

分散・並列スーパーコンピューティングのソフトウェアの研究

Software for Parallel and Distributed Supercomputing

(研究プロジェクト番号：JSPS-RFTF 96P00505)

プロジェクトリーダー

島崎 眞昭 京都大学大学院工学研究科・教授

コアメンバー

村岡 洋一 早稲田大学理工学部・教授

國枝 義敏 和歌山大学システム工学部・教授

佐藤 周行 東京大学情報基盤センター・助教授

笠原 博徳 早稲田大学理工学部・教授

山下 雅史 九州大学システム情報科学研究院・教授

城 和貴 奈良女子大学理学部・教授

福田 晃 九州大学システム情報科学研究院・教授

中島 浩 豊橋技術科学大学情報工学系・教授



1. 研究目的

分散・並列環境は、もはや一般的な計算環境として考慮すべき段階になっている。本プロジェクトの開始時には、「分散・並列」とは一部の最先端の計算環境を指す言葉であったことは確かであるが、一般的なネットワークで構築した Beowulf と呼ばれるクラスタシステムが普及しつつある今、そこに現れる計算資源をどのように表現し、かつ制御するかという問題の重要性がますます増大している。本プロジェクトではその開始当初から分散・並列環境が一般的になることを予想して、上記問題にさまざまなアプローチで解決することを試みてきた。その中でも重要なものとして

- (1)分散・並列環境におけるスケジューリング問題
 - (2)プログラマに対し使いやすいビューを提供するための言語処理系やそれに関連するツールの問題
 - (3)プラットフォームの多様性を吸収する実行時環境の構築の問題
- を選び、研究と実装を行うことを目的としてプロジェクトを推進した。

2. 研究成果概要

2.1 分散並列環境スケジューリング問題（九州大学、広島大学、早稲田大学）

アルゴリズム工学的な検討を行い、最適(近似)解の理論的検討を行なった（九州大学、広島大学）。また、GRASP (Greedy Randomized Adaptive Search Procedure) と呼ばれる解法の枠組みに基づくメタヒューリスティック解法の提案と評価を行った。実験の結果から、提案アルゴリズムは与えられたプログラムの並列性がプロセッサ数より大きいときに特に有用であって、決定性のリストスケジューリングが生成する解を最大で8%近く改善することが明らかになった。（広島大学）。

マルチプロセッサスケジューリングアルゴリズムの公平な評価のための一種のベンチマークとして標準タスクグラフセットを開発、提案し、アルゴリズムの評価を行った。またタスクグラフセットを Web で公開している（早稲田大学）。

2.2 自動並列化コンパイラ

2.2.1 学振 MIRAI コンパイラ

言語処理系におけるソースプログラムの内部

表現形式として、本プロジェクトにおいて新たに「統一的中間表現 UIR (Universal Intermediate Representation)」を提案し、同時に自動並列化、最適化のために UIR を操作・変形するための基本 API、応用 API をそれぞれ整備した。これらにより、コンパイラ各部の開発効率が大きく向上できた。本コンパイラは、Fortran77 のサブセットを上述の UIR に変換するフロントエンド、「並列処理可能性を調べる」依存解析部、目的コード生成部からなっている。データ/プログラムの自動分割法を提案し、コンパイラに実装し、検証した点にこのコンパイラの特徴がある（和歌山大学、奈良女子大学）。

2.2.2 プログラム構造の三次元視覚化サブシステム

コンパイラと一体化した統合並列処理環境の提供を目的として、並列化したいプログラムの構造やデータ依存の状況を三次元可視化するツールを作成した（奈良女子大学、岡山大学）。

2.3 メタコンピューティング環境用分散化コンパイラ、実行時環境のための技術開発（早稲田大学）

メタコンピュータとは、ネットワーク上に分散しつつ、かつ単一目的に利用できるコンピュータ群を指す。本研究では、プログラマに対してメタコンピュータを提供することを目標とし、そのための一つのシステムソフトウェアの構成法として、Java を分散処理に適した処理系に拡張する方式を示した。すなわち、Java にスレッド移送、分散オブジェクトの機能を付加することを提案し、その処理系を実装・評価し、その実用性を実証した。分散オブジェクトはネットワークの先にあるオブジェクトを操作する機能で、オブジェクト指向言語から複数コンピュータを活用するための一般的な手段である。実行状態の移送は実行中のプログラムをコンピュータ間で移動させる機能で、複数コンピュータの性能を活用するための動的負荷分散には欠かせない。

本研究で実現したスレッド移送システムは、Java 仮想マシンのプラグインとして実装されていて、Java 仮想マシンの再コンパイルなども不要となり、利用に際しての手間が小さく、実用的価値が大きいものとなっている。

また、プログラム変換で実行状態の移送を実装する既存の方式と異なり、本 Java スレッドの移送システムは、異機種間での非同期な移送を

達成しており、動的負荷分散や、複数スレッドの同時移送に必要な機能を実現している。さらに、性能上不可欠である JIT コンパイラとの共存方式についても提案を行なった。併せて、このスレッド移送機能を採用した応用として移動エージェントシステム (MOBA) を実装し、評価した。その結果、遠隔地にあるデータを利用する分散計算などの環境においては、遠隔データ参照よりもオブジェクト自身を移送させて遠隔地で計算する環境を実現する方が大きく有利であることを示した。

分散オブジェクトシステムの目標の一つは、オブジェクト指向言語のプログラミングモデルを分散処理に適用できるようにし、利用者にネットワークを意識させないようにすることである。オブジェクトがネットワーク上のどこにあっても、同一コンピュータ上のオブジェクトであるかのように扱えるとき、オブジェクトはネットワーク透過であるといえる。本研究では、メタコンピュータの構成要素として、ネットワーク透過な Java 分散オブジェクトシステム (MetaVM) を構築した。既存の分散オブジェクトシステムと異なり、本システムを用いると、ローカルオブジェクトとネットワークの先にあるオブジェクトをまったく同様に扱うことができ、プログラミングの際にコンピュータ群を単一コンピュータとみなすことができる。この、ネットワークおよびプログラミングモデルの透過性を達成するために、実行時コンパイラを用いてバイトコード命令の解釈を変更するという手法を提案し、実装、評価した。また、コストの小さいコード生成手法で実用的な性能を達成し得ることを示した。具体的には、MetaVM の実装を、分散オブジェクト機能のない既存 VM と比較評価した結果、MetaVM には分散オブジェクト機能が付加されているにも拘わらず、遜色のない性能が実現できていることが実証された。

2.4 分散共有メモリ機構及び科学技術計算問題解決環境 (九州大学、東京大学、京都大学)

分散計算環境に対してソフトウェア的に共有メモリビューを提供するソフトウェア方式分散共有メモリシステムを、プロジェクト開始当初は移植性を重視して通信ライブラリ PVM の上に構築することができた。研究後半では、性能を重視して厚いソフトウェア層をスリムにし、TCP レイヤの上に直接構築し、高い性能を得た。ソフトウェアで分散環境の高レイテンシ/低スループットにも対応できるシンプル、かつ高性能なキャッシュシステムの構築に成功した。問題解決環境に関しては、電磁界解析を対象とし WWW ベースのプロトタイプシステム KUPSE (Kyoto University Problem Solving Environment) を開発し、大規模連立一次方程式の並列反復型解法を中心に最新のアルゴリズムの組み込みを行なった (京都大学)。図 1 に KUPSE での電磁界解析結果の可視化例を示す。

2.5 非数値プログラミング環境の研究

不定型・動的データを多用する非数値分野のアプリケーション記述に適した新しい並列プログラミング言語 Orgel を設計し、共有メモリ並列計算機、PC クラスタシステム用の処理系、デバッグシステムを開発し、その有効性を確認した (豊橋技術科学大学)。

3. 結論

専用並列計算機から PC クラスタ、さらにインターネット上の計算機群までを対象とする分散並列環境を対象に、それぞれにおいて、ユーザーに単一システムビューを提供する言語、コンパイラ、問題解決環境などのシステムプログラムおよび関連アルゴリズムの研究とプロトタイプ試作による検証を行いその有効性を確認した。

主な発表論文

- (1) 南里豪志、佐藤周行、島崎真昭：分散共有メモリシステム上にソフトウェアによって構築されたキャッシュシステムの静的制御、情報処理学会論文誌 vol.38(1997), 1859-1868.
- (2) K. Shudo, Y. Muraoka: Asynchronous migration of execution context in Java Virtual Machines, Future Generation Computer Systems, vol.18(2001), 225-233.
- (3) 笹倉万里子、城和貴、國枝義敏、荒木啓二郎：変数オリエンテッドなデータ依存関係モデルの提案、情報処理学会論文誌, vol. 40, No.SIG2 (1999), 45-54.
- (4) Proc. ACM ICS'99 Workshop 1: Scheduling Algorithms for Parallel/Distributed Computing-From Theory to Practice, June 19, 1999.
- (5) T. Tobita, H. Kasahara: A Standard Task Graph Set for Fair Evaluation of Multiprocessor Scheduling Algorithms, to appear in Journal of Scheduling (2002).
- (6) 田中貴文、藤田聡：GRASP 法に基づくマルチプロセッサスケジューリングのためのメタヒューリスティック解法の提案と評価、電子情報通信学会論文誌 D-I vol. J83-D-1 (2000), 919-926.
- (7) T. Iwashita, M. Shimasaki: Parallel Processing of 3-D Eddy Current Analysis with Moving Conductor Using Parallelized ICCG Solve, IEEE Trans. Magnetics, vol.36(2000), 1504-1509.

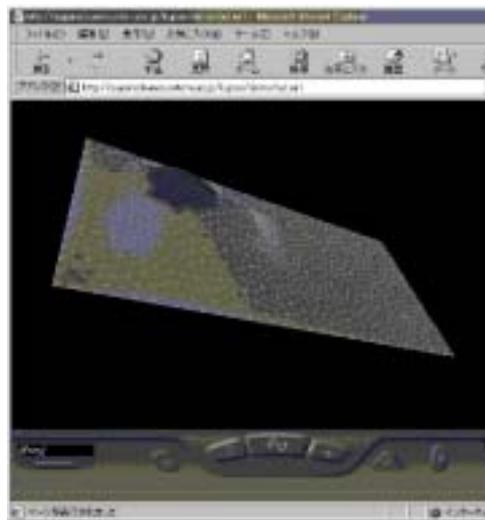


図 1 問題解決環境 KUPSE における計算結果の可視化画面例 (VRML)