情報処理学会論文誌　コンピューティングシステム　Vol. 2　No. 1　123–132 (Mar. 2009)

分散仮想ルータのための動的中継点制御機構

廣津　登志夫[†1]　福田　健介[†2]　栗原　聡[†3]
明石　修[†4]　菅原　俊治[†5]

データリンク層の仮想化技術である仮想 LAN（VLAN）は，ネットワーク構成の変更を容易にし設計の自由度を増すことから，ある程度の規模を持った組織内ネットワークでは広く使われている．しかし，1 つの論理的なデータリンクの範囲が広がるため，ネットワーク層でのパケット中継のトポロジとの間で不整合が発生することがある．本稿では，ネットワーク層のパケット中継点を動的に制御することにより，この不整合の問題を解消する機構について述べる．

Migration of the IP Routing Points for Distributed Virtual Routing

Toshio Hirotsu,[†1] Kensuke Fukuda,[†2]
Satoshi Kurihara,[†3] Osamu Akashi[†4]
and Toshiharu Sugawara[†5]

Virtual LAN (VLAN) is a useful technique for constructing the large scale enterprise network. Despite of the useful features given from the virtualization, it has possibility to cause some inefficient traffic transfer because of a mismatch between the topology of an overlaying logical network and one of an underlaying physical network. In this paper, we propose a distributed architecture to solve this problem, and evaluate the mechanism through the experiments and simulations.

1. はじめに

現在，企業や大学などの組織内ネットワークでは，仮想 LAN（VLAN）の技術を用いてネットワークを構築することが一般的になっている．VLAN はデータリンクレイヤの機能で物理的なネットワーク上に複数の論理ネットワークを重畳させる技術で，これを用いることによって物理的な位置に制約されずに論理ネットワークを構築できるという利点がある．実際，VLAN を用いた組織内ネットワークでは，物理的な機材と配線によるネットワークをインフラストラクチャとして事前に用意し，そのうえでの組織の変更や人の異動に応じて，論理的なネットワークを随時提供するということが行われている．

しかしながら，エンド間の通信の機能を提供するネットワーク層からはこの仮想化が透過であるため，ネットワーク層の通信を中継するルータの配置によってはパケットの転送経路に冗長が発生し，最適な経路での通信を行うことができない可能性がある．通常は，VLAN により提供されるネットワーク配置の自由度やセキュリティといった利点と秤にかけて，このような冗長転送の発生はあきらめている場合が多い．しかし，VLAN レベルの通信の状況を把握しルーティングを行う中継点（ルーティングポイント）を適切に設定することにより，この問題が軽減されたより効率の良い構成を作ることができると，VLAN のメリットと効率の両面の要求を満たすネットワークを実現することができる．本研究では，ネットワーク中の複数の機器が連携して適切なルーティングポイント制御を行う「分散仮想ルーティング」の実現を目指して，複数のレイヤにまたがるネットワークの情報を収集・解析し，ルーティングポイントの動的な制御を行う．本稿では，分散仮想ルーティングの概念について述べた後，実際のネットワークでの情報収集のモデルとそれに基づく制御法について提案する．さらに，実ネットワークデータを用いたシミュレーションを通じて本提案手法の効果を示す．

†1 豊橋技術科学大学
Toyohashi University of Technology
†2 国立情報学研究所
National Institute of Informatics
†3 大阪大学/JST CREST
Osaka University/JST CREST
†4 NTT 未来ねっと研究所
NTT Network Innovation Laboratories
†5 早稲田大学
Waseda University

2. 背　　景

初期の Ethernet によるネットワークは，同一の同軸ケーブルやハブに接続されたノードすべてが相互に直接通信可能となり，1 つの物理ネットワークで 1 つの論理ネットワークを構成していた．この時代には，ネットワークを 1 つ構築することは物理的な配線や機器の設置を意味した．これに対して，当初の VLAN は 1 つの Ethernet スイッチを内部的に複数に分離し，1 台のスイッチ中に複数の論理セグメントを構築とする技術であった（ポートベース VLAN）．これは，1 つの物理的なネットワーク装置中にブロードキャストすら届かないネットワークを複数構築できることを意味し，VLAN スイッチまでの配線を行えばソフトウェア的な設定でスイッチ内に任意のネットワークを構成できるようになった．

さらに複数の Ethernet スイッチにまたがって容易に VLAN を構築できるようにした技術がタグ VLAN（802.1Q VLAN[1),2)]）である．これは，Ethernet のヘッダを拡張してフレーム中に VLAN の識別子を付与するもので，Ethernet スイッチ上ではポートに付与されている VLAN 識別子を Ethernet フレームに埋め込んでスイッチ間で転送を行う．これにより，1 つのスイッチ間接続に複数の VLAN の通信を重畳（trunk，トランク）させることが可能となり，近くのスイッチまでの配線を確保すれば，ネットワーク中に自由に論理ネットワークを構築することができるようになった．

エンド間の通信を提供する IP 層からみると，この VLAN は従来の独立した物理機器によるセグメントとまったく同じものであり，VLAN 間の通信を実現するには，単一 Ethernet スイッチ上であっても個別にルータに配線するか 802.1Q による trunk に対応したルータに接続することが必要となる．そこで，VLAN 機能を持った L2 のスイッチに L3 のルーティング機能を持たせた L3 スイッチが利用されるようになった．多くの L3 スイッチは，L2 レベルで設定した VLAN に対してインタフェース定義をし，ルーティングプロセスを稼動させることで，スイッチ上にインタフェース定義した VLAN 間の通信を中継することが可能となり，1 つのネットワークインフラストラクチャ上でルーティングと論理セグメントの提供を同時に行うことができるようになる．

このような VLAN による仮想化で冗長なトラフィックが発生しうることは，著者らのグループが実トラフィックデータの解析によりその事実を明らかにし[3)]，また文献 4) では大規模な組織内ネットワークの機器の設定から調査した結果が示されている．著者らのグループの調査では，大学の学内ネットワークにおける主要なスイッチ間のトラフィックを解析した結果として，10%程度のノードの通信が冗長を招いていることが明らかになった．

3. 分散仮想ルーティング

ここでは VLAN ネットワーク上のトラフィック制御の問題について整理し，それを解決する手法である「分散仮想ルーティング」の概念について述べる．さらに分散仮想ルーティングを実現する手法について，その技術的課題などを検討する．

3.1 冗長トラフィックの問題

図 1 に，典型的な VLAN ネットワークの構成例を示す．これは，中心となる大型の L3 スイッチ（図中 SW-C，以下コアスイッチと称する）に，複数の L3 スイッチ（図中 SW-E1〜SW-E3，以下エッジスイッチと称する）が接続されている．このように VLAN ネットワーク中にルーティング機能を持つ機器が複数存在する場合，各論理ネットワークの中継点であるルータの配置によっては，冗長トラフィックを多発させネットワーク全体の効率を落とす可能性がある．たとえば，この図 1 の物理ネットワーク上に 3 つの VLAN（VLAN A，VLAN B，VLAN C）がコアスイッチをまたぐ形に構成されている場合を考える．ここで，SW-E1 に VLAN A のインタフェースを設定しルーティングさせる場合（以下，これをルーティングポイントの設定と呼ぶ），A-E1 と B-E1 の間の通信はエッジスイッチ内で効率的でに行われる（図 1 矢印 1）．これに対して，A-E3 と C-E2 の間の通信について考えてみる．A-E3 のノードは VLAN A のルーティングインタフェースがある SW-E1 に IP パケットを送信し，SW-E1 が VLAN C に転送したのちに，C-E2 に存在するノードに転送するために SW-E2 に向かって送られる．このため，1 つの IP データグラムが SW-C と SW-E1 の間の線を往復することになる（図 1 矢印 2）．

そこで，コアスイッチをまたぐ VLAN のルーティングポイントは SW-C に設定すると

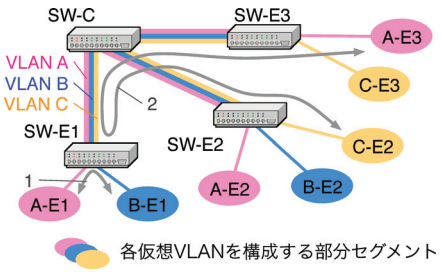

図 1　冗長トラフィック問題
Fig. 1　Redundant traffic.

いう戦略をとってみる．すると，前述の A-E3 と C-E2 の通信のような異なるエッジスイッチを通る通信に関しては冗長を回避できるが，本来ならスイッチ内で閉じていた A-E1 と B-E1 の通信のようなスイッチ内通信がコアスイッチ経由になってしまう．したがって，全体の効率を高めるためには，実際のトラフィックの流れなどの動的な要因を考慮したうえで制御することが必要となる．

以上では説明を簡略化するためにトラフィックのフローを中心にリンク帯域について説明したが，実際のネットワーク機器ではスイッチのバッファと L3 スイッチの変換テーブルについても考慮する必要がある．リンクを往復するということは，その両端のバッファを消費することになるので確率的にはトラフィックの棄却の可能性が上がることを意味する．また，多くの L3 スイッチは IP パケットの転送の高速化のために転送情報のキャッシュを持っており，1 回転送したアドレス対のパケットの転送を高速に行う仕組みになっている．これに対しては，ルーティングポイントを集中させると変換キャッシュが早く溢れることになるので，性能の低下を引き起こすことが予想される．

以上のことを考慮して仮想化ネットワークでとりうる現在の現実的な対処は，

(1)　複数のスイッチをまたぐ VLAN のルーティングポイントは SW-C に設置する，
(2)　閉じたトラフィックの多そうなスイッチにルーティングポイントを設置する，
(3)　SW-C をまたぐような広域の VLAN は構成しない，

のいずれかである．(1) はある程度の効率低下は容認してネットワーク構成の自由度と運用の容易さを得るものである．その対極が (3) で，これはネットワーク構成の自由度を放棄して，効率を確保する戦略である．(2) は，VLAN の構成当初は大まかな利用状況が予測できたとしても，使っている間に利用状況が変動したり短周期での変動が出たりした場合に手動で設定を変更する必要があり，現実的には対処が困難である．

なお，IP 層の動的ルーティングでは複数のルータからコストの低いものを動的に選択して中継させることが可能である．しかし，この例で示した各 L3 スイッチにルーティングポイントを設定し動的ルーティングを行ったとしても，適切な経路を選択することはできない．これは，すべての L3 スイッチで他のネットワークへのメトリックが同じに見えるため，ルーティングが混乱するからである．また，実際のネットワークでは，設定を単純にして安定運用させるために，各端末はデフォルトルートを使用することが一般的で，端末自身が動的ルーティングに参加するように設定することはほとんどない．これらの理由から，この問題は IP 層のルーティングだけでは解決することができない問題である．

3.2　ルータ協調による解決

このような問題の根本的な原因は，TCP/IP を中心とした現在のネットワーク技術[5] が設計・開発された当初は仮想 LAN のような技術が存在しなかったため，ネットワークの仮想化技術のことが考慮されていない点にある．当初のネットワークは先に述べたように，物理的な構造と論理的な構造が一致していた．つまり，端末やルータなどの通信ノードを「点」，通信ノードが接続されたネットワークを「線」と考えると，「線と線が点でつながる」形態であった．これに対して，現在の仮想化技術を利用したネットワークは，単一の物理ネットワーク全域に複数の論理ネットワークが「面」的に広がり，個々のデータリンク層のネットワークが層状に重なるような形となっている．つまり，「面と面が接する」形に変化しているのである．そのため，現在の仮想 LAN によるネットワークは，多数の面を接続するルーティング機能を持った複数の中継点がネットワーク中に分散して存在する，「面と面が多数の点でつながる」形態になっているといえる．

このような仮想ネットワーク環境においては，ネットワーク全体に分散した中継点（ルータ）が協調して，あたかも単一のルータのように振る舞い適切な中継を行う**分散仮想ルーティング**の考え方が必要である．この分散仮想ルーティングが目指すのは，仮想化技術が一般的になった現在のネットワークに適合したネットワーク制御・運用技術の確立および制御機構の実現であるが，その実現にはいくつかのアプローチが考えられる．

ルーティングポイントの設定支援

このアプローチでは，ネットワークトラフィックの状況から，静的なルーティングポイントの設定を予測する．ルーティングポイントの移動にともなう波及的な影響を考慮に入れようとすると，コア・エッジの全域でのトラフィック観測が必要になる．既存のネットワーク機器に対する設定だけで運用できるので，実環境への導入は容易である．

動的なルーティングポイントの移動

このアプローチでは，ネットワークのトラフィック状況の変動に対して，動的にルーティングポイントを移動する．ルーティングポイントの移動が通信に与える影響に配慮が必要であると同時に，トラフィック状況の変動に対する追従性の実現も必要となる．既存のネットワーク機器に対する動的な設定で運用可能なので，現状のネットワーク環境への適用は比較的容易である．

多点・並行的なルーティング

このアプローチでは，ネットワーク中の多点で並行的にルーティングするもので，つねに最適な転送を実現する．前述したようなルーティングの混乱が起きないようにするた

めには特殊なプロトコルやフィルタリングが必要であり，既存のネットワーク機器での実現は難しいと考えられることから，実環境への導入が最大の壁になると予想される．

本稿では，このうちの最初の 2 つに主眼をおいて，実際のネットワークのトラフィック情報をもとに動的にルーティングポイントを制御する技術について述べる．

4. 仮想化ネットワーク制御のモデル

前章までで述べたような仮想化ネットワークの効率的な制御のためには，IP 層だけでなく物理層・データリンク層も含めたマルチレイヤでの情報収集・管理と，それらの情報を用いた制御技術が必要になる．そこで，マルチレイヤに関わる情報収集のための仕組みについてまとめ，さらに観測情報を使った具体的な制御手法について検討する．

4.1 マルチレイヤの情報収集

仮想化されたネットワーク上のトラフィックの流れを把握するためには，以下のようなネットワーク情報が必要となる．

- 物理的なネットワーク機器の接続
- 論理的なスイッチトポロジ
- 仮想化された各ネットワークのトポロジ
- 仮想ネットワーク上の IP 層のトラフィックの流れ

まず，物理的なネットワーク機器の接続関係の情報については，ベンダによってはスイッチの管理情報として SNMP のプライベート MIB など経由して機器の接続構成情報を取得できるものもある．そのような機能を持つ機器であれば SNMP などを用いて，ソフトウェア的に構成することも可能になるが，一般的には管理者が環境の構成情報や設計情報として持っていることを仮定せざるをえない．本稿の実験においては，CISCO 社のプライベート MIB を経由して SNMP で接続関係の情報を取得した．

論理的なスイッチトポロジは，物理接続のうちのどれがアクティブになっているかという情報である．一般的な組織内ネットワークは，基幹付近のいくつかの機材が二重化されており，実際の利用状況を知るにはどの機器がアクティブに稼働しているかという情報を収集する必要がある．論理的なネットワークトポロジは多くの場合 Spanning Tree Protocol [1] などにより制御されており，SNMP の標準的な MIB で基本的な情報を取り出すことが可能である．一方，仮想化されたネットワークごとのトポロジは，通常スイッチ上の管理情報として保持されており，すべての機器に一様な手法を用いることは困難である．SNMP のプライベート MIB などを介して一元的に情報を収集できるものもあり，今回の実験では，

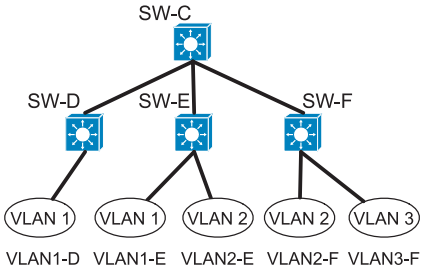

図 2　VLAN トポロジの例
Fig. 2　An example of VLAN topology.

CISCO 社のプライベート MIB により VLAN の接続情報を取得した．

仮想ネットワーク上の IP トラフィックの流れについては，ネットワーク全域の厳密なトラフィック情報を取得することは一般的には困難である．機器によっては，ポートミラーと呼ばれるスイッチ上を流れる全トラフィックを特定のポートに出力する機能を持つものがある．また，ネットワーク機器を接続する光ファイバなどの物理線のレベルで，機器間のトラフィックを直接取得する TAP と呼ばれる装置もある．さらに，ネットワークの主要な部分だけに限れば，高機能な L3 スイッチには netflow，sFlow，ipfix などのフロー情報のレベルでトラフィックを取得する仕組みを持っているものもある．本実験で用いた netflow の場合は，転送した IP フローの情報が，物理レベルの各ポートの情報が付与された形で供給される．この仮想化されたネットワーク上の IP トラフィックの情報については，単に IP 層のアドレスだけでなく，入出力インタフェースのような物理レイヤとひも付け可能な何らかの情報がなければ実際の仮想化ネットワークの制御には生かすことができない．

4.2 ルーティングポイントの制御

収集した情報からルーティングポイントを制御するための具体的な方法を検討する．まず，構成を単純化した**図 2** のような VLAN のトポロジを考える．このトポロジではコアスイッチ（SW-C）とエッジスイッチ（SW-D，SW-E，SW-F）の両方が L3 スイッチで，ルーティング機能を持っているとする．VLAN は 3 つであるが，説明の便宜上，それぞれのエッジスイッチの配下にある VLAN の断片に名前を付与した（VLAN1-D，VLAN1-E など）．

図中の VLAN3 のように，ある VLAN のセグメントが 1 つしかない場合はルーティングポイントをコア側（SW-C）に配置しようがエッジ側（SW-F）に配置しようが IP データグ

ラムの転送コストはさほど変わらない．SW-F に配置した場合は他スイッチ配下の VLAN との通信において，基幹 VLAN を通って SW-C を経由することになるので，IP 層で見たホップ数が 1 段増えるが，物理的なパケットの配送経路は同じである．ただし，SW-F 内に他の VLAN のルーティングポイントがある場合，トラフィックがエッジ内で収束できる可能性があるので，単一の VLAN セグメントの場合は基本的にエッジ側に配置することとする．この配置はネットワークの拡張などにより VLAN3 が他のスイッチに新設されるような構成変更を除けば，静的なトポロジ情報から実現することができる．

一方，複数のスイッチにまたがる VLAN の場合は，状況がもう少し複雑になる．図 2 において，VLAN1 と VLAN2 の間のルーティングについて考える．ここで，VLAN1 のルーティングポイントを SW-D に，VLAN2 のルーティングポイントを SW-F に配置すると，SW-E の配下にある VLAN1-E と VLAN2-E の通信は，SW-C との間を 3 往復することになる（VLAN1-E→SW-D→SW-F→VLAN2-E と流れる）．しかし，VLAN2-F と VLAN3-F の間の通信が頻繁であれば，全体としては VLAN2 のルーティングポイントを SW-F に配置する方が得策である．したがって，このような場合はネットワークの動的な状態に依存して決定する必要がある．これを制御するためには，すべての VLAN セグメント間の通信状況のマトリックスから最適になるような推定を行う．

ここでは，各 VLAN セグメントの使われ方の傾向はある程度の期間続くと考え，一定期間（周期）の通信状況マトリクスから次の周期の配置を決定することにする．ある VLAN のルーティングポイントをコアからあるエッジスイッチに移動するかどうかの決定は，そのエッジスイッチ配下で折り返した周期中のトラフィック量が，そのエッジスイッチ以外のすべてのエッジスイッチ配下にある当該 VLAN の部分セグメントが他のスイッチとの間で通信した量より多ければ，そのエッジスイッチに配置するという戦略をとる．たとえば，VLAN2 のルーティングポイントの設定を SW-E に移動するかどうかの判断は，VLAN2-E の SW-E 内の通信量と VLAN2-F の SW-E 以外のスイッチ向けの通信量を比べて，SW-E 内の通信量が多い場合にだけ移動する．

このような制御ポリシの実現には，各ルータでの IP 層のフロー情報が必要となる．ただし，上記のポリシをそのまま実現しようとすると，すべての L3 スイッチでフロー情報が既知でなければならない．実際にすべての L3 スイッチからフロー情報が得られるかどうかは機器構成に依存するので，エッジの L3 スイッチから情報が得られない場合には，定期的にルーティングポイントをコア側に戻して，トラフィック状況の変動を観測することが必要となる．

4.3　ルーティングポイントの移動

最後に，ルーティングポイントを実際に移動させる仕組みについて検討する．数カ月間隔の稀な設定変更であれば，定期的に手動で行うという手法も考えられるが，もっと頻繁に発生する変動に対する適応を考慮すると自動的にルーティングポイントを移動させるための仕組みが必要となる．そこで，VRRP や HSRP といったルータ冗長化のプロトコルを応用することを考える．VRRP はネットワークの出口ルータを冗長化し，障害時にスタンバイしている代替ルータに自動的に切り替える仕組みである．VRRP はデフォルトゲートウェイを設定している端末が接続されたネットワークにも対応できるように，IP 層とデータリンク層の境界で稼働する．具体的には，個々の機材に付与された実 VLAN インタフェースアドレス以外にゲートウェイアドレスとなる仮想アドレスを保持しており，そのネットワーク中の機器は中継が必要なトラフィックを仮想アドレス向けに送出する．MAC アドレスも仮想的なものを保持するため，クライアントに気付かれずにルータを切り替えることが可能となっている．ある仮想アドレスを担う機器が複数稼働しているときには，それぞれに設定された優先度情報をもとにアクティブな機器以外はスタンバイ状態となり，その VLAN についてのルーティングは 1 台のみが処理する．

現在多く見られるスター型のトポロジにおいては，VRRP のルータの組はネットワークトポロジ上の同一メトリックに水平に配置することが多い．分散仮想ルーティングで必要となる物理的なメトリックが異なる L3 スイッチ間でのルーティングポイントの移動を実現するには，コアとエッジの L3 スイッチ間で冗長化構成となるように設計すればよい．各 VLAN ではすべての L3 スイッチに実 VLAN インタフェースアドレスを付与したうえで，デフォルトゲートウェイとなる仮想インタフェースアドレスを L3 スイッチ間で冗長化して制御する．ルーティングポイントの制御は，各 VLAN ごとに各スイッチの実 VLAN インタフェースアドレスの優先度を変更することで，動的なルーティングポイントの移動が実現される．

5.　実験と評価

以上に述べたルーティングポイントの移動機構を実現する仕組みについて，実際のネットワーク機器を用いて試作するとともに，そのうえでルーティングポイントの移動が実トラフィックに与える影響を調査した．また，ルーティングポイント選択のポリシについては実環境から得た IP フローの情報をもとにシミュレーションで評価を行った．

実環境での予備実験と IP フロー情報の取得は，豊橋技術科学大学の学内ネットワークを用いて行った．このネットワークは**図 3** の模式図に示すように，二重化されたコアスイッチ

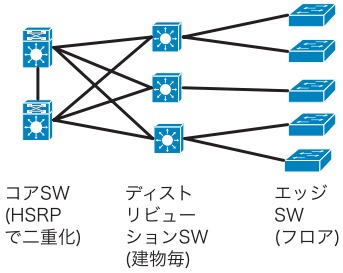

図 3　情報収集 NW
Fig. 3　Structure of the target network.

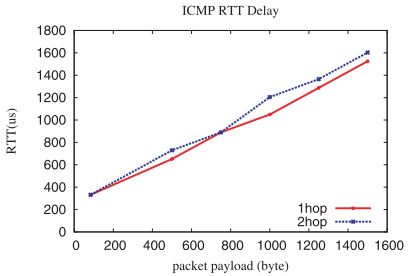

図 4　折り返しの影響の測定
Fig. 4　Effect of the routing hops.

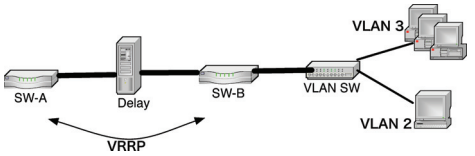

図 5　実験環境
Fig. 5　Experimental environment.

を中心にスタートポロジを構成している．中段のディストリビューションスイッチ（約 20 台）まではルーティング機能を持つが，実際にはコアスイッチで中継を行っている VLAN と，エッジスイッチ（約 100 台）の先に接続されているルータで中継を行っている VLAN が混在している．この上で約 100 の VLAN が稼働しており，コアスイッチでのみが IP フロー情報を取得する機能を持つ．コアスイッチでの IP フロー情報にはバックボーン VLAN を通る通信はすべて記録されるため，バックボーンを通らずルータ内で収束している通信を除くすべての情報が取得可能である．

5.1　折り返しの影響の測定

まず，予備実験として組織内ネットワークでの通信の遅延を測定した．ここでは，エッジスイッチの側からパケットを投入して，ディストリビューションスイッチで折り返してルーティングした 1 ホップの遅延とコアスイッチでルーティングした 2 ホップの遅延の違いを測定した．ICMP echo/reply を用い，ペイロードの大きさを 250 Byte〜1,500 Byte まで変化させた．1 試行につき 120 回測定を繰り返したうちの上位 100 回の平均を求め，これを 5 試行行った平均を**図 4** に結果を示す．縦軸の単位は μs である．測定の結果 1 ホップ増えると 10%〜20%の遅延が発生することが見てとれる．

5.2　ルーティングポイント移動の影響

次に，VRRP 対応 VLAN スイッチ（Yamaha RTX–1100）2 台と VLAN スイッチを接続した実験環境を構築した（**図 5**）．SW-A と SW-B の間で VRRP を構成し，優先度の初期値は SW-A を 192，SW-B を 128 とした．この構成で VLAN2 と VLAN3 の間で通信ができることを確認した後，SW-A の優先度を外部から 64 に変更すると，正常にルーティングポイントが SW-B に切り替わって通信は継続された．ここで SW-A の優先度を元に戻すと，再びルーティングポイントは SW-A に切り戻された．

次にこの環境で，ルーティングポイントの移動が通信に与える影響を調べた．VRRP は切替えに通常は 1 秒弱かかることと，SW-A と SA-B の間の通信線やバッファに存在するパケットが影響を及ぼすことが考えられる．まず，VLAN3 から VLAN2 に向けて iperf を用いて UDP の 90 Mbps と 80 Mbps の固定ビットレート（CBR）トラフィックを発生させ，IP パケットの識別子から見た到着タイミングを調べた．結果を**図 6** に示す．グラフでは 90 Mbps のものを上にずらしてプロットしてある．ここで，矢印のところが SA-A→SW-B→SW-A と切り替えた際のそれぞれの切替えのタイミングで，切替えの際に 500〜600 ms の通信の断が発生している．ここで特徴的なのは SW-B→SW-A に切り替えるときには断が発生しないことがあることである．今回様々な設定で 200 回以上の切替え実験を行ったが，切り戻しの際に断が発生したのは 20 回程度である．これは，スイッチ間のバッファや線路に滞留しているパケットが切替えの方向によって棄却されるためと推定される．

そこで，CBR の帯域を 50 Mbps〜90 Mbps まで変更した場合の切替え時連続棄却数の変

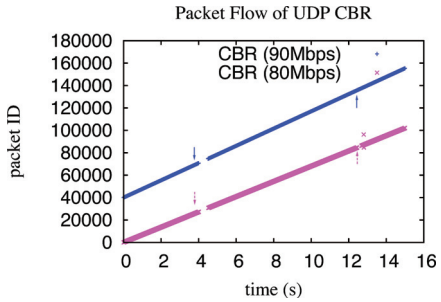

図 6　ルーティングポイント移動の影響（CBR）
Fig. 6　Effects of the migration on UDP CBR.

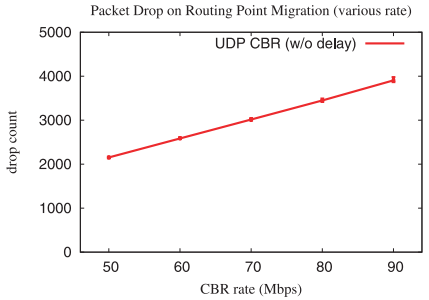

図 7　速度に対する連続棄却数の変化（CBR）
Fig. 7　Continual drops against CBR troughput.

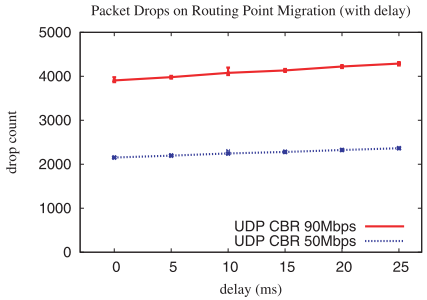

図 8　遅延に対する連続棄却数の変化（CBR）
Fig. 8　Continual drops against line delays.

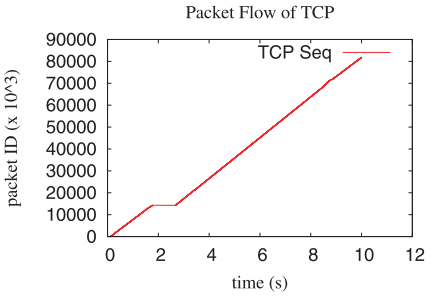

図 9　ルーティングポイント移動の影響（TCP）
Fig. 9　Effects of the migration on TCP.

化と，スイッチ間に 5 ms〜25 ms の遅延[*1]（図 5 中の Delay において，FreeBSD 6.3-R の dummynet で実現）を発生させた場合の連続棄却数の変化を見た．**図 7** と**図 8** に結果を示す．左が速度を変化させたもので，通信帯域が大きいほど連続棄却数が大きい．一方，右の遅延に対するグラフは，CBR を 90 Mbps と 50 Mbps の両方で調べたが，増加はするがわずかであった．以上のことから，切替えを行うスイッチ間のバッファに滞留するパケットが棄却されている可能性が高いものと考えられる．

次に，この棄却の TCP に対する影響を調べた．iperf を用い TCP セッションを張ったところで，ルーティングポイントの切替えを行った．受信側に到着するパケットのシーケンス番号の変化を追った結果を**図 9** に示す．結果は，ルーティングポイント切替えの 1 秒弱の間はパケットが届かず，シーケンス番号が延びていないが，切替え後は順調にウィンドウが回復し元の帯域に戻っている．切り戻しの際には影響は現れていない．以上の結果より，TCP セッションが切断することはないが帯域は一時的に影響を受けることが分かった．

5.3　ルーティングポイント選択ポリシの評価

最後にルーティングポイント選択ポリシの評価を行った．4.2 節に述べたポリシに従って，ルーティングポイントの移動を行った場合に，収集したデータから得られるトラフィック量，すなわちコアでルーティングを行っている場合のトラフィック量がどの程度削減されるかを

[*1] 折り返しで効くのでノード間の片方向で遅延は 2 倍になる．

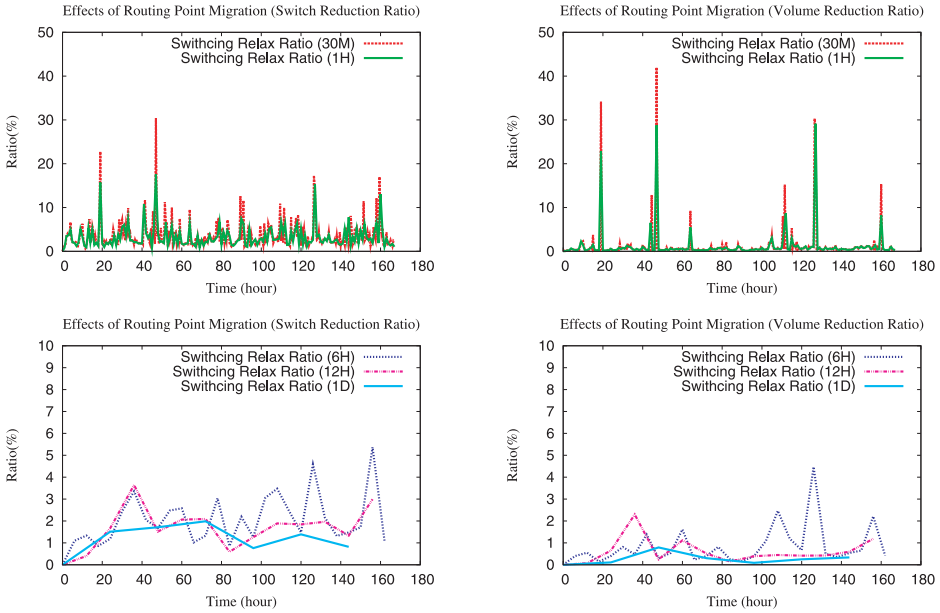

図 10　ルーティングポイント移動による削減効果
Fig. 10　Traffic reduction ratio.

シミュレーションにより調べた．ここではすべての L3 スイッチから IP フロー情報が得られることを仮定して，ルーティングポイントを移動した後の IP フロー情報についても，実環境から取得した情報をそのまま用いた．netflow のような IP フロー情報を供給できる機器は，通常外部の管理サーバにそれらの情報を蓄積する形になっているので，組織内ネットワークの運用を考えた場合にはこの仮定でも妥当である．トラフィックデータとしては，ネットワークのコアスイッチで 2008 年 6 月のある 7 日間に収集した netflow のデータから，IP フローと物理インタフェースの対応を抽出して用いた．このデータを事前に 1 時間単位で集計したところ，折り返しの冗長トラフィックが最大時で 40%，平均で 4%程度含まれていた．

今回のシミュレーションでは，一定時間情報を収集しその結果から次の周期のルーティング配置を変更するという手法をとり，変更周期として 30 分，1 時間，6 時間，12 時間，24 時間をパラメータとした．シミュレーションにおいては，IP フローの IP アドレスとインタフェース情報から，各インタフェースの先にある VLAN セグメント間の通信状態マトリックスを 1 周期の間蓄積し，周期の切替わり時にマトリックス情報から折り返しトラフィックの削減を見込むことができるインタフェースの先にルーティングポイントを移動するという処理を行う．**図 10** の上に短周期（30 分，1 時間）の結果を，下に長周期（6 時間，12 時間，24 時間）の結果を示す．このグラフは周期中の全トラフィックに対して，冗長度が削減されて減少したトラフィックの割合を示したもので，左がパケット数で見たもの，右がトラフィック量で見たものである．それぞれの周期がすべてプロットされているが，バースト的な大きなピークがあるのは 30 分と 1 時間の周期にした場合で，それ以外は下の方に緩やかな変化をしている．双方とも最大で 30%程度の削減効果が見てとれる．

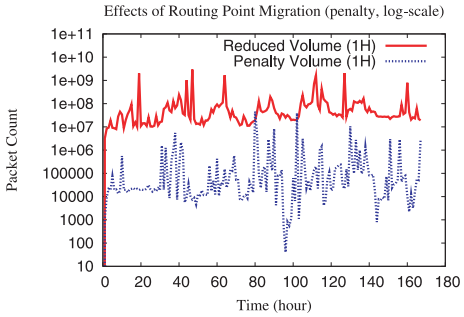

図 11　ルーティングポイント移動におけるペナルティ
Fig. 11　Penalty on the routing point migration.

一方，ルーティングポイントの移動ポリシでは VLAN 単位で移動を決めているために，他の VLAN に移動の余波として元の状態より多くのホップが生じる場合がある．この状況をペナルティと考え，1 時間周期で切り替えた際のペナルティの推移を**図 11** に示す．ここでは，冗長トラフィックの削減量を実線で，ペナルティにより増えたトラフィック量を点線で示している．グラフからほとんどの場合でペナルティの 10〜1,000 倍の削減効果が見られることが分かる．

6. 考　　察

実験とシミュレーションを通じて，分散仮想ルータとして連携して情報基盤を効率化することの効果と意義が確認できた．ルーティングポイントの設定については直前の情報だけを使って推定し，最大 30%程度の冗長の削減を実現できることが分かった．一方，ルーティ

ングポイント変更の周期は今回の手法では 30 分から 1 時間くらいの場合に大きな削減効果が得られているが，これについては推定の手法によって変わってくるものと思われる．

本稿で提案した手法では，VRRP を障害時の回復と通信の制御の両方に使うことになる．そのため本来の冗長化との競合が起きないかという懸念が生じるが，現在までに検討したところではプライオリティが十分な段数があれば問題ないと考えられる．ごく単純なポリシ付けとしては，

(1)　ルーティングポイントを移動して設定したルーティングポイント
(2)　通常のマスタルータ
(3)　通常のバックアップルータ
(4)　ルーティングポイントを移動が設定されていないルーティングポイント

の順にプライオリティを設定しておけば，ルーティングポイントの移動とバックアップが問題なく併用できる．

最後に，VRRP を用いたルーティングポイントの切替えで生じることのある数百 ms 程度の断の影響について考える．実験の結果からは，線路の遅延を増やしても CBR にさほど影響を与えないことから，機器類のバッファの影響が大きいのではないかと推察される．一般的な共用ネットワークのトラフィックは，組織の基幹近辺では多数のトラフィックが集約されるため，統計的多重効果により 1 つのフローに与える影響は小さくなる．一方で，広帯域のトラフィックに対する要求がある場合は，1 秒弱の遮断による影響は無視できない．切り戻しの際には多くの場合に影響が出ないことから，切替え時のバッファ処理などで改善できる可能性はあるとは考えられるが，これは今後の課題である．

7. ま と め

本稿では VLAN ネットワーク環境で起きうる冗長トラフックの動的制御について述べた．実際のネットワークにおいて冗長の影響を測定するとともに，VRRP を応用したルーティングポイント移送機構の実験を行い移送の影響を示した．この結果，性能の改善が見込めることと実際のネットワーク機器での運用が可能であることが明らかになった．今後はさらに大規模な実験環境を用意して複雑なトポロジや遅延のある環境での実験を進める予定である．

謝辞　この研究は科学研究費補助金基盤研究（C）「分散仮想マルチレイヤルーティング技術の研究」の支援を受けている．

参 考 文 献

1) IEEE: IEEE Standard 802.1D-1998, Information technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Common specifications (1998).
2) IEEE: IEEE Standard 802.1Q-1998, IEEE Standards for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks (1998).
3) 廣津登志夫，福田健介，菅原俊治：VLAN 環境における分散仮想ルーティングに関する一考察，情報処理学会 OS 研究会研究報告，Vol.2006, No.15 (2006–OS–101), pp.17–24 (2006).
4) Garimella, P., Sung, Y.-W., Zhang, N. and Rao, S.: Characterizing VLAN usage in an Operational Campus Network, *ACM SIGCOMM INM'07*, pp.305–306 (2007).
5) Tanenbaum, A.S.: *Computer Network*, *4th edition*, Pearson Education (2003).

(平成 20 年 7 月 23 日受付)
(平成 20 年 11 月 4 日採録)

廣津登志夫（正会員）

1995 年慶應義塾大学大学院理工学研究科計算機科学専攻博士課程修了．同年日本電信電話株式会社入社．2004 年より豊橋技術科学大学情報工学系准教授．分散システム，OS，ネットワーク，ユビキタスシステム等の研究に従事．博士（工学）．日本ソフトウェア科学会，ACM，IEEE-CS 各会員．

福田　健介

1999 年慶應義塾大学大学院理工学研究科計算機科学専攻後期博士課程修了（博士（工学））．同年日本電信電話株式会社入社以来，未来ねっと研究所に所属．この間 2002 年ボストン大学訪問研究員．2006 年より国立情報学研究所アーキテクチャ科学研究系准教授．2008 年より科学技術振興機構さきがけ研究員（兼任）．学術情報ネットワーク，インターネットおよびネットワーク科学に関する研究に従事．

栗原　　聡（正会員）

1992 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．同年日本電信電話株式会社入社．基礎研究所を経て未来ねっと研究所に所属．1998 年から慶應義塾大学大学院政策・メディア研究科専任講師（有期）．現在同大学環境情報学部非常勤講師．2004 年から大阪大学産業科学研究所知能システム科学研究部門准教授（同大学大学院情報科学研究科情報数理学専攻准教授兼務）．マルチエージェント，ネットワーク科学等の研究に従事．著書『社会基盤としての情報通信』（共立出版，共著）．翻訳『スモールワールド』（東京電機大学出版局，共訳）．編集『Emergent Intelligence of Networked Agents』（Springer in Computational Intelligence Series）等．博士（工学）．人工知能学会，日本ソフトウェア科学会，ESHIA 各会員．

明石　　修（正会員）

1987 年東京工業大学理学部情報科学科卒業．1989 年同大学院理工学研究科情報科学専攻修士課程修了．同年日本電信電話株式会社入社．以来，分散システム，ネットワークアーキテクチャ，マルチエージェントシステム等の研究に従事．現在，NTT 未来ねっと研究所主幹研究員（特別研究員）．博士（理学）．ACM，日本ソフトウェア科学会各会員．

菅原　俊治（正会員）

1982 年早稲田大学大学院理工学研究科数学専攻修士課程修了．同年日本電信電話公社入社（武蔵野電気通信研究所基礎研究部）．以来，知識表現，学習，分散人工知能，マルチエージェントシステム，インターネット等の研究に従事．1992〜1993 年マサチューセッツ大学アマースト校客員研究員．現在，早稲田大学基幹理工学部情報理工学科教授．博士（工学）．日本ソフトウェア科学会，電子情報通信学会，人工知能学会，AAAI，ISOC，IEEE，ACM 各会員．