メッセージブローカサーバを用いた適応的全順序マルチキャスト方式の提案とその評価

田 学軍[†] 井手口 哲夫[†]
安川 博[†] 奥田 隆史[†]

分散アプリケーションシステムにとって，全順序グループ通信サービス（TO-GCS）は非常に重要である．特に順序調整機能が付いていない分散アプリケーションに不可欠である．全順序マルチキャストの１つの方式であるシーケンス方式では，シーケンスサーバに負荷が集中する問題があり，規模拡張性が低い．All-Ack 方式は事前に決められた順番（Agreed mulicast）にメンバごとにそのメンバからのメッセージを１つずつ配送するため，各メンバによる送信頻度の変化に対応できなく，遅延が大きい．本論文では適応的全順序マルチキャストに基づいて ATOP/MBS（Adaptive Totally Ordering Protocol with the Message Broker Server）方式を提案し，シミュレーションで評価した．その結果からこの方式では動的に各メンバの送信頻度に応じてメッセージ配送を実現するとともに，ダミーメッセージを抑制することができる．また，遅延と遅延変動において改善されることが示された．

Adaptive Totally Ordered Multicast Communication Protocol with the Message Broker Server and Its Evaluation

Xuejun Tian,[†] Tetsuo Ideguchi,[†] Hiroshi Yasukawa[†] and Takashi Okuda[†]

In distributed appplication system, Totally Ordered Group Communication Service (TO-GCS) is very important, and it is a powerful infrastructure for building distributed fault-tolerant application such as distributed shared memory, Computer Supported Cooperative Work and distributed monitoring. Much work has been dedicated to Totally ordered multicast protocol. One of them, Adaptive Totally Ordered Multicast Communication(ATOP), is able to dynamically alter the message delivery order in response to changes in the transmission rates of group members. In this paper, we propose a protocol ATOP/MBS by improving ATOP, which use no dummy message and can adapt well in case of transmission frequency changing in comparatively short period. According to the simulation results, TO-GCS with ATOP/MBS can be carried out at low delivery latency and fluctuations of delivery latency.

1. はじめに

分散アプリケーションシステムにとって，全順序グループ通信サービス（TO-GCS）は非常に重要である．全順序グループ通信サービスでは，マルチキャストのメンバに対して，一連の更新イベントはすべてのメンバに同じ順序でアプリケーションに配送される[1)~3)]．

下位層によって提供される全順序グループ通信サービスにより，アプリケーションでは，順序を整えるために先行するメッセージを待つことなく，下位層から渡されたメッセージをただちに受理することができる．これは特に順序調整機能が付いてない分散アプリケーションに不可欠である．応用例としては共有仮想環境，電子会議，分散共有メモリ，コンピュータ支援協同作業アプリケーションなどがある．

マルチキャスト通信ではユニキャスト通信と異なり，パケットは受信者に到着するまでにその途中のルータによって複製される．この方式では異なるドメインのメンバからなるグループ通信に対して，ネットワーク資源の浪費，すなわち，トラヒックの増加を防ぐことができる．

全順序マルチキャストの実現方法として多くの検討が行われている[4)~8)]．シーケンス方式では，シーケンスサーバを用いてメッセージに順序番号を付ける方式である．すなわち，１つのグループに１つシーケンス

† 愛知県立大学情報科学部
Faculty of Information Science and Technology, Aichi Prefectural University

サーバがあり，マルチキャストメッセージを送ろうとするクライアントがサーバにシーケンス番号を申請する．サーバは連続的にマルチキャストメッセージに 1 つの番号を決め，クライアントにシーケンス番号を知らせる．クライアントはこのシーケンス番号をメッセージに付けてマルチキャストを行う．この方式はシーケンスサーバに負荷が集中する問題が発生する．

一方，All-Ack[9]アルゴリズムは，受信側は事前に決められたメンバの順番にそのメンバからの受信メッセージをアプリケーションに配送する．この方式では，全順序を実現するために基本的に 2 つのステップがある．1）メッセージの転送ではメンバは他の各メンバから少なくとも 1 つのメッセージを受信するまで待つ．2）ステップ 1）を満たしてから事前に決められたメンバの順番に各メンバからの受信メッセージを 1 つずつアプリケーションに配送する．この方式はシーケンスサーバのようなサーバを使わないため，負荷の集中が発生しにくい．しかし，上述したようにこの方式では，グループの各メンバは単位時間内にメッセージの送信要求の数または送信頻度が同じまたは近い場合に受信されたメッセージがスムーズにアプリケーションに配送される．あるメンバの送信頻度が増大する場合にはこのメンバからの送信されたメッセージの配送は遅れてしまう．このようにこの方式ではメンバの送信頻度の変化に対応できない．さらに各メンバからのメッセージを待つ必要があるため，すでに受信されたメッセージの配送遅延が大きくなる．

適応的全順序マルチキャストでは[10]，負荷集中を避けるため，All-Ack に似たアルゴリズムが採用されている．ただ事前に各メンバからのメッセージを配送する際にメンバの順番の決め方は異なる．この方式では各メンバの送信頻度に応じて動的メッセージ配送の順序を調節できる．しかし，全順序グループ通信をスムーズに実現するために空のダミーメッセージが使用される．空のダミーメッセージを受信し，順序付を行ったうえで配送処理を行われなければならないため，頻繁にこの空ダミーメッセージが使用されるとグループメンバに負担をかける．このような課題に対して，ATOP/MBS 方式[11]を提案，評価した．本論文では，さらに提案した方式を改良し，既方式[10]とのシミュレーションによる比較を行う．

2. 提案方式

本論文では，適応的全順序マルチキャストの改良方式を提案する．

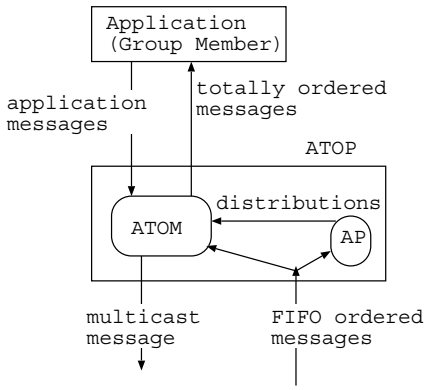

図 1 ATOP の基本構成
Fig. 1 Basic structure of ATOP.

2.1 ATOP 方式

適応的全順序マルチキャスト（ATOP）方式[10]ではグループの各メンバは**図 1** に示す構成となっている．グループの中から 1 つのメンバがブックキーパとして選ばれ，受信したマルチキャストメッセージと送信元 ID をチェックして，各メンバからの受信メッセージの数によって重み分布（分布）を作成し，定期的に各メンバに分布を送信する．ブックキーパ（サーバ）は送信要求が少ないメンバに割り当てることが適切である．他のメンバはクライアントとなる．ブックキーパにおけるプロセス p からの受信メッセージの分布式 $w[p]$ は下記のように表される．

$$w[p] = \frac{p\text{からのメッセージ数}}{\text{受信したメッセージ総数}} \quad (1)$$

ここで，メンバの集合に対して $\Sigma_{p\in P} w[p] = 1$ が成り立つ．

各メンバは最新の受信した分布を送信用分布としてマルチキャストメッセージに対応の送信用分布 ID（sending distribution, sndDistID）を付けてシーケンス番号を添付し，送信する．送信されるメッセージには順序制御のために使われる情報として，送信メンバ ID（senderID），分布 ID と分布 ID における順序番号（seqno）がある．この順序番号は配送の全順序番号（totalOrdered）と異なり，各メンバにおいて独立であり，送信用分布 ID を更新する際にリセットされる．

受信側メンバがメッセージの配送順序（全順序）を決め，アプリケーションに送信する．全順序を決めるにはまずどこからのメッセージを配送するかを決める．すなわちメッセージの送信元を決める．ATOM は配送用分布 ID（ordering distribution ID，ordDistID）に対応する分布と乱数生成器によって送信元を決定す

る．全順序を保証するために各メンバが同一乱数生成器を持つ．すなわち，各メンバは同じ乱数系列を生成する．式 (1) によって 1 つの分布に対してメンバを重みによって 0〜1 の範囲に対応させ，これを乱数発生器からの 0〜1 の乱数に対応させる．したがって配送にあたって送信元が決められる．決められた送信元からの該当の順序番号（seqno）が付いているメッセージが存在したら，配送する．また到着していなければ受信されるまで待つ．

このような処理機構を実現するために，ATOP（Adaptive Totally Ordering Protocol）では 2 つのモジュール ATOM（Adaptive Total Ordering Mechanism）と AP（Adaptive Policy）から構成される（図 1）．

(1) AP モジュール

AP モジュールは受信したメッセージとその送信元 ID を読み出し，ある間隔に受信したメッセージ総数に対して送信頻度表す分布 $w[p]$（重み）を計算し，重みとメンバ ID を対応付けたリストを定期的に ATOM に送る．

(2) ATOM モジュール

アプリケーションからのマルチキャストメッセージを送信する際に，最新の分布 ID と senderID，順序番号（seqno）をメッセージに添付して送信する．受信したメッセージをアプリケーションに配送する際にメッセージに添付された分布 ID と乱数生成器によってメッセージの順序付けを行い，メッセージを配送する．配送されるべきメッセージが一定時間到着しない場合には，ダミーメッセージをマルチキャストする．

基本動作として，グループの中からただ 1 つのメンバの AP がブックキーパとして選ばれて定期的に得た最新の分布を各メンバに送信する．受信側にある配送用分布 ID（ordDistID）と送信用分布 ID（sndDistID）との間に時間のずれがあり，これらの ID の値は発生順に上順の値が付与されるため，$sndDistID \geq ordDistID$ である．受信側は以下の条件が同時に成立するときに ordDistID を更新する．

1. 受信側メンバは配送されないメッセージバッファの各メンバからのメッセージの中に ordDistID が付いているメッセージが存在しない．
2. 受信側メンバは配送されないメッセージバッファの中に各メンバに対してそのメンバからのメッセージが少なくとも 1 つのメッセージが存在する．そのメッセージに付いている分布 ID は ordDistID より大きい．

また，ダミーメッセージの送信の時期に関して，メンバの送信頻度が変化し，送信用分布の中に当メンバの重みとその時点の実際の送信頻度と隔たりがある場合にダミーメッセージをマルチキャストする．これにより各クライアントでは配送すべきメッセージが受信バッファに存在せず，ブロックされる問題を避けることができる．しばらく送信の要求がないメンバは後に送信するために分布の中の当メンバの重みを維持する必要があり，ダミーメッセージもマルチキャストする．

2.2 ATOP/MBS 方式の提案

ATOP では，あるメンバの送信頻度の変化による配送ブロックを回避するためにダミーメッセージが送られ，各メンバの送信頻度の分布の変化が激しくなると，送受信されるダミーメッセージの数が増大する．分布の更新間隔を短くすることでより正確に通信頻度を反映することが考えられるが，負荷が高いときは，避けるべきである．また，分布の更新は古い分布 ID が付いたメッセージは全部配送されたあと行われるため，メッセージ送信遅延の変動は大きい．空のダミーメッセージはすべてのグループメンバに送信されるため，普通のメッセージと同じように受信，順序付などが行われるため，グループメンバに負担を掛ける．

この問題を改善するためにダミーメッセージの送信を抑えなければならない．よって，送信用分布と実際の送信頻度の相違を吸収するために 1 つの調節サーバとしてメッセージブローカサーバを設けてダミーメッセージを抑える ATOP/MBS（Adaptive Totally Ordered Protocol/Message Broker Server）方式を提案する．

2.2.1 システム構成

システム構成は**図 2** に示すように ATOP と同じ構成法を採用し，マルチキャストグループの 1 つのメンバがブックキーパとなり，通信状況によって新しい分布の発行を行う．それにマルチキャストグループの 1 つのメンバがメッセージブローカサーバ（MBS）になり，送信分布に基づいて実際の送信頻度と比較して，分布が更新されるまでの間に送信を調節する．

次の場合には調節する必要がある．すなわち，一定の頻度でメッセージを送信していたクライアントがある時点から突然多くのメッセージを送信しはじめる場合と，送信頻度が減ってしまう場合である．送信頻度は低下し，送信分布の当クライアントの重みより低くなるとブックキーパによって割り当てられた重みに相当する部分を他のクライアントに利用させるために必要な情報を送る．必要な情報は送信メンバ ID（senderID），分布 ID，シーケンス番号（seqno）などを含む．同様に，送信頻度が上昇した場合も送信分布

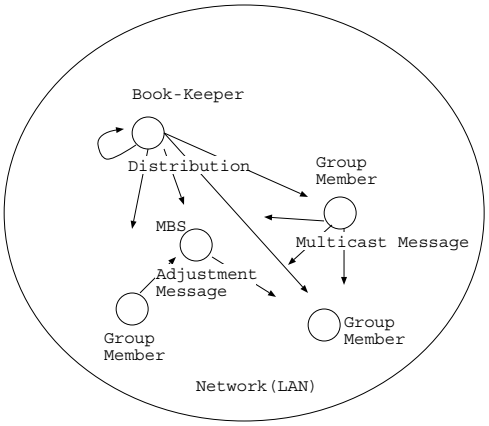

図 2 ATOP/MBS のシステム構成
Fig. 2 The ATOP/MBS system structure.

の重みより高くなるとメッセージブローカサーバに送信権を要求する．MBS が調節情報によって空番号を必要とするメンバに送信する．空番号を獲得したメンバがマルチキャストメッセージを送信する際に空番号に含まれる元の送信メンバ ID，分布 ID，シーケンス番号などの情報とともに当メンバの送信メンバ ID，分布 ID，シーケンス番号を添付して送信する．

2.2.2 送信頻度に基づく通信の調整法

次に MBS に知らせる調節メッセージの数を決める方法を述べる．まず，関連パラメータを次のように定義し，ATOP/MBS 方式の具体的なアルゴリズムを次の節で記述する．

Sliding Window：時間 T，ブックキーパが新しい分布を計算する時間間隔．

Sub-SW（Sub-Sliding Window）：時間 sbT，各メンバが調節メッセージの数を計算する時間間隔．

sbT_i：i 番目の Sub-SW 間隔．

$dmN_i(p)$：i 番目間隔の間に受信したプロセス p が提供した番号のダミーメッセージ数，p∈ P.

$bkN_i(p)$：i 番目間隔の間に受信したプロセス p が提供した番号を使ったメッセージ数．

$N_i(p)$：i 番目間隔の間に受信したプロセス p からのマルチキャストメッセージ数．

δ_i：sbT_i 開始の時刻に MBS に請求する調節メッセージの数．

w[p]：送信用分布においてプロセス p の重み．

空番号の提供または請求数は，Sub-SW ごとに計算する．たとえば，送信分布の重み $w[p]$ を持っているメンバ p において Sub-SW の間隔で実際に p が送信したメッセージの数に対する受信したメッセージ総数の割合（送信頻度）は $w[p]$ と比べ，両者の差を受信したメッセージ総数に掛けると空番号の提供または請求数が得られる．実際の送信頻度が $w[p]$ より大きい場合は番号請求，小さい場合は空番号の提供となる．式 (1) に示すように $\Sigma_{p\in P} w[p] = 1$ であるため，請求する空番号の総数が提供される空番号の総数と等しいアルゴリズムを設計した．sbT_i 開始の時刻に MBS へ請求する調節メッセージは sbT_{i-1} の間に送信されたマルチキャストメッセージと送信用分布の w[p] との相違への修正である．もし，空番号を獲得したメンバは送信要求がしばらくないと，空番号を使わなかった場合に TO-GCS 送信がブロックされる．このことを避けるため，空番号を獲得してから一定時間（dmT と呼ぶ）経過すると，その空番号を空のメッセージ（ダミーメッセージ）に付けて送信する．

何か 1 つのメンバ，たとえばメンバ p が受信したメッセージ全体の中に p からのメッセージの割合によって，sbT_i 開始の時刻に調節メッセージ数δ_i を計算し，MBS に知らせる．δ_i の計算式は下の式で表す．

$$\begin{aligned}\delta_i &= \delta_{i-1} \\ &+ (w[p] \times (\Sigma_{p\in P} dmN_{i-1}(p) + \Sigma_{p\in P} bkN_{i-1}(p) \\ &+ \Sigma_{p\in P} N_{i-1}(p)) - (dmN_{i-1}(p) + bkN_{i-1}(p) \\ &+ N_{i-1}(p))) \qquad (2)\end{aligned}$$

δ_i はマイナスの場合は当該メンバが空番号を請求し，プラスの場合は空番号を提供する．なお，$\delta_0 = 0$ とする．式 (2) の中で，$\Sigma_{p\in P} dmN_{i-1}(p) + \Sigma_{p\in P} bkN_{i-1}(p) + \Sigma_{p\in P} N_{i-1}(p)$ は sbT_{i-1} の間に発生するグループ全体のマルチキャストメッセージ総数であり，$dmN_{i-1}(p) + bkN_{i-1}(p) + N_{i-1}(p)$ はメンバ p からのメッセージの総数である．$N_{i-1}(p)$ は p からの通常メッセージの数であり，$dmN_{i-1}(p)$ は p からのダミーメッセージ数であり，$bkN_{i-1}(p)$ は他のメンバが p のシーケンス番号（seqno）を使って出したメッセージ数である．つねに実際の送信頻度が送信分布と一致する場合には，$N_{i-1}(p) = w[p] \times (\Sigma_{p\in P} N_{i-1}(p))$ が成り立ち，最初 $i = 1$ のとき，$\delta_0 = 0, dmN_0(p) = 0, bkN_0(p) = 0$ であり，$\delta_1 = 0$ となる．したがってこの場合には，$\delta_i = 0$ となり，調節が発生しない．実際の送信頻度が送信分布と一致しない場合ではメンバ p において $dmN_i(p)$ がメンバ p のシーケンス番号（seqno）を使うため，p からのメッセージに加える．同様，$bkN_{i-1}(p))$ は p から提供された p のシーケンス番号（seqno）を使うため，p からのメッセージとして計算する．

ブックキーパにおいて，新しい分布の計算は間隔 T ごとに，式 (1) に示すように行う．式 (1) には，実際の送信頻度を表すために受信したマルチキャストメッ

セージ数はダミーメッセージを含まない．各メンバは新しい分布を受信したら δ_0 を 0 に設定し，δ_i に関する統計パラメータを 0 に設定する．これより生じた配送バランス問題は分布更新の際に解消する．

2.2.3 ATOP/MBS 方式のアルゴリズム

（1）突然送信頻度が減り始めたクライアント（メンバ p）の場合

1. メンバ p の AP はその時点での当クライアントの送信頻度を定期的に ATOM に送る．
2. メンバ p の ATOM は δ_i を sbT_i の周期で計算し，δ_i の整数部分（Int(δ_i)）を提供する空番号数として調節メッセージに添付し，MBS に知らせる．この調節メッセージは空番号を特定するに必要な情報，現在送信に使っている分布 ID，メンバ p の ID，他のクライアントに使わせるメッセージ数 Int(δ_i)，開始シーケンス番号 n（現在 seqno=n 仮定する）からなる．メンバ p はシーケンス番号 n から Int(δ_i) 個番号を空ける．
3. 当クライアントは現在 seqno を n+Int(δ_i) と設定し，n+Int(δ_i) の後の番号を使ってマルチキャストメッセージの送信を続ける．

（2）現状の送信頻度より多くのメッセージを送信しはじめるクライアント（メンバ p）の場合

1. メンバ p の AP はその時点での当クライアントの送信頻度を定期的に ATOM に送る．
2. メンバ p の ATOM は $\delta_i(p)$ を sbT_i の周期で計算し，sbT_i の整数部分を請求する空番号数として調節メッセージに添付し，MBS に空番号を請求する．
3. MBS からの返事が来る前，メンバ p は通常にマルチキャストメッセージの送信を行う．MBS から空番号が来たら，通常マルチキャストメッセージに特別メッセージ（Brokered Message，BMG）の認識マーク，MBS からの空番号の情報，すなわち，空番号に関する分布 ID，クライアント ID のほか，メンバ p の ID，分布 ID，送信順序番号（seqno）を付けて送信する．

MBS の動作はただ持っている空番号を空番号の要求に応じて割り当てることである．

2.2.4 配送全順序調節の実現

空番号を使ったメッセージ BMG が受信された場合に ATOM はメッセージに付いたマークにより MBG であることを知り，以下の手順で全順序を保証する前提でこのメッセージを取り扱う．例としてメンバ p がメンバ q からの BMG を受信することにする．この BMG はメンバ r に提供された空番号を使っていると仮定する．

1. メンバ p はメッセージを受信する．
2. メンバ p の ATOM はメッセージが BMG であることを知ってそれを 2 つのメッセージに分ける．1 つは CBMG（cut BMG）であり，メンバ r の ID，分布 ID，シーケンス番号（seqno）およびメンバ q の ID からなる．もう 1 つはメンバ q からのもとのメッセージとなり，普通のメッセージとして取り扱われる．
3. 配送順番は CBMG となった場合に CBMG からメンバ q の ID を取り出し，次の配送メッセージの送信元をメンバ q にする．

CBMG を配送するとき，次の配送メッセージの送信元の決定は乱数シーケンスより発生するのではなく，直接にメンバ q にすることは $w[q]$ より多くのメンバ q からのメッセージを配送することができた．ブックキーパの AP は BMG から実際の送信メンバ q の ID を取り出し，新しい分布 ID 算出のために動的に送信頻度の重みを計算する．

2.2.5 変動調節ウィンドウ

提案する ATOP/MBS 方式は，各メンバのマルチキャスト通信頻度の変動を分布の重みとしてメッセージブローカサーバを介し，ダミーメッセージを抑制することにより，全順序マルチキャストメッセージの配送をスムーズに実現することが可能である．

ブックキーパにおける分布の計算と配布の間隔は，Sliding Window によって決められる．Sliding Window はメッセージの数，または時間の間隔にすることができる．マルチキャスト通信量が低いとき，一定のメッセージの数になるのは時間がかかるため，速く調節できない．ここは Sliding Window を時間 T とする．時間 T が経過するたびに，ブックキーパが新しい分布を計算し，閾値を超えると現時点のマルチキャスト通信を反映する新しい分布を発行する．

Sliding Window の間に時間 sbT（Sub-Sliding Window）ごとにメッセージブローカによって全順序マルチキャスト通信を調節する．メッセージブローカの調節で Sliding Window をもっと長い時間に設定することが可能である．Sliding Window と Sub-Sliding Window の設定はエンド・エンドの遅延などに影響がある．Sub-Sliding Window を短くするとメッセージブローカに負担をかける．一方，長くすると調節の効果が低くなる．時間 Window の設定はアプリケーションにも関係がある．

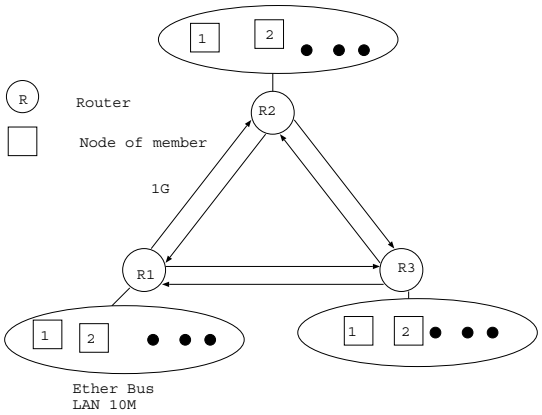

図 3 シミュレーションモデル
Fig. 3 Simulation model.

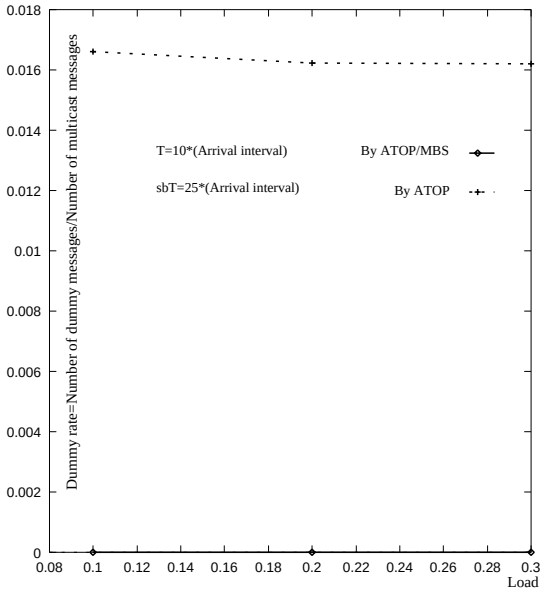

図 4 ダミーメッセージ発生率
Fig. 4 Dummy rate.

3. シミュレーション

ATOP/MBS 提案に対して，シミュレーションによりその性能を評価する．

3.1 シミュレーションモデル

シミュレーションのモデルは**図 3** に示すように 3 つのパーティションからなり，各パーティションはイーサバス型 LAN である．イーサバス型 LAN のノードは全部グループメンバと仮定する．イーサバス型 LAN の場合はブロードキャストが使えるが，非メンバがある場合を考えて各ノードがマルチキャストメッセージをメンバごとに送信するものとする．ここで，1 つのノード上には 1 つのメンバとする．ルータは当パーティション向けのマルチキャストメッセージを受信したら，メッセージを複製し，各メンバに送信する．このモデルではグループメンバ以外のノードを考えずすべてマルチキャストメッセージのトラヒックとする．関連パラメータは以下に示す．

パーティション数：3
パーティション中のメンバ数：$n = 10$
イーサバス型 LAN 速度：10 Mbit
パーティション間の通信速度：1 Gbit
固定フレーム長：512 byte.
空番号の持ち時間（dmT）：1.5×sbT

評価指標は，全順序マルチキャストにおける重要なエンド・エンドの遅延，ダミーメッセージ発生率，遅延変動とする．エンド・エンドの遅延はメンバの送信要求が生じてから受信され，配送されるまでの遅延とする．ダミーメッセージ発生率はダミーメッセージ総数を送信メッセージ総数で正規化した値，またはメッセージあたりのダミーメッセージ数である．遅延変動は平均遅延に対する遅延の分散であり，リアルタイムアプリケーションにとって重要な指標である．

ここでは特定のアプリケーションを想定しないため，マルチキャストメッセージの発生をポアソン到着とする．グループメンバの送信頻度が変化する場合に提案した ATOP/MBS を評価するために送信頻度の変化を示すイベントを以下のように発生させる．

送信頻度の変化するノード数：メンバ数の 10%
送信頻度の変化する継続時間：$10s \times r$
送信頻度の変化するイベントの到着率：平均到着率 $\times(1 + 40\% \times r)$

上の r は 0–1 の乱数であり，項目ごとに異なる．シミュレーションの初期化に送信頻度の変化を示すイベントを発生させ，シミュレーション中に終わるつど，ランダムに送信頻度の変化を示す次のイベントを決め，発生させる．

3.1.1 シミュレーション結果と評価

MBS による調節の効果にかかわるパラメータ sbT を異なる値に設定しシミュレーションを行った．シミュレーションの結果を**図 4**，**図 5**，**図 6** に示す．横軸は LAN の負荷を表す．負荷はマルチキャストでは 1 つのメッセージ（フレーム長：512 byte）が各メンバに送信されることを考えて算出されている．

図 4 はマルチキャストメッセージ数で正規化したダミーメッセージ発生率を示している．ATOP 方式では，1 つのメンバにおいて 1 回 δ_i を計算して，1 つ調節メッセージを発生する．ATOP 方式に対して，遅延とダミー発生率を考え，sbT を平均到着間隔の 25

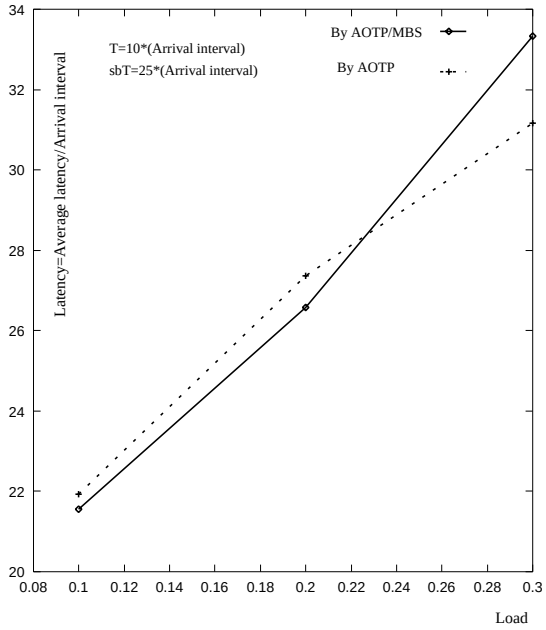

図 5 送信要発生から転送までの遅延
Fig. 5 The latency of messages from occurring to being delivered.

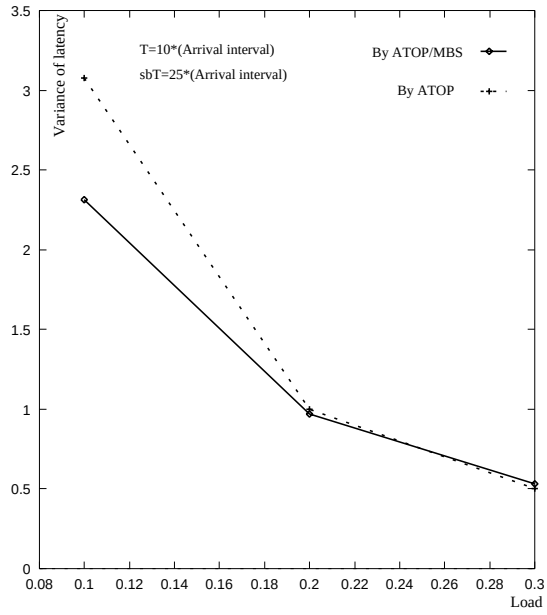

図 6 遅延の分散
Fig. 6 The variance of latency.

倍にしてシミュレーションを行った．ATOP 方式のダミーメッセージ発生率は負荷率にかかわらずほぼ一定（0.0165）である．ATOP/MBS の場合では，ダミーメッセージの発生は獲得した空番号を時間内（dmT）に使えない場合に起こる．図 4 に示すように同じ条件で ATOP/MBS 方式ではダミー発生率はほぼ 0 となり，ダミーメッセージによるトラヒックの増加は起きない．

図 5 は提案した方式と従来方式におけるメッセージのエンド・エンドの遅延を示す．エンド・エンドの遅延はメッセージ発生の間隔に影響されるため，エンド・エンドの遅延をメッセージ発生の間隔で正規化し，結果を図 5 に示している．この図から，負荷またはマルチキャスト発生率の増加にともなって，正規化された転送遅延も増えることが分かる．

ATOP/MBS 方式は負荷 0.2 以下の場合に ATOP 方式より転送遅延が改善される．負荷 0.3 を超えるとイーサバス型 LAN の送信遅延の増加のため，ATOP/MBS 方式での転送遅延が大きくなる．

ダミーメッセージを抑制しながら，MBS との間の調節のための通信が生じる．この調節メッセージはユニキャストであり，グループ全体宛のマルチキャストメッセージと比べてトラヒック量がかなり低いと考える．このメッセージの発生率を見ると ATOP によるダミーメッセージ発生率の約半分以下になっているため，マルチキャストメッセージの遅延への影響は少ないと思われる．

図 6 にリアルタイムアプリケーションに対して重要な遅延変動を示す．遅延の変動は低負荷時に大きく．負荷の増加とともに小さくなる．それは負荷による到着間隔と関係がある．図 6 に示すように ATOP/MBS 方式は従来方式よりエンド・エンドの遅延変動の分散は低負荷の場合に改善され，負荷 0.3 の場合には差が目立たない．

4. 結　　び

提案した ATOP/MBS 方式では，シーケンス方式より負荷の集中の軽減，ダミーメッセージの発生の抑制およびメンバの送信頻度変動への適応性などの面で改善できる．シーケンス方式ではすべてのマルチキャストメッセージが送信される前にシーケンスサーバにシーケンス番号が請求される．よってサーバに負荷が集中してしまう．シーケンスサーバの代わりに MBS は通信頻度の変化で分布に合わない部分だけを調節することで負荷の集中を軽減する．ATOP ではブックキーパは分布を計算し，新しい分布をすべてのメンバに配信しなければならない．頻繁に新しい分布の配信はネットワークに負担をかける．ネットワークを監視し，自メンバは分布によって送信頻度が低い場合にダミーメッセージを発信し，調節する．しかし，急激に送信頻度が増加する場合には調節できない．提案した

方式ではMBSを用いて式 $\Sigma_{p\in P} w[p] = 1$ に基づいて送信頻度の変化に応じて送信を調節することでダミーメッセージの発生を抑制することが可能となった．ダミーメッセージの発生は空番号を獲得したメンバの送信頻度が急に低下し，一定の時間が経って空番号を使わない場合に限定される．ATOPではグループのメンバは一時的にマルチキャストメッセージがなくても新しい分布計算にあたって自メンバの重みが0にならないように最低数のダミーメッセージを送信しなければならない．提案方式ではダミーメッセージの発生で自メンバの分布の重みを維持する必要がない．分布の更新の間にもっと短い周期 *sbT* でMBSによる調整は送信頻度の急激な変化に対して，全順序メッセージをスムーズに配送できる．本論文において比較のための従来方式のシミュレーションを行い，評価した結果では，送信頻度の急変の場合に低いダミーメッセージ発生率を抑制できた．

今後の課題として，MBSによる通信遅延が全順序メッセージ転送への影響および特定のアプリケーションによる適用性の検討が残されている．

本研究は平成13年度科学研究費基盤研究C（課題番号13680487）の補助を受けている．

参 考 文 献

1) Dolev, D., Kramer, S. and Malki, D.: Early delivery totally ordered broadcast in asynchronous environments, *Proc. 23rd Annual International Symposium on Fault-Tolerant Computing*, pp.544–553 (June 1993).
2) Guerraoui, R. and Schiper, A.: Total order multicast to multiple groups, *Proc. 17th IEEE Int. Conf. on Distributed Computing Systems*, Baltimore, pp.578–585 (May 1997).
3) 細谷 篤，佐藤文明，水野忠則：適応的全順序マルチキャストの拡張，情報処理学会論文誌，Vol.42, No.2, pp.138–146 (2001).
4) Chang, J. and Maxemchuck, N.: Reliable broadcast protocol, *ACM Trans. Comput. Syst.*, Vol.2, No.3, pp.251–273 (1984).
5) Birman, K., Schiper, A. and Stephenson, P.: Lightweight causal and atomic group multicast, *ACM Trans. Comput. Syst.*, Vol.9, No.3, pp.272–314 (1991).
6) Luan, S. and Gligor, V.: A fault-tolerant protocol for atomic broadcast, *IEEE Trans. Parallel and Distributed Syst.*, Vol.1, No.3, pp.271–285 (1990).
7) Peterson, L.L., Buchholdz, N.C. and Schlichting, R.D.: Preserving and using context information in interprocess communication, *ACM Trans. Comput. Syst.*, Vol.7, No.3 (1989).
8) Moser, L.E., Melliar-Smith, P.M. and Agrawala, V.: Asynchronous fault-tolerant total ordering algorithms, *SIAM J. Comput.*, Vol.22, No.4, pp.727–750 (1993).
9) Dolev, D. and Malki, D.: The design of the transis system, *Proc. dagstuhl workshop on unifying theory and practice in distributed computing* (Sept. 1995).
10) Chockler, G.V., Huleihwl, N. and Dolev, D.: An adaptive totally ordered multicast protocol that tolerates partitions, *17th ACM Annual Symposium on Principles of Distributed Computing* (1998).
11) 田 学軍，井手口哲夫，安川 博，奥田隆史：メッセージブローカサーバを用いる適応的全順序マルチキャストの提案，DICOMO'2001シンポジウム (Jun. 2001).

(平成13年 6 月 7 日受付)
(平成13年10月16日採録)

田 学軍

1991年中国天津紡績工業大学大学院修士課程修了，1998年名古屋工業大学大学院博士課程修了，博士（工学）．その後，愛知県立大学情報科学部情報システム学科助手．ネットワークアーキテクチャ，通信プロトコル，LANおよび環境電磁波の信号処理と評価等の研究に従事．電気学会会員．

井手口哲夫（正会員）

昭和24年生．昭和47年電気通信大学通信工学科卒業．同年三菱電機（株）入社．平成10年愛知県立大学情報科学部教授．工学博士．ネットワークアーキテクチャ，LAN，通信プロトコル設計方式，モバイルコンピューティング，タイムクリティカル通信等の研究に従事．著書として「コンピュータネットワーク概論」（ピアソン・エデュエーション），「分散システム入門」（近代科学社），「分散オペレーティングシステム」（科学技術出版，訳書）等．電子情報通信学会，IEEE各会員．

安川　博（正会員）

1972年静岡大学大学院工学研究科修士課程電気工学専攻修了．同年日本電信電話公社（現NTT）入社．同社研究所にて主にディジタル通信システムの研究実用化に従事．1998年愛知県立大学情報科学部教授．ディジタル信号処理の応用，情報通信，知的情報処理等に興味を持っている．IEEE，電子情報通信学会，日本音響学会，EURASIP等の会員．

奥田　隆史

1987年豊橋技術科学大学大学院修士課程修了，同年セイノー情報サービス（株）入社．88年より豊橋技術科学大学情報工学系教務職員，92年同助手，93年朝日大学経営学部講師，96年同助教授を経て，97年より愛知県立大学情報科学部地域情報科学科助教授．通信ネットワークの性能評価に関する教育研究に従事．この間，94〜95年Weber State University（米国）にて客員助教授．計測自動制御学会，電子情報通信学会，IEEE，経営情報学会，OR学会等の会員．博士（工学）．