Multicast Communication in Wormhole-Routed Star Graph Interconnection Networks with Multipath-Based Approach

Nen-Chung Wang and Chih-Ping Chu
Dept. of Computer Science and Information Engineering
National Cheng Kung University
Tainan, Taiwan, R.O.C.
E-mail: {wangnc, chucp}@csie.ncku.edu.tw

Tzung-Shi Chen
Dept. of Information Management
Chang Jung University
Tainan, Taiwan, R.O.C.
E-mail: chents@mail.cju.edu.tw

Abstract

The star graph interconnection network, when compared with hypercube network, being with low degree and small diameter, has been recognized to be an attractive alternative to the popular hypercube network. In this paper, we address a multipath-based multicast routing model for wormhole-routed star graph networks, propose two efficient multipath routing schemes, and present the performance of the proposed schemes in contrast with our previous work. Both of the two proposed schemes are proved deadlock-free. The first scheme, simple multipath routing*, uses multiple independent paths for concurrent multicasting. The second one,* two-phase multipath routing*, includes two phases: source-to-relay and relay-to-destination and, for each phase, the multicasting is proceeded using simple multipath routing. Experimental results show for small message startup latencies with short and medium messages the performance of our proposed schemes is evidently superior to that of previous schemes.*

Keywords: Multicast, multipath routing, parallel computing, star graphs, wormhole routing.

1 Introduction

Multicast is an important collective operation on multicomputer systems, in which the same message is delivered from a source node to an arbitrary number of destination nodes. Recently, as shown in the literature [4, 5, 7, 8, 12, 14, 17, 20], multicasting the message with the path-based routing on multidestination wormhole-routed networks has received considerable attention. The system architectures have been improved and are with multidestination message passing capacity to enhance the data transmission performance over that using unicast-based routing schemes. The path-based multicasting sends the source messages to all destinations according to the constructed paths.

The star graph [1, 2] interconnection network, when compared with the hypercube network, being with low degree and small diameter, has been recognized to be an attractive alternative to the popular hypercube network. In star graph networks, lots of solutions of communication problems with store-and-forward switching are proposed [3, 6, 10, 13, 15, 16, 18, 19]. In our previous work [5], for wormhole star graph networks, we addressed a path-based routing model, derived a node labeling formula based on a single hamiltonian path, and proposed four efficient deadlock-free multicast routing schemes.

In this paper, we address a multipath-based multicast routing model for wormhole-routed star graph networks, propose two efficient multipath routing schemes, and present the performance of the proposed schemes in contrast with our previous work. Both of the two proposed schemes are proved deadlock-free. The first scheme, *simple multipath routing*, uses multiple independent paths for concurrent multicasting. The second one, *two-phase multipath routing*, includes two phases: source-to-relay and relay-to-destination and, for each phase, the multicasting is proceeded using simple multipath routing.

The rest of this paper is organized as follows. Preliminaries are presented in Section 2. In Section 3, we address a multipath-based routing model and propose two multipath routing schemes. Simulation results of these algorithms are presented in Section 4. Finally, concluding remarks are drawn in Section 5.

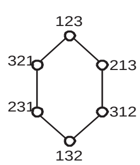
2 Preliminaries

2.1 System Model

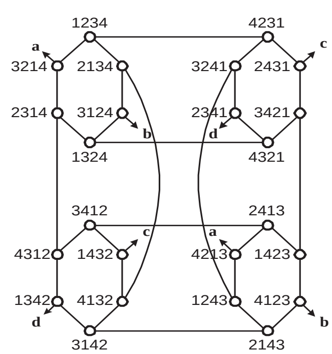
The topologies of mesh and hypercube are widely applied to parallel computer systems. Because of the simple characteristics, the mesh and the hypercube are easy to implement in hardware. The star graph proposed, which is symmetric and hierarchical, is a particularly attractive alternative to the hypercube [1, 2].

In the following, we first introduce some definitions and notations related to the star graphs. A permutation of n distinct symbols from the set $\{1, 2, \cdots, n\}$ is represented as $p = s_1 s_2 \cdots s_n$, where $s_i, s_j \in \{1, 2, \cdots, n\}, s_i \neq s_j$ for $i \neq j, 1 \leq i, j \leq n$. Given a permutation $p = s_1 s_2 \cdots s_n$, let the *generator* g_i be the function of p that interchanges the symbol s_i with the symbol s_1 in p for $2 \leq i \leq n$. Thus, $g_i(p) = s_i s_2 \cdots s_{i-1} s_1 s_{i+1} \cdots s_n$. An undirected *star graph* with dimension n is denoted as $S_n = (V_n, E_n)$, where the set of vertices V_n is defined as $\{v | v = s_1 s_2 \cdots s_n, s_i, s_j \in \{1, 2, \cdots, n\}, s_i \neq s_j$ for $i \neq j, 1 \leq i, j \leq n\}$ and the set of edges E_n is defined as $\{(v_p, v_q) | v_p, v_q \in V_n, v_p \neq v_q$, such that $v_q = g_i(v_p)$ for $2 \leq i \leq n\}$.

In other words, any two nodes v_p and v_q are connected by an undirected edge if and only if the corresponding permutation to the node v_q can be obtained from that of v_p by interchanging the symbol s_i of v_p with the symbol s_1 of v_p for $2 \leq i \leq n$. We also use the notation S_n to represent an n-dimensional star graph, called n-star graph, in this paper. Notice that star graphs are edge and vertex symmetric. Moreover, S_n is a regular graph with degree $n-1$, $n!$ vertices, and $\frac{(n-1)n!}{2}$ edges. A 3-star and a 4-star graphs are shown in Figure 1.



(a)



(b)

Figure 1: The topology of star graphs: (a) 3-star graph; (b) 4-star graph.

The star interconnection network system is composed of nodes, each node is a computer with its own processor, local memory, and communication links; each link connects two neighboring nodes through network [8]. The node architecture of a star network system is shown in Figure 2. A common component of nodes in a new-generation multicomputer is a *router*. It can handle the entering, leaving, and passing through the node of message. The architecture of the star network system provides the wormhole routing with multidestination message passing capability.

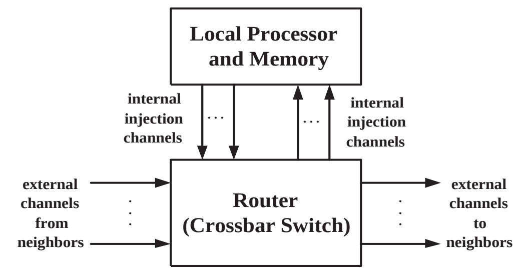


Figure 2: A node architecture with multidestination routing.

2.2 Path-Based Multicast Routing Model

In our previous work [5], for wormhole star graph networks, we addressed a path-based routing model, derived a node labeling formula based on a single *hamiltonian path* (HP), and proposed four efficient deadlock-free multicast routing schemes: dual-path, shortcut-node-based dual-path, multipath, proximity grouping. Generally, the dual-path scheme is simple and efficient. The multicasting in the dual-path routing includes two independent paths (toward high label nodes and low label nodes, respectively) and the next traversed node is the neighboring node with the label nearest to that of the next unvisited target node. The concept of the path-based routing model is described below.

2.2.1 Hamiltonian Paths and Channel Networks

The path-based routing method for meshes developed by Lin et al. [8] is based on a HP. In [5], we used the strategy in [11] to define a HP on the star graph. Because a star graph is embedded with more than one HP, the routing methods proposed in [5] is simply on basis of a specific HP of all possible HPs.

In an n-star graph, the number of nodes is $N = n!$ and each node s is with a label $\ell(s)$, where $0 \leq \ell(s) \leq N-1$ and $\ell()$ is node labeling function [5]. The labeling of a 4-star graph based on a HP is shown in Figure 3. For example, in a 4-star graph, $\ell(1234) = 0$, $\ell(4213) = 6$, $\ell(4312) = 13$, $\ell(4231) = 23$, and so forth.

According to the node labels, we can construct a specific HP, i.e., from the node with label 0, following the nodes with labels $1, 2, \cdots$, to the node with label $N-1$. When node labeling is completed, we can divide the network into two subnetworks, *high-channel network* and *low-channel network*. The *high-channel network*

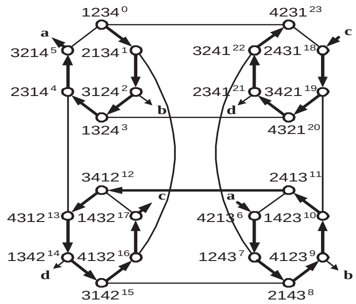


Figure 3: The labeling of a 4-star graph based on a HP.

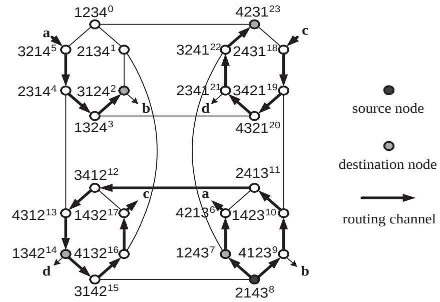


Figure 5: A sample multicast using hamiltonian-path routing.

contains all directional channels with nodes labeled from the lower to the higher, and the *low-channel network* contains all directional channels with nodes labeled from the higher to the lower. Then, a message routing can be performed along two legal paths, one along *high-channel network* and the other along *low-channel network*. The channel subnetworks of a 4-star graph are shown in Figure 4(a) and Figure 4(b), respectively.

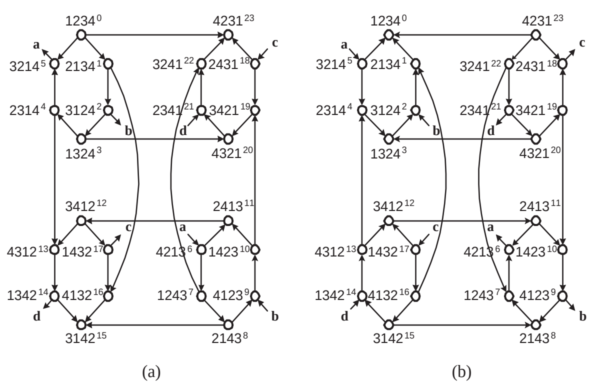


Figure 4: The channel networks of a 4-star graph: (a) high-channel network; (b) low-channel network.

2.2.2 Hamiltonian-Path and Dual-Path Multicast Routing

The unicast-based, the hamiltonian-path, and the dual-path routing strategies can be adopted in a lot of wormhole-routed interconnection networks. The unicast-based routing scheme uses one-to-one communication to achieve multicast, which requires startup latency in each intermediate node [9]. The disadvantage of this approach lies in that significant transmission latency is resulted from the required number of communication startup steps for multicast. In the hamiltonian-path routing, the source node sends the message to all destination nodes based on the constructed hamiltonian path. In this scheme, the multicast is divided into two submulticasts and that can be proceeded in parallel by two independent routing paths (one for high-channel routing and the other for low-channel routing). The disadvantage of this approach is that it always traverses nodes following the fixed path (hamiltonian-path) that requires more traverse links for multicast [8]. In the dual-path routing, the multicasting is similar to the hamiltomian-path routing except each router tries to find a shortcut node (the node with label closest to that of the next *unvisited target* node) for routing to reduce the average length of multicast paths [8].

A sample multicast using hamiltonian-path routing and dual-path routing respectively is shown in Figure 5 and Figure 6. The sample multicast is denoted as the multicasting set $R = \{\underline{2143}^8, 3124^2,\ 1243^7,\ 1342^{14}, 4231^{23}\}$, where the first element of R is the source node and the others are the destination nodes in arbitrary order. Notice that the source node is underlined, the label $\ell(u)$ of each node u in R is shown as a superscript to the node representation. In the hamiltonian-path and the dual-path routing, the multicasting set R can be completed by two submulticasting sets, R^h for high-channel routing and R^l for low-channel routing, i.e., $R^h = \{\underline{2143}^8, 1342^{14}, 4231^{23}\}$ and $R^l = \{\underline{2143}^8, 1243^7, 3124^2\}$, In R^h and R^l the first elements are source nodes and the others are destination nodes with label values higher and lower than source nodes and in ascending and descending orders, respectively. In hamiltonian-path routing as shown in Figure 5, the total number of channels traversed is 15+6=21, and the maximum routing distance is max(15,6)=15. In dual-path routing as shown in Figure 6, the total number of channels traversed is 11+6=17, and the maximum routing distance is max(11,6)=11.

The hold-and-wait property of wormhole routing is particularly susceptible to deadlock, and thus most wormhole-routed systems avoid messages routing to reach cycles of channel dependency. Deadlock can be prevented by the routing algorithm. By ordering network resources, such as nodes, and accessing resources

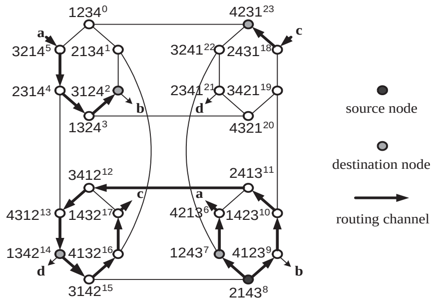


Figure 6: A sample multicast example using dual-path routing.

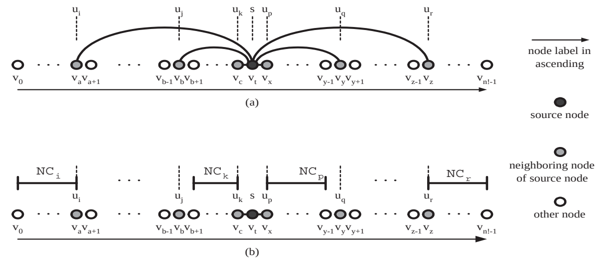


Figure 7: The system nodes except source node are partitioned into multiple node classes: (a) find neighboring nodes; (b) node classes.

according to a strictly monotonic order circular wait for resources will not occur and deadlock can be avoided [8].

3 Multipath Multicast Routing

In this section, we first address a multipath-based routing model. Then, we propose two efficient multipath multicast routing schemes.

3.1 Multipath-Based Multicast Routing Model

For the hamiltonian-path and the dual-path routing, the multicast always uses two independent routing paths (one for high-channel routing and the other for low-channel routing). However, in an n-star graph $S_n = (V_n, E_n)$ every node has degree $n-1$. That is, every node has $n-1$ neighboring nodes. Therefore, we can use $n-1$ independent paths for message routing concurrently to promote the performance of the multicasting.

For the necessity of our proposed multipath multicast routing model, the system nodes except source node are partitioned into $n-1$ node classes NC_i, where $1 \leq i \leq n-1$. All the node class NC_i can be obtained according to the following nodes-partiton rule.

Nodes-partition rule: Given an n-star graph $S_n = (V_n, E_n)$, the source node $s \in V_n$ and the neighboring node set of s, $NN = \{u_i | u_i = g_{i+1}(s)$ for $1 \leq i \leq n-1\}$. Then, based on each $u_i \in NN$ a corresponding NC_i can be obtained by either of the following two cases. Case 1 ($\ell(u_i) < \ell(s)$): Get $\ell(u_j) = max\{\ell(u_k) | \ell(u_i) - \ell(u_k) > 0, u_k \in NN\}$. If u_j exists then $NC_i = \{v | \ell(u_j) < \ell(v) \leq \ell(u_i), v \in V_n\}$; else $NC_i = \{v | 0 \leq \ell(v) \leq \ell(u_i), v \in V_n\}$. Case 2 ($\ell(u_i) > \ell(s)$): Get $\ell(u_j) = min\{\ell(u_k) | \ell(u_i) - \ell(u_k) < 0, u_k \in NN\}$. If u_j exists then $NC_i = \{v | \ell(u_i) \leq \ell(v) < \ell(u_j), v \in V_n\}$; else $NC_i = \{v | \ell(u_i) \leq \ell(v) \leq n-1, v \in V_n\}$.

Figure 7 shows the system nodes except source node are partitioned into multiple node classes. From Figure 7(a), the system nodes are $v_0, \cdots, v_{n!-1}$, where $\ell(v_i) = i, \ell(v_i) < \ell(v_{i+1})$, and $0 \leq i \leq n-2$. The neighboring node set is $NN = \{u_1, u_2, \cdots, u_{n-1}\}$. As shown in Figure 7(b), the system nodes except source node is partitioned into $n-1$ node classes $NC_1, NC_2, \cdots, NC_{n-1}$. Note that $NC_1 \cup NC_2 \cup \cdots \cup NC_{n-1} \cup \{s\} = V_n$.

Subsequently, the multipath-based multicast routing model is described as follows. First, the destination node set D is partitioned into m subsets D_i^α, where $D_i^\alpha \neq \emptyset$, $1 \leq i \leq m$ and $m \leq n-1$, according to each node in D belonging to the node class NC_i, $\alpha \in \{h, l\}$, $\alpha = h$ stands for high-channel routing and $\alpha = l$ stands for low-channel routing. Second, the destination nodes in D_i^h and D_i^l are sorted, according to the node labels, in ascending and descending order, respectively. Third, the multicast is partitioned into multiple submulticasts. Finally, those submulticasts send the message in parallel via multiple independent paths. As illustrated in Figure 8(a) and Figure 8(b), the dual-path and the multipath routing in the worst case are shown, respectively.

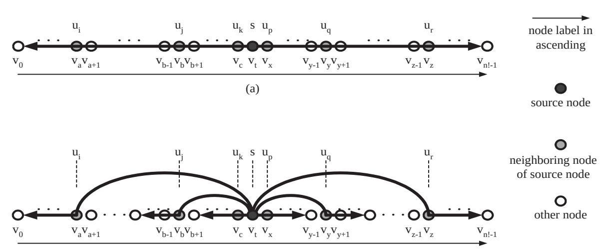


Figure 8: Schematic representation of the dual-path and the multipath routing in the worst case: (a) dual-path routing; (b) multipath routing.

3.2 Multipath Multicast Routing Algorithms

Before we introduce the proposed routing algorithms, let us first define a routing function RF.

Definition 1 (*The routing functions RF*). Let V, p, q be the node set, the source node, and the destination node of a star graph, respectively. The routing function RF is defined to be $RF : V \times V \rightarrow V$ and $RF(p,q) = x$, and if $\ell(p) < \ell(q)$, then $\ell(x)$ = max$\{\ell(u)|\ell(p) < \ell(u) \leq \ell(q)$, and p is adjacent to $u\}$; if $\ell(p) > \ell(q)$, then $\ell(x)$ = min$\{\ell(u)|\ell(p) > \ell(u) \geq \ell(q)$, and p is adjacent to $u\}$.

3.2.1 Simple Multipath Routing

Algorithm 1: *The simple multipath routing algorithm*
Input: Source node s, destination node set D, and node labeling function $\ell()$.
Step 1: // Destination-nodes partition
In an n-star graph S_n, the destination node set D is partitioned into m subsets D_i^α, where $D_i^\alpha \neq \emptyset$, $1 \leq i \leq m$, $m \leq n-1$, and $\alpha \in \{h, l\}$.
Step 2: // Destination-nodes sorting
For each destination node subset D_i^α do
if $(\alpha = h)$
  Sort the destination nodes in D_i^h according to the $\ell()$ values in ascending order.
else
  Sort the destination nodes in D_i^l according to the $\ell()$ values in descending order.
endif
Step 3: // Message preparation
Construct m messages M_i^α, where $1 \leq i \leq m$, $\alpha \in \{h, l\}$, and M_i^α contains D_i^α as part of the header.
Step 4: // Routing in parallel
For each message M_i^α do
if $(\alpha = h)$
  // The message M_i^h is sent to the nodes in D_i^h using
  // high-channel routing based on subnetwork N^h.
  $High\_Channel\_Routing(M_i^h)$
else
  // The message M_i^l is sent to the nodes in D_i^l using
  // low-channel routing based on subnetwork N^l.
  $Low\_Channel\_Routing(M_i^l)$
endif

The simple multipath routing scheme includes four steps. First, In an n-star graph S_n, the destination node set D is partitioned into m subsets D_i^α, where $D_i^\alpha \neq \emptyset$, $1 \leq i \leq m$, $m \leq n-1$, and $\alpha \in \{h, l\}$. Second, the destination nodes in D_i^h are sorted according to the $\ell()$ values in ascending order and the destination nodes in D_i^l are sorted according to the $\ell()$ values in descending order, respectively. Third, we construct m messages M_i^α, where $1 \leq i \leq m$, $\alpha \in \{h, l\}$, and M_i^α contains D_i^α as part of the header. Finally, the next traversed node from the source node for routing each message M_i^α is the node that has the nearest label to that of the next *unvisited target* nodes of their neighboring nodes. The simple multipath algorithm is shown in Algorithm 1, whereas the high-channel routing and low-channel routing algorithms are shown as Proce-

Procedure 1: $High\_Channel\_Routing(M_j^h)$
// high-channel routing proceeds on subnetwork N^h
begin
  For message M_j^h which contains D_j^h do
  $c := s$
  loop
    // for every current node c, and next traversed
    // destination node d
    if $(D_j^h = \emptyset)$
      exit
    else // find next node x to traverse
      get the destination node d with least $\ell()$ value from D_j^h
      while $(c \neq d)$
        // M_j^h routing along higher $\ell()$ value
        // $x = RF(c,d)$, where x is the next traversed
        // node and RF is the routing function
        $\ell(x)$:= max$\{\ell(u)|\ell(c) < \ell(u) \leq \ell(d)$, and c is adjacent to $u\}$
        $c := x$
      endwhile
    endif
    // remove node d from D_j^h and message M_j^h
    $D_j^h := D_j^h - \{d\}$
    Remove d from message M_j^h
  endloop
end

Procedure 2: $Low\_Channel\_Routing(M_j^l)$
// low-channel routing proceeds on subnetwork N^l
begin
  For message M_j^l which contains D_j^l do
  $c := s$
  loop
    // for every current node c, and next traversed
    // destination node d
    if $(D_j^l = \emptyset)$
      exit
    else // find next node x to traverse
      get the destination node d with greatest $\ell()$ value from D_j^l
      while $(c \neq d)$
        // M_j^l routing along lower $\ell()$ value
        // $x = RF(c,d)$, where x is the next traversed
        // node and RF is the routing function
        $\ell(x)$:= min$\{\ell(u)|\ell(c) > \ell(u) \geq \ell(d)$, and c is adjacent to $u\}$
        $c := x$
      endwhile
    endif
    // remove node d from D_j^l and message M_j^l
    $D_j^l := D_j^l - \{d\}$
    Remove node d from message M_j^l
  endloop
end

dure 1 and Procedure 2, respectively.

In the following, we use the same multicast example as the one used for the hamiltonian-path and the dual-path routing to demonstrate the better multicast performance of the simple multipath routing when compared with the hamiltonian-path and the dual-path routing. In the sample multicast, the multicasting set R can be completed by three submulticasting sets, R_1^l, R_2^h, and R_3^h, where $R_1^l = \{\underline{2143}^8, 1243^7, 3124^2\}$, $R_2^h = \{\underline{2143}^8, 1342^{14}\}$, and $R_3^h = \{\underline{2143}^8, 4231^{23}\}$. In R_1^l, the first element is source node and the others are destination nodes with lower label values than source node in descending $\ell()$ value order. In R_2^h and R_3^h, the first elements are source nodes and the others are destination nodes with high label values than source nodes in ascending $\ell()$ value order. Then, R_1^l routes the message using low-channel routing based on subnetwork N^l, whereas R_2^h and R_3^h route the messages using high-channel routing based on subnetwork N^h. Figure 9 shows simple multipath routing. From Figure 9, the total number of channels traversed is 6+6+5=17, whereas the maximum routing distance from the source to a destination is max(6,6,5)=6. So, the total number of channels traversed of simple multipath routing is smaller than that of hamiltonian-path routing and equal to that of dual-path routing. The maximum routing distance of simple multipath routing is smaller than that of hamiltonian-path and dual-path routing.

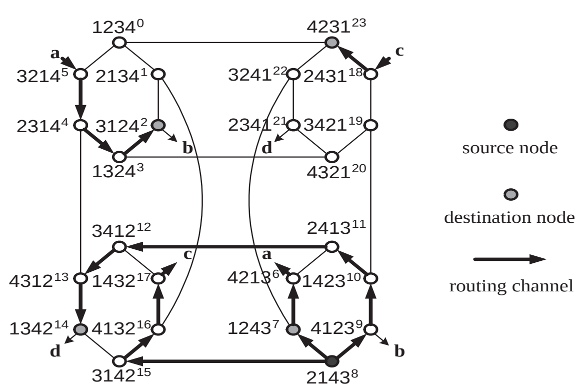


Figure 9: The sample multicast using simple multipath routing.

Now, let us discuss the time complexity of the simple multipath algorithm. Suppose n is the dimension of star graph, d is the number of destination nodes, and $N = n!$ is the number of nodes of star graph. In the destination-nodes partition step, the time complexity is $O(d)$. In the destination-nodes sorting step, the time complexity is $O(d \log d)$. In the message preparation step, the time complexity is $O(1)$. In the routing step, the time complexity is $O(\frac{3}{4}N)$ in the worst case. So, the total time complexity of the simple multipath algorithm in the worst case is $O(d) + O(d \log d) + O(1) + O(\frac{3}{4}N) = O(\frac{3}{4}n! + d \log d)$. On the other hand, the total time complexity of the hamiltonian-path and the dual-path algorithms are $O(d) + O(d \log d) + O(1) + O(N) = O(n! + d \log d)$ in the worst case.

To verify the correctness of the simple multipath routing algorithm, we derive the following lemmas and theorems.

Lemma 1. *For two arbitrary distinct nodes p and q in a star graph, the path from p to q selected according to the routing function RF is always existed.*

Proof. Suppose p and q are two arbitrary nodes in a star graph, without loss of generality, it can be assumed that $\ell(p) < \ell(q)$. Let the node c represent the source node or the intermediate node located in between source node p and destination node q on HP. Assume the next traversed node is x, according to the routing function RF, $x = RF(c, q)$, where $\ell(x)$ = max$\{\ell(u) | \ell(c) < \ell(u) \leq \ell(q)$, and c is adjacent to $u\}$. So, x is on HP going from c to q (including q) and adjacent (connected) to c. Then, the path from p to q selected according to the routing function RF is $(y_0, y_1, \cdots, y_k)$, where $y_0 = p$, $y_j = RF(y_{j-1}, q)$ for $0 < j \leq k$, and $y_k = q$. So, the path from p to q selected according to the routing function RF is always existed. □

Lemma 2. *The high-channel message routing, based on subnetwork N^h, in a star graph can always be completed.*

Proof. Based on Lemma 1, it is obvious. □

Lemma 3. *The low-channel message routing, based on subnetwork N^l, in a star graph can always be completed.*

Proof. Based on Lemma 1, it is obvious. □

Theorem 1. *The message routing using simple multipath algorithm in a star graph can always be completed.*

Proof. The message routing using simple multipath algorithm is performed by m submulticasts simultaneously. For each submulticast, the message routing can be completed via either high-channel subnetwork N^h or low-channel subnetwork N^l. According to Lemma 2 and Lemma 3, either high-channel or low-channel message routing can be completed. So, the message routing using multipath algorithm can always be completed. □

Theorem 2. *The simple multipath multicast routing is deadlock-free.*

Proof. Suppose that the destination node set D is partitioned into m subsets D_i^α, where $1 \leq i \leq m$ and $\alpha \in \{h, l\}$. The multicasting can be completed by m submulticasts simultaneously. For each submulticast, the multicasting is proceeded via one of the following two cases. Case 1 ($\alpha = h$): The message M_i^α is sent to the nodes in D_i^α using high-channel subnetwork N^h. Case 2 ($\alpha = l$): The message M_i^α is sent to the nodes in D_i^α using low-channel subnetwork N^l. Because the two subnetworks, N^h and N^l, are channel-disjoint, the multipath multicast routing is deadlock-free. □

3.2.2 Two-Phase Multipath Routing

In two-phase multipath routing, we intend to partition the destination nodes into multiple subsets. Then, the multicasting can be completed by two phases, source-to-relay and relay-to-destination. In each phase the simple-phase multipath routing is used.

This routing scheme includes four steps. First, in an n-star graph S_n, S_n can be partitioned into n disjoint $(n-1)$-dimensional substar graphs $S_{n-1}(1)$, $S_{n-1}(2)$, $\cdots$, $S_{n-1}(n)$ according to the nth symbol (the last dimension) of the nodes in S_n. The destination node set D is partitioned into n subsets θ_1, θ_2, $\cdots$, θ_n according to the nth symbol (the last dimension) of those destination nodes. In this way, the nodes of the same subset are located on the same $(n-1)$-dimensional substar graph. Second, for each subset θ_i, we can find a relay node r_i which is the node with the smallest label (value of $\ell()$) in the $(n-1)$-dimensional substar graph $S_{n-1}(i)$. Then, the message routing is proceeded by two phases: *source-to-relay* and *relay-to-destination*. In the *source-to-relay* phase, the message within the source node s is sent to the relay nodes r_i of θ_i using simple multipath routing. In the *relay-to-destination* phase, for each subset θ_i, the message received by relay node r_i is sent to all destination nodes in θ_i using simple multipath routing too. The two-phase multipath routing algorithm is shown in Algorithm 2.

For the sample multicast, introduced before for simple multipath routing, the two-phase multipath routing is proceeded as follows. In the destination-nodes partition step, we first define the destination node set θ as: $\theta = \{3124^2, 1243^7, 1342^{14}, 4231^{23}\}$. Then, θ is partitioned by the 4th symbol (the last dimension) of each node into four subsets: $\theta_1 = \{3124^2\}$, $\theta_2 = \{1243^7\}$, $\theta_3 = \{1342^{14}\}$, and $\theta_4 = \{4231^{23}\}$. In the relay-nodes finding step, for each subset θ_i we can find a relay node r_i which owns the smallest label in the 3-star graph $S_3(i)$.

Algorithm 2: *The two-phase multipath routing algorithm*
Input: Source node s, destination node set D, and node labeling function $\ell()$.
Step 1: // Destination-nodes partition
In an n-star graph S_n, S_n can be partitioned into n disjoint $(n-1)$-dimensional substar graphs $S_{n-1}(1)$, $S_{n-1}(2)$, $\cdots$, $S_{n-1}(n)$ according to the nth symbol (the last dimension) of the nodes in S_n. The nodes in destination node set D are collected into a set θ. θ is partitioned into n subsets θ_1, θ_2, $\cdots$, θ_n according to the nth symbol (the last dimension) of those destination nodes.
Step 2: // Relay-nodes finding
For each subset θ_i, we can find a relay node r_i which is the node with the smallest label (value of $\ell()$) in the $(n-1)$-dimensional substar graph $S_{n-1}(i)$.
Step 3: // Phase 1, *source-to-relay* routing
The message within the source node s is sent to the relay nodes r_i of θ_i using simple multipath routing.
Step 4: // Phase 2, *relay-to-destination* routing
For each subset θ_i, the message received by relay node r_i is sent to all destination nodes in θ_i using simple multipath routing too.

Thus, we obtain the relay nodes $r_1 = 1234^0$, $r_2 = 4213^6$, $r_3 = 3412^{12}$, and $r_4 = 2431^{18}$, respectively. Then, the multicasting is proceeded by following two phases. In the *source-to-relay* phase, the source node s routes a message to each of the relay nodes r_i. That is, the source node 2143^8 sends a multidestination message to the relay nodes 1234^0, 4213^6, 3412^{12}, and 2431^{18}. In this phase, the multicasting set is $R' = \{\underline{2143}^8, 1234^0, 4213^6, 3412^{12}, 2431^{18}\}$. R' is completed by routing three submulticasting sets, $R_1'^l$, $R_2'^h$, and $R_3'^h$, where $R_1'^l = \{\underline{2143}^8, 4213^6, 1234^0\}$, $R_2'^h = \{\underline{2143}^8, 3412^{12}\}$, and $R_3'^h = \{\underline{2143}^8, 2431^{18}\}$. In $R_1'^l$ the message is transmitted via low-channel routing based on subnetwork N^l, and in $R_2'^h$ and $R_3'^h$ the messages are sent through high-channel routing based on subnetwork N^h. In the *relay-to-destination* phase, each realy node r_i routes a message to destination nodes in the subset θ_i. That is, the relay nodes 1234^0, 4213^6, 3412^{12}, and 2431^{18} send a multidestination message to the destination nodes in each individual subset. In this phase, a multicasting set R'' is divided into four multicasting subsets: $R_1''^h = \{\underline{1234}^0, 3124^2\}$, $R_2''^h = \{\underline{4213}^6, 1243^7\}$, $R_3''^h = \{\underline{3412}^{12}, 1342^{14}\}$, and $R_4''^h = \{\underline{2431}^{18}, 4231^{23}\}$. Figure 10 shows the same multicast example of Figure 9 using two-phase multipath routing. For this multicast example, if we use two-phase multipath routing, the total number of channels traversed is (4+4+4)+(2+1+2+1)=18, and the maximum routing distance is max(4,4,4)+max(2,1,2,1)=6. So, the total number of channels traversed of two-phase multipath routing is smaller than that of hamiltonian-path routing but larger than that of dual-path routing. The maximum routing distance of two-phase multipath routing is smaller than that of hamiltonian-path and dual-path routing.

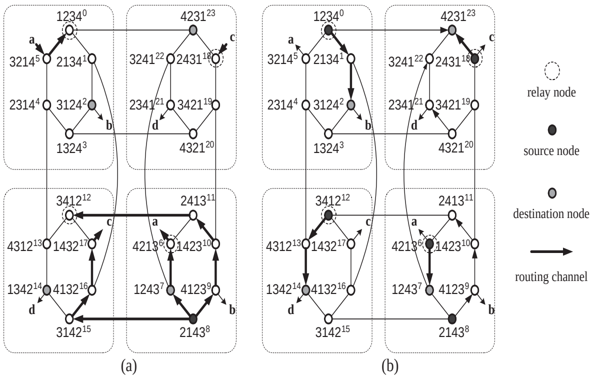


Figure 10: The sample multicast using two-phase multipath routing: (a) phase 1; (b) phase 2.

The time complexity of the two-phase multipath algorithm is computed as follows. Let n be the dimension of star graph, d be the number of destination nodes, and $N = n!$ be the number of nodes of star graph. In destination-nodes partition step, the time complexity is $O(d)$. In the relay-nodes finding step, the time complexity is $O(n)$ in the worst case. In *source-to-relay* phase, in the worst case the number of relay nodes is n, the time complexity is $O(n \log n)$ in the worst case. In *relay-to-destination* phase, the message is routed in the $(n-1)$-star graph, the time complexity is $O(\frac{3}{4}(n-1)! + d \log d)$ in the worst case. The total time complexity of the two-phase multipath algorithm in the worst case is $O(d) + O(n) + O(n \log n) + O(\frac{3}{4}(n-1)! + d \log d) = O(\frac{3}{4}(n-1)! + d \log d)$.

In Theorem 3 and Theorem 4, we prove that multicasting based on the two-phase multipath algorithm can always be completed and the routing is deadlock-free.

Theorem 3. *The message routing using two-phase multipath algorithm in a star graph can always be completed.*

Proof. The two-phase multipath routing is proceeded by two phases. In source-to-relay phase, the message is sent from source node to relay nodes using simple multipath routing. According to Theorem 1, the message routing can always be completed. In relay-to-destination phase, for each submulticast the message is sent from relay node to all destination nodes in the substar graph using simple multipath routing. Also according to Theorem 1, the message routing can always be completed. Thus, the message routing using two-phase multipath algorithm in a star graph can always be completed. □

Theorem 4. *The two-phase multipath multicast routing is deadlock-free.*

Proof. The two-phase multipath routing is proceeded by two phases. In source-to-relay phase, the message is sent from source node to relay nodes using simple multipath routing. According to Theorem 2, the message routing is deadlock-free. In relay-to-destination phase, for each submulticast, the message is sent from relay node to all destination node in the substar graph using simple multipath routing. Also according to Theorem 2, the message routing is deadlock-free. Thus, the two-phase multipath multicast routing is deadlock-free. □

In all our proposed multipath multicasting algorithms, we use the channel subnetworks that have been described in previous section. Because the subnetworks are disjoint and acyclic, no cyclic resource dependency can occur [8]. Thus, the proposed routing algorithms developed based on those subnetworks are deadlock-free.

Actually, the two-phase multipath multicasting can be conducted with various strategies. That is, for each phase the message can transmitted be either with multipath routing or with dual-path routing. Besides, the concept of two-phase multipath multicasting can also be extended to multiple-phase multipath multicasting.

4 Simulation Results

In this section, we shall present the performance of our proposed multicasting strategies by some simulation experiments. To evaluate the performance of the multicast schemes in an interconnection network, there are some parameters that must be considered: the multicast size, the message length, the startup latency, the link latency, and the router latency. The multicast size d is the number of destination nodes, and the message length f is the number of flits in a message. The message startup latency t_s includes the software overhead for buffers allocating, messages coping, router initializing, etc. The link latency t_l is the propagation delay of message through a link of network. The router latency t_r is the delay inside the router for handling multidestination messages.

We first give our assumptions to the parameters of system architecture in the simulations. All simulations were performed for a 720-node (6-dimension) star graph network. We examined the routing performance of our proposed schemes under various multicast sizes and message lengths. The source node and the destination nodes for each multicasting were randomly generated. The large message startup latency t_s is set to be 10.0 microseconds (5.5 microseconds for message sending latency, 4.5 microseconds for message receiving latency), and the small message startup latency t_s is 1.0 microsecond (550 nanoseconds for message sending latency, 450 nanoseconds for message receiving latency). The small message startup latencies were usually used for advanced network interface to improve the efficiency of latency time. The link propagation latency t_l is 5.0 nanoseconds. The router latency for handling multidestination messages t_r is 40.0 nanoseconds; however, it is set to 20.0 nanoseconds in unicast-based routing. For all of the multicasting, the message sizes of 6, 120, and 2400 flits were simulated.

4.1 Performance under Different Multicast Sizes

Figure 11 and Figure 12 present the performance of the various multicast schemes on a 6-star graph network with small and large message latencies, respectively. Results are shown for message lengths of 6,120, and 2400 flits, respectively. It is observed that, the performance of all path-based algorithms is superior to that of the unicast-based algorithm. This is because the unicast-based algorithm is a multiple-phase multicasting that needs more startup latency for processing.

In Figure 11, with small message startup latencies the performance of our proposed algorithms is superior to that of the unicast-based, the hamiltonian-path, and the dual-path algorithms except for very long messages. This is because the proposed algorithms uses multiple paths for simultaneous transmission that reduces the number of traversed links. The performance of the two-phase multipath algorithm is the best with short and medium message lengths. For long messages, the simple multipath algorithm performs the best. This is because in the two-phase multipath algorithm the message lengths plays a determining role on the performance of message transmission and its impact to transmission latency is larger for long messages, but smaller for short and medium messages. In general, for small message startup latencies with short and medium messages our proposed schemes are superior to the unicast-based, the hamiltonian-path, and the dual-path routing schemes.

Figure 12 shows the performance with large message startup latencies. With short and medium messages the performance of the simple multipath algorithm is better than that of the hamiltonian-path algorithm and worse than that of the dual-path algorithm for small number of destinations; however the performance of the simple multipath algorithm is worse than that of the hamiltonian-path and the dual-

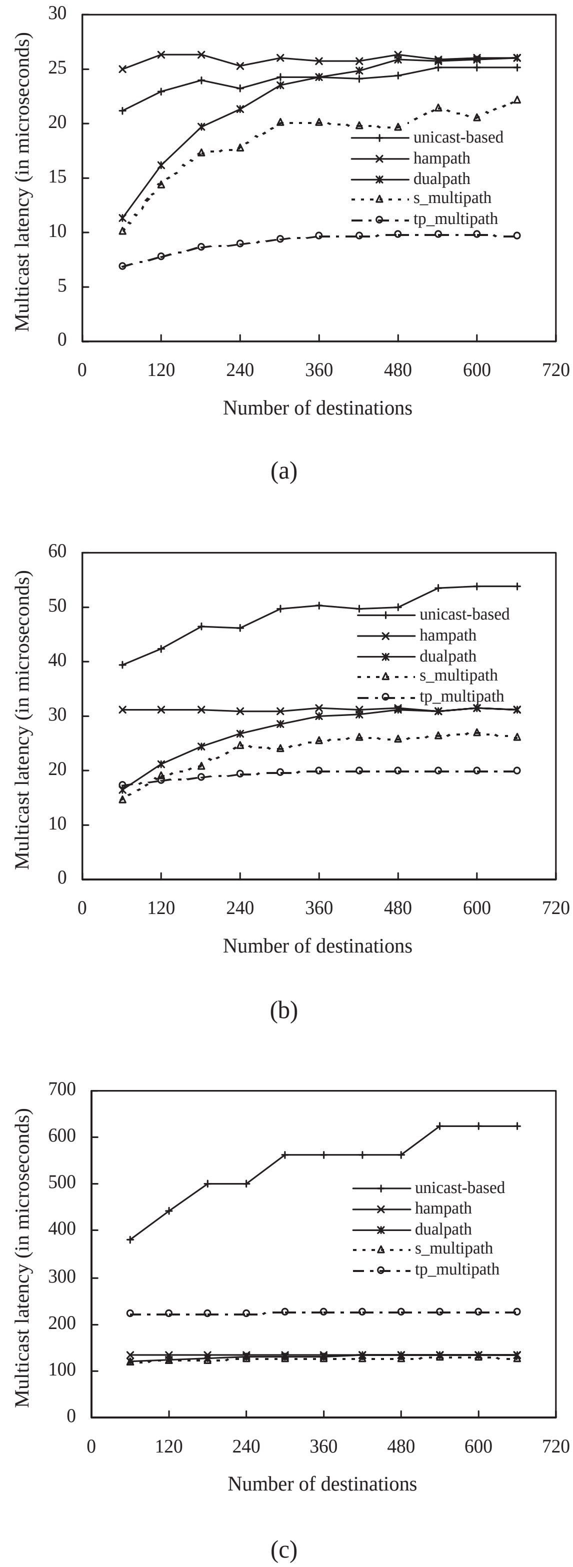
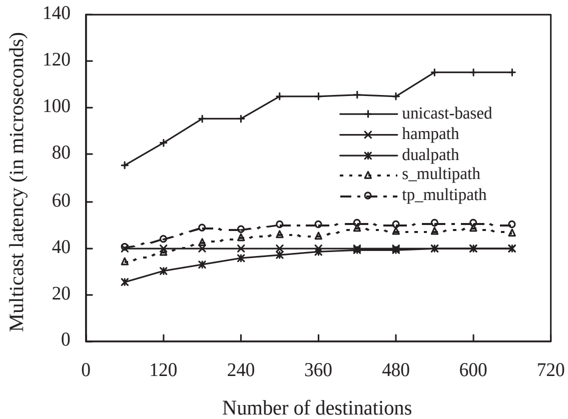
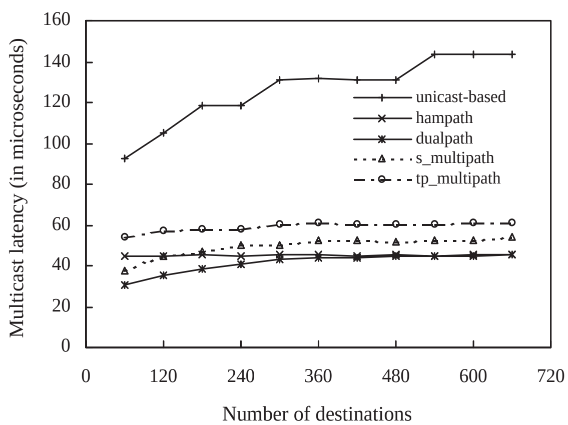


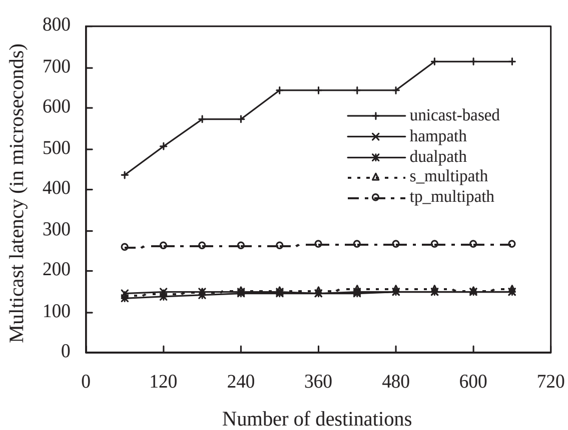
Figure 11: Multicast latency in a 6-star graph network with small message startup latency. (a) Message length = 6 flits. (b) Message length = 120 flits. (c) Message length = 2400 flits.



(a)



(b)



(c)

Figure 12: Multicast latency in a 6-star graph network with large message startup latency. (a) Message length = 6 flits. (b) Message length = 120 flits. (c) Message length = 2400 flits.

path algorithms for large number of destinations. With long messages the performance of the simple multipath algorithm is almost equal to the hamitlonian-path and the dual-path algorithms. This is because the simple multipath algorithm uses multiple paths for simultaneous transmission that needs more startup latency time (large message startup latency). The two-phase multipath algorithm performs worse than the hamiltonian-path and the dual-path algorithms for short, medium, and long messages. This is because the two-phase multipath algorithm is a two-phase routing strategy that needs more startup latency time (large message startup latency) for each phase.

4.2 Performance under Different Message Startup Latencies

In Figure 13, we show the influence of the message startup latency on the multicast latency. Here we set the number of destinations to be 120, and the flits to be 120 for each message. The latency increases faster with message startup latencies in the unicast-based algorithm. That is to say, the message startup latency has greater impact on the performance of the unicast-based algorithm. This is because the unicast-based algorithm needs multiple-phase to route the message. For the path-based algorithms, the impact of message startup latency of the proposed algorithms is almost equal to the hamiltonian-path and the dual-path algorithms.

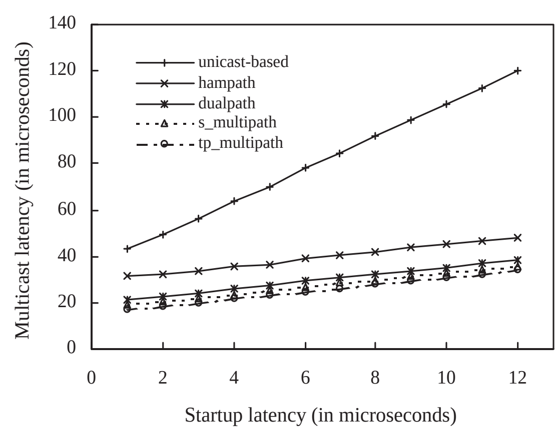
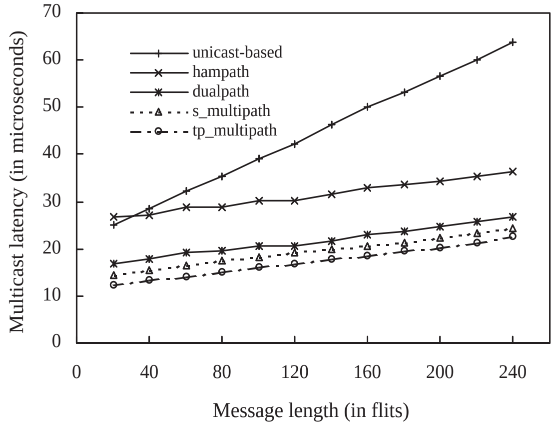


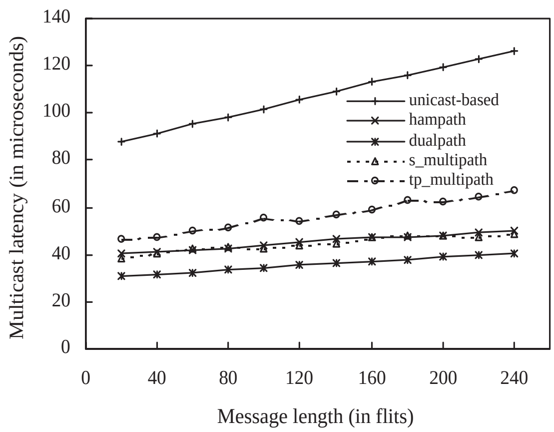
Figure 13: The effect of the startup latency in a 6-star graph network with 120 flits in message length and 120 nodes in multicast size on the multicast latency for various routing schemes.

4.3 Performance under Different Message Lengths

Figure 14(a) and Figure 14(b) show the performance with different message lengths under both



(a)



(b)

Figure 14: The effect of the message length in a 6-star graph network with 120 nodes in multicast size on the multicast latency for various routing schemes: (a) under small message startup latency; (b) under large message startup latency.

small and large message startup latencies, respectively. The number of destination nodes is assumed to be 120. The results show the multicast latencies are affected by the message length. As shown in Figure 14(a) and Figure 14(b), the path-based algorithms is least affected by the message length, while the unicast-based algorithm, requiring $\lceil \log_2 120 \rceil = 7$ phases, is most affected.

4.4 Utilization of Network Traffic

We then consider the traffic (in links) of interconnection networks. The network traffic may affect other communication in the network. We simulated the network traffic by the total number of links visited. Each link visited represents the use of one communication link by one message. Figure 15 presents the link usage for a 6-star graph network over various multicast sizes.

As shown in Figure 15, the proposed algorithms require fewer communication links than that of the unicast-based algorithm. For the path-based algorithms, the communication links of the simple multipath algorithm is almost equal to the dual-path algorithm, however the communication links of the two-phase multipath algorithm is larger than the dual-path algorithm.

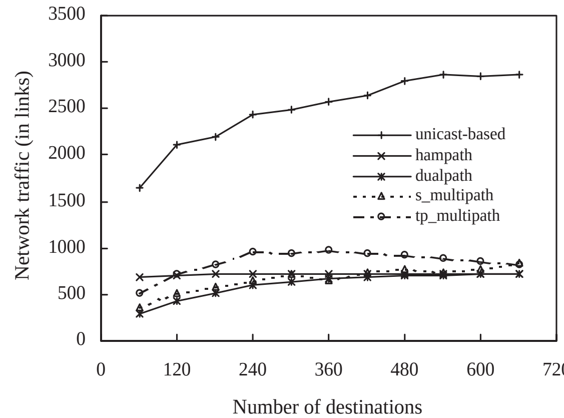


Figure 15: Network traffic in a 6-star graph network.

5 Conclusions

In this paper, we propose two efficient multipath multicast routing schemes for wormhole star graph networks. Both proposed schemes are proved deadlock-free. The first scheme, simple multipath routing, has the advantage of reducing the number of traversed links to improve the communication performance. The second one, two-phase multipath routing, has the advantage of reducing both the number of traversed links and parallel transmission. By the experimental results, for small message startup latencies with short and medium messages our proposed schemes are superior to the unicast-based, the hamiltonian-path, and the dual-path routing schemes significantly.

Acknowledgements

This work was supported by the National Science Council of Republic of China under Grant #NSC-90-2213-E-309-006.

References

[1] S. B. Akers, D. Harel, and B. Krishnamurthy, "The Star Graph : An Attractive Alternative to the n-Cube," *Proceedings of the 1987 International Conference on Parallel Processing*, pp. 393-400, August 1987.

[2] S. B. Akers and B. Krishnamurthy, “A Group-Theoretic Model for Symmetric Interconnection Networks,” *IEEE Trans. on Computers,* Vol. 38, No. 4, pp. 555-565, April 1989.

[3] T.-S. Chen, Y.-C. Tseng, and J.-P. Sheu, “Balanced Spanning Trees in Complete and Incomplete Star Graphs,” *IEEE Trans. on Parallel and Distributed Systems,* Vol. 7, No. 7, pp. 717-723, July 1996.

[4] T.-S. Chen, C.-Y. Chang, and J.-P. Sheu, “Efficient Path-Based Multicast in Wormhole-Routed Mesh Networks,” *Journal of Systems Architecture,* Vol. 46, No. 10, pp. 919-930, July 2000.

[5] T.-S. Chen, N.-C. Wang, and C.-P. Chu, “Multicast Communication in Wormhole-Routed Star Graph Interconnection Networks,” *Parallel Computing,* Vol. 26, No. 11, pp. 1459-1490, October 2000.

[6] K. Day, and A. Tripathi, “A Comparative Study of Topological Properties of Hypercubes and Star Graphs,” *IEEE Trans. on Parallel and Distributed Systems,* Vol. 5, No. 1, pp. 31-38, January 1994.

[7] K.-P. Fan and C.-T. King, “Turn Grouping for Multicast in Wormhole-Roued Mesh Networks Supporting the Turn Model,” *The Journal of Supercomputing,* Vol. 16, No. 3, pp. 237-260, July 2000.

[8] X. Lin, P. K. McKinley, and L. M. Ni, “Deadlock-Free Multicast Wormhole Routing in 2D Mesh Multicomputers,” *IEEE Trans. on Parallel and Distributed Systems,* Vol. 5, No. 8, pp. 793-804, October 1994.

[9] P. K. McKinley, H. Xu, A. H. Esfahanianm, and L. M. Ni, “Unicast-Based Multicast Communication in Wormhole-Routed Networks,” *IEEE Trans. on Parellel and Distributed Systems,* Vol. 5, No. 12, pp. 1252-1265, December 1994.

[10] V. E. Mendea and D. Sarkar, “Optimumal Broadcasting on the Star Graph,” *IEEE Trans. on Parellel and Distributed Systems,* Vol. 3, No. 4, pp. 389-396, July 1992.

[11] M. Nigam, S. Sahni, and B. Kirshnamurthy, “Embedding Hamiltonians and Hypercubes in Star Interconnection Graphs,” *Proceedings of International Conference on Parallel Processing,* Vol. 3, pp. 340-343, August 1990.

[12] D. K. Panda, S. Singal, and R. Kesavan, “Multidestination Message Passing in Wormhole k-ary n-cube Networks with Base Routing Conformed Path,” *IEEE Trans. on Parellel and Distributed Systems,* Vol. 10, No. 1, pp. 76-96, January 1999.

[13] K. Qiu, S. G. Akl, and H. Meijer, “On Some Properties and Algorithms for the Star and Pancake Interconnection Networks,” *Journal of Parallel and Distributed Computing,* Vol. 22, No. 1, pp. 16-25, July 1994.

[14] D. F. Robinson, P. K. McKinley, and B. H. C. Cheng, “Path-Based Multicast Communication in Wormhole-Routed Unidirectional Torus Networks,” *Journal of Parallel and Distributed Computing,* Vol. 45, No. 2, pp. 104-121, September 1997.

[15] J.-P. Sheu, W.-H. Liaw, and T.-S. Chen, “A Broadcasting Algorithm in Star Graph Interconnection Networks,” *Information Processing Letters,* Vol. 48, pp. 237-241, September 1993.

[16] J.-P. Sheu, C.-T. Wu, and T.-S. Chen, “An Optimal Broadcasting Algorithm without Message Redundancy in Star Graphs,” *IEEE Trans. on Parallel and Distributed Systems,* Vol. 6, No. 6, pp. 653-658, June 1995.

[17] Y.-C. Tseng, D. K. Panda, and T.-H. Lai, “A Trip-Based Multicasting Model for Wormhole-Routed Networks with Virtual Channels,” *IEEE Trans. on Parallel and Distributed Systems,* Vol. 7, No. 2, pp. 138-150, February 1996.

[18] Y.-C. Tseng and J.-P. Lai, “Toward Optimal Broadcast in a Star Graph Using Multiple Spanning Trees,” *IEEE Trans. on Computers,* Vol. 46, No. 5, pp. 593-599, May 1997.

[19] Y.-C. Tseng, “Multi-Node Broadcasting in Hypercubes and Star Graphs,” *Journal of Information Science and Engineering,* Vol. 14, No. 4, pp. 593-599, December 1998.

[20] Y.-C. Tseng, M.-H. Yang, and T.-Y. Juang, “Achieving Fault-Tolerant Multicast in Injured Wormhole-Routered Tori and Meshes Based on Euler Path Construction,” *IEEE Trans. on Computers,* Vol. 49, No. 3, pp. 1282-1296, March 2000.