분석 대상 SSD의 특성을 고려한 자동 블록 I/O 트레이스의 변환기의 구현

온준엽[0] 김지홍
서울대학교 컴퓨터공학부
{jyon, jihong}@davinci.snu.ac.kr

Design and Implementation of an SSD-Aware Automatic Block I/O Trace Converter

Junyub On[0] Jihong Kim
Seoul National University

요 약

블록 I/O 트레이스는 특정 어플리케이션을 수행할 때 발생하는 I/O 패턴을 기록하여 추후에 특정 시스템의 I/O 성능을 평가하기 위해 사용된다. 하지만 실제 연구에서 활발히 사용되고 있는 블록 I/O 트레이스는 대부분 오래전 하드디스크 기반의 시스템에서 수집되어 최근 SSD 기반의 시스템 평가에는 적합하지 않다. 본 논문에서는 기존의 트레이스를 특정 SSD의 성능에 따라 자동 변환하는 도구를 소개하고, 자동변환기에 의해 변환된 트레이스가 실제 SSD에서 어플리케이션을 수행하였을 때 발생하는 트레이스와 비슷한 특성을 가지고 있음을 실험적으로 검증하여 개발한 변환기의 유용성을 보인다.

1. 서 론

최근 활발히 사용되고 있는 낸드 플래시 기반의 SSD는 이전까지 주요 저장 장치로 사용된 하드디스크에 비해 높은 성능과 저전력, 그리고 높은 내구성 등의 특성을 지닌다. 유일한 단점이었던 가격 또한 지속적으로 하락하여 1$/1GB 이하의 수준에 이르렀다. 이에 따라 SSD는 개인용 PC 뿐만 아니라 모바일 및 서버에서도 널리 사용되고 있는 추세이다.

SSD의 연구에서 구현한 시스템의 성능을 평가하기 위한 한 가지 방법은 특정 어플리케이션을 수행하여 수행 시간을 측정하는 것이다. 이 방법은 특정 어플리케이션에 대한 성능을 비교적 정확하게 측정할 수 있다는 장점이 있지만, 직접 어플리케이션을 수행하기 위한 환경을 구축해야하는 번거로움이 있다. 이에 따라 실제 연구에서는 특정 어플리케이션의 I/O 패턴을 모사하는 블록 I/O 트레이스를 사용하는 경우가 많다. 즉, 특정 어플리케이션을 수행할 때 파일시스템과 저장 장치 사이에서 발생하는 트레이스를 기록하여 놓았다가 이를 다시 SSD에서 재생하는 방법으로 성능을 평가한다. 이 경우 연구자가 직접 어플리케이션을 수행하기 위한 환경을 구축하지 않아도 인터넷을 통해 배포되는 다양한 블록 I/O 트레이스들을 이용하여 성능을 평가할 수 있다. 이를 위해 SNIA와 같은 단체에서는 서버나 모바일 등 다양한 환경에서 수집된 블록 I/O 트레이스를 배포하여 연구에 이용할 수 있게 하고 있다.

하지만 SNIA에서 배포되는 트레이스들을 포함하여 연구에 사용되는 블록 I/O 트레이스들은 하드디스크 기반의 시스템에서 수집된 것이 대부분이다. 실제 어플리케이션이 수행될 때 각 I/O 요청들의 간격(I/O inter-arrival time)과 I/O에 대한 반응 시간(I/O response time)은 저장 장치의 성능에 크게 영향을 받기 때문에 하드디스크에서 수집된 블록 I/O 트레이스를 그대로 SSD에서 리플레이하는 경우, 정확한 성능의 평가가 이루어지기 어렵다. [1]

이에 본 논문에서는 특정 SSD의 반응 시간(response time)을 측정하고, 이에 따라 블록 I/O 트레이스를 변환하는 기법을 제시한다. 이 기법을 이용하여 하드디스크에서 수집된 블록 I/O 트레이스를 변환한 결과를 실제 SSD에서 측정한 데이터와 비교하였을 때, I/O 지연 시간이 비슷한 특성을 보이는 것을 확인하였다.

관련된 연구로는 블록 I/O 트레이스를 리플레이할 때, 각 I/O가 요청된 시각에 맞춰서 그대로 리플레이 하지 않고, 현재 처리되지 않은 I/O 개수에 맞춰서 리플레이하는 기법에 대한 연구와 [2], 하드디스크에서 수집된 블록 I/O 트레이스를 SSD 기반의 시스템에서 리플레이하여 재구성한 연구가 있다. 첫 번째 연구에서는 저장장치의 성능으로 인해 변하는 I/O 요청 속도를 반영할 수 있지만, I/O 사이의 간격이 긴 경우에는 정확한 리플레이가 어렵다는 단점이 있다. 두 번째 연구에서는 실제 연구에서 자주 사용되는 블록 I/O 트레이스를 SSD에 맞춰 재구성하였으나, 이를 위해서 큰 용량의 SSD로 구성된 시스템을 구축하는 비용과 리플레이를 위한 시간이 많이 든다는 단점이 있다. 또한 두 연구 모두 실제로 성능을 평가하고자 하는 SSD의 성능을 고려하지 않고 일반적인 상황을 가정하기 때문에 특정 SSD에서는 부정확한 평가가 이루어질 수 있다. 이에 비해 본 연구에서는 실제 SSD의 성능을 트레이스의 변환에 반영하기 때문에 보다 정확한 성능 평가가 이루어진다.

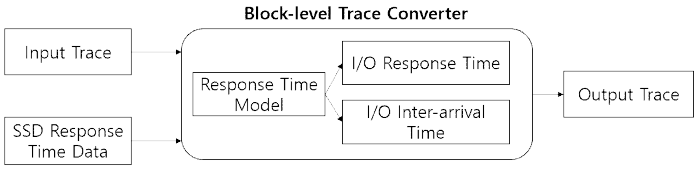

그림 1 트레이스 변환 과정

2. 블록 I/O 트레이스 변환 기법

이번 절에서는 블록 I/O 트레이스를 저장 장치의 성능에 따라 변환하는 기법에 대해 설명한다.

대부분의 블록 I/O 트레이스에는 각 I/O 요청이 시작된 시각, I/O 요청의 종류, I/O 요청의 논리주소, 요청의 크기, 각 I/O의 반응 시간 등이 포함되어 있다. 이 중에서 각 I/O의 반응 시간과 인접한 두 I/O 요청 사이의 간격은 저장 장치의 성능에 큰 영향을 받는다.

따라서 본 논문에서는 그림 1 과 같이 SSD의 성능을 고려하여 블록 I/O 트레이스를 변환하는 도구를 구현하였다. 이 도구에서는 트레이스에 대한 변환을 하기 위해 먼저 성능을 측정하고자 하는 SSD의 반응 시간의 데이터를 입력하여 이를 읽기와 쓰기에 대해 각각 통계적으로 모델링 한다. 이어서 입력으로 들어오는 블록 I/O 트레이스의 I/O 사이의 간격 및 I/O의 반응시간을 모델에 따라 변환하여 SSD의 성능에 맞는 트레이스를 생성한다.

2.1 SSD 반응 시간 모델링

트레이스의 변환 과정에 특정 SSD의 성능을 반영하기 위해 SSD에서 측정된 반응 시간 데이터를 모델링 하였다. SSD의 반응 시간은 GC 등의 동작으로 그림 2와 같이 여러 개의 피크를 나타내며, 꼬리가 매우 길다. 이러한 특성 때문에 정규분포나 지수분포와 같은 모델로는 반응 시간을 정확하게 모델링하기 어렵다. 본 논문에서는 여러 개의 정규분포의 합으로 특정 분포를 나타내는 가우시안 혼합 모델(Gaussian Mixture Model, GMM)을 사용하여 반응 시간을 모델링 하였다. 가우시안 혼합 모델은 수식으로 다음과 같이 나타낼 수 있다.

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$$

즉, 전체 분포를 K 개의 정규분포의 합으로 나타내고, 각 정규분포의 평균은 μ_k , 분산은 Σ_k이며 혼합가중치는 π_k인 분포이다. 혼합가중치는 다음과 같은 성질을 만족해야 한다.

$$0 \le \pi_k \le 1, \sum_{k=1}^{K} \pi_k = 1$$

가우시안 혼합 분포는 원하는 만큼의 정규 분포를 혼합할 수 있기 때문에, SSD의 반응 시간과 같이 여러 개의 피크를 가지고 꼬리가 길게 나타나는 분포를 표현하기 적합하다.

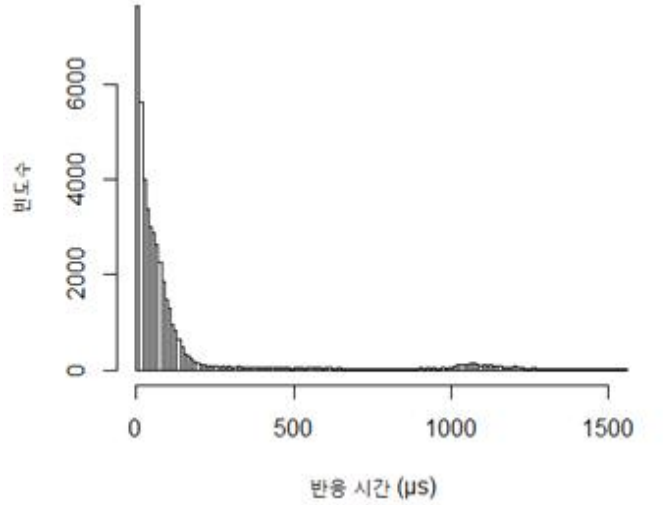

그림 2 읽기 반응시간 분포

반응 시간의 측정을 위해서 FIO를 통해 여러 번의 읽기와 쓰기 요청을 발생시키면서 SSD에서의 반응 속도를 측정한다. 이를 통해 얻어진 SSD의 반응 시간 데이터를 가우시안 혼합 모델으로 모델링하였다. 정확한 변환을 위해 읽기와 쓰기의 분포에 대해 각각 모델을 생성하였다.

생성한 모델의 적합도를 평가하기 위해 KS test를 통해 평가하였다. KS test는 분포 모델과 데이터를 CDF로 나타내었을 때 둘 사이의 차이가 가장 큰 값을 유의 확률(D)로 하여 이 값이 작을수록 모델이 정확하다는 것을 나타낸다. 읽기와 쓰기 모델 모두 $D < 0.03$으로 가우시안 혼합 모델을 이용하여 생성한 SSD의 반응 시간 모델이 정확한 것을 확인하였다.

2.2 I/O 반응 시간 변환 기법

특정 SSD에 맞춰 I/O 반응 시간을 변환하기 위해 입력으로 들어오는 트레이스를 읽어 들이면서, 읽기와 쓰기 요청들에 대해 앞에서 생성한 읽기와 쓰기 모델에 맞춰 난수를 발생시킨다. 이 값을 SSD에서의 I/O 응답시간으로 하여 변환을 진행한다.

2.3 I/O 요청 사이 간격 변환 기법

블록 I/O 트레이스에 나타나는 I/O 요청 사이의 간격은 인접한 두 I/O가 서로 의존하는 경우와 그렇지 않은 경우로 나눌 수 있다. 하지만 트레이스에는 이러한 정보가 포함되어 있지 않기 때문에, I/O 요청 사이의 간격을 정확히 변환하는 것은 어렵다.

다만 서로 독립적인 I/O의 경우, 요청 사이의 간격이 더 큰 경향을 가지고 있다. 이에 본 논문에서는 입력으로 들어오는 블록 I/O 트레이스의 인접한 요청 사이의 간격이 특정 임계값 이내이면 의존 관계가 있는 I/O로

판단하여 두 I/O 사이의 간격을 SSD에서 서로 의존하는 I/O 사이의 간격에 맞춰 감소시키고, 그렇지 않은 경우 두 I/O 사이의 간격을 그대로 유지하도록 하였다.

3. 실험 및 결과

3.1 실험 환경

실험에 사용한 트레이스는 MSR 트레이스이다. 변환을 위해 삼성 SSD 830에서 읽기와 쓰기에 대한 모델을 추출하였다. 생성된 모델을 기반으로 MSR 트레이스를 변환한 후, 읽기와 쓰기에 대한 I/O 반응 시간이 SSD에서 추출한 반응 시간 데이터와 어떤 차이를 보이는지 관찰하였다.

3.2 실험 결과

그림 3은 SSD 830에서 측정한 쓰기 반응 시간을 나타낸다. MSR 트레이스를 변환하여 MSR 트레이스 [3]의 쓰기 반응 시간의 분포를 구해본 결과 그림 4와 같이 실제 SSD의 반응 시간과 거의 일치하는 분포를 나타내었다. KS test 결과, $D < 0.03$으로 정확하게 변환이 이루어진 것을 확인하였다. 변환된 MSR 트레이스의 읽기 응답시간의 경우에도 $D < 0.007$로 읽기와 쓰기 모두 3% 이내의 오차를 나타냈다. 다만 일반적으로 배포되고 있는 블록 I/O 트레이스에는 I/O 간격에 대한 정보가 없기 때문에 이에 대해 변환이 올바르게 이루어지는 지에 대해서는 확인하기 어려웠다.

4. 결론 및 향후 연구

본 논문은 하드디스크에서 수집된 블록 I/O 트레이스를 특정 SSD의 성능에 맞춰 변환하는 기법을 제안하였다. 트레이스를 변환하기 위해 먼저 SSD에서 반응 시간을 측정하고, 이를 모델링하였다. 이후 I/O 사이의 간격과 I/O의 반응 시간을 생성한 모델에 따라 변환하였다.

실험 결과를 통해 제안된 기법을 통해 변환된 트레이스가 실제 SSD에서 측정된 I/O 반응 시간과 비슷한 특성을 가지는 것을 확인하였다.

널리 사용되고 있는 블록 I/O 트레이스에는 I/O 사이의 간격에 대한 아무런 정보가 포함되어 있지 않다. 이로 인해 변환된 트레이스의 I/O 간격이 제대로 변환되었는지를 검증하는 것은 어려운 문제이다. 실제 사용되고 있는 블록 I/O 트레이스들에 대해 제안한 기법이 I/O 간격에 대한 변환을 올바르게 하고 있는지를 검증하는 것이 앞으로 연구할 과제이다.

5. 감사의 글

이 연구를 위해 연구장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에 감사드립니다. 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015M3C4A7065645). (교신저자: 김지홍)

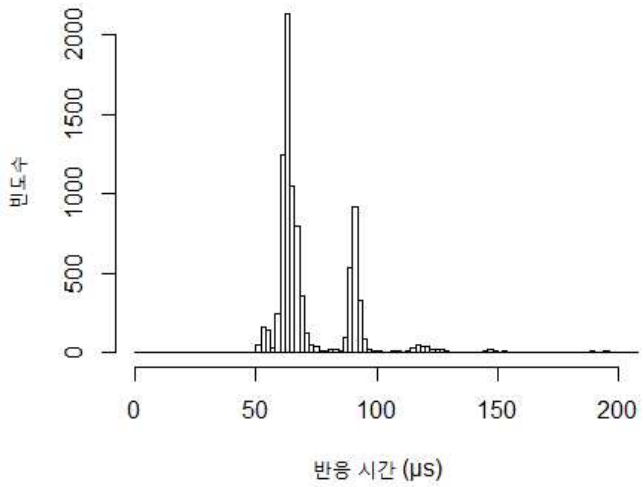

그림 3 쓰기 반응 시간 데이터

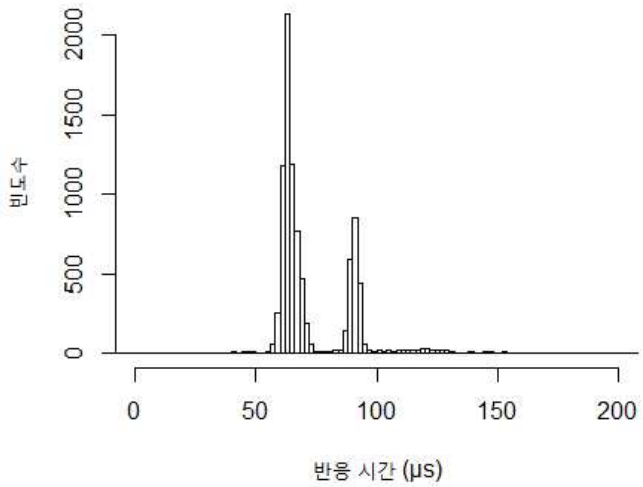

그림 4 변환된 쓰기 반응 시간

참 고 문 헌

[1] Mesnier, Michael P., et al, “Relative fitness modeling,” in *Communications of the ACM*, vol. 52, pp. 91-96, 2007.

[2] Sankaran Sivathanu et al., “Load-Aware Replay of I/O Traces,” in *Proc. of Int’l USENIX Conf. on File and Storage Technologies*, FAST, 2011.

[3] Narayanan, Dushyanth, Austin Donnelly, and Antony Rowstron. “Write off-loading: Practical power management for enterprise storage,” in *ACM Transactions on Storage*, vol. 4, 2008.