

분산 멀티미디어 시스템에서 개선된 QoS 협상과 자원 확보 방법

노 영 육
신라대학교 컴퓨터교육과
yulho@silla.ac.kr

An Advanced QoS Negotiation and Resource Reservation for Distributed Multimedia System

YoungUhg Lho
Department of Computer Education, Silla University

요약

분산 멀티미디어 시스템은 이식성을 보장하고 동적인 자원 변화를 수용하고 올바른 시간에 올바른 데이터 인도를 보장해야 한다. 따라서 분산 멀티미디어 시스템의 각 구성 요소들은 멀티미디어 응용에서 QoS(Quality of Service)를 제공하기 위한 자원 관리 기법과 다양하고 종류의 QoS 요구를 수용할 수 있어야 한다. 본 연구에서 가변 전송율을 갖는 비디오에 대한 여러 종류의 QoS 요구를 만족하기 위한 비디오 배치 및 검색 방법을 제안한다. 제안된 방법은 여러 종류의 해상도 요구에서 개선된 성능을 보였다.

1. 서론

분산 멀티미디어 시스템의 기본 목표는 다른 시스템에 이식성을 보장하고 동적인 자원 변화를 수용하고 올바른 시간에 올바른 데이터 인도를 보장하는 것이다. 따라서 분산 멀티미디어 시스템의 각 구성 요소들은 멀티미디어 응용에서 QoS(Quality of Service)를 제공하기 위한 자원 관리 기법이 필요하다.

어떤 응용에 특정 수준의 QoS 를 제공하기 위해, 시스템은 사용 가능한 자원을 충분히 가지고 있어야 한다. 그리고 이러한 자원은 응용이 필요할 때 사용할 수 있도록 스케줄링 되어야 한다. 이를 위해 응용의 QoS 요구가 만족되는지를 검사하기 위한 QoS 보장 계산, 주어진 QoS 보장을 위한 자원 확보, 주어진 QoS 보장이 자원 접근 스케줄링에 의해 만족되도록 하는 자원 스케줄링이 필요하다.

멀티미디어 응용을 쉽게 생성하기 위해 Apple Quicktime, Pandora[3], ACME, XMovie, Lancaster 시스템 [4] 등이 개발되었다. 그러나 이러한 시스템의 대부분은 응용에 신뢰성 있는 QoS 를 제공하지 못한다. QuickTime, Pandora, Xmovie 와 같은 시스템은 자원 관리 기법을 제공하지 않으며, ACME 와 Lancaster 시스템은 자원 관리에 대한 일부 기법이 제공되고 있다.

멀티미디어 응용을 위한 CPU 스케줄링 기법은 DASH[5], RT-MACH[6]와 같은 연구용 운영체제에서 개발되었으며, 이러한 연구에서는 연속매체를 위한 CPU 스케줄링을 표준 운영체제에 통합하였다[8,9]. 대부분의 QoS 와 관련된 기존 연구는 분산 멀티미디

어 시스템의 각 구성요소에 관해 연구가 이루어졌으며 통합된 시스템 환경에서의 연구[1,2]는 극히 미미한 상태이다. 그리고 전체 시스템 상에서의 자원관리 모델에 관한 연구가 있으나 자원관리에 대한 모델링 기술에 집중되어 있다.

그리고 SRP[10,11], ST-II, ST-II+, RSVP[12]와 같은 자원 확보 프로토콜은 QoS 요구와 자원 요구에 대한 정보를 네트워크 노드 사이에 교환하는 기법을 제공하나 분산시스템 전체 차원에서 모든 자원에 대한 요구를 누적 관리하는 기법을 제공하지 못하고 있다[8]. 특히 본 연구에서 가변 전송율을 갖는 비디오에 대한 요구를 만족하기 위한 시스템 구성 요소에 관한 기존 연구들은[13,14] 시스템 전체 차원에서의 자원 관리 기법에 관한 연구는 부족한 형편이다.

본 연구의 2 장과 3 장에서는 본 연구에서 고려하는 시스템 구조와 확장 가능한 압축 방법에 대해 기술한다. 4 장에서는 자원 확보 방법에 만족해야 할 시간 제한성과 공간 제한성을 정의하고 MZR 디스크에서 QoS 를 고려한 가변 비디오 배치 방법을 기술한다. 그리고 5 장과 6 장에서는 성능 측정 결과와 결론을 기술한다.

2. 시스템 구조

본 연구에서 고려하는 시스템 모델은 그림 1 과 같다. 비디오 서버는 비디오 데이터를 저장하는 디스크 배열, 버퍼로 사용되는 메모리 자원, 전송을 위한 네트워크 인터페이스로 구성된다. 디스크 검색 스케줄

려는 주기적으로 여러 비디오 스트림에 대한 데이터를 디스크 디스크로부터 읽어 메모리에 저장한다. 네트워크 스트림 스케줄러도 주기적으로 여러 비디오 스트림에 대한 데이터를 각 주기 동안에 네트워크에 전송한다. 디스크 배열의 각 디스크는 검색하는 동안에 SCAN 방법으로 디스크 헤드를 스케줄링한다고 가정한다. 디스크 배열에 데이터를 배치하고 데이터를 검색하는 방법은 중요한 문제이나 여기서는 특별한 방법을 가정하지 않는다. 비디오 서버는 완전한 대화성을 제공하여 시청자는 시청 기간동안에 언제든지 시청을 중지와 재시작 할 수 있다. 비디오 서버는 대화형 검색 전체 기간 동안에 고정된 대역폭과 버퍼 확보를 가진다. 그리고 상영 기간 동안에 각 비디오 스트림을 검색하기 위한 자원의 재협상은 없다고 가정한다.

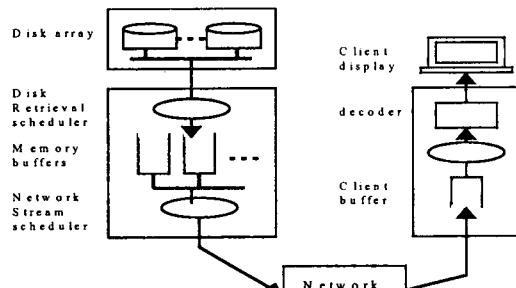


그림 1. VOD 서버의 시스템 구조

3. 확장 가능한 비디오 압축

일반적인 비디오 저장 서버는 하나의 비디오에 대한 최상의 해상도 하나만 저장하지만 확장 가능한 압축방법은 여러 개의 해상도를 저장한다. MPEG-2 표준은 공간적(spatial)인 방법과 SNR 확장성을 혼합하여 비디오 스트림을 기본 계층(base layer), 공간 증진 계층(spatial enhancement layer)과 SNR 증진 계층(SNR enhancement layer)으로 분리하여 압축한다.

확장성 있는 압축 방법으로 압축된 비디오를 디스크에 배치하는 기존 연구들은 MZR 디스크의 특성을 고려하지 않고 있다. VBR 비디오를 여러 사용자에게 동시에 서비스할 때 비디오 서버의 디스크 대역폭 용량을 일시적으로 초과하는 현상이 발생할 수 있다. 따라서 VBR 비디오 저장 서버는 모든 사용자의 다양한 QoS 요구를 항상 만족하기는 어려우므로 비디오 서버의 용량 초과 현상을 확률적으로 최대한 줄이는 방법을 사용한다. 그리고 비디오 서버의 처리 능력 보다 높은 요구에 대해서는 현재 서비스되고 있는 사용자들을 서비스 주기에서 제거하지 않고 서비스되고 있는 질을 저하하여 서비스되고 있는 사용자는 계속적으로 서비스하는 방법을 사용하고 있다.

기존 연구들에서는 pause, FF(Fast Forward), Reverse, FR(Fast Reverse) 등의 VCR 형 연산을 지원하기 위해 세그먼트 skip 방법, 디스크 이용률을 줄이는 방법으로 세그먼트를 skip하는 방법, 정상 속도보다 N 배 빠른 요구는 N 배 많은 대역폭을 요구하는 문제점이

있으므로 빠른 요구에 대해서 해상도가 낮은 스트림을 통제적으로 multiplexing하는 방법 등이 제안되었다.

4. 자원 확보 방법

VBR 비디오 데이터를 검색할 때 고정된 시간(CT: Constant Time) 단위로 검색하는 방법과 고정된 크기(CD: Constant Data) 단위로 검색하는 방법이 있다. CT 방법은 연속성을 유지하고 손실 없는 검색을 위해 검색하는 비디오 스트림의 최고 데이터 비율(최악의 경우)과 동일한 디스크 대역폭을 확보하는 것이 필요하다. 그러나 CT 검색에 기반한 비디오 서버는 지원 가능한 스트림의 수를 최대화할 때 메모리 자원을 완전히 사용할 수 없는 문제점이 있다(디스크 대역폭이 병목현상이 된다). CD 방법에서 VBR 비디오 데이터를 읽는 양은 고정되고 전송되는 데이터량은 가변이므로 버퍼 기아 상태가 발생될 수 있다. CD 검색 방법은 지원 가능한 스트림 수를 최대화 할 때 비디오 서버의 대역폭을 최대로 사용할 수 없는 문제점이 있다(메모리가 병목현상이 된다). 따라서 CD 와 CT 의 문제점을 동시에 극복할 수 있는 검색 방법과 자원 확보 방법이 필요하다. CT 와 CD 방법에서 디스크 대역폭을 확보할 때 각각 VBR 비디오 데이터의 최고 데이터 비율과 평균 데이터 비율이 확보한다.

4.1 검색 제한

비디오 서버의 검색 제한을 정의하기 위한 사용한 기호는 표 1과 같다.

표 1. 기호 정의

기호	정의
F	초당 재생되는 프레임 수
N_g	한 주기동안에 검색되는 프레임 수
N_v	비디오 서버에 저장된 비디오 수
R_b	서비스 주기 동안에 사용자 i 의 재생 요구률
$R_c(i)$	i 번째 존에 할당될 블록 크기
S_b^i	MZR 디스크 하나의 저장 용량
T_{max}	비디오 블록 하나의 전송 시간
T_d	MZR 디스크의 평균 탐색 시간
T_s	d 번째 디스크에서 서비스되고 있는 사용자 수
U_d	

미리 정의된 고정 길이의 주기 T_{period} 를 가지고 m 명의 사용자가 존재하고 주기 내에서 각 사용자를 서비스하는 활동을 세션이라 하자. 서버는 m 명의 사용자를 서비스하기 위해서 주기내에서 세션 사이에 전환되어야 한다. 그리고 다음 주기에서 세션이 서비스될 때 까지 사용자 요구를 지속적으로 서비스하기 위해 각 세션에 충분한 버퍼가 할당되어야 한다. 시간 제한성은 한 주기 동안에 모든 세션을 서비스하기 위한 디스크 접근 시간의 합이 주기보다 작거나 같아야 함을 나타낸다. 공간 제한성은 한 주기 동안에 각 세션을 서비스하기 위해 디스크에서 읽은 데이터를 저장하는 버퍼 크기는 한 주기 동안 서비스 가능해야 하고 모든 세션이 필요한 버퍼 크기 합은 서버에서 사용 가능한 버퍼 크기보다 작거나 같아야 함을 나타낸다.

낸다. 새로운 사용자 요구의 채택 여부는 새로운 사용자 요구를 처리하는 세션을 포함하여 시간 제한성과 공간 제한성을 모두 만족할 때 새로운 사용자 요구를 허용한다.

- 시간 제한성

$$T_{\text{ms}} + \sum_{i=1}^{v_e} \left(T_i + \frac{R_c(i)}{T_{\text{avg}}} \times \frac{N_g}{F} \right) \leq \frac{N_g}{F}, \quad d = [1, \dots, N_d]$$

- 공간 제한성

$$\sum_{d=1}^{N_d} \sum_{i=1}^{v_e} \frac{R_c(i) N_g}{F} \leq \frac{B_g}{2}$$

4.2 배치 방법

- Sequential Placement

동일한 q_j 를 갖는 비디오 블록을 디스크에 연속적으로 저장하는 방법으로 i 번째 비디오의 j 번째 전송 요구률을 갖는 비디오 블록 V'_i 은 디스크 $d (= (i-j) \% v)$ 의 j 번째 GOZ(j)에 배치한다. 이 방법은 사용자들이 요구하는 비디오가 균등할 경우에는 부하가 여러 디스크에 분산되어 모든 사용자에게 고정된 질의 서비스를 제공할 수 있다.

- Staggered Sequential Placement

전송율이 다른 비디오 블록을 다른 디스크에 라운드 로빈 방식으로 배치한다. GOZ(j)에 전송율 q_j 를 갖는 비디오 블록을 저장한 후에 동일한 전송율을 갖는 다른 비디오 블록을 연속적으로 저장한다. 비디오 블록 V'_i 은 디스크 $d (= (i+j-2) \% N_d)$ 의 GOZ(j)에 저장된다. 서비스 주기 동안에 i 번째 QoS(i)를 요구하는 사용자를 서비스하기 위해서 $m-i+1$ 개의 디스크를 접근한다. 예를 들어 사용자가 i 번째 비디오에 대해 QoS(j)를 요구하면 디스크 $(i+j-2) \% N_d$ 부터 $(i+j-1) \% N_d, \dots, (m+j-2) \% N_d$ 까지 $m-i+1$ 개의 디스크를 접근한다.

- Partitioned Striping Placement

이 방법은 RAID에서 사용하는 스트라이핑 방법을 적용한 것으로 엇갈린 배치 방법의 디스크 낭비 문제를 해결할 수 있다. m 개로 분리된 비디오 스트림을 각각 N_{gap} 개의 GOP(Group of Picture) 단위로 그룹화하여 N_d 개의 디스크에 스트라이핑하는 방법이다. N_{gap} 는 하나의 읽기 연산에서 접근되는 최소 GOP 수이다. 각 디스크에는 N_{gap}/N_d 크기의 비디오 블록이 저장되며 전송율 q_j 를 갖는 모든 비디오 블록은 모든 디스크의 GOZ(j)에 스트라이핑 된다. 사용자가 가장 높은 QoS(1)을 요구하면 모든 전송율을 갖는 비디오 블록을 읽기 위해 N_{gap} 단위로 m 번의 읽기 연산이 필요하다.

- Reverse Staggered Sequential Placement

확장성 있는 압축 방법의 특성상 OoS(i) 요구를 만족하기 위하여 q_i 이하의 전송율을 갖는 모든 스트림을 접근하여야 한다. 따라서 사용자의 QoS 요구가

다양할 때 전송율이 낮은 스트림은 전송율이 높은 스트림에 비해 접근 회수가 높다. 엇갈린 순차 배치 방법은 이러한 성질을 이용하여 디스크의 바깥쪽 GOZ에 접근 회수가 높은 스트림을 배치하고 디스크의 안쪽 GOZ에 접근 회수가 낮은 스트림을 라운드 로빈 방식으로 배치한다. 비디오 블록 V'_i 은 디스크 $d (= (m+i-j-1) \% N_d)$ 의 GOZ($m-j+1$)에 저장한다. QoS(i)를 요구를 처리하기 위해서 엇갈린 배치 방법처럼 $m-i+1$ 개의 디스크 접근이 필요하다.

4.3 QoS 를 고려한 MPEG 데이터 배치

MPEG-2은 비디오 압축에 관한 국제 표준으로 현재 VOD 서버의 응용 분야에서 가장 많이 사용되고 있다. MPEG-2은 GOP(Group of Picture) 단위로 독립적으로 압축되고 복원이 가능한 구조를 가지고 있다. GOP는 I(Intracode), P(Predicted), B(Bi-directional) 프레임으로 구성된다. GOP가 다음과 같이 I 프레임은 1 개, P 프레임은 3 개, B 프레임은 6 개로 구성되었다고 하자.

I B₁ B₂ P₁ B₃ B₄ P₂ B₅ B₆ P₃

a) MPEG 비디오의 구성 :

I₁B₁B₂P₁B₃B₄P₂B₅B₆P₃B₇B₈P₄B₉B₁₀P₅B₁₁B₁₂P₆ ...
I₁₂B₆₇B₆₈P₃₄B₆₉B₇₀P₃₅B₇₁B₇₂P₃₆

b) 저장 형태 :

I₁ I₂ I₃ ... I₁₂ P₁ P₄ P₇ P₁₀ P₁₃ P₁₆ P₁₉ P₂₂ P₂₅ P₂₈ P₃₁ P₃₄ P₂
... P₃₃ P₃₆ B₁ B₇ B₁₃ B₁₉ B₃ B₂₅ B₃₁ B₃₇ B₄₃ B₄₉ B₅₅ B₆₁ B₂ B₈ ... B₆₆ B₇₂

그림 2. MPEG 비디오의 구성과 저장 형태

B₂ 프레임을 복원하기 위해서는 시간 상으로 뒤에 존재하는 P₁ 프레임이 필요하다. 따라서 MPEG-1 비디오를 재구성하여 저장하는 방법이 필요하다. 초당 재생 프레임 비율이 $F = 30$ 이면 주기 길이 T_{period} 는 $10/30$ 초 단위의 배수가 된다. 주기가 $T_{\text{period}} = 4$ 초이면 3.3절의 시간 제한성과 공간 제한성의 식에서 $N_g = 120$ 이 된다.

[15]는 그림 2의 a)를 그림 2의 b)와 같은 형태로 재구성하여 저장하는 방법을 제안하였다. 재구성된 저장 형태는 I 프레임은 12 개, P 프레임은 36 개, B 프레임은 72 개로 구성된다. 저장되는 순서는 각 GOP의 I 프레임을 먼저 저장한다. I 프레임에 연속하여 각 GOP의 첫 번째 P 프레임들을 저장하고 이어서 두 번째 P 프레임을 저장한다. 세 번째 P 프레임을 저장한 후에 B 프레임을 P 프레임을 저장하는 방법과 동일한 방법으로 저장한다.

위와 같은 형태로 저장된 경우를 본 연구에서 제안한 방법에 적용하는 문제를 살펴 보자. 전형적인 비디오를 MPEG-2으로 압축하면 I, P, B 프레임의 크기는 보통 5:3:1을 갖는다고 알려져 있다. 프레임 크기 비율과 GOP의 프레임 구성 비율(I:P:B=1:3:6)을 사용하면 하나의 GOP의 저장 공간에서 I, P, B 프레임의 크기는 5.9:6의 비율을 가지고 그림 2의 b)도 I, P, B 프레임의 각각이 필요한 저장 공간 비율은 동일하다. MPEG와 확장성 있는 압축 방법을 비교하기 위해 확장성 있는 압축 방법은 m 개의 QoS 요구를 만족할 수 있도록 분리 압축되어 있고 QoS(I) \geq QoS(2)

$\geq \dots \geq QoS(m)$ 관계가 만족한다고 하자. MPEG을 재생하기 위해서 I 프레임은 반드시 필요하므로 확장성 있는 압축 방법의 가장 낮은 QoS에 대응된다. 기호 $\{X\}$ 는 X 종류의 모든 프레임의 집합을 나타낸다고 하자. MPEG 데이터를 I, P, B 프레임 별로 QoS를 분리하면 다음과 같이 된다.

$$\begin{aligned}QoS(3) &= \{I\} \\QoS(2) &= \{P\} + \{I\} \\QoS(1) &= \{B\} + \{P\} + \{I\}\end{aligned}$$

앞 절에서 제안한 방법을 적용하면 GOZ(1)에 $\{B\}$ 을, GOZ(2)에 $\{P\}$ 을, GOZ(3)에 $\{I\}$ 을 배치해야 한다. 그런데 $\{I\}$, $\{P\}$, $\{B\}$ 의 저장 공간 비율은 5:9:6을 가지므로 순차 배치, 엇갈린 배치와 분할된 스트라이핑은 GOZ(1), GOZ(2)와 GOZ(3)에 각각 $\{P\}$, $\{B\}$, $\{I\}$ 을 배치하고 역엇갈린 배치는 GOZ(1), GOZ(2), GOZ(3)에 각각 $\{I\}$, $\{P\}$, $\{B\}$ 프레임을 배치하면 된다.

5. 성능 측정 결과

성능을 측정하기 위해 디스크 모델 인수로 $N_d=6$, $m=6$, $T_{sr}=7.24ms$, $T_{seek}=20.4ms$ 을 사용하여 서비스 주기내에서 요구가 만족하는 최대 사용자 수를 구하기 위해 사용자를 하나씩 증가하면서 성능을 측정하였다. 그럼 3은 완전한 해상도를 요구한 경우의 최대 사용자를 나타낸다. 그럼 4는 부분 해상도를 요구하는 경우의 최대 사용자 수를 나타낸다.

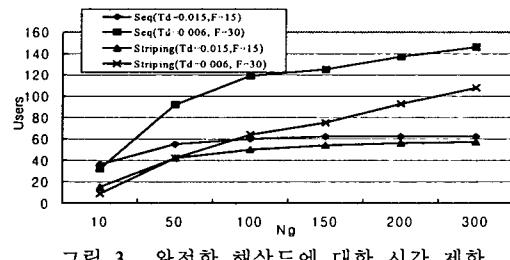


그림 3. 완전한 해상도에 대한 시간 제한

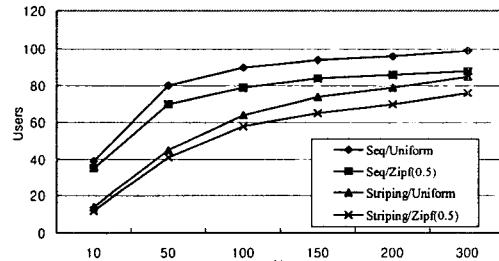


그림 4. 부분 해상도에 대한 시간 제한

5. 결론

본 논문에서는 분산 멀티미디어 시스템 환경에서 다양한 QoS의 요구를 수용할 수 있는 VBR 비디오의 배치와 검색 방법을 제안하였다. 다른 전송율을 갖는

비디오 스트림을 디스크의 다른 존에 배치함으로써 전체적인 전송율을 일정하게 유지하도록 한다. 본 연구의 성능 측정 결과 엇갈린 배치 방법이 완전한 해상도와 부분 해상도를 요구하는 모든 경우에서 좋은 성능을 나타내었다. 그러나 본 연구에서 제안한 방법은 고정된 전송율을 보장하기 위해 디스크 저장 공간이 낭비되는 문제점이 있다. 향후 연구 과제로 이러한 낭비 공간을 줄이고 편기된 접근의 경우에서 지원 가능한 최대 사용자 수를 최대화하기 위해 버퍼를 동적으로 할당하는 방법에 대한 연구가 필요하다..

참고 문헌

- [1] K. Nahrstedt and J.M. Smith, "The QoS Broker," IEEE Multimedia, Vol.2, No.1, 1995, pp.53~67.
- [2] K. Nahrstedt and R. Steinmetz, "Resource Management in Networked Multimedia Systems," IEEE Computer, Vol. 28 No.5, 1995, pp.52~63.
- [3] A. Hopper, "Pandora -an Experimental System for Multimedia Applications," ACM Operating System Review, Vol.24 No.2, 1990, pp.19~34.
- [4] P.Lougher and S.Sherpherd, "The Design of A Storage Server for Continuous Media," The Computer Journal, Vol.36 No.1, 1993, pp.32~42.
- [5] D.P. Anderson, S. Tzou, R. Wahbe, R. Govindan and M.Andrews, "Support for Continuous Media in the DASH System," Proc. of the 10th ICDCS, 1990, pp.54~61.
- [6] S.T.-C. Chou and H. Tokuda, "System Support for Dynamic QoS Control of Continuous Media Communication," proc. of the 3rd NOSSDAV, 1992, pp.363~168.
- [7] C.W. Mercer, S.Savage and H. Tokuda, "Processor Capacity Reserves : Operating System Support for Multimedia Applications," Proc. of the 1st ICMCS, 1994, pp.90~99.
- [8] G.Coulson, G.S. Blair, N.Davies and N.Williams, "Extension to ANSA for Multimedia Computing," Computer Networks and ISDN Systems, Vol.25, 1992, pp.305~323.
- [9] G.Coulson, G.S. Blair, P. Robin and D. Shepherd, "Extending the Chorus Micro-Kernel to Support Continuous Media Applications," Proc. of the 4th NOSSDAV, 1993, pp.3~5.
- [10] D.P. Anderson, R.G. Herrwitz and C.Schaefer, "SRP : A Resource Reservation Protocol for Guaranteed-Performance Communication in the Internet," TR-90-006, International Computer Science Institute, Berkeley, CA, USA, Feb. 1990.
- [11] R. Govindan and D.P. Anderson. "Scheduling and IPC Mechanisms for Continuous Medias," Proc. of the 13th Symp. on Operating System Principles, 1991, pp.68~80.
- [12] L.Zhang, S.Deering, et.al, "RSVP : A New Resource Reservation Protocol," IEEE Network, 1993, Sep. pp.8~18.
- [13] T. Nakajima, "A Dynamic QoS Control Based on Optimistic Processor Reservation," Proc. of ICMCS, 1996, pp. 95~103.
- [14] S. Paek and S-F. Chang, "Video Server Retrieval Scheduling for Variable Bit Rate Scalable Video," Proc. of ICMCS, 1996, pp. 108~112.
- [15] Chen, H.J, A Disk Scheduling Scheme and MPEG Data Layout Policy for Interactive Video Access From A Single Disk Storage Device, Ph.D Dissertation, Boston University, 1995