

비디오 서버를 위한 예약 기반 하이브리드 장치 대역폭 감소 정책

이경숙^o 배인한

대구효성가톨릭대학교 컴퓨터정보통신공학부

{g6721001, ihbae}@cuth.cat.aegu.ac.kr

An Reservation-based Hybrid Device Bandwidth Reduction Policy for Video Servers

Kyung-Sook Lee^o Ihn-Han Bae

School of Computer and Information Communication Eng., Catholic University of Taegu-Hyosung

요 약

주문형 비디오 시스템에서 비디오 서버의 입출력 대역폭은 지연 시간을 증가시키는 원인이 되는 중요한 자원이다. 공유를 통하여 비디오 서버의 입출력 요청을 감소시키는 다수의 방법들: 일괄처리, 브리징, 피기백킹이 제안되었다. 일괄처리는 같은 객체에 대한 요청들을 묶어서, 기억장치 서버에 대한 하나의 입출력 요청을 만드는 것이고, 피기백킹은 객체의 대응하는 입출력 스트림들을 그룹으로 서비스할 수 있는 하나의 스트림으로 병합하기 위하여 진행중인 요청들의 디스플레이율을 변경하는 정책이다. 본 논문에서는 인기 있는 비디오에 대한 요청들이 가능한 한 스케줄될 수 있도록 비디오 서버의 입출력 스트림 용량을 예약해 두고 상관관계에 있는 두 요소: 스트림 요청의 지연시간과 입출력 요청의 감소를 적절히 절충한 하이브리드 장치 대역폭 감소 정책을 제안하고, 시뮬레이션을 통하여 제안하는 정책의 성능을 평가한다.

1. 서 론

최근 정보통신 분야의 기술적인 발전으로 주문형 비디오, 홈쇼핑 등과 같은 여러 가지 주문형 멀티미디어 시스템들이 실용화되고 있다. 오늘날의 정보 시스템은 단순히 커다란 멀티미디어 객체를 저장하고 검색하는 것뿐만 아니라 그 객체를 일정한 대역폭에서 계속적으로 제공하는 엄격한 실시간 요구사항을 만족시킬 수 있어야 한다. 멀티미디어 시스템은 교육용 응용, 오락 기술, 도서관 정보 시스템 등에서 중요한 역할을 하고 있으며, 이러한 시스템에서 가장 중요한 것은 다수의 클라이언트들에게 주문형 서비스를 동시에 제공하는 것이다. 즉, 사용자들은 비디오 등과 같은 객체를 요청하고, 적절한 지연 시간 내에 요청한 객체를 관촬하기를 기대한다. 여기서 지연 시간은 요청이 도착한 시점으로부터 시스템이 디스크로부터 객체 읽기를 초기화하는데 까지의 시간으로 정의되고, 데이터가 실제적으로 디스플레이 장치에 전달될 때까지의 추가적인 지연은 디스크 지연에 비해 상대적으로 작기 때문에 무시할 수 있다. 이러한 지연은 서비스 요청을 위한 불충분한 대역폭, 디스크로부터 판독 내용을 스케줄링 하기 위한 불충분한 비퍼 공간, 불충분한 디스크 기억장치 등의 요인으로 발생한다. 이러한 지연 요소 중에서 입출력 대역폭(I/O bandwidth)은 매우 중요한 자원이므로 공유를 통하여 기억장치 서버의 입출력 요청을 감소시켜 동시에 서비스할 수 있는 사용자 요청의 수를 증가시키는 다수의 접근 방법들: 일괄처리(batching), 브리징(bridging), 피기백킹(piggybacking)이 제안되었다[1, 2, 3, 4].

본 논문에서는 인기 있는 비디오 요청들이 가능한 한 스케줄되도록 비디오 서버의 입출력 스트림 용량을 예약해 두고, 상관관계에 있는 두 요소: 스트림 요청의 지연 시간과 입출력 요청의 감소를 적절히 절충할 수 있는 하이브리드 장치 대역폭 감소 정책을 제안하고, 그것의 성능을 시뮬레이션을 통하여 평가한다. 본 논문의 구성은 다음과 같다. 2장에서는 공유를 통한 저장 서버의 입출력 대역폭 감소에 대한 관련 연구들을 살펴보고, 3장에서는 본 논문에서 제안하는 하이브리드 장치 대역폭 감소 정책을 설명하고, 4장에서는 시뮬레이션을 통

하여 제안하는 알고리즘의 성능을 평가한다. 그리고 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

공유를 통하여 저장 서버의 입출력 요청을 감소시켜 동시에 서비스되는 사용자 요청의 수를 증가시키는 기존의 정책들을 간략히 살펴보면 다음과 같다.

2.1 일괄처리

일괄처리는 같은 객체에 대한 요청들을 묶어서, 기억장치 서버에 대한 하나의 입출력 요청을 만드는 것이다. 일괄처리 방법에서 스트림 요청의 지연 시간과 기억장치 서버의 입출력 요청의 감소는 서로 상관관계에 있다. 일괄처리 정책은 크기에 의한 일괄처리와 시간에 의한 일괄처리로 분류할 수 있다. 각 객체 c 에 대해 미리 정해진 요청 개수 B 를 일괄처리 윈도우라 할 때, 크기에 의한 일괄처리는 객체 c 에 대해 B 만큼의 요청 개수가 누적되면 객체 c 에 대해 입출력 스트림을 초기화한다. 시간에 의한 일괄처리는 시간 단위로 일괄처리 윈도우를 설정하는 것이다. 하나의 요청이 기억장치 서버에 도착하고 같은 객체 c 에 대한 요청이 존재하지 않으면 타이머가 설정된다. 일괄처리 윈도우 B , 동안 모아진 같은 객체 c 에 대한 요청은 타이머가 만기되면 하나의 입출력 스트림으로 서비스된다[2]. 좋은 비디오 스케줄링 정책은 일괄처리 윈도우뿐만 아니라 관촬자의 이탈 확률과 대기 시간을 고려해야 한다.

2.2 피기백킹

피기백킹이란 다수의 디스플레이 스트림을 하나의 입출력 스트림으로 서비스해 주는 방법으로 일괄처리는 정적 피기백킹이다. 동적 피기백킹 방법은 다음과 같이 수행된다. 요구에 따라 각 디스플레이 스트림에 대한 하나의 입출력 스트림이 초기화되고, 하나의 디스플레이 스트림을 같은 객체에 대한 다른 디스플레이 스트림의 입출력 스트림을 향하여 적응적으로 피기백킹되는 것을 허용한다. 이것은 두 개의 입출

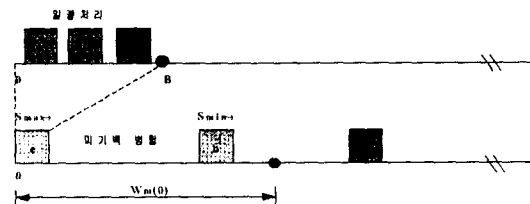
력 스트림을 하나로 병합하는 것으로 볼 수 있다. 병합 전에는 각각 한 개 이상의 디스플레이 스트림을 서비스하는 두 개의 입출력 스트림이 있다면, 병합 후에는 두 개의 디스플레이 스트림을 동시에 서비스하는 한개의 입출력 스트림이 있다. 이러한 병합은 요청의 디스플레이 속도를 조정함으로써 이루어질 수 있다. 병합이 빠를수록 더 많은 자원이 다른 요청들을 서비스하기 위해 사용될 수 있다[1].

3. 예약 기반 하이브리드 장치 대역폭 감소 정책

본 논문에서는 입출력 대역폭 감소 정책에서 상관관계에 있는 두 요소: 스트림 요청 지연 시간과 비디오 서버 입출력 요청 감소를 상호 보충하기 위하여 일괄 처리와 피기백킹을 혼합한 하이브리드 장치 대역폭 감소 정책을 제안한다.

3.1 하이브리드 장치 대역폭 감소 정책의 시나리오

본 논문에서 제안하는 예약 기반 하이브리드 장치 대역폭 감소 정책은 기존의 일괄처리 정책과 피기백킹 정책의 장점을 모두 수용한 정책이다. 일괄처리 정책은 일괄처리 윈도우가 만기가 된 다음, 비디오 요청의 스케줄을 시도하므로 많은 비디오 서버 입출력 요청을 감소시킬 수 있으나 스트림 요청 지연 시간이 길어진다 단점이 있고, 반면에 피기백킹은 도착하는 비디오 요청의 스케줄을 즉시 시도하므로 스트림 요청 지연 시간은 짧아지나 격차 해소 윈도우내에 같은 비디오에 대한 다수의 스트림이 존재하므로 감소되는 입출력 스트림 량이 일괄처리에 비해 작다는 단점이 있다. 따라서 제안하는 정책에서는 일반적인 일괄처리 정책에 비해 짧은 배치 윈도우를 설정하고, 일괄처리된 요청이 스케줄 될 때 격차 해소 윈도우내에 같은 비디오에 대해 앞서 느린 속도로 재생되고 일괄처리된 스트림이 있는지 검사하여 두 스트림의 병합을 시도한다. 그렇게 함으로써 스트림 요청 지연 시간과 서버 입출력 요청의 감소 모두에서 좋은 결과를 얻을 수 있다.



(그림 1) 하이브리드 장치 대역폭 감소 정책의 시나리오

하이브리드 정책의 시나리오는 그림 1과 같다. 여기서 i, j, k 는 요청 비디오를, a, b, c 는 입출력 스트림을, B 는 배치 윈도우를, 그리고 $W_m(0)$ 는 격차 해소 윈도우를 각각 나타낸다. 배치 윈도우안에 같은 비디오에 요청 i, j, k 는 하나의 비디오 요청으로 일괄처리 되어 입출력 스트림 c 로 스케줄 될 때, 같은 비디오에 대한 입출력 스트림 b 가 격차 해소 윈도우내에서 느린 속도(S_{min})로 재생되므로, 입출력 스트림 b 와 c 를 병합하기 위하여 입출력 스트림 c 를 빠른 속도(S_{max})로 재생한다.

3.2 예약 기반 하이브리드 장치 대역폭 감소 알고리즘

본 논문에서 제안하는 예약 기반 하이브리드 장치 대역폭 감소 알고리즘의 구조는 그림 2와 같다. 먼저 인기 있는 비디오에 대한 요청을 비디오 서버의 부하에 관계없이 가능하면 스케줄하기 위하여 인기 있는 비디오를 위한 비디오 서버 스트림 용량을 예약하고,

그리고 서버에 도착한 비디오 요청을 대기 큐에 넣고, 대기 큐의 선두에 있는 비디오 요청의 일괄처리 윈도우가 만기가 되면 대기 큐내의 같은 비디오에 대한 모든 요청을 하나의 요청으로 일괄처리 하여 하나의 입출력 스트림으로 스케줄을 시도한다. 이 때 요청 비디오가 인기 있는 비디오이면 서버의 예약 입출력 스트림을 할당하고, 아니면 서버의 비예약 입출력 스트림을 할당한다. 그리고 스케줄한 비디오와 같은 비디오 스트림이 느린 속도로 격차 해소 윈도우내에 재생되고 있으면 빠른 속도로 재생을 시작한다.

Algorithm Reservation-based Hybrid Device Bandwidth Reduction Algorithm(k)

```

Begin
  Reserve the I/O stream capacity of video server for
  k-hottest videos;
Case arrival of a video request:
  Put the arrival video request in waiting queue;
Case the batching window of the front request in
waiting queue is expired:
  If the requesting video is in the k-hottest video
  If the reserved stream capacity is not empty
  Begin
    Batch the same video requests in waiting queue
    into a single video request(i);
    If no stream within W is moving at S_min
      S_i=S_min;
    Else
      S_i=S_max;
  End
  Else
    Block the video request;
  Else
    If the unreserved stream capacity is not empty
    Begin
      Batch the same video requests in waiting queue
      into a single video request(j);
      If no stream within W is moving at S_min
        S_j=S_min;
      Else
        S_j=S_max;
    End
    Else
      Block the video request;
Case merge of i and j:
  Drop stream i;
  S_i=S_max;
End.
    
```

(그림 2) 예약기반 하이브리드 장치 대역폭 감소 알고리즘

4. 시뮬레이션 및 평가

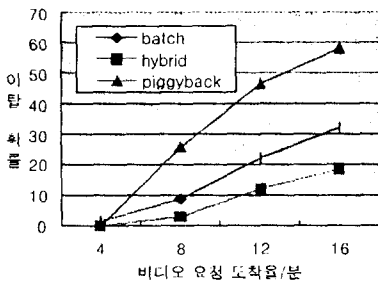
본 논문에서 제안하는 예약기반 하이브리드 장치 대역폭 감소 정책의 성능을 시뮬레이션을 통하여 평가하고, 그 결과를 예약기반 단순 병합 정책, 예약 기반 일괄처리 정책과 비교한다. 시뮬레이션에 사용된 매개변수는 표 1과 같다. 성능 평가를 위해 비디오 서버의 부하를 고려하여 3개의 인기 있는 비디오에 대한 비디오 서버의 입출력 스트림 용량을 동적으로 예약하는 예약기반 단순 피기백킹[6]과 배치 윈도우가 3분인 예약기반 일괄정책을 고려한다. 평가 항목은 비디오 요청 도착순에 따른 관찰자의 이탈 확률, 관찰자의 평균 대기 시간 그리고 프레임 전역 백분율이다. 여기서 i -번째 인기 있는 비디오의 인기도는 Zipf의 법칙[5]을 따른다고 가정한다. x 단위 시간 떨어진 디스플레이 스

트립들이 병합되는 시점은 $t_m = xS_{min} / (S_{max} - S_{min})$ 으로 계산된다. 그리고 시뮬레이션의 공정성을 기하기 위해 서비스를 시작할 후 100분~200분 사이의 실험 결과로 평가한다.

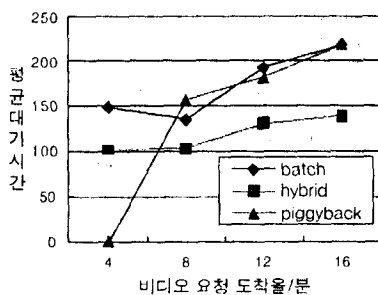
<표 1> 시뮬레이션 매개변수

매개변수	값
비디오의 개수	60 개
비디오 서버의 입출력 스트림 용량	400
일기 있는 영화의 개수	3 개
예약 입출력 스트림 용량	90
비디오 요청 도착율	4, 8, 12, 16 요청/분(포아송)
느린 디스플레이율 (S_{min})	28.5 프레임/초
빠른 디스플레이율 (S_{max})	31.5 프레임/초
영화의 길이	100 분
일괄처리 윈도우(B)	2분
최대 검색 해상도 윈도우(W_m)	9.5 분
서비스 이탈 시간	3~5분 (랜덤)
시뮬레이션 시간	100 분

그림 3은 비디오 요청 도착율에 따른 서비스의 이탈 확률을 나타낸다. 여기서 이탈 확률은 서비스 이탈 회수를 전체 비디오 요청 회수로 나눈 것이다. 비디오 요청 도착율에 상관없이 제안하는 예약기반 하이브리드 장치 대역폭 감소 정책(hybrid)이 낮은 이탈 확률을 제공함을 알 수 있다.



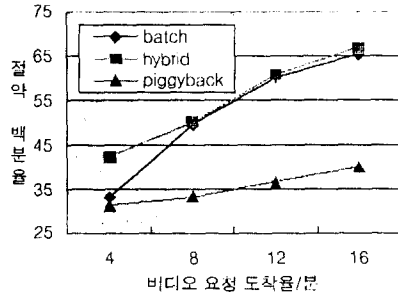
(그림 3) 비디오 요청 도착율에 따른 서비스 이탈 확률



(그림 4) 비디오 요청 도착율에 따른 평균 대기 시간

그림 4은 비디오 요청 도착율에 따른 서비스의 평균 대기 시간의 변화를 보여준다. 여기서 평균 대기 시간은 요

청된 비디오의 대기 시간의 합을 총 요청된 비디오의 개수로 나누어 계산하였다. 피기백킹은 비디오 요청이 도착하면, 일괄처리 윈도우 만큼 기다리지 않고 서버 스트림 용량이 있으면 즉시 스케줄링 되므로 짧은 대기 시간을 제공한다. 그러나 비디오 요청 도착율에 높아질수록 서비스 대기시간이 점점 길어진다. Hybrid는 저부하인 경우를 제외한 모든 서버 부하에서 좋은 서비스 대기시간을 제공한다.



(그림 5) 비디오 요청 도착율에 따른 프레임 길이 백분율

그림 5는 비디오 요청 도착율에 따른 정책별 프레임 길이 백분율을 보여준다. 프레임 길이 백분율은 절약된 프레임의 합을 전체 처리된 프레임의 합으로 나눈 백분율이다. 여기서도 제안하는 hybrid가 비디오 요청 도착율에 관계없이 항상 우수한 프레임 길이를 제공함을 알 수 있다.

5. 결론

본 논문에서는 일괄처리와 피기백킹의 장점들만 수용한 하이브리드 장치 대역폭 감소 정책을 제안하고, 그것의 성능을 시뮬레이션을 통하여 평가하였다. 그 결과 모든 성능 평가 항목: 이탈확률, 대기시간, 프레임 절약율에서 본 논문에서 제안하는 하이브리드 정책이 일괄처리 정책과 피기백킹 정책에 비해 좋은 성능이 제공하였다. 따라서 하이브리드 정책은 주문형 비디오 서버의 성능 뿐만 아니라 QoS도 향상시키는 우수한 장치 대역폭 감소 정책이라는 것을 확인하였다. 향후 연구과제는 비디오 서버의 부하에 상관없이 적절한 수준의 이탈율을 유지하는 동적 일괄처리 정책에 관한 것이다.

참고문헌

- [1] L. Golubchik, J. Lui, and R. Muntz "Reducing I/O Demand in Video-On-Demand Storage Servers." *In Proceedings of ACM Sigmetrics '95*, pp. 25~36, May 1995.
- [2] A. Dan, D. Sitaram, and P. Shahabuddin "Scheduling Policies for an On-Demand Video Server with Batching." *In Proceedings of the 2nd ACM Multimedia Conference*, pp.25~32, 1994.
- [3] M. Kamath, K. Ramamritham, and D. Towsley, "Continuous Media Sharing in Multimedia Database Systems." Department of Computer Science, University of Massachusetts, Technical Report 94-11, Feb. 1994.
- [4] A. Dan, D. M. Dias, R. Mukherjee, and D. Sitaram. "Buffering and Caching in Large-Scale Video Servers." IBM Research Division, Technical Report RC 19903, Jan. 1995.
- [5] G. K. Zipf, *Human Behavior and the Principles of Least Effort*, Addison-Wesley, 1949.
- [6] 김윤정, 배인한, "주문형 비디오 서버를 위한 예약 기반 Piggybacking Merging 정책," *한국정보과학회 '97 가을 학술발표논문집(IV)*, pp. 115~118, 1997. 10.