7D−7

並列オブジェクト指向トータルアーキテクチャA-NET
— ルータの構成 —

佐々木昌　茂木久　吉永努　馬場敬信
syo,mogi,yoshi,baba @infor.utsunomiya-u.ac.jp
宇都宮大学工学部

1 はじめに

A-NET 計算機 [1] は、1000 台規模のノードプロセッサを結合した中粒度高並列計算機である。各ノードプロセッサは、メソッドを実行する PE と通信を制御するルータ [2] からなる。A-NET 計算機では、そのネットワークの構造を静的可変とし、それぞれの応用分野に適したネットワークトポロジを実現するという特徴を持つ。

本稿では、オブジェクト及びメッセージの通信方式とネットワークトポロジ独立性を達成するためのルータのアーキテクチャについて述べる。

2 ルータの設計方針

ルータの設計方針は、次の3点に要約される。

- ネットワークトポロジ独立

  オブジェクト単位の比較的大きな粒度の並列処理を前提とした場合、動的な負荷分散はコストが大きくなる。そこで A-NET では、実行時のオブジェクトコード転送はユーザによって新しいオブジェクトの生成が明示的に指定された時のみに限定し、静的な負荷分散を有効利用する実行モデルを採用した。これに伴い、FPGA などのプログラマブル素子を用いることによりノードプロセッサの結合方式を静的に可変なものとし、ネットワークトポロジに応じた経路選択ができる構造とする。

- 2種類の通信方式

  オブジェクトレベルの並列性を引き出すためには、メソッドに相当するコードも各ノードプロセッサに独立して持つとともに、遅延の少ないメッセージ通信を達成する必要がある。また、動的オブジェクト生成に伴うオブジェクトコード転送と通常のメソッド起動のためのメッセージ転送を比較すると、オブジェクト生成とメッセージ通信の発生頻度や及びそれぞれのデータ転送量が大きく異なると考えられる。このことがシミュレータにより検証されたため、A-NET ルータでは、オブジェクト転送時の回線交換方式と通常メッセージ転送時の適応型バーチャルカットスルー方式の2種類の通信方式を実現する。

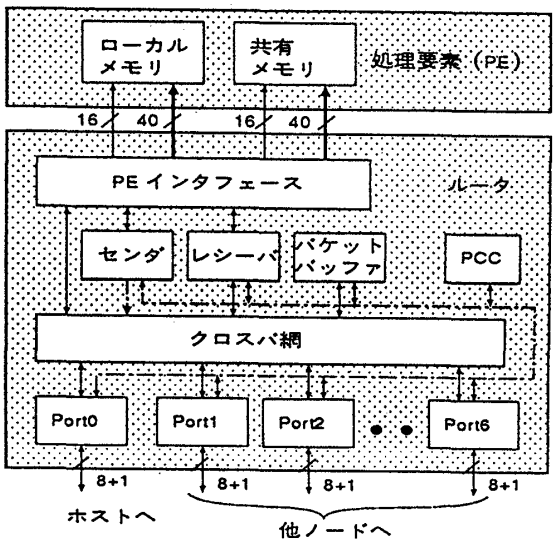

図 1: ルータのブロック図

- ルータ内並列動作の実現

  A-NET 計算機はノードプロセッサ間の並列性だけでなく、ノード内において PE のメソッド実行とルータのメッセージ交換の実行の並列動作による処理効率の向上を図っている。また一つのルータ内でもメッセージのオーバヘッドを減らすため、独立して動作する複数のブロックからなる設計とし、各ブロック間でのデータ転送を並列して行えるものとする。

3 ルータの構成

以下に各ブロックの構成と機能について述べる。図1にルータのハードウェア構成を示す。

3.1 PE インタフェース

ルータは、メッセージの送受信時に 40KB の PE-ルータ共有メモリ (CM) 上にある PE との共有情報へのアクセス、および出力キューからのメッセージ取り出す。またメッセージや動的オブジェクト生成による目的コードは PE のローカルメモリ (以下 LM) にアクセスする。この機能を実現するためのものが DMAC で、共有メモリアクセス回路・ローカルメモリの DMA 転送回路・バイト-ワード変換回路などから構成される。

ルータは自ノード宛メッセージを受信すると DMAC によりこれを直接 LM に書き込み、割り込みにより PE 上の OS を起

A Parallel Object-Oriented Total Architecture A-NET
— Organization of the Router —
Syo SASAKI,Hisashi MOGI,Tsutomu YOSHINAGA
and Takanobu BABA
Utsunomiya University

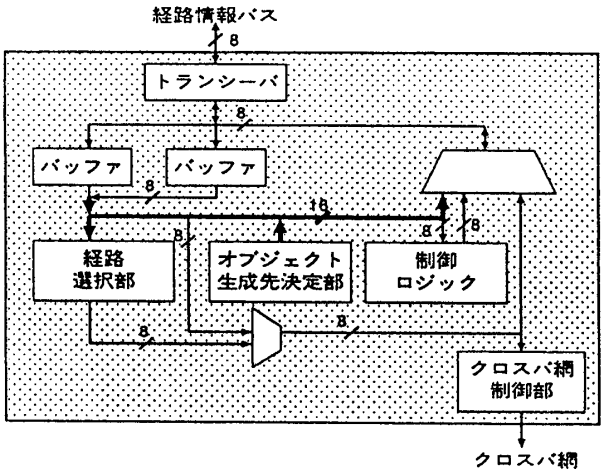

図 2: PCC ブロック図

動する。この機構は、メッセージ受信時に PE の動作を一時停止する必要があるが、OS によるメッセージコピーなどの処理が省け、メッセージの到着間隔と PE の内部処理のバランスを取ることで全体のスループットを上げることができる。

3.2 プログラマブル通信制御装置 (PCC)

トポロジ独立なルーティングをサポートするために各ポートに経路選択機能を持たせることは、ハードウェア量の観点から困難である。また A-NET ルータは、動的オブジェクトの割り付け先候補を生成する必要があり、これもトポロジにより変化する。これら 2 つのトポロジにより変化する機能を、プログラマブル通信制御装置 (以下 PCC) によりサポートすることとした。PCC をプログラマブルデバイスにより構成することにより、ある程度ルーティングをハードウェア化し高速化を図る。図 2に PCC のブロック図を示す。

PCC の機能をまとめると、メッセージ転送の経路選択・クロスバ網の設定・動的オブジェクト生成先候補の決定・経路選択などの割り込み要求の競合調停などがある。

3.3 メッセージレシーバ (MR)

メッセージレシーバ (MR) は、1 パケットを格納できるメモリ、CM のアドレス計算や PE-ルータ間共有情報、パケットデータの比較等を行う回路、制御ロジックからなる。

MR の主な機能は、自ノード宛であるパケットを受け取って、種類に応じた受信処理を行うことである。通常の受信パケットについては、CM への書き込みを DMAC に依頼する。DMAC を通して LM へ受信メッセージを格納した後、PE に割り込みを掛けて OS にその受理を示す。また、ルータデバッグ用メッセージを受信した場合は、MR 内のパケットメモリ上で、必要な情報を持つ返答用パケットを送り返すなどの処理を行う。このような複雑な制御は、ファームウェアで実現する。

3.4 メッセージセンダ (MS)

メッセージセンダ (MS) は、パケットヘッダを生成する回路を持つ以外は MR とほぼ同様なハードウェア構成である。

PE のメッセージ送信要求は MS で次のように処理される。PE がメッセージを CM 上の出力キューに書き込んだ後、送られてきた割り込み信号を MS が受信して、PE インタフェースを介して送出すべきメッセージを CM から読み出す。その後、パケットの種類などの情報を含むヘッダを生成して PCC に経路選択を依頼しクロスバ網を通してポートから転送する。

3.5 ポート (Port)

ルータは、隣接ルータ結合用ポートを 6 つ、ホスト結合用ポート 1 つ、計 7 つの半二重ポートを持つ。各ポート間のリンクは、データ 8 ビット＋パリティ 1 ビットの計 9 ビット長である。

ルータ-ルータ間のパケット交換は、送信側主導でハンドシェークを用いずに行う。受信ポート側では、外部から入力されたパケットをポート内 FIFO に格納するとともに PCC に経路選択要求を発する。要求が受け付けられるとパケットの宛先を PCC に送って、PCC が転送先となるポートや MR を確保しクロスバ網の設定を待つ。その後、PCC からのパケット送出コマンドによって、FIFO 内のパケットをクロスバ網を通して転送する。

PCC はポートの利用状況に応じて適応ルーティングを行うが、すべての送出候補ポートが使用不可能なときは、パケットをパケットバッファに退避し、送出ポートが使用可能になるのを待つ。ノード間転送中に、1 つのパケットが複数のノードにまたがることを可能とすることで、ストア&フォーワード方式よりも転送遅延を小さくすることができ、ブロッキングが起きなければワームホールルーティングに近い転送速度が達成できる。転送能力は、20MB／s／リンクである。

ホスト-ルータ間のパケット交換は、データ転送速度の違いを吸収するためハンドシェークによりデータ転送を行う。プロトタイプにおいては、ホストの VME インタフェース回路と各ルータのホスト用ポートは競合調停用の回路を介してバス結合される。

3.6 パケットバッファ (PB)

パケットバッファは、前述したポートや MS 、MR などの各ブロックから転送されるパケットのブロッキングを回避するためのブロックである。ハードウェアは一時退避用の FIFO と制御ロジックから構成される。

4 おわりに

現在 A-NET ルータは各ブロックの詳細設計を行なっている。パケットの格納や取り出しを効率良く実行するために、ルータ内で各ブロックがそれぞれ小規模なシーケンサを持つ構造とし、並列に動作を可能としている。それぞれのシーケンサは現段階で PLD が用いられているが、全体のハードウェア量を考慮し更に高密度デバイスを使用することも検討している。

参考文献

[1] Baba, T. et al.: A Parallel Object-Oriented Total Architecture: A-NET, *Proc. Supercomputing '90*, pp.278–285 (1990).

[2] 茂木ほか: A-NET 計算機におけるトポロジ独立なルータのアーキテクチャ, 電子情報通信学会, コンピュータシステム研究会, SWoPP 大沼 '91, CPSY91-4 (1991).