A BANDWIDTH ASSIGNMENT ALGORITHM ON A SATELLITE CHANNEL FOR VBR TRAFFIC*

Nedo Celandroni, Marco Conti, Erina Ferro,
Enrico Gregori, Francesco Potortì

CNUCE, Institute of National Research Council (C.N.R.)
Via S. Maria 36 - 56126 Pisa - Italy
Phone: +39-50-593111
Fax: +39-50-904052 - Telex: 500371
E-mail: {N.Celandroni, M.Conti, E.Ferro, E.Gregori, F.Potorti}@cnuce.cnr.it

SUMMARY

Variable Bit Rate (*VBR*) video is currently by far the most interesting and challenging real-time application. A VBR encoder attempts to keep the quality of video output constant, and at the same time reduces bandwidth requirements since only a minimum amount of information has to be transferred. On the other hand, as VBR video traffic is both highly variable and delay sensitive, high-speed networks (e.g. ATM) are generally implemented by assigning peak rate bandwidths to VBR video applications. This approach may however be inefficient in a satellite network based on a TDMA scheme. To overcome this problem, we have designed a demand-assignment satellite bandwidth allocation algorithm in TDMA, named V2L-DA (VBR 2 Level-Demand Assignment), which manages the VBR video traffic according to a dynamic bandwidth allocation algorithm. In this paper we discuss how to tune the proposed algorithm in order to optimise network utilisation when MPEG-1 VBR video traffic is being transmitted. Our results indicate that most of the time only 40% of the peak rate bandwidth is needed to satisfy the VBR source, and the remaining 60% of the peak rate bandwidth can be used to transmit the datagram traffic queued in the network stations.

Keywords: satellite, TDMA assignment, real-time traffic, non real-time traffic, VBR traffic, MPEG coding, traffic model

1. Introduction

A variety of new applications, such as the transport of pictures, teleconferencing, video, and a large volume of interactive computer data must be supported in an integrated

* Work carried out in the framework of the Italian co-ordinated project "*Advanced applications for next generation packet-switching networks*" funded by the National Research Council (C.N.R.)

manner by today's high speed networks. These applications have diversified quality of service (QoS) requirements and traffic statistics (ranging from the high burstiness of video applications to the smooth continuous traffic generated by large file transfers).

Variable Bit Rate (VBR) video is currently by far the most interesting and challenging real-time application. A VBR encoder attempts to keep the quality of video output constant and, at the same time, reduces bandwidth requirements, since only a minimum amount of information has to be transferred.

As VBR video traffic is both delay sensitive and has a high degree of burstiness, it is commonly believed that a bandwidth corresponding to the source peak-rate must be reserved for this application to satisfy its QoS requirements. In our TDMA satellite network the peak-rate allocation is extremely inefficient, because only the station reserving the bandwidth is authorised to utilise it. Taking into consideration that the ratio between the peak and the average bit rate for VBR video is generally high (for example, in the MPEG 1 movie used in this paper it is equal to five, see Section 3.2), this implies that a large portion of the network bandwidth remains unused unless the station transmitting the movie has enough low-priority traffic as well. To increase the efficiency in bandwidth allocation, we designed an algorithm which dynamically allocates bandwidth to VBR video on the basis of the actual source rate. This algorithm is integrated into a *centralised control* demand-assignment satellite access scheme, named V2L-DA (VBR 2-Level Demand Assignment). The results presented here show that V2L-DA can efficiently and simultaneously support two classes of traffic, called *datagram* and *stream,* respectively.

According to the traffic categories as defined in the ATM Forum TM4.0 ("ATM service categories") [12], the first class includes all the jitter-tolerant applications (unspecified bit rate, UBR , and available bit rate, *ABR*, service categories), while the second includes all the real-time applications (constant bit rate, CBR, and variable bit rate, VBR, service categories). This paper focuses on the efficiency of the V2L-DA when the stream service is used to transmit the VBR traffic, though the allocation scheme is also suitable for CBR traffic.

The paper is organised as follows. Section 2 presents the satellite bandwidth allocation scheme together with the criteria to dimension the buffers needed to compensate for the jitter of VBR data. Section 3 describes the MPEG traffic characteristics and its model, and discusses the tuning of the parameters used in the allocation algorithm. The tuning study is based on an MPEG-encoded Star Wars movie. Section 4 concludes the paper.

2. The stream bandwidth allocation on the satellite channel

The satellite bandwidth allocation policy which we studied has a centralised control. The channel dispatcher function is assumed by a control station which is responsible for allocating the bandwidth on the basis of the requests received from the traffic stations (requesters), where the traffic from various applications is collected. A distributed control algorithm would save the time needed to send the assignment request to the control station (one round trip time). On the other hand, it would be more complex, because of the need to prevent any possible misunderstanding among the stations from provoking collisions, or stopping the transmissions, thus paralysing the whole network [14, 15].

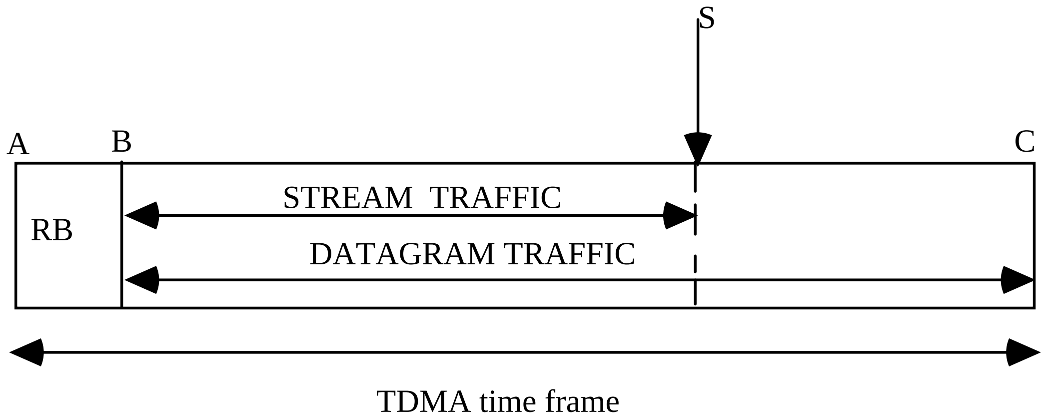


Figure 1. TDMA frame structure in V2L-DA

The TDMA frame structure used in V2L-DA is shown in Figure 1. RB is the reference burst sent by the channel dispatcher for allocations and synchronisation. The stream traffic cannot go over a fixed boundary S in the frame; the rest of the frame is devoted to datagram traffic. On the other hand, datagram traffic can temporarily expand in the TDMA frame if the stream traffic does not reach the S boundary. In Figure 1 the space in the frame devoted to the datagram traffic is expressed by $C - S + \varepsilon\ (S - B)$, where $0 \le \varepsilon \le 1$ is the fraction of stream allocation unused by the stream traffic.

The allocation algorithm for datagram traffic is not addressed in this paper; one of the policies proposed in the literature [2, 3, 4, 5, 6] is assumed to have been adopted.

A new request for a stream allocation is accepted by the channel dispatcher if and only if the sum of the new request plus the current stream allocations does not exceed the boundary S. A couple of parameters (R, P), where $R \le P$, are used by a stream application to declare two values of its throughput to the relative earth station: R, chosen as explained below, is the basic throughput required, while P is the booked rate. The earth station sends the request to the channel dispatcher to get the bandwidth allocation. The values of parameters R and P depend on the type of stream traffic that generates the request. If $R = P$, a request for CBR traffic is issued, which is easy to manage. If accepted, a set of time slots corresponding to P is reserved in all the time frames, until released. If $R < P$, a request for VBR traffic is issued, which represents our main interest. In fact, hereafter we will focus on the handling of VBR traffic. The channel dispatcher tries to book a bandwidth P. If it succeeds, in each frame a time slot corresponding to a bandwidth *R* is allocated to the requester, while still maintaining the booking for the *P* allocation. The channel dispatcher grants the booking for *P* only if the sum of all the bookings does not exceed the S boundary. Due to the variability in traffic intensity, the bandwidth allocated may not be completely used in every frame. The unused bandwidth can be employed by the requester to transfer its datagram data, if any is queued in the station (Figure 2).

Hereafter *P* will be referred to as the *booked* bandwidth, while we speak of *allocated* bandwidth when referring to the current value (either *P* or *R*) which is given to the station for its exclusive use. The allocated bandwidth is granted to a station by the master upon that station's request.

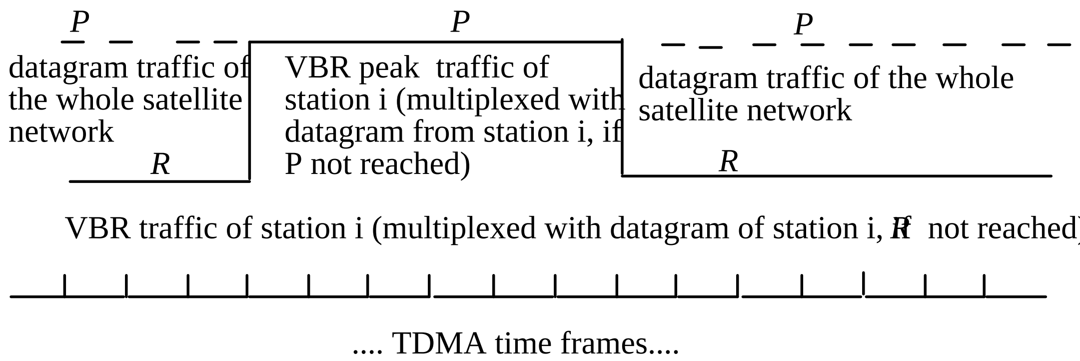


Figure 2. Relationship between the *booked* bandwidth *P* (dashed line) and the *allocated* bandwidth, either *R* or *P* (solid line).

When the VBR traffic fills the station's input queue beyond a given threshold T_u the station sends a request to the channel dispatcher to use the extra bandwidth *P* - *R* which it has booked beforehand. This request is needed because, when the bandwidth *P* - *R* is not used by the booking station, it is assigned by the channel dispatcher to all the active stations that require datagram allocations. When the station is allocated a bandwidth *R*, its input queue is empty as long as the throughput of the input traffic does not exceed *R*. We assume that the input traffic is smoothed by using a pre-buffering after the MPEG coder, so the traffic at the output of the MPEG source has a constant rate for the whole duration of a Group of Pictures (GOP, that is 12 frames, which corresponds to 0.5 seconds for a movie — see Section 3.1). As a side effect, the smoother delays the video stream by the time length of a GOP. After exiting the MPEG source, packets arrive at the input station after having crossed a network that may be quite complex, so they are generally affected by time jitter. In order to choose T_u in such a way as to avoid more than one switch per GOP from R to P, the threshold T_u must be big enough to absorb a spurious burst of input traffic. We use a threshold on the queue length rather than on the bandwidth to avoid a limiting case in which the latter would fail. In fact, suppose we choose a threshold $T > R$ in order to provide histeresys. A jitter-free input traffic with throughput higher than R but less than T would thus not cause any request for the high allocation level P, and would eventually fill the input buffer, whatever its size.

When the *throughput* of the traffic generated by the application goes below R the bandwidth P - R is released, and the station continues to use the bandwidth R, while still maintaining the booking for P. All the allocation is eventually released only on explicit request of the station or when the station itself is declared dead.

The input queue of the station never grows when the station is granted the high allocation level P, and once empty it remains empty. For this reason, in order to release the extra bandwidth, we use a threshold T_d based on a *negative virtual queue*, whose mechanism we will detail below. The choice of T_d must provide for a histeresys sufficient to avoid false triggers, that is, to avoid jumping to the low allocation level R because of the time jitter of input packets.

2.1 Video data delay and buffers

To analyse more closely the behaviour of this allocation scheme, we refer to Figure 3, using the subscripts i, o, r when referring to the input (station or link), output (station), and receiver (MPEG or link), respectively. We neglect the effect that the framing of the satellite channel has on the buffers' size and the data delays.

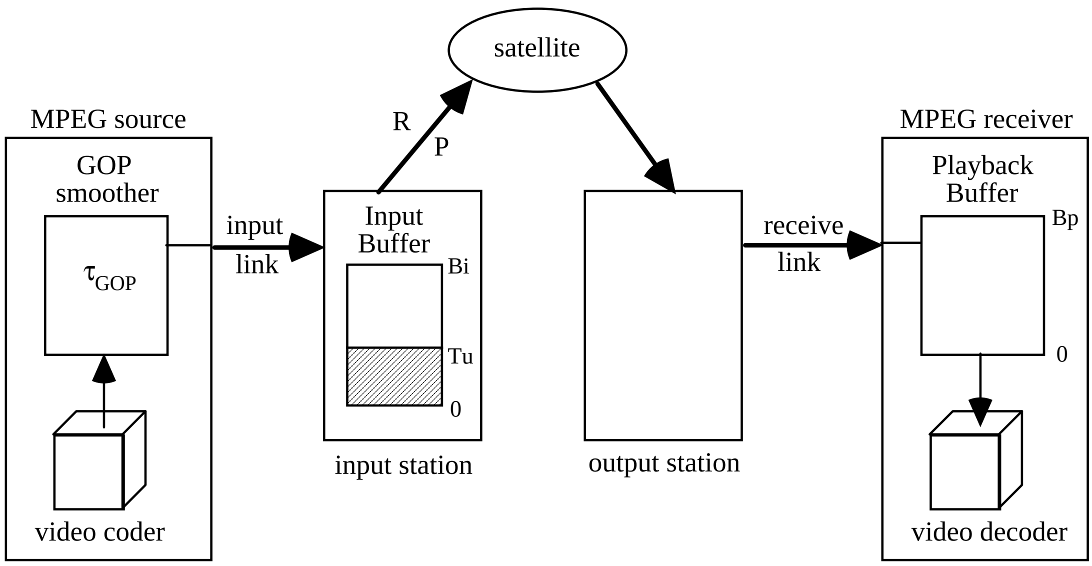


Figure 3. The communication chain.

When the current allocation is at the low level R, the station continuously monitors the length of its input queue, which is normally empty, and issues a request for the high allocation level P when the threshold T_u is exceeded.

We assume that there is a resource reservation on the input link which guarantees a bounded packet delay, and we call J_i the maximum jitter of the packets entering the input station, that is, the difference between their maximum and minimum delay. In order to avoid a false trigger of the allocation request for P, we must set

$$T_u = J_i \;\; R. \qquad (1)$$

On receiving the station’s request, the dispatcher grants the new allocation P. For a satellite network with centralised control, the delay D_a between request sending and allocation receiving is equal to 2τ. While waiting for its request to be granted, the station queues the excess input traffic in an input buffer, whose size must then be

$$B_i \;=\; T_u + D_a(P - R) + J_i P. \qquad (2)$$

The last term in (2) accounts for traffic bursts due to input jitter when the input traffic rate is P (see Figure 4).

Packets coming out of the output station experience variable delays in traversing the satellite link, from a minimum of τ to τ plus the maximum queuing time at the input station. The maximum jitter J_o of the packets exiting the output station is then equal to the maximum queuing time at the input station, that is (see Figure 4)

$$J_o = \frac{B_i}{P}. \qquad (3)$$

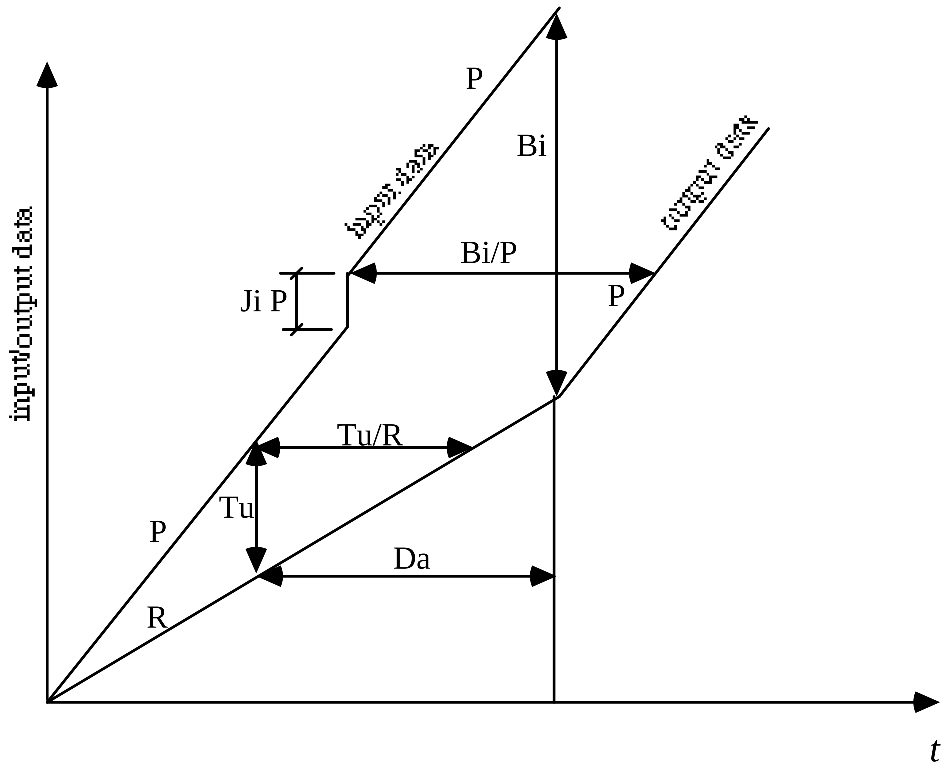


Figure 4. Buffer requirements and queuing times relationships in the transition from allocation R to allocation P (worst case).

The receiving MPEG application should begin playing the video after a time interval D_p, which is called the *playback delay*. D_p is equal to the maximum jitter of the arriving packets, that is

$$D_p = J_o + J_r, \qquad (4)$$

where J_r is the packet jitter caused by the receive link. The size of the *playback buffer* must accommodate twice the data received at maximum throughput during the playback delay:

$$B_p = 2\, D_p\, P. \qquad (5)$$

The factor 2 in (5) is necessary because the delay of the first packet received could be anything between 0 and D_p, since we want to provide for a receiver that can be turned on when the transmission has already begun.

The transmission delay D_t experienced by packets from the output of the MPEG coder to the input of the video decoder is evaluated as

$$D_t = \tau_{GOP} + \tau_i + J_o + \tau + \tau_r + J_r + D_p, \qquad (6)$$

where τ_{GOP} is the GOP duration time, and τ_i and τ_r are the latencies of the input link and receive link, respectively. To account for the image delay, from the moment a picture is taken to when it is shown, one should also add the coding and decoding delays of MPEG. Also, delays induced by framing and packetisation should be accounted for in (6) if a detailed estimation of the end-to-end delay is required.

We have followed the data path from the MPEG source to the MPEG receiver, and have computed the overall link delay taking into account what happens when the allocation switches from R to P. Now we consider the opposite switch, which happens when the current allocation is at the high level P. In this state, the station maintains a *virtual input queue*, that is, a counter which is incremented at the rate of the input traffic and decremented at rate R. The counter is never incremented above 0, so it always contains a nonpositive number. When the input traffic has a rate greater than R, the virtual queue is 0. When the virtual queue drops below the threshold $-T_d$, the station issues a request for the low allocation level R. This mechanism is specular with respect to the one used for switching from allocation R to allocation P, so the threshold is computed with the same criterion, and

$$T_d = J_i R. \qquad (7)$$

3. Tuning the parameters of the allocation algorithm for MPEG applications

Below we study how parameters P and R must be set to optimise the utilisation of the satellite network capacity when the stream traffic is an MPEG-1 encoded movie.

3.1. MPEG-1 traffic characteristics

An uncompressed video source may generate bits at rates as high as hundreds of Mbps. Data compression techniques are therefore used to reduce the video-source bit rate which is transmitted over the network.

MPEG-1 is a specification for coding video, developed by the ISO Joint Motion Pictures Experts Group [7, 13]. The standard is well suited for a large range of video applications at a variety of bit rates. Compression of a combination of video and audio information, particularly for “movie” applications, is also possible. Typical compression ratios are in the range of 50:1 to 200:1 [8].

MPEG-1 is an interframe coder. Coders in this class exploit, in addition to intraframe coding, the temporal redundancy that exists between adjacent frames by predicting the next frame from the current one. A key feature that distinguishes MPEG-1 from previous coding algorithms is bi-directional temporal prediction. For this type of prediction, some of the video frames are encoded using two reference frames, one in the past and one in the future, which leads to higher compression gains.

As indicated above, when applying MPEG-1 to video, one of three different coding modes can be used for each frame. The terminology used for the resulting frame is related to the model used as follows:

- I-frame: intra frame coded,
- P-frame: predictive coded with reference to the previous P or I frame,
- B-frame: bi-directional predictive coded.

I-frames provide access points for random access but only with moderate compression. Predictive coded frames are generally also used as a reference for future P-frames. The frames of type B provide the highest amount of compression but require both a past and future reference prediction.

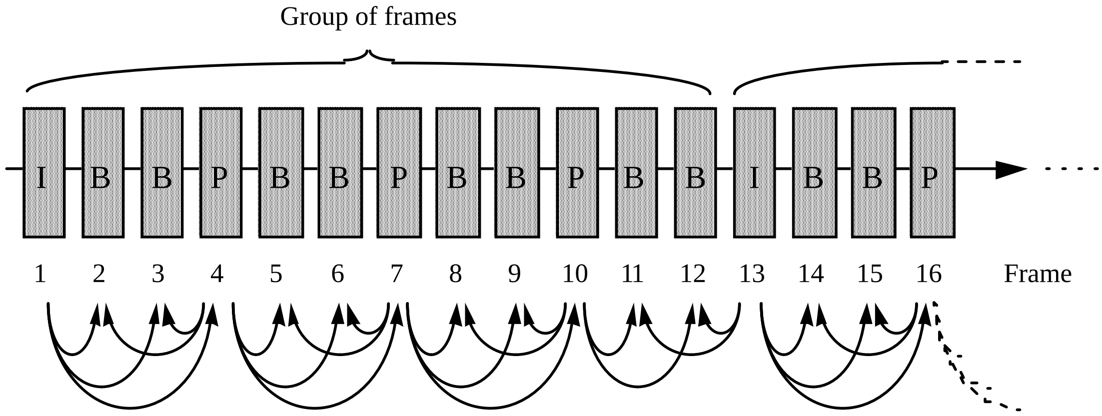


Figure 5. A sequence of MPEG-1 video frames and their relationship

In the encoded sequence, the frames are arranged into groups, as shown in Figure 5. In this case a group consists of 12 frames - one I-frame, three P-frames and eight B-frames. Figure 5 also shows the relationship between the frames. We can see that I-frames are independent, P-frames are predicted, and B-frames bidirectionally predicted.

Figure 6 shows a small extraction from the output of the MPEG-1 coded Star-Wars movie released by M. Garret at Bellcore. Specifically, frames are coded into GOPs as defined in Figure 5 (i.e. the frame pattern is IBBPBBPBBPBB).
As shown in Figure 6 the bandwidth required to transmit consecutive frames is highly variable and very much depends on the frame types, I, P and B. Furthermore, as expected (due to the coding scheme algorithm), the shape of the output is repeated every twelve frames.

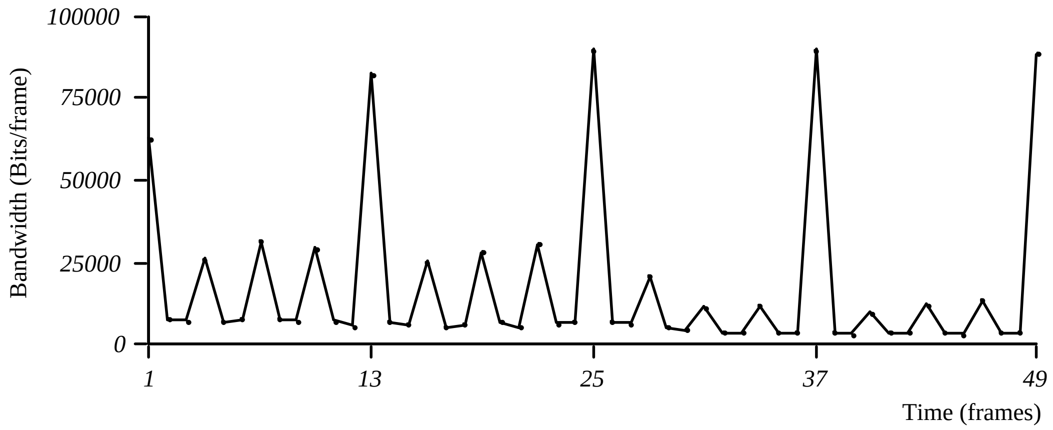


Figure 6. Part of the MPEG-1 coder trace, revealing group length and frame pattern.

To simplify the study of the bandwidth allocation algorithm we assume a twelve frames pre-buffering before the transmission, and we only look at the aggregate bit rate produced by the coding of a group of twelve frames. Thus, hereafter, we only consider the aggregate sequence obtained by summing the amount of bits generated in every GOP. This aggregate sequence has a period of twelve frames. In Table 1 the basic statistics of the MPEG-1 Star Wars aggregate sequence are presented.
Below we present a model developed to characterise the aggregate sequence obtained by the output of an MPEG-1 codec. More details on the modelling of an MPEG-1 video source can be found in [11]. The model is used as a synthetic traffic descriptor for analysing the performance of various bandwidth allocation schemes without requiring the huge amount of data describing the actual traces.

3.2. The model

The analysis presented in [11] shows that in the aggregate sequence there is both a short range dependency which lasts for a small number of groups (15 seconds), and long range dependencies which last for thousands of groups (10-20 minutes). To capture both types of dependencies a bidimensional Markov chain $\{L_k, H_k | k \ge 0\}$ is used. H_k is the k-th GOP size and L_k is the status of a low frequency process modulating the k-th GOP size.

$\{H_k | k \ge 0\}$ describes the bit rate per group of an MPEG encoder. To avoid unnecessary complexity (in the state space of $\{H_k | k \ge 0\}$) we quantize in a uniform way the bit rate into a number of levels. The number of quantization levels for the process will hereafter be denoted by N, i.e. $H_k \in \{0,1,\ldots,N-1\}$. Specifically, let *max* and *min* denote the maximum and minimum bit rates observed in the aggregate sequence, the possible bit rates are quantized with a constant step size $\Delta = (max - min)/N$. By applying this quantization procedure the average bit rate associated to $H = i$ is

$$min + (i+1) \cdot \Delta . \qquad (8)$$

The *min* and *max* values are reported in Table 1 together with the average, μ, and standard deviation, σ, of the GOP size.

Table 1: Star Wars basic statistics in Kbits

	μ	σ	*min*	*max*
GOP statistics	187.2	72.5	77.754	932.71

We use the GOP as the time unit. To represent the low-frequency component of our source, a modulating process $\{L_k | k \ge 0\}$ is included in the model as well $(L_k \in \{0,1,2,\ldots,M-1\})$. In the trajectories of the Markov chain, the H_k value frequently changes (every few time units, on average) while the L_k value changes on a much longer time scale (about 70-100 time units).

The transition probabilities of the Markov chain $\{L_k, H_k | k \ge 0\}$ are estimated from the MPEG 1 Star Wars trace by applying the procedure presented in [11].

Specifically, the model used for results presented in this paper is obtained with parameters M=8 and N=8. The accuracy of this model was investigated in [11]. The results obtained indicate that both the qualitative properties (i.e. burstiness and overall appearance of the traces) and the statistical properties (maximum, minimum, average, standard deviation and autocorrelation function) of the GOP size sequence generated with our Markov model are very similar to that of the real trace.

4. Tuning the allocation algorithm: a case study based on MPEG-encoded StarWars movie

In this section we study the setting of the P and R parameters to transmit the MPEG 1 "Star Wars" movie on the satellite link, together with some low-priority data. The choice of the P and R values is made by minimising the bandwidth allocation cost.

Minimising the end-to-end delay is not a primary target since the delays already caused by the MPEG coding, the pre-buffering and the satellite transmission make this technology unsuitable for interactive video applications. As described in Section 3, we divided the source bit rate into eight levels (i.e. H=0,.,7). However, as shown in [11], states with H equal to 5, 6 and 7 are rare (i.e. they only occur a few times in the two-hour sequence, and never consecutively), thus we do not consider these states for bandwidth allocation. In fact, a two-level allocation scheme (like the one presented here) would be extremely inefficient if P were set equal to the highest bandwidth level (i.e. H=7). Hereafter it is therefore assumed that when a group with throughput higher than H=4 occurs, the traffic exceeding the allocated bandwidth is transmitted with a best effort policy (e.g. as datagram traffic).

The bandwidth allocation problem, as shown in Table 2, is thus reduced to the study of four cases. P is set to the bit rate corresponding to H=4, while the R parameter can be set equal to the bit rate corresponding to one of the states 0, 1, 2, 3.

Case	*Allocation Level Positions*
1	R=0 and P=4
2	R=1 and P=4
3	R=2 and P=4
4	R=3 and P=4

Table 2. Cases of allocation positions

To identify the optimal allocation parameters setting we evaluate the cost of transmitting a VBR video source and some EDP data traffic. We assume a cost equal to 1 for each unit of bandwidth allocated to the VBR traffic. The cost for each unit of bandwidth reserved for the low-priority traffic is assumed to be less than 1 and will be denoted by β $(0 < \beta < 1)$.

As stated in Section 2, the unused stream bandwidth (i.e. booked for VBR traffic but not used for it) can be used to transfer datagram data. The maximum amount of such data is U, which is the difference between the peak and average bandwidth of a VBR video:

$$U = \sum_i \pi_i (P - i)$$

where π_i is the probability of a bit rate *i* for the VBR source.

When studying the bandwidth allocation we considered U as the maximum amount of data that could be transmitted by the station. The bandwidth exceeding U always needs to be allocated as a datagram bandwidth for all possible parameter settings (see Table 2), and thus its cost does not depend on the bandwidth allocation strategy. Hence, we assume that the station has to transfer a percentage *p* of U, and $p \cdot U$ is the amount of data traffic.

To simplify the presentation, in the computation of the bandwidth allocation costs we first focus on an ideal case in which the transients for switching between the two allocation levels (*P* and *R*) are negligible, i.e. when the bit rate is greater/lower than R the allocated level is *P*/*R*.

Under this assumption, a station with an MPEG-1 video source and $p \cdot U$ data traffic has to pay the following two costs for the required bandwidth.

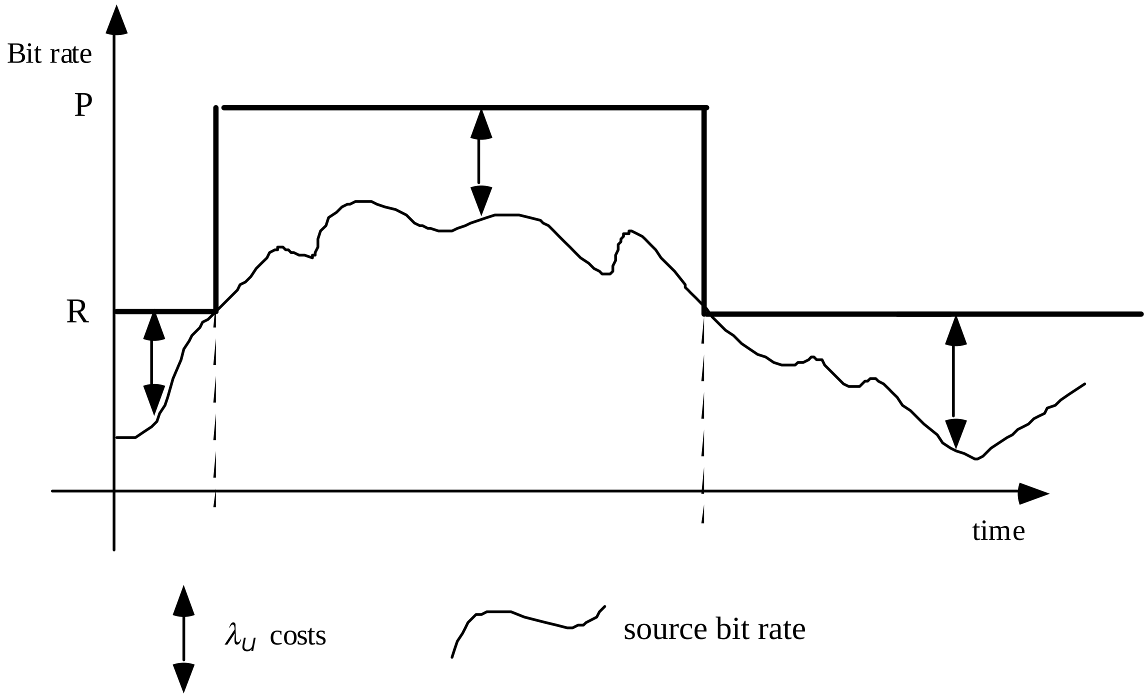


Figure 7. Allocation costs in the ideal case.

1) *Excess Stream cost.* This cost takes into account only the real-time bandwidth that is not used by video. As shown in Figure 7 this cost is the difference between the allocated bandwidth (bold line) and the source bit rate.

The amount of bandwidth used for transmitting the MPEG-1 traffic is obviously the same in all the allocation cases (see Table 2). Unused video bandwidth can be used by the same station to transfer its datagram data (if any), but it is allocated as stream bandwidth and thus it has a cost equal to 1. The excess stream cost is therefore given by

$$\lambda_u = \sum_{i=0}^{4} \pi_i \big(A(i) - i\big), \quad A(i) = \begin{cases} R & i \le \mathrm{R} \\ P & i > \mathrm{R} \end{cases} \tag{9}$$

where

π_i is the probability of bandwidth level *i*,

$A(i)$ is the allocated level.

The π_i probability is computed from the Markov chain which characterises the source (see previous section) and it is equal to

$$\pi_i = \sum_l P\{L_k = l, H_k = i\}.$$

2) *Requested Data Bandwidth Cost.* A portion of the unused bandwidth in each TDMA frame (see Figure 1), i.e. the datagram bandwidth, can also be used for data transfer if resources given by (9) are not sufficient to transmit all the low-priority data of the station. The additional bandwidth needed for datagram is then

$$\lambda_d = \left[p \cdot U - \lambda_u\right]^+ \qquad (10)$$

where $[y]^+$ is y for positive y, and zero otherwise.

In a real case, as explained in Section 2, transient intervals occur both to obtain and to release the (P - R) extra bandwidth. Figure 8 highlights the differences between the ideal case and the real one. Specifically, in the allocation and deallocation transient intervals, the dashed line represents the bandwidth allocation in the ideal case while the continuous line shows the real bandwidth-allocation level. Comparing Figure 7 and Figure 8 shows that there are differences only in the two dashed areas which correspond to the allocation and deallocation transients. To take into account the effect of these transients the allocation and deallocation costs must be added to the cost function.

3) *Allocation cost.* In the ideal case, the excess-stream cost was computed by assuming in this period an allocation level equal to P, and hence during this transient period, it provides a cost overestimation with respect to the real case in which the allocation level is still R. The real cost is negative. During this transient the buffer size increases and this backlog is transmitted by using future unused bandwidth. For this reason it must be considered as a negative cost with respect to the excess-stream cost computed in the ideal case. To remove this overestimation we need to subtract the allocation transient area (see Figure 8). The allocation transient lasts for the sum of a time interval, say x, required to fill the buffer up to the T_u level, plus the allocation time D_a.

Hence, the dashed area is:

$$\left(D_a + x\right) \cdot (P - R).$$

The exact computation of x is complex. However, using our source model it is possible to estimate the value of x when the source changes its level from i to j, $x_{i,j}$ (1). In this case

$$x_{i,j} = \frac{T_u}{H_j - R},$$

and hence, the average allocation cost for the bandwidth allocation $\left(\lambda_{up}\right)$ is:

---

(1) This computation is performed under the assumption that during the transient the source level j does not change.

$$\lambda_{up} \cong (P-R)\sum_{i=0}^{R}\sum_{j>R}\left(D_a + x_{i,j}\right)\cdot \pi_i q_{i,j}$$

where π_i is the steady-state probability for VBR video to be in level *i*, $q_{i,j}$ is the probability that the video moves from level *i* to level *j*, and hence $\sum_{i=0}^{R}\sum_{j>R}\pi_i q_{i,j}$ is the frequency of the *R* to *P* transitions in the bandwidth allocation levels.

4) *Deallocation cost*. In the real case an additional cost is introduced whenever a station wants to release the (*P* - *R*) bandwidth because a $\left(D_a + y\right)$ delay (at least 500 ms) must be kept into account, i.e. the time between the transmission of a relinquish request and confirmation from the satellite network.

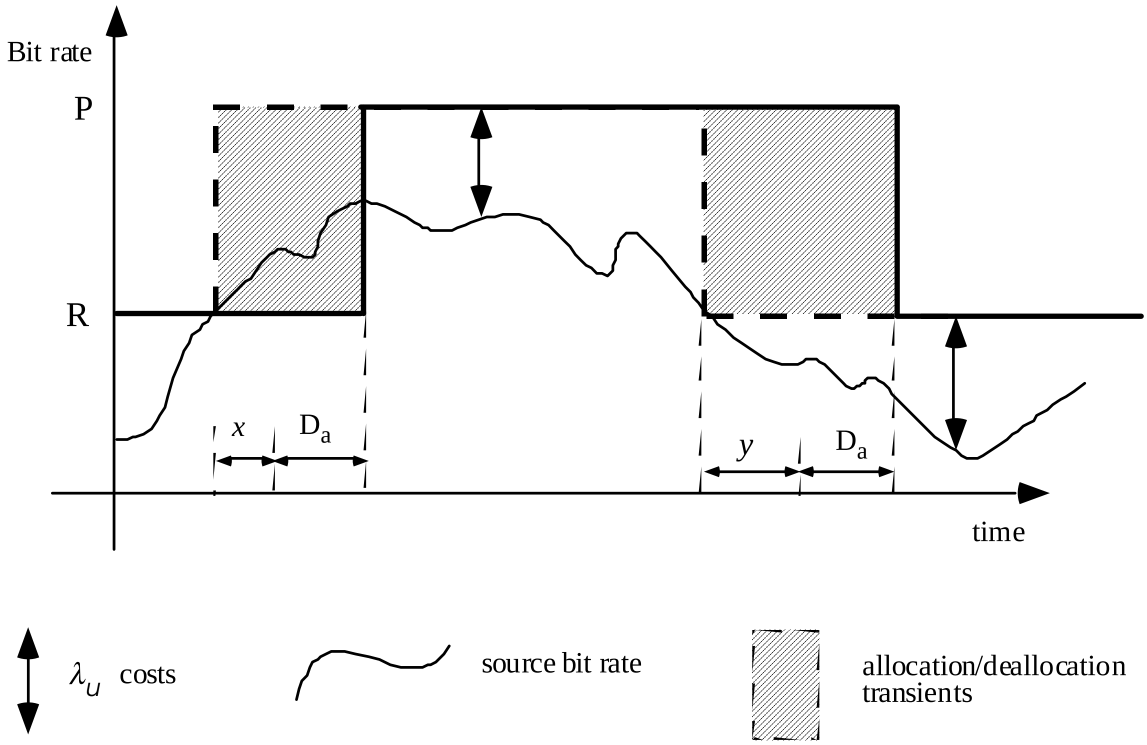


Figure 8. Real-time bandwidth lost due to dynamic allocation.

The situation is symmetrical with respect to the previous case, so the average cost for the bandwidth deallocation $\left(\lambda_{down}\right)$ is:

$$\lambda_{down} = (P-R)\sum_{i>R}^{P}\sum_{j\le R}\left(D_a + y_{i,j}\right)\pi_i q_{i,j} \qquad (11)$$

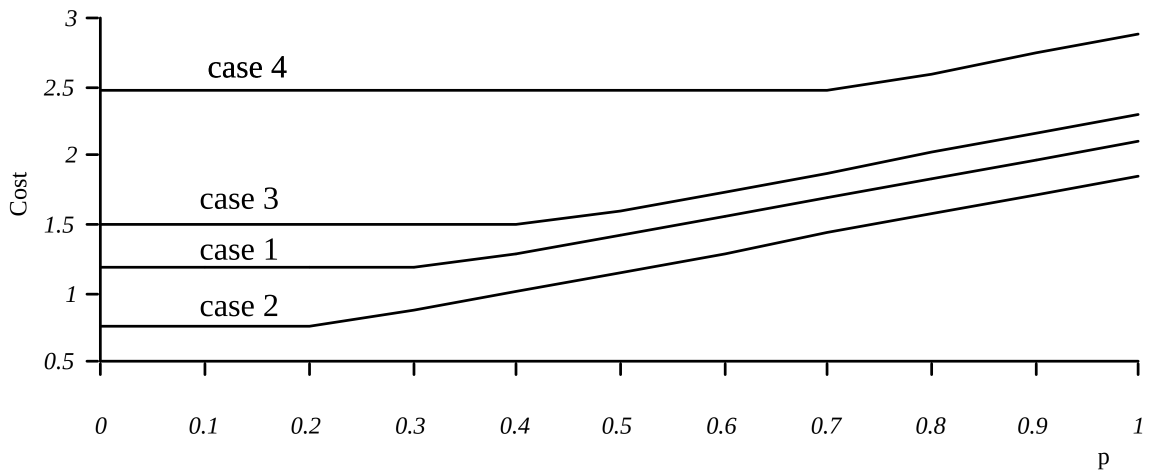
being $y_{i,j} = \dfrac{T_d}{R - H_j}$

By exploiting the cost functions defined in this section, we are now ready to compare the global cost for each of the bandwidth allocation parameter settings defined in Table 2. The cost is clearly the sum of the following components:

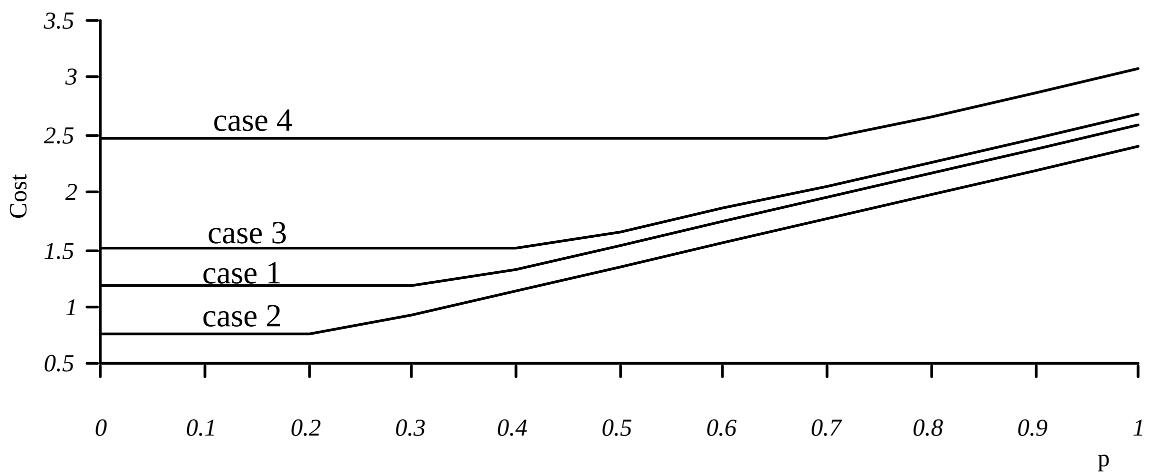
$$Cost = 1 \cdot \lambda_{down} + 1 \cdot \lambda_u + \beta \cdot \lambda_d - 1 \cdot \lambda_{up} \ . \qquad (12)$$

Results obtained by applying Formula (12) to the four allocation cases of Table 2 are presented in Figures 9a-9c. These figures assume D_a = 500ms and a price for a data bandwidth level, β, set to 0.4, 0.6 and 0.8, respectively.

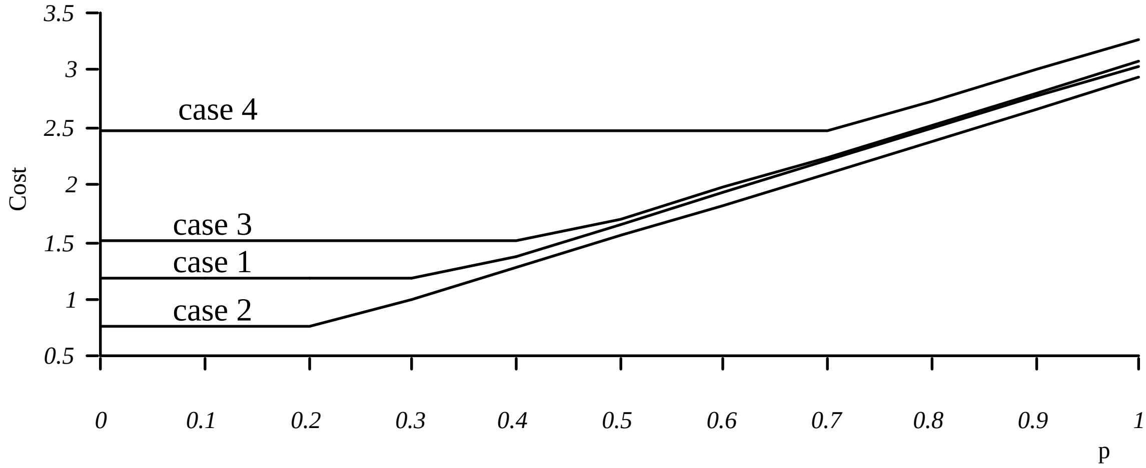
Clearly, case 2 is the winner in this comparison, irrespective of β and p. Hence we can conclude that allocating to the VBR video a level corresponding to H=1 as the minimum level (the R parameter) and a level corresponding to H=4 as the maximum level (the P parameter) is the optimal solution whatever the amount of data traffic to be transmitted by the station.



a)



b)



c)

Figure 9. Cost β equal to 0.4 (a), 0.6 (b), 0.8 (c) respectively.

The above results show that the best values for P and R must be calculated considering H=1 and H=4, respectively. Table 3 summarises the results obtained so far for our case study, in which 1 GOP is equal to 12 frames and the frame rate is 24 frames/s. We assume that some sort of resource reservation is made on the input and receive links (see Figure 3), so that packets traversing those links have a bound on their maximum delay. The values assumed for the latency and the max jitter are typical of a wide area terrestrial route.

Parameter	Reference	Value	
GOP peak rate of the stream	Table 1	1865	[Kbit/s]
GOP mean rate of the stream	Table 1	374	[Kbit/s]
Booked bandwidth P	Figure 2	1220	[Kbit/s]
Low level bandwidth R	Figure 2	583	[Kbit/s]
GOP length τ_{GOP}	Figure 3	500	[ms]
Link latencies $\tau_i = \tau_r$	Figure 3	50	[ms]
Satellite link latency τ	Figure 3	250	[ms]
Max jitter induced by links $J_i = J_r$	Figure 3	200	[ms]
Allocation delay D_a	Section 3	500	[ms]
Threshold levels $T_u = T_d$	Formulas (1), (7)	14	[KB]
Input station buffer size B_i	Formula (2)	83	[KB]
Playback delay at the MPEG receiver D_p	Formula (4)	757	[ms]
Playback buffer size at the MPEG receiver B_p	Formula (5)	226	[KB]
Reproduction delay D_s	Formula (6)	2360	[ms]

Table 3. Key parameter values for Star Wars case study.

It is interesting to note that the buffer sizes B_i and B_p are proportional to the data rate of the MPEG stream, while the delays are not.

5. Conclusions

The V2L-DA algorithm guarantees the peak bandwidth (P) for a VBR video application, while maintaining good efficiency in the overall channel bandwidth allocation. In fact, the throughput of a VBR application is often several times lower than its peak throughput (five times in our case), and this leads to an inefficient use of the channel bandwidth if peak rate allocation is adopted. To increase the efficiency in bandwidth allocation, when the throughput of a VBR application is below a certain threshold R, only a bandwidth up to R is actually allocated to this application, while the difference $P - R$ is booked for this application but (until requested) is used by the channel dispatcher to satisfy the datagram traffic of all the network stations. As soon as the throughput of the VBR encoder exceeds R, the channel dispatcher allocates all the bandwidth P already booked by this application. We have discussed the setting of parameters P and R in order to optimise the utilisation of the network capacity. Specifically, by considering the transmission of the trace of a movie produced by an MPEG-1 encoder, the optimal bandwidth allocation for this VBR video application is obtained by setting R to about 40% of the booked bandwidth P(2). Taking into consideration that most of the time the source bit rate is below R [11], it follows that 60% of the bandwidth booked by a VBR video application can be used to satisfy datagram transmissions of all the stations.

In this work we have assumed that, when the traffic exceeds the maximum allocable bandwidth (*spikes*), a best effort policy is used (e.g. by exploiting the datagram traffic). Although very rare, these events prevent professional quality transmissions. To avoid this limitation, future work will be devoted to extending the stream allocation policy to more than the two bandwidth levels currently used. The highest level should be used to manage the spikes.

REFERENCES

[1] N. Celandroni, E. Ferro, F. Potortì: "Satellite bandwidth allocation schemes for VBR applications", CNUCE Report C94-24, December 1994.

[2] N. Celandroni, E. Ferro, N. James, F. Potortì: "FODA/IBEA: a flexible fade countermeasure system in user oriented networks", International Journal of Satellite Communications, Vol.10, N.6, pp. 309-323, November-December 1992.

(2) As described in Section 4, the booked bandwidth P is set to 5/8 of the GOP peak rate.

[3] T. Zein, G. Maral, T. Brefort, M. Tondriaux: “Performance of the Combined/Fixed Reservation Assignment (CFRA) scheme for aggregated traffic”, Proceedings of the COST-226 Final Symposium, pp.183-198, Budapest (H), 10-12 May 1995.

[4] Dornier Deutche Aerospace: “Advanced Business Communications via satellite”, System Description, July 1992.

[5] Jahangir I. Mohammed, Tho Le-Noc: “Performance anlysis of Combined Free/Demand Assignment Multiple Access (CFDAMA) Protocol for Packet Satellite”, ICC 94, pp. 869-873.

[6] Tho Le-Gnoc, S.V. Krishnamurthy: “Performance of Combined Free/Demand Assignment Multi-Access (CFDAMA) Protocol with Pre-Assigned Request Slots in Integrated Voice/Data Satellite Communications”, Proceedings of the International Conference on Communications, ICC'95, pp. 1572-1576, 1995.

[7] D. Le Gall: "MPEG: “A Video Compression Standard for Multimedia Applications”, Communications of the ACM, Vol. 34, No. 4, April 1991, pp. 46-58.

[8] D. Minoli and R. Keinath: “Distributed Multimedia Through Broadband Communications Services”, Artech House, Inc., MA, 1994.

[9] P. Parcha and M. El Zarki: “MPEG Coding For Variable Bit Rate Video Transmission”, IEEE Communications Magazine, May 1994, pp. 54-66.

[10] G. K. Wallace: “The JPEG Still Picture Compression Standard”, Communications of the ACM, Vol. 34, No. 4, April 1991, pp. 30-44.

[11] M. Conti, E. Gregori, A. Larsson: “A study of the impact of MPEG-1 correlations on video-sources statistical multiplexing”, IEEE Journal on Selected Areas in Communications, Vol. 14, September 1996, pp.1455-1471.

[12] The ATM Forum: “ATM service cayegories: the benefits to the user”, White paper, European Market Awareness Committee.

[13] L. Chiariglione: “The development of an integrated audiovisual coding standard: MPEG”, IEEE Proceeding, Vol. 83, No. 2, February 1995, pp. 151-157.

[14] N. Celandroni, E. Ferro, F. Potortì: “DRIFS-TDMA. A proposal for a satellite access distributed-control algorithm for multimedia traffic in a faded environment”, to appear in the International Journal of Satellite Communications.

[15] N. Celandroni, E. Ferro, F. Potortì: “FEEDERS-TDMA. A distributed-control algorithm for satellite channel capacity assignment in a mixed traffic and faded environment”, to appear in the International Journal of Satellite Communications.