マルチメディア通信と分散処理　108－10
（２００２．６．６）

高信頼マルチキャストにおける再送木構築方式の改良

森田 悟史
静岡大学大学院
情報学研究科

古川 雄宣
静岡大学
情報学部

山田 和宏
静岡大学
情報学部

佐藤 文明
静岡大学情報学部

分散アプリケーションの開発において，高信頼マルチキャストは重要な基板技術である．様々なアプリケーションや規模に応じて既に多くの高信頼マルチキャストが開発されている．高信頼マルチキャストにおける重要な課題は，効率的なメッセージの再送処理である．本稿では木構造を再送に利用する高信頼マルチキャスト方式に注目する．この方式では，再送の負荷を分散させるため均等かつ，小さくまとまった木を構築する事を目標としており，この方式は良好な結果が得られている．しかし，木構造のあるメンバー間の接続が突然切断される場合については考慮されていない．そこで，このような場合にも対応できるようにバックアップリンクを保持する方式を提案する．この方式でシミュレーションを行った結果，従来の方式に比べ，メンバ間の接続が突然切断される場合において有効であることがわかった．

A Tree Configuration Algorithm for Reliable Multicast Protocols

Satoshi Morita　Takanori Hurukawa　Kazuhiro Yamada　Fumiaki Sato
Graduate School of Information, Shizuoka University
Faculty of Information Shizuoka University

In development of distributed applications, Reliable Multicast Protocols are the important infrastructure. For different applications and scales of the systems, many Reliable Multicast Protocols have been developed. The main theme of the Reliable Multicast is the efficient retransmission mechanism. In this paper, we have developed a new tree configuration algorithm for the Reliable Multicast, which has tree based retransmission mechanisms. This algorithm aims to configure a balanced and small tree, for distributed retransmission, and it got a good result. But, it has not considered for suddenly disconnect between members. So, we propose Backup-Link algorithm for this situation. Simulation of Backup-Link algorithm shows the effectiveness for that situation.

1　はじめに

近年，コンピュータネットワークの高速化により分散アプリケーションの使用が一般的になってきた．分散アプリケーションの例としては，共有仮想環境，電子会議システム [1] などがある．分散アプリケーションの設計において，高信頼マルチキャストは効果的かつ必要不可欠なネットワーク基盤となっている．また，参加しているユーザが増加しても少人数の場合と同様のサービスを提供できる性能が求められる．

高信頼マルチキャストについてはこれまで多くの研究が世界中の研究機関においてなされており，現在様々な全順序マルチキャスト方式が提案されている．高信頼マルチキャスト方式においてメッセージの欠落に対応する再送処理はシステム全体の性能において少なからず影響を与える．そのため，システム全体の性能を落さない効率的なメッセージの再送処理を行う工夫が必要である．

本研究では，メンバーが再送木を構築して，その構造に従って再送処理を行う方式に注目する．この方式では，送信者に再送要求が集中するのを避けらるのでスケーラビリティが高い．しかし木構造のメンバー管理など，複雑な処理が必要となる．

マルチキャストグループに参加するメンバーは動的に変化する木構造を構築し，それに従って再送を行い，メッセージを補完する方式が提案されてきた．この方式では，各メンバーは構築する木の規模を小さくし，再送の負荷を分散するためのデータを保持する．この方式は，良好な評価が得られているが，いくつかの課題が残っている．今回，それらの問題点に対応させる改良を加えた．メンバー間のリンクが突然切断される場合に対応するためのバックアップ方式，データの更新における木全体のタイムラグの考慮，またデータの更新頻度，参加要求パケットの送信範囲の限定という改良を加えシミュレーションで評価した．その結果，バックアップ方式を加えた場合はメンバ間のリンクが突然切断される場合においてもリンクは保たれ，データの更新頻度，参加要求パケットの送信範囲の限定においてはネットワーク資源をより消費しない再送木構築が行われていることがわかった．

2　高信頼マルチキャストと再送木

本章では高信頼マルチキャストの概要と従来の再送木構築方式について述べる．

2.1 高信頼マルチキャスト

マルチキャストされたパケットは欠落する可能性があり，すべてのマルチキャストメンバグループメンバーが受信する保証はされない．高信頼マルチキャストではすべてのメンバーがマルチキャストパケットを受信する必要がある．そのため，高信頼マルチキャストでは欠落したパケットの再送処理が問題となる．

再送処理には送信者が再送を行う場合と送信者以外が再送を行う場合がある．送信者が再送を行う場合，シンプルであるが送信者に再送処理のための負荷が集中する．受信者からの Nack のタイミングをずらすなどの改善策も存在する．

送信者以外が再送を行う場合では，雲形や木形，リング形に受信者を組織化する方法があり，それぞれ特徴は変わってくる．雲形は Nack が集中する事がなく再送の負荷が分散されスケーラビリティに優れているが，再送の最終責任者が特定されず，メンバー情報を管理しようとすると複雑になるといった欠点がある．木形では木に沿って Nack を送信し再送を要求する．受信者は送信元に直接コンタクトする事は無く，大規模な受信者グループにも対応できる．リング形ではトークンを回しデータ伝送および応答を同期化し，高パフォーマンスという利点がある．しかし，アルゴリズムの複雑さとプロトコルステータスの多さのため実装は困難である．また，各メンバーがメンバーシップ情報を維持しなければならずスケーラビリティは比較的限定される [2][3]．

以上のことをふまえ，信頼性が高くスケーラビリティがある木に基づく再送処理に着目して研究を行ってきた [4]．

2.2 従来の再送方式

ここでは従来の再送木構築方式について述べる．この方式でシミュレーションを行って，改良方式との比較を行う．

メンバーは，小さく均等な木構造を構築することを目的とする．木が小さくまとまてっいると再送要求である Nack がすべてのメンバーに短時間で行き渡り，再送が効率よく行われる．均等な木構造を作ることにより，ひとつのメンバーに処理が集中することを避ける．各メンバーの処理が分散されることでシステム全体の性能に影響を与えることを防ぎ，メンバー数が増大してもメンバー数が少数の時とほぼ同等のサービスを提供することができる．

木を構築するために必要な情報は各メンバーに分散させて管理するので，各自が保有するデータの独立度が高い．そのため，木に新しくメンバーが参加したり，途中でメンバーが脱退したとしても，それらの処理がアプリケーション全体に大きな影響を与えることはない．よって，木構造の操作が容易となる．

2.2.1 各メンバーが保持するデータ

システムに参加しているメンバーは，そのメンバーに何人のメンバーがリンクを張っているかを示す重みという値を持つ．また，メンバーに隣接するメンバー毎の通信遅延をノード間の距離値として保存する．それらの値とは別に，各ノードから見た各リンク方向の木の大きさに対応して与えられるノード深度という値がある．

ノード深度とは，そのノードから先の最も遠い末端のメンバーまでの距離を表す．メンバーはノードから先の木構造を完全に知ることは出来ないが，ノード深度によってどれだけの広がりを持っているのかおおよそ知ることが出来る．木構造では，1 人のメンバーからのびるリンクが複数存在する場合がある．それらのノードが持つノード深度の中で最も大きい値を最大ノード深度と呼ぶ．図 1 では 4 人のメンバー S，T，U，V がいる．それぞれのノードの距離値はΔ l，Δ m，Δ n(Δ l >Δ m >Δ n) である．U に繋がる 2 本のリンク方向のノード深度はΔ l +Δ m とΔ n であり，最大ノード深度はΔ l +Δ m となる (図 1)．

これらの値は参加，脱退処理で利用される．

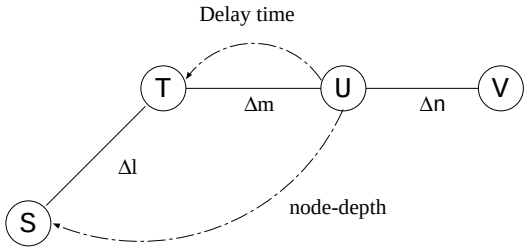

図 1: ノード進度

2.2.2 木構造への参加処理

新規に参加するメンバーはシステムがグループ通信に利用しているマルチキャストアドレスを知っており，メンバー全員に参加要求をマルチキャストする．この参加要求を受信したメンバーは，自身が持っている重みと最大ノード深度の情報を返す．また，新規メンバーは要求を送信した時刻を記録し，各メンバーからの応答が返ってくるまでの時間を計測し，これを通信遅延として保存する．新規メンバーは重み，最大ノード深度，通信遅延の合計が最も小さくなるメンバーを自分

の隣接ノードとして選択する．そして，再送管理の関係を結ぶためのノード接続要求メッセージを選択したメンバーに送信する (図 2).

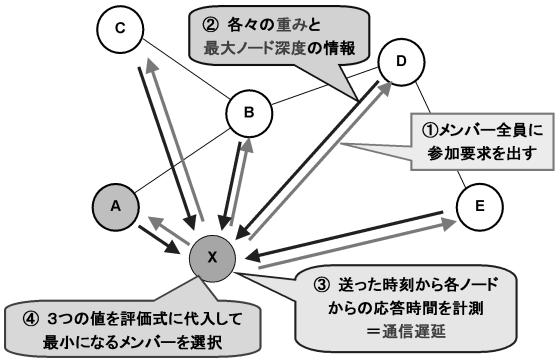

図 2: 参加処理

2.2.3　木構造からの脱退処理

メンバーが木構造から脱退する場合，木構造が切断されてしまうことがあるので，同時に木構造を再構築をしなくてはならない．また，新たに構築した再送管理関係で再送するべきメッセージの同期を取る必要がある．

それらの処理について説明する．木構造から脱退するメンバーは，自分にリンクを張っているメンバーの中から，重みと通信遅延の合計が最小の値を示すメンバーを選択して，脱退後の再接続先メンバーとして選択する．

次に脱退メッセージを隣接しているノードに対して送信する．脱退メッセージには先ほど選択したメンバーのアドレスが付加されている．脱退メッセージを受け取ったメンバーは，付加されてきたアドレスと自身のアドレスと比較する．自身と同じ場合，そのメンバーは待機する．異なるならば，そのアドレスを持つメンバーが新しい隣接メンバーになる．新しい隣接メンバーに再参加処理メッセージを送信し，再送されるべきメッセージの同期をとって再送管理の関係を新しい隣接メンバーに切り替える．再参加処理が終わったメンバーは，脱退するメンバーに対して脱退許可メッセージを送信する．脱退するメンバーに隣接していたメンバー全員から脱退許可メッセージを受け取った時点で，木構造からの脱退処理が終了する．

2.2.4　ノード深度の変更処理

以上のようにして参加，脱退をする場合，木の形が変わることがある．それに伴いノード深度の変更が加わる場合がある．新規参加処理の場合，新規メンバーまでの距離値Δ t がそのままノード深度Δ Q になる．再参加処理の場合，新規隣接メンバーから送られてきた最大ノード深度Δ Q と距離値Δ t を加えた値がノード深度Δ R(Δ R = Δ Q + Δ t) になる．新しくできたノードのノード深度の大きさによって，既存の隣接メンバーの中にノード深度を変更する必要があるメンバーが現れることがある．その場合，ノード深度の変更が必要と思われるメンバーに対して新しくできたノードのノード深度Δ R を，ノード深度変更メッセージとして送信する．ノード深度変更メッセージを受け取ったメンバーは，送信元のメンバーへの距離値Δ s と送られてきたノード深度Δ R を加算し，新たなノード深度Δ S = Δ R + Δ s として設定する．この処理を繰り返すことで，変更の必要があるメンバーの持つノード深度を変更する．

2.2.5　従来方式の課題

メンバ間の接続が突然切断された場合，後から木に参加したメンバーが再び参加要求を送信し再接続を行う．この時，リンクが確立されるまでは木が保たれていないため，再送が行われない場合が存在する．

また，従来方式はノード深度を更新する際の遅延を考慮せずに評価されていた．実際にはノード深度を更新するには新たにメンバーが参加･脱退することによってデータが変更された場所と，そこから末端の場所にはタイムラグが存在する．データを更新する際の遅延を考慮して実際と同等の環境で評価するべきである．

従来方式では参加処理の際に新規メンバーは，全てのメンバーに参加要求を送信している．参加要求を受信した全てのメンバーは応答しなければならない．しかし，新規メンバーから通信遅延の大きいメンバーと再送関係を築くことは非常に少ないため，数が増加した場合は大半のパケットが無駄となる．よって再送要求はグループ全体にマルチキャストせずに，新規メンバーの近隣のメンバーのみに送信するほうが大変効率がよいといえる．

3　提案方式

前述した従来の方式では，メンバ間における突然の切断による再送木の分断，データ更新の通信遅延，参加要求パケットの通信範囲という問題が残っていた．

これらを解決するため，従来の方式を改良する．

3.1 バックアップリンク

メンバ間のリンクが突然切断される場合，再送木が分断され，再送が正しく行われないことがあり，大変致命的である．そこで本来のリンクとは別にバックアップリンクを保持する方式を提案する (図 3)．リンクが突然切断される場合には，即座にバックアップリンクを利用し，再送木が分断されるのを防ぐ．

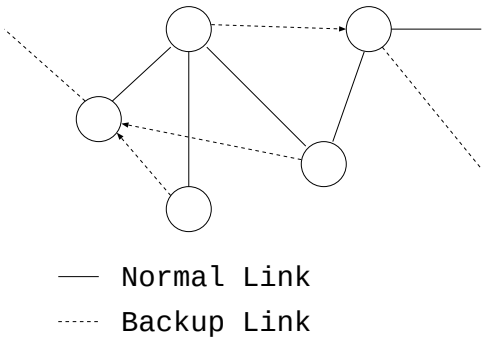

図 3: バックアップリンク

バックアップリンク先は通常のリンク先と同様の方法で選択される．つまり，新規メンバーは再送木における最適なメンバーと通常のリンクを保持し，次に最適なメンバーとバックアップリンクを保持する．

バックアップリンク方式では，バックアップリンク先も通常のリンク先と同様にパケットの再送管理を行が，実際に再送は行わない．つまり，受信したパケットの Ack はバックアップリンク先と通常のリンク先に送信するが，再送要求は通常のリンク先にだけ送信する．従来の方式に比べ，再送管理コストは多いが，突然のリンク切断が存在しても再送の信頼性は非常に高い．突然のリンク切断が一定の頻度で発生しても，従来方式に比べ再送木がどの程度稼動し続けるかシミュレーションで評価する．

3.2 参加要求の範囲設定

従来方式では参加処理を行う際に，再送木を構築している全てのメンバーに参加要求を送信する．しかし，メンバー数が増加するにしたがって参加要求に応答するパケットの数も増大し，ネットワーク全体に多大な負荷を与える．よって，新規参加者は一定の範囲にのみ参加要求を送信する．実際には参加要求パケットの TTL を制限する，または参加要求パケットに送信時間を付与し一定時間を過ぎて受信した場合は応答しないという方法が考えられる．

参加要求の範囲を制限した場合，従来に比べて再送木がどの程度劣化するかシミュレーションで評価する．

3.3 データ更新における通信遅延の考慮

従来方式では考慮されていないデータ更新における通信遅延を考慮して評価する．従来方式では，メンバーが参加・脱退を行い，木の構造が変化し，データ更新が行われると，その時，全てのメンバーでデータが更新された．実際には，メンバーが参加・脱退する場所とそこから木の末端まででは，ある程度の通信遅延が考えられる．つまり，データ更新が行われずまだ古い情報を持っているメンバーと新しい情報のメンバーが混在し，これは再送木の構築に影響を与える (図 4)．

通信遅延を考慮した場合，再送木がどの程度劣化するかシミュレーションで評価する．

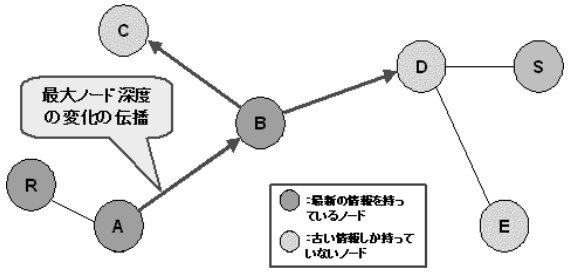

図 4: データ更新の遅延

3.4 データ更新の頻度調整

データの更新は，ノード深度が変更される時，つまり木全体においてメンバーが参加･脱退するたびに行われる．メンバー数が増大した場合，データ更新におけるネットワークの負荷はシステム全体に少なからず影響を与える．そこで木構造が一回変化する毎に更新するのではなく，一定回数ごとにデータ更新を行う．しかし，データ更新を控えることにより木構造のパラメータが最新の情報にはならず，再送木が最適に構築されるとは限らない．

データ更新頻度の削減がどのような影響を再送木に与えるかシミュレーションを行い従来方式と比較する．

4 シミュレーションの結果と評価

バックアップリンク方式とデータ更新における通信遅延を考慮した方式とでシミュレーションを行い従来方式と比較した．

4.1 バックアップリンク方式

100 のメンバーを配置した 1000*1000 の座標を使い再送木を構築させ，突然のリンク切断として，あるノード間の接続を切断させるイベントを一定の間隔で発生させた．シミュレーションは 100000 秒行い，木の稼動時間と平均接続時間を，従来方式とバックアップリンク方式とで比較した (図 5)．

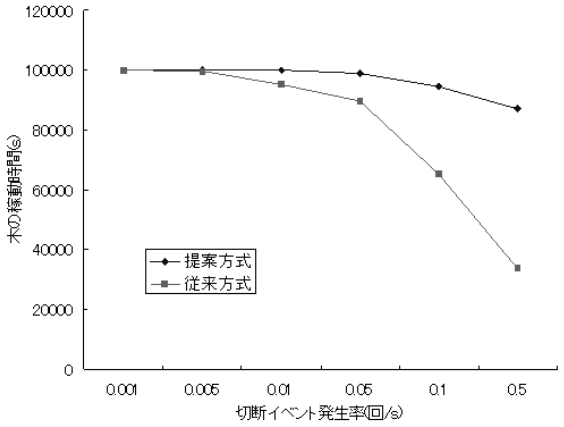

図 5: 木の稼動時間の比較

以上の結果からリンク間の突然の切断が多いほどバックアップリンク方式と従来の方式の差が出ることがわかる．

4.2 データ更新における通信遅延の考慮

100 のメンバーを配置した 1000*1000 の座標を使い再送木を構築させ，遅延を考慮した場合としていない場合でメンバーの参加頻度を変化させてシミュレーションを行った．遅延は距離 100 を平均 0.1 秒とする (図 6)．

ネットワークの遅延を考慮した結果，参加頻度が高くなるにつれ，古い情報に基づく木の構築が行われるため従来の評価に比べて劣化した木が構築されることがわかる．

4.2.1 参加要求の送信範囲の限定

50 のメンバーを配置した 1000 * 1000 の座標を使い再送木を構築させ，遅延を考慮した場合と，それに加え参加要求の送信範囲を設定した場合とでシミュレーションを行った．送信範囲を距離 500，300，100 以内と設定しない場合の 4 つで参加頻度を変化させた (図 7). メンバー数を 50 としたのは，木の構築具合が遅延を考慮しない場合とあまり変わらないからである．

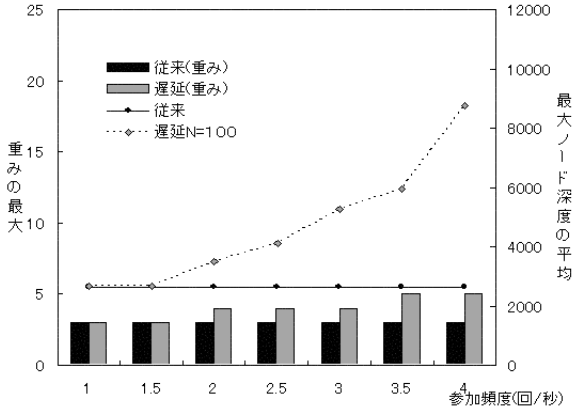

図 6: 参加頻度に対する最大ノード深度と重みの変化

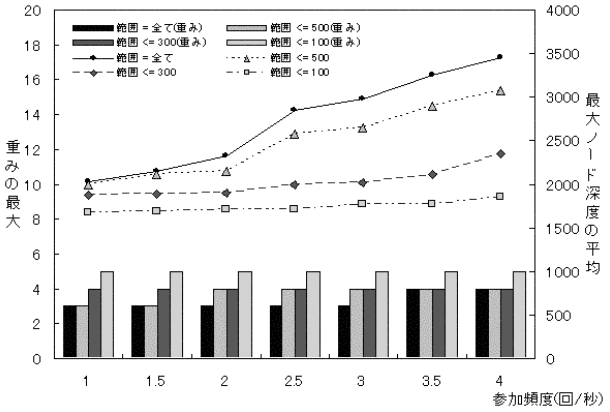

図 7: 参加要求の送信範囲を設定した場合との比較

送信範囲がもっとも狭い 100 以内では重みが全体的に多いが，木全体は大きく広がらないことがわかる．近くのメンバーにしか接続しないことで木は小さくなるがメンバー一人一人の負荷は増えるといえる．

4.2.2 データ更新頻度の設定

上記と同様の環境で，遅延を考慮した場合と，それに加えデータ更新頻度を設定した場合とでシミュレーションを行った．データ更新頻度は通常の場合は木が毎回変化するごとに更新しているが，設定した場合では 2 回に 1 回の頻度で更新している (図 8)．

更新頻度が変化しても重みの最大値はあまり変わらないことがわかる．また木全体のバランスは更新頻度を 2 回に 1 回としたほうが広がっていることがわかる．

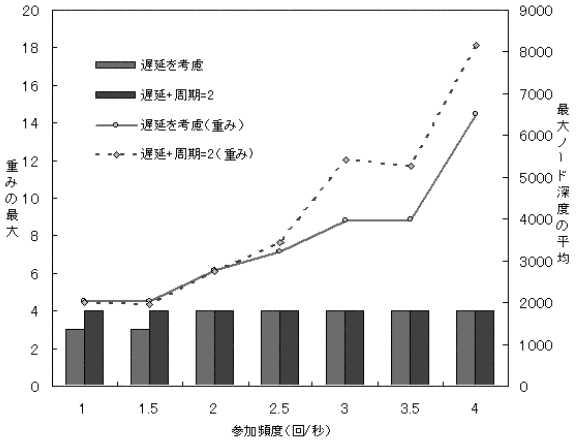

図 8: データ更新頻度を設定した場合との比較

5 結果の考察

従来の方式にバックアップリンク方式を追加した場合と，遅延を考慮した場合，遅延を考慮して参加要求の送信範囲を設定した場合，遅延を考慮してデータ更新頻度を設定した場合とで再送木の構築シミュレーションを行った．

バックアップリンク方式ではメンバー間のリンクが突然切断されることが多い環境では非常に有効であるといえる．しかし，通常の安定したネットワークでは，逆にバックアップリンクを保持するコストがかかり有効とはいえない．ネットワークの状態によってどちらの方式を使用するかを決定する必要がある．

遅延を考慮すると従来の評価より再送木は大きく広がることがわかった．メンバーの数が少ない場合はあまり変化は無いが，メンバーが増加するほどその差は顕著に現れることがわかった．

参加要求の送信範囲を設定した場合では，設定したほうが再送木が広がらずコンパクトに構築されることがわかった．メンバーの負荷は少ししか変わらず，ネットワーク資源の消費を軽減することからも参加要求の送信範囲を設定することは非常に有効であるといえる．

データ更新頻度を設定した場合では，参加頻度が大きくなると通常の場合と比べ再送木が広がることがわかった．参加頻度がそれほど大きくない場合では，どちらの場合でもあまり変わらないので，通信コストを軽減しているデータ更新頻度を設定する場合のほうが非常に有効であるといえる．

今後の課題としては，バックアップリンク方式では通常の方式と比べ，どの程度負荷がかかるか明確な値をだす必要がある．また，参加要求の範囲設定とデータ更新頻度の設定を組み合わせ，再送木のバランスを考慮しつつネットワーク資源の消費を抑えることができるか，検討する必要がある．また通常の方式における再送木を構築するための評価式を，遅延を考慮した場合にも最適な再送木を構築できるよう見直す必要がある．

6 おわりに

本稿では従来の再送木構築方式にバックアップリンク方式を追加した場合，通信遅延を考慮した場合，通信遅延を考慮し参加要求の送信範囲を設定した場合，通信遅延を考慮しデータ更新頻度を設定した場合とでシミュレーションを行い評価した．

バックアップリンク方式ではメンバー間の突然のリンク切断がよく発生するネットワークの場合に従来方式に比べ非常に安定した再送が行える事がわかった．通信遅延を考慮すると従来方式は再送木が広がることがわかった．参加要求の送信範囲を設定することで，再送木はよりコンパクトに構築され，かつ通信コストを軽減できることがわかった．データ更新頻度を設定した場合，参加頻度がそれほど多くない状況では通常の方式とあまり変わらずネットワーク資源の消費を軽減でき，非常に有効であることがわかった．

今後の課題としては参加要求の送信範囲設定とデータ更新頻度の設定を組み合わせ，最適な再送木を構築しつつどの程度まで通信コストを減少できるか検討する必要がある．

参考文献

[1] 山田善大, 池谷利明, 峰野博史, 太田賢, 水野忠則, “モーバイル電子会議システム PARCAE の提案”, 情報処理学会全国大会講演論文集 pp.3-543-3-544, (1997)

[2] J.Gemmell, “Scalable Reliable Multicast Using Erasure Correcting Resends”, Technical Report, MSR-TR-97-20, Microsoft Research, Redmond, WA, June 1997

[3] Brian Whetten,Talarian Corporaion, Gursel Taskale,Reuters, “An Overview of Reliable Multicast Transport Protocol II”, IEEE Network, January/February 2000

[4] 森田悟史，江崎美仁，佐藤文明，”高信頼マルチキャストにおける再送木構築方式の提案”，マルチメディア，分散，協調とモバイルシンポジウム論文集，June 2001