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　研究活動の概要

WIDE プロジェクトでは、IETF MPLS(Multi Protocol Label Switching) WG において検討されているラベルスイッチ技術に関する研究を 1996 年 4 月より LAST(LAbel Switch Technology) WG で行っている。

本報告書では、以下の内容に関する報告を行う。

- MPLS を用いた IX に関する研究
- ラベルスイッチシステムに必要なラベル数の評価
- 標準 ATM 上のラベルスイッチ技術 (VCID 通知手順)
- IETF MPLS-WG への技術貢献

第 2 章　MPLS を用いた IX に関する研究

2.1 背景

インターネットでは、複数のネットワークを相互に接続する手段のひとつとしてインターネット・エクスチェンジ (Internet eXchange: IX) が構築されるケースが多くなってきている。インターネット・エクスチェンジはひとつのネットワーク上で複数のネットワーク組織が相互に接続し、効率的にトラフィック交換を行なうことを目的としている。

一方、インターネット・エクスチェンジ技術にはまだ未解決の問題も多く残されている。例えば、いくつかのインターネット・エクスチェンジでは、ネットワーク間の相互接続に Ethernet や FDDI などの LAN 技術を利用しているが、これらの技術では通信の帯域が限界になりつつあること、あるいは、物理的に離れた場所での接続ができないことなどが問題になる。また、最近ではインターネット・エクスチェンジの相互接続に ATM 技術を利用することも多くなってきているが、この場合、個別に PVC の設定を行なう必要があるため、相互接続組織が増えると、運用コストが膨大になる、などの問題が残されている。

本稿では、インターネット・エクスチェンジにおいて、ラベルスイッチ (Label Switching)[45] の技術を応用して、分散された環境でも効率的にトラフィックの交換が可能なインターネット・エクスチェンジの設計を行う。

ラベルスイッチでは、特に ATM などのスイッチ網において、カットスルー (Cut Through) 技術を用いることにより高速なパケット転送を可能にするとともに、トラフィック・エンジニアリングの技術によって、ポリシに基づいたトラフィック制御が可能である。また ATM などの PVC の設定のほとんどを自動化することもできるため、運用コストを抑えたまま、分散された環境での高速なトラフィック交換も可能になる。

本稿では、この技術を応用し、インターネット・エクスチェンジにおいて、広域分散環境で複数のネットワークを相互に接続するための手法を提案する。本手法では、インターネット・エクスチェンジで必要とされるポリシの制御をラベルスイッチにおけるラベルを用いて行い、同時に、カットスルーを用いた効率的なトラフィック交換を実現する。また、ルートサーバの技術を用いることにより、経路制御における管理・運用を容易にし、インターネット・エクスチェンジ全体の運用コストを最小限に抑えながら効率的なトラフィック交換を可能にする。

2.2 インターネット・エクスチェンジ

インターネットは、数万ともいわれるネットワークが相互に接続された「ネットワークのネットワーク」である。これらの相互接続において、複数のネットワーク間を効率的に相互接続するための技術としてインターネット・エクスチェンジ (Internet eXchange: IX) の技術がある。図 2.1 は複数のネットワーク間で相互接続を行う場合の典型的な接続構成を示しているが、このように、個別に相互接続を行う場合には、各ネットワークは相互接続を行うネットワーク数に比例した相互接続コストを必要とする。

一方、インターネット・エクスチェンジ技術を用いる場合、図 2.2 に示されるように、ひとつのネットワークに複数のネットワークが接続し、集中的にトラフィックを交換することを可能にするため、各ネットワーク組織は相互接続のための回線コストや運用コストを抑えることが可能でる。

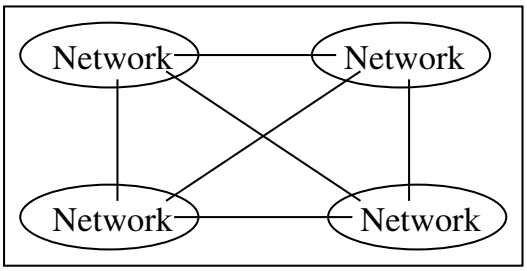

図 2.1　IX のない場合の相互接続

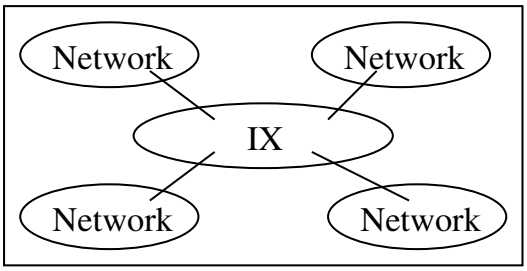

図 2.2　IX による相互接続

米国では CIX (Commertial Internet eXchange), MAE(Metroporitan Area Network), Chicago NAP (Network Access Point) などに代表される、多くのインターネット・エクスチェンジが運用されている。また、国内では WIDE Project により実験・研究目的で構築・運用されている NSPIXP (Network Service Provider Internet eXchange Point) や商用サービスの JPIX (JaPan Internet eXchange), MEX (Media EXchange) などが有名である。

2.2.1 インターネット・エクスチェンジにおけるポリシ

インターネット・エクスチェンジでは多数のネットワークが相互に接続されるが、経路情報の交換、すなわち、実際のトラフィックの交換は各ネットワークの運用ポリシに依存することが一般的になってきている。そのため、多くのインターネット・エクスチェンジでは全ネットワーク間で共通のポリシを強要されるマルチラテラル (Multi-Lateral) による相互接続協定よりも、個別にポリシを決定できるバイラテラル (Bi-Lateral) による相互接続協定を採用していることが多い。

2.2.2 Layer 3 のインターネット・エクスチェンジ

インターネット・エクスチェンジを構築する最も単純な方法は、IP データグラムを転送可能なルータ群で構成する (Layer 3 IX) 方法である。しかし、インターネット・エクスチェンジをルータ群で構成した場合、ルータのデータグラムの転送能力が全体のトラフィック交換性能に与える影響が大きく、性能上の問題が指摘されている。

また、前述の通り、近年ではインターネット・エクスチェンジにおいてバイラテラルによる相互接続協定を採用しているケースが増えてきているため、インターネット・エクスチェンジのネットワーク構成は、各ネットワークの運用ポリシを反映させることができなくてはならない。

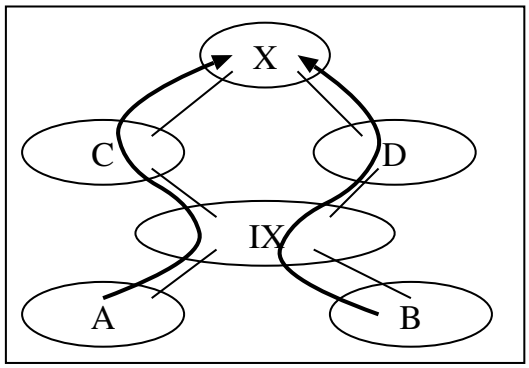

図 2.3　IX におけるポリシの反映

図 2.3 はインターネット・エクスチェンジの論理的な接続例を示している。例えば、インターネット・エクスチェンジに接続されるネットワーク A, B から X へデータグラムを送る場合にそれぞれ次のようなポリシが適用される場合、インターネット・エクスチェンジ内では、同じ宛先に対して異なる経路を選択する必要がある。

- A から X へは C を経由して送信する
- B から X へは D を経由して送信する

しかし、IP (Internet Protocol) においてルータがデータグラムの転送を行なう場合、その処理はルータ毎の Hop-by-Hop で宛先アドレスによって逐次処理されるため、上図において、インターネット・エクスチェンジがルータ群で構成されている場合 (Layer 3 IX) は、送信元が A, B のどちらかによって途中の経路を選択するということはできない。

このため、近年のインターネット・エクスチェンジではデータリンク層による接続を行なう Layer 2

IX を構築し、各ネットワーク間のトラフィック交換はデータリンク層のみで実現することにより、個別のポリシを反映させるケースがほとんどである。

2.2.3 LAN 技術を応用したインターネット・エクスチェンジ

これまでは、インターネット・エクスチェンジをデータリンク層 (Layer 2) で実現する手段として LAN 技術を利用するケースが多かった。例えば NSPIXP-2 や JPIX では、FDDI スイッチを用いて、ひとつのビル内で多数のネットワークのルータを相互に接続しており、NSPIXP-3 では Ethernet スイッチを用いて相互接続を行っている。

しかし、Ethernet や FDDI などの LAN 技術を用いた相互接続を行う場合、物理的に離れた場所での接続はできないため、各ネットワーク組織は、相互接続を行う場所へルータを持ち込み、物理的には一箇所、もしくは非常に近い場所での接続を強制される。

また、インターネットのバックボーン速度が LAN における通信速度を越えた現在においては、LAN 技術の通信速度はインターネット・エクスチェンジでのトラフィック交換には不適切だとの指摘もある。一部では Gigabit Ethernet などの利用実験を行っているケースもあるが、将来的には LAN 技術では帯域不足などの問題がより深刻になると思われる。

2.2.4 ATM によるインターネット・エクスチェンジ

近年では Chicago NAP や MAE-ATM のように ATM(Asynchronous Transfer Mode) による分散型のインターネット・エクスチェンジが構築されるケースも増えてきている。

ATM を用いた場合、Ethernet や FDDI などの LAN 技術を用いる場合とは異なり、容易に広域分散環境においてインターネット・エクスチェンジを構成することが可能であり、データ転送速度も Ethernet や FDDI に比較して容易に高速化が可能であることなどから、拡張性の高いインターネット・エクスチェンジが構築できると期待されている。

ただし、ATM 網上でインターネット・エクスチェンジを構成する場合、相互接続を行なうネットワーク間で PVC (permanent vertual circuit) を設定し、個別に経路情報の交換を行なう方法が一般的である。

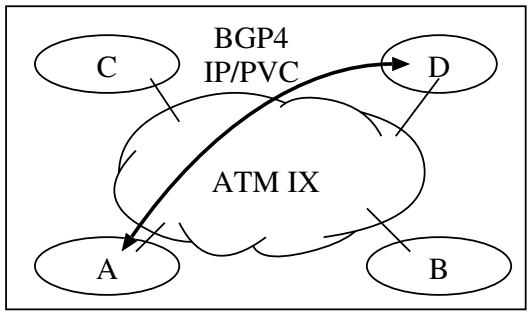

図 2.4 ATM による IX

この場合、各ネットワーク間の接続はデータリンク層で個別に確立・運用されるため、ネットワーク間で個別に経路制御を行なうことにより、それぞれのポリシを反映させることは容易であるが、各ネットワークでは相互接続ネットワーク数に比例した PVC の設定が必要であり、大規模なインターネット・エクスチェンジでは、管理・運用コストも膨大になる。

2.2.5 インターネット・エクスチェンジにおける経路制御とルートサーバ

多くのインターネット・エクスチェンジでは接続される個々のネットワーク間で個別のポリシを反映させるため、BGP4 (Border Gateway Protocol version 4)[46] による経路制御を行っている。この場合、ひとつのネットワークが他のネットワークとトラフィックの交換を行なうためには、それらのすべてのネットワークと BGP4 による経路情報の交換を行なう必要がある。接続組織の数を N とした場合、ひとつのネットワーク当たりの運用コストは $O(N)$ 、また、インターネット・エクスチェンジ全体での運用コストは $O(N^2)$ に相当することになり、運用上の負荷は膨大なものになる。また、接続組織数が多くなってきた場合、BGP4 を処理する各ルータの負荷も増大することになる。

一部のインターネット・エクスチェンジはこうした運用コストやルータの負荷の問題を解決するため、ルートサーバ (Router Server) を導入している。ルートサーバは、Ethernet や FDDI などの共有型データリンク層のネットワークで構成されるインターネット・エクスチェンジ上に設置され、そこへ接続されるネットワークとの経路制御を一括して行なう。この際、経路情報データベース (Routing Regisotry) を参照することにより各ネットワークの運用ポリシ

を反映し、運用負荷、あるいは各ルータの経路情報の処理負荷を最小限に抑えながら経路制御を行うことが可能になる。

ただし、ATM によるインターネット・エクスチェンジでは各ネットワーク間は PVC による個別の接続を行っており、全体で共通したネットワークが存在しないため、ルートサーバを用いて統括的な経路制御を行うことは難しい。

2.3 ラベルスイッチ技術を用いたインターネット・エクスチェンジ

本節ではラベルスイッチ技術を応用した、分散型のインターネット・エクスチェンジの構築手法を提案する。

本稿で提案する手法では、各ネットワークのポリシをラベルに対応付けることによりデータグラムの転送にポリシを反映させる。また、ラベルスイッチ技術により相互接続の設定を自動化し、ATM 上で構築されたインターネット・エクスチェンジであっても、その運用コストを最小限に抑えることが可能になる。

なお、本稿では、特に ATM によるインターネット・エクスチェンジ上でラベルスイッチ技術を応用する手法について述べるが、ラベルスイッチ技術は ATM 以外のデータリンク層技術でも利用が可能であり、将来的には POS や WDM などの次世代データリンク技術へも容易に応用が可能である。

2.3.1 ラベルスイッチを用いた IX モデル

図 2.5 はラベルスイッチを用いたインターネット・エクスチェンジの構成図を示している。Edge は各ネットワーク組織の管理下におかれ、インターネット・エクスチェンジ内の LSR (Label Switching Router) Core と接続される。

Edge-Core 間の接続に ATM 網を利用する場合、この間には PVC による直接接続を行い、LDP(Label Distribution Protocol)[47] 等によりラベルに関する情報を交換することになる。

なお、各ネットワーク間のトラフィック交換、すなわち Edge-Edge 間の通信においては、ラベルスイッチにおけるカットスルー (Cut Through) 技術によって高速なトラフィック交換を行う。

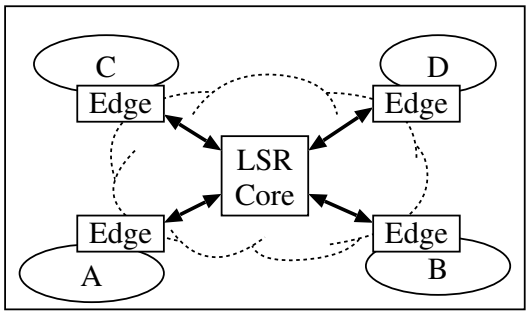

図 2.5　ラベルスイッチを用いた IX

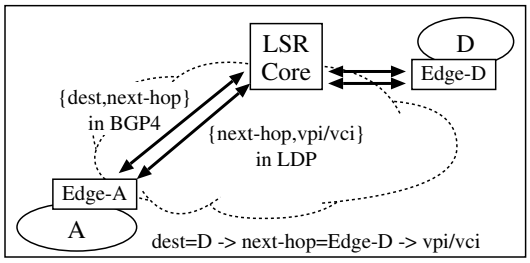

図 2.6　ラベルスイッチを用いたデータグラムの転送

labelfig:ix-lsr-send

2.3.2 経路情報の交換

ATM PVC によるインターネット・エクスチェンジでは、各ネットワーク間の相互接続は個別に行われるため、それぞれにどのようなアドレスが付与され、また、各リンクが現在利用可能かどうかをインターネット・エクスチェンジ側で把握することは難しい。

一方、本稿で提案されるインターネット・エクスチェンジでは各ネットワークの Edge ルータはインターネット・エクスチェンジ内の LSR Core と IP 層で接続されるため、各 Edge のアドレスや、リンクの接続性情報などを LSR Core で一括して管理が可能である。本モデルでは、これらの特徴を活かして、LSR Core にルートサーバ機能を実装し、Edge-Core 間では BGP4 および LDP による制御を行う。

BGP4 による経路情報には宛先アドレスと、それに対応する属性が含まれる。特に属性には Next-Hop アドレスが定義されており、Edge ルータは、この Next-Hop アドレスを用いて、その宛先へ向けたデータグラムを転送する、対向の Edge を決定する。

LDP によるラベルの対応情報は Next-Hop アドレスと、それに対応するラベル (ATM の場合 VPI/VCI) の対で表現される。実際にデータグラム

を転送する場合には、宛先アドレスから Next-Hop を検索し、さらに Next-Hop に対応するラベルを用いてインターネット・エクスチェンジ上でカットスルーを行なうことになる。

本稿で提案するインターネット・エクスチェンジでは、このように、ラベルスイッチ技術により BGP4 で取得した Next-Hop 情報をラベルに対応付けることにより、トラフィック交換におけるポリシの反映を可能にしている。

2.3.3 ラベルスイッチを用いたインターネット・エクスチェンジの特徴

本稿では、インターネット・エクスチェンジにおいてラベルスイッチを適用する手法について提案したが、本手法は次のような特徴を持つ。

1. 各 Edge において必要な設定は Core との間でのラベルスイッチの設定、および BGP4 の設定のみであり、管理・運用コストを低減させることができる。
2. 実トラフィックはラベルスイッチ技術により Edge-Edge 間で直接転送するため、カットスルーが可能であり、高速なトラフィック交換が可能である。
3. ATM, POS, WDM などの技術に対応可能であり、広域分散環境での相互接続、あるいは超高速化時代への対応も可能である。

2.3.4 既存のインターネット・エクスチェンジ技術との比較

本節では、ラベルスイッチ技術を応用したインターネット・エクスチェンジの利点を既存のインターネット・エクスチェンジ技術と比較する。

1. Layer 3 IX との比較。
   Layer 3 IX では、データグラムの転送はインターネット・エクスチェンジ内のルータに依存し、ポリシを反映させることができないため、ネットワーク毎にポリシが異なる場合などには適用できない。本稿で提案するインターネット・エクスチェンジではポリシをラベルに対応させることにより、基本的な接続や設定は Layer 3 と同等でありながら、各ネットワークのポリシを反映させたデータグラムの転送が可能である。
   また、ラベルスイッチを用いたインターネット・エクスチェンジでは、実トラフィックはカットスルー技術により極めて高速なパケット転送が可能である。
2. Ethernet, FDDI による IX との比較。
   LAN 技術を用いた場合、分散環境でインターネット・エクスチェンジを構成することが難しい。そのため、各ネットワークから接続拠点にルータを持ち込む必要があり、各ネットワークの運用負荷はさらに増大する。本稿によるインターネット・エクスチェンジは広域分散環境での相互接続を前提としており、各ネットワークの Edge ルータはそれぞれのネットワーク内に設置することも可能である。
   また、LAN 技術を用いた場合、Ethernet で 10Mbps〜1Gbps、FDDI で 100Mbps 程度が通信速度の上限になるが、ラベルスイッチ技術を応用する場合、ATM で 155Mbps〜2.4Gbps、POS で 10Gbps 程度など、次世代の超高速通信技術でも容易に応用が可能である。
3. 既存の ATM IX との比較。
   ATM PVC によるインターネット・エクスチェンジでは PVC の設定は接続する相手ごとに設定する必要があるため、その設定、管理、運用コストは、各ネットワーク当たり $O(N)$ になる。本稿におけるインターネット・エクスチェンジの場合、PVC の設定はラベルスイッチにより、経路制御の設定はルートサーバ機能により自動化されるため、個別の設定は 実質的にこれらのコストは $O(1)$ で抑えられる。

2.4 まとめ

本稿ではラベルスイッチ技術を応用したインターネット・エクスチェンジの構築手法を提案した。本提案によるインターネット・エクスチェンジは次のような特徴を持つ。

- 各ネットワーク組織の管理・運用コストが最小限に抑えられる。
- 広域分散環境において相互接続が可能である。
- カットスルーによるデータグラムの転送により、高速なトラフィック交換が可能である。

- ATM、POS や WDM などの技術へも応用可能である。

今後は、実証実験を通して性能評価等を行なうとともに、複数ルータの実装に関する相互接続試験などを行なうことを予定している。

第 3 章　OSPF リンクステートデータベースを用いたフローアグリゲーションの実現

3.1 背景

インターネットにおけるバックボーンインフラストラクチャの広帯域化によって，従来のネットワークトラフィックと異なる特性を持つ映像や音声などの広帯域ストリーム型トラフィックの利用が増加している．

このようなトラフィックの増加によって，ルータの高機能化，高性能化の必要性が高まった．

ラベルスイッチ技術はこのような要求に対応する技術の一つとして開発され標準化されようとしている．

ラベルスイッチ技術は，Layer 3 パケットストリームを固定長ラベルに対応付けすることで実現される．ラベルスイッチ技術を用いたラベルスイッチルータ (LSR) は，Layer 3 パケットを従来の Layer 3 アドレス情報を用いて転送するだけでなく，Layer 3 アドレス情報に対応させた固定長ラベルによってパケット転送する．

LSR では，1 つのパケットストリームに対して，1 つのラベルが必要である．必要なラベル数は，ラベル割り当てトリガとパケットストリームの粒度によって決まる．

現在，主に 2 つのラベル割り当てポリシーが提案/実装されている．1 つは，IFMP (Ipsilon Flow Management Protocol)[48] や FANP (Flow Attribute Notification Protocol)[49, 50] で定義され，トリガとなるパケットが来た時にラベルを割り当てる (トラフィックドリブン)．そのラベルに流すパケットストリームは，同一送信元アドレスと宛先アドレスを持つもの (ホストペア) である．もう 1 つのラベル割り当てポリシーは，TDP (Tag Distribution Protocol)[51] で用いられている．これは，パケットストリームの粒度として宛先ネットワークを選び，ラベル割り当てトリガとして経路情報が追加された時にラベルを割り当てるトポロジードリブン方式を使用している．両方式とも ISP などの大規模なネットワークのバックボーンで使用するとラベル数が多く必要であることが報告されている．

Layer 2 技術として ATM を用いた場合，ATM インターフェースのリアセンブルバッファ容量に限界があるために，利用可能なラベル数に制限がある．

本論文では，始点ルータが OSPF[52] リンクステートデータベースを用い，パケットが通過する終点 EdgeLSR を検知することによって，トポロジードリブン/始点-終点 EdgeLSR のラベル割り当て方式を実現する．この手法を用いることで EdgeLSR 数が N であるとき常に LSR 網の総 LSP (Label Switched Path) 数を $N*(N-1)$ におさえられる．また，この総 LSP 数は $O(N^2)$ で増加するため，上記の両方式よりも必要ラベル数を大幅に削減できる．

3.2 LSR の概要

LSR の概要を図 3.1 に示す．隣接ルータとして LSR だけと接続する LSR を内部 LSR(CoreLSR) と呼び，LSR 以外のルータと接続する LSR を境界 LSR(EdgeLSR) と呼ぶ．

CoreLSR(図 3.1:LSR) では，通常のルータと同様に Layer 3 のパケットヘッダを見て転送する Layer 3 転送 (図 3.1:L3 Forwarder) モジュールと，Layer 3 のパケットヘッダを見ることなく，固定長ラベル (ATM の場合，VPI/VCI) によりパケットを転送する Layer 2 転送 (図 3.1:L2 Forwarder) モジュールによって構成される．

EdgeLSR(図 3.1:Edge1, Edge2) では，Layer 3 転送モジュールと CoreLSR と接続可能なネットワークインターフェースによって構成される (また後述の LDP が実装されていなくてはならない)．

通常のルータと同じ Layer 3 転送は，デフォルト

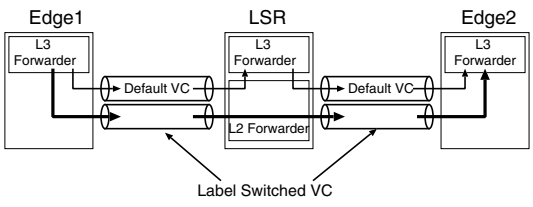

図 3.1　LSR の概要

VC を利用して行われる．Edge1 から送信したパケットは，デフォルト VC を通して，LSR で受信され，LSR では， Layer 3 転送モジュールでパケットヘッダの宛先アドレスをキーに IP ルーティングテーブルを検索し，Edge2 へデフォルト VC を通してパケットを送信する．

LSR の特徴である layer 2 転送は，まずパケットを送信する前に，あるラベルに流すパケットストリームを LDP(Label Distribution Protocol) を用いて隣接ノード間で通知する．例えば，図 3.1 の場合，Edge1 と LSR，LSR と Edge2 の間で LDP を動作させる．LDP を用いてラベルスイッチ VC に Edge 2 宛のパケットストリームを流すことを通知する．LSR は，通知したラベルを持つパケットを受信した場合，Layer 2 転送モジュールのみで転送できるようにラベルを Layer 2 モジュールに記憶する．

上記の通知が終わったあと，Edge1 から送信された Edge 2 宛のパケットは，通知してあるラベルをもつラベルスイッチ VC を通して，LSR に送られる．LSR は，L2 転送モジュールで，受信したパケットのラベルを検索し，予め登録されているものなので，そのラベルをもとに Edge 2 へのラベルスイッチ VC にパケットを転送する．これにより，Layer 3 転送処理を行わずに，パケットを転送できる．このように，ラベルスイッチ VC を通して，Layer 2 処理のみでパケットが転送される経路を LSP(Label Switched Path) と呼ぶ．

必要なラベル数は，次の 2 つの要素に依存する．

1. ラベル割り当てトリガ
2. パケットストリームの粒度

3.2.1 ラベル割り当てトリガ

LSP を制御するトリガとして，トラフィックドリブンとトポロジードリブンの 2 つがある．

トラフィックドリブン

トラフィックドリブンでは，ある特定のパケットを隣接 LSR に送信する時をラベル割り当ての契機とする．LSP を動的に設定するため，ラベル数は，そのときに存在しているパケットフロー数だけで十分である．一方，LSP を設定する間は，Layer 3 転送を行わないといけないため，その間，LSR の特徴である layer 2 転送は行えない．

なお，LSP にトラフィックが一定期間流れないときに解放される．

トポロジードリブン

トポロジードリブンでは，経路エントリが構築されたときに，その経路エントリにしたがって，データパケットが流れる前に予め LSP を設定する．利点と欠点は，トラフィックドリブンのほぼ逆となる．

なお，LSP に対応する経路エントリが削除された場合に，その LSP は解放される．

3.2.2 パケットストリームの粒度

LSP に流すストリームの種類としては，例えば，(送信元アドレス, 宛先アドレス) が同一のパケットを流すことが考えられる．この他にこれより粒度を荒くしたストリームや粒度を細かくしたストリームが考えられる．粒度を細かくすると必要なラベル数は多くなり，粒度を荒くすると逆にラベル数は少なくなる．本論文では粒度を荒くすることをフローアグリゲーションと呼ぶ．

ストリームの種類として以下のものが考えられる．上から順番に粒度が荒くなるため，必要なラベル数が少なくなると考えられる．以下の宛先ネットワークとは経路表に登録されているネットワークを指し，CIDR 環境では，様々なプレフィックス長を持つネットワークが存在することになる．

1. (送信元アドレス, 宛先アドレス) が同じストリーム．ホストペアと表現する．
2. (送信元アドレス, 宛先ネットワーク) が同じストリーム．
3. (送信元ネットワーク, 宛先ネットワーク) が同じストリーム．
4. (宛先アドレス) が同じストリーム．
5. (宛先ネットワーク) が同じストリーム．
6. (始点 EdgeLSR, 終点 EdgeLSR) が同じストリーム．同一の始点 EdgeLSR と終点 EdgeLSR を通過するストリーム．

(始点 EdgeLSR, 終点 EdgeLSR) が同じストリームの粒度で LSP を設定する場合には，各始点 EdgeLSR で経路表から LSR 網を通過して転送されると判断

できる宛先アドレスと通過する終点 EdgeLSR の対応情報が必要となる．

3.3 OSPF リンクステートデータベースを用いたフローアグリゲーション

本章ではリンクステート型経路制御プロトコル OSPF 内の情報を用いて，3.2.2 章の (始点 EdgeLSR, 終点 EdgeLSR) が同じストリームの粒度で LSP を設定するために必要な情報を取得する手法について議論する．

以下では，LSP とパケットストリームとの対応情報を管理し，ラベル情報を隣接 LSR と交換するモジュールを LDP モジュールと呼び，OSPF を利用して経路表の経路を操作するモジュールを OSPF モジュールと呼ぶ．

同一の OSPF エリアに所属する LSR 網において，各始点 EdgeLSR 内の OSPF モジュールは共通のリンクステートデータベースの情報を保持する．OSPF モジュールは，リンクステートデータベースから SPF 木 (Shortest Path First Tree) を作る経過で，他の EdgeLSR があること認識し，その EdgeLSR から先の経路情報を取得する．OSPF モジュールは LDP モジュールに対して，同じ EdgeLSR を通過する宛先アドレス情報の集合を渡す．

EdgeLSR の LDP モジュールは，他の EdgeLSR に対し 1 つの LSP を設定し，その LSP に対して，OSPF モジュールから渡された宛先アドレス情報の集合を対応付けする．

以上の操作によって，各終点 EdgeLSR へは一つのみ LSP が設定され，(始点 EdgeLSR，終点 EdgeLSR) が同じストリームの粒度というフローアグリゲーションが実現される．

このフローアグリゲーションを実現するために必要な機能を以下に示す．

1. 他の EdgeLSR の OSPF ルータ ID を検知する機能
2. OSPF モジュールから LDP モジュールへ通知する情報の定義と通知する機能
3. OSPF モジュール内に特定のルータを通過する経路を探索する機能

3.3.1 他の EdgeLSR の OSPF ルータ ID を検知する機能

各始点 EdgeLSR は，特定のルータの経路を探索するために，他の EdgeLSR の OSPF ルータ ID を保持する必要がある．

他の EdgeLSR の OSPF ルータ ID を取得するために，以下に示す手法が考えられる．

静的な設定

各 EdgeLSR において，他の EdgeLSR 全ての OSPF ルータ ID を静的に保持する．各 EdgeLSR は静的に保持しているため，LSR 網に変更があった場合 (新たな EdgeLSR が加わったり，既存の EdgeLSR が消失した場合) には，全ての EdgeLSR の保持する OSPF ルータ ID 表を変更しなければならない．

LDP の拡張

LDP を拡張し，LSR 網上の EdgeLSR の OSPF ルータ ID を交換する機能を追加する．各 EdgeLSR はこの拡張された LDP を通して他の EdgeLSR の OSPF ルータ ID を取得する．LSR 網の変更にも動的に対処できる．

Opaque LSA の利用

Opaque LSA[53] を利用し，各 EdgeLSR に他の EdgeLSR の OSPF ルータ ID を通知する．LSR 網の変更にも動的に対処できる．この場合，Opaque LSA を利用できる OSPF モジュールが必要である．

3.3.2 LDP モジュールに渡すべき情報

OSPF モジュールから LDP モジュールに対して通知するメッセージを以下に示す．

ADD

特定の EdgeLSR への LSP に対して新たな経路を追加する．以下の情報が必要になる．

1. 宛先ネットワークアドレス
2. 宛先ネットワークアドレスマスク
3. 終点 EdgeLSR の OSPF ルータ ID
4. 終点 EdgeLSR の LSR ネットワーク側ネットワークインターフェースのアドレス
5. その宛先ネットワークアドレスへのネクストホップアドレス，または出力ネットワークインターフェース情報

DELETE

無くなった経路を削除する．以下の情報などの経路と LSP を一意に識別できる情報が必要になる．

1. 宛先ネットワークアドレス
2. 宛先ネットワークアドレスマスク
3. 終点 EdgeLSR の OSPF ルータ ID

OSPF モジュールから ADD メッセージを受けとった LDP モジュールは，終点 EdgeLSR の OSPF ルータ ID から，その EdgeLSR への LSP を検索する．LSP が設定されていない場合，終点 EdgeLSR の LSR ネットワーク側ネットワークインターフェースのアドレスに対して設定する．検索された LSP に対して，宛先ネットワークアドレスと宛先ネットワークマスクの情報を対応付ける．

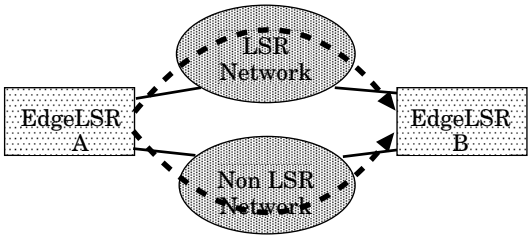

図 3.2 他の経路のある LSR ネットワーク例

OSPF の同一エリアに所属する，図 3.2 に示すネットワークにおいて，EdgeLSR A は EdgeLSR B に対して，LSP を設定する．EdgeLSR A は EdgeLSR A から EdgeLSR B への LSR 網以外を通る経路を EdgeLSR B への LSP に対応付けするべきでない．この場合 LDP モジュールが宛先ネットワークアドレスへのネクストホップアドレス，または出力ネットワークインターフェース情報を参照し，LSP に対応付けするべきか判断する．

3.4 特定のルータを通過する経路の探索手法

本章では，OSPF を利用して，同一エリア内の特定のルータを通過する経路の探索手法について議論する．

OSPF モジュールは，リンクステートデータベース内に保持される LSA(Link State Advertizement) を用いて，自分自身 (OSPF モジュールが動作するルータ) が根 (root) ノードとなる SPF 木を構成する．

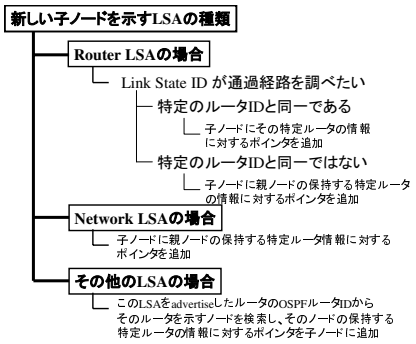

図 3.3 通過経路情報を取得するアルゴリズム

OSPF エリア内トポロジは Router LSA と Network LSA によって表現され，Summary LSA(Type 3, 4) や AS External LSA は他のエリアや AS の経路情報を示す．

SPF 木を構成する際にはまず，自分自身を示すルータ ID を持つ Router LSA を利用して，根ノードを作る．最も小さいリンクから順に，対応したリンクを持つ Router LSA と Network LSA を順に子ノードとして親ノード (既に SPF 木にある，対応したリンクを持つノード) に追加して行き，自分自身が所属するエリアのトポロジに関しての SPF 木を構成する．

Area 境界ルータ (ABR: Area Border Router)，または AS 境界ルータ (ASBR: AS Border Router) を示す Router LSA に対応するノードは，子ノードとして Summary LSA(Type 3, 4) や AS External LSA に対応したノード[1]を持つ．

ノードに，特定のルータの情報を参照するためのポインタを保持させた場合，新しい子ノードを親ノードに追加する際に図 3.3 に示すアルゴリズムを追加することで，そのルータを通過する経路を導き出すことができる．

Router LSA を保持する子ノードは，その Router LSA の Link State ID が特定の OSPF ルータ ID であるか調べ，一致すればその情報を子ノード内に追加する．一致しない場合と Network LSA を保持する子ノードの場合，親ノードの特定ルータの情報を

[1] 実際に SPF 木のノードとして表現されるかは実装依存であるが，ここではアルゴリズムを簡略化するためにノードとして表現した．

継承し，子ノード内に保持する．その他の LSA を保持する子ノードの場合, その LSA を広告したルータを示すノードを検索し，そのノードの保持する特定ルータの情報を継承し，子ノード内に保持する．

OSPF によって経路表に追加された経路に関しては，その経路の根拠となった LSA を持つ SPF 木のノードを参照することによって，特定のルータを通過するか判断できる．一方特定のルータの情報として，この情報を参照しているノードを全て管理することで，特定のルータを通過する全てのノードの参照もできる．

3.5 gated を用いた実装

実装は，ラベルスイッチ技術として Cell Switching Router(CSR)[49] を利用し，OSPF モジュールとして gated-3.5.9 を利用した．各 EdgeLSR が終点 EdgeLSR の OSPF ルータ ID を保持する機構として，静的な設定を用いた．また，gated と fanpd(CSR 上の LDP モジュール) の通信インターフェースに UNIX ドメインソケットを利用した．図 3.4 に実装環境を示す．

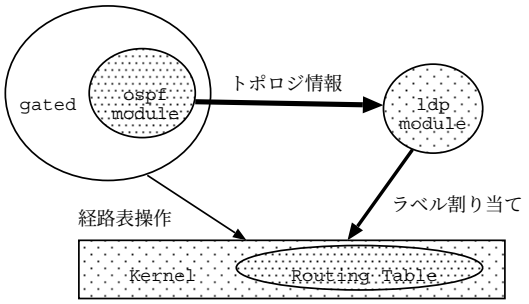

図 3.4　gated と fanpd の通信の様子

3.6 評価

1998 年 9 月に行われた WIDE Project 合宿において，3.5 章で実装した LSR を利用し，本論文の手法の有効性を評価する実験を行った．このネットワークのトポロジを図 3.5 に示す．ネットワークには 5 台の EdgeLSR(図 3.5: Edge) が CoreLSR(図 3.5: Core) に接続し，LSR 網を構成している．このネットワークの利用ホスト数は約 130 ホストであった．

このネットワークを利用し，トポロジドリブン/始点-終点 EdgeLSR に対して LSP を設定するポリシーで LSR 網を運用し，総 LSP 数を測定した．

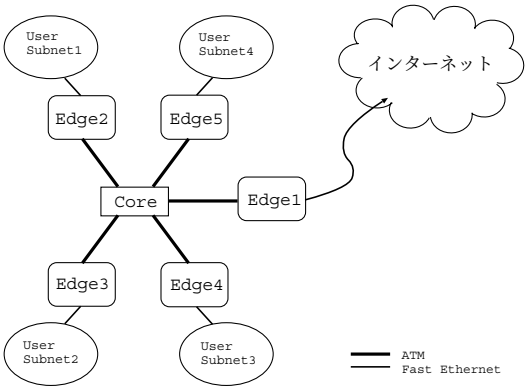

図 3.5　合宿ネットワークトポロジ

実験結果を図 3.6 に示す．EdgeLSR 数は 5 であり，この LSR 網では常に 20 本の LSP が必要であった．本論文の手法を用いることによって，必要ラベル数を常に $EdgeLSR$ 数 $*$ ($EdgeLSR$ 数 -1) におさえられることが分かった．

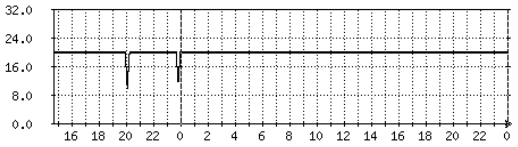

図 3.6　実験における総 LSP 数の推移

3.7 おわりに

本論文では，OSPF の情報を能動的に用いることでフローアグリゲーションする手法を設計・実装した．さらに本手法を用いて，実環境において必要ラベル数の測定をした．EdgeLSR 数が 5 である LSR 網において，必要ラベル数は常に 20 であった．

この手法によって，EdgeLSR 数が N である LSR 網において必要なラベル数を $N*(N-1)$ にできることが分かった．

今後は，動的な EdgeLSR の OSPF ルータ ID の検知機能や，OSPF トポロジ情報を利用するための汎用なインターフェースについて研究していく予定である．

第 4 章　標準 ATM 上のラベルスイッチ技術 (VCID 通知手順)

4.1 背景

IETF MPLS WG で標準化が行われているラベルスイッチ技術は、固定長のラベルによって、高速・高機能なパケット転送を行う。このラベルスイッチ技術は、様々なデータリンク層 を扱えるように設計されている。ATM を用いた ATM-LSR (ATM Label Switching Router) は、ラベルスイッチ技術を使った主要なルータである。この ATM-LSR は、IETF で標準化されたラベルスイッチ技術を ATM 上で使用する場合の名称である。

ラベルスイッチ技術では、2 つの隣接するルータ間でラベル配布プロトコル を用いて情報交換を行う。このラベル配布プロトコル では、転送パケットの集合を表す「パケットストリーム 」とそのパケットの集合に割り当てる「ラベル」の対応関係を隣接するルータ間で交換するためのプロトコルである。ラベル配布プロトコル では、このプロトコルで交換するラベルが隣接のルータ間で同じことを前提として作られている。ATM の場合は、ラベルは VPI/VCI であり、このラベルは隣接ルータ間の ATM スイッチによって、書き換えられるため、隣接ルータ間で同じ値を取るとは限らない。そこで、現状のラベルスイッチ技術で ATM を使う場合は、隣接ルータ間に ATM スイッチを置かない構成か、あるいは、隣接ルータ間の ATM スイッチで、VPI/VCI を書き換えないという制限を設けている。このようなリンクを「透過型リンク 」と呼ぶことにする。この制限のために従来のラベル配布プロトコル は、標準 ATM 上でラベルスイッチ技術を用いることができない。そこで、標準 ATM 上でラベルスイッチ技術を用いることができる新たな概念を提案する。

一般的な ATM 上でラベルスイッチ技術を用いるために、VCID (Virtual Connection IDdentifier) の導入を提案する。この VCID は、ATM VC に対してその VC の両端のルータ間で同一の ID となるように割り当てる。一般的に ATM VC の両端のルータでの VPI/VCI は異なる。それぞれのルータにおいて、VPI/VCI と VCID の対応関係を記憶し、ラベル配布プロトコル で交換するラベル (VPI/VCI) の代りに VC の両端で値が等しい VCID を用いることで、現状のラベル配布プロトコル を変更することなく、ラベルスイッチ技術を標準 ATM 上で利用することができるようになる。この概念とプロトコルは、ラベルスイッチ技術の標準として IETF(Internet Engneering Task Force) で採り入れられている [54]。

以下の節では、隣接ルータ間で 1 つの ATM VC に対して同一の VCID を割り当て、その VCID と VPI/VCI の対応関係をとるための VCID 割り当てプロトコルを提案する。また、そのプロトコルが実現可能性を評価するために、実装を行い、それを測定する。

表 4.1　VCID 通知手順とリンクタイプ

リンクタイプ	VCID 通知なし	インバンド	アウトバンド (small)	アウトバンド
透過型リンク	○	–	–	–
VP	○	○	–	–
PVC	–	○	–	–
SVC	–	○	○	○

4.2 VCID 通知手順概要

ATM-LSR システムは、種々のデータリンク層を扱うことができる。例えば、透過型リンク や VP, PVC, SVC である。ここで透過型リンク とは VC の両端で同一の VPI/VCI を持つものと定義する。例えば、2 つのノードが途中に ATM スイッチなしに接続されている場合や同一の VPI を持つ VP で接続されている場合を指す。

以下では、それぞれを説明する。

VCID 通知手順には 2 つの手順がある。インバンド方式とアウトバンド方式である。インバンド方式では、VCID 通知手順の制御メッセージがラベルスイッチ VC 上に流れる。アウトバンド方式では、VCID 通知手順の制御メッセージがラベルスイッチ VC 以外の VC に流れる。例えば、ATM シグナリング VC やデフォルト VC 上である。

以下にリンクタイプとそのリンクで用いる VCID 通知手順をまとめる。VCID 通知手順の詳細は、次の節で述べる。また、表 4.1 に、リンクタイプ毎に適用できる VCID 通知手順をまとめる。○印がついている部分が適用可能な領域である。

- 透過型リンク：VCID 通知手順不要
  VCID 通知手順は、必要ない。なぜならば、ラベル (例えば、VPI /VCI) が VC の両端で同じだからである。

- VP:
  VP の場合は、複数の VCID 通知手順を選択することができる。それぞれを以下に列挙する。

  – インバンド通知手順
    インバンド通知手順は、VC の両端で VPI の値が異なるときに用いる。それぞれの VC に対して、インバンド通知手順を行う。

  – VPID 通知手順
    VPID 通知手順は、VC の両端で VPI の値が異なるときに用いる。VP に対して 1 つの VPID 通知手順を行う。

  – VCID 通知手順不要
    隣接ノードに対して、1 つの VP のみが設定されている場合は、VCID 通知手順は、必須ではない。VCI を VCID として使用することができる。なぜならば、VCI の値が VP の両端で同一だからである。

- PVC: インバンド通知手順
  PVC の場合は、VC の両端で VPI/VCI が一般的に異なるので、インバンド VCID 通知を適用する。

- SVC:
  SVC の場合は、複数の VCID 通知手順を選択することができる。

  – アウトバンド通知手順
    ATM シグナリングメッセージが VCID を運ぶための十分なフィールドを持っている場合 (例えば、GIT フィールド [55]) は、この方法を適用する。

  – 小さいフィールドを使用したアウトバンド通知手順
    ATM シグナリングメッセージが VCID を運ぶために十分なフィールドを持っていない場合は、この方法を適用する。

  – インバンド通知手順
    ATM シグナリングメッセージがユーザ情報を運ばない場合、この方法を適用する。

4.3 インバンド通知手順

4.3.1 ポイントポイント VC の場合のインバンド通知手順

VCID 通知制御メッセージは、新しく LSP として使われる ラベルスイッチ VC 上で交換される。ラベルスイッチ VC は動的に VC が割り当てられる SVC と静的に割り当てられる PVC/VP の場合がある。図 4.1 は、インバンド通知手順のメッセージ交換を示している。ノード A とノード B の間の VCID 通知は以下の通りである。

1. ノード A が ラベルスイッチ VC をノード B へ設定する。(シグナリングで動的に設定するか、管理者が静的に設定する。)
2. ノード A が VCID の値を選択する。
3. ノード A は VCID 値とメッセージ ID を持つ VCID 制御メッセージ (VCID PROPOSE) を新しく設定した ラベルスイッチ VC を通してノード B に送信する。
4. ノード A は、新しく設定した ラベルスイッチ VC の出力ラベル (VPI /VCI) と選択した VCID 値の対応を記憶する。
5. ノード B は ラベルスイッチ VC からメッセージを受信するとメッセージ中の VCID とその制御メッセージの受信入力ラベル (VPI /VCI) との対応関係を記憶する。ノード B が LDP 要求メッセージを受信するまで、ノード B は、VCID PROPOSE メッセージを除いてその VC に流れるパケットを廃棄する。
6. ノード B は、ACK メッセージをノード A に送る。このメッセージは、受信したメッセージ中の VCID とメッセージ ID を含む。このメッセージは、LDP の制御用 VC で運ばれる。
7. ノード A が ACK メッセージを受信すると、VCID とメッセージ ID が登録されているものと同一かチェックする。もし、同じであれば、ノード A は、ノード B が ラベルスイッチ VC と VCID の対応関係を登録したと認識する。そう

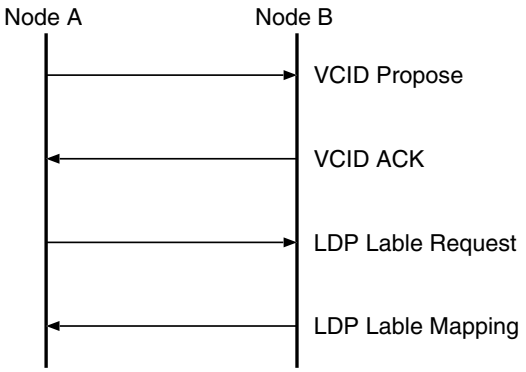

図 4.1 透過型リンク時のインバンド VCID 通知手順

でなければ、メッセージを無視する。もし、ノード A が一定時間内に期待している VCID とメッセージ ID を持つ ACK メッセージを受信しない場合は、VCID PROPOSE メッセージをノード B に再送する。

8. PROPOSE ACK メッセージを受信後、ノード A は、LDP REQUEST メッセージをノード B に送信する。そのメッセージは、VCID PROPOSE メッセージで使用したメッセージ ID を含む。ノード B が LDP REQUEST メッセージを受信すると、ノード A が PROPOSE ACK を正常に受信したと認識できる。この 3 つのメッセージ交換 (VCID PROPOSE, VCID ACK, LDP REQUEST) は 3 ウェイハンドシェークとなる。VCID PROPOSE メッセージがメッセージロスのある方式で転送されるので、3 ウェイハンドシェーク方式が要求される。3 ウェイハンドシェークが終わると、ノード B は、その VC から受信する全ての VCID PROPOSE メッセージを無視する。ノード B は、LDP に従って、VCID 値を含む LDP マッピングメッセージをノード A に送信する。

4.3.2 ポイントマルチポイント VC 接続時のインバンド通知手順

VCID 通知の制御メッセージは、新しく LSP として利用するポイントマルチポイントのラベルスイッチ VC を通して交換される。この手順は、マルチキャストパケット転送に適用できる。ポイントマルチポイント VC に 1 つの VCID を割り当てるために、上流ノードが VCID を割り当てる。上流ノードが 1 つで下流ノードが 2 つの場合のポイントマルチポイント VC を利用する場合の手順を以下に記述する。2 つの手順を記述する。1 つ目は、最初の下流ノードがマルチキャストグループに参加したときである。2 つ目は、新しい下流ノードが上記の既存マルチキャストグループに参加するときである。図 4.2 はポイントマルチポイント VC を利用する場合の VCID インバンド通知手順のメッセージ交換を示している。

まず、最初の VC の設定手順を説明する。これは、最初の下流ノードがマルチキャストグループに参加して時である。

1. 上流ノードが VCID を新しい ラベルスイッチ VC に割り当てる。
2. 上流ノードは、VCID と上流ノードのアドレスを含んだ VCID 制御メッセージを LSP として使う ラベルスイッチ VC を通して送信する。
3. 上流ノードは、 ラベルスイッチ VC のための出力ラベルと VCID の対応関係を記憶する。
4. VCID 制御メッセージを受信した下流ノードは、上流ノードのアドレスをチェックする。もし、上流ノードのアドレスが自分のアドレスと同一のネットワークプレフィックスを持たない場合は、VCID 制御メッセージを破棄する。そうでなければ、下流ノードが制御メッセージ中の VCID とその制御メッセージを受信した ラベルスイッチ VC の対応関係を記憶する。
5. 下流ノードは、VCID ACK メッセージを上流ノードに通知する。
6. 上流ノードが VCID ACK メッセージを受信すると、上流ノードは、下流ノードと同一の VCID を共有することができる。

次に、ポイントマルチポイント VC が存在している場合に、新たにマルチキャストグループに下流ノードが参加する場合を説明する。

1. 上流ノードは、ATM シグナリングを用いて下流ノードを追加する。
2. ポイントマルチポイントの ラベルスイッチ VC に既に割り当てている VCID を使う。
3. 上流ノードは、VCID と上流ノードのアドレスを含む VCID 制御メッセージを LSP として使うポイントマルチポイントの ラベルスイッチ VC を通して送信する。なお、この VCID 制御メッ

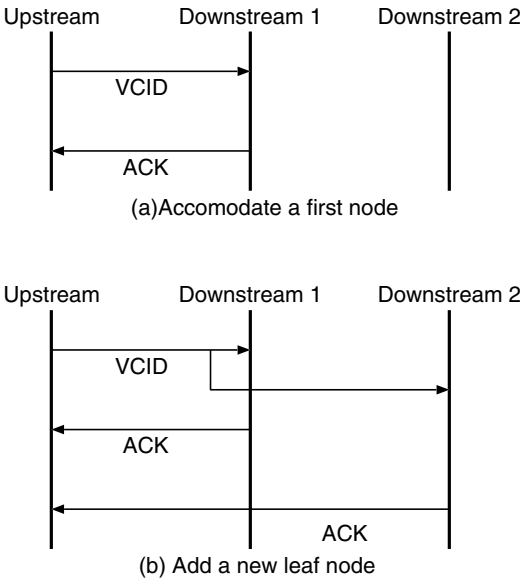

図 4.2　ポイントマルチポイント VC 利用時のインバンド通知手順

セージは、全ての下流ノードに届く。

4. VCID 制御メッセージを受信した下流ノードは、メッセージ中の上流ノードのアドレスを調べる。上流ノードのアドレスが下流ノードのアドレスと同じネットワークプレフィックスを持たない場合は、VCID 制御メッセージを廃棄する。そうでない場合は、下流ノードは、制御メッセージ中の VCID とそのメッセージを受信した ラベルスイッチ VC の対応関係を記憶する。
5. 下流ノードは、上流ノードに VCID ACK メッセージを送信する。
6. 上流ノードが VCID ACK メッセージを受信しすると、上流ノードとポイントマルチポイント VC に接続する全ての下流ノードの間で VCID が共有される。

4.3.3 小さいフィールドを使うアウトバンド通知

この手法は、 ラベルスイッチ VC が ATM シグナリングで設定され、ATM シグナリングメッセージが VCID を運ぶだけの大きなユーザ情報を運ぶフィールドを持たない場合に使う。

ATM フォーラム UNI 3.1/4.0 の SETUP メッセージは、ユーザ情報を運ぶ 7 ビットの必須フィールドを持つ。これは、BLLI IE(Broadband Low Layer Information-Information Element) 中の第 3 層プロトコルフィールドのユーザフィールドである。

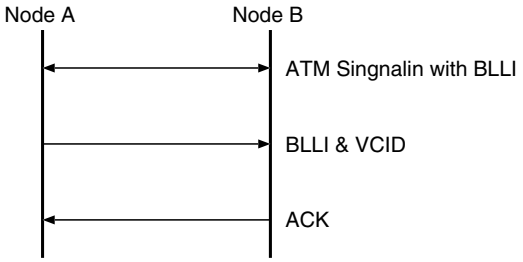

図 4.3　小さいフィールドを使うアウトバンド通知手順

BLLI は、VCID 通知手順の実行中の ラベルスイッチ VC の一時的な ID として使う。この手順を小さいフィールドを使うアウトバンド通知手順とする。ID の衝突を避けるために、他の VC の VCID 手順で使用中の BLLI 値は、新しい ラベルスイッチ VC に割り当ててはいけない。BLLI と VCID と VC の対応関係が確定すると、BLLI 値は新しい ラベルスイッチ VC に使用するために解放される。図 4.3 は、小さいフィールドを使うアウトバンド通知手順を示している。

ポイントマルチポイント ラベルスイッチ VC は、ATM シグナリングの ADD_PARTY メッセージを使って設定される。ADD_PARTY メッセージは、既存のポイントマルチポイント VC に新しいリーフを追加する。この手順で使う ADD_PARTY メッセージ中の BLLI 値は最初にポイントマルチポイント ラベルスイッチ VC を設定したときの値と同じものを使わなくてはならない。これは、ATM シグナリングで規定されているものである。複数のポイントマルチポイント ラベルスイッチ VC が存在していて、それぞれの ラベルスイッチ VC が使用する BLLI 値が同じ場合は、新しい下流ノードを同時に追加することはできない。一方の VCID 通知手順が終了したあとに、もう一方の VCID 通知手順を実行する。

4.3.4 ポイントポイント VC 使用時のアウトバンド通知手順

この節では、ポイントポイント ラベルスイッチ VC を設定する場合の手順を説明する。

1. 上流ノードは、ATM シグナリングを用いて下流ノードへのラベルスイッチ VC を設定する。ATM シグナリング中の BLLI フィールドには、他の VCID 手順で使用していない BLLI 値を選

び、それを記入する。
2. 上流ノードは、BLLI と VCID 値の対応関係を記入した VCID PROPOSE メッセージを下流ノードに送信する。
3. VCID PROPOSE メッセージを受信した下流ノードは、メッセージ中の BLLI 値を持つ ラベルスイッチ VC とメッセージ中の VCID の対応関係を記憶する。
4. 下流ノードは、VCID ACK メッセージを上流ノードに送信する。
5. 上流ノードが VCID ACK メッセージを受信すると、ラベルスイッチ VC に VCID が割り当てられたことになる。この状態の ラベルスイッチ VC はラベルスイッチング用に使用できる。このとき、BLLI 値は解放され、新たな VCID 通知手順で使用することができるようになる。

4.3.5 ポイントマルチポイント VC 使用時のアウトバンド通知手順

この節では、マルチキャスト用にポイントマルチポイントのラベルスイッチ VC を最初に設定する手順とその ラベルスイッチ VC に新たなノードが追加されたときの手順を説明する。

この手順では、上流ノードが VCID と BLLI 値を決定する。

まず、 ラベルスイッチ VC を設定する手順を説明する。

1. 上流ノードが BLLI 値と VCID 値を選ぶ。
2. 上流ノードは、BLLI 値を ATM シグナリングメッセージに入れてラベルスイッチ VC を設定する。
3. 上流ノードは、BLLI 値と VCID 値を含む VCID PROPOSE メッセージを下流ノードに送信する。
4. 下流ノードは、受信したメッセージの BLLI 値を持つ ラベルスイッチ VC と VCID の対応関係を記憶する。
5. 下流ノードは、VCID ACK メッセージを上流ノードに送信する。
6. 上流ノードが、VCID ACK メッセージを受信すると、ラベルスイッチ VC は使用可能状態になり、VCID 通知手順で使用した BLLI 値は解放され、他の ラベルスイッチ VC の VCID 通知手順で使用可能になる。

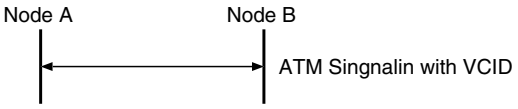

図 4.4 アウトバント通知手順のメッセージシーケンス

次に、上記で説明した、ポイントマルチポイント用 ラベルスイッチ VC に新たなノードを追加する場合の手順を説明する。

1. 上流ノードは、新しい下流ノードをポイントマルチポイント VC に参加させるために、ラベルスイッチ VC を ATM シグナリングを用いて設定する。このとき、ATM シグナリングメッセージで、BLLI を運ぶ。もし、他の ラベルスイッチ VC が同じ BLLI 値を使っている場合は、上流ノードは、その BLLI 値が使い終わるまで、VCID 通知手順を保留する。
2. 最初の ラベルスイッチ VC を設定する手順 3 から同様に行う。

4.4 アウトバンド通知手順

アウトバンド通知手順は、ラベルスイッチ VC を確立するときに ATM シグナリングを使用して、ATM シグナリングが VCID を運ぶのに十分なフィールド (例えば、GIT[55]) を持つ場合に使用する。図 4.4 はアウトバンド通知手順のメッセージのやり取りを示している。メッセージフォーマットの詳細は、[55] に書かれている。

ポイントマルチポイント ラベルスイッチ VC は、ATM シグナリングの ADD_PARTY メッセージを使って設定することができる。ADD_PARTY メッセージは、すでに存在している ラベルスイッチ VC に新しいラベルスイッチ VC のリーフノードを追加することができる。VCID は ADD_PARTY メッセージの GIT フィールドで運ばれる。

4.5 ラベル設定時間及び VCID 通知手順の評価

この章では、LSP 確立までの設定時間の評価を行う。ラベルスイッチルータの 1 種である、CSR プロトタイプシステムを実装し、そのプロトタイプを使

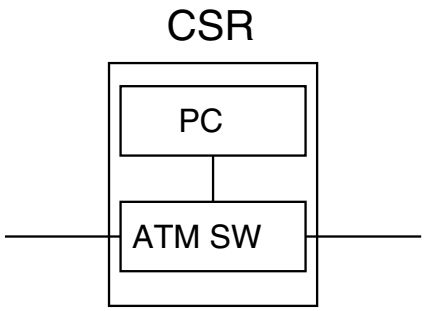

図 4.5　CSR プロトタイプシステム

図 4.6　VCID 通知手順測定ネットワークトポロジー

用して実測した。測定項目は、ラベル配布プロトコルに必要なラベル割り当て時間と ATM シグナリングを用いた場合の VCID 通知時間の両方を測定した。

4.5.1 CSR プロトタイプシステム

VCID 通知手順を LSR の 1 つの実装形態である CSR プロトタイプシステムに実装した [49](図 4.5)。この CSR は、インテル社製 Pentium PC とフォアシステム社製 ASX-200WG の ATM スイッチで構成される。VCID 通知手順とラベル配布プロトコルの FANP[50] がこの CSR プロトタイプシステム上に実装されている。

VCID 通知手順と FANP のソフトウェアは C 言語を使って書かれている。VCID 通知モジュールのソースコードは、457 行であり、FANP モジュールは 6742 行である。VCID 通知手順のソフトウェアモジュールは、ラベル配布プロトコル と比べて、7%の大きさしかない。ソースコードの観点から見ると、VCID 通知手順は、小さいモジュールだと考えられる。

4.5.2 LSP 確立パフォーマンス

評価ネットワークトポロジーを図 4.6 に示す。R1 と R2 は CSR のエッジノードである。それぞれインテル社製 Pentium PC を使って作成し、VCID 通知手順と FANP が搭載してある。CSR は、インテル社製 Pentium PC とフォアシステム社製 ASX-200WG の ATM スイッチで構成されるプロトタイプシステムである。CSR とエッジノードは、東芝社製 AX1500 の ATM スイッチで接続される。

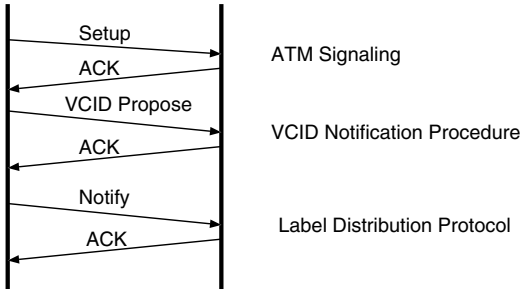

図 4.7　LSP 確立手順

この構成において、LSP 確立手順としてもっとも複雑な ATM SVC を使ったときの隣接ルータ間で動くインバンド VCID 通知手順に必要な時間を測定した。SVC を用いて、R1 が R2 へ VC を設定する。LSP 確立時間は、(1)ATM シグナリング、(2)VCID 通知手順、(3) ラベル配布プロトコルで構成される。以下で、それぞれの手順を説明する (図 4.7)。

1. ATM シグナリング手順
   R1 が R2 に対して LSP を確立したい場合は、R1 が ATM シグナリングを用いて隣接ルータである CSR に対してラベルスイッチ VC 用の VC を設定する。
2. VCID 通知手順
   R1 が VCID インバンドメッセージを CSR に送信する。CSR は VCID ACK メッセージを R1 に送信する。
3. ラベル配布プロトコル
   R1 がラベル配布プロトコル メッセージを CSR に送信する。CSR はラベル配布プロトコル ACK メッセージを R1 に送信する。

R1 と CSR へのメッセージ交換を R1 において測定することにより、上記のそれぞれの手順で必要な時間を測定した。表 4.2 は、それぞれの手順に必要な時間の測定結果を示している。ATM シグナリング手順は 88 ミリ秒必要であり、VCID 通知手順は 3 ミリ秒必要であり、ラベル配布プロトコル は 4 ミリ秒必要である。合計すると LSP 確立時間は、95 ミリ秒である。もし、R1 が VCID 通知手順を使わなかったとすると、LSP 確立時間は、92 ミリ秒必要となる計算である。

ATM シグナリングを用いる SVC の環境で、VCID 通知手順を使った LSP 確立時間は、VCID 通知手順

表 4.2 LSP 確立時間の評価

手順	時間 [ミリ秒]
ATM シグナリング	88
VCID 通知	3
ラベル配布プロトコル	4

を使わない場合の LSP 確立時間に比べて、3%必要な時間が増えるだけである。

4.6 まとめ

この章では、標準 ATM 上でラベルスイッチ技術を使うために VCID の概念を導入し、VCID 通知手順を提案した。透過型リンク あるいは単一 VP リンクの場合は、VCID 通知手順は必要ない。しかし、複数 VP や PVC や SVC のリンクでは、VCID 通知手順が必要である。複数 VP や PVC では、インバンド VCID 通知手順を使う。SVC の場合は、3 つの VCID 通知手順が許されている。

さらに、VCID 通知手順を含む LSP 確立手順を評価した。プロトタイプシステムを実装し、そのシステムを使って測定した。VCID 通知手順は 3 ミリ秒であった。ATM シグナリングを用いる SVC の環境で、VCID 通知手順を使った LSP 確立時間は、VCID 通知手順を使わない場合の LSP 確立時間に比べて、3%必要な時間が増えるだけである。

第 5 章　IETF MPLS WG への技術貢献

IETF のラベルスイッチ技術に関する WIDE プロジェクトの成果として、標準 ATM 上でラベルスイッチ技術を利用するためのプロトコル (VCID 通知手順) が上げられる。この Internet Draft は、標準トラックとして last call が終了し、Proposed Standard として RFC 化されることが決まった。