

사용자 이용 행태를 고려한 VoD 서버 부하 분산 방법[☆]

A VoD Load Balancing Algorithm Based on User's Behaviors

이 강 용*
Kang Yong Lee

장 욱**
Uk Jang

조 기 성***
Kee-Seong Cho

요 약

인터넷의 확산과 멀티미디어 콘텐츠의 활성화로 인해 최근 방송과 통신의 융합 서비스인 IPTV가 인터넷 서비스의 새로운 모델로 주목 받고 있다. 특히, IPTV 환경에서는 기존의 실시간 방송이나 주문형 비디오 (Video-on-Demand) 뿐만 아니라 양방향 멀티미디어 서비스나 고품질 디지털 콘텐츠 등과 같은 다양한 서비스를 제공한다. 하지만, IPTV 가입자수의 폭발적인 증가와 방대한 콘텐츠에 따른 트래픽의 증가는 IPTV 환경에서의 효과적인 네트워크 자원 관리에 어려움을 발생시키고, 결과적으로 품질 저하를 초래하게 된다. 본 논문에서는 IPTV 환경에서 네트워크 자원을 보다 효율적으로 이용하기 위해 사용자의 VoD 서비스 이용 행태를 고려하여 향후 잠재적인 서버의 부하 및 대기 시간에 대한 예상치를 계산하고 이를 서버 부하 분산에 적용하는 알고리즘을 제안한다.

ABSTRACT

During the last decade, Internet protocol television(IPTV), broadcasting of multimedia services using well understood Internet technologies, is a fast spreading technology for distributing multimedia content to consumers. As IPTV services advance, an increasing user demand and tremendous content volume of multimedia content cause some difficulties in network resource management. In this paper, we propose an improved load balancing algorithm for video-on-demand(VoD) service. Unlike existing algorithms, the proposed algorithm considers user's behaviors for VoD service, and determines the most efficient allocation of VoD requests by estimating expected server load and expected user waiting time.

☞ KeyWords : 자원 관리, 서버 부하 분산, 사용자 이용 행태 영문키워드: IPTV, VoD Server, Load Balancing, User Behaviors

1. 서 론

인터넷의 확산과 멀티미디어 콘텐츠의 활성화로 인해 방송 통신의 융합 서비스의 대표적인 IPTV(Internet Protocol Television)는 기존 아날로그 형태의 방송 서비스와는 다른 새로운 개념의 디지털 멀티미디어 전달 매체로서, 인터넷의 IP

(Internet Protocol)를 기반으로 TV를 통해 실시간 방송 및 VoD(Video-on-Demand)를 시청하고, 다양한 융합 서비스를 제공하는 양방향 멀티미디어 서비스이다. 현재 IPTV는 차세대 인터넷 서비스의 새로운 수익 모델로 주목 받고 있으며, 현재 전세계적으로 약 800여 개 이상의 IPTV 사업자들이 3,300만명 이상의 가입자에게 시범 또는 상용 서비스의 형태로 IPTV 서비스를 제공하고 있다 [1]. 또한, 국내에서도 KT, SK브로드밴드, LG유플러스 등을 비롯하여, 다수의 케이블 TV 사업자나 방송 사업자 등이 IPTV 서비스를 제공하고 있거나, 이를 위한 서비스를 준비하고 있다.

이처럼 IPTV가 본격적으로 확산됨에 따라 점차 서비스 품질에 대한 관심이 점차 커지고 있는데, IPTV는 그 서비스의 특성 상, 미디어 및 전달망에서 신뢰성(Reliability), 보안성(Security), 및 품

* 정 회 원 : 한국전자통신연구원 방송통신융합연구부문 연구원 kanglee@etri.re.kr(교신저자)

** 정 회 원 : 한국전자통신연구원 방송통신융합연구부문 연구원 ukjang@etri.re.kr

*** 정 회 원 : 한국전자통신연구원 방송통신융합연구부문 스마트서비스제어연구팀 팀장 chokis@etri.re.kr

[2010/08/31 투고 - 2010/09/01 심사 - 2010/09/27 심사완료]

☆ 본 연구는 지식경제부/방송통신위원회/한국산업기술평가원의 “차세대 IPTV 인프라 기반 Social TV 서비스 인에이블러 개발” 연구 과제의 지원을 받아 수행되었습니다.

질 보장성(Quality of Service)을 보장하여야 한다. 하지만, IPTV 가입자수의 폭발적인 증가와, 이에 따른 트래픽 양이 증가는 IPTV 환경에서의 효과적인 네트워크 자원 관리에 어려움을 발생시키고, 결과적으로 서비스 품질 저하를 초래하게 된다[2].

본 논문에서는 IPTV의 VoD 서비스에서 네트워크 자원을 보다 효율적으로 이용하기 위해 사용자의 VoD 서비스 이용 행태(User Behaviors) 정보를 활용한 VoD 서버의 효율적인 부하 분산 알고리즘에 대해 제안한다. 최근 IPTV에서 방송 서비스나 VoD 서비스에 대한 사용자의 서비스 이용 행태를 분석한 연구는 많이 진행 되어 왔으나 [3,4], 아직까지 이러한 정보를 활용한 네트워크 자원 관리 기법은 많이 연구되어 있지 않다. 본 논문에서 제안하는 알고리즘은 사용자의 서비스 이용행태를 고려하여, VoD 시청 종료 시간을 확률적으로 모델링하고 이를 통해 향후 VoD 서버의 잠재적인 부하와 가용 대역폭 예상치를 계산하여 서버 부하 분산 방법에 적용한다.

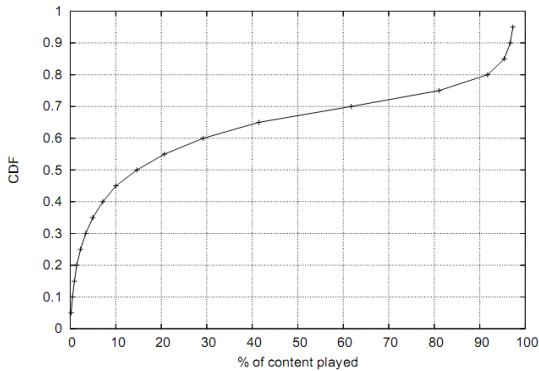
본 논문의 구성은 다음과 같다. 2장에서는 VoD 서버 부하 분산 기법에 관한 관련 연구 동향을 소개하고, 3장에서는 VoD 서비스 특성과 사용자의 서비스 이용 행태에 관한 내용을 다룬다. 4장에서는 본 논문에서 제안하고자 하는 VoD 서버 부하 분산 알고리즘에 대한 자세한 설명을 하고 5장에서는 제안한 알고리즘의 성능을 평가하기 위해 수행된 실험 결과를 분석한다. 마지막으로 6장에서는 본 논문의 결론을 맺는다.

2. 관련 연구 동향

VoD 서버 부하 분산 방법과 관련해서는 콘텐츠 관리(Cache management, placement, replacement), 디스크 스케줄링, 수락 제어(Admission control) 등 다양한 분야에 걸쳐서 많은 연구가 이루어지고 있다. VoD 서버 부하 분산 기법이란, 사용자의 VoD 서비스 요청을 미리 정해진 특정 기준에 따라 여러 미디어 서버들에 분산시키는 방법을 말한다. 특히, VoD 서비스의 경우, 사용자들의 서비

스 요청이 일정하게 발생하는 것이 아니라 서비스를 이용하는 시간과 사용자들의 위치 분포에 따라서 치우침 현상이 빈번하게 존재하게 된다. 따라서, VoD 서비스 요청이 특정 미디어 서버에만 집중되는 문제가 발생 할 수 있으며, 이 경우, 전체 네트워크 측면에서는 충분한 서버 자원이 존재함에도 불구하고 해당 서버의 용량 한계로 인하여 서비스 요청을 수용하지 못하고 차단하게 된다. 특히 IPTV 서비스와 같이 가입자의 수가 많고 다양한 콘텐츠에 따라 트래픽 양이 방대한 환경에서는 이러한 문제를 해결하기 위해 서비스 차단률(Blocking Probability)을 낮추고 네트워크 자원을 효율적으로 이용할 수 있도록 해주는 부하 분산 기법이 매우 중요하다.

부하 분산 기법과 관련해 최근의 연구들에 대해 살펴보면, [5]는 서버의 이용도(utilization)를 최대화하기 위한 부하 분산 알고리즘을 제안하고 있는데 부하를 분산시키기 위해서 전송 지연, 서버 별 수락제어와 콘텐츠의 인기도를 함께 고려하여 디스크 대역폭 측면의 효율성을 높이고자 하였다. [6]에서는 중앙 멀티미디어 서버와 트래커(tracker)를 통해서 관리되는 여러 지역의 프록시 서버 그룹(local proxy server group) 들로 구성된 네트워크 환경을 고려하고 있다. 지역 프록시 서버 그룹은 서로 인접하게 위치한 6개의 미디어 서버들로 이루어진 서버 그룹으로서, 콘텐츠들의 유효성과 인기도를 고려하여 분산/배치함으로써 중앙 미디어 서버에 트래픽이 집중되는 것을 방지한다. 이 논문의 경우, 네트워크 대역폭, 전송 지연 등을 고려하고 있지 않아 VoD 서비스에서 중요한 요소인 서비스 대기 시간 지연 등의 품질을 보장하지 못한다는 단점을 지니고 있다. [7]는 사용자들의 서비스 대기 시간을 줄이기 위한 부하 분산을 수행하고 있는 논문이다. 서버에 위치한 디스크 수량과 대역폭, 그리고 콘텐츠 인기도 등을 고려하고 있다. [10]에서는 클러스터 기반 VoD 서버에서 동적 버퍼 분할 방법을 통해 부하를 분산하는 알고리즘을 제안하였다. 제안한 방법은 사용자 요청을 처리하는 VoD 서버의 버퍼 성



(그림 1) VoD 콘텐츠 총 시간 대 재생 시간 비율 분포

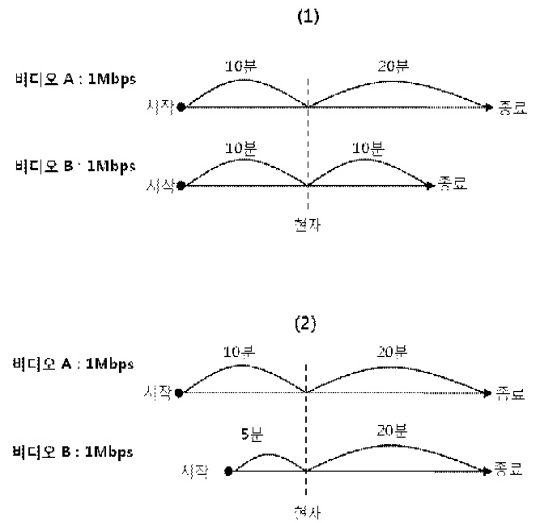
능과 디스크 접근 빈도를 고려하여 전체 부하를 고르게 분산하도록 하고, 동적 버퍼 분할을 통해 동일한 VoD 접근하려는 여러 사용자들의 평균 대기 시간을 감소하는 효과를 거두었다.

이 외에도 VoD 서비스의 부하 분산 방법과 관련하여 많은 연구들이 진행되어 왔지만 대부분의 부하 분산 기법들은 현재의 서버 부하 및 트래픽 상황만을 고려하고 있으며, VoD 서비스의 특성과 사용자의 서비스 이용 행태를 기반으로 하는 연구는 거의 이루어 지지 않았다.

3. 사용자의 VoD 서비스 이용 행태

[8], [9]에 의하면 VoD 서비스의 경우, 실시간 방송과는 다르게 다음과 같이 두 가지의 차별적인 특성을 갖는다.

첫째, VoD의 종료 시간은 알려져 있다는 것이다. VoD 서비스의 경우, 이미 미디어 서버에 비디오 콘텐츠가 저장되어 있고, 이를 사용자에게 전송하는 것이므로 해당 비디오 콘텐츠의 재생 시간은 고정된 값이다. 따라서, VoD 시청을 요청한 사용자가 중간에 임의로 시청을 중단하는 경우를 제외하면 종료 시간을 정확히 알 수 있다. 둘째, 대부분의 VoD 서비스 이용자들은 비디오 콘텐츠의 첫 부분을 시청 한 후, 해당 비디오의 시청 여부를 결정하는 경향이 있다. 따라서, 많은 경우에 VoD 서비스는 초반에 종료되는 경우가 많다.



(그림 2) 트래픽 특성에 따른 VoD 서버 부하 비교

그림 1은 VoD 서비스에서 비디오 콘텐츠의 총 재생 시간과 사용자가 시청을 중단하기 전까지 실제 재생된 시간의 비율에 따른 누적 분포 확률 분포(Cumulative Distribution Function)를 나타낸 그래프이다[8]. 그래프에 따르면 비디오 콘텐츠의 총 길이 중 15% 이내만 재생되고 시청이 종료되는 확률이 약 50% 정도 이고, 90% 이상 재생될 확률은 약 20% 미만으로 나타난다. 이를 통해 VoD시청 초반에 사용자가 시청을 종료할 확률이 매우 크며, 시청 시간이 길어질수록 시청을 종료할 확률이 점차 낮아짐을 알 수 있다. 이는 VoD 서비스를 이용하는 사용자들이 비디오 콘텐츠의 앞부분을 시청 한 후, 흥미를 느끼면 계속 시청하고 그렇지 않으면 비디오 시청을 종료하기 때문이다. 그림 1에서 그래프의 마지막의 급격한 기울기는 비디오 콘텐츠가 종료된 후, 엔딩 크레딧이 재생될 때 사용자들이 시청을 종료하기 때문이다. 이러한 사용자의 이용 행태는 VoD 서버 부하에 큰 영향을 미친다.

그림 2는 비디오 트래픽의 시작 시간과 종료 시간에 VoD 서버 부하를 비교하고 있다

위의 그림을 통해서 두 가지 트래픽 모두 같은 요구 대역폭을 가지고 있지만, 시작 시간과 종료

시간에 따라 잠재적으로 서버에 미치는 영향은 다름을 알 수 있다.

그림 2의 (1)은 트래픽의 시작 시간은 같지만 종료 시간이 틀린 경우이다. 이 같은 경우는 현재 시점에서 두 트래픽 모두 1Mbps의 대역폭을 점유하고 있지만, 비디오 A는 20분 후에 비디오 B는 10분 후에 종료될 예정이다. 따라서, 잠재적으로 트래픽 A가 트래픽 B보다 미디어 서버에 더 많은 부하를 주고 있음을 알 수 있다.

그림 2의 (2)는 트래픽의 종료 시간은 같지만 시작 시간이 다른 경우이다. 이때는 비디오 종료 시간이 같기 때문에 트래픽 A나 트래픽 B 모두 서버에 같은 부하를 주고 있다고 생각할 수 있다. 하지만 그림 1을 통해 재생이 시작된 지 얼마 안 되었을 수록 비디오가 종료될 확률이 높으므로, 트래픽 B가 트래픽 A 보다 더 빨리 종료될 가능성이 높다. 따라서, 이러한 측면에서 볼 때, 트래픽 A가 트래픽 B보다 잠재적으로 미디어 서버에 더 많은 부하를 주고 있다고 볼 수 있다.

본 논문에서는 이러한 사용자의 VoD 서비스 이용 행태를 고려하여, 미디어 서버의 유효 대역폭과 현재 서비스 중인 트래픽들의 잠재적인 영향을 함께 고려하고 이를 통해 서버 부하 분산을 수행하는 알고리즘을 제안한다.

4. VoD 서버 부하 분산 알고리즘

본 논문에서 대상으로 하고 있는 VoD 서비스 네트워크는 하나의 중앙 서버와 다수의 분산된 미디어 서버들이 존재하는 구조이다. 이 때 중앙 서버는 사용자로부터 VoD 서비스 요청을 받아서 사용자에게 가장 적합한 미디어 서버를 찾아 연결해 주는 역할을 수행하고, 각각의 미디어 서버에서는 사용자에게 해당 VoD 서비스를 제공한다. 본 논문에서 제안하는 VoD 서버 부하 분산 방법은 다음과 같이 크게 두 가지로 나눌 수 있다. 첫째, 미디어 서버들의 가용 대역폭에 여유가 있을 때, 각 서버들의 대역폭 기대치를 고려하여 부하를 분산한다. 둘째, 현재 미디어 서버들의 가용 대

역폭에 여유가 없는 경우, 각 미디어 서버에서의 사용자 응답 시간 기대치를 고려해서 사용자의 VoD 서비스 요청을 분산하는 방법이다.

4.1 미디어 서버의 가용 대역폭이 있는 경우

사용자의 VoD 서비스 요청을 받은 중앙 서버는 자신이 관리하는 미디어 서버들의 대역폭 가중치를 계산한 후, 대역폭 기대치(Expected Bandwidth, EBW)가 가장 큰 미디어 서버에게 사용자의 VoD 서비스 요청을 할당한다. 일반적으로 각각의 미디어 서버들은 주기적으로 자신의 상태 정보(가용 대역폭, VoD 세션 정보 등)를 중앙 서버로 보고한다. 따라서, 본 논문에서도 중앙 서버는 자신이 관리하는 모든 미디어 서버의 상태 정보를 이미 알고 있다고 가정한다. 중앙 서버는 미디어 서버의 EBW를 다음과 같이 계산한다.

$$EBW = \alpha \times ABW + (1 - \alpha) \times PBW \quad (1)$$

식(1)에서 가용 대역폭(Available Bandwidth, ABW)이란, 사용자의 VoD 요청이 들어온 시점에서 현재 미디어 서버에서 실제로 사용 가능한 부분의 대역폭을 의미한다. 또한, 잠재 대역폭(Potential Bandwidth, PBW)은 확률적인 값으로, 현재 미디어 서버가 서비스 중인 VoD 세션들의 종료 확률을 고려하여 각 세션들이 종료될 때 얻을 수 있는 예상 대역폭의 합을 의미한다. α 는 가중 계수로 0과 1 사이의 값을 갖는다. α 가 1이면 현재 가용 대역폭만을 고려하여 부하 분산 알고리즘을 수행하고, α 가 0이면 잠재 대역폭만을 고려하게 된다. 본 논문에서는 현재의 가용 대역폭이 작아질수록 미래의 잠재 대역폭보다 현재의 가용 대역폭을 더 많이 반영하도록 α 를 다음과 같이 정의한다.

$$\alpha = 1 - (ABW / \text{미디어 서버 총 대역폭})$$

잠재 대역폭 (PBW)은 다음과 같이 계산한다.

$$PBW = \sum_{i=1}^n \left\{ f\left(\frac{t_i}{T_i} \times 100\right) \times BW_i \right\} \quad (2)$$

식(2)에서 n 은 현재 미디어 서버가 서비스 중인 VoD 세션의 개수, 함수 f 는 VoD 세션 종료 시간 분포의 확률 밀도 함수를 나타낸다. t_i 는 VoD 세션 i 의 현재까지 재생된 시간, T_i 는 VoD 세션 i 의 총 길이, 그리고 BW_i 는 VoD 세션 i 를 서비스 하기 위해 요구되는 대역폭을 의미한다. VoD 세션 종료 시간의 분포를 나타내는 함수 f 를 통해 각각의 VoD 세션의 종료 확률을 예측할 수 있고, 여기에 BW_i 를 반영하여 해당 VoD 세션이 종료될 경우 얻게 되는 잠재 대역폭을 고려한다.

위와 같이, 중앙 서버는 자신이 관리하는 모든 미디어 서버들의 EBW를 구하고, 이것이 가장 큰 미디어 서버에게 사용자의 VoD 요청을 할당함으로써 VoD 서비스 부하를 분산하게 된다.

4.2 미디어 서버의 가용 대역폭이 없는 경우

중앙 서버는 자신이 관리하는 미디어 서버들의 가용 대역폭이 없는 경우, 미디어 서버에서의 예상 대기 시간이 가장 작은 서버에 사용자 요청을 할당한다. 다시 말해, 사용자의 VoD 요청을 서비스 해주기 위한 대역폭이 확보되기까지 사용자가 기다려야 하는 대기 시간이 가장 짧게 예상되는 미디어 서버에 사용자 요청을 할당하는 것이다. 본 논문에서는 이와 같이 미디어 서버에서 VoD 요청이 서비스 되기까지 사용자가 기다려야 하는 시간을 사용자 응답 시간 기대치(Expected Response Time, ERT) 라고 정의한다.

ERT를 계산하기 위해 중앙 서버는 우선 현재 미디어 서버에서 서비스 중인 VoD 세션들의 재생 시간 기대치(Expected Playing Time, EPT)를 계산한다. EPT란 각 미디어 서버에서 현재 서비스 제공 중인 VoD 세션이 종료될 때까지 남은 재생 시간에 대한 기대치이다. VoD 세션 i 의 EPT는 사용자의 VoD 세션 종료 시간 분포의 확률 밀도 함수를 이용하여 다음과 같이 계산한다.

(표 1) 알고리즘 예제

서비스		대기열	
VoD 요청 ID	EPT	VoD 요청 ID	EPT
S1	6	Q1	2
S2	12	Q2	4
S3	7	Q3	10
S4	14	Q4	5
S5	9		

$$EPT_i = \left\{ \int_{\frac{t_i}{T_i} \times 100}^{100} x f(x | x > \frac{t_i}{T_i} \times 100) dx \right\} \times \frac{T_i}{100} - t_i \quad (3)$$

식(3)에서 함수 f 가 VoD의 총 길이를 100으로 정규화하여 구해진 함수이므로 f 를 이용해서 구한 세션의 평균 재생 시간은 VoD의 총 길이가 100으로 정규화되어 있다. 이를 실제 평균 재생 시간으로 변환 한 후($T_i/100$), 현재까지 재생한 시간(t_i)을 참가하여, VoD 세션 i 의 남은 EPT를 구할 수 있다. 이 때, 해당 미디어 서버에서 아직 재생되지 않고 대기열에서 대기하고 있는 세션의 경우 t_i 를 0으로 설정하여 EPT를 계산한다.

또한, 중앙 서버는 ERT를 구하기 위해 미디어 서버의 대기열에서 대기 중인 VoD 요청들에 대해서 대기 시간 기대치(Expected Waiting Time, EWT)를 계산한다. 현재 대기열에 n 개의 서비스 요청이 대기 중인 미디어 서버에서, 대기열의 맨 처음에 위치한 VoD 요청의 EWT는 현재 미디어 서버가 서비스 중인 VoD 세션들의 EPT의 최소값이 된다. 대기열의 두 번째부터 n 번째까지 위치한 $n-1$ 개 VoD 요청들의 EWT는 현재 서비스 중인 VoD 세션들의 EPT와 자신보다 앞서 대기열에서 대기 중인 VoD 요청들의 EPT, EWT의 합을 비교하여 구한다. 결과적으로 해당 미디어 서버의 ERT는 $n+1$ 번째 VoD 요청의 EWT로 계산 된다. 좀 더 구체적인 설명을 위해서, 표 1을 통해 EWT를 구하는 방법을 예시한다.

그림 3에서 현재 미디어 서버에는 5개의 VoD 세션 (S_1, S_2, S_3, S_4, S_5)이 서비스 중이고, 4개의 VoD 요청 (Q_1, Q_2, Q_3, Q_4)이 대기 중인 상황이다.

또한, 위에서 설명된 식(3)을 통해 구한 각 VoD 요청들의 EPT가 나타나 있다. 대기열의 가장 처음에 위치하고 있는 Q_1 은 현재 서비스 중인 VoD 세션들 중에서 어느 하나라도 종료 되면 바로 서비스를 받을 수 있다. 따라서, EPT가 가장 작은 S_1 의 EPT 값을 EWT로 갖는다. Q_1 은 ($EWT_{Q_1}=6$) 대기열의 두 번째 위치하고 있는 Q_2 의 EWT를 구하기 위해서는 현재 서비스 중인 세션 중 EPT가 두 번째로 작은 S_3 와 자신보다 대기열 앞에 위치한 Q_1 을 비교한다. Q_1 의 EPT와 EWT의 합이 S_3 의 EPT보다 작을 경우, Q_1 이 종료됨에 따라 Q_2 가 서비스를 받을 수 있기 때문이다. 그림 3에서는 Q_1 의 EPT와 EWT의 합이 8이고 S_3 의 EPT가 7이므로, Q_2 의 EWT는 7이 된다($EWT_{Q_2}=7$).

마찬가지로, 대기열에서 세 번째 위치하고 있는 Q_3 의 EWT를 구하기 위해서는 현재 서비스 중인 요청 중 EPT가 세 번째로 작은 S_5 와 Q_1 , Q_2 를 비교해야 한다. S_5 의 EPT는 9고, Q_1 , Q_2 의 EPT와 EWT 합이 각각 8(2+6)과 11(4+7)이므로 Q_3 의 EWT는 8로 계산된다. 즉, Q_3 은 현재 서비스 중인 S_5 가 종료된 후 서비스가 되는 것이 아니라, 위에서 계산한대로 S_1 이 종료된 후 Q_1 이 서비스가 되고 그 이후에 바로 서비스가 될 수 있는 것이다($EWT_{Q_3}=8$).

위에서 Q_3 은 Q_1 이 종료된 후에 서비스가 되기 때문에, 대기열의 마지막에 위치한 Q_4 의 EWT를 구하기 위해서는 아직 서비스 중인 S_5 와 Q_2 , Q_3 를 비교해야 한다. S_5 의 EPT는 9고 Q_2 , Q_3 의 EPT와 EWT 합이 각각 11(4+7)과 18(10+8) 이므로 Q_4 의 EWT는 S_5 의 EPT와 동일한 9가 된다. ($EWT_{Q_4}=9$)

위와 같이 현재 대기열에서 대기 중인 VoD 요청들에 대한 EWT를 통해, 해당 미디어 서버의 ERT는 새로운 VoD 요청이 들어온 경우로 가정하여 구할 수 있다. 다시 말해, 새로운 VoD 요청 Q_5 의 EWT는 S_2 와 Q_2 , Q_3 , Q_4 를 비교해야 하며, Q_2 의 EPT와 EWT의 합이 최소이므로 ERT는 11이 된다.

위의 예제는 하나의 미디어 서버의 ERT를 구하는 방식이다. 이와 같은 방법으로 중앙 서버는 모든 미디어 서버들에 대해서 ERT를 구하고, 그 중에서 가장 작은 ERT를 갖는 미디어 서버에 새

로운 VoD 요청을 할당한다.

5. 성능 평가

본 논문에서는 네트워크 시뮬레이션을 통해 제안한 알고리즘의 성능 평가를 수행하였다. 5.1에서는 논문에서 사용한 네트워크 시뮬레이션 환경에 대해 설명하고, 5.2에서는 성능 평가 결과에 대해 분석한다.

5.1 시뮬레이션 환경

성능 평가를 위한 시뮬레이션에서 가정된 네트워크 구조는 1200 m * 1200 m 크기의 환경에서 미디어 서버들은 격자 형태로 위치해 있는 상황이다. 사용자의 VoD 요청들은 네트워크의 랜덤한 위치에서 포아송(Poisson) 분포에 의해 생성되고 각각의 VoD는 세션 종료 시간 분포에 따라 확률적으로 일정 시간만큼 재생된 후 종료된다. 시뮬레이션 환경의 구체적인 값들은 다음과 같다.

(표 2) 시뮬레이션 환경

시뮬레이션 환경	값
네트워크 크기	1200(m) × 1200(m)
미디어 서버 대역폭	60Mbps
미디어 서버 수	9 / 16 / 25개
VoD 재생 시간	1~120분
VoD 대역폭	6Mbps
트래픽 발생량(분당)	10 / 20 / 30
대기열 크기	2, 5, “무한”
최대 전송 거리	800m
시뮬레이션 시간	3000분

본 논문에서는 미디어 서버의 개수와 트래픽 발생량 그리고 대기열의 크기에 따른 제안 알고리즘의 성능을 평가하였다. 표 2에서 트래픽 발생량은 분당 발생하는 VoD 요청의 갯수를 의미한다. 그리고, 대기열의 크기를 넘어서는 VoD 요청은 차단되며, 대기열의 크기가 ‘무한’인 것은 모든

VoD 요청에 대해 차단 없이 수용함을 의미한다.

5.2 실험 결과

본 논문에서는 제안한 알고리즘의 성능 평가 대상으로 다음과 같이 Nearest 알고리즘과 Least load 알고리즘을 비교한다.

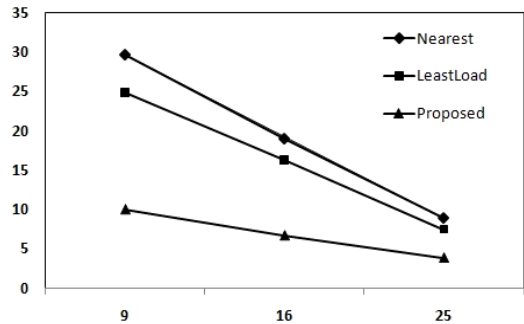
5.2.1 Nearest 알고리즘

VoD 요청을 사용자로부터 가장 가까운 거리에 위치한 미디어 서버에 할당하는 방식이다. 최대 전송 거리 반경 내 미디어 서버들 중에서 가용 대역폭에 여유가 있는 서버들 중, 사용자로부터 가장 가까운 거리에 위치한 미디어 서버에 VoD 요청을 할당한다. 만일 모든 미디어 서버들 중에서 가용 대역폭에 여유가 있는 서버가 존재하지 않는 경우에는 거리상 가장 가까운 미디어 서버에 할당한다.

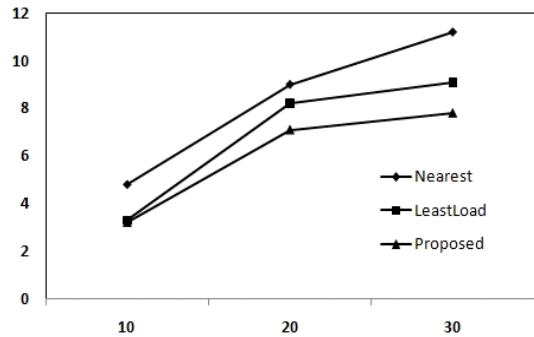
5.2.2 Least load 알고리즘

미디어 서버들 중에서 부하가 가장 작은 서버에 VoD 요청을 할당하는 방식이다. 최대 전송 거리 반경 내 가용 대역폭에 여유가 있는 미디어 서버들 중, 사용 대역폭 대비 서버의 총 대역폭 비율이 가장 작은 서버에 VoD 요청을 할당한다. 만일 가용 대역폭에 여유가 있는 서버가 존재하지 않는 경우에는 대기열에 있는 VoD 요청들의 총 재생 시간의 합이 가장 작은 미디어 서버에 요청을 할당하는 방식이다.

그림 3은 VoD 요청의 발생량을 분당 10개, 미디어서버의 대기열의 크기는 5로 고정한 환경에서, 미디어 서버 수를 9, 16, 25개로 변경할 때의 ERT 결과를 보여준다. 세 방법 모두 미디어 서버가 증가함에 따라 네트워크 용량이 증가하기 때문에 전체적인 ERT가 줄어든다. 본 논문에서 제안한 알고리즘이 가장 짧은 ERT를 보여주고 있으며, Least load 방식이 Nearest 방식보다 짧은 ERT를 갖는다. 이는 거리상 가까운 위치에 있는 미디어



(그림 3) 미디어 서버 수(Y축)에 따른 평균 대기 시간

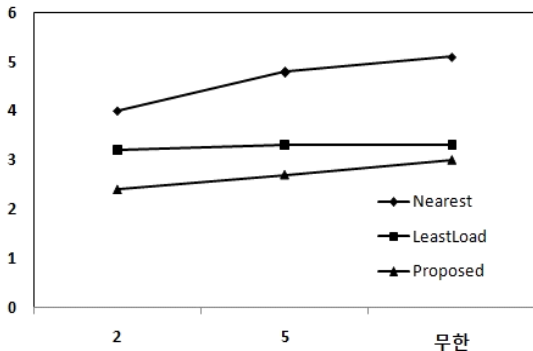


(그림 4) 발생 트래픽 양(Y축)에 따른 평균 대기 시간

서버보다 부하가 적은 미디어 서버에 VoD 요청을 할당할 때 응답이 빠르다는 것을 보여주는 것으로, 지리적 위치보다는 서버의 부하가 ERT에 더 큰 영향을 미친다는 것을 알 수 있다.

그림 4는 미디어 서버 수를 25개, 각 미디어 서버의 대기열 크기는 5로 고정한 환경에서, VoD 요청의 발생량을 분당 10, 20, 30으로 변화시키면서 ERT의 변화를 보여준다. 일반적으로 트래픽 발생량이 증가할수록 ERT도 비례하여 증가하지만 트래픽 발생량이 20에서 30으로 증가할 때는 대기열의 크기를 넘은 VoD 요청이 차단되기 때문에 ERT 증가가 정체되는 결과를 보여준다. 역시 제안한 알고리즘이 가장 좋은 성능을 보여주는데, 특히 트래픽 발생량이 클수록 ERT 차이가 늘어난다.

그림 5는 미디어 서버 수를 25개, 트래픽 발생량은 분당 10개로 고정한 환경에서, 대기열의 크



(그림 5) 대기열의 크기 (Y축)에 따른 평균 대기 시간

기를 2, 5, ‘무한으로 변화시키면서 ERT의 변화량을 관찰한 그래프이다. 본 논문에서 제안한 알고리즘과 Least load 같은 경우는 대기열의 크기가 커져도 ERT가 비슷한 수준의 값을 유지하고 있지만 Nearest 같은 경우는 ERT가 대기열의 크기에 따라서 증가한다. 이러한 결과는 제안한 알고리즘과 Least load 같은 경우는 미디어 서버들에 VoD 요청이 부하를 고려하여 골고루 분산되기 때문에 대기열의 크기 제한으로 인해 차단되는 요청이 많지 않지만, Nearest는 일부 서버에만 VoD 요청이 집중되어 대기열의 크기가 커지게 될 경우, 대기열의 뒤에 존재하게 되는 요청들이 많아지게 되고 이로 인해서 ERT가 지속적으로 증가하는 것으로 볼 수 있다.

6. 결론

본 논문에서는 IPTV 서비스 환경에서 네트워크 자원을 효율적으로 이용하고 서비스 품질을 향상 시키기 위해, 사용자 이용 행태에 기반한 VoD 서버 부하 분산 알고리즘을 제안한다.

기존의 서버 부하 분산 방법들은 현재 서버 부하와 트래픽 상태만을 고려한 반면, 본 논문에서 제안한 방법은 VoD 서비스의 특성과 이를 이용하는 사용자의 이용 행태를 고려하여 향후 잠재적인 서버 부하를 예측하고 이를 통해 서버의 부하를 효율적으로 분산한다. 이를 위해 사용자의 VoD 이용 종료 시간을 확률적으로 예측하고, 이

를 통해 대역폭 기대치(EBW)와 사용자 응답 시간 기대치(ERT)를 계산하여 네트워크 자원을 효율적으로 사용하면서 사용자 대기 시간을 최소화 한다. 제안한 알고리즘의 성능 평가는 미디어 서버 수, 발생 트래픽 양, 미디어 서버의 대기열 크기의 변화에 따른 성능 결과를 검증하기 위해 네트워크 시뮬레이션을 수행하였다. 시뮬레이션 결과, 제안한 알고리즘이 기존의 Nearest 알고리즘과 Least Load 알고리즘보다 모든 상황에서 훨씬 좋은 사용자 대기 시간을 보여준다.

본 논문에서 제안한 VoD 서버 부하 분산 알고리즘은 IPTV 환경처럼 가입자의 수가 방대하고 다양한 콘텐츠들이 존재하는 네트워크에서 한정된 자원을 보다 효율적으로 관리하고 서비스 품질을 향상하기 위한 서버 부하 분산 제어 방법으로 적용 될 수 있다.

참 고 문 헌

- [1] 한국 인터넷 진흥원, “Trend Report”, 2010년 7월
- [2] 박홍식, “IPTV 서비스 품질 제공기술”, *SK Telecommunication Reviews*, 19권 2호, 2009, pp.222-233.
- [3] M. Cha, P. Rodriguez, S. Moon, and J. Crowcroft, “On Next-Generation Telco-Managed P2P TV Architecture”, *Intl. Workshop on Peer-to-Peer Systems (IPTPS)*, Feb. 2008
- [4] M. Cha, P. Rodriguez, J. Crowcroft, S. Moon, and X. Amatrianin, “Watching Television over an IP Network”, *ACM SIGCOMM IMC 2008*, Oct. 2008.
- [5] Y.F. Huang and C.C. Fang, “Load Balancing for Clusters of VoD Servers”, *Information Science*, 2004, pp.113-138.
- [6] M. Dakshayini, H. Guruprasad, H. Maheshappa and A.S. Manjunath, “Load Balancing in Distributed VoD Using Local Proxy Server Group”, *Intl. Conf. on Computational Intelligence*

- and Multimedia Applications 2007*, Dec. 2007, Vol. 4, pp.162-168.
- [7] D. Sujatha, K. Girish, B. Rashmi, K. Venugopal, and L. Patnaik, "Load Balancing in Fault Tolerant Video Server", *LNCS 2007*, pp.306-315.
- [8] E. Kusmierek, M. Czymek, C. Mazurek, and M. Stroinski, "iTVP: Large-scale Content Distribution for Live and On-demand Video Services", *Multimedia Computing and Networking*, *SPIE-IS&T Electronic Imaging*, Jan. 2007, Vol. 6504, SPIE 2007.
- [9] H. Yu, D. Zheng, B. Zhao and W. Zheng, "Understanding User Behavior in Large-Scale Video-on-Demand Systems", in *Proc. of Eurosys 2006*, Apr. 2006.
- [10] 권춘자, 최황규, "동적 버퍼 분할을 이용한 클러스터 VoD 서버의 효율적 부하 분산 방법", *정보처리학회논문지C*, *Vo. 9, no. 5, pp.709-718*, 2002.

● 저 자 소 개 ●



이 강 용 (Kang Yong Lee)

2004년 아주대학교 정보 및 컴퓨터 공학과 졸업 (학사)
2006년 University of Science & Technology 광대역네트워크공학과 졸업 (석사)
2006년~현재 한국전자통신연구원 방송통신융합연구부문 연구원
관심분야 : 네트워크 최적화, 통신망 성능 분석, 시뮬레이션
E-mail : kanglee@etri.re.kr



장 욱 (Uk Jang)

2007년 연세대학교 전기전자 공학부 졸업 (학사)
2009년 연세대학교 전기전자 공학부 졸업 (석사)
2009년~현재 한국전자통신연구원 방송통신융합연구부문 연구원
관심분야 : Optimization, MIMO, Wireless, Multimedia
E-mail : ukjang@etri.re.kr



조 기 성 (Kee-Seong Cho)

1982년 경북대학교 전자공학과 졸업 (학사)
1984년 경북대학교 전자공학과 졸업 (석사)
1984년~현재 한국전자통신연구원 방송통신융합연구부문 스마트서비스제어연구팀 팀장
관심분야 : 서비스 제어 기술, 방송 통신 융합 서비스, IPTV, BcN
E-mail : chokis@etri.re.kr