멀티캐스트 ATM 스위치에서의 공정성 제어 방법

(A Fairness Control Scheme in Multicast ATM Switches)

손 동 욱 † 손 유 익 ††
(Dong-Wuk Son) (Yoo-Ek Son)

요 약 효과적인 멀티캐스트 트래픽 제어를 위하여 다단계 상호연결 네트워크에 기반한 ATM 스위치 구조에 대하여 언급한다. ATM 스위치의 많은 응용분야에서는 점대점 뿐만 아니라 멀티캐스트 연결도 요구되는 것으로, 이것은 하나의 시작지로부터 임의의 다수 목적지로 전달되는 멀티캐스트 연결은 온라인 화상회의, VOD, 분산 데이타 처리 등과 같은 응용분야에 중요하다. 이러한 서비스를 제공하기 위한 멀티캐스트 ATM 스위치 설계 시 고려해야할 사항으로서는 오버플로우 문제, 많은 복사본을 갖는 셀 처리 문제, 그리고 블록킹 문제들 외에도, 공정성 문제와 우선 순위 제어 문제 등이 있다. 특히 들어오는 입력셀들을 골고루 입력포트에 분산시키고자 하는 공정성 문제는 큰 복사본 수를 가진 셀이 상위 입력포트로 들어올 경우 발생된다. 이 경우 running sum을 계산하는 방법에 의해 상위 입력포트가 하위 입력포트 보다 우선적으로 전송됨으로써, 이로 인해 하위 입력포트에 도착하는 셀이 다음 사이클로 전송이 미루어지게 되어 전송 지연 시간이 길어지게 된다는 문제점이 발생된다. 이를 위해 본 논문에서는 셀 분할 및 그룹 분할 알고리즘을 제안하였으며, 또한 입력 패킷의 요구 수에 따른 적절한 복사와 넌블록킹 특성을 기반으로 공정성 제어 방안을 제시한다. 제안된 방법의 성능은 산출량과 셀 손실률, 셀 지연 시간으로 평가하였다.

키워드 : 멀티캐스트 ATM 스위치 , 반얀망, 복사망, 셀 분할

Abstract We present an ATM switch architectures based on the multistage interconnection network(MIN) for the efficient multicast traffic control. Many of these applications require multicast connections as well as point-to-point connections. Multicast connection in which the same message is delivered from a source to arbitrary number of destinations is fundamental in the areas such as teleconferencing, VOD(video on demand), distributed data processing, etc. In designing the multicast ATM switches to support those services, we should consider the fairness(impartiality) and priority control, in addition to the overflow problem, cell processing with large number of copies, and the blocking problem. In particular, the fairness problem which is to distribute the incoming cells to input ports smoothly is occurred when a cell with the large copy number enters upper input port. In this case, the upper input port sends before the lower input port because of the calculating method of running sum, and therefore cell arrived into lower input port is delayed to next cycle to be sent and transmission delay time becomes longer. In this paper, we propose the cell splitting and group splitting algorithm, and also the fairness scheme on the basis of the nonblocking characteristics for issuing appropriate copy number depending on the number of input cell in demand. We evaluate the performance of the proposed schemes in terms of the throughput, cell loss rate and cell delay.

Key words : multicast ATM switch, banyan network, copy network, cell splitting

† 정 회 원 : 혜천대학 컴퓨터통신계열 교수
psalm8@hcc.ac.kr
†† 종신회원 : 계명대학교 컴퓨터전자공학부 교수
yeson@kmu.ac.kr
논문접수 : 2002년 7월 9일
심사완료 : 2002년 11월 4일

1. 서 론

ATM 스위치는 점대점(point-to-point) 연결뿐만 아니라 음성/비디오 회의, 상업방송분배, LAN 브리징, 분산 데이타 처리와 같은 응용 서비스를 제공하기 위한 멀티캐스팅(multicasting) 기능이 요구된다. 이러한 멀티

캐스트 ATM 스위치의 예로는 A. Huang과 S. Knauer에 의해 제안된 Starlite System[1], J. S. Turner에 의해 제안된 Broadcast Packet Switch[2], T. Lee에 의해 제안된 Nonblocking self-routing copy network[3] 등이 있다.

멀티캐스트 전송을 위한 스위치들 중 Lee의 넌블로킹 복사망은 대표적인 멀티캐스트 구조로 인식되고 있다. 이것은 RAN(running adder network), DAE(dummy address encoder), BBN(broadcast banyan network), TNT(trunk number translator) 등 4개의 주요 요소로 구성되어지며, 전체적인 패킷 복사 과정은 인코딩 과정과 디코딩 과정을 거치며 이루어진다. 인코딩 과정은 RAN과 DAE에 의해 이루어지며, 입력 부에 들어오는 패킷의 헤더에 포함되어 있는 복사 요구 수를 새로운 패킷 헤더로 형성하는 monotone address interval의 집합으로 변환한다. 디코딩 과정은 BBN과 TNT에 의해 이루어지며, 패킷의 헤더에 포함되어 있는 monotone address interval에 따라 boolean interval splitting 알고리즘에 의해 패킷 복사를 수행하고 TNT에 의해 최종 목적지가 결정된다[3]. Lee의 스위치가 가지고 있는 문제점으로는 입력포트에서 복사 요구된 복사본 수의 총 개수가 출력포트의 개수를 초과하는 오버플로우와 큰 복사본 수(CN: copy number)의 처리, 그리고 입력의 공정성의 문제이다.

본 논문에서 다루고자하는 입력에 있어서 공정성은, 입력포트에 도착하는 패킷은 running sum을 계산하는 방법에 의해 상위 포트가 하위 포트보다 우선 하므로 상위 포트에서의 셀의 복사본 수가 클 경우, 하위 포트는 계속해서 다음 사이클로 미루어지는 문제로부터 발생된다. 결과적으로 블로킹을 피하기 위해 스위치는 전송에 보다 많은 클럭이 할당되게 된다. 이런 단점은 복사망의 사용 능력을 제한하게 되며, 아울러 hot spot일 경우에도 그와 같은 불공정성이 발생한다. hot spot에 의한 비균일 트래픽은 특정의 출력포트에 대해 동시에 많은 요구가 발생하는 것으로 일정 균일 트래픽에 부과된 접근률 보다 훨씬 높게 단일 출력포트에 집중되어지는 것을 말한다. 그러므로 hot spot에 의한 시스템 지연은 들어오는 셀의 비균일 분산의 결과로서 증가될 수 있다.

이러한 공정성의 문제를 해결하려는 여러 방안들이 제안되었는데, 복사망의 앞 끝에 SSP(shift sequence permutation)을 연결시킴으로서 복사망 접근에서 동일한 기회를 가지게 하거나[4], CDN(cyclic distribution network)을 통해 마스터 셀을 순환적으로 분배함으로써 공정하게 반얀망으로 들어가서 복사하는 방안과[5], 순환적인 방법으로 임의의 입력포트로부터 running sum을 계산하는 CRAN(Cyclic RAN)이 있다[7]. 제안된 방안들은 입력포트를 회전시키거나 순환적으로 입력을 분배하는 방법을 사용하고 있으나, 복사본 수에 대한 고려를 하고 있지 않다. 즉 작은 복사본 수를 가지고 하위 입력포트에 들어오는 셀과 큰 복사본 수를 가지고 상위 입력포트에 셀이 도착할 경우, 작은 복사본 수를 가진 셀 임에도 불구하고 전송 대기 시간은 길어진다.

본 논문은 입력에 대한 공정성 문제를 제어하기 위한 멀티캐스트 스위치를 제안하고자 한다. 제안된 스위치는 입력 셀 분할, 공유 메모리, 그룹 분할로 구성되어있다. 제안된 스위치는 입력 셀에 대한 공정성을 가지며, 큰 복사본의 수가 입력되는 경우에도 좋은 성능을 나타내고 있다.

2. 관련 연구

본 논문의 공정성 문제를 해결하기 위한 관련 연구로는 입력포트에서 순환적으로 복사본 수의 합을 계산하기 위한 CRAN 구조[7]와, Lee 스위치의 셀 분할과, Liu의 동적 셀 분할[8], 그리고 복사본 수의 요구 수가 출력포트의 수보다 클 경우에 셀들이 폐기되는 것을 막기 위해 사용되는 공유 메모리 구조[6]를 들 수 있다.

2.1 순환적 RAN 구조

Lee의 스위치에서의 문제점은 입력포트에서 복사 요구된 복사본 수의 총 개수가 출력포트의 개수를 초과하는 오버플로우 문제인데, 이것은 복사망의 산출량을 떨어뜨리고 입력 멀티캐스트 패킷간의 불공정 문제를 낳는 요인이 된다. 이러한 원인은 RAN의 고정된 구조에서 비롯된다. running sum의 계산이 모든 타임슬롯에서 입력포트 '0'에서 시작을 하며, 낮은 번호를 가진 입력포트가 높은 번호를 가진 입력포트(하위포트)보다 훨씬 높은 우선 순위를 가진다.

이를 위해 Lee의 스위치의 RAN을 수정한 CRAN은 순환적인 방법으로 임의의 입력포트로부터 running sum을 계산한다. 각 타임슬롯마다 running sum을 계산하기 위한 위치는 이전 타임슬롯에서 오버플로우의 발생 여부에 따라 동적으로 결정되어진다. 이러한 결정은 CRAN의 출력에서 만들어지며, 피드백 루프를 통하여 입력포트로 되돌아온다. shifter는 순환적으로 중복되어진 패킷을 이동함에 의해 멀티캐스트 복사본 패킷들은 라우팅망의 입력에 균일하게 분배함으로 출력의 공정성 문제를 해결한다. DAE는 새로운 패킷 헤더를 형성하기 위한 adjacent running sum을 가진다[7]. 그림

1은 CRAN을 갖는 스위치 구조를 보여준다. CRAN의 단점은 입력포트에 들어오는 셀의 복사본 수에 관심을 두고 있지 않으므로 복사본 수가 클 경우, running adder point에서 먼 포트의 전송 대기 시간이 길어지는 점이다.

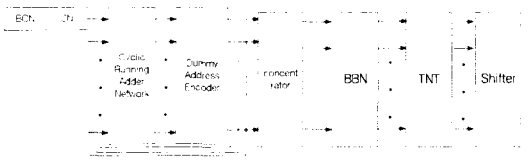

그림 1 CRAN을 포함한 스위치 구조

2.2 셀 분할

일반적으로 멀티캐스트 셀은 한 개의 타임슬롯으로 모든 복사 요구를 수용할 수 없으므로 여러 타임 슬롯을 걸쳐 분할 및 전송되어지는데, 이것을 셀 분할(cell splitting)이라 한다. 또한 팬아웃(fanout)는 복사본 요구 수를 의미하며, 팬아웃 분할 스케줄링은 동일 환경하에 팬아웃 분할을 하지 않는 정책보다 상대적으로 낮은 지연을 가진다. 이는 팬아웃 분할 정책이 훨씬 더 시스템 사용을 좋게 하기 때문이다. 또한 팬아웃 분할을 하지 않고 입력에 대한 순환 서비스를 하는 것과 팬아웃을 분할하지 않고 입력에 대한 순환 서비스를 하지 않는 정책의 경우, 전자의 경우가 더욱 지속적인 산출량을 개선하였음을 제안한 바 있다[9].

또한 일반적인 팬아웃 알고리즘에서 문제점 중의 하나는 큰 팬아웃 대한 블로킹 확률이 매우 크다는 사실이다. 그러므로 큰 팬아웃 요구는 스위칭 시스템을 거의 라우팅 되어지지 못하거나, 네트워크 사용률이 복사 요구에 대한 충분한 경로가 주어질 만큼 적어질 때까지 대기하여야 한다. 큰 팬아웃에 대한 보가 요구는 높은 블로킹 확률을 가지므로 바람직하지 못하다. 이러한 문제점을 해결하기 위해서는 입력 스위치에서 큰 팬아웃에 대한 복사 요구를 수용하기 위한 한 개 이상의 가운데 스위치에 우회 경로를 제공하는 방안을 제안한 바 있다[10].

셀 분할은 HOL 블로킹 문제를 해결하며, 평균 셀 지연을 줄여서 최대 산출량을 가져다준다. Waiting Time (WT)는 복사를 위해 반얀망으로 들어가기 위해 공유 메모리 FIFO 큐에서 대기하기 위해 필요한 각 셀의 타임 슬롯을 가리키며, 각 셀의 헤더에 삽입된다[6]. 셀 분할은 그림 2와 같이 running adder와 셀 분할 유닛으로 구성되며, running adder는 순서에 따라 입력 셀 헤더에 명시된 복사본 수의 running sum S_i를 생성한다. 각 셀에 대한 최대 복사본 수는 N이며, 최대 running sum은 N^2보다 작거나 같다. 이것은 출력당 $2log_2N$ 비트를 의미한다. Lee의 복사망에서는 N과 같거나 적다는 것을 가정하는 running sum은 복사된 셀의 라우팅 제어를 위해 사용한다. N보다 큰 running sum S_i를 가진 입력은 블록되어진다. 그러나 S_i의 가장 상위비트 log_2N은 오버플로우 정보와 오버플로우 범위를 주기 때문에 매우 중요하다. running adder 구조에서의 비트 수는 $(L + log_2 N)$ bit로 확장되어지며, L의 최상위 비트는 WT를 구성하며, 하위비트는 RS를 구성한다. 각 running sum은 셀 분할 유닛에 의해 읽혀진다.

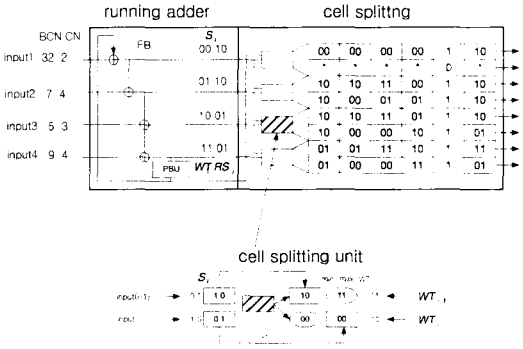

그림 2 셀 분할 구조

각 셀 분할 유닛(CSU)은 WT_{i-1}와 WT_i를 비교하여 그 결과에 따라 다음과 같이 셀 분할을 결정한다.

$[WT_i = WT_{i-1}]$, 셀 분할은 발생하지 않으며, 입력 셀은 레지스터1(R1)로 진행

$[WT_i > WT_{i-1}]$, 셀 분할은 발생하며, 두 개 분할된 셀은 레지스터 1과 2로 진행

$[WT_i < WT_{i-1}]$, 오버플로우가 발생하며, i 번째 셀은 폐기

기타 다음과 같은 추가적인 정보를 생성하여 각 셀 헤더에 삽입된다.

min : 각 출력 집합의 최소 주소로 같은 멀티캐스트 셀로부터 복사본이 항상 연속적인 출력포트에 나타나도록 한다.

max : 위에서 언급한 출력 집합에서 최대 주소

ID : 입력이 active(*1*) 인지 idle(*0*) 인지를 표시

WT : 큐에서 지체하는 셀의 대기 시간

IR : 각 셀에서 *min* 값과 같으며, 각 복사본의 출력 주소를 결정하는 TNT에 의해 사용

ID 값의 결정 규칙은 다음과 같다.
만약 $WT_i \neq WT_{i-1}$, $ID_2 = 0$이면,

$$\begin{cases} ID_1 = 0 & if\ CN = 0 \\ ID_1 = 1 & otherwise \end{cases}$$

만약 $WT_i > WT_{i-1}$, $ID_1 = 1$이면,

$$\begin{cases} ID_2 = 0 & if\ CN_2 = 0 \\ ID_2 = 1 & otherwise \end{cases}$$

running adder의 feedback signal FB는 WTf와 RSf로 구성된다.

$$\begin{cases} WT_f = WT_f - 1 \text{ and } RS_f = RS_N & if\ WT_N > 0 \\ WT_f = 0 \quad \text{and } RS_f = 0 & if\ WT_N = 0 \end{cases}$$

feedback unit(FBU)의 출력은 버퍼에 셀이 없으면 0으로 설정되어진다[8].

2.3 공유 버퍼

공유된 메모리 버퍼링은 그림 3과 같이 $2N \times 2N$ running adder, 역 반얀망, N개의 인터리빙 버퍼의 조합으로 구성되어진다. 이것은 셀을 압축하거나 버퍼링하는 기능을 수행한다. running adder는 ID 내에 명시된 모든 활성 셀을 $2N$ 큐잉 주소를 형성하기 위해 가산한다. 하단의 running sum은 다음 출발 포인트를 계산하기 위해 각 타임슬롯의 끝에서 running adder의 상단의 라인으로 피드백되어 진다. 그러므로 큐잉 주소는 압축적이고 순환적인 연속성을 가진다. 공유된 메모리 버퍼는 역반얀망의 각 출력포트에 위치한 물리적으로 분리된 N개의 메모리 버퍼이다. 역반얀망은 셀을 순환적으로 큐로 라우팅하기 위해 사용되어지며, $2N$개의 입력을 N개 인터리빙 큐로 순환적으로 멀티플렉싱하기 위해 마지막 단계에서 스위칭 요소는 2:1의 시간 분할 멀티플렉스로 수정되어진다[6].

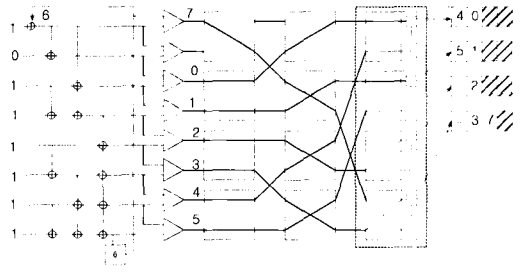

그림 3 공유 메모리

공유 메모리에 대한 멀티캐스팅 구조는 두 가지 형태로써, RAS(replicate-at-send)와 RAR(replicate at receive)가 있다. RAS는 멀티캐스트 셀이 직접적으로 공유메모리에 쓰여지고, 판독될 때 복사되는 반면에 RAR은 공유 메모리에 쓰여지기 전에 멀티캐스트 셀이 복사되어진다. RAS가 메모리에서 단지 한 위치를 점유하기 때문에 셀 손실 성능이 훨씬 더 우수하다[11]. RAS 메커니즘을 사용하는 공유 메모리의 형태로는 고속 스위칭에 의한 메모리의 접근과 횟수에 불합리성에 대한 문제점을 해결하기 위해 다중 병렬 서브메모리 구조를 가진 공통 메모리 형태 스위치가 제안된 바 있다[12].

3. 제안된 스위치 구조

3.1 스위치 구성

고속 멀티캐스트 스위치를 설계하기 위해서는 셀을 복사하는 방법, 셀의 충돌 및 오버플로우 해결 방법, 멀티캐스트 패턴 등을 고려해야 한다. 셀을 복사하는 방법은 복사망을 사용하는 경우와 셀을 라우팅하는 동안 복사하는 방법 등이 있고, 충돌 및 오버플로우 해결 방법은 셀 분할 알고리즘 적용, BBN 구조의 확장 및 병렬 연결 등 스위치 구조에 따라 다양한 방법으로 해결할 수 있다. 또한 원하는 목적지로 셀을 전송하기 위한 멀티캐스트 패턴은 셀의 헤더에 여분의 비트를 사용하는 방법과 멀티캐스트 패턴을 저장한 lookup 테이블을 사용하는 방법 등이 있다.

제안된 스위치는 매우 큰 복사본 수의 처리를 포함한 오버플로우와 블로킹, 그리고 공정성 중점을 두고 있다. 그러므로 제안된 스위치의 구성은 그림 4와 같이 공정성을 문제를 해결하기 위한 입력 셀 분할, 한 사이클 내 전송될 수 있는 셀의 그룹으로 분할 및 전송하기 위한 그룹 분할, 셀 순차와 패킷을 보존하기 위한 공유메모리, 반얀망으로 구성되어진다. 입력 셀 분할은 상하위 포트의 전송 순위를 제어하고 멀티캐스트 셀을 분할하여 전송하게 한다.

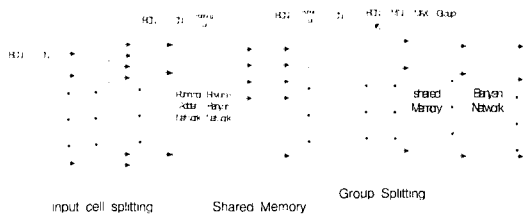

그림 4 스위치 구성

공유메모리는 복사본 수의 요구 수가 출력포트의 수보다 클 경우에 셀들이 폐기되는 것을 막기 위해 사용

된다. 공유메모리가 존재하는 경우에는 버퍼에 저장되어 지지만, 버퍼가 없는 경우에는 runnning sum을 계산한 후 공정하게 입력단에 도착하였더라도 패킷은 복사되지 못하고 폐기되어 차후 사이클에 다시 재 전송된다.

그룹 분할은 한번에 복사될 수 없는 셀을 그룹으로 나누어 복사하는 것으로 망의 대역폭을 최대한 활용하게 하여 망의 성능을 높이기 위한 것으로 반얀망에 들어가는 그룹을 분할하는 기능을 가진다.

3.2 입력 셀 분할

기존의 제안된 셀 분할[8]은 RAN에 의해 생성된 running sum에 의해 각 사이클에 전송되어질 수 있는 셀이 결정되어짐으로 공정성 문제가 제기되어진다. 본 논문에서는 이러한 문제점을 입력 셀 분할에 의해 해결하고자 한다. 입력 셀 분할은 하위 포트에 도착한 셀이 상위 포트에 도착한 셀에 비해 상대적으로 늦은 타임슬롯을 가지는 편중성을 해결하기 위해 복사본 수를 threshold 값으로 분할하여 입력포트의 셀이 적어도 최소 사이클 내에 셀을 전송하는 방안으로, 큰 복사본 수에 대해 작은 복사본 수를 가진 셀이 우선 전송하게 된다는 장점을 가지고 있다.

기존 셀 분할

	t	t+1	t+2	t+3	WT
input0	■■■				0
input1		■■■			t
input2		■	■		t
input3			■		t+1

입력 셀 분할

	t	t+1	t+2	t+3	WT
input0	■■	■	■		0
input1	■■		■		0
input2		■■			t
input3		■			t

그림 5 셀 분할과 전송 타임슬롯

기존의 셀 분할에 비해 제안된 셀 분할의 장점은 작은 대기 시간을 가지며 셀을 복사 전송할 수 있으며, running adder point에서 먼 거리에 있는 포트의 전송 대기 시간이 단축된다는 점이다. CN_i가 입력포트 i에서의 복사본 수라고 하면, $CN_0=4$, $CN_1=3$, $CN_2=2$, $CN_3=2$인 경우, 그림 5의 예를 통해 기존의 셀 분할과 제안된 셀 분할의 전송 대기 시간의 차이점을 보이고 있다. t는 타임슬롯을, WT는 대기 시간을 의미한다. 기존 셀 분할에서는 하위 포트로 내려갈수록 대기 시간이 길어지지만 입력 셀 분할에 의한 대기 시간의 거의 일정하다. 이는 각 포트에 대한 공정성 문제를 해결해준다.

입력 셀 분할은 복사본 수를 알고리즘에 따라 나뉘어져 먼저 전송되어지는 셀과 다음 사이클에 전송되어지는 셀로 나눈다. 먼저 전송되어지는 셀의 표시는 'state' 필드에 의해 표시되어지며, 'state' 필드는 '0', '1'의 값을 가진다. '0'은 분할된 상위 셀을 의미하며, 우선 순위를 가지고 FIFO 큐 먼저 진입한 다음, RAN으로 들어가서 running sum을 계산하게 된다. 'state' 필드의 '1'은 분할된 하위 셀로 FIFO 큐에 저장된 후 우선 순위 셀이 전송되어진 후, 다음 모듈로 진입하게 되어진다. 입력 셀 분할 알고리즘이 그림 6에 나와있다.

```
CN_i = Cell_i.CN
if CN_i > threshold_value  /* cell splitting */
      CN_i0 = threshold_value
      CN_i1 = N - threshold_value
      Cell_i0.CN = CN_i0
      if Cell_i1.state = '0' Cell_i0.state = '1'
      else Cell_i0.state = '0'
      Cell_i1.CN = CN_i1
      if Cell_i0.state = '0' Cell_i1.state = '1'
      else Cell_i1.state = '0'
      Cell_i.state =Cell_i0.state && Cell_i1.state
      if Cell_i0.state = '0'  Send(Cell_i0) to FIFO queue
                              Cell_i1.state = '0'
      if Cell_i1.state = '0'  Send(Cell_i1) to FIFO queue
else
   CN_i0 = CN_i
   Cell_i0.CN = CN_i0
   Send(Cell_i0) to FIFO queue
```

그림 6 입력 셀 분할 알고리즘

$Cell_i$는 i번째 입력포트에 도착하는 셀을 의미하며, 그 셀은 BCN(broadcast channel number), CN, 그리고 'state' 정보를 가지고 있다.

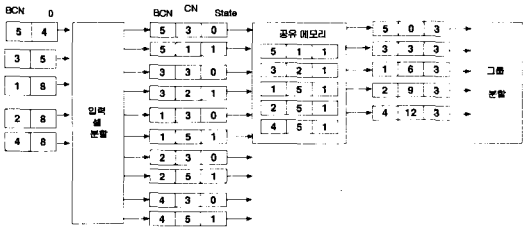

그림 7 입력 셀 분할

CN_{i0}에서 i는 입력포트, '0'은 상위 패킷 및 우선 순위를 가진 복사본 수를 의미한다. *threshold_value*는 복사본 수를 나누기 위한 threshold 값으로 N, $2/N$, $\log_2 N$을 사용한다. 여기서 N은 네트워크의 크기이다. 그림 7은 위 알고리즘에 의한 입력 셀 분할을 보여준다.

3.3 그룹 분할

그룹 분할은 RAN에서 계산된 running sum에 따라 한 사이클 내에 전송되어질 수 있는 셀을 분할하는 기존의 기능에 한 사이클에 전송되어질 수 있는 그룹을 분할하는 기능이 추가되어졌다. 즉, 공유 메모리 다음의 반얀망에 동시에 전송할 수 있는 그룹을 결정하는 기능이다. 그러므로 그룹 분할 알고리즘은 복사본 수를 나눔과 동시에 그룹을 결정하는 기능을 가진다. running sum에 의해 결정되어지는 그룹 번호는 한 사이클에 전송되어지는 그룹을 의미하며, 그룹 번호가 동일할 경우 같은 슬롯타임에 전송되어짐을 의미한다. 그룹 분할 알고리즘은 그림 8과 같다.

```
Gi = RSumi / 2N
if i= 0, Gi =0
        Mini =0
        Maxi = CNi
        return
if Gi = Gi-1   no CN splitting
  //Gi : 입력 포트 i의 그룹 번호
else if Gi > Gi-1 // CN splitting
        CNi0 = 2N - RSumi
        CNi1 = CNi  CNi0
        Gi0   = Gi-1
        Gi1   = Gi
else  overflow , i th cell discard
Mini = Maxi-1 + 1
Maxi = Mini + CNi + 1
```

그림 8 그룹 분할 알고리즘

Gi는 입력포트 i의 그룹번호이며, running sum, 즉 $RSum_i$에 의해 그룹 번호가 결정되어진다. 예를 들면 입력포트 i의 running sum이 6이면 $00\ 0110_{(2)}$으로 표시하여 앞의 두 자리가 그룹 번호(G_i=0)가 되어진다. 위의 알고리즘에 다른 셀 분할 방법이 그림 9에 나와 있다.

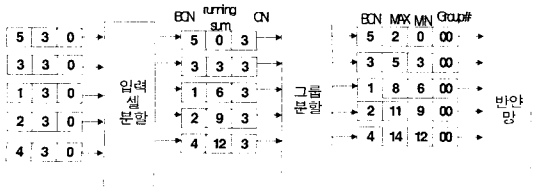

그림 9 그룹 분할

그룹 분할 모듈의 다음에 위치한 공유 메모리는 반얀망으로 들어가는 셀을 저장하는 기능을 가진다. 또한 반얀망에서 복사 가능한 복사본 수 보다 클 경우 최대 복사본 수 보다 큰 복사본 수를 가진 셀은 폐기시킨다. 그림 10은 그룹 분할 알고리즘에 의해 그룹 분할이 발생한 것을 굵은 선으로 보여주고 있다.

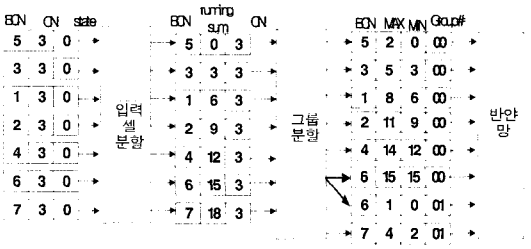

그림 10 그룹 분할의 예

4. Simulation

4.1 평가 요소

시뮬레이션을 수행하기 전에 보다 현실적인 환경에 근접하기 위하여 네트워크 환경과 모델에 대하여 네 가지 가정을 하였다.

ⓐ 네트워크는 동기적으로 운영된다. 즉, 셀은 주어진 시간간격의 시작점에서만 전송되며, 이때 시간은 이산적이다.

ⓑ 모든 입력과 출력 링크의 전송 속도는 동일하다.

ⓒ 각 입력단에 도착하는 셀의 분포는 실제적인 셀의 도착 분포에 근접하기 위하여 negative exponen tial distribution을 사용한다.

ⓓ 각 입력단에 입력되는 트래픽 형태는 균일 트래픽 환경을 가정하였으며, 입력되는 셀의 출력주소는 시뮬레이터가 무작위로 결정한다.

시뮬레이션 수행 시 복사본 수와 버퍼의 크기와 입력 부하를 변수로 설정하고 변수의 변화가 미치는 영향에 따른 성능의 변화를 관찰하였다. 시뮬레이션의 결과의 분석과 비교를 위하여 사용된 각 용어와 성능 평가의 정의는 다음과 같다.

ⓐ 입력 부하(offered traffic load) : 스위치의 각 입력 단에 대하여 매 주기마다 새로운 셀이 도착할 확률

ⓑ 산출량 (throughput) : 단위 시간당 처리할 수 있는 셀의 개수로 본 논문에서는 임의의 제한 시간 내 네트워크의 출력링크를 통과한 셀의 성공률로 정의한다. 산출량(throughput)은

$$\text{throughput, } T = \frac{\text{totaloutputcells}}{\text{totalinputcells}} \quad (\%)$$

ⓒ 셀 손실률 (cell loss rate) : 스위치에 입력된 총 셀의 개수에 대해 출력단으로 출력되지 못하고 손실되는 셀의 개수로서, 셀 손실(cell loss)은

$$\text{cellloss, } C = \frac{\text{total}\nabla\text{etedcells}}{\text{totalinputcells}} \quad (\%)$$

ⓓ 셀 지연(cell delay)은

$$\text{average waiting time in buffer, } D = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{N-1} buf_{ij}}{T}$$

으로 버퍼 속에서 셀이 지연되는 평균 시간(time slot)으로 정의한다. $\sum_{i=0}^{n-1}\sum_{j=0}^{N-1}$buf ij는 각 단계와 입력 포트가 연결되어진 스위치 요소의 버퍼의 총 합을 의미하며, I는 단계 수, j는 입력포트 수를 나타낸다.

이상과 같은 성능 평가의 기준을 분석하기 위해 시뮬레이션 방법을 사용하며, 제안된 셀 분할 스위치와 셀 분할이 없는 스위치를 비교 평가한다. 시뮬레이션 수행 시, 복사본 수와 제공된 입력 부하를 변수로 설정하고 변수의 변화가 미치는 영향에 따른 성능의 변화를 관찰하였다.

4.2 결과 분석

그림 11은 8×8 반얀망에서 제안된 입력 셀 분할을 적용한 모델과, 적용하지 않은 모델인 Lee의 스위치, CRAN 반얀 스위치의 산출량을 보여주고 있다. 셀 분할 모델은 입력 셀 분할의 threshold 값으로 $N/2$, log_2N, 최소 복사본 수, 복사본 수의 평균을 선택하여 실험하였다. 실험된 복사망은 망의 크기가 $N=8$ 인 8×8 반얀망으로 구성되어있고, 복사본 수와 목적지 주소는 임의적으로 선정되어지며, 10ms 속도로 입력포트로 입력되어진다. N/2와 logN의 표기는 고정된 threshold 복사본 수를 각각 $N/2$, log_2N으로 반얀망에 적용한 모델을 의미하며, 'AVE'는 복사본 수의 합을 입력포트의 수로 나눈 평균 복사본 수을 가지고 셀 분할을 적용한 모델을 의미한다. 'Lee'는 Lee의 스위치를, 'CRAN'은 CRAN 반얀 스위치를 뜻한다.

그림에서 log_2N으로 셀을 정적으로 분할하는 것이 복사본 수의 평균으로 셀을 동적으로 분할하는 것보다 우수한 것으로 나와있다. 동적인 분할인 복사본 수의 평균으로 분할하는 것은 부하에 별로 영향을 받지 않고 거의 일정한 산출량을 만들어낸다. 전체적으로 보면 log_2N이 상대적으로 우수하다. 부하가 0.6인 위치까지는 거의 같은 셀 손실율을 기록하지만 그 이상에서는 확연히 구분이 되어지고 있다. Lee의 스위치, CRAN 반얀 스위치 등 셀 분할을 하지 않은 모델은 부하가 커짐에 따라 산출량도 급격히 감소함을 보여준다.

그림 12는 16×16 반얀망에서의 산출량 대한 차트를 보여주고 있다. 그림에서 셀 분할을 하지 않는 모델에서는 부하가 1.0일 때 급격히 산출량이 떨어지고 있음을

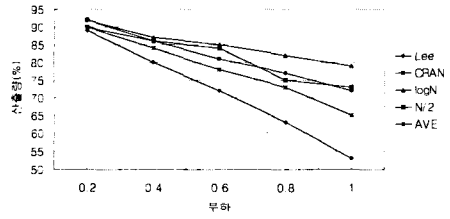

그림 11 셀 분할 계수에 따른 산출량 (N=8)

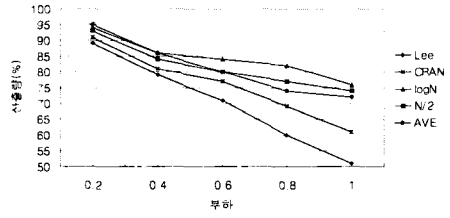

그림 12 셀 분할 계수에 따른 산출량 (N=16)

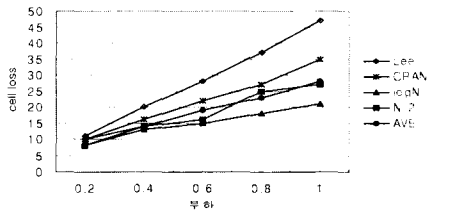

그림 13 셀 분할 계수에 따른 셀 손실율 (N=8)

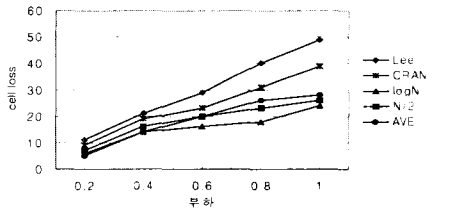

그림 14 셀 분할 계수에 따른 셀 손실율 (N=16)

알 수 있다. 그러므로 8×8 반얀망이나 16×16 반얀망에서 실험 결과는 셀 분할하는 것이 셀 분할을 하지 않은 모델에 비해 산출량에 효율적임을 나타내고 있다.

그림 13은 8×8 반얀망에서의 셀 손실율을, 그림 14는 16×16 반얀망에서의 셀 손실율을에 대한 그림은 복사본 수의 평균값에 의한 셀 분할이 상대적으로 적은 셀 손실율을 가짐을 보여준다. 셀 분할에 의한 멀티캐스팅은 부하에 크게 영향을 받지 않고 거의 일정한 산출량과 셀 손실율을 가지고 있다. 이는 입력포트로 들어오는 셀을 공정하게 분산하는 역할을 하게 되어 전체적으로 스위치 트래픽의 균형을 이루어주기 때문이다.

그림 15와 그림 16은 8×8 반얀망에서의 셀 분할 계수에 따른 지연율을, 16×16 반얀망에서의 셀 지연율을 보여주고 있다. 대체로 지연 정도가 부하에 따라 서서히 증가함을 보여주고 있다. 그러나 셀 분할을 하지 않은 모델인 CRAN은 트래픽의 부하에 따라 지연율이 상대적으로 크게 늘어남을 알 수 있다.

그림 17은 8×8 반얀망에서 복사본 수에 따른 산출량을 보여주고 있다. 주어진 복사본 수의 값은 각 입력포트에 동일하게 주어지며, 망 내부의 처리 속도에 따른 부하에 따른 산출량의 관계를 보여준다.

그림에서 'logN+0.8'은 셀 분할 계수를 log_2N으로 부하를 0.8로 적용한 것이며, 'CRAN+1.0'은 CRAN 반얀 스위치에 부하를 1.0로 적용한 경우를 나타낸다. 복사본 수에 따른 셀 분할과 셀 분할을 하지 않은 경우, 산출량은 현격히 차이가 남을 알 수 있다. 복사본의 요구의 개

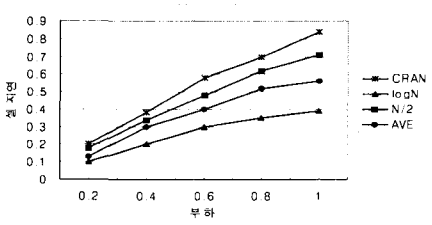

그림 15 셀 분할 계수에 따른 지연율 (N=8)

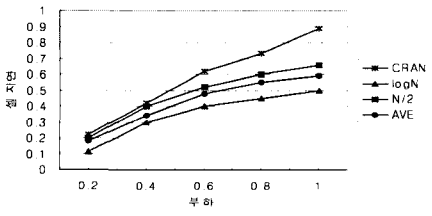

그림 16 셀 분할 계수에 따른 지연율 (N=16)

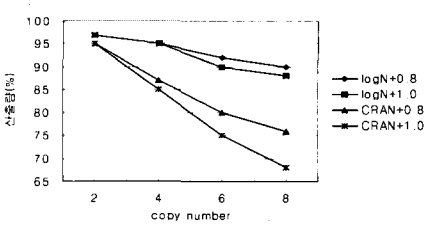

그림 17 복사본 수에 따른 산출량 (N=8)

수가 많아질수록 산출량은 다소 떨어질 수 있지만 셀 분할을 하지 않은 것은 입력포트의 불공정성에 의해 하위포트의 셀 지연 시간이 증가되어 산출량이 검토되고 있다.

5. 결 론

멀티캐스트 스위치에 대한 많은 연구가 있어 왔으나, 이들 제안된 멀티캐스트 스위치들이 가지고 있는 공통적인 과제로서는 오버플로우 문제, 큰 복사본 수의 처리, 블로킹 문제 외에 입력에 있어서 공정성 문제, 그리고 우선순위 제어를 들 수 있다. 특히 입력에 있어서 불공정성은, 입력포트에 도착하는 패킷은 상위 포트가 하위 포트보다 우선 하므로 상위 포트의 복사본 수가 최대가 되어지면 하위 포트의 셀은 계속해서 다음 사이클로 미루어진다는 점이다.

이런 문제점을 해결하기 위해 본 연구에서 제안된 스위치는 셀 분할과 공유 메모리, 그룹 분할, 그리고 반얀망으로 구성하고, 입력 패킷의 요구 수에 따른 적절한 복사와 넌블로킹 특성을 이용함으로써 산출량과 셀 손실에 있어서 많은 개선을 보였다. 특히 입력 셀 분할과 공유버퍼의 사용은 산출량에서 좋은 결과를 보여주며, 복사망이 출력 수보다 큰 복사본 수를 해결하고 있다..

결국 큰 복사본 수에 대한 작은 복사본 수를 가진 입력포트에 도착한 패킷의 불공정한 입력 부하 문제를 해결하여 시스템 전체 지연 시간을 줄여 산출량을 증가시켰다.

다만 스위치 내의 셀간 충돌이나 오버플로우, 공정성 문제를 해결하기 위한 멀티캐스트 스위치의 복잡도는 높아지고 있으나, 주지하는 바와 같이 VLSI 기술의 발전으로 하드웨어복잡도의 문제는 해결되는 것으로 간주되며, 차후의 과제로서는 네트워크 폭주 제어, 비디오 회의와 같은 실시간 서비스를 요구하는 우선 순위 제어 기능을 가진 스위치에 관한 연구이며, 이미 현재 진행중에 있다.

참 고 문 헌

[1] A. Huang, S. Knauer, "Starlite : A wideband digital switch," in Proc. Globecom'84, pp.121-125, Nov. 1984.

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

[3] Tony T. Lee, "Nonblocking Copy Networks for Multicast packet Switching," IEEE Journal on Selected Areas in Comm., Vol. 6, No. 9, pp.1455-1467, Dec. 1988.

[4] C. L. Tarng, J. S. Meditch, A. K. Somani, "Fairness and Priority Implementation in Non-Blocking Copy Network," International Conf. on Comm., pp.1002-1006, 1991.

[5] Wen De Zhong, Yoshikuni Onozato, Jaidev Kaniyil, "A Copy Network with Shared Buffers for Large-Scale Multicast ATM Switching," IEEE/ACM Trans. Networking, Vol. 1, No. 2, pp.157-165. 1993.

[6] Xinyi Liu, H. T. Mouftah, "Overflow Control In Multicast Networks," Proc. of Canadian Conf. on Electrical and Computer Engineering, Vancouver, B. C., pp.542-545, 1993.

[7] Jae W. Byun, Tony T. Lee, "The Design and Analysis of an ATM Multicast Switch with Adaptive Traffic Controller," IEEE/ACM Trans. on Networks, Vol. 2, No. 3, pp. 288-298, June 1994.

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

[9] Sikdar, B.; Manjunath, D., "Queueing analysis of scheduling policies in copy networks of space-based multicast packet switches," IEEE/ACM Transactions on Networking, Vol. 8, No.3, pp. 396 -406, June 2000.

[10] Dongsoo S. Kim, Ding-Zhu Du, "Performance of Split Routing Algorithm for Three-Stage Multicast Networks," IEEE/ACM Trans. on Networking, Vol. 8, No. 4, pp.526-534, August 2000.

[11] Ho, J. D.; Singh S.; Sharma, N. K. , "Modeling of replicate-at-send multicasting in shared-memory ATM switches," Globecom'00. IEEE, Vol. 1, pp.505-509, 2000.

[12] Kang, S. H,; Changhwan Oh, Sung, D. K. ,"Performance evaluation of a high-speed ATM switch with multiple common memories," IEEE Trans. on Comm., Vol. 50, No. 2, pp.332-340, Feb. 2002.

[13] 손동욱, 손유익, "그룹 분할 알고리즘을 이용한 멀티캐스트 스위치", 2001년 정보과학회 춘계학술발표논문집, 제28권 1호, pp 1232-234. 2001.

손 동 욱

1982.3~1986.2 계명대학교 전자계산학과 학사. 1986.3~1988.2 숭실대학교 전자계산학과 석사. 1998.3~계명대학교 컴퓨터공학과 박사수료. 1989.3~1993.2 LG정보통신 안양연구소 연구원. 1993.3~현재 혜천대학 컴퓨터통신계열 조교수

손 유 익

1976년 경북대학교 전자공학과 학사 1979년 경북대학교 대학원 전자공학과 석사. 1990년 경북대학교 대학원 전자공학과(컴퓨터공학전공), 공학박사. 1979~1984년 한국전자기술연구소 컴퓨터연구부 선임연구원. 1984~현재 계명대학교 컴퓨터공학과 교수. 1994~1995년 한국정보과학회 영남지부장. 1999~2001년 컴퓨터·전자공학부 학부장. 2002~현재 계명대학교 전산원장. 관심분야는 병렬 알고리즘 및 구조 MIN, ATM 스위칭 네트워크 등