비디오 데이터의 요청 정도에 따른 동적 분산 배치 기법

안 유 정○ 원 유 헌
홍익대학교 컴퓨터공학과
{yjahn, won}@cs.hongik.ac.kr

A Distributive Placement Policy according to the Degree of Requests for Video Data

You-Jung Ahn○ Yoo-Hun Won
Department of Computer Engineering, Hong-Ik University

요 약

주문형 비디오 서버와 같은 일기 작업을 주로 하는 멀티미디어 서버에서 검색 요청에 대한 보다 효율적인 서비스는 데이터 배치 기법에 의해 많은 영향을 받는다. 특히 사용자들에게 가장 인기있는 몇몇 비디오 데이터들은 많은 사용자들에 의해 동시에 요청되기도 하지만 짧은 시간 간격을 두고 연속적으로 요청되기도 하므로 인기있는 비디오 데이터일수록 데이터의 서로 다른 많은 부분이 동시에 검색될 수 있도록 효율적으로 배치시키는 것이 필요하다. 본 논문에서는 비디오 데이터의 인기도 즉 요청 성도에 따라 스트라이핑 폭을 동적으로 결정하는 배치 정책을 제안하고 실험을 통해 기존의 정책과 비교한다.

1. 서 론

본 논문에서는 읽기 작업을 주로 하는 주문형 비디오 서버에서 보다 많은 고객들에게 그들이 원하는 영화를 서비스해주기 위한 데이터 배치 정책을 제안한다. 주문형 비디오 서버의 경우, 검색 성능에 큰 영향을 미치는 것은 특정 시간대에 집중되는 인기있는 영화에 대한 요청들이다. 인기있는 영화의 경우, 동시에 많은 요청이 몰릴 뿐만 아니라 근소한 요청 간격 차이로 영화의 서로 다른 부분들이 동시에 검색되어 서비스되어야 한다. 따라서 저장 서버의 검색 성능은 데이터 배치 방법이 얼마나 효율적인가에 달려있다.

따라서 본 논문에서는 비디오 데이터의 인기도 즉, 요청 정도에 따라 데이터의 분산 배치 정도를 동적으로 결정하는 배치 방법을 제안한다. 또한 실험을 통해 제안된 방법의 효율성을 검증하고 결론을 맺는다. 2장에서는 제안된 배치 알고리즘에 대한 기술을, 3장에서는 제안된 방법을 위한 실험 결과를 보이고 4장에서 결론을 맺는다.

2. 요청정도에 따른 동적 분산 배치 정책

본 배치 정책에 사용된 비디오 데이터의 특성 정보로는 데이터 압축 과정에서 얻어지는 해당 데이터의 총 크기와 전송대역폭 및 상영 시간과, 저장 시스템 상에서 비디오 데이터가 3차 저장장치로부터 2차 저장장치로 교체되어 배치되는 과정에서 얻어지는 인기도, 즉 평균 요청 간격이다.

제안된 배치 정책에서는 기본적인 배치 기법으로, k-way uniform striping을 사용한다. 이 기법은 uniform striping과 k-way striping을 결합한 것으로서, k-way striping이란 데이터 분산 배치시 스트라이핑 폭을 k로 하여 분산 배치시킨다는 것을 의미한다. 이때 k는 각 데이터에 대한 평균 요청 간격 즉 인기도에 의해 결정되는데, 인기도가 큰 비디오 데이터일수록 k 값을 크게 함으로써 동시에 보다 많은 양의 데이터를 검색하여 전송할 수 있게 한다. 본 논문에서 제안한 배치 알고리즘은 다음과 같다.

· 인기도에 따른 분산 배치 알고리즘

단계 1 : 디스크 그룹화 과정

입출력 대역폭 성능이 유사한 디스크들을 논리적으로 같은 그룹으로 구성

단계 2 : 비디오 데이터 x(V_x)에 대한 평균 요구량 계산

$$V_x\text{평균요구량} = V_x\text{전송대역폭} \times \frac{V_x\text{상영시간}}{V_x\text{평균요청간격}} \quad (1)$$

단계 3 : 비디오 데이터 x(V_x)에 대한 분산 배치 폭 계산

V_x 배치량 = V_x 평균 요구량

$$\text{if } \textit{disk group } G_i = \min\left(\frac{\sum G_i\text{에 이미 할당된 비디오들의 평균 요구량}}{G_i\text{의 입출력대역폭}}\right) \quad (2)$$

while (V_x 배치량 ≠ 0)

{

if (G_i의 입출력대역폭< V_x 배치량) (3)
{ SW[x][i] = 디스크 그룹 G_i 의 총 디스크 수
V_x 배치량 = V_x 배치량 - G_i 의 입출력대역폭
}
else (4)

$$SW[x][i] = \frac{V_x \text{ 배치량}}{G_i \text{ 의 단위 디스크의 입출력대역폭}}$$

Striping_width[x] = Striping_width[x] + SW[x][i]
} /* end of while */
while ($G_i \in$ 비디오 데이터 x의 할당을 위해 (5)
선택된 그룹들의 집합)
{ Place video data x on G_i by Uniform Striping
with SW[x][i] width
}
모든 비디오 데이터에 대해 단계 3 반복
end. /* 분산 배치 알고리즘 끝 */

위의 알고리즘에서 가장 핵심은 각 비디오 데이터에 대한 분산 배치 폭, 즉 몇 개의 디스크에 분산 배치시킬 것인가를 결정하는 것이다. 단계 2에서는 먼저 각 데이터에 대한 평균 요구량을 구한다. 식 (1)에서 상영시간/평균요청간격은 해당 비디오 데이터가 평균적으로 늘 요청되고 있는 횟수를 의미하므로, 평균 요구량이란 각 비디오 데이터의 인기도를 고려했을때 해당 데이터가 단위 시간에 검색되어 전송되어야 할 총 데이터 양을 의미한다. 따라서 식 (1)에서 V_x 평균 요구량이란 비디오 데이터 x가 한 번 요청될 때 단위 시간에 검색되어 서비스되어야 할 데이터 량을 의미한다.

단계 3에서는 단계 2에서 구한 평균 요구량을 근거로 해당 비디오 데이터의 분산 배치 폭을 구한다. 단계 3에서 V_x 배치량이란 비디오 데이터 x가 디스크 상에 배치되는 과정에서 배치되어야 할 잔여 데이터 량을 의미하고 G_i는 디스크 그룹 i를 지칭한다. 식 (3)과 (4)에서 사용된 SW[x][i]는 비디오 데이터 x를 배치하기 위해 선택한 디스크 그룹 i가 있을 때 이 그룹 내에서 해당 비디오 k의 스트라이핑 폭을 값으로 갖는다. 이때 만일 디스크 그룹 i의 잔여 입출력대역폭이나 용량이 충분하다면 식 (4)에서처럼 한 디스크 그룹 내에 비디오 k가 모두 배치될 수 있으나, 식 (3)에서처럼 그렇지 못할 경우에는 여러 디스크 그룹을 선택하여 각각 조금씩 배치하여야한다. 단계 3의 과정이 끝나면 각 비디오 데이터 x의 스트라이핑 폭인 k의 값이 결정되는데, 위의 알고리즘에서 비디오 데이터 x에 대한 k 값은 최종적으로 Striping_width[x]에 들어가게 된다.

3. 실 험

본 논문에서 제안한 배치 정책의 검색 성능을 검증하기 위해 실험을 수행하였다. 실험에서 사용된 비디오 데이터들은 표 1과 같고[3] 디스크의 종류는 총 4가지로 표 2와 같다. 이 디스크들을 다양하게 조합하여 총 디스크 수를 20, 40, 60으로 구성하여 각각 실험하였으며 이들을 가지고 구성한 전체 저장 시스템의 구조는 표 3과 같다.

위에서 설명한 실험 환경을 통해 본 논문에서 제안한 배치 정책의 검색 성능을 측정하였다. 첫 번째 실험에서는 표 1의 비디오 데이터를 크기 20인 디스크 배열에 배치한 경우의 실험이다. 이때 본 논문에서 제안한 인기도를 고려한 배치 정책과 인기도를 고려하지 않은 배치 정책에 대해 동시에 서비스 가능한 고객의 수를 측정함으로써 검색 성능을 비교하였다. 먼저 표 4에서는 인기도를 고려한 배치 정책을 사용하여 데이터를 배치하였을 때 각 비디오 데이터가 어느 디스크 상에 어느 정도로 분산 배치되었는지를 나타내고 있다. 이 실험의 경우, 현재 사용된 저장 시스템의 성능이 배치될 전체 비디오 데이터들의 요구량 보다 작아서 일부의 비디오 데이터는 배치되지 못한다. 그림 4는 이 실험에 대한 결과를 보이고 있는데 인기도를 고려한 배치 정책을 사용하였을 때가 대체로 더 좋은 검색 성능을 내고 있다.

표 1. 실험에 사용된 비디오 데이터

비디오 제 목	총 크기 (GB)	평균 전송 대역폭 (Mbps)	상영 시간 (sec)	평균 요청 간격 (sec)
asterix	0.5587	0.59	7200	10
atp(tennis)	0.5473	0.55	3600	10
dino(movie)	0.327	0.33	7200	10
lambs(")	0.183	0.18	5400	20
mr.bean	0.441	0.44	3600	20
mtv	0.495	0.49	3600	20
news	0.384	0.38	3600	30
race	0.7687	0.77	7200	30
settop	0.15	0.15	3600	40
simpsons	0.4644	0.46	3600	60
soccer	0.6278	0.63	6000	300
starwars	0.39	0.36	7200	600
talk show1	0.3634	0.36	3600	600
talk show2	0.4479	0.49	3600	1800
terminator	0.2726	0.27	7200	3600

표 2. 실험에 사용된 디스크 종류

디스크 모델명	저장 능력 (MB)	평균 탐색 시간 (ms)	평균 회전율 (ms)	단위데이터 전송 시간 (ms/KB)
Elite9	9090	11	5.56	0.144
Medalist	631	14	7.87	0.2
Hawk	1050	9	5.54	0.142
Barracuda1	2150	8	4.17	0.13

표 3. 실험에 사용된 디스크 배열의 구성
(B : Barracuda1, H : Hawk, E : Elite9, M : Medalist)

총 디스크 배열의 크기	디스크 그룹의 구성			
	디 스 크 그룹 G1	디 스 크 그룹 G2	디 스 크 그룹 G3	디 스 크 그룹 G4
20	B 5개	H 5개	E 5개	M 5개
40	B 10개	H 10개	E 10개	M 10개
60	B 15개	H 15개	E 15개	M 15개

두 번째와 세 번째 실험 역시 디스크 배열의 크기가 40과 60으로 확장된 것 외에는 첫 번째 실험과 유사하며 결과도 역시 인기도를 고려한 경우 더 좋은 결과를 얻었다. 특히 세 번째 실험은 디스크 배열의 크기를 60으로 하였을 경우인데, 표6은 인기도를 고려한 배치 정책을 사용하여 데이터를 배치하였을 때 각 비디오 데이터가 어느 디스크 상에 어느 정도로 분산 배치되었는지를 나타내고 있

다. 이 실험의 경우에는 현재 사용된 저장 시스템의 성능이 배치될 전체 비디오 데이터들의 요구량 보다 크므로 전체 비디오 데이터들이 모두 배치된다.

표 4. 디스크 배열의 크기=20일 경우의 배치 상태

비디오 제 목	스트라이핑 폭	G1	G2	G3	G4
asterix	15	5	5	5	
atp(tennis)	×				
dino(movie)	×				
lambs(")	3				3
mr.bean	×				
mtv	×				
news	3				3
race	×				
settop	×				
simpsons	×				
soccer	×				
starwars	×				
talk show1	1			1	
talk show2	1				1
terminator	×				

표 5. 디스크 배열의 크기=60일 경우의 배치 상태

비디오 제 목	스트라이핑 폭	G1	G2	G3	G4
asterix	14	14			
atp(tennis)	8		8		
dino(movie)	10			10	
lambs(")	3				3
mr.bean	5				5
mtv	4		4		
news	3				3
race	8		5		3
settop	1			1	
simpsons	2			2	
soccer	1	1			
starwars	1			1	
talk show1	1			1	
talk show2	1	1			
terminator	1	1			

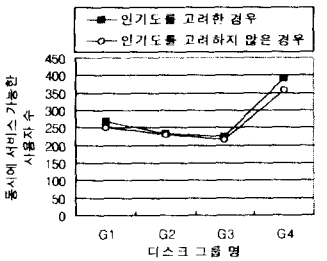

그림 4 인기도 고려 여부에 따른 배치 정책의 검색 성능 비교 (디스크 배열의 크기 = 20일 경우)

4. 결 론

본 논문에서는 주문형 비디오 서버에서 전체 검색 성능을 높이고자 요청이 많은 비디오일수록 비디오의 더 많은 서로 다른 부분을 동시에 읽어낼 수 있도록 분산 배치 시키는 배치 기법을 제안하였다. 인기있는 비디오일수록 많은 사용자들에 의해 동시에 요청되기도 하지만 또한 짧은 시간 간격을 두고 빈번한 요청이 일어나게 되므로 비디오의 많은 서로 다른 부분이 동시에 검색되기 때문이다.

제안된 배치 정책의 효율성을 평가하기 위해, 주문형 비디오 서버의 저장 시스템 환경을 구성하고 같은 환경하에서 비디오 데이터의 인기도를 고려하지 않은 기존의 배치 정책과 비교한 결과, 제안된 배치 방법은 비디오의 인기도를 고려하지 않은 기존의 배치 방법보다 더 검색 성능이 우수함을 알 수 있었다.

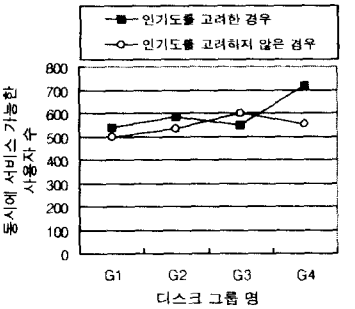

그림 5 인기도 고려 여부에 따른 배치 정책의 검색 성능 비교 (디스크 배열의 크기 = 40일 경우)

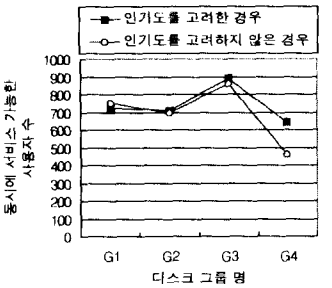

그림 6 인기도 고려 여부에 따른 배치 정책의 검색 성능 비교 (디스크 배열의 크기 = 60일 경우)

5. 참 고 문 헌

[1] Peter M. Chen, Edward K. Lee, David A. Peterson, "RAID: High-Performance, Reliable Secondary Storage", ACM Computing Surveys, Vol 26, No. 2, pp. 152-158, June 1994.

[2] Mathias Rantenberg, Helmut Rzehak, Interactive Distributed Multimedia Systems and Services, Springer, pp. 265-268, 1996.

[3] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems", Univ. of Wurzburg Research Report Series No. 101, Feb. 1995.

[4] Prashant J. Shenoy, Harrick M. Vin, "Efficient Striping Techniques for Multimedia File Servers", Univ. of Texas at Austin, Technical Report CS-TR-96-27, pp. 1-21, 1997.

[5] Sangyup Shim, Harish Vedavyasa, Liu, David H. C. Du, "An Effective Data Placement Scheme to Serve Popular Video-On-Demand", Pacific Workshop on Distributed Multimedia Systems, Hong Kong, June 1996.