User Satisfaction-Based Resource Allocation for GEO Satellites

C. Stamos*, D. Vasileiadou*, G. Koutsandria*, I. Spanou*, A. Vlachaki*, A. Lazaris† and P. Koutsakis*
*Department of Electronic and Computer Engineering, Technical University of Crete, Greece.
†Department of Electrical Engineering, University of Southern California, USA.
Corresponding Author Email: polk@telecom.tuc.gr

***Abstract*—In recent work we have introduced and evaluated a fair and dynamic joint Call Admission Control (CAC) and Multiple Access Control (MAC) framework, for Geostationary Orbit (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's revenue, and was shown to be highly efficient. In this paper we enhance FPRRA by talking into account the users' satisfaction for making scheduling decisions and we focus on its evaluation in the absence of accurate MPEG-4 and H.264 video traffic modeling. In addition, we discuss the efficiency of our proposed scheme in comparison to other efficient schemes from the literature.**

I. Introduction

Geostationary Orbit (GEO) satellite systems have attracted significant attention as a part of the global communication infrastructure. However, the long propagation delays (270 ms), the limited bandwidth, and the limitation in power make the need of an efficient MAC mechanism of paramount importance. In addition, a well-designed CAC scheme needs to be deployed in order for the provider to maximize its revenue while satisfying the users' Quality of Service (QoS) requirements. Hence, in [1] we have proposed FPRRA, a combined MAC and CAC framework for GEO satellite networks which makes decisions based on: 1) the model presented in our work in [2] for accurate modeling of multiplexed MPEG-4 videoconference traffic, and 2) the maximization of the provider's revenue. However, accurate traffic prediction is not always possible and even when it is, it imposes high computational complexity. For this reason, in [8] we presented a preliminary study of the efficiency of FPRRA in the absence of accurate multimedia traffic prediction. In this work, we first extend the study in [8] in order to focus on video traces with very high burstiness which leads to significantly worse modeling results compared to the modeling accuracy in [8]. Secondly, we introduce into FPRRA the notion of user satisfaction, formulated as user irritation, similarly to the work in [9] for cellular networks. In our study, however, we propose two different definitions than the ones in [9] for the Short-Term User Irritation Factor (SUIF) in GEO satellite networks, we compare their results and discuss their relation to our resource pricing approach. Thirdly, we use Jain's fairness index to evaluate the fairness of our framework, with the use of both SUIF definitions and we discuss the respective results.

II. Video Traffic Modeling

In this work, we study ten different long frame-size traces of both High Quality (HQ) and Low Quality (LQ) MPEG-4 and H.264 encoded videos taken from [3], [11]. In Table I, we present the statistics for each trace. We have investigated the possibility of modeling the traces with well-known distributions. Our results have shown that the best fit for MPEG-4 video traffic is achieved with the use of a) Lognormal for the P frames, b) Gamma/Negative Binomial for the I frames, and c) Lognormal for the B frames. Regarding the H.264 traces, the best fit for modeling the I, P and B frames was found to be the Pearson type V distribution. More importantly, this "best" fit for all of the traces under study was found not to be significantly accurate. Similarly to our work in [2] on videoconference traffic, we use the best distribution fit for each case in order to build a Discrete Autoregressive model of order 1 (DAR(1)) for each frame type that models the traffic more efficiently based only on 4 physically meaningful parameters (i.e. mean, peak, variance and the lag-1 autocorrelation coefficient). In Fig. 1, we present indicatively the modeling results of our approach for the I, P, and B frames of the LQ South Park MPEG-4 trace where we use Q-Q plots for assessing the fit. As shown in the figure, the modeling accuracy is only relatively good for the I frames, very mediocre for the P frames and very bad for the B frames, i.e, the points of the Q-Q plot fail to fall along the 45-degree reference line.

Trace Name (representing a service *mode*)	**Mean Bit Rate (Mbps)**	**Peak Bit Rate (Mbps)**	**Standard Deviation (Mbps)**	**Initial Revenue Weights and q value**
Tokyo Olympics (H.264, 16, 7, 22, *HQ**)	0.726	9.8	6.9	2.73 (q=5%)
South Park (MPEG-4, HQ)	0.68	8.6	0.49	2.52 (q=10%)
South Park (MPEG-4, LQ)	0.1	3.6	0.2	1.83 (q=50%)
Futurama (MPEG-4, HQ)	1	8.8	0.48	3.15 (q=1%)
Silence of the Lambs (MPEG-4,HQ)	0.58	4.4	0.46	1.96 (q=40%)
Silence of the Lambs (MPEG-4,LQ)	0.11	2.3	0.18	1.65 (q=65%)
Jurassic Park I (MPEG-4, LQ)	0.15	1.6	0.21	1.47 (q=80%)
Star Wars IV (MPEG-4, LQ)	0.053	0.94	0.09	1.23 (q=95%)
Silence of the Lambs (H.264, 16, 7, 48, *LQ**)	0.014	0.47	0.23	1.00
Star Wars IV (H. 264, 16, 3, 16, *HQ**)	0.714	7.84	7.8	2.18 (q=25%)

Table I: Trace Statistics. The numbers for the H.264 traces denote the GOP size, the number of B frames and the Quantization Parameters, respectively. Thick lines separate the traces into 3 mode groups ordered by decreasing quality. **We infer the quality of the traces based on the bandwidth requirements.*

III. Introducing User Irritation in FPRRA

A. Pricing - vs. Irritation- Based CAC

We proceed to use the modeling results from Section II (despite the lack of good prediction accuracy) in order to design a CAC mechanism that precomputes the bandwidth demands of various traffic scenarios based on the traffic parameters declared by the video sources at call setup. These parameters are used for the "identification" of the source as

978-1-4673-1239-4/12/$31.00 ©2012 IEEE

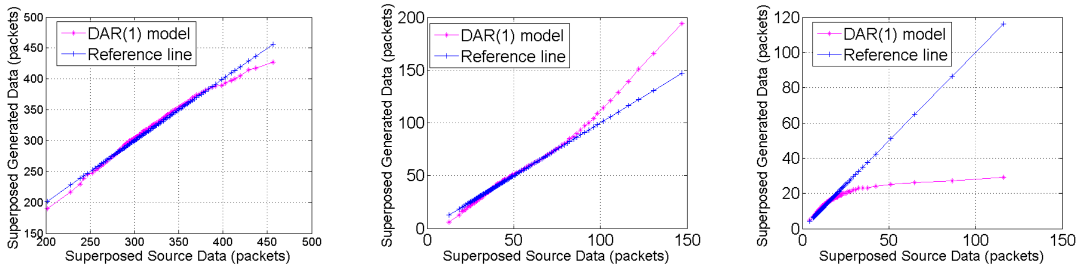

Figure 1: Q-Q plots of the 0.01, 0.02,, 1 quantiles of the DAR(1) model versus the respective quantiles of the actual video for the I, P, and B frames of the LQ MPEG-4 South Park trace, respectively. The results correspond to a superposition of 5 traces.

a user adopting a specific service "mode" (i.e. each row in Table 1). Users choose one of the ten "modes" with equal probability. For the computation of the profit, we assign and compute "revenue weights" for each one of the ten "modes". 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 calculated every T=26.5 seconds (i.e frame duration) using the demand function of [5]:

$$p_h = p_o + p_o\sqrt{-\ln(q)},\ p_h \geq p_o \qquad (1)$$

where p_o is the price for a LQ user, p_h is the price charged to HQ users and q is the percentage of HQ 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. By "degradation" here we refer to a "mode" being downgraded to the immediate next "mode" within one of the three groups noted in Table 1. In addition, we assume that users who accept degradation are degraded once. The current revenue R is computed as $R = \sum_i N_i W_i$, 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, with the use of the DAR(1) model, 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. 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.

B. *User Irritation- Based MAC*

Our proposed MAC component of FPRRA is based on a Multi-Frequency Time Division Multiple Access (MF-TDMA) approach. As in [1], the Network Control Center (NCC) should run a real-time simulation to predict the traffic volume from single and multiplexed videoconference sources. Hence, based on the "mode" declared by the terminals at call establishment, the NCC does not need to wait for a request from the terminals 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. In addition, the terminal will not need to send frequent requests to the NCC but it will only need to send a "corrective" request every superframe (defined as 11 channel frames, to account for the propagation delay) in order to help the NCC correct any mistakes. Similarly to the work in [9], we proceed to define SUIF and a Long Term User Irritation Factor (LUIF) to associate them with the bandwidth distribution, but in our case we give two different definitions of the SUIF. SUIF measures the delay that the user is ready to suffer prior to which the user decides to cancel a particular request, and LUIF determines the grade of irritation of the user resulting from continuous degradation. The two QoS metrics used in this work are that the video packet dropping probability should not surpass the 0.1% upper bound and that the mean video packet delay should not surpass the 0.6 seconds upper bound.

SUIF Definition 1: The Video packet transmission delay leads to packet dropping, which in turn leads to user irritation. So we can define SUIF for the $j-th$ user of mode-group 1 as $x_{1,j} = \tau \times P_{drop}$ where P_{drop} is the mean packet dropping probability (i.e. proportion of packets dropped) and $\tau < 1$ is the quantitative factor associated with irritation suffered due to a new or handoff call.

SUIF Definition 2: The decoding of P and B frames depends on the successful decoding of the I and P frames, respectively. Therefore, the successful transmission of an I frame is of paramount importance. Hence, we define SUIF as:

$$x_{1,j} = \tau \times P_{Thru\_P,B,GOP}, \text{where} \qquad (2)$$

$$P_{Thru\_P,B,GOP} = \begin{cases} \frac{P_{TRANS\_P,B}}{P_{GEN\_P,B}}, & \text{if } P_{drop\_I} \leq 0.1\% \\ 1, & \text{if } P_{drop\_I} > 0.1\%. \end{cases}$$

Thus, if the video packet dropping for the I frame within a Group-Of-Pictures (GoP) (i.e. a sequence of I,P,B frames) exceeds the 0.1% threshold, the transmission of P and B frames is of minor importance, since the basic information from the I frame is missing. If, on the other hand, the I frame has been transmitted, then we need to quantify the additional information that manages to be transmitted and without which GOP distortion and user irritation will increase. This is implemented via the calculation, in (2) of the ratio of the transmitted versus the total generated packets of P and B frames within a GOP. As in [9], we use an Exponentially

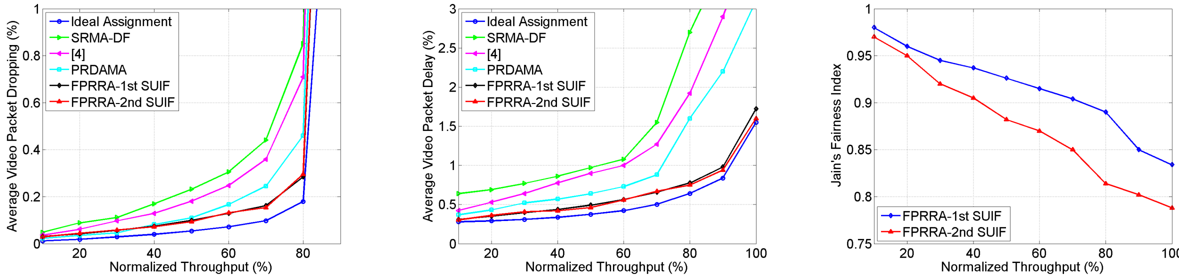

Figure 2: a) Average video packet dropping vs. System Utilization, b) Average video packet delay vs. System Utilization, c) Fairness Index vs. System Utilization.

Weighted Moving Average (EWMA) model to maintain continuous measure of the SUIFs for each user. By letting the stored LUIF be $U(k_{n-1})$, the LUIF to be computed be $U(k_n)$, and the current SUIF $U(x_i)$ be computed by either of our two definitions, then $k_n = \rho \times k_{n-1} + (1-\rho) \times U(x_i)$ where ρ is the weight assigned to the cumulative SUIF (0.2 in our case), k_n denotes the random variable used to measure the LUIF at the $n-th$ request and $U(x)$ is a Sigmoid function given by $U(x) = 1 - 1/(1+e^{-\alpha(x-\beta)})$ The value of α is set to 0.1, 0.3 and 0.9 for the groups which have three "modes" and 0.1, 0.3, 0.6, 0.9 for the group which has four "modes". The value of α indicates the users sensitivity to the QoS degradation, while β indicates the "acceptable" region of operation. Based on the above equation, we distribute the rest of the bandwidth by comparing the high quality users of each group in terms of their LUIF, and then continuing with the medium and low quality "modes". The remaining bandwidth is, each time, allocated to the user with the highest LUIF, in each quality, and the bandwidth distribution continues to the remaining users.

IV. Simulation Results and Discussion

We evaluate FPRRA by running Monte Carlo simulations (10 iterations with 95% t-confidence intervals), each simulating 3 hours of network operation. The connection lifetimes are exponentially distributed with mean 180 sec. The system parameters are [12]: frame duration = 26.5ms, 4 carriers, 128 slots/frame/carrier, 53 Bytes/slot, and 8 Mbps global rate. As in [1], we compare FPRRA with 4 other efficient schemes, from [4], [6], [7] and from an "ideal" framework, in which the NCC would "magically" know, what the video users bandwidth demands for the next video frame would be. These schemes do not take into account either pricing or user irritation, therefore they have the advantage over FPRRA that their only goal is the maximization of resource utilization. Fig. 2.a presents our simulation results for the average video packet dropping metric versus the system utilization. As shown in the figure, FPRRA outperforms the other three protocols from the literature, however it is clearly outperformed by the "ideal" framework. In addition, we can observe that the use of the 2nd SUIF definition leads to a fluctuation, in comparison to the results with the use of the 1st SUIF definition because the 2nd SUIF definition increases user "sensitivity" (i.e., irritation) in the cases where the loss of information is concentrated in specific GOPs, whereas the 1st SUIF definition "triggers" user irritation when packet dropping occurs anywhere in the video frames transmission. This is also confirmed by the results presented in Fig. 2.c. These results focus on fairness, based on the video packet dropping encountered by individual video streams when using each one of our two proposed SUIF definitions. The use of Jains fairness index [10] shows once again that the increased user irritation in the cases where the loss of information is concentrated in specific GOPs leads to a slightly smaller fairness (because videos experiencing this loss acquire larger portions of the free bandwidth). In addition, FPRRA achieves a high degree of fairness even for very high video traffic loads for both SUIF definitions. Finally, Fig. 2.b 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 2.a.

References

[1] P. Koutsakis, "Using Traffic Prediction and Estimation of Provider Revenue for a Joint GEO Satellite MAC/CAC Scheme", ACM Wireless Networks Journal, Vol. 17, No. 3, 2011, pp. 797-815.

[2] A. Lazaris, P. Koutsakis and M. Paterakis, "A New Model for Video Traffic Originating from Multiplexed MPEG-4 Videoconference Streams", Performance Evaluation Journal, Vol. 65, No. 1, 2008, pp. 51-70.

[3] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation", IEEE Network, Vol. 15, No. 6, 2001, pp. 40-54.

[4] A. Iuoras, P. Takats, C. Black, R. DiGirolamo, E. A. Wibowo, J. Lambadaris and M. Devetsikiotis, "Quality of Service-Oriented Protocols for Resource Management in Packet-Switched Satellites", Int. Journal of Sat. Comm., Vol. 17, No. 2-3, 1999.

[5] S.Yaipairoj and F.Harmantzis, "Dynamic Pricing with Alternatives for Mobile Networks", in Proc. of the IEEE WCNC 2004, Atlanta, USA.

[6] Z. Jiang and V. C. M. Leung, "A Predictive Demand Assignment Multiple Access Protocol for Internet Access over Broadband Satellite Networks", Int. Journal of Sat. Comm. and Netw., Vol. 21, No. 4-5, 2003, pp. 451-467.

[7] T-S. Yum and E. W. M. Wong, "The Scheduled-Retransmission (SRMA) Protocol for Packet Satellite Communications", IEEE Tran. on Info. Theory, Vol. 35, No. 6, 1989, pp. 1319-1324.

[8] P. Koutsakis, D. Vasileiadou and C. Stamos, "Performance Evaluation of the FPRRA Framework for GEO Satellites in the Absence of Accurate Multimedia Traffic Prediction", Int. Journal on Comm., Ant. and Prop., Vol.1, No. 1, 2011.

[9] S. Pal, M. Chatterjee and S. K. Das, "A Two-Level Resource Management Scheme in Wireless Networks Based on User-Satisfaction", ACM Mobile Computing and Comm. Review, Vol. 9, No. 4, 2005, pp. 4-14.

[10] R. Jain, "The Art of Computer Systems Performance Analysis", John Wiley&Sons, 1991.

[11] G. Van der Auwera, P. T. David, and M. Reisslein, "Traffic and Quality Characterization of Single-Layer Video Streams Encoded with H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension", IEEE Tran. on Broad., Vol. 54, No. 3, 2008, pp. 698-718.

[12] F. Chiti, R. Fantacci and F. Marangoni, "Advanced Dynamic Resource Allocation Schemes for Satellite Systems", in Proc. of the IEEE ICC 2005.