Proceedings of **ITC-CSCC 2000**, Pusan, Korea

A Dynamic Bandwidth Allocation and Call Admission Control Method for Quality of Service Control of VBR Video Traffic

Sang-Jo Yoo, Seong-Dae Kim
Division of Electrical Engineering, KAIST
373-1 Kusong-dong, Yusong-gu, Taejon 305-701, Korea
Tel: +82-42-869-5430, Fax: +82-42-869-8570
E-mail: sangjo@sdvision.kaist.ac.kr

Abstract: In this paper, we propose a new dynamic bandwidth allocation and call admission control method for the VBR video sources with QoS constraints to provide user's quality of service requirements and at the same time to achieve an efficient resource management in networks. The proposed mechanism dynamically adjusts the necessary bandwidth by the networks based on the provided quality of service satisfaction degree of each connection in respect to the user's requirements in terms of loss ratio and average delay. Simulation results show that our proposed dynamic method is able to provide the desired level of quality of service and high utilization.

1. Introduction

There have been a lot of researches in the literature on the resource allocation and call admission control for VBR video services in ATM networks. Most of the studies have addressed modeling of video sources with a few traffic parameters to characterize source traffic variations and to decide the necessary bandwidth and buffer size for the user's QoS requirements. However, these approaches have some fundamental problems: such as that the source traffic characteristics may not be known ahead of time and we may not achieve acceptable utilization due to the worst case considerations.

For real time interactive video services without any prior traffic descriptors, a few dynamic bandwidth allocation methods [1][2][3][4] using prediction techniques have been proposed. However, most of previous proposals are mainly interested in the accurate traffic prediction and their queueing behaviors. To date no dynamic method has the ability to tightly control the QoS for such complex traffic as transmission of compressed VBR video. Therefore, we need a new mechanism that not only dynamically allocates bandwidth to improve the network utilization, but also meets the user's desired QoS.

In this paper, we have distinguished a soft QoS control from a strict QoS control. In the strict QoS control, networks guarantee the user's QoS requirements strictly during the entire connection period if input traffic conforms to the declared traffic descriptions at the call set-up time. While in the soft QoS control, which is newly defined, networks can maintain the provided QoS to the desired QoS level very closely without prior traffic descriptions. However in the worst case, for short time intervals, networks may not provide the user's QoS requirements. For the soft QoS control, we propose a new dynamic bandwidth allocation and call admission control method for VBR video sources.

2. Dynamic Bandwidth Allocation Architecture

2.1 General

Against the conventional prediction-based dynamic approaches that allocate bandwidth only with the next predicted traffic and current queue size, in addition to them, our method allocates necessary bandwidth also using a real-time quality of service measurement to each connection. To compute the necessary bandwidth for each connection, in our proposed architecture we use the following three connection-based functional blocks.

- Traffic prediction functional block — using the current and previous input cell rates (in this paper, we only consider the fixed size packet like a cell), estimates the next input rate during a prediction period.
- Quality measurement functional block — measures the provided quality of service degree by the network to each connection.
- Dynamic request functional block — based on the results on the traffic prediction block and quality measurement block, computes and requests the necessary bandwidth for the next period.

One of the advantages of the prediction based dynamic bandwidth allocation methods is that it is possible to allocate the remained bandwidth of the trunk link capacity to other service classes after reserving the next necessary bandwidth for VBR video connections. Therefore, we can achieve high bandwidth utilization.

2.2. Traffic Prediction for VBR Video

For VBR video sources, some prediction methods including neural network, least mean square (LMS) linear filter, recursive least square filter, and adaptive wavelet have been proposed. Among them, the adaptive time-domain prediction using LMS algorithm is of particular interest due to its simplicity and relatively good prediction performance. In this paper, we employ the LMS adaptive filter to predict a VBR video traffic at regular interval – filter order p=5, prediction step k=1. However, it should be noted that our proposed dynamic bandwidth allocation method, which is explained in Section 3, is not limited by a specific prediction method.

3. Proposed Dynamic Resource Allocation and QoS Control Method

3.1 QoS Satisfaction Degree

In this paper, we use two quality of service parameters, average cell transfer delay (ACTD) and cell loss ratio (CLR). The cells that cause overflow of the connection buffer are dropped in a switch and considered lost cells. In the switch of the networks, to each connection the provided quality of service degree up to now is measured and used for the next frame bandwidth allocation. We define the following satisfaction degree indexes.

- **Loss satisfaction degree**: $ls_i(n) = \dfrac{CLR_i^{QoS}}{CLR_i^{Meas}(n)}$
- **Delay satisfaction degree**: $ds_i(n) = \dfrac{ACTD_i^{QoS}}{ACTD_i^{Meas}(n)}$

where, CLR_i^{QoS} and $ACTD_i^{QoS}$ are the QoS requirements of connection i; $CLR_i^{Meas}(n)$ and $ACTD_i^{Meas}(n)$ are the provided QoS by a switch up to the n-th frame time.

We newly define an index, $SD_i(n)$ — "*satisfaction degree of quality of service*", to express the overall QoS degree provided to connection i as in (1). It is simply determined by the minimum value of the loss satisfaction degree and the delay satisfaction degree. If $SD_i(n)$ is less than 1, then it means that a switch has not provided the desired quality of service to the user. Contrary, if $SD_i(n)$ is greater than 1, then it implies that a switch has provided a superfluous quality of service due to the over allocation of bandwidth.

$$SD_i(n) = \min\{ls_i(n), ds_i(n)\} \tag{1}$$

How to measure the defined satisfaction degree indexes is an important issue. In this paper, we present a simple measurement method for the defined indexes. Among the two satisfaction indexes, the loss satisfaction degree, $ls_i(n)$, can be easily computed as in (2).

$$CLR_i^{Meas}(n) = \frac{\sum_{j=1}^{n} lost(n)}{\sum_{j=1}^{n} x(n)} \tag{2}$$

where, $lost(n)$ is the number of lost cells from $x(n)$ input cells of the n-th frame.

Computing the difference between the arrival time and departure time for each input cell, and then averaging the differences do an ideal measurement of the average cell delay. However, this approach needs a complex implementation and even real time measurement may be impossible. Therefore, we take advantage of a simple theory for the average delay calculation. The queueing system in our study can be approximated as a batch arrival system —$D^X/G/1$. The average delay on this system is given by (3) with "*Little's Law*".

$$ACTD_i^{Meas}(n) = \frac{\overline{Lq}_i^{\,n}}{\lambda_{e_i}^{n}} \tag{3}$$

where, $\overline{Lq}_i^{\,n} = \sum_{j=1}^{n} Lq_i(j)/n$; $Lq_i(j)$ is a queue length at the end of the j-th frame time. The effective arrival rate $\lambda_{e_i}^{n}$ is given by (4). Where r_f is a frame rate.

$$\begin{aligned} \lambda_{e_i}^{n} &= r_f \times \lambda_i^n\left(1 - CLR_i^{Meas}(n)\right) \\ \lambda_i^n &= \sum_{j=1}^{n} x_i(j)/n \end{aligned} \tag{4}$$

Therefore, the average cell transfer delay $ACTD_i^{Meas}(n)$ can be easily computed by measuring the effective arrival rate and queue length at every frame time.

3.2 Dynamic Bandwidth Request and Allocation

Unlike the feedback control or re-negotiation methods between the user and networks, in our mechanism the bandwidth request and allocation are performed within a switch without any interaction with a source. It indicates that there is a very small signaling delay or overhead.

At each bandwidth request point, according to the following formula the necessary bandwidth for each connection is requested to a switch server.

$$R_i(n+1) = \max\left\{\left[\begin{array}{l} \hat{x}_i(n+1) + Lq_i(n) \\ -\kappa\left[\hat{x}_i(n+1) + Lq_i(n)\right] \times \ln\left[SD_i(n)\right] \end{array}\right], 0\right\} \tag{5}$$

The bandwidth request $R_i(n+1)$ for connection i is computed by adaptive adding or subtracting based on the quality of service satisfaction degree $SD_i(n)$ to or from the basis request $\hat{x}_i(n+1) + Lq_i(n)$, which is the amount of the next predicted input rate plus the current queue length. In (5), if a switch has provided the same amount of quality to the user with the user's QoS requirements, then $SD_i(n)$ will be 1 and the bandwidth request for the next frame only consists of the basis request. If the provided QoS is poor (i.e., $SD_i(n)$ is less than 1) or superfluous (i.e., $SD_i(n)$ is greater than 1), then some amount of bandwidth will be added or subtracted depending on the provided QoS to (or from) the basis request, respectively. In this case, the switch withdraws the redundant resources from the luxurious connections and uses them for poverty connections or other service classes such as best effort service. Our proposed dynamic bandwidth request scheme adaptively maintains every connection's QoS satisfaction degree to be close each other.

```
$N$ = the number of connections, $n$ = current frame time
For $(i = 1 \text{ to } N)$ Do
    $R_i(n+1) = \max\left\{\left[\begin{array}{l} \hat{x}_i(n+1) + Lq_i(n) \\ -\kappa[\hat{x}_i(n+1) + Lq_i(n)] \times \ln[SD_i(n)] \end{array}\right], 0\right\}$
Endfor
If $\left(\sum_{k=1}^{N} R_k(n+1) < C\right)$, Then
    $A_i(n+1) = R_i(n+1), \ \forall i$
Else
    $A_i(n+1) = \left\lfloor R_i(n+1) \times C \big/ \sum_{k=1}^{N} R_k(n+1) \right\rfloor, \ \forall i$
Endif
```

Fig. 1. The proposed bandwidth allocation procedure.

In Equation (5), κ is a scaling constant to control the amount of the bandwidth request $R_i(n)$. Using large κ results in a faster convergence to a suitable rate (bandwidth) to meet QoS requirements and quicker response to the input cell rate and measured quality changes. However after convergence, $R_i(n)$ will have large fluctuations. In the following simulations of this paper, we set $\kappa = 1.0$ intuitively. However, the optimality of this scaling constant remains to be justified. Fig. 1 shows the pseudo code of the proposed dynamic bandwidth request and allocation method.

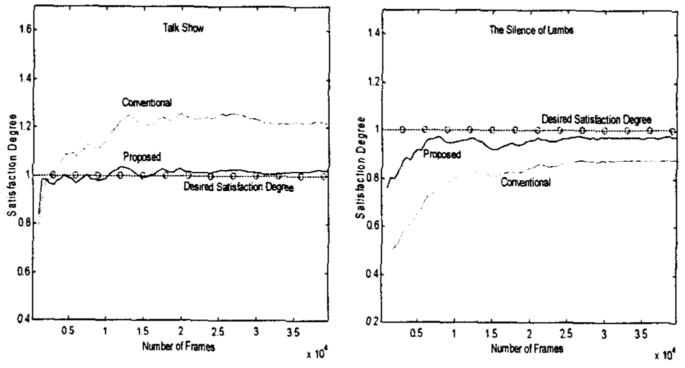
3.3 Performance Evaluation of the Proposed Method

To compare with conventional dynamic scheme, we have implemented a dynamic bandwidth allocation method of [2]. For simulation study, five MPEG-compressed video traces obtained from [5] (about half-hour each) are used. Since I, P, and B frame types have different statistical characteristics, we have predicted each frame type separately. In the proposed method, we set the scaling constant $\kappa = 1$ and buffer size for each connection $BUFF = 200$ cells. We assume that all sources start at the same time with I frame type for the worst case simulation.

In order to show how the proposed dynamic method provides the required QoS to each user when all users request the same QoS requirements and the trunk link capacity is limited, the following experiment is carried out. The required QoS is 5 msec average delay and 10^{-3} cell loss ratio; the trunk link capacity C is $AACR/0.3 = 4{,}225$ cells/sec. Here $AACR$ (average aggregated cell rate) is the sum of average input rates for all sources. As shown in Table 1, because the conventional method does not have QoS control, the values of QoS satisfaction degree (SD) of the connections are quite different. Even for some connections, network provides poor quality of service. However, in our method the satisfaction degrees of all sources can be maintained very closely to 1.0.

Table 1. Provided quality of service degree
(unit: msec (for delay), 10^{-3} (for cell loss ratio)).

Video	QoS Requirements		Measured Quality (Conventional)			Measured Quality (Satisfaction Degree-based)		
	ACTD	*CLR*	*ACTD*	*CLR*	*SD*	*ACTD*	*CLR*	*SD*
Star Wars	5 .0	1.0	5.48	1.37	**0.73**	3.28	1.03	**0.97**
Talk Show	5 .0	1.0	3.98	0.02	**1.22**	4.88	0.25	**1.02**
Terminator	5 .0	1.0	5.60	0	**0.89**	5.17	0	**0.97**
The Silence of Lambs	5 .0	1.0	5.70	0.53	**0.88**	5.12	0.53	**0.98**
Mr. Bean	5 .0	1.0	3.77	0.80	**1.31**	4.80	0.81	**1.03**



(a) Talk Show (b) The Silence of Lambs

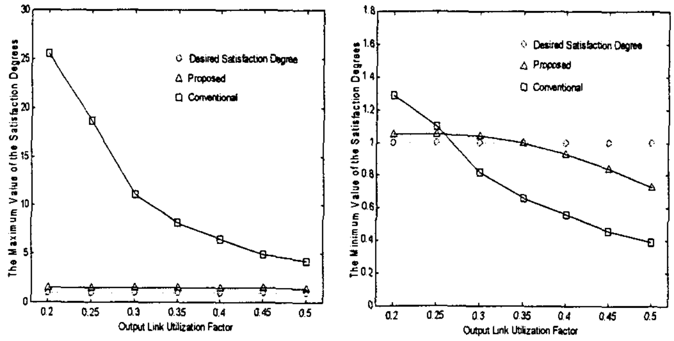
Fig. 2. Quality of service satisfaction degree.

Fig. 2 represents the QoS satisfaction degree as a function of time for "Talk Show" and "The Silence of Lambs" video sequences. We can observe that our proposed method provides very stable quality very close to the desired level (1.0).

As the second experiment, we have investigated the quality and allocated bandwidth changes in accordance with varying the output trunk link capacity. Different QoS requirements are used for simulation as shown in Table 2.

Table 2. The QoS requirements for each connection.

Connection	1	2	3	4	5
Video	Star Wars	Talk Show	Terminator	The Silence of Lambs	Mr. Bean
ACTD QoS	5 msec	5 msec	50 msec	50 msec	50 msec
CLRQoS	0.002	0.002	0.02	0.02	0.02



(a) maximum values (b) minimum values

Fig. 3. Maximum and minimum QoS satisfaction degrees.

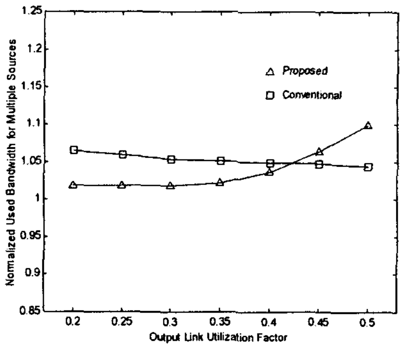


Fig. 4. Normalized used bandwidth.

Fig. 3-(a) and -(b) show the maximum and minimum values of the QoS satisfaction degrees of the five video connections, respectively. The output link utilization factor is defined as $AACR/C$. In the ideal dynamic bandwidth allocation, because the networks should meet all user's QoS requirements strictly and at the same time network resources should be used efficiently (i.e., high utilization), the maximum and minimum values of the QoS satisfaction degrees, which are close to 1.0, are preferable. Compared with the conventional method, we can see that the proposed scheme tries to keep the maximum and minimum values of QoS satisfaction degree to the desired level. Fig. 4 shows the used bandwidth as a function of output link utilization factor. The normalized used bandwidth is a ratio between the total allocated bandwidth for all connections and $AACR$. Fig. 4 indicates that our dynamic method needs less bandwidth than the conventional one

at the QoS supportable region (in Fig. 4, where the output link utilization factor is less than 4.0) through dynamic bandwidth adjustments between the connections. However, in accordance with decreasing trunk link capacity (i.e., increasing the output link utilization factor), QoS support is going to be more difficult so that our method allocates more bandwidth to the connections to meet the desired QoS.

However, even though our method adaptively controls the provided QoS, if the output trunk link capacity is too small to support the requested QoS for the superposed traffic, then as shown in Fig. 3-(b) our method also may not provide the desired level of quality of service for all connections. Therefore, to provide a consistent service of quality to a user during the connection period, we need a call admission control, which considers the quality of service satisfaction degree.

4. Call Admission Control

For on-line video applications, we assume that networks do not have any prior knowledge of the user's traffic statistics. Therefore, the call admission control using the traffic parameters can not be applicable. In this section we consider a suitable call admission scheme to the dynamic bandwidth allocation without any user traffic information. In fact, in the conventional dynamic approaches, the call admission control has not deeply considered and generally all incoming call requests are admitted. It leads to excessive quality of service variations.

When the current frame time is n and (in addition to the current existing N connections) the $(N+1)$-th new call is requested, in our proposed call admission control, only if the following two conditions in (6) are satisfied, would the new call be admitted.

$$\begin{cases} i) \;\; f_1(A_1, A_2, \cdots, A_{N-1}, A_N) < C, \;\; and \\ ii) \;\; f_2(SD_1, SD_2, \cdots, SD_{N-1}, SD_N) \ge 1 \end{cases} \tag{6}$$

In (6), function $f_1(\;)$ is defined as,

$$f_1(A_1, A_2, \cdots, A_{N-1}, A_N) = \sum_{i=1}^{N} E[A_i] \times \beta(N) \tag{7}$$

where, $E[A_i]$ is an average allocated bandwidth for connection i. As a result, $f_1(\;)$ is a multiplication of $\beta(N)$ and the sum of average allocated bandwidth for each connection. Here, $\beta(N)\,(>1)$ is a marginal value to leave some proportion of C for the new connection; as the number of existing connections N is increasing, a smaller value of $\beta(N)$, which is close to 1.0, can be used. By a network operator, $\beta(N)$ can be defined as a function of N or a constant value.

In the second condition of (6), function $f_2(\;)$ is defined as,

$$\begin{aligned} & f_2(SD_1, SD_2, \cdots, SD_{N-1}, SD_N) = \\ & \min\{SD_1(n), SD_2(n), \cdots, SD_{N-1}(n), SD_N(n)\} \end{aligned} \tag{8}$$

Namely, by the second condition of (6), a new call request is accepted only if the QoS satisfaction degrees for all existing connections are maintained over the desired level.

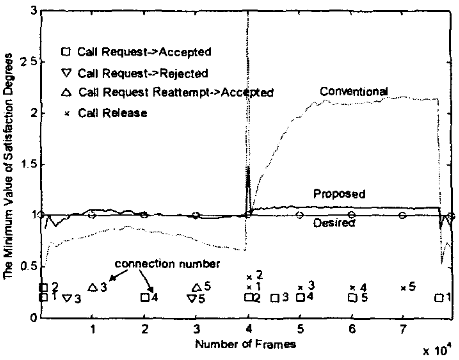


Fig. 5. QoS support using call admission control

Fig. 5 shows the variation of the minimum value of satisfaction degrees of all current existing connections according to the new call requests and releases. For simulation, five kinds of video sequences in Table 2 are used and the output trunk link capacity is set at 8750 cells/sec. In the conventional method, all requested calls are admitted immediately. For the proposed call admission control, '$\kappa = 1.0$, $BUFF = 200$ cells, and $\beta(N) = 1.2$' are used. As we can see in Fig. 5, with the proposed admission control, networks can very closely provide the desired quality of service for all admitted connections.

5. Conclusions

In this paper, we have proposed a dynamic bandwidth allocation and call admission control method, which can support user's QoS requirements in terms of loss ratio and average delay very well and at the same time can achieve high utilization. Some simulation results show that our proposed dynamic approach can provide efficient resource management and desired level of quality of service. Our future research in this area may involve more robust call admission scheme and the QoS control for maximum transfer delay, which is considered very important QoS parameter for real time videos.

References

[1] Song Chong, San-qi Li, Joydeep Ghosh, "Predictive dynamic bandwidth allocation for efficient transport of real-time VBR video over ATM", *IEEE Journal on Selected Areas in Communications*, Vol. 13, No.1, pp. 12-23, 1995.

[2] Abdelnaser Mohammad Adas, 'Using adaptive linear prediction to support real-time VBR video under RCBR network service model", *IEEE/ACM Transactions on Networking*, Vol. 6, No. 5, pp. 635-644, 1998

[3] Girish Chiruvolu, Ravi Sankar, N. Ranganathan, "An adaptive scheme for better utilization with QoS constraints for VBR video traffic in ATM networks", *3rd IEEE Symposium on Computer and Communications*, pp. 3-7, 1998.

[4] Xinyu Wang, Souhwan Jung, James S. Meditch, Dynamic bandwidth allocation for VBR video traffic using adaptive Wavelet prediction", *IEEE ICC'98*, pp. 549-553, 1998.

[5] ftp-info3.infomatik.uni-uerzbug.de/pub/MPEG/