

# 주문형 비디오 시스템에서 인기 비디오를 위한 예약 채널 용량의 결정

## (Determination of Reserved Channel Capacity for Popular Videos in Video-on-Demand Systems)

하 속 정<sup>†</sup> 배 인 한<sup>††</sup>  
(Sook-Jeong Ha) (Ihn-Han Bae)

**요약** VOD 서비스는 멀티미디어의 특성인 높은 대역폭과 긴 재생시간으로 인해 서버와 네트워크의 자원 소비가 높다. 패칭과 채널 예약 패칭 같은 여러 가지 기법들이 네트워크 임플렉 대역폭 요구를 감소시키기 위해 제안되었다. 패칭은 True VOD의 실현을 위하여 새로운 클라이언트들을 기존의 멀티캐스트와 결합하는 것을 허용하고, 채널 예약 패칭은 사용자들의 요청이 집중되어 있는 인기 비디오 요청을 서비스하기 위하여 비디오 서버의 전체 채널 용량의 일부를 예약한다. 본 논문에서는 채널 예약 패칭의 성능을 개선시키기 위하여 비디오 서버의 예약 채널 용량을 결정하는 방법을 제안한다. 제안하는 방법의 성능은 시뮬레이션을 통해 평균 서비스 지연시간, 이탈율 그리고 공평성으로 평가하고, 패칭과 다른 채널 예약 패칭의 성능과 비교한다.

**키워드**: 멀티미디어, 멀티캐스트, 스류, 예약 채널 용량, 인기 비디오, 패칭, 패칭 윈도우, True VOD

**Abstract** Video-on-Demand (VOD) service needs high resource consumption because of high bandwidth and long playback duration which are characteristics of multimedia. Various techniques such as Patching and Channel-Reserving Patching have been proposed to reduce the demand on the network-I/O bandwidth. Patching allows new clients to join an existing multicast for realization of true VOD, and Channel-Reserving Patching reserves the part of total channel capacity of video server in order to serve the requests for popular videos on which requests of users concentrate. To improve the performance of Channel-Reserving Patching, we propose the method that determines the reserved channel capacity of video server for the popular videos in this paper. The performance of our method is evaluated in terms of average service latency, defection ratio and fairness through simulation, and is compared with those of Patching and another Channel-Reserving Patching.

**Key words**: Multimedia, Multicast, Skew, Reserved Channel Capacity, Popular Video, Patching, Patching Window, True VOD

### 1. 서론

멀티미디어의 대표적인 형태인 주문형 비디오(Video on Demand) 서비스는 향후 가장 각광받게 될 서비스 중 하나로 가정 오락, 디지털 비디오 도서관, 주문형 뉴스, 원격리 교육, 전자 상업과 같은 많은 멀티미디어 용

용을 위한 중요한 기술이다[1, 2]. 전형적인 VOD 서비스는 서버 상에 대량의 비디오들이 저장되어 있으며, 원거리의 사용자는 장소와 시간에 제한받지 않고 자신이 원하는 비디오를 서버에 요청함으로써 서버로부터 전송된 비디오를 재생해서 보도록 해준다. 비디오 스트림은 비록 압축이 되더라도 대역폭 요구량이 높으며, 재생시간이 길고, 엄격한 응답시간을 요구하며, 연속적으로 재생되어야 하므로 전송 비용이 많이 든다. 하나의 비디오 스트림을 유지하는데 필요한 시스템 자원을 비디오 채널이라고 하는데 일반적으로 VOD 시스템에서 비디오 서버의 가용 채널 수는 통신 대역폭에 의해 결정되므로 한정되어 있다[2]. 비디오 서버의 가용 채널수가 한정되

· 본 연구는 2003학년도 대구가톨릭대학교 일반 연구비 지원에 의한 것임.

† 정희원 : 경북대학교 전자전기컴퓨터학부 교수  
siha@m80.knu.ac.kr

†† 정희원 : 대구가톨릭대학교 컴퓨터정보통신공학부 교수  
ihbae@cuth.cataegu.ac.kr

논문접수 : 2002년 10월 23일

심사완료 : 2003년 2월 6일

어 있으므로 여러 사용자들의 비동기적 비디오 요청에 대해 대용량의 비디오 스트림들을 경제적이며 적은 서비스 지연시간으로 많은 요청을 서비스할 수 있는 기법이 요구되었으며, 이를 위해 여러 사용자들이 비디오 스트림을 공유하는 방법들이 제안되었다[2, 3, 4, 5, 6, 7].

VOD 시스템에서 사용자들이 요청하는 비디오의 분포는 균등한 것이 아니라 대부분이 인기 비디오에 집중되어 있다. 그러므로 인기 비디오 요청에 대한 서비스 지연시간은 시스템의 평균 서비스 지연시간에 영향을 미칠 것이다. 자원 예약 패칭 방법[8, 9]에서는 이러한 인기 비디오 요청이 즉시 서비스될 수 있도록 비디오 서버의 채널 일부를 인기 비디오 요청 전용으로 예약한다. 채널 예약을 통해 인기 비디오 요청들의 이탈율을 감소시킴으로써 보다 많은 요청들을 서비스할 수 있으며 평균 서비스 지연시간을 개선시킨다. 그러나 서버의 예약된 채널 용량이 많다면 채널의 낭비가 발생하고 적다면 채널의 경쟁으로 인기 비디오를 즉시 서비스하기 어려워진다. 본 논문에서는 인기 비디오를 위한 보다 적절한 예약 채널 용량을 결정하는 두 가지 방법을 제안하고 이 두 방법의 시뮬레이션 결과를 통해 최적의 예약 채널 용량을 결정한다.

본 논문의 구성은 다음과 같다. 2장에서는 VOD 서비스를 위해 멀티캐스트를 이용하는 패칭과 비디오 서버의 채널 일부를 인기 비디오 전용으로 예약하는 단순 채널 예약 패칭을 소개하고, 3장에서는 단순 채널 예약 패칭의 성능을 개선하기 위한 두 가지 채널 예약 용량 결정 방법을 제안하고, 4장에서는 제안하는 예약 채널 용량 결정 방법들을 사용한 채널 예약 패칭 방법을 시뮬레이션을 통하여 평균 서비스 지연시간, 이탈율, 공평성을 근거로 패칭과 단순 채널 예약 패칭의 성능과 비교한다. 그리고 마지막으로 5장에서 결론을 맺는다.

## 2. 관련 연구

VOD 서비스의 성능을 개선하기 위해 비디오 서버의 네트워크 채널을 효율적으로 사용하기 위한 여러 가지 방법들이 제안되었다. 하나의 채널로 동일 비디오 스트림을 여러 비디오 요청자들에게 멀티캐스트함으로써 여러 사용자간에 비디오 스트림을 공유하며 서비스 지연시간 없이 즉시 서비스할 수 있는 패칭[2]이 있다. 최근 사용자 스테이션은 고성능화로 인해 서버로부터 동시에 두 개의 비디오 스트림을 전송 받을 수 있을 만큼의 대역폭을 가지게 되었으며, 수 분의 비디오 스트림을 버퍼링 할 수 있을 만큼의 충분한 디스크 공간을 가지게 되었다. 패칭은 이러한 특성을 이용하여 사용자 스테이션

에게 일반적인 멀티캐스트보다 두 배 높은 대역폭과 보다 많은 디스크 버퍼링 공간을 요구하지만 사용자에게 지연시간 없이 서비스할 수 있도록 제안된 방법이다. 한 개의 비디오 데이터 전체를 멀티캐스트하는데 사용되는 채널은 정규 채널이라고 하며, 비디오의 시작부터 일부 분만을 멀티캐스트하는 채널은 패칭 채널이라고 한다 [2]. 정규 채널과 패칭 채널을 통해 사용자가 수신하는 데이터 스트림들은 각각 정규 스트림과 패칭 스트림이라고 한다. 사용자가 비디오 서버에게 비디오  $i$ 를 요청하면 서버는 정규 채널을 통해 요청한 비디오  $i$ 의 전체 데이터를 전송한다. 그러나 이 정규 채널로 비디오  $i$ 의 서비스가 시작된 이후로부터 패칭 윈도우에 해당하는 시간이 지나가기 전에 도착한 동일 비디오  $i$ 에 대한 새로운 요청에 대해서는 비디오  $i$ 의 전체 데이터를 전송하는 것이 아니라 패칭 채널을 통해 스큐에 해당하는 비디오  $i$ 의 앞부분만을 사용자에게 전송한다. 여기서 스큐란 정규 채널상의 비디오 서비스 시작시간과 새 요청의 서비스 시작 시간간의 차이 시간이다. 이 때 사용자는 패칭 채널로 전송되는 비디오 스트림을 재생하면서 동시에 정규 채널로부터 멀티캐스트 중인 정규 비디오 스트림을 버퍼링함으로써 패칭 스트림을 모두 재생한 후 비디오의 뒷부분을 연속적으로 재생할 수 있다.

그림 1은 차례대로  $t_1, t_2, t_3$ 에 도착한 세 개의 비디오 요청  $R_1, R_2, R_3$ 가 서비스될 때의 채널 상태를 보여준다. 먼저  $t_1$ 에 도착한 요청  $R_1$ 은 정규 채널  $RCH_1$ 으로 서비스된다. 사용자 요청  $R_2$ 가  $t_2$ 에 도착했을 때 가용 채널인  $PCH_1$ 이 있다고 가정하자. 비디오 서버는 두 요청간의 차이 시간  $t_2 - t_1$ 이 패칭 윈도우 보다 작으므로 가용 채널  $PCH_1$ 을 패칭 채널로 할당할 후 스큐에 해당하는 비디오의 시작부분부터  $t_2 - t_1$ 까지의 비디오 데이터

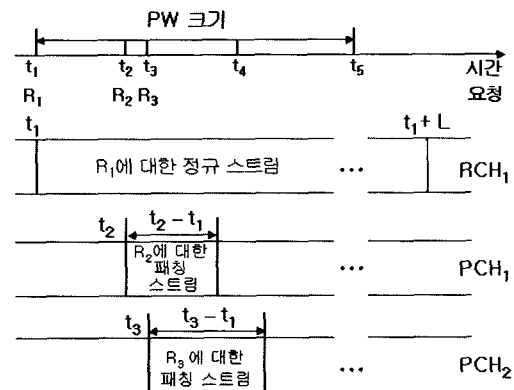


그림 1 패칭에서의 정규 채널과 패칭 채널

를 전송한다. 즉 요청  $R_2$ 를 위해 비디오 전체 재생시간 동안  $PCH_1$ 을 사용하여 비디오의 전체 데이터를 전송하는 것이 아니라,  $t_2$  이후부터는 정규 채널  $RCH_1$  상의 비디오 스트림을 멀티캐스트를 통해 공유함으로써  $t_2-t_1$  동안만 패칭 채널  $PCH_1$ 을 사용하면 되는 것이다.

그림 2는 그림 1의 시각  $t_4$  시점에서 각 사용자 스테이션의 상태를 보여준다. 사용자 스테이션의  $L_r$ 과  $L_p$ 는 각각 정규 채널과 패칭 채널로부터 전송되는 정규 스트림과 패칭 스트림을 다운로드하기 위한 다운로드이며 비디오 플레이어는 버퍼에 저장된 비디오 프레임들 조정하여 스크린에 상영하는 기능을 제공한다[2]. 요청  $R_1$ 의 사용자 스테이션에서는  $L_r$ 이 정규 채널  $RCH_1$ 으로부터 다운로드한 비디오 프레임을 비디오 플레이어가 상영하고 있다. 요청  $R_2$ 의 사용자 스테이션은  $L_p$ 가 패칭 채널  $PCH_1$ 으로부터 다운로드한 비디오 프레임을 모두 상영한 후, 버퍼에 저장되어 있는 정규 스트림을 상영하면서 계속해서  $L_r$ 이 정규 스트림을 다운로드하고 있다. 요청  $R_3$ 의 사용자 스테이션에서는  $L_p$ 가 패칭 채널  $PCH_2$ 로부터 패칭 스트림을 다운로드하여 비디오 플레이어가 이 패칭 스트림을 상영하며, 동시에 정규 채널  $RCH_1$ 으로부터 정규 스트림을 버퍼에 저장하고 있다.

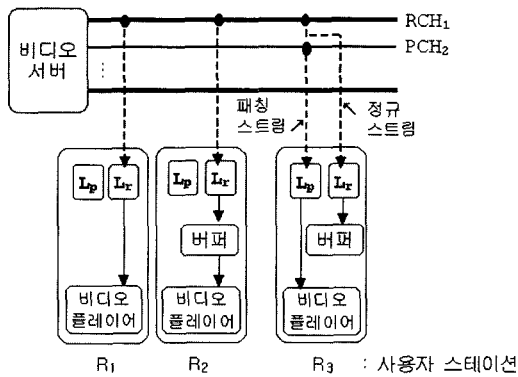


그림 2 그림 1의  $t_4$ 시점에서 사용자 스테이션의 상태

단순 채널 예약 패칭에서는 비디오 서버의 전체 채널 용량 중에서 일부를 인기 비디오 요청을 서비스하기 위해 미리 예약해두는 방법을 제안하였다[9]. 이 단순 채널 예약 패칭 방법은 다음과 같은 두 가지 사실에 근거하고 있다. 첫 번째, 사용자가 요청한 비디오에 대해 패칭을 적용하기 위해선 패칭 윈도우 안에 이미 동일한 비디오를 전송하고 있는 정규 채널이 존재해야 한다. 즉 정규 스트림이 존재하는 경우에만 패칭이 일어날 수 있

다. 그리고 새로운 사용자 요청이 도착했을 때 패칭 윈도우 안에 정규 스트림이 존재하더라도 비디오 서버에 가용 채널이 없다면 이 새로운 요청은 즉시 스케줄되지 못하고 대기 큐에서 기다려야 하며, 이후에 스케줄되었을 때는 패칭 윈도우 안에 정규 스트림이 존재하지 않아서 패칭 멀티캐스트가 수행되지 못하고 정규 멀티캐스트가 이루어져야 하는 상황이 될 수 있다. 두 번째, 사용자들의 요청은 모든 비디오에 대해 균등하게 분포되어 있지 않고 특정 인기 비디오에 편중되어 있다는 것이다[10, 11]. 요청율이 높은 인기 비디오에 대한 요청들은 자주 발생되므로 항상 새로운 요청의 패칭 윈도우 안에 전송 중인 정규 스트림이 존재하기 쉽다. 그러므로 단순 채널 예약 패칭은 인기 비디오의 요청이 항상 패치될 수 있도록 서버의 일부 채널을 인기 비디오 전용으로 예약해둠으로써 전체 요청 중에서 많은 비중을 차지하는 인기 비디오 요청들을 즉시 서비스하여 평균 서비스 지연 시간을 감소시키며 이탈율을 개선시켰다.

3. 채널 예약 패칭의 예약 채널 용량 결정 방법

단순 채널 예약 패칭에서 인기 비디오의 수  $k$ 는 서버가 저장하고 있는 모든 비디오 중에서 식 1을 만족하는 비디오 개수이며,  $k$ 개의 인기 비디오 요청을 위해 서버가 예약하는 채널 용량( $RCC_{hot}^1$ )은 식 2와 같이 결정된다[9].  $L$ 은 비디오의 길이,  $PW$ 는 패칭 윈도우의 크기,  $\lambda$ 는 전체 요청 도착율로서 분당 도착한 모든 비디오 요청 수를, 그리고  $p_i$ 는 비디오  $i$ 의 요청 확률로서 비디오의 인기도를 의미하며  $i$ 번 비디오의 요청 확률이  $i$ 번째로 높은 것으로 가정한다. 1분 동안 도착할 수 있는 비디오  $i$ 의 요청 수는  $p_i \times \lambda$ 이므로 식 1에 의해 1분 동안 도착 가능한 요청 수가 1개 이상이 되는 비디오가 인기 비디오로 간주된다.

$$p_i \times \lambda > 1.0 \tag{1}$$

$$RCC_{hot}^1 = (k \times \frac{L}{PW}) + \sum_{i=1}^k (p_i \times PW \times \lambda - 1) \tag{2}$$

비디오  $i$ 의 요청에 대해 정규 멀티캐스트가 시작된 후 패칭 윈도우만큼의 시간이 지나면 새로운 정규 채널로 정규 멀티캐스트가 시작되어야 한다. 비디오 길이  $L$ 분 동안 일어날 수 있는 정규 멀티캐스트의 회수는  $(L/PW)$ 이므로  $L$ 분 동안의 비디오 요청에 대해 지연 없이 즉시 서비스하기 위해 필요한 정규 채널 수는  $(L/PW)$ 이다. 비디오 요청을 서비스하기 위해 사용된 정규 채널은  $L$ 분이 지나면 다시 가용 상태가 되므로 비디오 서버가 작동하는 동안  $(L/PW)$ 개의 정규 채널만으

로 비디오  $i$ 에 대한 정규 멀티캐스트를 즉시 수행하는 것이 가능하다. 그러므로  $k$ 개의 인기 비디오를 위해 필요한 정규 채널 수는  $k \times (L/PW)$ 이 된다.

일단 비디오  $i$ 의 요청에 대해 정규 멀티캐스트가 시작되면 이 정규 멀티캐스트의 패칭 윈도우 안에 도착하여 서비스되는 요청들은 패칭 채널로 패칭 멀티캐스트가 수행된다. 패칭 윈도우  $PW$ 분 동안 도착 가능한 비디오  $i$ 의 총 요청 수는  $p_i \times \lambda \times PW$ 이지만 첫 번째 요청에 대해서는 정규 멀티캐스트가 수행되므로  $PW$ 분 동안 필요한 패칭 채널의 수는  $p_i \times \lambda \times PW - 1$ 이다. 이러한 패칭 채널들은 스큐에 해당하는 비디오 앞부분인 패칭 스트림을 전송한 후 가용 상태가 되므로 다음의 새로운 정규 멀티캐스트가 발생한 후부터 서비스되는 패칭 윈도우 안의 요청들을 위해 패칭 멀티캐스트를 수행하는데 다시 사용될 수 있다. 그러므로  $k$ 개의 인기 비디오를 즉시 서비스하기 위해 필요한 패칭 채널의 수는

$\sum_{i=1}^k (p_i \times \lambda \times PW - 1)$ 가 된다. 결국  $k$ 개의 인기 비디오를 위하여 정규 채널과 패칭 채널을 모두 합한 총 채널 예약 용량은 식 2가 된다.

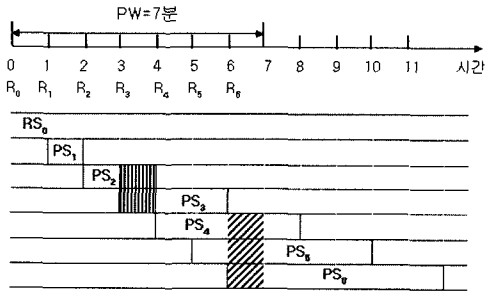
단순 채널 예약 패칭에서 인기 비디오 요청 전용으로 예약하는 서버의 채널 용량이 너무 많다면 채널의 낭비가 발생하고 너무 적으면 인기 비디오 요청자의 서비스 지연시간이 너무 커져서 요청자가 이탈할 수 있다. 그러므로 본 논문에서는 단순 채널 예약 패칭 기법의 성능을 개선시킬 수 있는 적절한 예약 채널 용량을 결정하는 두 가지 방법을 제안한다.

단순 채널 예약 패칭에서 채널 예약 용량을 결정하기 위해 사용한 식 2는 모든 인기 비디오의 요청이 서비스 지연시간 없이 즉시 서비스될 수 있도록 예약 채널 용량을 최대로 결정한 것이므로 실제적으로는 많은 낭비가 발생할 수 있다. 그림 3과 같이 패칭 윈도우가 7분이며 인기 비디오  $i$ 에 대한 요청이 매분마다 일정한 간격으로 도착하는 경우를 고려해보자. 먼저 서버는 요청  $R_0$ 에 대해 정규 채널 상으로 비디오 재생 시간동안 정규 스트림  $RS_0$ 을 전송함으로써 서비스한다. 그러나 요청  $R_1$ 부터  $R_6$ 는 패칭 윈도우 안에 요청  $R_0$ 에 대한 정규 멀티캐스트가 존재하므로 각자의 해당 스큐만큼의 비디오의 앞부분이 패칭 채널을 통해 전송된다. 그러므로 단순 채널 예약 패칭[9]에서는 비디오  $i$ 에 대해 1개의 정규 채널과 6개의 패칭 채널을 예약한다. 그러나 패칭 채널은 스큐만큼의 패칭 스트림만 전송하면 가용상태가 되어 재사용될 수 있으므로 최대로 패칭 채널을 예약한다면 낭비가 발생할 수 있다. 그림 3(b)는 요청  $R_1$ 부터

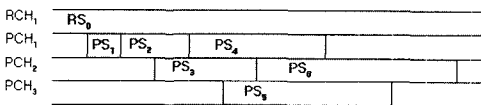
$R_6$ 를 위해 실제로 필요한 패칭 채널이 3개임을 보여 준다.  $R_1$ 은 패칭 윈도우 안에 정규 멀티캐스트가 존재하므로 스큐에 해당하는 1분 동안 패칭 채널  $PCH_1$ 으로 패칭 스트림  $PS_1$ 이 전송되고 이후 패칭 채널  $PCH_1$ 은 가용상태가 된다. 요청  $R_2$  역시 패칭 윈도우 안에 정규 채널  $RCH_1$ 상으로 정규 멀티캐스트가 존재하므로 가용상태가 된 패칭 채널  $PCH_1$ 으로 2분 동안 패칭 스트림  $PS_2$ 가 전송된다. 요청  $R_3$ 는 패칭 채널  $PCH_2$ 로 패칭 스트림  $PS_3$ 가 전송된다. 요청  $R_4$ 는 패칭 채널  $PCH_1$ 이 4분에 가용 상태가 되므로  $PCH_1$ 으로 패칭 스트림  $PS_4$ 가 전송된다. 요청  $R_5$ 는 새로운 패칭 채널  $PCH_3$ 로 패칭 스트림  $PS_5$ 가 전송된다. 마지막으로 요청  $R_6$  역시 패칭 윈도우 안에 전송중인 정규 스트림  $RS_0$ 가 존재하므로 가용상태가 된 패칭 채널  $PCH_2$ 으로 패칭 스트림  $PS_6$ 가 전송된다. 그러므로 실제로 필요한 패칭 채널은 6개가 아니라 3개이며 그림 3(a)에서 빗금이 쳐진 부분처럼 특정 시간에 동시에 전송되어야 하는 패칭 스트림 수와 동일함을 알 수 있다.

패칭 윈도우 내의 요청들을 즉시 서비스하기 위해 필요한 최소 패칭 채널 수는 그림 3을 관찰함으로써 패칭 윈도우의 크기  $PW$ , 비디오  $i$ 의 요청 확률  $p_i$ , 요청 도착률  $\lambda$ 를 이용하여 구할 수 있다. 그림 3(a)에서 패칭 윈도우가 4분이라고 가정한다면 요청 도착률  $\lambda=1$ 이므로 패칭 윈도우 안에 도착 가능한 요청 수는 4개이며 예약해야 할 최소 패칭 채널 수는 세로줄이 쳐진 부분인  $\lfloor 4/2 \rfloor = 2$ 개이며, 패칭 윈도우가 5분이라면 도착 가능한 5개의 요청에 대해  $\lfloor 5/2 \rfloor = 2$ 개, 패칭 윈도우가 6분이라면 도착 가능한 6개의 요청에 대해  $\lfloor 6/2 \rfloor = 3$ 개의 채널이 된다. 즉 패칭 윈도우  $PW$ 분 안에 도착할 수 있는 비디오  $i$ 에 대한 총 요청 수는  $N_R = p_i \times \lambda \times PW$ 이고 정규 멀티캐스트가 수행되는 첫 번째 요청을 제외한 패칭 윈도우 안의  $N_R - 1$ 개의 요청을 서비스하기 위해 패칭 멀티캐스트가 수행될 때, 동시에 전송이 수행되고 있는 패칭 채널의 최대 수는  $\lfloor N_R/2 \rfloor$  임을 알 수 있다.

이와 같은 관찰로부터 본 논문에서는 과다한 채널의 예약으로 인한 낭비를 제거하면서 서비스 지연시간을 최소화할 수 있도록  $k$ 개의 인기 비디오 요청을 위한 채널 예약 용량( $RCC_{hot}^2$ )을 식 3과 같이 구하는 예약 채널 용량 결정 방법을 제안한다. 여기서 첫 번째 항은 식 2의 첫 번째 항으로  $k$ 개의 인기 비디오를 위해 예약할 최대 정규 채널 용량을, 두 번째 항은 그림 3으로부터 얻어진 결과를 사용하여 예약할 패칭 채널 용량을 나타낸다.



(a) \$R\_0 \sim R\_6\$ 요청의 서비스 시작시간과 패칭 스트림



(b) \$R\_0 \sim R\_6\$ 요청을 서비스하고 있는 채널의 상태

그림 3 비디오 요청과 채널의 상태

$$RCC_{hot}^2 = (k \times \frac{L}{PW}) + \sum_{i=1}^k \left[ \left( \frac{p_i \times \lambda \times PW}{2} \right) \right] \quad (3)$$

첫 번째 예약 채널 용량 결정 방법은 인기 비디오 요청들이 일정 간격으로 도착한다고 가정했을 때 서비스 지연시간 없이 즉시 서비스하기 위해서 필요한 최소의 정규 채널과 패칭 채널을 예약하는 방법이다. 그러므로 최악의 경우로 비디오 요청들이 일정한 간격으로 도착하지 않고 어느 한순간에 집중적으로 도착한다면 단순 채널 예약 패칭과 같이 패칭 채널을 최대한 예약하지 않음으로써 패칭 채널이 부족하여 서비스 지연시간이 생길 수 있다. 그러나, 패칭 윈도우 안의 \$N\_R = p\_i \times \lambda \times PW\$ 개의 요청 중에서 서비스 지연 시간이 발생할 수 있는 최대 요청 수는 \$\lceil N\_R/2 \rceil - 1\$이며 패칭 채널의 데이터 전송시간이 스큐에 해당하는 작은 시간이므로 패칭 채널이 곧 가용상태가 될 수 있으므로 서비스 지연 시간이 크지 않을 것이다.

단순 채널 예약 패칭과 제안한 예약 채널 용량 결정 방법은 비디오 요청 도착율의 증가로 인기 비디오 수 \$k\$가 증가하면 예약 채널의 용량이 과다하게 증가함으로써 서버가 과부하일 때 비예약 채널의 용량이 너무 작아져서 비인기 비디오 요청들의 서비스가 제대로 수행될 수 없다는 문제점이 있다. 이를 개선하기 위해 두 번째로 제안하는 예약 채널 용량 결정 방법은 인기 비디오 요청들이 비인기 비디오 요청의 서비스로 인해 가용 채널이 없어서 패칭 기회를 놓치지 않도록 하면서도 비

인기 비디오 요청의 서비스를 위한 비예약 채널 용량을 확보할 수 있도록 하는 방법이다. 전체 비디오 요청 중에서 \$k\$개의 인기 비디오에 대한 요청이 차지하는 비율과 비디오 서버의 전체 채널 용량 중에서 예약 채널 용량이 차지하는 비율이 같도록 결정한다. \$CH\_{server}\$는 비디오 서버의 전체 채널 용량이라고 할 때, 전체 비디오 요청 중에서 \$k\$개의 인기 비디오 요청이 차지하는 비율은 \$\sum\_{i=1}^k p\_i\$이므로 두 번째 제안 방법에서 \$k\$개의 인기 비디오를 위한 예약 채널 용량(\$RCC\_{hot}^3\$)은 식 4와 같이 결정한다.

$$RCC_{hot}^3 = CH_{server} \times \sum_{i=1}^k p_i \quad (4)$$

#### 4. 시뮬레이션 및 성능 평가

본 논문에서는 시뮬레이션을 통하여 서버의 채널 일부를 예약하지 않는 단순 패칭, 인기 비디오 요청이 즉시 서비스될 수 있도록 최대한 서버의 채널 일부를 예약하는 단순 채널 예약 패칭 그리고 3장에서 제안한 두 가지 예약 채널 용량 결정 방법을 사용한 채널 예약 패칭의 성능을 평가하고 비교한다. 성능 평가에서 고려된 평가 항목은 VOD 서버의 성능 평가에 주로 사용되는 비디오 요청 도착율에 따른 사용자의 이탈율, 평균 대기 시간과 공평성이다.

시뮬레이션은 IBM PC-Pentium IV에서 Visual C++로 작성한 프로그램을 사용하여 수행되었으며 사용된 파라미터는 표 1과 같다. 비디오 관람 패턴에 대한 통계에 의하면 특정 비디오에 대해 인기도가 편중되며 이런 지역성은 Zipf 법칙을 사용하여 표현할 수 있기 때문에 비디오 \$i\$의 인기도(\$p\_i\$)는 식 5와 같이 Zipf 법칙을 따른다고 가정한다[10, 11].

$$p_i = \frac{1}{i^z \sum_{j=1}^N \frac{1}{j^z}} \quad (5)$$

여기서 \$N\$은 전체 비디오의 수이며 \$z\$는 스큐 인자(skew factor)로서 VOD 응용에서 일반적으로 나타나는 0.7을 사용하였다[4]. 비디오의 길이는 단순 패칭[2]에서와 마찬가지로 일반적인 영화 길이인 90분으로 설정하였으며 비디오를 요청한 사용자가 서비스를 기다리다가 이탈하는 시간은 관람자들의 일반적인 행동을 반영하여 3분~7분으로 랜덤하게 설정하였다.

패칭 윈도우의 크기는 Cai[3]의 시뮬레이션 결과에서 얻어진 5분으로 설정하였으며 이는 사용자가 5분 동안의 스트림을 저장할 수 있는 충분한 버퍼공간을 가진다

표 1 시뮬레이션 파라미터

파라미터	값
비디오 개수	100개
서버 네트워크 채널	600 채널
패칭 윈도우	5분
도착율/분	10, 20, 40, 60, 80 (포아송)
비디오 길이	90분
이탈시간	3분~7분 (랜덤)

는 것을 의미한다. 또한 사용자는 정규 스트림과 패칭 스트림을 동시에 전송받을 수 있는 충분한 대역폭을 가지며 비디오를 시작부터 끝까지 순차적으로 본다 고 가정한다. 시뮬레이션은 360분 동안의 사용자 요청에 대해서 수행되었으며 공평성을 위해 서비스를 시작한 후 90분과 320분 사이에 도착한 요청들에 대해 각 패칭의 성능을 평가하였다.

그림 4와 5는 패칭 정책들의 요청 도착율에 따른 이탈율과 예약된 채널 용량을 보여준다. 이탈율은 시뮬레이션 기간동안 도착한 전체 비디오 요청 중에서 서비스를 받지 못하고 이탈한 요청이 차지하는 비율로서 이탈한 요청 수를 전체 요청 수로 나눈 값이다. S-Patching은 단순 패칭, R-Patching은 단순 채널 예약 패칭, R-Patching1은 첫 번째로 제안한 예약 채널 용량 결정 방법을 사용하는 채널 예약 패칭, R-Patching2는 두 번째

로 제안한 예약 채널 용량 결정 방법을 사용하는 채널 예약 패칭을 나타낸다. 요청 도착율이 10일 때는 모든 패칭의 이탈율이 0으로써 사용자 전체의 요청이 서비스 지연시간 없이 즉시 서비스를 받았다. 단순 패칭은 요청 도착율이 20일 때에만 채널 예약 패칭들보다 조금 우수한 것으로 나타났으며 도착율이 40일 때부터는 가장 좋지 않음을 알 수 있다. 이처럼 서버가 과부하 상태일수록 단순 패칭의 이탈율이 높게 나타난 것은 단순 패칭에서는 서버가 과부하 상태일 때 모든 비디오 요청들이 대기 큐에서 블록되어 기다리고 있는 반면에 채널 예약 패칭들은 인기 비디오 요청들을 위한 충분한 서버 채널이 미리 예약되어 있기 때문에 요청 수가 많은 인기 비디오 요청들이 빨리 서비스될 수 있기 때문이다. 채널 예약 패칭들의 결과를 비교해보면 도착율이 40일때까지는 첫 번째 제안 방법인 R-Patching1이 가장 우수하고, 도착율이 더 높아지면 두 번째 제안 방법인 R-Patching2가 가장 우수함을 알 수 있다. 이러한 채널 예약 패칭의 이탈율 결과 그래프는 그림 5에서 나타난 것처럼 각 채널 예약 패칭에서 결정된 예약 채널 용량 결과 그래프와 같은 모양을 하고 있다. 즉 그림 5에서 도착율이 40이하일 때는 R-Patching1의 예약 채널 용량이 가장 작으며 도착율이 40보다 클 때는 R-Patching2의 예약 채널 용량이 가장 작으며 이탈율 또한 같은 결과를 보여준다. 이것은 인기 비디오를 위해 예약한 채널의 용량이 너무 많아지면 비인기 비디오 요청들이 사용할 수 있는 채널 용량이 너무 작아져서 비인기 비디오 요청들간의 채널 결정이 심해지고 이로 인해 이들의 이탈이 증가되어 전체 이탈율이 증가되기 때문이다. 그러므로 이탈율을 감소시키기 위해서는 예약 채널의 용량 ( $RCC_{hot}$ )을 식 3과 식 4의 최소 값을 취하는 식 6을 따라 결정해야 한다.

$$RCC_{hot} = \text{Min} \left[ \left( k \times \frac{L}{PW} \right) + \sum_{i=1}^n \left( \frac{p_i \times \lambda \times PW}{2} \right) \right], \quad (6)$$

$$CH_{server} \times \sum_{i=1}^n p_i$$

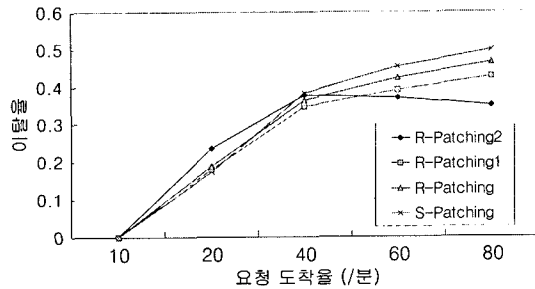


그림 4 요청 도착율에 따른 이탈율

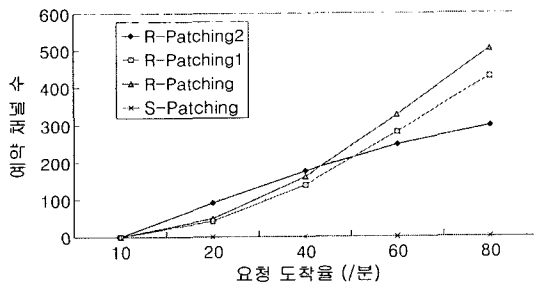


그림 5 요청 도착율에 따른 예약 채널 수

그림 6은 각 패칭 정책의 비디오 요청 도착율에 따른 서비스 지연 시간을 보여주고 있다. 제안하는 방법들이 단순 패칭보다 서비스 지연 시간이 적으며 도착율이 60일 때를 제외하곤 R-Patching1이 가장 우수함을 알 수 있다. 이탈율의 평가 결과와 다르게 서비스 지연 시간의 평가 결과에서는 도착율이 80일 때 R-Patching2의 서비스 지연 시간이 R-Patching1보다 7초 정도 더 많았다. 이는 예약된 채널의 수가 R-Patching1보다 작으며 이로 인해 인기 비디오 요청들의 평균 서비스 지연 시간이 0이 아닌 37.8초로 나왔기 때문이다. 그러나 요청

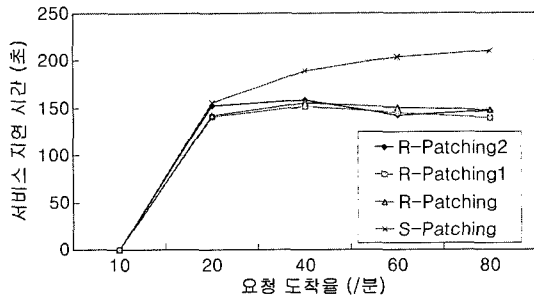


그림 6 요청 도착율에 따른 서비스 지연시간

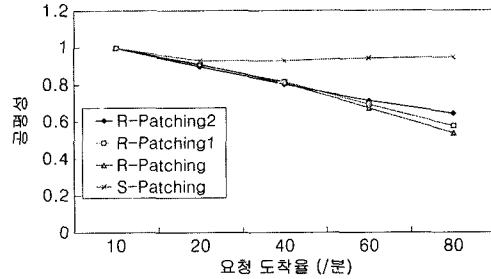


그림 8 요청 도착율에 따른 공평성

도착율에 따라 실제 서비스된 요청 수를 나타내고 있는 그림 7을 보면 도착율이 80일 때 R-Patching2에서 실제로 서비스된 요청 수는 R-Patching1보다 1,580개가 더 많으므로 평균 서비스 지연시간이 7초 정도 낮다 하더라도 서비스된 요청의 수가 훨씬 많은 R-Patching2를 더 좋은 방법으로 선택할 수 있을 것이다. 이와 같이 이탈율과 평균 서비스 지연시간을 비교해본 결과 요청 도착율에 따라 결정된 예약 채널의 용량이 적은 방법 즉 도착율이 40이하라면 첫 번째 제안 방법이 도착율이 40보다 높을 때는 두 번째 제안 방법의 성능이 우수함을 알 수 있다.

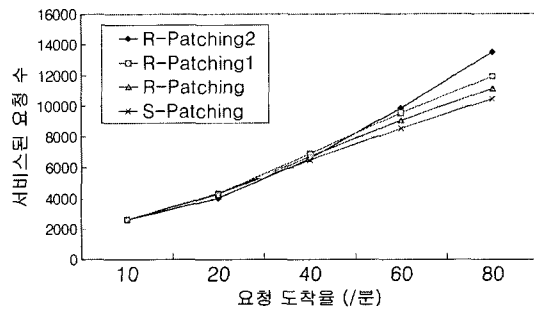


그림 7 요청 도착율별로 서비스된 요청의 수

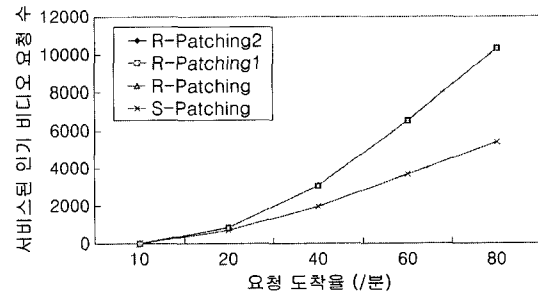


그림 9 서비스된 인기 비디오 요청 수

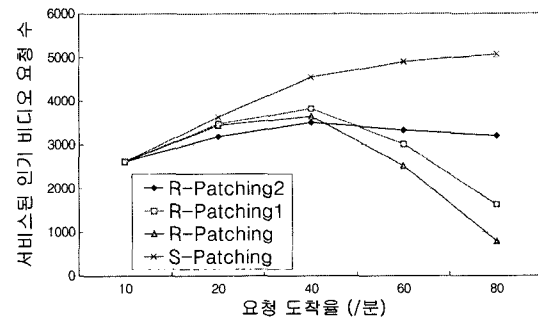


그림 10 서비스된 비인기 비디오 요청 수

그림 8은 비디오 요청 도착율에 따른 각 패칭 정책별 공평성을 보여주며, 그림 9와 그림 10은 각각 서비스 받은 인기 비디오 요청 수와 비인기 비디오 요청 수를 보여준다. 공평성(fairness)은 평균 이탈율이  $d$ 이고 각 비디오  $i$ 에 대한 이탈율이  $d_i$ 라고 할 때 식 7과 같이 계산된다[2].

공평성은 단순 패칭이 가장 좋으며, 서비스된 인기 비디오 요청수는 단순 패칭 보다는 채널 예약 패칭 방법들이 훨씬 많음을 알 수 있다. 그리고 그림 10으로부터

$$fairness = 1 - \sqrt{\frac{\sum_{i=1}^N (d_i - d)^2}{N-1}} \quad (7)$$

서비스된 비인기 비디오 요청수는 단순 패칭이 가장 많으며 채널 예약 패칭 방법 중에서는 이탈율의 평가 결과와 같이 도착율이 40이하일 때는 R-Patching1이, 도착율이 더 높을 때는 R-Patching2가 더 많음을 알 수 있다. 특히 비디오 요청 도착율이 80일 때 R-Patching2에서 서비스된 인기 비디오 요청수는 단순 패칭보다 4,930개의 요청이 추가로 서비스되었으며, 서비스된 비인기 비디오 요청의 수는 단순 패칭에 비해 1,860개가 감소되었다. 즉 채널 예약 패칭의 공평성이 떨어진 것은

인기 비디오의 이탈율이 패칭에 비해 최소화되었고 인기 비디오의 이탈율이 패칭보다 증가됨으로써 비디오 간의 이탈율의 차이가 심해졌기 때문이다. 그러므로 단순 패칭에 비해 비디오 전체에 대한 공평성은 많이 낮아졌으나 서버의 과부하로 인해 많은 요청들이 이탈될 수밖에 없는 상황에서 훨씬 더 많은 요청들을 서비스할 수 있다는 점을 고려할 때 공평성의 저하는 크게 문제 되지 않을 것이다.

지금까지의 시뮬레이션 결과를 분석해 볼 때 보다 많은 사용자 요청을 서비스할 수 있으면서도 서비스 지연 시간을 감소시키기 위해서는 단순 패칭보다는 인기 비디오 요청을 위하여 서버의 채널 일부를 예약하는 채널 예약 패칭이 좋은 방법이며 제안한 두 가지 예약 채널 용량 결정 방법을 혼합하여 둘 중에서 채널 예약 용량이 최소인 것을 선택하는 방법 즉 식 6에 의해 예약 채널 용량을 결정하는 방법이 가장 적절한 방법이라고 할 수 있다.

## 5. 결론

VOD 서비스의 성능을 높이기 위해서는 비디오 서버의 네트워크 대역폭에 의해 결정되는 한정된 채널 자원을 효율적으로 사용하면서도 서비스 지연시간과 이탈율을 감소시킬 수 있는 방법이 필요하다. VOD 서비스를 요청하는 사용자들의 비디오들을 분석해보면 모든 비디오가 균등하게 요청되는 것이 아니라 특정 인기 비디오에 집중된다는 특징을 가지고 있다. 본 논문에서는 이러한 사용자의 요청 특성에 근거하여 인기 비디오 요청을 위해 비디오 서버의 채널 일부를 미리 예약하는 단순 채널 예약 패칭의 성능을 개선하기 위해 보다 적절한 예약 채널 용량을 결정하는 두 가지 방법을 제안하고 두 방법들의 성능을 시뮬레이션을 통해 평가하였다. 시뮬레이션 결과 제안한 두 가지 방법에서 결정된 예약 채널 용량 중에서 최소인 값을 최종 예약 채널 용량으로 선택하는 것이 가장 좋은 방법임을 알 수 있었다. 이탈율과 평균 서비스 지연시간은 거의 모든 요청 도착율에 대해 제안하는 방법들이 우수하였으며 특히 서버가 과부하 상태일수록 전체 비디오 요청 중에서 인기 비디오 요청이 차지하는 비율과 동일하게 서버의 채널 일부를 예약하는 두 번째 제안 방법이 우수함을 알 수 있었다.

## 참고 문헌

- [1] Jani Huoponen and Thorsten Wagner, "Video on Demand A Survey", telecommunication Networks Project, 1, [http://fiddle.visc.vt.edu/courses/ee4984/Projects1996/huoponen\\_wagner/huoponen\\_wagner.html](http://fiddle.visc.vt.edu/courses/ee4984/Projects1996/huoponen_wagner/huoponen_wagner.html), 1996.
- [2] K. Hua, Y. Cai, and S. Sheu, "Patching: A Multicast Technique for True Video-on-Demand Services", *In Proc. ACM Multimedia*, pp. 191-200, 1998.
- [3] Y. Cai, K. Hua, and K. Vu, "Optimizing Patching Performance", *In Proc. SPIE/ACM Conference on Multimedia Computing and Networking*, pp. 204-215, 1999.
- [4] A. Dan, D. Sitaram, and P. Shahabuddin, "Scheduling Policies for an On-Demand Video Server with Batching", *In Proc. of the 2nd ACM Multimedia Conference*, pp. 25-32, 1994.
- [5] D. L. Eager, M. K. Vernon, and J. Zahorjan, "Optimal and Efficient Merging Schedules for Video-on-Demand Servers", *In Proc. of the 7-th ACM International Multimedia Conference*, pp. 199-203, 1999.
- [6] C. C. Aggarwal, J. L. Wolf, and P. S. Yu, "On Optimal Piggyback Merging Policies for Video-On-Demand Systems", *In Proc. 1996 ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems*, pp. 200-209, 1996.
- [7] K. Hua and S. Sheu, "Skyscraper broadcasting: A New Broadcasting Scheme for Metropolitan Video-on-Demand Systems", *In Proc. ACM SIGCOMM*, pp. 89-100, 1997.
- [8] 이주영, 이경숙, 오선진, 배인한, "True VOD 시스템을 위한 자원 예약 기반 패칭 방법", 인터넷정보학회 학술발표논문집, 제2권 2호, pp. 212-216, 2001.
- [9] 이주영, 하숙정, 배인한, "True VOD 시스템을 위한 채널 예약 패칭 방법의 설계 및 평가", 정보처리학회논문지, 제9-B권, 제6호, pp. 835-844, 2002.
- [10] A. Chervenak, D. Patterson, and R. Katz, "Choosing the Best Storage System for Video Service", *In Proc. of ACM Multimedia 95*, pp. 109-119, Aug. 1995.
- [11] Christer Samueleson, "Relating Turing's Formula and Zipf's Law", Proceedings of the Fourth Workshop on Very Large Corpora, 1996.

[1] Jani Huoponen and Thorsten Wagner, "Video on Demand A Survey", telecommunication Networks





하 숙 정

1988년 계명대학교 전자계산학과(학사).  
1990년 중앙대학교 전자계산학과(석사).  
1998년 대구가톨릭대학교 전산통계학과  
전산전공(박사). 2001년~2003년 경북대  
학교 전자전기컴퓨터학부 조교수(BK21).  
2003년~현재 경북대학교 전자전기컴퓨  
터학부 초빙교수. 관심분야는 모바일컴퓨팅, 분산컴퓨팅, 멀  
티미디어 시스템 등



배 인 한

1984년 경남대학교 전자계산학과(학사).  
1986년 중앙대학교 대학원 전자계산학과  
(석사). 1990년 중앙대학교 대학원 전자  
계산학과(박사). 1996년~1997년 Compu  
ter and Information Science, The Ohio  
State University(Post-doctoral). 2002  
년~현재 Department of Computer Science, Old Dominion  
University(Visiting Professor). 1989년~현재 대구가톨릭  
대학교 컴퓨터정보통신공학부 교수. 관심분야는 이동무선망,  
무선인터넷, M-Commerce, 멀티미디어 시스템 등