

RIPE Routing Working Groupによる経路アグリゲーションの推奨

Philip Smith
Rob Evans
Mike Hughes
(訳: 川村 聖一)

Document ID: ripe-399
Date: December 2006

要約

この文書では、今日のインターネットでのプレフィックスアグリゲーションの必要性を説明し、インターネットサービスプロバイダやインターネットに接続しているその他自律ネットワークが利用できる運用手法を推奨します。

1. 序章

インターネットは複数の自律ネットワーク(一般的にASまたはAutonomous Systemと呼ばれる)の集合体から成り立っています。これら自律ネットワークは、自身のネットワークで利用する目的、また顧客に割り当てる目的として、IPアドレス空間の割り振り/割り当てを受けることとなります。通常このIPアドレス空間は近隣のASに広告されます。ビジネス的な理由、または契約が原因で、このアドレス空間を近隣ASに広告したり、しなかったり、という事もあります。

インターネットを構成する団体が広告する経路の集合体を、インターネットルーティングテーブルと呼びます。

各ASがアドレス空間を広告し合い、経路を直接または間接的に学習する事により、インターネット上の各エンドシステムは互いに通信しあう事ができる。すなわち、グローバルな通信を可能にするInternetというシステムを形成します。

2. 背景

CIDRレポート[1]や同類の活動でも記載されているように、インターネットルーティングテーブルの経路数は、インターネットが90年代初頭に通信媒体として活用されだして以来インターネットサービスプロバイダ(以下ISP)やインターネットルーティング機器ベンダにとって常に中心的な話題のひとつとなっています。

2.1 インターネット初期

インターネット初期の頃、ASやエンドサイトへのアドレス割り当ては3つのカテゴリ: クラスA、クラスB、クラスCから成り立っていました。インターネットルーティングテーブルはこのたった3種類のアドレスから成り立っていました。小規模はクラスC、中規模はクラスB、大規模はクラスA。クラスフルアドレス割り当てと呼ばれ、ルーティングシステムはクラスを区別していました。

1994年、インターネットは大きな転機を向かえ、クラスフルプレフィックスからクラスレスプレフィックスへの移行が始まりました([2]、[3])。クラスフルアドレス空間が急速に枯渇しはじめた事がその理由で、特にクラスB空間(128.0.0.0から191.255.0.0)は切羽詰った状況となりました。

クラスレスへの移行の前のインターネットが一般普及し始めた頃、1つのクラスCアドレス空間を使い果たした組織は本来クラスBへのアップグレードされるはずが、クラスBの枯渇ペースを軽減するために、クラスCを複数持つという措置がとられました。

この結末として、多数の組織がクラスCブロック(今で言う/24プレフィックス)を保有する事になりました。クラスレスルーティングへの移行の一環として、CIDR Reportの元々のモチベーションは、連続的な/24を保有するネットワークオペレータに、経路の集約を促すものでした。これをアグリゲーションと呼び、後半詳細に解説します。

CIDR Reportは、アグリゲーションの推進には有効でした。オペレータ向けメーリングリストに毎週レポートが投稿され、アグリゲーション努力をしているISPを手助けしました

。この「周囲の目」戦略の効果は、CIDR Reportのサイトでグラフとして公開されています([1]、[4])。

2.2 今日のインターネット

今日のクラスレスインターネットでは、Regional Internet Registry(RIR)の会員である組織は、RIR(Afrinic、APNIC、ARIN、LACNIC、RIPE NCC)から割り振りを受けます。この割り振りサイズは、RIRへの要求サイズと、ネットワークオペレータの要求事項を元に決定され、一般的に最低プレフィックス長(例えば/21)に近いサイズが割り振られます。

1994年にクラスレスアドレスとルーティングが導入されて以降、ネットワークオペレータは、RIRから取得したアドレスブロックを、そのままインターネットに広告する事が期待されました。これはインターネット初期にクラスA、B、Cをそのまま広告していたのと同様に行われる想定でした。

ネットワークオペレータは、RIRから割り振られたアドレスをそのままインターネットに広告する、という暗黙の了解があるものの、/24を広告する事がいまだに慣例となっています(レガシークラスCと同様に)。インターネットルーティングテーブルの実に60%は/24のプレフィックスです。

/24広告の一部は、間違いなくマルチホーミングを目的としたトラフィックエンジニアリングの結果です(RFC1519-[3]の作者の視点では、業界の要求事項として定義されている)。しかしほとんどは、「RIRから32個クラスCを取得した」という思い過ぎのISPがそのまま経路広告している結果です(本来は/19に集約して広告する事が期待される)。

3. アグリゲーションとは何か

アグリゲーションとは、連続した複数のIPアドレスを、単一のブロックとして集約しIPルーティングシステムへ注入(BGPを使用して)する事である。例えば、とある企業がLANで利用する目的でISPから32個アドレスを取得した場合、ISPに向けて一つの経路として広告する事です。LAN上の各装置は、一つの経路として代表して広告されるため、個別の経路広告は必要ありません。

実際にアドレスを使った例ですが、企業が192.0.2.0から192.0.2.31を取得した場合、上流ISPに向けては、192.0.2.0/27として広告する事になります。この形式では、先頭27ビットがネットワークアドレスを意味し、後半5ビットがホストアドレスであることを明示しています。

より大きなネットワークでも同様に、ISPが4096個の連続したIPアドレスをRIRから取得した場合(例えば10.201.48.0から10.201.63.255)、10.201.48.0/20の単一の経路として近隣ネットワークへ広告する事になります。

上記2例はアグリゲーションの例です。末端のネットワークで、連続したアドレスを一つの経路として集約し、近隣のASに広告しています。

この例は、比較的小さなスケールを例にとっていますが、インターネットに参加するISPが、ネットワーク内の各アドレスを集約し、一つのブロックとしてまとめてインターネットに広告する行為を代表した例です。

ISPが、例えば他ネットワークから(通常BGPで)受けた経路広告で、連続しているブロックをより大きな経路として集約し、自身のネットワークをオリジンとした経路として広告しなおす事を、プロキシアグリゲーションとよびます。

4. インターネットルーティングテーブル

インターネットルーティングテーブルの現在の経路数には、いくつかの原因があり、以下に一つずつ分析します。

4.1 デアグリゲーションとは?

RIRがISPへアドレスを割り振る際に、そのまま変更せずにインターネットへ経路広告する事を期待して割り振りを行っています。ここで特筆すべきは、割り振るアドレスをどうインターネットに広告すべきか、ということに関してRIRは何もルールを設けていない事です。ISPに対してどのようにアドレスを広告すべきか、というのをRIRが指示するのは正しくない、というのが業界通念のようです。図書館で本を借りるときに、その本を

「どのように読めばよいか」を図書委員に指摘されるのは嫌う事と同じ考え方です。

しかし、ほとんどのISPは、割り振られたアドレスを一つのブロックとして広告せずに、時には/24まで分割して広告する事を好むようです。これをデアグリゲーションと呼びます。

4.2 一般的なデアグリゲーション

デアグリゲーションを行う理由はいくつかあります。プロバイダによっては、商業的な事情を理由とする人もいます。セキュリティを理由とするプロバイダもいれば、ウィルスや悪意あるトラフィックを排除し、トラフィックを削減できるとするプロバイダもいます。

ルーティングセキュリティは、インターネットに接続しているどのプロバイダにとって大きな関心事です。世界中を探しても、プロバイダが保持しているアドレスの経路広告を正式に許可するシステムというのは存在していません。(インターネットルーティングレジストリは、これに近いデザインですが、利用は必須ではないし、無くてもルーティングはそれなりに動いているという事実があります)。その結果、最小単位に経路を分割して広告する事でルーティングセキュリティシステムの欠如を回避しようとするプロバイダがでてきます。これにより、より長いプレフィックスのアドレスを他者が広告する事により発生する、正当なアドレス保持者ネットワークへのDoSを防ぐ事ができます。しかし、筆者の知る限りでは、このような事象は非常に稀であり、デアグリゲーションのリスクを考慮するといささか大げさな行為であるように思われます。

デアグリゲーションを行う理由の一つとして、サービスプロバイダネットワーク宛のDoS、および悪意を持った行為を軽減する効果が指摘されています。インターネットでは、経路広告の有無にかかわらず、連続したアドレス空間に対してスキャンを実施する事ができるワームやウィルスが存在します。これがいわゆる「ノイズ」というものを発生させます。筆者の経験では、/16の広告は2Mbps程度のノイズを吸い寄せてしまいます。インターネットが発展途上の地域にとっては、これは大変コストのかかる帯域になってしまうため、ISPは実際に利用しているアドレスしか広告しません。/24がインフラ、または顧客への割り当てで消費されるたびに、新たな/24をインターネットに広告します(困った事にこの/24は連続的に使用されるわけではありません)。元々/19だったアドレスが全て利用されたとしても、ISPはこれをアグリゲートしようとせず(彼らにとって、今現在ネットワークが正しく動いているため変更する必要がないと感じてしまう)、ルーティングテーブルの肥大化につながるわけです。

「商業的な理由により」デアグリゲーションしていると曖昧に宣言する事業者は、上記で述べたシナリオの両方を懸念しているようです。他にも理由があると推測されます。例えば、ここ数年で耳にする理由として、CIDRレポートのトップ10に登場する事はISPがビジネス的な成功を収めている証として良い意味で捉えられる事があるようです。

4.3 iBGPとeBGP

デアグリゲーションの理由をより深く探っていくと、サービスプロバイダ内のネットワークで利用されるBGP(iBGP)と、ドメイン間ルーティングで利用されるBGP(eBGP)との差分を理解しない事も利用のひとつとして見えています。iBGPの目的は、ISPの顧客のプレフィックスおよびインフラで使用するプレフィックス(ホスティングのアドレスなど)を全て広告する事です。eBGPは、ドメイン間のリーチャビリティを広告する事が目的で、単純にISPが、RIRから割り振りされたブロックをそのまま広告するだけで十分に達成できるのです。iBGPのルーティング情報を安易にeBGPへそのまま乗せてしまうISPは多く存在し、インターネットルーティングテーブルの肥大化に貢献しています。

ISPのiBGPとeBGPの差分を最もよくあらわしている実例は、オレゴン大学で実施しているRoute Viewsプロジェクト[5]で実際に見る事ができます。このプロジェクトでは、世界各地のISPが実際に保持しているインターネットルーティングテーブルが集まっていて、実際に参照する事ができます。eBGPをRoute viewsに送信するISPもいれば、iBGPを送信するISPもいます。このプロジェクトでは、「何を送信してよい」かは一切規定していません。ISPがアグリゲーションを行おうとする際に参考になるデータです。大規模なISPにとっては、そのままiBGPを送信しないようにする効果があり、他のISPにとっては、隣接ISPからもらった経路にiBGPが入っていれば、フィルタする事ができる情報を含んでいます。

4.4 マルチホーミングを目的としたデアグリゲーション

マルチホームしているネットワークは、トラフィックエンジニアリングを盾にデアグリゲーションを実施する事があり、インターネットルーティングテーブルが肥大化する原因として正当化されてしまっているようです。近年のマルチホーム手法で、多いのは割り振りもしくは割り当てアドレスを/24に分割し、/24をすべてのエクスターナルリンクで広告してしまうやり方です。

なぜ/24になってしまうか、というほとんどのISPは/24(レガシークラスC長)以上のプレフィックスを一律フィルタしてしまうという根拠の無い概念があるからです。CIDRレポートを軽く見るだけでわかりますが、これを裏付ける事実はありません。

さらには、/24に分割するとなんとなくマルチホームがうまくいく、という憶測を元を実施しているケースがあります。筆者の経験上、これは実際には上手くいきません。有効なトラフィックエンジニアリングやロードバランスは、割り振りされたブロック内の適切なサブプレフィックスを、トラフィックレベルに応じた適切な接続先で広告する事により達成できます。

この「ばらまき作戦」の結果、インターネットルーティングテーブルの肥大化に貢献してしまいます。

4.5 歴史的割り当て

歴史的割り当てが、インターネットルーティングテーブル肥大化の原因として指摘されてしまう事は少なくありません。RIRの仕組みが確立する以前に、IANAから割り当てされたアドレス帯の事です。元々、インターネットルーティングテーブルへの最大の「貢献者」は、クラスC時代以降最初の/8として利用された192/8ブロックでした。沢山のオペレータがデアグリゲーション努力を行い、クラスレスルーティングへの移行をがんばった「お掃除」プロジェクト以降、192/8のアドレスブロックからは、5,500プレフィックス程度がルーティングテーブルにのっています。実はこれは、RIRがISPへ割り振りする/8に比較するととても良い数値で、実際割り振りが完了したブロックでは8000から9000個のプレフィックスがルーティングテーブルを占めています。

旧クラスB空間(128/8から191/8)を見てみると、デアグリゲーションは顕著です。しかしこのほとんどは前述2セクション(iBGPとマルチホーミング)で記載されている内容が理由となっているようです。旧クラスA空間も同様です。

5. ルーティングテーブル肥大化のインパクト

インターネットルーティングテーブルの肥大化は我々にどう関係するのか?これまでの章では、インターネットへの経路広告はどうあるべきかを述べてきました。しかし、インターネットに参加しているASは様々な課題を抱えています。

5.1 ルーターメモリ

インターネットが始まって依頼、ルーターベンダーは「その時点でちょうど良い」仕様のルーターを提供してきました。急速に発展するインターネットの中で、このベンダーの開発思想はオペレーターにとって苦悩を与えています。インターネットが急速に広がるため、ある日フルルートを十分処理できるほどのメモリを搭載していても、次の年には足りなくなるという状況が発生しています。ルーターに最大値のメモリを搭載したとしても不足する事も多々あります。

より多くのメモリを搭載できる新しい筐体へのリプレースしか解がないため、インターネットに関連する機器の中では、ルーターが最も減価償却年数が短くなっています。装置リプレースはサービスプロバイダのネットワークを常に激変させる結果となっています。

5.2 ルーター処理能力

もう一つ影響を受けているリソースは、CPUです。インターネットルーティングテーブルが大きいほど、隣接ASとBGPをestablishする際の経路交換に時間がかかり、さらにトポロジ変化による経路再計算にかかる時間も長くなります。より高速なCPUを搭載すれば、この収束時間を短縮できるため、品質を維持するためには筐体毎交換しないといけない状況が多々発生しています。

5年後にどの程度のスペックのルーターが必要になるか、を一般的にあらわしたレポートが[6]で示されています。ルーティングテーブルの肥大化と、経路アップデートの増加と

ルーターの減価償却年数は直接的に関係しています。

5.3 経路収束

ネットワークが、ある特定のあて先への最適な経路を計算し終わる事を経路収束と呼びます。経路が取り消される、または再広告される都度経路収束が発生します。この計算には時間がかかります。

経路収束に必要なとする時間に影響を及ぼす要因はルーターのCPUとルーティングテーブルの経路数の二つです。インターネットルーティングテーブルのプレフィックス数が多いほどルーターの計算量は増えます。計算時間を短縮するためには、より高速なCPUが必要で、コントロールプレーンをアップグレードするお金と手間をかける事によって解決できます。ルーターの筐体交換はオペレーターにとって、より時間と費用のかかる作業でさらにネットワークの断時間も考慮しなければなりません。経路収束計算を短縮するもう一つの方法は、アグリゲーションを実施する事により、広告する経路数を最小限に抑え、インターネットルーティングテーブルの経路数を縮小する事です。

経路収束が遅いということは、ネットワークに障害が発生した場合の普及時間が増える事を意味し、障害をユーザ向けに「可視化」する事にもなります。

5.4 ネットワーク品質

ほとんどのネットワークオペレーターは、ネットワークの品質と経路広告は無関係だと思っているようです。

しかし、経路をアグリゲートせずに広告した場合、そのネットワーク内にいるエンドユーザにとっては、経路をアグリゲートした場合に比べてインターネットの利用品質が低下する事を証明するのは比較的簡単です。

筆者がよく目撃するケースは、ネットワークオペレーターがiBGPの経路をそのままインターネットに(eBGPで)広告してしまう場合です。顧客とのリンクがアクティブの場合、顧客のプレフィックスがiBGPに注入され、リンクが断になると、経路が取り消されます。iBGP経路をそのままインターネットに広告する問題は、顧客とのリンクが断になる度にフルルート保持するインターネット上の全ルータで経路が取り消されないといけない事です。経路取り消しは、瞬時にして行われるものではなく[7]、また均一に行われるものでもないため問題が発生します。顧客のリンクが再度アクティブになると、プレフィックスが再度iBGPに注入され、そのままインターネットに広告されます。こちらも同様に、アウトバウンドの経路広告は瞬時にして行われるものでもありません。

これによって起きることは、エンドユーザのリンクを接続してすぐにインターネット接続が完全に戻らないという現象です。プロバイダ内では、問題が無いように見えるためいくらユーザが問い合わせしたとしても、適切な回答を得られません。

もしプロバイダがiBGPではなく、アグリゲートされた経路をインターネットに広告していた場合、顧客はこのようにインターネット接続の復旧遅延を経験する事はなかったでしょう。

6. 解決策

過去数年にわたり、インターネットルーティングテーブルの肥大化を解決するためのソリューションが提案、実施されてきました。

6.1 CIDRレポート

CIDRレポートは元々、インターネットルーティングテーブルの肥大化を防止するためのテクニックでした。ISPにアグリゲーションを促す目的で考案され、実名を挙げて辱める事により効果をあげようというものでした。クラスレスインターネットへの移行の際には効果を発揮しましたが、近年では、CIDRレポートをむしろ、インターネットでの影響力を誇示するためのマーケティング材料として利用されている事例もあります。

CIDRレポートは過去数年にわたり大幅な改良が加えられており、[1]のWEBページでオペレーターが集約効果を検証できたりするツールなどが機能として付加されています。ルーティングテーブルの情報を用いて、任意のAS番号のアグリゲーション手法についてのアドバイスまで行う機能も付属しています。ネットワークオペレーターが、自身がインターネットにアナウンスしている経路を知らないわけもなく、またアグリゲーションする努

力を怠る理由は、全て言い訳でしかありません。

6.2 フィルタリング

IANAが、RIRへ割り振りするアドレス空間毎に、各RIRで設定している最小割り振りサイズでフィルタする手法も利用できます。例えば、あるRIRの/8からの最小割り振りサイズが/21であった場合、ネットワークオペレータは対外接続先からの経路広告で、該当する/8セグメント内では/21よりも長いサブネット広告は拒否する設定を入れる事ができます。フィルタを回避するために、より大きな経路に集約して広告させる、という効果が期待できます。

プレフィックスフィルタの効果は、CIDRレポートのウェブサイトで見ることが出来ます。さまざまなASの視点から見たインターネットルーティングテーブルのイメージが閲覧できます。

しかし、どれだけのオペレーターが最小割り振りサイズを基本としてフィルタを実施しているのか、またこの手法でルーティングテーブルサイズを減少させる効果が期待する程度のものなのかはわかっていません。もしRIRから最小割り振りサイズを取得した場合で、上流ISPが最小割り振りベースのフィルタを採用していた場合、残念ながらマルチホームを目的としたプレフィックスの分割広告は制限されてしまいます。

6.3 CIDRポリス

90年代後半から2000年前半までの間、少数のボランティアグループが、ルーティングテーブルを分析し、必要以上の経路を広告しているISPに協力の手を差し伸べていました[8]。連続の/24を広告しているISPを探し、集約する事でインターネットルーティングテーブルに大きく貢献できる事を助言していました。この活動に少々従う人もいれば、感謝されることもありましたが、相手から敵対されたりする事もありました。

2001年のITバブル崩壊を迎えると共に、CIDRポリスは本業に追われてしまいました。近年CIDRポリス復活の話は聞こえてきますが、執筆時点では具体的な活動にはなっていません。

6.4 BGPの機能

インターネットコミュニティ(主にISPと機器ベンダー)はBGPに機能を追加する事によりアグリゲーション手法の選択肢を増やしました。以下に説明します。

6.4.1 NO_EXPORT コミュニティ

アグリゲーションで一番最初に使える形の機能はBGPコミュニティアトリビュート仕様[9]で策定されているNO_EXPORTコミュニティです。このコミュニティが付与されたプレフィックスを受け取ったeBGPルータは、その経路を外に広告する事ができません。

サービスプロバイダが、上流またはPeerに対してトラフィックエンジニアリング目的でプレフィックスを分割して広告する際に、このコミュニティを付与する事で上流ISPに対して、他のASにその経路を広告しないように要求する事ができます。このコミュニティは、トラフィックエンジニアリング目的では非常によく利用されていますが、もっと普及してもいいはずで

6.4.2 NOPEER コミュニティ

トラフィックエンジニアリングとアグリゲーションに有効なもう一つの機能はNOPEERコミュニティです。比較的新しい機能で[10]、執筆時点では機器ベンダーの実装はまだ無く、ISPでこの機能を利用したい、という声もあまり聞こえてきません。

マルチホームトラフィックエンジニアリング目的のために経路を分割広告する場合、分割した経路にNOPEERコミュニティを付与して広告します。これを受け取った上流ISPはその経路を再広告する際に、Peerであるかそうでないかを判断した上でこの経路の広告可否を判断します。

ISPの対外接続先との関係は、通常トランジット、ピア、カスタマの3種類に分類することができます。NOPEERコミュニティを取り扱うには、BGPのPeerがトランジットであるか、ピアであるか、カスタマであるかを明記する必要があります。トランジットとカスタマに対しては、このコミュニティが付与された経路を送信しますが、Peerへはこの経路

は送信しません。これを利用する事により、インターネットの深い部分まで経路分割を流す事ができ、さらにインターネットのコアでこの経路が交換されてしまう事を防ぐことができます。

現在、インターネットの中心部分で「無償のピアリング関係」を結んでいるISPが10社程度あるといわれています[11]。インターネットでは、ほとんどのASは中心部分ではなくエッジ側に存在しており、NOPEERコミュニティを利用した場合、インターネットの中心部分のフルルートに多大な影響を及ぼすといわれています。

6.4.3 AS_PATHLIMIT

マルチホームのためのトラフィックエンジニアリング手法としてAS_PATHLIMIT属性[12]が最近提案されています。

ある特定のPrefixの伝播範囲を、AS_PATHLIMIT属性で定義する特定のAS「半径」内にとどめよう、という考え方です。この属性では、そのPrefixに付与する事のできる最大AS_PATH数を定義します（属性を付与したオリジンAS自身も含めて）。各ASは、経路に付与されているAS_PATHのAS数を数え、それを属性値と比較します。ASの数が属性値よりも大きい場合、その経路はルーティングテーブルにのらず、また他にeBGP Peerに広告される事ありません。これを使うと、インターネット上で「遠くのAS」が細かい経路を受ける事なく、ローカルな範囲でのトラフィックエンジニアリングを行う事ができます。

6.4.4 プロバイダ独自のコミュニティ

トランジット事業者の多くは、NO_EXPORTよりも範囲が広いが、グローバルルーティングテーブルにまで影響しない独自BGPコミュニティを顧客に提供しています。「地域限定で広告する」や「Peer全員に広告（トランジットには広告しない）」、などさまざまな種類があり、顧客がマルチホームしているPeerにのみ影響するものまであります。

プロバイダ独自のコミュニティは、通常そのプロバイダのWEBもしくはインターネットルーティングレジストリの中にルートオブジェクトとして記載されています。実際利用される値は、プロバイダ間調整は行われていませんが、InternetMCIが90年代半ばに利用していたコミュニティ利用方法が記載されているRFC1998 [3]を真似て、効果別の共通ルールに従っている例が多いです。[14]には、有名なプロバイダ独自コミュニティの例が記載されています。

ルータ上のBGPポリシーが適切に設定されている事はしっかりと管理するべきです。もし設定ミスを放置してしまった場合、せっかくコミュニティを付与しても経路が意図しないところへ流れてしまうからです。意図しない経路伝播はグローバルルーティングテーブルに影響を与えてしまい、このような経路広告制御機能を台無しにしてしまいます。

7. 推奨策

この文書の後半部分にあたるこの章では、RIPE Routing Working Groupが推奨する、インターネットへの経路広告手法を紹介します。全てのネットワークオペレータがこの推奨に沿った運用を行い、インターネットルーティングテーブルがコンパクトかつ必要最小限の大きさに留められる事を期待します。

7.1 初期割り振り

IPアドレスをRIRからの割り振り、または上流ISPからの割り当てを受けた場合、ネットワークオペレータがその経路を実現可能な限り集約し、インターネットに広告する事がインターネットコミュニティ全体の期待です。

例えば、ネットワークオペレータがRIRから/21を取得した場合、隣接ASへこの/21のみを広告するようBGPを設定するべきです。

7.2 追加割り振り

RIRは可能な限りLIRに対して連続したアドレス空間を割り振りします。もし過去の割り振りと同じ大きさでかつ連続した経路を割り振られた場合、ネットワークオペレータはこの二つのアドレスを集約し、一つの経路として広告するべきです。

例えば、/21の割り振りを過去に受けていた場合で、新たに連続した/21の割り振りを受けた場合、もしこの二つの/21がbit的に連続しているのであれば/20として結合する事が

できます。この/20を隣接ASへ広告し、古い/21は経路広告を取りやめるべきです。

もし割り振りが連続でない場合、またはサブネットbitの位置が異なる場合、またはサイズそのものが異なる場合は、経路集約は行えないため、二つの経路を独立の経路として広告するべきでしょう。

7.3 マルチホーム

もしマルチホームしているネットワークを運用している場合、トラフィックエンジニアリングのために経路を分割する必要性が生じる場合があります。もし経路分割が必須である場合でも依然として元々の経路をインターネットに広告する必要があります。さもなければ、片方のリンクが障害になった場合の冗長性が保てないからです。経路の分割広告は、注意深く行われなければなりません。過去数年にわたり、様々なネットワーク関連フォーラムで、トラフィックエンジニアリングの最大効果を得ながら、インターネットルーティングテーブルへのダメージを最小限に抑える手法を説明するチュートリアルが開催されています。

7.4 BGP拡張機能

6.4章で説明しているBGP拡張機能はルータベンダーが実装するべきで、これら機能は然るべきポイントで採用されるべきです。トランジット事業者以外のISPは、NO_EXPORTまたはNOPEERコミュニティを有効活用する事ができます。そして新たに開発された「AS_PATHLIMIT BGP:」アトリビュートを利用し、分割したサブプレフィックスがインターネット全体に伝播してしまわないようにする事ができます。

7.5 代理アグリゲーション

代理アグリゲーションは注意して実施しなければなりません。他ASがオリジンとなっている経路広告を勝手にアグリゲートする事は、予期しないマイナス効果を生むケースがあります。特に他ASがトラフィックエンジニアリングのために実施している場合、障害が発生した際に、トラフィックを「ブラックホール」してしまう事になる可能性があります。代理アグリゲーションは、実施するネットワークオペレータが悪影響がない事を確実に把握している場合のみに実施するべきです（例えば、マルチホーム先が同じ上流ISPの場合など）。

7.6 IPv6

今まで紹介した推奨策はIPv4インターネットを中心にお話しましたが、IPv6でも同様に当てはまります。IPv6インターネットの一部になるということは、IPv4インターネットの一部になる事と同等で、ASに対しての要求事項はまったく同じです。

8. 結論

アグリゲーションは、今日のインターネットに参加する全てのネットワークオペレータが取り組むべきです。90年初頭にクラスレスインターネットへの移行時に当たり前であった事は、もはや忘れられてしまったようです。その結果、インターネットルーティングテーブルが恐ろしいスピードで肥大化し、インターネットに参加する全プレイヤーに影響を与えています。

[16]などで紹介しているBGP分析活動は、インターネットルーティングテーブルがどれだけ削減の可能性があるのかを示しています。そしてこの削減率はインターネットルーティングテーブルや、測定方法によって異なりますが、30%-50%程度の驚異的な削減率がはじきだされています。

9. 謝辞

この企画の構想に携わったLINUXメンバー（[17]LINUXで実施されたルーティングアグリゲーションポリシー実験）、このドキュメントの基礎となる初期ドキュメントを作ったMike Hughesに感謝申し上げます。助言頂いたLeo Vegoda、Geoff Huston、Gaurab Raj Upadhaya、Duncan Rogersonに感謝申し上げます。

Y. 参考文献

[1] The CIDR Report | Original Idea: Tony Bates | Maintained by: Geoff Huston

- [2] RFC1518 - An Architecture for IP Address Allocation with CIDR | Tony Li and Yakov Rekhter
- [3] RFC1519 - Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy | Vince Fuller, Tony Li, Jessica Yu and Kannan Varadhan
- [4] Routing Table Status Report | Geoff Huston - APRICOT 2005/APNIC 19
- [5] The Route Views Project | University of Oregon
- [6] Routing Update | Geoff Huston - APRICOT 2006/APNIC 21
- [7] Delayed Internet Routing Convergence | Craig Labovitz, Abha Ahuja, Abhijit Bose and Farnam Jihanian - Sigcomm 2000
- [8] CIDR Police - Please Pull Over and Show Us Your BGP Announcements | Barry Greene and Hank Nussbacher - NANOG 27
- [9] RFC1997 - BGP Communities Attribute | Ravi Chandra and Paul Traina
- [10] RFC3765 - NOPEER Community for Border Gateway Protocol (BGP) Route Scope Control | Geoff Huston
- [11] AS Path Analysis | Vijay Kuhmar Adhikari, Gaurab Raj Upadhaya and Bill Woodcock - SANOG 8
- [12] AS-PATHLIMIT Attribute | Joe Abley, Tony Li and Rex Fernando
- [13] RFC1998 - An Application of the BGP Community Attribute in Multihome Routing | Enke Chen and Tony Bates
- [14] BGP Community Guide
- [15] BGP Multihoming Techniques | Philip Smith - NANOG 35
- [16] Aggregation Potential
- [17] The Routing Aggregation Policy - A Failed Social Experiment at the LINX | Nigel Titley - APNIC 21

Z. 筆者

筆者への連絡方法

Philip Smith | pfs@cisco.com
Rob Evans | rhe@nosc.ja.net
Mike Hughes | mike@linx.net