

新米エンジニアのための

初歩の

インターネット技術

《第9回 プロバイダー間のルーティング》

浅羽 登志也

asaba@ij.ad.jp

株式会社インターネットイニシアティブ

日本でプロバイダーの数がこれだけ増え、相互接続もますます複雑になってくるとプロバイダー間のルーティングが実際にどうなっているのか気になってきます。今回はプロバイダー間のルーティングをテーマに、ルーティングポリシーの問題、BGPと呼ばれるプロトコル、IXの形態、現在開発が進められているRouting RegistryやRoute Serverの実験について解説していきたいと思います。

はじめに

創刊一周年だそうである。月日の経つのは早いものである。近年のこの月日の経つスピードの早さは有史以来の感もあり、ふと1年経ってみて、確実に1つ年齢を重ねたこと以上に自分が一体何をなし得たのか疑問にさえ思う。

この連載は第2号からスタートした。途中春ごろの本誌月刊化の際には連載継続が危ぶまれながらも（ぐうたらな著者が勝手に危ぶんでいただけのような気もするが）、ここまで無事(?)に続けられている。これがいかに奇跡に近い傍伴であるかは、10月号の目次をご覧になれば御理解頂けるに違いない。

そもそも国内初めてのインターネット専門誌と銘打って、鳴り物入りでスタートしたインターネットマガジンであるが、1年経った今、同じような雑誌がそろそろ片手では足りなくなりそうなくらい（既に足りないのだろうか?）ともかく著者が数えられる限界を超えてしまっているのは確実である）あちらからもこちらからも創刊されている。このままのペースで増え続ければ、来年の今ごろには、インターネット関連雑誌ライター相関マップが毎号掲載されるようになるやも知れぬ。

冗談はさておき、この1年で星の数ほどプロバイダーがサービスを開始したのは確かである。創刊号の広告を見ると、当時はまだ、IJJ、AT&T Jems、InfoWebのほかには、リムネットが「誕生」したばかりであり、BEKKOAME/INTERNETが「筆を貸し」始めたところであり、ASAHIネットもまだインターネット「新時代」をスタートさせたところであった。もちろん、すべてのプロバイダーが広告を載せていたわけではないにせよ、たったのこれだけであ

る。それが今や、毎号のプロバイダー相関図（著者は個人的にこう呼んでいる）を見るにつけ、あの図のレイアウトをされているデザイナーの方の努力に涙を誘われるほどである。先月号から今月号にかけても、あそこがここに動いてさらにあっちにも繋がったはずだから、そうするとこの辺のスペースをもっと開けないと書き込めないし、うまくしないと他のプロバイダーの丸を線が横切ってしまうなあなどと要らぬ老婆心も湧いてくる。

さて、これだけ多数のプロバイダーが存在し、しかも相互に入り乱れて接続されるとなると、気になるのはプロバイダー間のルーティングである。なんと今月は予告どおり、この話題について触れたいと思う。しかし、以前にも一度プロバイダー間のルーティングの話をしたこともあるので、その時の話をちょびっとおさらいしつつも、まだ触れていない話題を中心に話を進めたい。

プロバイダーとAS

さて、読者は、現在インターネット上にいったいいくつの経路情報が流れているか想像がつくだろうか？

この原稿の執筆の時点で、実に31,675経路あり、これが約853のプロバイダーからアナウンスされている。これを結構多いと思われるだろうか、それとも、意外に少ないと思われるだろうか？

プロバイダーの数として約853と書いたのはなぜかというと、この数はこれらの経路をアナウンスしているAS番号の数だからである。以前この連載でも、プロバイダーのネットワークは、それぞれがAS（Autonomous System: 自律システム）とみなされ、それぞれ固有の番号を与えられていると書いた。だが正確に言えばAS番号の数がプロバイダーの数にはならない。プロバイダーによってはAS番号を持っていな

いところもあるし、管理の都合上複数のAS番号を持っているところもあるからである。しかし、おおむね1つのAS番号が1つのプロバイダーに対応していると考えて大きな間違いはない。

ASというと、記憶力の良い読者は憶えておいでだと思うが、先月号のOSPFの解説でも登場した。ASなどと難しい言葉を用いているが、その心は、インターネット全体を、ひとまとまりにできる範囲ごとに細かく区切ったものと考えればよい。

どうい範囲がひとまとまりにできるかというと、「ある1つの管理主体によって一貫した管理が行われている範囲」である。

これまた小難しい言い方をしてしまったが、たとえば、鉄道を例に取れば、JRとか、小田急電鉄というのは同じ鉄道会社であっても、それぞれが独立した経営母体を持ち、それぞれが一貫した運営方針やダイヤのもとで運営されている。鉄道の世界にASという概念を持ち込むとすれば、これらはそれぞれが独立したASとなるわけである。これらの独立した鉄道は、いくつかの大きな駅で相互に接続されていき、全国の鉄道路線ができあがる。相互接続する際の列車間の連絡をどうするかは、今度は1つの鉄道会社のみでは決定できず、相互に協議の未決定されるものと想像される。

同じように、インターネットの世界ではプロバイダーのそれぞれが1つのASと考えることができ、ASが集まって全体として1つのインターネットが構成されているのである。したがって、このAS同士のおつきあいをどうするかによって、インターネット全体の相互接続性が左右されることとなる。

AS間の相互接続は、2つのAS同士が直接接続されるやり方と、NAP（Network Access Point）やNSPIXP（Network Service Provider Inter-eXchange Point）、そ

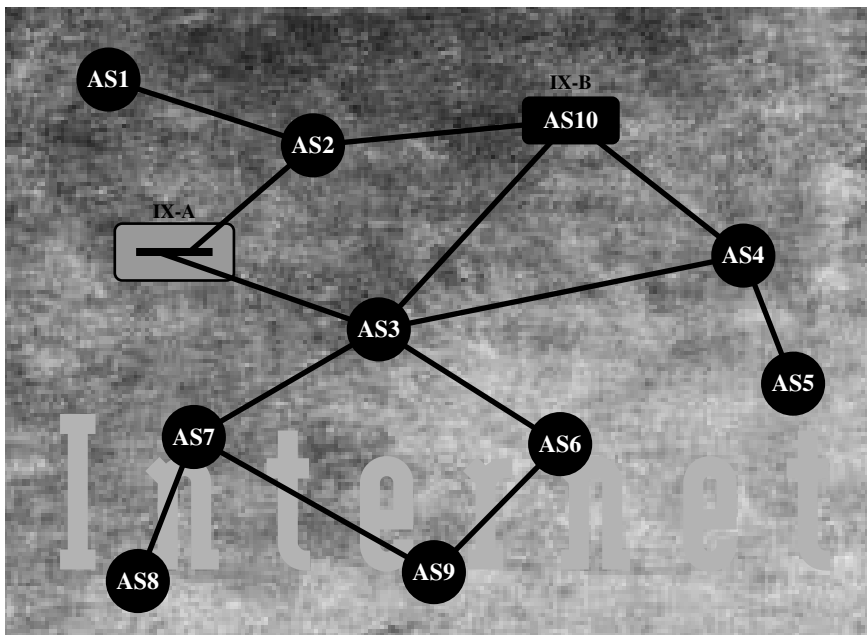


図1 プロバイダー（AS）同士が相互接続されてインターネットが形成されている

れから CIX (Commercial Internet eXchange) ような相互接続用に特別に作られた場所 (IX: Internet eXchange) を経由するやり方とがある。図1は、以前に掲載した図を書き直したものである。以下、この図を用いて解説を進める。

ルーティングポリシー

1つのASの内部であれば、ルーティングというのは比較的簡単である。前号で解説したOSPFのようなルーティングプロトコルを用いて、リンクのコストに応じて最適な経路を決めていけばよいのである。

しかし、AS間のルーティングとなるとそう簡単にはいかない。単純にこちらのASを経由したほうが近いからこちらを通そうと思っても、「おめえの所から来るトラフィックなんて通してやんね！」と言われてしまうかもしれない。また、「うちんとこは通し

てあげまっせえ、パチパチパチ (そろばんをはじく音) 月々XXXX円でどうでっしゃろ？」などと言われるかもしれない。

すなわち、AS間のルーティングでは単純にコストを定義して最適経路を選ぶというやり方が、必ずしも成立しないのである。これは、それぞれのASの運用上の都合や、ビジネス的なそろばん勘定、また、政治的理由などさまざまな要因によって左右される。つまり、そのASのルーティングポリシーに合わないトラフィックは受け入れてもらえないのである。

簡単な例として、図1のAS3、AS6、AS7、AS9の接続について考える。このような接続形態で、図2のように、AS6が、AS6を通過するトラフィックを許していないような場合がある。

このような場合は、AS9は、AS6以外のASに到達するためには、AS7経由の経路を選択しなければならない。これは、AS6

の「通過を許さない」というポリシーを受けて、AS9側がそれに合わせた形になる。

また別の例としては、図3に示すように、AS4からAS1にデータを送る際に、AS4では、AS4 AS3 AS2 AS1という経路よりも、AS4 AS10 (IX-B) AS2 AS1という経路を選択したいような場合がある。これは、AS4のポリシーである。

だが待てよ。そうすると、この場合AS5が、AS1との通信に関してAS4とはまったく逆のポリシーを持っていたらどうするのだろうか？ つまり、AS5は、AS1にデータを送るときに、AS4 AS3 AS2 AS1という経路を通りたいような場合である。(ここで即座に、インターネット上のデータグラム転送は、hop-by-hop だからそんなの無理じゃん、と思われた読者がいたら、すぐ履歴書を送ってほしい。:-))

と、例を挙げていけばきりがないが、ともかくこのようにAS間のルーティングには、

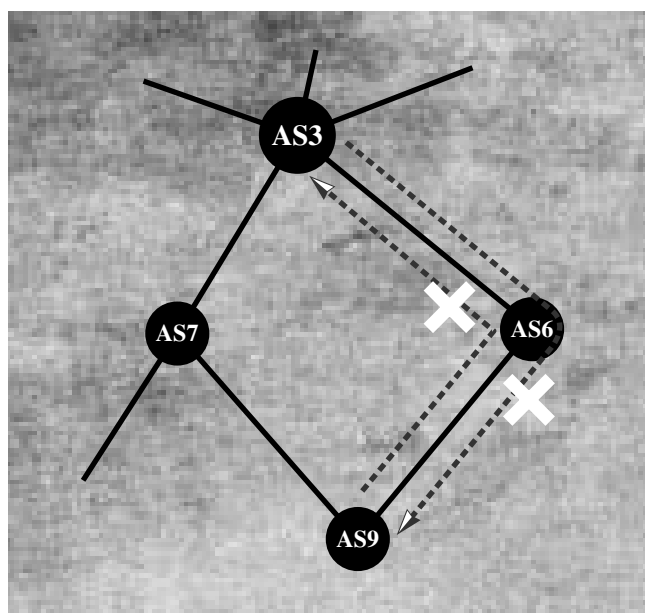


図2 途中にあるASのルーティングポリシーによって通過できない経路がある

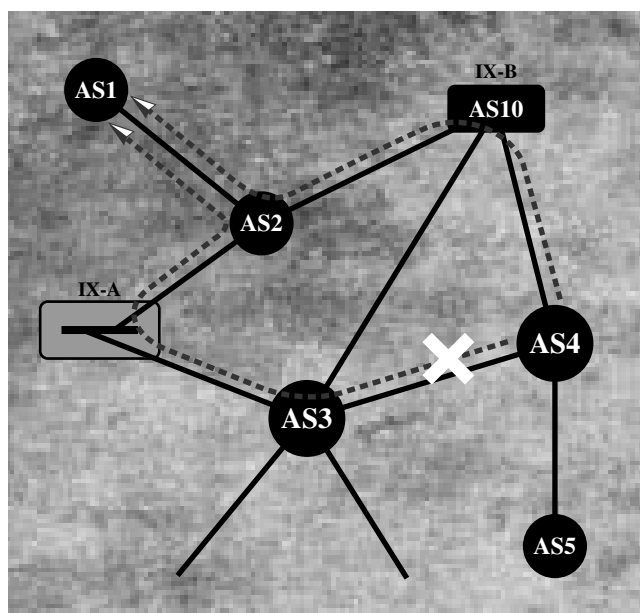


図3 ASが自らのポリシーに従って経路を選択する場合もある

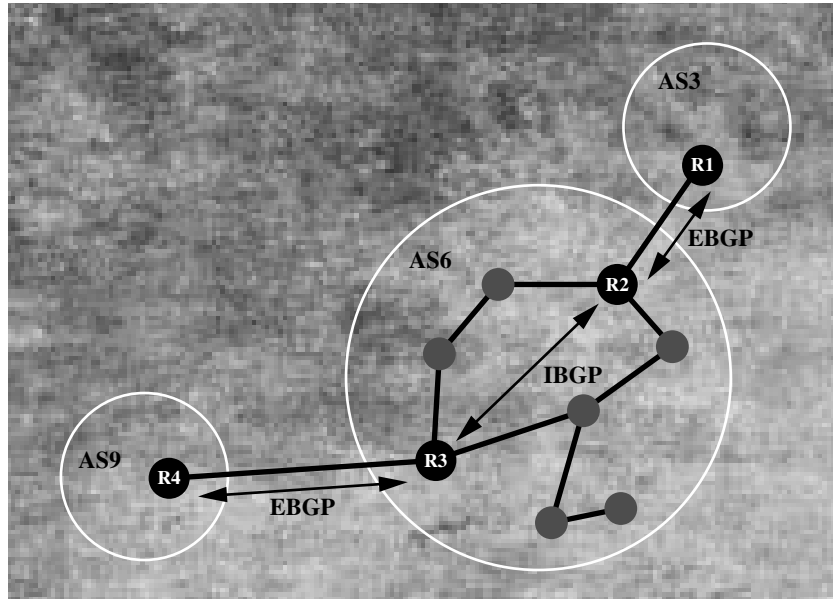


図4 AS間のルーティングのために、「BGP」というプロトコルが広く使われている

それぞれのASの思惑と絡んで、さまざまなポリシーが存在するのである。では、このようなポリシーを一体どのように実現していけばよいのだろうか？

BGP

AS内部のルーティングプロトコルとしてOSPFやRIPがあったように、AS間のルーティングを行うためのプロトコルというものも存在する。現在はBGP (Border Gateway Protocol) というプロトコルが広く用いられている。バージョンは4である。

BGPでは、前述のルーティングポリシーをある程度実現するために、Path Attributeという考え方を導入している。経路情報としてはもちろんデスティネーションとなるネットワークアドレスが伝えられるのだが、それに加えて、その経路情報が伝わって来た経路に関する情報も同時に伝えられるのである。

BGPで経路情報を受け取ったルーターは、各ネットワークに対する経路情報に付

加されているPath Attributeをもとにして経路の選択を行う。経路情報の伝達の方式としては、RIPなどと同様に、hop-by-hopに行われるため、このBGPでの方式をPath Vector方式と呼ぶこともある。

BGPは、図4に示すように、R1、R2、R3、R4などのASの境界にあるルーター間で用いられる。つまり、R1とR2の間、R3とR4の間でBGPによる経路情報の交換が行われるのである。また、R2では、AS3から受け取った経路情報をさらに先のASに伝える必要があるため、同じASに属するルーター同士であるR2とR3の間もBGPによる経路情報の交換が行われる。R1とR2や、R3とR4のように、異なるASに属するルーター間でのBGPセッションをEBGP (Exterior Border Gateway Protocol) と呼び、R2とR3のように同じASに属するルーター間でのBGPセッションをIBGP (Interior Border Gateway Protocol) と呼ぶ。

AS6の内部の他のルーター間では、OSPFなどのプロトコルを用いて、R2やR3が外部から学んだ経路情報を伝える(そう

しなくてもよい方法もあるが、本連載の守備範囲を越えそうなので省略する)。

さて、ではBGPで用いられるPath Attributeとしてはどのようなものがあるだろうか？いちばん分かりやすいものとしては、通過してきたASの列である。このAttributeをAS Pathと呼び、経由してきたASのAS番号を逆順に並べて表現する。

たとえば、図3で、AS1が202.232.0.0/16をBGPでアナウンスしたとすると、AS4は、AS10から"10 2 1"というAS Pathをもつ202.232.0.0/16に対する経路情報を受け取り、また、AS3から"3 2 1"というAS Pathを持つ202.232.0.0/16に対する経路情報を受け取ることになる。

AS4では、AS3とAS10から受け取った経路から、ポリシーに従って、経路の選択を行う。先の例のポリシーに従えば、AS4では、AS10から受け取った"10 2 1"というPath Attributeを持った経路を選択することになる。

ここで注意しなければならないことは、AS4がさらにAS5に経路を伝えるときには、

この202.232.0.0/16に対して"4 10 2 1"というAS Pathを持つもののみを送るということである。AS4で採用されなかった"3 2 1"というAS Pathを持つ経路情報を、さらに"4 3 2 1"とPath Attributeを更新して、AS5に送るようなことは意味がないのである。ほかにもいくつか経路の選択に用いられるPath Attributeが存在するが、ここでは割愛する。

Routing RegistryとRoute Server

さて、先の節で現在約853のASがインターネット上で経路のアナウンスをしていると言った。これだけの数のASが、プロバイダー関連図にみられるような複雑怪奇な相互接続を行っている（しかもあの図は日本だけであって、世界中を考えると、腕の良いデザイナーもはだして逃げるほどであることは想像に難くないだろう）と考える

と、BGPを用いた経路選択というものも真面目にやりだすとかなりの労力とルーターのCPUを消費するであろうことはお分かり頂けるだろう。

そこで、このAS間の経路選択をより簡単に行うための仕組み作りが現在進められている。それが、Routing Registry (RR) とRoute Server (RS) である。

Routing Registry とは、インターネット上でアナウンスされるネットワークの情報や、各ASに関する情報、また、それぞれのASの持つルーティングポリシーの情報などのデータベースである。インターネットに経路情報を流す際には、まずここへ、経路情報として流すネットワークに関する情報を登録する必要がある。ここに登録されている内容には、その経路情報を最初にアナウンスするASの番号 (home as) や、アナウンスされるデスティネーションネットワークが割り当てられている組織の情報などが

ある。この経路情報のRRへの登録は、home asの管理者が行う。

RRは、現状では、Merit、RIPE NCC、MCI、CA*netなどで別々に動作しており、各々の間でのデータベースの同期にはSMTPやFTPなどのプロトコルが用いられていると聞く。RR間の情報の同期のためのメカニズムはまだ開発の途中である。日本でもKDDでRRの実験的な運用が行われている。

一方、Route Serverは、RRからアナウンスされるネットワークの情報や各ASのポリシーの情報を取り出してきて、ポリシーによる経路選択などを行うサーバーである。これについて述べる前に、IXの構成の仕方について簡単な解説が必要であろう。

IXの作り方としては、CIXに代表されるlayer 3 (ネットワーク層)での相互接続を行う方式と、MAE-EAST/WESTやNAPやNSPIXPに代表されるlayer 2 (データ

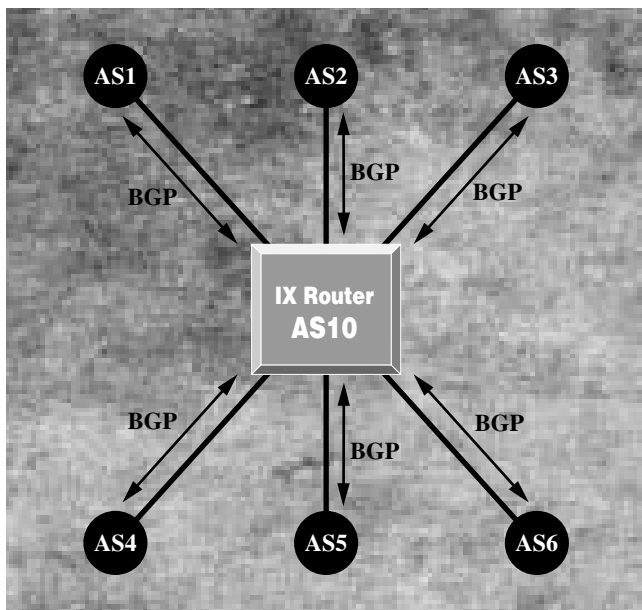


図5 layer 3での相互接続では、1つのルーターにASがそれぞれリンクを持つ

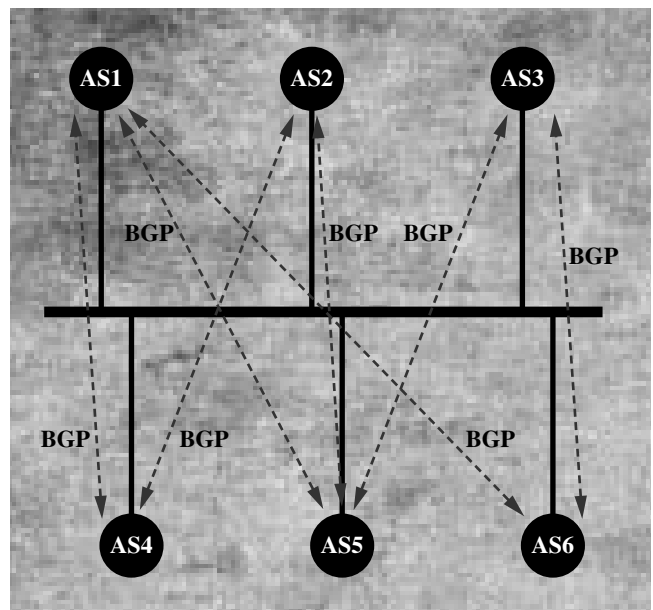


図6 layer 2での相互接続では、合意したAS同士でリンクを持つ

ンク層)での相互接続を行う方式の2通りがある。layer 3の相互接続とは、図5のように相互接続を行うポイントにルーターを1つ設置して、そこで相互接続を行うASがリンクを持つ方式である。この場合、その相互接続点は独自のAS番号を持ち、相互接続されている他のAS間の経路選択はすべてそのIXのルーターによって制御される。

一方、layer 2での相互接続とは、図6に示すように、相互接続ポイントには、FDDIやイーサネットのようなデータリンクメディアのみがあって、そこに相互接続を行うASがそれぞれルーターを設置する。そこでトラフィックの交換を行うことに合意したAS間でそれぞれ別個にBGPのセッションが張られる。この場合には、ポリシー制御は各AS間のルーターで行われることになり、layer 3での相互接続の場合に比べて、各AS単位でポリシーを実現することができるので、現在ではこちらの方式が主に用い

られている。

しかしこの方式では、そのIXでトラフィックを交換する相手のASが増えれば増えるほどBGPのセッション数が増え、またASの境界のルーターでのポリシーによる経路選択のオーバーヘッドが増えてしまう。

そこで、Route Serverの考え方が導入された。図7のように、layer 2の相互接続ポイントにRSを設置し、各ASのルーターは、このRSにのみBGPのセッションを張る。RSは、RRから各ASの持つルーティングポリシーを取り出してきて、各AS向けに経路選択を行い、選択された経路のみをそのASのルーターに渡す。こうすれば、BGPのセッション数も減り、また本来パケットの転送がメインの仕事であるルーターのCPUの負荷を減らしてやることができる。

しかし、現在はまだRSは実験段階であり、どのIXでもRSのみに頼った経路選択を行っているところはない。

おわりに

RRだRSだと言ったって、結局人間の管理者がきちんとRRに情報を登録していなければうまく動きはしない。したがって、やっぱり偉いのは人間様だということになるだろうか？ うーむ、逆に言えば、ちゃんと動かない場合にもやはり人間の責任ということになり、自分で書きながらちょっとドキドキしてしまう。

今回の話は、あまりエンドユーザーには関係ない話だったかもしれないが、自分の組織がつながっているプロバイダーの先のほうでは、いったいどんな風に経路制御が行われているか興味をお持ちの方も多いのではないだろうか？ こういう世界にドブブリと漬かってみたいと思ったあなた！履歴書を送るのを忘れなく。:-)

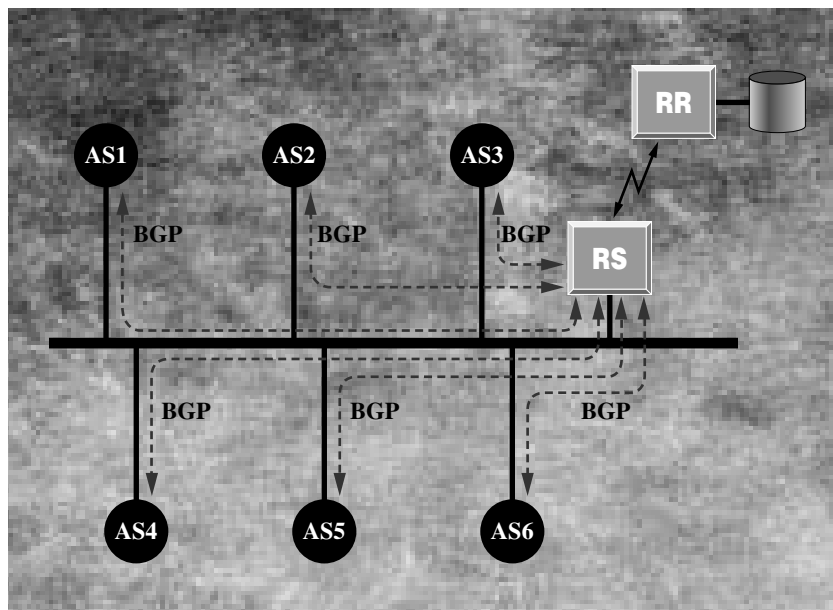


図7 Route ServerがRouting Registryからルーティングポリシーの情報を取り出し、選択された経路を各ASに伝える



[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の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