第 6 部

プロバイダ間の接続と経路制御技術

第 1 章

はじめに

国内で多数の ISP が急速に立ち上がり、現在、JPNIC により割り当てられた AS 番号の数も 100 を越えた。これらの ISP は、複数の対外接続をもつ場合が多く、同一 ISP に対して複数の経路が存在することももはや一般的になりつつある。このような状況のもとで、ISP にとっては、他の ISP に至る複数の経路の中からなんらかのポリシのもとで経路の選択を行なうことが不可欠となってきている。

また国内での ISP 間の相互接続をとりもつ NSPIXP などのいわゆる IX も 96 年度中にはより整備が進み、国内数個所にさまざまな IX が誕生することが予測されている。これにともない、複数の IX に接続を持ちそこで国内のトラフィックを交換する ISP も増えて来ることが予想される。

現在、このようなポリシールーティングを行なう為のプロトコルとして、BGP4(Border Gateway Protocol 4)[?] が広く用いられるようになって来ている。BGP4 では様々なパス属性を用いて、ポリシの実現を可能としているが、現在の複雑な要求の全てを表現し実装を行なう為には不十分であり、経路上の複数の ISP のオペレータ間で密なコーディネーションを必要とする。

このような状況下で本 WG は、ISP 同士の国内での安定した相互接続性を向上させて行くことを目標とし、複数の ISP 同士が複数の場所で相互に接続を持った場合に生じるさまざまな経路制御上の問題を、円滑に実現するための方法について考察を行なった。

第 2 章

ISP 間相互接続の変遷

2.1 学術研究時代

NSFNET(National Science Foundation Network) バックボーンは、1986 年にアメリカの 5 箇所のスーパーコンピュータセンタを相互に接続するネットワークとして運用を開始した。それ以来、学術研究コミュニティーのインターネットを介した活動を支援するという AUP(Acceptable Use Policy) のもとで、全米各地の大学などを中心に作られていた地域 (regional) ネットを接続したり、アメリカ国外の学術研究ネットワークを接続した。さらに商用 ISP のサービスが開始されると、そのユーザと学術コミュニティーとの間の相互通信をとりもつために、それらの商用 ISP とも相互接続を行った。

これにより、ネットワーク間の相互接続が NSFNET バックボーンを経由して行なわれることとなった。建前上は、商用のトラフィックは NSFNET バックボーンを通過できないことにはなっていたが、事実上は NSFNET バックボーンに接続される様々なネットワーク間の相互接続を提供するバックボーンとしての機能を果たすこととなった。

学術研究用の全米規模のバックボーンネットワークは NSFNET 以外にも存在した。例えば NSI(NASA Science Internet) や、ESnet(Energy Sciences Network) など特定のミッションをもったネットワークがある。これらのネットワークは NSFNET よりも厳しい AUP のもとで運営されており、一般的に商用のトラフィックの通過は許していなかった。NSFNET の AUP は、学術研究コミュニティーに対して何らかの形で貢献できる活動全てが許されていたため、どこかの企業と大学の共同研究のためなどの理由が認められていた。これらの政府系の学術研究ネットワークは FIX(Federal Internet eXchange) を介して相互に接続されていた。

一方、商用 ISP は、直接リンクを持って相互接続を行なったり、CIX(Commercial Internet eXchange) や、現在の NAP(Network Access Point) の原型ともなる MAE-East(Metropolitan Area Ethernet) を作り、そこに接続された ISP 間で商用トラフィックを交換するなどして、徐々に NSFNET バックボーンなどの学術ネットワークを経由しない商用 ISP 間の相互接続を進めた。

このような動きの中 NSF も、従来の IP のバックボーンネットワークである NSFNET バックボーンが地域ネットや商用 ISP の相互接続を提供するという形を止め、その代わり

に全米数箇所に ISP 同士がデータリンク層 (イーサネット、FDDI など) で相互接続を行なうポイントである NAP を作り、各 ISP は NAP に直接接続したり、また、NAP に直接接続した ISP から NAP 経由での他の ISP へのコネクティビティーを買うことにより、インターネット全体への接続性を得る形への移行を打ちだした。その後 NSFNET は、1995 年の 4 月の終りにバックボーンのサービスを停止し、その役割を終えた。

2.2 Internet Exchange

この NSF の NAP 政策により、Internet Exchange を介した ISP の相互接続の形態が推進された。NAP 以外にも、MAE-Chicago, MAE-LA, MAE-Dallas, MAE-NY, MAE-Houston などの MAE 系の IX や、Tucson NAP, Phoenix REP (Phoenix Regional Exchange Point) などの地域 IX のような、さまざまな規模の Internet Exchange(IX) ができ、それらを介した ISP 間の相互接続が行なわれるようになった。

この動きは、アメリカ以外にも次第に広がり、日本では WIDE プロジェクトが運営する NSPIXP(Network Service Provider Internet Xrossing Point) が IX の構築運用実験を 1994 年から行なっている。初期の NSPIXP-1 は、イーサネットスイッチを用いた ISP 間の相互接続を行なっていたが、日本での商用 ISP の急激な伸びに伴いそこを経由するトラフィックが急増し、現在では FDDI スイッチを複数台利用した NSPIXP-2 が運用されるに至っている。

2.3 NSFNET の停止と IX の台頭がもたらした変化

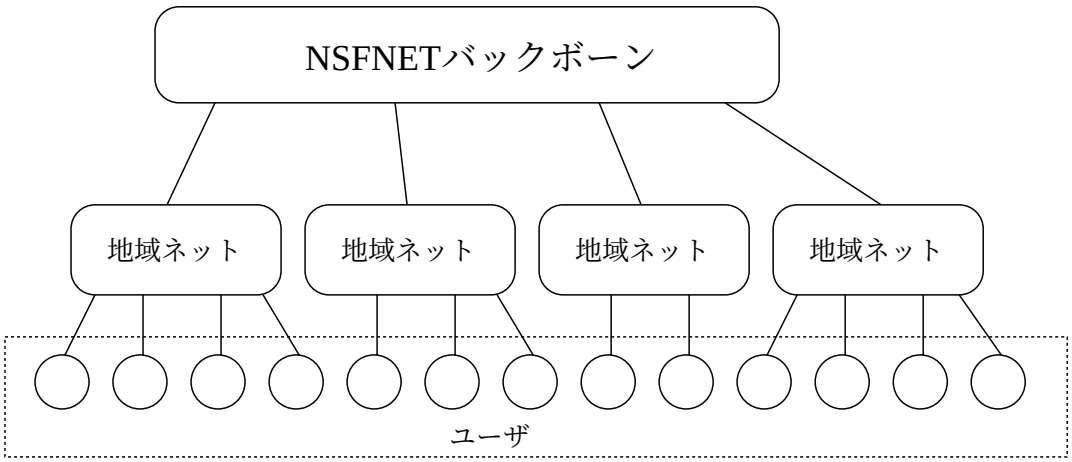

図 2.1: 学術研究時代

NSFNET バックボーンの停止と IX の林立が、ISP 間の相互接続の形態の変化をもたら

し、それによりインターネット全体の構造を大きく変化させた。

NSFNET を中心とした学術研究ネットワーク時代のネットワークの相互接続の構造を簡単に図示すると図 2.1のようになる。

ユーザは、大学などの研究機関を中心に構築されていた地域ネットに接続され、地域ネットは、NSFNETバックボーンに接続されていた。異なる地域ネットのユーザ間の通信は、多くの場合 NSFNET バックボーンを経由して行なわれていた。また、アメリカ国外のネットワークも、図の地域ネットと同様に何らかの形で NSFNET に接続をされていた。例えば日本の学術研究ネット WIDE や TISN などは、PACCOM(Pacific Computer COMmunication) プロジェクトのもと NSN に接続され、そこから FIX-West を経由して NSFNET などのアメリカの学術研究ネットワークに接続されていた。図では、地域ネットとだけ書いたが、さらに地域ネットの下に小さなネットワークが接続されている場合もある。ともかく、ネットワークの全体の構造はおおむね NSFNET を中心とした階層構造となっていた。

商用 ISP が登場し、これらが全米規模のバックボーンネットワークを構築するに至ると、この状況が若干変化してきた。

図 2.2にこの状況を示す。

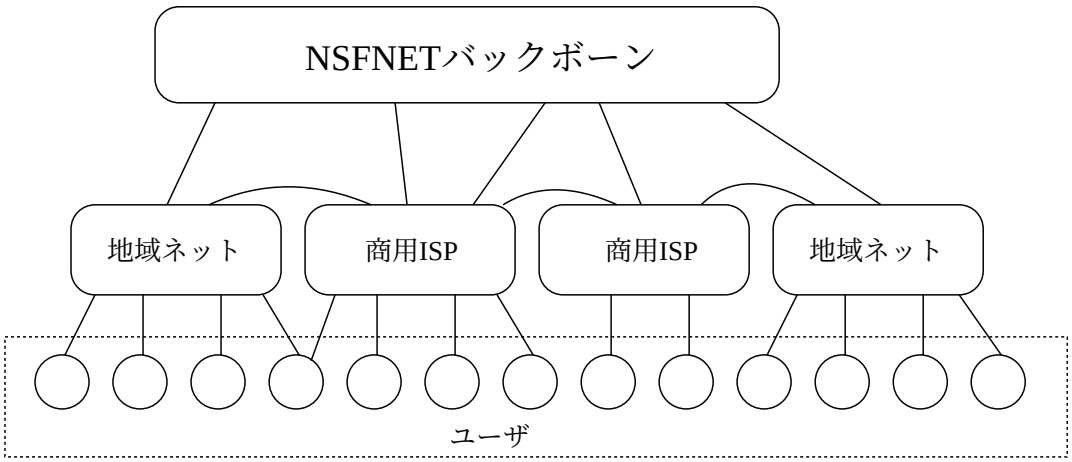

図 2.2: 商用 ISP の登場

商用 ISP は独自にバックボーンを構築し、ユーザを増やしていった。また、NSFNETバックボーンに接続を持ったり、地域ネットと相互接続を持ったり、また、他の商用 ISP と相互接続したりして、ユーザの全体的なコネクティビティーを広げていったのである。またこのころには NSFNET が NAP への移行の一環として、それまでの地域ネットの商用化を進めてきたこともあり、地域ネットの中には従来の学術研究目的のユーザに対するサービスを続けながらも、商用のユーザもつなぎ始め、いわば半商用とでもいうようなサービスを始めるネットワークも登場してきた。しかしこのころはまだかなりの部分の相互接続性が NSFNET バックボーンによって提供されていた。

さて、NAP が稼働を始め、地域ネットのほとんどが商用化し、NAP に接続を持つ ISP に接続されると、先に述べたように NSFNET バックボーンはその役目を終え、バックボーンサービスの停止を行なった。

この NSFNET バックボーン以降の状況が図 2.3の状況である。

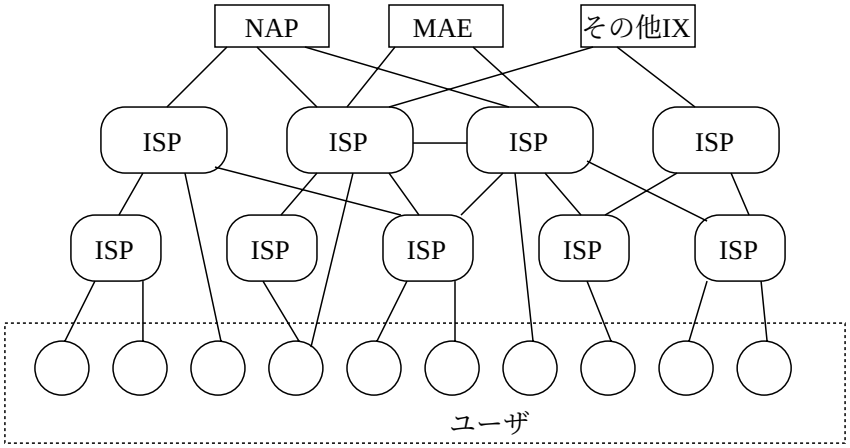

図 2.3: NSFNET 以降

ISP のいくつかは相互接続点である NAP や MAE やその他の IX の一つもしくは複数に接続を持ち、同じ IX に接続を持つ ISP との間でトラフィックの交換を行なう。また、IX に直接接続を持たない ISP は、IX に接続を持つ他の ISP に接続を持ち他の ISP との間のコネクティビティーを確保する。他の ISP との接続も一つではなく、複数持ったり、またユーザも、複数の ISP に接続する場合もある。

NAP などの IX は、NSFNET バックボーンとは異なり、データリンク層のサービスしか提供していない。従って、そこに接続したからといって、他の ISP と接続されたことにはならない。他の ISP と IX 経由で接続するためには、まず IX に接続し、次に同じ IX 上の ISP とトラフィックの交換をするためになんらかの契約を取り交わして、経路情報の交換を行なって始めて相互接続がされるのである。IX 自体は IP の階層からは見えなくなっている。

従って、NSFNET 以降の状況は、ISP 同士が IX 経由や直接リンクを張るなどの方法により、相互接続のメッシュを構成するようになった。つまり ISP 間の相互接続は、階層を持つツリー構造から、メッシュ構造へと大きく変化したことになる。

2.4　構造の変化の意味

インターネットの構造が変わるということは自ずとその上での経路制御の方法も変わることを意味する。インターネットがおおむね階層構造をしていた時代であれば、NSFNET

バックボーンを中心にした経路制御を考えれば良く、現在ほど細かなポリシーのもとでの経路制御も必要とされていなかった。しかし、ISP が複数の地点で相互に接続を始めるとそれぞれの相互接続点で接続を行なう ISP 間でポリシー制御を行なわなければならなくなってくる。現在のような同じ ISP 同士が複数の地点で相互接続するような状況になってくるにつれ、徐々にポリシー制御をいかに円滑に行なうかがインターネット全体の安定性や拡張性に影響を与えるようになる。

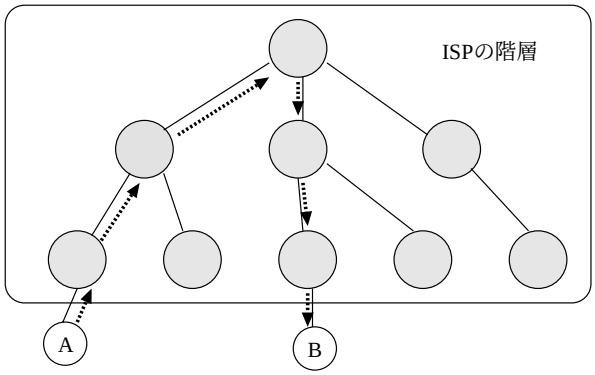

図 2.4: ISP が階層的に接続されている場合

経路制御ポリシーとは、簡単にいえば、どの相手とはどういう経路で通信をしたいかの方針と考えることができる。これはネットワーク全体が、ISP 間の相互接続の地点がそれほど多くなく、おおむね階層的な構造を持っており、ループなども持たなければこれは比較的簡単である。

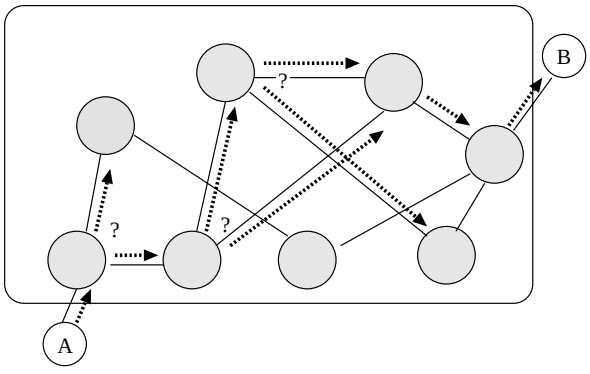

図 2.5: ISP がメッシュ状に接続されている場合

例えば極端な例として図 2.4のように、ISP がきれいに階層構造を持って相互接続しているような場合には、非常に簡単で、ユーザ A からユーザ B に至る経路はただ一つしかない

し、逆にユーザ B からユーザ A への経路も全く同じ経路を逆にたどることになる。ここでは難しいポリシー制御は必要とならない。

しかし、図 2.5のように ISP 同士の相互接続が特に階層を持つことなくまったく任意にメッシュを構成しているような場合には、ポリシー制御も複雑になる。

ユーザ A からユーザ B に至る経路は幾通りもあり、それぞれの ISP からその先の ISP にトラフィックが渡される地点で、複数の選択肢の中から好ましい経路を選んでいかなければならない。さらに、ユーザ B からユーザ A への経路はユーザ A からユーザ B への経路の逆になるとは限らず、多くの場合は、非対象な経路をとることになってしまう。

このように、ISP が自由に相互接続を行なった場合に、それぞれの ISP の持つ経路制御ポリシーを実装するための手段として、ISP 間の経路制御プロトコルである BGP(Border Gateway Protocol) が考案された。

第 3 章

BGP4 について

3.1 BGP4 の位置づけと開発経緯

BGP4(Border Gateway Protocol 4) は、インターネット上での経路制御を行うプロトコルのうち、EGP(Exterior Gateway Protocol) に分類されるものであり、これと対をなす IGP(Interior Gateway Protocol) に分類される RIP や OSPF とは、その目的や経路選択における考え方が大きく異なるものである。

BGP で用いられる重要な概念として、AS(Autonomous System) という概念がある。これは「他の AS からは、内部の経路制御管理が一貫したポリシーを持って行われているように見え、かつ、他の AS に対してどのネットワークがその AS を通して到達可能であるかを矛盾無く示すことのできるもの」と定義されている。これは非常に難しい概念ではあるが、大雑把に言って一つもしくは複数の ISP(Internet Service Provider) が一つの AS に対応すると考えるのが一般的である。

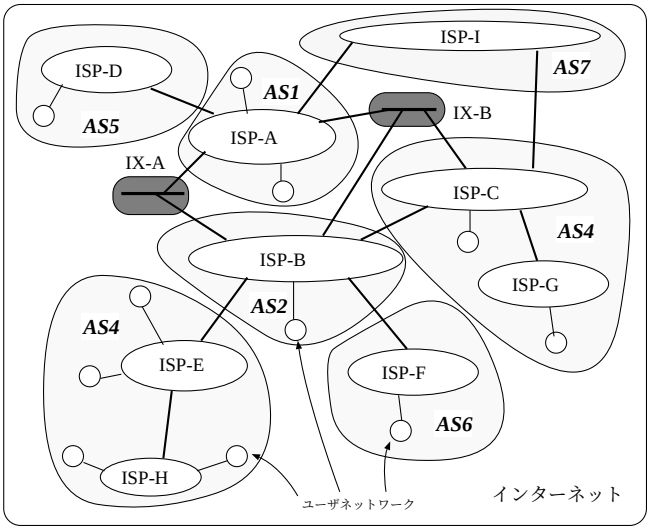

図 3.1: インターネットの全体像

この AS の概念を用いると、現在のインターネットは図 3.1 に示すように、複数の AS が相互に接続して構成されていると考えることができる。それぞれの AS の内部には、それを構成する ISP のネットワークと、その ISP に接続されるユーザのネットワークが含まれる。先に述べた IGP とは、この AS 内部の経路制御を行うためのプロトコルであり、EGP とは AS 間の経路制御を行うためのプロトコルである (図 3.2)

IGP と EGP との大きな違いは、経路選択を行う際に何をよりどころとするかという点である。IGP は一般的に組織の中のネットワークの経路制御に用いられるもので、ネットワークの各部分に至る経路にその部分までの「近さ」や「コスト」を表す重みをつけて、その重みが最小になるような経路を選択するような方式をとる。一方、EGP は ISP 等を単位とする大きなネットワーク間の相互接続を行う際の経路制御に用いられるもので、単に「近さ」や「コスト」を経路選択の指標として用いるのではなく、ある相手との通信のために、どの AS の資源をどう使いたいか、もしくは他の AS 間の通信に、自分の資源をどう使わせたいか、などの「ポリシー」に基づいた経路選択が行われる。この観点から BGP のような EGP による経路制御の事を「ポリシールーティング」と呼ぶこともある。

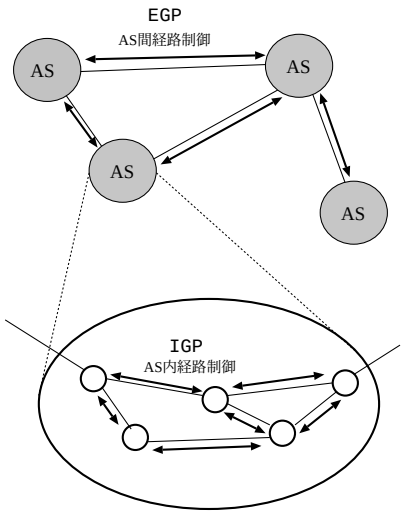

図 3.2: EGP と IGP

さて、図 3.1の IX(Internet eXchange) は、3 つ以上の ISP 同士を接続するための相互接続点である (図の IX-A には簡単のために 2 つしか ISP が接続されていない)。IX 自身は AS ではなく、経路制御的にはポリシーを持たない中立な地点とみなすことができる。通常は IX に各 ISP がルータを持ち込み、そこで FDDI や Ethernet などのデータリンク層のプロトコルを用いて相互の接続が行なわれる。IX は ISP 間の経路制御ポリシーを実際に実現する場所と考えることができ、ISP のルータの代わりに経路選択を行なう Route Server なども ISP のルータと同様に、IX のデータリンクメディア上に接続される。

ポリシーの例としては、例えば図 3.1の AS7 から AS2 にデータを送る時には、AS1 を経

由する経路と AS4 を経由する経路の 2 つの経路が考えられる。この 2 つの経路のうち一方を選ぶ際には、AS7 が AS1 と AS4 のどちらを利用したいかや、AS1 や AS4 が自分の AS を単に通過するトラフィックを AS7 や AS2 に対して許すか許さないかなどのさまざまなポリシーが関係する。

BGP は、このような AS 間のポリシーに基づく経路制御をある程度実現するための仕組みを提供するために開発されたプロトコルである。BGP は、相互接続された AS 同士の境界にあるルータ間で用いられる。この意味で、AS の境界にあり BGP を用いて他の AS のルータと経路情報の交換を行なうルータのことを、Border Gateway(これが BGP の名前の由来となっている) とか BGP スピーカーと呼ぶ。BGP では、BGP スピーカー間でポリシーコントロールを行なうためのヒントとなる「パス属性 (Path Attribute)」を経路情報とともに交換する。経路情報を受け取った側のルータでは、このパス属性を元に経路の選択を行なう。

BGP4 の 4 は、BGP というプロトコル自体のバージョンを表す数字である。即ち現在標準として用いられているのは、BGP のバージョン 4 ということになる。バージョン 3 と 4 の違いは主に CIDR(Classless Inter-Domain Routing) のサポートである。歴史的には、BGP4 が各ベンダのルータで実装されて、インターネット上の AS 間の経路制御に BGP4 が実際に用いられるようになって初めて CIDR が実現された (CIDR については、本稿では割愛する)。その意味ではインターネット全体の発展のために非常に重要な役割を果たしたプロトコルであるということができる。

以下、BGP4 のプロトコルの概要について主に ISP 間のポリシー制御に焦点を当てて解説する。

3.2 BGP4 の概要

BGP4 のプロトコル自体は RFC1771[?] により規定されているが、RFC1771 以降にいくつかのパス属性が追加されるなどして拡張が加えられている ([?][?])。現時点では RFC1771 以降に加えられた新しいパス属性を用いた AS 間経路制御もすでに広く用いられるようになっている。

BGP では、トランスポート層として TCP を利用しており、ポート番号 179 番を用いている。BGP により経路情報を交換するルータ同士は、BGP のセッションを開始するためにまず TCP のコネクションを確立する。セッションが確立されると次にそれぞれが持っている BGP の経路表の情報全てを交換する。その後は、例えば RIP のように定期的に全ての情報を交換する事はせず、かわりに、それぞれの経路表が変化するたびに、変化した分だけの情報の交換のみを行う。BGP プロトコルを用いて経路情報の交換を行う相手を peer と呼ぶ。

BGP では、AS を 2 オクテットの符号なし整数値で番号を付けて区別する。この値を AS 番号と呼ぶ。

3.2.1　IBGP と EBGP

図 3.3に、複数の AS の間でどのように BGP のセッションが張られるかを示す。

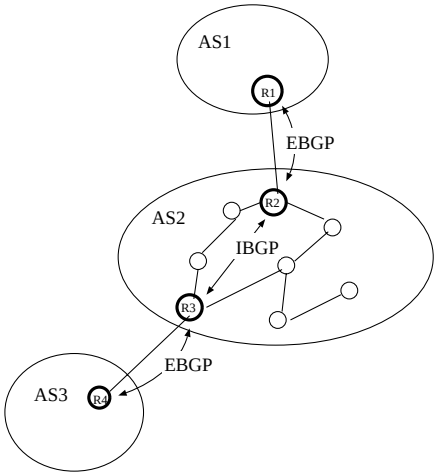

図 3.3: IBGP と EBGP

図では R1 と R2, R2 と R3, R3 と R4 の間にそれぞれ BGP のセッションが張られている。この時、R1 と R2 の間や、R3 と R4 の間のように、異なる AS の BGP スピーカー間の BGP セッションの事を EBGP(External BGP) と呼び、R2 と R3 の間のように同じ AS 内の BGP スピーカー間の BGP セッションを IBGP(Internal BGP) と呼ぶ。

R2 と R3 の間の IBGP は、R2、R3 がそれぞれ外部から受け取った BGP の経路情報を互いに交換し、AS2 が AS 全体として一貫した経路選択を行うために必要となる。この例では AS2 の中に 2 つの BGP スピーカーが存在するが、一般的には、一つの AS の中にある BGP スピーカー全ての間で完全グラフを構成するように IBGP のセッションを張る必要がある。

また、外部から BGP で受け取った経路情報は、それぞれの AS 内部で IGP を用いて BGP スピーカー以外の全てのルータに対して伝搬させる必要がある。

さらに AS2 が AS1 と AS3 の間のトラッフィクを中継するような場合には、AS1 の R1 から AS2 の R2 に伝えられた経路情報は、さらに AS2 の R3 から AS3 の R4 に BGP で伝えられる必要がある。この時、BGP の経路情報は R2 から R3 に IBGP を用いて伝えられるが、R3 は AS2 内部の IGP 経由で同じ経路情報を受け取るまでは、R4 に EBGP でその経路情報を流すことはできない。これはつまり IGP でこの経路情報が R3 に伝わって来るまでは、R2 と R3 の間にある BGP スピーカー以外のルータがこの経路情報を持っていないためで、AS2 内部で R3 から R2 に至る経路が確立されていないことになるからである。これを IGP と EGP の同期と呼ぶ。

3.2.2 RIB(Routing Information Base) と経路選択

各 BGP スピーカーは、現在コネクションを張っている全ての peer に対して自分が送った経路情報を保持しておく必要がある。これを行うために、BGP スピーカーは 3 種類の経路情報データベース (RIB: Routing Information Base) を用いる。また、自分自身の持つポリシーに関する情報は、ポリシーデータベース (PIB: Policy Information Base) に格納されている。経路選択は、複数の RIB と PIB の内容を用いて、決定プロセス (Decision Process) を通じて行われる。

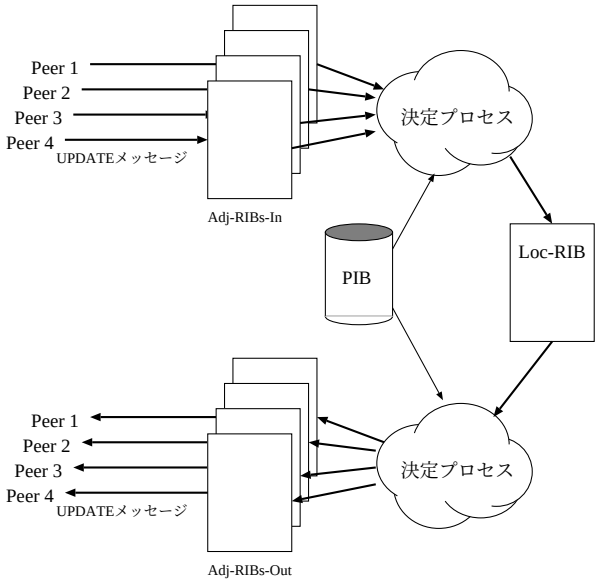

図 3.4: BGP での経路選択の流れ

以下は、3 種類の RIB の概要である。

Adj-RIBs-In 各 peer から UPDATE メッセージで受け取った経路情報をそのまま保持する。 この Adj-RIBs-In の内容が、決定プロセス (Decision Process) の入力となる。

Loc-RIB 決定プロセスにおいて、Adj-RIBs-In の内容に、ポリシー情報データベース (PIB: Policy Infomation Base) 中のポリシーを適用した結果選択された経路情報が保持される。

Adj-RIBs-Out 各 peer に対して送出するものとして選択された経路情報が保持される。この内容が、各 peer に対して UPDATE メッセージにより伝搬される。

Loc-RIB の内容は、さらに IGP 等の他のルーティングプロトコルで得られた経路情報との間で取捨選択され、最終的にルータの経路表が作られる。図 3.4に、これらの経路選択の流れの概略を示す。

3.2.3　パス属性

パス属性は BGP で伝搬される経路情報に付加される属性で、ポリシー制御のためのヒントとなる情報を含むものである。パス属性は、以下の 4 つのカテゴリに分類される。

- 標準的 (Well-known) 必須 (mandatory) パス属性
- 標準的任意 (discretionary) パス属性
- 選択的 (optional) 通過型 (transitive) パス属性
- 選択的非通過型パス属性

標準的 (well-known) パス属性とは、全ての BGP の実装でサポートされていなければならないものである。あるものは、さらに必須パス属性として、全ての UPDATE メッセージに含まれていなければならない。それ以外のものは任意なので、必要に応じて UPDATE メッセージに含まれていれば良い。

全ての標準的パス属性は、必要な変更を加えた後に他の BGP peer に対して伝達されねばらない。

一方、選択的 (optional) パス属性は全ての BGP の実装でサポートされている必要は無い。サポートしていない選択的パス属性を受け取った時には、それが通過型 (transitive) であった場合には、そのまま次の BGP peer に伝達しなければならない。その際に、UPDATE メッセージ中のパス属性のフィールドの中の、属性フラグの不完全ビットを 1 にしなければならない。非通過型であった場合には、単に無視し、次の BGP peer には伝達しない。

以下に、RFC1771 で定義されているパス属性の概略を示す。

ORIGIN 経路情報をどこから取ったかを示す属性。経路情報を BGP で最初に伝搬する BGP スピーカーがこの属性をセットしなければならない。可能な値としては、IGP(0), EGP(1),INCOMPLETE(2) の 3 種類ある。標準的必須属性。

AS_PATH その経路が経由して来た AS の一覧を示す属性。形態として AS_SET と AS_SEQUENCE があり、前者は単に通過して来た AS の集合であるのに対して、後者は AS の順序列を示す。標準的必須属性。

NEXT_HOP その経路に向かう経路上の次のボーダルータの IP アドレス。標準的必須属性

MULTI_EXIT_DISC(MED) 同一隣接 AS との間に複数のリンクがある場合、それらの間の経路選択に用いられる。値が小さい方が優先される。選択的非通過型属性。

LOCAL_PREF 同一 AS 内の他の BGP スピーカにその経路の好ましさを伝える。値が大きい方が優先される。 標準的必須属性。ただし EBGP では使えない。

ATOMIC_AGGREGATE BGP スピーカが受け取った中で、一方がもう一方を含むような複数の経路があった時により集成された経路を選んだ場合、ATOMIC_AGGREGATE パス属性を付加する。これにより、この属性が付いた経路を受け取った BGP スピーカに、集成されたこの経路をより細かい経路にばらすことができないという事を示す。標準的任意属性。

AGGREGATOR 複数の経路を集成して一つにまとめた場合、それを行った BGP スピーカの属する AS 番号とその BGP スピーカの IP アドレスから構成される属性。選択的通過型属性。

これらのパス属性は、決定プロセスで経路の選択をする際に用いられる。

3.2.4 決定プロセス

決定プロセス (Decision Process) では、Adj-RIB-In に格納された各 peer から伝搬された経路情報と、PIB を元にして、自分がどの経路を選択し、さらに他の peer にどの経路を伝搬するかを決定する。

決定プロセスは、以下の 3 つのフェーズに分かれる。

フェーズ 1 AS 外部の peer から受け取った各経路の優先度を計算し、各々異なるネットワークへの経路のうち最も優先度の高い経路の情報を AS 内部の他の BGP スピーカに伝達する。

フェーズ 2 フェーズ 1 終了後、各々異なるネットワークへの経路のうち最も優先度の高いものを選び、選んだ経路で Loc-RIB を更新する。

フェーズ 3 Loc-RIB が更新された後に、PIB に格納されているポリシーに基づいて、各外部の peer に対して経路情報を配布する。この時、経路情報の集成や、情報の削減を行うこともできる。

フェーズ 1 の処理は、外部の peer が新しい経路を伝達したり、すでに持っている経路を変更したり、経路を削除したりする UPDATE メッセージを送って来た時にはただちに実行される。新しい経路や既に持っている経路に変更がかかった時には、それらの経路の優先度の計算を行う。AS 内部の peer からそれらの経路が伝搬されて来た時には、既に設定されている local_pref パス属性の値を優先度として採用する。外部から学んだ経路に関しては、あらかじめ PIB に設定されているポリシーに応じて、優先度の計算を行い、その値を内部の peer に経路を伝搬する際の local_pref パス属性の値として利用する。これらの PIB をベースにした優先度の計算は、各ルータの実装依存である。

フェーズ 2 の処理は、フェーズ 1 の処理が終了した後に行われる。フェーズ 2 の処理では、内部と外部の全ての peer に対する Adj-RIB-In 内に存在する全ての経路情報が対象と

なる。この時、経路の NEXT_HOP パス属性として設定されているアドレスに対する経路が Loc-RIB 上に存在しないようなものは、考慮の対象から外される。その上で、全ての Adj-RIB-In 内に存在する全ての同じネットワークに対する経路に関して、最も高い優先度を持ったもの、もしくは、そのネットワークに対する唯一のもの、もしくは、以下に述べるタイブレイクルールを適用した結果残ったもののいづれかを選択する。選択された経路は Loc-RIB に格納される。

以下はフェーズ 2 でのタイブレイクルールである。これは、同じネットワークに対して優先度の同じ複数の経路が存在した時に用いられるものである。ルールを適用した結果一つの経路のみが残った時点で、タイブレイクの処理は終る。

- 同じ AS から受け取った経路を比較する場合には、MULTI_EXIT_DISC パス属性を比較して、より値の大きな経路を取り除く。MULTI_EXIT_DISC パス属性を持っていない経路に関しては、最も大きな値を持っているとみなす。
- 経路の NEXT_HOP パス属性で示されるアドレスに対する、IGP 上でのコストがより高いものを取り除く
- 外部の peer から受け取った経路があれば、その他の内部の peer から受け取った経路を取り除く
- その経路を伝えて来た peer のうち、最も BGP Identifier の値が低いものを選ぶ

フェーズ 3 の処理は、フェーズ 2 が終了した時、または、次のいずれかが起こった時に行われる。

- Loc-RIB 中の、自分の AS 内部のネットワークに対する経路が変更された時
- BGP 以外の方法で学んだ内部のネットワークに対する経路が変更された時
- 新しい BGP スピーカが登場し、その BGP スピーカとのコネクションが確立された時

Loc-RIB 中の全ての経路は Adj-RIB-Out に移される。この時必要に応じて経路の集成や、情報の削除が行われる。Adj-RIB-Out の更新が終り、実際にパケットフォワーディングに用いられている経路表の更新が終了した後に、peer に対する UPDATE メッセージの送出が行われる。

3.3　RFC1771 以降の拡張

BGP4 は現在インターネット上で AS 間のポリシールーティングを実現するために用いられている唯一のプロトコルと言う事ができる。EGP(一般名詞としての EGP ではなく、これそのものを名前に持つプロトコルが過去に存在した。) や古いバージョンの BGP がま

だ使われているところもあるかも知れないが、実質上 BGP4 でなければ現在のように複雑に AS が相互接続され、かつ、CIDR 化が進んだ環境での経路制御は不可能である。

しかし、BGP4 を用いても実現できないポリシーはまだたくさん存在する。これは一つには、新しいパス属性を導入する事により解決されて行くだろうが、インターネットのデータグラム転送方式そのもの (Hop-by-Hop) に制約を受けてしまうものもある。

RFC1771 以降に導入された新しいパス属性としては、一つの AS の内部を外部からは見えないようにしながら、さらに細かな部分に分割する事を可能とする AS 同盟 (Confederation)([?]) や、経路情報に特定の「色」(routing color という言い方がされる事がある) を付けて、離れた AS に伝搬されて行った時の経路選択方法をコントロールするなどを可能とする、コミュニティーパス属性 ([?])) などがある。

3.3.1 AS 同盟

AS 同盟は、大きな AS の内部でのポリシー制御を簡便化するために、一つの AS の内部を更に細かな AS に分割し、外部からは一つの AS に見えるけれども内部はいくつもの小さな AS の集まりであるような、階層的な AS の構成を可能にするパス属性である。

AS 同盟は、AS_PATH パス属性の新しいタイプとして定義されている。RFC1771 では、AS_SET と AS_SEQUENCE のみが AS_PATH パス属性のタイプとして定義されていたが、RFC1965 では新たに AS_CONFED_SET と AS_CONFED_SEQUENCE の 2 つのタイプが定義されている。

AS 同盟を用いる場合は、外部から見た時の AS を同盟 (Confederation)ID と呼びその同盟を構成する内部の小さな AS のことをメンバー AS と呼ぶ。同盟の内部で経路情報が伝搬される際には、どのメンバー AS を経由したかが、AS_CONFED_SET や AS_CONFED_SEQUENCE タイプの AS_PATH パス属性に付け加えられる。そのような経路情報が、同盟の外に出る時には、同盟の内部で付け加えられた AS_CONFED_SET や AS_CONFED_SEQUENCE タイプの AS_PATH パス属性は削除され、代わりに同盟 ID を AS_SET もしくは AS_SEQUENCE タイプの AS_PATH パス属性に付け加える。これにより同盟の内部では各経路情報がどのメンバー AS を通過してきたかを知ることができるが、外部からは、あくまでも同盟 ID で代表される 1 つの AS を通過しただけのように見える。

このように AS 同盟は、同盟の内部での細かな経路制御のために用いられる。

3.3.2 コミュニティー

コミュニティーパス属性は、ある性質を共有する経路情報のグループを表すパス属性で、32 ビットの値を用いている。いくつかのコミュニティーの値は良く知られたコミュニティー (Well Known Community) を表すものとして予約されている。

NO_EXPORT 0xFFFFFF01 の値で表されるコミュニティーで、このコミュニティーに属する経路情報は、AS 同盟の外部には伝搬されない (この場合 AS 同盟を用いてい

ない AS も、メンバー AS が 1 つだけの AS 同盟とみなされる)。

NO_ADVERTISE 0xFFFFFF02 の値で表されるコミュニティーで、このコミュニティーに属する経路情報は、他の BGP peer に対して伝搬されない。

NO_EXPORT_SUBCONFED 0xFFFFFF03 の値で表されるコミュニティーで、このコミュニティーに属する経路情報は、他の AS(AS 同盟の内部では他のメンバー AS) に対して伝搬されない。

上記以外のコミュニティー値のうち、0x0000000 から 0x0000FFFF までと、0xFFFF0000 から 0xFFFFFFFF までは予約されているが、それ以外の値は各 AS や AS 同士のとり決めに応じて自由に用いることができる。コミュニティー値の上位 16 ビットをそのコミュニティーを定義した AS の AS 番号を用い、下位 16 ビットをその AS でのコミュニティー識別番号として用いる方法が一般的である。

3.3.3　今後の拡張

これらのパス属性はまだ全てのルータで実装されているわけではないが、現実に ISP 間の経路制御に関する複雑なポリシーを実現するためには便利なものであると考えられている。だが、これらのパス属性を有効に使う方法や、コミュニティーパス属性のように ISP 相互間でこれらのパス属性をどう取り扱うべきかについての取り決めがきちんとなされる必要があるものもある。

また、IPv6 への対応に付いてもまだこれからの段階で、AS 間の経路制御が OSI の IDRP(Inter Domain Routing Protocol) が採用される事になるのか、または、現在の BGP4 をマルチプロトコル対応に拡張した BGP4+が採用される事になるのかはこれからの IETF などでの議論と、実際の運用での検証の結果を待たなければいけないだろう。

第 4 章

今日の ISP 間経路制御の実際

2章で述べたように、今日の ISP 間の相互接続は、以前のような階層的な構造からメッシュ状の相互接続へと移行しつつある。それにより ISP は複数の ISP と複数の地点で相互接続するようになってきた。

これにより各 ISP では、複数リンクの使い分けや、他の ISP との間の経路制御に関して複雑なポリシーを持つようになってきている。現在では 3 章で解説した BGP の各パス属性を用いてこれらの複雑なポリシーを実現していかなければならない。

特に、BGP では経路情報とともに伝搬されてくるパス属性に基づいてそれぞれの経路の優先度を決め経路選択を行なう。従ってポリシー制御とは、外部から受けとった経路に付加されたパス属性をどのように解釈し優先度を決めるかの方策と、さらに外部に公告する経路にどのようなパス属性を付加して流すかの方策と言うことができる。

本章では、今日の ISP 間の相互接続におけるポリシー制御の実際についていくつかのケーススタディーを通じて考察する。

4.1 複数経路間の経路選択

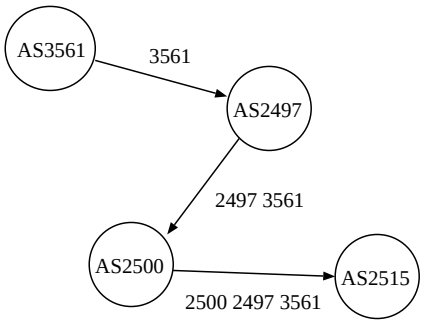

図 4.1: AS_PATH パス属性の例

AS_PATH パス属性では受け取った経路がどういう ISP を経由して届いたかと言う情報を、経由した ISP の持つ AS 番号を逆順に並べることにより表現している。つまり、図 4.1

のように、AS2515 が AS3561->AS2497-> AS2500->AS2515 という経路でとどけられた経路情報を BGP4 で受け取った時には、その経路情報には、2500 2497 3561 という AS_PATH パス属性が付加されている。

RFC1771 では、特にこの AS_PATH パス属性に関して優先度をどのように決めるべきかの指針は何も示されてはいないが、AS_PATH パス属性として持っている AS_PATH の長さ (AS ホップカウントということもある) が短い方が優先されるような実装が一般的である。しかし、複数の経路で同じ経路情報を受け取った場合に、特定の AS_PATH を持つ経路を、その長さに関係なく優先して採用することもできる。

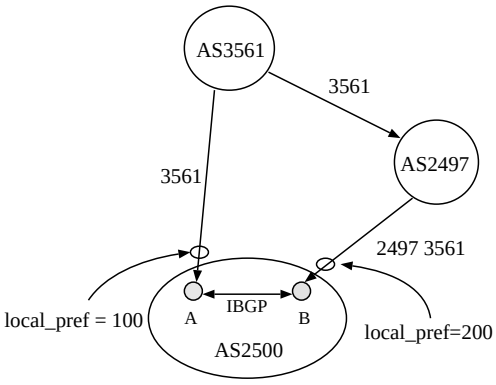

図 4.2: AS_PATH による経路選択

このような場合には、どの経路 (AS_PATH) で受けとる経路情報を優先したいかのポリシーを決め、それに応じて各 BGP スピーカで peer AS から受けとる経路情報に対して LOCAL_PREF パス属性の値を設定する。

例えば図 4.2のような場合、AS2500 では、AS3561 と AS2497 にマルチホームしているとする。ここで AS2500 には、AS3561 から最初にアナウンスされた経路が 2 通りの経路で届く。AS3561 に接続されているリンクを持つルータ A では、”3561” という AS_PATH パス属性を持つ経路情報を受け取る。また、AS2497 に接続されているリンクを持つルータ B では、”2497 3561” という AS_PATH パス属性を持つ経路情報を受け取る。ここで AS2497 経由で受けとる経路を優先したい場合には、ルータ A で”3561” という AS_PATH パス属性を持つ経路情報を受け取る時に LOCAL_PREF パス属性の値を 100 とし、ルータ B で”2497 3561” という AS_PATH パス属性を持つ経路を受けとる時に LOCAL_PREF パス属性の値を 200 とするよう設定する。LOCAL_PREF パス属性は、値の大きい方がより優先されるので、AS_PATH 長がより長い経路情報を優先して選択することができる。またこれらの値は特定の経路情報に対して相対的な優先度の差を与えるもので、利用可能な範囲で自由に設定して良いものである。また、ルータ A とルータ B は、その間の IBGP のセッションを経由してお互いにどういう経路をどういうパス属性で受け取っているかを知ることができるため、双方で矛盾することなく統一した経路選択ができる。

LOCAL_PREF パス属性はこれ以外にも、一般的に同じ経路情報を異なる経路で、かつ、複数の BGP スピーカで受け取った場合に、互いに設定した優先度を相手に伝える目的で用いることができる。これは、経路情報を受け取る側のポリシー、言い替えれば、その経路情報で示される特定の相手に対してデータを送る側のポリシーを実現するために用いられる場合が多いが、後に述べるように、コミュニティーパス属性と組み合わせて用いることにより、経路情報を発信する側のポリシーの実現にも用いることができる。要は特定の AS での LOCAL_PREF パス属性の設定を誰の意志で行なうかがポイントとなる。

さて、上記の例で見たように、ISP 間での経路制御において、複雑なポリシーの実現が必要になるのは、その ISP が他の ISP と複数の接続を持った時点だと考えて良い。

以下では、ISP が複数の接続を持った場合のポリシー制御について考察を進める。

4.2 非対称ルーティング

一般的に複数の接続を持ったからといって、インターネットとの接続性が向上するとは限らない。例えば図 4.3のように、ISP A から出ていくトラフィックは適度にリンク 1, リンク 2 の間で分散されて出て行くが、帰りのトラフィックが、リンク 1 に集中してしまい、結局そこがボトルネックとなり、全体としてのパフォーマンスが得られなくなるというようなことが発生する。しかしこれでは複数の接続を持った意味が無くなってしまう。

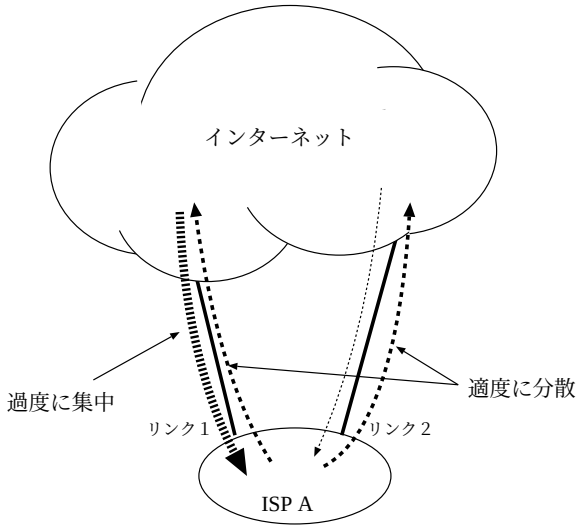

図 4.3: 非対称ルーティング

これは、インターネットの経路制御が片道ごとに行なわれることに起因しており、特定の相手との通信において経由する経路がかならずしも行きと返りで同じになるとは限らないのである。従って複数接続を持った場合には、複数のリンクの使いわけをうまく実現す

るためにさまざまな調整を行なわなければならないのが一般的である。しかし、行きのトラフィックはどのように外部から経路情報を受けとり、各目的ネットワークに対してどちらのリンク経由の経路を優先するかをコントロールし、ある程度効率の良いリンクの使い分けは可能であるが、帰りのトラフィックに関しては、なかなかコントロールする事が難しい。

しかし、前述のように今や ISP が複数リンクを持つことが日常茶飯事のように行なわれていることなので、これらの使いわけに関するポリシーを実現するためのなんらかの手段を持たなくてはならないことになる。

4.3　同一 ISP 間の複数接続

IX があちこちに設置され、それらに多くの ISP が接続を持つようになると、同じ ISP 同士が複数の IX で相互に接続される場合が増えてくる。

このような場合について以下の例で考察してみる。

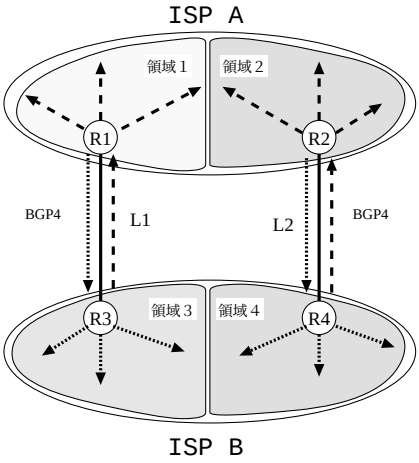

図 4.4: 同一 ISP 間の複数接続

図 4.4で、R1、R2 は ISP A 内部のルータ、 R3、R4 は ISP B 内部のルータ、L1 は R1 と R3 を結ぶリンク、L2 は R2 と R4 を結ぶリンクを表す。

R1 が R3 から受け取った経路や R2 が R4 から受け取った経路はそれぞれ ISP A の内部の他のルータに伝搬される。また、R3 が R1 から受け取った経路や R4 が R2 から受け取った経路はそれぞれ ISP B の内部の他のルータに伝搬される。

それぞれの ISP 内部から相手の ISP に向かうトラフィックがどちらのリンクを通るかは、ISP 内部でのどのようなルーティングプロトコルをどのような運用方針で用いているかに依存するが、一般的にはそれぞれの ISP 内部からリンク 1、2 に接続されているルータに到

達するコストが小さい方が選択される (ただし、BGP で学んだ経路を External Link として、OSPF の Type 2 メトリックを用いて内部に流している場合はこの限りではない)。図 4.4では、ISP A の内部で R1 に至る経路と R2 に至る経路のうち、R1 に至る経路の方がコストが低くなるような領域を領域 1、R2 に至る経路の方がコストが低くなるような領域を領域 2 とした。また ISP B の内部で R3 に至る経路と R4 に至る経路のうち、R3 に至る経路の方がコストが低くなるような領域を領域 3、R4 に至る経路の方がコストが低くなるような領域を領域 4 とした。

4.3.1 特に何も設定しない場合

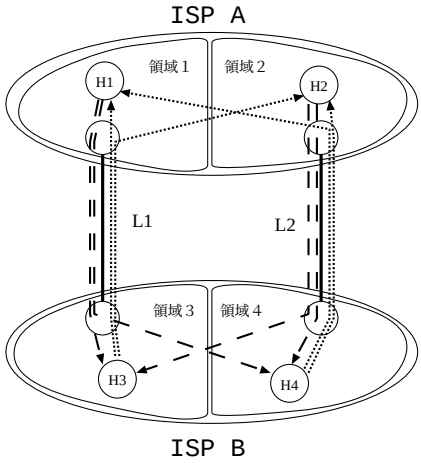

図 4.5: ISP A と ISP B 間のトラフィックの経路

この場合図 4.5のように、領域 1 から ISP B に向かうトラフィックはリンク 1 を、領域 2 から ISP B に向かうトラフィックはリンク 2 をそれぞれ通る。また、領域 3 から ISP A に向かうトラフィックはリンク 1 を、領域 4 から ISP A に向かうトラフィックはリンク 2 を通る。図では領域 1、領域 2、領域 3、領域 4 内のホストをそれぞれ H1、H2、H3、H4 で示している。

4.3.2 片方をバックアップに用いる場合

ISP A と ISP B は通常は L1 を利用し L2 はバックアップに利用したいような場合がある。

このように同じ ISP 同士が複数の接続を持つ場合のポリシー制御に用いられるのが MED パス属性である。MED は複数リンクに設定するメトリック的な意味を持つもので、LOCAL_PREF パス属性とは逆に値の小さい方が優先される。したがって、例えば図 4.6のように R1 と R3 で経路情報を交換する際に MED の値を 100 に設定し、R2 と R4 で経路情報を交換する際には、MED の値を 200 に設定すれば良い。これも LOCAL_PREF パス属

性と同様、相対的な大小関係を表すもので、絶対的な大きさには特に意味はないことに注意して欲しい。

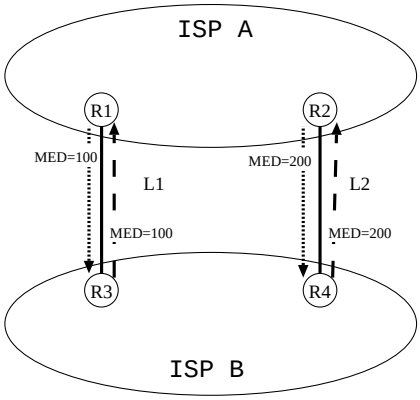

図 4.6: 片方をバックアップに用いる場合

この設定はリンク 2 が IX 経由での接続であるような場合に良く用いられる。IX はそこにつながっている全ての ISP の共有物的な意味あいが強いのでできるだけ必要の無いトラフィックは乗せないほうが良い。したがって直接のリンクを持っている相手に関しては直接のリンクの方を通常は使い、IX 経由の接続はバックアップ的に用いるのが一般的なようだ。

次にもう少し複雑なポリシーに関して考えてみる。

4.3.3　経路の対称化

図 4.5に示した例では、領域 1 と領域 3 の間のトラフィックはどちらの方向も L1 を、また、領域 2 と領域 4 の間のトラフィックはどちらの方向も L2 を経由していた。しかし、領域 1 と領域 4 の間のトラフィックや領域 2 と領域 3 の間のトラフィックが、方向によって L1 を通ったり L2 を通ったりしており非対称になっていた。これを対称にしたいような場合を考える。

現実にこのような接続形態でかつトラフィックをどうしても対称にしたいような一つの典型的な例としては、領域 1 と領域 3 が東京にあり、領域 2 と領域 4 が大阪にある場合などがある。

ここでは、ISP A の内部で領域 1 と領域 2 が 45Mbps の回線 (L3) で結ばれており、ISP B の内部では領域 3 と領域 4 が 1.5Mbps の回線 (L4) で結ばれているような場合がある。この場合 L1 も L2 も 1.5Mbps 程度もしくはそれ以上の容量を持っているものとする。

この場合、図 4.5の例では、領域 1 と領域 4 の間、および、領域 2 と領域 3 の間のトラフィックがどちらも行きか帰りかに ISP B 内の 1.5Mbps の回線を経由してしまい、ここがボトルネックになってしまう可能性がある。それでは双方にとってメリットが無くなるの

で、ISP A と ISP B の間で、相互接続にかかわるトラフィックは、できるだけ ISP A の持つ 45Mbps のリンクを使うようにする事で合意ができたとする。すると、各領域間のトラフィックの経路としては図 4.7に示したようにしなければならない。

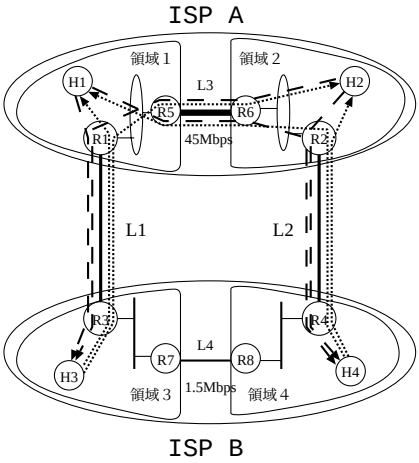

図 4.7: 経路の対称化

ここで、ひとつ前提条件をつけないとこの例は実現不可能である。すなわち、ISP B 側で、領域 3 に属するネットワークアドレスと、領域 4 に属するネットワークアドレスを明確に区別できること、が前提条件となる。これは、ISP B に接続され、ISP B 経由で通信を行うユーザのネットワークも含めてと言う事である。

仮にそれが明確にできたとして、領域 3 に属するネットワークアドレスの集合を NETS3、領域 4 に属するネットワークアドレスの集合を NETS4 と名付ける事にする。

さて、図 4.5と図 4.7を見比べると、ISP B から ISP A に向かうトラフィックに関しては同じである。したがって、ISP A から ISP B への経路情報の流し方は図 4.5の場合と同じで良い。

図 4.5と異なるのは、ISP A から ISP B へのトラフィックの流れであるが、ここで図 4.7を見ると、ISP A から ISP B の領域 3 に向かうトラフィックは必ず L1 を通り、ISP A から ISP B の領域 4 に向かうトラフィックは必ず L2 を通っている事が分かる。言い替えれば、NETS3 に向かうトラフィックに関しては L1 を優先させ、NETS4 に向かうトラフィックに関しては L2 を優先させる、という事になる。

であれば、図 4.8に示したように、ISP A では、NETS3 に関しては L1 経由で受け取る経路情報の MED の値の方が L2 経由で受け取る経路情報の MED の値よりも低くなるようにし、また NETS4 に関しては L2 経由で受け取る経路情報の MED の値の方が L1 経由で受け取る経路情報の MED の値よりも低くなるようにすれば良い。

これはいわば先にあげた 2 つの例の組み合わせといえるなものである。この場合あくまで前提として NETS3 と NETS4 を ISP B 側で明確に区別できなければならない。これは ISP

B 全体で一つの CIDR ブロックに属するアドレスを用いているような場合には結構難しいものとなる。さらに経路情報の集約の事を考えると、ISP A から他の ISP に ISP B の経路情報を伝える際には、ISP A が ISP B の持つ CIDR ブロックの経路情報を集約 (aggregate) して流さなければいけなくなってしまう。これは代理集約 (proxy aggregate) と呼ばれている。

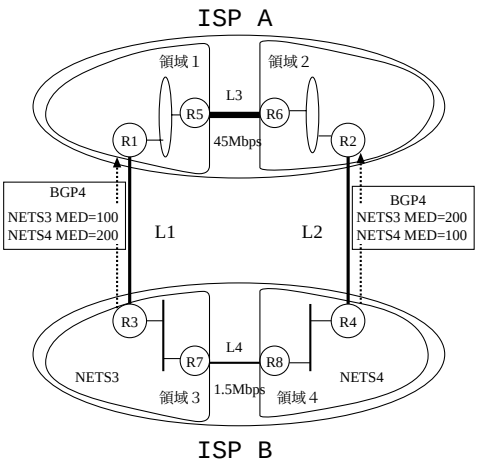

図 4.8: 領域毎に異なる MED の値を設定

この例の場合、各領域をメンバー AS として各 ISP で AS 同盟を用いることもできる。AS 同盟を用いた場合には、NETS3 と NETS4 の区別をより容易に行なうことができる。

4.3.4　他の ISP が絡んだ場合

これまでは、ISP A と ISP B の 2 者間のトラフィックのみに注目していたが、ISP A がさらに他の ISP C につながっていた場合を考察してみる。

図 4.9では、ISP C が ISP A の領域 1 内にあるルータ R9 に接続されている。またここで、新たな仮定として L1 が 1.5Mbps の回線、L2 が 512Kbps の回線であったとする。ISP A と ISP B の間のトラフィックに関するポリシーは先の図 4.7の場合と同じだとする。

こうなっても簡単なのは ISP C のトラフィックに関して、ISP A の領域 1 のトラフィックと同じポリシーで運用する場合である。すなわち、ISP C から ISP B に向かうトラフィックに関して、領域 3 向けには L1 を経由させ、領域 4 向けへは L2 を経由させる。ISP B から ISP C に向けては、領域 3 からは L1 を経由させ、領域 4 からは L2 を経由させるというものである。

これを行うには、単に ISP A は ISP C から来る経路情報を、ISP A の経路情報と同じ扱いで ISP B に流せば良い。

しかし ISP A のポリシーとして、ISP C と ISP B の間のトランジットのトラフィックは、通常のリンク使いわけのポリシーとは異なり、行きも帰りも L1 のみを用いるようにしたい場合も考えられる。

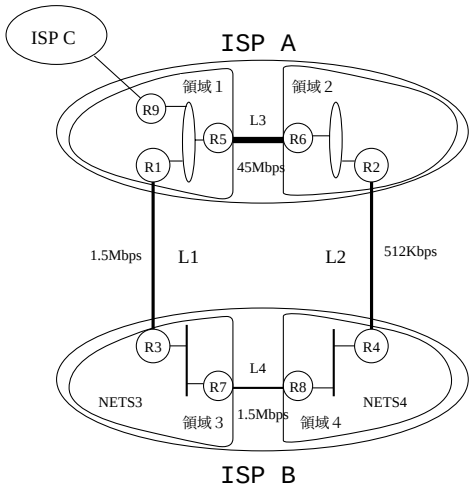

図 4.9: 他の ISP が絡んでいる場合

結論から言うと、図 4.7の状態を維持したままで、この ISP C に関するポリシーを実現するのは、現在の技術を普通に用いただけでは不可能である。ここでは、領域 1 から領域 4 に向かうトラフィックは L2 を経由させ、ISP C から領域 1 を経由させ領域 4 に向かうトラフィックは L1 を経由させないといけない。これはつまりソースアドレスを見て経路を決めないと行けないと言う事になるのだが、これは一般的には実現不可能である。

このように、複数のポリシーを組み合わせた場合には必ずしも全ての要求を満足することができない場合もある。

4.4 複数の ISP と接続する場合

複数の ISP と接続を持つ場合には、その複数の接続により、インターネットへの経路に冗長性を持ちたいというのが主な理由だろう。もちろん、ある程度の負荷分散という目的も考えられるが、1 つの隣接 ISP で何かトラブルがあってその ISP 経由のインターネット接続が失われたような場合でも、自分自身のインターネット接続は他の ISP 経由で維持できるようにしたい、というリスク分散的な意味合いが強いと考えられる。実際、先の 1 つの ISP と複数接続する場合よりも、こちらの場合の方が件数的には圧倒的に多い。

ここでは、複数 ISP と接続した場合の経路制御について考察するために、図 4.10に示すような例を考える。

ISP A は、リンク 1 で ISP B と、リンク 2 で ISP C と接続しており、リンク 1、リンク 2 の太さはそれぞれ、768Kbps, 1.5Mbps とする。

ここで ISP A とその他の ISP との間のトラフィックがどのような経路を経由するのかを考えてゆく。

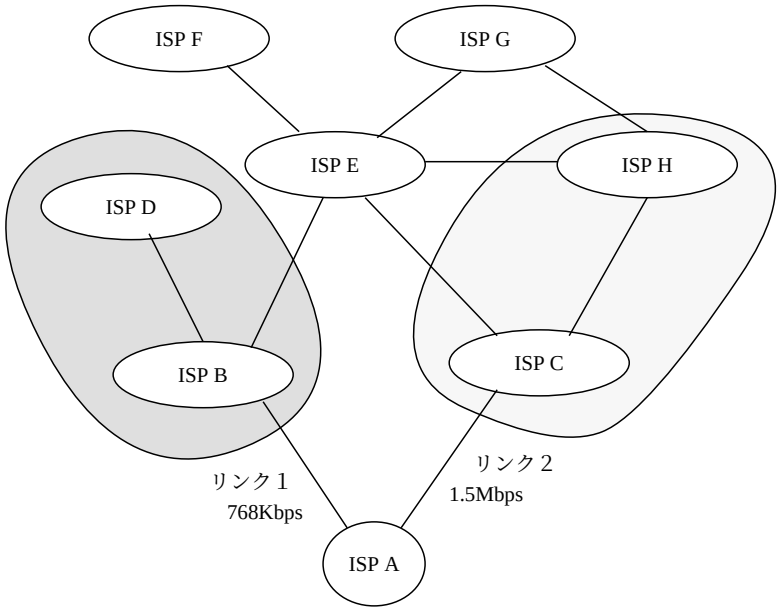

図 4.10: 複数の ISP に接続を持つ場合

AS ホップカウントが最短となる経路が選ばれると考えれば、ISP B や ISP D とのトラフィックはリンク 1 を通り、ISP C や ISP H とのトラフィックはリンク 2 を通る。図でリンク 1 を経由する ISP のグループを斜線で、リンク 2 を経由する ISP のグループを網かけで示すことにする。

では、それ以外の ISP E、ISP F、ISP G とのトラフィックはどのようにコントロールすれば良いのだろうか。

先に述べたように、ISP A からこれらの ISP に対して出て行くトラフィックは ISP A 側で受けとる経路情報の優先度を調整することによりコントロールすることも可能である。ISP A の BGP スピーカで ISP B, ISP C から受けとる経路情報に対して AS_PATH パス属性などに応じて LOCAL_PREF パス属性の値を決めてやるなりすれば、少なくとも相手 ISP 単位でどちらのリンクを使うのかをコントロールすることは容易である。

それに対して他の ISP から ISP A へのトラフィックをコントロールする事は難しい。しかし、前述のように行きと帰りのトラフィックが通る経路が非対称な場合には、片一方のリンクの太さに押さえられて、期待した通りのパフォーマンスが出なくなることがある。極端な場合、何らかの間違いで全ての ISP から ISP A に向かうトラフィックが全てリンク 1 に

集中してしまった場合には、全体として、リンク 1 の容量である 768Kbps 以上のパフォーマンスは出なくなる。

したがってうまくリンクの使い分けを考えたければある程度帰りのトラフィックもリンク 1 とリンク 2 でうまく振り分けてやる必要がある。

さて、では例えば ISP E の場合であるが、ISP A から ISP E までは、どちらのリンク経由でも別の ISP を一つ経由する事になる。つまり、BGP で ISP A の経路情報が ISP E に伝えられた場合、その経路に付加される AS_PATH パス属性の長さはどちらも 2 となる。この時 ISP E がどちらの経路を選択するかは、ISP A であらかじめ知る事が出来ない。ISP A 側でこれをコントロールしようと思った場合には、何らかの形で ISP E と調整する必要が出て来る。ISP G の場合も同様で、ISP G では、2 つの経路で、ISP A の経路情報を AS_PATH パス属性の長さ 3 で受け取る。従ってこれも何らかの形で ISP G と調整しないと、ISP G から ISP A へのトラフィックが ISP A の持つどちらのリンクを用いる事になるかは決められない。

さらにここで問題なのは、これらの経路選択が、トラフィックを交換する先の ISP とだけ調整しても決める事が出来ないと言う点である。一つには何ホップも先の ISP での経路選択に、経路情報を流した側のポリシーをうまく伝える手段が無いことと、さらに、対称となる ISP との間に存在する ISP 全てと調整をした上でないと望んだ通りの経路で相手から自分に向かうトラフィックの経路を指定することができないのである。

また、ISP F の場合は ISP A への経路上で必ず ISP E を通過しなければいけないので、基本的には ISP E が選択したのと同じリンクを用いる事になる。これは現在のインターネット上で用いられているデータ転送の仕組みによりどうしようもできない部分ではある。

4.5 直接接続していない ISP からの帰りのトラフィックの制御

しかし、何ホップも先の ISP から自分に対する経路をコントロールする手段が全く無いわけではない。RFC1998[?] で、コミュニティーパス属性を用いて、これを実現可能にする方法が提案されている。

図 4.11に示すようなネットワーク構成で、AS3 は、AS3561 に対して、AS1 経由の経路よりも、AS2 経由の経路の方を優先して採用して欲しいとする。

この時に、AS3 が AS1, AS2 に経路情報を送り出す際に、それぞれ 3561:100, 3561:90 というコミュニティー値を設定して送り出す。ただし、ここではコミュニティー値の上位 16 ビットと下位 16 ビットとをそれぞれ十進数で表記しそれらをコロンで区切る記法を用いている。3章で解説したように、設定可能な 32 ビットの値の上位 16 ビットはそのコミュニティーを定める AS の AS 番号であり、下位 16 ビットをその AS で決めたコミュニティー番号である。RFC1998 で紹介されている方法では、この AS ごとに決めることのできるコミュニティー番号に、その AS で経路情報を受けとった時に採用して欲しい LOCAL_PREF

パス属性の値を用いる。

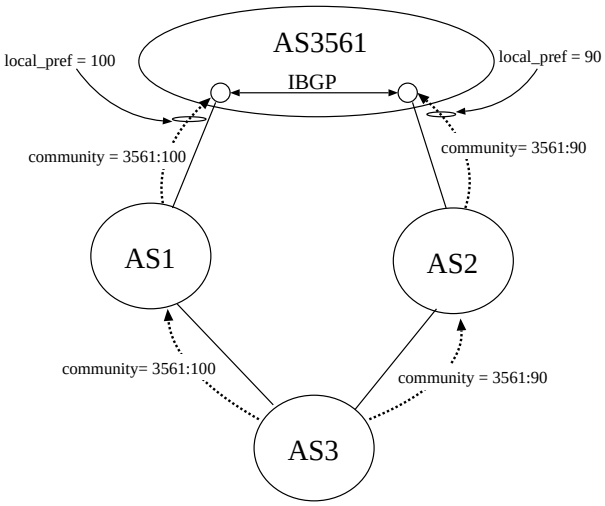

図 4.11: コミュニティーパス属性の応用

AS3561 側の BGP スピーカーで経路情報を受けとった時に、経路情報に付加されたコミュニティーパス属性の上位 16 ビットを調べて、自分の AS 番号が設定されていた場合には、下位 16 ビットの値をその経路の LOCAL_PREF パス属性の値として設定するのである。

これにより、特定の AS に対して自分がアナウンスしている経路情報をどのように扱って欲しいかを経路情報の中に表現する事ができ、よりこまかなポリシー制御が可能となるというものである。

ただ、現状では全ての AS でこの方法を採用しているわけではなく、一部の AS に対してしか利用することができない。

第 5 章

まとめ

本報告では ISP の相互接続の形態の変化の歴史を概観し、それによって生じた問題を明らかにした。また、現在 ISP 間経路制御で標準的に用いられている経路制御プロトコルである BGP4 に関して主にポリシーコントロールに関わる部分に関して概観し、それらを現実の問題にあてはめた場合にどのように利用すれば良いのかのケーススタディーを行なった。現在の技術では実現できない経路制御ポリシーについても考察した。

今後は、現在技術的な観点から実現できない経路制御ポリシーを実現するための技術的考察を進め、また、当然必要となる ISP 間の協調を円滑に行なうための仕組みについてさらに考察してく予定である。