3P-3

並列処理用ＯＳ　ＳＫＹ－１のスケジューリング方式

上脇　正　　山口　伸一朗　　齊藤　雅彦
日立製作所　日立研究所

１．はじめに

密結合マルチプロセッサ（ＴＣＭＰ）の多くは、各プロセッサがキャッシュメモリを持っており、このキャッシュメモリが、システムの性能上非常に重要な役割を果たす。

しかし、スレッドのスケジューリングの方法によっては、キャッシュメモリの性能が大きく低下してしまう問題がある。例えば、スレッドがプロセッサを移動すると、今まで蓄えられたキャッシュデータが無駄になり、キャッシュミスを連続しておこす[1)]。

また、キャッシュメモリがバスをスヌープしていて自分が記憶しているデータに対する書き込みを検出すると、そのデータを無効化するＴＣＭＰでは、共有データを持つスレッド同志を異なるプロセッサで実行すると、キャッシュメモリの無効化が頻発する問題がある[2)]。

そこで、我々が試作しているＴＣＭＰ用ＯＳ・ＳＫＹ－１(System Kernel for You)[3)]では、これらの問題を解決する手法をスレッドスケジューリング方式に組み込んでいる。

２．ＳＫＹ－１スケジューリング

ＳＫＹ－１スケジューリングの戦略は３つある。

2.1 同一プロセッサ優先

同一プロセッサ優先は、タイムスライスの終了、イベント待ちの発生等により一度中断されたスレッドを再開するときには、可能な限り前回実行されていたプロセッサで実行する方式である。この方式によれば、スレッドの不必要なプロセッサ間移動がなくなり、移動によるキャッシュミスが減少する。

2.2 スレッドのグループ化

スレッドのグループ化は、図１のようにスレッドをグループとして管理し、同一グループに属するスレッドを同一のプロセッサで実行するものである。スレッドT_0、T_1はG_0に属し、プロセッサP_0で実行される。

先に述べたようにデータを多く共有するスレッドを異なるプロセッサで実行すると、キャッシュ一致化のための無効化が頻発する。共有データの多いスレッド同士をグループにして、同一プロセッサで実行すれば、共有データが複数のキャッシュメモリに入ることが少なくなる。図１ではG_0のデータはキャッシュC_0に集中する。するとキャッシュ一致化のための無効化が減少、キャッシュヒット率が向上する。

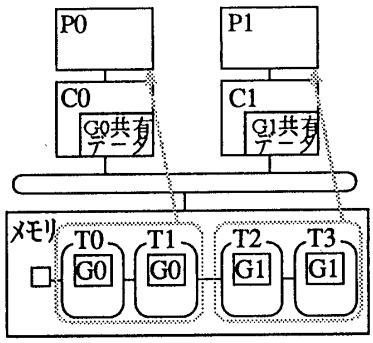

図１　スレッドグループ化

スレッドグループ化には、他の使いかたもある。同一タスクに属するスレッドをグループ化すると、スレッドスイッチのオーバーヘッドが減少する。つまり、タスクの異なるスレッド同志のスイッチは、同一タスクのスレッドのものに比べて遅い。同一タスクに属するスレッドをグループ化すると、スレッドスイッチが同一タスクのもの同志であることが多くなり、スレッドスイッチのオーバーヘッドが減少する。

また、資源を共有するスレッドをグループ化すると、排他制御を簡単化できる。ある資源を使用する全てのスレッドが一つのプロセッサのみで実行されることが判っている場合、スレッドスイッチと割り込みを禁止すれば、資源を使用するときにロックする必要がない。ロック処理は時間がかかるため、処理速度向上に役立つ。また、資源の排他制御を割り込み禁止のみで行っているシングルプロセッサ用ソフトウェアも、この機能によれば、改造なしにマルチプロセッサで実行可能となる。

2.3 負荷分散

同一プロセッサ優先とスレッドのグループ化を行っていると、あるプロセッサは処理するスレッドが沢山あるにも拘らず、アイドルとなるプロセッサが出来ることがある。そこで、プロセッサ間での負荷を平均化するため

Multiprocessor Operating System:SKY-1 -Thread Scheduling-
Tadashi KAMIWAKI, Shinichiro YAMAGUCHI, Masahiko SAITOH
Hitachi Research Lab., Hitachi Ltd.

に、一定時間実行されていないスレッドについては、プロセッサの移動を認めた（移動を禁止することも可能）。

実行されていないスレッドのデータは、徐々に他のスレッドのデータによりキャッシュメモリから追い出される。長く実行されていないスレッドは、キャッシュメモリにデータが少なく、移動による性能への影響が少ない。

3．性能評価

ＳＫＹ－１スケジューリング方式の性能を評価するために、図２のモデルを用いて解析的シミュレーションを行った。モデルには、68030(33MHz)を４台接続したマルチプロセッサを用いた。スレッドは、平均２５０ｍｓに１個生成され、発生したスレッドは実行待ち行列に入り、スケジューリングされるのを待つ。スケジューラがスケジューリングアルゴリズムに従い、実行待ち行列のスレッドをプロセッサに割り振る。タイムスライスは１ｓで、それを使いきったスレッドは次のスレッドとスイッチして実行待ち行列の最後に戻される。Ｉ／Ｏアクセスが発生したスレッドもスイッチを行う。キャッシュミス：ヒットのアクセス時間比は１０：１とした。

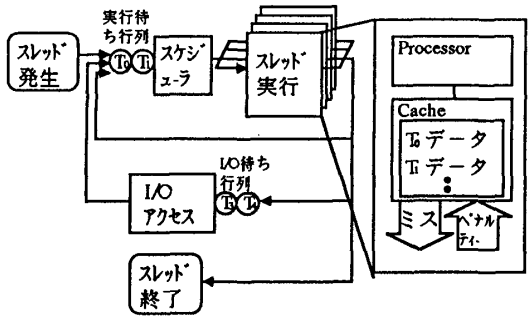

図２ シミュレーションモデル

ＳＫＹ－１方式を、スレッドスイッチの必要となったプロセッサが、単純に実行待ち行列の先頭から、スレッドを取り出して実行する従来方式と比較したところ以下のような結果が得られた。

(1) キャッシュミス率のキャッシュ容量依存性

図３はキャッシュメモリの容量を変化させたときのキャッシュミス率の変化を表している。キャッシュ容量が１２８ｋＢ時、ＳＫＹ－１方式は、キャッシュミス率が４５％（同一プロセッサ優先により１５％、スレッドグループ化により３０％）低減する。これは１３％のプロセッサ性能向上に相当する。また、キャッシュ容量が２ＭＢと多くなると、キャッシュミス率の低減も６５％と大きくなる。

(2) キャッシュミス率の共有データ量依存性

図４は、スレッド間の共有データ量（あるスレッドのライトアクセスが、他のスレッドのキャッシュデータを無効化する確率）とキャッシュミス率の関係のグラフで

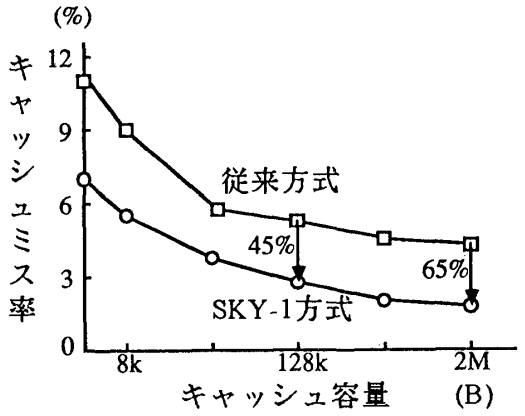

図３ キャッシュ容量依存性

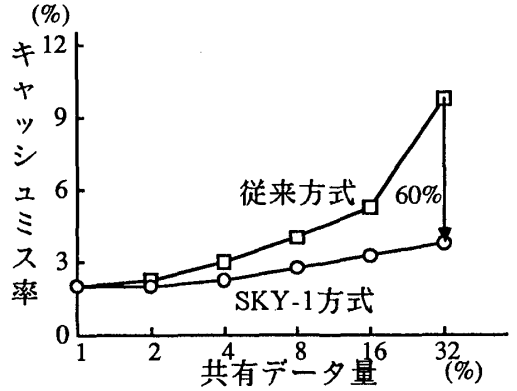

図４ 共有データ量依存性

ある。共有データ量が増加すると、従来方式はキャッシュミス率が急増するが、ＳＫＹ－１方式ではほとんど変化がない。共有データ量が３２％の時、ＳＫＹ－１方式によってミス率が６０％減少し、プロセッサ性能が２６％向上する。プロセスに代わりスレッドを使用するようになると、共有データ量は必然的に多くなるため、本方式が重要となってくる。

4．おわりに

並列処理ＯＳ・ＳＫＹ－１のスケジューリング方式について述べた。本方式の特徴は、同一プロセッサ優先とスレッドグループ化によりキャッシュメモリの性能を十分に引き出すことにある。

5．参考文献

1) Sites, R. L., et al.:Multiprocessor Cache Analysis Using ATUM: Proc.ISCA(1988)

2) Eggers, S. J., et al.:The Effect of Sharing on the Cache and Bus Performance of Parallel Programs: Proc.ASPLOS III(1989)

3) 山口　他：並列処理用ＯＳ　ＳＫＹ－１　－開発構想－：情報処理学会第39回全国大会(1989)

4) 齊籐　他：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(1989)