論　文

プレフィックス数の変動に基づくBGP障害リンク推定手法*

渡里　雅史[†a)]　立花　篤男[†]　阿野　茂浩[†]　山崎　克之[††]

A Method for Inferring BGP Link Failures Based on Prefix Variation*

Masafumi WATARI[†a)], Atsuo TACHIBANA[†], Shigehiro ANO[†], and Katsuyuki YAMAZAKI[††]

あらまし　インターネットにおいて，ドメイン間の接続リンクで発生する障害は，多数のユーザの通信品質を劣化させるとともに，該当リンクを経由する多数のサイトへの到達性を失う原因となる．このため，ISPにおいて，経路障害の発生原因となる不安定なリンクを特定することは，適切なルーチングポリシを設計し，経路制御を行う上で重要である．しかしながら，急速に大規模化・複雑化が進む今日のインターネットにおいては，経路変動が定常的に発生しているため，障害による経路変動を適切に抽出できず，発生箇所を正確に特定できない課題がある．これに対し，筆者らは，利用時間の長い最適経路上の各リンクにおいて観測される，一部のプレフィックスの変動に着目することにより，障害箇所を高精度に推定する手法を考案した．本論文では，提案手法の概要について述べるとともに，実ネットワークにおける評価実験を通して，提案手法が従来の手法に比べ高精度に発生箇所を推定可能であることを示す．

キーワード　インターネット，Border Gateway Protocol (BGP)，障害箇所推定

1. まえがき

家庭向け高速ブロードバンド環境の普及，更には3Gや Wi-Fi 等の無線機器を搭載した小型端末の普及により，ユーザのインターネットへのアクセスが益々便利になりつつある．これに加え，リッチコンテンツの流通及び利用拡大により，インターネットを流れるトラヒック量は，年々急増しており，インターネットがますます重要な社会インフラとして確立されつつある．一方，Autonomous System (AS) の相互接続により形成されるインターネットでは，ユーザの通信品質は，各AS間リンク（以降，リンク）の安定性に大きく左右される．不安定なリンクは，多数のユーザの通信品質を劣化させるとともに，該当リンクを経由する多数のサイトへの到達性を失う原因となる．このため，ISPにおいて，経路障害の発生原因となる不安定なリンクを特定することは，適切なルーチングポリシを設計し，経路制御を行う上で重要である．

これに対して，Border Gateway Protocol (BGP) [1] ネットワークにおいてリンク障害の発生箇所を推定する手法が数多く研究されてきた [2]〜[4]．これらの手法は，リンク障害により BGP 経路変動メッセージ（以降，BGP メッセージ）が発生することに着眼し，単一または複数の計測地点において観測する BGP メッセージ群から障害箇所を推定する．しかしながら，多数の AS が複雑に接続された今日のインターネットにおいては，経路変動が定常的に発生しており，また，経路収束の過程に伴い発生する多数の BGP メッセージが，障害による経路変動の抽出を困難とするため，発生箇所を正確に特定できない課題がある．一方，定期的に経路表を収集し，断片的な経路表の変化から障害箇所を推定する手法が提案されているが [5]〜[7]，この場合は，瞬間的に発生・復旧する経路障害を検出できない課題がある．

そこで，筆者らは，単一の計測地点で観測した BGP メッセージ群から，経路収束に伴い発生する BGP メッセージを区別した上で，BGP メッセージごとに経路表を作成し，その変化から経路障害の発生箇所を推定する手法を提案した [8]．本論文では，提案手法の概要

† （株）KDDI 研究所，ふじみ野市
KDDI R&D Laboratories Inc., Fujimino-shi, 356–8502 Japan
†† 長岡技術科学大学，長岡市
Nagaoka University of Technology, Nagaoka-shi, 940–2188 Japan
a) E-mail: watari@kddilabs.jp
* 本論文は，インターネットアーキテクチャ研究専門委員会推薦論文である．

と実ネットワークにおける評価実験を通して，その有用性について述べる．

本論文の構成は以下のとおりである．**2.** で従来の障害箇所推定手法の概要と課題について説明する．**3.** で課題を解決する提案手法について述べ，**4.** で実ネットワークにおける提案手法の推定精度を示す．**5.** で考察を行い，**6.** で結論を提示する．

2. BGP 障害箇所推定手法の課題

図 1 に BGP ネットワークにおける障害箇所推定手法の概要を示す．従来の障害箇所推定手法の多くは，単一または複数の計測地点において観測する BGP メッセージ群と経路表より障害箇所を推定する．BGP メッセージには，各 AS が広報するプレフィックスごとに広報元 AS からの通過経路を示す AS パス属性が含まれる．例えば，図 1 において，計測地点で観測する AS20 の広報プレフィックス $p20$ の AS パス属性は，障害発生前後でそれぞれ (0 10 20) と (0 30 40 20) となる．本論文では，本 AS パス属性に示される二つの AS 間の接続関係をリンクと呼ぶ．例えば，上記の AS パス属性により抽出されるリンクは，$(0, 10)$，$(10, 20)$，$(0, 30)$，$(30, 40)$，$(40, 20)$ の五つとなる．

BGP ネットワークにおいて経路障害の発生箇所を推定する手法として，一定時間ごとに作成される経路表を用いて，断片的な経路表の変化から障害箇所を推定する手法がある [5]〜[7]．提案手法は，本手法の拡張により障害箇所を推定するため，本論文では，はじめに従来手法の概要と課題について述べる．

2.1 従来手法の概要

従来手法 [5]〜[7] は，一定時間ごとに作成される経路表から各リンクを通過するプレフィックス数を計算し，前後の経路表からプレフィックス数が多く減少するリンクを障害箇所として推定している．各リンクを通過するプレフィックス数は，BGP の経路表において，それぞれのリンクが出現する回数より算出する．例えば，表 1 の経路表における各リンクの通過プレフィックス数を表 2 に示す．経路表は，α 秒（[5], [6] では 240 秒, [7] では 180 秒）ごとに作成し，前後の経路表においてプレフィックス数が多く減少したリンクを障害箇所として推定する．仮に図 1 のリンク $(10, 20)$ 及び $(30, 20)$ における経路障害の発生により，AS20 が広報する $p20$，$p21$ のプレフィックスが α 秒以内に AS40 経由で収束した場合，これらのリンクにおけるプレフィックス数が 0 個になるため，各リンクが経路障害の発生箇所として推定される．

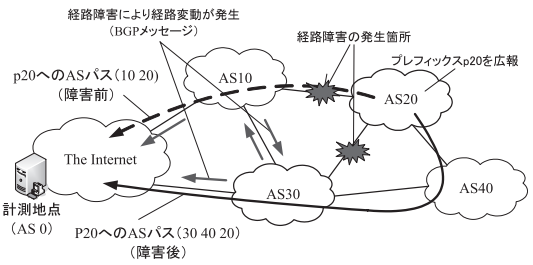

図 1 BGP メッセージのパッシブ計測による障害箇所推定
Fig. 1 Origin inference using passively collected BGP udpate messages.

2.2 従来手法の課題

従来手法は，一定のインタバル α ごとに作成される経路表から各リンクを通過するプレフィックス数を計算するため，α 秒以内に発生・復旧する経路障害を検出できない課題がある．仮に BGP メッセージごとに経路表を作成し，プレフィックス数を計算した場合では，正常なリンクや一時的に観測するリンクを多数誤検出する課題がある．

例えば，図 2 の AS10〜AS70 で構成された簡単な BGP ネットワークにおいて，リンク $(40, 60)$ の経路障害に伴い発生する BGP メッセージを用いて各リンクのプレフィックス数を算出する場合を考える．このとき，経路障害が発生する前の経路表を $RIB1$ とする．また，AS60 では，$p60$〜$p63$ を AS40 へ，$p64$ を

表 1 計測点における経路表
Table 1 Routing table at monitoring point.

Prefix	AS Path
p10	0 10
p11	0 30 10
p20	0 10 20
p21	0 30 20
p30	0 30
p31	0 30
p40	0 30 40

表 2 リンク別プレフィックス数
Table 2 Number of prefixes of each AS link.

AS Link	Number of Prefixes
0 10	2
0 30	5
10 20	1
30 10	1
30 20	1
30 40	1

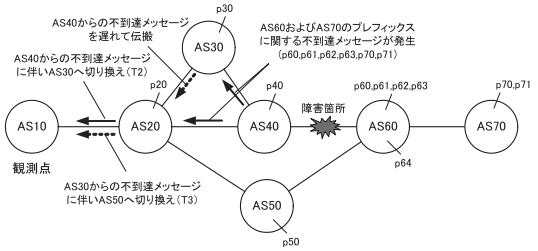

図 2　BGP における経路収束の特徴
Fig. 2　Convergence property of BGP.

表 3　従来手法における各リンクのプレフィックス数変動
Table 3　Prefix variation of each link using exisiting method.

経路表	10, 20	20, 30	20, 40	20, 50	40, 60	50, 60	60, 70	30, 40
$RIB1$	11	1	7	2	6	1	2	0
$RIB2$	11	7	1	2	6	1	2	6
$RIB3$	11	1	1	8	0	7	2	0
$RIB4$	11	1	7	2	6	1	2	0

AS50 へ向けて広報している．はじめに AS40 は，リンク (40, 60) の経路障害の検出により，隣接する AS20 及び AS30 に対して，AS60 及び AS70 の各 AS が広報するプレフィックス群の不到達メッセージを送信する．本メッセージを受信した AS20 は，代替経路として AS30 経由の AS パスを AS10 に通知し，計測点において経路表が更新される ($RIB2$)．しかしながら，AS20 は，直後に AS30 からも同じ不到達メッセージを受信するため，代替経路として AS50 経由の AS パスを AS10 に通知し，計測点において再び経路表が更新される ($RIB3$)．更に，リンク (40, 60) の経路障害が復旧した場合は，計測点における経路表が更新され ($RIB4$)，障害発生前と同じ経路表が作成される．

一連の経路変動に対して通知される BGP メッセージから，各リンクを通過するプレフィックス数を集計した結果を表 3 に示す．実際の障害箇所であるリンク (40, 60) に加え，一時的に観測するリンク (30, 40) においてもプレフィックス数が 6 個から 0 個に減少するため，本リンクを誤検出する場合がある．更に，リンク (20, 30)，(20, 40)，(20, 50)，(50, 60) のような正常なリンクにおいてもプレフィックス数が 6 個減少するため，変動量に基づき障害箇所を推定した場合では，これらのリンクを誤検出する可能性が考えられる．すなわち，従来手法では，経路収束の過程に伴い発生する BGP メッセージや障害箇所の復旧に伴い発生する BGP メッセージにより，正常なリンクや一時的に観測するリンクにおいてプレフィックス数が変動し，これらのリンクを誤検出する課題がある．

3. 提案手法

本論文では，従来手法における課題を解決するため，単一の計測地点で観測した BGP メッセージ群から，経路収束時や復旧時に発生する BGP メッセージを区別した上で，BGP メッセージごとに経路表を作成し，その変化に着眼することで，経路障害の発生箇所を高い精度で推定する手法を提案する．以下に，提案手法における推定手順を示し，各手順の詳細について述べる．

（1） プレフィックスごとに最適経路を推定

（2） プレフィックス数が 0 となるリンクを検出（候補箇所）

（3） 候補リンク群から障害箇所を推定

3.1 最適経路の推定

従来手法における最大の課題は，経路収束の過程に伴い発生する BGP メッセージや障害箇所の復旧に伴い発生する BGP メッセージにより，正常なリンクや一時的に観測するリンクにおいてプレフィックス数が変動し，これらのリンクを誤検出する点にある．文献 [9] によれば，n 個の AS がメッシュ状に接続されたネットワークでは，最大で $n!$ 個の AS パスを一時的に観測する可能性があり，多くのリンクが誤検出されることが考えられる．そこで，提案手法では，各プレフィックスに対して一つの最適経路を推定し，最適経路を示す BGP メッセージと経路収束や障害箇所の復旧時に発生する BGP メッセージを区別することで，これらのメッセージに対するプレフィックス数の集計方法を変更する．

最適経路の推定には，BGP の最適経路選択アルゴリズムで最優先される最小パス長を有する AS パス（例えば，AS パス 10 20 40 60 70 のパス長は 5）を最適経路とする方法が考えられるが，実運用では様々な運用ポリシに基づき経路選択が行われているため，本手法による最適経路の推定は困難である．そこで，本論文では，運用ポリシが反映された最適経路が，各 AS において最も長く利用される AS パスである可能性が高い点に着眼し [10]，各プレフィックスに対して観測する AS パスのうち，評価期間において観測時間が最も長い AS パスを最適経路とした．具体的には，プレフィックス p_i で観測する n 個の AS パス $\{path_1^i,\ path_2^i,\ \ldots,\ path_n^i\}$ に対して，それぞれの AS パスの累積観測時間を $\{usage_1^i,\ usage_2^i,\ \ldots,\ usage_n^i\}$ とした場合，$path_j^i$

の利用割合 P_j^i を次のとおり定義した．

$$P_j^i = \frac{usage_j^i}{\sum_{k=1}^{n} usage_k^i} \quad (1)$$

このうち，利用割合が最大 P_{max}^i となる AS パスをプレフィックス p_i の最適経路とした．

3.2 障害箇所候補の検出

正常なリンクや一時的に観測するリンクにおけるプレフィックス数の変動により，これらのリンクを誤検出するのを回避するため，**3.1** においてプレフィックス p_i に対して最適経路として推定した AS パス $path_{best}^i = (AS_m, \ldots, AS_0)$ において，本プレフィックスの集計対象とするリンクを，広報元 AS_0 とのリンク (AS_1, AS_0) に限定する．例えば，図 2 において，プレフィックス $p70$ 及び $p71$ における最適経路を $path_{best} = (10\ 20\ 40\ 60\ 70)$ とした場合，本プレフィックスの集計対象リンクは $(60, 70)$ となる．

表 3 の RIB1 を各プレフィックスの最適経路として，図 2 のリンク $(40, 60)$ で発生した経路障害に対して，本手法を用いてプレフィックス数を集計した結果を表 4 に示す．従来手法では，**2.2** で示したとおり，六つのリンクでプレフィックス数の変動が発生していたのに対して，提案手法では，障害箇所であるリンク $(40, 60)$ でのみプレフィックス数が減少するため，正常なリンクや一時的に使用されるリンクの誤検出を回避可能となる．

本論文では，本集計手法によりプレフィックス数が 0 となるリンクを障害箇所の候補とする．また，このときの時刻を検出時刻とする．ただし，BGP の経路収束により，連続した検出を回避するため，一度検出したリンクは，再び最適経路で復旧するまで対象外とする．次節において，これらの候補リンク群から障害箇所を推定する手法について述べる．

3.3 障害箇所の推定

提案手法により検出されるリンクは，一点観測の特性上，一部正常なリンクが含まれる場合がある．例えば，図 2 において，リンク $(40, 60)$ の経路障害に加え，リンク $(50, 60)$ で経路障害が発生した場合では，AS70 が広報するプレフィックス $p70$，$p71$ は計測点において不到達となるため，リンク $(60, 70)$ におけるプレフィックス数が 0 となり，本リンクも障害箇所として検出される．更に，本リンクの検出時刻は，AS 間の接続構成や BGP ルータの実装・タイマなどの影響により，BGP メッセージの伝搬に時間差が生じる結果，リンク $(40, 60)$ の検出時刻と異なる場合がある．このため，障害箇所の推定には，上記特性を考慮し，リンク $(60, 70)$ を障害箇所から除く必要がある．なお，この問題は，複数の計測点で観測した BGP メッセージを用いることで解決できる可能性が考えられるが，そのためには，全ての AS が計測点となる必要があり，インターネットでは現実的な解決策ではない．

表 4 提案手法における各リンクのプレフィックス数変動
Table 4 Prefix variation of each link using proposed method.

経路表	$10, 20$	$20, 30$	$20, 40$	$20, 50$	$40, 60$	$50, 60$	$60, 70$	$30, 40$
$RIB1$	1	1	1	1	4	1	2	0
$RIB2$	1	1	1	1	4	1	2	0
$RIB3$	1	1	1	1	0	1	2	0
$RIB4$	1	1	1	1	4	1	2	0

本論文では，各 AS に対して一つまたは複数の最適経路を決定し（プレフィックスに対しては一つの最適経路を決定），最適経路ごとに検出時刻の近いリンク群を同一の経路障害に伴う変動とみなし，障害箇所を推定する手法を提案する．以下にアルゴリズムの概要を示す．

（1） 各 AS が広報するプレフィックス群から一つまたは複数の最適経路を決定し，最適経路上で検出されるリンク群を一つの集合とする．

（2） 同集合のリンク群を検出時刻の順にソートし，最初のリンク検出から，BGP の経路収束時間を考慮し，T 秒以内に検出されるリンク群を同一の経路障害に伴う検出としてクラスタ化する．ただし，同一リンクで連続した経路障害が発生する可能性を考慮し，T 秒以内であっても同一リンクがクラスタ内に既に含まれる場合は，同リンク以降を別のクラスタとして分割する．

（3） 一点観測の特性上，クラスタごとに計測点に最も近いリンクを障害箇所として推定する．

4. 評　　価

提案手法の有用性を検証するため，米国オレゴン大学が提供する RouteViews [12] における公開 BGP データを用いた評価実験を行った．RouteViews では，2010 年 1 月 13 日現在，36 の計測点において収集された BGP 経路表及び BGP メッセージが公開されており，本評価では，アジア太平洋地域と米国を結ぶ国際的な学術ネットワークである AS22388 とインターネットにおける Tier 1 ISP である AS3356 の二つの計測点を対象とした．表 5 に，2009 年 9 月 1 日〜30 日に

表 5　評価期間におけるデータの概要
Table 5　Overview of BGP data for evaluation.

計測点	AS 名	プレフィックス数	AS 数	リンク数
AS22388	TRANSPAC	14,490	1,970	2,534
AS3356	LEVEL3	310,991	32,502	56,172

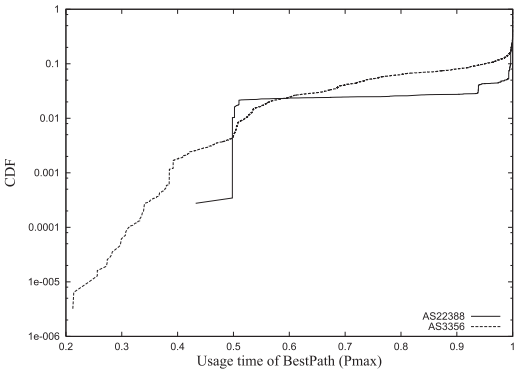

図 3　P_{max}の分布
Fig. 3　Distribution of P_{max}.

おいて各計測点で観測した BGP データの概要を示す．

本章では，はじめに提案手法の事前評価として，各計測点における最適経路の推定精度と最適な経路収束時間 T を検証する．

4.1　最適経路の推定

各計測点において，AS パスの観測時間から各 AS 間の最適経路を推定する手法の有効性を検証するため，評価期間に観測した全プレフィックスを対象に P_{max} を集計した（図 3）．その結果，各計測点ともに約 55%のプレフィックスでは $P_{max} = 1$ となり，一つの AS パスのみを利用しており，最適経路の可能性が高いことを確認した．更に，AS22388 では約 97%，AS3356 では 92%のプレフィックスが $P_{max} > 0.9$ と高く，大部分のプレフィックスに対して一つの AS パスを抽出できることを確認した．一方，一部のプレフィックスでは，P_{max} の値が低いことから，評価期間において運用ポリシの変更や観測期間が極端に短い場合などの原因が考えられる．しかしながら，全体の分布としては極めて少ないことが分かる．更に，これらプレフィックスによって検出される経路障害は，最適経路上のリンクではなく迂回経路上のリンクとなるだけであり，他の障害に対する推定への影響はないため，今後の分析課題とする．

4.2　経路収束時間の推定

提案手法は，**3.3** で述べた一点観測の特性上の課題を解決するため，同一最適経路上で最初に検出されるリンクから，BGP における経路収束時間を考慮し，T 秒以内に検出される他のリンク群を同一の経路障害と判断する．このとき，小さすぎる T の値は，収束時間の差により遅れて検出されるリンクを異なる障害として誤検出する可能性がある一方，大きすぎる T の値は，異なる経路障害を一つの障害として誤検出する可能性がある．そこで，本論文では，各計測点において，経路収束時間を推定し，T の値を決定した．具体的には，T の値を 60〜300 秒とした場合に作成されるクラスタから，各クラスタにおいて最初と最後に検出されるリンクの検出時刻の差（最大値は T）より，経路障害による経路変動の収束時間を推定し，T の値をこの収束時間以上とした．ただし，T 秒以内に同一リンクが検出される場合，異なる経路障害により他のリンクが同一クラスタに分類される可能性が高いと考えられるため，T 秒以内に同一リンクが検出される頻度が低いことも条件とした．

図 4 及び図 5 に各計測点における収束時間の分布と図 6 に T の値を変化させた場合の，同一リンクが検出される確率の分布を示す．まず，AS22388 では，各 T により作成されるクラスタの収束時間には大きな差がないことを確認した（図 4）．具体的には，$T = 60$ 及び $T = 300$ により作成されるクラスタ数は，それぞれ 12,115 個，12,084 個とその差は小さいことから，多くのクラスタが 60 秒以内に収束することを確認した．一方，図 6 においては，$T = 200$ 秒あたりから同一リンクの検出数が増加しており，期間内に異なる経路障害が発生している可能性が高いと考えられる．このため，AS22388 における T の値は $T = 60$〜200 秒が妥当であると考えられる．

一方，AS3356 においては，各 T により作成されるクラスタ内の経路収束時間に大きな差があることを確認できる（図 5）．具体的には，$T = 60$ 及び $T = 300$ により作成されるクラスタ数は，それぞれ 162,958 個，80,865 個と倍以上の差があり，$T = 60$ では経路収束していないと考えられる．また，$T = 240$ ではクラスタ数が 80,882 個となっており，多くのクラスタが収束していることを確認した．一方，図 6 より，T 内における同一リンクの検出確率は一定であり，また，全クラスタの 0.1%未満と十分に小さいことから，AS3356 における経路収束時間は，$T = 240$〜300 秒前後が妥当であると考えられる．

以上の結果より，以降の評価では，AS22388 と

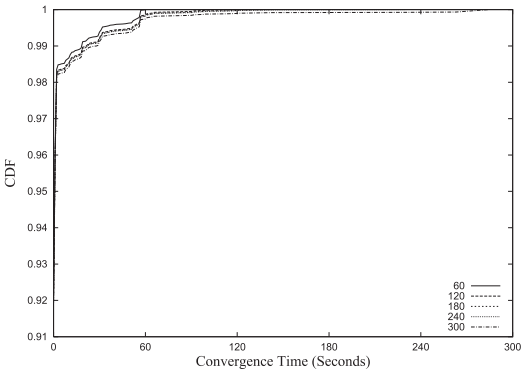

図 4　クラスタ内の収束時間差 (AS22388)
Fig. 4　Covergence time variance of each cluster. (AS22388)

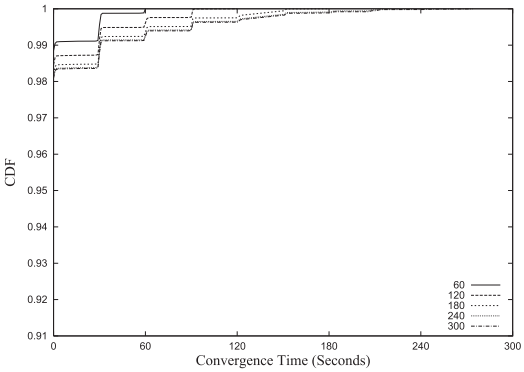

図 5　クラスタ内の収束時間差 (AS3356)
Fig. 5　Covergence time variance of each cluster. (AS3356)

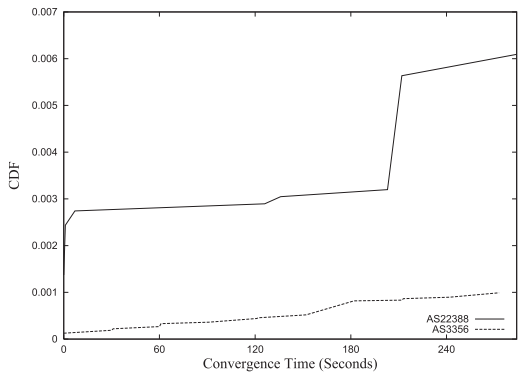

図 6　同一リンクの検出確率
Fig. 6　Distribution of ratio of detecting identical links.

AS3356 に対してそれぞれ $T = 120$ 秒，$T = 240$ 秒を採用した．

4.3　APAN 運用チケットを用いた評価

提案手法の有用性を評価するため，APAN-JP (AS7660) のネットワークオペレーションセンタ（大手町）において発行された 10 件の運用チケット [13] を用いて検出内容を検証した．本検証では，提案手法において，各運用チケットに記載された経路障害の発生時刻及び発生箇所に一致する検出の有無を確認した．なお，計測点には，AS7660 との接続を有する AS22388 を用いた．以下に検証した運用チケットの一部と検証結果を示す．

運用チケット 1

2009 年 9 月 3 日に AS7660 と AS37889 とのリンクにおいて，メンテナンス作業により 09:00 及び 12:05 に 2 回のセッション断が発生した．提案手法では，該当リンクにおいて 08:55:18 と 12:04:14 に該当リンク障害を検出しており，これらの検出時刻が運用チケットの記載時刻に近いことから，メンテナンス作業に伴うセッション断を正しく検出できていたと考えられる．

運用チケット 2

2009 年 9 月 15 日に AS22388 と AS7660 とのリンクにおいて，「Unexpected number of routes」により，14:10 にセッション断が発生した．提案手法では，該当リンクを最適経路上にもつ 403 の AS パスにおいて，14:11:21 に該当リンク障害を検出し，検出時刻が運用チケットの記載時刻に近いことから，本障害に伴うセッション断を正しく検出できていたと考えられる．なお，このうち 209 の AS パスでは最適経路上の他のリンクにおいても同時刻または 1 秒後にプレフィックス数が 0 となっており，提案手法が **3.3** に示したクラスタリング手順により，多数の候補リンク群における経路変動を同一障害に伴う経路変動として正しく推定できていることを確認した．

このほか，検証した全ての運用チケットにおいて，記載時刻に近い時間帯で経路障害を検出しており，提案手法が有効であることを確認した．

4.4　従来手法との比較評価

提案手法の効果を検証するため，提案手法において障害箇所の候補として検出されるリンク数と，提案手法がベースとした従来手法において，プレフィックス数が 0 となり検出されるリンク数を集計した．なお，評価の簡略化のため，各 AS に対する最適経路は一つとした．表 6 に示すとおり，提案手法における検出数は，従来手法に比べ約 64〜約 67%少なく，従来手法では経路収束により一時的に観測するリンクを多く誤

表 6　従来手法に対する提案手法の検出数
Table 6　The number of detected links using exisiting method.

手法	AS22388		AS3356	
	対象リンク数	検出数	対象リンク数	検出数
従来手法	2,533	22,941	56,172	208,391
提案手法	2,266	8,290	53,434	68,831

表 7　提案手法における生成クラスタ数と推定リンク数
Table 7　The number of detected links using proposed method.

計測点	評価パス数	検出数	生成クラスタ数	推定リンク数
AS22388	1,870	8,290	12,097	6,571
AS3356	32,502	68,831	80,882	63,576

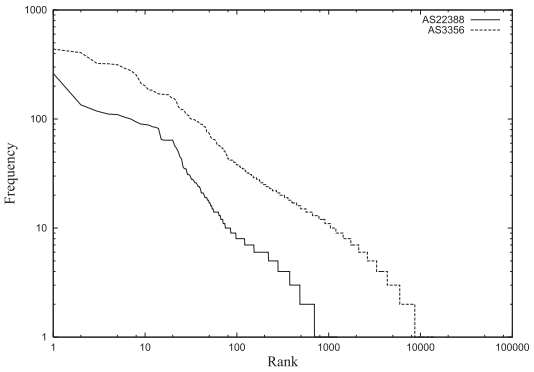

図 7　各リンクにおける障害回数の分布
Fig. 7　Number of link failures detected for each link.

表 8　インターネット (AS3356) における障害回数の多いリンク（トップ 10）
Table 8　Top 10 links detected over 1-month period at AS3356.

Rank	Link (X-Y)	Frequency	AS(X)	AS(Y)	Prefixes	Link Type
1	3356-42567	2480	Level 3 Communications	Simply Media TV	1	Edge
2	80-19981	438	General Electric Company	(Not Listed)	9	Edge
3	8218-49517	406	AS Confederation of Neotelecoms	Teikhos	2	Edge
4	34419-48728	323	Vodafone Group Services	Vodafone Qatar Q.S.C.	4	Edge
5	1239-23765	321	Sprint	Electronic Arts, Australia	1	Edge
6	12741-48922	315	Netia SA	ZK Technologie S.A.	1	Edge
7	47358-34618	292	NTRnet s.r.l.	Prometeo	3	Edge
8	22351-8668	278	Intelsat	TelOne Zimbabwe	10	Transit (for 3 ASes)
9	701-40345	254	MCI Communications Services	IP-Com, Inc.	2	Edge
10	1239-38861	213	Sprint	StarHub Internet Exchange 2	477	Transit (for 4 ASes)

検出していたと考えられる．

表 7 に提案手法において検出された候補リンク群から，最終的に障害箇所として推定したリンク数を示す．表 7 に示すとおり，提案手法は，障害箇所候補となる検出数のうち，約 8〜21%少ないリンクを障害箇所として推定することを確認した．すなわち，これらのリンク数が一点観測の特性上，従来手法では誤検出される可能性があることを確認した．また，最終的な推定リンク数は，従来手法に比べ，約 69〜約 71%少なく，推定結果を大幅に補正することを確認した．

一方，提案手法において，各プレフィックスの集計対象リンクを評価した結果，約 2.9〜約 6.5%のリンクでは，集計対象となるプレフィックスが存在しておらず，プレフィックス数の変動を検出できないことを確認した．このため，これらのリンクを最適経路上にもつ約 3.8〜約 8.7%の AS パスでは，障害箇所を誤推定する可能性がある．また，提案手法により生成される約 1.2〜約 7.2%のクラスタには，検出時刻が 1 秒以上異なるリンクが含まれており，これらのクラスタでは，本来異なる障害を同一障害としてクラスタ化することで，一部の障害を正しく検出できない可能性がある．しかしながら，一般に入手できる運用チケットは限られており，これらの事象を全て検証するためには，他の AS の管理者の協力が必要不可欠であり，今後の課題である．

5. 考　　察

提案手法による推定結果に基づき，各計測点において，障害回数の多いリンク順に発生頻度の分布を図 7 に示す．各ネットワークで検出される経路障害の多くが，少数のリンクで発生しており，各 BGP ネットワークで発生する経路障害が，自然界の様々な現象で見られる Zipf の法則に近い分布で発生していることを確認した．実際には，約 86〜約 95%の経路障害が，約 20%のリンクで発生していた．また，全検出数の約 87〜約 89%における発生箇所がエッジのリンクであった．一般的に，インターネットが経路障害に強いのは，スケールフリーの構成をもつためといわれているが [14]，実際は本特徴だけではなく，インターネットで発生する経路障害の発生箇所が接続性への影響が少ないエッジのリンクで発生しているからとも考えられる．

表 8 にインターネット (AS3356) において障害回

数の多い上位のリンクを示す．各リンクの AS 名は，CIDR-REPORT [15] から取得した．また，各リンクを通過するプレフィックス数を把握するため，リンクごとに評価期間において観測した最大プレフィックス数を集計した．インターネットにおける経路障害の多くが，トポロジー上の末端に位置するエッジリンクにおいて発生しているが，一部のリンクは，多数のプレフィックスに対して接続性を提供するトランジットリンクであった．これらのトランジットリンクにおいて発生する経路障害は，インターネット全体の経路を不安定にするだけでなく，大規模な通信品質の劣化やインターネット到達性の低下を招く原因となる．提案手法により，これらの不安定なリンクを特定することで，ISP において，これらのリンクを回避したルーチングポリシの設計と経路制御が可能となる．

6. むすび

本論文では，単一の計測地点で観測した BGP メッセージ群と，それに伴い作成される経路表の変化から，経路障害の発生箇所を高い精度で推定する手法を提案した．実ネットワークで収集した BGP データを用いた評価実験の結果，提案手法による検出結果が検証した運用チケットの記載内容と一致することを確認し，ネットワーク全体としては，従来手法の推定結果を約 69〜約 71%補正することを確認した．また，インターネットにおける障害状況を分析した結果，毎分約 1.47 件の経路障害が発生しており，その約 86〜約 95%の経路障害が，約 20%のリンクで発生していることを確認した．更に，全検出数の約 87%がエッジのリンクであることから，インターネットで発生する経路障害の多くが接続性への影響が少ない箇所で発生していることを確認した．

今後，実際の運用において更に事例を収集し，精度の評価及び提案手法の改善を進める予定である．

文　献

[1] Y. Rekhter, T. Li, and S. Hares, “A border gateway protocol 4 (BGP-4),” RFC 4271, Jan. 2006.

[2] A. Feldmann, O. Maennel, Z. Mao, A. Berger, and B. Maggs, “Locating Internet routing instabilities,” Proc. ACM SIGCOMM, 2004.

[3] M. Lad, A. Nanavati, D. Massey, and L. Zhang, “An algorithmic approach to identifying link failures,” Proc. 10th IEEE PRDC04, 2004.

[4] D. Chang, R. Govindan, and J. Heidemann, “The temporal and topological characteristics of BGP path changes,” 11th IEEE International Conference on Network Protocols (ICNP’03), 2003.

[5] M. Lad, L. Zhang, and D. Massey, “Link-rank: A graphical tool for capturing BGP routing dynamics,” IEEE/IFIP NOMS 2004, 2004.

[6] M. Lad, R. Oliveira, D. Massey, and L. Zhang, “Inferring the origin of routing changes using link weights,” IEEE ICNP 2007, 2007.

[7] A. Campisano, L. Cittadini, G. Battista, T. Refice, and C. Sasso, “Tracking back the root cause of a path change in interdomain routing,” IEEE NOMS 2008, 2008.

[8] 渡里雅史，立花篤男，阿野茂浩，山崎克之，“プレフィックス数の変動に基づく BGP 障害リンク推定手法の提案,” 信学技報，IA2009-119, March 2010.

[9] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, “Delayed Internet routing convergence,” Proc. ACM SIGCOMM, Aug. 2000.

[10] R. Oliveira, B. Zhang, D. Pei, and L. Zhang, “Quantifying path exploration in the Internet,” IEEE/ACM Trans. Netw., vol.17, no.2, pp.445–458, 2009.

[11] Asia-Pacific Advanced Network, http://www.jp.apan.net/, July 2010.

[12] Route Views Project, http://www.routeviews.org/, July 2010.

[13] APAN-JP NOC Tickets, http://www.jp.apan.net/NOC/tickets/, July 2010.

[14] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On power-law relationships of the Internet topology,” ACM SIGCOMM, 1999.

[15] CIDR Report, AS Names, http://www.cidr-report.org/as2.0/autnums.html, Jan. 2010.

（平成 22 年 10 月 8 日受付，23 年 1 月 21 日再受付）

渡里　雅史　（正員）

平 15 慶大・環境情報・環境情報卒．平 17 同大大学院修士課程了．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．

立花　篤男　（正員）

平 12 阪大・工・電子情報エネルギー卒．平 14 同大大学院修士課程了．同年 KDDI（株）入社．以来，研究所にて，IP ネットワーク計測・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主査．

阿野　茂浩　（正員）

昭 62 早大・理工・電子通信卒．平元同大大学院修士課程了．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．

山崎　克之　（正員：フェロー）

昭 55 電通大・通信卒．工博．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 18 から長岡技術科学大学教授．