社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

オーバレイネットワーク技術の非常時通信への適用に関する一検討

長谷川　剛†　　亀井　　聡††　　村田　正幸†††

† 大阪大学 サイバーメディアセンター　〒 560–0043 大阪府豊中市待兼山町 1–32
†† NTT サービスインテグレーション基盤研究所　〒 180–8585 東京都武蔵野市緑町 3-9-11
††† 大阪大学 大学院情報科学研究科　〒 565-0871 大阪府吹田市山田丘 1–5
E-mail: †{hasegawa,murata}@ist.osaka-u.ac.jp, ††kamei.satoshi@lab.ntt.co.jp

あらまし　本稿では、オーバレイネットワークを用いたトラヒックルーティング技術を用いることで、大規模災害等によって IP ネットワークに大きな障害が発生した際に、従来の BGP による AS 間ルーティングでは到達不可能となる AS 間通信の大部分を短時間で復旧することが可能となる、オーバレイルーティング手法を提案する。具体的には、オーバレイノードの設置場所、情報交換手法、AS の参加・離脱手法等の検討を行い、小さい通信オーバヘッドでより多くの AS が参加可能となるオーバレイネットワーク構築手法を提案する。提案手法の効果はインターネットにおける AS 間トポロジーを利用した性能評価を行うことにより確認する。
キーワード　非常時通信、オーバレイネットワーク、ルーティング、BGP

A study on emergency networking services based on overlay network technologies

Go HASEGAWA†, Satoshi KAMEI††, and Masayuki MURATA†††

† Cybermedia Center, Osaka University　1–32, Machikaneyama-cho, Toyonaka, Osaka, 560–0043 Japan
†† NTT Service Integration Laboratories, NTT Corporation　3-9-11, Midori-cho, Musashino-shi, Tokyo, 180–8585 Japan
††† Graduate School of Information Science and Technology, Osaka University　1–5, Yamadaoka, Suita, Osaka, 565-0871 Japan
E-mail: †{hasegawa,murata}@ist.osaka-u.ac.jp, ††kamei.satoshi@lab.ntt.co.jp

Abstract　In this report, we propose an architecture for emergency networking services, which is based on overlay routing technologies. We propose the improvement of existing overlay routing algorithm in terms of the amount of overhead in measuring reachability and exchanging informations between overlay nodes, and overlay-routing mechanism especially against the large-scale network failure caused by disasters, terrorism, routing software bugs, and so on. Through numerical examples with the actual AS-level network topology of the current Internet, we show that our approach can reduce the overhead in exchanging routing information up to 1/1000, and improve the network connectivity by up to 9 times.
Key words　Emergency networking services, Overlay networks, Routing, BGP

1. ま え が き

情報ネットワークにリンク・ノード障害が発生し、単一機器や複数の機器を含むエリアが通信不可能になる場合に対応するためにこれまでに考えられているのは、機器やリンク、およびそれらを制御するソフトウェアを冗長化し、障害発生時に冗長系へ切り替えるような手法である。これらの手法において重要となるのは制御コストと性能のトレードオフであり、既存研究の多くはこの点に着目している。そのため、大規模災害やテロ、大規模停電などによって引き起こされる大規模なネットワーク障害に対しては、発生確率が小さいにもかかわらずコストが大幅に増大するため、対応が極めて困難となる [1]。また、これまでのネットワークの高信頼化を目指した研究のほとんどは障害発生モデルとして単一障害を想定している。一方、大規模災害、テロ、ルータソフトウェア (OS) の不具合などによって発生すると考えられる、複数の構成要素が同時に故障するような大規模かつ面的なネットワーク障害に関する研究はほとんど行われていない。

また、IP ネットワークに対する同種の研究も少ない。この理

由として、IP そのものが軽度の障害発生に対しては代替経路の発見が比較的短時間に行われることが挙げられる。しかし、インターネットの AS 間経路制御を行っている Border Gateway Protocol (BGP) は、障害が大規模である場合や、ある特定のトポロジ環境においては、障害発生時のネットワーク接続性が低下し、代替経路発見および経路の収束に非常に長い時間 (数分〜数時間) を必要とすることが指摘されている [2, 3]。そもそも BGP には、経路収束にかかる時間の理論的上限は存在しない。そのため、BGP の経路収束時間を改善するための様々な手法が提案されている (例えば [4, 5]) が、そのほとんどは BGP や TCP/IP そのものの改変を必要とするため、導入には標準化作業が必要となり、現在のインターネットへの適用は困難であると考えられる。また、AS 間リンクには、トランジットリンクやピアリングリンクなどのコスト構造が異なるリンクが存在し、各 ISP はそれらの経済的コストや政治的思惑を考慮してトラヒックの経路制御を行っている [6, 7]。そのため、結果として得られる経路はエンド間遅延時間などの性能指標の面では必ずしも最適ではない [8, 9]。またこのことは、大規模ネットワーク障害の発生などの非常時におけるネットワーク接続性にも影響を与える。

そこで本稿では、既存の IP ネットワークを前提とし、近年着目されているオーバレイネットワーク技術を用いて、大規模ネットワーク障害の発生時に短時間で代替経路を発見し、非常時通信を実現するオーバレイルーティング技術の提案を行う。IP ネットワーク上に論理ネットワークを構築するオーバレイネットワーク技術は現在様々なアプリケーション (ファイル交換、音声通話、IP-VPN、コンテンツ配信など) で用いられ、サービスオーバレイネットワークと呼ばれている [10]。本稿では IP ネットワークとサービスオーバレイネットワークの中間に位置し、経路制御を行うルーティングオーバレイに着目する。従来提案されているルーティングオーバレイは適用できるネットワーク規模が限られているため、本稿ではスケーラビリティを高める改善手法を提案する。

提案手法の有効性は、CAIDA [11] が BGP トラヒックの計測を行い公開している AS ネットワークトポロジを用いて検証し、提案手法がオーバレイノード間の情報交換量を従来手法に比べて 1/10 - 1/1000 程度に削減できること、および大規模ネットワーク障害に対して、高いネットワーク接続性を維持し、BGP に比べて短時間で代替経路を発見することができることを示す。

以下、2 章では研究の背景として非常時通信およびオーバレイネットワークについて述べる。3 章で提案手法を説明し、4 章において AS 間トポロジデータを用いた性能評価例を示す。最後に 5 章で本稿のまとめと今後の課題を述べる。

2. 研究の背景

2.1 非常時通信における問題点

大規模災害、テロなどの発生によって、情報ネットワークにおいても障害が広範囲に渡って発生する。また、ルータを制御しているソフトウェア (OS) の不具合によって、同時に多数のルータ動作が不良となることも考えられる。このような非常時における通信において求められるのは、ネットワーク接続性のすばやい回復および重要通信の優先的処理である。本稿において着目している前者に関しては、多くの研究が行われているが、それらのほとんどにおいては単一障害、すなわち、ネットワークの構成要素の障害は同時には 1 つしか発生しないことが前提とされている。つまり、発生し得る障害をあらかじめ想定し、想定した障害に対して効率の良い手法が検討されている。したがって、一般的にそれらの手法は大規模かつ面的に発生するネットワーク障害に対しては有効ではない。

また、障害発生時の対応を含むネットワーク制御においては、コストと性能のトレードオフが重要となる [1]。このとき、非常時通信は大規模ネットワーク障害という発生確率の小さい事象に対するために必要となるため、コストが非常に大きくなる。例えば [12] では、MPLS ネットワークにおいて装置故障が発生

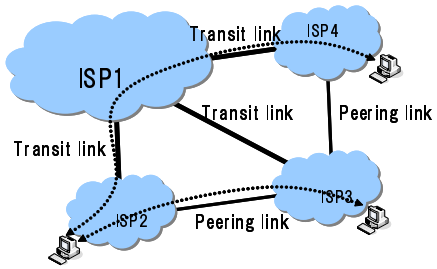


図 1　非常時通信におけるネットワーク接続性

した際に早期にデータ転送を復旧するネットワークアーキテクチャが提案されているが、ネットワークを現用面と予備面の 2 つ用意する必要があり、その導入コストは非常に大きいと考えられる。そのため、低コストで運用可能なアーキテクチャが求められる。

一方、現在のインターネットにおいて AS 間の経路制御を司っている BGP が、大規模な障害発生時やフルメッシュ構造などの特殊なトポロジ環境において不安定であることが近年指摘されている。この問題に対する改善案も多数提案されているが、それらは BGP そのものやルータの改変を必要とするため、その導入には標準化作業が必要となり、非常に長い時間かかる。さらに、Internet Service Provider (ISP) 間の接続性を提供するリンクには、上位 ISP からインターネット全体への接続性の提供を受けるためのトランジットリンクや、同程度の規模の ISP 間でトラヒックのやりとりを行うためのピアリングリンクなど、経済的コスト構造が異なるリンクが存在する。各 ISP はそれらのコスト構造に加えて政治的思惑なども考慮し、BGP を用いることによって経路制御を行っている。

このことが、非常時におけるネットワーク接続性に影響を与える例を図 1 に示す。ISP 2 – 4 は ISP 1 との間にトランジットリンクを持ち、インターネット全体への接続性を確保している。さらに ISP 2、3 間および ISP 3、4 間にはピアリングリンクが存在する。ピアリングリンクには通常、接続されている 2 つの ISP を始点および終点とするトラヒックのみが流れる。すなわち、ISP 2 から ISP 4 への通信はトランジットリンクおよび ISP 1 を経由して行われ、ISP 3 への通信はピアリングリンクを用いて行われる。この時、ISP 1 に障害が発生し、トランジットリンクの全てが利用できなくなる状況を考える。この時、ネットワークトポロジとしては ISP 1、3 間にピアリングリンクを 2 段経由する経路が存在するが、実際に用いることはできない。これは、ISP1、2 間のピアリングリンクは始点や終点が ISP 3 であるトラヒックは通過できず、さらに ISP2、3 間のピアリングリンクは始点や終点が ISP 1 であるトラヒックは通過できないためである。

この問題を解決し、非常時通信において接続性を向上させるためには、非常時にはルーティング設定を変更して複数のピアリングリンクを経由するような経路も利用可能にする必要がある。しかし、そのためには BGP 設定を注意深く行う必要があり、非常時の設定変更には通常 ISP のオペレータ同士の折衝が必要となるため、実現は困難であると考えられる。

2.2 オーバレイルーティング

オーバレイネットワークとは、下位層ネットワークである IP ネットワークの上に独自の論理ネットワークを構築するものであり、例えば P2P ネットワーク、Grid ネットワーク、IP-VPN サービスなどが挙げられる。これらのアプリケーションは、ある特定のサービスを前提として論理ネットワークを構築する。また、それぞれのアプリケーションのポリシーにしたがってアプリケーショントラヒックの制御を行う。例えば、P2P のファイル交換ネットワークは、コンテンツの所在場所に応じてダウンロードホストや中継ホストを選択する。

さらに、特定のアプリケーションを前提とせず、トラヒックのルーティングそのものを目的 (アプリケーション) とするオー

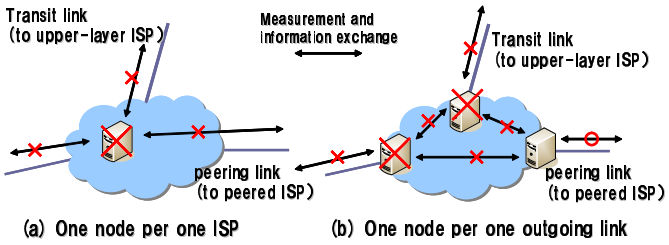


図 2　ISP へのオーバレイノード設置

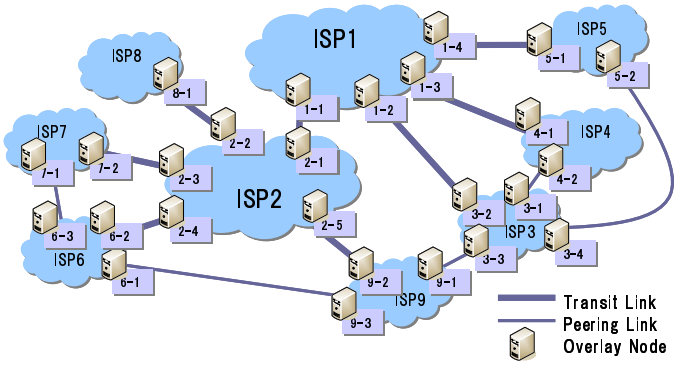


図 3　オーバレイノード配置例

	1	···	i	···	N
1					
⋮					
j			(connectivity, timestamp)		
⋮					
N					

図 4　オーバレイ到達性情報テーブル (ORT)

バレイルーティングと呼ばれる技術も登場しつつある。例えば Resilient Overlay Network (RON) [13] では、参加ノード間の伝送遅延時間やパケット廃棄率などを計測し、あるノード間のデータ転送を直接行うのか、他のオーバレイノードを経由して行うのかを判断している。これにより、IP ネットワークでのルーティングと比較して効率の良いデータ転送を実現し、また IP ネットワークの障害をすばやく検知し、迂回経路を選択することが可能である。この技術を非常時通信に適用し、RON のノードを各 ISP (AS) に配置しオーバレイルーティングを行うことで、下位層の IP ネットワークを変更することなく、ネットワーク障害発生時にすばやく経路切り替えを実現する手法を実現することができると考えられる。

しかし、RON は参加しているノード間でフルメッシュに計測および情報交換を行うため、計測オーバヘッドが大きく、数十ノード程度しか参加できないとされている [14]。したがって、このまま非常時通信に適用すると、参加できる ISP (AS) 数が限定されてしまう。また、迂回経路として 2 ホップ経路、すなわち、送信ノードと受信ノードの間に 1 つだけ中継ノードを挟む経路のみを考慮している。これは、エンド間遅延時間や空き帯域の観点では、3 ホップ以上の経路を選択することによるメリットは大きくないためである。しかし、非常時通信において最も重要となるのは接続性そのものであるため、接続性を確保するために 3 ホップ以上の経路を用いることは重要である。次章では、オーバレイルーティング技術を非常時通信に適用する際のこれらの問題点を解決する手法の提案を行う。

3. 提 案 手 法

本章では、本稿で提案する非常時通信のためのオーバレイルーティング手法 (以下、非常時オーバレイと呼ぶ) の説明を行う。

3.1 概　　要

提案する非常時オーバレイには、AS や ISP などの単位に相当する、ある程度の大きさを持ったネットワーク単位で参加することを前提としている。以下の説明では ISP 単位で参加することを仮定する。また、提案手法の核となるオーバレイノードは、ISP が他 ISP との間に持つ対外接続リンクが接続されているルータ上に設置する。これは、図 2(a) に示すように、各 ISP に 1 つずつオーバレイノードを設置すると、ISP がネットワーク障害によって部分的に通信不可能になりオーバレイノードがそれに含まれた場合に、残った部分も通信不可能となってしまうためである。一方、図 2(b) のように対外リンク毎にオーバレイノードを設置することによって、部分的障害が発生した場合においても、残った部分がオーバレイルーティングによって対外接続を維持することができる。

図 3 に、9 つの ISP から構成される IP ネットワークおよびオーバレイノード設置例を示す。この例では、各 ISP の対外接続リンク部分に 26 個のオーバレイノードが存在し、これらがフルメッシュにオーバレイリンクを設定し、オーバレイネットワークを構築している。

設置された各オーバレイノードは、他 ISP に設置されたオーバレイノードとの間で到達性を確認すると共にオーバレイ到達性情報テーブル (以降 Overlay Reachability Table: ORT と表記) の交換を行い、参加しているオーバレイノード間のフルメッシュ到達性情報を取得する。ネットワーク障害が発生した場合には、自身が保持している ORT に基いてトラヒックを運ぶ経路を決定する。

以降、オーバレイ到達性情報テーブル、到達性の計測とテーブル交換方法、経路探索手法、および ISP の参加および離脱方法について順に説明する。

3.2 オーバレイ到達性情報テーブル (ORT)

図 4 に、各オーバレイノードが持つオーバレイ到達性情報テーブル (ORT) を示す。テーブルのエントリ (i,j) には、オーバレイノード i が計測したオーバレイノード j への到達性の情報と、計測した時刻を表すタイムスタンプが含まれる。したがって、エントリ (i,j) とエントリ (j,i) は共にオーバレイノード i、j 間の到達性情報を示すが、計測を行った主体が異なる。到達性情報には、オーバレイノード間の計測によって得ることのできるパケット廃棄率、空き帯域、伝播遅延時間などが含まれる。タイムスタンプは情報の新しさを確認するために用いる。

本テーブルは、各オーバレイノードは非常時オーバレイに参加している全てのノード間の到達性情報を把握し、ネットワーク障害発生時に代替経路を発見するために用いる。そのためには、自ノード以外の全てのノードとの到達性情報を計測によって獲得するとともに、それを他ノードと交換することが必要となる。

3.3 オーバレイノード間の到達性計測とテーブル交換

前述のように、各 ISP の対外接続リンクごとにオーバレイノードを設置すると、非常時オーバレイに参加する ISP に比べてオーバレイノード数が増加するため、参加する ISP 数に対するスケーラビリティが低下する。特に RON と同様のフルメッシュ計測およびテーブル交換を行う場合、RON が良好に動作する規模が 50 ノード程度であることを考慮すると、非常時オーバレイに参加することのできる ISP 数は 10 程度に抑えられる。そこで提案手法においては、計測およびテーブル交換を行う他 ISP のオーバレイノード群を、ISP 内に設置された各オーバレイノードで分割することによって、オーバヘッドを削減し、より多くの ISP が参加できるようにする。

具体的には、他 ISP に設置されたオーバレイノードそれぞれに対して、自 ISP からそのノードへの IP ルーティング情報を参照し、自 ISP が持つ対外接続リンクのうち、どのリンクか

ら送出されるのかを調べ、そのリンクに設置されたオーバレイノードが担当するものとする。これは、自 ISP 内に設置された複数のオーバレイノードを仮想的に 1 つのオーバレイノードとみなし、他ノードとの間の到達性の計測とテーブル交換を行うことに相当する。

ノード間の到達性確認は、ノード間に TCP コネクションを確立することで行う。その際、確立した TCP コネクションを用いてお互いが持つ ORT を送ることによって、テーブル交換を行う。さらに、他ノードから獲得した ORT と自身が持つ ORT の各エントリのタイムスタンプを比較し、新しいものがあれば自身の ORT のエントリを更新する。その後、自 ISP 内のオーバレイノード同士でフルメッシュに ORT の交換を行い、到達性情報の共有を行う。これにより、自 ISP 内の全てのオーバレイノードが、他の全てのオーバレイノードへの到達性情報を持つことができる。

図 3 の ISP 2 を例にとると、ISP 2 内のオーバレイノード 2-1〜2-5 はそれぞれ下記に示すノードとの間で到達性計測とテーブル交換を行う。

- 2-1: 1-1〜1-4, 3-1〜3-4, 4-1, 4-2, 5-1, 5-2
- 2-2: 8-1
- 2-3: 7-1, 7-2
- 2-4: 6-1, 6-2, 6-3
- 2-5: 9-1, 9-2, 9-3

この場合、ISP 2 内のオーバレイノードが行う通信回数は 31 回となる。一方、この分割を行わず全てのオーバレイノードがフルメッシュに到達性確認およびテーブル交換を行う場合は 125 回となる。

3.4 経路探索

各オーバレイノードは、3.3 節に示した到達性確認およびテーブル交換によって得られた最新の ORT に基いて、ダイクストラ法に基いて各ノードへの経路探索を行う。一般にダイクストラ法はノード数の 2 乗の計算時間がかかるが、障害が発生していない部分はフルメッシュにオーバレイリンクが存在するため、実際の計算量は小さいと考えられる。

この手法により、RON と同様に、2 ホップパス (送信ノードと受信ノードの間に 1 つの中継ノードを経由させる経路) は固定時間で見つけることができる。3 ホップ以上のパスに関しては前節で示したテーブル交換の順序に依存するが、最悪の場合でも (ホップ数-1) と固定時間の積で発見可能である。BGP によるルーティング情報の伝播もホップ数に比例した時間が必要となるが、BGP は隣接ルータに更新情報を伝えるのに対して、提案手法はフルメッシュに張られたオーバレイパスを用いてテーブルを伝播させるため、BGP に比べて短時間でルーティング情報の伝播が可能となる。

さらに、本方式を用いることで、ネットワーク障害発生時に通常は取得できない情報を取得することができる。図 3 において ISP 2 に部分的なネットワーク障害が発生し、ノード 2-3、2-4 および 2-5 が設置されている 3 本の対外リンクが不通になった場合を想定する。この場合、ノード 5-1 からノード 6-3 およびノード 7-1 へは通常の IP ルーティングでは到達不可能となるため、ノード 6-3 とノード 7-1 間の接続性に関する情報を直接獲得することができない。しかし、(1) ノード 6-1 とノード 9-3 の間のテーブル交換、(2)ISP 9 内でのテーブル共有、(3) ノード 9-1 とノード 3-3 の間のテーブル交換、(4) ノード 5-1 とノード 3-3 の間のテーブル交換、というステップによって、ノード 5-1 は ISP 7 が持つ接続情報を獲得することができる。

また、この (1)-(4) 伝播経路は、そのままノード 5-1 とノード 7-1 が通信するための 4 ホップオーバレイパスとなる。この経路は 3 本のピアリングリンクを経由しており、通常の BGP ルーティングでは用いられることはない。このように、本提案手法によって障害発生時のネットワーク接続性が向上することが期待される。

3.5 ISP の参加および離脱

新たな ISP が非常時オーバレイに参加する場合の手続きは以下のようになる。

- 新規 ISP の作業
  - 既参加 ISP から、参加ノード情報および ORT を取得し、自 ISP 内に設置するオーバレイノード全てにコピーする
  - 自 ISP 内に設置するオーバレイノード間のフルメッシュ計測を設定する
  - 他ノードに対する到達性計測およびテーブル交換の担当を、IP ルーティング情報に基いて決定する (3.3 節参照)
  - 全ての既参加 ISP へ自 ISP の参加および設置したオーバレイノードに関する情報を通知する
- 既存 AS の作業
  - 新規 ISP 内に設置されるオーバレイノードへの IP ルーティング情報を基に、到達性計測およびテーブル交換を担当するノードを決定する

一方、ISP が非常時オーバレイから離脱する場合は下記のようになる。

- 離脱 ISP の作業
  - 全ての他ノードに対して離脱を通知する
- 既存 ISP の作業
  - 離脱通知を受けたノードのエントリを OTL から削除する
  - 離脱 AS に対する到達性計測およびテーブル交換を担当していたノードはその AS を対象から外す

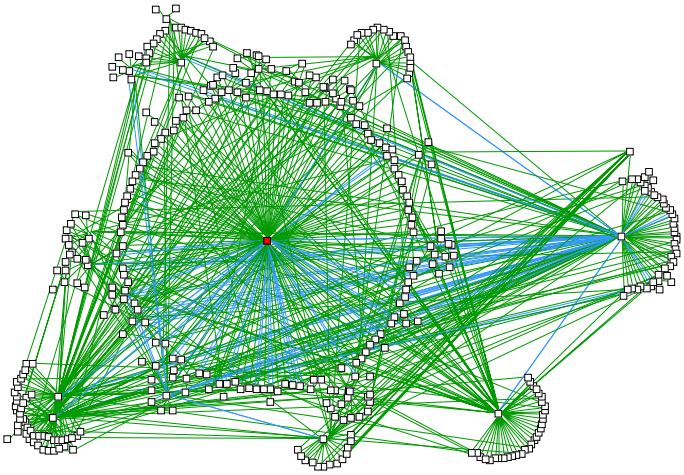


図 5　JP ネットワーク

4. 性能評価

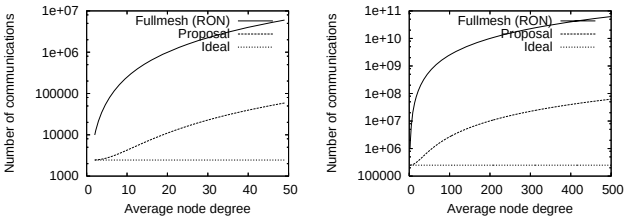
本章では、3. 章で提案した提案方式の性能評価を行う。まず、到達性確認およびテーブル交換のために必要となる通信量を評価し、従来手法に比べて通信量を大幅に削減できることを示す。また、大規模なネットワーク障害が発生した際に、代替経路を BGP に比べて短時間で発見できるとともに、ネットワーク接続性を高く維持できることを示す。

4.1 ネットワークモデル

性能評価に用いるネットワークは AS がランダムに他 AS と接続しているランダムネットワーク、および JP ネットワークとする。JP ネットワークは CAIDA [11] が BGP データの計測に基いて公開している AS 間の接続関係を示すトポロジデータのうち、JPNIC 管轄の AS のみを抽出したものを用いる。図 5 に JP ネットワークトポロジを示す。AS 数は 414、トランジットリンクが 692 本、ピアリングリンクが 50 本存在する。

4.2 通信量

まず、到達性確認およびテーブル交換のために必要となる通信量の評価を行う。ネットワーク内に存在する各 AS(ノード) が対外接続リンクにオーバレイノードを設置し、全ての他ノードと到達性確認およびテーブル交換を 1 回ずつ行うために必要



(a) ランダムネットワーク (AS 数: 50)　(b) ランダムネットワーク (AS 数: 500)

図 6　通信量 (ランダムネットワーク)

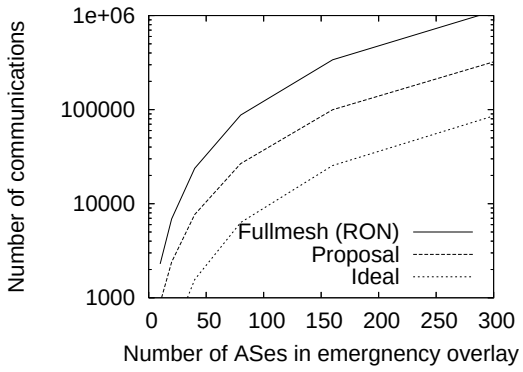


図 7　通信量 (JP ネットワーク)

となる通信回数 (TCP コネクションの確立回数) を求める。

図 6 は、AS 数が 50 (図 6(a)) および 500 (図 6(b)) の場合において、ネットワークの平均次数と通信回数の関係を表している。なおここでは、AS 間のリンクは全てトランジットリンクと考え、全てのオーバレイノード間のトラヒックが通過可能であるとしている。図中の「Ideal」は、各 AS にオーバレイノードを 1 つずつ設置した場合の情報交換量を表している。図から、ネットワークの平均次数が増加すると、各 AS が持つ対外接続リンクが増加するため、オーバレイノード数が増加し、通信量が増加することがわかる。また、提案手法を用いることによって、通信量を最大で約 1/1000 に削減できることがわかる。これは、ランダムネットワークを対象としており、かつリンクが全てトランジットリンクとして用いられるため、AS 内の各オーバレイノードが担当する他 AS のオーバレイノードがほぼ均等に分割されることに起因している。このような環境においては、提案方式の通信量削減効率は非常に高い。

次に、図 5 に示す JP ネットワークを用いた評価結果を示す。ここでは、トランジットリンクは全てのオーバレイノード間のトラヒックが通過可能であるとし、ピアリングリンクはそれによって接続されている 2 つの AS 内のオーバレイノード間のトラヒックのみが通過できるとしている。図 7 は、全 414 個の AS のうち非常時オーバレイに参加する AS 数を変化させた場合の、通信量の変化を示したものである。図中の「Ideal」は、非常時オーバレイに参加している各 AS にオーバレイノードを 1 つずつ設置した場合の情報交換量を表している。図から、提案手法を用いることによって通信量が約 1/10 に削減されることがわかる。前述のランダムネットワークの場合に比べて提案手法の効果が小さいのは、JP ネットワークの平均次数が 1.79 であることが主な原因であると考えられる。

4.3　経路切り替え時間およびネットワーク接続性

次に、ネットワーク障害が発生した後の経路切り替えにかかる時間、およびネットワーク接続性の評価を行う。ここでは図 5 に示す JP ネットワークを用い、BGP (IP) ルーティングは、トランジットリンクは全てのオーバレイノード間のトラヒックが通過可能であるとし、ピアリングリンクはそれが接続する 2 つの AS 内のオーバレイノード間のトラヒックのみが通過できるとし、最短経路を用いるものとする。障害発生時には、各ノードは (障害発生地点からのホップ数- 1)×(固定時間 (Mean Router Advertisement Interval: MRAI)) 後に新たな経路を発見できるものとする。これは、BGP ルーティングの経路が最も早く収束する場合に相当すると考えられる。一方提案手法は、3.4 節に示した時間で新たな経路を発見可能であるとする。なお固定時間部分に関しては IP ルーティングおよび提案手法とも同じ値であるとし、経路切り替えのために必要となる情報伝播のホップ数を用いた評価を行う。

表 1　経路切り替え時間とネットワーク接続性 (AS 障害)

Trouble: 1 AS							
	NE	0	1	2	3	>3	Connectivity
BGPa	0.44	0.18	0.22	0.04	0.0	0	0.88
EMa(10)	0.42	0.31	0.04	0.0	0	0	0.76
EMa(50)	0.40	0.46	0.02	0.0	0	0	0.88
EMa(100)	0.40	0.52	0.01	0.0	0.0	0.0	0.93
BGPp	0.26	0.52	0.01	0.0	0	0	0.80
EMp(10)	0.22	0.41	0	0	0	0	0.63
EMp(20)	0.21	0.68	0	0	0	0	0.89
EMp(40)	0.20	0.80	0	0	0	0	0.996
Trouble: 3 ASes							
	NE	0	1	2	3	>3	Connectivity
BGPa	0.19	0.16	0.16	0.04	0.0	0	0.54
EMa(10)	0.18	0.18	0.02	0.0	0	0	0.37
EMa(50)	0.18	0.38	0.03	0.01	0.0	0	0.61
EMa(100)	0.18	0.53	0.03	0.0	0.0	0	0.74
BGPp	0.06	0.32	0.02	0	0	0	0.40
EMp(10)	0.12	0.27	0	0	0	0	0.39
EMp(20)	0.12	0.60	0	0	0	0	0.72
EMp(40)	0.12	0.87	0	0	0	0	0.994
Trouble: 5 ASes							
	NE	0	1	2	3	>3	Connectivity
BGPa	0.05	0.13	0.14	0.03	0.01	0	0.35
EMa(10)	0.06	0.07	0.0	0	0	0	0.12
EMa(50)	0.05	0.21	0.02	0.01	0.0	0.0	0.30
EMa(100)	0.05	0.40	0.03	0.02	0.01	0.01	0.52
BGPp	0.03	0.24	0.02	0	0	0	0.29
EMp(10)	0.11	0.29	0	0	0	0	0.40
EMp(20)	0.10	0.64	0	0	0	0	0.74
EMp(40)	0.10	0.89	0	0	0	0	0.995

発生させるネットワーク障害は下記の 2 種類を想定する。

- AS 障害: 選択した AS ノード、およびその AS が持つ対外リンクが全て故障する。
- IX 障害: 選択した複数の AS 間を接続しているリンクが全て故障する。AS そのものは故障しないため、故障していないリンクは利用可能である。

また、障害の度合いを変化させるために、障害が発生する AS 数を変化させる。その際、より深刻な障害を想定するために、次数の高い AS から選択する。

表 1 は、AS 障害を想定し、障害が発生する AS 数が変化する場合における、ネットワーク接続性 (通信が継続できる AS ペアの割合) および代替経路発見のために必要となる情報伝播ホップ数の割合の変化を示している。なおここでは、BGP の場合 (BGPa および BGPp)、非常時オーバレイに参加する

表 2　経路切り替え時間とネットワーク接続性 (IX 障害)

Trouble: 4 ASes							
	NE	0	1	2	3	>3	Connectivity
BGPa	0.87	0.0	0.08	0.03	0.0	0.0	0.993
EMa(100)	0.85	0.15	0.01	0.0	0.0	0.0	0.996
BGPp	0.87	0.0	0.08	0.03	0.0	0.0	0.993
EMp(40)	0.84	0.16	0	0	0	0	1.000
Trouble: 10 ASes							
	NE	0	1	2	3	>3	Connectivity
BGPa	0.62	0.02	0.28	0.07	0.0	0.0	0.993
EMa(100)	0.58	0.40	0.0	0.0	0.0	0	0.994
BGPp	0.62	0.02	0.28	0.07	0.0	0.0	0.993
EMp(40)	0.55	0.44	0	0	0	0	0.995
Trouble: 20 ASes							
	NE	0	1	2	3	>3	Connectivity
BGPa	0.45	0.04	0.43	0.06	0.01	0.0	0.993
EMa(100)	0.42	0.53	0.02	0.0	0.0	0.0	0.98
BGPp	0.45	0.04	0.43	0.06	0.01	0.0	0.993
EMp(40)	0.43	0.56	0	0	0	0	0.995

AS を JP ネットワークの全ての AS からそれぞれ 10、50、100 (EMa(10)、EMa(50)、EMa(100)) 個ランダムに選択した場合、および非常時オーバレイに参加する AS を JP ネットワークにおいてピアリングリンクを持つ AS からそれぞれ 10、20、40 (EMp(10)、EMp(20)、EMp(40)) 個選択した場合を比較している。また、表中の「NE」は障害によって影響を受けない AS ペアの割合、「0」は情報伝達なしで代替経路を発見できる AS ペアの割合、および「1」、「2」、「3」はそれぞれ 1、2、3 ホップの情報伝播で代替経路が発見できる AS ペアの割合、および「>3 は」4 ホップ以上の情報伝播が必要となる AS ペアの割合をそれぞれ示している。

この表から、全ての AS から非常時オーバレイに参加する AS を選択する場合、非常時オーバレイに 10 個以上の AS が参加することで、代替経路を発見するまでの時間が大幅に改善されることがわかる。また、50 個以上の AS が参加することで、ネットワーク接続性が BGP に比べて改善することがわかる。一方、非常時オーバレイに参加する AS をピアリングリンクを持つ AS から選択する場合には、代替経路が発見できる全ての場合において、情報伝播なしで (オーバレイノード自身の到達性確認のみで) 代替経路が発見できる。また、接続性に関しては、10 個の AS の参加で接続性は BGP に比べて改善し、40 個の AS の参加で接続性は 99% 以上になる。これは、ピアリングリンクを持つ AS が非常時オーバレイに参加することで、代替経路の選択肢が増加するため、効果が高いことを示している。このことから、ピアリングリンクやトランジットリンクを多くもつ AS が提案する非常時オーバレイに参加することで、その効率が改善されるといえる。

表 2 は、IX 障害を想定した場合における、表 1 と同様の結果を示している。表から、IX 障害を想定する場合には、BGP を用いた場合においても最終的な接続性はほとんど劣化しないことがわかる。これは、IX 障害においては選択した AS 間のリンクのみの障害であるため、代替経路が容易に発見できるためであると考えられる。また、提案手法を用いることで、ほぼ 100%の代替経路を情報伝播なしで発見できることがわかった。

5. おわりに

本稿では、既存の TCP/IP ネットワークの上にルーティングを行うオーバレイネットワークを構築することで、大規模ネットワーク障害が発生した場合に短時間で代替経路を発見可能な非常時オーバレイネットワークの提案を行った。提案手法は既存のオーバレイルーティング手法を基盤としているが、ノード間の到達性確認およびルーティング情報の交換を ISP 内で分担して行うことで、オーバヘッドを 1/10 - 1/1000 程度に削減し、参加 ISP 数に対するスケーラビリティを向上している。また、オーバレイルーティング技術を用いることで、従来の BGP ルーティングでは用いることができなかった経路が利用可能となり、BGP ルーティングに比べてネットワーク接続性を最大で約 9 倍改善することができることが明らかとなった。

今後の課題としては、さらなる通信量削減の方法について検討することが挙げられる。また、本稿における提案方式は参加しているノードでフルメッシュのオーバレイネットワークを構築する手法に基いたものであり、ネットワーク障害が発生した場合には全てのオーバレイノードがオーバレイルーティングを開始する。今後は、障害発生地域の周辺だけでルーティングのためのオーバレイネットワークを動的に構築することにより、より効果的なルーティングを行う手法を検討したい。

謝　辞

本研究の一部は，文部科学省科学技術振興調整費「先端融合領域イノベーション創出拠点の形成：ゆらぎプロジェクト」の研究助成によるものである。ここに記して謝意を表す。

文　献

[1] 村田正幸, “サービスオーバレイによるネットワークの高信頼化,” **電子情報通信学会総合大会** *(BT-1-5)*, Mar. 2005.

[2] C. Labovitz, A. Ahuja, A. Abose, and F. Jahanian, “Delayed Internet routing convergence,” in *Proceedings of ACM SIGCOMM 2000*, Aug. 2000.

[3] B. Zhang, D. Massey, and L. Zhang, “Destination reachability and BGP convergence time,” in *Proceedings of GLOBECOM 2004*, Apr. 2004.

[4] C. Labovitz, A. Ahuja, R. Wattenhofer, and S. Venkatachary, “The impact of Internet policy and topology on delayed routing convergence,” in *Proceedings of INFOCOM 2001*, Dec. 2001.

[5] Dan Pei and Matt Azuma and Nam Nguyen and Jiwei Chen and Dan Massey and Lixia Zhang, “BGP-RCN: Improving BGP convergence through root cause notification,” Tech. Rep. TR-030047, UCLA CSD, Oct. 2003.

[6] William Norton, “Internet service providers and peering.” available at `http://www.equinix.com/pdf/whitepapers/PeeringWP.2.pdf`.

[7] William Norton, “A buisness case for peering.” available at `http://www.equinix.com/pdf/whitepapers/Business_case.pdf`.

[8] Y.Zhu, C. Dovrolis, and M. Ammar, “Dynamic overlay routing based on available bandwidth estimation: A simulation study,” *Computer Networks Journal*, vol. 50, pp. 739–876, Apr. 2006.

[9] D. G. Andersen, A. C. Snoeren, and H. Balakrishnan, “Best-path vs. multi-path overlay routing,” in *Proceedings of ACM SIGCOMM conference on Internet measurement*, pp. 91–100, Oct. 2003.

[10] Z. Duan, Z.-L. Zhang, and T. Hou, “Service overlay networks : SLAs, QoS and bandwidth provisioning,” in *Proceedings of IEEE ICNP 2002*, Nov. 2002.

[11] The CAIDA Web Site. available at `http://www.caida.org/home/`.

[12] 三堀英彦, 錦戸淳, 河村仙志, “Type-X による高信頼 mpls マネージドネットワークの実現,” *NTT* **技術ジャーナル**, June 2003.

[13] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, “Resilient overlay networks,” in *Proceedings of 18th ACM Symposium on Operating Systems Principles*, Oct. 2001.

[14] A. Nakao, L. Peterson, and A. Bavier, “Scalable routing overlay networks,” *ACM SIGOPS Operating Systems Review*, vol. 40, pp. 49–61, Jan. 2006.