

## 動的負荷分散制御を行う 相互結合網

2D-1

坂井 修一   田中英彦   元岡 達  
(東京大学 工学部)

### 1. はじめに

高並列計算機において、通信オーバーヘッドを低く抑えながら、効率良い動的負荷分散を実現する方式について述べる。本方式は、タスクの転送を行う相互結合網に、行先の負荷の情報をフィードバックする信号線を設け、これを使って行先プロセッサを自動的に決定するものであり、(1)負荷分散管理のオーバーヘッドが小さい、(2)多段結合網などを用いた場合、各プロセッサがグローバルな負荷情報を得ることができ、(3)高い転送スループットを保證することができ、などの利点をもつ。

本稿では、このような相互結合網(負荷分散適応型網)の構成要素であるスイッチング・ユニット(SU)の設計、本方式のシミュレーションによる評価に関して述べる。<sup>(1)(2)</sup>

### 2. 負荷分散適応型SU

負荷分散適応型網のSUには、タスクの転送と逆の向きに行先プロセッサの負荷の情報を伝送し、負荷の量が少ない行先を選択してそこにタスクを送りつける機能が必要である。SU(今回は $4 \times 4$ )設計の基本方針を記す。

(1)負荷分散向けの転送モードと、行先を指定する通常の通信向けの転送モードをもつ。どちらのモードでもマルチキャストができる。

(2)多段結合網(マルチパスの網も含む)を対象とする。

(3)網内のクロック・スキューを考慮して、SU内は同期制御、SU間是非同期制御とする。データ転送は、同期でも非同期でも行える。

(4)回線交換方式を採る。

我々はこのようなSUを2種設計した。一方は、閾値設定-タスク要求線方式(SU-1)であり、他方は最少負荷量方式(SU-2。図1)である。後者は半二重通信を行い、使用中でないデータ線には、その時点で閉塞なしに行けるプロセッサ群の中で、最も負荷の軽いものの負荷量が、逆方向に流れている。負荷分散の経路設定は、各段のSUで、最少負荷量を示す空き出力ポートを選択することで行われる。

### 3. 負荷分散シミュレーション

負荷分散方式の評価をイベントドリブン・シミュレーションによって行った。今回は、系ホクローズド・システムで、定常的な動作をする場合を扱う。シミュレーションのパラメータとして、①プロセッサのサービス時間、②タスクの生成分布、③プロセッサ台数(N)、④結合網の形状、⑤タスク転送時間、⑥プロセッサのキュー容量、⑦負荷情報の伝達・更新に要する時間、などが

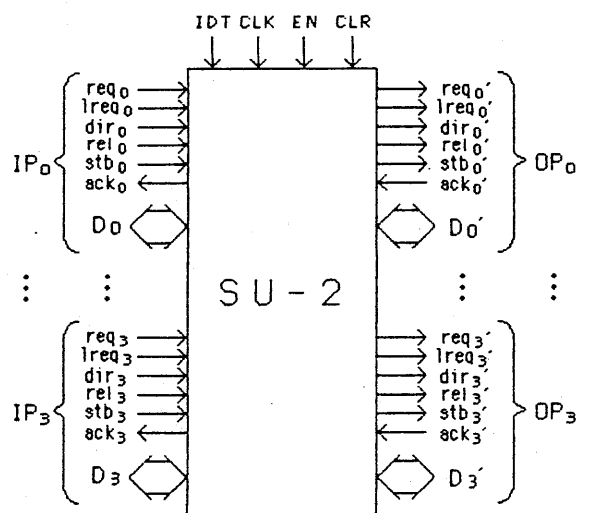


図1 SU-2の全体構成

挙げられる。ここでは、①平均 $T_p$ の指数分布（各プロセッサの処理能力は同じ）、②1タスクの終了ごとに1つ生成、③オメガ網、④SU 1段あたり $T_t$  ( $T_p / T_t = 0.25$ )、⑤無限大、⑥0、とした。

対象とした負荷分散方式は、(1)乱数分散方式、(2)負荷のみを集中管理する方式、(3)閉塞無しで到達可能な任意のプロセッサを行先とする方式、(4)SU-1を用いる方式（閾値 = 平均タスク数 + 5）、(5)SU-2を用いる方式、である。初期状態では、各プロセッサの出力キューに21個のタスクがあった。

シミュレーションの結果を図2（正規化スループット = 平均プロセッサ稼働率）、図3（キュー長の標準偏差 = 負荷配分の指標）に示す。網内の閉塞を考慮した(3)(4)(5)が正規化スループットの点で有利であり、正確な負荷情報を参照することのできる(2)(5)が均等な負荷の配分という点で有利であることがわかる。後者の点は、特にプログラムの立ち上げ時と終了時に意味をもつ。

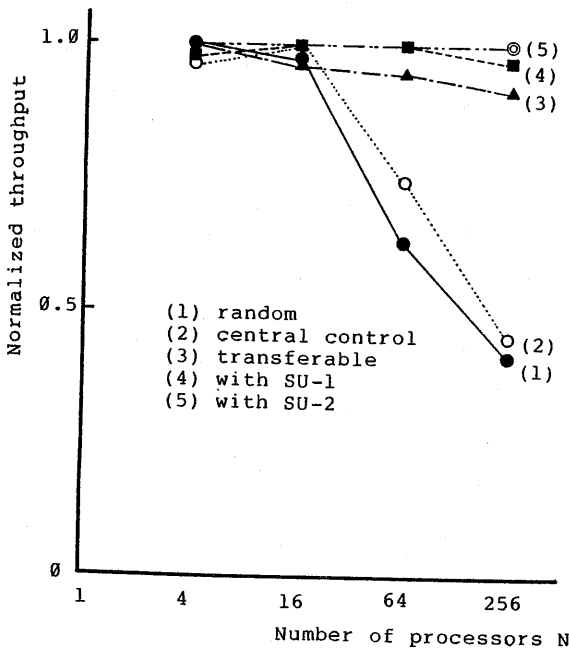


図2 オメガ網を用いた場合の正規化スループット

(5)の方式は、中央制御装置が不要で、しかも負荷情報の伝達・更新に要する時間が短い利点がある。以上の諸点から、SU-2によって構成される網を用いた動的負荷分散が、ここで採りあげた中でもっとも優れた方式であると結論づけられる。

4. おわりに

動的負荷分散を行う相互結合網の提案、構成要素であるSUの設計、負荷分散のシミュレーション評価に関して述べた。非定常的な問題を想定したシミュレーション評価、SU-2のLSIによる試作などが今後の課題である。

文 献

- (1)坂井他：“高並列推論エンジンPIEにおける相互結合網の構成”，信学技報，E C 84-46.
- (2)坂井他：“動的負荷分散を行う相互結合網”，信学技報，'85年8月.

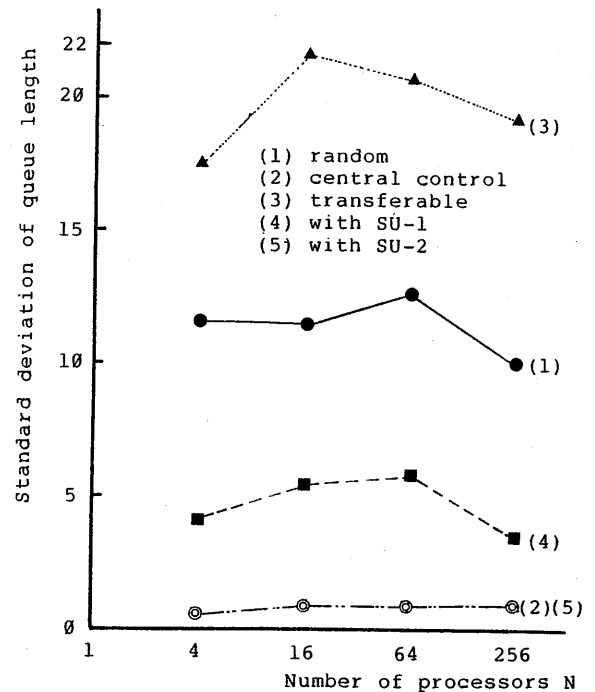


図3 オメガ網を用いた場合のキュー長の標準偏差