新米エンジニアのための

初歩のインターネット技術

《第12回 ネットワークの構造とプロトコル》

浅羽 登志也
asaba@iij.ad.jp
株式会社インターネットイニシアティブ

技術が歴史を変え、歴史が技術を形作っていくというプロセスは、インターネットの世界でも見ることができます。今回はアメリカのバックボーンネットワークの再編成によって、ルーティングにどのような問題が生じ、それがどのように解決されようとしているのかを見ていくことにします。

はじめに

世の中乱れている。昨年の4月にNSFNETバックボーンがオペレーションを停止して以来、混沌とした状態が続いている気がする。思えば、NSFNETバックボーンの停止が1つの新しい時代の幕開けとなったわけだが、これはそもそも学術研究主体のインターネットから商用サービス中心のインターネットへの変革の過程で進むべき道であった。したがって、現在の混沌とした状況はその過渡期のものであり、この混沌の中から新しい時代の安定した世界を作っていくのはわれわれインターネットエンジニアすべてに共通の課題だと認識される。

などといきなり堅苦しい始め方をしてしまったが、事実かなり状況は変わってきている。今回はこの辺の話から始めるとしよう。

ちょっとだけ歴史のお時間

NSFNET（National Science Foundation Network）バックボーンは、1986年にアメリカの5か所のスーパーコンピュータセンター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りしてきた。さらに商用ISPのサービスが開始されると、そのユーザー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てきた。

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さまざま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るのである。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。たとえばNSI（NASA Science Internet）や、ESnet（Energy Sciences Network）などがある。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すべてが許されていたので、大学がどこかの企業と共同研究をするなどという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。

さて、もちろんその間、商用ネットたちもその状況に甘んじていたわけではない。CIX （Commercial Internet eXchange）を作って商用ISP間の相互接続を行ったり、また現在のNAP（Network Access Point）の原型ともいえるMAE-East（Metropolitan Area Ethernet）を作るなどして、NSFNETバックボーンを経由しない相互接続も徐々に進んできていた。

ところで、MAEがNAPの原型だと書いたが、大きく違うのはNAPはNSF、つまり、政府主導で作られたのに対して、MAEは商用ISPが集まって作られたということだ。この辺が落ち着くまでにもいろいろな経緯があったようだ。

最初にMAE-Eastができたのは1992年の終わりごろで、これはちょどNAPの話が出始めたころである。MAE-Eastは当時からATMを利用したEthernetのブリッジサービスをしていたMFS（Metropolitan Fiber System）Datanet社のサービスを利用して、Washington DCを中心に作られたものである（現在はFDDIスイッチを中央に置き、離れた場所にはATMを利用してFDDIをブリッジして伸ばしていく形が主になっている。ちゃんと動いている技術をうまく組み合わせたサービスということができよう）。

さて、NSFのプランは、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やめて、「その代わり全米数か所にネットワーク同士がデータリンク層（イーサネット、FDDIなど）で相互接続を行うポイントとしてNAPを作るから、それで何とかしてちょうだいな。各NAPには、でっかいISPをつなげさせるから、NAP間の相互接続はそいつらからサービスを買ってちょ」というものである。

NAPの中で最初に動き出したのがSprint社のNew York NAPであり、これは1994年の秋頃である。NY NAPは、MAEと同様にFDDIスイッチを利用している。NY NAPは今でもFDDIスイッチを利用しており、ATMは仕様が安定して満足のいくサービスが提供できるようになるまでは見送る考えである。

Ameritech社のChicago NAP、PacBell社のSan Francisco NAPは、どちらもATMをダイレクトに提供する形のNAPを構築しており、安定性やパフォーマンスの点でNY NAPに遅れをとっていた。MAE-Westは、満足のいくパフォーマンスも出せないくせに、なおもATMに固執するPacBellにISPたちがプッツンして、「俺たちは俺たちでやる！」と叫んで作ってしまったものである。ISPたちがプッツンしてしまったのが1994年12月のIETF（Internet Engineering Task Force）のときであり、実際にサービスを開始したのは、1995年の春頃である。こちらはFIX-Westがある NASA Ames Research Center内にFDDIスイッチを構え、MFSがFDDIブリッジでカリフォルニアベイエリアでMAE-Westへの接続サービスを行っている。この辺の動きの早さはさすがアメリカといったところだろうか。

その後、「NAPも動き出したし、もうそろそろ本来の目的に専念してもええっちゃろう？」といいながら、1995年の4月の終わりにNSFNETバックボーンのサービスが停止したわけである。

1996年3月現在、ふと気がつくとあちこちにぼこぼこ同じようなものができあがっている。筆者の知るところでは、MAE-Chicago、MAE-LA、MAE-Dallas、MAE-NY、MAE-HoustonなどのMAE系や、Tucson NAP、Phoenix REP （Phoenix Regional Exchange Point）など、山のようにいわゆるInternet eXchange（IX）ができつつある。

日本にもIXは存在する。WIDEプロジェクトのNSPIXP（Network Service Provider Internet Xrossing Point）がIXの構築運用実験を1994年から行っている。現在のNSPIXP-1は、イーサネットスイッチを用いてISP間の相互接続を行っているが、日本での商用サービス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てきている。これを受けて、FDDIスイッチを利用したNSPIXP-2の計画も具体化してきている。

いったい何が変わったのか

さて、このNSFNETバックボーンの停止とIXの林立によりいったい何が変わってきたというのだろうか？

つまり、インターネットの構造が大きく変革しつつあるといってもよいのかもしれない。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1のようになる。

ユーザーは、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ー間の通信は、多くの場合NSFNETバックボーンを経由して行われていた。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に接続されていた。たと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で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。

さて、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にこの状況を示す。

商用ISPは独自にバックボーンを構築し、ユーザーを増やしていった。また、NSFNETバックボーンに接続したり、地域ネットと相互接続をしたり、また、他の商用ISPと相互接続したりして、ユーザー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勧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ーに対するサービスを続けながらも、商用のユーザーもつなぎ始め、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。NEARnet（New England Academic and Research Network）や、BARRNet（San Francisco Bay Area Regional Research Network）などがそれである（その後両者はBBNに買収され、BBN Planetとして統合されつつある）。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、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を停止した。このNSFNETバックボーン以降の状況を示したのが図3である。

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1つもしくは複数に接続をし、同じIXに接続しているISPとの間でトラフィックの交換を行う。また、IXに直接接続をしていないISPは、IXに接続を持つ他のISPに接続をして他の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1つではなく複数持ったり、またユーザーも複数のISPに接続する場合がある。

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していない。したがって、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、BGPのpeerを設定し、経路情報の交換を行って初めて相互接続がなされるのである。IX自体はIPの階層からは見えなくなっている。

したがって、NSFNET以降の状況は、ISP同士が思い思いにIX経由や直接リンクを張るなどして相互接続のメッシュを構成しているといってよい。かくして、おなじみのプロバイダー相関図のできあがりである。

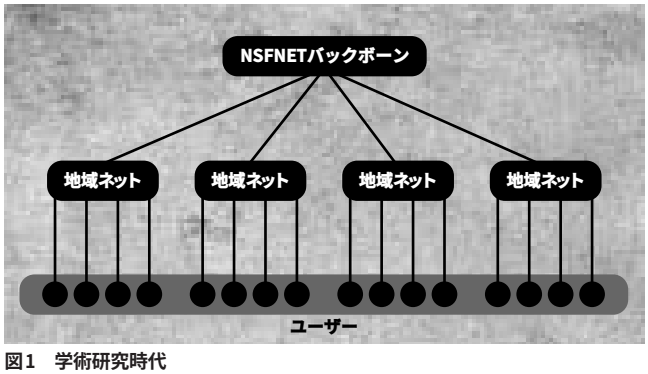

図1　学術研究時代

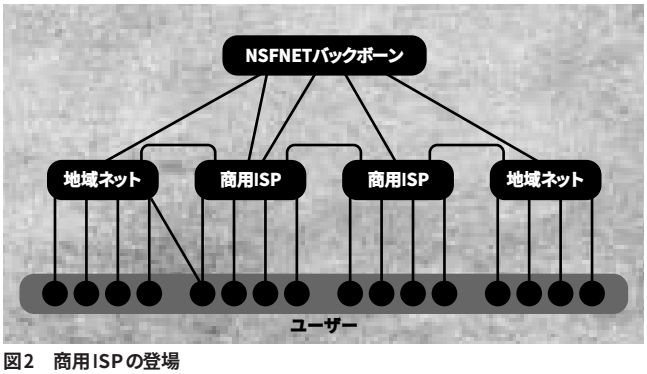

図2　商用ISPの登場

それが何を意味するのか？

さて、では一体それが何を意味しているのだろうか？

インターネットの構造が変わると、自ずとその上でのルーティングも変わってくるはずである。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ルーティングを考えればよく、ポリシーの制御も今ほど強くは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うISP間のポリシー制御を行わなければならなくなってくる。現在のような同じISP同士があっちでもこっちでもつながっていてなどという状況になってくると、だんだんこのポリシー制御に頭を悩ませることになるのである。

ルーティングポリシーとは、平たくいえば、どの相手とはどういう経路で通信をしたいかということである。ネットワーク全体が相互接続がそれほど多くなく、おおむね階層的な構造を持っていてループなどもなければ、これは簡単である。

たとえば極端な例として図4のように、ISPがきれいな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、ユーザーAからユーザーBに至る経路はただ1つしかないし、逆にユーザーBからユーザーAへの経路もまった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不要である。

しかし、図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、話が変わってくる。

ユーザーAからユーザーBに至る経路は幾通りもあり、それぞれのISPからその先に渡されるところで、複数の選択肢の中から経路を選んでいかなければならない。さらに、ユーザーBからユーザーAへの経路はユーザーAからユーザー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のである。

この図5のような状況が、NSFNETバックボーン停止以降の状況といってもよい。みんなで頭を悩ませなければいけない状態なのである。

BGPによるポリシー制御

ISP間の経路制御にはBGP4が用いられているという話は以前書いたと思う。ここでは、ポリシー制御という観点からもうちょっと突っ込んだ話をしよう。図5に示したようなISP間の相互接続がなされているとして、では、BGP4という経路制御プロトコルを用いることによっていったいどんなポリシーが実現できるのだろうか？

BGPでは、PATH ATTRIBUTEを用い

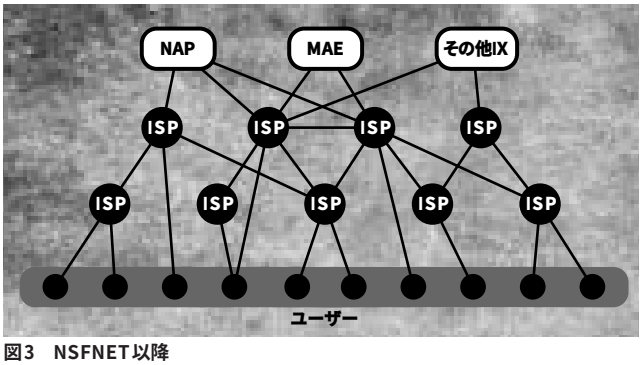

図3　NSFNET以降

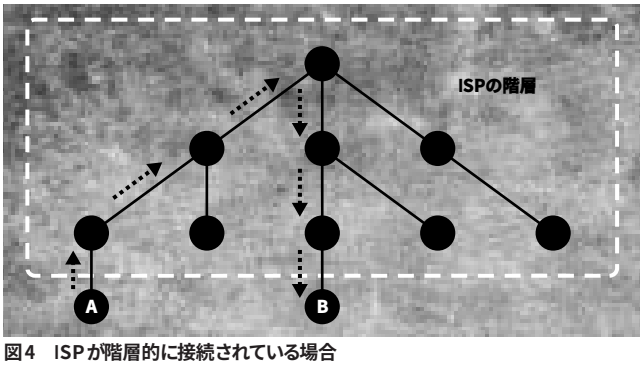

図4　ISPが階層的に接続されている場合

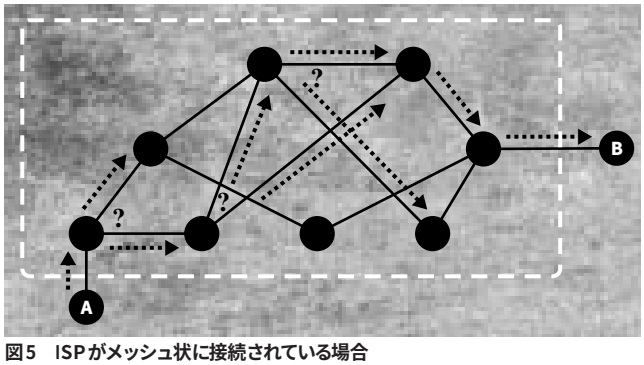

図5　ISPがメッシュ状に接続されている場合

て経路選択を行うと以前書いた。いくつか定義されているPATH ATTRIBUTEのうちポリシー制御に用いられるものは主に3つある。AS PATH、LOCAL_PREF（Local Preference）、MULTI_EXIT_DISC（MED: Multi Exit Discriminator）である。

AS PATHについては以前簡単に解説した。すなわち、受け取った経路がどういうISPを経由して届いたかという情報を、経由したISPの持つAS番号を逆順に並べたものであった。つまり、図6のように、AS2515がAS3561->AS2497->AS2500->AS2515という経路で届けられた経路情報をBGP4で受け取ったときには、その経路情報には、2500 2497 3561というAS PATH ATTRIBUTEがついているのである（本当は、AS PATH ATTRIBUTEは、さらにAS SEQUENCEとAS SETの2種類に分けられる）。しかし、AS PATH自体はポリシーを表しているわけではなく、ポリシーを決定するための材料を提供していると考えることができる。

一般的には、AS PATH ATTRIBUTEとして持っているAS PATHの長さが短いほうが優先される。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。

このような場合には、LOCAL_PREFというPATH ATTRIBUTEを用いることができる。たとえば図7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ーAでは、3561というAS PATH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ーBでは、2497 3561というAS PATHを持つ経路情報を受け取る。ここで、通常の場合には、ルーターAがAS3561から受け取ったAS PATH長の短い経路情報のほうが優先して採用される。ここで、あえてAS PATH長のより長い、ルーターBがAS2497から受け取る経路情報のほうを優先したい場合がある。この場合には、経路情報を受け取るときにLOCAL_PREFとしてルーターAで100、ルーターBで200と設定すればよい。

LOCAL_PREFは、値の大きいほうがより優先される。したがってAS PATH長がより長いにもかかわらず、ルーターBの受け取る経路情報を優先して選択することができるのである。ルーターAとルーターBは、IBGPのセッションを張っていて、お互いにどういう経路をどういうPATH ATTRIBUTEで受け取っているかを知ることができるので、双方で矛盾することなく統一した経路選択ができるのである。

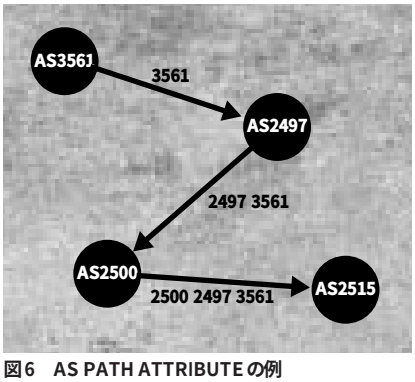

図6　AS PATH ATTRIBUTEの例

LOCAL_PREFは、上の例のAS PATHにより優先度を決めるということ以外にも、一般的に複数の経路で同じ経路情報を受け取った場合の優先度の決定に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、言い換えれば、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、経路情報を発信する側のポリシーの実現にも用いられる。要は特定のASでのLOCAL_PREFの設定を誰の意志で行うかがポイントとなる。

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。このような場合のポリシー制御に用いられるのがMEDである。MEDはLOCAL_PREFとは逆に、値の小さいほうが優先度が高いことを示している。

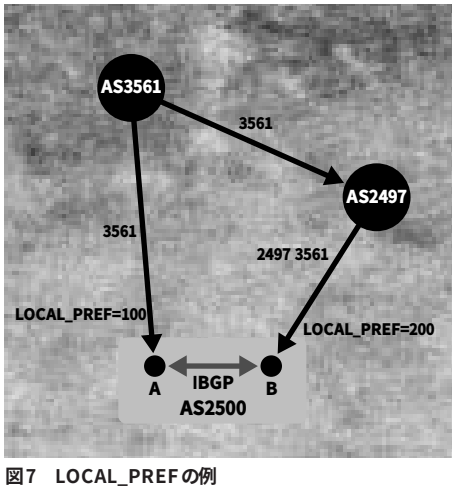

図7　LOCAL_PREFの例

図8のように、ISP AとISP Bが、IX-1とIX-2にそれぞれ接続していて、どちらのIXでもBGPのpeerを設定してトラフィックの交換をするが、IX-2経由の経路を優先的に利用して、IX-1経由をバックバップとして使いたいような場合がある。このような場合には、図に示すように、ISP AがIX-1とIX-2を経由してISP BにBGPで経路情報を送る際に、MEDの値をそれぞれ10, 1のように設定する。また、ISP BがIX-1とIX-2を経由してISP AにBGPで経路情報を送る際にも同様にMEDの値をそれぞれ、10, 1のように設定すればよい。

COMMUNITY

最近、BGP4の新しいPATH ATTRIBUTEとして、COMMUNITYが提案されている。これは、発信する経路情報にポリシーを表す“色”をつけておいて、経路制御に役立てるという目的で提案されたものである。MCIがCOMMUNITYの1つの利用法を提案し、実際に運用を行っている。その方法について簡単に解説してみよう。

図9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経由の経路のほうを優先して採用してほしいとする。ここで、AS3がAS1、AS2に経路情報を送り出すときに、それぞれ3561:100, 3561:90というCOMMUNITYを設定して送り出す。COMMUNITYとして設定可能な値は32ビットの数値である。この上位16ビットをAS番号、下位16ビットをそのASで設定してほしいLOCAL_PREFの値に設定する。AS3561側では、受け取る経路情報のCOMMUNITY ATTRIBUTEの上位16ビットを調べて、自分のAS番号が設定されている場合には、下位16ビットの値をその経路のLOCAL_PREFの値として設定する。

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ほしいかを経路情報の中に表現することができ、より細かなポリシー制御が可能となるというものである。

おわりに

BGP4では、経路制御のいろいろなポリシーコントロールを可能にするための道具が実装されているわけだが、まだまだ実現できないポリシーは山ほどある。これらは、さらに実際のインターネットを動かしながらエンジニアたちが運用経験を積み重ねてゆき、より便利で将来性のある技術を作り上げていかねばならない。21世紀の情報ハイウェイを実現するためには、今頑張って何が必要なのかを見極め、それを実現する技術を作り上げていかなければならないのである。

最後に、今回紹介したBGPの詳細について、参考文献を挙げておく。興味をお持ちの向きは是非ご一読あれ。RFCはftp://ftp.iij.ad.jp/pub/RFC/に、internet-draftはftp://ftp.iij.ad.jp/pub/internet-drafts/からそれぞれ入手可能である。

[1] Rekhter, T., and Li, T.,"A Border Gateway Protocol 4 (BGP-4)", RFC1771, March 1995
[2] Chandra, R., Traina, P., and Li, T., "BGP Communities Attribute", INTERNET-DRAFT, <draft-chandra-bgp-communities-00.txt>, April 1995.
[3] Chen, E., and Bates, T., "An Application of the BGP Community Attribute in Multi-home Routing", INTERNET-DRAFT, <draft-chen-community-usage-00.txt>, January 1996

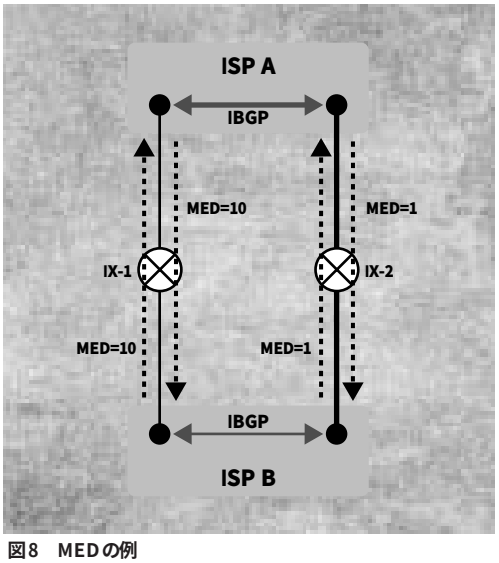

図8　MEDの例

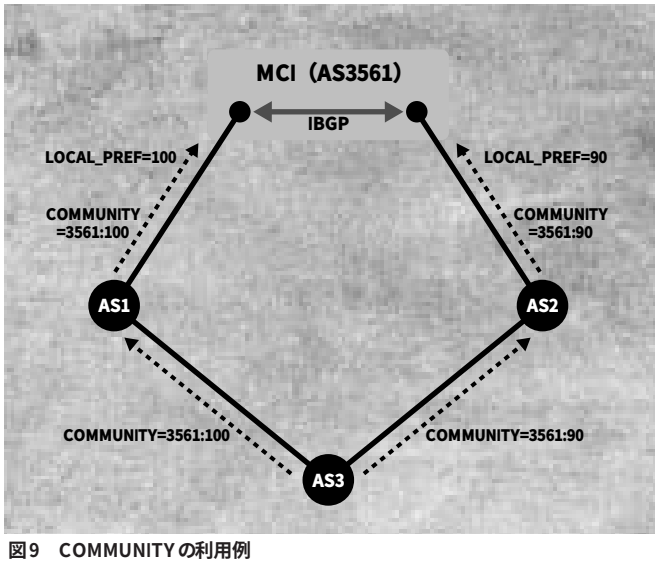

図9　COMMUNITYの利用例

［インターネットマガジン　バックナンバーアーカイブ］ ご利用上の注意

このPDFファイルは、株式会社インプレスR&D（株式会社インプレスから分割）が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン　バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

http://i.impressRD.jp/bn

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容（技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど）は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者（執筆者、写真の撮影者、イラストの作成者、編集部など）が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元（株式会社インプレスR&D）、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D（当時は株式会社インプレス）と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接的および間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先
株式会社インプレスR&D
All-in-One INTERNET magazine 編集部
im-info@impress.co.jp