マルチPSIシステムとその接続方式

7B-1

木村康則 瀧和男 内田俊一

(財)新世代コンピュータ技術開発機構

1．はじめに

第５世代コンピュータの研究開発プロジェクトでは，並列推論マシンＰＩＭの研究を主にアーキテクチャの観点から進めてきたが，その過程で，並列ソフトウエアに関する研究の重要性が明らかとなった［３］。マルチＰＳＩシステムは，将来の大規模並列推論マシンの実現を目指して並列ソフトウエアの研究開発を進めていくために，ＰＩＭの完成に先駆けて研究開発ツールとして使用される疎結合の並列マシンであり，以下に示すスケジュールで研究開発を進めている［１］。

（１）マルチＰＳＩ第１版（６１年度）

要素プロセッサ（以下ＰＥと略す）として前期に開発したＰＳＩを６～８台格子状に接続したシステム。並列言語として並列論理型言語ＫＬ１（核言語第１版）を使用する［２］。処理系やネットワークハンドラなどはＥＳＰで書かれ，ＰＥ内の処理は，逐次型ＯＳのＳＩＭＰＯＳ下で進む。速度よりも開発効率を重視したシステムである。

（２）マルチＰＳＩ第２版（６２～６３年度）

現在開発中のＰＳＩ－ＩＩ［６］を１６～６４台接続したシステム。処理系は，ＫＬ１の機械語ＫＬ１－Ｂをファームウエアで実行することにより高速実行を目指しており，本格的な並列ＯＳ（ＰＩＭＯＳ）を実装した実用的なシステムである。

（３）ＰＳＩ－ＩＩ（６１～６２年度）

マルチＰＳＩ第２版のＰＥとして使うためにＰＳＩを改良し，小型化，高速化したものである。また，単体でスタンドアロン型マシンとしても使えるように設計されている。

本稿では，マルチＰＳＩシステムの研究項目などについて説明した後，現在開発を進めている第１版のＰＥ間接続方式について報告する。

2．マルチＰＳＩシステムの研究項目と役割

マルチＰＳＩシステムで行おうとしている並列ソフトウエアの研究項目をまとめると以下のようになる［１］。

（１）論理型言語ＫＬ１とその処理系の開発

（２）並列ＯＳ（ＰＩＭＯＳ）の基本部分の開発

（３）ＰＩＭＯＳのうち大規模応用プログラムの効率的実行方式に関する研究

2．1　マルチＰＳＩ第１版の役割

第１版では，上記，（１），（２）に重点を置いて研究開発を行う。すなわち，並列論理型言語ＫＬ１の分散環境での効率的な実現方式，ＫＬ１の上位言語となるべきシステム記述言語の開発や，並列ＯＳのうち，メモリ管理，オブジェクトコードの分配，管理，入出力や割込みなどプログラム実行に不可欠な部分の開発である。このとき，必要ならばファームウエアも含め，ＰＳＩシステムの変更も行う予定である。

2．2　マルチＰＳＩ第２版の役割

第２版では，第１版で得られた知見をもとに，大規模なプログラムを走らせるに足る高速な並列マシンを作成し，（３）の研究を行う。すなわち，プログラムが本来持つ並列性をいかにして引き出すか，局所性を考慮した，効率的な負荷分散方式はどうすればよいか，などについて検討する。これは，従来の逐次ＯＳには無いが並列ＯＳには必須である機能の検討と言うこともできる。また，中期ＰＩＭ研究との関連においては，距離を意識したクラスタ間通信において，通信の局所性を保つ方法の検討と考えることができる［３］。

3．マルチＰＳＩ第１版の接続方式

3．1　概要

ＰＳＩは，元々スタンドアロン型のワークステーションとして開発したものであり，マルチプロセッサ用のハードウエア機能は持っていない。そこで，新たに接続用のハードウエア［５］を作成し，ＰＳＩを疎に結合した。

ＰＥ間の通信は，パケット交換によるメッセージ通信によって行われる。この処理の概要を図１に示す。

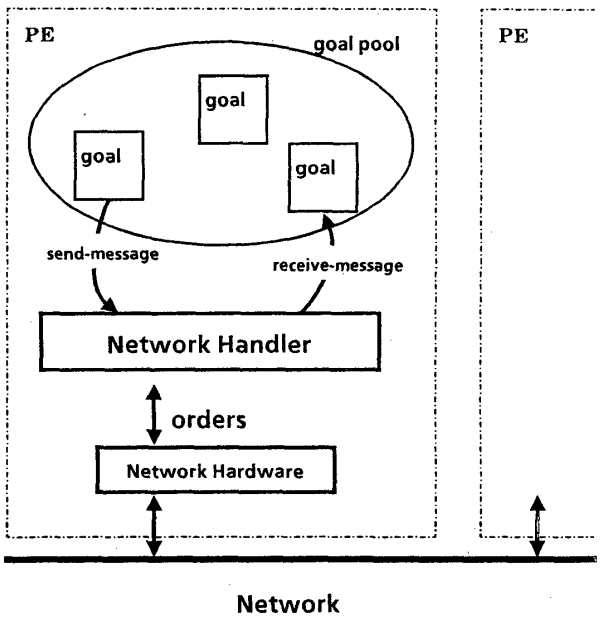

図１．処理の概要

MultiPSI system and its network

Yasunori KIMURA, Kazuo TAKI, Shun'ichi UCHIDA
Institute for New Generation Computer Technology (ICOT)

3．2　ＰＥ間通信処理

マルチＰＳＩ第１版システムにおいて或るＰＥの処理系から他のＰＥのデータをアクセスする場合の手順は以下の様になる。

（１）ＫＬ１処理系において，外部参照タグを持ったデータが現われると，処理系内で管理しているテーブルを引き，参照先ＰＥを求め，下位モジュールであるネットワークハンドラに対して参照要求をだす。

（２）ネットワークハンドラでは，参照要求とデータから図２に示す様なパケットを構築し，バッファに格納する。具体的には，ＫＬ０で言うバイトストリングをパケット格納領域として用い，その先頭バイトに行先ＰＥ番号を，終了バイトに送出元ＰＥ番号を格納し，その間に変換されたデータをいれる。表１にパケットの通信オーダを示す。

（３）次に，ＰＥ内の処理の隙間でためていたパケットをネットワークに送出する。具体的には，送出用に用意された組込述語に，パケットの格納されたバイトストリングと，その長さを与え，実行することにより行う。

（４）ネットワーク上のパケットは，目的のＰＥまで自動的に転送される。

（５）また，ネットワークよりパケットを読み込む時も組込述語により，バイトストリングに読み込まれたパケットをハンドラで処理系に意味のある形に再構成した後，処理系に渡す。

7		0
1	行き先ＰＥ番号	
パケット長（下位）		
パケット長（上位）		
ＰＥ間通信オーダ		
引数個数　N		
引数　1		
:		
引数　N		
0	送り元ＰＥ番号	

図２．パケット形式

3．3　ＰＥ間通信オーダ

パケットにうめ込む通信オーダはＫＬ１処理系と密接に関連しているものである。全部で１４種用意したが，ここではその一部について簡単に説明する。

read，read_answer は，他ＰＥに対する値の問い合わせと返答に用いられ，read_n_level は，構造体などの深さnレベルまでの値の問い合わせに用いる。unify は，他ＰＥに対する単一化要求である。throw_goal は，負荷分散処理などの結果，他ＰＥにゴールを投げるときに用いられる。第１版処理系では，ゴールを投げると，送出元ＰＥのＡＮＤ木には，“代理”が作られ，行き先ＰＥには，この代理を指す“里親”メタコールが一般的には作られる。しかし，行き先ＰＥに既にこの“里親”が存在するときには，二重作成をさけ，送出元ＰＥに対して今作った“代理”を消す要求を出さなければならない。このときに用いられるのが cancel である。このように，ＰＥ間にまたがるＡＮＤ木を管理するためのオーダを幾つか用意しており，kill，goal_dead，goal_terminates，goal_fails などがそれである（図３）。

表１．ＰＥ間通信オーダ

オーダ	引数
read	Var_id, Return_obj
read_n_level	Var_id, Return_obj
read_answer	Object_id, Value
unify	Var_id, Meta_call, Parent_record
throw_goal	Meta_call_id, Module_id,...
goal_terminates	Parent_record, Reduction_count
goal_fails	Parent_record, Reduction_count
goal_dead	Parent_record, Reduction_count
kill	Parent_record_id
cancel	Parent_record_id
wake_up_goals	Var_id, Value

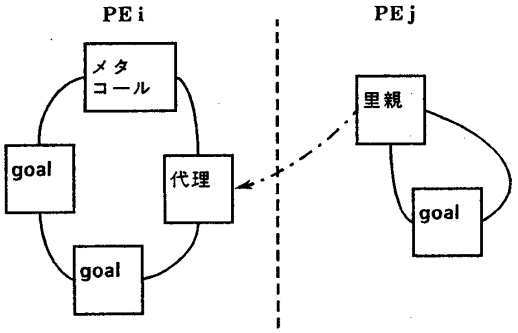

図３．代理と里親ノード

3．4　ネットワーク制御用組込述語

ネットワーク制御用組込述語としては，送出用に，check_write_buffer，write_buffer，読取用に，check_read_buffer，read_buffer，sense_packet_arrival の５種を用意した。

4．おわりに

現在，ＩＣＯＴにおいて，ＰＳＩを６台接続して第１版の試験，デバッグを行っているところである。今後は，第１版システム上でソフトウエアの開発を進めると共に，第２版システムに関する処理系及びネットワーク構成について検討を進める予定である。

＜参考文献＞

［１］瀧他：Multi-PSIシステムの概要，第32回情処全国大会，5Q-8，1986-3

［２］K. Ueda：GUARDED HORN CLAUSES，LPC '85，pp.225 1985-6，JAPAN

［３］後藤他：並列推論マシンPIM－中期構想－，本大会 3B-5

［４］宮崎他：Multi-PSI におけるGHCの実行方式，LPC '86，pp.83 1986-6，JAPAN

［５］益田他：マルチPSIのネットワークハードウエア構成，本大会 7B-2

［６］中島他：マルチPSI要素プロセッサPSI-IIのアーキテクチャ，本大会 7B-3