Dynamic Bandwidth Allocation for Stored VBR Video in ATM End Systems

S. Gumbrich, H. Emgrunt, T. Braun
IBM European Networking Center
Vangerowstr. 18, 69115 Heidelberg Germany
e-mail:(gumbrich,braun)@heidelbg.ibm.com
Phone: +49 6221 594369, Fax: +49 6221 593300

Abstract:

The paper introduces a scheme for delivery of variable bit-rate (VBR) coded video from a video server over ATM networks. To solve the problem of long term variations in bandwidth requirements, a dynamic bandwidth allocation scheme is proposed. This is based on renegotiation of bandwidth during the duration of the call.

An algorithm is presented, that determines optimized renegotiation points for a given VBR coded video by minimizing cost functions. Appropriate cost functions, and methods for reducing the run-time expense of this optimizing algorithm are outlined.

Results are presented for real VBR MPEG video. The results show, that dynamic bandwidth allocation reduces the bandwidth requirements significantly and that the presented method leads to lower costs compared to other known renegotiation schemes.

Keywords

ATM, MPEG, bandwidth allocation, video transmission, VBR

1 INTRODUCTION

Compressed video traffic is expected to be a significant source of the traffic mix in asynchronous transfer mode (**ATM**) networks. There are two ways to compress video streams before transmitting them over networks: Constant Bit Rate (**CBR**) and Variable Bit Rate (**VBR**) compression. Note, that compressed video data is variable bit rate in nature.

When a video source is compressed using CBR techniques, the result is a constant bit rate stream with varying quality. At times of peak activity, quality decreases since there is not enough bandwidth to transmit all information. In contrast, VBR compression results in constant quality but varying bandwidth requirements. This constant quality is advantageous for the user but more difficult to support by the network. In this paper we address transmission of VBR coded video over ATM networks.

One important characteristic of compressed video is its burstiness over multiple time scales (Rathgeb, 1993),(Garret,1994). There is not only the short term variation in the source rate due to different frame-types and variations within a scene, but also a variation over long time scales due to scenes with different information contents. For example, in an MPEG compressed version of the "Star-Wars" movie, there are episodes with a sustained peak rate of five times the long term average rate (Grossglauser, 1995). Those episodes last for more than 10 seconds.

Short term variations can be smoothed by shaping the video stream (Lam, 1994),(Graf,1994). This is not possible for the long term variations due to restrictions of delay and buffer space. Moreover, the long term variations lead to a ATM traffic descriptor that is mainly based on the episodes with the highest bandwidth demands. Therefore, large amounts of mostly unused bandwidth have to be reserved for the whole movie. Since billing in the future will be based on the reserved bandwidth this leads to unnecessarily high costs.

To solve the problems of long term variations in transmission of videos we propose the concept of dynamic bandwidth allocation via renegotiation. Renegotiation means sending a signalling message along the path, requesting an increase or decrease of the current service rate corresponding to a new traffic descriptor. If the request is feasible, the network allows the renegotiation, and upon completion of the request, the source is free to send data compliant to the new traffic descriptor.

Our scheme is based on the ATM VBR service, extended by renegotiation functionality. The traffic contract for the VBR service includes an extended traffic descriptor. This traffic descriptor has to be determined by the user. A policing algorithm (generic cell rate algorithm, **GCRA**) checks the conformance of the arriving source traffic with this traffic descriptor (ATM Forum,1995). Non-Conforming cells are discarded or tagged indicating low priority. This has to be avoided since it leads to an uncontrollable service degradation. On the other hand, the traffic descriptor should be as tight as possible to avoid over-reservation of bandwidth. Chapter 2 describes a way to obtain a deterministic traffic descriptor for stored video. Chapter 3 explains the concept of dynamic bandwidth allocation. The network model used for evaluation is described in chapter 4. In chapter 5 the algorithm for finding the renegotiation points and appropriate cost functions is presented. Schemes to accelerate the algorithm are discussed in chapter 6. Finally (chapter 7) we compare our algorithm to algorithms of other researchers.

2 DETERMINATION OF THE ATM VBR TRAFFIC DESCRIPTOR

The ATM Forum Traffic Contract (ATM Forum,1995) specifies the negotiated characteristics of an ATM Layer Connection at the UNI. It consists of a traffic descriptor and a set of QoS parameters. The traffic descriptor specifies the conforming cells of the ATM connection. It consists of the elements Peak Cell Rate (**PCR**), Sustainable Cell rate (**SCR**), Maximum Burst Size (**MBS**), Cell Delay Variation Tolerance (**CDVT**) and the Conformance Definition based on two GCRA algorithms.

The CDVT is independent of the source traffic and is therefore provided by the network adapter. It is the users task to determine the parameters PCR, SCR and MBS. An algorithm is necessary to select those parameters in such a way that no cells are tagged or discarded by the GCRA policing algorithm (i.e. to satisfy the conformance definition).

An *empirical envelope* is used (Cruz,1991), (Cruz,1995), (Knightly,1995b), (Wrege,1996) to give the most accurate, time-independent, deterministic characterization of the source traffic. We use the empirical envelope to determine the ATM traffic descriptor. The empirical envelope ε(I) for an arrival function b(t) is defined as shown in Figure 1. The quantity of data provided from the source within any interval of time I is less than a value ε(I) that depends on the interval length.

For our experiments we assume that the traffic shaping function used is “frame period averaging” (Graf,1994). This results in equally spaced cells over each frame period (33ms). Usage of more advanced traffic shaping functions leads to lower values of the empirical envelope, but has no impact on the schemes presented here.

$$\varepsilon(I) = \max_t \int_t^{t+I} b(\tau)\, d\tau$$

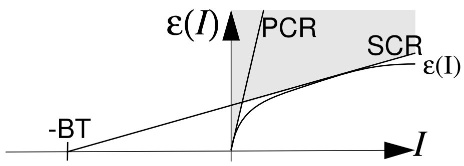


Figure 1: Empirical envelope and traffic descriptor

In Figure 1 it is shown how the source traffic descriptor is determined from a given empirical envelope. The PCR is determined by a tangent to the empirical envelope that goes through the origin. On the contrary, there is an infinite set of {SCR,BT} pairs that lead to minimal conformant traffic descriptors. The only requirements for the construction of the SCR line are to touch and not to cross the empirical envelope. This makes sure that the traffic descriptor is conform to the source traffic and as tight as possible. After selecting one of the SCR values the BT is defined by the crossing of the SCR line with the x-axis. The sum of BT and the time for sending all cells of a connection specify the duration of that connection. This term is relevant for the definition of SCR. The signalling Parameter MBS is determined from

$$MBS = \left\lfloor 2 + \frac{BT}{\frac{1}{SCR} - \frac{1}{PCR}} \right\rfloor$$

The value 2 is added to ensure that the GCRA does not tag or discard cells conformant to the given empirical envelope $\varepsilon(i)$. This is necessary, due to the convention of the ATM Forum (ATM Forum, 1995) to use the following equation to derive BT_{GCRA} from the signalling parameters (MBS,SCR,PCR).

$$BT_{GCRA} = (MBS - 1) \times \left(\frac{1}{SCR} - \frac{1}{PCR}\right)$$

To select the value for SCR we use a heuristic scheme introduced in Wrege et al. (1996). In this scheme an additional criterion to minimize the weighted difference between the traffic descriptor and the empirical envelope is used. Note that this way of constructing the empirical envelope and the determination of the deterministic traffic descriptor only works for frame sizes known at connection setup time (i.e. stored video). This deterministic source traffic descriptor is mainly determined by long scenes with high bandwidth requirements. Since those scenes are common in real VBR MPEG videos, this leads to over reservation of bandwidth. Dynamic bandwidth allocation accounts for this behavior.

3 DYNAMIC BANDWIDTH ALLOCATION

Bandwidth allocation techniques for video can be divided into two classes: dynamic and static. If static bandwidth allocation is used, the traffic descriptor is valid for the complete duration of the connection. If the source traffic descriptor can be changed without setting up a new connection the term dynamic bandwidth allocation is used.

Dynamic bandwidth allocation is necessary since scenes with different information contents can result in extreme changes in the bandwidth requirements. If only static bandwidth allocation is performed, the worst case scenes determine the empirical envelope.

This is due to the fact that the empirical envelope is equivalent to the maximum number of data provided from the source for each interval within the validity of the source traffic descriptor. This leads to over reservation of bandwidth during parts (scenes) of the video with lower bandwidth requirements.

A requirement for dynamic bandwidth allocation is that the network allows renegotiation of the source traffic descriptor. ITU-T Q.2963 defines an extension of ITU-T Q.2931 signalling to allow renegotiation messages. The ATM Forum has no renegotiation messages defined (ATM Forum, 1996), but renegotiation is planed as future work item for the next release (Kaufman,1996). With this knowledge we can expect that ATM networks will be capable of supporting renegotiation in the future.

During renegotiation a switch controller does not need to recompute routing, allocate a VCI or acquire housekeeping records. This reduces the renegotiation overhead. Nevertheless renegotiation is only feasible in time scales of several seconds (Lakshman, 1996) to avoid excessive signalling. The renegotiation frequency is a trade off between signalling overhead and a high utilization of the reserved bandwidth. A minimum of 1 second between two renegotiations and an average of more than 5 seconds is a good compromise.

The schemes dealing with video can be divided into two classes: off-line and on-line. For live video the frame sizes are not known at connection setup or renegotiation time. Schemes that deal with live video are called on-line schemes. In contrast, the frame sizes for stored video are known in advance. Schemes relying on stored video are called off-line schemes. There are two main problems that have to be solved for an off-line dynamic bandwidth allocation scheme. First, the positions have to be determined where the bandwidth requirements of the video change. An optimized solution for this topic is presented in this paper. Second, the bandwidth requirements of the video have to be adapted to the reserved traffic descriptor if the network denies the request for more bandwidth. We currently investigate this problem (Gumbrich,1995) .

Several scientists already addressed the problems of dynamic bandwidth allocation. Chong et al. (1995) discussed on-line schemes for CBR video. Their main interest is the prediction of the required bandwidth. Grossglauser et al (1995) developed on-line and off-line schemes for a CBR service. For off-line they defined constant costs per renegotiation and per bandwidth unit. This cost-function is then minimized to find the proper renegotiation points. This optimizing algorithm is intended for renegotiated CBR service. We compare this algorithm to our algorithm in chapter 7.2 of this paper. For on-line they predict the next frame size and start renegotiation depending on the predicted frame size and the CBR buffer level.

Renegotiation for on-line VBR video is addressed by Pancha et al. (1993) and Knightly et al. (1995). Pancha et al. rely on the Fast Reservation Protocol and allows renegotiation for each new frame. Their prediction is based on cell statistics. Knightly et al. expect the network to use the D-BIND model (Knightly,1994) as traffic descriptor. The ATM Forum Traffic descriptor is a subset of the D-BIND model. On-line and off-line schemes are addressed in their work. Their off-line algorithm for VBR video is very efficient but not optimal. We compare this algorithm to ours in chapter 7.1

(Reininger,1995),(Bansal, 1995),(Mark, 1996) concentrate on determining VBR traffic descriptors for on-line video. They come up with a complete system that is capable of determining on-line renegotiation points. This system increases the quantization values of the coder, if the renegotiation request is not granted by the network. El-Henaoui et al. (1996) investigate a renegotiation system for on-line MPEG 1 audio. Their system is based on fast reservation. Lakshman et al. (1996) implemented and evaluated a renegotiation system with FORE switches and SUN workstations. They use the ATM CBR service (PCR reservation) for transmission of VBR videos.

Our concept is different to others in some important points. A renegotiation scheme for off-line VBR video, based on the ATM VBR service is investigated. We present the only optimizing algorithm for finding renegotiation points in this scenario. A further advantage of our work is, that we adjust the minimal distance between two renegotiations and the number of renegotiations in a video by a cost function for renegotiations. This is necessary to avoid overloading the signalling system of the network.

4 NETWORK MODEL

We need a model for the evaluation of bandwidth requirements. This is straight forward for CBR traffic since there is only one numerical value the PCR. For VBR the set {PCR, SCR, BT} has to be mapped to one value with a strict order to be comparable. This value is the maximum number of deterministic connections that can be transmitted by a given switch. Therefore, we use the model in Figure 2..

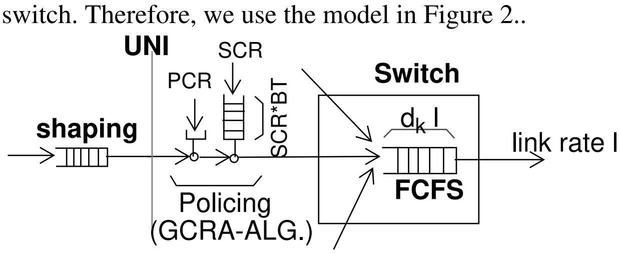


Figure 2: network model

In this model the switch uses a simplified **deterministic** call admission control (**CAC**) algorithm based on worst case assumptions. This allows for deterministic QoS guarantees, but leads to a lower network utilization compared to a statistical model. The problem with statistical models is, that it is still unsolved how to use the deterministic ATM source traffic descriptor to efficiently build a statistical descriptor for the CAC. The switch in Figure 2 has output buffers for multiplexing different incoming connections to outgoing links. We assume here a First-Come-First-Served scheduling (FCFS) strategy for the output buffer. This means that all packets are transmitted in the order of their arrival. The decision to use more advanced strategy like earliest deadline first or static priority (Knightly,1994), (Knightly,1995b), (Liebeherr,1994), (Cruz,1991) has no impact on our investigations.

Under the worst case assumption that all videos are identical and transmitted synchronously the conditions to guarantee a given delay bound d_k for each connection can be expressed as:

$$\{ max_{noc} \mid (noc \cdot b(t) - l \cdot t) \le d_k \cdot l \} \qquad \forall t$$

In this formula the link speed is denoted as l, b(t) is the source traffic descriptor and **noc** is the maximum number of connections that can be supported. Note, that the noc value is low if the bandwidth requirement of a given video stream is high. We use the noc value to determine the evaluation function and the cost functions.

An arbitrary shaping method can be chosen for shaping. For our investigations we assigned the method frame period averaging (Graf,1994). At the network edge a policing algorithm (GCRA) checks the conformance of incoming cells to the source traffic descriptor.

The evaluation function B(S) and the cost functions are defined based on this network model. B(S) is used for estimation of bandwidth requirements throughout this paper. The area between two renegotiations is called segment. Therefore, the off-line algorithm divides the video into the segments $\{S_1,...,S_m\}=S$. With this the evaluation function B(S) is defined as:

$$B(S) = \sum_{i=1}^{m} \frac{noc_i \times L_i}{L_S}$$

noc_i = number of connections for segment S_i
L_i = length of segment S_i
L_s = length of video

B(S) is equivalent to the weighted average of the number of admissible connections introduced in Knightly et al. (1995). With the knowledge of the long term average rate and the link speed this is converted to the average utilization in (Knightly, 1995).

5 OPTIMIZING OFF-LINE ALGORITHM O-RVBR

The goal of the off-line algorithm presented here is to find the optimal renegotiation points. Therefore we basically check all possible renegotiation points and decide which ones are associated with the lowest costs. Costs are considered for the amount of reserved VBR-bandwidth (source traffic descriptor) and the renegotiation itself. The cost function K(S) is defined as

$$K(S) = \left(\sum_{i=1}^{m-1} k(S_i) + r(S_i) \right) + k(S_m)$$

$k(S_i)$ = bandwidth cost for segment S_i
$r(S_i)$ = Cost for renegotiation at end of Segment S_i
m = number of segments

The goal of the algorithm is, to find the set of segments $S=\{S_1, ... ,S_m\}$ for which the cost function K(S) is minimized. We call the algorithm **O-RVBR** ("Optimizing algorithm for Renegotiated VBR video").

Since off-line algorithms rely on stored video, the frame sizes are known in advance. The determination of the renegotiation points and the corresponding traffic descriptors for each segment has to be done only once. As the results are stored together with the video at the video server, it is not necessary to do the calculations in real-time.

5.1 Bandwidth costs

ATM VBR traffic will be billed, depending on the amount of reserved bandwidth. Otherwise users could reserve huge amounts of unused bandwidth without paying for it. Therefore we considered the following aspect. For the network operator it is most advantageous to maximize the number of possible connections (with their requested QoS) for a given link speed. This reduces the cost per connection. We investigated different cost functions (Emgrunt,1996) based on the noc_i value defined in chapter 4. The most suitable function is

$$k(S_i) = \frac{l}{noc_i} \times L_i$$

l = link rate
L_i= Length of Segment S_i

The term l/noc_i is the amount of bandwidth that is reserved for Segment S_i in the CAC. This mainly determines the bandwidth cost. The term L_i avoids that cuts are detected at points with constant bandwidth requirement ($k(S_i)$ does not change). We proved [Emgr96] that L_i is necessary even if the renegotiation costs are decreasing linearly or exponentially with the length of the segment.

5.2 Renegotiation costs

The least bandwidth is reserved (lowest bandwidth cost) if renegotiation is performed after each frame. This is not suitable for common ATM switches since this will lead to an extraordinary high signaling overhead. Therefore, we introduce a cost function for renegotiation. In (Lakshman, 1996) it is shown by measurement that renegotiation in the range of “every few seconds” is suitable for ATM switches. To avoid unacceptable signaling overhead we introduced a minimal time interval MINTIME between two renegotiations. We use MINTIME=1sec in our investigations.

Ingenious renegotiation cost functions are constant or decrease with increasing segment length. We compared exponential with constant cost functions and drew the conclusion that with the same number of renegotiations the exponential cost function always leads to a lower rating value B(S) if the exponent is significantly larger than zero. Therefore only constant cost functions are investigated further. There are basically 3 possible ways to determine the constant value $r(S_i)=\gamma$. This value is responsible for the resulting number of segments.

- stream independent: This is the easiest way to determine γ. It is useful if there are real costs per renegotiation and for VBR bandwidth specified by the network providers. This also assumes a network service, that only restricts the number of renegotiations by the value of MINTIME.
- stream dependent: With this method the renegotiation costs depend on statistical characteristics of the video stream. Video streams with high variance in bandwidth requirements are assigned higher values for γ. The advantage of this method is, that the load on the signalling system of a switch increases slowly with increasing variance of bandwidth requirements.

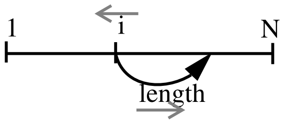
$$r(Si) = \begin{cases} \infty & ,\forall(\Delta t < MINTIME) \\ \frac{\sigma}{\mu} \cdot noc_{all} \cdot \alpha & ,\forall(\Delta t \geq MINTIME) \end{cases}$$

  In this formula σ is the variance and μ is the mean of the frame sizes of the video. σ/μ is a measure for the burstiness of the video stream. The number of connections if no renegotiation takes place is noc_{all}. The weight α determines how aggressive the algorithm segments the video stream. Our experiments showed that α=70 is a good trade-off between number of renegotiations and bandwidth requirements.
- fix number of renegotiations: With this method a fix number of renegotiations is allowed for a movie. This is very useful for comparing different schemes. Moreover in practical applications this makes sure that the signalling system of the switches is not overloaded. Our algorithm to calculate the renegotiation costs is based on approximations.

5.3 Optimizing algorithm O-RVBR

The algorithm to find the optimal renegotiation points is presented in Figure 3. The input for this algorithm is the list of frame sizes of the actual video. The output is a list of optimal renegotiation points.

The algorithm computes the traffic descriptor for the current segment S_i (i.e. the interval of frame i to frame i+length). Then, the bandwidth costs $k(S_i)$ and the renegotiation costs $r(S_i)$ for this segment are calculated. The costs for each segment plus the costs of the remaining segments up to the last frame are calculated. If this costs are the lowest for the current starting point, they are termed the minimal costs (cost(i)). This test is done for all possible segment sizes beginning from the current starting point (i). Then, the current starting point is moved. Note that this algorithm starts from the last frame and moves the current starting point into the direction of the first frame. The result of the algorithm is a chain of optimized renegotiation points for each start point (segment_len()).



N = number of frames in video
i = framenumber 0<=i<=N-1
S_i =Segment i with S={S_1, ..., S_R}
$r(S_i)$ = renegotiation costs (5.2)
$k(S_i)$ = bandwidth costs (5.1)
cost(i) = K(S') with S'= [S_i, $S_{i+segment\_len(i)}$, ...]
segment_len(i) = optimal length of segment starting at i

```
FOR( i=N-1; i>=1; i--)
        FOR (length=1; i+length <= N-1; length++)
                compute traffic descriptor for Segment S_i =[i, ..., i+length]
                cost = r(S_i) + k(S_i) + cost (i+length)
                if (costs < cost(i)
                        cost(i) = cost
                        segment_len(i) = length
        done
done
```

Figure 3: optimal off-line algorithm

In Figure 4 we show the rating B(S) for the video trailer depending on the number of segments (number of renegotiations +1) and the delay bound d_k. Increasing the number of renegotiations results in higher values of B(S). This is valid for all different delay bounds d_k. From this it gets obvious that with a given QoS requirement (d_k fix) a system without renegotiation (1 segment) reserves always more bandwidth (higher costs, B(S) lower) than a system with our renegotiation scheme. Moreover, it gets obvious that a high number of renegotiations is always superior to a lower number. The allowed number of renegotiations is unfortunately limited by the signalling system of the ATM switches [LakYa96]. To adjust the number of renegotiations our renegotiation cost functions r(S) are provided (chapter 5.2).

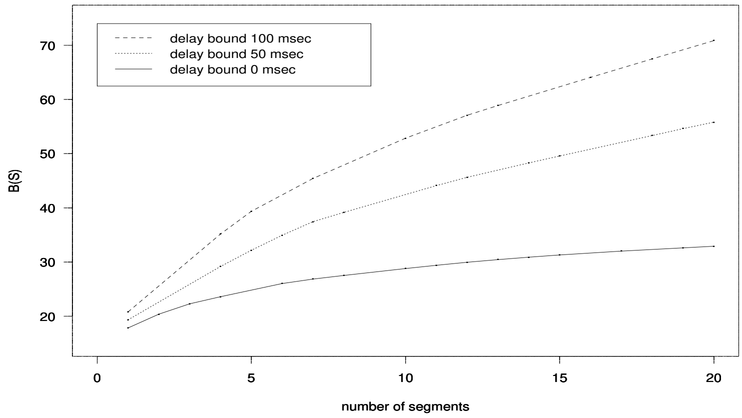


Figure 4: impact of numbers of renegotiation on B(S)

6 ACCELERATING THE ALGORITHM

The expense of the O-RVBR algorithm is $O(n^3)$ (Emgrunt,1996), which is a lot if movies with n= 162000 frames (90 minutes) are taken into consideration. There are several possibilities for speeding up the algorithm with little impact on the optimized points for renegotiation.We discuss the three schemes empirical envelope extrapolation, short time noc, mean estimation in the next subsections.

The scheme "short time noc" proved to be the most advantageous acceleration scheme since it lead to the highest run-time reduction and showed the lowest impact on the evaluation function B(S). We show that the combination of all three schemes further decreases the run-time.

6.1 Empirical envelope extrapolation

The operation with the highest expense in the inner loop of the optimal algorithm is the calculation of the empirical envelope. (This is part of computing the Traffic descriptor in Figure 3.) The expense for calculating the empirical envelope is $O(L_i^2)$ (L_i= segment length). By keeping the results from the last iteration step and checking only the difference introduced by the new frame the expense for calculating the empirical envelope is reduced to $O(L_i)$. This doesn't lead to any deviation from the optimal values and is therefore part of the optimal algorithm with expense $O(n^3)$.

The empirical envelope extrapolation is a scheme presented by Wrege (1996). The basic idea is to calculate the empirical envelope for a maximum interval length of k with k<= L_i. This reduces the expense from $O(L_i)$ to O(k). To get a complete envelope those first k parameters are repeated L_i/k times. Since the empirical envelope is subadditive this leads to a valid traffic constraint function. This traffic constraint function is not as tight as the empirical envelope which results in lower noc values and lower values of the evaluation function B(S). The expense of the algorithm is reduced to $O(k\ n^2)$ with Empirical envelope extrapolation.

6.2 Short time noc (k-noc)

This scheme basically reduces the number of runs through the inner loop by changing the end condition (N-1).

Our investigations showed that similar noc values for neighboring start points of segments in many cases result in the same endpoint for the optimal segment. Further it is obvious that the part of a segment with the highest bandwidth requirements is mainly responsible for the noc-value of the whole segment.

Based on our observations we developed a scheme called short-time noc. Short time noc calculates the empirical envelope from a given start point "a" of a segment to "a+k" (i.e. for the frames [a...a+k]). On the basis of this empirical envelope we determine the traffic descriptor and the noc value for this segment. This noc value is called $knoc_a$. The algorithm compares the value $knoc_a$ with the value $knoc_{a+1}$ and derives thereof the end condition of the inner loop. We distinguish three cases (Figure 5):

1. $knoc_a = knoc_{a+1}$: the segments are very similar in this case. Therefore the end condition of the inner loop is set to the optimal end e of the segment starting with a+1.
2. $knoc_a > knoc_{a+1}$: extending [a...a+k] to [a...a+k+1] will reduce $knoc_a$ to $knoc_{a+1}$. Therefore optimal endpoints won't differ much and the end condition is set identical to case 1.
3. $knoc_a = knoc_{a+1}$: in this case the next segment [i...i+k] (i>a) with $knoc_i < knoc_a$ is searched and the end condition of the inner loop is set to the optimal end of the segment starting with i.

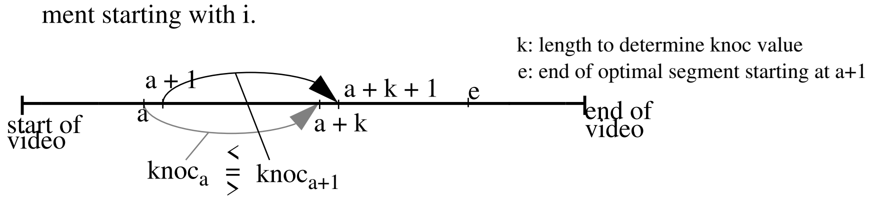


Figure 5: comparison of neighboring knoc values

6.3 Mean estimation

In contrast to the short time noc scheme this scheme accelerates the algorithm by completely skipping the inner loop if suitable. This is our only scheme that takes into account the frametype characteristics of MPEG streams. Note, that the other schemes work also for non MPEG VBR video streams.

Mean estimation calculates the mean of the frame sizes for each frametype (I,P,B) of the segment [a+1...e]. Frame a is compared to the mean of frames of the same type. If the difference of frame a's size to the mean is smaller than a threshold the segment [a+1...e] is extended to [a...e]. This is repeated with frames a-1, a-2, ... until the threshold is exceeded. Only in this case the inner loop is processed.

6.4 Results

We show in table 1 a run-time example for the O-RVBR algorithm with the different acceleration schemes. We used a MPEG2 coded version of the trailer from the video Terminator 2. The length of this video is 1800 frames (1 minute). The cost functions were chosen identical for all schemes and the computation was done at a PC with a 133MHZ Pentium processor under Linux 2.0.

Table 1: run-time for Terminator 2 trailer

	O-RVBR without acceleration	*Empirical envelope extrapolation*	*short time noc*	*mean estimation*
run-time	4h 23min	1h 34 min.	23 min.	1h 17 min.
expense	$O(n^3)$	$O(k\ n^2)$	$O(n^3)$	$O(n^3)$
B(S)	57,08	52,85	57,08	52,53
parameters		k=300	k=180	(I,P,B)=(10,20,40)cells

The results show that short time noc leads by far to the lowest run-time. At the same time short time noc has the highest values of the evaluation function B(S). This means that the renegotiation points found by this scheme are identical to the values found by O-RVBR. With appropriate parameters we can increase the evaluation function although for the other schemes, but this leads to an increased run-time. The value B(S) for k-noc is only decreased if the value for k is chosen extremely small (k = 30 => B(S)=56,25). Therefore k-noc is a good choice to reduce run-time. But keep in mind that reductions of B(S) are possible for certain videos.

To further reduce the run-time of the O-RVBR algorithm (especially for complete 90 minute videos) the acceleration schemes can be combined. Table 2 shows results for combinations of different schemes with the same parameters as in Table 1.

Table 2: combination of different schemes for video trailer

	Emp. Env. extr. + *mean estim.*	*short time noc* + *mean estim.*	*short time noc* + *Emp.Env. extr.*	*Combination of all three schemes*
run-time	43 min.	18 min.	13 min.	11 min.
B(S)	53,63	52,62	53,48	53,48

Considering the data of table 2 it becomes obvious that by combining the three schemes additional run-time savings are possible. The evaluation function B(S) shows results that are comparable to the single schemes mean estimation and empirical envelope extrapolation. But the combination leads to a much lower run-time, than for the single schemes mean estimation and empirical envelope extrapolation. If we consider run-time and B(S), the best results are achieved by the combination of all three schemes.

We conclude that the best choice is short time noc if the run-time of the optimal algorithm has to be reduced. If further reduction is necessary a combination of all three schemes is the best choice. For videos that are likely to be played very often we propose to go for the highest possible value of B(S) while the run-time reduction of the algorithm may be more convenient for videos that are requested not to often from the video server.

7 COMPARISON TO WELL KNOWN SCHEMES

This chapter compares our algorithm with the VBR off line algorithm for the DBIND model presented by Knightly et al. (1995) and with the CBR off-line algorithm RCBR of Grossglauser et al. (1995).

7.1 DBIND off-line Algorithm

The D-BIND model is a general traffic model developed by E.Knightly and H.Zhang (1994). We use their algorithm with the ATM traffic descriptor to make it comparable to our less general approach. We show that for this special usage with the ATM VBR service our algorithm leads to better results.

The DBIND off-line algorithm (Knightly, 1995) detects the renegotiation points much faster than our algorithm. The expense of this algorithm is O(n logn) compared to $O(n^3)$ or $O(k\ n^2)$ for our algorithm. On the other hand the renegotiation points of our algorithm lead to higher values of B(S) which is associated to reserving less bandwidth. Since the off-line algorithm is executed only once for a video, and the resulting renegotiation points are stored with the video on the video server our approach is advantageous for this class of applications.

The D-Bind algorithm finds the worst case part of the video that has the highest bandwidth requirement (Worst Case D-BIND parameter R_n). This part is increased to the right an left until the mean bandwidth of this part is lower than a threshold. This part is then termed a segment. The same procedure is repeated with the remaining parts of the video left and right of the segment.

For the DBIND algorithm we used the parameters I_n=1second and ψ is varied to get different numbers of segments. The tests were done with different videos. "Trailer" is a MPEG 2 video we captured on our own. The other videos are MPEG 1 traces kindly provided by the University of Würzburg (ftp-info3.informatik.uni-wuerzburg.de).Figure 6 and Figure 7 show that O-RVBR is superior to the D-BIND algorithm independent of the number of renegotiations (number of segments -1) and the delay bound d_k in the ATM switch.

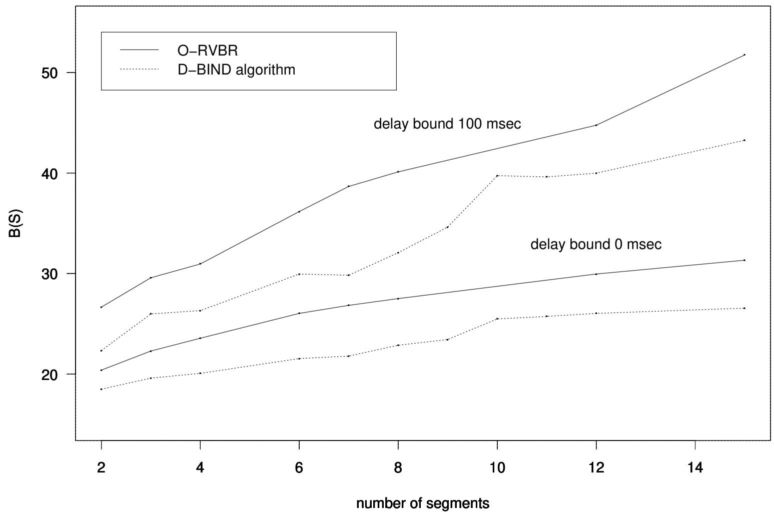


Figure 6: comparison of D-Bind and O-RVBR algorithm for video **trailer**

Figure 8 compares the O-RVBR algorithm and the D-Bind algorithm with a periodic scheme. The periodic scheme divides the video sequence into segments of equal size (length of segment = L_s / number of segments). It does not take the behavior of the video stream into consideration. In this comparison the value of B(S) is displayed for different delay bounds. The number of segments is fixed to 12. Here it gets obvious that the D-Bind algorithm does only little better than the periodic scheme and that O-RVBR is always better.

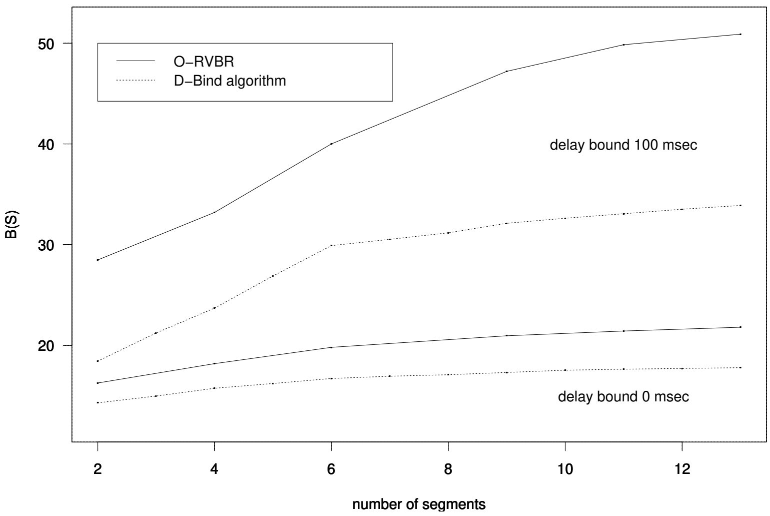


Figure 7: comparison of D-Bind and O-RVBR algorithm for video **asterix**.

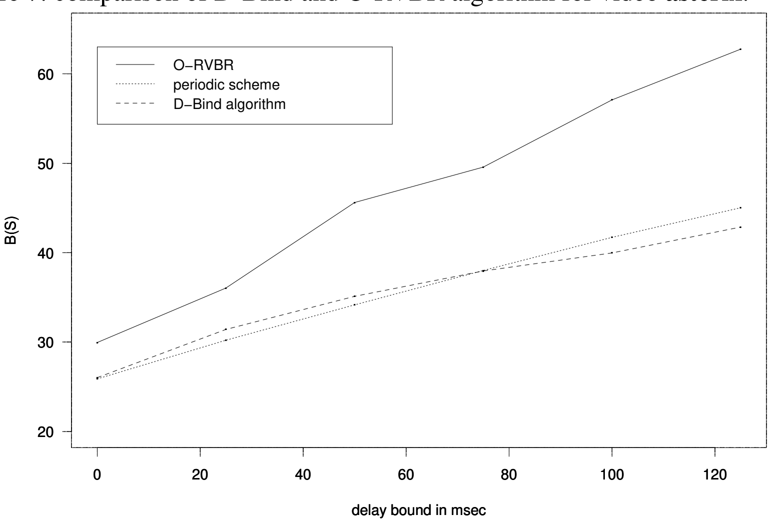


Figure 8: Comparison with periodic scheme for video **trailer**

7.2 RCBR algorithm

The off-line RCBR algorithm presented by Grossglauser et al. (1995) reserves a constant rate (PCR) for each segment. This is done under the requirement that there is no loss in a given smoothing buffer B. This results in transmission of VBR video using a renegotiated CBR ATM service (piecewise constant rate reservation).

Our goal is to compare the two algorithms by excluding the problems of comparing CBR versus VBR (Quality, delay, statistical multiplexing). Therefore, we chose the following approach.

- We use the same delay introduced by RCBR's smoothing buffer B for smoothing the input of our O-RVBR algorithm. We developed a shaping scheme for a renegotiation service. This shaping scheme is used instead of frame period averaging (chapter 4) for the comparison with the RCBR algorithm.
- The delay bound d_k equals zero to disable the multiplexing functionality for the VBR service. With d_k =0 only the PCR value is relevant for the traffic descriptor. Since SCR and BT have no impact the calculation of noc_i is reduced to noc_i = link rate /PCR_i

With this approach we compare the transmission of videos with equal quality and delay (QoS). The main difference are the algorithms for finding the renegotiation points. It has to be noted, that the run-time expense and the memory expense of the RCBR algorithm is O(K*B*N) with K=number of bandwidth levels, B= buffer size and N = length of video. The run-time expense is much lower than for our algorithm, but the memory expense may be critical since RCBR leads to memory demands of 10MByte for short examples with N=2000 frames (66 seconds), B=200 cells and K=20 levels.

For our investigations we use the same videos as for the comparison with the DBIND algorithm. The number of bandwidth levels is chosen to be K=20 as recommended in (Grossglauser, 1995).

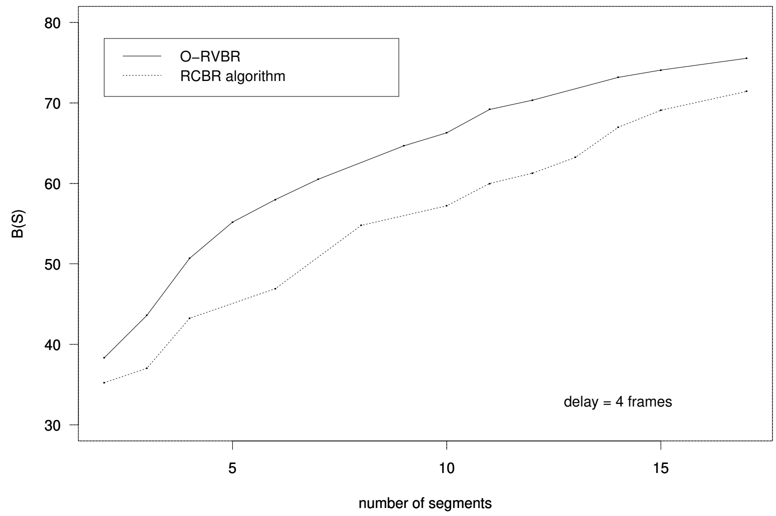


Figure 9: comparison of O-RVBR and RCBR for video **trailer**

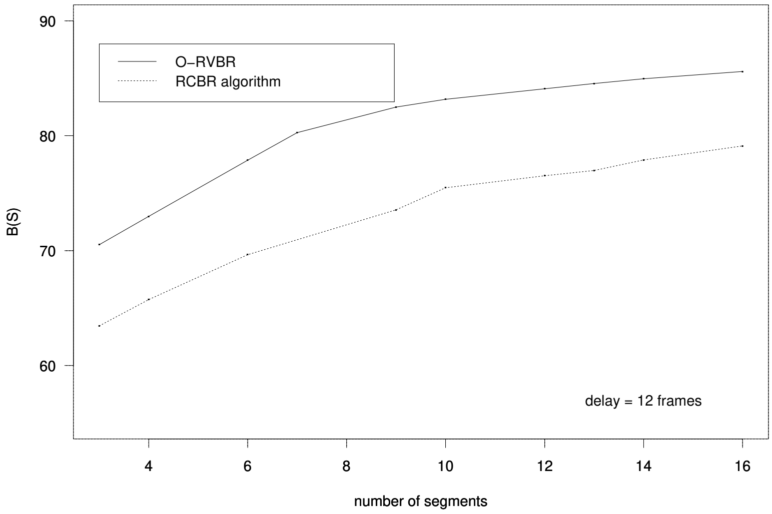


Figure 10: comparison of O-RVBR and RCBR for video **asterix**

We use different delays (4, 12 frame times) for the examples in Figure 9 and Figure 10. From these figures it gets obvious that O-RVBR leads to higher values of B(S) than the RCBR algorithm. The penalty for this advantage is the increased run-time. Again, the run-time is not very critical from our point of view, since the algorithm is started only once for each video. It stores the optimized renegotiation points together with the video on the sever. Moreover the RCBR algorithm is designed especially for CBR, and is not applicable for VBR services since all different combinations of PCR, SCR and BT are possible. This would lead to an unacceptable high value of bandwidth levels K.

8 CONCLUSION

In this paper, we presented a scheme for transmitting stored VBR video over ATM networks. The scheme is based on dynamic bandwidth allocation. We showed for a deterministic CAC scheme that dynamic bandwidth allocation increases the number of possible connections for a given link rate and delay bound considerably. This is made possible by reducing the inevitable over-reservation of static bandwidth allocation schemes.

We developed an algorithm (O-RCBR) that determines the optimized renegotiation points. This is done by minimizing a cost function based on costs for bandwidth and costs for renegotiations. Appropriate cost functions for a VBR service are presented. Since the optimizing algorithm results in high run-time expense we developed acceleration schemes. Especially the scheme k-noc proved to reduce the run-time considerably, while finding the same renegotiation points as the optimal scheme in most cases. The comparison of O-RVBR with existing off-line algorithms (Grossglauser, 1995), (Knightly, 1995) proved that O-RVBR leads to the lowest amount of reserved bandwidth. This advantage is obtained by using an algorithm with higher run-time. This is tolerable for the addressed application of a video server, since the algorithm runs only once per video. The resulting optimized renegotiation points are stored together with the video on the server.

9 REFERENCES

ATM Forum (1995), ATM User-Network Interface (UNI) Specification Version 3.1,

ATM Forum (1996), Traffic Managment Specification Version 4.0, February 96

Bansal V.,Siracusa R., Hearn J.,Ramamurthy G.,Raychaudhuri D.(1995), Adaptive QoS-Based API for ATM Networking, NOSSDAV 1995, Durham, New Hampshire

Chong S., Li S., Ghosh J. (1995), Predictive Dynamic Bandwidth allocation for Efficient Transport of Real Time VBR Video over ATM, IEEE JSAC, Vol 13, No1

Cruz R.L.(1991), A Calculus for Network Delay , Part I and II, IEEE Transactions on Information Theory, Vol37, No.1, January 1991

Cruz R.L. (1995), Quality of Service Guarantees in Virtual Circuit Switched Networks, IEEE Journal on Selected Areas in Communications Vol.13, No.6, August 95

El-Henaoui S., Coelho R., Tohme S. (1996), A Bandwidth Allocation Protocol for MPEG VBR TRaffic in ATM Networks, Infocom 96, San Francoisco

Emgrunt H. (1996), Dynamic bandwidth reservation for transmission of stored MPEG2 video, Diploma thesis, University of Karlsruhe, September 1996

Garret M.W., Willinger W. (1994), Analysis Modelling and Generation of Self Similar VBR Video Traffic, ACM SigComm, London, September 1994

Graf M. (1994), Traffic Shaping for VBR Video, Proceeding 4th OPen Workshop on High Speed Networks, Telecom Bretagne, Brest France

Grossglauser M., Keshav S., Tse D.(1995), RCBR: A Simple and Efficient Dervice for Multiple Time-Scale Traffic, ACM SIGCOMM 95

Gumbrich S., Sandvoss J., Comparison of Media Scaling Techniques for Real Time Packet Video Transport, Proc. of IEEE HPCS'95, Mystic Conneticut

ITU-T (1993) Recommendation Q.2931, B-ISDN Signalling System

ITU-T (1995), Recommendation Q.2963 (DRAFT), B-ISDN Connection Modification

Kaufman J.(1996), Status of ATM Forum Technical Work Groups, January, 1996

Knightly E., Zhang H.(1994), Traffic Characterization and Switch Utilization using a Det. Bounding Interval Dependent Traffic Model, U.C. Berkeley, TR-94-047

Knightly E., Zhang H.(1995), A New Approach to Support Delay-Sensitive VBR Video in Packet-Switched Networks, NOSSDAV 1995, Durham, New Hampshire

Knightly E.,Wrege D., Liebeherr J.,Zhang H.(1995), Fundamental Limits and Tradeoffs of providing det. guarantees to VBR Video traffic, ACM SIGMETRICS'95

LakshmanK., Yavatkar R.(1996), An Empirical Evaluation of Adaptive QoS Renegotiation in an ATM Network, NOSSDAV 1996, Zushi, Japan

Lam S., Chow S., Yau D. (1994), An Algorithm for Lossless Smoothing of MPEG Video, SIGCOMM 94, London England

Liebeherr J.,Wrege D.,Ferrari D.(1994), Exact Admission Control for Networks with bounded Delay Service, University of Virginia, Technical Report CS-94-92

Mark B., Ramamurthy G.(1996), Real Time Estimation and Dynamic Renegotiation of UPC Paramaetrs for Arbitrary Traffic Sources over ATM, 7th International Workshop on Packet Video, Victoria, Australia

Pancha P., Zarki M.(1993), Bandwidth Requirements of Variable Bit Rate MPEG Sources in ATM Networks, Modelling and Performance Evaluation of ATM Technology, ,Elsevier Science Publishers

Rathgeb E. (1993), Policing of realistic VBR Video Traffic in an ATM network, International Journal of Digital and analog Communication systems. Vol. 6, 213-226, 1993

ReiningerD., Ramamurthy G., Raychaudhuri D.(1995), VBR MPEG Video Coding with Dynamic Bandwith Renegotiation, IEEE International Conference on Comm.

Wrege D.,Liebherr J.(1996), Video Traffic Characterization for Multimedia Networks with Deterministic Service, IEEE Infocom 96

10 BIOGRAPHY

Stefan Gumbrich received his diploma degree in Electrical and Computer Engineering from the Fachhochschule Bingen and the University of Tennessee, Knoxville in 1991. In 1994 he received his diploma degree in Computer Science from the University of Paderborn. Since 1994 he has been a guest scientis at IBM European Networking Center in Heidelberg (Germany). His main research interests are the aspects of video transmission over high speed networks and the design and implementation of high speed network protocols.

Heinrich Emgrunt received his Diplom in Computer Science from the University of Karlsruhe in 1996. The subject of his diploma thesis was the transmission of stored video over ATM-networks.

Torsten Braun received his diploma degree and his doctoral degree in computer science from the University of Karslruhe (Germany) in 1990 and 1993,respectively. From 1990 to 1994 he was a research assistant at the Institute of Telematics, University of Karlsruhe. From 1994 to 1995 he was a visiting scientist at the Institut National de Recherche en Informatique et en Automatique (INRIA) in Sophia-Antipolis (France). Since September 1995, he has been a guest scientist at the IBM European Networking Center (ENC) in Heidelberg (Germany). His main research interests are the implementation of transport systems for multimedia communication and the integration of Internet protocols and ATM.

9 REFERENCES

(ATM Forum,1995) ATM User-Network Interface (UNI) Specification Version 3.1,ATM Forum, 1995

(Rathgeb, 1993) *E.Rathgeb, Policing of realistic VBR Video Traffic in an ATM network International Journal of Digital and analog Communication systems. Vol. 6 1993*

(Garret,1994) *M.W. Garret, W.Willinger, Analysis Modelling and Generation of Self Similar VBR Video Traffic, ACM SigComm, London, Sept 1994*

(Grossglauser, 1995) *M. Grossglauser, S.Keshav, D.Tse, RCBR: A Simple and Efficient Dervice for Multiple Time-Scale Traffic, ACM Sigcomm 95, August 95*

(Lam, 1994) *S.Lam, S. Chow, D.Yau, An Algorithm for Lossless Smoothing of MPEG Video, SIGCOMM 94 -8/94 London England*

(Cruz,1991) *R.L.Cruz, A Calculus for Network Delay , Part I and II, IEEE Transactions on Information Theory, Vol37, No.1, January 1991*

(Cruz,1995) *R.L.Cruz, Quality of Service Guarantees in Virtual Circuit Switched Networks, IEEE Journal on Selected Areas in Communications Vol.13, No.6, August 95*

(Knightly,1995b) *E.Knightly, D.Wrege,, J.Liebeherr,H.Zhang, Fundamental Limits and Tradeoffs of providing deterministic guarantees to VBR Video traffic, ACM Sigmetrics'95, Otawa, Canada, May 95*

(Wrege,1996) *D.Wrege, J.Liebherr, Video Traffic Characterization for Multimedia Networks with Deterministic Service, to appear in Proc. IEEE Infocom 96*

(Graf,1994) *Marcel Graf, Traffic Shaping for VBR Video, Proceeding 4th OPen Workshop on High Speed Networks, Telecom Bretagne, Brest France 1994*

(ITU-T,1995) *ITU-T Draft Standard Q.2963. Preliminary Draft, ITU-T, 1995*

(ITU-T,1993) *ITU-T Recommendation Q.2931, ITU-T, 1993*

(ATM Forum, 1996) ATM Forum Traffic Managment Specification Version 4.0, ATM Forum/95-0013, February 96

(Kaufman,1996) *Jill Kaufman, Status of ATM Forum Technical Work Groups, January, 1996*

(Chong,1995) *S. Chong, S.Li, J.Ghosh, Predictive Dynamic Bandwidth allocation for Efficient Transport of Real Time VBR Video over ATM, IEEE Journal on Selected Areas in Communication, Vol 13, No1, Jan.1995*

(Pancha, 1993) *P.Pancha, M.Zarki, Bandwidth Requirements of Variable Bit Rate MPEG Sources in ATM Networks, Modelling and Performance Evaluation of ATM Technology, ,Elsevier Science Publishers*

(Knightly, 1995) *E.Knightly, H.Zhang, A New Approach to Support Delay-Sensitive VBR Video in Packet-Switched Networks, NOSSDAV 1995, Durham,* *New Hampshire, 18-21 Apr.1995*

(Reininger,1995) *D.Reininger, G.Ramamurthy, D.Raychaudhuri, VBR MPEG Video Coding with Dynamic Bandwith Renegotiation, IEEE International Conference on Communications 1995, NJ,USA,p1773-1777*

(Mark, 1996) *B.Mark, G.Ramamurthy, Real Time Estimation and Dynamic Renegotiation of UPC Paramaetrs for Arbitrary Traffic Sources over ATM, 7th International Workshop on Packet Video, March 1996, Victoria, Australia*

(Bansal, 1995) *V.Bansal,R.Siracusa,J.Hearn,G.Ramamurthy,D.Raychaudhuri, Adaptive QoS-Based API for ATM Networking, NOSSDAV 1995, Durham, New Hampshire, 18-21 Apr.1995*

(El-Henaoui, 1996) *S. El-Henaoui, R.Coelho, S.Tohme, A Bandwidth Allocation Protocol for MPEG VBR TRaffic in ATM Networks, Infocom 96, San Francoisco, March 96*

(Knightly,1994) *E.Knightly, H.Zhang, Traffic Characterization and Switch Utilization using a Det. Bounding Interval Dependent Traffic Model, U.C. Berkeley School of Computer Science , TR-94-047*

(Liebeherr,1994) *J.Liebeherr,D.Wrege,D.Ferrari, Exact Admission Control for Networks with bounded Delay Service, University of Virginia, Technical Report CS-94-92*

(Gumbrich, 1995) *S.Gumbrich, J.Sandvoss, Comparison of Media Scaling Techniques for Real Time Packet Video Transport, Proc. of IEEE HPCS ,Aug.23-25 1995, Mystic Conneticut*

(Lakshman, 1996) *K.Lakshman, R.Yavatkar, An Empirical Evaluation of Adaptive QoS Renegotiation in an ATM Network, NOSSDAV 1996, Zushi, Japan*

(Emgrunt,1996) *Heinrich Emgrunt, Dynamic bandwidth reservation for transmission of stored MPEG2 video, Diploma thesis, University of Karlsruhe, 1996*