

인터넷 환경에서의 상용 연속미디어 서버의 부하 분석

김 기 완[†]·이 승 원^{††}·박 성 호^{†††}·정 기 동^{††††}

요 약

사용자의 접근 형태에 따른 서버 부하의 특성에 관한 연구는 연속미디어 캐싱 정책과 서버 및 네트워크의 부하 분산 정책에 대한 통찰력을 제공한다. 본 논문에서는 현재 인터넷을 통해 연속미디어 파일을 실시간 스트림 형식으로 제공하는 세개의 상용 사이트의 로그데이터를 이용하여, 각각의 서버가 보유한 연속미디어 파일의 특성과 각 연속미디어 파일에 대한 사용자 접근 요구의 특성을 분석하였다. 이들 서버는 이전에 발표되어진 연구의 서버보다 많은 연속미디어 파일을 보유하고 있으며, 매우 많은 사용자 접근 요구를 처리하고 있다. 각 서버가 보유한 연속미디어 파일의 특성은 파일의 크기, 재생 시간, 인코딩 대역폭 등을 통해 분석하였으며, 사용자 접근 요구의 특성은 연속미디어 파일에 대한 사용자 요구 분포, 사용자의 접근 시간 분포, 파일의 인기도에 따른 접근율 분포, 시리즈 연속미디어 파일에 접근 요구수 분포 등을 통해 분석하였다.

Analysis of Commercial Continuous Media Server Workloads on Internet

Ki-Wan Kim[†] · Seung-Won Lee^{††} · Seong-Ho Park^{†††} · Ki-Dong Chung^{††††}

ABSTRACT

A study on the characteristics of server workloads based on user access pattern offers insights for the strategies on continuous media caching and network workloads distribution. This paper analyses characteristics of continuous media files in each server and user access requests to each of them, using log data of three commercial sites, which are providing continuous media files in the form of real time streaming on the Internet. These servers have more continuous files than ones in the previously reported studies and are processing very large number of user access requests. We analyse the characteristics of continuous media files in each server by the size of files, playback time and encoding bandwidth. We also analyse the characteristics of user access requests by the distribution of user requests to continuous media files, user access time, access rate based on the popularity of the files and the number of access requests to serial continuous media files.

키워드 : 연속미디어(Continuous Media), 부하 분석(Workload Analysis)

1. 서 론

오늘날 인터넷 방송국과 VoD서비스, 연속미디어 포털 사이트, 연속미디어 광고등 다양한 형태의 연속미디어 서비스들이 증가하고 있으며, 연속미디어를 제공하는 업체들은 보다 많은 사용자에게 보다 높은 품질의 데이터를 제공하기 위하여 노력하고 있다. 최근의 컴퓨터 및 네트워크 관련 기술의 발전과 연속미디어 응용 기술의 급속한 발전은 고품질의 연속미디어를 인터넷을 통해 서비스할 수 있는 기술을 제공한다[9]. 그러나 인터넷 사용자의 폭발적인 증가는 서버의 부하, 네트워크의 혼잡 등을 유발하여 연속미디어

데이터의 고유한 특성인 실시간성을 보장하는데 매우 많은 문제점을 발생시키고 있다. 이러한 문제를 해결하기 위한 기초 단계로써 연속미디어 서버 부하의 특성에 관한 연구가 수행되고 있다. 사용자의 접근 형태에 따른 서버 부하의 특성에 관한 연구는 연속미디어 캐싱 정책과 서버 및 네트워크의 부하 분산 정책에 대한 통찰력을 제공한다.

최근의 한국인터넷정보센터의 조사결과[10]에 따르면 국내에서는 인구의 58%가 인터넷을 사용하고 있으며, PC를 보유하고 있는 가정의 82%가 ADSL, 케이블망 등의 초고속 인터넷망을 사용하고 있다. 또한 유료 콘텐츠의 사용자는 작지만 계속 증가 추세에 있으면, 이들 유료 사용자는 영화, 음악, 동영상 교육 콘텐츠 등을 선호하는 것으로 나타나고 있다. 그 외에 홈쇼핑, 공중파 방송 등에서도 인터넷을 통해 자사가 생성한 연속미디어 콘텐츠를 제공함으로써 인터넷을 통한 연속미디어의 서비스 환경을 넓혀가고

† 정 회 원 : 육군3사관학교 교수부 전산정보처리학과 교수
 †† 준 회 원 : 부산대학교 대학원 전자계산학과
 ††† 준 회 원 : 부산대학교 전자계산소 교수
 †††† 종신회원 : 부산대학교 전자계산학과 교수
 논문접수 : 2002년 8월 27일, 심사완료 : 2002년 10월 30일

있다. 그러므로 국내 연속미디어 서버 부하의 분석은 연속 미디어 서비스 환경을 연구하는 기초 자료가 될 것이다.

본 논문은 현재 국내에서 상용 서비스를 제공하는 사이트의 연속미디어 서버의 로그 데이터를 분석하였다. 이 서버들은 기존에 발표되어진 연구[1-4]와 달리 고품질의 데이터를 다수의 사용자를 대상으로 서비스를 제공하고 있다. 본 논문은 서버가 보유하고 있는 연속미디어 데이터의 특성, 각 서버별 파일 크기에 따른 요구분포, 인코더 대역폭에 따른 요구분포, 연속미디어 객체에 대한 접근 빈도, 각 사용자의 사용자 요구의 도착 시간 분포 등을 비교 분석한다.

본 논문의 구성은 다음과 같다. 2장에서 현재까지 진행되어 온 관련연구에 대하여 고찰하고 3장과 4장에서는 서버 부하의 특성과 사용자 접근의 특성을 각각 제시한다. 그리고 마지막 5장에서는 결론을 기술한다.

2. 관련 연구

현재, 텍스트나 이미지와 같은 웹 데이터가 인터넷을 통해 사용자에게 제공하는 웹 서버 및 웹 사용자의 부하에 관한 연구는 많이 이루어졌다[5-7]. 그러나 웹 서버 부하에 관한 연구는 연속미디어 데이터에 대한 사용자 접근의 특성을 반영하지 못하고 있다. 또한 최근에 발표된 연속미디어 데이터를 대상으로 하는 서버의 부하 분석은 대부분 대학 캠퍼스와 같은 제한된 공간에서 특정 사용자들에 의해 접근되는 서버의 로그 데이터를 중심으로 이루어졌다. [8]에서는 매사추세츠(Massachusetts) 대학의 MANIC(Multimedia Asynchronous Networked Individualized Courseware) 시스템의 부하 및 사용자 접근 패턴을 분석하였다. MANIC 시스템은 오디오로 저장된 강의 파일과 텍스트로 작성된 강의 노트를 200명 이상의 학생들에게 제공하였다. 이 서버는 개별 세션을 통해 인터랙티브 사용자에게 오디오의 시작, 정지, 일시 정지 그리고 오디오 스트림내에서의 인덱스된 점핑(Jumping)등 기능을 제공하며, 이 서버에서 측정된 데이터에 대한 분포를 다양한 수학적인 분포 함수와 비교하여 적절한 인수(Parameter) 값을 제시하였다. [3]에서는 오디오, 비디오 그리고 화이트 보드로 구성된 강의를 제공하는 Classroom 2000 Server 시스템의 부하, 제공되는 데이터의 특성 그리고 사용자 접근 패턴을 분석하였다. 이 Classroom 2000 Server는 이미지나 펜 스트로크 파일들과 같은 이산파일에 대한 접속과 연속미디어에 대한 접근들이 혼재하고 있으며, 평균적으로 하나의 학습 세션은 이산파일 32개, 오디오파일 1개, 그리고 전체 데이터의 약 60%를 차지하는 1개의 비디오 파일로 구성된다. 이 서버에서의 연속미디어 파일에 대한 접근은 전체 접근 중 아주 작은 횟수지만, 서버로부터 전송된 모든 데이터량의 79%이상을 차지한다. [1]에서는 대학 캠퍼스내의 서버에서 강의 비디오 파일과 오락 비디오 파일을 동시에 서비스하는 mMod(multicast Media On Demand)

시스템의 사용자 접근 로그 파일의 분석을 통하여 평균 크기 분포, 파일의 인코딩 대역폭, 지속시간 등의 부하를 분석하였다. 이 mMod 시스템의 대학 캠퍼스 내의 높은 네트워크 대역폭을 활용하여 다양한 대역폭의 비디오 파일을 스트리밍 형식으로 제공한다. [2]에서는 대학내의 인터넷 환경에서 트래이스 로그 데이터를 분석하여 연속미디어 서버 부하의 특성과 전통적인 웹 서버 부하의 특성을 조사하였다. [4]에서는 eTeach 시스템과 BIBS(Berkeley Internet Broadcasting System)의 로그데이터 분석을 통해 사용자 요구들간의 도달 시간의 분포, 인기도에 따른 접속빈도의 시간적, 공간적 지역성(Locality)에 대한 평가, 그리고 서버의 로드를 줄이기 위한 멀티캐스트 기술의 성능을 비교 분석하였다.

3 서버 부하의 특성

3.1 상용 서버의 환경

본 논문에서는 현재 인터넷을 통해 연속미디어 스트림을 실시간으로 제공하는 세 개의 사이트로부터 협조를 구하여 획득한 로그데이터를 분석하였다. A서버는 공중파 방송을 통해 방송한 연속미디어 데이터를 인터넷을 통해 제공하는 서버로 드라마, 음악프로그램, 오락 프로그램 등 다양한 분야의 데이터를 보유하고 있다. B서버는 공중파 교육 방송을 통해 방송한 연속미디어 데이터를 인터넷을 통해 제공하는 서버로 교육 분야의 콘텐츠를 보유하고 있다. C서버는 인터넷신문사에서 운영하는 서버로 뮤직 비디오, 영화 예고편 등을 인터넷을 통해 사용자에게 제공하고 있다. <표 1>은 각 사이트의 연속미디어 서버 환경을 보여준다. A서버의 로그 데이터는 20대의 연속미디어 서버중 임의의 1대를 선택하여 21일간의 트래이스 데이터를 수집하였다. B서버의 로그 데이터는 5대의 연속미디어 서버중 임의의 1대를 선택하여 9일간의 트래이스 데이터를 수집하였고, C서버는 한대의 연속미디어 서버를 운용하고 있으며, 19일간의 트래이스 데이터를 수집하였다. A서버는 Windows NT를, B, C서버는 Windows 2000 운영체제를 사용하고 있으며, 스트리밍 소프트웨어는 Windows Media Server를 사용한다.

<표 1> 각 사이트의 연속미디어 서버 환경

구 분	A 사이트	B 사이트	C 사이트
로그 서버	임의의 1대/ 20대 운용	임의의 1대/ 5대 운용	1대/1대 운용
서버의 운영체제	Windows NT	Windows 2000	Windows 2000
서버의 Streaming S/W	Windows Media Server	Windows Media Server	Windows Media Server
측정기간	2001. 5. 12~ 2001. 5. 31 (21 일간)	2002. 4. 17~ 2002. 4. 25 (9 일간)	2002. 2. 13~ 2002. 3. 3 (19 일간)
주요 콘텐츠	드라마, 오락, 뉴스, 등	교육, 강좌 등	영화예고편, 뮤직비디오

3.2 연속미디어 파일의 특성

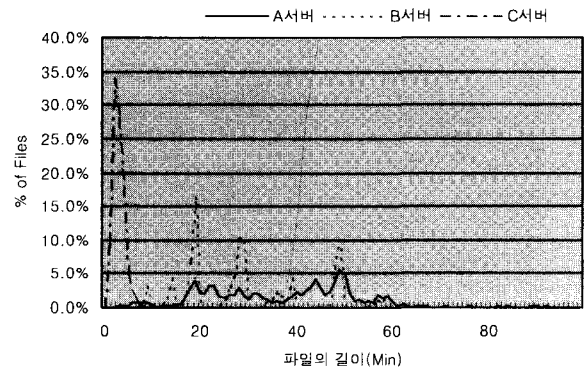
<표 2>는 각 서버가 보유한 연속미디어 파일에 대한 통계를 보여준다. A와 B서버는 매일 정기적으로 새로운 연속미디어 파일을 생성하며, C서버는 비정기적으로 새로운 연속미디어 파일을 생성한다. 각 연속미디어 서버의 로그는 시작 접근 요구(Start Access Request)와 임의 접근 요구(Random Access Request)를 모두 포함하고 있다. 시작 접근 요구는 사용자가 연속미디어 파일의 첫 위치부터 재생할 것을 요청하는 요구이며, 임의 접근 요구는 연속미디어 객체의 임의의 위치부터 재생할 것을 요청하는 요구이다. 또한 A, B, C 서버는 이전에 발표되어진 연구[4,8]보다 많은 연속미디어 파일을 보유하고 있으며, 매우 많은 요구를 처리하고 있다. 특히, C서버는 A, B서버에 비하여 비교적 짧은 연속미디어 파일을 제공하고 있으며, 요구에 대한 재생 비율이 높은 것으로 나타난다. 그리고 서버와 네트워크 환경의 부하 때문에 요구된 파일의 평균 재생시간 비율과 요구된 파일의 평균 전송 비율에 차이가 발생한다.

<표 2> 각 서버의 연속미디어 파일에 대한 통계표

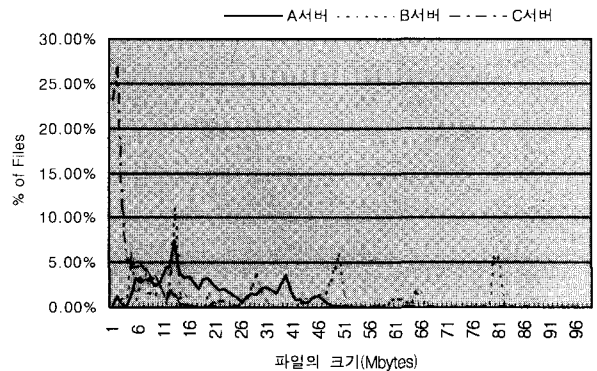
구 분	A 서버	B 서버	C 서버
측정기간에 접근된 연속미디어 파일의 수	5,058	10,868	1,488
측정기간(hour)	477	212	454
전체 요구의 수	1,969,737	556,663	164,376
시작 접근 요구수	452,216	132,699	151,762
임의 접근 요구수	1,517,521	423,964	12,614
시간당 평균 요구수	4,129	2,626	362
연속미디어 파일의 평균 길이(min)	38.4	30.2	3.1
연속미디어 파일의 평균 크기(Mbyte)	21.3	35.5	3.5
파일의 평균 전송 대역폭(Kbps)	73.9	156.8	147.0
요구된 파일의 평균 재생 시간(min)	3.4	3.2	0.7
요구된 파일의 평균 재생 시간 비율 (요구된 파일의 재생시간/파일의 길이)	9.26%	13.86%	37.92%
요구된 파일의 평균 전송량(Mbyte)	1.3	4.7	1.4
요구된 파일의 평균 전송 비율 (요구된 파일의 전송량/파일 크기)	5.69%	13.70%	25.94%

(그림 1)과 (그림 2)는 각 서버별 연속미디어 파일의 재생 시간 분포와 서버별 연속미디어 파일의 크기 분포를 보여준다. 비교적 짧은 비디오 클립을 제공하는 C서버는 1분~6분 사이의 재생 시간에 연속미디어 파일의 96%가 분포하고, A서버는 18분~60분 사이에 91%가 분포한다. 그리고, B서버는 99%의 연속미디어 파일이 5분~60분 사이에 분포되어 있으며, 5분 또는 10분간격의 이산적인 분포를 보인다. 또한 C서버는 98%의 연속미디어 파일이 1Mbyte에서 15Mbyte사이

에 분포한다. A서버는 2Mbyte~46Mbyte 사이에 95%의 연속미디어 파일이 분포하며, B서버는 3Mbyte~82Mbyte 사이에 97%의 연속미디어 파일이 존재한다. 그러나 동일한 대역폭으로 인코딩된 연속미디어 파일은 파일의 길이와 재생시간이 비례함에도 불구하고 (그림 1)과 (그림 2)에서 동일한 서버의 연속미디어 파일의 재생시간 분포와 연속미디어 파일의 크기 분포는 매우 상이함을 보인다. 이는 동일한 서버 내에 존재하는 연속미디어 파일의 인코딩 대역폭이 상이함을 나타낸다.



(그림 1) 각 서버별 연속미디어 파일의 재생 시간 분포

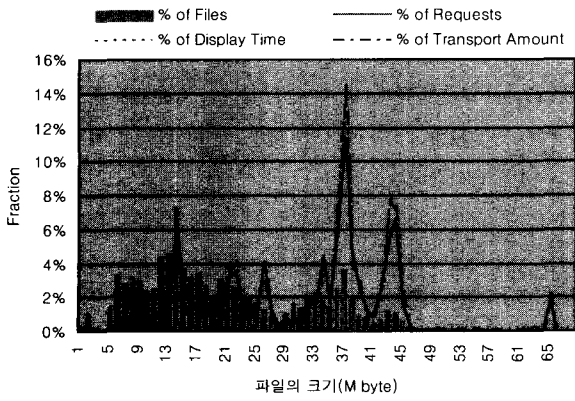


(그림 2) 각 서버별 연속미디어 파일의 크기 분포

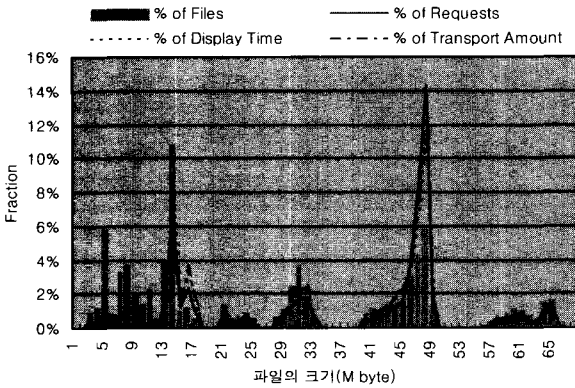
4. 사용자 접근의 특성

4.1 연속미디어 파일에 대한 사용자 요구 분포

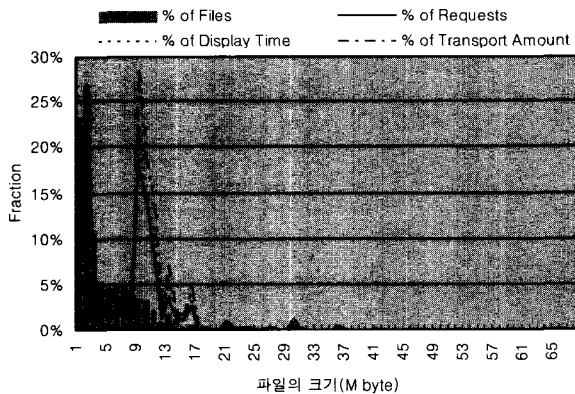
(그림 3)~(그림 5)는 파일의 크기에 따른 사용자 요구 빈도, 각 요구의 재생 시간, 그리고 전송된 데이터 량의 비율의 분포를 보여준다. 연속미디어 파일에 대한 사용자 요구 분포는 파일의 크기가 클수록 파일수의 비율 분포보다 사용자 요구수의 비율 분포가 높게 나타나며, 요구수의 비율 분포보다 전송된 데이터 량의 비율 분포가 높게 나타난다. eTeach 시스템과 BIBS 시스템의 분석[4]에서도 나타나는 이러한 분포는 파일의 크기가 클수록 대화형 사용자의 임의 접근 요구가 많음을 나타낸다.



(그림 3) A서버의 파일 크기에 따른 요구 분포



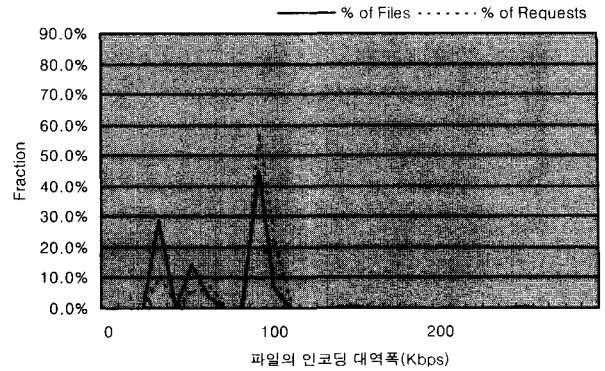
(그림 4) B서버의 파일 크기에 따른 요구 분포



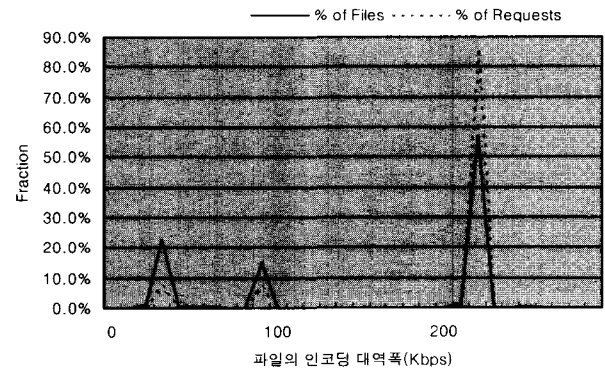
(그림 5) C서버의 파일 크기에 따른 요구 분포

(그림 6)~(그림 8)은 파일의 인코딩 대역폭에 따른 사용자 요구의 분포를 나타낸다. A서버의 연속미디어 파일은 약 30, 50, 100Kbps의 대역폭으로 인코딩 되었고 B서버의 연속미디어 파일은 약 30, 100, 220Kbps의 인코딩 대역폭을 가지고 있다. 그리고 C서버의 연속미디어 파일은 약 30, 80, 200, 300, 500Kbps의 다양한 대역폭으로 인코딩 되어 있다. 파일이 높은 대역폭으로 인코딩 될수록 파일 수의 비율에 비해 사용자 요구수의 비율이 높음을 알 수 있다. 이러한

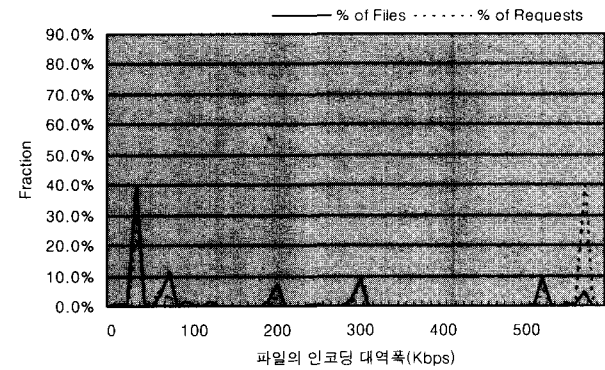
분포는 사용자들이 높은 대역폭으로 인코딩 된 파일에 대한 선호도가 높음을 보여준다.



(그림 6) A서버의 파일의 인코딩 대역폭에 따른 요구 분포



(그림 7) B서버의 파일의 인코딩 대역폭에 따른 요구 분포

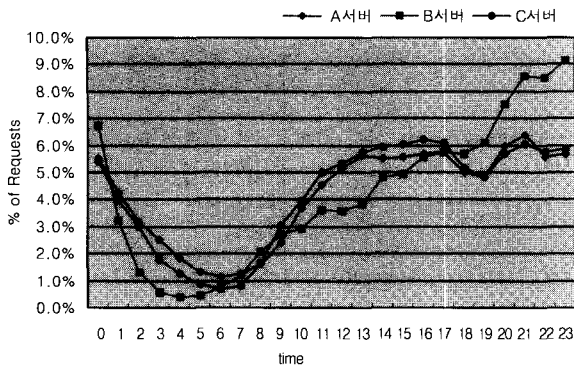


(그림 8) C서버의 파일의 인코딩 대역폭에 따른 요구 분포

4.2 사용자의 접근 시간 분포

(그림 9)는 측정기간 동안의 연속미디어 파일의 접근 시간 분포를 나타낸다. 사용자의 접근시간 분포는 오전 10시부터 익일 오전 2시 이전까지는 3%이상의 높은 접근율을 가지는 반면, 오전 2시부터 오전 10시 이전까지는 상대적으로 3%이하의 낮은 접근율을 보인다. 또한 콘텐츠의 내용이 주로 오락물인 A서버와 C서버는 유사한 분포를 보이는 반

면, 교육 콘텐츠를 서비스하는 B서버는 야간시간대(오후 6시~익일 오전 1시)의 요구수에 대해 다른 서버에 비하여 높은 편중을 보인다.

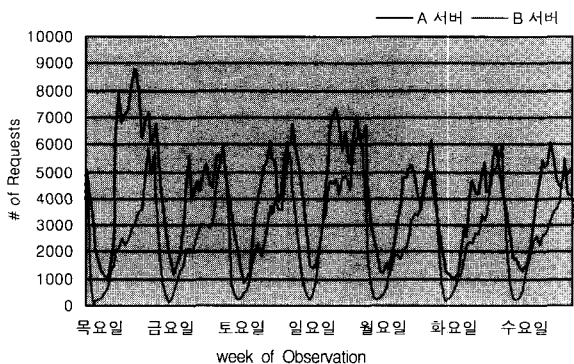


(그림 9) 측정기간 동안의 사용자 요구 발생 시간 분포

(그림 10)은 A서버와 B서버의 주(週) 단위 사용자 요구 빈도의 분포를 비교하여 보여준다. 이 분포의 측정 환경은 <표 3>과 같다. 교육용 콘텐츠를 서비스하는 B서버는 요일과 시간대에 따른 요구수의 변화가 적으나, 드라마 등 오락 콘텐츠를 제공하는 A서버는 요일과 시간대에 따른 요구수의 변화가 많다.

<표 3> 주 단위 관측의 환경

구 분	A 서버	B 서버
측정 기간	2001. 5. 17 (목)~ 2001. 5. 23 (수)	2002. 4. 18 (목)~ 2002. 4. 24 (수)
측정기간의 전체요구수	691,501	443,385



(그림 10) 주간 사용자 요구 빈도 분포

4.3 파일의 인기도에 따른 접근율 분포

(그림 11)과 (그림 12)는 각 서버별 파일의 인기도에 따른 접근율 분포를 나타낸다. 각 분포는 측정기간으로 1시간, 1일(24시간), 전체로 구분하여 측정하였으며, 측정기간에 따른 사용자 접근 통계표는 <표 4>와 같다. 그리고 인기도에 따른 접근율 분포를 수학적 분포와 비교하였다. 이 분포와 비교

된 수학적 분포는 지수분포¹⁾($\lambda = 1/16$)와 Zipf 분포($\theta = 0.271$)이다. 지수분포의 λ 와 Zipf 분포의 θ 는 χ^2 분포 검정 통계량 값의 합²⁾이 최소일 때의 값으로 설정하였다. 일반적으로 지수분포는 시스템이나 부품의 수명, 생존시간 또는 대기시간 등에서 볼 수 있는 분포이며, Zipf 분포[11]는 멀티미디어 파일의 선호도를 분석하는데 많이 사용되고 있다. Zipf 분포의 확률 밀도함수는 (그림 11) 같이 표현된다.

Skew(θ)를 가진 Zipf 분포

$$i \text{ 번째 값에 대한 선택 분포: } F(i) = \frac{1}{i^{(1-\theta)}} \quad (1 \leq i \leq n)$$

(n : 도메인(Domain) 원소의 개수)

$$i \text{ 번째 값에 대한 선택 확률: } P^{Zipf}(i) = \frac{F(i)}{\sum_{j=1}^n F(j)}$$

(그림 11) Zipf 분포

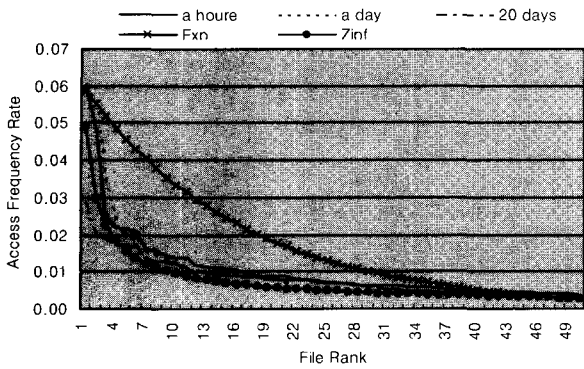
<표 4> 측정기간에 따른 사용자 접근의 통계표

구 분		A 서버	B 서버	C 서버
측정기간	1시간	2001/5/21 23:00~24:00	2002/4/21 23:00~24:00	2002/2/14 23:00~24:00
	1일	2001/5/21	2002/4/21	2002/2/14
	전체	2001/5/12 ~5/31(21일)	2002/4/17 ~4/25(9일)	2002/2/13 ~3/3(19일)
측정기간별 요구수	1시간	4,904	5,782	512
	1일	113,437	81,112	8,219
	전체	1,969,737	556,663	164,376
측정기간에 요구된파일수	1시간	708	705	130
	1일	2,939	3,562	623
	전체	5,058	10,868	1,488

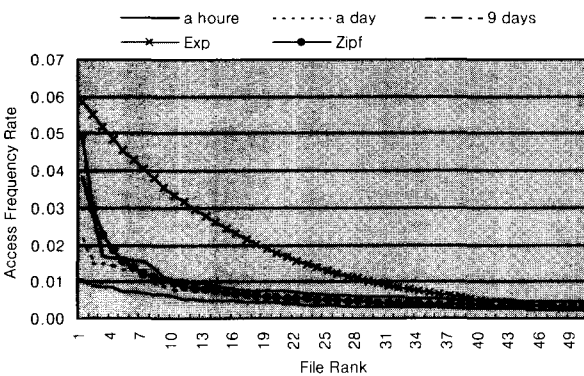
(그림 12)와 (그림 13)은 측정기간에 접근된 파일 중 인기도 순서로 50위까지의 분포를 나타낸다. A서버에서는 측정기간이 한시간 일 때의 접근율 분포와 하루(24시간) 일 때의 접근율 분포가 유사하며, 또한 이들 분포는 Zipf 분포($\theta = 0.271$)와 매우 유사함을 보여준다. 그러나 B서버는 측정기간이 한시간 일 때, Zipf 분포($\theta = 0.271$)와 유사한 반면 하루(24시간) 일 때의 접근율 분포는 많은 차이를 보인다. 즉, 드라마 등의 오락 콘텐츠를 제공하는 A서버는 하루(24시간) 내의 시간대에 따른 파일의 인기도의 변화가 작은 반면에, 교육 콘텐츠를 제공하는 B서버는 하루(24시간) 내의 시간대에 따른 파일의 인기도의 변화가 크다. 이는 시간대 별로 교육 콘텐츠를 사용하는 사용자 그룹이 변함을 알 수 있다.

1) 확률밀도함수는 $f(x) = \lambda e^{-\lambda x}$, $0 < x < \infty$ 이다.

2) $\sum_{i=0}^m T_i = \sum_{i=0}^m \frac{(O_i - E_i)^2}{E_i}$, O_i : 관측도수, E_i : 기대도수



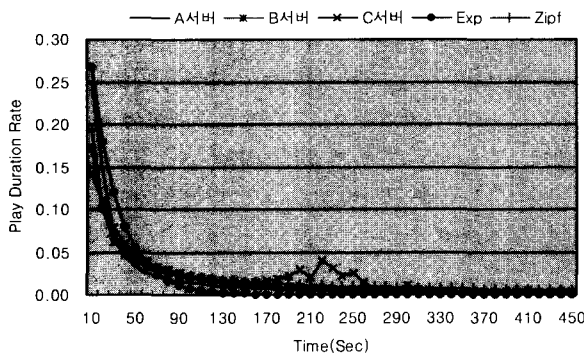
(그림 12) A서버의 파일의 인기도에 따른 접근율 분포



(그림 13) B서버의 파일의 인기도에 따른 접근율 분포

4.4 사용자 요구별 재생 시간 분포

사용자 요구별 재생 시간은 사용자의 요구에 의해 연속 미디어 파일이 재생을 시작하여 종료할 때까지의 시간 간격을 의미한다. 인터넷을 통한 연속미디어 사용자는 서버와 상호작용(Interaction)을 통하여 연속미디어 파일의 임의의 위치에서의 시작과 종료가 가능하다. (그림 14)는 각 서버의 사용자 요구별 재생 시간 분포를 수학적 분포와 비교하여 보여준다. 이 분포와 비교된 수학 분포는 지수분포($\lambda = 2/5$)와 Zipf 분포($\theta = 0.05$)이다. 지수분포의 λ 와 Zipf 분포의 θ 는 χ^2 분포 검정 통계량 값의 합이 최소일 때의 값으로 설정하였다. 각 서버의 사용자 요구별 재생 시간 분포는



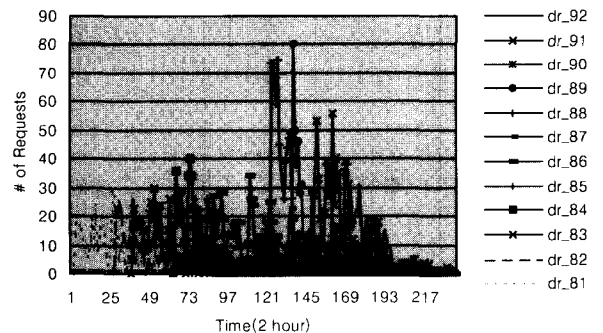
(그림 14) 사용자 요구별 재생 시간 분포

지수분포($\lambda = 2/5$)보다는 Zipf 분포($\theta = 0.05$)에 유사하다. 평균 파일의 길이가 186초인 C서버의 사용자 요구별 재생 시간 분포는 170초~270초 사이에 수학적 분포를 따르지 않고 불규칙하게 재생 시간율이 다시 증가함을 보인다. 이 부분은 전체 파일을 재생하는 사용자 요구의 분포를 나타내며, A서버와 B서버에서도 유사한 분포를 가진다.

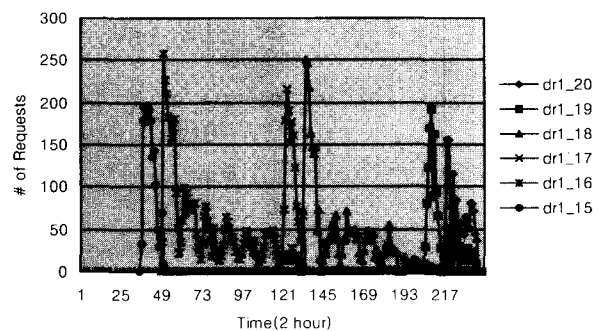
4.5 시리즈 특성을 지닌 연속미디어 파일의 접근 요구수 분포

인터넷에서 서버가 제공하는 연속미디어 파일은 영화, 뮤직비디오 등과 같이 내용적으로 독립적인 연속미디어 파일과 드라마와 같이 내용적으로 연속적인 연속미디어 파일이 존재한다. 본 논문에서는 내용적으로 연속적인 연속미디어 파일을 시리즈라 한다. A서버는 공중파 방송국 콘텐츠를 서비스함으로 시리즈 특성을 가진 연속미디어 파일을 서비스하고 있다. (그림 15)와 (그림 16)은 A서버에서 각각 주 5회, 주 2회 생성되는 시리즈 연속미디어 파일의 접근 요구수의 분포를 보여준다. 연속미디어 객체의 재생시간은 각각 20분과 50분이며 측정 시간간격은 2시간이다. (그림 15)와 (그림 16)에 보듯이 일정량 이상의 접근 수를 가진 연속미디어 객체의 접근 요구 수는 아래와 같은 특성을 가진다.

- 연속미디어 파일의 측정 시간간격 당 접근 요구수는 시리즈의 특성에 따라 큰 차이를 보인다.
- 연속미디어 파일이 생성된 후 1일~2일이 사이에 가장 많은 접근 요구수가 발생한다.



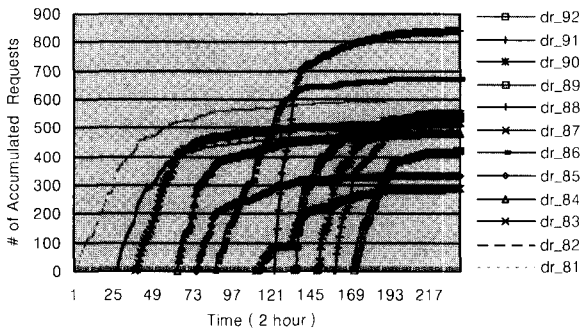
(그림 15) 주 5회 생성되는 시리즈 연속미디어 파일의 요구수 분포



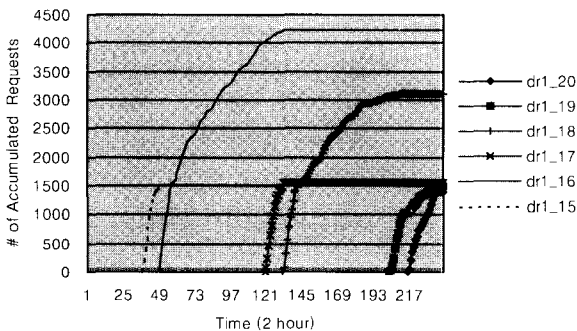
(그림 16) 주 2회 생성되는 시리즈 연속미디어 파일의 요구수 분포

(그림 17)과 (그림 18)은 각각 주 5회, 주 2회 생성되는 시리즈 연속미디어 파일의 접근 요구수의 누적 분포를 보여준다. 접근 요구수의 누적 분포는 연속미디어 파일이 생성된 이후 사용자의 접근 요구 수를 나타낸다. (그림 17)과 (그림 18)에서 보듯이 일정량 이상의 접근 수를 가진 연속미디어 파일의 접근 요구 수 누적 분포는 아래와 같은 특성을 가진다.

- 연속미디어 파일의 접근 요구 수 누적 분포는 파일이 생성된 이후에 일정기간 동안 급격한 증가를 하는 반면, 일정기간의 경과 후에는 증가량이 매우 둔화된다.
- 동일한 시리즈에 속한 연속미디어 파일의 접근 요구 수 누적 분포는 생성된 시점에 따라 약간의 차이는 있으나 유사한 분포를 보인다. 특히, 연속미디어 파일의 접근 요구수의 증가가 둔화되는 시점은 시리즈의 특성에 따라 큰 차이를 보인다.



(그림 17) 주 5회 생성되는 시리즈 연속미디어 파일의 요구수 누적 분포



(그림 18) 주 2회 생성되는 시리즈 연속미디어 파일의 요구수 누적 분포

5. 결 론

본 논문에서는 현재 인터넷을 통해 연속미디어 파일을 실시간 스트림 형식으로 제공하는 세 개의 상용 사이트의 로그데이터를 이용하여, 각각의 서버가 보유한 연속미디어 파일의 특성과 각 연속미디어 파일에 대한 사용자 접근 요구의 특성을 분석하였다. 이들 서버는 이전에 발표되어진 연구보다 많은 연속미디어 파일을 보유하고 있으며, 매우

많은 사용자 접근 요구를 처리하고 있다.

각 서버가 보유한 연속미디어 파일은 서버의 콘텐츠 내용에 따라 다양한 특성을 가지고 있다. 즉, 이들 서버가 보유한 연속미디어 파일은 파일의 크기, 재생 시간, 인코딩 대역폭 등에서 서로 상이한 특성을 가지고 있다. 연속미디어 파일에 대한 사용자 요구 분포 특성은 파일의 크기가 클수록 파일 수의 비율 분포보다 사용자 요구수의 비율 분포가 높게 나타나며, 요구수의 비율 분포보다 전송된 데이터 량의 비율 분포가 높게 나타난다. 그리고 파일이 높은 대역폭으로 인코딩 될수록 파일 수의 비율에 비해 사용자 요구수의 비율이 높음을 알 수 있다. 이러한 분포 특성은 사용자들이 높은 대역폭으로 인코딩 된 파일에 대한 선호도가 높음을 보여준다. 또한 사용자의 접근 시간 분포는 시간대에 따라 많은 격차를 보이며, 콘텐츠의 내용이 주로 오락물인 A서버와 C서버는 유사한 분포를 보이는 반면, 교육 콘텐츠를 서비스하는 B서버는 야간시간에 요구수가 다른 서버에 비하여 편중이 높게 나타난다. 그리고 교육용 콘텐츠를 서비스하는 B서버는 요일과 시간대에 따른 요구수의 변화가 적으나, 드라마 등 오락 콘텐츠를 제공하는 A서버는 요일과 시간대에 따른 요구수의 변화가 많다. 파일의 인기도에 따른 접근율 분포는 측정기간이 짧을수록 Zipf 분포 ($\theta = 0.291$)와 유사하다. 그러나 드라마 등의 오락 콘텐츠를 제공하는 A서버는 하루 내의 시간대에 따른 파일의 인기도의 변화가 작은 반면에 교육 콘텐츠를 제공하는 B서버는 하루 내의 시간대에 따른 파일의 인기도의 변화가 크다. 사용자 요구별 재생 시간 분포는 지수분포($\lambda = 2/5$)보다 Zipf 분포($\theta = 0.05$)에 유사한 것으로 관측되었다. 또한 내용적으로 연결되는 연속미디어 파일은 접근 요구수 분포에서 같은 시리즈 그룹에 속하는 연속미디어 파일 상호간에 밀접한 연관성을 가진다.

참 고 문 헌

- [1] S. Acharya, B. Smith and P. Parnes, "Characterizing User Access To Videos On The World Wide Web," in Proc SPIE/ACM Conf. on Multimedia Computing and Networking, San Jose, CA Jan., 2000.
- [2] M. Chesire, A. Wolman, G. Voelker and H. Levy, "Measurement and Analysis of a Streaming Media Workload," Proc 3rd USENIX symp. on Internet Technologies and Systems, San Francisco, March, 2001.
- [3] N. Harel, V. Vellanki, A Chervenak, G Abowd, U. Ramachandran, "Workload of a Media-Enhanced Classroom Server," Proc. IEEE Workshop on Workload Characterization, Oct., 1999.
- [4] Jussara M. Almeida, Jeffrey Krueger, Derek L. Eager and Mary K. Vernon, "Analysis of educational media server workloads," 11th International workshop on Network and Operating Systems support for digital audio and video, Janu-

ary, 2001.

- [5] V. Almeida, A. Bestavros, M. Crovella, and A. de Oliveira, "Characterizing Reference Locality in the WWW," in Proc. IEEE International Conference on Parallel and Distributed Information Systems, Miami Beach, Dec., 1996.
- [6] L. Breslau, P. Cao, L. Fan, G. Phillips and S. Shenker, "Web Caching and Zipf-like Distributions : Evidence and Implications," Proc. IEEE INFOCOM '99, New York City, NY, March, 1999.
- [7] P. Barford, A. Bestavros, A. Bradley and M. Crovella, "Changes in Web Client Access Patterns : Characteristics and Caching Implications," World Wide Web Journal, Special Issue on Characterization and Performance Evaluation, Dec., 1998.
- [8] J. Padhye and J. Kurose, "An Empirical Study of Client Interactions with a Continuous-Media Courseware Server," Proc. NOSSDAV '98, July, 1998.
- [9] J. E. Pitkow and Colleen M. Kehoe, "The GVU's WWW User Survey," http://www.cc.gatech.edu/user_surveys, 1998.
- [10] 송관호, "인터넷 이용자수 및 이용형태조사", <http://www.nic.or.kr/index.html>, 2002.
- [11] D. E. Kunth, "The art of computer programming," Sorting and searching, Addison-Wesley, Vol.3, 1973.



김기완

e-mail : wan1434@pusan.ac.kr
 1982년 성균관대학교 경영학과 졸업(학사)
 1987년 단국대학교 경영대학원 전자정보
 처리전공(석사)
 1993년 부산대학교 대학원 전자계산학과
 박사과정 수료

1982년~1984년 육군중앙경리단 전산실 프로그램장교, 자료과장
 1985년 육군 중앙전산소 개발실 제도분석장교
 1986년~현재 육군3사관학교 교수부 전산정보처리학과 교수
 관심분야 : 병렬처리 및 멀티미디어, VOD시스템



이승원

e-mail : swlee@pusan.ac.kr
 1997년 부산대학교 전자계산학과 졸업
 (학사)
 1999년 부산대학교 대학원 전자계산학과
 (이학석사)
 1999년~현재 부산대학교 대학원 전자계
 산학과(박사과정)

관심분야 : 멀티미디어, 이동통신, VOD

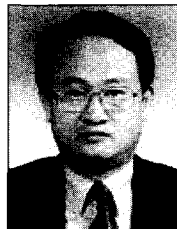


박성호

e-mail : shpark@pusan.ac.kr
 1996년 부산대학교 전자계산학과 졸업
 (학사)
 1998년 부산대학교 대학원 전자계산학과
 졸업(이학석사)
 2002년 부산대학교 대학원 전자계산학과
 졸업(이학박사)

2002년~현재 부산대학교 전자계산소 조교수

관심분야 : VOD 시스템, 인터넷 캐싱, 멀티미디어 이동통신



정기동

e-mail : kdchung@melon.cs.pusan.ac.kr
 1973년 서울대학교 졸업(학사)
 1975년 서울대학교 대학원 졸업(석사)
 1986년 서울대학교 대학원 계산통계학과
 졸업(이학박사)
 1990년~1991년 MIT, South Carolina 대학
 교환 교수

1995년~1997년 부산대학교 전자계산소 소장

1978년~현재 부산대학교 전자계산학과 교수

관심분야 : 병렬처리, 멀티미디어