第 6 部

WIDE Internet における経路制御

第 1 章

はじめに

WIDE Internet は主に OSPF を IGP として使用し、OSPF をサポートしていないルータは RIP あるいは static な経路制御を用いている。ほとんどの接続組織に対しては、RIP によって default 経路のみを送出し、またその組織のネットワークに対する経路情報を RIP で受信している。そのため、NOC 側ルータで RIP → OSPF ASE の変換を行っている。

当初、国内のインターネットの経路制御は単一の RIP によって行われていたが、経路数の増加とともに RIP パケットのオーバヘッドの増加および直径の増加による経路情報が届かない場所が増加してきた。さらに各ネットワークプロジェクト間の経路制御ではポリシを反映したものにすることが望まれてきたため、

- 経路制御ドメイン内部の経路制御プロトコルは、ドメインそれぞれが適当なものを使用する
- 経路制御ドメイン相互間の経路制御には BGP を用いる

ことが JEPG/IP で合意された。

WIDE Internet では、IGP を OSPF にするべく、version 3.0 の `gated` 等を用いて OSPF の運用を RIP による経路制御に並行して行い、徐々に WIDE Internet の基幹部分から OSPF による運用に切替えていった。それとともに、準備ができた隣接経路制御ドメインとの間の経路制御を RIP から BGP-3 に変更してきた。

以下、現在の WIDE Internet における経路制御の現状に関して報告する。

第 2 章

WIDE OSPF

WIDE Internet では、当初は全部を単一のエリア、すなわちバックボーンとして OSPF の運用を行ってきたが、単一エリア中のルータ数が増加すると次のような問題点が発生することが分かった:

- SPF の計算量が増加する
- SPF の計算を行うイベントの発生頻度が増加する

このため、ネットワークの発展とともにルータにおける処理が増大し、再計算の頻度が増加する結果になる。

そこで、全体をバックボーンとして運用するのではなく、適当にエリアに分割し、エリア相互を結合する部分をバックボーンとすることが考えられる。この場合、完全な SPF の計算はルータが所属しているエリア内部に限られる（複数のエリアに所属しているルータは、それぞれのエリアに対して SPF の計算が必要である）という利点がある。SPF の計算量は $O(nlogn)$ であるため、エリアを小さくするほど計算量も減少し、また、エリアに属しているルータ数も減少するため、経路が安定になるという利点がある。

経路制御ドメインをエリアに分割する場合の問題点として、全てのエリアはバックボーンエリアに隣接していなければならないという制約があることが挙げられる。言い替えれば、全てのエリアには一つ以上のバックボーンルータが存在していなければならない。バックボーンルータが含まれていないエリアが存在する場合には、仮想リンクによってバックボーンを拡張することで対処する。しかし、仮想リンク機能は必ず必要とされる機能ではないため、一般にバグのために正常に動作しない事が多く、WIDE Internet のように種類およびバージョンのバラエティに富んでいるネットワークの場合には避けるほうがよいと考えた。そのため、各エリアはバックボーンに必ず隣接するようになっている。

エリア分割のもう一つの利点は、各エリアからバックボーン方向へは経路のサマリをアナウンスすることが可能な点である。経路のサマリは、BGP-4 における経路の集成に似た概念であり、基本的にクラスレスである OSPF において、サブネット経路が存在する場合に、ナチュラルネット経路を生成するのに利用することができる。

WIDE Internet において、多くの組織との接続には、unnumbered なリンクが用いられているが、組織側のネットワークのサブネットが割り振られている場合も少なくない。こ

のような場合に経路をアナウンスする方法として、

1. WIDE 側のルータだけ、その point-to-point リンク上で OSPF を起動する。対抗側は OSPF を話さなくてよい。

2. すると、そのサブネット経路が Network LSA として、WIDE Internet に対してアナウンスされる。

3. サブネット経路だけでは困るので、point-to-point リンクはバックボーンではない別なエリアにする。

4. すると、エリア境界で Summary LSA を生成することができるので、ナチュラルネット経路を生成するように設定する。

という方法がある。この方法は、1. static なナチュラルネット経路を定義する。2. それを OSPF の External LSA としてアナウンスする。という方法に比べて、ルータの処理は増えるが、本来ドメイン外部の経路を表現するための External LSA を使わずに済むという利点がある。

以上のような点を考慮し、次のような方針で OSPF の運用を行っている。

共通事項:

1. 全てのエリアでは simple password による認証を行う。このパスワードをここに記すことはしないが、初期の OSPF の実装がなかなか動作しなかったことに由来するものが付けられている。gated では MD5 による認証がサポートされてきているが、他のルータでは最新版でサポートが始まったところであり、今後相互運用性を確認してから移行を検討する。

2. Router Dead Interval（この時間以上 Hello を受信しなかった場合には、相手ルータはダウンしていると判断してもよいタイムアウト値）は Ethernet 上では 30 ～ 40 sec, point-to-point リンク上では 40 ～ 60sec の範囲で各リンク毎に決める。当初は 30sec で運用していたが、Router Dead Interval を 20sec の整数倍にしか設定できないルータが存在するため、30sec から 40sec に変更したリンクもある。

3. 可能な限り、Summary LSA, Network LSA によって経路をアナウンスすることにし、External LSA は外部、すなわち WIDE Internet 以外の経路を示すようにする。また、External LSA のタグは、WIDE 組織の経路のものは 0、外部組織のものは、RFC1403 に従って生成されたものにする。ただし、外部組織のものとはいえ、BGP による経路制御が WIDE Internet との間で実施されていない NORTH および ORIONS の経路は WIDE 組織の経路に準じて取り扱う。

4. 原則として RIP および static 経路から得た WIDE 組織の経路、また BGP によって得られた外部の経路は、Type-1 External LSA としてアナウンスする。これは、WIDE

Internet に複数の地点で接続されている場合に、最寄りの接続点を経由するようにするためである。

5. その時点で一番信用できそうな実装のルータが指定ルータになるようにし、それ以外のルータは priority を 0 にして、指定ルータにならないようにする。

6. 各リンクのコストは OSPF の仕様で推奨されている値にすることを原則とする。すなわち、bps で表現した帯域に対して

$$cost = \frac{10^8}{bandwidth}$$

にする。また、External LSA を生成するときには、cost を 10000 とする。

現在 WIDE Internet において、OSPF による経路制御を行っているルータは、Cisco 10.0, Proteon 15.1, および `gated` 3.5α10 である。MAXAGE を越えた LSA の取り扱いに関して若干の実装の食い違いが見られるが、当初よりはずっと安定して動作するようになった。特に `gated` は 3.5α10 から OSPF に関して数多くのバグが修正されており、安定に動作するようになってきている。

第 3 章

BGP

WIDE Internet とその他のプロバイダとの経路制御はほとんどが RIP から BGP への移行が完了している。

ほとんどのプロバイダとの接続が BGP-3 になっているのは、単に WIDE Internet が完全に Classless になっていないからである。Backbone 部分で RIP に頼っている部分はすでに無くなり、OSPF に移行しているが、パケットの転送が Classless になっていない部分が多い。そのため、BGP-4 で集成された経路のアナウンスを受信した場合に、集成されていない経路も同時に受信しなければループが発生する可能性があるからである。ルータの経路情報交換という観点では、BGP を運用しているルータは全て既に BGP-4 をサポートしている。

BGP-3 を BGP-4 に変更するのは、単に BGP-3 を指定している設定を BGP-4 に変更するだけでよく、大きな設定変更は必要ない。事実、Cisco ルータ相互間を接続する IBGP セッションには BGP-4 が用いられており、一部の `gated` の設定だけの問題である。

現在、すでに実質的に IBGP セッションを維持する必要が無くなってしまったルータも含めて、13 個のルータが完全グラフ状に IBGP セッションを設定している。この 13 個という数字は決して少ない数ではないが、IBGP がもはや管理できないという程の数字でもない。IETF の IDR WG では、完全グラフ状の BGP セッションの問題を軽減するため、Route Server が提案されている。これは、単に BGP セッションの集線を行うもので、経路に対する積極的な解釈を行うものではない。Next Hop の計算に必要な BGP セッションの相手のアドレスが変わってしまうため、Route Server はその情報を新たに定義した属性によって伝達するため、現在の BGP-4 の実装そのままで運用することには大きな問題が発生すると考えられる。

BGP で学習した経路は OSPF External LSA に変換するが、この際、Metric Type は type-1 としている。これは、東京および京都で相互に接続されているプロバイダが存在し（TISN・IIJ）、type-2 を指定したとすると、近傍の接続点を利用するというポリシが実装できなくなってしまうためである。ただし、このことは経路が行きと帰りで非対象になってしまうことが多いが、行きあるいは帰りの片方の経路に障害が発生し、他方が正常な場合、正常な経路を往復とも利用するようになる（障害発生時からしばらくは不通になるが）ので、大きな問題にはならないと考えられる。

この方法の一つの問題点は、相手プロバイダの経路一つに対して、それぞれの接続点について OSPF LSA を生成する必要があり、経路数に比べて LSA 数が増えてしまうことである。現在の経路数は 2500 程度であるが、LSA 数は 4000 を越えており、ルータのメモリおよび CPU の消費、また、ルータが boot したときに定常状態になるまでの時間が掛かる、などの影響がある。

表 3.1: WIDE Internet の隣接プロバイダ

AS 番号	プロバイダ	プロトコル	接続地点
297	NSI	BGP-4	San Francisco
2497	IIJ	BGP-3	東京（大手町、NSPIXP）・京都
2498	JOIN	BGP-3	東京（一ツ橋）
2501	TISN	BGP-3	東京（弥生、大手町）・京都
2502	TRAIN	BGP-3	東京（弥生）
2503	TOPIC	BGP-3	仙台
2504	NCA5	BGP-3	京都
2506	CSI	BGP-4	広島
2508	KARRN	BGP-3	福岡（箱崎）
2510	InfoWeb	BGP-3	東京（NSPIXP）
2511	NTT Core	BGP-3	藤沢
2513	IMnet	BGP-4	東京（大手町）・奈良
2515	JPNIC	BGP-3	東京（弥生）
2518	MESH	BGP-3	東京（NSPIXP）
2519	NIS	BGP-3	東京（NSPIXP）
2520	ORIONS	RIP	大阪（阪大）
2521	TokyoNet	BGP-3	東京（NSPIXP）
2527	SinfoNY	BGP-3	東京（NSPIXP）
2907	SINET	BGP-3	東京（一ツ橋）
2915	SPIN	BGP-3	東京（NSPIXP）
3510	RWC	BGP-3	東京（大手町）
3561	InternetMCI	BGP-4	San Francisco
—	NORTH	RIP	札幌

第 4 章

経路数の問題

4.1 IP アドレスの枯渇

インターネットのネットワーク層が抱える問題として、アドレスの枯渇問題と経路表の大きさの問題が知られている。前者は IP アドレス空間のうち、Class A、Class B それぞれが半分程度割り当てられた 1992 年頃から問題になってきており、当座の解決策として、Class B の割り当てを抑制し、それに代えて複数の Class C アドレスの割り当てを行うようになった。

無論、急激にかつ世界的に広がっているインターネットをサポートするための究極的な解決は、32bit より広いアドレス空間を持つ次世代 IP を開発し、それに移行することであるが、そのためには相当な時間が必要である。そのため、当時 1% しか使われていなかった Class C アドレス空間を活用することにし、その次には Class A 空間の半分（実はこれだけでも Class B 空間全体、あるいは Class C 空間の倍のアドレスが収容できる）を割り当てることが想定された。

複数の Class C アドレスを割り当てる際、ばらばらに割り当てた場合、インターネット上で交換される経路情報数は爆発的に増大することが予想された。そのため、CIDR — Classless Inter-Domain Routing — という概念が導入され、２の冪乗個の連続する複数のネットワークアドレスを単一の経路で記述できるような枠組が定義された。そして、CIDR をサポートするために BGP-4 が開発され、また、基本的に Classless であった OSPF も若干の修正が行われた。

そして Class C アドレスに対する経路の集成を効率的に行うため、あるいは NIC におけるアドレス割り当ての負荷の軽減のため、一連のブロックアドレスをプロバイダに割り振り、顧客に対して必要量を割り当てることも行われるようになった。

4.2 経路表の増大

インターネット上の経路は 1987 年当初では 1,000 程度であったが、その後の特に米国におけるインターネットの普及によって経路数は次第に増加し、1993 年には 10,000 を越えるようになった。そのため、仮想記憶をあまり持たないルータでは、BGP や OSPF な

どの経路制御プロトコル毎のデータベースやパケット転送に用いられる経路表の大きさが増加し、通常実装されているルータの実メモリの限界に近付きつつあった。

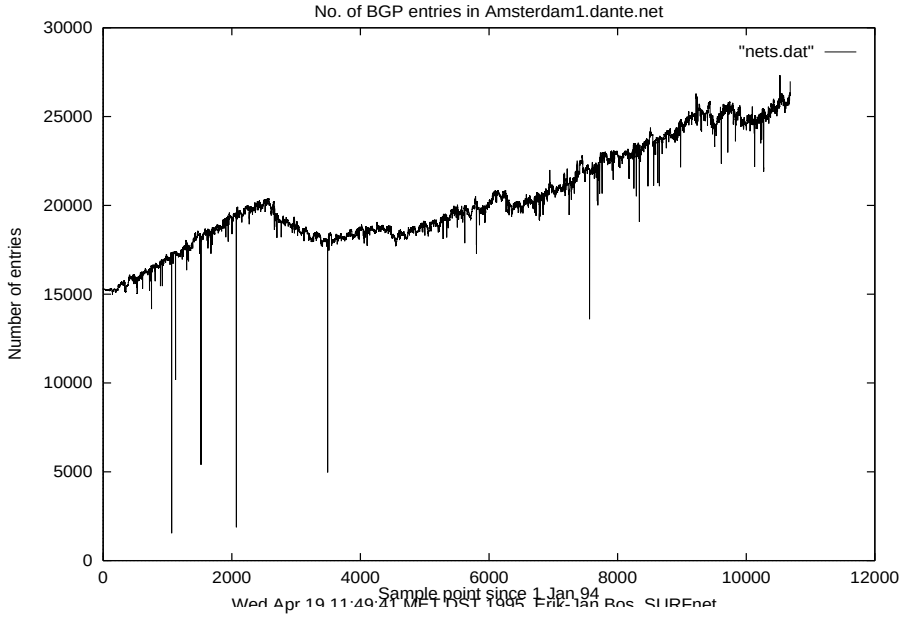

図 4.1: インターネットの経路数の増加（surfnet.nl） による

そのため、1993 年から BGP-4 の開発および実装がルータベンダを中心になって進められ、トンネルを利用した BGP-4 の運用実験も行われていた。そして、1994 年 3 月の Seattle IETF において、以下の事柄が申し合わされた:

- 速やかに BGP-4 に移行すること
- キャンパス単位での経路の集成（aggregate）を行うこと
- BGP-3 あるいは EGP を使用する場合には、default 経路を用いること

キャンパス単位での経路集成しか実施されなかったのは、経路制御上のポリシの問題に起因している。すなわち、一般のプロバイダの空間は、NSFNET AUP に合意した組織と合意していない組織が混在している場合が多く、その場合には経路の集成ができないためである。

また、RIPE の Tony Bates 氏によって週間ランキング（経路数を増やした “悪い” プロバイダ上位 20 位と減らした “良い” プロバイダ上位 20 位がリストアップされている）がメーリングリストで公表するという努力によって経路の集成は比較的スムーズに進み、ルータのメモリ問題は一旦は落ち着いた。図 4.1 において、サンプル点 2,500 のあたりが Seattle IETF であり、20,000 を越えていた経路が 18,000 程度に下がったことが確認できる。

しかし、しばらくするとインターネットの拡大により経路数が増加し、その度にメーリングリストに緊急の経路集成要請がアナウンスされたりして、依然として状況は楽になってはいない。ただ、NSFNET が 1995 年 4 月に終了して AUP を深刻に考える必要がなくなったことにより、プロバイダ単位での経路の集成が進んで行くと思われる。

また、インターネット上のセキュリティ問題は、アドレスおよび経路制御の問題にも大きく影響している。つまり、一般の企業がインターネットに接続する際には、セキュリティ問題からごく限られた数のホストしかインターネットと直接アクセスができないようにする、いわゆる防火壁を構築する場合が多い。この場合、企業内部のネットワークは RFC1597 で定義されている閉じたインターネットのためのアドレスを使用し、防火壁の部分にはプロバイダから割り当てられた Class C アドレスを一つ使用するというのが典型的な姿になってきている。このため、アドレス消費および経路の集成効率が高まっており、おそらくこのような背景によるものと思われるが、アドレス枯渇の予測も従来の 2008 ± 5 年から 2013 ± 8 年に修正されている。

4.3 国内のインターネット

国内のインターネットの経路数に関しては、古い部分に関しては統計が残っていないので接続年月日から推定するしかないが、1994 年夏からの経路の変動は図 4.2 に示す通りである。このデータは `endo.wide.ad.jp` における経路数を一日４回カウントし、その最大数をプロットしたものである。ただし、1994 年 10 月はそのマシンが crash し、代替マシンで運用されていたため、データが得られていない。

図 4.2 は `netstat -r -n` の出力を、WIDE Internet のサブネット経路などを除いてカウントしたものである。Hash 方式の経路表の場合、経路を変更している最中に `netstat -r -n` を実行すると、しばしば正常な出力が得られないことがある。そのため、時々経路数が異常に増加している部分があるが、それは無視して差し支えない。

この図によると、1995 年 5 月末時点の経路数は 2,500 を越えており、毎月の増加率は 100 ぐらいである。そのため、1995 年秋には 3,000 経路に到達することが予想される。

このような経路をプロバイダ単位の集成を可能な限り行った場合の経路数は、約 1,000 であり、半分以下になる。さらに今後は多くの、特に商用プロバイダは JPNIC からアドレスブロックの割り振りを受ける傾向にあるので、経路集成の効率は高まることが予想される。

現在の経路数約 2,500 は、ルータのメモリおよび CPU 能力を考えると、直ちに経路の集成を行わなければならないという程ではない。しかし、経路数は毎月のように増加しており、いずれ経路制御プロトコル上の問題が発生するか、オーバヘッドが増加して、正常な運用に絶えられなくなる。そのため、早期に経路情報の集成を行い、経路数および経路数の増加を抑えることが必要になる。

前述のように、国内のプロバイダ間の経路制御はほとんどが BGP に移行しており、新

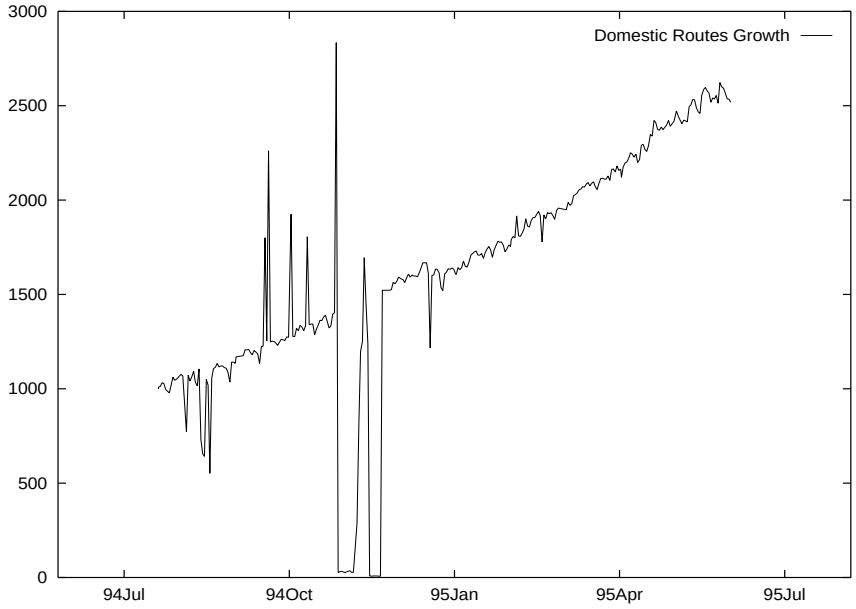

図 4.2: 国内の経路数の増加

規に運用を開始した商用インターネットではすでに BGP-4 を使用しているものも少なくない。ところが、旧来から運用を行っている学術系プロバイダでは、BGP-4 をサポートするためにはルータの更新や少なくともソフトウェアのバージョンアップが必要であったり、classless な経路制御を実施するための開発が必要である。そのため、特に商用プロバイダ側から早期に経路の集成を行うように求める声が高まっているが、まだ実現はされていない。ただし、国際的には経路の集成は必須であるため、国際ゲートウェイでは概ね経路の集成が実施されている。

4.4 WIDE Internet

WIDE Internet における経路数の問題に関するアプローチは３つの努力が行われている。一つは国際ゲートウェイにおける経路の集成、二つ目は WIDE Internet に対してアドレスブロックを JPNIC から割り当ててもらい、それを利用するように接続組織にアドレス変更をお願いすることである。三つ目は、Sun Workstation を用いたルータはそのままでは classless routing は実施できないが、それを可能にするような修正を行うことである。

4.4.1 代理集成

最初の国際ゲートウェイのおける経路の修正は 1994 年秋から、NSI の要請によってスタートした。具体的には、８つ以上の連続する Class C アドレスを有する組織に対して、

国際ゲートウェイで代理集成（Proxy Aggregate）を実施した。さらに 1995 年 1 月に、集成可能な経路は全て集成を行うように設定を行った。

1995 年 5 月時点で、WIDE から MCI および NSI にアナウンスされている経路数は 257 であり、もし集成が行われないとした場合に、全ての経路がアナウンスされていると仮定した場合の経路数は、約 1,000 となる。現在アナウンスされている経路のうち、代理集成されている経路は 97 である。

4.4.2 Delegate されたアドレス空間

２つ目は WIDE Internet の集成後の経路総数を減らすことである。そのため、JPNIC から 202.249.0.0/16 のアドレスブロックを 1994 年 5 月に取得し、それを利用することにした。具体的には、新規に WIDE Internet に接続してきた組織および既存の組織で防火壁を設けるか、あるいは会社全体のインターネット接続性は商用ネットワークに依存し、研究部門の一部だけ WIDE を利用するような場合で、それらのホストが単一の Class C で十分カバーできる場合に、従来のアドレスからの移行をお願いすることである。

WIDE Internet における予想で、２年間で必要とされるアドレス数は 40 程度であると推定されており、202.249.0.0/18 で十分であったが、JPNIC では当時は 256 個単位での delegation を行っていたため、不要分は後日返却することとした。

現在、11 組織が 202.249.0.0/16 の下のアドレスを用いるように移行を行なっており、さらに 14 組織にアドレスが割り当てられている。実際の移行は DNS の登録や NOC 側との協調作業が必要なため、

4.4.3 100 校プロジェクト

その後、通商産業省のいわゆる「100 校プロジェクト」が実施されたことにより、これらの小中高等学校へのアドレス割り振りが問題になった。そこで、JPNIC と協議し、主に 202.249.64.0/18 および 202.249.128.0/18 の空間を、国際的に考えた場合の経路数を最小にするように割り当てを行うことになった。各接続地点は地域ネットワークに依頼されているが、その単位での集成も可能にすること、接続学校数にも若干の増加が考えられるので、割り当て効率を高めることよりも、経路数を抑えることを主眼においた。その結果、一部の学校は上記の空間に収まらず、WIDE Internet に直接接続されている学校を中心に 202.249.0.0/18 からのアドレス割り当てを行っている。

この結果、「100 校プロジェクト」全体の経路を国際的に評価した場合、４経路ですむことになり、また、202.249.0.0/17 は基本的に国際回線は WIDE Internet 経由であるため、この部分の集成を行った場合、３経路で済むことになる。

4.4.4　Classless 化

WIDE Internet では、ベンダで標準的にサポートされていない機能に対する実験や測定を可能にするため、いわゆる専用ルータとともにワークステーションに高速シリアルインターフェースを追加したものをルータとして用いてきた。現在は主に Sun SparcStation-2 に HSI/S を挿したものが多く利用されている。このため、IP Multicast や NTP などの機能をルータに持たせることができており、WIDE Project の研究を行う上でこのことは重要である。

しかしながら、SunOS 4.1.x のネットワーク部分は概ね 4.3BSD Tahoe リリースに準拠しており、いわゆる classless な経路制御は実装されておらず、ハッシュを利用した経路制御アルゴリズムが実装されている。これは WIDE Internet の経路制御を classless にする上で大きな障害になっているが、その他にも、経路数の増大に伴うハッシュの効率低下が問題になってきている。

1995 年春の国内の経路数は約 2,500 であるが、これを標準では８つのハッシュバケツを用いてハッシュしている。そのため、平均的にハッシュされた場合、一つのハッシュバケツには 300 経路が格納されることになり、経路が存在する場合で平均 150 回、経路が存在しない場合には 300 回程度の経路の比較が必要になり、OS のオーバヘッドが大きくなっている。木状の経路表を利用した場合、実装方針や経路表の状態にも依存するが、比較回数は最悪 IP アドレスの bit 長である 32 回であり、多くの場合、この半分程で済むのではないかと考えれている。比較１回あたりのオーバヘッドが異なるため、簡単には比較は難しいが、最悪値が経路数に依存しないことは、さらに発展が見込まれているインターネット上の技術としては重要な要素になっている。

これに対しては、4.3BSD Reno リリースのネットワークコードを導入する案、SunOS ではなく 4.4BSD にする案も検討された。Reno リリースの経路表検索アルゴリズムはアドレス長が制限されていない点や連続性なネットワークマスクを扱うことができる点で優れているが、アルゴリズムが難解であり、また 4.3BSD Reno リリースでは Routing Socket の導入や mbuf の変更など変更点が多いことから、OS ソースコード無しでこれらのコードの導入を行うのは困難であることが予想された。また、4.4BSD は今後の IPv6 などのサポートを考えた場合もっとも望ましい方式であるが、高速シリアルカードのドライバを作成しなければならず、十分なデータがないため、現実的ではなかった。

結局、SunOS に簡単にソースコードなしに導入でき、当面 IPv4 のしかもネットワークマスクは連続なものに限るという制約があったも、効率的で簡潔なコードが望ましいとされ、そのため、二分木に基づいた経路の検索および管理プログラムが作成された。後日、これは Trie と呼ばれている構造であることが判明したが、現在 SunOS への実装および評価が行われている。

このコードが稼働し、`gated` のカーネルとのインターフェースにネットワークマスク情報を渡す機構を追加すると、`gated` はすでに BGP-4、OSPF、RIP-2 などの Classless 経路制御プロトコルをサポートしているので、CIDR に対応した経路制御か可能になる。

第 5 章

経路の登録

アドレスの割り当ては NIC（APNIC, JPNIC を含む）あるいはそれらから委託されたプロバイダによって行われており、必要な情報はそれぞれのデータベースに登録され、whois などのコマンドで参照することができる。

割り当てられたネットワークに対しては、基本的には経路情報をアナウンスすることで到達性が得られるが、ポリシなどの問題によって単純に隣接 AS からの経路を採用するのではなく、フィルタあるいは優先順位を設定している場合がある。このために必要なデータベースを Internet Routing Registry (IRR) と呼ぶ。

5.1 PRDB

RR でもっとも有名なものは、1995 年 5 月で運用を終了したが、NSFNET の PRDB で、Michian 大学に併設された Merit Inc. によって設計・運用が行われていた。基本的には NACR と呼ばれるフォーマットがあり、NSFNET（AS690）の直接の隣接プロバイダから、申請を行うことになっていた。WIDE Internet では、NSFNET への通信は NASA Science Internet を経由していたため、NSI に対して NACR を送り、Merit に申請をしていた。海外（米国からみて）の経路は NSF の承認を必要としていた時代もあるが、あまりに繁雑なので、一部の国に対してのみ承認手続きを残し、その他の国はほぼ事務的に作業が行われていた。

NSFNET は NSFNET AUP が定義されており、それに従って利用することが必要であった。そのため基本的には、NSFNET AUP を遵守する旨の誓約書に署名することが必要であったが、WIDE Internet の場合、NSFNET とほぼ同等な AUP があり、全ての通信はこれに従うことになっていたので、ネットワーク単位での署名は不要であった。商用ネットワーク経由のアクセスに関しては、署名を原則として必要としていた。

Merit では送られてきた NACR に対して、種々のチェックをした後、週２回、具体的には月曜日と木曜日の早朝、データベースに登録を行い、その内容をもとにそれぞれの ENSS と呼ばれるルータの設定ファイルを生成し、各ルータに送って、ルータの経路制御プロセスを再起動していた。ルータの経路制御プロセスの再起動に関して、経路情報が途切れることがあるため、影響の少ない早朝が選ばれていたが、日本時間では午後９時前後であり、

```
%begin nsfnet nacr v7.1

netnum:          133.4.0.0/16
netname:         WIDE-BB
netcc:           JP
orgname:         WIDE Project
orgaddr:         5322 Endo
orgcity:         Fujisawa
orgstate:        Kanagawa
orgzip:          252
orgcc:           JP
orgtype:         N
bbone:           T3
homeas:          2500
aslist:          372  297
aup:             N
action:          A
comment:

%end nsfnet nacr
```

図 5.1: 133.4.0.0/16 に対する NACR

この時間帯には海外への到達性が不安定になることがあった。

NACR で指定するポリシは aslist フィールドによって指定できる。つまり、AS690 に対してその経路をアナウンスする隣接 AS 番号と、必要に応じて経路を受理する ENSS 番号を付加したもののリストを指定する。NSI の AS 番号は、西海岸では 372、東海岸では 297 を用いていたため、WIDE Internet の経路に対しては、

```
aslist:     1:372   2:297
```

と記されていた。これは、その経路は優先順位一番が AS372、二番が AS297 であることを意味し、その他の AS からアナウンスされていても受理されなかった。

1994 年秋に NSI の AS 番号は 297 に統一され、その結果、AS690 は AS297 と二箇所で接続されることになった。この場合、ENSS 番号を ( ) に付記し、優先順位を指定した:

```
aslist:     1:297(144)   2:297(147)
```

1995 年 1 月に WIDE Internet が InternetMCI 経由でのアクセスを開始したのに伴い、

AS690 に対しては、InternetMCI (AS3561) を優先し、NSI をバックアップとしたため、aslist は次のように長くなっている:

```
aslist: 1:3561(11) 2:3561(144) 3:3561(147) 4:3561(218) 5:297(144) 6:297(147)
```

Merit の PRDB はこのように経路がアナウンスされる隣接 AS を制限するフィルタと、複数の隣接 AS から経路がアナウンスされている場合にどちらを優先するかということが記述されていた。そのため、隣接 AS が NSI のように大規模であった場合（日本の WIDE Internet, TISN, IMnet を始め、オーストラリアやニューランドなどを含む)、その設定ファイルは巨大になる。この設定ファイルは基本的には Gated の設定ファイル `gated.conf` であり、NSI に隣接する ENSS の `gated.conf` は 3MB を越える巨大なものになっていた。

5.2 RIPE-181

NSFNET の終了が 1995 年 4 月に予定されており、その後継としては、各プロバイダの接続地点 NAP — Network Access Point – を NSF が支援することになり、３箇所の Priority NAP が、New York, Chichago, San Francisco に設定されることになった。後日、Washington D.C. も追加されたが、各 NAP における経路制御の問題に関して、Routing Arbiter が計画され、その開発は IBM, Merit, ISI などによって行われることになった。これに対するデータベースはヨーロッパの RIPE/NCC で運用されいた RIPE-181（RIPE-81+とも呼ぶ）に基づいたものに変更され、`ra.net` によって運用されることになった。この IRR は RADB と呼ばれている。

移行期の混乱を避けるため、PRDB のデータは逐次 RADB に反映されてきていた。現在は PRDB は完全に運用を終了したため、経路の登録は RIPE-181 に基づいた形式で依頼しなければならない。ただし、従来は AS690 の隣接 AS の管理者から申請することになっていたが、RADB の場合、AS690 の隣接 AS あるいはその経路の homeas の管理者のいずれからでもよくなった。また、RIPE-181 では AS690 の経路設定に十分な情報が含まれていなかったため、若干の拡張が行われている。

現在、RIPE-181 に基づいた IRR を 運用しているのは、`ripe.net`, `ra.net`, `ca.net`, および `mci.net` である。そのほか、APNIC でも運用を行うことが計画されており、また IMnet の研究の一環として実装するという話もある。

RIPE-181 の特徴は、Maintainer、AS、Route という３つのオブジェクトによって管理される点である。AS オブジェクトは各 AS の状態およびポリシを記述したもの、Route オブジェクトは各経路に関する記述であり、Maintainer オブジェクトは AS や Route オブジェクトが誰によって管理されているかを記述するものである。

データベースにオブジェクトが登録される際には、必ず対応する Maintainer オブジェクトが参照される。データベースの登録には電子メールが用いられる（データベースの登録プロトコルは RFC822 である、と表現されることもある）が、Maintainer オブジェクトは

```
mntner:      WIDE-MAINT-MCI
descr:       WIDE maintainer object
admin-c:     Jun Murai
tech-c:      Akira Kato
upd-to:      kato@wide.ad.jp
auth:        MAIL-FROM Akira Kato <kato@nezu.wide.ad.jp>
auth:        CRYPT-PW 76poJR4XPkrm6
notify:      kato@wide.ad.jp
mnt-by:      WIDE-MAINT-MCI
changed:     mci-rr@mci.net 950326
source:      MCI
```

図 5.2: MCI IRR における WIDE の Maintainer オブジェクト

その電子メールの From: フィールドの値の集合を記述し、誰からのメールならばその変更を受理して差し支えないことが示されている。さらに、crypt されたパスワードを登録しておき、更新メール本文の先頭に

password: *plain-password*

を指定することにより、From: に加えてチェックが行われる。

現在、WIDE Internet において、新たなネットワークが接続された場合、MCI の IRR にその情報を登録する必要がある他、AS690 を運用している ANS も通過できるように RADB にも登録を行う必要がある。両者の登録は基本的な同じ形式であるが、MAINTAINER の名前の付け方が異なっているため、全く同じデータを送れば良いというわけではない。それぞれの形式に従ってデータを用意し、登録を行う必要がある。

ところで、複数の IRR へ登録を行う場合、手続きが繁雑である以上に、IRR 相互間でデータの一貫性を維持することは困難である。そのため ANS では、MCI や CA*NET、RADB のいずれかの IRR に登録されていればよいようにする予定である。この場合、MCI IRR のみへの登録で済み、事務手数量が減少する。

ところで、これらの IRR への登録は、予め登録された MAINTAINER が電子メールで申請書を送った場合、自動的に文法誤りなどがチェックされ、重要なものに関しては再提出が必要になるが、マイナーな問題の場合には、訂正して登録してくれる。たとえば、updt: にはそのオブジェクトが更新された日付を記入することが必要である。ところが、日付変更線があるため、アメリカではまだ前日である場合があり、日本時間で日付を記入すると、エラーになる。このエラーは深刻なものではないため、警告メッセージが送付されるが、受理される。通常、電子メールで情報を送ってから、それに対する返答があるまでの所用時間は数分であり、データベースの登録は即時に行われる。一方ルータの設定に対する実

装は、翌日早朝（現地時間）に行われる。

これらのデータベースの内容は whois コマンドで参照することができる。whois コマンドでしているホスト名（-h）は、RADB に関しては whois.ra.net、MCI IRR は whois.mci.net である。

```
aut-num:     AS2500
as-name:     ASN-WIDE
descr:       WIDE Internet AS
as-in:       from AS297 100 accept ANY
as-in:       from AS2497 100 accept AS2497 AS2514 AS2518
as-in:       from AS2498 100 accept ANY
as-in:       from AS2501 100 accept AS2501
as-in:       from AS2502 100 accept AS2502
as-in:       from AS2503 100 accept AS2503
as-in:       from AS2504 100 accept AS2504
as-in:       from AS2506 100 accept AS2506
as-in:       from AS2508 100 accept AS2508
as-in:       from AS2510 100 accept AS2510
as-in:       from AS2511 100 accept AS2511
as-in:       from AS2513 100 accept AS2513 AS2522
as-in:       from AS2515 100 accept AS2515
as-in:       from AS2518 100 accept AS2518
as-in:       from AS2519 100 accept AS2519
as-in:       from AS2520 100 accept AS2520
as-in:       from AS2521 100 accept AS2521
as-in:       from AS2523 100 accept AS2523
as-in:       from AS2907 200 accept AS2907 AS2505 AS2517 AS3510
as-out:      to AS297 announce AS-WIDE
as-out:      to AS3561 announce AS-WIDE
default:     AS3561 100 static
default:     AS297 200 static
guardian:    as2500@wide.ad.jp
admin-c:     JM292
tech-c:      AK3
mnt-by:      WIDE-MAINT-MCI
changed:     kato@wide.ad.jp 950328
source:      MCI
```

図 5.3: WIDE AS オブジェクト

```
route:        133.5.0.0/16
descr:        Kyushu University
origin:       AS2500
advisory:     AS690 1:3561(11) 2:3561(218) 3:297(144) 4:297(147)
mnt-by:       WIDE-MAINT-MCI
changed:      kato@wide.ad.jp  950329
source:       MCI
```

図 5.4: 九州大学の Route オブジェクト

第 6 章

おわりに

アドレスおよび経路制御の分野は、アプリケーションに比べると華々しさはないが、インターネットを支える重要な部分である。しかも、現実にインターネットを動かさなくてはならない要求と、新たな接続による拡大の両面からの対応が必要であり、インターネットの今後を決定する一つの要因になる部分である。

特に Classless 化に対しては WIDE Internet は他のプロバイダに遅れており、対策を早急に行なうとともに、その経験から IPv6 の実験にも取り組んで行きたいと考えている。