Collective Communication AAB for Regular and Irregular Topology Based on Prediction of Conflicts

Miloš Ohlídal, Josef Schwarz
Brno University of Technology,
Faculty of Information Technology
Božetěchova 2, 612 66 Brno, CZ
Tel.: +420-5-41141210, fax: +420-5-41141270
e-mail: {ohlidal, schwarz}@fit.vutbr.cz

***Abstract* - Collective communications involving all processors are frequently used in the solution of demanding parallel problems and their time complexity has a dramatic impact on the performance. This paper deals with scheduling of collective communications in multiprocessor networks using the Store-and-Forward switching technique resulting in minimum number of communication steps. We designed novel technique of communication conflict prediction, which significantly increases the success rate of optimal communication schedule.**

I. Introduction

Multi-core processors in parallel computers are often interconnected with networks that are node-symmetric, i.e. each node has the same view of the network. Such regular networks like ring, 2D-torus or a hyper-cube have the advantage that the same relatively simple routing function can be used, identical in (translated to) all nodes. All-to-all communications then require that all nodes communicate simultaneously without conflict, i.e. no channel can be used at any time in one direction by more than one message. However, there are cases when irregular networks have a certain advantage and are given priority over the regular networks. An example of such irregular network topology is the AMP (A Minimum Path) configuration [1].

In the rest of this paper, we focus especially on All-to-All Broadcast (AAB) [2] operations realized on interconnection networks with full-duplex links, Store-and-Forward (SF) switching with non-combining nodes and all-port communication networks

.

II. Prediction of Conflict

We proposed the partitioning of scheduling problem of a group communication into two subproblems. The first subproblem is defined as a search for appropriate paths between source and destination node. The second subproblem is stated as a finding of confict-free communication step to each channel for each founded path.

The time of finding the optimal schedule can be reduced by usage of conflicts prediction. It is possible to discover during the solution of the first subproblem, whether the communication schedule will be conflict-free or not. The conflict needn't to appear in two possible cases:

1. Utilization of each channel in one direction equals at most the number of communication steps S of the whole schedule
2. Utilization of investigated channel in one step in one direction equals at most or just one. It can be described by (1).

$$S >= P_C \tag{1}$$

where S is the desired number of communication steps of whole schedule and P_C is the number of all paths, which utilize the investigated channel.

The next case of the conflict detection in the designed schedule expresses the situation that it is not possible to assign a communication step to the channel to be conflict-free although (1) is true. This case can be described by the equation (2) with the interval (3), which includes all possible communication steps for the investigated channel.

$$bound = (S - (L - O)) \tag{2}$$

where *bound* is the upper border of communication steps of the investigated channel, L is the length of the investigated path and O is the channel position on the path. Finally, the communication step of an investigated channel on the path is chosen from the interval (3):

$$\begin{aligned} step &\in \langle O, bound \rangle, \; O \neq 1 \\ step &= 1 \qquad\qquad , \; O = 1 \end{aligned} \tag{3}$$

It is tried to assign to each channel on the investigated path the communication step from (3). If it is impossible to choose two different values from (3) for the investigate channel into two different paths, the conflict appears.

These equations (1), (2) and the interval (3) perform only to detect the conflict and also to evaluate the fitness function.

III. The Routing Algorithm

The goal of proposed algorithms is to find a schedule of a group communication with the number of steps as

close as possible to the lower bounds (4) stated analytically [4].

$$\lceil (P-1)/d \rceil \quad (4)$$

where P is the number of node in the interconnection network, d is node degree.

We developed a routing algorithm HGSA [3] to optimize collective communication SF AAB.

A. Solution encoding

The gene consists of two integer components. The first component is an index of the shortest path from source to destination. The second component is a step sequence of communication channels on the path.

B. The fitness function

The fitness function is based on conflict counting. In the first phase the prediction is used to find out how many conflicts appear in the whole schedule. This is based on (5), (2) and (3).

$$conflict_{new} = (P_C - S)$$
$$conflict = \Sigma\, conflict_{new} \quad (5)$$

The parameters S and P_C are defined in Chapter 2, $conflict_{new}$ is the number of conflicts detected for the investigated channel, *conflict* is the number of conflicts of the whole schedule.

If the conflict occurs, the schedule cannot be used in real application.

IV. Experimental Results

In the first experiment the proposed routing algorithm was verified on some multiprocessor topologies (e.g. Midimew, K-Ring, Octagon...). All topology had 8 nodes and the regular and the irregular (AMP with SC and Ladder) topology were examined. It was achieved optimal communication schedule in all tested topologies, which is illustrated by the third column in the Table 1.

TABLE 1
Experimental Result of the Collective Communication AAB for the Topology with 8-Nodes

Topology	AAB HGSA	Theoretical lower bounds
Hypercube	3	3
Hypercube with body diagonal	2	2
AMP with SC	3	-
AMP without SC	2	2
K-ring	2	2
Midimew	2	2
Moore	3	3
Octagon	3	3
Ladder	4	-

In the second experiment the proposed routing algorithm was tested with higher number of node in two different architectures, a hypercube and AMP. The number of nodes varied from 8 to 64. AAB communication complexities, measured by the number of communication steps in schedules found by HGSA so far, are shown in Table 2 (columns three and four).

TABLE 2
Results of the AAB Optimization

Number of nodes	Hypercube minimal	Hypercube HGSA	AMP without SC HGSA
8	3	3	2
16	4	4	-
23	-	-	6
32	7	7	8
42	-	-	11
64	11	12	-

V. Conclusions

In this paper we have focused on finding of optimal schedule communications for the arbitrary topology and AAB communication schedule.

Parallel algorithm HGSA has been developed for optimization of AAB communication pattern and tested with high quality of achieved results. The ability of HGSA to find good or even optimal solutions was proven by a hypercube benchmark. It can schedule group communications for various networks with unknown optimal (minimal) number of steps. It is useful especially for irregular topologies, where analytical approach cannot be applied. The probability rate of achievement of the optimal solution is increased, if the conflict prediction utility is used.

Acknowledgement

This research has been carried out under the financial support of the research project FRVS 2848/2006/G1 “Memetic Evolutionary Algorithms Applied to Designed of Group Communication between Processors” (Ministry of Education).

References

[1] Chalmers, A.,Tidmus, J.: Practical Parallel Processing. International Thomson Computer Press, 1996.

[2] Defago, X., Schiper, A., Urban, P.: Total Order Broadcast and Multicast Algorithms: Taxonomy and Survey, technical report DSC/2000/036, 2003.

[3] Jaroš, J., Ohlídal, M., Dvořák, V.: Evolutionary Design of Group Communication Schedules for Interconnection Networks, In: Proceedings of 20th International Symposium of Computer and Information Science, Berlin, DE, Springer, 2005, pp. 472-481.

[4] Dvořák, V.: Scheduling Collective Communica-tions on Wormhole Fat Cubes, In: Proc. of the 17th International Symposium on Computer Architecture and High Performance Computing, Los Alamitos, US, IEEE CS, 2005, pp. 27-34