修士論文 2019 年度 (令和元年)

IPv6 シングルスタックネットワークにおける ダイナミックなアドレス変換テーブル広告手法

慶應義塾大学大学院 政策・メディア研究科
豊田安信

修士論文要旨 - 2019 年度 (令和元年度)

IPv6 シングルスタックネットワークにおける ダイナミックなアドレス変換テーブル広告手法

2019 年現在，IANA が保有する IPv4 アドレスプールは既に枯渇しており，各地域レジストリからも 2020 年頃には新規割当が行えなくなることが予想されている．一般に新規に IPv4 アドレスの取得するためにはこのような民間取引市場を利用する方法が考えられるが，1 アドレスあたりの単価は年々上昇しており，新規に IPv4 ネットワークを構築するためのコストは日々上昇していくことが考えられる．IDC 事業者・コンテンツ事業者がビジネスを拡大するためには，IPv4 アドレスを極力使用しない IPv6 シングルスタックネットワークの活用が不可欠になっている．

一方で 2019 年現在においても IPv4 によるアクセス・トラフィックは依然としてインターネット全体の大きな割合を占めていることから，IPv6 シングルスタックネットワークでありながら IPv4 によるサービスを継続して提供可能なネットワーク設計が必要になってくると言える．

IPv6 のみ構築された IPv6 シングルスタックネットワークにおいて既存の IPv4 クライアントに対してサービスを提供する方法として、ステートレスアドレス変換を利用した”SIIT-DC” と呼ばれるネットワークデザインがインターネット標準として標準化されている．SIIT-DC では BR(Border Relay) と呼ばれる変換ノードを IPv4 ネットワーク・インターネットとの境界点ごとに設置し，明示的アドレス変換テーブル (EAMT: Eplicit Address Mapping Table) を参照してプロトコル変換を行い，IPv6 ノードでの IPv4 サービス提供を可能にする．しかしながら SIIT-DC では EAMT の動的な交換方法についての定義がなされておらず，対外接続点が複数存在する場合の冗長性の維持が難しい点や，IPv4 でサービス提供を行なうサーバーの構成変更が行われた場合に運用負荷が非常に高くなる点が課題に挙げられる．

本研究では BGP を利用したアドレス変換テーブルの広告・更新技術と，それを適切に運用するために必要なノード群の設計手法を提案する．これにより，SIIT-DC の課題であった冗長性の維持や構成変更へのに対して，ダイナミックに対応することが可能になる．

この手法を評価するために，新たに BGP によるアドレス変換テーブル制御機構を実装したソフトウェアルーターを実装し，多くの対外接続点を持つ学術 ISP である WIDE Project のバックボーンネットワークをモデルケースに，エミュレータを用いて概念検証実験を行った．考えられる他の手法と比較し，本手法が冗長性と柔軟性の点で優位であることが証明された．

キーワード:
1. IPv6, 2. データセンターネットワーク, 3. ネットワークオペレーション, 4. IPv6 移行技術

慶應義塾大学大学院 政策・メディア研究科
豊田安信

Abstract of Master's Thesis - Academic Year 2019

Dynamic advertising method of
Explicit Address Mapping in IPv6 single stack network.

Dynamic advertising method of Explicit Address Mapping in IPv6 single stack network.
Keywords :
1. IPv6, 2. Data center network, 3. Network operation, 4. IPv6 transition mechanism

Keio University Graduate School of Media and Governance
Yasunobu Toyota

目 次

図 目 次

表 目 次

第1章　序論

本章では本研究の背景とモチベーション，および全体の構成について記述する．

1.1　IPv6 シングルスタックネットワークに求められる役割

1.1.1　IDC ネットワークを取り巻く環境

IDC 市場の広がり

近年，ライブ映像配信のようなリアルタイムなサービスに対するニーズが年々高まっている．例えば Cisco 社の調査 [2] によれば，2022 年には全てのアプリケーショントラフィックのうちインターネットビデオが有する割合が 82 %を超え，そのうち 17 %がライブ映像配信が占めると予想されている．リアルタイムな高品質サービスを提供するためには，ユーザーの地理的に近いサービス拠点から配信を行うことが有効であるため，今後 IDC・コンテンツ事業者が各地域拠点を介したコンテンツ配信基盤を活用するしていくことが予想される．

一方で，インフラストラクチャに対する災害や地政学的リスクの軽減は，コンテンツ事業者の継続的な事業の成長のためには避けては通れない課題である [3]．2011 年に発生した東日本大震災以降，国内の IDC 事業者やコンテンツ事業者を中心に，関東大都市圏に集中していたサービス拠点への依存性を解消するために，東京圏以外の各地域にサービス拠点を分散する取り組みが活発だ [4]．大阪・名古屋の他の都市圏の IDC は 2019 年現在満床状態が続いているほか，他の地方拠点都市も含めた IDC 建設も並行して行われている．

特に近年では VXLAN や SRv6 のような新しいネットワーク仮想化技術の標準化も進み，サービス拠点のマルチテナンシーと柔軟性を両立するネットワークデザインの障壁が低くなってきているため，今後より多くの IDC・コンテンツ事業者のサービス拠点の拡大が続くと想定できる．

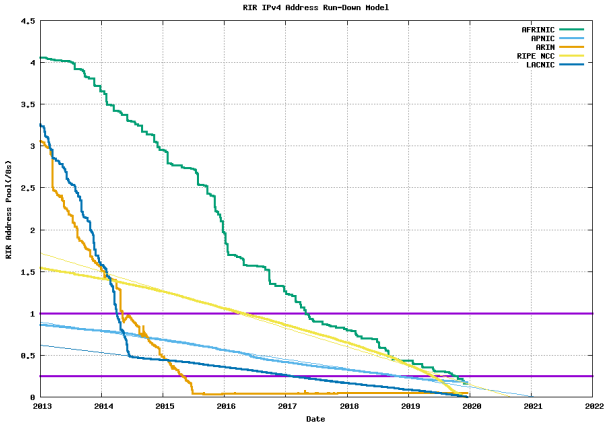

図 1.1: Projection of consumption of Remaining RIR Address Pools. `potaroo.net` より引用 [1]

IPv4 アドレスの枯渇

2019 年現在，IANA[1]が保有する IPv4 アドレスプールは既に枯渇しており [5]，各 RIR[2] からも 2021 年頃までには新規割当が行えなくなることが予想されている [1].

一方で近年は民間事業者間アドレス取引も盛んに行われている．一般に新規に IPv4 アドレスの割当を受けるためにはこのような民間取引市場を利用する方法が考えられるが，1 アドレスあたりの単価は年々上昇傾向にあり [6]，新規に IPv4 ネットワークを構築するためのコストは日々上昇していくことが考えられる．

IPv6 への移行

1998 年に初めての IPv6 標準仕様が策定されて以降 [7]，IPv6 インターネットと IPv4 インターネットの独立した二つのインターネットが並行して存在する状態が 2019 年現在まで続いている．

一方でインターネット技術は IPv6 を前提とした設計が行われる段階を迎えている．2016 年には IAB[3]により，”IAB Statement on IPv6”が発表され，インターネット標準は IPv6 に最適化標準策定を行う方針が確認されている [8]．例えば既に SRv6 のような新しい標準は IPv6 の拡張ヘッダーを利用した技術として策定が進められており，IPv4 を前提とした長期的なネットワーク運用は限界を迎えていると言える．

[1]Internet Assigned Numbers Authority．インターネットに利用される様々な資源を一元的に管理する組織．`https://www.iana.org/`

[2]Regional Internet Registry.

[3]Internet Architecture Board. `https://www.iab.org/`

IPv4/IPv6 デュアルスタックネットワークの問題

IPv6 プロトコルの導入に主に用いられていた手法として IPv4/IPv6 デュアルスタックネットワークが挙げられる [9]．IPv4/IPv6 デュアルスタックネットワークとは，IPv4 ネットワークと IPv6 ネットワークを同一機器群上に並行して運用する手法であり，企業・一般家庭向けアクセスネットワークを中心に IPv4/IPv6 デュアルスタック環境の整備が進んでいる．

一方でコンテンツ事業者が運用する IDC では以下の主な 3 つの理由からデュアルスタック環境の導入はデメリットが大きい．

- **IPv4 アドレスの継続的調達が困難**
  先に述べたように，IPv4 アドレスをサービスの成長にあわせて継続的に調達していくことは困難である．民間市場の市況に調達コストが左右されるため長期的な見通しが立てにくい．

- **オペレーションコストの肥大化**
  デュアルスタック環境では 2 つの異なる IP プロトコルを同時に運用する必要があるため、シングルスタック環境と比べて運用コストの上昇が見込まれる [10]．

- **ネットワーク機器の性能要件の上昇**
  デュアルスタック環境では，シングルスタック環境よりも多くの経路をネットワーク機器が保持しなければならないため，より高性能な機器を導入する必要がある．

1.1.2 IPv6 シングルスタックネットワーク

IDC 事業者・コンテンツ事業者がビジネスを健全に拡大するためには，IPv6 ネットワークのみで機器間を接続した IPv6 シングルスタックネットワークの利用が不可欠である．

IDC の IPv6 シングルスタックネットワークには以下のような働きが期待される．

IPv4 サービスの提供

Google 社が定常的に行っている調査 [11] によれば，2019 年 12 月現在全世界のインターネットトラフィックの 7 割程度を IPv4 トラフィックが依然として占めている．将来的には IPv6 によるアクセスの割合が徐々に大きくなることが予想されるが，今後しばらくは IPv4 クライアントに対しても IPv6 クライアントと同等にサービス提供を行っていくことが望ましい．

コンテンツ事業者の IPv6 シングルスタックネットワークにおいても，何らかの手段を用いて IPv4 サービスを継続して提供する機構を備える必要がある．

シングルスタック運用による OPEX/CAPEX の削減

第 1.1.1 項で述べたように，IPv4/IPv6 デュアルスタックネットワークではオペレーションコストの肥大化が問題視されていた．IPv6 シングルスタックネットワークでは IPv4 ネットワークを廃止することが出来るため，OPEX[4] と CAPEX[5] の軽減が期待される．また IPv6 アドレスは IPv4 アドレスと比較して広大なアドレススペースを有するため，アドレススペースに依存しない柔軟なネットワーク設計が可能になる．

IPv4/IPv6 デュアルスタックネットワークと同等以上の性能

IPv6 により提供されるサービスはもちろんのこと，IPv4 によるサービスにおいても IPv4/IPv6 デュアルスタックネットワークと同等の耐障害性・サービス品質・サービス容量が保証されることが望ましい．

とりわけネイティブな IPv4 ネットワーク以外の手段を用いて提供される IPv4 サービスの性能の担保が運用課題になると予想される．

1.2 本研究のモチベーションと取り組み

第 1.1.2 項で述べたような IPv6 シングルスタックネットワークに求められる要件のうち，IPv4 サービスの提供における冗長性や構成変更への追従性の向上を促す手法の確立を目指す．

本研究では IPv6 シングルスタックネットワークにおける IPv4 サービスの提供手法のうち，アーキテクチャがシンプルで広範な利活用が期待される SIIT-DC[12] に着目した．SIIT-DC とは IPv6 ネットワークと IPv4 ネットワークの各境界部に，BR[6] を配備することにより，IPv6 ネットワークのみに属するホストで仮想的に IPv4 サービスを提供するネットワーク設計を定めたインターネット標準である．SIIT-DC において各 BR は静的に定義されたアドレス変換テーブルを利用してネットワークプロトコル変換を行うため，BR を複数配備する場合における一貫性の確保や冗長性，IDC 内の構成変更に対する追従性の面で課題があった．

本研究では動的経路アルゴリズムの一つである BGP[14] を利用したメッセージングによるアドレス変換テーブルの動的な広告手法を提案する．エミュレータを利用した概念実証実験により，本提案手法がこれらの課題に対して効果的に作用することが証明された．

1.3 本論文の構成

本論文の構成を以下に示す．

[4] Operating expense. 運用に掛かる継続的なコスト．
[5] Capital expenditure. 設備配備に掛かる初期投資コスト．
[6] Border Relay. IP/ICMP 変換アルゴリズム [13] を実装した IPv4/IPv6 トランスレーション機器．

第 2 章では，IPv6 シングルスタックネットワークにおける IPv4 サービス提供手法に関してそれぞれの特徴や利点を紹介し比較する．

第 3 章では，IPv4/IPv6 プロトコル変換を利用した IPv4 サービス提供手法の一つである SIIT-DC のアーキテクチャと，解決すべき課題について述べる．

第 4 章では，SIIT-DC の課題を解決するために考えられる手法を比較・検討する．

第 5 章では，本研究において提案するダイナミックなアドレス変換テーブル広告手法の要件と構成について記述する．またメッセージングプロトコルとして採用した BGP の技術的利点について述べる．

第 6 章では，本提案手法の BGP メッセージペイロードの設計と第 7 章でも評価実験に用いる PoC の具体的な実装について紹介する．

第 7 章では，第 3 章で述べた課題に対して，本提案手法が有用であることを検証するための実証実験の概要及び具体的なシナリオについて述べ，結果を考察する．

第 8 章では，本研究のまとめと今後のロードマップについて検討する．

第2章　IPv6シングルスタックネットワークでのIPv4サービス提供手法

本章ではIPv6シングルスタックネットワークでのIPv4サービス提供手法を比較し，検討する．

2.1　概要

第1.1.2で述べたように，コンテンツ事業者が運用するIPv6シングルスタックネットワークの重要な役割の一つに，IPv4クライアント端末に対するサービス提供がある．

関連して，アクセスネットワーク網ではIPv6シングルスタックネットワーク上でIPv4によるインターネット接続をクライアントエッジに提供する手法はをIPv4aaS[1]と呼称し，様々な手法が検討されている[15]．

一方でコンテンツ事業者が運用するネットワークでのIPv4サービス提供においては下のような要件を満たす必要があるため，必ずしもアクセスネットワークでのIPv4aaSと同様の方法が適切であるとは限らない．

2.1.1　IPv4サービス提供機構に求められる要件

IPv4クライアントからのアクセス

IPv4クライアントに対して透過的にサービスを提供する機構を備える．一般的なサーバークライアントモデルを想定した場合，インターネット上のIPv4クライアントからサービス提供サーバーに能動的に接続するためには，FQDN[2]もしくはIPv4アドレスをIPv4クライアントが指定出来る必要がある．

[1] IPv4 as a Service

[2] Fully Qualified Domain Name. 完全就職ドメイン名

スケーラビリティ

近年のコンテンツ事業者のネットワークでは，サービスのニーズに合わせて柔軟にスケールアウト[3]可能な設計であることが重要視されている [16]．同様に IPv4 サービスの提供手法に関しても，事業者の IPv4 サービス規模の変化にあわせて柔軟に拡大・縮小可能なアーキテクチャが求められる．

例えば，第 1.1.2 でも述べたように，将来的に IPv4 クライアントの占める割合が IPv6 クライアントに相対して低下していった場合に，既設の IPv6 ネットワークへの影響を最小限にしつつ，IPv4 サービス提供機構を縮小可能であると望ましい．

IPv4 ネットワークへの非依存性

第 1.1.2 項で述べたように，IPv6 シングルスタックネットワークのメリットを最大限に活かすためには IPv4 サービスを提供する場合においても IPv4 ネットワーク及びアドレスに極力依存しないことが望ましい．

2.2 IPv4 サービス提供手法の分類

想定される IPv4 サービス提供機構をその技術的差異や狙いを基に以下の 3 つの手法に分類した．

2.2.1 L7 リバースプロキシ

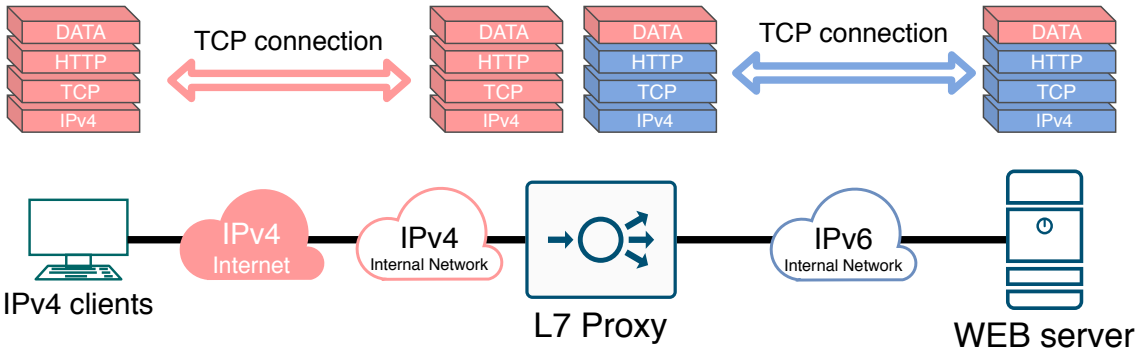

図 2.1: L7 リバースプロキシによる IPv4 サービス提供

L7 リバースプロキシとは，クライアントからの接続をプロキシサーバーがアプリケーション層レベルで終端し，プロキシサーバーがクライアントに代わってサーバーと接続す

[3]水平スケール．同等性能の機器を増減させることでサービス容量を拡大・縮小可能なモデル．

る機構である [17]．図 2.1 に本手法の構成を簡便に示す．主に WEB サーバーへの HTTP 接続を負荷分散するための手法として広く採用されている．

IPv6 シングルスタックネットワークにおいて IPv4 サービスを提供するためには，IPv4 インターネットとの接続点からプロキシサーバーまでの間に IPv4 ネットワークを配備する必要がある．

IPv4・IPv6 間のプロトコル仕様の差を考慮する必要が無いため互換性に留意する必要が無い点や，MTU[4]を減らさずにアプリケーショントラフィックを伝送可能である点が利点に挙げられる．

一方でアプリケーションレイヤーでのコネクション終端やそのステート管理を行う必要があるため，プロキシサーバーに負荷が掛かるため高性能な機器の導入が必要になる．

またスケールアウトを可能にするために L4LB と組み合わせた 3 ステージのアーキテクチャ利用する手法が近年主流である [18, 19] が，この手法を採用するためには，ある程度の IPv4 ネットワークを配備する必要があり，第 2.1.1 項で述べた要件に合致せず，IPv6 シングルスタックネットワークのメリットを損なうことになる．

2.2.2 IPv4/IPv6 トンネリング

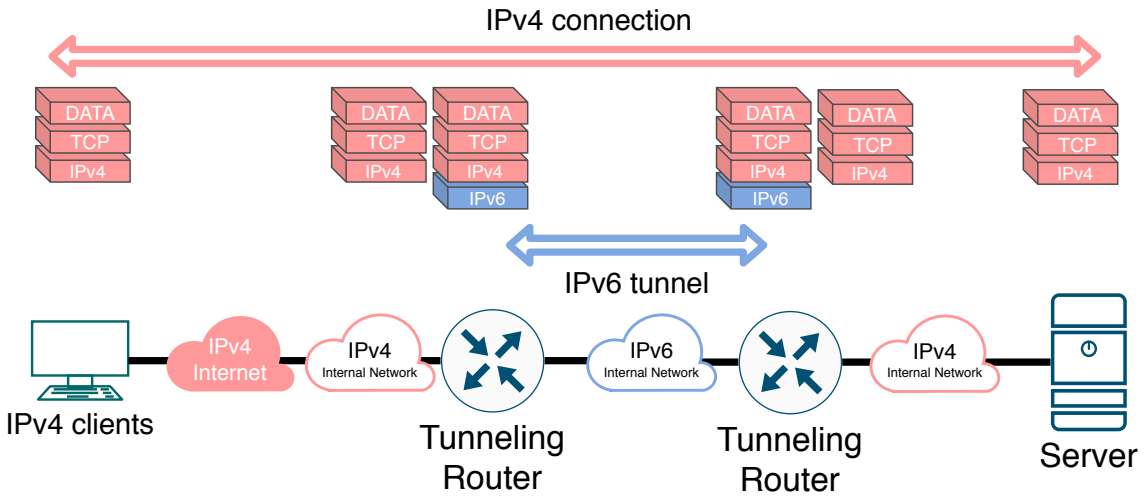

図 2.2: IPv4/IPv6 トンネリングによる IPv4 サービス提供

IPv4/IPv6 トンネリングとは，IPv4 パケットを IPv6 パケットによってカプセリングすることで IPv6 ネットワークを通過させる手法である．IPv4 トラフィックを透過的に利用することが出来るため，アクセスネットワークで最も一般的に利用されている IPv4aaS 手法である [15]．図 2.3 に本提供手法の構成を簡便に示す．

IPv6 シングルスタックにおける IPv4 サービス提供手法としては，IPv4 クライアントから到達したパケットをトンネルルーターによって一度 IPv6 パケットでカプセリングし，

[4]Maximum Transmission Unit. ここでは一つのパケットに搭載可能なデータ量を指す．

IDC 内の IPv6 シングルスタックネットワークを通過させ，IPv4 サービス提供サーバー上もしくはその直前で再びでカプセリングを解くことで，IPv4 提供サーバーまでネイティブな IPv4 トラフィックを通過させる運用が考えらえる．IPv4 ネットワークをサーバーで利用できるため，多種多様なアプリケーションでの採用が期待できる．

しかしながら，トンネルルーターと IPv4 サービスサーバー間にある程度の IPv4 ネットワークを配備しなければならず，ToR[5]及びサーバーでは IPv4/IPV6 デュアルスタック運用が必要になるため，第 2.1.1 項で上げた要件である「IPv4 ネットワークへの非依存性」に合致しない．また，トンネルプロトコルの多く [20] はは基本的に 1:1 もしくは 1:N の接続が基本となるため，水平スケールさせることが困難である．

2.2.3 IPv4/IPv6 トランスレーション

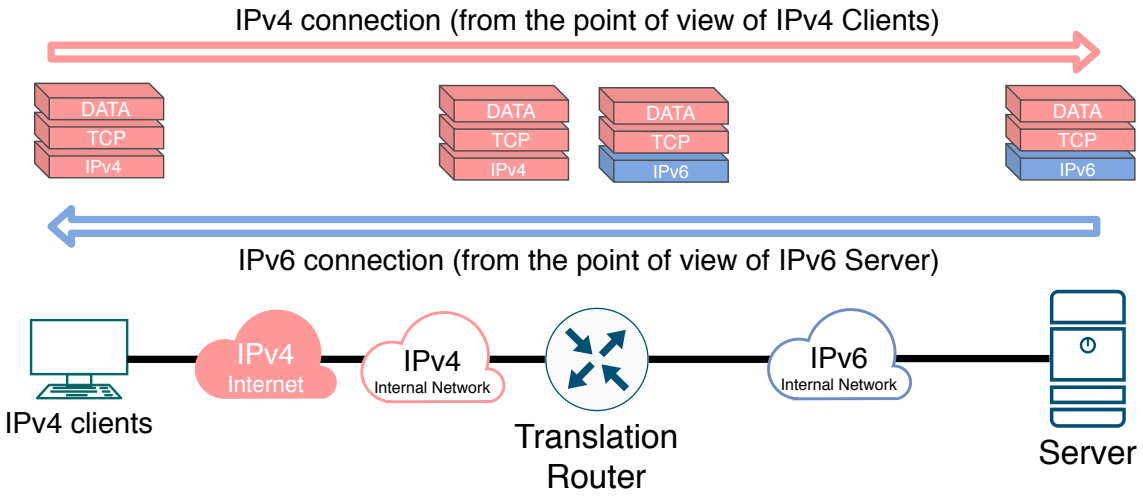

図 2.3: IPv4/IPv6 トランスレーションによる IPv4 サービス提供

IPv4/IPv6 トランスレーションとは，IPv4 パケットと IPv6 パケットを IP/ICMP 変換アルゴリズムを利用して相互に変換する手法である．$1:N$ の関係でアドレス・ポート変換を行うステートフルな NAT64[21] と，1 : 1 でアドレス変換を行うステートレスな SIIT[13] が定義されている．IPv4 ネットワークと IPv6 ネットワークの境界に位置する変換ルーターにより，相互にプロトコル変換が行われる．

IPv4/IPv6 トランスレーションでは IPv4 アドレスを IPv6 アドレスとして表現することが要求されるが，変換プレフィックスと呼ばれる IPv6 ネットワークプレフィックスに IPv4 アドレスを埋め込むことで，任意の IPv4 アドレスを IPv6 ホストから認識可能な形で表現する．変換プレフィックスには RFC6052 で定義された 64:ff9b::/96 の他に，運用者が専有可能な GUA[6]の/96 の IPv6 プレフィックスを利用することが想定されている [22]．

[5]Top of rack switch．ここではサーバーの L2 終端を行うルーターを指す．
[6]Global Unicast Address.

図 2.3 で示すように，変換ルーター以外のホストが IPv4 ネットワークに属する必要が無いため，第 2.1.1 項で述べた「IPv4 ネットワークへの非依存性」の面で，他の 2 手法より優れていると言える．また，IPv4 サービスを行うサーバーから変換ルーターの間はネイティブな IPv6 ネットワークで接続可能なため，ECMP[23] による経路の冗長化が可能なほか，ステートレスモードでは変換ルーターの水平スケールが可能な点で，IPv6 シングルスタックネットワークにおける IPv4 サービス提供に求められる要件を満たしやすい．

一方で IPv4 と IPv6 のプロトコル実装に差があるため，コンテンツ事業者のサービスの内容によってはサービス影響を考慮する必要がある点は留意すべきである．

第3章　SIIT-DCのデザインと現状の課題

第2.2.3で述べたPv4/IPv6トランスレーションを用いたIPv4サービス提供手法の一つとして，SIIT-DCがインターネット標準化されている．本章ではてSIIT-DCのデザインとメリット及び考えられる運用，そして現状の課題について述べる．

3.1　SIIT-DC

3.1.1　概要

SIIT-DCとは，ステートレスIP/ICMP変換アルゴリズム[13]を利用して，IPv4インターネット・ネットワークからのアクセスをIPv6シングルスタックネットワーク上のホストに提供するためのネットワークデザインである．2016年よりIETF IPv6 Operations WG[1]によりインターネット標準化(Informational RFC)されている[12].

3.1.2　用語

SIIT-DCで利用される用語及び特殊な役割を有する機器・技術について述べる．

SIIT

SIIT[2]とはIPv4/IPv6トランスレーションに用いられるプロトコル変換機能の略称である．RFC2765[24]で初めて標準化され，その後RFC6145[25]により一部の仕様が実運用のユースケースに合わせて変更され，現在はIPv6拡張ヘッダーを扱う機構などが追加されたRFC7915[13]が現行の標準仕様である．

[1]IPv6 ネットワークの運用要件や関連する技術仕様の策定を行うワーキンググループ．`https://datatracker.ietf.org/wg/v6ops/about/`

[2]Stateless IP/ICMP Translation Algorithm

BR

BR[3]とは，SIIT-DC ネットワークにおいて IPv4 インターネットと IPv6 ネットワークとの間で SIIT による IPv4/IPv6 トランスレーションを行う機器[4]である．IPv4 インターネットと IDC 内の IPv6 シングルスタックネットワークの各境界部に所在し，後述する EAMT を参照した 1:1 のアドレス変換を行う．IDC ネットワークに IPv4 インターネットとの接続点が複数ある場合，接続点ごとに最低一つの BR を配備する．

ER

ER[5]とは，IPv4 ネットワークと IPv6 ネットワークとの境界点において多:多の IPv4/IPv6 トランスレーションを行う機器である．SIIT-DC ではそのオプションとして，IPv4 ネットワーク内の IPv4 しか利用出来ないホストが，SIIT-DC を利用して IPv4 サービスを提供するユースケースをサポートする SIIT-DC Dual Translation Mode[26] が定義されており，ER はその中での利用が想定されている．

通常，ER が参照する EAMT 中の EAM は IDC ネットワーク内の IPv4 ネットワーク全体を包括的に指定した IPv4 ネットワークアドレスと，その IPv4 ネットワークを表す IPv6 サービスアドレスにより構成される．

IPv4 サービスアドレス

IPv4 サービスを提供する IPv6 シングルスタックネットワークに属するホストに割り当てる IPv4 アドレス (群) を IPv4 サービスアドレスと呼称する．このアドレス宛に送信されたパケットは，BR/ER によって対応する IPv6 サービスアドレスに変換される．

なお，IPv4 サービスアドレスは BGP[14] によって IPv4 インターネットに経路広告されている必要がある．

IPv6 サービスアドレス

ER/BR を介してアプリケーションやホストに割り当てられた IPv6 アドレス (群) を IPv4 サービスアドレスと呼称する．IPv4 クライアントは SIIT-DC のアーキテクチャをを介して，この IPv6 サービスアドレスが割り当てられたホストと通信することが出来る．

変換プレフィックス

変換プレフィックス[6]とは，RFC6052[22] で定義されたプロトコルに従って全ての IPv4 アドレスをマッピングするために用いられるネットワークプレフィックスが 96bit の IPv6

[3]Border Relay
[4]専用機器もしくは他の役割を有する機器の一機構.
[5]Edge Relay
[6]Translation Prefix.

プレフィックスである．IANA によって主に WKP[7] として 64:ff9b::/48 が予約 [27, 28] されているが，運用者の裁量で ISP 自身に割り当てられた NSP[8] を利用する事ができる．

IPv4 アドレスと IPv6 アドレスの間で変換を実行する際に，B は BR/ER は変換前の IP ヘッダーのアドレスフィールドを、変換プレフィックスが挿入・削除された状態に書き換える。

なお SIIT-DC ネットワークにおいて，変換プレフィックス宛のパケットは各 BR/ER の IPv6 インターフェース宛に IGP[9] などでルーティングされる必要がある．

EAM

EAM[10] とは，EAM アルゴリズム [29] によって結びつけられた IPv4 サービスアドレスと IPv6 サービスアドレスのペア[11] を表す．

EAM において，IPv4 サービスアドレスと IPv6 サービスアドレスは同数[12] である必要がある．

また，BR 及び ER が変換を行う際に参照する EAM 群が記録されたテーブルを EAMT[13] と定義している．以後 EAMT もしくは変換テーブルと呼称する．

3.1.3 ネットワーク設計

基本的な SIIT-DC ネットワークを図 3.1 に示す．

BR は IPv4 インターネットとの各接続点に配置される．各 IPv4 サービスアドレスは BGP により接続先の AS[14] に対して経路広告される．変換プレフィックス宛のパケットは各 BR に広告される．

BR が複数ある場合，それぞれの BR が IGP 等で広告する変換プレフィックスを分けるか，エニーキャスト [30] によって複数の BR が同一の変換プレフィックスを広告するようにする．エニーキャストを使用した場合，BR の障害時に別の BR へとトラフィックを迂回させることが可能になる．

ER は IDC 内の IPv4 ネットワークとの接続点に配置され，IPv4 のみを持つホストが IDC 内の IPv6 ネットワークを介して IPv4 インターネットにサービス提供を行う場合に利用される．

[7] Well Known Prefix.
[8] Network Specific Prefx. 主に RIR から割り当てられた IPv6 Global Unicas Address を指す.
[9] Interior Gateway Protocol
[10] Ecplicit Address Mapping
[11] サービスアドレスはそれぞれネットワークプレフィックスとして指定することも想定されている.
[12] 標準では結び付けられた IPv6 サービスアドレスが IPv4 サービスアドレスより多い状態が想定されているが，IPv6 サービスアドレスのホスト部が若いものから優先して変換するため，余剰分のアドレスは無視される.
[13] Ecplicit Address Mapping
[14] Aunotoumous System. インターネットを構成する自律した組織.

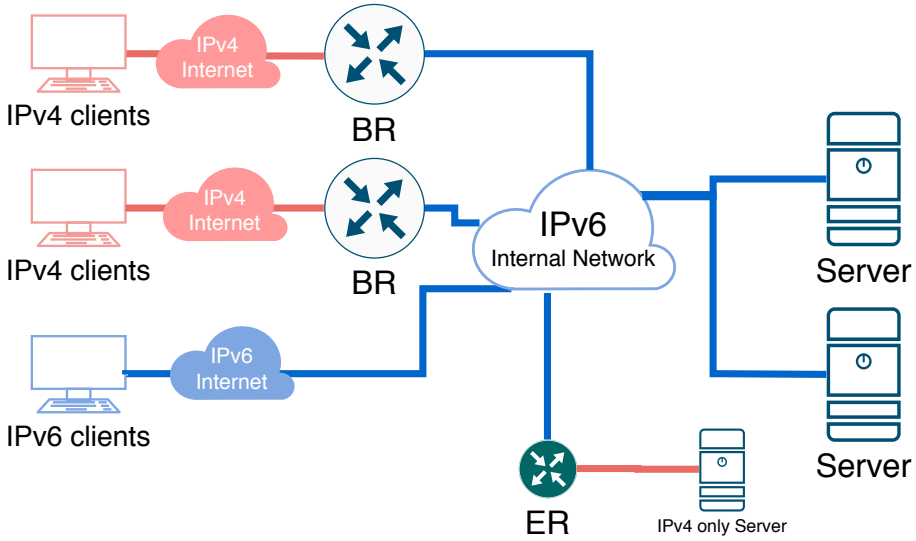

図 3.1: SIIT-DC ネットワーク

3.1.4 SII-DC のメリット

SIIT-DC を用いた IPv4 サービスの提供によるメリットとして，以下の様な物が挙げられる，

デプロイメントが容易

SIIT-DC は，IDC の IPv6 ネットワークと IPv4 インターネットとの接続点に BR を設置を行うのみで，対外的な基本的な IPv4 サービスの提供が可能だ．そのため IDC のネットワークトポロジーに限定されないシンプルな IPv4 サービス提供が期待できる．

アドレス単位での IPv4 アドレスの効率的な利用が可能

通常の IP ネットワークでは，サブネット単位での割当が必要である．従来，事前に同一サブネットに属するホスト数を見積持った上で不足が生じないようになネットワークサイズを設定しする必要があったため，ネットワークサイズをを超えるサービスの拡大が必要になった場合，サブネット全体の再設計が不可欠であった．また IP ネットワークには，ネットワークアドレスやブロードキャストアドレス，そしてデフォルトゲートウェイとなるルーターのインターフェースのアドレスを確保する必要があり，ネットワークサイズが断片化されるほど，実質的に利用できないアドレスの割合が大きくなる問題が会った．

しかしながら SIIT-DC ではアドレス単位での IPv4 提供サーバーへの割当が可能である．すなわち，従来利用できなかった IPv4 アドレスを再利用することで，IPv4 アドレスの効率的な活用を実現できる．第 1.1.1 で述べたように今後益々IPv4 アドレスの調達が難しくなっているため，IPv4 アドレスの効率的な利用は事業者の負担軽減に繋がる．

スケーラビリティ

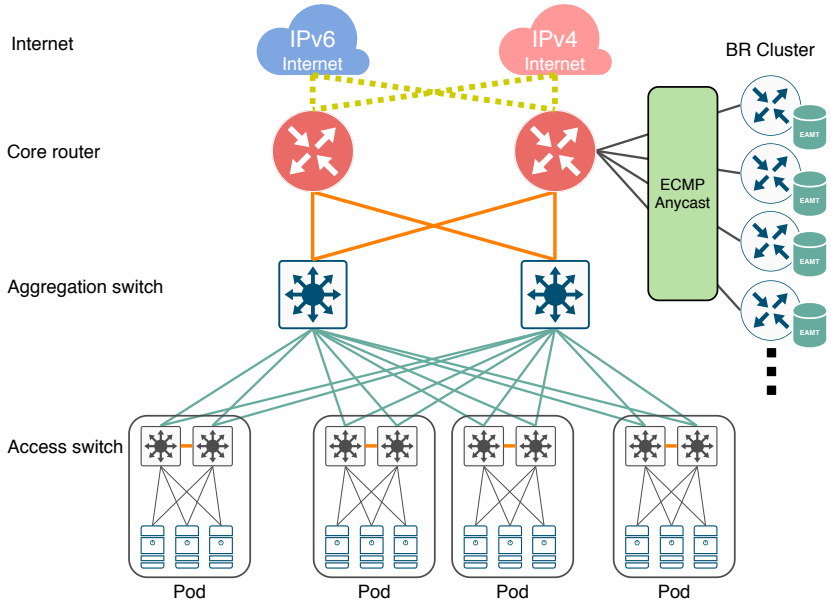

図 3.2: BR を水平スケールすることが出来る SIIT-DC ネットワーク

BR は IDC の IPv6 ネットワーク IPv4 インターネットに接続するためのネットワーク疎通性を用意することでサービスを行う．そのため，ECMP 及びエニーキャスト [30] を利用して，BR の数を追加するのみによって，IPv4 サービスの提供容量を増加させる事ができる．

図 3.2 にスケーラビリティを意識した BR 配置を行ったネットワークの例を示す．IPv4 クライアントからのアクセスは一度いずれかの BR にフォワーディングされた後，IPv6 プロトコルに変換された上で再度コアルーターを介して IDC ネットワークに到達する．

3.1.5 基本的なパケットの流れ

SIIT-DC における基本的な IPv4 クライアントからのトラフィックの流れは以下の様になる．一連のパケットの送信元・送信先のアドレスの遷移を図 3.3 に示す．

IPv4 クライアントの IPv4 サービスアドレス宛のパケットは IPv4 インターネットに接続する BR に到達後，当該 BR が有する EAMT に従って IPv6 サービスアドレス宛の IPv6 パケットに変換される．このパケットの送信元アドレスは変換プレフィックスに埋め込まれた IPv6 アドレスとして表現される．IDC 内の IPv6 ネットワークを介して IPv6 サーバーに到達した後，IPv6 サーバーは送信元アドレスへの応答パケットを送信する．第 3.1.2 項で述べたように，変換プレフィックス宛のパケットは IPv6 ネットワークを経由して BR にルーティングされる．IPv6 サーバーからの応答を受け取った BR は EAMT を参照し，送信元アドレス (IPv6 サービスアドレス) を IPv4 サービスアドレスに書き換え，送信先ア

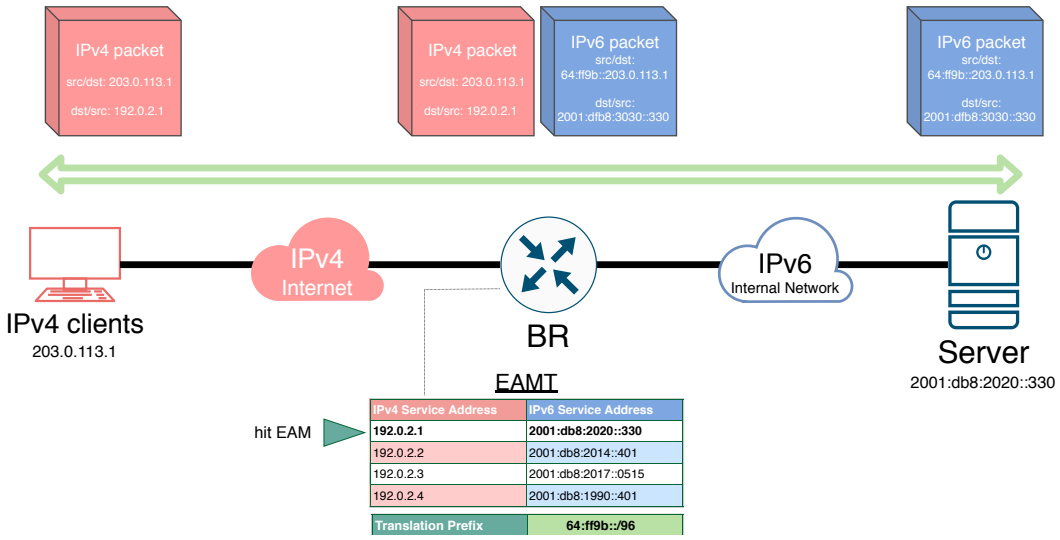

図 3.3: SIIT-DC パケットの流れ

ドレス (IPv4 クライアントの IPv4 アドレス) から変換プレフィックスを除去書き換えたのち，IPv4 インターネットを介して IPv4 クライアントに返送される．

3.2 SIIT-DC の課題

本節では SIIT-DC の現状の課題及びそれに起因して起こる事象に関して述べる．

3.2.1 一貫した EAMT の必要性

第 3.1.2 で述べたように，SIIT-DC では対外接続点ごとに BR を配置するネットワークデザインを採用することで，IPv6 シングルスタックネットワークに最小限の IPv4 ネットワークを追加するのみにより IPv4 サービスの提供を可能にしている．また第 3.1.3 項や第 3.1.4 項で触れたように，複数の BR で共通した変換プレフィックスをエニーキャストで IDC ネットワーク内に広告する運用を行うことにより，BR 及び対外接続点の障害時に他の BR を用いて IPv4 サービスの提供を継続することが出来る．この機構を有効に作用させるためには，SIIT-DC ネットワーク内の全ての BR で一貫した EAMT の保持が求められる．

しかしながら現状の SIIT-DC 及び EAMT の仕様 [12, 26, 29] では，BR は他の BR との間で EAMT を共有するためのメッセージング機構を有さない，これは BR 間で EAMT の不一致が発生した場合に，差異となった EAM に該当する IPv4 サービス宛のトラフィックの別の BR への迂回が出来なくなるケースが発生することを意味する．

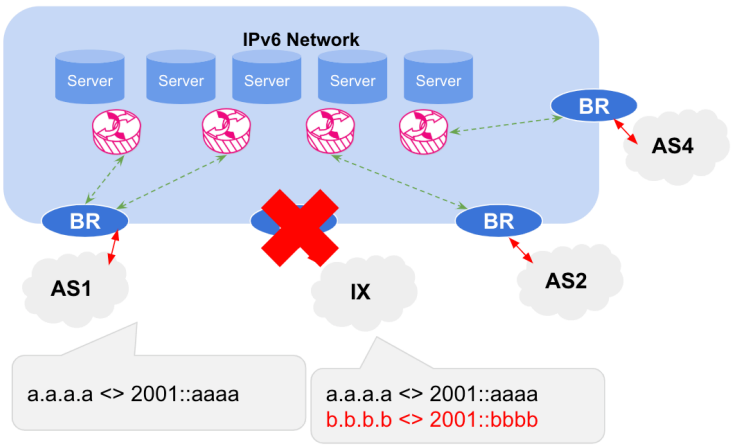

図 3.4: BR に障害が発生した場合に適切にフェイルオーバーが出来ないケース

3.2.2 変更追従性の欠如

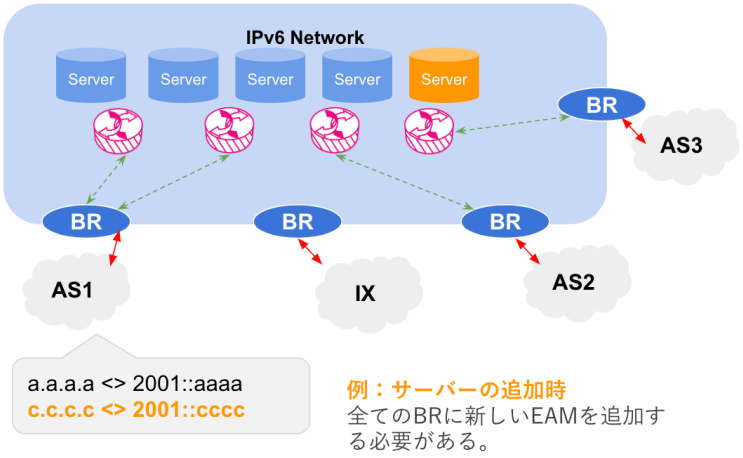

図 3.5: サーバーを追加した際，全ての BR への設定追加が必要になる．

プライベートクラウド環境が一般的に利用される IDC ネットワークでは，日々多くのサーバーやアプリケーションが追加・廃止・変更される．一方で第 3.2.1 で触れたように，SIIT-DC で IPv4 提供サービスを冗長に運用するためには，IPv4 提供サービスに該当する EAM が BR の EAMT に保持されることを要求している．IPv4 提供サービスの構成に変更があった場合，全ての BR の EAMT に対して EAM の更新を行わなくてはならない．

しかしながら現状 SIIT-DC 及び EAMT の仕様 [12, 26, 29] において，IPv4 サービスを行うサーバーの存在や状態によってダイナミックに EAMT を更新する機構は存在しない．そのため，IDC ネットワークにおける IGP などによって IPv6 サービスへの到達性が検証

されていたとしても，IPv4 サービスの場合はリアルタイムな構成変更に追従することが出来ない．

第4章　手法の検討

4.1　概要

第 3.2 項で述べたように，現状の SIIT-DC 及び EAMT の仕様は EAMT の一貫性を担保する手法の検討がなされておらず，それに起因した障害時の適切なフェイルオーバーの実行や IPv4 サービスの増減時の変更追従に関しての課題がある．NPO 日本ネットワークセキュリティ協会 (JNSA) らの調査によれば IT システムの障害の原因の約半数は人為ミスに分類されるものにあり [31]，サービスの安定的な稼働を実現するためには単調な繰り返し動作を含む運用をシステムによって減らす必要がある．

本研究では SIIT-DC におけるダイナミック EAMT[1] の実現を目指す．本章では考えられる手法を大別した上でその特徴と利点及び欠点を挙げ，最も適した手法を検討する．

4.2　求められる要件

第 2.1.1 で述べた IPv4 サービス提供手法の機能要件と，第 3.2 項で挙げた SIIT-DC の現状の課題を総合し，EAMT を動的に制御する手法に求められる要件を下記のように定義した．

BR 間の EAMT の一貫性

第 3.2 項で述べたように，障害時の適切なフェイルオーバーを実現するためには，各 BR の EAMT の一貫性が保証される必要がある．

変更追従性

近年の IDC では多数の物理サーバーを統合的に管理するプライベートクラウド環境やコンテナオーケストレーション環境[2] が普及しており，アプリケーション・サービスの追加及び削除が頻繁に行われている．サービスの障害時に適切にそれを検知し，適切に冗長系に移行するはは SLB[3] を中心として広く利用されている．SIIT-DC の IPv4 サービス提供の場でも，サービスの状態の変動にリニアに対応しフェイルオーバーできるような働きが求められる．

[1] BR 群の EAMT をシステムにより動的に制御する機構
[2] Container Orchestration. コンテナ型仮想化統合管理環境
[3] Servver Load Balancer

スケーラビリティ

第 2.1.1 で述べたように IPv6 シングルスタックネットワークにおける IPv4 サービスの提供では水平スケールが容易に行える仕組みを備える必要がある．IPv4 サービスを行うサーバーの増設や，対外接続点が増えた場合の BR の拡大に十分に適用するスケーラビリティを有することが望ましい．

次項ではスケーラビリティの評価のために，制御に必要な通信コネクション数による比較を行う．以後 BR の数を M，IPv4 サービスを提供するサーバーの数を N とし，総通信コネクション数を C として表現する．

デプロイメントの容易さ

第**??**で述べたように，SIIT-DC の最も特筆すべきメリットの一つにデプロイメントの容易さが挙げられる．これを損なうこと無くダイナミック EAMT が導入されることが望ましい．

4.3 アプローチの分類と比較

本説ではダイナミック EAMT を実現するアプローチとして，二系統のアプローチを考案する．それぞれのアプローチで考えられる実装と実際の構成，及び第 5.5 節で述べた各要件への適合を定性的に評価する．

4.3.1 中央管理型アプローチ

中央管理型アプローチとは，複数の BR の EAMT を統合的に管理する「コントローラー」を IDC ネットワーク上に配置し，各 BR がネットワークを介してこれを参照する機構である．図 4.1 に中央管理型アプローチによってダイナミック EAMT を実現した SIIT-DC の各コンポーネントの関係図を表す．

中央管理型アプローチではコントローラーが各 BR に投入する EAM が記録された「マスターテーブル」を保持し，それを元に各 BR のデータプレーンにルールを書き込む手法を取る．マスターテーブルに記載される EAM はオペレーターがネットワークの構成変更に合わせて追加・削除・更新を行い，それぞれの IPv4 サービスを提供するサーバー群に対してはコントローラからプル型[4]の外部監視[5]によりサーバーの状態変化を検知しマスターテーブルを更新する．

本アプローチの実装手法としては，OpenFlow[6]を用いた集中コントローラー型 SDN フレームワークを利用する方法が考えられる [32]．類似事例として，Sheng らによって Open

[4]pull-based monitoring. コントローラから各サーバーに能動的に情報を取得する

[5]External monitoring

[6]Open Networking Foundation により標準化されているデータプレーン制御用通信プロトコル．`https://www.opennetworking.org/`

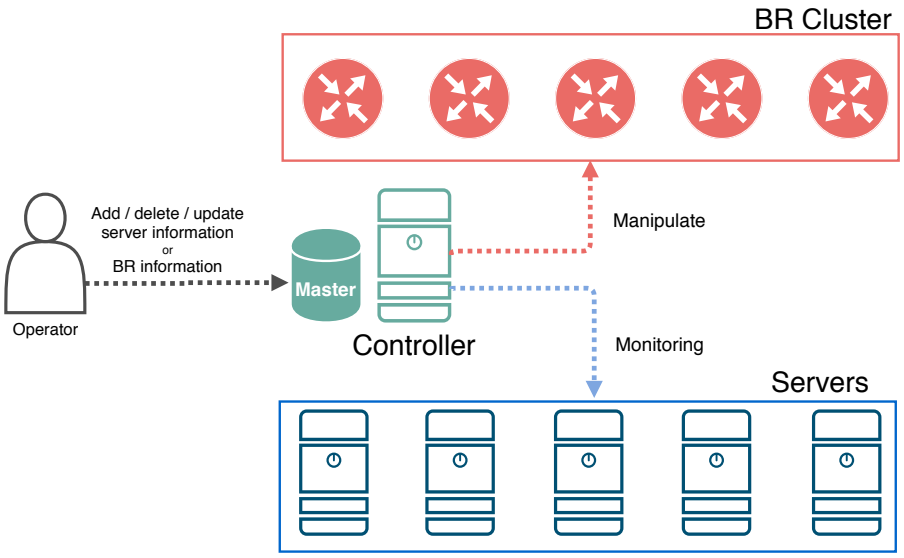

図 4.1: 中央管理型アプローチによるダイナミック EAMT

Flow を利用して各アクセススイッチに IPv4/IPv6 トランスレーション機構をデータプレーンとして導入するデータセンターネットワークデザインの提案がなされている [33].

要件評価

- BR 間の EAMT の一貫性
  本アプローチでは各 BR の EAMT が一つのマスターテーブルからレプリケーションされるために，十分な一貫性が保証される.

- 変更追従性
  基本的には EAM 情報の更新はオペレーターのマスターテーブルへの記入までの時間はコントローラーのサーバー監視性能に依存する.

- スケーラビリティ
  コントローラーの数を L とすると，EAMT の制御に必要とする総通信コネクション数 C は以下の通りになる.

$$C = L(M + N) \tag{4.1}$$

  一方，変更追従性と同じくどこまでの筐体を収容できるかはコントローラーの実装・性能がボトルネックになる設計となる.

- デプロイメントの容易さ
  コントローラーに求められる機器の性能・機能要件が大きいため，標準的な SIIT-DC よりデプロイメントのコストは向上する.

4.3.2 分散管理型アプローチ

分散管理型アプローチとは，IPv4 サービスを提供するサーバーがエージェントプロセスを介して自身の IPv4 サービスアドレスと IPv6 サービスアドレスを広告し，その広告情報を受け取った BR が自身の EAMT に反映させる機構である．図 4.2 に中央管理型アプローチによってダイナミック EAMT を実現した SIIT-DC の各コンポーネントの関係図を表す．

サーバー群は各 BR と EAM を広告するためのコネクションを確立する．IPv4 サービスを提供するサーバーと BR の間の IP ネットワークが何らかの原因により疎通不能になると，当該サーバーの広告も同時に停止されるため，該当 BR の EAMT から該当する EAM のレコードが削除される．

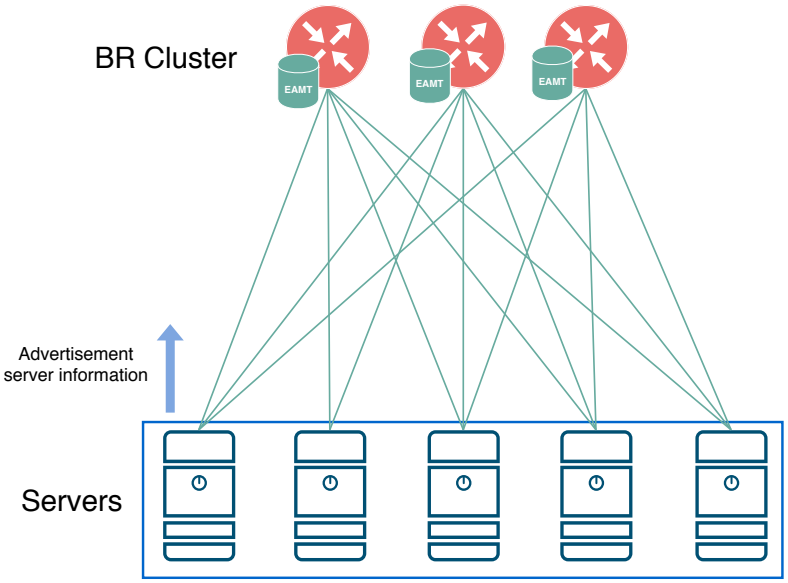

図 4.2: 分散管理型アプローチによるダイナミック EAMT

要件評価

- BR 間の EAMT の一貫性
  各 BR 間で EAMT 一貫性を保証する機構は無いが，当該 BR と疎通できないサーバーは障害時に自身の IPv4 サービスアドレう宛のトラフィックを当該 BR に経由させることが出来ないため，問題にならない．

- 変更追従性
  サーバー自身のエージェントプロセスが直接 BR に広告を行うため，実際の変更にリニアに対応出来る．

- スケーラビリティ
  EAMT の制御に必要とする通信コネクション数 C は以下の通りになる．

$$C = M \cdot N \tag{4.2}$$

  サーバー群・各 BR 間でフルメッシュでのコネクションが必要なため，SIIT-DC ネットワーク自体が小規模の場合のみ採用可能である．

- デプロイメントの容易さ
  各サーバー・BR にエージェントを導入する必要があるが，システム自体の機能は軽量である．

4.4 アプローチの検討

中央管理型アプローチが各 BR 間での EAMT 一貫性，スケーラビリティの二要素で優位であるが，コントローラーの役割が非常に大きくなり機能要件が高くなるため，変更追従性とデプロイメントの容易さの面での障壁が高いという問題も抱えている．一方で分散管理型アプローチはシンプルな構成であるためデプロイメントが比較的容易であり変更への追従がリニアであるが，各サーバーが通信コネクションを多量に貼らなくてはならない点でスケーラビリティに難がある．

本提案手法では両アプローチを総合した動的経路制御プロトコルである iBGP を利用したハイブリッド型アプローチを提案する．第 5 章において，本提案手法を中央管理型・分散型の両アプローチと比較する．

第5章　本提案手法の設計

第4章では，DynamicEAMTを設計する上で考えられる二種類のアプローチについて，求められる要件に照らし合わせて評価・検討を行った．本章では検討結果の得られた内容を基に，本研究において提案するDynamicEAMTの実現手法の設計に関して論じる．

5.1　概要

本提案手法では動的経路制御プロトコルであるBGPを利用したサーバー・BR間のメッセージングにより，SIIT-DCネットワークにおけるDynamicEAMTを実現する．加えて4.3で論じた中央管理型・分散管理型の両アプローチの利点を活用可能な，iBGP[1]・RR[2]構成を採用することにより，デプロイメントが容易且つスケーラビリティに優れたメッセージングが可能になる．

5.2　BGP

5.2.1　概要

BGPとはインターネットにおいて自律システム[3]間の経路情報交換に用いられるパスベクタ型の動的経路制御プロトコルである．現在有効なバージョンはBGP4であり，RFC4271で定義されている[14]．

5.2.2　用語

BGPにおいて利用される用語のうち，本提案手法において重要なものを以下に列記する．

BGPスピーカー

BGPを実装された機器をBGPスピーカーと呼ぶ．

[1]Internal BGP
[2]Route Reflector
[3]Autonomous System．インターネットを構成するネットワークをそれぞれ独立的に運用する組織群を指す．

BGP ピア

BGP で経路交換を行う関係にある機器をぞれぞれ BGP ピアと呼称する．

そのうち，自律システム間での接続関係にある BGP ピアを EBGP[4] ピア，同一自律システム内の BGP スピーカー同士の経路交換に用いられるピアを IBGP ピアと呼ぶ．

BGP コネクション

BGP コネクションとは BGP で経路交換に用いられる接続関係を指す．各機器は 1 対 1 の関係で BGP コネクションを確立する．BGP コネクションにはトランスポート層のプロトコルとして TCP/179[5] が利用され，フラグメンテーションや再送制御，応答確認，誤り制御等，TCP による高信頼なメッセージングが可能である．

また，BGP コネクションを維持・管理するために，BGP では以下のような 4 つのメッセージが定義されている．

BGP コネクションは BGP ピア間で TCP コネクションを確立したのちに OPEN メッセージにより各機能の対応関係を確認することにより確立され，KEEPALIVE メッセージによりセッションが維持される．UPDATE メッセージにより，BGP ピアへ広告する経路[6] に変更が生じたことを通知する．何らかの理由により BGP コネクションが確立出来なくった場合，NOTIFICATION メッセージを利用して切断を通知する．

Adj-RIB-In/Adj-RIB-Out/Loc-RIB

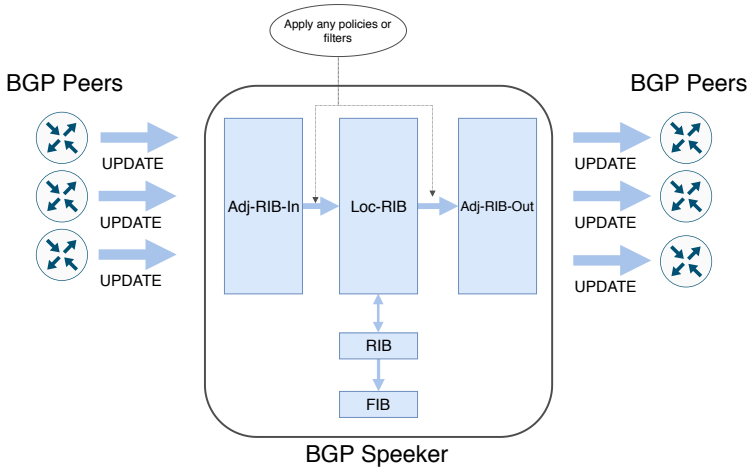

図 5.1: BGP スピーカーの経路の扱い

[4] External BGP
[5] TCP[34] のポート番号 179
[6] Adj-RIB-Out.

図 5.1 に BGP における経路受信・保持・送信の流れを示す．BGP ピアから受信した経路は Adj-RIB-In と呼ばれ，BGP スピーカーの任意のフィルターやポリシーを適用した上で Loc-RIB と呼ばれるテーブルに保存される．BGP スピーカーは Loc-RIB から任意のフィルターを適用した経路を BGP ピアに広告する．この広告する経路を Adj-RIB-Out と呼ぶ．

5.2.3 特徴

本提案手法において DynamicEAMT を実現するためのメッセージングプロトコルとして BGP を選択するに至った要素について記述する．

マルチプロトコル

現行版である BGP4 では，OPEN メッセージにオプション値[7]を挿入することで，IANA によって定められた任意のネットワークプロトコル [35, 36] の経路を交換することが想定されている [37]．本提案手法で利用している IPv6 ユニキャスト経路もこの機構により実装されている．

実装が一般的

BGP は自律システム間の経路交換プロトコルとしてインターネットで利用されている実質的に唯一のプロトコルであり，OSS[8]にも多くの実装が存在する．広く普及したプロトコルを利用することにより，特別な実装を最小限にして本提案手法を実現することが出来る．

中継ネットワークに非依存

本提案手法で採用している IBGP では，TTL[9]が 255 に設定されており，BGP ピア間で IPv4/IPv6 による到達性があればメッセージングを行うことが可能である．すなわち本提案手法は既存の SIIT-DC ネットワークに非依存であり，これは第 5.5 項で述べた要件の一つである，デプロイメントの容易さを充足する．

5.3 基本的なネットワーク設計

図 5.2 本提案手法の各要素の関係を示す．

[7] Capabilities Optional Parameter.
[8] Open Source Software
[9] Time to Live. そのパケットが宛先ホストに到達するまでに許容される中継ルーター数．IPv6 プロトコルでは Hop Limit として同一の機能が実装されている [38].

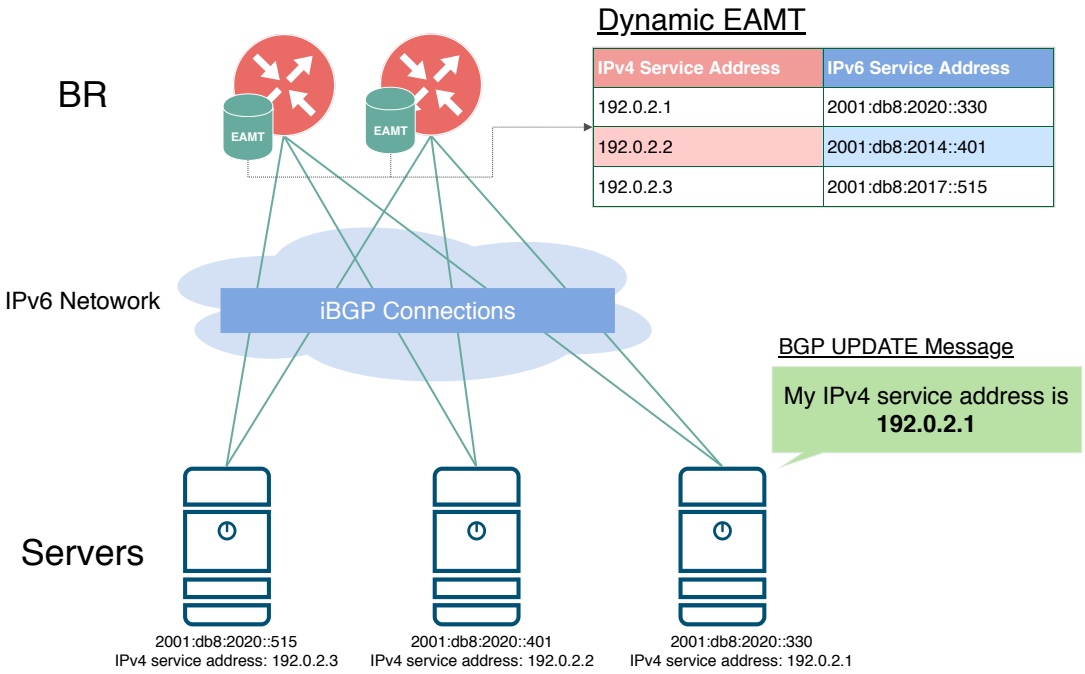

図 5.2: 本提案手法の基本機能を実装した SIIT-DC ネットワークの例

BR 数を N，サーバー数を M とした，ルートリフレクターを利用した本提案手法での必要な BGP コネクション数 C_a は式 5.1 のように表現できる．

$$C_a = M \cdot N \tag{5.1}$$

5.3.1 各ノードの役割と機能要件

BR

BR では下記のような 3 つの機能が必要となる．

- BGP デーモン
  各サーバーと iBGP コネクションを確立し，Loc-RIB を生成する．
- SIIT 機構
  EAMT を保持し，それを参照して IPv4/IPv6 プロトコル変換を行う．
- EAMT 制御機構
  BGP デーモンが有するの Loc-RIB を参照し，EAMT を更新する．

IPv4 サービス提供サーバー

IPv4 サービス提供サーバーでは以下の 2 つの機構が求められる．

- IPv4 サービス
  IPv4 によりインターネットに提供したいサービスを稼働させる．
- BGP デーモン
  自身が提供する IPv4 サービスアドレスを含んだ情報を広告する．

5.4 ルートリフレクターを活用したネットワーク設計

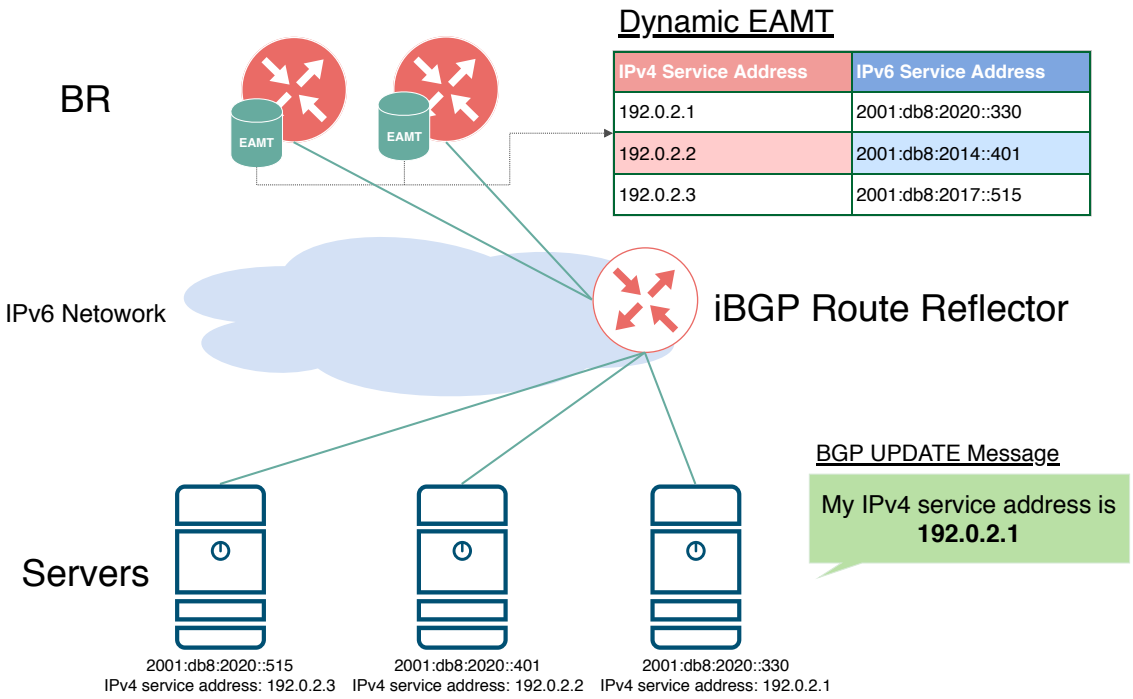

図 5.3: ルートリフレクターを採用した SIIT-DC ネットワークの例

通常，iBGP ではルートループを防ぐために異なる BGP ピアから受信した経路は他の BGP ピアに広告されない．そのため一つの iBGP スピーカーが広告する経路を他の iBGP スピーカーが受信するためには，BGP コネクションをフルメッシュで確立する必要がある [39]．

ルートリフレクターとは，Originator-ID と呼ばれる特殊な属性を Adj-RIB-Out に付与することでルートループを防ぎながら，iBGP ピアから受信した経路を他の iBGP に対して広告する特殊な BGP スピーカーである [40]．ルートリフレクターは iBGP のコネクション数を削減するために広く利用されている．

ルートリフレクターを複数設置することで，負荷分散・冗長化構成を容易に実現することが出来る．一般的にはルートリフレクター間はフルメッシュでの BGP コネクションを確立する設計を行うが，Gutiérrez らによればツリー型の BGP コネクション関係を一部で

選択することにより，よりルートリフレクターに掛かる負荷出来ることが明らかになっている [41].

BR 数を N，サーバー数を M，ルートリフレクターの数を L とした，ルートリフレクターを利用した本提案手法での必要な BGP コネクション数 C_b は式 5.2 のように表現できる．なおルートリフレクター間の BGP コネクションはフルメッシュを想定している．第 5.3 節で述べた基本的なネットワーク設計を行う場合と比較して，SIIT-DC ネットワークが大きくなった場合に BGP コネクションが大幅に削減できることがわかる．

$$C_b = \frac{L(2M + 2N + L - 1)}{2} \tag{5.2}$$

5.4.1 各ノードの役割と機能要件

BR 及び IPv4 提供サーバー

BR 及び IPv4 サービス提供サーバーは，ルートリフレクターとのみ BGP コネクションを確立する．複数ルートリフレクターを配備する場合，それぞれとコネクションを確立することで冗長性を高めることが出来る．その他の機能は第 5.3 で述べたものと同様に配備する．

ルートリフレクター

ルートリフレクターでは，ルートリフレクター機能が有効となった BGP デーモンを配備する必要がある．各サーバー・BR と BGP コネクションを確立する．

5.5 各アプローチとの比較

第 4.3 節で検討した各アプローチと本提案手法を，第節で挙げた各性能要件に関して比較する．

表 5.1 にそれぞれの項目における比較結果を表す．なお，コントローラーもしくはルートリフレクターの導入数を L，サーバー数を M，BR の数を N としている．

表 5.1: 各手法の比較

手法	EAMT の一貫性	変更追従性	コネクション数	デプロイメントの容易さ
参考: オペレーターによる手動設定	無し	無し	——	——
中央管理型アプローチ	有り	(監視機構の実装依存)	$\frac{L(2M+2N+L-1)}{2}$	(コントローラーの実装依存)
分散管理型アプローチ	無し	有り	$M \cdot N$	有り
提案手法 1: iBGP	有り	有り	$M \cdot N$	有り
提案手法 2: iBGP + ルートリフレクター	有り	有り	$\frac{L(2M+2N+L-1)}{2}$	有り (RR は容易に水平スケール可能)

第6章　プロトコル設計と実装

本章では，第5章で述べた提案システムのメッセージ設計と実装について述べる．

6.1　BGP UPDATEメッセージの設計

本提案手法ではサーバー・BR・ルートリフレクター間のメッセージングにBGPを利用する．本節ではBGP UPDATEメッセージの設計を行う．

6.1.1　要件

Dynamic EAMTを実現するにあたって，EAMとして広告すべきに必要な属性は1) IPv4サービスアドレス，2)IPv6サービスアドレス, 3) 変換プレフィックスの3種が想定される．表6.1.1に各属性の情報を列記する．

表 6.1: EAMに必要な情報

属性名	型	備考	例
IPv4 サービスアドレス	IPv4 ネットワークアドレス	IPv6 サービスアドレスとホストアドレス長が一致	192.0.2.1/32
IPv6 サービスアドレス	IPv6 ネットワークアドレス	IPv4 サービスアドレスとホストアドレス長が一致	2001:db8:200::1/128
変換プレフィックス	IPv6 ネットワークアドレス (/96)		64:ff9b::/96

6.1.2　実装

本提案手法では，IPv6ユニキャスト経路[1]として，BGPを利用してEAMを広告・交換する．UPDATEメッセージ以外の扱いは標準的なBGPメッセージに準ずる．

BGP UPDATEメッセージ

本提案手法におけるBGP UPDATEメッセージに含有するパス属性[2]を図6.1.2のように定義した．

[1]アドレスファミリー番号 2, サブアドレスファミリー番号 1[35, 36]
[2]Path Attributes

表 6.2: BGP UPDATE メッセージにおける各パス属性

タイプコード値	パス属性	必須	値	備考	例
1	ORIGIN	必須	2(IMCOMPLETE)	本実装においては利用しない.	2
2	AS_PATH	必須	AS 番号	iBGP のみで広告するため，自身の AS 番号を記載する	65001
5	LOCAL_PREF	任意	1 ～65535	EAM の優先度	200
8	COMMUNITY	任意	[0～65535]:[0～65535]	BGP コミュニティ名	2500:200
9	ORIGINATOR_ID	必須	BGP Identifier	自身のルーター ID	192.0.2.1
10	CLUSTER_LIST	任意	クラスター ID	ルートリフレクターを利用する場合，要指定 同じ EAMT を共有する範囲を指定する	192.0.2.1
14	MP_REACH_NLRI ->NLRI	必須	IPv6 アドレス+プレフィックス長	変換プレフィックス + IPv4 サービスアドレス/128	64:ff9b::192.0.2.1/128
14	MP_REACH_NLRI ->NEXT_HOP	必須	IPv6 アドレス	変換プレフィックス + IPv4 サービスアドレス/128	2001:db8:200::1
15	MP_UNREACH_NLRI	必須		MP_REACH_NLRI と同様	

6.1.3 実装時に留意すべき事項

BGP を利用した Dynamic EAMT において留意すべき事項を述べる.

ルーティングテーブルの隔離

通常の IGP・EGP 経路とは用途が異なるため，何らかの仮想化技術を利用してそれらと EAMT を BGP スピーカーが区別する必要がある．具体的には VRF[3]などのルーティングテーブル仮想化技術の利用が想定される.

ホストルートでの利用に限定

本提案手法では MP_REACH_NLRI 及び MP_REACH_NLRI のアドレスファミリーとして IPv6 ユニキャスト経路を利用している．そのため実装上の問題から，IPv6 サービスアドレス及び IPv4 サービスアドレスがそれぞれ 1 アドレスの場合のみをサポートしている.

6.2 PoC の実装

第 7 章で行う概念検証実験にもちいる PoC[4]について，各ノードで必要なコンポーネントとその役割及び具体的な実装について記述する.

6.2.1 各コンポーネントの実装

第 5.3.1 項で述べたコンポーネント群は以下の様にそれぞれ実装した．BR における BR に必要なコンポーネント群の関係図を図 6.2 に示す．表 ref にコンポーネント群の情報の概要を示す.

[3]Virtual routing and forwarding
[4]Proof of Concept. 概念検証実装

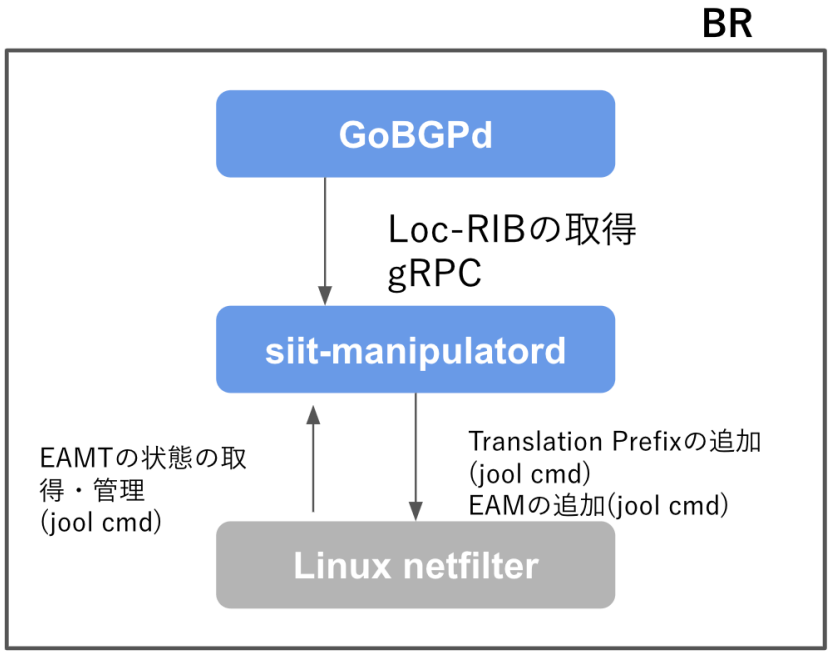

図 6.1: BR に必要なコンポーネント群の関係図

BGP デーモン

BGP デーモンには，OSS の BGP デーモンである GoBGP[5]を利用する．

GoBGP では gRPC[6]を用いた操作機構が実装されており，同期・非同期を問わず他のアプリケーションとの連携が容易に行える．RouteReflector 機構もサポートされているため，本 PoC では全てのノードの BGP デーモンとしてこれを利用する．

SIIT

SIIT には Jool[7]の SIIT モードを利用する．

Jool は LinuxOS で利用できる NAT64/SIIT 環境で，Lunux Netfilter によって実装されており，汎用的に様々なプラットフォームでの利用が可能である [42, 43]．EAMT の変更には専用の CLI コマンドを利用する．

EAMT 制御機構

BGP 上で受信した Loc-RIB を EAMT に反映するために，EAMT 制御機構”siit-manipulatord”を実装した．gRPC によって GoBGP の Loc-RIB の変化を Jool の CLI コマンドを利用し

[5] https://osrg.github.io/gobgp/
[6] gRPC Remote Procedure Calls. https://www.grpc.io/
[7] Jool. https://jool.mx/en/index.html

て Linux NetFilter 反映するほか、Translation Prefix の定義など SIIT に必要な情報を管理する.

6.2.2 メッセージングと状態遷移

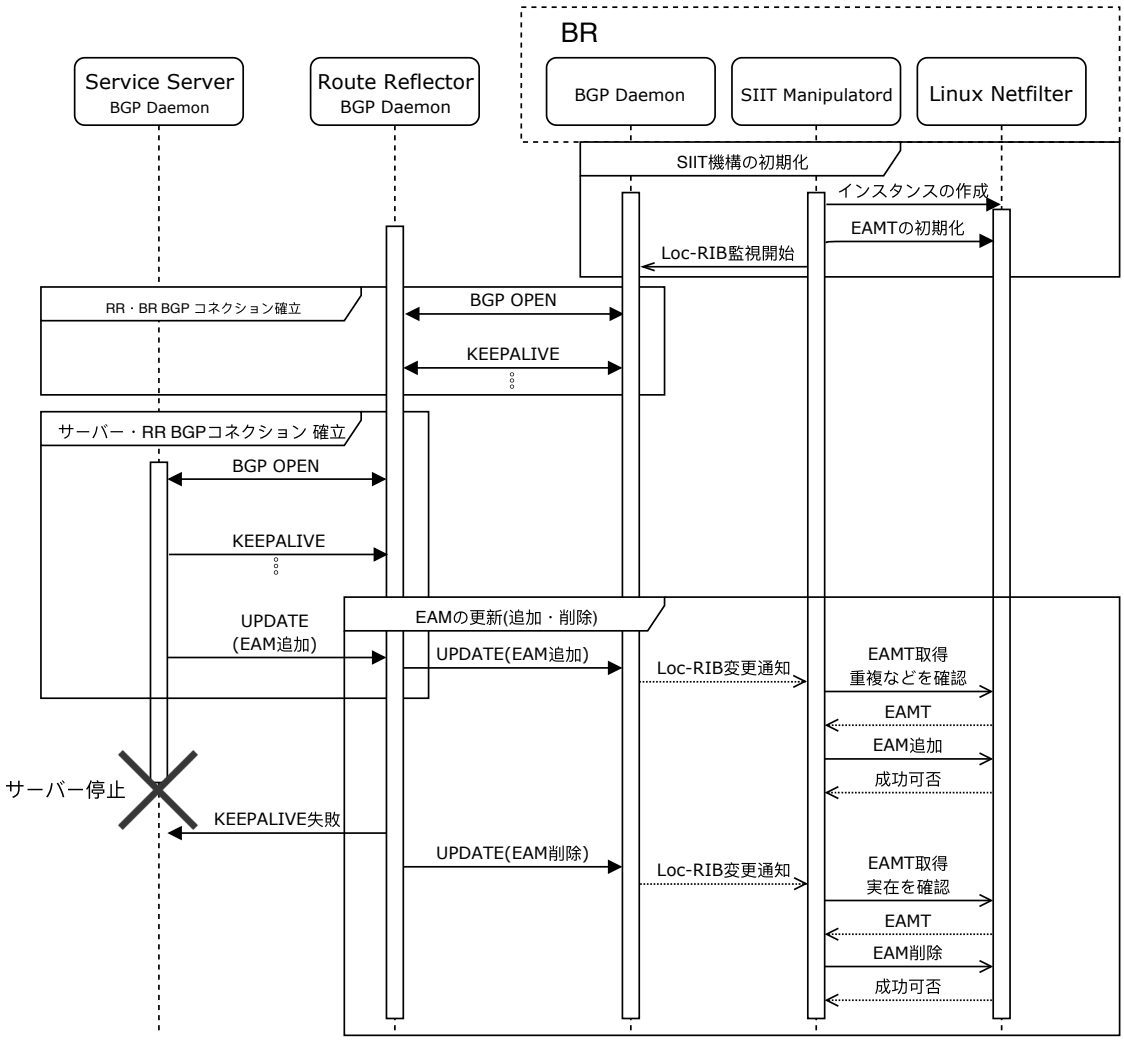

図 6.2: 本 PoC における各ホスト・コンポーネントの相互作用と状態遷移

図 6.2 に IPv4 サービス提供サーバー・ルートリフレクター・BR の相互作用と状態遷移の概要を示す.

以下に各状態にあるホスト・コンポーネント間の相互作用に関して記述する.

6.2.3 SIIT 機構の初期化

BR は起動後ネットワーク環境の準備が出来次第，BGP デーモンと SIIT 制御機構のサービスを開始する．EAMT 制御機構は初めに Linux Netfilter を操作し，SIIT によるプロトコル変換を行うインスタンス[8]を作成する．このインスタンスは BR が利用する変換プレフィックスを保持する．その後 EAMT を初期化し，gRPC を利用して BGP デーモンから Loc-RIB の監視を開始する．以後 Loc-RIB に変更があるまで待機する．Loc-RIB の変更はベストパスの変更に伴っても発生するため，BGP NOTIFICATION メッセージなどにより，ベストパスとなる経路情報を発信していた BGP ピアとのコネクションが切断されたような場合にも，SIIT 制御機構に変更通知が送信される．

6.2.4 ルートリフレクター・BR 間の BGP コネクションの確立と維持

BR の BGP デーモンは起動次第，事前に登録されたルートリフレクターに対する BGP OPEN メッセージの送信を開始し，BGP コネクションの確立を試みる．RR からの BGP OPEN メッセージの回答を受けて BGP コネクションを確立し，以後 BGP KEEPALIVE メッセージによって接続性の死活監視を行う．

6.2.5 IPv4 サービス提供サーバー・ルートリフレクター間の BGP コネクションの確立と維持

IPv4 サービス提供サーバーの BGP デーモンは起動に伴って，自身の Loc-RIB に EAM(変換プレフィックスと IPv4 サービスアドレスによって構成された NLRI を埋め込んだ経路情報) を登録する．BGP コネクションが確立次第，IPv4 サービス提供サーバーは UPDATE メッセージにより EAM の追加をルートリフレクターに通知する．このコネクションでも同様に，以後 BGP KEEPALIVE メッセージによって接続性の死活監視を行う．

6.2.6 EAM の追加

IPv4 サービス提供サーバーから UPDATE メッセージはルートリフレクターにより各 BR に伝達され，UPDATE メッセージを受信した BR の BGP デーモンは自身の Loc-RIB を更新する．

Loc-RIB の更新に伴い Loc-RIB の監視を行っていた SIIT 制御機構に gRPC を介し更新通知が送信される．以後の SIIT 制御機構の処理はコルーチンを利用した非同期処理によって行われるため，EAMT 走査にまつわる処理を行っている最中であっても Loc-RIB の監視はブロッキングされない．

SIIT 制御機構は EAMT の更新を滞りなく行うため，LinuxNetfilter に登録された現在の EAMT の状態を取得し，競合する EAM が登録されていないことを確認する．IPv4 サー

[8]Jool Instance. ネットワーク名前空間ごとに一つ存在可能である．以後変換インスタンスと呼称する．

ビスアドレスと IPv6 サービスアドレスが一致する EAM が存在していた場合や事前に登録された変換プレフィックスに一致しない場合，以後の処理をスキップする．

新しく受信した EAM に問題がない場合，SIIT 制御機構は EAM の追加を試みる．Linux Nefilter は新しい EAM が追加された EAMT を参照し，直ちにパケットのフォワーディングを開始する．

6.2.7 EAM の削除

ルートリフレクターは，何らかの理由で IPv4 サービス提供サーバーとの BGP KEEPALIVE メッセージが失敗した場合や BGP NOTIFICATION メッセージを受信した場合，自身の Loc-RIB から該当の EAM を削除し，Adj-RIB-OUT の情報を更新する．

ルートリフレクターから EAM 削除を広告する BGP UPDATE メッセージを受け取った BR の BGP デーモンは Loc-RIB を更新し，SIIT 制御機構に対して Loc-RIB 変更通知を送信する．EAM の追加時と同様に，EAMT を取得して該当する EAM の実在を確認後，SIIT 制御機構は EAM 削除を試みる．

6.2.8 EAM の更新

BGP 経路情報の属性値の変更などに伴って BR の Loc-RIB に登録されるベストパスが変更になる場合がある．

BR の BGP デーモンは EAM 追加・削除時同様に，SIIT 制御機構に対して Loc-RIB 変更通知を送信する．この変更通知に伴って EAMT 上の古い EAM は削除され，新しい EAM に更新される．

第7章　評価

本章では，第5章及び第6章で設計・実装に関して述べた本提案手法に関して，第3.2節で指摘したSIIT-DCの課題に対して有効性があることに対して定量的に評価を行う．

7.1　評価要件

SIIT-DCにおけるDynamicEAMT機構では，第5.5節で述べた事項が求められる．それらを定量的に評価するための指標・尺度について記述する．

7.1.1　BR間のEAMTの一貫性

SIIT-DCネットワークにおいて，各BRが一貫したEAMTを持つまでの時間を計測する．

7.1.2　変更追従性

サーバー構成が変更になった場合に，どれだけの時間でサービス提供が行えるかを計測する．

7.1.3　スケーラビリティ

第5.5節で示したように，本提案手法はルートリフレクターを導入することにより，ネットワーク全体でのBGPコネクション数を低減出来ることがわかっている．本実験においてはネットワーク規模の拡大[1]がBR間のEAMTの一貫性と変更追従性の2指標に対して，どのような影響を及ぼすかを検証する．

7.2　想定するネットワークトポロジー

第3.1.4節で述べたように，SIIT-DCは様々なネットワークトポロジーでの活用が可能である．

[1]IPv4サービスを提供するサーバーとBRがそれぞれ多くなった場合

7.3 実験環境

第項で記述した PoC 実装を利用する.
ネットワークエミュレータ環境として，GNS3 を利用する.

7.4 実験シナリオ 1: SIIT-DC ネットワークの構築

このシナリオにおける，BR の数, サーバーの数を変数とした時に変化を調べる.

7.4.1 ネットワーク構成

図を張る． BR とサーバーは水平スケールさせる.

7.4.2 実験結果

EAMT が一貫するまでの時間を調べる.

7.4.3 考察

一貫性保持が成功.
線形に変化しており，O(n) の十分なスケーラビリティがある.

7.5 実験シナリオ 2: サーバーの構成変更

このシナリオにおける，BR の数, サーバーの数を変数とした時に変化を調べる.
サーバーの削除からの Failover(LP で重み替え)

7.5.1 ネットワーク構成

図を張る． BR とサーバーは水平スケールさせる.

7.5.2 実験結果

追加したサーバー 1 台がサービス開始出来るまでの時間を，サーバー台数と BR 台数が増えていった場合にも変わらず出来ることを証明.

7.5.3 考察

リニアに追従.
線形に変化しており，O(n) の十分なスケーラビリティがある.

第8章　結論

本章では，本研究のまとめと今後の課題を示す．

8.1　本研究のまとめ

IPv6 のみ構築された IPv6 シングルスタックネットワークにおいて既存の IPv4 クライアントに対してサービスを提供する方法として、ステートレスアドレス変換を利用した”SIIT-DC”と呼ばれるネットワークデザインがインターネット標準として標準化されている．SIIT-DC では BR(Border Relay) と呼ばれる変換ノードを IPv4 ネットワーク・インターネットとの境界点ごとに設置し，明示的アドレス変換テーブル (EAMT: Eplicit Address Mapping Table) を参照してプロトコル変換を行い，IPv6 ノードでの IPv4 サービス提供を可能にする．しかしながら SIIT-DC では EAMT の動的な交換方法についての定義がなされておらず，対外接続点が複数存在する場合の冗長性の維持が難しい点や，IPv4 でサービス提供を行なうサーバーの構成変更が行われた場合に運用負荷が非常に高くなる点が課題に挙げられる．

本研究では BGP を利用したアドレス変換テーブルの広告・更新技術と，それを適切に運用するために必要なノード群の設計手法を提案する．これにより，SIIT-DC の課題であった冗長性の維持や構成変更へのに対して，ダイナミックに対応することが可能になる．

この手法を評価するために，新たに BGP によるアドレス変換テーブル制御機構を実装したソフトウェアルーターを実装し，多くの対外接続点を持つ学術 ISP である WIDE Project のバックボーンネットワークをモデルケースに，エミュレータを用いて概念検証実験を行った．考えられる他の手法と比較し，本手法が冗長性と変更追従性の点で優位であることが証明された．

8.2　本研究の課題

- SIIT-DC Dual Translation Mode へのデザインと対応
- ホストルート以外に対応
- 実環境でのテスト

謝辞

俺に関わった全てに感謝

参考文献

[1] potaroo. Ipv4 address report. https://ipv4.potaroo.net/. 最終閲覧: 2019-12-17.

[2] Cisco. Cisco visual networking index: Forecast and trends, 2017–2022 white paper, 2017. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html.

[3] Felipe Alonso and John Boucher. Business continuity plans for disaster response. *The CPA Journal*, 71(11):60, 2001.

[4] 石田慶樹, 吉田友哉, and 西田圭. 日本のインターネットは本当にロバストになったのか？ In *JANOG 44* ミーティング, 2019. https://www.janog.gr.jp/meeting/janog44/application/files/7715/6577/5523/janog44-robust-ishida-01.pdf.

[5] IANA. Internet protocol version 4 address space. https://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xml, 2019. 最終閲覧: 2019-12-17.

[6] Lee Howard and Time Warner Cable. Internet access pricing in a post-ipv4 runout world. *White Paper*, 2013.

[7] Bob Hinden and Dr. Steve E. Deering. Internet Protocol, Version 6 (IPv6) Specification. RFC 2460, December 1998.

[8] Cindy Morgan. Iab statement on ipv6. https://www.iab.org/2016/11/07/iab-statement-on-ipv6/, 2016. 最終閲覧: 2019-12-17.

[9] Alain Durand. Deploying ipv6. *IEEE Internet Computing*, 5(1):79–81, 2001.

[10] 北口善明, 近堂徹, 鈴田伊知郎, 小林貴之, 前野譲二, et al. クライアント os の ipv6 実装検証とネットワーク運用における課題. 研究報告インターネットと運用技術 *(IOT)*, 2017(13):1–8, 2017.

[11] Google. Ipv6 statistics. https://www.google.com/intl/en/ipv6/statistics.html. 最終閲覧: 2019-12-18.

[12] Tore Anderson. SIIT-DC: Stateless IP/ICMP Translation for IPv6 Data Center Environments. RFC 7755, February 2016.

[13] Congxiao Bao, Xing Li, Fred Baker, Tore Anderson, and Fernando Gont. IP/ICMP Translation Algorithm. RFC 7915, June 2016.

[14] Yakov Rekhter, Susan Hares, and Tony Li. A Border Gateway Protocol 4 (BGP-4). RFC 4271, January 2006.

[15] Jordi Palet, Hans M.-H. Liu, and Masanobu Kawashima. Requirements for IPv6 Customer Edge Routers to Support IPv4-as-a-Service. RFC 8585, May 2019.

[16] A. Vahdat, M. Al-Fares, N. Farrington, R. N. Mysore, G. Porter, and S. Radhakrishnan. Scale-out networking in the data center. *IEEE Micro*, 30(4):29–41, July 2010.

[17] Katja Gilly, Carlos Juiz, and Ramon Puigjaner. An up-to-date survey in web load balancing. *World Wide Web*, 14(2):105–131, Mar 2011.

[18] Patrick Shuff. Building a billion user load balancer. Dublin, May 2015. USENIX Association.

[19] Daniel E. Eisenbud, Cheng Yi, Carlo Contavalli, Cody Smith, Roman Kononov, Eric Mann-Hielscher, Ardas Cilingiroglu, Bin Cheyney, Wentao Shang, and Jinnah Dylan Hosein. Maglev: A fast and reliable software network load balancer. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, pages 523–535, Santa Clara, CA, 2016.

[20] N. Chuangchunsong, S. Kamolphiwong, T. Kamolphiwong, R. Elz, and P. Pongpaibool. Performance evaluation of ipv4/ipv6 transition mechanisms: Ipv4-in-ipv6 tunneling techniques. In *The International Conference on Information Networking 2014 (ICOIN2014)*, pages 238–243, Feb 2014.

[21] Philip Matthews, Iljitsch van Beijnum, and Marcelo Bagnulo. Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers. RFC 6146, April 2011.

[22] Xing Li, Mohamed Boucadair, Christian Huitema, Marcelo Bagnulo, and Congxiao Bao. IPv6 Addressing of IPv4/IPv6 Translators. RFC 6052, October 2010.

[23] Christian Hopps. Analysis of an Equal-Cost Multi-Path Algorithm. RFC 2992, November 2000.

[24] Erik Nordmark. Stateless IP/ICMP Translation Algorithm (SIIT). RFC 2765, February 2000.

[25] Xing Li, Fred Baker, and Congxiao Bao. IP/ICMP Translation Algorithm. RFC 6145, April 2011.

[26] Tore Anderson and S.J.M. Steffann. Stateless IP/ICMP Translation for IPv6 Internet Data Center Environments (SIIT-DC): Dual Translation Mode. RFC 7756, February 2016.

[27] Tore Anderson. Local-Use IPv4/IPv6 Translation Prefix. RFC 8215, August 2017.

[28] IANA. Internet protocol version 6 address space. `https://www.iana.org/assignments/ipv6-address-space/ipv6-address-space.xhtml`, 2019. 最終閲覧: 2019-12-17.

[29] Tore Anderson and Alberto Leiva. Explicit Address Mappings for Stateless IP/ICMP Translation. RFC 7757, February 2016.

[30] Kurt Erik Lindqvist and Joe Abley. Operation of Anycast Services. RFC 4786, December 2006.

[31] NPO 日本ネットワークセキュリティ協会 (JNSA). 情報セキュリティインシデントに関する調査報告書. `https://www.jnsa.org/result/incident/2018.html`, 2018. 最終閲覧: 2019-12-21.

[32] Evangelos Haleplidis, Kostas Pentikousis, Spyros Denazis, Jamal Hadi Salim, David Meyer, and Odysseas Koufopavlou. Software-Defined Networking (SDN): Layers and Architecture Terminology. RFC 7426, January 2015.

[33] S. Maojia, B. Congxiao, and L. Xing. A sdn for multi-tenant data center based on ipv6 transition method. In *2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*, pages 190–195, May 2016.

[34] Transmission Control Protocol. RFC 793, September 1981.

[35] IANA. Address family numbers. `https://www.iana.org/assignments/address-family-numbers/address-family-numbers.xml`, 2019. 最終閲覧: 2019-12-21.

[36] IANA. Subsequent address family identifiers (safi) parameters. `https://www.iana.org/assignments/safi-namespace/safi-namespace.xhtml`, 2019. 最終閲覧: 2019-12-21.

[37] Tony J. Bates, Ravi Chandra, Yakov Rekhter, and Dave Katz. Multiprotocol Extensions for BGP-4. RFC 4760, January 2007.

[38] Dr. Steve E. Deering and Bob Hinden. Internet Protocol, Version 6 (IPv6) Specification. RFC 8200, July 2017.

[39] Mythili Vutukuru, Paul Valiant, Swastik Kopparty, and Hari Balakrishnan. How to construct a correct and scalable ibgp configuration. 2005.

[40] Enke Chen, Tony J. Bates, and Ravi Chandra. BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP). RFC 4456, April 2006.

[41] E. Gutiérrez, D. Agriel, E. Saenz, and E. Grampín. Rrloc: A tool for ibgp route reflector topology planning and experimentation. In *2014 IEEE Network Operations and Management Symposium (NOMS)*, pages 1–4, May 2014.

[42] Jool. Subsequent address family identifiers (safi) parameters. `https://jool.mx/en/intro-jool.html`, 2019. 最終閲覧: 2019-12-23.

[43] Adira Quintero, Francisco Sans, and Eric Gamess. Performance evaluation of ipv4/ipv6 transition mechanisms. *International Journal of Computer Network and Information Security*, 8(2):1, 2016.