한국산학기술학회논문지
Vol. 8, No. 1, pp. 65-72, 2007

e-Learning 서버 작업부하 분석

손세일[1], 안효범[2*], 김흥준[3]

Analysis of e-Learning Server Workload

Sei-Il Son[1], Hyo-Beom Ahn[2*] and Heung-Jun Kim[3]

요 약 본 논문은 단국대학교의 e-learning 서버의 부하 분석을 통해 교육용 서버의 통계적 부하 모델 작성을 위한 기초 자료 제공을 목적으로 한다. 분석 결과로 파일의 크기 분포, 파일 유형별 접속 빈도와 전송량, 접근 간격, 선호도 변화, 네트워크별 접근 비율 등을 제시하였다. 특히, 동영상 파일의 크기 분포와 접근 빈도별 파일 분포는 이전의 연구들과는 다른 결과가 나타났다. 이것은 동영상 강의를 제작하기 위해 저작 도구를 사용하고 수강생 수의 제약이 없는 온라인 교육의 특성이 반영되었기 때문이다. 본 논문에서 제시된 분석 결과는 e-learning 시스템 구조와 서버 성능 향상을 위한 연구에 기초 자료로 이용될 수 있다.

Abstract This paper aims to provide information to generate a statistical load model of an educational server by analyzing workload of an e-Learning server at Dankook University. The result of the analysis shows file size distribution, access frequency and transmission volume for each file type, access interval, changes in preference and clients access rate by networks. In particular, it had different results from previous studies about video file's size distribution and file distribution based on access frequency. This is because the characteristics of e-learning are influenced by using authoring tools for making into video file and by freeing the number of students who register for a course. The result in this paper can be used as a basic data for studies designed to improve e-learning system architecture and server performance.

key Words : e-Learning, Server Workload, Network Server

1. 서론

최근 인터넷을 기반으로 교육 서비스를 제공하는 e-Learning이 보편화되면서 사용자들이 급속히 증가하고 있다. e-Learning 서버는 일반적인 웹 서버와는 달리 오디오, 비디오로 형태의 강의와 텍스트 형태의 강의 자료를 제한된 사용자들, 즉, 수강생들에게 제공한다. e-Learning은 멀티미디어 콘텐츠로 수강생들을 교육하기 때문에 서버와 네트워크에 많은 부하를 발생시킨다. 특히 수강생들의 서버 접속이 특정 시간대에 집중되면 과부하 발생, 응답 지연 등과 같은 문제들이 발생한다. 이 같은 문제들은 일반적인 미디어 서버에서도 발생되었기 때문에 프록시 서버(proxy server), 캐싱 전략, CDN(Content Delivery Network)과 같은 다양한 연구들이 있어왔다.

이 같은 연구의 결과로 제안된 방법들은 타당성을 검증받기 위해 시뮬레이션을 수행한다. 시뮬레이션에 이용되는 서버의 부하는 통계적 부하 모델에 따라 생성되거나 운영 중인 서버의 실제 트레이스(trace)를 적용한다. 통계적 부하 모델을 이용하여 서버의 부하를 생성하기 위해서는 무엇이 부하를 결정하는 인자들(parameters)인지를 파악하고, 이들이 어떤 분포를 따르는지를 부하 모델로 표현하고, 부하 생성기를 제작하여 시뮬레이션에 이용한다. 트레이스를 이용하면 실제 환경에서 예상되는 부하를 보다 정확히 발생시킬 수 있지만, 트레이스를 수집하기 위해서 긴 시간이 필요하며, 수집된 자료의 양도 매우 거대하다. 또한 시뮬레이션에 사용하기 위해서 트레이스를 가공하는 경우 많은 비용이 요구된다[15]. 따라서 대부분의 서버 성능 향상과 관련된 연구들은 통계적 부하 모델을 이용한 시뮬레이션을 통해 자신의 제안을 검증하고 있다.

[1]단국대학교 정보컴퓨터과학공학부
[2]공주대학교 정보통신학부학대학교
[3]진주산업대학교 컴퓨터공학부
*교신저자: 안효범(hbahn@kongju..ac.kr)

본 논문에서는 e-Learning 서버의 부하 분석을 통해 교육용 서버의 통계적 부하 모델을 작성할 수 있는 기초 자료 제공을 목적으로 한다. 이를 위해 2004년 8월 30일부터 2004년 12월 17일 사이의 단국대학교 e-Learning 서버 로그를 분석하였다. 분석된 e-Learning 서버는 이전의 연구들에서 다룬 정적, 동적 웹 문서들뿐만 아니라 많은 멀티미디어 파일들을 가지고 있다. 분석된 서버는 대학의 정규 과목의 전부 또는 일부를 지원하고 있으며, 사용자들은 사이버 강좌 또는 웹 보조수업 과목을 수강 신청한 단국대학교 재학생들로 제한된다. 서버에서 지원하는 과목의 운영은 대학의 학사일정에 따른다.

본 논문에서는 분석 결과로 파일의 크기 분포, 파일 유형별 접속빈도와 전송량, 시간대별 사용자 접근 빈도, 접근 간격, 네트워크별 접근 비율, 선호도 변화 등을 제시하였다. 본 논문의 연구 결과는 사용자가 급증하고 있는 e-Learning 시스템의 구조에 관한 연구와 서버의 성능 향상을 위한 연구들에 기초 자료로 이용될 수 있다. 또한 제안된 방법들의 타당성을 검증할 수 있는 통계적 부하 모델을 작성하는 토대가 된다.

2. 관련 연구

[5]에서는 Massachusetts 대학에서 오디오 형태의 강의 파일과 텍스트 형태의 강의 자료를 수강생들에게 제공하는 코스웨어 시스템을 대상으로 부하 특성을 분석하였다. 이 논문에서는 서버에서 측정한 데이터에 대한 다양한 통계적 분포 함수와 함께 여러 인자들의 값을 제시하였다. [4]에서는 스웨덴의 Luleå University of Technology에서 동영상 강의를 포함한 다양한 스트리밍 서비스를 제공하는 mMod(multicast Media On Demand) 시스템의 사용자 접근 로그 파일의 분석을 통해 시간적 지역성, 접근 간격, 파일의 크기 분포, 파일의 인코딩 대역폭 분포, 전송량 등을 분석하였다. [2]에서는 eTeach 시스템과 BIBS(Berkeley Internet Broadcasting System)의 로그 데이터 분석을 통해 사용자 요청의 밀집도, 접속 빈도의 시간적, 공간적 지역성 평가와 함께 서버의 부하를 감소시키기 위한 멀티캐스트 기술의 성능을 분석하였다. [13]에서는 연속형 미디어 파일을 실시간으로 제공하는 세 개의 사용 사이트의 로그를 이용해 각 서버가 보유한 연속 미디어 파일의 특성과 각 연속 미디어 파일에 대한 사용자 접근 요구의 특성을 분석하였다. 이 논문에서는 사용자 요구 분포 특성을 파일의 크기가 클수록 파일 수의 비율 분포보다 사용자 요구 수의 비율 분포가 높게 나타나며, 요구 수의 비율 분포보다 전송된 데이터 량의 비율 분포가 높게 나타남을 보였다. 또한 사용자들이 높은 대역폭으로 인코딩 된 미디어 파일들을 선호하고 있음을 알 수 있었다. [1]에서는 두 기업의 미디어 서버들의 작업 부하를 분석하였다. 이 논문에서는 기업 미디어 서버의 작업부하에 대한 특성들을 기술하기 위해 미디어 서버의 특성을 정적(static)인 것과 시간적(temporal)인 것으로 분류하였다. 정적 특성들은 클라이언트들의 참조에 대해 양적 분석을 통해 미디어 파일들에 대한 기본적 특성들을 보여주었다. 시간적 특성들은 시간이 지남에 따라 동적으로 변화하는 접근 특성들을 반영하였다. 시간적 특성들을 도출하기 위해 서버에 새로운 파일이 도입될 때의 영향과 클라이언트들의 새로운 파일에 대한 접근 패턴의 변화를 척도로 삼아 분석하였다.

3. 분석 시스템 배경

3.1 개요

본 논문은 2004년 8월 30일부터 2004년 12월 17일 사이에 단국대학교 e-learning 서버에 기록된 로그를 기초로 다양한 분석을 수행하였다. 이 기간 동안 개설된 유형별 강좌 수와 수강생 수는 [표 1]과 같다. [표 1]에서 사이버 강의란 수업의 전 과정이 온라인을 통해 이루어지는 강좌를 말하며, 웹 보조 수업이란 전통적인 오프라인 수업에서 수강생들의 편의를 돕기 위해 강의 자료, 과제물 제출 및 평가, 게시판 등이 웹 기반으로 운영되는 수업을 말한다.

[표 2] 유형별 강좌 수 및 수강생 수

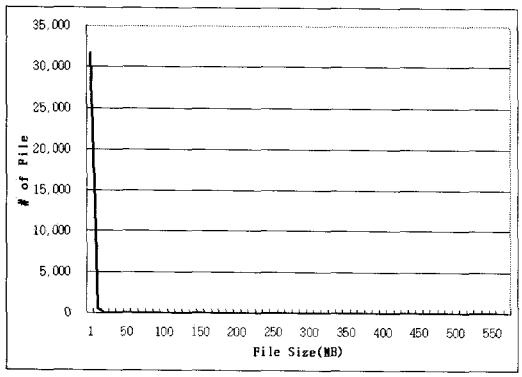
유형	캠퍼스	교과목	수강생
사이버강의	서울	6	637
	천안	11	516
	소계	17	1,153
웹보조수업	서울	64	3,284
	천안	102	4,964
	소계	166	8,248
	총계	183	9,401

단국대학은 서울과 천안에 캠퍼스를 두고 있으며, 두 캠퍼스의 네트워크는 전용선을 통해 연결되어 있다. e-learning 시스템은 서울 캠퍼스 내에 위치하며 서울과 천안 두 캠퍼스에 개설된 강좌들과 수강생들에게 통합된 서비스를 제공한다. 운영 중인 e-learning 시스템은 웹 서버, 미들웨어, 데이터베이스 서버, VOD 서버로 구성된다. 사용자들은 웹 서버를 통해 필요한 서비스를 제공받으며,

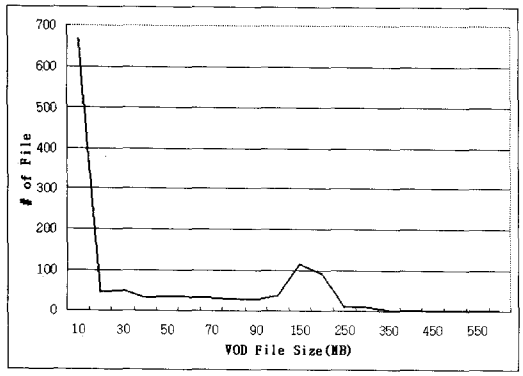
웹 프로그램 언어로 JSP가 사용되었다. 데이터는 데이터베이스 서버에서 관리되고, 지금까지 진행되었던 모든 동영상 강의 파일들은 VOD 서버에 저장되어 있다.

사용자가 웹 서버에 접속하여 요청한 모든 작업들은 수행과 동시에 로그 파일에 기록된다. 로그 파일은 웹 서버가 수행한 작업에 대한 정보 및 성공 여부를 저장하며, Access 로그, Error 로그, Referral 로그 및 Agent 로그 등 다양한 종류의 로그 파일들이 있다. 본 논문에서는 이들 중 Access 로그를 분석하였으며, 이것은 NCSA의 CLF(Common Log file Format) 형식으로 기록되어 있다.

3.2 파일 크기의 분포



(a) 전체 파일



(b) 동영상 파일

[그림 1] 파일의 크기 분포

[그림 1]은 분석기간 동안 사용자들이 접근한 웹 서버와 VOD 서버의 총 31,566개 파일의 크기 분포를 보여준다. 분석에 따르면 접근된 파일의 평균 크기는 1.77MB이고, 표준 편차는 15.4MB이다. 표준 편차가 매우 큰 것을 볼 때 파일의 크기가 넓은 범위에 걸쳐 분포하고 있음을 알 수 있다. 또한 왜도(skewness)는 11.9로 이것은 파일 크기의 분포를 정규 분포와 비교할 때 양의 방향으로 긴 꼬리를 갖는 비대칭 분포임을 보여준다. 첨도(kurtosis)는 187.4로 파일의 크기 분포가 정규 분포와 비교하여 매우 뾰족한 형태임을 알 수 있다.

[그림 1]의 (a)는 전체 파일에 대한 크기 분포를 보여주고 있으며, 이것은 이전 연구들과 유사한 후미성(厚尾性)이 있는 분포(꼬리가 두꺼운 분포, Heavy Tailed Distribution)를 나타낸다[8],[11],[15]. 파일의 크기 분포에 후미성이 있다는 것은 대부분의 파일들이 크기가 작고, 소수의 파일들만 크기가 크다는 것을 의미한다. 따라서 파일 크기 분포는 정규 분포가 아닌 파레토 분포(pareto distribution)로 나타낸다[15][16].

하지만 동영상 파일들만을 선택하여 크기 분포를 보여주는 [그림 1]의 (b)는 일반적인 파레토 분포(pareto distribution)를 따르지 않으며, 100MB부터 200MB 사이의 크기를 갖는 파일들이 많이 존재하고 있다. 동영상 강의 파일은 관련 부서의 도움을 받아 저작 도구를 이용해서 제작하기 때문에 인코딩 방법과 압축률은 동일하며, 파일의 크기는 강의 시간에 영향을 받는다. 또한 20MB 이하의 동영상 파일들은 과목의 소개, 슬라이드 등에 삽입되는 것들이다.

4. 작업 부하 특성

4.1 처리 상태에 따른 분석

서버에 제출된 HTTP 요청의 처리 결과를 로그에 기록된 상태코드를 이용하여 분석하였다. 상태코드가 2xx와 3xx는 요청한 정보를 서버가 클라이언트에게 성공적으로 처리한 경우이고, 4xx는 클라이언트 오류를 5xx는 서버 오류를 의미한다. [표 2]는 월별/상태코드별 응답 현황으로 조사기간 내의 로그에 기록된 모든 요청을 정리한 것

[표 2] 월별/상태코드별 응답 현황

상태코드	8월	9월	10월	11월	12월	전체	비율
2xx	189,026	7,048,321	9,057,037	7,402,704	5,094,755	28,791,843	55.64%
3xx	111,005	5,275,571	6,804,994	5,547,251	4,014,220	21,753,041	42.04%
4xx	7,631	300,713	368,740	298,796	211,547	1,187,427	2.29%
5xx	13	2,445	4,281	2,168	1,225	10,132	0.2%
계	307,675	12,627,050	16,235,052	13,250,919	9,321,747	51,742,443	

이다. 이 표를 보면 클라이언트 요청에 대해 성공적으로 응답한 경우가 97.68%로 이전 연구[8]에서 분석 서버들의 성공적 응답률과 비슷한 결과를 보이고 있다.

[표 3]은 상태코드가 2xx를 갖는 응답들에 대해 참조된 파일의 유형별 참조횟수와 전송량을 보여준다. 여기서 문서 파일이란 .htm, .css, .xml, .xls, .txt, .hwp 등의 확장자를 갖는 파일이고, 이미지 파일이란 .gif, .jpg, .bmp, *.ico, .swf 등의 확장자를 갖는 파일이다. 멀티미디어 파일이란 .avi, *.asf, .asx, .smi, .mp3, .wav, .wmz 등의 확장자를 갖는 파일이고, 프로그램 파일이란 .jsp, .asp, .php, .pl, .cgi 등의 확장자를 갖는 파일이고, 그 외 파일은 기타로 분류하였다. 멀티미디어 파일에 대한 접근 비율은 0.04%에 불과하지만, 전송량의 비율은 17.92%로 접근 횟수에 비해 큰 차이를 보인다. 이것은 요청된 멀티미디어 파일의 크기가 상대적으로 다른 파일들에 비해 대단히 크기 때문이다.

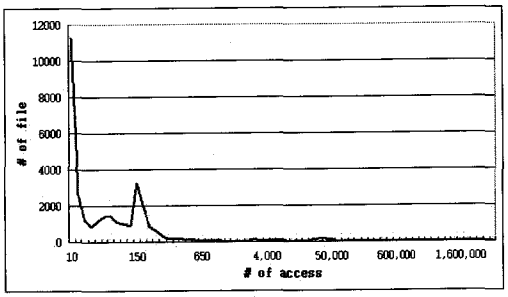
[표 3] 파일 유형별 접근횟수 및 전송량 (상태코드 2xx만)

파일 유형	접근횟수	접근비율	전송량(MB)	전송비율
문서	4,073,272	14.15%	15,501	5.29%
이미지	14,827,326	51.50%	98,592	33.64%
멀티미디어	12,932	0.04%	52,511	17.92%
프로그램	9,546,246	33.16%	121,969	41.62%
기타	332,067	1.15%	4,484	1.53%
계	28,791,843		293,057	

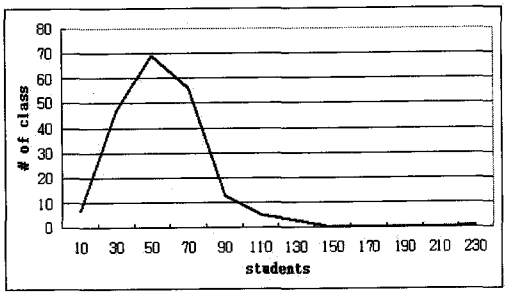
4.3 파일 접근 특성

4.3.1 파일 접근 빈도

e-learning 서버에 요청된 파일들에 대해 접근 빈도를 분석하였다. 서버 내 모든 파일들이 동일한 접근 빈도로 요청되지 않았으며, 일부 파일들이 집중으로 요청되었음을 확인하였다. 이 같은 특성은 이전의 관련 연구들 [8],[9],[13]에서도 지수 분포나 파레토 분포로 표현되어 왔다. [그림 2]의 (a)는 접근 횟수별로 파일 수를 그래프로 나타낸 것이다. 가장 빈번하게 접근한 파일들은 프로그램 파일들로 사용자들에게 각종 메시지를 전송하고, 공지 사항과 게시판의 내용을 보여주고, 사이버 강의의 시작하는 html과 jsp 파일들 이다. 본 논문의 분석 결과에 따르면, 이전의 연구 결과와는 달리 접근 횟수가 100부터 200번 사이였던 파일들의 비율이 16.92%나 되며, 이것은 단순한 지수 분포나 파레토 분포로 설명할 수 없다. 온라인 강의 특성 상 물리적 공간의 제약이 없어 수백 명의 수강생들이 한 과목을 수강할 수 있기 때문에 수강생이 많은 과목의 관련 파일들의 접근 횟수가 증가했다. 따라서 파일의 접근 빈도를 모델링 할 때는 [그림 2]의 (b)와 같은 과목별 수강생들의 분포를 고려해야만 한다.



(a) 접근횟수별 파일 수 분포

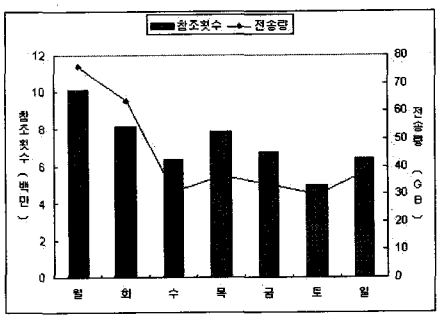


(b) 수강생 분포

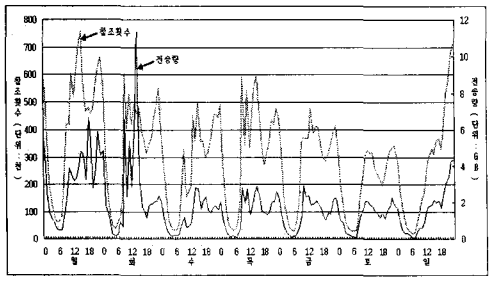
[그림 2] 접근 횟수에 따른 파일 분포와 수강생 분포

4.3.2 시간별 접근 비율

[그림 3]은 요일별 접근 횟수와 그에 따른 전송량을 보여주고 있다. [그림 3]의 (a)를 보면 학생들의 접근이 월요일, 화요일, 목요일이 높으며, 토요일이 가장 접근 횟수가 낮음을 알 수 있다. [그림 3]의 (b)를 보면 월요일 오후 12시 무렵과 일요일 야간 시간대에 접근이 빈번함을 알 수 있다. 데이터 전송량은 접근 횟수와 비례하는 모습을 보여준다. 흥미로운 점은 사용자들이 주중인 수요일 오후보다 일요일 야간에 보다 많이 서버에 접근한다는 것이다.

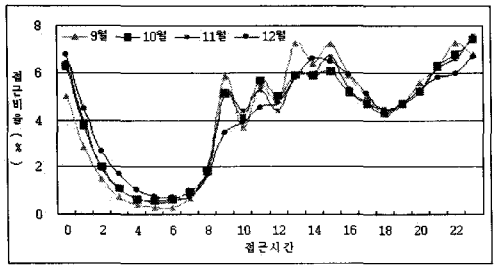


(a) 요일별 접근과 전송량

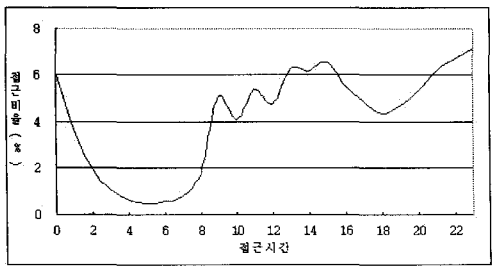


(b) 요일별 시간대별 접근과 전송량

[그림 3] 요일별 접근 횟수와 전송량



(a) 월별 시간대별 접근 비율



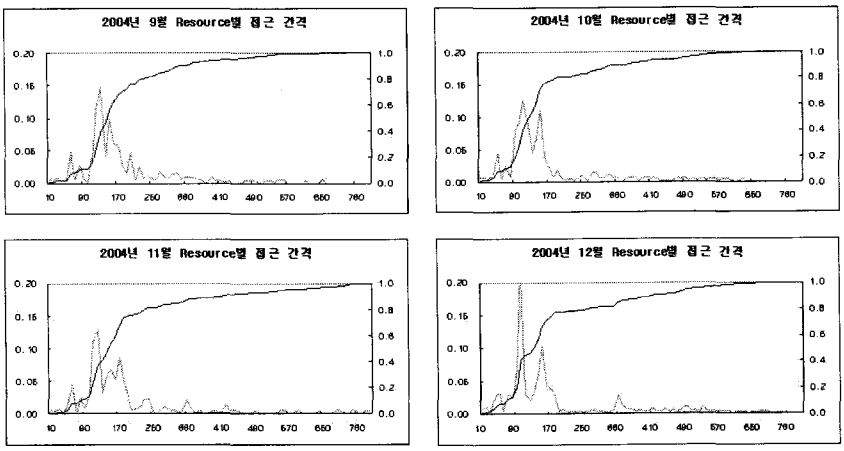
(b) 평균 시간대별 접근 비율

[그림 4] 시간대별 접근 현황

[그림 4]는 사용자들이 서버에 자원을 요청한 시간대를 기준으로 각 시간대별 접근 비율을 보여준다. [그림 4-a]는 2004년 9월 1일부터 2004년 12월 17일까지의 사용자들의 월별 시간대별 접근 비율을 계산하고, 그 결과를 그래프로 나타낸 것이다. 이 그림을 보면 매월 사용자들의 시간대별 접근 비율이 상당히 유사함을 알 수 있다. [그림 4-b]는 2004년 2학기, 즉 2004년 8월 30일부터 2004년 12월 17일사이의 시간대별 접근 빈도를 구한 후, 매 시간별로 접근 비율을 계산하고, 그 결과를 그래프로 그린 것이다. 이 그림을 보면, 사용자들은 13시부터 15시 사이와 21시부터 자정사이에 집중적으로 서버에 접근하고 있으며, 가장 집중적으로 서버에 접근한 시간은 23시로 접근 비율이 7.13%로였다. 오전 1시부터 사용자들의 서버 접근이 급격히 줄어 오전 2시부터 8시 사이의 접근 비율은 평균 약 1%이고, 오전 9시부터 다음날 새벽 1시까지는 평균 5.56%의 접근 비율을 보였다.

4.3.3 파일의 접근 간격

[그림 5]는 2004년 9월부터 12월 사이의 사용자 요청에서 대해 자주 접근되는 상위 90%의 파일에 대해 접근 간격과 누적 분포를 보여주고 있다. 그래프에서 X축은 접근 요청이 발생한 간격을 나타내며 단위는 초(second)이다. 좌측 Y축은 파일의 접근 비율로 2004년 9월 그래프를 보면 15%의 파일들이 130초 간격으로 요청되었다. 우측 Y축은 X축과의 교차점까지의 접근된 파일들의 누적 비율을 보여준다. 2004년 9월 그래프를 보면 0부터 130초 사이의 간격으로 요청된 파일들의 누적 접근 비율이 41%임을 알 수 있다. 또한 그래프들을 보면 평균 접근 간격 180초 이내로 요청되는 파일들의 비율이 70% 이상이며, 접근되는 파일들의 90%는 평균 360초 간격으로 요

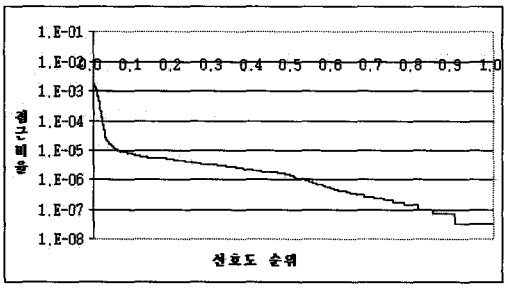


[그림 5] 월별/Resource별 접근 간격과 누적분포

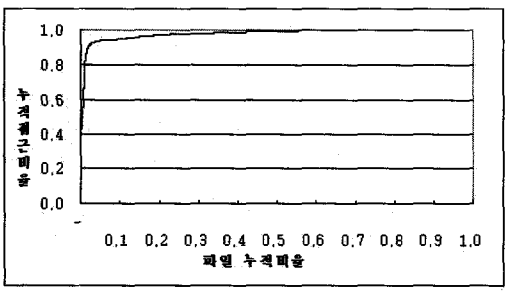
청된다. 파일 접근 간격에 대한 지식은 캐싱 알고리즘의 재배치(replacement) 전략을 세우는데 도움을 준다.

4.3.4 파일의 선호도

파일의 선호도란 서버에 존재하는 개별 파일의 상대적 요청수를 말하며, 웹 서버 내의 파일에 대한 선호도 분포는 통상적으로 Zipf의 법칙을 따른다고 알려져 왔다[12]. Zipf의 법칙을 웹 서버에 적용하면, 소수의 파일들에 사용자 접근이 대부분 집중되고, 나머지 파일들은 거의 참조되지 않는다.



(a) 선호도 순위에 따른 접근 비율



(b) 선호도 순위에 따른 누적접근비율

[그림 6] 파일의 선호도

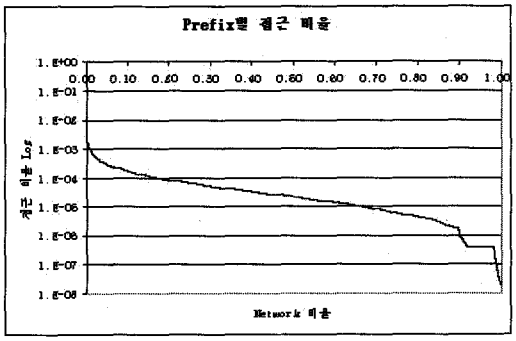
[그림 6-a]는 파일의 선호도 순위에 따른 접근 비율을 $\log_{10}$그래프로 그린 것이다. 이 그림을 보면 선호도가 높은 소수 파일들의 접근 비율이 10^{-3}이상이고, 나머지 파일들은 10^{-5} 이하의 낮은 접근 비율을 갖는 것을 알 수 있다. [그림 6-b]는 선호도 순위에 따라 정렬한 파일들의 접근 비율을 누적한 결과를 보여준다. 선호도 순위 상위 10% 이내의 파일들의 누적 접근 비율이 95%이 되며, 선호도가 높은 상위 100개 파일들의 누적 접근 비율이 51%나 되었다.

파일들의 시간에 따른 접근 특성이 어떻게 변화하는지를 알아보기 위해 2004년 9월부터 12월까지 4개월간의 데이터에 대해 매월 선호도 상위 1000등까지의 파일 목록을 구하여, 다른 달의 목록과 비교하고, 이를 [표 4]와 같이 정리하였다. 이 표를 보면 2004년 9월 상위 1000등 내의 파일들 중 81.7%의 파일들이 2004년 10월에도 선호도 상위 1000등내에 있음을 알 수 있다. 하지만 그 비율은 2004년 11월에는 66.7%, 2004년 12월에는 66%로 줄었다. 이 같이 파일의 선호도가 변화하는 현상은 다른 달의 상위 1000등의 파일을 기준으로 분석한 경우에 관찰된다. 분석된 서버가 제공하는 교육용 콘텐츠가 대학의 정규 과목과 동일한 학사 일정으로 운영된다. 학기 초에는 1강, 2강의 콘텐츠들이 자주 접근되지만, 학기 말에는 14강, 15강의 콘텐츠들이 자주 접근되기 때문이다. 본 논문의 분석 결과에 따르면 파일의 선호도는 월 30% 내에서 변화하고, 변화의 폭은 시간이 지날수록 줄어든다.

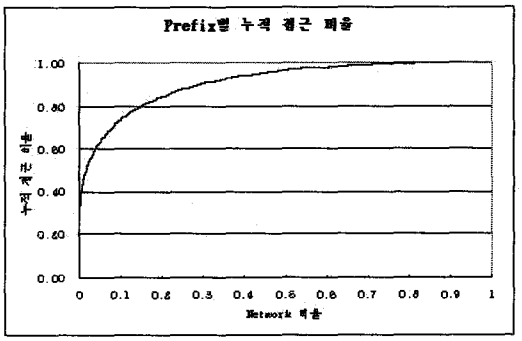
[표 4] 월별 선호도 상위 1000개 파일들의 전이

	2004년 10월	2004년 11월	2004년 12월
2004년 9월	817	667	660
2004년 10월	×	739	696
2004년 11월	×	×	923

4.3.5 네트워크별 접근 빈도



(a) 네트워크별 요청 비율 $\log_{10}$



(b) 네트워크별 누적 요청 비율

[그림 7] 네트워크별 클라이언트 전송 요청 분석

[그림 7]의 (a)는 전송 요청을 한 클라이언트가 속한 네트워크를 IP주소의 프리픽스(prefix)를 이용하여 분류한 후, 각각의 네트워크별로 전송 요청 비율을 계산하여 $\log_{10}$ 그래프로 나타내었다. [그림 7]의 (b)는 클라이언트로 전달된 요청들을 네트워크별로 구분하고, 네트워크별 요청 비율에 대한 누적 그래프를 그린 것이다. 분석 결과 전송 요청이 많은 상위 10%의 네트워크들로부터 전체 요청의 78%가 제기되었음을 알 수 있었다. 소수의 네트워크들로부터 제기된 요청이 서버 부하의 대부분을 차지하고 있으므로 많은 요청을 발생하는 네트워크 내에 프록시 서버가 존재하면 e-learning 서버의 부하를 상당히 줄일 수 있다.

5. 결론 및 향후 연구과제

본 논문에서는 e-learning 서버의 부하 분석을 통해 교육용 서버의 통계적 부하 모델을 작성할 수 있는 기초 자료를 제공하였다. 분석 결과를 정리하면 다음과 같다.

첫째, 파일의 크기 분포는 조사 기간 내에 사용자가 접근한 파일들 전부를 대상으로 하면, 파레토 분포를 보였다. 하지만 동영상 파일들만을 분석한 결과를 보면, 100MB에서 200MB 사이의 구간에 많은 수가 존재하여, 파레토 분포를 따르지 않았다. 이것은 이전의 관련 연구들에서는 관측되지 못한 것이다. 분석 대상이 된 파일들을 생성하기 위해 온라인 강의 저작 도구를 사용했기 때문에 강의 시간에 따라 파일의 크기가 결정되었기 때문이다.

둘째, 사용자 요청에 따른 처리 결과와 요청된 파일들의 유형에 따른 접근 빈도와 전송량을 제시하였다. 멀티미디어 파일의 접근 비율이 0.04% 였지만, 전송량 비율은 17.92%가 되는 것을 볼 때, 멀티미디어 파일의 접근이 서버와 네트워크에 많은 부하를 준다는 것을 알았다.

셋째, 파일의 접근 빈도는 이전 연구들에서는 파레토 분포를 따르는 것으로 알려졌지만, 본 논문에서는 접근 횟수가 100에서 200번 사이의 파일들의 접근 비율이 높은 것으로 측정되었다. 이것은 대학에서 온라인으로 수업을 진행하는 경우 한 과목의 수강생 수가 물리적 공간의 제한을 받지 않기 때문에 수백 명의 수강생들을 가진 과목들이 존재하기 때문이다.

넷째, 시간에 따른 접근 빈도를 보면, 오전 9시부터 오후 4시까지 높은 접속 빈도를 보이다가 점차 낮아진 뒤, 오후 7시부터 자정까지 다시 증가하는 패턴을 보였다. 또한 요일별로는 수요일 오후보다 일요일 야간의 접속이 많은 것이 관찰되었다.

다섯째, 한 세션에서 파일의 접근 간격을 살펴보면, 사용자들이 180초 이내에 다른 파일을 요청하는 비율이 70%이상 되는 것을 알았다,

여섯째, 파일의 선호도를 분석한 결과 이전 연구들에서 나타난 Zipf 법칙을 따르는 것을 확인하였다. 또한 파일의 선호도가 시간에 따라 변화됨을 월별 파일의 선호도 순위 분석을 통해 확인하였다. 분석된 서버가 제공하는 콘텐츠가 대학의 정규 학사일정에 종속되기 때문에 수강생들이 학기 초와 학기 말에 주로 접근한 파일들이 달라졌기 때문이다.

일곱째, 서버에 접근하는 클라이언트들을 네트워크별로 나누어 접속 비율을 측정한 결과 요청이 많은 상위 10%의 네트워크로부터의 접근 비율이 78%에 이름을 확인하였다. 따라서 프록시 서버를 적절히 배치하면 서버의 부하를 효과적으로 감소시킬 수 있다.

본 논문에서는 단국대학교의 e-learning 서버의 부하 분석을 통해 일반적인 웹 서버나 교육용 서버의 부하 분석들에서 관측되지 못한 몇 가지 사실을 확인하였다. 본 논문에서 제시한 분석 결과는 지속적으로 사용자가 증가하고 있는 e-learning 시스템의 구조와 서버 성능 향상을 위한 연구의 기초 자료를 제공한다. 또한 e-learning 서버의 통계적 부하 모델을 작성을 위한 토대를 마련하였고, 이 자료를 바탕으로 e-Learning 시스템을 운영하거나 설계하는 산업체의 운영모델을 구축하는데 많은 도움을 줄 것이다. 향후 연구 과제로는 여러 대학들의 e-learning 서버의 작업부하를 분석하여, e-learning 서버의 부하 모델을 정의하고 부하 생성기를 제작하여 서버의 작업부하를 예측하는 시뮬레이션툴을 제작하는 것이다.

참고문헌

[1] L. Cherkasova, M. Gupta, "Analysis of enterprise media server workloads: access patterns, locality, content evolution, and rates of change", IEEE/ACM Transactions on Networking (TON), Vol.12, Issue 5, pp.781-794, October, 2004.

[2] J. M. Almeida, J. Krueger, D. L. Eager, and M. K. Vernon, "Analysis of Educational Media Server Workloads", In Proc. 11th Int'l. Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV 2001), June 2001.

[3] M. Chesire, A. Wolman, G. Voelker, H. Levy, "Measurement and Analysis of a Streaming Media

Workload", In Proc. of the 3rd USENIX Symp. on Internet Technologies and Systems (USITS), March 2001.

[4] S. Acharya, B. Smith, P. Parnes, "Characterizing User Access To Videos On The World Wide Web", In Proc. Multimedia Computing and Networking 2000(MMCN00), San Jose, CA, January 24-26, 2000.

[5] J. Padhye, J. Kurose, "An empirical study of client interactions with a continuous media courseware server", In Proc. of ACM Workshop on Network and Operating System Support for Digital Audio and Video(NOSSDAV), July 1998.

[6] V. Almeida, A. Bestavros, M. Crovella, A. de Oliveira, "Characterizing Reference Locality in the Web". in Proc. IEEE Conference on Parallel and Distributed Information Systems, pages 92-107, Miami Beach, Florida, December 1996.

[7] L. Breslau, P. Cao, L. Fan, G. Phillips, S. Shenker, "Web Caching and Zipf-like Distributions: Evidence and Implications", in Proc. IEEE Infocom'99, New York City, NY, March, 1999.

[8] Web Server Workload Characterization : The Search for Invariants (Extended Version)

[9] Characterization of a Large Web Site Population with Implications for Content Delivery

[10] T.Spangler, "Promising Satellite Services Emerge as Alternative to Earthbound Lines," Internet World, Mar.9 1998.

[11] P. Barford, M. Crovella, "Measuring Web performance in the wide area", ACM SIGMETRICS Performance Evaluation Review Vol.27, Issue 2, pp.37-48, Sept. 1, 1999.

[12] G. K. Zipf, "Human Behavior and the Principle of Least Effort", Addison Wesley Cambridge, MA, 1949

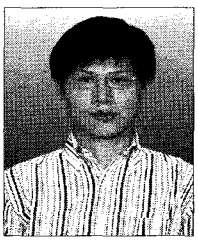
[13] 김기완, 이승원, 박성호, 정기동, "인터넷 환경에서의 상용 연속 미디어 서버의 부하 분석", 정보처리학회 논문지 제10-B권 제1호, pp. 87-94, 2003. 2.

[14] 김병섭,김홍연,김영철,원종호,이미영, "VOD 전용 파일 시스템 개발 및 성능 분석", 제22회 한국정보처리학회 추계학술발표대회 논문집 제11권 제2호 pp.191-194, 2004.11

[15] 정성무, 이상용, 장중순, 송재신, 유해영, 최경희, "TTT 타점법을 이용한 웹서버 파일 분포의 후미성 분석", 정보처리학회논문지A 제10-A권 제3호, pp. 189-197, 2003. 8.

[16] Weisstein, Eric W. "Pareto Distribution.", http://mathworld.wolfram.com/ParetoDistribution.html

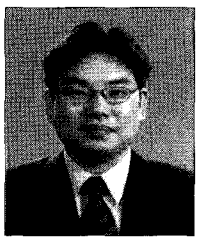
손 세 일(Sei-Il Son) **[정회원]**



- 1993년 2월 : 유한전문대 전자계산과
- 1997년 2월 : 방송통신대 전자계산학과 (이학사)
- 1999년 2월 : 단국대학교 전산통계학과 (이학석사)
- 2007년 2월 : 단국대학교 전산통계학과 (이학박사)
- 1993년 5월 ~ 1996년 5월 : 상지전산(주) 개발부
- 2002년 3월 ~ 2006년 2월 : 단국대학교 정보컴퓨터학부 강의전임강사
- 2007년 2월 ~ 현재 : (주)씨디네트웍스 연구소 차장

<관심분야>
P2P, CDN, IPTV, 트래픽 분석, 데이터 동기화

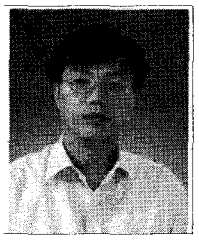
안 효 범(Hyo-Beom Ahn) **[정회원]**



- 1992년 2월 : 단국대학교 전자계산학과(이학사)
- 1994년 2월 : 단국대학교 전산통계학과(이학석사)
- 2002년 2월 : 단국대학교 전산통계학과(이학박사)
- 1997년 9월 ~ 2005년 3월 : 천안공업대학 정보통신학부 부교수
- 2005년 3월 ~ 현재 : 공주대학교 정보통신학부 부교수

<관심분야>
네트워크 보안, RFID 보안, Ad-hoc 네트워크 보안, 시스템 성능분석

김 흥 준(Huenh-Jun Kim) **[정회원]**



- 1989년 : 단국대학교 전자계산학과 졸업(학사)
- 1993년 : 단국대학교 대학원 전산통계학과 (석사)
- 1999년 : 단국대학교 대학원 전산통계학과 (박사)
- 1999년 3월 ~ 현재 : 진주산업대학교 컴퓨터공학부 부교수

<관심분야>
컴퓨터구성, 모바일 네트워킹, 전자상거래