社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2004－DPS－117 (6)
2004－CSEC－ 24 (6)
2004／3／4

端末モビリティ対応の因果順序保存放送プロトコルのネットワーク階層化による拡張

松田 哲史
三菱電機株式会社 情報技術総合研究所
〒247-8501 神奈川県鎌倉市大船 5-1-1
E-mail: tmatsuda@isl.melco.co.jp

要約 端末の移動に対応した因果順序保存放送プロトコルに関する既存の提案方式では、因果順序を保存するためにメッセージに付与するベクトル時計のサイズが移動支援局数であるため、ネットワーク規模が大きくなり移動支援局数が増えた場合に、メッセージのオーバーヘッドが増加するという問題がある。本報告では、因果順序保存放送型通信を行うネットワークを、ツリー状に接続されるサブネットワークに分割し、既存の提案方式でのプロトコル処理に拡張を行うことで、ネットワークに存在する移動支援局数が増えても、メッセージに付与するベクトル時計のサイズを一定値以下に抑えることが可能となることを示す。

An Extension of causally ordered broadcast protocol supporting terminal mobility by way of subdividing a network hierarchically

Tetsushi Matsuda
Information Technology R&D Center, Mitsubishi Electric Corporation
5-1-1, Ofuna, Kamakura-shi, Kanagawa, 〒247-8501
E-mail: tmatsuda@isl.melco.co.jp

Abstract In a previous research on causally ordered broadcast protocol supporting terminal mobility, the size of vector clock attached to one broadcast message is equal to the number of Mobile Support Station (MSS) in the network. Hence, the more MSS is in the network, the larger is the overhead for each broadcast message. In this report, we propose to keep the overhead less than a constant value by subdividing the network into sub-networks which are connected in a tree topology and modifying the protocol described in the previous research.

1．はじめに

因果順序保存放送プロトコルは、分散処理アルゴリズムや、複数の人と人の間のインテラクションを実現する電子会議システム等の実現に有効であり複数の研究がなされている[3][4]。近年、ネットワーク環境のモバイル化が進んでいることに対応するために、端末モビリティをサポートするための拡張を施した因果順序保存放送プロトコルの提案も行われている[1][2]。いずれの方式も、因果順序保存を実現するために、ネットワーク規模に比例して増加するサイズのベクトル時計データを放送メッセージに付与する。このため、ネットワーク規模を拡大する場合に、放送メッセージに付与するベクトル時計データのサイズが大きくなり、ネットワーク帯域の利用効率が悪くなるという問題がある。

本論文では、ネットワークをツリー状に接続されるサブネットワークに分割して、ネットワーク規模の拡大にはサブネットワーク数を増やすことで対応し、因果順序を保存するために放送メッセージに付与するベクトル時計データのサイズを、サブネットワークの規模に比例する値に抑えることで、放送メッセージに付与するベクトル時計データのサイズをネットワーク規模に依らない一定値に抑えることを可能とするための、[1]の因果順序保存放送プロトコルに対する拡張を提案する。

2．関連する研究

[3],[4]は、端末の移動性を考慮しない場合の因果順序保存マルチキャスト通信プロトコルを提案している。[3]では、ネットワークを、何らかの手段(例えば[4]のプロトコル)で因果順序保存通信を行う複数のサブネットワークと、各サブネットワークを接続する1つのグローバルネットワークの2段の階層に分け、各サブネットワーク内とグローバルネットワーク内で送信するメッセージに、サブグループ数を次元とするベクトル時計データを付与する様にすることで、因果順序保存マルチキャスト通信のネットワーク規模に対するスケーラビリティ改善を行っている。サブネットワーク内で送信されるメッセージには、サブネットワーク内での因果順序保存通信を実現するために必要なベクトル時計データと、サブグループ数を次元とするベクトル時計データの両

方を付与する必要があるので、1 グローバルネットワーク＋サブネットワーク群を全体で 1 つのサブネットワークとみなして、グローバルネットワークを接続する更に上位のグローバルネットワークを設けることで多段の階層化を行った場合に、ネットワーク中のノード数に対する log オーダーではあるが、ネットワーク規模拡大に従ってメッセージに付与するオーバーヘッドが増加することとなる。

[1],[2]は、端末移動に対応した因果順序保存放送プロトコルを提案している。[1],[2]共に、ネットワークが移動端末(Mobile Host, 以下 MH と略す)と移動支援局(Mobile Support Station, 以下 MSS と略す)から構成され、MH は接続している MSS 経由で他の MH との間で因果順序保存放送通信を行うとしてモデル化している。ネットワークに存在する MSS と MH に増減はないと仮定する。MH と MSS 間は、メッセージ損失のない FIFO 型の通信が行われる動的チャネルで接続される。(図 1)

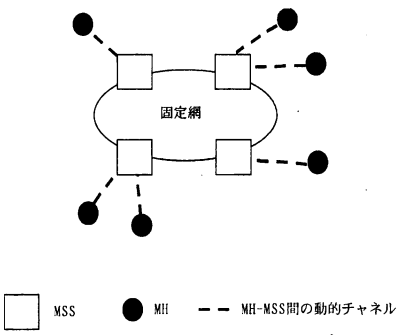

図 1 ネットワークのモデル

MSS は固定網で相互接続されており、MH から受信した放送メッセージの他 MSS への因果順序保存放送プロトコルによる送信と、他 MSS から受信した放送メッセージの因果順序保存放送プロトコルによる処理と MH への送信を行う。MH は、ある瞬間では最大 1 つの MSS と接続しており、移動により接続先の MSS を切り替えること(ハンドオフ)が可能である。

[1]では、ネットワークに存在する MSS 数 Nmss を次元とするベクトル時計データ 2 つを各放送メッセージに付与することで、因果順序保存放送プロトコルを実現しており、MH 数よりも MSS 数の方が少ないことから、MH 数が増加してもメッセージに付与するオーバーヘッドが増加しないこととなる。しかし、MSS 数が増加した場合にはメッセージに付与するオーバーヘッドが増加することとなるため、ネットワーク規模拡大に対して充分なスケーラビリティがあるとはいえない。

3．提案手法

3．1 提案手法の考え方

提案手法では、[1]のプロトコルに拡張を施す。提案手法の説明のために、[1]のプロトコルの基本アイディアの要旨をまとめると以下の様になる。以下で、複数ベクトルデータの最小値/最大値という表現を用いる場合は、ベクトルデータの各要素の最小値/最大値を要素とするベクトルデータを意味することとする。

- MH-MSS 間は FIFO 型通信チャネルで放送メッセージの転送を行い、MSS 間で因果順序保存放送プロトコルを実行する。MH では自分が受信済みの放送メッセージについて管理不要。
- MSS は、ネットワーク中に存在する MSS 数を次元とする、SENT と DELIV の 2 つのベクトル時計データを管理する。SENT は放送メッセージの前後関係の管理に使用し、送信する放送メッセージに付与する。DELIV は、各 MSS から受信した配達済みの放送メッセージの数を記録する。
- MH のハンドオフに対応するため、MSS は自局に接続している各 MH に対して、各 MSS から受信した配達済みの放送メッセージの数を記録する RECV ベクトルデータを管理する。MH ハンドオフ時の放送メッセージの欠落，重複発生を防ぐため、その MH に対する RECV データを移動元 MSS から移動先 MSS へ通知する。また、ハンドオフした MH が送信するメッセージの因果順序保存のために、ハンドオフ時点での SENT の値も移動元 MSS から移動先 MSS へ通知する。移動先 MSS では、自局の SENT の値を、現在値と通知された SENT の最大値に更新する。
- ハンドオフ中の MH へ配達が必要な可能性があるため保持する必要が有る放送メッセージを格納する DELIV_MES というリストを管理する。DELIV_MES から放送メッセージを削除するタイミングを決めるために、MSS が送信する放送メッセージに、MSS 数次元の REDUCE ベクトルデータを付与する。REDUCE の値は、ハンドオフ中の MH に対する RECV と DELIV の最小値として求める。MSS は、ネットワーク上の各 MSS から受信済みの REDUCE の最大値を MSS 毎に記録するために、RECV_RDC データを管理する。DELIV_MES に含まれる放送メッセージの中で、各 MSS の RECV_RDC 値の最小値以下となる SENT ベクトル

が付与されたものを削除する。

今回提案する拡張の考え方を、端末モビリティを考慮しない因果順序保存放送プロトコルにも適用可能なネットワークの階層化と、端末モビリティを考慮した場合に固有のハンドオフ時の処理に分けて説明する。

3．1．1 ネットワークの階層化

因果順序保存放送型通信を行うネットワークを、以下の条件を満たす様にサブネットワークに分割する。(図2)

- サブネットワークに属するMSS数が一定値以下
- 1つのサブネットワークにのみ属するMSSと、2つのサブネットワークを接続するために2つのサブネットワークに属するMSSの2種類のみが存在する様に限定した接続形態でツリー状に接続される

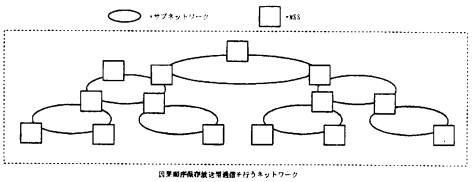

図2 サブネットワークへの分割例

サブネットワークへの分割方法としては、因果順序保存放送型通信を行うために必要な、単なる放送型通信で放送メッセージ転送を行うためにMSSをツリー状に接続し、そのツリーのサブツリーをサブネットワークとして、2つのサブネットワークに属するMSSが行う放送メッセージのサブネットワーク間転送は、因果順序保存放送プロトコル処理でのサブネットワーク間転送タイミングに合わせて行うことが考えられる。(図3)

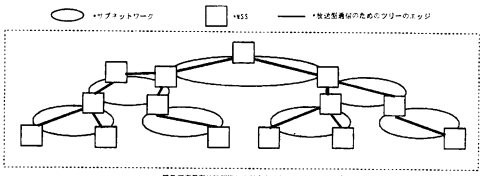

図3 サブネットワークへの分割方法例

端末移動性を考慮しない場合は、次のように処理を行う。

(1) [4]や[1]に示される因果順序保存放送プロトコルを、各サブネットワークを1ネットワークとみなして、サブネットワーク内のMSS間で実行する。

(2) 2つのサブネットワークA, Bを接続するMSSは、以下の様に処理を行う。

- サブネットワークA上で受信する放送メッセージは、その放送メッセージがサブネットワークA上で配達可能となった時点でサブネットワークB上で送信する。
- サブネットワークB上で受信する放送メッセージは、その放送メッセージがサブネットワークB上で配達可能となった時点でサブネットワークA上で送信する。
- 自局に接続するMHか自局が送信する放送メッセージは、サブネットワークAとBの両方に個別に送信し、両方で配達可能となって初めて自局に接続するMHと自局に配達可能として扱う。

上記の処理では、各サブネットワーク内で放送メッセージに付与するベクトル時計データの次元は、サブネットワークに存在するMSS数になる。サブネットワークに存在するMSS数が一定値以下になるように分割することで、放送メッセージのオーバーヘッドが、ネットワーク内に存在するMSS数に依存しないようにできることとなる。

3．1．2 ハンドオフ時の処理

MHのハンドオフに対応するために、3．1．1の方式で各サブネットワーク内では[1]のプロトコル処理を実行する方式に対して、以下の拡張を行う。以下では、[1]で使用するデータについて、サブネットワークA用データには添字aを、サブネットワークB用データには添字bを付ける。

(1) MHは以下の様に処理を行う。

- 各MHが送信する放送メッセージには、送信元MHを示す識別子と、MHが1ずつインクリメントするシーケンス番号を付与する。
- MHは、[送信元MHの識別子, そのMHから受信した最大のシーケンス番号値]を、受信した放送メッセージに対して記録するためのリストを保持する。

MHが放送メッセージをMSSから受信した時に、放送メッセージに付与される[送信元MHの識別子, シーケンス番号]の送信元識別子をキーとしてリストを検索し、データが見つからないか、見つかったデータ(データ1と呼ぶ)のシーケンス

番号が受信メッセージのシーケンス番号より小さい場合は、新規メッセージを受信したとしてメッセージを処理し、データ1をリストから削除した上で、放送メッセージに付与される[送信元MHの識別子, シーケンス番号]の情報をリストに追加する。そうでなければ、重複メッセージの受信として廃棄する。

・ MHがハンドオフを実行し、移動先MSSに接続した場合は、移動先MSSから通信許可を受信するまでは、放送メッセージの送受信を停止する。

(2) 2つのサブネットワークA,Bを接続するMSSは、以下の様に処理を行う。

・ [1]のRECVに類似のデータとして、サブネットワークA上で受信した放送メッセージについて、サブネットワークB以遠のサブネットワーク群に存在するMSSと、それらMSSに接続するMHに配達済みのものがいくつかを、サブネットワークA上の各MSSについて記録するベクトルデータRECV-Aを管理する。サブネットワークB上で受信した放送メッセージについても、AとBを逆にした同様のベクトルデータRECV-Bを管理する。

RECV-A, RECV-B更新のために、RECLIST-AとRECLIST-Bの2つのデータを管理する。RECLIST-Aは、サブネットワークAへ放送メッセージを送信する際に、その放送メッセージに付与するSENTaと、その時点でのサブネットワークB側でのDELIVbの組[Sa, Db]を記録するリスト。RECLIST-Bは、AとBを逆にした同様のデータ。サブネットワークAで放送メッセージを受信した時に、メッセージに付与されるREDUCEベクトルでRECV_RDCaを更新した後に、サブネットワークAの各MSSに対するRECV_RDCaの最小値ベクトルMIN_Raを求める。RECLIST-Aに含まれる[Sa, Db]の中で、MIN_Ra以下のSaの中で最大のものを[SA, DB]とすると、RECV-B=DBと更新し、SAよりも小さなSaを持つ[Sa,*]をRECLIST-Aから削除する(RECLIST-Aに含まれる[Sa,*]のSaは、全順序関係を満たすことに注意)。サブネットワークBで放送メッセージを受信した場合も、AとBを逆にして同様。

・ サブネットワークAへ放送メッセージを送信する際に付与するREDUCEベクトルは、ハンドオフ中のMHに対するRECVとDELIVaとRECV-Aの最小値として求める。

・ MHがハンドオフした場合に、移動元MSSは、MHに対するRECVの情報をハンドオフ中MHの情報として記録し、MHの識別子と自局の識別子を含む移動通知メッセージを因果順序保存放送メッセージとして送信する。

移動先MSSは、移動通知メッセージが配達され、かつ、MHが自局に新たに接続してきた時に、MHに通信許可を与えて、DELIV_MESに含まれる放送メッセージをMHへ送信し、移動通知メッセージに示されるMHの識別子と移動元MSSの識別子を含む移動確認メッセージを、因果順序保存放送メッセージとして送信する。

移動元MSSは、移動確認メッセージを受信した時に、そこに示される移動元MSSの識別子が自局であれば、ハンドオフ中MHの情報からMHの識別子に対応する情報を削除する。

(3) 1つのサブネットワークに属するMSSは、MHのハンドオフ時の処理を除いては、[1]に記述のプロトコル処理を行う。

これらの拡張の背景となる考え方を以下に説明する。

提案の拡張で追加したRECV-AとRECV-Bは、サブネットワークによる階層化の影響のため、配達されていることが保証される放送メッセージの情報となり、実際にはより多くの放送メッセージが配達済みの可能性はあるが、定常状態では、RECV-AとRECV-Bは、実際に配達済みの放送メッセージの情報と一致する。RECV-AとRECV-Bの使用目的が、全MHとMSSとに配達済みなので廃棄可能な放送メッセージを判断するための入力であることから、一時的に配達済みの放送メッセージが未配達として扱われることに不都合はない。

MHのハンドオフ時の処理で、MHがハンドオフ前に移動元MSSへ送信した放送メッセージと、MHがハンドオフ後に移動先MSSへ送信した放送メッセージの間に因果順序関係が成り立つようにするために、[1]では移動元MSSから移動先MSSへSENTの値を通知している。提案の方式では、ネットワークを階層化したため、MHが異なるサブネットワークに属するMSS間をハンドオフした場合に、SENTの値を移動元MSSから移動先MSSへ通知しても意味を持たなくなってしまう。ハンドオフしたMHが送信する放送メッセージ間の因果順序関係が成り立つことを保証するためには、ハンドオフ前にMHが移動元MSSから送信した放送メッセージと移動元MSSで配達されていた放送メッセージが、移動先MSSで配達された後に、MHが

ハンドオフ後の放送メッセージ送信を行うことが保証できればよい(条件1)。移動元MSSが、MHのハンドオフ後に移動通知メッセージを因果順序保存放送メッセージとして送信しているので、移動先MSSが移動通知メッセージを配達した時点では,移動元MSSがハンドオフ前に送信した放送メッセージと移動元MSSで配達されていた放送メッセージは、移動先MSSで配達済みになっている。移動先MSSは、移動通知メッセージを配達した後に、ハンドオフしてきたMHに通信許可を与え、通信許可を受け取るまではMHは放送メッセージの送信を行わないので、条件1が成り立つ。

また、MHハンドオフ時の処理で、移動先MSSでMHに配達する放送メッセージの欠落，重複を防ぐために、[1]ではMHに対するRECVを移動元MSSから移動先MSSへ通知している。提案の方式では、ネットワークを階層化したため、MHが異なるサブネットワークに属するMSS間をハンドオフした場合に、RECVの値を移動元MSSから移動先MSSへ通知しても意味を持たなくなってしまうので、正確にMHへ配達済みの放送メッセージが何かを管理することができなくなっている。移動先MSSで、ハンドオフ中のMHを含めて全てのMHに対する配達がまだ済んでいない可能性が有る放送メッセージをDELIV_MESに保持しているので、ハンドオフ後に移動先MSSからMHへDELIV_MESに保持する全ての放送メッセージを送信し、MHが送信する放送メッセージにMHの識別子とシーケンス番号を付与して、送信元MHの識別子とシーケンス番号で受信済みメッセージかどうかを判断可能とすることで、メッセージの欠落，重複を回避する。

紙数の関係でプロトコル処理の詳細記述は割愛する。

4．正当性の証明の概要

ネットワークを階層化した場合に因果順序保存放送通信が行われることは、以下の定理と補題を用いて証明できる。以下で用いる記法を示す。

- 因果順序保存放送メッセージmをMHjが送信するイベントをCBCASTj(m)
- 因果順序保存放送メッセージmがMHjに配達されるイベントをDELIVERj(m)
- 因果順序保存放送メッセージmをMSSjが送信するイベントをMSS_CBCASj(m)。2つのサブネットワークを接続するMSSについては、mは自局に接続するMHか自局が送信する放送メッセージのみとし、MSSがサブネットワーク間を転送する放送メッセージを含めない。
- 因果順序保存放送メッセージmがMSSjに配達されるイベントをMSS_DELIVj(m)。2つのサブネットワークを接続するMSSについては、mは、いずれかのサブネットワークで受信するメッセージと、自局に接続するMHか自局が送信する放送メッセージの両方を含む。
- イベント間の因果順序関係を→と記す。
  同一MH, MSSで生起したイベントには、生起順と同じ因果順序関係がある。
  CBCASTj(m)→DELIVERk(m) (kは任意のMH)
  MSS_CBCASTj(m)→MSS_DELIVk(m) (kは任意のMSS)
  CBCASTj(m)→MSS_CBCASTk(m) (MSSkにMHjが接続)
  MSS_DELIVk(m)→DELIVERj(m) (MSSkにMHjが接続)
  →には推移則が成り立つ。
- 因果順序保存放送通信が行われているというのは、CBCASTj(m1)→CBCASTk(m2)ならば、任意のMHnについてDELIVERn(m1)→DELIVERn(m2)が成り立つことと定義する。

本提案で導入したRECV-A, RECV-Bについて、以下の定理1が成り立つ。

(定理1)

(a) RECV-A以下のSENTを付与される、サブネットワークA上で送信される放送メッセージは、サブネットワークB以遠で配達済みであることが保証される。同様に、RECV-B以下のSENTを付与される、サブネットワークB上で送信される放送メッセージは、サブネットワークA以遠で配達済みであることが保証される

(補題1)

(a) あるMSSxでMSS_CBCASTx(m1)→MSS_CBCASTx(m2)ならば、任意のMSSyでMSS_DELIVy(m1)→MSS_DELIVy(m2)

(b) あるMSSxでMSS_DELIVx(m1)→MSS_CBCASTx(m2)ならば、任意のMSSyでMSS_DELIVy(m1)→MSS_DELIVy(m2)

が成り立つ。

(補題2)

MHの移動がない場合、

(a) ある MHx で CBCASTx(m1)→CBCASTx(m2)ならば、任意の MHy で DELIVERy(m1)→DELIVERy(m2)
(b) ある MHx で DELIVERx(m1)→CBCASTx(m2)ならば、任意の MHy で DELIVERy(m1)→DELIVERy(m2)

(補題 3)
(a) 任意の MSSm で MSS_DELIVm(m1)→MSS_DELIVm(m2)が成り立つ時に、MHx で DELIVER(m1)が発生した後に MHx がハンドオフしても、DELIVERx(m1)→DELIVERx(m2)が成り立つ。

(補題 4)
(a) MHx が、CBCASTx(m1)後にハンドオフし、その後に CBCASTx(m2)した場合、任意の MH で DELIVERx(m1)→DELIVERx(m2)

(補題 5)
(a) MHx が、DELIVERx(m1)後にハンドオフし、その後に CBCASTx(m2)した場合、任意の MH で DELIVERx(m1)→DELIVERx(m2)

(定理 2)
(a) CBCASTx(m1)→CBCASTy(m2)ならば、任意の MH で DELIVER(m1)→DELIVER(m2)が成り立つ

(証明)
CBCASTx(m1)→CBCASTy(m2)の関係を構成する連鎖には、
−CBCASTa(mi)→CBCASTa(mj)
−CBCASTa(mi)→DELIVb(mi)→CBCASTb(mj)
の 2 種類の基本的な連鎖が連なっている。
- CBCASTa(mi)→CBCASTa(mj)ならば、任意の MHx で DELIVERx(mi)→DELIVERx(mj)
- DELIVb(mi)→CBCASTb(mj)ならば、任意の MHx で DELIVERx(mi)→DELIVERx(mj)

の 2 つが成り立てば、因果順序関係が推移的であることから、CBCASTx(m1)→CBCASTy(m2)ならば、任意の MH で DELIVER(m1)→DELIVER(m2)が成り立つことがいえる。上記 2 項は、補題 2 から 5 で示されているので、定理 2 が成り立つ。

5．今後の課題

提案の方式では、ネットワークをサブネットワークに分割し、各サブネットワーク内では[1]に示されるのと同様な因果順序保存放送通信を行い、2 つのサブネットワークを接続する MSS で、1 つのサブネットワーク上で受信した放送メッセージが配達可能となった時点で、他方のサブネットワークへ送信を行うため、ネットワーク全体にメッセージを配達するための遅延時間が増大すると考えられる。また、2 つのサブネットワークを接続する MSS での処理において、本来存在しなかった因果順序関係が発生する。これら 2 つのことが、メッセージ配達に対する遅延時間にどの様な影響を及ぼすかの評価を、シミュレーション等の方法で行うことが必要と考える。

今回提案の範囲では、ネットワークに存在する MSS や MH の増減を考慮できていない。MSS が増減した場合のサブネットワーク構成方法も含めて、今後の検討が必要と考える。

6．謝辞

日頃より御指導頂いております IP 通信システム部 伊藤部長、土田チームリーダ、及び、本報告の内容について議論頂いた IP 通信システム部の皆様に感謝致します。

参考文献

[1] 大堀，井上，増澤，藤原，”分散移動システムのための前後関係保存放送プロトコル”，電子情報通信学会論文誌 D-I，Vol.J82-D-I，No.2 pp425-435，1999
[2] S. Alagar and S. Venkatesan, “Causal ordering in distributed mobile systems”, IEEE Trans. On Computers, vol.46, no.3, pp353-361, 1997
[3] K. Taguchi and M. Takizawa, “Group Communication Protocol for Hierarchical Group”, IPSJ Journal, vol.44, No.3, pp674-681, 2003
[4] K. Birman and T. Joseph, “Lightweight causal and atomic group multicast”, ACM Trans. on Computer Systems, vol.9, no.3, pp272-314, 1991
[5] L. Lamport, “Time, clocks and the ordering of events in a distributed system”, CACM, vol.21, no.7, pp558-565, 1978