「ポストペタスケール高性能計算に資するシステムソフトウェア技術の創出」
平成 23 年度採択研究代表者

H27 年度
実績報告書

南里　豪志

九州大学情報基盤研究開発センター
准教授

省メモリ技術と動的最適化技術によるスケーラブル通信ライブラリの開発

§1. 研究実施体制

(1)「インタフェース」グループ
①研究代表者:南里　豪志 （九州大学情報基盤研究開発センター、准教授）
②研究項目
・隣接通信インタフェースの実装
・非ブロッキング集団通信インタフェースの実装
・隣接・集団通信の動的最適化技術の開発
・スケーラブルな通信ライブラリの実装と公開

(2)「プロトコル」グループ
①主たる共同研究者:住元　真司(富士通株式会社次世代 TC 開発本部、シニアアーキテクト)
②研究項目
・通信バッファを削減した通信モデルにもとづいた通信プロトコル

(3)「通信路制御」グループ
① 主たる共同研究者:柴村　英智 （財団法人九州先端科学技術研究所次世代スーパーコンピュータ開発支援室、研究員）
②研究項目
・パケット送信間隔動的最適化技術
・Exa FLOPS 環境のアプリケーション性能予測技術

(4)「アプリケーション」グループ(平成 27 年 8 月 7 日より、「インタフェース」グループに統合)
① 主たる共同研究者:高見　利也 （九州大学情報基盤研究開発センター、准教授）

②研究項目

・非ブロッキング集団通信と遠隔 Atomic 通信を活用した OpenFMO の開発と評価

・隣接通信を活用した電磁流体プログラムの開発と評価

・既存アプリケーションの隣接通信、非ブロッキング集団通信による改良

・ExaFLOPS 環境に向けた高スケーラブルなアプリケーション作成技術の確立

§2. 研究実施の概要

本年度実施した主な研究活動は、省メモリ型通信ライブラリ ACP (Advanced Communication Primitives)の実装改良、およびアプリケーションへの応用技術の研究開発、通信動的最適化技術の研究開発、ネットワークシミュレータ NSIM-ACE の精度評価、パケットペーシングの有効性検証である。

ACP の実装改良については、まず、ACP 基本層の上に構築する分散データ構造インタフェースの見直しと改善を行った。特に、グローバルメモリアロケータのアルゴリズムについて、初期実装では片方向リストとフリーリストで保持していたため、メモリ解放時に空きメモリ領域数に比例する計算量が必要になっていた。これに対し本年度は、隣接領域を直接調べて空き領域かどうかを判定するアルゴリズムに改良することで、計算量を O(1)に削減できた。また、省メモリ型のメッセージパッシング通信を実現するチャネルインタフェースについて、以前は、メッセージサイズが大きい場合のプロトコルとして Rendezvous を採用していたため、領域の登録と解放に要する時間により十分な帯域幅が得られていなかった。そこで本年度は、メッセージサイズが大きい場合のプロトコルとして、パイプライン型 Eager を採用した。これは、複数のスロットにデータを分割して送付するもので、これによりネットワークのピークバンド幅に近い性能が得られることを確認した。

アプリケーションへの応用技術については、まず電磁流体プログラムにおける stencil 計算の隣接通信について、簡単に記述するための Halo 通信モデルを作成し、さらにその通信を計算とオーバラップさせるための通信スレッドの実装を行った。これにより、スケーラビリティの向上を確認した。現在の実装は、MPI によるプロトタイプであり、今後 ACP での実装を予定している。次に重力 N 体シミュレーションについては、ACP の片側通信を活用して、粒子の移動に伴う通信を効率よく実現することを示した。さらに、従来の MPI で記述されたアプリケーションに対して必要最小限の書き換えで省メモリでの大規模並列実効を可能とするため、複数の MPI 並列プログラムを ACP により接続する機構を提案し、実装した。

通信動的最適化技術については、複数の NIC を活用した隣接通信最適化技術について、評価と性能解析を行い、有用性を検証した。また、集団通信の実装アルゴリズム選択技術について、プログラム実行中の状況を片側通信でモニタリングし、性能変動を検知したらアルゴリズムを再選択する非同期動的最適化機構を構築した。実験の結果、提案手法により、負荷バランスの変動などによる性能の変化を低オーバヘッドで検知することができることを示した。これにより、実行中に挙動や環境が変化する場合の動的アルゴリズム選択を効率的に行える。

ネットワークシミュレータ NSIM-ACE については、RDMA 通信のシミュレーション機能についての有用性を検証するため、重力 N 体シミュレーションにおける通信を ACP で実装した際の通信パターンについてシミュレートし、実測値との比較を行った。その結果、実測値と同等の傾向をシミュレーションで再現できることを示した。

パケットペーシングについては、Fujitsu FX10 上のランダムリング通信、および NICAM の通信パターンをシミュレートし、パケットペーシングにより通信衝突の影響を低減できることを示した。

【代表的な原著論文】

- Yuichiro Ajima, et al, "ACPdl: Data-Structure and Global Memory Allocator Library over a Thin PGAS-Layer", Proceedings of the First International Workshop on Extreme Scale Programming Models and Middleware, pp. 11-18, 2015
- 森江 善之, 南里 豪志, "直接網において複数の通信デバイスを有効に使用する隣接通信アルゴリズムの提案", 2015 ハイパフォーマンスコンピューティングと計算科学シンポジウム論文集, 2015.
- Kin'ya Takahashi, Sho Iwagami, Taizo Kobayashi, Toshiya Takami, “Theoretical Estimation of the Acoustic Energy Generation and Absorption Caused by Jet Oscillation”, J. Phys. Soc. Jpn., Vol.85, No.4, Article ID: 044402