An Overview of Some Techniques for Cross-Layer Bandwidth Management in Multi-Service Satellite IP Networks

Nedo Celandroni*, Franco Davoli°, Erina Ferro*, Alberto Gotta°
*ISTI-CNR, Area della Ricerca del C.N.R., Via Moruzzi 1, I-56124 Pisa, Italy
nedo.celandroni@isti.cnr.it, erina.ferro@isti.cnr.it
°DIST-University of Genoa and CNIT (Italian National Consortium for Telecommunications),
University of Genoa Research Unit, Via Opera Pia 13, 16145 Genova, Italy
franco.davoli@cnit.it, alberto.gotta@cnit.it

Abstract. The main purpose of the paper is to present some cross-layer parametric optimization problems in bandwidth allocation over a rain-faded satellite channel, in a unified setting. The fade countermeasure regards bit and coding rate adaptation, and the traffic types addressed are both inelastic and elastic ones. Bandwidth allocation is performed in an adaptive fashion, according to fade and traffic variations. A numerical example for elastic traffic is presented and briefly discussed.

I. INTRODUCTION

Two main aspects characterize multi-service networks: the presence of sources with different statistical behavior, and a multiplicity of performance requirements. In general, especially in the Quality of Service (QoS) IP world, one main point associated with multiple services is the identification of some types of *flows*, with specific statistical and performance characteristics. A flow may be identified differently, according to the particular context: a 5-tuple (source and destination IP address, source and destination TCP/UDP port number, protocol) or a marking in the TOS (Type of Service) field in IPv4, the *flow label* in IPv6, a virtual pipe in MPLS. A very basic distinction that can be made is between inelastic (or guaranteed bandwidth) and elastic (a part of Best-Effort traffic, whose behavior is governed by TCP congestion control) flows. The presence of both traffic types in the network increases its heterogeneity and, as a consequence, complicates the task of network control, which is exerted to distribute the resources (in particular, the bandwidth), in order to achieve some desired goals. In wireless networks, where bandwidth may be relatively scarce with respect to cabled networks and adverse signal propagation conditions may affect channel quality, the dynamic control of allocated resources becomes a challenging task. In particular, satellite systems not only have to face variable load multimedia traffic, but also variable channel conditions and large propagation delays. The variability in operating conditions is due both to changes in the traffic loads and to the signal attenuation on the satellite links, owing to bad atmospheric events, which particularly affect the transmissions in the Ka band (20-30 GHz). It is therefore crucial to make use of adaptive network management and control algorithms to maintain the Quality of Service (QoS) of the transmitted data. The combined action among various layers of the network is likely to improve the performance of the overall system, through a coordination of control actions, QoS mapping and cross-layer information exchange for control purposes. However, architecting the layers, possibly with the adoption of ad-hoc solutions, becomes of paramount importance in a heterogeneous network, which may require the separation of the satellite component (e.g., by the introduction of *performance-enhancing proxies* or *relay entities*), in order to work effectively. Much care has to be taken in these approaches, in order not to disrupt wise layering principles, possibly leading to unstable system performance [1]. In general (though rarely in a cross-layer context), the dynamic control of resource allocation has been considered widely in the literature (e.g., [2-12] in the satellite environment and [13-17] in different settings).

The work summarized here addresses some specific dynamic resource allocation problems over a geostationary satellite link loaded with multi-service IP traffic. The main objective is to describe some classes of parametric optimization problems in this context, as well as possible solution techniques, in a unified setting. In all cases, we attempt to coordinate the actions taken on the satellite link at the physical layer (where the fade countermeasure technique is applied) with those done at the data link layer (where the satellite bandwidth is allocated), in a cross-layer fashion. Besides the cross-layer interaction, another feature characterizes our control architecture: a more or less hierarchical multilevel structure of the bandwidth allocation and Call Admission Control (CAC) strategies. These aspects will carry over the operative scenarios that we consider. Such scenarios may be divided into two categories: one assumes the presence of both *reserved bandwidth* (RB) flows and best-effort (BE) traffic, without explicitly characterizing the elastic behavior of the latter (as induced by TCP congestion control mechanisms); the other considers only best-effort services, but explicitly models the effects of TCP congestion control. RB flows may come from *Continuous Bit Rate* (CBR) calls (i.e., voice or video with fixed coding rate), aggregate traffic from DiffServ classes (to which a certain amount of bandwidth has to be dedicated, possibly between

MPLS-enabled routers), or from individual flows in an IntServ paradigm. In essence, all these categories may be characterized by bandwidth requests, with a statistical behavior that may cover a wide range of dynamics (from minutes to hours or days).

The paper will be organized as follows. In the next Section, we introduce the fade countermeasure and the traffic models adopted. In the third Section, we treat the first case of mixed inelastic and BE traffic. Section IV is devoted to the analysis of long-lived TCP connections, experiencing diverse fading conditions, and to their fair bandwidth allocation, performed by taking into account the tradeoff between Bit Error Rate (BER) and Information Bit Rate (IBR) upon TCP goodput. Section V presents and discusses an instance of numerical results, based both on analytical modeling and discrete event simulation. Section VI concludes the paper, also giving some directions for further research.

II. Fading and traffic models

A. Fading and fade countermeasure.

We consider a fully meshed satellite network that uses the Ka band (20-30 GHz) of a geostationary satellite transponder as a bent-pipe channel, and we counteract the fade attenuations of the signals, due to bad weather conditions, by applying adaptive FEC (Forward Error Correction) codes and bit rates. This means that the fade is countered by applying redundancy to the data before their transmission on the satellite channel, according to the detected attenuation level of the signal.

In many cases, it may happen that the attenuation experienced by each station is independent of the destination of its traffic; this is the case where the fading is of the up-link-predominant type, or when all the traffic sent by a station is addressed to destinations affected by the same environmental conditions. More generally, a station might be in the *multi-class* case, where the connections that utilize a given bandwidth portion may have different statistical parameters and peak rates, and they may be affected by different fading conditions, because they are addressed to destinations that experience diverse down-link attenuations.

Since the signal fade may vary in very short time intervals (even less than a second), in order to avoid too many oscillations in applying the fade countermeasures, according to each single fade level variation, the measured value of the signal attenuation is categorized in a class f, f=1,…,F, where F is the number of fade classes, equal for all stations. The countermeasure strategy adopted remains unchanged for all those levels of signal attenuation that belong to the same class. Thus, for each type of traffic with a given BER requirement, a fade class aggregates those fade levels that need the same data redundancy (in order to keep the BER below a given threshold), expressed at station i by redundancy coefficients $r_f^{(i)}$, f=1, 2,…, F. They represent the ratio between the IBR in clear sky and the IBR in the specific working condition. As non-guaranteed and RB (usually real-time) traffic have different QoS requirements in terms of BER, we indicate the respective redundancies with $r_{f,ng}^{(i)}$ and $r_{f,rt}^{(i)}$. The operative environment we refer to is a Multi-Frequency Time Division Multiple Access (MF-TDMA) system, i.e., a network where the total capacity of the satellite transponder is divided into carriers at different frequencies, each one accessed in TDMA. We also assume that a traffic station cannot transmit at different frequencies in the same temporal slot. It is worth noting that the inverse of the redundancy coefficients can be seen as multiplicative factors that express the bandwidth reduction due to the countermeasure adopted.

B. Traffic models and related performance indexes in the mixed (RB/BE) traffic case.

As regards RB traffic, we model bandwidth requests coming from the stations for flows (voice or MPEG4 video, bandwidth reserved for DiffServ aggregates, etc.), which may be carried within some specific DVB (Digital Video Broadcasting) service class [18]. The dynamics of interest are therefore at the connection level, and the relevant performance index is the steady-state blocking probability (P_{block}). We assume that blocked calls are lost (not re-attempted). For this type of traffic, we adopt the usual birth-death model with exponential distribution of call inter-arrival and duration times (Poissonian traffic). We assume that all connections of station i have the same peak rate $B^{(i)}$. If they also belong to the same fade class, we face a particular single-class case, where the expression of the blocking probability is given by the classical Erlang B loss formula [19]. At station *i*, given the Erlang traffic intensity $\rho^{(i)}$ and a desired upper bound on the blocking probability $\eta^{(i)}$, the maximum number of acceptable calls $N_{\max}^{(i)}$ can be easily derived. In this particular case, under a specific fading condition, requiring redundancy $r_{f,rt}^{(i)}$, the total bandwidth needed to carry the maximum number of calls at station *i* would be $B^{(i)} r_{f,rt}^{(i)} N_{\max}^{(i)}$.

In the case of multiple fade classes per station, the transmission of the data would require the simultaneous application of different redundancy values. In this multi-class case, the blocking probability would result from a stochastic knapsack problem [19]. The situation can be kept always in the single class case if the bandwidth is assigned separately *per fading* or *per traffic class* inside the earth station.

As far as the best-effort traffic is concerned, it was assumed in [7, 20] that it originates from non-real-time data packets that are fragmented into fixed-size cells (ATM or DVB [18]) before transmission on the satellite channel. This traffic stems from the aggregation of packet bursts, generated by a high number of sources, and may include TCP/IP short-lived connections and UDP/IP flows with no particular bandwidth reservation. At each station *i*, cells are queued in a finite buffer of capacity $Q^{(i)}$. In this context, the quantity of interest is the *cell loss probability* (P_{loss}) in the queue of station i. An approximate evaluation of this quantity has been used in [7, 20], under a discrete-time self-similar traffic model, derived from [21]. This model represents the superposition of on-off sources, whose active periods (bursts) have Pareto-distributed 'on' time. In our case, we have to take into account that the extraction rate is determined by the residual capacity $C_{nrt}^{(i)}(t)$, out of the total capacity $C^{(i)}$ allocated to station i, available for the non-guaranteed traffic, after serving all RB connections in progress at the required transmission rate (as determined by the peak or by an equivalent bandwidth figure). In general, the residual bandwidth is a random variable; as a consequence, the loss probability at fixed capacity can be considered only as conditional on the number of connections in progress, and its average must be computed with respect to the statistics of the Markov chain that describes the connection dynamics. Thus, the performance measure for BE traffic is the average P_{loss}. Its expression may be different, according to the presence of a single-class (i.e., single redundancy coefficients) per station or of multiple classes (i.e., multiple redundancy coefficients) per station. The first case has been treated in [7, 20]. The multi-class case, so far, has been considered in [22], where, however, no BE traffic is present.

The dependence on the redundancy factors, which are time-varying quantities, deserves a further comment. In fact, all previously discussed calculations regarding the performance indexes have considered the current values of these coefficients as lasting forever. Whereas this is true in a "static" case (where, for the purpose of bandwidth allocation, the redundancies are treated as unknown random values with known probabilities), it is not in the "dynamic" case, where they are recomputed at each change of fade class. The resulting dynamic control scheme is thus a sort of "open-loop feedback" repetitive control [23], where the initial time continuously shifts ahead.

C. Traffic model for long-lived TCP connections

In this sub-section we outline the model used to derive analytical expressions of the goodput of long-lived TCP connections, which is the performance measure to be used in the bandwidth allocations to be considered in Section IV. For the sake of simplicity, we refer to a single satellite link.

In order to avoid time consuming simulations, reasonable estimations can be constructed for the goodput of a TCP Reno agent. A first relation that can be used is the one taken from [24], which is estimated for infinite bottleneck rate, and thus it is valid far apart the approaching of the bottleneck rate itself. Let μ be the bottleneck (the satellite link) rate expressed in segments/s, n the number of TCP sources (which experience the same delay and get an equal share of the link [25]), and d the delay between the beginning of the transmission of a segment and the reception of the relative acknowledgement, when the satellite link queue is empty. Moreover, let us assume the segment losses to be independent, with rate q. We have $d = c_l + 1/\mu$, where c_l is the channel latency. The TCP connections that share the same link also share an IP buffer, inserted ahead of the satellite link, whose capacity is at least equal to the product μd. By exploiting the expression of the send rate derived in [24], a slightly modified expression of the relative goodput T_g (normalized to the bottleneck rate) has been used in [25]. This relation is rather accurate for high values of q, i.e., far apart the saturation of the bottleneck link. For low values of q, it was found by simulation that, given a fixed c_l, for fixed values of the parameter $y = q\left(\mu/n\right)^2 d^5$, the goodput has a limited variation, with respect to individual variations of the parameters q, μ and n. Simulation results have been obtained for the goodput estimation, with a 1% confidence interval at 99% level, over a range of values of μ/n between 20 and 300, and n between 1 and 10. For $0 \le y \le 1$, goodput values corresponding to the same y never deviate for more than 8% from their mean. We then interpolated such mean values with a 4-th order polynomial approximating function, whose coefficients have been estimated with the least squared errors technique [25].

In essence, by combining the analytical result (for $y > 1$) and the numerical approximation (for $0 \le y \le 1$), it is possible to obtain reasonably accurate analytical expressions of the goodput of a connection, over a wide range of values of the parameter y.

Assuming to operate on an Additive White Gaussian Noise (AWGN) channel (a reasonable approximation for geostationary satellites and fixed earth stations), the segment loss rate q can also be expressed analytically [26], as a function of the BER p_e, of the segment length in bits l_s, and the average error burst length (ebl) l_e. These quantities have been plotted versus E_c / N_0 (channel bit energy to one-sided noise spectral density ratio) in [26]. In order to make q computations easier, we interpolated such data, and expressed p_e and l_e analytically, as functions of the coding rates considered and the E_c / N_0 ratio [25]. The TCP goodput relative to the bottleneck rate is a decreasing function of the segment loss rate q, which, in its turn, is a decreasing function of the coding redundancy applied in a given channel condition C / N_0 (carrier power to one-sided noise spectral density ratio, related to E_c / N_0) and for a given bit rate b_r. As already noted, the combination of channel bit rate and coding rate gives rise to a "redundancy factor", which, in this case where no ambiguity is possible, will be indicated simply by r_f. The absolute goodput of each TCP connection $\hat{T}_g$ is obtained by multiplying the relative value by the bottleneck rate, i.e., $\hat{T}_g = T_g(\mu/n) = T_g(1/n)\cdot(C/r)$, where C is the link rate in segments/s in clear sky conditions.

III. Bandwidth allocation and CAC in the presence of RB and BE flows

A number of policies have been derived for resource sharing in Call Admission Control. Ross [19] provides a classification and an extensive discussion about alternative different solutions. The simplest allocation policy is the *complete sharing (CS)* one, where connections are admitted simply if sufficient resources are available at the time of the request, without considering their importance. As an almost opposite situation, in the set of policies of *complete partitioning* (CP) type, each class of traffic is allocated a set of resources that can be used only by that class. Actually, optimal approaches should be based on the application of Markov Decision Processes, given a certain cost functional to be minimized (or maximized) as a performance index; however, they must take into detailed account any allowable network state and state transition, which is impractical even for networks of modest complexity. The functional form of the optimal policies is usually unknown. Therefore, a set of generally suboptimal policies with fixed structure (often described by a set of parameters), have been developed, which are simpler to implement and, in some special cases, do correspond to the optimal one. Obviously, once one of such fixed-structure policies has been selected, parametric optimization can be adopted, in order to choose the "best" values of parameters that minimize a given cost function (or maximize a performance index). This is the approach we have taken in [7, 20, 22]. Moreover, whereas the presence of best-effort traffic is most often neglected in this case, we have taken it explicitly into account in [7, 20], when designing the cost function that determines the bandwidth partitioning, which is based on the performance indexes P_{block} and P_{loss}, defined in II.*A*.

Such allocation policies fall in the CP category, but it is important to highlight that, as the partitions are *adaptively* changed in response to traffic or fading variations, they try to match the traffic load and channel conditions as closely as possible. In this respect, the philosophy of this approach is much in the same line as that of [17]; however, the major novelty introduced is in the cross-layer optimization, which stems from the explicit presence of the redundancy factors representing the fade countermeasure in the cost function.

As we mentioned, all control schemes we are considering exhibit a hierarchical structure. A lower level performs CAC (and implicit bandwidth allocation between CBR and BE traffic) locally at each station within a given bandwidth, whereas a higher one (master) is devoted to bandwidth allocation among the earth stations. The lower level allocation algorithm acts with a faster timing than the upper one, and it is located at each earth station *i*. It shares the bandwidth $C^{(i)}$, allocated to station *i*, between guaranteed ($C_{rt}^{(i)}$ [bits/s]) and non-guaranteed ($C_{nrt}^{(i)}$ [bits/s]) traffic, by imposing a constraint on the call blocking probability. It performs Call Admission Control (CAC) of the incoming guaranteed calls and measures the statistics necessary for successive allocations.

Given the bandwidth $C^{(i)}$, allocated to station *i*, the lower level optimization problem evaluates the threshold $N_{\max}^{(i)}$, so that the call blocking probability be, if possible, below $\eta^{(i)}$. It is worth noting that, owing to the dynamic fade changes, there may be time intervals where the assigned bandwidth $C^{(i)}$ is insufficient to carry the currently ongoing number of CBR connections in the station (i.e., $C^{(i)} < B^{(i)} r_{f,rt}^{(i)}(t) n^{(i)}(t)$); since we are considering *inelastic* traffic, in such situations one or more ongoing calls would be dropped. This is a situation over which we have no direct control (except immediate signaling to the master station in the dynamic case), and a good design should minimize its probability.

As regards the master station, the bandwidth allocation is formulated as an optimization problem in a discrete setting (with the assignment's granularity determined by the minimum bandwidth unit – *mbu*); if the performance index is a separable function of the stations' parameters (e.g., a sum of terms, each depending only on the bandwidth to be assigned to a station), the problem can be solved numerically by applying Dynamic Programming over the stations [7, 20]. Moreover, within a cross-layer approach that takes into account fade countermeasures based on adaptive code and bit rate (ultimately, redundancy) assignment, the allocation can be made adaptive. On the basis of the measured fading attenuation values, a receding-horizon open-loop feedback control problem is stated, with the optimization performed on-line upon request from the stations. Versions of this problem that take explicitly into account the different losses caused by UDP and TCP traffic [31] have also been considered in [32].

Finally, it is worth noting that these model-based approaches can be by-passed by using a fluid approximation and by treating the bandwidth partitions as continuous variables. A gradient descent technique can be adopted, in conjunction with Infinitesimal Perturbation Analysis (IPA) for gradient estimation [33-35]. The advantage of these technique is that they are measurement-based and do not require any functional form of the performance index.

IV. Bandwidth allocation for TCP flows

The application of adaptive FEC techniques was investigated in [26], to optimize the efficiency of TCP connections when transmitted over rain-faded geostationary satellite channels, with fixed user antennas, as in the environment that we

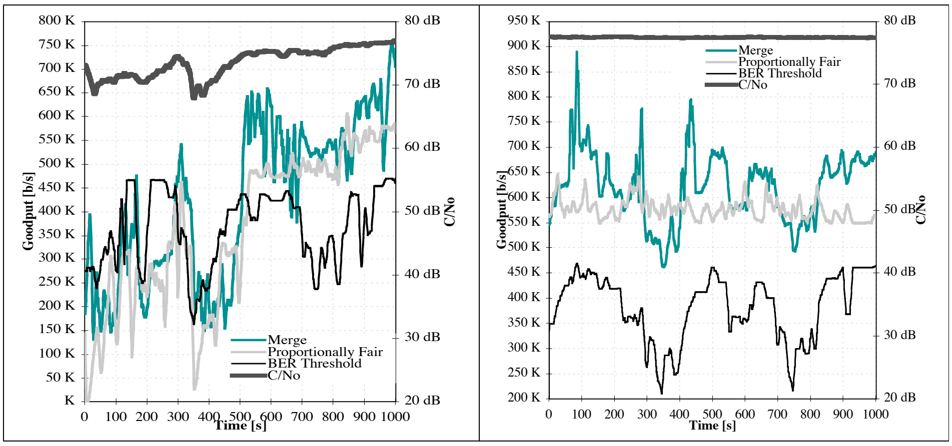

Fig. 1. Sample traces over 1,000 s simulation of the goodput (averaged over a 10-sample window), under different allocation strategies, for a faded (left) and a clear-sky (right) class.

consider now. We have applied the same philosophy in [25], by trading the bandwidth of the satellite link for the packet loss rate due to data corruption. The FEC techniques adopted do not interfere in any way with the normal behavior of the TCP stack, as they are applied just before the transmission over the satellite link. It has been shown in [26] that, given an available radio spectrum, antenna size, and transmission power, the selection of an appropriate modulation scheme and a FEC type allows choosing the BER and the IBR of the link that maximize the goodput of a TCP connection. This optimization can be done for different channel quality conditions caused, for example, by atmospheric events. The optimal transmission parameters, for each channel condition, can be reported in look-up tables and then applied in an adaptive fashion.

If connections take place within different source-destination (SD) pairs over satellite links, they may be generally subject to diverse fading conditions, according to the atmospheric effects at the source and destination stations (i.e., on the up- and down-links). We refer to connections on the same SD pair, which experience a specific channel condition, as belonging to the same "class"; they feed a common buffer at the IP packet level in the traffic station, which "sees" a transmission channel with specific characteristics (that may differ, in general, from those of other SD pairs originating either from the same or from other stations). The bandwidth allocated to serve such buffers is shared by all TCP connections in that class, and, once fixed, it determines the "best" combination of bit and coding rates for the given channel conditions. The goal of the allocation is to satisfy some global optimality criterion, which may involve goodput, fairness among the connections, or a combination thereof. Therefore, in correspondence of a specific channel situation, determined by the various up- and down-link fading patterns, and a given traffic load, we face a two-criteria optimization problem, whose decision variables are the service rates of the above mentioned IP buffers for each SD pair, and the corresponding transmission parameters. We have referred to these allocation strategies as TCP-CLARA (Cross Layer Approach for Resource Allocation), and we have analyzed a few different criteria. In all cases, the indexes chosen for the performance evaluation of the system were the TCP connections' goodput and the fairness of the allocations. The optimal allocations were derived numerically on the basis of the analytical model described in II.*C*, under different fade patterns. The different strategies have been compared first in a static fading scenario and then in a dynamically varying one [25], with fading traces taken from real-life samples. In this case, the allocation is applied adaptively, following the fading and traffic variations, similarly to the dynamic situation considered in the previous section.

V. A Numerical Example

In this Section, we present one among the numerous numerical results that have been obtained in the analysis of the above-mentioned techniques. We refer to the cross-layer optimization for the achievement of the "best" compromise between the maximization of TCP goodput and fairness. The numerical details are the same as in [25], with a combination of long-lived TCP NewReno connections, sharing various bottleneck links, determined by 10 different fading classes, under the HB6 link budget [32] and real fading traces. Fig. 1 shows the behavior of the overall goodput for two classes operating under different fading conditions. The "instantaneous" goodput is determined by the dynamic bandwidth and redundancy allocation, which aims to counter the fading effects, and to achieve a compromise between maximizing the total goodput and maintaining fairness among the connections. The strategies indicated attempt to do so in different ways (all described in detail in [25]): the “merge” strategy is the best choice between two alternative methods (“tradeoff” and “range”, respectively) that establish a balance between goodput and fairness; the “proportionally fair” one maximizes the sum of the logarithms of the individual goodputs, so as to attain a Nash Bargaining Solution (NBS); the “BER threshold” simply adjusts the redundancy to always keep the BER below a given limit, and assigns the bandwidths proportionally to the redundancy and the number of connections of each class.

The average values of the goodput, over all classes are reported in Table I below. For comparison, also the values referring to TCP Westwood+ [33] with BER threshold adaptation are shown. It can be noted that the gain achieved by the optimized cross-layer strategies (merge, range, tadeoff, proportionally fair) in conjunction with NewReno is superior to that of Westwood+, if the BER threshold imposed to the latter is 10^{-6}. This is, indeed, a relatively favorable BER for operating with NewReno; on the hand, Westwood+, which is much more robust with respect to the BER, takes advantage of the higher achievable IBR, in the presence of a less stringent BER constraint (10^{-5}). Thus, the cross-layer optimization (tailored on NewReno, because of the model we have used) guarantees a performance that is almost comparable to that of a modified TCP, working under a favorable BER. The main advantage of working with NewReno is that no prior modification is necessary on the end-to-end devices.

TABLE I. TOTAL GOODPUT AVERAGED OVER ALL CLASSES

Strategy		Goodput [Mbit/s]	Gain %
Merge		5.921	33.39
Range		5.858	31.97
Tradeoff		5.437	22.48
Proportionally Fair		5.412	21.91
	BERthr1e-6	4.626	4.21
Westwood+	BERthr1e-5	6.491	46.22
	BERthr1e-4	6.307	42.09
Reno BERthr1e-6		4.619	4.05
Reno BERthr1e-7		4.439	0.00

VI. CONCLUSIONS

The paper has dealt with resource allocation in multi-service satellite networks, by adopting a general cross-layer approach, where fade countermeasures (in terms of bit and coding rate adaptation) at the physical layer are combined with adaptive bandwidth assignment, in order to optimize some specified figures of merit, for a given situation of the transmission channels and of the traffic sources. In this context, a few optimization problems have been addressed, even with different specific techniques. The two main scenarios considered have been that of mixed RB and BE traffic classes, and that of elastic TCP connections, established over channels with different fading conditions. In the second scenario, we have presented a numerical example to highlight some characteristics of the dynamic cross-layer allocation.

ACKNOWLEDGMENT

This work was partially supported by the Ministry of Education, University and Research (MIUR), in the framework of the “*IS-MANET*” project, and by the European Commission in the framework of the NoE project “*SatNEx*” (contract no. 507052).

REFERENCES

[1] V. Kawadia, P. R. Kumar, “A cautionary perspective on cross layer design", *IEEE Wireless Commun. Mag.*, vol. 12, no. 1, pp. 3-11, Feb. 2005.

[2] S. Alouf, E. Altman, J. Galtier, J-F Lalande, C. Touati, “Quasi-optimal resource allocation in multi-spot MFTDMA satellite networks”, in M. Cheng, Y. Li, D.Z. Du (Eds.), *Combinatorial Optimization in Communication Networks*, Kluwer Academic Publishers, 2005.

[3] F. Alagöz, D. Walters, A. AlRustamani, B. Vojcic, R. Pickholtz, “Adaptive rate control and QoS provisioning in direct broadcast satellite networks”, *Wireless Networks*, vol. 7, no. 3, pp. 269-261, 2001.

[4] F. Alagöz, B. R. Vojcic, D. Walters, A. AlRustamani, R. L. Pickholtz, “Fixed versus adaptive admission control in direct broadcast satellite networks with return channel systems”, *IEEE J. Select. Areas Commun.*, vol. 22, no. 2, pp. 238-249, Feb. 2004.

[5] L. Chisci, R. Fantacci, T. Pecorella, “Predictive bandwidth control for GEO satellite networks”, *Proc. IEEE Internat. Conf. Commun. (ICC 2004)*, Paris, France, June 2004, pp. 3958-3962.

[6] S. Kota, M. Marchese, “Quality of Service for satellite IP networks: a survey”, *Internat. J. Satell. Commun. Network.*, vol. 21, no. 4-5, pp. 303-348, July-Oct. 2003.

[7] N. Celandroni, F. Davoli, E. Ferro, “Static and dynamic resource allocation in a multiservice satellite network with fading”, *Internat. J. Satell. Commun. Network.*, vol. 21, no. 4-5, pp. 469-487, July-Oct. 2003.

[8] N. Iuoras, T. Le-Ngoc, M. Ashour, T. Elshabrawy, “An IP-based satellite communication system architecture for interactive multimedia services”, *Internat. J. Satell. Commun. Network.*, vol. 21, no. 4-5, pp. 401-426, July-Oct. 2003.

[9] A. Durresi, P.K. Jagannathan, R. Jain, “Proportional bandwidth allocation in satellite networks”, *Proc. 2003 IEEE Aerospace Conf.*, Big Sky, MT, March 2003, pp. 1253-1264.

[10] A. Iera, A. Molinaro, S. Marano, “Traffic management techniques to face the effects of intrinsic delays in geostationary satellite networks”, *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 145-155, Jan. 2002.

[11] A. Fu, E. Modiano, J. N. Tsitsiklis, “Optimal energy allocations and transmission control for communications

satelittes," *Proc. IEEE INFOCOM* 2002, vol. 2, New York, NY, June 2002, pp. 648-56,

[12] A. Iera, A. Molinaro, S. Marano, "Call admission control and resource management issues for real-time VBR traffic in ATM-satellite networks," *IEEE J. on Select. Areas in Commun.*, vol. 18, pp. 2393-2403, Nov. 2000.

[13] M. H. Ahmed, "Call Admission Control in wireless networks: a comprehensive survey", *IEEE Commun. Surveys*, vol. 7, no. 1, pp. 50-69, 1st Quart. 2005.

[14] C.W. Leong, W. Zhuang, Y. Cheng, L. Wang, "Call admission control for integrated on/off voice and best-effort data services in mobile cellular communications", *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 778-790, May 2004.

[15] H. J. Chao, X. Guo, *Quality of Service Control in High-Speed Networks*, John Wiley & Sons, New York, NY, 2002.

[16] H. Yaïche, R. R. Mazumdar, C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks", *IEEE/ACM Trans. Networking*, vol. 8, no. 5, pp. 667-678, Oct. 2000.

[17] R. Bolla, F. Davoli, "Control of multirate synchronous streams in hybrid TDM access networks", *IEEE/ACM Trans. Networking*, vol. 5, no. 2, pp. 291-304, April 1997.

[18] *ETSI EN 300 421 V1.1.2*, Digital Video Broadcasting, frame structure, channel coding and modulation for 11-12 GHz satellite services, 1997.

[19] K. W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*, Springer-Verlag, London, 1995.

[20] N. Celandroni, F. Davoli, E. Ferro, A. Gotta, "Adaptive cross-layer bandwidth allocation in a rain-faded satellite environment", *Internat. J. Commun. Syst.* (to appear).

[21] B. Tsybakov, N.D. Georganas, "Self-similar traffic and upper bounds to buffer-overflow probability in an ATM queue", *Performance Evaluation*, vol. 32, pp. 57-80, 1998.

[22] A. Gotta, "Resource partitioning and Call Admission Control in a rain-faded satellite channel with real-time connections", to be presented at *Internat. Workshop on Satellite and Space Commun. 2005 (IWSSC 2005)*, Siena, Italy, Sept. 2005.

[23] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont, MA, 1995.

[24] J. Padhye, V. Firoiu, D. F. Towsley, J. F. Kurose, "Modeling TCP Reno performance: a simple model and its empirical validation", *IEEE/ACM Trans. Networking*, vol. 8, pp. 133-145, 2000.

[25] N. Celandroni, F. Davoli, E. Ferro, A. Gotta, "Adaptive optimal bandwidth partitioning among TCP elephant connections over multiple rain-faded satellite channels", *Lecture Notes in Computer Science*, 3375, Springer-Verlag, Berlin, pp. 559-573, 2005.

[26] N. Celandroni, F. Potortì, "Maximising single connection TCP goodput by trading bandwidth for BER", *Internat. J. Commun. Syst.*, vol. 16, no. 1, pp. 63-79, Feb. 2003.

[27] I. Bisio, M. Marchese, "Analytical expression and performance evaluation of TCP packet loss probability over geostationary satellite channels", *IEEE Commun. Lett.*, vol. 8, no. 4, April 2004, pp. 232-234.

[28] I. Bisio, M. Marchese, "A bandwidth allocation scheme in satellite channels with Internet traffic sources: strategy and performance evaluation", *Proc. 2004 Internat. Symp. on Performance Evaluation of Computer and Telecommun. Syst. (SPECTS'2004)*, San Jose, CA, USA, July 2004, pp. 321-328.

[29] F. Davoli, M. Marchese, M. Mongelli, "Optimal resource allocation in satellite networks: certainty equivalent approach versus sensitivity estimation algorithms", *Internat. J. Commun. Syst.*, vol. 8, no. 1, pp. 3-36, Feb. 2005.

[30] F. Davoli, M. Marchese, M. Mongelli, "Discrete stochastic programming by infinitesimal perturbation analysis: the case of resource allocation in satellite networks with fading", *IEEE Trans. Wireless Commun.* (to appear).

[31] M. Baglietto, F. Davoli, M. Marchese, M. Mongelli, "Neural approximation of open loop–feedback rate control in satellite networks", *IEEE Trans. Neural Networks* (to appear).

[32] http://www.eutelsat.com/satellites/13ehb6.html

[33] A. Dell'Aera, L. A. Grieco, S. Mascolo, "Linux 2.4 implementation of Westwood+ TCP with rate-halving: a performance evaluation over the Internet", *Proc. IEEE Internat. Conf. Commun. 2004 (ICC 2004)*, Paris. France, June 2004, pp. 2092-2096.