차세대 인터넷 서버를 위한 스트리밍 가속장치

A Streaming Accelerator for the Next Generation Internet Server

김 성 운*, 김 명 준*, 김 보 관**

* 한국전자통신연구원 (전화(042)860-5745, E-mail : ksw@etri.re.kr)

** 충남대학교 전자공학과 (전화(042)821-6585, E-mail : bgkim@cnu.ac.kr)

Abstract

The requirements for the high quality service is increased according to the changing of the internet environment. But, the computer system can not satisfy the requirement because of the limitation of the computer's own problem. This paper describes the network storage accelerator which can perform the high quality streaming service. We also show the implementation result of the streaming accelerator.

서 론

지난 수년간 네트워크 대역폭의 빠른 증가와 컴퓨터 성능의 향상으로 인터넷이 급속히 보급되었다. 이러한 인터넷 환경의 변화로 네트워크를 통한 멀티미디어 데이터 송수신이 급증하게 되었고, 이에 따른 네트워크를 통한 스트리밍을 빠르게 처리할 수 있는 시스템이 요구되고 있다[1]. 차세대 인터넷 서버(이하 NGIS)는 수 천 내지 수 만 명이 동시에 20Mbps의 대역폭으로 HDTV급 동영상 서비스할 수 있도록 지역서버와 광역서버를 포함하는 계층적 구조를 갖는 서비스 시스템이다[2]. (그림 1 참조)

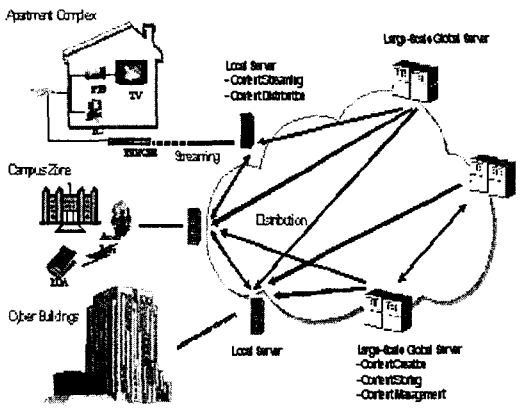

그림 1. 차세대 인터넷 서버의 개념도

NGIS는 인터넷 데이터센터와 같은 곳에 대용량 컨텐츠를 저장하고 서비스하는 광역 서버를 두고, 아파트 단지, 학교 캠퍼스 또는 빌딩과 같은 서비스가 직접 일어나는 곳에 지역 서버를 두는 계층적 구조로 수만 명에게 동시에 고품질의 서비스를 제공할 수 있다.

전통적인 컴퓨터 구조는 고성능 연산을 목적으로 만들어져 있기 때문에, 컴퓨터의 네트워킹 기능은 운영체계 위에 올라가는 하나의 어플리케이션으로 다루어져 왔다. 이러한 컴퓨터 구조는 빠르게 발전하고 있는 인터넷 환경에 적절하게 대응하기가 쉽지 않다. 즉, 대량의 데이터가 네트워크를 통해 전송되고 이를 컴퓨터 내부의 메모리에 복사를 하고 다시 저장장치로 복사하는 일련의 일들은 전통적인 컴퓨터 구조에서는 빈번하게 데이터 복사가 반복되기 때문에 컴퓨터 성능을 크게 저하시킨다. 즉 대량의 데이터를 처리할 수 있는 하드웨어 및 운영체제 고유의 기능이 없기 때문에, 대용량의 네트워킹 관련 서비스는 단지 몇 개의 HDTV급 스트리밍 서비스만을 수행해도 시스템의 부하가 급속히 늘어나서 QoS를 만족시킬 수 없게 된다. 이를 위해 TCP/IP 오프로드 엔진 등이 제안되기도 하였다.[3][4]

NGIS의 지역서버는 일정한 지역에 있는 수백명의 사용자들에 고품질의 서비스를 제공하기 위해 대용량의 스토리지에서 고품질의 멀티미디어 데이터를 네트워크로 직접 내 보낼 수 있는 스트리밍 가속장치가 내장되어 있는 새로운 개념의 차세대 인터넷 서버이다. 지역서버의 운영체계는 이러한 대량의 데이터 처리를 가능하게 하는 스트리밍 가속 장치의 특성에 맞도록 수정되어 있으며, 또한 멀티미디어 데이터 처리를 수행하는 고유의 파일 시스템을 가지고 있다.

응용 프로그램의 직접적인 제어와 zero-copy 메커니즘을 제공하는 스트리밍 가속장치는 PCI 버스를 통해서 접근 가능한 PCI 메모리를 제공한다. 또한 빠른 스토리지 접근 장치와 Gigabit 이더넷 TOE (TCP/IP Offload Engine)을 내장하고 있다.

본 논문은 NGIS 지역서버에서 사용되는 스트리밍의 가속장치 구조 및 각 구성 요소의 기능을 간단히 살펴보고, 스트리밍 가속장치의 구현된 모습과 네트워킹 가속장치를 이용하여 실제의 스트리밍 처리 성능을

실측치를 바탕으로 살펴 본다.

1. 스트리밍 가속장치 구조

스트리밍 가속장치는 20Mbps의 대역폭으로 동시에 200명의 HDTV급 스트리밍 서비스를 할 수 있어야 한다. 스트리밍 가속장치는 이런 성능의 스트리밍 서비스를 위해 TCP, UDP 등의 네트워크 동작의 오버헤드를 최소화 할 수 있도록 부분적인 TCP/IP 오프로드 처리 기능을 가지고 있다. 또한, 사용자에게 스트리밍 서비스를 위해 대량의 데이터가 반복적으로 복사하는 것을 막기 위해 zero-copy 기능을 가지고 있다.

또한 대량의 스트리밍 데이터를 빠르게 디스크로부터 읽어 오기 위해 파이프라인드 I/O 처리를 수행하는 스토리지 제어 기능을 포함하고 있다. 스토리지 제어기는 최대 2 테라바이트의 디스크 용량을 보유하여 수백 편의 동영상을 하나의 지역서버에 저장할 수 있다.

스트리밍 가속장치는 그림 2와 같이 PCI 브릿지, TOE, 스토리지 제어기와 디스크, 메모리 제어기로 구성되는 하드웨어와 이러한 하드웨어를 제어하는 드라이버들로 이루어져 있다. 이외에도 스트리밍 가속장치를 진단하고 감시하는 진단 및 관리 소프트웨어로 구성된다.

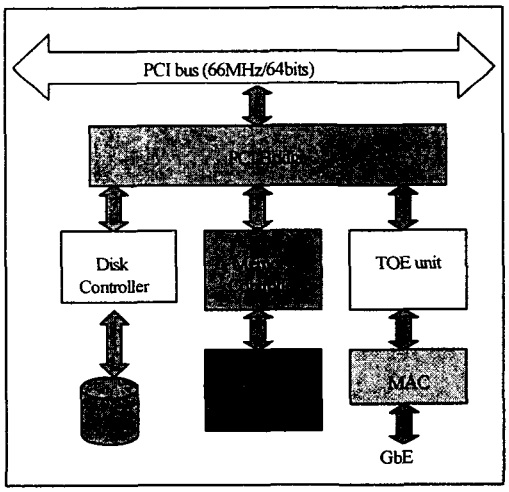

그림 2. 스트리밍 가속장치 내부 구성

응용 프로그램의 직접적인 제어와 zero-copy 메커니즘을 제공하기 위해서 스트리밍 가속장치는 PCI 버스를 통해서 접근 가능한 PCI 메모리를 제공한다. PCI 메모리는 스토리지와 네트워크의 데이터 흐름에서 주 프로세서의 간섭을 줄일 수 있는 완충 역할을 수행하고, 응용 프로그램의 융통성을 높일 수 있다.

PCI 메모리 드라이버는 PCI 메모리의 물리적인 정보를 획득하고 유지한다. 이러한 정보에는 물리적인 메모리 시작 주소, 메모리 크기를 포함한다. PCI 메모리 관리를 위한 데이터 구조는 그림 3과 같다. 물리적인 메모리 공간은 블록 단위로 나누어서 관리를 하는데, 전송되어야 할 멀티미디어 매체에 따라 블록 크기를 2MB, 1MB, 512kB, 256kB로 선택할 수 있다.

PCI 메모리를 사용하기 위해서는 메모리 드라이버에서 제공하는 PCI 메모리 블록 할당, 블록 해제, 사용자 주소 맵핑, 가용한 PCI 메모리 블록의 수를 알려 주는 인터페이스를 사용한다.

스트리밍 가속장치의 PCI 메모리는 여러 사용자 프로세스에 의해 사용된다. 사용자 프로세스는 운영체제를 통해서 한번에 연속된 하나의 PCI 메모리 블록을 얻을 수 있다. 또한 할당된 메모리 블록의 해제를 요청할 수 있다. 더욱이 여러 PMEM 메모리 블록을 소유한 프로세스가 제대로 해제하지 않고 죽을 수 있다. 이러한 다양한 형태를 지닌 사용자 프로세스를 위해서 한정된 PCI 메모리는 자원의 효율성과 편리성을 위해서 체계적으로 관리된다.

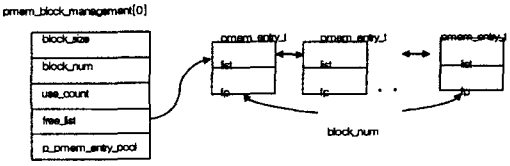

그림 3. PCI 메모리 관리 데이터 구조

TOE는 NGIS 지역 서버에서 고성능의 네트워크 기능을 제공하여 차세대 인터넷 서버의 주요 요구사항을 만족시켜 주는 중요한 역할을 담당한다. NGIS 지역서버의 요구사항 중에 가장 핵심이 되는 것은 멀티미디어 스트리밍 데이터의 네트워크를 통한 전송 성능이므로 TOE의 성능은 NGIS 지역서버 전체 성능을 좌우한다고 볼 수 있다. NGIS 지역서버의 전체 구조에서 TOE의 구성 부분은 실제 Linux 커널로부터 명확하게 분리할 수는 없으나, 구현의 관점에서 보면 Linux커널의 TCP/IP 스택 일부분과 네트워크 컨트롤러 디바이스 드라이버로 구성되어 있다.

TOE 가장 중요한 요구 사항은 데이터 전송 속도로서 한 개의 스트리밍 가속장치는 50명의 동시 사용자에게 20Mbps의 스트리밍 데이터를 전송할 수 있어야 한다. 이러한 고속의 데이터 전송이 가능하도록 스트리밍 제어장치는 CPU의 부하를 줄이기 위하여 저장장치로부터 직접 네트워크로 스트리밍 데이터를 전송할 수 있도록 PCI 메모리를 이용한다. 기존의 TCP/IP 스택만으로는 PCI 메모리를 지원하면서 스트리밍 가속장치에 요구되는 성능을 만족시키기가 불가능하다. 따라서,

TOE는 PCI 메모리의 사용을 지원하면서 데이터 이동 경로를 최적화하여 PCI 메모리로부터 직접 네트워크로 스트리밍 데이터 전송이 가능하도록 하고, TCP/UDP/IP checksum offload와 Scatter/Gather I/O 기능을 제공한다.

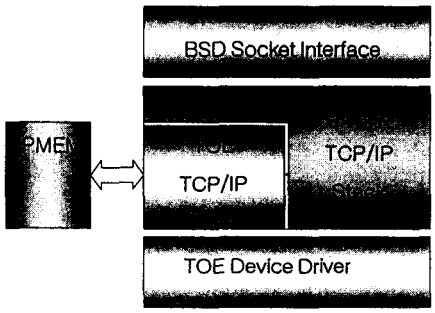

그림 4. TOE 프로그램 구조

디스크 제어기는 최소한 1Gbps 대역폭으로 스트리밍 데이터를 전송할 수 있도록 초고속의 획기적인 디스크 입출력 방식을 사용하고 있다. 멀티미디어 데이터는 블록 분할 방식에 의하여 디스크 어레이를 구성하는 모든 디스크들이 동일하게 병렬적으로 저장하거나 읽어 낼 수 있도록 구성되어 있다. 이러한 블록 분할 방식과 더불어서 디스크 입출력 시에 파이프라인 I/O 방식을 사용하여 어떠한 방식보다 획기적으로 디스크 접근시간을 줄인다.

디스크 제어기는 블록 스플릿 방식을 사용한다. 이 방식은 기존의 Linux 블록 I/O 드라이버와 같이 사용할 수 있도록 구현되어 있으며, raw 블록 입출력을 가능하게 한다.

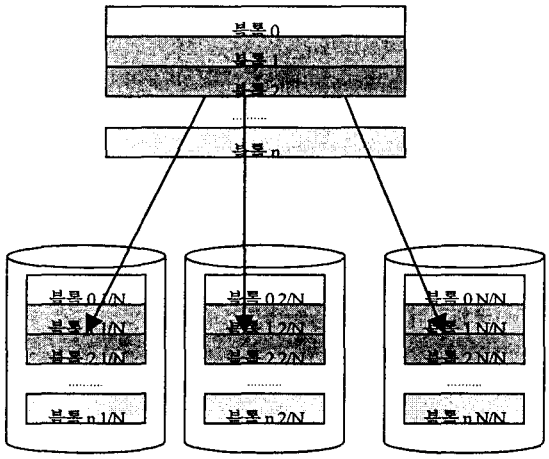

그림 5. 디스크 블록 Split 방법

그림 5는 블록 스플릿 기능을 제공하는 복수의 디스크들로 이루어진 저장 장치에서 구현된 스플릿 방법을 나타낸 것이다. 즉, 여러 개의 디스크 어레이의 모습이다. 이러한 디스크 제어에 멀티미디어 스트리밍 데이터를 저장할 때에는 파일 시스템의 할당 단위로 파일을 조각화하여 해당 크기의 블록 단위로 저장한다. 그림 5의 위의 부분은 멀티미디어 스트리밍 데이터가 이러한 블록 단위로 나누어진 논리적인 블록이다.

이러한 논리적 순서의 블록들이 블록0, 블록1, 블록2, 블록3, 블록N의 순서로 배치되었을 때, 이러한 블록들을 복수의 디스크에 나누어 저장하기 위하여, 각 블록들을 디스크 어레이에 포함된 디스크의 개수(N)로 나눈다. 예를 들어 블록0만을 설명한다. 블록0을 디스크의 개수(N)로 나누면, N개의 블록으로 쪼개지게 된다. 이렇게 쪼개진 N개의 블록 중 첫 번째 조각 블록은 첫 번째 디스크에 저장하고, 두 번째 조각 블록은 두 번째 디스크에 저장하고, N번째 조각 블록은 N번째 디스크에 저장한다. 나머지 블록들도 위와 마찬가지 방법으로 디스크 어레이에 나누어 저장한다.

디스크 제어기는 블록 스플릿 기능을 포함하여 파이프라인드 I/O 기능도 제공하고 있다. 즉 디스크 접근 명령을 파이프라인드 방식으로 연속적으로 접근할 수 있는 기능을 제공하여 대량의 멀티미디어 데이터를 빠르게 읽어 오기 위한 방법이다.

멀티미디어 스트리밍 데이터가 저장된 디스크 어레이로의 읽기 명령이 발생하면, 첫 번째 스플릿 블록 조각 읽기 명령과 두 번째 스플릿 블록 조각 읽기 명령들은 순서대로 디스크 어레이의 각 디스크에 전달한다.

그러면, 읽기 명령을 받은 디스크 어레이의 각각의 디스크들은 첫 번째 스플릿 블록 조각의 위치까지 헤드를 움직인 후 읽어 들여 내부전송하고, 상기 읽어 들인 첫 번째 스플릿 블록 조각을 디스크 어레이 제어 장치에 외부전송한다. 이와 같이, 디스크 어레이 제어 장치가 디스크 어레이에 스플릿 블록 조각 읽기 명령을 큐잉하여 전달함으로써, 모든 외부전송을 하나의 시간 축으로 보면 외부전송이 쉼 없이 일어나는 것을 알 수 있고, 그만큼 디스크 어레이 읽기가 높은 성능을 가지게 된다.

II. 스트리밍 가속장치 구현

고성능 멀티미디어 스트리밍 가속장치는 그림 6과 같이 실제로 구현되었다. 현재 스트리밍 가속장치는 모든 Linux 드라이버 및 관리 진단 기능이 구현 완료된 상태이며, 이 장치를 사용한 NGIS 지역서버를 시험하고 있다. 그림 6에 보이는 바와 같이 PCI 버스 카드 형태로 스트리밍 가속장치가 구현되었으며 내부에 512 Mbyte의 PCI 메모리를 내장하여 지역서버의 동작에 영향을 미치지 않고 내부적으로 스트리밍 데이터를 전송할 수 있는 구조를 띄고 있다.

그림 6. 스트리밍 가속장치 구현 모습

스트리밍 가속장치는 SCSI 160 Dual 채널 커넥터, 1 Gb 이더넷 포트, 내부 플래시 메모리를 프로그램 할 수 있는 포트, 시리얼 모니터 포트가 연결되어 있다.

III. 스트리밍 가속장치 성능

고성능 멀티미디어 스트리밍 가속장치의 실제 성능을 측정하기 위하여 그림 7과 같은 환경을 구성하였다.

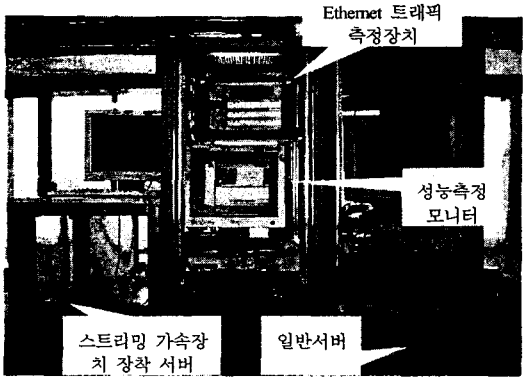

그림 7. 스트리밍 가속장치 성능 측정 환경

스트리밍 가속장치 성능을 알아보기 위해서 스트리밍 가속장치를 장착한 서버와 일반적인 서버를 비교하여 측정하였다. 일반서버는 1Gbps의 네트워크 카드가 장착되어 있는 Supermicro 마더보드에 두 개의 Intel Xeon 2.4GHz 프로세서를 장착하고 있으며, 메인 메모리는 2GB 이다. 스트리밍 서버는 이와 동일한 구성에 단지 그림 6과 같은 스트리밍 가속 장치가 내장되어 있는 서버이다.

대용량의 멀티미디어 데이터를 대상으로 시험을 진행하기 위하여 2GByte 크기의 영상 데이터 10개를 50명의 사용자가 1초 간격으로 20Mbps의 대역폭으로 연속적으로 요청하는 경우에 실제로 네트워크를 통해서 전송되고 있는 패킷을 측정하여 그래프로 작성하였다.

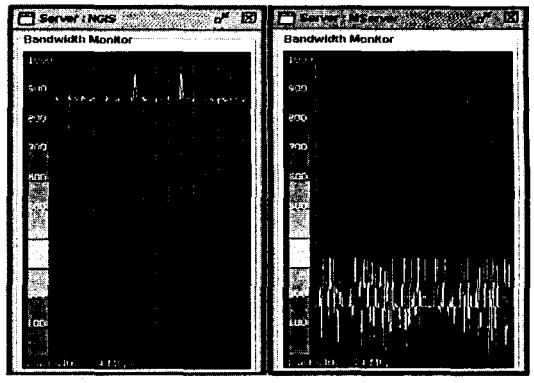

그림 8. 일반서버와 NGIS 서버의 성능측정 결과

NGIS 서버는 평균 성능이 890Mbps를 일정하게 유지하는데 반하여, 일반서버는 평균 약 190Mbps 정도의 대역폭으로 성능이 일정하지 않고 계속 변한다. 이는 UDP 패킷 처리 시에 한꺼번에 네트워크로 데이터가 몰리는 현상이 발생되어 일정한 성능을 내지 못하는 것을 알 수 있다.

결 론

차세대 인터넷 서버는 세계적으로 전례가 없는 서비스를 제공하는 것을 목표로 한국의 인터넷 환경에 맞는 서버를 개발하고 있다. 현재 하드웨어는 구현이 완료되어 목표로 하는 성능을 내고 있으며, 계속하여 어플리케이션 소프트웨어가 올리는 작업을 하고 있다. 본 논문에서는 스트리밍 가속 장치가 이러한 차세대 인터넷 서버의 요구사항에 맞게 동작하고 있음을 보여 주었다.

참고문헌

[1] 윤석한, " 5 대 국책시리즈 - 차세대 인터넷 서버 기술 개발", 한국전자통신연구원 소식지, 2002 년 6 월호, pp.64-67, 2002.

[2] 김명준, 임기욱, "차세대 인터넷 서버(SMART 서버) 기술 개발", 한국콘텐츠학회지 창간호, 2003. 06

[3] B. Boucher, Stephen E. J," Intelligent network Interfaced Device and System For Accelerated Communication", USA Patent Number : 6,427,173 B1, 2002. 6

[4] Eric Yeh, Herman Chao, Venu Mannem, Joe Gervais, Bradley Booth, "Introduction to TCP/IP Offload Engine", Version 1.0, 10 Gigabit Ethernet Alliance, April 2002

[5] Alan Earls, "Integrating TCP Offload Engines : A look at vendors and trends", TidalWire, http://www.tidalwire.com/, May 200