

新米エンジニアのための

初歩の

# インターネット技術

《第14回 もっと難しい「マルチホーム」》

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

7月号では、マルチホームのルーティングの基本について解説しましたが、ISP間のマルチホームはもっと難しい問題が起こります。今回はマルチホームをさらに深めてみることにしましょう。

はじめに

最近面白くないという声がちらほら聞こえる。しかしよく声を聞いてみると、どうやら本題の技術的な話に入る前の雑談の部分のことをいっているらしい。しかもどうやら、時々そこにそれとなく埋め込んでいる「知らない人には笑えない内輪ネタ」。しかもインターネットマガジン20万読者のうち3人笑えば儲けものくらいのマニアックなギャグのことをいっているらしい。うむ、これをもっと面白くしろというのは難しい要求だ。イチローちゃんて満足できないというのならもうあとはジローちゃんを出すしかないのだが、イチローちゃんとジローちゃんでは受ける読者層が異なるのでそうは簡単ではない。唯一IJとXXXXネットにマルチホームしている読者のみが笑えるのかもしれないが、そう考えるとマルチホームもあながち悪くないものなのかもしれない。

冗談はさておき、今回は、ある組織がマルチホームする場合について書いたら紙面が尽きてしまった。今回は引き続き前回書き損ねたISPが他のISPにマルチホームする場合について書いてみようと思う。

ところで、人々の行動にはかならず動機というものがあるらしい。テレビの刑事ドラマを見たり、推理小説を読んだりしても、かならず犯行の動機というものが問題にされているようだ（実際の犯罪捜査に関してどうかはよく知らないが）。だが胸に手を当てて考えてみるに、はて、本当にそうかしらとも思う。自分の行動パターンを考えてみると、どうもいきあたりぱったりで動いているようなことも多く観測される。観測されるなどと表現すること自体、なんらかのインテリジェントな理由づけがなされたうえで何かをしているわけではないということを暴露しているようなものではある。

さて、では人はなぜマルチホームするの  
 だろう？それはそこに線が2本あるからさ、  
 という答えが返ってきそうであるが、これ  
 は、人はなぜ原稿を書くのかという問いに  
 対して、そこに切があるからだと答える  
 のと同じくらい本末転倒な答えである。筆  
 者は決してマルチホームを否定しているわ  
 けではないのだが、あまり深く考えずにと  
 もかく線を2本張ってしまえばそれで何とかな  
 るだろうと思うのはあまりにもいきあたり  
 ばったりではなからうか？

ISPが他のISPとマルチホームする場合  
 も、組織がマルチホームする場合と同様、  
 1つのISPにマルチホームする場合と、複数  
 のISPにマルチホームする場合がある。そ  
 れぞれの場合で得られる効果や、それを行  
 うに当たって考慮すべき事項が異なる。以  
 下では順番に解説していく。

ISPが1つのISPにマルチホームする場合

図1に、ある2つのISP同士が異なる2本  
 のリンクで相互に接続される場合の形態を  
 簡単に示す。

図1で、R1、R2はISP A内部のルーター、  
 R3、R4はISP B内部のルーター、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。また経路情報の交換はBGP4  
 で行われるものとする。

何もこだわらないとどうなるか

まず一番簡単な場合について考える。こ  
 れは何ら特殊な設定をしない場合、つまり、  
 R1とR3の間とR2とR4の間でまったく同一  
 の経路情報を交換し、さらに交換する経路  
 情報のPath Attributeに関して、Router ID

とNext Hop以外はまったく一致させるよ  
 うな場合である。

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  
 部の他のルーターに伝搬される。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  
 のルーターに伝搬される。

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  
 ルーターに到達するコストの小さい方が選  
 択される(ただし、BGPで学んだ経路を  
 External Linkとして、OSPFのType2メト  
 リックを用いて内部に流している場合はこ

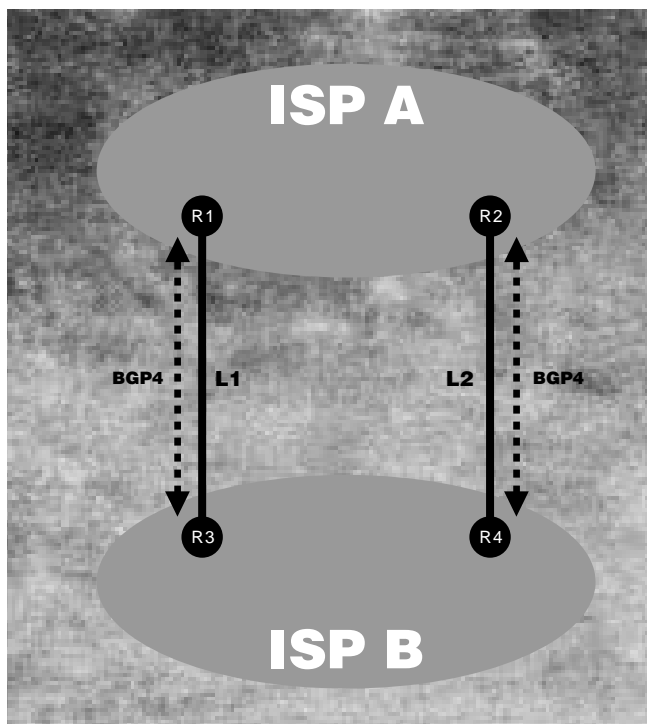


図1 同一ISP間のマルチホームの例

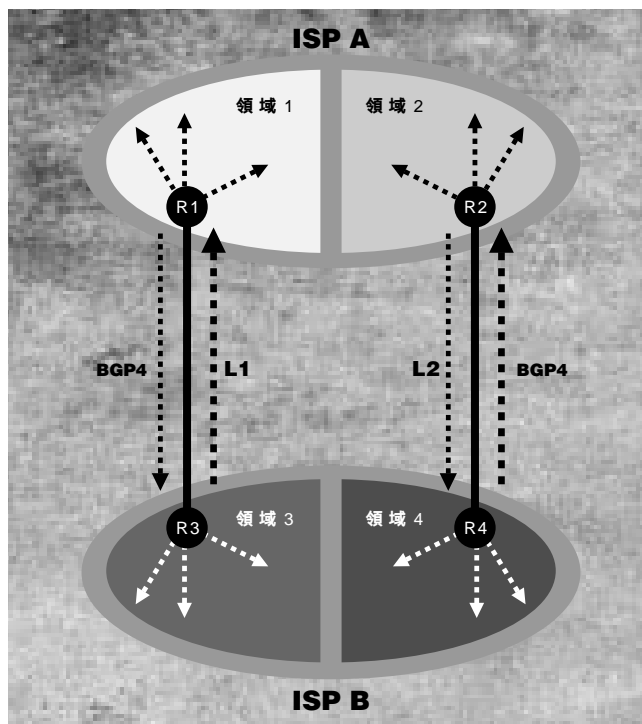


図2 何もこだわらない場合

の限りではない)。図2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。

この場合、図3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で示した。

H4で示した。

### 片方はバックアップにすると...

さて次に簡単なのは、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するような場合である。

これはBGPのMULTI\_EXIT\_DISC (MED) Path Attributeを用いることにより簡単に実現することが可能である。同じ相手から複数経路で経路情報を受け取った場合には、MEDの値がより小さい経路の方が優先される。したがって、たとえば図4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に設定すればよい。

この設定はリンク2がIX経由での接続で

あるような場合によく用いられる。IXはそこにつながっているすべてのISPの共有物的な意味合いが強いのでできるだけ必要のないトラフィックは乗せないほうがよい。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。

### ちょっと難しい場合

さて、いままでは簡単な例を考えてきたが、つぎにちょっと難しい例を考えてみよう。

図3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

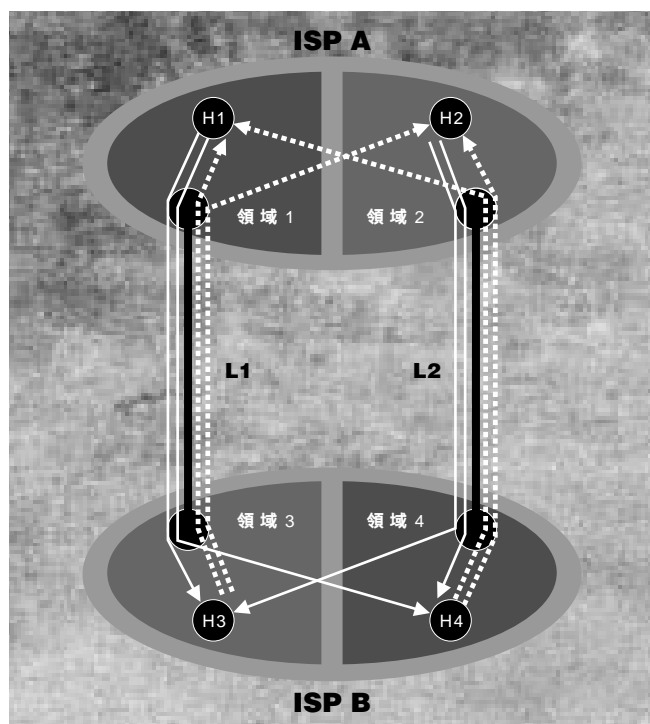


図3 ISP A とISP B の間のトラフィックの経路

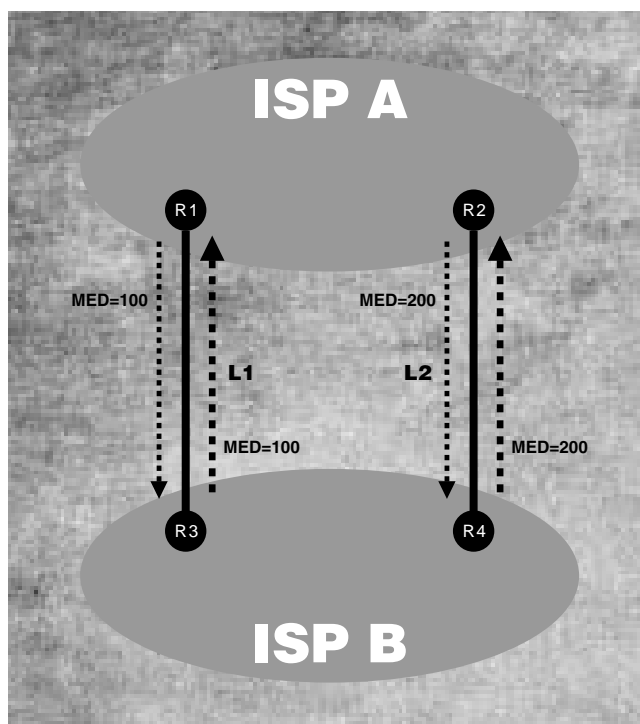


図4 片方をバックアップにのみ用いる場合

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よう。現実にも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場合というのはそんなにあるケースではない気がするが、1つ典型的な例を考えるとすれば、領域1と領域3が東京にあり、領域2と領域4が大阪にあり、ISP Aの内部で領域1と領域2が45Mbpsの回線(L3)で結ばれていて、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。このとき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3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が帰りにISP B内の1.5Mbpsの回線を經由してしまい、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なくなるので、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こと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6に示したようにしなければならない。

ここで、1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ーのネットワークも含めてということである。

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と名付けることにする。

さて、図3と図6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3の場合と同じでよい。

図3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6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かならず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かならずL2を通っていることが分かる。言い換えれば、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、ということになる。

であれば、図7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

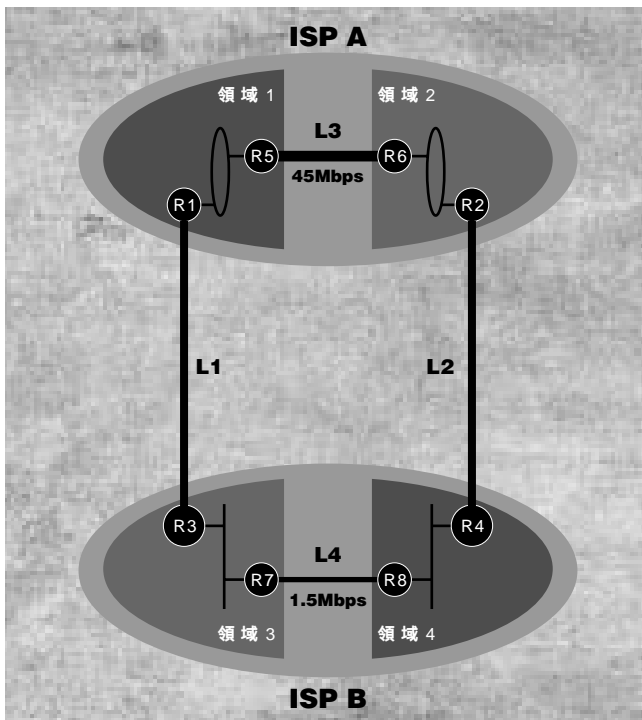


図5 ISP A内部のリンクが太い場合

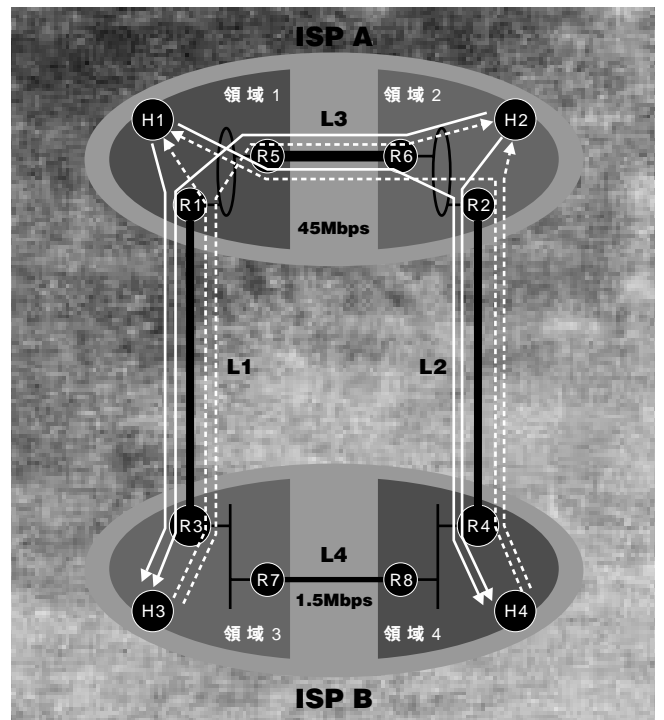


図6 経路の対称化

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よい。

これは、いわば先にあげた2つの簡単な例の合わせ技みたいなものである。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1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厄介かもしれない。さらに経路情報の集約のことを考えると、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。

ISP B側でNETS3とNETS4を区別することを考え出すと、本当はここでBGP Confederationなんていうものも考えたくなくなるのだが、ちょっとまだそれを書く勇気が持てないので、ここでは割愛する。あしからず。

### もっと難しい場合

さて、結構疲れてきたのでもういいやという気にもなるが、そうもいってられない。もうちょっと誌面があるし、さらに難しい場合もあったりするので、まだまだ書かなきゃいけない。辛いところである。

これまでは、ISP AとISP Bしか登場していなかったのだが、ここでさらに別のISPが絡んできたりすると話がさらに複雑になって楽しいこときわまりない。

たとえばISP Aがさらに他のISPCにつながっていた場合を考えてみよう。

図8では、ISP CがISP Aの領域1内にあるルーター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6の場合と同じだとする。

さて、この場合はどうだろうか？

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。

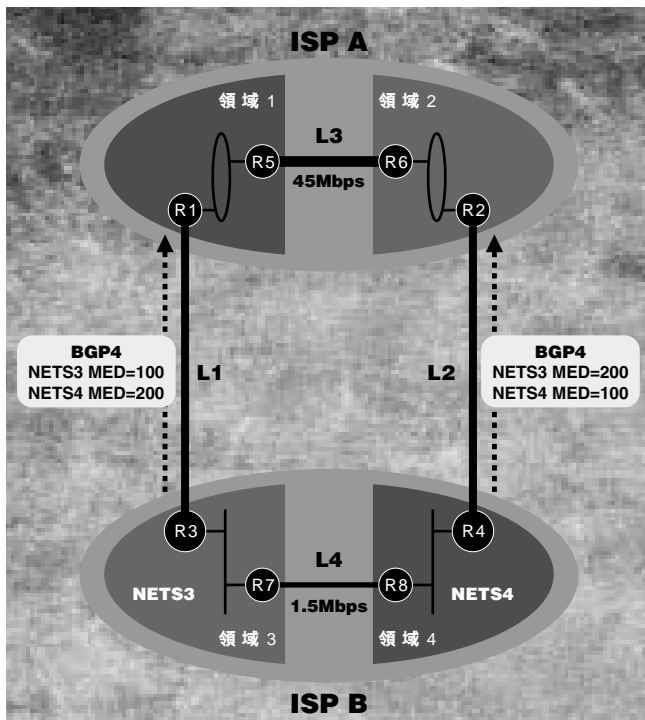


図7 領域ごとに異なるMEDの値を設定

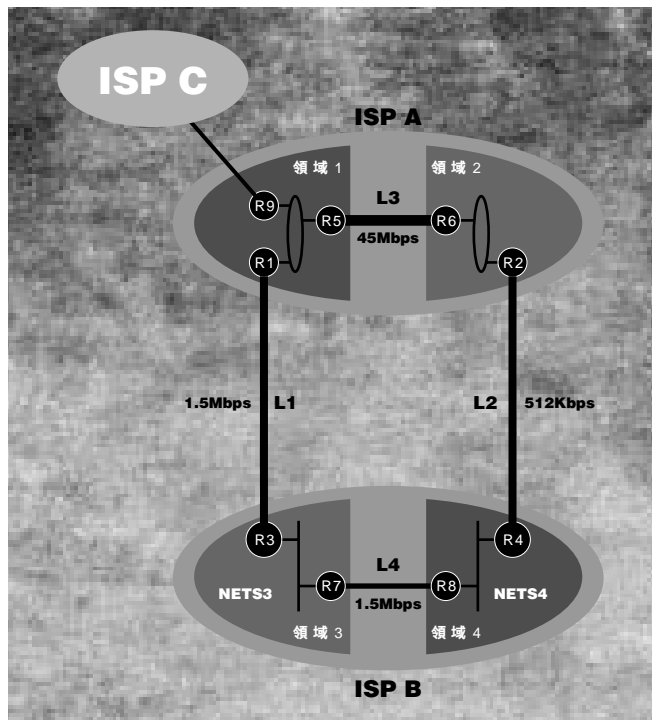


図8 さらに他のISPが絡んでいる場合



これを行うには、単にISP AはISP Cからくる経路情報を、ISP Aの経路情報と同じ扱いでISP Bに流せばよい。簡単でハッピーである。しかし世の中そううまくいかない。これを行ったところ、L2が真っ黒（業界用語で、回線が非常に混んでいる状態であることを指す）になってしまった。さて、どうしよう...

なぁに、簡単。ISP CとISP Bの間のトラフィックは行きも帰りもL1のみを用いるようにすればよい。ふむ。確かにそれができるならそれでよい。でもどうやるの？

あーら、何をいっているんだか。MEDですよ、MED。ISP BにISP Cの経路情報を送るときにL1経由の場合はMEDの値を100にして、L2経由の場合はMEDの値を200にしてやればよいんですわん。

いやいやいや。確かにそれでISP BからISP Cに向かうトラフィックをL1経由にすることはできるけれど、ISP CからISP Bの領域4に向かうトラフィックは依然L2を通過してしまう。うーむ。

結論からいうと、図6の状態を維持したまま、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いけないということになるのだが、これは一般的には実現不可能である（ただし、深緑色のルーターを使えば最近はこのことも可能であると聞いているが、すべての場合に対応できるかどうかは分からない）。

こういう場合は素直にL2の増速を考えたのがよいだろう。

できないこと

さて、先の例にも書いたが、現在の技術

ではできないことも山ほどあることを理解しておくことは重要である。実際にどういう要求があって、そのうちいまの技術で何ができ何ができないかをきちんと理解しておくことが今後の技術の発展にも結び付くのである。

いまの技術ではできないことで、できるようになると嬉しいことといえば、マルチホームしたときに複数のリンク間でうまくload-balanceすることがまず思い付く。

そもそも、現在用いられているルーティングプロトコルではリンクにかかっている負荷を元に経路を選択するための機構を持つものがほとんどない。深緑色のルーターは、リンクの負荷を経路のメトリックの1つとして用いるプロトコルを持ってはいるが、これは組織の中や比較的狭い範囲で用いられることを目的としたIGP (Interior Gateway Protocol) の範疇にはいるものなので、複数の組織やプロバイダをまたぐような状況では用いることができない。

それでも、それぞれのリンクを用いて通信する相手を細かく制御することができればある程度外部とのトラフィックをバランスよく複数のリンクに振り分けることもできないわけではない。複数のリンク経由で受け取る経路情報に対してきめ細かくlocal\_prefなどを設定して行けば何とかなる場合もある。だがこの方法ではあくまで自分の組織から外部に出ていくトラフィックしか制御できない。

入ってくるトラフィックを複数のリンク上にバランスよく振り分けるためには、自分がアナウンスしている経路情報が先の方のネットワークにどのように伝わって行くのかをコントロールできなければならない。この機構は現在のBGPには備わっていない。これをやろうと思ったときには、それぞれのリンクから自分がアナウンスしている経路情報がどのプロバイダーをどう経由してインターネット全体に伝搬して行って

いるかをかなりきめ細かく知っていて、かつ、自分が直接つながっているプロバイダよりも先にあるプロバイダに特殊な経路のアナウンスをするような設定をお願いしなければならない。最近では、AS Path Prepend (深緑色ルーター用語)とか、AS Path Stuffing (ねずみ色ルーター用語)という、AS Path Attributeに余分にAS番号を付加することによって外に出ていく経路情報のアナウンスをコントロールするような方法も可能になってきている。しかしこれは、たとえば昔RIPでMetricを適当にいじってトラフィックをコントロールしていたのとほとんど同じ発想で、あまりきめ細かなことはできない。あくまで次の技術ができるまでのその場しのぎでしかないだろう。

筆者自身、ぼんやりとこうすればよくなるんじゃないかなと言うアイデアはあるのだが、きちんとまとまる前に書いてしまうと人に先にやられてしまうので口をつぐむことにする（我ながらうまいはったりのかましかただ :))。

おわりに

またも全部語り尽くす前に誌面が尽きてしまった。ともかく、マルチホームは難しいということがお分かり頂ければ本望なのである。

本当はあと、複数ISPへのマルチホームと、それに伴うlocal\_prefの利用とか最近出てきたcommunityの考え方、DPAの利用などについて書きたかったのだが、これはまた次回以降にゆずることにする。

今回はいつになるか分からないがIPアドレスの割り当てに関する最近の動向について書いてみたいとも思っている。

な～んていっておきながら、全然違う内容になるかもしれないので、その辺はあらかじめ御了承を :)



## [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](mailto:im-info@impress.co.jp)