

실시간 멀티미디어 저장 서버에서 초기 지연시간 최소화를 위한 동적 스트림 합병 기법

김 근 혜, 최 황 규

dragon4@miraе.kangwon.ac.kr, hkchoi@cc.kangwon.ac.kr

강원대학교 컴퓨터·정보통신공학부

Dynamic Stream Merging Technique for Reducing Initial Latency in Real Time Multimedia Storage Servers

Geun Hye Kim, Hwang Kyu Choi

Division of Computer, Information & Communications Engineering

Kangwon National University

요 약

본 논문은 다수의 사용자에 대한 실시간 멀티미디어 서비스에서 문제가 되는 초기 지연시간을 최소화하는 새로운 동적 스트림 합병 기법을 제안한다. 제안된 기법은 멀티미디어 서비스, 특히 비디오 서비스의 경우 약간의 QoS 변화가 서비스의 질에 큰 영향을 미치지 않는 점을 이용하여, 시간적으로 서로 인접한 여러 개의 스트림을 점차적으로 합병하여 서비스함으로써 짧은 초기 지연시간 유지를 위한 버퍼의 양을 최소화 할 수 있다. 성능분석 결과에서 제안된 기법은 기존의 방법들에 비하여 버퍼 활용면에서 우수한 성능을 나타냄을 보인다.

I. 서 론

실시간 멀티미디어 저장 서버는 동영상상을 포함한 대용량의 멀티미디어 데이터를 실시간으로 검색하여 각각의 세그먼트들이 끊임 없이 디스플레이 장치에 도착하도록 해야 하는 실시간 특성을 보장해야 한다. 이러한 멀티미디어 저장 서버에서 가능한 많은 사용자들에게 동시에 실시간 서비스를 지원하기 위해서는 보다 정교한 디스크 스케줄링과 데이터 배치기법이 필요하다[1]. 이들 기법 중에서 제약된 분할 할당(constrained and partitioned allocation)과 라운드로빈 스케줄링의 조합된 기법은 간단하면서도 이러한 조건을 비교적 잘 만족한다[3].

그러나 이 기법에서 가장 큰 문제점은 초기 지연시간(initial latency time)이 매우 길다는 것이다. 여기서 초기 지연시간이란 서버에 새로운 요구가 도착해서 첫 번째 데이터 세그먼트가 서버의 메모리로 이용 가능하게 적재되는데 걸리는 시간을 의미한다. 초기 지연시간의 증가는 비교적 짧은 상영시간을 갖는 비디오 클립이나 비디오게임 같은 실시간 대화형 서비스에서 큰 문제가 될 수 있다.

이러한 초기 지연시간 문제를 해결하기 위한 주된 연구로는 데이터 배치 기법이나 디스크 스케줄링 기법을 변화시키는 방법과 시작 세그먼트들을 별도의 디스크나 메모리에 중복 저장하는 방법 등이 있다[2]. 그러나 이 방법들도 초기 지연시간을 줄이기 위해 많은 양의 버퍼를 요구하는 등의 문제점이 있다.

본 논문은 멀티미디어 서비스, 특히 비디오 서비스의 경우 약간의 QoS 변화가 서비스의 질에 큰 영향을 미치지 않는 점을 이용하여 짧은 초기 지연시간을 유지하면서도 적은 양의 버퍼를 필요로 하는 동적 스트림 합병 기법을 제안한다. 제안된 기법의 성능은 수식과 시뮬레이션을 통하여 분석되며, 성능분

석 결과 제안된 기법이 기존의 기법에 비하여 짧은 초기 지연시간을 유지하면서 적은 양의 버퍼가 사용됨을 보인다.

본 논문의 구성은 2장에서 관련되는 최근 연구들에 대하여 기술하고, 3장에서는 본 논문에서 제안하는 기법에 대하여 설명한다. 그리고 4장에서는 제안된 기법의 성능 분석에 대해 기술하며, 마지막으로 5장에서는 결론에 대해 기술한다.

II. 관련 연구

실시간 멀티미디어 저장서버에서 가능한 많은 사용자들에게 동시에 실시간 서비스를 지원하기 위한 디스크 스케줄링과 데이터 배치 기법들 중에서 비교적 우수한 성능을 보이는 제약된 분할 할당(constrained and partitioned allocation) 기법[3]은 초기 지연시간이 길다는 문제점을 가지고 있다. 이 초기 지연시간 문제를 해결하기 위한 지금까지의 연구는 다음과 같으며, 여기에서 사용되는 파라미터들은 표 1에 나타내었다.

파라미터	설 명
N	처리량
Mem	총 메모리 크기, bytes
TR	전송률, bytes/s
S	세그먼트 크기, bytes
R	나뉘진 영역 수
Cyl	실린더 수
r(d)	d 에 대한 탐색 오버헤드 계산 함수
T	요구의 한 주기 서비스 시간
X_i	i 번째 스트림
$X_{i,j}$	i 번째 스트림의 j 번째 세그먼트

표 1. 파라미터

1. 양방향 배치 기법 (Scheme Bidirectional)

이 기법은 하나의 스트림을 n개의 세그먼트들로 나누어 R개의 동등한 영역으로 분할한 디스크에 저장하는 기법으로 연속적인 세그먼트들은 인접 영역에 배치한다. 이때 연속된 세그먼트의 최대 탐색 거리는 $2/R$ 이 된다. 동시에 여러 프리젠테이션 요구가 들어오면, 디스크 헤드가 각 영역에서 요구된 모든 세그먼트들을 추출하여 서비스하므로 서로 다른 스트림의 세그먼트간 탐색거리는 Cyl/R 이다. 그림 1은 3개의 스트림을 저장하고 있는 디스크를 4개의 영역으로 분할하였을 때의 양방향 배치 기법을 보인다.

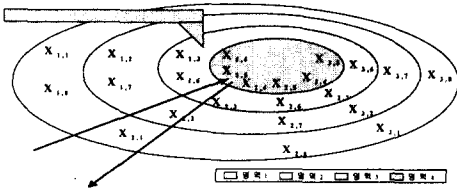


그림 1. 양방향 배치 기법

2. 단방향 배치 기법 (Scheme Unidirectional)

양방향 배치 기법과 유사한 기법으로 세그먼트들을 가장 바깥쪽의 영역으로부터 가장 안쪽의 영역으로 위치시켜 양방향 배치 기법의 초기 지연시간을 반으로 감소시킬 수 있다. 그림 2는 3개의 스트림을 저장하고 있는 디스크를 4개의 영역으로 분할하였을 때의 단방향 배치 기법을 보인다.

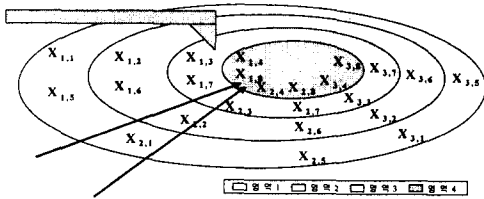


그림 2. 단방향 배치 기법

3. 순차적 기법 (Scheme Sequential)

이 기법은 단방향 배치 기법을 사용하면서 동시에 여러 스트림의 세그먼트 검색 시에 순차적으로 요구된 세그먼트들을 추출하는 기법이다. 디스크 헤드가 순차적으로 세그먼트들을 추출하여 탐색 거리는 $Cyl/(R*N)$ 으로 감소하나 요구하는 버퍼의 양이 $2*N*S$ 로 증가한다.

4. 복사 기법 (Replication Schemes)

복사 기법들은 기본적으로 단방향 배치 기법과 순차적 기법을 따르며 다음과 같은 두 가지 기법이 있다.

1) 메모리에 복사본을 두는 기법

디스크에 저장되어 있는 스트림들의 초기 일부 세그먼트들을 메모리에 복사하여 저장해 두는 기법으로 초기 지연시간을 0으로 줄일 수 있으나 메모리 비용은 매우 크다.

2) 디스크에 복사본을 두는 기법

스트림의 초기 일부분을 메모리가 아닌 별도의 디스크에 복사해 두는 기법으로 그림 3과 같이 시작 세그먼트를 서비스하고 있는 중 새로운 요구가 들어왔을 때 기다리지 않고 다른 디스크에 복사해둔 시작 세그먼트를 서비스함으로써 초기 지연시간을 줄인다. 그러나 새로운 사용자 요구들이 시간의 차이를 두고 연속적으로 들어오면 초기 지연시간을 줄이기 위해 많은 양의 세그먼트를 버퍼로 읽어 들인 후 서비스가 끝날 때까지 이 세그먼트들을 버퍼에 유지시켜야하며 이를 위한 많은 양의

버퍼를 필요로 한다.

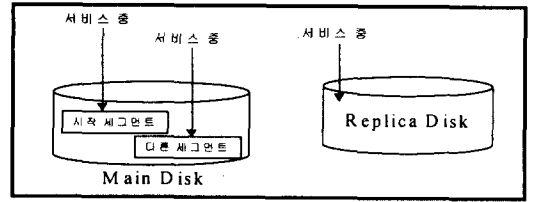


그림 3. 디스크 복사 기법

III. 동적 스트림 합병 기법

본 논문은 실시간 멀티미디어 저장 서버에서 가능한 많은 사용자들에게 동시에 실시간 서비스를 지원하기 위한 디스크 스케줄링과 데이터 배치기법으로써 단방향 배치 기법과 순차적 기법을 기반으로 하며, 초기 지연시간을 감소시키기 위하여 디스크 복사 기법을 사용한다. 본 논문은 여기서의 문제점을 개선하여 적은 양의 버퍼의 사용하여 초기 지연시간을 최소화하고 사용자의 수용 능력을 최대화 할 수 있는 새로운 기법을 제안한다.

제안된 동적 스트림 합병 기법은 기존의 서비스에 대하여 일정 시간 내에 들어온 새로운 서비스 요구에 대하여 디스플레이 속도를 약간 증가시켜 이전의 서비스와 합병시킴으로써 적은 초기 지연시간의 유지에 필요한 버퍼의 양을 줄이고 디스크 능력에 최대한의 사용자를 수용할 수 있다.

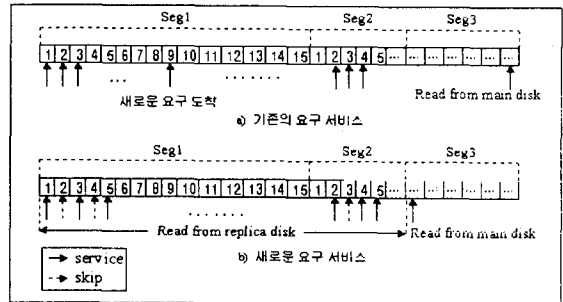


그림 4. 동적 스트림 합병 기법

그림 4는 본 논문에서 제안하는 동적 스트림 합병 기법을 보인 것으로 기존에 들어온 사용자 요구를 서비스하고 있는 도중 새로운 요구가 도착했을 때, 디스크 헤드가 새로운 요구의 시작 세그먼트를 읽어 들이지 못하는 영역에 있다면 복사해 둔 디스크에서 시작 세그먼트의 복사본을 읽어온다. 이때 시작 세그먼트뿐만 아니라 기존의 요구와의 영역차이 만큼 세그먼트를 읽어오고, 원래의 데이터가 있는 디스크의 헤드는 기존의 요구를 위한 세그먼트를 읽어 들인 후 같은 영역에 있는 새로운 요구를 위한 세그먼트를 읽어 들인다. 새로운 요구는 버퍼의 사용량이 최소가 될 때까지 일정 비율로 프레임을 삭제하여 서비스함으로써 버퍼의 사용량을 줄인다.

그림 5는 그림 4와 같은 조건에서 접근하는 영역이 같은 영역이 될 때까지 프레임 삭제를 서비스하는 것을 보인 것으로, x축은 서비스 단위 시간을 의미하고, y축은 요구된 세그먼트가 저장되어 있는 영역을 의미한다.

본 논문에서는 디스크에 복사본을 두는 기법을 적용하여 적은 초기 지연 시간을 갖도록 하고, 요구의 간격에 따라 처음 읽어 들이는 많은 양의 세그먼트들이 차지하는 버퍼를 서비스가 끝날 때까지 유지하는 단점을 세그먼트의 일부분을 전송하

지 않는 기법으로 버퍼 사용량을 최소로 줄인다. 또한 최소 버퍼 요구량으로 줄어드는 데까지 걸리는 시간은 불과 몇 초에 지나지 않고, 줄어든 후에는 더 이상 프레임들을 삭제하지 않기 때문에 서비스의 질을 많이 낮추지 않고도 버퍼 사용량을 감소시킬 수 있다.

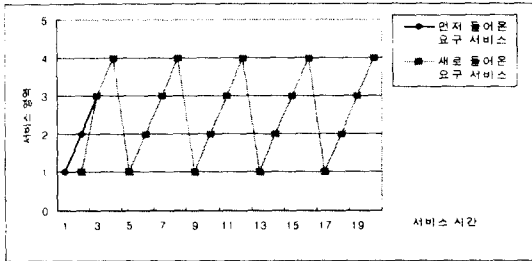


그림 5. 동적 스트림 합병 기법에서의 프레임 삭제

IV. 성능분석

1. 수식적 성능분석

위에서 기술한 배치 기법들에서 동시에 N개의 사용자 요구를 서비스하기 위해서 디스크 헤드가 한 영역에 도착하면, T시간 안에 다음 세그먼트뿐만 아니라, 다른 스트림의 세그먼트들도 읽어야하므로 N개의 세그먼트를 읽는 총 시간을 고려해야 한다. 이때 동시에 N개의 요구를 서비스하는데 걸리는 시간 T는 영역에서 세그먼트를 탐색하고 전송하는 시간과 N개의 세그먼트를 전송하는 시간의 합과 같으므로 $N*S/TR$ 와 같다.

이를 이용하여 순차적 기법과 본 논문에서 사용하는 디스크에 복사해 두는 기법의 초기 지연 시간 T와 디스크에 복사해 두는 기법과 본 논문에서 제안하는 기법의 버퍼 요구량을 구하면 다음과 같다.

1) 순차적 기법

디스크에서 단방향 배치기법으로 놓여 있는 것을 순차적으로 읽어 들이므로 초기 지연 시간 $T = Cy/(R*N)$ 이고 버퍼 요구는 다음과 같다.

$$((R+1)*(N-1)+2)*S \leq Mem$$

2) 디스크에 복사해 두는 기법

복사 디스크에 초기 몇 개의 세그먼트들을 복사해 두므로 초기 지연 시간 $T = \alpha(Cy)$ 이 된다. 또한 순차적 기법을 따르므로 버퍼 요구는 다음과 같다.

$$((R+1)*(N-1)+2)*S \leq Mem$$

3) 동적 스트림 합병 기법

디스크 복사 기법을 적용하므로 초기 지연시간 $T = \alpha(Cy)$ 이 된다. 버퍼 요구는 처음에는 디스크에 복사해 두는 기법과 동일하나 점차 감소하므로 최종 버퍼 요구량은 다음과 같다.

$$R+1+N*2*S \leq Mem$$

2. 시뮬레이션

본 논문은 제안된 기법의 성능을 시뮬레이션을 통하여 실험적으로 분석한다. 실험에서 사용자의 요구는 지수분포로 들어오며 스트림의 합병을 위한 프레임의 삭제 비율은 50%(2개의 프레임 중 1개의 프레임은 디스플레이 하지 않음)로 하였으며, 하나의 서비스 단위시간에 약 16개의 프레임을 서비스한다.

실험 결과로 그림 6, 7은 디스크 영역의 수를 4와 10으로 했을 때 버퍼 요구량을 비교한 것으로 디스크 복사 기법과 비교하여 제안된 기법이 적은 양의 메모리를 요구함을 볼 수 있다.

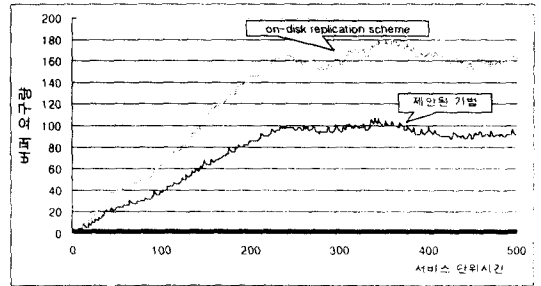


그림 6. 영역 수가 4일 때 버퍼 요구량 비교

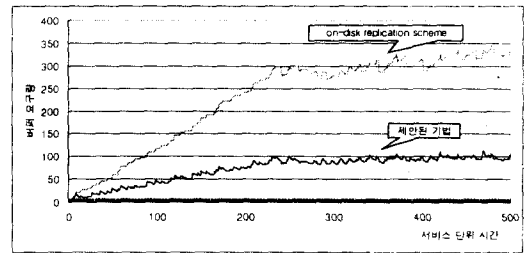


그림 7. 영역 수가 10일 때 버퍼 요구량 비교

V. 결론

실시간 멀티미디어 저장 서버에서 가능한 많은 사용자들에게 동시에 실시간 서비스를 지원하기 위해서는 정교한 디스크 스케줄링과 데이터 배치기법이 필요하다. 이들 기법 중에서 제약된 분할 할당(constrained and partitioned allocation)과 라운드 로빈 스케줄링의 조합된 기법은 간단하면서도 이러한 조건을 비교적 잘 만족한다.

본 논문에서는 이 기법의 가장 큰 문제점인 초기 지연시간을 감소시키기 위한 방법으로 동적 스트림 합병 기법을 제안하였다. 제안된 기법은 멀티미디어 서비스, 특히 비디오 서비스의 경우 약간의 QoS 변화가 서비스의 질에 큰 영향을 미치지 않는 점을 이용하여, 시간적으로 서로 인접한 여러 개의 스트림을 점차적으로 합병하여 서비스함으로써 짧은 초기 지연시간 유지를 위한 버퍼의 양을 최소화 할 수 있다. 성능분석 결과에서 제안된 기법은 기존의 방법들에 비하여 버퍼 활용면에서 우수한 성능을 나타냄을 보였다.

참고문헌

[1] A.L. Narasimha Reddy, Jim Wylie, "I/O issues in a multimedia system" IBM Almaden Research Center, CA 95120
 [2] Edward Chang, Hector Garcia-Molina, "Reducing Initial Latency in Media Servers" IEEE Multimedia July-September, Vol.4, No.3, 1997, pp.50-61
 [3] S. Ghandeharizadeh, S. Kim, C. Shahabi, "On Configuring a Single Disk Continuous Media Server" Sigmetrics Performance Evaluation, Vol.23, No.1, 1995, pp.37-46