

第VII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7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 MPLS 実装 AYAME

あやめプロジェクト [121] 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。

本節では、あやめプロジェクトにおいて 2002 年度に実施した

- MPLS を用いた IPv6 伝送実験と 6PE 対応
 - MPLS のマルチキャスト対応実装と応用
- に関して報告する。

1.1 AYAME MPLS 実装の IPv6 対応実装と運用

既存の MPLS 網に対し、エッジルータを拡張することで IPv6 トラフィックを扱うことを実現した。この手法では、拡張対象が IPv6 を必要とする利用者が収容されているエッジルータに限られるため、既存の IPv4 を対象とした MPLS 網に容易に導入することが可能である。特に、MPLS-IX においては、エッジルータは IX 利用者側に置かれるため、IPv6 を必要とする IX 利用者側の拡張のみで対応可能な本手法の持つ意味は大きい。ここでは、MPLS 網においてどのように IPv6 を扱うことができるかを議論した上で、MPLS 実装 AYAME への IPv6 対応拡張の組み込みについて述べる。さらに、本実装を用いて実運用環境上で行った運用実験に関して述べる。

1.1.1 MPLS-IX と IPv6

インターネット・エクスチェンジ (IX: Internet eXchange) を MPLS を用いて構築する手法 (MPLS-IX) が提案されている。

一般的な MPLS 網の構築形態は、特定ネットワークを MPLS 化した上で、エッジ LSR に利用者回線を収容する形態である。しかし、MPLS-IX においては、IX の利用者側に MPLS 網のエッジルータを配置するアーキテクチャとなっている。これは、

MPLS-IX アーキテクチャの特徴的な点である。つまり、MPLS-IX においては、MPLS 網は、

- IX 提供者の設置するコアルータ群
 - IX 利用者の設置するエッジルータ
- から成る。

このような網において、LDP/RSVP-TE/CR-LDP などのシグナリングを用い、トラフィック交換を行う IX 利用者間に LSP を作成する。その LSP 上で eBGP [165] を用いた経路交換を行い、それに基づいて LSP を介したトラフィック交換を行う (図 1.1、図 1.2)。

MPLS-IX の特徴として、以下の点が挙げられる。

- IX 提供者側ルータ (コアルータ) は、MPLS-IX 外の経路を必要としない。
- IX 利用者を収容するデータリンクメディア、さらには、コアルータ間に用いるデータリンクメディアに制約がない。

1.1.2 MPLS 網での IPv6

MPLS 網における IPv6 サポートには、いくつかの手法・段階がある。どの手法を選択するかは、網を構成する LSR の機能に依る。

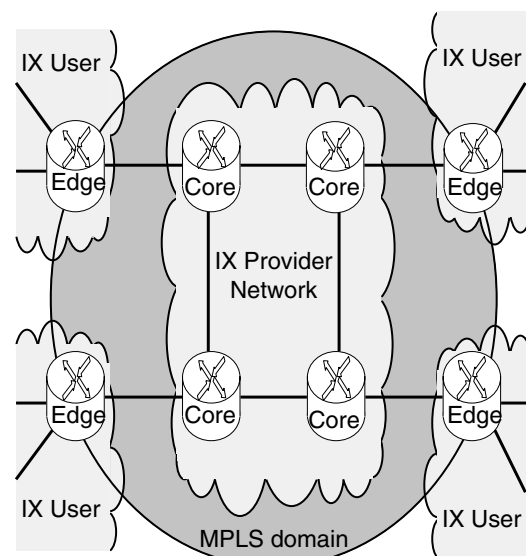


図 1.1. MPLS-IX アーキテクチャ

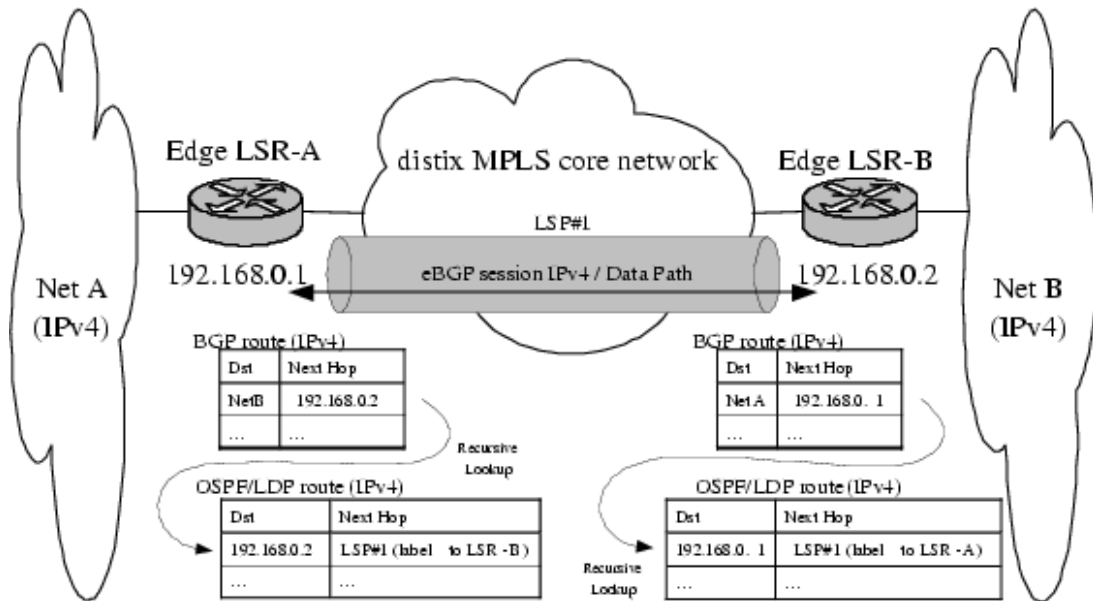


図 1.2. MPLS-IX での制御構造

1. エッジ LSR のみで IPv6 対応を行い、シグナリング等は IPv4 のものをそのまま用いる。
コアルータなどでの対応が全く必要ない。このため、既存 MPLS 網を容易に IPv6 に対応させることが可能である。
2. シグナリングトランスポートには IPv4 を用い、IPv6 に関する情報も交換する。
シグナリングプロトコルの IPv6 対応拡張が必要だが、未だ仕様が確定していない。コアルータを含め、IPv6 トラフィックを扱う全てのルータでの対応が必要である。
3. シグナリングトランスポートも含め、全ての制御に IPv6 を用いる。
現状の IPv4 と全く同等の扱いで IPv6 を扱うことができる。特別な処理が必要なく IPv6 のみを対象と考えると最も単純な形態だが、全てのルータを IPv6 対応させなければならない。
各項の必要条件から分かるとおり、順に導入されることになると思われる。

中でも、1 の方式は、IPv6 を必要とするエッジルータのみの対応で導入でき、既に構築されている MPLS 網を必要に応じて IPv6 対応させることができ、IPv6 への移行期である現状には重要な技術である。

MPLS-IX において、IPv6 の経路交換・トラフィック交換を実現した。

我々は、既設の MPLS-IX において IPv6 対応を実現するために、前節の 1 の手法を選択した。これにより、コアルータが IPv6 に対応していない状況においても、IPv6 対応を望む IX “利用者” のエッジルータのみの対応で実現できる。

この設計にあたり、我々は、

- BGP ルータ間で IPv4 で eBGP セッションを確立したうえで、MP-BGP (Multiprotocol Extensions for BGP-4) [153] を用いる。
- 広告する IPv6 経路情報の BGP next hop 情報には IPv4 アドレスを IPv4 写像アドレス (IPv4-mapped IPv6 address) を用いる。

こととした。

さらに、LSR で BGP 経路の IPv4 写像アドレスを再帰探索する際に IPv4 経路表を探索することとした (図 1.3)。これにより、最終的な探索結果に IPv4 での経路制御と MPLS シグナリングの結果が用いられることとなる。つまり、IGP の経路制御と MPLS シグナリングには手を加えることなく、IPv6 の経路交換・トラフィック交換を実現できた。

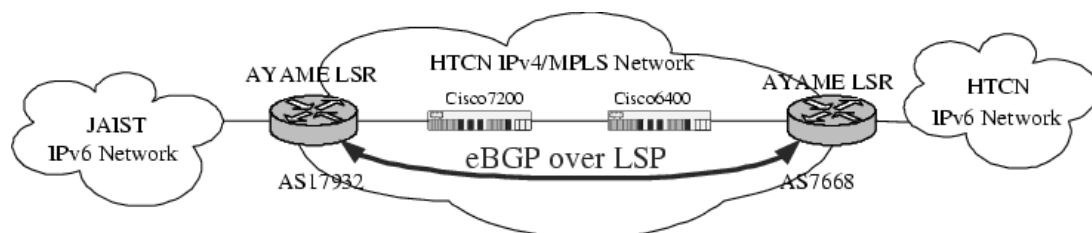


図 1.4. 実験トポロジ

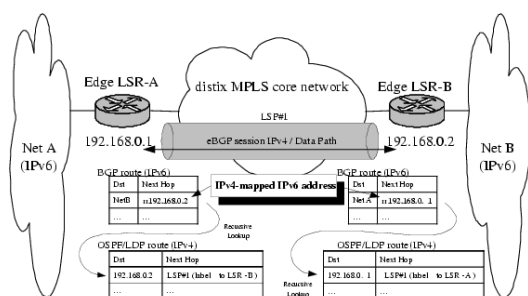


図 1.3. MPLS-IX での制御構造 (IPv6 拡張)

この方式は、“Connecting IPv6 Islands across IPv4 Clouds with BGP” [42] における MP-BGP over IPv4 アプローチ (Tunneling over MPLS LSPs) の MPLS-IX アーキテクチャへの適用である。

1.1.3 AYAME の IPv6 拡張実装

前章で述べた IPv6 拡張の実現のために、AYAME に対して行った拡張は、大分すると以下ようになる。

- zebra 経路制御デモンの IPv4 写像アドレスへの対応
- IPv6 スタックの MPLS 対応
- MPLS スタックの IPv6 対応

まず、zebra 経路制御デモンの IPv6 写像アドレスへの対応では、すでに述べたとおり、BGP によって得た経路の再帰探索にかかわる処理を拡張し、IPv4 経路表を参照するための拡張を行った。

次に、IPv6 スタックの MPLS 対応であるが、入力パケットをどの LSP を用いて転送すべきかを決定するためのパケット区分器を拡張し、IPv6 パケットを LSP へ転送することを可能とした。これは、ingress LSR で必要となる機能である。

最後に、MPLS スタックの IPv6 対応では、LSP

から入力されたパケットを正しい L3 プロトコルスタックへ渡すための拡張である。既存の AYAME では、IPv4 を扱う部分のみ実装されていたが、IPv6 パケットを IPv6 スタックへ渡すための機能を追加した。これは、egress LSR で必要となる機能である。

これらの拡張により、AYAME で、前章で述べた IPv6 対応拡張機能を扱うことができるようになった。

1.1.4 AYAME を用いた運用実験

この IPv6 対応拡張を施した AYAME を、MPLS-IX に接続し、運用実験を行った。

今回の実験では、MPLS-IX として、北陸通信ネットワーク株式会社 (HTnet)¹ が実験的に構築している MPLS 網を用いた。このネットワークに接続する北陸先端科学技術大学院大学 (JAIST) と HTnet に AYAME LSR を設置し、両 LSR 間で eBGP による IPv6 経路交換とトラフィック交換を行った。

このネットワークのコアは、IPv6 には対応していない LSR 製品で構成されていた (図 1.4)。このような網上でも、今回我々が設計実装した手法により、IPv6 トラフィックを LSP を介して問題なく流すことができた。

上記の実験の際の、JAIST 側 LSR の経路表は図 1.5 のようになった。

eBGP セッションは、

- JAIST 側: 211.120.192.2
- HTnet 側: 211.120.192.1

の間で張られている。この 2 つの LSR 間には、それぞれの LSR を両端とする LSP が LDP を用いて確立されている (経路表の 1 行目)。eBGP で交換された IPv4 経路 (経路表の 3~4 行目) は、nexthop が HTnet 側の LSR(211.120.192.1) に向いており、再起探索により LSP を用いて転送することが示されている。

¹ <http://www.htcn.ne.jp/>

```

S - static, B - BGP, M - MPLS,
> - selected route, * - FIB route

[IPv4 Routing Table]
M>* 211.120.192.1/32 [10/0] via LSP 0:20
S- 211.120.192.1/32 [15/0] via 211.120.193.5, fxp0
B>* 150.65.0.0/16 [20/0] via 211.120.192.1 (recursive via LSP 0:20)
B>* 203.178.128.0/19 [20/0] via 211.120.192.1 (recursive via LSP 0:20)

[IPv6 Routing Table]
B>* 2001:308::/48 [20/0] via ::ffff:211.120.192.1 (recursive via LSP 0:20)
    
```

図 1.5. JAIST 側 LSR の経路表 (抜粋)

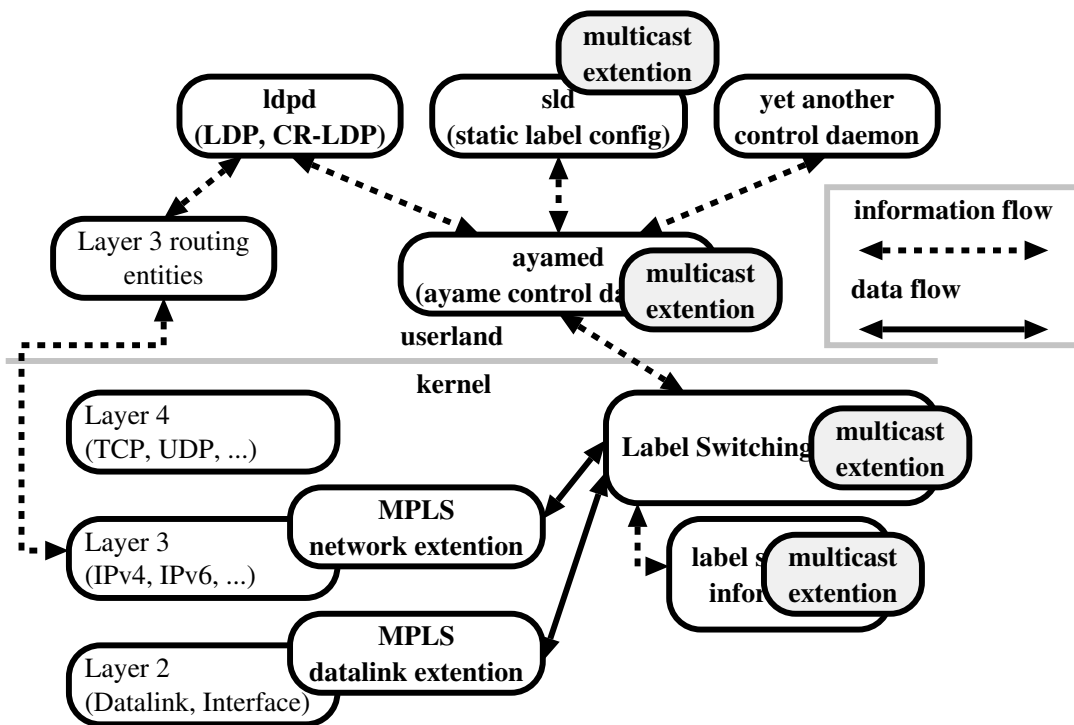


図 1.6. マルチキャスト機構のアーキテクチャ

本実験の焦点である IPv6 経路 (IPv6 経路表の 1 行目) については、nexthop が IPv4 写像アドレス (::ffff:211.120.192.1) となっている点が重要である。このアドレスは、もちろん、HTnet 側 LSR を示すものである。この経路の再帰探索は、通常の IPv6 経路表ではなく IPv4 経路表を用いて行われる。つまり、211.120.192.1 宛の経路を IPv4 経路表から探索し、IPv4 経路表 1 行目のエントリが用いられる。これにより、図中に例示した 2001:308::/48 宛のトラフィックは LSP を介して転送されることとなる。

1.2 AYAME MPLS 実装のマルチキャスト基本機能拡張

現在のインターネットにおける多点間通信 (マルチキャスト) は IP マルチキャスト [142] 技術およびその派生技術が用いられている。IP マルチキャストを実現するためには、パケットを多点間に配送するための技術と同時に、多点間でのパケットの配送経路を規定する配送木の確立などの配送制御技術が必要である。そのため、現時点では主にユニキャスト配送に関する仕様化がなされているマルチプロトコル

ラベルスイッチング (MPLS) を IP マルチキャストで利用するためには、多くの技術的課題を解決する必要があります。我々は MPLS におけるマルチキャストの実現を目指して、我々が設計開発している MPLS 実装 AYAME 上に MPLS でのマルチキャストパケット配送機構を実装した。本実装は IP マルチキャストを包含した汎用的な多点間通信機構を MPLS 上に実現するための試みである。ここでは、MPLS におけるマルチキャスト配送のフレームワークをまとめたうえで、MPLS パケット配送機構の設計と実装を論じる。

1.2.1 MPLS マルチキャスト転送機構の設計

MPLS では制御と配送は完全に分離しており、ユニキャストだけでなくマルチキャストにおいてもその構造が適用可能である。現在 MPLS を用いたマルチキャストとして IP マルチキャストと MPLS の併用に関する議論が IETF において行われているが、MPLS を用いたマルチキャストは IP マルチキャストを併用する方法だけでなく、経路制御には TE、配送するパケットは非 IP というような方法も存在する。以下は、MPLS におけるこのような制御と配送の特徴を IP と IP 以外という観点からまとめたものである。

1. IP 経路制御および IP パケットの配送:
LDP 等によって IP 経路制御に追従したラベルマッピングを配布し、ラベルを用いて IP パケットを配送する。バックボーンネットワーク等における一般的な MPLS の利用形態。
2. IP 経路制御および IP 以外のパケットの配送:
LDP (Label Distribution Protocol) 等によって IP 経路制御に追従したラベルマッピングを配布し、ラベルを用いて IP 以外のパケットを配送する。
3. TE 経路制御および IP パケットの配送:
RSVP-TE や CR-LDP 等のシグナリングプロトコルによって IP 経路制御とは独立したラベルマッピングを配布し、ラベルを用いて IP パケットを配送する。TE を目的とした一般的な MPLS の利用形態。
4. TE 経路制御および IP 以外のパケットの配送:
RSVP-TE や CR-LDP 等のシグナリングプロトコルによって IP 経路制御とは独立したラベ

ルマッピングを配布し、ラベルを用いて IP 以外のパケットを配送する。

特に上記の 1. については、RFC3353 [49] においてその経路制御とパケット配送に関する要求事項がまとめられている。同 RFC では、上記の 3. および 4. については明示的経路制御の方法として IP マルチキャスト経路制御プロトコルによる配送木を LSP ベースの配送木で上書きする方法、および明示的にユニキャストの経路を用いる方法の二種が考えられると言及するにとどまっているおり、IP 以外の経路制御や IP 以外のパケット配送については言及していない。

このように MPLS を用いたマルチキャストについては、単に IP マルチキャストと MPLS の併用だけでなく極めて多数の制御と配送の組み合わせが考えられる。

MPLS マルチキャストの制御機能は、IP マルチキャスト技術を対象とした経路制御や TE を用いたマルチキャスト技術でそれぞれ異なる可能性が極めて高く、それらの異なる複数の制御スキームで同一の MPLS マルチキャスト配送機構を利用できるアーキテクチャが要求される。このような中核的な MPLS マルチキャストの基本機能を実現し利用するためには、配送部分は基本的な機能を提供し、その制御インターフェース部分では汎用性の高いインターフェースを提供する必要がある。

このような種々の制御スキームでの応用を目的とする基本的な MPLS マルチキャスト配送機構に求められる基本機能を以下にまとめる。

- 入力されたパケットの複製機能
- 入力されたパケットの複数インターフェースへの同時複数出力機能
- 入力ラベルおよび各種情報によるフィルタを用いたパケット区分化機能

これらに加えて、MPLS マルチキャストを行う LSR に対して単一の LSR が単一の FEC に対して入口ノード (ingress)、転送ノード (transit)、出口ノード (egress) のすべての役割を同時に満たすことが要求される。ユニキャストの MPLS の場合、単一のパスについては各ルータで役割が明確に異なるが、マルチキャストの場合、LSP はツリー状であるのである枝については transit、ある枝については egress、そのルータがマルチキャストトラフィックの発生源の

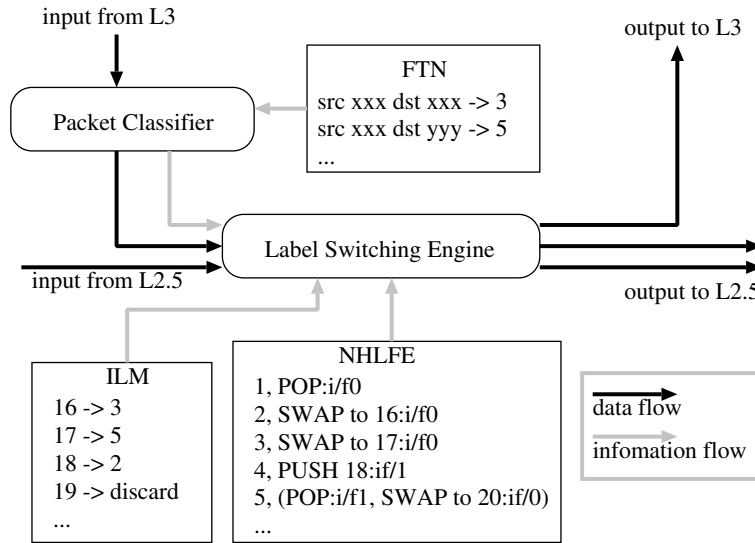


図 1.7. マルチキャスト配送機構

場合には ingress としても動作する必要がある。これは、MPLS マルチキャスト LSR は複数インターフェースへの同時複数出力機能として L2.5 および L3 に対する出力機能を持たなければならないということを示している。

1.2.2 AYAME における MPLS マルチキャスト転送機構の実装

AYAME をベースとした MPLS マルチキャスト機構の全体像を図 1.6 に示す。我々は、マルチキャスト配送機構をマルチキャスト拡張 LSE、マルチキャスト拡張 FIB、マルチキャスト拡張制御デーモンから構築という設計を選択した。

マルチキャスト配送に必要な情報は、カーネル空間に存在するマルチキャスト拡張された FIB に保持される。マルチキャスト拡張 FIB は、標準的な MPLS の FIB と同様 NHLFE、ILM、FTN から構成される。マルチキャスト拡張 NHLFE には、パケットに対する挙動が設定されている。ユニキャストの場合には、POP、SWAP、PUSH の操作、出力インターフェースが設定される。マルチキャスト拡張 NHLFE には、この操作と出力インターフェースのペアがリスト構造で格納される。

パケット複製機能および配送機能は、LSE の拡張として実現される。マルチキャスト拡張された LSE は、マルチキャスト拡張 FIB に基づいて、L3 および L2.5 から入力されたパケットを必要に応じて複製し L3 および L2.5 に適切に出力する。ラベルが

付加されたパケットが入力された場合、マルチキャスト拡張 LSE はそのラベル値から NHLFE を取得し、リストの要素の数にパケットを複製し、リストの各要素に従ってパケットに操作を行い各要素にしたがってパケットを出力する。このリストは出力が L2.5 であっても L3 であっても POP と SWAP というラベル操作として区別なく記述できるという特徴がある。一方、ラベルの付加されていないパケット、すなわち L3 のパケットが入力されてきた場合、L3 の AYAME 拡張によって LSE は FTN を用いてそのパケットに対する操作が設定されている NHLFE を取得する。この NHLFE に従って、ラベルが付加されたパケットが入力された場合と同様の処理を行う。マルチキャスト拡張 LSE とマルチキャスト拡張 FIB の詳細を図 1.7 に示す。

マルチキャスト拡張 FIB の確立は、ayamed を介してマルチキャスト拡張ラベル制御デーモンによって行われる。マルチキャスト用の FIB エントリはユニキャスト用の FIB エントリと同時に存在でき、ユニキャストされるパケットとマルチキャストされるパケットは同一の入出力インターフェースを介して処理されその管理テーブルも同一の管理インターフェースを用いて設定される。また、マルチキャスト拡張ラベル制御デーモンは制御スキーム毎に異なる実装が考えられるのでこの ayamed の管理インターフェースは種々のラベル制御デーモンで汎用的に利用できる必要がある。

我々は、AYAME の LSE、FIB、ayamed、sld にマル

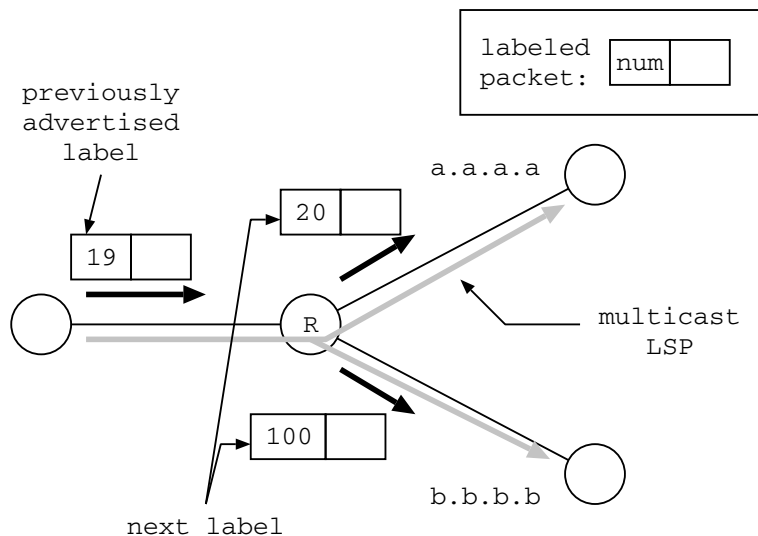


図 1.8. MPLS マルチキャストの例

マルチキャスト拡張を実装することで MPLS マルチキャスト機構を実現した。また、カーネルと ayamed のインターフェース、ayamed と sld のインターフェースにもそれぞれマルチキャスト拡張を行った。

我々は、LSE に内部的にラベルスイッチング動作を行わせることでマルチキャスト拡張 NHLFE に出カインターフェースと操作のペアのリストを格納するという機能を実現した。マルチキャスト拡張 NHLFE でインターフェースと操作のペアに加えて内部的に用いるラベルを保持し、パケットを複製しつつ LSE 内で内部的にラベルスイッチング動作を繰り返すことでマルチキャストラベルスイッチング動作を実現した。この方式は内部的にラベルスイッチング動作を繰り返すためラベルの PUSH、SWAP、POP などの操作を自由に組み合わせることが可能で、非常に柔軟な制御が可能であるという利点を持つ反面、ラベル空間を内部的に消費するという欠点を持つ。我々は、本実装には研究的利用における柔軟な設定が要求されるという点を重視し、このような実装方式を選択した。

このマルチキャスト拡張 NHLFE を構築するための制御デーモンとして我々はマルチキャスト拡張 sld を実装した。sld は静的な LSP を設定するためのデーモンで、マルチキャスト拡張 sld は静的な一対多の LSP を設定するためのデーモンである。以下は、マルチキャスト拡張 sld の設定構文である。

1. label { ラベル空間 } { ラベル値 } { 操作 } { 次

ホップ } { 次ラベル値 }

2. label { ラベル空間 } { ラベル値 } filter { プロトコル } src { 始点アドレス } { 始点アドレスマスク長 } { 始点ポート } dst { 終点アドレス } { 終点アドレスマスク長 } { 終点ポート }

1. のコマンドは ILM と NHLFE を同時に設定するためのもので、2. のコマンドはラベル空間とラベル値を NHLFE と対応させるキーとして FTN を設定するためのものである。これらのコマンドは、ユニキャストの場合もマルチキャストの場合も構文自体に差異はない。しかしながら、ユニキャストの場合には、上記の 1. のコマンドはラベル空間とラベル値のペアに対して一意でなければならないが、本論文で提案するマルチキャスト拡張 sld では 1. のコマンドの複数記述を許容できるようにした。これにより、あるラベルのついたパケットを受け取った際のそのパケットに対する LSR の挙動を複数記述でき NHLFE のリストに相当する FIB の構築が可能となった。図 1.8 に MPLS マルチキャストの例を示す。また図 1.9 に図 1.8 の LSR の設定例を示す。

1.2.3 考察と今後の課題

ここでは、MPLS を用いたマルチキャスト配送の実現に必要な、今後解決しなければならない主な問題点に着目して、前章までで論じた我々の実装について論じる。

マルチアクセスリンク

```

! configuration for unicast MPLS
label 0 18 SWAP y.y.y.y 10101
label 0 18 filter 1 ANY src 0.0.0.0 0 0 dst x.x.x.x 32 80
label 0 18 filter 2 ANY src 0.0.0.0 0 0 dst x.x.x.x 32 8080
! configuration for multicast MPLS
label 0 19 SWAP a.a.a.a 20
label 0 19 SWAP b.b.b.b 100
label 0 19 POP 0.0.0.0 0
label 0 19 filter 1 ANY src 0.0.0.0 0 0 dst x.x.x.x 32 0
    
```

図 1.9. MPLS マルチキャストの静的設定の例

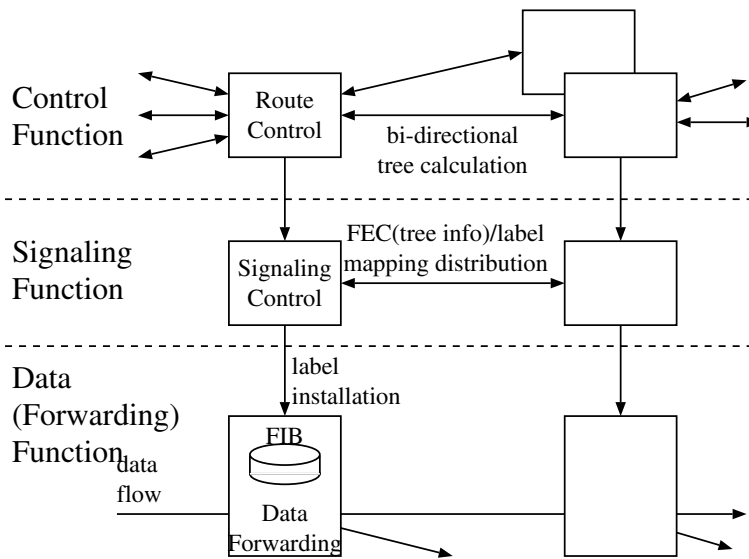


図 1.10. The architecture of the BLAST-CAST using MPLS

IP では同一のマルチアクセスリンク上でデータリンクレベルでのマルチキャスト技術が用いられるケースでも、現在の標準的な MPLS では個々の隣接ノードに対して個別のラベルを付加して送信する必要があり、帯域幅を浪費する可能性がある。我々が実装した MPLS マルチキャストの基本配送機能もこれを踏襲しており、運用の方法によっては帯域幅を浪費する可能性がある。これを本質的に解決するためには、マルチアクセスリンク上での単一ノード対複数ノードの LSP を用いる必要がある。しかしながら、このような point-to-multipoint の LSP については未だ研究段階である。point-to-multipoint の LSP はマルチアクセスリンク上で用いられるだけでなく、網全体に対する適用の可能性もあり、また、ユニキャストの point-to-point の LSP と同様の抽象化に基づいた場合には point-

to-point と point-to-multipoint の LSP を混在させたマルチキャスト技術など、応用の幅が広がることが期待できる。point-to-multipoint の LSP は潜在的に IP でのマルチキャスト技術とは根本的に異なるマルチキャスト技術の可能性を保持していると言え、応用の可能性は極めて高く、今後の研究が期待される。我々が実装した MPLS マルチキャスト機構では、マルチアクセスリンク上で単一のマルチキャストフレームを複製して出力することも複製せず出力することも可能であり、柔軟な設定が可能である。いえる。

非対象リンク

標準的な MPLS のシグナリングプロトコルではシグナリングの信頼性を高めるため、ピア LSR 間での双方向の情報伝達を必要とするケースが多い。衛星リンク等の非対称リンクを用いたマルチ

キャスト技術を MPLS に適用する場合、なんらかの方法でラベルマッピングを設定する必要があるが、現在の標準的なシグナリングプロトコルは双方向のリンクを前提としているケースが多く問題が多いと言える。我々が実装したラベルマッピングの設定方法は静的な設定でありラベルマッピングの変更が時間的に細粒度で運用する場合には非現実的であるが、非対称リンクなどのシグナリングが困難な場合に関する研究への応用の可能性を示している。

IP マルチキャスト

我々は、MPLS マルチキャストのフレームワークを示し、さらにその一部として IP マルチキャストを MPLS に適用するための要求事項のフレームワークを示した。前述のとおり、IP マルチキャストには種々の要素技術が用いられているうえ、伝統的な IP の特徴とも言える制御と配送の非分離を応用している側面もあるため、MPLS に IP マルチキャストを完全に適用するには極めて多くの問題が存在すると言える。MPLS においてマルチキャスト型の配送を実現するには、IP マルチキャストで利用されてきた種々の技術をそのまま MPLS に適用するだけでなく、MPLS の特徴を利用した新しいマルチキャスト技術を含めて段階的に MPLS マルチキャストを実現していく必要があると考えられる。

1.3 多対多マルチキャストの提案と AYAME での実装計画

経路表のエントリ数の増加を抑制しつつ多対多のマルチキャスト配送を実現可能なマルチキャスト手法として、双方向木を用いた手法が提案されている。ここでは、双方向マルチキャスト木の経路制御スキームとは独立に双方向マルチキャスト配送が可能な双方向マルチキャスト配送系のモデルであり、MPLS (Multiprotocol Label Switching) を用いたその実現として我々が提案した BLAST-CAST について論じる。

1.3.1 BLAST-CAST のアーキテクチャモデル

双方向型マルチキャストとは送信者に非依存に単一のマルチキャスト配送木を用いてマルチキャスト配送するというモデルに基づいたマルチキャスト技術である。このモデルでは、ルータは送信者毎の経

路を保持する必要がなく、多対多のマルチキャストにおける経路表の爆発的な増加を効果的に抑えることができる。あるグループに向けて送信されたパケットが入力された際にはそのグループに対応する枝に出力するだけであるので、グループ毎に単一の経路エントリを保持することとなる。言い換えると、送信者依存の経路エントリを送信者非依存な経路エントリとしてマージして保持するとも言える。双方向マルチキャストの本質的な利点はこの経路エントリのマージにあるとも言える。

BLAST-CAST は、双方向マルチキャスト技術の適用状況や、ポリシー状況などに応じて柔軟に制御できる双方向マルチキャスト配送系のモデルとそのインプリメントである。配送と制御を分離することで制御に関する問題を双方向マルチキャストの利点から明確に分離するというアプローチにより双方向マルチキャストの柔軟な制御を実現する。配送と制御の分離パラダイムは、ユニキャストについては既に確固たるパラダイムとして確立されており、そのパラダイムのインプリメントである MPLS 技術を用いた種々の TE や QoS 要求への対応が実現されつつある。

BLAST-CAST のアーキテクチャモデルは、FIB(Forwarding Information Base) を中心に、FIB を構築するための制御部と、任意の制御機構によって構築された FIB を用いてパケット転送するフォワーディング部を分離したモデルである。このように分離することで、フォワーディング機構は FIB だけに基いてフォワーディングするだけの機能、制御部は FIB を構築するための API を介して FIB を構築する機能だけがそれぞれに対する requirement となり、それぞれの独立性を高めることができると考えられる。

制御部では何らかのアルゴリズムやポリシーに基づいた計算や通信が行われ、その結果双方向マルチキャスト木の枝の構成を意味する情報が集積される。制御部はこの情報を FIB インターフェースを介して FIB に設定する。パケット転送部は、あらかじめ構築された FIB に基づいてパケットフォワーディングを行う。

1.3.2 MPLS を用いた BLAST-CAST の実装

既に述べた通り、MPLS のアーキテクチャは制御部、配送部、およびシグナリング部から構成されて

いる。BLAST-CASTのモデルも配送と制御の分離パラダイムに基づいたものである。MPLSのアーキテクチャとBLAST-CASTのモデルの親和性は高く、MPLSアーキテクチャのBLAST-CASTモデルへの適用は比較的容易であると考えられる。図1.10にMPLSを用いたBLAST-CASTのアーキテクチャを示す。以下は図中のそれぞれの機構の説明である。

BLAST-CAST 制御部

何らかのアルゴリズムに基づいて双方向配送木を制御する層。BLAST-CASTでは特定のアルゴリズムや特定の制御機構を想定しない。BIDIR-PIMの木生成アルゴリズムや双方向マルチキャスト木確立TEなどの「制御」が、BLAST-CAST制御部に属するものである。

BLAST-CAST シグナリング部

制御部での制御に基づいて、BLAST-CASTルータ間でシグナリングを行うプロトコル。このシグナリングプロトコルによって双方向マルチキャスト木を意味するFIBが構築、維持される。このプロトコルは、特定のプロトコルに限定されない。BLAST-CASTのシグナリング要件を満たすラベル配布プロトコル群をBLAST-CAST LDP(BLDP)と呼ぶ。

BLAST-CAST ラベルスイッチング部

FIBに基づいて双方向マルチキャスト配送を行うラベルスイッチング機構

MPLSのフレームワークでは、ラベル配布のおおまかな方針として要求駆動方式(on demand mode)と能動的駆動方式(unsolicited mode)を提案している。BLAST-CASTは基本的に多対多のマルチキャスト配送を目的とした機構であるので不要な情報の配布は極力避けるべきであり、必要なときにラベルを要求する要求駆動方式が望ましいと考えられる。BLAST-CASTの制御には様々なものが考えられるので能動的駆動方式が有用なケースも存在する可能性もあるが、それらに関する考察は今後の課題とする。また、MPLSのフレームワークでは、ラベル配布制御のおおまかな方針として依存型(orderd mode)と非依存型(independent mode)が提案されているが、受信者の要求駆動で木を構築する場合と、TEなどでコアバックボーンを構築する場合などで適切な配布制御方法は異なると考えられる。

現在MPLSのシグナリングプロトコルとして、

LDP[95]やCR-LDP[25]、RSVP-TE[47]などが提案されている。MPLSのシグナリングプロトコルは複数存在し、L3の経路制御に追従してラベルマッピングを配布する機能、ある束縛に基づいてラベルマッピングを配布する機能、など機能面で分類可能である。複数の異なるシグナリングプロトコルが同機能を実装していることもある。しかしながら、どのラベル配布プロトコルもFECとラベルのマッピングの配布が目的であり、その点については本質的な差はない。

BLAST-CASTで用いられるシグナリングプロトコルもMPLSの各種シグナリングプロトコルと同様FECとそれに対するラベルのマッピングの配布が目的である。しかしながらBLAST-CASTのシグナリングプロトコルには、既存のシグナリングプロトコルとは異なるいくつかの要求事項が存在する。

第一に、BLAST-CASTではLSPは常に双方向になる必要がある。BLAST-CASTのシグナリングプロトコルには双方向のLSPの確立および維持のための手順が必要である。第二に、MPLSのシグナリングプロトコルではパケットが属するFECを判別するために用いられる情報はdestination addressやLSP IDなどであったが、BLAST-CASTではFECを判別するための情報としてパケットが転送される双方向マルチキャスト木を判別するための情報をシグナリングする必要がある。第三に、BLAST-CASTではLSPは木構造になるので、従来のMPLSのシグナリングプロトコルには存在しなかった分岐点LSRという概念を導入する必要がある。

BLAST-CASTのシグナリング制御用インターフェースに必要な基本機能は、枝の接合、枝の切断、枝の状態の通知である。枝の接合操作には、引数として木の識別子と接合するための経路情報が必要である。枝の切断には木の識別子が必要である。木の状態の通知には、木の識別子と、木の状態に関する情報が通知される。

これらをLSPに対する操作に写像すると、「枝の接合」はAからBへのLSPとBからAへのLSPの確立、「枝の切断」はAからBへのLSPの破棄とBからAへのLSPの破棄である。双方向LSPの一方のLSPが切断された場合、他方のLSPを切断し、BLAST-CASTの双方向木の枝が切断されると制御部に通知する。以下に各操作とそれに必要な情報、および通知される情報をまとめる。

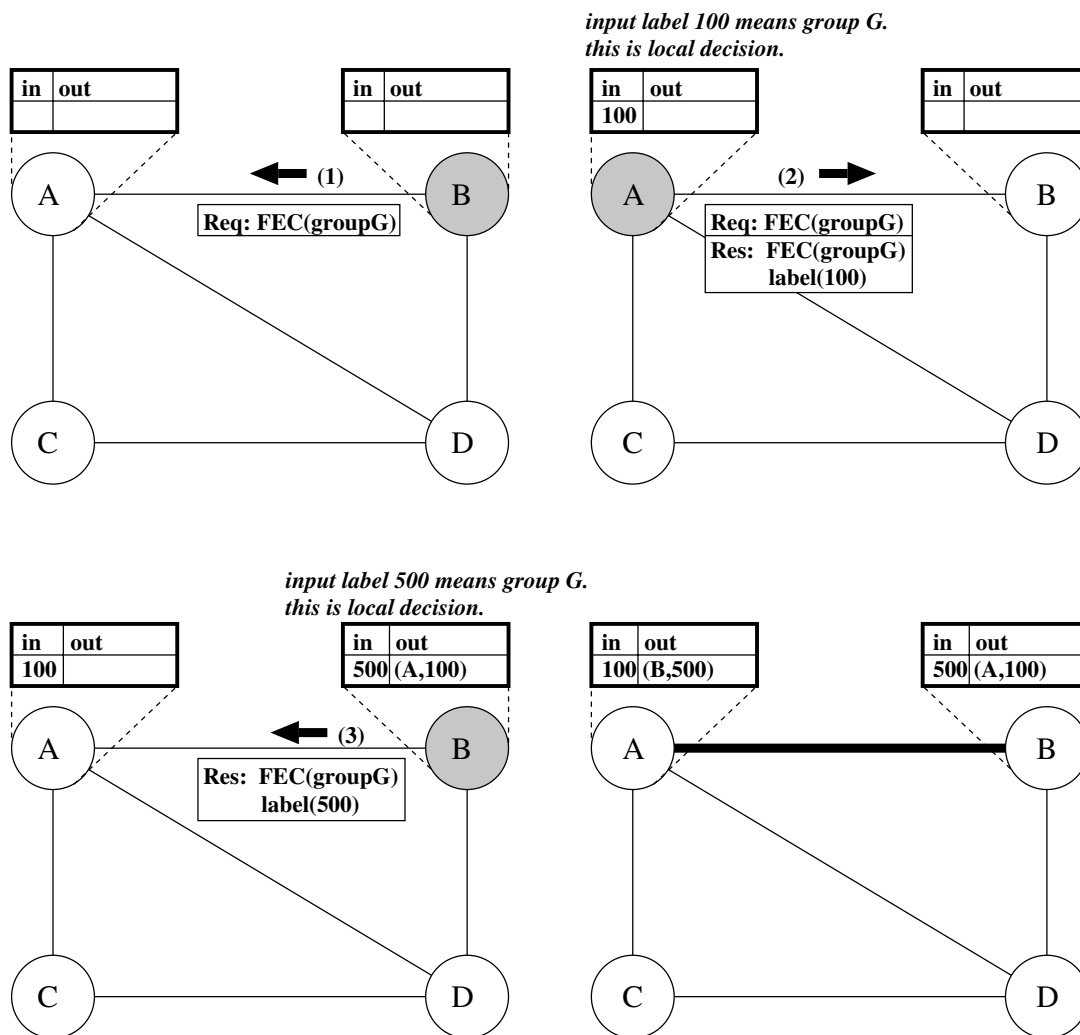


図 1.11. BLAST-CAST signaling example

枝の接合操作

引数は木の識別子と経路情報を引数とする

枝の切断操作

引数は木の識別子を引数とする

枝の切断通知

切断された木の識別子、および木の状態が通知される

図 1.11 に、BLAST-CAST シグナリングプロトコルによる双方向 LSP 確立の例を示す。まず LSR B から LSR A に対して、グループ G に対するラベルを要求する。A はグループ G に対するラベル値を 100 と決定し、(group G, label 100) というマッピングで応答する。この際、マッピングと同時に逆方向 LSP 用のラベルを要求する。B は、グループ G パケット用に配布するラベル値を 500 と決定し、A から受け

取ったマッピングを group G = 500 : (A, 100) として保持するとともに、A から受け取ったラベル要求への応答として (group G, label 500) というマッピングで応答する。このようにして双方向の LSP を確立できる。自 LSR が隣接 LSR に配布する、グループにマップされたラベル値は配布元 LSR が独自に決定する。この値を複数の隣接 LSR に対して同一の値とするか、各隣接毎に異なる値とするかは実装上選択可能である。つまり、A がこの後グループ G は 100 あるいは 200 として、C に (group G, 200) というマッピングを配布することも可能である。

第2章 MPLS を応用した VLAN 設定

本章では、VLAN タグに MPLS のラベルをマップして伝送することで VLAN に関するオペレーションのコストを軽減する手法について、これまでに行った議論をまとめる。

2.1 VLAN LSR の提案

現在、高速で長距離伝送可能なイーサネットの登場により、広域イーサネットによる L2VPN のサービス等が行われるようになっている。このような広域イーサネットにポイントツーポイントの L2 トンネル (仮想パス) を組み合わせることで、ユーザと ISP を接続する高速、広帯域で安価なアクセスネットワークを提供することが可能である。

具体的には、広域イーサネット上に VLAN を用いることで仮想パスを構成することが可能である。広域イーサネット上に VLAN セグメントがパスを構成するように設定することで、このアクセスネットワークに接続するユーザの packets を、接続先の ISP まで IP アドレスに依存することなくスイッチすることが可能になる。また packets のフォワーディングにレイヤ 2 のスイッチを用いるため、広帯域網を非常に安価に構成することが可能となる。

このようなモデルの特徴を以下に示す。

- レイヤ 2 での伝送を行うため、ネットワーク層の protocol (IPv4/6) やそのアドレス体系に依存しない。
- ネットワーク層の protocol の多重が可能であり、さらに、複数の ISP を接続することが可能である。
- VLAN の設定を集中管理することで、自由にパスを構成し、経路を管理することが可能である。この為、トラフィックエンジニアリングや経路の障害時のリカバリーをすばやく行うことが可能である。

しかし、現状では、各装置に対してそれぞれ、そのつど VLAN の設定を実行することが必要となり、オペレーションの負荷を非常に上げる原因となる。このため、このような VLAN によりパスを構成す

るネットワークに対して、MPLS の概念を導入し、VLAN タグにラベルをマップして伝送する方法と、シグナリングによるラベルの配布方法について提案した。また、トラフィックエンジニアリングを行うための明示経路を指定して MPLS パスを構成する方法とともに、VLAN スイッチを用いて通常のホップバイホップの MPLS パスを構成する方法についても議論した。

2.1.1 定義

LSR は [55] に記述されるラベルスイッチコントロールコンポーネントとフォワーディングコンポーネントを実装したデバイスである。

ラベルコントロール VLAN (LC-VLAN) インタフェースはラベルスイッチコントロールコンポーネントによってコントロールされるイーサネットインタフェースである。このようなインタフェースを通過する packets を受信したとき、それは通常の MAC スイッチでなく、VLAN タグによるラベルスイッチとして処理することもできる。また、このような packets は VLAN タグ中の VID フィールド中でラベルを運ぶ。

VLAN-LSR は VLAN タグ中で運ばれるラベルを用いて、これらのインタフェース間をフレーム転送する多くの LC-VLAN インタフェースを持つ LSR である。

VLAN-LSR では、コントロール用リンクには異なるネットワークアドレスが割り当てられている。ラベルスイッチコントロールコンポーネントは、この異なるネットワーク上で隣り合う VLAN-LSR と隣接を構築する。

VLAN-LSR のドメインは LC-VLAN インタフェースによって相互に接続された一連の VLAN-LSR である。

VLAN-LSR ドメインのエッジセットは LC-VLAN インタフェースによってドメインのメンバに接続される LSR のセットである。VLAN-LSR ドメインのエッジセットのメンバである LSR はエッジ LSR と呼ばれるかもしれない。

加えて、ここでは [55] に記述されるターミノロジを使用する。

2.1.2 オペレーションモデル

VLAN スイッチを VLAN-LSR として構成、管理

するネットワークモデルを2種類示す。

ルーティングモデル

VLAN-LSRにより構成されるMPLSドメインがルーテナットワークとして他のルータおよびLSRに接続される。

VLAN-LSRドメイン内の経路情報はドメイン外にも伝達され、IPルーティングによりパケットの伝送を行うことも可能である。このためこの様な網においてLSPを構成する場合、ホップバイホップでのルーティング情報に基づくパスの構成と、明示経路の指定によるパスの構成が可能である。

トランスペアレントモデル

VLAN-LSRにより構成されるMPLSドメインがブリッジネットワークとして、他の網を相互接続する。VLAN-LSRドメイン内の経路情報はドメイン外には伝達されない。このため外部の網からはレイヤ2をトランスペアレントに伝送するトンネルとして認識される。この様な網においてISPを接続する場合、外部から受信するIPパケットに基づいてホップバイホップでパスを構成することは不可能で、明示経路の指定によるパスの構成を行い、インGRESSでのクラシファイにより、必要な経路に伝送データをマッピングすることになる。この様なネットワークモデルとしては、ユーザとISPを接続するアクセス網や、L2VPN網などがある。このような網の構成概要を図に示す。

2.1.3 機能

MPLSアーキテクチャーがLSRインプリメンテーション中の相当な柔軟性を許している一方、VLAN-LSRは、ハードウェアおよびVLAN(IEEE802.1q)の基準によって課されたVLANタグフォーマットのような制限を有する。これらの制約のために、いくつかの特別の手続きがVLAN-LSRに要求される。

LSRとしての動作に影響するイーサネットVLANスイッチのキーとなる特徴のいくつかは：

ラベルスワッピング機能をサポートしない。ただし、通常のVLAN対応イーサネットスイッチをフォワーディングコンポーネントとして使用する場合、ラベルコントロールコンポーネントを実装してもラベルスワップの機能はない。この場合入力ラベルと同じ値の出力ラベルを使用することになる。また、VLAN-LSRとしてラベルスイッチ可能なフォワーディングコンポーネントを実装するのであれば上記制約の限りではない。

VLANタグ中のVIDでラベルを運ぶ為、パケット中のラベルのサイズと位置は固定される。

ラベルスワップ機能の無いVLAN-LSRドメイン内のLSPは双方向である。ただし、ラベルスワッピングを行わないため入力、出力で使用されるラベルに区別は無い。このため1つのラベルは2つのポートにマッピングされそれぞれのポートの入力・出力ラベルは同一のものとなる。このため、トラ

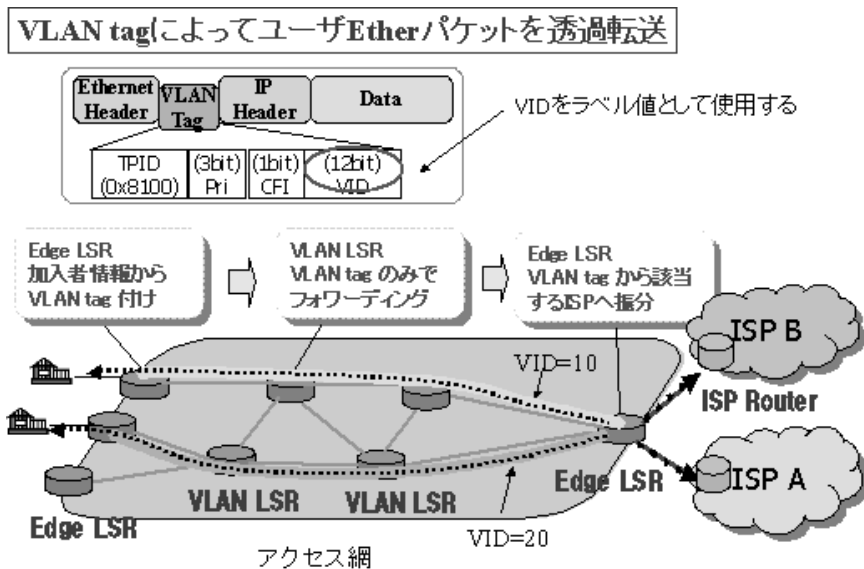


図 2.1. トランスペアレントモデル

トランスパレントモデルでは、ラベルスワップ機能の無い VLAN-LSR のみで構成したドメインの LSP は 1 本の LSP を張ることで双方向 LSP として使用することが可能である。また、ラベルスワップ機能の無い VLAN-LSR 以外の LSR(ATM-LSR や FR-LSR 等) を含んだドメインにおいて、ラベルスワップ機能の無い VLAN-LSR では LSP 方向を認識しないが、その他の LSR によって方向が識別されているため、この LSP を双方向として使用することはできない。

ルータ中の IP ヘッダで実行されるような TTL 減算機能を実行する能力はない。ただし、トランスパレントモデルの場合、このドメインは外部からブリッジとして見えるため、TTL 減算は必要ない。また、TTL の減算が必要なネットワークを構成する場合、シグナリング中のホップカウントオプションを使用することで Ingress の LSR により TTL 減算を実行することは可能である [14] [148]。

2.1.4 コンポーネント

ラベルスイッチングをサポートするために VLAN スイッチはラベルスイッチングのコントロールコンポーネントを実装しなければならない。これは、第一にラベルの割り当て、配布とメンテナンスの手続きからなる。

ラベルバインディング情報はいくつかのメカニズム、特に LDP[95],CR-LDP[25] によってやり取りされる。このドキュメントは LDP,CR-LDP にある要求を課す。

ただし、ここではラベル配布プロトコルとして LDP、CR-LDP に対する要求を記述するが、使用可能なラベル配布プロトコルをこれらに限定するものではない。その他のプロトコル (RSVP-TE 等) を使用してパスを構成することは可能である。

このような場合、LSR はラベルバインドを配布するために他のプロトコル (RSVP) を使用することができる。これらのケースでは、VLAN-LSR はこれらのプロトコルに参加する必要がある。しかしながら、これらはここでは明示的には考慮されない。

2.1.5 VLAN タグの利用

ラベルスイッチは、FEC にラベルを関連づけ、パケットを転送するためにラベル値を使用する [55]。VLAN-LSR では、ラベルは VLAN タグ中の VID

フィールドで運ばれる。

加えて、2 つの LDP ピアが LC-VLAN インタフェースで接続されるなら、コントロールパケットを運ぶことが可能な MPLS でない接続が利用可能でなければならない。この MPLS でない接続は 2 つのピア間でシグナリングパケットを転送するために使用される、またその他のラベル無しパケット (OSPF などの) のためにも使用することもできる。

コントロール情報やラベル無しパケットを運ぶために、追加の VLAN タグを用いて LC-VLAN インタフェースを設定することも可能である。

ノン MPLS 接続の VID の値は双方で、設定した値を認識する限り、さらに 1 つの VLAN-LSR における複数のインタフェースで VID の値が重複することが無い限り、自由に設定することは可能である。ただし、ユーザパスとして使用する MPLS 接続の VID の値との間でスペースの区分を行わない場合、VLAN-LSR ドメイン内においてノン MPLS 接続の VID の値とパス用 VID の値がラベルスワップ機能を有さない VLAN-LSR で重複しないよう管理する必要がある。また、ノン MPLS 接続の VLANID のみに関係していえば、VLAN-LSR ドメイン内で VLANID の重複があってもかまわない。ただし、これもラベルスワップ機能を有さない VLAN-LSR において重複がないことが条件である。

2.1.6 ラベル配布とメンテナンス手順

エッジ LSR の動作

オペレーションモデルで記述したルーティングモデルとトランスパレントモデルについて、それぞれのエッジ LSR で動作が異なる。

トランスパレントモデル

まず、トランスパレントモデルの VLAN - LSR ドメインのエッジセットのメンバを考察する。トランスパレントモデルの場合、転送するパケットのあて先 IP アドレスとは無関係にトンネルのパスを構成することが可能であるため、網管理機能等によりあらかじめパスを算出し、このパスを構成する VLAN-LSR の IP アドレスのリストをイングレスとなるエッジ LSR に通知することが可能である。このように、エッジ LSR は、与えられたリスト上の次ホップのアドレスを持つ VLAN-LSR に対してラベルバインドをリクエストするために CR-LDP 等の明示経路を構成可能な

シグナリングプロトコルを使用する。エッジ LSR はラベルバインド情報を一度受信すると、その指定される LSPID のパスを確定する。この LSP に対する、クラシファイ情報を指示されると、出力ラベルとして指定されたラベルを用いて転送するために、MPLS 転送手順を使用する。クラシファイ情報としては、入力ポート番号や、入力される VLAN タグの値等を用いることが可能である。また、ラベルスワップ機能を有さない VLAN-LSR のみで構成された LSP の場合、LSP は双方向に使用することが可能であるため、クラシファイ情報は LSP の両端となるエッジ LSR 双方に対して指示することが可能である。

ルーティングモデル

次に、ルーティングモデルの VLAN-LSR ドメインのエッジセットのメンバを考察する。ルーティングモデルの場合、明示経路の構成も、ホップバイホップでのパスの構成も可能である。また VLAN-LSR ドメインがパスを構成する MPLS ドメインの一部 (部分集合) となることも可能である。エッジ LSR が VLAN-LSR ドメインのエッジであるとともに MPLS ドメインのエッジでもある場合、明示経路の構成に関しては前述のトランスペアレントモデルのエッジ LSR と同様の動作を行う。ホップバイホップでのパスの構成に関しておよび、ラベルスワップ機能を有する VLAN-LSR を使用したパスの構成に関しては、[24] に記述されるのと同様のエッジ LSR の動作を行う。エッジ LSR が MPLS ドメインの中継 LSR として動作する場合、ラベルスワップ機能を有さない VLAN-LSR で VLAN-LSR ドメインが構成され明示経路のパスを張る場合、VLAN-LSR ドメイン内で使用するラベルが重複することがないように、egress のエッジ LSR がラベルを割り当てるときに、網管理機能に対し使用するラベルを問い合わせるなどの手法ととることが必要である。

内部 VLAN-LSR

VLAN-LSR が LC-VLAN インタフェース上で、VLAN-LSR に接続するピアからラベルバインド要求を受信したとき、VLAN-LSR は下記の動作をする：

- ラベルを割り当てる。ラベルスワップをしない VLAN-LSR で構成された VLAN-LSR ドメイ

ンの場合、リクエストに含まれる LSPID をラベルとして割り当てることも可能である。その他、egress のエッジ LSR でのラベル割り当て時に網管理機能に対し使用するラベルを問い合わせる等のラベルの管理を行うことが必須である。マッピングを待って、同じ値を割り当てる。ただし、LSPID をラベルとして割り当てる場合、正確には LSPIDTLV に含まれる LocalCR-LSPID の 16 ビット中の 12 ビット分をラベルとして VLAN タグの VID にマッピングする。このため、LSPID も 12 ビットまでの値の使用に限定する。ラベルのスワッピングを行わない VLAN-LSR では、ラベルが LSP を示す。通常ラベルは各 LSR に独立のスペースを持つが、ラベルスワッピングを行わない VLAN-LSR においては VLAN-LSR ドメインで共通のスペースを持つことになる。このため網管理機能によりドメイン内で重複の無いように LSPID を管理する必要がある。また、ドメイン内で 2 つの LSP が全く分離された状態 (2 つの LSP が同一の VLAN-LSR を通ることがない) である場合、網管理機能で管理する限りにおいてはこの 2 つの LSP に同一のラベル (LSPID の 12 ビット分) を使用することは可能である。

FEC に対する次ホップの VLAN-LSR もしくは明示経路を指定するリクエスト中の次アドレスの VLAN-LSR に対してラベルバインドを要求する。

要求してきたピアに割り当てたラベルを含むバインドを返信する。

(VLAN-LSR は下流からバインドが帰ってくるまで待つ) これは、([55] および [95] で定義される) “ordered control” の形式であり、特に “ingress-initiated ordered control” の形式である。

VLAN-LSR が他の (上流)LSR からラベルバインドリクエストを受信した結果、その次ホップの LSR にラベルバインドリクエストを生成して、次ホップ LSR へのリクエストが満足されないときはいつも、VLAN-LSR は受信したリクエストに関連して生成したバインドを破棄してリクエスト要求者に通知するべきである。

LSR がもう一方の LSR との LDP セッションを失ったと判断したとき、下記の動作が行われる。

r
t

o
p
r
e
s
s
i
o
n
a
n
n
o
2
0
0
2

T
C
E
J
O
R
P
E
D
I
W

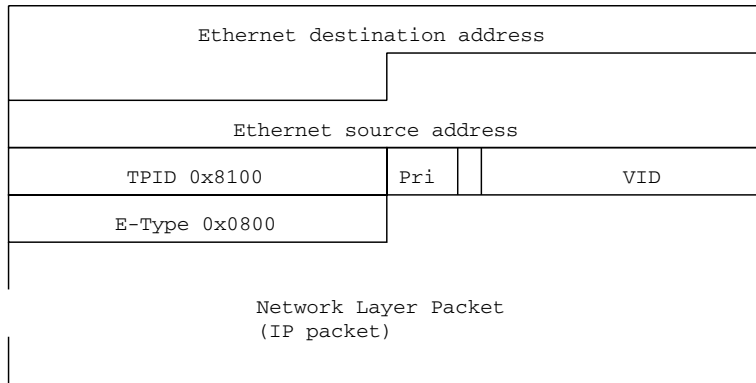


図 2.2. ラベルスタックがない場合

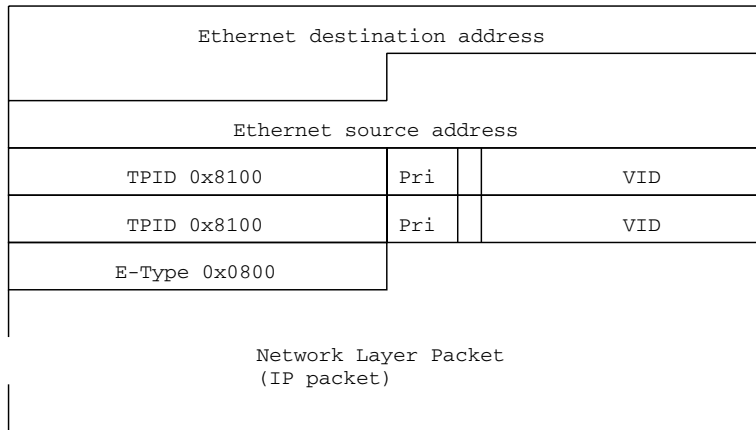


図 2.3. ラベルスタックがある場合

この接続を通じて学習したバインド情報の全ては破棄されなければならない。ピアからラベルバインドリクエストを受信した結果として生成された全てのラベルバインドに対して、LSRはこれらのバインドを破棄するかもしれない(そしてこれらのバインドに関連するラベルを開放する)。

2.1.7 カプセル化

ここで記述する手順はVLAN-LSRドメインのエッジLSRのみに影響する。VLAN-LSRそれら自身はいかなる方法でもカプセル化を修正しない。ラベルパケットはVLANタグを使用して、LC-VLANインタフェース上で転送される、このとき、パケットはシムヘッダを含んでいる必要はない。

図 2.2 および図 2.3 に VLAN-LSR でのカプセル化の方法を示す。

Ethernet destination/source address
IEEE802.3 で規定されるデータリンク層ヘッダア

ドレス
TPID
Tag Protocol Identifier タグプロトコル識別子 = 0x8100
Pri
ユーザ・プライオリティ 3 bits
CFI
Canonical Format Identicator = 0
VID
Virtual LAN Identifier 12bit
TCI
Tag Control Information 16bit + Pri+ CFI + VID
E-Type
Ethernet Type 16bit
もしパケットが n エントリのラベルスタックを持っている場合で、全てのエントリが VLAN-LSR ドメイン内で終端されたパスを表す場合、n エントリの VLAN タグ (TPID+TCI) を運ばばよい、ラベルの

0	0	VLAN label (0x0203)	Length
Reserved			VID

図 2.4. ラベル TLV の拡張

値は VID フィールド中にエンコードされる。エントリのスタックは TCI フィールドの後に置かれるイーサタイプで判断する、イーサタイプが TPID(0x8100) であれば、スタックである。このように、IEEE802.1q で規定される VLAN タグフォーマットでは、VLAN タグ (TCI フィールド) の後ろにイーサタイプがくるよう規定されているため、スタックビット等を用いることなく標準のタグを用いて、スタックを示すことが可能である。また、同様にイーサタイプによりネットワーク層プロトコルを示すことが可能であるため、一般的なシムヘッダ [56] で規定されるネットワーク層プロトコルを示す明示的ラベルを使用する必要も無い。

パケットの出力 TTL はネットワーク層ヘッダで運ばれる。本ドキュメント内のオペレーションモデルで、VLAN-LSR ドメインがブリッジネットワークとして他の網の相互接続を行うためのトンネル網を形成する場合、各 LSR では L2 トンネルのデータリンク層を終端することはないので、VLAN-LSP では TTL の減算を行う必要が無く TTL はネットワーク層ヘッダで運ばれる。一方、ルーティングネットワークとしてネットワーク層の終端を行う場合、[14] および [24] に示されるようにシグナリングのホップカウントオプションを利用して VLAN-LSR ドメインのエッジで TTL の処理を行うことも可能である。

2.1.8 ラベル TLV の拡張 (LDP)

VLAN-LSR は LC-VLAN インタフェース上で使用するためのラベルをコード化するために、VLAN Label TLV を使用する。図 2.4 に拡張ラベル TLV のフォーマットを示す。

Reserved

このフィールドは予約。もし送信時は、0 をセットしなければならないそして受信時は無視しなければならない。

VID

Virtual lan Identifier 12bit

