

Design of Scalable Inter-Domain IP Multicast Architecture

大森, 幹之
九州大学大学院システム情報科学研究科

岡村, 耕二
九州大学情報基盤センター

荒木, 啓二郎
九州大学大学院システム情報科学研究院

<http://hdl.handle.net/2324/6334>

出版情報 : 2000-11-01
バージョン :
権利関係 :



スケーラビリティを考慮した IP マルチキャストアーキテクチャの設計

大森幹之[†] 岡村耕二[‡] 荒木啓二郎^{*}

[†] 九州大学大学院システム情報科学研究科

[‡] 九州大学情報基盤センター

^{*} 九州大学大学院システム情報科学研究院

本論文では、経路情報量を削減することによってインターネット全体に適應できる IP マルチキャストの経路制御を実現することを目的とし、新しい IP マルチキャストアーキテクチャを提案する。提案するアーキテクチャでは、まず、効率的にアドレスを集約できるようにアドレスを動的に割り当てることを考え、経路情報量を削減することを実現した。次に、より柔軟にアドレスを割り当てるために、マルチキャストアドレスのアドレス空間を分離することを考え、マルチキャストアドレスを VMA (Virtual Multicast Address) と MAR (Multicast Address for Routing) の 2 種類に分割した。また、提案したアーキテクチャによって削減することのできる経路情報量を示し、現在の IP マルチキャストのアーキテクチャとの相互運用について考察した。

Design of Scalable Inter-Domain IP Multicast Architecture

Motoyuki OHMORI[†] Koji OKAMURA[‡] Keijiro ARAKI[†]

[†] ... Graduate School of Information Science and Communication Engineering, Kyushu University

[‡] ... Computing and Communications Center, Kyushu University

In this paper, we design the scalable inter-domain IP multicast architecture and achieve IP multicast routing over the whole Internet. In our architecture, we propose the dynamical address allocation in order to aggregate addresses efficiently. And for the sake of more flexible address allocation, we define two kind of multicast addresses, which are called VMA (Virtual Multicast Address) and MAR (Multicast Address for Routing). We present amount of routing information with our architecture, and consider inter-operability with other routing protocols.

1 はじめに

近年、インターネットの普及とともに、インターネット上でさまざまなサービスが提供されるようになってきた。WWW (World Wide Web)、E-mail といったテキストベースの 1 対 1 のユニキャスト通信に加え、音声や動画といったマルチメディアデータの 1 対多または多対多のマルチキャスト通信も広まりつつある。

また、インターネット上でのマルチキャスト通信の実現に向け、Mbone と呼ばれる仮想的なマルチキャストネットワーク上で盛んに実験されてきたが、解決されていない問題は多い。重要な問題としては、経路情報の量が膨大になることがあげられるが、この問題は現在の IP マルチキャストアーキテクチャと経路制御プロトコルの仕様から生じている。

現在の IP マルチキャストアーキテクチャでは、あ

るマルチキャストグループに参加しているホストはインターネット上のどこにでも存在し得る。そのため、ユニキャストの場合とは異なり、マルチキャストでは、アドレスをドメイン毎に割り当てることができず、マルチキャストアドレスを集約することが困難となっている。

また、Mbone上で主に使われていたDVMRP (Distance Vector Multicast Routing Protocol) [1]、MOSPF (Multicast extension to OSPF) [2] といった経路制御プロトコルでは、送信ホストごとに配送木を構築する。これはルータが、配送木の構築と維持のために、送信ホストとマルチキャストグループに対して1つの経路情報を保持しなければならないことを意味している。そのため、ネットワークの規模に応じてルータで保持しておく経路情報が増大するので、大規模ネットワークには適していない。

最近では、経路情報量を減少するために、共有木型の配送木を構築するPIM-SM (Protocol Independent Multicast-Sparse Mode) [3] が提案された。PIM-SMでは、あるグループに対して1つの根となるルータ (*RP*:Rendezvous Point) を決定しておく。そして、多くのマルチキャストパケットは *RP* から各受信ホストに対して配送され、これによってPIM-SMは経路情報量を削減している。しかし、我々がJBネットワーク [4] 上でPIM-SMを用いて行なった実験では、送信ホストと *RP* の経路上に受信ホストがいた場合には、効率的な配送ができないことが明らかとなり、*RP* を決定することが、重要で、かつ、非常に難しい問題であることがわかった [5]。

そこで、本研究では経路情報量を削減することによって、インターネット上で世界規模に適應できるIPマルチキャストの経路制御を実現することを考え、これを実現する新しいIPマルチキャストアーキテクチャを提案する。なお、提案するアーキテクチャでは、主にAS間でのマルチキャストの経路制御について議論し、各AS内では、DVMRP、MOSPF、PIM-SMなどの経路制御プロトコルを用いることを前提としている。

2 IPマルチキャストの問題点

IPマルチキャスト技術が提案されてから長い年月がたっているにもかかわらず、IPマルチキャスト通信はそれほど普及していない。これは、現在のIPマルチキャストの問題点がまだ完全には解決されていないためである。本章では、これらのIPマルチキャストの問

題点について述べる。

2.1 経路制御

マルチキャストアドレスは、マルチキャストグループに関連づけられており、マルチキャストパケットの配送は、宛先のマルチキャストアドレスに基づいて行われる。そして、マルチキャストパケットは、マルチキャストグループに参加しているホストすべてに配送されなければならない。そこで、マルチキャストパケットを配送する経路を制御するマルチキャスト経路制御プロトコルが提案された [1][2][3]。しかし、マルチキャストの経路制御は、ユニキャストとは違い複数地点への経路を制御する必要があるために、非常に複雑なものとなっており、いくつかの問題点がある。

この問題としては、まず、経路情報量が膨大になるために、規模適應性に欠けていることがあげられる。グループに参加しているホストは、インターネット上のあらゆる地点に存在しえるので、すべてのルータが、1つのマルチキャストアドレスに対して経路情報を保持する必要がある。また、DVMRPやMOSPFでは、マルチキャストグループと送信ホストごとに配送経路を決定するため、経路情報量が膨大になってしまう。経路情報を削減するためには、経路情報を集約することが重要であり、これを実現するためにマルチキャストアドレスを割り当てる手法が提案されている。

次に、広域ネットワーク上での異なるドメイン間の経路制御の問題があげられる。DVMRP、MOSPFでは、経路制御のために必要となる制御パケットによるトラフィックが増大してしまうため、広域ネットワーク上の経路制御には適していない。そこで、制御パケットによるトラフィックを削減したPIM-SMが提案され、PIM-SMでドメイン間の経路を制御することが考えられた。しかし、PIM-SMでは、マルチキャストグループ群に対して1つの *RP* しか存在できないため、*RP* へのトラフィックの集中する問題や経路が最適なものにならないという問題がある。そのため、ドメイン間の経路制御には適していないといわれている。

2.2 アドレス割り当て

ユニキャストでは、ドメイン内でサブネットワークを階層化して構成し、アドレスをドメインごとに割り当てることによって、経路情報を集約し、経路情報量を削減している。一方、マルチキャストにおいては、特

定のマルチキャストグループに参加しているホストがインターネット上のどこにでも存在し得るため、アドレスを集約できない。そこで、アドレスを集約するために、マルチキャストアドレスをドメインごとに割り当てる手法が提案されてきた [6][7][8][9]。

文献 [6] は、マルチキャストアドレス割り当てのアーキテクチャを定義しており、文献 [7][8][9] では、そのアーキテクチャに基づいたアドレス割り当てが提案されている。しかし、これらの提案では、マルチキャストアドレスがドメイン毎に割り当てられるため、あるマルチキャストグループに参加しているホストが、自身が属していないドメインに依存してしまう問題がある。これは、あるホストが自身が属していないドメインに割り当てられているマルチキャストアドレスに対応したマルチキャストグループに参加する場合に、そのホストの参加と離脱の状態の管理がホストが属していないドメインに依存してしまうということである。このような依存がある場合には、あるドメインに障害が発生するとその影響が、障害が生じたドメインに割り当てられたマルチキャストアドレスに関するマルチキャスト通信ができなくなってしまう。

3 IP マルチキャストアーキテクチャの概要

2 節で述べたように現在の IP マルチキャストアーキテクチャにはいくつかの問題があり、インターネット全体でマルチキャスト通信を実現することは困難である。本論文では、これらの問題を解決する新しい IP マルチキャストアーキテクチャを提案し、インターネット全体でマルチキャスト通信を実現するための基盤を提供する。

本節では、提案する IP マルチキャストアーキテクチャの目的とその効果、および、実現方法について述べる。

3.1 本手法の目的

提案するアーキテクチャでは、2 節で挙げた問題を解決し、インターネット全体でのマルチキャスト通信を実現することを目的としている。2 節で述べたように、経路情報量が膨大になってしまうことが現在の IP マルチキャストアーキテクチャの重要な問題の 1 つであり、この問題が IP マルチキャストの規模適応性を損なわせている。そこで、本論文では経路情報量を削減

し、規模適応性に優れた IP マルチキャストアーキテクチャを提案する。また、経路制御が他ドメインに依存することも現在の IP マルチキャストアーキテクチャの重要な問題の 1 つである。提案するアーキテクチャでは、この問題を解決するための手法も提供している。

3.2 本手法の効果

本手法によって得られる効果として、経路情報量が削減されることと他のドメインに依存しない経路制御が可能になることがあげられる。まず、本手法によって削減される経路情報量について述べる。本手法では、経路情報が集約できるマルチキャストアドレス割り当て方式を提案している。そのため、経路情報量はアクティブになっているマルチキャストグループの数や送信ホストの数に比例しない。これは、将来 IP マルチキャストが普及し、ネットワークの規模が大きくなって経路情報量が急激に増加しないことを意味する。このため、本手法を用いることによってインターネット全体で IP マルチキャスト通信を実現することができる。

次に、あるドメインが他のドメインに依存しない経路制御が可能となることについて述べる。本手法のマルチキャストアドレス割り当てでは、マルチキャストアドレスはドメインに対してではなく、配送木に対して割り当てられる。これは、特定のマルチキャストアドレスに対応したドメインが存在しないことを意味し、これによって、あるドメインが他のドメインに依存することを防ぐことができる。

3.3 IP マルチキャストアーキテクチャの実現方法

経路情報量を効率的に削減するためには、経路情報を集約しなければならない。そこで、提案するアーキテクチャでは、経路情報を集約できるように動的にアドレスを割り当てることを考える。さらに、アドレスをより柔軟に割り当てるために、経路制御用マルチキャストアドレス (MAR: Multicast Address for Routing) と仮想マルチキャストアドレス (VMA: Virtual Multicast Address) の 2 つにマルチキャストアドレス空間を分離し、アドレス割り当てでは MAR を動的に割り当てることを考える。MAR はデータを転送するために使われ、VMA はエンドホストがグループを識別するために使われる。また、各 AS では、これらの 2 つのアドレ

スを変換するために、アドレス変換表 (AMT: Address Mapping Table) を保持している。一方、AS があるグループに対してデータを送信する場合には、そのグループに対応したマルチキャスト木への経路を各 AS が知っておく必要がある。これは、ある AS が経路情報を広告し、その経路情報に基づいて経路の計算することによって実現される。

実際のマルチキャスト通信において VMA と MAR を用いてデータが配送される様子を図 1 に示す。図 1 において、網線部分は AS 内であることを示し、網線部分の境界にあるルータは AS 境界ルータである。また、 R_{1+2} は、 R_1 と R_2 を集約したアドレスである。まず、送信ホストは、IP ヘッダのアドレスに VMA を指定して送信を行なう。送信ホストが存在する AS 境界ルータでは、AMT を見て VMA に対応する MAR を検索し、IP アドレスとして MAR が指定されている IP ヘッダでパケットをカプセル化する。そして、広告されていた経路情報から作成した経路表に基づいて他の AS へパケットを転送する。VMA に参加している受信ホストが存在する AS の境界ルータは、カプセル化されたパケットから MAR を含む IP ヘッダを取り除いて受信ホストへ転送する。このようにして、MAR と VMA が AS 境界ルータによって変換され、パケットが配送される。

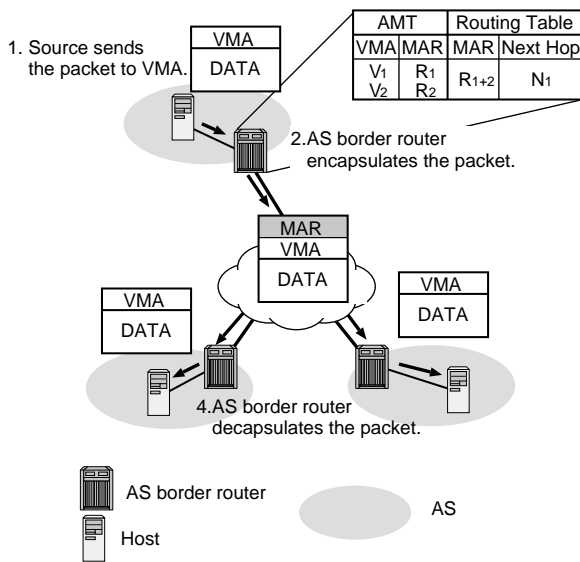


図 1: VMA と MAR によるマルチキャストパケットの転送

また、本手法では、ドメイン間の経路制御において経路情報量を削減するために、AS が新しくグループに参加する際に、効率的にアドレスが集約できるよう

に動的に MAR を割り当てることを考える。図 2 は、既に 224.3.1.1/32 というマルチキャスト木が存在していたときに、 F という AS が、あるマルチキャストグループに参加する様子を表している。なお、図 2 の $A \sim H$ は、AS を表す。

$AS-E$ と $AS-F$ を結ぶリンクを N 、実線で表されるリンクを T とし、 T が 224.3.1.1/32 のマルチキャスト木であるとする。ここで、 $AS-F$ があるマルチキャストグループに参加するとき、 T に含まれる AS が F に 224.3.1.2/32 のアドレスを割り当てる。これによって、 T に N を継ぎ足して、224.3.1.0/30 のマルチキャスト木を構築でき、アドレスの集約ができる。

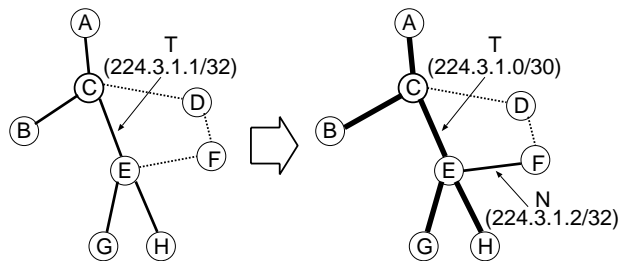


図 2: アドレス割り当て

4 IP マルチキャストアーキテクチャの設計

4.1 経路情報の広告

AS があるグループに対してデータを送信する場合には、そのグループに対応したマルチキャスト木への経路を各 AS が知っておく必要がある。これは、ある AS が経路情報を広告し、その経路情報に基づいて経路の計算することによって実現される。

経路情報の広告は、1 つのマルチキャスト木の中で最も経路情報を集約している AS が行なう。図 3 では、 $AS-C$ が集約された経路情報である 224.3.1.0/28 をすべての AS に対して広告している。実線は、既に構築されているマルチキャスト木を表し、破線は、マルチキャスト木に含まれていないノードを表す。また、矢印は C が広告する経路情報の流れを示している。 C が広告する経路情報には、ホップ数も付加されており、これは、構築されているマルチキャスト木からのホップ数を表す。このホップ数は、実線のノードを通過する際には変化しないが、マルチキャスト木に含まれないノードを通過する際に増加する。

ここで、 $AS-F$ が受け取る経路について考える。 $AS-$

*E*では、ホップ数を変更せずに経路情報を *AS-F*、*AS-G*、*AS-H* に広告する。*AS-D*では、*AS-D*がマルチキャスト木に含まれていないため、ホップ数に1を加え経路情報の広告を行なう。これによって *AS-F*は、224.3.1.0/28のマルチキャスト木に最も近い経路を得ることができる。

また、*AS-D*が224.3.1.0/30のマルチキャスト木に含まれていて、*AS-E*が含まれていない場合を考える。この場合、*AS-F*はマルチキャスト木への最短経路として *AS-E*を選出するので、ここで述べたものとは異なったマルチキャスト木が構築される。本手法では、異なったマルチキャスト木が構築されても問題にならないように、4.4節で述べている手法を用いて完全に受信ホストがいなくなるまでマルチキャスト木を保持している。

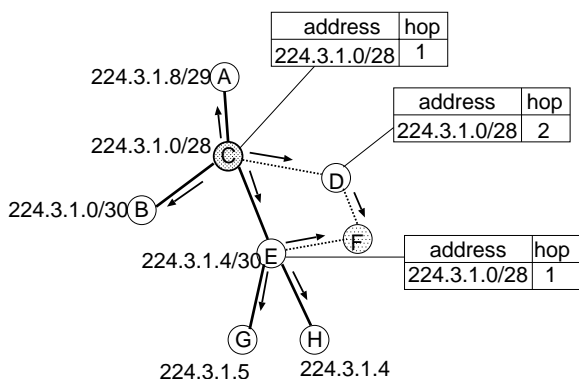


図 3: 経路の広告

このような手法を用いて経路情報を広告することで、マルチキャスト木のノードに含まれていない *AS*が、最も近いマルチキャスト木までの経路を得ることができる。

4.2 マルチキャスト木の構築

*AS*が新しく *VMA*に参加すると、*VMA*に対応した *MAR*が *AS*に割り当てられ、マルチキャスト木が構築される。

ここでは、224.3.1.1/32のマルチキャスト木 *T*がすでに構築されているとし、*AS-F*が新しく *VMA*へ参加することを考える(図4)。*AS-F*は、以下のような流れで *VMA*へ参加する。

1. マルチキャスト木までの最短経路の計算

節4.1で述べた手法によって広告されている経路情報に基づいて各 *AS*は224.3.1.1/32のマルチ

キャスト木までの最短経路を計算する。図4では、*AS-C*が経路の広告を行なう。

2. *VMA*に対応する *MAR*の検索

*AS-F*は、*VMA*に参加するとき、その *VMA*に対応する *MAR*のエントリを *AS-F*が持っている *AMT*の中から検索する。ここでは、簡単のために、*VMA*に対応する *MAR*が存在しないものとする。これは、既に *VMA*に対応したマルチキャスト木が存在しないことを意味する。

3. マルチキャスト木の構築

*AS-F*は、マルチキャスト木 *T*に含まれている *AS*の中から *AS-F*に最も近い *AS-E*を選出し、*AS-E*に対して Graft Request メッセージを送信する。

4. *MAR*割り当て

Graft Request メッセージを受信した *AS-E*は、*AS-F*に対して新しい *MAR*が割り当て可能かどうかを式1によって判断する。

$$L > \max_netmask_len \quad (1)$$

式1中の *L*は、集約しているアドレスのネットマスクの長さを表し、*max_netmask_len*は、アドレスを集約できる最小のネットマスクの長さを表す。このようにして集約できるアドレスの数を制限するのは、マルチキャスト木が1つになることを防ぐためであり、これによって、マルチキャスト木に含まれる *AS*に障害が発生してもすべてのマルチキャスト木に影響が及ぶ可能性を減少させている。

もし、式1が成立すれば、集約しているアドレスの数が上限に達していないことになり、アドレスが割り当て可能であることを示す。そして、割り当てが可能であれば、*AS-E*は効率的にアドレスを集約できる *MAR*を選択し、その *MAR*を Graft Ack メッセージの中に入れて、*AS-F*に送信する。割り当て不可能であれば、*AS-F*自身が特定の *MAR*を選択し、そのマルチキャスト木を構築する。

5. *AMT*への *MAR*の登録

*AS-E*と *AS-F*は、*AS-E*によって選択された *MAR*と *AS-F*が参加する *VMA*とのエントリを *AMT*に登録する。

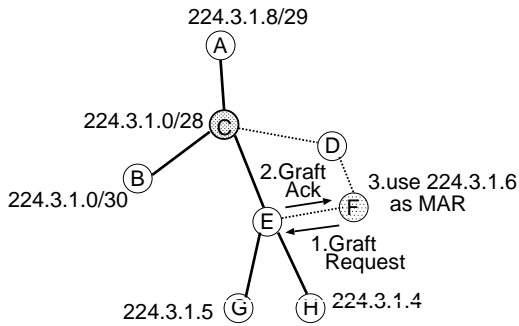


図 4: マルチキャスト木の構築

4.3 VMA からの離脱

VMA に参加している AS は、その VMA からいつでも脱退できるため、MAR を集約すると VMA に参加している AS が存在しない VMA 宛のパケットを受信してしまう。そこで、VMA に参加している最後の AS が VMA から離脱する際に、VMA に対応する MAR を解放することを考える。ここで、MAR を解放するとは、AMT に存在する VMA エントリを削除し、集約しているアドレスから MAR を取り除くことを指す。MAR の解放は、以下のような流れで行なわれる。

1. 解放する MAR を集約していたマルチキャスト木の末端の AS は、もし、その MAR を含まないようにアドレスを集約できれば、集約を行なう。
2. MAR の集約が可能、不可能に関わらず、AS は VMA エントリを削除する。

もしアドレスの集約ができない場合には、利用されていない MAR が存在してしまうが、新しく AS が VMA に参加する時に再利用可能であるために、それほど問題にはならない。しかし、参加している AS が存在していない VMA 宛のパケットを受信してしまう問題は解決できていない。この問題を解決するために、経路情報にアクティブフラグと呼ばれるフラグを導入する。アクティブフラグについては、後述のマルチキャスト木の保持の節で詳しく述べる。

また、利用されていない MAR の数が多くなると、アドレス空間を浪費してしまう問題も生じる。そこで、提案するアーキテクチャでは、MAR の再割り当てを行なう。MAR の再割り当てについては後述する。

4.4 マルチキャスト木の保持

VMA への参加と離脱が頻繁に生じると経路表を変更する回数やアドレスの再割り当てを行なう回数が増

加する。これを防ぐために、経路情報にアクティブフラグを設け、VMA に参加している AS が完全にいなくなってもマルチキャスト木をある一定期間維持することを考える。

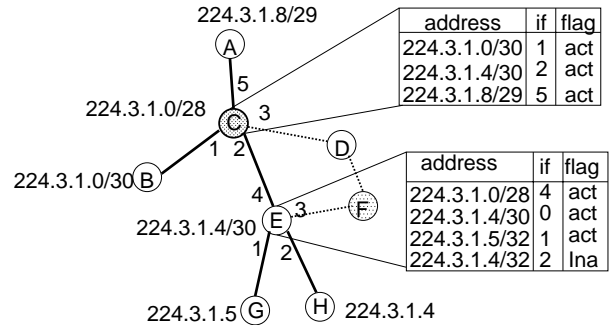


図 5: マルチキャスト木の保持

たとえば、図 5 では、AS-F が 224.3.1.5 から離脱したときの AS-C、AS-E における経路情報の変化の様子を表している。図 5 の address、if、flag は各 AS での経路表であり、address は、MAR、if はインターフェース、flag はアクティブフラグを表す。AS-F が離脱しても、AS-C の経路表は何も変化はないが、AS-E の経路表には、224.3.1.5 に参加している AS がいないため、アクティブフラグが、act から ina に変化している。このとき、AS-E は、224.3.1.5 宛のマルチキャストパケットを AS-F に配送しない。しかし、AS-F が以前まで参加していた同じ VMA に再び参加するときには、AS-E の経路表の 224.3.1.5 のアクティブフラグが act に設定される。

4.5 アドレス再割り当て

AS の参加と離脱が頻繁に起こる場合、経路の集約によって、使われていない経路が存在してしまう、アドレスのフラグメントと呼ばれる問題が生じる。これらを防ぐためにアドレスの再割り当てが必要となる。

AS がマルチキャストグループから離脱してから一定時間過ぎてから、実際にアドレスの再割り当てが行なわれる。アドレス再割り当ての処理の流れは以下のとおりである。なお、 $AS_{realloc}$ は、再割り当てが必要な AS を表し、 AS_{upstrm} は $AS_{realloc}$ の上流の AS を表す。

1. $AS_{realloc}$ がグループから離脱し、タイマをセットする。
2. タイマが切れると、 $AS_{realloc}$ は再割り当てが必要

要かどうかを式 2 に基づき判断する。

$$\frac{G_{inact}}{G_{alloc}} > threshold \quad (2)$$

アクティブでないループの数 : G_{inact}

AS に割り当てられているグループの数 : G_{alloc}

もし、式 2 が成り立つ場合には、以降のアドレスの再割り当ての処理を行なう。

3. $AS_{realloc}$ は G_{inact} を含んだ Realloc request を AS_{upstrm} へ送信する。
4. AS_{upstrm} は Realloc Request に含まれた G_{inact} が解放できアドレスが再割り当てできるかどうか判断する。

もし、再割り当てができれば AS_{upstrm} は $AS_{realloc}$ に割り当てられているマルチキャストアドレスのうち、その上位 G_{inact} だけを解放し、アドレスの集約を行なう。

一方、割り当てができない場合には、 AS_{upstrm} は、さらに自身よりも上流の AS に対して Realloc Request メッセージを送信する。

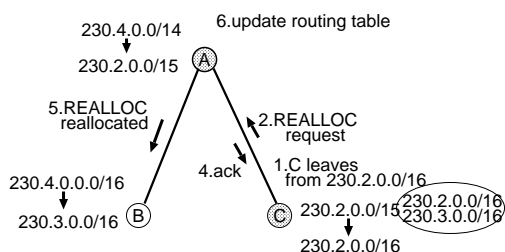


図 6: アドレス再割り当て (グループからの離脱時)

なお、ここで述べているタイマや閾値 $threshold$ の値は、アドレス再割り当てに要する時間やアドレス空間の使用率に基づいて決定される必要があるが、その具体的な決定方法については、今後検討していく。

5 考察

5.1 経路情報量

本アーキテクチャでは、4.2 節で述べているように、1 つのマルチキャスト木に割り当てられるアドレスの数の上限が決められている。ここでは、このアドレス

の数の上限によって経路情報量が増加することを示し、この上限の値を決定するための指針を提供する。

まず、ある AS での経路情報量の最大値を求める。 N_{nbrs} をある AS に隣接している AS の数とし、1 つのマルチキャスト木に関してすべての隣接している AS 毎に経路が存在するとすると、存在する経路情報量は N_{nbrs} に等しくなる。また、AS が集約できるアドレスの最小のネットマスクの長さを L とすると、存在するマルチキャスト木の数は 2^L に等しいので、経路情報量の最大値 R_{info} は、式 3 で表すことができる。

$$R_{info} = 2^L \cdot N_{nbrs} \quad (3)$$

式 3 において、 N_{nbrs} は定数であるので、経路情報量の最大値 R_{info} を削減するためには、 L の値を小さく設定する必要があることがわかる。

次に、他の経路制御プロトコルでの経路情報量と比較し、 L の値を決定するための指針を示す。DVMRP や PIM-DM では、マルチキャスト木は送信ホストごとに構築されるために、 $O(S \cdot G)$ の経路情報が必要となる。また、MOSPF では、送信ホストから各受信ホストまでの経路を保持しておかなければならないので、 $O(R \cdot G)$ の経路情報が必要となる。また、PIM-SM や CBT では、マルチキャストグループが RP もしくはコアルータによって集約されるため、マルチキャスト木に含まれないルータでの経路情報量は $O(G/C)$ となる。しかし、マルチキャスト木に含まれているルータでは、グループ毎にマルチキャスト木を維持しなければならないので、必要となる経路情報量は $O(G)$ となる。なお、 S 、 R 、 G および C は、それぞれ、送信ホストの数、受信ホストの数、グループの数、RP もしくはコアルータの数である。

これらのことから、IPv4 では IP マルチキャストアドレス空間は 28 ビットであるので、各プロトコルの経路情報量の最大値は、表 1 のようになることがわかる。なお、マルチキャストアドレスはアプリケーションや経路制御プロトコルに対していくつか既に静的に割り当てられており、本手法で使われるアドレス空間は実際には 28 ビットよりも少なくなるが、本論文ではその点について厳密には考慮しない。

表 1 から、本論文で提案しているアーキテクチャが DVMRP、PIM-DM、MOSPF に比べ、大幅に経路情報量を削減できていることがわかる。また、PIM-SM や CBT よりも経路情報量を削減するためには、式 4 を満たすように L を決定する必要がある。

表 1: 経路情報量の比較

DVMRP,PIM-DM	$2^{28} \cdot 2^{32} = 2^{60}$
MOSPF	$2^{28} \cdot 2^{32} = 2^{60}$
PIM-SM,CBT	2^{28}
提案するアーキテクチャ	$2^{(L-4)} \cdot N_{nbrs}$

$$2^{L-4} \cdot N_{nbrs} < 2^{28} \quad (4)$$

例えば、 $N_{nbrs} = 128$ だと仮定すると、式 5 を満たすように L を決定すれば良い。なお、実際にマルチキャストアドレスとして使えるアドレス空間は 28 ビットよりも少ないため、より小さな値をとる必要がある。

$$L < 25 \quad (5)$$

式 5 から L の値を小さくすればするほど経路情報が削減できることがわかる。たとえば、経路を集約するアドレスのマスクの長さ L を 8、ある AS に隣接している AS の数 N_{nbrs} を 128 だと仮定すると経路情報量は、式 8 で表される。

$$2^{L-4} \cdot N_{nbrs} = 2^4 \cdot 128 \quad (6)$$

$$= 2^{14} \quad (7)$$

$$= 16384 \quad (8)$$

現在、AS 間で交換されている IPv4 のユニキャストの全経路数は、現在 88,325 個であることから [10]、式 8 で得られたマルチキャストの経路数がマルチキャストをインターネット全体で実現するのに十分な値であると思われる。

しかし、 L を小さな値にすればするほど、マルチキャスト木の数が増えることになる。それにとまって、マルチキャスト木に含まれる AS で障害が発生するとマルチキャスト木が途切れることが頻繁に生じる可能性がある。このことから対障害性を考慮して、 L の最小値を決めることも重要であることがわかる。今後、この L の最小値を決定していく予定である。

5.2 アドレス再割り当てのオーバーヘッド

ここでは、アドレス再割り当てにかかるオーバーヘッドの最大値について考察する。オーバーヘッドが最も大きくなるのは、マルチキャスト木に属するすべての AS

に対してアドレス再割り当てが行なわれるときである。つまり、最も経路を集約している AS まで再割り当てが行なわれ、マルチキャスト木の末端の AS に再割り当てまで再割り当てが必要となるときである。

図 7 では、AS-A が最も経路を集約している AS であり、AS-D が再割り当てを要求した AS である。再割り当てが実際に行なわれるのは、AS-D から Realloc Request メッセージが出されて、AS-A にまでそのメッセージが到達してからである。そのため、AS-A が再割り当ての処理を始めてから、すべての AS に対する再割り当てが完了するまでにかかる時間 T_{total} は、最後の AS に割り当てが完了するまでの時間 T_{last} に等しい。

次に、図 7 において、AS-Z が最後に割り当てが完了する AS であったとして T_{total} を求めることを考える。 $m_1 \sim m_n$ をそれぞれのリンクを再割り当てメッセージが通過するのにかかる時間とし、各 AS で再割り当ての処理にかかる時間を $T_{realloc}$ とすると、 T_{total} は式 10 のように表せる。

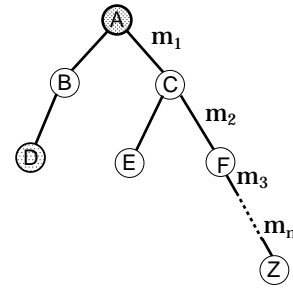


図 7: 再割り当てのオーバーヘッド

$$T_{total} = T_{last} \quad (9)$$

$$= \sum_{i=1}^n T_{realloc} \cdot m_i \quad (10)$$

式 10 から、再割り当てに要する時間を削減するためには、最も経路を集約している AS から末端の AS までのホップ数を減らすことが必要で、また、リンクの通信速度を考慮に入れてマルチキャスト木を構築しなければならないことがわかる。このようなリンクごとのコストやホップ数を考慮にいれた最小コストのマルチキャスト木の構築は、Steiner 問題と呼ばれ、NP-完全問題であることが証明されている [11][12]。すでに、Steiner 問題を近似的に解決する手法も提案されており [13]、今後、それらの手法を参考にして、マルチキャスト木を構築する手法を改善していく予定である。

また、アドレス再割り当てがマルチキャスト全体で完了しなければ、マルチキャストパケットが正しく配送されることはない。つまり、アドレス再割り当ての処理が始まってから T_{total} だけパケットが配送されなくなってしまう、パケットが損失されるおそれがある。パケット損失は、AS のボーダルータでパケットをバッファに貯めておくことで防ぐことができる。しかし、バッファに貯めることによって遅延時間が発生するため実時間性が重要となるアプリケーションにとっては、バッファに貯めることはあまり望ましくない。

このため、上位のアプリケーションが必要とする QoS パラメータに応じてアドレス再割り当ての実行を判断しなければならない。

5.3 他の経路制御プロトコルとの相互運用

ここでは、提案するアーキテクチャの他の経路制御プロトコルとの相互運用性について考察する。

現在、Mbone 上でのマルチキャスト通信で使われているマルチキャストアドレスは、各ホストが通信を始める際に適当に選択されたものが多い。そこで、これらのアドレスを本アーキテクチャにおける VMA とみなし、現在 Mbone 上で使われている経路制御プロトコルとの境界ルータで、VMA と MAR の変換を行なうことで相互接続が可能となる。しかし、本アーキテクチャと他の経路制御プロトコルでは、マルチキャスト木を構築する手法が異なるために、境界ルータでアドレス変換以外のプロトコル変換が必要になることが考えられる。

今後、この点をより詳細に考察していく予定である。

6 おわりに

本論文では、スケーラビリティを考慮した IP マルチキャストアーキテクチャを提案した。提案したアーキテクチャでは、マルチキャストアドレス空間を VMA と MAR の 2 つに分離することによって、より効率的なアドレス割り当てを実現し、経路情報を削減することができた。今後、本論文で明らかになった問題点を解決し、実装を行なう予定である。

参考文献

[1] D. Waitzman, C. Partridge, and S.E. Deering. *Distance Vector Multicast Routing Protocol*,

November 1988. RFC 1075.

- [2] J. Moy. *Multicast Extensions to OSPF*, March 1994. RFC 1584.
- [3] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, and L. Wei. *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification*, June 1998. RFC 2362.
- [4] Masaki Minami, Masafumi Oe, Akimichi Ogawa, Kengo Nagahashi, Kouji Okamura, Youki Kadobayashi, Hiroshi Esaki, and Akira Kato. *Jb: Design and architecture of next generation internet infrastructure in japan*, September 1999.
- [5] 大森幹之, 岡村耕二, 荒木啓二郎. 次世代インターネット上でのマルチメディア通信処理の課題について. 九州大学大型計算機センター第 17 回計算機科学研究集会.
- [6] D. Thaler, M. Handley, and D. Estrin. *The Internet Multicast Address Allocation Architecture*, July 2000.
- [7] Deborah Estrin, Ramesh Gobindan, Mark Handley, Staish Kumar, Pavlin Radoslavov, and Dave Tahaler. *The Multicast Address-Set Claim(MASC) Protocol*, January 2000.
- [8] Mark Handley. *Multicast Address Allocation Protocol(AAP)*, March 2000.
- [9] S. Hanna, B. Patel, and M. Shah. *Multicast Address Dynamic Client Allocation Protocol(MADCAP)*, December 1999.
- [10] <http://www.employees.org/~tbates/cidr-report.html>. *The Cider Report*.
- [11] S.Hakimi. Steiner's problem in graphs and its applications. *Networks*, Vol. 1, pp. 113-133, 1971.
- [12] R.M. Karp. *Reducibility among combinatorial problems*. Complexity of Computer Computations. Plenum, New York, 1972.
- [13] H. Takahasi and A. Matsuyama. An approximate solution for the steiner problem in graphs. *Mathematics Japonica*, Vol. 6, pp. 573-577.