이중 스트라이핑을 이용한 Fault Tolerant 병렬서버의 설계와 구현

이 민 홍, 김 경 훈, *김 서 균, 남 지 승
전남대학교 컴퓨터공학과, *정보통신진흥원
전화 : 062-530-0422 / 핸드폰 : 016-209-4012

Design and Implementation of Fault Tolerant parallel server with double striping

Min Hong Lee, Kyung Hun Kim, Seo kyun Kim, Ji seung Nam
Dept. of Computer Science, Chonnam National University
E-mail : leemh@mdclab.chonnam.ac.kr

Abstract

In this paper, we present the design and implementation of a Fault Tolerant VOD system. The system uses the double striping policy that a original file is sequentially striped to multiple streaming server with constant data block size and the data block of one streaming server is re-striped to the other streaming server.

When a fault occurs in one streaming server, automatically the other streaming server will service.

I. 서론

최근 초고속 통신망의 발달로 인하여 텍스트, 이미지, 오디오, 비디오 등과 같은 다양한 매체들을 하나의 종합된 멀티미디어 응용서비스 형태로 실현하기 위해 많은 연구가 활발히 진행되고 있다. 대표적인 서비스로 주문형 비디오 (VOD : Video On Demand), 원격 화상 회의, 홈쇼핑 등을 들 수 있으며, 그 활용 범위가 매우 광범위하다. 이러한 서비스 중 인터넷상에서의 스트리밍 기술을 기반으로 한 주문형 비디오 서비스가 증가하고 있다.

VOD 서비스 시스템은 고용량이어야 하며 추가되는 새로운 저장공간에 대한 확장성, 고속의 네트워크 자원, 사용자와 자원의 분배 및 활용, 네트워크와 저장장치의 고장 등 시스템 성능과 안정을 위한 많은 부분이 고려되어 설계되어야 한다.[1]

기존 솔루션은 상대적으로 취약한 네트워크 자원을 최대로 활용하고 보다 많은 사용자를 지원하기 위한 구조 위주로 연구개발이 행하여 졌다. 그러나 현재의 멀티미디어 서비스 시스템은 이러한 시스템 요구에 대한 고려와 늘어나는 사용자 부하, 그리고 분산 저장된 데이터 저장장치의 고장에 대한 반영이 이루어지지 못하여 증가되는 사용자의 서비스 만족도를 충족시키지 못하고 있다.

이제는 서비스에 대한 사용자 만족도, 즉 지속적으로 늘어나는 사용자에 대한 끊김 없는 서비스에 대한 신뢰성을 제공하여야 하며, 고장이 발생한 상황에서도 사용자에게 지속적인 서비스를 유지시켜 전체 VOD 서비스에 대한 만족도를 향상시킬 수 있도록 Fault tolerance를 감안한 고성능 미디어 서버로 최적화 되어야 한다.

II. 시스템 구조 및 서비스 방식

2.1 시스템 구조

기존의 범용 서버를 Fault tolerance를 감안한 미디어 서버로 이용하기 위해서는 하드웨어와 미디어 전송 및

관리 SW의 두 가지의 구성 요소가 필요하다. H/W는 많은 용량과 빠른 전송 속도를 갖는 저장 장치가 가장 중요한 요소이며 충분한 데이터 전송을 위한 I/O 장치 역시 중요하다. 저장 장치의 물리적인 성능뿐 만 아니라 데이터의 효율적인 배치 및 SW를 통한 능률적 관리가 필수적인데 이것은 매우 임의적인 사용자 서비스 요구에도 불구하고 예상치 못한 다양한 변수가 존재하기 때문이다. 영화 서비스를 위한 서버의 경우 동시에 요구되는 미디어는 전체 저장 미디어의 극히 일부일 가능성이 있고, 장시간 서비스를 받는 사용자의 특성에 따라 초기지연 시간을 조정하는 등 부족한 시스템 자원을 최대로 이용할 수 있는 다양한 제어를 통하여 그 같은 상황에 맞추어 시스템 운영을 탄력적으로 할 수 있기 때문이다. 또한 새로운 형식의 미디어 파일 타입과 로컬 타입의 컨텐츠를 수정 없이 네트워크 상에서 이용하기 위하여 거기에 맞는 제어 역시 미디어 솔루션의 주요 포함 요소가 되었다.

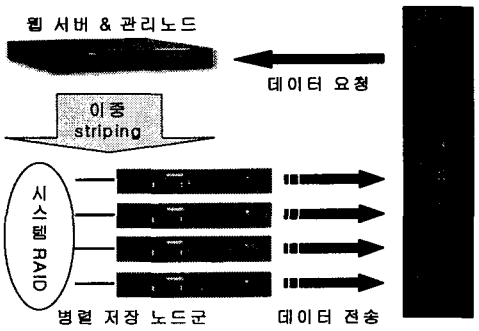

그림 1. 전체 시스템 구조

그림 1은 .전체적인 시스템 구조를 나타낸 것이다. 구현된 시스템은 기본적으로 여러 대의 서버가 동시에 하나의 클라이언트를 위해 다중의 접속 경로를 갖는 형태로 구성되었다. 미디어 타입과 관계없이 모든 파일 타입을 지원하기 위하여 기본적으로 TCP가 전송 프로토콜로 선택되었고, 미디어 특성에 맞는 데이터 흐름을 자율적으로 조정하고, 여러 서버의 제어흐름을 조정하고 관리하기 위한 제어 서버 노드가 별도로 존재한다. 이는 다른 타입의 데이터 전송 예를 들어 인터페이스용 웹 데이터, 데이터 베이스 트랜잭션 데이터 등을 미디어 전송 서버로부터 분리시키기 위한 목적으로 존재한다. 또한 사용자 인터페이스로서의 웹 서버와 미디어 관리용 데이터 베이스, 각 저장 서버에 데이터를 분산 저장하고 콘텐츠를 관리 재배치하는 관리자용 부 프로그램들로 전체 시스템이 구성된다.

위와 같은 구조로 인하여 사용자 부하를 효율적으로 각 서버 노드에 분산시키고 새로운 노드와 저장 공간 추가 시에 생기는 중복저장의 단점을 해결하는 새로운 형식의 시스템 구조를 가지게 되었다.

2.2 시스템 인터페이스 및 서비스 방식

구현된 시스템은 데이터의 제어 흐름을 조정 관리하기 위한 하나의 컨트롤 서버와 이의 제어를 받는 최소 2대 이상의 저장 노드 전송 서버들로 구성되었다.

컨트롤 서버는 사용자 인터페이스를 위한 웹 서버와 데이터베이스 서버가 설치되며 각 노드들의 부하 상태를 체크하는 프로그램, 그리고 컨트롤 노드의 데이터를 분산 저장 관리하는 프로그램들로 구성된다.

저장 서버에는 클라이언트가 요구하는 데이터를 전송하는 프로그램이 설치된다. 이러한 기능적 분배는 다른 특성을 갖는 즉, 미디어데이터와 다른 제어 메시지 데이터 전송으로 인하여 디스크와 I/O장치를 소모하는 overhead 감소를 위해 고려되었다. 클라이언트에는 다중 접속 경로를 갖는 데이터 수신 모듈이 탑재되고 이는 기존 플레이어에 수신된 데이터를 공급하는 역할을 수행한다. 전체적인 서비스 방식은 그림 2 에 나타내었다.

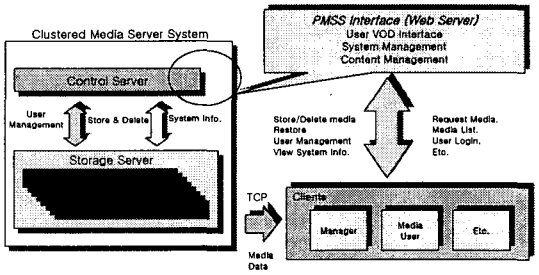

그림 2. 시스템 인터페이스 및 서비스 방식

사용자가 미디어 서비스를 이용하는 과정은 다음과 같다. 사용자는 컨트롤 서버에 접속하여 관리자가 데이터를 분산 저장하면서 자동으로 생성된 미디어 정보를 통해 서비스 요구를 결정하고 컨트롤 서버로부터 실제 데이터가 저장된 분산 서버들과 저장 스트라이핑 사이즈 정보 등 스트리밍 서비스를 위한 정보를 가진 메타 데이터를 전송 받는다. 전송된 메타 데이터를 통하여 사용자 클라이언트는 각각의 저장 서버와 연결을 설정하고 저장 정보와 분산 데이터 사이즈에 따라 적합한 버퍼링과 전송 데이터 블록을 전송 받고 이를 조합하여 미디어를 재생하게 된다. 사용자 클라이언트의 수신 모듈은 기존의 Windows Media Player의 Source 필터의 형태로 구현되어 있다. 따라서 미디어 플레이어가 지원하는 모든 표준 형식의 데이터를 지원한다.[2]

III. 시스템 관리와 Fault Tolerance

3.1 데이터 저장 관리와 시스템 확장

새로운 저장 노드가 추가됨으로써 전체 시스템은 동시 사용자 로드를 전체 시스템에 분산한다. 또한 새로운 노드에 장착된 저장 디스크는 기존 저장 데이터의 재 분산 저장을 통하여 전송 부하를 저장 노드간에 분산 할 수 있다. 극단적으로 전체 데이터이 재 분산을 통하여 전체 시스템 부하를 분산 할 수 있고 필요에 따라 앞에서 언급한 바와 같이 일부 인기 데이터에 대한 재 분산만 실시하여 로드 균형을 맞출 수도 있다. 이것은 전체적으로 시스템 가용성에 중점을 둔 시스템의 설계 원칙을 반영하는 것으로써 관리자는 충분한 시스템 상황과 서비스 상황을 모니터링하고 이와 같은 정책에 따라 시스템 환경을 구축한다. 데이터 저장 관리와 시스템 확장을 위한 웹 기반 시스템 인터페이스를 그림 3에 나타내었다.

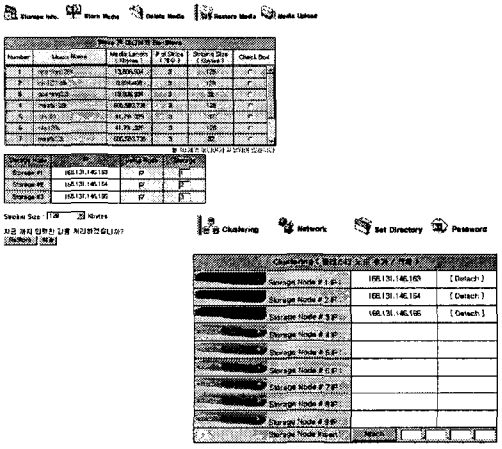

그림 3. 데이터
저장 관리와 시스템 확장

데이터의 분산 저장정책에 의하여 전체 시스템을 운용하며, 시스템 클러스터를 통하여 RAID 수준의 성능을 기대할 수 있고 거기에 시스템 확장을 추가 노드의 증설로 간단히 확장되고, 이에 대한 별도의 부하 분배기가 필요하지 않는 점등이 개발 시스템이 갖는 중요한 장점이다.

3.2 Backup

개발된 시스템은 관리자와 사용자 모두 웹 인터페이스를 통한 접근을 하는데 관리자는 미디어 데이터를 컨트롤 서버로 업로드하여야 하고 이를 각 저장 서버에 스트라이핑 정책에 따라 분산 저장하여야 한다. 이와 같은 이중 데이터 저장은 컨트롤 노드에 주요한 원시 미디어 데이터를 저장하여 여러 노드에 걸쳐 데이터를 분산 배치라는 데서 생기는 데이터 손실과 에러에 대비하는 백업 및 복구 데이터로 사용할 수 있다. 이 같은 구조 때문에 컨트롤 서버에는 전체 저장 노드 데이터 용량만큼의 저장 공간이 필요하게 되나 사용자 데이터의 중요성에 따라 이러한 백업 대상 데이터의 보관 여부를 결정할 수 있다. 시스템 오류에 대비한 백업 공간의 낭비는 어느 시스템에서나 발생하는 문제로서 단일 시스템으로 구성되었다 하다 라도 이와 같은 디스크 사용을 해결 할 수 없는 문제이다. RAID의 물리적 복구 가능한 레벨을 적용하거나 SAN 과 같은 저장장치를 채용하더라도 짧은 시간에 시스템을 복구하거나 서비스 보장을 위해서는 별도의 백업 데이터가 필요하다. 구현 시스템에도 이러한 저장 장치를 채용하면 같은 효과를 기대할 수 있으나 많은 비용을 소모함으로 효율적이지 않다.

3.3 가용성과 Fault Tolerance

병렬서버의 설계에 있어서 서버의 높은 확장성은 분명 장점이지만, 노드추가를 통한 확장은 이에 상응하는 저장노드 고장에 대한 가능성이 높아짐을 의미한다. 하나 이상의 노드에 고장이 발생했을 경우에도 지속적인 VOD 서비스를 제공하기 위해 서버는 반드시 부가적인 정보를 저장하고 있어야 한다. 이러한 부가적인 정보의 양과 배치 정책은 디스크 고장의 수와 고장나지 않은 서버들간의 Load Balancing의 관점에서 매우 중요하다.

기존 시스템의 대부분은 이중 서버로 구성되어 있으며 서비스 도중 고장이 발생한 서비스 노드에 대해 일대일 대응되는 여분의 백업 노드가 기능을 대체하는 방식을 채택하고 있다. 하지만 이러한 방식은 모든 서버를 동시에 돌려야 하는 비용, 고장 발생 전에는 서버의 절반만 서비스에 참여한다는 낮은 가용성, 원 서버와 추가 서버간의 감시에 따른 부하 발생 등으로 인한 부담이 존재한다.

본 시스템에서는 이러한 부담을 줄이고, 발생하는 고장에 대응하는 보다 나은 서비스를 제공하기 위해 다음과 같은 이중 스트라이핑 정책을 적용하였다. 이러한 이중 스트라이핑은 기존에 많이 연구 개발되었으며, 본 논문에서는 이러한 연구 개발된 부분들을 적용

하여 시스템을 설계하였다.

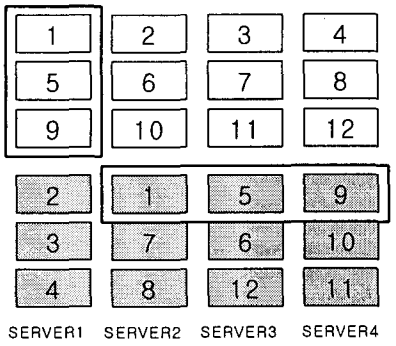

그림 4는 이중 스트라이핑 정책을 도식화한 것이다. 스트라이핑은 비디오 데이터를 몇 개의 디스크에 스트라이핑 하는 비디오 오브젝트 스트라이핑과 하나의 비디오 세그먼트 데이터를 몇 개의 디스크에 스트라이핑 하는 비디오 세그먼트 스트라이핑의 두가지 개념으로 나누어진다. 본 논문에서는 비디오 오브젝트 스트라이핑 개념과 고장이 발생하지 않은 모든 서버가 고장이 발생한 서버의 역할을 균일하게 분담하는 이중 스트라이핑 미러링 방식을 이용하여 Fault Tolerance를 보장하도록 하였다.[3,4,5,6]

비디오 데이터를 여러 개의 저장서버로 일차적으로 스트라이핑하여 분산 저장한다. 그림 4의 상위 부분으로 1번에서 12번 블록으로 스트라이핑 되었다. 하나의 서버에서 고장이 발생할 경우에 대비해 다른 서버로부터 동일한 데이터를 수신 받을 수 있도록 각각의 스트라이핑 블럭을 다시 한번 스트라이핑 하게 된다.. 그림 4의 하위 부분에 해당한다. 1번 서버에 저장된 1, 5, 9번 스트라이핑 블록은 서버2에 1번, 서버3에 5번, 서버4에 9번 블록으로 다시 스트라이핑 되는 것이다. 각각의 서버는 고장 발생 전까지 모두 서비스에 참여하게 되고, 고장 발생 시 서버는 이러한 이중 스트라이핑을 통하여 백업된 데이터를 사용자에게 제공할 수 있게 된다. 이와 같은 구조를 통하여 사용자의 입장에서는 끊김 없는 서비스를 제공받을 수 있게 되며, 서비스 운영의 입장에서는 모든 노드가 서비스 운영에 참여함으로써 각 노드에 대한 높은 가용성과 서비스 안정성을 갖추게 되는 것이다.

Ⅳ. 결론

VOD 서비스 시스템은 고용량이어야 하며 추가되는 새로운 저장공간에 대한 확장성뿐만 아니라 지속적으로 늘어나는 사용자에 대한 끊김 없는 서비스에 대한 신뢰성을 제공하여야 한다. 구현된 시스템은 기본적으로 병렬 전송과 분산 저장을 통하여 사용자 부하를 효율적으로 각 서버 노드에 분산시킴으로써 성능향상과 부하 분산을 구현하였고, 신뢰성 향상을 위해 데이터 교차 스트라이핑 정책의 적용과 관리도구들을 활용해 VOD 서버의 성능을 최대화할 수 있도록 제작되었다.

서비스에 대한 신뢰성은 이러한 미디어 서버뿐만 아니라 서버 시스템의 운영에도 무척 중요시되며, 본 시스템 역시 이와 같은 안정성 보장을 위해 다수의 서버가 하나의 클라이언트에게 공동으로 작업하는 현 서비스 형태의 보완을 위한 연구를 수행하고 있다. 보다 나은 시스템의 구성 방법과 서비스 방식에 대한 연구가 지속적으로 수행되어야 한다.

참고문헌(또는 Reference)

[1] Huang-Jen Chen T.D.C Little, "Storage allocation policies for Time-Dependent multimedia data" ,IEEE Trans. On Knowledge and Data Engineering, Vol. 8, No.5. pp. 865-864, 1996

[2] Microsoft, "Microsoft Media Service SDK", 2001

[3] S.Chandeharizadeh and S.H.Kim, "Striping in multi-disk video servers" in Proc. High-Density Data Recording and Retrieval Technologies Conference, SPIE, Oct. 1995

[4] B.Ozden et al., "Disk striping in video server environments" in Proc. of the IEEE Conf. On Multimedia System, (Hiroshima, Japan), pp. 580-589,jun,1996

[5] B.Ozden et al., "Fault-tolerant architectures for continuous media servers" in SIGMOD International Conference on Management of Data 96, pp 79-90, June 1996

[6] A. Mourad, "Doubly-striped disk mirroring : Reliable storage for video servers" Multimedia, Tools and Application, Vol.2, pp. 253-272, MAY 1996