

대규모 병렬컴퓨터에 적합한 교차메쉬구조에서의 경로설정

김 종 진[†] · 윤 성 대^{††}

요 약

본 논문에서는 대규모 병렬컴퓨터의 구현에 적합한 위상구조인 교차메쉬구조에서, 균등한 메시지분포를 갖고 메시지의 경쟁이 있는 실제적 상황에서 상호접속망내의 메시지의 경로를 선정하기 위한 방법들을 제안하고, 이의 성능을 검증하기 위해 우회 경로설정 알고리즘을 이용하여 시뮬레이션을 하였다. 교차메쉬의 특성상 최적경로의 수가 다른 구조들에 비해 다양하다는 특징을 이용하여 최적경로의 수를 우선순위에 반영한 우회조건 및 대각방향의 링크를 효율적으로 활용하기 위한 링크선정방법에 따른 교차메쉬의 최대지연(maximum delay), 평균지연(average delay) 및 메시지처리율(throughput)을 구하고 이를 비교 고찰하였다.

메시지 전송시 최적인 경로상의 링크에 경쟁이 생길 경우 최적 경로의 수가 적은 메시지가 높은 우선순위를 가지며 만약 같은 조건이라면 우회한 횟수가 많은 메시지가 높은 우선순위를 갖는 우회조건 LD를 사용하며 이 우선순위에 따라 경로를 선정할 차례가 된 메시지가 선택할 수 있는 최적경로의 수가 많은 경우 대각방향의 링크로 우선적으로 전송할 경우, 오래된 메시지가 높은 우선순위를 갖는 우회조건 A에 의한 방법에 비해 최대지연, 평균지연 및 메시지처리율에 있어서 각각 이상값에 대한 개선목표치의 약 58%, 70% 및 31%의 성능개선이 있었다.

Routing Strategy on the X_MESH Topology for the Massively Parallel Computer Architecture

Jong-Jin Kim[†] · Sung-Dae Youn^{††}

ABSTRACT

We propose the routing strategies on the X_MESH topology for the massively parallel computer architecture with uniform message distribution and in the presence of link contention. And to confirm these strategies, the deflection routing algorithm and self-routing algorithm for the X_MESH is used to simulate the maximum delay, the average delay, and the throughput in the presence of contention.

For the deflection routing, we propose a new deflection criterion LD that the message with fewer optimal directional links has a higher priority, but for the messages with the same number of optimal directional links, the message with more deflections has a higher priority. And if the message that has the right to select link by the criterion has many optimal directional links, we adopt the strategy that the message has the preference for the diagonal link. The simulation results show that the maximum delay, the average delay, and the throughput using the proposed strategies are enhanced by about 58%, 70%, and 31%, respectively, compared with them using criterion A that the older message has a higher priority.

[†] 정 회 원 : 부경대학교 전자공학과 교수

^{††} 정 회 원 : 부경대학교 전자계산학과 교수

논문접수 : 1998년 9월 7일, 심사완료 : 1998년 10월 27일

1. 서 론

수천 혹은 수만 개의 프로세서로 구성되는 고성능의 대규모 병렬컴퓨터시스템을 구현하기 위해서는 효과적인 상호접속망을 선택하는 것이 필수적이다. 과제를 여러 개의 부태스크로 분할하고 이를 각 프로세서에 최적 할당하여 병렬 처리하는 병렬컴퓨터 시스템에서 과제가 매우 클 경우, 부태스크의 수가 증가하여 프로세서의 수가 증가하게 되면 상호접속망을 통한 이들 프로세서간의 통신부담(communication overhead)이 시스템의 성능을 저해하는 큰 요인중의 하나가 된다. 따라서 대규모 병렬컴퓨터에서는 통신대역폭이 커서 통신지연이 적어 원활한 프로세서간 통신을 해결해 줄 수 있는 상호접속망 및 경로설정방법이 필요하다[1].

최근 반도체 기술의 발달로 고집적 병렬컴퓨터의 제조가 기술적 및 경제적으로 가능하게 되었다. 대규모 병렬컴퓨터를 칩에 구현하는 VLSI 또는 WSI(wafer scale integration)에서는 인접 노드끼리 연결되는 메쉬 구조가 적합하다고 알려져 있다.[2] 토로이달메쉬(toroidal mesh) 즉 토로이드(toroid)는 그 위상구조가 유체역학 및 유한요소법에 의한 구조해석 등의 과제의 병렬태스크 그래프와 잘 정합되므로 이런 분야의 과학 기술 계산에 널리 활용되고 있으나 과제가 대단히 클 때는 노드수가 증가하여 망의 직경 및 노드간 평균거리의 증가에 따른 통신지연이 커진다.

Zohorjan등[3]은 균등메시지분포를 가지며 일정한 수의 메시지를 가진 시스템에서 메시지처리율의 상한을 구하여 접속망의 구조적 성능을 해석하는 방법을 제안하였으며 Reed등[4]은 균등메시지분포의 경우보다 더 현실적인 가정 하에서 시스템의 성능을 분석하기 위해 메시지 routing분포라는 개념을 도입하여 좀 더 실질적인 성능 해석 방안을 제시하였다. 그러나 이들 방법은 접속망의 구조해석만에 의한 성능 평가 방법이므로 컴퓨터의 각 노드 및 링크 등 한정된 자원에 대한 경쟁이 생겼을 때의 경로설정정에 따른 실제 부담은 고려되지 않은 성능 평가 방법이다. Tang등[5]은 Barnes[6]가 제안한 우회경로(deflection routing)알고리즘을 이용하여 직사각형구조의 대각메쉬(diagonal mesh)[7]가 동일 규모의 토로이드에 비해 성능이 우수함을 보였다. 그러나 정방형 대각메쉬는 토로이드에 비해 성능이 개선되지 않는다.

Kim등[8]은 토로이드의 수평방향의 링크는 그대로

사용하고 수직방향링크 대신 대각선방향으로 교차 접속시킴으로써 토로이드나 대각메쉬에 비해 망의 직경 및 노드간 평균거리를 줄인 교차메쉬(crossed mesh: X-MESH)구조를 제안하였다. 그리고 교차메쉬를 위한 최적 자기경로설정 알고리즘(optimal self-routing algorithm)을 개발하였으며, 구조의 성능을 우회경로알고리즘을 사용하여 시뮬레이션하고 이를 검토 비교하였다.

본 논문에서는 대규모 병렬컴퓨터의 구현에 적합한 위상구조인 교차메쉬구조의 경우, 균등한 메시지분포를 갖고 메시지의 경쟁이 있는 실제적 상황에서 상호접속망내의 메시지의 경로를 선정하기 위한 방법들을 제안하고, 이의 성능을 검증하기 위해 우회 경로설정 알고리즘을 이용하여 시뮬레이션을 하였다. 교차메쉬의 특성상 최적경로의 수가 다른 구조들에 비해 다양하다는 특징을 이용하여 최적경로의 수를 우선순위에 반영한 우회조건 및 대각방향의 링크를 효율적으로 활용하기 위한 링크선정방법에 따른 교차메쉬의 최대지연(maximum delay), 평균지연(average delay) 및 메시지처리율(throughput)을 구하고 이를 비교 고찰하였다.

2. 교차메쉬의 위상구조 및 경로설정

2.1 위상구조(topology)

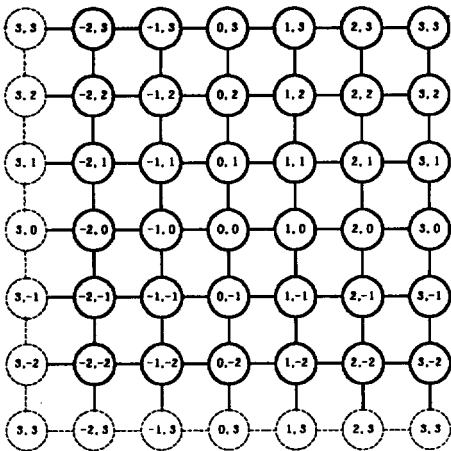
교차메쉬의 노드수 N 은 위상구조가 정방형의 한 축상의 노드수를 w 라고 할 때 $N=w \times w$ 가 되며, 정방형이 아닐 때는 수직방향으로 n 개 수평방향으로 k 개씩 노드를 배치하여 $N=n \times k$ 로 주어진다. 이 때 각 노드의 주소는 순서쌍 (x, y) 로 표현된다. 여기서 x 및 y 는 각각 수평 및 수직축 좌표로서 $-k/2 < x \leq k/2$ 및 $-n/2 < y \leq n/2$ 이며, 위상구조적 특성상 n 및 k 는 짝수이다.

(그림 2)에서 1행의 노드들을 7행에, 7열의 노드들을 1열에 점선으로 이중표기함으로써 종단접속링크를 나타내었다. 여기서 1행 및 7행의 노드와 7열 및 1열의 노드들 중 동일한 주소를 갖는 노드는 동일노드를 의미한다. 그리고 임의의 노드 (x, y) 의 4개의 인접노드를 (그림 3)에 나타내었다. (그림 3)에서 종단접속링크로 연결되는 노드를 표현하기 위하여, 이 구조의 대칭성 및 상대주소개념에 따라 기호 $\{x\}_k$ 를

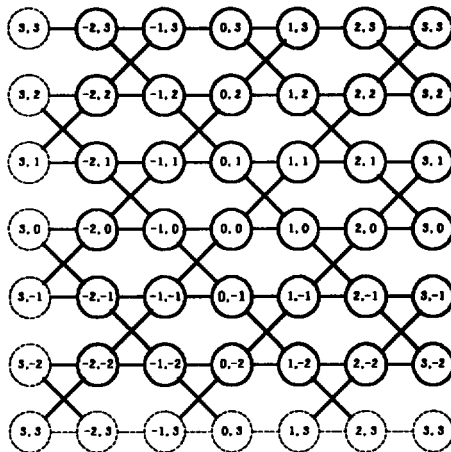
$$\{x\}_k = \begin{cases} x, & \text{if } -k/2 < x \leq k/2 \\ x - k, & \text{if } x > k/2 \\ x + k, & \text{if } x \leq -k/2 \end{cases} \quad (1)$$

와 같이 정의하였다. (그림 3)에서 보는 바와 같이 임의의 노드 (x, y) 에서 $x+y$ 가 짝수일 때는 $((x+1)_k, y)$, $((x-1)_k, y)$, $((x+1)_k, (y+1)_n)$ 및 $((x-1)_k, (y-1)_n)$ 노드와 직결되어 동쪽(E), 서쪽(W), 북동쪽(NE) 및 남서쪽(SW)으로 연결링크를 가지며 $x+y$ 가 홀수일 때는 $((x+1)_k, y)$, $((x-1)_k, y)$, $((x-1)_k, (y+1)_n)$ 및 $((x+1)_k, (y-1)_n)$ 노드와 직결되어 동쪽(E), 서쪽(W), 북서쪽(NW) 및 남동쪽(SE)으로 4개의 링크를 갖는다.[8]

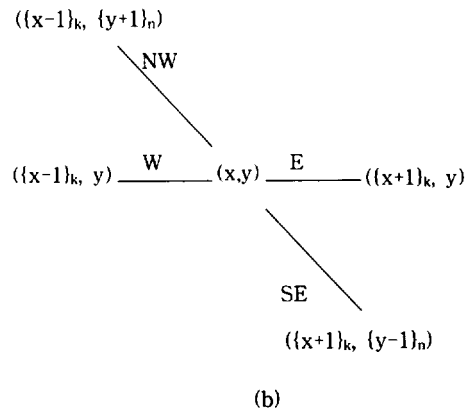
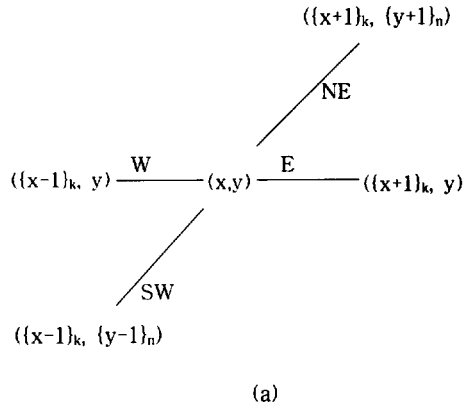
이 구조의 수평방향 링크는 (그림 1)의 토로이드와 동일하나, 토로이드의 수직방향 링크대신 대각선방향으로 교차접속한 것이다.



(그림 1) 6×6 토로이드
(Fig. 1) 6×6 toroid



(그림 2) 6×6 교차메쉬
(Fig. 2) 6×6 XMesh



(그림 3) (a) $x+y$ 가 짝수인 경우와 (b) $x+y$ 가 홀수인 경우에 노드 (x, y) 의 4개의 인접 노드
(Fig. 3) The four neighboring nodes of the node (x, y) when (a) $x+y$ is even and (b) $x+y$ is odd.

2.2 경로설정(Routing)

상호접속망에서 경로설정 알고리즘은 한 메시지를 망내의 임의의 한 발신 노드로부터 목적지(또는 수신) 노드까지 안내하는 경로를 설정하는 과정으로서, 이 경로설정 목적은 각 메시지의 통신지연이 적은 경로를 선택하는 것이다. 한 노드가 다른 노드로 메시지를 전송함에 있어서 그 메시지의 발신노드에서 수신노드까지의 경로상에 있는 각 노드는 이 전송에 참여하게 된다. 즉, 각 노드는 메시지를 받은 다음 이 메시지를 어느 방향으로 전송할 것인지를 결정하여야 한다.

XMesh 위상구조의 대칭성에 근거하여 이 구조가 갖는 최단 경로를 최대한으로 활용할 수 있는 최적 경로설정 알고리즘을 (그림 4)에 나타내었다. 이 최적 경

로설정 알고리즘에서는 교차메쉬의 각 노드에서 메시지의 상세한 과거이력이나 망 전체의 정보를 알 필요 없이 수신노드의 주소만으로 최단거리로 그리고 독자적으로 경로설정할 수 있도록 고안되었다.

발신노드에서 메시지가 생성될 때 수신 노드의 주소가 메시지 내에 함축된다고 가정하였다. 메시지가 경로상의 중간노드에 도달했을 때, 중간노드의 주소와 수신노드의 주소를 비교하여 다음의 진행방향을 결정한다. 최적 진행방향의 링크가 여러 개가 존재할 경우 임의로 한 링크를 선택하도록 하였다. 만약 메시지의 경쟁이나 링크결합 등의 이유로 해당 링크에서 블럭킹(blocking)이 발생하는 경우에는 다른 최적링크 중 하나를 선택하게 되며, 다른 최단 경로가 없다면 우회하는 링크를 선택한다. 다음의 중간노드에서는 이 과정을 반복함으로써 경로설정을 하여, 결국 수신노드까지 메시지를 전송하게 된다.

(그림 4)에서는 $n \times k$ 교차메쉬의 발신노드 (x_1, y_1) 에서 수신노드 (x_2, y_2) 로 메시지를 전송함에 있어서 최적 진행방향의 링크를 설정하기 위한 경로설정 알고리즘을 기술하였다. 대칭성을 가지는 상호접속망의 임의의 노드 (x_1, y_1) 에서 노드 (x_2, y_2) 로의 경로설정 은 노드 $(0, 0)$ 에서 노드 $(\{x_2-x_1\}_k, \{y_2-y_1\}_n)$ 로의 경로설정과 동일하다. 따라서 메시지의 진행방향을 결정하기 위해 먼저 현재 노드와 수신노드의 주소를 비교하여 상대주소 $x = \{x_2-x_1\}_k, y = \{y_2-y_1\}_n$ 및 진행방향 결정상수 z 를 계산한다. 이 때 교차메쉬구조에 대칭성을 부여하기 위하여 x_1+y_1 이 홀수일 경우 y 대신 $(-y)_n$ 으로 대체함으로써 모든 노드에서 동일한 경로설정알고리즘을 사용할 수 있게 하였다. 계산된 x, y 및 z 의 값에 따라 최적 진행방향을 결정한다. 이 때 (그림 4)에서 최적방향의 기호표기는 x_1+y_1 이 짝수일 경우를 기준한 것이며, 만약 x_1+y_1 이 홀수일 경우에는 최적방향 중 NE방향 및 SW방향은 SE방향 및 NW방향으로 각각 해석하도록 하였다. 예를 들어 노드 $(0, 0)$ 에서 노드 $(-1, 1)$ 로 메시지를 전송하기 위한 방향을 조사하면, $x = -1$ 및 $y = 1$ 이며 $x*y < 0$ 이므로 $z = x+y = 0$ 이다. 따라서 $x < 0, y > 0$ 이며 $z = 0$ 이므로 (그림 4)에서 나타난 바와 같이 NE, W 및 E방향 등 세 개의 최적방향을 갖는다.

제안된 알고리즘에서 보듯이 교차메쉬의 경우 경로 설정시 1~3개의 최적방향을 가지며 이는 (그림 1)에서의 토로이드의 경우 1~2개에 비해 달리 택할 수 있는 최적경로의 수가 많음을 알 수 있다. 이는 특정 경

로상에 경쟁이 생겼을 때 또 다른 최단경로가 존재함을 의미하며 각 노드별로 링크자원이 접속도에 의해 제한되므로 같은 상황에서 달리 택할 최단경로가 많은 것은 체증으로 인한 통신지연이 적음을 의미한다.

```

/* Routing between (x1, y1) and (x2, y2) in an XMesh */
/* with N=n×k nodes(n, k are even). */
Step 1: Evaluate  $x = \{x_2-x_1\}_k$  and  $y = \{y_2-y_1\}_n$ .
        If  $x_1+y_1$  is odd, replace  $y$  with  $(-y)_n$ .
Step 2: Calculate the direction decision parameter  $z$ .
        
$$z = \begin{cases} x-y, & \text{if } x*y > 0 \\ x+y, & \text{if } x*y \leq 0 \end{cases}$$

Step 3: Determine the optimal direction.
        If  $x_1+y_1$  is odd, NEs and SWs in step 3
        must be replaced with SEs and NWs, respectively.
1) When  $x > 0$  and  $y > 0$ ,
   if  $(z \geq 2)$ , take NE or E direction.
   elseif  $(|z| \leq 1)$ , take NE direction.
   elseif  $(z \leq -2)$  and  $(z = \text{even})$ , take NE, W, or E direction.
   elseif  $(z \leq -2)$  and  $(z = \text{odd})$ , take NE direction.
2) When  $x < 0$  and  $y > 0$ ,
   if  $(z \geq 3)$  and  $(z = \text{odd})$ , take NE direction.
   elseif  $(z \geq 0)$  and  $(z = \text{even})$ , take NE, W, or E direction.
   elseif  $(z = -1)$ , take NE or E direction.
   elseif  $(-2 \leq z \leq -1)$ , take W direction.
   elseif  $(z \leq -3)$ , take W or SW direction.
3) When  $x < 0$  and  $y < 0$ ,
   if  $(z \geq 3)$  and  $(z = \text{odd})$ , take SW direction.
   elseif  $(z \geq 2)$  and  $(z = \text{even})$ , take SW, E, or W direction.
   elseif  $(|z| \leq 1)$ , take SW direction.
   elseif  $(z \leq -2)$ , take SW or W direction.
4) When  $x > 0$  and  $y < 0$ ,
   if  $(z \geq 3)$ , take E or NE direction.
   elseif  $(1 \leq z \leq 2)$ , take E direction.
   elseif  $(z = -1)$ , take SW or W direction.
   elseif  $(z \leq 0)$  and  $(z = \text{even})$ , take SW, E, or W direction.
   elseif  $(z \leq -3)$  and  $(z = \text{odd})$ , take SW direction.
5) When  $x \neq 0$  and  $y = 0$ ,
   if  $(z \geq 3)$ , take E or NE direction.
   elseif  $(1 \leq z \leq 2)$ , take E direction.
   elseif  $(-2 \leq z \leq -1)$ , take W direction.
   elseif  $(z \leq -3)$ , take W or SW direction.
6) When  $x = 0$  and  $y \neq 0$ ,
   if  $(z > 1)$  and  $(z = \text{odd})$ , take NE direction.
   elseif  $(z > 1)$  and  $(z = \text{even})$ , take NE, W, or E direction.
   elseif  $(z = 1)$ , take NE or E direction.
   elseif  $(z = -1)$ , take SW or W direction.
   elseif  $(z < -1)$  and  $(z = \text{odd})$ , take SW direction.
   elseif  $(z < -1)$  and  $(z = \text{even})$ , take SW, E, or W direction.
    
```

(그림 4) $n \times k$ 교차메쉬의 경로설정 알고리즘
(Fig. 4) Routing algorithm in an XMesh with $N = n \times k$ nodes.

3. 성능평가

3.1 노드간 거리

네트워크의 노드간 최대거리(diameter)란 임의의 두 노드간 거리(number of hops)중 최대값을 말한다. 여기서 노드간 거리란 한 노드에서 다른 노드로 가는 여러 경로중 최단거리를 의미한다. 반면, 평균노드간 거리(mean internode distance)는 한 메시지가 목적지까지 도달하기 위해 가로질러야 할 평균 링크의 수이며, 보통 메시지의 평균 지연시간 측면에서는 노드간 최대거리보다 더 우수한 척도라 할 수 있다.

각 노드는 균등한 메시지 routing분포(uniform message routing distribution)를 갖는다고, 즉, 임의의 노드 i 에서 노드 j 로 메시지를 보내는 확률이 모든 $ij(i \neq j)$ 및 $ij \in V(G)$ 에 대하여 동일하다고 가정한다. 이와 같이 균등메시지분포를 갖는 대칭망의 임의의 한 노드에서 거리가 정확히 k 인 노드의 수를 $N(k)$ 그리고 최대거리를 D_m 이라 정의하면, 평균노드간거리 d 는 다음 식으로 주어진다.

$$d = \frac{\sum_{k=1}^{D_m} k \cdot N(k)}{N-1} \quad (2)$$

$$= \frac{4w^3 + 9w^2 + 2w - 24}{12(w^2 - 1)} \quad (3)$$

여기서 $D_m = (w/2 + 2)$ 이다. 동일한 방법으로 정방형 ($w \times w$)일 경우의 교차메쉬 및 토로이드에 대해 노드간 최대거리 및 노드간 평균거리 등을 구하여 <표 1>에 나타내었다.

<표 1> 링크접속도, 노드간 최대거리 및 평균거리의 비교
<Table 1> Comparison of the degree, the diameter and the mean internode distance.

	교차메쉬	토로이드
노드수, N	w^2	w^2
접속도	4	4
직경, D_m	$\frac{w}{2} + 2$	w
평균노드간거리, d	$\frac{4w^3 + 9w^2 + 2w - 24}{12(w^2 - 1)}$	$\frac{w^3}{2(w^2 - 1)}$

<표 1>에서 보듯이 교차메쉬는 노드의 수가 16000 개 정도가 되면 노드간 최대거리 및 평균노드간거리에 있어서 토로이드의 경우보다 각각 약 0.52배 및 0.68배에 불과하다. 네트워크에서 두 노드간의 전송시간은 이 값들과 비례하는 경향이 있으므로 토로이드에 비해 교차메쉬구조가 통신속도면에서 더 우수함을 입증한다. 이는 토로이드에서의 대각방향의 노드사이의 큰 통신지연을 줄이기 위해, 교차메쉬에서 대각방향의 링크를 도입한데서 기인한다.

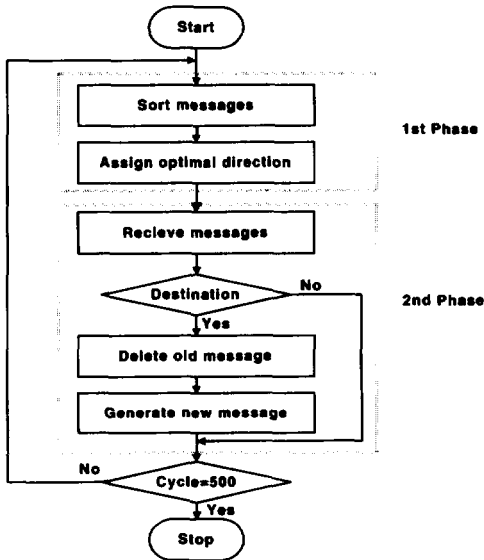
3.2 시뮬레이션 및 결과

본 장에서는 메시지의 경쟁이 있는 실제적 상황에서 상호접속망의 성능을 평가하기 위하여, 제2장에서 기술된 최적 경로설정 알고리즘 및 Barnes가 제안한 최적 경로설정 알고리즘을 이용하여 컴퓨터 시뮬레이션을 하여, 각 최적조건에 따른 교차메쉬의 최대지연, 평균지연 및 메시지처리율을 구하고 이를 비교 고찰한다.

병렬컴퓨터의 한 노드에서 계산(처리)결과 발생된 메시지는 몇 개의 통신 링크와 경유지의 노드를 지나서 수신노드에 도달하게 된다. 이 메시지는 수신노드에서 필요한 계산에 사용되며 그후 시스템에서 제거된다. 이때 수신노드에서 계산이 끝나면 새로운 메시지가 발생되어 다음 수신노드로 전송된다. 이 때 발생된 메시지가 수신노드까지 도달하기 위하여 지나가는 링크수가 이 메시지의 지연시간이 되며, 수신노드에서 소멸된 메시지들의 지연중 최대값 및 소멸된 메시지들의 평균지연 그리고 각 사이클마다 수신노드에 도달한 평균메시지의 수를 시뮬레이션하여 구한다.

시뮬레이션을 함에 있어 교차메쉬는 양방향의 링크를 가지며 초기 단계에서 각 노드마다 메시지를 한 개씩 할당한다. 각 메시지의 목적지주소를 발생시키기 위해 균등한 메시지분포(uniform message distribution)를 갖는 난수발생기(random number generator)를 사용하였으며 본 시뮬레이션에서는 2단계 스케줄링 알고리즘을 이용하였다.

첫번째 단계에서는 각 노드에 존재하는 모든 메시지들을 정렬하여 그 우선순위에 따라 최적의 출력링크(optimal output link)를 할당한다. 최적의 경로상의 링크에 경쟁이 생길 경우 높은 우선순위를 가진 메시지가 (그림 5)의 경로설정 알고리즘에 따라 먼저 최적의 링크를 할당받고 낮은 우선순위의 메시지는 다른 최적



(그림 5) 2단계 스케줄링 알고리즘
(Fig. 5) Two Phase scheduling algorithm

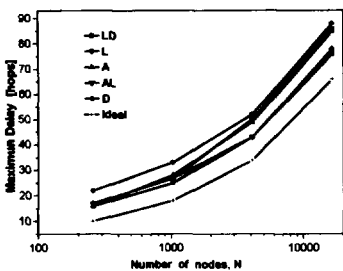
의 경로를 할당받는다. 만약 최적인 경로가 남아 있지 않다면 그 메시지는 다른 모든 메시지가 최적경로를 할당받은 후 남아 있는 경로중 하나를 할당받아 우회하게 된다. 두번째 단계에서는 출력 링크에 할당된 각 메시지를 전송하게 되며 이 메시지를 받은 노드는 메시지의 목적지 주소를 조사하여 목적지노드에 도달한 메시지가 있으면 이를 시스템에서 제거하고 새 메시지를 발생시킨다. 따라서 시스템내의 메시지의 수는 항상 N개로 고정된다.

이 2단계 스케줄링 알고리즘은 구현하기 간단하지만 우회조건(deflection criteria)을 잘못 선정할 경우 방향상태(livelock) 즉, 메시지가 목적지에 도달하지 못

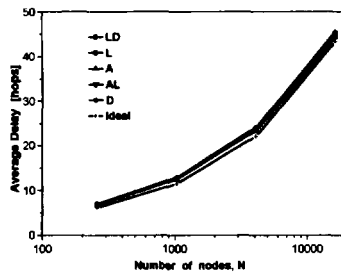
하고 무한히 시스템 내를 돌아다니는 경우가 발생할 수도 있다. 상호접속망의 구조적 특징을 최대한 활용하기 위해서는 우회경로설정 알고리즘에서의 우회조건을 적절히 선택하여야 함이 필수적으로 요구된다. 여러 우회조건에 영향을 파악하기 위해 N노드를 가진 교차메쉬에 대하여 다음 6가지의 조건을 적용하여 고찰하였다.

- i) random: 임의의 순서로 경로설정이 되는 경우
- ii) LD(path_num+deflection): 최적 경로의 수가 적은 메시지가 높은 우선순위를 갖는데 만약 같은 조건이라면 우회한 횟수가 많은 메시지가 높은 우선순위를 갖는 경우
- iii) L(path_num): 최적 경로의 수가 적은 메시지가 높은 우선순위를 갖는 경우
- iv) A(age): 오래된 메시지가 높은 우선순위를 갖는 경우
- v) AL(age+path_num): 오래된 메시지가 높은 우선 순위를 갖는데 만약 같은 조건이라면 최적경로의 수가 적은 메시지가 높은 우선순위를 갖는 경우
- vi) D(deflection): 우회한 횟수가 많은 메시지가 높은 우선순위를 갖는 경우

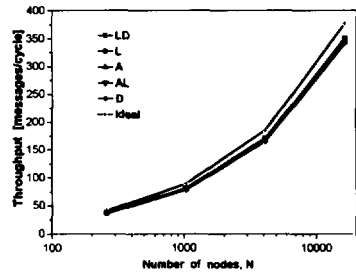
교차메쉬 및 토로이드에 대하여 사이클이 진행됨에 따른 최대지연, 평균지연 및 메시지처리율을 조사하였다. 그 결과 임의경로설정조건(random)의 경우 livelock현상으로 무한지연이 발생하였으며 나머지 5가지 우회조건에 따른 시뮬레이션 결과를 (그림 6), (그림 7) 및 (그림 8)에 나타내었다. 실험의 결과를 검토하기 위해 노드간 최대거리, 평균노드간거리 및 메시지처리율의 이상적인 이론치를 <표 1>에 의한 계산 및 (그림 4)의 알고리즘에 의하여 구하여 그림에 나타내었다.



(그림 6) 최대지연
(Fig. 6) Maximum delay



(그림 7) 평균지연
(Fig. 7) Average delay



(그림 8) 메시지처리율
(Fig. 8) Throughput

지금까지의 연구[5,8]에 오래된 메시지가 높은 우선순위를 갖는 우회조건 A가 많이 쓰여왔으나, 우회조건 A의 경우 메시지 발생 후 초기단계에 많은 우회가 발생할 수 있으며 이 우회로 인한 지연이 전체 메시지 지연에 큰 영향을 미치게 되어 우회한 회수가 많은 메시지에 높은 우선순위를 주는 우회조건 D를 고려하였으며, 또한, 교차메쉬의 특성상 최적경로의 수가 다른 구조들에 비해 다양하다는 특징을 이용하여 최적경로의 수를 우선순위에 반영한 우회조건 LD를 제안하게 되었다. (그림 6)에서 보듯이 우회조건 LD가 우회조건 A에 비해 최대지연, 평균지연 및 메시지처리율에 있어서 각각 이상값에 대한 목표치의 약 47%, 47% 및 21% 개선되었으므로 시스템의 통신시간 측면에서 우수함을 알 수 있다.

교차메쉬는 대각방향으로의 링크를 도입함으로써 통신지연을 줄일 수 있으며 최적경로인 방향의 수도 많아짐은 전술한 바 있다. 메시지를 전송함에 있어 최적경로가 많을 경우 어느 방향의 링크를 택하는 것이 유리한가를 조사하기 위해, 대각방향으로의 메시지 전송확율을 수평방향의 링크에 비해 1배(1X), 4배(4X), 16배(16X) 및 무한대배(Inf) 즉, 무조건 우선 등 5가지 경우에 대해 시뮬레이션하고 그 결과를 <표 2>에 나

타내었다. 또한 검토에 참고하기 위해 노드간 최대거리, 평균노드간거리 및 메시지처리율의 이상적인 이론치(Ideal) 즉, 우회가 전혀 발생하지 않은 가상치를 표에 나타내었다.

최적경로의 수가 많을 경우 대각방향으로의 전송확율을 높일수록 성능이 개선되었으며, 대각방향의 링크로 우선적으로 전송하는 경우(Inf)가 대각방향의 링크와 수평방향의 링크로의 전송을 동일 확률로 하는 경우(1X)에 비해, 최대지연, 평균지연 및 메시지처리율에 있어서 각각 이상값에 대한 개선목표치의 약 20%, 43% 및 13% 개선되었다.

따라서, 메시지 전송시 최적인 경로상의 링크에 경합이 생길 경우 최적 경로의 수가 적은 메시지가 높은 우선순위를 가지며 만약 같은 조건이라면 우회한 횟수가 많은 메시지가 높은 우선순위를 갖는 우회조건 LD를 사용하며, 이 우선순위에 따라 경로를 선정할 차례가 된 메시지가 선택할 수 있는 최적경로의 수가 많을 경우 대각방향의 링크로 우선적으로 전송함이 바람직하다는 사실을 알 수 있다. 이 경우 단순히 우회조건 A에 의한 방법에 비해 최대지연, 평균지연 및 메시지처리율에 있어서 각각 이상값에 대한 개선목표치의 약 58%, 70% 및 31%의 성능개선이 있음을 알 수 있다.

<표 2> 대각선방향의 링크로의 전달확율을 증가함에 따른 최대지연, 평균지연 및 처리율의 비교

<Table 2> Comparison of maximum delay, average delay, and throughput against the increase of the probability of the transmission on the diagonal link.

	노드수 전달확율	256	1024	4096	16384
		최대지연 [hops]	1X 16 4X 16 16X 14 Inf 14 Ideal 10	25 24 24 25 18	43 43 41 41 34
평균지연 [hops]	1X 6.78 4X 6.69 16X 6.64 Inf 6.64 Ideal 6.11	12.47 12.36 12.25 12.18 11.43	23.36 23.26 23.12 23.01 22.09	44.42 44.29 44.13 43.99 43.42	
메시지 처리량 [messages/ cycle]	1X 37.47 4X 38.01 16X 38.29 Inf 38.33 Ideal 41.90	81.03 81.78 82.58 83.02 89.58	170.89 171.71 172.68 173.60 185.42	350.66 351.91 352.95 354.07 377.33	

4. 결 론

상호접속망에서 경로설정 알고리즘은 한 메시지를 망내의 임의의 한 발신 노드로부터 목적지(또는 수신) 노드까지 안내하는 경로를 설정하는 과정으로서, 이 경로설정의 주목적은 각 메시지의 통신지연이 적은 경로를 선택하는 것이다.

본 논문에서는 대규모 병렬컴퓨터의 구현에 적합하게 제안된 새로운 위상구조인 교차메쉬구조의 경우, 균등한 메시지분포를 갖고 메시지의 경쟁이 있는 실제적 상황에서 상호접속망내의 메시지의 경로를 선정하기 위한 방법들을 제안하고 이의 타당성을 검증하기 위해, 제2장에서 기술된 최적 경로설정 알고리즘 및 Barnes가 제안한 우회 경로설정 알고리즘을 이용하여 시뮬레이션을 하였다. 교차메쉬의 특성상 최적경로의 수가 다른 구조들에 비해 다양하다는 특징을 이용하여 최적경로의 수를 우선순위에 반영한 우회조건 및 대각방향의 링크를 효율적으로 활용하기 위한 링크선정방법에 따른 교차메쉬의 최대지연, 평균지연 및 메시지

처리율을 구하고 이를 비교 고찰하였다.

실험 결과, 교차메쉬의 특성상 최적경로의 수가 다른 구조들에 비해 다양하다는 특징을 이용하여 최적경로의 수를 우선순위에 반영한 우회조건 LD가 오래된 메시지가 높은 우선순위를 갖는 우회조건 A에 비해 최대지연, 평균지연 및 메시지처리율에 있어서 각각 이상값에 대한 개선목표치의 약 47%, 47% 및 21% 개선되었다. 최적경로의 수가 많을 경우 대각방향으로의 전송확률률 높일수록 성능이 개선되었으며, 대각방향의 링크로 우선적으로 전송하는 경우가 대각방향의 링크와 수평방향의 링크로의 전송을 동일 확률로 하는 경우에 비해, 최대지연, 평균지연 및 메시지처리율에 있어서 각각 이상값에 대한 개선목표치의 약 20%, 43% 및 13% 개선되었다.

따라서, 메시지 전송시 최적의 경로상의 링크에 결합이 생길 경우 최적 경로의 수가 적은 메시지가 높은 우선순위를 가지며 만약 같은 조건이라면 우회한 횟수가 많은 메시지가 높은 우선순위를 갖는 우회조건 LD를 사용하며, 이 우선순위에 따라 경로를 선정할 차례가 된 메시지가 선택할 수 있는 최적경로의 수가 많을 경우 대각방향의 링크로 우선적으로 전송함이 바람직하다는 사실을 확인하였다. 이 경우 오래된 메시지가 높은 우선순위를 갖는 우회조건 A에 의한 방법에 비해 최대지연, 평균지연 및 메시지처리율에 있어서 각각 이상값에 대한 개선목표치의 약 58%, 70% 및 31%의 성능개선이 있었다. 제안된 경로설정방법은 메시지의 경쟁이 있는 실제적 상황에서 통신지연이 적어 원활한 프로세서간 통신을 해결해 줌으로써 시스템의 성능을 최대한 활용하는데 응용할 수 있으리라 기대된다.

향후 대규모 병렬처리의 대표적인 과제를 선정하여 병렬화, 최적태스크할당 및 시뮬레이션을 통하여 실제 응용에 대한 구체적인 연구검토가 필요하다.

참고 문헌

[1] K.Hwang, 'Advanced Computer Architecture: Parallelism Scalability Programmability', McGraw-Hill, 1993.
 [2] P.Mazumder, "Evaluation of On-Chip Static Interconnection Networks," IEEE Trans. Comp. Vol. C-36, pp.365-369, Mar. 1987.
 [3] Zahorjan, K.C.Sevick, D.L.Eager, and Galler,

"Balanced Job Bound Analysis of Queueing Networks," Commun. Ass. Comput. Mach., Vol.25, pp.134-141, Feb. 1982.

[4] D.A.Reed and H.D.Schwetman, "Cost-performance Bound for Multimicrocomputer Network," IEEE Trans. on Comp., Vol.C-32, No.1, pp.83-95, Jan. 1983.
 [5] K.W.Tang, and S.A.Padubidri, "Diagonal and Toroidal Mesh Networks," IEEE Trans. on Comp., Vol.43, pp.815-826, July 1994.
 [6] P.Barnes, "On Distributed Communication Networks," IEEE Trans. Commun. Syst., Vol.12, pp. 1-9, 1964.
 [7] B.W.Arden and F.Li, "Simulation of Interconnection Networks for Massively Parallel Systems," Tech. Rep., Dept. of Elec. Eng., Univ. of Rochester, Rochester, NY, 1991.
 [8] J.J.Kim and H.M.Choi, "XMESSH Interconnection Network for Massively Parallel Computers," IEEE Proc.-Comput. Digit. Tech., Vol.143, No.6, pp.401-143, Nov. 1996.



김종진

e-mail : kimjj@pine.pknu.ac.kr

1983년 경북대학교(학사)

1985년 한국과학기술원 전기 및 전자공학파(공학석사)

1995년 경북대학교 대학원 전자공학과(공학박사)

1987년~현재 부경대학교 전자공학과 부교수

관심분야 : 컴퓨터구조, 병렬분산처리, 상호접속망 등



윤성대

e-mail : sdyoun@pine.pknu.ac.kr

1980년 경북대학교 졸업(공학사)

1984년 영남대학교 대학원 졸업(공학석사)

1992년 부산대학교 대학원 전자계산학과 박사과정 수료

1986년~현재 부경대학교 전자계산학과 부교수

1992년~1995년 부산공업대학교 전자계산소 소장

관심분야 : 병렬처리, 계산기구조 등