

# 가변 비트율 주문형 비디오 서버에서 선반입과 캐싱을 이용한 버퍼 관리 기법

## (A Buffer Management Scheme Using Prefetching and Caching for Variable Bit Rate Video-On-Demand Servers)<sup>†</sup>

김 순 철\*  
(Soon-Cheol Kim)

**요 약** 주문형 비디오 시스템에서 가변 비트율로 압축된 데이터들은 압축 대상이 되는 데이터의 내용 변화와 압축 기법의 특성으로 인해 단위 시간 당 처리해야 할 데이터 크기가 일정하지 않다. 그러나, 대부분의 주문형 비디오 서버들은 가변 비트율로 압축된 데이터를 실시간에 처리하기 위해 가변 데이터 크기의 최대값으로 시스템 자원을 예약하기 때문에 자원의 활용률이 떨어진다. 본 논문에서는 주문형 비디오 서버에서 시스템 자원의 활용률을 향상시켜 더 많은 사용자를 수용할 수 있는 버퍼 관리 기법을 제안한다. 제안된 버퍼 관리 기법은 데이터에 대한 캐싱 기법과 함께 가변 비트율로 압축된 데이터를 예약할 때 선반입 기법을 적용함으로써 비디오 데이터의 가변성을 줄이고 디스크 대역폭과 서버 버퍼에 대한 활용률을 극대화하였다. 본 논문에서 제안한 버퍼 관리 기법의 효율성은 모의 실험을 통해 확인하였다.

**Abstract** Video-On-Demand servers have to provide timely processing guarantees and reduce the storage and bandwidth requirements for continuous media. However, compression techniques used in Video-On-Demand servers make the bit rates of compressed video data significantly variable from frame to frame. Consequently, most previous Video-On-Demand servers, which use constant bit rate retrieval to guarantee deterministic service, under-utilize the system resources and restrict the number of clients. In this paper, I propose a buffer management scheme called CAP(Caching And Prefetching) for Video-On-Demand server using variable bit rate continuous media. By caching and prefetching the data, CAP reduces the variation of the compressed data and increases the number of clients simultaneously served and maximizes the utilization of system resources. Results of trace-driven simulations show the effectiveness of the scheme.

## 1. 서 론

### 1.1. 연구 배경

처리의 성능 향상과 고속 대용량 저장 장치의 개발 및 컴퓨터 통신망의 발전은 데이터 처리(data processing)와 데이터 통신(data communication) 분야에서 이러한 멀티미디어 데이터의 사용을 가능하게 하였다. 주문형 비

디오 시스템, 전자 도서관, 화상회의와 같은 멀티미디어 응용들이 이미 실생활에서 사용 중이거나 개발 중에 있다. 이와 같은 멀티미디어 응용들은 일상 생활 속에서 컴퓨터의 이용을 더욱 가속화시킬 것이다.

멀티미디어 데이터 중에서 주문형 비디오 시스템에서 주로 처리해야 하는 오디오 데이터나 비디오 데이터는 정해진 시간 내에 처리되어야 한다는 시간적인 특성을 가지고 있기 때문에 연속 매체(continuous media)라고 한다. 대용량의 데이터가 요구하는 저장 공간과 전송 대역폭을 줄이기 위해 사용되는 압축 기법과 데이터의 실시간 처리를 보장하기 위한 자원의 예약 기법은 모두 연속 매체를

<sup>†</sup> 이 논문은 1999년도 대구대학교 학술연구비 지원에 의한 논문임  
\* 대구대학교 공과대학 컴퓨터정보공학부 전임강사

처리하는데 있어 필수적이다. MPEG과 같은 표준 압축 기법에 의해 가변 비트율로 압축된 데이터들은 압축 데이터 내용의 변화로 인해 단위 시간에 재생해야 할 비디오 프레임의 크기가 일정하지 않다. 따라서, 압축된 비디오 데이터량의 시간에 따른 가변성은 주문형 비디오 시스템에서의 자원 관리를 매우 어렵게 한다 [1,2]. 고정 비트율(Constant Bit Rate)의 비디오 데이터 검색 기법을 사용하는 기존의 주문형 비디오 서버의 경우 가변 비트율의 비디오 데이터를 실시간에 처리하기 위해서는 가변 데이터의 최대값으로 자원을 예약해야 하기 때문에 자원 활용률이 떨어지게 된다.

## 1.2. 관련 연구

가변 비트율(Variable Bit Rate)로 압축된 비디오 데이터의 크기는 시간에 따라 매우 심하게 변하기 때문에 비디오 데이터를 실시간에 처리하기 위해서는 비디오 서버와 통신망 그리고 사용자 시스템 등 데이터가 전달되는 모든 경로 상에서 자원들에 대한 효율적인 관리가 필요하다. 지금까지 통신망 상에서 가변 비트율의 연속 매체를 처리함에 있어 대역폭과 지연을 보장하는 자원 관리 기법에 대해서는 많은 연구들이 수행되었다 [3,4]. 그러나, 서버에서 가변 비트율의 데이터를 처리하기 위한 자원 관리 기법에 대한 연구는 매우 미진하였다. 현재 개발된 대부분의 주문형 비디오 시스템들은 가변 비트율의 비디오 데이터를 실시간에 처리하기 위해 비디오 데이터의 최대값으로 자원을 예약함으로써 자원을 최대한 사용할 수 없었다 [5,6,7].

## 1.3. 논문의 범위와 구성

주문형 비디오 시스템에서는 제한된 자원으로 연속 매체들의 실시간 처리를 보장하면서 수용 사용자 수를 극대화할 수 있어야 한다. 본 논문에서는 주문형 비디오 서버에서 연속 매체의 실시간 처리 조건을 만족하면서 시스템 자원을 최대한 활용하여 더 많은 사용자를 수용할 수 있는 버퍼 관리 기법을 제안한다. 제안하는 버퍼 관리 기법은 기 저장된 데이터에 대한 참조 패턴을 오프라인(off-line)으로 분석하여 이를 기반으로 하여 교체할 버퍼를 결정하게 된다. 이러한 버퍼 교체 기법은 과거의 데이터 참조 패턴을 분석하여 교체할 버퍼를 결정하는 버퍼 교체 기법보다 더 높은 버퍼 활용률을 나타낸다. 제안하는 버퍼 관리 기법은 이러한 버퍼 교체 기법과 함께 가변 비트율로 압축된 데이터를 예약할 때 디스크 대역폭에 대한 활용률을 극대화할 수 있는 선반입 기법을 함께 사용한다. 선반입 기법은 미리 알 수 있는 데이터 참조 패턴

을 이용하여 디스크 대역폭에 의한 수용 사용자의 제한을 서버의 버퍼에 의한 제한으로 전환함으로써 더 많은 사용자들을 수용할 수 있었다.

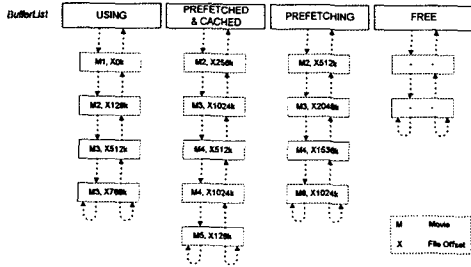
본 논문의 구성은 다음과 같다. 2 장에서는 주문형 비디오 서버에서 자원 활용률을 향상시킬 수 있는 새로운 버퍼 관리 기법을 제안한다. 3 장에서는 모의 실험을 통해 새로운 버퍼 관리 기법의 성능을 분석하였으며, 마지막으로 4 장에서 결론을 맺는다.

## 2. CAP 버퍼 관리 기법

본 장에서는 주문형 비디오 서버에서 시스템 자원의 활용률을 향상시키고 주문형 비디오 시스템이 더 많은 사용자를 수용할 수 있는 새로운 버퍼 관리 기법을 제안한다. 제안한 버퍼 관리 기법은 데이터에 대한 참조 패턴을 기반으로 하여 데이터 블럭에 대한 캐싱(caching)과 선반입(prefetching) 기법을 모두 사용함으로써 주문형 비디오 서버에서 효율적인 버퍼 관리를 할 수 있다. 이후 본 논문이 제안하는 버퍼 관리 기법을 CAP(Caching And Prefetching) 이라 부른다.

### 2.1. CAP 버퍼 관리 기법의 개요

주문형 비디오 서버에서 버퍼를 관리하는 기법은 크게 각 사용자들에게 독립적인 버퍼를 할당하여 사용하도록 하는 지역 버퍼 관리 기법과 모든 사용자들이 버퍼 캐시를 공유하는 전역 버퍼 관리 기법으로 나눌 수 있다. 지역 버퍼 관리 기법의 문제점은 각 사용자마다 버퍼를 할당하고 지역적으로 관리함으로써 같은 데이터 블럭을 참조하는 다른 사용자가 있을 경우에도 이들 사용자간의 데이터 공유를 할 수 없다는 것이다. 전역 버퍼 관리 기법의 경우 데이터 소비율이 크게 차이가 나는 사용자들이 함께 버퍼를 사용할 때 데이터 소비율이 높은 사용자의 데이터 블럭들이 버퍼를 차지할 수 있다. 그러나 주문형 비디오 시스템의 경우, 여러 사용자들의 데이터 소비율이 비슷하고 데이터 블럭에 대한 미래 참조 패턴을 비교적 정확하게 예측할 수 있기 때문에 한 사용자의 데이터 블럭에 의해 버퍼가 독점되는 경우는 매우 드물다. 이러한 이유로 최근의 연구들에서는 전역 버퍼 관리 기법을 사용하고 있으며 [8,9], 본 논문에서 제안하는 CAP 기법도 전역 버퍼 관리 기법을 사용하고 있다.



〈그림 1〉 CAP 버퍼 관리 기법의 버퍼 관리 구조

〈그림 1〉은 CAP 기법에서 관리하는 버퍼 풀(buffer pool)의 구조를 보이고 있다. 주문형 비디오 서버에서는 연속 매체의 재생이 가지는 주기적인 특성으로 인하여 주어진 주기동안 필요로 하는 데이터를 저장 장치로부터 읽어와서 버퍼에 저장해 두어야 한다. 동시에 사용자들이 요구하는 데이터 블록을 사용자 시스템으로 전송하여야 한다. 따라서 주문형 비디오 서버는 한 주기동안 저장 장치로부터 읽어들이는 데이터 블록과 사용자 시스템으로 전송되는 데이터 블록을 유지할 수 있어야 한다. 저장 장치로부터 읽어들이는 데이터 블록을 저장하는 입력 버퍼와 사용자 시스템에 전송되는 데이터 블록을 유지하는 출력 버퍼의 역할을 주기 단위로 바꾸어 가면서 버퍼를 관리하는 이중 버퍼링(double buffering) 구조는 주문형 비디오 서버에서 사용할 수 있는 간단한 버퍼 구조이다. CAP 기법에서는 〈그림 1〉의 prefetching 버퍼 리스트와 using 버퍼 리스트가 이중 버퍼링 구조 하에서의 입력 버퍼와 출력 버퍼에 해당한다. CAP 기법에서는 prefetching 버퍼 리스트와 using 버퍼 리스트와 함께 prefetched & cached 버퍼 리스트를 유지하고 있다. prefetching 버퍼 리스트는 오프 라인으로 알려진 데이터 블록에 대한 참조 패턴을 바탕으로 앞으로 사용될 데이터 블록들 중 미리 비디오 서버의 버퍼에 읽혀져진 데이터 블록들과 사용된 데이터 블록들 중에서 다시 다른 사용자에게 사용될 것이 결정된 데이터 블록들을 유지하고 있다. prefetching 버퍼 리스트는 현 주기동안 저장 장치로부터 읽혀지고 있는 데이터 블록들을 유지하는 버퍼들인데, 여기에 저장된 데이터 블록들은 다음 주기 이후에 사용되어진다. 따라서 CAP 기법에서 유지하는 버퍼 리스트들은 사용 시점에 따라 free → prefetching → prefetched & cached → using의 순서로 관리된다.

## 2.2. 최소 선반입 기법

MPEG 기법으로 압축된 자원의 데이터 소비 형태는

매우 가변적인 크기의 프레임들로 구성된다. 가변적인 데이터 소비 형태의 실시간 처리를 보장하기 위해 자원을 예약하는 방법은 크게 3 가지로 나누어 생각할 수 있다 [10]. 첫 번째 방법은 데이터의 실시간 처리를 보장하기 위해 최고의 데이터 소비율에 맞춰서 자원을 예약하는 방법이다. 이 방법은 자원 예약이 매우 간단하다는 장점이 있지만 데이터 소비량의 최고값으로 예약함으로써 많은 자원들을 낭비하게 된다. 주문형 비디오 시스템에서 주어진 자원을 효율적으로 사용하지 못하는 것은 바로 수용할 수 있는 사용자의 수를 제한하게 된다. 두 번째 방법은 시스템 자원을 데이터 소비량에 맞춰 매우 동적으로 할당하는 방법이다. 이 방법은 매 주기별 데이터 소비량에 따라 자원을 할당하기 때문에 최적으로 자원을 사용할 수 있다. 세 번째 방법은 몇 개의 단위 시간 간격을 합쳐서 자원 할당 구간을 적당한 구간으로 정하고 그 구간 별로 자원을 할당하는 방법이다. 이 방법은 앞의 두 방법의 절충식이라 할 수 있다.

CAP 기법은 최적 자원 예약을 사용하면서 서버의 버퍼를 이용하여 데이터 요구 형태를 완만한 형태로 변환시켜 자원의 사용을 예약하게 된다. 압축된 비디오 데이터의 가변성은 오프라인으로 미리 알려진 과부하 영역의 데이터를 선반입함으로써 줄일 수 있다. 서버의 버퍼를 이용해 평활화된 데이터 스트림을 저장 장치에 요구하는 것은 디스크 대역폭의 활용률을 증가시킨다. 이 때, 비디오 서버는 데이터 블록에 대한 참조 형태를 미리 알고 있기 때문에 각각의 데이터 블록들을 사용될 순서에 따라 버퍼로 읽어들이 수 있다.

CAP 기법은 가변 비트율의 비디오 데이터를 처리하기 위한 버퍼 예약을 위해 시간을 일정한 간격인 슬롯(slot)으로 나누어 슬롯 단위로 버퍼의 예약 여부를 검사하게 된다. CAP 기법을 설명하기 위해 비디오 서버가 현재  $n$  명의 사용자를 수용하고 있고 각 사용자  $C_i$  ( $0 \leq i \leq n-1$ )는 서로 다른 영화  $M_i$ 를 검색하고 있다고 가정한다.  $C_i$ 를 위해 슬롯  $t$  동안에 제공되어야 할 데이터 블록의 수를  $C_i(t)$ 라고 할 때, 슬롯  $t$  동안 모든 사용자들에 의해 소비되는 데이터 블록의 수  $C(t)$ 는 다음과 같은 식으로 나타낼 수 있다.

$$C(t) = \sum_{i=0}^{n-1} C_i(t) \quad (1)$$

한 슬롯 동안 비디오 서버가 읽어들이 수 있는 최소 데이터 블록의 수를  $D_{min}$ 이라고 하고 모든 사용자들을 위한 서비스 시작 슬롯이  $\alpha$ , 서비스가 끝나는 슬롯이  $\omega$ 라 하면, 모든 사용자들의 요구를 지연 없이 실시간에 처리하기 위해서는 다음 식과 같은 디스크 대역폭에 대한 조건이 만족되어야 한다.

$$C(t) \leq D_{\min} \quad (\alpha \leq t \leq \omega) \quad (2)$$

또 비디오 서버가 가지고 있는 총 버퍼의 수를  $B_{\max}$  라고 하면, 비디오 서버의 버퍼는 사용자들을 서비스하는 모든 슬롯동안 사용자들이 소비하고 있는 데이터 블록들과 저장 장치로부터 읽혀지고 있는 데이터 블록들을 저장할 만큼의 공간은 확보할 수 있어야 한다. 이러한 조건은 다음 수식으로 표현될 수 있다.

$$C(t) + C(t+1) \leq B_{\max} \quad (\alpha \leq t \leq \omega-1) \quad (3)$$

따라서 비디오 서버에 새로운 사용자  $C_n$ 이 도착했을 때  $C_n$ 의 수용 여부는 디스크 대역폭과 비디오 서버의 버퍼에 대한 검사를 모두 만족해야하기 때문에 수용 제어 기준(admission control criterion)은 새로운 사용자의 데이터 소비량을 포함하여 새로 계산된 식 (1)을 사용하여 식 (2)와 식 (3)을 동시에 만족되어야 한다.

본 논문에서 제안하는 선반입 기법을 적용하기 위해 먼저 과부하 슬롯(overflow slot)을 다음과 같이 정의한다.

**정의 1 (과부하 슬롯)** 모든 사용자들을 위한 서비스 시작 슬롯이  $\alpha$ , 서비스가 끝나는 슬롯이  $\omega$ 라 하면, 슬롯  $t$  ( $\alpha \leq t \leq \omega$ )가 다음 식을 만족하면 슬롯  $t$ 를 영역  $[\alpha, \omega]$ 에서의 과부하 슬롯이라 한다.

$$C(t) \geq D_{\min} \quad (4)$$

선반입 기법으로 과부하 슬롯을 포함하는 데이터 요구를 수용하기 위해서는 과부하 슬롯들의 데이터를 버퍼에 적재되는 시간을 최소로 하면서 언제, 얼마나 선반입해야 할 것인가를 결정해야 한다. 이 때 선반입되는 데이터 블록들은 사용되기까지 최소한의 시간동안 적재되어 있도록 선반입되어야 한다. 데이터 블록의 버퍼 적재 시간을 최소화하면서 과부하 슬롯의 데이터 소비량을 만족하기 위해 선반입 작업을 시작해야 하는 슬롯은 다음과 같이 정의할 수 있다.

**정의 2 (과부하 슬롯의 선반입 시작 슬롯)** 슬롯  $t$  ( $\alpha \leq t \leq \omega$ )가 영역  $[\alpha, \omega]$ 에서의 과부하 슬롯일 때 다음 식을 만족하는 최대의 슬롯  $\phi$  ( $\leq t$ )를 과부하 슬롯  $t$ 의 선반입 시작 슬롯이라 한다.

$$(t - \phi + 1) \times D_{\min} \geq \sum_{i=\phi}^t C(i) \quad (5)$$

과부하 슬롯을 가지는 주어진 영역 내에서 모든 과부하 슬롯을 처리하기 위해 선반입을 시작해야 할 슬롯은 구하기 위해서는 다음과 같이 정의되는 임계 슬롯(critical slot)을 구해야 한다.

**정의 3 (임계 슬롯)** 슬롯  $\alpha, \omega$  ( $\alpha \leq \omega$ )에 대해 영역  $[\alpha, \omega]$  내의 과부하 슬롯들 중에서 시간적으로 가장 앞서는 선반입 시작 슬롯을 가지는 과부하 슬롯  $c'$ 을 영역

$[\alpha, \omega]$ 의 임계 슬롯이라 한다.

승인 제어 기준 중에서 식 (2)에 의한 디스크 대역폭에 대한 제한 조건은 과부하 슬롯의 초과된 데이터 블록들을 선반입함으로써 완화할 수 있다. 즉, 디스크 대역폭을 최대한 활용하였을 때 비디오 서버의 버퍼에 적재하여야 하는 데이터 블록의 양을 결정하고 이를 비디오 서버의 버퍼가 수용할 수 있는 지 여부를 결정함으로써 새로운 사용자에 대한 승인 여부를 정한다.

**정의 4 (최소 필요 버퍼량)** 사용자들의 데이터 소비량의 집합  $C$ 에 대해 영역  $[\alpha, \omega]$  내의 슬롯  $t$ 에서 데이터 블록의 버퍼 적재 비용을 최소화하면서 모든 사용자들의 서비스를 보장할 수 있는 최소의 버퍼량을 영역  $[\alpha, \omega]$ 에서 집합  $C$ 의 최소 필요 버퍼량이라 한다.

사용자들의 데이터 소요량  $C(t)$ 로부터 최소 필요 버퍼량  $B_{\min}$ 을 구할 수 있고,  $B_{\min}(t)$ 를 이용하여 슬롯  $t$ 에서 비디오 서버에 의해 버퍼에 읽혀져야 할 최소의 블록 수인  $P_{\min}(t)$ 를 다음 식에 의해 구할 수 있다.

$$P_{\min}(t) = \min ( B_{\min}(t+1), D_{\min} ) \quad (6)$$

### 2.3. 최대 선반입 기법

비디오 서버는 데이터 블록에 대한 참조 형태를 미리 알 수 있으므로 서버의 버퍼와 디스크 대역폭이  $B_{\min}(t)$ 를 만족하고 남은 경우 이를 다음 블록의 선반입을 위해 사용할 수 있다.  $B_{\min}(t)$ 를 만족하기 위한 선반입이 사용자들의 요구를 만족시키기 위해 반드시 필요한 선반입인 것에 비해, 여분의 디스크 대역폭과 버퍼를 이용한 선반입은 비디오 서버의 상태를 더 많은 사용자 요구를 수용할 수 있는 상태로 만들기 위한 선반입이다. CAP 기법은 현실적으로 가능한 최대의 블록들을 사용 순서에 따라 버퍼에 반입함으로써 비디오 서버의 자원 활용률을 최대화하였다. 따라서 CAP 버퍼 관리 기법은 다음과 같은 두 가지 목표를 만족하도록 버퍼의 상태를 유지하게 된다.

- 현실적으로 가능한 최대 블록수 반입
- 소비 시점이 먼 블록부터 반출, 소비 시점이 가까운 블록부터 반입

여분의 디스크 대역폭과 버퍼를 이용하여 최대한의 블록을 반입하기 위해서는 비디오 서버의 버퍼가 허용하는 한도 내에서 최대의 디스크 대역폭으로 데이터 블록들을 버퍼로 읽어들이어야 한다. 슬롯  $t$ 에서 버퍼 한도 내에서 최대한 읽을 수 데이터 블록의 수는 다음 식에 의해 계산될 수 있다.

$$P_{\max}(t) = \min ( B_{\max} - B(t), D_{\min} ) \quad (7)$$

또 각 슬롯에 대해 버퍼에 적재 가능한 최대 블록 수

$B_{max}(t)$ 는 다음의 재귀 수식으로 나타낼 수 있다.

$$B_{max}(t+1) = B_{max}(t) + P_{max}(t) - C(t), \quad B_{max}(0) = 0 \quad (8)$$

### 3. 실험 및 평가

#### 3.1. 실험 환경

본 장에서는 CAP 기법의 성능을 평가하기 위해 트레이스 데이터(trace data)를 기반으로 한 모의 실험의 결과를 보인다. 모의 실험에서는 MPEG-1으로 압축된 비디오 데이터에 대한 트레이스 데이터를 사용하였다 [11]. 모의 실험에서는 주문형 비디오 시스템에 포아송 분포(poisson distribution)를 따르는 사용자 요구가 도착하여 원하는 영화를 선택하면 비디오 서버는 이를 정해진 시간 동안 사용자 시스템에 전달하여 사용자가 출력 장치를 통해 영화를 상영하는 환경을 가정하였다. 일반적으로 영화에 대한 인기도에 따라 그 검색 정도가 결정되기 때문에 비디오 서버에 저장된 영화들이 서로 다른 인기도를 가지도록 함으로써 영화에 대한 참조 회수를 달리 가지도록 하였다.

비교 대상으로 사용된 버퍼 관리 기법은 LRU 기법이며, 가변 비트율로 압축된 비디오 데이터에 대한 검색 방법(항등 비트율 검색과 가변 비트율 검색)에 따른 성능 비교와 사용자 시스템의 버퍼를 이용한 대역폭 평활화 기법의 사용 여부에 따른 성능 비교를 함께 수행하였다. LRU 기법이 데이터 블록이 참조된 시점이라는 과거의 정보를 사용하여 교체할 데이터 블록을 정하는 데 비해, CAP 기법은 미리 분석한 데이터 블록에 대한 참조 패턴을 근거로 하여 교체할 데이터 블록을 결정한다. 또한 CAP 기법은 여분의 디스크 대역폭과 서버 버퍼를 이용하여 앞으로 사용할 데이터 블록들을 선반입함으로써 요구가 물리는 구간의 데이터 블록에 대한 요구를 고르게 분포시킬 수 있다. 대역폭 평활화 기법은 사용자 시스템의 버퍼를 이용하여 전송 데이터의 가변성을 줄일 수 있는데, 이 기법은 서버의 버퍼 관리 기법과 무관하게 사용될 수 있다. 본 모의 실험에서는 대역폭 평활화 기법의 사용 여부에 따라 CAP 기법과 LRU 기법의 성능이 어떻게 변화하는지를 조사하였다. 또한, CAP 기법을 LRU 기법과 비교함에 있어 가변 비트율로 압축된 비디오 데이터를 검색할 때 검색 방법이 본 논문에서와 같이 슬롯 단위로 예약하는 가변 비트율 검색(Variable Bit Rate retrieval)을 사용할 때와 자원을 최악의 경우에 맞추어 예약하는 항등 비트율 검색(Constant Bit Rate retrieval)을 사용할 때의 성능도 함께 비교하였다. 따라서 본 모의

실험에서는 CAP 기법과 LRU 기법을 대역폭 평활화 기법의 사용 여부 및 가변 비트율 검색의 사용 여부에 따라 <표 1>과 같은 8가지의 경우에 대해서 성능을 비교하였다.

<표 1> 모의 실험에서의 정책 비교

정 책	CAP/LRU	대역폭 평활화 기법	검색방법
CAP-S-V	CAP	○	VBR
CAP-S-C	CAP	○	CBR
CAP-NS-V	CAP	×	VBR
CAP-NS-C	CAP	×	CBR
LRU-S-V	LRU	○	VBR
LRU-S-C	LRU	○	CBR
LRU-NS-V	LRU	×	VBR
LRU-NS-C	LRU	×	CBR

#### 3.2. 성능 평가 항목

주문형 비디오 시스템의 가장 중요한 성능 평가 항목은 동시에 수용 가능한 사용자 수인데, 주어진 시스템 자원으로 최대한 많은 사용자들을 수용할 수 있어야 한다. 또한 본 모의 실험에서는 동시에 수용 가능한 사용자 수와 함께 비디오 서버의 대표적인 자원인 디스크 대역폭과 비디오 서버의 버퍼에 대해 이들 자원이 고정된 값으로 주어졌을 때 각각의 자원 활용률도 함께 조사함으로써 자원 활용률과 수용 가능한 사용자 수 간의 상관 관계를 조사하였다.

##### • 자원 활용률 Uresource

자원 활용률이란 사용 가능한 자원에 대한 실제 사용한 자원의 비율로 정의한다. 비디오 서버에서 중요한 자원은 데이터 블록들을 저장 장치에서 읽어들이는 디스크 대역폭과 읽은 데이터 블록이 저장되는 서버의 버퍼이다. 따라서 본 논문의 모의 실험에서는 성능 평가 항목으로 디스크 대역폭의 활용률과 서버 버퍼의 활용률을 사용하였다.

##### • 동시에 수용 가능한 사용자 수 Nclient

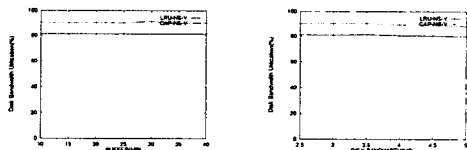
일반적으로 주문형 비디오 시스템에 도착하는 사용자들의 도착 분포는 시간의 변화에 따라 다르다. 즉, 아침보다는 저녁이나 밤에 더 많은 사용자들이 영화를 요구할 수 있다. 본 모의 실험에서는 주문형 비디오 시스템에 도착하는 사용자 요구들의 도착 분포가 포아송 분포를 따른다고 가정하였으며, 시간 간격을 몇 개의 구간으로 나누어 각 시간 간격 내에서는 서로 다른 평균 도착 시간을 가지는 포아송 분포를 가정하였다. 또한 슬롯의 크기는 독립적으로 복원이 가능한 단위인 GOP를 재생하는 데 필요한 시간으로 하였으며, 10 ~ 30 슬롯 간격으로 새로운 요구가 시스템에 도착하도록 하였다. 모의 실험에 사용한 비디오 데이터의 특성으로 인해 특정 시간에 주문형

비디오 서버가 동시에 수용할 수 있는 사용자 수는 항상 일정하지 않기 때문에 실험에서는 각 시간에 수용된 사용자 수의 평균을 계산하여 이를 평가 항목으로 사용하였다.

### 3.3. 자원 활용률

#### 3.3.1. 디스크 대역폭의 활용률

<그림 2>는 사용자 시스템에서 대역폭 평활화 기법을 사용하지 않고 가변 비트율로 검색할 때 비디오 서버에서 버퍼 관리 정책으로 CAP 정책과 LRU 정책을 적용하였을 경우의 디스크 대역폭의 활용률을 보이고 있다. <그림 2>(a)는 일정한 크기의 서버 버퍼에 대해 디스크 대역폭의 변화에 따른 각 정책의 디스크 대역폭에 대한 활용률을 보이고 있으며, <그림 2>(b)는 디스크 대역폭이 일정할 때 비디오 서버의 버퍼의 크기 변화에 따른 디스크 대역폭의 활용률을 보이고 있다. 두 경우 모두 CAP 기법이 LRU 기법보다 디스크 대역폭을 약 10 % 정도 더 사용하고 있음을 알 수 있다.

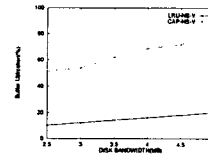


(a) 서버 버퍼 크기가 일정 (40 MB)      (b) 디스크 대역폭이 일정 (2.5 MB)

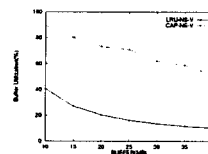
<그림 2> CAP-NS-V 기법과 LRU-NS-V 기법의 디스크 대역폭 활용률

#### 3.3.2. 서버 버퍼의 활용률

<그림 3>은 대역폭 평활화 기법을 사용하지 않으면서 가변 비트율 검색을 할 때 비디오 서버에서 버퍼 관리 정책으로 CAP 정책과 LRU 정책을 적용하였을 경우의 서버 버퍼의 활용률을 보이고 있다. <그림 3>(a)는 서버 버퍼의 크기가 일정할 때 디스크 대역폭의 변화에 따른 각 정책의 서버 버퍼의 활용률을 보이고 있으며, <그림 3>(b)는 디스크 대역폭이 일정할 때 비디오 서버의 버퍼의 크기 변화에 따른 서버 버퍼의 활용률을 보이고 있다. <그림 3>의 서버 버퍼 활용률을 <그림 2>의 디스크 대역폭 활용률과 비교해 볼 때 디스크 대역폭 활용률보다 서버 버퍼 활용률에서 CAP 기법이 LRU 기법과 큰 차이를 보이고 있다. 이는 CAP 기법이 새로운 사용자에 대한 수용 가능성을 높이기 위해 서버 버퍼에 가능한 최대의 데이터 블록들을 선반입하기 때문이다.



(a) 서버 버퍼 크기가 일정 (40 MB)



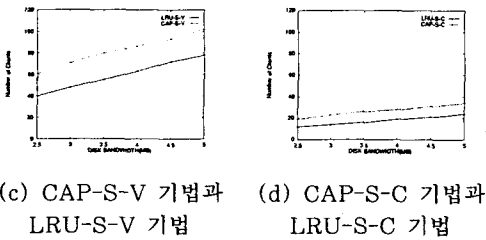
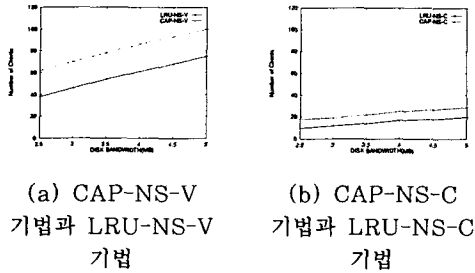
(b) 디스크 대역폭이 일정 (2.5 MB)

<그림 3> CAP-NS-V 기법과 LRU-NS-V 기법의 서버 버퍼 활용률

### 3.4. 수용 사용자 수

주문형 비디오 서버의 가장 중요한 성능 평가 항목은 서버가 처리할 수 있는 사용자 수이다. <그림 4>는 비디오 서버 버퍼의 크기가 일정할 때 디스크 대역폭이 변함에 따라 CAP 기법과 LRU 기법이 동시에 수용할 수 있는 사용자 수가 어떻게 변하는지를 보이고 있다.

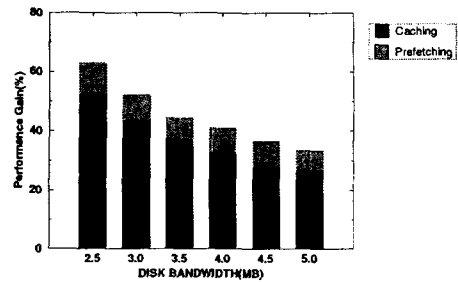
서버 버퍼의 크기가 40 MB로 일정할 때, CAP 기법과 LRU 기법을 비교하였다. <그림 4>(a)는 대역폭 평활화 기법을 사용하지 않고 가변 비트를 검색을 사용하였을 경우이고 <그림 4>(b)는 대역폭 평활화 기법을 사용하지 않고 고정 비트를 검색을 사용하였을 경우이다. 고정 비트를 검색의 경우 자원을 과예약함으로써 실제 사용할 수 있는 자원의 양이 줄어들기 때문에 가변 비트를 검색에 비해 사용자들을 30 % 정도만 수용하였다. 가변 비트를 검색에 대해 대역폭 평활화 기법만 사용했을 때에는(<그림 4>(c)에서 LRU-S-V의 경우), CAP 기법을 사용하였을 때 보다 25 ~ 50 % 정도 수용 사용자가 감소함을 알 수 있다.



<그림 4> 디스크 대역폭의 변화에 따른 CAP 기법과 LRU 기법의 수용 사용자 수

### 3.5. 선반입과 캐싱의 효과 분석

<그림 5>는 <그림 4>(a)에서 CAP 기법의 LRU 기법에 대한 성능 향상 부분을 캐싱에 의한 부분과 선반입에 의한 부분으로 나누어 보여주고 있다. <그림 5>에서 알 수 있듯이 일정 크기의 서버 버퍼를 사용하였을 경우 디스크 대역폭이 증가함에 따라 동시에 수용 가능한 사용자 수는 증가하지만 CAP 기법의 LRU 기법에 대한 상대적인 성능은 줄어들고 있음을 알 수 있다. 그 이유는 디스크 대역폭은 증가하지만 서버 버퍼의 크기가 일정하기 때문에 버퍼에 의한 제한으로 성능이 완만하게 증가하게 된다. 또한, <그림 5>에서는 캐싱을 사용하지 않고 선반입만 사용하였을 경우와 캐싱과 선반입을 모두 사용한 경우를 함께 보이고 있는데, 선반입 기법에 의한 효과는 CAP 기법을 사용함으로써 LRU 기법을 사용했을 때 디스크 대역폭에 의한 제한으로 수용할 수 없었던 사용자들을 더 수용할 수 있게 된다. CAP 기법의 성능 향상 요인 중 캐싱에 의한 효과가 선반입에 의한 효과보다 크다는 것을 알 수 있다. 선반입에 의한 효과는 디스크 대역폭이 증가함에 따라 거의 일정한데 비해 캐싱에 의한 효과는 감소하는데, 선반입에 의한 효과는 2.5 MB의 디스크 대역폭에서 이미 충분히 이루어졌기 때문에 디스크 대역폭이 증가한다고 해도 데이터 가변성에 의한 디스크 대역폭 제한을 받지 않는다.



<그림 5> CAP 기법의 LRU 기법에 대한 성능 향상(서버 버퍼 크기가 일정)

### 4. 결론

통신과 방송 및 인쇄 매체간의 벽이 점차 무너지면서 기존의 정보 매체들이 새로운 매체들로 바뀌고 있다. 현재 사용중인 아날로그 방송은 점진적으로 디지털 방송으로 바뀌어 될 것이고 대화형 텔레비전이나 주문형 비디오 시스템의 사용도 보편화될 것이다. 주문형 비디오 시스템에서 비디오 서버의 성능은 전체 시스템의 성능을 크게 좌우하게 된다. 연속 매체를 저장하고 처리하는 비디오 서버에서 주어진 자원을 효율적으로 관리함으로써 더 많은 사용자들을 동시에 수용하는 것은 주문형 비디오 서버에서의 궁극적인 자원 관리의 목표일 수 있다.

본 논문에서는 가변 비트율의 비디오 데이터를 지원하는 주문형 비디오 서버에서의 버퍼 관리 기법을 제안하였다. 제안된 기법은 비디오 데이터의 가변성을 오프라인으로 미리 계산하고 이를 이용하여 디스크 대역폭을 초과하는 과부하 영역의 데이터를 선반입함으로써 비디오 데이터의 가변성을 줄일 수 있을 뿐만 아니라 가능한 최대의 데이터 불럭을 선반입함으로써 관련 자원들의 활용률을 극대화하여 비디오 서버가 더 많은 사용자들의 요구를 수용할 수 있었다. 본 논문에서 제안한 버퍼 관리 기법은 주문형 비디오 시스템 뿐만 아니라 기 저장된 데이터를 사용하는 시스템에서 자원의 활용률을 높이는 데 사용될 수 있다.

### 참고 문헌

[1] Marwan Krunz and Herman Hughes, "A Traffic Model for MPEG-Coded VBR Streams", In Proceedings of the 1995 ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems,

pages 47-55, May 1995.

[2] Thomas D. C. Little and Dinesh Venkatesh, "Prospects for Interactive Video-On-Demand", IEEE Multimedia, 1(3):14-24, 1994.

[3] Lixia Zhang, Stephen Deering and Deborah Estrin. RSVP: A New Resource ReSerVation Protocol. IEEE network, 7(5):8-18, September 1993.

[4] D. Ferrari and D. C. Verma. A Scheme for Real Time Channel Establishment in Wide Area Networks. IEEE Journal on Selected Areas in Communications, 8(3):368-379, April 1990.

[5] Phillip Lougher and Doug Shepherd. The Design of a Storage Server for Continuous Media. The Computer Journal (special issue on multimedia), 36(1):32-42, February 1993.

[6] P. V. Rangan and H. M. Vin. Designing File Systems for Digital Video and Audio. ACM Operating Systems Review, 25(5):81-94, October 1991.

[7] D. Anderson, Y. Osawa and R. Govindan. A File System for Continuous Media, ACM Transactions on Computer Systems, 10(4):311-337, November 1992.

[8] B. Ozden and R. Rastogi and A. Silberschatz, "A Framework for the Storage and Retrieval of Continuous Media Data", In International Conference on Multimedia Computing and Systems, IEEE Computer Society, May 1995.

[9] Craig S. Freedman and David J. DeWitt, "The SPIFFI Scalable Video-on-Demand System", In Proceedings of the ACM SIGMOD International Conference on Management of Data, pages 352-363, ACM Press, 1995.

[10] Hung-Ju Lee and Jyh-Cham Liu, "Real-Time Guaranteed Video Transport for Video-On-Demand Systems", In Proceedings of the International Conference on Multimedia Computing and Systems, pages 156-159, June 1996.

[11] Oliver Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems, Technical Report No. 101, University of Wuerzburg, February 1995.



김 순 철

1990년 2월 서울대학교 공과대학  
컴퓨터공학과 졸업(공학사)

1992년 2월 서울대학교 공과대학  
컴퓨터공학과 졸업(공학석사)

1998년 8월 서울대학교 공과대학  
컴퓨터공학과 졸업(공학박사)

1998년 9월 ~ 1999년 9월 서울대

컴퓨터신기술공동연구소 특별연구원

1999년 3월 ~ 현재 대구대학교 공과대학 컴퓨터정보공학  
부 전임강사

관심분야는 운영체제, 멀티미디어 시스템