広域ネットワークにおける経路の監視・管理方法

高村 真俊† 多田 卓央§ 門林 雄基‡ 山口 英†

†奈良先端科学技術大学院大学情報科学研究科
§シャープ株式会社
‡大阪大学大型計算機センター
†*masato-t@is.aist-nara.ac.jp*, *suguru@is.aist-nara.ac.jp*
‡*youki@center.osaka-u.ac.jp*

現在、多くのネットワーク管理システムでは、独立した個々のネットワーク機器、つまり管理者の権限の及ぶ範囲のネットワーク機器のみを監視・管理することが目的となっている。しかし広域ネットワークでは、隣接する独立したネットワークから受け取る経路情報の不具合が、管理しているネットワークの経路情報に影響を与えることがしばしばあるため、権限の及ぶ範囲のみを監視したとしても問題を解決することはできない。よって広域ネットワークを安定運用していく上で、経路情報を監視することは極めて重要な意味を持つ。

本研究は、経路情報を監視し、解析することで広域ネットワークにおけるトラブルシューティングをより容易に短時間で行うことを目的とする。監視する項目として、経路テーブルの大きさの変化、ピアが切れる頻度、経路がフラップする頻度、ルータの設定情報の変化、インターフェースの状態などがある。これらの情報の量は膨大なものであるため、できるだけ自動的に解析し、少量の情報のみを管理者に通知するシステムが必要になる。

本稿では、広域ネットワークにおける経路を管理する手法を提案する。さらに、それに基づいたシステムの実装と現段階での運用・評価を行い、その有効性について考察する。

A mehod for monitoring routing information in large-scale internetworks

Masatoshi Takamura† Takuo Tada§ Youki Kadobayashi‡ Suguru Yamaguchi†

†Graduate School of Information Science, Nara Institute of Science and Technology
§Sharp Corporation
‡Computation Center, Osaka University

Stability of large-scale internetworks heavily depend upon the routing information exchanged through multiple administrative domains. Since routing information in single administrative domain are prone to human errors in other administrative domains, it is important to monitor both internal and external routes. In large-scale internetworks, the traditional method of managing reachability within single administrative domain is not helpful for most cases.

This paper focus on management of large-scale internetworks through periodic monitoring and analysis of routing information that are exchanged across the administrative border. Most of the state changes in internetworks can be detected by periodically monitoring the routing table size, the frequency of dropped peering connection, route flaps, changes in router configuration, and the transition of interface state. Automated analysis and summarization of these observation are necessary, since the amount of information that are obtained through monitoring can be quite huge.

This paper proposes a method for monitoring routing information in large-scale internetworks. We evaluate its effectiveness through its implementation, operational experience and evaluation.

1 目的

経路制御は、インターネットを円滑に運用していく上で最も重要な項目の一つである。インターネットが急激に拡大したことにともない、制御しなければならない経路数は膨大になっている。今日のインターネットではインターネット全体の経路制御プロトコルとしてBGP4[1]を用いており、BGP4によってネットワーク組織間で交換される経路の数は4万6000を越える。

膨大な経路情報から、経路に関するトラブルの原因をつきとめるのは容易なことではない。このような状況であるにもかかわらず、経験を積んだ一握りのネットワーク管理者がこの問題に対処しているというのが現状である。BGP4ではネットワーク組織(AS: Autonomous System)ごとに経

路を管理しており、それぞれの AS で経路の障害に対応している。

経路に障害が生じた場合、問題になっている経路上のルータだけを調査しても障害を除去することができない場合が多い。他の AS のルータから管理対象 AS にアナウンスされた経路情報、および受け取った経路情報を管理対象 AS でどのように取り扱っているかなどを総合的に判断し、障害の発生原因を突き止めなくてはならない。障害の発生時に、原因を特定するためには多くのルータを調査しなければならず、たいへんな労力を要する。

経路情報の管理とは、経路情報に関連した情報を収集し、統計をとり、それらの情報から障害を発見し、障害を取り除くことである。さらに、収集した統計情報から、経路テーブルの大きさの変化や、障害の発生頻度を把握できる。これらは今後のネットワークの設計・運用に有用な情報となる。本研究では、インターネットにおける標準的な経路制御プロトコルである BGP4 における経路情報を取り扱う。

本研究は、次のことを目的としている: 1) 広域ネットワーク上で使用される経路制御プロトコルを定期的に監視し、統計情報を管理者に分かりやすいかたちで報告できること、2) 障害の早期発見ができること、3) 障害の発生原因を突き止めるために、調査対象となる経路やルータなどを指定するだけで自動的に関連する情報を入手できること。

本稿は以下のように構成される。まず、経路の管理手法を述べ、その手法を用いた経路監視システムを提案する。次に、現状での実装について述べ、現状の実装における運用と評価について述べる。最後に今後の課題について検討する。

2 経路管理手法

提案手法は 1 つの AS に注目して AS 内にサーバを設置し、AS ボーダルータを監視するものである。

今までのネットワーク管理システムは、図 1の (a) のように権限の及ぶネットワーク内の機器の状態を管理することが目的であり、別のネットワークとの到達性 (reachability) まで監視するものはなかった。また、管理対象機器の情報は SNMP によって収集されるため、管理対象機器までの経路が失われた場合、情報収集することができなかった。

それに対し、本提案手法は、図 1の (b) のように、AS 間の経路制御で使用される標準的な経路制御プロトコルである BGP によって得られる経路情報を監視し、AS 間の経路情報を管理する。さらに、 telnet を用いることで管理対象機器までの経路上に障害が発生した場合も、情報収集することが可能である。

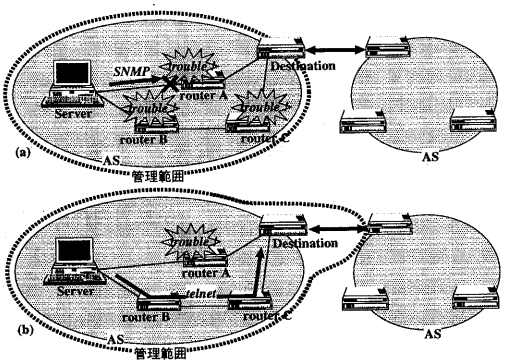

図 1: 経路テーブルの収集

本提案手法では、以下にあげる BGP4 の障害を管理・監視対象とする。

- **ピアの切断**
  BGP4 ではルータ間で TCP コネクションを確立し、TCP コネクション上で経路情報を交換する。これをピア (peering connection) と呼ぶ。ピアが切断された場合、経路情報が経路テーブルから消え、該当するネットワークへ到達できなくなる。

- **経路情報の欠落**
  ピアが張られていても、特定のネットワークへの経路情報を誤って抹消される場合がある。この場合も、該当するネットワークへ到達できなくなる。

- **誤った経路情報の挿入**
  パラメータ (origin AS など) の設定を誤った経路情報が流される場合がある。この場合も、該当するネットワークへ到達できなくなる。

- **不要な経路の混入**
  海外からの経路情報などの不要な経路情報が誤って流される場合がある。この場合、ルータに負荷がかかり、通信品質の劣化や通信不可の原因となる可能性がある。

- **メトリックの変化**
  BGP4 では各々の経路情報に対して MED (Multi Exit Discriminator), Local Preference と呼ばれるメトリック値が定義される。

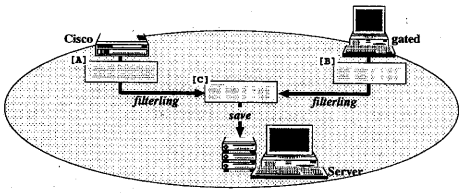

図 2: 経路テーブルの収集

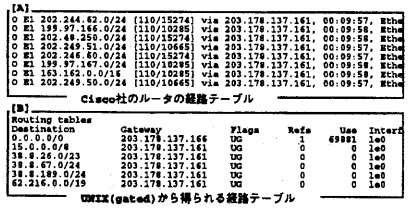

図 3: 経路テーブルの例

これらの設定を誤った場合、該当するネットワークへの通信品質が劣化する場合がある。

これらの障害は複数のルータにおいて以下の情報を収集することによって検出することができる：経路テーブル、ルータの設定情報、インターフェイスの状態。

提案手法の基本的な手順は次の通りである: 1) 他の AS と BGP によって経路制御を行っている AS ボーダルータから得られる経路に関する情報を AS 内のサーバが定期的に収集する。2) 前の情報と収集した情報との差分をとり、障害検出を行う。3) 差異があればネットワーク管理者に報告する。

3 経路監視システムの設計

本章では、2章で述べた手法を実現するためのシステムを提案する。

さらに、統計情報へのアクセスや過去の障害履歴の検索を行う具体的な方法についても述べる。

3.1 収集

ルータから収集する情報は、経路テーブル、ルータの設定情報、BGP テーブル、インターフェイスの状態である。経路テーブルの収集の流れを図 2 に示す。また、Cisco 社のルータ、gated の稼働しているルータからそれぞれ得られる経路テーブルの例を図 3に示す。

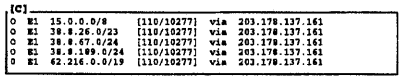

図 4: フィルタリング後の経路テーブル

収集は、Unix 上で cron を利用して 30 分毎に行う。定期的に収集された情報はデータベースに蓄積され、統計情報として利用する。統計情報はネットワーク上の慢性的な障害や今後のネットワークの運用・設計に役立つと考えられる。

ルータからの情報の取得は telnet を用いて行う。これは、データ収集するサーバと情報を取得したいルータとの間の経路上の障害が発生したとき、情報を得たいルータに SNMP では到達できないためである。

telnet でデータを取得するため、ルータの種類によって得られる情報の形式が違う。Cisco 社のルータおよび gated の稼働しているルータから得られる経路テーブルの情報の一部分を図3の［A］、［B］に示す。これらの情報を、収集後の情報検索を容易にするためにデータ形式を図 4の［C］のように統一し、データベースに保存する。

また、経路テーブルのデータ量は膨大なため、サーバ側で効率の良い保存方法を考える必要がある。データを効率よく保存することでネットワーク管理者への報告や、統計情報の作成、過去の障害履歴の検索などを容易に行うことができる。

これを実現するため、カリフォルニア大学バークレイ校で開発された POSTGRES をサーバへ導入し、データベースを構築する。

3.2 障害検出

定期的に収集した情報を用いて、以前の状態と現在の状態の差分をとり障害検出を行う。

但し、Cisco 社のルータに限り以下の障害が発生した場合、差分をとらなくても現在の状態だけで障害を検出できる。

- **BGP のピアが切れた場合**

  BGP summary によって得られる情報から、BGP のピアが切れたことを検出できる。ピアが切れている状態を表している BGP summary の例を図 5 に示す。図 5の場合、*202.249.47.18* のピアが Active 状態であるため、ピアが確立され経路の交換途中であることがわかる。

```
Neighbor        V    AS Up/Down  State
133.186.1.2     4 65531 9w44
202.241.1.97    4  2513 07:04:14
202.249.47.18   4  4717 00:00:21 Active
203.178.136.1   4  2500 17:15:04
203.178.136.2   4  2500 16:33:51
203.178.136.3   4  2500 3d17h
203.178.136.4   4  2500 1w3d
.........
```

図 5: ピアが切れているときの BGP summary

- AS ループ

  BGP テーブルには各経路の AS Path が含まれているので、AS のループを検出することができる。

なお、gated を使用した場合でも特定の経路を監視することによって特定のピアが切れたことを検出することは可能である。

しかし、ネットワークに障害が生じた場合、定期的に収集したデータと過去のデータとの差分だけでは、障害の原因を発見するための情報としては不十分である。経路制御における問題は、ほとんどの場合において、複数のルータを調査しなければ原因がわからない場合が多い。ネットワーク管理者は、障害の報告の後、図 6の（a）のように多数のルータから一つ一つ手動で障害状況を調査しなければならない。

障害を検出後、ネットワーク管理者が必要であると思われる情報を収集し、通知するシステムが必要である。しかし、障害の傾向を知り、原因を割り出す作業は豊富な経験と知識が必要であり、すべての障害に対して対応できる手順を導き出すのは不可能である。

これを解決するために、本システムのユーザであるネットワーク管理者が、諸々の障害に対してどのような情報が必要なのか、その情報の中でどこの部分を抽出したいのかといった手順をデータベースに登録し、登録した手順を自動的に実行するしくみが必要である。

例えば、図 6の（a）で管理者がルータ A、B、C に対して行っていた調査手順をサーバのデータベースに登録しておく。それ以降に同じ障害が検出されたときは、図 6の（b）のようにサーバに登録した手順に従って自動的にルータ A、B、C の情報を入手し、障害の報告と同時に調査した情報をネットワーク管理者に提供する。

手順をデータベースに登録することにより、ネットワーク管理者が手動で行ってきた数多くの作業を自動化することができ、原因を割り出すまでの時間の短縮につながる。

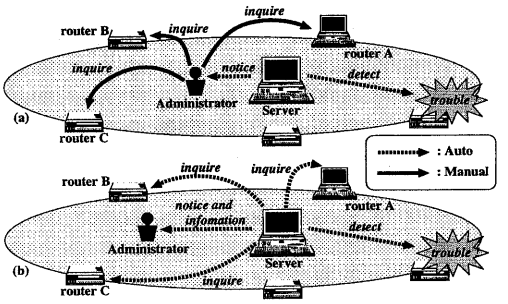

図 6: 手順を登録することによる効果

3.3 対象を指定して調査

データベースに登録した手順によって得られる情報だけでは、障害の原因を突き止めるための情報が不足している場合が考えられる。そのような場合、ネットワーク管理者は不足している情報をやはり手動で補うことになる。

例えば、原因を割り出すために “ルータ A からルータ B までの経路上のルータの情報” が必要だとすると、ネットワーク管理者は一つ一つのルータに telnet し、情報を収集しなければならない。

本システムでは、“ルータ A からルータ B までの経路上のルータの情報” といった抽象的な調査対象を指定するだけで、対象となるルータを検索し、情報を収集する。それによりネットワーク管理者の負担の軽減を図る。

3.4 ユーザインターフェース

障害報告時には電子メールを使用する。また、過去の障害履歴の検索や統計情報へのアクセスは Web ブラウザ上から行う。調査のリクエストや障害検出時の情報収集手順のデータベースへの登録なども Web ブラウザ上から行う。

Web ブラウザ上からデータにアクセスできることで、ネットワーク管理者が遠隔地にいる場合や、電子メールが受け取れない状況にあっても、管理対象 AS の状態を知ることができる。また、他の AS の管理者から自分の管理対象 AS の状態に関する問い合わせがあった場合、Web サーバを公開するだけで良い。

3.5 報告

ルータの設定情報、インターフェイスの状態に変化があった場合とピアの切断が報告された場合

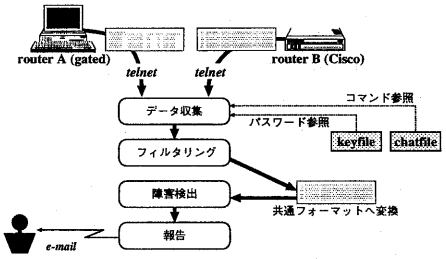

図 7: 現状の実装

Address	hostname(.wide.ad.jp)	Peering AS	Address	hostname(.wide.ad.jp)	Peering AS
133.186.1.2		65531	203.178.136.7	cisco1.sezu	2500
202.241.1.97		2513	203.178.136.9	cisco4.kyoto	2500
202.249.27.18		2500	203.178.136.10	cisco18.kyoto	2500
203.178.136.1	cisco1.sanfrancisco	2500	203.178.136.11	cisco1.hiroshima	2500
203.178.136.2	cisco2.fujisawa	2500	203.178.136.12	sun1.fukuoka	2500
203.178.136.3	cisco9.tokyo	2500	203.178.136.13	cisco3.fukuoka	2500
203.178.136.4	cisco12.tokyo	2500	203.178.136.14	sun1.sendai	2500
203.178.136.5	cisco1.tokyo	2500	203.178.136.15	sun1.sapporo	2500
203.178.136.6	cisco1.otemachi	2500	203.178.136.16	osaka	2500

表 1: cisco2.nara.wide.ad.jp とのピア

Address	hostname(.wide.ad.jp)	Peering AS	Address	hostname(.wide.ad.jp)	Peering AS
203.178.136.8	cisco2.nara	2500	202.249.47.61	jc-hk.ai3	4717
202.249.47.33	ix-id1.ai3	4717			

表 2: ix-jp1.ai3.nara.wide.ad.jp とのピア

は、状態の差分と登録した手順によって得られた情報を電子メールでネットワーク管理者に報告する。

報告された情報は電子メールと同時に Web ブラウザでアクセスできるようにすることで、電子メールが届かない環境でも報告された内容にアクセスできるようにする。

3.6 過去の障害履歴の検索

過去の障害履歴には、ネットワークに障害が発生した場合に役立つ情報が含まれている。効率よく障害履歴を検索するために、特定の経路、ピア、ルータ等といった検索したい対象を指定することによってデータベースから自動的に必要な履歴だけを抽出できるようにする。

また、データベースの内容を Web ブラウザからアクセス可能にし、遠隔地のシステムからサーバへアクセスできるようにする。

4 実装

現状では、本システムの基本的な動作である、定期的な情報収集、障害の検出、管理者への報告を実装しており、対象となるルータは、Cisco 社のルータと gated の稼働しているルータである。システム構成を図 7に示す。なお、実装言語は Perl である。

定期的な収集は cron で 30 分毎にシステムを起動し、telnet でルータの情報を収集する。このときルータのパスワードを keyfile から読み出し、ログインする。その後、情報収集のためのルータ別の操作方法を記述した chatfile を読込む。chatfile に従ってそれぞれのルータの情報を収集する。

収集したデータは共通フォーマットに変換し、対象のルータ、情報の種類、取得した時間別にファイルに保存する。

収集した情報と、その 30 分前に収集した情報との差分をとることによって障害検出を行う。また、Cisco の場合にわかる障害についても検出する。

管理者へ電子メールによって差分のみを報告している。

データベースおよび Web ブラウザによるアクセスについては現状では実装されていない。

5 運用

5 本システムを用いて、Cisco 社製のルータ 3 台と gated の稼働しているルータ 1 台を監視した。実験環境を図 8に示す。

表 1、2は、*cisco2.nara.wide.ad.jp*、および *ix-jp1.ai3.wide.ad.jp* から張られているピアを示す。特に、*ix-jp1.ai3.wide.ad.jp* は AS4717 におけるハブの役割をしているので、注意して監視する必要があった。

本研究では、4 台の各ルータから張られているピアを監視し、定期的に経路に関する情報を取得した。この取得した情報をもとに、30 分おきの経路数の変化、24 時間以内のピアが切れた回数の統計情報を取った。

また、統計情報を取るほかに、AS ループや、インターフェイスの状態の変化、ピアが切れた場合などの障害検出も行った。

これらの問題を検出した場合、問題の内容を要約したものを電子メールでネットワーク管理者に報告した。

6 評価

約 2 週間の観測を行った結果、ピアに関してはいくつかの IBGP ピアを除いてほとんど切れることなく安定していた。

ピアが切れる頻度の統計から AS2500 内の特定の IBGP ピアが不安定であることがわかった。切れる頻度の高いピアは、San Francisco、仙台、福

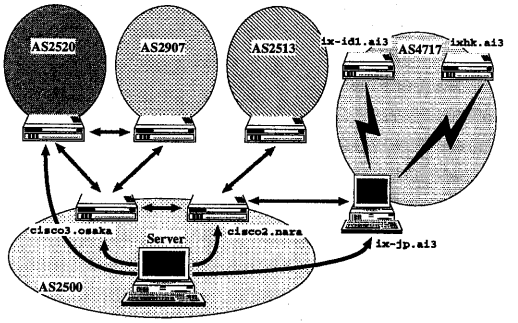

図 8: 運用した環境

岡といった地理的に遠いルータであった。この結果から、BGP の hold timer の値を調節し、ピアが切れる頻度を減らすべきであることがわかった。

AS4717 に関しては、衛星リンクになっている *ix-jp1.ai3.wide.ad.jp* ⇔ *id.ai3.wide.ad.jp*、*ix-jp1.ai3.wide.ad.jp* ⇔ *ix-hk.ai3.wide.ad.jp* の間でピアの切れる頻度が多いことが分かった。

AS ループについてもいくつか発見された。しかしながら本障害の原因がネットワーク上のトラブルなのか、ソフトウェアのバグなのかは現在のシステムでは解析できない。

設定ファイルの変更、インターフェイスの状態の変化も監視したが、これらが原因になって障害が発生するようなことはなかった。

また、ネットワーク管理を行っていると、実務のレベルで過去の情報に関する問い合わせに対して答えなくてはならない場合も生じる。このような場合に過去の情報の履歴が重要なことがわかる。

7 今後の課題

7.1 障害発生の検出

現状の設計では、本システムが障害発生を検出するタイミングは、定期的にデータを収集したときであり、障害発生時に自動的に情報収集を行う機能はない。

これを実現するためには、クライアントであるルータから、能動的に障害を通知するしくみが必要である。クライアント側に新たな実装をするのは困難であるため、SNMP トラップによるアプローチが現実的であると考える。

7.2 他のベンダーのルータへの対応

現在、情報収集が可能なルータの種類は、Cisco 社のルータと gated が稼働しているルータである。しかし、単一のベンダーでネットワークを構築するケースは稀であり、ほとんどのネットワークは複数のベンダーで構築されている。よって今後、他のベンダーのルータへ対応させることは実用レベルで重要なことであると考える。

また、4章の図 7からわかる通り、ベンダー固有の chatfile とフィルタリングのモジュールを追加するだけでよい。よって他のベンダーのルータへの対応は容易に行える。

8 まとめ

インターネットを正常に動作させるためには経路制御の管理が必要不可欠である。経路制御の管理は、一部の極めて限られたネットワーク管理者が手動で管理を行ってきたが、インターネットが急速に拡大しつつある現在、ネットワーク管理者の負荷を軽減するシステムが必要となってきている。

本研究ではこのような状況を改善するために広域ネットワークにおける経路情報の監視を行う手法を提案した。

この提案手法に基づくシステムを実装し、提案手法の有効性を運用に基づいて評価した。

現状では定期的に経路情報を収集し、統計情報をとり、以前に収集したデータとの差異があった場合に電子メールでネットワーク管理者に報告するというシンプルな実装となっている。それにも関わらず、ネットワーク管理者の労力をはるかに軽減できることがわかった。

さらに統計情報を定期的に収集するだけでも、慢性的に問題のある箇所を発見し、障害を未然に防止することが可能であることや、利用者からの問い合わせに対応する場合にも参考になることから、本システムが、実際にネットワークを管理していく上で有効な手段であることが実証された。

今後は、データベースの導入と、Web ブラウザから障害履歴の検索および統計情報へアクセスするシステムを実装し、他のベンダーのルータへの対応を行う。

参考文献

[1] Y.Rekhter and T.Li. *A Border Gateway Protocol 4*. RFC1771, March 95.