

시뮬레이션 기법을 이용한 완전 비동기 디스크 어레이 성능 평가*

A Performance Evaluation of a Fully Asynchronous Disk Array Using Simulation Method

오유영**, 김성수**, 김재훈**

Yuyoung Oh, Sungsoo Kim, Jae-Hoon Kim

Abstract

As real-time processing of data with large storage space is required in the era of multimedia, disk arrays are generally used as storage subsystems which be able to provide improved I/O performance. To design the cost-effective disk array, it is important to develop performance models which evaluate the disk array performance. Both queueing theory and simulation are applicable as the method of performance evaluation through queueing modeling. But there is a limit to the analytical method using queueing theory due to the characteristics of disk array requests being serviced in the parallel and concurrent manner. So in this paper we evaluate the disk array performance using simulation method which abstract disk array systems in the low level. Performance results were evaluated through simulation, so that mean response time, mean queueing delay, mean service time, mean queue length for disk array requests and utilization, throughput for disk array systems, can be utilized for capacity planning in the phase of disk array design.

* 본 연구는 1999년도 정보통신부 정보통신 우수시범학교 지원사업에 의한 결과임.

** 아주대학교 정보통신대학 정보및컴퓨터공학부

1. 서론

최근 수십 년 사이에 마이크로 프로세서의 성능이 급속도로 향상되었고 다중 프로세서 구조에 대한 연구가 진전됨으로서 멀티미디어 서비스를 제공하기 위한 컴퓨터 시스템의 계산 처리 능력(processing power)은 상상할 수 없을 정도로 강력해졌다. 반면에 입출력(I/O) 성능은 계산 처리 능력의 발전 속도만큼 크게 향상되지 못함으로써 전체적인 컴퓨터 시스템의 성능은 심한 불균형 상태에 놓이게 되었다[1, 2, 3, 4]. 이에 초고속 정보 통신망에서 대용량 데이터의 실시간 처리를 요구하는 멀티미디어 시대의 도래와 함께 계산 처리 능력에 견줄 수 있도록 입출력 성능을 개선시키기 위한 연구가 활발히 진행 중이다. 그 중에서도 두드러진 연구 성과는 고성능(high performance)과 고신뢰도(high reliability)를 동시에 제공할 수 있는 저장 서브시스템(storage subsystem)으로 디스크 어레이(disk array) 시스템을 사용하는 것이다. 특히, 데이터 스트라이핑(data striping)에 의한 병렬(parallel) 및 병행(concurrent) 접근에 의해서 고성능의 입출력 서비스와 미러링(mirroring) 및 패리티(parity) 정보를 사용하여 결함 허용(fault-tolerance)을 부여함으로써 고신뢰도를 제공할 수 있는 디스크 어레이의 한 유형인 RAID(Redundant Array of Inexpensive Disks)[4, 5, 6, 7]에 관한 연구가 많은 성과를 획득함으로써 근래에 상용화된 RAID 관련 제품들이 출시되고 있다.

디스크 어레이에 관한 연구가 많이 수행되었지만, 비용 효율적인 디스크 어레이 설계 및 용량 산정(capacity planning)을 위한 디스크 어레이 시스템에 대한 성능을 평가할 수 있는 모델링에 관한 연구는 미진한 실정이다[8, 9]. 디스크 어레이의 성능 평가 모델링 방법으로 큐잉 모델(queueing model)을 사용할 경우에, 큐잉 이론(queueing theory)을 적용하여 모델을 수학적으로 해석하거나 시뮬레이션 기법(simulation method)을 이용해서 평균적인 성능 평가 매트릭(metric)을 산출할 수 있다. 큐잉 이론을 사용할 경우에는 저비용으로 신속하고 간단하게 성능을 분석할 수 있지만, 상위 수준에서 추상화된 디스크 어레이 시스템에 적용되므로 정확한 결과를 산

출하는 것이 어렵다. 반면에 시뮬레이션 기법을 이용할 경우에 고비용이 요구되지만 디스크 어레이 시스템을 구성하고 있는 디스크의 물리적인 특성을 반영할 수 있는 하위 수준의 추상화가 가능하다. 특히, 디스크 어레이 시스템의 디스크 어레이 요구(disk array request)에 대한 병렬 및 병행 서비스 처리의 특성상 큐잉 이론으로 디스크 어레이의 큐잉 모델에 대한 분석적인 성능 평가가 불가능하다. 디스크 어레이 요구의 병렬 및 병행 서비스 특성을 무시하고 큐잉 이론으로 해석될 수 있도록 단순화된 수학적 가정을 하더라도, 산출할 수 있는 성능 평가 매트릭의 수가 제한적이고 비현실적이기 때문에 디스크 어레이 시스템의 비용 효율적인 설계 및 용량 산정을 위해서 활용되기에는 부적합하다[8].

큐잉 이론의 분석적인 방법의 한계성을 인식하고, 본 논문에서는 낮은 수준에서 디스크 어레이 시스템을 추상화하고 디스크 어레이 요구가 다수의 디스크 요구(disk request)로 분리(fork)되어서 해당 디스크 서버에서 병렬 및 병행 서비스되고 다시 디스크 어레이 요구로 합체(join)되는 특성을 반영할 수 있도록 개발된 시뮬레이터를 사용하여 디스크 어레이 시스템의 성능을 평가한다. 시뮬레이션을 통해서 산출된 디스크 어레이 요구에 대한 평균 응답 시간(mean response time), 평균 큐잉 지연(mean queueing delay), 평균 서비스 시간(mean service time), 평균 큐 길이(mean queue length) 및 디스크 어레이 시스템의 이용률(utilization), 처리율(throughput) 등은 디스크 어레이 설계 시에 시스템의 용량 산정에 활용될 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구 결과를 살펴보고, 3장에서는 시뮬레이션하고자 하는 완전 비동기 디스크 어레이 시스템의 구조, 서비스 메카니즘, 그리고 단순화된 수학적 가정을 갖는 큐잉 모델을 설명하고, 4장에서는 완전 비동기 디스크 어레이 시스템의 시뮬레이션 모델을 제시하고, 5장에서는 시뮬레이션을 통해서 산출된 결과를 분석하고, 마지막으로 6장에서는 본 논문에 대한 결론을 맺고 향후 연구 방향 및 개발한 시뮬레이터의 활용 방안에 대해서 논한다.

2. 기존 연구

최근에 MOD(Movie On Demand)나 VOD(Video On Demand)와 같은 멀티미디어 시스템에서 대용량의 저장 공간을 제공하고 동시에 다수의 사용자를 서비스하기 위한 저장 서브시스템으로 단일의 대용량 디스크를 사용하기보다는 디스크 어레이를 사용하여 병렬 및 병행 처리를 지원함으로써 고성능의 입출력 서비스가 가능하게 되었다. 디스크 어레이는 다수의 디스크들로 구성되므로 평균 고장 발생 시간(MTTF: Mean Time To Failure)이 단일 디스크에 비해 짧기 때문에 디스크 어레이에 결합허용 특성을 부가한 고성능 및 고신뢰도를 보장하는 RAID 기법을 사용한 결합허용 저장 서브시스템 설계에 관한 많은 연구가 있었다[10]. 또한 디스크 어레이는 병렬 서비스를 제공할 수 있도록 데이터를 배치(placement)하는 알고리즘과 디스크의 결합허용을 위한 여분(redundancy)을 다루는 메카니즘에 주로 초점을 맞춰서 디스크 어레이 시스템의 고성능과 고신뢰도 사이에 연관된 Trade-off를 연구하였다[1, 3, 4, 5, 6, 7].

심각한 불균형 상태에 놓인 마이크로 프로세서와 입출력 시스템의 성능 차이를 극복하기 위해서 고성능 대용량 저장 서브시스템으로 디스크 어레이가 사용되는 추세를 고려해 볼 때, 디스크 어레이가 미래의 고성능 입출력 시스템에서 중요한 위치를 차지한다는 사실은 의심의 여지가 없다. 그러므로 비용 효율적인 디스크 어레이 시스템의 설계 및 용량 산정을 위한 체계적인 성능 평가 모델의 개발은 중요하다. 성능 평가 모델은 분석적인 수학적 모델, 시뮬레이션 모델 그리고 측정 모델 등이 있다. 일반적으로 비용 측면에서 살펴보면 분석적인 수학적 모델을 이용하는 것이 가장 유리한 방법이다. 분석적인 수학적 모델로 산출된 결과가 정확하다면 문제가 없지만, 디스크 어레이 시스템에서 디스크 어레이 요구가 다수의 디스크 요구로 분리되어 병렬 및 병행 서비스되고 다시 디스크 어레이 요구로 합체되는 메카니즘을 고려할 경우에 디스크 어레이 큐잉 모델에 대해서 큐잉 이론을 적용하여 성능 모델을 해석하는 것이 현재로서는 불가능하다. 3개 이상의 서버로 구

성되는 시스템에서 일(job)이 포아송(Poisson) 분포로 도착해서 서버 수만큼의 태스크(task)로 분리되어 병렬로 지수(exponential) 분포로 서비스되고 다시 합체되어 서비스 받는다고 단순화된 수학적 가정을 하더라도, 이 큐잉 모델에 대해서 근접된(approximate) 결과를 얻을 수밖에 없다. 그렇다고 프로토타입(prototype)을 만들어서 실제로 성능을 측정한다는 것은 고비용이 들고 유동적이지 못한 단점이 있다[7].

디스크 어레이 시스템을 일반적인 큐잉 시스템으로 간주하고 큐잉 모델링을 통해서 큐잉 이론을 적용하여 성능 평가를 수행한 연구에 대해서 살펴보면, [11]에서는 데이터 스트라이핑을 하지 않은 n 개의 독립적인 디스크로 구성된 디스크 어레이와 데이터 스트라이핑을 적용한 n 개의 디스크가 디스크 어레이 요구를 서비스 시에 개별 디스크 요구 사이에 동기화(synchronization)되어서 서비스가 이루어진다고 가정한 후에, 이 두가지 경우의 디스크 어레이에 대해서 성능 비교를 했다. 여기에서 n 개의 디스크로 구성된 디스크 어레이를 하나의 논리적인 디스크로 간주하고 $M/G/1$ 큐잉 이론을 적용해서 성능을 평가했다. 그리고 [12]에서는 비동기(asynchronous) 디스크 어레이에 대해서 디스크 어레이 요구에 대한 최소 응답 시간을 산출하는 식을 유도했다. 하지만 디스크 어레이 요구가 디스크 요구로 분리되어서 병렬 및 병행 서비스 받는다는 사실을 고려하지 않았다. [5]에서는 RAID의 최대 처리율에 대한 수학적 식을 유도했다. 디스크 어레이 시스템에 대한 성능 모델이 큐잉 모델링을 통해서 큐잉 이론으로 해석된 식이 의미를 갖기 위해서는, 디스크 어레이 요구가 디스크 요구로 분리되어서 병렬 및 병행 서비스 받고 다시 디스크 어레이 요구로 합체된다는 사실이 모델링에 포함되어야만 의미가 있다. 하나의 일(job)이 다수의 태스크(task)로 분리되어 서비스 받은 후에 다시 하나의 일로 합체되는 서비스 메카니즘을 주로 병렬 컴퓨팅과 다중 프로그래밍 분야에서 큐잉 이론으로 해석하고자 했었다[13]. 하지만 서버의 수가 2개일 경우에만 정확한 Closed 형태의 성능 분석 식을 유도할 수 있지만 아직까지 3개 이상일 경우에는 정확한 결과를 갖는 식을 유도할 수 없다[14]. 근

접한 결과를 산출하기 위해 서비스 시간에 대해서 지수(exponential) 분포를 가정하고 있지만 디스크 어레이에서 서비스 시간이 지수 분포를 갖는다고 단정하는 것이 비현실적이므로, 당분간은 큐잉 이론을 통한 디스크 어레이 시스템에 대한 정확한 성능 평가는 사실상 불가능한 실정이다.

디스크 어레이 시스템의 전체적인 성능 평가에 대한 연구 외에도 디스크 어레이의 성능 향상을 위해서 중요한 시스템 구성체인 캐쉬(cache), 디스크, 버스(bus), 그리고 인터페이스(interface)에 대한 연구가 있었다. [15]에서는 캐쉬가 계층적으로 호스트 컴퓨터의 캐쉬, 디스크 어레이 컨트롤러의 캐쉬, 그리고 디스크 버퍼로 구성될 때 성능이 향상된다는 결과를 얻었고, [16]에서는 한 디스크 드라이브의 성능 모델 개발 시에 탐색 시간(seek time)에 대해서 지수 분포를 적용하는 것이 비현실적이고 분석적인 수학적 모델에 의한 성능 평가가 부적합하기 때문에 시물레이션 기법을 이용해야 한다고 결론을 내렸고, [17]에서는 디스크 어레이를 구성하는 디스크 수가 많아질 경우에 통합된 디스크들의 데이터 전송률이 SCSI(Small Computers System Interface) 버스의 대역폭(bandwidth)보다 클 경우도 고려해야 하고, SCSI 버스 프로토콜이 우선 순위 제어(priority-arbitrated) 방식이기 때문에 우선 순위가 낮은 경우에는 서비스를 받지 못할 확률이 커지므로 라운드 로빈(round robin) 혹은 타임 슬롯(time slot)을 할당하는 프로토콜을 사용하는 것이 바람직하다고 말한다.

3. 완전 비동기 디스크 어레이 시스템

본 장에서는 본 논문의 대상이 되는 일반적인 디스크 어레이 시스템의 구조, 완전 비동기 디스크 어레이 시스템의 특징을 설명한다. 그리고 완전 비동기 디스크 어레이 시스템의 서비스 메카니즘, 디스크 어레이 요구의 병렬 및 병행 서비스의 특성을 반영한 큐잉 모델을 살펴본다.

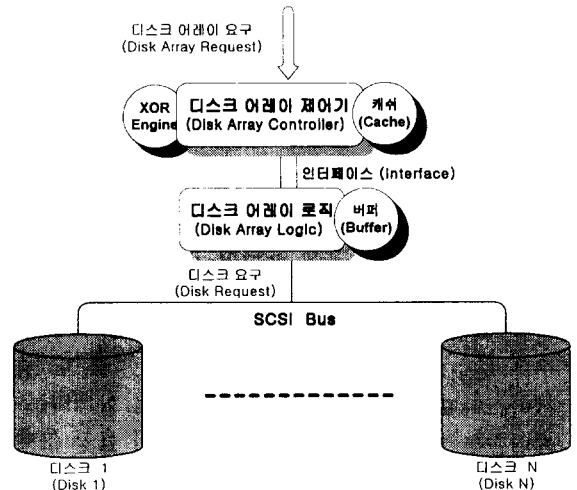
3.1 디스크 어레이 시스템 구조

본 논문에서 고려하고 있는 디스크 어레이 시스

템은 그림 1과 같이 N개의 동일한 디스크, 디스크 어레이 제어기, 디스크 어레이 로직, SCSI 버스, 그리고 인터페이스로 구성된다. 디스크는 고정된 대역폭을 갖는 SCSI 버스와 인터페이스에 의해서 디스크 어레이 제어기에 연결된다. 디스크 어레이 제어기는 선인출한 데이터를 저장하거나 나중에 요청된 디스크 어레이 요구에 대한 응답 시간을 줄임으로써 성능을 향상시키기 위해서 캐쉬로 사용되는 지역 메모리를 가지고, 고신뢰도를 보장하기 위해서 결함허용을 위한 패리티 정보를 사용할 경우 디스크 결함으로 인해서 액세스가 불가능한 데이터를 생성하는 XOR Engine을 갖는다. 디스크 어레이 로직은 개별적인 디스크들이 논리적으로 디스크 어레이가 될 수 있도록 하는 소프트웨어 모듈로서 각 디스크와 디스크 제어기에 존재한다. 디스크 어레이 버퍼는 각 디스크에서 디스크 요구에 대한 데이터를 인터페이스를 통해서 디스크 제어기로 보내기 위한 완충 역할을 수행한다.

3.2 완전 비동기 디스크 어레이 시스템

다수의 디스크를 하드웨어 혹은 소프트웨어에 의해서 어떻게 하나의 디스크 어레이로 만들 것인가에 의해서 디스크 어레이 시스템을 다음과 같이 4가지로 분류할 수 있다[17].



<그림 1> 디스크 어레이 시스템의 구조

● **완전 동기 디스크 어레이(Fully Synchronous Disk Array):** 모든 디스크의 헤드 탐색과 회전이 하드웨어에 의해서 동기화가 맞춰진다. 이 때에 인터페이스 대역폭이 충분히 크다면 단일 디스크의 전송 대역폭보다 디스크 어레이를 구성하는 디스크 수만큼 확장된 대역폭을 갖지만, 동일한 탐색 시간 및 회전 지연(rotational latency)을 초래한다.

● **부분적 동기 디스크 어레이(Partially Synchronous Disk Array):** 모든 디스크의 헤드 탐색은 하드웨어에 의해서 동기화가 맞춰지지만, 회전은 각 디스크에서 독립적으로 수행된다. 데이터는 모든 디스크의 동일한 물리적 위치에 스트라이핑된다.

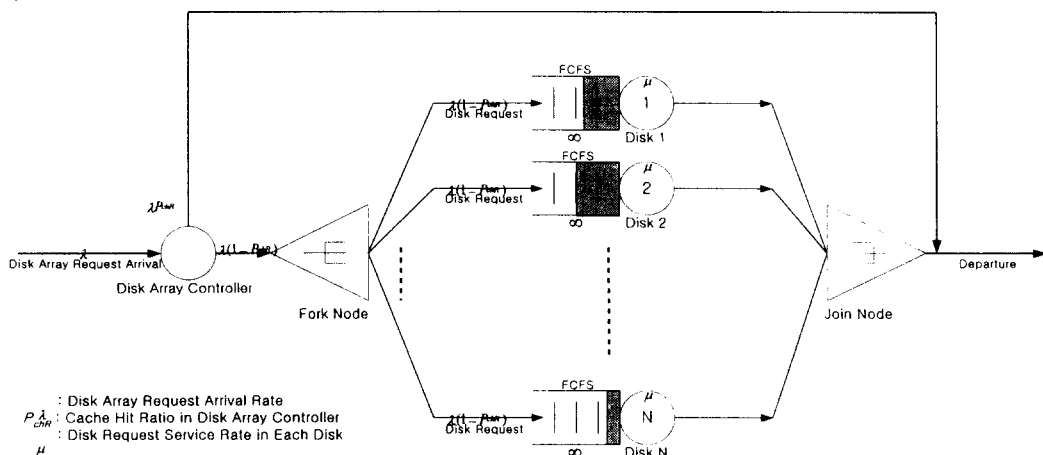
● **완전 비동기 디스크 어레이(Fully Asynchronous Disk Array):** 각 디스크의 헤드 탐색 및 회전이 독립적으로 수행된다. 모든 디스크가 한 디스크 어레이 요구에 대해서 병렬적으로 서비스에 참여하는 동기 디스크 어레이와 다르게 디스크 어레이 요구들에 대해서 데이터를 갖는 디스크들이 병렬 및 병행적으로 서비스한다.

● **완전 비결집 디스크 어레이(Fully Decoupled Disk Array):** 모든 디스크가 완전히 독립적으로 동작하지만 데이터가 스트라이핑되어 저장되지 않는다. 일반적으로 데이터를 스트라이핑하는 디스크 어레이와 비교를 목적으로 모델링되는 디스크 어레이이다. 특히, 디스크 어레이 요구가 동일한 비율로 각 디스크에 의해서 서비스 받는다는 비현실적인 가정을 한다.

동기 디스크 어레이의 성능 평가 모델은 $M/G/1$ 큐잉 모델링을 통해서 큐잉 이론을 적용하여 수학적으로 정확한 성능 평가 결과를 얻을 수 있지만, 완전 비동기 디스크 어레이는 디스크 어레이 요구의 병렬 및 병행 서비스 특성으로 인해서 분석적인 방법으로는 성능을 평가하기가 어렵다. 그러므로 본 논문에서 시뮬레이션 대상으로 삼고 있는 디스크 어레이는 모든 디스크를 하드웨어적으로 동기화할 필요가 없는 완전 비동기 디스크 어레이이다.

3.3 완전 비동기 디스크 어레이 시스템의 큐잉 모델

<그림 2>는 완전 비동기 디스크 어레이 시스템에서 디스크 어레이 요구가 포아송 분포로 도착, 하나의 디스크 어레이 요구에 대해서 모든 디스크로 디스크 요구가 분리되어서 지수 분포에 의해서 서비스, 각 디스크 서버의 큐 길이는 무한대, 그리고 서비스 정책이 FCFS(First Come First Service)라고 가정할 경우의 큐잉 모델($n M/M/1$ with fork/join)을 보여준다. 한 디스크 어레이 요구가 Fork Node에서 N 개의 디스크 요구로 분리되어서 병렬 및 병행적으로 서비스 받은 후에 마지막으로 서비스 받는 디스크 요구의 종료시점에 Join Node에서 다시 디스크 어레이 요구로 합체되는 특성으로 인해서 큐잉 이론으로는 정확한 성능 평가가 불가능하다.



<그림 2> 완전 비동기 디스크 어레이 시스템의 큐잉 모델($n M/M/1$ with fork/join)

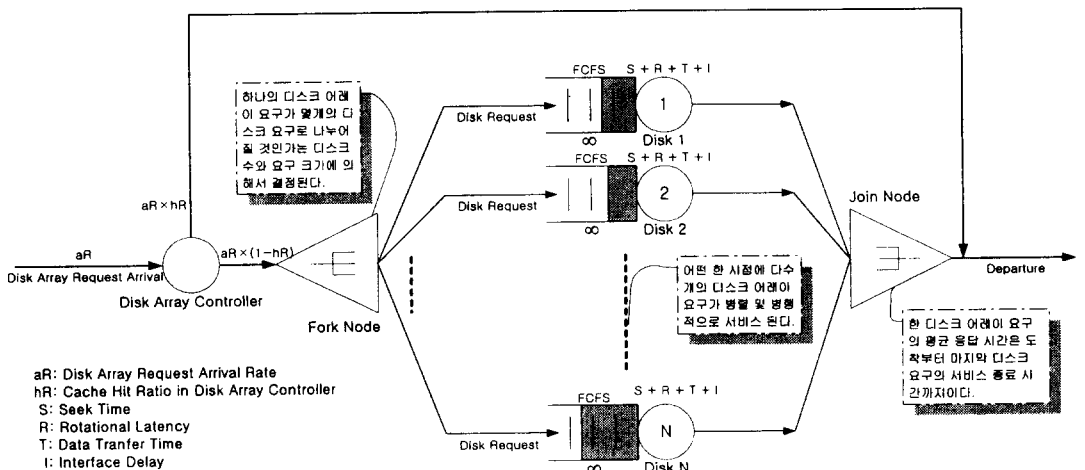
4. 시물레이션 모델

완전 비동기 디스크 어레이에 대해서 단순화된 수학적 가정을 갖는 큐잉 시스템조차도 큐잉 이론을 적용하는 분석적인 방법을 사용할 수 없기 때문에 시물레이션 기법을 이용하여 성능 평가를 수행한다. 본 장에서는 단순화된 수학적 가정을 배제한 완전 비동기 디스크 어레이 큐잉 시스템에 대한 시물레이션 모델을 제시하고 완전 비동기 디스크 어레이 시물레이터를 구성하는 모듈들을 살펴본다. 그리고 디스크 서버에서 디스크 요구의 서비스 시간 계산 시에 어떻게 디스크의 물리적/기계적 특성이 반영되는지를 설명한다.

4.1 시물레이션 모델

3장에서 살펴본 완전 비동기 디스크 어레이 시스템에 대한 큐잉 모델은 단순화된 가정을 갖는다. 즉, 디스크 어레이 요구가 포아송 분포로 도착하고, 디스크 어레이 요구가 Fork Node에서 디스크 요구로 분리될 때 디스크 어레이를 구성하는 모든 디스크로 각각 하나의 디스크 요구가 할당되고, 각 디스크에서 서비스 시간이 지수 분포를 따른다는 비현실적인 가정을 했었다. 실제적인 완전 비동기 디스크 어레이 시스템에 대한 시물레이션을 위해서는 <그림

3>과 같이, 디스크 어레이 요구가 Fork Node에서 몇 개의 디스크 요구로 분리될 것인가는 디스크 수, 디스크의 특성 파라미터(가령, 실린더 수, 실린더 당 트랙 수, 트랙 당 섹터 수, 섹터 크기, 스트라이핑 유닛 크기 등), 디스크 어레이 요구 크기 등에 의해서 서비스에 참여하는 디스크 요구 수가 결정될 것이고, 또한 서비스가 시작되는 디스크도 이산 항등(discrete uniform) 분포에 의해서 선택되도록 해야 한다. 그리고 각 디스크 요구가 디스크 서버에서 서비스 받는 시간은 디스크 헤드가 원하는 데이터가 저장된 실린더의 트랙까지 이동하는 탐색 시간, RPM(Rotation Per Minute) 속도에 따른 디스크 회전 에 의해서 데이터 블록이 저장된 섹터를 찾는 데 걸리는 지연 시간, 데이터를 디스크 전송 속도로 디스크 어레이 로직의 일시적인 캐쉬로 사용되는 트랙 버퍼로 전송하는 시간, 그리고 인터페이스를 통해서 버퍼의 데이터를 디스크 어레이 제어기로 전송하는 시간을 합산한 것이다. 앞서 설명한 서비스 시간을 데이터 읽기(read)에 대한 디스크 어레이 요구의 시간을 순서적으로 설명한 것이다. 하지만 쓰기(write)에 대한 디스크 어레이 요구의 경우도 탐색 시간, 회전 지연, 인터페이스 지연, 그리고 데이터 전송 시간 순서로 서비스되는 것만 다르다고 볼 수 있다. 특히, 이 때에 디스크 어레이 제어기의 지역 메모리가 캐쉬로 사용되지 않고 임시적인 쓰기 데이터의 버퍼로



<그림 3> 완전 비동기 디스크 어레이 시스템의 시물레이션 모델

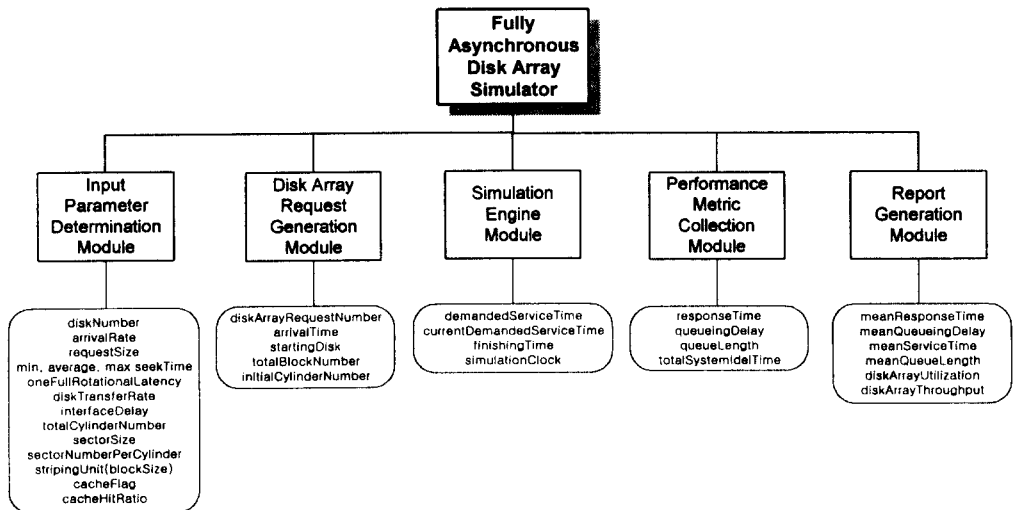
서 인터페이스를 통해서 디스크 어레이 로직의 버퍼로 전송되기 전에 다만 일시적으로 데이터를 저장하는 곳이다. 디스크 어레이 요구의 도착률, 각 디스크 서버의 큐 길이, 그리고 서비스 정책 등은 큐잉 모델과 같이 각각 포아송 분포에 의한 디스크 어레이 요구의 도착, 무한대 큐 길이, FCFS 정책에 의한 디스크 요구의 서비스로 시뮬레이션된다.

4.2 완전 비동기 디스크 어레이 시뮬레이터

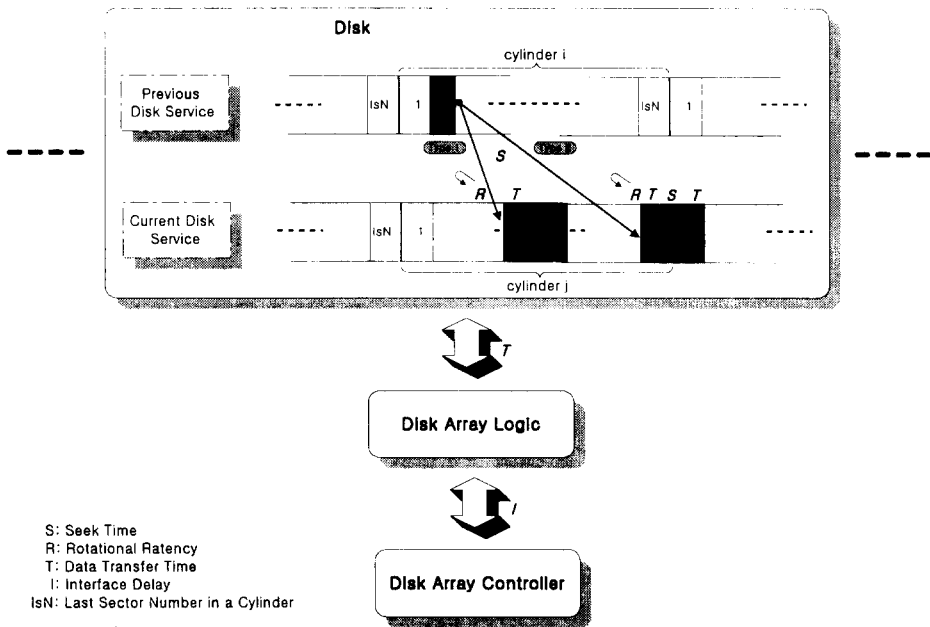
완전 비동기 디스크 어레이에 대한 시뮬레이터는 C 프로그래밍 언어[18]에 의해서 구현되었고, Event-driven 방식[19, 20]으로 시뮬레이션된다. 그림 4는 시뮬레이터를 구성하는 모듈 및 주요 관련 파라미터들을 보여준다. 완전 비동기 디스크 어레이 시스템은 Input Parameter Determination Module에 의해 입력되는 디스크 수, 디스크 어레이 요구 도착률, 디스크 어레이 요구 크기, 디스크 특성 파라미터, 인터페이스 전송 속도, 그리고 디스크 어레이 제어기의 캐쉬에 대한 적중률(hit ratio) 등에 의해서 시스템이 결정된다. 디스크 어레이 요구는 Disk Array Request Generation Module에 의해서 포아송 분포 도착 시간에 따라 발생되고, 생성된 디스크 어레이 요구는 유일하게 식별될 수 있도록 요구 번호를 배정받게 되고, 서비스 요청한 데이터에 대한 총 블록

수(스트라이핑 유닛 수), 서비스 시작 디스크, 각 디스크에서 서비스 시작 실린더 등에 대한 정보를 갖는다. Simulation Engine Module에 의해서 디스크 어레이 요구가 Fork Node에서 분리된 디스크 요구를 디스크 서버에서 병렬 및 병행적으로 서비스되는 것을 시뮬레이션하게 된다. 이 때에 디스크 서버의 디스크 요구는 물리적/기계적 수준을 고려하여 서비스 시간이 계산되고, Simulation Clock의 진행에 의해서 가장 먼저 서비스 받아야 할 디스크 요구는 currentDemandedServiceTime에 의해서 결정된다. 각 디스크 요구가 끝나는 시점에 Performance Metric Collection Module에 의해서 디스크 요구에 대한 응답 시간, 서비스 시간, 큐잉 지연, 큐 길이에 대한 정보가 수집된다. 마지막으로 Report Generation Module은 서비스 받은 모든 디스크 요구에 대해서 수집된 정보를 근거로 산출된 디스크 어레이 요구에 대한 평균 응답 시간, 평균 큐잉 지연, 평균 서비스 시간, 평균 큐 길이 및 디스크 어레이 시스템의 이용률과 처리율을 모니터나 파일로 출력한다.

디스크 서버에서 디스크 요구에 대한 서비스 시간을 계산할 때에, <그림 5>와 같이 디스크 헤드의 스케줄링은 이전 디스크 요구에 대한 서비스가 종료될 때에 디스크 헤드는 데이터 서비스가 끝난 실린더에 위치하고 현재 디스크 요구를 서비스하기 위해



<그림 4> 완전 비동기 디스크 어레이의 시뮬레이터 모듈



<그림 5> 디스크 요구에 대한 서비스 시나리오

서 이산 항등 분포에 의해서 선택된 새로운 실린더로 이동하게 된다. 이 디스크 스케줄링에 의해서 실린더 거리가 계산됨으로써 탐색 시간이 계산된다. 이 때에 디스크 요구 크기에 따라 두 가지 서비스 시나리오를 고려해야 한다. 디스크 서버에 배정된 디스크 요구 크기가 한 실린더 범위 안에서 서비스가 가능한 경우(Type 1)와 디스크 요구가 한 개 이상의 실린더에 걸쳐서 서비스되어야 하는 경우(Type 2)로 나눈다. 후자의 경우에 최초의 탐색 시간과 데이터의 서비스 중에 인접 실린더로의 탐색 시간도 고려해야 한다. 회전 지연 시간은 항등 분포에 의해서 한번 디스크 회전에 소요되는 시간 이내의 값이 된다. 이 때 한번의 회전 내에서 원하는 섹터를 찾는 데에 실패할 경우는 고려하지 않았다. 디스크 요구 크기에 따른 데이터 전송 시간 및 인터페이스 지연은 디스크 전송 속도와 인터페이스 대역폭에 의해서 계산된다.

5. 시물레이션을 통한 성능 평가

완전 비동기 디스크 어레이 시스템의 성능을 평가하기 위해서, 디스크 어레이는 동일한 디스크 (IBM 0661 SCSI 디스크)로 구성되고 모든 디스크는 표 1과 같은 디스크 특성 파라미터를 갖는다고 가정한다. 그리고 시물레이터에 입력되는 시스템 결정 파라미터는 표 2, 출력되는 성능 파라미터는 표 3과 같다. 특히, IBM 0661 디스크에 대한 탐색 시간 함수는 식 (1)과 같다[7]. 식(1)에서 x 는 탐색 거리(실린더 단위), a, b, c 는 한 실린더에 적용되는 상수 값 ($a = 0.4623, b = 0.0092, c = 2$)이다. 이 식에 의해서 최초 탐색 시간 및 인접 실린더간의 탐색 시간(최소 탐색 시간)이 포함되어 디스크 요구에 대한 탐색 시간이 계산된다. 본 장에서는 시물레이터의 Report Generation Module에 의해서 얻어진 결과, 즉 디스크 어레이 요구 도착률 및 크기, 디스크 어레이를 구성하는 디스크 수의 변화에 따른 디스크 어레이 요구의 평균적인 성능 메트릭을 분석함으로써 비용 효

올직한 완전 비동기 디스크 어레이 시스템의 설계 및 용량 산정에 활용되도록 한다.

$$\text{탐색시간}(\text{seekTime}) = \begin{cases} 0, & \text{if } x=0 \\ a(x-1)^{0.5} + b(x-1) + c, & \text{if } x>0 \end{cases} \quad (1)$$

5.1 평균 응답 시간

디스크 어레이 제어기에서 캐쉬를 사용할 경우와 캐쉬를 사용하지 않을 경우로 나누어 시뮬레이션 실험을 수행함으로써, 완전 비동기 디스크 어레이 시스템의 디스크 어레이 요구에 대한 평균 응답 시간을 획득했다. 캐쉬를 사용할 경우에 [15]에 의해서 LRU (Least Recently Used), RR(Random Replacement), SLRU(Segmented LRU) 등의 캐쉬 제어 정책을 사용한다고 가정할 때에 캐쉬 적중률(hR)에 대해서 캐쉬로 사용되는 지역 메모리의 크기는 결정된다. <그림 6>은 디스크 수(dN)의 변화에 따른 디스크 어레이 요구 도착률(aR), 디스크 어레이 크기(darS)의 조합에 대해서 산출된 디스크 어레이 요구에 대한 평균 응답 시간(mRT)을 보여준다. aR이 크면 큐잉 지연이 발생할 확률이 커지므로 mRT가 증가하지만 변화의 폭은 그렇게 크지 않다. 반면에 darS가 4KB보다 128KB일 때, dN이 10 이하일 경우에 변화의 폭이 크다. <그림 7>은 aR이 10/sec이고 darS가 4, 64, 128, 256KB일 때, dN이 작을 경우에는 디스크 어레이 서비스 시간 중에서 디스크 전송 시간과 인터페이스 지연이 차지하는 비율이 높기 때문에

mRT는 dN이 큰 경우보다 상대적으로 높다는 것을 보여준다. <그림 8>은 aR이 10/sec, darS가 64KB일 때에 캐쉬를 사용할 경우와 사용하지 않을 경우에 mRT를 보여준다. 캐쉬를 사용하지 않을 경우보다 캐쉬 적중률(hR)이 0.2, 0.2보다 0.4일 경우에 mRT는 감소한다.

<표 4> IBM 0661 SCSI 디스크의 특성 파라미터

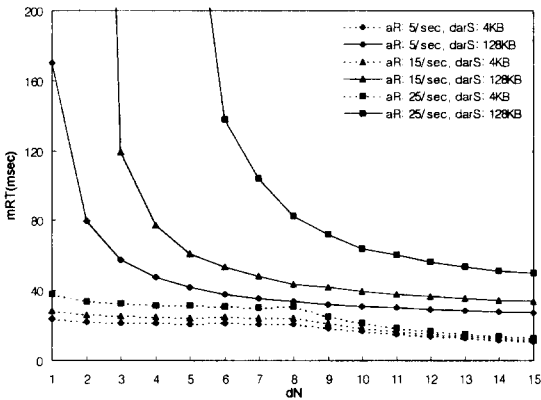
Cylinders per disk	949
Tracks per cylinder	14
Sectors per track	48
Bytes per sector	512
RPM	4320
Single cylinder seek time	2.0 msec
Average seek time	12.5 msec
Max seek time	25.0 msec
Max transfer rate	1.7 MB/sec

<표 3> 성능 평가에 사용된 출력 파라미터

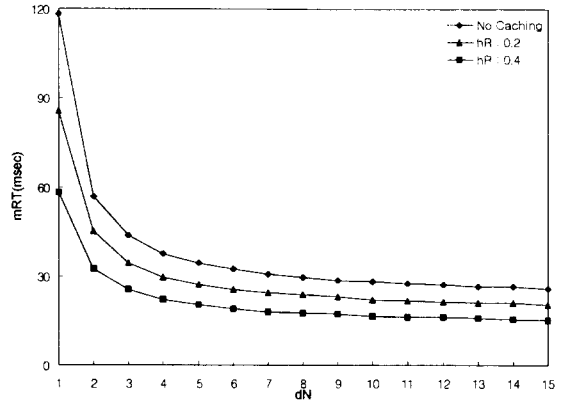
항 목	설 명
meanResponseTime(mRT)	평균 응답 시간(msec)
meanQueueingDelay(mQD)	평균 큐잉 지연(msec)
meanServiceTime(mST)	평균 서비스 시간(msec)
meanQueueLength(mQL)	평균 큐 길이(디스크 어레이 요구 수)
diskArrayUtilization(daU)	디스크 어레이 시스템의 이용률
diskArrayThroughput(daT)	디스크 어레이 시스템의 처리율 (KB/sec)

<표 2> 성능 평가에 사용된 입력 파라미터

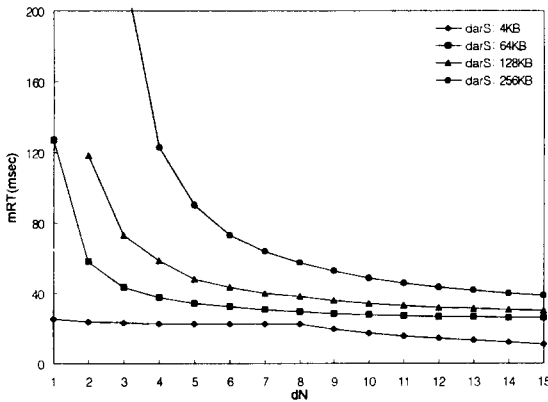
항 목	값	설 명
arrivalRate(aR)	5, 10, 15, 20, 25/sec	초당 도착하는 디스크 어레이 요구 수
diskArrayRequestSize(darS)	4, 64, 128, 256 KB	디스크 어레이 서비스 크기
diskNumber(dN)	1 ~ 15	디스크 어레이를 구성하는 디스크 수
cylinderNumber(cN)	949	디스크의 실린더 총 수
sectorNumberPerCylinder(sN)	686	실린더의 섹터 총 수
sectorSize(sS)	512 B	한 섹터의 크기
stripingUnitSize(sU)	512, 1024, 2048 B	스트라이핑 크기(한 블록)
min, average, maxSeekTime (msT, asT, xsT)	2.0, 12.5, 25.0 msec	최소, 평균, 최대 탐색 시간
oneFullRotationalLatency(rL)	13.9 msec	회전 지연 시간
diskTransferRate(dtR)	1.7 MB/sec	디스크 전송 속도
interfaceSpeed(iS)	10.0 MB/sec	인터페이스 전송 속도
cacheHitRatio(hR)	0.2, 0.4	캐쉬 적중률



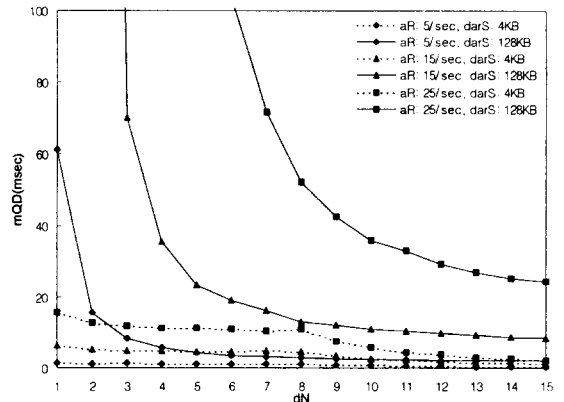
<그림 6> 디스크 어레이 요구에 대한 평균 응답 시간 (aR: 5, 15, 25/sec, darS: 4, 128KB, N: 변화, sU: 512B, No Caching)



<그림 8> 캐쉬 적중률 변화에 따른 디스크 어레이의 평균 응답 시간 (aR: 10/sec, darS: 64KB, sU: 512B, Caching)



<그림 7> 디스크 어레이 요구 크기의 변화에 따른 평균 응답 시간 (aR: 10/msec, darS: 4, 64, 128, 256KB, N: 변화, sU: 512B, No Caching)

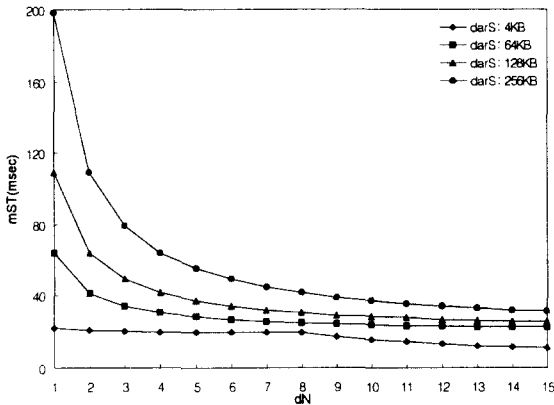


<그림 9> 디스크 어레이 요구에 대한 평균 큐잉 지연 (aR: 5, 15, 25/sec, darS: 4, 128KB, N: 변화, sU: 512B, No Caching)

5.2 평균 큐잉 지연 및 평균 서비스 시간

완전 비동기 디스크 어레이 시스템에서 디스크 어레이 요구에 대한 평균 큐잉 시간(mQD)은 한 개의 디스크 어레이 요구가 Fork Node에서 다수개의 디스크 요구로 분리되어서 해당 디스크 서버에서 다른 디스크 서버와 병렬 및 병행적으로 서비스 받기

때문에, 분리된 디스크 요구 중에서 가장 긴 큐잉 지연을 디스크 어레이 요구에 대한 큐잉 지연으로 간주하고 모든 디스크 어레이 요구에 대해서 평균값을 취한 것이다. 디스크 어레이 요구에 대한 평균 서비스 시간(mST)은 앞서 산출된 mRT에서 mQD를 뺀 값이다. <그림 9>와 <그림 10>은 각각 mQD와 mST에 대한 결과를 보여준다. mQD는 dN이 증가할수록 감소하고, aR의 변화에 대해서는 변

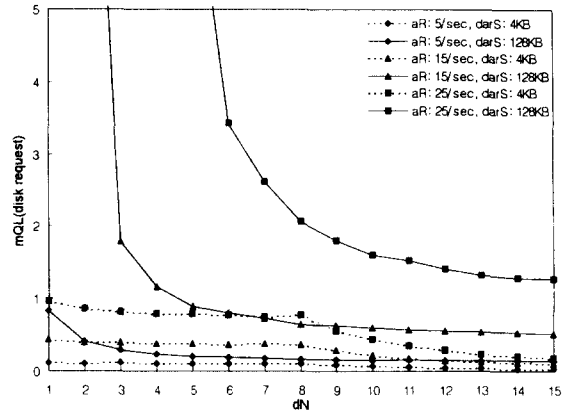


<그림 10> 디스크 어레이 요구 크기의 변화에 따른 평균 서비스 시간
(aR: 5/msec, darS: 4, 64, 128, 256KB, N: 변화, sU: 512B, No Caching)

화의 폭이 크지 않지만 darS의 변화에 대해서는 상당히 크게 변화하는 양상을 보인다. mST는 darS가 작을 경우에는 dN에 무관하게 비슷한 값을 보이지만, darS가 크고 dN이 작은 경우에는 증가한다. 디스크 요구에 대한 서비스 시간 중에서 데이터 전송과 인터페이스 지연이 차지하는 비율이 상대적으로 증가하기 때문이다.

5.3 평균 큐 길이

완전 비동기 디스크 어레이 시스템에 대한 시뮬레이션 모델에서 각 디스크 서버의 큐 길이는 무한대로 가정했다. 여기에서 큐 길이의 기본 단위는 디스크 요구이다. 그러므로 디스크 어레이 요구 중에서 캐쉬 적중률에 의해서 서비스 받는 것을 제외한 모든 디스크 어레이 요구는 Fork Node에서 디스크 요구로 분리되어 해당 디스크 서버에서 서비스를 받게 되는데, 만약 디스크 서버가 다른 디스크 요구를 서비스하고 있을 때에는 큐에 대기하게 된다. 여기에서 한 디스크 어레이 요구에 대한 큐 길이는 분리된 디스크 요구 중에서 가장 긴 큐 길이가 디스크 어레이 요구에 대한 큐 길이가 되기 때문에, 디스크 어레이 요구에 대한 평균 큐 길이(mQL)는 모든 디스크



<그림 11> 디스크 어레이 요구에 대한 평균 큐 길이
(aR: 5, 15, 25/sec, darS: 4, 128KB, N: 변화, sU: 512B, No Caching)

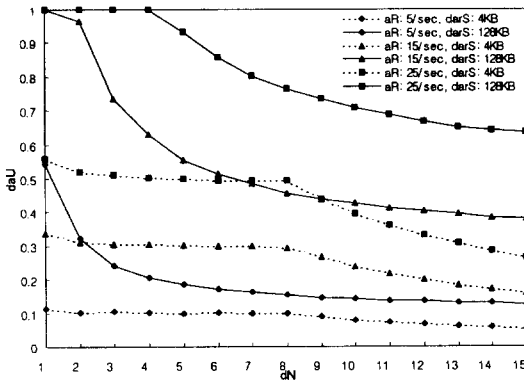
크 어레이 요구에 대해서 평균값을 취한 것이다. 시뮬레이터에서 한 디스크 요구에 대한 큐 길이는 디스크 요구가 디스크 서버에서 서비스가 종료되는 시점에 서버의 큐에 대기 중인 디스크 요구의 수를 확인함으로써 산출된다. 그림 11은 dN의 변화에 따른 aR, darS의 조합에 대해서 얻어진 디스크 어레이 요구에 대한 mQL를 보여준다. darS가 작을 경우에는 aR의 변화에 민감하지 않지만, darS가 크고 aR이 15, 25/sec일 경우에 N이 10이하에 대해서는 mQL이 길고 dN의 변화에 민감하게 변한다.

5.4 이용률 및 처리율

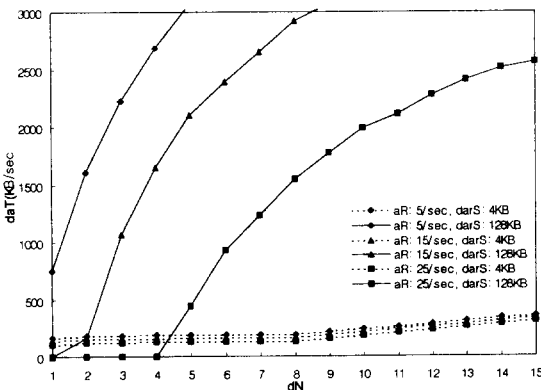
시뮬레이터에서 완전 비동기 디스크 어레이 시스템의 이용률(daU)은 어떤 디스크 서버도 서비스하지 않은(idle) 시간을 Performance Metric Collection Module에서 시뮬레이션이 종료될 때까지 total SystemIdleTime이라는 파라미터에 유지함으로써 시스템의 비이용률이 계산됨으로써 구해진다. 그림 12는 dN의 변화에 따른 aR, darS의 조합에 대해서 산출된 디스크 어레이 시스템의 daU를 보여준다. dN이 작을수록, aR과 darS가 클수록 daU는 높다. 또한 darS가 클수록 dN의 변화에 따른 daU의 변화

의 폭도 크다.

완전 비동기 디스크 어레이 시스템의 처리율 (daT)은 초당 얼마만큼의 데이터를 처리할 수 있는가를 의미한다. 그림 13은 dN의 변화에 따른 aR, darS의 조합에 대해서 산출된 완전 비동기 디스크 어레이 시스템의 daU를 보여준다. darS가 작으면 aR의 변화에 민감하지 않지만, darS가 크면 aR이 5, 15, 25/sec일 경우에 각각 N이 1, 2, 4부터 급격히 높아진다. 하지만 dN이 어느 정도 커지면 처리율은 한계에 다다르게 된다.



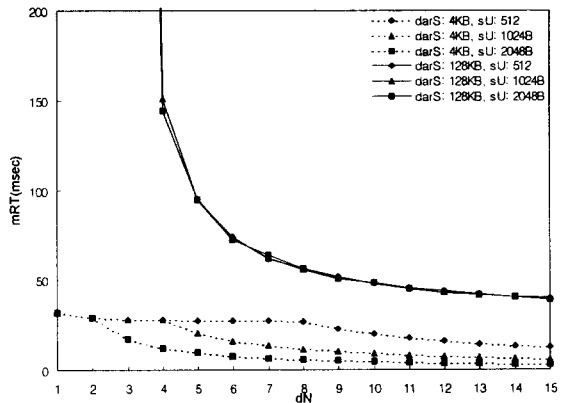
<그림 12> 완전 비동기 디스크 어레이시스템의 이용률 (aR: 5, 15, 25/sec, darS: 4, 128KB, N: 변화, sU: 512B, No Caching)



<그림 13> 완전 비동기 디스크 어레이 시스템의 처리율 (aR: 5, 15, 25/sec, darS: 4, 128KB, N: 변화, sU: 512B, No Caching)

5.5 스트라이핑 유닛의 변화에 따른 평균 응답 시간

지금까지는 스트라이핑 유닛(sU) 크기를 일반적으로 디스크 어레이 시스템에서 많이 사용되는 한 섹터의 크기인 512B로 고정시키고 성능을 평가했다. 여기서는 sU 크기의 변화(기본 단위는 512B이다)에 따른 디스크 어레이 요구의 mRT가 어떻게 변화되는지를 살펴본다. 그림 14는 dN의 변화에 따른 darS와 sU의 조합에 대해서 산출된 완전 비동기 디스크 어레이 시스템의 디스크 어레이 요구에 대한 mRT를 보여준다. darS가 4KB일 때에 dN이 증가함에 따라서, sU의 크기 변화에 불규칙적으로 서서히 mRT가 감소하지만, darS가 128KB일 때에 N이 10이하에서 sU의 크기에 상관없이 급격하게 mRT가 감소하다가 dN이 10보다 커지면 변화폭이 둔화되는 양상을 보인다.

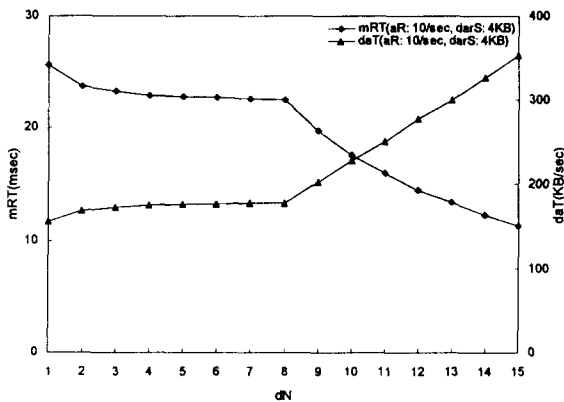


<그림 14> 스트라이핑 유닛의 크기 변화에 따른 평균 응답 시간 (sU: 512, 1024, 2048B, aR: 20/sec, darS: 4, 128KB, N: 변화, No Caching)

5.6 완전 비동기 디스크 어레이 시스템의 용량 산정

사용자가 원하는 성능을 발휘할 수 있는 완전 비동기 디스크 어레이 시스템을 비용 효율적인 측면에서 용량 산정하는 것은 중요하다. 디스크 어레이 요

구의 도착률과 크기에 대한 정보를 알고 있을 경우에 사용자가 원하는 디스크 어레이 요구에 대한 mRT와 디스크 어레이 시스템의 daT를 만족시킬 수 있는 dN을 결정함으로써 최소의 디스크 비용으로 사용자가 원하는 수준의 서비스를 제공할 수 있다. 그림 15에서 dN이 10일 때에 mRT와 daT가 균형 상태를 이룬다. 만약 dN이 10보다 크면, 디스크 어레이 요구에 대한 mRT는 감소하고 디스크 어레이 시스템의 daU가 증가하므로 전체적인 디스크 어레이 요구에 대한 서비스가 좋아진다. 그러나 사용자가 원하는 성능을 dN이 10보다 작은 값으로 만족시킬 수 있는 경우에 일부러 10이상인 dN을 사용할 필요는 없다. 그러므로 사용자가 원하는 성능에 대한 mRT와 daU에 대한 정보를 알 수 있다면, 최적의 디스크 수를 예측할 수 있다. <그림 15>에서 mRT와 daT의 상관 관계만을 다루었지만 그 밖의 성능 메트릭들을 조합함으로써 보다 광범위한 용량 산정도 가능하다.



<그림 15> 평균 응답 시간과 처리율에 의한 용량 산정 (aR: 10/sec, darS: 4KB, N: 변화, sU: 512B, No Caching)

6. 결론 및 향후 계획

일반적인 디스크 어레이 시스템의 성능 평가 방법으로 큐잉 이론을 이용한 분석적인 수학적 해석, 하위 수준에서 시스템을 추상화한 시뮬레이션, 그리고 실제 프로토타입을 만든 후에 성능을 측정하는 방법이 있다. 하지만 기존의 큐잉 모델링을 통해 큐잉 이론을 적용한 경우에 상위 수준의 단순화된 수학적 가정을 한 후에, 동기 디스크 어레이 시스템의 근접된 성능만을 평가할 수 있었다. 완전 비동기 디스크 어레이 시스템의 경우에 디스크 어레이 요구가 디스크 요구로 분리되어 디스크 서버에서 병렬 및 병행적으로 서비스 받은 후에 다시 합체되는 특성을 갖기 때문에 큐잉 이론으로 해석하는 것이 불가능하기 때문에, 분리된 디스크 요구가 병렬 및 병행적으로 서비스되는 것을 반영할 수 있는 시뮬레이션 기법을 이용해서 완전 비동기 디스크 어레이 시스템에 대한 성능을 평가했다. 본 논문에서는 시뮬레이션을 통해서 산출된 디스크 어레이 요구에 대한 평균 응답 시간, 평균 큐잉 지연, 평균 서비스 시간, 평균 큐 길이 및 디스크 어레이 시스템의 이용률, 처리율 등의 성능 메트릭을 획득함으로써 디스크 어레이 설계를 위한 시스템의 용량 산정에 활용될 수 있다.

본 논문에서 개발된 시뮬레이터가 실제로 완전 비동기 디스크 어레이 시스템의 설계 시에 보다 효과적으로 이용되기 위해서는 이질적인(heterogeneous) 특성을 갖는 개별 디스크의 집합으로 이루어진 디스크 어레이를 고려해야 하고, 디스크 어레이 요구의 도착 분포가 포아송이 아닌 실질적인 트레이스(trace)에 근거한 워크로드(workload)가 생성되어야 한다. 또한 상세한 물리적 수준의 시뮬레이션이 되어야 하고, 디스크 어레이 시스템의 중요한 구성체인 버스 및 인터페이스에 대한 프로토콜도 고려함으로써 디스크 어레이 요구의 충돌 현상을 시뮬레이션에 포함해야 하고, 마지막으로 보다 광범위한 용량 산정에 의해 비용 효율적으로 최적화된 완전 비동기 디스크 어레이 시스템의 설계를 위해서는 적절한 비용함수를 유도해야 한다.

참고문헌

- [1] G.R. Ganger, B.L. Worthington, R.Y. Hou and Y.N. Patt, "Disk Arrays: High- Performance, High-Reliability Storage Subsystems," *IEEE Computer*, pp. 30-36, March 1994.
- [2] 권택근, 연속 매체 저장 시스템에서 디스크 입출력 성능 향상 기법. 박사학위논문, 서울대학교 컴퓨터공학과, 1996. 2.
- [3] 오유영, 권원석, 김성수, 강창훈, "신뢰도와 임의 속도 탐색을 고려한 연속 매체 디스크 배치 기법," 1998년 한국정보과학회 춘계학술발표논문집, 제25권 제1호(A), pp. 45-47, 1998. 4.
- [4] 오유영, 김성수, "결함복구율을 고려한 연속 매체 디스크 배열의 신뢰도 분석," 1998년 한국정보과학회 추계학술발표논문집, 제25권 제2호(A), pp. 9-11, 1998. 10.
- [5] D. Patterson, G. Gibson and R. Katz, "A Case for Redundant Array of Inexpensive Disks(RAID)," in *Proc. of ACM SIGMOD'88*, June 1988, pp. 109-116.
- [6] P. Chen, E. Lee, G. Gibson and D. Patterson, "RAID: High-Performance, Reliable Secondary Storage," *ACM Computing Surveys*, pp. 145-186, June 1994.
- [7] G. Gibson, *Redundant Disk Arrays: Reliable, Parallel Secondary Storage*. PhD thesis, Univ. of California at Berkeley, December 1991.
- [8] E. K. Lee, *Performance Modeling and Analysis of Disk Arrays*. PHD thesis, Univ. of California at Berkeley, December 1993.
- [9] C-S Lee and T-M Parng, "Performance Modeling and Evaluation of a Two-Dimensional Disk Array System," *Journal of Parallel and Distributed Computing* 38, pp. 16-27, October 1996.
- [10] 오유영, 김성수, "임의 속도 탐색을 지원하는 연속 매체 저장 시스템," 1999년 한국정보과학회 춘계학술발표논문집, 제26권 제1호(A), pp. 193-195, 1999. 4.
- [11] M. Y. Kim, "Synchronized Disk Interleaving," *IEEE Transactions on Computers*, pp. 978-988, November 1986.
- [12] M. Y. Kim and A. N. Tantawi, "Asynchronous Disk Interleaving: Approximating Access Delays," *IEEE Transactions on Computers*, pp. 801-810, July 1991.
- [13] R. Nelson and A. N. Tantawi, "Approximate Analysis of Fork/join Synchronization in Parallel Queues," *IEEE Transactions on Computers*, pp. 739-743, June 1988.
- [14] C. Kim and A. K. Agrawala, "Analysis of the Fork-join Queue," *IEEE Transactions on Computers*, pp. 250-255, February 1989.
- [15] R. Karedla, J.S. Love and B.G. Wherry, "Caching Strategies to Improve Disk System Performance," *IEEE Computer*, pp. 38-46, March 1994.
- [16] C. Ruemmler and J. Wilkes, "An Introduction to Disk Drive Modeling," *IEEE Computer*, pp. 17 - 28, March 1994.
- [17] Christopher L. Elford and Daniel A. Reed, "Technology Trends and Disk Array Performance," *Journal of Parallel and Distributed Computing* 46, pp. 136 - 147, November 1997.
- [18] A. Kelley and I. Pohl, *A Book on C*. The Benjamin/Cummings Publishing Company, Inc., 1989.
- [19] K. Kobayashi, *Modeling and Analysis*. Addison-Wesley, Reading Mass., 1978.
- [20] J. Banks, J.S. Carson and B.L. Nelson, *Discrete-Event System Simulation*. Prentice-Hall, New Jersey, 1996.

● 저자소개 ●



오유영

1998년 아주대학교 정보 및 컴퓨터공학부(공학사)

1998년~현재 아주대학교 컴퓨터공학과 석사과정 재학 중

관심분야 : 컴퓨터 구조, 결합 허용, 멀티미디어, 이동 컴퓨팅, 시뮬레이션



김성수

1982년 서강대학교 전자공학과(공학사)

1984년 서강대학교 전자공학과(공학석사)

1995년 Texas A&M University, 전산학과(공학박사)

1983~1986년 삼성전자(주) 종합연구소 컴퓨터연구실(주임연구원)

1986~1996년 삼성종합기술원 수석연구원

1991~1992년 Texas Transportation Institute 연구원

1993~1995년 Texas A&M University, 전산학과, T.A.&R.A.

1997~1998년 한국정보과학회, 한국정보처리학회 논문지 편집위원

1996~현재 아주대학교 정보통신대학 정보 및 컴퓨터공학부 교수

관심분야 : 결합 허용, 멀티미디어, 이동컴퓨팅, 성능평가, 시뮬레이션



김재훈

1984년 서울대학교 제어계측공학과 (학사)

1993년 Indiana University 전산학과 (석사)

1997년 Texas A&M University 전산학과 (박사)

1995~1997년 Texas A&M University 전산학과 Graduate Research Assistant

1984~1991년 대우통신(주) 컴퓨터 연구실

1997~1998년 삼성전자 컴퓨터시스템팀

1998~현재 아주대학교 정보 및 컴퓨터공학부 조교수

관심분야 : 분산시스템, 실시간시스템, 운영체제, 결합 허용