情報処理学会第40回(平成2年前期)全国大会

可変構造型並列計算機のオペレーティング・システム ーメモリ管理ー

6G－1

草野和寛　福澤祐二　福田晃　村上和彰　富田眞治
(九州大学大学院総合理工学研究科)

1. はじめに

現在我々が開発中の「可変構造型並列計算機」は，128台のプロセッシング・エレメント（PE）を相互結合網（クロス・バー網）によって結合したマルチプロセッサ・システムである[1]．本システムは，密／疎結合型マルチプロセッサのいずれの形態も実現できる．現在，密結合型マルチプロセッサと見た，並列オペレーティング・システム（OS）を開発している[2]．本稿では，このOSにおけるメモリ管理の概要について述べる．

2. メモリ・アーキテクチャ

2.1 概要

可変構造型並列計算機では，物理メモリを128台の各PEに分散させるメモリ構成を採用している[3]．この分散メモリの共有／私有をページ単位に制御するために，3つのアドレス空間（仮想アドレス空間，実アドレス空間，物理アドレス空間）を導入している．実アドレス空間はプライベート空間とコモン空間の2つから構成されている．プライベート空間は各PEの私有領域であり，そのまま各PEの物理メモリにアクセスを行う．コモン空間は，全PEに共通の領域であり，各PEに対応した256個の共有メモリ・ウィンドウ（SMW）に分割されている（現在，128個は未使用）．各PEはこのウィンドウ内にアクセスすることで，他PEの物理メモリへネットワークを介してアクセスすることが可能となっている．仮想アドレス空間から実アドレス空間へのマッピングを様々に変化させることによって物理メモリの共有／私有を制御している．

2.2 アドレス変換

仮想アドレス空間から実アドレス空間への変換では，セクション・テーブルとページ・テーブルを使用する．この変換は，MMU（Memory Management Unit）によって行われる．変換されたアドレスがコモン空間の場合，目的のPEに相互結合網を通してアクセスを行う．各PEではSMWによる他PEからのアクセスに対して，SMW変換テーブルを用いて物理アドレスへの変換を行う．この変換はハードウェアが行う．

3. メモリ管理の概要

仮想アドレスをコモン空間（SMW）へマッピングすることによって，全PEの物理メモリをネットワークを介してアクセスが可能になることを用いて，密結合型マルチプロセッサを実現できる．この時にメモリ管理では，主に以下のことを行う必要がある．

① ユーザに仮想空間を提供するための変換テーブルの管理．
② 物理メモリのページの確保，ページイン／ページアウト等の操作に伴うTLBの一貫性の保証．
③ 二次記憶領域の管理．

現在，可変構造型並列計算機にはディスクは備えられておらず，ファイル・システムはホストであるSUN－4のものを利用する予定である．

次章では，可変構造型並列計算機で密結合を実現する場合の，実アドレス空間におけるコモン空間の使用方法，および共有メモリ・ウィンドウ（SMW）の管理について述べる．

4. SMWの管理

4.1 SMWエントリの使用法

本システムにおける，論理的な共有メモリ・ウィンドウ（SMW）の大きさ（8MB）のうち，実際に使用可能である領域（4MB）は物理メモリと同じ大きさがある．したがって，物理メモリを全て共有することも可能であるが，物理メモリには各PE固有の領域（物理メモリの管理テーブルやカーネルのスタック等）も存在する．このPE固有の領域はSMWエントリを使用しないので，このSMWエントリを利用することが考えられる．これを利用することで，仮想空間に対して割り当てる実アドレス空間のSMWの大きさを，実際に共有している物理メモリにより制限されないようにすることができる．この利点として，次のようなことが考えられる．共有可能な物理メモリを超える物理メモリのページ要求に対して，未使用のSMWエントリを割り当てる．この時，ページアウトした物理メモリに対応していたSMWエントリ内のValidビットをインバリッドにしておくだけでよい．このようにすると，他PEと共有可能な物理メモリを超える割り当て要求の度にTLBを無効化する必要がない．TLBの無効化は，SMWエントリが全て使用中の時に，新たにエントリの内容を入れ換える場合のみでよい．以上に述べたことから，使用可能な全てのSMWエントリを，仮想空間へ割り当てる実アドレスとして使用する．

4.2 SMWの管理概要

SMWのエントリの割り当てと解放を行う方法としては，大きく以下の2つに分けられる．

① 仮想空間へのSMWエントリの割り当ては，物理メモリの割り当て時に行い，そのSMWエントリの解放は割り当てられ

An Operating System for the Reconfigurable Parallel Processor －Memory Management－
Kazuhiro KUSANO, Yuji FUKUZAWA, Akira FUKUDA, Kazuaki MURAKAMI, and Shinji TOMITA
Kyushu University

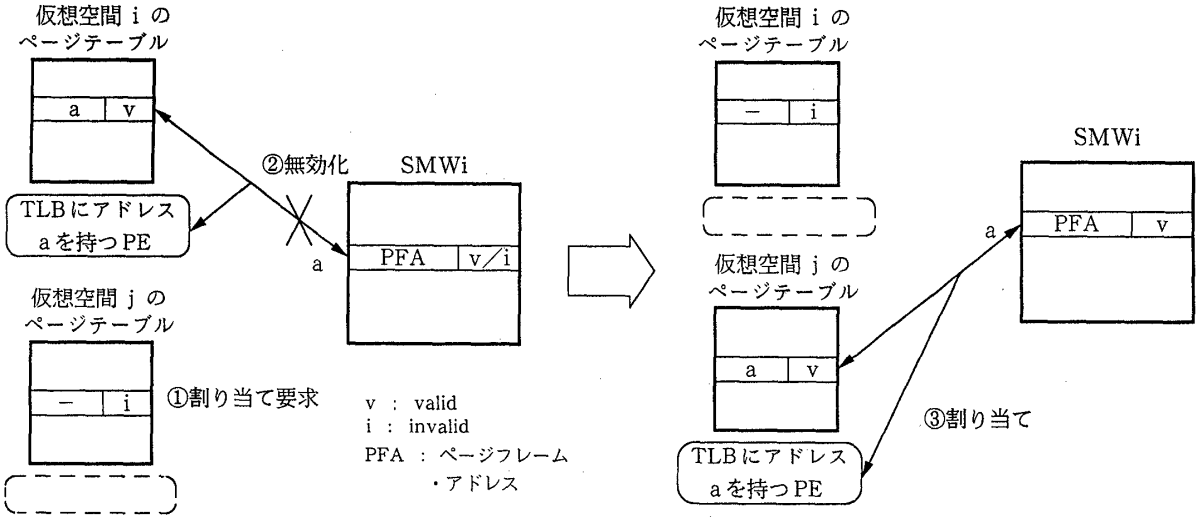

図1. SMWエントリの解放と割り当て

た仮想空間の消滅時にのみ行う. この方法では, SMWの全エントリを割り当て済みの状態で新たな要求があった場合には, その要求に対して対応できない. この場合は, エントリに空きがある他のPEで対応するようにすれば良いが, 全PEの全SMWエントリが使用されていれば結局対応できないことになる.

② ①と同様にSMWエントリの割り当てを行い, そのSMWエントリの解放は仮想空間が消滅する時以外にも行う. つまり, SMWエントリが全て割り当てられている時に新たに要求があった場合は, あるアルゴリズムによって選択されたエントリを解放し, 要求に対応する. この場合はSMWエントリを解放するために, そのエントリが割り当てられている仮想アドレスへの逆変換テーブルが必要になる. エントリの解放を行う時には, 対応する物理メモリを解放する必要があり, さらにそのエントリを参照可能である全てのPEに対して, そのエントリのTLBの無効化を通知しなければならない (図1). このエントリの無効化方法では, SMWのエントリとページ・テーブルの内容および, 参照可能PEのTLBの内容の全ての一貫性を保証する必要がある. このためのアルゴリズムは, 現在検討中である.

①の方法では, 他PEのメモリを共有メモリとして使用できる上限がSMWのエントリ数で決定されることになる. これは作成中である密結合型のシステムを作る上において大きな制約となる. これに対して②の方法では, このような制約はないので, ①の方法よりも柔軟性は高いと考えられる. したがって, 我々はSMWエントリの割り当てと解放の方法として②の方法を採用する.

②の方法を採用した場合に必要となるTLBの無効化を通知するPEの決定には, 以下の2つの方法が考えられる.

① 全PEに無効化を通知する方法.
この場合には, ブロードキャストを用いるにしても, 全PEに対して無効化を通知するオーバヘッドが問題になる.

② PEを限定して無効化を通知する方法.
SMWエントリを参照するのはそのエントリが割り当てられた仮想空間を実行したPEに限られるので, それらに限定することで, 無効化を通知する必要のあるPEの数を減らすことができる. この方法では, プロセス管理やスケジューラから, それぞれの仮想空間が実行されたPEについての情報を得て, 参照可能PEを動的に管理する必要がある. この管理情報と実際の状況の不一致等があると致命的なエラーにつながる可能性があるので, その一貫性は常に保証しておく必要がある. この管理を, ソフトウェアのみで行うことも可能であるが, 制御も複雑になり, かなりのオーバヘッドを伴う. また, 可変構造型並列計算機で提供している, SMWのエントリ単位に参照可能PEを設定できる機能を用いると, この管理はソフトウェアのみで行う場合と比べて比較的容易に実現できると考える.

②の方法を用いる場合, 無効化するPEの数がある程度増えると①の方法よりも遅くなる可能性が高い. これは, ①ではブロードキャスト機能を用いて無効化を通知するのに対して, ②では各PEに1つずつ無効化を通知する必要があるためである. したがって, あるしきい値により2つの方法を使い分けることがよいと考える. このしきい値を決定することは, これからの検討課題である.

5. おわりに

以上, 現在開発中である「可変構造型並列計算機」のOSのメモリ管理の概要を述べた. 今後は, 本稿で述べたことの詳細化を進め, 早期の実現を目指す.

参考文献

[1] K.Murakami et al.: The Kyusyu University Reconfigurable Parallel Processor −Design Philosophy and Architecture− , Proc. IFIP 11th World Computer Congress , pp.995−1000 (1989).

[2] 福田ほか: 可変構造型並列計算機の並列／分散オペレーティング・システム, 情報処理学会研究報告, 89−OS−43 (1989).

[3] 蒲池ほか: 可変構造型並列計算機のメモリ・アーキテクチャ, 情報処理学会第38回全国大会講演論文集, 3T−2 (1989).