ALL-TO-ALL BROADCASTING IN 1-D HYPERCYCLES

R. Sivakumar *N. J. Dimopoulos*

Technical Report ECE 94-4

Department of Electrical and Computer Engineering
University of Victoria
Victoria, B.C.
CANADA - V8W 3P6

April 1994

Abstract

In this work, we present an efficient algorithm for performing all-to-all broadcasting in 1-D Hypercycles. We show that this scheme takes exactly $\left\lceil \frac{m-1}{2\rho} \right\rceil$ using 2ρ-port communication where m is the number of nodes in the 1-D Hypercycles and ρ is the connectivity.

1. Introduction

Broadcasting is an important communication protocol in message passing concurrent computers [1]. All-to-all broadcasting is a broadcasting scheme in which every node in the network receives messages from all the remaining nodes. This work focuses on hypercycles which are a class of multidimensional, symmetric graphs that are generalizations of *hypercubes*. They embody a cyclic interconnection topology with variable connectivity, incremental expandability, simple routing and enhanced fault-tolerant capabilities. These networks have the ability, given a fixed degree, to form a number of alternate size graphs and therefore can match the topology of a given application. In this report, an efficient all-to-all broadcasting scheme using simultaneous 2ρ-port communication for 1-D Hypercycles.

1.1. Definitions and Preliminaries

An r-D hypercycle [3] is defined using the mixed radix number system and is represented by a simple, finite, undirected graph $\mathcal{G}_{\vec{m}}^{\vec{\rho}} = \{\mathcal{V}_{\vec{m}}^{\vec{\rho}}, \mathcal{E}_{\vec{m}}^{\vec{\rho}}\}$ where $\mathcal{V}_{\vec{m}}^{\vec{\rho}} = \{0, 1, 2, \ldots M-1\}$ is the set of nodes, $\mathcal{E}_{\vec{m}}^{\vec{\rho}}$ is the set of edges, $\vec{m} = [m_1\ m_2\ \ldots\ m_r]$ is the radix vector such that $M = \prod_{i=1}^{r} m_i$, $m_i \geq 2$, and $\vec{\rho} = [\rho_1\ \rho_2\ \ldots\ \rho_r]$ is the connectivity vector where $1 \leq \rho_i \leq \lfloor \frac{m_i}{2} \rfloor$. The diameter k of the graph is given by $k = \sum_{i=1}^{r} \left\lceil \frac{\lfloor \frac{m_i}{2} \rfloor}{\rho_i} \right\rceil$. The degree of the graph d is given by $d = \sum_{i=1}^{r} d_i$ where

$$d_i = \begin{cases} 2\rho_i & \text{if } m_i \neq 2\rho_i \\ 2\rho_i - 1 & \text{otherwise} \end{cases}$$

Therefore, the number of ports for each node is defined by the degree of the network as above.

2. All-to-All Broadcasting

We make the following assumptions in the broadcasting scheme for 1-D Hypercycle $\mathcal{G}_{\vec{m}}^{\vec{\rho}}$ where $\vec{m} = [m]$ and $\vec{\rho} = [\rho]$.

1. Each message at node N_i is represented by D_i, $0 \leq i \leq m-1$, and all messages are of identical size.

2. Node numbers are assigned in a clockwise direction for convenience. When $m > 2\rho$, we logically partition the ports as *right ports* and *left ports* of node i. such that $n_r = n_l = \rho$ where n_r and n_l represent the cardinality of the right ports and left ports of node i respectively[1]. Each right (left) port of node i is named as RP_j^i (LP_j^i), $1 \leq j \leq \rho$, $0 \leq i \leq m-1$, such that RP_j^i (LP_j^i) is connected to LP_j^{i-j} (RP_j^{i+j}).

[1] All node/port address are evaluated as modulo m.

3. When $m = 2\rho$, we partition the ports of node i as *right ports*, *center port* and *left ports* such that there is exactly one center port with $n_r = n_l = \rho - 1$. Then RP_j^i (LP_j^i) is connected to LP_j^{i-j} (RP_j^{i+j}), $1 \leq j \leq \rho - 1$, $0 \leq i \leq m - 1$, while CP^i is linked to $CP^{i+\rho}$. An example is shown in Fig. 1 which illustrates the port naming conventions for a 1-D Hypercycle.

4. There are enough buffer registers and local memory for storing the intermediate messages.

5. Each port permits full duplex transmission. i.e., messages can be transmitted and received through the ports of a given node simultaneously.

We need a multinode broadcasting algorithm (MBA) where every node sends a message to all the other nodes (N simultaneous broadcasts). We first derive the lower bound, to guide us through the performance assessment of any algorithms we design.

Lower Bound of MBA

Every node must receive in total $M - 1$ messages. Since each node has 2ρ incoming links incident to it, the minimum number of steps needed to have all messages received is $(M-1)/2\rho$. Thus we have

$$T(MBA) \quad \geq \quad \left\lceil \frac{M-1}{2\rho} \right\rceil \tag{1}$$

Link utilization at the lower bound

Since each node has to receive $M - 1$ messages, we have $M(M-1)$ receptions in total in the network. If this has to be accomplished by a given algorithm that with a lower bound given previously, then on the average there are

$$\frac{M(M-1)}{\left\lceil \frac{M-1}{2\rho} \right\rceil} \approx 2M\rho$$

receptions per step. Hence, on the average there are $2M\rho$ links active in each step. But the number of edges in the graph is

$$\left(\sum_{i=1}^{M} \text{degree of } i\text{th vertex} \right) /2 = (M \times 2\rho)/2 = M\rho$$

Since each edge corresponds to 2 links, we have in total $2\rho M$ links. Hence, *all* the links should be active in every step and therefore an optimal MBA should have 100% link utilization.

2.1. Discussion

For the hypercube and simple rings ($\rho = 1$), there exists an algorithm [2] that achieves the lower bound defined by Eq. (2). In Fig. 2, we present an optimal algorithm that runs at the lower bound

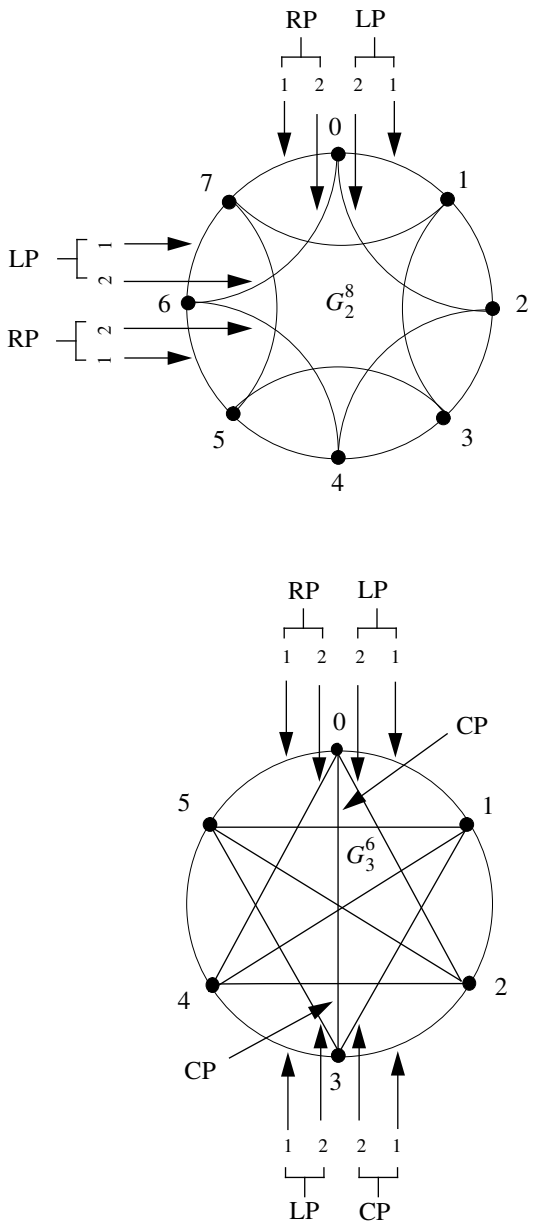


Figure 1 *Port naming conventions in* $\mathcal{G}_8^2$

for 1-D hypercycles. The input to the algorithm consists of the network parameters, namely, ρ and m and also the message D_i originating at Node N_i, $i = 0, \ldots, m-1$. Three counter variables w_i, x_i and y_i are introduced as shown in Line (1) of Fig. 2 and these are essential to track the messages broadcast without introducing redundant steps. We shall now prove the correctness of the algorithm and explain its operation through Theorem 1.

Theorem 1 *Given a 1-D hypercycle $\mathcal{G}_m^\rho$, an all-to-all broadcasting scheme shown in Fig. 2 is completed in $\left\lceil \frac{m-1}{2\rho} \right\rceil$ steps.*

Proof. To establish the correctness of the algorithm, let us consider the following cases.

Case 1: $\frac{m-1}{2\rho}$ is an integer and $m > 2\rho$

Subcase 1: $\frac{m-1}{2\rho} = 1$

The variables $w_i = \frac{m-1}{2\rho} = 1$ and $x_i = 0$ are computed in Line (1) and control branches to Line (5) since $m > 2\rho$. At the end of Line (7), messages accumulated in the right and left ports of N_i in the first step is given by

$$N_i^1 \leftarrow \{\underbrace{D_{i-1}, D_{i-2}, \ldots, D_{i-\rho}}_{RP}, \quad \underbrace{D_{i+1}, D_{i+2}, \ldots, D_{i+\rho}}_{LP}\} \tag{2}$$

since

$$LP_j^i \leftarrow D_{i+j} \tag{3}$$

$$RP_j^i \leftarrow D_{i-j} \tag{4}$$

where $1 \leq j \leq \rho$. As the hypercycle is fully connected ($\rho = \frac{m-1}{2}$), each node N_i receives messages D_j from all the other nodes N_j such that $j \neq i$ and $\forall j$, $0 \leq j \leq m-1$.

Subcase 2: $\frac{m-1}{2\rho} > 1$

At the end of the first iteration, node N_i contains messages which are transmitted by other nodes and Eq. (3) defines its constituent messages. The counter w_i is decremented in Line (24) and in the second iteration, control branches to Line (25) since $x_i = 0$. In Line (20), if we replace the index i by $(i-j) \bmod m$, $1 \leq j \leq \rho$, it follows that $RP_j^i \leftarrow RP_\rho^{i-j}$(i.e, Node i receives the data present in the ρ^{th} right port of Node $i-j$). From Eq. 5, it is easy to note that RP_ρ^{i-j} contained $D_{i-j-\rho}$ at the end of the first step. Hence RP_j^i stores $D_{i-j-\rho}$ during the second step. Similarly, in Line (21), it can be seen that $LP_j^i \leftarrow LP_\rho^{i+j}$. From Eq. (4), we can infer that LP_ρ^{i+j} contained $D_{i+j+\rho}$ during the first step. Hence $LP_j^i \leftarrow D_{i+j+\rho}$. Therefore, at the end of the second step, we have

$$N_i^2 \leftarrow \{\underbrace{D_{i-\rho-1}, D_{i-\rho-2}, \ldots, D_{i-2\rho}}_{RP}, \quad \underbrace{D_{i+\rho+1}, D_{i+\rho+2} \ldots, D_{i+2\rho}}_{LP}\} \tag{5}$$

The above procedure repeats itself and after some t steps where $1 \leq t \leq \frac{m-1}{2\rho}$ and $w_i = \frac{m-1}{2\rho} - t$, $RP_j^i \leftarrow D_{i-j-(t-1)\rho}$ and $LP_j^i \leftarrow D_{i+j+(t-1)\rho}$

$$N_i^t \leftarrow \{\underbrace{D_{i-(t-1)\rho-1}, D_{i-(t-1)\rho-2}, \ldots, D_{i-t\rho}}_{RP}, \quad \underbrace{D_{i+(t-1)\rho+1}, D_{i+(t-1)\rho+2} \ldots, D_{i+t\rho}}_{LP}\} \tag{6}$$

Algorithm: All-to-All broadcasting in 1-D Hypercycles

Input: Total number of nodes m, Connectivity ρ, Data D_i at node N_i, $i = 0, \ldots, m-1$

Output: Message D_i broadcast to all other nodes N_j, $j \neq i$, $j = 0, \ldots, m-1$

Execute the following code in parallel at each node N_i, $0 \leq i \leq m-1$

(1) **Initialize** counters $w_i = \left\lceil \frac{m-1}{2\rho} \right\rceil$, $x_i = \left\lceil \frac{(m-1) \bmod 2\rho}{2} \right\rceil$, $y_i = \left\lfloor \frac{(m-1) \bmod 2\rho}{2} \right\rfloor$
(2) **While** $w_i > 0$ **begin**
(3) **If** $w_i = \left\lceil \frac{m-1}{2\rho} \right\rceil$ **then begin**
(4) **If** $m > 2\rho$ **then begin**
(5) Transmit D_i to the j^{th} right port of Node $i+j$, $\forall j$ such that $1 \leq j \leq \rho$
(6) Transmit D_i to the j^{th} left port of Node $i-j$
(7) **end**
(8) **else begin**
(9) Transmit D_i to the l^{th} right port of Node $i+l$, $\forall l$ such that $1 \leq l \leq \rho - 1$
(10) Transmit D_i to the l^{th} left port of Node $i-l$
(11) Transmit D_i to the center port of Node $i+\rho$
(12) **end**
(13) **end**
(14) **else begin**
(15) **If** $(w_i = 1)$ and $(x_i \neq 0)$ **then begin**
(16) Transmit data received at the ρ^{th} right port of Node i (along the j'^{th} left port) to Node $i+j'$, $\forall j'$ such that $1 \leq j' \leq x_i$
(17) Transmit data received at the ρ^{th} left port of Node i (along the j''^{th} right port) to Node $i-j''$, $\forall j''$ such that $1 \leq j'' \leq y_i$
(18) **end**
(19) **else begin**
(20) Transmit data received at the ρ^{th} right port of Node i (along the j^{th} left port) to Node $i+j$, $\forall j$ such that $1 \leq j \leq \rho$
(21) Transmit data received at the ρ^{th} left port of Node i (along the j^{th} right port) to Node $i-j$, $\forall j$ such that $1 \leq j \leq \rho$
(22) **end**
(23) **end**
(24) $w_i = w_i - 1$
(25) **end**
(26) Stop /* All-to-All broadcasting is complete */

Figure 2 *An algorithm for all-to-all broadcasting in 1-D Hypercycles*

It can be inferred that the messages received at N_i during each iteration is unique and when $t = \frac{m-1}{2\rho}$, N_i has a copy of all messages D_j such that $j \neq i$, $0 \leq j \leq m-1$.

Case 2: $\frac{m-1}{2\rho}$ **is not an integer and** $m > 2\rho$

The all-to-all broadcasting scheme is exactly identical as discussed above till $w_i = 1$ or equivalently until $t = \left\lfloor \frac{m-1}{2\rho} \right\rfloor$ steps. Then Eq. (7) defines the messages received at N_i after the t^{th} step with $RP_j^i \leftarrow D_{i-j-(t-1)\rho}$ and $LP_j^i \leftarrow D_{i+j+(t-1)\rho}$. Note that the variable x_i computed initially at Line (1) is non zero for Case 2. When Lines (16) and (17) are executed, we obtain $RP_{j'}^i \leftarrow D_{i-j'-t\rho}$ where $1 \leq j' \leq x_i$ and $LP_{j''}^i \leftarrow D_{i+j''+t\rho}$ where $1 \leq j'' \leq y_i$. Therefore, at the $t+1^{th}$ step, we have

$$N_i^{t+1} \leftarrow \{\underbrace{D_{i-t\rho-1}, D_{i-t\rho-2}, \ldots, D_{i-t\rho-x_i}}_{RP}, \underbrace{D_{i+t\rho+1}, D_{i+t\rho+2} \ldots, D_{i+t\rho+y_i}}_{LP}\} \quad (7)$$

Observe that each of the above messages is distinct and there is no overlapping with previously received messages. Hence the total number of messages received by N_i at the $t+1^{th}$ step is given by $x_i + y_i = (m-1) \bmod 2\rho$.

Case 3: $m = 2\rho$

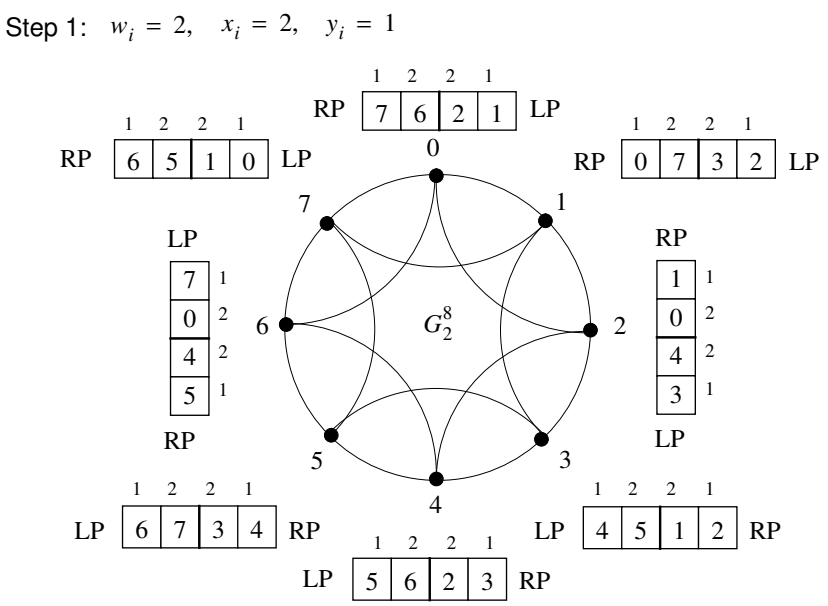
The hypercycle is fully connected and all-to-all broadcasting is completed in one step similar to subcase 1 above. However, the only major difference is that the number of ports and their naming conventions differ compared to the case $m > 2\rho$. Lines (9)-(11) of the algorithm are executed when $m = 2\rho$ and node N_i holds the following messages.

$$N_i^1 \leftarrow \{\underbrace{D_{i-1}, D_{i-2}, \ldots, D_{i-\rho-1}}_{RP}, \underbrace{D_{i+\rho}}_{CP}, \underbrace{D_{i+1}, D_{i+2}, \ldots, D_{i+\rho-1}}_{LP}\} \quad (8)$$

Note that $D_{i+\rho}$ is obtained from the center port of $N_{i+\rho}$. Hence, all $2\rho - 1 = m-1$ messages are received by each node at the end of the first step.

■

As an example, Fig. 3 illustrates the trace of the all-to-all broadcasting algorithm in $\mathcal{G}_8^2$. As can be seen, after 2 steps, the all-to-all broadcasting is complete and all messages received at each node are unique.



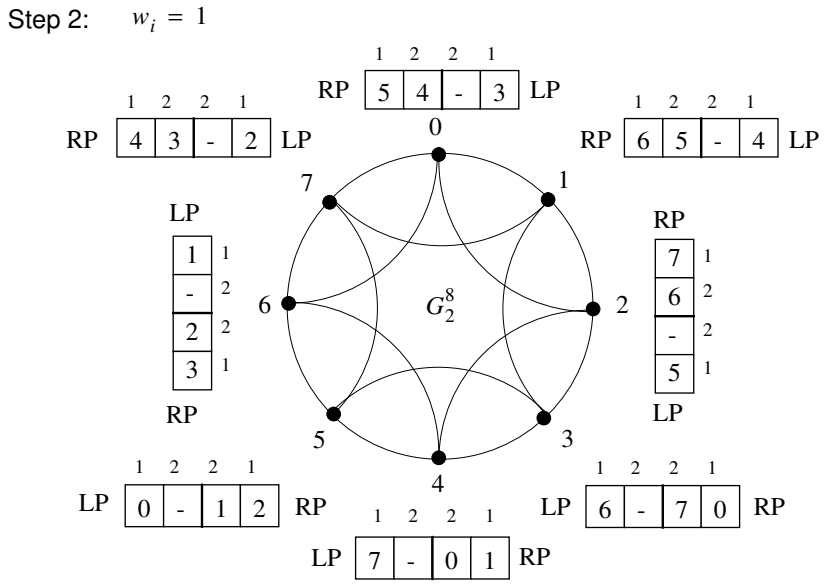


Figure 3 *All-to-All broadcasting in* $\mathcal{G}_8^2$

References

[1] S. L. Johnsson and C-T. Ho, "Optimum broadcasting and personalized communication in Hypercubes", *IEEE Trans. Comp.*, Vol. 38, Sept, 1989, pp. 1249-1268.

[2] D. P. Bertsekas, J. H. Tsitsiklis, *Parallel and Distributed Computation*, Prentice-Hall, New Jersey, 1989.

[3] N. J. Dimopoulos, R. D. Rasmussen, G. S. Bolotin, B. F. Lewis, and R. F. Manning, "Hypercycles, Interconnection Networks with simple Routing Strategies", *Proceedings of the Canadian Conference on Electrical & Computer Engineering*, Nov. 1988, pp. 577-580.