータリンクでIXを用いてISP間を接続することはできない。しかし、OC-192など高速なデータリンクが存在して来ているため、これに見合う高速なデータリンクでISP間を接続する必要がある。

本報告書では、これら典型的なIX技術の問題点を解決する新しいIX技術であるMPLSを用いたIX技術（MPLS-IX）を提案する。このMPLS-IXは、主に以下の特長を持つ。

- 様々なデータリンクメディアを使用してIXに接続することができ、異なるデータリンクメディアで接続しているISP間でピアリングができる
- 広域に適したデータリンクメディア（POS等）を利用することで、IXを広域に分散させることができる
- 高速なデータリンクメディア（POSやDWDM等）を使用することで、その時代にあった高速なIXを実現することができる

まず始めに、MPLS IXに対する要求を列挙し、必要な機能を整理する。次に、MPLS IXのアーキテクチャの基本概念を説明する。最後に、MPLS IXに必要なルータの機能を整理する。

2.2 MPLSを用いたIX技術

本節では、我々が提案しているMPLSを用いた新

しい IX 技術（MPLS-IX）を説明する。

2.2.1 MPLS-IX の概要

MPLS-IX のアーキテクチャは、図 2.1 に示すように、MPLS コアルータ を持つ IX 提供者と MPLS エッジルータ を持つ ISP が存在する。従来の IX と同様に ISP の境界ルータ同士が IX を通して BGP のピアリングを行う。従来の IX との大きな違いは、ISP の境界ルータ同士が同一のサブネットに属さず、MPLS-IX 内に複数のルータが存在していることである。以下では、図 2.1 を用いて、MPLS-IX の概要を説明する。

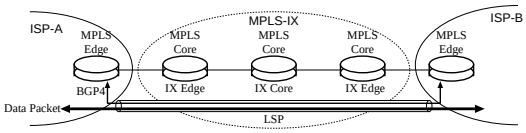

図 2.1.　MPLS-IX

MPLS-IX は、1 つ以上の MPLS コアルータで構成される。以下の説明のために MPLS コアルータを「IX エッジルータ」と「IX コアルータ」という用語で区別することにする。IX エッジルータは、MPLS-IX の境界ルータであり、MPLS エッジルータである ISP の境界ルータと接続するルータである。IX コアルータは、MPLS-IX のルータ（すなわち、MPLS コアルータ）であり IX エッジルータではないルータである。IX エッジルータには、ISP の境界ルータである MPLS エッジルータが接続する。IX エッジルータと MPLS エッジルータ間のデータリンクは、MPLS がサポートされているデータリンクメディアであればどんなデータリンクメディアでも構わない。理論的には全てのデータリンクメディアで MPLS が利用可能であるので、どんなデータリンクメディアも使うことができる。ISP-A と MPLS-IX 間のデータリンクメディアと ISP-B と MPLS-IX 間のデータリンクメディアは、異なっていても構わない。また、MPLS-IX 内部も任意のデータリンクメディアを使用することができる。この図では、簡単のために 2 つの ISP しか書いていないが一般的に MPLS-IX には、複数の ISP が接続することができる。以下の説明では、図中の 2 つの ISP の MPLS エッジルータ間のピアリングのみを説明するが、複数の ISP が接続しているときも同様にピアリングを行うことができる。

上記の様な接続状態で ISP-A の MPLS エッジルータと ISP-B の MPLS エッジルータの間でピアリングを行う。MPLS エッジルータ間は、ラベル配布プロトコルを用いて LSP を設定する。LSP が設定されるとその LSP を通して BGP による経路情報交換を行う。BGP によって経路情報が交換されるとそれにしたがって、データトラフィックが LSP 上を流れる。

ここで重要なのは、LSP で運ばれるパケットは、専用線やトンネルと同じように途中の MPLS コアルータでは IP ヘッダの宛先アドレスではなく、ラベル情報を参照して転送されるため、MPLS-IX のルータ（MPLS コアルータ）において BGP で交換される経路を持つ必要がない。また LSP は、MPLS エッジルータ間で設定できるため、MPLS-IX は介在せずに、ISP 同士でパケット転送のポリシを決めることができる。これにより現在の IX の基本ポリシであるバイラテラルの実現が可能になる。

2.2.2 MPLS-IX の動作手順

本節では、MPLS-IX の実現方法を説明する。

図 2.2 は、MPLS-IX に接続する 2 つの ISP 間に LSP を設定し、BGP4 による経路情報交換を行う手順を示している。MPLS-IX では、実際のデータを流すための LSP を設定するために以下の手順を実行する。

1. ルータ間の物理接続を行う
2. MPLS エッジルータ間で LSP を設定するための経路情報を持つ
3. MPLS エッジルータおよび MPLS コアルータで MPLS シグナリングプロトコルを動作させる
4. MPLS エッジルータ間で LSP を設定する
5. MPLS エッジルータ間の BGP4 で経路情報交換を行う

まずはじめに、MPLS コアルータ間および MPLS

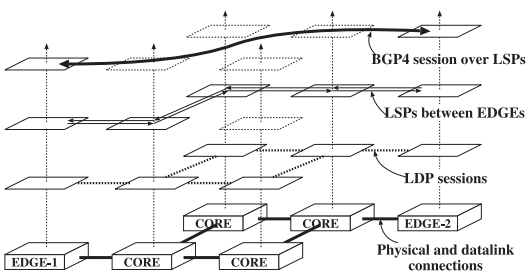

図 2.2.　MPLS-IX アーキテクチャ

コアルータと MPLS エッジルータをあるデータリンクメディアで接続する必要がある。このデータリンクメディアは、MPLS が動作するものであればどのようなデータリンクメディアでも良い。理論的には、どんなデータリンクメディアも選ぶことができる。例えば、MPLS コアルータ間は、POS やギガビットイーサネットで接続することもできるし、MPLS コアルータと MPLS エッジルータ間を ATM で接続することもできる。それぞれのルータ間は任意のデータリンクで接続することができる。

次に、MPLS エッジルータ 間で LSP を設定するために必要な IP 経路情報を持つ必要がある。MPLS コアルータ間では、OSPF や IS-IS などの動的経路制御プロトコルを動作させ、IP 経路情報交換を行う。MPLS エッジルータと MPLS コアルータ間では、基本的にスタティック経路設定を行う。これは、MPLS コアルータ をもつ IX 提供者と MPLS エッジルータを持つ ISP の管理者が異なるため、運用上それぞれの IGP 経路情報交換を避けたいためである。MPLS エッジルータでは、LSP を設定する相手の MPLS エッジルータへのスタティック経路を設定し、MPLS コアルータでは、自分が接続している MPLS エッジルータへの経路を OSPF や IS-IS によって MPLS コアルータに配布する。

次に、それぞれのルータで MPLS 機能を動作させ、MPLS のシグナリングプロトコルを起動する。現在の MPLS のシグナリングプロトコルは大きく分けて 3 つ存在する。LDP[5]、RSVP[9]、CR-LDP[74] である。これらのシグナリングプロトコルにより、MPLS エッジルータ間に LSP を設定する。これらのシグナリングプロトコルで設定される LSP は片方向なので、両方向のデータ交換ができるように 2 つの LSP を設定する。図 2.2 では、Edge-1 から Edge-2 への LSP と Edge-2 から Edge-1 への LSP の 2 つを設定することになる。

MPLS エッジルータ間に LSP を設定した後、その LSP を通して BGP4 で経路情報交換を行う。

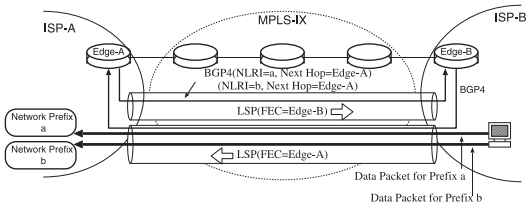

図 2.3. BGP による経路交換と LSP 上のデータ通信

図 2.3 で BGP による経路情報交換とその後のデータトラフィックの流れを説明する。ISP-A の MPLS エッジルータである Edge-A と ISP-B の MPLS エッジルータである Edge-B 間でピアリングを行う。この図では簡単のために、Edge-A から Edge-B へ BGP による経路情報通知を示している。Edge-A と Edge-B の間には、LSP が 2 本設定されている。1 本は、Edge-A から Edge-B へのデータ通信用であり、FEC は Edge-B が設定されている。もう 1 本は Edge-B から Edge-A へのデータ通信用であり、FEC は Edge-A が設定されている。現状の MPLS での LSP は、片方向通信にのみ使えるので、ある MPLS エッジルータ間では両方向通信のために 2 本の LSP が必要であるために、このような状況となる。

ISP-A には、2 つのネットワークプレフィックスがあり、それぞれ「a」と「b」である。Edge-A は、Edge-B に BGP で「ネットワークプレフィックス a のネクストホップが Edge-A」であること「ネットワークプレフィックス b のネクストホップが Edge-A」であることを通知する。この BGP パケットは、LSP (FEC=Edge-B) を通して送信される。この BGP パケットを受信した Edge-B は、ネットワークプレフィックス a および b のパケットを LSP (FEC=Edge-A) で送信できるように設定する。これは、ネットワークプレフィックス a および b のネクストホップが Edge-A であるため、Edge-A へパケットを配送できる LSP を通してデータパケットが転送される。

このように、MPLS エッジルータ間で片方向 1 つづつの LSP を設定し、BGP パケットとデータパケットの両方をその LSP で転送する。ネットワークプレフィックス毎に LSP を 1 本設定する方法もあるが、IX の様なフルルートを持つルータでプレフィックス毎に 1 本の LSP を設定するとなると莫大な LSP が必要となるため、ネクストホップに対して 1 本の LSP を設定することにより、MPLS-IX 中の LSP 数を少なく保つことができる。

2.2.3 MPLS-IX の特長

前節で説明した MPLS 技術を使った MPLS-IX は、以下のような特長を持つ。

- 様々なデータリンクメディアで IX に接続可能
  従来の IX では、1 つのデータリンクメディアを使用している。そこで、IX に接続するために

は、指定されたデータリンクメディアを使う必要があった。しかし、MPLS は、ATM、POS、GbE など様々なデータリンクメディアを扱うことができる特長を持つ。このため、MPLS を利用した MPLS-IX では、様々なデータリンクメディアによって IX に接続することができる。

- 異なるデータリンクメディアで接続している ISP 同士でピアリング可能
  MPLS では、異なるデータリンクメディア間で LSP を作成することができる。この LSP を用いて BGP およびデータパケットを交換することにより、異なるデータリンクメディアで接続している ISP 同士で BGP によるピアリングが可能になる。
- 従来の IX との接続性がある（マイグレーションが容易）
  MPLS-IX では、異なるデータリンクメディアで接続することが可能である。このため、従来の IX と MPLS-IX を接続し、従来 IX に接続している ISP が MPLS 対応ルータを準備すれば、MPLS-IX に接続している ISP と従来の IX に接続している ISP 間でピアリングを行うことが可能である。
- 高速なデータリンクでの接続が可能
  MPLS-IX では、様々なデータリンクメディアを使えるために、POS やギガビットイーサネット等の高速なデータリンクメディアを使うことが可能である。また、IETF MPLS WG では、高速なデータリンクメディアである DWDM 等に MPLS を適用する方式が検討されているため、今後の高速性が期待できる。
- 広域分散が可能
  MPLS-IX では、データリンクメディアを選ばないため、ATM や POS のように広域で使うように設計されたデータリンクメディアを使うことができる。このデータリンクメディアを MPLS コアルータ間で使用することにより、MPLS-IX のバックボーンを全世界に拡張することも可能である。また、ISP と IX の接続を広域で使えるデータリンクメディアにすることで、IX に接続する ISP のルータを IX が存在するコロケーションに置く必要がなくなる。

2.2.4 MPLS IX に対する要求条件

本節では、MPLS IX に必要な要求条件をまとめる。MPLS IX では、図 2.4 の様に MPLS IX の提供者とその MPLS-IX を使用する利用者（例えば、ISP）に分類される。MPLS IX 提供者とその利用者は、それぞれ別のネットワーク管理ポリシで運用されることが一般的であるため、MPLS IX 提供者の境界ルータと MPLS IX 利用者の境界ルータとの間では、できるだけお互いに影響を与えないような構成にする必要がある。また、運用管理を容易にする必要がある。上記の要求条件は、以下のようにまとめられる。

- MPLS IX 提供者と利用者の間では共通の IGP (Interior Gateway Protocol) を使わない。
  MPLS IX 提供者と利用者の間で共通の IGP を利用すると、相手のネットワーク内の経路変更が自分のネットワーク中に伝わることになり、相手の経路変更の影響を受け、ルータの経路計算の負荷が大きくなったり、経路トラブルを発生する可能性がある。そこで、MPLS IX 提供者と利用者の間で共通の IGP を使わないようにすることで、上記の問題を回避する。図 2.4 では、MPLS IX 提供者内ネットワークの経路制御を OSPF とし、MPLS IX 提供者の境界ルータ（図 2.4 Core-B、Core-D）と利用者の境界ルータ（図 2.4 Edge-A、Edge-E）では、スタティック経路を使うことにより、同一の IGP を使わない例を示している。
- MPLS IX の境界ルータで LSP のフィルタを設定できるようにする
  LDP を MPLS のシグナリングとして用いる MPLS IX 利用者は、自分のアドレスを FEC

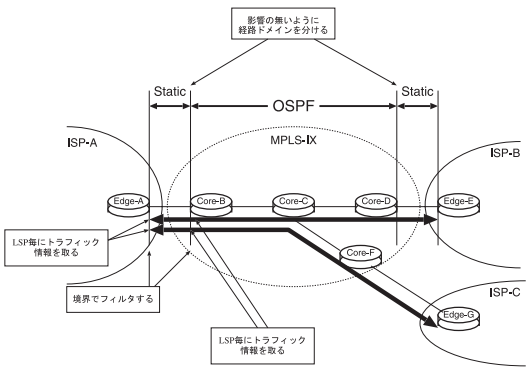

図 2.4.　MPLS-IX アーキテクチャ

としてラベル割当メッセージを MPLS 提供者の境界ルータに送信する。ところが、設定ミスや悪意のある利用者からは、利用者のアドレス以外を FEC としてラベル割当メッセージが送信される可能性がある。もし、このメッセージを MPLS IX 提供者が受信してしまうと、正常ではない LSP が設定されることになり、パケットが期待していないところに転送される可能性がある。このような MPLS IX 利用者の設定ミスによる影響を、他の利用者に与えないようにするため、自分のアドレス以外の FEC を受理しないように MPLS IX 提供者の境界ルータに LSP のフィルタを設定する必要がある。

- LSP 毎のトラフィック統計情報を取得可能にする。

  IX の利用者である ISP 等は、どの ISP 間でトラフィックが交換されているのかを知ることによりネットワーク設計を容易にしたり、課金のための情報を得ることができる。MPLS IX では、相手先毎に LSP を設定するため、この LSP 毎にトラフィック情報を取得すれば、この目的を達成できる。そこで、LSP 毎にトラフィック情報を取得する必要がある。

2.3 MPLS IX に必要なルータの機能

本節では、前節で説明した MPLS IX のアーキテクチャと MPLS IX の要求条件を満たすための MPLS IX で使うルータに必要な機能を整理する。

2.3.1 コアルータの機能

本節では、MPLS IX 提供者のルータで必要な機能を示す。

- エッジルータへの経路を MPLS IX 内部の IGP（例えば OSPF）で MPLS IX 内部のルータに伝搬できること。

  RSVP の場合は、エッジルータとコアルータ間のネットワークの経路情報あるいはエッジルータのループバックインターフェイスを伝搬できれば良い。図 2.5 では、Edge-A のループバックアドレスを A、Edge-A のインターフェイスアドレスを X.A とする。ここで、X は IP アドレスのネットワーク部を示し、A は IP アドレスのホスト部を示すとする。この場合には、Core-B は、MPLS IX 中にネットワークプレフィックス X と Edge-A のループバックインターフェイス A を伝搬できれば良い。

  シグナリングプロトコルに LDP を利用する場合は、エッジルータのインターフェイスアドレスかエッジルータのループバックアドレスを伝搬する必要がある。例えば、図 2.5 では、Edge-A のループバックアドレス A と Edge-A のインターフェイスアドレス X.A を伝搬させないといけない。

  RSVP と LDP とも LSP の FEC は、エッジルータのループバックアドレスかインターフェイスアドレスである。しかし、LDP の場合は、プロトコル仕様上、FEC と全く同じ経路情報をコアルータで持つ必要があるため、FEC をインターフェイスアドレスとする場合にインターフェイスアドレスと全く同じ経路情報を持つ必要があるため、インターフェイスアドレスを経路プロトコルで広告する必要がある。RSVP の場合は、宛先まで到達できる経路があれば良いので、インターフェイスアドレスではなく、そのインターフェイスが属しているネットワークプレフィックスを広告すれば十分である。

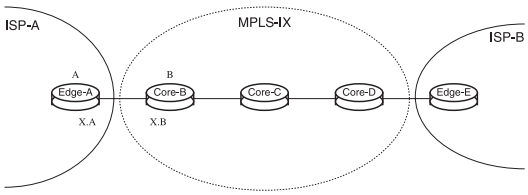

図 2.5. MPLS-IX ネットワーク構成

- LSP のフィルタができること。

  MPLS のシグナリングプロトコルである LDP や RSVP を MPLS エッジルータから受信したときに拒否することができるようなフィルタを設定できる必要がある。例えば、特定の FEC のみを受理するように設定したり、RSVP Path メッセージの送信 IP アドレスにより拒否できるように設定する。

- LSP 毎の統計情報が取れること。

  MPLS IX の入口および出口の境界ルータで LSP につながれるトラフィック統計情報を取得できる必要がある。これは、課金やパケットロスを監視するため等に使う。

2.3.2 エッジルータの機能

この節では、MPLS IX 利用者であるエッジルータで必要な機能を示す。以下に必要な機能を示す。

- スタティック経路で設定した相手エッジルータへのアドレスに対して LSP を設定できること。
  要求条件で示したように、エッジルータとコアルータの間では、同一 IGP を動かさない。そこで、エッジルータでは、スタティック経路等で相手先エッジルータへの LSP を設定できる必要がある。一般的な MPLS の使い方では、エッジルータとコアルータ間で OSPF を用いることが想定されている。MPLS IX での使い方は一般の方法と異なるので、注意が必要がある。
- BGP パケットが LSP に流れること。
  BGP のピアアドレスは、上記で設定した LSP の FEC アドレスである。この BGP のパケットを LSP に流す必要がある。これは、LSP が切断されたときに、LSP を通るデータパケットを迂回させるために必要である。この機能が実現されない場合は、BGP による経路情報は交換できるが、実際のデータパケットを交換する LSP が切れているため、データパケットがすべて廃棄されてしまうことになる。
- BGP で受信した経路にマッチするデータパケットが LSP に流れること。
  この機能は、MPLS IX の基本動作として必要なものである。BGP で受信した経路にマッチするデータパケットを LSP に流すことにより、コアルータで IP を見る必要がなくなるので、コアルータにフルルートを持つ必要がなくなり、エッジルータ間でバイラテラルポリシーでパケット交換ができる。
- IP TTL を MPLS TTL にコピーしない設定ができること。
  BGP パケットが MPLS IX 外部を経由して相手エッジルータに到達しないように IP TTL を 1 として送出する。通常、エッジルータでは、IP TTL を MPLS TTL にコピーして送信するが、この動作を行うと、コアルータでパケットが廃棄されてしまう。そこで、IP TTL を 1 としながら、コアルータでパケットを廃棄せずに転送できるために、エッジルータで挿入する MPLS TTL を IP TTL からコピーせずに設定値（例えば、最大値の 255）で送信できるように設定できる必要がある。また、出口のエッジルータでも通常の動作で MPLS TTL を IP TTL にコピーするようになっているが、これも同様に MPLS TTL を IP TTL にコピーしないように設定できる必要がある。
- LSP のフィルタができること。
  MPLS のシグナリングプロトコルである LDP や RSVP をコアルータから受信したときに拒否することができるようなフィルタを設定できる必要がある。例えば、特定の FEC のみを受理するように設定したり、RSVP Path メッセージの送信 IP アドレスにより拒否できるように設定する。
- LSP 毎の統計情報が取れること。PHP 無のラベル割当ができること。
  エッジルータでの LSP 毎の統計情報をとれる必要がある。送信側エッジルータでは、必ずラベルが割り当てられているので、そのラベルのトラフィック情報がとれれば良いが、受信側エッジルータでは、PHP[128] を用いるとラベルがエッジルータの前段コアルータで削除されてしまうため、受信側エッジルータでトラフィック統計情報をとることができない。そこで、LSP 毎の統計情報をとるためにも PHP 無のラベル割当ができる必要がある。

第 3 章　MPLS 実装 AYAME

あやめプロジェクトでは、1999 年から Sub-IP 技術および MPLS 技術に関連したネットワーク研究開発環境の構築を目的として、MPLS スタックを独自に実装している。

本章では、あやめプロジェクトにおいて 2001 年度に実施した

- AYAME MPLS 実装のインターオペラビリティ試験
- MPLS を用いた IPv6 伝送実験

に関して報告する。

3.1 AYAME MPLS 実装の相互接続性実証実験

インターネットのプロトコル実装を作成する上で、同一の仕様に則った他実装との相互接続性（インターオペラビリティ）を提供、維持することは非常に重要である。

あやめプロジェクトでは、MPLS 技術による次世代 IX 技術の検証、開発を行っている “次世代 IX 研究会” に参加している。次世代 IX 研究会には国内外の MPLS ベンダも多く参加しており、各ベンダ間の接続検証を行うために “ルータ分科会” が下部組織として存在する。MPLS 実装は各ベンダとも現時点で開発途上の部分も多く、機能の向上や動作の変更がおこわれることが多いため、ルータ分科会では、定期的に相互接続実験を行っている。

本節では、次世代 IX 研究会のルータ分科会によって開催された、MPLS 実装相互接続実験の結果を報告する。

3.1.1 次世代 IX 研究会 MPLS 相互接続実験概要

次世代 IX 研究会では、主に研究会が構築・運用している MPLS 広域 IX への接続機器間の動作検証を目的として、以下の 2 回の相互接続実験を開催した。

第 1 回

2001/10/15〜2001/10/19。通信放送機構 (TAO) 大手町 IPv6 システム運用技術開発センター。

第 2 回

2002/1/28〜2002/2/1。通信放送機構 (TAO) 幕張ギガビットリサーチセンター。

第 1 回、第 2 回を通してのべ 12 ベンダ、17 実装の検証が行われた。相互接続実験に参加したベンダおよび実装を表 3.1 に示す。

相互接続性の検証実験は

- MPLS-IX で要求する仕様機能を充足できること
- MPLS LSR としての動作が正しく、異種混合環境下で相互接続性が維持できること。

を目的としている。それぞれの検証項目毎に概略を説明する。

3.1.2 検証項目 (MPLS-IX)

MPLS-IX における要求仕様は MPLS-IX のアーキテクチャに依存している。図 3.1 に MPLS-IX が想定している、接続アーキテクチャおよびシグナリングフローを示す。

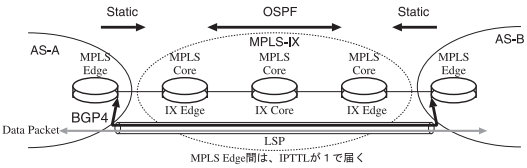

図 3.1. MPLS-IX のアーキテクチャおよびシグナリングフロー

MPLS-IX では、IX に接続した AS 間の経路交換を BGP で行い、交換された経路に対応する LSP を確立するアーキテクチャで動作する。そのため、MPLS-IX に接続する MPLS LSR は以下の機能を提供する必要がある。

- Multihop BGP を用いて IX をはさんだピア間で BGP セッションを確立し、経路を交換する

表 3.1. MPLS 相互接続実験参加ベンダ

ベンダ名	第 1 回	第 2 回
Cisco	C7200	C12406/C7200
Extream	(N/A)	BlackDiamond6808
Juniper	M20/M5	←
Unisphere	(N/A)	ERX-700
DML (IP-infusion)	Zebos	←
NEC	IX5010	IX5005
日立	GR2000-6H	←
Foundry	NetIron400	←
富士通	Geostream R940	←
古川電工	FITELnet-G10	FITELnet-G12
Riverstone	RS8000	RS8000/RS3000
あやめプロジェクト	AYAME	←

- BGP セッションは MPLS LSP の上で確立する。その際の TTL は 1 とする。
- BGP4 で交換した経路に対応する FEC を生成し、その FEC に対して LSP を確立する

相互接続検証実験では、これらの要求仕様に関して、各ベンダ間の実装で相互接続性があること検証した。

3.1.3 検証項目 (MPLS 汎用)

相互接続検証実験では、MPLS-IX に特有な機能だけでは無く一般的な MPLS 実装の相互接続性の検証も行っている。

第 1 回の相互接続検証実験では主に

- エッジ LSR としての動作検証

に主眼をおいた検証をおこなった。

2 台の LSR を接続したコアネットワークに対して各ベンダの実装を 2 台づつ接続したうえで、接続可能な全組合わせに関して以下の項目を検証した。

- MPLS シグナリングプロトコルの相互接続性の検証
- MPLS 転送系の相互接続性の検証
- MPLS 冗長機構の動作検証
- MPLS LSR の第 3 層経路制御プロトコルと MPLS シグナリングプロトコルの協調動作に関する検証

現時点では仕様化されている MPLS シグナリングプロトコルは

- LDP (RFC3036)[5]
- RSVP (RFC3209)[9]
- CR-LDP (RFC3212)[74]

の 3 種類が存在する。第 1 回、第 2 回ともに、LDP および RSVP を対象とした相互接続実験を行った。

AYAME では CR-LDP も実装を進めているが、

- 実装しているベンダ少ないこと
- コアネットワーク用のルータで対応していないこと

から実験は行わなかった。

第 2 回実験ではこれらに加えて以下の検証も行った。

- Core LSR としての MPLS 基本動作の検証
- Core LSR としての MPLS 冗長機構の検証

それぞれの検証項目の詳細を表 3.2 に示す。AYAME MPLS では RSVP をサポートしていないため、RSVP 関連の検証項目は割愛した。

3.1.4 相互接続実験結果

第 1 回および第 2 回の相互接続実験の結果のうち、AYAME MPLS スタックに関連する部分を報告する。

MPLS シグナリングプロトコルである LDP は複数の動作モードが仕様化されている。各ベンダの LDP プロトコルの動作モードを表 3.3 に示す。

相互接続実験のトポロジを図 3.2 および図 3.3 に示す。前者は LSR Edge 試験時のトポロジであり、後者は LSR Core 試験時のトポロジである。

AYAME を Edge LSR として Juniper (M20) 2 台によって構成される MPLS Core ネットワークに接続した際の、Juniper との LDP セッションの相互接続性の検証結果を表 3.4 に示す。また、対向の Edge LSR と接続した際の相互接続性の検証結果を表 3.5 に示す。

更に、AYAME を Core LSR として他実装との相互接続性を検証した結果を表 3.6 に示す。

3.1.5 性能計測

第 2 回実験では、MPLS LSR の相互接続性の試験に加えて

- MPLS LSR の転送性能計測、高負荷時動作検証

を行った。

転送性能計測および高負荷動作検証はアジレント社の RouterTester を用いた。検証項目を以下に示す。

- パケット長毎の最大スループット/パケットロス率/転送遅延の計測
- IP の経路表の大きさに関する性能相関の計測

性能測定では MPLS LSR の転送機構の性質を考慮し、

- あらかじめ複数のラベル束縛を与える
- テストパケットの FEC が分散するようにトラフィックを生成

してある。このようにすることで、LSR 内部の転送機構の挙動を一般のバックボーン部分のルータに近似した状態におけると考えられる。

スループットはパケット長 64 バイト、512 バイト、1500 バイトのそれぞれに関して、パケットロス率が 0 になる状態の最高値を計測した。

3.1.6 性能計測結果

AYAME の性能計測結果を表 3.7 および表 3.8 に

示す。

AYAME は PC アーキテクチャで動作する LSR であるので、測定に利用された PC の仕様の概略を以下に示す。

CPU

AMD Athlon-MP 1.2GHz

Memory

DDR SDRAM 512MB

Chipset

AMD-760（Tiger MP:64bit PCI）

GbE

Nortel GA620（alteon チップ）∗2

OS

NetBSD-1.5.2/AYAME0.3

表 3.7 はラベルスイッチングした際の性能を、パケット長毎に

- 最大パケット転送速度
- パケット長 × パケット数

表 3.2. MPLS LSR の検証項目

項目	MPLS エッジ	MPLS コア
LDP セッション確立		
（Transport Address=インターフェイスアドレス）	○	○
（Transport Address=ループバックアドレス）	○	○
LDP RouterID をインターフェイスアドレスに設定	○	○
OSPF 設定		
（MPLS エッジルータのインターフェイスアドレス/32 を広告）	–	○
（MPLS エッジルータのループバックアドレス/32 を広告）	–	○
LSP 設定		
（FEC=MPLS エッジルータのインターフェイスアドレス）	○	○
（FEC=MPLS エッジルータのループバックアドレス）	○	○
LSP 設定（PHP なし）	○	○
Conservative Mode のルータからの LSP 設定	○	○
Multihop BGP		
（Transport Address=インターフェイスアドレス）	○	○
（Transport Address=ループバックアドレス）	○	○
LSP でのデータ転送		
（BGP4 パケットの転送）	○	○
（BGP4 で受信した経路のパケット）	○	○
入口ルータで IP TTL を MPLS TTL にコピーしない	○	–
出口ルータで MPLS TTL を IP TTL にコピーしない[1]	○	–

1 PHP ありの場合には要コアルータでの試験

表 3.3. 各 LSR 実装の LDP モード

ベンダ	コンサバティブ/リベラル	DU/DOD	Ordered/independent
AYAME	リベラル	DU	Ind
Cisco	リベラル	DU	Ind
Juniper	リベラル	DU	Ord
Hitachi	リベラル	両方（試験は DU）	Ord
Furukawa	コンサバティブ	両方（試験は DU）	Ord（DOD 時）/Ind（DU 時）
Riverstone	リベラル	DU	Ord
Fujitsu	リベラル	DU	Ord
DML	両方（試験はリベラル）	両方（試験は DU）	両方（試験は Ind）

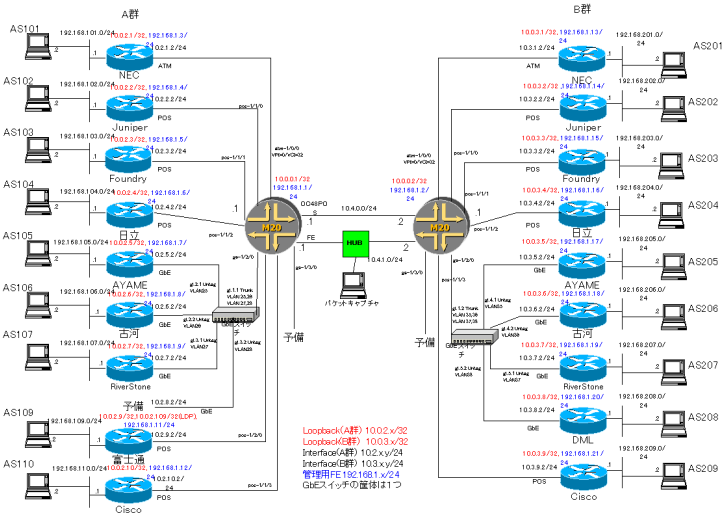

図 3.2.　相互接続実験トポロジ（LSR edge 検証）

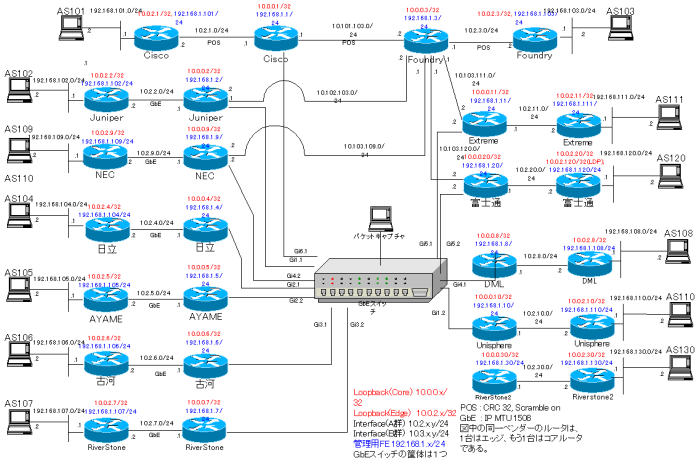

図 3.3.　相互接続実験トポロジ（LSR core 検証）

表 3.4.　相互接続実験の結果（LDP セッション）

CORE ルータとの LDP セッション	
LDP セッションのトランスポートアドレスをループバックアドレスにできる	(1)
LDP セッションのトランスポートアドレスをインターフェイスアドレスにできる	○
Passive/Active の両方のモードでの隣接確立ができる	○
LDP RouterID をループバックアドレスにできる	○
LDP RouterID をインターフェイスアドレスにできる	○

Edge-Core 間のリンク切断時処理	
Edge-Core 間のリンク切断時に LSP が消される	○
Edge-Core 間のリンク切断時の LDP セッションか切断される	○
Edge-Core 間のリンク再接続時に LDP セッションおよび LSP が復活する	○

(注)　(1) src address はインターフェイスアドレス

表 3.5. 相互接続実験の結果（AYAME: Edge LSR）

項目	あやめ	cisco	Juniper	zebos	日立	富士通	古川電工	Riverstone
LSP 設定								
FEC を Edge ルータのループバックアドレスとする LSP が設定できる	○	○	○	○	○	○	○	○
FEC を Edge ルータのインターフェイスアドレスとする LSP が設定できる	○	○	N/A	○	○	N/A	○	N/A
PHP（Implicit Null ラベル）を使わない LSP	○	○	○	○	○	○	○	○
BGP ピアの確立								
IP TTL を 1（PHP） or 2（非 PHP）で送出する	○	○	○	○	○	○	○	○
IP TTL を大きくしても BGP ピアが確立できる	○	○	○	○	○	○	○	○
上記設定での BGP 経路情報交換	○	○	○	○	○	○	○	○
LSP でのデータ転送								
BGP で受信した経路情報に対応するデータパケットを LSP で転送できる	○	○	○	○	○	○	○	○
LSP を介した BGP パケット転送	○	○	○	○	○	○	○	○
MPLS TTL と IP TTL								
入口ルータで IP TTL を MPLS TTL にコピーしない設定ができる	○	○	○	○	○	○	○	○
出口ルータで MPLS TTL を IP TTL にコピーしない設定ができる	○	○	○	○	○	○	○	○

(注) 1: Cisco は IP TTL の最低値が 2 なので 2 で試験

- 最大スループット

を計測した。参考までに IP 転送した場合の計測結果も併記してある。

表 3.8 は L3 の経路数すなわち LSR 内の FEC 数と転送性能の相関を計測した。BGP で一定数の経路を注入した際の MPLS 転送に対して、以下の要素を計測している。

- 最大スループット（Mbps および pps）
- インターフェイスの上限帯域に対する転送効率
- 転送遅延

3.1.7 まとめ

2 回にわたる相互接続性検証テストによって、あやめプロジェクトで作成している MPLS 実装は多くの実装との相互接続性があることが検証された。また、ソフトウェア実装としては性能も十分であり、安定した動作をすることが確認された。

AYAME は次世代 IX 研究会が構成している MPLS 網に実装に投入され、運用に関する経験も蓄積されつつある。今後も更なる性能/機能の改善を行っていく予定である。

3.2 MPLS を用いた IPv6 伝送実験

DISTIX アーキテクチャにもとづいた MPLS 網における IPv6 伝送実験を北陸先端科学技術大学/あやめプロジェクトと北陸通信ネットワーク株式会社（HTCN）の共同で行った。本実験においては、HTCN のネットワーク内に MPLS 網を構築したうえで、この網に対してあやめプロジェクトの LSR 実装をエッジルータとして接続した。この実験に用いるためあやめプロジェクトでは、IPv6 伝送に関する AYAME LSR の機能拡張を行った。拡張の詳細については後述する。

3.2.1 実験 MPLS ネットワーク

本実験のトポロジを図 3.4 に示す。

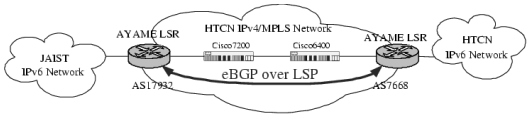

図 3.4. AYAME（IPv6）実験トポロジ

図 3.4 の MPLS 網では、コアルータ側では IPv6 のスタックは含まれておらず、MPLS 網の制御は

表 3.6.　相互接続実験の結果（AYAME: CORE LSR）

検証項目	cisco	古川電工	日立	juniper	unisphere
LDP セッション					
ループバックアドレスをトランスポートアドレスとした LDP セッションの確立	○	○	○	○	○
PHP 時、シグナリングで ImplicitNULL を受けられる	N/A	N/A	N/A	N/A	○
非 PHP 時、シグナリングで ExplicitNULL もしくはラベルを受けらる	○	○	○	○	N/A
OSPF 設定					
MPLS エッジルータのループバックアドレス/32 を広告できる	○	○	○	○	○
LSP 設定					
FEC を Edge ルータのループバックアドレスとするエッジ間の LSP が設定できる	○	○	○	○	○
PHP（ImplicitNull ラベル）を使わない LSP	○	○	○	○	○
コア間 LSP からパケット取りだし Edge LSR に送信する際の IP TTL の扱い	(1)	(1)	(1)	(1)	(1)
FEC のフィルタを設定できる	×	×	×	×	×
BGP ピアの確立					
BGP ピアアドレスがループバックアドレスでピア先 Edge LSR と BGP ピアを確立できる	○	○	○	○	○
IPTTL を 1（PHP なし）or 2（PHP）で送出できる	○	○	(2)	○	○
上記設定での BGP 経路情報交換	○	○	○	○	○
LSP でのデータ転送					
BGP で受信した経路情報に対応するデータパケットを LSP で転送できる	○	○	○	○	○
BGP のパケットが LSP で転送できる	○	○	○	○	○

(注) (1) MPLS TTL はコピーせず IP TTL をそのまま送信
(2) IP TTL を 2 にしてピア確立

IP 経路制御

OSPF

MPLS シグナリング

LDP

が利用されている。

実験においては、DISTIX アーキテクチャに基づき、エッジルータ間に LSP を構成し、LSP 上で eBGP を用いた経路交換とトラフィック交換を行った。

3.2.2 実験における IPv6 経路制御

本実験では、IPv6 経路の交換を

- BGP ルータ間で IPv4 により eBGP セッションを確立したうえで、MP-BGP (Multiprotocol Extensions for BGP-4)[18] を用いる
- 広報する IPv6 経路情報の BGP next hop 情報には eBGP セッションの IPv4 アドレスを IPv4-mapped IPv6 address を用いて表現したものを指定する

方式を採用した。

この方式は、“Connecting IPv6 Islands across IPv4 Clouds with BGP” (draft-ietf-ngtrans-bgp-tunnel-04.txt) における MP-BGP over IPv4 アプローチ（Tunneling over MPLS LSPs）の DISTIX アーキテクチャへの適用である。

本方式の経路制御およびデータ転送の構成を図 3.5 および図 3.6 に示す。図 3.5 は DISTIX アーキテクチャにおける IPv4 経路制御の際の構成を示しており、図 3.5 が、その上での IPv6 の経路の扱いを示している。これらの図の関係を見てわかるように、これらのアーキテクチャはエッジルータ間における交換経路の形式および各エッジルータの保持経路の状態といった点で親和性が極めて高いのが特徴である。

そのため、本方式では、

- 経路制御用の eBGP パスと実際のデータパスが同一の LSP を用いる
- 何らかの原因により LSP が切断された場合、

表 3.7. L3/L2.5 の転送パフォーマンス測定

パケット長	最大パケット数/秒	パケット長 × パケット数	最大スループット（MB/s）
IP packet（L3 転送時）			
64	1201923	76923072	615.38
256	422297	108108032	864.86
512	226449	115941888	927.54
1024	117481	120300544	962.41
1496	81380	121744480	973.96
MPLS packet（L2.5 転送時）			
68	1157416	78704288	592.60
260	416670	108334200	853.34
516	224822	116008152	920.87
1028	117042	120319176	958.81
1500	81170	121755000	971.44

表 3.8. L3 経路数による転送パフォーマンスの変化の計測

パケットサイズ（byte）	最大スループット（Mbps）	最大スループット（pps）	効率（%）	遅延（μs）
BGP 注入経路=1000				
64	53	103515.625	8.94	1347
256	214	104492.188	25.08	1468.57
512	430	104980.469	46.69	1347
1024	695	84838.8672	72.49	1257
1496	765	63920.4545	78.75	1713
BGP 注入経路=10000				
64	53	103515.625	8.94	1360
256-1024		–		
1496	765	63920.4545	78.75	1690
BGP 注入経路=100000				
64	53	103515.625	8.94	255
256-1496		–		
BGP 注入経路=200000				
64	53	103515.625	8.94	260
512	420	102539.063	45.61	–
1024-1496		–		

BGP セッション断が発生する

など、DISTIX アーキテクチャにおいて重視されている点において IPv4 の伝送場合と同等の効果が得られる。

3.2.3 AYAME/IPv6 実装と設定例

AYAME MPLS スタックに対して本方式を実装するために以下の変更を行った。

- FEC クラシファイアに対する拡張
- 経路制御ソフトウェアに対する拡張

後者は IPv6 経路制御機構によって交換された IPv6 プレフィックスに対する FEC を認識して LSP へ投入するために必要である。AYAME は複数のプロトコルに対するサポートを想定した設計となっているため、変更は小規模であった。

AYAME では L3 の経路制御ソフトウェア機構とし

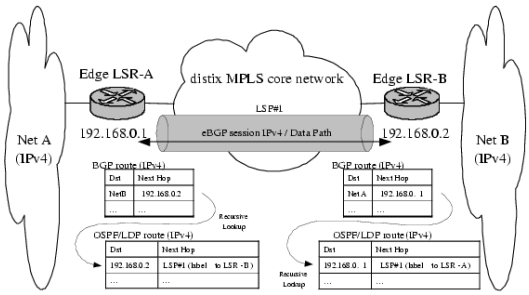

図 3.5.　IPv4-mapped IPv6 address を用いた eBGP での経路交換（IPv4 部分）

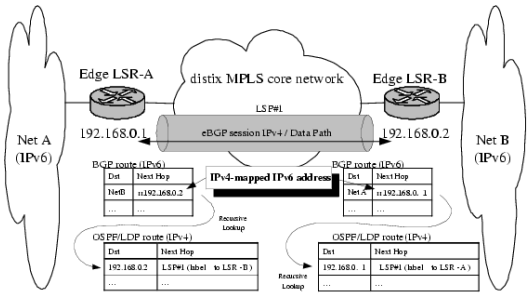

図 3.6.　IPv4-mapped IPv6 address を用いた eBGP での経路交換（IPv6 部分）

て、zebra を採用している。zebra に対しては DISTIX モデルによる経路交換をサポートするために以下に示す拡張を行ってある。

- LDP と IGP（static route を含む）の連携機構の導入
- BGP での TTL 1 ebgp multihop のサポート
- BGP での recursive lookup の際に LSP を利用可能とする拡張

その上で、IPv6 サポート用の拡張として新たに以下の拡張をおこなった。

- IPv4-mapped IPv6 address の取り扱いのための拡張
- IPv4-mapped IPv6 address に対応した、BGP での recursive lookup の機構の拡張

AYAME における IPv6 経路交換のための BGP 設定を図 3.7 に示す（本設定は今後変更される可能性がある）。DISTIX モデルによる接続を介して BGP 接続による IPv6 経路交換を行うには以下の点に留意しなければならない。

- MPLS 網の先の eBGP peer に経路を広告する際には、nexthop に MPLS 網で用いる自 IPv4 アドレス（IPv4-mapped IPv6 address）を指定する。
- MPLS 網の先の eBGP peer から受けた経路を iBGP にて内部 IPv6 ネットワークへ広報する際には、nexthop-self を指定する必要がある。

3.2.4 実験経過と今後

本実験において確立した、IPv6 ネットワークを MPLS 網を介して接続する実験ネットワークは、現在、実装の安定性の検証を兼ね定常運用ネットワー

```
bgp router-id 211.120.192.2
 neighbor 2001:308:5::2 description ** iBGP peer (in local network)  **
 neighbor 2001:308:5::2 remote-as 17932
 neighbor 211.120.192.4 description ** eBGP peer across MPLS network **
 neighbor 211.120.192.4 remote-as 7668
 neighbor 211.120.192.4 ebgp-multihop 1
 neighbor 211.120.192.4 update-source 211.120.192.2
!
 address-family ipv6
  network 2001:308:5::/48
  neighbor 2001:308:5::2 activate
  neighbor 2001:308:5::2 nexthop-self
  neighbor 211.120.192.4 activate
  neighbor 211.120.192.4 route-map set-nexthop-myv4 out
 exit-address-family
!
ipv6 access-list v6-all permit any
!
route-map set-nexthop-myv4 permit 10
 match ipv6 address v6-all
 set ipv6 next-hop global ::d378:c002  (※注 <=::211.120.192.2)
!
```

図 3.7.　AYAME における bgpd の設定例

クとして稼働中であり、概ね安定動作している。

次年度以降、分散 IX 研究会をはじめとして MPLS 網における IPv6 の展開には大きな期待が寄せられている。そこで、今回実験した方式のより広域での実験、および、他の実現方式との比較検討していく予定である。

3.3 あやめプロジェクトの活動と今後

あやめプロジェクトでは今後更に研究活動を継続していく予定である。以下にいくつかの重点トピックに関して説明する。

3.3.1 MPLS IPv6

MPLS 技術と IPv6 技術は転送機構という点では直行した概念ではあるが、LSR 実装、アドレッシングアーキテクチャなどに関して検討しなければならない点は多い。

前章では IPv4 で運用されている MPLS 網の外部に IPv6 網を接合するモデルに関する試験実装に関して説明した。今後は、

- それ以外のアーキテクチャに関する考察、実装
- MPLS 制御プロトコル群の IPv6 化に関する考察、実装

を行っていく予定である。

3.3.2 MPLS Multicast

MPLS ではいまだマルチキャスト配送に関する仕様は明確には決定されていない。MPLS におけるマルチキャスト配送技術の実現はあやめプロジェクトの研究課題の一つであり、マルチキャスト配送対応の LSR の開発や MPLS に対応したマルチキャスト配送木の構成機構の研究などを行っている。

あやめプロジェクトでは、双方向木型マルチキャスト技術にパケット配送機能と制御機能の分離パラダイムを適用することで制御上の問題に起因する制約を軽減できる可能性を持つ双方向木型マルチキャストのアーキテクチャモデルと、MPLS を利用したそのインプリメントとして BLAST-CAST: A Bi-directional Label Switched Multicasting を提案し、議論を進めてきた。

今後は、BLAST-CAST のアーキテクチャモデルとそのインプリメント、あやめをベースとした BLAST-CAST のソフトウェア実装などに関する議論および実装を進めていく予定である。

3.3.3 汎用 LSP

MPLS は LSP をベースとした転送を行う機構であり、上位プロトコルとは独立した操作が可能である。しかし、既存の実装では一般に IP プロトコルの伝送を前提としているものが多く[1]、非 IP トラフィックの扱いはまだ発展途上である。そこであやめプロジェクトでは、非 IP トラフィックの伝送を想定とした場合を考慮し、非 IP トラッフィックを含めた汎用的に LSP を扱うための機構の研究、開発を行っている。

3.3.4 ハードウェア転送機構

MPLS 実装 AYAME は BSD 系ネットワークスタック（NetBSD）の拡張としてラベル配送機構を実現している。ソフトウェアで転送を実現するアーキテクチャは柔軟性が極めて高い一方で、一般に性能が低いという問題がある。一方、ハードウェア転送は高速ではあるが、改変性に欠るため研究用プラットホームとしては不適切である。

これらの問題に対して、プログラム可能なネットワーク用のハードウェアデバイスとして最近入手可能となったネットワークプロセッサを用いる方法が考えられる。あやめプロジェクトでは、いくつかのネットワークプロセッサを用いた BSD スタックの高速化手法を検討している最中である。

[1] EoMPLS: Ether over MPLS などのアクティビティは存在する