2000년 한국정보처리학회 추계 학술논문발표집 제7권 제2호

STC104망을 위한 멀티캐스팅 알고리즘

이효종*, 정우찬*
*전북대학교 전자공학과
e-mail:hlee(wcjeong)@sel.chonbuk.ac.kr

Multicasting Algorithm for the STC104 Network

Hyo Jong Lee*, Woo-Chan Jeong*
*Dept. of Electronic Engineering, ChonBuk National University

요 약

STC104 라우터칩은 점대점 방식으로 자료를 전송하도록 설계되어 있어서 하드웨어 기반의 멀티캐스팅을 할 수 없다. 멀티캐스팅을 하기 위해서는 각 노드간의 관계를 파악하여 동시에 데이터를 전송하는 알고리즘을 개발하여 소프트웨어적인 방법을 이용해야 한다. 본 논문에서는 멀티캐스팅 트리를 이용하여 STC104 라우터칩으로 이루어진 그물 구조 망에서 멀티캐스팅 전송 알고리즘을 제시하고 그 성능을 실험하였다. 멀티캐스팅 트리를 이용한 전송 성능은 전체 hop수가 적을수록 그리고 발원 노드가 전송할 때 적절히 분산시켜 전송할수록 개선된 성능이 나타났다.

1. 서론

병렬시스템의 성능은 시스템에 사용되는 프로세서의 성능과 수에 따라 많이 좌우되나 병렬성이 강한 응용분야에서 무엇보다 중요한 것은 프로세서간의 통신 성능이다. 현재 사용되는 프로세서는 고속으로 작업을 수행한다. 하지만 이에 비해 프로세서간의 통신 속도는 상대적으로 빈약하다. 하나의 소스가 여러 개의 목적지로 자료를 전송해야 하는 멀티캐스팅 방식의 통신인 경우 시스템의 성능은 통신량에 따라 심하게 좌우된다.

SGS-Thomson사에서는 STC104라는 분산 제어 능력을 가진 통신 전용의 고속 라우팅 스위치 칩을 개발하였다. 이 칩은 32개의 양방향(Bidirectional) 링크를 동시에 이용하여 메시지를 송·수신할 수 있으며 각 링크당 전송 속도는 최대 100Mbps까지 될 수 있다. STC104는 32개의 입력포트로부터 하나의 라우팅 지점인 크로스바 스위치로 들어가며 여기에서 크로스바 스위치에 의해 다시 32개의 출력포트로 나간다. DS-Link를 이용하는 각각의 링크는 하나의 입력포트와 출력포트로 구성된다. 여기에 추가적인 버퍼와 라우팅 기능을 수행하는 소자를 포함하여 완전한 하나의 입출력 경로를 제공하는 링크부라는 구조를 이룬다. STC104칩의 응용 분야로는 프로세서 망으로 구성된 병렬 시스템을 비롯하여 멀티미디어 자료의 고속 입출력을 다루는 시스템이나 수요자 요구 비디오(VOD) 시스템 등과 같이 광범위하게 적용될 수 있다.

본 논문에서는 STC104칩으로 이루어진 그물 구조 망을 모델링하여 멀티캐스팅이 이루어질 경우, throughput과 latency를 중심으로 멀티캐스팅 전송량이 다양하게 변화할 때 전송 알고리즘을 시뮬레이션하여 성능을 측정, 비교 분석하였다.

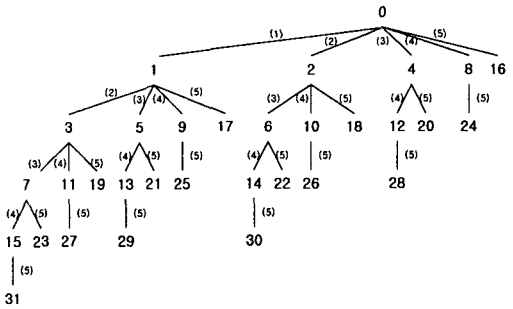
본 논문의 구성으로 제 2절에서는 소프트웨어 멀티캐스팅 기법 중 멀티캐스팅 트리에 관해 기술하였다. 3절에서는 이번 실험에서 사용한 전송 알고리즘에 대해 기술을 하였고 4절에서는 실험 결과를, 5절에서는 결론과 앞으로의 과제를 제시하였다.

2. 멀티캐스팅 트리

멀티캐스팅(multicasting)은 하나의 발원 노드로부터 다수의 확정된 목적지로 데이터를 전송하는 것을 의미한다. 일반적으로 멀티캐스팅의 구현은 하드웨어적인 방법과 소프트웨어적인 방법으로 구분할 수 있다. 하드웨어적인 방법은 네트워크를 구성하는 스위치에서 지원하는 멀티캐스팅 기능을 사용하는 방법이다. 소프트웨어적인 방법은 멀티캐스팅을 원

하는 노드에 대한 소프트웨어 트리를 구성하여 각 노드에 멀티캐스팅 트리의 하나 또는 그 이상의 복사본을 전송하는 방법이다. 그러나, 본 논문에서 사용하는 STC104 라우터칩은 하드웨어적인 멀티캐스팅을 지원하지 않는다. 따라서 STC104 칩을 위한 소프트웨어 멀티캐스팅에 관한 것만 다룬다.

최적 멀티캐스트팅 트리의 구성 방법은 멀티캐스팅이 적용되는 망의 구조와 라우팅 알고리즘에 따라 달라지며 반드시 패킷이 동시에 동일한 링크를 사용하지 않도록 구성되도록 한다.



(그림 1) 멀티캐스팅 트리의 예

(그림 1)은 노드 0이 노드 (1, 2, ..., 31)의 31개의 노드에 멀티캐스팅을 한다고 가정할 경우 멀티캐스팅을 하기 위해 구현한 멀티캐스팅 트리의 예이다. (그림 1)에서 초기 발원 노드인 노드 0이 노드 1에 패킷을 보내면(step 1) 발원 노드는 노드 0과 노드 1 두 개가 된다. 노드 0은 노드 2로 노드 1은 노드 3으로 자료를 전송한다(step 2). 전송 단계 2가 완료되면 발원 노드는 총 4개가 된다. 이와 같이 전송 단계가 진행될수록 발원 노드의 수는 2^n(n: 전송 단계)개가 된다. 위와 같은 방법으로 이후 전송 단계에서 전송 받은 2^n개의 노드는 전송 받지 못한 노드로 자료를 전송하며 멀티캐스팅 트리를 구성한다. 이렇게 하여 목적하는 모든 노드에 대한 멀티캐스팅은 모두 5 단계에서 끝나게 된다.

3. 전송 알고리즘

그물 구조에서 멀티캐스팅 트리를 구성하는 기본 방법은 발원 노드에서 가까운 목적 노드로 계속해서 자료를 전송하는 방법(NEAR)이다. 하지만 이 방법은 트리가 늘어날수록 전체 hop수도 증가하지만 전송시 충돌이 발생된다. 전체 hop수도 줄이고 충돌도 발생되지 않게 전송하는 방법은 발원 노드가 전송할 때 멀리 분산시켜서 전송하는 것이다. 분산 정도가 크면 충돌 발생이 줄어드나 전체 hop수가 증가하여 전체적인 전송 성능이 저하된다. 본 논문에서는 발

1. NEAR :
   All step : MIN[Weight(src, $dest_i$)] i=0, 1, ,...

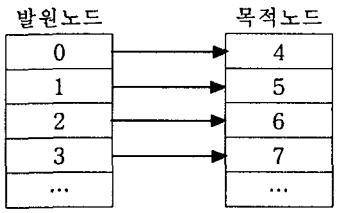
2. FAR :
   step1 : MAX[Weight(src, $dest_i$)] i=0,1, ...
   other step : MIN[Weight(src, $dest_i$)] i=0,1,...

3. FAR FAR :
   step1, 2 : MAX[Weight(src, $dest_i$)] i=0,1, ...
   other step : MIN[Weight(src, $dest_i$)] i=0,1,...

4. FAR MIDDLE :
   step1 : MAX[Weight(src, $dest_i$)] i=0,1, ...
   step2 : MID[Weight(src, $dest_i$)] i=0,1, ...
   other step : MIN[Weight(src, $dest_i$)] i=0,1,...

(그림 2) 멀티캐스팅 전송 방법

원 노드를 분산시키는 정도에 따라 처음 전송 단계에서 발원 노드가 가장 멀리 있는 목적 노드로 자료를 전송하는 방법(FAR), 처음과 두 번째 전송 단계에서 발원 노드가 가장 멀리 있는 목적 노드로 자료를 전송하는 방법(FAR FAR), 처음 전송 단계에서는 가장 멀리 있는 목적 노드로 전송하고 두 번째 전송 단계에서는 처음 단계에서 발원 노드와 목적 노드의 hop의 중간값에 가장 가까운 목적 노드로 전송한 방법(FAR MIDDLE), 세가지 방법과 NEAR 방법으로 전송하여 성능을 비교 분석하였다. 각 네 가지 전송 방법의 특성을 (그림 2)에 요약하였다.

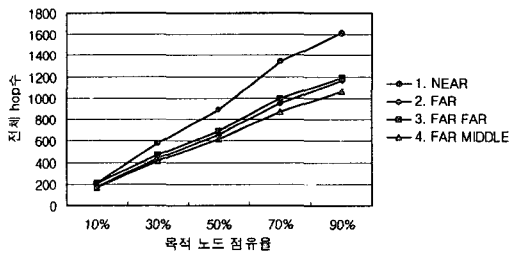


(그림 3) 목적 노드 결정의 예

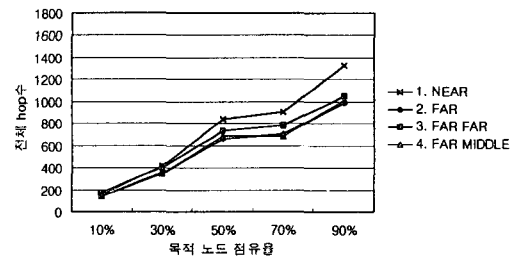
멀티캐스팅 트리는 전송이 진행되면서 발원 노드의 수는 2배씩 증가한다. 멀티캐스팅 트리는 자료를 얻은 순서로 목적 노드를 정하게 된다. (그림 3)에서 노드 0이 목적 노드를 정하고 다음에는 노드 1이 목적 노드를 정하는 순서이다. 하지만 이와 같이 순차적으로 목적 노드를 정하다 보면 노드 0이 정한 목적 노드가 다른 발원 노드에서 전송할 경우 hop수가 더 작을 수 있다. 따라서 최단 거리의 멀티캐스팅 hop수를 구할 수 있는 다른 방법이 고려되야 한다. 두 번째 전송 방법은 이 같은 것을 정정해주는 방법이다. 노드 0이 목적 노드를 결정하였을 경우 결정된 목적 노드를 가지고 다른 발원 노드들과

hop수를 비교한다. 만일 노드 0보다 적은 hop수를 갖는 발원 노드가 있다면 노드 0이 결정한 목적 노드는 보다 적은 hop수를 갖는 발원 노드로 결정된다. 이 방법은 순차적으로 목적 노드를 정하는 방법보다 많은 계산량을 요구하나 프로세서는 고속으로 동작함으로 그 시간차는 극히 미비하다.

(그림 4)와 (그림 5)는 앞에서 설명한 바와 같이 순차적인 방법과 최적 노드 선별 방법에 따라 멀티캐스팅 트리를 구성할 경우의 전체 hop수이다. 그물망의 크기는 16x16이고 각 스위치 당 프로세서는 하나씩 연결된 것으로 가정하였다. 목적 노드의 수는 전체 망의 크기인 256의 10%인 25개, 30%인 76개, 50%인 128개, 70%인 179개, 90%인 230개로 정하였다. 목적 노드는 256보다 작은 값에서 멀티캐스팅 그룹 멤버 수만큼 난수를 발생하였다. 그 값을 위 전송 방법에 적용하여 나온 전체 hop수를 저장, 이를 수 차례 반복하여 나온 값들의 평균값에 가장 가까운 것을 표본으로 삼아 실험하였다.



(그림 4) 순차적 결정 알고리즘의 전체 hop수



(그림 5) 최적 노드 선별 알고리즘의 전체 hop수

(그림 4)와 (그림 5)에서 보는 바와 같이 전체 hop수는 NEAR, FAR FAR, FAR, FAR MIDDLE 순으로 감소하고 있다. 순차적 결정 전송 방법에 의해 얻어진 값에 비해 최적 노드 선별 전송 방법에 의해 얻어진 값이 전반적으로 20% 적은 값이 나왔다.

4. 실험 결과

위에서 얻어진 멀티캐스팅 트리를 가지고 실제 STC104를 모델링한 네트워크를 통하여 자료를 전송하였다. 전체망에서 메시지 발생률은 1Mbytes/sec로 가정하였다. 그물 망의 크기와 목적 노드의 수, 스위치당 프로세서의 수는 위와 동일한 값으로 설정했다. STC104칩은 32개의 양방향 링크를 가지고 있다. 본 논문에서는 32개의 링크를 최대한 사용하는 Multi Link 환경을 적용하였다. 이것은 프로세서가 사용하는 하나의 링크를 제외한 31개의 링크를 스위치의 상, 하, 좌, 우에 같은 수의 링크를 사용하여 스위치을 연결하는 방법이다. 본 논문에서 멀티캐스팅 알고리즘의 성능 측정에 사용되는 특성 변수는 다음과 같다.

(1) 멀티캐스팅(Multicasting) latency :
발원 노드가 메시지의 복사본을 처음 보내기 시작한 때부터 마지막 목적 노드가 메시지를 받을 때까지의 시간

(2) 멀티캐스팅(Multicasting) Throughput :
목적 노드 당 단위 시간동안 수신된 메시지의 양

(3) 멀티캐스팅 블록킹 시간 :
전송시 발생하는 Packet 정체시간의 평균값

기능 개선을 보여주기 위해 개별 전송한 것과 위 알고리즘에 의해 전송한 값을 비교 분석하였다. (표 1)과 (표 2)는 목적 노드 점유율에 따른 멀티캐스팅 latency를 순차적인 방법과 최적 노드 선별 방법에 의한 경우를 비교한 것이다.

단위 : sec	10%	30%	50%	70%	90%
1. NEAR	0.523	0.733	0.837	0.837	0.837
2. FAR	0.523	0.733	0.837	0.837	0.837
3. FAR FAR	0.523	0.733	0.837	0.837	0.837
4. FAR MIDDLE	0.523	0.733	0.837	0.837	0.837
5. 개별 전송	2.617	7.958	13.403	18.743	24.083

(표 1) 순차적 결정 알고리즘 멀티캐스팅 latency

단위 : sec	10%	30%	50%	70%	90%
1. NEAR	0.433	0.628	0.733	0.733	0.785
2. FAR	0.450	0.628	0.733	0.733	0.743
3. FAR FAR	0.478	0.628	0.733	0.733	0.767
4. FAR MIDDLE	0.511	0.628	0.733	0.785	0.824
5. 개별 전송	2.617	7.958	13.403	18.743	24.083

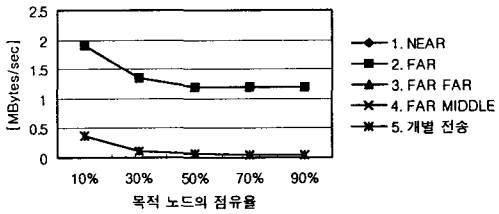
(표 2) 최적 노드 선별 알고리즘 멀티캐스팅 latency

(표 1)과 (표 2)에서 보여지는 것과 같이 개별 전송한 것과 멀티캐스팅 트리를 이용한 전송과는 커다란 차이를 보이고 있다. 멀티캐스팅 트리를 이용하는 경우 성능에 영향을 끼치는 중요한 요소 중의 하나는 전송 단계의 수(S)이다.

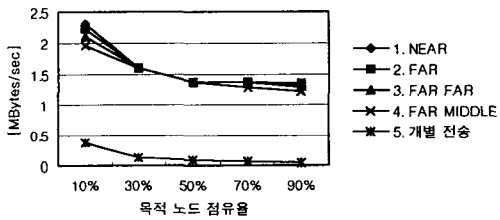
$S = \log_2(n)$ n: 목적 노드의 수 … (1)

(식 1)에 값을 적용하여 보면 목적 노드 점유율이 10%인 경우엔 5단계, 30%는 7단계, 그리고 50%와 70%, 90%인 경우에는 8단계를 갖는다. 50%와 70%, 90%인 경우는 같은 수의 단계를 갖음으로 거의 같은 성능이 나왔다.

(그림 6)과 (그림 7)는 목적 노드 점유율에 따른 멀티캐스팅 throughput을 비교 분석한 것이다. 개별 전송과 비교해서 많은 차이를 보이고 있다.



(그림 6) 순차적 결정 알고리즘 multicast throughput

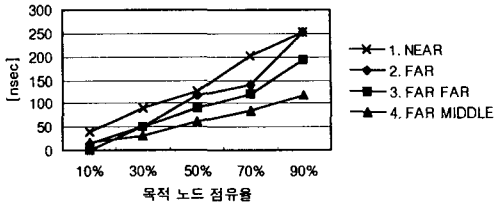


(그림 7) 최적 노드 선별 알고리즘 multicast throughput

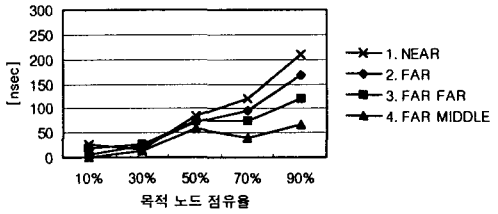
(그림 8)과 (그림 9)은 목적 노드 점유율에 따른 전송시 발생한 블록킹 시간을 비교 분석한 것이다. 3절에서 보여준 전체 hop수의 그림과 비교하면 가장 적은 전체 hop수를 가진 FAR MIDDLE 전송방법이 순차적 결정 알고리즘과 최적 노드 선별 알고리즘 모두 가장 적은 블록킹 시간을 보여주고 있다. 그리고 순차적 결정 알고리즘보다는 최적 노드 선별 알고리즘에서 전반적으로 더 적은 블록킹 시간을 나타내었다.

5. 결론

본 논문은 여러 전송 방법에 따른 그물 망에서의 멀티캐스팅 성능을 측정하였다. 멀티캐스팅 latency는 전송 방법보다는 목적 노드의 수의 영향을 많이 받았다. 전체 hop수가 적은 경우에도 성능 향상은 나타났다. FAR FAR 전송방법은 FAR 전송방법보다 많은 전체 hop수를 가지고 있으나 FAR 전송방



(그림 8) 순차적 결정 알고리즘 멀티캐스팅 블록킹 시간



(그림 9) 최적 노드 선별 알고리즘
멀티캐스팅 블록킹 시간

법보다 전송 노드의 분산이 잘 되어 있어 더 적은 블록킹 시간을 나타내었다. FAR MIDDLE 전송 방법은 모든 면에서 가장 높은 성능을 나타내었다. 최적 노드 선별 전송 방법을 이용할 경우 전송로의 최적화를 통해 모든 면에서 개선된 성능이 나타났다.

본 논문에서는 그물 망에서의 실험만 다루었다. 앞으로는 환원체, 초입방체등 다른 중요한 망구조에서의 특성을 비교하는 연구가 필요하다.

참고문헌

[1] The STC104 *Asynchronous Packet Switch, Engineering Data sheet*, SGS THOMSON Microelectronics, April 1995

[2] 송병열, 이효종 ,"다중 연결 STC104망의 성능 분석", 한국정보과학회 가을 학술발표논문집 Vol.23 No. 2, pp1361-1364 October 1996

[3] J. Robert Jump, *NETSIM Reference Manual Version 1.0*, ECE Dept., Rice University, March 1993

[4] Philip K. McKinley, Hong Xu, Abdol-Hossein, "Unicast-Based Multicast Communication in Wormhole-Routed Networks," *IEEE Transactions on parallel and distributed systems*, Vol.5., NO. 12 pp. 1252-1265, December 1994

[5] Hong Wang, Douglas M. Blough, "Tree-Based Multicast in Workhole-Routed Torus Networks" PDPTA '98 International Conference pp. 702-709