인피니밴드 기반의 네트웍 저장장치에서 SRP 프로토콜 개발 및 성능평가

*김영환, 전기만, 박창원
전자부품연구원 지능형정보시스템 연구센터(yhkim93,kmjeon,parkcw@keti.re.kr)

1. 서 론

현재 컴퓨터 시스템에서 사용되고 있는 입출력 버스 방식은 디스크 접근, 특히 고 성능의 서버에 있어서 병목 현상의 주요원인으로 나타나고 있다. 이러한 버스 방식은 구조가 단순하다는 큰 이점을 가지고 있어 지금까지 산업 전반적으로 사용되어 왔지만 버스 기반의 입출력 시스템은 현재의 디바이스 장치들이 요구하는 데이터 전송 대역폭을 처리할 수 있을 만큼의 시스템 입출력 성능을 가지고 있지 않다.

앞서 설명한 시스템 상의 문제는 현존하는 모든 시스템 기반의 저장장치가 공통적으로 가지고 있다. 이를 해결하기 위한 방안으로 채널 기반의 네트웍 저장장치가 대두되게 되었고, 관련 업계에서는 IBTA라는 기술 표준화 단체를 통해 표준화를 진행해 왔다. 그 대표적인 기술로 인피니밴드가 있다. 본 논문에서는 인피니밴드 기반의 저장장치를 구성하여 SRP 프로토콜을 통해 저장장치에 연결된다. 본 논문에서는 SRP를 구현하고 이를 기존의 FC(Fibre Channel)기반의 i-SCSI와 성능을 비교 분석할 것이다.

2. 관련연구

2.1 TCP/IP기반의 저장장치 접근 프로토콜

TCP/IP 기반 저장장치 기술의 가장 큰 이점은 지금 당장 사용 가능할 뿐만 아니라 비용 절감효과가 높다는 것이다. 이와 같은 장점을 발휘하는데 가장 큰 공을 세운 것은 이더넷의 급성장이다. 파이버 채널에 근접하는 수준으로 고속을 구현했기 때문이다. TCP/IP 기반 저장장치의 이점으로는 우선 상호 연동성이 있다. 비용 절감 측면에서는 파이버 채널 HBA(Host Bus Adapter)와 파이버 채널 스위치가 필요 없기 때문에 비용 지출이 줄어든다. 기존 IP 네트워크의 장점을 물려받는 TCP/IP 기반 저장장치 기술의 혜택으로는 QoS(Quality of Service)와 보안을 빼놓을 수 없다. IPSec, 3DES, 방화벽, ACL(Access Contol List), VPN(Virtual Private Network) 등 표준 IP 보안기능을 그대로 사용할 수 있다.

2.2 인피니밴드 기반의 저장장치 접근 프로토콜

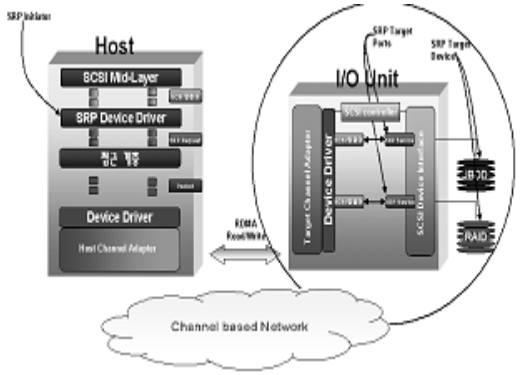
인피니밴드 저장장치는 서버 클러스터 시장에서 우수한 성능을 인정받고 인피니밴드 저장장치에 대한 많은 요구를 이끌어 내게 되었다. 이와 함께 저장장치 성능에 매우 민감한 응용 서비스에 대해서도 많은 해결책을 내놓았다.

인피니밴드 기반 저장장치 기술을 구현하는 프로토콜은 SCSI 저장장치에 대한 접근 인터페이스 기술을 정의하는 Technical Commitee T10 의 SRP(SCSI RDMA Protocol)가 있다[.

- SRP(SCSI RDMA Protocol)

SRP(SCSI RDMA Protocol)는 원래 ANSI NCITS T10 워킹 그룹에 의해 개발되었다. SRP 는 원격의 SCSI 장비를 제어하기 위한 프로토콜로 제안되었고, 인피니밴드 기술의 특성에 맞게 사용되도록 구현되었다.

그림 1 은 SRP 프로토콜에 대한 전체 블럭도이다. SRP 는 초기자가 SCSI 작업을 생성하고 이를 타겟에서 수행하도록 하는 기본적인 서버-클라이언트 모델이 가능한 전송 서비스를 제공하는 프로토콜이다. 또한 SRP 와 관련한 모든 통신은 신뢰성을 기반으로 한 연결 서비스를 제공해야 한다. SRP 는 메시지 흐름 제어 메커니즘을 제공하는데 초기자에 의해 생성된 작업 요구에 대한 디스크립터를 큐에 넣을 수 있는 수를 타겟이 제한할 수 있도록 하고 있는데, 이 메커니즘은 다중 초기자에 의해 필요한 메시지 버퍼를 동적으로 할당할 수 있어 내부 자원을 관리하는데 사용된다. 따라서 제한된 자원에 대한 적절한 이용을 통해 전체 시스템 성능을 향상시킬 수 있다.



<그림 1> SRP 전체 블럭도

SRP 타겟은 모든 데이터 전송을 초기자 메모리에 직접 읽고 쓰기가 가능하도록 RDMA 기능을 포함하고 있다. 초기자는 자신의 데이터 버퍼를 등록하고, 그 내용을 전송할 SRP 명령어 내에 포함 시킴으로서 타겟으로부터 RDMA 접근이 가능하다. 다음은 SRP 프로토콜의 I/O 과정을 단계별로 설명한다.

1. 초기자는 SCSI 미들웨어로부터 SCSI 명령어와 LUN(Logical Unit Number) 그리고 데이터 버퍼 디스크립터를 포함한 SRP 요구 메시지를 생성하고, 타겟으로 해당 메시지를 전송한다.

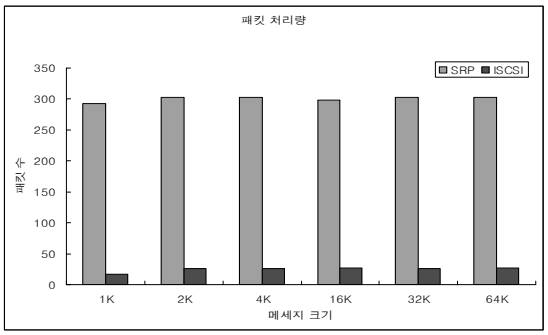
2. 타겟은 SRP 요구 메시지를 받고 메시지에 포함되어있는 초기자의 버퍼 공간 정보를 기반으로 RDMA 전송을 수행한다.

3. 타겟은 해당 요구 작업에 대한 완료 내용을 담은 SRP 응답 메시지를 생성하고 초기자에게 전송한다.

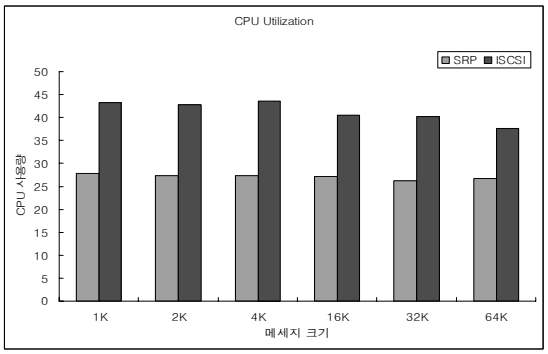
또한, 초기자는 타겟 상에 존재하는 작업(task)을 무시할 수 있는 SRP 작업 관리에 동작을 수행할 수 있다. 게다가, 타겟은 새로운 미디어 추가와 같은 비동기적으로 발생하는 이벤트에 대한 메세지를 초기자에게 전송할 수 있다.

3. 성능평가

본 논문에서 성능 평가를 위해 메시지를 임의로 패킷 처리량의 변화에 따라 작은 사이즈의 메세지, 중간 사이즈의 메세지 그리고 큰 사이즈의 메시지로 구분하여 측정하였다. 그리고 Raw Disk IO 방식을 이용해 Disk Cache 에 읽기와 쓰기 작업을 각각 50% 할당하였다. 또한 읽기와 쓰기 작업은 3 분 동안 순차적인 방식으로 수행하도록 했다.



<그림 2> 패킷 처리량



<그림 3> CPU Utilization

[참 고 문 헌]

[1]Infiniband Architecture Specification, Release 1.1 Infiniband Trade Association, 2002.

[2]S. Bhattacharya, S. Pratt, B. Pulavarty, and J. Morgan. Asynchromous I/O Support inLinux 2.5. In Proceedings of the Linux Symposium, pp.371-386, July 2003

[3]*Cohen, A*, "A performance analysis of the sockets direct protocol (SDP) with asynchronous I/O over 4x infiniband", 2004 IEEE International Conference , pp. 241-246, April 2004.