8Q-08 PCクラスタにおける共有メモリの実装実験(3)
－UDPによる実装－

成蹊大学大学院工学研究科情報処理専攻
大橋 祐介・緑川 博子・飯塚 肇

1. はじめに

筆者らは、クラスタ型並列コンピュータシステム上に共有メモリプログラミング環境を実現する、ソフトウエア分散共有メモリシステムの設計、実装を行ってきた。すでにバリア、ロックなどの同期機構を提供する分散共有メモリシステム(sms-0.3.x)を構築し、その概要について報告した[1] [2]。

今回、通信方式を従来のTCPからUDPに変更し、通信性能の向上を目指した実装(sms-0.4.1)を試みたので報告する。

2. sms-0.4.1の特徴

sms-0.4.1は以下のような特徴を持っている。

・ユーザレベルソフトウェアによる実装
・仮想共有メモリに対するページベース管理
・同期機構としてバリア、ロック、条件変数を提供
・仮想共有メモリの一貫性を保つ更新データとしてページのコピーとの差分を採用
・通信方式としてUDPを採用
・ソケットは、全プロセス同士が結合する全二重結合

3. 性能評価

3.1 評価環境

評価は以下のようなPCクラスタ環境で行った。

CPU	Intel Celeron 400MHz
メモリ	128MB
ネットワーク	100Mbps Ether
OS	RedHat Linux 6.0
台数	1～8

表1:評価環境

3.2 基本関数の評価

図1はsms-0.4.1の提供する関数のTCP版に対するUDP版の速度向上比である。8台で実行したときの典型的な値を示す。通信プロトコルの変更により各関数とも3倍から10倍以上の性能向上を得ている。特に、システムの初期化関数startupの性能向上が大きい。これはTCP版におけるプロセスごとのconnectとacceptがUDP版では不要なためである。表2に実際の関数の典型的な実行時間を示す。

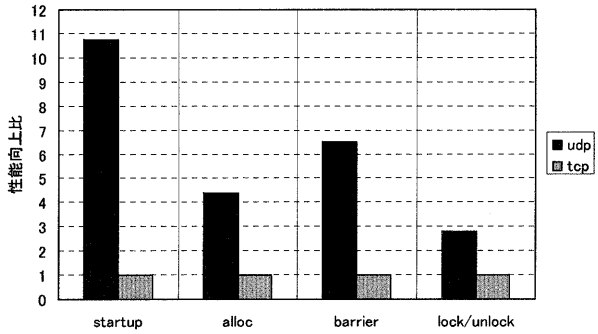

図1:基本関数の評価(8台実行時)

関数名	UDP版	TCP版
startup	0.143(sec)	1.537(sec)
alloc	0.812(msec)	3.564(msec)
barrier	1.794(msec)	11.718(msec)
lock/unlock	8.733(msec)	24.318(msec)

表2:関数の実行時間

3.3 応用プログラムによる評価

3.3.1多体問題による評価

この問題は宇宙の星の動きを万有引力の法則を用いて計算する問題で、計算量は星の数n個に対して$O(n^2)$である。星の集合を各プロセスで静的に分割し、並列処理を行うと、通信量はO(n)となる。それぞれのプロセスが受け持つ計算量は均等で、並列化効率のよい問題である。

星の数が100個、1000個におけるTCP版に対する性能向上比を図2、図3に示す。問題の性質上、星の数が多くなるほど並列化の効果が高く、100個程度の処理では、TCP版では並列化効果は得られない。しかしUDP版では1.8程度の性能向上比が得ら

UDP Implementation of Shared Memory Systenl on PC Cluster
Yusuke Ohashi, Hiroko Midorikawa, Hajime Iizuka
Department of Information Sciences, Seikei University

れている。UDP版の並列性能向上比は、8プロセスの場合、星の数1000個で6.7、2000個で8.0と高い。また実行時間は、TCP版に対し、星の数100個で31%、1000個で81%に短縮化されている。

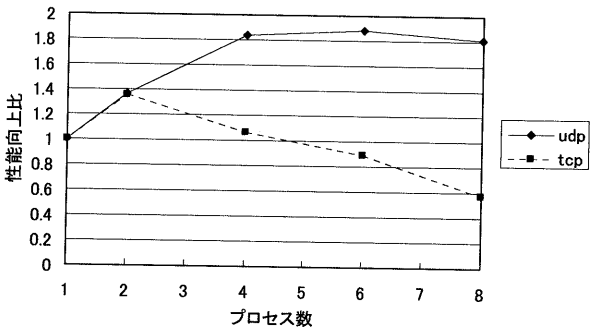

図2:多体問題の性能評価(星の数100)

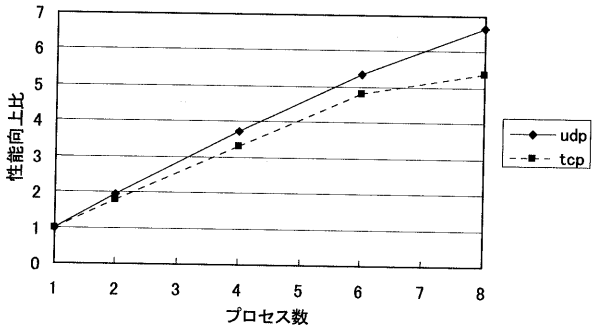

図3:多体問題の性能評価(星の数1000)

3.3.2 マンデルブロー集合問題による評価

マンデルブロー集合とは、$z_0=0$, $z_{n+1}=z_n^2-c$ (n=0,1, …)で定義される複素数の数列$\{z_n\}$が有界である(n→∞で$|z_n|$が発散しない)ような複素数cの集合のことである(図4)。この問題は計算量が領域によって異なるため、全領域を小領域に分割し、プロセスに動的に割り当てることによってプロセス間の負荷の均一化を図っている。前述の多体問題がバリアのみを用いた同期的な処理であるのに対し、この問題はロックを多用した非同期な通信が発生するのが特徴である。

図4:マンデルブロー集合

並列化性能向上比を図5に示す。8プロセスの場合の並列速度向上比は、TCP版の4.9から6.7へと向上した。また、実行時間は、72%に短縮化した。

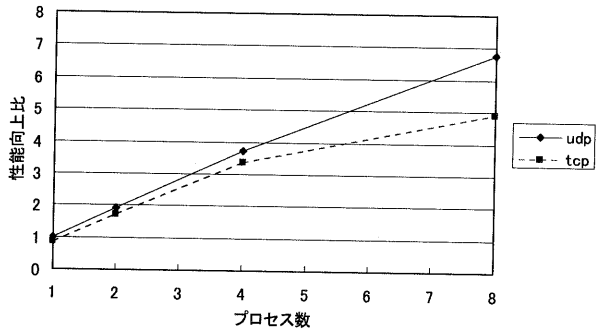

図5:マンデルブロー集合による評価
(0.3<x<0.4, 0.5<y<0.6, 4096*4096, 分割数512)

3.3.3 FFTによる評価

各次元のサイズが64の3次元高速フーリエ変換を行った場合の速度性能向上比を図6に示す。TCP版では並列効果が全く得られなかったが、UDP版では逐次処理に比べ若干の速度向上が見られた。

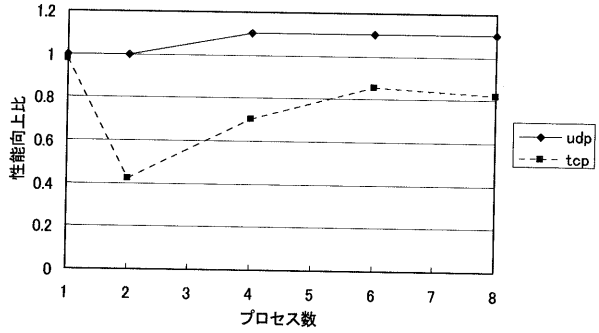

図6:FFTによる評価

4. おわりに

分散共有システムsmsにおけるUDP実装の効果について述べた。今回の通信プロトコルでは、通信バッファがあふれることのないようなsmsの通信手順を用いることにより、流量制御や再送などの手順を含まない単純なUDP実装を行った。PCクラスタとスイッチハブなどのように比較的近距離で、安定したネットワークを用いる場合、単純なUDP通信であっても十分実用に耐える通信が可能で、通信性能が向上することがわかった。

参考文献

[1]緑川, 伊藤, 大橋, 飯塚:"PCクラスタにおけるユーザレベルソフトウエア分散共有メモリSMS",並列処理シンポジウムJSPP '00 論文集 p.165 (2000,5)

[2]伊藤, 緑川, 飯塚:"PCクラスタにおける共有メモリの実装(2)ーロックの実装ー",情処全国大会 2H-2(2000,3)