PIE64の通信機能の測定

5G−7

佐藤充, 小池汎平, 田中英彦
東京大学工学部

1 はじめに

現在、我々の研究室では、並列推論マシンPIE64[1]の開発を行なっている。PIE64の特徴として、その強力な通信機構が挙げられる。PIE64は相互結合網として、自動負荷分散機能を備えた回線交換方式の多段網を採用し、通信をサポートする構成要素としてネットワーク・インタフェース・プロセッサ(NIP)を搭載している。

本稿では、PIE64での実機での測定結果を基に、回線交換方式のネットワークで特徴的な輻輳の問題と、NIPの機能によるメッセージ送信におけるレイテンシ低減効果について調査した結果を報告する。

2 PIE64

PIE64は、記号処理の高速実行を目的とした並列推論エンジンである。PIE64は推論ユニット(Inference Unit: IU)と呼ばれる基本ユニットが2系統の相互結合網で接続される構成をとっている。2系統の相互結合網はそれぞれ自動負荷分散機能を備えた回線交換方式の3段の多段網である。

各IUには、メインプロセッサであるUNIRED[2]、通信・同期機構を主に扱うNIP[3]、さらに並列管理用プロセッサとしてSPARCが搭載されている。

3 測定

3.1 相互結合網の特性

PIE64の相互結合網は回線交換型の多段網であるため、通信が頻繁に発生し、相互結合網が混雑してくると各段での衝突が生じる。今回、この衝突の状態を表すパラメータとして、相互結合網の入口における接続までの待ち時間を選んだ。

相互結合網を混雑させる要因としては、

1. IU台数
2. メッセージ長(N[word])
3. 通信要求出現頻度(p[%])
4. 接続パターン

が考えられる。ここでは一般的な通信機能を測定するために、2と3を変化させて測定を行なった。

なお、他の条件については、1は現在稼働可能なすべてのIU(54台)、4はランダムに通信先を選ぶ方式を選んだ。

測定は、プログラム中でのメッセージ転送命令の発行間隔を9 clock〜∞ clockで変化させ、その時の

- プログラム全体の実行時間: T_{total} [clock]
- 転送回数: n [times]
- メッセージ転送を行なわない場合のプログラム実行時間: T_{inst} [clock]
- 衝突がない場合のメッセージ転送時間: T_{trans} [clock]

を測定し、$p = n/T_{total}$ を横軸に、$T_w = T_{total} - T_{inst} - T_{trans}$ を縦軸にプロットした。その結果を図1に示す。

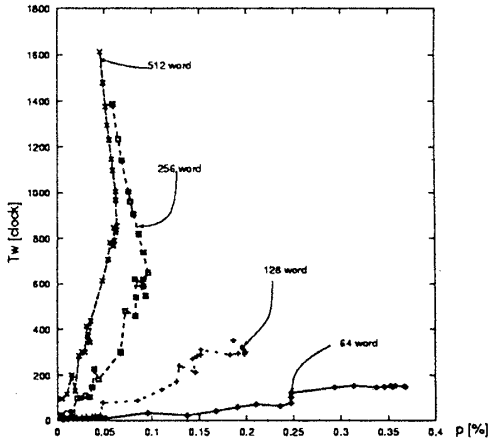

図1: 待ち時間の変化

図1を見ると、p と T_w が比例する領域と、反比例する領域とにわかれることがわかる。

反比例領域では、プログラム中での転送命令発行の割合をあげていくと、待ち時間(T_w)が急激に増大し、それに伴ってメッセージ転送にかかる時間が大きくなる。処理系では、このような状態に至らないようデータの配置や負荷分散に気をつける必要がある。

3.2 NIPの効果

PIE64では、通信は送信側Master NIPと受信側Slave NIPの間で行なわれる。ここで重要なことは、受信側ではSlave NIPが動作するのみで、UNIRED/SPARCは通信に関与する必要がないということである。特にメッ

Evaluation for Communication System of PIE64
Mitsuru SATO, Hanpei KOIKE, Hidehiko TANAKA
Faculty of Engineering, University of Tokyo

セージ通信で送信側が送り先のアドレスを知らない場合、一般には

1. 送信側から送信要求をだす
2. 受信側では受信バッファを用意して、送信側に acknowledge(Ack) を返す
3. データの転送を行なう

のような send/recieve 方式を用いるが、PIE64 の場合は受信バッファのアロケートも Slave NIP がハードウェア的に行なうので、メインの計算を中断することなく通信処理を行なうことができる。

ここでは、この NIP の有効性を確認するため、通信の方式を管理プロセッサである SPARC を用いた send/recieve(MP-Write) を使って同様の測定を行なった。条件は前節と同様である。その結果が図 2である。

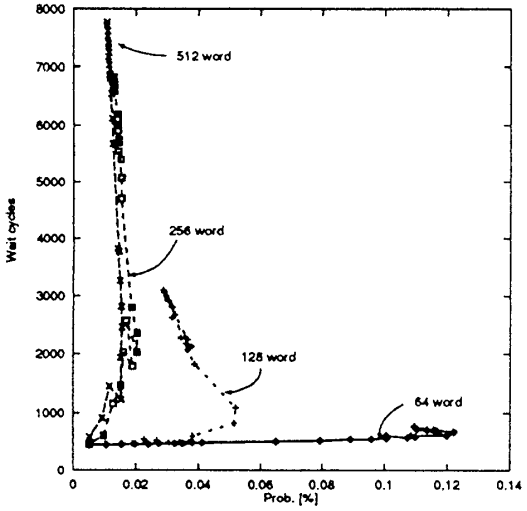

図 2: 待ち時間の変化 (MP-Write)

図 1と図 2を比較すると

- MP-Write ではかかる待ち時間が多い
- MP-Write では比例領域から反比例領域へ移るポイントが早く現れる (通常の転送に比べて、生起確率で 1/4 程度の位置)

ということがわかる。さらに、NIP を用いた場合に比べて、どの部分が特に時間がかかっているかを調べるために、さらにその内訳を調べた。その結果が図 3である。

図 3より、MP-Write では接続要求、Ack、データ転送と 3 回ネットワーク に接続しなくてはならないので、その分ネットワークでの待ち時間が多くなっていることがわかる。

また、時間的に大きな部分を占めているのが相手からの Ack を待つ時間である。このなかでも特に、相手 IU において、転送要求を受け取ってから Ack を生成する部

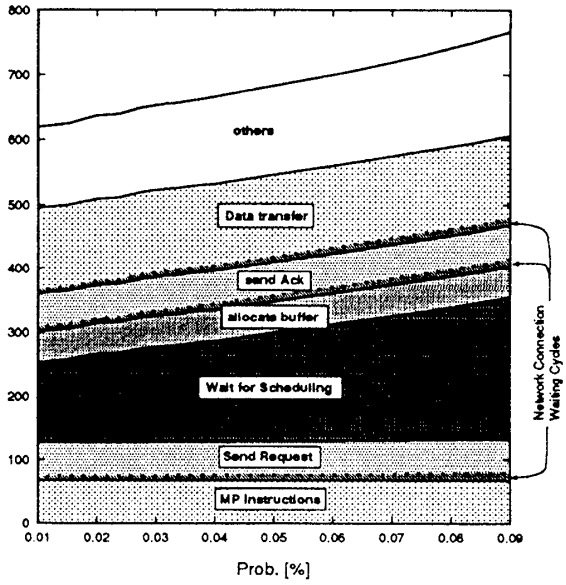

図 3: MP-Write の内訳

分がスケジューリングされるまでの時間が大きく影響していることがわかる。今回用いたテストプログラムは、データを転送するだけの非常に軽い処理しか行っていなかったが、実際のアプリケーションでは複雑な処理を行う必要があり、この部分がさらに大きくなるものと予想される。

4 まとめ

本稿では PIE64 の通信機能について、実機での測定を行ないその特徴について調べた。また、PIE64 で採用されている NIP について、そのメッセージ通信におけるレイテンシ低減効果について確認した。

今後はさらに実際のアプリケーションに沿った条件で調査を行なっていく必要がある。また、NIP にはこのようなローカルメモリを自由にアクセスする機能に加えて、Fleng の実行をサポートする様々な機能を備えている。これらの機能についても、実際の fleng 処理系をもとに測定する必要があるだろう。

参考文献

[1] 日高康雄, 小池汎平, 高橋栄一, 島田健太郎, 清水剛, 田中英彦. 高並列推論エンジン PIE64 研究経過報告 – ハードウェア –. 情報処理学会第 46 回全国大会, 1993.

[2] Kentaro Shimada, Hanpei Koike, and Hidehiko Tanaka. UNIRED II: The high performance inference processor for the parallel inference machine PIE64. In *Proceedings of FGCS 1992*, pp. 715–722, 1992.

[3] 清水剛, 小池汎平, 田中英彦. 並列推論マシン PIE64 のネットワーク・インタフェース・プロセッサ. 並列処理シンポジウム JSPP'89, pp. 99–106, 1989.