MPLSを用いた広域分散IXの実現

中川郁夫[†] 江崎浩[††]
菊池豊[†††] 永見健一[†]

著者らは，IX（Internet eXchange）にMPLS（Multi-Protocol Label Switching）を適用した次世代IXアーキテクチャ**MPLS-IX**を提案してきた．IXは複数のISP（Internet Service Provider）間での相互接続を実現し，効率的にトラフィック交換を行うことを目的として運用されている．**MPLS-IX**では，IXに接続するISPがMPLSによる仮想的なパスを確立し，データリンク層に依存しない相互接続環境を提供する．さらに，**MPLS-IX**では，柔軟なIXトポロジを構築することが可能であるという特徴を持つ．本稿では，**MPLS-IX**を用いて階層型IXを実現し，地域IXなどの既存のIXを相互に接続することにより，広域分散IXを実現する手法について提案する．本稿で提案する手法は，既存のリソースを有効に活用し，既存のIXにおける相互接続に影響を与えることなく，広域分散環境での相互接続を可能にする．

An Implementation of Widely Distributing IX Using MPLS Technology

Ikuo Nakagawa,[†] Hiroshi Esaki,[††] Yutaka Kikuchi[†††]
and Kenichi Nagami[†]

Authors have proposed a next generation IX architecture, called **MPLS-IX**, in which we apply MPLS (Multi-Protocol Label Switching) technology to an IX. Many IXes exist in the current Internet, so that numerous ISPs interconnect and exchange traffic efficiently, each other. **MPLS-IX** establishes virtual paths between participating ISPs. **MPLS-IX** provides data-link medium independancy for interconnection, and enables an IX to be flexible in consiting IX topology. In this paper, we propose an implementation of widely distributing IX. Our implementation has hiarachical IX architecture based on **MPLS-IX** architecture. Deployment of our distributing IX is easy, because we connect existing IXes in hiarachical model, and use existing resouces without any effect to current communications.

1. はじめに

本稿では，IX（Internet eXchange）にMPLS（Multi-Protocol Label Switching）[10]技術を適用した**MPLS-IX**アーキテクチャ[6]を用いて，広域分散環境における階層型IXを実現する手法を提案するとともに，同技術を用いて展開している広域分散IXの実証実験の報告を行う．

IX（Internet eXchange）は，自律的に運用されている複数のネットワークどうしを相互接続する仕組みである．IXには多数のIXが接続し，お互いの経路情報の交換を行うことにより相互接続を行っている[1]．これを「ピアリング」と呼ぶ．

現在のインターネットでは数百にも及ぶIXが運用され[12]，ISP間のトラフィック交換を実現するうえで，IXはきわめて重要な役割を果たしている．たとえばPAIX[13]やMAE[14]，LINX[15]などは世界でも最大級のトラフィック交換点として位置付けられている．また，国内ではNSPIXP2[16]やJPIX[17]が著名である．一方，世界的に地域IXの構築を進める活動も進められている．日本でもTRIX[20]，OKIX[21]，TOYAMA-IX[22]，BeX-J[19]など，数々の地域IXが構築され，運用されている．

一方，著者らは，MPLS（Multi-Protocol Label Switching）技術をIXに応用した次世代IXのアーキテクチャ**MPLS-IX**に関する研究を進めている[6]．この技術は，IXに接続するISP間にMPLSを用いて仮想パスを確立し，その上で相互接続を行う．これ

† 株式会社インテック・ネットコア
Intec NetCore, Inc.
†† 東京大学
University of Tokyo
††† 高知工科大学
Kochi University of Technology

により，データリンク層に非依存な IX を実現できる．また，**MPLS-IX** では，仮想的なパスを確立するための制御に IP を基礎とするネットワーク技術を用いており，IX 全体のトポロジを柔軟に組めることも大きな特徴である．

本稿では，**MPLS-IX** アーキテクチャを用いて，広域分散環境で階層型の相互接続環境を実現する手法について提案する．近年，インターネットでは次のような目的のため ISP 間の相互接続を広域分散環境で実現するための広域分散 IX の仕組みが求められている．

- コンテンツ提供者と地域のアクセスプロバイダを高品質な環境で相互接続する．
- 大手 ISP が地域の ISP にトランジットサービスを提供する．
- 地域間において，高速・広帯域の通信を行う．
- 商用 IX の地方展開を行う．

本稿で提案する手法は，地域 IX などの既存の IX をリーフとして相互に接続するような “メタ” な IX を **MPLS-IX** により実装するものであり，広域分散環境に階層型の相互接続環境を実現するものである．本手法を用いることにより，既存のリソースを有効に活用しながら，容易に広域分散型の相互接続環境を実現することが可能になる．

また，著者らは通信・放送機構の委託研究を受けて **MPLS-IX** アーキテクチャの研究を進めるとともに，次世代 IX 研究会を設立し，広域分散 IX の実証実験を実施している．本稿では，実証実験の状況についても報告を行う．

本稿では，まず 2 章において IX の仕組みと技術について述べる．従来用いられている L2-IX と呼ばれる IX 技術として，特に LAN 技術を用いるものと ATM 技術を用いるものについて述べる．

次に，3 章では **MPLS-IX** アーキテクチャの概要について述べる．ここでは，**MPLS-IX** の基本的な仕組みと，広域分散環境へ適用する際に必要になる特徴について述べる．

そして，4 章では **MPLS-IX** アーキテクチャを用いて，広域分散 IX を実現するための手法について提案する．特に IX を相互に接続する場合に，現存の相互接続の仕組みを変えることなく，シームレスに階層性を持った広域分散環境へ拡張するための手法について述べる．

著者らは **MPLS-IX** 技術の確立とその実証を行うために次世代 IX 研究会を設立した．5 章では，同研究会で構築・運用を行っている広域分散 IX の実証実験について報告する．

なお，本稿では広域分散 IX を実現するための技術的な側面からの研究を対象としている．IX によるビジネスモデル，あるいは社会的な影響は本稿の対象外とする．

2. IX—Internet eXchange

本章では，IX の特徴を明確にするため，プライベートピアリングと IX の仕組みについて述べる．また，従来の IX で用いられている技術として，LAN（Local Area Network）スイッチを用いるもの，および ATM（Asynchronous Transfer Mode）スイッチを用いるものについて述べる．

2.1 プライベートピアリングと IX

ISP が相互接続を行う場合，何らかの手段で物理的な接続を行い，BGP4（Border Gateway Protocol version 4）[8] による経路情報の交換を行ったうえで，トラフィックの交換を行う．この際，相互接続に用いる物理的な回線の形態により，相互接続の方法はプライベートピアリング，および IX に分類される．

プライベートピアリングでは，相互接続を行おうとする 2 つの ISP 間に専用の回線を準備して直接的に相互接続を行う．プライベートピアリングはほかから独立した環境で 2 つの ISP 間のみの相互接続を行うため，物理的な構成やトラフィック制御などで自由度が高い．半面，1 つの ISP は相互接続先の ISP ごとに専用の回線を準備することになる．

ISP 数が増加し，さらにピアリングがさかんに行われるような状況では，プライベートピアリングを行うのはスケーラビリティの観点で問題がある．たとえば，ISP の数を N とし，1 つの ISP は N に対して一定割合 p の ISP に対してピアリングを行うとすると，全体では $N \times (N-1)/2p$ 本の回線を準備する必要がある．極端な場合，ISP 間の完全なメッシュ状の相互接続環境を実現するためには，全体で $N \times (N-1)/2$ 本の回線を準備する必要がある．すなわち，全体の回線数は $O(N^2)$ のオーダである．

IX はプライベートピアリングに比較して効率的に ISP 間の相互接続環境を実現する．IX は相互接続の「場」を提供し，各 ISP は IX に接続するための回線を用意する．IX 内では各 ISP 間の相互接続を行うことができ，機能的に前述のプライベートピアリングによる完全メッシュの相互接続と同等の環境を実現できる（**図 1**）．このため，IX 上での相互接続はパブリックピアリングとも呼ばれる．本稿では IX の提供者を IXP と呼ぶ．ISP は IXP からみてユーザに相当する．

IX を用いて ISP 間の相互接続を実現する場合，回

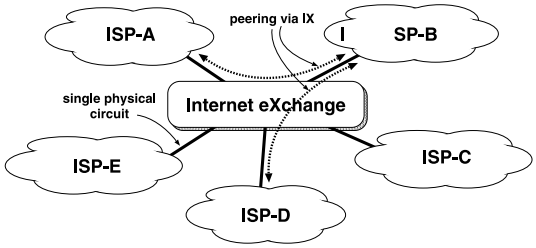

図 1 IX の概念図
Fig. 1 IX model.

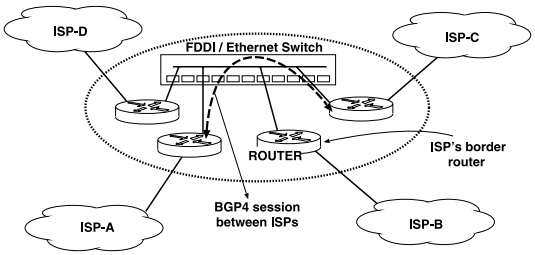

図 2 LAN-IX の例
Fig. 2 An Example of LAN-IX.

線数は全体で $O(N)$ であり，この意味で，プライベートピアリングと比較して効率的で拡張性に優れている．また，1 つの ISP からみた場合，単一の物理回線で複数の ISP と相互接続することでトラフィックを集約できるため，回線コストを抑えることが可能である．

2.2 既存の IX 技術

IX で利用されている既存の技術は，OSI 参照モデル第 2 層での交換技術を用いて ISP 間の相互接続を行うのが一般的である．本稿ではこれらの IX を L2-IX と呼ぶ．L2-IX には LAN（Local Area Network）の技術を用いるものと，ATM（Asynchronous Transfer Mode）の技術を用いるものに分類される．ここでは，それぞれの技術の特徴について述べる．詳細は文献 4) を参照されたい．本節で述べる技術は，後述の 4 章で階層型の IX を実現する手法について述べる際に参照する．

2.2.1 LAN-IX

現在，多くの IX ではイーサネットなどの LAN の技術を用いて相互接続環境を実現している．本稿では同技術を用いた IX を「LAN-IX」と呼ぶ．

IX に接続を行う ISP はそれぞれのルータ（境界ルータ）を LAN スイッチに接続する．LAN スイッチは論理的に 1 つの LAN セグメントと見なされ，IX に接続するすべてのルータに共通のサブネットとして機能する．各 ISP のルータは同サブネット上で BGP4 による経路情報の交換を行い，トラフィックの交換を行う．

現状で，LAN-IX に用いられるのはイーサネットスイッチが多い．100 Mbps イーサネットスイッチが出現する前には，LAN スイッチとして FDDI スイッチが用いられることが多かったことから，現在でも FDDI スイッチを用いている場合がある．原理的にはバッファードリピータやダム HUB も利用可能であるものの，現実にはほとんど利用されない．

図 2 は LAN-IX の基本的な構造を示したものである．これは ISP の境界ルータが IXP にロケーションされている場合である．

2.2.2 ATM-IX

ATM 技術を用いた IX は ATM スイッチ，もしくは複数の ATM スイッチからなる ATM 網から構成される．接続 ISP は ATM インタフェースを持ったルータを IX に接続する．IX は接続 ISP のルータ間に PVC（Permanent Virtual Circuit）と呼ばれる仮想的な回線を設定し，ISP 間の相互接続を実現する．本稿では本技術による IX を「ATM-IX」と呼ぶ．

ATM-IX では，ISP のルータ間に確立された PVC はポイントツーポイントの仮想的な回線として見なすことができる．各ルータは仮想回線上で BGP4 による経路制御を行うとともに，トラフィック交換を行う．

3. MPLS-IX

著者らは MPLS（Multi-Protocol Label Switching）技術を IX に応用した次世代の IX アーキテクチャ**MPLS-IX** を提案している[6]．本章では，MPLS の概念，**MPLS-IX** の仕組み，および特徴について簡単に述べる．本稿で提案する広域分散 IX の構築手法は **MPLS-IX** に基づいている．

3.1 MPLS の概念

MPLS は，IP パケットに固定長のラベルを付加することで網の内部で柔軟なトラフィック制御を行うための技術である．

MPLS 網は LSR（Label Switching Router）と呼ばれるルータ集合のネットワークとして構築される．MPLS 網と外とを接続するルータを Edge LSR と呼び，そうでないルータを Core LSR と呼ぶ．MPLS 網内は OSPF や IS-IS といった IGP により経路制御が行われる．

MPLS 網でトラフィックを交換する場合には，まず Edge LSR 間に仮想的なパスを確立する．これを LSP（Label Switched Path）と呼ぶ．そして LSP の上にラベル付けされたパケットが流れる．

LSP は 1 つの Edge LSR から 1 つ以上の Core LSR

を経由して他の Edge LSR に至る仮想パスである．LSP の確立には，LDP（Label Distribution Protocol）や RSVP-TE や BGP4 といったシグナリングプロトコルが用いられる．LSP が確立すると各々の LSR 間で，その LSP で用いるインタフェースと固定長のラベル番号が決定される．

データを MPLS 網内で転送する際は，まず Edge LSR でパケットにラベルが付与される．ラベルは Edge LSR の持つ IP 経路とラベル番号の対応表によって決定される．ラベルが付与されたパケットは，対応表に基づいて選択されたインタフェースに送信される．Core LSR では，ラベル表に基づいて転送が行われる．この場合，入力パケットに付与されたラベルのみで，出力パケットのラベルとインタフェースが決定される．最終 Edge LSR に到達するとラベルは除去され，対応表に従って IP の経路制御に従い MPLS 網から出力される．

このように，データトラフィックはあらかじめ確立された LSP に沿って転送される．また，LSP は一方向であり，双方向の通信をするには 2 本の LSP を必要とする．

3.2 MPLS-IX の仕組み

MPLS-IX では，ISP 間の接続に MPLS 網を用いる．IX の構造全体のうち，IXP は MPLS 網の Core LSR すべてを，ISP は MPLS 網の Edge LSR の 1 台を保持する．そのうえで，ISP は Edge LSR を IXP の Core LSR のどれかに接続する．

このうえで，ISP 側の Edge LSR 間に双方向に LSP を確立する．これがピアリング，すなわち相互接続の成立であり，この LSP 上でデータトラフィックを交換する．

通常，MPLS は単一の管理ドメイン内で使用されるのに対し，**MPLS-IX** では，Core LSR と Edge LSR とが IXP と ISP に分散しており，なおかつ Edge LSR は複数の異なる ISP 間に存在する．**図 3** に MPLS-IX の構造を示す．

ピアリングが確立する場合の各プロトコル間の関係を表記したのが**図 4** である．一番下が物理層とデータリンク層を示しており，両端が ISP の保持する Edge LSR で，中央部の 4 台は IXP の保持する Core LSR である．

MPLS 網では IP の到達性を確保し，LDP あるいは RSVP-TE といったシグナリングプロトコルを利用可能にする（図の下から 2 段目）．ピアリングする場合には，Edge LSR 間で LSP を確立する（図の上から 2 段目）．LSP が確立できたなら，ピアリングを

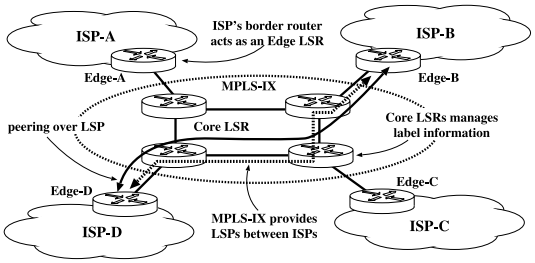

図 3 MPLS-IX の構造
Fig. 3 Structure of MPLS-IX.

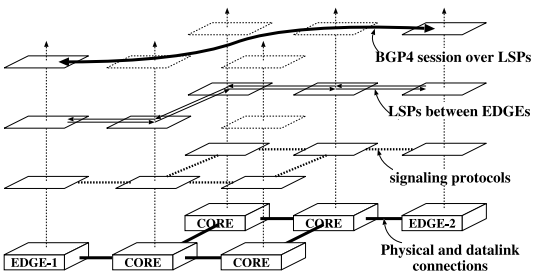

図 4 MPLS-IX の仕組み
Fig. 4 Architecture of MPLS-IX.

行う ISP 間で BGP4 のセッションを張り経路情報の交換を行う（図の一番上）．ISP はこの経路制御に従って LSP を経由してデータトラフィックを交換する．

なお，ここでプロトコルが構成する層は，階層型 IX の階層とは直接関係ないことに注意してほしい．

3.3 MPLS-IX の特徴

MPLS 技術を IX に応用することの最大の利点は，データリンクメディアに非依存な相互接続環境を実現できることである．MPLS はデータリンク層，およびネットワーク層に非依存に設計されており，異なるデータリンクメディアで接続するルータ間でも，仮想的なパス（LSP）上でデータを交換することができる．MPLS-IX ではイーサネット，ATM，POS（Packet over Sonet）など任意のデータリンクメディアを用いて相互接続が可能である．

L2-IX 技術は LAN や ATM などのデータリンク層に強く依存した仕組みであり，接続 ISP はデータリンクメディアを自由に選ぶことはできない．一方，MPLS はデータリンク層に非依存であり，たとえば，現在事実上最大の通信速度を提供する OC-768 POS を用いることにより IX 上で 40 Gbps の通信速度を実現することも可能である．

また，**MPLS-IX** は IX 機能を MPLS 網として実現することで，IX のトポロジを柔軟に構成することができる．LAN-IX のような単一のスイッチによる実

現手法と異り，極端な地理的制約を発生しない．さらに，MPLS の動作が IP ネットワーク技術を用いていることにより，**MPLS-IX** を設計する際に大きな自由度を持ち拡張性に優れている．

管理上の特徴としては，**MPLS-IX** の ISP と IXP との責任分界点を明確に定めることができることがある．MPLS 網内の IGP 経路制御は，Core-Core LSR 間と Core-Edge LSR 間とで独立しており，Core-Edge LSR 接続回線を責任分界点とすることで，責任範囲を装置の所有関係とほぼ同等に定義できる．ただし，回線費用の分担や運用の形態によっては回線のどちらかの端点を責任分界点としてかまわない．

4. 広域分散 IX の実現手法

本章では **MPLS-IX** アーキテクチャを用いて広域分散型の IX を実現する手法について述べる．**MPLS-IX** アーキテクチャを採用した理由は，3.3 節に述べた特徴が広域分散型の IX に好都合と判断したためである．

以下では，まず，広域分散 IX が求められている背景について紹介する．次に，**MPLS-IX** アーキテクチャを用いて階層型 IX を実現する方法について述べ，同技術を用いて相互に IX を接続することにより，階層性を持った広域分散 IX を実現する手法について提案する．最後に，広域分散 IX を実現するためのアーキテクチャとして，**MPLS-IX** を採用した場合と，他のアーキテクチャを採用した場合との得失について議論する．

4.1 広域分散 IX の必要性

国内のインターネットでは広域分散 IX の実現が強く求められている．これまで，国内のインターネットは市場規模の差異に起因して，東京一極集中型のトポロジを維持したまま成長してきた．このため地域間，もしくは地域と東京の通信は制約が多いことが一般的である．広域分散 IX の技術は，これらの遠距離の通信においても，通信経路，帯域，通信遅延などの設計が容易で，より高速で安定した通信の実現を可能にする．

広域分散 IX 技術は，次のような利用方法が期待されている．

- コンテンツプロバイダと各地アクセスプロバイダとの直接接続が可能になる．ゲーム[24] やストリーム[25] などのコンテンツを広域分散 IX を通してエンドユーザに直接配信することは，高スループット・低遅延かつジッタなどの揺らぎをおさえ，高品質なコンテンツの配信を可能にする．
- 映像伝送，ファイル共有など地域間の通信の実現が可能になる．これまでも，国体映像や CATV の映像を地域間で交換する取り組みが実験的に行われている[3),5)]．
- 大手プロバイダが地域でトランジット（IP の接続性を提供する）サービスを提供することが可能になる．これまで，外資系プロバイダは国内の回線を有していないため，地域でトランジットサービスを提供することはできなかった．広域分散 IX は，東京を拠点にするプロバイダが，東京のユーザと同等に近いサービスを国内全域に提供可能にする．JANOG9（JApan Network Operators Group）の会議では，後述の広域分散 IX の実験網を介して AboveNet，日本テレコムの 2 社がトランジット提供の実験を行った[26]．
- 商用 IX の地方展開が可能になる．現在も JPIX，JPNAP などの商用 IX が大阪や名古屋に拠点を増やそうとしている．これらの商用 IX はこれまで「点」の IX として，特定のビル内でのみサービスを提供してきたが，広域分散 IX を用いることにより容易に地方展開が可能になる．

上述の要求を実現するために，以下の目標を達成することを目指した．

- L2-IX を相互に接続することを目的とした IX を構成できる技術を提案する．これにより，L2-IX と “メタ” IX とで階層的な構造を持たせ，全体として巨大な仮想 IX を構成する．
- 1 つの接続装置や特定の地理的な場所に依存しないような，広域分散 IX の実現手法を提案する．また，運用・管理上の課題を抽出・整理し，より実用性を高めるための実証実験を行う．
- ISP と IXP との責任分界点をを明示し，管理・運用を容易にする．

4.2 MPLS-IX による階層型 IX

本節では，**MPLS-IX** を用いて階層型 IX を実現する手法について述べる．本稿では，IX を構成するネットワークを以下の 2 つに分類する（**図 5**）．

- IX バックボーン
- IX リーフ

IX バックボーンは，Core LSR および Core LSR 間のネットワークとして構成する．IX バックボーンは IX リーフの情報を含め，IX 全体の接続情報を管理する．Core LSR 間では OSPF や IS-IS などのリンクステート型の経路制御プロトコルで経路情報を交換し，IX 全体のトポロジ情報を更新する．

IX リーフは，Edge LSR と接続するための Core LSR におけるインタフェース，およびそのインタフェー

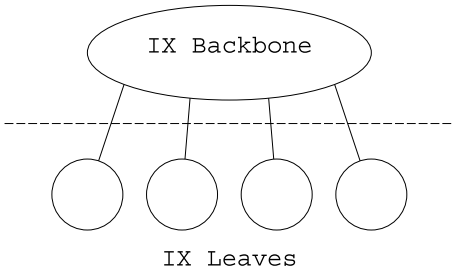

図 5 MPLS-IX による広域分散 IX のモデル
Fig. 5 A model of widely distributed MPLS-IX.

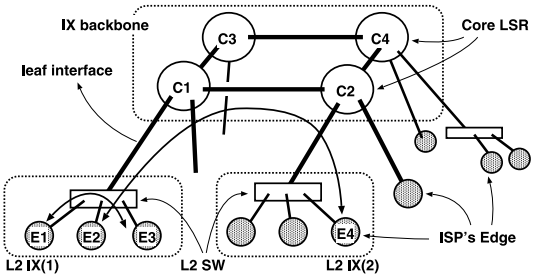

図 6 MPLS-IX による広域分散 IX
Fig. 6 An example of widely distributed MPLS-IX.

スの IP アドレスが所属するサブネットを示す．各リーフは，独立したサブネットを構成する．

本提案において，IX リーフは L2-IX（ないしは ISP）を，IX バックボーンはそれらを接続する “メタ”IX として機能する．この構造をもって階層型 IX と命名した．

4.3 広域分散 IX の実現手法

IX バックボーンは IP ネットワークである．したがって，IX バックボーンを WAN（Wide Area Network）の技術を利用して広域に分散させて構成することにより，階層型 IX を広域分散環境に適用することができる．

階層型 IX では，IX リーフを構成する際，Edge LSR の接続形態により次のいずれかの構造をとる．

- 直接接続
- データリンク層のスイッチを介しての接続

前者は，Core LSR と Edge LSR が 1 対 1 に物理的に直接接続されるケースである．この場合，1 つの Edge LSR が 1 つの IX リーフを占有する．後者は，1 つの Core LSR に対し複数の Edge LSR がスイッチ経由で接続される場合である．**MPLS-IX** はデータリンク層に非依存であるため，ATM-IX や LAN-IX といった L2-IX を IX リーフとして接続可能である．

図 6 に階層型 IX を用いて構成される広域分散 IX を模式的に示す．白丸で表された C1〜C4 は IX の Core LSR を表している．Core LSR とこれらの間を接続するネットワークが IX バックボーンを構成している．L2-IX および Core LSR への接続回線，あるいは L2-IX なしの Edge LSR と Core LSR への接続回線が IX リーフである．

広域分散 IX では，MPLS-IX を提供する IXP の管理・運用範囲を IX バックボーンとし，L2-IX を提供する IXP の管理・運用範囲を IX リーフとする．すなわち MPLS-IX の Core LSR での Edge LSR 側のネットワークインタフェースが責任分界点となる．こうすることにより，ハードウェアの管理と回線管理および経路制御の管理とがほぼ一致し，管理・運用を行うことが容易となる．

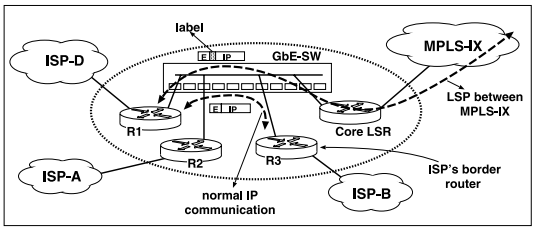

図 7 MPLS-IX による階層構造
Fig. 7 Hierarchy of MPLS-IX.

4.4 L2-IX との共存

本稿で提案する広域分散 IX の構築手法において，L2-IX 内での通信は **MPLS-IX** を用いる場合の通信と共存が可能である．すなわち，L2-IX に対して，広域分散 IX の IX リーフを接続するだけで，それまでの環境を維持したまま広域分散 IX 対応にすることができる．このことは，すでにあるリソースを有効に活かすことで広域分散対応へのコストや移行への手順を容易にするばかりではなく，トラフィックの集約効率を上げる意味でも重要な意味を持つ．

図 7 は L2-IX での通信と **MPLS-IX** を介した通信が共存可能なことを LAN-IX を例にして示している．L2-IX 内での通信は，イーサネットスッチを介して，IP パケット（図で IP と記述されている箱）にイーサネットヘッダ（E と記述されている箱）が付与されて通信を行う．一方，L2-IX に接続する ISP のルータが **MPLS-IX** を介して通信を行う場合には，IP パケットとイーサネットヘッダの間にラベル情報が挿入されることにより通信を行う．L2-IX で利用されるスイッチでは，通常の IP による通信と，MPLS による通信は互いに干渉しない．

4.5 他の手法との比較

以下では，**MPLS-IX** を用いて広域分散 IX を実現する場合と，それ以外の技術を用いて広域分散 IX を

実現しようとする場合について，技術的な面からの比較を行う．

- ATM 網
  IX 構造として広域 ATM 網を用いることにより，広域分散環境で ISP 間の相互接続が可能である．ただし，ATM のみを用いて大規模な広域分散 IX を実現しようとした場合，PVC を設定する運用の手間が非常に大きいことが問題になる．また，他のデータリンク技術との相互接続性はないため，多くの IXP が利用している LAN-IX との接続はできない．
- 広域 LAN/広域イーサネット
  近年，広域 LAN や広域イーサネットなどのサービスが提供され始めている．同技術を用いることにより，広域で LAN-IX を構築することが可能である．しかし，イーサネットによるサブネットを共有するため，接続する ISP の数がサブネットのアドレス空間の広さの制限を受けるなど，拡張性に問題がある．また，共通のサブネットを広域分散環境で共有することになり，管理上の責任分界点の定義が困難になるという問題がある．また，別の L2-IX と接続を行う場合には，境界ルータの IP アドレスの変更が必要になるなどの問題がある．さらに，L2-IX と接続を行って運用する場合に，IX 用の共通サブネットのアドレス割当ての管理，あるいはブロードキャストパケットによる障害の切り分けが困難となる．加えて，ATM-IX との接続性はないため，ATM で L2-IX を構築している場合には接続は不可能である．
- IP-VPN
  RFC2547[9] は MPLS 上で VPN（Virtual Private Network）を実現する手段を定義している．この技術はすでに通信事業者がサービスを提供しており，企業などが社内網を構築する際に利用している．MPLS による VPN は，Core LSR においてユーザ経路をすべて保持管理する．これを IX として用いると，ISP の持つ多量の経路情報を IX の Core LSR が交換することになる．このため，VPN による広域分散 IX は事実上に不可能である．
- Optical-IX
  光スイッチの技術を応用した **Optical-IX** が提案されている[7]．光スイッチ技術により高速広帯域による相互接続が可能である．しかし，ルータのインタフェースが DWDM（Dense Wavelength Division Multiplexer）に対応していることを想定しており，各組織が準備するルータが非常に高価なものになる．また，Optical-IX で必要とされる GMPLS（Generalized MPLS）[11] や DWDM 対応のルータの実装は数年先といわれており，当面の実現技術としては不適切である．

5. 広域分散 IX の実証実験

著者らは，通信・放送機構の委託を受けて，次世代広域分散 IX 技術の研究を行うために，次世代 IX 研究会を設立した[23]．同研究会では，**MPLS-IX** アーキテクチャを用いた相互接続技術の確立と，同技術を用いた広域分散 IX の実証実験を展開している．本章では次世代 IX 研究会で進めている実証実験の概要とその結果について報告する．

5.1 実証実験ネットワーク

実証実験は JGN（Japan Gigabit Network）上に構築された広域分散環境のテストベッド上で行っている．**図 8** にテストベッドの物理層のトポロジを示す．

JGN と書かれた雲型が本実験の IXP を構成する物理網である．楕円が Core LSR を，角の丸い箱型が Edge LSR を示している．Edge LSR は AS（Automonous System）であり，AS 名と AS 番号を示している．点線の箱型は今後の予定である．以下で，この構造について詳しく述べる．

なお，図で下の 2 つの雲型は，他の **MPLS-IX** 事業者との相互接続を示している．現在，日本テレコム社と MCI WorldCom 社が試験的に **MPLS-IX** による相互接続環境を提供している．これについては本稿では詳細を述べない．

5.1.1 Core LSR

現在，JGN に構築した広域分散 IX のテストベッドは 6 つの Core LSR を持つ．Core LSR には Juniper 社製の M10 もしくは M20 を用いている．Core LSR は東京（u-tokyo，notemachi，kotemachi），大

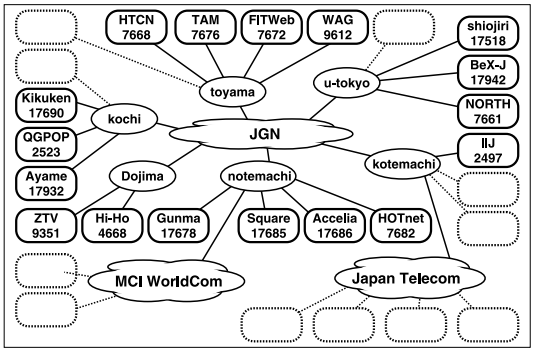

図 8 広域分散 IX のテストベッド
Fig. 8 The testbed of distributed IX.

阪（dojima），富山（toyama），高知（kochi）に分散配置され，その間を JGN 上の ATM PVC を用いて接続している．

東京では，東京大学，NTT 大手町ビル，KDDI 大手町ビルにそれぞれ Core LSR を設置している．これは，実証的な相互接続実験を行うため，国内の ISP が特に集中しているこれらの通信拠点に接続点を置くことが有効と判断したためである．

5.1.2 Edge LSR

実験参加組織は，まず Edge LSR を Core LSR へ接続する．接続は以下のいずれかの方法で行う．

- JGN を用いて接続を行う．接続組織が JGN に接続し，いずれかの Core LSR まで ATM PVC を確立する．
- Core LSR に直接接続する．接続組織がルータを持ち込むか，もしくは回線を引き込むことにより，いずれかの Core LSR に対して直接接続を行う．接続可能なデータリンクメディアは ATM，POS，GbE である．

次に，Edge LSR の設定を行う．設定はおおよそ以下のステップで行う．

- Edge LSR を用意する．
  現状では Juniper 製ルータ，Cisco 製ルータ，そして PC ルータを用いるという 3 つの例がある．PC ルータは OS に NetBSD [27]，経路制御に Zebra [28]，MPLS 機能に Ayame [29] を導入したものを用いる．
- Edge LSR の一方に自組織のネットワークを接続する．
  自組織側では任意の IGP を用いて良い．静的経路制御でもかまわない．
- Edge LSR の他方に Core LSR への接続を行う．
  このとき MPLS のラベルを付与することによるトラブルを避けるため，MTU を 1508 octet にすることを推奨している．直接接続する Core LSR には静的経路制御を用いる．
- LDP か RSVP-TE により LSP を確立する．
  この場合に利用するアドレスとして，ルータのループバックアドレスかインタフェースアドレスかを選択する．LSP を通過する際に IP の TTL と MPLS の TTL とが独立に減らされるように，特に LSP 通過により IP の TTL が 1 だけ減るような設定をする．
- BGP4 のセッションを確立する．
  この場合，LSP が消失した際に BGP4 のセッションが切断するように，BGP4 の multi-hop 数が 1 になるようにする．
- データの交換が可能か試験する．
  ping，ftp，netperf などのツールを用いて，LSP 上にデータが流れることを確認する．

なお，接続の際には事前にピアリングの相手を確認し，相手の AS 番号を得ているものとする．

5.1.3 階層型 IX 実験

富山（toyama）において階層型 IX の仕組みに関する実験を進めている．現在，富山地域では ATM スイッチ，およびギガビットイーサネットスイッチを介して本テストベッドに接続できるようになっており，それぞれ，県内のフリーウェイ（ATM ネットワーク）および富山地域 IX 研究会（イーサネットによる L2-IX）に接続されている．県内の ISP や研究機関は，それぞれに接続しているネットワークを介して広域分散 IX の実証実験に参加することが可能である．

5.2 実証実験の状況

現在，次世代 IX 研究会の実証実験では 22 組織が相互接続に参加してトラフィック交換を行っている．実証実験には大学や研究機関のほか，ISP，CSP などの民間企業も接続している．現在は，研究目的を主としているため，特定のアドレス空間での通信や，地域イベントのコンテンツ配信用のトラフィックなど，実験用のトラフィック交換が中心となっている．これは，**MPLS-IX** という新技術を用いて相互接続を行っているため，安定運用に至るまでの経過的な措置である．

図 9，**図 10** は実証実験において交換されているトラフィックの合計値を示している．これは 2002 年の計測データである．図は，テストベッドを介して交換されたトラフィックの合計を 5 分間隔で 2 カ月間にわたって計測したものである．図 9 はトラフィックの推移を見やすくするため，Y 軸を 200 Kbps で抑えている．機能検証を目的とした実験トラフィックが主であ

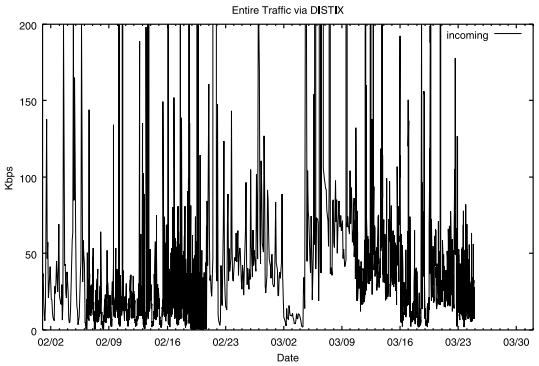

図 9 トラフィック状況（1）
Fig. 9 Traffic over the testbed (1).

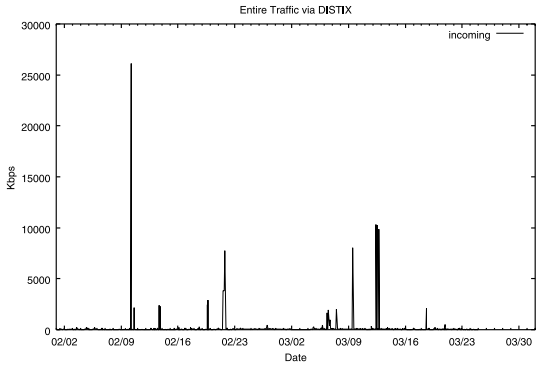

図 10 トラフィック状況（2）
Fig. 10 Traffic over the testbed (2).

るためトラフィックの総和は必ずしも多くはないものの，継続的なトラフィック交換が行われていることが分かる．

一方，図 10 は，おなじ期間のトラフィックを Y 軸を 30,000 Kbps（30 Mbps）にして示したものである．図に示すように，広域分散 IX では瞬間的に 10〜30 Mbps 程度のトラフィックが交換されていることが分かる．なお，これらの瞬間的な高トラフィックの通信は，映像伝送などのイベント時のトラフィックである．次世代 IX 研究会では，実験目的のアプリケーションとして特定のストリームを定期的に流している．これらの実験から，本テストベッドにおいて，広域分散 IX により遠距離での相互接続環境上で高帯域アプリケーションを利用することが可能になったことが分かる．

5.3 MPLS-IX の課題

次世代 IX 研究会では，これまで広域分散 IX としての基本的な機能の検証を中心に実証実験を行ってきた．これまでの実証実験では，多くの組織間で順調に相互接続ができている．実証実験では相互接続実験を開始してから約 3 カ月間，計画的な停止を除いて，相互接続上の深刻な問題は発生していない．このことからも **MPLS-IX** アーキテクチャを用いた広域分散 IX の仕組みが非常に安定して動作していることが分かる．

一方，実証実験を通していくつかの細かな課題が指摘されている．以下では，これまでの実証実験で明らかになった **MPLS-IX** の課題についてまとめる．

- MPLS ではパケットにラベルを付与するため，通信経路上の MTU（Maximum Transfer Unit）の設定に注意が必要である．実証実験でも，MTU の設定ミスにより，ストリームによる映像を見ることができないなどの問題が起こった．一般に，各 LSR 間では MPLS のラベル部分を除いて MTU が 1500 Octets 確保されるよう設定すること必要である．
- ルータの実装の差異を原因とする問題の整理が必要である．多くの MPLS ルータは **MPLS-IX** に必要な基本機能の実装を終えているが，一部のルータでは，たとえば，BGP4 の TTL（Time To Live）の値が正しく指定できないなど，運用上の支障，もしくは混乱を招きやすい実装が残っているケースがある．次世代 IX 研究会では，これらのルータの実装上の問題についてまとめ，同時に各ベンダに対して対応を依頼している．

6. 考　　察

本章では，特に 4.1 節で述べた技術上の目的についての考察を行う．

6.1 階層型 IX

大きな目的の 1 つは，L2-IX を相互に接続するための適切な IX 技術を提案することであった．本提案は，データリンクメディアへの非依存性や，L2-IX との非干渉性，明快な責任分界点の定義といった MPLS-IX の特性を活かすことができる．さらに，MPLS-IX の構造より，IXP が維持する経路情報は MPLS 網内の到達性を維持すればよく，各 ISP が BGP4 で交換する経路情報を管理しなくてよい．このため，**図 4.5** 節で示したように，MPLS-VPN アーキテクチャと異りスケーラビリティが高い．以上により，L2-IX を相互接続するために MPLS-IX を用いることは適切であると結論づける．

一方で，MPLS-IX が複数存在する場合に，それらをどのように接続するのが適切かは明らかでなく，今後の研究を行う必要がある．

6.2 広域分散 IX

MPLS-IX の Core LSR を WAN 技術で広域に分散することにより，従来の 1 点型の IX とは異る仮想的な広域 IX を構成することが可能になった．これにより IXP や ISP がかかえる問題の一部を解決することができる．

一方で，広域分散 IX に必要な冗長性や QoS といった機能の実現は完成していない．たとえば，ISP に対する保証伝送速度や IXP に対する最大伝送速度などを契約によって設定することができるようにしたいという要求がある．これは IX が QoS 制御機能を持つことで実現できるはずであるものの，現状では扱うことができないでいる．

広域分散 IX の実証実験については，管理・運用上の多くのノウハウが得られた．これにより，現状では管理に必要な情報が不十分であることも分かってき

た．たとえば，今存在する LSP やラベル交換の状況を確認することが難しいことや，LSP のトラフィックを SNMP で計測したくても MIB がないなどの問題が発見できている．

6.3 責任分界点

階層型の IX を提案するにあたり，どこに IXP と ISP との責任の分界点を置くかは管理・運用上重要な問題である．MPLS-IX を用いることにより，より自然な管理上の境目を定義できた．これはトラブルシューティングの効率に深く関係し，実用上で大きな利便性を得られると考える．

7. おわりに

本稿では，**MPLS-IX** アーキテクチャを用いて広域分散 IX を実現する手法について述べた．本稿で提案する手法は，IX を階層性を持たせて相互接続することにより，既存のリソースを有効に利用しながら，スムーズに広域分散環境での相互接続環境へ移行することを可能にする．

また，著者らは次世代 IX 研究会において，広域分散 IX の実証実験を行っている．本稿では，これらの実証実験の概要について紹介するとともに，これまでの実験の状況についても報告した．

広域分散 IX は，コンテンツ事業者とアクセスプロバイダの直接接続や，地域間での広帯域アプリケーションの利用，大手 ISP のトランジットサービスの地域での利用，あるいは商用 IX の地方展開など，さまざまな利用方法が期待されている．本研究は，これらの要求に対して，技術的な面から実現方法を示した．また，実証実験を通じて，その実用性についても示した．

今後は，広域分散 IX 技術において，IPv6 などの次世代プロトコルへの適用，あるいは QoS（Quality of Service）などの品質保証に関する技術についても研究を進めていく．

謝辞 本研究の実施にあたって有益なご意見をいただいた林英輔教授，および全国の IX 関係者に感謝いたします．また，本研究は通信・放送機構の委託研究（委 121-401）に基づいて実施しています．

参考文献

1) Huston, G.: Interconnection, Peering and Settlements, *The Internet Protocol Journal*, Vol.2, No.1 (Mar. 1999).
2) 中川郁夫，江崎　浩，永見健一：ラベルスイッチを用いた分散 IX の設計，分散システム/インターネット運用技術研究会研究報告，99-DSM-14 (Jul. 1999).
3) 中川郁夫，林　英輔，樋地正浩，八代一浩，菊池　豊，西野　大：ギガビットネットワークを用いた地域間相互接続の試み，分散システム/インターネット運用技術研究会研究報告，99-DSM-15 (Sep. 1999).
4) 中川郁夫，林　英輔，高橋　徹，江崎　浩：次世代インターネットエクスチェンジの技術動向，情報処理，Vol.42, No.7 (Jul. 2001).
5) Kikuchi, Y., Nakagawa, I., Hiji, M., Yatsushiro, K., Nishino, D. and Hayashi, E.: A trial for reconstructing the ground design of the Internet architecture in Japan, *Proc. 2nd International Conference on Advances in Infrastructure for Electoronic Business, Science and Education on the Internet* (Aug. 2001).
6) Nakagawa, I., Esaki, H. and Nagami, K.: A Next Generation IX Architecture using MPLS, *SAINT2002*, Nara (Jan. 2002).
7) 中川郁夫，江崎　浩，菊池　豊，永見健一：光スイッチを用いた次世代インターネットエクスチェンジの設計，電子情報通信学会学会誌 (May 2002).
8) Rekhter, Y. and Li, T.: A Border Gateway Protocol 4, IETF RFC1771 (Mar. 1995).
9) Rosen, E. and Rekhter, Y.: BGP/MPLS VPNs, IETF RFC2547 (Mar. 1999).
10) Rosen, E., Viswanathan, A. and Callon, R.: Multiprotocol Label Switching Architecture, IETF Internet-Draft (April, 1999).
11) Ashwood-Smith, P., Banerjee, A., et al.: Generalized MPLS—Signaling Functional Description, IETF Internet-Draft (Sep. 2001).
12) Manning, B.: Exchange Point Information. http://www.ep.net/
13) PAIX: Palo Alto Internet eXchange. http://www.paix.net/
14) MCI WorldCom: MAE Information. http://www.mae.net/
15) LINX, LINX. http://www.linx.net/
16) WIDE Project, NSPIXP. http://jungle.sfc.wide.ad.jp/NSPIXP/
17) JPIX, JaPan Internet eXchange. http://www.jpix.ad.jp/
18) JPNAP, JPNAP Service. http://www.mfeed.ad.jp/jpnap/main.html
19) BeX-J, Business eXchange Japan. http://www.bex-j.net/
20) Tohoku Regional IX. http://www.tia.ad.jp/trix/
21) OKayama IX. http://www.okix.ad.jp/
22) Toyama Regional IX Consortium. http://www.toyama-ix.net/
23) Next Generation IX Consortium. http://www.distix.net/

24) PlayOnline. http://www.playonline.com/
25) CRN Forum. http://www.crnf.net/
26) JANOG9 Meeting.
http://www.janog.gr.jp/meeting/janog9/’
27) The NetBSD Project.
http://www.netbsd.org/
28) GNU Zebra.
http://www.zebra.org/
29) Ayame Project.
http://www.ayame.org/

(平成 14 年 4 月 2 日受付)
(平成 14 年 9 月 5 日採録)

中川 郁夫（正会員）

1968 年 8 月 26 日生．1991 年東京工業大学理学部数学科卒業．1993 年東京工業大学大学院総合理工学研究科システム科学専攻修士課程修了．同年（株）インテック入社．同社研究所にてネットワーク管理，大規模経路制御技術，次世代インターネットに関する研究に従事．2002 年（株）インテック・ネットコア取締役．理学修士．

江崎 浩

1963 年 1 月生．1987 年九州大学大学院工学研究科修士課程修了．同年（株）東芝入社．1998 年東京大学情報基盤センター助教授．2001 年東京大学大学院情報理工学系研究科助教授に就任．MPLS および IPv6 に関する研究開発に従事．WIDE プロジェクトボード メンバ．IPv6 普及・高度化推進協議会専務理事．電子情報通信学会会員．工学博士（東京大学，1998）．

菊池 豊（正会員）

1992 年東京工業大学大学院博士課程単位取得退学．同年より同大学情報工学科助手．1997 年より高知工科大学情報システム工学科助教授．地域指向型のインターネットトラフィック交換の研究を行う．情報処理学会 DSM 研究会幹事．KPIX 実験研究協議会会長．博士（工学，東京工業大学，1994）．

永見 健一

1992 年 東京工業大学大学院理工学研究科修士課程修了．同年（株）東芝入社．IETF MPLS WG で標準化活動を行い，CSR および MPLS に関する RFC を提出．東芝開発センターで MPLS および IPv6 の研究に従事．2002 年（株）インテック・ネットコア入社．工学博士（東京工業大学，2001）．