소규모 VOD 시스템의 저장 서버로서 디스크 배열 구조의 분석

고 정 국† · 김 길 용††

요 약

대용량의 저장 공간과 고속 통신망을 갖춘 고성능 저장 장치를 필요로 하는 멀티미디어 응용에서는 데이터 전달 속도와 입출력 성능을 향상시키기 위해 디스크 배열이 사용되고 있다. 디스크 배열은 구성 방법 및 데이터 할당 방법 등에 따라 성능의 차이를 보이므로 디스크 배열을 설계할 때 해당 응용에 적합한 디스크 배열 특성 변수가 결정되어야 한다. 본 논문에서는 소규모 VOD 시스템의 저장 서버로서 사용될 디스크 배열의 구조를 결정하기 위해 연속 매체 파일 시스템의 데이터 블럭 크기와 입출력 요구의 크기가 주어질 때 디스크 배열 구성 디스크 수, 디스크 배열 구성과 디클러스터링 정도를 결정하기 위해 시뮬레이션을 통해 성능을 평가하였다. 시뮬레이션을 통해 6 Mbps의 MPEG-2 파일을 제공하는 디스크 배열의 구조는 스트라이핑 단위가 64 KB이며 데이터 블럭이 연속 배치되어 있는 5개의 디스크로 구성된 RAID-5가 가장 적합한 것으로 나타났다.

Analysis of Disk Array Architecture as a Storage Server of a Small-Scale VOD Server

Jeong-Gook Koh† · Gil-Yong Kim††

ABSTRACT

Disk arrays are using to enhance data transfer rate and I/O performance in multimedia applications which need a high-performance storage device with large storage capacity and high-speed network. As performance varies with configuration and data layout scheme, disk array characteristic variables must be appropriately determined in designing disk array architecture for a specific application. In this paper, in order to design a disk array architecture as a storage server of a small-scale VOD system, we evaluate performance of a disk array to choose the number of disks in the array, disk array configuration, a degree of declustering for a given data block size of continuous media file system and I/O request size through simulation. Simulation result shows that RAID level 5 with 5 disks is a suitable candidate for the disk array architecture which provides MPEG-2 files with a rate of 6 Mbps. Moreover, we show that stripe unit is 64 KB and a layout scheme is contiguous placement.

1. 서 론

컴퓨터 하드웨어와 통신 기술의 급속한 기술 발달은 비디오, 오디오, 이미지와 문자 정보 등과 같은 다양한 멀티미디어 정보를 온라인 서비스를 가능케 하였다. 기존의 텍스트 데이터에 비해 엄청난 저장 공간과 고속 전송이 필수적인 멀티미디어 데이터는 전

† 준 회 원:부산대학교 컴퓨터공학과
†† 정 회 원:부산대학교 컴퓨터공학과
논문접수:1996년 9월 12일, 심사완료:1997년 2월 3일

송 지연에 민감하기 때문에 실시간 요구 조건(real-time constraints)을 충족시키면서 연속적으로 제공될 때만 의미 전달이 가능하다[1]. 주문형 비디오(Video-On-Demand)와 같은 멀티미디어 응용에서 다양한 형태의 멀티미디어 정보를 실시간으로 서비스하기 위해서는 대량의 데이터를 효율적으로 저장하고 실시간으로 재생할 수 있는 고성능 저장 장치가 필요하다[2]. 현재 처리기와 메모리의 급속한 발전 속도에 비해 입출력 장치는 이에 미치지 못하여 시스템의 성능 향상에 병목 지점이 되고 있으므로, 전체적인 성능 향상을 위해서는 입출력 시스템의 성능이 향상되어야 한다. 가격대 성능면에서 여러 응용 분야에 적합한 저장 장치로 자기 디스크가 많이 사용되고 있는데, 데이터 전달 속도는 디스크 헤드의 물리적인 이동 속도에 의해 제한되므로 급속한 성능 향상을 기대하기는 어렵다. 이러한 문제를 해결하기 위해 다수의 디스크들을 묶어서 강력한 디스크 시스템을 구성하려는 연구가 진행되어 왔으며 그 결과 디스크 배열이 등장하게 되었다[5, 8, 9, 17]. 디스크 배열은 다수의 디스크상에 데이터를 분산 저장함으로써 대용량의 저장 공간을 얻을 수 있고, 입출력에 참여하는 디스크 수에 비례하여 데이터 전송률도 증가하므로 멀티미디어 응용 분야에 적합하다. 다수의 디스크를 액세스하는 디스크 스트라이핑(disk striping) 기법은 부하 균배(load balancing) 효과, 병렬(parallel) 데이터 전달과 접근 병행성(access concurrency)을 제공한다. 데이터 중복(redundancy) 메카니즘은 디스크의 고장에 대한 하드웨어 신뢰성(reliability)을 높여준다. 디스크 배열은 데이터 분산 구조와 중복 메카니즘에 따라 RAID-0~RAID-6까지 분류된다. 디스크 배열 사용시 디스크 배열 구성 디스크 수, 디스크 배열 구성 그리고 파일을 구성하는 블럭들이 분산 저장되는 정도를 나타내는 디클러스터링 정도(degree of declustering)가 미리 결정되어야 하는데, 이러한 것들을 디스크 배열 특성 변수라고 한다[3].

본 논문에서는 멀티미디어 응용의 예로서 최대 30명 내외의 사용자들에게 재생률이 6 Mbps인 MPEG-2 파일을 제공하는 소규모 VOD 서버를 고려한다. 연속 매체 파일 시스템의 데이터 블럭 크기와 디스크 배열에 대한 입출력 요구의 크기가 주어질 때 디스크 배열 구성 디스크 수, 디스크 배열 구성, 디클러스터링 정도를 결정하기 위해 시뮬레이션을 통해 성능을 측정하고 결과를 분석한다.

본 논문의 구성은 다음과 같다. 2장에서 디스크 배열 특성 변수 결정 방법에 대한 관련 연구들을 기술한다. 3장에서는 데이터 블럭 크기에 따른 연속 매체 파일 시스템의 성능 변화와 시뮬레이션 모델, 작업 부하 특성 및 디스크 모델 특성 등을 기술한다. 4장에서는 시뮬레이션을 통한 성능 측정과 결과에 대한 분석을 기술하며, 5장에서 결론을 맺는다.

2. 관련 연구

디스크 배열의 성능을 최대한 활용하기 위해서는 디스크 배열 설계시 디스크 배열의 구성 방법 등 특성 변수를 해당 응용 분야에 적합하게 결정해야 한다. 이러한 특성 변수를 결정하는 방법들에 대한 연구는 현재 활발하게 이루어지고 있다.

[8]는 16개의 디스크로 구성된 RAID-0를 대상으로 시뮬레이션을 통해 스트라이핑 단위와 병행성간의 이율배반성(tradeoff)을 조사하였으며, 스트라이핑 단위의 크기는 작업 부하의 평균 크기보다 작업 부하의 병행성에 밀접하게 관련되어 있음을 보였다. [9, 10]는 비중복(non-redundant) 디스크 배열의 분석 모델(analytic model)을 제시하고 최적의 스트라이핑 단위를 구하는 방정식을 유도하고, 이 방정식을 이용하여 스트라이핑 단위는 디스크 성능 곱(performance product)과 병행성에 종속적임을 보였다. 그러나, 이러한 종속성도 5% 이내의 성능 손실(penalty)을 감안하면 무시할 수도 있음을 [8]에서 보였다. [12, 13, 15]에서는 비동기 디스크들로 구성된 디스크 배열에서 디스크 배열 구성 디스크 수가 주어진 상태에서 디스크 배열 특성 변수를 결정하기 위한 분석 모델을 제시하였는데, [12, 13]는 평균 요구의 크기와 요구 도착률이 주어졌을 때 하나의 요구가 몇 개로 나누어져 서비스되어야 주어진 요구 처리율(throughput)을 만족시키면서 응답 시간이 최소가 되는지를 결정하고자 하였다. [15]는 요구 도착률(arrival rate)과 두 가지 종류의 요구 크기가 주어질 때 디스크 배열의 평균 응답 시간을 계산하는 분석 모델을 제시하였으며 디클러스터링 정도는 미리 정해져 있었다. [14]는 동기화 디스크와 비동기화 디스크가 조합된 디스크 배열 구조에서 디

스크 수가 주어졌을 때 디스크 배열의 동기/비동기 구조에 따른 응답 시간의 변화를 시뮬레이션을 통해 보였다. [19]는 평균 요구 크기, 요구 도착률, 그리고 평균 요구 응답 시간이 주어질 때 입출력 요구를 만족하는 디스크 배열 구성을 결정하는 분석 모델을 제시하였다.

본 논문에서는 입출력 요구의 크기가 주어질 때 입출력 요구를 충족시킬 수 있는 디스크 배열의 구조를 결정하기 위해 [12, 13, 14, 15]의 제한적인 결과를 보다 확장하고, 일반적인 상황에서 디스크 배열의 특성 변수를 결정하고자 하는 [14, 18]과 달리 특정 응용 분야를 가정함으로써, 해당 분야에 적합한 디스크 배열 특성 변수들을 결정하고자 한다.

3. 시뮬레이션 모델

본 장에서는 우선 데이터 블럭의 크기에 따른 연속 매체 파일 시스템의 성능 변화를 기술한다. 다음으로, 소규모 VOD 시스템의 저장 서버로서 적합한 디스크 배열 구조를 결정하기 위해 사용할 시뮬레이션 모델과 작업 부하 특성 및 디스크 모델 특성, 블럭 배치 정책들을 기술한다.

3.1 데이터 블럭의 크기에 따른 파일 시스템 성능 변화

연속 매체 데이터의 연속적인 재생을 보장하기 위해서는 각 데이터 블럭들이 재생되기 전에 도달해야 한다[4]. 스트림 수의 증가에 따른 입출력 요구의 증가는 디스크의 빈번한 사용과 응답 시간 증가를 유발하는데, 응답 시간 증가는 현재 서비스중인 활성화된(active) 스트림의 QoS(Quality of Service)를 저하시킬 수 있다. 일반적으로 입출력 시스템의 성능 비교시 단위 시간당 평균 데이터 전달률이 높을수록, 응답 시간이 짧을수록 성능이 우수하다고 평가한다. 디스크 시스템의 응답 시간은 서비스 시간과 큐 대기 시간의 합으로 정의되는데, 탐색 시간과 회전 지연 시간 등 물리적인 요인들이 응답 시간을 결정하는 주요한 요인들이다. 따라서, 동일한 상황하에서도 디스크 입출력 횟수를 줄이고 한 번의 디스크 접근으로 전달되는 데이터량을 늘려주면 단위 시간당 데이터 전달률이 높아지고, 응답 시간이 감소되어 디스크 시스템의 입출력 성능이 향상된다.

이 점은 [18]와 단일 디스크에 1시간 분량의 6 Mbps MPEG-2 파일을 저장한 소규모 VOD 서버에서 데이터 블럭의 크기에 따른 연속 매체 파일 시스템의 입출력 성능 측정 결과에서도 알 수 있다. 성능 평가 기준은 연속 매체 파일 시스템이 10개의 활성화된 스트림에게 연속 매체 데이터를 제공하는 상황에서 데이터 블럭 크기를 8 KB~512 KB까지 변경하면서 매 크기마다 100 MB 분량의 데이터를 디스크로부터 연속적으로 읽은 후 측정된 평균 데이터 전달률을 사용하였다. 〈표 1〉은 데이터 블럭의 크기에 따른 연속 매체 파일 시스템의 평균 데이터 전달률을 측정한 결과이다.

〈표 1〉 데이터 블럭의 크기에 따른 파일 시스템의 평균 데이터 전달률

〈Table 1〉 Average data transfer rate of CMFS according to data block size

데이터 블럭 크기	평균 데이터 전달률(Mbps)
8 KB	9.47
16 KB	16.32
32 KB	25.67
64 KB	24.91
128 KB	36.80
256 KB	43.20
512 KB	50.00
8 KB(ufs)	4.37

성능 측정 결과 512 KB에서 가장 우수한 성능을 보였으나 512 KB 데이터 블럭을 사용할 때 메모리 관리의 어려움이 있다. 즉, 인덱스 블럭이나 디스크로부터 읽어온 데이터를 보관하기 위해 512 KB 크기의 버퍼 공간을 할당해야 하는데, 연속 매체 파일 시스템이 동작하는 동안 지속적으로 512 KB 크기의 메모리 영역 할당이 어렵다. 이로 인해 256 KB를 연속 매체 파일 시스템의 데이터 블럭 크기로 선정하였다.

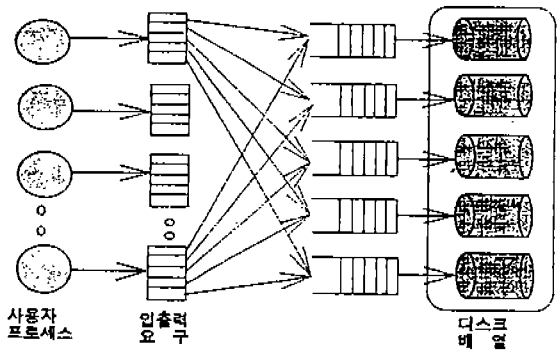
3.2 디스크 배열의 시뮬레이션 모델

(그림 1)은 디스크 배열의 성능 측정을 위한 시뮬레이션 모델로서 입출력 요구가 주어질 때 1)스트라이

핑 단위의 크기 변화에 따른 평균 응답 시간, 2)사용자 프로세스(스트림) 수 증가시 디스크 접근 유형과 디클러스터링 정도에 따른 디스크 배열의 입출력 성능 변화를 측정한다. 시뮬레이션은 디스크 배열 시뮬레이터인 raidSim[6]을 본 논문의 목적에 맞게 수정하여 Sparc 10 워크스테이션상에서 수행하였다. 시뮬레이션 모델은 디스크 배열에 대한 입출력 요구(disk array request)를 생성하는 다수의 사용자 프로세스들과 디스크 배열을 구성하는 여러 개의 디스크로 이루어지며, 디스크 헤드 스케줄링 정책으로 SCAN을 사용한다.

사용자 프로세스는 스트라이핑 넓이(stripe width)가 n인 디스크 배열에 대한 입출력 요구를 한번에 하나씩 생성한다. 디스크 배열의 전역 큐에 도착한 입출력 요구는 n개의 디스크 요구(disk request)들로 나뉘어지며, 나뉘어진 디스크 요구들 중에 첫번째 요구를 서비스할 디스크는 임의로 선정된다. 나머지 디스크 요구들은 먼저 선정된 디스크를 기준으로 라운드 로빈 방식으로 각 디스크의 지역 큐에 들어간다. 각 디스크는 자신의 지역 큐에서 대기하고 있는 디스크 요구들을 한번에 하나씩 서비스한다. n개의 디스크 요구들이 모두 서비스되면 디스크 배열에 대한 요구를 생성한 사용자 프로세스는 새로운 입출력 요구를 순환적으로 계속 생성한다.

디스크 배열은 하나의 입출력 요구에 대해 다수의 디스크를 병렬로 액세스하거나 여러 개의 독립적인 입출력 요구들을 병행 서비스한다. 각 디스크는 자신의 큐를 스케줄링하여 디스크 헤드의 움직임을 최적화하는 입출력 요구를 선정한 후 수행한다. 각 디스크들로부터 읽어 온 데이터는 입출력 요구의 크기에 따라 다음과 같이 해당 사용자 프로세스에게 전달된다. 입출력 요구의 크기가 스트라이핑 단위와 같을 때는 읽어온 데이터를 해당 프로세스에 직접 전달하며, 스트라이핑 넓이가 n인 입출력 요구는 n개의 디스크로부터 읽어온 데이터를 동기화시켜 논리 블럭을 형성한 후 해당 프로세스에 전달한다.



(그림 1) 시뮬레이션 모델
(Fig. 1) Simulation Model

3.3 작업 부하 특성

시뮬레이션에 사용된 소규모 VOD 시스템의 저장 서버에 대한 작업 부하의 특성은 〈표 2〉와 같다. VOD 시스템은 저장 서버에 준비된 다수의 영화나 TV 쇼 등을 고객들에게 제공하며, 고객들은 자신의 취향에 따라 여러가지 영상물을 선택할 수 있다. 일반적으로 고객들의 선택 유형은 인기가 있는 특정 영상물에 편중이 되는 Zipf's Law 분포를 보이며, 일단 선택한 영상물은 끝까지 보게 되므로 고객들에 의해 생성되는 모든 작업 부하는 순차적인 읽기 요구이며 쓰기 요구는 존재하지 않는다. 영상물의 길이는 100분 정도로 가정하며, 압축 기법으로 6 Mbps의 재생률을 갖는 MPEG 2를 사용할 때 영화 1편당 약 4.6 GB의 저장 공간이 필요하다.

〈표 2〉 VOD 서버의 작업 부하
〈Table 2〉 The VOD server's workload

Request placement distribution	Zipf
Percentage sequential accesses	0%
Percentage write accesses	0%
Response time goal	1/3 sec
Request size distribution	Constant
Request size means for simulation runs	4.6 GB

작업 부하로서 디스크 배열에 대한 입출력 요구의 도착률은 포아송 분포를 따르며, 초기의 시스템이 비어있는 상태에서 시뮬레이션할 때 나타나는 준비 효과(warmup effect)를 줄이기 위해 초기의 통계량을 제거한다. 즉, 시뮬레이션 초기의 모든 요구들을 서로 시차를 두어(staggered) 발생시키면서, 초기에 발생시킨 요구들이 모두 종료된 이후부터 통계량을 수집한

다. 시뮬레이션의 결과 분석에는 일괄 평균법(Batch mean method)을 사용하는데, (전체 스트림 수 * 길이 10)의 서브런(subrun) 집합으로 나누고 각 일괄마다 표본 평균을 계산하여 총 평균과 신뢰 구간을 계산한다. 본 논문에서 사용된 값들은 95%의 신뢰 구간(confidence interval)을 가지며 폭(width)은 5%이하이다.

3.4 디스크 모델 특성

디스크 배열을 구성하는 디스크 모델은 Seagate사의 ST-11950W 디스크이며, 디스크의 물리적 특성을 나타내는 디스크 매개변수는 〈표 3〉과 같다. 8 bit SCSI-II 버스의 대역폭이 10 MB/sec이므로 2~3 MB/sec의 데이터 전달률을 갖는 SCSI 디스크로 디스크 배열을 구성할 경우 5개 이상의 디스크를 장착시 SCSI 버스가 병목 지점이 된다. [18]의 VOD 서버에서 사용한 2940 Ultra Wide SCSI-II 버스가 20 MB/sec의 전송 대역폭을 가지므로 최대 8대까지 디스크를 장착할 수 있지만, 본 논문에서는 디스크 수를 최대 5개로 제한한다.

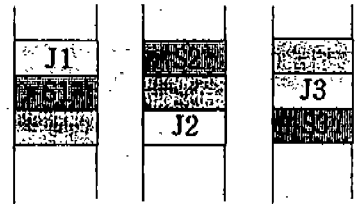
〈표 3〉 디스크 매개변수
〈Table 3〉 Disk parameter values

디스크 매개변수	수 치
Cylinders Disk	2,706
Tracks per Cylinder	15
Sectors Track	81
Bytes per Sector	512
Single Track seek time	0.9 ms
Max Full seek time	19 ms
Spindle Speed	7,200 rpm
Average Latency	4.17 ms
Interface	SCSI-2 Fast Wide

3.5 데이터 블럭 배치 정책

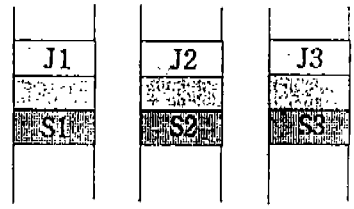
데이터를 저장 공간에 배치하는 블럭 배치 방법들로는 임의(random) 배치, 연속(contiguous) 배치와 제한적(constrained) 배치가 있다[7]. 임의 배치 방법은 효율적인 공간 활용은 가능하지만 대량의 데이터를 고속으로 전달해야 하는 연속 매체 특성상 만족할 만한 입출력 성능을 내지 못한다. 또한, 데이터를 읽기 위해 인덱스 블럭을 자주 참조해야 하며 특정 디스크에 부하가 집중될 경우 디스크 배열의 성능이 저하될 우려가 있다. 연속 배치 방법은 연속적인 디스크 접근시 인덱스 블럭에 대한 참조가 불필요하므로 디스크 사용 횟수가 감소하며, 데이터 블럭이 스트라이핑 되어 있으면 각 디스크들에 대한 부하 균배 효과도 얻을 수 있어 입출력 성능이 우수하다. 단점으로는 데이터에 대한 삽입, 삭제 등 편집 작업을 수행하는 동안 대량의 데이터를 복사해야 하는 부담이 따른다. 제한적 배치 방법은 데이터 액세스 시간을 매체의 연속 재생을 위한 응답 시간 요구 조건 이내로 유지하면서 데이터를 분산 배치(scattered placement)한다. 이 방법은 부하 분산 측면에서는 연속 배치와 동일하지만, 인덱스 블럭을 참조해야 하며 새로운 스트림 추가시 기존 스트림과의 배치 문제가 발생할 수 있다[16].

본 논문에서는 디스크 배열의 데이터 블럭 배치 정책으로 연속 매체 데이터를 여러 디스크상에 스트라이핑시키는 배치 구조를 사용한다. 디스크상의 스트라이핑 단위 배치 방법으로는 (그림 2)와 같은 임의



J Blocks : Jurassic park
S Blocks : StarWars

(a) random placement



J Blocks : Jurassic park
S Blocks : StarWars

(b) sequential placement

(그림 2) 스트라이핑된 데이터 블럭 배치 정책
(Fig. 2) Striped data blocks placement policy

배치와 연속 배치의 2가지 방법을 가정한다. 디스크 내에 임의 배치되어 있는 데이터 블럭을 읽어내기 위한 디스크 헤드의 움직임은 임의 접근 유형과, 연속 배치되어 있는 데이터 블럭을 읽어 내기 위한 디스크 헤드의 움직임은 순차 접근 유형과 유사하므로 디스크내 데이터 블럭 배치 방법과 디스크 접근 유형을 연관지어 생각할 수 있다.

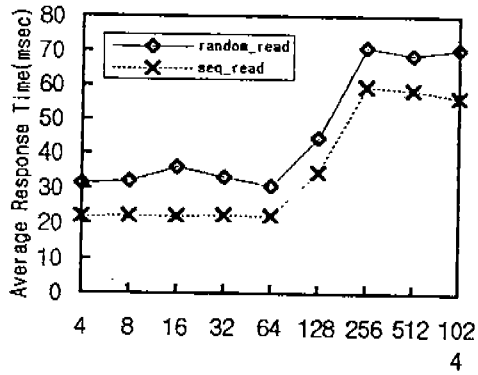
4. 성능 측정 및 결과 분석

VOD 시스템과 같은 멀티미디어 응용 환경에서 입출력 요구는 일반적으로 쓰기 요구보다 읽기 요구의 비율이 높다. 본 논문에서는 이러한 입출력 요구 특성을 반영하여 사용자 프로세스의 모든 입출력 요구를 디스크에 대한 읽기 요구로 가정한다. 디스크 접근 형태에 따라 입출력 요구는 임의 접근과 순차 접근으로 구분된다.

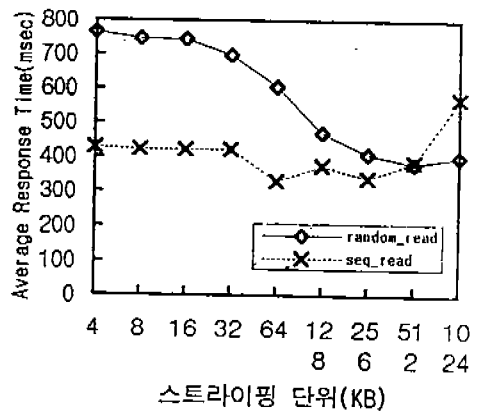
재생률이 6 Mbps인 MPEG-2 파일의 연속 재생을 보장하기 위해서는 초당 768 KB의 데이터가 디스크로부터 제공되어야 한다. 3.1절에서 연속 매체 파일 시스템의 데이터 블럭 크기로 256 KB를 선정하였는데, 256 KB 블럭을 이용하여 매초 768 KB의 데이터를 제공하기 위해서는 초당 3번씩 데이터 블럭을 읽어야 한다. 본 장에서는 데이터 블럭 크기와 입출력 요구의 크기가 주어질 때 적용 분야에 적합한 디스크 배열 구성 디스크 수, 디스크 배열 구성 및 디클러스터링 정도를 결정하는 시뮬레이션 결과를 제시한다.

4.1 디스크 배열의 스트라이핑 단위

스트라이핑 단위는 디스크 배열의 성능을 결정하는 중요한 요소이므로 디스크 배열의 성능을 극대화하기 위해서는 최적의 스트라이핑 단위가 선택되어야 한다. 스트라이핑 단위가 데이터 블럭보다 클 때는 디스크 배열이 다수의 입출력 요구를 병행 서비스하므로 디스크 큐 대기시간이 감소한다. 스트라이핑 단위가 데이터 블럭 크기 이하일 때는 하나의 입출력 요구에 대해 다수의 디스크들을 병렬로 접근하므로 입출력 요구에 대한 응답 시간이 감소한다. 매초 768 KB의 데이터 제공에 적합한 스트라이핑 단위를 선정하기 위해 5개의 디스크로 구성된 RAID-5에서 입출력 요구의 크기가 256 KB일 때 스트라이핑 단위의 변화에 따른 디스크 배열의 평균 응답 시간을 측정하였다. (그림 3)은 단일 스트림과 27개의 스트림을 서비스할 때 스트라이핑 단위에 따른 디스크 배열의 성능 측정 결과를 도시하고 있는데, 다음과 같은 2가지 대조적인 양상을 보인다. 첫째, 단일 스트림을 서비스할 때는 디스크 접근 형태에 관계없이 스트라이핑 단위가 64 KB일 때 응답 시간이 가장 짧아지는데, 이것은 전체 디스크에 작업 부하가 균등 분배되었고 각 디스크로부터 데이터가 병렬로 전달되기 때문이다. 둘째, 스트림 수가 증가하면 디스크 접근 유형에 따라 최대 서비스가능 스트림 수가 달라진다. 최대 서비스가능 스트림 수가 임의 접근 형태에서는 21개,



(a) service a single stream



(b) service 27 streams

(그림 3) 스트라이핑 단위에 따른 평균 응답 시간
(Fig. 3) Average response time according to stripe unit

순차 접근 형태에서는 27개까지 가능하였다. 스트림 수가 27개이며 순차 접근 유형을 보일 때 평균 응답 시간은 64 KB와 256 KB에서 가장 짧았다. 256 KB에서는 입출력 요구를 하나의 디스크가 처리할 수 있기 때문에 스트림 수가 디스크 수보다 작을 때는 유휴(idle) 디스크가 존재하여 성능이 나쁘지만, 스트림 수가 증가하면 다수의 입출력 요구를 독립적으로 병행 처리할 수 있어 성능이 향상된다. 본 논문에서는 스트림 수가 1개와 27일 때 가장 짧은 응답 시간을 보인 64 KB와 256 KB를 스트라이핑 단위로 선정한다.

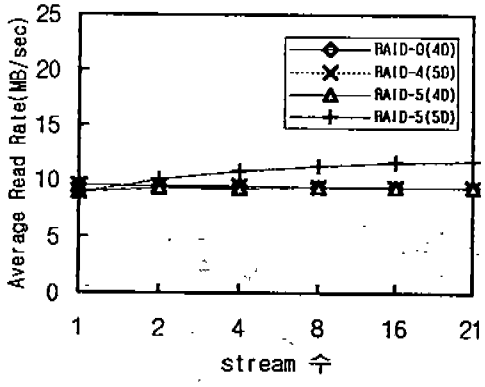
4.2 디스크 접근 유형에 따른 성능 분석

본 절에서는 입출력 요구의 디스크 접근 유형에 따른 디스크 배열 성능을 분석하기 위해 RAID-0, RAID-4 및 RAID-5를 대상으로 스트림 수 증가시 평균 데이터 전달률 측정 결과를 기술한다. (그림 4)은 입출력 요구가 전형적인 OLTP 형태의 임의 접근 유형을 취할 때 성능 측정 결과이다.

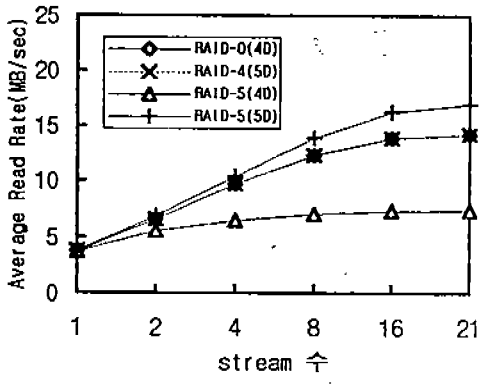
스트라이핑 단위가 64 KB일 때는 디스크 수가 가장 많고 각 디스크에 부하가 고루 분배되는 RAID-5(5D)가 가장 성능이 우수하였고, 다른 디스크 배열 형태들은 거의 동일한 성능을 보였다. RAID-5(5D)는 다른 배열 형태들에 비해 최대 25%의 성능 우위를 보였으며, 256 KB일 때는 RAID-5(4D)에 비해 약 2.3배, 다른 배열 형태에 대해서는 최대 19%의 성능 우위를 보였다. RAID-4(5D)는 RAID-0(4D)에 비해 디스크가 하나 더 있지만 읽기 동작에서는 패리티 디스크가 불필요하므로 데이터 디스크 수가 같아져서 성능이 동일하다. 스트라이핑 단위가 64 KB일 때 RAID-5(4D)는 RAID-0(4D)보다 데이터 디스크의 수가 적지만 성능은 유사하다. 즉, RAID-5(4D)가 하나의 입출력 요구를 서비스할 때 3개의 디스크를 사용하지만 스트림 수가 증가하면 데이터와 패리티 디스크를 구분하지 않는 RAID-5의 특성상 전체 디스크들이 균형있게 사용된다. 그러나, 스트라이핑 단위가 256 KB이고 스트림 수가 3개 이상이 되면 최대 병행 스트림 수는 3개가 되므로 RAID-0(4D)보다 평균 데이터 전달률은 낮다.

(그림 5)는 VOD 시스템의 입출력 요구 특성을 반영한 것으로서, 각 스트림은 선택된 영상물의 디스크 내 시작 위치로부터 데이터를 순차적으로 읽어 나간다. 임의 접근 형태에 비해 순차 접근 특성 반영시 상대적으로 성능이 우수한데, RAID-5(5D)를 대상으로 (그림 4(a))와 (그림 5(a))를 비교하면 순차 접근시 최대 1.8배, (그림 4(b))와 (그림 5(b))에서는 순차 접근시 최대 1.4배의 성능 우위를 보인다. 이러한 결과는 임의 접근 방식의 입출력 특성을 지닌 응용 분야보다 VOD와 같은 멀티미디어 응용 분야의 저장 장치로서 디스크 배열을 사용하는 것이 효율적임을 나타낸다.

디스크 접근 유형에 따른 성능 측정 결과는 데이터 블럭 배치 방법에 따른 디스크 배열의 성능 변화로 생각할 수 있다. 따라서, 디스크 접근 유형에 따른 성능 측정 결과로부터 디스크 배열을 이용한 연속 매체 파일 시스템의 블럭 배치 정책에는 연속 배치 정책이 적합함을 알 수 있다.

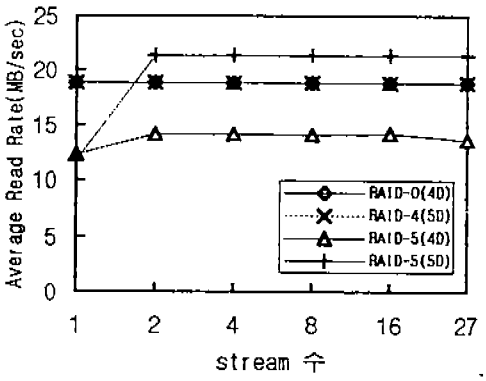


(a) stripe unit : 64 KB

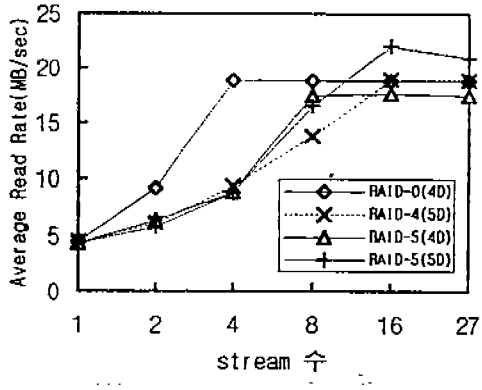


(b) stripe unit : 256 KB

(그림 4) 임의 접근시 평균 데이터 전달률
(Fig. 4) Average data transfer rate in random access



(a) stripe unit : 64 KB



(b) stripe unit : 256 KB

(그림 5) 순차 접근시 평균 데이터 전달률
(Fig. 5) Average data transfer rate in sequential access

4.3 디클러스터링 정도에 따른 성능 분석

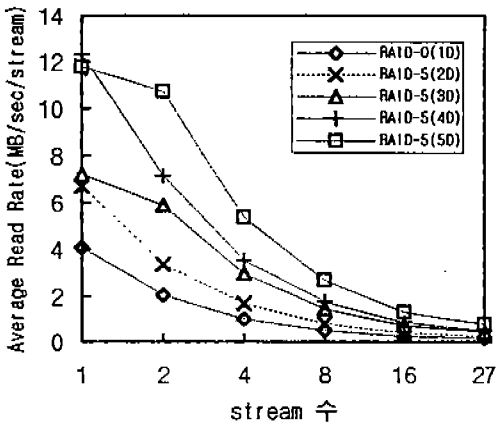
디스크 배열에서는 디스크 접근 단위가 스트라이핑 단위이므로, 입출력 요구 크기가 스트라이핑 단위보다 클 때는 하나의 입출력 요구를 다수의 디스크 요구로 나누어 처리한다.

(그림 5(a))는 스트라이핑 단위가 동일한 디스크 배열 형태들에서 디클러스터링 정도에 따른 성능 변화를 보여준다. 성능 측정 결과 디클러스터링 정도가 높을수록 입출력 요구에 더 많은 디스크가 참여하므로 병렬성이 높아져서 성능이 향상된다. (그림 6)은 RAID-0과 RAID-5를 대상으로 디스크 수를 1개~5개까지 변화시키면서 디스크 배열의 성능 변화를 측정한 결과이다. 저장 장치로 단일 디스크를 사용할 때와 입출력 성능을 비교하기 위해 디스크 배열의 성능을 스트림당 평균 데이터 전달률(MB/sec/stream)로 정규화하였다. (그림 6)에서 보듯이 스트림 수가 늘어나면 응답 시간이 증가하여 스트림당 평균 데이터 전달률은 감소한다. 그러나, 입출력 요구의 크기를 스트라이핑 단위로 나누어 그 몫을 디스크 배열에 장착할 수 있는 최대 데이터 디스크 수로 설정하였을 때, 최대 데이터 디스크 수의 범위 내에서 디스크 수의 증가는 병렬성을 증대시켜 스트림당 평균 데이터 전달률은 향상된다. 즉, 병렬성 이용시 디스크 배열의 성능은 디스크 수에 비례한다. 또한, 전체 저장 용량이 동일할 때에는 적은 수의 대용량 디스크들을 이용하여 디스크 배열을 구성하기보다는 다수의 소용량 디스크를 이용하는 것이 효율적임을 알 수 있다.

스트라이핑 단위에 따른 성능 차이는 (그림4(a))와 (그림 4(b)), (그림5(a))와 (그림 5(b))를 비교함으로써 알 수 있다. 스트라이핑 단위가 64 KB이며 스트림 수가 적을 때에는 짧은 응답 시간과 높은 병렬성으로 256 KB에 비해 성능 우위를 보이지만, 스트림 수가 증가하면 디스크 시스템이 포화상태가 되어 응답 시간이 급격히 증가한다. 스트라이핑 단위가 256 KB일 때는 각 디스크가 독립적으로 입출력 요구를 처리하기 때문에 다수의 입출력 요구를 병행 처리할 수 있다. 그러나, 입출력 요구 수가 디스크 수보다 적으면 유휴 디스크로 인해 디스크 활용도(utilization)가 낮아지고 응답 시간이 길어져서 성능이 나쁘지만, 스트림 수가 증가하면 병행성과 디스크 활용도가 높아져서 스트림 수가 27일 때는 우수한 성능을 발휘하였다. (그림 5)를 통해 다음과 같은 점을 알 수 있다. 첫째, 스트라이핑 단위가 입출력 요구의 크기보다 작을 때에는 디스크의 병렬성을 이용할 수 있는 디스크 배열 구조가 우수한 성능을 발휘한다. 결과적으로 디클러스터링 기법을 적용하면 디스크에 대한 부하 균배를 보장할 수 있으며, 성능 향상을 얻을 수 있다. 디클러스터링 정도는 스트라이핑 단위로 입출력 요구의 크기를 나누어 몫을 구한 다음 그 값으로 설정한다. 둘째, 스트라이핑 단위가 입출력 요구의 크기와 동일하고, 사용자 수가 많은 경우에는 디스크의 병행성을 이용할 수 있는 디스크 배열 구조도 채택할 수 있다.

이상의 결과를 종합해 볼 때, 매초 3번의 256 KB 데이터를 제공해야 하는 VOD 서버의 저장 장치로

적합한 디스크 배열구조는 전반적으로 우수한 성능을 보여주는 스트라이핑 단위가 64 KB이며 데이터 블럭이 연속 배치되어 있는 5개의 디스크로 구성된 RAID-5가 가장 적합함을 알 수 있다.



(그림 6) 순차 접근시 디스크 수에 따른 성능 변화 (스트라이핑 단위 : 64 KB)

(Fig. 6) Performance variation according to number of disk in sequential access

5. 결 론

본 논문에서는 30명 내외의 사용자를 지원할 수 있는 소규모 VOD 시스템의 저장 서버로서 디스크 배열 구조를 결정하기 위해, 연속 매체 파일 시스템의 데이터 블럭 크기와 입출력 요구 크기가 주어질때 디스크 배열 구성 디스크 수, 디스크 배열 구성, 디클러스터링 정도를 결정하기 위해 시뮬레이션을 통한 성능 측정과 분석을 수행하였다.

시뮬레이션 결과 입출력 요구에 대해 최소 응답 시간을 제공하는 스트라이핑 단위는 64 KB이며, 순차 접근이 임의 접근 형태에 비해 약 1.8배 정도의 성능 우위를 보였다. 입출력 요구의 디스크 접근 형태와 블럭 배치 정책을 관련지어 생각하면 연속 매체 파일 시스템의 블럭 배치 정책으로는 연속 배치 정책이 적합함을 알 수 있다. 스트라이핑 단위가 입출력 요구의 크기보다 작을 때는 데이터 전달의 병렬성을 이용할 수 있는 디스크 배열 구조가 유리하며, 디클러스터링 정도는 입출력 요구의 크기를 스트라이핑 단위로 나누어서 그 몫으로 설정한다. 스트라이핑 단위와 입출력 요구 크기가 동일하고 사용자 수가 많은 경우에는 병행성을 이용할 수 있는 디스크 배열 구조가 적합하다. 결론적으로, 6 Mbps의 MPEG-2 파일을 제공하는 소규모 VOD 시스템의 저장 서버로 적합한 디스크 배열 구조는 스트라이핑 단위가 64 KB이며, 데이터 블럭이 연속 배치되어 있는 5개의 디스크를 사용하는 RAID -5가 가장 적합하다.

향후 연구 과제로는 본 논문의 제한적인 요소들을 보다 일반화시켜 디스크 배열의 구조를 결정하려는 연구가 필요하다. 즉, VOD 응용에서도 약간의 쓰기 요구가 발생하므로 입출력 요구의 종류에 쓰기 요구를 추가하고, 디스크에 국한된 작업 부하를 CPU 시간과 운영체제의 오버헤드까지 고려하여 보다 실제적인 작업 부하를 생성하면 다양한 응용 분야에 적합한 디스크 배열을 설계하고 구성하는데 도움이 될 것이다.

참 고 문 헌

[1] H. M. Vin, P. Goyal, A. Goyal and Anshuman Goyal, "A Statistical Admission Control Algorithm for Multimedia Servers," *Proc. of the ACM Multimedia 94*, pp. 33-40, Oct 1994.

[2] D. James Gemmell, Harrick M. Vin, Dilip D. Kandlur, P. Venkat Rangan and Lawrence A. Rowe, "Multimedia Storage Servers: A Tutorial," *IEEE Computer*, pp. 40-49, May 1995.

[3] Peter M. Chen and Edward K. Lee, "Maximizing Performance in a Striped Disk Array," *SIGARCH 17th Annual International Symposium on Computer Architecture*, 1990.

[4] P. Venkat Rangan and Harrick M. Vin, "Designing File Systems for Digital Video and Audio," *Proc. of the 13th Symposium on Operating Systems Principles(SOSP '91), Operating Systems Review*, Vol. 25, No. 5, pp. 81-94, Oct 1991.

[5] D. A. Patterson, G. Gibson, and R. H. Katz, "A case for redundant arrays of inexpensive disks (RAID)," *Proc. ACM SIGMOD*, pp. 109-116, Jun 1988.

[6] Edward K. Lee, "Software and Performance Issues in the Implementation of RAID Prototype," *Technical Report UCB/CSD 90/573*, Berkeley CA, 1990.

[7] A. L. Narasimha Reddy and Prithviraj Banerjee, "An Evaluation of Multiple-Disk I/O Systems," *IEEE Transactions on Computers*, Vol. 38, No. 12, pp. 1680-1690, Dec 1989.

[8] Peter M. Chen and David A. Patterson, "Maximizing Performance in a Striped Disk Array," *Proc. of the 1990 International Symposium on Computer Architecture*, pp. 322-331, May 1990.

[9] Edward K. Lee and Randy H. Katz, "An Analytic Performance Model of Disk Arrays and its Application," *Technical Report UCB/CSD 91/660*, Berkeley CA, 1991.

[10] Edward K. Lee and Randy H. Katz, "An Analytic Performance Model of Disk Arrays," *Proc. of the 1993 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pp. 90-109, May 1993.

[11] Edward K. Lee and Randy H. Katz, "The Performance of Parity Placement in Disk Arrays," *IEEE Transactions on Computers*, Vol. 42, No. 6, Jun 1993.

[12] Peter Scheuermann, Gerhard Weikum and Peter Zabback, "Automatic Tuning of Data Placement and Load Balancing in Disk Arrays," Database Systems for Next-Generation Applications: Principles and Practise, 1991. DBS-92-91.

[13] Gerhard Weikum and Peter Zabback, "Tuning of Striping Units in Disk Array-Based File Systems," *Proc. of the 2nd International Workshop on Research Issues on Data Engineering: Transaction and Guery and Processing*, pp. 80-87, 1992.

[14] A. L. Narasimha Reddy and Prithviraj Banerjee, "An Evaluation of Multiple Disk I/O Systems," *IEEE Trans. Computers*, Dec 1989.

[15] Shenze Chen and Don Towley, "A Queueing Analysis of RAID Architecture," *Technical Report 91/71*, U. of Massachusetts, 1991.

[16] P. V. Rangan and H. M. Vin, "Efficient Storage Techniques for Digital Continuous Multimedia," *IEEE Trans. On Knowledge and Data Engineering : Special Issue on Multimedia Information Systems*, Aug 1993.

[17] Fouad A. Tobagi, Joseph Pang, Randall Baird, and Mark Gang, "Streaming RAID: A Disk Array Management Systems for Video Files," *Proc. of the ACM Multimedia 93*, pp. 393-400, Aug 1993.

[18] 고정국, 남경규, 김길용, "멀티미디어 데이타 스트림을 위한 효율적 화일시스템의 설계 및 구현," 한국정보과학회 '96 봄 학술발표논문집, 제23권 1호, pp. 407-410, 1996.

[19] 황 경숙, 박 찬익, "입출력 요구 특성을 지원하기 위한 디스크 배열 구조의 성능 분석," 한국정보과학회 논문지, 제21권 제12호, pp. 2344-2353, 1994.

고 정 국

1992년 부산대학교 컴퓨터공학과 졸업(학사)
1994년 부산대학교 대학원 컴퓨터공학과(공학석사)
1994년~현재 부산대학교 대학원 컴퓨터공학과 박사과정
관심분야: 분산 시스템, 멀티미디어 시스템, 실시간 시스템

김 길 용

1981년 서울대학교 수학교육과 졸업(학사)
1983년 서울대학교 대학원 컴퓨터공학과(공학석사)
1988년 서울대학교 대학원 컴퓨터공학과(공학박사)
1983년~1986년 금성반도체(주) 연구원
1994년~1995년 Univ. of Southern California(USC) 객원교수
1988년~현재 부산대학교 컴퓨터공학과 부교수
관심분야: 분산 시스템, 멀티미디어 시스템, 실시간 시스템