マルチメディア通信と分散処理 111−35
コンピュータセキュリティ 20−35
（２００３．２．２８）

ゲートウェイ間の冗長性と負荷分散に関する提案

清水崇史† 小原泰弘††
南政樹† 村井純†

本研究では, インターネットにおける LAN(Local Area Network) の外部への接続性の冗長化と負荷分散に関する提案を行う.

LAN からの外部接続地点は通常一箇所で, その回線やゲートウェイとなるルータの障害によって下流のノード群の外部接続性が断たれてしまう場合がある. このように一点での障害によりサービス全体が提供できなくなるという問題を single point of failure と呼ぶ. この問題はインターネットにおいて一般的であるが, 複数の外部接続性を用意し, その間での冗長性が保証されるべきである. また, 理想では回線を有効活用するために負荷分散が行われるべきである.

本研究では, ある一セグメントにおける外部接続性の冗長化, 負荷分散の実現, さらにそのユーザへの透過性および規模性について考察する. また, 既存の技術の問題点を述べ, これらの要素を含む独自のシステムを設計する.

Proposal of redundancy and load balancing between gateways

Takashi Shimizu,† Yasuhiro Ohara,†† Masaki Minami† and Jun Murai†

In our research, we focus on redundancy of connectivity and load sharing of traffic from a LAN to outside networks.

Typically, a LAN has only a single outgoing connection, where its component router and/or circuit failure causes the entire LAN to disconnect. Such points where single failure causing the entire service to terminate is referred to as "single point of failure". Although the single point of failure is very common in today's Internet, the redundancy of outgoing connections should be ensured by provisioning multiple connections to the LAN.

In our research, we study the redundancy of outgoing connectivities and the traffic load sharing, together with its transparency and scalability. Furthermore, we design a new system with consideration of the potential problems in existing researches.

1. 研究背景

一般家庭で LAN を構築しており, 外部への接続性として二種類以上の ISP(インターネットサービスプロバイダー) へと接続している環境を想定する (図 1). このような環境において LAN は単一のアドレスブロック, 単一のセグメントとして動作し,LAN 内から外部への回線の違いを意識することなくインターネットへの接続を提供することが求められる. しかし, 現状では外部への接続性が 2 本以上存在しても, 動的経路制御プロトコルが動作していなければ, 複数の外部接続性を有効活用することは出来ない.

2. 実現したい環境

外部への接続性が複数ある状況において, そのことをエンドノードが意識することなく有効に利用できる状況が理想である. 有効に利用するということは, 具体的には以下を指す.

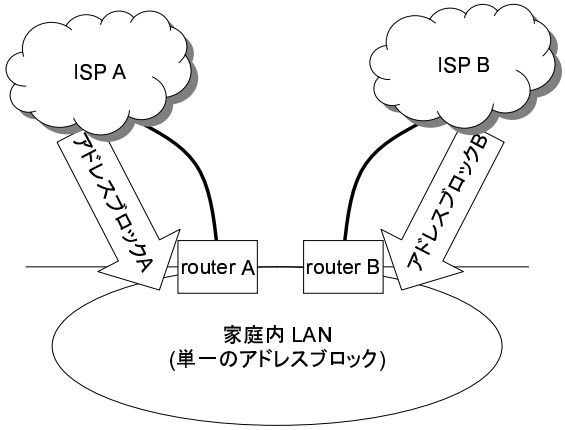

図 1 2 つの ISP に接続性のある家庭

[1] 冗長性
複数の外部接続性があるとき, その片方の回線やルータに障害が起きたとしても, 他の接続性が代わりに利用されることで, その外部接続性が継続されなければならない.

[2] 負荷分散

† 慶應義塾大学 環境情報学部
Faculty of Environmental Information, Keio University
†† 慶應義塾大学 政策・メディア研究科
Graduate School of Media and Governance, Keio University

複数の外部接続性があるとき, トラフィックが流れることによって掛かる負荷は各回線に分散される必要がある.

[3] 外部接続性の透過性
複数の外部接続性は, ユーザに透過であることが求められる. 複数ある外部接続インターフェイスを一つの仮想的なインターフェイスと見せ, エンドノードからはどの回線を利用しているかを意識させないべきである.

[4] 規模対応性
LAN の規模が拡大したり回線数が増加したときに, ネットワーク構成やシステムに極端な変化が起こらないことが求められる.

本研究では, 示したような項目を実現することを目的とする. その中でも冗長性, 負荷分散に特に重点を置く. 本研究で目的とする環境を図 2 に示す.

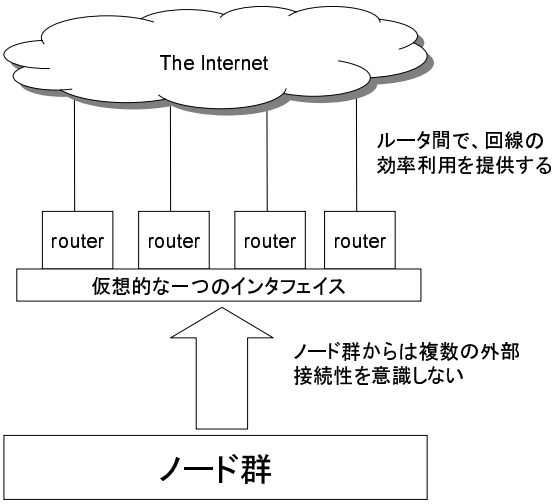

図 2 実現したい環境

3. 関 連 研 究

本節では, 本研究と関連のある研究, プロトコルの紹介を行う.

3.1 Virtual Router Redudancy Protocol[1)]

動的経路制御が利用できない環境で複数のルータの冗長化を行うためのプロトコルとして VRRP がある.VRRP に対応した複数のルータを一つのグループに所属させ, 通常はそのうち一つのルータが通信を行うが, そのルータが障害を起こした時には, 同グループに属するルータが自動的に通信を受け継ぐ. 一つのグループの ID を VRID という.

複数台のルータでの負荷分散を行うためには図 3 のように複数のグループを作成する.LAN 上に複数の VRRP グループが存在するとき, お互いは全く独立に動作する. そして一つのルータを複数の VRRP グループに所属させる. 図 3 の例だとルータ R1 は VRRP グループ 1 と 2 に所属しグループ 1 のマスタルータ, グループ 2 のバックアップルータとして設定する. このとき, エンドノードの所属させるグループを分け, デフォルトゲートウエイを各グループのマスタルータにする. このような設定をすることで, 複数台の VRRP ルータは冗長化を兼ねつつ負荷分散を行う事が出来る.

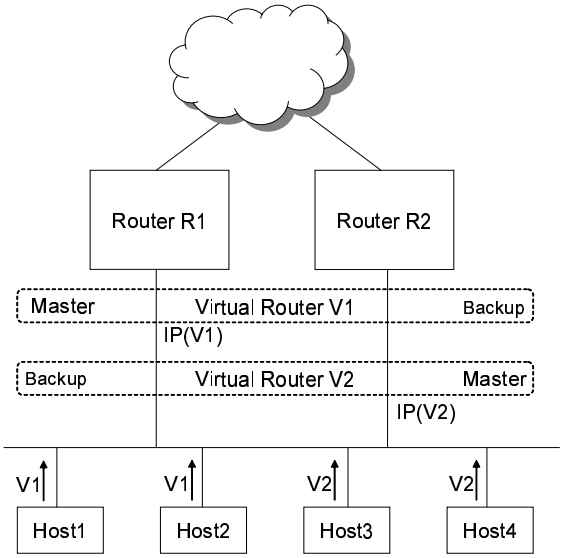

図 3 VRRP ルータ間の負荷分散

3.2 VRRP の類似機構

上記の VRRP は標準化されたホットスタンバイのプロトコルであるが, それ以外にベンダ独自仕様の技術がある. 各ベンダごとにまとめると以下のようになる.

- Cisco → HSRP[2)] (Hot Standby Routing Protocol)
- Foundry → FSRP[3)] (Foundry Standby Router Protocol)
- eXtreme → ESRP[4)] (Extreme Standby Router Protocol)

各ベンダの基本機能は VRRP と同様であるが, 互換性はない.
このなかで,ESRP はルーティングの冗長化をすると同時に, スパニングツリープロトコルよりも高速なレイヤ 2 での冗長化を提供する.

3.3 Gateway Load Balancing Protocol

株式会社シスコシステムズが提供する Gateway Load Balancing Protocol は冗長化と負荷分散を同時に実現するプロトコルである. 図 4 のようなネットワーク構成を組み, ルータの手前にあるレイヤ 2 スイッチによって上流ルータへの負荷分散を行う.

3.4 マルチホーム技術

本研究では, インターネットへの接続において複数の ISP との接続を行うことをマルチホームと呼ぶ. マルチホームな環境での通信をするために, 以下に発生する問題を解決する必要がある. この環境において発生する問題として復路の経路問題がある. インターネットにおける経路は往路と復路で別々に計算される. 本システムで考慮する経路は往路, つまり外向きの経路についてのみである. ルータと回線についてのみの冗長化や負荷分散を考え, ルータの上流の ISP が同じである場合, 送信元となるアドレスがその ISP から配られたアドレスとなる. そのため, 復路の経路は ISP 内の経路制御に依存している.

図 5 にマルチホーム環境の例を示す.

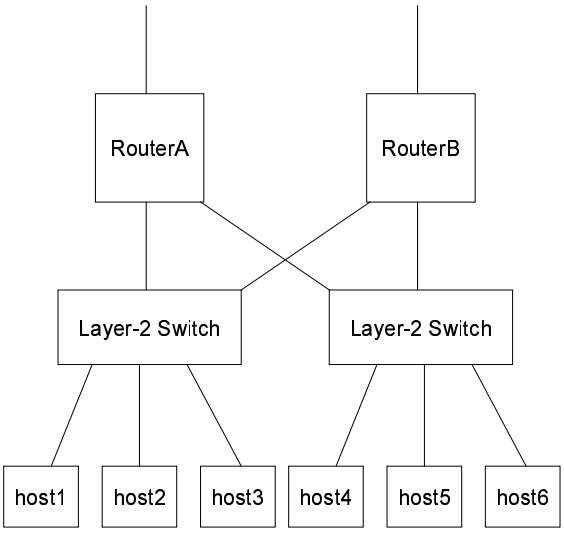

図 4 Gateway Load Balancing Protocol

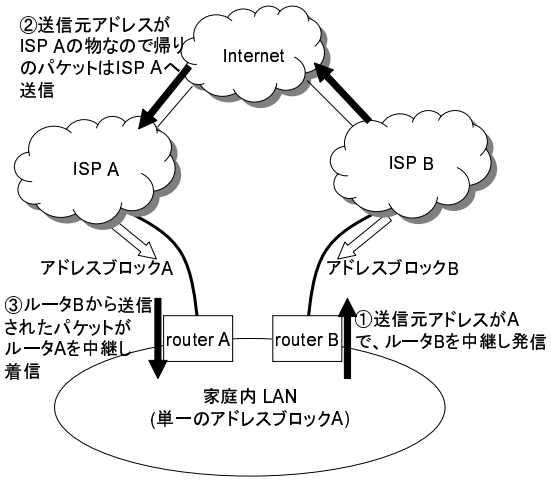

図 5 マルチホームの例

このような場合, ゲートウェイ B から発信されるパケットは, 送信元のアドレスが ISP A から配られたものであれば復路は ISP A を通ることになる.

平常時, 行きの経路の負荷分散を行っても, 帰りは全て片方の経路を通ることになる. また障害時には冗長性を失うことになってしまう. 例えば ISP A に障害が起きたとき, 外向けの経路は冗長化されているので全てのパケットはルータ B から送信される. しかし, 送信アドレスが A のままだと帰りのパケットは ISP A に向けて送信されるので, 障害のある ISP A で通信が止まってしまう.

このようなマルチホームの問題について IETF(Internet Engineering Task Force) の”Site Multihoming in IPv6” ワーキンググループ[5] ではルータによる発信元アドレスの書き換え (NAT) による解決方法が提案されている.

図 6 にアドレスの書き換えの図を示す.

ISP A と ISP B 二つの組織からの回線を持つとき,LAN 内部では片方の ISP から配られたアドレスブロックを使う. そしてアドレスを使わない側の ISP の回線から送信する場合はルータでアドレスの書き換えを行い送信する. また, 帰りの経路でもアドレスを逆に書き換える. この方法を利用

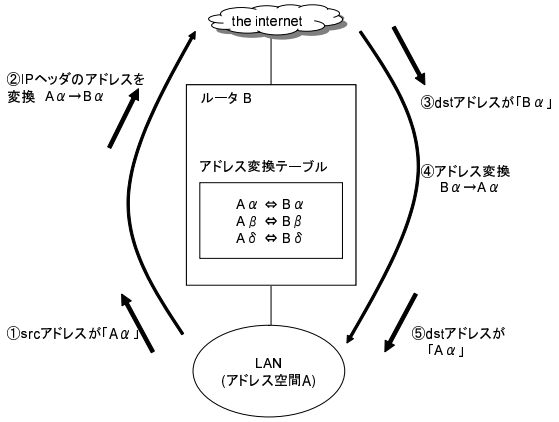

図 6 アドレスの書き換え

することで, 送信元アドレスに依存することなく往路と復路で同じ回線を利用することが出来る.

4. 問 題 点

本節では, 関連研究を踏まえた上でのその問題点を述べる.

4.1 冗 長 性

VRRP で負荷分散を行う場合, 次のような問題点が起こる.

VRRP ルータに起こる可能性のある障害は大きく次の 5 種類に分ける事が出来る.

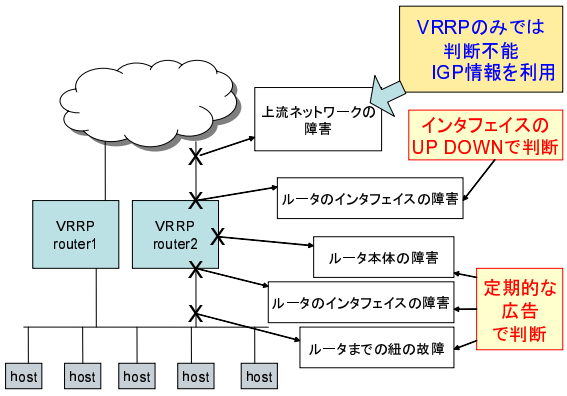

図 7 VRRP における障害発生箇所

[1] 下流ネットワークからルータまでの回線の障害

[2] 下流ネットワークへのインターフェイスの障害

[3] ルータ本体の障害

[4] 上流ネットワークへのインターフェイスの障害

[5] 上流ネットワークの障害

マスタールータから下流ネットワークへの定期的な広告を利用してバックアップルータはマスタールータの生存確認を行う. 上の 1,2 の場合は広告がバックアップルータへ届かないためマスタールータに障害が起こったことをバックアップルータが判断できる.

3 のルータ本体の障害は本体が物理的に故障した場合は広告が届かないため障害検知可能である. しかし経路エント

リの欠落等のソフトウエア的な障害の場合は, 予め障害の可能性を予測しておき, 障害判断のアルゴリズムを実装する必要がある.

4 の上流ネットワークへのインターフェイスの障害はインターフェイスの UP,DOWN を監視することで,DOWN したときに下流へ広告を停止することを実現できる.

5 のネットワークの障害に関しては VRRP のみでは障害検知をすることが出来ない. ルータの上流でダイナミックルーティングプロトコルによる径路の交換が行われている場合,VRRP ルータはその経路交換情報から上流ネットワークの状況を解析し障害を見付けなければならない.

4.2 負 荷 分 散

VRRP での負荷分散を行うにはエンドノード毎に別々の VRID を設定する必要が生じる. また, ホストベースでの負荷分散しか行う事が出来ず, フローベースやプロトコルベースでの負荷分散は行うことが出来ない.

フローベースやプロトコルベースでの制御が出来ると,QoS(Quality of Service) の考えを用いたより粒度の細かい負荷分散が行われる.

4.3 規模対応性

GLBP の問題点は物理的制約が厳しいということである. ルータ数が増えることにより GLBP に対応したレイヤ 2 スイッチの数を増やす必要がある. また, 完全な冗長性を実現するためにルータとスイッチ間をフルメッシュに接続する必要がある. 例えば外への接続性が 4 本あり, それに伴ってルータ、スイッチをそれぞれ 4 台用意する. 各ルータ, スイッチ間をそれぞれ配線すると各機器からは 4 本の線が延びていることになる. ルータは上流へのインターフェイスも持っていることから合わせて 5 本のインターフェイスを持つことになる. ルータ数の増加に伴って各ルータのインターフェイス数が増加する. このシステムは大規模なネットワークを構築する上での規模対応性に適しているとは言えない.

5. 解決へのアプローチ

本節では, 前述の関連研究を踏まえた上で, 負荷分散と冗長性を同時に実現する方法として以下のようなモデルを提案する.

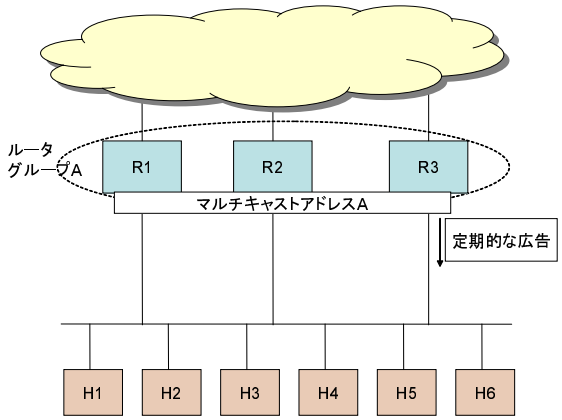

図 8 新しいモデル

5.1 システム概要

- エンドノードから見て, デフォルトゲートウェイとなる複数のルータインターフェイスを仮想的に一つのインターフェイスに見せる.
- エンドノードからのデフォルトゲートウェイをその仮想的なインターフェイスに向ける.
- 全てのルータは下流ノードからの全外向けパケットを受信する.
- ルータ間では広告を送り状態を確認しあう. 広告の内容は, 生存を確認する意味の他に, お互いにネットワーク特性を知らせる意味を持たせる.
- ネットワーク特性には, 帯域・遅延・ジッタ (遅延のゆらぎ)・パケット損失率などが含まれる.
- このお互いのネットワーク特性から最適な各ルータの転送ポリシを決定する.
- この転送ポリシに当てはまる一台のルータのみが受信したパケットを転送する.

5.2 仮想的なインターフェイス

ルータ群のインターフェイスを一つの仮想インターフェイスに見せるときに次の二通りの方法について考察した.

[1] 単一 MAC アドレスおよび IP アドレスの共有
複数台あるルータにその中の一台と同じ IP アドレスと MAC アドレスをつける. これによってエンドノードからの外向けのパケットは全てのルータで受けとられる. しかし, この方法は各アドレスは同じネットワークでは唯一無二であるという原則を破っていることになる. また, 途中にスイッチが挟まれるネットワークだとスイッチの arp テーブルで同じ MAC アドレス,IP アドレスが複数存在するという矛盾が生じてしまう. そのため, この方法が動作するかはネットワーク設定依存になってしまう.

[2] マルチキャストの利用
全ルータがエンドノード側のインターフェイスにおいて, あるマルチキャストアドレスに JOIN する. エンドノードがデフォルトゲートウェイをそのマルチキャストアドレスに向けることで, 全てのルータは外向けのパケットを受信することが出来る. このとき, ネットワーク途中のスイッチがマルチキャストに対応している必要がある.

上の二通りの方法のうち, マルチキャストを利用するシステムを採用する.

5.3 ポリシのタイプ

各ルータがどのパケットを転送するかというポリシのタイプしては以下のようなものが考えられる.

[1] ホストベース
ホストごとに通すルータを選択する. 主に帯域によってトラフィックを分散するときにその帯域にあった量のホスト数を割り当てる.Ether header の src アドレスフィールドを参照する.

[2] プロトコルベース
プロトコルの種類でフォワーディングするルータを分けるポリシを決定する. 主に各回線のネットワーク特性が異なる場合に用いることができる. 例えば帯域は広い

が遅延が大きい衛星回線と帯域は狭いが遅延は少ない地上回線がある場合,HTTP のような遅延のによる影響の少ないプロトコルは衛星回線を使い,SSH のような帯域による影響は少ないが遅延による影響の大きいプロトコルは地上回線を使う等のポリシを設定する. IPv4 環境ならば IP ヘッダのプロトコルフィールドで判断する. ただ,IPv6 の next header フィールドではパケットの中身が具体的には判別不可能なので実際のプロトコルの種類を調べるためにはレイヤ 4 以上のフィールドの中身を参照することになる. これはルータがレイヤ 3 の機器だという前提にたいして, レイヤバイオレーションになる.

[3] フローベース
発信元, 送信先そしてプロトコルの組によるフローごとにポリシを決定する. 一つのフローが同じ経路を通ることが保証されると, 到達順序の不整合の問題が起こらない.

5.4 ネットワーク特性情報の入手方法

ネットワーク特性情報の入手方法は大きく二通りある.

[1] あらかじめ管理者が手動で入力する
この方法は管理者が回線の特性を調査し静的に設定する.

[2] ルータが上流回線を監視し動的に現在の状況を把握する
この方法を用いるためには特別なネットワーク監視方法が必要になる. たとえば, 帯域を測定する方法としては Packet Pair 方式[6], Packet Train 方式[6] などがある.

5.5 上流ネットワークのルーティングプロトコルとの連係

耐故障性を高めるためには各経路の障害検知精度を高める必要がある. ルータ周りの障害の中で上流の障害を検知するためには上流のルーティングプロトコルから情報を入手する必要がある.

6. 課　　題

本システムを実現したときに発生する可能性のある問題点を述べる.

6.1 パケットロス, パケットの重複

本システムは動的にポリシを変更するためポリシ変更時にルータ間でのポリシに矛盾が生じ, パケットロスやパケットの重複が起こってしまう恐れがある.

6.2 ルータのコスト増加

従来のルータにかかるコストは大きくわけて以下の二つである.

- パケットの転送 (受信→経路表参照→転送)
- ダイナミックルーティングプロトコルによる動的経路表の交換

それに対して, 本システム実現に伴うコストは次のようになる.

- パケットの転送 (受信→自ルータが転送すべきかポリシ参照→経路表参照→転送)
- 上流回線の状態を監視
- ルータ間の広告の送受信
- ダイナミックルーティングプロトコルによる動的経路表の交換

この中でも毎回の経路選択にかかるポリシ参照がもっとも大きな負荷になると考えられる. 経路表のデータ構造には radix tree が用いられており経路検索の高速化が計られている. 同様にポリシデータも高速なデータ構造に入力されるべきである.

6.3 コントロールトラフィック

各ルータ間での定期的な広告を送信するため, 下流の LAN でのトラフィックが増加する. しかし, ルータ間の広告はマルチキャストを利用することで VRRP と比較しても極端なトラフィックの増加にはつながらない.

6.4 問題特定の複雑化

現在, ネットワークの到達性を調べるときに ping というコマンドが通常使用される.ping は icmp echo を送信し echo reply の有無で到達性を調べる. 本システムでプロトコルベースの負荷分散が行われているとき, 例えば 3 本の接続性があるとき,icmp を転送するルータ A 回線 A,TCP を転送するルータ B 回線 B,UDP を転送するルータ C 回線 C のように分けたとする. この場合, 回線 B の上流に障害が発生し, バックアップに失敗すると TCP での通信を行うことが出来なくなる. しかし, 通常のように ping を用いても回線 B の到達性を調べることはできない. このような場合に備え,TCP や UDP を用いた ping の代替品を用意するべきである.

7. 結　　論

LAN における外部接続性は single point of failure になってはならない. またユーザからみて複数の接続性を透過的に利用できるべきである. 本論文では,LAN から外部への接続性の冗長性, 負荷分散を効率良く実現するシステムの提案を行った.

参 考 文 献

1) S.Knight, et.al *Virtual Router Redundancy ProtoclApril 1998*RFC2338
2) Cisco Systems*Hot Standby Routing Protocol*
http://www.cisco.com/
3) Foundry*Foundry Standby Router Protocol*
http://www.foundrynet.com/
4) Extreme Networks*Extreme Standby Router Protocol*
http://extremenetworks.com/
5) B. Black et.al*Site Multihomeing in IPv6*
http://www.ietf.org/html.charters/multi6-charter.html
6) Constantions Dovrolis and Parameswaran Ramanathan and David Moore
What Do Packet Dispersion Techniques Measure?2001