MPEG Video Traffic on a MetaRing Network: Complexity Reduction of a "Worst-Case" Model for Bandwidth Allocation Analysis

G. Anastasi, P. Foglia, L. Lenzini

University of Pisa, Dept. of Information Engineering
Via Diotisalvi 2 - 56126 Pisa, Italy
email: {anastasi, foglia, lenzini}@iet.unipi.it

Abstract
This paper deals with the problem of selecting the optimal bandwidth allocation scheme for MPEG video transport on a MetaRing network. The problem is addressed by analyzing a simplified "worst case" network model faded by an MPEG video traffic model which only captures short range correlations. To assess to what extent the video traffic model still provides useful information, a methodology based on the busy time distribution of the buffer is introduced. According to this methodology, the MPEG video traffic model still provides accurate results until the portion of bandwidth allocated to each MPEG video source is greater than or equal to 50% of the peak rate.

1. Introduction

In high speed networks of the future most of the traffic will be generated by multimedia sources such as teleconferencing terminals and video-on-demand servers. Given the proportion of bandwidth requirements of the various media (e.g., data, speech and video), most of the bandwidth provided by a high speed network is likely to be used by video traffic.

There are two main ways to reduce costs of video transport in current and future high speed networks. From an application point of view, efficient Variable Bit Rate coding techniques allow bandwidth saving, while keeping the quality level constant. In this context, the MPEG (ISO Moving Picture Expert Group) coding scheme seems to be the most promising ([LEG91]). It allows an average to peak bit rate ratio in the order of 0.1 ([ROS95a]).

From a network point of view, efficient resource allocation algorithms, based on statistical multiplexing, can be designed to take advantage of the VBR behaviour of video sources. Statistical multiplexing increases bandwidth utilization by allocating, to each VBR video source, a bandwidth lower than its peak rate. Temporary congestion phenomena are handled by storing extra information in a buffer. With an appropriate design of the resource (bandwidth and buffer) allocation scheme it is possible to provide users with probabilistic bounds on the Quality of Service (QoS) they achieve.

Unfortunately, VBR video applications can only tolerate extremely low packet loss rates (in the order of 10^{-9} and less). This makes it extremely difficult to select the optimal resource allocation scheme. In fact, the problem cannot be dealt with by traditional simulation due to the difficulty of estimating very low tail probabilities. Analysis thus becomes the most suitable technique in this case.

This paper addresses the problem of efficient bandwidth allocation for MPEG video transport on a MetaRing LAN (Local Area Network). The MetaRing ([OFE94]) is a simple and effective MAC (Medium Access Control) protocol for high speed Local and Metropolitan Area Networks (MANs), which has received a lot of attention. It is capable of integrating real-time (synchronous) and non real-time (asynchronous) traffic while providing fairness to non real-time users ([WU92]).

Due to the complexity of the MetaRing MAC protocol, the problem of selecting the optimal bandwidth allocation scheme is analyzed by using a simplified "worst-case" model of the network. This model is analytically tractable and yet still provides useful information on the QoS achieved by users.

As in any performance analysis, the choice of the appropriate traffic model is essential for the reliability of the results. An ideal traffic model should be sufficiently simple to preserve analytical tractability. Furthermore, it should be able to take into account all the aspects which are relevant to the goal of the analysis. Specifically, an MPEG video traffic model should capture the correlations exhibited by MPEG video sequences over any time scale. However, to our knowledge, an MPEG video traffic model with both of the above properties does not exist. The traffic model used in this paper only captures short term correlations. The results are thus only meaningful when parameters characterizing the bandwidth allocation scheme range in a certain region of values ("model reliability region"). A methodology is developed in the paper to identify such a reliability region.

The paper is organized as follows. Section 2 briefly introduces the MetaRing MAC protocol. In Section 3 the MetaRing worst-case queuing model is derived and a description of MPEG video traffic model is given. Section 4 is devoted to the model solution. The results are discussed in Section 5, while the methodology to assess the reliability region for the video traffic model is introduced in Section 6. Section 7 concludes the paper.

2. MetaRing MAC Protocol Description

The MetaRing is a Medium Access Control (MAC) protocol proposed for high speed LANs and MANs. It connects a set of stations by means of a bi-directional ring made up of full duplex point-to-point serial links. A MetaRing can operate under two basic access control modes: *buffer insertion* for variable size packets, and *slotted* for fixed length cells. In this paper only the slotted access mode is considered.

The MetaRing MAC protocol provides two types of services: synchronous for real-time users and asynchronous for non real-time users. This section outlines the MAC protocol aspects that are relevant to the study reported in the paper. Details on the MAC protocol can be found in [CID93], [WU92], and [OFE94].

According to the slotted access mode, information is segmented into cells and each cell is transmitted in one slot. Slots are structured into a header and an information field. The header includes a busy bit which indicates whether the slot is empty (busy bit=0) or busy. Since the ring is bidirectional the MetaRing MAC protocol uses the *shortest path* criterion to choose one of the possible directions. Cells are removed by the destination station which frees the slot by resetting the busy bit to zero. After cell removal the slot can be reused by the same station or by its downstream stations. Slot reuse increases the aggregate throughput beyond the capacity of single links, but it can cause starvation. This happens if a station is constantly being “covered” by an upstream ring traffic and is thus unable to access the ring for a very long period of time.

In order to prevent this, the MetaRing MAC protocol includes a fairness mechanism which is based on a control signal, called SAT (*SATisfied*), circulating in the opposite direction from the information flow it regulates. Each station has a counter which increases by one each time the station transmits a cell. The station can transmit cells whenever it observes an empty slot unless the counter is equal to the value of a MAC protocol parameter denoted by K. When the counter reaches the value of K, the station must refrain from sending new cells until the SAT signal arrives at the station. Upon SAT arrival, the station resets the counter to zero if it has been sent a number of cells greater than or equal to an additional protocol parameter denoted by L (with $K \geq L \geq 0$) and releases the SAT towards its upstream station. When the counter is less than L the station is not satisfied, and the following two mutually exclusive events can occur:

- if the station has data to transmit, it holds the SAT until the counter reaches the value of L, then the station resets the counter and releases the SAT;
- if the station has no data to transmit, the station resets the counter and releases the SAT signal.

Each station has two queues, one for synchronous traffic and the other for asynchronous traffic. The traffic in the synchronous queue has priority over asynchronous traffic: the station can transmit synchronous traffic regardless of its asynchronous queue state. For example, a station which holds the SAT (because it is not satisfied) will first send traffic from the synchronous queue, and only then does the station send its asynchronous quota and releases the SAT.

Synchronous traffic is transmitted with a guaranteed bandwidth: each synchronous station is given the opportunity to transmit a predefined quota of cells in a certain time interval. In order to achieve this the MAC protocol cyclically needs to refrain from transmitting the asynchronous traffic. The asynchronous traffic transmission will be resumed later when synchronous traffic requirements are met. The integration mechanism employs three control signals, *ASYNC-EN(GR)*, *ASYNC-EN(YL)*, *ASYNC-EN(RD)* (Asynchronous Enable - Green, Yellow, Red, respectively) to enable or disable the transmission of the asynchronous traffic. These signals circulate in the opposite direction from the information they regulate. When the ASYNC-EN(GR) rotates around the ring the asynchronous traffic transmission can occur. Furthermore, each station will forward the ASYNC-EN(GR) immediately after receiving it. After the ASYNC-EN(GR) has completed at least *r* rounds (*r* is a protocol parameter), a station with a backlog of synchronous traffic (i.e., a station for which the first packet in the synchronous queue has been waiting for more than a predefined threshold *Thres*) can change the attribute from GREEN to YELLOW. When stations observe the ASYNC-EN(YL) signal they must refrain from transmitting asynchronous cells and forward the signal to adjacent stations. When after a round the ASYNC-EN(YL) reaches its origin station, the station itself switches the signal attribute from YELLOW to RED. The ASYNC-EN(RD) is transferred once around the ring. A station forwards the ASYNC-EN(RD) to its upstream neighbour if it has no backlog of synchronous traffic, otherwise it holds the ASYNC-EN(RD) until its backlog of synchronous traffic has been depleted. When the ASYNC-EN(RD) signal returns to its origin station it will change its attribute back to GREEN. The ASYNC-EN(GR) signal should complete at least r $(r \geq 0)$ rounds before a new cycle can start again. Therefore, the integration mechanism of asynchronous and synchronous traffic uses two protocol parameters: r and *Thres.*

3. Complexity Reduction and Model Definition

The modelling and performance analysis of our environment is a very complex problem, due to

- the MetaRing MAC protocol;
- the number of stations and their location along the ring;
- the workload characterization.

An exact model of the MetaRing MAC protocol should take into consideration the behaviour of each (synchronous and asynchronous) station therein. In fact, time intervals between consecutive control signal arrivals (SAT and/or ASYNC_EN) at a specific station are strongly influenced by the behaviour of all the other stations. This is the typical situation in any LAN. However, in networks where slot reuse is adopted, this dependency is further complicated by the slot occupancy pattern observed by stations. For a specific station this pattern depends on the destination addresses of the upstream stations. In order to obtain a model of our environment that can be solved analytically, we identify a specific network configuration (hereafter 'scenario') which is still significant from a practical standpoint.

As far as the workload characterization is concerned two important issues need to be pointed out. First, due to the Group of Picture (GoP) structure used in the video coding process (see Section 3.3), the autocorrelation function of MPEG video sequences exhibits a pseudo-periodic behaviour. This makes all models based on only one picture type unsuitable for modelling MPEG video traffic [ROS94b]. Second, MPEG video sequences are characterized by long range dependence. Models which do not capture it provide optimistic predictions of the network performance [BLE96]. On the other hand, it is almost impossible to define a very accurate model while still preserving analytical tractability.

3.1 Network Complexity Reduction

The network model is reduced in complexity by considering a worst-case scenario in which the network congestion is stressed, and by focusing on the least favourite station.

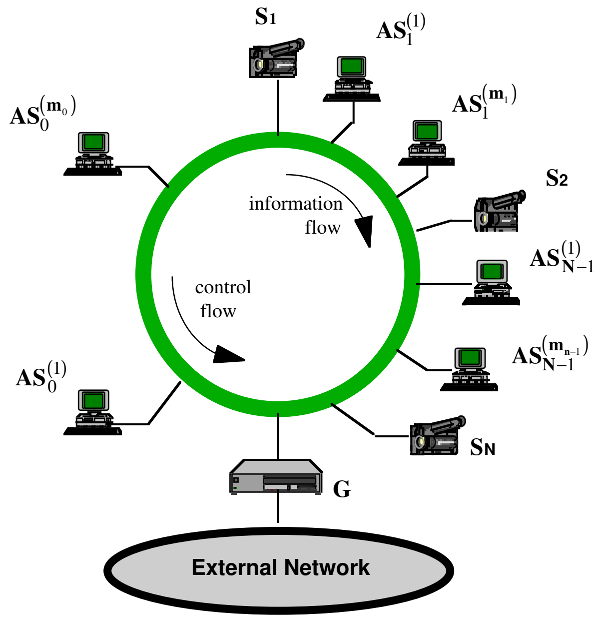

Figure 1: Network Scenario

With reference to Figure 1, where the network scenario analyzed in the paper is depicted, the following terminology will be used throughout:

$S_1, S_2 \ldots S_N$ are synchronous stations transmitting video traffic;
$AS_0^{(1)} \ldots AS_{N-1}^{(m_{N-1})}$ are asynchronous stations transmitting data traffic;
G is a gateway interconnecting the MetaRing with an external network;
$d_{i,i+1}$ is the distance (in slots) between synchronous stations S_i and S_{i+1};
T_{RING} is the ring length (in slots).

In order to identify the least favourite synchronous station in the network scenario shown in Figure 1 we assume that all the synchronous stations are concentrated in one half of the ring and that each of them transmits its traffic only to the gateway G. Also asynchronous stations located between the synchronous station S_1 and the gateway (by considering the direction of the information flow) transmit only to the gateway G. On the other hand, all the other asynchronous stations send their cells to a multicast address.

When the destination address is a multicast address, according to the MetaRing MAC protocol, cell removal is operated by the source station itself ([CID93]). Furthermore, we assume that asynchronous stations operate in asymptotic conditions (i.e., they always have cells to send). Hence, slots used by asynchronous stations are never reused by synchronous stations.

We also assume that video frame arrivals occur at the same time instant for all the synchronous stations, i.e., video sources are synchronized.

Under the above assumptions station S_N is the least favourite synchronous station as it observes empty slots only when the asynchronous traffic transmission is disabled and all the other synchronous stations have transmitted their preallocated quota. Thus, throughout we will focus on station S_N which will be referred to as the "tagged station".

The number of empty slots observed by the tagged station is clearly strongly influenced by the arrival process and the amount of bandwidth preallocated to each other synchronous station. However, we can assume that $S_1, S_2 \ldots S_{N-1}$ are CBR (Constant Bit Rate) video sources, i.e., they generate fixed length video frames. Furthermore, if T_v is the video frame generation period (expressed in slots), we assume that the preallocated quota of cells B_i that station S_i $(i = 1,2,\ldots,N-1)$ can transmit in a time interval of duration T_v is exactly equal to the (constant) number of cells in each video frame.

To derive a worst-case model of the network we need to know

- the maximum amount of bandwidth B_N (i.e., number of slots in a T_v interval) which can be guaranteed for the tagged station, while still preserving the quota allocated to the other synchronous stations;
- the slot occupancy pattern observed by the tagged station.

The first question is quite easy to answer. In fact, the sum of quotas preallocated to each synchronous station must be less than or equal to the number of slots between two consecutive video frame arrivals, i.e., T_v. Hence,

$$B_N \le T_v - \sum_{i=1}^{N-1} B_i \qquad [1]$$

On the other hand, the slot occupancy pattern observed by the tagged station depends on many factors. First of all, it is influenced by the values of the parameters which regulate the traffic integration mechanism: r and $Thres$. However, in [MOT94] it was found that the lower the value of each of the two parameters, the more reactive the traffic integration mechanism. Thus, in order to minimize the access delay experienced by traffic transmitted by synchronous stations (which is delay-sensitive in most cases) both r and $Thres$ values should be as low as possible. In this paper we will follow this criterion by putting $r = 0$ and $Thres = 0$.

Condition [1] guarantees that the tagged station achieves its quota of slots, B_N, every T_v interval. However, this is true only after the asynchronous traffic transmission has been disabled. In fact, suppose a video frame arrival occurs at time t_0 and the asynchronous transmission is enabled at that point in time. Since the traffic integration mechanism takes some time to inhibit all the asynchronous stations in the network, starting from t_0 there is a transient period during which the tagged station may observe less empty slots (for each interval of duration T_v) than its preallocated quota. In the following we will try to characterize the duration of the transient period and the amount of bandwidth achieved by the tagged station during this transient period. To this end we make some claims related to the scenario under consideration.

Claim 3.1

In the scenario under consideration, if a video frame arrival occurs while the asynchronous transmission is enabled then station S_1 observes the first empty slot after a time interval

$$B_{AS} \le 2T_{RING} + d_{AS,1} \qquad [2]$$

since the video frame arrival. In [2] $d_{AS,1}$ is the distance (in slots) of station S_1 from the furthest asynchronous station ($AS_0^{(1)}$ in Figure 1).

Proof. *See Appendix/*

Due to the delay B_{AS} some synchronous stations may not have finished transmitting their quota of cells when the next video frame arrives. However, this can never occur if the condition specified in the following claim holds:

Claim 3.2
If the asynchronous traffic transmission is enabled when a video frame arrives then condition

$$\sum_{i=1}^{N} B_i \le T_v - 3T_{RING} \qquad [3]$$

implies that all the synchronous stations complete the transmission of their quota before the next video frame arrival.

Proof. *See Appendix*

The previous claim states that whenever condition [3] holds the tagged station is guaranteed to achieve its quota of slots even immediately after the arrival of a video frame which finds the asynchronous transmission enabled. However, condition [3] may represent a very strict constraint especially when the value of T_{RING} is high (e.g., in a metropolitan environment). In the following we will analyze what happens when the above condition is relaxed.

It is easy to verify that if there exists a (synchronous) station index $N-j$ $\left(j \in [0, N-1]\right)$ such that

$$\sum_{i=1}^{N-j-1} B_i \le T_v - B_{AS} - \sum_{i=1}^{N-j-1} d_{i,i+1} < \sum_{i=1}^{N-j} B_i \qquad [4a]$$

then S_{N-j} is the station that is still transmitting (i.e., not having completed its quota) when the next video frame arrives. Otherwise if there exists a (synchronous) station index $N-j$ $\left(j \in [0, N-1]\right)$ such that

$$\sum_{i=1}^{N-j-1} B_i \le T_v - B_{AS} - \sum_{i=1}^{N-j-2} d_{i,i+1} < \sum_{i=1}^{N-j} B_i + d_{N-j-1,N-j} \qquad [4b]$$

then station S_{N-j-1} has completed the transmissions of its quota, but station S_{N-j} has not yet started transmitting when the next video frame arrives.

Depending on the value of index j for which either [4a] or [4b] are satisfied the slot occupancy pattern observed by the tagged station in the transient period may be quite different. Specifically, the following cases may occur:

a) condition [4a] holds and $j=0$, i.e., the tagged station is the station still transmitting when the next frame arrives;

b) condition [4b] holds and $j=0$, i.e., stations from S_1 to S_{N-1} have finished transmitting their quota but the tagged station hasn't begun transmitting when the next frame arrives;

c) condition [4a] ([4b]) holds and $j \ne 0$, i.e., a synchronous station different from the tagged station is still transmitting (waiting for transmission) when the next frame arrival occurs.

The following claims characterize the slot occupancy pattern observed by the tagged station during the transient period whenever one of the cases specified above occurs.

Claim 3.3
If condition [4a] holds and $j=0$ then the transient period lasts just one T_v period and during this interval the tagged station observes a number of empty slots given by

$$B'_N = B_N - \left(B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1}\right) \ge B_N - 3T_{RING} \qquad [5]$$

Proof. *See Appendix*

Claim 3.4
If condition [4b] holds and $j=0$ then the transient period lasts two T_v periods. In the former period the tagged station observes no empty slots, while in the latter it achieves a number of slots equal to

$$B'_N = T_v - \sum_{i=1}^{N-1} B_i - d_{N-1,N} + \left(T_v - B_{AS} - \sum_{i=1}^{N-1} B_i - \sum_{i=1}^{N-2} d_{i,i+1}\right) \qquad [6]$$

Proof. *See Appendix/*

If condition [4a] or [4b] holds and $j > 0$ each station behaves as in Figure 11 (see appendix) with the difference that the station which is still transmitting (or which still has to start transmitting) when the new frame arrives is located before the tagged station. Thus, the tagged station itself does not achieve any slots in the period between t_0 and $t_0 + T_v$ (see Figure 12 in the appendix).
The number of empty slots observed by the tagged station after $t_0 + T_v$ depends on the values of quotas for the other stations. A closed formula for the general case is hard to find. However, this is possible once the values of quotas and distances between synchronous stations are known. For example, it is possible to prove claim 3.5 which relates to the case when the quota is the same for all the stations and such that condition [4a] holds with $j = 1$.

Station index	$(t_0,$ $t_0 + T_v]$	$(t_0 + T_v,$ $t_0 + 2T_v]$	$(t_0 + 2T_v,$ $t_0 + 3T_v]$
1	B	B	B
2	B	B	B
3	B	B	B
:	:	:	:
N-2	B	B	B
N-1	B'	$(B - B') + B$	B
N	0	B''	B

Table 1: Number of empty slots observed by each synchronous station in successive intervals of duration T_v when condition [4a] holds with $j = 1$.

Claim 3.5

If $B_i = B$ $(i = 1,2,\dots N)$, $\sum_{i=1}^{N} B_i = NB = T_v$ and B is such that condition [4a] holds with $j = 1$ then the number of empty slots observed by each synchronous station in successive intervals of duration T_v is as reported in Table 1 where

$$B' = T_v - B_{AS} - (N-2)B - \sum_{i=1}^{N-2} d_{i,i+1} \qquad [7]$$

$$B'' = B' - d_{N-1,N} \qquad [8]$$

Proof. *See Appendix*

Claims 3.3, 3.4, and 3.5 characterize the slot occupancy pattern observed by the tagged station for different conditions which can occur as a consequence of the operation parameter values. This slot pattern is observed until the tagged station buffer becomes empty before the next frame arrival. When this occurs the process that resumes the asynchronous traffic transmission is activated. How exactly this process evolves is hard to describe, since it depends on the specific network state (i.e., value of counters in each asynchronous station, positioned along the ring of SAT and ASYNC_EN signals) when the process starts. However, as we are interested in the worst case behaviour of the tagged station we can assume that the interval between the time instant when the process is activated and the time instant at which the next video frame arrives is long enough to enable all the asynchronous stations before the frame arrival[1].
Whenever the buffer of the tagged station becomes empty before the next frame arrival, a new cycle starts immediately after the video frame arrival. During each cycle the slot occupancy pattern observed by the tagged station is the same, and for the cases reported above it is the one specified by claim 3.3, 3.4, or 3.5.

Claims 3.3, 3.4, and 3.5 clearly show that the slot occupancy pattern observed by the tagged station during each cycle depends on conditions [4] and hence, on the quota values. However, if the conditions specified in the following claim are matched, only one slot occupancy pattern is possible, whichever the specific values of quotas are.

Claims 3.6

If the tagged station quota is

$$B_N \geq 3T_{RING} \qquad [9]$$

and under the hypothesis that all the available bandwidth is allocated, i.e.,

$$\sum_{i=1}^{N} B_i = T_v \qquad [10]$$

then conditions underlying claim 3.3 ([4a] with $j = 0$) hold.

Proof. *See Appendix*

[1]This assumption minimizes the number of empty slots observed by synchronous stations.

Claim 3.6 guarantees that whenever conditions [9] and [10] are both satisfied the only possible slot occupancy pattern observed by the tagged station becomes the one specified in claim 3.3.

3.2 Network Model Definition

For the worst case scenario under consideration, claims and considerations reported in Section 3.1 allow us to identify the exact slot occupancy pattern observed by the tagged station once all the necessary operation parameters (distances between stations, ring length and, if conditions underlying claim 3.6 are not matched, quota values) have been specified. Therefore, the behaviour of the tagged station can be analyzed in isolation from the rest of the network.

In the following we will assume that the tagged station transmits MPEG video traffic and we will describe its dynamics by using the embedding points technique. Specifically, we will observe the state of the tagged station just before the arrival of each video frame.

Observe, that if we operate in a LAN environment the value of T_{RING} is in the order of some units. On the other hand, MPEG video sequences typically exhibit an average number of cells per video frame which is in the order of some tens[2] (see Table 2 and [ROS95a]). To guarantee system stability the quota of slots allocated to the tagged station in a T_v period must be greater than the average bit rate (i.e., number of cells per video frame) of the MPEG source. This implies that, in a LAN context, condition [9] is satisfied and hence, if the bandwidth is fully allocated to synchronous stations (i.e., [10] holds) the slot pattern observed by the tagged station is the one defined by claim 3.3. More precisely, between two consecutive embedding points (i.e., during each T_v interval), the tagged station observes a number of empty slots equal to its quota B_N if the asynchronous traffic was disabled at the former embedding point. Otherwise the number of empty slots observed is

$$B'_N = B_N - \left(B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \right) \geq B_N - 3T_{RING}$$

[2]The cell size was based on the ATM standard.

In the following we will minimize the value of B'_N by assuming $B'_N = B_N - 3T_{RING}$. This allows us to define a worst-case network model which only depends on the ring length.

In the proof of claim 3.3 we showed that the tagged station is the last synchronous station to transmit within a video period (see, in particular, Figure 11 in the appendix). Furthermore, we have assumed that if the tagged station empties its buffer before the next video frame arrives then all the asynchronous stations are enabled when the frame itself arrives. Therefore, at a given embedding point, the asynchronous traffic transmission is enabled (disabled) if and only if at that time instant the buffer of the tagged station is empty (not empty).

From the above considerations it follows that, if Q_n indicates the number of cells in the buffer of the tagged station at the $n-$th embedding point, then it is

$$Q_{n+1} = \begin{cases} max\left[\left(A_n - P'_{max}\right),\ 0\right] & if\ \ Q_n = 0 \\ max\left[\left(Q_n + A_n - P_{max}\right),\ 0\right] & if\ \ Q_n > 0 \end{cases} \qquad [11]$$

where A_n is a random variable representing the number of cells in the video frame that arrived just after the $n-th$ embedding point, while $P_{max} = B_N$ ($P'_{max} = B_N - 3T_{RING}$) indicates the maximum number of cells the tagged station can transmit between the $n-th$ and $(n+1)-th$ embedding points, provided that the asynchronous traffic transmission was disabled (enabled) at the $n-th$ embedding point. Of course, the actual number of transmitted cells may be less. Specifically, this occurs if the tagged station empties its buffer before the maximum value has been reached.

3.3 MPEG Video Traffic Characterization

MPEG (ISO Moving Picture Expert Group) is currently the most important compression algorithm used in any type of video application. The MPEG compression is done by reducing both the spatial and temporal redundancy of the video stream. This is achieved by using three different types of frame, I, P, and B frames, each coded in a different way. Typically, I frames require more bits than P frames, while B frames have the lowest bandwidth requirement. These frames are then arranged in a deterministic sequence, e.g. “IBBPBBPBBPBB”, which is called Group of Picture (GoP). The number of frames in a GoP is referred to as the GoP size ([ROS95c], [LEG91]).

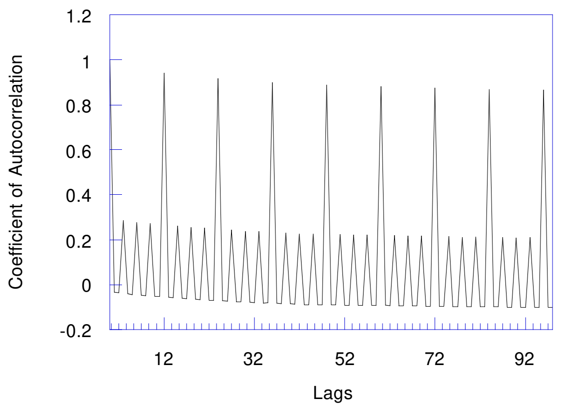

Figure 2: Coefficient of autocorrelation for the frame length of the "dino[3]" sequence

Due to the GoP, the autocorrelation function of an MPEG video sequence exhibits a pseudo periodic behaviour as shown in Figure 2. Therefore, it is not possible to model an MPEG video source by using a stochastic process based on only one frame type ([ROS94b]).

Furthermore, MPEG video sequences are also characterized by long range dependence ([ROS95a]). The video traffic model should thus capture this long range dependence. On the other hand, the complexity of the source model needs to be kept low to preserve analytical tractability.

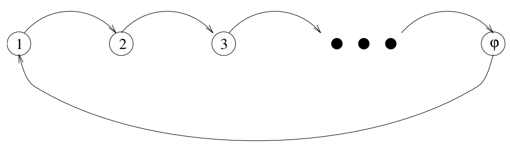

Figure 3: Markov chain underlying the MPEG video source model

In this paper we model MPEG video sources by a Markov chain with a number of states equal to the GoP size. Each state of the Markov chain is assigned with a different frame size distribution. Whenever the Markov chain is in a given state, the corresponding

[3]The "dino" sequence is described later.

frame size is obtained by a sample from the associated distribution. Transitions between states of the Markov chain occur just after frame arrivals. From a given state $\{i\}$ the Markov chain transits, with probability 1, to state $\{i+1\}$ if $i<\varphi$ and to state $\{1\}$ if $i=\varphi$ (see Figure 3).

As a consequence of its definition the model only captures the short term autocorrelation introduced by the GoP pattern.

4. Model Solution

On the basis of the considerations outlined in Section 3.0, the state of the system at the $n-th$ embedding point can be described by a couple (Q_n, Ph_n), where Q_n is the queue length at the tagged station and Ph_n indicates the state of the Markov chain underlying the video arrival process. It can easily be verified that process $\{Q_n, Ph_n;\ n=0,1,2,...\}$ is a Markov chain.

By introducing two new variables L and R so that $Q = LP_{\max} + R$ the system dynamics can be represented by Markov chain $\{L_n, R_n, Ph_n;\ n=0,1,2,...\}$ which is an M/G/1-type Markov chain ([NEU89]) with level L and phase (R, Ph). The resulting transition probabilities matrix P of this Markov chain is as shown in Figure 4.

$$P = \begin{bmatrix} B_0 & B_1 & B_2 & B_3 & \dots \\ A_0 & A_1 & A_2 & A_3 & \dots \\ 0 & A_0 & A_1 & A_2 & \dots \\ 0 & 0 & A_0 & A_1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Figure 4: Transition probability matrix P.

Blocks B_0, $B_k(k>0)$, and $A_k(k\ge 0)$ are square matrices of $P_{max}\cdot\varphi$ size. By operating the following re-arrangement of states $[R, Ph] \rightarrow [Ph, R]$, the B_0, $B_k(k>0)$, and $A_k(k\ge 0)$ matrices assume a block periodic structure, as shown in Figure 5 for B_0. The structure is exactly the same for $B_k(k>0)$, and $A_k(k\ge 0)$ matrices.

$$B_0 = \begin{bmatrix} 0 & \mathrm{B}_0^{(1)} & 0 & \dots \\ 0 & 0 & \mathrm{B}_0^{(2)} & \dots \\ \dots & \dots & \dots & \dots \\ \mathrm{B}_0^{(\varphi)} & 0 & 0 & \dots \end{bmatrix}$$

Figure 5: Block periodic structure for B_0

The internal structure of blocks which constitutes B_0, $B_k(k>0)$, and $A_k(k \ge 0)$ is shown in Figures 6, 7, and 8, respectively. In these figures $a_k^{(\phi)}$ indicates the probability that a frame of size k (i.e. including k cells), arrives given that the state of the Markov chain underlying the arrival process is $Ph=\phi$, and it is

$$a_k^{(\phi)} = 0 \quad \textit{if} \quad k<0$$

$$\tilde{a}_k^{(\phi)} = \sum_{i=0}^{k} a_i^{(\phi)}$$

$$B_0^{(i)} = \left[\begin{array}{c|cccc} & 0 & 1 & \cdots & P_{\max}-1 \\ \hline 0 & \tilde{a}^{(i)}_{P'_{\max}} & a^{(i)}_{P'_{\max}+1} & \cdots & a^{(i)}_{P'_{\max}+P_{\max}-1} \\ 1 & \tilde{a}^{(i)}_{P_{\max}-1} & a^{(i)}_{P_{\max}} & \cdots & a^{(i)}_{P_{\max}-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{\max}-1 & \tilde{a}^{(i)}_{1} & a^{(i)}_{2} & \cdots & a^{(i)}_{P_{\max}} \end{array}\right]$$

Figure 6: Structure of block components in matrix B_0.

$$B_k^{(i)} = \left[\begin{array}{c|cccc} & 0 & 1 & \cdots & P_{\max}-1 \\ \hline 0 & a^{(i)}_{P'_{\max}+kP_{\max}} & a^{(i)}_{P'_{\max}+kP_{\max}+1} & \cdots & a^{(i)}_{P'_{\max}+(k+1)P_{\max}-1} \\ 1 & a^{(i)}_{(k+1)P_{\max}-1} & a^{(i)}_{(k+1)P_{\max}} & \cdots & a^{(i)}_{(k+2)P_{\max}-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{\max}-1 & a^{(i)}_{kP_{\max}+1} & a^{(i)}_{kP_{\max}+2} & \cdots & a^{(i)}_{(k+1)P_{\max}} \end{array}\right]$$

Figure 7: Structure of block components in matrix $B_k(k>0)$.

$$A_k^{(i)} = \left[\begin{array}{c|cccc} & 0 & 1 & \cdots & P_{\max}-1 \\ \hline 0 & a^{(i)}_{KP_{\max}} & a^{(i)}_{KP_{\max}+1} & \cdots & a^{(i)}_{(K+1)P_{\max}-1} \\ 1 & a^{(i)}_{KP_{\max}-1} & a^{(i)}_{KP_{\max}} & \cdots & a^{(i)}_{(K+1)P_{\max}-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{\max}-1 & a^{(i)}_{(K-1)P_{\max}+1} & a^{(i)}_{(K-1)P_{\max}+2} & \cdots & a^{(i)}_{KP_{\max}} \end{array}\right]$$

Figure 8: Structure of block components in matrix $A_k(k \ge 0)$.

4.1 . Stability Analysis

If we calculate the vector $\beta = \left(\sum_{k=0}^{\infty} kA_k\right)\cdot \mathbf{e}$ where $\mathbf{e}=[1,1,\ldots 1]^T$, then the stability condition requires that ([NEU89])

$$\rho = \pi\beta < 1 \qquad [12]$$

where vector π is the invariant probability vector of the matrix $A = \sum_{k=0}^{\infty} A_k$. Vector π satisfies the following matrix system

$$\begin{cases} \pi = \pi A \\ \pi\mathbf{e} = 1 \end{cases} \qquad [13]$$

and A is a block periodic matrix:

$$A = \begin{bmatrix} 0 & \tilde{A}_1 & 0 & \ldots \\ 0 & 0 & \tilde{A}_2 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \tilde{A}_\varphi & 0 & 0 & \ldots \end{bmatrix} \qquad [14]$$

where the size of blocks $\tilde{A}_i$, $(1 \le i \le \varphi)$ is $P_{\max}$, and they satisfy the properties:

$$\mathbf{e} = \tilde{A}i \cdot \mathbf{e} \qquad [15a]$$
$$\mathbf{e^T} = \mathbf{e^T} \cdot \tilde{A}i \qquad [15b]$$

$$\tilde{A}_i = \left[\begin{array}{c|cccc} & 0 & 1 & \cdots & P_{\max}-1 \\ \hline 0 & \sum_{k=0}^{\infty} a^{(i)}_{kP_{\max}} & \sum_{k=0}^{\infty} a^{(i)}_{kP_{\max}+1} & \cdots & \sum_{k=0}^{\infty} a^{(i)}_{(k+1)P_{\max}-1} \\ 1 & \sum_{k=0}^{\infty} a^{(i)}_{kP_{\max}-1} & \sum_{k=0}^{\infty} a^{(i)}_{kP_{\max}} & \cdots & \sum_{k=0}^{\infty} a^{(i)}_{(k+1)P_{\max}-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{\max}-1 & \sum_{k=0}^{\infty} a^{(i)}_{(k-1)P_{\max}+1} & \sum_{k=0}^{\infty} a^{(i)}_{(k-1)P_{\max}+2} & \cdots & \sum_{k=0}^{\infty} a^{(i)}_{kP_{\max}} \end{array}\right]$$

Figure 9: Structure of $\tilde{A}_i$, $(1 \le i \le \varphi)$ blocks.

[15a] and [15b] can be proved by observing that matrices $\tilde{A}_i$, $(1 \le i \le \varphi)$ have the structure shown in Figure 9.

Figure 9 clearly shows that the sum of elements in each row and column is equal to 1. From the structure of matrix A and properties [15a] and [15b], after structuring π in the same manner as A (i.e. $\pi = [\pi_1, \pi_2, ..., \pi_\varphi]$), it follows that the linear system [13] can be written as :

$$\begin{cases} \pi_1 = \pi_1 \tilde{A}_1 \tilde{A}_2 .... \tilde{A}_j \\ \pi_2 = \pi_2 \tilde{A}_2 \tilde{A}_3 .... \tilde{A}_\varphi \tilde{A}_1 \\ \pi_i = \pi_i \tilde{A}_i \tilde{A}_{i+1} .... \tilde{A}_\varphi \tilde{A}_1 .... \tilde{A}_{i-1}, \; 3 \le i \le \varphi - 1 \\ \pi_\varphi = \pi_\varphi \tilde{A}_\varphi \tilde{A}_1 \tilde{A}_2 .... \tilde{A}_{\varphi-1} \\ \pi_i \mathbf{e} = \dfrac{1}{\varphi}, \; 1 \le i \le \varphi \end{cases} \qquad [16]$$

By applying properties [15b] it can be easily shown that vector $\pi = \frac{1}{P_{\max} \cdot \varphi} \mathbf{e}^{\mathrm{T}}$ satisfies equation [16]. Hence, $\pi = \frac{1}{P_{\max} \cdot \varphi} \mathbf{e}^{\mathrm{T}}$ is the invariant probability vector of the matrix A.

From the structure of matrices A_k, after some algebraic manipulations, the following expression can be derived

$$\rho = \frac{1}{P_{max}} \left[\frac{1}{\varphi} \sum_{j=1}^{\varphi} \sum_{i=0}^{\infty} i a_i^{(j)} \right] \qquad [17a]$$

Thus, system stability requires that

$$\frac{1}{P_{max}} \left[\frac{1}{\varphi} \sum_{j=1}^{\varphi} \sum_{i=0}^{\infty} i a_i^{(j)} \right] < 1 \qquad [17b]$$

This relation can be easily interpreted. In fact, terms in brackets represent the average number of cells arriving at the tagged station between two consecutive embedding points; clearly this number must be less than $P_{\max}$, i.e. the amount of bandwidth reserved for the tagged station. Throughout condition [17b] will always hold.

4.2. Computation of the steady state queue length distribution.

Following the standard methodology for the solution of an M/G/1-type Markov chain, our first step was to calculate the G matrix whose (i,j) entry represents the probability that starting at level $K+1$ with $(R, P_h) = i$ the system enters, for the first time, level K with $(R, P_h) = j$. A general technique to derive the G matrix is the following recursive algorithm [NEU89]:

$$G = \lim_{n \to \infty} G_n$$

where

$$G_n = \begin{cases} A_0 & n = 0 \\ \sum_{j=0}^{\infty} A_j G_{n-1}^j & n > 0 \end{cases} \qquad [18]$$

This algorithm can be computationally demanding and hence we used the recursive scheme reported in [LAT94] which gives the same result as [18] but is characterized by a higher order of convergence. According to this scheme it is

$$G = \lim_{n \to \infty} G_n$$

$$G_1 = \left(I - \sum_{i \ge 1} A_i \right)^{-1} A_0$$

$$G_{n+1} = \left(I - \sum_{i \ge 1} A_i G_n^{i-1} \right)^{-1} A_0$$

After computing the G matrix, we have to calculate the stationary probability vector π of the transition probability matrix P. By partitioning π in a manner congruent to the partition of matrix P we have $\pi = [\pi_0, \pi_1, \pi_2 ...]$. To derive the subvector π_0 we applied the Latouche algorithm ([LAT94]) which requires the solution of the following linear system

$$\begin{cases} \pi_0 = \pi_0 K \\ \pi_0 \kappa = 1 \end{cases}$$

where

$$K = \sum_{h=0}^{\infty} B_h G^h$$
$$\kappa = \mathbf{e} + \sum_{i \ge 1} B_i \sum_{0 \le k \le i-1} G^k \mu$$
$$\mu = \left(I - \sum_{i \ge 1} A_i \sum_{0 \le k \le i-1} G^k \right)^{-1} \mathbf{e} \qquad [19]$$

All the other components of vector π are obtained by applying the Ramaswami algorithm ([NEU89], [RAM88])

$$\pi_i = \left[\pi_0 \overline{B}_i + \sum_{j=1}^{i-1} \pi_j \overline{A}_{i+1-j} \right] \left(I - \overline{A}_1 \right)^{-1} \qquad i \ge 1 \qquad [20]$$

where

$$\overline{A}_n = \sum_{i=n}^{\infty} A_i G^{i-n} \ , \ \overline{B}_n = \sum_{i=n}^{\infty} B_i G^{i-n} \qquad n \ge 0$$

By observing that subvector π_i $(i \ge 0)$ provides the steady state probability of being at level i at an embedding point then the steady state distribution of the buffer occupancy is given by

$$P_q = \lim_{n \to \infty} Prob\{Q_n = q\} = \sum_{j=0}^{\varphi} \lim_{n \to \infty} Prob\{L_n = l, R_n = r, Ph_n = j\} \qquad [21]$$

where $l = \left\lfloor \frac{q}{P\max} \right\rfloor$ and $r = q \mod P_{\max}$.

5. Results

This section discusses the results obtained from the model solution for different bandwidth allocation schemes. Each bandwidth allocation scheme is characterized by the value of a parameter α defined as the ratio between the portion of bandwidth allocated to the tagged station and the peak rate generated by the tagged station itself.
By observing that the peak rate of a VBR (MPEG) video source can be expressed as the maximum number of cells in a frame, $N_{\max}$, divided by the video frame interarrival period, T_v, and that the bandwidth allocated to the tagged station is equal to $P_{\max}$ slots for T_v period, then α can be written as

$$\alpha = \frac{P_{\max}}{N_{\max}} \qquad [22]$$

Hence, $\alpha = 1.0$ identifies a peak allocation while a bandwidth allocation under the peak rate is characterized by a value of α less than 1.0.

Frame size distributions associated with the video traffic model have been estimated from the MPEG sequence "dino" available at [WUE]. This sequence was obtained by coding the first 40000 frames (corresponding to a 20 minute duration) of the film "Jurassic Park" with a GoP size of 12 $(\varphi = 12)$. Table 2 shows some statistical data of the above sequence. The minimum (N_{min}), maximum (N_{max}) and average (N_{avg}) number of cells[4] per frame together with the ratio between the average and maximum number of cells (N_{avg}/N_{max}) are reported both for the entire sequence and for I, B and P frames only. The ratio N_{avg}/N_{max} computed on the entire sequence provides the minimum value for α which guarantees system stability.

Frame type	N_{min} (cells)	N_{max} (cells)	N_{avg} (cells)	$\frac{N_{avg}}{N_{max}}$
ALL	3	312	34.55	0.11
I	46	312	143.92	0.46
B	3	181	19.52	0.11
P	4	203	38.14	0.18

Table 2: Statistical data of the "dino" sequence.

Link capacity = 150 Mbit/s
Slot size = 48 payload + 5 control bytes
Slot duration ≈ 2.83 μsec
Length of each ring = 5 slots ≈ 2.8 Km

Table 3: Network parameter values.

In our analysis the network capacity was assumed to be equal to 150 Mbit/sec, while the length of each ring was set to 5 slots corresponding to 2.5 ÷ 3 km. These values are

[4] Cell size is assumed to be equal to the ATM cell size, i.e., 48 payload bytes and 5 control bytes.

typical for a LAN environment. The list of all operation parameters characterizing the network configuration is reported in Table 3.

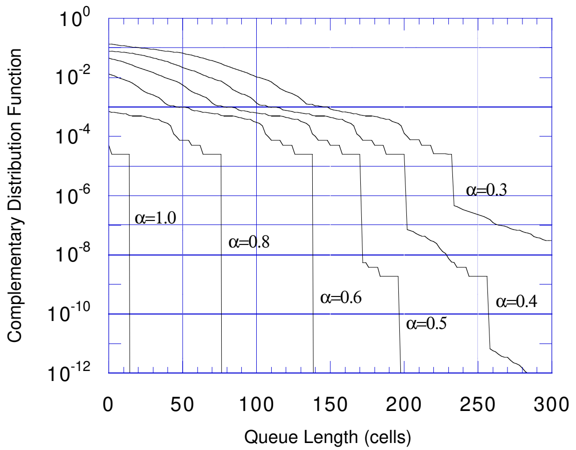

Figure 10: Complementary distribution function of the buffer occupancy.

Figure 10 shows the Complementary Distribution Function (CDF) of the queue length at the tagged station as a function of the α value. As expected, the buffer occupancy dramatically increases as the portion of bandwidth allocated to the tagged station decreases.

Probability	**Quantiles (α=1)**
$1-10^{-2}$	0
$1-10^{-4}$	0
$1-10^{-5}$	16
$1-10^{-6}$	16
$1-10^{-7}$	16
$1-10^{-8}$	16
$1-10^{-9}$	16
$1-10^{-10}$	16

Table 4: Quantiles of the buffer occupancy distribution for $\alpha = 1$.

An interesting point to analyze is the influence of the asynchronous traffic on the behaviour of the tagged station. Clearly, when $\alpha = 1$ (i.e., peak allocation) and the asynchronous traffic is disabled, it can easily be proved that the queue length, at embedding points, is null with probability 1. In fact, as embedding points are located just before video frame arrivals, this means that each video frame is completely transmitted before the next frame arrival. Hence, a buffer size equal to the maximum video frame size guarantees no cell loss.

On the other hand, Table 4, which reports quantiles of the buffer occupancy for $\alpha = 1$, highlights that, even under a peak allocation, when the asynchronous traffic is enabled an additional buffer is needed to reduce cell losses to an acceptable level.

6. Reliabilty of the video traffic model

As already stated, MPEG video sequences exhibit correlations over a wide range of time scales. On the other hand, the video traffic model used in our analysis only captures the short-term correlations introduced by the GoP pattern in the MPEG coding algorithm. Figure 10 highlights that the tagged station buffer occupancy increases significantly when α decreases. This means that, for low α values the influence of medium and long-term correlations can no longer be neglected. Thus, it can be reasonably expected that the traffic model provides optimistic predictions when α drops below a certain threshold.

In order to assess the range of α values where results provided by the MPEG video traffic model can be assumed as being reliable we estimated the tagged station buffer occupancy using “trace-driven” simulation. Specifically, we simulated the MetaRing network by again using the worst-case network model derived in Section 3.1. However, it was faded by the real “dino” sequence instead of the video traffic model.

A comparison of the quantiles of the buffer occupancy is shown in Table 5 and is limited to probability values up to $1-10^{-4}$ due to the limited size of the video sequence[5]. As can be seen, analytical results and simulative estimates agree perfectly for α values less than or equal to 0.5, start diverging when $\alpha = 0.4$, and completely disagree for lower α. Specifically, for low α values (i.e., $\alpha \le 0.4$), as expected, the video traffic model provides optimistic predictions.

[5]The lack of large sets of empirical data is another problem that makes the modeling of VBR video traffic difficult.

	"Dino" Trace					Model				
Probability	α=.9	α=.6	α=.5	α=.4	α=.3	α=.9	α=.6	α=.5	α=.4	α=.3
$1-10^{-1}$	0	0	0	0	28	0	0	0	0	22
$1-10^{-2}$	0	6	38	72	114	0	6	38	68	102
$1-10^{-3}$	0	48	80	135	470	0	48	80	110	148
$1-10^{-4}$	15	109	140	203	721	15	108	142	172	204

Table 5: Comparison of quantiles of the buffer occupancy obtained by analysis and trace-driven simulation.

The model reliability region, i.e., the range of α values where analytical results matches simulative estimates, can also be assessed by using the following methodology based on the busy time distribution of the tagged station buffer.

The busy-time is defined as the duration in which the buffer is not empty and, as such, it increases as α decreases. Dependencies between video frames separated by a time interval greater than the maximum busy-time can obviously be neglected by the video traffic model without affecting the accuracy of the analysis. In fact, the system regenerates itself before these dependencies can produce their effects.

Hence, from the estimation of the busy time distribution (by means of trace-driven simulation) it is possible to identify the temporal range within which correlations must be captured by the video traffic model in order to obtain reliable results.

Alternatively, given a video traffic model, it is possible to establish the minimum α value for which the video traffic model still provides useful information. This occurs until the maximum estimated busy-time is less than the range of correlation captured by the model

As the video traffic model used in our analysis only captures the short-term correlations introduced by the GoP pattern and the GoP size is equal to 12, according to the previous methodology, it still provides good results until the busy time (expressed in the number of consecutive video frame arrivals) is less than 12.

Quantiles of the busy-time distribution estimated by trace-driven simulation experiments for different α values are reported in Table 6. Time is expressed in numbers of video frame arrivals. Hence, a busy-time equal to zero indicates that the buffer becomes empty before the next video frame arrival.

Probability	α=.5	α=.4	α=.3
$1-10^{-1}$	0	0	1
$1-10^{-2}$	1	1	2
$1-10^{-3}$	1	2	5
$1-10^{-4}$	2	5	35
1	6	22	103

Table 6: Quantiles of the busy-time distribution for different α values.

On the basis of the above methodology, the results shown in Table 6 confirm that the minimum α value for which the traffic model still provides useful results is $\alpha = 0.5$.

Due to the low value of the N_{avg}/N_{max} ratio (i.e., average to peak bit rate ratio) exhibited by MPEG video sequences (in the order of 0.1 as shown in Table 2), setting $\alpha = 0.5$ probably does not absolutely identify the optimal bandwidth allocation scheme. Hence, more complicated MPEG video traffic models capable of capturing medium and long range correlations would be required to extend the analysis to lower α values. Unfortunately, due to their complexity, such models ([ROS94a], [ROS94c], [GAR93]) are not analytically tractable and hence are only useful for simulation.

7. Conclusions

This paper has focused on the problem of selecting the optimal bandwidth allocation scheme for MPEG video transport on a MetaRing LAN. The QoS (i.e. the queue length distribution) experienced by MPEG video users was derived as a function of the bandwidth allocation scheme adopted. As VBR video users only tolerate packet loss rates in the order of 10^{-9} and less, and to overcome the complexity of the MetaRing MAC protocol, the problem was addressed by analyzing a simplified "worst case" network model. This network model was faded by an MPEG video traffic model which only captures short range correlations. Hence, as soon the bandwidth allocation scheme is such that effects of long term correlations cannot be neglected the results obtained are no longer reliable.

To assess to what extent the video traffic model still provides useful information, a methodology based on the busy time distribution of the buffer occupancy was introduced.

According to this methodology, the MPEG video traffic model still provides accurate results until the portion of bandwidth allocated to each video source is greater than or equal to 50% of the peak rate. As the average to peak bit rate ratio exhibited by MPEG video sequences is in the order of 10% an analytically tractable MPEG video traffic model capable of capturing correlations over more time scales could be very useful to fully exploit burstiness in MPEG video sequences. This will be the subject of further study.

Appendix - Proofs of Claims

This appendix reports the proof of claims in Section 3.

Claim 3.1

In the scenario under consideration, if a video frame arrival occurs while the asynchronous transmission is enabled then station S_1 observes the first empty slot after a time interval

$$B_{AS} \le 2T_{RING} + d_{AS,1} \qquad [2]$$

since the video frame arrival. In [2] $d_{AS,1}$ is the distance (in slots) of station S_1 from the furthest asynchronous station ($AS_0^{(1)}$ in Figure 1).

Proof. *The claim is easily proved by observing that the unluckiest situation for station S_1 occurs when the ASINC_EN(GR) is released by station S_1 itself immediately before the video frame arrival. Hence, the ASINC_EN attribute will be changed to YL at most after T_{RING} time slots since the frame arrival. Another T_{RING} interval will, at most, elapse before all the asynchronous stations are inhibited. Eventually, the first empty slot will be observed by the first synchronous station (S_1) after a further time interval equal to $d_{AS,1}$.*

Claim 3.2

If the asynchronous traffic transmission is enabled when a video frame arrives then condition

$$\sum_{i=1}^{N} B_i \le T_v - 3T_{RING} \qquad [3]$$

implies that all the synchronous stations complete the transmission of their quota before the next video frame arrival.

Proof. *To prove the claim it is sufficient to consider the order in which synchronous stations access the medium. As stated in claim 3.1) S_1 begins to transmit B_{AS} time slots after the frame arrival. As soon as station S_1 has completed its quota it refrains from transmitting new cells. Thus, after a delay equal to $d_{1,2}$, empty slots are observed by station S_2 which can thus transmit its quota. Then, after time interval $d_{2,3}$ it is station S_3's turn, and so on. The tagged station, which is the last station to observe empty slots, should complete the transmission of its quota exactly $B_{AS} + \sum_{i=1}^{N} B_i + \sum_{i=1}^{N-1} d_{i,i+1}$ time slots after the video frame arrival. Hence, the tagged station actually transmits the whole quota of preallocated slots if the following condition holds*

$$B_{AS} + \sum_{i=1}^{N} B_i + \sum_{i=1}^{N-1} d_{i,i+1} \le T_v$$

or, equivalently

$$\sum_{i=1}^{N} B_i \le T_v - \left(B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \right).$$

Please note that conditions $B_{AS} \le 2T_{RING} + d_{AS,1}$ and $d_{AS,1} + \sum_{i=1}^{N-1} d_{i,i+1} \le T_{RING}$ together imply $B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \le 3T_{RING}$.

Hence, as condition $\sum_{i=1}^{N} B_i \le T_v - 3T_{RING}$, implies $\sum_{i=1}^{N} B_i \le T_v - \left(B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \right)$ the claim is proved.

Claim 3.3

If condition [4a] holds and $j = 0$ then the transient period lasts just one T_v period and during this interval the tagged station observes a number of empty slots given by

$$B'_N = B_N - \left(B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \right) \ge B_N - 3T_{RING} \qquad [5]$$

Proof. *To prove the claim we will make use of Figure 11 which shows the behaviour of each synchronous station. In this figure synchronous stations indices are reported on the vertical axis, while time is represented on the horizontal axis. Each station alternates between transmission periods (represented by solid lines) and periods during which it is not transmitting (dotted lines). A station may refrain from*

transmitting either because it is observing busy slots or because its quota has been exhausted.
At time instant t_0*, when the video frame arrival occurs, the asynchronous traffic transmission is enabled. Hence, as stated by [2], station* S_1 *observes the first empty slot after a time interval equal to* B_{AS}*. From that point in time onward, all the stations but the tagged one, in turn, transmit their quota of cells and then stop. The tagged station is still transmitting its quota when the next video frame arrives. Hence, as can be verified by Figure 11, the tagged station transmits only* $B'_N = B_N - \left(B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \right)$ *cells in the time interval* $(t_0,\ t_0 + T_v]$*. The inequality in [5] immediately follows from [2] (*$B_{AS} \le 2T_{RING} + d_{AS,1}$*) by observing that* $d_{AS,1} + \sum_{i=1}^{N-1} d_{i,i+1} \le T_{RING}$ *and hence* $B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \le 3T_{RING}$.

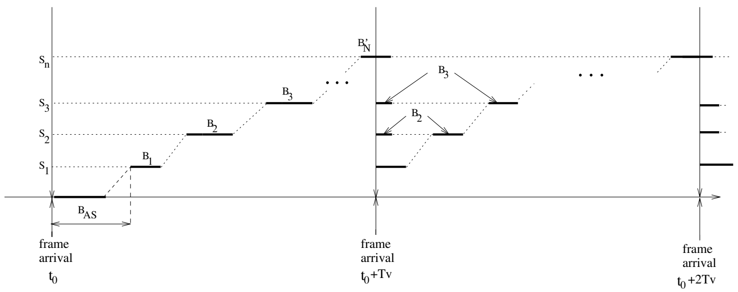

Figure 11: Synchronous station behaviour when $j = 0$ *and condition [4a] holds.*

To completely prove the claim it is now sufficient to show that, starting from time $t_0 + T_v$*, the tagged station achieves its quota of slots during each* T_v *interval. This can be accomplished by observing that at time* $t_0 + T_v$ *all the stations have cells in the buffer and the asynchronous traffic is disabled (due to the Thres=0 and r=0 assumptions). Furthermore, immediately after the same time instant, each station observes a number of empty slots equal to its distance from the closest upstream synchronous station. Hence, all stations experience an initial period of transmission as shown in Figure 11. Then stations from* S_2 *to* S_N *stop, while* S_1 *continues to transmit until its quota* B_1 *has been exhausted and then stops. Hence, empty slots are observed by station* S_2 *which can complete its quota* B_2*. Then it is station* S_3*'s turn and so on. Eventually empty slots are observed by the tagged station. The tagged station has a credit of* $B_N - B'_N$ *slots from the previous* T_v *period and thus, in principle, it should transmit* $(B_N - B'_N) + B_N$ *slots in the period* $(t_0 + T_v,\ t_0 + 2T_v]$*. In practice, due to [1] the tagged station only achieves* B_N *slots in this period and thus, the credit is traded to the following periods.*
Starting from time $t_0 + 2T_v$*, during successive* T_v *intervals each station behaves exactly like in the time interval* $(t_0 + T_v,\ t_0 + 2T_v]$ *unless the tagged station buffer becomes empty before the next video frame has arrived. The claim is thus proved.*

Claim 3.4
If condition [4b] holds and $j = 0$ then the transient period lasts two T_v periods. In the former period the tagged station observes no empty slots, while in the latter it achieves a number of slots equal to

$$B'_N = T_v - \sum_{i=1}^{N-1} B_i - d_{N-1,N} + \left(T_v - B_{AS} - \sum_{i=1}^{N-1} B_i - \sum_{i=1}^{N-2} d_{i,i+1} \right) \qquad [6]$$

Proof. *The claim that the tagged station observes no empty slot during the first* T_v *period follows directly from conditions [4b] and* $j = 0$.
To prove [6], observe that at time $t_0 + T_v$ *all the stations have cells in the buffer and the asynchronous traffic is disabled. Furthermore, immediately after time instant* $t_0 + T_v$ *each station but the tagged one observes a number of empty slots corresponding to its distance from the closest upstream synchronous station. On the other hand, it can be verified that the tagged station observes a number of slots equal to*

$$T_v - B_{AS} - \sum_{i=1}^{N-1} B_i - \sum_{i=1}^{N-2} d_{i,i+1} < d_{N-1,N} \qquad [6a]$$

Hence, from $t_0 + T_v$ *on, each synchronous station behaves as in Figure 11 with the only difference that, immediately after time instant* $t_0 + T_v$*, the tagged station observes a number of empty slots less than* $d_{N-1,N}$.
Specifically, stations from S_2 *to* S_N *experience a second period of transmission and it can be verified that, during this second period, the tagged station observes a number of empty slots given by*

$$T_v - \sum_{i=1}^{N-1} B_i - d_{N-1,N} \qquad [6b]$$

Furthermore, in the following T_v periods the tagged station always observes a number of empty slots equal to its quota until its buffer becomes empty before the next frame arrival.

Claim 3.5

If $B_i = B \ (i = 1,2,\ldots N)$, $\sum_{i=1}^{N} B_i = NB = T_v$ and B is such that condition [4a] holds with $j = 1$ then the number of empty slots observed by each synchronous station in successive intervals of duration T_v is as reported in Table 1 where

$$B' = T_v - B_{AS} - (N-2)B - \sum_{i=1}^{N-2} d_{i,i+1} \quad [7]$$

$$B'' = B' - d_{N-1,N} \quad [8]$$

Proof. *To prove claim 3.5 we will make use of Figure 11 which shows the behaviour of each synchronous station.*

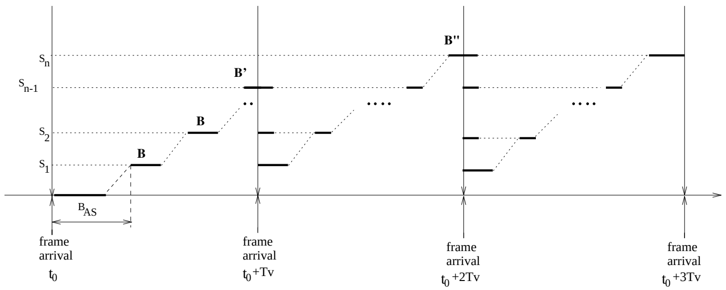

Figure 12: Synchronous station behaviour when the quota is the same for each synchronous station and such that condition [4a] holds with $j = 1$.

By hypothesis, S_{N-1} is the station that is still transmitting when the video frame arrives at time instant $t_0 + T_v$. Let B' be the number of empty slots observed by S_{N-1} during the interval $(t_0, t_0 + T_v]$. Figure 12 highlights that the following relation must be satisfied during interval $(t_0, t_0 + T_v]$

$$T_v = B_{AS} + (N-2)B + \sum_{i=1}^{N-2} d_{i,i+1} + B'$$

and hence [7] immediately follows.

Similarly, with reference to the interval $(t_0 + T_v, t_0 + 2T_v]$ the following relation must hold

$$T_v = (N-2)B + (2B - B') + d_{N-1,N} + B''$$

which implies [8] after some algebraic manipulations and taking into account that $NB = T_v$.

Eventually, since station S_{N-1} has no more credits from previous intervals, starting from the interval $(t_0 + 2T_v, t_0 + 3T_v]$ all the station transmit their quota. This situation lasts until the tagged station buffer becomes empty before the next frame arrival.

Claims 3.6

If the tagged station quota is

$$B_N \geq 3T_{RING} \quad [9]$$

and under the hypothesis that all the available bandwidth is allocated, i.e.,

$$\sum_{i=1}^{N} B_i = T_v \quad [10]$$

then conditions underlying claim 3.3 ([4a] with $j = 0$) hold.

Proof. *Observe that for $j = 0$ condition [4a] becomes*

$$\sum_{i=1}^{N-1} B_i \leq T_v - B_{AS} - \sum_{i=1}^{N-1} d_{i,i+1} < \sum_{i=1}^{N} B_i \quad [23]$$

Therefore we have to prove that [9] and [10] together imply [23]. The second inequality in [23] immediately follows from [10] by observing that

$$T_v - B_{AS} - \sum_{i=1}^{N-1} d_{i,i+1} < T_v = \sum_{i=1}^{N} B_i$$

To prove the first inequality note that [10] can be written as $\sum_{i=1}^{N-1} B_i = T_v - B_N$ and hence, by [9]

$$\sum_{i=1}^{N-1} B_i \leq T_v - 3T_{RING} \quad [24]$$

By claim 3.1 it is $B_{AS} \leq 2T_{RING} + d_{AS,1}$. Furthermore, Figure 1 highlights that $d_{AS,1} + \sum_{i=1}^{N-1} d_{i,i+1} \leq 3T_{RING}$. By merging together the latter two relationships we obtain

$B_{AS} + \sum_{i=1}^{N-1} d_{i,i+1} \le 3T_{RING}$ *and hence* $T_v - B_{AS} - \sum_{i=1}^{N-1} d_{i,i+1} \ge T_v - 3T_{RING}$. *By introducing this relationship into [24] it yields* $\sum_{i=1}^{N-1} B_i \le T_v - 3T_{RING} \le T_v - B_{AS} - \sum_{i=1}^{N-1} d_{i,i+1}$ *which proves the first equality in [23] and hence, the claim .*

References

[BLE96] N. Blefari-Melazzi, “Modelling and Analysis of MPEG Video Sources for Performance Evaluation of Broadband Integrated Networks”, International Zurich Seminar IZS’96, Zurich (CH), February 1996; also in Lecture Notes in Computer Science, No. 1044.

[CID93] I. Cidon, Y. Ofek, "MetaRing, a Full Duplex Ring with Fairness and Spatial Reuse", IEEE Transaction on Communications, Vol. 41, No. 1, January 1993.

[GAR93] M. W. Garrett, "Contributions Toward Real-time Services on Packet Switched Networks", PhD thesis, Columbia University, 1993.

[LAT94] G. Latouche, "Newton's Iteration for Non-linear Equations in Markov Chains", IMA J. of Numerical Analysis, 1994 (14), pp 583-598.

[LEG91] D. Le Gall, “MPEG: A Video Compression Standard for Multimedia Applications”, Communication of the ACM, Vol. 34, No. 4, pp. 46-58, April 1991.

[MOT94] P. Motta, "Integrazione dei Servizi nel MetaRing: Valutazione delle Prestazioni", (in Italian), Laurea thesis, University of Pisa, Pisa (I), 1994.

[NEU89] M. F. Neuts, "Structured Stochastic Matrices of M/G/1 Type and their Applications", Marcel Dekker, Inc., 1989.

[OFE94] Y. Ofek, “Overview of the MetaRing Architecture”, Computer Networks and ISDN Systems, Vol. 26, No. 6-8, March 1994, pp. 817-830.

[RAM88] V. Ramaswami, "A stable recursion for the steady state vector in Markov chains of M/G/1 Type", Stochastic Models, 4, pp. 183-188, 1988.

[ROS94a] O. Rose, M. R. Frater, "A comparison of models for VBR video traffic sources in B-ISDN", in IFIP Transactions C-24: Broadband Communications, II, pages 275-287, North Holland, 1994.

[ROS94b] O. Rose, M. Ritter, “MPEG-Video sources in ATM-Systems - A new approach for the dimensioning of policing functions”, Submitted to the International Conference on Local and Metropolitan Communication systems: LAN & MAN, December 1994, Kyoto, Japan.

[ROS94c] O. Rose, “Delivery of MPEG Video Services over ATM”, Institute of Computer Science Research Report Series 86, University of Würzburg, Würzburg (D), January 1994.

[ROS95a] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modelling in ATM systems", Proceedings of the 20th Annual Conference on Local Computer Networks, Minneapolis (USA), October 1995.

[ROS95c] O. Rose, "Simple and Efficient Models for Variable Bit Rate MPEG Video Traffic", Institute of Computer Science Research Report Series 120, University of Würzburg, Würzburg (D), July 1995.

[WU92] H. T. Wu, Y. Ofek, K. Sorhaby, "Integration of Synchronous and Asynchronous Traffic on the MetaRing Architecture and its Analysis", Proceedings of ICC’92, to appear on IEEE/ACM Transactions on Networking.

[WUE] MPEG video sequence available via anonymous ftp at IP address `ftp-info3.informatik.uni-wuerzburg.de`, file `/pub/MPEG/traces/dino.tar.gz`