

非常時における通信確保のための BGP 接続手法の提案と評価

渡里 雅史^{†1} 屏 雄一郎^{†1}
阿野 茂浩^{†1} 山崎 克之^{†2}

インターネットは、非常時における通信確保のための重要な社会インフラとなった一方で、インフラを支える Border Gateway Protocol 4 (BGP) の仕様は、標準化から 10 年以上経った今も変更されていない。一般的な BGP を用いたインフラの安定化・高信頼化手法としてマルチホームがあるが、確保する代替経路の数に応じて運用コストが高くなるとともに、インターネット全体の経路数を増加させることが問題となる。本課題解決のため、本稿では、新たな高信頼化手法として、あらかじめ代替経路を確保せず、障害時にオンデマンドでドメイン内外の他の BGP ピアを自律的かつ動的に探索し、接続性を確保するオンデマンド型 BGP 接続手法を提案する。提案手法は、代表 BGP スピカ間で AS 番号を変換することで、ドメイン間の収容変更を可能とする。本稿では、提案手法の BGP 拡張に関わる詳細設計ならび既存ネットワークへの適用性について述べる。

The Evaluation and Proposition of BGP Peering Method for Emergency Communications

MASAFUMI WATARI,^{†1} YUICHIRO HEI,^{†1} SHIGEHIRO ANO^{†1}
and KATSUYUKI YAMAZAKI^{†2}

While the Internet has become an important infrastructure for providing a reliable and dependable connectivity, the Border Gateway Protocol 4 (BGP) specification remains unchanged since its standardization. A common approach for providing stability and fault tolerance in BGP is through multi-homing. However, the way multi-homing requires pre-configuration of alternative path raises concern on the operational cost and brings scalability issues on the size of the global routing table. As an alternative approach to provide fault tolerance, this paper presents a novel on-demand BGP peering method that allows ASes to automatically and dynamically restore from failures. The method allows restoration of ASes with different domains with limited changes to the current AS path modification rule. The detail extensions made to BGP and its applicability to the current BGP networks are discussed.

1. はじめに

近年、IP 電話に代表されるライフラインサービスの普及にともない、インターネットは、非常時における緊急通信確保のための重要な社会インフラとなった。一方、IP 網における障害多発を背景に、インターネットに対しては、ベストエフォート型の通信インフラから、安定性、信頼性のある高品質なサービスインフラへの変化を求められつつある。すでに総務省は、IP インフラを用いた重要通信確保のための技術基準の検討¹⁾を開始しており、今後は IP インフラに対するさらなる高品質化・高信頼化などの要求が想定される。

一方、インターネットのインフラを支える Border Gateway Protocol 4 (BGP)²⁾は、標準化から 10 年以上経った今も、その基本的な仕様は変更されていない。本プロトコル仕様に従った運用では、オペレータによる静的な設定を前提としており、収容変更の際は、手動で設定変更する必要があるなど、非常時の早期接続性回復には適していない。BGP を用いたインフラの安定化・高信頼化手法として、あらかじめ異なる Autonomous System (AS) を経由した代替経路を確保するマルチホーム³⁾がある。多くの BGP の運用では、本手法により、障害からの早期回復を実現している。しかしながら、マルチホームにより代替経路を確保する手法は、接続先ごとに設定・監視・管理する必要があるため、運用コストを増加させるとともに、集約されない形での経路広報により、インターネット全体の経路数を増加させることが問題となる。

^{†1} 株式会社 KDDI 研究所
KDDI R&D Laboratories Inc.

^{†2} 長岡技術科学大学
Nagaoka University of Technology

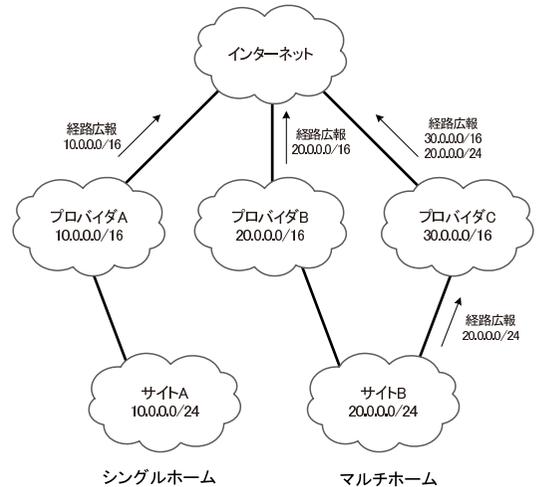
そこで、本稿では、特に非常時における BGP インフラの耐障害性向上のため、ネットワークの大規模障害回復を実現する BGP 接続手法について提案する。提案手法は、BGP の経路情報に基づいて隣接 AS との接続性を監視し、障害時において、オンデマンドでドメイン内外の他の BGP ピアを自律的かつ動的に探索し、接続性を回復する。本手法は、全 BGP スピーカを置き換える必要がなく、代表 BGP スピーカ間において AS 番号を変換することで、ドメインをまたがる収容変更を可能とする。本稿では、提案手法の BGP 拡張に関する詳細設計を述べ、既存ネットワークとの互換性および収容変更に関する実装評価を行う。

本稿の構成は以下のとおりである。2 章で関連研究について説明し、3 章で非常時におけるネットワーク収容形態について述べる。4 章で提案するオンデマンド型の BGP 接続手法について述べ、5 章で実装評価による提案方式の動作検証ならびに収容時間の評価結果を示す。6 章で有用性に関する考察を行い、7 章で結論を提示する。

2. 関連研究

BGP ネットワークの耐故障性の向上には、あらかじめ代替経路を確保するマルチホーム³⁾が広く用いられている。図 1 にシングルホームとマルチホームの概要を示す。シングルホームとは、サイト A の接続形態であり、単一のプロバイダと接続する構成を指す。一方、マルチホームとは、サイト B の接続形態であり、複数のプロバイダと接続する構成を指す。サイト A では、プロバイダ A がインターネットとの接続性を保つ唯一のプロバイダとなる。一方、サイト B では、プロバイダ B およびプロバイダ C を経由し、インターネットへの接続性を保つことが可能となる。このため、シングルホームに対して、マルチホームは、より冗長性の高い接続性の確保が可能となり、信頼性の向上が図れる。一般に、より多数のプロバイダと接続することで、より信頼性の高いネットワークが構築可能となる。

一方、マルチホームにより代替経路を確保する手法にはいくつかの問題がある。まず、マルチホームは、接続先の数に応じて運用面で高コストとなる。たとえば、実際の運用では、プロバイダごとに運用ポリシーが異なるため、そのポリシーに応じた制御を実現するためには、正しい知識と経験に基づいたネットワークの設計、接続先とのピアリングの交渉、実際の制御のためのフィルタの設定などが必要となる。また、ネットワークの信頼性を高めるには、BGP ピアごとの接続性監



シングルホーム マルチホーム

図 1 シングルホームとマルチホーム概要

Fig. 1 Overview of single-homed and multi-homed.

視やトラフィック量監視など、安定運用のための管理が必要となる。一般に、接続先ごとにこれらの設定・監視・管理が必要となるため、接続先の増加にともなって運用コストは増加する。また、マルチホームすることで、パンチングホールと呼ばれる、より細かい単位での経路広報が発生する。図 1 では、20.0.0.0/24 の経路広報がパンチングホールとなる。このため、マルチホームサイトの増加は、インターネット全体の経路数増加を加速させる⁴⁾。経路数の増加は、経路変動にともなうインターネット安定性の低下、経路収束の遅延⁵⁾、ハードウェア処理能力の圧迫⁶⁾など、サービスの品質を低下させかねない深刻な問題となる。

また、BGP の運用は、基本的にオペレータによる手動設定が中心であるため、ネットワークの信頼性はオペレータの監視管理体制に委ねられている。すなわち、マルチホームした複数のプロバイダが同時に障害を起こした場合は、オペレータによる復旧作業が必要となるため、非常時の通信確保手段として、マルチホームだけで十分とはいえない。たとえば、BGP 接続では、双方で共通の設定が必要であるため、管理ドメイン外の BGP ピアとの接続は、オペレータ間の連携対応が必要不可欠であり、接続性の再確保に多くの時間を要する。特にオペレータが遠隔から対象機器を制御できない場合は、現場対応が必要となり、復旧により多くの時間を要することが想定される。災害時では交通機関の混乱や道路の閉鎖などにより、現場へ到達が困難な状況も想定されるため、深刻な問題となる⁷⁾。

すなわち、ネットワークの信頼性向上の観点からは、マルチホームが望まれる一方、運用コストおよびパン

表 1 シングルホームとマルチホームの特徴
Table 1 Single-homed vs Multi-homed.

項目	シングルホーム	マルチホーム
耐故障性の向上	図れない	図れる
運用コスト	小さい	大きい
経路数への影響	一定	増加
非常時の通信確保	不十分	不十分

パンチングホールによる経路数増加への影響を考慮した場合には、シングルホームが望まれる。また、非常時の通信確保では、あらかじめ代替経路を確保するだけでは不十分であり、各ネットワークが自律的かつ動的に接続性を確保する手法が必要となる。特に大規模な災害では、必要に応じてプロバイダ間の連携による迅速な障害対応が望まれる。表 1 に上述のシングルホームとマルチホームの特徴をまとめる。

そこで、本稿では、ネットワークの耐故障性向上手法として、マルチホームのようなあらかじめ接続性を確保する手法ではなく、シングルホームを拡張し、障害時はオンデマンドで自律的かつ動的に代替経路を確保するオンデマンド型 BGP 接続手法を提案する。

3. 非常時におけるネットワーク収容形態

非常時には、利用可能な通信インフラを用いて、孤立したネットワークの接続性を迅速に復旧させることが望まれる。特に障害による重要サービスの停止は、事業の信頼性に関わる問題となるため、また、企業の社会的責任が問われるため、多くの企業が障害対策に積極的な設備投資を行いつつある。本稿では、非常時の通信確保を目的に、各サイトが、通常時は使用しない、非常時専用のバックアップ回線を保持することを前提とする。バックアップ回線は、同様に他のサイトが多数接続した、非常時用 IX のようなレイヤ 2 ネットワークとする。各サイトは、本ネットワークへの接続手段を用意することで、運用中の BGP のネットワーク構成を大幅に変更することなく、非常時の接続手段を確保することが可能となる。

このような非常時用回線を想定した場合、前述のシングルホーム拡張による、障害時の通信確保のためのネットワーク収容方法として、トンネルを用いて障害前と同じネットワークに再収容する手法と、経路広報により別ネットワークへ収容する手法が考えられる⁸⁾。図 2 に両手法の詳細を示す。

前者のトンネル手法は、ネットワークモビリティ技術⁹⁾に近い収容形態である。障害前のサイト A は、直接プロバイダ A に対して経路を広報することで接続性を確保する。その後、障害が発生した場合、サイト

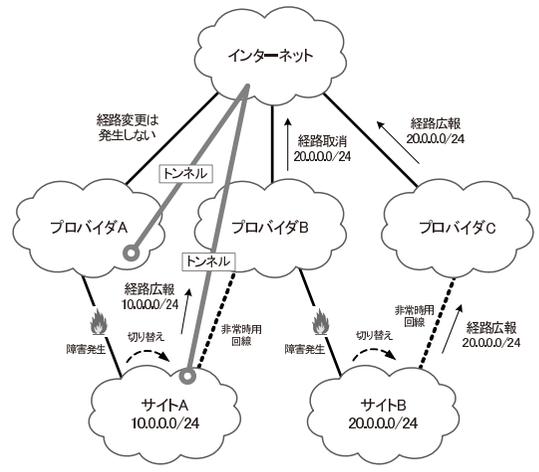


図 2 ネットワーク収容形態
Fig. 2 Configuration for re-homing networks.

A は、プロバイダ B 経由で、障害前の接続先プロバイダ A までトンネルを確立する。その後、確立されたトンネルを用いて経路交換を行い、接続性を再確保する。収容変更にとまなう経路変更はプロバイダ A 内で完結するため、上流に対する経路変動の伝播を抑制できる。一方、後者の経路広報手法は、機内インターネットサービスを提供した Connexion by Boeing¹⁰⁾に近い収容形態である。障害前のサイト B は、直接プロバイダ B に対して経路を広報することで接続性を確保する。その後、障害が発生した場合、サイト B は、新たな接続先であるプロバイダ C に対して直接経路を広報する。インターネットに対しては、プロバイダ C から本経路が伝播されるとともに、プロバイダ B からはサイト B に対する経路取消が広報される。結果として、経路変動が広域にわたって伝播されるとともに、トンネル手法に比べ、経路の収束に時間を要する。

一方、2章で述べたとおり、収容変更にとまなう経路変動は、インターネット全体に伝播されるため、結果として、インターネット全体の安定性を低下させる可能性があり、最小限にとどめることが望ましい。一方、非常時のような重要通信確保が最優先される場合には、孤立したネットワークが迅速に再収容されることが望まれる。前者のトンネル手法は、トンネルを終端するノードへの到達性が前提となるが、災害時にはプロバイダ全体の到達性が失われる可能性があるため、信頼性に欠ける。したがって、非常時におけるネットワーク収容には、マルチホームと同様、パンチングホールにより一時的な経路数の増加をとまなうが、重要通信

の確保という観点から後者の経路広報手法による収容変更手順が望ましいと考える。

また、ネットワークの収容変更の際には、経路制御プロトコルの大原則である、ループのない接続構成を実現する必要がある。BGP では、受信経路更新メッセージにおける AS パス属性を用いて AS 番号の重複を確認し、ループの検知を行う。図 2 において、各サイトがグローバル AS 番号を所持する場合では、収容変更後も AS 番号は重複しないため、BGP がループと誤認する可能性は低い。一方、一部プロバイダでは、各サイトに対してプライベート AS 番号を割り当てる AS コンフェデレーション¹¹⁾を用いる。プライベート AS 番号は、インターネット全体へ伝播されないため、プロバイダ内で自由に割当てが可能となる。このため、非常時における収容変更では、プロバイダ間におけるプライベート AS 番号の衝突により、BGP がループと誤認する課題がある。

4. 提案手法

以上の 2 章、3 章の課題を解決する手法として、障害時にオンデマンドでドメイン内外の他の BGP ピアを自律的かつ動的に探索し接続性を回復する手法を提案する。はじめに概要を述べた後、機能詳細を示す。

4.1 概要

より複雑なネットワーク構成に対応した収容変更を実現するため、以降では、各 AS が、AS コンフェデレーション¹¹⁾を用いた接続構成を対象とする。AS コンフェデレーション内の各 AS は、BGP の経路情報に基づき、隣接 AS との接続性を監視し、障害時において、AS コンフェデレーション内外の他の BGP ピアを自律的かつ動的に探索・接続し、接続性回復を実現する。本手法は、全 BGP スピーカを置き換える必要がなく、代表 BGP スピーカ間において AS 番号を変換することで、ドメインをまたがる収容変更を可能とする。

4.2 隣接 BGP ピアの探索

障害を検知した BGP スピーカは、接続性を再確保するため、非常時用回線を用いて、代替経路を探索する。探索手法として、あらかじめ各 BGP スピーカにおいて代替経路をすべて保持しておく手法が考えられるが、障害後の到達性が保証できないため、障害直後に再度探索する必要があり、効率的ではない。このため、障害検知に合わせて、BGP スピーカが自律的かつ動的に他の BGP スピーカを探索することで代替経路を確保することとした。一方、探索では、非常時回線に接続する他の BGP スピーカを 1 度に把握すると

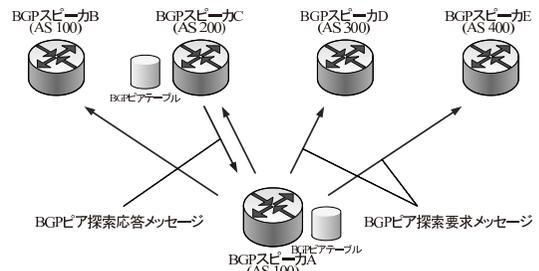


図 3 隣接 BGP ピア探索概要

Fig. 3 Neighbor BGP Discovery overview.

ともに、特別な設定なく探索のための通信を開始できることが望まれる。そこで、探索部分は、IPv6 を必須とし、プラグ&プレイ機能により自動生成されたリンクローカルアドレスと、Neighbor Discovery Protocol (NDP)¹²⁾を用いて、非常時回線に接続するすべての BGP スピーカをいっせいに把握可能とした。具体的には、交換するメッセージ数を減らすことを目的に、Neighbor Solicitation メッセージを必要としない Unsolicited Neighbor Advertisement メッセージを用いて、隣接ノードに対して、BGP の接続に必要なプライベート AS 番号、グローバル AS 番号、IP アドレスの通知を可能とした⁸⁾。また、新たに応答フラグを設けることで、Unsolicited Neighbor Advertisement メッセージに対して Unsolicited Neighbor Advertisement メッセージで応答させ、隣接ノードの接続情報を取得できる仕組みとした。

図 3 に隣接 BGP ピア探索概要を示す。障害を検知した BGP スピーカ A は、全ノードマルチキャストアドレスに対して、BGP ピア探索要求メッセージを送信する。本メッセージを受信した BGP スピーカ C は、BGP ピア探索応答メッセージを用いて接続の可否を応答する。接続を許可する場合は、受信した BGP ピア探索要求メッセージのピア情報を新たに保持する BGP ピアテーブルで管理するとともに、自身の BGP 接続情報を応答する。

4.3 最適 BGP ピアの選択

BGP は、TCP 接続の確立により、Update メッセージを用いた経路情報の交換を行う。経路情報の交換により、インターネットへの接続性を確保することが可能となる。複数の BGP ピアを発見した場合は、全 BGP ピアと経路を交換することで、複数の接続性を同時に確保することが可能となる。しかしながら、前述のとおり、経路変動の伝播は、インターネット全体の安定性低下につながるという問題がある。そこで、提案手法では、最適な BGP ピアのみを選択し、接続性を再

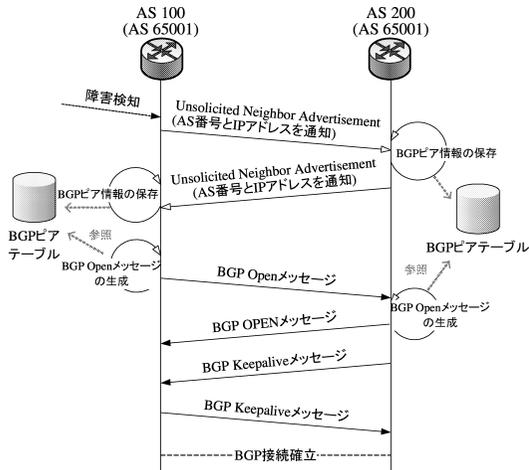


図 4 提案手法のメッセージ交換手順
Fig. 4 Message exchange sequence.

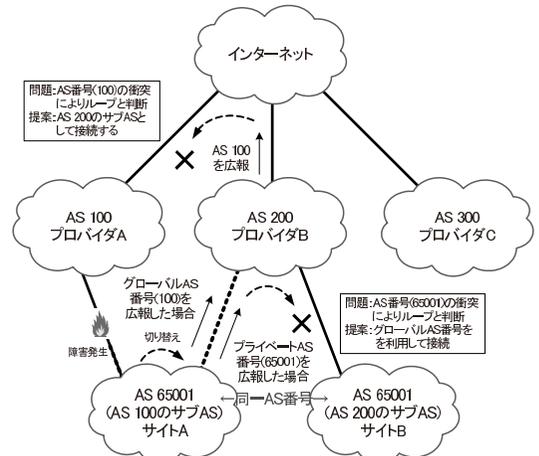


図 5 ループ誤認防止のための BGP 接続手法
Fig. 5 BGP peering method for loop prevention.

確保する手法とした。最適な BGP ピアとは、上流の任意 AS までの最短 AS パスを所有する BGP ピアとし、複数の BGP ピアを発見した場合は、各 BGP ピアから取得した経路情報を解析したうえで経路交換を行い、他の BGP ピアとの接続は遮断する仕組みとした。

4.4 オンデマンド BGP 接続のための拡張

通常、BGP では、あらかじめピアリングが設定されていない BGP ピアからの接続要求は無条件で遮断する。そのため、接続に際しては、あらかじめ BGP ピア間において、ピアリングに必要な設定を完了しておく必要がある。しかしながら、障害に連動してオンデマンドで接続を試みる場合は、接続要求を受けてはじめて BGP ピアを把握するため、事前に設定しておくことが困難となる。そこで、提案手法では、BGP の接続要求に合わせて、BGP ピアテーブルを参照し、必要なピアリングの設定を動的に注入することで接続を可能とした。BGP ピアテーブルには、BGP ピア探索により、非常時に接続性を必要とする BGP ピアのみ含まれるため、本テーブルを参照し、要求元のエントリの有無を確認することで、必要な BGP ピアに対してのみ接続を許可することを可能とした。

障害発生から BGP 接続確立までのメッセージ交換の具体的な手順を図 4 に示す。BGP スピーカ A は、隣接 BGP ピア探索により発見した BGP スピーカ C に対して、BGP 接続要求を行う。本要求を受けた BGP スピーカ C は、BGP ピアテーブルを参照し、該当するピア情報が存在する場合のみピアリングの設定を注入する。これにより、BGP Open メッセージおよび Keepalive メッセージの交換が可能となり、接続が確立される。同様の手順を追うことで、必要に応じて、

同時に複数の BGP スピーカと接続することが可能となる。

4.5 ループ誤認防止のためのネットワーク収容手法

3章で述べたとおり、AS コンフェデレーションを用いた接続構成では、各 AS に対してプライベート AS 番号を割り当てるため、収容変更において、プロバイダ間でプライベート AS 番号が衝突し、BGP がループと誤認する可能性がある。図 5 に詳細を示す。

図 5 のサイト A は、プロバイダ A のプライベート AS であり、サイト B は、プロバイダ B のプライベート AS とする。また、障害はサイト A とプロバイダ A 区間を想定し、サイト A のプロバイダ B への収容変更を想定する。このとき、サイト A が自身の AS 番号として、プライベート AS 番号 (65001) を用いた場合には、サイト B のプライベート AS 番号 (65001) と衝突し、BGP はループと誤認する。一方、サイト A が自身の AS 番号として、グローバル AS 番号 (100) を用いた場合には、サイト B とのループ誤認は回避できるものの、プロバイダ A のグローバル AS 番号 (100) と衝突し、ループと誤認する。

そこで、提案手法では、サイト A は、グローバル AS 番号をプライベート AS 番号として広報するとともに、プロバイダ B の AS コンフェデレーションに加わることにする。これにより、プロバイダ B の AS コンフェデレーション内での衝突が回避できるとともに、上流のプロバイダ A に対する経路広報は、プロバイダ B においてグローバル AS 番号 (200) に変換されるため、ループ誤認を回避した接続が可能となる。

5. 評価

提案する BGP 接続手法の評価のため、テストベッドを用いた動作検証を行うとともに、障害発生から回復までの時間を測定した。以下、実験構成について説明し、評価結果を示す。

5.1 実験構成

実験ネットワークは、標準 BGP ルータとの相互接続性を検証するため、商用ルータと開発ルータで構成した。開発ルータは、汎用 PC と UNIX OS で動作する Quagga Routing Software Suite¹³⁾ を拡張した。図 6 に実験ネットワーク構成を示す。

図 6 の AS100, AS200, AS400 の各 BGP スピーカは、AS コンフェデレーション¹¹⁾ を有効とした接続構成とし、それぞれにプライベート AS 番号を割り当てた。また、AS パス長の調整のため、R1-R2 区間には AS パスプリベンド機能により仮想 AS301 を設けた。

5.2 提案 BGP 接続手法の動作検証

提案する BGP 接続手法が、障害時の接続性回復手法として有効であるかを検証するため、実験ネットワークを用いて動作検証を行った。図 7 に、R1-R3 区間においてリンク障害を発生させた場合の、R3 の動作ログ内の時間に基づく、8 回の試験時における個々のイベント(1-7)の発生時間を示す。各線は、1 回の測定結果を示す。R3 では、リンク障害の発生にともない、BGP ホールドタイマがタイムアウトし(1)、隣接 AS への接続性を失うことで障害を検知する(2)。その後 R3 では、接続性確保のため、非常時回線を用いて隣接 BGP ピアを探索する。探索の結果、R2 および R5 を発見し(3)、それぞれに対して BGP 接続を要求する(4, 5)。接続が正常に確立すると、各 BGP ピアから経路情報を取得し、解析により、上流 AS まで最短 AS パスを有する BGP ピアを選択する(6)。その後、選択した BGP ピアに対して、経路広報のための BGP Update メッセージを送信する(7)。

また、図 8 に隣接 BGP ピア探索と、上流 AS を AS300 とした場合の経路解析時の動作ログを示す。実装では、AS コンフェデレーションのパス数を 1 としたため、経路解析の結果、R5 (AS100) を経由した AS パス [(65002 65001) 300] のパス数が 2、R2 (AS400) を経由した AS パス [(65002) 301 300] のパス数が 3 となり、AS パス数の少ない R5 を選択した。

図 9 に R3 における収容変更前後の経路表を示す。収容変更前の経路表では、対外経路はすべて AS300 を経由しているのに対して、収容変更後では、AS100

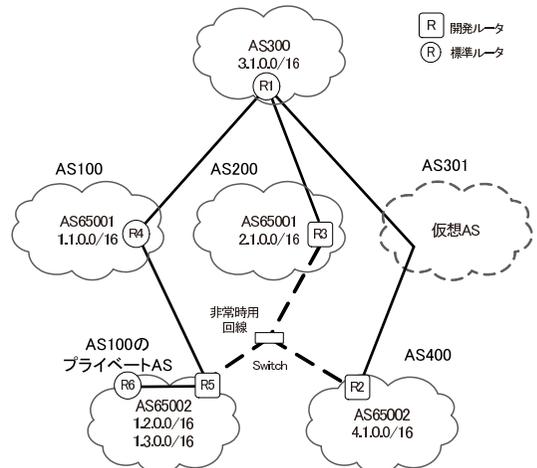


図 6 実験ネットワーク構成
Fig. 6 Network topology for experiment.

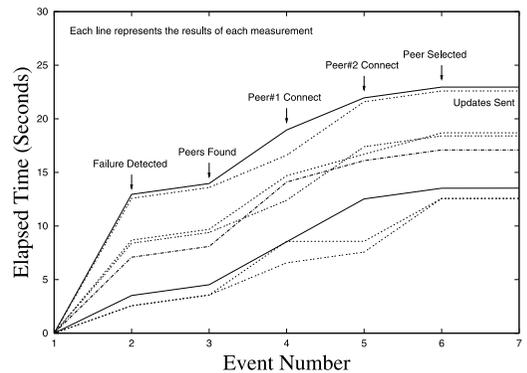


図 7 各種イベント発生時の経過時間
Fig. 7 Elapsed time of each event.

```
02:16:41.641 ND6 found Peer: AS 65001, ConfID 400, IF 2.2.2.3
02:16:41.641 ND6 found Peer: AS 65002, ConfID 100, IF 2.2.2.1
..... <Receive BGP Updates> .....
02:16:42.647 Peer 100, 3.1.0.0/16, plen 2, path (65002 65001) 300
02:16:42.647 Peer 400, 3.1.0.0/16, plen 3, path (65002) 301 300
02:16:42.647 Selected Peer: 100, AS 65002, plen 2
```

図 8 隣接 BGP ピア探索および経路解析の動作ログ
Fig. 8 Neighbor BGP peer discovery and route analysis.

を経由するとともに、プライベート AS 番号をグローバル AS 番号に変換し、AS コンフェデレーション内におけるプライベート AS 番号の衝突を回避した。

続いて、障害発生箇所での接続性が回復した場合の動作を検証するため、リンク障害を発生させた R1-R3 区間を復旧させた。その結果、R3 では、BGP の再接続処理により、R1 と経路の交換を開始し、その後、隣接 AS への到達性を確認することで、障害回復を検知する。その後、R3 では、非常時回線の接続性を遮断するため、R5 に対して拡張 BGP Notification メッ

切替前の経路表						
R3# show ip bgp						
Network	Next Hop	Med	Lpf	Weight	Path	
*> 1.1.0.0/16	5.5.5.2			0	300 100 i	
*> 1.2.0.0/16	5.5.5.2			0	300 100 i	
*> 1.3.0.0/16	5.5.5.2			0	300 100 i	
*> 2.1.0.0/16	0.0.0.0	0		32768	i	
*> 3.1.0.0/16	5.5.5.2	0		0	300 i	
*> 4.1.0.0/16	5.5.5.2			0	300 301 400 i	

切替後の経路表						
R3# show ip bgp						
Network	Next Hop	Med	Lpf	Weight	Path	
*> 1.1.0.0/16	2.2.2.1	0		0	(100) i	65002に相当
*> 1.2.0.0/16	2.2.2.1			0	(100) i	65001に相当
*> 1.3.0.0/16	2.2.2.1			0	(100 100) i	
*> 2.1.0.0/16	0.0.0.0	0		32768	i	
*> 3.1.0.0/16	2.2.2.1			0	(100 100) 300 i	
*> 4.1.0.0/16	2.2.2.1			0	(100 100) 300 301 400 i	

図 9 R3 における切替え前後の経路表

Fig. 9 Routing table at R3.

メッセージを送信し、BGP の再接続処理を停止した。

同様の手順で R1-R2 区間、R1-R4 区間、R4-R5 区間において、リンク障害とリンク回復を発生させた結果、提案手法がすべての区間において有効であることを確認した。また、提案手法が一部の BGP スピーカのみを置き換えることで、標準の BGP 商用ルータとの互換性を保ったドメインをまたがる収容変更が可能であることを確認した。

5.3 収容時間短縮に関する考察と実験結果

非常時における通信確保では、ダウンタイムを最小限にとどめることが要求される。提案手法では、図 7 の結果に示すとおり、接続性の再確保までに多くの時間を要している。具体的には、障害検知から経路広報までに、以下 5 つのパラメータが関係する。

- BGP ホールドタイム
BGP 接続の障害を検知。接続性は Keepalive メッセージにより監視：3 秒 (Quagga 初期値は 180 秒, RFC 推奨は 90 秒)。
- 接続性監視間隔 (本提案の追加機能)
隣接 AS への到達性を監視する間隔：15 秒 (動作検証が目的のため、安全側の値とした)。
- BGP ピア探索待ち時間 (本提案の追加機能)
隣接 BGP ピア探索後、応答を待つ時間：1 秒。
- BGP 接続開始タイム
BGP 接続 (TCP) を開始するまでの待ち時間：Quagga は 3~8 秒でランダムに設定。
- 経路解析開始時間 (本提案の追加機能)
経路情報の解析を開始する時間：9 秒 (BGP 接続開始タイム (d) が最大 8 秒のため、実装上、+ 1 秒とした)。

したがって、これらのパラメータを調整することで、提案方式を用いた場合の収容時間の短縮が図れる。そこで、より迅速な障害回復を目指すため、これらのパ

表 2 収容時間短縮のためのパラメータ (秒)

Table 2 Modified Parameters for Fast Restoration.

項目	内容	変更前	変更後
a	BGP ホールドタイム	$t = 3$	$t = 2$
b	接続性監視間隔	$0 < t \leq 15$	$0 < t \leq 1$
c	BGP ピア探索待ち時間	$t = 1$	$t = 1$
d	BGP 接続開始タイム	$3 \leq t \leq 8$	$t = 0$
e	経路解析開始時間	$t = 8 + 1$	$t = 1$

ラメータの調整を行った。まず、BGP ホールドタイム (a) は、Keepalive メッセージの送信間隔に依存するため、最短の障害検知には、本メッセージの送信間隔を最小にする手法が考えられる。しかしながら、Keepalive メッセージは、BGP ホールドタイムの 1/3 間隔で BGP ピアが送信する仕組みであり、さらに、本メッセージの最小送信間隔は 1 秒と規定されている。そこで、最小送信間隔 1 秒で Keepalive メッセージを送信させるため、BGP ピアに通知するホールドタイムを 3 秒とする一方で、開発ルータ自身に設定するホールドタイムを、実装上、2 秒とした。また、接続性監視間隔 (b) および BGP ピア探索待ち時間 (c) は、それぞれ最短とするため 1 秒とした。また、Quagga の仕様では、多数の BGP ピアに対する Open メッセージの同時送信を避けるため、各 BGP スピーカは、接続開始に 3~8 秒のランダムな遅延を挿入している。これは、同時に複数の接続が開始された場合の、ルータに対する負荷を考慮したタイムであるが、非常時用回線では、限られた BGP ピアとのみ接続するため、また、早期回復が望まれるため、BGP 接続開始タイム (d) は 0 秒とした。これにより、経路解析開始時間 (e) を 1 秒とした。

表 2 に収容時間に関わるパラメータをまとめる。BGP 標準機能 (a, d) は、最大で 9 秒 (変更前は最大 11 秒, 変更後は 2 秒) の時間短縮が可能となる。また、これにより、(e) の時間短縮が可能となり、本提案に関わる追加機能 (b, c, e) は、 $10 < t \leq 25$ から $2 < t \leq 3$ まで短縮可能となる。そこで、本パラメータを用いた動作検証を行うため、図 7 と同等の試験を行った。その結果、図 10 に示すとおり、本考案のパラメータは、収容時間の短縮に有効であることを確認した。

ただし、上記収容時間には、障害発生から BGP ホールドタイムのタイムアウトまでの時間、また、経路情報の伝播に掛かる時間は含まれない。そこで、障害発生から接続性確保までの総時間を検証比較するため、ping によるダウンタイムを測定した。表 3 に R3-R6 区間における、ping の 10 ミリ秒間隔の測定結果を示

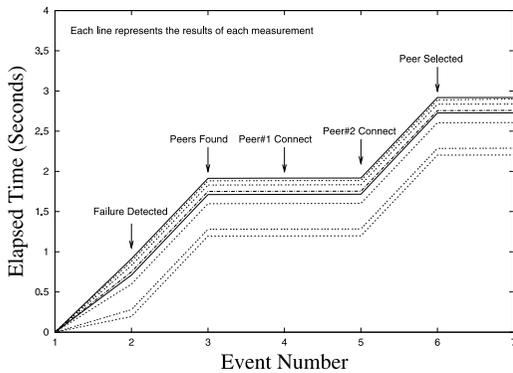


図 10 各種イベント発生時の経過時間 (パラメータ変更後)

Fig. 10 Elapsed time of each event (Modified Parameters).

表 3 障害時のダウンタイムの比較

Table 3 Comparison of Downtime after a Failure.

収容モード	ダウンタイム		
	最小	最大	平均
変更前	14.87 秒	26.99 秒	21.24 秒
変更後	3.73 秒	8.08 秒	5.88 秒

す。変更後のダウンタイムは、変更前に比べ、計 10 回の平均で 15.36 秒短縮可能であった。また、障害前の接続構成への切替え時のダウンタイムは、あらかじめ非常時用回線の優先度を下げしておくことで、0 秒で切替え可能であり、提案手法が、非常時の接続性確保として有効であることを確認した。

6. 考 察

提案手法は、災害時などプロバイダ間を結ぶ地上回線が断線した場合において、衛星回線¹⁴⁾などの利用による迅速な通信確保手法としても有効である。提案手法は、代表 BGP スピーカ間で AS 番号を交換し、ドメインをまたがる収容変更を実現するため、大規模な障害時は、衛星回線を経由したプロバイダ間の収容変更による通信確保が可能となる。

今後の検討課題としては、実インターネットを用いた場合の実用性検証がある。接続構成によっては、より多くの経路数を扱うため、経路取得および解析により多くの時間が必要となる。また、複数の隣接 BGP ピアから同時に多数の経路を受信するには、処理能力の高いハードウェアが必要となる。特に、経路数およびピア数の増加は、表 2 の BGP ホールドタイム、BGP 接続開始タイム、経路解析時間に大きく影響する可能性がある。これら課題解決には、交換する経路情報を制御し、解析に必要な経路のみ受信するなどの対策が必要である。また、本手法では、プロバイダに接続す

る複数のサイトが孤立した場合は、サイト間における通信確保のため、それぞれのサイトを異なるプロバイダで収容する必要がある。そのためには、4.2 節の探索プロセスにおいて、プロバイダ側で収容の可否を応答する制御が必要となる。

7. ま と め

本稿では、非常時における通信確保のため、BGP を用いたインフラの耐障害性を向上させる BGP 接続手法について提案した。従来のマルチホームに代表されるネットワーク耐障害性技術は、確保する代替経路の数に応じて運用コストが高くなるとともに、パンチングホールによりインターネット全体の経路数を増加させることが問題であった。また、BGP の運用は、オペレータによる手動設定が中心であるため、ネットワークの信頼性はオペレータの監視管理体制に委ねられており、接続性の再確保に多くの時間を要することが問題であった。提案手法は、これら問題の解決のため、BGP の経路情報に基づき、隣接 AS との接続性を監視し、障害時において、オンデマンドで自律的かつ動的に接続性回復を実現した。提案手法を既存の BGP 環境上で実装した結果、30 秒内の収容変更を得た。実験・評価により標準の BGP との互換性を保ちつつ、パラメータの最適化を検討実施した結果、最短で 3 秒台の収容変更が可能であることを確認した。本手法は、シングルホームを拡張するため、通常時はインターネット全体の経路数を増加させることなく、運用コストを抑えた障害対策が可能となる。

参 考 文 献

- 1) 総務省：大規模地震などの災害時における重要通信確保に向けて—「電気通信事業における重要通信確保の在り方に関する研究会」報告書の公表および宮城県沖の地震を踏まえた今後の取組，報道資料 (July 2003).
- 2) Rekhter, Y., Li, T. and Hares, S.: A Border Gateway Protocol 4 (BGP-4), RFC 4271 (Jan. 2006).
- 3) Abley, J., Lindqvist, K., Davies, E. and Black, B.: IPv4 Multihoming Practices and Limitations, RFC 4116 (July 2005).
- 4) Bu, T., Gao, L. and Towsley, D.: On characterizing BGP routing table growth, *The International Journal of Computer and Telecommunications Networking*, Vol.45, Issue 1 (May 2004).
- 5) Labovitz, C., Ahuja, A., Bose, A. and Jahanian, F.: Delayed Internet Routing Convergence, *IEEE/ACM Trans. Networking*,

Vol.9, No.3 (June 2001).

- 6) Meyers, D., Zhang, L. and Falls, K.: Report from the IAB Workshop on Routing and Addressing, Internet Draft (work in progress) (Apr. 2007).
- 7) 中沢淳一, 高橋謙三: 情報通信ネットワークの災害対策, 電子情報通信学会誌, Vol.89, No.9, pp.782-276 (2006).
- 8) 渡里雅史, 屏雄一郎, 長谷川輝之, 阿野茂浩, 山崎克之: 高信頼ドメイン間経路制御に向けた BGP 接続手法の提案, 情報処理学会研究報告, QAI, Vol.2007, No.3, pp.43-48 (2007).
- 9) Devarapalli, V., Wakikawa, R., Petrescu, A. and Thubert, P.: Network Mobility Basic Support, RFC 3963 (Jan. 2005).
- 10) Dul, A.: Global IP Network Mobility using Border Gateway Protocol (BGP), Technical Report, The Boeing Company (Mar. 2006).
- 11) Traina, P., McPherson, D. and Scudder, J.: Autonomous System Confederations for BGP, Internet Draft (work in progress) (Feb. 2007).
- 12) Narten, T., Nordmark, E. and Simpson, W.: Neighbor Discovery for IPv6, RFC 2461 (Dec. 1998).
- 13) Quagga Routing Software Suite.
<http://www.quagga.net/>
- 14) インターネットにおける WINDS 利用検討委員会 . <http://i-space.jaxa.jp/i-winds/index.html>
(平成 19 年 6 月 11 日受付)
(平成 19 年 12 月 4 日採録)



渡里 雅史

平成 15 年慶應義塾大学環境情報学部環境情報学科卒業 . 平成 17 年同大学大学院修士課程修了 . 同年 KDDI (株) 入社 . 以来, 研究所にて, IP ネットワーク制御の研究に従事 . 現

在, (株) KDDI 研究所 IP 品質制御システムグループ研究員 .



屏 雄一郎 (正会員)

平成 8 年東京大学工学部電子情報工学科卒業 . 平成 10 年同大学大学院修士課程修了 . 同年 KDD (株) 入社 . 以来, 研究所にて, IP ネットワークの構築・運用管理の研究に従事 . 現在, (株) KDDI 研究所 IP 品質制御システムグループ研究主査 . 平成 15 年度電子情報通信学会学術奨励賞受賞 .



阿野 茂浩 (正会員)

昭和 62 年早稲田大学理工学部電子通信工学科卒業 . 平成元年同大学大学院修士課程修了 . 同年 KDD (株) 入社 . 以来, 研究所にて, ATM 交換方式, IP ネットワーク管理・制御, 次世代インターネットの研究に従事 . 現在, (株) KDDI 研究所 IP 品質制御システムグループリーダー . 平成 7 年度本会学術奨励賞受賞 .



山崎 克之 (正会員)

昭和 55 年電気通信大学通信学部卒業 . 工学博士 . KDD (現 KDDI) (株) において ISDN, SDH, ATM, L2, IP の情報通信ネットワークとマルチメディア通信の研究開発・実用化, 国際標準化, に従事 . (株) KDDI 研究所・研究戦略室長を経て, 平成 18 年から長岡技術科学大学教授 .