

並列処理用OS SKY-1

— 開発構想 —

3P-1

山口 伸一郎 上脇 正 齋藤 雅彦 小林 芳樹 坂東 忠秋
(株) 日立製作所 日立研究所

1. はじめに

我々は、汎用マイクロプロセッサを用いた小規模並列処理用の密結合型マルチプロセッサ (TCMP :Tightly Coupled Multiprocessor) OS・SKY-1 (System Kernel for You -1)を開発中である。本稿では、その開発構想及びハード・ソフトの概要を報告する。

2. 開発構想

2.1 開発の動機と目的

高度ユーザインターフェイス・オンライン処理・CAD・AI等の、一般的でありながらプロセッサ消費型のアプリケーションが多くなり、シングルプロセッサで対応できる限界が見えてくるにつれて、汎用的に使える並列処理用計算機へのニーズが高まってきた。そして、密結合型マルチプロセッサによる並列処理がこれを解決する手段として期待されている。

一方、計算機を支えるマイクロデバイス技術の進展には目覚ましいものがある。10Mトランジスタ/チップ、64MDRAMも遠い話ではなく、100MIPS以上のシングルチップ・マルチプロセッサも実現可能領域となる。そこで我々は、図1に示すような超大容量メモリを持ち、超高速シングルチップ・マルチプロセッサからなるTCMPを想定して、この実現に向けての研究に着手した。

2.2 技術課題

図1のTCMPで並列処理を実現するには、ハードの技術開発もさることながら、システムソフトウェアの面で解決すべき課題が多い。

並列の粒度：

並列処理での問題は、並列の粒度をどれくらいに設定するかである。細かい粒度の大規模並列処理は、高い性能を引き出す可能性があるものの、汎用性に乏しく、またプログラミングも今までと違ったスタイルが要求されるため、一般のユーザが使いこなすのは難しい。また、プロセス単位の粗い粒度の並列処理は、一つの処理を細分化した時に、プロセス間のアイソレーションが却って弊害となり、オーバーヘッドの増大を招く。

これに対して、Mach [1]などで用いられている軽量なプロセス(スレッド)は、オーバーヘッドの少な

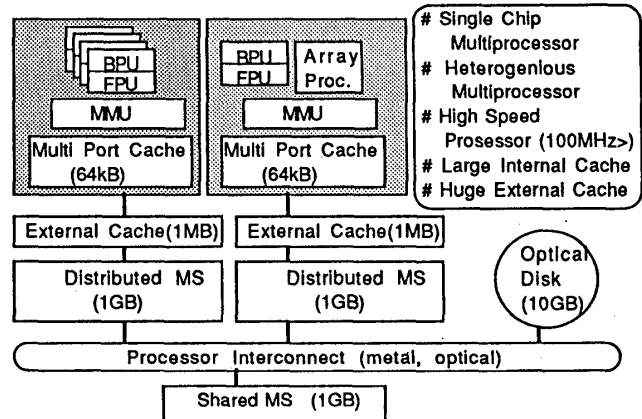


Fig. 1 Super Multiprocessor

さから、中程度の粗さの並列処理に適している。かつ要求されるプログラミングも現行のものとは大きな違いがなく、汎用の並列処理に対して有効なモデルと考えられる。しかしスレッドを計算機制御やOLTP等の厳しい環境で使うには、信頼性や使い勝手の点で問題がある。

異種プロセッサ接続：

汎用のプロセッサだけで高性能を維持するのが難しく、特化プロセッサのいくつかをシステム内に含むことは必須である。そこで同種プロセッサのTCMPだけでなく、異種・異種プロセッサを透過にする仕掛けを考える必要がある。

ハードウェア活用：

ハードウェア性能を最大限に引き出すというのが、OSの重要課題である。このためにはキャッシュやTLBといったシステム性能に大きく影響するリソースの管理をハードウェア任せでなく、ソフトウェアで制御して効率をあげる必要がある。

2.3 開発のアプローチ

以上の課題は、いずれも実用的な高性能TCMPに必須の技術であり、その効果を知るには実システムで評価することが最良である。

そこで我々はこれらに対して検討を行い、これを盛り込んだ並列処理用OS・SKY-1をハードウェアに実装してTCMPシステムを構築することにした。ハードウェアは一般的なTCMPであり、100MB/sの共

有バスに、モトローラ社のマイクロプロセッサ68030の33MHz版を4台と64MBの共有メモリを接続する。各プロセッサは、内蔵の命令キャッシュと外部に256kBの大容量ライトスルーキャッシュを持つ。

3. 並列処理用OS・SKY-1の概要

SKY-1はUNIX互換であり、これにマルチプロセッサ支援及びスレッドを並列処理の単位とした並列処理支援機能を付加したOSである。SKY-1の特徴は次の点にある。

- ・リアルタイムスレッド機能
- ・スレッド用メモリ管理
- ・プロセッサ・アフィニティ制御
- ・高効率スレッドスケジューリング
- ・並列処理用C言語ライブラリ

(1) リアルタイムスレッド機能

スレッドは、Machで用いられているのと基本的に同等であるが、そのままではリアルタイム性に乏しい。これを補うために、ユーザがスケジューリングを制御するシステムコールも準備されているが、返ってユーザの負担を増すことになりかねない。そこでSKY-1では、OLTPや計算機制御にスレッドを使用する事を考えて、スレッド単位で固定優先度スケジューリングやプロセッサ先取り機能をカーネル自体に設ける事で、リアルタイム性を向上させた。

(2) スレッド用メモリ管理

スレッドは、固有のスタック領域を同一空間内に持ち、コンテキスト・スイッチを高速化しているが、その代償としてハードウェアによるスタック間の保護が失われてしまった。そこでスタック伸長/縮退時に、他のスタック領域を侵していないかチェックする新しいスタック管理方式[2]を開発した。本方式は、一般的な仮想メモリ管理ハードウェアに変更を加える事無く、コンパイラとOSの連携だけでUNIXのプロセス間のスタック保護と同程度の保護が可能である。

(3) プロセッサ・アフィニティ制御

TCMPが同種プロセッサ以外で構成されるものとして、例えば信頼性をあげるために特別に設計したプロセッサとか、AI・アレイ計算用のコプロセッサを持つ特化プロセッサが、TCMP内に存在する場合が考えられる。そしてこの場合、スレッド側にもその性質によって、特化プロセッサを割当てる必要性が生ずる。

そこで、プロセッサとスレッドに属性を付加して、スレッドの属性がプロセッサのそれを含むときに限って、そのプロセッサをスレッドに割当てるようにして、ユーザが特別意識しなくても異種TCMPに対応できるように考慮した。

(4) 高効率スレッドスケジューリング

現行のプロセスのスケジューリングは、主メモリ上の状態(例えばプロセスがスワップアウトされているか否か)までは考慮しているが、性能に大きく影響するキャッシュやTLBの状態を考慮していなかった。このためプロセッサ間のプロセスマイグレーションによってキャッシュヒット率が低下し、性能が悪化すると予想される。

そこでSKY-1では、キャッシュの状態を考慮し、スレッドに関するデータを最も多く持つプロセッサをそのスレッドに割当てる等の新しいスケジューリング[3]を開発した。

(5) 並列処理用C言語ライブラリ

並列処理では、同じ機能のスレッドを複数個使うことが多く、またこれらをまとめて制御する必要性が大きい。SKY-1では、スレッドグループの考えを導入し、複数のスレッドを一度に生成/終了させたり、図2に示すようにグループにメッセージを送りいづれかのスレッドから応答を貰う等の68関数をC言語ライブラリとして設けて、スレッドの使い勝手を向上させている。

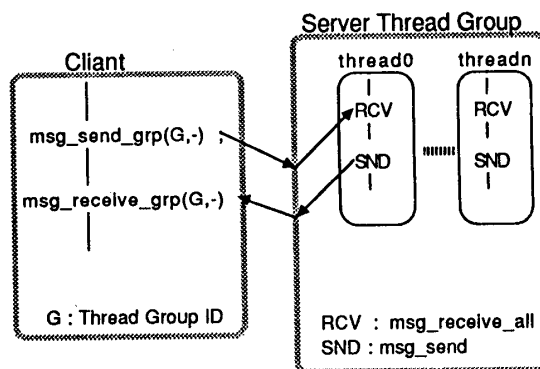


Fig.2 Message To/From Thread Group

4. おわりに

以上、開発構想及びハードウェアと並列処理用OS・SKY-1の概要を述べた。今後並列処理の環境の下でアプリケーション開発を行い、並列処理のデータ収集を行って行きたい。

参考文献

- [1] R.Rashid, et al.: "Mach: A New Kernel Foundation For UNIX Development", proc. of USENIX '86 summer
- [2] 齋藤ほか, "並列処理用OS・SKY-1のメモリ管理方式", 情報処理学会第39回全国大会
- [3] 上脇ほか, "並列処理用OS・SKY-1のスケジューリング方式", 情報処理学会第39回全国大会