

大規模並列システムのための動的負荷分散における 重なりのあるドメイン分割方針の検討

1B-6

五明 則人 芦原 評 清水 謙多郎
電気通信大学 情報工学科

1 はじめに

並列システムにおいて、プロセッサ資源を有効に利用するための技術として動的負荷分散がある。しかしながら大規模並列システムにおいてはシステム全体の負荷状態を把握し、その状態に基づいて動的に負荷分散を行うことはほとんど不可能である。従って、システムの規模が拡大してもそれに応じた性能が得られるスケーラビリティが要求される。そこで、我々は、スケーラブルな動的負荷分散方式として、システムを複数のグループ(ドメイン)に分割し、そのグループ内またこれらのグループを相互に組み合わせて、情報収集および負荷分散を行う方式について考えている。本稿では、このような方式をもとに、考えられる方式を分類・整理し、特にドメインを互いにオーバーラップさせる方式に着目して、性能をシミュレータを用いて比較・評価した結果について報告する。性能評価はプロセス移送や情報収集のコストを考慮した現実のシステムに近いモデルで行う。ドメインはあるPE(Processor Element)を中心に構成され、中心となるPEとドメイン内のPEの距離の最大値をドメインの大きさとする。

2 動的負荷分散の方針

- ドメイン間の関係
オーバーラップ、タイリング
- 移送判定の基準
自PEの負荷を利用、負荷が最小のPEの負荷を利用、負荷が最大のPEの負荷を利用
- 情報収集のタイミング
定期的、負荷の変動時、移送判定時
- 移送判定の主体
送り手主導、受け手主導
- 移送先の選択
ランダム、負荷が最小なもの
- 移送するプロセスの選択
ランダム、転送コストが最小なもの、残り実行時間が最大なもの

Overlapped Domain Decomposition for Adaptive Load Balancing in Massively Parallel Systems
Norihiro Gomyo, Hyo Ashihara, Kentaro Shimizu
The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182, Japan

3 性能評価

本研究では、相互結合網のトポロジが2次元メッシュ(トーラス)である並列システムを想定して性能評価を行った。表1はシミュレータで用いたパラメータとその既定値(特にことわらない限りこの値を用いる)をまとめたものである。性能指標には平均応答時間を使用した。

表1: パラメータの既定値

パラメータ	既定値
ドメインの大きさ	2
ドメイン間の関係	オーバーラップ
移送判定の基準	負荷が最大のPE
情報収集のタイミング	負荷の変動時
移送判定の主体	送り手
移送先の選択	負荷が最低なもの
移送するプロセスの選択	移送コストが最小なもの
プロセスの平均処理時間	1秒
プロセスの平均転送コスト	0.01秒
負荷情報の転送コスト	0.0001秒
ローカルスケジューリング	ラウンドロビン
PE数	256

3.1 スケーラビリティ

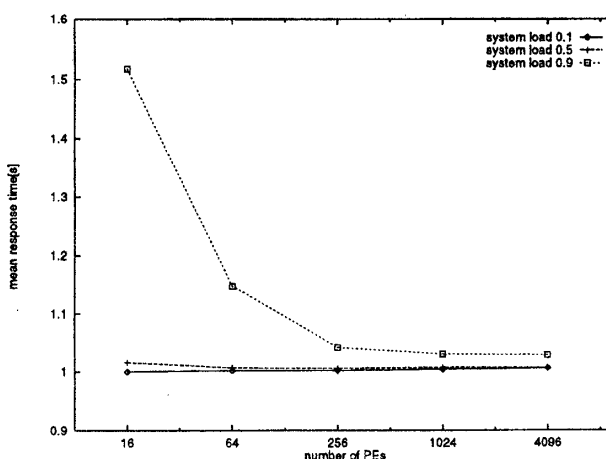


図1: スケーラビリティ

図1はPE数を変化させたときの性能を示したものである。PE数の違いによる性能の変化はほとんどな

く、ドメインに分割することによってスケーラビリティが得られていることがわかる。

システム負荷 0.9 のとき、PE 数が多い場合の方が平均応答時間は小さくなっているが、これは PE の数が増えることによって、負荷を分散できる可能性が増えるからである。

3.2 ドメイン間の関係

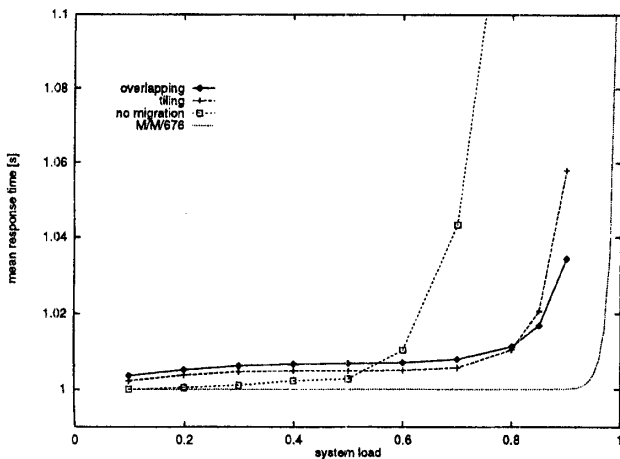


図 2: オーバラップとタイリングの比較

図 2 は PE 数が 676、ドメインの大きさが 2 のときのオーバラップとタイリングの性能を示したものである。システムの負荷が低いとき、情報収集のコストが低い分だけタイリングの方が性能が若干良くなっている。しかしタイリングの場合ドメイン間の負荷に不均衡が発生し得るため、負荷が高くなるとシステム全体の性能はオーバラップよりも悪くなっている。

3.3 ドメインの大きさ

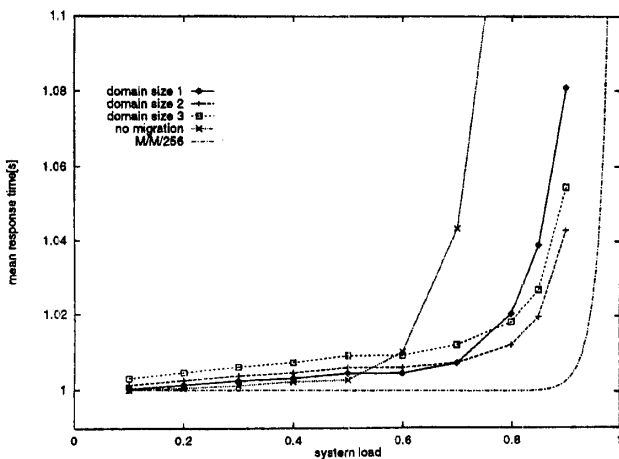


図 3: ドメインの大きさ影響

図 3 はドメインの大きさを変えたときの性能を示したものである。ドメインの大きさが 1 のときと比べて大きさが 2 のときの性能が良いのは、近くに負荷の低い PE が存在しない場合でも、遠くのより負荷が低い PE にプロセスを移送することによって性能向上が図れるためである。またドメインの大きさが 3 のときに性能が悪化しているのは、情報収集やプロセス移送のオーバーヘッドが大きくなり、遠くのより負荷が軽い PE にプロセスを移送することによって得られる性能向上の効果が薄れるためである。

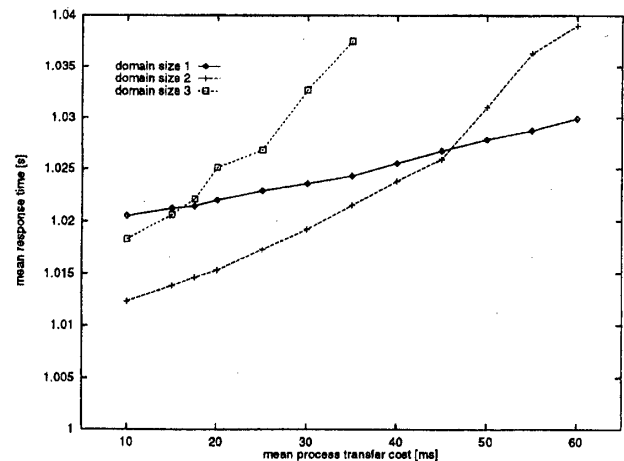


図 4: プロセスの転送コストの影響

図 4 は大きさ 1, 2, 3 のドメインに対しシステム負荷を一定 (0.8) としたときの、プロセスの平均転送コストの変化を示したものである。転送コストが大きいほど小さいドメインの性能が良くなっていることがわかる。また、ドメインが大きくなるにつれてプロセスの転送コストの影響もそれだけ大きくなっている。

4 まとめ

本研究で、システムを複数のドメインに分割することによってスケーラブルな動的負荷分散を実現できること、動的負荷分散の効果を高めるためにはネットワークの転送速度に応じた適当なドメインの大きさの設定が必要なこと、などを示すことができた。

参考文献

- [1] M.H.Willebeek-LeMair and A.P.Reeves, "Strategies for Dynamic Load Balancing on Highly Parallel Computers", *IEEE Trans. on Parallel and Distributed Systems*, Vol.4, No.9, pp.979-993, 1993.
- [2] A.Elleuch and T.Muntean, "Process Migration Protocols for Massively Parallel Systems", *Proc. of the First International Conference on Massively Parallel Computing Systems*, pp.84-95, 1994.