第 6 部

マルチキャスト通信

第 1 章

はじめに

ここでは、マルチキャスト通信ワーキンググループによる、IP マルチキャストを用いた通信技術について報告する。

計算機を接続するために各物理媒体が提供する通信形態は、以下のようなものがある。

- 通信相手を 1 つに特定して通信を行なう 1 対 1 型通信
- 通信相手を 1 つの計算機グループに特定して通信を行なう 1 対 n 型通信 (マルチキャスト型通信)
- 通信相手を特定せず、1 つの物理媒体で接続される全ての計算機に対して通信を行なう同報型通信 (ブロードキャスト型通信)

多くの物理媒体では通信相手を特定するために、物理媒体固有のアドレス体系を使って通信を行なっている。つまり Ethernet における 48 ビットの MAC アドレスとかである。

インターネットにおいては、様々なネットワークの物理媒体の特徴を生かすために、インターネットのアドレス体系を別に作り、各物理媒体において物理媒体固有のアドレス体系に変換している。これにより、外見上は物理媒体を意識することなく通信が行なえるような仕組みとなっている。

特に、インターネットにおけるネットワークプロトコルの一つである IP では、32 ビットからなる IP アドレスをインターネットアドレスとして用い、異なる物理媒体間の通信を実現している。

IP アドレスは、内部構造がネットワーク部とホスト部に分かれ、管理レベルのネットワークや、ホストを識別できるようになっている。これにより、以下のような識別が可能である。

- 各ホストが接続されている物理媒体への入口 (インターフェース)
- あるネットワークに属している全てのホスト

前者は、1 対 1 型の通信を提供するために使われ、後者は同報型の通信を提供するために使われている。

このようなアドレス体系に基づき、1 対 1 型の通信、および同報型の通信を利用した、様々な分散型のアプリケーションが開発されてきた。 これらのアプリケーション、特に

分散型のデータベースシステム等においては、同一の情報を複数のホスト (n 個) に伝達する必要のある状況が生じる。しかしながら、現在運用中のインターネットの提供する通信形態では、1 対 1 型の通信を n 回行なうか、同報型の通信を利用する以外に実現方法がない。

1 対 1 型の通信を n 回利用する場合には、同一情報を伝えるためだけにネットワークのバンド幅が消費され、他のネットワーク利用の効率に影響を与える。バンド幅の小さい物理媒体を利用している場合には、特に影響が大きいと考えられる。 また、同報型の通信を利用する場合には、n 個以外のホストに対して、無駄な情報を与えることになるので、他のホストのデータ処理時間に影響を与える。したがって、同一情報を複数のホストに伝達するときに、最小限のデータ量で済むような通信形態の必要性が高まっている。

前述のように、物理媒体によっては 1 対 n 型 (マルチキャスト型通信) の通信を提供している場合もあるが、インターネット規模ではそのような通信形態の利用は進んでいない。。そこで、インターネット、つまり異なる物理媒体 間、においても 1 対 n 型 (マルチキャスト型通信) の通信形態を提供できるように、インターネットを拡張する動きが出てきた。

TCP/IP プロトコル体系では、n 個のホストをグループとして扱い、グループに対してアドレスを割り当てることが可能となるようにアドレス体系を拡張している。グループアドレスを元にした経路制御アルゴリズムは、グループメンバーホストの所属状況等の情報を必要とするので、1 対 1 型、および同報型の通信形態の経路制御アルゴリズムに比べ、さらに複雑となることは容易に想像できる。

実際、マルチキャストの経路制御アルゴリズムについての議論はまだ十分とはいえない。TCP/IP プロトコル体系においても、マルチキャストの既存の経路制御アルゴリズムの実装例が示されているが、広域でマルチキャストを利用する際には多くの問題が残されている。

マルチキャスト型の通信を提供する物理媒体の中には、放送型広域通信媒体のように、異なる物理媒体に所属するホスト間を直接接続することのできるものもある。WIDE マルチキャスト通信ワーキンググループでは、TCP/IP プロトコル体系において、広域マルチキャストを実現するために、従来の経路制御アルゴリズムの問題点を指摘すると共に、広域で利用できるような物理媒体を有効に利用した経路制御アルゴリズムを研究している。

本稿では、まず、既存の IP マルチキャスト通信についてアドレス体系および経路制御アルゴリズムについて概観する。その後、マルチキャスト型通信を広域に拡張する際の問題点について議論する。さらに、衛星通信等の広域の放送型通信媒体があった場合の、効率的な経路制御機構、マルチキャスト通信の伝搬機構について述べる。

第 2 章

IP マルチキャスト

この章では、既存の IP マルチキャストのプロトコル体系について概観する。

2.1 経路制御アルゴリズム

ここでは、マルチキャスト機能を有する LAN 間を接続して構成されるインターネットワークにおいて、マルチキャストの利用を可能とするための既存の経路制御アルゴリズムについて述べる。このプロトコルは DVMRP(Distance Vector Multicast Routing Protocol) としてまとめられおり、RFC1075 に定義されている [?], [?]。

2.1.1 Distance–Vector Routing を元にしたマルチキャスト経路制御アルゴリズム

Distance–Vector routing アルゴリズムは、多くのネットワークで長年に渡り適用されてきたアルゴリズムである。現在でも XNS、routed、gated を使う UNIX のネットワークで使われている [?]。

Distance–Vector routing アルゴリズムは隣接ルータ間で経路制御情報を交換することにより、経路を学習する。このような経路制御情報の交換によって、終点への最短距離の経路を選択するようにする。経路制御情報は (V, D) というリストを含んでいる。Vは終点を表し、Dは終点への距離を表す。距離とは、理想的には個々の中継ルータ間 (1 ホップ) にかかるコストの合計によって定義される。実際には、例えばデータが通過する中継ルータ数 (ホップ数) で定義されている。これについては、データの伝搬遅延の総量、あるいは、送信コストによって定義される場合もある。

中継ルータを介さない、直接接続された 2 つのホスト i_1から i_2への距離を $d(i_1, i_2)$ とすると、これは個々のホップにおけるコストを表すことになる。また、ホスト iから jへの最短距離を $D(i, j)$ と表すと、ホスト i と直接接続しているホスト kとして、次のように表現される。

$$\begin{aligned} D(i,i) &= 0 & all\ i \\ D(i,j) &= \min_k[d(i,k) + D(k,j)] & otherwise \end{aligned}$$

Distance–Vector routing アルゴリズムは、ユニキャストの経路制御アルゴリズムとしてよく使われているため、ユニキャストの経路制御情報を応用したマルチキャストの経路制御アルゴリズムが考えられている。マルチキャストの経路は tree を形成するため、この tree の構築方法が重要となる。そこで、まず、マルチキャストの経路を構築するための方法として 2 つの方法が挙げられている。

- 全てのリンクに対する single–spanning–tree を計算し、ブリッジで使われる single–spanning–tree routing アルゴリズムを利用する。
- reverse path broadcast アルゴリズム により得られる shortest–path broadcast tree を元にして、各送信ホストから各グループに対する shortest–path multicast tree を計算する。

一般に、同じグループに対するパケットであっても、どの送信ホストからのパケットであるかによって、グループの各メンバーに対する最短経路は異なる。ネットワークの規模が大きくなるにつれ、グループの各メンバーにできるだけ小さい遅延で配送をすることが LAN のマルチキャスト機能を有効にするために重要となってくる。

前者の single–spanning–tree を計算する方法では、送信ホスト毎の柔軟な経路制御を行なうことが難しく、遅延をできるだけ小さくする方針に反することになる。

したがって後者の reverse path broadcast アルゴリズムを元に、柔軟なマルチキャストの経路制御を行なう方針が選択されている。

$$shortest - path\ multicasttree \subseteq shortest - path\ broadcasttree$$

であることから、shortest–path broadcast tree を効率良くたどる方法から broadcast tree を切り詰めて shortest–path multicast tree を構築する方法が考えられている。

ここでは、reverse path broadcast algorithm を元にした以下の 4 つの既存のマルチキャスト経路制御アルゴリズムとその問題点について述べる。

- Reverse Path Flooding (RPF)
- Revese Path Broadcasting (RPB)
- Truncated Reverse Path Broadcasting (TRPB)
- Reverse Path Multicasting (RPM)

4 つのアルゴリズムは、

RPF $\longrightarrow$ RPB $\longrightarrow$ TRPB $\longrightarrow$ RPM

の順に問題点に改良を加えたブートストラップ型のアルゴリズムとなっている。これらは、以下の 2 種類に大別される。

- shortest–path broadcast tree を効率よく辿り、LAN のインターフェースで提供されているアドレスフィルタリング機構を利用して、各グループのメンバーがパケットを受け取るアルゴリズム。従って、グループメンバーの把握を行う必要がない。(RPF,RPB)

- shortest–path broadcast tree の切り詰めを行い、グループメンバー以外にパケットを送信しないようにするアルゴリズム。tree の切り詰めを行うために、マルチキャストルータによるグループメンバーシップ情報の交換が必要となる。(TRPB,RPM)

2.1.2 RPF

まず、shortest–path broadcast tree を構築するが、これは reverse path broadcast algorithm によって行われる。reverse path broadcast algorithm では、ブロードキャストパケットをルータが転送するかどうかと、転送する場合にはその方向を与える。転送するかどうかの判断として、ブロードキャストパケットを受け取ったルータからブロードキャストパケットの送信ホストに戻る経路を考える。これが reverse path と呼ばれる由縁である。

その経路が shortest–path (shortest–reverse–path) であるなら、ブロードキャストパケットを転送する。その方向は、ブロードキャストパケットが届いたリンク以外の全てのリンクに対してである。shortest–reverse–path の判断は、ユニキャスト型の経路制御アルゴリズムで Distance Vector を採用している場合には、経路制御テーブルを変更なく利用することが可能である。

ここではマルチキャストの経路制御を行うために、インターネットマルチキャストアドレスを利用して、broadcast tree をたどり、アドレスフィルタリング機構を利用してグループメンバーホストのみがパケットを受け取る。マルチアクセス型の LAN においては、マルチキャストルータは LAN のマルチキャストアドレスに変換して隣接マルチキャストルータ間とマルチキャストパケットの転送を行う。従って、マルチキャ ストルータは全てのマルチキャストパケットを受け取る必要がある。

このアルゴリズムの問題点は次のとおりである。

> 1 つの LAN にマルチキャストルータが複数存在する場合には、その LAN に同一のマルチキャストパケットが複数回配送される可能性がある。例えば、図 2.1 の接続状況において、ネットワーク c にはその可能性がある。

2.1.3 RPB

RPF の問題点を解決するために、RPB では 1 つの LAN に同一のマルチキャストパケットが 1 回だけ配送されるように改良するアルゴリズムとなっている。

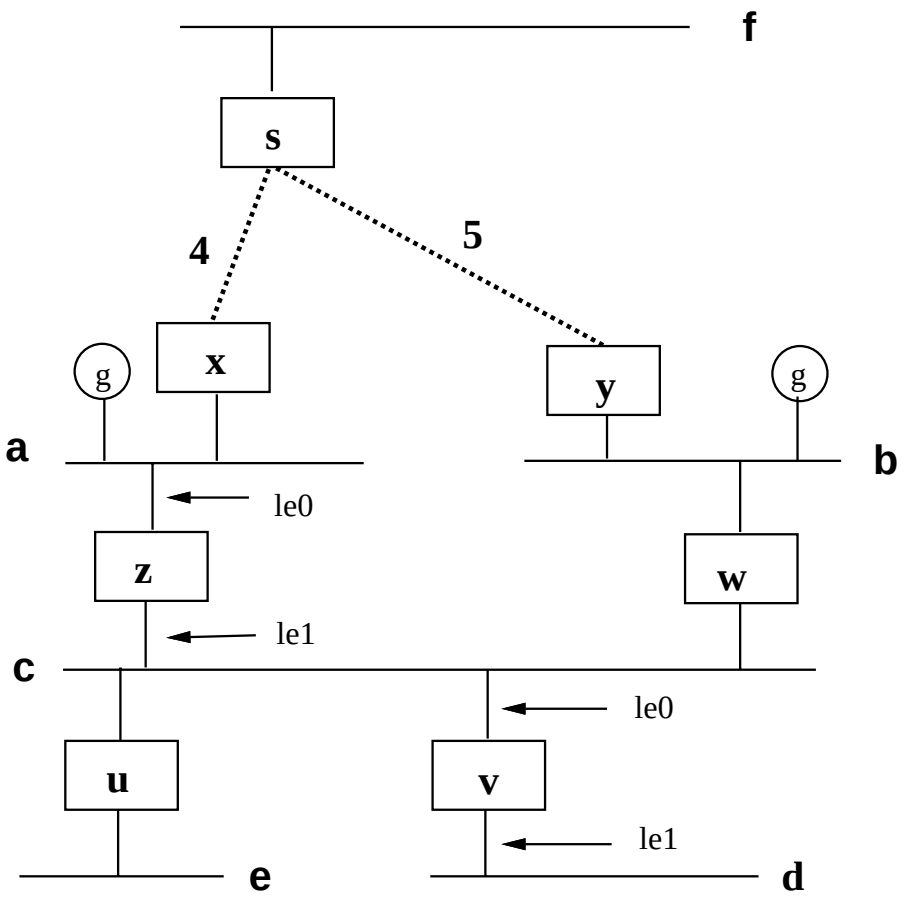

図 2.1: LAN 間接続例

RPB では、1 つの LAN にマルチキャストパケットを配送するマルチキャストルータが 1 つになるようにする。そのために、マルチキャストルータによる child link の識別を RPF のアルゴリズムに付加している。

マルチキャストルータは、直接接続しているリンク上の他のマルチキャストルータと比較して、マルチキャストパケットのある送信ホストに対し、自分が最短距離で到達できる場合にはそのリンクを child link と判断する。そして、マルチキャストルータは、マルチキャストパケットが届いた以外の child link に対してのみ、マルチキャストパケットの転送を行う。あるマルチキャストルータが、ある送信ホストに対して child link と判断するリンクを所有するとき、逆にそのリンクから見てそのマルチキャストルータを parent router と呼ぶ。

child link の識別の際は、Distance Vector の経路制御アルゴリズムを利用する。隣接ルータ間と定期的に経路制御情報を交換しているため、マルチキャストパケットの任意の送信ホストから隣接ルータへの距離がわかり、自分の経路制御テーブルに書かれた距離との比較を行うことによって、child link の識別を行うのである。他の隣接ルータと等しい距離であるときには、例えばインターネットアドレスの小さい方を選択するなどの対処が必要である。

child link の識別情報を保持するために、経路制御テーブルに children という新しい欄を設けて記録しておく。これは、経路制御テーブルの各エントリに対し、そのマルチキャストルータに直接接続している各リンクが child link であるか否かをビットで示すものである。例えば 図 2.1のような接続状態において、ルータ z の 経路制御テーブルを表 2.1 に示す。

表 2.1: RPB におけるルータ“ z ”の経路制御テーブル

destination	distance	next-gateway	interface	age	children	
					a	c
a	0	z	le0	..	0	1
b	1	w	le1	..	1	0
c	0	z	le1	..	1	0
d	1	v	le1	..	1	0
e	1	u	le1	..	1	0
f	5	x	le0	..	0	1
...						
...						

RPF の改良による経路制御自体のコストの増加は、 経路制御テーブルに children 欄を設けることにより、経路制御テーブルの容量がわずかに大きくなるという点のみである。

このアルゴリズムの問題点は次のとおりである。

> shortest–path broadcast tree を辿る方法としては RPF と比較して効率がよいが、マルチキャストパケットの受信ホストが少ない場合には、無駄なトラフィックが発生する。これは broadcast tree をそのまま辿るアルゴリズムの欠点といえる。

2.1.4 TRPB

TRPB では、shortest–path broadcast tree の葉 (末端) に当たるリンク (leaf link) の中で、グループメンバーに属しているホストが 1 つも存在しないリンクにはマルチキャストパケットを転送しないようにする。これにより、従来の shortest–path broadcast tree の葉の部分の切り詰めが可能となり、無駄なパケットを多少防ぐことになる。

shortest–path broadcast tree を切り詰めるために、各マルチキャストルータは、直接接続しているリンクに対して、RPB のアルゴリズムに加え次の 2 つの新たな認識を行う必要がある。

- 各送信ホストに対する shortest–path broadcast tree の葉の部分 (leaf link) の認識
- 直接接続しているリンク上のホストのグループ所属状況の検出 (グループメンバーシップ情報の把握)

leaf link とは、あるマルチキャストルータにとって、マルチキャストパケットのある送信ホストに到達する経路を考える際に、他のどのマルチキャストルータもそのリンクを使わないような child link のことである。従って、あるマルチキャストルータにとって child link でないようなリンクについては leaf link とは呼ばない。

leaf link の識別のためには、ある送信ホストに戻る経路において、あるリンクを他のマルチキャストルータが利用するかどうかをそのリンクの parent router に伝える必要がある。つまり、定期的に交換する経路制御情報の内容に加え、各経路 (reverse–path) がどのリンクを次のホップとしているかという情報を付加する必要が生じる。

Distance Vector 経路制御アルゴリズムの実装例の中には、インターネットワークのトポロジーが変化したときに古い経路情報が早く無効になるように、各経路がどのリンクを次のホップとしているかという情報を利用している場合がある。これは、split horizon と呼ばれる技術で、各経路で次のホップとなるリンクに対しては、その経路の距離を無限大にして経路情報を与えるものである。

すると、ネットワークのトポロジーに変化が生じて従来の経路が使用不可能となった場合に、それまで自分のリンクを通じて経路を確保してきたルータの経路制御情報から、誤ってそのルータを通る経路を新しい経路であると判断しないようにすることができ、経路制御の混乱を防ぐことができる。split horizon の詳細については [?] にある。この実装が既に行なわれている環境では、leaf link の識別のために新しい経路制御情報を付加する必要はない。

leaf link の識別の情報は RPB における child link の識別の際と同様に、経路制御テーブルの各エントリに対し、leaves というビット欄を設けて記録する。従って、経路制御テーブルの容量のコストが RPB の経路制御テーブルよりも leaves 欄分だけさらに増加する。例えば、図 2.1の接続状態でのルータ v の経路制御テーブルの具体例を示すと、表 2.2 のようになる。

グループメンバーシップ情報の把握に関しては、グループメンバーホストが同一 LAN 上の parent router にグループメンバーシップの報告を定期的に行うことによって実現される。parent router は送信ホストごとに異なるので、グループメンバーシップの報告を行う方向は、全ての送信ホストからの shortest–path broadcast tree を遡る方向であるといえる。この報告を、各グ ループの全てのグループメンバーホストが個々に行った場合には、定期時間ごとに、O(グループ数 $\times$ 発信ホスト数) のバンド幅やメモリを必要とする。従って、グループ数や発信ホスト数が増加すると、broadcast tree の切り詰めにかかるコストが増大する。このコストを最小限とするために、リンクレイヤブリッジのグループメンバーシップの報告方法を利用している。

表 2.2: TRPB におけるルータ“ v ”の経路制御テーブル

destination	distance	next-gateway	interface	age	children c	children d	leafs c	leafs d
a	0	z	le0	..	0	1	0	1
b	1	w	le0	..	0	1	0	1
c	0	v	le0	..	0	1	0	1
d	1	v	le1	..	1	0	1	0
e	1	z	le0	..	0	1	0	1
f	6	z	le0	..	0	1	0	1
...								
...								

具体的には、各グループに所属するホストが LAN のマルチキャスト機能を利用して、そのグループアドレスを受信ホストとするマルチキャストパケットを送信する。マルチキャストパケットで送信するため、同一グループに所属する他のホストにもグループメンバーシップの報告が行われたことがわかる。ということは、他のグループメンバーは同じグループについての報告を次の定期間隔になるまで行う必要がない。また、マルチキャストルータはその LAN 上の全てのマルチキャストパケットを受信し、leaf link のグループメンバーシップ情報を把握する。

この方法では、グループメンバーシップの報告にかかるコストは定期時間ごとに O(グループ数) に軽減される。このコストは報告を行う時間間隔に依存するが、リンクレイヤブリッジのグループメンバーシップの報告方法においても述べたように、時間間隔を大きく取ることによる影響は小さいため、分単位のオーダーで十分であるといえる。

このアルゴリズムにおける問題点は次のとおりである。

broadcast tree の切り詰めが行われるのは tree の葉の部分だけに限られている。そのため、broadcast tree の規模が大きく、グループメンバーホストの構成が送信ホストの近距離の範囲に集中している場合には、broadcast tree をだとる場合との差が小さい。

また、全ての送信ホストを想定してグループメンバーシップ情報の報告を行うので、マルチキャスト型の通信が頻繁に行われないような環境では、グループメンバーシップ情報のほとんどが実際には有効に利用されない。

2.1.5 RPM

RPM では、実際に使用されている shortest–path broadcst tree に関して、TRPB で行うよりもさらに細かい tree の切り詰めを行うものである。ある送信ホストが最初にあるグループに対してマルチキャストパケットを送信するときには TRPB のアルゴリズムで配送を行う。

1 回目の送信で、あるマルチキャストルータの全ての child link が全て leaf link であり、かつそれらのリンク上にグループメンバーホストが存在しない場合には TRPB よりもさらに木を切り詰められる可能性があると考える。そして、その場合には NMR(Non–Membership report) を、マルチキャストパケットの送信ホストに戻る経路上の次のホップとなるマルチキャストルータ (one–hop–back router) に対して送信する。(図 2.2参照) NMR を受け取ったマルチキャストルータは、全ての child link から NMR を受け取ったら、さらに NMR を one–hop–back router に対して送信する。

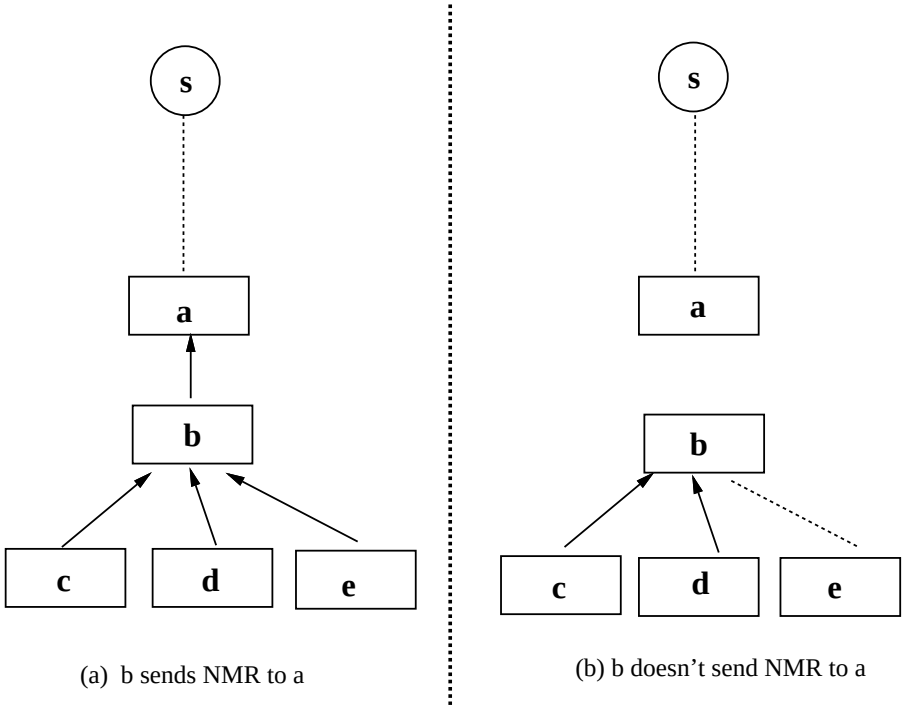

図 2.2: NMR の伝搬方向

2 回目からのマルチキャストパケットは、shortest–path broadcast tree の途中のマルチキャストルータが全ての child link から NMR を受け取っている場合には、そこから下の subtree にパケットが転送されない。従って完全なマルチキャスト tree を構築することになる。

NMR のメッセージ中には *age* フィールドを含め、NMR が tree を逆方向に辿るごとに増やしていく。この *age* フィールドの値がある敷居値 T_{maxage} を越えた場合には、その NMR を無効として捨てる。NMR の *age* は出発時は 0 とし、複数の NMR を受け取るマルチキャストルータが NMR を発行するときには、受け取った NMR の *age* フィールド

の値の最大値を採用する。これにより、T_{maxage}後には新たなメンバーがそのグループに所属することが可能となる。

しかし、新たにメンバーが加わった場合にはできるだけ早くそのグループに対するマルチキャストパケットを受け取ることが可能となるよう、NMR の取消し (Non–Membership Report Cancellation) を送信できるようにする等の工夫が必要となる。NMR の取消の方向は NMR と同様である。NMR の取消自身が伝搬途中で失われると、新しいグループメンバーがマルチキャストパケットを受け取るまでの遅延が大きくなるので、NMR の取消はそれを受け取ったという確認応答を返すようにし、T_{maxage}(つまり NMR の発生する間隔) は長い間隔で済むような設計が、NMR によって生じる付加的なコストを軽減するために重要である。

RPM では NMR による付加的なコストが、TRPB のコストに積算される。マルチキャストルータが保存すべき NMR の量は T_{maxage}毎に、

活動中のマルチキャストパケットの送信ホスト×
各送信ホストが受信ホストに指定するグループの平均数 × 隣接マルチキャストルータ数

のオーダーになると予想される。この量を軽減するための対応策として、以下の方法が挙げられる。

- 送信ホストをネットワーク単位等にまとめて情報量を減らす。
- マルチキャストパケットの生存時間を短くして大量の NMR が発生しないようにする。

このアルゴリズムの問題点は次のとおりである。

shortest–path broadcast tree を切り詰め、完全なマルチキャスト tree を構築することが可能であるが、tree の切り詰めにかかるコストが TRPB よりもさらにかかり、経路制御アルゴリズムも複雑となる。broadcast tree の規模が大きく、マルチキャスト型通信を頻繁に行うような状況では、NMR による通信コスト (バンド幅) や NMR 情報の保持に必要なメモリの量が増大し、scalability の面で問題が生じる。

また、インターネットワークのトポロジーの変化が生じた場合には、NMR を一度すべて取消し、トポロジーを再度把握しないと経路制御の混乱を招きやすい。

2.1.6　4 つのアルゴリズムの具体例

図 2.1に示すようなネットワークの接続例において、各アルゴリズムにより得られる tree　の状態を図 2.3、2.4 に示す。図で、太い実線で囲った○は、グループメンバーが存在するリンクであり、点線で囲った○はグループメンバーが存在しないリンクである。

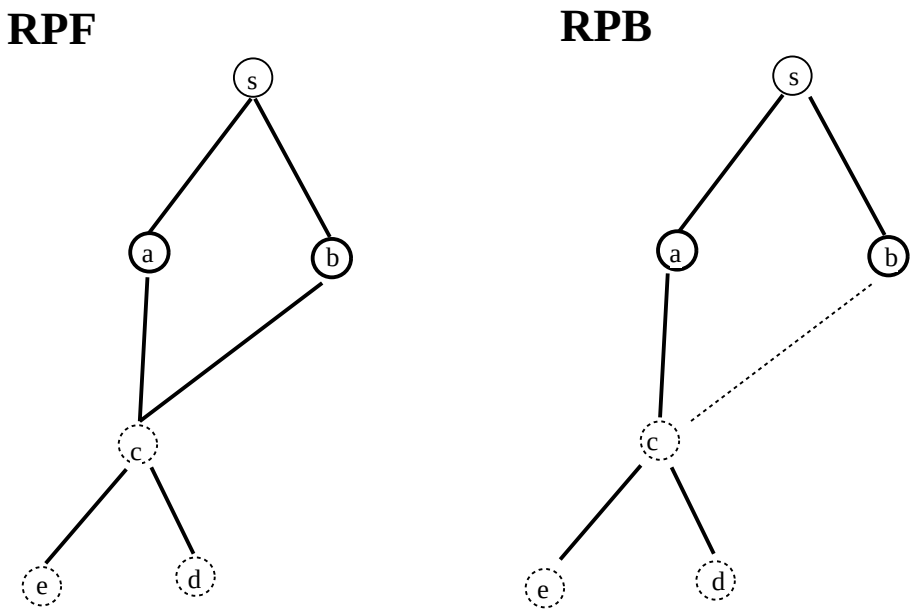

図 2.3:

ここでは、Distance Vector アルゴリズムを元にして構築される reverse path broadcast tree を出発点としたマルチキャストの既存の経路制御アルゴリズムについて示した。ブロードキャスト tree を効率よくたどる方法では、グループメンバーの存在しないリンクに無駄なトラフィックが生じるという欠点があり、また、マルチキャスト tree を構築するアルゴリズムでは、正確なマルチキャストを構築しようとする程、そのための通信コストが増大するという欠点が生じることが明らかにされた。

一般に reverse path broadcast tree を元にしたアルゴリズムでは、両方向の shortest path が同一であるという前提がある。実際のネットワークでは往復で経路の異なる場合が存在するので、reverse path が最短距離であっても、真に最短距離の tree が得られるとは限らない。

また、Distance Vector を元にした経路制御アルゴリズムはネットワークのトポロジーの変化への対応が速やかではない。マルチキャストでは、グループの構成ホストが短時間で変化することは十分考えられる。そこで、次にネットワークのトポロジーの変化への対応を得意とする、既存のアルゴリズムを元にしたマルチキャストの経路制御アルゴリズムについて述べることにする。

2.1.7 Link–State Routing を元にしたマルチキャスト経路制御アルゴリズム

Link–State Routing は New Arpanet、Shortest Path Frst とも呼ばれ、Arpanet だけでなく、ドメイン内経路制御のための ISO の標準として ANSI によっても提案されているアルゴリズムである。

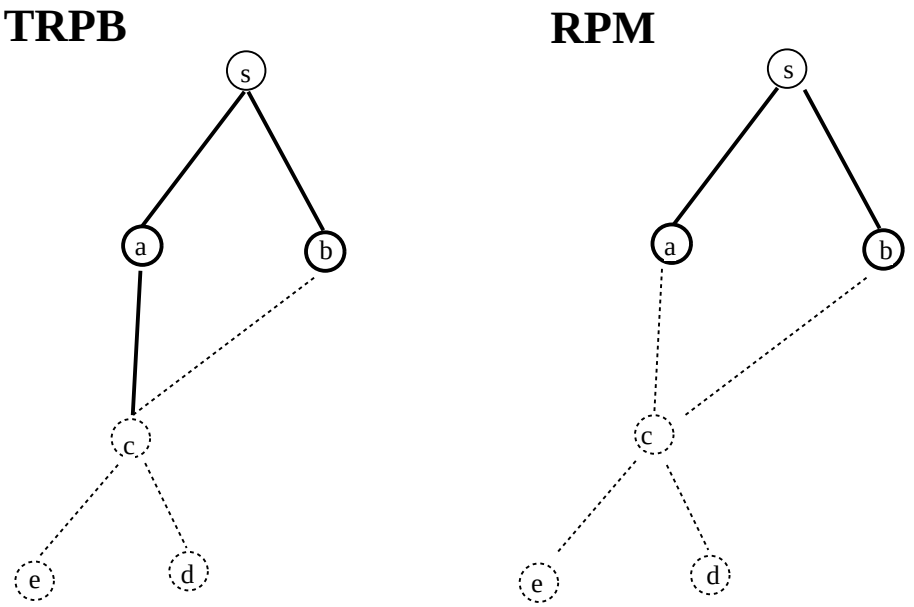

図 2.4:

Link–State Routing では、ルータが直接接続しているリンクの状態 (運用状態、トラフィックの状況等) を監視し、状態の変化が生じた場合には全ての他のルータにブロードキャストを使って報告を行う。このブロードキャストは特別な目的として行われ、優先度が高く全てのルータに即座に伝えられる。そして、どのルータもインターネットワーク上の全てのリンクの状態を知っているため、他のルータと同一の tree を構築することが可能である。

ユニキャスト型の経路制御においては、各ルータを根とする tree をそのルータに近い順に構築し (shortest path first と呼ばれる由縁である)、パケットの転送の際に利用する。このアルゴリズムはリンクの状態の変化、つまりインターネットワークのトポロジーの変化に柔軟に対応できる点が特徴である。詳細については [?] で述べられている。

このアルゴリズムでは、任意のリンクの状態を知っているため、任意のルータを根とする tree を構築することが可能である。すると、マルチキャストのパケットの任意の送信ホストを根とする shortest path tree を構築することも可能であると考えられる。特に、shortest–path multicast routing に応用するためには、各ルータがリンクの状態に関する付加情報として、そのリンク上のグループ所属状況を把握するよう拡張することが必要となる。グループの所属状況に変化が生じたら、ブロードキャストを使って全てのマルチキャストルータに新しい状態を伝えるようにする。従ってグループ所属状況の変化にも柔軟に対応することが可能である。

各マルチキャストルータが直接接続しているリンク上のグループメンバーシップ情報を把握することができるようにするためには、他のアルゴリズムと同様にリンクレイヤブリッジのグループメンバーシップ情報の把握方法を利用する。これは、定期時間毎に O(グループ数) のコストが必要とされる。グループメンバーシップ情報を他のルータに

ブロードキャストする回数を最小限とするために、各リンクのグループメンバーシップ情報の把握は、1 つのマルチキャストルータによって行われることが望ましい。

一般に、全ての送信ホスト毎に全てのグループに対する shortest–path multicast tree は異なる可能性がある。このアルゴリズムでは全ての shortest–path multicast tree を計算することは可能であるが、計算時間や計算結果を保存するのに必要な容量が膨大となるため、RPM と同様、実際に必要とされている shortest–path multicast tree のみを計算し、さらにキャッシュとして保持する方法が考えられている。キャッシュの形式は以下のようになる。

```
(source, subtree, (group,link-ttls),
                  (group,link-ttls),...)
```

source マルチキャストパケットの送信ホスト

subtree その送信者を根とする shortest–path spanning tree 中のこのルータ以下の subtree

group マルチキャストグループアドレス

link–ttls 直接接続しているリンク毎の TTL 値 (生存時間値) のベクトル。このルータからそのリンクを通じて最短で到達するグループメンバーへの最小限必要な TTL を表す。この値が無限大の時には、グループメンバーがいないことを表す。

マルチキャストパケットの転送アルゴリズム

マルチキャストルータがマルチキャストパケットを受け取ったときには、キャッシュに該当するエントリが存在するかを探す。ある場合には、$(group, link - ttls)$ を単位とするリスト中に目標とするグループが存在するかどうかを探す。これも見つかった場合には、link–ttls 欄から得た、各リンクを通じてグループメンバーに到達するのに最小限必要な TTL 値がマルチキャストパケットヘッダの TTL 値に等しいかまたは小さい場合にのみ、そのリンクへパケットを転送する。

$(group, link - ttls)$ を単位とするリスト中に目標とするグループが存在しなかった場合には、subtree の情報を利用して $(group, link - ttls)$ を計算する。新しいグループや link–ttls が得られたらそれもキャッシュに加え、転送に必要な情報として利用する。

キャッシュに情報が存在しない場合には、shortest–path spanning tree を計算し、この情報を新たにキャッシュに加える。キャッシュの内容が飽和した場合には、最近使われていないエントリを消すようにする。そして、ネットワークのトポロジーに変化が生じた場合にはキャッシュの内容を全て消すようにする。グループの所属状況に変更があった場合には、そのグループについての $(group, link - ttls)$ の部分を全て消すようにする。

このアルゴリズムによる経路制御自体のコストは RPM の場合と同様に、インターネットワークにおけるマルチキャストパケットのトラフィックパターンへの依存性が高い。というのは、実際に使われる multicast tree に着目したアルゴリズムであるためである。多

くの送信ホストから頻繁にマルチキャストパケットが送信される場合には、shortest–path multicast tree の計算機会も増加し、計算結果を保持するための記憶領域も広くなることが望ましい。

また、そのような状況ではグループメンバーシップ情報を保持するために必要なメモリの量も増大する。ほとんどのマルチキャストパケットが、それを転送するのに必要とするマルチキャストルータの数が、インターネットワーク上のマルチキャストルータに比較してわずかであるという仮定の元では、このアルゴリズムは RPM よりも少ないメモリ容量で実現できると予想されている。

このアルゴリズムの問題点は次のとおりである。

最初にマルチキャストパケットを転送するときには、キャッシュに情報がないため shortest–path single spanning tree から計算しなければならず、伝搬遅延が大きくなる可能性がある。tree の構築にかかる計算の複雑度は、インターネットワーク上のリンク数に依存する。問題点の解決のためには、ANSI においても提案されているように、インターネットワークをいくつかのドメインに分割し、ドメイン内のリンク数を調整できるようにすることが必要となってくる。

2.1.8　アルゴリズムに関するまとめ

Distance Vector Routing を元にした経路制御アルゴリズム、Link–State Routing を元にしたアルゴリズムともに、各送信ホストから各グループに対する効率の良い shortest–path multicast tree を構築する場合には、グループメンバーシップ情報の把握が必要であることがわかった。これは、マルチキャスト型通信にのみ必要とされる情報である。ユニキャスト型通信でも、ブロードキャスト型通信でも必要とされない情報であるため、グループメンバーシップ情報の把握方法によって、マルチキャスト型通信の効率が左右されるといえる。

既存の経路制御アルゴリズムは、マルチキャスト機能を有する LAN を接続して構成されるインターネットワークに前提を置いて考えられている。しかし、実際のインターネットワークでは、マルチキャスト機能のない LAN も多く接続されており、これらの LAN も考慮した実用的なアルゴリズムを考案することが求められる。

さらに、既存の経路制御アルゴリズムでは、マルチキャスト機能を適用するインターネットワークの範囲が比較的狭い場合を前提として考えられていた。これに関しても、実際のインターネットワーク規模を考慮したアルゴリズムが必要となってくる。

[?] においても、これらの問題点に対する記述がある。そこでは、インターネットワークを幾つかのドメインに分割し、各ドメイン内でそれぞれに適した前述の経路制御アルゴリズムを選択して独立運用を図るべきであると指摘されている。そして、ドメイン間のグループメンバーシップ情報をまとめる、スーパードメインの必要性に関しても記述されている。

2.2 IP マルチキャストアドレス

ここでは、標準的なネットワークプロトコル体系の 1 つである TCP/IP におけるマルチキャスト機能を付加するためのアドレス体系について述べる。

IP アドレスは従来、ホストの各ネットワークインターフェースを識別する目的で利用され、32 ビットで構成される。基本的には内部でネットワーク部とホスト部に領域を分けて利用している。ネットワーク部とホスト部の領域は、ネットワークの規模に応じた 3 つのクラスにより決定される。クラスは図 2.5 に示すように、IP アドレスの先頭数ビットにより識別されるようになっている。ホスト部が全て 1 のアドレスは、ネットワーク部で表現されるネットワークへのブロードキャストアドレスとして割り当てられている。

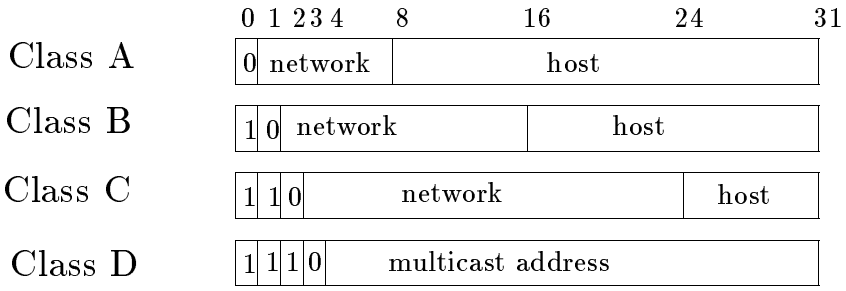

図 2.5: IP アドレスのクラス

このような IP アドレス体系にマルチキャストアドレスを組み入れるために、クラス D のアドレスが導入されている。クラス D とは、IP アドレスの先頭 4 ビットが “1110” となるアドレスで、224.0.0.0 から 239.255.255.255 までの範囲となっている。

マルチキャストアドレスの割り当てに関しては基本的に、アプリケーション毎、あるいはマルチキャストにおいてよく使われるような機能を示すアドレスが予約され、使用されている。例えば、224.0.0.1 は直接接続されるネットワーク上の全てのマルチキャストホスト (IP マルチキャスト機能を有するホスト) を表すアドレスとして割当て済みである。

また、アプリケーションによっては一時的にマルチキャストアドレスを利用するだけで十分な場合もあり得るが[1]、一時的なグループアドレスの割り当てに関しては、特に規定されていない。現状としては、アプリケーションがアドレスの割当てを予約する際に、一時的に利用するマルチキャストアドレスも含めて予約し利用している状態である。

IP マルチキャストアドレス体系において, グループに所属するホストに関する制限は特にないため、送信ホストの予期していないホストが IP マルチキャストデータグラムを受信する可能性がある。したがって、受信ホストの制限を行ないたい場合には、アプリケーション毎に行うようにしなければならない。

また、受信ホストの制限が特にないため、配送ミス等により受信に失敗した受信ホストに関する認識は不可能である。これに関しては次節でも述べるが、送信ホストと受信

[1] 例えば VMTP などでは一時的なプロセスグループを組む場合がある。

ホスト間のデータグラムの信頼性の保証を必要に応じて IP より上のレイヤで行なうことが求められる。

2.3　TCP/IP の各プロトコル層における拡張機能

TCP/IP は、そのネットワークプロトコル体系が 4 つの概念的プロトコル層で構成されている。4 つのプロトコル層をハードウェアに近い順に並べると以下のようになる。

- ネットワークインターフェース層
- インターネット層
- トランスポート層
- アプリケーション層

ここでは、マルチキャスト機能の付加に伴い、これらの各プロトコル層において必要とされる拡張機能の概略を述べる [?]。

まず、マルチキャスト機能は、送信機能および受信機能に分けられる。マルチキャスト機能を有する、あるいは有しないインターネットワーク上のホストは、次の 3 つのレベルに分類することができ、それぞれ以下のように機能を拡張する必要がある。

- **レベル 0** IP マルチキャストをサポートしない。
  クラス D の IP アドレスを使った IP データグラムを捨てるような機能が必要。
- **レベル 1** IP マルチキャストの送信機能のみをサポートする。
  クラス D の IP アドレスを使った IP データグラムを配送できるような機能が必要。送信機能のみを有するので、グループに所属することはできない。
- **レベル 2** IP マルチキャストの送受信機能をサポートする。
  グループメンバーシップ情報の問い合わせや応答をするための機能 (インターネット層。IGMP(Internet Group Management Protocol) の付加) や、ローカルマルチキャスト機能をサポートする機能 (ネットワークインターフェース層) 等の機能拡張が必要である。

次にレベル 2 のホストに必要とされる各プロトコル層および層間 (図 2.6参照) の拡張機能を具体的に示すことにする。

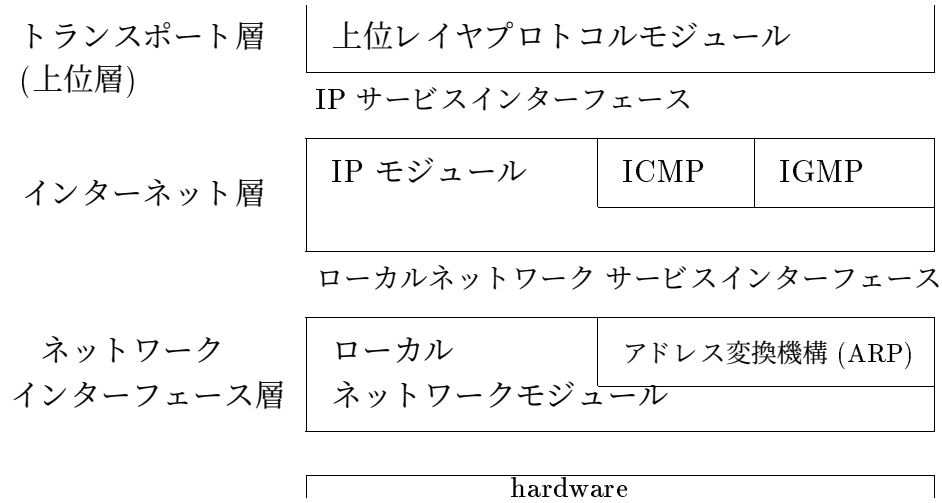

図 2.6: TCP/IP のプロトコル層間関係

2.3.1 IP マルチキャストデータグラム送信機能

IP サービスインターフェースに対する拡張

IP モジュールより上位層が以下の機能を利用できるように拡張される。

- マルチキャストデータグラムの生存時間 (TTL(time–to–live)) の規定。
- マルチキャストデータグラムの転送を行なうネットワークインターフェースの選択
- マルチキャストデータグラムの送信ホストが同じグループに属している場合、送信ホスト自身がそのデータグラムを受信するか否かの選択

IP モジュールに対する拡張

送信の際には、まずクラス D アドレスを認識して経路制御が可能になるように拡張される。 さらに、経路制御を行なうための論理を以下のように変更する必要がある。

IF データグラムの目的地アドレスが同じローカルネットワーク上にある
　　または データグラムの目的地アドレスがクラス *D*アドレスである
THEN ローカルネットワーク上にデータグラムを送信する
ELSE データグラムの目的地アドレスへの次のホップとなるゲート
　　ウェイ (ルータ) にデータグラムを転送する。

ローカルネットワークモジュールに対する拡張

マルチキャスト機能を有するローカルネットワークの場合

イーサネットのようにマルチキャスト機能を有するローカルネットワークに関しては、IP マルチキャストアドレスとローカルマルチキャストアドレスの対応づけを行なう機能が必要とされる。

例えばイーサネットでは、イーサネットマルチキャストアドレスとして、01:00:5E:00:00:00 が予約されているが、このうち下位 23 ビットは IP マルチキャストアドレスの下位 23 ビットを用いる。

この場合、イーサネットマルチキャストアドレスが異なっていても、IP マルチキャストアドレスに対応づけされたときに同一のマルチキャストアドレスとなる可能性がある。

マルチキャスト機能のないローカルネットワークの場合

ブロードキャスト機能がある場合には、IP マルチキャストアドレスをローカルブロードキャストアドレスに対応させればよい。

2 つのホストのみで構成される point–to–point 型のネットワークの場合にはユニキャスト型と同様に送信を行なわなければならない。

また、ARPANET や 公衆 X.25 網のような store–and–forward 型のネットワークの場合には、IP マルチキャストアドレスを、そのネットワーク上の IP マルチキャストルータのローカルネットワークアドレスに対応させ、IP マルチキャストルータに配送を委ねることが考えられる。

IP マルチキャストデータグラム受信機能

IP サービスインターフェースに対する拡張

IP マルチキャストデータグラムを受信する前に、グループに所属する必要があるので、これを IP より上位層から要求するような機能を提供する必要がある。そのためには、グループの所属および離脱に関する以下の 2 つの操作が提供される。

```
JoinHostGroup(group-address, interface)
LeaveHostGroup(group-address, interface)
```

interface については必ずしも指定する必要はない。指定しない場合には IP マルチキャストデータグラムの送信の際に利用されるインターフェースが設定される。

同一ホストの 2 つ以上のインターフェースが同じグループに所属する場合もあり得る。

IP モジュールに対する拡張

自ホストのどのインターフェースがどのグループに所属しているかを示すリストを保持することが必要である。そして、入ってきたデータグラムが以下の条件を満たす場合にはデータグラムを捨てるようにする。

- 所属していないグループ宛の IP データグラム
- グループに所属しているインターフェースと異なったインターフェースから入ってきたデータグラム
- 送信アドレスにマルチキャストアドレスが書かれているデータグラム

IP マルチキャストデータグラムに対する ICMP(Internet Control Message Protocol) エラーメッセージは生成しないようにする。

自ホストに関するグループの把握に関しては、上位層のプロトコルによるグループ所属および離脱に関する要求に応じて情報を更新することが必要となる。

また、ローカルネットワークのグループメンバーシップ情報の把握のために、マルチキャストルータが定期的にグループ所属状況に関する問い合わせを行ない、これに答えるようなプロトコル (IGMP[?]) を実装する必要がある。グループ所属状況に関する問い合わせには、ローカルネットワーク上のマルチキャスト受信機能を有するホストグループ、224.0.0.1 をアドレスとして利用する。

ローカルネットワークサービスインターフェースに対する拡張

どのローカルマルチキャストパケットを受け取り IP モジュールに渡すかを IP モジュールが、ローカルネットワークモジュールに指示することができるように、以下の 2 つの操作を追加する必要がある。

```
JoinLocalGroup(group-address)
LeaveLocalGroup(group-address)
```

但し、同じローカルネットワークモジュールから送信されたパケットは IP モジュールに渡さないようにしなければならない。

ローカルネットワークモジュールに対する拡張

マルチキャスト機能を有するローカルネットワークの場合

マルチキャストを有するローカルネットワークモジュールでは、ローカルマルチキャストアドレスに変換されたパケットを受け取れるようにする必要がある。特に、受け取るべきローカルマルチキャストアドレスを識別できるようなアドレスフィルタリング機構が求められる。

例えばイーサネットの場合、ハードウェアが識別可能なアドレス数が少ないため、ローカルネットワークモジュールのソフトウェアにアドレスフィルタリング機構を入れることが提案されている。

マルチキャスト機能のないローカルネットワークの場合

ブロードキャスト型機能を有している場合には、全てのローカルマルチキャストパケットを IP モジュールに渡すことになる。

point–to–point 型や store–and–forward 型の場合には、ローカルネットワークアドレスはユニキャストアドレスとなっているため、特に追加すべき機能はない。

このように、実際のインターネットワーク上では、マルチキャスト型機能のない物理媒体も十分考慮に入れてマルチキャスト機能を考えていく必要がある。次に、このように TCP/IP にマルチキャスト機能が追加された環境で実装された、既存の IP マルチキャスト経路制御アルゴリズムについて述べ、問題点を指摘する。

IP header	IGMP header	DVMRP message

図 2.7: DVMRP データグラムの構成

2.4　既存の IP マルチキャスト経路制御アルゴリズム (DVMRP)

設計段階では、前述の RPM アルゴリズムを採用しているが、現状の実装としては、TRPB アルゴリズムを元にしている。ここでは、`MUTICAST 1.2 Release` における `mrouted` の実装について述べる [?]。

実際に運用されている TCP/IP インターネット上において混乱なく TRPB アルゴリズムを実装可能とするために、以下の工夫がなされている。

- IP マルチキャストを認識しない (2.3で述べた拡張機能を持たない) ルータがあっても対応可能とする。
  −→ IP マルチキャストを認識しないルータを通過する際はトンネリングという技術を用いて IP マルチキャストデータグラムを一時的に IP ユニキャストデータグラムに変換する。

- ユニキャストの経路制御アルゴリズムで Distance Vector アルゴリズムを採用していないルータへの対応も可能とする。
  −→ ユニキャストの経路制御で得られる経路制御テーブルとは独立に DVMRP の経路制御テーブルを持つ。

DVMRP は、IP マルチキャストデータグラムの配送に必要な経路制御情報を得るためのプロトコルであり、グループメンバーシップ情報の把握は別のプロトコル、IGMP によって行なわれる。ここでは、トンネリング技術や DVMRP による経路制御テーブルの作成アルゴリズムについて述べるとともに、具体的な転送アルゴリズムについても述べる。 そして、DVMRP の運用上の問題点について挙げる。

2.4.1　DVMRP で交換される情報の形式

現段階の実装では、DVMRP で行なわれる情報交換では、IGMP により規定されたデータ形式 [?] が用いられる。IGMP メッセージは IP データグラムの中にカプセル化されているため、DVMRP データグラムの構成は図 2.7 のようになっている。

DVMRP メッセージ部分は名札付きデータのストリームから構成される。名札付きデータとして定義されているコマンドは、表 2.3で示す通りである。経路情報は、これらのコマンドを組み合わせることによって構成される。通常、DA1 つにつき、Metric・Infinity・Flags0・Subnetmask コマンドをセットして 1 つの経路を表すことになっている。

表 2.3: DVMRP メッセージ中で扱われるコマンド

Value	Name	
0	Null	Padding に使われる。
2	AFI	address family を示す。
3	Subnetmask	subnet mask を示す。
4	Metric	metric の数値を表す。
5	Flags0	split horizon などの flag を表す。
6	Infinity	無限大を表す数値を入れる。
7	DA	destination address
8	RDA	requested destination address
9	NMR	non membership report
10	NMR-cancel	non membership report cancel

2.4.2 トンネリング

マルチキャストルータ間に、IP マルチキャスト機能をサポートしていないルータが存在する場合には、IP マルチキャストデータグラムを一時的にユニキャストデータグラムの中にカプセル化し、IP マルチキャスト機能をサポートしていないルータを経由する。これがトンネリングと呼ばれる技術である。

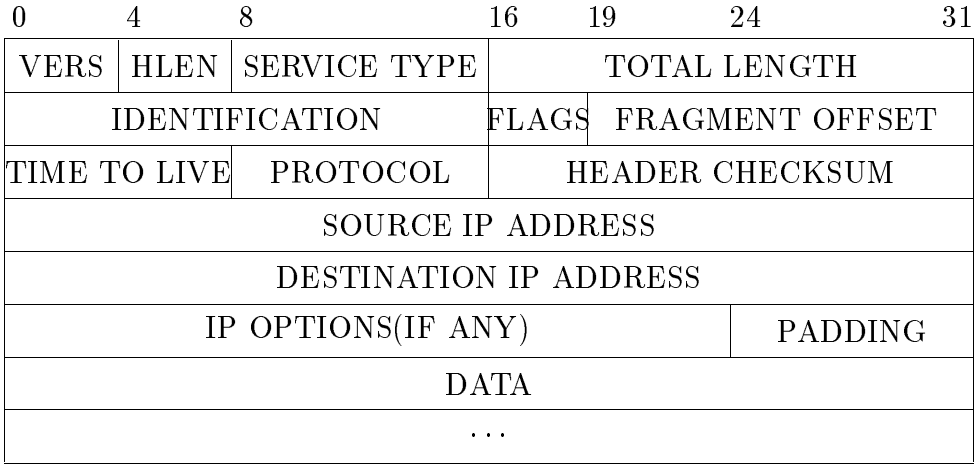

図 2.8: IP データグラムのヘッダの構成

ユニキャストデータグラム中へのカプセル化には、IP データグラムヘッダ (図 2.8) に付加することができる既存のオプションのうち、loose source route を用いる (図 2.8)。本来、送信者が loose source route は IP データグラムの経路を、あらかじめ指定できるよ

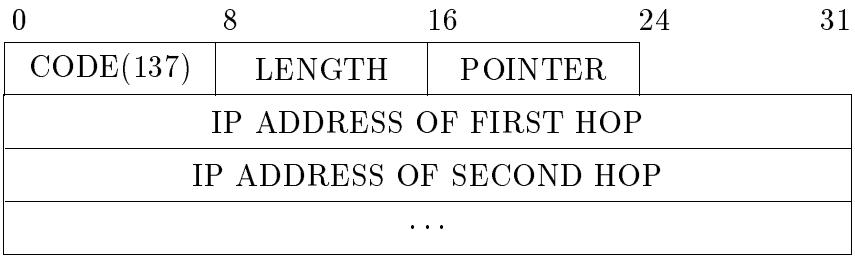

図 2.9: IP loose source route オプションの構成

うにするためのオプションである。“loose” というのは、source route として連続して書かれた経路を通過する時、複数のネットワークを経由しても構わないという意味である。マルチキャストのトンネリングでは、IP マルチキャストデータグラムの送信ホストアドレス、目的ホストアドレス欄を退避させるためにこのオプションを利用する。

トンネルには、トンネルの終端と距離、敷居値がある。トンネルの終端のルータは、ローカルとリモートで区別される。トンネルのローカルの終端ルータで、IP マルチキャストデータグラムをカプセル化する方法は次のようになる (これは、IP モジュールで行なわれる) 。

1. マルチキャストデータグラムに 2 つの loose source route IP オプションを入れる。
2. source route pointer を source route の 2 番目の要素を示すように設定する。(図 2.10(1))
3. source route の 1 番目の要素は、マルチキャストデータグラムの本来の送信ホストアドレスを入れる。(図 2.10(2))
4. source route の 2 番目の要素は、マルチキャストデータグラムの本来の目的ホストアドレスを入れる。(図 2.10(3))
5. マルチキャストデータグラムの送信ホストアドレスには、トンネルのローカルの終端となるルータのアドレスに置き換える。(図 2.10(4))
6. マルチキャストデータグラムの目的ホストアドレスには、トンネルのリモートの終端となるルータのアドレスに置き換える。(図 2.10(5))

トンネル上で配送エラーが生じた場合には、ICMP エラーメッセージはトンネルのローカルの終端ルータへ送られることになる。

一方、トンネルのリモートの終端で、カプセル化されたデータグラムを受け取った後の loose source route オプションの処理は次のようになる。

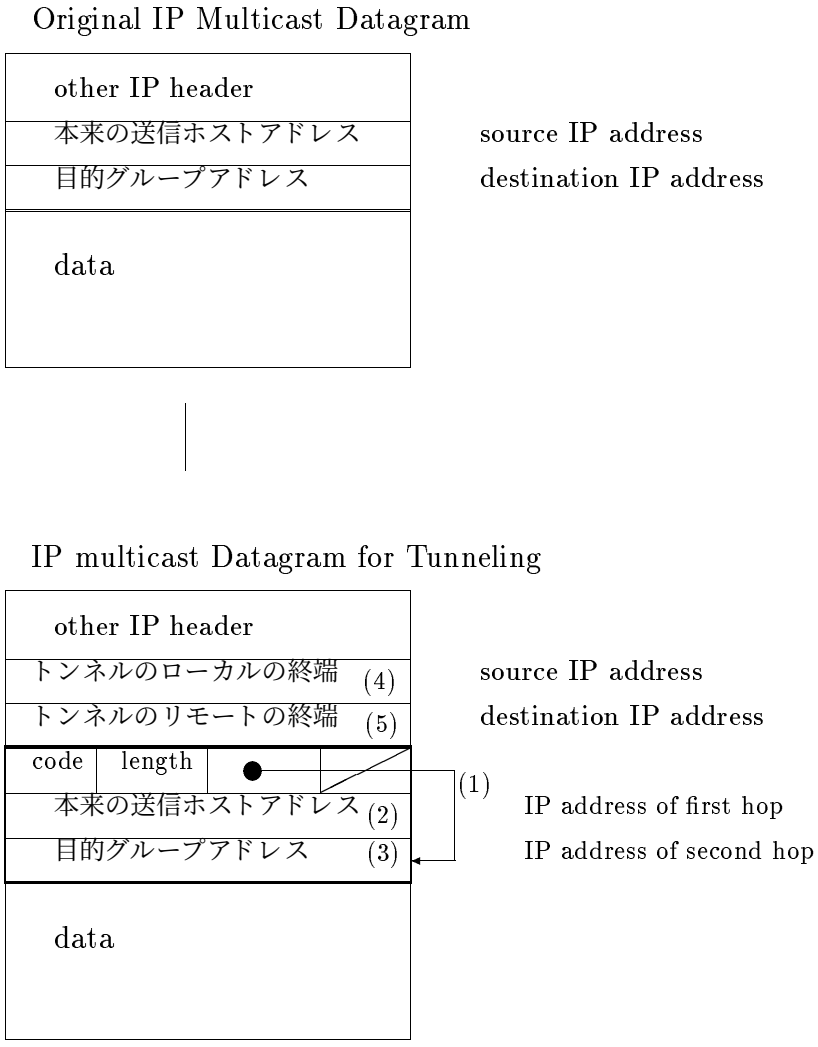

図 2.10: トンネリングにおけるアドレス操作

1. マルチキャストデータグラムの送信ホストアドレスを loose source route に書かれた 1 番目の要素に置き換える。

2. マルチキャストデータグラムの目的ホストアドレスを loose source route に書かれた 2 番目の要素に置き換える。

3. loose source route オプションを取り除く。

トンネル内の、マルチキャスト機能をサポートしていないルータでは、データグラムの目的ホストアドレスが自分でないので、loose source route オプションを見ない。 トンネルのリモートの終端ルータでは、データグラムの目的ホストアドレスが自分であるので、loose source route オプションを見て上記のような処理を行なう。通常の loose source route では、マルチキャストアドレスを書かないことになっているので、トンネルのリモートの終端ルータはこのデータグラムがトンネルを通過してきたことが判断できる仕組みとなっている。

2.4.3　経路制御テーブルの作成

ここでは、経路制御情報の交換方法と、経路制御テーブルの作成方法に関して述べる。

DVMRP は、Distance Vector を元にしているので基本的にはユニキャストの経路制御アルゴリズムとしてよく使われている RIP[?] と似た経路制御情報の交換を行なう。しかし、RIP では目的地アドレスとしているエントリが DVMRP ではデータグラムの送信ホストを意味している点で異なっている。

これに加えて、DVMRP では TRPB アルゴリズムの実装のために、child link および leaf link の識別に必要な情報を集める必要があるため、経路制御テーブルのエントリは以下のような構成となる。

- 目的地アドレス (マルチキャストデータグラムの送信ホスト)
- 目的地アドレスのサブネットマスク
- 目的地アドレスへ配送するための次のルータ
- 次のルータへ到達するためのインターフェース
- child となるインターフェース (child link)
- leaf となるインターフェース (leaf link)
- 各インターフェースの親ルータ (child link の識別に使う)
- 各インターフェースで直接接続されているルータのうち、自分を親と考えているルータ (leaf link の 識別に使う)

- エントリの古さを検出するためのタイマー
- そのルートのエントリの状態を示すフラグ
- 距離
- 無限大値

経路制御情報の送信方法

マルチキャストルータは、DVMRP データグラムを全てのインターフェースに定期的に送信する。この時 DVMRP データグラムの生存時間 (TTL) は 1 とする。

また、以下の状況では定期時間間隔以外にも経路制御情報を送信する。

- 経路に変更が生じた時
  その度に、その経路についての情報を送る。この情報だけでネットワークが溢れないように、わざと遅延をつけて送信する必要がある。
- ルータが再起動され経路制御情報の要求が送られた時
- ルータが DVMRP の実行を終える時
  その経路の全てに対して距離を無限にした経路情報を送ることが望ましい。

マルチキャスト機能をサポートしているルータに送信する時には、マルチキャストアドレス 224.0.0.4 を用いる。トンネルへ経路変更の情報を送る時はユニキャストデータグラムを使ってトンネルのリモートの終端ルータに送られるようにする。

また、DVMRP メッセージで送信する経路の距離に関しては、leaf link の識別のために、poisoned split horizon の処理が行なわれている必要がある。これは、TRPB の説明のところでも述べたが、ネットワーク X を使う経路がある時、ネットワーク X へその経路の情報を送信する時に距離を無限大とする方法である。

経路制御情報の受信方法

経路制御情報を受信したマルチキャストルータは、その経路制御情報を元に経路制御テーブルを更新する。その際、経路制御情報の内容以外に、その経路制御情報がどのインターフェースから届いたかという情報も利用される。経路制御情報を受け取った時の処理は、次のようなアルゴリズムで行なわれる。

IF そのルートへの距離が与えられている
THEN メッセージが届いたインターフェースの距離を加える
経路制御テーブル内のその経路の *destination address* を探す

IF その経路がない
THEN 同じネットワークに対する経路が存在するかを探す
　　IF それがある
　　THEN

IF　前と同じルータから届いた情報である
THEN　次の経路に対する処理をする。
IF　その経路が無限大の距離になっていない
THEN　経路制御テーブルにその経路を加える
次の経路について続ける

IF　この経路が同じルータから届いている
THEN　経路のタイマーを消す
IF　距離を受け取って、前の経路の距離が変わっている
THEN　その経路の距離を新しい距離に変える
IF　新しい距離が無限大になっている
THEN　その経路に、*EXPIRATION_TIMEOUT*のタイマーにかける
次の経路について処理を続ける
IF　無限大の数値が前の無限大の数値と違っている
THEN 新しい無限大の数値に変え、距離については、
min(新しい無限大の数値、前の距離) にする。

ELSE　(同じルータから届いた情報ではない時は、)
IF　新しく書かれた距離が前のよりも小さい、
または
(経路のタイマーが *EXPIRATION_TIMEOUT*の少なくとも半分経過している
かつ 前の距離と同じである
かつ 無限大になっていない)
THEN　新しい距離に変える。その経路にかけていたタイマーを止める。
次の経路について処理を続ける

隣接ルータについて
マルチキャストルータは、直接接続しているネットワーク上の隣接ルータのリストを保持し、定期時間内にこれらのルータから経路情報が届かない場合には、その隣接ルータが停止していると判断するようにする。

2.4.4　IP マルチキャストデータグラムの転送アルゴリズム

DVMRP では、TRPB を元にした経路制御アルゴリズムを採用しているため、ルータは、受け取った IP マルチキャストデータグラムの送信ホストから見て、直接接続しているネットワークが child link であって leaf link でない時、データグラムの転送を行なう。
child link および leaf link 欄の操作は、経路制御情報を受け取った際に行なわれ、転送アルゴリズムは次のようになる。

IF　*IP*の *TTL* が 2 より小さい

THEN 次のデータグラムを処理する
*IP*データグラムの送信ホストへの経路を探す
IF 見つからない
THEN 次のデータグラムを処理する
IF データグラムの入ってきたインターフェースが
経路に書かれている $next - hop virtual interface$ と異なっている
THEN 次のデータグラムを処理する

IF データグラムが *tunnel*を通って来た

THEN *tunneling*の説明で出てきたように、送信ホストと目的地ホストを
loosesourceroute のところに書かれたアドレスに置き換え、
*IP*データグラムのオプション欄、ヘッダの長さを調整する。
IF データグラムの目的地アドレスが 224.0.0.0
または 224.0.0.1 である
THEN 次のデータグラムを処理する

FOR 各インターフェース V に対して以下の処理を行なう
IF Vがデータグラムの *source* にとっての *childlist* に含まれている
THEN
IF Vが送信ホストにとっての *leaf*でない
または *leaf*だが Vに目的グループのメンバーがいる
THEN
IF TTL が Vで決められている値よりも大きい
THEN TTL の値を 1 減らしてインターフェース Vから
データグラムを転送する。

2.4.5 DVMRP の問題点

TRPB アルゴリズム自体の問題点、Distance Vector 経路制御アルゴリズムの問題点の他に、実際のインターネットワーク上で DVMRP を運用する際の問題点について考える。

まず、DVMRP は、基本的にマルチキャストルータが IP マルチキャストデータグラムを受け取った時の転送に関するプロトコルであるので最初に IP マルチキャストデータグラムの送信に関しては厳密な規定がない。最初に IP マルチキャストデータグラムを送信する際の経路決定では経路制御テーブルを見ず、あらかじめ指定されたインターフェースを通しているため、無駄な経路制御を行なっている場合がある。

これについての解決方法の 1 つは ICMP Router Discovery Messages の RFC [?] で述べられているが、IP マルチキャス トデータグラムを送信する際にも DVMRP によって得

た経路制御テーブルを利用できるように、マルチキャスト機能を有するレベル 1 のホストについても経路制御テーブルを与える必要があると考えられる。これによって、データグラムの最初の送信に関しても最適な経路選択をすることが可能となる。DVMRP は、本来ユニキャスト型の Distance Vector の経路制御アルゴリズムと融合させることにより、経路制御情報によるオーバーヘッドを軽減することが期待されているが、これが実現されれば IP マルチキャストデータグラムの送信に関しても柔軟な対応が可能となるかも知れない。

また、DVMRP ではマルチアクセス型ではない物理媒体として、point–to–point 型や store–and–forward 型についても考慮されているため、種々の物理媒体を接続して構成される一般的なインターネットワークにおいてもマルチキャスト機能を提供することが可能となっている。しかし、DVMRP で仮定している物理媒体はいずれも通信方向が双方向であると考えられる。マルチアクセス型の物理媒体の中には、衛星通信型のように通信方向が一方向に限られている媒体も存在する。衛星通信型の媒体は、距離の大きいマルチキャストルータ間を直接接続することが可能であるため、IP マルチキャストを広域に拡張する際には重要な媒体として利用されることが十分に考えられる。

このような媒体を含んでいる場合、特に reverse–path–algorithm を利用しているため、reverse path が全く認識されないという問題が起こり得る。具体的にどのような問題が起こるかについては次章以降に譲ることにする。

第 3 章

WIDE における IP マルチキャストの実装・実験

この章では WIDE における IP マルチキャスト通信の実装および実験について述べる。とくに、マルチキャスト機能を広域に拡張する場合の経路制御ではどのような点に注意すべきか、また既存の経路制御アルゴリズムではどのような点が問題となるかを示す。そして、新しい経路制御アルゴリズムを考える際に考慮すべき点としてまとめることにする。

3.1 広域 IP マルチキャストの必要性

マルチキャストを利用して有効となるアプリケーションについては、分散データベースの更新、分散ファイルシステム、ニュースシステム、FTP の代替等、いろいろ考えられるが、このようなアプリケーションの中には、例えばニュース配送のように、大規模な範囲に渡り大量のデータ転送を行なうようなアプリケーションがある。現状としては、大規模な範囲に渡って同一の情報を複数の目的地へ転送するようなアプリケーションは多くはない。しかし、このようなアプリケーション、特にニュース配送においては、それ 1 つだけでも全体のネットワークのバンド幅を消費し得るため、このようなアプリケーションに対してマルチキャストを適用すると、ネットワーク全体のバンド幅消費の減少効果も高くなると予想される。

また、ニュース配送によって転送されるデータは、現在テキストデータが中心であるが、今後は、音声・画像等、テキストと比較してデータ量が飛躍的に増大するようなデータ転送の必要性も高まってくると思われる。それに従い、マルチキャストの利用がさらに重要となる。

したがって、IP マルチキャストを広域に拡張することが必要となってくる。具体的にここでは、TCP/IP により接続されている組織の集合状態を広域として捉え、定量的にはクラス B アドレス数十個程度からなる範囲を「広域」として定義する。例えば日本全域へのマルチキャストの適用を想定する。

3.2 LAN 拡張型のマルチキャストとの相違点

一般に、計算機ネットワークにおいては、距離の近いホスト間を接続する媒体の提供する通信速度は、例えばイーサネットの場合は 10Mbps であるように高速であり、距離

の遠いホスト間を接続する媒体では、専用線の数百 Kbps のように、低速である。したがって広域のネットワークは、高速のネットワーク間を低速のネットワークを用いて接続する形態となっている。低速のネットワークに高速のネットワークと同様の効率で通信を行うことは期待できないのは明らかである。従って、低速のネットワークには負荷ができるだけ小さく、最低限のトラフィックで済むような工夫が広域ネットワークを利用する際には必要となってくる。

実際的には低速のネットワークを利用するのは、組織間を接続する場合が多い。例えばユニキャストの経路制御アルゴリズムにおいては、経路制御情報の scalability(大規模性) を改善する目的で、組織間で交換される経路制御情報をネットワーク単位にまとめる等の工夫がなされているが、結果的には低速のネットワークへの負荷を小さくすることにもなっていると考えられる。マルチキャストの場合も同様に、広域になるとマルチキャストデータグラムを送信する可能性のあるホストの数が増えるため、交換する経路制御情報の量は増大する。したがって、経路制御情報の scalability の改善と共に、低速のネットワークへの対応をも考慮することが必要となってくる。つまり、組織間で交換する情報の省略化を考慮する必要がある。

また、広域化にともなって、送信ホストからグループメンバーホストへの距離が広がり、送信ホストからグループメンバーホストに到達するまでの中継を行うマルチキャストルータの数も増大する。同一 LAN に接続されていないマルチキャストルータ間は基本的には point–to–point で接続されているため、LAN のマルチキャストと比較して効率が落ちると考えられる。つまり、マルチキャストにかかる通信コストは大きくなると予想される。これは逆に、point–to–point で接続されている距離の離れたマルチキャストルータ間をマルチキャスト型媒体で直接接続することが可能となれば、広域に渡るマルチキャストの効率を良くする可能性が出てくると判断することもできる。

さらに、不必要なトラフィックを与えることによる影響も広域になるに従って大きくなる。例えば、広域に渡るブロードキャストはネットワークのふく湊に大きな影響を与える為、ほとんど利用されることがない。広域のマルチキャストでは、ブロードキャストとの差をつけるためにも、グループメンバーホストの存在しないネットワークにマルチキャストデータグラムを配送しないような仕組みが必要である。

LAN、あるいは LAN を拡張したマルチキャスト機能との相違点をふまえた上で、広域マルチキャスト機能に適した経路制御アルゴリズムに必要とされる条件を以下にまとめる。

- 経路情報の scalability の改善
  ルータが格納する経路制御テーブルが管理しきれないほど大きくならないように経路制御情報をまとめること。

- 低速ネットワークへの考慮
  低速ネットワーク (組織間の接続に通常用いられる) には最小限のトラフィック量で済むような工夫をすること。

- 中継ルータに関する工夫
  中継ルータ数を少なくする、あるいは中継ルータ間を直接接続することが可能な媒体を利用すること。

- グループメンバーシップ情報の有効利用
  グループメンバーシップ情報交換の範囲を拡張し、グループメンバーホストに関係のないネットワークの範囲を正確に把握して、不必要なトラフィックが少なくなるような工夫をすること。

次に、既存の経路制御アルゴリズムを広域に拡張した場合の問題点を前述の条件等に当てはめて追求する。

3.3 既存の経路制御アルゴリズムを広域に拡張した場合の問題点

Reverse–Path Forwarding アルゴリズム、特に TRPB を元にしたマルチキャストの経路制御アルゴリズムを広域に拡張しようとした場合、以下の問題点を指摘することができる。

- 構築されるマルチキャスト tree が大きくなる。
  TRPB では、ブロードキャスト tree の葉の部分を切り詰めるだけであるので、マルチキャストデータグラムの送信ホストを根とするマルチキャスト tree はほとんどブロードキャスト tree に近い。ブロードキャスト tree の葉の数が少なく、グループメンバーホストが送信ホストに十分近い場合には、不必要なマルチキャストデータグラムが広範囲に及ぶため、ネットワークのバンド幅を消費してしまう。また、中継ルータ数もマルチキャスト tree が大きくなるにつれ、増大する。

- 交換される経路制御情報の量が増大する。
  マルチキャストデータグラムを送信するホスト数が増大するため、これらのホストへの reverse–path の情報量が増大する。これらの情報交換は隣接ルータ間で行われるが、隣接ルータ間がどのようなネットワークで接続されているか等についての考慮はなされていない。隣接ルータ間で交換される経路情報量はネットワーク数 (マルチキャストデータグラムの送信ホスト数) に比例して多くなるため、低速ネットワーク、あるいはスループットの小さいネットワークでは、経路情報の交換自体による負荷も高くなる。

これらの問題点は、広域マルチキャストのための経路制御アルゴリズムに必要な前述の条件のうち、経路制御情報量の scalability、低速ネットワークへの考慮、不必要なトラ

フィックの防止という点を満たしていない。ということは、既存の経路制御アルゴリズムは広域マルチキャストの経路制御アルゴリズムとしては、適しているとは言えない。

したがって、広域マルチキャスト機能を提供するためには、少なくとも既存の経路制御アルゴリズムに改良を加えることが必要である。

3.4　広域マルチキャストに適した物理媒体の利用

広域マルチキャストに適した経路制御アルゴリズムに必要な条件として、中継ルータに関する工夫を挙げたが、特に中継ルータ間の通信を効率よくするための工夫についてここでは述べ、新しい経路制御アルゴリズム考案のためのヒントとする。

前述したように、マルチキャスト型の LAN によって直接接続されていない中継ルータ間を、別の特別なマルチキャスト型媒体で直接接続することが可能であれば、以下の点で広域ネットワークに対して影響を与えると予想される。

- 特別なマルチキャスト型媒体を利用することによって、それらのルータ間を接続する従来のネットワークが、マルチキャストによるトラフィックの影響を受けない。もし、従来のネットワークが低速である、あるいはバンド幅やスループットが小さい場合には、マルチキャストデータグラム以外のデータグラムの伝搬に影響を与えずに済む可能性がある。マルチキャストデータグラムのデータ量が大量である場合には特に、従来のネットワークを回避することによるアプリケーションの実行効率への影響が大きくなると予想される。

- 特別なマルチキャスト型媒体で接続することによって、送信ホストから中継ルータへの距離が、従来の距離よりも短縮される場合には送信ホストからグループメンバーホストへの距離が小さくなり、マルチキャスト自体の効率が改善される可能性がある。

- 特別なマルチキャスト型媒体を利用すると、従来のマルチキャスト tree が再構築され、特別なマルチキャスト型媒体によって接続されたマルチキャストルータは、マルチキャスト tree 中に並列に位置することになる。従来のマルチキャスト tree でそれらのマルチキャストルータが上下方向に位置していた場合にはマルチキャスト tree 全体の高さを低化させる場合もある。

  特別なマルチキャスト型媒体を経由することで従来のマルチキャスト tree よりも低くなり、しかも従来の経路よりも伝搬遅延が小さくなる場合には、マルチキャスト tree の高さが、マルチキャストの伝搬遅延に対応しているということが言える。ということは、ある送信ホストから最も距離の離れたグループメンバーホストとの伝搬遅延が小さくなると考えることができる。マルチキャストの RTT(Round Trip Time) を

「送信ホストから最も距離の離れたグループメンバーホストとの RTT」

と定義すれば、送信ホストにとってマルチキャストデータグラムの RTT を小さく見積もることができるようになる

これは例えば、送信ホストがマルチキャストデータグラムの到達を確認してから、次のマルチキャストデータグラムを送信するような場合や、送信ホストとグループメンバーホスト間で、データの信頼性を保証し再送を必要とするような場合にマルチキャストの効率を左右する上で重要になると考えられる。

このような特徴を持つ特別なマルチキャスト型媒体としては、衛星通信型の物理媒体が考えられる。したがって、広域マルチキャストを提供するために中継ルータ間をサテライト型の物理媒体で接続することを考える。衛星通信型の物理媒体を利用するに当たり、経路制御アルゴリズムにこの媒体を組み入れる際の留意点は以下の通りである。

- 単方向性であること。
- 送信専用ホストの数は、受信専用ホストの数に比して少ないこと。
  したがって、衛星通信型の物理媒体を利用する場合には、まず送信専用ホストへの経路を確保することが必要である。

したがってまず既存のマルチキャスト経路制御アルゴリズムが一方向性の物理媒体を含んだ場合でも十分に機能するかどうかを検証する。

3.5 一方向性のマルチキャスト型媒体を含んだ既存の経路制御アルゴリズムに関する検証

ここでは、既存のマルチキャスト経路制御アルゴリズムにおいて、ネットワークとして一方向性のマルチキャスト型媒体を含んだ場合にどのような問題点が生じるかについて考察する。さらに、その場合でも十分に機能するように、既存の経路制御アルゴリズムの改良を提案し、問題点について考察を行う。

まず、一方向性のマルチキャスト型媒体が DVMRP のアルゴリズムではどのように扱われるかについて順番に考えてみる。

3.5.1 前提

ここでは、以下のような前提を設ける。

- 一方向性のマルチキャスト型媒体で接続されているホストはマルチキャストルータである。
- 一方向性のマルチキャスト型媒体で接続されているホスト間には、他の双方向経路も存在する。

- 一方向性のマルチキャスト型媒体の受信専用ホストは、任意の送信ホストに対して、少なくとも leaf link でない child link を持つ。つまり、受信専用ホストは常に parent router となるリンクを持っているようにする。

3.5.2 DVMRP の個々の機能に一方向性のマルチキャスト型媒体を適用した時の影響

- 隣接したマルチキャストルータ間での経路制御情報の交換
  一方向性のマルチキャスト型媒体では、送信専用ホストから受信専用ホストへの方向のみが隣接マルチキャストルータとして把握される。
  つまり、受信専用ホストは送信専用ホストを隣接ルータとして把握できるが、送信専用ホストは受信専用ホストを隣接ルータとして把握することが不可能である。

  また受信専用ホストは、他の隣接マルチキャストルータに、一方向性のマルチキャスト型媒体を利用する経路をアナウンスすることはできない。

- child link の識別
  一方向性のマルチキャスト型媒体を child link として把握するホストは、データグラムを送信することが可能な送信専用ホストのみである。したがって、マルチキャストデータグラムの送信者によらず、常に送信専用ホストが paret router となる。

- leaf link の識別
  一方向性のマルチキャスト型媒体の parent router である送信専用ホストは受信専用ホストから経路情報を得ることができないので、送信専用ホストを next hop router とするような経路は存在しないと判断する。したがって、マルチキャストデータグラムの送信者によらず、常に送信専用ホストから見て、一方向性のマルチキャスト型媒体は leaf link である。

- local group membership の把握
  送信専用ホストから受信専用ホストに対しては、グループメンバーシップについての問い合わせをすることは可能であるが、それに対する解答を得ることができない。したがって、送信専用ホストは一方向性のマルチキャスト型媒体について常にグループメンバーがいないと判断してしまう。

3.6 問題点および解決方法

3.6.1 問題点その 1

衛星通信型の媒体上のホストに関する local group membership で、常にグループメンバーがいないと判断するので、一方向性のマルチキャスト型媒体にマルチキャストデー

タグラムが流れない。

3.6.2 改良点その 1

送信専用ホストはグループメンバーシップの把握を行なわずデータグラムを受け取ったら必ず受信専用ホストに (衛星通信型の媒体を必ず利用して) データグラムを転送するように改良する。

3.6.3 問題点その 2

「改良点その 1」により、受信専用ホストは、通常の経路制御情報から得られる経路からと、一方向性のマルチキャスト型媒体の送信専用ホストからの 2 つのデータグラムを必ず受け取ることになる。

ところが、一方向性のマルチキャスト型媒体から受け取ったデータグラムは、通常受け取るべきインターフェースからのデータグラムではないため、捨てられる。

3.6.4 問題点解決へのアプローチ

既存の経路制御アルゴリズム (reverse path を扱うアルゴリズム) をそのまま利用すると、 一方向性のマルチキャスト型媒体は有効な経路として認識されないということが明らかとなった。

既存の経路制御アルゴリズムを元に、これらの問題の解決方法を考える際に目標とする点は以下のようになる。

- 通常の経路と比較して、一方向性のマルチキャスト型媒体を通る方が有利な場合 (ホップ数が減るなど) には、通常の経路からのデータグラムの方を捨てるようにする。
- 一方向性のマルチキャスト型媒体から受信専用ホストがデータグラムを受け取った後の経路に関しては、通常の経路制御情報から得られる経路をそのまま使うようにする。したがって、通常の経路の情報を保つことができるように、経路制御情報交換のアルゴリズムに関して変更がないようにする。

3.6.5 解決方法その 1

まず、一方向性のマルチキャスト型媒体の受信専用ホストは、すべてのマルチキャストデータグラムを一方向性のマルチキャスト型媒体だけから受け取るようにする。

この時は、通常の経路から受け取ったデータグラムは捨てる。そして、マルチキャストデータグラムの転送に関しては、通常の経路制御情報交換から得られるルーティングテーブルを参照して決定する。

これの問題点は次のとおりである。

一方向性のマルチキャスト型媒体の送信専用ホストを経由した経路よりも、通常の経路の方が距離が短い (ホップ数が小さい) 場合には、一方向性のマルチキャスト型媒体の受信専用ホスト以下への伝搬遅延が大きくなる。つまり、ホップ数等を基準にした柔軟な経路制御が困難になると考えられる。

3.6.6　解決方法その 2

解決方法その 1 よりも、柔軟に経路を決定できるように改良を加える。改良点は、一方向性のマルチキャスト型媒体を経由した時の距離と従来の経路の距離を比較し、距離の短い方を選択するという点である。

(1) 通常の経路制御情報交換として、一方向性のマルチキャスト型媒体の送信専用ホストが受信専用ホストに送るルーティングテーブルを、受信専用ホストが経路制御決定の材料にはせず、独立して保存する。保存の際に、ルーティングテーブルの距離の欄に、一方向性のマルチキャスト型媒体を通過するコストを足しておく。これにより、一方向性のマルチキャスト型媒体を経由した場合の距離を知ることが可能となる。

送信専用ホストのルーティングテーブルを保存する方法を簡単にするために、送信専用ホストから受信専用ホストにルーティングテーブルを送るということを、一方向性のマルチキャスト型媒体での、グループメンバーシップの方法として改めて定義する。

(2) そして、一方向性のマルチキャスト型媒体の受信専用ホストがデータグラムを受け取った時に、送信専用ホストのルーティングテーブルと、自分のルーティングテーブルを用いた以下の比較を行なう。

IF　マルチキャストデータグラムの入ってきたインターフェースが、
　一方向性のマルチキャスト型媒体のインターフェース
　または ルーティングテーブルから得られる正しいインターフェースである
THEN マルチキャストデータグラムの送信者からの距離について
　　IF (マルチキャストデータグラムを受け取ったインターフェースを
　　経由した場合の距離)
　　≤ (もう一方[1] のインターフェースを経由した場合の距離)
　　THEN　通常のルーティングテーブルを見て、転送する
ELSE データグラムを捨てる

これの問題点は次のとおりである。

データグラムが来る度に距離の比較を行なうので、インターフェースの選択の際のオーバーヘッドが大きくなる。同一の送信者からマルチキャストデータグラムが連続してくる場合は、毎回比較を行なわなくてもすむように比較の結果を保存しておくような工夫も必要となる。

[1]「一方」とは、マルチキャストデータグラムを受け取ったインターフェースが一方向性マルチキャスト型媒体に対応する場合は通常の経路制御から得られるインターフェース、通常の経路制御から得られるインターフェースである場合には、一方向性マルチキャスト型媒体へのインターフェースということである。

解決方法その 1、その 2 では、既存のルーティングアルゴリズムで一方向性の媒体を利用するように改良する方法を提案した。しかし、既存のルーティングアルゴリズムでは、全てのマルチキャストルータは必ずマルチキャストデータグラムを受け取るので、全体のトラフィック量に関しては、一方向性の媒体を利用しない場合と変化がない。

理想としては、一方向性の媒体を利用することによって、全体のトラフィック量が減ることが望ましい。全体のトラフィック量を減らすためには、一方向性の媒体を経由してデータグラムを受け取るルータが、他の媒体からデータグラムを受け取らないようにする工夫が必要になる。これに関する解決へのアプローチとして考えられることは、一方向性の媒体を経由してデータグラムを受け取るルータが、他の媒体からの経路を断ち切り、一方向性の媒体を中心とする地域を形成することである。 このためには、経路制御情報の交換を、特定のリンクに対して制限するような工夫が必要となる。

DVMRP は、送信ホスト毎に柔軟な経路を動的に構築する。もし、地域分けを行なうとすると、一方向性の媒体を中心とする地域の境界の決定も動的となるはずである。よって、DVMRP で構築される動的な経路に対して、経路制御情報交換の制限を行なうためには、複雑なプロトコルが必要となり、DVMRP の大幅な改良が必要となることが予想される。

一方向性の物理媒体を利用するために、DVMRP への適用を考えてきたが、一方向性の物理媒体を有効に利用しようとすると、問題点が多く、それに対する改良点も多くなることがわかった。したがって、一方向性の媒体を利用する場合には、その媒体の部分だけは固定的な経路となる点を考慮し、一方向性の媒体を有効利用することを前提とした、新しいプロトコルの考案を試みる必要がある。

3.7 広域 IP マルチキャストに適した経路制御アルゴリズムの提案

ここまで述べてきたように、一方向性のマルチキャスト型媒体を有効に利用するためには、既存の経路制御アルゴリズムに代わる新しいアルゴリズムが必要である。ここでは、一方向性のマルチキャスト型媒体の利用を前提とした、広域マルチキャスト機能のための経路制御アルゴリズムについて提案を行なう。

LAN を拡張した従来の経路制御アルゴリズムでは、各送信ホストを根とする動的なマルチキャスト tree を構築する方法が中心であった。しかし、衛星通信型の物理媒体を利用しようとすると、送信専用ホストへの経路を確保しなければならないので、例えば送信専用ホストへの伝搬遅延が大きい場合には効率が良くならない可能性もある。つまり、動的な経路全てに対して効率が良くなるような経路制御は非常に難しくなると考えられる。

したがって、衛星通信型の媒体を取り入れた経路制御アルゴリズムではある程度固定的な経路を提供することを考えていくことにする。

3.8　既存の経路制御アルゴリズムの組み合わせによる広域化

前述の DVMRP に関する問題点は、DVMRP に限った問題ではなく、一般的に 1 つの経路制御アルゴリズムのみでインターネット規模のマルチキャストを全て扱うことによる問題点としても捉えることができる。例えば、Link–State Routing を元にした経路制御アルゴリズムを広域に拡張することを考えた場合でも、構築されるマルチキャスト tree が大きくなるため、以下の問題点が生じる。

- ある送信ホストからあるグループへのマルチキャスト tree を、最初に計算する際の計算時間が増大し、最初のマルチキャストデータグラムの転送の遅延が増大する。
- マルチキャスト tree を計算するために交換される、リンクの状態やグループ所属状況等の情報に変更が生じた場合、全てのマルチキャストルータに対して報告を行なうため、ネットワークのバンド幅 を消費し、経路制御自体のオーバーヘッドが大きくなる。これは、グループ所属状況の変更が頻繁に起こる場合には重大な欠点となり得る。
- 広域になることによって、実際にマルチキャストデータグラムを送信するホストや、利用されるグループ数が増大するため、計算したマルチキャスト tree をキャッシュとして記憶するメモリ容量も大容量であることが求められる。キャッシュのための容量が小さいと、マルチキャスト tree を計算する回数が増え、データグラムの伝搬遅延が問題となる。

また、DVMRP や Link–State Routing を元にしたマルチキャストの経路制御アルゴリズムは、マルチアクセス型の LAN を拡張した LAN としては、十分に機能すると考えられる。したがって、広域ネットワークを分割し、各分割部分 (以後、この部分を地域と呼ぶことにする。) 内では、必要に応じて既存の経路制御アルゴリズムを適用することにする。そして、ここでは衛星通信型の物理媒体を、地域間を接続するために用いることを前提とする。

3.9　衛星通信型の媒体を前提としたアルゴリズムの段階

まず、衛星通信型の媒体はブロードキャスト型であることから、送信専用ホストと受信専用ホストの間でのマルチキャストデータグラムの扱いは以下の 2 つの方法が考えられる。

- 送信専用ホストは必ず受信専用ホストにマルチキャストデータグラムを転送する。
- 送信専用ホストは、どの受信専用ホストもマルチキャストデータグラムを必要としない (衛星を使う必要がないと判断する) 場合には転送を行なわない場合もある。

後者では、受信専用ホストが地域内のグループメンバーシップ情報を把握し、送信専用ホストにその情報を報告する必要がある。しかし、グループメンバーシップ情報の把握を、どのように行なうかにより、広域に適した経路制御アルゴリズム全体の効率に大きな影響を与えると考えられ、様々な方法を検討する価値がある。したがってここでは、グループメンバーシップ情報の方法に制限を与えないことにする。そして、衛星通信型の送信専用ホストから受信専用ホストまでのトラフィックは、地上網に影響を与えないことから、前者の方法で問題がないと判断する。

衛星通信型の送信専用ホストおよび受信専用ホストは、ともにマルチキャストデータグラムを受信して、受信専用ホストへデータグラムを転送するため、マルチキャストルータであるという前提をおく。

送信専用ホストが必ず受信専用ホストに転送する方法では、経路制御アルゴリズムを次の 3 つの段階に分けることができる。

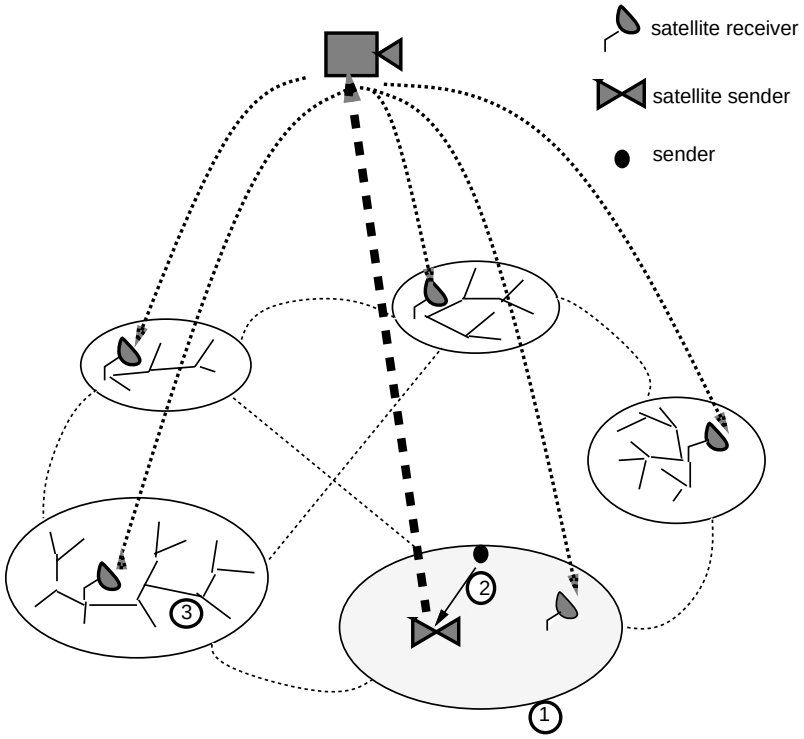

図 3.1: 衛星を前提とした広域マルチキャスト

- 地域の定義と衛星通信型の受信専用ホストの配置 (図 3.1 1 参照)
  地域毎に独立した経路制御アルゴリズムを適用できるように、地域の境界を認識する方法が必要となる。
  送信専用ホストの配置に関しては、受信専用ホスト数に比較して、送信専用ホスト

数は少ないという想定であるので、例えば、日本国内に数個程度の規模で配置されると仮定する。

- マルチキャストデータグラムの送信ホストから衛星通信型の送信専用ホストまでの経路制御アルゴリズム (図 3.1 2 参照)
- 衛星通信型の受信専用ホストから各地域内への経路制御アルゴリズム (図 3.1 3 参照)

ここでは、経路制御の各段階において様々な提案を行い、各段階のアルゴリズムの組み合わせによっていくつかの経路制御アルゴリズムを構成することにする。その後、各アルゴリズムの問題点等を指摘する。

3.10　衛星通信型の受信専用ホストの配置

経路制御の段階から見ると、衛星通信型の受信専用ホストの配置は、地域分けの段階に属していると考えられる。これは、地域の境界の基準をどのように定め、また、地域内で適用する経路制御アルゴリズムの個数に関する制限をどのようにするかを決定する段階である。地域の境界の基準により、地域の規模が決定されるので、この基準は地域内で適用する経路制御アルゴリズムの個数の決定にも影響を与える。

そこで、地域の境界の基準から考えることにする。

- IP ユニキャストアドレスのネットワーク部が共通であるホストの集合を地域とし、地域ごとに 1 つ受信専用ホストを配置する。
- 低速ネットワークにより接続されている部分を区切りとして、高速ネットワークで接続されている部分を地域とし、地域毎に 1 つ受信専用ホストを配置する。
  地理的に離れたホスト間を接続するための物理媒体は、低速であるというのが技術的に一般的なことである。ここでいう低速ネットワークとは、例えばイーサネット程度の転送速度と比較して低速で、地理的に離れたホスト間を接続するネットワークであり、例えば専用線等のことである。
- バックボーンに直接接続しているルータ毎に地域を分け、その地域内に 1 つ受信専用ホストを配置する。
  ここでいうバックボーンとは、例えば日本のインターネットの例である WIDE ネットワークにおいて根幹にあたる部分であり、特に地理的な地区間 (例えば東京– 大阪間等) を接続する部分である。バックボーンのルータは地区間の経路制御や各地区の管理を行なう存在である。この場合、受信専用ホストの配置については以下のように考えられる。
  - バックボーンに直接接続しているルータを受信専用ホストとする
  - バックボーンに直接接続しているルータ以外を受信専用ホストとする

バックボーンに直接接続しているルータが、ハードウェアその他の理由により衛星通信型の受信専用ホストとして稼働可能でない場合には、前者は不可能となる。また、逆に、前者を前提とすることにより、バックボーンに直接接続しているルータのハードウェア等の選択を制限するのは実用的とは考えられない。したがって、地域内の任意のルータが衛星通信型の受信専用ホストになり得ると想定し、考えを進める。

地域内で適用する経路制御アルゴリズムについては、

- 1 つの経路制御アルゴリズムに制限する場合
- 複数の経路制御アルゴリズムを利用する場合

が考えられる。

地域の規模が、1 つの経路制御アルゴリズムを効率良く適用することができない範囲となる場合は、地域内で複数の経路制御アルゴリズムを適用することが考えられる。その場合には、地域内でさらに経路制御適用範囲を認識することが必要であり、複雑な機構となることが予想される。ここでは、衛星通信型の媒体を中心として考えるために、地域内では 1 つの経路制御アルゴリズムを適用することを中心として考えていくことにする。この場合には、地域間の境界を認識することが必要であるが、その認識方法を先に考えることは、地域内で複数の経路制御アルゴリズムを適用する場合に、地域内の分割の境界を考える際の参考となるはずである。

3.10.1 地域の境界の認識方法

隣接する地域毎に異なる経路制御アルゴリズムを適用する場合には、プロトコルが異なるため、地域の境界のルータがどちらの地域に所属しているかを混同する可能性は小さい。

しかし、隣接する地域で、同一の経路制御アルゴリズムを適用する場合には、地域の境界のルータが自分の所属する地域からのみ、マルチキャストデータグラムを受信するような工夫が必要となる。隣接する地域毎に異なる経路制御アルゴリズムを選択するように制限を設けることは実用的とは言えない。よって、隣接する地域で、同一の経路制御アルゴリズムを適用する場合にも十分に対応できるようにしなければならない。

まず、地域の境界のルータが隣の地域の経路制御情報を受け取らないようにする方法が考えられる。

IP アドレスを基準に地域を分ける場合

地域の境界のマルチキャストルータというのは、この場合、IP アドレスのネットワーク部が異なるインターフェースを持つルータである。経路制御情報の交換の際は、地域を表現するネットワークアドレスと異なるインターフェースには経路制御情報を流さないようにすることで、地域の境界を決定することが可能である。

point–to–point 型のネットワークと、マルチアクセス型のネットワークに所属するルータは、point–to–point 型のネットワークの方が低速である場合が多いので、マルチアクセス型のネットワーク側の地域に所属する方が地域内のマルチキャストを有効に利用できるという意味で一般的であると考えられる。

マルチアクセス型のネットワークのみに所属しているルータは、所属する地域内のネットワーク数を考慮し、そのネットワークだけで孤立することがないようにした方が効率が良い。なぜならば、ネットワーク数が極端に少ない場合は、ブロードキャストとマルチキャストの有効性の差が小さくなるからである。

境界のマルチキャストルータが所属する地域は、地域ネットワークアドレスとして設定を行なうことによって識別を行なうことが可能である。

このアルゴリズムの問題点は次のとおりである。

> 例えば、クラス B とクラス C のネットワークが隣接する場合には、クラス C の地域とクラス B の地域の規模の差が大きくなる。クラス B で扱えるホスト数も多いので、クラス B を 1 地域と扱うのではなく、数地域に分ける方が実用的であるかも知れない。クラス B の場合、ホスト部からさらにネットワーク部として数ビットをとるサブネット技術が利用されているため、サブネット単位で地域を分ける方法も考えられる。しかし、ホスト部から何ビットをネットワーク部として扱うかについては、管理上の問題である。サブネットが細かい場合には、地域の規模が小さくなり、マルチキャストの効果も小さくなる。

低速ネットワークを基準に地域を分ける場合

低速ネットワークは、一般に point–to–point 型で 2 つのホスト間を接続する場合が多い。よって、地域の境界の識別は、point–to–point 型のインターフェースに対して経路制御情報を流さないようにすればよい。

低速ネットワークを基準とすると、地理的な基準で地域を分けていると捉えることもできる (図 3.2参照)。地理的な地域と IP アドレスの関係は規則的なパターンで捉えることはできないので、地域内か地域外かの識別には、地域内の IP ネットワークアドレスのリストを利用することが実用的な方法として挙げられる。

バックボーン上のルータを基準に地域を分ける場合

バックボーン上のルータは複数の組織を管理下におく状態となっているため、地域の規模は、IP アドレスや低速ネットワークを基準とする場合と比較して非常に大きくなる。すると、1 つの経路制御アルゴリズムの有効適用範囲を越える可能性が高い。したがって、地域内をさらに数地域に分けて管理を行なうことが一般的であると考えられる。分割の基準としては、前者の方法、つまり IP アドレスや低速ネットワークとする。

地域毎の規模の差が大きいと、地域内の分割数が異なってくる。ここでは、地域内に衛星通信型の受信専用ホストを 1 つに限定しており、地域内では、部分間を衛星通信型の媒体で直接接続することは考えていないため、地域内の部分間のマルチキャストは point–to-point 型を想定して行なわなければならない。また、地域毎の規模の差が大きいと、送信ホストと各グループメンバーホストの伝搬遅延時間の偏差が大きくなる。

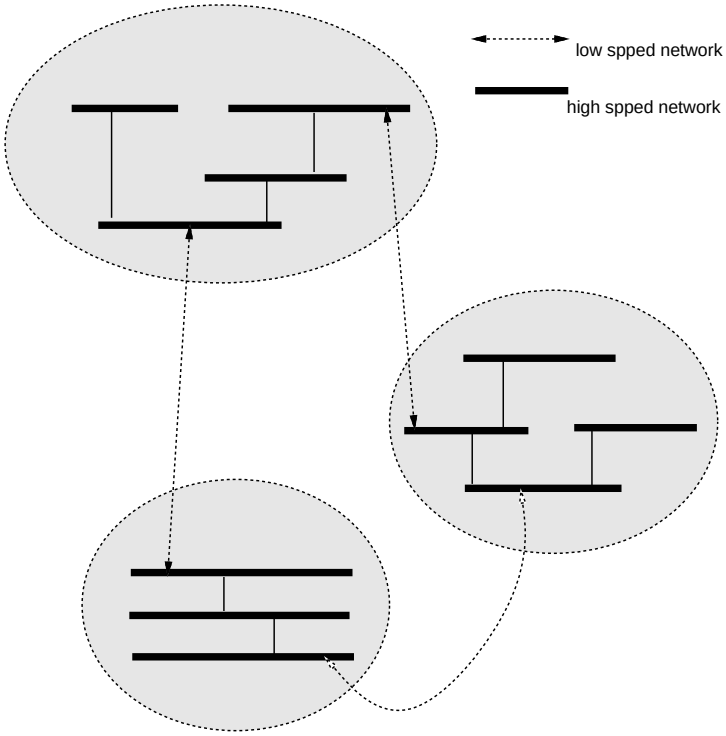

図 3.2: 低速ネットワークを基準とした地域分割

どのような基準で地域分けを行なう場合でも、地域の規模と 1 つの経路制御アルゴリズムの適用範囲との関係によっては、地域内で複数の経路制御アルゴリズムを扱うことが必要な場合がある。その時は、衛星通信型の受信専用ホストから地域内の経路制御方法についても、地域内の各分割部分への経路制御という要素が新たに必要となる。

地域の境界の決定に関しては、自動的に決定するのは困難であり、初期設定によって固定的に決定する方法が地域境界決定の柔軟性を与える意味で実用的な方法であるといえる。このような観点から、低速ネットワークを基準とする地域分けの方法は、IP アドレスを基準とする場合と比較して実用的であると考えられる。したがって、今後は低速ネットワークを基準とする地域分けを前提とし、さらに、1 地域内で経路制御アルゴリズムを 1 つ適用すると仮定して、考えを進めていくことにする。

3.11 送信ホストから衛星通信型の送信専用ホストまでの経路制御

この段階は、経路制御の面から見て、広域のマルチキャストを利用するかどうかの判断と、どのホストが判断を行なうか、そして、利用すると判断した場合にはどのような方法で衛星通信型の送信専用ホストまで到達させるかを決定する段階に属している。また、この段階で送信ホストと同一地域内のグループメンバーホストへの配送を行なうか否か

に関しても考える必要がある。この方法に依存して、「広域」マルチキャストの判断を行なうホストの決定や送信専用ホストまでの通信形態も変わってくると予想される。したがって、次のような順序で考えていくことにする。

- マルチキャストデータグラムの転送範囲の基準
- 送信ホストと同一地域内のグループメンバーホストへの配送
  - 「広域」マルチキャストの判定ホストの決定
  - 送信専用ホストまでの通信形態

3.11.1　マルチキャストデータグラムの転送範囲の基準

まず、送信ホストから衛星通信型の送信専用ホストまで経路制御を行なう必要があるのは、送信ホストが「広域」のマルチキャストデータグラムを転送したいときである。「地域内」のマルチキャストで十分な場合には、地域内の通常の経路制御に従うべきである。よって、広域のマルチキャストであるのか、地域内のマルチキャストであるのかを区別する必要が生じてくる。

広域のマルチキャストを利用するかどうかの判断は、IP マルチキャストデータグラムの転送範囲というマルチキャストの一般的な問題に関わる。したがって、マルチキャストデータグラムの転送範囲の基準を決める必要がある。これに関して、次の 2 つの方法が考えられる。

その 1　データグラムの生存時間を基準とする場合
その 2　IP マルチキャストアドレスを基準とする場合

その 1–データグラムの生存時間を基準とする場合

既存の経路制御アルゴリズムでは、マルチキャストデータグラムの転送範囲はデータグラムの生存時間を利用して指定される (2.3.1参照)。したがって、まず考えられる方法としては、生存時間の値を利用して、「広域」か「地域内」かを判断することが挙げられる。「広域」を表す生存時間を定義し、送信ホストがが生存時間を指定できるようにすれば、データグラムの生存時間の値によって、適用するアルゴリズムを変えることが可能であると考えられる。

例えば、DVMRP の実装 [?] ではデータグラムの生存時間の指定の参考値として 128 まで示されているので、それ以上の値で、かつ生存時間の最大値以下となる数を指定することが考えられる。

データグラムの生存時間は通常、中継ルータが処理する際に値を減らすことになっているため、「広域」であることを正確に判断するためには、送信ホストの指定した生存時間の値が大きく変化しないところ、つまり送信ホストにできるだけ近いところで、「広域」か「地域内」かを判断することが望ましい。

この方法の欠点は、隣接地域間のように、さらに細かい転送範囲の指定が難しいことである。というのは、各地域で規模が異なる場合、データグラムの生存時間の指定の仕方によっては、 到達させたい地域の全てにマルチキャストデータグラムが行き渡らない場合もあり得るからである。

その 2–IP マルチキャストアドレスを基準とする場合

「広域」であることを判断する方法として、IP マルチキャストアドレスを利用することも考えられる。ユニキャストアドレスについては 3 つのクラスに構造化され、アドレスからネットワークの規模を知ることが可能であるが、マルチキャストアドレスについては、アドレスの構造化については特に規定されていない。

そこで、IP マルチキャストアドレスに関して、マルチキャストデータグラムの転送範囲を判断できるような構造化を行なう方法が考えられる。IP マルチキャストアドレスは、32 ビットのうち上位 4 ビットまではクラスの識別のために予約されている。すると、IP マルチキャストアドレスの上位 4 ビットに関しては転送範囲を判別する材料とはならない。そこで、残りの 28 ビットを構造化することを考える。

ところで、IP マルチキャストアドレスは、転送されるべき各ローカルネットワークにおいて、ローカルネットワークのアドレスへ対応づけられる。ローカルネットワーク内で行なわれるマルチキャストは、「広域」でも「地域内」でも区別を行なう必要はないと考えられる。したがって、ローカルネットワークのアドレスに対応させる時には、IP マルチキャストアドレスの中で「広域」か「地域内」かを識別している部分を反映させる必要はない。同一のグループに関して、「広域」の場合と「地域内」の場合で、ローカルネットワークの異なるアドレスに対応づけられると、ローカルネットワークモジュールにおいて各ホストが把握しなければならないグループの数が大きくなってしまう。

例えば、イーサネットでは IP マルチキャストアドレスの下位 23 ビットのみをイーサネットマルチキャストへ変換する仕組みとなっている。すると、イーサネットのレベルでは IP マルチキャストアドレスの上位 5 ビットから 9 ビットまでは識別しないことになる。ということは、この部分のビットを利用して「広域」か「地域内」かの判断に利用しても、イーサネットのマルチキャストアドレスには反映されない。そこで、これらのビットを利用することが 1 つの方法として挙げられる。

もちろん、イーサネット以外の媒体で IP マルチキャストアドレスの全てをローカルネットワークのマルチキャストアドレスに内包する場合もあり得る。また、同一のグループについて、「広域」か「地域内」かによって複数のマルチキャストアドレスを利用することも考えられるため、利用できるグループの種類は構造化することによって少なくなる。

また、「広域」と「地域内」以外に「隣接地域」等の柔軟な転送範囲の制御を行なうことも可能となる。この点はデータグラムの生存時間を基準とする場合と比較して利点として考えることができる。

別の方法として、広域 IP マルチキャストのためのマルチキャストアドレスを予約する方法も考えられる。広域のマルチキャストの応用例が多くない点を考えると、現実的な方法としては有効であると考えられる。しかし、広域のマルチキャストの応用例が多くなってくると、広域 IP マルチキャストのためのアドレスの判断は、アドレスのビットの

パターンで把握できなくなる可能性がある。そうなると、広域IP マルチキャストのアドレスリストから比較しなければならなくなる。

3.11.2　送信ホストと同一地域内のグループメンバーホストへの配送について

広域マルチキャストの場合、衛星通信型の送信専用ホストへの経路制御を行なうが、その方法は次の 2 つに分類することができる。以下、各々の方法について、「広域」マルチキャストの判定ホストや、衛星通信型の送信専用ホストへの通信形態等について検討を行なう。

- 地域内グループメンバーホストへの配送はこの段階では行なわない。

  これはすなわち、送信ホストと同一地域に存在するグループメンバーホストにも、必ず衛星経由でマルチキャストデータグラムを転送することを意味する。

- 地域内グループメンバーホストへの配送を行なう。
  送信ホストから衛星通信型の送信専用ホストに到達するまでに、地域内にグループメンバーホストが存在するときには先にマルチキャストデータグラムを転送する。すなわち、送信ホストと同一地域内に存在するグループメンバーホストには、先にマルチキャストデータグラムを転送する。したがって、送信ホストと同一地域に存在するグループメンバーホストはサテライト経由でマルチキャストデータグラムを受信しないような工夫が必要となる。

3.11.3　送信ホストと同一地域内グループメンバーホストへの配送を行なわない場合

ここでは、送信ホストがどの地域に所属しているかによらず、まずマルチキャストデータグラムを、衛星通信型の送信専用ホストへ到達させる方法を考える。(図 3.3 参照)

地域内のグループメンバーホストはここでは関係がないので、送信ホストから衛星通信型の送信専用ホストに到達するまでに地域内にマルチキャストデータグラムが伝搬しないようにしたい。また、地域間でマルチキャストの経路制御情報は交換しないので、マルチキャストデータグラムの送信者と衛星通信型の送信専用ホストが異なる地域に存在している場合には、通信形態はユニキャストかブロードキャストのどちらかを利用することになる。この場合衛星通信型の送信専用ホストが一意に決まっているので、当然ユニキャストを利用することになる。

すると、「広域」マルチキャストの判断を行なった後、衛星通信型の送信専用ホストまでユニキャストを使って到達するのがよいことになる。これには、マルチキャストデータグラムを一時的にユニキャストデータグラムに変換するトンネリングの方法を利用す

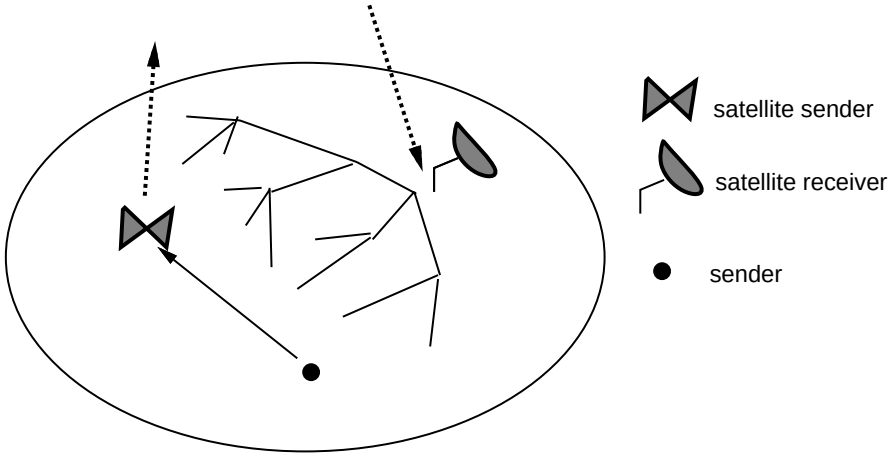

図 3.3: 送信ホストと同一地域内グループメンバーへ配送しない例

ることが考えられる。すなわち、「広域」マルチキャストの判断を行なったホストが、衛星通信型の送信専用ホストまでトンネリングを行なうということになる。(ここでいうトンネリングとは、DVMRP のトンネリング機能 (2.4.2参照) を利用することを考える。トンネリングで行なうデータグラムの IP オプションの追加は、IP モジュールで行なわれる。) したがって、送信ホストから衛星通信型の送信専用ホストまでマルチキャストの経路制御は不要である。

「広域」マルチキャストの判断を行なうホストとは、トンネリング技術から見ると、ローカルの終端となるホストとなる。したがって、これらのホストには、マルチキャストデータグラムをユニキャストデータグラムに変換することが必要である。ローカルの終端として次の 2 つが考えられる。

(1) マルチキャストデータグラムの送信ホストがなる場合

(2) その送信ホストに一番近いマルチキャストルータがなる場合

また、トンネリングのリモートの終端は衛星通信型の送信専用ホストであるので、そこではマルチキャストデータグラムに戻してからマルチキャストデータグラムを転送する必要がある。

(1) の場合

マルチキャストデータグラムの送信ホストは、2.3で述べたように、レベル 1 とレベル 2 の場合がある。レベル 1 のホストはクラス D に対する経路制御はほとんど行なわず、レベル 2 と比較して各プロトコル層の拡張点が少ない。トンネリング機能はマルチキャストデータグラムの送信のための経路制御アルゴリズムの一部と考えているので、レベル 1 のホストがトンネリング機能を持つと、プロトコル層の構造に混乱が生じることになる。

また、送信ホストにトンネリング機能を付加する場合には、クラス D アドレスを目的地アドレスとするデータグラムに対して、2.3.1で述べた IP モジュールに関する拡張点を

変更する必要が生じる。1 つの経路制御アルゴリズムの実現のために、IP モジュールの変更を行なうのは望ましいとは言えない。IP モジュールを変更しないでトンネリング機能を付加しようとすると、上位レイヤでクラス D アドレスの検出を行なわなければならず、各プロトコル層の役割に混乱をもたらす。

(2) の場合

この場合には、((1) の場合のような) 意味的な問題はないと考えられる。

前述のように、「広域」と「地域内」のマルチキャストを区別した上で、「広域」と判断された場合には、判断を行なったマルチキャストルータは、サテライト型の送信専用ホストまでトンネリングを行なう。衛星通信型の送信専用ホストの IP アドレスに関しては、あらかじめ設定しておく必要がある。これに関しては衛星通信型の送信専用ホストの情報を個々のマルチキャストルータが持っていると、情報に変更が起きた時の修正が容易ではないという問題が生じる。 工夫すべき点としては、衛星通信型の送信専用ホストに役割としての「名前」をつけ、既存の名前サーバを利用して参照することも考えられる。まとめると、次のようなアルゴリズムとなる。

受け取ったマルチキャストデータグラムが「広域」を指定しているか
「地域内」を指定しているかの判断を行なう。
IF「地域内」である
THEN　地域内の経路制御アルゴリズムに基づいて経路制御を行なう。
ELSE　トンネリングを指定し、トンネルのリモートの終端として
　　サテライト型の送信専用ホストを指定する。
　　(*IP*モジュールによって、ユニキャストデータグラムに変換する。)

問題点

地域内のグループメンバーホストに対してこの段階で配送を行なわない問題点としては、送信ホストから距離が短いグループメンバーホストが、衛星を経由することによって、距離が逆に長くなることである。また、ここで論じたアルゴリズムでは、全てのマルチキャストルータが衛星通信型の送信専用ホストに関する情報を保有しなければならない。しかし、経路制御アルゴリズムとしては、比較的単純であるといえる。

3.11.4　送信ホストと同一地域内のグループメンバーホストに先にデータグラムを配送する場合

この方法では、送信ホストの所属する地域内のマルチキャストの経路制御を最初に行なうことになるので、「広域」マルチキャストの判断を行なうホストの候補は、次のように考えることができる。(図 3.4 参照)

◁ 1 ▷ 送信ホストに一番近いマルチキャストルータが「広域」マルチキャストの判断を行なう。

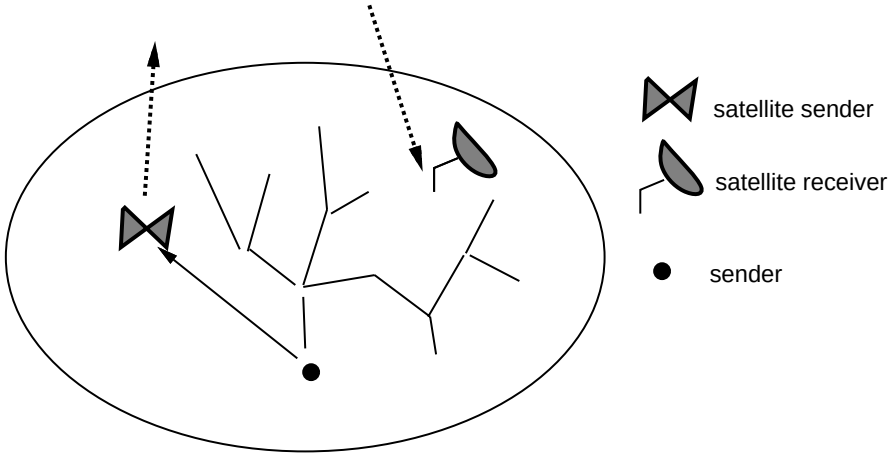

図 3.4: 送信ホストと同一地域内のグループメンバーホストに先にデータグラムを配送する例

◁ 2 ▷ 送信ホストと同一地域内の、特別なマルチキャストルータが「広域」マルチキャストの判断を行なう。

◁ 1 ▷ の場合

「広域」マルチキャストであると判断した場合には、3.11.3 で述べたように、送信ホストに近いルータが、衛星通信型の送信専用ホストに対してトンネリングを行なうという方法が挙げられる。

この場合、3.11.3で述べたトンネリング技術の他に、地域内のマルチキャストルータがその地域内の経路制御アルゴリズムに基づいてデータグラムを転送する方法となる。すると、送信ホストからマルチキャストデータグラムを最初に受け取るマルチキャストルータの動作は以下のようにまとめることができる。

IF 「広域」指定のマルチキャストデータグラムである
THEN 衛星通信型の送信専用ホストへのトンネリングを行なう
地域内の経路制御アルゴリズムに基づいてデータグラムの転送を行なう

◁ 1 ▷ では、全てのマルチキャストルータは衛星通信型の送信専用ホストの情報を知る必要がある。

◁ 2 ▷ の場合

この場合には、地域内のマルチキャストルータの中から、特別なマルチキャストルータを 1 つ選択し、このルータが「広域」マルチキャストの判断を行なって、衛星通信型の送信専用ホストまでトンネリングを行なうことになる。

送信ホストから、この特別なマルチキャストルータへは地域内のマルチキャストアルゴリズムに従って到達するようにする。衛星通信型の送信専用ホストは、その所属する地域内から送信されるマルチキャストデータグラムの全てを受信可能でなければならな

い。というのは、マルチキャストデータグラムが「広域」の場合には、地域内にそのグループメンバーホストが存在しなくても、衛星通信型の送信専用ホストまでは転送しなければならないからである。

例えば、2.1で述べたアルゴリズムにおいては、マルチキャストルータは全てのマルチキャストデータグラムを受け取ることを前提としているため、機能拡張等の必要は特にない。しかし、地域内の経路制御アルゴリズムとして DVMRP アルゴリズムを適用しているような場合には、衛星型の送信専用ホストのみが受け取るべきであるようなマルチキャスト データグラムであっても、全てのマルチキャストルータに転送されてしまう。これは、無駄なトラフィックを生じることとなるため、好ましくない。

また、地域内で Link–State Routing を元にしたアルゴリズムを適用している場合には、衛星通信型の送信専用ホストが全てのグループに所属することを Link–State の情報として、他のルータに知らせる必要がある。この時、各グループ毎に衛星通信型の送信専用ホストをグループメンバーとして報告するのは、scalability の点で問題がある。したがって、「全てのグループ」を意味する 1 つのマルチキャストアドレスを確保し、グループメンバーシップ情報の scalability を改善する必要がある。「全てのグループ」に関しては、[?] には“ ワイルドカードグループ ”として記述されている。

地域内にグループメンバーホストの存在しない状態で、衛星通信型の送信専用ホストへの転送をマルチキャストで行なうことは、効率の良いことではない。また、その際に発生する無駄なトラフィックや、グループメンバーシップ情報の報告にかかるコストは、地域内のマルチキャスト経路制御アルゴリズムに依存している。この要素が地域内の経路制御アルゴリズムの選択の際に制限を与えるのは、地域内で独立して経路制御アルゴリズムを選択するという前提上、望ましいことではない。

但し、この場合特別なマルチキャストルータだけが「広域」マルチキャストの判定を行なう機能を持ち、衛星通信型の送信専用ホストの IP アドレスを知っていればよいという特徴がある。つまり、他のマルチキャストルータは「広域」か「地域内」かを判断する機能を付加する必要がないのである。

さらに、送信ホストと衛星通信型の送信専用ホストが同一地域内に存在する場合には、「特別な」マルチキャストルータを衛星通信型の送信専用ホストと考えることも可能である。しかし、その地域内のマルチキャストデータグラムは全て衛星通信型の送信専用ホストに集中することになる。サテライト型の送信専用ホストへは地域外からの広域マルチキャストデータグラムも集中するため、送信専用ホストの負担が大きくなると考えられる。

衛星通信型の受信専用ホストによる地域内外の識別方法

ところで、送信ホストと同一地域内のグループメンバーホストに先にデータグラムを転送する方法では、前述の通りその地域に所属する衛星通信型の受信専用ホストから再度データグラムが転送されないようにする必要がある。

これには、マルチキャストデータグラムの送信ホストを指標として、受信専用ホストが地域内外のホストを識別することが必要となる。この指標による識別方法は地域の区

切り方に依存するため、地域の区切り方毎に、IP モジュールの経路制御アルゴリズムに対して、次のような付加機能を考えることができる。

- 地域の境界が IP アドレスを基準に決定されている時 (3.10.1参照)
  受信したマルチキャストデータグラムの送信ホストの IP アドレスと自分の IP アドレスを比較し、同一のネットワークアドレスである時はデータグラムを捨てる。(1地域内に複数の IP アドレスが存在する場合には、次の場合と同様に考えることができる。)

- 地域の境界が低速ネットワークを基準に決定されている時 (3.10.1参照)、または、バックボーン上のルータを基準に決定されている時 (3.10.1参照)
  受信したマルチキャストデータグラムの送信ホストの IP アドレスのネットワークアドレスと、自分と同一地域内に存在するネットワークアドレスのリストを比較して、同一のアドレスがあればデータグラムを捨てる。

ここまでで、送信ホストと同一地域内のグループメンバーホストに先にデータグラムを転送しない場合とする場合に分けて考察を行なった。トンネリングを部分的に使う方法では、多様なアルゴリズムが考えられるが、トンネリングのみを使う方法に比較して複雑なアルゴリズムとなっている。

それから、送信ホストから各グループメンバーホストへの到達時間の分散について比較してみると、送信ホストと同一地域内のグループメンバーホストに転送しない時は必ず衛星通信型の送信専用ホストを経由するため、送信ホストと同一地域内のグループメンバーホストに転送する時よりも到達時間の分散は小さくなると予想される。これは、アプリケーションレベルで送信ホストが各グループメンバーホストからの返事を待つことが必要な場合に、待ち時間の幅を見積もる問題として影響する点である。

広域のマルチキャストを利用するアプリケーション (送信ホスト) が、グループメンバーホストとデータの信頼性を保障する、つまり各データの到達を確認しながら連続してデータのやりとりを行なう場合には、1 回のデータ転送の際の各グループメンバーホストへの到達時間の分散 ($T_{member1}, T_{member2}, \cdots, T_{membern}$の分散) は小さい方が望ましいと考えられる。

3.12 衛星通信型の各受信専用ホストから地域内への経路制御

経路制御の段階から見ると、ここでは地域内のマルチキャストを利用するかどうかの判断と、利用する場合にはどのような方法で地域内のマルチキャスト経路制御へ渡すかを決定する段階である。この段階の経路制御は、衛星通信型の受信専用ホストが地域内のグループメンバーシップ情報をどの程度把握するかに依存して、効率が左右されるところである。

まず、衛星通信型の受信専用ホストが地域内のグループメンバーシップ情報を把握しないで、そのまま地域内の経路制御に委ねると、アルゴリズムによっては無駄なトラフィックを大量に生じる危険性が高い。無駄なトラフィックにより、地域内のネットワークの使用効率に影響が出るような方法は避けるべきである。そこで、地域内に関するグループメンバーシップ情報の把握を行なうことを前提とする。つまり、地域内のマルチキャストを利用するかどうかの判断機構が必要である。

そして次に、地域内の経路制御に委ねられた場合、地域外の送信ホストに対する経路制御情報を簡略化することも必要となってくる。

以下では、これらの 2 点について詳しい考察を行なうこととする。

3.12.1　衛星通信型の受信専用ホストのグループメンバーシップ情報の把握方法

2.1 で述べたアルゴリズムで、Distance–Vector routing を元にした経路制御アルゴリズムでは、マルチキャトルー タは直接接続しているローカルネットワークに関するグループメンバーシップ情報を把握するだけで、それをネットワーク間で交換する訳ではない。また、Link–State routing を元にした経路制御アルゴリズムでは、2.1.7 で述べたように、ローカルネットワークのグループメンバーシップ情報を他の全てのマルチキャストルータに報告する方法を適用している。このように、地域内で適用する経路制御アルゴリズムによって、衛星通信型の送信専用ホストが地域内のグループメンバーシップ情報を把握できる場合とできない場合がある。

地域内は複数のネットワークを接続して構成されている。そのため、衛星通信型の受信専用ホストが地域内のグループメンバーシップ情報を把握するためには、ローカルネットワークのグループメンバーシップ情報を把握しているマルチキャストルータから情報を収集する必要がある。つまり、ローカルグループメンバーシップ情報が 1 つの物理的なネットワークを越える必要がある。したがって、受信専用ホストと地域内のマルチキャストルータ間で、地域内の経路制御アルゴリズムによらない、独立したグループメンバーシップ情報の把握方法を確立しておく必要がある。但し、Link–State routing を元にした経路制御アルゴリズムのように、地域内の経路制御アルゴリズムだけで地域内のグループメンバーシップ情報を把握可能な場合には、改めてグループメンバーシップ情報の交換を行なう必要はないと考えられる。

地域内全体のグループメンバーシップ情報を地域内のマルチキャストルータから衛星通信型の受信専用ホストへ伝達する方法は、まず、定期的に行なうか、行なわないかに分類できる。以下、この分類にしたがって、幾つかのグループメンバーシップ情報の把握方法について提案を行なう。

ところで、地域内のグループメンバーシップ情報は、「広域」マルチキャストを利用するグループについてわかればよい。そこで、地域内グループメンバーシップ情報の量を軽減する方法について、 3.11.1 と関連させながら考える。

3.12.2 グループメンバーシップ情報の交換を定期的に行なう場合

定期的にグループメンバーシップ情報を把握しようとする場合には、グループメンバーシップ情報を把握する側が問い合わせをする場合と、グループメンバーシップ情報を提供する側が一定の時間間隔で報告を行なう場合とが考えられる。しかし、後者の場合も、結局はグループメンバーシップ情報を把握する側が一定の時間間隔で報告が行なわれるのを知っていなければならない。というのは、情報が得られない場合に、グループメンバーが存在しないから情報がないのか、ルータ、あるいはネットワークの異常等により、情報が到達しないのかを判断する必要があるからである。また、後者ではグループメンバーシップ情報を把握する側が、情報を要求しない場合でも報告が行なわれることになるので、無駄なトラフィックが定期的に流れる場合もあり得る。したがって、グループメンバーシップ情報を把握する側が定期的に問い合わせを行なうことを前提とする。

問い合わせを行なう場合は、

- 問い合わせを行なう対象
- 問い合わせに利用する通信形態

によって、幾つかの方法が考えられる。そして、問い合わせに対してどのような報告を行なうかについても、以下のように分類される。

- 報告を行なう対象
- 報告に利用する通信形態

まず、最も単純な方法から示す。

グループメンバーシップ情報交換方法その 1

衛星通信型の受信専用ホストが定期的に、地域内のすべてのマルチキャストルータに対してマルチキャストで問い合わせを行う。
問い合わせに対して、各ネットワークの代表マルチキャストルータ (例えば、ローカルグループメンバーシップの問い合わせを行なうルータ) はそれぞれのローカルグループメンバーシップ情報をユニキャストで衛星通信型の受信専用ホストへ報告する。

問い合わせのコストは O(地域内のネットワーク数)、報告のコストは
$O(\sum^{\text{地域内のネットワーク}}$(各ネットワークから受信専用ホストのホップ数 $*$ 各ネットワークのグループ所属数)) となる。

この方法では、各ローカルグループメンバーシップ情報を加工せずに、サテライト型の受信専用ホストへ直接報告が行なわれるため、報告のコストが大きいと考えられる。そこで、報告のコストを改善する方法としては、以下の方法が考えられる。

グループメンバーシップ情報交換方法その 2

衛星通信型の受信専用ホストが定期的に、地域内の全てのマルチキャストルータに対してマルチキャストで問い合わせを行う。

問い合わせに対して、各ネットワークの代表マルチキャストルータは、ローカルグループメンバーシップ情報として把握しているグループアドレスで、マルチキャストを用いて報告する。したがって、衛星通信型の受信専用ホストは全てのグループに属している必要がある。

受信専用ホストは地域内の全てのグループに所属しなければならないが、受信専用ホストは地域内の全てのマルチキャストを受信することになるので、これは地域内グループメンバーシップ情報の交換の際だけに適用するような工夫が必要となる。

この方法は、ローカルグループメンバーシップ情報の把握方法を元にしている。マルチキャストで報告を行うので、グループメンバーのすべてが報告を行う必要はなく、各グループにつき 1 つのマルチキャストルータが報告を行うだけですむ。報告のコストは
O(地域内の所属グループ数) となる。

グループメンバーシップ情報交換方法その 3

衛星通信型の受信専用ホストが定期的に、地域内のすべてのマルチキャストルータに対してマルチキャストで問い合わせを行う。

問い合わせに対して、問い合わせのデータグラムが通過した経路 (受信専用ホストを根とするマルチキャスト tree) の末端部の (つまり、データグラムを他のネットワークに転送する必要のないネットワーク) マルチキャストルータから、その経路を遡る方向でグループメンバーシップ情報の報告を行う。このとき、マルチキャストルータは受け取ったグループメンバーシップ情報の中に重複するグループが存在するときは、そのグループに関する情報はつけ加えないようにする。これによって衛星通信型の受信専用ホストが受け取るグループメンバーシップ情報の量を最小限におさえる。

複数のインターフェースを持つマルチキャストルータが複数のマルチキャストルータからグループメンバーシップ情報を受け取るとき、それらを 1 つのグループメンバーシップ情報としてまとめるには、同じ問い合わせに対する報告であることを識別する必要があるので、グループメンバーシップ情報の報告に識別番号を入れる必要がある。また、あるインターフェースから来るべきグループメンバーシップ情報が得られない場合に、どの程度待つかを決定する必要もある。

この方法は、2.1.1で述べたアルゴリズムのうち、RPM における Non–Membership report の方法を参考にしている。地域内で Distance–Vector routing を元にしたアルゴリズムを採用している場合には、マルチキャスト tree の末端部の認識の機能が含まれているので問題はない。但し、Distance–Vector routing を元にしたアルゴリズムや Link–State routing を元にしたアルゴリズム以外のアルゴリズムの場合では、

マルチキャスト tree の末端部の認識について新たに定義することになる可能性がある。

報告のコストは O(受信専用ホストを根とする $single-spanningtree$ のホップ数) となる。受信専用ホストが受け取るグループメンバーシップ情報の量は [その 2] の場合と等しい。

また、[その 1] の方法で受信専用ホストによる問い合わせのコストを軽減するための改良点として以下の方法が考えられる。

グループメンバーシップ情報交換方法その 4

衛星通信型の受信専用ホストにグループメンバーシップ情報を提供するマルチキャストルータを制限する。衛星通信型の受信専用ホストはこれらのマルチキャストルータにだけ定期的に問い合わせを行う。

そして、それらのマルチキャストルータは地域内の各部分のグループメンバーシップ情報を把握し、その代表として衛星通信型の受信専用ホストにグループメンバーシップ情報を報告する。

この方法では、[その 1] と比較して、受信ホストによる問い合わせのコストの軽減だけでなく、受信専用ホストが受け取るグループメンバーシップ情報量も軽減する。しかし、[その 2] [その 3] と比較すると、報告のコストは大きくなる。

問い合わせは、それを行うべきマルチキャストルータの数が数個である場合には、地域内の経路制御によっては、ユニキャストで行う方が関係のないマルチキャストルータへのトラフィックが減少し、有効となる場合もあり得る。

地域内の部分的なグループメンバーシップ情報の把握の方法としては、[その 1][その 2][その 3] が考えられる。この時、マルチキャストを利用する場合には、地域内の部分にだけマルチキャストが到達するような工夫が必要である。地域内のマルチキャストの制限を行なう場合には、3.11.1 で述べたような、マルチキャストデータグラムの転送範囲の問題を、「広域」か「地域内」かだけでなく、「地域内のある部分」かという点にまで拡張しなければならなくなる。また、この方法では地域の規模が大きい場合には有効となる可能性があるが、3.10.1 で述べた地域分けの方法にも関わってくる問題である。

地域内の経路制御アルゴリズムについては、1 つの経路制御アルゴリズムを仮定しているため、衛星通信型の受信専用ホストと地域内のマルチキャストルータの間に仲介ルータを入れる必要は特にないと考えられる。というのは、1 つの経路制御アルゴリズムが十分有効に利用される範囲内では、[その 4] を採用することによるコストの軽減効果は、[その 4] を採用するためにかかるコストより大きいとはいえないからである。

定期的に交換する場合、衛星通信型の受信専用ホストによる問い合わせの頻度は、グループメンバーシップ情報交換自体のオーバーヘッドに影響を与えるが、ローカルグループメンバーシップ情報交換の時間間隔よりも大きくて十分である。なぜなら、ローカル

グループメンバーシップ情報が更新されないうちにネットワーク間のグループメンバーシップ情報を更新しても正しい情報を得ることにならないからである。

しかし、定期的な場合でもグループメンバーシップに変化が起きた時にはできるだけ早く更新する方が良い。

3.12.3　グループメンバーシップ情報の交換を定期的に行なわない場合

次に、グループメンバーシップ情報の交換を定期的に行なわない場合には、Link–State のマルチキャスト経路制御アルゴリズムを元にした方法が考えられる。

グループメンバーシップ情報交換方法その 5

地域内の各ネットワークの代表ルータが、ローカルグループメンバーシップ情報に変化が起こった場合はいつでも、受信専用ホストへユニキャストを用いて報告を行なう。

この場合には、3.12.2で述べたのと同様に、受信専用ホストがグループメンバーシップ情報を必要としていない時にも行なわれる可能性があるので、そのような場合には無駄なトラフィックが生じる可能性がある。また、受信専用ホストの起動時には、受信専用ホストから問い合わせを行なってグループメンバーシップ情報を把握する必要がある。

[その 1] から [その 5] の方法では、「広域」マルチキャストを利用する可能性のある全てのグループに関する情報を把握することになる。そこで、実際に利用している「広域」マルチキャストについてのみ、グループメンバーシップ情報を把握する方法として、[その 6] を挙げることができる。

グループメンバーシップ情報交換方法その 6

衛星通信型の受信専用ホストは、広域マルチキャストのデータグラムを受け取ったときに、そのデータグラムの目的グループに対するグループメンバーシップ情報のみを、そのグループに対するマルチキャストを利用して問い合わせる (マルチキャストポーリングと呼ぶことにする。)

この問い合わせに対して、そのグループの所属状況を把握しているマルチキャストルータが (問い合わせを受け取り次第) マルチキャストで応答する。したがって問い合わせのコストは O(そのグループの所属状況を把握しているマルチキャストルータ数) となる。

一度行った問い合わせに対しては、キャッシュ等の一時的な記憶領域に保存し、問い合わせの回数を減らすことが考えられる。

キャッシュの情報が正しい情報を保存するために、

1. キャッシュの情報を時間で管理する
2. 地域内でそのような問い合わせを受けたことのあるグループから離脱し

た場合には、衛星通信型の受信専用ホストへ知らせるようにする
3. キャッシュ内の情報に関しては、衛星通信型の受信専用ホストが定期的にポーリングを行う

のような方法が必要である。この中で、キャッシュの情報を最も正しく保存できると考えられる方法は 3 の方法である。

[その 6] の問題点としては、あるグループに対する最初のデータグラムについてマルチキャストポーリングをおこなうため、伝搬遅延が増大することが挙げられる。また、問い合わせに対する報告が来ない場合、グループメンバーがいないと判断する最適時間の決定が困難であるといえる。しかし、現状として広域ネットワークが低速である点を考慮すると、広域ネットワークを利用する際には小伝搬遅延を必要条件としては要求していないと考えられる。もし、広域マルチキャストを利用するアプリケーションが、n 回のデータ転送で、あるグループメンバーホストへの到達時間の分散 ($T_{member1}(1), T_{member1}(2), \cdots, T_{member1}(n)$ の分散) を厳密に要求していないならば、この問題点は深刻な問題とはならないはずである。それでも、地域内のマルチキャストや、ユニキャストの配送には影響を与えないようにする工夫が必要である。

また、「広域」マルチキャストを利用するグループが増えてくるとキャッシュのために必要なメモリ量が増大する。キャッシュのためのメモリ量を越える程グループ数が増えた場合には、、キャッシュ内のグループリストから最近使われていないグループの情報を消す (LRU(Least Recently Used) 方式) ことにより、対処する必要がある。

利点としては、問い合わせ、報告ともにコストが小さいことが挙げられる。また、実際に広域マルチキャストとして利用される場合にのみグループメンバーシップ情報を把握するため、受信専用ホストが保存するグループメンバーシップ情報の量は「広域」マルチキャストの実際の利用数に対応する。「広域」マルチキャストの利用が少ない場合には、他の方法と比較してグループメンバーシップ情報の量が少なくなる。

ここまでで、地域内のグループメンバーシップ情報の把握に関しては、6 つの方法について述べた。このうち、地域内の経路制御アルゴリズムとの独立性や、グループメンバーシップ情報の情報量の観点から、[その 2][その 6] の方法を地域内グループメンバーシップ情報の交換方法として利用することにする。

[その 2] は個々のグループメンバーについての到達時間の分散が小さいことを要求するような場合、[その 6] はそのような分散を特に考慮していないような場合に適した方法といえる。

3.12.4 地域内グループメンバーシップ情報の制限

受信専用ホストが把握すべきグループメンバーシップ情報は、「広域」マルチキャストを利用するグループについてだけで十分である。マルチキャストデータグラムの転送範囲指定が「地域内」か「広域」かを判別する方法に関しては 3.11.1で述べたが、この判別

方法を利用して、受信専用ホストが把握すべきグループメンバーシップ情報の情報量を軽減することが可能な場合もある。

例えば、IP アドレスを基準として転送範囲を判別する場合には、広域用のマルチキャストアドレスに関してだけ、グループメンバーシップ情報の報告を行なうようにすることができる。

前述の、地域内グループメンバーシップ情報の把握方法 [その 1] から [その 5] については、グループメンバーシップ情報の報告のコストに直接影響が出てくる。

実際には、この点を考慮し、マルチキャストデータグラムの転送範囲の基準を IP アドレスと決定できるほど、単純ではない。しかし、IP アドレスを基準とした場合に、このような利点が付加的に生じる点については、全体的なアルゴリズムを考える上で加味すべき点である。

3.12.5　地域内の経路制御に関する付加機能

地域外の送信ホストからデータグラムが届いた場合、地域内で構築されるマルチキャスト tree は受信専用ホストが根となると考えると、経路情報数を減少させることができる。これによって、地域外から来たマルチキャストデータグラムは、受信専用ホストを根とする single–spanning tree に従って経路制御されることになる。この single–spanning tree を「デフォルトマルチキャスト tree」と呼ぶことにする。したがって、地域内のマルチキャスト経路制御テーブルには、デフォルトマルチキャスト tree に関するエントリを 1 つ加えれば良い。

逆にこのような工夫を行なわないと、地域外の経路制御情報も地域内の経路制御情報と同様の形式で含められるので、経路制御テーブルが大きくなるという問題点が生じる。よって、このような付加機能を地域内の経路制御に加えることが重要である。

3.13　広域マルチキャストのための経路制御アルゴリズム

ここでは、各段階に関する考察のまとめを行ない、広域マルチキャストのための経路制御アルゴリズムを幾つか提案する。その際、広域マルチキャストのための総合的な視点から各段階のアルゴリズムを選択して全体のアルゴリズムを構成し、最終的にはインターネットの現状を考慮した場合に実用的と思われるアルゴリズムを 1 つ選択することにする。

3.13.1　各段階のまとめ

各段階のアルゴリズムに関して述べてきたが、各段階は必ずしも独立に考えられるわけではない。そこで、各段階の依存関係を明らかにするために、図 3.5にまとめる。依存関係は図中の点線で示している。ここでいう、依存関係とは、「A のアルゴリズムによって B のアルゴリズムが異なってくる」ような関係のことである。マルチキャストデータ

グラムの転送範囲の基準によって、地域内グループメンバーシップ情報のコストに影響を与えるという場合は、アルゴリズム自体の変化が起きないので依存関係としては示さないことにした。

段階 1

衛星通信型の受信専用ホストの配置域 (段階 1) に関しては、低速ネットワークを基準として地域を分ける方針とした。

段階 2

衛星通信型の送信専用ホストまでの経路制御の段階 (段階 2) に関しては、転送範囲の基準と広域マルチキャストの利用を判定するホストの組み合わせにより、6 通りが考えられる。このうち、転送範囲の基準を TTL とする場合には、広域マルチキャストの利用は送信ホストに最も近いマルチキャストルータであることが望ましいため 2 通りに絞られる。転送範囲の基準を IP アドレスとする場合には 3 通りとなり、合計 5 通りの方法を以下に挙げる。

《2.1》

転送範囲の基準	TTL
地域内へのグループメンバーへの配送	×
広域マルチキャストの判定ホスト	任意のマルチキャストルータ
送信専用ホストまでの通信形態	トンネリング

《2.2》

転送範囲の基準	TTL
地域内のグループメンバーへの配送	○
広域マルチキャストの判定ホスト	任意ののマルチキャストルータ
送信専用ホストまでの通信形態	トンネリング

《2.3》

転送範囲の基準	IP アドレス
地域内のグループメンバーへの配送	×
広域マルチキャストの判定ホスト	任意のマルチキャストルータ
送信専用ホストまでの通信形態	トンネリング

《2.4》

転送範囲の基準	IP アドレス
地域内のグループメンバーへの配送	○
広域マルチキャストの判定ホスト	任意のマルチキャストルータ
送信専用ホストまでの通信形態	トンネリング

《2.5》

転送範囲の基準	IP アドレス
地域内のグループメンバーへの配送	○
広域マルチキャストの判定ホスト	特別なマルチキャストルータ
送信専用ホストまでの通信形態	トンネリング

特別なマルチキャストルータは全てのグループに所属する必要がある。

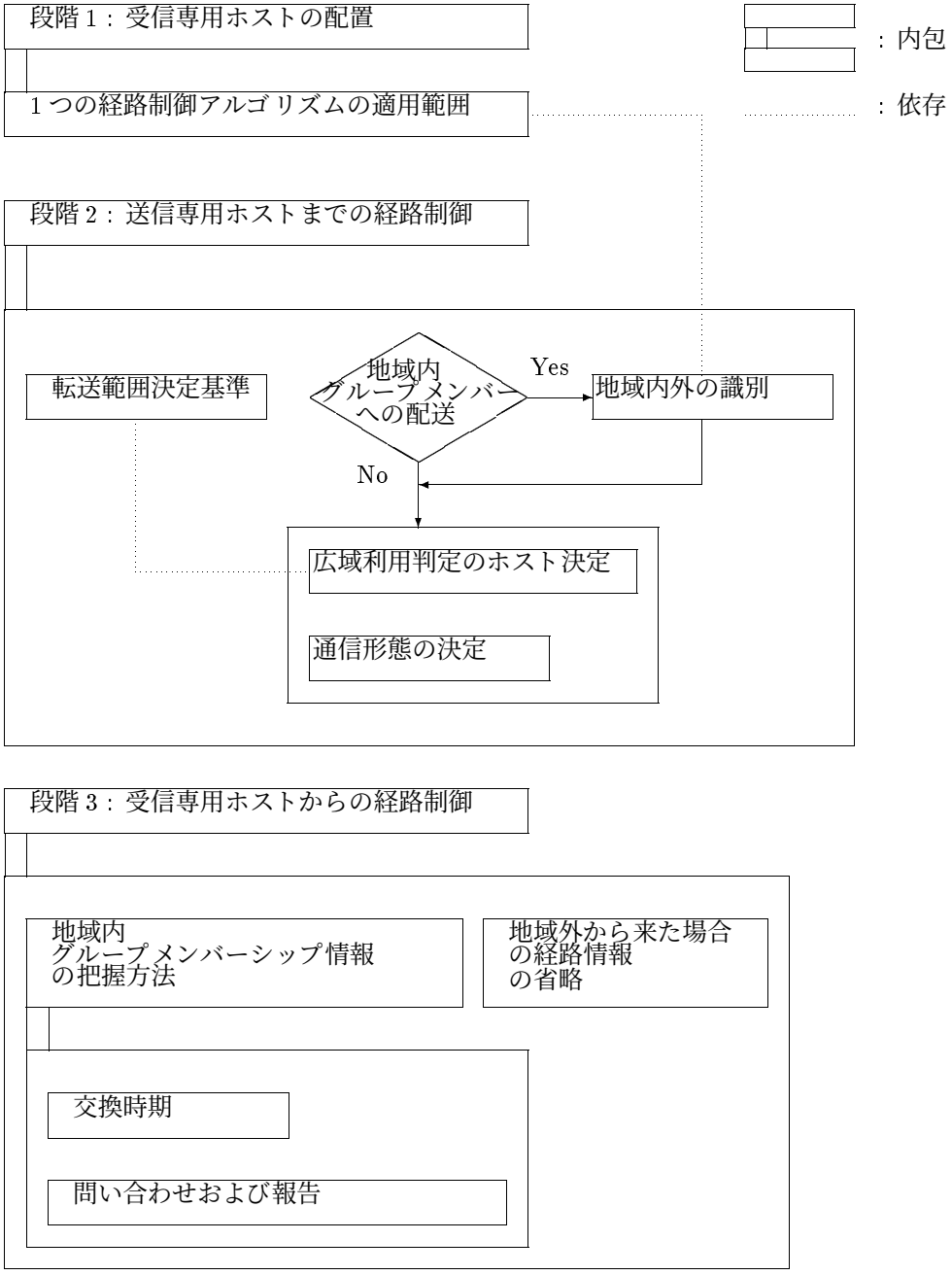

図 3.5: 広域マルチキャスト経路制御の各段階に関するまとめ

段階 3

衛星通信型の受信専用ホストからの経路制御の段階 (段階 3) では、受信専用ホストが、地域内のグループメンバーシップ情報を把握することによって、転送するかしないかを決定する。

この段階は、各地域内に閉じた段階であるので、基本的にはこの段階で必要とされる要素は独立に決定することができる。但し、地域内グループメンバーシップ情報の把握に関しては、その通信コストがデータグラムの転送範囲の基準によって影響を受ける。また、マルチキャストデータグラムの小伝搬遅延を特に要求しない場合は、《3.2》の方法が通信コストの合計量から考えて有用と考えられる。

《3.1》

地域内グループメンバーシップ情報の交換	定期的
受信専用ホストによる問い合わせ	
対象	地域内の全てのマルチキャストルータ
通信形態	マルチキャスト
グループメンバーシップ情報の報告	
報告者	地域内の所属グループのメンバーの 1 つ
通信形態	マルチキャスト

受信専用ホストは全てのグループに所属する必要がある。

《3.2》

地域内グループメンバーシップ情報の交換	非定期的
受信専用ホストによる問い合わせ	
対象	受信した広域マルチキャストデータグラムの目的グループ
通信形態	マルチキャスト
グループメンバーシップ情報の報告	
報告者	そのグループの所属状況を把握している各マルチキャストルータ
通信形態	マルチキャスト

一度得たグループメンバーシップ情報はキャッシュに保存し、キャッシュに保存されたグループに対しては定期的に問い合わせを行なう。

どちらの方法でも、地域外から来たマルチキャストデータグラムに対しては、受信専用ホストを根とする single–spanning tree に沿って地域内の経路制御を行なう。

3.13.2　広域マルチキャストの実現のための総合的条件を考慮したアルゴリズムの提案

広域マルチキャストのためには、以下の条件を満たすことが必要であることを 3.2 で述べた。

- 経路情報の scalability の改善
- 低速ネットワークへの考慮
- 中継ルータに関する工夫
- グループメンバーシップ情報の有効利用

このために衛星通信型の媒体を利用することを前提として、各段階においてのアルゴリズムについて考えてきた。ここで、広域マルチキャストの実現のために必要とされる総合的な条件について考えてみる。

- 広域マルチキャストのための地域内の経路制御アルゴリズムに対する付加機能はどの程度の規模に渡って必要か
  - マルチキャストルータの全てに必要か
  - マルチキャストルータの一部に必要か

  もちろん、最低限の変更で済むことが望ましい。
- 広域マルチキャスト機能を付加する対象の負荷の増加によって、他の通信に影響を与えないか
  一部のマルチキャストルータに極端に負荷が増大することがないように、一部のマルチキャストルータに多くの付加機能を持たせないようにする。
- 広域マルチキャスト機能の付加の結果、ユニキャストやブロードキャストと比較して、マルチキャストの有効範囲を維持することが可能であるか
- 広域マルチキャスト機能を利用する上位レイヤ、最終的にはアプリケーションが、広域に分散するグループメンバーとのエンド–エンドの通信で要求する条件を満たすことができるか
  エンド–エンド間で要求する条件とは例えば次のようなことである。
  - データの信頼性の保障
  - グループメンバーへ近い順に配送すること
  - グループメンバーへの到達時間の分散が小さいこと (どのグループメンバーにも同程度の遅延で到達すること)

もちろん、これらの条件には、相反する条件もある。例えば、広域マルチキャストの利用条件としてグループメンバーへ近い順に配送することを満たすように考えると、段階2のアルゴリズムは複雑となり、広域マルチキャストの機能自体が増えることになる。すると、広域マルチキャスト機能を付加する対象の負荷が増加する可能性がある。

広域マルチキャストの利用条件は、結局は、広域マルチキャストデータグラムの伝搬遅延に関わる問題である。そこで、付加機能と伝搬遅延に焦点を当て、広域マルチキャストのためのアルゴリズムを提案することにする。

広域マルチキャストアルゴリズム 1 付加機能が少なく、グループメンバーへの伝搬遅延に関しても特に考慮しないアルゴリズム
伝搬遅延を考慮しないことから、{《2.1》または《2.3》}+{《3.1》または《3.2》}が考えられるが、《2.3》ではマルチキャストアドレスの構造化を行なわなければならないので、《2.1》よりも付加機能が増える。また、伝搬遅延を特に考慮しないので、地域内グループメンバーシップ情報のコストが小さい方を重視すると、《2.1》+《3.2》となる。

広域マルチキャストアルゴリズム 2 付加機能を与える対象数が少なく、各グループメンバーへの伝搬遅延が最小限となることを重点とするアルゴリズム
広域マルチキャストの判定ホストが特別なマルチキャストルータだけで済むというアルゴリズムは、《2.5》である。伝搬遅延を重視するので、地域内グループメンバーシップ情報に関しては、《3.1》がよいと考えられる。よって、《2.5》+《3.1》となる。

広域マルチキャストアルゴリズム 3 各グループメンバーへの伝搬遅延が最小限となることを重点とし、無駄なトラフィックを最小限にするアルゴリズム
段階2では、《2.2》、《2.4》、《2.5》が考えられるが、《2.5》では、特別なマルチキャストルータに必ずマルチキャストデータグラムが到達するので無駄なトラフィックが生じる。したがって、段階2では、《2.2》または《2.4》が考えられる。また段階3の、地域内グループメンバーシップ情報の通信コストが小さいのは、《3.2》である。但し、《2.4》の場合では、《3.1》でもグループメンバーシップ情報の制限を行なうことが可能であるため、《3.2》との通信コストの差は小さくなる。《3.2》では、最初のマルチキャストデータグラムの伝搬遅延は大きくなるという点を考慮すると、《2.4》+《3.1》が得られる。

3つのアルゴリズムは、重視する条件が異なるので、一般的な比較を行なうことは困難である。しかし、現状のインターネット、ここでは特にTCP/IPにおいて、広域のマルチキャスト機能に関する実現がまだ行なわれていないことを考慮すると、まず広域マルチキャストを実現することが重要である。伝搬遅延を調整するアルゴリズムの是非については実際に広域マルチキャストを利用しながら検討していく必要がある。

したがって、伝搬遅延については特に考慮しないこととし、地域内の経路制御アルゴリズムに最低限の付加機能を加えるだけで実現されるような広域マルチキャストアルゴリズム 1

の方法をここでは最終的なアルゴリズムとして選択することにする。すなわち次のアルゴリズムである。

段階 1

低速ネットワークを基準として地域分けを行ない、各地域に受信専用ホストを配置する。衛星通信型の送信専用ホストは、数地域に 1 つ配置する。

段階 2

マルチキャストデータグラムの転送範囲の基準	TTL
地域内のグループメンバーへの配送	×
広域マルチキャストの判定ホスト	送信ホストに最も近いマルチキャストルータ
送信専用ホストまでの通信形態	トンネリング

段階 3

受信専用ホストが、地域内のグループメンバーシップ情報を把握することによって、転送するかしないかを決定する。

地域内グループメンバーシップ情報の交換	非定期的
受信専用ホストによる問い合わせ	
対象	受信した広域マルチキャストデータグラムの目的グループ
通信形態	マルチキャスト
グループメンバーシップ情報の報告	
報告者	そのグループの所属状況を把握している各マルチキャストルータ
通信形態	マルチキャスト

一度得たグループメンバーシップ情報はキャッシュに保存し、キャッシュに保存されたグループに対しては定期的に問い合わせを行なう。

地域外から来たマルチキャストデータグラムに対しては、受信専用ホストを根とする single–spanning tree に沿って地域内の経路制御を行なう。

次では、このアルゴリズムの特に 段階 2 の実装について述べることにする。

3.14　実装環境

ここでは、前章で述べた広域マルチキャストの経路制御アルゴリズムのうち、段階 2 の「送信ホストから衛星通信型の送信専用ホストまでの経路制御」の部分の実装について、実装のための環境、および具体的な設計・実装方法等を述べることにする。

3.14.1 ハードウェア

実装のために利用するハードウェアとして、オムロン株式会社のワークステーション LUNA–II 5 台を利用した。オペレーティングシステムは Mach 2.5 で、UNIX 4.3BSD–tahoe に準拠している。ネットワークプロトコルは TCP/IP を採用している。

5 台の接続は図 3.6のように、4 つのセグメントからなるようにイーサネットを用いて接続した。図中 * 印のマシンはマルチキャストルータであり、hitomaro というマルチキャストルータを衛星通信型送信専用ホストと想定し、buson というマルチキャストルータが、衛星通信型の送信専用ホスト (hitomaro) にマルチキャストデータグラムを配送する部分 (段階 2) を実装するための環境とした。

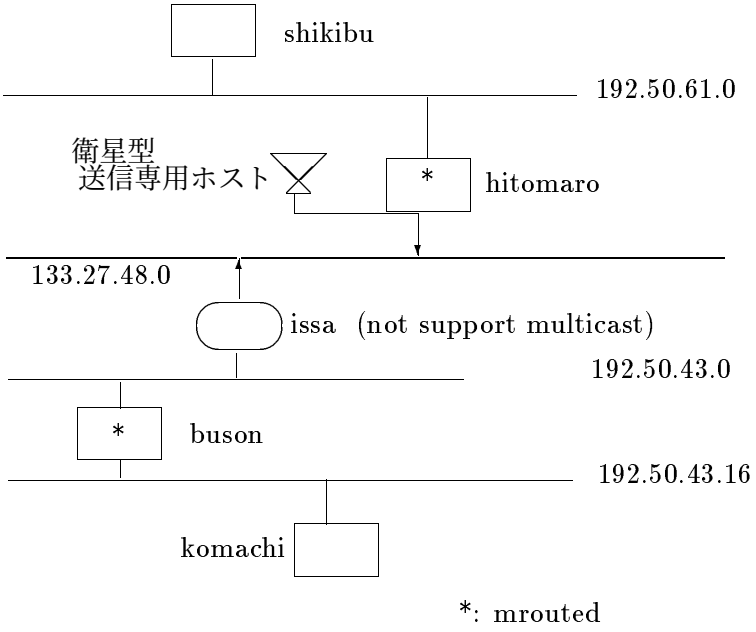

図 3.6: ハードウェアの接続構成

3.14.2 ソフトウェア

広域のマルチキャスト経路制御アルゴリズムは、既存の経路制御アルゴリズムを組み合わせて利用するアルゴリズムであるので、実装のためには既存の経路制御のソフトウェアが必要となる。そこで、前述のハードウェアに対して、既存のマルチキャスト経路制御アルゴリズム DVMRP の実装版である `mrouted` というソフトウェアを利用することにした。このソフトウェアは、トランスポートプロトコルの 1 つである VMTP[?, ?, ?] の 4.3 BSD 対応の実装パッケージの一部として含まれており、Stanford 大学の Steve Deering 氏によって開発された [?, ?]。特に `mrouted` に関する部分は、MULTICAST 1.2 Release

と呼ばれている。

mrouted をインストールするためには、2.3で述べた TCP/IP に対する拡張を必要とするため、カーネル部分に対する変更部分もパッケージの中に含まれている。これは、2.3で述べたレベル 2 のホストのための変更である。このうち、ローカルネットワークサービスインターフェース (特にイーサネットのインターフェース) に対するマルチキャスト機能の拡張に関しては、 LUNA–II で扱うイーサネットインターフェースには対応していなかったため、 LUNA–II のイーサネットインターフェースに対応するように別に作られたものを利用した。

3.14.3　カーネル変更部分について

パッケージに含まれる具体的なカーネルの変更部分のうち、新しく加えられたコードは以下の通りである。

- IGMP(ローカルグループメンバーシップ情報の把握) モジュール
  `netinet/{igmp.c,igmp.h,igmp_var.h}`

- DVMRP のカーネル部分
  `netinet/{ip_mroute.c,ip_mroute.h}`

これらのコードが加えられることにより、以下の既存のコードについても変更がなされている。

```
netinet/if_ether.{c,h}      net/if.{c,h}
netinet/in.{c,h}            net/if_loop.c
netinet/in_pcb.{c,h}        net/if_sl.c
netinet/in_proto.c          net/raw_cb.{c,h}
netinet/in_var.h
netient/ip_icmp.c           h/ioctl.h
netinet/ip_input.c          h/mbuf.h
netinet/ip_output.c
netinet/ip_var.h
netinet/raw_ip.c
netinet/udp_usrreq.c
```

また、LUNA–II のイーサネットインターフェースにマルチキャスト機能を拡張したコードは　`luna2if/if_am.c`　である。

LUNA–II の 4.3BSD–tahoe のネットワークソースコードを利用して、上記のコードに変更を加えたカーネルを (ルータ用と、ルータ機能を持たないホスト用に) 再コンパイルし作成した。 図 3.6中 * 印のホストにルータ用のカーネル、shikibu,komachi にはルータ機能を持たないカーネルをインストールした。issa にはマルチキャスト機能を付加しなかった。

マルチキャストデータグラムのカーネル内での経路制御方法について次にまとめる。

3.14.4 IP モジュールにおける通常のマルチキャストデータグラムの経路制御ルーチン

mrouted は、マルチキャストの経路制御情報を交換することによって、マルチキャストのための経路制御テーブルを (ユニキャストとは別に) 作成している。実際に、経路制御を行なうのは mrouted ではなく、カーネルの IP モジュールである。そこで、まず、カーネル内のマルチキャストデータグラムの経路制御方法について述べることにする。

IP モジュールでは、通常のマルチキャストデータグラムの経路制御は、以下に示す a)、b) それぞれの場合について、以下のように処理される。

a) ローカルネットワークモジュールからデータグラムを受け取った場合
つまり、ネットワークを通じて (他のホストから) データグラムを受け取ったと考えられる場合
ipintr() において
IF (*IP*オプション *LSRR*の処理)
　自分の持つアドレスがトンネリングのリモートの終端アドレスと一致する
THEN *IP*ヘッダの目的地ホストをグループアドレスに戻す。

IF 目的地アドレスがグループアドレス
THEN *ip_mforward*() を呼び出し転送に関する処理を行なう
　　自ホストも受け取るかどうか決定する
　　上のレイヤの *input* ルーチンへ渡す

b) 上位レイヤモジュールからデータグラムを受け取った場合
ip_output() において
ユニキャストのルーティングテーブルからマルチキャストデータグラムを送信
する際のデフォルトのインターフェースの情報をとってくる
(3.14.5 で述べた送信インターフェースの設定)
IF　目的地アドレスがグループアドレス
THEN
　　IF　そのマルチキャストデータグラムの送信インターフェース、
　　TTL 等が別に指定されている
　　THEN　デフォルトの設定から指定された値へ変更する
　　IF　自分がデータグラムの目的とするグループに所属している、
　　かつ
　　自分もデータグラムを受け取る設定となっている
　　THEN　自分にもマルチキャストデータグラムを転送する
　　ELSE
　　IF　自分がマルチキャストルータである
　　かつ
　　転送フラグが立っていない (この関数が呼び出されるのが 1 回目
　　である)
　　THEN *ip_mforward*() へ
　　指定されたインターフェースへの *output* ルーチンへ渡す。

a) および b) で呼び出される *ip_mforward()* はマルチキャストルータが、マルチキャス

トのための経路制御を行なうルーチンである。マルチキャストルータでないホストは、*ip_mforward()* では何もしない。

マルチキャストルータの場合
ip_mforward() において
IF　マルチキャストデータグラムがトンネリングされている
THEN　*IP*ヘッダから、トンネリングのために使った *IP* オプション部分
　　を削る (1**)

経路制御テーブルから送信ホストに関するエントリを探す (*mrtfind*())(2**)
FOR　全てのインターフェースについて
　IF　(*children* ビットが立っている)
　　または
　　(*leaf* ビットが立っていて、グループメンバーが存在する)
　THEN
　　IF　インターフェースにトンネリングフラグが立っている
　　THEN　*tunnel_send*() へ
　　ELSE　*phyint_send*() へ

3.14.5　mrouted の起動および関連ソフトウェアについて

ここでは、 mrouted を起動する際に必要な設定等について述べる。

まず、マルチキャストルータでないが、レベル 2 のマルチキャスト機能を付加したホストは、マルチキャストルータによるローカルグループメンバーシップ情報交換に参加するために、224.0.0.1 のグループに自動的に所属することになる。

そして、ローカルグループメンバーシップ情報の問い合わせをマルチキャストルータから受け取るためには、マルチキャスト機能を持つインターフェースに対してマルチキャストパケットを受け取るフラグ (PROMISC, ALLMULTI 等) をあらかじめ設定しておく必要がある。また、この問い合わせに対して答えるには、マルチキャストデータグラムの送信のためのインターフェースの初期設定も必要となる。設定には以下のコマンドを利用する。

```
#/etc/route add group-address interface-address
例: #/etc/route add 224.0.0.0 192.50.43.18
```

今回利用したソフトウェアパッケージではレベル 2 用の機能付加が行なわれているが、例えばレベル 1 の (マルチキャストデータグラムの送信のみが可能である) ホストでは、

マルチキャストデータグラムの送信のためのインターフェースを設定するだけで十分となるはずである。

次に、mrouted の起動は、前述のように図3.6中 * 印のついたルータに対して行なった。

mrouted では、インターフェースについて、マルチキャスト機能を持つインターフェースを“ phyint ”、トンネル機能 (2.4.2参照) を持つインターフェースを“ tunnel ”として仮想的に扱っている。(その仮想的なインターフェースは vif と呼ばれる。)phyint はそのルータが持つマルチキャスト機能付のインターフェース数となる。tunnel は point–to–point 型のネットワークインターフェースにも適用できる。したがって、phyint、 tunnel に対応する実際のインターフェースは同一でもよい。また、同一のインターフェースから複数の tunnel を指定することも可能である。

そして、phyint に対する metric (そのインターフェースでかかるコスト) および threshold (そのインターフェースからマルチキャストデータグラムを転送する際のデータグラムの生存時間の閾値) に関する設定や 使わないようにする設定を設定ファイルによって行なうことが可能である。そして、tunnel のローカルおよびリモートの終端に関する設定についても同様である。

mrouted でのトンネリング () を利用する場合は、ローカルの終端およびリモートの終端となる各々のホストにおいて、互いに明示的に設定ファイルに書いてく必要がある。設定ファイルの形式は以下のようになっている。

```
書式：phyint [disable] [metric <m>] [threshold <t>]
書式：tunnel <local-addr> <remote-addr> [metric <m>] [threshold <t>]
例　：tunnel 36.8.0.77    36.2.0.8     metric 3
```

mrouted を起動すると、設定ファイルを元に vif を管理するための情報を格納し、さらに 各 vif の情報を元にして経路制御に関する情報の初期化を行なう。

MULTICAST 1.2 Release には、mrouted の他に表 3.1 に示される関連ソフトウェアが含まれており、全て動作を確認した。

3.14.6 DVMRP と mrouted の相違点

2.4 で述べた DVMRP の設計と、その実装版 mrouted には幾つかの相違点がある。以下では、その相違点について挙げる。

- DVMRP メッセージの形式を簡潔にしていること。
- 経路制御情報交換で必要な split horizon には 2 種類の方法があるが、ここでは無限大の距離を利用する“ split horizon with poisoned reverse ”を利用している。到達不可能な経路との区別を行なうため、別のフラグを用意している。
- 無限大を表す距離はコンパイル時に組み込まれる定数とし、経路間あるいはルータ間で異ならないようにしている。

表 3.1: MULTICAST 1.2 Release に含まれるソフトウェア

mtest	マルチキャスト機能の動作確認のためのツール。インターフェースのフラグの設定やグループへの所属・離脱の試行が可能。
netstat	DVMRP により得られる経路制御テーブルや各インターフェースのグループ所属状況を確認するためのツール
ping	マルチキャストを利用したアプリケーション例。マルチキャストアドレスを利用したグループメンバーへの応答確認が追加されている。
rwhod	マルチキャストを利用したアプリケーション例。既存の rwhod はブロードキャストを利用していたが、それをマルチキャスト対応に変更している。

- サブネット化されたネットワークを外部のルータから隠さないようにしている。このようにしないと、child link や leaf link の識別のアルゴリズムがうまく起動しない。
- mrouted を稼働させるネットワークのトポロジは、マルチキャスト機能を持つネットワークとトンネルのみからなることを前提としている。ブロードキャスト機能だけのような他のネットワークは、IP マルチキャストデータグラムを制御することが不可能なため、排除する。
- タイマーの設定値が DVMRP で規定された値とは変更している。
- タイマー処理の 1 つ、leaf timeouts に関しては、vif 毎ではなく経路毎に行なわれる。
- ある vif の隣接ルータから応答がない場合は、経路制御情報ではなく、隣接ルータへの検査メッセージが送られる。これは両方向で成立しない経路が確立されないようにするためである。(特にトンネリングの場合その可能性がある。)
- トンネリングされたデータグラムの送信ホストアドレス欄には、トンネルのローカルの終端アドレスではなく、本来の送信ホストアドレスがそのまま保持される。
- leaf link の認識のアルゴリズムの詳細に関しては、認識速度が早くなるように変更している。

3.15　段階 2 の設計および実装

段階 2 のアルゴリズムは以下のようであった。

マルチキャストデータグラムの転送範囲の基準	TTL
地域内のグループメンバーへの配送	×
広域マルチキャストの判定ホスト	任意のマルチキャストルータ
送信専用ホストまでの通信形態	トンネリング

全てのマルチキャストルータが送信専用ホストの IP アドレスを知っているようにする。

段階 2 の具体的な設計に当たり、現在のマルチキャスト経路制御への追加機能をトンネリングを中心として次のように分類して考えることにする。

- トンネルのローカルの終端側 (任意のマルチキャストルータ) の機能
  - TTL を基準とした広域マルチキャストの判定機能
  - トンネリングの設定
- トンネルのリモートの終端側 (衛星通信型の送信専用ホスト) の機能
  - トンネリングの設定
  - 通常のトンネリングとの判別機能
  - 衛星通信のインターフェースへの転送機能

ここでいうトンネリングとは、次のような性質を持っている。

3.15.1 衛星通信型の送信専用ホストへのトンネリングの特徴

トンネリングとは、2.4.2で述べたように、マルチキャストデータグラムをユニキャストデータグラムに変換し、マルチキャスト機能を有しないルータに混乱を与えずに転送を行なう技術である。

ここで要求するトンネリングとは、トンネリングによって通過するルータがマルチキャスト機能を有するかどうかは問題ではなく、「広域」マルチキャストである場合のみに利用する。したがって、広域マルチキャストのためのトンネリングの特徴は、以下のようにまとめられる。

(0) このトンネリングの設定は各ホストにつき 1 回のみである。

(1) マルチキャストルータから見て、トンネルのリモートの終端は送信専用 ホストに固定される。

(2) 送信専用ホストから見て、トンネルのリモートの終端は任意のマルチキャストルータの可能性がある。したがって、トンネリングされたデータグラムを受け取った時、どのルータから受け取った場合でも受け入れる必要がある。

(3) トンネリングの終端同士では経路制御情報の交換を行なわない。(経路制御情報の交換は地域内で閉じるようにするため)

(4) トンネルの終端となる各インターフェースは、広域マルチキャスト時以外は、地域内のマルチキャストのためのインターフェース (mrouted では、phyint, tunnel) として機能する必要がある。

上記の特徴に対し、 mrouted で提供されているトンネリングの機能には以下の相違点がある。

<0> トンネリングの設定の個数に特に制限はない。

<1> トンネリングを行なう際には、トンネルのローカルとリモートの終端となるホスト (インターフェース) 間で互いに固定する必要がある。(上記 (2) との相違点)

<2> トンネリングのリモートの終端では、トンネリングされたデータグラムのローカルの終端が、逆に自分から見てリモートの終端に当たっていない場合には、データグラムを転送しない。(上記 (2) との相違点)

<3> トンネリングの終端同士で経路制御情報の交換を行なう。(上記 (3) との相違点)

したがって、広域マルチキャストのために送信専用ホストまで行なうトンネリングは mrouted で提供されるトンネリング機能とは多くの相違点があることがわかった。

特に、新しいトンネリング機能では、経路制御情報をトンネル間で交換する必要がないため、トンネリング機能に関する情報を 新しい vif として phyint , tunnel と並列に扱うように定義すると、 mrouted (DVMRP) 自体の経路制御方法に変更を加えることになる。新しいトンネリング機能は、DVMRP 以外の経路制御アルゴリズムに対する適用をも想定しているため、 mrouted に対しても比較的簡単に移植可能でなければならない。よって、 mrouted で行なわれている通常のトンネリング機能とは別の形式で扱う必要がある。但し、マルチキャストデータグラムからユニキャストデータグラムへの変換方法自体に関しては、2.4.2 で述べたのと同様である。

以下では、このような相違点を踏まえた上で、広域マルチキャストで利用するトンネリング機能の設計および実装について述べることにする。

3.15.2　トンネルのローカルの終端側の機能

広域マルチキャストであることを示す TTL 値について

IP マルチキャストデータグラムの TTL の設定方法は、マルチキャストを利用するアプリケーションレベルで指定できるようになっている。具体的には、プロセス間通信で利用されるシステムコール socket に対するオプションとして TTL 値を指定するようになっている。TTL 値を特に指定しない場合は、1 とされる。TTL 値の指定可能範囲は、0 以上 255 となっている。(0 を指定した場合は、ローカルホスト内でループバックインターフェースを利用してマルチキャストが行なわれる。) 3.11.1 で述べたように、[?] ではマルチキャストデータグラムの転送範囲の参考値として 128 まで示されている。

広域マルチキャストが指定されていることを判断するのは、送信ホストに近いマルチキャストルータである。TTL は通常、ルータがデータグラムの処理を行なう度に減らすことになっている。送信ホストに近いマルチキャストルータに到達した時は、送信ホストが指定した TTL 値のままであるので、広域マルチキャストと判断される TTL 値は域である必要がなく、固定定数で判断することが可能である。

128 以内の値を広域マルチキャストを指定する値として決めると、既存の mrouted に対して混乱を招く可能性がある。したがって、ここでは、*200* を広域マルチキャストを指定する値とする。

```
#define WIDE_MULTI_TTL 200
```

広域マルチキャストの判定機能のカーネルへの追加

広域マルチキャストの判定を行なった結果、広域マルチキャストでなかった場合は、mrouted によって作成された経路制御テーブルを見てルーティングを行なうことになる。広域マルチキャストである場合には、静的なトンネリングを行なうので、マルチキャストの経路制御テーブルを参照する必要はない。従って、広域マルチキャストの判定を行なうのは、*ip_mforward()* ルーチンのマルチキャストの経路制御テーブルを参照する前、(つまり 3.14.4 の (1**) と (2**) の間) に行なうこととする。

衛星通信型の送信専用ホストの IP アドレスに関する初期設定

全てのマルチキャストルータは、広域マルチキャストのトンネリングのリモートの終端として、送信専用ホストを知っていなければならないが、このアドレスは送信専用ホスト数が増加すると変化する可能性があるため、 mrouted の設定時にユーザが行なえる方法とする。その場合、方法としては 2 つ考えられる。

- ファイルに送信専用ホストの IP アドレスを記述し、 mrouted の起動時に読み込む。
- 名前サーバに送信専用ホストを登録しておき、 mrouted の起動時に名前サーバを利用して送信専用ホストの IP アドレスを呼び出す。

今回の実装環境では、送信専用ホストは 1 台であり、また実装に利用するホスト数も少ないことから、名前サーバを利用しなくても十分であると考え、ファイルによる初期設定を選択する。

また、広域マルチキャストのためのトンネルのローカルの終端アドレスの指定についても自由に選択できるようにするために、3.14.5 で説明した mrouted のトンネリングの初期設定と同一形式のファイルを利用して、新しいトンネリングの初期設定を行なうことにする。

```
形式： wide-tunnel <local-addr> <remote-addr>
例　： wide-tunnel 192.50.43.18 133.27.48.154
```

衛星通信型の送信専用ホストへのトンネリングに利用する情報の格納方法

前述のように、広域のマルチキャストのためのトンネリングに利用する情報は、独立した 1 つの vif として扱わない方が mrouted の経路制御情報の交換に関する変更が少ないため、1 つの vif に対する追加情報として扱うことにする。

したがって、具体的には仮想インターフェースの情報の構造体に、新しいトンネリングの情報のための構造を追加する。仮想インターフェースの情報の構造体は、カーネルに格納するための vif 構造体、およびユーザーレベルで格納するための uvif 構造体がある。また、ユーザーレベルからカーネルレベルへの情報交換のために利用する構造体 vifctl に関しても同様の追加が必要となる。変更部分は以下のコメントをつけた部分で、vif の状態を表すフラグの追加と新しいトンネリングのリモートの終端アドレスを格納する構造の追加を行なっている。広域のマルチキャストのためのトンネリングの設定情報は、初期設定ファイルに記述するため、 mrouted の起動時にこれらの構造体へ値を格納する。

よって、「新しいトンネリングのローカルの終端アドレスと vif の実際のインターフェースアドレスが等しい」という条件を満たすような vif には全て VIFF_WIDE_TUNNEL フラグが設定されることになる。

```
#define  VIFF_WIDE_TUNNEL  0x2 /* 広域マルチキャストのためのトンネル */

struct vif {
  u_char          v_flags;       /* VIFF_WIDE_TUNNEL フラグを追加 */
  u_char          v_threshold;
  struct in_addr  v_lcl_addr;
  struct in_addr  v_rmt_addr;
  struct ifnet    *v_ifp;
  struct mbuf     *v_lcl_groups;
  u_long          v_cached_group;
  int             v_cached_result;
  struct in_addr  v_wide_rmt_addr;  /* WIDE_TUNNEL のリモートの終端 */
};


struct uvif {
  u_short         uv_flags;   /* VIFF_WIDE_TUNNEL フラグを追加 */
  u_char          uv_metric;
  u_char          uv_threshold;
  u_long          uv_lcl_addr;
  u_long          uv_rmt_addr;
  u_long          uv_wide_rmt_addr; /* WIDE_TUNNEL のリモートの終端 */
  u_long          uv_subnet;
  u_long          uv_subnetmask;
  u_long          uv_subnetbcast;
  char            uv_name[IFNAMSIZ];
  struct listaddr *uv_groups;
```

```
  struct listaddr *uv_neighbors;
};

struct vifctl {
  vifi_t          vifc_vifi;         /* VIFF_WIDE_TUNNEL フラグを追加 */
  u_char          vifc_flags;
  u_char          vifc_threshold;
  struct in_addr  vifc_lcl_addr;
  struct in_addr  vifc_rmt_addr;
  struct in_addr  vifc_wide_rmt_addr; /* WIDE_TUNNEL のリモートの終端 */
};
```

広域マルチキャストと判定された場合には、マルチキャストデータグラムをユニキャストデータグラムに変換するが、通常のトンネリングと異なるのは、トンネルのリモートアドレスの指定の部分のみである。しかし、*tunnel_send()* では、その引数の 1 つとして vif のポインタを渡すため、その vif が通常の tunnel の設定となっている場合には、*tunnel_send()* のルーチン内でどちらのトンネリングが必要とされているかを判断することが不可能である。判断のための情報は、新たに *tunnel_send()* に追加されなければならない。*tunnel_send()* が呼び出される以前にどのトンネリングが必要かは判断されているため、*tunnel_send()* で再び判断を行なうのは冗長と考え、新しいトンネリングのための関数 *wide_tunnel_send()* を別に用意することにした。したがって、以下の機能を追加することになる。

ip_mforward() において
IF マルチキャストデータグラムがトンネリングされている
THEN IPヘッダから、トンネリングのために使った IPオプション部分
を削る (1**)
IF IPヘッダの $TTL == WIDE\_MULTI\_TTL$
THEN VIFF_WIDE_TUNNEL のフラグの立っている vifを探す
wide_tunnel_send() **へ**
経路制御テーブルから送信ホストに関するエントリを探す (*mrtfind*())(3.14.4 2**)

wide_tunnel_send() において
IPヘッダのオプション部分を削除する
ip_output() へ

3.15.3 トンネルのリモートの終端側の機能

トンネリングの設定

トンネリングのリモートの終端 (衛星通信型の送信専用ホスト側) では、どのマルチキャストルータからトンネリングされても受信できるようにしなければならないので、ファイルの設定時にトンネルのリモートの終端アドレスを“　＊　”で表すことにする。実際に vif 構造体の v_wide_rmt_addr には NULL を入れることとした。つまり、

> VIFF_WIDE_TUNNEL フラグが立ち、かつ v_wide_rmt_addr が NULL であるようなトンネルを持つマルチキャストルータが衛星通信型の送信専用ホストであると判別することが可能となるようにした。

```
wide-tunnel 133.27.48.154 *
```

通常のトンネリングとの判別方法

広域マルチキャストを指定したマルチキャストデータグラムは、トンネリングを通じて、衛星通信型の送信専用ホストに到達する。このとき、通常のトンネリングで到達したか、広域マルチキャストのためのトンネリングで到達したかを判別し、それによって地域内の経路制御を行なうか、衛星通信を利用するかを決定しなければならない。広域マルチキャストのトンネリングの設定は vif の情報に含まれているので、*ip_mforward()* で vif のフラグを検査することによって行なうことになる。

これについては、マルチキャストデータグラムの入ってきたインターフェースに対応する vif やトンネルのリモートアドレスに対応する vif を検索する方法が考えられる。

しかし、トンネリングによって到達するインターフェースは、トンネリングのリモートアドレスと同一のインターフェースであるとは限らない。したがって、マルチキャストデータグラムが入ってきたインターフェースアドレスによってトンネリングの判別を行なうことは困難である。

また、トンネリングは IP オプションの LSRR を利用しているが、このオプションに対する処理 (IP アドレスの入れ換え) は *ipintr()* において、*ip_mforward()* が呼び出される前に行なわれるため、トンネルのリモートアドレスとして指定されたアドレスが保存されるとは限らない。すると、マルチキャストデータグラムのトンネルのリモートアドレスによっても、トンネリングの判別を正しく行なうことは難しい。

したがって、送信専用ホストの持つ vif を元に、トンネリングの判別を行なう方法が考えられる。トンネリングによってデータグラムを受け取った場合に、そのルータが持つ vif の状態としては以下の場合が考えられる。

	通常のトンネルフラグ	広域のためのトンネルフラグ	広域のためのトンネルのリモートアドレス
(1)	○	×	NULL
(2)	○	○	NULL
(3)	×	○	NULL

(1) や (3) の場合は判別は容易である。しかし、(2) の場合、特にその vif で設定されている通常のトンネルのリモートアドレスと、マルチキャストデータグラムのトンネルのローカルの終端アドレスが一致する場合には、判別が不可能である。したがって、さら

にマルチキャストデータグラムの TTL 値によって判断する必要がある。トンネリングを行なう間に TTL 値は減少しているため、ここでの TTL 値の判断を定数で行なうことが不可能である。今回の実装は実験的に TTL 値が WIDE_MULTI_TTL - 15 以上で広域と判断することとした。

よって、トンネルの判別アルゴリズムは以下のようになる。

ip_mforward() において
IF マルチキャストデータグラムがトンネリングされている
THEN
　　FOR 各 vifについて
　　　　IF 送信専用ホストの持つ vifの通常のトンネルのリモートアドレスが
　　　　マルチキャストデータグラムのトンネルの送信アドレスと一致する
　　　　THEN
　　　　　　IF その vifに広域マルチキャストのフラグが立っている
　　　　　　かつ
　　　　　　広域マルチキャストのトンネルのリモートアドレスが **null** である
　　　　　　かつ
　　　　　　データグラムの TTL が閾値以上である
　　　　　　THEN 広域のトンネリングである
　　　　　　　　BREAK
　　　　　　ELSE 通常のトンネリングである
　　　　　　　　BREAK
　　IF まだ判別されていない
　　THEN
　　　　FOR 各 vifについて
　　　　　　IF 広域のトンネルである
　　　　　　かつ
　　　　　　広域のリモートアドレスが **null** である
　　　　　　THEN 広域のトンネリングである **BREAK**
　　　　IF まだ判別されない
　　　　THEN (トンネルの設定ミスであるので) データグラムを捨てる
　　IPヘッダから、トンネリングのために使った IPオプション部分を削る
　　(3.14.4 1**)

衛星通信のインターフェースへの転送機能

送信専用ホストは広域マルチキャストのトンネルから来たデータグラムに対して、今度は衛星通信のインターフェース (送信専用) に対して静的に転送を行なうことになる。そのためには、衛星通信のインターフェースに関する情報が必要となる。実際の衛星通

信のインターフェースに衛星通信であることを示すような情報が用意されるかどうかは未定であるため、ここでは vif の情報として付加することを考えた。そして、vif の構造自体に変更を加えずに済む方法として、衛星通信のインターフェースに関する情報をフラグとして追加することにした。また、衛星通信は基本的にマルチキャスト機能を有するため、このフラグの設定は phyint に対してのみ設定可能となるようにし、wide-tunnel の設定と同様に、ファイルから設定する方法を選択した。

ファイルからの設定

```
phyint 192.50.61.1 sat-send
```

vif に格納するフラグ

```
#define VIFF_WIDE_SAT_SEND 0x4
```

まとめると、以下のように転送されることになる。

ip_mforward() において

IF　マルチキャストデータグラムのトンネリングの判別を行なう

THEN　IPヘッダから、トンネリングのために使ったIPオプション部分を削る (3.14.4 1**)

IF　広域のトンネルで到達した

THEN　$VIFF_WIDE_SAT_SEND$のインターフェースを検索し、転送する経路制御テーブルから送信ホストに関するエントリを探す (*mrtfind*())(3.14.4 2**)

ところで、衛星の送信専用ホストに衛星以外で直接接続しているホストが広域マルチキャストを指定した場合は、衛星の送信専用ホストはそのマルチキャストデータグラムを衛星へのインターフェースに転送することになる。この拡張は *ip_mforward()* におけるトンネルのローカルの終端部分の機能に実装する。

ip_mforward() において

IF　マルチキャストデータグラムのトンネリングの判別を行なう

THEN　IPヘッダから、トンネリングのために使ったIPオプション部分を削る (3.14.4 1**)

IF　IPヘッダの$TTL{=}{=}WIDE_MULTI_TTL$

THEN　$VIFF\_WIDE\_TUNNEL$ のフラグの立っている vifを探す

IF　**トンネルのリモートアドレスが NULL でない**

THEN　*wide_tunnel_send*() へ

ELSE　**VIFF_WIDE_SAT_SEND のインターフェースを検索し、転送する**経路制御テーブルから送信ホストに関するエントリを探す (*mrtfind*())(3.14.4 2**)

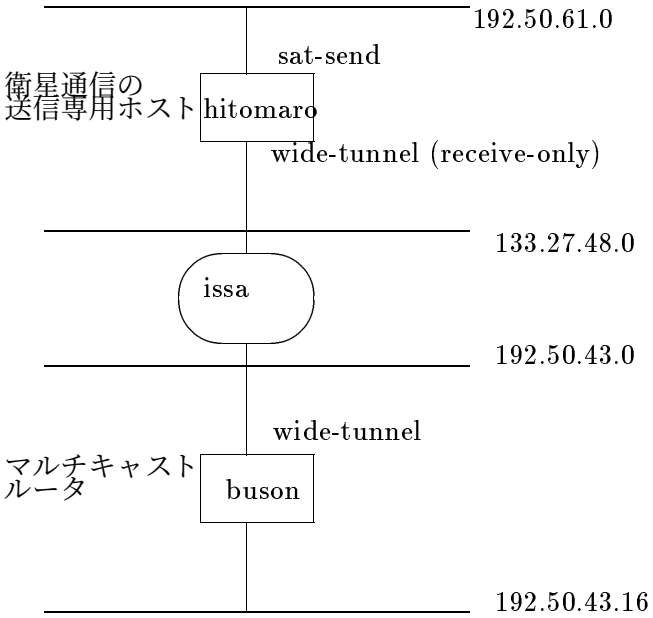

図 3.7: マルチキャストルータの設定

3.16 実行例

ここでは、実装部分の稼働状況を示すことにする。

3.16.1 グループへの所属設定

パッケージに含まれている mtest コマンド (表 3.1参照) を用いて、図 3.6の shikibu と komachi で 224.10.10.10 のグループに所属させた。

各ホストのグループ所属状況は、パッケージに含まれている netstat コマンド (表 3.1参照) でインターフェースの状態を参照することにより、確認することができる。

```
mine@shikibu<~>[20]netstat -nia
Name  Mtu   Network     Address            Ipkts Ierrs    Opkts Oerrs  Coll
am0   1500  192.50.43.1 192.50.43.20       44967     0     3397     0     0
                        224.10.10.10
                        224.0.0.1
lo0   1536  127         127.0.0.1            398     0      398     0     0
                        224.0.0.1
```

3.16.2 mrouted の初期設定

各マルチキャストルータに対して図 3.7 に示すような初期設定を行なった。

これらの初期設定の結果は、 netstat コマンドによって参照できるように改良を加え

た。buson, hitomaro の設定は以下のようになっている。Groups とあるのは、その仮想インターフェースにおいて把握しているグループアドレスである。また、マルチキャストのルーティングテーブルも同時に参照できる。

buson の場合

```
mine@buson</usr/src/vmtp-ip/netstat.bsd>[20]netstat -M
Virtual Interface Table
 Vif   Threshold   Local-Address   Remote-Address   Groups
  0          1      192.50.43.2
                    wide-tunnel     133.27.48.154
                                                    224.0.0.4
  1          1      192.50.43.18
                                                    224.0.0.4
                                                    224.10.10.10

Multicast Routing Table
 Hash  Origin-Subnet  In-Vif  Out-Vifs
  16   192.50.43.16       1    0*
  43   192.50.43          0    1*
```

hitomaro の場合

```
mine@hitomaro</usr/src/vmtp-ip/netstat.bsd>[20]netstat -M
Virtual Interface Table
 Vif   Threshold   Local-Address   Remote-Address   Groups
  0          1      192.50.61.1
                    wide-sat-send
                                                    224.0.0.4
                                                    224.10.10.10
  1          1      133.27.48.154
                    wide-tunnel     receive-only
                                                    224.0.0.4

Multicast Routing Table
 Hash  Origin-Subnet  In-Vif  Out-Vifs
  48   133.27.48          1    0*
  61   192.50.61          0    1*
```

3.16.3　ping による動作確認

パッケージに含まれる ping コマンド (表 3.1参照) は オプションによって、ping のマルチ キャストデータグラムの TTL 値を指定することが可能である。そこで、komachi から ping を実行させることによって、広域マルチキャスト の実装部分の動作確認を行なった。

図 3.6 において、issa はマルチキャストルータではないため、issa を境界として地域が形成された状態となっている。shikibu と komachi は別の地域に所属していることになるため、komachi から TTL 値を 200 未満にして 224.10.10.10 に対して ping を行なうと komachi(192.50.43.20) からしか応答が得られない。

```
mine@komachi</usr/src/vmtp-ip/ping>[35]./ping -t 199 224.10.10.10
PING 224.10.10.10 (224.10.10.10): 56 data bytes
64 bytes from 192.50.43.20    icmp_seq=0 time=0 ms
64 bytes from 192.50.43.20    icmp_seq=1 time=0 ms
64 bytes from 192.50.43.20    icmp_seq=2 time=0 ms
64 bytes from 192.50.43.20    icmp_seq=3 time=0 ms
64 bytes from 192.50.43.20    icmp_seq=4 time=0 ms
^C

----224.10.10.10 PING Statistics----
5 packets transmitted, 5 packets received, 0% packet loss
round-trip (ms)  min/avg/max = 0/0/0
```

しかし、マルチキャストデータグラムの TTL 値を 200 に設定した際には、広域マルチキャストのトンネリングが行なわれ、以下のように shikibu(192.50.61.2) からも応答が得られる。

しかし、マルチキャストデータグラムの TTL 値を 200 に設定した際には、広域マルチキャストのトンネリングが行なわれ、以下のように shikibu(192.50.61.2) からも応答が得られる。

第 4 章

今後の課題

この章では、IP マルチキャストを用いた通信の今後の研究課題について議論する。

4.1 今回の実装に対する問題点および考察

4.1.1 TTL を基準とした広域マルチキャストの判定について

送信ホストから、最初にマルチキャストデータグラムを受信したマルチキャストルータは、TTL の値の比較を単純な定数の比較によって行なうことができた。

しかし、衛星の送信専用ホストにおいてはトンネリングによって受信したマルチキャストデータグラムを、再び TTL の値を基準として広域かどうかを判断する必要が生じた。というのは、今回利用した既存のマルチキャストの経路制御方法では、トンネリングを広域マルチキャスト以外にも利用することができるようになっていたためである。既存のマルチキャストの経路制御方法で用意されているトンネリングは、地域内で行なわれるため、その場合には、地域内の経路制御に委ねる必要がある。衛星の送信専用ホストに到達するまでに、TTL の値は不確定に減少するため、広域のためのトンネルと、地域内で利用されるトンネルの判別を行なう TTL の閾値を決定するのは難しい。

これは、TTL を基準とした広域マルチキャストアルゴリズムの問題点の 1 つであると考えられる。

4.1.2 衛星へのインターフェースの表現について

衛星の送信専用ホスト、あるいは受信専用ホストが持つ衛星へのインターフェースは、一方向である。今回の実装では、一方向性という情報は 仮想インターフェースの情報に含めたが、本来はインターフェースの情報として得られるべきであると考えられる。実際に衛星を利用して、広域マルチキャストを行なう場合には、衛星のインターフェースであることを実際のインターフェースの情報から検出するように変更するのが望ましいと考えられる。

4.1.3 不必要な広域マルチキャストデータグラムの配送の可能性について

広域マルチキャストを利用可能とするための方法について実装を行なったが、TTL 値を 200 にすることだけが広域マルチキャストを利用するための条件となっている。すると、誤って、あるいは故意に、実際に利用されていないグループを利用して広域マルチキャストを指定した場合、各地域内の衛星受信専用ホストまでは到達してしまう。このようなことが故意に行なわれると、衛星の送信専用ホストに向かう経路のトラフィックが増加し、その経路を利用する他の通信にも影響を与えかねない。これに対する解決策としては、衛星送信専用ホストにおいて、登録済みのグループアドレスリストを用いてアドレスの検査を行なうことが挙げられる。

4.2 まとめ

世界規模に発展しつつある計算機ネットワークをより効率的に利用するために、既存の通信形態 (1 対 1 型通信および一斉同報型通信) に加え 1 対 n 型通信 (マルチキャスト型通信) の広域ネットワークへの適用の必要性、有効性、そしてその実現のための技術について述べてきた。

既存の技術では、物理媒体としてのレベルでは、1 対 1 型通信と同様のコストでマルチキャスト型通信を行なえるところまで発展している。さらに、そのような機能を持つ物理ネットワークの接続によって構成される、比較的狭域の異機種間ネットワークにおけるマルチキャストの技術についても幾つかの経路制御アルゴリズムが提案され、実装も行なわれてきている。既存のマルチキャスト経路制御においては、ネットワークホストの集合を表すグループアドレスをデータグラムの目的地アドレスとして利用する、グループアドレッシングの方法がとられている。

本研究では、このようなグループアドレッシングを前提とした上で、低速ネットワークによって接続されるような広域ネットワークにおいてもマルチキャスト機能が必要であることを述べた。その際物理的に距離の離れたホスト間を直接接続することの可能な放送型広域通信媒体の利用が、既存の低速ネットワークのトラフィックを回避するためにも不可欠であると判断し、放送型広域通信媒体を利用したマルチキャストの経路制御アルゴリズムに対する様々な提案を行なった。

放送型広域通信媒体としては、特に衛星を取り上げた。衛星は、計算機ネットワークとして利用する視点から見ると、基本的に一斉同報型の物理媒体であるという特徴の他に、送信局と受信局が固定された、一方向性の物理媒体という特徴も見られる。そこで、まず既存の経路制御アルゴリズムについて、一方向性のマルチキャスト型媒体を想定してうまく機能するかどうかの検証を行なった。

この結果、既存の経路制御アルゴリズム (DVMRP) に一方向性の物理媒体を利用しようとすると、多くの変更を行なわなければならないことがわかった。また、1 つの経路制御アルゴリズムで広域に渡るマルチキャストを経路制御すること自体にも経路制御情報の大規模性という点で問題があるため、既存の経路制御アルゴリズムの組み合わせと衛

星の利用による広域マルチキャストの経路制御アルゴリズムの考案を行なった。その際、特に衛星の受信専用ホストの配置とそれにともなう地域分け、衛星の送信専用ホストまでの経路、および衛星の受信専用ホストからの経路の 3 段階に分けて様々なアルゴリズムを挙げ、考察を行なった。

その結果として得られた幾つかの広域マルチキャストのための経路制御アルゴリズムについて 1 つのアルゴリズムを選択した。これは、マルチキャストデータグラムの生存時間を基準として、広域か地域内かを判別し、経路制御を行なうアルゴリズムである。そして特に、送信専用ホストから衛星の送信専用ホストまでの経路制御の部分について、実装を行なった。

この実装は、主に広域マルチキャストのためのトンネリング機構を構築するもので、既存のマルチキャスト経路制御情報の交換方法に、できるだけ変更を加えない方針で行なった。結果として、衛星の送信専用ホストについての経路制御機構の追加は、既存の経路制御アルゴリズム内に衛星を組み込む場合と比較すると、既存の経路制御のアルゴリズム自体への影響が小さいという意味で、自由度の高い実装が可能となることを示した。

4.3　今後の課題

- グループメンバーの認証機構について
  本論文で扱ったマルチキャストの伝搬機構は、グループアドレッシングを前提としていたが、グループへの所属および離脱に関しては、他のグループメンバーとの関連がないため、受信されるべきでないホストがマルチキャストデータグラムを受け取る可能性もある。これは、グループアドレッシングの方法で、一般的な問題として取り上げられる認証問題であるが、やはり基本的には配送を行なうレベルではなく、それ以上のレベルにおいて、送信ホストとグループメンバー間 (エンド–エンド間) で認証を行なうのが望ましいと考えられる。

- マルチキャストデータグラムの信頼性の保障について
  本研究では今回考慮しなかったが、エンド–エンド間におけるマルチキャストデータグラムの信頼性の保障に関しては、マルチキャスト機能を利用するアプリケーションを考える上で重要な問題である。衛星を利用する場合、他の物理媒体と比較して、媒体自体の伝送誤りの確率が高いため、データの信頼性の保障は特に必要性が高いといえる。

- アプリケーションが要求するマルチキャスト型通信の提供について
  データの信頼性の保障の方法にも関わってくるが、広域マルチキャストはまだ実用段階に至っていないため、広域マルチキャストを利用するアプリケーションがグループメンバーに対してどのような通信を要求するかは一概に決定することは不可能である。従って、これらの傾向に関しては、広域マルチキャストを実際に運用しながら、序々に解決していくべき問題である。

- 衛星の送信専用ホスト数について
  衛星通信の利用に当たっては、現状では衛星の送信専用ホスト数が受信専用ホスト数に比較して少ないという想定であるが、将来送信専用ホストに関する設置コストがさらに低下すれば、送信専用ホスト数と受信専用ホスト数の比率が同程度になることも予想される。その場合、計算機ネットワーク上の任意のホストから衛星の送信専用ホストへの距離の平均が小さくなるように利用すれば、広域マルチキャストの有効性はさらに高まると考えられる。

- データ量の増加とマルチキャストの関連について
  将来、通信されるデータに、音声・画像等の新しいメディアデータが加えられることを考慮すると、マルチキャストを利用してトラフィック量を最低限に押えることは、計算機ネットワークを利用したアプリケーションの多様化に影響するという点でますます重要となってくる。

- 実際に衛星を利用した広域マルチキャストの実験について
  平成 4 年 7 月に、実際に衛星を利用した広域マルチキャストの実験運用が開始される予定である。図 4.1に示すように、衛星送信局は市ヶ谷に配置し、WNOC 東京との間を当初 INS64 で接続する。また、衛星受信局に関しては、まず、電気通信大学・ASTEM・東京大学・OMRON・慶応義塾大学 SFC に配置される予定である。

  この実験運用によって、実際の広域マルチキャストの利用の特性等についての考察を行なうことが可能となる。そして、それらの考察を元に、広域マルチキャスト経路制御方法を最適な方向へ改良していくことが求められる。

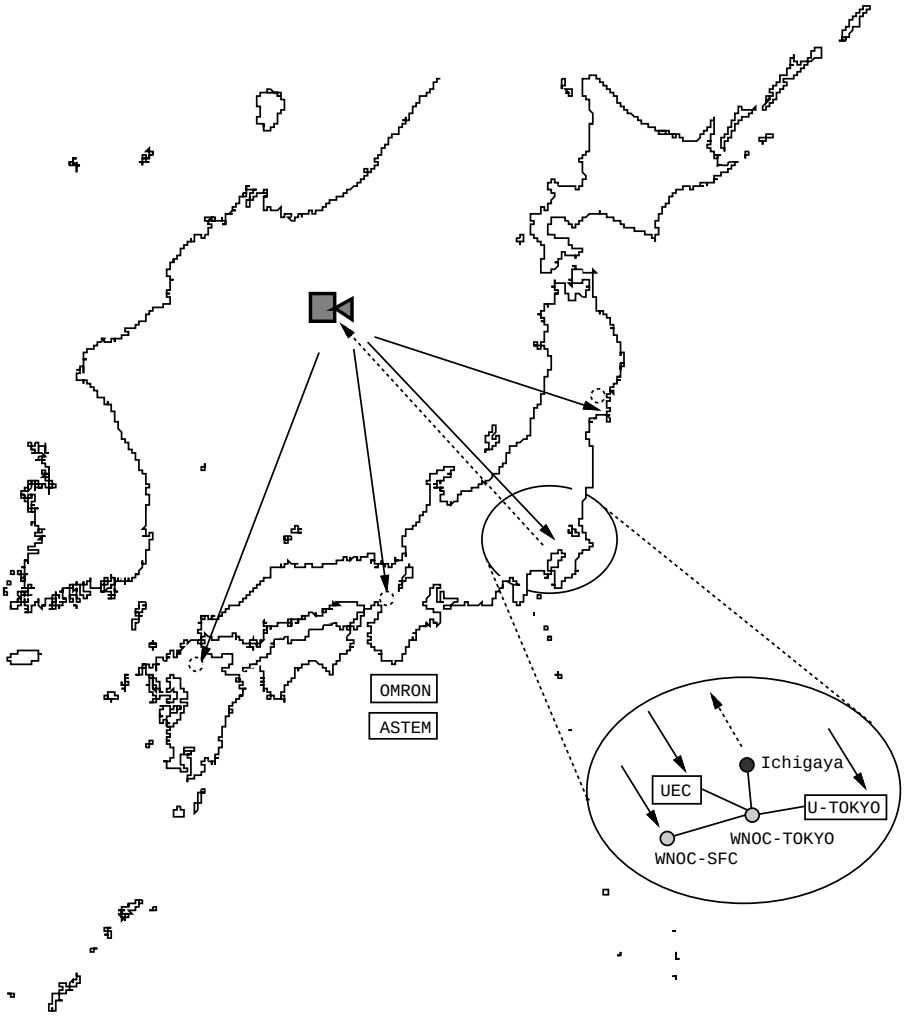

図 4.1: 衛星を利用した実験予定図