An Adaptive Unicast/Multicast Routing Algorithm for MPSoCs

Masoumeh Ebrahimi, Masoud Daneshtalab, Pasi Liljeberg, Hannu Tenhunen

Department of Information Technology, University of Turku, Turku, Finland

{masebr, masdan, paslil, hanten}@utu.fi

Abstract. *Several parallel applications in MPSoCs take advantage of multicast communication. Path-based multicast scheme has been proven to be more efficient than the others multicast schemes in on-chip interconnection network. We present a new adaptive path based model for both the multicast and unicast wormhole routing protocols. The proposed model under mixed traffic models has lower latency than the previous path-based methods with negligible hardware overhead.*

1. Introduction

The performance of the MPSoCs is highly dependent on the underlying communication mechanism. The Communication in MPSoC (NoC) can be either unicast or multicast. The multicast communications are frequently employed in many application of MPSoC such as replication, barrier synchronization, cache coherency in distributed shared-memory architectures, and clock synchronization. Multicast routing algorithm can be classified as unicast-based [4][5], tree-based [4], and path-based [6]. In unicast-based, the multicast operation is performed by sending a separate copy of the message from the source to every destination. The drawback of this scheme is the fact that multiple copies of the same message are injected into the network. In tree-based multicast approach, a spanning tree is constructed that the source is indicated as the root and the messages are sent down the tree. In this way a message might be replicated at some of the intermediate nodes and forwarded along multiple outgoing channels toward disjoint subsets of destinations. If one branch of the tree is blocked, all are blocked. Branches must proceed forward in lock step, which may cause a message to hold many channels for extended periods, which results increasing network contention [2]. A solution to overcome the tree-based disadvantages is to utilize the path based multicast wormhole routing [2]. In this method, a source node prepares a message for delivery to a set of destinations by first sorting the addresses of the destinations in the order in which they are to be delivered, and then placing this sorted list in the header of the message. When the header enters a router with address A, the router checks to see whether A is the next address in the header, If so, the address A is removed from the message header and a copy of data flits will be delivered to the local core and the flits are forwarded to the next node on the path. Otherwise, the message is forwarded only to the next node on the path.

In this work, for improving the path-based method we proposed an adaptive, deadlock-free wormhole routing algorithm based on Hamiltonian path for both unicast and multicast mechanisms in 2D-mesh NoCs. The proposed algorithm restricts the locations where some types of turns can be taken. Unlike the other adaptive turn model which is applicable just for unicast messages, both unicast and multicast routing methods are supported by our adaptive and deadlock-free algorithm. In addition, none of the path-based multicast techniques has addressed the issue of using an efficient adaptive algorithm for both unicast and multicast approaches. By applying the proposed adaptive algorithm we can achieve better performance compared to the traditional path-based multicast algorithm with very low hardware overhead. Experimental results show that the performance can be improved by using proposed adaptive model in the traditional path-based multicast algorithms such as Multi Path, and Column path. Additionally, the unicast approach of our proposed model outperforms the traditional adaptive unicast models. The chip area overhead of the proposed scheme is negligible, less than 0.5%. The paper is organized as follows. In Section 2 and 3, the traditional and proposed path-based multicast algorithms are discussed while the proposed switch architecture is presented in section 4. The results are discussed in Section 5 with the summary and conclusion given in the last section.

2. Hamiltonian Path-based Structure

In this paper, we consider two-dimension mesh with wormhole switching technique. Formally, in $m \times n$ 2D-mesh two nodes with coordinates (x_i, y_i) and (x_j, y_j) are connected by a communication channel if and only if $|x_i - x_j| + |y_i - y_j| = 1$. The path-based routing is established as the Hamiltonian path algorithm [1]. In this method an undirected Hamilton path of the network is constructed; A Hamilton path visits every node in a graph exactly once [7]. In this algorithm, for each node in an $m \times n$ mesh a label $L(x, y)$ is assigned as follows: $L(x, y) = y \times n + x$, if y is even, and $L(x, y) = y \times n + n - x - 1$, if y is odd. As shown in Fig. 1, two directed Hamilton paths (or two subnetworks) are constructed by the labeling [1]. The high channel subnetwork (H_u) starts at (0, 0) ,and the low channel subnetwork (H_l) ends at (0, 0). In case the label of the destination node is greater than the label of the source node, the routing always takes place in the H_u subnetwork; otherwise it takes place in the H_l subnetwork. The destinations are placed into two groups. One group contains all the destinations that could be reached using the H_u subnetwork, and the other contains the remaining destinations that could be reached using the H_l subnetwork. To reduce the path length the vertical channels that are not part of the Hamilton path (the dashed lines in the Fig. 1) could be used in appropriate directions. The proposed adaptive model designed for both unicast and multicast messages, uses the Hamiltonian path strategy.

3. The Proposed Adaptive Model

The former path-based routing models such as Multi-Path (MP) [1] and Column-Path(CP) [2] algorithms that have been described in the following subsections, route the unicast and multicast messages by deterministic routing algorithms. Therefore, the network performance has been degraded by these models. The proposed minimal adaptive scheme takes the place of the deterministic model in the path-based routing algorithms to route both of the unicast and multicast messages through the destination(s). For breaking all of cycles in the proposed adaptive scheme, similar to the odd-even model, the locations at which certain turns can be taken are restricted so that deadlock can be avoided. The rules regulating the proposed scheme are categorized in the up channel subnetwork and the down channel subnetwork as follows:

For the *up channel subnetwork*:

Rule1: NW turn is not allowed in even rows.

Rule2: NE turn is not allowed in odd rows.

For the *down channel subnetwork*:

Rule1: SE turn is not allowed in even rows.

Rule2: SW turn is not allowed in odd rows.

Notice that the message will be forwarded to the destination as is done in the deterministic Hamiltonian algorithm, when the current node is located one row to the south (north) of the destination row in the up channel network (down channel network). Inasmuch as the proposed rules keep the messages traveling through the Hamiltonian paths, it prevents the occurrence of the deadlock [2]. In addition, both minimal and non-minimal paths are possible with the proposed adaptive path-based model. However, our implementation is based on minimal path and does not support the non-minimal paths. Now we describe how the proposed adaptive model affects the path-based multicast routing algorithm. The Multi-Path (MP) [1] and Column-Path (CP) [2] algorithms are the most important path-based routing methods that use the Hamiltonian path strategy.

MP Multicast Routing: In Multi-Path (MP) routing algorithm the destination node set is partitioned into two subsets, D_H and D_L, where every node in D_H has a higher label than that of the source node and every node in D_L has a lower label that of the source node. Thus, multicast messages from the source node will be sent to the destination nodes in D_H using the H_u subnetwork and to the destination nodes in D_L using H_l subnetwork. To reduce the path lengths D_H and D_L are also partitioned. The set D_H is divided into two subsets. One consist of the nodes whose x coordinates are greater than or equal to that of the source and the other subset contains the remaining nodes in D_H. The set D_L is partitioned in a similar way. Hence, all destinations of a multicast message are grouped into four disjoint. Consider the example illustrated in Fig. 2(a) for a 8×8 mesh network where node 27 (3, 4) send its multicast messages to destinations 0, 1, 7, 8, 9, 19, 26, 31, 32, 37, 50, 55, 57, 59, 62, and 63. Accordingly, two subsets are organized. The first subset (D_H) that has all the destinations that could be reached from the source node using H_u subnetwork which are 31, 32, 37, 50, 55, 57, 59, 62, and 63 in sequence and the second one (D_L) has the remaining destinations that could be reached using the H_l subnetwork which are 0, 1, 7, 8, 9, 19, and 26. As exhibited in Fig. 2(a), D_H is divided into two subsets, which are D_{H1}= {31, 32, 50, 62, 63} and D_{H2}= {37, 55, 57, 59}. In the same way D_L is divided into two subsets, with D_{L1}= {0, 1, 19} and D_{L2}= {7, 8, 9, 26}. The multi-path is deadlock free and could be used for unicast and multicast routing simultaneously. AMP, Adaptive MP, is the adaptive model of the MP algorithm after the proposed model is applied in the MP algorithm. Consider the example used for MP in Fig. 2(b). The multicast message can be forwarded in three different ways from the node 37 through the node 55 (32 through 50, 19 through 1, and 26 through 9).

CP Multicast Routing: In this method, the destination node set is partitioned to 2k subsets. K is the number of columns in the mesh, and at most two messages will be copied to each column. If a column of the mesh has one or more destinations in rows above the source, then one copy of the message is sent to service all of those destinations. Similarly, if a column has one or more destinations in the rows below the source, then another copy of the message is sent to service all of those destinations. One copy of the message is sent to a column if all destinations in that column are either below or above the source node. Otherwise, two messages are sent to that column. For instance, to send a message to destinations 0, 1, 7, 8, 9, 19, 26, 31, 32, 37, 50, 55, 57, 59, 62, and 63 from the source node 27 using the Column-Path (CP) routing algorithm based on the Hamiltonian path-base is shown in Fig. 2(c). Thirteen copies of the message are used to achieve the desired multicast operation. Though destinations 1 and 62 are in the same column, two message copies are sent to this column, since two of the destinations are above the source node's row and the other below. The routing algorithm used in this scheme is based on the XY routing algorithm. Therefore, the CP routing algorithm is compatible with the unicast routing method and it is deadlock-free and livelock-free [2]. The ACP, stood for the Adaptive CP is the adaptive method of the original CP by taking advantage of the proposed adaptive model. To indicate how the adaptive scheme affects the CP algorithm, as illustrated in Fig. 2(d), again thirteen copies of the multicast message must be used to achieve the desired multicast operation. But in this figure for simplicity, we only consider two subsets A and B. Due to utilizing the proposed adaptive scheme in the CP, each multicast message can be delivered to its subset through different paths indicated by dashed lines.

4. Hardware Implementation

Message Format: The multicast message format is shown in Fig. 3. As can be seen; it includes a header flit and a parametric number of payload flits. Each flit is *n* bit wide and the *n*th bit is the EOM (End Of Message) sign and the *(n-1)*th bit is the BOM (Begin Of Message) sign. In the header, the third field is used to describe the type of the message. There are two types of message: unicast (T=0) and multicast (T=1), represented by T. The specific addresses of the source node and the destination node(s) are placed in the last field of the header, respectively and the content of the message is located in the rest of flits (Payload).

Switch Structure: Each input port has a controller for handshaking and an input buffer. After receiving the message header, first the routing unit determines which output should be used for routing this message and then the arbiter request for a grant to inject the message to a proper output using a crossbar switch. The router has a crossbar which establishes the connection path from an input port to an output port. Since the crossbar can only serve a single output port at a time, it uses an arbiter for arbitrating among simultaneous input requests to access the same output port. When a new message reaches the input port, it waits until the previously arrived messages leave the port. Then the new message header is delivered to the routing unit and routed to the appropriate output port. The Congestion Flag (CF) of the buffer becomes active when the number of empty cells of the buffer is less than a threshold value. In this case, warning for the full status, the signal CF, is activated indicating that most buffer cells are occupied. Each input port has a CF through which it informs its adjacent router about its congestion condition. Therefore, the routers which uses that input port for forwarding a message to the next router should consider this router as a congested one (congestion area or hotspot) and should not send messages to this router until the congestion is over.

Header Processing Mechanism: The router employs a routing unit which decodes the header of messages coming from an input port. If the header belongs to a unicast message (T=0), the minimal path adaptive routing algorithms is used to determine the output port to which the message should be sent. In the proposed adaptive routing algorithm there could be more than one minimal output directions where to route messages. In this case the address

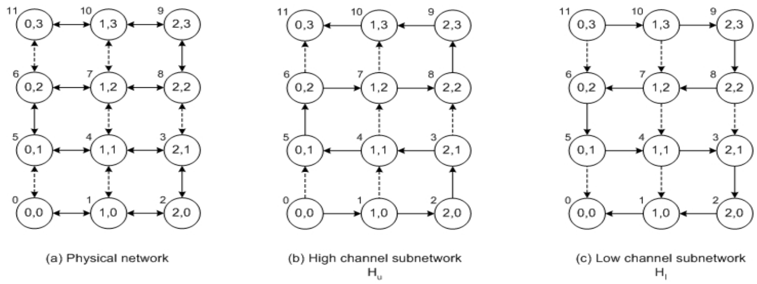

Fig. 1. (a) A 3×4 mesh physical network with the label assignment and the corresponding (b) high channel and (c) low channel networks. The solid lines indicate the Hamiltonian path and dashed lines indicate the links that could be used to reduce path length in routing.

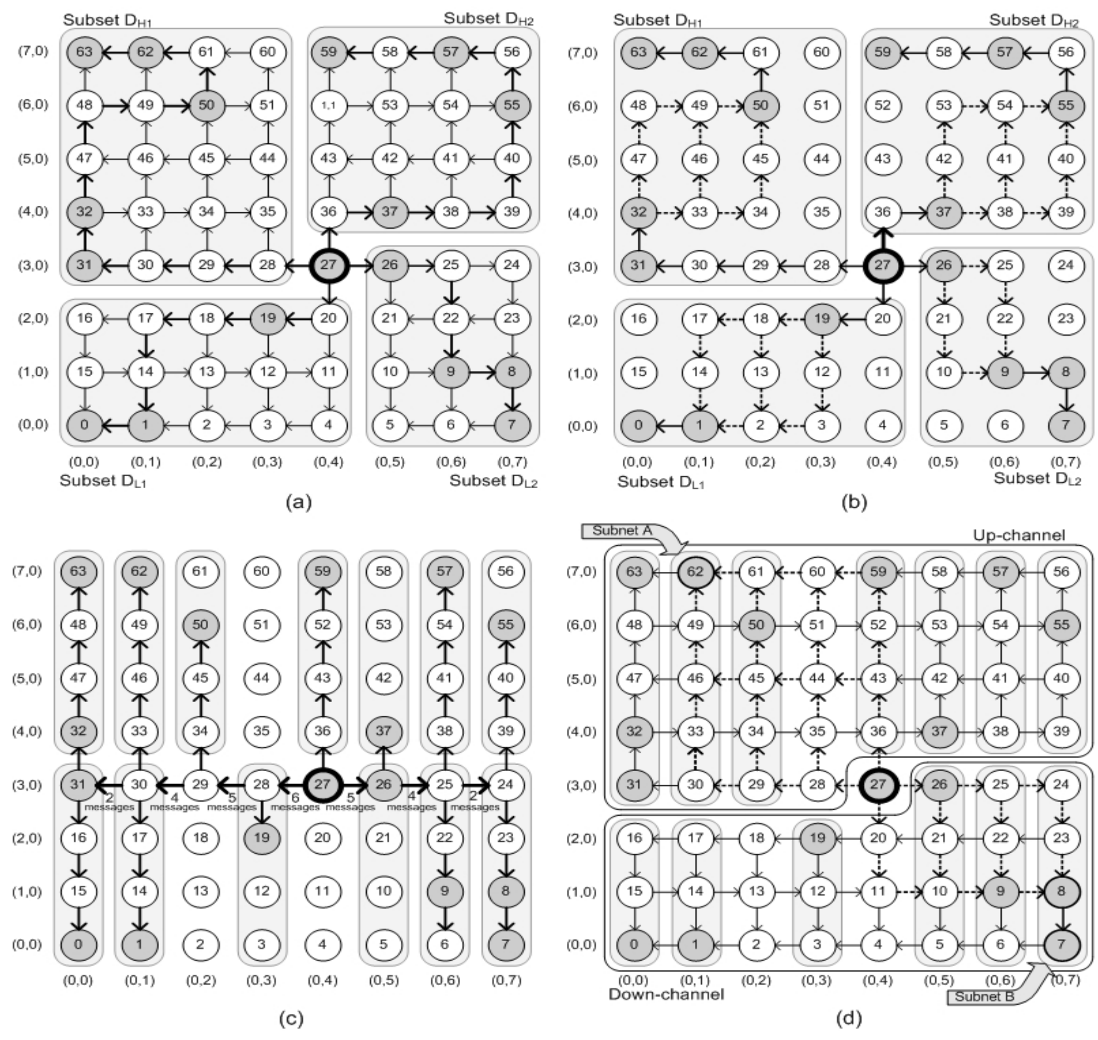

Fig. 2. Examples of (a) Multi-path (MP), (b) Adaptive Multi-Path (AMP), (c) Column-Path (CP), and (d) Adaptive Column-Path (ACP) multicast routings from the node 27 (to subsets A, and B for column path example). The unused links are not indicated.

decoder will choose the direction where the corresponding downstream router has not raised its congestion flag. For instance, if a message with a given source and destination could be routed to both outputs p1 (CF=0) and p2 (CF=1), then it will be routed to p1. If p1 and p2 happen to have both their congestion flag raised, the message will be routed to p1. On the other hand, if the header type is a multicast message (T=1), the routing unit fetches the destination address which the pointer in the header points. Afterward, the routing unit increases the pointer value of the header, and if it overflowed, the routing unit would remove the corresponding flit header from the message. In a word, whenever a destination address is fetched from the header, the pointer value will be increased. After fetching the destination address from the header, if the destination address is the current node, the routing unit will request the local output port. Meanwhile, the routing unit fetches the next destination address from the header and runs the adaptive routing procedure to determine the output port(s) corresponding to the next destination address.

5. Results and Discussion

To assess the efficiency of the proposed adaptive model, two multicast routing algorithms were also implemented. These algorithms include MP and CP. We have developed a synthesizable wormhole NoC simulator implemented in VHDL to assess the efficiency of the proposed adaptive method. This simulator can be used for wormhole switching in two dimensional mesh configuration for the NoC. The simulator inputs include the array size, the routing algorithm, the link width length, buffer size, and the traffic type. The simulator can generate different traffic profiles. To calculate the power consumption, we have used power compiler. For all switches, the data width was set to 32 bits, and each input channel has a buffer (FIFO) size of 12 flits with the congestion threshold set at 75% of the total buffer capacity. The message size was assumed to be 16 flits. For the performance metric, we use the multicast latency defined as the number of cycles between the initiation of multicast message operation and the time when the tail of the multicast message reaches all the destinations. The array size is considered to be 8×8.

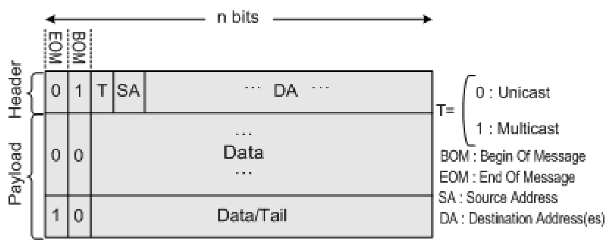

Fig. 3. Multicast message format

Unicast and Multicast (Mixed) Traffic Profile: We have employed a mixture of unicast and multicast traffic, where 80% of injected messages are unicast messages and the remaining 20% are multicast messages. This pattern may be representative of the traffic in a distributed shared-memory multiprocessor where updates and invalidation produce multicast messages and cache misses are served by unicast messages [2]. The unicast messages are also routed using the proposed adaptive scheme. Uniform traffic model [3] has been taken into account for unicast traffic generation. In the uniform traffic profile, each processing element sends a message to any other PE in an equal probability. This is determined randomly using a uniform distribution. In Fig. 4 the average communication latency of different algorithms under the uniform traffic model for unicast traffic is shown. As depicted in these figures, for this traffic, the adaptive routing algorithms perform better.

Hardware Overhead: The switches were synthesized by Synopsys D.C. using the TSMC 0.09µm standard cell library. For all switches, the data width was set to 32 bits (flit size), and each input channel had a buffer size of 12 flits. Comparing the area cost indicates that the hardware overhead of implementing the proposed adaptive scheme in both the MP and CP switches is less than 0.5% and that can be considered negligible.

6. Summary and Conclusion

In this paper, a new adaptive model based on Hamiltonian path in mesh interconnection networks for NoCs was proposed. In this scheme, three facets have been considered such as utilization of network partitioning, and taking advantage of the proposed adaptive model for routing both the multicast and unicast messages through the network. A synthesizable VHDL NoC-environment was developed to evaluate the efficiency of the proposed multicast routing algorithm under the mixed traffic models.

References:

[1] X. Lin et al., “Multicast Communication in Multicomputer Networks,” IEEE Transactions on Parallel and Distributed Systems, 1993, 4, pp. 1105-1117

[2] R. V. Boppana et al., “Resource deadlock and performance of wormhole multicast routing algorithms,” IEEE Transactions on Parallel and Distributed Systems, pp. 535-549, 1998

[3] G. Chiu, “The Odd-Even Turn Model for Adaptive Routing,” IEEE Tran. On Parallel and Distributed System, pp 729-738, July 2000

[4] N. E. Jerger et al., “Virtual Circuit Tree Multicasting: A Case for On-Chip Hardware Multicast Support,” in proc. int. conf. Computer Architecture, China, 2008, pp. 229-240

[5] P. McKinely et al., “Unicast-based multicast communication in wormhole-routed networks,” IEEE Trans. of Parallel and Distributed Systems, 1994, 5, pp. 1252-1265

[6] X. Lin et al., “Multicast Communication in Multicomputer Networks,” IEEE Transactions on Parallel and Distributed Systems, 1993, 4, pp. 1105-1117

[7] F. Harary, “Graph Theory” (Readings, Massachusetts: Addson-Wesley, 1972)

[8] C.J. Glass et al.,“The Turn Model for Adaptive Routing,” in Proc. 19th Ann. Int'l Symp. Computer Architecture, pp. 278, May 1992.

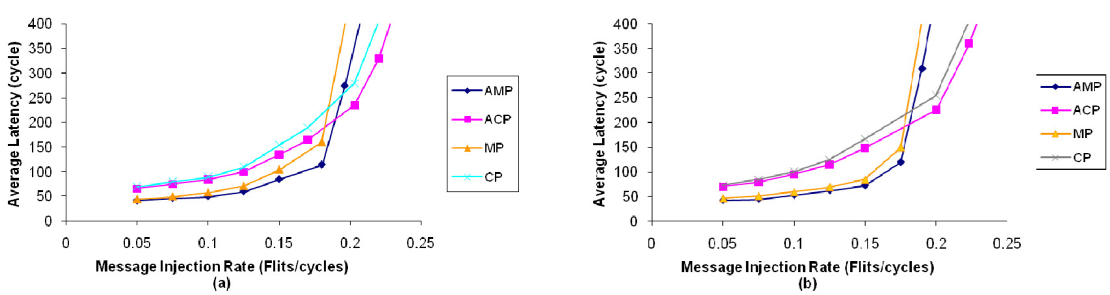

Fig. 4. Performance under different loads in 8×8 2D-mesh with (a) 10 destinations, (b) 25 destinations under mixed traffic (20% multicast and 80% unicast). Unicast traffic is based on the uniform traffic model.