

博士研究計画：次世代インターネット経路制御に関する研究

政策・メディア研究科 後期博士課程三年 小原泰弘 yasus@sfc.wide.ad.jp

2003 年 10 月 21 日

概要

現在のインターネット経路制御の手法では、TE や QoS の機能を有したまま、規模性、耐故障性を実現することができない。MPLS などの付加機能での TE や QoS の実現は、インターネットの規模性や耐故障性を制限し、ネットワークを効率的に利用することができない。本研究では、経路制御の手法を根本から見直すことによって、トラフィックを自在に制御でき、大規模であっても頑健なインターネットを実現する。

1 問題点

1.1 TE や QoS の実現

インターネット通信では、通信データは IP パケットに分割される。この IP パケットがたどる通信経路は、中継ルータの経路表にしたがって決定される。経路表はパケットの終点 IP アドレス毎に、パケットを次に転送すべき中継ルータを保持している。

経路表はパケットの終点 IP アドレスだけを考慮して転送方向を決定してしまう。そのため、中継ルータでは基本的に、同じ終点に対する複数種類のパケットを区別することができない。この結果、パケットがどのようなサービスを実現するためのものかを考慮して経路決定を行う、QoS (Quality of Service Routing) という概念は実現できていない。

さらに、ある中継ルータで合流してしまったトラフィックを、ネットワーク回線の混雑状況によって、複数の経路に分散させることも不可能だ。ここから、ネットワーク回線の混雑状況によってトラフィックを自在に制御する、TE (Traffic Engineering) の概念も、実現が困難となっている。

TE や QoS は、MPLS (Multi-Protocol Label Switching) [1] を利用して実現しようとするのが一般的となっている。これは、トンネルなどの概念と似た、パケット交換網上に回線交換ネットワークを仮想的に構築する技術である。ある任意の一点で、条件に合うパケットを仮想回線に乗せることによって、特定のトラフィックだけをショートカット (もしくは、迂回) させ、TE や QoS を実現する。

MPLS の問題点は、条件設定が局所的であることである。条件が設定されているルータ以外は、どのようなトラフィックを仮想回線に乗せるかということを知らない。理想では、ネットワーク全体の状態情報なども加味し、このような条件設定は動的に更新されるべきである。しかし、現在利用されている MPLS の設定はいくらか静的であり、ネットワークトポロジ変更の際にはその設定が無効になってしまったり、効率的な利用ができない場合がある。

もうひとつの問題点は、プロトコルの階層化を一段

増やしていることである。インターネットでは、規模性の問題を解決するために経路制御の階層化を実現しているが、この副作用として最適な経路が選択されない可能性がある。経路制御階層の境界では、何かしらの経路制御情報の間引きが行われるからである。最適な経路選択という目的だけを見れば、全世界をフラットな経路制御ドメインにし、階層化を行わないのが理想である。ここで、MPLS は TE や QoS の実現のために、経路制御の階層を一つ増やしている。規模性の問題に対応するために必要な階層化の上に余分な階層化を増やすため、最適な経路選択のためには余分な設定が必要になる傾向が強い。その結果、ネットワークが複雑になって運用が困難になる可能性がある。また、ネットワークの状態が変わったときに自動的に設定を変更するようなシステムを構築するのが困難である。

1.2 耐故障性の向上

インターネットの耐故障性は、経路制御機構の耐故障性に大きく依存している。経路制御機構が故障すると、パケットが転送できなくなるため、インターネットの故障に直接つながる。

現在のネットワーク回線レベルでの経路制御の手法は、リンクステート型アルゴリズムを利用するものが主流となっている。リンクステート型アルゴリズムでは、回線の状態をネットワーク全体で共有し、それぞれの中継ルータが経路を独立に計算する。このとき、回線の状態情報が完全に共有できていないと、経路制御ループが起こってしまう。これは、リンクステート型アルゴリズムは、共有 (同期) されている情報に対して特定のアルゴリズムを適用することによって、経路ループを防止するような手法だからである。

またリンクステート型アルゴリズムには、回線の状態情報を共有するためのメッセージを、それぞれのルータが勝手に拒否してはいけない、という制限がある。そのため、ネットワークの中に一つでも不正なメッセージを広告するルータがいれば、ネットワーク全体の経路制御が容易に故障する。そして、これを改善することは、リンクステート型アルゴリズムである限り本質的に不可能である。

このように、リンクステート型アルゴリズムを利用する経路制御機構では、頑健なインターネットは実現できない。あるルータが不正なメッセージを投げているかどうか周りのルータがそれぞれ独立に見張り、不正であると判断された場合にはネットワークから動的に切り離すような機構の実現が必要である。

2 実現する世界と解決案

インターネット経路制御は、パケットレベルでの経路を詳細にコントロールできるべきである。TE や QoS の概念は、これにより達成される。また、故障時には、2 ~ 3 秒で次善の経路を発見できるべきであるし、明らかに異常が発生したルータは、自動的に経路制御ドメインから切り離されるべきである。これにより、機能と性能の両面から、次世代インターネット経路制御の要件を満たすことになる。

現在までの経路制御機構は、終点へ到達できれば良いという思想のもと、終点への一意な経路を計算してきた。そのため、経路表は終点のみをキーとしてネクストホップルータを指し示すものとなっている。

個々のパケットレベルでの経路制御は、経路表を検索する際の入力キーを、終点のみではなく、始点アドレスなども含めるようにしなくてはならない。インターネットルータの経路表は、パトリシアツリー [2] を利用して構築されるのが一般的である。本研究では、これを变形し、複数の入力キーに対して有効となるような経路表を提案する。実現される新しい構造を持つ経路表は、フローベースの経路制御を実現する。

次に、新しい経路制御プロトコルを設計し、実現する。設計では、通信フロー毎に経路を計算できるようにし、各経路には通信遅延やジッタ等の QoS パラメータおよび、帯域等の TE パラメータを付随する経路制御情報として組み込む。

プロトコルは、リンクステート型アルゴリズムの脆弱性を排除するため、パスベクタ型アルゴリズムを用いる。パスベクタ型アルゴリズムは、AS (Autonomous System) レベルで現在のインターネット全体の経路制御を実現している経路制御アルゴリズムであり、規模性が実証されている。加えて、経路制御メッセージのフィルタを任意の地点で実行することができ、故障したルータを経路制御ドメインから切り離すことができる。

また、インターネットの耐故障性を向上するために、切断の検知は短時間で行われなければならない。これまでの経路制御プロトコルでは、ネットワーク回線の up/down の繰り返しが経路制御全体に与える影響を比較的少なく保つために、切断の検知に 40 秒などの比較的長い間隔を要するように設計された。しかし、BGP [3] の Flap Damping [4] のアルゴリズムを用いることにより、プロトコルの切断検知間隔とは無関係に、経路制御全体に与える影響を減衰させることができる。これを用いて、短時間での切断検知を実現するように、プロトコルを設計する。

3 期待される成果

本研究により、ネットワーク管理者は、これまで実現できなかった「パケット毎の経路選択」が実現できるようになる。ネットワーク管理者が最適な経路を選択でき、また特定のトラフィックを乗せるための回線設計などが容易に実施できるようになるため、ネットワーク回線の利用効率が向上し、ユーザの通信環境向上につながる。

ネットワーク回線の故障や、ルータ故障によって、これまで経路制御機構が受けて来た影響は、本研究で設計された経路制御プロトコルにより最小限に抑えることができる。これは、故障時にもユーザの通信セッションを迅速に回復させるという強力な機能である。

本研究の実現により、最善努力型のインターネット

の上に、品質保証型のサービスを実現するための第一歩が実現される。例えば、品質保証型サービスの最も顕著な例は、IP 電話による 110 番通報などである。詳細な経路制御の設定により、いつでも、どのネットワーク回線も、帯域の 10 % を 110 番通報用の IP 電話サービスのために空けておく、といったことが実現できる。また、110 番通報中にネットワークが変化しても、2 ~ 3 秒、将来的には 1 秒以内に通話が再開できる、という状態を実現する。

これらのことから、本研究はインターネットの社会基盤としての信頼性を向上する。

参考文献

- [1] E. Rosen, A. Viswanathan, and R. Callon. *Multiprotocol Label Switching Architecture*, January 2001. RFC 3031.
- [2] D.R. Morrison. Patricia-practical algorithm to retrieve information coded in alphanumeric. *Journal of ACM*, 15(4):514-534, January 1968.
- [3] Y. Rekhter and T. Li. *A Border Gateway Protocol 4 (BGP-4)*, March 1995. RFC 1771.
- [4] C. Villamizar, R. Chandra, and R. Govindan. *BGP Route Flap Damping*, November 1998. RFC 2439.