日本オペレーションズ・リサーチ学会
2019年 秋季研究発表会

クラウドHPCのネットワークパフォーマンスがParaNUOPTの性能に与える影響の比較結果

NTT データ数理システム　*石橋保身　ISHIBASHI Yasumi
01207140　Zuse Institute Berlin　品野勇治　SHINANO Yuji

1. 概要

スーパーコンピュータのような分散メモリアーキテクチャ上で動作するソフトウェアの性能は単一計算ノードの性能だけでなく，計算ノードを繋ぐインターコネクトの性能も重要である．コンピュータの浮動小数点演算の性能を計測する High Performance LINPACK (HPL) の結果をランキングしている TOP500 の上位 100 件によると，有線 LAN で最も普及している規格の Ethernet が使われているコンピュータは僅か 3 件であり，Ethernet よりも広帯域で低遅延な特徴を持つ Infiniband が使われているコンピュータは 32 件である [1]．一方，クラウドの普及により誰でも PC クラスタを構築することが可能になっているが，Infiniband のような高性能なインターコネクトを提供しているベンダーは少ないため，通信の少ないソフトウェアを設計して開発することは重要である．本発表では Infiniband を提供している Microsoft Azure を使い，ネットワークパフォーマンスが分散並列分枝限定法ソルバ ParaNUOPT の性能に与える影響を確認し，その結果を報告する．

2. UG と ParaNUOPT

品野によって開発された Ubiquity Generator framework (UG) は分枝限定法ソルバを並列化させるソフトウェアである．Zuse Institute Berlin の SCIP [2] や FICO Xpress [3] といった混合整数線形計画 (MIP) ソルバが UG によって既に並列化されていて，それぞれ ParaSCIP，ParaXpress と呼ぶ [2, 3, 4, 5]．ParaSCIP や ParaXpress による計算は高性能なネットワークパフォーマンスを持つスーパーコンピュータ上でおこなわれており，MIP のベンチマーク問題を整備するプロジェクト MIPLIB2017 に登録されている未解決問題をいくつも解いている．

ParaNUOPT は NTT データ数理システムの MIP ソルバ Numerical Optimizer [4] を UG によって並列化したものであり，ParaSCIP や ParaXpress と同じく，MIPLIB2017 の未解決問題を解くことに成功している[5]．特に，計算環境は Ethernet を用いた PC クラスタであり，後述する UG の仕組みも考慮すると，ネットワークパフォーマンスを意識することなく ParaNUOPT を使用できると考えられる．

3. UG の仕組み

UG は分枝木にある未探索な部分問題を解く Worker と，負荷分散と Worker 間のメッセージを仲介する Supervisor が協調動作する並列化手法を提供している．Worker は分枝限定法ソルバであり，Supervisor から送られてくる部分問題を分枝限定法で解く．Supervisor は Worker に部分問題を渡す役割を担っているが，Supervisor が部分問題を生成するわけではなく，計算負荷が大きい Worker から未探索な部分問題を受け取り，それを他の Worker に渡す．つまり，Supervisor は一時的に部分問題を保持するバッファとして機能し，各 Worker が分枝木の部分木を持っている (図 1)．

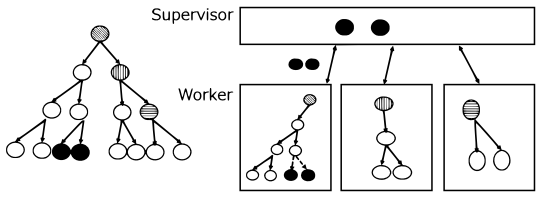

図 1: UG による並列処理

ネットワークパフォーマンスが ParaNUOPT の性能に与える影響を確認するため，UG が Super-

[1] https://www.top500.org/list/2019/06/?page=1
[2] https://scip.zib.de/
[3] https://www.fico.com/jp/products/fico-xpress-solver
[4] https://www.msi.co.jp/nuopt/
[5] https://miplib.zib.de/instance_details_rococoC11-010100.html

表 1: UG によって通信されるメッセージ

メッセージ	サイズ	タイミング
Worker Status	60 bytes	定期
上界値	8 bytes	不定期
暫定解	変数の数に依存	不定期
部分問題	主に変数の数に依存	不定期

visor と Worker の間で通信させるメッセージを表 1 にまとめた．Worker Status は Supervisor が動的な負荷分散をおこなうために下界値や探索済みの部分問題数といった情報を含んでいる．このメッセージはデフォルトの設定では 1 秒に 1 回送信されるようになっているが，そのメッセージサイズは 60 bytes であり，非常に小さい．上界値や暫定解は Worker がより良い解を発見した際に通信されるメッセージであり，頻繁に通信されるようなメッセージではない．部分問題は Worker が処理するタスクであり，Supervisor が保持する部分問題が少ない時やアイドルな Worker が存在する時に通信されるメッセージである．このメッセージには Worker による前処理や分枝によって変化した変数の上下限が含まれている．

部分問題の定義はソルバに応じて変更可能である．ParaNUOPT の場合は元の問題との変数の上下限の差分に加えて Worker によって追加された切除平面も含んでいるため，インスタンス毎にメッセージサイズを正確に見積もることは困難ではあるが，少なくとも変数の数に依存している．

以上をまとめると，定期的に通信されるメッセージは Worker Status のみであり，そのメッセージサイズは非常に小さい．従って，ネットワークパフォーマンスが ParaNUOPT の全体的な性能に与える影響は小さいと考えられる．

4. 計算実験

ネットワークパフォーマンスが ParaNUOPT の性能に与える影響を確認するため，クラウド上にネットワークパフォーマンスの異なる PC クラスタを構築する．具体的には Microsoft Azure が提供している HPC 用途の仮想マシン H16 と H16r を用いる．どちらも Intel Xeon E5-2667 v3 Haswell 3.2 GHz (16 コア) の CPU と DDR 4 メモリを搭載している．H16 の正確なネットワークパフォーマンスは公表されていないが，H16r では Message Passing Interface (MPI) トラフィックに特化された Remote Direct Memory Access (RDMA) を利用した InfiniBand が提供されている [6]．実際に H16r を使った HPL の計算実験によると 32 個の計算ノードで 28.33 倍の加速が達成されているため [1]，高性能な PC クラスタをクラウド上に構築できると考えられる．細かな実験の設定や実際の計算結果については当日に報告する．

参考文献

[1] M. Mohammadi and T. Bazhirov. “Comparative benchmarking of cloud computing vendors with high performance linpack”. In: *Proceedings of the 2nd International Conference on High Performance Compilation, Computing and Communications*. ACM. 2018, pp. 1–5.

[2] Y. Shinano, T. Achterberg, T. Berthold, S. Heinz, and T. Koch. “ParaSCIP: a parallel extension of SCIP”. In: *Competence in High Performance Computing 2010*. Springer, 2011, pp. 135–148.

[3] Y. Shinano, T. Achterberg, T. Berthold, S. Heinz, T. Koch, and M. Winkler. “Solving open MIP instances with ParaSCIP on supercomputers using up to 80,000 cores”. In: *Parallel and Distributed Processing Symposium, 2016 IEEE International*. IEEE. 2016, pp. 770–779.

[4] Y. Shinano, T. Berthold, and S. Heinz. “A first implementation of ParaXpress: Combining internal and external parallelization to solve MIPs on supercomputers”. In: *International Congress on Mathematical Software*. Springer. 2016, pp. 308–316.

[5] Y. Shinano, T. Berthold, and S. Heinz. “ParaXpress: an experimental extension of the FICO Xpress-Optimizer to solve hard MIPs on supercomputers”. In: *Optimization Methods and Software* 33.3 (2018), pp. 530–539.

[6] https://azure.microsoft.com/ja-jp/blog/availability-of-h-series-vms-in-microsoft-azure/