

논문 2004-41TC-1-3

인터넷 및 방송서비스의 QoS보장을 위한 10Gbps급 스트리밍 패킷 스케줄러 구조 및 제어방법

(A architecture and control method of Streaming Packet Scheduler at 10Gbps for Guaranteed QoS of Internet and Broadcasting Services)

김 광 옥*, 박 완 기*, 최 병 철*, 곽 동 용*

(Kwang-Ok Kim, Wan-Ki Park, Byeoun-Chul Choi, and Dong-Yong Kwak)

요 약

본 논문에서는 초고속 패킷 스위치 네트워크에서 VoD나 HDTV, VoIP같은 고품질 스트리밍 서비스의 QoS를 보장하는 패킷 스케줄러의 구조 및 제어방법을 제시한다. 스트리밍 서비스는 버스트 데이터 응용서비스보다 더욱 엄격한 QoS(jitter, delay, packet loss)보장을 요구한다. 또한 스트리밍 서비스는 다른 플로우들의 동작에 상관없이 끊임 없는 서비스를 제공하기 위해 각 플로우별로 최소 대역 보장과 종단간 지연조건을 보장해야 한다. 이들 요구조건들을 만족하기 위해, 패킷 스케줄러는 플로우들이 다른 플로우의 영향을 받지 않도록 분리하고, 각 플로우들에게 종단간 지연 보장을 제공해야 한다. 그리고 각 플로우들에게 요구되는 최소 대역폭을 할당해야한다. 지금까지 많은 벤더들이 10Gbps급 트래픽 관리기 칩을 개발하였지만 대부분 칩들은 고품질 스트리밍 서비스를 지원하지 못하는 단점이 있다. 따라서 본 논문에서는 상용 TM칩들의 단점 및 스트리밍 서비스의 트래픽 특성을 조사하고, 제안한 패킷 스케줄러의 하드웨어 구조를 제시한다. 그리고 마지막으로 제안한 스케줄러의 시뮬레이션 결과를 분석하였다.

Abstract

This paper presents architecture and control method of packet scheduler to guarantee QoS of high quality streaming services in high-speed packet-switched networks. Since streaming services need far more stringent QoS requirements than the typical sort of burst data applications, they should be guaranteed minimum bandwidth and end-to-end delay bound to each flow, regardless of the behavior of other flows. To meet these requirements, a packet scheduler isolate a flow from the undesirable effects of other flows and provides end-to-end delay guarantees for individual flow and divides stringently the available link bandwidth among flows sharing the link. Until now, many vendors are developing traffic management chips running at 10Gbps, but most of chips have drawbacks to support high quality streaming services. In this paper, we investigate the drawbacks of commercial TM chips and traffic characteristic of streaming services and present implementation frameworks of the proposed packet scheduler. Finally, we analyze the simulation results of the proposed scheduler.

Keywords: Delay Guaranteed Packet Scheduler, QoS, Traffic Manager, Per-Flow Shaping

1. 서 론

최근 통신 서비스 사업자들은 인터넷 및 전화, 디지털 방

송 등 Triple Play서비스를 제공하기 위해 다양한 콘텐츠의 개발 및 가입자 액세스 망에서의 QoS보장을 중요한 이슈로 생각한다. 특히 가입자 액세스 망에서 액세스 장비는 고품질 스트리밍 방송서비스를 제공하기 위해 엄격한 QoS를 보장해야하며, 동시에 IP 단일 망을 통해 인터넷 서비스도 제공해야 된다. 또한 수십 ~ 수백 채널의

* 정회원, 한국전자통신연구원
(Electronics and Telecommunications Research Institute)
접수일자 : 2003년11월8일, 수정완료일 : 2003년12월30일

SDTV(2.5Mbps)나 DVD(4.2Mbps), HDTV(20M bps) 등의 서비스를 가입자들에게 제공하기 위해 광대역 시스템이 요구된다.

가입자 액세스 장비에서 인터넷 서비스는 실 시간성을 요구하지 않기 때문에 안정적인 Throughput만 제공되면 큰 문제는 없다. 그러나 VoD나 고품질 HDTV와 같은 스트리밍 서비스는 실 시간성을 요구하며, 방송을 보는 내내 최소의 Delay, Jitter, Loss를 보장하여 끊김 없는 화질을 제공해야한다. 또한 이웃 가입자의 채널변경에 따른 화질 저하를 제거하기 위해 isolation 기능이 제공되어야 한다.

현재 EZChip, Agere, ZettaCom, AMCC등 상용 벤더들은 가입자망에서 QoS보장을 제공하기 위해 10Gbps급 트래픽 관리기를 개발 중이거나 이미 상용화를 시작하였다. 그러나 상용 트래픽 관리기는 WRR이나 WFQ 계열의 대역보장형 패킷 스케줄러를 통해 인터넷 서비스에 대한 QoS만을 제공할 뿐 Flow별 Shaper나 Queuing/Scheduling 구조를 사용하지 않기 때문에 고품질 스트리밍 서비스에서 요구되는 지연이나 지터 등을 엄격하게 보장하지 못한다.

따라서 본 논문에서는 인터넷 버스트 서비스와 고품질 스트리밍 방송서비스에 대해 QoS를 하나의 단일망을 통해 동시에 제공하는 지터 보장형 10Gbps급 스트리밍 패킷 스케줄러 구조를 제안하였다. 스트리밍 패킷 스케줄러는 Heap Manager기반의 EDF (Earliest Deadline First)스케줄러를 이용하여 인터넷 서비스에 대한 QoS를 보장하고, Per-Flow Shaper기반의 EDF 스케줄러를 이용하여 가입자별 스트리밍 서비스에 대한 지연, 지터, 셀 지연변이, 대역보장 등 QoS를 제공한다.

본 논문의 구성은 다음과 같다. II장에서는 방송서비스용 스케줄러 고려사항 및 상용 트래픽 관리기에서 패킷 스케줄러의 구조를 설명한다. 그리고 상용 패킷 스케줄러의 문제점 분석한다. 또한 통신 및 방송서비스를 제공하기 위한 가입자 정합카드 구조 및 동작에 대해서 간략히 논의한다. IV장에서는 본 논문에서 제안한 지터보장형 10Gbps급 스트리밍 패킷 스케줄러 구조 및 제어 방법에 대해서 논의하고, V장에서는 시뮬레이션을 통한 제안한 패킷 스케줄러에 대한 성능 평가를 수행한다. 그리고 마지막으로 결론을 맺는다.

II. 통신방송 패킷 스케줄러 고려사항 및 상용 패킷 스케줄러 구조와 문제점

2.1. 통신방송서비스를 위한 패킷 스케줄러 고려사항

통신방송융합서비스에서 패킷 스케줄러는 기존 인터넷 서비스에 대한 QoS보장 및 방송서비스에 대한 보다 엄격한 QoS보장을 수행해야한다. 인터넷 서비스에서는 QoS보장을 위해 버스트 트래픽에 대한 대역보장 및 안정적인 Throughput을 제공해야하며, 방송서비스에서는 QoS보장을 위해 스트리밍 트래픽에 대한 지속적인 Delay나 Jitter, Loss보장 및 절대적인 대역보장을 제공해야한다. 특히 스트리밍 패킷 스케줄러는 끊김 없는 방송 서비스를 제공하고, 가입자에게 일정한 전송품질을 제공해야한다.

따라서 스트리밍 패킷 스케줄러는 가입자별로 엄격한 대역 제한을 위한 Per-Flow Shaping기능이 반드시 요구되며, 방송 채널 Flow에 대한 대역폭을 Fair Queuing이 아닌 망의 혼잡 상태에 상관없이 대역을 절대적으로 보장해주는 Strict Queuing 기능이 필요하다. 스트리밍 패킷 스케줄러의 주요 고려사항은 다음과 같다.

- 방송서비스를 위해 가입자에게 절대적인 대역보장이 고려.
- 인터넷 서비스를 위해 안정적인 Throughput이 고려.
- 끊김 없는 방송을 위해 가입자별 Shaper고려.
- 지연 민감 트래픽에 대해 최소의 시 지연이 고려.
- 지연 기반의 우선순위 정렬기능이 고려.

2.2. 상용 패킷 스케줄러 구조와 문제점

현재 상용 트래픽 관리기에서 사용되는 패킷 스케줄러는 할당된 대역(Bandwidth) 내에서 각 가입자에게 인터넷 데이터 서비스를 효율적으로 제공하도록 서비스 클래스별로 최대한(Best effort) 대역 보장을 제공하기 위해 노력한다. 따라서 상용 트래픽 관리기는 망의 혼잡 상태가 발생하면 요구된 대역 보장을 제공하지 못한다. 상용 트래픽 관리기의 패킷 스케줄러는 대역 보장을 위해 WRR(Weight Round Robin)이나, DRR(Deficit Round Robin), WFQ(Weight Fair Queuing)등의 스케줄링 알고리즘을 사용한다. WRR이나 DRR은 대역 보장형 스케줄링 알고리즘으로서, Delay Bound를 보장하지 못한다. WFQ는 요구되는 Delay Bound를 보장하나 가상시간 계산 및 유지가 너무 복잡하여 고속 망 환경에 부적합하며, 지터를 보장하지 못하다. 또한 WFQ는 Rate기반으로 지연 조건을 만족하기 때문에, VoIP(64Kbps~5.3Kbps)같이 Service rate이 작고 지연에 민감한 서비스에는 작은

지연 조건을 만족하기 위해 rate을 높여 많은 대역을 할당하는 결과를 초래한다. 상용 TM 패킷 스케줄러의 일반적인 내부 구조는 <그림 1>과 같다.^[3]

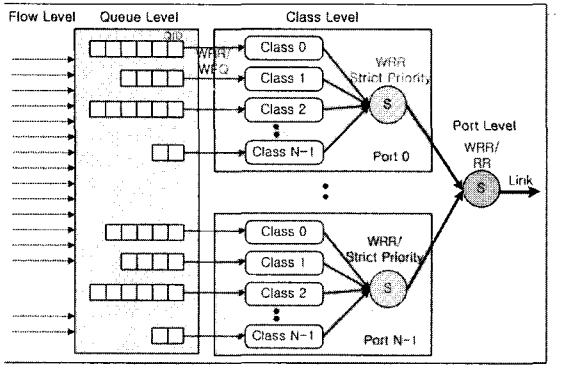


그림 1. 상용 트래픽 관리기의 패킷 스케줄러 구조.
Fig. 1. A architecture of Packet Scheduler in the commercial traffic manager.

상용 트래픽 관리기의 패킷 스케줄러는 크게 플로우 레벨, 큐 레벨, 클래스 레벨, 포트 레벨로 구성된다. 스케줄러는 매 셀 타임마다 포트->클래스->큐->플로우 중 하나를 선택하여 서비스를 수행한다. 많은 플로우들은 특정 큐 속에 집합되므로 각 개별 플로우에 대한 대역보장이나 지연, 지터를 보장하지 못한다. 큐 레벨 스케줄러는 WRR이나 WFQ을 사용하여 각 큐에 대한 대역 보장을 제공한다. 클래스 레벨 스케줄러는 Strict Priority 스케줄러를 사용하여 우선순위에 따라 서비스를 수행한다. 클래스 그룹은 최대 128개 까지 구성 가능하지만, 스위치와의 VoQ서비스를 위해 4 클래스를 사용한다. 포트레벨 스케줄러는 WRR, RR을 사용하여 순차적으로 각 포트에 대한 서비스를 수행하여 Fairness를 제공한다.^{[7][8][9]}

따라서 현재 상용 트래픽 관리기는 인터넷 서비스에 대한 대역보장은 가능하나, 스트리밍 방송서비스에서 요구되는 가입자별 QoS를 보장하지 못한다. 또한 Per-Flow Shaper를 사용하지 않으므로 가입자들에게 지터를 제공하지 못한다. 현재 주요 벤더들의 상용 트래픽 관리기는 스트리밍 방송서비스를 가입자에게 제공할 때 <표 1>과 같은 문제점이 있다.

따라서 본 논문에서는 각 가입자에 끊임 없는 스트림이 방송서비스와 인터넷 서비스를 제공할 수 있도록 Per-Flow Queuing/Scheduling/Shaping기능을 갖는 스트리밍 패킷 스케줄러의 구조 및 트래픽 처리방안을 제안하였다. 또한 현재 한 노드에서 시 지연 시간이 최적으로 증명된 EDF스케줄러

를 이용하여 delay bound 보장을 제공하였다.

표 1. 상용 트래픽 관리기의 문제점
Table 1. A defect of common traffic manager

상용TM	방송서비스 적용 시 문제점
Agere -TM10	-포트별 셰이핑 기능을 수행하므로 가입자별 CDV나 Jitter보장이 안됨.(치명적) -16K Queue에 2M Flow들이 aggregate되어 가입자별 엄격한 대역보장이 제공 안됨. -Flow당 최소 1.5Mbps이므로 VoIP전송 시 대역낭비 초래.
EZchip -QX-1	-셰이핑 기능이 없어 가입자별 Jitter보장이 안되고, delay bound 보장 안됨.(치명적) -1K Queue에 4M Flow들이 aggregate되어 가입자별 엄격한 대역보장이 제공 안됨.
ZettaCom -ZTM552	-4K Queue에 1M Flow들이 aggregate되어 가입자별 엄격한 대역보장이 제공 안됨.

2.3. 가입자 액세스 장비의 라인카드 구조

가입자 액세스 장비의 라인카드 구조는 <그림 2>와 같다. 라인카드는 가입자들에게 광대역 서비스를 제공하기 위해 라인카드 당 10Gbps이상의 트래픽을 처리를 제공한다. 라인카드는 전기신호를 패킷으로 변환하는 Framer와 플로우 구분 및 다양한 패킷 처리를 수행하는 네트워크 프로세서 그리고 패킷에 대한 QoS를 보장하는 트래픽 관리기로 구성된다. 시스템 구조에 따라 추가로 상용 트래픽 관리기나 큐잉 엔진을 사용한다. Framer는 1Gbps 이더넷 10포트 또는 10Gbps 이더넷 1포트로 구성되며, 네트워크 프로세서와 SPI-4.2나 XAUI 인터페이스를 통해 정합된다. 네트워크 프로세서는 트래픽 분류 기능 및 목적지 포트로 패킷을 전송하는 기능을 수행하며, L4 또는 L7까지 룩업 하여 인터넷 서비스 및 방송 서비스 플로우들을 구분한다.

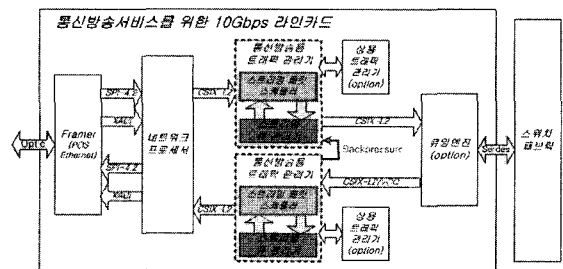


그림 2... 가입자 액세스 장비의 라인카드 구조.
Fig. 2... A architecture of line card in the Subscriber Access Equipment.

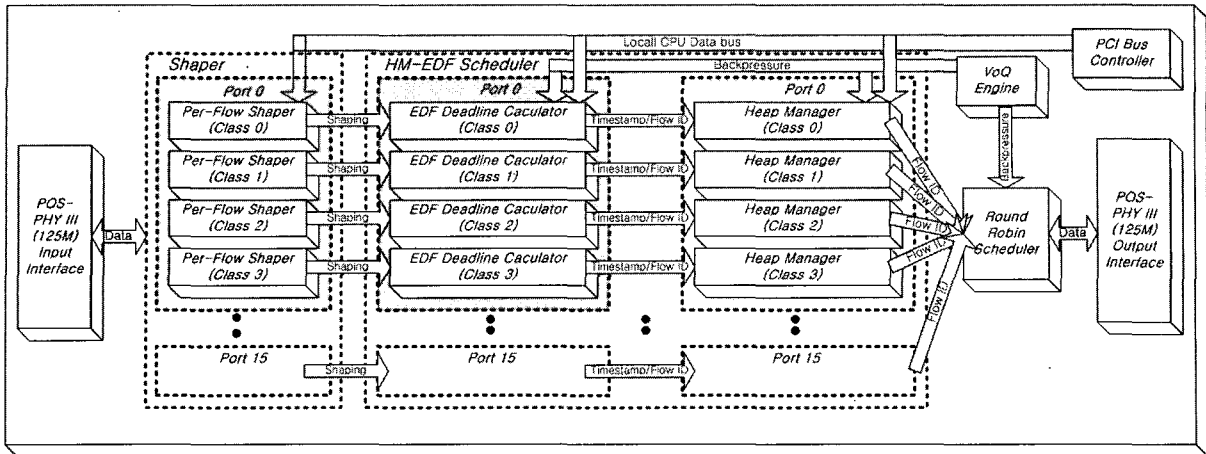


그림 3. 10Gbps 스트리밍 패킷 스케줄러 구조.
Fig. 3. A architecture of Streaming Packet Scheduler at 10Gbps.

QoS보장을 제공하는 통신방송용 트래픽 관리기는 Ingress/Egress 각각에 위치하며, 스트리밍 패킷 스케줄러와 큐 관리기로 구성된다. 스트리밍 패킷 스케줄러는 인터넷 및 방송서비스 플로우에 따라 요구되는 QoS 보장을 제공한다. 상용 트래픽 관리기는 통신방송용 트래픽 관리기를 방송서비스 전용으로 사용할 경우 인터넷 서비스에 대한 QoS를 제공한다.

III. 제안한 10Gbps급 스트리밍 패킷 스케줄러 구조 및 제어방법

3.1. 10Gbps 스트리밍 패킷 스케줄러 구조.

본 논문에서는 인터넷 및 방송서비스에 대한 엄격한 QoS를 제공하는 10Gbps급 스트리밍 패킷 스케줄러 구조를 제안하였다. 제안한 스트리밍 패킷 스케줄러는 <그림 3>같이 입출력 인터페이스, 가입자별 Shaper, HM-EDF스케줄러, VoQ엔진, 포트 스케줄러, PCI인터페이스로 구성된다. 주요 모듈의 기능을 다음과 같다.

① **입출력 인터페이스 모듈** : POS-PHY III 인터페이스를 사용하며 큐 관리기와 인터페이스 된다. 입력 인터페이스는 큐 관리기로부터 Enqueue Flow ID, Destination Port, Service Rate, Service Priority의 스케줄링 정보를 받고 출력 인터페이스는 큐 관리기로 Dequeue Flow ID 스케줄링 정보를 전송한다.

② **Shaper 모듈** : 각 Flow마다 선택 플래그에 따라 Shaping 기능을 수행한다. Shaper는 PCR파라미터를

이용해 각 Flow들의 Maximum 대역 제한 기능을 수행하며, 셀 들 간의 서비스 간격을 일정하게 유지하여 버스트나 지터 등을 제거한다. 방송서비스 가입자 Flow는 Shaper에서 Shaping을 수행 후 스케줄러로 전송되며, 통신서비스 가입자 Flow는 Shaper를 거치지 않고 바로 스케줄러로 전송된다.

③ **HM-EDF 스케줄러 모듈** : EDF스케줄러는 Shaping된 Flow나 Non-Shaping된 Flow들의 셀 전송 지연 조건을 기반으로 Deadline을 만들어 가장 Deadline이 적은 셀을 전송함으로써 Minimum 대역 보장을 제공한다. HM-EDF스케줄러는 가장 Deadline이 적은 Flow를 찾기 위해 Heap Manager를 이용한다.

Heap Manager는 10Gbps에서 Deadline, 우선순위 정렬을 제공하며, 파이프라인 하드웨어 구현이 용이하며 확장성이 뛰어나다. Deadline은 PCR파라미터를 이용하여 계산한다.

④ **포트 스케줄러 모듈** : 포트 스케줄러는 라운드 로빈 알고리즘을 사용하며, 활성화된 포트들 사이에서 순서적으로 서비스 기회를 제공한다.

⑤ **VoQ엔진 모듈** : VoQ엔진 모듈은 스위치 패브릭의 클래스 또는 포트의 Backpressure상태에 따라 스위치 패브릭으로의 패킷 흐름 제어를 수행한다.

즉, 스트리밍 패킷 스케줄러는 인터넷 서비스의 트래픽은 EDF스케줄러를 통해 delay bound를 최소화하여 대역보장을 제공하고, 방송 서비스의 트래픽은 Shaper와 EDF를 이용해 end-to-end delay bound나 지터보장을 제공한다.

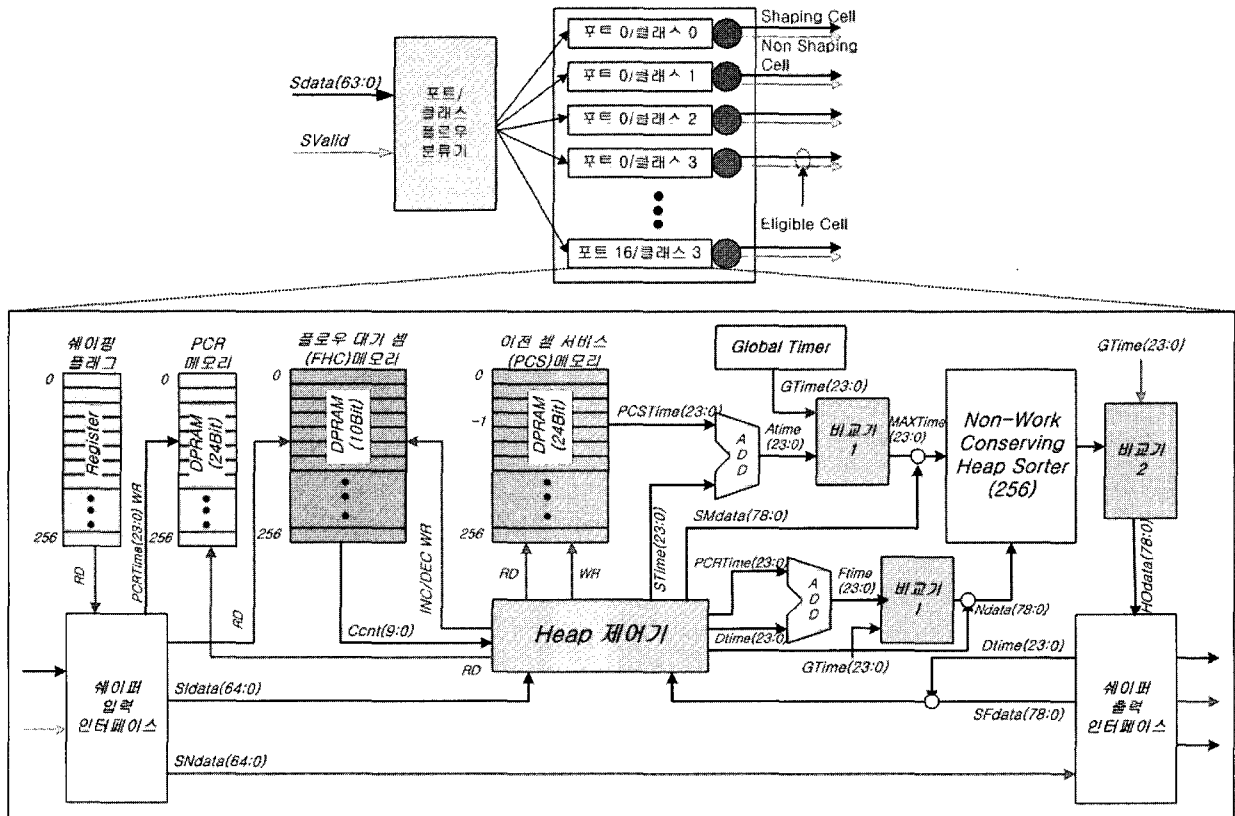


그림 4. Per-Flow Shaper 모듈 구조.
Fig. 4. A architecture of per-flow shaper.

4.1. Shaper 모듈 구조 및 동작원리

Shaper 모듈은 스트리밍 방송서비스 Flow들에 대해 일정한 속도로 패킷을 전송하여 끊임 없는 방송서비스 및 각 가입자들에게 엄격한 최대 대역제한 기능을 제공한다. Shaper 모듈의 구조는 <그림 4>와 같으며 RC-EDF 스케줄러 기능을 수행한다, Shaper는 입력되는 Flow들을 포트와 클래스로 구분하는 포트/클래스 Flow분류기와 각 클래스 모듈 안에서 Flow들에게 Per-Flow Shaping 기능을 제공하는 클래스별 Shaper 모듈로 구성된다.

각 클래스별 Shaper 모듈은 셰이퍼 입력력 인터페이스, 각 Flow의 Shaping 설정 상태를 나타내는 레지스터, 각 Flow의 이전 셀의 서비스 시간을 저장하는 메모리, Shaping Rate과 셀 도착 시간을 비교하는 비교기, 이전 셀 서비스 시간과 Flow의 서비스 시간을 더하는 연산기, Eligible Time 조건일 때 Shaping Flow를 서비스하는 Non-Work Conserving Heap Sorter, 각 Flow별로 현재 대기 중인 셀 수를 나타내는 대기 셀 운터 메모리, Shaper 모듈들을 제어하는

Heap 제어기로 구성된다. Shaper 모듈은 각 Flow들에 할당된 Service Rate (bps)을 다음 식 (1)에 의해 Service Time(sec) 변환하여 하드웨어 클럭 카운터에 따라 정확한 시간에 Shaping 서비스를 수행하게 한다.^{[1][2][4]}

$$Service\ Time(sec) =$$

$$\frac{Cell\ Segment\ Size(bit)}{Cell\ Time(sec) \times Service\ Rate(bps)} \quad (1)$$

식 (1)에서 Cell Time은 하나의 패킷의 세그먼트를 16Gbps(250MHz*64Bit) 속도로 전송하는데 걸리는 시간으로 다음 식 (2)에 의해 구하며, 본 논문에서는 32ns가 된다. Service Rate은 각 Flow들의 PCR값을 이용한다.

$$Cell\ Time = \frac{Scheduler\ 용량}{Bit수 / Clock\ Cycle당} \quad (2)$$

Shaper 모듈은 계산된 Service Time을 이용하여 각 Flow들이 일정한 셰이핑 속도로 서비스하도록 다음 식 (3)에 의해 Eligible Time을 계산한다. 계산된

Eligible Time은 스케줄링 정보와 함께 Non-Work Conserving Heap Sorter에 입력되며, 최 상위 노드의 값이 Global Time보다 크거나 같으면 유효한 셀이 되므로 Flow의 스케줄링 정보를 HM-EDF 스케줄러로 전송한다. 그렇지 않으면 Heap Sorter에 대기한다. Heap Sorter는 Eligible Time이 제일 적은 값을 최 상위 노드에 위치하므로 Heap Sorter 서브 노드에 대기 중인 모든 Flow들도 시간에 따라 Per-Flow Shaping 기능을 수행한다. 만약 Heap Sorter는 동일시간에 동일한 Eligible Time을 갖는 셀이 존재하면, Deadline이 적은 셀을 먼저 스케줄러로 전송한다. 따라서 모든 Flow들은 요구된 셰이핑 간격으로 일정하게 셀을 전송하게 된다.

$$\begin{aligned} Eligible\ Time(F_k^i) = \\ MAX\{(F_k^{i-1} + Service\ Time), G(A_k^i)\} \end{aligned} \quad (3)$$

F_k^{i-1} : K번째 노드에서 Flow i가 Shaper에서 서비스된 시간

$G(A_k^i)$: K번째 노드에서 Flow i가 Shaper에 도착한 시간..

Global Time은 셀 타임(32ns) 단위로 증가된다. 따라서 Global Time +1 증가는 하나의 셀이 전송되는 것과 같다. Shaper모듈의 동작은 원리는 다음과 같다.

① 입력 인터페이스로 스케줄링 정보와 Valid 신호가 입력되면 목적지 주소 필드에 따라 해당 클래스 셰이퍼로 데이터를 전송하고, 클래스 필드를 이용해 Flow가 셰이핑 상태인지 또는 현재 대기 중인 셀이 있는지 체크한다. 그리고 입력된 Service Time(PCR) 값은 PCR메모리에 저장된다.

② Flow가 Shaping Flow이면 Heap제어기로 데이터를 전송하고, 그렇지 않으면 셰이퍼 출력 인터페이스로 전송한다.

③ Heap제어기는 입력된 Flow에 대기 중인 셀이 존재하면 Heap Sorter에 동일한 Flow의 셀이 대기 중이므로 셀 카운터를 +1증가시킨다. 그렇지 않으면 이 Flow에서 이전 셀 전송 시간을 체크하고 입력된 데이터의 Service Time 가산하고 비교기를 통해 Global Time과 비교하여 큰 값을 스케줄링 정보와 함께 Non-Work Conserving Heap Sorter로 전송한다. 각 Flow는 서비스 전송을 대기 중인 셀 카운터를 가지고 있다.

④ Non-Work Conserving Heap Sorter는 매 셀 타임마다 최 상위 노드의 Eligible Time과 Global Time을 비교하여 크거나 같으면 Eligible 조건이 되므로 셰이퍼 출력 인터페이스로 스케줄링 정보를 전송한다. 그리고 대기 중인 다음 셀을 전송하기 위해 전송 시간과 함께 스케줄링 정보를 다시 Heap 제어기로 전송한다.

⑤ Heap 제어기는 대기 중인 셀이 있는지 체크하여 대기 중인 셀이 있으면 다시 Non-Work Conserving Heap Sorter로 스케줄링 정보를 전송하고 셀 카운터를 -1 한다.

4.2. HM-EDF 패킷 스케줄러 구조 및 동작원리

EDF 알고리즘은 프로세서에서 실시간 태스크 스케줄링 알고리즘으로 사용되어오다 최근 패킷 스케줄링 알고리즘에 사용되었다. 현재 EDF 패킷 스케줄러가 한 노드에서 최적의 패킷 지연 시간을 제공한다고 증명되었다. 또한 Shaper와 함께 제공됨으로서 종단간 지연 조건을 보장한다. EDF 스케줄링 알고리즘은 패킷이 입력된 arrival time과 패킷의 delay bound를 이용해 deadline을 계산하고, minimum deadline 순으로 서비스를 수행한다. 즉 EDF는 WFQ와 달리 delay를 기반으로 우선순위를 제공한다. EDF는 선점방식을 사용한다.^[5]

본 논문에서는 Heap Manager를 이용해 deadline 우선순위 정렬을 제공한다. HM-EDF 패킷 스케줄러는 다음 식 (4)에 의해 deadline을 계산한다.

$$Deadline_k^i = A_k^i(t) + d_k^i \quad (4)$$

$A_k^i(t)$: K번째 노드에서 Flow i의 패킷 도착시간.

$$d_k^i = \frac{Packet\ Length}{g_i} + \frac{Packet\ Length}{R^k}$$

g_i : Flow i에 할당된 속도(PCR)

R^k : K번째 라우터에 할당된 링크 속도

Packet Length : 고정된 세그먼트 셀 길이 (= '1')

스트리밍 패킷 스케줄러는 다음 식 (5)에 의해 end-to-end delay bound를 제공한다..

$$D_i = \sum_{k=1}^{K_i} S(E_i(t)) + \sum_{k=1}^{K_i} d_k^i \quad (5)$$

HM-EDF 스케줄러는 Shaping된 방송서비스 Flow

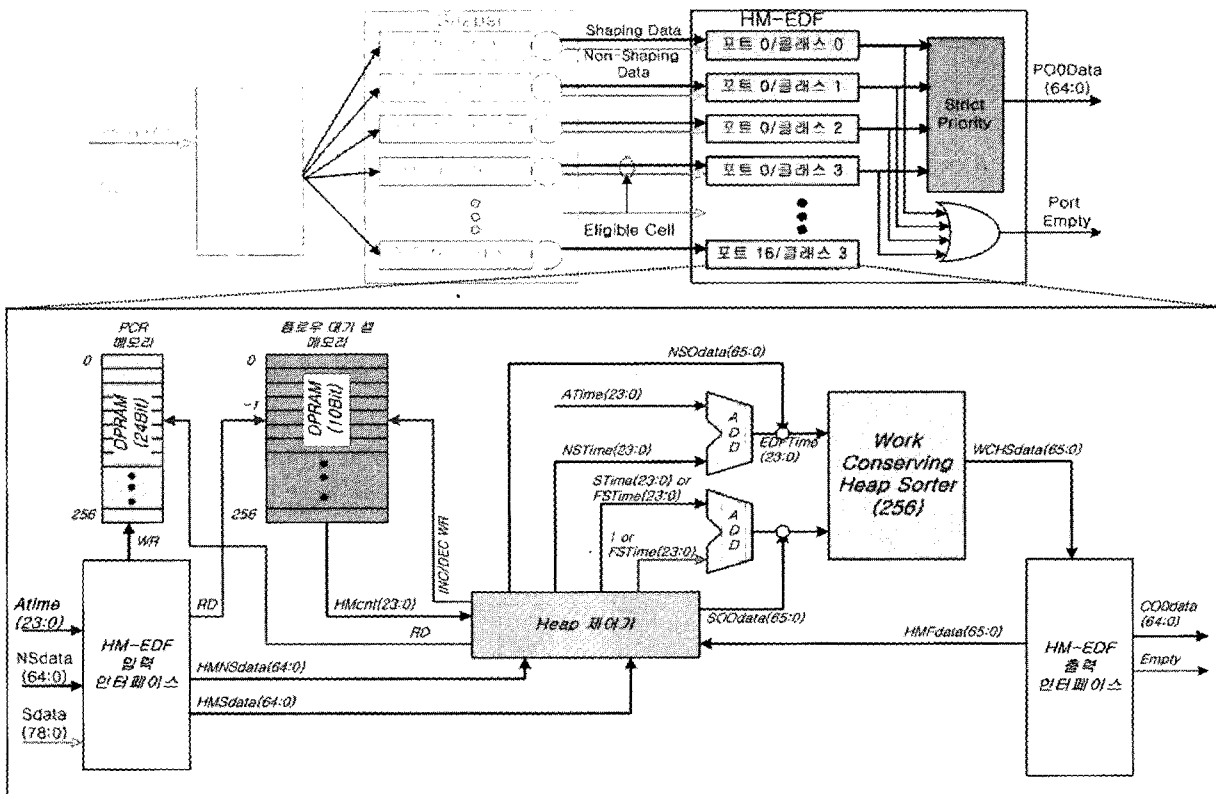


그림 5. HM-EDF 스케줄러 모듈구조.
Fig. 5. A architecture of HM-EDF Scheduler.

와 Non-Shaping된 인터넷 서비스 Flow가 입력된다.

Shaping된 데이터는 Eligible Time 필드를 제거하고 식 (4)와같이 deadline을 구해 Work Conserving Heap Sorter에 입력한다. Shaping되지 않은 데이터도 마찬가지로 식 (4)와같이 deadline을 계산하여 Work Conserving Heap Sorter에 입력한다.

Heap Sorter는 매 셀 타임마다 deadline을 우선순위 별로 정렬하고, 최 상위 노드의 스케줄링 정보를 전송한다. Global Time은 Shaper의 Global Time과 동일한 값을 사용한다. HM-EDF 스케줄러 모듈은 <그림 5>와 같이 Deadline 가산기, Heap제어기, Heap Sorter, PCR메모리, 플로우 대기 셀 메모리, Strict Priority 스케줄러로 구성된다. HM-EDF 스케줄러 동작원리는 다음과 같다.

- ① 입력 인터페이스를 통해 셰이핑 된 스케줄링 정보와 셰이핑 되지 않은 스케줄링 정보가 입력된다. 또한 Global Time이 입력된다. 입력된 데이터는 Heap제어기로 전송하고 동시에 입력된 Flow에 대기 중인 셀이 있는지 체크한다.
- ② 대기 중인 셀이 있으면 셀 카운터를 +1 증가시키

고, 그렇지 않으면 Shaping된 데이터는 Eligible Time 필드를 제거하고 Heap Sorter로 전송된다. Shaping되지 않은 데이터들은 service time과 delay bound를 계산하여 Heap Sorter로 전송한다.

- ③ Work Conserving Heap Sorter는 매 셀 타임마다 최 상위 노드의 스케줄링 정보를 출력 인터페이스로 전송하고, 출력된 스케줄링 정보는 대기 중인 다음 셀을 전송하기 위해 Heap 제어기로 전송한다.
- ④ Heap제어기는 피드백 된 플로우에 대기 중인 셀이 있는지 체크하고, 셀이 존재하면 다시 Heap Sorter에 입력하고 셀 카운터를 -1 감소한다.
- ⑤ 동시에 Heap제어기에 2개의 스케줄링 정보가 입력되고 1개의 스케줄링 정보가 피드백 될 때, 만약 피드백 된 스케줄링 정보가 Shaping된 Flow이면 입력된 Shaping Flow의 셀 카운터를 +1 증가시키고 대기 중인 스케줄링 정보를 Heap Sorter에 저장한다. Shaping되지 않은 Flow이면 입력된 Non-Shaping Flow의 셀 카운터를 +1 증가시키고 대기 중인 스케줄링 정보를 Heap Sorter에 저장 한다.

4.3. Heap Manager 모듈

스트리밍 패킷 스케줄러에서 Heap Manager는 각 플로우들의 10Gbps속도로 우선순위 정렬을 제공한다. 스트리밍 스케줄러에서 Heap Manager는 Shaper모듈과 HM-EDF 스케줄러 모듈에 각각 사용된다. 각 클래스별로 Heap Manager모듈은 총 9레벨 / 511개의 노드로 구성된다.^[6]

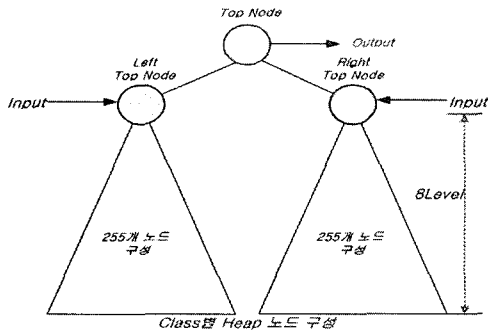


그림 6. Heap Manager 전체 구조.
Fig. 6. A top architecture of Heap Manager.

Heap Manager는 새로 입력된 플로우와 대기 중인 플로우를 동시에 받아들이기 위해 <그림 6>과 같이 구성된다. Heap Manager에서 Heap 정렬은 왼쪽 노드부터 채워진다. 2개의 데이터가 동시에 입력될 경우 서브 노드 양쪽에 입력되고, 출력은 최 상위 노드에서 출력된다. 각 서브 노드의 내부구조는 <그림 7>과 같다. 맨 하위레벨 노드를 제외한 각 노드는 자신을 포함한 하위 노드 Non-empty 수를 나타내는 카운터 정보가 존재한다. 이 카운터 정보를 이용해 유효한 empty노드를 찾는다. 최 상위 노드의 카운터가 서브 노드의 개수와 동일하면 왼쪽 노드들이 모두 채워진 것이므로 오른쪽 루트로 데이터를 보낸다. 각 노드들의 어드레스는 왼쪽 루트는 '0', 오른쪽 루트는 '1'로 매핑한다.

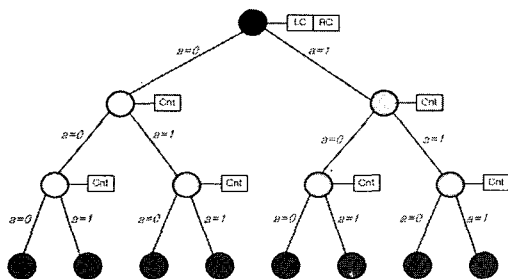


그림 7. Heap Manager 서브트리 구조.
Fig. 7. A architecture of Sub tree in the HM.

Heap Manager에서 우선순위 정렬 규칙은 다음 식

(6), (7), (8)와 같다.

$$L_i^{count} \geq L_j^{count} \tag{6}$$

$$L_j^{non-empty} \Rightarrow L_i^{non-empty} \tag{7}$$

$$L_i^{non-empty} \wedge L_j^{non-empty} \Rightarrow L_i^{Deadline} \leq L_j^{Deadline} \tag{8}$$

각 레벨의 노드는 <그림 8>과 같이 구성된다. Non-empty 카운터를 이용한 Heap Manager는 메모리 액세스 수를 현저하게 줄이며, 정렬을 간단하고 Calendar방식보다 확장성이 뛰어나다. 또한 Heap Manager는 파이프라인 구조가 용이하여 고속 정렬을 수행한다.

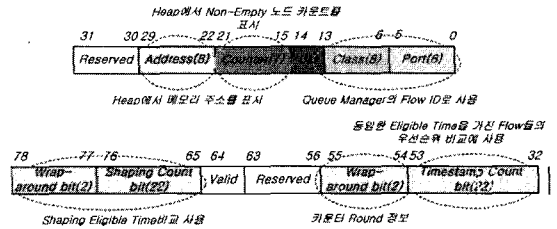


그림 8. Heap Manager의 각 노드 구조.
Fig. 8. A architecture of each node in the HM.

V. 스트리밍 패킷 스케줄러 시뮬레이션 및 성능평가

5.1. 시뮬레이션 시스템 구조 및 환경

본 논문에서는 제안한 스트리밍 패킷 스케줄러의 성능을 검증하기 위해 C 모델링을 통해 시뮬레이션을 수행하였다. 트래픽 발생은 <그림 9>와 같이 On-Off모델을 사용하였다. 방송서비스는 스트리밍 트래픽을 생성하고, 인터넷 서비스는 버스트 트래픽을 생성하여 동시에 입력하였다.

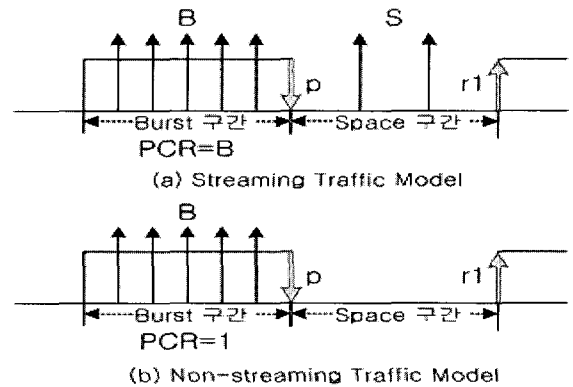


그림 9. 시뮬레이션 트래픽 모델 구조.
Fig. 9. A architecture of simulation traffic model.

<그림 9-(a)>와 같이 멀티미디어 스트리밍 트래픽은 음성을 전달하면서 화상을 보낼 수 있기 때문에 Burst 구간뿐만 아니라 Space 구간에서도 어느 정도 패킷이 전송된다. 스트리밍 트래픽의 셀 발생확률은 다음 식 (9)에 의해 구해진다. B는 버스트한 구간에서 셀이 발생할 확률을 나타내며 PCR값으로 설정된다. MCR(Mean Cell Rate)은 평균 시스템 노드로 0.2에서 0.8까지 0.2단위로 설정하였다. S는 Space한 구간에서 셀이 발생할 확률을 나타낸다. 또한 MBL은 20으로 설정하였다.^{[10][11]}

$$MCR = \frac{\left(\frac{B}{p} + \frac{S}{r1}\right)}{\left(\frac{1}{p} + \frac{1}{r1}\right)}, \quad 1 > B > S > 0 \quad (9)$$

$$p = \frac{1}{\text{Mean Burst Length}}$$

<그림 9-(b)>와 같이 비 스트리밍 데이터는 Burst한 구간에서만 트래픽을 발생한다. 비 스트리밍 트래픽의 셀이 발생확률은 식 (10)에 의해 구해진다.

$$MCR = \frac{\left(\frac{1}{p}\right)}{\left(\frac{1}{p} + \frac{1}{r2}\right)} \quad (10)$$

<그림 10>은 성능 측정을 위한 전체 시스템 구조를 보여준다. IBM스위치 패브릭과의 연동을 통해 제안한 스트리밍 패킷 스케줄러가 얼마나 QoS를 제공하는지 성능을 검증하였다. 2개의 트래픽 모델은 동시에 ingress TM으로 입력되어 목적지 주소에 따라 egress TM으로 전송된다.

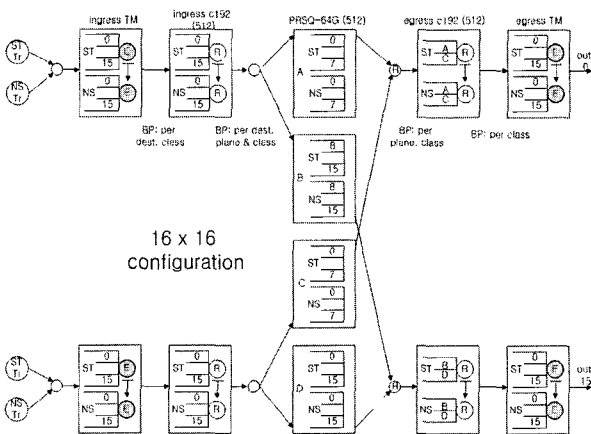


그림 10. 성능시험 시뮬레이션 시스템 구조.
Fig. 10. A system architecture for performance test.

5.2. 성능평가

스트리밍 트래픽과 비 스트리밍 트래픽은 2:8, 4:6, 6:4, 8:2의 비율로 발생시켜 스트리밍 트래픽 로드 증가에 따른 Mean Delay나 Cell Delay Variance를 측정하였다. <그림 11, 12, 13, 14>에서는 스트리밍 트래픽의 로드가 증가하더라도 초기에 여러 단계를 거쳐 발생되는 6셀 타임의 기본 latency만 발생되고, 거의 평균대기 지연시간 없이 일정한 Mean Delay를 유지함을 보여준다.

반면에 비 스트리밍 데이터는 스트리밍 트래픽보다 우선순위가 낮다. 따라서 비 스트리밍 데이터는 스트리밍 서비스를 하고 남은 대역을 이용하므로 비 스트리밍 트래픽 로드가 감소함에도 불구하고 Mean Delay가 증가됨을 보여준다.

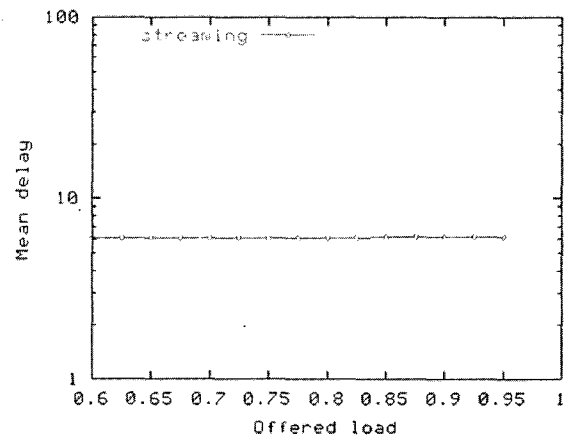


그림 11. 2:8 Mean delay 시뮬레이션 결과
Fig. 11. A result of simulation at 2:8 traffic.

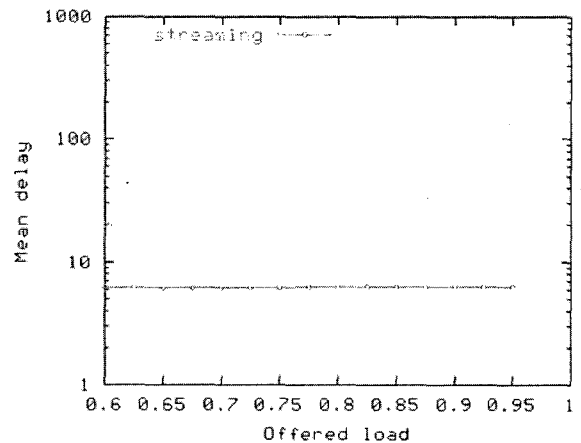


그림 12. 4:6 Mean delay 시뮬레이션 결과
Fig. 12. A result of simulation at 4:6 traffic.

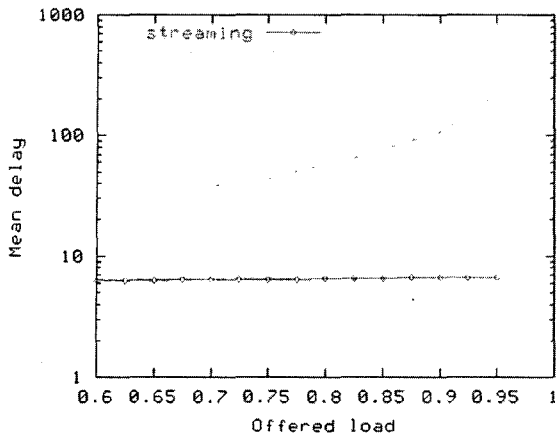


그림 13. 6:4 Mean delay 시뮬레이션 결과
Fig. 13. A result of simulation at 6:4 traffic.

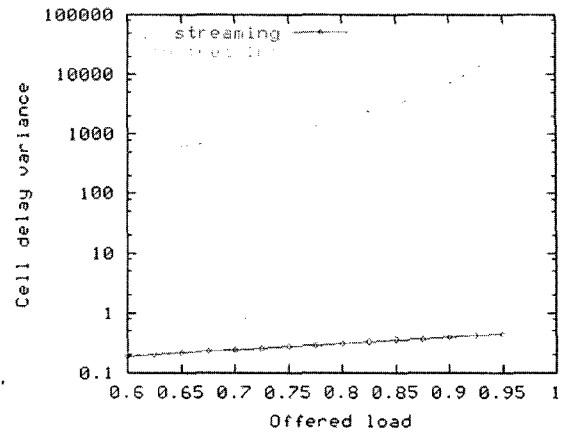


그림 16. 4:6 CDV 시뮬레이션 결과
Fig. 16. A result of CDV simulation at 4:6.

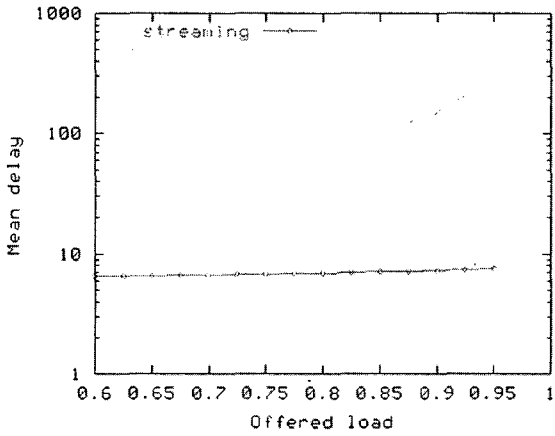


그림 14. 8:2 Mean delay 시뮬레이션 결과
Fig. 14. A result of simulation at 8:2 traffic.

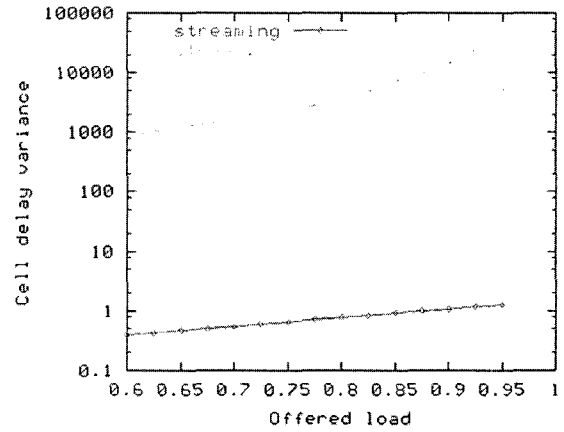


그림 17. 6:4 CDV 시뮬레이션 결과
Fig. 17. A result of CDV simulation at 6:4.

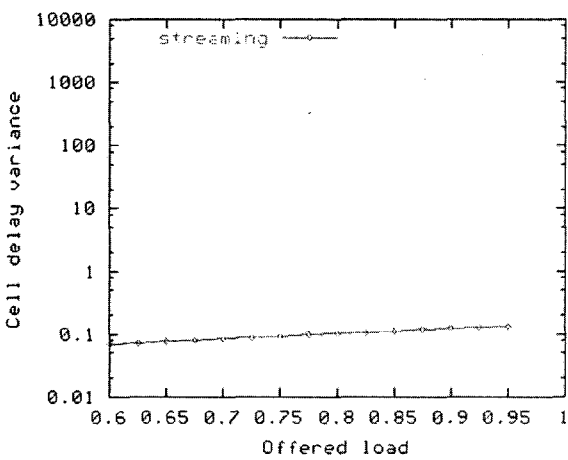


그림 15. 2:8 CDV 시뮬레이션 결과
Fig. 15. A result of CDV simulation at 2:8.

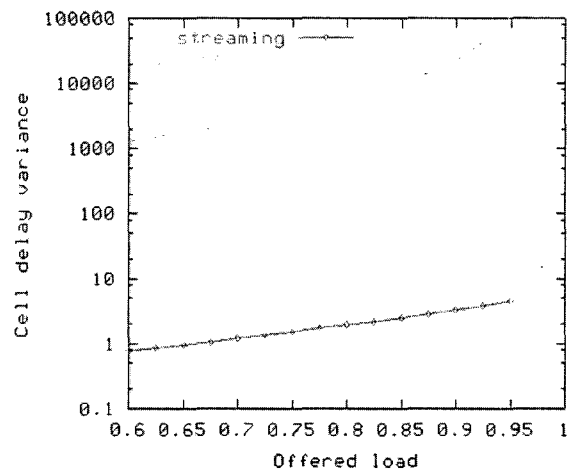


그림 18. 8:2 CDV 시뮬레이션 결과
Fig. 18. A result of CDV simulation at 8:2.

또한 <그림 15, 16, 17, 18>은 스트리밍 트래픽이 CDV 측면에서도 4:6 트래픽 모델까지는 시스템 로드 에 상관없이 1 셀 타임 이하의 CDV간격을 유지함을 보여준다. 그리고 6:4나 8:2에서는 시스템 로드가 증가 시 약간 CDV가 증가하지만 이 정도는 VoIP의 음성 Toll 서비스 수준의 트래픽 처리를 제공할 수 있다.

VI. 결 론

본 논문에서는 인터넷 및 방송서비스를 요구하는 가입자들에게 지터 및 셀 지연변이, 지연 등을 제공할 수 있는 10Gbps급 스트리밍 패킷 스케줄러를 제안하였다. 특히 제안한 스트리밍 패킷 스케줄러는 끊임 없는 방송서비스를 위해 delay 기반의 EDF스케줄링 알고리즘을 사용하였으며, Per-Flow Shaper와 함께 연동하여 end-to-end delay bound를 제공한다.

시뮬레이션을 통해 제안한 스트리밍 패킷 스케줄러가 스트리밍 방송서비스에서 요구되는 지터나 지연 등을 만족함을 보여주었다. 시뮬레이션 결과 Mean Delay는 로드 증가에 상관없이 거의 Zero의 지연 시간을 제공한다. 따라서 본 논문에서 제안한 스트리밍 패킷 스케줄러는 인터넷 및 방송서비스에서 요구되는 QoS보장을 제공하며, 가입자 액세스 장비의 라인카드에서 QoS보장을 위해 사용된다. 현재는 delay bound를 PCR값만을 고려하였지만, 향후 좀더 정확한 delay bound를 위해 MCR, SCR등을 고려 할 계획이다.

참 고 문 헌

- [1] "Providing end-to-end statistical delay guarantees with earliest deadline first scheduling and per-hop traffic shaping", Sivaraman, V.; Chiussi, F. INFOCOM2000. 19th Annual Joint Conference of the IEEE Computer and Communication Societies. Proceedings. IEEE, Vol: 2, 2000 Page: 631-640.
- [2] "Achieving End-to-End Delay Bounds by EDF Scheduling without Traffic Shaping", Kai Zhu; Yan Zhuang; Yannis Viniotis, In Proceedings of IEEE INFOCOM, Vol. 2, pp. 1113-1122, April 2001.
- [3] "End-to-End Statistical Delay Guarantees under GPS and EDF Scheduling": A Comparison Study, Sivaraman, V; Chiussi, F.M.; Gerla, M. In Proceeding of IEEE INFOCOM, Vol. 3, pp. 1493-1501, April 2001.
- [4] "Traffic shaping for end-to-end guarantees with EDF scheduling", Sivaraman, V.; Chiussi, F.M.; Gerla M. Quality of Service, 2000. IWQOS. 2000 Eighth International Workshop on, 2000 page: 10-18.
- [5] "Enhancing the end-to-end schedulability condition of EDF scheduling for real-time application", Elsayed, K.M.F. ATM Workshop Proceedings, 1998 IEEE, 1998 page:75-79.
- [6] "An ASIC Core for Pipelined Heap Management to support scheduling in High Speed Networks", Aggelos D. FORTH-ICS / TR-278 Report, 2000, 11.
- [7] "Technical guide to the TM10 Preliminary Data Book", Agere TM10 Manual, September, 2002.
- [8] "QX-1 OC-192 Traffic Manager Architectural Specification Document Version 2.1", Ezchip, 2002.
- [9] "ZTM552A/F Traffic Manager Preliminary Functional Specification Rev.2.0", ZettaCom, October, 25, 2002.
- [10] "Traffic description and performance evaluation in multimedia ATM environment" A. Lombardo, G. Schembra, Proc.IEEE GLOBECOM'96, 1996.
- [11] "The pitfalls of scaling VoIP", BCR, vol 32. March 2002.

저 자 소 개



김 광 옥(정회원)

1999년 조선대학교 정보통신공학과 학사. 2001년 전남대학교 전자공학과 석사. 2001년 ~ 현재 한국전자통신연구원 네트워크연구소 연구원근무. <주관심 분야: 네트워크 프로세서응용, 트래픽 스케줄링 기술>

<주관심 분야: 네트워크 프로세서응용, 트래픽 스케줄링 기술>



최 병 철(정회원)

1987년 한양대학교 전자공학과 학사. 1997년 한남대학교 전자공학과 석사. 1997년 ~ 현재 부산대학교 컴퓨터공학과 박사과정, 1987년 ~ 1993년 삼성 전자통신연구원 주임연구원 근무, 1993년 ~ 현재 한국전자통신연구원 네트워크연구소 선임연구원 근무. <주관심 분야: 고속통신망, IP록업, 네트워크프로세서응용>

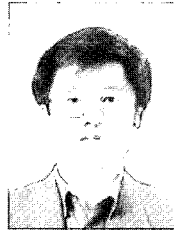
<주관심 분야: 고속통신망, IP록업, 네트워크프로세서응용>



박 완 기(정회원)

1991년 충남대학교 전자공학과 학사. 1993년 충남대학교 전자공학과 석사. 2001년 ~ 현재 충남대학교 전자공학과 박사과정, 1993년 ~ 2000년 국방과학연구소 연구원 근무, 2000년 ~ 현재 한국전자통신연구원 네트워크연구소 선임연구원 근무, <주관심 분야: MPLS, Gigabit Ethernet, EPON>

<주관심 분야: MPLS, Gigabit Ethernet, EPON>



곽 등 응(정회원)

1983년 동국대학교 전산학과 공학사. 1985년 동국대학교 전산학과 공학석사. 1998년 ~ 현재 한국정보통신대학교 공학부 공학박사, 1985년 ~ 현재 한국전자통신연구원 네트워크연구소 책임연구원 근무 <주관심 분야: 트래픽 스케줄링, 네트워크 프로세서응용>

<주관심 분야: 트래픽 스케줄링, 네트워크 프로세서응용>