

日本の IP インターネットの経路制御方式†

加藤 朗¹ 山本 和彦² 中村 素典³ 相原 玲二⁴
東京大学 九州大学 京都大学 広島大学

従来わが国のインターネットの相互接続における経路制御には RIP が用いられてきたが、経路制御上の能力の限界やオーバーヘッドの増加が問題になりつつあり、CIDR に基づいたアドレス割り当てがこの状況を加速している。また、障害発生時にはバックアップ経路が自動的に確立することも求められている。

本論文では、ネットワーク相互間の経路制御に BGP-3 を用いた場合の利点や問題点について考察する。さらに、WIDE インターネットを始めとして各ネットワークとの間の経路制御を BGP-3 に移行する作業を行っているが、この作業に関連する問題点についても述べる。また、WIDE インターネットでは RIP から OSPF に移行を行っており、この際の問題点についても報告する。

1 はじめに

わが国のインターネットは、1988 年頃から WIDE や TISN、JAIN といった 3 つのネットワークプロジェクトとしてスタートした。当時、キャンパス全域をカバーしているキャンパスネットワークは日本でも数少なく、計算機センターや情報関係学科に対する接続というのが主な形態であった。経路制御に関しても、各大学で使用されていた RIP [1] を単純に広域ネットワークに適用し、国内のインターネット全体で単一の RIP による経路制御ドメインが形成されていた。

近年のインターネットの普及とともに接続される組織が増大し、また CIDR に基づく IP アドレス割り当てを受けた組織の接続の増加とともに、経路制御の対象となるネットワーク数の増加には著しいものがある。また、このような規模の増加は、RIP では対応できないものになってきている。

JEPG/IP (Japanese Internet Engineering Planning Group for IP) では、このような事態を解決するため、従来の単一の RIP の運用に代えて、ネットワークプロジェクト間の経路制御

を BGP (Border Gateway Protocol) に移行することを勧奨している [2]。本論文では、このような移行の背景と IGP (Interior Routing Protocol) を利用した経路制御上のポリシーの表現に関する問題点について述べる。そして、WIDE インターネットを中心に行われてきている経路制御移行の実験について報告し、現時点での日本のインターネットの経路制御方式を紹介する。そして今後の方針について述べる。

2 ポリシーを導入した経路制御

2.1 日本のインターネットのポリシー

初期の日本のインターネットでは、経路制御は単一の RIP で運用されていた。各大学のキャンパスネットワークを専用線で接続し、全体がキャンパスネットワークの拡張として運用されていた。そこには経路に関するポリシーはなく、接続性が最優先であった。

その後、複数のネットワークプロジェクトが出現し、相互に接続されるようになってきた。そして、おおむね次のようなインターネット上の経路制御に関するポリシーが生じてきた：

- 1 あるネットワークプロジェクトに参加しているネットワーク間の通信は、他のネット

† The Routing System on the Japanese IP Internet

¹ kato@wide.ad.jp

² kazu@csce.kyushu-u.ac.jp

³ motonori@kuis.kyoto-u.ac.jp

⁴ ray@csl.hiroshima-u.ac.jp

ワークプロジェクトを経由しないこと。

は

この場合、両方のネットワークプロジェクトに参加している組織相互の通信はどちらを利用しても差し支えないが、一般的にはパフォーマンスの良い方を用いることが好まれる。従って、どちらを使うかを設定できる必要がある。

また、トラブルの際に相互に協力する体制が確立してくると、次のようなポリシーも加わってきた：

2. トラブルの際は 1. の限りではなく、接続性を優先する。

このようなポリシーが明らかになってくるに従って、それぞれのネットワークプロジェクトの独立性や協調性も明確になり、単一の RIP で経路制御を行っているが、それぞれは一つの AS(Autonomous System) として考えることができるようになる。以下の部分では、各ネットワークプロジェクトを経路制御上のポリシーの表明主体として捉え、AS と記すことにする。

2.2 ポリシの表現

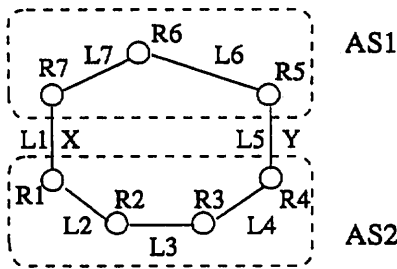


図 1: 2つの AS 間のポリシー

今2つの AS を考え、図 1 のように AS1, AS2 とする。相互接続地点 X, Y の間でポリシー 1 を満足するために必要な条件は、 m_n はリンク L_n に対する経路制御上の metric を表現するものとする、ルータ R_7 とルータ R_5 間の AS1 側の経路の条件は、

$$m_1 + m_2 + m_3 + m_4 + m_5 > m_6 + m_7 \quad (1)$$

であり、一方ルータ R_1 と R_4 の AS2 側の条件

$$m_1 + m_7 + m_6 + m_5 > m_2 + m_3 + m_4 \quad (2)$$

となる。従って、ポリシー 1 を実現するためには、上の二つの不等式を満足する必要がある。一般に経路制御上の metric m_i は正の整数であり、かつ経路が有効であるためには、ネットワーク上の任意のルータ間の metric の和が経路制御プロトコルによって定まる一定値 m_∞ 未満であることが要求される。

2つの AS の相互接続地点が X, Y だけではなく多数存在する場合や AS が多数存在する場合には、そのすべての AS の組合せについて同様な不等式を満足することが要求される。従って、一般的にはポリシー 1 は、全ネットワーク中のリンクの数を n としたときに、 n 元連立不等式で表現することができる：

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ & \ddots \end{pmatrix} \cdot \begin{pmatrix} m_1 \\ m_2 \\ \vdots \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \\ \vdots \end{pmatrix} \quad (3)$$

ただし、 a_{ij} は -1, 0, 1 のいずれかの値をとり、 $0 < i \leq n$ の i に対して $0 < m_i < m_\infty$ であるものとする。

さらに、経路が有効に機能するためには、全ての経路に対して、経路上の各リンクに付された metric の合計が m_∞ に達してはならないという制約も加わる。

ポリシー 2 の相互バックアップの問題は、ポリシー 1 が実現できている環境において、経路のスクリーニングを行わないことによっておおむね実現できる。図 1 において、 L_3 が切断し、AS2 が二つに分断された場合、 R_2 と R_3 の間の接続性は $L_2, L_1, L_7, L_6, L_5, L_4$ という経路で保たれている。 L_3 が復旧した場合には

$$m_2 + m_1 + m_7 + m_6 + m_5 + m_4 > m_3 \quad (4)$$

という条件が必要であるが、これは式 (2) から得ることができ、ポリシー 1 が実現されていれば式 (4) も満たされる。

ポリシー 2 を考慮する場合の問題点は、経路が正常時に比べて長くなる場合があり、経路上の metric の合計がポリシー 2 を考慮しない場合に比

較して m_{∞} に近づく。従って、metric の組合せによっては m_{∞} を越えてしまい、一部でバックアップが不可能になる場合もありえる。

2.3 RIP によるポリシーの表現

RIP はもともと中小規模のネットワークの経路制御を目的としており、IP データグラムの中継 1 hop に対して metric を 1 とするのが基本である。従って、3 hop で到達する場合には、経路に対する metric は 3 となるが、多くのルータでは 1 より大きな metric を設定することも可能である。

RIP における m_{∞} は 16 と小さく、metric = 16 は目的経路が利用不可能であることを表明していると解釈される。このため、ポリシー 1 を満足する metric の組合せとしての解が存在するためには、ネットワーク全体の規模が比較的小さい場合に限られる。

ポリシー 2 の相互バックアップ問題であるが、RIP では m_{∞} が小さいため、トポロジ上では到達できる経路が存在しても、経路制御上有効な経路とはみなされない場合が多くなってくる。そのため、ポリシー 2 を完全に実施することは現実的には不可能である。

さらに、両方の AS を利用できる組織間の通信は、AS1 経由、AS2 経由いずれを選択してもよく、一般的には高いパフォーマンスが得られる方が選択される。RIP において、ホップ数は少ないが帯域の狭い経路と、ホップ数は多いが高帯域が得られる経路の使い分けをする場合には、ホップ数の少ない方の経路上の metric にオフセットを加える操作が必要になる。このことは、ポリシー 1, 2 で考えてきた m_{∞} の問題をより深刻なものにしている。

2.4 OSPF によるポリシーの表現

RIP の場合、 m_{∞} が小さいため、RIP によって経路制御上のポリシーを表明することは非常に大きな制約がある。OSPF [3] は Link-State 型の経路制御プロトコルであり、Distance-Vector 型のそれに対してループの問題やトポロジ変化への追従性などの点で優れているとされている。

OSPF では、内部の経路に対する metric は、各 TOS (Type of Service) 毎に定義することができるが、TOS0 と呼ばれる最も一般的な TOS に対する metric で代表することが多い。

また OSPF では、内部経路の metric は 16bit の符号無し整数で表現されている。つまり $m_{\infty} = 65535$ であり、RIP による場合に比べて対応可能なネットワークは格段に多くなる。

具体的にどの程度のトポロジになった場合に解が存在しなくなるかということは現時点では明らかではないが、現在の日本のインターネットの相互接続のトポロジ (付録 A 参照) では可能な metric の組合せは存在するのではないかと思われる。

しかしながら現実の運用に際して、ネットワーク全体のトポロジが変化した場合、連立不等式をその度に解き、可能な metric の組合せでネットワークが動作するようにそれぞれのルータに必要な設定を行うことは、運用上問題であり、現実的ではない。

2.5 EGP を用いた経路制御

ある AS 内のトポロジの変化に対して、管理母体の異なった AS 内部の経路制御上のパラメータまで再設定が必要になることは、ネットワークの運用管理上好ましくない。そこで、経路制御を AS 内部のものと AS 間のものを別々に運用することが考えられた。AS 内部の経路制御プロトコルを IGP と呼び、AS 相互間の経路制御プロトコルは EGP (Exterior Routing Protocol) と呼ばれている。前章で述べた RIP や OSPF は IGP の一種であり、EGP としては、EGP [4] や BGP [5] などが知られている。

EGP を利用した場合、経路の決定は EGP と IGP の 2 段階構成で行うことができる。従って、

EGP で得られた経路と IGP で学習した経路の両方が存在する場合には IGP の方を採用する

という方針によって、ポリシー 1 を容易に実現することができる。

またバックアップの問題に関しても、図 1 において、ルータ R_2 では、 R_3 などの経路は通

常は AS2 における IGP で得られるのでこれが採用されるが、 L_3 で AS2 が切断された場合には、AS1 からの EGP で経路を得ることができる。従って、AS2 の経路を AS1 が相互接続点 X, Y でアナウンスする場合には、バックアップの問題も解決する。

複数の AS に接続されているマルチホームネットワーク間の通信において、どちらの AS を経由するべきかという経路選択の問題は、EGP の導入だけで解決することはできない。この一つの理由として、EGP を導入した場合、AS1 側の IGP は AS2 側の IGP と全く独立に運用されるため、双方の metric を比較することは意味がなくなってしまうということをあげることができる。Metric が経路の質を相対的に表現しているとしても、どちらの経路が優れているかという判断はできない。

マルチホームネットワークの問題は、ポリシを考慮した経路制御に関係するだけではなく、CIDR [6, 7] における経路の集成的問題や双方の AS 間の経路情報の漏洩など多くの問題に関係しているので、別に議論したい。

3 日本のインターネットにおける経路制御

3.1 RIP の問題点

当初キャンパスネットワークの延長として運用されてきた日本のインターネットは、経路制御には RIP を用いていたが、種々の問題点が明らかになってきた。その一つは経路制御に対するポリシの実現に関する問題で、前章で述べた通りである。現実には RIP の限られた metric 空間を駆使してポリシ 1 を実現しているため、ネットワークのトポロジを変更した場合でも、隣接 AS にアナウンスする経路の metric には極力変更がないようにする努力が必要になる。

また RIP は m_{∞} が 16 と小さく、ポリシを考慮しない場合でも、インターネットの発展に伴って 16 hop では届かない場所も生じてきていた。このような場合には、キャンパスネットワークの入口で default 経路を生成するという運用を行うことで接続性は失なわれることはなかったが、生

成された default 経路が漏洩した場合には経路のループが発生することになり、運用上好ましくない。

RIP は 30 秒毎に経路情報を各ルータ間で交換を行っており、経路の増加とともに必要な帯域も増加している。1993 年 10 月の時点で国内のインターネットに接続されているネットワーク数は約 340 あり、一回の RIP の交換には約 7kB のトラフィックが発生する。これは 64kbps のリンクを約 1 秒占有する量である。このトラフィックは、CIDR 方式の IP アドレスの割り当て方式 [8] の普及につれ、急激に増加する傾向にある点に注意が必要である。

また、broadcast をサポートしていない X.25 パケット交換網のようなマルチアクセスネットワーク上で RIP を運用する場合、各 VC に対して経路情報交換のトラフィックが発生する。一つの物理インターフェースで複数の VC を収容する場合、経路情報交換によって長時間リンクが占有されることになり、好ましくない。

3.2 BGP

国内のインターネットの相互接続に関連する技術的問題を検討するグループ JEPG/IP は、これまで述べてきた背景を元に、早期に RIP から EGP を用いた経路制御方式に移行することを考えてきた。EGP としては、ポリシの表現能力があり、インターネットの標準化としての検討が行われているプロトコル BGP を AS 間の経路制御に用い、AS 内部の経路制御はそれぞれが任意のものを運用すればよいという結論に達した。

BGP は継続的に改訂作業が続けられており、現在広く利用可能なものは BGP-3 と呼ばれる version 3 のものである [5]。経路の集成を行うためには、現在仕様が IETF で議論されている version 4 の BGP を用いる必要がある。しかしながら、国内のインターネットが経路の集成なしには運用に支障がでてくるまでにはまだ時間がかかるだろうという見通しのもとに、一旦 BGP-3 に移行し、BGP の運用の関連する問題点をクリアにすることが合意されている。BGP-4 の実装が一般的になった時点で、BGP-4 への移行に関しては議論が行なわれる予定である。

3.3 IGP

AS 間の経路制御を BGP-3 で行う場合、AS 内部の経路制御は任意の IGP によって行うことができる。しかしながら、他の AS から別の AS へのトラフィックを中継する中継 AS (Transit AS) では、経路のアナウンスのタイミングに関して、AS 内部における BGP と IGP との間の同期の問題が指摘されている [9]。

事実上の国内の中継 AS として機能している WIDE インターネットでは、BGP と同期をとることが容易な IGP を内部の経路制御プロトコルとして使用する方法を考え、IGP として OSPF を採用することにした。OSPF と BGP-3 の相互の問題に関しては、既にガイドラインが提示されている [10]。また、BGP-3 への移行の議論を行っているときに誕生したネットワーク JOIN では、内部の経路制御を当初から OSPF で行っている。

4 経路制御の移行

経路制御方式を単一の RIP から BGP と各 AS 毎の IGP を組み合わせて運用する方式に移行する場合、種々の問題が発生する。ここでは、今までの移行に関する経過について報告する。

4.1 WIDE インターネットにおける IGP の変更

BGP に対して整合性のよい IGP の条件は、経路に tag を付けることができることであり、OSPF や RIP-2 [11]などを挙げることができる。プロトコルの実装の容易さという観点では RIP-2 より OSPF の方が複雑である。しかしながら、TCP/IP での標準としての作業が進められているために OSPF は近年多くのルータに実装されており、より応答性のよい経路制御が可能である。そこで、WIDE インターネットでは OSPF を IGP として採用することにした。

1992 年に入り Cornell 大学で作成が本格化してきた gated version3 には、OSPF のサポートが含まれている。また、Proteon や Cisco でも OSPF を利用することができるようになって

きた。そこで、実際の経路制御に関する gated プロセスは経路情報交換を RIP のまま行ない、別の gated プロセスによって OSPF による経路情報交換のみの実験を行なった。ここでは、next hop に関する取り扱いなどに関して、gated の不都合がいくつか発見され、修正が行なわれた。

その次の段階として、OSPF が独立に安定して動作するようになった 1992 年 10 月頃に、実際の経路制御を担っている gated プロセスで OSPF による経路情報交換を、RIP による経路情報交換と併せて行なうようにした。この段階では、WIDE インターネット内部の経路のみが OSPF で配送され、その他のほとんどの経路は依然として RIP で得られたものが使用されていた。

そして、実際の経路制御は OSPF で行ない、OSPF の実装の不都合があった場合には RIP で補うことができるように、OSPF が主で RIP を副とした運用に切替えたのが 1993 年 2 月頃である。

この時点での問題点は、他の AS との経路情報交換であった。AS 間の経路情報は RIP で行なわれており、経路の RIP metric は前述のポリシーにより決定されており、WIDE インターネットが OSPF によって制御されている場合でも、アナウンスする RIP metric は以前と同じものにするのが求められていた。しかし、RIP → OSPF → RIP と経路は変換されるので以前の RIP metric と同じ値で経路をアナウンスすることはできない。

そこで、隣接 AS 毎に OSPF AS external 経路に異なった tag を付け、RIP に変換する際にはこの tag に基づいて metric を決定することにした。この方法では、ネットワーク単位での metric の設定はできないが、他の AS との調整の結果、おおむね問題なく従来と同じ経路を設定できることが分かった。この問題が解決した時点で、WIDE インターネットのバックボーン部分では、一部を除いて OSPF によって経路制御が行なわれている。

一部のルータでは、OSPF AS external 経路に自由に tag をつけることができないため、現在ではアナウンスする際にネットワークの集合を指定して metric を加減するような設定になって

いる。

における AS 番号の割り当てを表 1 に、また、BGP の運用状況を図 2 に示す。

4.2 BGP への移行

WIDE インターネットが OSPF に移行したあと、1993 年春から AS 間の経路制御を RIP から BGP への移行作業が始まった。具体的な手順としては、RIP から OSPF への移行と同様に、別プロセスによる単なる経路情報交換の段階、RIP が主で BGP が副となる段階、BGP を主として RIP を副として運用する段階、RIP を廃し BGP のみでの運用の段階というフェーズによって比較的スムーズな移行が可能である。

BGP に対する運用の準備ができた AS と共同で移行作業を行っており、まだ完全に BGP に移行は完了してはいない。一つの問題は、複数の接続点を持つ 2 つの AS の間の経路制御において、最適な AS 間リンクが選択されるような制御方式が確立していない点である。図 1 の 2 つの AS において、X 側のリンクを相互接続として選択した場合、 $R_4 \rightarrow R_5$ という経路は従来に比べて冗長になってしまう。

AS2 側で、AS1 の部分を指定してそれぞれ X, Y のいずれを使用するかという設定を行なうことは可能であるが、この場合、相手 AS の内部トポロジに関する情報が必要であり、エレガントな解決ではない。BGP-3 における INTER_AS Metric を活用することが考えられるが、現在の実装では INTER_AS Metric と IGP の関係は明確ではなく、直ちにこれを利用することはできない状態である。IETF の BGP-4 開発グループで、目下経路選択の問題について議論がなされており、片方の相互接続点のみを専ら使用する方法と、複数の接続点を効率的に使い分ける運用の両方の可能性について検討が行なわれている。

このような理由で、複数の接続点を持つ AS 間 — 例えば WIDE と TISN — は、まだ BGP への移行は行なわれていない。

4.3 現在の経路制御

現在は経路制御の移行中の段階であり、一部分は依然として RIP で運用が行われているが、多くの AS で BGP-3 の接続実験が行われていたり準備中である。1993 年 10 月時点での国内

表 1: BGP-3 を利用した経路制御の状況

AS 番号	AS 名	状態
2497	IIJ	接続実験中
2498	JOIN	計画中
2499	SINET	計画中
2500	WIDE	接続実験中
2501	TISN	計画中
2502	TRAIN	計画中
2503	TOPIC	接続実験中
2504	NCA5	接続実験中
2505	HEPNET-J	計画中
2506	CSI	接続実験中
2507	TSUKUBA-RIC	計画中
2508	KARRN	接続実験中

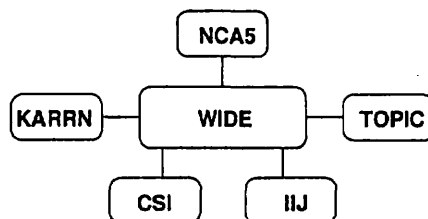


図 2: 接続実験中の AS のトポロジ

5 まとめと今後の方針

本論文では、RIP を用いてポリシを含む経路制御を実現する場合の問題点について述べた。また、RIP に比べて metric の自由度が大きい OSPF を利用する場合でも、ポリシに従った経路設定は可能であるが運用上の問題があることを示した。より正確な議論は今後さらに検討しなければならないが、EGP を導入して EGP・IGP の 2 段階の経路決定によって、ポリシに関するいくつかの問題が解決できることを述べた。

そして、WIDE インターネットや JPEG/IP が中心になって行なわれている RIP から OSPF

や BGP への移行の経過について報告し、現在の国内のインターネットの経路制御方式について紹介した。

今後の問題点として次のような事項を挙げる事ができる：

- 複数の接続点の使い分け
- 複数の AS の使い分け
- マルチホームネットワークに対する経路選定の問題
- 発信元ポリシーの導入に関連する問題点
- 経路の集成を行なう場合の問題点の検討
- 経路のポリシーの粒度の改善

これらはある程度複雑なトポロジのネットワークのもので問題となるもので、なかなか実験室のテストネットワークでは表現できない種類のものが多い。シミュレーションやあるいは実際のインターネットを利用して、検討・解決を行なっていきたい。

謝辞

本論文の内容は非常に多くの方々への協力なしには得られなかったものである。特に、JEPG/IP の諸氏、WIDE Project 関係各位の援助と助言に感謝する。

参考文献

- [1] C. Hedrick. Routing Information Protocol. RFC 1058, June 1988.
- [2] K. Yamamoto, A. Kato, M. Hirabaru, Y. Murayama, M. Nakamura, and N. Demizu. Autonomous System Partitioning and Policy Routing in the Japanese Internet. In *Proceedings of INET'93*, 1993.
- [3] J. Moy. OSPF version 2. RFC 1247, July 1991.
- [4] D. L. Mills. Exterior Gateway Protocol Formal Specification. RFC 904, April 1984.
- [5] K. Loughheed and Y. Rekhter. A Border Gateway Protocol 3 (BGP-3). RFC 1267, October 1991.
- [6] V. Fuller, T. Li, J. Yu, and K. Varadhan. Classless Inter-Domain Routing(CIDR): an Address Assignment and Aggregation Strategy. RFC 1519, September 1993.
- [7] Y. Rekhter and C. Topolcic. Exchanging Routing Information Across Provider Boundaries in the CIDR Environment. RFC 1520, September 1993.
- [8] E. Gerich. Guidelines for Management of IP Address Space. RFC 1466, May 1993.
- [9] Y. Rekhter and P. Gross. Application of the Border Gateway Protocol in the Internet. RFC 1268, October 1991.
- [10] K. Varadhan. BGP OSPF Interaction. RFC 1403, January 1993.
- [11] G. Malkin. RIP Version 2 Carrying Additional Information. RFC 1388, January 1993.

本図は 1993 年 10 月時点での日本のインターネットの相互接続の様子を JPEG/IP 経路制御グループの協力の元でまとめられたもので、相互接続に関連する部分のみ描かれている。

