

다중 디스크 기반 모바일 시스템 대상의 에너지 효율적인 연관 데이터 배치 기법

(Energy-efficient Correlated Data Placement Techniques for Multi-disk-based Mobile Systems)

김 영 진 [†] 권 권 택 [‡] 김 지 흥 ^{‡‡}
(Young-Jin Kim) (Kwon-Taek Kwon) (Ji-Hong Kim)

요 약 하드디스크는 가장 널리 사용되는 보조 기억 장치로 여겨져 왔으며 최근에 그 용도가 멀티미디어 응용 프로그램이나 게임과 같은 I/O 집중적인 응용 프로그램으로 인해 더욱 이동성 컴퓨팅 시스템에서 중요해지고 있는 실정이다. 그러나 디스크 드라이브에서의 상당한 전력 소모로 인해 이동성 시스템에서의 배터리 수명이 여전히 치명적으로 제약되고 있다. 본 논문에서는, 일반적인 모바일 플랫폼에서 (하나의 큰 디스크를 사용하는 대신에) 다중의 작은 디스크를 사용하는 것이 에너지 효율적이면서도 큰 성능 저하 없는 보조 저장 장치 해법이 될 수 있다는 것을 보인다. 또한, 본 논문에서는 디스크 데이터를 유관한 그룹들로 분류하고 이 그룹들을 같은 디스크로 이동시키는, 새로운 에너지 절감 기법을 제안한다. 또한, 이 기법을 기존의 데이터 집중 기법과 비교하고 나아가 이 기법들을 결합한다. 실험 결과, 제안한 기법을 단일 2.5" 디스크에 적용한 경우 보다 한 쌍의 1.8" 디스크에 적용한 경우, 충분히 작은 I/O 응답 감소 증가만을 가지면서도 최고 34%의 에너지 절감을 얻을 수 있었다. 또한, 기존의 데이터 집중 기법에 대해서도 최고 14.8% 이상의 에너지 절감과 최고 11배 이상의 I/O 응답 시간 개선을 얻을 수 있었다.

키워드 : 에너지 절감, 다중 디스크, 이동성 시스템, 연관 데이터 배치 및 이동

Abstract Hard disks have been the most prevalent secondary storage devices and these days their usage is becoming more important in mobile computing systems due to I/O intensive applications such as multimedia applications and games. However, significant power consumption in the disk drives still limits battery lifetimes of mobile systems critically. In this paper, we show that using several smaller disks (instead of one large disk) can be an energy-efficient secondary storage solution on typical mobile platforms without a significant performance delay. Also, we propose a novel energy-efficient technique, which clusters related data into groups and migrates the correlated groups to the same disk. We compare this method with the existing data concentration scheme, and also combine them. The experiments show that our technique saves the energy consumption up to 34% when a pair of 1.8" disks is used instead of a single 2.5" disk with a negligible increase in the average response time. The results also show that our method also saves up to 14.8% of disk energy consumption and improve the average I/O response time by up to 10 times over the existing scheme.

Key words : energy conservation, multiple disks, mobile systems, correlated data placement

1. 서 론

이동성 및 유비쿼터스 컴퓨팅 시스템은, 도처에서 연

결 가능한 네트워크 통신과 고성능의 컴퓨팅 수행 환경을 바탕으로 최종 사용자들에 의해 이제 폭넓게 받아들여지고 있다. 특히, PDA(Personal Digital Assistant), PMP(Portable Multimedia Player), MP3 재생기 및 다른 이동성 컴퓨팅 기기의 소유하고 사용하는 추세는 해마다 크게 증가하고 있다. 이에 따라서 그러한 플랫폼에 내장되거나 연결될 수 있는 하드 디스크 드라이브에 수요 또한 지속적으로 증가하고 있다. 현재는 서버 또는 데스크탑 PC용 디스크에 대한 수요가 작은 크기의 디

[†] 정 회 원 : 서울대학교 전기. 컴퓨터공학부
youngjk@davinci.snu.ac.kr

[‡] 정 회 원 : 삼성종합기술원
kwontaek.kwon@samsung.com

^{‡‡} 종신회원 : 서울대학교 전기. 컴퓨터공학부 교수
jihong@davinci.snu.ac.kr

논문접수 : 2006년 8월 3일
심사완료 : 2007년 1월 2일

스케에 대한 수요보다 많지만 조만간 이 두 시장이 비등해질 것으로 예측되고 있다[1].

일반적으로 하드 디스크 드라이브는 컴퓨터 시스템에서 10~20%의 전력 소모를 가지면서 상당히 큰 전력 소모 주체로 자리 매김해 오고 있다. 특히, 대부분의 이동성 컴퓨팅 시스템은 배터리에 전원을 의존하므로 디스크 드라이브의 에너지 소모를 제어하는 것은 매우 중요하다. 그러므로 디스크 드라이브에서 소모되는 에너지를 줄이는데 많은 노력이 기울여져 왔다.

하지만, 기존의 에너지 효율적인 기법들은 주로 단일 디스크를 가진 이동성 시스템에 한정되어 연구되어 왔는데, 큰 고성능 및 고전력의 디스크 하나를 크기가 작으면서 전력 소모가 낮은 디스크를 다수 개 사용함으로써 얻을 수 있는 에너지 절감의 가능성에 대해서는 잘 고려하지 않았다. 이러한 이동성 저장장치 시스템에 대한 연구 경향에 비하여 서버 저장장치 시스템에서는 최근에 다중 디스크들의 배열 사용에 기반 하는 에너지 절감 기법들이 많이 연구되었다. 그러한 기법들은 데이터 센터나 네트워크 서버를 대상으로 하는 응용 프로그램을 목표로 하고 있지만, 다중 디스크 간에 I/O 접근을 적절히 분배함으로써 에너지 절감을 이끌어 낼 수 있다는 기본 생각은 이동성 저장장치 시스템에 충분히 적용 가능하다.

이러한 기반 배경을 바탕으로 크고 전력 소모가 많은 디스크(가령, 2.5" 디스크)를 작고 전력 소모가 적은 디스크(가령, 1.8" 디스크)로 치환한다면 시스템에서의 에너지 절감은 명백할 것이다. 그러나 그러한 치환은 동일한 데이터 처리량(throughput)과 디스크 용량(capacity)이 치환 전후에도 유지되도록 보장이 되어야한다. 이러한 제약을 만족시키려면 디스크간의 1:1 치환보다는 1:N ($N>1$)의 치환, 즉 1개의 큰 고전력 디스크를 다중의 작은 저전력 디스크를 치환할 수밖에 없다. 이러한 목적을 위해서는 다양한 이동성 기기용 디스크들을 후보로 하여 에너지, 성능, 용량 등의 관점에서 조사하는 것이 필요하다. 그리고 다중의 디스크들 간에 데이터가 어떻게 위치되고 또 I/O 접근이 어떻게 분배되어야 하는 것도 중요하게 된다.

본 논문에서의 목표는, 하나의 노트북용 디스크를(보통, 2.5" 디스크) 대신하여 다중의 더 작은 크기를 갖는 디스크들을 가지는 이동성 저장장치 시스템에 대해, 수행되는 이동성 기기용 워크로드(workload)로부터 분류되는 연관 데이터 그룹을 동적으로 디스크 간에 새롭게 배치하는 기법을 통해 실제적이고 많은 에너지 절감을 얻는 것이다.

원하는 해법을 위해서, 먼저 작은 크기의 디스크들에 대해 이동성 시스템에서 에너지 절감을 위해서 어떤 후

보들이 사용될 수 있는지를 면밀히 살펴보았다. 또한, 다중의 디스크들을 사용하므로 데이터의 적절한 배치가 I/O의 분산 또는 집중과 크게 관련되며 결과적으로 디스크에 대한 접근으로 나타나므로, 이러한 데이터의 디스크 간 배치가 에너지 절감과 성능 보장에 영향을 미치는 지를 파악하였다. 특히, 이동성 기기들을 대상으로 하므로 이동성 워크로드에서 나타나는 디스크 접근 패턴에서의 데이터 간 연관 관계를 파악하여 같은 그룹으로 분류하여 관리하고 이 그룹들 중 많이 접근되는 그룹들은 동일한 디스크로 이동함으로써 디스크에 대한 I/O를 집중하도록 기법을 제안하였다. 이 기법을 사용함으로써, 이동성 기기용 워크로드에 따라 자주 접근되는 디스크는 활성 상태(active state)로, 덜 자주 접근되는 디스크는 유휴 상태(idle state) 또는 휴면 상태로 됨으로써 많은 에너지 절감을 얻을 수 있도록 하였다. 트레이스 기반의 실험을 수행한 결과, 제안한 기법을 단일 2.5" 디스크에 적용한 경우 보다 한 쌍의 1.8" 디스크에 적용한 경우, 충분히 작은 I/O 응답 감소 증가만을 가지면서도 최고 34%의 에너지 절감을 얻을 수 있었다. 또한, 기존의 데이터 집중 기법에 대해서도 수행 시나리오에 따라서 최고 14.8% 이상의 에너지 절감과 최고 11배 이상의 I/O 응답 시간 개선을 얻을 수 있는 것으로 나타났다.

본 논문의 구성은 다음과 같다. 2절에서는 다중 디스크를 사용하는 기존의 저전력 연구 내용을 소개한다. 3절에서는 본 논문의 연구 동기를 가져오는, 작은 크기의 디스크의 다중 사용에 의한 에너지 절감 가능성 타진과 이동성 시스템에서의 워크로드에 대한 관찰 및 분석에 대한 내용을 언급한다. 4절에서는 다중 디스크를 사용하는 기존의 에너지 절감 기법과 제안하는 기법을 소개 및 비교하고 이들의 에너지 효율적인 결합에 대해 설명한다. 5절에서는 제안한 기법을 검증하기 위해 개발된 시뮬레이터와 시뮬레이션 결과들을 보인다. 마지막으로 6절에서 본 논문의 결론을 맺고 향후 연구 과제를 논의한다.

2. 관련 연구

하드 디스크에서의 저전력 기법은 많은 연구자들에 의해 연구되어 왔는데, 특히 근래에 다중 디스크의 배열을 사용하여 보조 저장장치 시스템에서의 에너지 소모를 절감하는 기법들이 발표되었다[2-6]. [2]와 [3]에서는 높은 속도의 서버 디스크를 작은 크기의 다중 디스크 배열로 치환함으로써 거의 같은 전체 I/O 처리량을 유지하면서도 에너지 절감을 얻는 방법을 연구하였다.

[2]의 연구에서는 노트북용 디스크와 SCSI 디스크의 조합이 네트워크 서버에서 에너지 절감을 가져오는 것

을 확인하였다. 이 연구에서는 고성능 디스크를 작은 저전력 디스크들로 교체하는 것을 검토하였으나 깊이 들여다보지 않았으며, 단순히 고성능 디스크와 저전력 디스크의 내용이 동일하게 유지된다고(mirroring) 가정하였고 데이터의 디스크 간 이동에 대해서는 고려하지 않고 있다. 이에 대해 본 연구는, 서버가 아닌 모바일 시스템에 대해서 실제적인 디스크의 교체에 대해 고찰하였으며 데이터의 배치와 더불어 데이터의 동적 이동을 고려하고 있다. 또한, 전통적인 디스크 모델보다는 현실적으로 사용이 불가한 이중 속도 디스크¹⁾ 모델과 여기에서의 전력 관리 기법을 적용하여 에너지를 절감하는 것에 주력하였다. 이에 대해 본 연구에서는 전통적인 디스크 모델 및 실제적인 전력 관리 기법을 사용하고 있다.

[3]에서는, 에너지 효율성과 최악 시간 응답 시간(worst-case response time)간의 타협점을 고려하면서 서버용 디스크를 노트북용 디스크의 배열로 치환하여 에너지 절감을 얻는 것을 제안하였다. 그러나 이 연구에서는 서버용 디스크의 원래의 모든 데이터가 이미 다른 세 노트북용 디스크들에 각기 복사되어 있다고 가정하고 있으며 동적인 데이터 배치 또는 이동에 대한 고려가 없다. 디스크 교체에서는 본 연구와 비슷한 접근을 취하고 있으나, 본 연구에서는 임의의 디스크 간 데이터 초기 배치에 관계없이 동적 이동을 고려하고 있으며 제안한 데이터의 동적 이동 기법은 모바일 플랫폼상의 작업 부하 특성을 반영하고 있다.

서버용 디스크 워크로드에서는 보통 접근간의 유희 시간이 짧아서 디스크를 저전력 상태로 진입시키는 것이 에너지 측면에서 비효율적인 때가 많다. [4]에서는 이러한 점을 극복하기 DRPM(Dynamic Rotation per Minute)라는 기법을 디스크 배열을 위해 제안하였다. 이 기법은 다중 속도 디스크²⁾ 모델을 사용하여 I/O 처리량보다는, 요구되는 평균적인 응답 시간과 I/O 처리량의 길이에 따라서 디스크의 속도를 변화시킴으로써 짧은 유희 시간에 대해서도 에너지 절감을 효율적으로 할 수 있도록 하는 기법이다. 그러나 DRPM에서도 디스크간의 데이터 이동에 대한 고려가 없다.

PDC(Popular Data Concentration) 기법은 디스크 배열중 일부의 디스크에 자주 접근되는 디스크 데이터를 동적으로 이동시키는 기법이다[5]. 이 연구에서는 I/O 부하를 일부의 디스크들에 집중시키고 다른 디스크들을 저전력 상태로 진입하도록 하여 저전력을 얻는 것

을 목적으로 하고 있다. 또한, 이 기법은 웹 서버를 포함한 네트워크 서버에서 전통적인 디스크 모델이 아닌 이중 속도 디스크에서 보다 많은 에너지 절감을 얻을 수 있는데 이것은 디스크 데이터의 동적 배치 필요성에 따른 이동의 에너지 부하가 이중 속도에 대해서 전체 저장장치 시스템에 미치는 영향이 작게 나타남을 의미한다. 시뮬레이션 결과에서 나타나듯이 PDC 기법은 I/O 요구율이나 데이터 이동 주기 등의 인자가 바뀌에 따라서 에너지 효율과 성능 지연이 상당히 가변하는 것으로 나타났다. 이에 비해 본 연구는 연관된 파일들을 이동시킴으로써 부하를 분산하면서도 에너지 절감을 얻는 방식을 채택하고 있어 가변하는 작업 부하에 대해서도 덜 영향을 받는 구조를 취하고 있다.

한편, Hibernator 기법은 보다 개선된 속도 설정과 데이터 이동 방법을 결합하여 여러 속도가 가능한 다중 디스크들의 배열에 대해 SLA(Service-Level Agreement) 사양에 의해 보장되는 응답 시간 기준을 만족시키면서도 에너지를 절감하고 있다[6]. 이 기법은 중복성(redundancy)을 위해서 RAID5와 유사한 분산 블록 저장(striping)을 취하고 있다. 그러나 사용된 데이터 이동 기법은 분산 블록 저장된 데이터중의 일부를 이동하게 함으로써 주로 데이터베이스 서버용에 사용되도록 설계되었다. 이에 비해 본 연구에서는 다중 속도 디스크 모델 대신에 전통적인 디스크 모델을 사용하며 데이터 서버 대신에 모바일 시스템을 대상으로 데이터 전체를 이동하는 기법을 사용한다.

정리하면, 위에서 살펴본 연구들은 모두 서버용 워크로드를 대상으로 하고 있으며 본 연구에서는 이동성 기기상에서의 워크로드를 대상으로 하고 있다. 그러나 앞에서 언급한 바와 같이 데이터 배치 및 이동을 통해서 I/O를 집중시킴으로써 에너지를 절감한다는 기본 생각은 다중 디스크를 가진 이동성 컴퓨팅 시스템에 대해서도 충분히 적용 가능하다. 또한, 기존의 일부 기법들은 아직까지 실제적으로 사용이 불가한 이중 속도 또는 다중 속도 디스크 모델과 그에 따른 전력 관리 기법을 사용함으로써 실제적인 디스크 저장장치 시스템에서의 에너지 및 성능과는 괴리가 있었다. 본 논문에서는 전통적인 디스크 모델을 기반으로 하는 전력 및 성능 모델을 이용한다.

더불어, 본 연구에서는 다중 디스크를 사용하되 중복성에 의한 신뢰성(reliability)나 유용성(availability)의 증진에 초점을 맞추기 보다는 에너지 절감과 허용가능한 정도의 응답 시간에 초점을 맞추고자 한다. 이것은 바로 이동성 시스템에서는 배터리 제한과 공간 및 무게 제한 등으로 인해 한정된 수의 디스크만을 수용할 수 있기 때문이며, 보통 높은 수준의 데이터 처리량이 잘

1) 이중속도 디스크(2-speed disk): 2가지 속도가 가능한 디스크이며 현재 산업계에 나와 있지는 않음.

2) 다중 속도 디스크(Multiple-speed disk): 다수개의 속도를 가질 수 있도록 하는 디스크로 현재 산업계에 나와 있지 않음.

요구되지 않는 특성이 있기 때문이다.

3. 연구 동기

본 연구는 이동성 디스크 저장장치와 워크로드에서 다음의 두 가지 주요 관찰들을 연구 기반으로 하여 이루어진다. 첫째, 큰 고성능 및 고전력 디스크를 작은 저전력 디스크들을 다중으로 사용함으로써 큰 성능 저하 없이 상당한 에너지 절감을 얻을 수 있었다. 둘째, 구분이 가능한 디스크 접근 패턴들을 산출하는 파일 I/O 요청간의 현저한 연관 관계가 존재하였다.

3.1 이동성 기기용 디스크의 특성

표 1에서는 주요 디스크 제조업체들로부터의 자료로부터 얻어진 다섯 개의 최신의 작은 크기(form factor)의 디스크들의 특성을 보이고 있다[7-11]. 이 디스크들은 용량, 속도, 전력 그리고 크기에서 각기 다른 값을 가지고 있다. Travelstar 80GN는 보통 노트북 컴퓨터에서 사용되는 2.5" 디스크이며 나머지 디스크들은 휴대폰이나 PMP와 같은 휴대용 기기에서 사용되고 있다. 표 1로부터 플래터³⁾의 크기가 큰 디스크일수록 더 많은 전력을 소모하지만 더 좋은 성능을 보임을 알 수 있다. 이것은 디스크의 플래터가 더 빨리 회전하므로 데이터 접근을 더 빨리 할 수 있기 때문이다. Travelstar 80GN의 경우는 2.3W를 활성 상태에서 소모하는데, 순수하게 활성 상태에서의 전력 소모끼리 비교하더라도 이 값은 MK4004GAH의 전력 소모의 1.6배에 해당하며 MK4001MTD보다 3.8배가 많다. 성능 측면에서는 Travelstar 80GN는 모든 다른 작은 디스크들보다 최소한 25% 작은 지연 시간을 가지는데, 이는 디스크 속도와 평균 탐색 시간으로부터 쉽게 추정이 가능하다.

표에서 보는 바와 같이, MK4004GAH는 전력과 성능 측면에서 다른 디스크들에 비해 가장 큰 디스크인 Travelstar 80GN에 대적할 만한 유일한 디스크이다. 더구나, MK4004GAH는 나머지 세 디스크들과는 달리 이동성 기기를 위한 충분한 용량을 가지고 있다. 따라서 하나의 노트북용 디스크(즉, Travelstar 80GN)를 다중의 작은 크기의 디스크로 치환하려고 할 때, 그 노트북용 2.5" 디스크에 버금가는 성능과 용량 그리고 보다 적은 전력 소모로 인해 MK4004GAH는 강력한 후보가 될 것이다. 하지만, 이러한 치환은 특정한 디스크 모델에 한정될 필요는 없으며 본 연구에서는 단지 주변에서 쉽게 구할 수 있는 2.5"와 1.8" 디스크를 사용하였다. 본 논문에서는 이동성 플랫폼에서 하나의 큰 디스크를 다중의 작은 디스크로 치환하는 초기의 시도로서 하나의

2.5"를 2개의 1.8" 디스크로 치환하고자 한다.

먼저, 하나의 2.5" 디스크가 I/O 요청을 주어진 시간의 15%동안 처리하고 나머지 85%의 시간동안은 유휴 상태에 있다고 가정하자. 표 1의 값들로부터 10분 동안의 전체 에너지 소모는 691.5J이 될 것이다. 만약, 2.5" 디스크를 2개의 1.8" 디스크로 치환하고 또 같은 워크로드가 두 디스크간에 균등히 분배가 된다고 하면 이 두 디스크는 각각 주어진 시간의 9.38%동안 활성 상태로 I/O를 처리하게 될 것이다. 이것은 고려중인 1.8" 디스크가 2.5" 디스크보다 25% 더 큰 응답 성능을 보이기 때문이다. 두 디스크가 소모하는 전체 에너지는 592.56 J이 될 것이며 이것은 하나의 2.5"가 사용하는 에너지보다 약 14%를 절감하게 되는 것이다. 만약 I/O가 하나의 1.8" 디스크에만 집중될 수 있다면 나머지 디스크는 더 낮은 저전력 상태인 대기 상태(standby state)로 유지될 수 있다. 이때 에너지 절감율은, I/O 집중이 없이 하나의 2.5" 디스크를 사용했을 때에 비해 무려 32%에 이르게 된다. 이러한 관찰 결과는 바로 디스크간의 치환을 통해서 얻을 수 있는 에너지 이득의 정도를 제시해주는 것으로 볼 수 있다.

3.2 이동성 플랫폼상의 워크로드 특성

데이터베이스나 웹 서버에서의 워크로드와 비교하면 이동성 컴퓨팅 시스템에서의 워크로드는 응용 프로그램들의 수행시에 더 자주 변하는 경향이 있으며 또한 프로그램들간의 잦은 전환에 따라 불규칙적으로 움직인다. 그럼에도 불구하고, 노트북을 포함한 이동성 컴퓨팅 환경에서 규칙적인 워크로드 패턴을 발견할 수 있는데 공통적인 응용 프로그램들이 자주 사용되거나 순서를 가지고 사용되는 경우에서 바로 그러하다. 예를 들면, 노트북 컴퓨터를 사무를 위한 여행에 가지고 간다면 그 주 용도는 아마도 이메일을 주고받거나 발표 문서를 편집하거나 파일 운용을 하기 위해 원격의 서버에 연결을 하는 일 등이 될 것이다. 비슷한 사용 패턴은 노트북을 주로 데스크탑 PC 대신에 사용하는 사용자에게서도 이메일 클라이언트 프로그램, 프로그래밍 개발 도구 프로그램, 문서 편집기, 멀티미디어 재생기 등을 반복적으로 사용하는 형태로 나타날 수 있다. 이동성 플랫폼에서 자주 발견되는 워크로드를 특징짓기 위해 [12]에서는 9개의 전형적인 이동성 플랫폼용 응용 프로그램을 선택하여 *programming*과 *networking*으로 지칭하는 2개의 워크로드 수행 시나리오를 작성하였다. 이 각각의 시나리오는 각각의 응용 프로그램 수행 행태를 모사한다. Programming 시나리오는 8개의 단계로 이루어져 있으며 각 단계동안 프로그램과 관련된 응용 프로그램들, 가령 make, grep 등이 수행된다. Networking 시나리오는 3개의 단계로 이루어져 있으며 thunderbird(이메일 클

3) platter, 하드 디스크 드라이브에서 실제로 정보를 저장하는 부분. 알투 마늘 등의 재질로 된 금속 원판으로 자성체가 코팅되어 있음

표 1 노트북용 디스크와 작은 크기 디스크들의 특성

Form factor / Model		2.5"	1.8"	1"		0.85"
		Travelstar 80GN (HGST)	MK4004GAH (Toshiba)	ST1.3 (Seagate)	Microdrive 3K8 (HGST)	MK4001MTD (Toshiba)
Capacity (GB)		80	40	12	6, 8	4
Rotational speed (rpm)		4,200	4,200	N/A	3,600	3,600
Avg. seek time (ms)		12	15	N/A	12	16
Power (W)	Active	2.3	1.4	0.792	0.627	0.6
	Idle	0.95	0.4	0.254	N/A	0.45
	Standby	0.25	0.2	0.0429	N/A	0.12
Physical size	Weight (g)	99	62	14	13	8.5
	Area (cm ²)	70	42.39	12	12	7.68

라이언트 프로그램), ftp, xmms(MP3 재생기)와 같은 네트워크를 이용한 통신 프로그램이나 멀티미디어 프로그램이 수행된다.

일반 사용자들이 그들의 노트북 컴퓨터로 이러한 종류의 작업을 할 때, 데이터 파일은 응용 프로그램들 간에 공유되는 경우가 많게 된다. 이런 공유된 파일들은 그렇지 못한 파일들에 비해 보다 자주 접근되는 경향을 보인다. 예를 들면, 디스크 내에서 어떤 파일을 검색하여 찾은 다음에 사용자가 그 파일을 편집하고 컴파일을 하려고 시도할 수 있다. 서버에서는 잘 일어나지 않는 이러한 패턴들은 연관된 파일 요청들, 즉 하나의 응용 프로그램 내에서 관련된 파일들에 대해 또는 여러 응용 프로그램들간에 공유된 파일들에 대해 동일한 순서로 실행되는 요청들은 많은 에너지 절감을 얻기 위해서 다중 디스크 저장장치 시스템 내의 하나의 디스크로 I/O 부하를 집중하도록 사용될 수 있다는 것을 제시한다. 그러나 PDC와 같은 기존의 저전력 I/O 부하 집중 기법들은 서버 워크로드를 대상으로 해서 설계되었으므로 이동성 컴퓨팅 환경에서의 워크로드에 대해서는 만족할 만한 결과를 낼 수 없을 가능성이 높다.

4. 다중 디스크를 사용하는 에너지 절감 기법

3장에서 제안한 바와 같이, 단일 2.5" 디스크를 2개의 1.8" 디스크들로 교체하는 것은 상당한 에너지 절감을 가져올 수 있을 것이다, 그러나 여기에는 이동성 기기용 워크로드에 대한 적절한 데이터 배치 기법이 요구된다. 이장에서는 연관된 파일 요청들을 한쌍의 디스크들 중의 하나에 집중시키고 나머지 디스크는 더 많은 여유 시간을 가지게 하는 새로운 에너지 절감 기법을 제안한다. 그다음, 에너지와 성능의 관점에서 기존 기법인 PDC 기법을 재검토할 것이다. 마지막으로, 좀더 건설하고 효율적인 에너지 절감을 위해 이 두 기법을 결합하는 것에 대해 기술을 하고자 한다. 이후로 제안하는 기법을 연관 데이터 집중 기법으로 지칭하기로 한다.

4.1 연관 데이터 집중 기법

연관 데이터 집중 기법의 기본 개념은 2개의 디스크 중 하나에 연관 데이터(correlated data)를 집중하는 것이다. 연관 데이터는 접근이 고정된 같은 순서로 반복해서 발생하는 데이터를 지칭한다. 이 데이터는 하나의 응용 프로그램에 의해 고정된 같은 순서로 데이터 요청이 발생하거나 (가령, 라이브러리 또는 상황 설정 파일들에 대한 요청) 응용 프로그램들 간에 파일들이 공유되도록 디스크 저장 시스템으로 보내어 지는 경우 (가령, 하나의 응용 프로그램의 출력이 다른 응용의 입력이 되는 경우)에 발생한다. 예를 들면, 파일들이 A, B, C, D, E, B, C, D의 순서로 접근된다고 가정하자. 그러면, B, C와 D는 연관 데이터의 하나의 그룹으로 식별될 것이다. B, C와 D가 자주 접근 되고 모두 하나의 디스크 상에 존재한다면, 이들 파일들이 접근될 동안 다른 디스크는 긴 여유 시간을 가질 수 있게 되고 더 낮은 전력 소모 상태에서 유지될 수 있을 것이다. 이동성 플랫폼상의 워크로드에서는, 3절에서 기술한 것처럼, 이러한 연관 데이터의 그룹들이 많이 발견되므로 제안하는 기법은 상당한 양의 에너지를 절감할 수 있게 한다. 연관 데이터 집중 기법의 목적은 연관 데이터의 대부분을 하나의 디스크가 저장할 수 있도록 두개의 디스크 간에 데이터를 분배하여 배치하는 것이다.

다음은 연관 파일 배치 기법의 알고리즘을 의사 코드(pseudo code)로 나타낸 것이다. 연관 기법은 전체 알고리즘 코드에서 보는 바와 같이 크게 2가지 단계로 동작한다. 첫째, 같은 순서로 발생하는 파일 접근들을 식별하여 연관 그룹 리스트내의 원소로 등록한다 (3-6줄). 둘째, 미리 설정된 기준값보다 자주 접근된 연관 데이터 그룹에 대해 디스크 간 이동에 의한 데이터 재배치가 수행된다 (7-16줄). 모든 연관 데이터 그룹은 연관 파일들의 확인자(identifier)들과 연관 그룹의 접근 횟수 카운터로 이루어져 있다.

<전체 알고리즘 코드>

1. 매 파일 접근마다 현재 파일에 이어서 접근되는 파일들의 포인터를 유지함
 2. 매 파일 접근마다 FIFO 리스트에 파일 접근 정보를 push함
 3. FIFO 리스트내에서 파일 접근 순서를 추적함
 4. if 최근 파일 접근 순서가 반복
 5. if 반복된 접근 길이 > COR_FILEGRP_SIZE 또는 반복된 접근이 끝
 6. 연관 그룹 리스트에 이 파일들의 열을 원소로 등록함
 7. 연관 그룹 리스트의 모든 원소들에 대해 FIFO 리스트의 현재 파일 접근 열을 비교함
 8. if FIFO 리스트 길이 > COR_FILE_HISTORY_SIZE
 9. FIFO의 머리 원소를 pop함
 10. if 동일한 접근 열 (즉, 연관 그룹) 발견
 11. if 열의 접근 횟수 > COR_MIGRATION_THRESHOLD
 12. 대상 디스크 선택 루틴 호출
 13. 재배치 루틴 호출
 14. 열 접근 횟수 카운터 = 0;
 15. else
 16. 열 접근 횟수 카운터++;
 17. else
 18. Goto 7
-

<대상 디스크 선택 루틴>

1. 가능하다면, 해당 연관 그룹의 파일 전체가 옮겨 갈 수 있는 디스크 선택
2. 그렇지 않으면, 해당 그룹의 파일들의 블록이 가장 많이 존재하는 디스크 선택

<재배치 루틴>

1. 재배치 대상 그룹의 파일들(i = 처음부터 끝)을 처음부터 끝까지 확인함
 2. if 파일 i의 디스크 == 대상 디스크
 3. Goto 1
 4. else
 5. if 대상 디스크의 여유 공간 < 파일 i의 크기
 6. 대상 디스크 내에서 희생 파일 검색
 7. if 희생 파일 발견
 8. 희생파일을 대상 디스크로부터 파일 i의 디스크로 이동
 9. Goto 12
 10. else
 11. 파일 이동 실패 and goto 1
 12. 파일 i를 i의 디스크로부터 대상 디스크로 이동
-

COR는 연관 데이터 그룹들을 식별하고 분류해내기 위해서 바로 이전에 접근된 각 파일로부터 현재 접근되는 파일에 대한 포인터를 만들어낸다. 물론 파일들은 여러 개의 다른 고정된 순서로 섞여서 접근 되겠지만 이 기법에서는 오직 미래에 같은 순서가 발생하는 지는 바로 직전의 접근들이 같은 순서를 따르는지에 대한 정보만을 기록하고 조사한다 (1줄). [13]에서는 현재의 파일 접근에 대해 마지막으로 접근되었을 때 뒤따르던 파일이 다시 뒤따라 접근될 확률이 현대 I/O 시스템에서 72%에 이른다고 보고하고 있다. 이 사실은 많은 파일 접근 순서들을 가장 최근의 파일 접근들을 추적하는 것

만으로도 충분히 발견가능하다는 것을 제시한다. 향후 연구에서는 더 많은 과거 정보를 바탕으로 하여 많고 다양한 연관 데이터를 얻는 것을 고려하고 있다.

COR는 또한 파일 접근 과거 이력을 관리하기 위해서 FIFO(First In First Out) 리스트를 유지하는데 (줄 2) 그 과거 이력의 길이는 COR_FILE_HISTORY_SIZE 값에 의해 제한된다 (줄 8-9). 동일한 접근 열이 FIFO 리스트 내에서 반복되면 이 접근 열의 파일들은 연관 데이터 그룹으로 등록된다. 반복 열의 최대 길이를 제한하기 위해서 COR_FILEGRP_SIZE 값을 사용한다 (줄 5). 동시에, 매 파일 접근 마다 FIFO 리스트의 파일 접근 과거 이력은 이미 등록되어 있는 연관 그룹 리스트 내의 연관 그룹들과 비교된다. 이때, 만일 현재 분류한 연관 그룹이 이미 등록되어 있고 이 연관 그룹이 COR_MIGRATION_THRESHOLD를 넘어서는 접근 횟수를 가진다면 재배치 루틴을 수행하게 된다 (줄 11-13). 그렇지 않다면 그 접근 횟수 카운터를 그냥 1 증가한다 (줄 16). 이러한 동작을 모든 연관 그룹 리스트 내의 모든 원소들에 대해 수행을 하며 (줄 18) 끝나게 되면 다시 알고리즘의 처음으로 돌아가서 각 파일의 접근을 추적하게 된다. 이러한 과정들을 반복함으로써 자주 접근된, 같은 그룹에 있는 파일들은 같은 디스크 상에 위치하게 되고 I/O 들은 이 데이터에 집중되게 될 것이다.

COR는 연관 그룹 파일들을 이동 시에 이동하는 데이터 블록의 개수를 최소화하기 위해 대상 디스크를 선택하는 2가지 기준을 가지고 있다 (대상 디스크 선택 루틴, 줄 1-2). 첫째, 전체 그룹의 파일들이 모두 옮겨 갈 수 있는 디스크를 선택한다. 그렇지 않으면, 해당 그룹의 파일들의 블록이 가장 많이 존재하는 디스크 선택한다. 대상 디스크가 선택되면 COR는 위 재배치 루틴에서 보는 바와 같이 해당 연관 그룹내의 모든 파일들에 대해 재배치를 수행하게 된다. 각 파일이 옮겨갈 대상 디스크 내에 여유 공간이 충분히 있으면 그대로 이동을 시키나 (재배치 루틴, 줄 12) 여유 공간이 부족한 경우는 대상 디스크 내에서 희생 파일을 선택하고 이를 대상 디스크에서 옮길 파일이 있는 디스크로 이동한 다음에 (재배치 루틴, 줄 5-9) 다시 파일 이동을 수행한다. 희생 파일을 선택하는 기준은, 파일 크기가 옮겨야 할 파일 크기보다 크면서 이동 대상 파일의 디스크 여유 공간보다는 작아야 하며 연관 그룹으로 분류되지 않은 파일을 선택하도록 하고 있다. 이러한 동작을 반복함으로써 점진적으로 같은 디스크 상에 같은 연관 그룹 내의 파일들이 같은 디스크 상에 배치되게 될 것이다.

4.2 기존 데이터 집중 기법

PDC기법은 일부 네트워크 서버상의 워크로드에서 나타나는 매우 편중된 파일 접근들의 횟수를 다루기 위해

제한된 기법이다[5]. 예를 들면, 웹 서버에서의 파일 접근들의 횟수는 매우 높은 계수를 가지는 Zipf 분포를 따르는 것으로 알려져 왔다. Zipf 법칙은 접근의 횟수를 예측하는데 파일에 대한 접근 횟수는 파일별 접근 횟수 순위의 역수에 비례한다. 이때, 워크로드에서의 계수가 높을수록(1에 가까울수록) 높은 순위에 위치해있는 파일 접근 횟수가 낮은 순위에 비해서 월등히 많게 되는 집중성이 나타나게 된다.

PDC의 기본 개념은 이렇게 자주 접근되는 파일 데이터를 다중 디스크들 중의 일부로 이동시켜서 I/O를 집중하는 것이다. 자주 접근되는 디스크들이 계속 활성 상태인 동안 자주 접근되지 않는 디스크들을 그만큼 더 여유 시간을 가지게 되고 에너지를 절감하기 위해 저전력 상태로 보내 질 수 있게 된다.

PDC는 수행 동작을 간단히 기술하면 다음과 같다. PDC는 접근 횟수를 순서별로 배열하여 파일 정보를 저장하면서 이 정보를 기반으로 다중 디스크들로 된 디스크 배열 간에 데이터를 재배치시킨다. 이때, 첫 디스크는 가장 접근수가 많은 데이터를 저장하며 두 번째 디스크는 그 다음으로 접근이 많은 데이터를 저장하는 식으로 데이터 배치를 하게 된다. 가장 접근이 적거나 전혀 접근되지 않는 데이터들은 최종 몇 개의 디스크에 저장될 것이다. 접근 수가 큰 디스크들은 “활성 디스크”들로 분류되며 접근 수가 작은 “비활성 디스크”로 분류된다. 파일들은 대상 디스크로 그 디스크 용량이 다 차거나 데이터 처리량이 최대 한도에 이를 때 까지 이동하게 된다. 본 논문에서는 디스크의 개수를 2개로 한정하고 있으므로 하나의 디스크에 상대적으로 많이 접근되는 데이터를 다른 디스크에 덜 접근되는 데이터를 저장하도록 데이터 이동이 일어나게 된다.

하지만, PDC기법은, 파일 접근이 시간에 따라 가변하는 경우에 상당한 수의 데이터 이동을 유발할 수 있으며 이 때문에 이동 자체에 의한 에너지 소모 뿐만 아니라 이로 인해 유휴 상태에 대한 디스크에 대한 접근 가능성이 커져서 에너지 절감 가능성을 떨어뜨리게 된다. 더구나, 새로이 파일들이 생성되는 경우 이들은 모두 가장 덜 접근되는 데이터가 모인 디스크에 저장되어야 하는데 이는 이 디스크의 저전력 상태를 방해하게 된다. 그러므로 PDC기법이 이동성 플랫폼용 워크로드에 적용된다면 파일 접근 횟수의 가변성으로 인해 에너지 절감의 폭이 줄어들 것이다. 우리가 제안하는 기법은 이런 단점을 극복하고 있다.

4.3 혼합 기법

PDC기법은 가장 자주 접근되는 데이터를 다중 디스크들의 일부로 옮기므로 불필요하게 많은 수의 데이터 이동을 유발할 수 있다. 예를 들면, 과거에 자주 접근되

었던 한 디스크상의 파일들이 현재 자주 접근되는 경우를 고려하자. PDC기법은 현재 다수의 접근 횟수를 가지면서 접근되고 있는 디스크를 활성 디스크로 유동적으로 결정하기보다는 이미 결정되어 있던 활성 디스크로 자주 접근되는 파일들을 이동시키려고 시도한다. 즉, 유연하지 않게 기존 디스크 분류(활성, 비활성 디스크)를 유지하는 것은 상당한 수의 추가적인 디스크 간 데이터 이동을 야기할 수 있다.

이 상황에서, 덜 활성인 디스크 상에서 최근에 접근된 데이터가 연관 데이터로 식별된다면, 제한한 연관 데이터 집중 기법은 덜 활성이었던 디스크를 활성 디스크로 지정할 것이며 데이터를 이동시키지 않을 것이다. 하지만, 연관 데이터 집중 기법의 현재 버전은 이미 지정된 디스크들에 대해 연관 데이터 그룹별로 배치하는 방법은 가지고 있지 않으므로 추가적인 디스크 I/O를 집중할 기회를 잃을 수도 있다. 더구나, 최근에 접근된 데이터가 각 파일은 자주 접근되었을 지라도 전체가 낮은 연관성을 보인다면 이 기법은 에너지를 절감하기 위해 부하를 집중하는데 실패를 할 수도 있다. 이것은 이 기법에서는 각 파일에서의 접근 횟수에 대한 정보를 가지지 않기 때문이다.

따라서, 우리는 각각의 단점을 극복하기 위해 연관 데이터 집중 기법과 PDC기법을 결합하는 혼합 기법을 제안한다. 이 기법은 두 기법을 간단히 상보적으로 결합하는 것으로서 다음과 같이 동작한다. 먼저 PDC기법에 의해 파일 이동이 각 파일이 접근되는 횟수에 기반하여 일어나게 된다. 동시에 연관 데이터 집중 기법은 같은 디스크로 연관 데이터의 그룹들을 반복적으로 이동시키도록 노력한다. PDC기법과 연관 데이터 집중 기법에 의한 파일 이동 요청이 상충하게 되면 연관 데이터 집중 기법이 우선순위를 가지게 된다. 그래서 접근 횟수에 의해 파일 이동이 발생하면 현재 연관 데이터의 그룹으로 등록되지 않은 파일들만이 이동될 것이다.

5. 실험 및 결과

이 절에서는 먼저 우리의 시뮬레이션 플랫폼과 디스크 모델을 기술한다. 그런 다음 시뮬레이션을 수행하여 다양한 에너지 절감 기법들이 이동성 시스템의 다중 디스크들을 대상으로 얼마나 에너지 효율적으로 파일을 배치하고 부하를 집중시키는 지를 평가한다. 현재의 연구는 단지 2개의 디스크만을 사용하지만 더 많은 디스크의 사용을 통해 디스크 저장장치 시스템에서의 일반성과 견실성을 향상시킬 수 있을 것이다. 향후 연구에서 이를 수행하는 것을 계획하고 있다.

5.1 시뮬레이션 플랫폼

본 연구에서는 실험 보드로부터의 실제 워크로드를

사용하는 다중 디스크 전력 및 성능 시뮬레이터를 구현하였다. 이 시뮬레이터는 그림 1에서 보는 바와 같이 크게 세 부분으로 구성되어 있다: 1) 이동성 시스템에서 흔하게 나타나는 응용 프로그램의 패턴을 모사하는 워크로드 생성기(workload generator) 2) 다중의 디스크들에 대해 가변하는 파일 배치와 I/O 분배를 모사하는 다중 디스크 파일 I/O 시뮬레이션 모듈(multi-disk file I/O simulation module) 3) 디스크 접근 패턴을 기반으로 한 각 디스크의 에너지 및 성능을 추정하는 다중 디스크 전력 및 성능 시뮬레이션 모듈(multi-disk power and performance simulation module).

워크로드 생성기. 이 모듈은 특정한 이동성 플랫폼상의 사용자 응용 프로그램 수행 패턴을 모델링하며 그에 따른 I/O 트레이스를 생성해낸다. 지정한 이동성 기기 상에서 수행되는 응용 프로그램들은 [12]와 유사하게 미리 정의된 수행 시나리오를 반복적으로 실행하는 것을 가정한다. 또한, 응용 프로그램들의 종류는 파일 전송, 이메일 프로그램, 파일 검색, 미디어 재생 및 유희로 한정시킨다. 표 2는 PDA와 PMP에서의 대상 응용 프로그램들에 대한 단순화된 사용자 수행 패턴을 보여주고 있다. File transfer는 디스크 파일을 전송하거나 네트워크로부터 전달받으며 Email은 메일 메시지를 디스크상의 메일 박스로부터 읽고 다른 메일 박스로 옮긴다. File search는 디스크들에 저장되어 있는 파일들을 읽고 특정한 문자열에 대해 검색을 수행한다. Madplay는 멀티미디어 파일들을 기설정된 재생목록을 사용하여 재생한다. Sleep은 사용자가 잠시 동안 아무런 파일 운용도 하지 않는 동안을 모델링하므로 아무런 I/O 요청도 발생시키지 않는다.

워크로드 생성기는 단일 1.8" 디스크 (MK4004GAH)에 대해서 위의 시나리오를 따르면서 I/O 요청들을 생성한다. I/O 요청들은 페이지 캐시, 파일 시스템 및 리

표 2 단순화된 사용자 수행 패턴

이동성 기기 종류	응용 프로그램 수행 순서
PDA	file transfer → email → file search → sleep
PMP	file transfer → madplay → sleep

눅스 2.4 커널을 통해 디스크에 전달된다. 이러한 과정 동안 페이지 캐시를 통과하여 디스크에 도달하는 I/O 요청들이 이러한 과정동안 페이지 캐시를 통과하여 디스크에 도달하는 I/O 요청들이 I/O 트레이스로 저장된다. 이 트레이스들은 타임 스탬프, 파일 id, 파일내 블록의 오프셋, I/O 데이터 크기 및 접근되는 파일의 크기의 정보 들을 가진다.

다중 디스크 파일 I/O 시뮬레이션 모듈. 이 모듈은, 트레이스 들이 다중의 가상 디스크를 가지는 이동성 플랫폼으로 전달될 때 이 디스크들에 대해 정적인 파일 배치와 동적인 파일 이동을 모사한다. 정적인 파일 배치는 파일들이 디스크 접근이 일어나기 전에 디스크들 상에 분포되어 있도록 하는 방식을 언급한다. 그림 1에서 보는 바와 같이, 다중 디스크 할당 맵이 이 과정 중에 만들어지며 이 맵은 디스크들 상에 파일들이 어떻게 배치되어 있는지를 알려 준다. 동적인 파일 이동은 앞에서 기술한 에너지 절감 기법들이 디스크 간 파일 이동을 유발시킬 때 항상 일어나게 된다.

다중 디스크 전력 및 성능 시뮬레이션 모듈. 다중 디스크 I/O 파일 시뮬레이션 모듈에 의해 다루어지는 다중 디스크 I/O 트레이스들은 어떻게 각 디스크에 대한 I/O 동작들이 분배가 되는지를 나타낸다. 이 트레이스들로부터 이 모듈은 디스크 모델과 전력 제어 정책을 사용하여 각 디스크의 성능과 에너지 소모를 추정한다. 우리는 전력 제어 정책으로 전통적인 임계값 기반의 전력 관리 정책(Traditional threshold-based Power Mana-

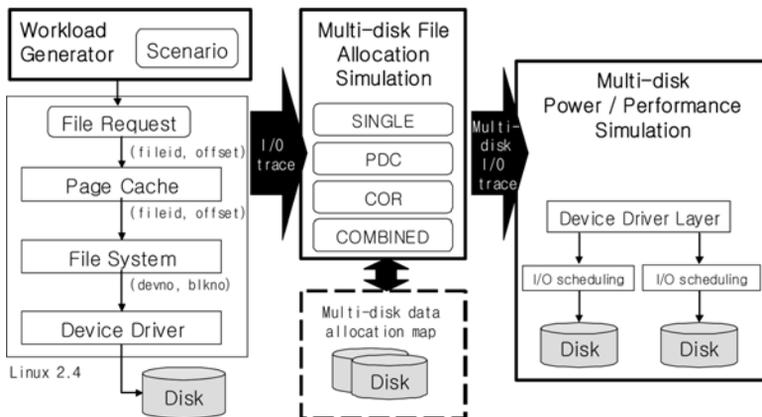


그림 1 시뮬레이터 전체 구조

gement)을 사용한다. 임계값 기반 전력 관리 정책은 고정된 임계값 기간동안 디스크가 유휴한 상태로 있으면 디스크를 더 낮은 전력 상태로 진입하게 한다. 다중 속도 디스크의 속도 제어 정책이 더 짧은 유휴 시간을 가지는 워크로드에 대해서도 에너지 절감의 효율성이 뛰어나지만 현재 산업계에서 이러한 디스크를 생산하고 있지 않으므로 우리는 기존의 디스크 모델과 임계값 기반의 전력 제어 정책을 사용한다.

사용하는 디스크 성능 모델은 다음과 같이 각 디스크의 전체 응답 시간을 디스크 큐 지연, 디스크 자체 지연, 그리고 요청당 서비스 시간의 합으로 추정한다.

$$T_{total\ response\ time} = \sum T_{queue\ delay} + \sum T_{disk\ delay} + \sum T_{service\ time}$$

$$T_{avg.\ response\ time} = T_{total\ response\ time} / N$$

평균 요청 응답 시간은 전체 응답 시간을 I/O 트레이스 내의 총 요청수 N으로 나누어서 얻는다. 디바이스 드라이버 레벨에서는 대기하는 I/O 요청들이 요청 큐에서 더 큰 하나의 I/O 요청으로 통합될 수 있는데 새로이 추가된 요청들이 기존에 대기하고 있는 요청들에 연속적인 경우에 가능하다. 이것을 I/O 클러스터링이라고 부르며 리눅스 커널에서는 이를 통해 I/O 요청의 수를 감소시킴으로써 성능을 개선하고 있다. 그러나 요청들이 큐에서 너무 오래 대기하면 응답 시간이 아주 커질 수 있으므로 구현한 시뮬레이터에서는 최대 70ms로 제한하였다.

큐로부터 디스크로 보내지는 각 응답은 서비스 시간에 의해 처리되며 서비스 시간은 시작하는 b 블록에서 l 개의 블록을 읽어내기 위해 다음의 요소로 이루어진다.

$$T_{service}(b, l) = T_{seek}(b, b_{last}) + T_{sfer}(l) + T_{avg.\ rotation\ delay}$$

탐색 시간 $T_{seek}(b, b_{last})$ 은 I/O 동작이 발생할 때, 블록 b_{last} 에 위치한 디스크 헤드를 블록 b 로 이동하는 데 걸리는 시간이다. 이 시간의 모델로는 $T_{seek}(b, b_{last}) = a_1|b - b_{last}| + a_0$

의 간단한 선형 근사모델을 사용한다. 이때, a_0 와 a_1 는 상수로서 디스크의 특성에 따라 가변하는 값들이다. 전달 시간 $T_{sfer}(l)$ 은 l 개의 블록을 읽거나 쓸 때 걸리는 시간으로 단위 전달 시간에 l 을 곱해서 구한다. 회전 지연 시간은 불규칙적으로 가변하므로 평균 회전 지연 시간인 $T_{avg.\ rotation\ delay}$ 를 사용한다.

디스크 i 의 전체 에너지 소모는 각 디스크에 의해 디스크의 각 전력 상태에서 소모되는 에너지와 모든 상태 전이 기간(transition period)동안에 소모된 에너지의 총합으로 구한다.

$$E_i = \sum_j P_{ij} \cdot T_{ij} + \sum_k \sum_l N_{ikl} \cdot E_{ikl}$$

디스크 전력 상태 j 는 활성(active), 유휴(idle) 그리고 휴면(standby) 상태중의 하나이며 각 상태에서 각 디스크는 다른 전력 소모값 P_{ij} 를 가진다. T_{ij} 는 디스크 i 가 각 상태에서 머무는 동안을 나타낸다. E_{ikl} 은 디스크 i 가 상태 k 에서 상태 l 로 전이할 때 소모되는 에너지를 나타내며 N_{ikl} 은 디스크 i 가 상태 k 에서 상태 l 로 전이하는 전이 횟수를 나타낸다.

5.2 시뮬레이션 결과

시뮬레이션 환경 설정. 시뮬레이션은 5.1절에서 기술한 디스크 모델과 I/O 트레이스를 사용하여 수행되었다. 두 종류의 다른 디스크 크기를 갖는 디스크들에 대한 자세한 전력 및 성능 상수 값들은 표 3에 주어졌다. 이 디스크들에 대한 성능 및 전력 상수값 들은 [7]과 [8]에 있는 값들과 같으며 다만 용량만 1.8" 디스크는 400MB로, 2.5" 디스크는 800MB로 한정하였다.

워크로드 생성기를 수행하고 I/O 트레이스를 추출하기 위한 실험 보드로는 1.8" Toshiba MK4004GAH 디스크가 장착되어 있고 PXA255가 탑재된 실험 보드를 사용하였다. 보드의 운영체제로는 리눅스 2.4를 사용하였다. 이동성 플랫폼상의 워크로드를 생성하기 위해서는

표 3 실험에 사용된 디스크 상수

디스크 크기 (")	2.5	1.8	
용량 (MB)	800	400	
회전 속도 (RPM)	4200	4200	
평균 탐색 시간 (ms)	12	15	
전력 (W)	활성	2.3	1.4
	유휴	0.95	0.4
	휴면	0.25	0.12
단위 시간당 활성에서 유휴 전이 시 소모 에너지 (J/s)	1.15/0.5	0.7/0.5	
단위 시간당 유휴에서 활성 전이 시 소모 에너지 (J/s)	1.15/0.5	0.7/0.5	
단위 시간당 활성에서 휴면 전이 시 소모 에너지 (J/s)	2.94/2.3	2.05/3.1	
단위 시간당 휴면에서 활성 전이 시 소모 에너지 (J/s)	5.00/1.6	1.84/1.7	
전력 제어 임계값: 유휴/휴면 (s/s)	1/3.379	1/5.979	
탐색 시간 모델값 (a_1, a_0)	($2.9^{-10}, 0.0072$)	($3.6^{-10}, 0.0090$)	

표 2의 PMP 사나리오와 PDA 사나리오에 따라서 각각 71분과 81분을 응용 프로그램들을 반복적으로 수행하였다. 요청 비율은 각 트레이스 별로 약 3(요청/초)와 약 10(요청/초)로 나타났다.

우리는 실험을 통해 디스크 간에 서로 다른 파일 배치와 I/O 분배를 하는 5가지 방법을 에너지 효율성과 성능 면에서 평가하고 비교하였다. 이 5가지 방법은 표 4에서 주어져 있다. SINGLE은 단일 2.5" 디스크를 이용하며 데이터 이동이 존재하지 않는다. 나머지 다른 방법들은 모두 2개의 1.8" 디스크들을 이용한다. SPAN 또한 데이터 이동이 존재하지 않으며 PDC, COR 및 COMBINED는 각각 PCD기법, 연관 데이터 집중 기법 그리고 혼합 기법을 적용한다. PDC에 대해서는 매 5분마다 파일 이동이 일어나도록 설정하였다. 다른 모든 방법들의 에너지 절감량과 평균 요청 응답 시간은 SINGLE을 기준으로 하여 비교하였는데 SINGLE은 기본 이동성 기기의 디스크 저장 장치 시스템을 나타내기 때문이며 에너지 에 대해서는 상한을, 응답 시간에 대해서는 하한을 제시한다. 모든 방법들은 이전에 언급한 것처럼 임계값 기반의 전력 제어 정책을 적용한다. 그리고 파일들은 디스크 수에 관계없이 수행 초기에 임의로 균등하게 디스크(들)상에 배치되는 것으로 하였다.

시뮬레이션 결과. 그림 2는 PMP 트레이스에 대해서 5 가지 방법에 의한 에너지 소모값을 보여 준다. 예상한 것처럼, SINGLE은 7874J의 가장 높은 에너지 소모를 보이고 있다. SPAN의 경우, 하나의 2.5" 디스크를 2개의 1.8" 디스크로 단순히 바꿈으로써, 6155J을 소모하면서 SINGLE에 비해 22%의 에너지 절감을 보이고 있다.

표 4 실험에 사용된 에너지 절감 방법과 설명

방법 이름	설명
SINGLE	단일 2.5" 디스크, 파일 이동 없음
SPAN	2개의 1.8" 디스크들, 파일 이동 없음
PDC	2개의 1.8" 디스크들, PDC 기법 적용
COR	2개의 1.8" 디스크들, COR 기법 적용
COMBINED	2개의 1.8" 디스크들, COMBINED 기법 적용

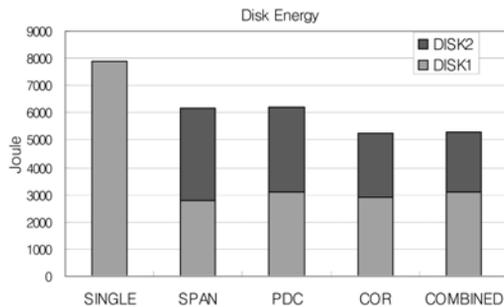


그림 2 에너지 소모값 (PMP 시나리오)

PDC는 이 시나리오에 대해 충분히 긴 유휴 시간을 얻지 못하고 있어서 에너지 절감량이 SPAN보다도 조금 작게 나타났다. COR와 COMBINED는 SPAN보다도 약 15%의 에너지를 더 절감하는 것으로 나타났으며 SINGLE과 비교해서는 34%의 에너지 절감을 얻는 것으로 나타났다. 이러한 결과들은 2가지 관찰을 이끌어낸다. 첫째, 단일 고전력 디스크를 2개의 작은 디스크로 교체하는 것 그 자체로도 에너지 절감에 이득을 가져오며, 둘째, 데이터 배치에 따른 I/O집중은 추가적으로 상당한 에너지 절감을 가능하게 하고 있다.

이러한 에너지 절감 효과는 각 디스크들이 각 상태에서 보낸 시간들을 분석함으로써 확인될 수 있다. 그림 3은 각 디스크가 각 상태에서 에너지 절감 방법별로 PMP 시나리오에서 얼마나 머물렀는가를 나타내고 있다. SINGLE은 좀처럼 더 낮은 전력 상태로 진입하기 위해 충분한 유휴 시간을 얻지 못하는 것을 볼 수 있다. SPAN의 경우는 I/O가 2개의 디스크로 분산되면서 각 디스크가 약 30% 및 18%의 휴면 시간을 가진다. PDC의 경우는 파일 이동에도 불구하고 I/O 부하를 하나의 디스크로 제대로 집중을 하지 못하고 있어서 결과적으로 지속적으로 저전력 상태로 디스크를 유지시킬 만큼의 유휴 시간을 충분히 갖지 못하고 있다. COR는 연관 데이터 기반으로 파일 이동을 수행함으로써 각 디스크에서 36%와 50%의 휴면 시간을 얻고 있다. 이것은 PMP 시나리오에서는 파일의 access 수보다는 연관성이 중요하게 에너지 절감과 관계된다는 것을 보여주는 것으로 판단할 수 있다. COMBINED는 COR와 유사한 휴면 시간을 가지는 것으로 나타났다.

표 5는 요청에 응답하는데 걸린 시간인 평균 응답 시간을 보여주고 있는데 각 방법의 성능 측정의 지표로 사용된다. 2.5"와 1.8" 디스크 간 원래의 성능 차이로 인해 SPAN은 SINGLE보다 2배의 응답 시간을 보이고 있다. PDC는 SINGLE에 비해 무려 2.6배나 긴 응답 시간을 보이고 있는데 이것은 많은 수의 파일 이동으로 인한 과잉부하 때문이다. PDC보다 더 적은 파일 이동

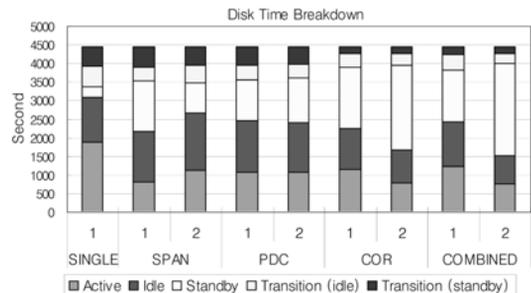


그림 3 디스크 시간 분포 (PMP 시나리오)

표 5 평균 요청 응답 시간 (PMP 시나리오)

방법	평균 요청 응답 시간 (s)
SINGLE	0.133
SPAN	0.279
PDC	0.342
COR	0.167
COMBINED	0.166

을 가지는 COR와 COMBINED는 SINGLE에 비해 단지 30ms정도의 지연만 더 있는 것으로 나타나고 있다.

그림 4, 5와 표 6은 각각 PDA 시나리오에서 수행한 트레이스에 대해 에너지 소모값, 각 상태에서 보낸 시간 그리고 평균 요청 응답 시간을 보여주고 있다. 이 결과들은 전반적으로 PMP의 결과들과 유사하게 보이지만 몇 가지 언급할 만 사항은 다음과 같다.

첫째, 그림 4에서 PDC는 이 시나리오에서는 SPAN이나 심지어 COR보다도 더 좋은 에너지 절감율을 나타내고 있다. 그러나 표 6에서 보는 것처럼 PDC는 엄청

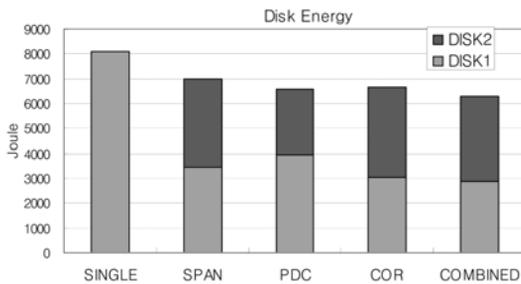


그림 4 에너지 소모값 (PDA 시나리오)

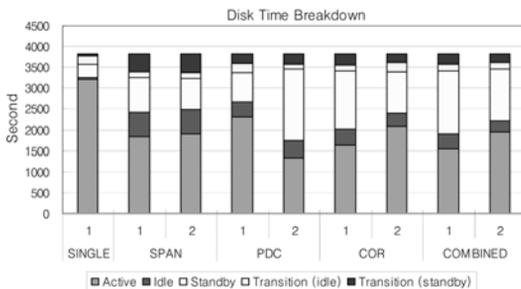


그림 5 디스크 시간 분포 (PDA 시나리오)

표 6 평균 요청 응답 시간 (PDA 시나리오)

방법	평균 요청 응답 시간 (s)
SINGLE	0.047
SPAN	0.123
PDC	1.176
COR	0.112
COMBINED	0.104

나게 증가한 파일 이동수로 인한 I/O 과잉부하 때문에 COR보다 무려 10배 이상의 평균 요청 응답 시간을 가진다. COMBINED에 대해서도 비슷한 성능 차이를 보인다. (실제로 PDC는 80138번, COR는 10491번 그리고 COMBINED는 9192번의 파일 이동 횟수를 기록하였다.) 이 결과는 에너지 측면에서 PDA 워크로드가 연관 데이터 기반의 알고리즘보다는 파일 접근 횟수 기반의 알고리즘에 더 유리하다는 것을 보인다. 그러나 한편으로는 여전히 파일 이동에 따른 과잉부하가 전체 디스크 시스템의 성능에 영향을 크게 주고 있음을 알 수 있다.

둘째, PMP 시나리오와는 달리 COMBINED가 확연히 가장 좋은 에너지 측면에서 가장 좋게 나타나고 있으며 또한 평균 요청 응답 시간에서도 SINGLE을 제외하고 가장 좋게 나타나고 있다. COMBINED는 약 SINGLE에 비해 23%의 에너지를 절감하고 있으며 60ms의 적절한 응답 시간 지연을 보인다. 이것은 바로 혼합 기법이 위에서 언급한 PDC기법의 장점을 수용함으로써 많은 에너지 절감을 얻으면서도 연관 데이터 집중 기법에 의한 파일 이동 횟수 감소에 따른 응답 시간 개선의 결과로 나타난 것으로 분석된다. 결과적으로, COMBINED는 PDC에 비해서 4.4%의 에너지를 절감하며 11배 이상의 응답 시간 개선을 얻고 있다.

5. 결론

플래시 메모리의 상승세에도 불구하고 하드 디스크 드라이브는 용량과 가격으로 인해 향후 몇 년간은 가장 중요한 보조 저장 장치로서의 역할을 유지할 것으로 예측된다. 또한 이에 따른 기술도 더욱 진보할 것이며 조만간 다중 디스크를 채택한 이동성 기기도 시장에 나타날 것으로 기대한다.

본 논문에서는 다중의 작은 디스크들을 하나의 큰 디스크 대신에 사용함으로써 이동성 컴퓨팅 시스템에서의 에너지 절감할 수 있는 기법을 제안하였다. 이 논문의 기여는 첫째, 이동성 기기에서의 다중 디스크 사용에 의해 상당히 에너지를 절감하는 방법을 제안하였고 둘째, 이동성 플랫폼의 워크로드 특성을 분석하고 이에 기반하여 다중 디스크 간에 데이터를 배치하고 이동하여 에너지 절감을 추가적으로 얻는 기법을 새로이 제안하였다는 것이다. 디스크 트레이스 기반의 실험을 통해서 제안한 기법을 단일 2.5" 디스크에 적용한 경우에 비해서 한쌍의 1.8" 디스크에 적용한 경우, 충분히 작은 I/O 응답 감소 증가만을 가지면서도 최고 34%에 이르는 에너지 절감을 얻을 수 있었다. 또한, 기존의 데이터 집중 기법에 대해서도 최고 14.8%이상의 에너지 절감과 최고 11배 이상의 I/O 응답 시간 개선을 얻을 수 있었다.

향후 연구로는 더 많은 디스크를 도입함으로써 이동

성 저장 장치 시스템에서의 신뢰성, 에너지 및 성능을 더 깊이, 다각도로 제안된 알고리즘을 확장하고자 한다. 또한, 이동성 컴퓨터를 위한 저전력 이종(heterogeneous) 저장장치 및 저장 기법에 대해서도 연구를 수행할 예정이다.

참고 문헌

- [1] L. D. Paulson, "Will hard drives finally stop shrinking?," *IEEE Computer*, vol. 38, no. 5, pp.14-16, 2005.
- [2] E. V. Carrera, E. Pinheiro, and R. Bianchini, "Conserving disk energy in network servers," in *Proc. of the 17th International Conference on Supercomputing*, June 2003.
- [3] A. Papathanasiou and M. Scott, "Power-efficient server-class performance from arrays of laptop disks," Technical Report 837, Department of Computer Science, University of Rochester, May 2004.
- [4] S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke, "DRPM: Dynamic speed control for power management in server class disks," in *Proc. of the International Symposium on Computer Architecture*, June 2003.
- [5] E. Pinheiro and R. Bianchini, "Energy conservation techniques for disk array-based servers," in *Proc. of the 18th International Conference on Supercomputing (ICS'04)*, June 2004.
- [6] Q. Zhu, Z. Chen, L. Tan, Y. Zhou, K. Keeton, and J. Wilkes, "Hibernator: helping disk arrays sleep through the winter," in *Proc. of the 20th ACM Symposium on Operating Systems Principles*, Oct. 2005.
- [7] Hitachi GST, Travelstar 80GN. http://www.hitachigst.com/tech/techlib.nsf/products/Travelstar_80GN.
- [8] Toshiba, MK4004GAH. <http://www3.toshiba.co.jp/storage/english/spec/hdd/mk4004gs.htm>.
- [9] Seagate, ST1.3. http://www.seagate.com/products/consumer_electronics/st1series.html.
- [10] Hitachi GST, Microdrive 3K8. http://www.hitachigst.com/tech/techlib.nsf/products/Microdrive_3K8.
- [11] Toshiba, MK4001MTD. <http://www3.toshiba.co.jp/storage/english/spec/hdd/mk4001.htm>.
- [12] F. Chen, S. Jiang, and X. Zhang, "SmartSaver: turning flash drive into a disk energy saver for mobile computers," in *Proc. of 11th ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED'06)*, Tegernsee, Germany, October 4-6, 2006.
- [13] T. M. Kroeger and D. D. E. Long, "The case for efficient file access pattern modeling," in *Proc. of the Seventh Workshop on Hot Topics in Operating Systems*, March 1999.



김 영 진

1997년 서울대학교 전기공학부 학사. 1999년 서울대학교 전기공학부 석사. 1999년~2003년 ETRI 연구원. 2003년~현재 서울대학교 전기·컴퓨터공학부 박사 과정 재학중. 관심분야는 임베디드 시스템 및 소프트웨어, 저전력 기법, 성능 및 전

력 분석 도구



권 권 택

2004년 서울대학교 컴퓨터공학부 학사
2006년 서울대학교 컴퓨터공학부 석사
2006년~현재 삼성종합기술원 Computing Technology Lab 연구원. 관심분야는 임베디드 시스템, 시스템 소프트웨어, 컴퓨터 아키텍처, 저전력 기법



김 지 홍

1986년 서울대학교 계산통계학과 학사
1988년 University of Washington 컴퓨터과학과 석사. 1995년 University of Washington 컴퓨터과학 및 공학과 박사. 1995년~1997년 미국 Texas Instruments 선임연구원. 1997년~현재 서울대학교 전기·컴퓨터공학부 부교수. 관심분야는 임베디드 소프트웨어, 저전력 시스템, 멀티미디어 시스템, 컴퓨터 구조