IniniBand 링크 계층 설계

*권영민, 박창원, 정하중
전자부품연구원 지능형정보시스템 센터
Email : youngminy@keti.re.kr

1. 서 론

현재의 연결망 기술은 인터넷 시대에 접어들면서 폭발적으로 증가하는 데이터 전송량을 처리하기 위해 보다 넓은 대역폭과 짧은 대기시간을 요구하게 되었다. 대역폭의 문제를 국소적으로 해결하기 위해 AGP, PCI-X와 같은 기술들이 발표되었으나 이들 역시 여러 I/O 디바이스가 버스를 공유하는(multi-droop)하는 병렬버스(parallel bus technoligy)구조로 되어있어 성능향상의 한계를 가지고 있다.

InfiniBand architecture(IBA)는 기존 연결망기술의 한계를 극복하기 위해서 point- to-point 방식을 이용한 직렬연결 방식으로 디자인 되어 lane 당 2.5GB/s, 최대 30Gbps의 대역폭을 가진다. 또한 IBA는 하드웨어 트랜스포트 프로토콜을 정의함으로써 소프트웨어 독립적으로 reliable transaction과 RDMA (remote DMA)를 지원한다. 이러한 장점으로 인해 IBA는 블레이드 서버, 스토리지 서버의 고속 연결망 표준으로 자리 잡을 것으로 보인다.

본 논문에서는 IBA의 링크계층에 대해 소개하고, 최소한의 buffer로써 효율적으로 data 처리 가능한 링크 계층 구조를 제안하고자 한다.

2. 요 약

IBA system area network(SAN)은 그림 1과 같이 프로세서 노드와 I/O 유닛들이 직렬 스위치 구조로 연결되어 있는 형태이다.

그림2는 IBA의 계층구조로써 물리계층, 링크계층, 네트웍계층, 전송계층으로 구성되어 있다.

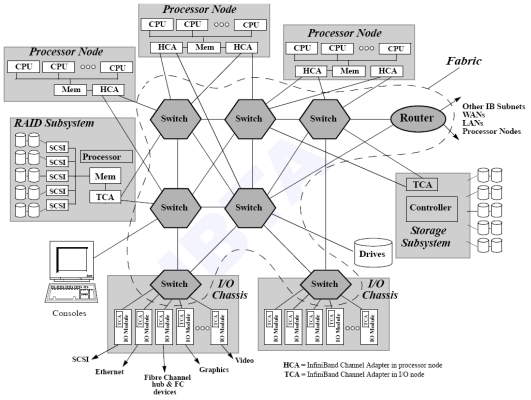

그림 1 IBA System Area Network

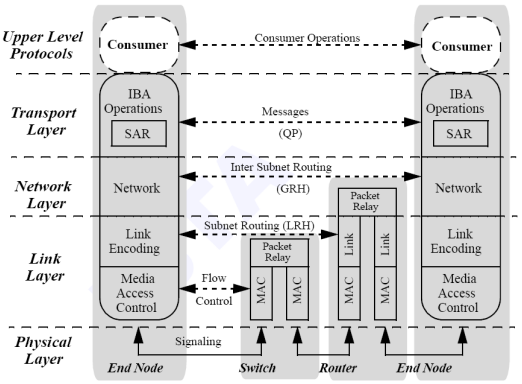

그림 2 IBA Architecture Layer

링크계층은 물리계층과 네트웍 계층 사이에서 addressing, error detecting and switching 그리고 flow control을 하며 그림 3은 본 논문에서 설계한 링크계층의 top block diagram이다. 링크 계층은 link state machine(LSM), error detection logic, virtual lane(VL) buffer, service level(SL)to VL mapping unit, VL arbitration unit 등으로 구성되어 있으며 상위로는 네트웍계층과 management entity, 하위로는 물리계층과 인터페이스를 가지지고 있다.

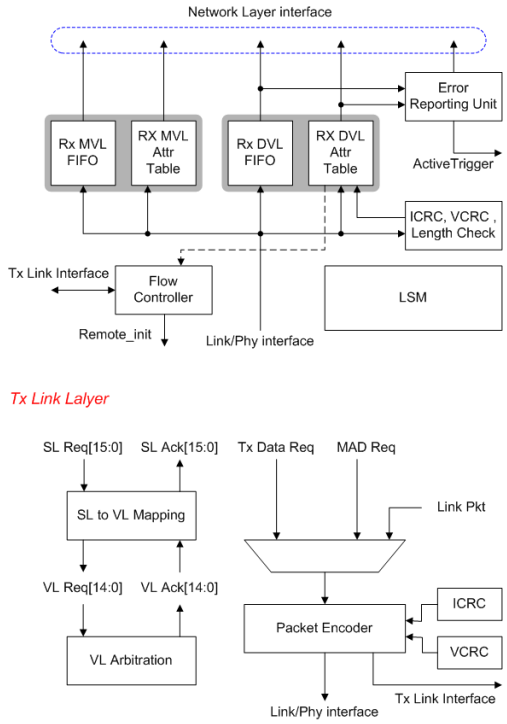

그림 3 Link Layer Top Block Diagram

그림 4는 링크계층의 주요 기능중 하나인 Rx data VL buffer이다. Rx data VL buffer는 물리계층에서 전송된 data의 error 검출과 Rx VL flow control을 담당한다.

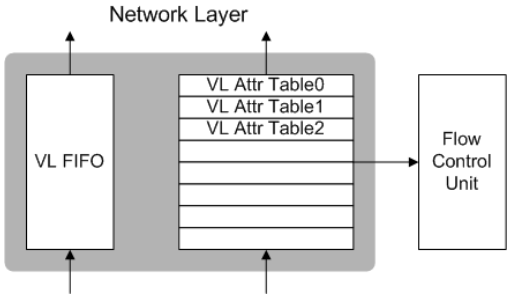

그림 4 Rx data VL buffer

Data VL buffer는 data 저장을 위한 VL FIFO와 VL attribute table로 구성되어 있다. VL attribute table은 그림 5와 같은 필드로 구성되어 있다. Flow control unit은 그림 6과 같이 VL attribute table에 저장되는 packet size와 VL 정보를 이용해서 각 VL의 flow control을 수행함으로써 하나의 data VL FIFO 만으로도 효과적인 data 전송과 기능의 중복을 피할 수 있다.

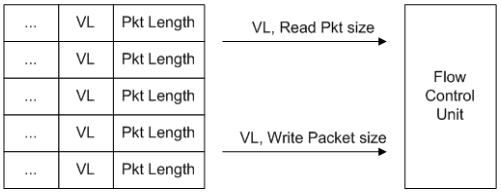

그림 5 Flow Control Mechanism

Tx link layer의 구조를 단순화하기 위해 그림 6과 같은 전송 mechanism이 사용되었다.

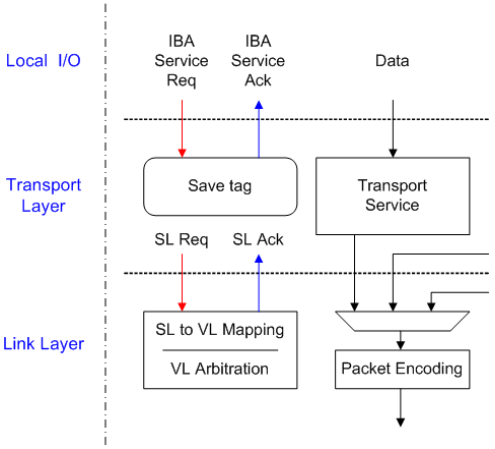

그림 6 Tx Mechanism

Data 전송이 일어나기 이전에 Tag 정보를 이용해서 SL to VL mapping과 VL arbitration을 수행함으로써 data를 저장하기 위한 별도의 VL buffer를 필요로 하지 않는다.

3. 실험 결과

본 논문을 통해 최소한의 buffer를 사용한 효율적인 링크 계층의 설계를 제한했다.

현재 functional test와 프로토콜 검증을 진행 중이다.

Reference

[1] InfiniBand Architecture release 1.1 November 6, 2002

[2] MindShare, Inc. Tom Shanley, InfiniBand Network Architecture