

自律システム相互接続点における経路制御運用の自動化システムに関する研究

渡邊秀文 土本康生 中村修 村井純
{robo,tsuchy,osamu,jun}@sfc.wide.ad.jp
慶応義塾大学政策メディア研究科

概要

自律システム相互接続点における経路制御において、接続 AS 数の増大はルータ管理コストの増大、ポリシー設定作業の複雑化をもたらしている。また、AS を超えた正しい経路情報伝播の監視も困難なものとなる。本稿では、経路制御サーバ、経路制御レジストリ、及びそれらを応用した運用技術を用いて、BGP Peering セッションの削減、BGP ルータ自動設定、及び、効果的な AS 間経路監視を実現する経路制御運用の自動化システムの考察を行った。

Survey for Automation System of Routing Operation in Exchange point of the Autonomous System

Hidefumi Watanabe Yasuo Tsuchimoto Osamu Nakamura Jun Murai
Keio Univ.
{robo,tsuchy,osamu,jun}@sfc.wid.ad.jp

Abstract

As the number of connected autonomous system increased, it causes the increase of cost for router administration, and complication of configuration for policy of routing in the exchange point of autonomous system. In this paper, the automated operational system with Routing Registry and Route Server for inter-domain routing is examined. Therefore, BGP peering session is decreased, BGP router can be configured automatically, and effective monitoring system is realized.

1. はじめに

インターネットは、自律システム (Autonomous System: AS) と呼ばれる独自に運用されているネットワークが相互接続された形で構成されている。AS は、単一の明確に定義された運用ポリシーを持っている

BGP(Border Gateway Protocol)[6]などの外部経路制御プロトコルは、ポリシーに基づいて経路制御情報を計算するための手段を提供している。しかし、それらはポリシー情報を共有するための仕組みを提供していない。したがって、ネットワーク運用者間での協調作業や障害の検知が非常に困難である。これは、インターネット全体に及ぶ問題ともなりうる。

そうした問題に対処する方法として、いくつかの運用技術が開発されている。本稿では、これらの運用技術に関する現状調査及び提案を行う。

2. 問題点

BGP を前提にした AS 間経路制御を運用するには以下の問題点が生じる。

- mesh peering(図 1)によるルータ設定・管理のための作業量の増大
- BGP のポリシーに基づくフィルタ設定が複雑なものとなる
- 伝播(propagate)されている BGP 経路監視が困難である

その解決策として、経路制御レジストリ (Routing Registry:RR)、経路制御サーバ

(Routing Server : RS)、ルータ設定ツール、及び、経路監視のための運用技術が開発されている。

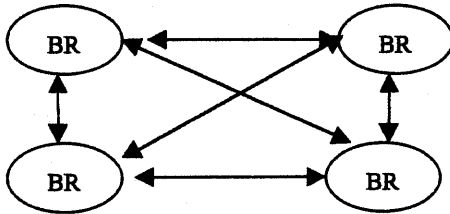


図1 Mesh Peering

2.1. mesh peering によるルータ設定・管理のための作業量

AS が新たに接続点に接続された場合、他の AS のルータすべてが新しく接続された AS のルータと peer の設定を行う必要がある。3 層モデルでは 1 対 N の運用管理となり、N の増減に対する作業量は $O(N)$ となるのに対して、2 層モデルでは N 対 N の運用管理となり、N の増減に対する作業量は $O(N^2)$ となる。そのため、2 層モデルでは、より一層ルータ設定の作業は困難なものとなる[7]。

2.2. BGP ルータ設定

ルータの BGP 設定が複雑となる原因には以下の 3 つが挙げられる[2]。

1. 各 AS に数多くの BGP スピーカが存在する場合、すべてのルータで一貫性を持たせるための設定作業は困難となる。
2. AS の相互接続点では、各 BGP スピーカはピアとなる AS(以下、「ピア AS」と略)に対して別々にポリシーの設定を行なう。そのため、相互接続点では AS 数に比例して設定作業が増大する。
3. 使用するルータによって、別々の仕組みや文法で設定を行う必要がある。

第一の問題点は、各 AS に数多くの BGP スピーカが存在する場合、すべてのルータで一貫性を持たせるための設定作業は困難とな

る。なぜなら、各 AS が外部 AS と複数箇所で相互接続する場合、AS 内において一貫したポリシーで経路制御を設定しなければ、AS 内の経路制御に矛盾が生じる。

ここで AS に外部 AS と接続する境界ルータ (Border Router : BR) が一台追加された場合、他の BR 全てがその設定を変更しなければならない、BR が数多く存在する AS では、その変更作業は困難となる。

第二の問題点は、各 AS は他の AS に対して別々にポリシーの設定をしなければならないので、AS 間を接続する相互接続点ではピア AS が増大するに比例して、設定作業は困難となる。

第三の問題点は、使用するルータによって、別々の仕組みや文法で設定を行わなければならないことである。そのため、第一の問題点と第二の問題点で指摘した原因に加えて、ネットワーク運用者は、ルータの種類によって別々の仕組みや文法を習得しなければならないため、ルータの設定が困難になる。

2.3. BGP 経路監視

自 AS から外に向かう経路の安定性については把握できるが、他 AS から自 AS への経路の安定性については保証できない。図 2 を例にすると、AS 2 が保有する R2 や AS 3 が保有する R3 に AS1 の管理者は、アクセスできない。そのため、AS2 から流れてくる経路、AS3 から流れてくる経路についての安定性は把握できないのである。

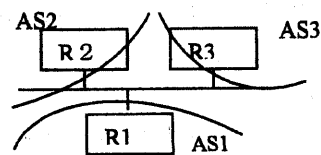


図2

経路フラップを含めて正しく経路制御情報が伝播されない異常を検出するためには、他 AS からの経路の監視が必要になる。

3. 既存の自律システム間経路制御運用自動化へのアプローチ

3.1. 経路制御レジストリ

経路制御レジストリ(Routing Registry: RR)とは、AS の経路制御ポリシー情報を含む AS のネットワークに関する情報が登録されたデータベースである。RR は、経路制御ポリシー情報を共有し、AS を超えてネットワーク運用者の協調作業を支援するための仕組みを提供している。

RIPE181[3]

RIPE RR は、以下のようなデータベースオブジェクトで構成されている。

- Autonomous System オブジェクト
- Route オブジェクト
- AS Macro オブジェクト
- Community オブジェクト

RR を構成する主なオブジェクトは、

“Autonomous System (aut-num)”及び“Route”オブジェクトである。aut-num オブジェクトは、AS へのコンタクト(連絡先)情報及び AS の経路制御ポリシーが記述される。このオブジェクトで記述されている情報は、各 AS が直接制御するポリシーの全体像がグローバル・ポリシーとなる。そのため、すべての aut-num オブジェクトを結合した情報から、理論上正しい任意の AS 間の AS_PATH[6]を推測できる[3]。Route オブジェクトは各 AS で外部に告知する個々の経路について記述されている。

Routing Policy Specification Language [1]

Routing Policy Specification Language (RPSL) は、RIPE81 や RIPE181 の限界を克服するかたちで、IETF の Routing Policy System ワーキング

グループで開発が進められた経路制御ポリシー記述のための言語である。RPSL はポリシーの詳細及びルータの設定に必要な各種パラメータを十分に記述できるため、RPSL で記述されたポリシーからルータ設定の生成を可能としている。

3.2. 経路制御サーバ

各プロバイダーの BR は相互接続点で他の全ての BR と Peering session を確立している。このような状況は“mesh peering”[5]と呼ばれる(図 1)。一方、“star peering”[5]は接続点で各 BR が経路制御サーバ(Route Server : RS)とのみピアする状態を指す(図 3)。

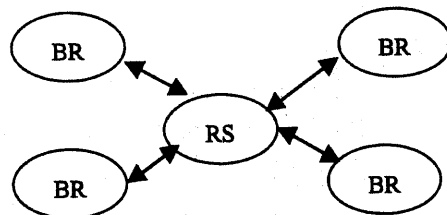


図 3 Star Peering

RS は前述 の RR からデータを取得し、各 BR 用の経路受け入れ及び経路告知のポリシーを設定する。そして、各 BR は経路制御サーバに対して選択した経路を告知する。BR から告知された情報を使用して、経路制御サーバは BR に代わって BGP 経路の計算を行なう。その結果、star peering は、mesh peering と同様の機能を実現できる。

経路制御サーバの実装・設計例を以下に示す。

RSd

RSd¹は、南カリフォルニア大の ISI(Information Sciences Institute)で開発されている経路制御サーバである。現在、アメリカのいくつかの IX (Internet eXchange) で実際

¹ <http://www.isi.edu:80/div7/ra/>

に運用されている。RSdは、BGP ルータと peer を張るだけに留まらず、IX に接続されている各 AS のルータそれぞれに対して “VIEW”[4]と呼ばれる RIB (Routing Information Base) [6]を構築する機能を持つ。そのため、接続点に接続されている各ルータは、mesh peering を設定せずに、mesh peering と同等な経路制御情報の交換が可能になる。つまり、上記の star peering を実現する機能を持つ。現在、Merit Network では、“Route Server Next Generation” Project²として商用ベースで Merit の提供する RSd への接続サービスを開始している。

BGPd

BGPd は、zebra³と呼ばれる分散経路制御システムの一部として開発されている。Zebra は、以下の特徴を持つソフトウェアである。

- 1プロトコルを1プロセスで処理する
- カーネルの経路表及び異なるプロトコル間の経路情報の再配布をzebra daemonによって提供する
- zebra daemon以外のdaemonはプロトコル操作に用いられる。
- 新しいプロトコル用のdaemonを容易に追加できる
- BGPdにBGP経路制御サーバとRoute Reflector[9]が統合されている。

RSdがGateD⁴を元に開発されていたのに対して、BGPdは新しい観点から開発されている。

² <http://www.rang.net/>

³ <http://www.zebra.org>

⁴ <http://www.gated.org>

現在、ZebraはNSPIXP⁵において実験的に接続ASのルータとPeerが行われているが、RSdのようにVIEWを構築してStar Peeringを提供するサービスには対応していない。

RFC1863方式の経路制御サーバ[4]

full mesh routing の問題を解決する方法として、経路制御サーバの技術を提唱している。

RFC1863 は、経路制御サーバ (RS : Route Server) と経路制御サーバクライアント (RC : RS Client) の通信だけではなく、同一の RC を共有する経路制御サーバ群 (RSC: Route Server Cluster) 間の通信も定義している。

4. 既存の運用自動化システムの検証

以下に AS 間経路制御の運用で自動化すべき点を示す。

4.1. BGP Peer Session の削減

前述した通り、BGP Peer Session の削減は、経路制御サーバの VIEW を用い Star Peering の仕組みを利用することで実現できる。現在は、RSd のみ VIEW に対応している。自律システムの接続点に RSd と同様な VIEW 機能を持つ RS を設置することで BGP Peer Session の削減を実現できる。

4.2. BGP ポリシー設定自動化

第 2.2 節で指摘した通り、BGP ポリシーの設定は困難である。この解決方法として、BGP ルータ設定の自動化が有効である[2]。自動化の流れとして、第一に登録されている自 AS のポリシーを RR で検索し、その検索結果を BGP 自動設定システムによって設定に反映する。第二に、各ピア AS に対するポリシーを RR で

⁵ <http://xroads.sfc.wide.ad.jp/NSPIXP/>

検索し、検索結果からルータ設定に必要なデータを抽出する。第三に、各ベンダ別の設定の仕組みや文法を低レベル・ルータ設定記述言語と位置付け、RPSLなどを高レベル・ルータ設定記述言語と位置付ける。そして、RRではRPSLのような高レベル記述言語でデータを保持し、それを各ベンダ別のルータ設定に変換する仕組みを提供する。設定者は、ピアASのAS番号のみを入力するだけでよい。

上記の仕組みを実現したシステムとしてRtconfig[1]やbgpcf[10]などが挙げられる。RtconfigはRSdと連動して、いくつかのIXで実際に利用されている。bgpcfは、WWWブラウザ上で利用可能なルータ設定ツールである。

4.3. BGP 経路監視

BGP 経路を効果的に監視する仕組みとして、以下の手法が考えられる[8]。

- BGP ルータを公開し監視支援を行う
DIGEXのLooking Glass⁶などのようにNAP(Network Access Point)に接続されたBGP ルータが公開 BGP ルータとなり、WWWをインターフェースとしてBGPの経路情報、経路フラップに関する統計、ping/traceroute コマンドが利用可能である。
- RSが持つ情報の活用
IPMA(Internet Performance Measurement and Analysis)⁷では、RSにピアを設定しているASの接続トポロジや経路の安定性を視覚化したASExplorerや受け取った経

路更新回数を統計化するRoute Flap Status Generatorを提供している。

5. 自動化システムの提案

5.1. VIEW 機能を持つ経路制御サーバの活用

経路制御サーバの導入により、これまでに述べてきたPeering sessionが削減され、以下の効果を実現できる。

- オペレータの管理コスト削減
管理コストの削減とともに煩雑な作業を削減するため、設定ミスを防ぐ効果もある。
- BRの負荷削減
BRはパケット転送機能とBGP経路計算機能を分割できるため、BRの負荷を削減できる。
- BGP経路情報の取得
経路フラップ等の経路制御の異常検出効果や経路数等の経路制御に関する統計情報を取得できる。

但し、経路制御サーバの実装は以下の2点を考慮しなければならない。

- 接続AS数に応じた耐久性及び負荷分散の実現
 - BGPdやMRT⁸のような柔軟的な新しいプロトコルの追加機能の実現
- これら2点の実現により、一層安定し拡張性の高い接続点における経路制御を実現できる。

5.2. BGP 自動ルータ設定システム

BGP自動ルータ設定は以下の効果をもたらす。

- 設定に必要な情報のみ自動的に検出できる
- BGP Filter 設定の自動化が実現できる。

⁶ <http://nitrous.digex.net/>

⁷ <http://www.merit.edu/~ipma/>

⁸ <http://www.merit.edu/~mrt/>

- 新モジュールの追加により、BGP の仕様の更新やルータベンダ別の文法への対応が容易になる。

RtConfig では、特定のルータベンダの文法にしか対応していないため、汎用性を持つ RPSL からルータ別文法への変換の仕組が必要となる。

5.3. BGP 経路監視のための仕組の確立

BGP 経路監視を効果的に行う仕組を以下に示す。第一に、接続点における BGP ルータを公開し、AS のオペレータが自分以外の AS のルータにアクセスできる環境を用意し、他 AS から伝播される経路情報の取得を可能とする。ただし、経路情報公開を公開したために、情報取得のためのルータへのアクセスが、ルータの packets 転送や経路情報交換などの処理に悪影響を及ぼさないための制限が必要となる。

第二に、RS の経路情報を活用する場合、接続点管理組織が接続 AS のオペレータが必要とする情報提供を可能とし、経路監視システムの構築を実現できる。

6. 考察

経路制御サーバは、BR の管理・設定コストの削減、及び、経路監視システムの構築が可能とする。また、RR は、BR の自動設定システムを実行する際のポリシー・データベースとして活用することや、登録されている経路情報と実際の経路情報を比較することに使用し、経路監視にも利用できる。

RS や RR の活用は、自律システム相互接続点における経路制御運用自動化システムを実現し、安定的かつ効率的な運用体制を支える重要な基盤となる。

7. 結論

本稿では、RS 及び RR を利用した自律システム接続点における経路制御運用手法について述べた。RR や RS を利用した運用システムは、接続 AS のオペレータ間における情報共有の仕組を提供し、AS を超えた経路制御運用の協調作業の促進し、AS 間経路制御の安定性を実現する。

現在、BGP ルータ設定ツールについてはプロトタイプの実装を行った。今後の課題として、第五節で述べた提案を実現する実装を進める。

謝辞

本研究の機会を与えていただき、有益な助言をくださった慶応義塾大学村井純教授、中村修博士、および土本康生氏に感謝します。

参考文献

- [1] Cengiz Alaettinoglu, Tony Bates, Elise Gerich, Daniel Karrenberg, Marten Terpstra, Curtis Villamizer, "Routing Policy Specification Language(RPSL)", draft-ietf-rps-rpsl.txt, 1996
- [2] Cengiz Alaettinoglu, "Scalable Router Configuration for the Internet", Information Sciences Institute University of Southern California, 1996
- [3] T. Bates, E. Gerich, L. Joncheray, D. Karrenberg, M. Terpstra, J. Yu, "Representation of IP Routing Policies in a Routing Registry (ripe-81++)", RFC 1786, 1995
- [4] D. Haskin, "A BGP/IDRP Route Server alternative to a full mesh routing", RFC 1863, 1995
- [5] Ramesh Govindan, Cengiz Alaettinoglu, Kannan Varadhan, Deborah Estrin, "Route Servers for Inter-Domain Routing", USC/Information Sciences Institute, 1995
- [6] Y. Rekhter, T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, 1995
- [7] WIDE Project, "商用インターネット相互接続実験", WIDE プロジェクト報告書 1995 年度第 2 章, 1996
- [8] 上水流由香, "AS 間経路制御監視システムに関する一考察", 情報処理学会研究報告マルチメディア通信と分散処理, 1996
- [9] T. Bates, R. Chandra, "BGP Route Reflection", RFC 1966, 1996
- [10] 渡辺秀文, "自律システム相互接続点における経路制御運用技術に関する研究", 卒業論文, 1996