

토러스에서 다중전송을 위한 효율적 워홀 라우팅의 성능 분석

김소은[°] 김창수 윤성대
부경대학교 전자계산학과

A Performance Analysis of Efficient Wormhole Routing for Multicast Communication in Torus Networks

So-Eun Kim[°] Chang-Soo Kim Sung-Dae Youn
Department of Computer Science
Pukyong National University

요 약

본 논문에서는 동일한 목적지 노드들의 집합을 가지는 다중 멀티캐스트(multiple multicast)가 동시에 일어날 경우 노드간의 경쟁을 최소화하고 메시지 전송 경로를 단축시키기 위한 알고리즘을 제시하고, 2차원 네트워크인 토러스 구조에서 제시된 본 논문의 알고리즘을 기존 알고리즘과 비교한다. 기존 논문에서는 멀티캐스트 메시지 패싱 시스템(multicast message-passing system)에서 여러 개의 소스노드로부터 목적지노드가 중첩되는 다중 멀티캐스트를 구현하기 위한 방법으로 2차원 메쉬 구조의 특징을 이용한 SCHL(Source-Centered Hierarchical Leader) 기법[7,8]이 제안되었다. 본 논문에서는 이를 토러스 구조에 적합하게 변형하여, 기존의 메쉬 구조에 적용된 SCHL기법[7]과 성능을 분석함으로써 제안된 알고리즘이 메쉬 구조보다 토러스 구조에서 구현될 때, 더 향상된 성능을 가진다는 것을 보인다. 본 논문에서의 메시지 전송기법은 결정적 워홀 라우팅 방식을 이용한다.

1. 서 론

최근 들어 정보의 교통량이 많아짐에 따라 여러 개의 프로세서들이 상호 연계하여 많은 메시지들을 전달하는 병렬 컴퓨터에서는 하나의 시작 노드로부터 같은 메시지를 임의의 수의

목적지 노드로 전송하는 멀티캐스트망의 중요성이 점점 더해가고 있다. 특히 대규모의 병렬 컴퓨터는 많은 수의 노드로 이루어지기 때문에, 경우에 따라서는 여러 개의 소스 노드로부터 중첩된 목적지 노드들에 대해 동시에 각각의 다중전송이 일어나는 경우가 발생하게 된다. 이때, 효율적인 메시지 패싱을 위해서는 중첩된 목적지 노드들로 인해서 발생하게 될 노드 경쟁을 최소화하고 메시지 전송 경로를 단축시켜 메시지를 신속하게 전달하기 위한 방법이 제시되어야 한다.

다중 컴퓨터에서 사용되는 스위칭 기술로서는 store-and-forward 방식, virtual-cut-through 방식, wormhole 방식이 있다[2]. 본 논문에서 적용된 워홀(wormhole) 라우팅 방식은 각각의 패킷을 flit(flow control digit)이라는 작은 단위로 나누어 첫 번째 flit에는 목적지 노드의 주소를, 그리고 두 번째 flit부터는 메시지를 포함하여 각각의 flit들이 파이프라인 형식으로 전송되는 방식이다. 또한 워홀 라우팅 방식은 가상 채널을 사용함으로써 메시지를 각 노드로 전송할 때 발생할 수 있는 교착 상태를 회피[1]하도록 하였다.

라우팅 알고리즘은 네트워크의 상태에 따라 적응적(adaptive) 라우팅과 결정적(deterministic) 라우팅으로 구분한다[2]. 본 논문에 적용된 결정적 라우팅은 전송 메시지들이 최단(minimal) 경로를 따라 라우팅을 수행한다는 이점이 있는 반면 노드간의 최단 경로 설정에 따른 네트워크 경쟁이 발생하게 되는 단점이 있다. 이러한 단점은 다중 컴퓨터의 성능을 저

하시키는 주요 원인이 된다.

하드웨어적으로 메시지의 다중전송(multi-cast communication)을 허용하는 다중전송 메시지 패싱 시스템은 하나의 소스 노드로부터 단 하나의 multidestination worm을 이용하여 여러 개의 목적지 노드로 한번에 동일한 메시지를 전송할 수 있다.

본 논문은 이러한 특성을 이용하여 토러스 구조를 가진 다중전송 메시지 패싱 시스템에서 다중 멀티캐스트(multiple-multicast)를 수행할 경우 결정적 라우팅의 특성으로 인해 발생하는 네트워크간의 노드 경쟁을 최소화하기 위한 방법으로 멀티캐스트 메시지 패싱 시스템(multicast message-passing system)을 위한 SCHL 기법을 제시하고, 또한 제시한 알고리즘을 수행하여 그 성능을 기존에 제안된 메쉬 구조와 비교 분석한다.

2장에서는 본 논문이 제시하는 알고리즘에 관련된 기존의 SCHL 기법에 대해 그 개선점과 장단점을 살펴보고, 3장에서는 2장에서 설명된 알고리즘의 단점을 수정 보완하여 토러스 구조를 위한 새로운 알고리즘을 제시한다. 그리고 마지막으로 4장에서는 제안된 기법에 대한 성능 평가 후 결론을 맺는다.

II. 관련 연구

이 장에서는 본 논문과 관련하여 멀티캐스트가 제공되는 다중컴퓨터에서 다중 멀티캐스트 발생 시 빈번히 발생하는 노드간의 경쟁을 회피하기 위해 제안된 기존 알고리즘[7,8]을 살펴보고자 한다.

SCHL(Source Centered Hierarchical Leader) 기법은 다중 멀티캐스트 전송시에 노드 경쟁을 최소화하기 위해 소스 노드의 위치를 고려한 그룹 분할과 같은 방향의 목적지 노드들에 대해 그 각각의 리더 노드를 설정함으로써 멀티캐스트를 수행하도록 하는 방법이다.

<정의1> 만일 $\Phi = \langle D_0, D_2, \dots, D_n \rangle$

차원순 체인이고 소스노드 D_i 가 Φ 의 요소이면, $\Phi_1 = \langle D_i, D_{i-1}, \dots, D_n, D_0, D_1, \dots, D_{i-1} \rangle$ 은 R-chain이다[6].

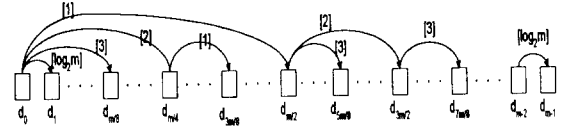


그림 1: 최소시간 멀티캐스트

d_0 를 소스노드, $d_1 \dots d_{m-1}$ 을 목적지 노드라고 가정할 때, 그림 1은 $m-1$ 개의 목적지 노드로 메시지를 전송하는데 있어 노드경쟁을 회피하기 위한 목적지 노드간의 수행 우선 순위를 부여하는 방법이다. 이는 최소시간 멀티캐스트를 수행하기 위한 오일러 디트레일 경로 [4,5]에 기반을 둔 것이며, 기존의 SCHL 기법에서 소스노드가 가진 메시지를 리더노드 L_2 로 전송하기 위해 목적지 노드간의 우선 순위를 설정하는데 사용되는 방법이다.

소스노드와 리더노드 L_2 가 사전 순 체인(Φ)으로 정렬이 되어 있다고 할때, R-chain(Φ_1)은 소스노드(s)가 Φ_1 의 맨 처음 위치에 올 때까지 좌회전(rotate-left) 연산에 의해 시프트 된 노드들의 집합이다. 이와 같은 실행은 Φ_1 을 입력으로 가장 좌측의 노드(D_{left})와 우측의 노드(D_{right})를 비교한 후 최소시간 멀티캐스트를 수행하기 위해 그림 1과 같은 방법으로 체인의 가운데 노드에게 먼저 우선 순위를 부여[6]하여 단계별로 단일전송에 기반한 다중전송(unicast based multicast)를 수행한다.

2차원 메쉬 구조에서 SCHL 기법은 소스노드를 s, 목적지 노드를 D, $L_0 = D \cup \{s\}$ 이라 할 때, L_0 를 소스노드를 중심으로 차원을 따라 수직으로 양분(g_1, g_2)한 후, 그들 각 그룹 내에서 SCHL 기법에 따른 리더노드 L_1 의 집합을 설정하고, 다시 리더노드 L_1 의 집합을 소스노드를 중심으로 수평으로 양분하여 리더노드 L_2 를 설정한다. 그 이후 소스노드로부터 각 리더노드로 메시지를 전송하게 되는데, 우선 정의 1[6]에 의해 각 L_2 리더 노드들과 소스노드의 집합을 배열한 후 그림1과 같은 방법으로 소스노드 s로부터 L_2 리더 노드들에 대한 방문 우선 순위를 부여한 후, XY 라우팅을 수행하여 단일 전송 메시지 패싱에 의해 메시지를 전송한다. 그리고 다시 L_2 리더 노

드들로부터 수직방향으로 L_1 리더 노드들을 향해 다중 전송 메시지 패싱을 수행한다. 이후, L_1 리더 노드들에 의해 나머지 목적지 노드들로의 다중 전송 메시지 패싱을 수행함으로써 SCHL 기법에 의한 전체 목적지 노드로의 메시지 전송을 끝낸다.

이와 같은 방법은 메쉬 구조에서 각 소스노드의 위치에 기반하여 그룹핑을 행하게 되면서 서로 상이한 소스노드의 위치에 의해서 리더노드가 각각 다르게 설정되어 그에 따른 특정 리더 노드에게 집중될 노드 경쟁을 분산시킬 수 있게 되므로 다중 멀티캐스트시의 노드경쟁을 최소화 할 수 있다는 사실에 기반을 둔 것이다.

III. 토러스에서 SCHL 기법

2장에서 설명한 메쉬에서의 SCHL 알고리즘은 소스 노드를 중심으로 2개의 그룹으로 양분할 때, 각 그룹의 크기를 고려하지 않았다. 따라서 만약 소스 노드가 그림 2의 (5,3)과 같이 한쪽으로 치우쳤을 경우 목적지로의 메시지 전송시에 긴 패스 경로를 요구하게 된다. 최악의 경우 $k \times k$ 메쉬에서 최대 목적지로의 메시지 전송거리는 $k-1$ 이 된다.

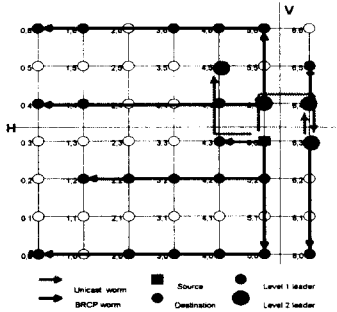


그림 2: 메쉬 네트워크에서 SCHL 기법을 사용한 병행 멀티캐스트

따라서 이 장에서 제시하는 알고리즘은 메쉬에서의 이러한 단점을 보완하고 토러스의 특징인 웨어라운드(wraparound)를 최대한 활용하여 소스노드의 위치를 중심으로 메쉬 구조에서와는 달리 그룹을 비슷한 크기로 양분 (g_1, g_2)하여, 각 그룹의 리더노드를 설정함으로써 다중 멀티캐스트 통신에서 노드 경쟁을 최소화함은 물론 통신경로를 단축시키고자 하였다.

그림 3는 SCHL 기법을 토러스 구조에 적용하여 리더노드를 선택하는 알고리즘이다.

The SCHL scheme

Input: Source $s=(s_1, s_2, \dots, s_n)$, s_i is the coordinate of s in the i th dimension; $\forall_i 1 \leq s_i \leq k_i$, k_i is the radix in the i th dimension.

A grouping in the r th dimension $g=(d_1, d_2, \dots, d_w)$, $d_i(1 \leq i \leq w)$ is the tuple $(t_{i,1}, t_{i,2}, \dots, t_{i,n})$, $t_{i,j}$ is the coordinate of d_i in the j th dimension ($\forall_j 1 \leq t_{i,j} \leq k_j$); k_j is the radix in the j th dimension)

Output: Leader nodes for g such that they can cover all the members of g in a single startup

Procedure:

```

/* Partition g into g1 and g2 */
if  $s_r \leq \lfloor \frac{k_r}{2} \rfloor$  then
   $\forall_i d_i \in g_1$   $t_{i,r} \leq s_r$  or  $t_{i,r} > s_r + \lfloor \frac{k_r}{2} \rfloor$ 
  and  $\forall_i d_i \in g_2$   $s_r < t_{i,r} \leq \lfloor \frac{k_r}{2} \rfloor + s_r$ 
else
   $\forall_i d_i \in g_1$   $s_r < t_{i,r}$  or  $t_{i,r} < s_r - \lfloor \frac{k_r}{2} \rfloor$ 
  and  $\forall_i d_i \in g_2$   $s_r - \lfloor \frac{k_r}{2} \rfloor \leq t_{i,r} \leq s_r$ 
endif
/* Choose the leader node in g1 and g2 */
if  $s_r \leq \lfloor \frac{k_r}{2} \rfloor$  then
  if  $g_1$  is non-empty then
    Choose  $d_i (d_i \in g_1)$  as the leader for  $g_1$ 
    such that  $\forall_i d_i \in g_1$   $t_{i,r} \leq t_{i,r} \leq s_r$ 
  if  $t_{i,r}$  is empty then
     $\forall_i d_i \in g_1$   $t_{i,r} \geq t_{i,r}$ 
  endif
endif
if  $g_2$  is non-empty then
  Choose  $d_i (d_i \in g_2)$  as the leader for  $g_2$ 
  such that  $\forall_i d_i \in g_2$   $t_{i,r} \geq t_{i,r}$ 
endif
else
  if  $g_1$  is non-empty then
    Choose  $d_i (d_i \in g_1)$  as the leader for  $g_1$ 
    such that  $\forall_i d_i \in g_1$   $s_r \leq t_{i,r} \leq t_{i,r}$ 
  if  $t_{i,r}$  is empty then
     $\forall_i d_i \in g_1$   $t_{i,r} \leq t_{i,r}$ 
  endif
endif
if  $g_2$  is non-empty then
  Choose  $d_i (d_i \in g_2)$  as the leader for  $g_2$  such
  that  $\forall_i d_i \in g_2$   $t_{i,r} \leq t_{i,r}$ 
endif
endif

```

그림 3: SCHL 기법에 의한 리더노드 선택 알고리즘

소스노드와 목적지 노드의 집합을 각각 s 와 D 라 할 때, $L_0 = D \cup \{s\}$ 에 대해 각 차원에서 두 개의 그룹으로 그룹핑을 수행하여 리더노드들을 선정한다. 즉, k_r 이 임의 차원의 차수이고, s_r 이 r 번째 차원에서 소스노드 s 의 좌표라고 할 때 s_r 의 위치를 중심으로 소스노드가 포함된 g_1 그룹과 다른 한쪽의 g_2 그룹으로 양분하는데, 그 크기는 기껏해야 r 차원에 있어서 노드수가 ± 1 정도가 되도록 함으로써 최악의 경우 최대 전송 경로가 길어야 $\lfloor k_r/2 \rfloor + 1$ 홉수를 넘지 않도록 하였다. 이때, 토러스 구조에서는 웹어라운드 특성을 이용함으로써 메쉬에서 최악의 경우 $k_r - 1$ 홉수를 가지게 되는 것을 배제할 수 있도록 하였다.

그림 3의 알고리즘에서는 이차원 토러스에서 리더노드 L_1, L_2 를 설정하기 위해 두 개의 레벨을 거치게 되는데, 레벨1에서는 L_0 에 대해 소스노드를 중심으로 수직 양분 (g_1, g_2)하여 레벨1의 리더노드의 집합 L_1 을 소스노드와 근접한 노드들로 설정하고, 레벨2에서는 L_1 에 대해 다시 수평으로 양분 (g_1, g_2)해서 소스노드와 근접한 노드들로 레벨2의 리더노드 집합 L_2 를 설정한다. 이 과정에서 L_2 는 소스노드와 근접한 위치의 노드들로 구성되는 것을 볼 수 있다. 이와 같은 그룹핑을 통해 리더노드들의 집합 L_1, L_2 를 설정 후 HL(Hierarchical Leader)기법[3]과 동일하게 소스노드 s 는 정의1에 의한 SPUMesh 알고리즘에 의해 L_2 노드들로 단일 메시지 패싱 방식으로 메시지를 전송하고, L_2 노드들은 다시 L_1 노드들로, L_1 노드들은 나머지 L_0 노드들로 다중 메시지 패싱을 행함으로써 모든 목적지 노드들의 메시지 전송을 완료하게 된다.

그림 4는 토러스 네트워크상에 각각 소스노드가 (3,4)와 (5,3)인 두 개의 멀티캐스트를 실행할 때, 그 각각의 수행 결과를 나타낸 것이다. 위의 수행 결과로부터 메쉬구조와 마찬가지로 노드경쟁이 일어나지 않을 뿐만 아니라 (5,3)을 소스노드로 할 때, 메쉬구조에서 SCHL 기법을 이용하여 메시지를 전송(그림2)한 경우 패스 길이가 35인데 반해 토러스에서 웹어라운드를 사용하여 SCHL 기법으로

메시지를 전송(그림4-(b))한 경우 패스 길이가 23으로 웹어라운드를 사용한 토러스 구조가 메쉬 구조와 비교하여 패스길이가 짧아졌음을 볼 수 있다.

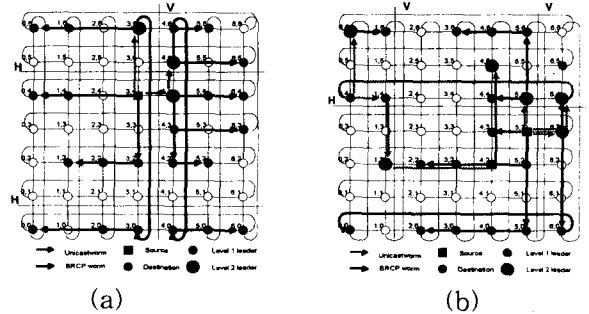
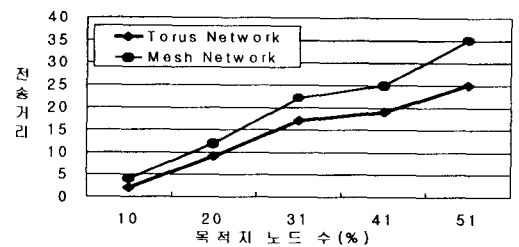


그림 4: 토러스 네트워크에서 SCHL 기법을 사용한 병행 멀티캐스트 사이의 노드 경쟁

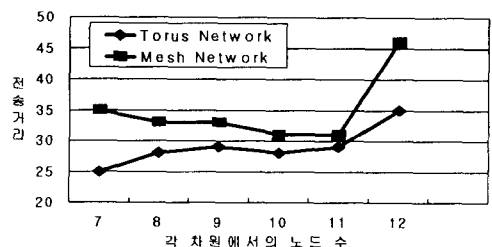
IV. 성능 평가

본 장에서는 SCHL기법을 메쉬 구조와 토러스 구조에 적용시켰을 때, 목적지 노드 수와 일반 노드 수를 증가시킴으로써 그 실행 결과에 따른 메시지 전송길이를 비교하여 토러스 구조가 메쉬 구조보다 더 나은 성능을 가짐을 보이고자 한다.

그림 5는 메쉬 구조와 토러스 구조에서 목적지 노드의 수와 각 차원에서의 노드 수를 늘임으로써 전체 노드의 수를 각각 점차적으로 증가시켰을 때, 메시지의 전송 길이를 비교한 것이다.



(a) 목적지 노드 수와 전송거리



(b) 각 차원에서의 노드 수와 전송거리

그림 5: 메쉬와 토러스 구조에서의 성능 비교

그림 5-(a)는 7×7 메쉬 구조와 7×7 토러스 구조에서 각각의 목적지 노드수를 증가시켰을 때, 메시지가 전송되는 거리를 비교한 것으로 각각의 목적지 노드 수가 증가함에 따라 메쉬와 토러스 구조간 전송 거리의 차도 따라서 증가함을 볼 수 있다. 그림 5-(b)는 메쉬와 토러스 구조에서 각각의 전체 네트워크의 노드 수를 증가시켰을 때, 메시지가 전송되는 거리를 비교한 것으로 각 구조가 같은 수의 노드를 가질 때, 토러스 구조가 메쉬 구조에 비해 전송 거리가 짧음을 볼 수 있다.

V. 결 과

기존의 알고리즘은 멀티캐스트 메시지 패싱 시스템의 멀티캐스트 특성을 이용한 SCHL 기법을 메쉬구조에 적용함으로써 목적지 노드간의 노드 경쟁을 최소화 할 수 있었지만 소스노드의 위치에 따라 최악의 경우, 메시지 전송 경로가 $k-1$ 이 되는 경우가 발생했다.

본 논문은 이러한 메쉬에서의 단점을 보완하기 위해 SCHL 기법을 웨어라운드를 가진 토러스 네트워크 상에 적용하여 r 차원에 있어서 두 그룹간의 노드수의 차가 많아야 ± 1 정도가 되도록 그룹핑함으로써 최악의 경우에 있어서도 최대 전송 경로가 $\lfloor k_r/2 \rfloor + 1$ 이 되어 다중 멀티캐스트 통신을 수행하는데 있어 발생하는 노드 경쟁을 최소화함은 물론 메시지 전송경로를 단축시킨 라우팅 메카니즘의 구체적인 수행 능력을 보였다.

참고문헌

- [1] W.J. Dally and C.L. Seitz, "The torus routing chip" J. Distrib. Computing, vol. 1, no. 3, pages 187-196, 1986.
- [2] L.M. Ni, P.K. McKinley, "A Survey of Wormhole Routing Techniques in Direct Networks" IEEE Computer, Pages 62-76, Feb. 1993.
- [3] D.K. Panda, S. Singal, and P. Prabhakaran. "Multidestination Message Passing Mechanism Conforming to Base wormhole Routing Scheme" In Proceedings of the parallel Computer Routing and Communication Workshop, pages 131-145, 1994.
- [4] P.K. McKinley, H. Xu., A.H. Esfahanian, L.M. Ni, "Unicast-based Multicast Communication in Wormhole Routed Networks" IEEE Transactions on Parallel and Distributed Systems, vol. 5, no. 12, pages 1029-1042, Dec. 1994.
- [5] L. De Coster, N. Dewulf, C.T. Ho, "Efficient Multipacket Multicast Algorithms on Meshes with Wormhole and Dimension Ordered Routing" In International Conference on Parallel Processing, pages III:137-141, Aug 1995.
- [6] D.F. Robinson, P.K. McKinley, B.H.C. Cheng, "Optical Multicast Communication in Wormhole-Routed Torus Networks", IEEE Trans. Parallel Distrib. Syst. vol.6, no. 10, Oct. 1995.
- [7] R. Kesavan, D.K. Panda, "Minimizing Node Contention in Multiple Multicast on Wormhole K-ary n-cube Networks" Technical Report OSU-CISRC-4/9-TR26, TheOhio State University, 1996.
- [8] R. Kesavan, D.K. Panda, "Minimizing Node Contention in Multiple Multicast on Wormhole K-ary n-cube Networks" In International Conference on Parallel Processing, Aug 1996.
- [9] 김소은, 김창수, "이차원 토러스에서 다중 전송을 위한 효율적 워홀 라우팅", 한국통신학회 하계 학술발표논문집, 1997.