

分散共有メモリサーバにおけるページ配送の効率化

1 B-1

峠田 睦 斎藤 彰一 大久保 英嗣
立命館大学工学部情報学科

1 はじめに

主記憶の容量を越える大きさのデータを処理する方法として処理分割やデータ分割といった方法がある。複数のマシンをネットワークを介して接続した環境においてこれらの処理を行うためには、ユーザがデータの位置管理やマシン間の移送を意識しなければならない。これらを意識させない機構を実現するために分散共有メモリ(DSM:Distributed Shared Memory)を用いることが考えられる [1].

DSMを使用することで、主記憶の容量を越える大きさのデータを処理できる。また、DSMではメモリ空間を共有することができるため、複数のマシン間におけるネットワーク通信プログラムの作成が容易となる [2].

本稿では、分散共有メモリサーバ(DSMサーバ)を利用するクライアントプログラムの高速化のための、ページ配送の効率化について述べる。

DSMサーバにおいてページ配送コストを削減する方法として「ページングメモリの使用」と「マルチページング」が考えられる。「ページングメモリの使用」とは、ページングのための領域(ページングオブジェクト)としてUNIXファイルではなく主記憶を使用する方法である。「マルチページング」とは、プログラム毎に適切なページサイズに変更してメモリアクセスを行う方法である。本稿では、このうちマルチページング方式について述べる。

2 マルチページングの利点

マルチページングとは、ページ要求時に当該ページを含む複数ページの同時配送を行うことを意味する(図1参照)。マルチページングを行うことによって、要求されたページがローカルサイトに存在する割合を増加させる

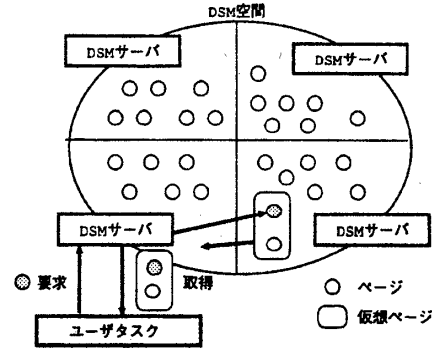


図1: マルチページング

ことが可能となる。したがって、クライアントタスクにおける処理の待ち時間が短縮されることになる。

また、リモートサイトにページ要求を行う割合を減少させることで、そのための制御メッセージの数も減少させることができる。このことによって、DSMサーバ自身の負荷を軽減することができる。

3 マルチページングの実現法

マルチページングの実現法として、カーネルインタフェースによる方法と仮想ページによる方法が考えられる。

カーネルインタフェースによる方法とは、マイクロカーネルとメモリマネージャとのインタフェースである外部メモリ管理インタフェースや既存のメモリアクセス関数を使用せず新しくこれに相当するインタフェースを作成し、ページサイズの変更を可能にする方法である。この方法では、カーネルやハードウェアに依存する部分で実現することになる。この作業にはカーネルの変更が必要なため実現が困難である。

仮想ページによる方法とは、連続したアドレスを持つページの集合を一つの仮想的なページとみなす方法である。仮想ページサイズの変更は仮想ページに含まれるページサイズの倍数単位で行う。この方法を用いると、各ページ単位で送られて来るカーネルからの要求を、仮

想ページの単位に変換する必要がある。したがって、仮想ページ内のページの一貫性制御が困難になる。

しかし、カーネルインタフェースの変更による方法よりも実現が容易である。さらに、基本となるページの管理は従来そのままなので、連続したページを配送するだけでなく、連続でないページの集合を配送するという応用も考えられる。以上のことから我々は仮想ページによる方法を採用した。

4 ページ情報の管理

前章で述べたように、カーネルとのインタフェースを変更しない方法を採用しているため、ページフォルトは各ページ単位に発生する。この場合には、当該ページが所属する仮想ページの情報にしたがって、一貫性制御などの処理を行わなければならない。以下では、仮想ページを DSM ページグループと呼ぶことにする。

DSM ページグループサイズの動的変更を容易に行えるように、ページ毎の情報は各ページが持つ。また、マルチページ化の情報である DSM ページグループサイズは各共有オブジェクトが持つ。このことによって、同一の DSM サーバを使用しながら、ユーザタスクにおいて最適と思われる DSM ページグループサイズで各共有メモリオブジェクトにアクセスすることが可能となる。

5 評価

評価は、CPUi486DX2(66MHz)、外部キャッシュ256キロバイト、主記憶16メガバイト、SCSI HDD から構成されるマシン2台を Ethernet で接続したシステムで行った。なお、1ページは4キロバイトとなっている。評価用プログラムは、double型(8バイト)のデータをソートするプログラムである。このプログラムでは、まずデータをマシン台数分に分割し各マシンでクイックソートによるソートを行う。その結果をさらに、各マシンでマージを行う。測定結果を図2に示す。

DSM ページグループサイズが通常の2倍の大きさ(8192バイト)の場合には、DSM ページグループサイズが通常の大きさのときより処理時間が増加する場合がある。これは、DSM サーバにおける処理のオーバーヘッドがプリフェッチングの効果を上回っていると考えられる。一方、ページサイズが通常の4倍および6倍のときは、処理時間が大幅に削減されていることが分かる。評価プログラムでは、分割後のページ参照は連続的になる場合が多い。また、分割して各マシンで処理を行うため、他のマシンからの参照が行われない。このためにマルチペー

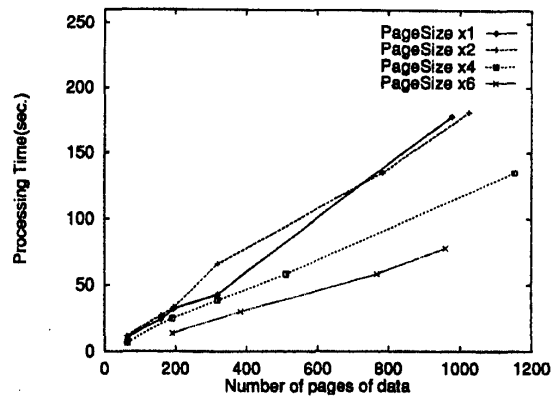


図2: DSM ページグループサイズと処理時間

ジングによるプリフェッチの効果が大きく出ていると考えられる。このことから、ページへのアクセスが連続的である割合が高く、他からの書き込みが少ない場合にマルチページングは有効であることが分かる。

6 おわりに

今回の評価結果より、連続するページへのアクセスの場合には、マルチページングによるプリフェッチは有効であることが分かった。しかし、このようなアクセスの逐次性が常に存在するとは限らない。このため、DSM ページグループサイズを大きくすると無駄なデータの転送を引き起こし、かえって処理効率を低下させることになる。したがって、動的に DSM ページグループサイズを変化させ、無駄な転送を引き起こさないようにする必要があると考えられる。これらの処理を行うことで、分散処理に DSM を利用した形態がより有効なものとなることが予想される。

参考文献

- [1] 斎藤 彰一, 中村素典, 大久保英嗣, 大野豊, 白川洋充: Mach の外部ページャを利用した分散共有メモリサーバの評価, 情報処理学会研究会報告 94-OS-65, Vol.94, No.64, pp.145-152(1994).
- [2] 斎藤 彰一, 中村素典, 大久保英嗣, 大野豊: 分散共有メモリサーバの大規模データ処理への適用と評価, 情報処理学会研究会報告 95-HPC-55, Vol.95, No.28, pp.25-32(1995).
- [3] 乾 和志, 菅原 圭資: 分散 OS Mach がわかる本, 日刊工業新聞社 (1992).