

# 실시간 스트리밍을 위한 대역폭 적응에 관한 연구\*

정재훈, 양제현  
한국과학기술원 전산학과

## On the Bandwidth Adaptation for Realtime Streaming

Jae-Hun Jung, Jae-Heon Yang  
Department of Computer Science,  
Korea Advanced Institute of Science and Technology

### 요 약

본 논문은 폭주적인(bursty) 트래픽을 생성하는 스트리밍에서 사용자가 요구하는 최소 품질(Quality of Service)의 보장 및 공정한 대역폭 활용을 위하여 개선된 대역폭 적응 기법을 제안한다. LAN상의 서버/클라이언트 유니캐스트 환경에서 수행된 시뮬레이션은 제안된 기법에 의한 전송율의 패턴이 네트워크의 일시적인 혼잡에 의한 영향이 적고 안정하며 대역폭을 공유하는 기타 커넥션들에게 공정하도록 대역폭을 활용함을 보여준다.

제안한 대역폭 적응 기법에서 사용한 윈도우 기반 전송을 추정된 윈도우 크기에 따라 스트리밍 어플리케이션의 graceful degradation과 대역폭 활용의 trade-off를 갖고 있지만 세션 중 윈도우 크기를 상황에 따라 동적으로 변화시킴으로써 요구사항을 충족하도록 응용될 수 있다.

## 1. 서론

스트리밍은 영상, 음성과 같은 연속성과 실시간성을 지닌 미디어 스트림을 전송 및 수신하여 재생하는 기술로 VOD(Video On Demand), 영상회의, 멀티미디어 원격 프리젠테이션 등 네트워크 기반의 멀티미디어 서비스를 위한 필수 기술 중의 하나이다. 스트리밍에서 제공하는 스트림은 전송하는 미디어의 형태에 따라 저장된 스트림(stored stream)과 활성 스트림(live stream)으로 나뉜다. VOD는 단일 송신자가 파일 서버에 저장된 이질 스트림을 다중 수신자에게 스트리밍하고 영상회의는 다중 송신자가 단일한 활성 스트림을 다중 수신자에게 스트리밍하는 경우이다.

시간적 명세(temporal specification)가 지켜지기 위해서는 네트워크 레벨에서 혼잡(congestion) 등으로 인한 패킷 손실과 지연이 최소화되어야 한다. 이를 위해 스트리밍에 필요한 대역폭을 미리 예약하는 방법이 연구되었다. 이 방법은 라우터 등 네트워크 내부의 구조적인 변경이 필요하다. 이와 달리 각 어플리케이션이 네트워크의 상태를 파악하고 전송율을 조절하여 혼잡을 제어하거나 적응하도록 하는 연구가 수행되었다.

대표적인 예로 additive increase/multiplicative decrease 알고리즘이 있다. 이는 전송율 제어를 수행하는 어플리케이션이 네트워크의 상태를 혼잡(congestion)과 비혼잡(non-congestion)으로 구분하고 혼잡시에는 대역폭이 1/GAIN에 의해 기하급수적으로 낮추어지고, 비혼잡시에는 INCR만큼 대역폭이 상향 조절되도록 하여 전체 네트워크의 성능이 급격히 낮아지는 것을 막아주는 기법으로 [7]에서 동정성과 효율성이 증명되었다. [1][2]는 위 알고리즘을 적용하여 멀티캐스트 네트워크에서 선택된 수신자들로부터 패킷손실율에 대한 피드백 정보를 받아 전체 수신자들에 대한 혼잡을 겪는 수신자 또는 혼잡을 겪지 않는 수신자의 비율이 10%를 초과하는 경우 송신자의 최대 전송율을 조절하도록 하였다. [2]에서 이 기법의 실험결과에 의하면 패킷손실율이 키질에 따라 송신자의 최대 전송율과 실제 전송율이 급격히 떨어짐을 볼 수 있다. 이는 네트워크의 일시적인(transient) 혼잡에 대해 송신자가 민감하게 전송율을 떨어뜨려 어플리케이션에서 패킷지연을 일으키게 된다.

본 논문은 VOD에서와 같이 유니캐스트로 비디오 등 실시간 스트림을 전송

할 때 다른 TCP커넥션들과 대역폭을 공정하게 나누어 사용하는 대역폭 적응 기법을 제안한다. 제안하는 기법은 혼잡을 줄여 네트워크의 혼잡 제어 및 방지에 기여한다. 2장에서는 대역폭 적응 시 전송율을 조절하는 기준으로 TCP대역폭을 이용하는 방법과 안정된 전송율을 유지하기 위한 방법을 설명하고 이들을 이용한 대역폭 적응 기법을 소개한다. 이를 토대로 수행한 실험 결과가 3장에서 기술되며 마지막으로 4장에서 결론을 맺는다.

본 연구는 여러 RTP(Real-time Transport Protocol) 스트리밍 커넥션과 여러 TCP(Transmission Control Protocol) 커넥션이 대역폭을 공유하는 네트워크를 가정한다.

## 2. 대역폭 적응 기법

### 2.1 네트워크 피드백

수신자로부터의 피드백은 실시간 미디어 스트림을 멀티캐스트하기 위하여 주로 사용되는 RTP(realtime transport protocol)의 제어 프로토콜인 RTCP의 RR(receiver report) 패킷으로부터 구할 수 있다[11]. 특히, 송신자가 RTCP의 SR(sender report) 패킷을 보낸 NTP(network time protocol) 시간(LSR - last sender report)과 수신자가 SR패킷을 받은 후 RR패킷을 전송할 때까지 걸린 시간(DLSR - delay since last sender report), 수신자가 겪는 패킷손실율 p가 본 연구를 위한 피드백 정보로 이용될 수 있다. 위 정보로부터 종단간의 rtt(round trip time)는 송신자가 RR패킷을 받은 NTP시간에서 LSR을 제한 값이 된다

### 2.2 TCP커넥션에게 공정한 대역폭의 계산

[5][9]는 전송율 기반의 혼잡제어(rate-based congestion control)를 수행하는 어플리케이션이 동일 상황에서의 가상의 TCP커넥션이 요구하는 전송율보다 크지 않도록 데이터를 전송하여 RED(random early detection)와 같은 혼잡제어를 수행하는 게이트웨이에서 패킷 손실을 줄이도록 권장하고 있다 RED 게이트웨이에서의 패킷 손실로 인하여 TCP의 혼잡된 윈도우 크기가 반으로 줄어들기 때문에 패킷 손실을 유발하는 과도한 전송율은 대역폭을 공유하는 기타

\* 본 연구는 한국통신의 연구지원을 받았다

TCP커넥션들에게 공정하지 못하다 [5]의 Appendix는 패킷의 최대전송단위 MTU와 적절한 피드백 기법을 통해 구한 현재 네트워크의 중단간 RTT와 패킷손실률  $p$ 가 주어졌을 때 (1)식을 이용하여 스트리밍이 겪고 있는 동일 환경에서 가상의 TCP커넥션이 요구하는 대역폭을 구하는 방법을 소개하고 있다<sup>1)</sup>

$$bandwidth_{TCP} = 1.22 \times \frac{MTU}{RTT \times \sqrt{p}} \quad (1)$$

### 2.3 윈도우기반 대역폭 추정

네트워크의 상태를 파악하여 데이터 전송을 조절을 위하여 [3][13]에서는 누적된(cumulative) low-pass filter를 이용한다. 데이터 전송율에 대한 누적된 low-pass filter는  $r_{estimate} = (1 - \alpha) \times r_{estimate} + \alpha \times r_{measured}$  로 표현된다. 본 연구에서는 스트리밍 어플리케이션이 데이터 패킷을 전송할 때 윈도우를 이용한 점추정기법과 CLT(central limit theorem)[4]를 이용하여 현재까지의 네트워크 상태에 따른 최대 전송율의 분포를 구하고 다음 전송을 위한 최대 전송율을 추정한다.

$(r_1, r_2, \dots, r_w)$ 이 최근  $w$  개의 측정치이고 서로 독립이며 동일 분포를 따르고 평균을  $\mu$  라고 하면,  $\hat{\mu}$  은 다음과 같이 정의될 때 평균  $\mu$  의 불편추정량이 된다[4].

$$\hat{\mu} = \frac{1}{w} \sum_{i=1}^w r_i$$

본 연구에서는 네트워크의 최근의 동적인 특성을 고려하기 위하여  $w$  개의 최근 rate 측정치에 대해 위의 불편추정량(unbiased estimator)을 구하는 식을 적용한다.<sup>2)</sup> 스트리밍 어플리케이션의 최대 전송율을 구하는 식은 (2)식과 같다.

$$r_{estimate} = \frac{1}{w} \sum_{i=1}^w r_i \quad (2)$$

### 2.4 대역폭 적용 알고리즘

스트리밍 시 송신자의 안정된 전송율을 통해 수신자에서의 자연스러운 제정이 이루어지도록 하기 위하여 네트워크의 혼잡성을 나누는 기준을 제시할 때 스트리밍 도중 전송율의 급격한 반전을 줄이고 전송율의 변화량이 적절한 범위내 바운드로되도록 하여야 한다. 본 논문에서는 네트워크의 혼잡성을 구분하기 위하여 (2)식에서 추정된 전송율과 (1)식에서 구한 가상의 TCP커넥션이 요구하는 대역폭을 비교한다. 추정된 대역폭이  $B_{TCP}$  보다 크면 대역폭을 공유하는 가상의 TCP커넥션들이 패킷을 drop하거나 손실을 입을 확률이 높아지므로 네트워크의 혼잡을 유발하고 기타 커넥션들에게 공정하지 못한 활용도를 보이게 된다 따라서 다음의 전송율이  $B_{TCP}$  이하에 머물도록 GAIN으로 나누어준다. 추정된 대역폭이  $B_{TCP}$  보다 작으면 전송율을 INC만큼 상향 조절하면서 스트리밍의 효율을 높인다.

알고리즘은 크게 전송율을 제어하는 부분과 제어된 전송율을 평활화하는 부분으로 구성된다. 전송율의 제어를 위하여 additive increase/multiplicative decrease 알고리즘을 이용하고 전송율의 평활화를 위하여 2.3절의 윈도우 기반 추정기법을 이용하였다.  $w$ 는 전송율을 추정하기 위하여 이용하는 모의집단의 수를 나타낸다. 모의집단의 수가 클수록 전송율은 더욱 안정해서 네트워크의 상태변화에 대해 반응하는 속도가 줄어들고 반면 모의집단의 수가 작을수록 빠

```

get RTT, p from feedback mechanism;
B_TCP = 1.22 * (MTU / (RTT * sqrt(p))); // by Appendix A of [5]
if B_REQ >= B_TCP
    R = max( (R_ESTIMATE / GAIN), MIN_RATE );
else
    R = min( R_ESTIMATE + INC, MAX_RATE );
if w is expanding
    R_ESTIMATE = (R_ESTIMATE * w + R) / (w + 1);
expand w;
else
    R_ESTIMATE += (R - R_OLDEST) / w;
B_REQ = R_ESTIMATE;
    
```

르게 반응하여 대역폭이 가용할 때 적절히 활용하는 결과를 보일 것이다. 따라서, graceful degradation과 대역폭 활용의 trade-off를 가지고  $w$  값을 결정할 수 있다. 더우기 세션동안 상황에 따라 동적으로  $w$  값을 변화하여 최대의 효율을 얻을 수도 있다.

스트리밍을 수행하는 송신자는 실제 전송율이  $R_{ESTIMATE}$  이하가 되도록 하여야 한다. 또한 실제 전송율의 잦은 진동(oscillation)을 방지하기 위하여 위 알고리즘의 계산 주기를 스트리밍하는 미디어의 특성에 맞게 고려하여야 한다. 예를 들어, 고대역폭 네트워크 환경에서 MPEG을 스트리밍하는 경우, 세션에 있어 의미있는 단위는 픽처단위이므로, 한 픽처를 구성하는 패킷들의 전송이 완료될 때 위 알고리즘을 수행하여 전송율을 조절할 수 있다. 또한, 동일한 경우에서, 하나의 GOP(group of pictures), 즉 픽처간의 시공간적 참조단위를 구성하는 패킷들의 전송이 모두 완료되는 경우도 가능하다. 전송하는 미디어의 특성을 고려하여 의미단위의 전송이 완료되면 참여자들은 각각 SR패킷과 RR패킷을 전송하여 패드백 정보를 교환하고 송신자는 현재의 네트워크 상태와 수신자의 상태에 맞게 전송율을 조절하도록 한다.

$B_{TCP}$  가 네트워크의 일시적인 현상에 민감하게 반응하고 급격히 저하되어 사용자가 요구하는 최소한의 품질이 제공되지 않는 현상에 대비해 알고리즘은 스트리밍이 패킷을 내보내는 전송율이  $MIN\_RATE$  에 하향 바운드로되도록 한다. 따라서, 사용자가 요구하는 최저 QoS에 해당하는 전송율을  $MIN\_RATE$  로 환산하여 설정하도록 하면 최소한의 QoS를 지켜줄 수 있다.

### 3. 실험 및 평가

실험은 LAN상의 서버/클라이언트 유닉스 환경에서 1000개의 4000~4096 바이트크기의 데이터패킷을 전송 및 수신하도록 하였고, 실험에 필요한 상수 값으로는  $MTU=4KBps$ ,  $GAIN=2$ ,  $INC=10KBps$ ,  $MAX\_RATE=100KBps$ ,  $MIN\_RATE=1KBps$ 로 하였다.<sup>3)</sup>

Figure 1은 [3][13]에서 사용한 누적형 low-pass filter와 2.3절의 윈도우를 이용한 전송율 추정기법을 각각 이용하였을 때 동일상황에서의 가상의 TCP 커넥션이 요구하는 대역폭과 비교한 결과이다 두 방법 모두 TCP대역폭의 일시적인 변화에 영향을 적게 받고 있음을 볼 수 있다. 400~700번째 제어 루프에서는 패킷손실율을 높혀 지속되는 혼잡에 대해 전송율 추정결과와 변화 형태를 살펴 보았다. 높은 패킷손실률로 인한 하강된 TCP 대역폭에 따라 전송율도 함께 따라 감을 볼 수 있는데 혼잡의 시작부분에 평활화로 인하여 graceful degradation을 보이고 네트워크의 상태가 좋아지는 시점에서 다시 적용하는 것을 볼 수 있다 Graceful degradation은 스트리밍 시 수신자에서 화질이 천천히 degrade되도록 하는 등 네트워크의 상황을 사용자에게 간접적으로 알려주게 된다 Figure 2에서는 누적형 low-pass filter의  $\alpha$  값을 변화시키면서 위 실험과 동일 상황에서의 전송율 적용형태를 비교하였다.  $\alpha$  값이 0.2, 0.5, 0.8일 때 전송율이 거의 비슷함을 알 수 있다.

3) GAIN과 INC의 값에 따라 전송되는 비트율이 틀려지기 때문에 실험 상수준 결정하는 것은 신중하게 고려되어야 한다. 또한, MIN\_RATE와 MAX\_RATE는 네트워크와 전송하는 미디어의 특성에 따라 다르게 설정될 수 있다

1) TCP는 윈도우기반 혼잡제어를 수행한다[12]. TCP는 혼잡을 감지하면 혼잡원도우의 크기(W)를 반으로 줄이고 혼잡이 없으면 W의 1배 증가시킨다 따라서, 혼잡 발생후 W가 원래 크기보다 돌아올때까지 처리할 수 있는 최소의 패킷 개수는  $W/2 + (W/2 + 1) + \dots + W$ 개가 된다. TCP는 평균 0.75WB/R의 대역폭을 요구하는데 W의 근사치와 R(rtt)을 구할 수 있으므로 혼잡발생 후부터 혼잡원도우가 원래크기보다 돌아올때까지 혼잡이 발생하지 않도록 하는 대역폭을 구할 수 있다 자세한 유도과정은 [5]의 Appendix. A에 참고하기 바란다.

2) 최근  $w$ 개의 rate 측정치들은 선제로 독립이 아니다. 따라서, (2)식을 이용하여 구한 전송율은 최근 측정치에 대해 biased된 값이 될 것이다 하지만 본 연구에서의 주목적은 불편추정량을 구하는 식을 이용, rate 측정치들을 평활화하여 정확한 추정량보다는 안정된  $r_{estimate}$ 를 얻고자하는 것이므로 독립성 여부에 대한 논란은 본 연구에서 중요하지 않다

Figure 3은 역시 동일 상황에서 윈도우 기반 전송율 추정기법을 이용할 때  $w$ 값의 크기에 따른 전송율의 변화를 비교하였는데 그 중  $w$ 값이 30, 60, 90개 일때의 전송율을 나타낸다. 세선의 시작부분과 네트워크의 상태가 다시 좋아지기 시작하는 부분에서  $w$ 값이 작을수록 네트워크의 상태에 보다 민감하게 반응하여 전송율이 더 높지만 변화가 더욱 심한 것을 볼 수 있다. 오히려  $w$ 값이 클수록 더 많은 데이터로 추정하므로 전송율은 더 낮지만 혼잡이 계속될 시 더욱 graceful한 degradation을 보이고 있음을 볼 수 있다

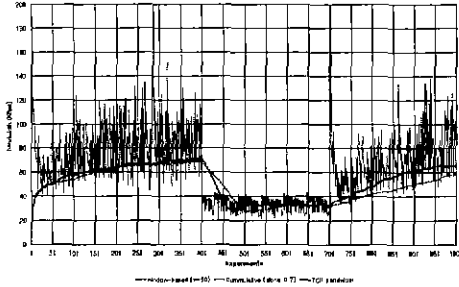


Figure 1 Congestion avoidance using TCP bandwidth

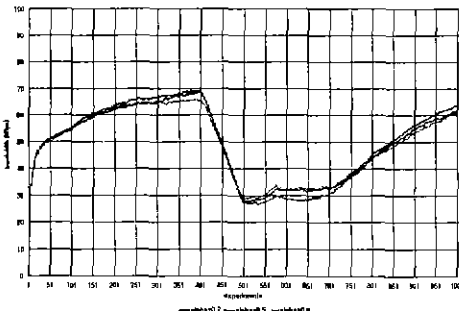


Figure 2 Bandwidth using cumulative low-pass filter

따라서, 우리는 스트리밍시 대역폭 활용과 graceful degradation의 trade-off라는 관점에서  $w$ 값을 선택할 수 있다 즉, 네트워크의 대역폭이 쉽게 가용되는 환경에서는  $w$ 값을 작게 하여 대역폭 낭비를 줄이고, graceful degradation이 중요한 어플리케이션에서는  $w$ 값을 크게 할 수 있다. 그리고,  $w$ 값이 세션동안 네트워크의 상황에 따라 동적으로 변화되도록 할 수 있다 즉, 사용가능한 대역폭이 제공되면  $w$ 값을 작게 하여 가용한 대역폭을 활용하도록 하고 사용자가 graceful degradation을 원할 시  $w$ 값을 크게 하는 등 어플리케이션의 특성에 따라  $w$ 값을 동적으로 조절되도록 하여 대역폭활용과 graceful degradation이라는 두 요구사항을 충족시킬 수 있다.

4. 결론

본 논문에서는 스트리밍시 기존의 알고리즘을 이용하였을 때 안정하지 못한 전송율로 인해 발생하는 수신자측에서의 부자연스러운 재생문제와 대역폭 활용에서의 비공정성 등을 보완하기 위해 개선된 대역폭 적응 기법을 제안하였다 실험을 통해 제안한 알고리즘이 적어도 스트리밍과 동일한 상황에서 가상의 TCP커넥션이 요구하는 대역폭에 가깝고 전송율이 안정된 패턴으로 유지되는 것을 확인하였다 또한 스트리밍 어플리케이션의 특성에 따라 윈도우의 크기를 동적으로 조절하도록 하여 graceful degradation과 더불어 효율적으로 대역폭을 활용할 수 있음을 확인하였다

본 연구에서는 유니캐스트 환경에서의 스트리밍을 고려하였다. 스트리밍은 다중수신자에게 동일스트림으로 제공되는 영상회의와 같은 어플리케이션에서

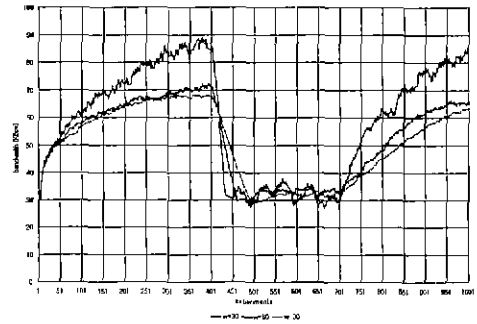


Figure 3 Bandwidth using window-based rate estimation

멀티캐스트를 하는 것이 유니캐스트할 때 보다 대역폭 활용면에서 효율적이다. [2][8] 등은 멀티캐스트 환경에서의 혼잡제어에 관한 연구결과를 제시하고 있다. [2]는 멀티캐스트 트리를 구성하는 수신자들이 혼잡한 링크에 연결된 10%의 수신자들에게 의해 저하된 QoS의 스트림을 서비스 받게 되므로 사용자의 입장에서 부적절하다. [8]은 수신자가 네트워크의 혼잡에 따라 송신자가 멀티캐스트하는 계층화된 스트림을 받거나 받지 않음으로써 네트워크의 혼잡에 적용하는 방법을 제안하고 있다. 그러나 대역폭을 공유하는 다른 커넥션들과의 공정성에 관한 문제는 다루지 않고 있다. 앞으로 멀티캐스트 네트워크 환경에서의 공정한 대역폭 활용도와 안정된 전송율의 제공을 위하여 본 논문에서 제안한 대역폭 적응 기법을 멀티캐스트 환경으로 확장하는 연구를 수행할 계획이다

참조 문헌

- [1] J-C. Bolot, T. Turletti and I. Wakeman. "Scalable feedback control for multicast video distribution in the Internet", SIGCOMM94, pp. 58-67, September 1994.
- [2] J-C. Bolot and T. Turletti. "Experience with control mechanisms for packet video in the Internet", ACM, Computer communication review, January 1998
- [3] S. Cen, C. Pu, R. Staehli, C. Cowan and J. Walpole, "A distributed real-time MPEG video audio player".
- [4] J. L. Devore. "Probability and statistics for engineering and the sciences (4th edition)", Chap.5-6, 1995.
- [5] S. Floyd and K. Fall. "Router mechanisms to support end-to-end congestion control", February 1997.
- [6] J. F. Gibbon and T. D. C. Little "The use of network delay estimation for multimedia data retrieval", Journal of selected areas in communications, V.14, N.7, September 1996.
- [7] D-M. Chiu and R. Jain. "Analysis of the increase/decrease algorithms for congestion avoidance in computer networks", Computer networks and ISDN systems, V 17, pp.1-14, January 1989.
- [8] S. McCanne and V. Jacobson. "Receiver-driven layered multicast", SIGCOMM96, August 1996.
- [9] J. Mahdavi and S. Floyd. "TCP-friendly unicast rate-based flow control", submitted to the end2end interest mailing list, January 1997.
- [10] L. L. Peterson and B. S. Davie "Computer networks: A systems approach", Chap 8, pp.393-424, 1996
- [11] H. Schulzrinne, S. Casner, R. Frederck and V. Jacobson. "RTT: A transport protocol for real-time applications", RFC1889, January 1996
- [12] V. Jacobson and M. J. Karels. "Congestion avoidance and control", SIGCOMM88, August 1988
- [13] 고통환, 나승근, 안중석 "화상회의 시스템에서 RTCP를 이용한 네트워크 대역폭 예측", 한국정보과학회 '97 가을학술발표논문집(III), 제 24권 2호, 1997.
- [14] E. Arnar, S. McCanne and R. Katz. "Receiver-driven bandwidth adaptation for light-weight sessions", ACM Multimedia, November 1997