

インターネット上のマルチキャストサーバの 設計と実装

石川 憲洋

NTT 情報通信研究所

インターネット上の仮想的なマルチキャスト通信ネットワークである MBone 上で、様々なマルチメディアツールを用いた実験が実施されている。その経験は、IP マルチキャスト通信が、インターネット上でのマルチメディア通信サービスのための基盤技術であることを示している。しかしながら、IP マルチキャスト通信は、今のところ実験段階にあり、実用レベルには達していない。

本稿では、IP マルチキャスト通信をサポートしていない現在のインターネット上で、IP マルチキャスト通信と同じ機能を実現するために、アプリケーションプロトコルレベルでマルチキャスト通信の機能を実現する、インターネット上のマルチキャストサーバの設計と実装について報告する。

Design and Implementation of Multicast Servers over the Internet

Norihiro Ishikawa

NTT Information and Communication Systems
Laboratories

The various multimedia tools have been tested on MBone, a virtual multicast network on the Internet. The experience shows that IP multicast is a core technology for multimedia services over the Internet. However, IP multicast is now at the experimental stage, and hence has not been widely deployed over the existing Internet at this time.

In this paper, we describe the design and the implementation of multicast servers over the Internet, which provide multicast communication functions over the existing Internet that does not support IP multicast. Multicast servers provide the same services as IP multicast, by using the application level protocol that provide multicast communication functions.

1 はじめに

インターネット上の仮想的なマルチキャスト通信ネットワークである MBone[1]上で、様々なマルチメディアツールを用いたマルチメディア通信実験が実施されている。その経験は、IP マルチキャスト通信が、インターネット上でのマルチメディア通信サービスのための基盤技術であることを示している。

しかしながら、IP マルチキャスト通信は、運用の難しさ、スケーラブルな IP マルチキャストルーティング技術の未確立などの理由から、今のところ実験段階にあり、実用レベルには達していない。

本稿では、IP マルチキャスト通信をサポートしていない現在のインターネット上で IP マルチキャスト通信と同じ機能を実現するために、アプリケーションプロトコルレベルでマルチキャスト通信の機能を実現する、インターネット上のマルチキャストサーバ[2]の設計と実装について報告する。さらに、IP マルチキャスト通信に対応した既存アプリケーションのマルチキャストサーバ上への移植実験を行ったので、その結果についても述べる。

2 アーキテクチャ

提案するインターネット上のマルチキャストサーバのアーキテクチャを図1に示す。図1に示すように、提案するアーキテクチャは、スケーラビリティを確保するために、1台のマルチキャストサーバで構成するのではなく、複数台のマルチキャストサーバが木構造で接続される構成とした。木構造を構成するマルチキャストサーバ間には上下関係がある。受信ホストは、通常、最下位のサーバと接続する。また、既存の IP マルチキャスト通信ネットワーク (MBone など) との相互接続も可能としている。

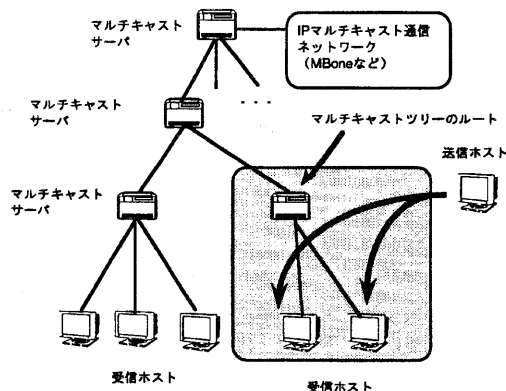


図 1 マルチキャストサーバのアーキテクチャ

マルチキャストされるデータ (以下、マルチキャストデータと呼ぶ) は、ホストグループ (マルチキャストデータの受信を要求する 0 台以上の受信ホスト) に対して送信される。受信ホストは、いつでもホストグループに参加 (join) 又は離脱 (leave) できる。ホストグループは、グループ識別子により識別される。グループ識別子は、IP マルチキャスト通信のグループアドレス (クラス D の IP アドレス) に相当するもので、その長さは、IP マルチキャスト通信との相互接続を考慮して、クラス D の IP アドレスと同じ 32 ビットとした。

マルチキャストデータの送信先であるホストグループ毎に、マルチキャストツリーとそのルートとなるマルチキャストサーバが定義される。送信ホストは、ホストグループに対して定義されたマルチキャストツリーのルートに対して、マルチキャストデータを送信する。マルチキャストデータは、マルチキャストツリーのルートから、中間のマルチキャストサーバを経由して、最終的には、受信ホストまでルーティングされる。

社内ネットワーク (イントラネット) の例で考えると、最上位のサーバは全社レベルのサーバ、最下位のサーバはオフィスレベルのサーバと考えることができる。全社レベルのサーバがルートとして指定された場合は、全社レベルのマルチキ

キャスト通信が実現できる。オフィスレベルのサーバがルートとして指定された場合は、オフィスレベルのマルチキャスト通信が実現できる。このように、木構造のアーキテクチャを採用することにより、階層的かつスケーラブルなマルチキャスト通信が実現できる。

3 プロトコル (MDSP)

マルチキャストサーバを実現するために、MDSP (Multicast Distribution Server Protocol) と呼ぶプロトコルを UDP/IP 上で設計した (図 2)。

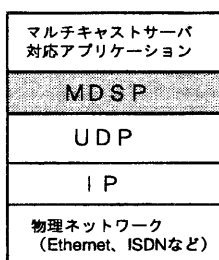


図 2 MDSP の位置付け

MDSP で定義しているメッセージと主なパラメータを表 1 に示す。MDSP メッセージは、ユニキャスト UDP/IP を使用して、送信ホスト、マルチキャストサーバ、受信ホスト間で送受信される。

表 1 MDSP のメッセージ

メッセージ	主なパラメータ
Upstream	グループ識別子、ペイロードタイプ、マルチキャストデータ
Downstream	グループ識別子、ペイロードタイプ、マルチキャストデータ
Join	グループ識別子、認証、タイマ
Join Ack	グループ識別子
Join Nak	グループ識別子、理由
Leave	グループ識別子
Alive	グループ識別子
Alive Ack	グループ識別子

4 方式概要

提案するインターネット上のマルチキャストサーバ方式の概要について述べる (図 3)。

4. 1 マルチキャストデータの送信

送信ホストは、マルチキャストデータの送信先であるホストグループに対して定義されたマルチキャストツリーのルートであるサーバに対して、そのホストグループを識別するグループ識別子を設定した Upstream メッセージを使用して、マルチキャストデータを送信する。

4. 2 マルチキャストデータの受信開始

マルチキャストデータの受信開始を要求する受信ホストは、参加するホストグループを識別するグループ識別子を設定した Join メッセージを、その受信ホストが接続するマルチキャストサーバに送信する。

Join メッセージは、中間のマルチキャストサーバを経由して、ホストグループに対して定義されたマルチキャストツリーのルートであるサーバまで、順次伝搬される。

但し、中間のマルチキャストサーバが、上位サーバに Join メッセージを送信するのは、最初に、受信ホスト又は下位サーバから Join メッセージを受信した場合のみで、以降、受信ホスト又は下位サーバからそのグループ識別子を設定した join メッセージを受信した場合は、上位サーバには再度送信しない。

各マルチキャストサーバでは、各グループ識別子毎に、グループ識別子と、そのグループ識別子を持つマルチキャストデータを受信中の受信ホスト及び下位サーバのリストを、ルーティングテーブルにより管理する。

受信ホスト又は下位サーバから Join メッセージを受信した時、各マルチキャストサーバは、そのノードの IP アドレスを、ルーティングテーブルに追加する。

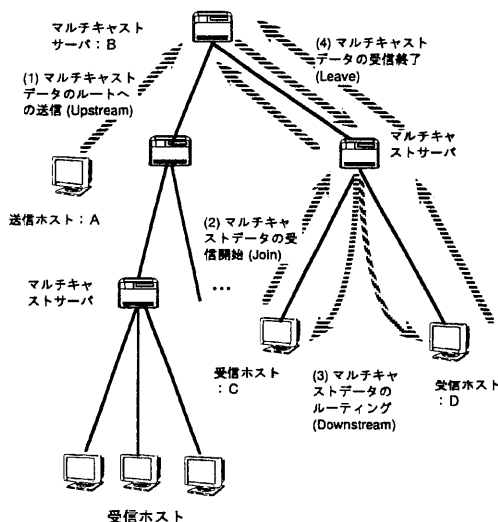


図 3 マルチキャストサーバの方式概要

4. 3 マルチキャストデータのルーティング

ホストグループに対して定義されたマルチキャストツリーのルートであるサーバは、送信ホストから、マルチキャストデータを、Upstream メッセージとして受信する。

マルチキャストサーバは、ルーティングテーブルを検索し、受信した Upstream メッセージに設定されたグループ識別子を持つマルチキャストデータを受信中の受信ノード及び下位サーバが存在する場合、各ノードに対して、受信したマルチキャストデータを、Downstream メッセージに変換して送信する。

同様に、上位サーバから Downstream メッセージとしてマルチキャストデータを受信したマルチキャストサーバは、ルーティングテーブルを検索し、受信した Downstream メッセージに設定されたグループ識別子を持つマルチキャストデータを受信中の受信ノード及び下位サーバが存在する場合、各ノードに対して、受信したマルチキャストデータを、Downstream メッセージとして送信する。

このように、マルチキャストデータは、ルーティングツリーのルートであるサーバから、中間のマルチキャストサーバを経由して、最終的には、

受信ホストまで、Downstream メッセージとしてルーティングされる。

4. 4 マルチキャストデータの受信終了

マルチキャストデータの受信終了を要求する受信ホストは、離脱するホストグループを識別するグループ識別子を設定した Leave メッセージを、その受信ホストが接続しているマルチキャストサーバに送信する。

受信ホスト又は下位サーバから Leave メッセージを受信した時、マルチキャストサーバは、そのノードの IP アドレスを、ルーティングテーブルから削除する。

IP アドレスを削除した結果、そのグループ識別子を持つマルチキャストデータを受信中の受信ホスト又は下位サーバが存在しなくなった場合、マルチキャストサーバは、そのグループ識別子を設定した Leave メッセージを上位サーバに送信する。

このように、Leave メッセージは、受信ホストから、中間のマルチキャストサーバを経由して、最終的には、ホストグループに対して定義されたマルチキャストツリーのルートであるサーバまで、順次伝搬される。

4. 5 キープアライブ (Keep Alive)

マルチキャストツリーを構成する、マルチキャストサーバ-マルチキャストサーバ間、マルチキャストサーバ-受信ホスト間は、LAN、ルータなどで構成された IP ネットワーク経由で接続されている。従って、マルチキャストサーバは、接続している上位サーバ、下位サーバ、受信ホストが障害等でダウンした場合、それを検出することは困難である。結果として、障害等でダウンした上位サーバからのマルチキャストデータを待ち続けるなどのデッドロックが生じる可能性がある。

上記デッドロックを回避するために、マルチキャストサーバ-マルチキャストサーバ間、マルチキャストサーバ-受信ホスト間に、いわゆるキープアライブ機能を導入した。

マルチキャストサーバ及び受信ホストは、接続している（上位）サーバに対して、定期的に Alive メッセージを送信する。Alive メッセージを受信した（上位）サーバは、Alive Ack メッセージで応答する。

一定期間、Alive メッセージを受信しなかった場合、マルチキャストサーバは下位サーバ又は受信ホストがダウンしたとみなし、接続関係を解除する。同様に、一定期間、Alive Ack メッセージを受信しなかった場合、受信ホスト又はマルチキャストサーバは（上位）サーバがダウンしたとみなし、接続関係を解除する。

4. 6 IP マルチキャスト通信ネットワークとの相互接続

IP マルチキャスト通信ネットワークと接続する場合、IP マルチキャスト通信のグループアドレスの値を、そのまま、マルチキャストサーバのグループ識別子の値として使用する。

また、IP マルチキャスト通信ネットワークとの接続点は、マルチキャストツリーのルートであるサーバのみとした。

4. 6. 1 IP マルチキャスト通信ネットワークへの送信

マルチキャストツリーのルートであるサーバは、送信ホストから Upstream メッセージとして受信したマルチキャストデータを、IP マルチキャストデータグラムに変換して、IP マルチキャスト通信ネットワークに送信する。

4. 6. 2 IP マルチキャスト通信ネットワークからの受信

マルチキャストツリーのルートであるサーバは、受信ホスト又は下位サーバから Join メッセージを受信すると、IP マルチキャスト通信ネットワークに対して、IP マルチキャストデータグラムの受信開始を要求する。

その後、IP マルチキャスト通信ネットワークから受信した IP マルチキャストデータグラムを、

Downstream メッセージに変換して、マルチキャストデータを受信中の下位サーバ及び受信ホストに送信する。

受信ホスト又は下位サーバから最後の Leave メッセージを受信して、マルチキャストデータを受信中の受信ホスト又は下位サーバが存在しなくなった場合、IP マルチキャスト通信ネットワークに対して、IP マルチキャストデータグラムの受信終了を通知する。

5 実装

提案したマルチキャストサーバの実装を WS、PC 上で進めている（図 4）。

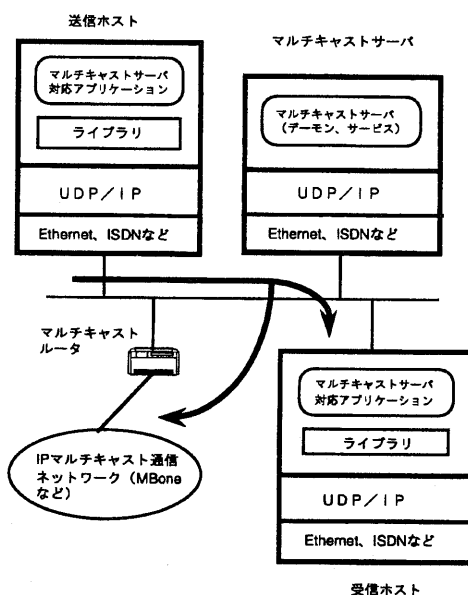


図 4 マルチキャストサーバの実装

図 4 に示すように、送信ホスト、受信ホストの機能はライブラリとして、マルチキャストサーバの機能は、デーモン（UNIX）及びサービス（Windows NT）として実装した（表 2）。マルチキャストサーバは、通信インタフェースとして、UNIX の標準的なソケットインタフェースのみを使用して実装しているため、UNIX の種別に依存

することなく動作可能である。また、UNIX 環境から Windows NT/95 環境への移植も容易に実現できた。

表2 マルチキャストサーバの実装環境

機能	実装環境
送信ホスト、 受信ホスト	UNIX、Windows NT、 Windows 95 (ライブラリ)
マルチキャスト サーバ	UNIX (デーモン)、 Windows NT (サービス)

送信又は受信を行うアプリケーションとマルチキャストサーバが、同一システム内に存在してもよい。この場合、送信又は受信アプリケーションとマルチキャストサーバの間は、ソケットインタフェースを利用したローカルなプロセス間通信を行う。特に、同一システム内に複数の受信アプリケーションが存在する場合は、各アプリケーションにマルチキャストデータを分配するために、同一システム内でマルチキャストサーバを動作させる必要がある。

送信及び受信アプリケーションに提供している主なライブラリ関数を表3に示す。

表3 送信及び受信アプリケーションに提供しているライブラリ関数

名前	機能
mds_get_root_address	マルチキャストツリーのルートとして定義されたサーバの IP アドレスを返す
mds_send	マルチキャストデータを送信する
mds_recv	マルチキャストデータを受信する
mds_join	マルチキャストデータの受信を開始する
mds_leave	マルチキャストデータの受信を終了する

ソフトウェアの規模は、現在までのところ、マルチキャストサーバが約 800 行、ライブラリが約 400 行、ヘッダ等が約 100 行であり、非常にコンパクトに実現することができた。

6 IP マルチキャスト通信対応アプリケーションの移植実験

提案したインターネット上のマルチキャストサーバの有効性を確認するために、以下に示す IP マルチキャスト通信対応アプリケーションのマルチキャストサーバ上への移植実験を行った。

- vic [3]、vat、RMTP [4]

表3のライブラリ関数は、IP マルチキャスト通信のためにソケットインタフェースで定義しているシステムコールとほぼ1対1に対応付け可能である。従って、移植作業は、各アプリケーションが IP マルチキャスト通信のためのシステムコールを呼び出している部分を、表3のライブラリ関数で置き換えることで、ほぼ機械的に行うことができた。結果として、移植作業は、各アプリケーション毎に、ソースコードの解析も含め、1人月以内で完了した。

7 まとめ

インターネット上のマルチキャストサーバの設計と実装について述べた。

今後は、性能評価、マルチキャストサーバの新しいアプリケーションへの適用等について検討を進める予定である。

参考文献

- [1] V. Kumar: MBone: Interactive Multimedia on the Internet, New Riders Publishing, 1996.
- [2] 石川: インターネット上のマルチキャストサーバの提案, 情報処理学会第 53 回 (平成 8 年後期) 全国大会, 3-435, 1996.
- [3] S. McCanne and V. Jacobson: vic: A Flexible Framework for Packet Video, ACM Multimedia '95, Nov. 1995.
- [4] 城下、高橋、佐野、山下、中村、山内、串田: インターネットに適用可能な高信頼一斉分配システム, 情報処理学会研究報告 95-AVM-11, Dec. 1995.