

VOD 시스템에서 일괄전송을 위한 채널 스케줄링 기법

박 호 균[†] · 유 황 빈^{††}

요 약

주문형 비디오 시스템은 광대역 통신 네트워크를 통하여 사용자가 원하는 비디오를 용통성 있게 제공하는 전자 비디오 대여 시스템이다.

일괄전송 기법에 의하여 동일한 비디오 스트림을 서비스한다고 가정할 때, 하나의 비디오 스트림이 server-end에서 이용 가능할 때마다 어떤 비디오가 어느 순간에 스케줄링할 것인지의 문제가 제기된다. 따라서, 각 스트림마다 변동하는 자원 대역폭을 처리하기 위해서 채널 할당 전략에 관한 것이 중요한 문제로 대두되었다.

본 논문에서는 큐잉을 고려한 세 가지 스케줄링 정책 즉, 누적 대기 시간 기법(Cumulative Waiting Time), 평균 대기 시간(Mean Waiting Time) 기법, 그리고 예측 인기도 가중치(Predictable Popular Factor)에 의한 방법을 제안하였다. 누적 대기 시간 기법은 인기있는 비디오에 대하여 우수한 반면, 평균 대기 시간 기법은 비인기 비디오에 대하여 우수하였다. 또한, 예측 인기도 가중치에 의한 기법은 비디오에 대한 평균 대기 시간과 사용자의 요구 취소율에 대하여 다소 공평함을 알 수 있었다.

Channel Scheduling Policies for Batching in VOD System

Ho-Kyun Park[†] · Hwang-Bin Ryou^{††}

ABSTRACT

Video-On-Demand system provides electronic video rental services from remote video servers on a broadband communication networks.

When we assume some situation that serve the same video stream by batching, each time single video stream became available at server-end side, we face the problem that what video stream to be scheduled at what time instant. Hence, in order to manage the resource bandwidth which is different at each stream, the policy of channel allocation became important factor.

In this work, we proposed three type of scheduling policies with queueing, as are Cumulative Waiting Time(CWT), Mean Waiting Time(MWT) and Predictable Popular Factor(PPF). While CWT policy is good for popular video, MWT policy was good for non-popular video. Also, we know that PPT policy has some fairness for average wait time and renegeing rate of user requests.

* 본 논문은 1998년도 광운대학교 교내연구비 지원에 의해 연구되었음.

† 정 회 원 : 신홍대학 컴퓨터정보계열 교수

†† 종신회원 : 광운대학교 컴퓨터학과 교수

논문접수 : 1998년 4월 6일, 심사완료 : 1998년 9월 11일

1. 개 요

거대한 통신 대역폭과 컴퓨터 기술의 발전으로 인하여 실시간으로 사용자에게 멀티미디어 데이터를 처리하여 전송하는 것이 가능하게 되었다. 특히, ATM을 기반으로 하는 B-ISDN과 대용량 저장 서버를 구현하기 위한 RAID(Redundant Arrays of Inexpensive Disks) 기술 그리고 동영상 압축 기술 등의 발전은 VOD(Video-On-Demand)를 가장 가까운 미래에 실현 가능성이 높은 주거용 서비스 시스템으로 발전시켰다[1, 2, 3, 4, 5].

VOD시스템은 서버에 저장되어 있는 대용량의 비디오 라이브러리를 지리적으로 분산되어 있는 사용자들에게 고속의 통신망을 통하여 재생 서비스를 제공하는 시스템으로 사용자는 언제든지 이 라이브러리로부터 시청하고자 하는 비디오 프로그램을 선택할 수 있다.

T-VOD(True-VOD)에서 사용자는 그들이 원하는 프로그램을 선택하여 원하는 시간에 시청할 수 있으며, 마치 VCR 재생 장치를 이용하여 비디오 테이프를 보듯이 fast-forward, rewind, pause, resume, random positioning 등과 같은 일련의 VCR 동작을 서버에게 요청할 수 있다. 반면, N-VOD(Near-VOD)는 비디오 프로그램을 일정간격으로 방송하여 사용자가 시청하고자 하는 시간에 가장 근접한 프로그램 시간에 채널을 할당하여 서비스하는 방법이다. 따라서, 기존의 비디오 대여점에서 테이프 대여 서비스와 상업적 경쟁력을 유지하기 위해서는 사용자들에게 실시간 대화성을 제공하는 T-VOD로 설계되어야 한다.

T-VOD를 해결하기 위해서는 각각의 사용자들에게 전용 비디오 스트림을 갖는 것이다. 그러나, 이러한 시스템은 동일한 비디오 스트림을 요구하는 다수의 사용자에게 전용 전송 채널로 각각 서비스하므로 비효율적이며 확장성이 없다는 단점을 가지고 있다. 특히, 다수의 비디오 스트림들이 동일한 비디오를 참조하는 다수의 사용자에게 전송할 때, 이것은 각각의 스트림이 고속의 데이터 전송을 요구하기 때문에 많은 비용이 든다. 즉, 시스템에서 사용자 수가 증가함에 따라 할당된 채널 수도 증가하여 제한된 사용자만이 비디오 전송 서비스를 받기 때문이다. 이러한 문제를 해결하기 위하여 동일한 비디오 스트림을 요구한 여러 사용자에게 하나의 스트림을 이용하여 멀티캐스트 전송을 수행하는 일괄전송(batching) 기법이 연구되었다.

그럼에도 불구하고, 일괄전송 기법은 동시에 서비스될 수 있는 스트림의 수가 네트워크 I/O 대역폭에 의하여 여전히 제한된다. 동일한 비디오 프로그램을 시청하고자 하는 많은 사용자 요구들이 일괄전송에 의하여 공통의 비디오 스트림을 이용하여 서비스된다고 가정할 때, 하나의 비디오 스트림이 server-end에서 이용 가능할 때마다 어떤 비디오가 어떤 순간에 스케줄링할 것인가라는 의문이 제기된다. 따라서, 각 스트림마다 변동하는 자원 대역폭을 처리하기 위해서 채널 할당 전략이 필요하다. 일괄전송을 이용하여 server-end에서 스케줄링 기법을 어떻게 설계해야 하는지, 가급적 많은 VOD 사용자들이 비디오 선택 요구를 만족할 수 있도록 스케줄링 전략에 관한 것이 중요한 문제로 대두되었다.

일괄전송 기법을 구현하기 위한 하나의 일반적인 기법은 FCFS 기법이다. 이 기법은 가장 오랫동안 대기했던 요구를 갖는 비디오를 우선적으로 스케줄링한다. 두 번째 일반적인 기법은 가장 많은 대기 요구를 갖는 비디오를 선택하여 서비스하는 MQL(Maximum Queue Length)기법이다. 이들 두 기법은 완전하게 만족한 결과를 얻지 못한다. MQL기법은 서비스될 수 있는 스트림의 수를 최대로 하기 위하여 단지 대기행렬 길이만을 고려함으로써 인기있는 비디오들을 스케줄링하는데 보다 유리한 반면에, FCFS는 대기행렬 길이를 완전히 무시하고 취소율을 감소하도록 도착 시간에 초점을 맞추으로써 반대의 효과를 갖는다. 또한, 그 밖의 몇 가지 개선된 방법들이 연구되었다. 기존의 제한된 채널 할당 전략은 많은 수의 사용자 수용과 공평한 서비스의 두 가지 조건을 만족시키지 못하였다.

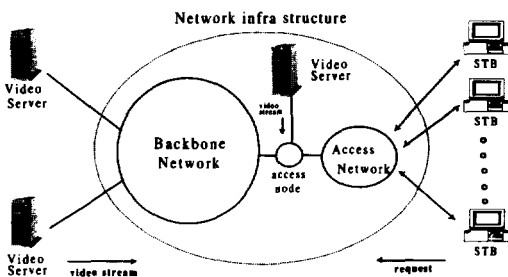
본 논문에서는 비디오를 시청하기 위한 요구들을 다루는 스케줄링 기법에 대하여 연구하였다. 일괄 전송 방식에 의하여 멀티캐스트 기법으로 처리할 비디오를 스케줄링하기 위하여 큐잉을 고려하지 않은 채널 할당 기법과 큐잉을 고려한 채널 할당 기법으로 나누어 제안하였다. 큐잉을 고려하지 않은 채널 할당 기법은 현재 일괄 전송 시간 간격 내에 서비스되지 못한 요구들을 다음 시간 간격에 다시 고려하지 않는 기법이며, 큐잉을 고려한 기법은 현재 일괄 전송 시간 간격 내에 서비스되지 못한 요구들을 다음 시간 간격 내에 다시 스케줄링하기 위하여 요구 버퍼에 저장하는 기법이다. 전자의 스케줄링 기법에는 인기도에 따른 기법과 고정 채널 기법으로 나누어 제안하였으며, 후

자의 스케줄링 기법에는 누적 대기 시간 기법, 평균 대기 시간 기법, 그리고 예측 인기도 가중치 기법을 제안하였다. 사용자 요구가 적은 비디오는 충분치 못한 대역폭으로 인하여 대기 행렬 내에서 서비스 받지 못하는 기아(starvation) 현상이 발생할 수 있으므로, 본 논문은 공평성 문제에 대해서도 고려하였다.

2. VOD 시스템

2.1 VOD 시스템과 서비스 특성

VOD 시스템이란 사용자가 선택한 비디오를 통신망을 통해서 서버로부터 전송 받아 사용자의 요구에 따라 VCR기능을 이용하여 대화형 서비스를 제공하는 시스템이다. (그림 1)은 전형적인 VOD 시스템 구조로 VOD 서버(VOD server), 대용량의 비디오 저장 장치, 광대역 기간 통신망(broadband backbone network), 액세스망(access network), STB(Set-Top Box)등의 요소들로 구성된다[2].



(그림 1) VOD 시스템의 구조
(Fig. 1) VOD system architecture

VOD 서버는 VOD 시스템에서 가장 중요한 구성 요소이고, 서버는 대용량의 크기를 갖는 비디오 데이터를 저장하고 연속적으로 전송하기 위해, 여러 가지 동영상 압축 기법 중에 가장 일반적인 MPEG으로 비디오 데이터를 압축하고 디스크 어레이(disk array)를 사용하여 압축된 데이터를 저장하며, 사용자에게 대화형 서비스를 제공하기 위하여 전송할 비디오 스트림을 설정하고 이를 유지하는 한편 VCR 기능 제공 방법도 포함한다. 광대역 기간 통신망과 액세스 망은 사용자와 서버 사이에 대용량의 데이터를 고속으로 전송해야 하는 고수준의 상호 연결성을 제공하기 위해 각각의 네트워크 요소들을 연결시키고 데이터의 전송, 교환, 경

로 배정 및 네트워크 자원의 관리와 서비스 제어 등의 기능을 제공한다. 그리고, STB는 VOD 서버로부터 전송 받은 비디오 정보를 사용자 디스플레이 장치에 재생하는 기능과 사용자의 요구를 서버에게 알려주는 기능을 수행하며, 사용자가 서버에게 제어 신호를 전송할 수 있도록 사용자 인터페이스도 포함해야 한다. 또한, 통신망 지터(jitter) 등으로 인한 지연으로부터 재생의 연속성을 보장하기 위한 버퍼도 필요하다.

VOD 서비스에서 서비스 대상이 되는 비디오의 서비스 요구 빈도의 전형적인 분포는 <표 1>과 같다. 비디오 클래스 1과 2에 속한 소수의 비디오가 사용자의 서비스 요구 빈도의 대부분을 차지하고, 비디오 클래스 3과 4에 속하는 나머지 대다수의 비디오들은 아주 적은 서비스 요구 빈도를 나타낸다[6]. 이러한 경향은 서비스 요구 빈도의 집중도(skew level)가 높을 경우에 더욱 두드러지게 나타나고, 서비스되는 비디오의 수가 더욱 많아지는 경우에도 유사한 분포를 나타낸다. VOD 서비스는 이러한 서비스 요구 빈도의 특성을 가지고 있기 때문에, 서비스 요구 빈도가 낮은 비디오들을 대용량의 지속 저장 장치에 저장하여 서비스하는 방법들도 연구되었다.

<표 1> 전형적인 서비스 요구 빈도 분포
<Table 1> Typical distributions of service request probability

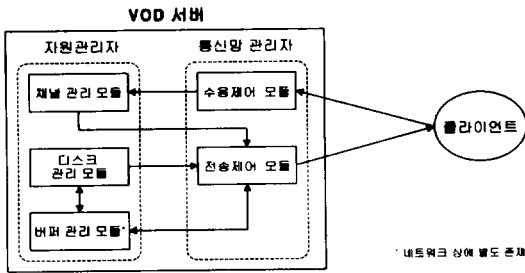
집중도 (Skew level)	영화 클래스 (Movie class)	영화 범주 (Movie range)	서비스 요구 빈도
Low	1	1 - 10	0.5
	2	11 - 30	0.25
	3	31 - 100	0.15
	4	101 - 200	0.1
High	1	1 - 10	0.8
	2	11 - 30	0.15
	3	31 - 100	0.04
	4	101 - 200	0.01

2.2 시스템 모델

본 논문에서 제안된 VOD 시스템 모델은 (그림 2)와 같다. 서버 구조는 자원 관리자(resource manager)와 통신망 관리자(network manager)로 구성된다. 자원 관리자는 채널 관리 모듈(channel management module)과 디스크 관리 모듈(disk management module) 그리고 버퍼 관리 모듈(buffer management module)로 구성되며, 통신망 관리자는 수용 제어 모듈(admission

control module)과 전송 제어 모듈(transmission control module)로 구성된다. 클라이언트의 요구가 발생되면 수용 제어 모듈에 의해 접속이 이루어지며, 이 정보는 채널 관리 모듈에게 전달되어 새로운 멀티캐스트 채널을 생성 받거나 기존의 채널에 합류된다. 이때 채널 관리 모듈은 본 논문에서 제안하는 동적 버퍼 기반의 일괄 전송 알고리즘에 의해 클라이언트의 요구를 처리한다.

버퍼 관리 모듈은 동적 버퍼 관리를 위해 존재하며, 다중 클라이언트로부터 발생되는 VCR 요구를 효과적으로 처리하기 위해 사용된다. 따라서, 채널 관리 모듈에 의해 설정된 클라이언트들의 멀티캐스트 채널에 대해 일괄 전송을 담당하는 모듈로 존재하며, 이것은 실제 구현상 효율을 높이기 위해 서버와 분리시켜 별도의 서버로 구현할 수 있다.



(그림 2) 제안된 VOD 시스템 모델
(Fig. 2) Proposed VOD system model

2.3 시스템 구현 환경

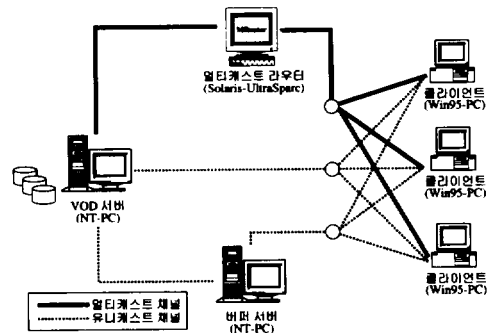
본 논문에서 구현한 일괄 전송 방식을 이용한 멀티캐스트 VOD 서버는 고속 이더넷(100Mbps)을 기본 통신망으로 하여 펜티엄 프로 서버를 사용하였으며, 운영체제로는 Windows 95와 Windows NT 플랫폼을 사용하였다. 또한, 멀티캐스트를 구현하기 위해 Mbone 통신망을 사용하였다. 효율적인 통신망의 자원 관리를 위해 본 VOD 시스템에서는 RSVP(Resource Reservation Protocol)를 적용하였으며, RSVP는 서버와 클라이언트간에 일정한 서비스 품질을 유지하도록 위해 사용된다.

VOD 서버는 Windows NT 상에서 IP 멀티캐스트(UDP)와 유니캐스트(TCP)를 혼용한 형태로 서비스를 수행하도록 설계하였으며, TCP 메시지는 VCR 기능과 같은 제어 메시지의 송·수신 채널로 사용되고, UDP

메시지는 실제 MPEG 비디오 데이터 전송을 위해 사용된다. 따라서, 일괄 전송 방식을 사용한 멀티캐스트 기반의 VOD 서비스가 가능해진다. 그리고 앞서 언급한 바와 같이 버퍼 서버는 서버 시스템과 통신망 상에 분리된 형태로 존재함으로써 서버의 부하를 최소화하고 서비스 성능을 높일 수 있도록 구현하였다.

클라이언트는 Windows 95 플랫폼을 사용하는 PC에서 구현하였고, GUI(Graphic User Interface)를 통하여 서버에게 접속을 요청한 후 비디오 리스트들을 수신하게 된다. 클라이언트는 RSVP 채널을 통해 QoS(Quality of Service)를 예약한 후 비디오 리스트중 하나를 선택함으로써 비디오를 시청하게 된다. 또, 사용자의 VCR 기능 요구를 수행하기 위해서는 선택된 VCR 기능 메시지를 비디오 서버에게 전송해 줌으로써 기능을 수행하게 된다.

(그림 3)은 본 논문에서 설계한 VOD 시스템의 실제 구현 환경이다.



(그림 3) VOD 시스템 구현 환경
(Fig. 3) VOD system implementation environment

3. 스케줄링 기법

3.1 채널 할당 기법의 목적

채널 스케줄링 기법의 선택은 사용자가 요구를 취소하기 전에 허용할 수 있는 대기 시간에 달려있다. 즉, 자주 서비스되는 핫 비디오와 드물게 요구되는 콜드 비디오에 따라서 채널 할당의 요구 조건이 바뀌게 된다. 즉, 채널 할당 기법은 어떤 이용 가능한 채널로 어떤 비디오를 선택하여 재생하느냐에 대하여 고려한 것이다. VOD 서버에서 제공하는 비디오는 사용자가 많이 요구하는 핫(hot) 비디오와 적게 요구하는 콜드(cold) 비디오로 크게 나눌 수 있다. 핫 비디오는 짧은

시간 내에 많은 요구가 도착할 수 있기 때문에 여러 요구들을 일괄전송 함으로써 충분한 이익을 얻을 수 있다. 이와 반대로, 콜드 비디오는 동일한 비디오들을 위한 새로운 요구의 도착이 매우 작기 때문에 사용자가 짧은 주기 동안 기다리는 것으로부터 이득을 얻을 수 없다 그러므로, 모든 비디오에 대하여 고정된 일괄전송 간격은 부적당하다. 또한, 대기 시간이 길수록 비디오를 요구한 사용자가 취소하여 시스템으로부터 떠나는 원인이 될 수 있다.

일괄전송 기법은 동시에 서비스될 수 있는 스트림의 수가 I/O대역폭에 의하여 여전히 제한되며, 하나의 일괄전송 간격 내에 먼저 도착한 요구들은 불공평하게 나중에 들어오는 요구들을 기다려야 한다. 따라서, 이러한 시스템의 많은 사용자들은 서비스를 받기 위하여 오랫동안 대기하여야 한다. 그리고, 사용자의 흥미가 저하될 수 있으므로 너무 오랫동안 기다릴 경우에는 그들의 요구를 취소할 수 있는 제한을 여전히 갖는다.

채널 할당 기법을 평가하는 주요 객체는 세 가지로 나눌 수 있다. 첫째, 취소 확률(renegeing probability)의 최소화는 매우 중요한 평가 기준이다. 왜냐하면, 서버 용량(server capacity)에 대한 취소확률을 최소화하는 기법은 주어진 수용율(acceptance rate)을 만족하는 더 낮은 서버 용량을 요구하기 때문이다. 대기행렬 이론에서 고객이 서비스를 받기 전에 대기행렬을 떠나는 것을 사용자의 "renege"라 한다. 이것은 사용자 요구 후 오랫동안 채널을 할당받지 못하고 오랜 시간으로 인하여 요구를 취소한 사용자를 말한다. 취소 확률을 최소화하여 시스템의 처리율을 최대화하여야 한다. 둘째, 평균 대기 시간(average waiting time)을 최소화하는 것이다. 비디오 요구가 도착해서 실제로 재생 장치에 서비스가 시작될 때까지의 경과된 시간을 의미한다. 평균 대기 시간과 서비스를 받을 사용자의 대기 시간의 분산을 최소화하는 것이다. 그리고, 셋째는 모든 비디오들에 대한 요구에 공평성(fairness)을 최대화하는 것이다. 공평성은 모든 비디오 요구에 대한 균등한 취소 확률로서 정의된다. 비디오 i 에 대한 취소 확률을 d_i 라 할 때, 평균 취소 확률은 식 (1)과 같다.

$$\bar{d} = \sum_{i=1}^N \frac{d_i}{N} \tag{1}$$

따라서, 불공평성(unfairness)은 식 (2)와 같이 구할 수

있다. 공평성은 이것의 표준 편차와 반비례한다[4][5].

$$\sqrt{\sum_{i=1}^N \frac{(d_i - \bar{d})^2}{(N-1)}} \tag{2}$$

3.2 사용자 요구별 기법

기존의 논문에서 일괄전송을 위하여 비디오를 선택하는 기법에는 예약(reservation) 기법, FCFS(First-Come-First-Service)기법 그리고 MQL(Maximum Queue Length)기법 등이 있다.

예약 기법은 비디오를 요구하는 사용자가 발생할 때마다 스트림을 채널에 할당한다. 한 사용자가 각각의 채널을 소유하게 되므로 시스템의 구성에 따라서 사용자 수가 제한되는 단점이 있지만, 디스크 I/O 대역폭의 확보에 따라서 VCR 기능을 제공할 수 있다. 또한, FCFS 기법은 모든 비디오들에 대한 요구는 요구 대기행렬이라고 하는 단일 대기행렬로 들어간다. 각 사용자는 요구를 취소하기 전까지의 최대 대기 시간 즉, 취소 시간에 입각하여 다른 사용자와는 별도로 대기행렬을 떠날 수 있다. 일단 하나의 스트림을 전송하기 위하여 서버 용량이 이용 가능하다면, 요구 대기행렬 앞에 있는 사용자가 서비스를 받는다. 또한, 멀티캐스트 기능으로 인하여 동일한 비디오에 대한 모든 사용자의 요구는 동일한 스트림에 의하여 서비스된다. 직관적으로 FCFS기법은 독립적으로 비디오 요구를 선택하기 때문에 공정한 기법으로 생각할 수 있다. 그리고, MQL 기법은 각 비디오에 대한 요구는 별도의 대기행렬에 들어가며, 최대 대기행렬 길이를 갖는 비디오가 멀티캐스트를 위하여 선택된다. 그러나, 이 기법의 결점은 짧은 시간 주기(한 사용자의 renegeing time) 내에 콜드 비디오들에 대해서는 극소수의 요구가 있기 때문에 핫비디오 만이 선택할 수 있다는 것이다. 따라서, 이 기법은 불공평성의 증가로 인하여 콜드 비디오들에 대한 요구의 취소확률이 상당히 증가할 것이다. 하지만, 이 기법은 전체 취소확률을 감소시키기 위하여 작은 서버 용량을 더 활용할 수 있다.

이에 대한 변형 기법으로 FCFS- n 기법은 서버 용량의 일부가 미리 예약되어, n 개의 가장 인기있는 비디오에 대한 일괄전송 요구를 위하여 미리 할당하여 서비스하며, 나머지 비디오에 대해서는 FCFS기법과 유사하다. 이 기법은 콜드 비디오에 대하여 다소 불공평할 수 있다. 하지만, 핫 비디오에 대하여 서버 용량을 미리 할당하는데 많은 장점을 가지고 있다. 첫째로

최대 대기 시간에 대한 보장은 핫 비디오에 대한 요구를 서비스할 수 있으며, 사용자 습성에 영향을 보장할 수 있다. 둘째로, 이 기법은 비교적 작은 서버 용량으로 높은 수용율을 보장할 수 있다. 콜드 비디오의 경우는 FCFS기법에 의해 서비스되며 FCFS-0는 FCFS 기법과 동일하다. 만약 핫 비디오의 요구가 없을 경우에 콜드 비디오를 멀티캐스트 전송함으로써 가능할 것이다. 다른 방법으로 MQL- n 기법을 생각할 수 있다. n 개의 핫 비디오를 서버에 할당하고 나머지 요구들은 MQL기법에 따라 서비스한다.

3.3 큐잉을 고려하지 않은 스케줄링 기법

일괄 전송을 위한 시간 간격 내에 도착한 사용자 요구들은 스케줄링 기법에 의하여 서비스를 받거나 또는 거절된다. 큐잉을 고려하지 않은 스케줄링 기법은 자유 채널의 수가 이미 할당되었을 경우 사용자 요구가 다음에 스케줄링을 위하여 대기 행렬에 저장되지 않고 거절되는 블로킹(blocking) 기법으로 인기도에 따른 스케줄링 기법과 고정 채널 스케줄링 기법을 고려하였다.

3.3.1 인기도에 따른 스케줄링 기법

이 기법은 비디오를 시청하기 위하여 요청한 요구 수에 따라서 사용자 요구를 스케줄링하는 기법이다. 특히, 자유 채널의 수가 n 개라면, 서버엔드는 전송을 위하여 사용자의 요구 수에 따라 순위가 매겨져 있는 n 개의 비디오 프로그램들이 선택된다. 요구한 비디오 프로그램의 전체 수가 n 개 이상이라면, 적은 요구를 갖는 비디오 프로그램을 요구한 요구들이 거절된다. 이러한 스케줄링 하에서, 만약 요구의 도착 시간이 시간 간격 내에 일정 분포를 갖는다면, 선택된 비디오 프로그램을 시청하기 위한 평균 대기 시간은 $\frac{\Delta t}{2}$ 이며 최대 대기 시간은 Δt 이다.

이 기법은 N-VOD 시스템이 수행한 것과 같은 방법으로 핫 비디오 프로그램들에게 VCR 기능을 제공할 수 있다. 핫 비디오 프로그램은 각 시간 간격에서 전송하기 위하여 스케줄링되는 것이 매우 좋을 것이라 기대할 수 있다. 그러므로, 시간 간격 내에 동일한 프로그램을 갖는 다른 채널로 교환함으로써 VCR 기능이 제공될 수 있다.

3.3.2 고정 채널 스케줄링 기법

Nussbaumer[13]는 T-VOD에서 골든 시간 동안 모

든 채널이 같은 시간대에 모두 할당되며, 시스템은 상당한 시간 동안 포화상태로 유지되는 비정상적인 현상을 연구하였다. 그러므로, 시스템이 포화되어진 후에 대부분의 새로운 요구들은 거절된다. 비록 요구가 정해진 시간 간격에 스케줄링된다 하더라도 인기도에 따른 스케줄링 기법과 비슷한 현상이 발생한다.

비디오 프로그램의 평균 지속 시간이 시간 간격의 길이와 비교할 때 다소 길며, 요구의 도착율이 클 때, 모든 자유 채널들이 바로 할당될 것이다. 그 결과, 다음 시간 간격 동안 도착한 매우 작은 새로운 요구들이 수용될 것이다. 만약 모든 비디오 프로그램들의 시간이 거의 동일하며 요구율이 높게 유지된다면, 모든 채널은 거의 동일한 시간에 반환될 것이며 그다음 다시 빠르게 채널이 할당된다. 따라서, 시스템은 때로는 서로 다른 시간 동안 대다수의 요구들을 수용하며, 때로는 대다수의 요구들을 거절하는 비안정적인 행동을 한다.

이 기법하에서, 각 시간 간격에서 스케줄링을 위한 자유 채널의 수는 상한치를 갖는다. 즉, 오직 고정된 비디오 프로그램의 수만이 각 시간 간격에서 방송을 위하여 스케줄링될 수 있다. 적절한 임계치를 선택함으로써, 채널을 각 시간 간격에 공평히 할당될 것이라 예견할 수 있다. 따라서, 소수의 시간 간격에 의하여 채널이 점유되어 포화되는 것을 피할 수 있다.

일정한 시간 간격 동안 할당할 수 있는 최대 채널의 수 C_{max} 는 사용자가 하나의 비디오를 시청하는 시간인 평균 채널 보유 시간(average channel holding time) T_{view} 에 달려있다. 그리고 시스템내의 전체 채널의 수를 C_{total} , 그리고 시간 간격을 $T_{interval}$ 이라 한다면 일정 시간 간격 동안 할당될 수 있는 채널의 전체 수 C_{max} 는 식 (3)과 같이 계산될 수 있다.

$$C_{max} = \frac{(T_{interval} * C_{total})}{T_{view}} \quad (3)$$

예를 들면, 100개의 채널이 VOD 시스템에서 이용 가능하며, 스케줄링을 위한 시간 간격이 매 5분이라 가정하고, 비디오 프로그램의 길이는 100분에서 120분 사이의 일정분포를 갖는다고 가정한다. 이러한 가정 하에서, 22 간격(110분/5분)이 하나의 사이클을 형성한다. 각 간격 동안 사용 가능한 채널의 상한치는 5개([100채널/22채널])이다. 다시 말해서 하나의 간격 동안 수신된 모든 요구는 거의 5개 채널에 이른다.

이 기법은 요구의 거절율을 줄일 수 있다. 왜냐하면, 인기있는 비디오 프로그램이 스케줄링될 수 있는 기회를 더 많이 얻기 때문이다. 하지만, 반대로 비인기 비디오 프로그램은 전송될 기회를 훨씬 덜 갖게된다. 이것은 어떤 특정한 비디오를 즐기기를 원하는 사용자들에게 매우 불공평하다. 인기도에 따른 스케줄링 기법과 고정 채널 스케줄링 기법은 공정성 문제에 단점을 가지고 있다.

3.4 큐잉을 고려한 스케줄링 기법

큐잉을 고려한 스케줄링 기법에서 사용자 요구는 자유 채널을 할당받기 전에 요구 대기 행렬 내에 대기한 후 자유 채널을 할당받아 선택한 비디오 스트림을 전송하는 기법이다. 즉, 자유 채널이 이미 할당되어 없을 경우에도 사용자가 시스템을 떠나지 않는 한 요구가 대기 행렬에 저장되는 기법이다.

3.4.1 누적 대기 시간 기법

누적 대기 시간 기법은 대기 시간과 대기 행렬에 저장된 요구 수에 대하여 고려한 기법이며, 비디오를 스케줄링하기 위한 우선 순위로서 모든 요구들의 대기 시간을 합한 것을 사용한다. 비디오 i 의 대기 행렬 내에 j 번째 요구에 대한 대기 시간을 $w_{i,j}$ 라하며, N_i 는 대기 행렬내의 요구 수라 가정할 때 비디오 i 에 대한 대기 시간을 누적한 시간 W_i 는 식 (4)와 같이 나타낼 수 있다.

$$W_i = \sum_{j=1}^{N_i} w_{i,j} \tag{4}$$

3.4.2 평균 대기 시간 기법

우선 순위로서 대기 시간을 누적한 값을 사용하는 누적 대기 시간 기법과는 달리 평균 대기 시간 기법은 누적된 대기 시간을 요구 수로 나누어 비디오를 스케줄링하는 기법이다. 따라서, 비디오 i 에 대한 평균 대기 시간 M_i 는 식 (5)와 같다.

$$M_i = \frac{W_i}{N_i} = \frac{\sum_{j=1}^{N_i} w_{i,j}}{N_i} \tag{5}$$

3.4.3 예측 인기도 가중치 기법

기존의 MQL 기법은 콜드 비디오에 대한 취소 확률이 증가하는 단점을 가지고 있으며, FCFS는 여러 사용

자들에게 비디오 서비스를 우선적으로 제공하지 못하여 채널 할당이 비효율적이라는 단점을 가지고 있다.

이 기법에서는 비디오의 서비스 요구 빈도를 계산하기 위하여 선형 예측법(linear prediction)을 사용한다.

예측 인기도 가중치(Predictable Popular Factor) 기법을 이용한 스케줄링 기법은 상대적 요구 빈도에 의하여 인기도를 구하기 위해서는 가까운 과거 기간 내에 발생한 각 비디오의 요구 빈도 수를 계수함으로써 얻어진다. 이 예측 인기도 가중치 기법은 현재 대기행렬 길이를 상대적 인기도에 따른 요구 빈도에 대하여 가중치를 구하고 이를 평균하여 적용할 수 있다.

전체 비디오의 수를 N 개, 그리고 각 비디오 i 요구 인기도를 p_i 라고 하면, 인기도에 대한 벡터 값은 식 (6)과 같다.

$$P = [p_1, p_2, p_3, \dots, p_N], \quad \sum_{i=1}^N p_i = 1 \tag{6}$$

선형 예측법은 이전 시간 간격 동안 비디오의 요구 빈도를 기준으로 현재의 요구 빈도를 결정하는 방법이다. $q_i(t-2)$ 와 $q_i(t-1)$ 을 각각 $(t-2)$ 시간과 $(t-1)$ 시간의 비디오 i 를 요구한 사용자 수라고 한다면, 시간 t 에 요구한 사용자 수는 식 (7)과 같이 표시할 수 있다.

$$q_i(t) = 2q_i(t-1) - q_i(t-2), \quad q_i(t) \geq 0 \tag{7}$$

따라서, 비디오 i 에 대하여 요구된 사용자 요구 빈도는 식 (8)과 같다.

$$p_i = \frac{q_i(t)}{\sum_{m=1}^N q_m(t)} \tag{8}$$

만약 비디오 i 의 길이를 L_i 라 하면, 평균 비디오 길이 L 은 다음과 같은 식 (9)으로 표현할 수 있다.

$$L = \frac{\sum_{i=1}^N p_i \cdot L_i}{\sum_{i=1}^N p_i} \tag{9}$$

서버에 존재하는 서버 용량 즉, 스트림 수를 S 라고 하면, 전체 용량에서 매분마다 서버에 의해 스케줄링되는 평균 스트림 수는 $\frac{S}{L}$ 이다.

비디오 i 가 일괄 전송에 의하여 전송될 때 평균 시간 간격을 t_i 라고 하면, 비디오 i 에 대한 평균 지연 시간은 $\frac{t_i}{2}$ 가 된다. 따라서, 비디오 i 에 대하여 사용자 요구 빈도가 p_i 이므로, 하나의 일괄 전송을 위한 시간 간격 동안 도착하는 모든 요구의 지연 시간의 합은 식 (10)과 같다.

$$W = \sum_{i=1}^N p_i \cdot \frac{t_i}{2} \quad (10)$$

4. 시뮬레이션 및 성능분석

제안된 스케줄링 기법들에 대한 성능이 시뮬레이션을 통하여 평가되었다. 우선, 시뮬레이션을 위한 가정을 기술하였다. 그런 다음 제안된 스케줄링 기법에 대한 성능을 평가 및 비교하였다. 제안된 기법의 성능을 비교, 분석하기 위하여 CSIM18 시뮬레이션 엔진(C버전)[12]으로 시스템 이용도가 가장 높게 예상되는 오후 6시부터 12시까지 끝든 아워 동안 시뮬레이션을 수행하였다.

4.1 가정 및 파라미터

모든 비디오의 길이는 90분에서 120분 사이의 일정분포를 갖는 것으로 가정하였으며, 이용 가능한 비디오의 수 L 은 100개로 가정하였다. 비디오의 인기도는 Zipf 분포를 따른다. Zipf 분포의 선택은 고정된 객체들로부터 몇 개의 객체가 선택될 확률이 다른 객체에 비해 현저히 클 때 특정한 객체가 선택될 확률을 구하기 위해 사용되는 분포이며, 실제로 비디오 대역 분포와 가장 근접하게 일치되기 때문이다[4].

L 개의 비디오는 그 비디오들의 인기도에 따라서 순서화되어 있으며, i 번째 비디오에 대한 액세스 빈도는 $f_i = \frac{c}{i^{1-\theta}}$ 에 의하여 선택되는 확률을 갖으며, 사향(skewed) 분포로 가정하였다. 이때, $c = \frac{1}{\sum_{i=1}^L (\frac{1}{i^{1-\theta}})}$

이며, $0 \leq \theta \leq 1$ 이다. 파라미터 θ 는 사향 요소(skew factor)라 부른다. θ 가 1에 접근하면 일정분포에 접근하며, 반면 θ 에 0에 가까우면 집중도가 증가하는 아주 높은 사향 분포가 된다. 일괄전송 요구는 분포가 높은 사향일 경우 보다 우수한 성능을 나타낸다. 임의의

비디오 i 를 요구할 확률은 q_i 로 가정하였다. 본 시뮬레이션에서 q_i 에 대하여 θ 는 0.271로 설정하였다[8].

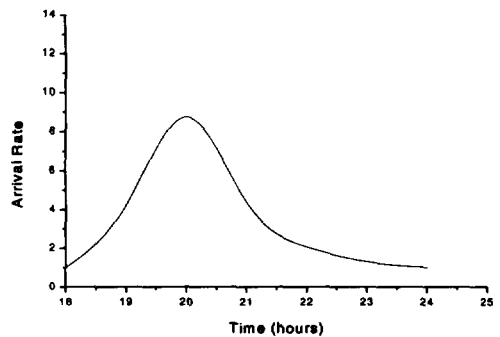
사용자 요구가 파라미터 $\lambda(t)$ 를 갖는 포아송 분포에 따라서 server-end에 도착한다고 가정하였다. 요구율의 분산은 σ 에 의해서 제어된다. σ 가 큰 값을 가지면 시간상으로 더 완만한 곡선을 낳는다. 본 시뮬레이션에서는 사용자들을 위하여 180개의 채널을 예약한 시스템으로 가정하였다. 시스템의 확률론적 특성으로 인하여, 시스템은 많은 시뮬레이션을 실행하였다.

4.2 성능 분석

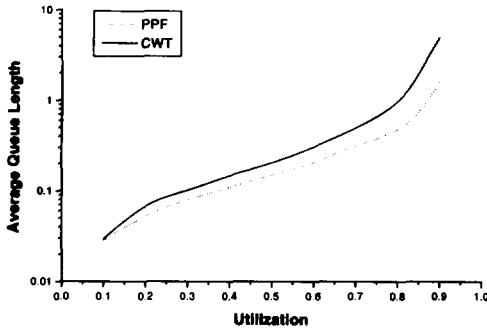
본 논문은 스케줄링 기법들의 성능에 관한 몇 가지 시뮬레이션 파라미터의 결과를 우선 검토하였다. 그 다음 서로 다른 파라미터 설정 하에서 스케줄링 기법들의 성능을 비교하였다.

그리고, 각 사용자로부터 평균 요구 횟수는 주당 2회이며, 전체 사용자 수는 1600명이다. 따라서, 일반적으로 하루에 요구하는 평균 횟수는 약 457명이 된다. (그림 4)는 시간에 따른 도착율을 표시하였다. 이때 X-축은 하루의 시간을 표시하였다. 예를 들어, 20은 오후8시를 의미한다. (그림 5)는 이용률과 평균 대기행렬 길이의 관계에 대하여 도식화하였다. 실선은 CWT기법이며, 점선은 PPF기법을 표시한 것이다. 이용률이 증가함에 따라 CWT 기법의 평균 대기행렬 길이가 PPF기법 보다 증가함을 알 수 있었다.

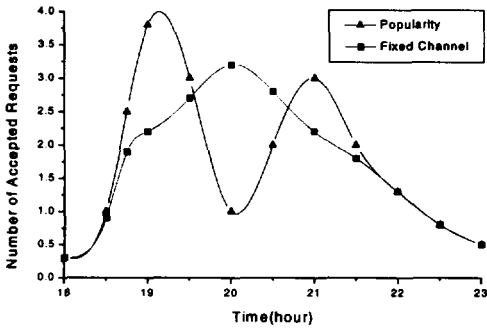
(그림 6)은 평균 도착율이 2이고, 스케줄링 간격이 5분일 때, 인기도에 따른(Popularity-based) 기법과 고정 채널(Fixed Channel) 기법에 대하여 하루 중



(그림 4) 끝든 아워 동안의 요구 도착
(Fig. 4) Request arrival for peak hour



(그림 5) 이용율과 평균 대기 행렬 길이
(Fig. 5) Average queue length



(그림 6) 골든 시간에 수용된 요구 수
(Fig. 6) Number of accepted requests in peak hour

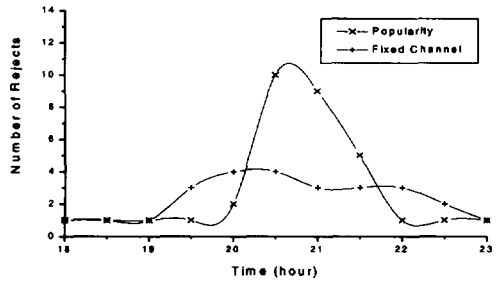
여러 시간에 수용된 요구수를 나타낸 것이다. 그림에서 알 수 있듯이, 인기도에 따른 기법이 고정 채널 기법보다 하루에 요구를 훨씬 더 많이 수용할 수 있다.

(그림 7)은 큐잉을 고려하지 않은 스케줄링 기법인 인기도에 따른 기법과 고정 채널 기법에 대하여 골든 시간에 따른 사용자 요구 취소율을 나타낸 것이다. 채널을 일찍 할당하기 때문에 고정 채널 기법은 인기도에 따른 기법보다 이른 시간에 취소율을 줄일 수 있었지만, 고정 채널 기법은 골든 시간 경에 취소율을 줄일 수 있었다.

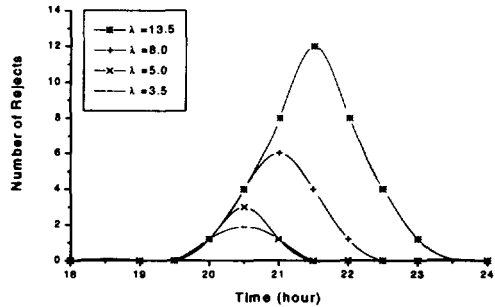
(그림 8)과 (그림 9)는 각각 누적 대기 시간 기법과 예측 인기도 가중치 기법에 따라 스케줄링 간격을 5분, θ 를 0.271로 가정할 때, 도착율 λ 가 3.5, 5.0, 8.0 그리고 13.5로 고정할 때의 사용자 요구 취소율을 나타낸 것이다. 누적 대기 시간 기법보다 상대적으로 예측 인기도 가중치

기법이 사용자 요구 취소율이 적음을 알 수 있다.

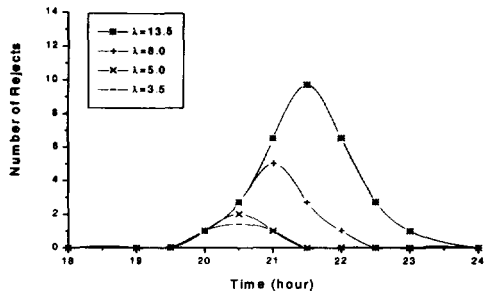
(그림 10)은 큐잉을 고려한 세 가지 기법 각각에 대하여 골든 시간 동안의 평균 대기 시간을 나타낸 것이다. 큐잉을 갖는 세 가지 기법들은 모두 인기도에 따른 스케줄링 기법보다 훨씬 낮은 취소율을 나타내며, 평균 대기 시간에 대하여 누적 대기 시간 기법과 평균 대기 시간 기법보다 예측 인기도 가중치 기법에 평균 대기 시간이 작게 요구됨을 알 수 있었다.



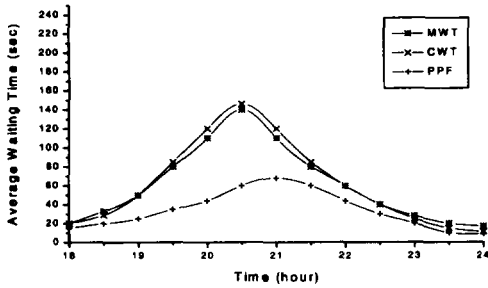
(그림 7) 큐잉을 고려하지 않은 기법과 취소율
(Fig. 7) Number of rejects without queueing policies



(그림 8) 누적 대기 시간 기법과 취소율
(Fig. 8) Number of rejects under CWT policy



(그림 9) 예측 인기도 가중치 기법과 취소율
(Fig. 9) Number of reject under PPF policy



(그림 10) 큐잉을 고려한 기법들과 평균 대기 시간 (Fig. 10) Average waiting time with queuing policies

5. 결 론

VOD에서 일괄 전송 기법은 일반적으로 I/O요구를 감소시키거나 처리율을 향상 시키는데 사용된다. 일괄 전송 기법에 의하여 동일한 비디오 스트림을 이용하여 서비스한다고 가정할 때, 하나의 비디오 스트림이 server-end에서 이용 가능할 때마다 어떤 비디오가 어떤 순간에 스케줄링할 것인지 문제가 제기된다. 따라서, 각 스트림 마다 변동하는 자원 대역폭을 처리하기 위해서 채널 할당 전략에 관한 것이 중요한 문제로 대두되었다.

본 논문에서는 큐잉을 고려한 세 가지 스케줄링 정책을 제안하였다. 일괄전송 기법을 바탕으로 제안된 채널 스케줄링 기법에 대하여 성능을 분석하였으며, 그 결과 누적 대기 시간 기법은 인기있는 비디오에 대하여 우수한 반면, 평균 대기 시간 기법은 비인기 비디오에 대하여 우수하였다. 누적 대기 시간 기법은 가장 낮은 취소율을 나타냈으며, 예측 인기도가중치에 의한 기법은 비디오에 대한 평균 대기 시간과 사용자의 요구 취소율에 대하여 다소 공평함을 알 수 있었다.

참 고 문 헌

[1] W. D. Sincoskie, "Video On Demand : Is It Feasible?," *GLOBECOM '90*, Vol.1, pp.201-205. 1990.
 [2] K. Almeroth, M. Ammar, "The Use of Multicast Delivery to Provide a Scalable and Interactive Video-On-Demand Service," *IEEE Journal on Selected Areas in Communications*, August

1996.
 [3] K. Almeroth, M. Ammar, "On the Performance of a Multicast Delivery Video-On-Demand Service with Discontinuous VCR Actions," *ICC*, June 1995.
 [4] A. Dan, D. Sitaram, P. Shahabuddin, "Scheduling policies for an On-Demand Video Server with Batching," *ACM Multimedia '94*, (San Francisco, CA), Oct 1994.
 [5] Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu, "On Optimal Batching Policies for Video-On-Demand Storage Servers," *IBM Research Report (RC20621)*.
 [6] T. Little, D. Venkatesh, "Prospects for Interactive Video-On-Demand," *IEEE Multimedia*, pp.14-23, Fall 1994.
 [7] K. Almeroth, M. Ammar, "Long Term Resource Allocation in Video Delivery Systems," *INFOCOM '97*, pp.1335-1342, 1997.
 [8] R. Flynn, W. Tetzlaff, "Disk Striping and Block Replication Algorithms for Video File Servers," *IEEE Proc. of Multimedia '96*, pp.590-597, 1996.
 [9] V. Li, W. Liao, X. Qiu and E. Wong, "Performance Model of Interactive Video-on-Demand Systems," *IEEE Journal on Selected Areas in Communications*, August 1996.
 [10] Hatem Ghafir, Henry Chadwick, "Multimedia Servers-Design and Performances," *GLOBECOM '94*, pp.886-890, 1994.
 [11] Emmanuel L. Abram_Profeta, Kang G. Shin, "Scheduling Video Programs in Near Video-on-Demand Systems," *ACM Multimedia '97 Proceedings*, Nov. 1997.
 [12] User's Guide : CSIM18 Simulation Engine (C Version), *Mesquite Software, Inc.*, 1997.
 [13] J. P. Nussbaumer, B. V. Patel, F. Schaffa, "Multimedia Delivery on Demand : Capacity Analysis and Implications," *Proc. of the 19th Conference on Local Computer Networks*, Oct. 1992.
 [14] H. Park, H. Ryou, "Multicast Delivery for Interactive Video-On-Demand Service," *IEEE Proc*

of ICOIN-12, Jan. 1998.

- [15] 박호균, 이승윤, 유황빈, "VOD상에서 프리젠테이션 방향성을 이용한 동적 캐쉬 스케줄링 방법", 한국통신학회논문지 제20권 제9호, 1995.
- [16] 박호균, 홍명준, 유황빈, "VOD 상에서 VCR 기능 제공 방법 설계 및 구현", 한국정보처리학회논문지 제4권 제10호, 1997.
- [17] 박호균, 유황빈, "주문형 비디오 서비스 시스템에서 VCR 기능을 위한 Batching 전송을 위한 채널 스케줄링 기법에 관한 연구", 한국통신학회논문지 제22권 제12호, 1997.
- [18] 박호균, 이현찬, 유황빈, "IP 멀티캐스트 상에서 VOD 서버의 설계 및 구현", 한국정보처리학회 춘계학술발표논문집, 1998.



박 호 균

e-mail : hkpark@tiger.shinheung-c.ac.kr
 1987년 광운대학교 전자계산학과 (이학사)
 1989년 광운대학교 대학원 전자계산학과(이학석사)
 1998년 광운대학교 대학원 전자계산학과(이학박사)

1992년~현재 신홍대학 컴퓨터정보계열 조교수
 1999년~현재 신홍대학 정보산업연구소장
 관심분야 : VOD, 멀티미디어통신, 멀티캐스트, 성능평가



유 황 빈

e-mail : ryou@cs.kwangwoon.ac.kr
 1975년 인하대학교 전자공학과(공학사)
 1977년 연세대학교 산업대학원 전기전자공학과(공학석사)
 1989년 경희대학교 대학원 전자공학과(공학박사)

1994년~1995년 미국 UCSD 교환교수
 1975년~1980년 금성반도체(주)
 1981년~현재 광운대학교 컴퓨터학과 교수
 1997년~현재 광운대학교 중앙도서관장
 관심분야 : 정보보호, VOD, 초고속통신망, 성능평가