

多対多マルチキャストにおけるメンバークラスタリング手法

大崎 功一 川田 雅人 中内 清秀 森川 博之 青山 友紀

東京大学大学院情報理工学研究科

〒113-8656 東京都文京区本郷7丁目3番1号

Tel:03-5841-6710 Fax:03-5841-6776 Email:koichi@mlab.t.u-tokyo.ac.jp

あらまし

ネットワークゲームなどの多対多参加型アプリケーションでは、アプリケーションを構成する複数のデータに対する受信者の要求に多様性があるため、この多様性に対応したデータ配信を行なう必要がある。また、多数の受信者にデータを送信するため、マルチキャスト通信が効率的である。このとき、受信者の要求の重複が大きくなるように受信者を幾つかのクラスタに分類し、使用するマルチキャストアドレス数を節減する手法がある。しかし、受信者をクラスタリングした場合、複数のクラスタに送信すべき送信者が生じ、通信性能が劣化する恐れがある。

そこで本稿では、受信者のデータに対する要求に応じつつ、送信先アドレス数が多い送信者の数を抑えるメンバークラスタリング手法を提案する。本手法をシミュレーションで評価した結果、不要データ量が少し増加するものの、マルチキャストツリー数を制限しながら送信負荷の高い送信者数を抑えることができることを示す。

A Member-Clustering Scheme for Many to Many Multicast

Koichi OHSAKI Masato KAWADA Kiyohide NAKAUCHI
Hiroyuki MORIKAWA Tomonori AOYAMA

School of Information Science and Technology,

The University of Tokyo 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN

Tel: +81-3-5841-6710 Fax: +81-3-5841-6776 Email: koichi@mlab.t.u-tokyo.ac.jp

Abstract

The efficiency of using multicast in multi-party application such as networked games is constrained by preference heterogeneity of receivers. It is desirable to cluster receivers into approximately similar groups to maximize overlap in preference and satisfy the constraint of the limited number of multicast addresses. This method, however, increases transmission burden of some sources by sending data to multiple clusters. In this paper we present the member clustering method which takes into account both receiver preferences and source burden. We found by simulation that our method improves the burden of sources much at the cost of a little decrease of data goodput.

1. はじめに

インターネット利用者数の急増により、ビデオ会議[1]やネットワークゲーム[2]、分散シミュレーション[3]などの多対多参加型アプリケーションが注目されてきている。多数の受信者にデータを送信する必要があるこのようなアプリケーションでは、効率的な1対多通信を実現するマルチキャスト技術を用いて効率的に通信をすることが望ましい。

しかし、多種多様なデータからなるこれらのアプリケーションでは、送信データに対する受信者の要求も多種多様になることが考えられる。例えば、ネットワークゲームの場合、仮想世界上で近隣に位置する参加者とはインタラクションが強く、遠隔に位置する参加者よりも詳細なデータを要求する[4]。

このような状況下でマルチキャストを適用する場合、全ての送受信者を1つの共有木で結ぶと、受信者が不

要なデータを大量に受信してしまう恐れがある。不要なデータを防ぐために、中間ルータで不要なデータを取り除く[5,6]という手法もあるが、現在のルータを高機能化する必要があり現実的でない。また、送信データ毎に1対多のマルチキャストツリーを作成する手法もあるが、送信データ数が増えるにつれてマルチキャストツリー数も増加し、ルータや受信者の負荷増大の弊害が生じてしまう[4]。このため、不要なデータの受信を避け、かつマルチキャストツリー数を抑えることが求められる。

これに対し、受信者のデータ要求に適應して参加者をいくつかのクラスタに分類し、各クラスタにマルチキャストアドレスを与えてデータを送信するクラスタリング手法[4,7,8]が提案されている。クラスタリング手法により、従来のネットワークに大きな改良を加えずに、受信者の不要データ受信量を小さく保ちながら必要なマルチキャストアドレス数を減らすことができる。本稿で想定する多

対多参加型アプリケーションでは、受信者をクラスタリングする方法が、不要なデータ量の削減、join/leave メッセージの削減の面で適している[4]。しかしこのとき、送信者は自分のデータを要求する受信者のいる全てのクラスタのマルチキャストアドレス宛にデータを送信するため、多くの受信者が要求するデータを持つ送信者は、多くのアドレス宛に送信することになってしまい、その送信者には、過大な送信負荷がかかることになる。

そこで本稿では、受信者の要求に適応し、かつ送信者の負荷を抑えるメンバークラスタリング手法を提案する。この手法は、受信者をクラスタリングする際に送信者の送信先アドレス数を参照し、送信先アドレス数の多い送信者群が存在する場合はその送信者群に対して新たにマルチキャストアドレスを割り当て、再度クラスタリングすることにより送信者の負荷を抑えるというものである。

以下、2. では、多対多アプリケーションにおいて、マルチキャスト適用の際の問題点とその解決へのアプローチを紹介し、送信者の負荷を減少させる必要性を論じる。3. では本稿で提案するメンバークラスタリング手法の具体的方法を説明し、4. では ns (Network Simulator) によるシミュレーションにより、本手法の有効性を評価する。最後に5. で結論を述べる。

2. 多対多マルチキャストアプリケーションの運用

多数の参加者がコミュニケーションを行なうアプリケーション[1,2,3]では、多数の受信者にデータを送信する必要があるため、マルチキャスト技術により効率的に通信をすることが望ましい。しかし、受信者は必ずしも全ての送信データを要求するとは限らないため、受信者の要求に従って適切にマルチキャスト配信することが求められる。

2.1 受信者の要求に従ったマルチキャスト配信

多数の参加者へマルチキャスト配信を行なう場合、まず全ての参加者を1つの共有木で結ぶ方法と、送信データ毎に送信者を根とするマルチキャストツリーを構成する方法が考えられる。

前者の方法では、全ての送信者のデータが受信者に到達するため、受信者は不要なデータをフィルタリングする必要があり、また不要なデータによるネットワーク資源の浪費を引き起こす。

後者の方法では、送信データの数だけマルチキャストアドレスが必要になるため、送信データ数が膨大なアプリケーションでは必要なマルチキャストアドレス数も膨大になる。一般的に、マルチキャストはアドレス空間の枯渇、ルーティングテーブルにおいて集約できないアドレス形態、ルーティングパケットや join / leave パケットといった大量の制御パケットを必要とする、といった固有の問題を抱えているため、マルチキャストツリー数が増大することは望ましくない[9]。また、受信者は要求が変わるたびに join/leave メッセージの交換を行なう必要がある、非常に制御が複雑になる。

そこで、不要なデータの受信を避けつつ、上記の問題点を解決するために、受信者のデータに対する要求

に従って効率よくマルチキャストツリー数を減らす手法がいくつか提案されている。これらの手法は、ネットワーク内で不要なデータを破棄するフィルタリング手法と、参加者を要求に応じて分類するクラスタリング手法の2つに分類される。

フィルタリング手法には、中間ルータが下流の受信者の要求データを記憶し、不要なデータを破棄するルータレベルフィルタリング[5]と、サブキャストツリーを生成する AIM[6]がある。しかしこれらはルータを高機能化する必要があり、データ数が増えたとルーティングエントリーが増大しルータの処理負荷が高まってしまふ

クラスタリング手法[4,7,8]では、受信者の要求データや参加者のネットワーク環境を基に、不要なデータ配信を避けるように参加者をクラスタに分類し、ツリー数を削減する。これらはネットワーク設備に手を加えずに、限られたネットワーク資源の利用効率を高めることができる。

2.2 クラスタリング手法とその問題点

以降本稿では、実装における制約条件が少ないクラスタリング手法に注目して議論を進める。

代表的なクラスタリング手法として、部分最適に収束する k-means 手法[10]を用いて、限られたグループ数の中で受信者の要求の重なりを大きくするクラスタリングアルゴリズム[4]がある。このアルゴリズムでは、送信者をクラスタリングする Grouping-Sender (GS) Scheme と受信者をクラスタリングする Grouping-Receiver (GR) Scheme がある。

GS では、送信者は自分が属するクラスタのマルチキャストアドレス宛にデータを送信し、受信者は自分の要求するデータが属するクラスタに割り当てられたマルチキャストアドレス宛に join メッセージを送信し、データを受信する。GS は配信型アプリケーションに向いている。

GR では、受信者は自分が属するクラスタのマルチキャストアドレス宛に join メッセージを送信し、送信者は自分のデータを要求するクラスタに割り当てられたマルチキャストアドレス宛にデータを送信する。一般に、ネットワークゲームなどの参加型アプリケーションでは、仮想空間上で近隣に存在する受信者達は、似通ったデータを要求することから、GR を使用することが望ましい。

しかし GR では、送信者は自分のデータを要求している受信者の属する全てのクラスタのマルチキャストアドレス宛にデータを送信しなければならない。図 1 は、参加者 100 人のネットワークゲームを想定して GR を用いた際の、使用マルチキャストアドレス数と最大送信先アドレス数の関係を示している。このように GR では多数のクラスタに対してデータを送信する送信者が生じてしまふこれにより、送信者の送信負荷増大、送信者付近のリンクの消費帯域増大、それに伴うデータ送信完了にかかる時間の増大、といった弊害を引き起こす。

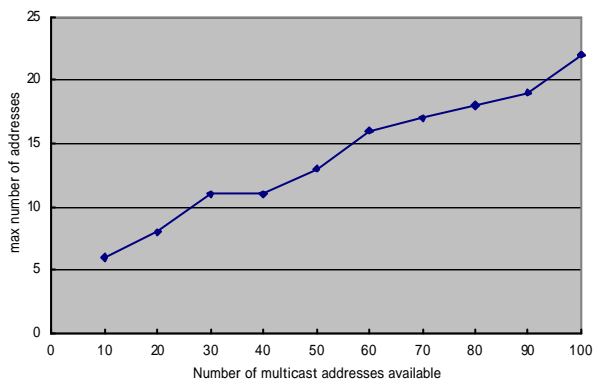


図 1 GR における使用マルチキャストアドレス数と最大送信先アドレス数

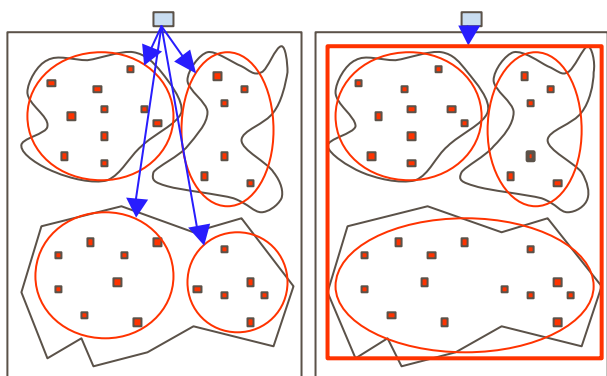


図 2 GR の問題点

図 3 提案手法

3. メンバークラスタリング手法

3.1 概要

2.2 で挙げた問題点を解消するために送信者の負荷を抑えるメンバークラスタリング手法を提案する。まず、送信先アドレス数に対する閾値を設定しておく。そして GR を用いて受信者を使用出来るマルチキャストアドレス数分のクラスタに分類し、それぞれのクラスタに対してマルチキャストアドレスを割り当てる。各受信者は自分の属すクラスタに割り当てられたマルチキャストアドレス宛に join メッセージを送信し、送信者は自分のデータを要求している受信者の属すクラスタに割り当てられたマルチキャストアドレス宛にデータを送信する。ここで送信先アドレス数が閾値を超えている送信者が存在する場合、送信者を閾値を超えていないものと超えているものに分ける。つまり、閾値を超えていない送信者群と全ての受信者、閾値を越えている送信者群と全ての受信者、という2種類の送受信者の組み合わせが可能になる。このそれぞれの組み合わせに対して、再度 GR によるクラスタリングを行なう。この際のマルチキャストアドレス数は、全体で使用出来るマルチキャストアドレスから各組み合わせの送信者数の割合ずつ割り当てられる。この手法を繰り返し行なうことにより、送信先アドレス数が閾値を超える送信者の送信先アドレス数は低減される。

図 2, 図 3 は、1 つの送信者に注目して、送信先アドレス数の低減を示したものである。枠内の四角を受信者、

枠外の四角を送信者とし、使用可能マルチキャストアドレス数を4、閾値を3とする。まず、GR を用いて受信者を4つのクラスタに分類する(受信者を囲む楕円がクラスタを示す)。この際、図 2 に示す送信者のデータを全受信者が要求する場合、その送信者は全てのクラスタ(ここでは4つ)に対してデータを送信しなければならない。閾値は3であるから、その送信者の送信先アドレス数は閾値を超えることになる。ここで、送信先アドレス数が閾値を超える送信者は図に示す送信者のみとすると、その送信者に対して1つのマルチキャストアドレスが、その他に3つのマルチキャストアドレスが割り当てられることになる。これにより、送信先アドレス数が閾値を超える送信者は1つのマルチキャストアドレス宛にデータを送信し、その送信者のデータを要求する受信者(ここでは全受信者)は、その送信者に割り当てられたマルチキャストアドレス宛に join メッセージを送信し、データを受け取る。これにより、その送信者の送信先アドレス数は4つから1つに低減することになる。

また、本手法を実際のアプリケーションで適用するには、クラスタリングを実行する者が必要である。それには2つのパターンが考えられる。

(1) サーバまたは代表者が実行する

アプリケーションを管理しているサーバまたは代表者に対して各参加者が自分の要求データリストを送信し、それを用いてサーバまたは代表者がクラスタリングを行なう。そして結果(join メッセージ及びデータを送るマルチキャストアドレス)を各参加者に送信し、各参加者は結果に従って join し、自分のデータを送信する。

(2) 全ての参加者が自立分散的に実行する

各参加者が全ての参加者宛に自分の要求データリストを送信し、それを用いて各参加者はクラスタリングを行ない、その結果を用いて join メッセージ及びデータの送信を行なう

3.2 アルゴリズム

本手法のアルゴリズムを表 1 に示す。以下、表 1 について説明する。ここで、 K は使用できるマルチキャストアドレス数、 P は要求データ行列(行が受信者、列が送信者を表す)を表す。 P の行ベクトル P_{r_m} は受信者 r_m の送信者に対する要求ベクトルを表し、列ベクトル P_{s_n} は送信者 s_n に対する受信者の要求を表す。また、 G_k はクラスタを表す。

(1) 初期クラスタ中心には受信者をランダムに選ぶ。ただし、送信者のデータを全く必要としない受信者はクラスタリングする必要がないため、クラスタ中心とはならない。また、要求データの一致する受信者が複数いる場合、その中からは1つしかクラスタ中心とならない。

(2) 各受信者は全てのクラスタ中心に対して距離関数 $d(r_m, r_n)$ による比較を行ない、 d の値が最も小さくなるクラスタに属す。ここで距離関数 d とは各受信者の要求データベクトルのユークリッド距離であり、

表 1 メンバークラスタリング手法のアルゴリズム

(1)	初期クラスタ中心を決める
(2)	受信者をクラスタリング
(3)	新たにクラスタ中心を決める
(4)	新たなクラスタ中心が前と違うなら(2)に戻る
(5)	送信先アドレス数が閾値を超えていなければ終了
(6)	送信者を送信先アドレス数が閾値を超えていないものと超えているものに分ける
閾値を超えていない送信者群について	
(7)	(1)～(5)を適用する
(8)	送信先アドレス数が閾値を超えている送信者がいる場合、閾値を超えている送信者群に加える
閾値を超えている送信者群について	
(9)	(1)へ戻る

$$d(r_m, r_n) = \left| P_{r_m} - P_{r_n} \right|$$

で表される。

(3) 各クラスタにおいてクラスタ重心を求める。クラスタ重心とは、クラスタに属している受信者の要求ベクトルの平均を各要素とするベクトル P_{G_k} で、その各要素は

$$P_{G_k, s_n} = \frac{\sum_{r_m \in G_k} P_{r_m, s_n}}{|G_k|}$$

で表される。そして P_{G_k} と P_{r_m} ($r_m \in G_k$) との間で距離関数 d による比較を行ない、 d の値が最も小さくなるものを新たなクラスタ中心とする。

(4) 新たなクラスタ中心が以前のクラスタ中心と異なるならば(2)に戻り再度クラスタリングを行なう。

(5) 全ての送信者に対して送信先アドレス数を参照し、値が閾値を超えているものがないければ、送信者の負荷をそれ以上減少させる必要がないものとして終了する。また、値が閾値を超えているものがあるならばその送信者の負荷を減少させるように再度クラスタリングする必要がある。

(6) 送信者を閾値を越えていないもの(Aグループとする)と超えているもの(Bグループとする)に分ける。ここで、Aグループ、Bグループの要求データ行列をそれぞれ

$$P_A = P \Big|_{s_n \in A}$$

$$P_B = P \Big|_{s_n \in B}$$

とする。ここで、 $P \Big|_{s_n \in A}$ とは P から A 以外の s_n の列を掃き出した行列を表す。また、クラスタリングに使用するアドレス数は、使用出来るマルチキャストアドレス数のうち、 A 、 B それぞれのグループに属す送信者の人数の割合分ずつ割り当てられ、

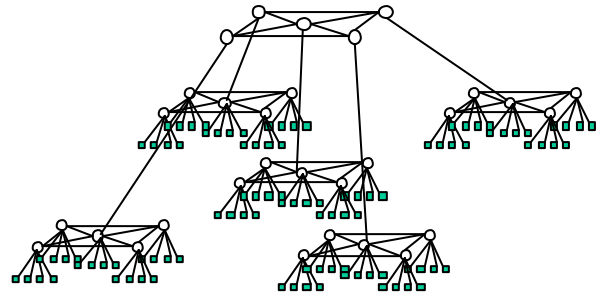


図 4 シミュレーショントポロジー

$$K_A = \begin{cases} \left\lfloor \frac{K|A|}{|A| + |B|} \right\rfloor & (\text{if } K_A \geq 1) \\ 1 & (\text{if } K_A < 1) \end{cases}$$

$$K_B = K - K_A$$

となる。

(7) Aグループに対して(1)～(5)を適用してクラスタリングする。ただし、要求データ行列は P_A 、使用マルチキャストアドレス数は K_A とする。

(8) Aグループをクラスタリングする際に送信先アドレス数が閾値を超える送信者が出てくる可能性があるため、そのような送信者いる場合は、その送信者も B グループに加える。

(9) Bグループに対しても要求データ行列は P_B 、使用マルチキャストアドレス数は K_B として、(1)に戻りクラスタリングする。

4. シミュレーションによる評価

4.1 シミュレーションモデル

本稿では、前述したように、多対多参加型アプリケーションの環境例として、ネットワークゲームを想定している。そこで、本章でもネットワークゲームを想定したモデルによるシミュレーションにより、本手法の評価・検討を行なう。ネットワークゲームを想定したモデルというのは、すなわち、参加者が仮想世界内で自由に存在し、参加者は送信者にも受信者にも成り得るようなモデルである。今回のシミュレーションでは、参加者数は 100 人とし、仮想世界全体を 32×32 の 1024 個のエリアに分け、参加者はそのセルのいずれかに属す。仮想世界における受信者の位置は全くのランダムとし、受信者の要求データとしては仮想世界内で自分の周り 10% 内にいる参加者のデータを要求する。

4.1.1 シミュレーショントポロジー

今回のシミュレーションにおけるネットワークトポロジーは、図 4 のようになっている。ここで、 \circ はルータ、 \square はユーザを示す。ネットワークトポロジーは 2 階層に分かれた形を取っており、下位層のルータそれぞれに、ユーザが 4 人ずつ接続する形をとっている。

表 2 使用マルチキャストアドレス数と閾値

アドレス数	10	20	30	40	50
送信先アドレス数	4	6	7	8	10
アドレス数	60	70	80	90	100
送信先アドレス数	11	12	14	15	16

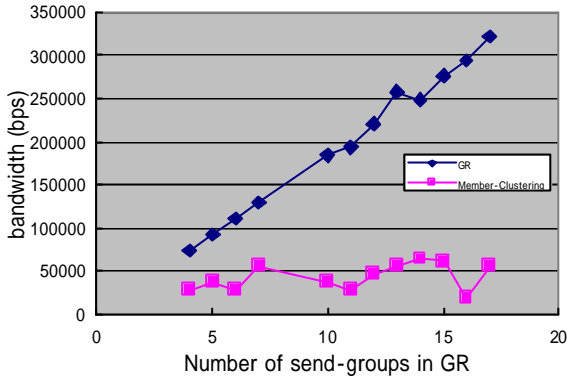


図 5 送信先アドレス数の多い送信者付近の使用帯域

シミュレーションパラメータについては、リンク帯域と伝播遅延はそれぞれ、ユーザから 1 ホップのルータまでは 640kbps、遅延は 1msec、下位層における帯域は 10Mbps、遅延は 1msec、下位層から上位層へのリンク帯域は 100Mbps、遅延は 5msec、上位層における帯域は 1Gbps、遅延は 10msec とし、各送信者の送信レートは 18.4kbps とする。ただし、パケットロスには起きないものとする。また、ルーティングプロトコルとして PIM-SM を用いた。

4.1.2 閾値

本手法の適用の際には、閾値を決定しなければならない。本来、送信先アドレス数の減少が必要な送信者は、送信処理負荷または近接リンクのネットワーク負荷の高い送信者である。つまり、処理能力が高く、リンク帯域も広い送信者にとっては、送信先アドレス数が多くなって問題のない場合も考えられる。しかし、送信者の処理負荷及びリンク帯域やその状況を知ることは困難である。よって、本稿では、数回のシミュレーションにより統計的に送信先アドレス数が閾値を超える送信者が全体の 5 ~ 15%程度になるように設定した。閾値の値を表 2 に示す。

4.2 シミュレーション結果

本手法の適用により GR と比較して送信者の負荷は大幅に低減される。図 5 は、GR において閾値を超える送信者の送信先アドレス数とその送信者の送信帯域を表したものである。GR において送信先アドレス数の非常に多い送信者付近の使用帯域は送信先アドレス数にほぼ比例して増加している。このため、当該送信者及び、その付近のネットワークには過大な負荷がかかる。これ

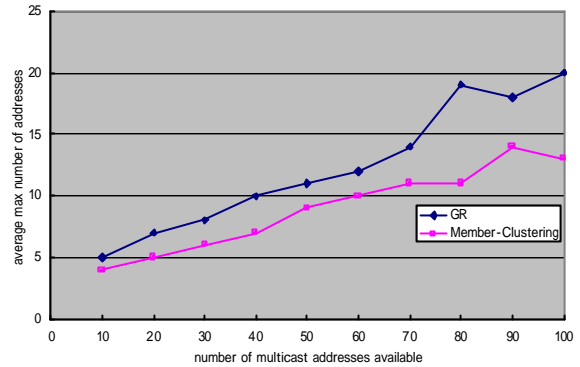


図 6 使用マルチキャストアドレス数と最大送信先アドレス数の関係

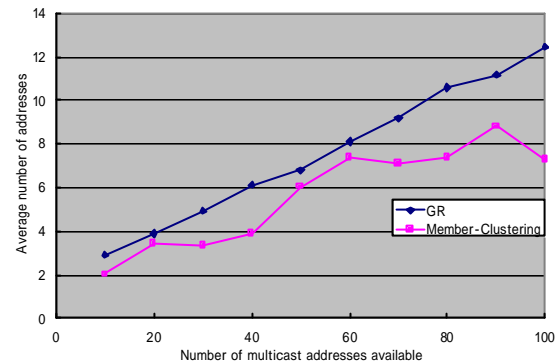


図 7 使用マルチキャストアドレス数と平均送信先アドレス数の関係

に対し、本手法を適用した場合、送信先アドレス数の多い送信者に対して送信先アドレス数を減少させるように再度クラスタリングされるため、使用帯域が大幅に低減されることになる。

また、図 6 は、送信者の最大送信先アドレス数と使用マルチキャストアドレス数の関係を表したものである。本手法の適用により、送信者の最大送信先アドレス数が GR と比較して減少していることが分かる。この傾向は使用マルチキャストアドレス数の増加とともに大きくなっている。また、これに伴い、図 7 のように平均送信先アドレス数も減少している。このことは、アプリケーションによってネットワークに送出される総データ量が減少していることも意味する。

さらに、データの到着時間の差が大きくなる、という問題点も解消されている。図 8 は、GR と本手法におけるデータの到着時間の差を示したものである。使用マルチキャストアドレスにかかわらず到着時間の差が減少している。特に使用マルチキャストアドレスが多い場合に効果が高い。これは、使用マルチキャストアドレスが多い場合には、送信先アドレス数も多くなるためだと考えられる。

また、図 9 に使用可能マルチキャストアドレス数と goodput の関係を示す。本手法における goodput は GR と比較して数% ~ 10%程度減少している。本手法では、送信者の負荷を減少させるために送信者を送信先アドレス数が閾値を超えるものと超えないものに分ける。

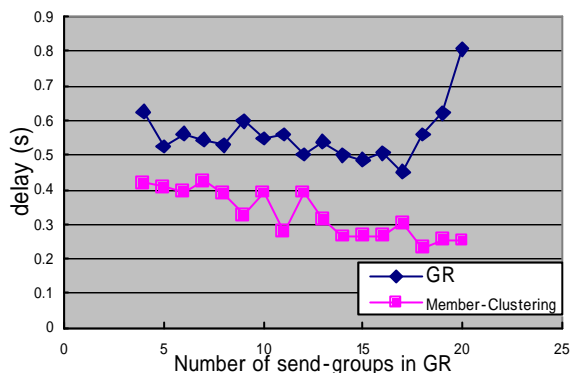


図 8 到着時間の差

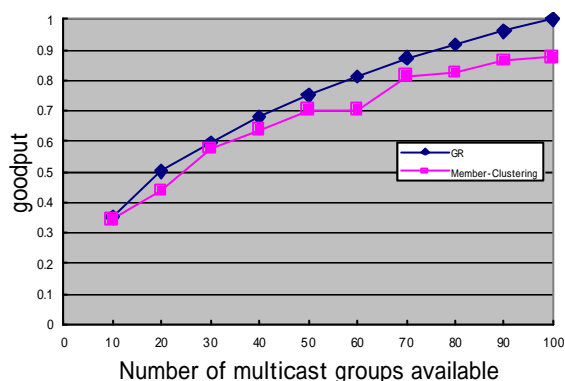


図 9 使用マルチキャストアドレス数と goodput の関係

そして、その人数に応じてマルチキャストアドレスも割り当て、それぞれのグループに対して再度 GR によるクラスタリングを行なう。その際、各グループにおいて、送信者数は減るものの、それぞれのグループに割り当てられたマルチキャストアドレス数も減少する。しかし、クラスタリングに使用する受信者数は変わらない。GR では、受信者をクラスタリングするため、受信者数が変わらない場合、送信者数が減ることよりも、マルチキャストアドレス数が減ることの方が影響が大きく、goodput の低下を招いたものと考えられる。

例えば、100 人の参加者で 50 個のマルチキャストアドレスが与えられた場合を考える。全体の 10% に当たる 10 人の送信先アドレス数が閾値を超えたとすると、送信者は 90 人のグループ(A グループ)と 10 人のグループ(B グループ)に分けられる。その際、マルチキャストアドレスは A グループに全体の 90% に当たる 45 個、B グループには全体の 10% に当たる 5 個が割り当てられる。この後、本手法では、各グループに対して GR を適用するが、グループ A は送信者 90 人、受信者 100 人、マルチキャストアドレス数 45 という条件で、グループ B は送信者 10 人、受信者 100 人、マルチキャストアドレス数 5 という条件でクラスタリングを行なうこととなる。このように、本手法を適用した場合、A、B どちらのグループに対しても受信者数は変化せずに用マルチキャストアドレスが減少しているため、goodput の減少を招く。

ここで分かるように、送信者の負荷と、goodput はトレードオフの関係にある。しかし、送信者の過大な負荷によりデータが到着しない場合を考えると、多少の不要なデータの受信があるとしても送信者の負荷を減少させることは重要である。また、本手法ではデータの遅延を減少させる、ネットワークに送出されるデータ量を減少するなどの利点も存在する。これらの結果により、本手法は多対多ユーザ参加型のアプリケーションにおいて有効であるといえる。

5. 結論

本稿では、要求データが受信者毎に異なるような多対多参加型アプリケーションにおいて、マルチキャスト適用手法の 1 つである GR における問題点を指摘した。その解決手法として、送信先アドレス数の多い送信者がいる場合、送信者を 2 つのグループに分け、それぞれのグループで GR を行なうことにより、送信者の送信先アドレス数を抑えるメンバークラスタリング手法を示した。

参考文献

- [1] S. McCanne, et al, "Toward a Common Infrastructure for Multimedia Networking Middleware", In Proceeding of NOSSDAV'97, May 1997.
- [2] C. Diot and L. Gautier, "A distributed architecture for multiplayer interactive applications on the internet", IEEE Network magazine, July/August 1999.
- [3] M. Pullen, M. Myjak, and C. Bouwens, "Limitations of internet protocol suite for distributed simulation in the large multicast environment", RFC 2502, 1999.
- [4] T. Wong, R. Katz and S. McCanne, "An Evaluation of Preference Clustering in Large-Scale Multicast Applications", in Proceedings of IEEE INFOCOM 2000, March 2000.
- [5] M. Oliveira, J. Crowcroft and C. Diot, "Router Level Filtering for Receiver Interest Delivery", proceeding of NGC 2000. November 2000.
- [6] B. Levine, and J. Garcia-Luna-Aceves, "Internet Multicast Based on Group-Relative Addressing", Tech. rep., UCSC, 1999.
- [7] B. N. Levine, J. Crowcroft, C. Diot, J.J. Garcia. L. Aceves and J. F. Kurose, "Consideration of Receiver Interest for IP Multicast Delivery", in Proceedings of Infocom 2000. March 2000.
- [8] T. Wong, R. Katz, and S. McCanne, "A Preference Clustering Protocol for Large-Scale Multicast Applications", In Proceedings of NGC1999, November 1999.
- [9] C. Diot, et al, "Deployment Issues for the IP Multicast Service and Architecture", IEEE Network Magazine, January/February 1999.
- [10] J. HARTIGAN, "Clustering Algorithms", John Wiley and Sons, 1975.