

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 29 日現在

機関番号：12201

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24500055

研究課題名(和文) 超大規模相互結合網における自律分散先行制御方式の研究

研究課題名(英文) Study on Self-Lookahead Control Mechanisms in Large-Scale Interconnection Networks

研究代表者

横田 隆史 (Yokota, Takashi)

宇都宮大学・工学(系)研究科(研究院)・教授

研究者番号：90334078

交付決定額(研究期間全体)：(直接経費) 4,200,000円

研究成果の概要(和文)：結合網の輻輳状況に関する情報をメッセージパケットの転送と並行して流通させる機構およびこれにより先行的に輻輳緩和の制御を行う手法(2手法)を考案した。シミュレーションにより各々の有効性を示した。さらにセルオートマトンの計算原理を導入したシミュレーション手法を考案し、GPGPUにより100万ノード規模の相互結合網でも実用的な時間で模擬・評価できる基盤技術を確立した。

研究成果の概要(英文)：We have proposed two novel control methods that efficiently reduce congestion situation. The proposed methods are based on the mechanisms that virtually conduct congestion information in parallel with message-packets. Both methods show outstanding performance in reducing congestion situation and, thus, enhance the network performance. Furthermore, we have also introduced the cellular automata principle into large-scale network simulation, which is suitable with GPGPU environment and the resulting simulator allows us to simulate upto 1 million nodes within practical time.

研究分野：計算機システム

キーワード：計算機システム 並列計算機 計算機アーキテクチャ 相互結合網 輻輳制御

1. 研究開始当初の背景

絶対的な性能を指向している大規模並列スーパーコンピュータ・システムでは、性能向上のために並列性の向上が求められ続けている。現在の超大規模並列システムでは計算ノードを数 10 万の規模で擁することが求められており、将来的には数 100 万ノード規模での構築を可能にする技術の研究開発が不可欠である。

このような超大規模並列システムでは、計算ノード間を単に集積するだけでは解決しない現実的な課題が山積している。特に通信系は並列計算機の成否の大きな鍵を握る重要な要素であり、その課題は主に以下の 3 点にまとめられる。(1) 実行性能: 通信の性能、制御・管理のための機能性能を高くすること、(2) 実現性: 物理的に構築でき、コストや消費電力を現実的な範囲に抑えられること、(3) 耐故障性: 多少の構成要素が故障しても運用を続けられること。

現実の超大規模システムでは、上記の各課題に対してリッチな通信系を提供することで解決を図ってきた。具体的には二分幅 (bisection width) に優れた fat-tree や高次元のメッシュ/トラスのトポロジを採用することで性能や耐故障性の問題に対処してきた。しかしこれらの技術もノード数の増大に伴い見直しの必要に迫られている。

研究代表者はこれまで一貫して超大規模化に伴う通信系の問題の解決に取り組んできた。なかでも輻輳の制御は中心的なテーマであり、大規模化の観点から集中制御を排しながら同時にシステム全体での大域的な観点での最適制御を目指して検討を進めてきた。その結果、相互結合網の具体的な制御手法として、局所的な輻輳の状況を通信パケットと並行して伝播させることを基本アイデアとした適応ルーティング手法 "Cross-Line" を考案し、様々な発展手法も検討してきた。

また、モデル化や他分野のアプローチも積極的に導入し問題の解決を図ってきた。評価手法の標準化をはかり定量的評価指標を提案した。さらに、セルオートマトン等による物理学的アプローチを導入し輻輳の数理的モデルを導出する試みも行い成果を得ることができた。また、シミュレータによる実験的な手法により、規模と通信パターンの組み合わせに対し、ルーティングアルゴリズムごとの性能および特徴を抽出した。

相互結合網における輻輳は、その拡大と解消の速度に桁違いの差があり規模の拡大とともに深刻化する。一旦発生した輻輳は容易には解消しないため、最大性能を得るには相互結合網全体を輻輳が発生しない上限の条件で運用する必要がある。研究代表者による上述の先行研究でも一定の効果を得ることができたが、ルーティングアルゴリズムを初めとする従来型の制御手法では、輻輳の程度を緩和することは可能であるがその効果は

限定的であることが明らかになった。つまり症状が生じてからの応答になり対応が後手になり十分な効果を得にくい。将来的に必要なが見込まれる超大規模相互結合網にはなおブレークスルーが必要である。

以上の経緯から、本研究では、症状を検出してからの対症療法的なアプローチを改め、先行して制御を行うことにより症状の発生そのものを抑止する着想に至った。すなわち、症状が現れる前に適切な制御を施す (先行制御) ことである。超大規模並列環境への適用を想定し相互結合網の各構成要素による自律分散機能の集積として実現する条件の下で、症状が現れる前に先行的に制御する機能を求める。これまででない先行制御を自律分散により実現することが本研究でのチャレンジである。

2. 研究の目的

大規模レギュラーネットワークにおける通信の輻輳制御において、様々な通信状況に対応可能であり、かつ、効果的な手法を創出することを第一の目的とする。またこれに付随して、大規模網でのシミュレーション評価実験の過程を加速するためのシミュレーションの高速化を第二の目的としている。

第一の目的の検討にあたり核概念となるのは、上述の「先行制御」である。本課題で対象としている大規模レギュラーネットワークに先行制御を適用するにあたって、有効な制御指標を明らかにしたうえで、効果的な制御手法を検討・創出することを目的とする。

第二の目的については、代表者の既往研究により基本部分の着想を得ている。このため、本課題では、GPGPU への適用など実践的な観点からの検討を進め、100 万ノード規模のシミュレーションを現実的な時間で行うことを目的とする。

3. 研究の方法

先行制御により相互結合網を輻輳が発生する限界の領域で運用するアイデアは、既往研究にもほとんど例がないため、特に本課題の初期段階では、代表者の既往研究の成果に基づきながら探索的に検討を行い、必要に応じてシミュレーション実験の環境を整備し実験評価を行う。本課題の全研究期間を通して、後述する 5 項目の検討を順次行っている。

本課題での検討にあたっては、様々な通信状況に対応することを前提とする。相互結合網の研究では、特定の通信パターン (ランダム通信など) の下で継続的に通信を行う状況のみを前提にするなど、限られた状況のみで実験評価を行う事例が非常に多い。本課題においては、通信パターンを転置、シャッフル、ビット反転、ビット逆順、ビット回転、トルネード、ランダム、ランダム・ペアの 8 種類を想定し、すべてのパターンで実験評価を行うこととする。また、継続的な定常通信だけ

ではなく、各ノードが一定数のパケットを送信するだけの一斉同期通信も対象とする。

本研究課題は相互結合網の輻輳制御について基礎的な知見を得ることを主眼とする。このため複雑な構成を前提とせずシンプルで規則性の高い接続形態による大規模網をもとに検討を行う。具体的には2次元トラス網を用いる。

本課題での検討項目は以下のとおりである。

(1) 輻輳情報の伝播高速化の基本検討

検討の第一段階として、輻輳制御にかかわる情報のみを通常パケットの2倍・4倍・8倍の速さで伝播させるときの効果を、適応ルーティングを対象として調べる。これにより、ルーティング制御に必要な情報を高速に伝播させること自体の効果を検証する。

(2) 流量制御の基本的検討

高速道路における交通流の制御手法である「ランプメータリング」に着目し、同様の流量制御手法を大規模相互結合網に適用する手法を検討する。シミュレーション実験により効果を検証する。

(3) 超大規模相互結合網シミュレーションの高速化

相互結合網のシミュレーションの考え方をセルオートマトンの考えに基づいて大幅に変更することで、GPGPUでの処理に適合した高効率のシミュレーション手法を検討する。

(4) 準広域情報に基づく流量制御

流量制御の基本的検討の結果に基づき、準広域情報に基づいた流量制御手法を検討する。物理学の粒子系の研究において、時空間図(space-time diagram)を導入する。この視覚効果により、遠方で発生した輻輳の種(seed)が時間の経過とともにパケットの進行方向とは逆の方向に移動すること、したがってこの輻輳の種を検知することで予防的な制御が可能であることを示す。

(5) 広域情報に基づく流量制御

大規模な相互結合網全体にわたる広域情報(局所的(準広域)ではない)を効果的に収集・通知する手法を検討し、効果的な流量制御を行う。本項目については、代表者による先行研究の成果であるエントロピーの概念を導入する。継続的にエントロピー値の収集し、その経時変化を検知することでシステム全体の輻輳状況を把握・予想し、適切な輻輳制御を行うメカニズムを創出する。

4. 研究成果

(1) 輻輳情報の伝播高速化の基本検討

代表者による既往研究の成果である適応ルーティング手法"Cross-Line"では各ノ

ードのバッファの状況をビットマップ表現にまとめ、次元方向に伝播することでパケットの進行方向の輻輳状況を検出できる仕組みがある。この情報を2倍・4倍・8倍の速度で伝播できるようシミュレータを改変し、その効果を確認した。

情報の伝播速度が速いほど現在の状況に迅速に対応することが可能であるため、効果が大きいものと期待されたが、評価実験の結果、転送パターン等の条件にもよるが、必ずしも劇的な効果が得られないことが判明した。このため、本研究課題の以降において、適応ルーティング(通信経路制御)ではなく、流量制御を中心に議論することとした。

(2) 流量制御の基本的検討

交通流に対するランプメータリングから着想した流量制限の手法について基礎的な検討を行った。2次元トラス網において、次元方向に直進する経路を、高速道の本線に対応するものとし、本線への流入制限を行う手法をシミュレータにより検討した。

ノードからのパケット流入制限、つまりノードで生成されたパケットを結合網に投入するのを待たせることが最も効果がある。次いで次元進行方向が変わる経路での流入制限となるが、この場合、デッドロックを発生する可能性があることが判明した。

(3) 超大規模相互結合網シミュレーションの高速化

相互結合網のルータをセルオートマトンにおけるセルと対応させる。これにより近接ルータ間でのパケットのやり取りは、セルオートマトンでの近傍セル間でのやり取りに置き換えられる。基本的に、各セルの動作は独立であるから巨大な並列性が抽出できる。また、各セルの状態更新を、現在のセルの状態に基づき表形式で求める手法を確立した。この手法は、全演算装置が同一の操作を行う並列処理形式(SIMD形式)と良好に適合する。

この処理形式はGPGPUと大変よく適合するため、専用の記述言語環境CUDAを用いて相互結合網シミュレータを構築した。実際のGPUデバイスの構成や処理方式に合わせて種々の最適化を検討し、適用した。この結果、大幅な性能向上を果たせることを実証し、100万ノード規模のシミュレーションでも実用的な時間で終了することを示した。

(4) 準広域情報に基づく流量制御

局所的な輻輳が発生した場合でも、その輻輳は同じ場所に留まるとは限らない。一般的には、進行方向と逆の方向に伝播する。この事実は物理学における粒子系の研究で知られていたことであり、時空間図により直感的にも把握されていたことである。本研究課題では、こうした物理学での成果を相互結合網の輻輳制御のために導入した。

すなわち、遠方で発生した局所的な輻輳現

象は、輻輳伝播の現象により近い将来に当該地点に伝播してくる可能性が高い、との予測が可能になる。前出の"Cross-Line"では輻輳状況を次元軸方向に伝播させる機能があるが、この機能を発展させ、流量制限に用いることにした。遠方で発生した輻輳が当該地点に波及してくるのは、当該地点を通過し輻輳地点の方向に向かうパケットがあるためである。この方向のパケットの流量を制限すれば輻輳は伝播せずに解消の方向に向かう。

こうした原理に基づく流量制限手法を State Propagation Throttling (SPT: 状態伝播スロットリング)として提案し、その有効性をシミュレータにより示した。

(5) 広域情報に基づく流量制御

相互結合網内での転送負荷が大きくなりパケット間での衝突の頻度が高くなると、それに応じてルータ内での待ちが生じ多くのパケットが結合網内に滞留することになる。この度合いはパケットの平均自由行程をもとにエントロピーの概念で定量化することができる。相互結合網でのエントロピーは既に代表者の先行研究により提案されており、これを基にした輻輳制御手法の基本形(エントロピースロットリング)が提案されていた。本課題では、先行制御の概念に基づいてエントロピースロットリング手法を拡大した。先行研究ではランダム・定常通信のみに適用していたが、本課題での研究方針に基づき、種々の転送パターンならびに定常・非定常通信に適用し、ほとんどの場合で顕著な改善が得られることを示した。

5. 主な発表論文等

[雑誌論文](計10件)

Takashi Yokota, Kanemitsu Ootsu, Takeshi Ohkawa, Relaxing Heavy Congestion by State Propagation, Journal of Information Processing, 査読有, Vol.23, No.5, pp.730-743, 2015.

DOI: 10.2197/ipsjip.23.730

<http://id.nii.ac.jp/1001/00145375/>

Takashi Yokota, Kanemitsu Ootsu, Takeshi Ohkawa, Entropy Throttling: Towards Global Congestion Control of Interconnection Networks, Proc. 2015 Third International Symposium on Computing and Networking (CANDAR 2015), 査読有, pp.40-49, 2015.

DOI: 10.1109/CANDAR.2015.58

Yuki Suzuki, Takashi Yokota, Kanemitsu Ootsu, Takeshi Ohkawa, Performance Improvement of Large-Scale Interconnection Network Simulator by Using GPU, Proc. 2015 Third International Symposium on Computing and Networking (CANDAR 2015), 査読有, pp.585-589, 2015.

DOI: 10.1109/CANDAR.2015.98

鈴木裕樹, 横田隆史, 大津金光, 大川猛, 大規模相互結合網シミュレータのGPUを用いた高速化, 電子情報通信学会技術報告, 査読無, Vol.115, No.174,

pp.97-102, 2015.

<http://www.ieice.org/ken/paper/20150804ybb8/>
Takashi Yokota, Kanemitsu Ootsu, Takeshi Ohkawa, Cellular Automata as Acceleration Kernel of Interconnection Network Simulation, Proc. 2014 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2014), 査読有, Vol.2, pp.561-567, 2014.

http://worldcomp-proceedings.com/proc/proc2014/pdpta/PDPTA_Papers.pdf

Takashi Yokota, Kanemitsu Ootsu, Takeshi Ohkawa, Performance Impact of Arbitration Functions of Interconnection Network Router for Steady/Unsteady Communications, Proc. 2014 International Conference on Computational Science and Computational Intelligence (CSCI 2014), 査読有, Vol.1, pp.28-33, 2014.

DOI: 10.1109/CSCI.2014.13

鈴木裕樹, 横田隆史, 大津金光, 大川猛, セルオートマトン手法を用いた大規模相互結合網シミュレータのGPUによる高速化, 情報処理学会第76回全国大会講演論文集, 査読無, Vol.2014, No.1, pp.199-200, 2014.

<http://id.nii.ac.jp/1001/00104316/>

Takashi Yokota, Kanemitsu Ootsu, Takeshi Ohkawa, A Cellular Automata Approach for Large-Scale Interconnection Network Simulation, Proc. 2013 First International Symposium on Computing and Networking (CANDAR 2013), 査読有, pp.545-551, 2013.

DOI: 10.1109/CANDAR.2013.97

横田隆史, 大川猛, 大津金光, 馬場敬信, 流量制限による相互結合網の実効性能改善の検討, 情報処理学会研究報告, 査読無, Vol.2012-ARC-202, pp.1-12, 2012.

<http://id.nii.ac.jp/1001/00087389/>

澤野彦太郎, 横田隆史, 大津金光, 大川猛, 馬場敬信, GPGPUによる大規模相互結合網シミュレーション, 情報処理学会第75回全国大会講演論文集, 査読無, Vol.2013, No.1, pp.105-106, 2013.

<http://id.nii.ac.jp/1001/00110856/>

[学会発表](計10件)

Takashi Yokota, Entropy Throttling: Towards Global Congestion Control of Interconnection Networks, 2015 Third International Symposium on Computing and Networking (CANDAR 2015), 2015-12-09, 札幌市産業振興センター(札幌市)

Yuki Suzuki, Takashi Yokota, Performance Improvement of Large-Scale Interconnection Network Simulator by Using GPU, 6th International Workshop on Advances in Networking and Computing (IWANC 2015), 2015-12-09, 札幌市産業振興センター(札幌市)

鈴木裕樹, 横田隆史, 大規模相互結合網シミュレータのGPUを用いた高速化, 電子情報通信学会コンピュータシステム研究会, 2015-08-04, 別府国際コンベンションセンター(別府市)

Takashi Yokota, Cellular Automata as

Acceleration Kernel of Interconnection Network Simulation, 2014 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2014), 2014-07-24, Las Vegas (米国ネバダ州)

Takashi Yokota, Relaxing Heavy Congestion by State Propagation ---Preliminary Discussion---, Annual Meeting on Advanced Computing System and Infrastructure 2015 (ACSI 2015), 2015-01-28, つくば国際会議場(つくば市)

Takashi Yokota, Performance Impact of Arbitration Functions of Interconnection Network Router for Steady/Unsteady Communications, 2014 International Conference on Computational Science and Computational Intelligence (CSCI 2014), 2014-03-10, Las Vegas (米国ネバダ州)

鈴木裕樹, 横田隆史, セルオートマトン手法を用いた大規模相互結合網シミュレータのGPUによる高速化, 情報処理学会第76回全国大会, 2014-03-11, 東京電機大学(足立区)

Takashi Yokota, A Cellular Automata Approach for Large-Scale Interconnection Network Simulation, 1st International Workshop on Applications and Fundamentals of Cellular Automata(AFCA 2013), 2013-12-06, ひめぎんホール(松山市)

横田隆史, 流量制限による相互結合網の実効性能改善の検討, 情報処理学会計算機アーキテクチャ研究会, 2012-12-13, 北海道大学(札幌市)

澤野遼太郎, 横田隆史, GPGPUによる大規模相互結合網シミュレーション, 情報処理学会第75回全国大会, 2013-03-06, 東北大学(仙台市)

[その他]

ホームページ等

<http://aquila.is.utsunomiya-u.ac.jp/~yokota>

6. 研究組織

(1) 研究代表者

横田 隆史 (YOKOTA Takashi)

宇都宮大学・大学院工学研究科・教授

研究者番号：90334078

(2) 連携研究者

大津 金光 (OOTSU Kanemitsu)

宇都宮大学・大学院工学研究科・准教授

研究者番号：00292574

大川 猛 (OHKAWA Takeshi)

宇都宮大学・大学院工学研究科・助教

研究者番号：80392596

馬場 敬信 (BABA Takanobu)

宇都宮大学・大学院工学研究科・教授

研究者番号：70092616

(平成24年度まで)