

第 6 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 プロジェクトで現在、IETF(Internet Engineering Task Force)において、国際標準化に向けての検討が推進されている ラベルスイッチ (Label Switch) 技術に関する研究を、1996 年 4 月よりワーキンググループ (LAST-WG; LAbel Switch Technology Working Group) として正式に開始した。 IETF においては、1997 年 3 月に設立された MPLS-WG(Multi-Protocol Label Switching Working Group) により推進されている。 WIDE プロジェクトにおけるラベルスイッチ技術に関する先駆的な研究開発および実証実験に基づいたシステム評価を、約 1 年以上前に着手していたことになる。 この先駆的な研究開発および実証実験に基づいたシステム評価の成果を元に、IETF MPLS-WG においては、積極的な貢献活動を継続して行っている。

LAST-WG においては、IETF MPLS-WG における標準化の基礎的なアーキテクチャである 2 つの技術に関する研究開発活動ならびにシステム評価活動を行ってきた。 具体的には、株式会社東芝殿と株式会社日立製作所殿との共同実験としてセルスイッチルータ (CSR; Cell Switch Router) 技術に関する研究、そして株式会社 CISCO System 殿との共同実験としてタグスイッチ (Tag Switch) 技術の研究を行なっている。

今回の報告書では、以下の内容に関する報告を行う。

1. セルスイッチルータ (CSR) システムにおけるラベルスイッチ転送の性能評価
2. ラベルスイッチシステムに必要なラベル数の評価
3. CSR システムの IPv6 化に関する研究と実験運用結果
4. ラベルスイッチシステムのシステムアーキテクチャに関する研究
5. IETF MPLS-WG への技術貢献

性能評価は、WIDE 研究会における実験ネットワークや、WIDE インターネットバックボーンにおける 実トラフィックを用いて評価解析し、実ネットワーク環境におけるラベルスイッチ技術の適応性および導入形態に関する研究を行った。

第 2 章

ラベルスイッチ技術の概要

2.1 ラベルスイッチルータ技術の動作原理

インターネットおよび計算機技術の急速な発展により、インターネットが処理しなければならないパケット量は、指数関数以上の速度で増加している。ラベルスイッチ技術は、現在、IETF の MPLS(Multi-Protocol Label Switching) ワーキンググループにおいて検討が進められている技術で、高性能・高機能のパケット転送を実現することができる。

ラベルスイッチ技術を適用したルータ(ラベルスイッチルータ; LSR)では、IP パケット流に対して固定長のラベルを割り当て、割り当てた固定長ラベル情報を用いて、パケットの転送処理を行う。固定長ラベルとしては、データリンクヘッダ内の情報(例えば、ATM (Asynchronous Transfer Mode) における VPI/VCI 値、あるいは、フレームリレーにおける DLCI 値)を用いる方法と、データリンクヘッダと IP ヘッダの間にシムタグフィールド(Shim Tag Field; 図 2.1 参照)と呼ばれる固定長ラベルを書きこむことのできるヘッダを挿入する方法の 2 つの方法がある。固定長ラベルのスイッチングをハードウェア処理などの方法を用いて高速処理可能なモジュールを用いることができれば、従来のソフトウェアによるパケット処理に比べて、高速大容量のパケット処理を実現することができる。代表的なラベルスイッチルータの実装として、タグスイッチ [RFC2105] やセルスイッチルータ (CSR) [RFC2098] が上げられる。これらは、固定長のラベルとして、VPI/VCI 値を用いており、ATM スイッチで使用されているセルスイッチモジュールを、ラベルスイッチモジュールとして使用することで、高速大容量のパケット処理を実現している。ラベルスイッチルータ技術は、IETF の MPLS (Multi Protocol Label Switch) ワーキンググループ('97 年 3 月 Routing Area の WG として発足)において、仕様・プロトコルの標準化検討が進められている。

ラベルスイッチルータ (LSR) を用いたネットワークでは、現在のインターネットモデル CATENET モデルを守り、ルータでのパケット転送情報として、レイヤ 3 情報 (e.g., IP アドレス) より下のレイヤの情報 (以下ラベルと呼ぶ) を利用する。ATM スイッチをパケットの転送エンジン (e.g., ラベルスイッチエンジン) として利用する場合には、ラベルとして VPI/VCI が使われる。すなわち、レイヤ 3 情報をラベル (e.g., データリンクヘッダ) にマッピングして、ラベル情報を用いて実際のパケット転送を行う。ラベルに対応するレイヤ

3 情報としては、アプリケーションフロー (e.g., 動画フロー) から宛先ネットワーク (e.g., Address Prefix) まで、さまざまな粒度のパケットフローを定義することができる。ラベルにマッピングするレイヤ 3 のパケットフローを FEC (Forwarding Equivalence Classes) と呼ぶ。LSR では、FEC とラベルをマッピングするためのプロトコル (以下ラベルマッピングプロトコルと呼ぶ) が必要となり、現在いくつかの方式が提案されている。ラベルマッピングプロトコルにより、ラベルと FEC とのマッピングが確立すると、レイヤ 3 情報 (e.g., IP アドレス) を用いてパケット転送を行う必要がなくなり、IP アドレスのルーティングエントリを検索することなく、ラベル情報のみでパケットの転送を行うことができる (これをカットスルー転送呼ぶ)。カットスルーパスの経路、すなわち実際にパケットが転送される経路は、ルーティングプロトコルで示される経路と常に同一である。すなわち、カットスルーパスが、ラベルバインディングプロトコルにより確立されたあとでも、ネットワークの状況の変化に伴い、ルーティングプロトコルにより指示される経路が変化した場合には、エンドエンドのアプリケーションが継続中でも、カットスルーパスの経路も自動的にルーティングプロトコルにより指示される経路と同じ経路に変更される。既存のルータでは、ユーザデータパケットの転送処理とネットワーク制御機能処理 (例えばルーティングプロトコル処理) とを、共通のプロセッサを用いて実行するのが一般的で、ユーザデータパケットの増加により、プロセッサが過負荷状況に陥り、ネットワークの制御機能が適正に動作しなくなる場合があった。LSR システムでは、ユーザパケット処理とネットワーク制御機能処理を、分離して処理することが可能となり、より信頼性の高いルータシステムを構築できる。

2.2 ラベルマッピングプロトコル

ラベルスイッチ技術を適用するためには、IP アドレス情報を固定長ラベルにマッピングする必要があり、このマッピング情報を隣接 LSR 間で共有する必要がある。ラベルに対応させる IP パケット流としては、個別のアプリケーションフロー (例えば、動画フロー) のような細かなフローから、宛先ネットワーク (Network Address Prefix) が同じパケット流のような集約 (Aggregate) されたフローまで、さまざまな、集約度・粒度のパケット流に対して、ラベルを割り当てることができる。パケット流の集約はシステムの大規模化のために必要であり、一方、個別パケット流へのラベルの割り当てはユーザあるいはアプリケーションごとの細かな品質制御のために必要となる。LSR 間で、FEC とラベルのマッピングを確立するためのプロトコルが必要となるが、これを、IETF MPLS ワーキンググループでは、(Label Distribution Protocol) と呼んでいる。上述の LSR の例では、タグスイッチでは TDP (Tag Distribution Protocol) が使用され、CSR では FANP (Flow Attribute Notification Protocol) [RFC2129] が使用されている。

2.2 に、セルスイッチモジュールを用いた LSR システムにおけるパケット転送の概念図を示した。LDP により、ラベル (VPI/VCI 値) と FEC のマッピングが確立すると、IP アドレ

スを用いてパケット転送 (図中では Default VC を用いたパケット転送) を行う必要がなくなり、IP アドレスのルーティングエントリを検索することなく、ラベル情報 (VPI/VCI) のみでパケット転送を行うことができるようになる。ラベルを用いてパケット転送を行うことのできる経路 (パス) を、ラベルスイッチパス (LSP; Label Switched Path) と呼ぶ。LSP は、ルーティングプロトコルが指示する経路と常に同一である。したがって、経路の変更が発生した場合には、エンドエンドのアプリケーションが継続中でも、自動的に、LSP は新しい経路と同じ経路に変更される。

2.3 ラベルマッピング確立ポリシー

ラベルと FEC のマッピング確立 (i.e., ラベルスイッチパス LSP の確立) のポリシーとしては、実際のパケットの到着を契機とする Data-driven、ルーティング情報のようなコントロール情報を用いる Control-driven、さらに、運用上で必要となるノード間に必要に応じて LSP を生成するエンジニアリングドリブンの 3 つの動作モードがある。具体的には、ラベルスイッチパス (LSP) を確立する契機としては、次の 4 つのものがある。

1. データドリブン

実際のパケットの到着を契機に、LSP の確立を行うかどうかをオンディマンド判断する。各アプリケーションあるいは宛先ごとに排他的に通信資源を割り当てることができるので、RSVP などの資源予約プロトコルがなくても排他的に高品質な通信品質を提供することができる。カットスルーパスは、ルーティングドメインに関係なく生成することができ、ルーティングドメインをまたがったカットスルーパスを提供することができる。また、帯域の確保をデータパケットの実際の到着により実行するので、Topology-Driven のみでの運用と比較して、トータルで必要となるルータ間の帯域幅を小さくすることが可能となる。一方、Data-Driven のみで運用した場合には、多数のラベルが必要となる。ラベルは有限な資源であり (ATM の場合には VPI/VCI 値)、特にバックボーンで必要となるラベル数が大きくなってしまう。また、ラベルの生成の遅延 (ATM では VC の設定遅延) も問題となる場合もある。さらに、ある割合のパケット (セッションの最初のパケットや短い寿命のセッションのパケット) は、通常のソフトウェア処理によるホップホップでのパケット転送を行う必要がある。

2. 制御メッセージドリブン

(a) トポロジードリブン

ルーティングプロトコルで得られるネットワークのトポロジー情報 (ネットワークアドレス情報あるいはボーダルータアドレス) をもとに、LSP の確立を行う。すなわち、ルーティングプロトコルにより生成されるルーティングエントリを FEC とする方式で、LSR で新しいルーティングエントリが生成された場合に、カットスルーパスを生成すべきルーティングエントリ先に対応するノード

(e.g., ボーダルータ) にカットスループスの生成を行う。Topology-driven の適用により、必要となるラベルの数を少なくすることができる。カットスループスは、実際のパケットの到着を待たずに確立されるので、セッションのすべてのパケットをカットスルー転送することができる。宛先ネットワーク (i.e., Address Prefix) 行きのアプリケーションフローがすべて共通のカットスループスを共有するので、アプリケーションごとに排他的に通信資源を割り当てることができず、ルーティングドメインのボーダルータでは通常のソフトウェアによるホップホップでのパケット転送を行う必要がある。さらに、Topology-driven にみでの運用を行った場合には、帯域の確保をデータパケットの実際の到着により行わないので、Data-Driven のみでの運用と比較して、トータルで必要となるルータ間の帯域幅が大きくなってしまう。

(b) 予約メッセージドリブン

RSVP (Resource ReSerVation Protocol) のような、資源予約用の制御メッセージをネットワークに通知するパケット流に対して、資源予約制御メッセージ (例えば、RESV メッセージ) を用いて LSP の確立を行う。なお、ラベルと FEC のマッピングを、資源予約制御メッセージに相乗り (Piggy Back) させることも可能となっている。

3. エンジニアリングドリブン

ネットワークの良好な運用のために、手動設定により、特定のパケット流に対して LSP の設定を行います。LSP の確立により、同一宛先への複数の経路を容易に作りだすことができるようになる。LSP の確立の方法には、ルーティング制御が指示する経路と同じにする方法と、絶対経路 (Source Routing あるいは Explicit Routing) を指定する方法とがある。複数の経路を提供する目的としては、以下の 2 つのものがあげられる。

- 負荷分散

OSPF などでは、同じコストの経路についてのみ複数の経路を提供することができるが、複数の経路に負荷を良好に分散させることは容易ではない。また、複数の経路を提供可能なのは、同じコストの経路が存在するときのみであり、一般的には、1 つの宛先に対して、唯一の経路しか提供されない。複数の経路にパケットを分散させることで、中継ノードでのパケット処理の分散を行うことができるようにできる。

- CoS/QoS 提供

ノード間に複数の経路を提供し、各経路に対して、特定の CoS/QoS を提供することができるので、より容易に、同じ宛先に対して、ことなる CoS/QoS 品質を提供することができる。

表 2.1: ATM スイッチを用いた LSR の機能比較

	ARIS	CSR	IP Switch	Tag
Data-Driven	No	Yes	Yes	No
Control-Driven	Yes	Yes	No	Yes
ATM I/F	No	Yes	No	No

2.4 ラベルスイッチルータ (LSR) の実現方法

高性能なラベル情報を用いてデータを転送するエンジンを利用することができれば、より高性能な LSR を実現させることが可能となる。ATM スイッチは高性能、スケーラブルかつ安価なラベルスイッチエンジンと見ることができ、LSR のラベルスイッチエンジンとして ATM スイッチを利用する方式がいくつか提案されている。表に ATM スイッチを用いた LSR の機能比較を行った。

ARIS、IP スイッチ、タグスイッチでは、データリンクラベルが各レイヤ 2 ノードで書き換え (swapping) が行われない Transparent なリンクを用いなければ相互接続できない。つまり、VPI/VCI を FEC にマッピングするラベルとし、ATM スイッチを LSR のラベルスイッチエンジンとして利用しているのであるが、LSR 間には標準の ATM スイッチを設置することができず、したがって、ATM リンクを使ったネットワーキングを行うことができない。さらに、ATM スイッチをスイッチハブとして利用し、例えば高性能な ATM サーバークラスタを構成した場合に、LSR と ATM サーバークラスタを高性能に相互接続をすることができない。

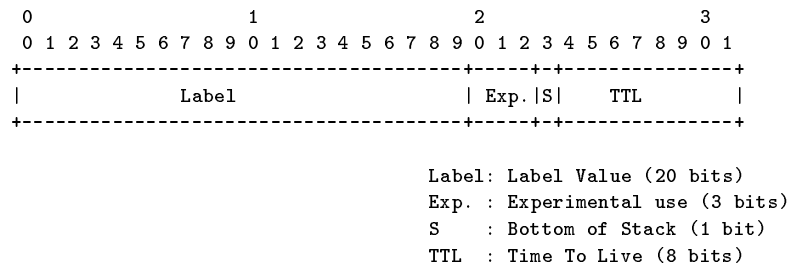


図 2.1: Shim Header フォーマット

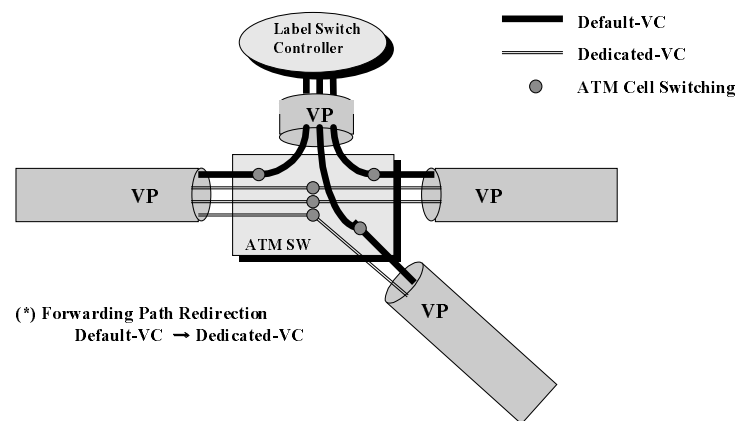


図 2.2: セルスイッチモジュールを用いた LSR システムの概念図

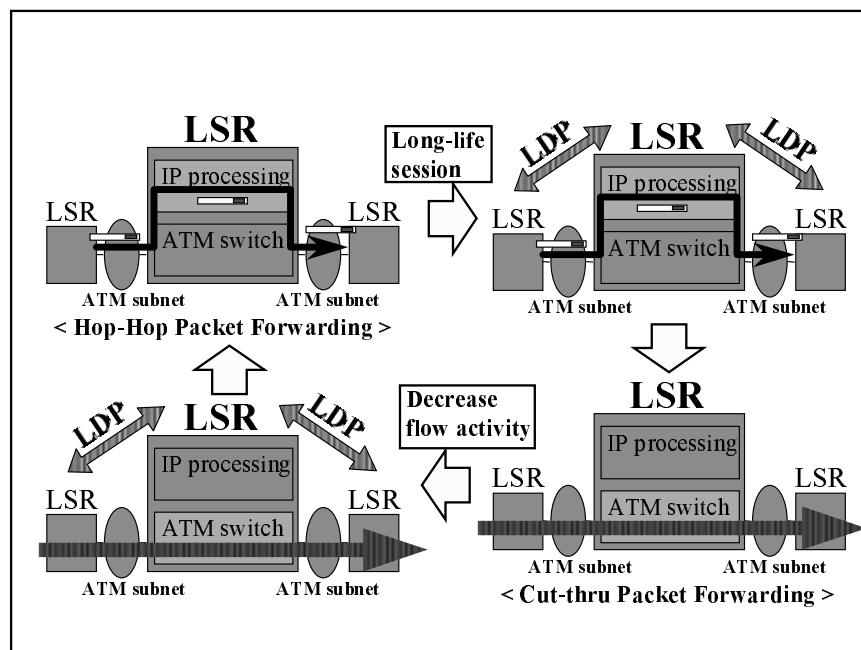


図 2.3: データドリブンでのパケット転送

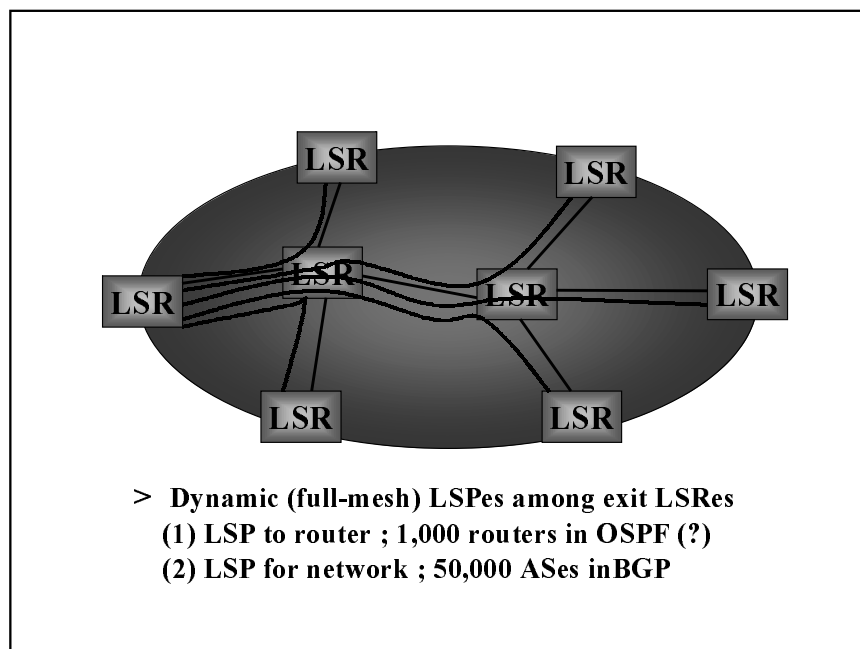


図 2.4: Topology-Driven での LDP の生成

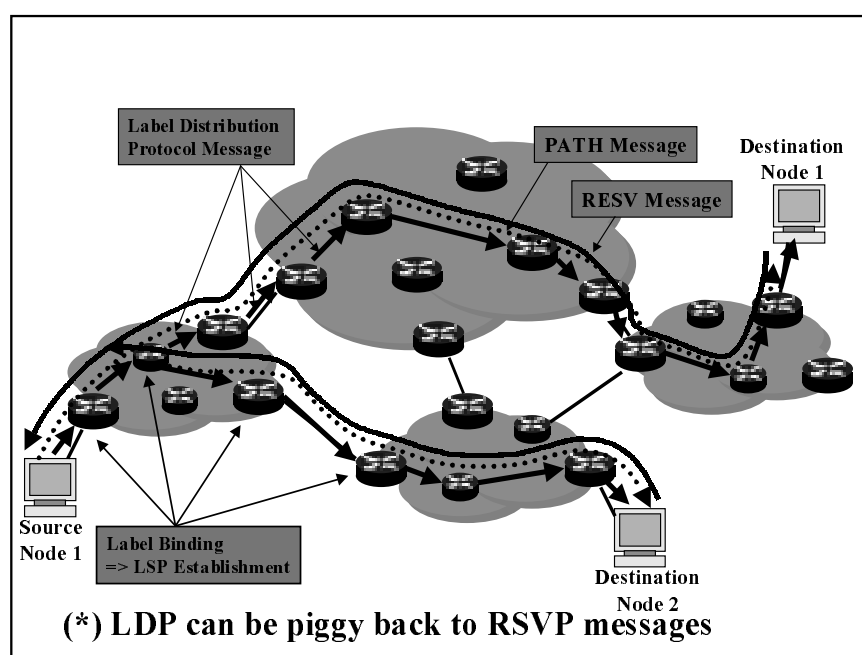


図 2.5: Reservation Driven での動作例

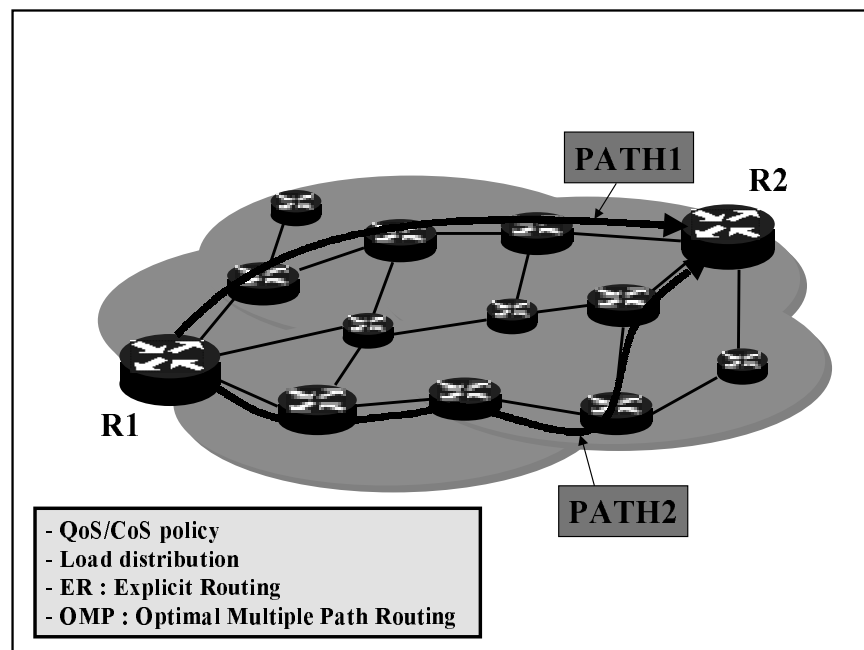


図 2.6: Engineering Driven での動作例

第 3 章

ラベルスイッチルータ技術の今後の課題と方向性

3.1 マルチプロトコル対応 (レイヤ 2 / レイヤ 3)

ラベルスイッチ技術は、ATM だけではなく、すべての L2 技術 (e.g., フレームリレー) およびすべての L3 技術 (e.g., IPX) に適用可能である。現在は、主に L2 として ATM/SONET、L3 として IP の場合の検討が行われているが、今後はその他のデータリンクへの適用方法の検討が必要である (e.g., 100M/ギガイーサネット)。現在、IETF ではイーサネットへの適用方式として、2 つの方式が議論されている。すなわち、IP ヘッダとイーサネットヘッダの間に固定長のラベルを挿入する方式と、ラベル情報を Encoding した Private イーサネット MAC アドレスを使用する方式である。現在のイーサネットスイッチを有効に利用することができれば、高性能で安価なイーサネットフレーム対応の LSR を構成することが可能となる。

3.2 カットスループス運用形態・ポリシー

Flow-driven、Topology-driven および Reservation-driven の 3 つのカットスループス確立ポリシーの実際のネットワークでの運用形態の検討・確立が必要である。LSR の設置場所およびハードウェアの制約を考慮にいたった検討が必要である。本報告書では、Flow-Driven と Topology-Driven の実トラヒックを用いた性能評価ならびに、必要なラベル数の評価を行っている。さらに、2 つの方式を融合することにより、より大規模化に適したラベルの割り当て方式を提案し、その評価を行っている。さらに、2 つの運用ポリシーの適用形態を、大規模なシステム構成の観点から検討している。

3.3 カットスループスの管理方法

IP レベルのフローのデータリンク VC へのマッピングルールの確立 (例えば、IP Integrated Service Model サービスの具体的なデータリンクへのマッピング)、および、カットスループスの管理・制御方式の確立が必要である。

3.4 ルーティンググループ形成への対応

LSR においては、従来のレイヤ 3 のパケット処理がカットスルーされるので、IP ヘッダ中の TTL(Time To Live) 値が減少されない。従来のルータ網では、TTL 機能によりパケットが永久にループすることを回避することができていた。したがって、ラベルマッピングプロトコル中に何らかのルーティンググループ検出 (回避) 機能を持つ必要がある。WIDE プロジェクトでは、簡単でかつ有効なルーティンググループの検出方式を研究開発し、これを IETF に提案している。

3.5 VPN(Virtual Private Network) 機能

LSP の中では、各 IP パケットの IP アドレスを用いてパケットの転送処理を行わないので、LSP を用いて転送される IP パケット流の IP アドレス (送信元 IP アドレス、宛先 IP アドレス) は、必ずしも、グローバルな IP アドレスである必要がなくなる。すなわち、3.1 に示したように、プライベート IP アドレスを用いた同じ VPN に属する複数のサイトを、アドレス変換なしに、相互接続することが可能となる。なお、ルーティングプロトコルは、IPv6 への対応や、IP 以外のプロトコルへの対応のために、マルチプロトコル対応化が行われており (OSPFv2 や BGP4+ など)、各 VPN ごとに異なるネットワークプロトコルを用いていると定義することで、プライベート IP アドレスを用いたインターネット上での経路制御を実現することができる。いったん、プライベート IP アドレスを用いて経路が決まった後は、LSP により、IP アドレス (プライベート IP アドレス) を用いず、ラベル情報 (例えば VPI/VCI 値) を用いて、プライベートアドレスを持った IP パケットを、インターネット上で転送することができるようになる。

3.6 通信品質制御

ラベルスイッチを適用したネットワークは、LDP を適用するコア部分にあるコアルータと、既存のネットワークとコア部分の境界に存在するエッジルータから構成される。エッジルータは既存のラベルスイッチ技術を適用していないネットワークと、ラベルスイッチ技術を適用したネットワークの間を相互接続する境界ルータとなる。このネットワーク構成は、通信品質の提供を行う Differentiated-Service のネットワークアーキテクチャモデルとほぼ同じである。

Differentiated-Service ネットワークも、エッジルータという概念を持っており、エッジルータにより、CoS/QoS 提供のための制御 (DS ビットの制御) を行う。Differentiated-Service ネットワーク内のすべてのルータは、DS ビットを用いてパケットの転送スケジューリングを行う。ところが、ラベルスイッチ技術と Differentiated-Service を融合することで、コアルータでのパケット転送のスケジューリング処理を、より、簡素化、大容量化、高速化すること

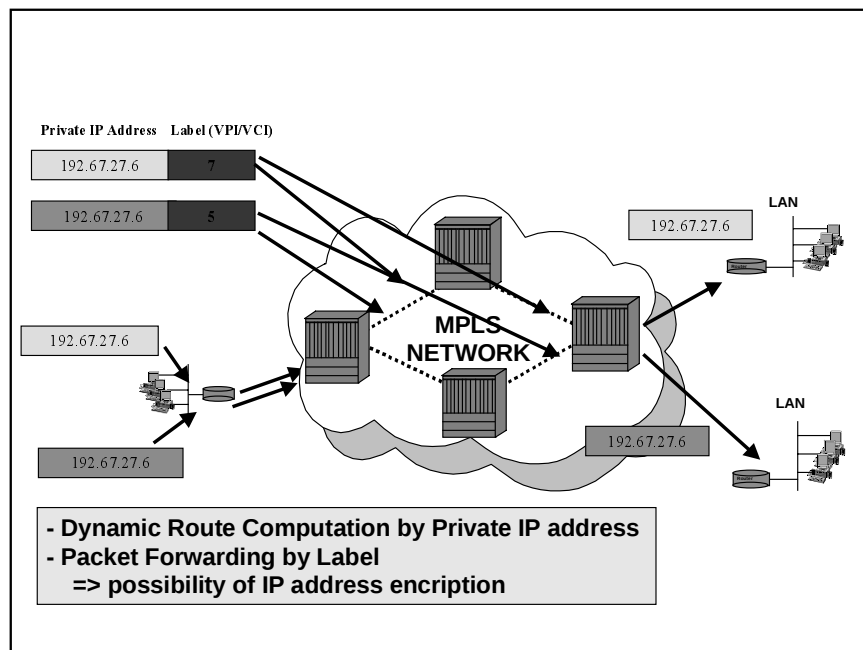


図 3.1: ラベルスイッチ技術を用いた新しい VPN の提供

ができるようになる。3.2に、ラベルスイッチ技術が、現在の典型的な ISP ネットワークに適用された場合の構成を示した。POP(Point Of Presence) ルータが、Differentiated-Service およびラベルスイッチ技術的にエッジルータとなり、コアネットワークでは、ラベルスイッチ技術が適用され、ラベルに Differentiated-Service の DS ビットにより処理ポリシーの指示がマッピングされることで、高速大容量の Differentiated-Service の提供が可能となる。

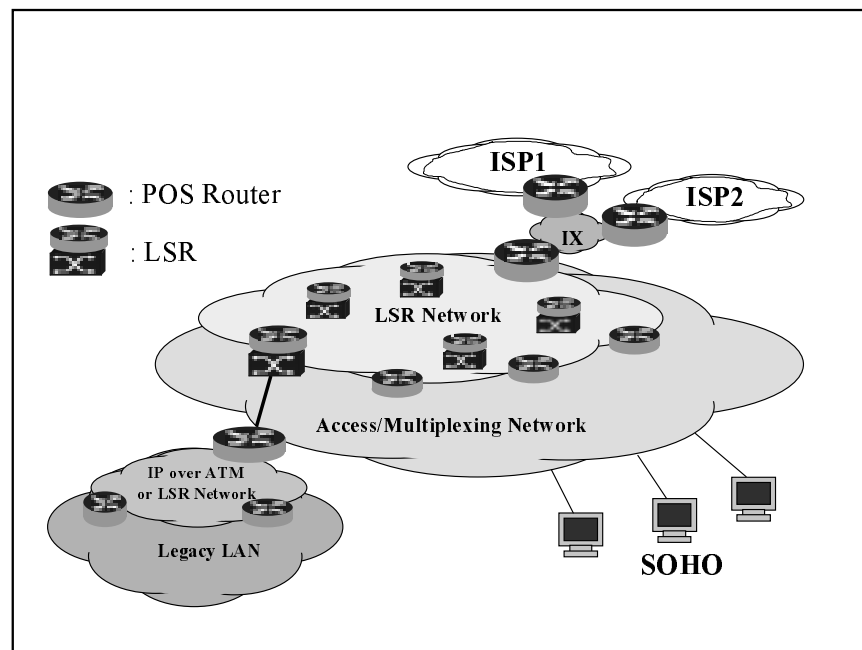


図 3.2: ラベルスイッチ技術と Differentiated-Service

第 4 章

セルスイッチルータ技術の概要

セルスイッチルータ (CSR; Cell Switch Router) は、ATM フォーラム準拠の ATM スイッチおよび ATM ネットワークとの相互接続が可能な、唯一のラベルスイッチ技術を適用したラベルスイッチルータ (LSR) である。CSR は、ATM セルをデータ転送の単位とし、ATM セルヘッダ中の VPI/VCI 値 (計 24 ビット) をラベルスイッチング用ラベルとして使用し、ATM スイッチモジュールをラベルスイッチモジュールとして利用します。CSR システムは、データパケットの実際の到着を契機としてラベルスイッチパス (LSP; Label Switch Path) の確立を行うトラヒックドリブンと、経路制御プロトコルから獲得するネットワーク情報を元に LSP の確立を行うトポロジードリブンを同時にサポートしている (デュアルモード運用)。さらに、マルチキャストパケットに対するラベルスイッチング、2 レベルのパケット転送品質 (CoS; Class of Service) の提供、およびシスコ社タグスイッチとの相互接続機能を実現することができる。今後、IPv6 対応、ラベルスイッチ技術を用いた新しい VPN 機能、RSVP を用いた Guaranteed Service の提供、さらには、Differentiated Service の提供が研究開発課題となる。以下が、CSR システムの特徴である。

1. ATM システムとの相互接続
2. ATM ネットワーク上でのラベルスイッチング
3. デュアルモード運用
4. マルチキャストパケットのラベルスイッチング
5. 2 レベルのパケット転送品質提供
6. タグスイッチとの相互接続機能

4.1 CSR の技術概要

CSR は、ハードウェアによるラベルスイッチングを行うセルスイッチモジュールと、従来のルータ機能およびセルスイッチモジュールの制御を行う IP コントローラモジュールの組み合わせから構成される。LSP の制御・管理は、東芝が独自に開発し IETF にも提

案されている、FANP(Flow Attribute Notification Protocol) を用いる。RFC2129 に記述されている FANPv1 (FANP version 1) は、トラヒックドリブンのみのサポートであるが、FANPv2(FANP version 2) の導入により、トポロジードリブン (=デュアルモード) 機能、マルチキャストパケットのラベルスイッチ転送が可能となった。以下が、CSR で使用されている FANP の特徴である。

- ソフトステートによる LSP 管理

LSP は、隣接 CSR 間で定期的に転送されるレフレッシュメッセージにより維持されます。一定期間、レフレッシュメッセージが受信されない場合には、LSP が開放される

- 上流側ノード起動による LSP 生成

LSP の生成は、経路制御的に上流側に存在する CSR により起動されます。ユニキャスト通信では、下流 CSR からの LSP の生成も可能ですが、ATM のマルチキャストコネクションを用いたマルチキャスト通信では、下流 CSR からの LSP の生成が不可能となる。したがって、CSR システムでは、上流側の CSR が、LSP の生成を起動する。

- SDH のようなトランスペアレントなリンクでの動作

ATM 以外のリンク (例えば、SONET/SDH リンク) では、いわゆる Cell-in-the-wire モデルを適用し、ATM シグナリングなしに、ラベルスイッチングを実現することができる。

- ATM のようなラベルスワップリンクでの動作

CSR システムでは、ATM ネットワーク上でラベルスイッチングを実現するために、VCID 通知手続き (VCID Notification Procedure) を実行する。VCID 通知手続きは、WIDE プロジェクトの活動の一貫として、IETF MPLS ワーキンググループに提案した。その結果、IETF MPLS の標準の Protokol として、採用される予定である。

- トラヒックドリブンとトポロジードリブンを同時サポート

CSR システムでは、トポロジードリブンとトラヒックドリブンを同時にサポートすることができる。トポロジードリブンは、CSR システムが ISP バックボーンのような大規模システムのバックボーンにおいて適用されるために使用される。トポロジードリブンの適用により、CSR システムで必要となるラベル数 (VPI/VCI 数) を削減することができる。一方、トラヒックドリブンにより、特定のアプリケーション (あるいはユーザ) に対して、個別の LSP を割り当てることができる。トラヒックドリブンの LSP のパケット転送優先度を、トポロジードリブンの LSP のパケット転送優先度よりも高くすることで、特定のアプリケーション (あるいはユーザ) に対して、より高品質のパケット転送サービスを提供することが可能となる。

Type of VC	no VCID	inband	outband (small)	outband
Point-to-point	Yes	—	—	—
VP	Yes	Yes	—	—
PVC	—	Yes	—	—
SVC	—	Yes	Yes	Yes

図 4.1: ATM リンクタイプと VCID 通知手続き

- Raw IP および UDP 上での FANP の動作

隣接 CSR ノード間での FANP 制御メッセージの交換には、Raw IP および UDP が使用されている。

4.2 VCID 通知手続き

CSR システムでは、標準の ATM インターフェースとの相互接続を持つ、唯一のラベルスイッチルータである。すなわち、CSR システムでは、ATM 網の SVC(Switched VC) および PVC(Permanent VC) を CSR 間を相互接続するデータリンクとして使用することができる。したがって、CSR システムでは、CSR 間を相互接続するデータリンクとして、ポイントポイントの ATM リンク (Point-to-Point ATM-VC) を利用できるだけでなく、サブネット技術として、Classical IP (RFC1577) や、NHRP/MPOA を適用して、隣接 CSR 間を相互接続することを可能とする。

VCID 通知手続きには、4 種類の手続きが定義されています。図 4.1に、ATM リンクの種類ごとに適用可能な VCID 通知手続きを示した。

”No VCID”においては、VCIDを使用する必要はなく、VPI/VCI 値を直接 FEC(Forwarding Equivalent Class) にマッピングすることができる。

1. インバンド方式

ATM の ARP メッセージを拡張することで、VCID 値の通知を、LSP として使用する ATM-VC を用いて行う。下図に、インバンド方式を示した。VCID 通知メッセージは、FEC 情報と VCID 値を持つ。下流ノード B は、VCID 通知メッセージの受信により、(VCID, FEC, LSP) として使用する ATM-VC とのマッピングを確立する ((FEC, VCID, Port, VPI/VCI))。下流ノードは、デフォルト VC を用いて、上流ノードに VCID 通知 ACK メッセージを転送する。VCID 通知 ACK メッセージには、FEC と VCID 情報が書かれており、下流ノードでのラベルマッピングが確立したことを確認することができます。上流ノードおよび下流ノードともに、(FEC, VCID, Port, VPI/VCI) のマッピングを確立する。なお、Port と VPI/VCI は、各ノードでの

ローカルな値で、FEC と VCID が隣接 CSR で共通に使用される識別子となる。上流ノード A で、VCID 通知 ACK メッセージの受信により、LSP を用いたパケット転送を開始することができるようになる。

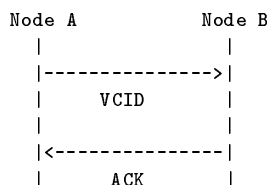


図 4.2: インバンド方式

2. アウトバンド方式

ATM シグナリング (Q.2931/Q.2941) において、エンド ノード (=隣接 CSR ノード) 間で交換可能な情報フィールド (BLLI および GIT フィールド) を用いて、VCID 値を交換する方法である。

● BLLI 方式

Q.2931 ATM-VC セットアップメッセージ中の BLLI (Broadband Low Layer Information) フィールド (7 ビット) を、テンポラリーな識別子として隣接 CSR ノード間で交換する。その後、デフォルト VC を用いて、BLLI 値 (7 ビット) に対応する VCID 値の通知および確認 (ACK) を行う。下図に示した手続きにより、 $\{FEC, VCID, Port, VPI/VCI\}$ のマッピングを確立する。

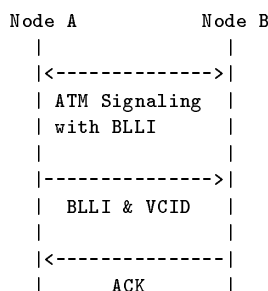


図 4.3: アウトバンド (BLLI) 方式

● GIT 方式

Q.2941 ATM-VC セットアップメッセージ中の GIT フィールドを用いて、VCID 値の通知を行う。VCID 通知および VCID 通知確認 (ACK) メッセージの交換により、 $\{FEC, VCID, Port, VPI/VCI\}$ のマッピングを確立する。

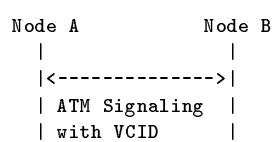


図 4.4: アウトバンド (GIT) 方式

第 5 章

ラベルスイッチルータにおけるラベルマッピング方式

5.1 概要

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、ルータ高性能化を実現させる技術である。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 アドレス情報によってパケット転送だけでなく、layer-3 アドレス情報を対応づけたラベル情報によりパケット転送を行う。LSR では、1 つのパケットストリームごとに 1 つのラベルが必要であるため、必要なラベル数の評価は重要である。本セクションでは、従来提案されていたラベル割り当て方式を用いて、LAN 環境における必要ラベル数と WAN 環境における必要ラベル数を評価した。LAN 環境においては、必要ラベル数は最大 50 であり、トラヒックドリブンでの実現が可能であることがわかった。WAN 環境では、従来提案されていた方式では、多くのラベル数が必要であることがわかった。そこで、LAN および WAN 環境において実用なラベル数を少なくするための方法をそれぞれ提案し、実データトラヒックを用いて、評価を行った。LAN 環境における必要ラベル数の削減の方法として、宛先ネットワークではなく、OSPF ドメインのボーダルータ間にラベルスイッチパスを生成する方法 (本セクションではこれを OSPF を用いたアグリゲーションと呼んでいる) を提案している。一方、WAN 環境における必要ラベル数の削減方法として、宛先ネットワークの粒度のパケットストリームを持つトラヒックドリブンを用いたラベル割り当て方式 (フロ - アグリゲートトラヒックドリブン方式) を提案している。

実トラヒックデータを用いた評価の結果、それぞれの提案方式の有効性を確認することができた。従来のラベル割り当て方式に比べて、必要なラベル数を削減し、さらにラベルスイッチ転送率の向上を実現できることがわかった。

- OSPF アグリゲーション

必要ラベル数を約 56%削減し、カットスルー転送率をほぼ 100 ができた。

- フロ - アグリゲートトラヒックドリブン

必要ラベル数を 1/10 ~ 1/100 程度、カットスルー転送率を 30%程度) を向上させることができた。

第 6 章

OSPF を用いたフローアグリゲーション

このセクションでは OSPF の情報を用いたボーダルータ間でのカットスループス (LSP) の設定方法を述べる。

6.1 OSPF による経路決定

まず、OSPF での経路情報決定方法を簡単に説明する。

OSPF による経路プロトコルを動作させているルータは、そのルータと同じ OSPF エリア内のネットワークトポロジーとそれぞれのリンクのコストを保持している。この情報は隣接ルータ同士が OSPF で情報交換することにより生成される。ルータがパケットを転送するための次段ノードは、この情報により決定される。ルータはこの情報を用いて最小木を作りその経路に従ってパケットを転送する次段ノードを決定する。この動作をそれぞれのルータが行なうことにより、パケットが目的地まで転送される経路が決定する。OSPF エリア外のネットワークに関する経路計算は、エリア内のルータからエリア外のネットワークへの到達性とコストの情報を使う。コストが一番小さいエリア内のルータから到達できるものとして、そのエリア内のルータに到達できる次段ノードにパケット転送を行なうように設定できる。

例を図 6.1 に示す。この図は、R1 が保持しているトポロジー情報とリンクのコストを示している。R1 からそれぞれのノードへの経路は、太線で書いた矢印のような最小木となる。R1 は、R2 宛のパケットを送信する場合は、R2 へ転送すれば良いことがわかり、R3, R4, R5 へパケットを送信する場合は、R1 へ転送することがわかる。次に、エリア外のネットワークである Network A と Network B への経路計算を示す。Network A は、エリア内のルータ R4 から到達性があること、Network B は、エリア内のルータ R5 から到達性があることを R1 が OSPF 情報として保持している。そこで、Network A は、R4 宛の経路と同じ経路を通り、パケットは次段ルータである R3 へ転送すれば良いことがわかる。また、Network B は、R5 宛の経路と同じ経路をとり、次段ルータとして R3 を選ぶ。

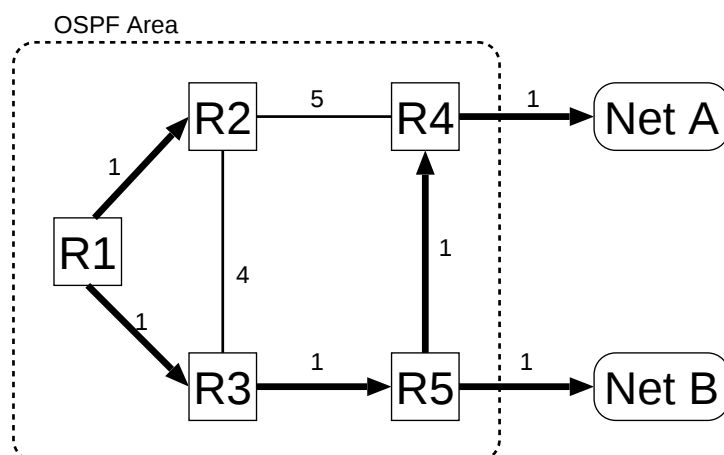


図 6.1: OSPF Topology

6.2 OSPF 情報を用いたアグリゲート LSP

この節では、OSPF 情報を利用してラベル数を減少させる方法を示す。これは、OSPF 情報を用いて、入口ルータと出口ルータの間に 1 つの LSP を設定することにより実現する。このためには、以下の 2 点を実現する必要がある。

1. 入口ルータが LSP を設定すべき出口ルータを認識できること
2. 入口ルータが出口ルータまでの LSP を設定できること
3. 入口ルータが出口ルータまでの LSP を通して転送するパケットを特定できること

(1) を実現する方法として、LSR ドメイン¹の出口ルータを自動判別する方法と出口ルータを手設定で設定する方法がある。本セクションでは、前者のアプローチによる方法を提案・評価している。今回は、CSR 上で動作している FANP を利用して実装したが、IETF で標準化中の LDP を使用しても同様のことができる。本方式を実現するためには、LSP の出口ルータより先に接続しているネットワークを入口ルータが知る必要がある。OSPF を扱うすべてのルータは、OSPF の同一エリア内のトポロジーと経路情報を知っている。そのため、入口ルータと出口ルータが OSPF の同一エリアに属している場合には、これを比較的容易に実現することができる。

具体例として、図 6.2 を用いて、OSPF 情報を用いて入口ルータ R1 からの LSP の設定方法を説明する。R1 には予め R4, R5 がエッジルータであることを設定しておく。R1 はそれぞれのエッジルータに対して LSP を設定する。エッジルータである R4 への経路は、R1→R3→R5→R4 とわかるので、この経路で LSP を設定する。図 6.3 に示しているように

¹LSR のみで接続されている領域

R4 を経由して到達する宛先は、R4 と Network A であることがわかる。そこで、R4 宛の LSP では R4 と Network A 宛のパケットを転送すれば良いことがわかる。もう 1 つのエッジルータ R5 へは、R1→R3→R5 の経路で LSP を設定する。この LSP では、R5 と Network B 宛のパケットを転送する。上記のように OSPF の情報を利用して、エッジルータまでの LSP を設定し、その LSP に流すパケットを設定することで、入口-出口ルータ間の LSP を設定することができる。

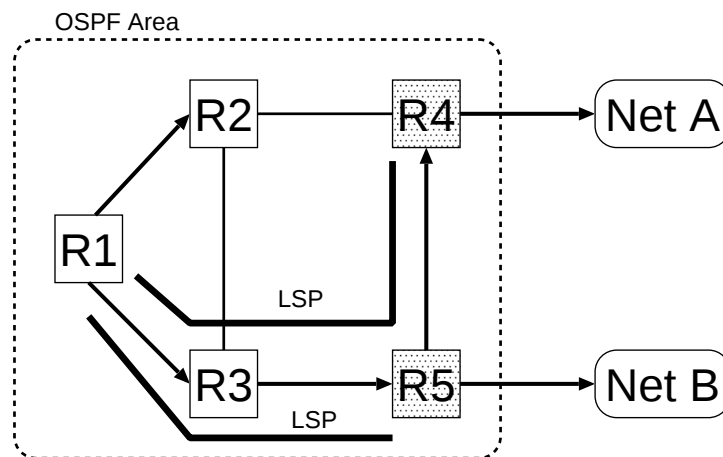


図 6.2: LSP 設定例

6.3 評価環境

1998 年 3 月に行なわれた 4 日間の WIDE Project 研究会において、LSR の 1 実装例である CSR のバックボーンネットワークを構築した。研究会に参加した人が、このネットワークを利用して、メール、WWW 等の実トラヒックを流していた。今回は、このネットワーク内の CSR を観測することにより、CSR に実トラヒックを流したときの動作を測定した。

今回の実験に用いた CSR のネットワークポロジ図を図 6.3 で示す。CSR コアノード 1 台に CSR エッジノード 4 台を 155M ATM でスター状に接続した。1 台のエッジ (エッジ 1) の先でインターネットに接続し、他の 3 台のエッジ (エッジ 2,3,4) から 10M/100M イーサネットにより、ユーザホストが接続している。

CSR の動作パラメータとして、以下ものを使った。

ラベル割り当てトリガ: トラヒックドリブン

http/ftp/telnet/nntp のパケットを送信するときにラベルを割り当てる。また、60 秒トラヒックが流れないときにラベルを解放する。

パケットストリーム: 同一ホストペア

このネットワーク構成で、4 日間の実運用トラフィックをかけ、CSR に関する統計情報を測定した。測定項目は、CSR CORE ノードの LSP 数、ラベル数、ユーザサブネットのホスト数である。ホスト数は、CSR エッジノードで動いている DHCP サーバからの貸し出したアドレスの数であり、実際にパケット転送を行うアクティブなホスト数と見ることができる。これらの項目を 5 分間隔で収集した。統計情報収集プログラムとしては、MRTG を使用した。

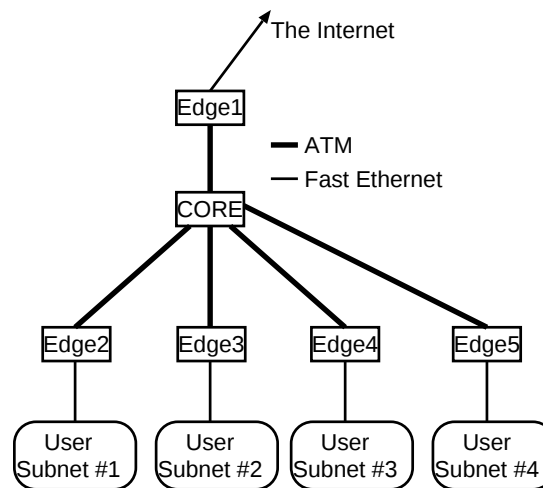


図 6.3: 実験トポロジー

評価結果

図 6.3 の実験トポロジーで測定した統計情報を議論する。図 6.4, 6.5, 6.6 は、約 1 日間のホスト数の推移を示している。横軸が時刻であり、縦軸がホスト数である。

この図から、4 時から 9 時までは、利用ホスト数が少なく、10 時から翌日 1 時までの利用ホストが多いことがわかる。ホスト数は 22 時頃が最大で、合計 50 台のホストが存在していた。

図 6.7 は、CSR コアノードで測定した LSP 数である。図 6.8, 6.9, 6.10, 6.11 は、CSR コアノードからそれぞれの CSR エッジノードへのラベル数 (VC 数) である。横軸が時刻で、縦軸がそれぞれ LSP 数とラベル数を示している。これらは、約 1 日分の測定結果であり、上記のホスト数と同じ 1 日の結果である²。

² 全てのユニキャスト用 LSP が CSR コアノードを通過している場合は、1 つの LSP に対して入力側 VC と出力側 VC が対応するので、LSP の数の 2 倍のラベルが存在することになる。今回の結果は、ラベル数が LSP 数の 2 倍以下であるが、これは、CSR コアノードで LSP が終端しているものがあり、1 つの LSP に対して 1 つの VC しか無い場合があるからである。

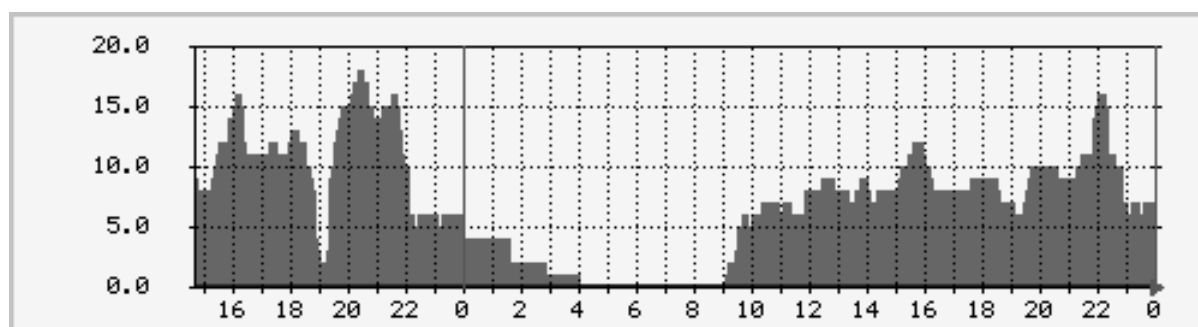


図 6.4: サブネット 1 のホスト数

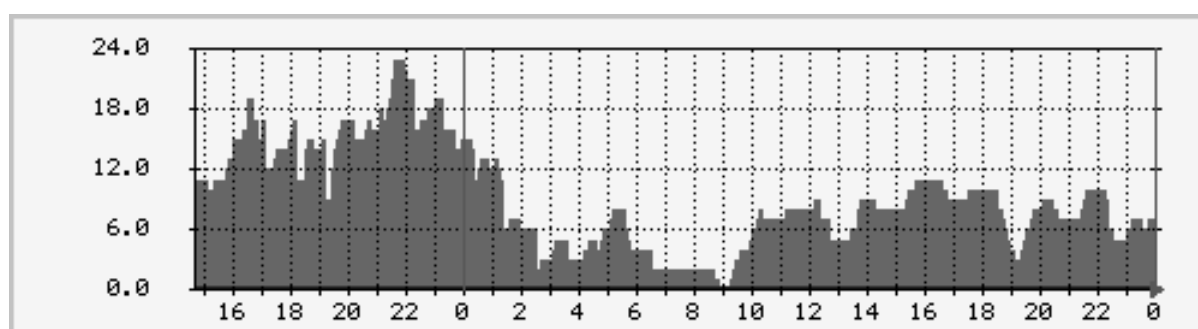


図 6.5: サブネット 2 のホスト数

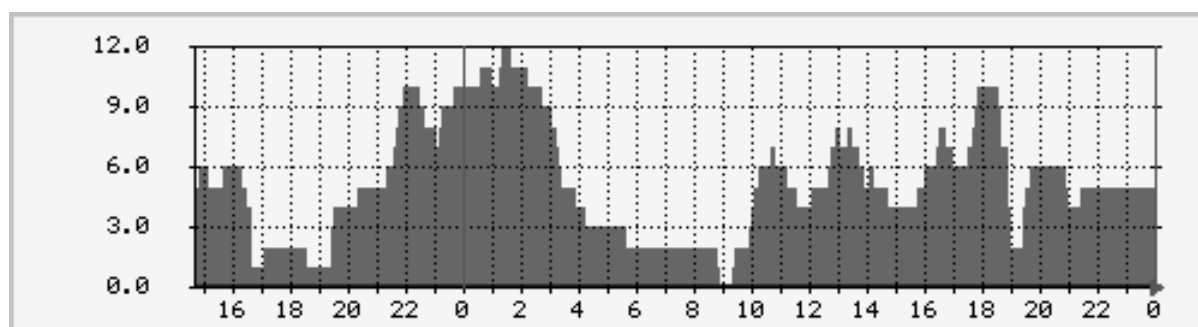


図 6.6: サブネット 3 のホスト数

LSP 数は、19.5 時及び 22 時に最大であり、36 であった。ホスト数がもっとも多い 22 時に LSP 数も最大になっているが、ホスト数がそれほど多くない 19.5 時にも LSP 数も最大になっている。ラベル数の合計の最大値も、LSP 数が最大の時であり、50 であった。この規模の LAN では、トラヒックドリブン/ホストペアのラベル割り当てポリシーでラベルを割り当てたとしても、実装可能な値に収まることがわかった。

6.4 ラベル数の評価

前章のネットワークにおいて、今回実装した OSPF の情報を用いた入口-出口ルータのパケットストリームの粒度の LSP を設定した場合のラベル数を測定した。測定結果を図 6.12 に示す。理想的には、入口-出口ルータの粒度で設定した LSP 数は、 $N \times (N-1)$ (N はエッジルータ数) になる。今回の結果では、いくつかの時間でこの数と異なる場合もあるが、これは、経路が不安定であったときに、このようになったと考えられる。前節のトラヒックドリブン/ホストペアのラベル割り当てポリシーの最大ラベル数が 45 本であるのに比べて、OSPF を用いた入口ルータ - 出口ルータの LSP でのラベル数は、20 本であった。この図では、OSPF による LSP アグリゲーションにより、56%のラベルを削減することができた。

6.5 まとめ

LAN 環境において LSR を使用し、従来のラベル割り当て方式であるトラヒックドリブン/ホストペアで LSP を設定した場合と、ラベル割り当て方式をトポロジードリブン/入口-出口ルータのアグリゲート LSP を設定した場合のラベル数を比較した。

この環境では、フロードリブン/ホストペアの LSP では、最大 45 本のラベルが必要であった。OSPF を用いた入口ルータ - 出口ルータの LSP でのラベル数は、20 本であった。OSPF による LSP アグリゲーションにより、56%のラベルを削減することができた。

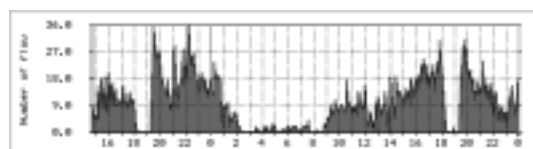


図 6.7: コアノードの LSP 数

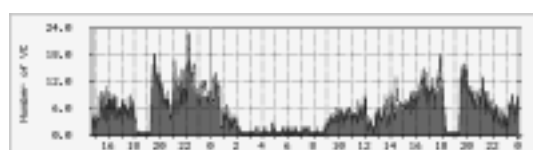


図 6.8: コアノードのラベル数 (Edge 1 宛)

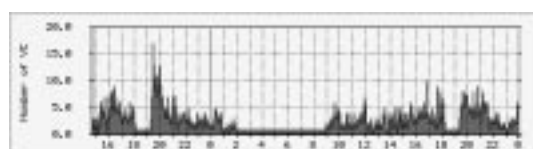


図 6.9: コアノードのラベル数 (Edge 2 宛)

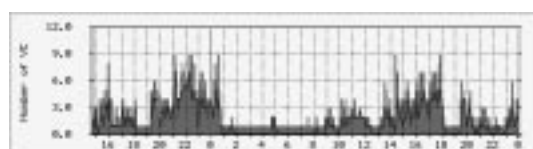


図 6.10: コアノードのラベル数 (Edge 3 宛)

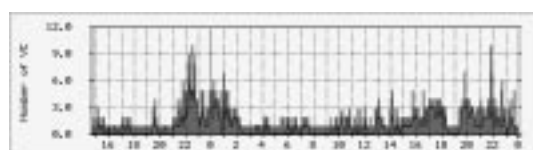


図 6.11: コアノードのラベル数 (Edge 4 宛)

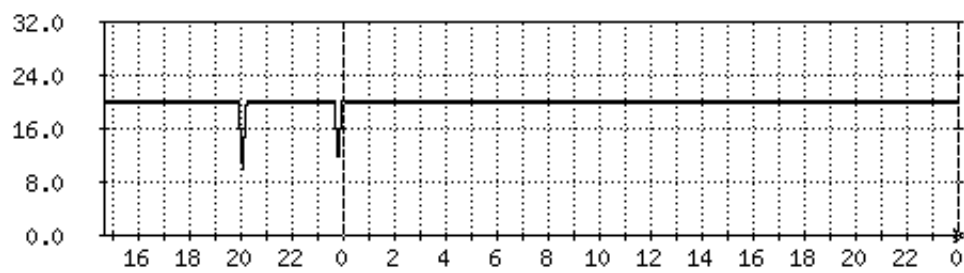


図 6.12: OSPF 情報を用いた LSP のラベル数

第 7 章

Aggregated Flow-Driven 方式を用いたラベルマッピング

7.1 LSR におけるラベル割り当て方法の概要

SR では、通常のルータと同じ様に Layer 3 のパケットのヘッダを見て転送する Layer 3 転送モジュールと、Layer 3 のパケットのヘッダを見ることなく、固定長ラベル (例えば ATM の場合は、VPI/VCI) によりパケットを転送する Layer 2 転送 (図 7.1:L2 Forwarder) モジュールを持つ。Layer 3 転送では、パケットヘッダの宛先アドレスをキーとするベストマッチ検索が行われるが、Layer 2 転送では、固定長ラベルでの検索を行えるため、低遅延での転送が行える。

まず、通常のルータと同じ Layer 3 転送を図 7.1を用いて説明する。Edge1 から送信したパケットは、デフォルト VC を通して、LSR で受信される。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て、Edge2 ヘデフォルト VC を通してパケットを送信する。

次に、LSR の特徴である layer 2 転送の場合を説明する。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 7.1の場合は、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は、Layer 2 転送モジュールのみで転送できるように、そのラベルを 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 転送処理を行うことなく、パケットを転送することが出来る。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。

LSR の 1 つの実装形態として ATM-LSR がある。これは、Layer 2 転送モジュールとし

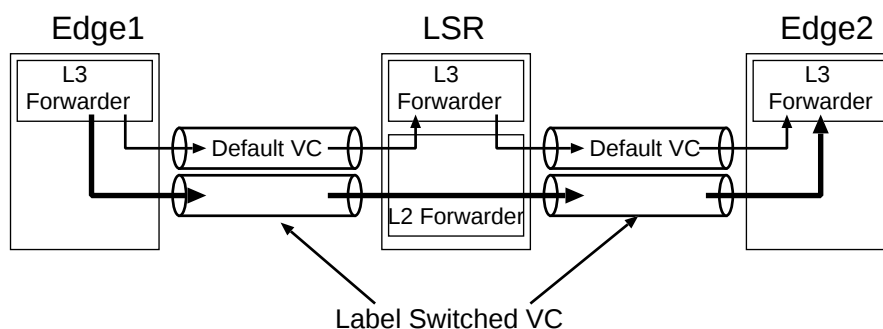


図 7.1: LSR の動作概要

て ATM スイッチを使うものである。ATM-LSR では、それぞれのラベルの値毎にリアセンブルバッファを割り当てる必要がある。したがって、ATM-LSR では、リアセンブルバッファ容量の大きさにより、割り当てられる最大ラベル数が決定される。ATM-LSR の拡張性を評価するためには、必要なラベル数を評価することが重要である。必要なラベル数は、以下の 3 つのパラメータに依存する。

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. 運用システムの規模

前述の通り、LSP を制御するトリガとして、大きくトラヒックドリブンとトポロジードリブンの 2 つがある。

トラヒックドリブン

トラヒックドリブンとは、ある特定の packets を隣接 LSR に送信する時をラベル割り当ての契機とするものである。利点としては、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 packet 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ため、その間、LSR の特徴である layer 2 転送は行うことができない。

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とは、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、トラヒ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。

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) が同じパケットを流すことが考えられる。これは、ホストからホストへのフローである。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。

ストリームの種類として以下のようなものが考えられる。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。

- (送信元アドレス, 宛先アドレス) が同じストリーム。以下では、ホストペアあるいは (src,dst) と表現する。
- (送信元アドレス, 宛先ネットワーク) が同じストリーム。以下では、(src,dstnet) と表現する。
- (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。以下では、(srcnet,dstnet) と表現する。
- (宛先アドレス) が同じストリーム。以下では、(*,dst) と表現する。
- (宛先ネットワーク) が同じストリーム。以下では、(*,dstnet) と表現する。
- (始点 LSR, 終点 LSR) が同じストリーム。LSP の始点 LSR と終点 LSR が同じストリーム。

ラベル割り当てポリシーは、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、様々なものが考えられる。現在提案されているものとして、2 つのラベル割り当てポリシーがある。1 つは、IFMP (Ipsilon Flow Management Protocol) [RFC1953] and FANP (Flow Attribute Notification Protocol) [RFC2129] で定義されているものでパケットストリームの粒度としてホストペアを選びラベル割り当てトリガとしてトラヒックドリブンを使うものがある。以下では、トラヒックドリブン/ホストペアのラベル割り当てポリシーと表現することとする。もう 1 つのラベル割り当てポリシーは、TDP (Tag Distribution Protocol) [TDP] 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トポロジードリブンを使っている。トポロジードリブン/宛先ネットワークのラベル割り当てポリシーと呼ぶこととする。

7.2 提案方式; Flow Aggregated Traffic Driven

本セクションでは、共通の宛先ネットワークを持つパケットフローを定義し、このパケットフローに対するラベルスイッチパス (カットスルーパス) の割り当てを、トラヒックドリブンポリシーを適用して行うという方式を提案し、評価を行っている。すなわち、特定の宛先ネットワークに転送されるパケットが実際に到着したときにのみ、カットスルーパスの生成を行う。

7.3 評価環境

ラベル割り当てポリシーがトポロジードリブン/宛先ネットワークの場合とトラヒックドリブン/ホストペアの場合の LSR のラベル数とカットスルー率を実際のトラヒックトレースを用いて評価した。トラヒックドリブンのパラメータとして、LAN と評価の場合と同様に http/ftp/telnet/nntp のパケットを送信する場合にラベル割り当て、60 秒トラヒックが流れない場合にラベルを解放することとした。また、ラベル割り当ては、100ms かかるものとして評価した。

以下では、シミュレーションに使ったトラヒック情報の取得環境及びパケット転送レートを述べる。

トラヒック取得点 図 7.2 は、トラヒック取得点を示している。トラヒック取得ホストは、WIDE (Widely Integrated and Distributed Environment) プロジェクトの国内と海外を結ぶイーサネットに接続している。トラヒックトレースは、tcpdump プログラムにより取得した。トラヒックは、約 2 時間 (7437 秒: 1997 年 11 月 4 日 14:08 から同日 16:11 まで) 取得した。トラヒックトレースは、国内向、海外向けの両方向別々に取得した。シミュレーションで使用した経路表は、全世界の経路表 (フルルート経路表) を用いた。

パケット転送速度 表 7.1 は、トラヒック取得点での平均パケット転送速度を Mbps (Mega Bits Per Second) と pps (Packet Per Second) で示している。WIDE プロジェクトバックボーンネットワーク (以下 AS (Autonomous System) と呼ぶ) に入るトラヒックは、1.225Mbps で 329pps であり、AS から出るトラヒックは、0.926Mbps で 564pps であった。

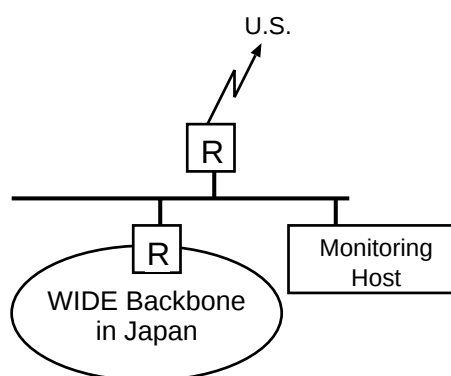


図 7.2: トラヒック取得点

方向	Transfer Rate [Mbps]	Transfer Rate [pps]
AS 内へ	1.225	329
AS 外へ	0.926	564

表 7.1: 平均転送速度

7.4 評価結果

表にトラヒックドリブンでパケットストリームがホストペアの場合とトポロジードリブンの場合のラベル数とカットスルー率を示す。また、トラヒックドリブンに関しては、Disconnect Timer に対するカットスルー率および必要となるラベル数の評価も行っている。

トラヒックドリブンに関する評価条件は以下の通りである。

- カットスループスの生成は、HTTP/FTP/TELNET/NNTP のセッションに属する IP パケットが、LSR から転送された時とした。
- Disconnect Timer Interval の期間、IP パケットが転送されなかった場合に、カットスループスが解放される。評価では、10,30,60,120,240,300 および 600 秒の Disconnect Timer Interval を用いて評価した。
- カットスループス (LDP) の生成遅延は、100 ミリ秒とした。

図 7.3は、WIDE プロジェクトバックボーンから他の AS に向かって流れ出るパケット流に対して、必要となるラベルの数とカットスルー率を評価したものである。一方、図 7.4 は、他の AS から、WIDE プロジェクトバックボーンに向かって流入するパケット流に対して、必要となるラベルの数とカットスルー率を評価したものである。

どちらの結果も、Disconnect Timer が 60 ミリ秒以上の領域では、パケットのカットスルー率の向上があまり期待できない(飽和状態になる)ことが分かる。つまり、この評価結果から、Disconnect Timer の値は、60 ミリ秒程度が最適であることを意味している。そこで、以下の評価では、Disconnect Timer の値を、60 ミリ秒として評価を行っている。

トラヒックドリブンを用いた場合は、AS 内へのトラヒックに関するラベル数は 365 で、カットスルー率は 61%である。AS 外へのトラヒックに関するラベル数は 1131 でカットスルー率は 68%であった。

トポロジードリブンを用いた場合は、経路表のエントリの数だけのラベル数が必要である。全世界の経路表(フルルート)を使った場合には、AS 内方向は、ラベル数は 2518 であり、AS 外方向は、ラベル数は 48385 である。トポロジードリブンで LSP を設定した場合は、全てのトラヒックがその LSP を通るのでカットスルー率は、100%となる。

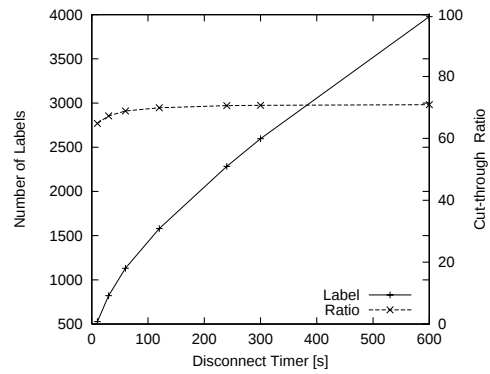


図 7.3: Number of labels and cut-through ratio directed outside of AS

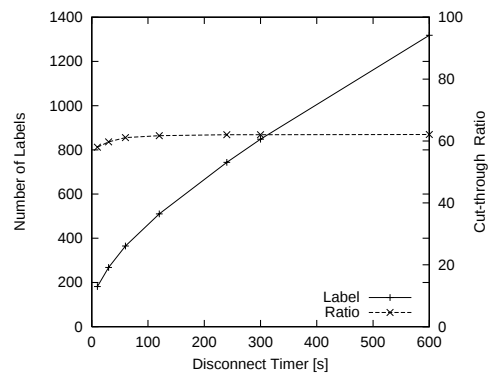


図 7.4: Number of labels and cut-through ratio directed inside of AS

この結果から、WAN で LSR を利用した場合で典型的なラベルマッピングポリシーを使った場合には、多くのラベル数が必要であることが分かった。

	ラベル数	カットスルー率 [%]
トラヒックドリブン (AS 内へ)	365	61
(AS 外へ)	1131	68
トポロジードリブン (AS 内へ)	2518	100
(AS 外へ)	48385	100

表 7.2: 典型的なラベルマッピングポリシーを用いたラベル数とカットスルー率

7.5 フローアグリゲーションの評価

前セクションまでの評価において、典型的なラベルマッピングポリシーを用いて LAN 及び WAN における必要ラベル数とカットスルー率を述べてきた。LAN においてトラヒックドリブンを用いると、多くのラベル数が必要無いことがわかった。WAN においては、典型的なトラヒックドリブンとトポロジードリブンのラベルマッピングポリシーでは、多くのラベル数が必要なが分かった。

トポロジードリブンでラベル数が多く必要な理由は、全ての経路エントリに対して、LSP を設定しているためである。表に示すように、実際の経路エントリ数に比べて、参照されるエントリ数は、AS 内部宛で 4.2% であり、AS 外部宛で 15.6% と非常に少ない。そこで、トポロジードリブンと同じ宛先ネットワークの様なアグリゲートフローをパケットストリームに選び、そのネットワーク宛のパケットを送信するときのみ LSP を設定するトラヒックドリブンのラベルマッピングポリシーを提案する。

この方式では、パケットストリームの粒度を大きくし、必要なときに LSP を設定するトラヒックドリブンを選んでいるため、ラベル数の減少が予想される。

まず、外部 AS とのパケットの交換において、少数 AS に対するの固定経路情報を使用する場合 (i.e., ほとんどは Default Routing となる) について、実際に参照されるルーティングエントリーの評価を行った。評価では、与えられた時間内に、実際にパケットの到着・転送時に、参照されたルーティングエントリーの解析を行った。評価結果は、図 7.5、7.6 および 表に示した。

図 7.5 および図 7.6 は、実際に参照されたルーティングエントリーの頻度を示している。横軸は、ルーティングエントリーに存在するネットワークアドレスを IP アドレスの順番

に並べたものである。例えば、“192.69.251.56”宛てのパケットを LSR が受信した場合には、ベストマッチポリシーに基づいたルーティングテーブルの検索を行い、対応するルーティングエントリーを探す（例えば、“192.69.251.0/24”など）。したがって、このパケット（192.69.251.56 宛て）が受信されたというイベントは、図の横軸の “192.69.251.0” 上のデータにカウントされる。また、横軸の “0.0.0.0” は、Default Routing でのパケットの転送に対応させている。

図 7.5 で分かるように、外の AS に転送されるパケットのうち 約 96 % のパケットが Default Routing で転送されている（図中の “0.0.0.0”）。

表は、評価期間終了時に存在した ルーティングエントリーの数と、評価期間中に実際に、参照されたルーティングエントリーの数を示している。評価に使った LSR ノードは、他の AS 行きのルーティングエントリーが 5 つ存在している。1 つが Default Routing なので、残りの 4 つが個別ネットワーク (AS) に対するルーティングエントリーである。5 つのエントリーは、評価期中にすべて (100%) 参照されている。ここで、図 7.5 では、5 点以上のルーティングエントリーが参照されていることになっている。これは、経路の不安定化により、評価期間中に消えていたルーティングエントリー行きのパケットが、他の AS 経由で転送されたために、発生したものであると考えられる。

自 AS に流れ込むパケットに関しては、合計で 209 のルーティングエントリーが実際に参照されている。最も、参照されているルーティングエントリーは全体の 33% 以上のパケットが参照していることが分かる。実際に、必要になるラベルの数は、この数（例えば 209）よりも少ない数になるはずである。いづれにしても、通常の Topology-Driven での運用では、2860 のラベルが必要になってしまうことと比較すると、提案している、宛先ネットワークに関してアグリゲートされたフローに対してトラヒックドリブンのポリシーを適用することで、約 7% のラベル数で対応することが可能となる。

次にフルルートを使用した場合の、評価結果を示す。合計のルーティングエントリーの数は、50,903 である。表に示したように、自 AS 内へのエントリー数が 2518、自分以外の AS に対するエントリー数が 48385 である。やはり、横軸上の “0.0.0.0” は Default Routing を示している。図中の Default Routing で転送されるパケットは、LSR のルーティングエントリーに存在しない AS に対するパケットの転送である。

106 のラベルが自 AS 宛てのパケット転送のために必要となる。一方、他 AS へのラベルスイッチ転送のためには、7530 のラベルが必要となる。なお、従来の Topology-Driven での運用では、50903 のラベルが必要となる。一方、提案しているラベル割り当て方法を適用することで、合計で 7636 (= 7530 + 106) のラベル数で十分となる。つまり、従来の Topology-Driven での運用に比べて、約 15% 程度のラベル数を提供すればよいことになる。

最後に、必要なラベル数とカットスルー率をトラヒックドリブンで様々なパケットストリームの粒度の場合を評価した。トラヒックドリブンのパラメータは、前セクションと同じである。

図 7.8 に、WIDE プロジェクトバックボーンから他の AS に向かって流れ出るパケット流に対して、必要となるラベルの数とカットスルー率を評価したものである。一方、図 7.9

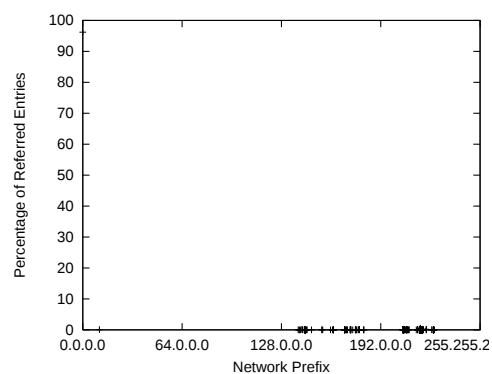


図 7.5: Percentage of referred entries directed outside of the AS

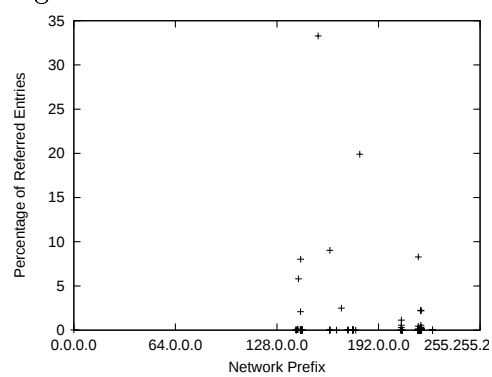


図 7.6: Percentage of referred entries directed inside of the AS

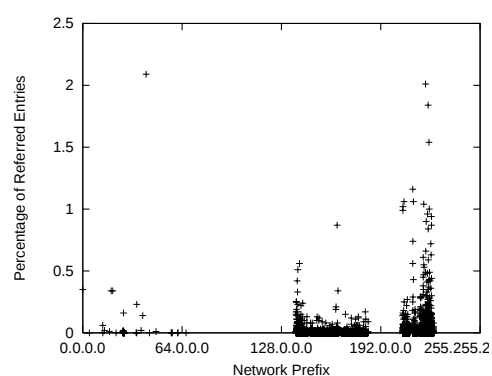


図 7.7: Percentage of referred entries directed outside of the AS

表 7.3: 経路表エントリ数と参照エントリ数

	AS 内部宛	AS 外部宛	合計
経路エントリ数	2518	48385	50903
参照エントリ数	106	7530	7636
参照率	4.2 %	15.6 %	15.0 %

に、他の AS から、WIDE プロジェクトバックボーンに向かって流入するパケット流に対して、必要となるラベルの数とカットスルー率を評価したものである。

どちらの結果も、Disconnect Timer が 60 ミリ秒以上の領域では、パケットのカットスルー率の向上があまり期待できない(飽和状態になる)ことが分かる。つまり、この評価結果から、Disconnect Timer の値は、60 ミリ秒程度が最適であることを意味している。

表 7.4 に示したように、AS の外に転送されるパケットに対して必要となるラベル数は、ホストペアに対してラベルを割り振る場合には、1,131 となる。一方、宛先ネット (*, dstnet) に対するパケットフローに対してラベルを割り振る場合には、99 のラベルで十分である。すなわち、必要なラベル数が、1/10 以下になっている。さらに、パケットのカットスルーによる転送率も、68 % から 99 % に向上している。

表 7.5 は、宛先ネット (*, dstnet) に対するパケットフローに対してラベルを割り振った場合の、AS に流れ込むパケットフローに対して必要になるラベルの数を示した。従来のトラヒックドリブンの場合に比べて、必要となるラベル数は 1/20 以下になっている。さらに、パケットのカットスルーによる転送率も、61 % から 95 % に向上している。

次に、自 AS 以外の領域へのパケット転送に Full-Route を適用した場合の評価を行った。図 7.12 と図 7.13 は、自 AS から他の AS へ転送されるパケットフローに対する、必要なラベルの数とカットスルー率の評価結果を示している。図 7.14 と 7.15 は、他 AS から自 AS へ転送されるパケットフローに対する、必要なラベルの数とカットスルー率の評価結果を示した。本評価でも、Disconnect Timer の値は、60 ミリ秒としている。

送受信ホスト間のパケットフロー (src, dst) の細かな粒度の場合 (Data-Driven に対応)、1,131 のラベルが必要となる。一方、宛先ネットワークについてアグリゲートした場合、つまり、(*, dstnet) の粒度のパケットフローにラベルを割り当てた場合には、542 のラベルが必要となることがわかる。すなわち、他の AS に向かうパケットフローに対して、宛先ネットワークによるアグリゲーションを行うことで (*, dstnet)、送受信ホスト対にラベルを割り当てる場合 (src, dst) に比べて必要なラベル数を約半分に削減することができる。

送受信ホスト間のパケットフロー (src, dst) の細かな粒度の場合 (Data-Driven に対応)、365 のラベルが必要となる。一方、宛先ネットワークについてアグリゲートした場合、つまり、(*, dstnet) の粒度のパケットフローにラベルを割り当てた場合には、16 のラベルが

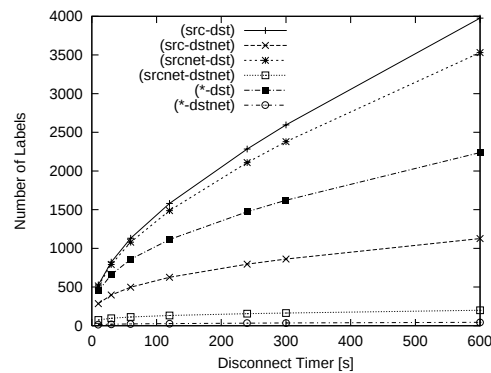


図 7.8: Number of labels directed outside of the AS

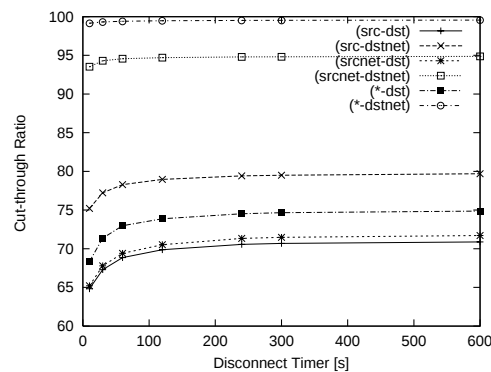


図 7.9: Cut-through ratio directed outside of the AS

必要となることがわかる。すなわち、他の AS に向かうパケットフローに対して、宛先ネットワークによるアグリゲーションを行うことで (*, dstnet)、送受信ホスト対にラベルを割り当てる場合 (src, dst) に比べて必要なラベル数を $1/22$ に削減することができることがわかる。

以上の評価から、宛先ネットワークに対するパケットフローのアグリゲーションを行い、このパケットフローに対して、トラヒックドリブンの運用ポリシーを適用することで、必要なラベルの数を大きく削減するとともに、カットスルー率の向上を実現することができることが明らかとなった。

表 7.4 に AS 外へのトラヒックトレースを使ったシミュレーション結果を示す。パケットストリームにホストペア (src, dst) を選んだ場合は、ラベル数 1131 でカットスルー率が 68% であるのに対して、パケットストリームを宛先ネットワーク (*, dstnet) にするとラベル数 99 でカットスルー率が 99% となる。ラベル数は約 $1/10$ になり、かつ カットスルー率は 31% 向上した。

トポロジードリブンの場合は、ラベル数が、48385 必要であるので、今回の結果である

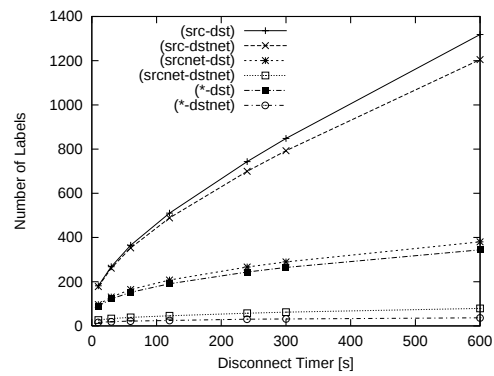


図 7.10: Number of labels directed inside of the AS

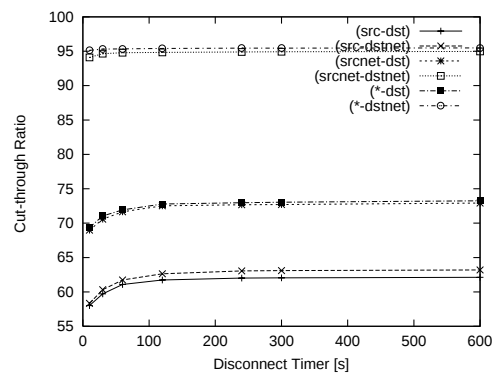


図 7.11: Cut-through ratio directed inside of the AS

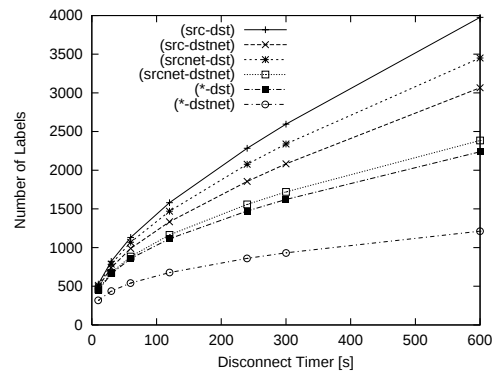


図 7.12: Number of labels directed outside of the AS using full-route

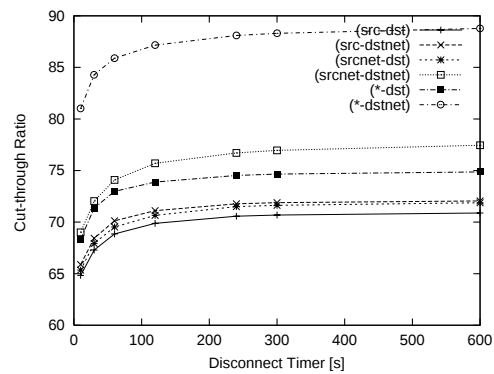


図 7.13: Cut-through ratio directed outside of the AS using full-route

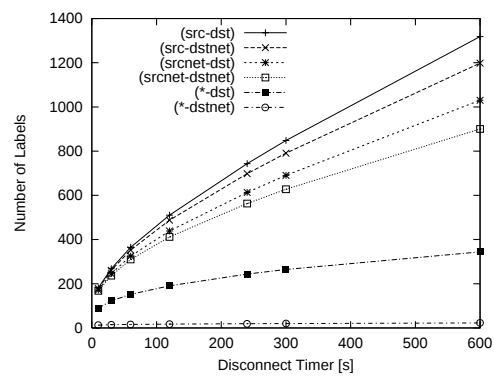


図 7.14: Number of labels directed inside of the AS using full-route

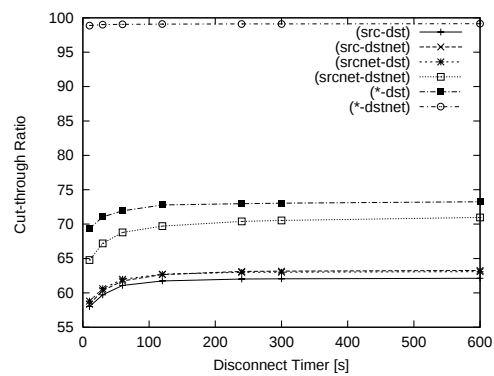


図 7.15: Cut-through ratio directed inside of the AS using full-route

99 と比べるとラベル数は $1/500$ に減少する。

表 7.5 には、AS 内へのトラヒックトレースを使ったシミュレーション結果を示す。パケットストリームにホストペアを選んだ場合は、ラベル数 365 でカットスルー率が 61% であるのに対して、パケットストリームを宛先ネットワークにするとラベル数 21 でカットスルー率が 95% となる。ラベル数は約 $1/20$ になり、かつ カットスルー率は 34% 向上した。

トポロジードリブンの場合は、ラベル数が、2518 必要であるので、今回の結果である 21 と比べるとラベル数は約 $1/100$ に減少する。

これらの評価結果より、トラヒックドリブンでフローアグリゲートを用いたラベルマッピングポリシーが、実運用上かなり有効であることが明らかとなった。

Granularity	Number of Labels	Cut-through Ratio
(src, dst)	1131	68
(*, dst)	859	73
(src, dstnet)	498	78
(srcnet, dstnet)	113	94
(*, dstnet)	99	99

表 7.4: ラベル数とカットスルー率 (AS 外へ)

Granularity	Number of Labels	Cut-through Ratio
(src, dst)	365	61
(src, dstnet)	353	62
(*, dst)	152	72
(srcnet, dstnet)	38	95
(*, dstnet)	21	95

表 7.5: ラベル数とカットスルー率 (AS 内へ)

第 8 章

ラベル割り当てアーキテクチャの考察

以上の考察および実トラヒックを用いた、評価から、以下のような結果が明らかとなった。

- キャンパスバックボーン規模のネットワークでは、純粋なトラヒックドリブンでも十分対応可能な数のラベル数が要求されることが、実ネットワーク環境での実験運用により明らかとなった。
- 同一 OSPF ドメイン内においては、ボーダルータ間に LSP を OSPF の経路情報データベース情報を用いて、自動的に生成・解放することが実現可能であることを実ネットワーク環境での実験運用により確認することができた。また、OSPF を用いたフローのアグリゲーションにより、OSPF ドメイン内で必要となるラベル数を約半数以下に削減可能であった。OSPF が適用されている広域ネットワークにおいても、本手法 (OSPF ボーダルータ間の LSP の提供) が有効であると考えられる。
- 宛先ネットワークに対するパケットフローを定義することで、パケットフローのアグリゲーションを実現できる。さらに、このアグリゲートしたパケットフローに対して、トラヒックドリブンの運用ポリシーを適用することで、ラベルスイッチシステムが提供しなければならないラベルの数を、大幅に削減させることができる。

上記の結果から、大規模システムにおける、ラベル割り当ての運用方式としては、以下のようなものが適切であると考えられる。

1. 小規模キャンパスネットワーク

トラヒックドリブンで十分に運用可能である。ボーダルータにおいては、宛先ネットワークごとにパケットフローのアグリゲーションを行う。アグリゲートしたパケットフローの転送は、バックボーンのルーティングドメインで形成されるボーダルータ間の LSP を使ってトンネリングさせる。すなわち、ラベルの階層的な割り当て機能を使い、ルーティングドメイン内 (e.g., OSPF ドメイン内) では、ボーダルータ間の LSP を一種のレイヤ 2 トンネルのように使用する。

2. 広域バックボーンルータ

単一のルーティングドメイン内においては、ボーダルータ間に LSP の生成を自動的

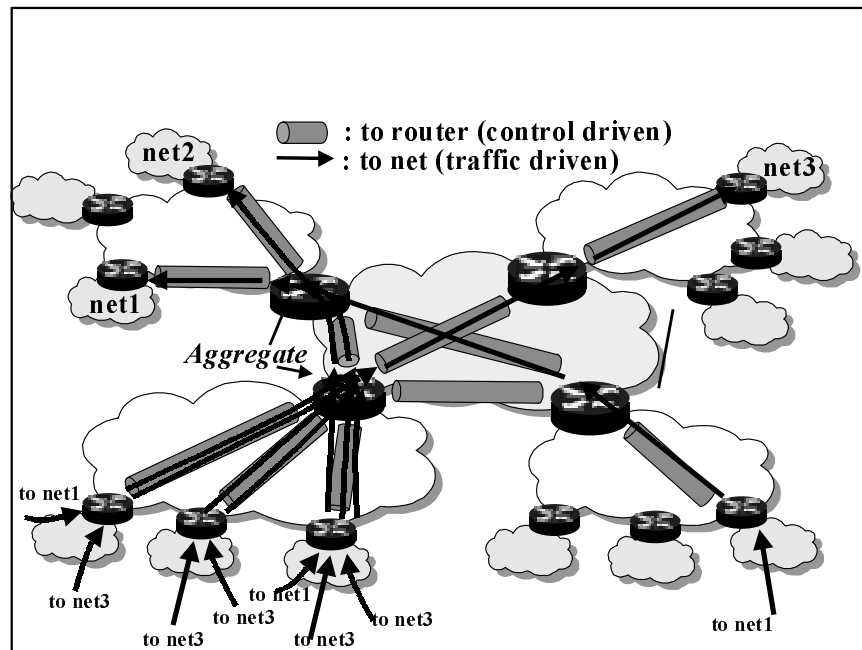


図 8.1: 大規模広域システムにおける LDP の運用方法

に行う。このボーダルータ間の LSP はボーダルータにとっては、一種のレイヤ 2 トンネルとして利用することで、ルーティングドメインにまたがるパケット転送の効率的な転送をラベルスイッチ技術を用いて行う。ラベルの階層的な割り当て機能を用いることにより、1 つの LSP の中に、複数の LSP をトンネリング (多重化) させることができる。

3. AS 間ボーダルータ

ボーダルータ間には、レイヤ 2 トンネル的に見える LSP が提供され、この LSP の中に、宛先ネットワークごとにアグリゲートされたパケットフローに対応する LSP が多重化される。この宛先ネットワーク単位にアグリゲートされたパケットフローに対する LSP の生成は、トラヒックドリブンのポリシーが適用される。したがって、必要となる LDP の数を削減させることが可能となる。

以上のポリシーに基づいた LDP の制御管理運用の形態の概念図を 図 8.1 に示した。いわゆる、トラヒックドリブンとトポロジードリブンのハイブリッド形態での運用である。トポロジードリブンを AS 内のルータ間に適用し、トラヒックドリブンをアグリゲートされたパケットフローに対して適用する。この 2 つの異なる、LSP は、ラベルの階層化割り当て機能により有機的に利用されることになる。

トポロジードリブンに必要な LSP の数は、最大でも $N(N-1)$ である。N は、ルーティングドメインの中に存在するルータの数であり、最大でも数百というのが実運用上での上限である。のルータ数が数百のルーティングドメインの場合、各ルータに必要なラベルの数は、最大でも $2(N-1)$ であるので、十分に実装可能な数に押さえられると考えられる。

以上、OSPF アグリゲーションによるラベル数の削減、および、アグリゲーションパケットフローへのトラヒックドリブンの適用によるラベル数の削減について、実ネットワークトラヒックを用いて定量的な評価を行った。評価の結果、両方式とも、良好に動作することを、確認することができた。大規模な広域起きラベルスイッチシステムを考えた場合、この2つの方式は、競合するものではなく、むしろ、協調して動作するものであることが明らかとなった。また、実際に必要となるラベル数は、数百本のオーダーに押さえることができるが、実トラヒックを使った評価によりわかり、十分に実装可能な数であることも明らかにすることができた。

第 9 章

IETF での標準化に関する成果

IETF におけるラベルスイッチ技術に関する WIDE プロジェクトの成果および貢献の具体的な成果物としては、以下にあげるような Internet-Draft が上げらる。

- VCID メカニズムを用いた ATM リンク上でのラベル
- ループ検出メカニズム

それぞれの、技術的な背景と内容については、既に、説明を行ったので、ここでは行わない。本年度は、以下の IETF 会合において、それぞれ、プレゼンテーションを行った。その結果、VCID メカニズムに関する Internet-Draft は、今後標準化のトラックとなることになっている。

- 夏; Chicago 会合
- 冬; Orlando 会合

第 10 章

今後の研究

WIDE プロジェクトでは、今後のラベルスイッチ技術の重要性を考慮し、引き続き実証実験を含めた、総合的かつ実践的な研究開発活動を継続する予定である。ラベルスイッチ技術は、次世代のインターネットバックボーン技術として重要な技術であり、システムアーキテクチャおよび要素技術に関する研究開発のみならず、実ネットワークを用いた運用技術の確立が必要とされている。

したがって、平成 11 年度も引き続き ラベルスイッチ技術に関する 総合的かつ実践的な研究開発ならびに実験運用活動を行う予定である。平成 11 年度以降の研究開発活動の方向性と課題は以下の通りである。

- 研究開発活動

- IPv6 への対応

次世代インターネットプロトコルである IPv6 への対応を行う。次世代インターネットバックボーンルータとして、IPv4 と IPv6 の両方のサポートが必要である。

- Diff-Serve への対応

米国 Internet2 で構築されている Q-Bone で使用される Differentiated Service アーキテクチャへの対応を行う。高速性のみならず、品質制御を行う。

- 高速マルチキャストへの対応

放送型アプリケーションの本格化に対応するために、大容量高速な IP マルチキャストに対応する必要がある。新しいマルチキャストプロトコル (PIM や Simple-Multicast など) への対応を行う。

- ATM 以外のリンクへの対応

Ethernet や FDDI などの ATM リンク以外へのラベルスイッチの拡張ならびに適用を進める。Shim タグを使用する。

- VPN 機能への対応

Private IP アドレスを用いた VPN の提供のための ラベルスイッチの拡張。

- 高機能ルーティング制御

付加分散および通信品質の提供のための機能を、高機能ルーティング制御に関する研究と連携させながら進める。

－ ラベル割り当て方式の詳細検討

ラベル割り当て方式の検討評価を引き続き進め、ラベル割り当て方式の確立を行う。

- 実験運用活動

WIDE バックボーンを用いた実験運用を継続して行う。ITRC 殿や CKP(Cyber Kansai Project) 殿との相互接続運用によるより広域環境での実験運用を進める予定である。

- IETF 標準化活動

MPLS(Multip-Protocol Label Switching) ワーキンググループでの 技術標準化への積極的な貢献活動を継続する。特に、ループ検出方式の確立および VCID 方式を確立し、標準化技術とするよう活動する。

