

複数 AS 連携によるドメイン間通信の耐障害性向上

屏 雄一郎[†] 中尾 彰宏[‡] 大岸 智彦[†] 長谷川 亨[†] 山本 周^{*}

KDDI 研究所[†] 東京大学[‡] NICT^{*}

1. はじめに

インターネットで広く使われている BGP (Border Gateway Protocol) [1]は、障害時の経路収束に時間がかかるため、リンク障害からの復旧に時間を要する。この問題は、通信相手先との間で経由する組織 (AS: Autonomous System) が多いほど深刻化する。これに対して、筆者らは特定の AS 間でこの問題を回避するため、複数の AS が連携して、それらの AS 間で重なるの少ない複数の経路を構築することにより、連携 AS 間の通信の可用性や耐障害性を向上させる手法を提案している [2]。本稿では、提案手法とその経路制御設計について述べる。また提案手法の実用例として、地理的に分散したデータセンター間での通信を想定したシナリオについて述べる。

2. 提案手法

提案手法では複数の AS が連携し、連携 AS 間で重なるの少ない複数の経路を構築する。重なるの少ない経路とする理由は、一つの障害により複数経路の同時消失を防ぐためである。

図 1 に提案手法における経路の種類を示す。図 1 は、AS A, B, C が連携する場合での、AS A から AS B への経路の種類を示している。なお各連携 AS は、自 AS や自身の顧客 AS 以外からのトラフィックは他の AS に転送しないとする。

通常経路は BGP で取得できる経路である。自 AS 用オーバーレイ経路は、他の連携 AS を経由して宛先 AS に到達する経路である。AS A は AS

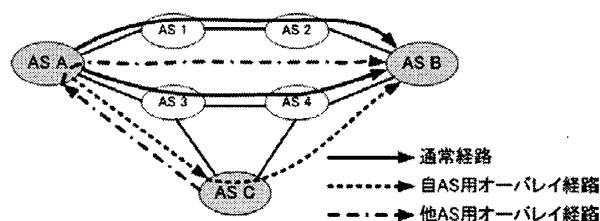


図 1: 提案手法における経路の種類 (AS A から B への経路)

B への通常経路が全て消失したとしても、通常経路と重ならない自 AS 用オーバーレイ経路経由で、B との通信を継続することができる。

他 AS 用オーバーレイ経路は、他の連携 AS からのトラフィックを宛先 AS に転送する経路である。なお AS A から見た AS B への他 AS 用オーバーレイ経路は、AS C から見ると AS B への自 AS 用オーバーレイ経路の一部となる。他 AS 用オーバーレイ経路を定義した理由は以下の通りである。図 1 で AS C が自 AS 用オーバーレイ経路を利用して AS B にパケットを送信する場合、そのパケットは AS A を経由させる必要がある。これは AS A へのトンネリングにより実現できる。しかし AS A の上位 AS が Ingress filtering を行う場合、AS C からのパケットを AS A 発と見せるために、そのパケットを AS A から AS B へのトンネルに通す必要がある。そこで連携 AS 間のパケット転送において通常経路と区別するため、上述のトンネル経路を他 AS 用オーバーレイ経路と定義する。

また提案手法では、連携 AS を経由するオーバーレイ経路を構築するため、連携 AS 間で multihop BGP により経路交換する。これにより例えば図 1 で、AS A は AS C から AS B への AS パス等を取得することができる。

3. AS 内構成と経路制御設計

連携 AS 間で複数経路を構築するため、各連携 AS は AGF (Alliance Gateway Function) と APCF (Alliance Path Computation Function) を持つ。AGF は外部との経路制御を担う機能、APCF は連携 AS 間の経路を算出する機能を持つ。AS 内では AGF と APCF が経路交換し、連携 AS 間では APCF が経路交換することで複数経路を構築する。図 2 に提案手法における AS 内機能構成と経路交換を示す。

AGF は隣接 AS から BGP で経路を受信し、APCF には他の連携 AS を起源とする経路を広告する (図 2(1), (2))。また AGF は通常の BGP ルータとしての役割も持つ。すなわち隣接 AS から受信した経路を自 AS 内の他の BGP スピーカ (ルートルフレクタ等) に広告する (図 2(2'))。

APCF は自 AS の AGF から受信した経路を、他の連携 AS の APCF と BGP で交換する (図 2(3))。この経路交換により、APCF は任意の連

Empowering Interdomain Communication with Resilience through AS alliance

[†] Yuichiro Hei, Tomohiko Ogishi, Toru Hasegawa, KDDI R&D Laboratories, Inc.

[‡] Akihiro Nakao, The University of Tokyo

^{*} Shu Yamamoto, NICT

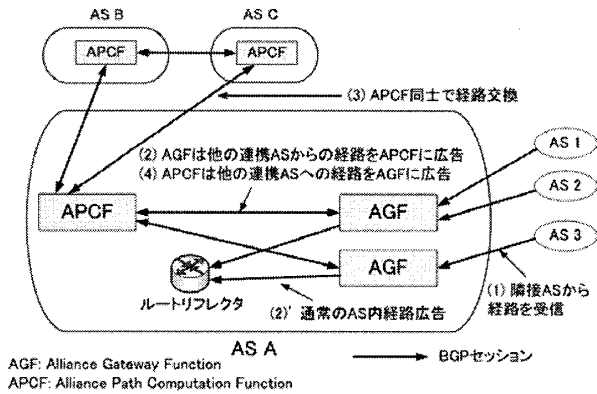


図 2: 連携 AS 内機能構成と経路交換

携 AS 間の経路情報 (AS パス) を取得することができる。すなわち APCF において、図 1 に示すような連携 AS 間の AS パストポロジーを認識することができる。その後 APCF は連携 AS 間で、重なるの少ない複数経路を算出し[2]、その結果を AGF に広告する (図 2(4))。

AGF は APCF から連携 AS への経路を受信すると、連携 AS 間通信用の経路表を作成する。なお先に述べた通り、AGF は通常の BGP ルータとしての役割も持つ。そのため AGF は、通常の経路表と連携 AS 間通信用の経路表を保持する。AGF はパケット転送の際にパケットの送信元アドレスを確認し、そのアドレスに応じて参照する経路表を決定する。AGF においてそのルールを設定することで、特定のアドレスからのパケットのみ、オーバーレイ経路を含む複数の経路を利用したパケット転送を行う等の経路制御を適用することができる。

4. 実用例

提案手法の実用例として、地理的に分散したデータセンター (DC) 間における通信、例えばデータ複製やデータベース同期等の堅牢化を想定したシナリオについて考察する。

近年、Google Public DNS[3]など、地理分散した DC を利用して、ユーザが世界中どこからアクセスしても品質を変えずに提供可能なネットワークサービスが増加しつつある。そのようなサービスを提供する方法の一例として、世界各地の DC に同一のデータを配置する方法が考えられるが、その場合 DC 間での迅速かつ確実なデータ複製や同期が必要となる。提案手法を地理分散した DC 間に適用すれば、インターネット経由の DC 間通信を、複数経路を同時に利用することで迅速に行うことができる。またリンク障害により、DC 間の通常経路が全て消

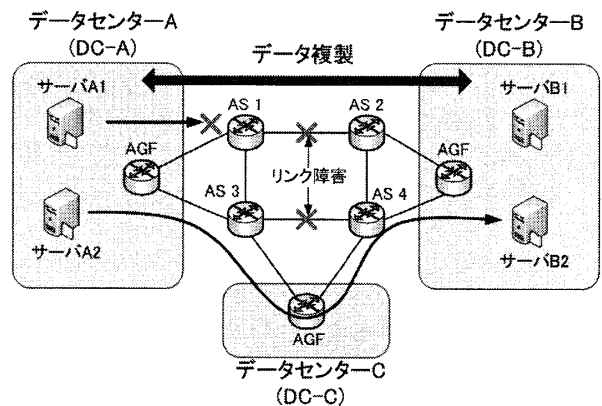


図 3: シナリオ構成

失したとしても、連携する他の DC 経由で通信を継続することが可能となるため、DC 間通信の耐障害性を高めることができる。

図 3 にシナリオの構成例を示す。DC-A, B, C に提案手法が導入されており、各 DC は複数の上流 AS と BGP 接続しているとする。ここで DC-A, B 間でデータ複製を行うとする。また DC-A の AGF において、サーバ A1 からのパケットは通常の経路表を参照し、サーバ A2 からのパケットは連携 AS 間通信用の経路表を参照するように設定する。すなわちサーバ A2 の通信のみ提案手法が適用されるとする。図 3 に示すように、AS1-AS2 間と AS3-AS4 間でリンク障害が発生した場合、DC-A, B 間の通常経路が全て消失するため、サーバ A1 からの通信は中断するが、サーバ A2 からの通信は DC-C 経由で継続できる。このように提案手法では、特定のサーバ間でのデータ複製を堅牢化させるような利用形態も可能である。

5. おわりに

本稿では、複数 AS が連携してその AS 間で重なるの少ない複数の経路を構築することにより、連携 AS 間通信の可用性や耐障害性を向上させる手法と経路制御設計について述べた。今後は実ネットワークでの検証評価を行う。

参考文献

- [1] Y.Rekhter, et al, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, Jan. 2006.
- [2] Y.Hei, et al, "AS alliance: Cooperatively improving resilience of intra-alliance communication", In Proc. of ROADS Workshop, Dec. 2008.
- [3] Google Public DNS, <http://code.google.com/intl/ja/speed/public-dns/index.html>