

트래픽 유형에 따른 QoS 보장을 위한 지연 적응적인 스케줄링에 관한 연구

준회원 이 상 호*, 정회원 오 영 환**

A Study on the Delay Adaptive Traffic Scheduling for QoS of Traffic Type

Sang Ho Lee*, Young Hwan Oh** *Regular Members*

요 약

ATM 망에서 제공되고 있는 음성, 영상, 데이터와 같은 다양한 서비스는 사용자의 만족도를 수용할 수 있어야 한다는 것을 전제조건으로 한다. 이러한 기본적인 요구사항을 충족시키기 위해서는 노드대 노드간의 자원관리와 오류제어 및 다양한 트래픽의 특성을 고려한 전송 순서의 결정에 해당하는 스케줄링 방법이 요구되어 진다. 본 논문에서는 이러한 기술 요소 중에서 트래픽 설정 단계에서 제공되는 트래픽 특성 및 QoS(Quality of Service) 정보를 바탕으로 교환 노드에서 발생하는 전달 지연 시간에 적응적인 스케줄링 방식을 제시하였다. 이 방식은 멀티미디어 서비스와 같이 혼합된 트래픽 특성을 갖는 구조에서 트래픽 구성비율에 따라 적용되는 지연 여유치를 매우 융통성 있고 효과적으로 조절할 수 있다. 성능분석을 위하여 기존의 스케줄링 방식인 WFQ (Weighted Fair Queueing) 방식과 제안한 스케줄링 방식의 수학적 분석을 수행하였으며 이 두 방식의 결과식을 비교하여 교환 노드에서의 평균 지연과 셀 처리에 관한 성능을 분석하였다. 그리고 수학적 분석에 대한 검증으로는 Simulation tool ARENA 3.0을 이용하여 제안한 알고리즘의 Worst case와 기존의 알고리즘의 성능을 비교하였다.

ABSTRACT

A service offer from ATM network is variety service as voice, video, data. It should be premised on acceptance of user's satisfaction. For satisfaction of fundamental demand, resource management in node to node and applicable determination of the transmission sequence consider traffic characteristic in variety traffic for the scheduling method could be required. At follow technical request, this thesis presents the adaptive scheduling method of occurrence transfer delay time from network about traffic characteristics and QoS information offered from traffic creation phases. This kind of method makes an offer of very adaptable and effective control from mixed traffic characteristic structure adapted to the delay margin of traffic arrangement rate alike the multimedia service which gets into the emerging method. I measured a performance of compared with two scheduling method's result equation in mathematical analysis of WFQ existing scheduling method and accomplish mathematical analysis of suggested scheduling method for suggested performance analysis scheduling method. Finally, we compared the normal case and that of the proposed algorithm in a worst case to verify the performance by using the Arena 3.0.

* 광운대학교 전자통신공학과 (lshxxx@hanmail.net)

** 광운대학교 전자통신공학과 (yhoh@daisy.kwangwoon.ac.kr)

논문번호 : 00250-0704, 접수일자 : 2000년 7월 4일

* 이 논문은 2000년도 광운대학교 교내 학술 연구비 지원에 의해 연구 되었음.

I. 서론

통신망의 성능은 고속 스위칭 기술의 발달과 광 전송 기술의 발전으로 인하여 광대역 서비스가 가능해 졌으며 이로 인해 네트워크는 고속 전달과 수백Mbps부터 수십 Gbps까지의 대역폭으로 운용될 것이다.

ATM을 기반으로 하는 B-ISDN은 현재 VOD (Video on Demean)와 VCF(Video Conference)등과 같은 광대역 멀티미디어 서비스(Broadband Multimedia Service)를 제공한다. 이를 지원하기 위한 통신망 설계시 가장 중요한 쟁점은 광대역 멀티미디어의 원활한 서비스를 보장해 주기 위한 대역폭을 보장해 주는 것에 있다. 망 연결시 망 서비스 제공자와 사용자는 대역폭 사용에 최적화를 해야한다. 이때 트래픽 유형에 맞는 QoS를 보장하기 위한 협약을 할것이고 평균 지연은 QoS 요구에 대한 사항을 결정하는 중요한 요소가 된다. 따라서 중간 교환 노드에서의 지연 경계(Delay Bound)가 중요한 특징으로 대두된다.

ATM 망은 각 소스의 요구에 따라 대역폭 배치를 해주기 위한 스케줄링 알고리즘과 고속, 능률적이며 성능 Bound에 적합한 알고리즘들이 제안되어 왔다^{[1][2]}.

대부분의 스케줄링 알고리즘은 스케줄링 동안 Time Sensitive Traffic과 Loss Sensitive Traffic을 구분하지 못함으로, Time Sensitive Traffic이 요구하는 종단간의 지연경계를 보장해 주지 못한다. 따라서 Time Sensitive Traffic의 지연경계를 보장해 주기 위해 대역폭의 낭비를 초래하게 된다^[11]. Round-Robin방식으로 동작하는 Fair Queueing(FQ) 같은 알고리즘은 다른 소스들에게 각각 다른 가용 대역폭을 제공할 수 없고, 이러한 단점을 보완하기 위해 나온 방법들이 Weighted Fair Queueing (WFQ)^[3]이며, Generalized Process Sharing (GPS)^[4]로도 표현된다. 또한 보다 정확한 Fluid Fair Queueing(FFQ) 모델을 따를 수 있도록 개선한 방법이 WF2Q^[5]이다.

이러한 Queueing Model을 이용한 Input Queueing Packet Switching System^[6]에서 항상 일어나는 HoL(Head of Line) Blocking은 종단간 지연을 증가시켜 처리율(Throughput)을 감소시키는 문제점을 유발시킨다. 위와 같은 문제점을 해소하기 위해 제안된 LA(Look Ahead) 기법과 PIM(Parallel Iterative

Matching) 알고리즘은 스케줄링시 계산의 복잡도를 증가시킨다는 단점이 있다^[7]. 따라서 본 논문에서는 입력 트래픽을 세분화하여 처리해 줌으로서 각각의 트래픽 유형에 적합한 QoS보장과 효율성을 높이고, WFQ 모델을 기반으로한 알고리즘을 사용하여 Queue의 공정한 전송을 보장함으로써 다른 소스들에게 각각 다른 가용 대역폭을 제공해 주었다. HoL Blocking의 문제점을 해결하는 방안으로 스케줄러 종단에 Reader를 달아 순차적으로 셀을 출력해 고효율을 실현하는 한편 스케줄링시 계산의 복잡도는 낮추고 지연을 줄일 수 있는 알고리즘을 제안하였다.

II. 관계이론

2.1 Fair Queueing Algorithm

WFQ의 한 형태인 GPS의 환경에서 N개의 session이 연결 되어있는 서버에서의 해당 연결 가중치를 $\Phi_1, \Phi_2, \dots, \Phi_N$ 로 특징지어지면 다음과 같다.

$$\frac{W_j(t_1, t_2)}{W_i(t_1, t_2)} \geq \frac{\Phi_j}{\Phi_i} \quad j=1, 2, \dots, N \quad (1)$$

이 경우 $W_j(t_1, t_2)$ 는 시구간 (t_1, t_2) 에서 해당연결에 대한 전송량을 나타내며 이상적으로 공평한 서비스를 제공하려면 (t_1, t_2) 에서 식(1)의 좌우변이 같은 값을 가질때이다. 이것은 할당된 가중치에 비례한 만큼 전송이 이루어 졌다는 것을 나타낸다. 그리고 이때 특정 흐름에 대하여 보장되는 서비스율은 다음과 같다.

$$r_i(t_1, t_2) = \frac{\Phi_i}{\sum_{j \in B_{GPS}(t_1)} \Phi_j} r \quad (2)$$

여기서, r 은 서버의 전송율을 나타내며, $B_{GPS}(t_1)$ 은 시간 (t_1) 에서의 링크상의 모든 연결의 집합을 나타낸다.

$r_i(t_1, t_2)$ 의 범위는 다음과 같다.

$$r_i(t_1, t_2) \geq r_i \quad (3)$$

여기서, r_i 는 session i 에 대하여 서버가 보장해 주는 최소 대역폭이다. 따라서 r_i 는 다음과 같이 정의할 수 있다.

$$r_i = \frac{\Phi_i}{\sum_{j=1}^N \Phi_j} r \quad (4)$$

그리고 임의의 시간 간격에 있어 개별적인 연결의 실효 전송량은 식 (2)를 이용하여 해당 시간 간격을 곱한 값으로 다음과 같다.

$$W_i(t_1, t_2) = \frac{\Phi_i}{\sum_{j \in B_{GPS}(t)} \Phi_j} r(t_2 - t_1) \quad (5)$$

여기서 $B_{GPS}(t)$ 는 시구간 (t_2, t_1) 에서는 변경되지 않는다. 이때 시구간의 길이를 미소구간으로 보내면 서비스율의 순간 증가율은 다음과 같이 주어진다.

$$r_i^{inst} = \frac{\Phi_i}{\sum_{j \in B_{GPS}(t)} \Phi_j} r \quad (6)$$

이상의 경우는 비트 단위의 전송을 고려한 경우로서 실제로 적용하기 위해서는 패킷 단위로 고려되어야 한다. 여기서 가상 시간들은 패킷들에 대한 전송 순서를 결정하기 위해 사용되고 있으며, 이때 주어지는 가상 시간 및 종료 시간은 각각의 패킷 단위로 계산되어지며, 이에 대한 일반적인 수식은 다음과 같다.

$$F_i^k = \frac{L_i^k}{\Phi_i} + \max(F_i^{k-1}, v(a_i^k)) \quad (7)$$

여기서, L_i^k 는 class i 의 k 번째 패킷의 길이이며 a_i^k 는 class i 의 k 번째 패킷의 도착 시간, $v(t)$ 는 가상시간 함수이다. 그리고 서버의 용량(Capacity)을 C 라 하면 $v(t)$ 의 순간 변화율은 다음과 같다.

$$\frac{dv(t)}{dt} = \frac{C}{\sum_{j \in B_{GPS}(t)} \Phi_j} \quad (8)$$

이상과 같은 기본적인 개념을 바탕으로 구현상의 복잡성을 단순화시키고 트래픽의 특성을 만족하는 형태의 여러 가지 유형들이 있다^{[8][9]}. 그리고 위의 식에서 볼 수 있듯이 이러한 Fair Queuing 유형들은 임의의 시간 t 에서 전송할 패킷이 있는 backlog queue들을 대상으로 전송이 이루어지는 work conserving 형태이며, 제한된 지연과 주어진 최소한의 처리율을 제공하는 특징을 가지고 있다. 그리고 이러한 요소들의 보장을 통하여 QoS를 제공할 수 있는 방법을 제공한다. 그러나 패킷 단위의 반복적인 많은 계산을 필요로 하기 때문에 현실적으로 적용하기 어려우며, 동일한 개념을 바탕으로 보다 간단한 방법을 통하여 효과적으로 적용하고 있다.

2.2 Multi-class에서의 WFQ방식

Multi-traffic환경하에서의 각 트래픽 유형의 QoS를 보장하기 위해 ATM Forum 4.0^[10]에서는 표 1과 같은 서비스 분류를 해놓았다.

표 1. ATM service Category Attributes

Attribute	ATM Layer Service Category				
	CBR	rtVBR	nrtVBR	UBR	ABR
Traffic Parameter					
PCR and CDVT	specified				
SCR,MBS,CDVT	n/a	specified		n/a	
MCR	n/a			specified	
QoS parameter					
pick to pick CDV	specified		unspecified		
max CTD	specified		unspecified		
CLR	specified		unspecified	specified	
Other Attributed					
Feedback	unspecified			specified	

그리고 이러한 Multi-Traffic의 QoS를 보장하기 위해 각 유형마다 우선순위(Priority)를 인가해주고 이러한 우선순위 값이 스케줄링시 중요한 전송순서 결정 기준이 되었다.

그림 1은 실제적으로 스케줄링이 수행되는 관계를 설명한 것으로서 CBR, rt-VBR, nrt-VBR, ABR, UBR 서비스를 요구한 backlog queue를 대상으로 한 것이다. 여기서 서비스 종료시간 정보는 일반적

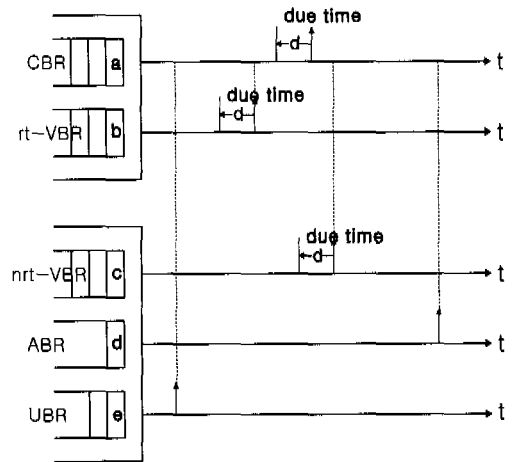


그림 1. 전송 순서 결정

인 Fair Queuing 방법에서의 종료시간인 F_i 에 해당된다. 그리고 상위 우선순위에 해당되는 셀의 서비스 시간이 이 서비스 종료시간을 넘을때 QoS에 영향을 미치게 된다. 따라서 그림 1과 같은 셀이 queue에 있을 때 기존의 알고리즘에 따른 서비스는 a, b, c, d, e순이며 이렇게 상위 우선순위의 셀이 있을때 이 상위 우선순위의 셀이 서비스될 때까지 하위 우선순위 셀이 queue에서 대기함으로 가용의 대역폭을 낭비하는 경우가 발생한다.

III. 제안한 스케줄링 알고리즘

3.1 알고리즘

제안한 스케줄링 서버(Scheduling Server)의 구조는 그림 2와 같다.

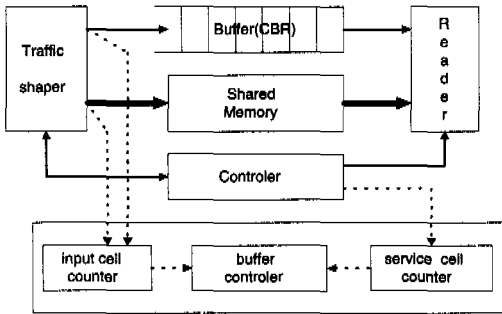


그림 2. 입력 버퍼형 스위치의 스케줄링 서버의 구조

그림 2의 스케줄링 서버의 구조는 다음과 같이 세부분으로 나뉘고 각 부분의 알고리즘은 다음과 같다.

3.1.1 셀분리부, 저장부

단계 0

- 입력 Data를 트래픽 유형에 따라 순차적으로 분류

단계 1

- 입력 셀 개수(Input Cell Counting)를 해준 후 결과값을 buffer controller에 통보

단계 2

- Buffer controller는 buffer의 현재 용량을 임계값과 비교
- 임계값 초과시 폭주 예방제어 수행

단계 3

- 분류된 CBR 셀을 우선순위 비트(bit)와 서비스 종료시간을 tag로 표시

- 분류된 rt-VBR, nrt-VBR, ABR, UBR 셀을 우선순위 비트, 서비스 종료시간, WFAT address를 tag로 표시
- 위 tag정보를 제어부로 통보

단계 4

- 각각의 memory로 셀저장

위의 수행절차는 입력되는 data를 각 트래픽 유형으로 분류한 후 전송순서를 결정해주는 tag를 생성하는 과정이다. 이 tag에는 QoS 파라미터를 이용하여 얻을 수 있는 이론적인 전송주기를 바탕으로 실제로 셀 전송이 수행되는 과정에서 실시간 트래픽의 특징인 제시간에 서비스 해줘야하는 특징을 만족해 주는 서비스 종료 시간을 tag에 표시하여 제어부로 알려준다.

3.1.2 스케줄러 제어부

단계 0

- 입력된 tag정보를 가지고 각 트래픽 유형의 유무를 순차적으로 조사

단계 1

- 검색된 상위 우선순위를 가진 셀의 서비스 종료시간을 R_n 와 비교하여 서비스 유무결정

단계 2

- 다음 서비스가 결정된 셀의 저장 주소(Memory address)를 출력부에 통보

단계 3

- 서비스된 셀에 대해 서비스 셀 계수 (service cell counting)를 해준 후 결과를 buffer controller에 통보

위의 수행절차는 상위 우선순위부터 서비스 종료 시간과 R_n 값을 비교하여 가용대역폭에 대한 하위 우선순위의 서비스의 가부를 결정해 주는 단계이다. 여기서 R_n (Round Trip Time)는 시스템이 스케줄링 알고리즘을 한번 수행하는 시간을 나타내며, 이 시간은 스케줄링 서버의 성능에 의해 결정된다. 위의 단계에서 여유 대역폭이 생겨났어도 하위 우선순위의 셀이 있는 backlog queue가 존재하지 않을 때에는 다시 최상위 우선순위의 queue부터 다시 비교를 해준다.

3.1.3 출력부

단계 0

- 제어부에서 보내는 출력 정보에 따라 저장부에서 셀 reading

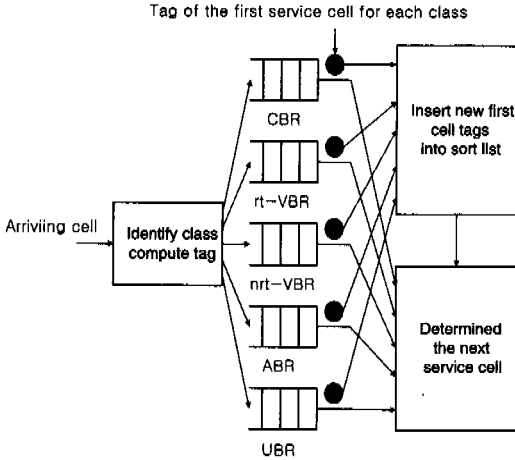


그림 3. 서비스 등급별 스케줄링 방법

단계 1

- reading한 셀을 다음단의 Self routing ATM switch로 출력

이상에서 제한한 알고리즘에서 그림 1과 같은 셀의 처리 순서를 보면 e, b, a, c, d의 순으로 셀 전송이 이루어 지므로 최적의 대역폭 이용을 실현할 수 있다.

IV. 수학적 분석

수학적 분석을 위해 다음과 같은 파라미터를 정의한다.

- k : 시스템에 존재하는 셀의 총수
- k_c : 시스템에 존재하는 CBR 셀의 수
- k_r : 시스템에 존재하는 π -VBR 셀의 수
- k_n : 시스템에 존재하는 nrt-VBR 셀의 수
- k_a : 시스템에 존재하는 ABR 셀의 수
- k_u : 시스템에 존재하는 UBR 셀의 수
- τ_{ps} : 분리부에서의 처리시간
- τ_{pc} : 제어부에서의 처리시간
- τ_{pr} : 출력부에서의 처리시간
- τ_p : 한 셀의 처리 시간 ($\tau_p = \tau_{ps} + \tau_{pc} + \tau_{pr}$)
- τ_x : 다음 셀을 서비스 해주기 위한 시스템 대기시간
- $k = k_c + k_r + k_n + k_a + k_u$
- p_c : 마진(Margin) 확률

4.1 Nomargin에서 셀의 Mean delay
그림 1의 입력 버퍼형 스위칭 시스템에 k 개의

셀이 존재할 때의 CBR 셀에 대한 Mean queuing time은 다음과 같다.

$$C_T = \frac{\sum_{c=1}^{k_c} (c-1)(\tau_{pc} + \tau_{pr}) + \tau_x}{k_c} = \frac{\sum_{c=1}^{k_c} (c-1)(\tau_p + \tau_x)}{k_c} \tag{9}$$

Nomargin에서의 하위 우선순위 셀에 대한 서비스는 상위 우선순위 셀의 서비스가 모두 끝나고 난 후에 실행 될 것이다. 따라서 π -VBR의 셀에 대한 Mean queuing time은 다음과 같다.

$$r_T = \frac{\sum_{r=1}^{(k_c+k_r)} (r-1)(\tau_p + \tau_x) - \sum_{c=1}^{(k_c+k_r)} (r-1)(\tau_p + \tau_x)}{k_r} = \frac{\sum_{r=k_c+1}^{(k_c+k_r)} (r-1)(\tau_p + \tau_x)}{k_r} \tag{10}$$

식(10)과 같은 방식으로 nrt-VBR, ABR, UBR 셀에 대한 Mean queuing time은 식(11), (12), (13)과 같다.

$$n_T = \frac{\sum_{n=1}^{(k_c+k_r+k_n)} (n-1)(\tau_p + \tau_x)}{k_n} \tag{11}$$

$$a_T = \frac{\sum_{a=1}^{(k_c+k_r+k_n+k_a)} (a-1)(\tau_p + \tau_x)}{k_a} \tag{12}$$

$$u_T = \frac{\sum_{u=1}^{(k_c+k_r+k_n+k_a+k_u)} (u-1)(\tau_p + \tau_x)}{k_u} \tag{13}$$

시스템의 전체 셀에 대한 Mean queuing time은 식(9), (10), (11), (12), (13)를 이용하여 식(14)와 같이 나타낼 수 있다.

$$QT_{nm} = \frac{\sum_{i=1}^k (i-1)(\tau_p + \tau_x)}{k} \tag{14}$$

식(14)에 의해 시스템에서의 전체 셀에 대한 Mean delay는 다음과 같다.

$$MD_{nm} = \frac{\sum_{i=1}^k (i\tau_p + (i-1)\tau_x)}{k} \tag{15}$$

4.2 Margin에서 셀의 Mean delay

시스템의 셀에 대한 Mean queuing time은 식(14)을 이용하여 다음과 같다.

$$QT_{mg} = \frac{\sum_{i=1}^k (i-1)(\tau_p + \tau_x(1-p_d))}{k} \quad (16)$$

식(16)에 의해 시스템에서의 Mean delay는 다음과 같다.

$$MD_{mg} = \frac{\sum_{i=1}^k (i\tau_p + (i-1)(1-p_d)\tau_x)}{k} \quad (17)$$

V. 시뮬레이션 및 성능분석

5.1 시뮬레이션

본 절에서는 수학적 분석에 의해 얻어진 식을 토대로 다음과 같은 환경하에서 시뮬레이션을 수행하여 두 알고리즘 간의 성능을 비교한다. 시뮬레이션을 위한 도구로는 "Systems Modeling"사가 개발한 Arena Version 3.0을 이용하였고, 주어진 환경은 다음과 같다.

- 음성 트래픽 : on-off 소스 모델 적용
 - 유음구간 : 평균 1000msec인 지수분포
 - 무음구간 : 평균 1500msec인 지수분포
- 프로세싱 지연 : 2.74 μ sec
- CTD : 5msec
- rt-VBR 트래픽의 SCR : 1Mbps
- nrt-VBR 트래픽의 SCR : 1Mbps
- ABR 트래픽 셀 전송율
 - PCR : 1Mbps
 - MCR : 100kbps
- UBR 트래픽 : 트래픽의 특성상 고려하지 않음
- 입력 cell은 한번 이상의 service contention을 수행

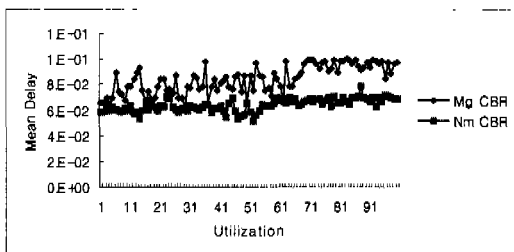


그림 4. CBR 트래픽에서의 셀 평균지연

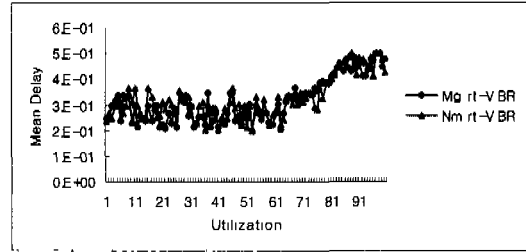


그림 5. rt-VBR 트래픽에서의 셀 평균지연

그림 4와 그림 5는 실시간 트래픽에서의 Nomargin 알고리즘과 Margin 알고리즘의 Worst case에서의 셀 지연을 측정된 그래프이다. 그림에서 알 수 있듯이 제안한 알고리즘에서의 CBR 트래픽의 지연이 기존 알고리즘에서의 CBR 트래픽의 지연보다 QoS를 만족하는 범위 안에서 다소 큰 값을 갖는 것을 알 수 있고, rt-VBR 트래픽의 경우에는 두 알고리즘 모두다 거의 같은 지연 시간을 갖는 것을 알 수 있다.

그림 6과 그림 7은 비실시간 트래픽에서의 셀 지연을 측정된 그래프이다. Nomargin 알고리즘의 경우 시스템의 부하가 높아질 수록 상위 우선순위 class 셀에 대한 서비스를 위해 시스템 대기시간이 길어지며, 이러한 현상으로 시스템에 존재하는 하위 우선순위 class 셀에 대한 지연 시간이 급격한 증가

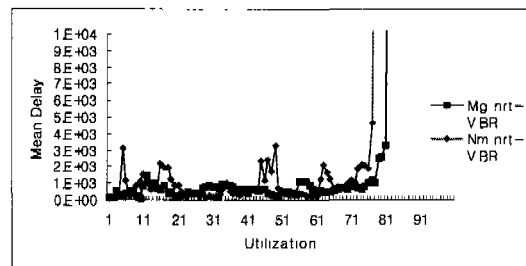


그림 6. nrt-VBR 트래픽에서의 셀 평균지연

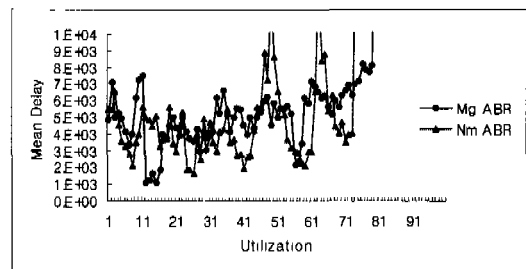


그림 7. ABR 트래픽에서의 셀 평균지연

를 보이는 반면, 제안한 알고리즘에서의 경우 시스템의 부하가 높아져도 상위 우선순위 class의 서비스 여유시간동안 하위 우선순위 class 셀을 서비스 해 줌으로서 지연 시간의 완만한 증가를 볼 수 있다.

5.2 성능분석

기존의 스케줄링 알고리즘과 제안한 스케줄링 알고리즘의 성능 비교는 두 알고리즘의 Mean delay와 처리 효율을 비교하였다.

시스템의 전체 셀에 대한 두 알고리즘의 Mean delay는 식 (15)와 (17)에 의해 다음과 같다.

$$\frac{MD_{mg}}{MD_{nm}} = \frac{\tau_p + \tau_x(1-p_d)}{\tau_p + \tau_x} \therefore MD_{mg} \leq MD_{nm} \quad (18)$$

위의 식(18)에서 보여지는 결과를 그래프 상으로 도식하면 다음 그림 8과 같다.

그림 8은 OC-3의 전송속도를 갖는 non-blocking ATM self routing switch에서 80%의 부하가 걸렸을 경우를 고려하여 M/D/1으로 분석한 결과 그래프이다. 도시된 바와 같이 기존의 알고리즘보다 제안한 알고리즘에서의 셀 평균 지연 시간이 적은 것을 알 수 있으며 이 평균 지연 시간 비는 마진율이 증가함에 따라 감소됨을 알 수 있다.

그림 9에서의 TP_{mg} 은 제안한 알고리즘에서의 시구간 (t_1, t_2)에서의 셀 처리율이고, TP_{nm} 은 기존의 알고리즘에서의 시구간 (t_1, t_2)에서의 셀 처리율을 나타낸다. 이 그래프의 값은 그림 8과 동일한 환경에서의 기존의 알고리즘과 제안한 알고리즘의 셀 처리율의 비에 대한 그래프이다. 그림의 결과값에서 알 수 있듯이 동일한 시구간에서의 셀 처리율이 제안한 알고리즘에서 개선된 것을 볼 수 있다.으며, 이러한 두 가지 결과로 제안된 알고리즘의 성능에

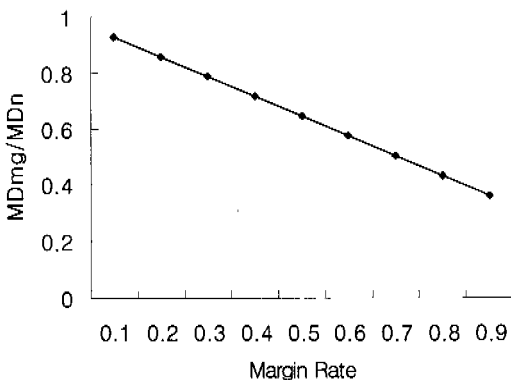


그림 8. 마진율에 따른 두 알고리즘간 평균 지연시간 비율

대한 검증을 할 수 있다. 즉 주어진 마진율에 의해 상위 우선순위 class에 대한 QoS에 영향을 미치지 않는 범위에서의 여유 대역폭에서 하위 우선순위 class에 대한 서비스를 수행한다는 것을 의미한다.

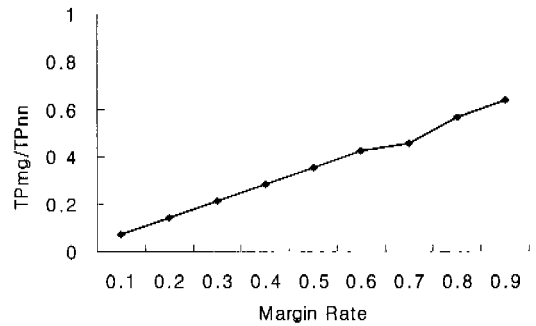


그림 9. 마진율에 따른 두 알고리즘간 셀 처리 성능비

VI. 결론

본 논문에서는 유선 ATM 멀티미디어 서비스 구조에서 중간 교환 노드에서 각 트래픽 유형의 특성에 맞는 QoS를 보장해 주는 범위 내에서 일정 분의 마진을 적용하여 셀 전송 순서를 결정하는 스케줄링 알고리즘을 제안하였다. 실제의 트래픽의 흐름에 있어서 트래픽의 구성비에 따라 트래픽 class별로 할당된 우선순위만을 가지고 셀의 전송순서를 결정하는 기존의 스케줄링 알고리즘에서는 하위 우선순위 class 셀의 burst가 발생할 경우 상위 우선순위 셀에 대한 서비스를 위하여 가용의 대역폭에서도 시스템 대기가 발생하고, 이러한 현상으로 인해 대역폭의 낭비를 초래하게 된다. 반면 본 논문에서 제안한 스케줄링 알고리즘은 최초 망 연결시 주어지는 QoS정보를 바탕으로 트래픽의 지연 특성을 고려한 적정 마진율을 적용하여 상위 우선순위 class의 QoS에 영향을 주지 않는 범위에서 하위 우선순위 class의 최적의 셀 서비스에 의해 처리 효율을 높일 수 있는 스케줄링 방법을 제공한다. 시뮬레이션 결과에서 알 수 있듯이 제안한 알고리즘의 Worst case의 실시간 트래픽에서 CBR은 QoS를 만족하는 범위에서 기존의 알고리즘보다 좀 더 큰 지연값을 나타내었고, rt-VBR은 두 알고리즘이 거의 같은 특성을 보였다. 비실시간 트래픽의 경우 시스템의 부하가 높아질수록 기존의 알고리즘에서는 상위 우선순위 class의 QoS를 보장해 주기 위해 시스템의 대기시간이 늘어난다. 따라서 하위 우선순위 class의 평균지연이 급격히 커짐진다. 그러나 제안한

알고리즘에서는 시스템의 부하가 높아져도 상위 우선순위 class의 서비스 여유시간동안 하위 우선순위 class의 셀을 처리해 줌으로써 셀 지연시간의 완만한 증가를 볼 수 있다. 이러한 결과로 시스템이 안정적으로 셀을 서비스해 줌을 알 수 있다. 이에 따라 일반적으로 지연시간이 큰 하위 우선순위 class 셀에 대한 지연시간이 줄어들어서 시스템의 전체적인 지연시간이 줄어들게 된다.

참 고 문 헌

[1] Mark W. Garrett, A Service Architecture for ATM : From Applications to Scheduling, IEEE Network, pp.6-14, May/June 1996.

[2] Pawan Goyal, Harrick M. Vin, and Haichen Cheng, Start-Time Fair Queueing: A Scheduling Algorithm for Integrated Services Packet Switching Networks, IEEE/ACM Transaction on Networking, Vol.5, No.5, pp.690-704, Oct. 1997.

[3] Yuhua Chen, Jonathan S. Turner, Design of a Weighted Fair Queueing Cell Scheduler for ATM Network, IEEE Globecom98, Vol.1, pp.405-410, Nov. 1998.

[4] G. De Veciana, George Kesidis, Bandwidth Allocation of Multiple Qualities of Service Using Generalized Processor Sharing, IEEE Transactions on Information Theory, Vol.42, No.1, pp.268-272, Jan. 1996.

[5] Jon C. R. Bennett, H. Zhang, WF2Q: Worst-Case Fair Weighted Fair Queueing, IEEE INFORCOM 96, pp.3-12, March 1996.

[6] Bin Li, Mounir Hamdi, and Xi-Ren Cao, An Efficient Scheduling Algorithm for Input-queueing ATM Switches, IEEE International Workshop on Broadband Switching Systems, pp.148-154, Dec. 1997.

[7] 손장우, 입력 버퍼링 ATM 스위치에서 최대 처리율 향상 및 QoS 제공방안: 연세대학교 공학 박사 학위 논문, 1996.

[8] S. Golestani, A Self-Clocked Fair Queueing Scheme for Broadband Application, IEEE INFORCOM94, pp.636-646, June 1994.

[9] Jon C. R. Bennett, H. Zhang, Hierarchical Packet Fair Queueing Algorithms, ACM

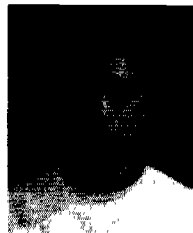
SIGCOMM, pp.143-156, Aug. 1996.

[10] Traffic Management Specification V4.0, ATM Forum Document AF-TM-0056.000, April 1996

[11] 류효용, 효율적인 출력 버퍼형 ATM 스위치의 Port Manager에 관한 연구: 광운대학교 공학 박사 학위 논문, 1998.

[12] Jean Waleand, Pravin Varaiya, High Performance Communication Networks, Morgan Kaufmann Publishers, 2th Edition.

이 상 호(Sang-Ho Lee)



1999년 2월 : 홍익대학교
전자전산공학과 졸업
1999년 3월~현재 : 광운대학교
전자통신공학과 석사
<주관심 분야> High speed
Network, QoS in
Networks

오 영 환(Young-Hwan Oh)

정회원

한국통신학회 논문지 제25권 제9호 참조