

A Scalable Routing Methodology for Low-Latency Interconnection Networks

Ryuta Kawano

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

School of Science for Open and Environmental Systems
Graduate School of Science and Technology
Keio University

December 2017

報告番号	Ⓐ 乙 第	号	氏 名	河野 隆太
主 論 文 題 名： A Scalable Routing Methodology for Low-Latency Interconnection Networks (低遅延相互結合網のためのスケーラブルなルーティング手法)				
(内容の要旨) ネットワークのエンドノード間通信遅延は、高性能計算機上で実行される並列アプリケーションにおける重要な課題となっている。次世代の HPC システム上で実行される大規模並列アプリケーションの場合、MPI 通信遅延は 1 マイクロ秒未満であることが必要とされている。スイッチ経由による遅延は InfiniBand QDR では約 100 ナノ秒と、シリアル/パラレルコンバータを含む配線遅延・フリット注入遅延よりも通常大きい。このため、スイッチ間トポロジは、直径や平均最短距離が小さいことが求められている。 この要求に対処するため、近年の研究により、スイッチ間ネットワークに適用可能なランダムトポロジが開発された。従来の Torus や Fat-tree といったネットワークトポロジと比較して、これらのトポロジはホップ数を大幅に削減し、HPC システムやデータセンター、及びメニーコアシステムに効率的に適用できることが示されている。 こうしたランダムトポロジの実用化に向けては、スケーラビリティの面で課題が存在している。第 1 に、ランダムトポロジはキャビネット間の総配線延長を増加させる傾向があり、コスト増加につながってしまう。第 2 に、各スイッチが保持するテーブルエントリの数が増加する。第 3 に、デッドロックフリーを保証するために、仮想チャネル (Virtual Channels; VCs) をルーティング経路に割り当てるための時間計算量が増加する。 これらのスケーラビリティに関する問題を解消するため、本論文では、配線長制限を課したランダムリンクを用いたトポロジに着目する。このネットワークは、完全にランダムなネットワークとほぼ同等のホップ数を達成しながら、そう配線延長を大幅に削減可能である。 本論文の目的は、先述の配線長制限ランダムトポロジに適用可能なスケーラブルなルーティング手法の探求であり、低遅延な相互結合網の実用化に向けた取り組みである。まず、配線長制限ランダムトポロジ向けのスケーラブルなルーティング手法である LOREN (Layout-Oriented Routing with Entries for Neighbors) を提案する。この方法は、配線長制限ランダムトポロジの持つ不規則性と局所性を利用し、ノード間のホップ数を少なくしつつ、必要なルーティングテーブルサイズを削減可能である。				

第 2 に，仮想チャネルの割り当て方法である ACRO (Assignment of Channels in Reverse Order) を提案する．この手法は，従来の VC 割り当て方法と比較して，同程度の VC 数を必要としつつ，時間計算量を大幅に削減可能である．

本論文で提案されるルーティング手法は，通信遅延と必要テーブルエントリ数とのより良いトレードオフを達成する．提案ルーティング手法である LOREN に提案 VC 割り当て手法である ACRO を適用し，デッドロックフリーを保証した上で，サイクル・アキュレートなネットワークシミュレータを用いて性能評価を行った．その結果，提案手法は従来のコンパクトルーティング方法と比較して，ネットワークスループットを最大 67.9%向上させた．さらに，必要なルーティングテーブルエントリ数が最大 91%削減され，実装上のスケーラビリティと柔軟性が向上した．

Thesis Abstract

No. 1

Registration Number	<input checked="" type="checkbox"/> “KOU” <input type="checkbox"/> “OTSU” No. _____ *Office use only	Name	Ryuta Kawano
Thesis Title A Scalable Routing Methodology for Low-Latency Interconnection Networks			
Thesis Summary <p>End-to-end network latency has become an important issue for parallel application on large-scale High Performance Computing (HPC) systems. For large parallel applications executed on the next generation of HPC systems, MPI communication latency should be lower than one microsecond. Switch delays (e.g., about 100 nanoseconds in InfiniBand QDR) are typically larger than the wire and flit injection delays, even when including serial and parallel converters. This is why inter-switch topologies should have a low diameter and low average shortest path length, both of which can be measured in terms of a number of switch hops.</p> <p>In order to cope with this requirement, recently researchers have developed random topologies applicable for inter-switch networks. Compared with the conventional Torus or Fat-tree networks, these topologies can drastically reduce the number of hops and be efficiently applied to HPC systems, data centers, and many-core systems.</p> <p>These irregular topologies often suffer from their lack of scalability for feasible interconnection networks. Firstly, they tend to increase the total amount of cable length between cabinets and thus to increase the cost. Secondly, the number of table entries consumed on each switch is increased. Thirdly, they increase the time complexity to compute assignment of Virtual Channels (VCs) to the routing paths for deadlock-freedom.</p> <p>To improve their scalability, this thesis focuses attention on layout-conscious random topologies that contain randomly connected links with length limitation. These networks achieve drastic reduction of the total cable length with a minimal increase in the number of hops compared with completely random networks.</p> <p>This thesis aims to explore a scalable routing methodology that can be applied to the layout-conscious random topologies to achieve feasible low-latency interconnection networks. Firstly, a scalable routing method for the layout-conscious random topologies, LOREN (Layout-Oriented Routing with Entries for Neighbors), is proposed. This method exploits irregularity and locality in these topologies to achieve both the small number of hops between nodes and small routing table sizes required. Secondly, an advanced method for the VC assignment, ACRO (Assignment of Channels in Reverse Order), is proposed. This approach has a small time complexity, yet with the same number of VCs compared with conventional VC assignment methods.</p> <p>The proposed routing methodology provides the better trade-offs between an achieved network latency and the number of required table entries. It is evaluated by using a cycle-accurate network simulator. In the evaluation, the proposed routing method LOREN is enhanced by the proposed VC</p>			

Thesis Abstract

assignment method ACRO to guarantee the deadlock-freedom. The results show that the proposed methodology can improve the network throughput by up to 67.9 % compared to a conventional compact routing method. Moreover, the number of required routing table entries is reduced by up to 91 %, which improves scalability and flexibility for implementation.