ATMを用いた複製型共有メモリシステム

3P－2

林 博之，小口 正人，相田 仁，齊藤 忠夫
東京大学 工学部

1　はじめに

分散共有メモリには様々なタイプがある。これらはいずれも各ノード間で頻繁にデータのやりとりが行なわれ、またノード間距離として長距離まで利用するとなると、ネットワークとしては高速・高帯域のものが望まれる。一方で近年の技術進歩により伝送速度は著しく向上している上、このようなネットワークもATMネットワークの様に容易に利用できる状況にある。

本稿では、当研究室で提案している複製型共有メモリシステムの概要と、そのATM上への実装に関する研究について述べる。

2　複製型共有メモリシステムの概要

概要

本複製型共有メモリシステム[1]は、主に広域ネットワーク環境をターゲットとしている。以下にその特徴を列挙する。

- システムがソフトウェアで記述され、ハードウェアの追加なしに、マシン・通信ネットワークとも既存の標準的な環境で実現できる。
- グローバルメモリ(共有メモリ)領域がマシン内部のメモリに確保されるためアクセス時間が速い。
- セマフォ機構を用いることで release consistency をサポート。メモリの整合性は write update プロトコルで維持される。
- ロック解除要求・共有メモリの update は別CPU上でバックグラウンドで動作する送信・受信サーバによって行なわれるため、ユーザプログラムがブロックされることが少ない。

以上に挙げた機能を効率良く実現するのに不可欠であり、システム上想定しているのが高速ネットワークとマルチCPUマシンである。これらは近年の技術進歩により決して特異な環境ではないといえる。

"A Study of Replicated Shared Memory System Implimented on ATM Networks"
Hiroyuki Hayashi, Masato Oguchi, Hitoshi Aida and Tadao Saito
Faculty of Engineering, The University of Tokyo

性能評価

仮想共有メモリ、外部バス上に共有メモリを持つ複製型共有メモリ、本複製型共有メモリの3方式をSCRAMNet™上で実現し性能評価を行なった結果、ネットワーク距離に関わらず、またネットワーク距離が長いほど、他のシステムよりも性能が良いことが示されている[1]。

3　ATM上への実装

3.1　ATMを用いることの利点

本複製型共有メモリシステムでは、まず広域環境までターゲットとしているということ、また共有メモリの更新データを全て他ノードへリアルタイムに送信するということから、高速高帯域な通信路・高効率なデータ転送が必要となる。

そこでATMを利用するわけだが、ATMではLANからWANまで境界なく利用でき、光ファイバ伝送路による高速・高帯域伝送が可能であり、またハードウェアによるマルチキャスト機能をサポートしているため他ノードへのマルチキャストがスムーズ且つ効率良く行なうことができるなど、複製型共有メモリにとって多くの利点がある。

3.2　ATMスイッチのマルチキャスト

ATMを用いることの利点の一つであるATMスイッチのハードウェアマルチキャストのスループットの測定を、図1に示すテストネットワークで行なった。

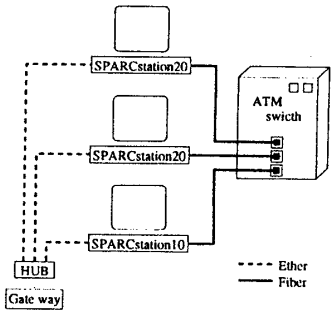

図1: テストネットワーク構成

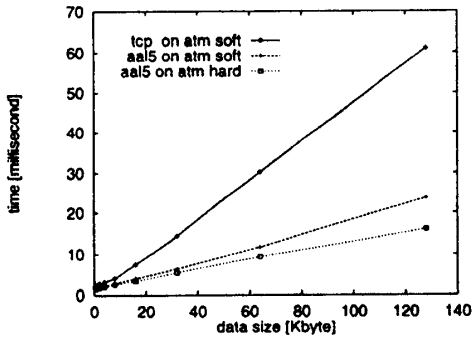

図 2: マルチキャスト通信 (1000 回の平均)

データ転送は、ATM 上で TCP を用いた場合と AAL5 を用いた場合とで行ない、2 台に対してマルチキャストし ACK を返して終了する。ハードウェアマルチキャスト以外は 2 度の送信命令を発行する。図 2 に結果を示す。

このテストでは、わずか 2 ノードでネットワーク距離も無視できる程度だが、広域・多ノードになるとこの差はより顕著になるであろう。

3.3 実装システムの評価

現在のテスト環境では、直接 AAL5 を用いた通信では信頼性が確保できないため、以下では ATM 上で TCP を用いたシステムを使用している。

データ転送時間

ユーザプログラムにとってのデータ転送時間は、送信指令を発行して送信サーバと繋がっている FIFO キューにデータをつめる間の時間となる。データ範囲だけの場合は範囲を大きくすれば、いくらでも速くすることができ、一回は数μsec で終了する。データを乗せる場合は、さらにメモリコピーの時間を要する。

FIFO バッファ

送信はユーザプログラムは転送したいデータの範囲を FIFO キューに挿入するだけで良いのだが、送信サーバでの処理の方が重いため、FIFO キューが溢れたり同じ領域のデータを連続して更新した場合に実際に送信される前に送るべきデータを書き換えてしまう、といった問題が生じる。

FIFO キューの長さは有限であるため溢れることは仕方がなく、使用可能なメモリを最大限利用するしかない。二つ目の問題の解決策として、FIFO キューにデータそのものを乗せるための領域を確保し、FIFO キューに送る際はデータの範囲のみの場合と、データ自体を格納する場合とを用意した。

それゆえ FIFO キューの長さは FIFO パケット中のデータ領域の大きさで決まる。送信サーバ内では一つの FIFO パケットに対して一つの送信命令を発行しているため、できるだけ大領域を確保すべきだが、これでは FIFO キューの長さが少なくなる上、使用効率が悪くなる。このトレードオフの条件に関して実際のシステムで測定を行なった。結果を図 3 に示す。

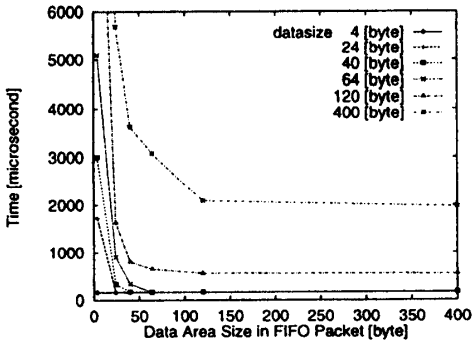

図 3: FIFO パケットの大きさによる送信速度

一回の明示的な送信指令で送られるデータ長は、かなりアプリケーションに依存するであろうが、大領域の場合は時間的余裕を持ってデータ範囲のみを FIFO キューに入れる (送信サーバ内に数 Kbyte のバッファがある) ことにすれば良いだろう。結局この結果から、ATM のセル長を考えて FIFO パケット中のデータ領域は 24〜64 バイト (送信時は 40〜80 バイト, 1〜2 セル) 程度とするのが妥当であるといえる。

4 終わりに

本稿では ATM スイッチのハードウェアマルチキャストのテストと、本複製型共有メモリシステムの ATM 上への実装について述べた。現段階の実装プログラムにも、またシステム自体のアーキテクチャに関しても改善の余地があろう。しかし、ATM を用いることで本システムの有効性がより高められており、今後は ATM の特徴をより効果的に利用したシステム作りを進めていく。

参考文献

[1] M.Oguchi, H.Aida, and T.Saito, "A Propoasl for a DSM Architecture suitable for a Widely Distributed Environment and its Evaluation", 4th IEEE International Symposium on High-Perfomance Distributed Computing, August 1995