

NDMP를 사용한 디스크 기반 백업 소프트웨어

(A Disk based Backup Software using NDMP)

윤 종 현 [†] 북 경 수 ^{**} 유 재 수 ^{***}
 (Jong Hyeon Yun) (Kyoung Soo Bok) (Jae Soo Yoo)

요 약 최근 데이터 통신 기술 및 저장 장치의 발전과 더불어 대용량의 데이터를 가공 처리하기 위한 많은 연구들이 진행되고 있다. 이와 함께 NAS와 SAN 등의 보급으로 저장 매체의 대용량화가 가속화되면서 대용량의 데이터를 효과적으로 백업하기 위한 요구가 증가되고 있다. 본 논문에서는 네트워크 기반의 백업을 위해서 제안된 개방형 표준 프로토콜인 NDMP를 이용하여 디스크 기반 백업 소프트웨어를 설계하고 구현한다. 제안하는 백업 소프트웨어는 다양한 백업 및 복구 타입을 지원한다. 또한 네트워크에 연결된 다수의 장치들을 관리하고, NDMP 프로토콜을 확장하여 디스크 백업을 지원한다.

키워드 : 백업 소프트웨어, SAN, NAS, NDMP

Abstract Recently, with the development of data communication techniques and storage devices, many researches have been done to process a large amount of data. As the size of storage media has been increased because of the storage technologies such as NAS and SAN, the efficient backup for a large amount of data is required. In this paper, we design and implement a disk based backup software that fully supports NDMP(Network Data Management Protocol) which is an open standard protocol for network based backup. The proposed backup software supports various types of backup and restore. In addition, it manages multiple network connected devices and supports a disk based backup by extending the NDMP protocol.

Key words : Backup Software, SAN, NAS, NDMP

1. 서 론

최근 산업 전 부분이 인터넷 환경으로 전환되면서 개인의 데이터는 물론 기업의 정보 데이터가 기하급수적으로 증가하고 있다. 또한, 이들 데이터를 효과적으로 관리하기 위한 데이터웨어하우스, 전사적 자원 관리, 고객 관계 관리, 지식 관리 등 첨단 엔터프라이즈 컴퓨팅 환경의 구축 및 증설이 활발히 진행되고 있다. 이로 인해 금융, 방송, 학교, 병원, 기업과 같은 업종의 데이터 저장 장치는 하루에도 수백 메가바이트에서 수십 기가바이트의 저장 공간 증설이 이루어지고 있다[1]. 이러한

기업 환경에서 임무 결정적(mission critical) 데이터가 천재지변 또는 시스템의 장애로 하루 또는 수 일간 제 공되지 않는다면, 해당 기업의 비즈니스는 큰 손실을 입을 수밖에 없다. 또한 이들 데이터에 대한 백업 데이터가 존재하지 않는다면 막대한 손실이 발생할 것이다. 따라서 방대해진 데이터를 유지, 보호하는 백업의 필요성이 더욱 중요해지고 있다[2-5].

백업의 필요성이 중요해지면서 백업 데이터를 저장, 관리하는 저장 장치 관리 기술의 중요성 역시 높아지고 있다. 백업 데이터를 저장, 관리하는 저장 장치 관리 기술로는 DAS(Direct Attached Storage), NAS(Network Attached Storage), SAN(Storage Area Network)등이 있다. 백업 데이터의 크기가 크지 않았던 과거에는 DAS만으로 사용자의 백업 요청을 충족시킬 수 있었다. 하지만 개인 사용자가 관리하는 데이터의 크기가 기가바이트 단위로 커지고, 백업 데이터의 공유 및 백업 성능 향상에 대한 요구가 증가되면서 점차 NAS와 SAN 등의 기술이 사용되고 있다[6-8].

NAS 기반 백업은 백업을 위한 전용 소프트웨어 및 백업 장치를 공유하여 실제 백업 장치가 없는 서버들도

· 본 연구는 한국과학재단 목적기초연구(특정기초연구 과제번호 R01-2003-000-10627-0) 지원 및 산업자원부의 지역혁신 인력양성사업의 연구 결과로 수행되었음

[†] 학생회원 : 충북대학교 정보통신공학과
 bluette@netdb.cbnu.ac.kr

^{**} 정 회 원 : 한국과학기술원 전산학과
 ksbok@dbserver.kaist.ac.kr

^{***} 통신회원 : 충북대학교 정보통신공학과 교수
 (corresponding author 임)

논문접수 : yjs@cbucc.chungbuk.ac.kr

심사완료 : 2005년 4월 12일

2005년 9월 5일

다른 서버의 백업 장치를 이용해 백업이 가능하다. 또한 네트워크를 이용해 백업을 수행함으로써, 실제 백업에 대한 중앙 집중화가 가능하며, 백업 관리 및 백업 장치의 효율적인 운영으로 백업에 대한 효율을 높일 수 있다. 하지만 NAS 기반 백업은 백업에 대한 관리 및 백업 장치에 대한 비용을 줄여주는 효과가 있지만, 네트워크를 이용하기 때문에 네트워크에 부하가 가중되고 애플리케이션 성능에 영향을 주기도 한다. 또한 백업 성능이 DAS 기반 백업에 비해 많이 떨어진다는 단점도 있다. 이러한 단점을 극복하고, 네트워크 백업의 장점 및 고성능의 백업을 수행하기 위해 새롭게 등장한 것이 SAN 백업이다. SAN 백업은 DAS 백업의 성능상 장점과 네트워크 백업의 중앙 집중 관리에 대한 장점을 지닌다. 또한 LAN-Free 백업과 Serverless 백업을 지원하며, SAN 상의 장비 공유를 통해 백업 장비의 활용을 높인다[4].

오늘날의 백업 솔루션은 보다 적은 비용과 시간 안에 기업의 중요한 스토리지 자산을 보호하고, 복구할 수 있는 기능을 요구하고 있다. 이러한 백업 솔루션의 성능은 백업 데이터를 저장 관리하는 백업 저장 장치에 의해 결정되며, 크게 테이프와 디스크 장치가 사용되고 있다 [9]. 테이프 장치를 이용한 백업 및 복구 작업은 저장 장치의 특성상 백업 데이터를 읽거나 쓰기 위해 순차적인 접근을 수행해야 한다. 따라서 백업 및 복구, 특히 특정 파일을 복구하는데 있어서 많은 시간이 필요하다는 단점을 지니고 있다. 이에 비해 디스크 장치는 백업 데이터의 임의 접근이 가능하여 테이프 장치에 비해 I/O속도가 빠르므로 백업 및 복구 작업의 속도를 향상시킬 수 있다. 특히 복구에 있어서 적은 양의 데이터, 즉 일부 파일만을 복구하는 경우 기기의 특성상 순차적인 접근만 가능한 테이프 기반 복구보다 매우 빠른 복구 시간을 갖는다. 디스크의 경우 테이프와 비교하여 단일한 취급 및 데이터 전송 방법을 제공한다는 점에서 저장 매체로서의 신뢰성이 높다. 또한 백업 장치를 관리하는데 있어서도 디스크 기반 백업은 테이프 기반 백업보다 쉽게 관리될 수 있다는 장점을 갖는다[10,11].

한편 IntelliGuard Software사와 Network Appliance사에서 제안한 개방형 표준 프로토콜인 NDMP(Network Data Management Protocol)는 다른 호스트에서 운영되는 관리 소프트웨어와 네트워크 파일 서버 또는 파일 서버로 동작하는 네트워크에 연결된 스토리지 장치간의 plug-and-play 백업을 가능하게 한다[12, 13]. NDMP는 사용자들이 범용 운영체제 없이도 네트워크에 연결된 저장 장치 또는 네트워크 파일 시스템 서버의 데이터를 백업할 수 있도록 한다. 게다가 백업된 데이터를 범용 서버를 거치지 않고 파일 서버의 디스크에서 테이프로 직접 이동하게 함으로써 신속한 백업이 가능하다.

메릴랜드 대학에서는 대용량 테이프 장치에 대한 백업을 수행하기 위해 AMANDA(Advanced Maryland Automated Network Disk Archiver)라는 공개용 백업 소프트웨어를 개발했다[14]. AMANDA는 네트워크에 존재하는 다양한 시스템에서 효과적인 백업이 가능하도록 설계된 백업 소프트웨어이다. 그러나 AMANDA는 현재 표준 프로토콜로 제안된 NDMP를 지원하지 못하며, SAN 환경에도 적합하지 못하다는 문제점이 있다. AMANDA 이외에도 다수의 상용 백업 소프트웨어들이 개발되었다. 하지만 대부분의 제품들이 외국 기업에 의해 개발된 제품이며, 이들 제품들의 상당수가 NAS 기반의 백업을 위해서만 NDMP를 지원하거나 별도의 옵션 형태로 NDMP 기능을 제공하고 있다. 또한 대부분의 제품이 테이프 위주의 백업 환경에서 설계되어 디스크 기반의 백업을 원활하게 지원하지 못한다.

따라서 본 논문에서는 테이프 위주의 백업 환경을 제공하는 NDMP를 확장시켜 디스크 기반의 백업을 지원할 수 있도록 백업 소프트웨어를 설계하고 구현한다. 본 논문에서 제안하는 백업 소프트웨어는 이 기종 간의 백업을 위해 NDMP를 지원하며, 기존 NDMP가 제공하는 테이프 서비스를 확장하여 테이프 및 디스크에 대한 백업 및 복구 기능을 모두 지원한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 분야 및 기존에 개발된 백업 소프트웨어의 특징에 대해 기술하고, 3장에서는 제안하는 백업 소프트웨어의 구조에 대해서 크게 시스템 구조와 각 모듈별 특징으로 나누어 설명한다. 4장에서는 제안하는 백업 소프트웨어의 백업 및 복구 처리 과정을 설명한다. 5장에서는 제안하는 백업 소프트웨어의 구현 내용 및 성능 평가와 타 상용 백업 소프트웨어와 비교 분석한 내용을 기술한다. 마지막으로 6장에서는 결론에 대해서 기술한다.

2. 관련 연구

2.1 백업 유형

백업은 전산장비의 고장이나 다른 불의의 사고에 대비하여 파일 또는 데이터베이스를 복사해 두는 행위를 말한다. 데이터의 백업은 대형 컴퓨터를 운영하는 대규모 사업체는 물론 개인 컴퓨터에서도 필수적이고 일상적인 업무이다. 백업 유형은 표 1과 같이 백업의 형태, 기능, 백업 시스템의 구조 등에 따라서 백업 방식을 구분할 수 있다[3,4].

2.2 NDMP

IntelliGuard Software사와 Network Appliance사에서 제안한 개방형 표준 프로토콜인 NDMP(Network Data Management Protocol)는 NAS 기반의 이 기종 간 백업을 위한 프로토콜이다[12]. NDMP는 백업과 복

표 1 백업 방식의 구분

구분	백업 방식	
형태	전체백업	백업을 수행할 시스템에 존재하는 모든 파일 백업
	부분 백업	이전의 백업 데이터와 비교하여 변경된 파일만 백업
대상	파일 기반 백업	파일 구조를 파악하여 파일과 디렉토리 구조를 백업
	장치 기반 백업	파일 구조를 무시하고 디스크블록을 백업
기능	온라인 백업	작업 수행 중에도 백업 지원
	오프라인 백업	백업을 위해 일반적인 작업을 중지
시스템 구조	S/W 구조	중앙집중식백업 지역관리백업
	H/W 구조	LAN 백업 LAN-Free 백업 Serverless 백업

구 과정 동안 데이터를 전송할 수 있는 인터페이스 및 데이터 포맷을 정의하고, 장치와 플랫폼에 관계없이 표준화된 통신 채널을 사용하여 백업 장치로 데이터를 송수신할 수 있다[13].

NDMP는 클라이언트/서버 구조로 구성되며, NDMP 클라이언트에 해당하는 DMA(Data Management Application)와 NDMP 서버에 해당하는 DSP(Data Service Provider)가 있다. DMA는 NDMP 세션(session)을 생성하고 관리하며, DSP가 제공하는 NDMP 데이터 서비스와 테이프 서비스 간의 데이터 전송을 관리한다. 또한 DMA는 백업 및 복구를 위한 NDMP의 정책, 클라이언트 인터페이스 그리고 DSP에서 발생하는 여러 예외 상황을 관리한다. DSP는 백업 및 복구를 수행하기 위한 테이프 서비스(tape service), 데이터 서비스(data service), SCSI 창구 서비스(Small Computer System Interface pass through service) 그리고 전송 서비스(translation service) 등을 제공한다. 데이터 서비스는 원본 데이터 저장 장치에 대한 NDMP 인터페이스를 제공하며, 테이프 서비스는 백업 데이터 저장 장치에 대한 NDMP 인터페이스를 제공한다. SCSI 창구 서비스는 DMA가 SCSI 장치를 사용할 수 있는 기능을 제공한다. 전송 서비스는 DSP간의 데이터 전송 서비스를 제공한다.

NDMP에서 백업과 복구를 수행하려면 DMA와 DSP 간에 NDMP 세션을 설정해야 한다. NDMP 세션이란 데이터 송신자와 수신자들 사이의 데이터 흐름의 집합으로, DSP의 데이터 서비스와 테이프 서비스들 사이의 데이터 연결(data connection)과 DMA와 DSP가 제공하는 서비스들 사이의 제어 연결(control connection)을 의미한다. 그림 1은 NDMP의 구조를 보여준다. DMA는 백업 서버가 있는 호스트에 위치하며, DSP의 데이터 서비스와 테이프 서비스는 서로 다른 호스트에서 제

공되거나 같은 호스트에서 제공할 수 있다. DMA는 제어 연결을 통해 각 DSP를 제어하고, DSP의 데이터 서비스와 테이프 서비스는 데이터 연결을 통해 데이터를 주고받는다.

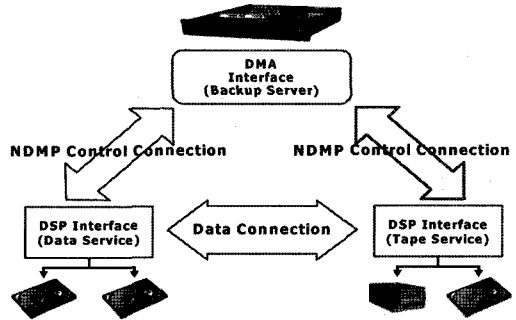


그림 1 NDMP의 구조

2.3 상용화된 백업 소프트웨어의 특징

NDMP는 NAS 시스템에서의 백업을 위해 제안되었으나 버전 5에 이르러 SAN에 대한 지원이 가능하도록 확장되었다. 현재 Veritas사의 NetBackup, Legato사의 NetWorker, HP의 OmniBack-II 등이 NDMP를 지원하고 있다. Veritas사에서 개발된 NetBackup은 분산 클라이언트/서버 환경에서 대규모 서버에 이르기까지 네트워크 환경에서 자동화된 백업 시스템을 구축할 수 있는 소프트웨어이다. 다양한 운영체제 및 데이터베이스 옵션을 통해 백업 및 복구 솔루션을 제공한다. 또한 미디어 관리 기능을 통해 라이브러리 공유 및 미디어 관리를 수행할 수 있다. NetBackup은 디스크에서 테이프 로 직접 데이터 쓰기를 수행하여 서버의 부하와 백업시간을 줄이는 Serverless 백업을 지원한다.

Legato사에서 개발한 NetWorker는 DAS, NAS, SAN 환경에서 다양한 운영체제를 탑재한 서버와 일반 데스크탑 PC의 데이터 보호가 가능하다. NetWorker는 온라인 백업, 무인 자동백업을 지원하며, 백업 데이터의 효과적인 관리 및 처리 기능을 제공한다. 또한 데이터베이스와 애플리케이션 서비스 모듈을 추가함으로써 보다 명확한 애플리케이션 보호 기능을 제공한다. NetWorker는 기존의 네트워크를 거치지 않고 파이버(Fiber) 채널을 경유하여 저장 장치에서 테이프 로 직접 백업할 수 있는 Serverless 백업과 LAN-Free 백업 등을 완벽하게 지원한다.

HP OpenView OmniBack-II는 이 기종 간 플랫폼들이 혼재해 있는 환경에서 수집된 데이터를 스토리지 영역 네트워크를 통해 관리할 수 있는 백업 소프트웨어이다. OmniBack-II 소프트웨어는 다수의 이 기종 간 호

스트에서 SAN 및 NAS 환경에 관계없이 백업 및 복구 과정을 자동화하고 모니터링 할 수 있는 단일 인터페이스를 제공한다. 데이터베이스 및 애플리케이션 서비스의 가용성 보장을 위해 온라인 백업을 사용하여 백업으로 인한 애플리케이션 종료 및 애플리케이션의 서비스 성능 저하를 방지한다. 또한 고속 백업 환경 지원을 위해 SAN 상에 위치한 라이브러리 및 테이프 드라이브의 공유를 통한 백업 환경을 제공하고, NDMP 지원을 통한 NAS 장비의 로컬 백업 등을 지원한다.

3. 제안하는 백업 소프트웨어

본 논문에서 제안하는 백업 소프트웨어는 NDMP의 기능을 확장하여 NAS와 SAN 환경에서 테이프 기반 백업 뿐만 아니라 디스크 기반 백업이 가능하도록 설계한다. 네트워크상에서 로컬 및 원격에 위치한 데이터의 백업이 가능하다. 또한, 여러 사용자들의 다양한 백업 요구를 스케줄링 하여 자동적으로 백업 및 복구를 처리하는 기능을 지원한다. 제안하는 백업 소프트웨어는 하나의 백업 서버가 백업 및 복구를 중앙 집중적으로 관리하기 때문에 일관된 정책을 유지할 수 있고 수행 중인 작업을 효과적으로 제어한다. 다양한 유형의 백업 및 복구 타입을 지원하며, 백업 파일들에 대한 생명주기(life cycle) 기능, 스크립트 기능, 백업 장치 관리 기능을 지원한다.

제안하는 백업 소프트웨어는 로컬 및 원격에 위치한 데이터에 대해 작업을 요청하는 백업 클라이언트, 백업 관련 메타 정보를 중앙 집중식으로 관리하는 백업 서버, 백업 서버 또는 LAN 상에 부착된 백업 장치를 관리하는 백업 장치 서버로 구성된다. 또한 NDMP 프로토콜을 구현한 DMA 모듈과 DSP 모듈을 각각 백업 서버와 백업 장치 서버에 부착하여 NAS 및 SAN 기반 백업을 지원

한다. NAS 백업의 경우 NAS 장치를 관리하는 백업 장치 서버를 두고, 백업 장치 서버에 NDMP의 DSP 모듈을 부착하여 로컬 및 원격에서도 백업이 가능하다. 또한 NDMP에서 제공하는 LAN-Free 및 Serverless 기능을 구현하여 SAN 환경에서도 효과적인 백업이 가능하다.

3.1 시스템 구조

그림 2는 본 논문에서 제안하는 백업 소프트웨어의 시스템 구조를 나타낸 것이다. 제안하는 백업 소프트웨어는 크게 백업 클라이언트(backup client), 백업 서버(backup server), 그리고 백업 장치 서버(backup device server)로 구성되어 있다. 백업 클라이언트는 작업을 요청하는 사용자로서 사용자의 로컬 컴퓨터에 저장된 데이터뿐만 아니라, 원격지에 존재하는 데이터에 대해서도 백업 및 복구 작업 요청을 수행할 수 있다. 백업 서버는 서버 모듈 혹은 백업 장치 서버의 작업 요청을 수신하고 처리하는 데몬 모듈, 클라이언트의 작업 요청을 수신하여 처리하는 서버 모듈, 그리고 NDMP에 따라 백업 장치 서버에 위치한 NDMP 데몬인 DSP를 관리하는 DMA 모듈로 구성되어 있다. 백업 장치 서버는 백업 장치를 논리적 또는 물리적으로 관리한다. 백업 장치 서버는 DSP 모듈과 백업 서버와 작업 메시지를 송수신하는 데몬으로 구성된다. 제안하는 백업 소프트웨어에서 사용자는 백업 클라이언트 인터페이스를 통해 작업 요청을 백업 서버로 전송한다. 백업 서버는 백업 클라이언트의 작업 요청을 수신하고 분석한다. 분석된 정보에 따라 작업을 처리할 서버 모듈이 결정되고, 백업 클라이언트의 요청을 처리할 서버 모듈이 호출되어 요청된 작업을 처리하거나, 백업 장치 서버에 적절한 명령을 전송하여 작업을 처리한다. 이후 작업의 처리 결과를 백업 클라이언트에게 반환한다.

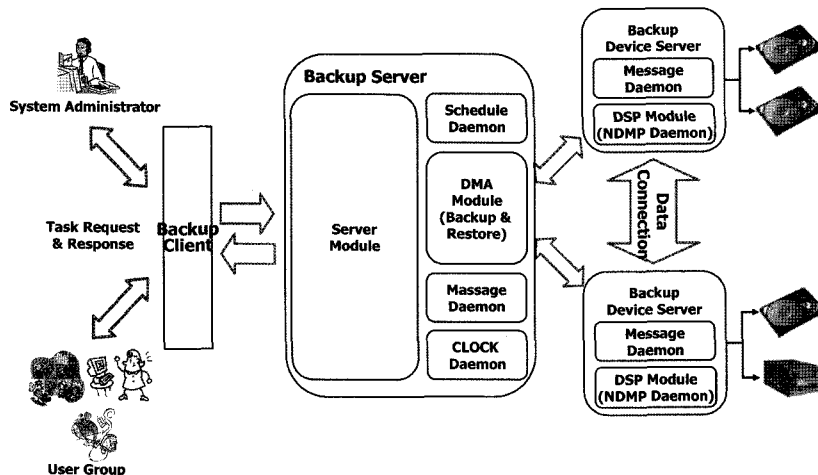


그림 2 제안하는 백업 소프트웨어의 구조

3.2 백업 클라이언트

백업 클라이언트는 사용자에게 작업 명령을 입력받고, 백업 서버와 필요한 정보를 송수신하는 역할을 수행한다. 백업 클라이언트는 C언어를 이용하여 CLI(Command Line Interface) 방식으로 구현하였다. 백업 클라이언트는 TCP/IP 소켓을 사용하여 백업 서버에 작업 명령을 전달하며, 필요에 따라 백업 서버로부터 알림 메시지와 작업에 대한 상태 정보를 받는다. 백업 클라이언트가 백업 또는 복구를 수행하려면, 백업 서버에 접속하여 로그인 과정을 통해 사용자 권한에 대한 인증을 받아야 한다. 로그인 과정에서 클라이언트의 권한 레벨이 부여되며, 클라이언트가 사용할 수 있는 장치 목록이 설정된다. 로그인 과정이 완료되면 클라이언트는 실제적인 백업과 복구에 대한 명령을 서버로 전송한다.

그림 3은 백업 클라이언트의 명령 구조를 나타낸 것이다. 각 명령어는 백업 서버의 서버 모듈에 의해 처리된다. 백업 클라이언트의 각 명령어는 실행 권한이 부여되며, 실행 권한을 갖지 못한 사용자는 해당 명령어를 실행할 수 없다. 백업 클라이언트가 지원하는 명령어의 기능은 표 2와 같다.

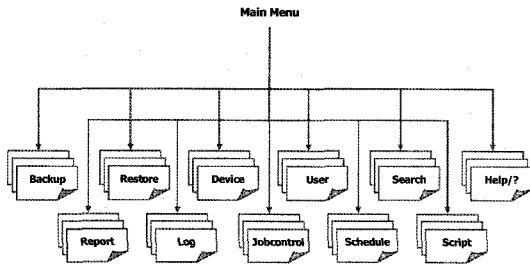


그림 220 백업 클라이언트의 명령어 구조

표 2 백업 클라이언트가 지원하는 명령어

명령어	기능
backup	예약 및 즉시 백업 기능 제공
restore	예약 및 즉시 복구 기능 제공
device	백업 장치(미디어, 디스크) 관리 기능 제공
search	백업된 데이터 검색 기능 제공
report	백업 및 복구 작업의 리포트 기능 제공
user	백업 소프트웨어 사용자 추가 및 관리 기능 제공
jobcontrl	백업 및 복구 작업의 상태(일시정지, 강제종료...)를 관리하는 기능 제공
schedule	예약 백업을 위한 작업 스케줄 정보 작성 및 관리 기능 제공
script	스크립트 작성 및 관리 기능 제공
log	백업 및 복구 로그 관리 기능 제공
help	명령어별 도움말 기능 제공

3.3 백업 서버

백업 서버는 백업 클라이언트의 작업 요청을 분석하고 처리한다. 클라이언트의 작업 요청을 처리하기 위한 서버 모듈을 수행시키고, 백업 장치 서버에 적절한 작업 명령을 전송하여 백업 및 복구 작업을 처리한다. 그림 4는 제안하는 백업 소프트웨어의 백업 서버 구조를 나타낸 것이다. 제안하는 백업 소프트웨어의 백업 서버는 클럭 데몬, 스케줄 데몬, 메시지 데몬으로 구성된 데몬 모듈과 NDMP를 구현한 DMA 모듈, 그리고 클라이언트 인터페이스 및 각 작업 관리자들로 구성된 서버 모듈을 갖는다. 서버 모듈은 백업 클라이언트의 작업 요청을 분석하고 처리한다. 서버 모듈의 구성과 기능은 표 3과 같다.

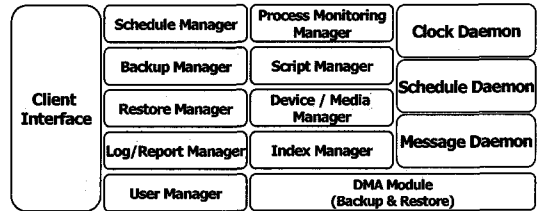


그림 4 제안하는 백업 소프트웨어의 백업 서버 구조

표 3 제안하는 백업 소프트웨어의 서버 모듈 구성

서버 모듈 이름	기능
Client Interface	사용자의 작업 요청을 입력받아 각 서버 모듈에 전달하고, 작업 수행 결과를 사용자에게 전달하는 기능 제공
User Manager	백업 소프트웨어를 사용하는 사용자의 계정 정보 관리(관리자 기능) 기능 제공
Backup Manager	백업 클라이언트의 자동/수동 백업 요청 처리 기능 제공
Restore Manager	백업 클라이언트의 전체/부분 복구 요청 처리 기능 제공
Schedule Manager	백업 작업을 스케줄링 하는 스케줄 파일 관리 기능 제공
Index Manager	백업된 파일 정보(파일 히스토리) 관리와 복구 수행을 위해 백업된 내용의 검색 및 복구 목록 작성 기능 제공
Process Monitoring Manager	수행중인 작업 프로세스의 상태를 관리하고 완료되거나 중지된 작업 프로세스 관리 기능 제공
Device / Media Manager	백업 및 복구를 위한 장치 및 미디어 관리 기능 제공
Log / Report Manager	작업 수행 과정에서 생성된 작업 정보 관리 및 작업 정보의 통계 보고서 작성 기능 제공
Script Manager	백업 및 복구 작업의 전, 후처리 작업을 수행하는 스크립트 관리 기능 제공

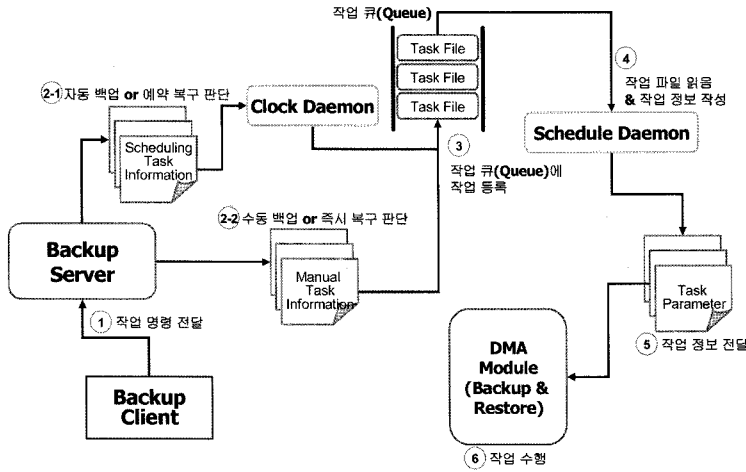


그림 5 클럭 데몬과 스케줄 데몬의 동작 과정

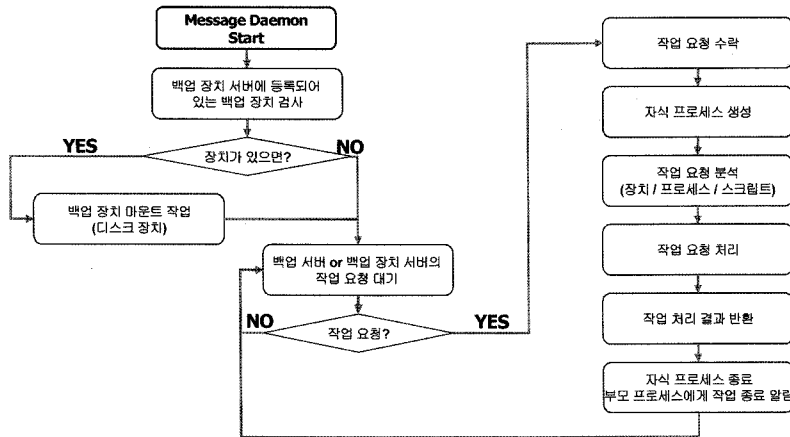


그림 6 메시지 데몬의 동작 과정

3.3.1 데몬 모듈

클럭 데몬은 백업 서버에서 동작하는 데몬으로 백업 클라이언트의 자동 백업 및 예약 복구 요청을 주기적으로 검사하고, 스케줄 데몬에게 작업 요청을 전달하는 역할을 수행한다. 클럭 데몬은 지정된 특정 주기마다 각 사용자의 스케줄 파일을 검사하고, 현재 시점에 수행해야 할 작업 요청을 스케줄 데몬에게 전달한다. 스케줄 데몬은 클럭 데몬과 마찬가지로 백업 서버에서 동작하는 데몬이다. 스케줄 데몬은 클럭 데몬 또는 백업 서버의 백업 및 복구 관리자가 직접 전달한 작업 요청을 수행하기 위해 필요한 백업 및 복구 파라미터를 생성하여 백업 서버의 DMA 모듈에게 전달하는 역할을 수행한다. 그림 5는 백업 서버의 클럭 데몬과 스케줄 데몬의 작업 수행 과정을 나타낸 것이다. 클럭 데몬은 서버 측에서 정해진 시간마다 스케줄 파일을 검사하여 수행해야 할

작업(자동 백업 또는 예약 복구) 요청에 대한 정보를 메시지 큐에 삽입한다. 만약 클라이언트가 즉시 실행해야 할 작업(수동 백업 또는 즉시 복구)을 요청하는 경우에는 클럭 데몬을 거치지 않고 메시지 큐에 직접 삽입된다. 스케줄 데몬은 메시지 큐를 검사하여 DMA에 전달할 작업 파라미터를 생성한다. 스케줄 데몬은 작성된 작업 파라미터를 DMA에게 전달하여 실제적인 작업을 수행하도록 한다.

메시지 데몬은 백업 서버 및 백업 장치 서버에서 동작하는 데몬이다. 메시지 데몬은 백업 장치 서버에 등록되어 있는 백업 데이터 저장 장치를 관리하고, 사용자가 설정한 작업 관련 스크립트를 처리하는 기능을 담당한다. 또한 백업 장치 서버에서 수행되는 작업 프로세스를 관리한다. 그림 6은 메시지 데몬의 동작 과정을 나타낸 것이다. 백업 서버의 장치 및 미디어 관리자에 의해 등

특된 백업 장치 정보는 백업 서버와 백업 장치 서버의 메시지 데몬 간 통신을 통해 전송된다. 백업 장치 서버의 메시지 데몬이 수행되면, 먼저 백업 서버의 메시지 데몬과 통신하여 마운트(mount)시킬 디스크 기반 백업 장치가 있는지 확인한다. 마운트 할 장치가 있는 경우 마운트 작업을 수행한다. 마운트 작업이 완료되면 메시지 데몬은 백업 서버의 요청에 따라 디스크 기반 백업 장치 관리, 스크립트 실행, 실제 백업 및 복구 작업을 수행하는 DSP 프로세스의 관리 등을 수행하게 된다.

3.3.2 DMA 및 DSP 모듈

DMA 모듈은 NDMP의 NDMP 클라이언트 부분을 구현한 것으로, 백업 및 복구를 위한 NDMP의 정책을 구현하고 DSP에서 발생하는 여러 예외 상황을 관리한다.

그림 7은 DMA의 동작 과정을 나타낸 것이다. DMA 모듈은 스케줄 데몬으로부터 넘겨받은 작업 파라미터를 분석하고, 작업을 수행할 백업 장치 서버를 선택한다. 또한 작업 파라미터의 백업 정보를 분석하여 DSP에게 전달할 작업 명령을 작성한다. 작업 명령이 작성되면 DMA 모듈은 작업이 수행될 백업 장치 서버의 DSP와 NDMP 세션을 생성한다. 로컬에서 수행되는 작업의 경우 하나의 DSP가 데이터 서비스와 테이프 서비스를 제공하므로 한 번의 연결만 수행한다. 원격에서 수행되는 작업의 경우 원본 데이터가 있는 백업 장치 서버의 DSP가 데이터 서비스를 제공하고, 백업된 데이터를 저장할 백업 장치 서버의 DSP가 테이프 서비스를 제공하므로 두 번의 연결을 수행한다.

NDMP 세션 설정 과정에서는 백업 장치 서버에 대한 접근 권한 인증 과정이 수행된다. 인증 과정이 성공적으로 수행되고, NDMP 세션이 설정되면 DMA 모듈은 테이프 서비스를 제공하는 백업 장치 서버에 대해 작업 준비 명령을 전송한다. 백업 작업의 경우 백업 정보에 명시되어 있는 백업 장치 서버 정보를 가지고 사용 가능한 백업 저장소를 검색하여 백업 파일을 작성한다. 복구 작업의 경우 사용자가 선택한 백업 파일을 가지고 있는 백업 저장소의 백업 데이터 파일을 선택한다. 테이프 서비스를 제공하는 백업 장치 서버에서 준비 작업이 모두 완료되면 DMA 모듈은 데이터 서비스를 제공하는 DSP와 테이프 서비스를 제공하는 DSP와의 데이터 전송 인터페이스를 설정한다. 데이터 전송 인터페이스가 설정되면 DMA 모듈은 백업 및 복구 작업의 진행을 데이터 서비스를 제공하는 DSP 모듈에게 명령한다. DSP가 제공하는 데이터 전송 인터페이스를 통해 데이터가 이동하며, 이때 데이터의 이동은 백업 서버의 DMA를 거치지 않고 수행된다. 작업 진행 과정에서 발생하는 예외나 상태 정보만 DMA 모듈에게 전송되며, DSP가 전송한 정보에 따라 DMA 모듈은 적절한 DSP 제어 작업을 수행한다. 또한 백업 서버의 모니터링 관리자를 이용하여 DSP에서 수행 중인 작업을 제어할 수도 있다. DSP는 작업이 완료되거나 진행 중에 실패하면 작업 상태 정보를 DMA에게 전달하여 작업이 종료되었음을 알린다. DSP가 전송한 작업 종료 상태에 따라 DMA는 DSP와 연결된 세션을 종료하고, 작업의 성공, 실패를

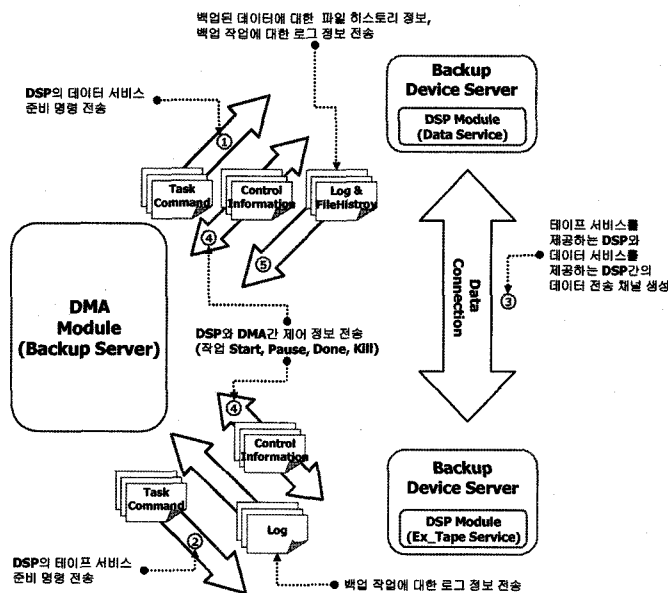
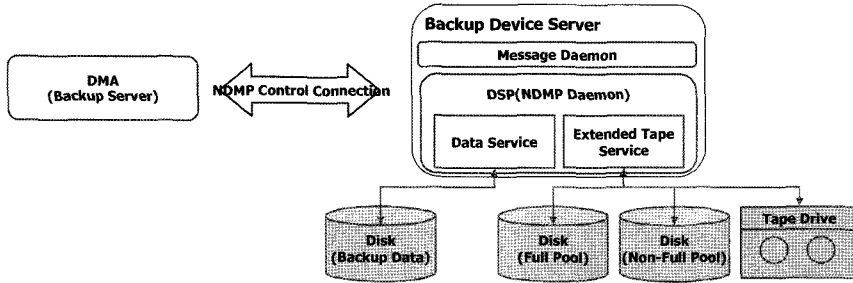
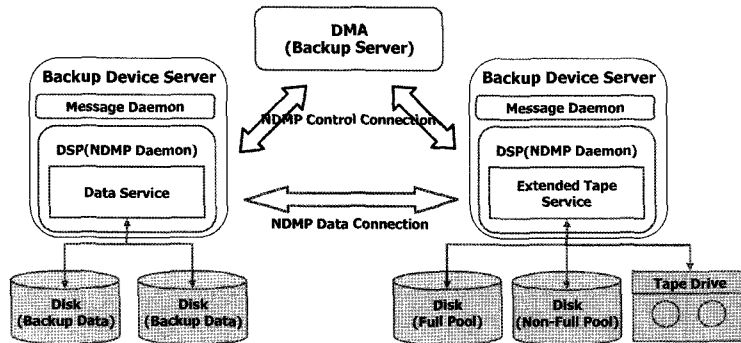


그림 7 DMA의 작업 요청 처리 과정



(a) Local 환경에서 DSP의 기능



(b) Remote 환경에서 DSP의 기능

그림 8 백업 장치 서버의 구조

백업 클라이언트에게 전송한다.

3.4 백업 장치 서버

백업 장치 서버는 백업 서버의 작업 요청을 수신하고 처리하는 메시지 데몬과 NDMP 데몬인 DSP로 구성된다. DSP는 NDMP의 서버 부분을 구현한 것으로, DMA가 명령한 작업을 수행하기 위해 데이터 서비스와 테이프 서비스를 제공하는 데몬이다. 데이터 서비스는 백업 과정에서 원본 데이터를 전송하고, 복구 과정에서 백업 데이터를 지정된 위치로 복구하는 기능을 제공한다. 테이프 서비스는 백업 과정에서 백업 데이터를 기록하고, 복구 과정에서 백업 데이터를 읽는 기능을 제공한다. NDMP에서 제공하는 테이프 서비스는 테이프 장치만을 백업 데이터 저장 장치로 사용한다. 제안하는 백업 소프트웨어는 NDMP가 제공하는 테이프 서비스를 확장시켜 디스크를 백업 데이터 저장 장치로 사용할 수 있도록 하고 있다. 본 논문에서는 이를 확장 테이프 서비스(extended tape service)라고 부른다.

그림 8은 백업 장치 서버의 구조를 나타낸 것이다. 백업 클라이언트가 설정한 파라미터에 따라 백업 장치 서버의 DSP는 데이터 서비스와 확장 테이프 서비스를 제공한다. 원본 데이터가 위치한 저장 장치와 백업 데이터를 저장할 저장 장치가 하나의 백업 장치 서버에 위치하면, DSP는 데이터 서비스와 확장 테이프 서비스를

동시에 제공한다. 원본 데이터가 위치한 저장 장치와 백업 데이터를 저장할 저장 장치가 서로 다른 백업 장치 서버에 위치하면, 원본 데이터가 있는 백업 장치 서버의 DSP는 데이터 서비스를, 백업 데이터를 저장할 저장 장치가 있는 백업 장치 서버의 DSP는 확장 테이프 서비스를 제공한다.

백업 장치 서버의 백업 저장 장치는 디스크와 테이프 장치로 구성된다. 제안하는 백업 소프트웨어는 백업 장치 서버에 존재하는 백업 디스크 장치 및 미디어를 관리할 수 있는 장치 및 미디어 관리 기능을 제공하고 있다. 장치 및 미디어 관리 기능은 백업용 장치를 효과적으로 사용할 수 있게 하고, 백업 과정에서 발생할 수 있는 백업 저장소의 저장 공간 부족으로 인한 에러를 줄일 수 있게 한다. 이를 위해 관리자는 각 사용자 별 백업용 장치 및 풀(pool)을 할당할 수 있게 한다[9]. 풀은 다수의 디스크 장치를 논리적으로 하나의 단위처럼 묶어서 관리하는 것이다. 관리자는 이러한 장치 및 미디어 관리 기능을 사용하여 사용자에게 자신이 관리하는 시스템의 데이터가 백업될 저장 공간을 미리 확보하도록 하고, 이를 사용자 스스로 관리하게 하여 백업 수행에 대한 신뢰성을 높일 수 있다.

그림 9는 제안하는 백업 소프트웨어의 백업 풀 할당 알고리즘을 나타낸 것이다. 제안하는 백업 소프트웨어에

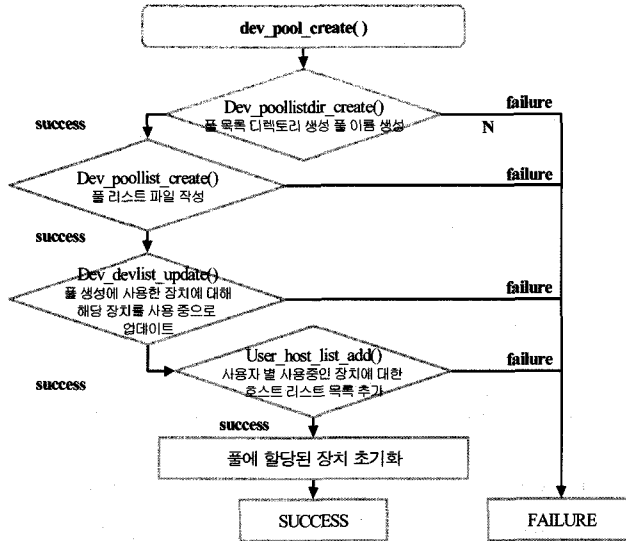


그림 9 백업 풀(pool) 할당 알고리즘

서는 미리 등록된 백업용 디스크 장치 목록 중에서 풀로 등록할 장치를 선택한다. 풀로 등록할 장치를 선택하면, 선택한 장치에 대한 풀 정보를 입력하여 풀 등록 정보를 생성한다. 또한 풀로 등록한 장치는 사용 중인 장치임을 표시하도록 디스크 장치 목록의 정보를 갱신한다. 마지막으로 생성된 풀을 초기화시킨다.

백업 장치 서버의 메시지 데몬은 백업 장치 서버의 장치 및 미디어를 관리하고 프로세스 관리 및 스크립트를 처리하기 위한 목적으로 사용된다. 백업 서버는 백업 장치 서버의 메시지 데몬과 지정된 작업 메시지를 사용하여 백업 클라이언트가 요청한 장치 및 미디어 관리, 스크립트 처리, 프로세스 관리 작업을 처리하도록 명령한다. 백업 장치 서버의 메시지 데몬은 백업 서버가 명령한 작업을 처리하고, 그 결과를 백업 서버에게 반환한다.

4. 백업 및 복구 처리

제안하는 백업 소프트웨어는 자동 백업과 수동 백업, 예약 복구와 즉시 복구를 지원한다. 자동 백업은 사용자가 백업 서버의 스케줄 관리자를 통해 작성한 백업 스케줄 정보를 클럭 데몬이 검사하여 스케줄 정보의 지정된 시간에 수행되는 백업이다. 수동 백업은 사용자가 요청하는 즉시 수행되는 백업이다. 예약 복구는 사용자가 백업 서버의 복구 관리자를 통해 작성한 복구 작업 수행 시간을 클럭 데몬이 검사하여 지정된 시간에 수행되는 복구이다. 즉시 복구는 사용자가 요청하는 즉시 수행되는 복구이다. 제안하는 백업 및 복구 작업은 NDMP를 따라 설계된 백업 서버의 DMA와 백업 장치 서버의 DSP간 통신을 통해 처리된다.

4.1 백업

제안하는 백업 소프트웨어는 자동 백업과 수동 백업을 지원한다. 또한 다양한 백업 유형을 제공하며 로컬 및 원격에 있는 데이터를 백업할 수 있는 기능을 제공한다. 제안하는 백업 소프트웨어의 백업은 백업 클라이언트의 작업 요청과 백업 서버의 서버 모듈인 백업 관리자 및 각 데몬, 그리고 NDMP를 따라 설계된 백업 서버의 DMA 모듈과 백업 장치 서버의 DSP간 데이터 전송을 통해 수행된다.

그림 10은 제안하는 백업 소프트웨어의 백업 처리 구조를 나타낸 것이다. 사용자는 백업 클라이언트를 통해 백업 서버에 백업 작업을 요청한다. 사용자의 백업 요청 정보는 백업 서버의 백업 관리자에 의해 분석된다. 사용자의 백업 요청은 스케줄 관리자가 작성한 백업 스케줄에 의해 수행되는 자동 백업과 백업 스케줄을 사용하지 않고 사용자가 요청한 즉시 수행되는 수동 백업으로 구분된다. 자동 백업의 경우 사용자가 요청한 백업 작업 정보는 클럭 데몬을 통해 스케줄 데몬에게 전달된다. 수동 백업의 경우 백업 작업 정보는 클럭 데몬을 거치지 않고 바로 스케줄 데몬에게 전달된다. 스케줄 데몬은 사용자가 요청한 작업 정보를 분석하여 백업 서버의 DMA 모듈에게 전달할 작업 파라미터를 작성한다. DMA 모듈은 스케줄 데몬으로부터 전달받은 파라미터를 분석하여 백업 사용자가 설정한 백업 장치 서버의 DSP와 연결을 수행하고, NDMP에 따라 백업이 수행되도록 한다. 이후 DMA 모듈은 NDMP의 메시지와 여러 관리 정책에 따라 DSP를 관리하고 제어한다.

그림 11은 백업 과정에서 호출되는 함수의 흐름을 나

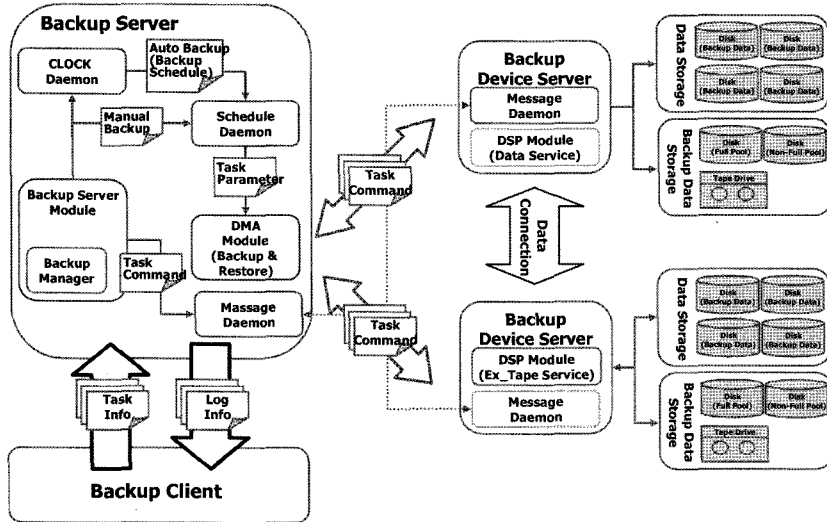


그림 10 제안하는 백업 소프트웨어의 백업 처리 구조

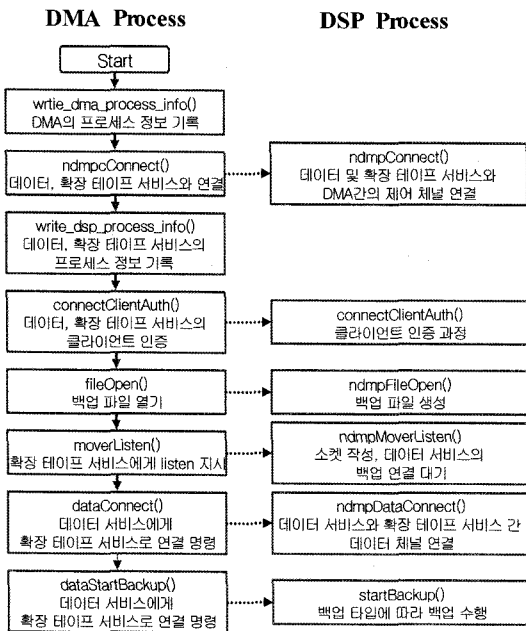


그림 11 백업 관련 함수 호출 과정

타낸 것이다. 백업을 하기 위해 DMA 프로세스는 DSP 프로세스의 데이터 서비스와 확장 테이프 서비스에 연결을 수행한다. 확장 테이프 서비스는 NDMP의 DSP 프로세스가 제공하는 테이프 서비스를 확장시킨 서비스로, 백업 저장 장치로 디스크의 사용이 가능하도록 하고 있다. 연결이 완료되면 각 DSP 프로세스에 대한 인증 과정을 거친 후, DMA는 확장 테이프 서비스에게 백업 파일을 작성하도록 명령한다. 백업 파일을 작성한 후

확장 테이프 서비스를 제공하는 DSP 프로세스는 데이터 서비스를 제공하는 DSP의 연결을 대기 한다. DMA는 데이터 서비스를 제공하는 DSP에게 확장 테이프 서비스를 제공하는 DSP로의 연결을 명령하고, 연결이 완료되면 데이터 서비스를 제공하는 DSP는 테이프 서비스를 제공하는 DSP로 백업 데이터를 전송한다.

DMA는 DSP가 전송하는 메시지를 비동기적으로 수신하여 처리한다. DSP는 백업 작업 중 발생하는 다양한 예러나 경고를 DMA에게 알려 처리하도록 한다. 그림 12는 백업 과정 중 백업 파일의 크기가 2GB가 넘거나, 혹은 백업 디스크에 여유 공간이 없을 경우의 처리 과정을 나타낸 것이다. 리눅스 환경에서는 파일 크기가 2GB를 넘지 않도록 제한하고 있다. 따라서 백업 파일의 크기가 2GB가 넘으면, 새로운 백업 파일을 작성하여 백업을 수행해야 한다. 또한 백업 시 선택한 백업 디스크 풀에 여유 공간이 부족한 경우, 여유 공간이 충분한 백업 디스크 풀로 교체하여 백업이 진행 될 수 있어야 한다. DSP는 이러한 예러 상황이 발생하면 백업을 일시 중지시키고, DMA에게 이를 통보한다. DMA는 DSP가 통보한 예러 정보를 검사하여 어떤 타입의 예러가 발생했는지를 확인한다. DMA는 예러의 타입에 따라 해당 예러를 처리할 수 있는 예러 핸들러를 호출하여 DSP가 예러를 처리할 수 있도록 명령한다. 예러가 처리되면 일시 정지된 백업이 다시 수행되도록 백업 진행 명령을 DSP에게 전송한다.

DSP가 모든 작업을 종료하고 이를 DMA 모듈에게 알리면, DMA 모듈은 로그 및 리포트 관리자를 통해 백업 과정에서 발생한 로그와 백업된 파일들에 대한 파일 히스토리 정보를 기록한다. 파일 히스토리 정보는 복

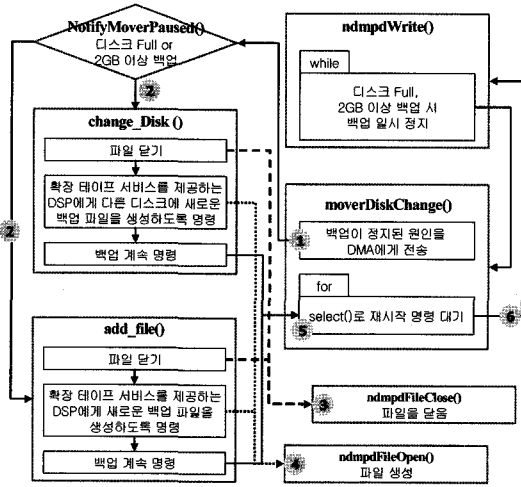


그림 12 백업 과정 중 에러 처리 흐름

구를 수행하기 위해 백업된 파일 정보를 검색할 목적으로 사용된다. 또한 사용자가 지정한 백업 파일의 생명주기에 따라 백업된 데이터에 대한 생명주기 정보를 기록한다. DMA는 작업 종료 상태에 따라 작업 리포트를 작성하며, 백업 작업을 요청한 백업 사용자의 백업 클라이언트 인터페이스에게 작업 종료 상태 정보를 메시지로 전송한다.

백업 사용자는 백업의 이전 또는 이후에 수행할 작업을 스크립트 관리자를 통해 작성할 수 있다. 스크립트는 백업 및 복구 작업의 전반부에 수행되는 pre 스크립트와 백업 및 복구 작업의 후반부에 수행되는 post 스크립트로 분류할 수 있다. pre/post 스크립트를 사용함으로써 사용자는 백업 소프트웨어에서 자동적으로 지원하기 힘든 작업 또는 사용자의 의사가 많이 반영되는 작업들을 수행할 수 있다.

제안하는 백업 소프트웨어에서는 백업 디스크 장치의 저장 공간을 효과적으로 사용하기 위해 생명주기 기능을 제공한다. 디스크 기반의 백업의 경우 더 이상 사용되지 않는 백업 데이터들이 디스크에 남아있을 수 있다. 이들 백업 데이터들은 저장 공간을 낭비하여 디스크의 가용성을 떨어뜨리고, 추가적인 저장 공간을 요구하게 된다. 따라서 이들 백업 데이터마다 적절한 보관 기간을 명시하고, 보관 기간이 지난 백업 데이터들에 대해서는 2차 백업 혹은 백업 데이터 삭제와 같은 관리 작업을 수행하는 기능이 필요하다. 제안하는 백업 소프트웨어에서는 백업 작업을 수행할 때 백업 데이터에 대한 보관 기간을 명시하고, 보관 기능이 지난 백업 이미지를 자동으로 검사하여 삭제하는 기능을 제공한다. 사용자가 백업에 대한 생명주기를 설정하면 백업 서버는 생명주기

가 종료되는 시점을 확인하고, 해당되는 백업 파일 및 백업 데이터와 관련된 파일 히스토리 정보를 모두 삭제한다. 이는 백업 디스크 장치의 저장 공간 낭비를 줄이고, 저장 공간의 가용성을 높인다.

4.2 복구

제안하는 백업 소프트웨어는 예약 복구와 즉시 복구를 지원한다. 로컬 및 원격에 존재하는 백업 데이터에 대한 복구가 가능하다. 제안하는 백업 소프트웨어의 복구는 백업 클라이언트의 작업 요청과 백업 서버의 서버 모듈인 복구 관리자와 인덱스 관리자, 각 데몬, 그리고 NDMP를 따라 설계된 백업 서버의 DMA 모듈과 백업 장치 서버의 DSP간 데이터 전송을 통해 수행된다.

그림 13은 제안하는 백업 소프트웨어의 복구 처리 구조를 나타낸 것이다. 백업 작업과 마찬가지로 복구 작업은 클럭 데몬을 통해 스케줄 관리자가 지정한 스케줄 정보에 따라 수행되는 예약 복구와 사용자가 요청한 즉시 스케줄 데몬을 거쳐 수행되는 즉시 복구로 나눌 수 있다. 또한 백업된 데이터를 전부 복구하는 전체 복구와 백업된 데이터의 일부를 선택하여 복구하는 부분 복구가 가능하다.

백업 클라이언트를 통해 전달된 백업 사용자의 복구 요청은 인덱스 관리자와 복구 관리자를 통해 분석되고 복구할 정보를 선택하게 된다. 인덱스 관리자와 복구 관리자를 통해 선택된 복구 정보에 따라 백업된 내용에 대한 전체 또는 부분 복구를 수행할 수 있다. 인덱스 관리자는 백업 작업 후 기록된 파일 히스토리 정보를 검색하여 백업 클라이언트에게 전달한다. 백업 사용자는 인덱스 관리자가 검색한 백업 데이터의 파일 히스토리 정보를 검색하여 복구할 백업 데이터의 목록을 작성할 수 있다. 사용자가 작성한 백업 데이터 목록이 서버의 복구 관리자에게 전달되면, 복구 관리자는 사용자로부터 전달받은 정보를 이용하여 복구 작업 정보를 작성한다.

작성된 복구 정보는 스케줄 관리자가 설정한 스케줄 정보에 따라 클럭 데몬에게 전달되어 지정된 시간에 스케줄 데몬으로 보내지거나, 즉시 스케줄 데몬으로 전달된다. 스케줄 데몬은 사용자가 요청한 작업 정보를 분석하여 백업 서버의 DMA 모듈에게 전달할 복구 파라미터를 작성한다. 스케줄 데몬은 작성된 복구 파라미터를 가지고 DMA 모듈을 호출한다. DMA 모듈은 전달받은 파라미터를 분석하여 사용자가 설정한 백업 장치 서버의 DSP와 연결을 수행하고, NDMP에 따라 복구가 수행되도록 한다. DMA 모듈은 백업 데이터가 저장된 백업 장치 서버의 사용자 플로부터 백업 데이터를 읽어 지정된 백업 장치 서버의 복구 위치에 복구 작업을 수행도록 DSP에게 명령한다.

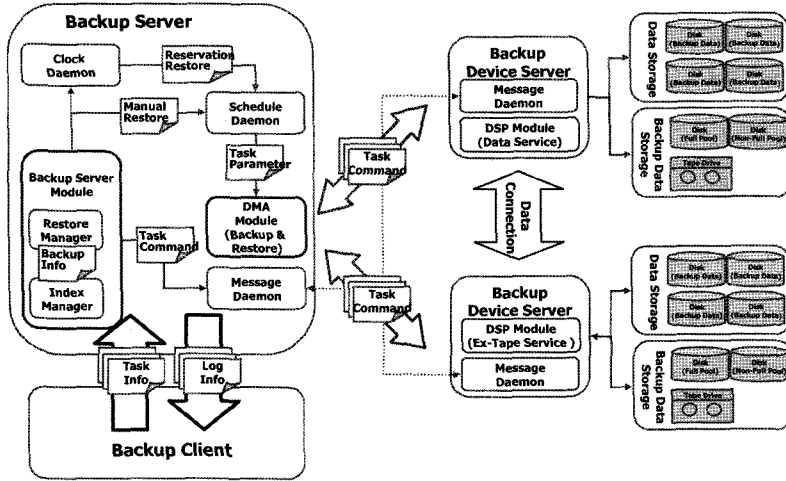


그림 13 제안하는 백업 소프트웨어의 복구 처리 구조

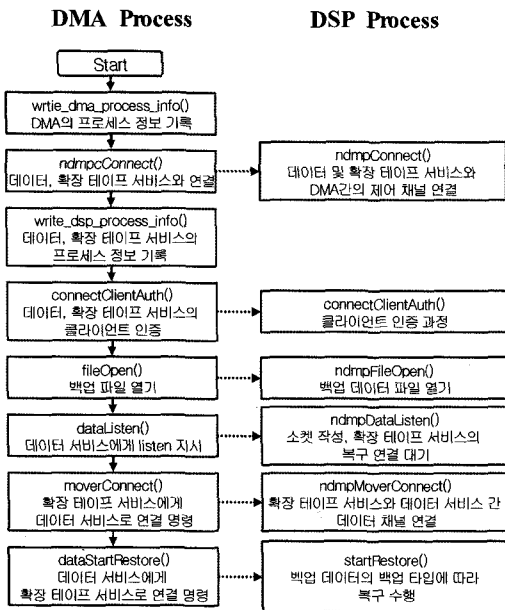


그림 14 복구 관련 함수 호출 과정

그림 14는 복구 과정에서 호출되는 함수의 흐름을 나타낸 것이다. 백업 과정에서와 마찬가지로 복구를 하기 위해 DMA 프로세스는 DSP 프로세스의 데이터 서비스와 확장 테이프 서비스에 연결을 수행한다. 연결이 완료되면 각 DSP 프로세스에 대한 인증 과정을 거친 후, DMA는 확장 테이프 서비스에게 복구할 백업 파일을 열도록 명령한다. 복구할 백업 파일을 연 후, DMA는 확장 테이프 서비스를 제공하는 DSP에게 데이터 서비스를 제공하는 DSP로의 데이터 전송 채널 연결을 명령하고, 연결이 완료되면 확장 테이프 서비스를 제공하는

DSP는 데이터 서비스를 제공하는 DSP로 백업된 데이터를 전송한다.

백업과 마찬가지로 DMA 모듈은 NDMP의 메시지와 여러 관리 정책에 따라 DSP를 관리하고 제어하며, DSP가 전송하는 메시지를 비동기적으로 수신하여 처리한다. DSP가 모든 작업을 종료하고 이를 DMA 모듈에 알리면, DMA 모듈은 로그 및 리포트 관리자를 통해 복구 과정에서 발생한 로그 정보를 기록한다. 복구 작업이 종료되면 DMA는 작업 종료 상태에 따라 작업 리포트를 작성하고, 복구 작업을 요청한 백업 사용자의 백업 클라이언트 인터페이스에게 작업 종료 상태 정보를 메시지로 전송한다.

5. 구현 및 성능 평가

5.1 구현 환경

본 논문에서는 네트워크 기반의 백업을 위해 제안된 개방형 표준 프로토콜인 NDMP를 이용하여 NAS와 SAN 환경에 적합한 디스크 기반의 백업 소프트웨어를 설계하고 구현한다. 본 논문에서 제안하는 백업 소프트웨어의 구현 환경은 표 4와 같다. 백업 소프트웨어의 구성 모듈 대부분은 gcc를 이용하여 작성하였으며, 실제 파일 시스템이나 장치의 백업 및 복구를 수행하는 부분은 tar, cpio, dd와 같은 명령을 사용한다. 데이터베이스 백업은 Oracle 데이터베이스에 한해 구현하며, Oracle의 Pro* c와 Oracle에서 지원하는 import, export 명령, 그리고 스크립트 언어를 사용하여 구현한다. 또한 백업 및 복구의 전, 후 스크립트 작업을 위해 Shell 스크립트를 사용한다. 백업 서버와 백업 장치 서버간의 데이터 통신을 위해 NDMP를 이용하여 DMA와 DSP를 구현한다. 본 백업 프로그램은 Red Hat Linux 9에서 구현한다.

표 4 제안하는 백업 소프트웨어의 구현 환경

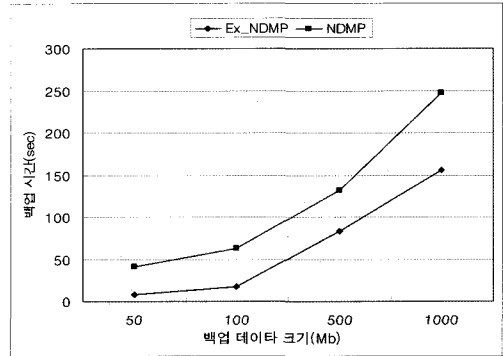
환경	특징
OS	Red Hat Linux release 9 (Shrike)
개발 언어	gcc shell script(GNU Bourne Again Shell)
파일/장치 백업	tar, cpio, dd, shell script
데이터베이스 백업	tar, export, import, shell script
NDMP	NDMP SDK Version 3.2

5.2 성능 평가

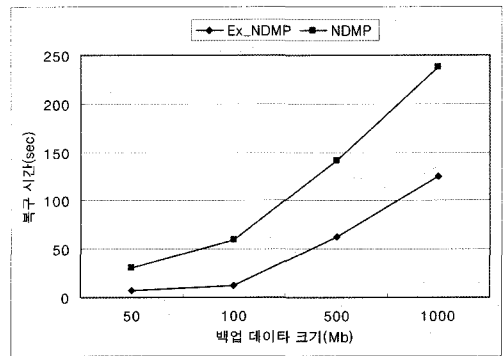
구현한 백업 소프트웨어의 성능을 평가하기 위해 다양한 종류의 백업 데이터를 사용하여 제안하는 백업 소프트웨어의 성능을 비교, 분석한다. 또한 백업 데이터 압축 기능을 사용한 백업 서비스와 압축 기능을 사용하지 않은 백업 서비스간의 성능을 비교, 분석한다. 성능 평가는 백업 작업의 평균 작업 시간을 기준으로 수행한다. 성능 평가에 사용된 시스템은 펜티엄III 1GHz 프로세서에 256Mbytes의 메모리를 가지며, 운영체제로 Red Hat Linux release 9를 사용한다. 백업 데이터는 다양한 동영상 및 텍스트 파일을 혼합하여 구성하며, 백업 데이터를 다양한 크기로 조정하여 성능 평가를 수행한다.

그림 15는 제안하는 백업 소프트웨어에서 다양한 유형의 실험 데이터에 대해 백업 및 복구를 수행한 결과를 비교한다. 실험 결과는 NDMP에서 제공하는 테이프 기반 백업 서비스와 제안하는 백업 소프트웨어의 디스크 기반 확장 테이프 서비스의 백업 및 복구 작업 시간으로 나누어 비교한다. 실험 결과, 제안하는 백업 소프트웨어에서 제공하는 디스크 기반 확장 테이프 서비스가 NDMP에서 제공하는 테이프 기반 서비스에 비해 2~5배의 빠른 작업 시간을 보인다. 이는 테이프 장치의 하드웨어적인 특성에 기인한 것으로 분석된다. 즉, 테이프 장치는 디스크 장치와 달리 순차적인 접근만 가능하므로 백업의 경우 테이프 장치를 인식하고 백업할 위치로 테이프 장치의 헤드가 이동하는데 걸리는 시간이 있다. 또한 복구의 경우에도 백업 데이터가 있는 위치까지 테이프 장치의 헤드가 이동해야 복구 작업을 진행할 수 있다. 따라서 테이프 서비스만 제공하는 기존 NDMP 서비스에 비해, 제안하는 백업 소프트웨어는 디스크 기반의 백업 및 복구를 지원하는 확장 테이프 서비스를 제공하여 성능을 향상시킬 수 있다.

그림 16은 제안하는 백업 소프트웨어에서 다양한 크기의 백업 데이터에 포함된 일정 크기의 데이터를 부분 복구한 실험 결과를 비교한다. 실험 결과는 NDMP에서 제공하는 테이프 서비스와 제안하는 백업 소프트웨어의 디스크 기반 확장 테이프 서비스의 평균 작업 시간으로 나누어 비교한다. 실험에서는 다양한 크기의 백업 데이터에 대해 50Mb의 데이터를 부분 복구하여 성능을 비



(a) 백업 시간



(b) 복구 시간

그림 15 제안하는 백업 소프트웨어의 데이터 크기에 따른 백업 및 복구 시간

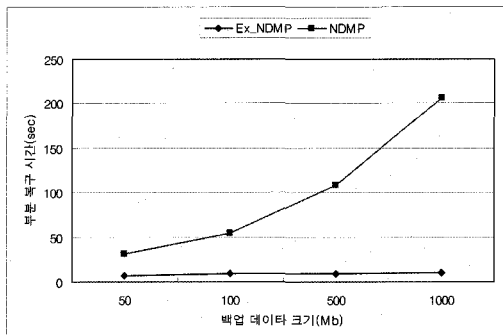


그림 16 제안하는 백업 소프트웨어의 백업 데이터 크기에 따른 부분 복구 작업 시간

교한다. 실험 결과 테이프 서비스만을 제공하는 NDMP의 경우 백업 데이터의 크기가 변화함에 따라 높은 작업 시간 차이를 보인 반면, 제안하는 백업 소프트웨어는 일정한 수준의 작업 시간을 유지하고 있음을 확인할 수 있다. 이번 실험 역시 테이프 장치가 갖는 하드웨어적인 특성에 기인한 것으로 분석된다. 즉 NDMP의 테이프 서비스는 테이프만을 지원하므로 부분 복구를 수행하기

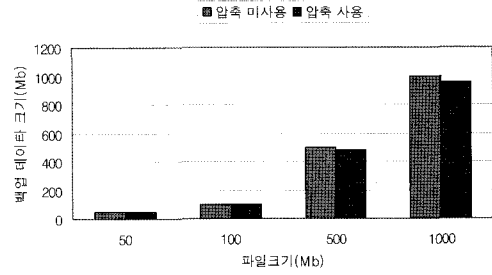
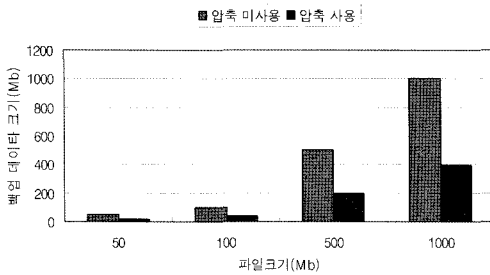
위해 복구 위치까지 테이프의 헤드를 이동시켜야 한다. 하지만 제안하는 백업 소프트웨어의 확장된 테이프 서비스는 디스크를 지원하므로 백업 데이터에 대한 즉시 접근이 가능하다. 따라서 기존 NDMP 서비스에 비해 빠른 부분 복구 성능을 보장한다.

그림 17은 제안하는 백업 소프트웨어에서 압축 기능 적용 여부에 따른 백업 데이터의 크기를 비교한다. 실험 데이터는 이미지 및 동영상과 텍스트 및 문서 데이터로 다양한 크기로 나누어 실험한다. 비교 결과 압축 옵션을 적용한 경우, 텍스트 기반의 데이터는 60% 이상의 압축률을 보이는 반면, 이미지나 동영상과 같은 데이터의 경우 5% 미만의 압축률을 보인다. 이는 데이터의 특성에 기인한 것으로, 이미지나 동영상의 경우 이미 높은 수준의 압축이 되어 있는 데이터이기 때문에 일반 텍스트나 문서 데이터와 비교하여 낮은 수준의 압축률을 보인다고 분석된다.

그림 18은 제안하는 백업 소프트웨어에서 압축 기능 적용 여부에 따른 백업 작업 시간을 I/O 시간과 데이터 압축 시간으로 나누어 비교한다. 실험 데이터는 텍스트 및 문서 데이터와 이미지 및 동영상 데이터를 다양한 크기로 나누어 실험한다. 실험 결과 텍스트 데이터의 경

우 높은 압축률에 의해 I/O 시간이 크게 줄어들지만, 데이터 압축에 드는 작업 시간에 의해 전체적인 작업 시간은 증가함을 알 수 있다. 이미지 및 동영상 데이터의 경우 데이터의 압축률이 텍스트 데이터에 비해 크게 떨어지므로 I/O 시간은 압축 기능을 사용하지 전과 큰 차이를 보이지 않는다. 하지만 텍스트 데이터에 비해 데이터 압축에 드는 시간은 약간 감소함을 알 수 있다.

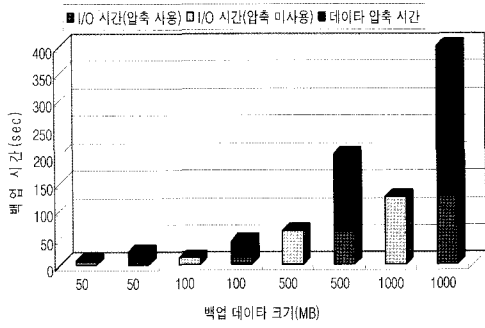
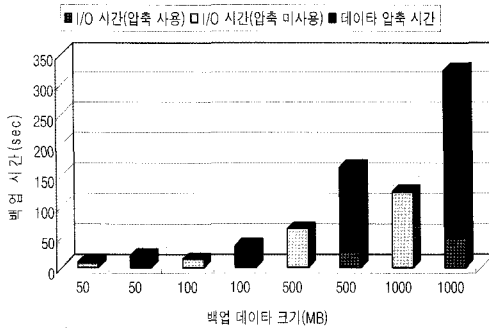
표 5는 제안하는 백업 소프트웨어와 메릴랜드 대학의 AMANDA, 그리고 Veritas의 NetBackup, Legato의 NetWorker, HP의 OmniBack-II 제품을 비교 분석한다. 비교 결과 제안하는 백업 소프트웨어는 다른 소프트웨어들에 비해 완벽한 NDMP 호환 기능과 다양한 백업 유형을 지원하고 있다. 제안하는 백업 소프트웨어는 중앙 집중식 관리를 통해 원활한 백업 및 복구가 가능하며, LAN 백업과 LAN-Free 백업을 지원한다. 하지만 NetBackup과 같은 상용 소프트웨어와 비교하여 온라인 백업 기능 지원과 다양한 DBMS 백업 및 복구 기능에 대한 지원이 부족하다. 또한 Snapshot, Mirroring, 바이러스 검사 기능등과 같은 고급 기능을 제공하지 못한다.



(a) 텍스트 / 문서 데이터

(b) 이미지 / 동영상 데이터

그림 17 제안하는 백업 소프트웨어의 압축 기능 사용에 따른 백업 데이터의 크기



(a) 텍스트 / 문서 데이터

(b) 동영상 / 이미지 데이터

그림 18 제안하는 백업 소프트웨어의 압축 기능 사용에 따른 백업 작업 시간 비교

표 5 백업 소프트웨어의 특징 비교

기능	Veitas NetBackup	Legato NetWorker	HP OmniBack-II	메릴랜드 AMANDA	제안하는 백업소프트웨어
중앙 집중식 관리	O	O	O	O	O
NDMP 호환	△	O	△	X	O
파일시스템 백업	O	O	O	O	O
데이터베이스 백업	O	O	O	X	△
Raw 장치 백업	O	O	O	X	O
장치 및 미디어 관리	O	O	O	O	O
다양한 압축 기능	△	△	△	X	O
온라인 백업	O	O	O	X	△
LAN-Free 백업	O	O	O	X	O
LAN 백업	O	O	O	O	O

6. 결론 및 향후 연구

본 논문에서는 NAS와 SAN에 적합한 디스크 기반의 백업 소프트웨어를 설계하고 구현했다. 구현한 백업 소프트웨어는 이 기준 간의 백업을 위해 NDMP를 지원한다. 여러 클라이언트의 다양한 백업 요구를 스케줄링 하여 처리가 가능하고, 스케줄링에 따른 자동 백업 및 예약 복구와 사용자의 요청 즉시 수행되는 수동 백업 및 즉시 복구를 지원한다. 하나의 백업 서버가 백업 및 복구를 중앙 집중적으로 관리하기 때문에 일관된 정책을 유지할 수 있고, 수행 중인 작업에 대한 제어를 효과적으로 수행할 수 있다. 테이프 기반의 백업 서비스뿐만 아니라 디스크 기반의 백업 서비스도 제공하여 백업 장치의 선택 범위를 넓히고, 보다 빠른 백업이 가능하도록 한다.

구현한 백업 소프트웨어는 다른 상용 소프트웨어와 비교하여 유사한 기능을 제공하지만 일부 고급 기능에 대한 지원은 아직 제공되지 않고 있다. 향후 백업 및 복구 성능의 개선과 상용 소프트웨어와 비교하여 지원이 미비한 기능 등의 추가 구현을 위해 연구를 계속할 것이다.

참고 문헌

[1] Backup/Recovery Tutorial, Storage Networking Industry Association, 2001.
 [2] A. L. Chervenak, V. Vellanki, Z. Kurmas and V. Gupta. "Protecting File System, A Survey of Backup Techniques," Pro. Joint NASA and IEEE Mass Storage Conference, 1998.
 [3] 왜 백업이 필요한가?, <http://www.veritas.co.kr/whitepaper.asp>
 [4] 백업 및 복구 서비스 구성, <http://www.microsoft.com/korea/technet/security/guidance/secmod201.asp>
 [5] W. C. Preston, UNIX Backup and Recovery, pp. 1-65, O'REILLY, 1999.
 [6] W. C. Preston, Using SANs and NAS, pp. 66-188, O'REILLY, 2002.
 [7] 서대화, 민병준, 임기욱, "네트워크 연결형 스토리지의 기술 동향", 제 19권, 제3호, pp. 6-13, 2001.
 [8] SAN이란 무엇인가?, <http://www.nds.co.kr/tech.htm>
 [9] Kimberly Keeton, Eric Anderson, "A Backup

Appliance Composed of High-Capacity Disk Drives", p. 171, HotOS, 2001.

[10] Disk to Disk Backup-Backup and Restore, <http://www.stonefly.com/applications/whitepapers.asp>
 [11] Michael Kaczmariski, Tricia Jiang, David A. Pease, "Beyond backup toward storage management," IBM Systems Journal 42(2), pp. 322-337. 2003.
 [12] NDMP, <http://www.ndmp.org>
 [13] NDMP를 이용한 백업 솔루션, <http://emc.webinar.co.kr/>
 [14] AMANDA, <http://www.amanda.org/>



윤 중 현

2003년 충북대학교 전기전자및컴퓨터공학부 정보통신공학 전공(공학사). 2005년 충북대학교 정보통신공학과(공학석사) 2005년 ~ 현재 충북대학교 정보통신공학과 박사과정. 관심분야는 데이터베이스 시스템, 저장 시스템, 시공간 색인 구조, 이동 객체 데이터베이스, 센서 네트워크 등



복 경 수

1998년 2월 충북대학교 수학과(이학사) 2000년 2월 충북대학교 정보통신공학과(공학석사). 2005년 2월 충북대학교 정보통신공학과(공학박사). 2005년 3월~현재 한국과학기술원 전산학과 Postdoc. 관심 분야는 시스템 소프트웨어, 이동 객체 데이터베이스, 자료 저장 시스템, 멀티미디어 데이터베이스, 센서 네트워크 및 RFID 등



유 재 수

1989년 2월 전북대학교 컴퓨터공학과 공학사. 1991년 2월 한국과학기술원 전산학과 공학석사. 1995년 2월 한국과학기술원 전산학과 공학박사. 1995년 2월~1996년 8월 포도대학교 전산통계학과 전임강사. 1996년 8월~현재 충북대학교 전기전자컴퓨터공학부 및 컴퓨터정보통신연구소 부교수. 관심분야는 데이터베이스 시스템, 정보검색, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅 등