7L−7

スーパーデータベースコンピュータ (SDC) のバケット平坦化ネットワークにおける縮退動作時の動作特性

田村孝之　中村稔　喜連川優　高木幹雄
東京大学 生産技術研究所

1　はじめに

“スーパーデータベースコンピュータ (SDC)” は, 現在我々が開発中の高並列関係データベースサーバである. SDC は, 数台(4〜6台)のマイクロプロセッサと磁気ディスク装置とを共有バスで密結合して処理モジュールとし, さらに複数の処理モジュールをネットワークで疎結合したハイブリッドアーキテクチャをとる [1]. また, 結合演算に対するアルゴリズムとして, “バケット分散 GRACE ハッシュ” 法 [2] を採用し, これをハードウェアで支援するために, “バケット平坦化機能” を有するオメガネットワークの提案がなされている [3].

このネットワークは, 各スイッチ素子自体が局所的な履歴に基づいて適応的なルーティングを行ない, 競合によるスループットの低下と処理モジュール毎の負荷の偏りに起因する性能向上の限界とを同時に解決することを目指したものであり, その有効性はすでにシミュレーションにより確認されている [3]. また, バケット平坦化機能にはこれまでにいくつかの拡張が施されてきたが [4], 処理モジュール数はネットワークの大きさに等しいと仮定され, ネットワークの性質から 2^n に限られてきた.

しかし, 各処理モジュールの故障に対するロバスト性を向上させ, また, 処理の規模に応じて徐々にシステムを拡張できるようにするには, 任意のモジュール数が許されることが望ましい. そこで今回, これまでの制限を取り除き, ネットワークの大きさと異なる任意数の処理モジュールの間で負荷分散を可能にするアルゴリズムを開発した. 本論文では, この新たなアルゴリズムを用いた時のネットワークの動作特性を, シミュレーションによる解析結果に基づいて述べる.

2　バケット平坦化アルゴリズムの拡張

ネットワークの一部のポートに接続された処理モジュールの間でバケット平坦化を行なうには, ポート毎に送られるデータの量を変える必要がある. したがって, 各スイッチの2つの出力ポートから出力されたタプル数 $C_i(x)$ の分散に基づいて, バケット x の分布の平坦さを表すことにすると,

$$V(x) = \frac{1}{4}\left\{\frac{1}{N_{p_0}(m)}C_0(x) - \frac{1}{N_{p_1}(m)}C_1(x)\right\}^2 \equiv \frac{1}{4}D^2(x) \quad (1)$$

で定義される $V(x)$ を用いて, $F = \sum_x V(x)$ を評価関数にすればよい. ただし, m はスイッチの段数を表し, p_i はスイッチの出力ポート i のネットワーク中での通し番号である.

また, $N_p(m)$ はスイッチの出力ポート i から到達可能なアクティブなモジュールの総数を表し,

$$N_p(0) = n_p \quad (2)$$
$$N_p(k+1) = N_{2p \bmod N}(k) + N_{2p+1 \bmod N}(k) \quad (3)$$

で定義される. ここで, n_i はモジュール i が アクティブな時は 1, ダウンしている時は 0 であり, $N = 2^n$ はネットワークの大きさを表す. N_p の設定される様子を図1に示す.

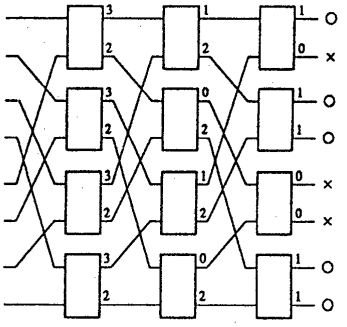


図 1: アクティブなモジュール数の設定法

このように設定された評価関数 F に基づき, スイッチに新たに入力されたタプルについて, 2つの接続状態, straight (0) と crossed (1) のうち F の増加が小さくなる方を求めると,

$$state = \begin{cases} 0 & \cdots \quad D(b_0) - D(b_1) < 0 \\ 1 & \cdots \quad \text{otherwise.} \end{cases} \quad (4)$$

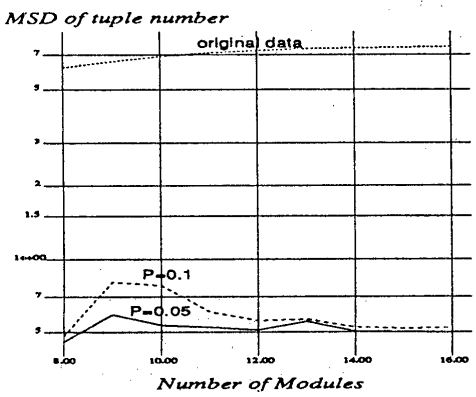
となる. ただし, b_j はスイッチの入力ポート j に入力されたタプルが属するバケットである.

つまり, 一部の処理モジュール間でのバケット平坦化を実現するには, 各スイッチ毎に全てのバケットについてカウンタを用意し, タプルを出力する度に, ポート 0 から出力したタプルの属するバケットについてはカウンタを $+1/N_{p_0}$ し, ポート 1 から出力したタプルの属するバケットについては $-1/N_{p_1}$ すればよいことになる.

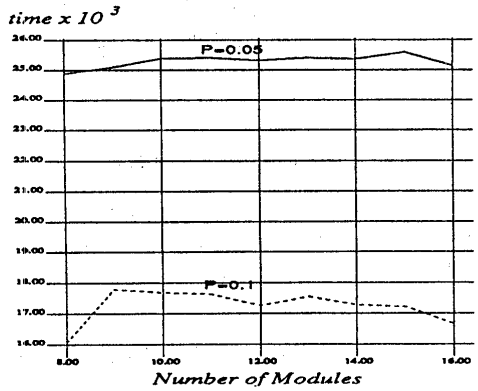
3　シミュレーションによる評価

ここでは, ネットワークのサイズを固定し, アクティブな処理モジュール数を変化させた時のバケット分布の平坦度と処理時間について, シミュレーションを行なった結果を述べる. このシミュレーションにおいては, 全てのタプルの発生間隔は指数分布に従うものとし, 処理モジュールはネットワークの端から1

Behaviors of bucket flattening network of the Super Database Computer under reduced configuration.
T.Tamura, M.Nakamura, M.Kitsuregawa, M.Takagi.
Institute of Industrial Science, University of Tokyo.

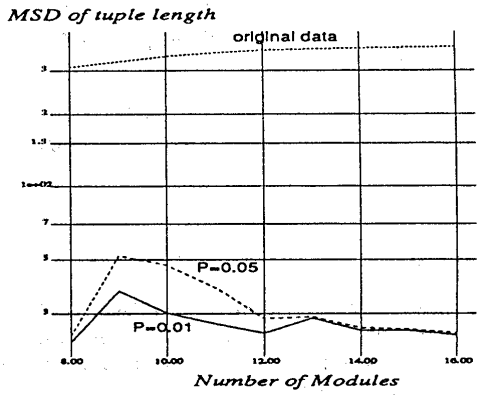


(a) モジュール数 vs. 平均標準偏差

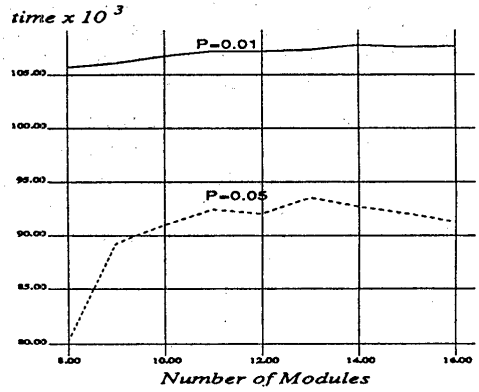


(b) モジュール数 vs. 処理時間

図 2: 固定長データに対する動作特性



(a) モジュール数 vs. 平均標準偏差



(b) モジュール数 vs. 処理時間

図 3: 可変長データに対する動作特性

つずつ, ネットワークの大きさの半分になるまで減らしていった.

図 2(a), (b) は各タプルを 10 ワードの固定長データとしたときのバケット分布 (属するタプル数) の平均標準偏差と処理時間である. ただし, ネットワークの大きさ N, バケット数 B, 処理モジュールあたりのタプル数 T はそれぞれ, 16, 128, 1024 とした. また, 任意の時点でのタプルの発生確率 P は 0.05, および, 0.1 の両方について調べた.

同様に, 図 3(a), (b) は各タプルが 20 ワードから 80 ワードまでの間に一様に分布する可変長データの場合について示している. この場合, バケットの分布は, 属するタプルの総ワード数の平均標準偏差で表している. N, B, T のパラメータは固定長の時と同じとし, P は 0.01, および, 0.05 とした.

平坦度についての結果を見ると, $\frac{3}{4}N$ 台よりモジュール数が多い時はほぼ一定の平坦度が達成されているのに対し, それよりモジュール数が少ない時, 特に $N/2 + 1$ 台のときには平坦度が若干悪化し, さらにトラフィックの影響を受けやすくなっているのが分かる. また, 処理時間については, トラフィックが多い時に $N/2$ の場合だけ小さくなっているが, 全体としてはモジュール数に対する大きな変化は現れていない. したがって, $2^n + 1$ 台になる場合を避ければ, どのようなモジュール数についても通常時に比べて遜色のない性能が得られると言える.

4 まとめ

SDC のバケット平坦化ネットワークにおいて, 利用可能なモジュール数の変化に対応できるアルゴリズムを示し, シミュレーションによる性能評価を行なった. その結果, 新たに提案した方式が, モジュール数の変化に対してもほぼ一定の負荷分散能力を有することが確認された.

参考文献

[1] 平野, 原田, 中村, 小川, 楊, 喜連川, 高木: “スーパーデータベースコンピュータ SDC のアーキテクチャ”, 並列処理シンポジウム, pp.137, 1990.

[2] 平野, 原田, 中村, 楊, 喜連川, 高木: “スーパーデータベースコンピュータ SDC のソフトウエア”, 電子情報通信学会技報, pp.7-12, 1990.

[3] 喜連川, 小川: “バケット平坦化機能を有するオメガネットワーク”, 情報処理学会論文誌, 30(11) pp.1494, 1989.

[4] 相場, 喜連川, 平野, 高木: “スーパーデータベースコンピュータにおけるバケット平坦化オメガネットワークの動作特性”, 電子情報通信学会技報, 1991.