멀티캐스트 스위치를 위한 확장된 B-Tree 복사망

신재구° 손유익
계명대학교 컴퓨터공학전공
shin@tgedu.net yeson@kmucc.kmu.ac.kr

Extended B-Tree(EBT) Copy Network for Multicast Switches

Jae-Gu Shin, Yoo-Ek Son
Department of Computer Eng. Keimyung University

요 약

본 논문은 멀티캐스트 패킷 스위치의 성능 향상에 관하여 언급한다. 네트워크에 요구된 복사본의 수가 네트워크의 크기보다 클 경우 발생되는 오버플로우 문제를 해결하기 위해 Lee의 브로드캐스트 반얀 네트워크(BBN)를 기반으로 하여 다중경로와 다중출력을 제공하는 기능이 추가된 구조를 제안하였으며, 여기에 입력에서 다음 처리해야할 패킷의 fanout 값이 남아있는 BBN의 출력포트 수보다 클 경우 패킷이 복사될 수 없게됨으로서 발생되어질 수 있는 네트워크의 성능이 저하되는 문제를 해결하기 위하여, 셀분할 알고리즘을 이용한 수정된 DAE(dummy address encoder) 방식을 제안하였다.

1. 서론

멀티캐스트는 B-ISDN을 지원하기 위한 스위칭 네트워크의 중요한 특성 중 하나이다. 이런 서비스를 제공하기 위한 멀티캐스트 스위치 구조는 입력부에 들어온 셀을 요구된 복사본 수 만큼 복사하여 출력부로 보내 주는 복사망과 복사망에서 복사되어 나온 셀들을 원하는 목적지로 보내 멀티캐스트 통신을 가능하게 해주는 라우팅망으로 구성 되어있다.

현재까지 연구되어 온 주요 구조는 Huang과 Knauer에 의해 제안된 Starlite System[1]과 Turner에 의해 개발된 Broadcast Packet Switch[2], Lee의 Non-blocking copy network[3] 등이 있다.

Lee의 복사망을 비롯하여 제안된 대부분의 복사망이 안고 있는 문제점은 요구된 복사본의 수가 출력부의 포트 수 보다 많을 경우 발생하는 오버플로우 문제와 함께 큰 fanout을 가진 패킷의 처리 문제로써, 남아 있는 네트워크의 출력부 포트 수보다 fanout이 더 클 경우 이 패킷은 전송할 수 없게되어 네트워크의 성능을 떨어뜨리게 되는 요인이 된다.

이와 같은 문제점을 해결하기 위해 많은 스위치 구조가 제안되었다. Turner[4]는 N개 출력포트보다 크게 확장된 복사망을 사용하여 오버플로우를 해결하려고 하였다. Tagle과 Sharma는 dilated banyan network[5]을 사용였으며, Alimuddin, Alnuweiri과 Donaldson은 Fat-Banyan(FAB) network[6]을 사용하여 오버플로우를 줄이고자 하였다.

본 논문은 Lee가 제안한 복사망의 특성을 기반으로 하여 제안된 EBT 네트워크와 셀 분할 알고리즘을 사용하여 스위치의 성능을 향상시키고자 한다. Lee가 사용했던 BBN을 확장시켜 사용 가능한 출력포트 수와 내부 경로 수를 확장시킴으로서 발생될 수 있는 충돌을 줄였으며 이에 따라 오버플로우 발생 확률을 줄였다. 또한 새로운 셀 분할 알고리즘을 제안하여 주어진 대역폭을 확장시켜 네트워크의 성능을 높이고자 한다.

2. B-tree 네트워크

B-tree 네트워크[7]는 임의의 입력에서 임의의 출력으로 최소의 셀 손실을 갖는 다중경로를 제공하기 위한 다중 연결의 베이스라인 네트워크에 기초한다. 이것은 4개의 입력(I_0, I_1, I_2, I_3)과 2개의 *formal output*(f_0, f_1)과 2개의 *redundant output*(r_0, r_1)을 갖는 4 × 4 SE(스위칭 소자)들로 구성된다.

도착한 셀은 정상적으로는 *formal output*으로 라우팅 되지만, *formal output*에서 충돌이 발생하면 *redundant output*으로 라우팅 되게 된다. B-tree의 각 stage는 $[i,j]$ $(0 \le i \le n-1,\ 0 \le j \le n-1)$로 표현되고 각 stage는 N/2개의 4 × 4 SE로 구성된다. *stage*$[i,j]$에서 각 스위칭 소자는 $SE[i,j,k]$, $(k=d_{n-2}d_{n-3}\ldots d_0)$로 표현된다. B-tree 네트워크에서의 *formal output* 및 *redundant output connection* 알고리즘은 다음과 같다.

formal output connection :

$SE[i,0,d_{n-1}d_{n-2}\ldots d_1]$

$\Rightarrow SE[i+1,0,d_{n-1}d_{n-2}\ldots d_{n-i}dd_{n-i-1}\ldots d_2],\ 0 \le i \le n-2$

redundant output connection :

$SE[i,0,d_{n-1}d_{n-2}\ldots d_1]$

$\Rightarrow SE[i,1,d_{n-1}d_{n-2}\ldots d_{n-(i+j)}dd_{n-(i+j)-1}\ldots d_2],\ 0 \le i \le n-2$

이때, d 값이 1이면 $f_1(r_1)$로 연결되고, 0이면 $f_0(r_0)$로 연결된다. 같은 방법으로

formal output connection :

$SE[i,j,d_{n-1}d_{n-2}\ldots d_1]$

$\Rightarrow SE[i,j+1,d_{n-1}d_{n-2}\ldots d_{n-(i+j)}dd_{n-(i+j)-1}\ldots d_2],\ 1 \le i+j \le n-2$

redundant output connection :

$SE[i,j,d_{n-1}d_{n-2}\ldots d_1]$

$\Rightarrow SE[i+1,j,d_{n-1}d_{n-2}\ldots d_{n-(i+j)}dd_{n-(i+j)-1}\ldots d_2], 1\le i+j\le n-2$

$stage[S_i,S_j]$와 $[T_i,T_j]$ 사이의 상호연결 패턴은 베이스라인 네트워크에서 $stage[S_i+S_j]$와 $[T_i+T_j]$의 상호연결 패턴과 같다($|T_i-S_i|\le 1$, $T_j=S_j$ 또는 $T_i=S_i$, $|T_j-S_j|\le 1$).

이러한 B-tree 스위치 구조는 셀을 출력부에 전송시키기 위해 log_2N개의 stage만을 거치게되므로 전송이 빠르며, 또한 N개의 다중 경로에 의해 내부충돌을 피할 수 있으며, $2(log_2N)$개의 access point를 가지므로 출력충돌을 피할 수 있다. 이로 인해 산출량과 셀 손실율, 셀 지연 성능을 높일 수 있다.

3. 제안된 복사망

Lee의 복사망에서는 복사본 합의 값이 N 이하인 경우 한번만에 패킷이 복사될 수 있지만 이를 초과하는 경우에는 오버플로우가 발생하여 해당 패킷은 폐기된다. 본 논문에서는 요구된 셀을 복사하는 BBN에 다중경로와 다중출력을 제공하는 EBT 네트워크를 제안함으로서 이러한 문제를 해결하고자 한다. 또한 입력에서 다음 처리해야 할 패킷의 fanout 값이 남아 있는 BBN의 출력포트 수보다 클 경우 패킷은 복사될 수 없다. 이를 위해서는 DAE 부분에 셀 분할 알고리즘[8]을 사용하여 네트워크의 성능을 높이고자 한다.

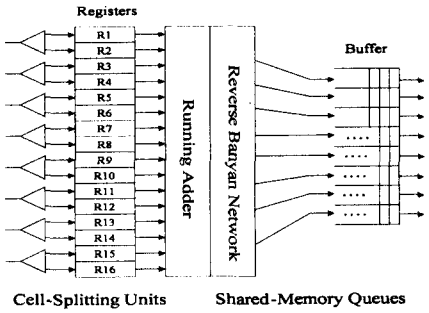
3.1 B-tree 네트워크의 확장

제안된 EBT 네트워크의 구조는 B-tree 네트워크에 4 × 4 SE의 η개 베이스라인 네트워크를 추가하고, $stage[0,0]$의 입력부에 연결된 디멀티플렉서를 통해 입력 셀을 베이스라인 네트워크와 기본 B-tree 네트워크로 분산시켰다. 따라서 BBN에 B-tree 네트워크를 사용할 경우 동시 처리할 수 있는 복사본의 최대수는 N이 되지만 ($N \times N$ 스위치인 경우), EBT 네트워크는 각 입력과 출력 사이에 ηN 개의 다중경로를 가지게 되어 제안된 복사망의 성능과 신뢰성을 η배 증가시키게 된다.

3.2 셀분할 알고리즘

셀분할은 한번에 복사할 수 없는 패킷을 여러 번에 나누어 복사하는 것으로 망의 대역폭을 최대한 활용하게 하여 망의 성능을 높이기 위한 것이다. Lee의 복사망에서 DAE는 dummy address interval ([max,min])과 index reference을 생성하여 새로운 패킷 헤더를 구성하게 된다. 이 경우 단순히 복사본 합에 기초하여 새로운 헤더를 구성하기 때문에 패킷을 복사하기 위한 BBN의 상황을 고려하지 않았다. 따라서 남은 출력 포트보다 큰 fanout을 가진 패킷은 처리할 수 없게 되어 망의 성능을 떨어뜨리게 된다. 이러한 점을 고려하여 Lee의 복사망의 DAE를 그림1과 같이 수정 적용하였다.

수정된 DAE 구조는 입력 셀 분할 단계와 공유 메모리 버퍼 단계로 나뉘어 진다. 입력 셀 분할 단계는 가산기에서 생성된 복사본 합을 기초로 하여 셀 분할이 이루어지며, 이때 각 패킷의 최대 CN(Copy number)은 N이고, 최대 복사본의 합은 N^2이 된다. 그래서 가산기의 비트수는 $L+log_2N$으로 확장된다. 상위 비트 L은 패킷의 DT (Departure time) 스케줄링을 위해 사용되며 하위비트 log_2N은 dummy address interval([max,min]) 생성을 위해 사용된다.



[그림1] 수정된 DAE 구조

CN과 DT의 값에 따라 이루어지는 셀 분할 방법은,

$[DT_i=DT_{i-1}]$ 경우, 셀 분할은 이루어지지 않으며 출력 셀의 형태는 아래와 같다.

R_i	BCN	CN	DT	$Active\ ID(=1)$
R_{i+1}				$Inactive\ ID(=0)$

$[DT_i>DT_{i-1}]$ 경우, 셀 분할이 이루어지고 출력 셀의 형태는 아래와 같으며

R_i	BCN	CN_1	DT_1	$Active\ ID(=1)$
R_{i+1}	BCN	CN_2	DT_2	$Inactive\ ID(=0)$

CN과 DT는 다음과 같이 변경된다.

$CN_{i1} = N - CN_i$, $\quad CN_{i2} = CN_i - CN_{i1}$

$DT_{i1} = DT_{i-1}$, $\quad DT_{i2} = DT_i$

$[DT_i<DT_{i-1}]$ 경우, S_i는 오버플로우 발생으로 셀 손실이 일어난다. 그리고 각 셀의 dummy address interval의 최대값, 최소값은 아래와 같다.

$min_i = CN_{i-1}$, $\quad max_i = CN_i - 1$

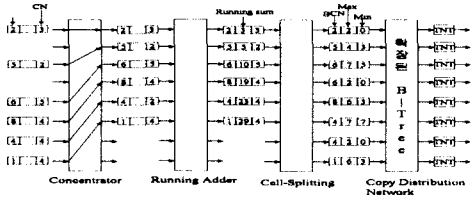
공유 메모리 버퍼링 단계는 $2N \times 2N$ 가산기, 역 반얀 네트워크와 N 인터리빙 버퍼로 구성되어 있으며, 모든 활성 패킷을 선택하여 주소별로 연속적이고 순환적으로 정렬하여 N개의 공유 메모리 버퍼에 저장하게 된다. 저장된 패킷은 출발 시간 스케줄링 알고리즘에 따라 BBN으로 보내져 복사하게 된다.

3.3 제안된 EBT copy network

제안된 복사망은 기본적으로 Lee의 복사망의 특성을 유지하고 있으나 DAE와 BBN에 대한 새로운 적용을 통한 성능 향상을 시도하였다. 제안된 복사망의 구조는 그림2와 같이 4개의 구성요소로 되어있다.

concentrator는 입력된 패킷이 running adder 입력부

에 연속적으로 놓이게 정렬을 해주며, running adder는 입력된 패킷의 CN필드 값에서 복사본 합을 구한다. 이 단계에서 구해진 복사본 합을 기초하여 셀분할 단계에서는 BBN의 크기에 따라 셀분할이 이루어지고 dummy address interval ([max,min])을 할당한다.



[그림2] 제안된 복사망의 구조

셀 분할 단계에서 생성된 DT는 BBN에 보내질 패킷의 그룹을 정하게 된다. 패킷을 복사하기 위한 EBT network에서는 전 단계에서 생성된 DT의 값에 따라 동시에 전송할 그룹을 할당하고, 할당된 그룹은 Boolean interval splitting을 알고리즘을 사용하여 요구된 수량만큼 패킷을 복사하게 된다. 이 단계에서 EBT network의 산출량을 분석해 보면 다음과 같다[7]. 각 SE의 4개의 입력 (I_0, I_1, I_2, I_3)에 걸리는 평균 부하를 $\lambda_{I0},\lambda_{I1},\lambda_{I2},\lambda_{I3}$로 표현하고, formal outputs f_0, f_1과 redundant outputs r_0, r_1에 걸리는 평균 부하를 $\lambda_{f0},\lambda_{f1},\lambda_{r0},\lambda_{r1}$로 표현한다. 각 SE의 출력부하는 입력부하에 따라 유지되며, 4개의 입력이 독립적이라 가정할 때 입력부하에 따른 출력 부하는

$$\lambda_{f0} = \lambda_{f1} = 1-(1-\lambda_{I0}/2)(1-\lambda_{I1}/2)(1-\lambda_{I2}/2)(1-\lambda_{I3}/2)$$

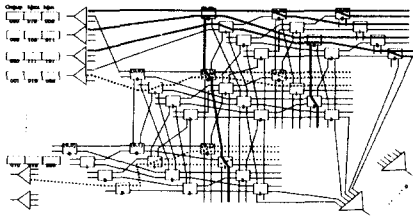
$$\lambda_{r0} = \lambda_{r1} = \lambda_{f0} - (\lambda_{I0}/2)(1-\lambda_{I1}/2)(1-\lambda_{I2}/2)(1-\lambda_{I3}/2) - (1-\lambda_{I0}/2)(\lambda_{I1}/2)(1-\lambda_{I2}/2)(1-\lambda_{I3}/2) - (1-\lambda_{I0}/2)(1-\lambda_{I1}/2)(\lambda_{I2}/2)(1-\lambda_{I3}/2) - (1-\lambda_{I0}/2)(1-\lambda_{I1}/2)(1-\lambda_{I2}/2)(\lambda_{I3}/2)$$

와 같다.

또한, SE의 stage별로 걸리는 트래픽 부하를 보면

stage[0,0], [i,0] : $\lambda_{I0} = \lambda_{I1} = \lambda$, $\lambda_{I2} = \lambda_{I3} = 0$

된다. 왜냐하면, 이 stage의 SE에서 I_2, I_3은 사용하지 않기 때문이다. 이상에서 알 수 있듯이 stage[0,0], [i,0]속의 SE에서는 패킷 손실이 일어나지 않는다. 그 이유는 항상 다음 단계로 전송되기 위한 입력 패킷의 경로가 충분하기 때문이다. 따라서 EBT network에서는 완전부하(λ=1.0)가 주어지도 최대의 산출량을 유지할 수 있다.



[그림3] 그룹화된 복사본의 전송

그림3는 EBT network에서 셀이 복사되는 예를 보여주는 것으로, BBN이 확장비율 $\eta=2$인 EBT network 경우 동시에 전송할 수 있는 그룹이 $(\eta+1)$개 즉, $3N$까지 전송이 가능하다. EBT network은 ηN개의 경로와 $2(1og_2N)$개의 출력부를 제공하기 때문에 산출량, 셀 손실율 및 셀 지연율 면에서 높은 성능을 가진다.

4. 결론

본 논문은 Multicast packet switch의 성능 향상을 위해 Lee의 복사망을 확장한 새로운 스위치 구조를 제안했다. Lee의 복사망이 갖는 오버플로우 문제와 네트워크에 대한 대역폭을 최대로 활용하여 네트워크의 성능을 높이는 문제를 함께 고려하였다. 이를 위해 DAE에 변경한 셀 분할 알고리즘을 적용하여 동시에 전송할 수 없는 패킷을 나누어 전송하게 하였으며, 패킷 분할시 DT필드를 추가하여 동시에 전송할 수 있는 패킷들을 그룹화 하여 공유 메모리 버퍼에 저장하고, 출발 시간 스케줄링에 따라 전송하게 했다.

패킷 복사시 오버플로우 문제를 해결하기 위해서 Lee의 복사망에서 사용했던 BBN대신 확장된 EBT 네트워크를 사용했다. 이것은 ηN개의 다중경로와 $2(1og_2N)$개의 출력부를 제공해 주며, 동시에 전송할 수 있는 패킷 수를 N개에서 $(\eta+1)N$개로 확장시킴으로서 오버플로우를 줄이고 산출량과 셀 지연 및 셀 손실에 대한 성능을 향상시켰다.

5. 참고문헌

[1] A. Huang and S. Knauer, "Starlite : A wideband digital switch," in Proc. IEEE GLOBECOM'84, pp121-125, 1984.

[2] J. S. Turner, "Design of a Broadcast Packet Switching Network," IEEE Trans. Comm., p.734-743, June 1988.

[3] Tony T. LEE, "Nonblocking Copy Networks for Multicast Packet Switching," IEEE Journal, Selected Areas in Communications. Vol.6, No.9, p.1455-1467, December 1988.

[4] J. S. Turner, "A Practical Version of Lee's Multicast Switch Architecture," IEEE Trans. on Comm., Vol.41, No.8, p.1166-1169, August 1993.

[5] Pierre U. Tagle, Neeraj K. Sharma, "Multicast Packet Switch based on Dilated Network," IEEE Global Telecommunications Conf. Vol.2, p.849-853, November 1996.

[6] M. Alimuddin, H. M. Alnuweiri and R. W. Donaldson, "Efficient Multicast Copy Network," Broadband Switching Systems Proceedings, 1997. IEEE BSS '97., 1997 2nd IEEE Int'l Workshop on, pp. 169-172, 1997.

[7] J. J. Li, C. M. Weng, "B-tree : a high-performance fault-tolerant ATM switch," IEEE Proc.-Comm., Vol.141, No.1, pp 20-28, February 1994.

[8] Xinyi Lju and H. T. Mouftah, "A Dynamic Cell -Splitting Copy Network Design for ATM Multicast Switching," Global Telecommunications Conf., 1994. GLOBECOM '94. Comm.: The Global Bridge., IEEE , pp 458 -462 vol.1, 1994.