氏　　名　Ke Cui

学位(専攻分野)　博士(情報学)

学 位 記 番 号　総研大甲第 2495 号

学位授与の日付　2024 年 3 月 22 日

学位授与の要件　複合科学研究科　情報学専攻
学位規則第6条第1項該当

学 位 論 文 題 目　Efficient Collective Communication in Interconnection Networks with Shortcut Network Topologies

論 文 審 査 委 員　主　査　鯉渕 道紘
情報学コース 教授
五島 正裕
情報学コース 教授
合田 憲人
情報学コース 教授
福田 健介
情報学コース 教授
八巻 隼人
電気通信大学 大学院情報理工学研究科
准教授

Summary of Doctoral Thesis

Name in full : Ke Cui

Title : Efficient Collective Communication in Interconnection Networks with Shortcut Network Topologies

Many applications running on large parallel computers are designed to solve various challenging problems, such as weather forecasting and drug discovery. To solve these challenging problems more efficiently, modern large parallel systems (supercomputers) rely on hundreds of thousands of compute nodes to provide powerful computational capabilities. Applications running on parallel computers usually require the collaborative work of many compute nodes to complete the computational task; the collaboration between compute nodes involves a great deal of mutual communication. Collective communication, a mode of communication involving multiple processes or compute nodes, is widely used in applications running on parallel computers.

For many parallel applications, the collective communication operations dominate the execution time. Improving the performance of collective communication operations is a crucial way to shorten the execution time of the applications. The performance of collective communication operations is affected by both hardware and software. At the hardware level, a crucial factor is the interconnection network, especially network topology. Network topology dictates the connectivity among compute nodes within parallel computers. A low latency network topology reduces the communication overhead between compute nodes and thus improves the performance of collective communication operations. At the software level, the implementation of algorithms for collective communication operations and the mapping strategy of jobs and processes can affect the performance of collective communication operations.

In this dissertation, we present efficient collective communication in parallel computers with shortcut network topologies. The shortcut network topology consists of a baseline topology with shortcut links, which shorten the diameter and average shortest path length (ASPL). We propose the implementation of efficient collective communication operations on two types of network topologies: the random shortcut topology and the non-random shortcut topology.

The first approach proposes implementing efficient collective communication using random shortcut network topologies. The random network topologies can be used to achieve low hop counts between nodes and, thus, low latency on average. We describe three process mapping strategies on random shortcut topologies: random mapping, hierarchical-tree mapping, and ring-based continuous mapping. Then, we apply the two-opt algorithm to the mapped compute nodes to optimize the rank placement for building

efficient collective communication operations on random shortcut network topologies. The two-opt approach minimizes the total path hops or possible communication contention of point-to-point communications that form the target collective communication by implementing the process rank re-placement for efficient collective communication. Our proposed two-opt approach can significantly reduce the number of hops for collective communication operations such as Broadcast, Allreduce, and Alltoall. SimGrid discrete-event simulation results show that the two-opt approach can dramatically improve the performance of collective communication and, in parallel applications where collective communication operations dominate, the two-opt approach can improve the overall performance of the application.

The second approach proposes using circulant network topologies. Unlike random shortcut topologies, a circulant topology is obtained by adding non-random links to a ring topology. The circulant network topologies provide algorithmic features that reduce the total hop counts of some typical collective communication algorithms; these features make them ideal for collective communication operations. We propose two process mapping strategies for circulant network topologies, which are ring-based consecutive mapping and circulant mapping. These two mapping strategies result in very low path hops of collective communication operations, and it is even possible to achieve the optimal hops. SimGrid discrete-event simulation results show that ring-based consecutive mapping and circulant mapping significantly improve the performance of collective communication operations. We also evaluate the parallel applications on circulant network topologies; the application evaluation results also show that when communication operations dominate the performance, the low hops mapping strategies on circulant network topologies achieve better performance.

Finally, we compare these two shortcut network topologies. Circulant network topologies have a higher diameter and average shortest path length (ASPL) than random shortcut network topologies in the case of having the same degree. However, the collective communication operations (e.g., Broadcast, Allreduce, and Alltaoll) on circulant network topologies have far fewer hops than random shortcut network topologies, which results in better performance for collective communication operations on a circulant network topology. Moreover, we compare the cable length of random shortcut topologies and circulant topologies with the same degree; the circulant network topologies have much less cable length than random topologies, which makes the circulant network topologies less costly than the random shortcut network topologies in constructing network interconnection for parallel computers.

（別紙１ Separate Sheet 1）

Results of the doctoral thesis defense

博士論文審査結果

Name in Full
氏　　名　Ke Cui

Title
論文題目　Efficient Collective Communication in Interconnection Networks with Shortcut Network Topologies

本学位論文は、「Efficient Collective Communication in Interconnection Networks with Shortcut Network Topologies」と題し、英文で記述され、全６章から構成されている。

第１章「Introduction」では、並列計算機システムにおける集合通信の研究分野の概況と本研究の目的を述べている。本章では相互結合網における集合通信の重要性について述べ、現状の集合通信技術の問題点を指摘している。そして、本論文の目的がこれらの問題点を解決するために、ネットワークトポロジとプロセスマッピングに関する提案であると述べている。

第２章「Background」では、並列計算機システムにおける相互結合網のネットワークトポロジ、集合通信アルゴリズム、プロセスマッピングについて解説するとともに、各々に関してスループットの向上を実現する研究動向をまとめている。

第３章「Efficient Collective Communication using Random Network Topology」では、直径の面で優れているランダムネットワークトポロジにおいて、集合通信で発生するパケットの総転送ホップ数を削減する点を特徴とする手法を提案している。出願者は、ランダムネットワークトポロジにおいて集合通信の実行時間の短縮効果を示した上で、典型的な並列ベンチマークの実行時間の短縮効果をシミュレーションにより示している。

第４章「Efficient Collective Communication using Circulant Network Topology」では、一定間隔にてショートカットリンクを追加したネットワークトポロジにおいて、一部の集合通信で発生するパケットの総転送ホップ数を理論下限に抑える手法を提案している。出願者は、本ネットワークトポロジにおいて集合通信の実行時間の短縮効果を示した上で、典型的な並列ベンチマークの実行時間の短縮効果をシミュレーションにより示している。

第５章「Comparison of Random and Non-Random Network Topologies」では、第３章で提案した集合通信手法と第４章で提案した集合通信手法の比較を行っている。具体的には、集合通信と１対１通信が混在した場合のパケットの総転送ホッ

プ数、並列ベンチマークの実行時間などに関して評価が成され、集合通信が支配的な通信パターンでは後者の手法が優れているが、その他の場合には前者の手法が優れていることが報告された。

第 6 章「Conclusions and Future Direction」では、本研究を総括し、得られた成果および今後の課題について述べている。

公開発表会では博士論文の章立てに従って発表が行われ、その後に行われた論文審査会及び口述試験では、審査員からの質疑に対して適切に回答がなされた。質疑応答後に審査委員会を開催し、審査委員で議論を行った。審査委員会では、出願者の博士研究が相互結合網の集合通信のスループットの向上に貢献することが評価された。

以上を要するに本学位論文は、並列計算機システムにおいて重要となる集合通信の性能向上を実現する手法を提案し、ネットワークシミュレーションの性能評価によりその有効性を示したものである。また、本学位論文の成果は、学術雑誌論文 1 件、フルペーパー査読付き国際会議論文 1 件として発表され、学術的な貢献も認められる。以上の理由により、審査委員会は、本学位論文が学位の授与に値すると判断した。