第 12 部

ポリシールーティング

第 1 章

はじめに

この部では、新しい形の経路制御として、研究とインターネット網での実用化が進んでいる政策的経路制御 (policy routing) についての研究のこの１年の成果をまとめた。

従来の経路制御の研究では, 網上の各ノード間のコスト (cost) あるいは距離 (distance) と呼ばれるパラメータの値を基にして, 情報の流れの発信元から宛先ノードまでのなんらかの総コストを最小にするような経路を作成するものがほとんどであった. これらは, 一つの経路制御方針のもとに運用されている網内では, 確かに有用である. ここで紹介する政策的経路制御 (policy routing) は, 従来の経路制御では把握できなかった, 運用方針の異なる網の相互接続のためのものである.

WIDE ポリシー・ルーティング ワーキング・グループは、昨年３月に研究活動を始め、以来、米国における研究の追従とともに、わが国のインターネット網環境を考慮にいれながら、経路制御に反映されたい具体的な方針（policy) を考えてきた。

本章では、この研究分野への紹介を行なう。以下、1.1. では政策的経路制御の背景を探り、1.2. では、この１年の当ワーキング・グループにおける活動状況を報告する。最後に 1.3. では、次章以降の構成を述べる。

1.1 背景

米国 ARPANET から始まった広域網環境は, 大西洋域衛星網 (SATNET) [111] など他の広域網との接続のためのインターネット・プロトコル (IP) の出現により, さらに大きく発展し [112], 現在の地球規模に近いインターネット網にいたるまでになった. こうした環境の中で, いくつかの政策的経路制御手順の出現の背景には２つの流れがあった； 網規模の拡大問題解決とアクセス・コントロールの必要性である.

７０年代から８０年代にかけてのローカル・エリア・ネットワーク (LAN) やワークステーションの発達によるローカル・サイトの網環境の充実に伴い, いまや, インターネット網は広域だけでなく, ローカルな物まで含め, その規模の拡大は, 従来の網制御および管理機能では, 把握できなくなっていった. 経路制御においては, ノード数の増加などの規模の拡大問題は, 階層化された域 (domain) に分けることによって回避される [113]；すなわち, 域内 (intra-domain) では, ある経路制御がローカルに行なわれ, 域間 (inter-domain) の経路制御は, 各域を１ノードとみなすような１段階上のレベルで行なわれる. インターネット

網においては, 初期には, Exterior Gateway Protocol (EGP) [114] が, 域間 (inter-domain) の経路制御手順として生み出された. 当時は, まだ, ARPANET 時代から相変わらず, 網管理は一貫して Bolt Beranek and Newman Inc. (BBN) で行なうという状況 [115] で, EGP では, 域間 (inter-domain) 経路制御の階層も, 最上階は BBN 管理下のノードがコア・システムを形成し, その下に各サイト管理下の網が, 木構造 (tree structure) でつながるように決められていた. ループ問題を回避するための, この強制された物理的な木構造 (engineered tree structure) は, その後発展を続けるインターネット網の任意な相互接続の構造にそぐわなくなっていった. それまでのインターネット網は, 米国の国防省の援助によって構築されていたが, 学術系の The National Science Foundation(NSF) による主要網 (backbone) が構築されるに至り, 異なる運用方針を持つ主要網や地域網間の経路制御は単純に階層化できず, 対等な自律した網システム (Autonomous System) 間の制御が必要となった. こうした背景から, EGP を発展させた形で, Border Gateway Protocol(BGP) [116] は生まれた.

もうひとつの流れは, 運営や管理の方針の異なる網が相互接続されている環境での, 各網での網資源 (network resources) の使用制御 (access control) の必要性である. あるホストから, 世界中のサイトの網資源へのアクセスが可能という地球規模での広域環境は, 様々な使用制御の問題を生み出してきた. この分野では, Deborah Estrin が, MIT 時代から続けている異組織間の網 (inter-organizational networks) における使用制御の研究 [117] が根底にあり, また MIT ／ LCS(Laboratory for Computer Science) の David Clark の提唱した policy routing (ここでは政策的経路制御と翻訳) [118] がある. この背景から, BGP よりも, さらに徹底した形の政策的経路制御手順の Inter-Domain Policy Routing (IDPR) [119] が作成された.

EGP も BGP も, 中間システム経路決定型 (hop-by-hop) の経路制御のための情報交換手順である. 経路情報は, distance vector と呼ばれる形式のもので, 情報提供ルータは, 自分が到達可能な網のリストと各網への次のノードを指定する. これに対して, IDPR は, 域単位での発信元経路決定型 (source-specified routing) の経路制御のための経路制御手順体系である. パケットの発信元域から宛先域までの道があらかじめ準備され, その道に沿って, パケットが送られる. 経路情報は, link state 形式で, 各域の情報が, 基本的には flooding で他の域に届けられる. この他, この２つの流れの妥協案 [120] として, 最近, Source Demand Routing Protocol (SDRP) [121] という手順が指定された. これは, BGP を使用する環境下で, 発信元が宛先までの経路を域単位で指定する経路制御手順である. SDRP のための情報交換手順などは, まだ検討されている最中である. また, BGP のさらにスーパーセットとして, Inter-Domain Routeing Protocol(IDPR) が現在, ISO およびインターネット網環境で, 作成されている.

1.2 ワーキング・グループの活動状況

本ワーキング・グループは、作年３月下旬から活動を開始した。この１年は、我々にとっては未知の分野の研究への出発準備期間であった。WIDE 内では、その創設以来、網

での実用に沿った実践的な研究が中心であった。本グループでの研究もそのような流れに乗るような期待も周囲にはあったように思う。しかし、結果としては、我々はそれまでの研究姿勢といささか異なる形の研究—問題認識を中心に行なった。他の網研究分野に比べ、政策的経路制御は新しく、また色々な意味で、未知数の部分が多い。従ってそこでは実践的なアプローチ以外の方法が必要であった。今でこそ、このように解説できるが、始めた頃は、まさに暗中模索の状態であった。以下に、我々の活動状況を報告する。

当初、ポリシー自体の研究を目指したが、これはきわめて困難であった。わが国のインターネット網では、まだ網資源に対するアクセス・コントロールの必要性についての認識は低く、網を介しての各地への接続性の方が強く求められていたからである。

当時のわが国のインターネット網 [122] の状況は、IP 主要網の増加などによる規模拡大に伴い, それまで使われていた Routing Information Protocol(RIP) [123] は適当でなくなってきていた。次世代の経路制御が求められていた。こうした短期的な需要の面からも、米国などで既に実用化が進んでいた BGP などの政策的経路制御に関する調査は必要とされていた。

そこで、我々は、漠然とした政策や方針についての研究はさておき、具体的な経路制御であった BGP についての学習から始めた。ここで問題となったのは、わが国のインターネット網を、どのように BGP 経路情報を交換し合う AS に区切っていくかであった。これは、今もって、解決されていない。理由は、国内での BGP 実用化にあったって、BGP から、現在の域内経路制御 RIP への接続が難しいため、すぐに BGP へ進むわけにはいかなかったからである。米国のインターネット網では、BGP を使うところでは、域内として、Open Shortest Path First (OSPF) という経路制御手順が使われており、BGP から OSPF への接続については、実装の面で充分な支援がある。従って、現在の段階では、先ず各主要網で OSPF の使用に移り、そして、一時的な域分割で BGP を使ってみることが、試みられようとしている。このような、作業は、実際には、国内インターネット網を運営および管理する人々の集まる組織 Japan Internet Engineering and Planning Group (JIEPG) で、進められている。域分割などの点については、今後、JIEPG、WIDE インターネット運用グループおよび本ワーキング・グループで、連絡をとって進めていくことになるであろう。

こうした実用面とは独立に、本ワーキング・グループでは、必要な政策や方針を解明していく過程として、問題認識を進めていった。実際に網の運用において生じた様々な問題点をみていくことで、それらに潜んでいる網の運用の方針や政策が浮かび上がってくることを期待した。これらについては、3 章で報告されるが、期待以上の成果がでた。そのひとつとして、強制的な経路制御の必要性から、仮想トンネル — Virtual Tunnel (VT) の研究が始まった。これは、現在、独立したワーキング・グループとして活動している。また、速度の違う網接続の際におこるユーザあるいは終端サイトの希望をかなえようとするところから、Routing by Preference が生まれた。興味深いことに、海外に支社をもつ組織などインターネット網に複数の出入口を持つ組織においても、同様な需要があることが判明した。この Routing by Preference は、来年度の本ワーキング・グループの主要研究課題の一つとなる。

また、既存の政策的経路制御のプロトコルの調査の一環として、IDPR も試みられたが、その複雑性と実用化の遅れにより、BGP よりもやや理解が遅れている。しかし、ここにきてようやく米国で実験段階にはいっていることもあり、このプロトコルの調査は継続され、本ワーキング・グループも実験にも加わっていく。この他、米国では、BGP 体系の中の発信元指定経路のプロトコル、SDRP の研究も進んでいることから、この調査もされている。

Routing by Preference については、この 3 月、特別のチームを結成し、我々独自の具体的な解決法を検討しはじめている。

1.3 部の構成

以下、2 章では既存の政策的経路制御のプロトコルを紹介し、3 章では、我 WG 独自の研究成果として、網運用における経路制御の諸問題を提起する。この問題認識を原点として、我々の研究は進んでいる。4 章ではさらに今後の研究課題を検討し、5 章でこの部のまとめと、今後の研究計画を報告する。

第 2 章

既存のプロトコル

2.1 概要

本章では、既存の政策的経路制御のプロトコルのうち、実装段階にあるもの、BGP、SDRP、および IDPR を紹介する。

2.2 用語について

2.2.1 域 (domain)

政策的経路制御では, 網資源の使い方に対する政策や方針を唱える域の定義が必要となる. この域間で, 経路情報が交換されからである.

EGP や BGP の仕様書の中では, この域は, 自律システム (Autonomous System), 略して AS と呼ばれている. もともと, EGP の中では, AS とは, その中では, 一つの AS 内経路制御手順をもつものとして定義付けられている. しかし, その後, BGP では, インターネット網環境の発展と共に, 意味が広がり, 複数の AS 内経路制御手順が存在しても, その AS の外からみると, 一つの経路制御手順で動いているかのように, まとまりのある群として見えるものとして再定義されている.

これに対し, IDPR では, この域は, 管理域 (Administrative Domain), 略して AD と呼ばれている. これは, 網資源の運用や管理についての方針や政策が, 一つのところで決められているものを指す.

この 2 つの言葉の違いには, それぞれの手順の出てきた背景の違いが, 反映されている [124]. EGP や BGP は, 網単位の経路制御から生まれてきた背景があるのに対し, IDPR は, 資源使用制御 (access control) の背景からでてきたので, 管理組織単位の意識が強くでている. しかし, 現在では, AS も AD も, 各手順の仕様書のなかでは, ひとつの管理政策に基づくまとまりとして扱われている. 厳密にいうと, この言い方も正しくない. なぜなら, 例えば, 異なる管理組織がまとまってひとつの地域 AD を形成している例もある. 従って, 正しくは, AS あるいは AD は, インターネット網を構成する連続な一部分で共通の網資源管理方針のもとに運営されているものと解釈できるであろう；ここで, 連続な一部分とは, その部分内では, AS 内経路だけで, ある地点から任意の地点まで到達できる範

囲という意味である.

以下では, 政策的経路制御の単位域を AS あるいは AD と呼ぶが、どちらもひとつの運用・管理方針の下の網集合の域という意味を持つものする。

2.2.2 ホストとルータ

本論文では, インターネット網環境の経路制御について述べているので, 終端システムをホストと呼び, 中間システムをルータと呼ぶ. ルータの内, 特別な経路制御手順動作を行なうものをそのプロトコル名をつけた形で呼ぶ. すなわち, BGP の仕様書で指定してあるボーダ・ゲートウェイ (Border Gateway) を BGP ルータと呼び, IDPR のポリシ・ゲートウェイ (Policy Gateway) を IDPR ルータと呼ぶ.

2.3 経路制御に反映されたい政策や方針の種類

網経路制御に反映されたい政策や方針には, 通過パケットに対するものと, 終端 AS として自 AS から外へ発信されたパケットに対するものがある.

通過パケットに対する方針には, 例として網資源の使用に対する制約や課金の条件などがある. 終端 AS の方針には, 安全性, 遅延時間, 速度などの他, 信頼度, 使用可能な度合い, 使用料金などを含めた網サービスの種類 (Type of Service) による通過網の選択や, もっと人間レベルでの政策的な通過網の選択などがある.

Estrin によると [125], 通過, 終端の両タイプの AS には, 一般に, 次のような種類の方針・政策がある:

- **終端サイト指定による規制:** パケットの発信元と宛先の組合せにより, 網資源使用を制限する. この発信元や宛先の単位は, AS であったり, 終端システムであったり, ユーザクラスであったりする.

- **部分経路指定:** 終端 AS の方針例として, 宛先までの経路に通ってほしい, あるいは避けたい AS や複数の AS を指定する. 通過 AS の方針例として, 自分の好まない AS や複数の AS を通過したあるいは通過しようとするパケットは通さない.

- **網資源の質 (Quality of Service):** 例えば, 高速, 低遅延時間の特別なリンクは, 選ばれたところだけが使用できるであろう. このように, 網資源の質によって, その資源への規制は行なわれる.

- **ユーザ・クラス:** 様々なユーザのクラス識別ができることにより, AS 単位よりも細かい, あるいは, AS から独立した単位での資源使用制御を行なうことができる. このクラスとは, 使用制御のためのグループである.

- **網資源保証:** 通過 AS 方針例として, 選ばれたところへは, ある程度の網資源の質を保証する.

- **時間に関わる制限:**　時間帯によって資源の使用を制限する.
- **アプリケーション別使用制御:**　パケットが, どの種類のアプリケーションに関わっているものかによって, 網資源の使用を制御する.
- **限られた網資源**限られた資源を活用するために, ある時間内に, 発信元から出されるパケットの量を制限する.
- **認証と完全性 (integrity):**　特別な資源使用の制限には, 鍵などを使用した本格的な認証を必要とする場合もある. また, 完全性が検証された情報に基づく経路しか通さないという方針もあるだろう.
- **課金についての方針:**　課金に際しては次のような要因を考慮されるであろう:
  - 課金の単位 (例： 日本円, 米国ドル)
  - 課金対象の単位 (例： 一様なレート, キロ・バイトごと, キロ・パケットごと)
  - 実際の値段（例：.50 ドル／メガ・バイト）
  - 支払う人と支払われる人（例：パケットの発信元, あるいは宛先, パケットを送りこんできた隣接 AS, 第三者など）
  - だれの測定したパケット数を課金のために使うか（例： 発信元, 宛先, 通過 AS, 経路上のその他の AS）
  - 課金の限度（例： 支払者が使える金額の限度, あるいは, 通過 AS が転送するパケット量の限度）

これらの要因を組み合わせて, 各 AS の方針や政策は表すことができるであろう. 可能な組合せのすべての方針が, 実践的とは限らない. 仮に, ある AS が, 「自分の AS 内での使用に邪魔にならない程度なら, 他の AS に通過 AS として使ってもらってもよい. 」というような方針をもったとしよう. この場合**邪魔にならない程度**は, 動的に変化し, 大きなインターネット網では, 実践的でない. 細か過ぎるユーザ・クラスも問題である. それらは, 終端システム間あるいは AS 内の網レベルで解決されたほうが良いだろう. また, セキュリティーの問題も完全に解決することも難しい. この点については, 後で検討する.

2.4　BGP — Border Gateway Protocol

BGP は AS(Autonomous System) 間で経路制御情報を交換するためのプロトコルである. BGP は EGP(Exterior Gateway Protocol) の経験に基づいて設計されているので, EGP のいくつかの欠点を解決している. また, 初歩的な運営方針を反映した経路制御を実現するのに十分な手段を提供している.

インターネットは, AS が任意のトポロジで相互に接続されたものであるとモデル化できる. この時, 経路制御を行なうプロトコルは, AS 内で用いられる経路制御プロトコル

と, AS 間で用いられる経路制御プロトコルに分類できる. 例えば, 現在日本で用いられている RIP(Routing Information Protocol) は AS 内で用いられる経路制御プロトコルである. (日本の各サービスプロバイダは唯一の AS 内で用いられる経路制御プロトコル –RIP– を用いているため, 結果として日本のインターネットは 1 つの AS を構成している.) この分類からいえば, BGP は AS 間で用いられる経路制御プロコトルである. AS 内で用いられる経路制御プロトコルの主な役割は, AS 内において安定した通信を確保することである. また, AS 間で用いられる経路制御プロコトルの主な役割は, 各サービスプロバイダの運営方針 (ポリシー) を反映した経路制御 (ポリシールーティング) を実現することである.

従来は AS は、一つの経路制御プロトコルを用いている範囲として定義されていたが、現在では複数の経路制御プロトコルが使用されていることも珍しくない。そこで、AS は、一貫性のある管理下におかれたネットワークの範囲として定義されている。つまり、AS の境界のゲートウェイへの到達性と AS 内部への到達性が等しいということである。

AS 間で用いられた初期の経路制御プロトコル –EGP– には以下のような欠点があった.

(1) EGP は IP の上位層として実現されていた. TCP 等の安定したプロトコルを用いていなかったために, EGP のパケットが失われることがあった.

(2) EGP は ARPANET のコアモデルを念頭に設計されたプロトコルである. よって, AS 間の接続は木構造に制約されていると仮定している. この制約のもとでは, EGP は十分に機能する. しかしながら, EGP はループを検出する機能を提供していないため, AS 間で任意の接続をとることができない. AS 間の接続を木構造に制約することで, EGP はループを防いでいたということもできよう。

(3) EGP のメトリックには取り決めが無い. よって, あるネットワークに対して, 複数の経路から異なったメトリックを得たとしても, それらを比較することができない.

BGP はこれらの EGP の欠点を次のように解決している.

(1) BGP は TCP を用いる. よって, BGP を用いているゲートウェイ間の通信は保証されている.

(2) BGP はメトリックとして, BGP のメッセージが経由した AS の列 (AS Path) を用いている. あるゲートウェイが受け取った AS Path の中に, そのゲートウェイが属している AS 番号があれば, そのメッセージを用いた経路はループすることが分かる. つまり, AS Path はループを検出するために十分な情報を提供するので, AS に任意の接続を許すことができる.

(3) 複数の経路から受け取った AS Path は, その AS の運営方針に基づいて比較することができる.

さらに, BGP には次のような長所がある.

(4) BGP は変化のあった情報のみを交換する. よって, 全ての情報を交換するプロトコルに比べ, 帯域的に有利である.

(5) AS Path は初歩的なポリシールーティングを実現するのに十分な情報を提供している.

現在の日本のインターネットは全体で 1 つの AS を構成しているために, 各サービスプロバイダ間の独立性が低い. また, 日本のインターネットの直径は, RIP において到達不可能を意味する 16 ホップを越えている. そこで, 日本インターネットを複数の AS に分割する必要がある. 各サービスプロバイダの接続形態は木構造に制約されていない. また, 初歩的なポリシールーティングを実現する必要がある. よって, 現時点では AS 間の経路制御プロトコルとして BGP を選択することが良いと考えられる.

現在 BGP は version 3 が規定されており [126] Draft Standard Protocol として認定されている [127]。また、CIDR [50] に対応するために version 4 も IETF で提案されており、Internet Draft としてまとめられている。

2.5 SDRP

SDRP(Source Demand Routing Protocol) は、BGP 及びそのスーパーセットである IDRP(Inter-Domain Routing Protocol) によって提供されるルーティング情報を補足し、発信元によるドメイン間のルーティングをサポートするプロトコルであり、現在 Internet Draft である。現在の draft では、パケットフォーマットと SDRP パケットの生成、パケットの forwarding について規定してあり、SDRP route の生成については未定義となっている。

SDRP での IP パケットの転送を説明する。ドメイン間の接続を行なっている Router を Border Router(BR) とすると、図 2.1のようなモデルとなる。

domain A の一般ホスト a から発信したパケットは、domain A 内の IGP によるルーティングにより、ホストの属するドメインの外部とのゲートウェイである BR c に送られる。

BR c は、SDRP によって転送するか、従来の BGP/IDRP により転送するか決定し、SDRP の場合、SDRP Route(ドメイン単位の source route) を決定し、もとの IP パケットに SDRP Route を組み込み、SDRP パケットにカプセル化する。SDRP Route は domain B, domain D, domain E のようなものになる。

BR c は SDRP パケットを SDRP Route のドメインの BR(この例では domain B の BR d) に送る。 SDRP パケットを受けとった BR は、ドメインの transit policy を調べ、transit policy を侵している場合はエラーにする。

domain E の BR j にて SDRP Route をすべて終了した場合、 SDRP ヘッダをとり、もとの IP パケットを通常の IP routing にて行き先マシン k に送る。 SDRP Route を完了していない場合、SDRP Route 上の次のドメインへパケットを送ることになる。

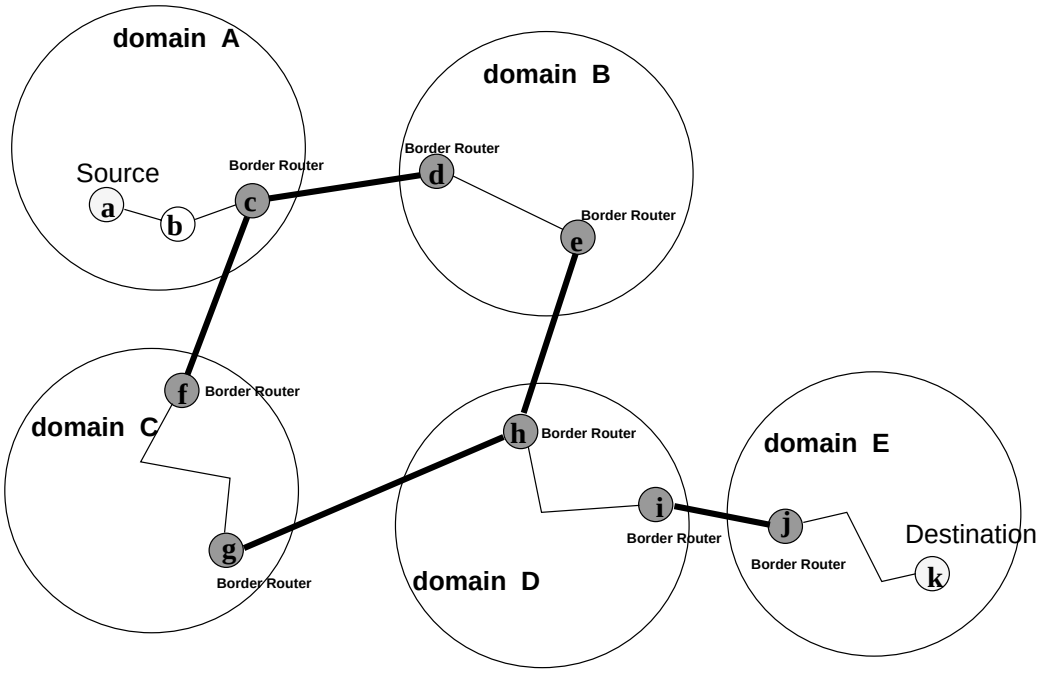


図 2.1: SDRP のモデル

BR は BGP/IDRP により学んだ情報と設定情報から Domain-Forwarding Information Base(ドメインレベルのルーティングテーブル,D-FIB) を作る。SDRP パケットは FIB(BGP,IDRP でのルーティングテーブル), D-FIB により複数のドメインを転送される。

SDRP route の決定方法全般はまだ指定されていない。transit policy は、IP ポート番号、IP アドレスの組をもとに調べると書かれている。

fragment オーバーヘッドを減らすため、すでに MTU を学んでいて、細分化する必要があるときは、SDRP Packet をつくる前に細分化 (fragmentation) する。

SDRP の特徴としては、経路情報の交換については BGP/IDRP に依存していること、まだ詳細が決まっていない点、ドメイン単位な点があげられる。

SDRP の問題点としては、SDRP Route の決定方法がまだみえないこと、ユーザーごとの経路制御などについては言及していない点であろう。

2.6 IDPR

IDPR(inter-domain policy routing) は、データメッセージを送信する AD(administrative domain) が、自らのポリシー (Source Policy) とこのデータメッセージを通過させる AD のポリシー (Transit Policy) の制約の範囲内で宛先 AD までの経路を確立し、この経路を通してデータメッセージを宛先 AD に届ける手段を提供するプロトコルであり、Internet Draft として IETF に提案されている。IDPR は, 網資源への使用制御 (access control) の観点から, それまでの経路制御手順から全く独立に生まれた経路制御手順体系である. 経路制御にリンクステート、メッセージ転送に送信元経路指定のアルゴリズムを用いる。

IDPR の機能は次のとおりである:

- 経路制御情報の配布と収集
- 経路制御情報と Source Policy に基づくポリシールート (AD 単位での経路) の生成と選択
- 選択したポリシールートに沿ったパス (paths: IDPR ルータ単位での経路) の立上げ
- 立上げたパスを通したデータメッセージの転送
- 経路制御情報、既に立ち上げられたポリシールート、転送制御情報、自 AD の構成情報よりなるデータベースの維持

各 AD はその AD を通過する外からのパケットに対する AD の網資源についての使用制御の方針 (Transit Policy) を基本的に flooding により他の AD へ配布する。flooding 以外にも, request-reply 形式の経路情報収集方法も装備されている. 各 AD は, 自 AD の transit policy と隣接 AD への接続 (connectivity) 情報よりなる経路制御情報を生成し, 配布する. 各 AD は経路制御情報の配布先を制限することにより, 自 AD の網資源を利用する AD を制限できる. また、経路情報の配布制限を確実に行うため、その転送経路を指示する (source specified fowarding) こともできる。その AD 内から外へ出ていくパケットに対する方針 (Source Policy) は他の AD には公開されない.

各 AD が主張できる policy の種類は BGP よりもかなり多い. Transit Policy として、次のようなものが指定できる：

- 一定の AD またはユーザクラス (研究, 商用) を送信元または宛先とするトラフィックに適用されるような、アクセス制限
- 遅延、スループット、エラー特性等の提供できる網サービスの質 (Quality of Service),
- バイト、メッセージ、単位時間当りの課金等の金銭コスト

Source Policy には, 次のようなものが考えられている：

- 経路として望ましい、または避けるべき AD 等のアクセス制限
- 許容できる遅延、スループット、信頼性等の要求品質
- 許容できるセッションコスト等の金銭コスト

IDPR の経路制御を司るものには、仮想ゲートウェイ (Virtual Gateway, VG), IDPR ルータ — 正式にはポリシー・ゲートウェイ (Policy Gateway, PG), パスエージェント (path agent)、経路サーバ (Route Server)、マッピングサーバ (Mapping Server)、構成サーバ (Configuration Server) がある。VG は、IDPR における AD 間の仮想的なリンクで、実際には、直接接続された 2 隣接 AD 内の IDPR ルータの集合。隣接 AD 間に 2 つ

以上の VG がありうる。IDPR ルータは、VG 内の物理的ゲートウェイで、経路制御情報の収集・分配、パス立上げへの参加、データメッセージ転送、転送制御情報データベースの維持、を行う。パスエージェントは、ホストに代って経路を選択し、パスを立上げて管理し、転送制御情報データベースを維持する。この機能は IDPR ルータに組み込まれていることが多いと考えられている。

経路サーバは、経路制御情報及びポリシールートのデータベースを維持するとともに、経路制御情報と、構成情報または直接パスエージェントから得られる送信元ポリシーを用いてポリシールートを生成する。Mapping Server は、IP 名とアドレスを AD の ID にマッピングするためのデータベースを維持する。既存のネットワーク管理システムに組み込むことができる。構成サーバは、AD 内の IDPR ルータ、パスエージェント、経路サーバにおいて使用される構成情報のデータベースを維持する。構成情報には、送信元ポリシー、通過ポリシー、AD 内 IDPR エンティティの IP アドレスとのマッピング情報などがあり、Domain Name System (DNS) 等の既存のネームサーバに組み込むことができる。

パケット転送は, 送信元経路指定で行なわれる. 送信元 AD はデータメッセージ転送の前にパス立上げを次のように行う。AD 内からのパケットが外の目的地へ送り出される際には, 1) そのパケット自体の要求品質 (Quality of Service), 2) その AD 内の方針, 3) 目的地までの通過 AD の方針などを考慮して, 経路サーバが経路を通過すべき AD と VG を指定した AD レベルでの経路を設定する. その AD 経路に沿って各ホップとなる IDPR ルータが次のように決定される. 送信元 AD の path agent は中間 AD の IDPR ルータに対して、パス識別子 (ID)、要求サービス、構成 AD とその通過ポリシー等よりなるポリシールート情報を送信してパス立上げを要求する。中間経路制御エンティティはデータベースの転送制御情報を変更し、パス ID を next hop に変換できるようにする。各データメッセージはパス ID を含む。中間経路制御 エンティティはこのパス ID により next hop を決定する。

この経路は閉鎖されるまで, 発信元 AD から宛先 AD までのパケット配送に使われる. いつ閉鎖されるかは, 各 AD の動作方針によるが, 一般にこれは一回ごとののパケット転送とは独立に行なわれる. すなわち, あるパケット転送時には, 目的地までの経路がすでに存在していれば, それを使用することができ, これによって, 経路の開閉のためのオーバーヘッドが少なくなる. 発信元ホストからでた IP パケットは, 発信元 IDPR ルータで, 経路 ID などのヘッダをつけた IDPR データメッセージとして宛先 AS まで送られる.

経路制御情報の配布や, 経路設定など, 網経路管理用のメッセージは, IDPR 用のトランスポート層プロトコル, Control Message Transport Protocol (CMTP) を使って送られる. CMTP は, request-reply 形式の connectionless なプロトコルで、再送機能がある.

2.7 章のまとめ

本章では、既存の政策的経路制御のプロトコルを解説した。BGP は hop-by-hop の経路制御系の経路情報交換のプロトコルで、IDPR は AD レベルでの発信元経路設定の経

路制御系のプロトコル体系である。現在のところ、米国インターネット網では、BGP が広く実用化されており、IDPR は実験段階にある。

この他、Inter-Domain Routeing Protocol(IDRP) が BGP のスーパセットとして開発されているので、今後もこうした動きを追っていこうと考えている。

第 3 章

問題提起

3.1 概要

本章では、今年度の我ワーキング・グループの最大の研究成果である経路制御における諸問題の提起を行なう。ここに上げる 5 問題は、IP パケット網上に仮想トンネルを作成し経路の強制指定の必要性を訴える問題、そして続く 3 問題は、最適経路選択の問題で、最後にこれらを統合して、AS 間 (inter-AS) の環境で一般化した最適経路選択問題、別名もとのり問題を定義する。

3.2 金魚モデル問題 (Virtual Tunnel)

3.2.1 金魚モデルにおける経路制御

ふたつのネットワークがリンクを共有する場合の経路制御を考えてみる。図 3.1 において、a, b 発 h 宛のパケットはできるだけ自ネットワークを通過させたいというポリシー、つまり以下のポリシーを実現したいとする。

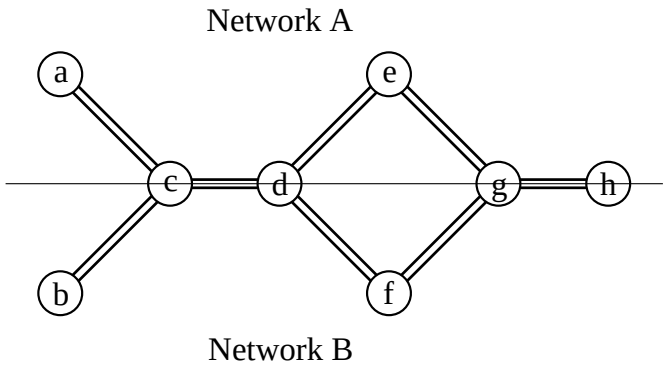
(1) a ⇌ h 間の経路: **a** – c – d – **e** – g – h

(2) b ⇌ h 間の経路: **b** – c – d – **f** – g – h

このポリシーを実現する方法として次が考えられるが、それぞれ問題点をかかえている。

1. d においてソースアドレスを考慮した経路制御を行なう。
   問題点 1: この機能がないルータが多い。
   問題点 2: この機能を有効利用できる経路制御プロトコルがみあたらない。

2. 例えば b – f 間に新たにリンクを張る (図 3.2)。
   問題点: リンクを共有している意味がなくなる。



Network A のゲートウェイ:　**a**, c, d, **e**, g, h
Network B のゲートウェイ:　**b**, c, d, **f**, g, h

図 3.1: 金魚モデル

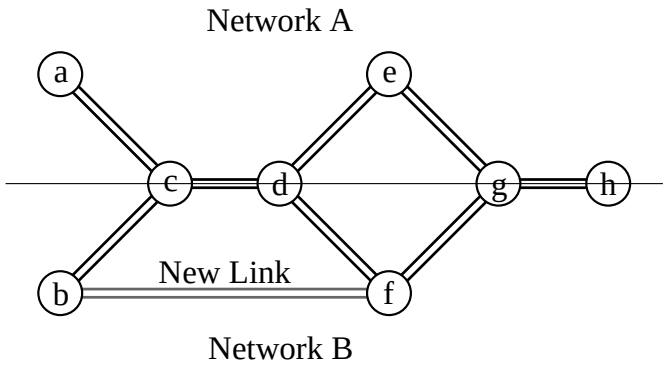


図 3.2: 金魚モデル問題を新たなリンクで解決した場合

3.2.2 トンネリングによる解決

上記 2 項において b – f 間に新たにリンクを張る代わりに、トンネリングを利用して仮想的なリンクを張る方法が考えられる。トンネリング機能をネットワーク・インタフェースの形で提供できれば、b, f における経路制御アルゴリズムを全く変更しないでトンネルを仮想リンクとして利用することができるようになる。この方法により上記の問題は解決できるであろう。

ただし、利用に際しては例えば以下のような注意事項が発生する。

1. b – c または c – d がダウンしたとき、b – f を代替経路として用いることはできない。
2. d – f がダウンしたとき、b – f のトンネルは予想外の経路を通ってしまい、ポリシーを実現できなくなる。
3. トンネルを実現するために、b – f 間で直接交換されるパケットは、トンネルを通すのではなく b – c – d – f の経路を通らなければならない。

以上のように、トンネリングによる仮想リンク (ネットワーク) を提供する仮想インタフェースを、初期には “vt” (Virtual Tunnel) と呼び、実装も行なった。“vt” という名称は Virtual Terminal と紛らわしいため、現在は用いていない。

3.2.3 ddt-WG に発展

考えてみると、トンネリング技術はインターネットワーキング技術の随所で利用されながらも裏方の汚い技術とされてきた。ところが上記のアイデアによりトンネリング技術に光を当てて表舞台の技術とすることができる可能性が出てきた。また、トンネリング技術は Layer Violation の技術でもあるため、プロトコルの階層構造モデルに対する挑戦となる可能性もある。

そこで vt のアイデアを理論的に発展させることを目的として ddt-WG が設立された。ddt-WG では理論面の研究だけでなく、既存の実装を用いた実験ネットワーク構築も計画している。詳細は 11 部を参照されたい。

3.3 RIP における最適経路選択問題 (The original Motonori Problem)

日本の IP インターネットは、JAIN, TISN, WIDE 等のネットワークを中心に発達し、現在では SINET, JOIN その他地域ネットワークも稼働を開始している。これらのネットワークは、それぞれ相互接続を行ない IP の相互乗り入れを可能としている。

これらのネットワークのうちのいくつかは複数箇所で相互接続を行なっているが、複数箇所で相互接続を行なう場合には、通信の速度や容量などのコストを考慮した最適な経

路選択や、各リンクに流れるトラフィックを均衡化することによる負荷分散が行なわれることが望まれる。

経路選択への要件の表現であるポリシーは経路制御プロトコル上に実現されることになるが、日本のインターネットにおいては、その創世期から今日まで経路制御プロトコルとして RIP (Routing Information Protocol; RFC 1058) を利用しており、暫くの間はこのまま RIP を利用し続けていく可能性が高いため[1]、RIP を用いたポリシーの実現に対する考察について述べる。

3.3.1　RIP メトリックの調整

RIP では、あるネットワーク[2]に到達するためのコストを 1 から 15 までの値を持つメトリックによって表現する。メトリックは目的ネットワークに到達するまでに通過するゲートウェイの数 (ホップ数) に対応し、目的ネットワークまでの経路が複数存在する場合には、メトリックが小さい方の経路が選択される。

したがって、複数存在する経路のうち特定の経路が優先的に選択されるようにするためには、そのリンクの方向から来る RIP のメトリックが他の経路から来る RIP のメトリックよりも小さくなくてはならない。

この条件が満たされていない場合には、

- 優先させたい経路に関する RIP のメトリックを他の経路のメトリックより小さくなるように調整する (redistribute)。

- バックアップとさせたい経路に関する RIP のメトリックを優先させたい経路のメトリックよりも大きくなるように調整する。

このうち redistribute では、経路情報にループが発生し、正しくない経路情報が流れてしまう可能性があるため避けるべきである。

また、RIP のメトリックは 15 が最大値であり 16 が到達不能であることを示すため、RIP を用いているネットワークの規模が大きくなるにしたがってメトリックを増加させて対処するという方法は現実的でなくなる。現在、日本のインターネットではすでに最遠のネットワークまでのホップ数が 16 を越えるところが存在するが、海外への出口のある東京をほぼ中心としたネットワークのトポロジーになっていることから、海外を示す default やその redistribute 等によって中心圏まで出てこれれば、なんとか到達できるという状況にある。したがって、メトリックを増加させる方法もあまり勧められない。

さらに、RIP のメトリックは単に通過するゲートウェイの数を示す値であり、ネットワークトポロジに変化が発生するたび容易に変動する。このため、メトリックが変動す

[1] WIDE Internet では 1993 年 4 月から順次内部プロトコルを OSPF に、対外プロトコルを BGP に変更していく予定である。

[2] ここでネットワークとは、RIP の経路情報アナウンスの単位であるひとつの IP ネットワークアドレスをもつホストの集合体を意味する。

る毎にネットワーク管理者がその時点での各経路のゲートウェイの数を考慮して RIP のメトリックの再調整を行なわなければならないという問題がある[3]。

3.3.2 RIP preference の導入

ワークステーション上のルーティングソフトウェアとして広く利用されている gated では Release 2 から preference という概念が導入された。これは、同一のネットワークに対する経路情報が複数存在する場合、特定の経路情報が到達可能を示していればメトリックにかかわらずその経路情報を優先的に採用できるようにするものである。

ただ、gated の仕様では、RIP による経路情報に関しては preference の値は固定されており、RIP の経路情報は得られたゲートウェイにかかわらずメトリックの値のみによって経路選択が行なわれるようになっている。そのため、RIP 同士の経路情報の間に対しても preference に基づく経路選択をするためには、図 3.3のパッチを当てる必要がある。

```
*** rip.c.ORIG  Sat Sep  5 17:14:35 1992
--- rip.c       Sat Sep  5 17:15:59 1992
***************
*** 635,641 ****
--- 635,647 ----
                        if ((metric >= RIPHOPCNT_INFINITY) || (rt->rt_state
 & RTS_HOLDDOWN)) {
                            continue;
                        }
+ #ifdef NO_RIP_PREFERENCE
                        if ((metric < rt->rt_metric) ||
+ #else
+                       if ((preference < rt->rt_preference) ||
+                           (preference == rt->rt_preference) &&
+                           (metric < rt->rt_metric) ||
+ #endif
                            ((rt->rt_timer > (rt->rt_timer_max / 2)) &&
                             (rt->rt_metric == metric) && !(rt->rt_state &
 (RTS_CHANGED | RTS_REFRESH)))) {
                            if (rt_change(rt,
```

図 3.3: RIP 間で preference に基づく経路選択を行なうための gated 2.1 へのパッチ

パッチを当てた gated では、gated.conf において図 3.4のような記述が意味を持つようになる。

preference は値が小さいものほど高い優先度を表し、図 3.4の例では 130.54.4.2 から送られてきた RIP の経路情報に含まれる到達可能な経路については、130.54.20.192 からの情報によらず 130.54.4.2 経由の経路が選択されることになる。

[3]WIDE, TISN, JAIN では複数地点で相互接続が行なわれているが、実際にトポロジーの変化が発生する度にメトリックの再調整が行なわれている。

```
accept proto rip gateway 130.54.20.192 {
        listen all preference 80 ;
} ;
accept proto rip gateway 130.54.4.2 {
        listen all preference 60 ;
} ;
```

図 3.4: RIP preference を用いた gated.conf の例

RIP preference を導入する上での前提条件

オリジナルの RIP には経路が変化する際に安定で迅速な状態遷移を可能とするため、sprit horizen with poisoned reverse (RFC1058) という手法がある。

RIP に preference を導入した場合、メトリックが大きくとも到達可能であることを示している場合にはその方向の経路を選択させることが可能になるが、これは逆向きの経路情報をも到達可能を示しているとすると、実際には到達不能である方向であるにもかかわらずその方向の経路を選択してしまい、経路がなかなか収束せずに振動が発生する可能性がある。

したがって、RIP preference を導入する場合には

- split horizen with poisoned reverse (RFC1058) が実装されている
- metric 16 (unreachable) は preference の値にかかわらず他の reachable な経路に優先しない

という前提条件が必要となる。実際に gated はこの条件を満たしているため、この条件を以下の議論における前提とする。

3.3.3　ポリシーの実現

ここで 2 つの AD (Administrative Domain) [4]が複数のリンクで相互接続を行なっているモデルを考える (図 3.5)。図中の L_nは AD 間のリンク、Gx_nはリンクに接続された AD Xのバックボーンを構成するゲートウェイ、Nx_nは AD Xの各ゲートウェイに接続されたネットワークを示す。

ここでは簡単のため、両 AD は一直線のバックボーン構成で、バックボーンを構成するゲートウェイはそれぞれ反対側の AD の対向するゲートウェイに対してリンクを持つものとする。

このとき、少なくとも次に上げる 2 つのポリシーが考えられる。

1. 始点ホストから最も近いリンクを経由させる

[4]ここで AD とは、同一 IP ネットワークアドレスをもつホストの集合であるところのネットワークが、複数集まって構成する一つのネットワーク組織に対応する。

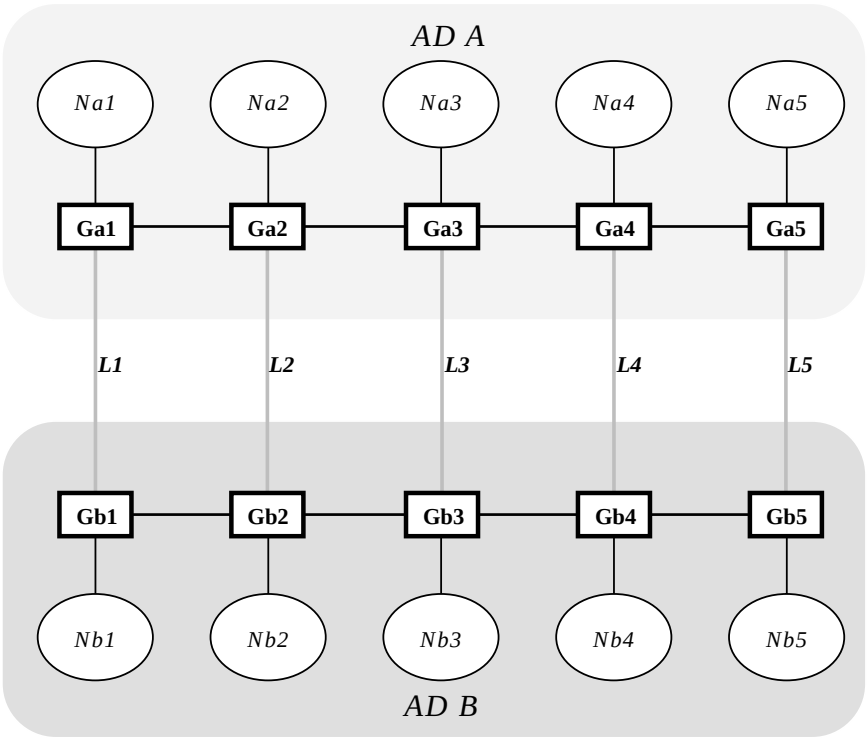


図 3.5: 2 つの AD が複数のリンクを持つモデル

2. 終点ホストに最も近いリンクを経由させる

以下ではこの 2 つのポリシーを RIP preference に基づいて実現するための方法について述べるが、実現にあたっては RIP メトリックの操作は行なわないものとする。なお、どちらの場合においても AD A から AD Bへの経路に注目するものとする。

始点ホストから最も近いリンクを経由

始点となるホスト Ha_nに最も近いリンクを経由して AD Bに到達させる場合には、各リンクのゲートウェイ Ga_nは、対となっている相手側ゲートウェイ Gb_nから送られてくる、AD Bに属するネットワークに関する RIP を高い優先度で受け、他のリンクを経由して到達可能であることを示す、両隣のゲートウェイ $Ga_{(n-1)}$, $Ga_{(n+1)}$から送られてくる RIP は低い優先度で受けとるように設定する。両隣のゲートウェイから受けとる RIP に対する preference が等しく設定してあれば、最寄りのリンクがダウンした場合にメトリックに従って両隣のゲートウェイのうちのどちらかの方向が選択される。

両隣のゲートウェイから受けとる RIP に対する preference の設定が等しい場合、各ゲートウェイは単に各々の持つリンクの方向を高い優先度で選択し、リンクがダウンしている場合にはメトリックのみに基づいて経路選択を行なうことになるため、AD A の内部トポロジーは一直線のバックボーン構成でなく任意のトポロジーでよい。

セカンダリ以下のリンクに対しても優先順位を与えたい場合には、両隣のゲートウェイから受けとる RIP についても順位付けをすることになるが、この場合は AD A の各ゲートウェイが連携した経路選択動作をする必要があるため、次項の議論に帰着できると考えられる。

終点ホストに最も近いリンクを経由

AD A 内部で最も終点ホストに近いリンクまで近付いてから AD Bに進入させたるためには、AD A のゲートウェイは、各々の保持するリンクに最も近い相手のネットワークに関する経路情報を高い優先順位で受けとり、他のリンクに近いネットワークに関しては、保持するリンクから送られてくる経路情報よりも AD A 内部からの経路情報を高い優先順位で受けとる。このようにすることで、各々のネットワークに対して特定のリンクを経由させることができる。

また、どのような組合せでリンクがダウンした場合であっても、終点ホストに極力近いリンクを経由させるためには、それぞれのリンクに順位付けを行ない、AD A 全体としてその順にリンクが選択されるような preference を設定する必要がある。

図 3.5を例にとると、まず相手の AD B内部の各ネットワークがどのリンクに最も近いかを調べ、ネットワーク毎に利用するリンクの優先順位を決定する (表 3.1, 値が小さいほど優先度が高いことを示す)。

この表に基づき、AD Bの各ネットワーク Nb_mに関して Ga_nでは、リンク L_nからの RIP を preference $P(Nb_m, L_n)$ で、$Ga_{(n-1)}$からの RIP を preference $P(Nb_m, L_{(n-1)})$ で、$Ga_{(n+1)}$からの RIP を preference $P(Nb_m, L_{(n+1)})$ で受けとるように gated.conf を設定する。Ga_2における具体的な設定は、図 3.6のようになる。

表 3.1: リンクの優先順位

	L_1	L_2	L_3	L_4	L_5
Nb_1	10	20	30	40	50
Nb_2	20	10	20	30	40
Nb_3	30	20	10	20	30
Nb_4	40	30	20	10	20
Nb_5	50	40	30	20	10

```
accept proto rip gateway Ga1 {
        listen Nb1-list preference 10; # P(Nb1, L1)
        listen Nb2-list preference 20; # P(Nb1, L2)
        listen Nb3-list preference 30; # P(Nb1, L3)
        listen Nb4-list preference 40; # P(Nb1, L4)
        listen Nb5-list preference 50; # P(Nb1, L5)
};
accept proto rip gateway L2 {
        listen Nb1-list preference 20; # P(Nb2, L1)
        listen Nb2-list preference 10; # P(Nb2, L2)
        listen Nb3-list preference 20; # P(Nb2, L3)
        listen Nb4-list preference 30; # P(Nb2, L4)
        listen Nb5-list preference 40; # P(Nb2, L5)
};
accept proto rip gateway Ga3 {
        listen Nb1-list preference 30; # P(Nb3, L1)
        listen Nb2-list preference 20; # P(Nb3, L2)
        listen Nb3-list preference 10; # P(Nb3, L3)
        listen Nb4-list preference 20; # P(Nb3, L4)
        listen Nb5-list preference 30; # P(Nb3, L5)
};
```

図 3.6: Ga_2の gated.conf の設定

AD A のゲートウェイが一直線のバックボーン構成になっている場合、リンクの優先順位は一般にあるリンクを中心として両側に順に単調増加していくように振られることになる。最も優先度の高いリンクの両側に振られる preference 値は等しいものである必要はないが、もし等しく設定してあれば最も優先度の高いリンクがダウンした場合には、メトリックの値のみに従って代替経路が選択される。

理想的には、ダウンしているリンクの両側のリンクに負荷分散されることが望ましいが、例えば L_3がダウンした場合に Ga_3が Ga_4方向の経路を選択したとすると、Ga_1, Ga_2 は Ga_3の方向を選択することになり、この方法ではあるネットワークへの経路として AD A 全体として 1 つの経路を選択することになる。

中央のリンクに最も近いネットワークにあるホストへの経路において、もし中央のリンクがダウンした場合、両側のリンクに負荷分散されることが理想であるが、この場合には両側のリンクのうち片方のみが選択されることになる。

また、AD A のゲートウェイがループを形成している場合、利用されるリンクの優先順位は期待通りとなるが、AD A 内部においてリンクに到達するまでの経路が冗長なものになってしまう恐れがある。例えば、図 3.7において L_1, L_2, L_3の順に高い優先度を与えた場合、L_1がダウンすると Ga_3からは Ga_1, Ga_2という経路を経由することになる。さらに L_2もダウンしている場合についても同様である。

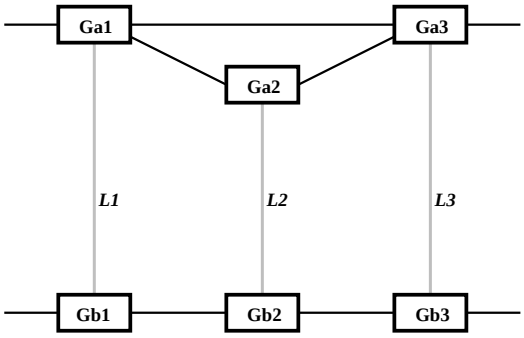


図 3.7: ゲートウェイがループを形成している例

終点ホストに最も近いリンクを経由するような設定を行なう場合は、AD のポリシーを全てのゲートウェイに preference として矛盾なく表現する必要があるので gated.conf の一貫性が重要となる。

3.3.4　RIP preference の問題点

状態遷移の安定性

RIP preference を導入した場合、複数のルータの間で次のような状況が発生すると、経路が不安定になると考えられる。

- 隣り合うルータが相互に相手からの経路情報を他より高い優先度で受けとるようになっている場合
- 3 つ以上のルータがループを形成しており、それぞれが高い優先度で経路情報を受けとる方向が巡回する関係になっている場合

前者の場合は、経路情報の送受のタイミングが完全に同期していることがなければ、割と早くどちらかの方向のみが選択された状態で安定すると考えられる。しかし、後者の場合は際限なく不安定な状態で振動を続けることになるため、このような設定が行なわれないように注意することが必要である。

2 つの安定状態の存在

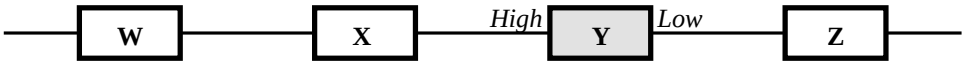


図 3.8: 2 つの安定状態を持つ例

RIP 間で preference による選択ができるようにした場合、ルーティング情報の到達するタイミングの違いにより 2 つの安定状態が存在する。

図 3.8においてルータ W, X, Z は preference による経路選択は行なわず、ルータ Y のみが preference によりルータ X 方向を優先しようとしているものとする。

ルータ Y の起動後、あるネットワーク N に関する RIP を最初に受けとるとき、その RIP がルータ X から送られてきたものであった場合は、preference の設定により X 方向が到達可能である限りは Z 方向から如何なるメトリックの RIP が送られてきても X 方向を選択し続ける。

ネットワーク N に関する RIP をルータ Z から最初に受けとった場合には、ルータ Y は Z 方向を選択し、ネットワーク N に関する RIP をルータ X にフォワードする。ルータ X は preference による経路選択を行なっていないので、純粋にメトリックのみによる比較を行なう。ここで、ルータ Y から送られてきている RIP のメトリックがルータ W からのものよりも小さかった場合には、ルータ X は Y 方向の経路を選択し Y 方向にネットワーク N に関する到達可能な RIP を送ることはない (split horizen)。したがって、ルータ Y は Z 方向の経路を選択したままとなる。

このような現象が発生する条件は、ルータ X の受けとるネットワーク N に関するメトリックが

$$(\text{ルータ } W\text{からのメトリック}) > (\text{ルータ } Y\text{からのメトリック})$$

となるときである。

W 方向の経路が到達可能である場合に必ずそちらの経路を利用させるようにするためには、上の条件が成り立たなくなるまで W 方向に遡っていく際に通過するすべてのルータにおいて preference を用いた経路選択の設定をしておく必要がある。

また、preference を用いて経路選択を行なうルータ Y が、隣のルータ X, Z に対してネットワーク N に関する経路情報を送らない、あるいは、隣のルータ X, Z がルータ Y からの経路情報を無視するような場合、すなわちルータ Y が transit をさせない multi-homed であるような場合は、ルータ Y は近隣のルータの経路制御に影響を与えることなく preference によって自由に経路選択を行なうことが可能である。

ルータ X がルータ Y 方向を優先し、ルータ Y がルータ X 方向を優先するような状況が発生すると、立ち上がりの状態が非常に不安定となる。sprit horizen with poisoned reverse が前提であるので、ルータの RIP の送出間隔の微妙な差によりしばらくすると安定状態に移行すると考えられるが、安定状態に達するのに要する時間が長くなると予想される[5]。したがって、この様な状況が発生するようなルータの設定は避けなければならない。

RIP の限界

あるゲートウェイが選ばれるべき優先順位の高いリンクを通る経路をアナウンスしている場合と、無視されるべき優先順位の低いリンクを通る経路をアナウンスしている場合とがあるにもかかわらず、隣のゲートウェイではそのどちらであるかがアナウンスされている情報からでは判断できない。このため、RIP preference を導入したとしても最適な経路を正しく選択することが困難な場合がある。これは、RIP が link state model でなく distance vector model に基づいた経路制御プロトコルであることによると考えられる。

また、RIP ではネットワークアドレスとそのネットワークまでのコストを示すメトリックのペアの情報しか配布されないため、あるネットワークがどの AD に属するネットワークであるかが経路情報のみからは判断できない。そのため、ネットワークがどの AD の属するものであるかを区別するためには、AD ごとに属するネットワークを明示的に列挙しなければならないという問題もある。

3.4　最適経路選択問題２ (メールの配送経路問題)

富士通は、アメリカに Fujitsu America(FAI) という関係会社を持っている。UUCP 時代には、海外から日本に送られてくるメールは uunet.uu.net を通って日本に入ってきていたので、uunet.uu.net の postmaster に頼んで fujitsu.co.jp 宛のメールは、当時 FAI のゲートウェイだった fai.fai.com に投げてもらうようにし、さらに fai.fai.com から UUCP を使って富士通のメールのドメインマスタまで送ってもらうようにしていた。

その後世の中が IP 接続され、メールの転送に name server の mx レコードを使うようになってきた。富士通では、fujitsu.co.jp ドメインの mx レコードを富士通の WIDE とのゲートウェイマシンである fwide.fujitsu.co.jp というホストに向けて出すようにした。この結果、fujitsu.co.jp 宛のメールは、世界中どこからでも fwide.fujitsu.co.jp にダイレクトに送られてくるようになった。この場合、日本以外の国から送られてくるメールは当

[5] 2 つのルータの間に別のルータが存在し、両方の経路情報を preference なしに受けとるように設定されているならば不安定さは解消する?

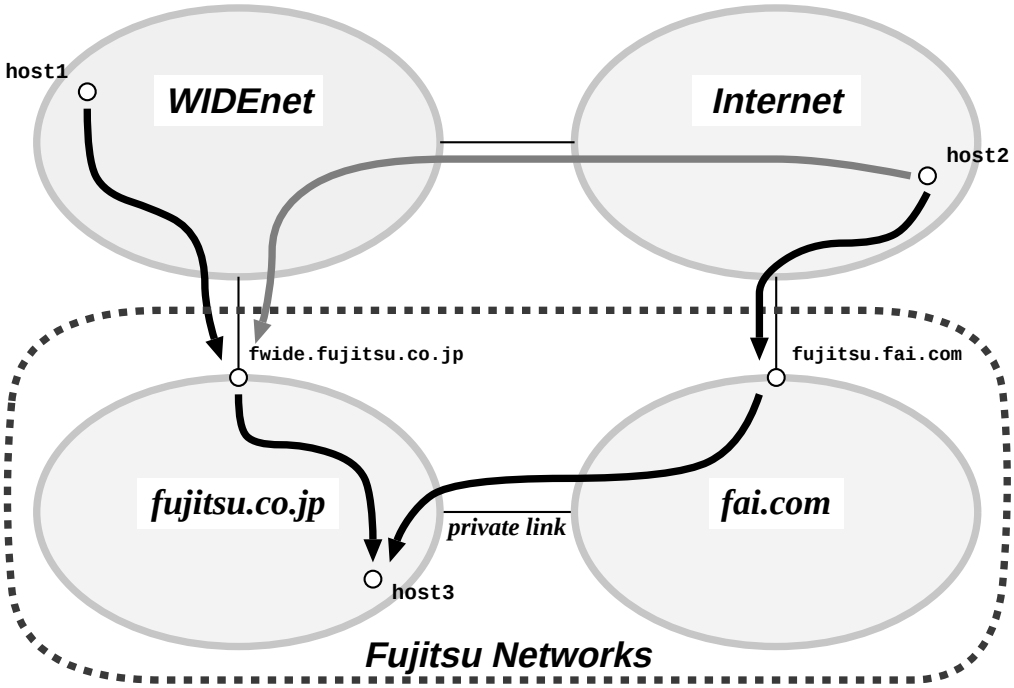


図 3.9: 2 つの AD が複数のリンクを持つモデル

然 WIDE の国際回線を通ることになる (図 3.9 参照).

ところが、FAI もまた BARRNET(米国西海岸のリージョナルネットワーク) に IP 接続し、富士通と FAI は専用回線を使って内部的に直接 IP のリンクを持つようになったので、日本以外の国から fujitsu.co.jp 宛に送られてくるメールを、UUCP 時代のように、fwide ではなく FAI の BARRNET とのゲートウェイである fujitsu.fai.com に送ってもらうようにできれば、メールの転送時間はほとんど変わらずに、WIDE の国際回線にかける負担を減らすことができると考えた。

しかし、UUCP 時代と違い、メールは世界中のメールサーバから fwide.fujitsu.co.jp にダイレクトに送られてくるため、どこか一つのメールサーバの配送ルールだけを修正すれば良いというわけにはいかない。

何らかの方法で、日本以外の国に出す富士通向けの mx レコードだけを fujitsu.fai.com に向けることができれば、海外から fujitsu.co.jp 宛のメールを FAI 方向に送ることができる。幸い、WIDE では日本国内向けの name server と、海外向けの name server が異なっていたので、それぞれの name server に登録する内容を変えることで、この問題が解決できることがわかった。

しかし、同様の問題は fai.com 宛のメールを日本国内から出す場合にも起こり、この場合には、米国の name server が日本向けとそれ以外の国向けで異なっていないため、海外から fujitsu.co.jp 宛のメールの場合のようにうまい逃げ手がない。

メールの場合には mx レコードという特別な機能を使っていたために、name server が

二つ用意されているという特殊な事情ではあるけれども、なんとか解決する方法があった。しかし、一般的な IP 接続の場合には接続先は単なる IP アドレスなので、根本的に IP パケットのルーティング自体を工夫しなければならない。

もし IP パケットのルーティングが、日本国内から fwide.fujitsu.co.jp に接続する場合は WIDE から、海外から fwide.fujitsu.co.jp に接続する場合には BARRNET〜FAI から、というふうにできるなら、上記のメールの転送経路問題も、mx レコードの小細工なしに解決することができる。

3.5　最適経路選択問題３

ここでは、RIP を用いた IP パケットの経路制御により Policy Routine を実現する一実施例について報告する。

3.5.1　ポリシーの定義

まず、図 3.10に示すような、Network-A と Network-B とが、hop count 1 で隣合う Backbone 上の NOC-W と NOC-E のそれぞれに対し IP Link を張る他、互いを接続する Private Link を持つネットワーク構成を考える。

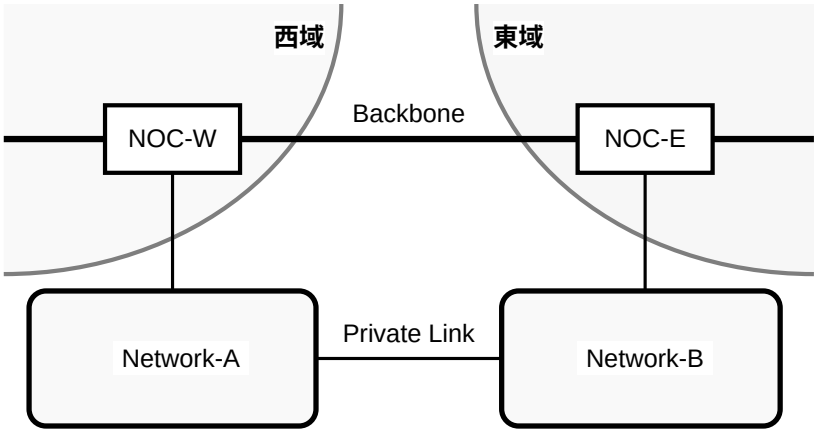


図 3.10: ネットワークの構成

この構成で実現するポリシーを

> Network-A と Network-B とのいずれかが source または destination となる IP パケットは、NOC-W と NOC-E 間の Backbone を通過しない。

と定める。つまり、Backbone 上で NOC-W と NOC-E との間を境として二つの領域に分け、NOC-W を含む領域を西域、NOC-E を含む領域を東域とし、

1. Network-A と Network-B との間の IP パケットは、Private Link を経由する、
2. Network-A と東域との間の IP パケットは Private Link、Network-B、NOC-E を経由し、NOC-W を経由しない、
3. Network-B と西域との間の IP パケットは Private Link、Network-A、NOC-W を経由し、NOC-E を経由しない、
4. Network-A と西域との間の IP パケットは NOC-W を経由し、NOC-E を経由しない、
5. Network-B と東域との間の IP パケットは NOC-E を経由し、NOC-W を経由しない、

である。

なお、Backbone 側の設定は現在の WIDE Backbone で実施されている、

- NOC-W、NOC-E は、Backbone から受けとった全経路情報に関し、接続先に通常どおりアナウンスする、
- NOC-W は、Network-A から Network-A と Network-B の経路情報のみを受け取り、Backbone に対しアナウンスする、
- NOC-E は、Network-B から Network-A と Network-B の経路情報のみを受け取り、Backbone に対しアナウンスする、

を前提とする。

3.5.2 Backbone と Private Link のゲートウェイが同一の場合

Backbone 上の NOC と Private Link 先 GW との接続を一台のゲートウェイで実現する場合について考察する。この場合、Network-A 内並びに Network-B 内のホストは必ずそれぞれの GW を経由して外部と接続するため、GW の経路管理のみを考慮すれば良い。

図 3.11は経路情報伝達の初期状態を示し、西/W は NOC-W から西域に属するネットワークの経路情報、西/E は NOC-E から西域に属するネットワークの経路情報、東/W は NOC-W から東域に属するネットワークの経路情報、東/E は NOC-E から東域に属するネットワークの経路情報で、括弧内は metric 値を示す。

受けとる経路情報の metric 値より、GW-A から西域への経路は NOC-W へと常に定まるが、東域への経路は最初 NOC-W となり、後に NOC-W と GW-B とのぞれぞれから受けとる同一 metric 値の東域への経路情報により、NOC-W 経由と GW-E 経由の二通りに振れる。これは、GW-B における西域への経路に関しても同様である。

ここで先に述べたポリシーを実現するために、以下の metric 調整を行なう。

> それぞれの GW は、Private Link 経由で受けとる経路情報の metric 値から 1 を減算したものを適用する。

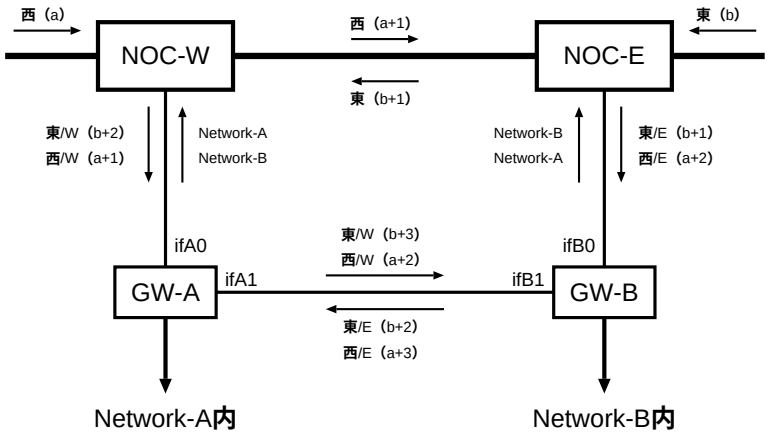


図 3.11: 一つのゲートウェイにより Backbone への Link と Private Link とを管理

これにより、

- GW-A から東域への経路は GW-B 経由となり、その後 GW-A から GW-B へは西/W(a+2) のみが経路情報としてアナウンスされる。
- GW-B から西域への経路は GW-A 経由となり、その後 GW-B から GW-A へは東/E(a+2) のみが経路情報としてアナウンスされる。

なる定常状態に落ち着く。

一方、それぞれの GW から NOC にアナウンスする Network-A と Network-B の経路情報の metric 値は 1 となっており、これにより外部から Network-A または Network-B への経路は Backbone を経由することはない。

また、いずれかのリンクが落ちた場合、復帰後もこの定常状態に落ち着くことは容易に確認でき、提起したポリシーを本手段により実現できることが示される。

ここで、先に述べたポリシーを

> それぞれの GW が、Private Link 経由でアナウンスする経路情報では、受けとった metric 値から 1 を減算したものを適用する。

によっても実現可能であることは容易に類推される。これらに対し、それぞれの GW が NOC から受け取る段階で経路情報の metric 値に関して操作すると、それが Private Link 経由でアナウンスする経路情報の metric にも同様の影響が及ぶので、本ポリシーを実現することは不可能である。

3.5.3　Backbone と Private Link のゲートウェイが個別の場合

Backbone 上の NOC と Private Link 先 GW との接続を別のゲートウェイで実現した構成における本ポリシーの実現手段について考察する。この場合、図 3.12から明らかなよ

うに、Network-A 内、並びに Network-B 内へアナウンスする経路情報の metric に関して考慮する必要がある。なお、図 3.12中の意味は、図 3.11と同様である。

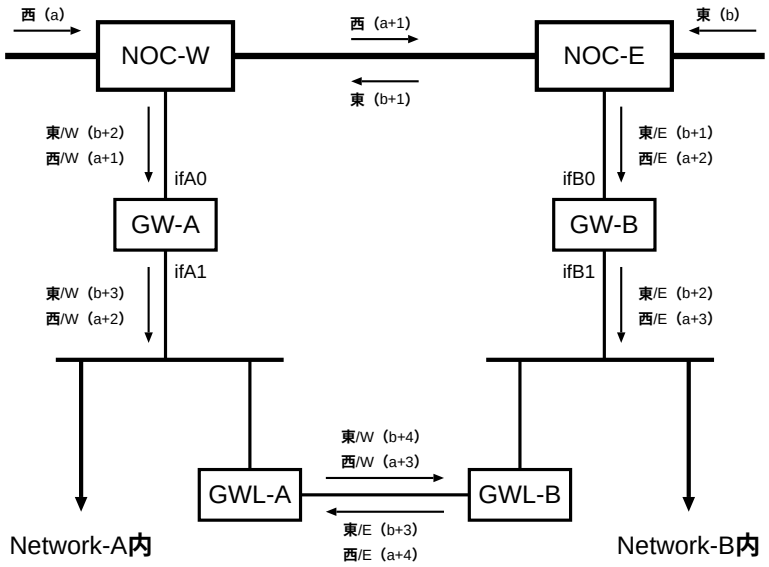


図 3.12: Backbone への Link と Private Link とを異なるゲートウェイで管理

ここでそれぞれのゲートウェイでの経路では、以下に示す状態変移が発生する。

1. 初期状態では、NOC-W から受けとった経路情報により、GW-A は西域、並びに東域への経路を NOC-W へと定め、GWL-A を含む Network-A 内に東/W（b+3）、西/W（a+2）の経路情報をアナウンスする。

2. 初期状態では、NOC-E から受けとった経路情報により、GW-B は西域、並びに東域への経路を NOC-E へと定め、GWL-B を含む Network-B 内に東/E（b+2）、西/E（a+3）の経路情報をアナウンスする。

3. GWL-A では GW-A から受けとった経路情報により西域、並びに東域への経路は GW-A へと定まり、Private Link 側に東/W（b+4）、西/W（a+3）の経路情報をアナウンスする。

4. GWL-B では GW-B から受けとった経路情報により西域、並びに東域への経路は GW-B へと定まり、Private Link 側に東/E（b+3）、西/E（a+4）の経路情報をアナウンスする。

5. GWL-A は Private Link 経由と GW-A とから受けとる経路情報より、西域への経路は常に GW-A を向くが、東域への経路は GW-A と GWL-B との二通りに振れる。

6. GWL-B は Private Link 経由と GW-B とから受けとる経路情報より、東域への経路は常に GW-B を向くが、西域への経路は GW-B と GWL-A との二通りに振れる。

7. Network-A 内の各ホストは GW-A から定常的に東/W（b+3）、西/W（a+2）、GWL-A での東域への経路が GWL-B を向いている時点でそこから東/E（b+4）を受けとり、その結果 Network-A 以外の全ての経路は GW-A を向く。また、GW-A の経路に関しても、Network-A 以外の全てに関して NOC-W を向く。

8. Network-B 内の各ホストは GW-B から定常的に東/E（b+2）、西/E（a+3）、GWL-B での東域への経路が GWL-A を向いている時点でそこから西/W（b+4）を受けとり、その結果 Network-A 以外の全ての経路は GW-B を向く。また、GW-B の経路に関しても、Network-A 以外の全てに関して NOC-E を向く。

ここで本ポリシーを実現するためには、先の同一のゲートウェイの場合と同様に Private Link 経由で受けとる経路情報に対し以下の metric 調整を行なう。

> それぞれの GW は、Private Link 経由で受けとる経路情報の metric 値から 2 を減算したものを適用する。

これにより、

- GW-A を含めた Network-A 内の各ホストでの東域への経路は GWL-B、GWL-A を含めた Network-A 内のホストでの西域への経路は GW-A となり、GWL-A を含む Netowrk-A 内の各ホストへは GW-A から西/W（a+2）が、GW-A を含む Netowrk-A 内へは GWL-A から東/E（b+1）が、GWL-A から Private Link 側へは西/W(a+3) が、それぞれ経路情報としてアナウンスされる。

- GW-B を含めた Network-B 内の各ホストでの西域への経路は GW-A、GWL-B を含めた Network-B 内のホストでの西域への経路は GWL-B となり、GWL-B を含む Netowrk-B 内の各ホストへは GW-B から東/E（a+2）が、GW-B を含む Netowrk-B 内へは GWL-B から西/W（b+1）が、GWL-B から Private Link 側へは東/E(a+3) が、それぞれ経路情報としてアナウンスされる。

なる定常状態に落ち着く。

ここで、GWL-A 並びに GWL-B のそれ Private Link 経由で受けとり適用する Network-A と Network-B の経路情報の metric 値は-1 となるため、それぞれで負の値を取る metric 値を 0 に換算する工夫が必要である。

一方、それぞれの GW から NOC にアナウンスする Network-A と Network-B の経路情報の metric 値は 1 となっており、これにより外部から Network-A または Network-B への経路は Backbone 経由することはない。

ここで、いずれかのリンクが落ちた場合、復帰後もこの定常状態に落ち着くことは容易に確認でき、以上により提起したポリシーを上記手段により実現できることが示された。

また、提起したポリシーを

> それぞれの GW が、Private Link 経由でアナウンスする経路情報では、受けとった metric 値から 1 を減算したものを適用する。

によっても実現可能であることは同一のゲートウェイの場合と同様である。但し、それぞれの GW が NOC から受け取る段階で経路情報の metric 値に関して操作すると、それが Private Link 経由でアナウンスする経路情報の metric も同様の影響が及ぶので、本ポリシーを実現することは不可能である。

3.5.4 まとめ

以上、対 NOC と対 Private Link のゲートウェイの構成に関して二通りの場合について考察したが、いずれの場合も Private Link 経由で受けとる、もしくはアナウンスする経路情報の metric 値に操作を加えることにより、提起したポリシーを実現できることが明らかになった。

また、本ポリシーを実現する手段として、これ以外に

- static route の適用
- RIP preference を導入した gated の適用
- Virtual Tunnel の適用

が考えられる。しかし、これらの解決は以下の理由により本問題の解決策としては適当ではないと考える。

- static route では dyanmic routing による代替経路適用の利点を享受できない。
- RIP preference を導入した gated は二つの安定状態を招き、本問題の解決は不可能。
- Virtual Tunnel では、ネットワーク間を接続するゲートウェイの構成によっては、ダブルトラフィックを発生させる。

なお、現在 WIDE Backbone の経路制御が RIP から BGP と OSPF への移行が進められている。OSPF の経路情報中の TAG を用いることで、これらのポリシーが明確な形で記述できることを最後に述べておく。

3.6 もとのり問題

ここでは、3.3 から 3.5 の３つの最適経路選択問題をまとめ、AS 間 (inter-AS) の環境での問題定義を行なう。以下、この問題をもとのり問題を呼ぶ。

問題を AS 間環境に適用させるため、先ず、発信元と宛先を独立した AS に置く。事実、複雑な相互接続のインターネット網環境では、発信元や宛先が主要網などの通過のための中間網とは別である方が一般的であろう。

前出の最適経路選択問題１は次のようになる。図 3.13 にあるように、今、AS X にある X0 から AS Y にある Y0 へトラフィックが送られると考える。X から Y へ行く経路に

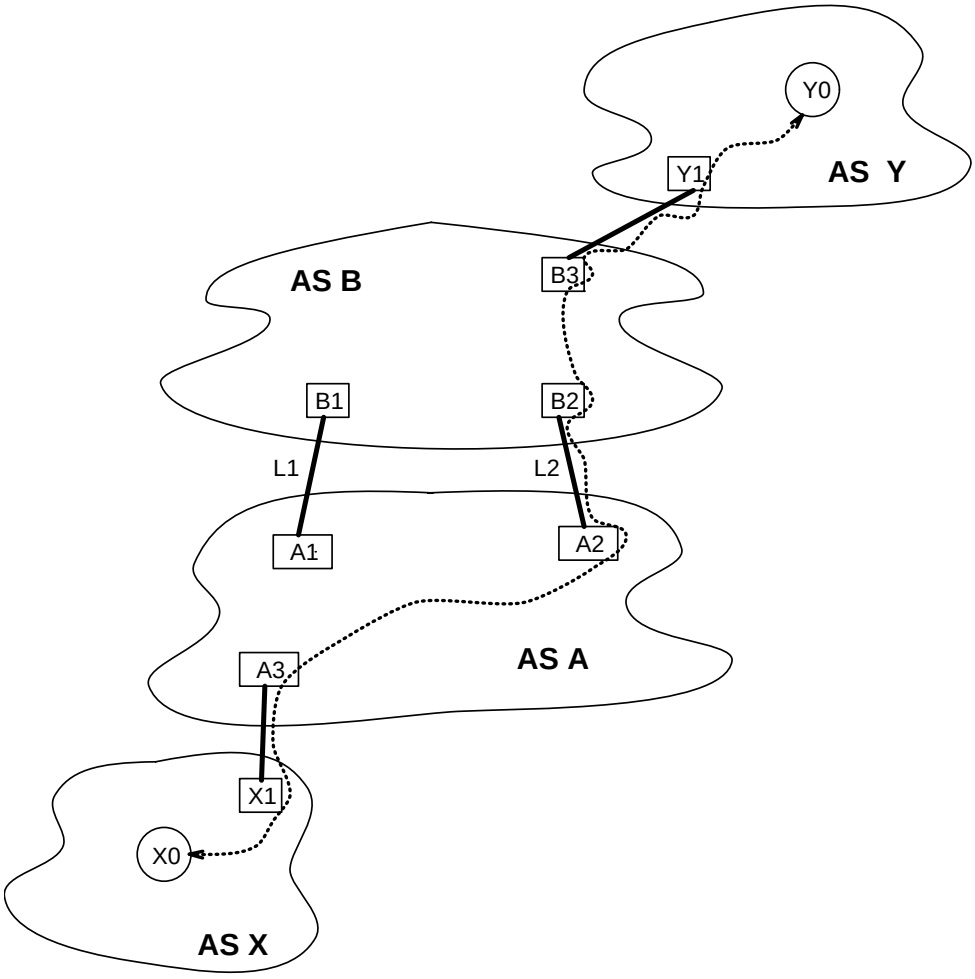


図 3.13: もとのり問題

介在する２つの通過 AS はそれぞれある通過コストの値が異なっているとする。例えば、X にとっては、A は B よりあるコストが低いとする。議論を易しくするために、Y からみてもこのコスト差は同じだとする。この場合、問題はどのように次の２つの方針が施行されるかである：

1. X0 から Y0 へ送られるトラフィックは、A にできるだけ長くいて、B ではできるだけ短い経路で通過する。

2. Y0 から X0 へ送られるトラフィックは、B をできるだけ早く出て、できる限り A を通って行く。

もう少し一般化すると、問題は発信元がどのようにして宛先までの経路を構成する n 個の AS の内、自分が望む m 個の AS がある場合、望まない AS ではできるだけコストを低く通り抜けたいという方針を、いかに実現できるかということである。Routing by Preference は、嫌いな AS ではできるだけ通過コストを抑えることで、達成されるのである。すなわち、好きな AS ではどんなにコストがかかっても気にしないが、嫌いな AS ではあるコスト面での最短経路を通りたいという要求である。この発信元からの要求は経路設定の際に実現されるであろう。

これに対して最適経路選択問題２および３は、宛先 AS の受け入れトラフィックに対する要望である。これについては、当該 AS が経路制御情報を他に流す際に制御できると考える。さらに、この宛先 AS の要望の背景には、公的な主要網の一部のリンクの使用軽減につながることもあるので、他の AS も協力した形での経路情報制御が可能かもしれない。

これら発信元の意志からの経路選択および宛先の意志からの経路選択の問題を、もとのり問題と呼び、今後、Routing by Preference のチームで具体的な解決策を研究していきたいと考えている。

3.7 既存の政策的経路制御での解決法

仮想トンネル問題は、政策的経路制御に限らず、もっと一般的な問題といえる。もともと、仮想トンネル技術自体は、様々なアプリケーション別に一時凌ぎ的な発想で行なわれてきた。現在行なわれているインターネット網上での teleconference の一環の Audio Cast の実験でも、使用されている。現在のどの経路制御プロトコルによっても強制経路設定がせいぜい発信元経路決定 (source routing) 程度に留まっている。ここで紹介した仮想トンネルは、汎用性のあるツールとして、既存の網上に自由な網を構築できるものを目指している。

もとのり問題の内、発信元の意志による経路選択は、BGP の枠組では、SDRP により、AS 単位での経路までは設定できるのだが、AS 間リンク指定ができないことと、「経路上のある AS では、コストを低くしたい。」というような要求はだせない。また、SDRP のための情報をどのようにして収集するのかも以前はっきり決められていない。IDPR に

おいては、新たに要求項目を追加することで、発信元経路設定が可能である。そのためには、経路情報交換において AS 間リンクの情報の追加が必要なのだが、それがただでさえ複雑なプロトコルをさらに複雑にすることとなる。このように、もとのり問題は既存のプロトコルに修正を加えてできないことはないであろう。しかし、もともと、それらのプロトコルはこの Routing by Preference の目的のためのプロトコルではない。そこで、我々は、先ず Routing by Preference だけのためのプロトコルを作り、そこで得た結果から、最低必要項目をどのように既存のプロトコルに挿入できるかを検討していこうと考えている。

3.8　章のまとめ

問題認識は研究の原点である。本章で取り上げた諸問題には、様々な意味で新しさがある。Virtual Tunnel 問題は、今までに、認識はされていたが、それ自体が研究の対象と扱われたことはなかった。最適経路選択問題は、政策的経路制御の世界に具体的かつ実践的なポリシーを定義した。

仮想トンネル問題は、独立したワーキング・グループとして活動を開始している。もとのり問題は後で紹介する Routing by preference のプロジェクトなって、本ワーキング・グループのなかの実働チームにより、今後研究を進めていく。

その他、九州工業大学や九州大学などで発生している、複数の主要網接続の拠点となっている地点における AS 分割問題があり、今後も諸問題の認識を行なっていきたいと考えている。

第 4 章

今後の研究課題

4.1 わが国のインターネット網での政策的経路制御

前章の RIP による最適経路選択についてのセクションでもあるように、わが国のインターネット網では, Routing Information Protocol(RIP) [123] を使ってきたが, コストの最大値が１６であることや, すべての網で共通のコストを使わなければならない制約が, 主要網や参加組織の数の増加による網規模の増大に合わなくなってきた. 現在, 次世代の経路制御手順が必要な時期に来ている. 現段階での解決法としては, BGP への移行が有力である. これは, 米国で広く使用されている実績からである.

インターネット網全体としても, IDPR は BGP のような実用に至っていない. 理由として, 今までの経路制御手順から大きく逸脱していること, そして, 現在のところまだ実験段階にあることがあげられる. また, このような積極的な形での資源使用制御の必要性の認識が, 今一つ低いことも事実であろう.

資源使用制御の認識の低さは, インターネット網構築の背景にある. もともと, インターネット網は, より容易に各部分網が接続できるように設計されている. 接続性(connectivity) が資源使用制御 (access control) よりも優先されていた. この思想は, 現在, まだ初期段階を脱しかけている, わが国のインターネット網にも, 強く反映されている. こうした環境下では, BGP への移行は, 今の段階では, 短期的解決としては, 妥当であろう.

BGP への移行に際して, 問題になるのは, AS の分け方である [128]. どこまでが, ひとつの政策で, ひとまとめにできるかといった問題は, 今後の現実的な研究課題である.

4.2 Routing by Preference

もとのり問題の解決のための、Routing by Preference は、今後、具体的なプトロコル設計に入っていく。当初、これについては、既存の BGP や IDPR を修正しながらの解決方法も考えられたが、preference 自体、新しい考え方なので、そのための経路制御を先ず我々で作りあげ、その結果から、既存のプロトコルへの修正案を主張できるであろうと考えている。

経路制御のためには、経路情報交換手順、転送アルゴリズムなどの設計をしなければならい。

発信元の要求は、経路決定の際に考慮され、宛先域の要求は、経路情報交換において実現されるであろう。

現段階では、次のような経路制御プロトコル体系設計のための要因を考えている:

- 発信元および宛先域の要求の種類
- 誰が経路の preference を要求し、その要求を誰が認可し、誰が経路を決定するのか。
- 域レベルの経路や部分経路決定はどこで、誰が決定するのか。
- 経路情報の配布におけるアクセス・コントロールはどのように行なうか。その際に宛先域の要求をどのように反映するか。

今後はこれらに基づいて様々な手順を設計していきたいと考えている。

4.3　セキュリティー問題

米国では政策的経路制御のためのセキュリティも研究されている [129]. 真の網資源の使用制御には, 使用者の認証が不可欠である. また, 経路情報が正しいものであるかどうか, その完全性 (integrity) も保ちたい [130] [131]. しかし, 現在のところ, ネットワーク層でのセキュリティーの対策は, インターネット網ではほとんど行なわれていない. これは, ただでさえ, 複雑化する経路制御に, セキュリティのためのオーバーヘッドを付加できないでいるからである.

したがって, どのような政策的経路制御手順が使用されようと, 現段階では, 使用制御が正しく行なわれるかどうかは疑問の余地がある. 経路制御における使用制御と共にセキュリティの問題の研究は, 今後の大きな研究課題である.

第 5 章

むすび

5.1 まとめ

ここでは, 今, 新しい種類の経路制御である, 政策的経路制御を紹介した. 具体的な例として, 現在実用段階に入りつつある 2 つのプロトコル, BGP と IDPR を取り上げた.

BGP は基本的に網規模の拡大問題の解決から生まれてきた一連の経路制御手順の一つであり, IDPR は網資源の使用制御の研究を背景としている. 前者は, 中間システム経路決定形式, 後者は発信元経路決定形式のなかでの, 経路情報交換や転送形式を指定している.

IDPR は BGP にくらべ, 把握できる方針の種類が多い. しかし, 実用面では, 既存の経路制御手順を逸脱して IDPR に比べ, 既存の AS 間経路制御手順の延長上にある BGP が優先している.

わが国のインターネット網では, その規模増大に伴い, 現段階では, BGP への移行を短期的解決策としている. 実用面での実績がほとんどの理由であるが, 他の理由として, 網資源の使用制御についての認識の薄さがあげられる. これは, わが国だけではなく, インターネット網全体について言えることである. インターネット網は, そのプロトコル体系の容易な接続性 (connectivity) のために, 急速な成長を遂げてきた. 従って, これに相反する思想の網資源使用制御 (access control) は, なかなか実現されにくい. IP アドレス空間の枯渇問題 [132] と共に, インターネット網の構造自体が, 今, 過渡期にある. この段階で, 資源使用制御問題が, 例えば終端 AS による経路選択要求 (routing by preference) [133] などの, もっと具体的な形の政策的制御となって, 次世代のインターネット網に採り入れられていくのではないだろうか. 使用制御に伴いセキュリティ問題も今後の課題である.

今後の本ワーキング・グループでは, 研究課題の一つとして, この routing by preference を取り上げている. この他, 当グループでは, 経路制御レベルで実現できる具体的な政策を模索中である. また, BGP 移行に際しての, AS 分割問題なども, 興味深い研究課題である.

5.2 来年度の研究計画

研究課題の筆頭は、routing by preference の研究があげられる。

また、米国の研究の追従の一環として、今年は、IDPR の実験などにも参加していき

たい。

この他、AS 分割問題は時間をかけて手掛けたい。国際リンクなどの政策・方針なども調査なども必要であると思われる。網の境界、組織の境界、国の境界などが入り乱れた中での網資源の運用についてのポリシーはどのように生み出されていくかは、面白い研究課題のひとつとなろう。九州工業大学や九州大学などで発生している、複数の主要網接続の拠点となっている地点における AS 分割問題も、来年度の問題認識の研究課題となろう。

5.3　今後のワーキング・グループの構造

今後、このワーキング・グループでは、ポリシー・ルーティングに限らず、様々な経路制御に関する問題を取り扱っていきたいと考えている。大きな柱として、次の３つの仕事をしていかなくてはならない：

1. 研究
2. 米国などの研究の追従
3. 実践

研究は、今までのような、問題認識や、それらを原点に始まった我々独自の研究活動を意味する。この研究活動の成果を将来 IETF(Internet Engineering Task Force) などにフィード・バックしていきたいと考えている。

米国や ISO などで現在行なわれているこの分野における活動の追従も必要であろう。

実践では、Working Operation Team (WOT) で実施されていく日本のインターネット網での経路制御のための、ブレインとなる仕事を意味する。これには、今後のインターネット網の AS 分割などが上げられる。