Performance Evaluation of the FPRRA Framework for GEO Satellites in the Absence of Accurate Multimedia Traffic Prediction

Polychronis Koutsakis, Dimitra Vasileiadou, Charalabos Stamos

Abstract – In recent work [1] we have introduced and evaluated a fair and dynamic joint Call Admission Control (CAC) and Multiple Access Control (MAC) framework, for Geostationary (GEO) Satellite Systems, named Fair Predictive Resource Reservation Access (FPRRA). The framework was based on accurate videoconference and data traffic prediction, made decisions after taking into account the provider revenue, and was shown to be highly efficient. However, the accurate modeling of multimedia traffic is not always possible, especially in the case of video (i.e., not videoconference) traffic. In this paper we evaluate the performance of FPRRA in the absence of accurate video traffic modeling and we discuss the efficiency of the scheme in comparison to other efficient schemes from the literature.

Keywords: GEO Satellite Systems, Multiple Access Control, Call Admission Control, Video Traffic, Traffic Modeling

I. Introduction

GEO satellite systems are less complicated to maintain, in comparison with Low Earth Orbit (LEO) and Medium Earth Orbit (MEO) systems. The reason is that their fixed location in the sky requires relatively little tracking capability in ground equipment. In addition, their high orbital altitude allows them to remain in orbit longer than LEO and MEO systems, which operate closer to Earth. For these reasons, they have attracted significant attention as a part of the global communication infrastructure.

Despite their advantages, GEO systems also pose significant network design problems. These have to do with the long propagation delays (270 ms for a GEO satellite network), the limited bandwidth to be shared among many users, and limitation in power. For these reasons, a well-designed MAC protocol is of paramount importance, in order to efficiently allocate network resources. The need for efficient allocation will grow even further for next generation networks, which will have to cope with bursty multimedia users transmitting data from various types of applications. Still, as efficient as the MAC protocol design might be, it cannot suffice on its own. It needs to be combined with an equally well-designed CAC scheme, which will not only serve the traditional role of CAC mechanisms (i.e., to prevent traffic overload) but also to maximize the satellite provider’s profit without jeopardizing the Quality of Service (QoS) offered to multimedia users. Hence, in [1] we have proposed FPRRA, a combined MAC and CAC framework for GEO satellite networks which makes decisions after taking into account the provider revenue.

Our work considered an on-board processing (OBP) system [2]-[3], where the NCC is located onboard the satellite so that it takes the requesting earth station only one round-trip time (RTT) plus the processor/queuing processing time to receive the reply for its reservation request. The reason that in most current broadband satellite access systems the scheduler is on the ground has to do with the computational requirements being less restrictive [4], but in [1] we have shown that our framework adds low computational complexity, therefore the “ideal” choice of having the scheduler on-board can be implemented.

However, our work in [1] exploited our results in [5]-[6] where we had shown that MPEG-4 and H.264 videoconference traffic, respectively, can be accurately modeled with the use of a Discrete Autoregressive Model of order 1 (DAR(1) model) [7]-[8]. This accurate prediction is not possible for all types of video sequences, and even when it is, it often involves a higher degree of complexity (e.g., [9]) which would incur additional computational requirements for an OBP system.

For this reason, in the present work we study the efficiency of our scheme in the absence of accurate multimedia traffic prediction, i.e., we implement our DAR(1) modeling approach on MPEG-4 and H.264 encoded video traces, and we study the effect that the lack of modeling accuracy has on our scheme.

The paper is organized as follows. Section II presents our new work on traffic modeling for MPEG-4 and H.264 video traffic, as well as a brief presentation of our CAC and MAC framework, FPRRA (the interested reader can find the detailed presentation of FPRRA in
II. Fair Predictive Resource Reservation Access (FPRRA)

II.1. Video Traffic Modeling

A substantial portion of the traffic carried by emerging wireless networks will be traffic from video services [10], hence presenting a challenge for satellite system designers. Video packet delay requirements are strict, because delays are annoying to a viewer. Whenever the delay experienced by a video packet exceeds the corresponding maximum delay, the packet is dropped, and the video packet dropping requirements are equally strict. Therefore, a good statistical model can be very useful in evaluating network performance under various video traffic loads. For satellite networks, in particular, an accurate traffic model can significantly enhance the system’s ability of providing an effective resource allocation policy. Such a technique is imperative to be developed in order for the system to cope with the large propagation delay [11].

II.1.1. MPEG-4 Traffic Modeling

MPEG-4 is an ISO standard that leverages existing digital video content by supporting the MPEG-1 and MPEG-2 coding standards, and enables richer development of new digital video applications and services. An MPEG-4 scene is composed by a number of audiovisual objects, which can be either static or time-varying in nature [12]. The MPEG-4 standard is particularly designed for video streaming over wireless networks [13]-[14].

In the present work we have studied six different long sequences of MPEG-4 encoded videos, from the publicly available library of frame size traces of long MPEG-4 and H.263 encoded videos provided by the Telecommunication Networks Group at the Technical University of Berlin [15]-[16]. Table 1 presents the statistics for each trace.

<table>
<thead>
<tr>
<th>Movie</th>
<th>Mean Bit Rate (Mbps)</th>
<th>Peak Bit Rate (Mbps)</th>
<th>Standard Deviation (Mbps)</th>
<th>Initial Revenue Weights and (q) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silence of The Lambs</td>
<td>0.58</td>
<td>4.4</td>
<td>0.46</td>
<td>1.83 (q=50%)</td>
</tr>
<tr>
<td>(HQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formula 1</td>
<td>0.84</td>
<td>2.9</td>
<td>0.35</td>
<td>2.17 (q=25%)</td>
</tr>
<tr>
<td>(HQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simpsons</td>
<td>1.3</td>
<td>8.8</td>
<td>0.54</td>
<td>2.73 (q=5%)</td>
</tr>
<tr>
<td>(HQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Star Wars IV</td>
<td>0.05</td>
<td>0.94</td>
<td>0.09</td>
<td>1</td>
</tr>
<tr>
<td>(LQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silence of The Lambs</td>
<td>0.11</td>
<td>2.3</td>
<td>0.18</td>
<td>1.47 (q=80%)</td>
</tr>
<tr>
<td>(LQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Die Hard III</td>
<td>0.25</td>
<td>1.6</td>
<td>0.21</td>
<td>1.71 (q=60%)</td>
</tr>
<tr>
<td>(MQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We have investigated the possibility of modeling the traces with a number of well-known distributions (uniform, exponential, gamma, negative binomial, lognormal, weibull, geometric) and our results have shown that, unlike the modeling results on videoconference traffic in [5], the best fit for MPEG-4 video traffic is not achieved with the use of the Pearson type V distribution. More specifically, the best fit for modeling the I, P, and B frames of the MPEG-4 traces used in our study varies, depending on the trace; the lognormal distribution provides the best fit for P frame sizes of all the traces, but regarding the modeling of I and B frames’ sizes the best fits were found to be the Pearson V, gamma/negative binomial (the negative binomial distribution is a discrete analogue of the gamma distribution), and lognormal distribution, respectively. Also, this “best” fit was often found not to be significantly accurate in the majority (4 out of 6) of the traces.

For each one of the six videos under study we have used one of the three “qualities” with which they have been encoded in [15]-[16]: high quality (HQ), medium quality (MQ) or low quality (LQ). We have also experimented with other MPEG-4 traces from the same online library, without any qualitative change in our results. New video frames arrive every 40 msecs. The length of the videos varies from 20 to 60 minutes and the data for each trace consists of a sequence of the number of cells per video frame (we use packets of ATM cell size throughout this work, but our mechanism can be used equally well with other packet sizes, as the nature of our modeling results would not be altered at all).

A simple example on the fitting approach is provided below, with the use of the Pearson type V distribution, indicatively (the respective procedure is followed when another distribution is the best fit).

The Probability Density Function (PDF) of a Pearson type V distribution with parameters (α, β) is: $f(x) = x^{-\frac{\alpha+1}{\beta}} e^{-\frac{x}{\beta}} \Gamma(\alpha)$, for all $x>0$, and zero otherwise [18].
The mean and variance of a Pearson type V distribution are given by the equations:

\begin{align*}
\text{Mean} &= \beta/\left(\alpha - 1\right) \\
\text{Variance} &= \beta^2/\left[\left(\alpha - 1\right)^2\left(\alpha - 2\right)\right]
\end{align*}

from which \((\alpha, \beta)\) are computed for use in the Pearson V fit (the same procedure is followed, respectively, to compute the parameters of all the other distributions examined as possible fits).

The last column of Table I will be explained in Section II.2.

As mentioned above, the degree of accuracy of the fits varied, depending on the trace and the type of frames. However, even in the cases where the fit was quite accurate, the autocorrelation between successive video frames can never be perfectly "captured" by a distribution generating frame sizes independently, according to a declared mean and standard deviation, and therefore none of the fitting attempts can achieve perfect accuracy. Hence, similarly to our work in [5], we used the fitting results in order to build Discrete Autoregressive Models of order 1 (DAR(1) models) [7, 8] which inherently use the autocorrelation coefficient of lag-1 in their estimations; DAR(1) provides an easy and practical model based only on four physically meaningful parameters (mean, peak, variance and the lag-1 autocorrelation coefficient), therefore it uses parameters which are either known at call set-up time or can be measured without introducing much complexity in the network.

We build a separate model for each video frame type (I, P, B).

A Discrete Autoregressive model of order \(p\), denoted as DAR\((p)\)[8], generates a stationary sequence of discrete random variables with an arbitrary probability distribution and with an autocorrelation structure similar to that of an Autoregressive model. DAR\((1)\) is a special case of a DAR\((p)\) process and it is defined as follows: let \(\{V_n\}\) and \(\{Y_n\}\) be two sequences of independent random variables. The random variable \(V_n\) can take two values, 0 and 1, with probabilities \(1-\rho\) and \(\rho\), respectively. The random variable \(Y_n\) has a discrete state space \(S\) and \(P\{Y_n = i\} = \pi(i)\).

The sequence of random variables \(\{X_n\}\) which is formed according to the linear model:

\[X_n = V_n X_{n-1} + (1-V_n) Y_n\]

is a DAR\((1)\) process.

A DAR\((1)\) process is a Markov chain with discrete state space \(S\) and a transition matrix:

\[P = \rho I + (1-\rho) Q\]

where \(\rho\) is the autocorrelation coefficient, \(I\) is the identity matrix and \(Q\) is a matrix with \(Q_{ij} = \pi(j)\) for \(i, j \in S\).

Autocorrelations are usually plotted for a range \(W\) of lags. The autocorrelation can be calculated by the formula:

\[\rho(W) = E[(X_i - \mu)(X_{i+w} - \mu)]/\sigma^2\]

where \(\mu\) is the mean and \(\sigma^2\) the variance of the frame size for a specific video trace.

In our model the rows of the \(Q\) matrix consist of the probabilities of the best distribution fit \((f_0, f_1, \ldots, f_k, F_k)\), where \(F_k = \sum_{i < K} f_k\) and \(K\) is the peak rate. Each \(k\), for \(k < K\), corresponds to possible source rates less than the peak rate of \(K\).

We present, in Figures 1-6, some indicative results of our modeling approach. We use Q-Q plots [18], and we plot the 0.01-, 0.02-, 0.03-,… quantiles of the actual I, P and B video frames’ types versus the respective quantiles of the respective DAR\((1)\) models, for a superposition of traces. All values are in packets. As shown in the Figures, once again the accuracy of the model varies; in Figures 1, 2, 4 the model provides highly accurate results (the points of the Q-Q plot fall almost completely along the 45-degree reference line), in Figures 3, 6 only medium accuracy is achieved, while in Figure 5 the DAR\((1)\) model fails to emulate the behavior of the actual traffic.
Fig. 3. Q-Q plot of the DAR(1) model versus the actual video for the B frames of Die Hard III MQ traces, for 20 superposed sources

Fig. 4. Q-Q plot of the DAR(1) model versus the actual video for the I frames of Star Wars IV LQ traces, for 10 superposed sources

Fig. 5. Q-Q plot of the DAR(1) model versus the actual video for the P frames of Star Wars IV LQ traces, for 10 superposed sources

Fig. 6. Q-Q plot of the DAR(1) model versus the actual video for the P frames of Silence of the Lambs HQ traces, for 10 superposed sources

II.1.2. H.264 Traffic Modeling

H.264 is the latest international video coding standard. It was jointly developed by the Video Coding Experts Group (VCEG) of the ITU-T and the Moving Picture Experts Group (MPEG) of ISO/IEC. It uses state-of-the-art coding tools and provides enhanced coding efficiency for a wide range of applications, including video telephony, video conferencing, TV, storage (DVD and/or hard disk based, especially high-definition DVD), streaming video, digital video authoring, digital cinema, and many others [17].

In the present work we have studied four different long sequences of H.264 Variable Bit Rate (VBR) encoded videos, from the publicly available Video Trace Library of [19]-[20]. Table II presents the statistics for each trace.

<table>
<thead>
<tr>
<th>Movie</th>
<th>RES, GOP Size, B, F</th>
<th>Mean Bit Rate (Kbps)</th>
<th>Peak Bit Rate (Mbps)</th>
<th>Standard Deviation (Mbps)</th>
<th>Initial Revenue Weights and (q) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo Olympics [CIF, 16, 7, 28]</td>
<td>330</td>
<td>6.8</td>
<td>3.87</td>
<td>2.27 (q=20%)</td>
<td></td>
</tr>
<tr>
<td>Sony Demo [CIF, 16, 3, 42]</td>
<td>60</td>
<td>1.62</td>
<td>1.29</td>
<td>1.54 (q=75%)</td>
<td></td>
</tr>
<tr>
<td>Star Wars IV [CIF, 16, 7, 42]</td>
<td>31</td>
<td>0.74</td>
<td>0.48</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Terminator 2 [CIF, 12, 2, 48]</td>
<td>252</td>
<td>3.38</td>
<td>3.01</td>
<td>1.96 (q=40%)</td>
<td></td>
</tr>
</tbody>
</table>

As in the case of MPEG-4 video traffic, we have investigated the possibility of modeling the H.264 traces.
with a number of well-known distributions and our results have shown that, unlike the modeling results on videoconference traffic in [6], the best fit for H.264 video traffic is not achieved with the use of the Pearson type V distribution. Again, the best fit for modeling the I, P, and B frames of the H.264 traces varies, depending on the trace; the lognormal and gamma/negative binomial distribution, alternate in providing the best fits, which again are not significantly accurate. New video frames arrive every 33.3 msecs. The length of the videos varies from 10 to 70 minutes and the data for each trace consists of a sequence of the number of cells per video frame. The last column of Table II will be explained in Section II.2. Again, we built a separate model for each video frame type (I, P, B), and the degree of accuracy of our DAR(1) modeling approach varied, similarly to the cases presented in Figures 1-6 for MPEG-4 traffic. In order to avoid repetition, we do not present here these similar results.

II.2 The proposed CAC Scheme

In [22] we have proposed a new efficient CAC scheme for multimedia traffic transmission over wireless cellular networks. The scheme focused on H.263 videoconference traffic and web traffic, and was shown to provide significantly better results in comparison to the traditional “effective bandwidth” approach. The reason for which our scheme excelled was the use of a highly accurate H.263 videoconference traffic model in order to precompute various traffic scenarios and combine that knowledge with online simulation, in order to be able to make accurate decisions on the acceptance or rejection of a new call. The precomputation, along with the online simulation, was made based on the traffic parameters declared by the video sources at call setup. These parameters are used for the “identification” of the source as a user adopting a specific “mode”, i.e., a set of traffic parameters. Therefore, we used in our work each user’s declared set of parameters in order to examine the respective precomputed traffic scenario, based on our H.263 model for a source with such a set of parameters.

Similarly to our work in [22] we denote as “modes” for the satellite MPEG-4 video users the sets of traffic parameters presented in Table I, and for satellite H.264 video users the sets of traffic parameters presented in Table II. Hence, we use six “modes” for MPEG-4 video traffic and four “modes” for H.264 traffic. Each “mode” represents a specific quality of service that a user wishes to get. For example, it is clear from the trace statistics in Table I that a user adopting the “Star Wars IV mode” requires less bandwidth than a user adopting the “Die Hard III mode”, who, in turn, requires less bandwidth than users adopting the “Simpsons mode”, which has the highest mean and peak bit rate. Similarly, from the H.264 trace statistics in Table II, it is clear that the modes, in descending order of bandwidth demands are: {Tokyo Olympics, Terminator2, Sony Demo, Star Wars IV}. Users choose one of the six MPEG-4 or one of the four H.264 “modes” with equal probability (16.6% and 25%, respectively).

The CAC scheme proposed in [22] focused only on the system’s ability to accommodate a newly arriving user in terms of the total channel capacity which is needed for all terminals after the inclusion of the new user. In the case when the channel load, with the admission of a new call, was precomputed (or computed online) to be higher than the channel information rate, users were gradually degraded up to the point where the new call could be admitted. One parameter not included in our study in [22] was that in a real-life scenario, the decision of admitting or rejecting a new call in the network will be made by the provider not only based on the capacity needed to accommodate the call, but also on the revenue that the admission of the new call will provide. That is, in the case that the admission of a new call (and the subsequent increase in bandwidth utilization) can only be made with the degradation of a higher-paying customer who enjoys higher QoS, the CAC module should compute whether this is a profitable decision.

For this reason, in FPRRA we not only adopt the idea of the scheme in [22] for precomputation and online computation of various traffic scenarios and implement it for MPEG-4 and H.264 traffic, but we also assign and compute “revenue weights” for each one of the six MPEG-4 modes and the four H.264 modes, thereby differentiating them into different service classes. To define what the revenue weights should be, based on network congestion and the type of users present in the network at any given time, we use dynamic pricing. We utilize the formula for the demand function from [23], since it is implemented for different priority users, which fits our system’s assumptions. From [23], we derive:

\[ p_h = p_o + p_o^* \sqrt{-4\ln(q)/2}, \quad p_o \geq p_o^* \]  

where \( p_o \) is the price for a low quality user, \( p_o^* \) is the price charged to high quality users and \( q \) is the percentage of high quality users who accept dynamic pricing (i.e., they do not accept degradation and are willing to pay more for their calls during network congestion periods). Without loss of generality, we have set both for MPEG-4 and for H.264 traffic the value of \( p_o \) equal to 1 for the Star Wars IV mode (revenue weight=1); that is, for each type of encoding we have set the revenue weight equal to 1 for the mode with the lowest bandwidth requirements. With these \( p_o \) values, we then proceed to dynamically calculate the values of \( p_h \) for all the other modes, using Equation (6). The dynamic calculation is based on the value of \( q \) in every time interval of \( T=26.5 \) seconds, equal to the frame duration (as shown in the system parameters in Table III). The initial revenue weights, shown in Tables I and II, are calculated based on the \( q \) values for each mode; these values have been selected indicatively, based on the rationale that the modes with the highest bandwidth requirements will be the ones with the least “loyal” users (users who are willing to pay more in order to keep
transmitting at a high rate). Hence, for example, Table II shows that of the 25% of H.264 users who initially choose, on average, the “mode” Tokyo Olympics, 20% are willing to pay more in case of network congestion, in order to keep transmitting at this “mode’s” rate. Users who accept degradation are degraded once, to the immediate next mode, based on the algorithm that follows.

Still, \( q \) varies at any given time, depending on the traffic mix of the moment. This creates the need of the dynamic calculation of \( p_h \) with Equation (6). We have also conducted simulations for other values of \( q \) and other percentages of “mode” selection (i.e., not 25% for each “mode”). The change in the values had no influence on our results.

Our CAC scheme uses the traffic models presented in Sections II.1.1 and II.1.2 to precompute or compute online (if a non-precomputed scenario occurs) a number of traffic scenarios.

The current revenue \( R \) is computed as:

\[
R = \sum_i N_i \cdot W_i
\]

where \( N_i \) is the total number of video users of “mode” \( i \), and \( W_i \) is the revenue from each user of “mode” \( i \).

The logic of the CAC algorithm is that, when a new video user arrives, the system first checks whether it can be accommodated in terms of the total bandwidth which will be needed when the user is multiplexed with the existing users in the system. If this is not possible, the algorithm attempts to degrade the user, if the user accepts degradation. The rationale behind this decision is that the arrival of a new user should cause the minimum possible number of degradations to users who are already in the system, therefore it is preferable that the new user is accepted with degradation. If after degradation the acceptance of the call is still not possible, the CAC scheme will not degrade a higher priority user, but it will check all possibilities of degrading users of the same or lesser priority of the new call in order to accommodate it. However the new call will be accommodated only if its acceptance will lead to higher revenue; otherwise, even if the total bandwidth that will be used with the acceptance of the new call is larger than the bandwidth previous used, there is no reason to degrade a significant number of users and cause their irritation if the provider will receive no extra revenue. In the case that the new call does not accept any degradation, the attempt to degrade lesser or equal priority users who are already in the system is still made, and the new call is again accepted only if it leads to higher revenue.

II.3 The Proposed MAC Scheme

The Digital Video Broadcasting Return Channel Satellite (DVB-RCS) standard [24] develops a communication system for the return channel (uplink channel), i.e., the link from the user terminal to the network gateway. Due to the expected services features and large delay-bandwidth product of satellite networks, DVB-RCS represents a proper test bed for the proposed resource allocation schemes. The most important components of the DVB-RCS Network are: a) RCSTs (Return Channel Satellite Terminals), i.e., a generic access terminal, b) the NCC (Network Control Center), a device in charge of managing the access and bandwidth allocation for RCSTs, c) Gateways and Feeders, which are the elements that receive and transmit information outside the network [25].

As in [25], our proposed satellite medium access scheme is based on a Multi-Frequency Time Division Multiple Access (MF-TDMA) approach, according to which a carrier is divided in timeslots (grouped in frames and superframes). The MAC component of FPRRA can also be applied to a regular TDMA frame structure, however we chose MF-TDMA since it is an attractive framing scheme for satellite uplinks with low power terminals and small antennae [26] and MF-TDMA schemes are capable of providing efficient bandwidth utilization [25]-[27]. The system parameters are taken from [25] and are presented in Table III.

<table>
<thead>
<tr>
<th>SYSTEM PARAMETERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame Duration</td>
</tr>
<tr>
<td>Carriers</td>
</tr>
<tr>
<td>Slots/frame/carryer</td>
</tr>
<tr>
<td>Bytes/slot</td>
</tr>
<tr>
<td>System global rate</td>
</tr>
</tbody>
</table>

The NCC allocates to each active RCST a set of timeslots, each characterized by a frequency, bandwidth, start time and duration time. The DVB-RCS standard provides five allocation types: Continuous Rate Assignment (CRA), Rate Based Dynamic Capacity (RBDC), Volume Based Dynamic Capacity (VBDC), which is especially suited for bursty traffic [27], Absolute Volume Based Dynamic Capacity (AVBDC), which is a request similar to VBDC, with the difference that the last request from a RCST overwrites all previous AVBDC requests from the same RCST, and Free Capacity Assignment, which is not a true request, but allocates automatically the otherwise unused capacity without involving any signaling from the RCST to the NCC.

In FPRRA’s MAC scheme, in addition to the MF-TDMA frame structure, we adopt the idea that after all requests have been satisfied the rest of the bandwidth is distributed freely following a certain algorithm. This approach was proposed for the first time in the relevant literature in the CFDAMA scheme [28]-[29]. The algorithm implemented in our scheme is a simple round-robin assignment algorithm to all RCSTs which are currently active, and is fair to video terminals who have been allocated a smaller amount of bandwidth than other video terminals of the same priority type. This distribution of the remaining bandwidth is especially useful in allocating slots to video users, as the difficulty in providing them with adequate bandwidth due to their
frequent changes in bandwidth needs could be somewhat alleviated by their acquiring the unused channel bandwidth freely in a round-robin manner. Similarly to [29], user stations send their capacity requests embedded in the header of their packets. The free capacity distribution performed by the protocol brings the end-to-end delay performance at low loads close to that obtained with random access protocols, while the demand-based bandwidth allocation at the beginning of each frame guarantees the protocol’s stability, robustness and efficient utilization of transmission bandwidth at high loads.

Based on the models for single MPEG-4 and H.264 video traces and the respective DAR models of multiplexed traffic, we propose that the “burden” of traffic prediction for the RCSTs should fall on the NCC instead of the RCSTs. This idea was also proposed in [26] (for data traffic prediction) as a possibility for an OBP system, but in the simulations conducted in that paper the NCC was finally considered to be part of the terrestrial segment of the satellite network. To the best of our knowledge, our proposal in [1] is the first in the relevant literature to be made in the context of video traffic transmission. More specifically, we propose that the NCC should run a real-time simulation, both for single and for multiplexed video sources. Hence, based on the “mode” declared by the RCSTs at call establishment, the NCC does not need to wait for a request from the RCSTs every channel frame (which would arrive with a delay of more than 10 channel frames, due to the propagation delay). Instead, it can start allocating resources to the video terminals, by simulating the single source models with the sources’ mean rate as a simulation start point, and by computing the free slots in each channel frame; this can be done by using the DAR(1) models for multiplexed video traffic, and subtracting the estimated used slots from the total number of slots in the system. This subtraction gives the estimated number of free slots, which will be allocated in a round-robin manner to all active RCSTs. With this slot allocation scheme, the RCST will not need to send frequent requests to the NCC but it will only need to send a “corrective” AVBDC request every superframe (defined in our work as equal to 11 channel frames, to account for the propagation delay). The reason for sending this request will be for the RCST to help the NCC correct any mistakes (due to overassignment or underassignment of slots) of the models produced at the NCC via online simulation. After receiving the AVBDC request, the NCC will resume its simulation with the current RCST state (in terms of bandwidth requirements) as a start point.

In order to quantify the above description of our protocol, we denote by \( N \) the number of information slots in the system, \( L \) the number of active video stations, \( D_i(s) \) the amount of bandwidth that the NCC estimates as needed by the \( i \)th RCST at the start of frame \( s \), and \( A_i(s) \) the amount of bandwidth that the NCC assigns to the \( i \)th RCST at the start of frame \( s \) (i.e., the virtual request of the RCST). If \( N \sum_{i=1}^{L} D_i(s) < 0 \), i.e., if the NCC estimates that there will be no amount of bandwidth left after all video terminals are granted bandwidth equal to their predicted demands, then the use of the following equation ensures the fair sharing of the available bandwidth resources to all active video terminals:

\[
A_i(s) = N \cdot D_i(s) / \sum_{i=1}^{L} D_i(s) \tag{8}
\]

It should be noted here that, although generally \( D_i(s) \) refers to the estimation made by the NCC of each video terminal’s upcoming bandwidth requirements, it also refers, every 11 frames, to the “corrective” AVBDC request sent by the video terminals every superframe. With the use of Equation (8) for bandwidth allocation, our protocol also guarantees that, in the case of traffic overload, all users experience equal video packet dropping probability.

If \( N \sum_{i=1}^{L} D_i(s) > 0 \), i.e., if the NCC estimates that there will be an amount of bandwidth left after all video terminals are granted bandwidth equal to their predicted demands, then the round-robin allocation (which is max-min fair [21], [30] for packets of equal size) takes place by starting with the video terminal which was estimated by the NCC to have the smallest bandwidth demand, and continuing with the next terminals in an increasing order of estimated bandwidth demands. The major reason behind this policy is again fairness - we want to alleviate any mistakes caused by underassignment to a terminal due to a mistaken estimation of its actual larger bandwidth demands. This policy also serves in the case where the NCC has made correct estimations (or slight overestimations) for all users, as it helps to maximize the number of satisfied users.

### III. Simulation Results and Discussion

We use computer simulations (the code is written in C) executed on Pentium-IV workstations to study the performance of FPRRA. Each simulation point is the result of an average of 10 independent runs (Monte-Carlo simulation), each simulating three hours of network operation. All our results have been derived for 95% t-confidence intervals (constructed in the usual way [18]). Connection lifetimes are exponentially distributed with mean value equal to 180 seconds.

As in [1], we compare FPRRA with four other efficient schemes. A more detailed analysis of the advantages and disadvantages of each protocol can be found in [1]; the focus of the present work is on assessing FPRRA’s performance when an accurate and simple video traffic model is not available.
The first scheme we compare with was introduced in [27]. In [27], the authors propose a satellite MAC protocol based on MF-TDMA (i.e., they use a similar approach to that of the MAC component of FPRRA). After receiving the terminals’ requests, the scheduler determines the amount of resources which should be allocated to each terminal. The protocol, however, proposes the use of a CRA-type assignment for rt-VBR traffic such as the video traffic used in our work, with the difference that the assignment is fixed for the duration of the connection (no new negotiation is needed), and equal to the rt-VBR user’s peak transmission rate (declared at call establishment); the reason for this choice in [27] is that rt-VBR traffic has strict delay and packet dropping constraints, and no traffic prediction mechanism is provided in [27], hence leading to the “defensive” choice of peak cell rate assignment, which in turn leads to significant bandwidth waste, as the assignment is most of the time larger than the rt-VBR user’s actual needs. The scheme in [27] achieves good performance results, but takes into consideration a case where all rt-VBR users steadily offer 25% of their peak rate, and therefore Available Bit Rate (ABR) traffic can take advantage of the rest of the bandwidth assigned to a terminal, by statistical multiplexing done within terminals. If, however, the terminal does not need the remaining bandwidth for other types of traffic than the rt-VBR one, this bandwidth is lost; this choice leads to inferior performance, especially since almost all of the traces used in our study have a peak-to-mean ratio larger than 6 (only the Formula 1 trace has a ratio of 3.5), which means that the constant allocation of the peak rate to all video sources leads to significant loss of valuable bandwidth resources. This will also be shown through the simulation comparison of FPRRA and [27]. Finally, it should be noted that video traffic in [27] is generated with the use of a Markov Modulated Poisson Process (MMPP) model, whereas we use actual video traces in our work. For the simulation comparison of the two schemes, the MMPP traffic generation is substituted here by the real video traces.

The second scheme with which we compare is the Scheduled-Retransmission Multiaccess (SRMA) Protocol [31], which improved on the well-known Announced Retransmission Random Access (ARRA) [32] satellite MAC protocol. More specifically, in ARRA the frame is divided into a number of K slots and each slot is divided into a data slot and K minislots; minislots have a much smaller length than the data slot. When a terminal attempts to transmit in a data slot, it also sends a reservation request by choosing one of the minislots of that slot at random. If the data transmission is successful, the reservation that has been made through the use of the minislot will be cancelled; otherwise, the reservation made is valid and the terminal will transmit again its failed data packet in the data slot corresponding to the used minislot. Additionally, each frame begins with an additional group of K minislots which serve as the common minislot pool. SRMA achieved improved throughput performance results in comparison to ARRA by not using the common minislot pool (and thus saving this bandwidth for transmissions) and, instead, assigning a status vector to each packet to eliminate the reservation collisions from different slots in the frame. Two versions of SRMA were designed and evaluated in [1], one with a fixed frame duration (SRMA-FF) and one with a dynamic frame duration (SRMA-DF), containing a fixed-length Aloha frame and a variable-length reserved frame with length equal to the number of successfully reserved packets in the corresponding Aloha frame. Since SRMA-DF was shown to provide higher throughput and smaller packet delay results than SRMA-FF, we choose SRMA-DF for comparison with our framework.

The third scheme with which we compare is the Predictive Demand Assignment Multiple Access (PRDAMA) protocol, proposed in [26], which focuses on Internet data traffic; despite the fact that it is not optimized for video traffic, it is of interest due to its prediction algorithm, which is similar in nature to ours. We discuss this protocol and its efficiency later in this section.

The fourth scheme with which we compare is an “ideal” framework, as we want to compare FPRRA with a similar scheme in which the NCC would “magnetically” know, without any information exchange (therefore, no contention is necessary among video RCSTs), exactly what the video RCSTs’ bandwidth demands for the next video frame would be. Therefore, in the case of the “ideal” framework, the knowledge of the RCST’s needs is used both for the CAC and the MAC component of the framework; hence, both our proposed MAC scheme and CAC algorithm use the actual RCSTs’ bandwidth demands instead of the estimates that FPRRA uses, based on the video traffic models.

Figure 7 presents our simulation results for MPEG-4 traffic, for the average video packet dropping metric versus the system utilization. The results have been derived for all possible total numbers of video users in the system (1-11 for MPEG-4 traffic, as it has been shown from our simulation study and confirmed by our traffic modeling that for 12 users the average generated load is larger than the system capacity of 8 Mbps). When 11 users are present in the system, the average generated load (over the 10 simulation runs, with each user in each run choosing a “mode” with equal probability) has been found to be 94.9%, and as shown in our results, this is a load that the system cannot handle (the video packet dropping probability surpasses the 0.1% upper bound and the mean video packet delay surpasses the 0.6 seconds upper bound). Utilization indicates the traffic load normalized to the uplink capacity, e.g., a traffic load equal to 40% represents 40% of the 8 Mbps uplink capacity, i.e., 3.2 Mbps system throughput. As it is shown in the Figure, even with the use of a traffic model which provides only mediocre accuracy in its predictions, FPRRA outperforms the other three protocols from the literature, however it is clearly outperformed by the “ideal” framework. FPRRA can handle up to 60% system
load (corresponding to 8 video users present in the system) while at the same time satisfying the strict QoS requirement of maximum video packet dropping equal to 0.1%; the respective maximum system load which the “ideal” framework can handle is 72%, while SRMA-DF achieves only a 27% maximum throughput, [27] achieves a 34% maximum throughput and PRDAMA a 51% maximum throughput for the same QoS requirement. With the use of the CAC component of FPRRA the system is either “pushed” to its limits in terms of bandwidth utilization in order to allow as many users as possible into the network, or no change is made to bandwidth utilization because the acceptance of a new user is deemed non-profitable for the provider. FPRRA manages to achieve higher network utilization with smaller video packet dropping due to the combination of the “aggressive” CAC with the efficiency of the scheduling ideas included in the MAC component of our framework.

PRDAMA makes the allocation based on the RCST’s request at the previous channel frame. This choice in PRDAMA is not optimal for video traffic, the bandwidth demands of which can fluctuate significantly from one video frame to the next. Also, in PRDAMA requests need to be sent every channel frame (therefore bandwidth needs to be dedicated to them, and lost in terms of transmitting information) to the NCC, in order to check whether the NCC’s prediction was correct and enough resources were obtained.

b. For the allocation of any free bandwidth, we use in FPRRA a round-robin algorithm that starts from the RCST which has been allocated the smallest amount of bandwidth, whereas PRDAMA makes the allocation based on the positive varying trend of each station’s traffic, i.e., on the number of packets in the RCST’s buffer during the previous two video frames. Therefore, the authors in [26] attempt to estimate the traffic varying trend at each station based on the first- and second-order variations of traffic. This allocation method, which is optimized for data traffic as mentioned above, fails to capture the dynamics of video traffic. An important feature of common MPEG encoders is the manner in which frame types are generated. Typical video encoders use a fixed Group-of-Pictures (GOP) pattern when compressing a video sequence; the GOP pattern specifies the number and temporal order of P and B frames between two successive I frames. A GOP pattern is defined by the distance K between I frames and the distance M between P frames. In practice, the most frequent value of M is 3 (two successive B frames) while the most frequent values of K are 6, 12, and 15, depending on the required video quality (K=12 for the traces used in our study) and the transmission rate. Therefore, due to the cyclic GOP format, the lag-1 and lag-2 autocorrelation for the MPEG-4 traces are very low (this is true even for videoconference traffic, as shown in [1]); this means that an estimate which would be based on the immediately previous demands of video traffic is unreliable.

However, in order to make a comparison with PRDAMA which will be as fair as possible, and in order to use a method similar in principle with our approach of modeling each video frame type separately, we have implemented PRDAMA by using its free bandwidth allocation algorithm separately on I, P and B frames. The results presented in all our Figures have been derived for PRDAMA with the use of this approach. Hence, in our implementation of PRDAMA the positive varying trend of the station’s traffic is estimated based on the number of packets in the RCST’s buffer during the previous two video frames of the same type. This change improved significantly the results which were initially obtained by [26]. This was expected, given that the lag-1 autocorrelations and lag-2 autocorrelations for I, P and B frames, respectively, are higher than 0.7. The much higher autocorrelations of the traces when focusing separately on each video frame type lead PRDAMA to significant improvement both in the initial (predictive)
bandwidth allocation and in the free bandwidth allocation. Still, in many cases, the autocorrelation is not high enough to guarantee the accuracy of the estimation made by PRDAMA; this is the reason that FPRRA, which provides a relatively accurate estimation of video traffic in order to allocate the proper bandwidth for the current frame demands of the RCSTs, outperforms PRDAMA. The improvement of FPRRA over PRDAMA is not as large in this case as it was in [1], where the accuracy of the video traffic model helped FPRRA achieve a performance close to the “ideal” framework, but it is still substantial.

Regarding the following results, it should be noted that all the results concerning packet delays include the assumption that, after the packet is transmitted from the transmitter station to the satellite, the minimum additional delay of one hop (0.27 seconds) will be needed for the packet to be received by the receiver station (i.e., the 0.27 additional seconds are included in our results).

Figure 8 presents our simulation results for the average video packet delay versus the system utilization. The results are generally similar in nature with those of Figure 7. It should be noted once again that the performance gap between FPRRA and the “ideal” framework is much larger in the present work than it was in [1].

Figures 9 and 10 present the respective results for H.264 traffic (the maximum number of users that the system can accommodate in this case is 28). The nature of the results is the same and the comments made above regarding MPEG-4 traffic stand for these results as well. The only difference between Figures 7-8 and 9-10, respectively, is that for H.264 traffic our framework is able to handle up to 57% channel load in order to satisfy both the H.264 video QoS requirements (in maximum packet dropping and delay) while the “ideal” framework can handle up to 68% channel load; SRMA-DF can achieve just a 10% throughput, [27] achieves a 31% throughput and PRDAMA 49% throughput. The reason that the achievable maximum throughputs for all protocols are smaller than those in the case of MPEG-4 traffic is the higher burstiness (peak/mean) of the H.264 traces used in our study.

Finally, in Figure 11 we study the efficiency of FPRRA in terms of its fairness, by comparing our scheme with a case where no fair sharing is enforced, but instead the procedure is First Come First Served (FCFS). Although our framework was shown to be fair in [1], this comparison is important, as our CAC scheme introduces the possibility of degrading a number of video users in order to allow a new video user, not accepting degradation, into the network (if this is a profitable decision for the provider). Hence, in the absence of a very accurate video traffic model, we need to re-evaluate our framework’s fairness.

It is clear from the Figure that once again our fair scheduling scheme significantly improves the individual QoS of each video user in comparison to the FCFS approach, as no user is allowed to dominate the channel. Video packet dropping is almost uniform for video users of the same “mode”; when the average video packet dropping exceeds 0.1% (for a channel load larger than 60%) then the vast majority of video users experience video packet dropping larger than 0.1% in FPRRA; on the contrary, with the use of an FCFS procedure, some of the users experience large video packet dropping while the video packet dropping of other users is well below 0.1%, thus causing unfairness. The results presented in Figure 11 concern one of the H.264 “modes”, the “Terminator 2 mode”. Similar results were derived for the other H.264 and MPEG-4 “modes” as well.
Figure 11. Percentage of video users of the same mode who experience packet loss larger than 0.1% vs. System Utilization

IV. Conclusions

In this paper, we have re-evaluated the performance of our recently proposed (in [1]) Call Admission Control and Medium Access Control framework for multimedia traffic transmission over GEO satellite networks; the re-evaluation has been made for the case when bursty video traffic is present in the system and no accurate video traffic model is available. The CAC component of our scheme makes acceptance/rejection decisions about new calls based not only on the predicted capacity that the users will consume, but also on the revenue gained for the provider when degrading current users in order to accommodate new ones. The MAC component of our framework focuses on the fair allocation of bandwidth among video users. In order to implement both components, we have used a simple but moderately accurate video traffic model, in order to predict the future bandwidth needs of satellite users. In this way, the system can cope with the significant disadvantage of the large propagation delay. Our framework, FPRRA, is shown to excel in comparison to three other schemes of the literature, in accommodating high loads of video traffic.

References


Authors’ information
Dept. of Electronic and Computer Engineering
Technical University of Crete, Greece
e-mail: polk@telecom.tuc.gr, dvasileiadou@isc.tuc.gr, estamou@isc.tuc.gr

Dr. Polychronis Koutsakis was born in Hania, Greece, in 1974. He received his 5-year Diploma in Electrical Engineering from the University of Patras, Greece and his MSc and Ph.D. degrees in Electronic and Computer Engineering from the Technical University of Crete, Greece. From July 2006 till December 2008 he was an Assistant Professor (tenure-track) at the Electrical and Computer Engineering Department of McMaster University, Canada, with his research being funded by the National Sciences and Engineering Research Council of Canada (NSERC). In January 2009 he joined the Electronic and Computer Engineering Department of the Technical University of Crete, as an Assistant Professor. His research interests focus on the design, modeling and performance evaluation of computer communication networks, and especially on the design and evaluation of multiple access schemes for multimedia integration over wireless networks, on call admission control and traffic policing schemes for both wireless and wired networks, on multiple access control protocols for mobile satellite networks, wireless sensor networks and powerline networks, and on traffic modeling. Dr. Koutsakis has authored more than 90 peer-reviewed papers in the above mentioned areas, has served as a Guest Editor for an issue of the ACM Mobile Computing and Communications Review, as TPC co-chair for the 4th ACM WMUNEP 2008, as a project proposal reviewer for NSERC and for the Dutch Technology Foundation, and serves as Editor for a number of journals and as a TPC member annually for many major IEEE conferences in his research fields. He is a Voting Member of the IEEE Multimedia Communications Technical Committee (MMTC).

Dimitra Vasileiadou was born in Alexandroupoli, in 1987. She is a final year undergraduate student at the Department of Electronic and Computer Engineering at the Technical University of Crete. Ms. Vasileiadou is currently working on her final year thesis, which focuses on H.264 video traffic modeling. This is her first journal publication.

Charalabos Stamos was born in Athens, in 1987. He is a final year undergraduate student at the Department of Electronic and Computer Engineering at the Technical University of Crete. Mr. Stamos is currently working on his final year thesis, which focuses on Call Admission Control for wireless networks. This is his first journal publication.