
종단간 지연 큐 서비스 방식을 이용한 실시간 트래픽 QoS 개선에 관한 연구

김광준* · 나상동*

A Study on the Improvement of Real-Time Traffic QoS using the Delay Guarantee Queue Service Discipline of End-to-End

Gwang-Jun Kim* · Sang-Dong Ra*

이 논문은 2000년도 조선대학교 연구년제 해외파견 지원에 의하여 연구되었음

요 약

ATM망에서 큐 서비스 방식에서는 상위 프레임과 하위 프레임을 따로 정의하여 대역폭 할당과 전송 지연 보장을 분리할 수 있기 때문에 종단간 전송한계에서 QoS를 보장 할 수 있는 단순화한 큐 서비스 구현을 제안 하였다. 제안한 큐 서비스 방식은 실시간 트래픽과 비실시간 트래픽을 나누어 저장하기 때문에 연결별 트래픽 조절기 없이도 지연 지터 특성이 개선되었다.

제안된 큐 서비스 방식을 이용하여 연결 수락 제어는 트래픽 왜곡의 영향을 계산할 수 없는 트래픽 왜곡의 영향이 실제로 존재한다는 가정 하에 트래픽 왜곡에 대한 영향에 종단간 지연을 보장함으로써 연결 수락 가능 범위를 다양한 프레임의 길이에 대해 나타내었으며, 프레임의 길이가 클수록 제안된 큐 서비스 방식의 연결 수락 범위가 지연 EDD방식과 비교해서 많은 차이를 보이지만 프레임의 길이를 점차 줄여갈수록 연결 수락 가능 범위가 지연 EDD 방식에 근접함을 나타내었다.

ABSTRACT

We propose a cell-multiplexing scheme for the real-time communication service in ATM network and a new service discipline guarantee end-to-end delay based on pseudo-isochronous cell switching. The proposed scheme consists of two level frame hierarchy, upper and lower frame, which is used to assign the bandwidth and to guarantee the requested delay bound, respectively.

Since the proposed algorithm employs two level frame hierarchy, it can overcome the coupling problem which is inherent to the framing strategy. It can be comparatively reduce the complexity, and still guarantee the diverse delay bounds of end-to-end. Besides, it consists of two components, traffic controller and scheduler, as the input traffic description model and regulates the input traffic specification.

The function of the traffic controller is to shape real-time traffic to have the same input pattern at every switch along the path. The end-to-end delay is bounded by the scheduler which can limit the delay variation without using per-session jitter controllers, and therefore it can decrease the required buffer size. The proposed algorithm can support the QoS's of non-real time traffic as well as those of real time traffic.

키워드

Real-Time Traffic, End-to-End Queue Service, Delay Guarantee

1. 서론

ATM망에서 트래픽에 대해 개개의 서비스 품질(QoS: Quality of Service)을 보장할 수 있는 트래픽 제어가 핵심 기술 중의 하나로서 트래픽 제어에는 연결 수락 제어, 체증 제어, 흐름 제어, 우선 순위 제어 등이 포함된다. 큐 서비스 방식은 연결 수락 제어 과정에서 결정된 사용자 매개 변수를 이용하여 대역폭과 버퍼 공간을 이용하여 사용자가 요구하는 QoS를 만족시킬 수 있도록 여러 연결로부터 들어 온 셀들의 우선 순위를 결정하고 그 순위에 따라서 전송하는 기법이다^[1,3].

정렬 우선 순위 큐 서비스는 대역폭 할당과 전송 지연 한계를 독립적으로 보장할 수 있다는 장점이 있지만, 정렬 과정이 필요하고 동작과 연결 수락 제어가 복잡한 단점이 있다. 이에 비해 프레임 방식은 구현이 단순하고 트래픽을 재구성하는 과정이 있기 때문에 중단간 트래픽 특성이 좋은 장점이 있으나 대역폭 할당과 전송 지연 한계가 결합되어 QoS를 만족시켜야 하기 때문에 QoS의 보장에 있어서 유연성과 효율성이 떨어지는 단점이 있다^[2,12].

제안한 큐 서비스는 상위 프레임과 하위 프레임의 두 계층 프레임에 의해서 이루어지게 되며, 상위 프레임은 대역폭을 할당하기 위하여 사용되고, 하위 프레임은 셀 교환이 발생하는 곳으로서 하위 프레임의 길이가 모두 같기 때문에 동기화되어 있다. 모든 노드가 동기화되어 있지 않을 경우에는 노드 사이에 프레임의 결합 문제가 발생하게 되어 그에 따른 추가 지연이 발생하게 되나 프레임 결합에 따르는 추가 지연의 최대는 하위 프레임의 길이가 되므로 하위 프레임의 길이가 작을 경우 대기 지연에 비하여 충분히 무시할 수 있음을 수학적인 연산을 통해 증명하고, 지연 EDD

(Earliest Due Date)방식의 연결 수락 범위와 거의 근접함을 시뮬레이션 결과를 통해 QoS됨을 입증한다. 또한 노드간의 프레임 연동방식을 이용하여 연결별로 독립된 상위 프레임과 하위 프레임을 정의하여 사용함으로써 유연한 대역폭을 할당하여 전송 지연 특성에 대해 분석하고, 실시간 트래픽 사용자 파라미터를 통해 전송 지연 번이를 개선하여 지연 분석에 이용되는 트래픽 특성 모델과 지연 한계를 유도함으로써 프레임 방식의 문제점을 해결한다.

II. 중단간 지연 한계를 보장하기 위한 큐 서비스 방식

1. 제안한 큐 서비스 방식의 구조

지연 EDD 방식이나 지터 EDD 방식과 같은 EDD 방식군은 연결 별 전송 지연을 보장해 주기 위한 방식으로는 우수하지만, 지연 한계를 기반으로 하여 우선 순위를 결정하기 때문에 교환을 위해서는 상태가 변화할 때마다 정렬이 필요하여 동작이 복잡한 단점을 가지고 있다.

본 논문에서 제안한 큐 서비스 방식은 동시성 셀 교환과 같이 두 단계의 프레임을 사용하여 대역폭과 전송 지연을 분리시키고, 지연 EDD 방식과 같이 연결에 대한 전송 지연 요구에 따라 서로 다른 셀 교환 방식을 사용함으로써 프레임 방식의 한계를 개선한다. 상위 프레임은 Stop-and-Go 방식과 같이 대역폭을 할당하기 위하여 사용되고, 하위 프레임은 셀 교환이 일어나는 기본 단위로 전송 선로별로 정의되는데 모든 전송 선로에서 하위 프레임의 길이가 모두 같으며 동

기화되어 있다.

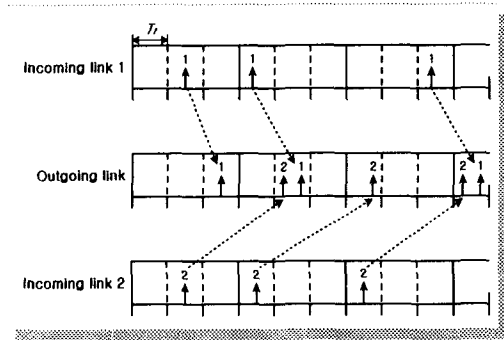


그림 1. 제안한 큐 서비스 방식의 구조
Fig.1 Proposed queue service discipline scheme

대역폭 할당을 위해서 독립된 상위 프레임을 사용하여, 연결별로 서로 다른 프레임을 사용하고 노드간의 프레임 이동 방식은 제안된 방식을 사용함으로써 전송 지연과 필요한 버퍼의 크기가 달라지게 된다.

그림 1은 제안된 큐 서비스 방식으로서 연결 1의 요구 전송 지연이 $2T_i$ 이므로 즉시적 전송 방식을 사용하고 연결 2는 요구 전송 지연이 $4T_i$ 이므로 3 프레임 전송 방식을 사용한다. 첫번째는 TAT_i 가 $AT_i^k + \tau_{i,s}$ 보다 적거나 같은 경우에 트래픽이 버스트 허용 한도 내에서 발생하는 경우이므로 마감 시간과 TAT_i 는 식(1), (2)와 같이 나타낼 수 있다.

$$DL_i^k = \max \{ AT_i^k, DL_i^{k-1} + T_i \} \dots \dots \dots (1)$$

$$TAT_i = \max \{ AT_i^k, TAT_i \} + T_{i,s} \dots \dots \dots (2)$$

두 번째는 TAT_i 가 $AT_i^k + \tau_{i,s}$ 보다 큰 경우 트래픽이 버스트 허용 한도를 초과하여 망에 유입되는 경우이기 때문에 마감 시간과 TAT_i 는 식(3), (4)와 같다.

$$DL_i^k = TAT_i + d_i \dots \dots \dots (3)$$

$$TAT_i = TAT_i + T_{i,s} \dots \dots \dots (4)$$

단 AT_i^k 는 연결 i 의 k 번째 셀의 실제 도착시간이며, d_i 는 연결 i 의 요구하는 전송 지연 한계, T_i 는 연

결 i 의 셀의 최소 간격으로 PCR(Peak Cell Rate)의 역수이다. $T_{i,s}$ 는 연결 i 의 SCR (Sustainable Cell Rate)을, $\tau_{i,s}$ 는 연결 i 의 버스트 허용 한도(BT: Burst Tolerance)를 나타내며, TAT_i 는 트래픽이 SCR의 전송율로 입력된다고 가정했을 때의 예상 도착 시간으로 초기에는 0으로 되어 있다.

큐 서비스 방식을 ATM망에 적용하기 위해서는 실시간 트래픽뿐만 아니라 비실시간 트래픽의 QoS도 보장할 수 있어야 한다. 비실시간 VBR 트래픽의 경우 PVR, SCR을 보장해 주어야 하며, ABR 트래픽의 경우에는 MCR (Minimum Cell Rate)이라는 최소 전송율을 보장해 주어야 한다. 기존의 큐 서비스 방식들은 대부분 실시간 서비스의 QoS 보장에 초점을 맞추었기 때문에 비실시간 서비스는 최선 노력 서비스로 분류하여 가장 낮은 우선 순위를 할당하는 방식을 사용하였다. 큐 서비스 방식 역시 실시간 서비스에 중점을 두고 지연 한계에 의한 우선 순위에 의해서 셀을 전송하기 때문에 이를 그대로 비실시간 트래픽에 적용하면 QoS의 보장을 하지 못한다. 그러나 제안된 큐 서비스 방식은 프레임에 의한 방식이기 때문에 비실시간 트래픽을 서비스하는데 상위 프레임을 이용하여 실시간 서비스 방식과 조화시킬 수 있다.

실시간 서비스와 동일한 큐 서비스 방식을 사용하기 위해서는 비실시간 서비스 역시 우선 순위로 마감 시간을 사용한다. 비실시간 연결 i 의 k 번째 셀의 마감 시간은 식(5)와 같다.

$$DL_i^k = \left(\lceil \max \left\{ \frac{(AT_i^k - \phi_i)}{T_{i,c}}, \frac{k}{n_i} \right\} \rceil + 1 \right) \cdot T_{i,c} + \phi_i \dots \dots \dots (5)$$

단 AT_i^k 는 연결 i 의 k 번째 셀의 실제 도착시간이며, $T_{i,c}$ 는 연결 i 의 프레임 길이를 n_i 는 연결 i 의 프레임당 전송할 수 있는 셀의 개수를 나타내고, ϕ_i 는 연결 i 의 프레임이 처음 시작한 시간을 나타낸다.

$\lceil \cdot \rceil$ 는 \cdot 보다 작지 않은 최소의 정수를 나타내는 함수이므로 큐 서비스 방식에 결정론적인 전송 지연 특성에 대해 분석하고 실시간 트래픽에 대한 전송 지연 변이를 개선하게 된다.

2. 큐 서비스 방식의 전송 지연 특성

연결 i 의 셀들이 k 개의 노드를 경유하여 목적지에 도착한다고 하면 연결 i 의 k 번째 셀이 겪게되는 중단간 지연은 식(6)과 같다.

$$D_i^k = \sum_{j=0}^{k-1} d_{i,j}^k + \sum_{j=1}^k \tau_j \dots\dots\dots (6)$$

단 $d_{i,j}^k$ 는 연결 i 의 k 번째 셀이 j 번째 노드에서 겪는 지연을, τ_j 는 $j-1$ 번째 노드와 j 번째 노드 사이에서 경험하는 전송 선로에서의 전달 지연이다. j 번째 노드에서 겪는 대기지연은 제안하는 큐 서비스 방식이 프레임을 사용하여 전송하기 때문에 연결 수락 제어 과정에서 요구하는 전송 지연 한계보다 항상 작게된다. 따라서 연결 수락 제어과정에서 연결 i 에 대해 요구한 전송 지연 한계가 $d_{i,j}$ 라고 하면 연결 i 의 k 번째 셀이 j 번째 노드에서 겪는 대기 지연의 한계는 식(7)과 같다.

$$d_{i,j}^k \leq d_{i,j} \dots\dots\dots (7)$$

라운드 로빈 큐가 비어 있을 때 도착한 셀은 즉시 전송되어 대기 지연의 하한은 j 번째 노드의 한 셀의 전송 시간이 된다. j 번째 노드에서 한 셀을 전송하는 시간이 s_j 이라면 식(7)은 식(8)과 같다.

$$s_j \leq d_{i,j}^k \leq d_{i,j} \dots\dots\dots (8)$$

식(7)과 식(8)을 이용하면 중단간 지연의 한계는 식(9)와 같다.

$$\sum_{j=0}^{k-1} s_j + \sum_{j=1}^k \tau_j + D_i^k \leq \sum_{j=0}^{k-1} d_{i,j} + \sum_{j=1}^k \tau_j \dots\dots\dots (9)$$

식(9)는 큐 서비스 방식이 지연 EDD 방식과 마찬가지로 사용자가 요구하는 지연 한계를 보장할 수 있다는 것을 나타낸다. 실시간 트래픽 지연 특성을 개선하기 위해서는 그림 2와 같이 제어기에 도착한 셀의 마감 시간을 계산하여 라운드 로빈 큐로 전송하여 실시간 트래픽과 비실시간 트래픽을 나누어서 저장한다. 실시간 트래픽 큐와 비실시간 트래픽 큐의 동작은 T_j

가 지날 때마다 우선 순위가 변하고 실시간 트래픽 큐는 작업 비보존 방식과 비실시간 트래픽 큐는 작업 보존 방식을 사용한다. 실시간 트래픽 큐에 저장되어 있는 셀들은 우선 순위가 1이 큐에 저장되어 있는 셀만 전송이 가능하며, 나머지 큐에 저장된 셀들은 그 큐의 우선 순위가 1이 될 때까지 기다려야 한다. 따라서 전송 우선 순위는 실시간 트래픽 큐의 우선 순위가 1인 큐, 비실시간 트래픽 큐, UBR 트래픽 큐의 순서로 된다.

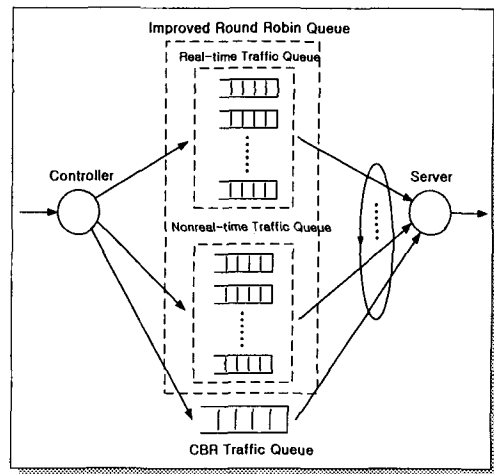


그림 2. 지터 특성을 개선하기 위한 큐 서비스 방식의 구조
Fig. 2 The new queue service discipline to improve jitter characteristic.

실시간 트래픽의 지연 특성은 우선 연결 i 의 k 번째 셀이 j 번째 노드에서 겪는 대기 지연의 상한이 식(7)과 같고, 연결 i 의 k 번째 셀이 저장된 큐의 우선 순위가 1이 되는 순간의 최소는 $d_{i,j} - 2T_j$ 이므로, 연결 i 의 k 번째 셀이 j 번째 노드에서 대기 지연 한계는 식(10)과 같다.

$$d_{i,j} - 2T_j \leq d_{i,j}^k \leq d_{i,j} \dots\dots\dots (10)$$

식(6)과 식(10)을 이용하면 중단간 지연의 한계는 식(11)과 같다.

$$\sum_{j=0}^{k-1} (d_{i,j} - 2T_j) + \sum_{j=1}^k \tau_j \leq d_{i,j}^k \leq \sum_{j=0}^{k-1} d_{i,j} + \sum_{j=1}^k \tau_j \dots (11)$$

식(11)을 식(9)와 비교해보면 종단간 지터가 현격하게 감소하고 경유하는 중간 노드의 수와 하위 프레임의 길이에 의해서 결정되므로 쉽게 예측할 수 있다.

3. 실시간 트래픽 채널 설정

마감시간을 우선 순위로 하는 큐 서비스 방식의 연결 수락은 두 가지의 요건을 만족시켜야 이루어진다. 첫 번째 요건은 요구하는 망 자원을 할당해 줄 수 있어야 한다는 안정성 조건이고, 두 번째는 지연 한계를 보장해 줄 수 있어야 한다는 스케줄링 가능 조건으로서 두 가지의 요건 중 하나라도 만족시킬 수 없으면 그 연결은 거부된다.

지연 EDD 방식에서 안정성 조건과 스케줄링 가능 조건은 식(12)와 식(13)과 같다.

$$\lim_{t \rightarrow \infty} \sum_{j=1}^N A_j^*(t) < 1 \dots\dots\dots (12)$$

$$t \geq \sum_{j=1}^N A_j^*(t - d_j) + \max s_k \quad t \geq d_1 \dots\dots\dots (13)$$

단, $A_j^*(t)$ 는 t 동안에 연결 j 에서 발생하는 셀의 전송 시간의 총합을 $A_j(t)$ 라고 했을 때 모든 t, r 에 대해서 $A_{j,r,t+1} \leq A_j^*(t)$ 를 만족시키는 함수로 트래픽 제한 함수라 한다. s_k 는 연결 k 의 셀의 전송 시간을 나타내며, d_j 는 연결 j 의 요구시 전송 지연 한계를 나타낸다. 현재 시간을 t 라고 할 때 하위 프레임에 전송되어야 하는 셀의 총 전송 시간은 $\sum_j A_j^*(t - d_j + T_j)$ 가 된다. 그리고 처음에 마감시간이 남아 있는 셀이 전송되었을 가능성이 있기 때문에 스케줄링 가능 조건은 식(14)와 같다.

$$t \geq \sum_{j=1}^N A_j^*(t - d_j^*) + \max s_k \quad t = nT_j, \quad n \geq 2 \dots\dots\dots (14)$$

단 $d_j^* = \left\lfloor \frac{d_j}{T_j} \right\rfloor \cdot T_j$ 이며 안정성 조건은 식(12)와 같다. 식(13)과 식(14)를 비교 해보면 효율적인 큐 서비스 방식은 d_j 를 사용하였기 때문에 지연 EDD 방식에 비해서 스케줄링이 가능한 연결의 수가 적어지

는 것은 프레임 사용하기 때문이다.

연결 수락 제어 방식을 사용하기 위해서는 먼저 트래픽 제한 함수를 적절한 트래픽 모델로 나타내야 하며 흔히 분석에 사용되는 (σ, ρ) 모델을 사용한다. 이 모델은 트래픽의 버스트성을 나타내는 σ 와 일정한 트래픽 증가 비율을 나타내는 ρ 를 사용하여 버스트성이 있는 트래픽을 나타낸 것으로서 이 모델에 의한 연결 j 의 트래픽 제한 함수는 다음 식(15)와 같다.

$$A_j^*(t) = \sigma_j + \rho_j t \dots\dots\dots (15)$$

(σ, ρ) 모델은 간단하여 분석에 쉽게 사용할 수 있는 장점이 있지만 실제의 트래픽을 너무 과장해서 나타내기 때문에 분석 결과가 실제와 차이가 많이 나는 단점이 있다. (σ, ρ) 모델을 사용했을 때 지연 EDD에서 보장할 수 있는 한계는 연결들을 요구하는 지연 한계가 작은 순서부터 연결 번호를 붙이면 식(16)과 같다.

$$d_j = \frac{\sigma_j + \sum_{i=1}^{j-1} (\sigma_i - \rho_i d_i^*) + \left\{ \sum_{i=1}^j \sigma_i - 1 \right\} \cdot T_j + \max s_k}{1 - \sum_{i=1}^j \rho_i} \dots\dots\dots (16)$$

(σ, ρ) 모델을 요구하는 지연 한계가 다른 트래픽을 나타낼 수 있으며, 트래픽의 버스트 특성은 같다.

종단간 지연을 보장하기 위해 연결은 트래픽 특성을 요구하며, 망에서는 지연 한계를 유도하기 위해 버스트성 있는 트래픽을 기술하는 모델 ($X_{min}, X_{avg}, I, S_{max}$)을 이용한다. 이 모델은 최고 전송율과 일정 기간의 평 사용하여 큐 서비스 방식의 연결 수락 범위를 균 전송율 등의 두 가지 파라미터로 버스트성을 나타내는 VBR 트래픽과 같은 경우에 적합한 모델로써 트래픽 특성을 이용한다.

III. 종단간 지연보장을 위한 실시간 트래픽 변수 특성

1. 실시간 트래픽 사용자 파라미터

실시간 채널은 트래픽 성능 파라미터와 관계되어 각

각의 망 접속은 다음의 파라미터 ($X_{min}, X_{ave}, I, S_{max}$) 에 의해 구성된다. X_{min} 은 패킷의 최소 상호도착시간, X_{ave} 는 평균 상호 도착 I 를 초과하여 연산된 평균 상호 도착시간으로 최악의 경우에 의해 X_{ave} 는 I 기간의 가장 버스트한 기간 동안 평균 상호 도착시간, I 는 평균적인 간격 즉, X_{ave} 가 연산되는 평균 기간, S_{max} 는 최대 패킷의 크기이다.

파라미터는 중단간 지연과 최대 셀 손실율과 같은 실행 파라미터를 제공하며, 트래픽 모델에서 최대 이용율은 식 (17)과 같이 된다. I 를 초과하는 평균율 상에서의 상위 한계는 식 (18)과 같이, 평균 대 피크율의 비 즉 실시간 트래픽의 버스트 비율은 식 (19)와 같다.

$$S_{max}/(X_{min} * I) \dots\dots\dots (17)$$

$$S_{max}/X_{ave} \dots\dots\dots (18)$$

$$X_{ave}/X_{min} \dots\dots\dots (19)$$

트래픽 표현자의 파라미터 ($X_{min}, X_{ave}, I, S_{max}$)를 만족하는 트래픽 소스는 I 기간동안 「 I/X_{ave} 」보다 많은 패킷을 생성하지 않는다. 식(19)에서 X_{ave}/X_{min} 는 평균율 대 최대율의 비로서 이를 트래픽의 버스트성의 정도가 어떠한가를 나타내는 버스트 비율이다. 즉, 버스트 비율이 클수록 더욱 버스트한 트래픽이다. X_{ave}/X_{min} 와 I 이러한 두 가지 변수가 트래픽의 버스트성에 기여하는데 즉, X_{ave}/X_{min} 와 I 가 크면 클수록 트래픽은 더욱 큰 버스트성을 갖으며, 최대 버스트 길이는 식(20)과 같다.

$$I/X_{ave} * X_{min} \dots\dots\dots (20)$$

버스트한 트래픽은 연결의 최대 대역폭이 망 링크의 속도와 비교하여 낮고, 버스트 기간이 짧을 때 통계적 서비스 모드로 다중화할 수 있다.

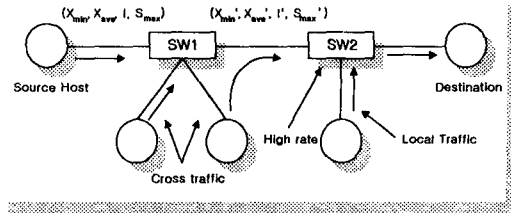


그림 3. 트래픽 왜곡 패턴
Fig. 3 Traffic distortion pattern.

그림 3과 같이 소스 호스트에서 트래픽 표현자의 파라미터 ($X_{min}, X_{ave}, I, S_{max}$)를 만족할 수 있는 트래픽을 고려하여 처음 스위치는 나타낸 트래픽 표현자의 파라미터는 ($X_{min}', X_{ave}', I', S_{max}'$)이기 때문에 트래픽 왜곡의 패턴은 원래의 표와 같지 않으므로 X_{min} 이 어떻게 변하는가를 알 수 있어서 몇 개의 노드를 통과해야만 함으로 최소 상호 도착시간이 X_{min} 보다 더욱 작을 경우 최악에 두 개의 연속적인 패킷이 0으로 도착한다.

평균 상호도착시간 X_{ave} 가 I 의 상관성 때문에 연산되는 평균간격에 의존하여 변하므로 지연 한계 분석은 평균 상호도착시간의 인수(factor)로 계산을 한다. 흐름 모델에서 간격 u 동안 그 출력의 양이 $\sigma + u \times \rho$ 에 의해 한계 되면, 트래픽 소스는 (σ, ρ)를 만족한다고 말할 수 있다. δ 는 최대 버스트 크기이고, ρ 는 트래픽의 평균율로서 망의 유입점에서 (σ_j, ρ_j)를 만족하는 연결 트래픽은 스위치 i 의 유입점에서 ($\sigma_i + \sum_{k=1}^n (\rho_j \times d_{kj}) \times \rho_j$)로써 특성화된다. 여기서 $d_{k,j}$ 는 스위치 k 에서의 연결 j 의 근거리 지연 바운드로 재특성화된 트래픽을 가짐으로써 근거리 지연 분석이 된다.

제안된 큐잉 서비스 구조를 이용하여 데이터가 버퍼에서 겪는 지연 및 지터의 특성 즉 큐잉을 분석하여 ATM망에 들어오는 다양한 데이터 특히, 양도 많고 버스트적인 데이터들이 출력으로 가기 전에 잠시 버퍼에 대기하면서 겪는 데이터의 상태를 분석해 많은 데이터를 다중화 함으로써 지연 및 지터에 대한 분석을 수행하고, 이러한 분석을 토대로 데이터 트래픽 제어 및 모델링 망 자원의 효율적 이용 그리고 셀 손실에

대한 정보 및 대책을 세울 수 있으며, 지연 분석은 단일 소스일 때와 다중 소스가 다중화 될 때의 지연분포를 나누어서 분석한다.

지연분석을 연산하기 위해 사용된 변수 D 는 종단간 지연, j 는 접속수, i, k 는 교환기의 수, $d_{i,j}$ 는 노드 j 에서 채널에 의해 발생된 연결 j 의 종단간 지연, $d_{min}^{k,i}$ 는 k 번째 스위치에서 접속 j 의 지역 낮은 한계 지연, u 는 시간 간격, S_{max} 는 링크를 초과하여 전송될 수 있는 최대 패킷의 크기, $N_{max}^{i,j}(u)$ 는 u 동안 i 노드에서 j 접속에 도달하는 패킷의 최대수, $N_{max}^{src}(u)$ 는 u 기간 동안 소스에 의해 생성되는 패킷의 최대 수를 의미한다.

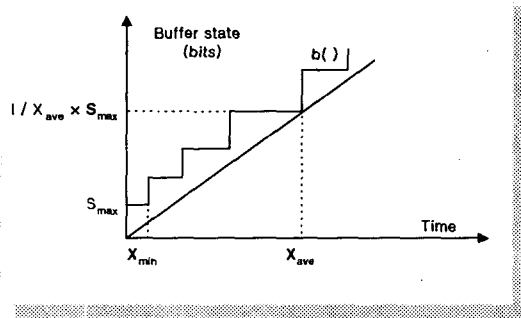


그림 4. 트래픽 억압 함수
Fig. 4 The Constraint function of traffic.

그림 4 에서 계단과 같은 곡선은 트래픽 억압함수로서 이는 어떤 시간 간격동안에 도달 할 수 있는 비트들의 수로 정의하고 $\bar{b}(u)$ 로 표현하기 때문에 j 접속의 트래픽 소스의 표현자 $(X_{min}^j, X_{avg}^j, I_j, S_{max}^j)$ 를 만족하면 간격 u 에 도달 될 수 있는 비트들의 최대 수 $\bar{b}_j(u)$ 은 다음과 같다.

간격 u 와 평균 시간 간격 I 와의 관계에 따라서 $\bar{b}_j(u)$ 는 세 가지 경우로 나눌 수 있으므로 $u=I$ 일 때 $\bar{b}_j(u)$ 는 식(21), $u < I$ 일 때 $\bar{b}_j(u)$ 는 식(22), $u > I$ 일 때 $\bar{b}_j(u)$ 는 식(23)과 같다.

$$\bar{b}_j(u) = (\lceil I_j / X_{avg}^j \rceil) \times S_{max}^j \dots\dots\dots (21)$$

$$\bar{b}_j(u) = (\lceil u / X_{min}^j \rceil) \times S_{max}^j \dots\dots\dots (22)$$

$$\bar{b}_j(u) = \lceil \min(\lceil u \text{ mod } I_j / X_{min}^j \rceil, \lceil I_j / X_{avg}^j \rceil) + \lfloor u / I_j \rfloor \lceil I_j / X_{avg}^j \rceil \times S_{max}^j \rceil \dots\dots\dots (23)$$

그림 4는 백로그와 활성기간을 트래픽 표현자 $(X_{min}, X_{avg}, I, S_{max})$ 를 이용하여 트래픽 억압 함수가 된다.

2. 실시간 트래픽 연결 수락 제어 분석

제안된 큐잉 서비스를 이용하여 지연 범위 연산을 구하고 FIFO 서비스를 위한 연결 수락 제어는 두 가지 단계로 나누어서 실행된다. 첫 번째 충분한 자원들이 이용되는 단계와 채널에 대해 필요한 망의 자원을 비축해 두는 단계로 이루어진다.

지연에 이용될 수 있는 휴지기 $\Delta_j = D_j - \bar{D}_j$ 로 되기때문에 D_j 는 요구된 종단간 지연 바운드로서 비율 제어된 망들에서 노드 i 에 대해 할당된 Δ_j 의 일부분으로 $d_{i,j}$ 를 증가시키는 노드 상에서 휴지기를 분할하는 것이다. 그러나 $d_{i,j}$ 는 노드 i 에서 지터 지연을 증가시키기 때문에 망들을 보존하는 작업에 불충분하며, 지터 지연은 i 에서 뿐만 아니라 i 노드와 목적노드 사이에서도 증가한다. 그러므로 휴지기의 분할은 초기 노드에서 증가되는 지터에 의해 야기되는 지연뿐만 아니라 현재의 망에서 증가되는 지연도 고려해야 한다.

σ_{i-1} 에 의한 상위 지연 바운드의 증가 $d_{i-1,j}$ 는 $\sigma_{i-1} * S_{max}^j / (I * X_{max}^j)$ 으로서 $d_{i,j}$ 를 증가시켜 식(6-6)의 \bar{d} 을 이용하여 노드 $i+1$ 에서 \bar{d}_{i+1} 은 $[\sigma_{i-1} + \sigma_{i-1} * S_{max}^j / (I * X_{max}^j)] * S_{max}^j / (I * X_{min}^j)$ 에 의해 증가된다. 노드 $i+1$ 에서의 지연 증가는 이전 노드 지연에 의해 다중화될 때 지연증가에 대한 합이다. 노드에서 지연 영향에 의한 증가를 Δ_m 으로 나타내고 목적 노드에서 채널 j 에서 발생된 종단간 지연은 식(24)와 같다.

$$D_j + \Delta_j = \sum_{i=1}^n \bar{d}_i + \sigma_n + \sum_{m=1}^n \Delta_m \times \frac{S_{max}^j}{I} \times \frac{1}{X_{min}^j} \dots\dots\dots (24)$$

식(24)에서 나타낸 σ_n 는 노드 n 에서 트래픽의 버스트성을 나타내고, Δ_m 은 노드 m 에서 $d_{m,j}$ 이 증가함으로써 발생하는 추가적 지연이다. 식(24)는 중단간 지연을 두 개의 부분으로 분리하는데 이는 최소 중단간 지연과 추가적 중단간 지연이 트래픽의 버스트 특성에 따라 증가하며 채널 j 에 대해 노드 i 에서의 추가적 지연은 식(25)와 같이 쓸 수 있다.

$$\Delta_{i,j} = \sigma_{i-1} + \sum_{m=1}^{i-2} \Delta_{m,j} \times \frac{S_{\max}^j}{l} \times \frac{1}{X_{\min}^j} \dots\dots (25)$$

노드 i 에서 σ_i 에 대한 값은 버스트 특성에 따라 달라지고 σ_n 의 값이 식(25)에 의해 제한되는데 더 많은 트래픽의 버스트에 의해 목적지 이전 노드에서의 지연을 증가시키면 중단간 지연을 명백히 위반하기 때문이다. 다음의 노드에서 마찬가지로 $\sum_{j=1}^{m-1} \Delta_{m,j} \geq 0$ 이므로 $\sigma_i \leq \Delta_{i,j}$ 이다. 그러므로 각각의 노드 i 에서 채널 j 에 대한 할당된 지연 한계 $d_{i,j}$ 는 $\bar{d}_i + \sigma_i$ 와 일치함을 알 수 있다.

$\Delta_{i,j}$ 의 값에 대한 또 하나의 제한함수가 존재하는데 이 값은 노드 i 가 전체적으로 제어될 수 없는 i 노드 이외의 노드에서 실시간 트래픽 수락 범위를 제한하는데 \bar{d}_i 에서의 잠재적인 지연 증가 때문이다. \bar{d}_i 의 새로운 값은 채널 j 에 의해 링크를 공유할 때 링크를 공유하는 그 외 채널들의 지연 한계보다 작다. 그러므로 버퍼 요구량은 \bar{d}_i 의 새로운 값으로 만족될 수 있으며, 실시간 트래픽의 버스트 특성에 따라 중단간 지연을 보장하기 위해서는 트래픽 억압함수에 의해 선택된 $d_{i+1} = \Delta_{i+1,j} - \Delta_{i,j}$ 을 만족시키는 $\Delta_{i,j}$ 의 가장 큰 값을 선택함으로써 충족시킬 수 있다. σ_i 의 개별적 값의 선택은 그 이상의 버스트 트래픽을 전송하기 위해 망의 능력에 영향을 끼치며 망의 중앙을 향한 노드와 또는 더 높은 트래픽 부하를 갖는 노드에 대해 선택된 σ_i 의 높은 값들에 의해 중단간 지연 보장은 민감해진다.

IV. 시뮬레이션 결과 및 분석

제안된 중단간 지연 보장 큐 서비스 방식의 연결 수락 범위를 위해 두 종류의 트래픽을 이용했다. 각각의 트래픽의 평균 비트율은 76.8 Kbit/s이고, 최대 비트율을 15.36Mbit/s과 30.72Mbit/s로 설정하여 두 종류 트래픽의 평균 비트율에 비례한 평균 셀율은 200 cell/s이고, 최대 셀율은 각각 40,000 cell/s과 80,000 cell/s이다. 다른 두 종류의 트래픽에 대해서 요구하는 버스트 트래픽의 크기와 지연한계는 표 1과 같이 두 종류의 트래픽 특성을 이용하여 지연 EDD 방식과 비교하였다.

표 1. 실험에 사용한 연결의 트래픽 특성
Table 1. Traffic characteristic of connection for simulation

| | Delay Bound (ms) | BurstSize σ (Cell) | Transfer Rate ρ (Mbps) |
|-----------------|---------------------|------------------------------|--------------------------------|
| Traffic Class 1 | 17 | 2000 | 1 ~ 155.520 |
| Traffic Class 2 | 35 | 4000 | 1 ~ 155.520 |

그림 5는 제안된 큐잉 서비스 구조에서 다양한 버스트 특성을 가지고 있는 실시간 VBR 트래픽이 버퍼에서 대기되는 지연 및 지터 특성 즉 셀 지연 변이 허용치를 변화시킴으로서 수락 가능한 망의 부하율을 나타낸 것으로서 버스트성 트래픽이 큐잉 시스템에 유입될 때 버퍼 내에서의 누적량을 증가시켜, 지연을 증가시킨다. 그림 5의 결과에서 나타낸 바와 같이 버스트성 트래픽 유입은 요구되는 버퍼의 용량을 증가시킴으로서 버퍼 용량 증가에 따른 셀 지연이 증가하여 망내의 트래픽 혼잡이 발생하게 된다. 이러한 혼잡을 해결하기 위해서는 버스트성 트래픽 특성에 따른 셀 지연 변이 허용치를 증가시켜 줌으로서 망의 부하율을 줄일 수 있었고, 또 ATM 망의 중단간의 지연보장을 통한 실시간 서비스의 품질을 보장할 수도 있었다. 셀 지연 변이가 상대적으로 적을 때 수락 가능한 부하율의 감소가 거의 없음을 알 수 있기에 최대 셀율이 적은 영역에서는 망의 부하율에 비례하여 셀 손실율이 감소한다는 것도 알 수 있었다.

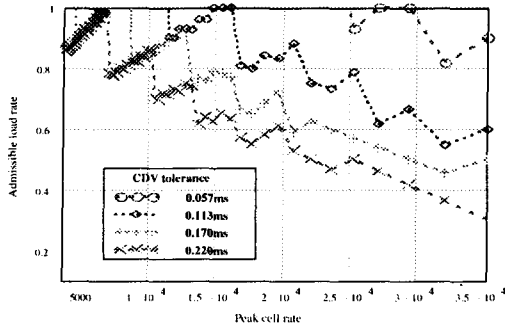


그림 5. 셀 지연 번이 허용치에 따른 VBR 트래픽의 수락 가능한 부하율
 Fig. 5 Admissible load rate for VBR traffic source with different CDV tolerance.

그림 6은 최대 버스트 크기에 대한 이용 율을 나타내므로 위 그림에서 셀 손실 율 요구가 10^{-4} 에서부터 10^{-10} 까지 변화한다고 가정할 때 링크 버퍼의 용량은 50cell, 셀 슬롯 율은 3532 cell/s로 설정하고 지속적인 셀 율은 3532 cell/s로 선택하였다. 최대 셀 율은 최대 버스트 크기와 평균 지속 셀 율로 계산되어질 수 있으며, 이는 수락 가능한 망의 부하에 중요한 효과를 가지기 때문에 최대 버스트 크기의 값을 정확하게 구하는 것이 중요하다.

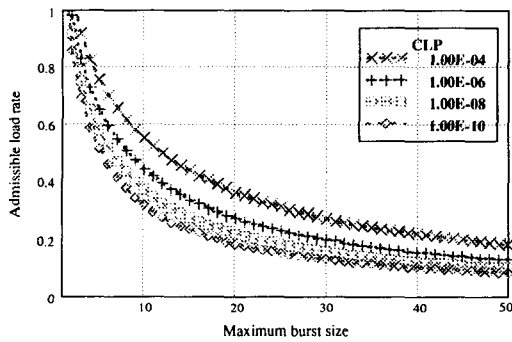


그림 6. 셀 손실 율에 따른 최대 버스트 크기에 대한 수락 가능한 부하율
 Fig. 6 Admissible load rate for maximum burst size with different cell loss rate.

위 그림에서 나타낸바와 같이 버스트 크기의 전체 영역을 통해 버스트 트래픽 특성에 따른 버퍼를 오버

플로우 하는 셀 손실 율이 감소할수록 수락 가능한 망의 부하율에 비례하여 감소하였다.

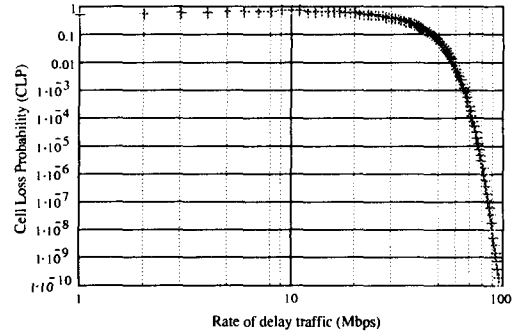


그림 7. (a) 지연 한계 조건이 없는 경우의 연결 수락 가능 범위
 Fig. 7 (a)Connection admission possibility region without delay tolerance condition

그림 7(a)는 전송지연 요구가 없는 경우에 연결 수락 가능 범위를 나타낸것으로서 셀 손실 확률을 10^{-10} 으로 할 때 연결 수락 범위는 100Mbps로 지연 번이 특성이 전혀 고려되지 않는 이상적인 연결 수락 가능 범위를 나타내고 있으며, 그림 7(b)는 지연 EDD 방식의 큐잉 서비스를 이용하였을 경우의 연결 수락 가능 범위가 69Mbps 정도임을 알 수 있다. 표 1에서 나타낸 두 가지 형태의 트래픽의 지연한계와 버스트 크기에 따른 트래픽 클래스 1과 트래픽 클래스 2의 프레임의 길이에 따른 실시간 트래픽 연결 수락 가능범위를 그림 8부터 그림 10까지 나타내었으며, 각각 그래프의 아래 부분이 연결 수락 가능한 범위이다. 제안된 큐 서비스 방식에서 프레임의 길이에 따른 실시간 트래픽의 연결 수락 가능 범위의 영향을 살펴보기 위해서 전송 프레임의 길이를 6ms, 2ms, 0.5ms로 변화시켜 가면서 나타낸 연결 수락 범위 시뮬레이션 결과를 그림 8에서부터 그림 10까지 나타낸 것이다.

제안된 큐 서비스 방식을 이용하여 실시간 트래픽 연결 수락 제어는 트래픽 왜곡의 영향이 실제로 존재한다는 가정 하에 트래픽 왜곡의 영향에 대한 종단간 지연을 보장함으로써 연결 수락 가능범위를 프레임의 길이를 다양하게 변화시켜가면서 시뮬레이션 결과를 나타내었다.

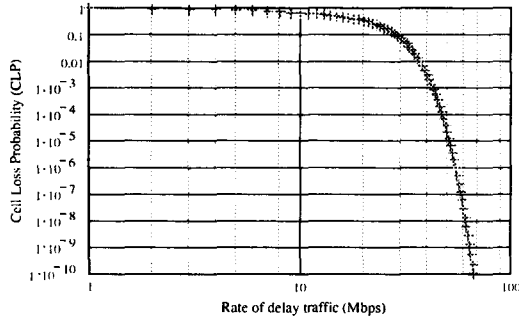


그림 7(b). 지연 EDD 방식의 연결 수락 가능범위
Fig. 7 (b) Connection admission possibility region of delay-EDD discipline.

그림 8은 제안된 큐 서비스 방식에서 셀 손실 확률을 10^{-10} 으로 설정하여 프레임의 길이 T_f 가 6ms일 때 트래픽 클래스 1과 2의 연결 수락 가능 범위는 각각 10.7 Mbps와 9.5Mbps로서 지연 EDD방식의 연결 수락 가능 범위 69Mbps와 비교해서 많은 차이를 보인다.

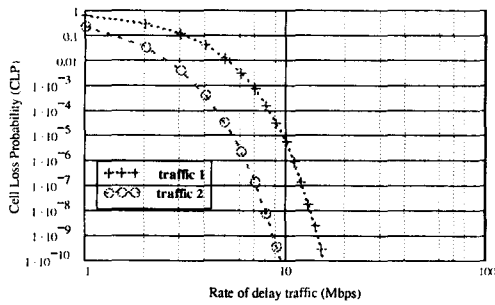


그림 8. 제안한 큐서비스 방식의 연결 수락 가능 범위 ($T_f=6ms$)
Fig. 8 Connection admission possibility region of proposed queue service discipline ($T_f=6ms$)

그림 9는 프레임의 길이 T_f 를 점차적으로 줄인 2ms로 설정하였을 경우의 트래픽 클래스 1과 2의 연결 수락 범위로서 각각 12 Mbps와 11.4 Mbps로서 그림 8과 비교해서 거의 차이가 없음을 알 수 있고, 또한 지연 EDD 방식의 연결 수락 범위와 비교해서도 많은 차이를 나타내고 있다.

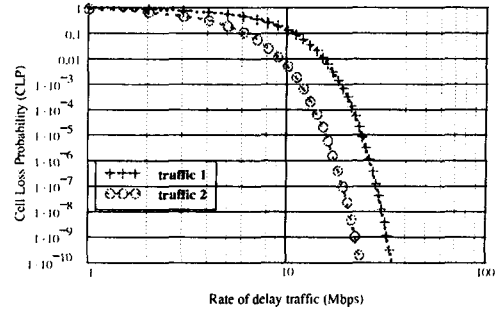


그림 9. 제안한 큐서비스 방식의 연결 수락 가능 범위 ($T_f=2ms$)
Fig. 9 Connection admission possibility region of proposed queue service discipline ($T_f=2ms$)

그림 10은 프레임의 길이 T_f 를 0.5ms로 설정한 경우의 트래픽 클래스 1과 2의 연결 수락 범위의 시뮬레이션 결과로서 각각 69 Mbps와 61 Mbps임을 알 수 있다. 위 그림의 결과로부터 트래픽 클래스 1의 연결 수락 가능 범위와 지연 EDD 방식의 큐 서비스를 이용한 연결 수락 가능 범위의 비교하면 69 Mbps로서 일치함을 알 수 있고 트래픽 클래스 2의 경우 60.1 Mbps로 지연 EDD방식의 연결 수락 가능 범위에 거의 근접하여 차이가 없음을 알 수 있다.

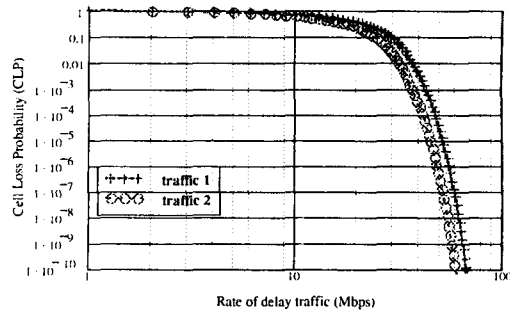


그림 10. 제안하는 큐서비스 방식의 연결 수락 가능 범위 ($T_f=0.5ms$)
Fig. 10 Connection admission possibility region of proposed queue service discipline ($T_f=0.5ms$)

위 시뮬레이션 결과로부터 제안된 큐 서비스 방식의 성능은 전송 프레임의 길이에 좌우되며, 전송 프

임의 길이가 짧을수록 제안하는 서비스 방식의 연결 수락 가능 범위는 지연 EDD방식에 점차적으로 접근하게 된다.

Ⅴ. 결론

ATM망에서 큐 서비스 방식에서는 상위 프레임과 하위 프레임을 따로 정의하여 대역폭 할당과 전송 지연보장을 분리할 수 있기 때문에 종단간 전송계층에서 QoS를 보장할 수 있는 단순화한 큐 서비스 구현을 제안한다.

큐 서비스 방식에서 연결별 큐를 사용하지 않고 라운드 로빈 큐와 CBR 트래픽 큐만으로 구성하여 구조를 간단하게 하므로 구현이 용이하도록 하였으며, 지연 EDD 알고리즘과 유사한 방법으로 마감 시간을 이용 다양한 전송지연을 허용할 수 있도록 하였다.

제한한 큐 서비스 방식은 실시간 트래픽과 비실시간 트래픽을 나누어 저장하기 때문에 연결별 트래픽 조절기 없이도 지연 지터 특성을 개선할 수 있었다. 또한 큐 서비스 방식은 기존의 큐 서비스 방식과 달리 실시간 서비스뿐만 아니라 비 실시간 서비스의 대역폭을 보장해 줄 수 있다.

종단간 지연을 보장하기 위해 연결은 트래픽 특성을 요구하며, 망에서는 지연 한계를 유도하기 위해 버스트성있는 트래픽을 기술하는 모델 ($X_{min}, X_{avg}, I, S_{max}$)을 이용하였으며, 최고 전송 율과 일정 기간의 평균 전송 율 등의 두 가지 파라미터로 버스트성을 나타내는 VBR 트래픽과 같은 경우에 적합한 모델로써 트래픽 특성을 이용하였다.

제안된 큐 서비스 방식을 이용하여 연결 수락 제어는 트래픽 왜곡의 영향을 계산할 수 없는 트래픽 왜곡의 영향이 실제로 존재한다는 가정 하에 트래픽 왜곡에 대한 영향에 종단간 지연을 보장함으로써 연결 수락 가능범위를 다양한 프레임의 길이에 대해 나타내었으며, 프레임의 길이가 클수록 제안된 큐 서비스 방식의 연결 수락 범위가 지연 EDD방식과 비교해서 많은 차이를 보이지만 프레임의 길이를 점차 줄여갈수록 연결 수락 가능 범위가 지연 EDD 방식에 근접하여 T_f 의 길이가 1 ms이면 지연 EDD방식의 그래프와 거의 일치하는 것을 알 수 있으며, T_f 의 길이가 0.5 ms이면

지연 EDD 방식과 같은 연결 수락 가능 범위의 차이가 없음을 시뮬레이션 결과를 통해 알 수 있었다. 제안하는 큐 서비스 방식의 성능은 전송 프레임의 길이에 좌우되고, 전송 프레임의 길이가 짧을수록 제안하는 서비스 방식의 연결 수락 가능 범위는 지연 EDD방식에 접근하게 된다. 또한 시뮬레이션 결과에서 트래픽 왜곡의 영향이 최악의 경우를 고려하여 모든 채널이 일정한 비율로 셀을 생성하는 경우에 트래픽 왜곡의 영향이 적음을 알 수 있었고 제시된 지연 예측도 정확하게 지연 EDD방식과 일치함을 알 수 있었다.

참고 문헌

- [1] C. Li, Y. Ofek, A. Segall, and K. Sohraby, "Pseudo-isochronous Cell Switching in ATM Networks," in Proc. IEEE infocom'94, pp.428-437, 1994.
- [2] C. Parris, G. Ventre and H. Zhang, "Graceful Adaptation of Guaranteed Performance Service Connections," In Proceedings of IEEE GLOBECOM'93, Houston, TX, November, 1993.
- [3] H. Kanakia, P. Mishra and A. Reibman, "An Adaptive Congestion Control Scheme for Real-time Packet Video Transport," In Proc.of ACM SIGCOMM'93, San Francisco, California, pp.20-31, Sept. 1993.
- [4] H. Kroner, "Comparative Performance Study of Space Priority Mechanisms for ATM Networks," In Proc. of the IEEE INFOCOM, pp.1136-1143, June 1990.
- [5] J. C. R. Bennett and H. Zhang, "Hierarchical Packet Fair Queueing Algorithm," Proc.of SIGCOMM '96, August 1996.
- [6] L. Zhang, "Virtual Clock: a New Traffic Control Algorithm for Packet Switching," In Proc. ACM Sigcomm'90, pp.19-29, 1990.
- [7] M. Juiano et al. White Paper, "ATM Traffic Management and Congestion Control Fore System," Inc., February 1994.
- [8] R. Chipalkatti, J. F. Kurose and D. Towsley,

"Scheduling Policies for Real-time and Non-real-time Traffic in a Statistical Multiplexer," In Proc. of IEEE INFOCOM, April 1989.

- [9] H. Zhang and S. Keshav, "Comparison of Rate-based Service Disciplines," In Proc. ACM Sigcomm'91, pp.113-121, 1991.
- [11] Ian R. Philp and Jane W. S. Liu, "End-to-End Scheduling in Real-time Packet Switched Networks," In Proc. International Conference on Network Protocol, Columbus, Ohio, pp.23-30, October 1996.
- [12] J. C. R. Bennett and H. Zhang, "WF2Q : Worst-case Fair Weighted Fair Queueing," In Proc. IEEE Infocom'96, pp.120-128, 1996.
- [13] S. Jamaloddin Golestani, "A Stop-and-Go Queueing Framework for Congestion Management" In SIGCOMM' 90Symposium, Communication Architecture & Protocols, Philadelphia Pennsylvania, Sept. 1990.



나상동(Sang-Dong Ra)

1968년 조선대학교 전기공학과 졸업(공학사)

1980년 건국대학교 대학원 졸업(공학석사)

1995년 원광대학교 대학원 졸업(공학박사)

1995년~1996년 Dept. of Electrical & Computer Eng. Univ. of California Irvine 연구교수

1998년 조선대학교 전자계산소 소장 역임

1973년~현재 조선대학교 컴퓨터공학부 교수

2001년-2002년 Dept. of Electrical & Computer Eng. Univ. of California Irvine 연구교수

※ 관심분야 : 실시간 통신, 디지털 통신망, 데이터 및 이동통신, TMO, 적응 신호처리 등임.

저 자 소 개

김광준(Gwang-Jun Kim)

1993년 조선대학교 컴퓨터공학과 졸업(공학사)

1995년 조선대학교 대학원 컴퓨터공학과 졸업(공학석사)

2000년 조선대학교 대학원 컴퓨터공학과 졸업(공학박사)

2000년~2001년 Dept. of Electrical & Computer Eng. Univ. of California Irvine Postdoctoral

2001년~현재 베네테크(주) 연구실장

※ 관심분야 : ATM망, 실시간 통신 프로그래밍, 이동통신