今月のテーマ

BGP

今回の10分間講座は、
BGPについて解説します。

■BGPとは

相互通信を行うインターネットを使った通信では、通信を行うパケットのあて先を正確に把握し、維持していく必要があります。この、パケットのあて先を正確に把握し、維持していくための技術を経路制御と言い、経路制御を行うためのプロトコルを経路制御プロトコルと呼びます。今回の10分用語解説では、代表的な経路制御プロトコルであるBGP（Boarder Gateway Protocol）について解説します。

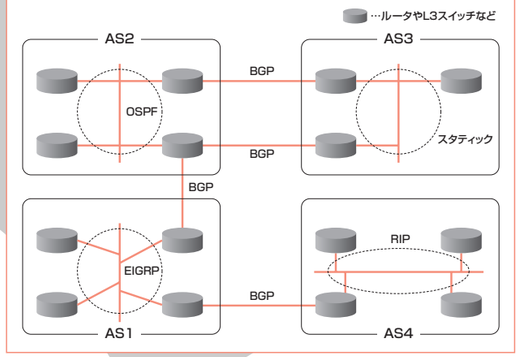
■BGPの概要

▼EGP/IGP

経路制御プロトコルの分類方法はいくつかありますが、制御しようとする経路の対象範囲によって、EGP（Exterior Gateway Protocol）、IGP（Interior Gateway Protocol）の二つに大別することができます。EGPはインターネット上で組織間の経路情報をやり取りする経路制御プロトコルであり、BGPはEGPに分類されます。組織の内部で完結する経路制御プロトコルをIGPと呼び、RIP（Routing Information Protocol）やOSPF（Open Shortest Path First）、EIGRP（Enhanced Interior Gateway Routing Protocol）などはIGPに分類されます。また、古くはEGPという名称の経路制御プロトコルも存在しましたが※1、今日ではほとんど利用されていません。経路制御プロトコルのEGPと組織間での経路情報をやり取りするプロトコルの総称としてのEGPが混在する場合には、後者をEGPsと表記し、区別することがあります。

現在のBGPはRFC4271で定義されるBGP-4（BGP version 4）を指し、BGPとBGP-4という単語はほぼ同じ意味であると言えます。

図1　BGPと他の経路制御プロトコル



▼AS番号

BGPは、他の経路制御プロトコルと違い、組織内での設計と機器の準備だけでは利用することができません。BGPでは経路制御を行う組織ごとにインターネットの世界で唯一の番号が割り当てられ、個々の経路を識別します。このインターネットの世界で唯一の番号をAS（Autonomous System）番号と呼び、AS番号はIANA（Internet Assigned Numbers Authority）※2が管理しています。IANAはAS番号をある程度のブロック単位でRIR（地域インターネットレジストリ:Regional Internet Registry）へ割り振りを行っており、RIRからNIR（国別インターネットレジストリ:National Internet Registry）やLIR（ローカルインターネットレジストリ:Local Internet Registry）へさらに割り振られるという階層構造で管理されています。

BGPによる経路制御を始めるには、このAS番号の割り当てを受けることが必要です。AS番号はRIRやJPNICなどのNIRから割り当てを受けることができます。

AS番号は原則として2バイトの大きさを持ち、0から65535までの整数で表現されます。しかし、最近はAS番号の数が足りなくなってきたため、後述するように4バイト化が進んでいます。AS番号は、数種類に分類することができ、組織に割り当てられる番号、IANAが予約する番号、組織内に閉じ外部に直接接続しないネットワークのためのプライベートAS番号などに分類されます。

表1 2007年2月現在のAS番号のリスト

0	予約
1〜43007	RIRへ割り振り
43008〜48127	Held by IANA
48128〜64511	Reserved by IANA
64512〜65534	Private AS番号
65535	Reserved by IANA
23456	4バイトASのために利用

http://www.iana.org/assignments/as-numbersより

▼BGPの拡張

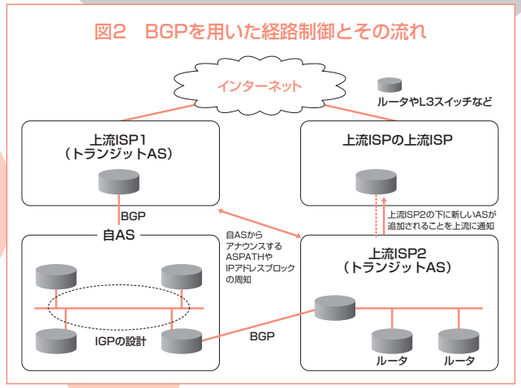
BGP4は元々、IPv4アドレスの経路制御のために利用する経路制御プロトコルとして設計されました。しかし、IPv4アドレスだけではなく、IPv6アドレスやマルチキャストアドレス、MPLS（Multi Protocol Label Switching）のラベルなど、さまざまなプロトコルの経路制御に対応すべく、拡張が加えられ続けています。代表的なRFCにRFC2545「Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing」、RFC2858「Multiprotocol Extensions for BGP-4」などが存在します。これらの拡張をまとめてBGP4+やMBGPと称する場合があります。

▼BGPを用いた経路制御の実際

BGPを用いた経路制御では、AS番号とIPアドレスの割り当てを受けた後、上流ISPやIXへの接続が必要です。その後、自組織のBGPルータと接続先のBGPルータ間でピアと呼ばれる経路の交換を行う設定を行います。BGPによる経路制御はこのピアを通して行います。ピアを設定後、自組織のIPアドレスを接続先へピアを通じて通知する必要があり、このことを「アドレスをアナウンスする」と言います。自組織のアドレスをアナウンスするだけではなく、インターネット上の経路を上流ISPから取得することも必要です。

自組織以外のアドレスを接続先にアナウンスするASを「トランジットAS」と言い、自組織のアドレスだけを接続先にアナウンスするASを「非トランジットAS」と言います。非トランジットASは一般的に、フルルートを上流ISPから取得する必要があり、このことを「トランジットを取得する」と言います。トランジットを単独の上流ISPから取得することを「シングルホーム」と言い、トランジットを複数ISPから取得することを「マルチホーム」と言います。

図2　BGPを用いた経路制御とその流れ



※1 BGPなどを含む経路制御プロトコルの総称としてのEGPとは別に、EGPという名称を持つプロトコルが存在し、RFC904で仕様が定められています。このような事情からEGPsという表記が生まれました。

※2 それぞれ、地域、国および地域といった単位でIPアドレスの管理を行っているインターネットレジストリです。

■BGPプロトコルの簡単な紹介

▼BGPプロトコルの流れ

BGPはTCPポート179番を利用して通信を行います。BGPルータはピアを確立するために、メッセージと呼ばれる形式で機能情報を交換します。メッセージはBGPの状態により数種類存在します。ピアが確立した後は、経路情報を伝達するためのメッセージを定期的に交換し、お互いの持つ経路情報を保持、更新し続けます。

BGPではピアを確立するルータが同じAS間である場合と、異なるAS間である場合とで異なる方式で経路情報を処理します。同じAS間でのピアをiBGP（内部BGP）ピアと呼び、異なるAS間でのピアをeBGP（外部BGP）ピアと呼びます。

以下にBGPプロトコルで交わされるメッセージと、メッセージが利用するパス属性について簡単に解説します。

▼メッセージ

・OPENメッセージ

OPENメッセージはTCPセッションの確立後、最初に交換されるメッセージです。OPENメッセージはお互いのAS番号やピアの認証を行い、拡張機能などについて情報の交換を行います。ピアの確立に問題が無ければ、後述のKEEPALIVEメッセージやUPDATEメッセージの交換に移ります。

・KEEPALIVEメッセージ

KEEPALIVEメッセージは、ピアが確立されていることを確認するために定期的に交換するメッセージです。KEEPALIVEメッセージを交換する頻度は、OPENメッセージのホールドタイマの値から設定されます。一般的にKEEPALIVEメッセージはホールドタイマの1/3の間隔で交換されます。ホールドタイマの時間が過ぎてもKEEPALIVEメッセージやUPDATEメッセージが交換されないと、ピアがダウンしたと判断されます。

・UPDATEメッセージ

UPDATEメッセージは実際の経路情報を伝達するメッセージです。UPDATEメッセージには新規にアナウンスする経路情報と削除する経路情報のリストが格納されています。UPDATEメッセージの中には経路情報ごとにパス属性という属性情報が存在し、経路制御に利用されます。パス属性の詳細については後述します。

・NOTIFICATIONメッセージ

NOTIFICATIONメッセージはピアに対し、ピアの継続不可能なエラーが発生した場合に送信されます。NOTIFICATIONメッセージを送信した送信者側は、NOTIFICATIONメッセージを送信後、TCPのセッションを切断します。

▼代表的なパス属性

UPDATEメッセージで利用される、代表的なパス属性について説明します。

・ORIGIN属性

ORIGIN属性はその経路情報の生成元を表します。

ORIGIN属性にはIGP/EGP/INCOMPLETEの3種類があります。

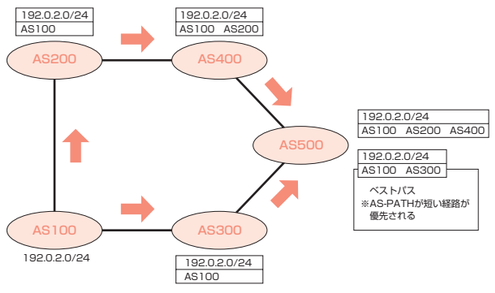
表2 ORIGIN属性

優先順位高い↑	IGP	AS内部で生成された経路
	EG	AS外部で生成された経路
優先順位低い↓	INCOMPLETE	IGP/EGP以外の手段で取得した経路

・AS_PATH属性

AS_PATH属性はその経路情報が通過してきたAS番号を並べたAS番号のリストです。ピアからUPDATEメッセージを受け取った時にAS_PATH属性中に自分のAS番号を含む経路は全て無視します。このことによりルーティングループを防ぐことができます。AS_PATH属性の長さで、経路の優先順位を制御することが可能です。

図3 AS-PATH属性と経路の選択



・NEXT_HOP属性

NEXT_HOP属性は、スタティックルートやOSPFなどのIGPでのパケットホップ先とは異なり、複雑な決定プロセスを経てネクストホップが決まります。NEXT_HOP属性はeBGPピアやiBGPピアなどの経路取得方法によって異なります。

・MULTI_EXIT_DISC属性

MULTI_EXIT_DISC属性は外部のASとの接続口を複数持っているASで機能し、外部のASから自ASへ向かってくる、内向きの通信を制御します。自ASへの入り口の経路として利用することを希望する経路に対して、低い値をMULTI_EXIT_DISC属性に設定します。

しかし、経路を受け取った外部ASでMULTI_EXIT_DISC属性の値を書き換えてしまう場合があり、MULTI_EXIT_DISC属性だけで戻りの経路を制御することは難しい場合もあります。

・LOCAL_PREF属性

LOCAL_PREF属性は、外部のASとの接続口を複数持っているASで機能し、自ASから外部へ送信される、外向きの通信を制御します。LOCAL_PREF属性はAS内部独自の属性で、iBGPピア間でのみ交換されます。

その他にもさまざまな属性がありますが、その他の属性は参考文献をご参照ください。

▼経路の優先順位

BGPで取得した経路が複数存在する場合、次の順番で利用する経路が決定されます。

表3 BGP経路優先順位

1. ネクストホップへのIGPルートを持っていない経路は無視されます。
2. weightパラメータを持つルータはweightパラメータ値が最大の経路を選択します。
3. LOCAL_PREF属性の値の最も高い経路を選択します。
4. AS_PATH属性のリストの長さが最も短い経路を選択します。
5. ORIGIN属性のタイプが最も低い経路を選択します。（IGP<EGP<INCOMPLETEの順）
6. ルートが同じASから取得し、複数存在する時にはMULTI_EXIT_DISC属性の低い経路を優先します。
7. iBGPよりもeBGPで取得した経路を優先します。
8. ネクストホップへIGPで最も近い経路を優先します。
9. ルータIDが最も低いピアから学習した経路を優先します。（ルータIDは通常、ルータのインタフェースから自動的に生成されます）

実際のルーティングでは、BGP以外で取得した経路との優先順位付けもあり、BGP内での優先順位だけでは、利用する経路を決定することはできません。

▼コンフェデレーション/ルートリフレクタ

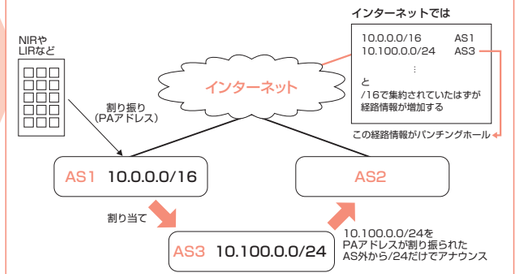
大規模で複雑なASを制御する仕組みとして、コンフェデレーションとルートリフレクタの存在があります。ルートリフレクタは、AS内部でiBGPピアを集約することができます。コンフェデレーションはAS内部を複数のサブASに分割し、それぞれのサブAS単位で経路制御を行います。

■BGPの最近の話題

▼経路数の増大

2006年度中にインターネットのフルルートは20万を超えました。経路数の増大は、BGPルータのCPUやメモリの負荷を増加させます。急激に経路数が増加した場合にはルータのCPUやメモリの必要量が足りなくなり、BGPによる経路制御が破綻する可能性がありますが、現在まで、ルータも経路数増加に耐えられるよう、その性能を継続的に高めており、問題なくBGPによる経路制御が継続しています。また、経路数の増大を招くパンチングホールについても是非が議論されています。

図4 パンチングホールの例



▼AS番号の枯渇

AS番号は2バイト長であるため、利用できる番号は0～65535しかありません。また、プライベートAS番号などの予約分もあり、全体を利用できるわけではありません。2007年1月現在、IANAでは1番～43007番までをRIRへ割り振っていて、残りは43008～64511の番号となっています。そのため、世界でAS番号を4バイトに拡張する技術が実装され始めており、2007年1月から、AS番号を申請する際に、希望する組織へは4バイトAS番号の割り当てが開始されています。2007年現在では4バイトAS番号を理解できるルータは非常に少ない状況ですが、これからは4バイトASを理解できるルータが増えると見込まれます。

▼経路情報の信頼性

現在のBGPを利用した経路制御では、UPDATEメッセージに含まれる経路情報を認証する仕組みが存在しません。そのため、BGPルータのオペレータの設定間違いなどにより、本来設定するべきでない経路がインターネットにアナウンスされる場合があります。

そのような経路が流れた場合、設定間違いを行ったASだけでなく、全インターネットに影響を与えることになります。経路情報を識別し、安全な経路制御を実現する仕組みとして、以下のような技術が存在します。

・IRR

BGPでアナウンスするAS番号や経路について、事前に登録しておく手法としてIRR（Internet Routing Registry）があります※3。IRRでは自分のアナウンスするAS番号や経路を簡単に登録できる反面、登録するAS番号や経路のそもそもの信頼性が担保されていないなどの問題があります。

代表的なIRRとして、米国Merit※4が運営するRADBがあります。
また、日本国内におけるIRRとしては、2006年8月より正式サービス化した、当センターが提供するJPIRRがあります。

IRRの正当性を担保する仕組みとして、IRRに登録するオブジェクトへ電子署名を行う手法が提案されています。この手法では、RIRであるAPNICから提案されている、「Route Origination Authorization（ROA）with IRR」をはじめ、日本国内でも検討がなされています。

・S-BGP/soBGP

BGPのプロトコル内部で、電子署名技術を利用する試みにS-BGP（Secure BGP）とsoBGP（Secure Origin BGP）があります。どちらの技術も経路情報のORIGIN ASと経路情報のAS_PATH属性が正しいかどうかを検証することを主目的としています。

▼障害検知方法

BGPでの経路制御はインターネットの基盤を構成するため、経路上の障害を迅速に検知し、経路を切り替えることが望まれています。この経路上の障害検知と切り替えを行う方法の一つとして、それぞれ次の技術が議論されています。

・Bidirectional Forwarding Detection

BFD（Bidirectional Forwarding Detection）はBGPに限らず、さまざまな経路制御プロトコルにおいて、障害が発生した経路を検出、通知する仕組みです。BGPではデフォルトの状態で最大、180秒間、通信断後もパケットを通信断となったインタフェースへ送出する場合があります。このようなパケットロスを引き起こす事象を、根本から解決する仕組みとして、BGPでのBFDが議論されています。

BFDは隣接ノード間のリンクやパスの障害を、高速に検知することを目的としています。BFDのベンダによる実装では、UDPパケットを使った生存確認を高速で行い、障害発生時には経路制御プロトコルへすばやく障害を通知することが可能です。BFDは現在もIETF BFD Working Group※5で議論が継続中です。

・Micro Allocation for IPv6

BGPの経路選択アルゴリズムでは、経路集約とネクストホップアドレスの関係で、障害発生時、経路の切り替わりが瞬時に行われない場合があります。この事象を予防するために、AS内の基幹ネットワークだけに利用することを目的としたIPv6アドレスが、ARINやRIPE NCC、LACNICでポリシーとして採用され、実装がなされています。

Policy Proposal 2006-2: Micro-allocations for Internal Infrastructure
http://www.arin.net/policy/proposals/2006_2.html

▼エニーキャスト

IPアドレスは、一般的に特定のインタフェースへ一意に割り当てて利用します。 この利用形態のIPアドレスをユニキャストアドレスと呼んでいます。それに対して、エニーキャストアドレスは、 複数のインタフェースに割り当てられた同じIPアドレスです。複数のインタフェースに割り当てられたIPアドレスを含むIPアドレスブロックを、分散してアナウンスするためにBGPが利用されます。

BGPを用いたエニーキャストでは、エニーキャスト用のASを構築し、さまざまなASの下に接続します。エニーキャスト用のIPアドレスブロックを受け取ったASはAS_PATH属性の最も短いASを選択します。

エニーキャストの利用形態の一つにDNSサーバの分散配置などが挙げられます。ルートDNSサーバやccTLDのDNSサーバでは、エニーキャストを用い、拠点間分散により負荷分散、耐障害性の向上を図っています。

■おわりに

BGPは自由度が高く、運用される現場でも高い信頼性を求められることが多い経路制御プロトコルです。そのため、BGPオペレータとして成長するために、いわゆる「壊して覚える」ということが非常に難しいのが現状です。そのため、検証環境などで、擬似的なインターネット環境を作成し、BGPオペレータは日夜研鑽に励んでいます。BGPは非常に奥が深く、とてもこの10分用語解説という限られた誌面で解説しきれるものではありませんが、みなさまがインターネットを利用される際にBGPのことを思い出し、経路制御について理解を深めていただくきっかけとなればと考えております。

（JPNIC IP 技術部 岡田雅之）

■参考文献

RFC-2545 “Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing”
RFC-2858 “Multiprotocol Extensions for BGP-4”
RFC-3392 “Capabilities Advertisement with BGP-4”
RFC-4271 “Border Gateway Protocol 4（BGP-4）”
RFC-4272 “BGP Security Vulnerabilities Analysis”

※3 IRRの詳細については、過去の記事をご参照ください。

JPNIC NewsLetter Vol.27
インターネット10分講座「IRR」
http://www.nic.ad.jp/ja/newsletter/No27/100.html

JPNIC NewsLetter Vol.34
特集「JPIRRサービス正式サービス化」
http://www.nic.ad.jp/ja/newsletter/No34/0210.html

※4 米国にある研究機関で、RADB（Routing Assets Database）と呼ばれるIRRを運営しています。
Merit Network, Inc.
http://www.merit.edu/

※5 **IETF BFD（Bidirectional Forwarding Detection）WG**
http://www.ietf.org/html.charters/bfd-charter.html