

5X-8 ブロードキャストメモリ方式による分散処理システムのセマフォとその管理

黒岩 実^{*1}、 呉 江荔^{*1}、 馬場 光弘^{*2}、 相田 仁^{*1}、 齊藤 忠夫^{*1}
 *1 東京大学 工学部、 *2 ㈱日立製作所

1. はじめに

当研究室は分散処理システムの形態として、汎用のワークステーションを用いた、ブロードキャストメモリに基づくシステムを提案している^[1]。ブロードキャストメモリはシステム内の各プロセッサにバスを通じて接続されており、プロセッサがデータを書き込むと、そのデータを他のすべてのブロードキャストメモリに結合バスを通じて放送し、それぞれのメモリで同じ書き込みを実行することにより、それぞれメモリの内容を同一に保つものである。

本稿では、現在製作中である実験システムのハードウェアとソフトウェアの構成、およびシステムの特徴について述べる。

2. ハードウェアの構成

本システムは、プロセッサとしてのワークステーションと共有メモリ（実際はその内容が共有される）、ブロードキャストを制御するブロードキャストコントローラ、排他制御を行なうためのセマフォコントローラ、それぞれのブロードキャストコントローラおよびセマフォコントローラを接続する結合バスから構成される（図1）。以下、それらについて順に説明する。

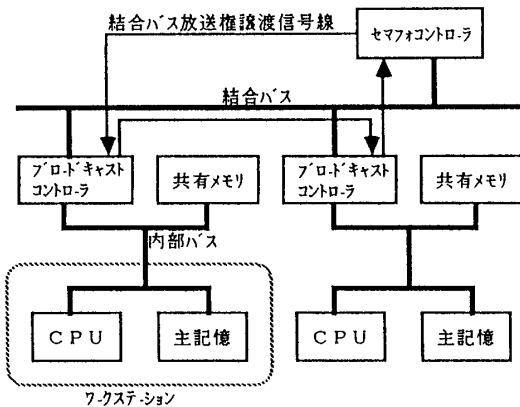


図1 システム構成

共有メモリ：

ワークステーションの外部に8Mbyte（可変）配置され、内部のローカルメモリとは区別される。また、先頭にセマフォ操作のための領域が8byte設けられており、その領域に対するアクセスは通常のブロードキャストの対象外である。

ブロードキャストコントローラ（図2）：

プロセッサの共有メモリへの書き込みを検出してブロードキャストする機能と、他のブロードキャストコントローラによって結合バスに放送されたデータを共有メモリに書き込む機能をもつ。

プロセッサが共有メモリに書き込みを行なうと、そのデータとアドレスを内部のFIFOに格納し、結合バスの使用権が与えられたときにFIFOに蓄えられたデータを結合バスに放送すると同時に、結合バスからデータを受信するFIFOにも書き込む。また、セマフォ領域に対する書き込みにはIDを付加して放送する。

一方、結合バスを通じて放送されたデータは、ブロードキャストコントローラのFIFOに取り込まれたのち、プロセッサのアクセスに優先して、共有メモリに書き込まれる。その際、セマフォのIDが付加されたデータは共有メモリには書き込まない。また、FIFO(1024byte)のデータが512byteに達した時点で、half-full信号を結合バスに出力して、結合バスの使用を一時的に停止させる。

結合バスの使用権は、ブロードキャストコントローラとセマフォコントローラをリング状に接続した結合バス放送権譲渡信号線によって与えられ、その間を順番に移行する。結合バスの使用権を得たコントローラはすべてのデータを放送し終わったのち、次のコントローラに使用権を譲る。

動作速度は、結合バスのサイクルタイムが最小125ns、共有メモリへの書き込みサイクルが340nsに設計している。

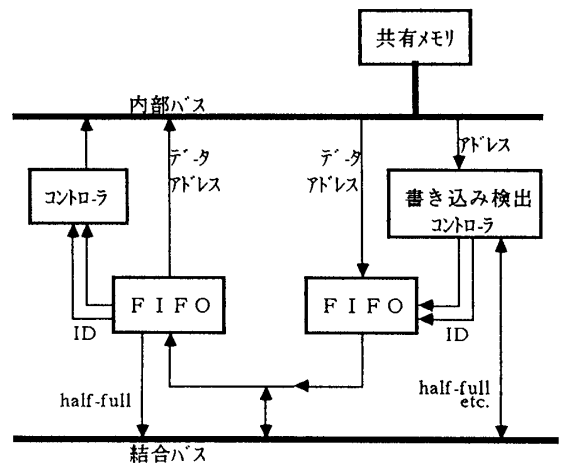


図2 ブロードキャストコントローラの概略

セマフォコントローラ (図3) :

先着判定のできるレジスタを1つもつ。このレジスタへの更新要求は共有メモリ先頭の4byteに0以外の値(プロセッサID+プロセスID)を書き込むことを行ない、いったんセットされると0を書き込むことでセマフォをリセットしないかぎり更新されない。また、セットされた値は、セマフォコントローラが結合バスの使用权を得た時点で、共有メモリの先頭からのオフセット4byte目にブロードキャストされる。各プロセッサは書き込んだ値とブロードキャストされた値とを比較して、等しければセマフォを獲得できたことを確認できる。それらの処理は他のデータと同一の経路で行なわれるため、セマフォ操作とデータ更新の順序が逆転することはない。

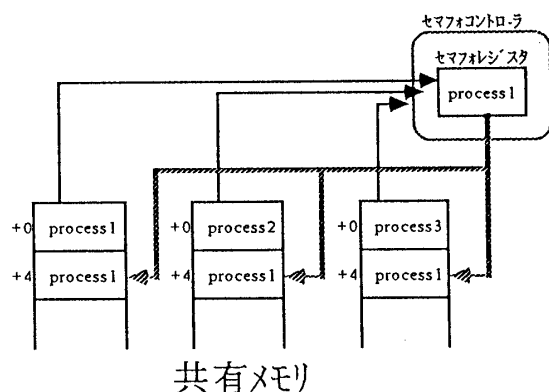


図3 セマフォコントローラの動作

3. ソフトウェアの構成

OSにはUNIXを使用し、その上に各種の機能を実現する。最初の段階ではOSの変更は行なわない。また、共有メモリ領域は、UNIXのプロセス空間とは区別し、仮想記憶の対象とはならない。メモリは、数Kbyte単位でユーザープロセス空間に割り付け、ローカルメモリと同様にポインタによってアクセスする。その際、IDやアドレスを含む管理情報を共有メモリ上に配置して、参照できるようにすることが必要であろう。

本システムではメモリの内容は複数のプロセッサで共有されているから、一般の共有メモリシステムと同様に、際どい区間に対して注意する必要がある。本システムでは、不可分なメモリアクセスをセマフォによって保証する。セマフォ変数は共有メモリ上に確保し、その値のテストとセットをセマフォコントローラを用いて不可分にすることで実現できる(図4)。

ただし、OSの変更を行なわないという条件のもとでは異なるプロセッサで実行されているプロセスにシグナルを送ることは困難であり、セマフォの獲得を失敗したプロセスのブロックとその再開を効率的に行なうことができない。また、共有メモリへ値を書き込んでから、セマフォコントローラによってその値がブロードキャストされるまで待たねば結果が分らないため、このセマフォ操作はあまり効率的ではなく、応用プログラムを作成するときにはそのことに注意しなければならない。

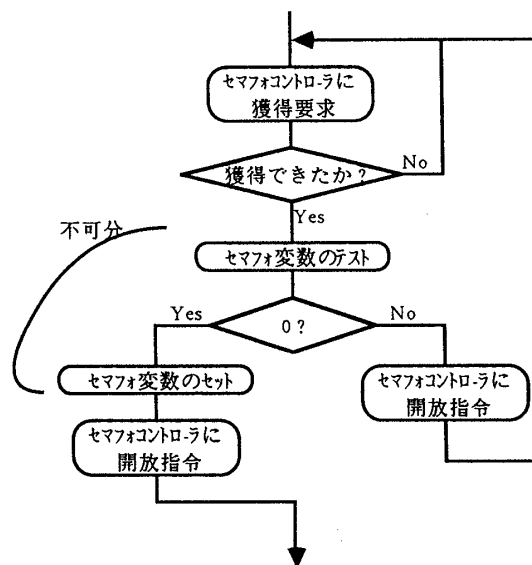


図4 セマフォの獲得

4. システムの特徴と応用

本システムの性能を決定する大きな要因に共有メモリに対する'読み出し・書き込み比'がある。共有メモリからの読み出しは結合バスの動作に影響を与えないため、読み出しの比率が大きいほどシステムの性能は理想的になり、分散処理が効果的なものになる。

また、共有メモリへの書き込みがすべての共有メモリで実行されるのに時間がかかるため、複数のプロセッサで同一の領域を読み書きすると同期に時間を消費し、応答が悪化することが考えられる。

このような特徴を考えると、データを供給するプロセッサが特定され、そのデータに対して複数のプロセッサで異なった処理をほどこす機能分散的な応用がもっとも適していると考えられる。連立方程式等の並列演算も考えられるが、プロセッサ間の同期を頻繁に行なうアルゴリズムでは不利になる。

5. おわりに

'読み出し・書き込み比'の大きい応用に対して、本システムは非常に有効であると考えられる。今後、システムが完成次第、性能評価を行なう予定である。

参考文献

- [1] 榎田、斉藤、後藤: "ブロードキャストメモリ方式による分散処理コンピュータシステム"、情報処理学会第37回全国大会、pp. 155-156(1988)。