EGPsの負荷検出による動的経路制御*

井澤 志充†　　篠田陽一‡
北陸先端科学技術大学院大学§
情報科学研究科

概要

現在のEGPs(Exterior Gateway Protocols)は、経路上の到達性のみを扱っている。しかし、マルチホームのASが増えるようになると、全ASから構成されるネットワークも複雑になるため、最適な経路制御のためにはEGPsはリンク状態に応じて経路を変更するような仕組みが必要となる。各経路における負荷も考慮した経路制御を行なうことは、ネットワークにおける負荷集中を避けるために必要となる。そこで本研究では、EGPsに対する動的負荷分散機構の拡張を提案し、そのしくみを提案するものである。

1　背景

インターネットは数々の組織のネットワーク同士を繋ぎあわせることによって巨大なネットワークを形成している。この場合の一組織とは、ネットワークにおけるポリシによって区別される。ポリシとは、ネットワークの運用に関する方針のことである。たとえば商用ネットワークとは接続しないなどである。インターネットにおいてポリシを共にする組織、あるいはその集合を指してAS(Autonomous System)と呼ぶ。例えばJAIST(北陸先端科学技術大学院大学)はWIDE Projectというの一部である。つまりインターネットはASの集合から成っているといえる。また、ASにはそれぞれ固有の番号が振られている。

1.1　AS間経路制御プロトコル

あるAS内のホストから別のAS内のホストへ宛てたIPパケットが到達するためにはそれらのASが隣接していない限り、いくつかのASを経ることになる。つまりあるASへ向けたパケットをどこへ向けて転送すればいいかという情報が必要になる。この情報をやり取りするために用いるのがEGP(Exterior Gateway Protocol)とよばれる経路制御プロトコルである。以前はEGPと呼ばれるプロトコルが使用されていたが、現在はBGP(Border Gateway Protocol)が使用されている。また、あるASは、隣接するあるASからは直接パケットをやり取りしたくないかも知れない。このプロトコルの一つの機能は、ルーティングにポリシを強いることが出来ることである。BGPでは、

*Dynamic routing with load detection in EGPs
†Yukimitsu IZAWA
‡Yoichi SHINODA
§School of Information Science, Japan Advanced Institute of Science and Technology

AS番号の列挙であらわされるAS-PATHでその到達性情報を保持する。この情報はインターネット上でのASの接続状態を再現するに充分な情報である。また、このBGPを用いて通信するプロセスをBGPスピーカと呼ぶ。一つのASに複数のBGPスピーカを設置することもでき、そのスピーカそれぞれが通信し合うことによってスピーカ間の情報の同一性はとることができる。この際に使用するBGPをIBGPと呼ぶ。

BGPを用いた現在のAS間の経路制御プロトコルは、AS-PATHによって表現される経路の「到達性」とポリシのみによって経路を制御している。

1.2 インターネットの変化

近年、インターネットは発展するに従ってその形態を変えてきた。以前はバックボーンがあり、それに各組織が接続するといった形態が主だったので図1のような木構造を形成していた。

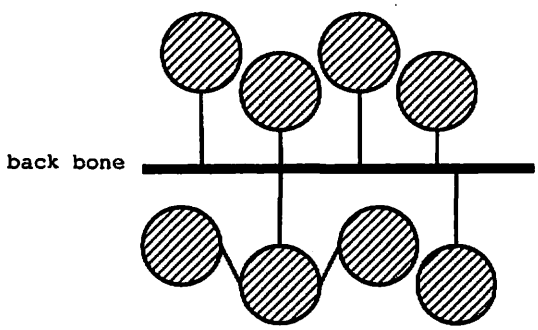

図 1: バックボーンを中心とした木構造

しかし今日ではより複雑に組織間が接続している。この傾向は多数のISP(Internet Service Provider)の台頭や各組織のポリシの変化によるところが大きい。ISP同士の接続や、以前ならばポリシによって接続しなかった組織同士もポリシが変化することで接続するようになってきている。インターネット自体が発展する限り、網としての複雑化はより一層強まると思われる。図2。

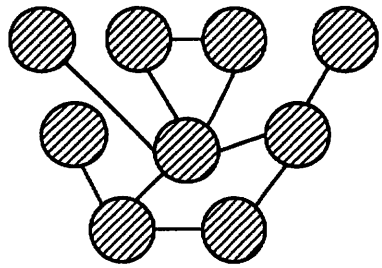

図 2: 複雑化するネットワーク

2 研究の目的

前述したようにAS間の接続が複雑になるに従って、到達性のみで経路を制御するのは限界がある。現状では、あるASを通過するのに時間がかかったり、パケットロスを起こしても、AS間ルーティングにはフィードバックされない。従って、たとえ混雑しているルートでも、そちらが目的地への到達性において最も良いとされれば、そこを通ってパケットがルーティングされてしまう。それを回避するには人間がこれを判断し、手作業で別のPATHを使用するように設定する必要がある。

しかし、本研究ではBGPに対しリンク情報という概念を追加することによりルートが混雑している場合、自動的に代替ルートの発見及び経路の再設定を行なうことを目的とする。

図3に概念図をしめす。

実線によって1〜5のASが接続していることを表している。実線の波線はパケットの通過をしめす。破線はリンク状態が混雑

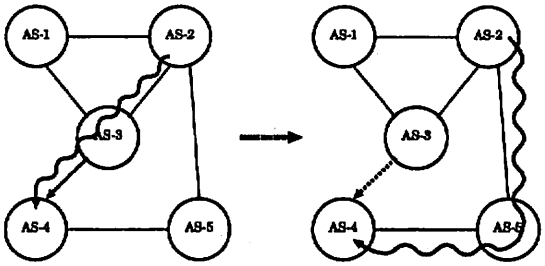

図 3: 代替経路を用いた負荷分散

していることをしめす。左側の図では AS2 から AS4 へパケットが到達するのに AS3 を通過している。しかし、AS3 と AS4 の間のリンクが混雑している。そこで当方式の機能によって右側のような状態となる。AS2 から AS4 へは AS5 を経由するように変更されている。混雑したリンクを使用しない代替経路を発見・適用している。

3 概要

現在、AS 間の経路情報の交換に使われている「到達可能である」という情報だけでは最適なルーティングを行うのには不足で「現在のリンク状態」という概念も同時に取り入れる必要がある。当方式で用いるモデルでは「リンクステートデータベース」という形で「現在のリンク状態」を保持する。

当方式は BGP をリンクステート型の経路制御プロトコルにすることを目的としているわけではない。最適 PATH 選択の際に用いる情報を増やし、より現在の状態に見合った PATH 選択を行えるようにするのが目的である。

従って、この仕組みは BGP への機能追加という形で実現する。この仕組みの基本的な概念を説明する。まず AS 内部で BGP スピーカ同士がその負荷情報を交換し合う。次に、その負荷情報を「現在のリンク状態」として隣接 AS の BGP スピーカへ伝達する。各 AS でこの情報を交換し合うことによって、各 AS はこの情報を元により適切な AS-PATH の選択を行うことが出来る。つまり、負荷の高い AS を通らないような「代替 PATH」を発見し、そちらを使用することが出来るようになる。

また、AS-PATH を変更したことの履歴をもち、これの評価をフィードバックすることにより、際限なく経路変化が起こり続けないようにする。

4 アーキテクチャ

BGP スピーカは BGP の枠組のなかで「現在のリンク状態」を、隣接する AS の BGP スピーカに伝達する。当方式では BGP の KEEP ALIVE メッセージを使用して伝達する。「現在のリンク状態」を伝達する為の新しい attribute を定義する。この新しい attribute を便宜的に「リンクアトリビュート (以下、LSA)」と名付ける。

このメッセージを受けとった BGP スピーカは内部に持つ「リンクステートデータベース (以下、LSDB)」へ受け取った情報を追加する。

この情報によって伝達されるリンク状態によって、使用する AS-PATH を変更する。

負荷の高い経路を使用している場合には RIBs(Routing Information Bases) を検索し、自分のポリシに従った代替経路を検索する。代替経路がある場合は、特定のパケットに対しては現在の経路のかわりとするなどの方法によってある特定の AS への流量を減少させ、結果的にネットワーク全体の負荷を減少させる。

図 4 と 5に、IBGP と LSA の伝達の仕方を表す。 この図は 3 つの AS に接続している

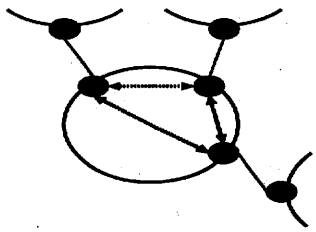

図 4: IBGP でリンク情報を交換

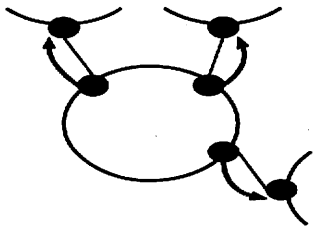

図 5: LSA を用いて隣接 AS へ伝達

AS を表している。斜線の小さい円は BGP スピーカを表す。また、斜線の矢印は同じ AS の BGP スピーカ同士がリンクステート情報を交換し合う様子を表している。この後、図 5のように他の AS の BGP スピーカへこの情報が伝達される。

経路の変更については変更履歴に基づき閾値を決定する。この閾値を越える変更要求にのみ、経路を変更する。

4.1 LSA とルーティングポリシ

AS 間のルーティングは、ポリシに沿ったルーティングであることが最も優先される。これは AS の性質上ポリシルーティングが破られてはならないためである。これは到達性のみならず、リンクステートにも当てはまる。つまり、空いている経路でも、ポリシを破るような経路選択は行なわない。

4.2 リンクステート情報について

同じ AS の BGP スピーカは IBGP によってリンクステート情報を交換し、それぞれが同じリンクステート情報を持つことが求められる。この時に交換する情報は、

- 隣接する AS の番号
- 負荷

である。この情報を元に LSA を生成する。LSA はある AS における全ての隣接リンクの負荷情報であるといえる。隣接する AS へ伝達する LSA は、以下のような構成になる。

- この LSA を発信する AS の番号
- この LSA の ID
- この AS の負荷情報
  - 隣接する AS の番号
  - 負荷
    - 負荷の原因になっているパケットの発信元の AS 番号
    - ⋮

LSA の ID は、この情報が生成された時刻に基づく情報である。複数の ID を比較することによってどちらがより新しいかを区別するために用いる。

4.3 負荷の検出について

負荷の検出方法はいくつか考えられるが、本研究ではルータ (BGP スピーカを兼ねる) の出力に関する負荷を対象とする。その方法として出力キューを監視するやリンクによってはパケットの衝突率を監視する方法も考えられる。

4.4 LSDB について

BGP スピーカにはルーティングの為にいくつかの AS-PATH が保持されている。この AS-PATH に現れる AS に対する LSA を LSDB に保持する。LSDB は、AS 番号の対とその負荷、そしてその負荷の原因となるパケットの発信元 AS 番号の組の集合で構成される。

- あるリンクの、負荷を検出した側の AS の番号
- そのリンクの相手側の AS の番号
- 負荷
- 負荷の原因となるパケットの発信元 AS 番号のチェーン

LSDB に、高い負荷のリンクを検出したら、使用する AS-PATH の再選択のアルゴリズムを呼び出す。

4.5 AS-PATH の選択

閾値を越える負荷を LSA によって検出した場合、以下の手順で AS-PATH 選択の再計算を行なう。

1. RIBs から、ポリシによってフィルタリングした結果を得る。
2. その中から負荷のかかっている AS-PATH 以外の経路がある場合、現在経路と代替経路との選択のための閾値を代替経路側へずらす。
3. 代替経路のない場合は経路変更はしない。

このうち 2 のフェーズはさらに以下に続く。

- 閾値が移動した結果、別の経路を使用することになった場合、
  1. その LSA によって知らされた AS 番号で表される発信元からのパケットは代替 AS-PATH を通すように変更する。
  2. 変更履歴に変更事項を追加する。以前この経路が選択されたことがある場合、その評価を行ない、ヒステリシスをもつ制御関数の 2 つの値をそれに応じて変更する。
- そうでなければ経路の変更は行なわない。

4.6 BGP スピーカの内部構造

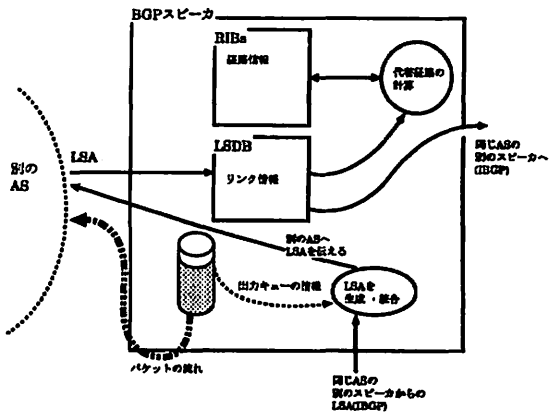

図 6: BGP スピーカの内部構造

図 6はあるルータ (BGP スピーカ) の内部構造を表している。一定時間に一回、BGP スピーカは出力キューの状態から LSA を生成する。これを同一の AS 内の全ての BGP スピーカで交換する。これは IBGP を用いる。その後、この情報を LSA として隣接 AS へ伝達する。

また、隣接 AS より LSA が伝達されると、それを元に LSDB を再構成する。その

時に負荷のかかっている経路を自分が使用している場合、代替経路の発見に努める。その結果、代替経路を発見した場合はこれをIBGPを用いて各BGPスピーカで交換する。

5 スケーラビリティについて

ここで、あるASがLSAが、全AS(平均距離40とする)に行き渡るのに要する最大の時間を試算する。KEEP ALIVEメッセージは通常は30秒毎に発行されるメッセージである。従って、

$$30 * 40 = 1200(秒)$$

20分で全てのASに伝播する計算である。つまり、あるLSAを発信してから次のLSAを発信するまでの時間的間隔はこれ以上であれば十分であることがわかる。

次に、このメッセージが費やすバンド幅に関して計算する。このLSAのみを含むKEEP ALIVEメッセージのサイズは、以下の式で表すことが出来る。隣接するASの数をNとし、負荷の原因になっているパケットの上位Sまで含めるとする。

$$M = 19 + 2 + 4 + 3N * 2S$$

3つの隣接ASがあり、負荷の原因になっているパケット、上位3つを伝達する場合は、

$$19 + 2 + 4 + 3 * 3 * 2 * 3 = 79(byte)$$

となる。これは十分にスケールするサイズである。

6 フィードバック

BGPスピーカは、経路の変更履歴を保持する。この履歴を持つことによって、ある経路変更が有効だったかどうかの判断をすることが出来る。これによって経路変更のための閾値を決定する。図7に閾値の変更のための制御方法を表す。

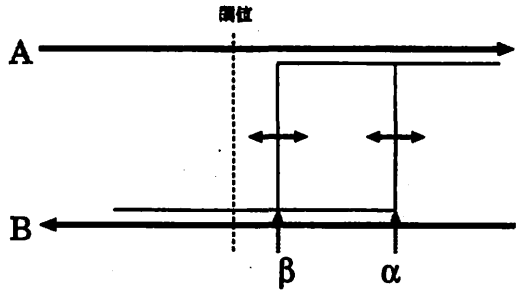

図 7: ヒステリシスをもつ制御

AとBの二つの経路間において閾値と図のような制御によって経路の変更の判断を行なう。現在は経路Bが選択されている。閾値がαを越えないと経路はAに変更されない。また、経路がAからBになるには閾値がβより小さくなる必要がある。また、経路変更に対する評価関数によってαとβは変化する。このようなヒステリシスをもつ制御によって、AS-PATHの変更による全体の系としての挙動を安定させる。

7 今後の予定

今後は、さらにこれらの仕組みについて考察し、これらの仕組みを実現する。

- 負荷検出機能およびリンクステートによる動的な経路制御を、実現するBGPスピーカの実装
- 経路変更への評価関数の作成
- 評価方法に関する検討およびシステムの評価

8 参考文献

- Radia Perlman, "Interconnections-Bridges and Routers",Addison Wesley Publishing Company, 1992
- Y. Rekhter T. Li , "A Border Gateway Protocol 4 (BGP-4)", RFC1771,March 1995
- Y. Rekhter P. Gross , "Application of the Border Gateway Protocol in the Internet",RFC1772, March 1995
- J. Moy, "OSPF Version 2", RFC1583, March 1994
- P. Traina, "Experience with the BGP-4 protocol", RFC1773, March 1995
- P. Traina, "BGP-4 Protocol Analysis", RFC1774, March 1995