DYNAMIC BANDWIDTH ALLOCATION WITH OPTIMAL NUMBER OF RENEGOTIATIONS IN ATM NETWORKS

Fatih Murat Porikli Zafer Sahinoglu

Mitsubishi Electric Research Labs
Murray Hill, NJ 07974

Abstract

This paper introduces a scheme for online dynamic bandwidth allocation for variable bit rate (VBR) traffic over ATM networks. The presented method determines the optimum bandwidth renegotiation time and bandwidth amount to allocate to VBR traffic source by minimizing predefined cost functions. Traffic rate predictor designed by wavelets is provided as feedback to the system. The results show that the introduced scheme minimizes both under-utilization of the available capacity and queuing delays. The proposed method can also be deployed in Weighted Fair Queuing disciplines to dynamically update a weight coefficient assigned to an application.

1 Introduction

ATM networks provide connection-oriented services with guaranteed bandwidth. To carry an IP datagram in such networks, a virtual circuit (VC) has to be setup with an indicated bandwidth requirement. Once a VC is open, the adaptation layer has to decide how long to keep the VC open with the initial bandwidth assignment. If the rate of the incoming packets matches the specified bandwidth allocation, VC is kept open [1], [2], [3]. However, if packets arrive at a higher or lower rate, there is a need to readjust the allocated resource or even to close the VC. Periodic algorithms adjust the bandwidth allocation in fixed time intervals. On the other hand, adaptive algorithms respond whenever a change is necessary as long as the updating process is not frequent. Readjusting can be done in two different ways either closing the existing VC and opening a new one with new allocation or changing the allocation of the current VC in lieu of closing it [2]. The second option, if preferred, must be supportable by the network. Indeed, the Q.2963 series of recommendations belongs to the DSS 2 family of ITU-T Recommendations and specifies the procedure of the modification of traffic parameters of a call/connection in the active state. Recommendation Q.2963.3 defines the procedure of the ATM Traffic Descriptor modification with renegotiation that is equivalent to that specified in Recommendation Q.2962. Therefore, we can expect that network provides such a support and bandwidth allocation to a VC can be updated without closing it. To follow temporal variations in bandwidth demand of VBR sources, we propose a method for dynamic bandwidth allocation with minimal number of renegotiations. Each renegotiation process involves a signaling between network and a source. The renegotiation frequency is a trade-off between signaling overhead and high bandwidth utilization. High renegotiation frequency loads the network with heavy overhead. On the other hand, long inter-renegotiation intervals make the follow-up of the traffic bit rate pattern difficult. Renegotiation is only feasible in time scales of several seconds [4]. In [5], it is suggested that minimum of 1 sec and an average of 5 seconds or more for renegotiation is a good compromise. It is crucial that optimal number of bandwidth renegotiations must be performed under predetermined cost constraints such as under utilization ratio and packet/cell transmission delay.

The proposed method consists of two main parts:

the bandwidth predictor and the renegotiation decision unit, RDU. Section II elaborates on the predictor design which uses wavelets and signal energy distribution in frequency domain to predict the bandwidth demand of the source for the next discrete time slot. Section III explains the cost functions and strategies in determination of the resource renegotiation time and amount. Section IV gives performance results of the introduced approach when applied to a real time MPEG-4 coded video trace.

2 Bandwidth Prediction

There are several different strategies in dynamic bandwidth allocation to predict the future bandwidth demand of a traffic source. Each new allocation consists of a prediction and a correction term based on previous updates [6]. The simplest bandwidth predictor is the previous value of the bit rate as an estimate. In our simulations, we use a bandwidth predictor proposed in [7]. Briefly, the predictor decomposes the time series traffic data, each element of which consists of bit arrival information, into different frequency bands by applying dyadic tree filter banks and separates the low and high frequency components in the arrival process. The energy distribution in each frequency band gives information about the contribution of these components to the main traffic pattern. There is an analogy between that the optimum fixed quantizer allocates bits based on the computed variances in the spectral coefficients and that the bandwidth for the next time slot based on the computed variances in the signal energies computed by the spectral coefficients returned by the wavelet transform. Energy distribution information is used as a feedback parameter in prediction of the new arrival rate. Passing the original traffic data through scaling and wavelet filters in the dyadic tree returns the coarse component, and the details of the original trace revealing irregularities and sharp changes in traffic behavior.

3 Cost Function

In order to avoid under allocation (buffering) and under utilization of the bandwidth, the allocated bandwidth should be dynamically adapted to follow-up of the traffic bit rate pattern (Fig.1). However, each adaptation (renegotiation) process involves a signaling between the network and the source. High renegotiation frequency loads the network with heavy overhead. On the other hand, long inter-renegotiation intervals make the follow-up of the traffic bit rate pattern difficult. Therefore, optimal number of bandwidth renegotiations must be provided under constraints such as the cost of under utilization, cost of renegotiation, and buffer size as shown in Fig. 2.

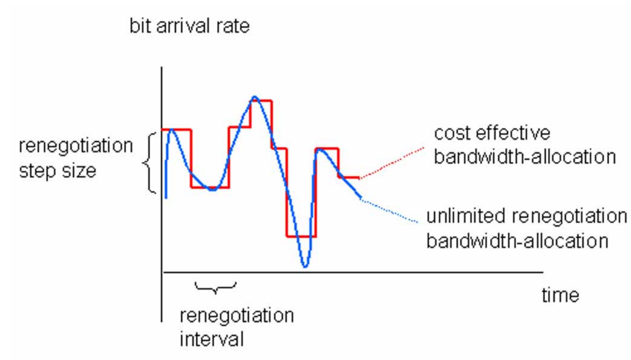


Figure 1: Bandwidth renegotiation problem for VBR.

We determine an optimum bandwidth allocation $a(n+1)$ for real-time traffic at a future time $n+1$ given a current traffic bit arrival rate $r(n)$, and current allocated bandwidth $a(n)$ at time n. To attain the optimum solution, we design a total cost function J that includes costs of under utilization in terms of $u(n)$, under allocation in terms of $b(n)$, and renegotiation as

n	time
β	buffer size
$a(n)$	bandwidth allocation at time n
$r(n)$	bit arrival rate at time n
$\hat{r}(n)$	predicted arrival rate at time $n+1$
$w(e)$	bandwidth cost function
$b(n)$	size of the queue (buffered bits)
$u(n)$	size of the under-utilized bandwidth (bits)
$e(n)$	bandwidth error function (bits)
$T(n)$	cost of the renegotiations
$J(n)$	total cost function

Table 1: Notation of parameters and functions

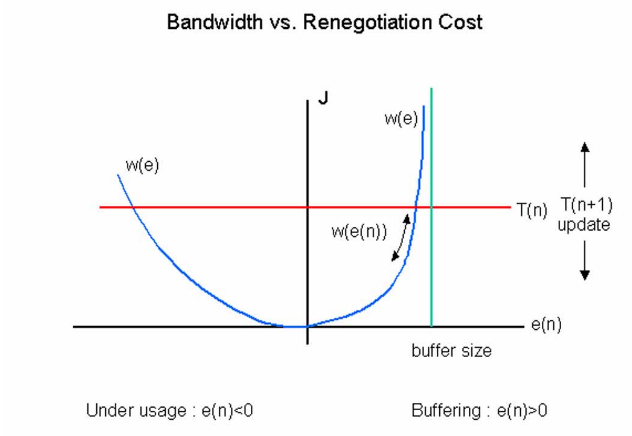


Figure 2: Analytic representation of cost functions.

$T(n)$. The cost function J is defined as

$$J = w_b b(n) + w_u u(n) + T(n) \quad (1)$$
$$= w(b(n) + u(n)) + T(n) \quad (2)$$
$$= w(e(n)) + T(n) \quad (3)$$

where the $e(n)$ is the bandwidth error such that

$$e(n) = \sum_{i=0}^{n} r(i) - a(i) + \hat{r}(n+1) - a(n) \quad (4)$$

In the cost function, the size of the queue $b(n)$ and the size of the under used bandwidth $u(n)$ are weighted by shaping functions w_b and w_u, and then added to the cost of renegotiation $T(n)$. Using separate cost terms for under utilization and under allocation enables us to adapt the optimization method for various types of applications; weighted fair queuing (WFQ) algorithms, ATM switches, etc. In addition, one cost term can be preferred to the other with respect to the changing network conditions, i.e., buffer cost can have nonlinear dependence on the current queue size, cost of bandwidth can be alternating at the certain times of the day. The under allocation happens if the allocated bandwidth is not enough to handle the bit arrival rate. Incase of under allocation, the excess bits are queued in the buffer. The buffered bits are sent when the arrival bit rate is less than the allocated bandwidth, thus, there is available bandwidth to forward bits from buffer. The under utilization $u(n)$ occurs when the allocated bandwidth is greater than the bit arrival rate and the buffer is empty. Therefore, the allocated bandwidth is not fully used. The bandwidth error function $e(n)$ is the $u(n)$ for under utilization, and $b(n)$ for under allocation, and it also includes the predicted bandwidth error for time $n + 1$ to include the effect of keeping the same bandwidth allocation level. Obviously, the bandwidth cost function $w(e(n))$ corresponds to the under allocation cost if $e(n)$ is more than zero, and under utilization cost vice versa.

While optimizing J, the renegotiation step size and time are obtained. The determination for $a(n + 1)$ is made by minimizing the cost function

$$a(n+1) = \arg\min J$$
$$= \arg\min[w(e(n)) + T(n)] \quad (5)$$

To understand the properties of minimization, let us investigate the impact of each cost term on J. It is worthwhile to realize that the renegotiation cost $T(n)$ should be high if there was another bandwidth renegotiation made recently at time $n - \delta$ where δ is a small time period. By the increasing values of δ, which also means that the last renegotiation was made long past; the cost of renegotiation should be decreasing because renegotiation becomes more affordable. Therefore, the time period δ between the current time and the very last renegotiation determines the magnitude of the variable cost function $T(n)$:

$$T(n) = \begin{cases} \alpha^{+} T(n-1) & a(n) \neq a(n-1) \\ \alpha^{-} T(n-1) & a(n) = a(n-1) \end{cases} \quad (6)$$

If the bandwidth cost function $w(e(n))$ becomes larger than renegotiation cost for the predicted traffic, it becomes advantageous to renegotiate to prevent from the expansion of $w(e(n))$. Fig. 3 presents this analogy. Incase a constant renegotiation cost is preferred over to variable cost term, the number of renegotiations may multiply if the newly allocated bandwidth is incapable of reducing the under allocation or under utilization costs quickly. The J fluctuates close to the decision boundary; each time bandwidth cost function becomes higher than renegotiation cost, a new renegotiation is made.

Considering the under utilization and under alloca-

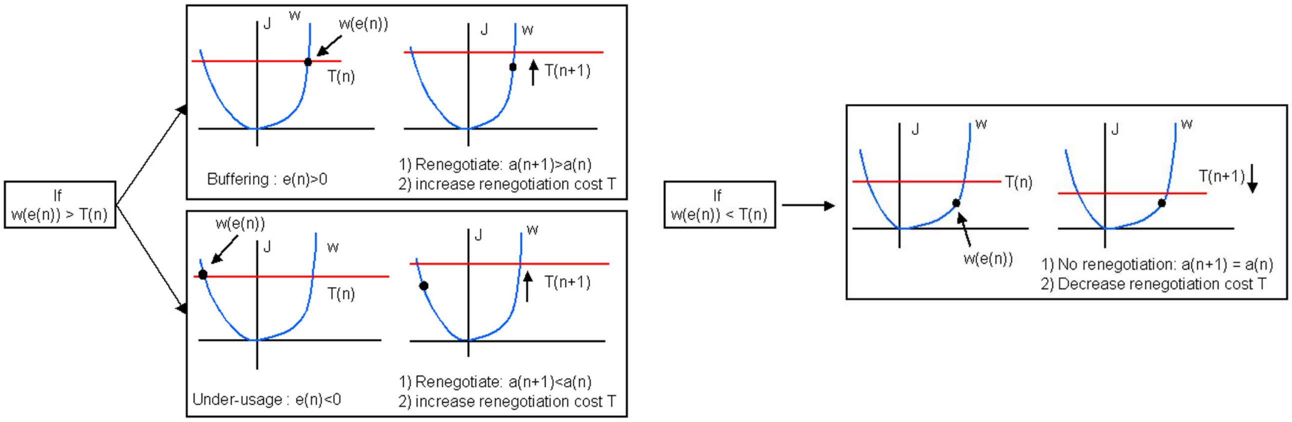


Figure 3: Relation between $w(e(n))$ and $T(n)$.

tion, we form w(e(n)) as

$$w(e(n)) = \begin{cases} e^K & e(n) > 0 \\ |e^L| & e(n) \leq 0 \end{cases} \quad (7)$$

We choose $K > L > 1$ to weight under allocation cost more. For the hard buffer size constraint case (i.e. no buffer overload permitted), $w(e(n))$ becomes infinity at β by asymptotically converging to the $e(n) = \beta$. In the above equations, the cost of bandwidth $w(e(n))$ is assigned as a combination of polynomial functions, still it can be defined by piece-wise continuous or exponential functions.

We summarized solving of the minimization problem in terms of $a(n)$, $r(n)$, and $\hat{r}(n)$ in Fig. 4.

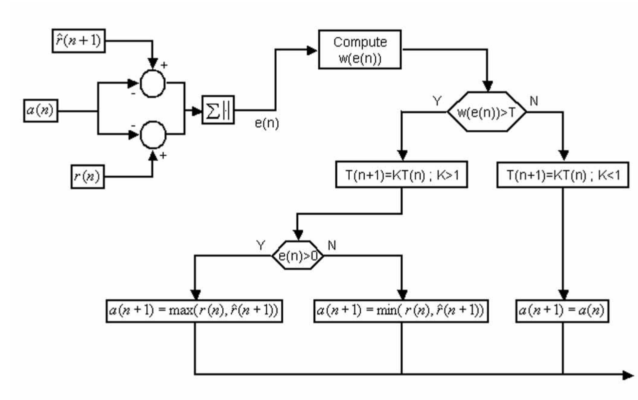


Figure 4: Real-time determination of bandwidth.

4 Conclusion

Sample simulation results for a typical MPEG-2 video sequence is presented in Fig.5. The first figure (Fig. 5-a), shows the bit arrival rate and its prediction by the wavelet method. The original data is also plotted in the rest of the graphs for comparison. The computation time of the bandwidth renegotiation method is negligible; the process is real-time. A hard buffer model that does not permit overshoot of the queue size is utilized. We simulated bandwidth allocation for different severity degrees of the renegotiation cost and buffer size. Fig.5-b presents the very high renegotiation cost $\alpha^+ = 4.0$ and $\alpha^- = 0.99$ scenario result. The number of renegotiations is 6 for 250 seconds of data sequence, and the bandwidth utilization ratio is 71.10%. This is the ratio of the total arrived data bits to the total allocated bandwidth; the ratio of areas under each functions in the figures. The following rows, Fig.5-c,d,e, are the results for high ($\alpha^+ = 2.0$, $\alpha^- = 0.95$), medium ($\alpha^+ = 1.3$, $\alpha^- = 0.95$), and low ($\alpha^+ = 1.3$, $\alpha^- = 0.85$) renegotiation cost scenarios. The number of renegotiations increases to 15, 33, 64, respectively because the renegotiation cost is assumed to be lower. The utilization ratios are found as 83.14%, 82.56%, 86.05%. We observed that by selecting smaller renegotiation costs, we increase the number of renegotiations that leads the higher utilization ratios. Furthermore, using larger buffer size allow the network to renegotiate less. We are able to allocate optimal bandwidth to the variable bit rate video traffic over the ATM switches dynamically in real-time. In addition, the method attains high utilization ratios while achieving the minimum total cost. Another significant advantage of the method is the ability of adapting to the network conditions as the constraints may change, i.e., buffer cost can have nonlinear dependence on the current queue size, cost of bandwidth can be alternating at the certain times of the day, etc.

References

[1] H. Saran, S. Keshav, "An empirical Evaluation of Virtual Circuit Holding Times in IP over ATM Networks", Proc. of INFOCOM, 1994.

[2] Y. Afek, M. Cohen, E. Haalman, Y. Mansour, "Dynamic Bandwidth Allocation Policies", 0743-166X/96 IEEE.

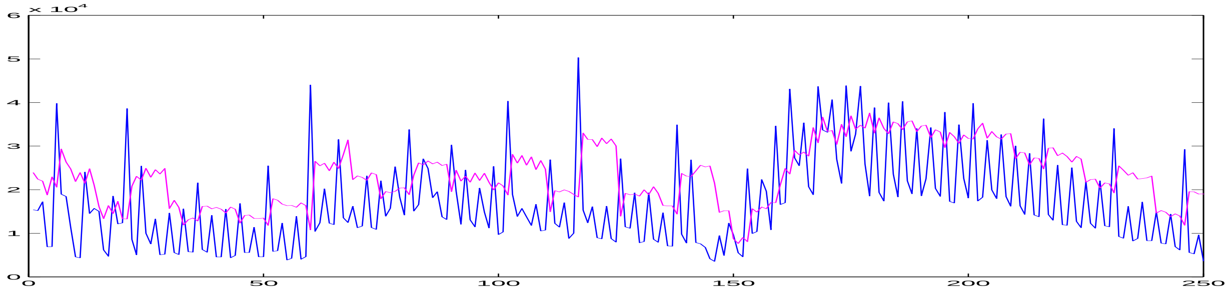
[3] S. K. Biswas, R. Izmailov, "Design of a fair Bandwidth allocation Policy for VBR Traffic in ATM Networks", IEEE/ACM Trans. On Networking, V:8, N:2, April 2000.

[4] K. Lakshman, R. Yavatkar, "An Empirical Evaluation of Adaptive QoS Renegotiation in an ATM Network", NOSSDAV 96, Zuhst, Japan, 1996.

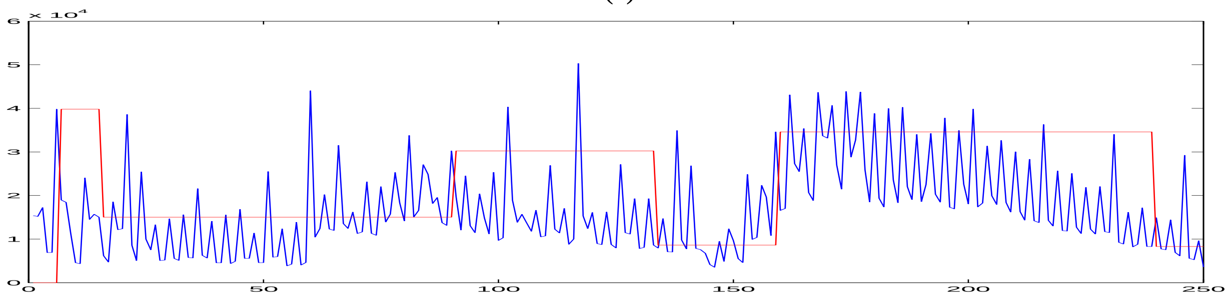
[5] S. Gumbrich, H. Emgrunt, and T. Brown, "Dynamic Bandwidth Allocation for Stored VBR Video in ATM End Systems", Proceedings of IFIP'97, 297-317, May 1997.

[6] S. Chong, S. Li, J. Ghosh, "Efficient Transport of real time VBR Video over ATM via Dynamic Bandwidth Allocation", Uni. Of Texas at Austin, Austin, Texas 78712, August 1995.

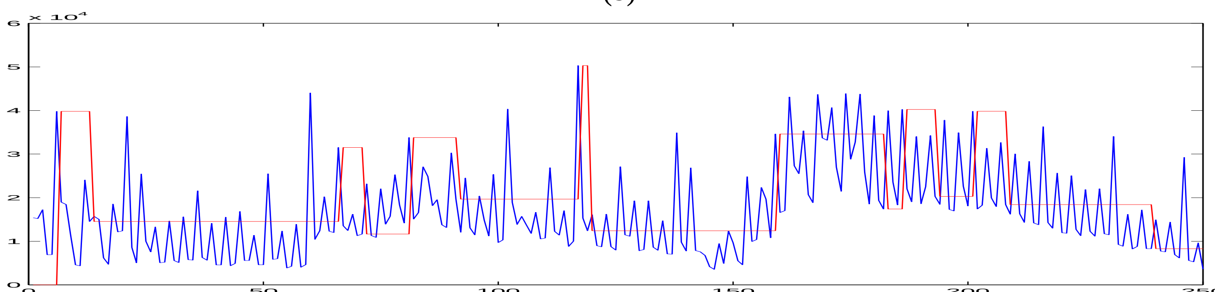
[7] Z. Sahinoglu, S. Tekinay, "A Novel Adaptive Bandwidth Allocation: Wavelet Decomposed Signal Energy Approach", GLOBECOM 2001.



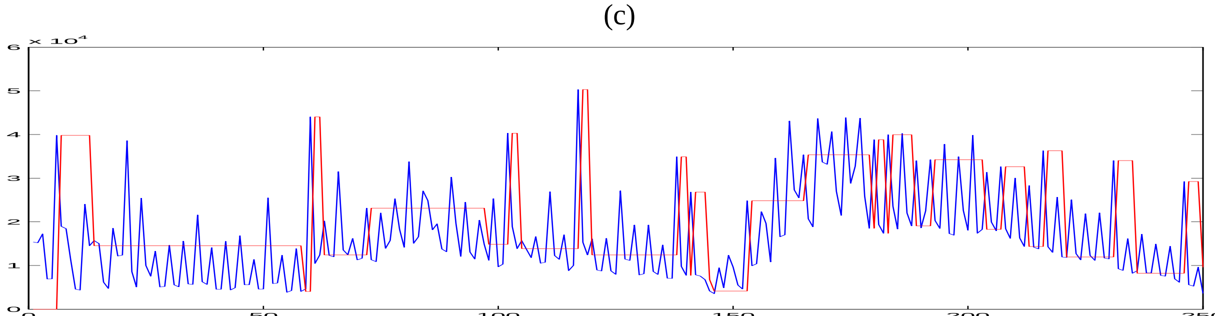
(a)



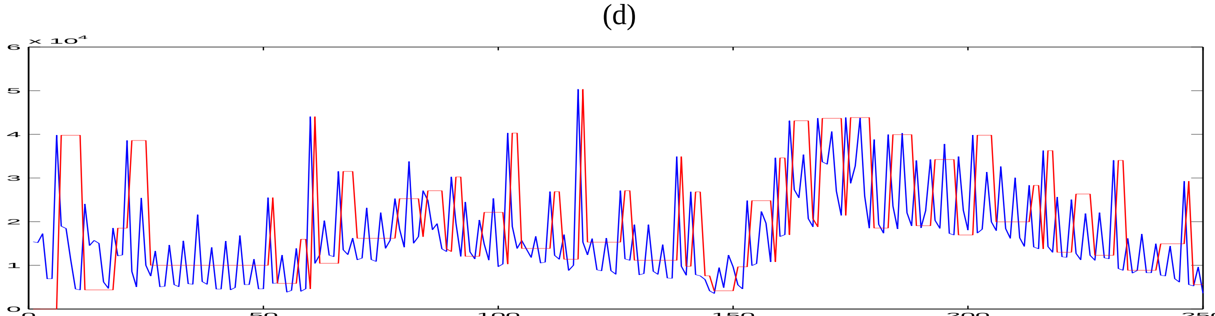
(b)



(c)



(d)



(e)

Figure 5: (a) Video traffic and its wavelet-based prediction, (b-c-d-e) optimum bandwidth allocations for different levels of renegotiation cost scenarios; very high (α^+=4.0, α^- =0.99), high (α^+=2.0, α^- =0.95), medium (α^+=1.3, α^- =0.95), and low (α^+=1.3, α^- =0.85). The number of renegotiations are 6, 15, 33, 64 respectively.