分散システム特論レポート
BGP-4(Border Gateway Protocol-4)
[RFC1771]、[Internet-Draft]

坂本 弘達 hirotatu@swlab.csce.kyushu-u.ac.jp

2001年9月21日(金)

1 背景

今日のインターネットは、複数のAS(Automous System、自律システム)が相互に接続することで成り立っている。各ASはそれぞれ独自の運営方針を持ち、外部から見るとそれぞれが固有の識別番号(AS番号)を持った一つの実体として見える。ASを単位としてインターネットを見るとき、経路情報を扱うプロトコルには、AS内部で経路情報を交換するためのプロトコル群(IGPs)と、AS間で経路情報を交換するプロトコル群(EGPs)が存在する。

ASには独自の運営方針が存在する。経路の広告に関する運営方針としては、自分の経路を他に報告し、他からの経路の広告も受け取り、他から受けた経路の内の一部は他に広告するなどが考えられる。自ASからの出力トラフィックの制御に関する運営方針としては、基本的に、ある境界ルータからパケットを出力し、それ以外の境界ルータはバックアップに使うなどがある。ルーティングに関する運営方針としては、パケットの送信元もしくは送信元と送信先によって経路を制御するというものがある。EGPsでは、上で挙げたようなASの運営方針を実現できる。

EGPsとしては、EGP(Exterior Gateway Protocol)、IDRP(InterDomain Routing Protocol)、BGP(Border Gateway Protocol)などがあり、事実上の標準として使用されているものがBGPのバージョン4(BGP-4)である。本レポートでは、BGPについて、BGP-4に基づいて記述している。

2 BGPの概要

BGPが前述の通り、EGPsの一つ、すなわち、AS間の経路情報、到達可能性情報を交換するプロトコルである。BGPでは、BGPを話す装置をBGPスピーカと呼ぶ。(特にこの装置がルータの場合はBGPルータと呼ぶ。)このBGPスピーカは経路情報を格納するデータベースを持つ。また、運営方針を格納するデータベースも持つ。BGPスピーカはOpen、Update、Notification、KeepAliveという4種類のメッセージを他のBGPスピーカと交換することにより、データベース内の経路情報を更新したり、経路情報を交換したりする。

BGPは、他の経路制御プロトコルと比較すると、以下の特徴がある。

- 『パスベクトル型プロトコル』: BGPの経路情報には、目的ASから自ASまで経由してきたASを表すAS番号のリストが含まれる。『距離ベクトル型プロトコル』であるRIPでは、ルーティングループができるという問題があるが、BGPでは排除される。

- TCP 上に実装、ポート番号 179 を使用：経路制御用のパケット再送などの処理は TCP に任せている。
- CIDR(Classless InterDomain Routing) のサポート
- 経路の集約機能: BGP-4 でのみサポートされており、古いバージョンの BGP ではサポートされていない。

3 経路情報の交換

経路情報の交換には以下の場合が存在する。

- 異なる AS 内に存在する BGP スピーカにおける場合
  AS 内のルータは、矛盾のないトポロジーの管理を行う。そのために、異なる AS 内に存在する BGP スピーカは外部リンク (External Link) で接続され、経路情報を交換する。特にこのときのプロトコルを E-BGP(External-BGP) と呼ぶ。E-BGP を話す BGP スピーカは、同じ物理ネットワークセグメントに存在しなければならないという制限がある。
- 同一 AS 内に存在する BGP スピーカにおける場合
  外部 AS から伝搬してきた経路情報は、自 AS 内の BGP スピーカで整合性の取れたものである必要がある。また外部 AS に広告する経路情報も同様である。そのために、同一 AS 内に存在する BGP スピーカは内部リンク (Internal Link) で接続され、経路情報を交換する。特にこのときのプロトコルを I-BGP(Internal-BGP) と呼ぶ。

4 BGP のメッセージ

BGP スピーカは 4 種類のメッセージを交換する。

- Open: BGP セッションを確立、BGP スピーカ間の情報交換を可能にする。
- Update: BGP スピーカ間で経路情報の更新内容を広告する。
- Notification: エラー条件が検知されると送信、その後、セッションを終了する。
- KeepAlive: 相手がしていることを確認する。

以下、メッセージの詳しいパケットフォーマットについて記述する。そして BGP スピーカ間でメッセージを交換する例を示す。

4.1 メッセージパケットフォーマット

- 固定長のヘッダ部と可変長のデータ部から構成
- 最小 19bytes、最大 4096bytes

4.1.1 ヘッダ部

図 1 に、メッセージパケットに共通であるヘッダ部のフォーマットを示す。

- マーカ: メッセージの受信側で推測可能な認証値
  - オープンメッセージの場合や、認証情報が無い時は、全て 1
- メッセージ長: メッセージ全体の長さ

マーカ(16)	メッセージ長(2)	メッセージタイプ(1)	データ(可変)

図1: BGPメッセージフォーマット:ヘッダ部

4.1.2 Openメッセージ

BGPスピーカはBGPセッションを確立するために、最初にOpenメッセージを転送する。以降、BGPスピーカ間の情報交換が可能となる。図2にOpenメッセージのフォーマットを示す。

- バージョン
  - 送信者と受信者が同じバージョンを使っているかの確認に使用
  - 現在の値は4
- 送信者のAS: 送信者のAS番号
- 保持時間: 宛先が機能していないと決定してしまう前に、メッセージを受信せずに待つ事のできる最大秒数
  - BGPスピーカが自ら設定した値と受け取った値を比較し、最小のものに設定しなおす
  - 保持時間を0にすると、KeepAliveメッセージを送信しない
- BGP識別子: 送信側のBGP識別子(IPアドレス)
  - BGPスピーカ起動時に、全てのローカルインターフェースと全てのBGPピアで同じ値になる
- オプションパラメータ長: オプションパラメータフィールドの長さ
- オプションパラメータ: オプションパラメータのリスト
  - RFC1771で定義されているのは認証情報
    - ∗ 認証コード: 使われている認証の種類
    - ∗ 認証データ: 認証機構用のデータ

バージョン(1)	自律システム(2)	保持時間(2)	BGP 識別子(4)	オプション パラメータ長(1)	オプション パラメータ(4)

図2: BGPメッセージフォーマット:Openメッセージ

4.1.3 Updateメッセージ

BGPスピーカが経路情報の更新内容を広告するためにはUpdateメッセージを使用する。図3にUpdateメッセージのフォーマットを示す。

- 使用不能経路の長さ: 削除された経路フィールドの合計長
- 削除された経路: 削除された⟨長さ、IPアドレスのプレフィックス⟩のリスト
- パス属性長: パス属性フィールドの長さ
- パス属性

  - 送信元 (必須)
  - AS パス (必須): 経由してきた AS のリスト
    - リストの要素は ⟨ 要素のタイプ、長さ、値 ⟩
    - 要素のタイプは、AS_SEQUENCE(AS 固有の番号) または AS_SET(AS 番号の集合)
  - ネクストホップ (必須): 目的ネットワークへのネクストホップの IP アドレス
  - 複数出口の識別 (オプション): 隣接する AS に対する複数出口の識別点
  - ローカルでの最適性 (任意): 同 AS 内での評価に基づく経路の良さ
  - アトミック集約経路 (任意): 経路か集約されている事を示す
  - 集約経路 (オプション): 集約された経路の情報

- ネットワーク層の到達可能性情報: 広告される目的ネットワークに関する ⟨ 長さ、IP アドレスのプレフィックス ⟩ のリスト

使用不能経路長(2)	削除された経路(可変)	パス属性長(2)	パス属性(可変)	ネットワーク層の到達可能性情報(可変)

図 3: BGP メッセージフォーマット:Update メッセージ

4.1.4 Notification メッセージ

BGP スピーカにおいてエラーが検知されると送信される。その後 BGP セッションは終了し、TCP のコネクションも切断される。また BGP スピーカからの要求により、意図的に送信することも可能である。図 4 に Notification メッセージのフォーマットを示す。

- エラーコード
  - 致命的なエラーにより発生: メッセージヘッダエラー、Open メッセージエラー、Update メッセージエラー、保持時間の超過、有限状態マシンエラー (予期しないイベントが発生した時)
  - BGP スピーカからの要求により発生: 中止
- エラーサブコード
- エラーデータ

例) オープンメッセージにサポートしていないオプションパラメータを発見

- エラーコード: OPEN Message Error
- サブエラーコード: Unsupported Optional Parameters

エラーコード(16)	エラーサブコード(2)	エラーデータ(可変)

図 4: BGP メッセージフォーマット:Notification メッセージ

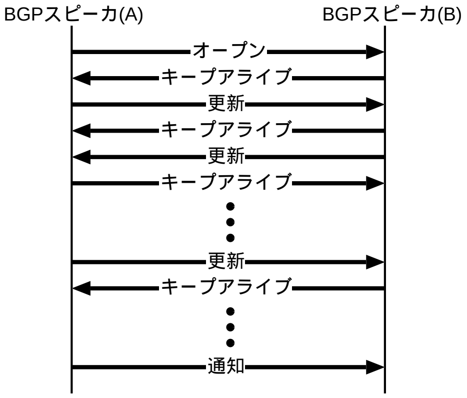
4.1.5 KeepAlive メッセージ

相手が生きているを確認するためにBGP スピーカはKeepAlive メッセージを適度に送信する。このメッセージは保持時間 (Open メッセージ内のフィールド内に格納) で示される有効期限が来ない程度に送信される。またこのメッセージは他の3 つのメッセージと異なり、特別なフィールドは存在せず、ヘッダ部のみで構成される。尚、TCP にはKeepAlive オプションがあるが、BGP ではこれを利用しない。

4.2 BGP スピーカ間のメッセージ交換の例

以下に、BGP スピーカがセッションを開き、メッセージを交換した後、セッションを終了する例を示す。

1. (A) が Open を送信、BGP セッションを開く
2. (B) がそれに対して KeepAlive を送信
3. (A) が Update で全ての経路情報を送信
4. (B) が KeepAlive を送信
5. (B) が Update で全ての経路情報を送信
6. (A) が KeepAlive を送信
7. (A) が Update で更新された情報を送信
8. (B) が KeepAlive を送信
9. (A) が Notificatino で BGP セッションを終了



5 AS 間の経路情報広告の例

図5 に AS 間の経路情報広告の例を示す。

1. 初期状態で AS (E) は次のような経路情報を持っている
2. AS (F) と AS (E) の間で BGP セッションを確立、更新メッセージで AS (E) の経路情報を AS (F) に広告
3. AS (F) は経路情報を更新
4. AS (F) から AS (E) も同様の過程で経路情報を更新
5. AS (E) は、 AS (B) と AS (D) に経路情報を広告
6. AS (C) は、両方の AS から経路情報を受け取り、経路情報から最適な経路を選択、保存

6 経路決定の過程

BGP スピーカは、経路情報を RIB(Routing Information Base) というデータベースに格納する。また、AS の運営方針を PIB(Policy Information Base) に格納する。RIB では更に経路情報を3 種類に分類して格納する。

- Adj-RIB-In: 他の BGP スピーカから広告された情報を格納
- Loc-RIB: 自 BGP スピーカが扱う情報を格納
- Adj-RIB-Out: 他の BGP スピーカに広告する情報を格納

これらのデータベースを用いて、BGP スピーカは経路情報を決定する。以下に

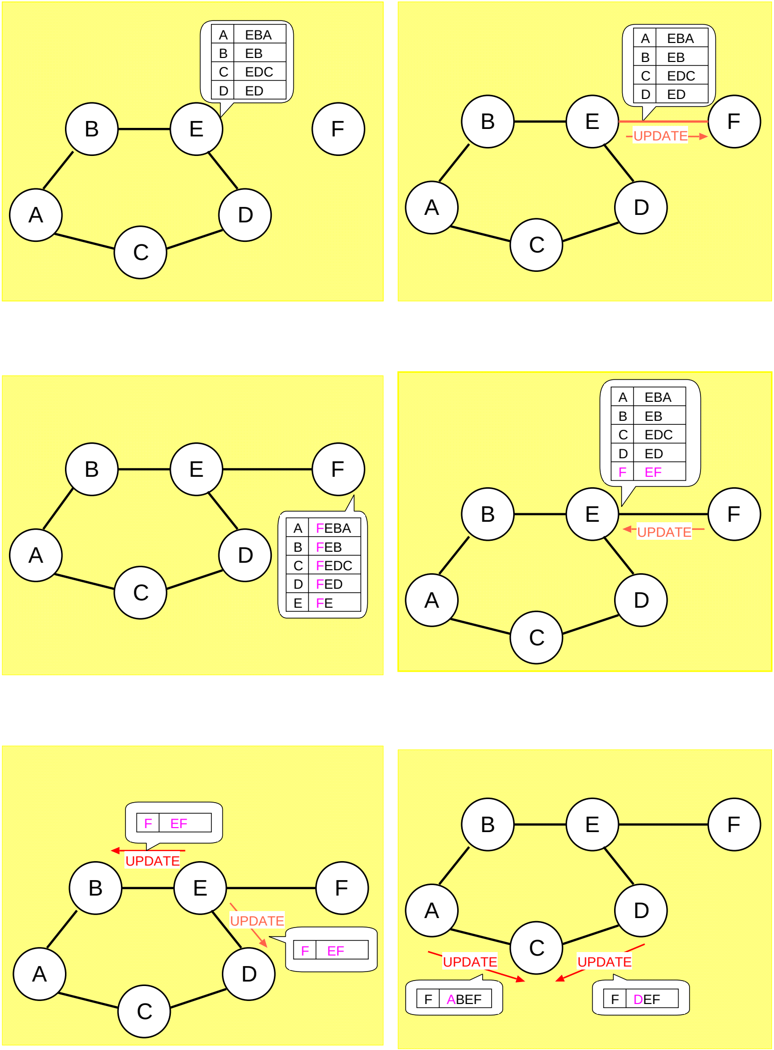
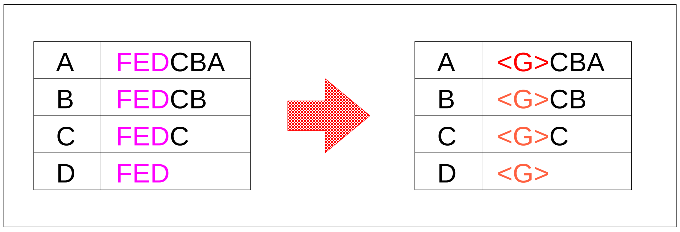


図 5: AS 間の経路情報広告

1. ピアから広告された経路情報を Adj-RIB-In に格納する。
2. 他の AS からの広告に従って、変更のある経路の良さを計算、最良の経路を同 AS 内の他の BGP スピーカに広告する。この経路の良さとは、「ローカルでの最適性」を意味する。
3. それぞれの目的ネットワークに対して、自 AS のポリシーに従って最良の経路を選択する。
4. 選択した経路を Loc-RIB に格納する。
5. 経路の集約を行い、PIB 内の運営方針に従って、他の AS に広告する経路を選択する。
6. 選択した経路を Adj-RIB-Out に格納し、その後、他の AS に広告する。

6.1 経路の集約

BGP は経路情報として AS パスを保持する。図 6 に示すように、それぞれの経路に対する AS パスには共通部分が多い。そのためこの共通部分を 1 つの集合として扱うことで経路情報を集約可能である。これにより Adj-RIB-Out に格納するデータ量が削減できる。ただし、最適な経路を選択できない可能性が生じる。この経路の集約は、BGP-4 の特徴の一つである。



A	FEDCBA
B	FEDCB
C	FEDC
D	FED

A	<G>CBA
B	<G>CB
C	<G>C
D	<G>

図 6: 経路集約の例

7 補足

7.1 BGP に関する RFC

- RFC1105: バージョン 1 について規定
- RFC1163: バージョン 2 について規定
- RFC1267: バージョン 3 について規定
- RFC1654: バージョン 4 の最初の仕様について規定
- RFC1771: 現バージョンについて規定

7.2 BGP-4 を実装しているソフトウェア

BGP-4 を実装しているソフトウェアには例えば以下のものがある。

- gated(http://www.gated.org/)
- zebra(http://www.zebra.org/)
- radix

講義では、gated の実装について簡単に説明した。

8 まとめ

本レポートでは、AS 間の経路情報、到達可能性情報を交換するプロトコルである BGP-4 について説明している。本レポートは、RFC1771 に基づき、かつ分散システム特論の講義で用いた発表資料をベースに作成したものである。発表資料の内容と比較すると、岡村先生からご質問があった「I-BGP と E-BGP」に関する記述が若干だか追加されている。