

コモディティグリッド技術によるテラスケール大規模数 理最適化

東京工業大学 学術国際情報センター ○松岡 聡

Tera-scale mathematical optimization using the commodity grid technology
Satoshi Matsuoka, Tokyo Institute of Technology

Abstract:

We have constructed the TeraFlops test-bed, 886 processors and 3 or more TeraFlops by the participating research organization (Tokyo Institute of Technology, Tokyo Denki University, and University of Tokushima) total, using the commodity grid technology by the PC technology and the high-speed network, and researched the fault tolerant and scalability of Ninf-G which is a GridRPC system on our test-bed. On our test-bed, we have examined the scalability and performance of our various software of mathematical optimization. Our various software of mathematical optimization is controlled by Ninf-G between each organization and by Ninf-1 in each organization. In an experiment, it was able to succeed in the parallel computing of a maximum of 770 processors among 886 processors, and a scalability and stability were able to be shown.

1. はじめに

PC技術と高速ネットワークによるコモディティグリッド技術により、参加研究機関(東京工業大学・東京電機大学・徳島大学)合算で886プロセッサ・3TeraFlops以上のTeraFlopsテストベッドを構築・整備し、Ninf-Gの耐故障性およびスケーラビリティ拡張の研究開発を行った。また、アプリケーション班が開発した各種数理最適化ライブラリ・アプリケーションを構築したTeraFlopsテストベッド上で実行し、そのプログラミングの容易性、Ninf-GのAPIの検証、並びにそれぞれの最適化問題の並列性の粒度による現状でのスケーラビリティ、絶対性能、並びに資源の不安定性・故障に対するリカバリの種々のスキームを評価した。

2. 研究開発項目とその成果概要

2.1 TeraFlopテストベッド構築

下表のように参加研究機関(東京工業大学・東京電機大学・徳島大学)合算で886プロセッサ・3TeraFlops以上のテストベッドを構築・整備した。各機関内および各機関間はCPU性能・メモリ容量・Linuxカーネルのバージョンなどが異なるヘテロ環境となっているが、各機関間を繋いだ大規模実験に必要なミドルウェアなど(Globus Toolkit, Ninf-G, Ninf, Java, gcc, glibc)のバージョンおよびCAは、実験に支障が生じない程度に合わせた。各機関は機関内のクラスタのゲートウェイノードにGlobusをインストールし、機関間の通信はGlobusを介したインターネット経由で行った。また、実験に用いる各種数理最適化ライブラリ・アプリケーションの計算の制御は、計算の効率を考慮し、全体の制御(各機関間の制御)をNinf-Gで行い、クラスタのローカルな制御(各機関内の制御)はより効率の高いNinf-1で行うハイブリッド制御とした。

	東工大大岡山 キャンパス	東工大すずかけ台 キャンパス	東京電機大学	徳島大学
全プロセッサ数	256nodes/512CPUs	36nodes/72CPUs	56nodes/112CPUs	95nodes/190CPUs
ノード性能	Node-A (160nodes) CPU: Athlon MP 1900+ Memory: 768MB 100Base-T NIC Kernel: 2.4.19 Node-B (96nodes) CPU: Opteron 242 Memory: 2GB 1000Base-T NIC×2 Kernel: 2.4.21	CPU: PentiumIII 1.4GHz Memory: 512MB 100Base-T NIC×2 Kernel: 2.4.2	Node-A (40nodes) CPU: Athlon MP 2400+ Memory: 1 or 2GB 100Base-T NIC×2 Kernel: 2.4.22 Node-B (16nodes) CPU: Opteron 240 Memory: 2GB 1000Base-T NIC×2 Kernel: 2.4.22	Node-A (65nodes) CPU: Athlon MP 2000+ Memory: 512MB 100Base-T NIC Kernel: 2.4.20 Node-B (30nodes) CPU: Athlon MP 2800+ Memory: 512MB 100Base-T NIC Kernel: 2.4.20
Globus	2.4.3	2.2.3	2.4	2.4
Ninf-G	1.1.1	1.1.1	1.1.1	1.1.1
Java	1.4.1	1.4.1	1.4.1	1.4.1
gcc	2.95.4	2.95.3	2.95.3	2.95.3
glibc	2.2.5	2.2.2	2.2.5	2.2.4
CA	ApGrid	ApGrid	ApGrid	ApGrid

2.2 GridRPC/Ninf-Gスケーラブル・高信頼性拡張の研究

GridRPCシステムを用いても、アプリケーションプログラマにとって、組み合わせ最適化問題のアルゴリズム、データ構造などを一から分散実装するという煩雑さの問題は解決されず、その負担は依然として大きい。この問題を解決するために、代表的な数種の並列組み合わせ最適化アルゴリズムのテンプレートを提供し、プログラムの並列化、安全性などの問題をプログラマから隠蔽するシステムjPoPを開発した。jPoPを用いることで、プログラマは問題領域依存なデータ構造や操作を定義するだけで、グリッド上で組み合わせ最適化アプリケーションを容易に開発でき、かつ安全に実行することができる。さらに、実装にJavaを用いることで、さまざまなプラットフォーム上での実行を可能にする。本研究では、jPoPの分枝限定法用のテンプレートクラスであるjPoP-BBの設計およびマスタ・ワーカ方式を用いたプロトタイプの並列実装を行った。さらに、このプロトタイプ上で数値実験を行い、1000プロセッサ環境に向けての改良点などの知見を得た。また、産総研グリッド研究センターで開発されている蛋白質の立体構造に基づく薬物設計を目的とした分子シミュレーションソフトウェア(REMD Toolkit)に対し、負荷分散等の機構を取り入れ、計算機間に性能があるようなグリッド環境においても、効率のよいシミュレーションを行えるような改良を行った。また、計算機性能が均一な環境、不均一な環境、異なるサイト間を広域なネットワークでつなげた環境の3点に関して実験を行い、その改良の有効性を確認した。

2.3 グリッド上数理最適化ライブラリ・アプリケーション(遺伝子ネットワーク推定)の研究

ゲノムプロジェクトの成果として、多くの生物の全DNA配列が決定されてきているが、これは遺伝子の構造や位置などを解析できたに過ぎず、遺伝子の機能、特に複数の遺伝子間の相互作用(遺伝子ネットワーク)が解析できたわけではない。近年、DNA マイクロアレイや DNA チップなどの mRNA の発現量測定技術の発達により、一度に大量の遺伝子の発現の有無を調べることが可能になったことで、発現量のタ

タイムコースデータを得られるようになった。この得られたタイムコースデータから、それを再現する数理モデルを構築することができれば、その数理モデルをもとに遺伝子ネットワークを推定することができると考えられる。これまで数理モデルにべき乗則に基づいた S-system を用いた遺伝子ネットワーク推定手法が考案されているが、この数理モデルは近似を用いたモデルのため、すべてのネットワーク構造を表現できない。本研究では一般質量作用則 (Generalized Mass-Action law; GMA) に基づいた数理モデルを採用し、数理モデルの最適化に人工生命などでよく用いられる遺伝的プログラミング (Genetic Programming; GP) を用いた手法を提案した。数理モデルの解法において、与えられた非線形連立微分方程式をいかに高速に、かつ高精度に解くかが重要なポイントとなる。提案手法のプロトタイプをJavaを用いて実装し、実際の発現量タイムコースデータを用いたテストベッドでの大規模実験を視野に入れ、グリッド環境でJavaプログラミングを支援する並列実行環境Jojoを用いて並列化を行った。小規模の仮想タイムコースデータを用いて計算機実験を行ったところ、問題の規模に応じてスケーラビリティが得られた。また、小規模な問題ではあるが、与えたタイムコースを非常に良く再現するタイムコースが得られた。

2.4 グリッド上数理最適化ライブラリ・アプリケーション(多変数多項式方程式系)の研究

多変数多項式方程式系の全ての孤立解(実根及び複素根)を求めることは、工学的にはシステムと制御、最適化、ロボット工学、計算幾何学、化学工学等に応用を持つ重要な問題である。本研究では、多変数多項式方程式系のすべての孤立解を求める多面体的ホモトピー法の開発を行い、これまでにこの方法を構成する3つのフェイズ(多面体的ホモトピー関数族の構築、予測子・修正子法によるホモトピーパスの追跡、および、解の検証)に対する個々のアルゴリズムの開発とソフトウェアPHOM (Polyhedral Homotopy Method)の公開を行った。また、PHOMの並列版をNinfを用いて、PC クラスタ上に実装し、孤立解の個数が百万を超えるようなより大規模な多変数多項式方程式系を解くことができるようになった。構築したTeraFlopテストベッド(東工大、東京電機大、徳島大の3箇所 で 564CPU) 上でNinf-GとNinf 用い、Cyclic Problem を対象として世界記録($n=14$)の更新を達成した。 $n=14$ の問題をそれぞれのクラスタ単体で解いた場合には、55,207 秒から 98,569 秒程度かかるが、4つのクラスタで問題を分割して解くと、16,197 秒で実行を完了することが出来た。

2.5 グリッド上数理最適化ライブラリ・アプリケーション(並列GAおよびゲノム蛋白質構造最適化)の研究

NMR蛋白質立体構造解析においては、測定データを処理して立体構造を構築していくフェーズにかなりの試行錯誤を要し、専門家の負担が大きいことが問題となっている。本研究では、本フェーズを自動化するための遺伝的アルゴリズムに基づく立体構造解析手法を提案した。提案手法を大規模な問題へ適用するためには、高速化が必要不可欠であり、提案手法において使われている世代交代モデルの並列化およびグリッド化を行った。グリッド化は、グリッドRPCミドルウェアであるNinf-GおよびNinfを階層的に組み合わせることにより実現している。開発したアプリケーションを用いて、78残基の構造決定問題を対象に大規模グリッドテストベッド上での性能評価実験を行った。性能評価実験は、東京工業大学すずかけ台キャンパス、同大岡山キャンパス、東京電機大学、徳島大学の4サイトからなるグリッドテストベッド上で行った。その結果、4サイト770CPUからなるグリッドテストベッド上で、PentiumIII 1.4GHzを用いた逐次計算では約81時間の計算時間を要するところを、約5分33秒で実行することに成功した。

2.6 グリッド上数値最適化ライブラリ・アプリケーション(BMI)の研究

数値最適化問題の一つであるBMI固有値問題求解アプリケーションをベンチマークプログラムとして使用し、大規模グリッドテストベッド上での性能評価実験を行った。本実験では、東京工業大学すずかけ台キャンパス、同大岡山キャンパス、東京電機大学、徳島大学の4サイトに跨るグリッドテストベッド上で、グリッドRPCミドルウェアであるNinf-GおよびNinfを階層的に組み合わせて実装したアプリケーションプログラムを実行した。本実験の結果、4サイト344CPUからなるグリッドテストベッド上で、逐次計算では約9時間半の計算時間を要するベンチマーク問題を約5分で求解することに成功した。

3. ネットワークの活用について

各参加研究機関(東京工業大学大岡山キャンパス・東京工業大学すずかけ台キャンパス・東京電機大学・徳島大学)間でそれぞれ一日数回程度(一回あたり数十MB～数百MB程度)、データのネットワーク転送を行った(尚、東工大大岡山キャンパスーすずかけ台キャンパス間はSuperTITANET経由、東工大ー電機大間・東工大ー徳島大間・電機大ー徳島大間はSINET経由で転送した)。

4. まとめ

本研究では、PC技術と高速ネットワークによるコモディティグリッド技術により、参加研究機関(東京工業大学・東京電機大学・徳島大学)合算で886プロセッサ・3TeraFlops以上のTeraFlopsテストベッドを構築・整備した。各機関内および各機関間はCPU性能・メモリ容量・Linuxカーネルのバージョンなどが異なるヘテロ環境となっているが、各機関間を繋いだ大規模実験に必要なミドルウェアなど(Globus Toolkit, Ninf-G, Ninf, Java, gcc, glibc)のバージョンおよびCAは、実験に支障が生じない程度に合わせた。各機関は機関内のクラスタのゲートウェイノードにGlobusをインストールし、機関間の通信はGlobusを介したインターネット経由で行った。構築したテストベッド上でGridRPCシステムであるNinf-Gの耐故障性およびスケーラビリティ拡張の研究開発を行った。さらに、アプリケーション班が開発した各種数値最適化ライブラリ・アプリケーションを構築したTeraFlopsテストベッド上で実行し、それぞれの最適化問題の並列性の粒度によるスケーラビリティ、絶対性能などを評価した。実験における各種数値最適化ライブラリ・アプリケーションの計算の制御は、計算の効率を考慮し、全体の制御(各機関間の制御)をNinf-Gで行い、クラスタのローカルな制御(各機関内の制御)はより効率の高いNinf-1で行うハイブリッド制御とした。実験では参加研究機関合算で886プロセッサ中、最大770プロセッサの並列計算に成功し、スケーラビリティおよび安定性を示すことができた。現在、関連研究機関である産業技術総合研究所グリッド研究センターの協力を得て、4サイトおよび産総研合算で1000プロセッサ超のテストベッド上で各種数値最適化ライブラリ・アプリケーションの大規模並列実験を行っており、得られた結果を検証中である。

5. 研究開発実施体制

代表研究者 東京工業大学 学術国際情報センター 松岡 聡

研究分担

研究開発項目：TeraFlopテストベッド構築

東京工業大学 教授 松岡 聡

東京工業大学 教授 青木 尊之

東京電機大学 助教授 藤澤 克樹

徳島大学 助教授 小野 功

研究開発項目：GridRPC/Ninf-Gスケーラブル・高信頼性拡張の研究

東京工業大学 教授 松岡 聡

東京工業大学 客員助教授 中田 秀基 (産業技術総合研究所より)

お茶の水女子大学 助手 竹房 あつ子

研究開発項目：グリッド上数理最適化ライブラリ・アプリケーション(遺伝子ネットワーク推定)の研究

科学技術振興機構 研究員 田中 康司

研究開発項目：グリッド上数理最適化ライブラリ・アプリケーション(多変数多項式方程式系)の研究

東京工業大学 教授 小島 政和

東京電機大学 助教授 藤澤 克樹

研究開発項目：グリッド上数理最適化ライブラリ・アプリケーション(並列GAおよびゲノム蛋白質構造最適化)の研究

徳島大学 助教授 小野 功

研究開発項目：グリッド上数理最適化ライブラリ・アプリケーション(BMI)の研究

東京工業大学 助教授 合田 憲人