

# Neural Approximation of Open-Loop Feedback Rate Control in Satellite Networks

Marco Baglietto, *Member, IEEE*, Franco Davoli, *Senior Member, IEEE*, Mario Marchese, *Senior Member, IEEE*, and Maurizio Mongelli, *Student Member, IEEE*

**Abstract**—A resource allocation problem for a satellite network is considered, where variations of fading conditions are added to those of traffic load. Since the capacity of the system is finite and divided in finite discrete portions, the resource allocation problem reveals to be a discrete stochastic programming one, which is typically *NP-Hard*. In practice, a good approximation of the optimal solution could be obtained through the adoption of a closed-form expression of the performance measure in steady-state conditions. Once we have summarized the drawbacks of such optimization strategy, we address two novel optimization approaches. The first one derives from Gokbayrak and Cassandras and is based on the minimization over the discrete constraint set using an estimate of the gradient, obtained through a “relaxed continuous extension” of the performance measure. The computation of the gradient estimation is based on infinitesimal perturbation analysis (IPA). Neither closed forms of the performance measures, nor additional feedbacks concerning the state of the system and very mild assumptions about the stochastic environment are requested. The second one is the main contribution of the present work, and is based on an open-loop feedback control (OLFC) strategy, aimed at providing optimal reallocation strategies as functions of the state of the network. The optimization approach leads us to a functional optimization problem, and we investigate the adoption of a neural network-based technique, in order to approximate its solution. As is shown in the simulation results, we obtain near-optimal reallocation strategies with a small real time computational effort and avoid the suboptimal transient periods introduced by the IPA gradient descent algorithm.

**Index Terms**—Functional optimization, neural networks, resource allocation, satellite networks, sensitivity estimation.

## I. INTRODUCTION

### A. Optimization Problems in Telecommunication Networks

**I**N computer networks extending over large geographical areas and in multiservice packet switching communication networks, in the presence of limited resources (buffers, bandwidth, or processing capacity), several forms of control are exerted to maintain a desired level of performance for all users and traffic types. The technology evolution and the users’ needs have brought to study and implement algorithms and mechanisms able to provide quality of service (QoS) in the Internet [3]. Heuristics often help finding efficient and simple solutions, but the strong market competition often requires more optimized performance.

Several optimization problems in telecommunication networks have a discrete stochastic programming nature. We

disregard, for the time being, the system’s dynamics and consider a general resource allocation environment in which some stochastic variables influence the system’s evolution. Decision variables (denoted by  $\theta$ ) are encountered in the form of nonnegative integers, which must be modified along time as parameters of the control policies, in order to optimize the system performance. Mathematically speaking, the aim of any network manager is to look for the best resource allocation  $\theta^*$  so that

$$\theta^* = \arg \min_{\theta \in \Theta} E_{\omega} \{L[\theta, \omega]\} \quad (1)$$

where the system performance can be expressed according to specific metrics  $L[\theta, \omega]$ , such as *blocking probability* of connection requests, *packet loss probability*, *packets’ mean delay* or *delay jitter*, *Service Provider’s revenue* [4]–[14]. The expectation in (1) is over all the feasible sample paths (i.e., the realizations of the stochastic processes)  $\omega$  of the system. Such optimization problems fall in the well-known area of the so-called stochastic discrete resource allocation problems (SDRAPs). This class of problems is *NP-hard* [15], even if no stochastic variables affect the system’s dynamics. In the presence of a stochastic environment, when no closed-form expression of  $L[\theta, \omega]$  is possible, the topical problem reveals to be the estimation of  $E_{\omega} \{L[\theta, \omega]\}$ . This generally leads to brute force offline simulations or direct measurements made on the system.

In the context of telecommunication networks, such problems are sometimes solved by means of centralized approaches, in which the control systems are based on closed-form expressions for the performance measure (see, for example, [8]–[11], [14], [16] for what concerns *Call Admission* and *bandwidth Control, routing and pricing* issues). For instance, in [5], [10], and [16], the Tsybakov-Georganas formula for the packet loss probability in the presence of self-similar traffic [17] is used. In [11] and [14], the blocking probability of call requests is computed through the well-known *Erlang B* formula.

The main drawback of these approaches is due to the fact that conditions for the applicability of closed-form functional costs are difficult to implement in real-life contexts. Actually, such approaches act according to a parameter adaptive *certainty equivalent* control [18]. A mapping between the current statistical behavior of the system and the parameters of the functional costs must be periodically performed online, in order to maintain good performance of the resource allocation algorithms. In the presence of self-similar traffic, closed forms for important performance measures (e.g., mean delay and delay jitter of the packets) are not always available; but “[...] *even under Markovian assumptions for processes of queuing systems, there are*

Manuscript received November 24, 2003; revised April 13, 2005.

The authors are with the Department of Communications, Computer, and Systems Science (DIST), University of Genoa, Genoa 16145, Italy (e-mail: mbaglietto@dist.unige.it)

Digital Object Identifier 10.1109/TNN.2005.853424

only limited cases where closed-form expressions can be obtained' [19], and, in general, it is very difficult to assure that in a real application scenario some strict hypotheses are verified. Moreover, most of these techniques rely on the application of dynamic programming algorithms, whose online implementation in a real context may be quite impractical, due to the well-known *curse of dimensionality*.

### B. Need of Sensitivity Estimation Algorithms

The application of algorithms, able to estimate the sensitivity of the performance measure, could help in providing "good" control decisions without the adoption of closed-form functional costs and the application of computationally expensive algorithms.

Sensitivity estimation algorithms are based on perturbation analysis (PA). PA relies on the observation of the sample paths followed by the stochastic processes of discrete event systems (DESes) and gives an estimation of the derivative of the performance index [19]–[22], thus, allowing the application of a gradient-based algorithm to optimize the resource allocation [6], [22]–[24]. The related optimization techniques are known in the literature as *online surrogate optimization methodologies*, because they act online, with a gradient-based algorithm, by applying a "surrogate" relaxation of the discrete functional cost [1], [2].

For the purpose of the present work, we recall the results of [24] to highlight how an online surrogate methodology is able to improve the performance over a control strategy based on a closed-form expression of the performance index. As a main drawback, online surrogate methodologies do not have any guarantee concerning the time needed to reach the optimal resource allocation. They lead to control decisions characterized by suboptimal transient periods, in which the DES reaches sub-optimal performance. If the stochastic processes are stationary, at least in wide sense (i.e., their moments are fixed during the temporal evolution of the system), the transient periods can be properly reduced, by means of a proper tuning of the gradient stepsize. Nevertheless, since, in practice, it is difficult to assure a stationary behavior of the stochastic processes, dynamic control reactions, able to quickly converge to the optimal reallocation, must be investigated.

### C. Functional Optimization Approach

The aforementioned criticism about pure gradient-based algorithms leads us to the principles of the theory of optimal control, which give instruments for the choice of control variables as functions of the state of the system. In doing so, however, the dynamic evolution of the system over time must be described explicitly, and we are no longer facing a parametric optimization problem of resource allocation, but rather a dynamic control problem, where the decision variables are functions of information acquired in successive instants. Nevertheless, the theory of optimal control can reveal more than one obstacle in facing such a dynamic stochastic discrete resource allocation problem (D-SDRAP).

Almost all of the optimal control techniques in the literature make use of closed-form expressions of the system equations of the controlled dynamic system and of the cost functional. In [25]

and [26], an optimization technique is presented to solve continuous functional optimization problems to obtain approximate solutions at any degree of accuracy. Such a method, called the *Extended Ritz* method, consists in assigning a fixed structure to the control functions, where a fixed number of parameters have to be optimized. By exploiting the approximating properties of neural networks [27], it is possible to face the curse of dimensionality in the approximating structure of the control strategies with respect to the dimension of the state-space. The LQG assumptions (i.e., the controlled dynamic system and the observation channels on the system state being linear, the cost function to be minimized being quadratic, and the stochastic variables being Gaussian) are not necessary. In [28], this methodology was applied to decentralized routing in packet switched networks. The technique is able to compute off line near-optimal control strategies, relaxing the necessity of heavy online computational efforts. However, such optimization technique cannot manage control variables that lie on a discrete constraint set and needs analytical expressions for the dynamic equations of the controlled system and for the functional cost.

In this paper, we take a somewhat simpler approach, which is based on an open-loop feedback strategy [18]. More specifically, we consider a D-SDRAP consisting in the minimization, at each time instant  $t$ , of a cost of the type

$$E L[\theta(t), \omega] = E \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} \gamma[\theta(t), \omega(\tau)] d\tau \quad (2)$$

where we have explicitly indicated the dependence on time of the decision parameter and of the stochastic environment, and  $\gamma[\cdot, \cdot]$  represents the "instantaneous" performance measure. We state the problem in the satellite environment and face it by means of a functional optimization approach, where the system dynamic equations are not required explicitly. Hence, we modify the optimization methodology proposed in [25] and [26], in order to solve such functional optimization problem.

In this perspective, the major concern is to adapt the related training algorithm to a discrete event simulation-based environment, in which only estimates of the functional cost and of its gradient can be obtained. This can be done by exploiting the principle of PA and applying the infinitesimal perturbation analysis (IPA) proposed in [21] and [22] to the solution of our problem. Finally, we shall show how our approach guarantees better performance with respect to the corresponding online surrogate methodology.

Considering the resource allocation in the context of broadband satellite networks makes the optimization even more difficult with respect to a terrestrial environment, because channel degradation effects must be taken into account, together with the traffic changes. As will be clear from the following, however, the proposed model can be generalized for other application scenarios and functional costs.

The remainder of paper is organized as follows. In the next section, we formulate the D-SDRAP arising in a satellite environment; then, in Section III, we briefly describe the certainty equivalent approach adopted to solve it in [5] and [10], and in Section IV we show a possible alternative solution with an online surrogate optimization methodology. In Section V, we formulate the solution of the D-SDRAP with our open-loop

feedback control (OLFC) strategy and approximate it through a proper neural network-based technique. In Section VI, we compare such optimization strategies according to different simulation scenarios, showing the capability of the OLFC to achieve the best performance. Finally, conclusions and future work are outlined in Section VII.

## II. DYNAMIC STOCHASTIC DISCRETE RESOURCE ALLOCATION PROBLEM IN A SATELLITE SYSTEM

### A. State of the Art

Satellite systems (Low Earth Orbit, Medium Earth Orbit, and Geostationary satellites) have been proposed to support worldwide multimedia and interactive services. Many research issues are currently under investigation to improve the performance of multimedia satellite systems: integrated satellite architectures, beam scheduling, on board signal generation, adaptive modulation and coding, multiple access, flow control, and resource allocation [29].

In such systems a major concern is related to variable fading conditions over the channel that can heavily affect the transmission quality, especially when working in *Ka band*, where the effect of rain over the quality of transmission is more significant [10], [29].

In the literature, it is possible to find optimal policies developed in the case of a finite quantity of transmission energy for satellite network devices. [30], [31] show a dynamic programming formulation of the problem that leads, for special cases, to a closed-form optimal policy, in order to find a tradeoff between the minimization of the energy required to send a fixed amount of data and the maximization of the throughput over a fading channel. Power allocation for fading multiuser broadcast channels is a popular topic also in information theory [32]. Error recovery techniques, such as automatic retransmission request (ARQ) and forward error correction (FEC), are employed in wireless environments to face adverse channel conditions. Usually ARQ is not adopted in the presence of real-time traffic with stringent latency constraints. FEC mechanisms allow recovering erroneous packets, despite channel degradation, but, owing to their overhead, they may cause further congestion, thus, increasing the packet loss probability. A tradeoff must be found out between the resource allocation and the FEC redundancy.

In the aforementioned works, the problem is analyzed and solved at the physical layer: power allocation is performed, in order to obtain good reactions to variable fading conditions. In [5], [10], and [24] and in this work, a FEC mechanism is located at the physical layer and adaptive bandwidth allocation strategies are provided at the data link (or upper) layers to minimize the loss probability of the overall system.

Resource management schemes are often considered for satellite systems with respect to the call admission control (CAC) issue [29]. The satellite system is managed at the call level, adopting CAC techniques with dynamic reactions performed by the resource management agent, to face variable system condition. In this work, we investigate bandwidth allocation strategies disregarding the CAC problem and account for the effect of variable traffic and fading level conditions, by considering the network at the packet level. In this perspective, two

novel optimization algorithms are investigated to counteract variable fading and traffic load. Stochastic processes are assumed to be nonstationary. The optimization algorithms we are going to investigate have to dynamically adjust the bandwidth allocation, by following adaptively the current behavior of the stochastic processes and optimally distributing the available channel capacity among the satellite stations.

### B. Model of the Satellite Network

The satellite environment under investigation consists of a fully meshed satellite network that uses bent-pipe geostationary satellite channels, joining  $N$  stations. This means that the satellite only performs the function of a repeater, without onboard processing of data. The system operates in multifrequency-time division multiple access (MF-TDMA) mode, which allows us to divide the system capacity  $K$  into a number of channels, so that the traffic stations can be downsized with respect to a pure TDMA system. TDMA allows the transmission of digital data streams from many sources sequentially assigned to different time slots. Each earth station has to know when to transmit and it must be able to recover the carrier and the clock for each received burst in time to sort out all wanted channels; this can be accomplished through preambles at the beginning of each burst. The integer  $K$  represents the number of available bandwidth units, where a unit consists of the smallest quantity of bandwidth assignable to a station. TDMA is easy to reconfigure for changing traffic demands and is robust with respect to noise and interference. In particular, for satellite systems, it maximizes the downlink *Carrier/Noise Power* ratio.

We further suppose that the stations generate IP packets, which, in turn, are divided at the data link layer for transmission on the satellite channel into smaller, fixed size data units. The latter may typically be asynchronous transfer mode (ATM) or digital video broadcasting (DVB) cells.

A master station is supposed to maintain the system synchronization and is responsible for the capacity allocation to the traffic stations. The master station performance is the same as the slave stations' one; thus, the role of master can be assumed by any station in the system.

This assures that the master operates in clear sky conditions for almost all of the time, because when the current master's attenuation exceeds a given threshold, its role is assumed by another station that is in good conditions. In other settings, the satellite itself could be responsible of the allocation: in this case, it should be equipped with an onboard processing unit, in order to receive information periodically from the stations and calculate the next bandwidth allocation.

With a notation that slightly differs from [21], each station  $i = 1, \dots, N$  is assumed to have a finite-capacity buffer of fixed size  $Q_i$  and a single server. Fig. 1 illustrates the main components of a fluid model of the earth station.

The buffer receives variable bit rate (VBR) traffic from different sources. The stochastic processes associated with our optimization problem are the following:

- $\alpha_i(t)$ : the input flow (*inflow*) rate process at the  $i$ th station;
- $x_i(t)$ : the buffer *workload* process, namely, the fluid volume in the  $i$ th buffer;
- $\theta_i(t)$ : the *service capacity* of the buffer at station  $i$ ;

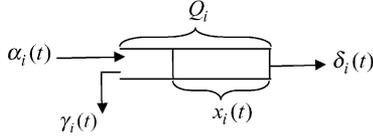


Fig. 1. Model of the  $i$ th satellite station.

- $\delta_i(t)$ : the output flow rate;  $\delta_i(t) = \theta_i(t)$  if the buffer is not empty, otherwise:  $\delta_i(t) = \alpha_i(t)$  if  $\alpha_i(t) < \theta_i(t)$ ,  $\delta_i(t) = \theta_i(t)$  if  $\alpha_i(t) \geq \theta_i(t)$ ;
- $\gamma_i(t)$ : the loss rate (*overflow*) process due to a full buffer.

### C. Traffic Model

Even if the optimization techniques we are going to formulate are not related to a specific behavior of the traffic sources, we shall adopt a specific traffic model in order to specify a closed-form functional cost of the performance index for the purpose of performance comparison. Hence, we now introduce the traffic model adopted for each station's inflow process  $\alpha_i(t)$ ,  $i = 1, \dots, N$ .

In the recent years, packet-based traffic has been demonstrated to show statistical characteristics very close to "self-similar" processes [33]. We suppose that each inflow process  $\alpha_i(t)$ ,  $i = 1, \dots, N$ , is composed by a self-similar stochastic process, due to the aggregation of some VBR sources. For each station  $i$ , the statistical parameters that describe such process are: the peak bit rate  $B_p^i$  of the on-off sources and the burst arrival rate  $\lambda_{\text{burst}}^i = (M_i)/(\bar{\tau}_i + \bar{\zeta}_i)$  of the aggregated flow, namely, the average number of bursts "seen" by station  $i$ . We denote by  $\bar{\tau}_i$  and  $\bar{\zeta}_i$  the mean time durations of the burst and the silence periods, respectively. They are both Pareto distributed. It is shown in [17] that the aggregation of  $M$  independent sources with such probability distribution over  $\tau$  determines an aggregated flow with self-similar properties. This has a dramatic impact over the resources that must be reserved to such flow in order to guarantee packet-level QoS constraints [13], [17].  $M_i$  is the maximum number of on-off sources in the flow  $\alpha_i(t)$ .

### D. Performance Measure

We take the minimization of the loss volume as the final goal to pursue. However, the generality of the proposed technique would allow us to address also other network metrics. The loss volume for each station  $i = 1, \dots, N$  is given by

$$L_i[\theta_i^d(\cdot), \alpha_i] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \gamma_i[\theta_i^d(\tau), \alpha_i(\tau)] d\tau \quad (3)$$

where  $\theta^d(t) = [\theta_1^d(t), \dots, \theta_N^d(t)]$  is the vector of the service capacities allocated to each station at time  $t$ . (From here on, we shall use the superscript  $d$  in order to stress that a quantity belongs to a "discrete" set).

It is worth noting that the cost in the form (3) is fairly general and far more complex than that expressed inside the expectation in (2). Indeed, its minimization with respect to  $\theta^d(t)$  would

imply finding an infinite sequence of decision vectors. Its solution requires suitable approximations that will lead us to the form of (2).

As to  $\theta^d(t)$ , we have  $\theta^d(t) \in \Theta_d$

$$\Theta_d = \left\{ \theta^d(t) \in \mathbb{N}^N : \theta_i^d(t) = h_i(t) \cdot \text{MAU}, h_i(t) \in \mathbb{N} \right. \\ \left. \sum_{i=1}^N \theta_i^d(t) = K \right\} \quad (4)$$

namely, the allocated service rate for each station is a discrete number of minimum allocation units (MAUs), i.e., the smallest portion of bandwidth that can be assigned to a station. As mentioned previously,  $K$  is the maximum number of MAUs available for the system.

The total loss volume of the overall system becomes the sum of the contributions of each station

$$L[\theta^d(t), \alpha] = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^N \int_0^T \gamma_i(\theta_i^d(\tau), \alpha_i(\tau)) d\tau \quad (5)$$

where  $\alpha(t) = \text{col}\{\alpha_1(t), \dots, \alpha_N(t)\}$  is the aggregate vector of the inflow processes.

### E. Fading Effect and Problem Formulation

The effect of fading is modeled as a reduction in the bandwidth actually "seen" by a traffic station. The fading effect is represented by a variable  $\phi_i(t)$ , which shows how the bandwidth  $\theta_i^d(t)$  is reduced [5], [10]. For each station  $i$ , at time  $t$ , the "real"  $\theta_i(t)$  is

$$\theta_i(t) = \phi_i(t) \cdot \theta_i^d(t); \quad \phi_i(t) \in [0, 1]; \quad i = 1, \dots, N. \quad (6)$$

In our model, the fading effect involves a reduction of the bandwidth actually "seen" by the station, or, equivalently, an increase in the bandwidth required by the traffic sources to maintain the same bit error rate (BER). We suppose the presence of fade countermeasures located at the physical layer, totally managed by the earth station. They are expected to provide the desired BER through FEC codes [4], [5], [10], [29]. Whenever the fading effect causes errors over the packets, an adaptive control can monitor the carrier/noise power (C/N) ratio and, on the basis of this measure, increase the redundancy of the packets sent introduced by the FEC. In this way, for each station, the available bandwidth is reduced: since more bits are necessary to transmit a single packet (because of FEC coding), the outflow cell rate can be considered as modified by the fading effect. Clearly, heavier fading conditions will involve a more consistent decrease of the allocated bandwidth, because more coding protection of data will be necessary, and vice versa.

So, we are going to discuss the combination of a BER-related fade countermeasure technique with a resource allocation problem, where a master control station is responsible of the bandwidth reallocations.

The optimization problem can now be stated. It consists of finding the optimal sequence of bandwidth allocations

$\theta^{d*}(t), t \geq 0$ , so that the overall loss volume of the system is minimized.

*Problem* dynamic-stochastic resource allocation problem (D-SDRAP): Find  $\theta^{d*}(t) \in \Theta^d, t \geq 0$  such that the cost function

$$\omega_1, \dots, \omega_N \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^N \int_0^T \gamma_i [\theta_i^d(\tau) \cdot \phi_i(\tau), \alpha_i(\tau)] d\tau \right\} \quad (7)$$

is minimized. We denote by  $\omega_i$  is the generic sample path for station  $i$ , namely, a realization of the stochastic processes characterizing the temporal evolution of station  $i: \phi_i(t), \alpha_i(t), i = 1, \dots, N, t \geq 0$ . In the following, we denote by  $\varphi(t) = \text{col}\{\phi_1(t), \dots, \phi_N(t)\}$  the aggregate vector of the fading processes and by  $\varphi$  a single realization of  $\varphi(t)$  over all  $t \in [0, +\infty]$ . Similar definitions hold for  $\alpha(t)$  and  $\alpha$ .

As mentioned in the introduction, the expectation  $E_{\omega_1, \dots, \omega_N}[\cdot]$  is over all the feasible sample paths  $\omega_i \in \Omega_i$  for each station  $i$ .

### III. CERTAINTY EQUIVALENT OPTIMIZATION APPROACH

Also in the presence of a self-similar behavior of the traffic sources, it is possible to obtain analytical models for the computation of the loss probability performance [17]. Such closed-form expressions could be used in the aforementioned resource allocation framework in order to optimize the resource allocation [5], [10]. Anyway, they need to assume a perfect knowledge of the system's state and a strong consumption of computing power, due to the continuous online minimization of a global cost through the adoption of a proper dynamic programming algorithm.

#### A. Closed-Form Formula for the Loss Probability

We now describe in some detail such optimization strategy. Following the model employed in [5], [10], and [16], we adopt the Tsybakov-Georganas formula, reported in (8), shown at the bottom of the page, for the packet loss probability  $P\text{Loss}_i$  of each station  $i$ , given a static allocation  $\theta_i^d$  that is supposed to last indefinitely.

The parameter  $\bar{\tau}_i$  is the average burst length and  $a_i$  and  $\chi_i$  are the shape and the normalization parameters, respectively, of a discrete Pareto distribution over the burst length. The other notation is explained in the following.

Let  $\hat{T}$  be a reference time interval (slot), to which we shall refer all the relevant parameters of the cell queue of each station  $i$ . The slot also represents the minimum duration of a burst, and the burst length  $\tau$  is expressed as an integer number of slots. Let  $B_p$  be the peak generation rate [bits/s] of each source, supposed for simplicity, and without loss of generality, to be equal for each transmitting station, and let  $L$  be the number of bits in a cell. Then,  $R = \lceil ((\hat{T} \cdot B_p)/L) \rceil$  is the number of cells generated by

an active burst in a slot ( $\lceil o \rceil$  being the smallest integer greater than or equal to  $o$ ). Suppose that the number of new sources becoming active in each slot is i.i.d. Poissonian (which is true for the model of Section II, asymptotically in the number of sources [17]), with parameter  $\lambda_i = \lambda_{\text{burst}}^i \cdot \hat{T}$ . If  $H$  is the cell's header length in bits, then  $X_i = \lfloor (\theta_i^d \cdot \phi_i) / (L + H) \cdot \hat{T} \rfloor$  represents the bandwidth  $\theta_i^d$ , assigned to station  $i$  and degraded according to the current value of fading coefficient  $\phi_i$ , expressed in cells per slot ( $\lfloor o \rfloor$  being the largest integer less than or equal to  $o$ ).

Equation (8) computes the steady-state loss probability of the  $i$ th station, resulting from a stationary configuration of the inflow rate process, i.e., when the input bit rate of buffer  $i$  follows a stationary random process producing, on average, a fixed value of the  $\lambda_{\text{burst}}^i$  parameter. In this perspective, the closed form resembles the static functional cost (1) and allows implementing a parameter adaptive certainty equivalent (PACE) control. It consists of mapping the current behavior of the sources to the parameters of (8) and periodically performing the service rates reallocations. Actually, it is necessary to counteract both time varying  $\lambda_{\text{burst}}^i, i = 1, \dots, N$ , and fading values. The stochastic environment under investigation is therefore not stationary. Real time adjustments of (8) are needed. As a result, the mentioned mapping is mandatory to update the PACE control with respect to the current state of the network. That means the state is perfectly known in real time or, in practice, a traffic estimator should be employed to capture the behavior of the sources.

The PACE control acts as follows. The optimization problem formulated in (7) is approximated by a sequence of finite-horizon OLFC problems, where stationarity in the loss probability distribution is supposed to be reached in each interval, in order to apply (8). The minimization of the overall loss probability is therefore stated at the time instants  $k = 1, 2, \dots$ , where a new bandwidth reallocation is performed

$$\theta^{d*}(k) = \arg \min_{\theta^d(k) \in \Theta_d} J[\theta^d(k)] \quad (9)$$

$$J[\theta^d(k)] = \sum_{i=1}^N P\text{Loss}_i(\theta_i^d), \quad k = 1, 2, \dots \quad (10)$$

The index  $k$  denotes the reallocation time instants at which a new solution of (9) is computed, according to the current state of the network (in terms of traffic load and fading levels). We must note that the self-similar assumption holds asymptotically in the number of sources [17] and that (8) holds asymptotically in  $Q_i$ . In practice, the number of sources is finite and, as a consequence, the bandwidth needs are not exactly predicted by the PACE approach [24].

#### B. Online Computational Effort

An efficient computational algorithm to solve problem (9) can be based on *Dynamic Programming* over the stations [34]. This is polynomial in the number of stations and in the total

$$P\text{Loss}_i(\theta_i^d) = \begin{cases} \min \left\{ \frac{\chi_i \cdot \lambda_i \cdot R^{a_i} \cdot Q_i^{-a_i+1}}{a_i \cdot (a_i-1) \cdot (X_i - \lambda_i \cdot R^{a_i} \cdot \bar{\tau}_i)}, 1 \right\} & \text{if } X_i > \lambda_i \cdot R^{a_i} \cdot \bar{\tau}_i \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

number of MAUs and has complexity  $O(K^2 \cdot N)$ ,  $K$  being the maximum number of available MAUs. Hence, its computational burden seems to be not so heavy as expected. Unfortunately,  $K$  may be very large in the presence of a high capacity link and with the adoption of small MAUs (from 100 kb/s down to lower values). So, if also the number of active stations is high, such optimization procedure might need a huge computational time to terminate successfully. This drawback can severely degrade its performance. Since the adoption of high values for the MAUs is not recommendable (as it leads to poor performance of the optimization algorithm, too), if the dynamic programming algorithm has to be employed, a proper tradeoff must be found out in order to limit its computational burden.

In the following, we shall denote the application of this technique as closed-form functional cost with dynamic programming (CF&DP).

It is worth noting that the optimization algorithms proposed in Sections IV and V will reveal to be computationally lighter than the CF&DP technique. Nevertheless, we do not insist on their suitability due to the involved lower computational complexity, but we highlight how they are capable to outperform the CF&DP approach in terms of performance.

#### IV. ONLINE SURROGATE OPTIMIZATION ALGORITHM

In this section, we summarize the online surrogate optimization technique, taken from [1] and [2], and detail its application to our problem.

In order to generate a gradient descent of bandwidth reallocations, a derivative estimation of the performance index is needed. We employ a derivative estimation technique that assumes mild a-priori hypotheses concerning the stochastic processes involved in the system.

##### A. Performance Derivative Estimation

Since our purpose is to obtain a gradient estimate of the performance metric with respect to the service rate  $\theta \in \mathbb{R}$ , we apply IPA [21], [22]. It consists in finding queuing systems performance derivative estimators at the packet level. The topical problem is that such derivatives are unknown, even for basic queuing systems (see, e.g., [19] for an overview concerning the principles of PA). Only estimators can be provided. They are derived as functions of the sample paths of the system. “*The form of the IPA estimators is obtained by analyzing the system as a Stochastic Fluid Model, but the associated values are based on real data*” [22]. The derivative needed here is related to the loss volume with respect to the service rate. In brief, IPA gives analytical instruments to derive an estimator of the mentioned derivative as function of the lengths of the measured congestion periods of a traffic buffer. This requires real time measurements over the system. Details about its derivation are briefly summarized in the following.

As proved in [21] and [22], the durations of the so-called “busy periods” of the buffer are the key to derive the IPA estimator. Let  $B_k$  be a busy period of the buffer between two observation time instants (e.g., the bandwidth reallocation time instants), namely, a period of time in which the buffer is nonempty. In particular, two time instants within  $B_k$  reveal to be topical. Let  $\xi_k$  be the starting instant of  $B_k$ . Let  $\nu_k$  be the instant of time

when the last loss occurs during  $B_k$ . Then, for every  $\theta_i, i = 1, \dots, N$ , it can be shown that

$$\frac{\partial L_i^k(\theta_i)}{\partial \theta_i} = -(\nu_k(\theta_i) - \xi_k(\theta_i)) \quad (11)$$

where  $L_i^k(\theta_i)$  is the value of the loss volume related to the given busy period  $B_k$  at station  $i$ . This means the contribution to the derivative of each active period  $B_k$ , during which some losses occurred, is the length of the time interval from the start of  $B_k$  until the last time point in  $B_k$  at which the buffer is full. At the time of a bandwidth allocation, denoting by  $N_i^B$  the number of active periods between two consecutive bandwidth allocations at station  $i$ , we get an estimation of the performance index derivative as

$$\frac{\partial L_i(\theta_i)}{\partial \theta_i} = \sum_{k=1}^{N_i^B} \frac{\partial L_i^k(\theta_i)}{\partial \theta_i}; \quad i = 1, \dots, N. \quad (12)$$

IPA estimators are “nonparametric,” since they are computable directly from an observed sample path  $\omega$ , without any knowledge concerning the probability distributions of the involved stochastic processes [19], [21], [22]. The mathematical conditions requested for the validity of (11) are quite general:  $L_i(\theta)$  must be *Lipschitz continuous* and the function  $\alpha_i(t) - \theta_i(t)$  must be *piecewise continuously differentiable* [21].

##### B. Allocation Algorithm Based on Sensitivity Estimation

The previously described derivative estimation, available for each station  $i$  at time  $t$ , lets us establish a bandwidth reallocation procedure, which consists of an exchange of information phase followed by the resource allocations, performed by the master.

First of all, it is necessary to “relax” the discrete constraint set  $\Theta_d$  into a continuous one  $\Theta_c$

$$\theta^c \in \Theta_c, \quad \Theta_c = \left\{ \theta_i^c \in \mathbb{R}^+; \sum_{i=1}^N \theta_i^c = K \right\}. \quad (13)$$

As proposed in [1] and [2], the discrete functional cost defined over  $\Theta_d$  is transformed into a “surrogate” one over  $\Theta_c$ . Then, the sensitivity estimation procedure is decentralized. Let  $\hat{t}$  be the time interval necessary for each station  $i$  to reach all other stations with the “reallocation messages” (denoted by  $\hat{\theta}_i^c$ ). Initially, the bandwidth resources are equally distributed among the stations and, during the system evolution, each station  $i$ , for every  $t = k\hat{t}, k = 1, 2, \dots$ , has to:

- 1) **observe** the buffer temporal evolution during the time interval  $[(k-1)\hat{t}, k\hat{t}]$  according to the current sample path  $\omega_i$  and bandwidth allocation  $\theta_i^d[(k-1)\hat{t}], \theta^d[(k-1)\hat{t}] \in \Theta_d$ ;
- 2) **compute** the gradient estimation  $(\partial L_i(\theta_i[(k-1)\hat{t}])) / (\partial \theta_i)$  according to (11) and (12);
- 3) **adjust** the value of its “bandwidth allocation need” using the gradient method

$$\hat{\theta}_i^c[k\hat{t}] = \theta_i^c[(k-1)\hat{t}] - \eta \cdot \phi_i \cdot \frac{\partial L_i[\theta_i(k-1)\hat{t}]}{\partial \theta_i}; \quad (14)$$

- 4) **communicate** such  $\hat{\theta}_i^c[k\hat{t}]$  to each master station;

- 5) (for each station that has the role of master station) by looking at the information received by the other stations (i.e.,  $\widehat{\theta}_j^c[(k-1)\hat{t}]$ ,  $j = 1, \dots, N; j \neq i$ ), and on the basis of the local bandwidth need  $\widehat{\theta}_i^c[(k-1)\hat{t}]$ , **obtain**  $\theta^c[(k-1)\hat{t}]$  through  $\theta^c[(k-1)\hat{t}] = \Pi[\widehat{\theta}^c[(k-1)\hat{t}]]$  and **convert**  $\theta^c[(k-1)\hat{t}]$  to the nearest discrete feasible neighbor  $\theta^d[k\hat{t}]$  so that  $\theta^d[k\hat{t}] \in \Theta_d$ ; such conversion defines the bandwidth allocation for the satellite system in the time interval  $[k\hat{t}, (k+1)\hat{t}]$ .

The projection operator  $\Pi[\widehat{\theta}]$  is defined as  $\Pi[\widehat{\theta}] = \arg \min_{\theta \in \Theta_c} \|\widehat{\theta} - \theta\|$ .

The previous procedure stems from the usage of  $(\partial L_i[\theta_i(k-1)\hat{t}]) / (\partial \theta_i)$ , which is the measurable quantity derived from IPA, in (14). Actually, we are trying to approximate the gradient with respect to  $\theta_i^c$ , which would require the application of PA on the discrete parameter set. However, since we are considering a service rate, which can admit *infinitesimal perturbations*, whose effect is evaluated from the sample path through (11) and (12), we have chosen to apply IPA and to perform what would be a descent step on the loss function without discretization.

As is shown in [1], the nearest feasible neighbor  $\theta^d[k\hat{t}] \in \Theta_d$  of  $\theta^c[(k-1)\hat{t}] \in \Theta_c$  can be determined, at step 5, by using an algorithm based on the simplex method. However, it is possible to apply a simpler algorithm based on the  $N+1$  discrete neighbors of  $\theta^c[(k-1)\hat{t}]$ , not necessarily all feasible, and on the selection of one of them, which satisfies the discrete constraint set  $\Theta_d$  [2].

### C. Decentralized Sensitivity Estimation Versus Centralized Reallocation

The gradient-descent algorithm of step 3 allows a decentralization of the optimization procedure. We suppose that a personal processor is assigned to each station  $i$ ; in this way, the optimization procedure runs in parallel on each independent processor located in each station [20]. Such distributed computation is a very attractive property, as it enables each station to compute its “optimized bandwidth need” locally on the basis of the temporal evolution of the local stochastic processes.

### D. Computational Complexity

The computational effort required by the IPA sensitivity estimation procedure is very low. The algorithm adopted in step 5 for the “surrogate to discrete” conversion is  $O(N+1)$ . For this reason, and owing to the mild assumptions requested for the applicability of the IPA technique, the proposed optimization algorithm can be efficiently applied in real time.

### E. System Signaling

The reallocation period  $\hat{t}$  depends on the satellite round trip time (RTT). The master control station has to get the sensitivity estimation of each earth station to compute the next reallocation, and, particularly with geostationary satellites, which are located at a distance of about 36 000 km from the earth with a RTT close to 500–600 ms, it is necessary to consider the relevance of the propagation delay for this information. If, at time  $t = k\hat{t}$ , station  $i$  generates its new bandwidth need  $\widehat{\theta}_i^c[k\hat{t}]$ , such value becomes available for all other stations only at the time  $t = (k+1) \cdot \hat{t}$ .

So, a reasonable value for the reallocation time period  $\hat{t}$  is 1.0 s [5], [10], [24].

In [6], a similar approach is applied for the optimization of the CAC in a circuit switched network, and, at the best of our knowledge, such technique has been adopted for the first time to optimize the performance of a telecommunication network at the packet level in [24].

In the following, we shall denote the application of this technique as sensitivity estimation and gradient descent (SE&GD) optimization approach.

## V. FUNCTIONAL OPTIMIZATION APPROACH

As we shall show in the simulation results, the SE&GD can be successfully applied online if the step size  $\eta$  in (14) is properly refined by means of an offline performance evaluation. If the optimal solution of the discrete stochastic optimization problem reaches a steady state, also good performance for the DES can be obtained, even if the gradient algorithm of step 3 needs suboptimal transient periods to reach the optimal steady states. Clearly, the investigation necessary to evaluate the best value for the gradient stepsize  $\eta$  could be difficult, since several simulations should be performed with different values of  $\eta$ . Moreover, if the behavior of the stochastic processes of the DES does not allow the decision variables to reach a steady state, it is quite difficult to verify if the proposed optimization algorithm offers good performance.

In order to avoid these drawbacks and since it is impossible to solve (7) analytically, we formulate a closed-loop control strategy based on a functional optimization approach. The idea is to provide resource reallocations as functions of a suitable “information vector”  $\mathbf{I}(t)$ , which summarizes the “recent history” of the DES. In particular, new resource reallocations are performed for  $t = k\hat{t}$ ,  $k = 0, 1, \dots$

Let

$$\mathbf{I}(k\hat{t}) = \text{col}\{\theta^d((k-\Xi)\hat{t}), \dots, \theta^d((k-1)\hat{t}), \varphi((k-\Xi)\hat{t}), \dots, \varphi((k-1)\hat{t}), \alpha((k-\Xi)\hat{t}), \dots, \alpha((k-1)\hat{t})\} \quad (15)$$

be an aggregate vector that maintains a finite horizon memory (of depth  $\Xi$ ) over the values assumed by the decision variables and the measurable stochastic variables involved in the DES during the time interval  $[(k-\Xi)\hat{t}, (k-1)\hat{t}]$ . We denote by  $\hat{t}$  the reallocation period as outlined in the previous section. Let  $J_{k\hat{t}}(\theta^d(k\hat{t})) = E_{\omega} L^{k\hat{t}}[\theta^d(k\hat{t}), \varphi, \alpha]$  be the functional cost after such bandwidth reallocation along an infinite time horizon, where

$$L^{k\hat{t}}[\theta^d(k\hat{t}), \varphi, \alpha] = \lim_{T \rightarrow \infty} \frac{1}{T} \times \sum_{i=1}^N \int_{k\hat{t}}^{k\hat{t}+T} \gamma_i(\theta_i^d(k\hat{t}) \cdot \phi_i(\tau), \alpha_i(\tau)) d\tau. \quad (16)$$

Let  $\mathbf{f}(\mathbf{I}(k\hat{t}))$  be a reallocation law, which provides a surrogate continuous bandwidth allocation  $\widehat{\theta}^c(k\hat{t}) \in \Theta_c$  as a function of the current information vector

$$\widehat{\theta}^c(k\hat{t}) = \mathbf{f}(\mathbf{I}(k\hat{t})). \quad (17)$$

Then, by converting  $\widehat{\theta}^c(k\hat{t})$  into its discrete feasible neighbor  $\theta^d(k\hat{t}) \in \Theta_d$  (by means of the algorithm mentioned in the pre-

vious section), Problem D-SDRAP now becomes a functional optimization problem.

*Problem* functional dynamic-stochastic discrete resource allocation problem (FD-SDRAP): Find the optimal bandwidth reallocation function  $\mathbf{f}^*(\cdot)$ , such that  $\mathbf{f}^*(\mathbf{I}(k\hat{t})) \in \Theta_d, k = 0, 1, \dots$ , and the cost

$$J^{k\hat{t}}(\mathbf{f}(\mathbf{I}(k\hat{t}))) = E_{\omega} L^{k\hat{t}}[\mathbf{f}(\mathbf{I}(k\hat{t})), \boldsymbol{\varphi}, \boldsymbol{\alpha}] \quad (18)$$

is minimized.

Such formulation yields a closed-loop resource allocation law able to perform online dynamic control reactions to variable traffic and fading levels conditions. However, at time  $k\hat{t}$ , new reallocations are performed “as if” they were to be applied without changes in the future, i.e., possible changes in the statistics of the random variables are disregarded. Then, the control strategy resulting from the solution of Problem FD-SDRAP for successive time instants  $k\hat{t}, k = 0, 1, \dots$  can be considered as an OLFC one [18]. The problem is now to formulate a technique aimed at solving such functional optimization problem.

It is worth noting that the infinite horizon functional cost (16) can be only computed through a simulation-based receding-horizon approximation. The neural approximating scheme we are going to investigate follows this direction, by exploiting a finite horizon estimate of (16) for each possible stationary configuration of the stochastic processes involved in the problem.

#### A. Modified Extended Ritz Method

In order to approximate the optimal OLFC resource allocation law  $\mathbf{f}^*(\cdot)$ , we propose a modification of the Extended Ritz method of [25].

The Extended Ritz method approximates the solution of a functional optimization problem by fixing the structure of the decision functions. Namely, decision functions are constrained to take on the structure of one hidden layer networks, i.e., linear combinations of basis functions containing free parameters to be optimized

$$\bar{\mathbf{f}}(\mathbf{I}, \mathbf{w}) = \text{col} \left\{ \sum_{l=1}^v c_l^i p(\mathbf{I}, \tilde{\mathbf{w}}_l) + c_0^i; i = 1, \dots, N \right\} \quad (19)$$

where  $\mathbf{w} = \text{col}\{\tilde{\mathbf{w}}_l, l = 1, \dots, v; c_l^i, l = 1, \dots, v, i = 1, \dots, N; c_0^i, i = 1, \dots, N\}$  and  $p(\cdot, \cdot)$  represents a suitable basis function. The presence of the vector  $\tilde{\mathbf{w}}$  in the argument of the basis functions provides powerful approximation properties and allows facing the well-known curse of dimensionality, i.e., the possible exponential growth in the number of free parameters needed to obtain a growing degree of accuracy [25], [26].

Among the possible basis functions in (19), we choose sigmoidal functions. Then, we have

$$\bar{\mathbf{f}}(\mathbf{I}, \mathbf{w}) = \text{col} \left\{ \sum_{l=1}^v c_l^i \cdot \sigma(\tilde{\mathbf{w}}_{1l}^T \mathbf{I} + \tilde{w}_{0l}) + c_0^i; i = 1, \dots, N \right\} \quad (20)$$

where  $\sigma(x) = (1/(1 + e^{-x}))$  is a sigmoidal activation function. With such a choice, we have  $\tilde{\mathbf{w}}_l = \text{col}\{\tilde{w}_{1l}, \tilde{w}_{0l}\}, l = 1, \dots, v$ . In the following, we shall define  $\bar{\boldsymbol{\theta}}$  as the output of the neural approximator, i.e.:  $\bar{\boldsymbol{\theta}} = \bar{\mathbf{f}}(\mathbf{I}, \mathbf{w})$ . In order to guarantee the fulfillment of the “continuous” channel constraints (13) (i.e.,  $\hat{\boldsymbol{\theta}}^c(t) \in \Theta_c, \forall t$ ), we compose the output of the neural network with a

“normalization operator”  $\mathbf{n}(\cdot)$ . We, thus, obtain the surrogate continuous bandwidth allocation  $\hat{\boldsymbol{\theta}}^c(k\hat{t})$  at any time  $k\hat{t}$  as

$$\hat{\boldsymbol{\theta}}^c(k\hat{t}) = \mathbf{n}[\bar{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w})] \quad (21)$$

where

$$\mathbf{n}(\bar{\boldsymbol{\theta}}) = \text{col} \left\{ K \cdot \sigma(\bar{\theta}_i) / \sum_{j=1}^N \sigma(\bar{\theta}_j); i = 1, \dots, N \right\}.$$

As to the capacity of the neural bandwidth allocation function to approximate the unknown optimal solution to Problem FS-SDRAP, the reader is referred to [28], where the composition of a feedforward neural network with a normalization operator is considered. Density properties are provided; moreover, it is shown that the number of parameters required to achieve an integrated square error of order  $O(1/v)$  is  $O(\psi)$ , i.e., it grows linearly with the dimension  $\psi$  of the input vector of the neural approximator (20).

The surrogate continuous bandwidth allocation  $\hat{\boldsymbol{\theta}}^c(k\hat{t})$  is then converted to the feasible bandwidth allocation  $\boldsymbol{\theta}^d(k\hat{t}) \in \Theta_d$ . In the following, we shall call “neural approximators” the functions (21) obtained as composition of the neural networks (20) and the normalization operators  $\mathbf{n}(\cdot)$ . Moreover, we shall call “neural bandwidth allocation function” the mappings made up of the composition of the neural approximators together with the application of the algorithm that projects each  $\hat{\boldsymbol{\theta}}^c(k\hat{t})$  to its discrete feasible neighbor  $\boldsymbol{\theta}^d(k\hat{t})$  and denote it as  $\hat{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w}) = \boldsymbol{\theta}^d(k\hat{t})$ .

It follows that a cost function is obtained by substituting the fixed structure of such neural bandwidth allocation functions into cost (18), which then depends on the parameter vector  $\mathbf{w}$  and, thus, leading to the following mathematical programming problem.

*Problem FD-SDRAP<sub>w</sub>*: Find the optimal parameter vector  $\mathbf{w}^*$  such that the cost

$$E_{\omega} L^{k\hat{t}}[\hat{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w}), \boldsymbol{\varphi}, \boldsymbol{\alpha}] \quad (22)$$

is minimized.

In this way, the functional optimization Problem FD-SDRAP has been reduced to an unconstrained nonlinear programming one.

#### B. Training Algorithm

To solve (22), we apply a gradient-based algorithm

$$\mathbf{w}^{h+1} = \mathbf{w}^h - \zeta \nabla_{\mathbf{w}} E_{\omega} L^{k\hat{t}}(\hat{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w}^h), \boldsymbol{\varphi}, \boldsymbol{\alpha}) \quad (23)$$

where  $\zeta$  is a fixed stepsize. However, the explicit computation of the expected cost and its gradient is a very hard task, even if closed-form formulas for the functional cost  $L^{k\hat{t}}(\cdot)$  were available. Following [25] and [28], we choose to compute a realization  $\nabla_{\mathbf{w}} L^{k\hat{t}}(\hat{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w}^h), \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$  instead of the gradient  $\nabla_{\mathbf{w}} E_{\omega} L^{k\hat{t}}(\hat{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w}^h), \boldsymbol{\varphi}, \boldsymbol{\alpha})$  and we apply the updating algorithm

$$\mathbf{w}^{h+1} = \mathbf{w}^h - \zeta_h \nabla_{\mathbf{w}} L^{k\hat{t}}(\hat{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w}^h), \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h) \quad (24)$$

where the index  $h$  denotes both the steps of the iterative procedure and the generation of the  $h$ th realization of the stochastic processes  $\boldsymbol{\varphi}$  and  $\boldsymbol{\alpha}$ . The probabilistic algorithm

(24) is related to the concept of *stochastic approximation* [35] and conditions for its convergence are related to the decreasing behavior of the step size  $\zeta_h$ . We have taken  $\zeta_h$  as  $\zeta_h = (c_1)/(c_2 + h)$ ,  $c_1, c_2 > 0$ , and we have also added a “momentum” term  $\rho \cdot (\mathbf{w}^h - \mathbf{w}^{h-1})$ ,  $\rho > 0$ , as is usually done in training neural networks. The components of the gradient  $\nabla_{\mathbf{w}} L^{k\hat{t}}(\hat{\mathbf{f}}(\mathbf{I}(k\hat{t}), \mathbf{w}^h), \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$  can be obtained by using the classical *backpropagation equations* for the training of neural networks [36]. For the sake of simplicity, in the following, we omit the dependence on the time variable ( $k\hat{t}$ ) and define  $z_l^i = \tilde{\mathbf{w}}_{0l}^{iT} \mathbf{I} + \tilde{w}_{0l}^i$ ,  $l = 1, \dots, v$ ,  $i = 1, \dots, N$ , as the input of each neural unit. Hence, we have (for  $i = 1, \dots, N$ )

$$\begin{aligned} \frac{\partial L}{\partial c_0^i} &= \frac{\partial L}{\partial \bar{\theta}_i} \\ \frac{\partial L}{\partial c_l^i} &= \frac{\partial L}{\partial \bar{\theta}_i} \cdot \sigma(z_l^i), \quad l = 1, \dots, v \\ \frac{\partial L}{\partial \tilde{w}_{0l}^i} &= \frac{\partial L}{\partial \bar{\theta}_i} \cdot c_l^i \cdot \sigma' [z_l^i], \quad l = 1, \dots, v \\ \frac{\partial L}{\partial \tilde{w}_{1l}^i} &= \frac{\partial L}{\partial \bar{\theta}_i} \cdot c_l^i \cdot \sigma' [z_l^i] \cdot I^T; \quad l = 1, \dots, v. \end{aligned} \quad (25)$$

The backpropagation procedure must be initialized by means of the quantities  $(\partial L / (\partial \bar{\theta}_i))$ ,  $i = 1, \dots, N$  (i.e., the gradient  $\nabla_{\bar{\theta}} L(\boldsymbol{\theta}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$ ). Unfortunately, in our case, such quantities cannot be obtained analytically as in [25], [26], and [28], because no closed form for the functional cost is available.

In order to avoid such a heavy drawback, during the offline training phase (24), we estimate the gradient  $\nabla_{\bar{\theta}} L(\boldsymbol{\theta}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$  by means of the IPA technique outlined in Section IV.

Operatively, at step  $h = 0$ , a random initialization of the vector  $\mathbf{w} = \mathbf{w}^0$  is performed. Then, a sample path on a time horizon  $[0, \Xi\hat{t} + \Sigma T]$  is generated according to its probability distribution. At time  $\bar{t} = \Xi\hat{t}$ , the information vector  $\mathbf{I}(\bar{t})$  is collected and the bandwidth reallocation  $\boldsymbol{\theta}^d(\bar{t}) = \hat{\mathbf{f}}(\mathbf{I}(\bar{t}), \mathbf{w})$  is computed by means of the neural bandwidth allocation function. Such a reallocation is applied over a time horizon  $[\bar{t}, \bar{t} + T]$  and the pseudo-gradient  $\nabla_{\boldsymbol{\theta}^c} L(\boldsymbol{\theta}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$  is approximated through IPA (11)–(12), namely,  $(\partial L_i(\boldsymbol{\theta}^c)) / (\partial \theta_i^c) = \phi_i \cdot (\partial L_i(\theta_i)) / (\partial \theta_i)$ ,  $\theta_i = \theta_i^d \cdot \phi_i$ ,  $\boldsymbol{\theta}^c \cong \boldsymbol{\theta}^d$ . The partial derivatives  $(\partial L / (\partial \theta_i))$ ,  $i = 1, \dots, N$  are, thus, obtained by applying the “chain rule,” i.e.

$$\begin{aligned} \frac{\partial L}{\partial \bar{\theta}_i} &= \sum_{j=1}^N \frac{\partial L^{k\hat{t}}}{\partial \hat{\theta}_j^c} \cdot \frac{\partial \hat{\theta}_j^c}{\partial \bar{\theta}_i}, \quad i = 1, \dots, N \\ \frac{\partial \hat{\theta}_j^c}{\partial \bar{\theta}_i} &= \begin{cases} K \cdot \sum_{m=1}^N \bar{\theta}_m / \left[ \sum_{m=1}^N \bar{\theta}_m \right]^2, & \text{if } i = j \\ -K \cdot \bar{\theta}_j / \left[ \sum_{m=1}^N \bar{\theta}_m \right]^2, & \text{if } i \neq j. \end{cases} \end{aligned} \quad (26)$$

Once the quantities  $(\partial L / (\partial \bar{\theta}_i))$ ,  $i = 1, \dots, N$  have been computed, we are able to apply (25) and calculate  $\mathbf{w}^1$  by means of (24). Then (step  $h = 2$ ), we move to  $\bar{t} = \Xi\hat{t} + T$ , collect the information vector  $\mathbf{I}(\bar{t})$  and repeat the same operations, thus, obtaining  $\mathbf{w}^2$ . The same procedure is applied until  $\bar{t} = \Xi\hat{t} + \Sigma T$  (step  $h = \Sigma$ ). Then, a new sample path is generated and a new simulation is performed. The same steps are repeated until some stopping criterion is verified.

### C. Remarks on the Extended Ritz Mathematical Framework

1) *Other Approximate Techniques*: Other approaches are known in the literature to approximate the solution of NP-hard optimization problems. Among them, reinforcement learning-neuro dynamic programming approaches (NDP) are employed to overcome the computational burden introduced by dynamic programming [7]. Through an offline Montecarlo simulation approach, NDP is able to estimate the corresponding optimal differential reward function, which is adopted online to optimize the performance with some further computational effort. In [7], a NDP technique is investigated to solve both the CAC and the routing issues for a multiservice broad-band network.

In this work, we have taken a somewhat different approach concerning the approximating structure. The idea is to estimate the optimal control law coming from a functional optimization formulation of the problem, as highlighted by (18).

2) *Computational Complexity*: The previously described training procedure (24) can be performed off line. In this respect, the required computational burden does not influence the online performance of the system. In real time, the optimized neural bandwidth allocation function  $\hat{\mathbf{f}}(\cdot, \mathbf{w}^*)$  is applied, thus, obtaining new resource allocations “almost instantly”. The only online computational effort is related to the  $O(N + 1)$  mapping of the surrogate bandwidth allocations  $\hat{\boldsymbol{\theta}}^c(t) \in \Theta_c$  to the discrete constrain set  $\Theta_d$ .

Moreover, if a PA technique is obtainable for the performance index of interest, no further conceptual difficulty is involved if one would want to employ the data collected online to perform real time “adjustment” of the parameters’ vector  $\mathbf{w}$  through (24).

3) *Generality of the Control Technique*: The proposed neural-based OLFC strategy is quite general, as it does not assume any hypothesis concerning the stochastic properties of the DES under investigation and, since the training algorithm can be performed through simulation inspection (or directly over the real system), it can be generalized for the optimization of other performance indexes and according to different traffic load statistics.

In the offline training phase, the behavior of the system can be described by a suitable *discrete event simulator*, which mimics the possible evolutions of the DES, thus, guaranteeing the possibility of computing gradient estimates of the functional cost. The system dynamic equations are not required explicitly. Hence, the technique reveals to be suitable for optimizing DESes whose dynamics and performance metrics are unavailable in closed form.

4) *On Sensitivity Estimation Literature*: Other approaches are known in the literature concerning DESes optimization through gradient estimates of the performance index of interest. PA has been already mentioned in this work. Moreover, dynamic control reactions based on gradient estimates, acting over Markov decision processes are studied in [37], [38]. It is worth noting that all these techniques act online. Therefore, suboptimal transient periods could arise in any case, even if variance reduction methods were applied [38]. On the other hand, after the training phase, the proposed neural OLFC approach yields near-optimal control laws online, thus, avoiding the need of further optimization steps. The following performance evaluation is devoted to highlight this promising capability.

5) *Offline Training Without PA*: Also without any PA technique, during the offline training phase, it is possible to obtain an estimation of  $(\partial/(\partial\theta_i^c))L(\boldsymbol{\theta}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$  by means of simulation analysis, namely, each  $(\partial/(\partial\theta_i^c))L(\boldsymbol{\theta}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$  can be computed as

$$\frac{\partial}{\partial\theta_i^c}L(\boldsymbol{\theta}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h) = L(\boldsymbol{\theta}_{i+}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h) - L(\boldsymbol{\theta}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h) \quad (27)$$

where  $\boldsymbol{\theta}_{i+}^d$  means that one further MAU has been added to  $\theta_i^d$  (the  $i$ th component of the discrete resource vector  $\boldsymbol{\theta}^d$ ). Clearly,  $L(\boldsymbol{\theta}_{i+}^d, \boldsymbol{\varphi}^h, \boldsymbol{\alpha}^h)$  is computed through simulation inspection only, as suggested in [1]. In this way, the PA technique reveals to be an instrument for accelerating our offline training phase, whereas, in the online surrogate methodologies, it is the main component of the optimization algorithm [6], [20], [23], [24].

In the following, we shall denote the application of this technique as “*ExtRitz*.”

#### D. Remarks on the Networking Environment

1) *Traffic Model*: Further considerations are needed concerning the adoption of the traffic model introduced in Section II-B, also used for the following performance evaluation.

In recent years, analyzes of packet-based traffic have demonstrated that its main statistical characteristics have good affinity to Self-Similar processes. Intuitively, self-similar traffic is supposed to present the same statistical behavior over largely different time intervals. W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson analyzed Ethernet traffic [39], highlighting its self-similar nature, as V. Paxson and S. Floyd [40] extended these observations to the TCP protocol over WANs. Besides, M.W. Garrett and W. Willinger [41] proposed this model also for video VBR traffic. Self-similar traffic arises for wireless networks, too. The self-similarity nature of traffic is mostly related to users’ behavior than to the underlying technology. For instance, in [42] and [43], four types of traffic profiles are modeled as self similar sources, on the basis of the most frequently used wireless applications: e-mail, www, file transfer protocol, and telemetry traffic. The adopted self-similar model allows us to mimic the statistical behavior of queuing systems of current telecommunication networks.

2) *Protocol Architectures and Performance Constraints*: Some additional words are necessary concerning the application of the proposed techniques in different networking scenarios. We mentioned a satellite system is dealt with in this work. The related network protocol architectures are typically based on ATM or DVB. In this context, we firstly must note that the proposed techniques are independent of the corresponding encapsulation formats of the packets (ATM or DVB). Actually, the measured  $i$ th loss volume is a function of the packet-based frame produced at station  $i$ . More specifically, our optimization framework could be related to a Best Effort network (based on the TCP/IP suite over DVB or on the ATM ABR service class), where no priority traffic exists and resources have to be shared among all the users. We therefore obtain the minimization of the overall loss probability without taking into account any specific QoS differentiation. On the other hand, QoS requirements could be taken into account, by formulating the network optimization problem at the call level (as done, for instance, in [6], [7], and [10]). In this case, the metric under investigation reveals to be the blocking

probability of the connection requests. Blocking probability (or other performance constraints, such as delay or delay jitter) can be managed without affecting the control design. The considered functional costs and their gradients being computed through active measurements, the proposed control algorithms are suitable for managing system dynamics whose temporal behaviors can be monitored through real time sampling.

Another important aspect is related to the effect of large delays over the system performance, in particular when a single centralized unit is responsible for the resource management. We recall that, in the network we are considering, a master station periodically implements the reallocations for all components of the system, but the calculations needed are performed at the earth stations. Thus, the only impact of delays in our allocation mechanism regards the signalling phase, required by the master station to perform the normalization and the subsequent projection onto the discrete domain. However, this delay in our specific context is the one introduced by the satellite, which is relatively low (in the order of 500 ms for a geostationary satellite), but fixed. As our allocation is performed at the data link layer, and regards the sharing of a single link, variable delays in the network would not influence it. Performing the allocation in one satellite round-trip delay is generally acceptable.

## VI. SIMULATION RESULTS

### A. Bandwidth Allocation Strategies

We now summarize the bandwidth allocation strategies adopted for comparison in the following simulation results. At the end of each simulation, the final loss volume of the overall satellite system is computed in terms of the average loss probability.

1) *CF&DP*: The certainty equivalent approach formulated in Section III is employed. For all stations in the satellite system, a perfect knowledge on both the state of traffic sources and the fading levels is assumed. The sensitivity of the solution obtained through CF&DP is investigated, with respect to possible estimation errors over the parameters of the traffic sources. The delay due to the computational burden introduced by dynamic programming is disregarded, fixing the reallocation time interval always to 1.0 s, independently of the related computational effort.

2) *SE&GD (Sensitivity Estimation and Gradient Descent)*: The optimization algorithm described in Section IV is adopted, and derivative estimations are computed. Different values of the gradient stepsize are taken into account. We denote with “SE&GD $\eta$ ” the SE&GD technique, in which the gradient stepsize  $\eta$  is set to a fixed value, e.g., “SE&GD.1E3” means that  $\eta$  is fixed to  $1 \cdot 10^3$ .

3) *ExtRitz*: the neural OLFC strategy of Section V is applied. The optimal parameters’ vector  $\mathbf{w}^*$  characterizing the neural bandwidth allocation function is obtained off line by means of the stochastic gradient technique described in Section V.

The bandwidth allocation strategy is implemented by adopting a neural network with 1 hidden layer with 15 sigmoidal neural units. The  $c_1, c_2, \rho$  parameters of the training algorithm (24) are fixed to 10.0, 1.0, 0.3, respectively. The depth  $\Xi$  of the time horizon of the information vector  $\mathbf{I}(\cdot)$  has been dimensioned in order to achieve a satisfying performance.

TABLE I  
TRAFFIC LOAD CHANGES

Time Interval (s)	0.0 -60.0	60.0-120.0	120.0-180.0
Station1, bursts/s	50.0	25.0	50.0
Station2, bursts/s	25.0	50.0	25.0
Time Interval (s)	180.0-240.0	240.0-300.0	300.0 -360.0
Station1, bursts/s	25.0	50.0	25.0
Station2, bursts/s	50.0	25.0	50.0

The IPA formulas (11) and (12) have been used in order to initialize the backpropagation procedure. The gradient descent (24) iterates the training steps with several independent replications of the following simulation scenarios until the stopping criterion is satisfied.

We developed a C++ simulator for the queues of the satellite network and for the optimization algorithms proposed in this paper. We verified the performance of a satellite system composed of two earth stations. The following loss measures are averaged over 15 independent replications of the simulation scenario. Initially, the number of stations has been intentionally kept to such low value, in order to better understand the behavior of the various techniques compared under different traffic pattern and fading situations. Increasing the number of stations to simulate more complex scenarios would not substantially change the evaluation, but it would make the comparison not so immediate and easy. Finally, however, we address the scalability of the neural control, by increasing the number of traffic stations.

### B. Traffic Load Changes

For the time being, we suppose no fading attenuation acting over the system (i.e.,  $\phi_i(t) = 1; i = 1, \dots, N; \forall t \in [0, T_s]$ ,  $T_s$  being the simulation time). Each source is supposed to transmit at a peak bit rate  $B_p = 1.0$  Mb/s, the total capacity  $K$  is fixed to 80.0 Mb/s ( $K = 800$  MAUs) and  $T_s$  to 6 min. The number of sources for each station is fixed at  $M_i = 100$  for all  $i$ . During the simulation scenario, the average values of active and silence periods of the sources (respectively,  $\bar{\tau}$  and  $\bar{\zeta}$ ) are changed according to Table I. Two different burst arrival rates (i.e., the average number of bursts “seen” by the stations) are taken into account in order to generate variable traffic conditions, leaving all the other traffic parameters equal for each station. In this way, the station in high traffic conditions (i.e., with  $\bar{\tau} = \bar{\zeta} = 1.0$ ) receives a burst arrival rate of  $\lambda_{\text{burst}}^i = (M_i)/(\bar{\tau}_i + \bar{\zeta}_i) = 50.0$  bursts/s,  $i = 1, 2$ , while, for the one in low traffic conditions, the burst arrival rate is  $\lambda_{\text{burst}}^i = 25.0$  bursts/s,  $i = 1, 2$ . (see Table I).

Both stations are provided with a finite buffer of 100 ATM cells, thus, guaranteeing a reasonable bound for the mean delay and delay jitter; e.g., with an allocation of 25.0 Mb/s (that is a lower bound of the following bandwidth allocations), such bound is around 2.0 ms for each ATM cell.

The reallocation period  $\hat{t}$  has been kept fixed at 1.0 s as outlined in Section IV. The master control station has to get the sensitivity estimation of each earth station to compute the next allocation of bandwidth in the SE&GD technique and the information vector  $\mathbf{I}(\cdot)$  in the ExtRitz strategy. The MAU is set to 100 kb/s as in [5], [10], and [24].

The depth  $\Xi$  of the time horizon of the information vector  $\mathbf{I}(\cdot)$  in the ExtRitz technique was set to 5, namely, the neural

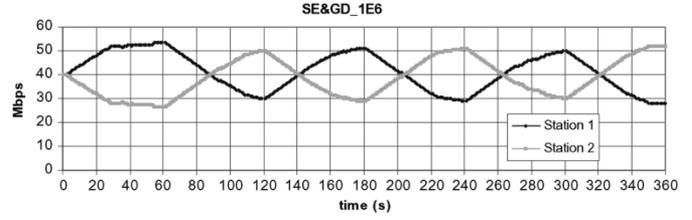


Fig. 2. Traffic load changes. SE&GD bandwidth allocation with different gradient step sizes.

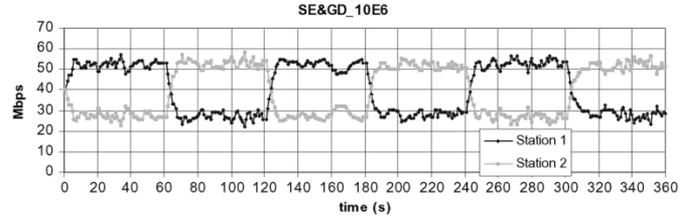


Fig. 3. Traffic load changes. SE&GD bandwidth allocation with different gradient step sizes.

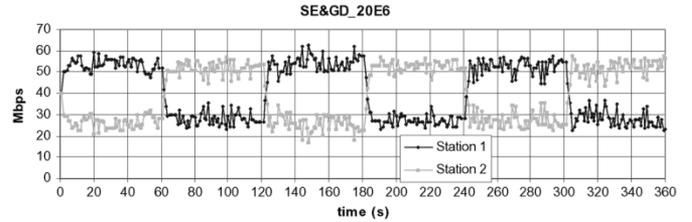


Fig. 4. Traffic load changes. SE&GD bandwidth allocation with different gradient step sizes.

bandwidth allocation strategy at time  $k\hat{t}$ ,  $k = 0, 1, \dots$  depends on the values assumed by the inflow processes  $\alpha(t)$  at the time instants  $(k-1)\hat{t}, (k-2)\hat{t}, \dots, (k-5)\hat{t}$ . Up to 200 independent replications of the proposed simulation scenario have been necessary to achieve a steady state in the ExtRitz’s bandwidth allocation. The simulation time of the training phase was around 10 h with an AMD Athlon at 1.8 GHz.

A sample path in the bandwidth allocation for the SE&GD technique is depicted in Figs. 2–4 where, for each station, the fraction of the total system’s capacity assigned by the SE&GD technique is visualized. It is clear how SE&GD is able to react to traffic variations: in the first minute, station 1 suffers of a heavier traffic load and a larger quantity of MAUs has been allocated to it. The situation is inverted in the second minute, and it is clear, in this case, how SE&GD provides more resources to station 2. In the third minute, the situation is brought back to the one of the beginning and the same behavior arises in the following 3 min. Looking at Figs. 2–4, it is also clear that different values of the stepsize  $\eta$  allow different behaviors in the bandwidth allocations; in particular, the suboptimal transient periods are much more evident with  $\eta = 10^6$ .

Looking at the ExtRitz bandwidth allocations (Fig. 5), it is clear how the ExtRitz is able to provide the optimal bandwidth allocations without any suboptimal transient period. This fact, clearly, has an impact on the system performance.

In Figs. 6–7, the CF&DP’s bandwidth allocations are depicted together with the ones obtained through ExtRitz. Also

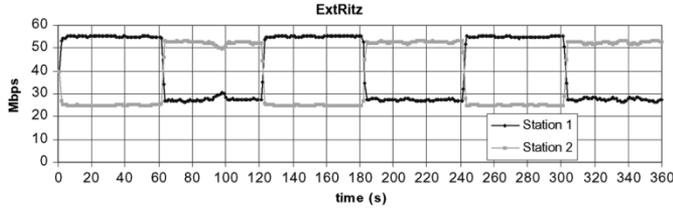


Fig. 5. Traffic load changes. ExtRitz bandwidth allocation.

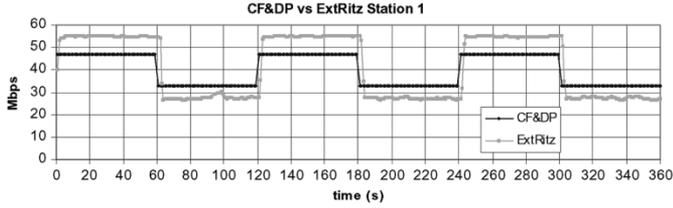


Fig. 6. Traffic load changes. ExtRitz versus CF&DP.

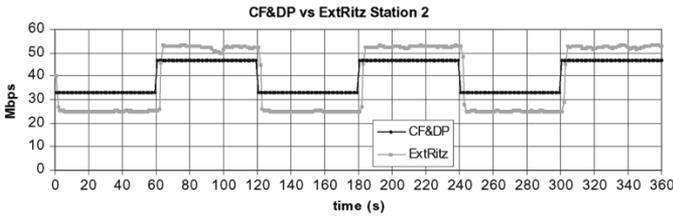


Fig. 7. Traffic load changes. ExtRitz versus CF&DP.

the CF&DP reveals to be a good heuristic for the bandwidth allocation, as it is able to follow the variable traffic conditions. Nevertheless, it does not maintain the best resource allocation: the bandwidth allocation to the station in heavier traffic load is lower than the one obtained by applying the ExtRitz (around 47 Mb/s using CF&DP in front of the 53 Mb/s obtained with the ExtRitz). This has an impact on the system performance, too.

In Fig. 8, the overall loss probability of the system is shown.

We denote by CF&DP\_ErrX the application of the CF&DP technique with a percentage error over the traffic load foreseen at station 1, i.e., we highlight the performance of the CF&DP technique in which the feedback over the state of the first station underestimates the real traffic load with a percentage error that amounts to X% over the real value. For example, with  $X = 10\%$ , a  $\lambda_{\text{burst}}$  of 45 bursts/s is employed in the functional cost (8) of the CF&DP technique, instead of the real value of 50 bursts/s. Such estimation errors severely decrease the CF&DP's performance, especially if they affect the feedback with a percentage error around 30%. On the other hand, with a perfect feedback on the system's state, the CF&DP technique reaches good performance, but only the application of the SE&GD and the ExtRitz techniques guarantees the best approximation of the optimal solution. They are able to achieve the optimal equilibrium in the bandwidth allocation, thus, guaranteeing better performance than the one reached by the CF&DP approach, even in the presence of a traffic load, whose statistical behavior is close to the assumptions adopted to provide a closed-form expression of the loss performance metric.

The intuition behind such efficiency relies on the sensitivity of the gradient descents (14) and (24). They are driven by IPA

to capture the system performance through measurements provided on the real system. We verified in our simulations that they follow, for each stationary configuration of the system (e.g., for each combination of  $\lambda_{\text{burst}}^i, i = 1, \dots, N$ ), an equilibrium point characterized by the equalization of all the loss probabilities of the stations. This equilibrium is also characterized by the equalization of all components of the gradient used in (14) and (24). On the other hand, the CF&DP technique, even if a perfect knowledge over the traffic sources' state is available, does not yield the optimal resource allocation. In fact, the  $P_{\text{Loss}}$  formula (8) holds *asymptotically* in the number of sources ([17]) (i.e., the number of sources in the aggregated flow should tend to infinity). In a realistic scenario with a finite number of on-off sources, it provides only a heuristic indication about the bandwidth needs of the stations.

Fig. 8 outlines how ExtRitz guarantees the best performance. The SE&GD<sub>10E6</sub> (i.e.,  $\eta = 10^7$ ) is able to reach a loss probability very close to the best one. It is worth noting that either SE&GD with too low gradient stepsize values (i.e.,  $\eta \in [10^1, 10^6]$ ) and with too high gradient stepsize values (i.e.,  $\eta \in [10^8, 10^9]$ ) fails in optimizing the system performance. With  $\eta \in [10^1, 10^6]$  the DES reaches the same loss probability of an equal bandwidth distribution among the stations (around  $9 \cdot 10^{-2}$ ) and with  $\eta \in [10^8, 10^9]$  loss probability values over 20% are reached. This latter case is due to the fact that, with  $\eta \geq 10^8$ , too high bandwidth oscillations are produced, thus, leading to a strong decrease in the system performance.

### C. Fading Changes

We consider now the effect of the fading phenomenon. The time horizon of the simulation scenario  $T_s$  has been increased to 1000 s (around 16 min) and the channel capacity  $K$  is fixed at 53.0 Mb/s ( $K = 530$  MAUs). Again, a group of on-off sources with Pareto distributed burst periods of activity constitutes the inflow process  $\alpha_i(t)$  for each station  $i$ . The peak bit rate  $B_p$ , the mean burst  $\bar{\tau}$  and silence period  $\bar{\zeta}$  of such on-off sources are fixed to 1.0 Mb/s, 1.0 s and 1.0 s, respectively. The number of on-off sources for each station is fixed at  $M_i = 10, i = 1, 2$ . All of the other system's parameters (buffer dimensions, MAU values and reallocation time interval's length) are equal to the ones of the previous simulation scenario. This time, no traffic changes take place, namely, each inflow process generates a  $\lambda_{\text{burst}}^i = 5$  bursts/s for each station  $i$  of the satellite system,  $i = 1, 2$ .

The employed fading processes come from [10], where real-life fading attenuation samples are taken from a data set chosen from the results of experiments, in *Ka band*, carried out on the *Olympus* satellite by the CSTS (*Centro Studi sulle Telecomunicazioni Spaziali*) Institute, on behalf of the *Italian Space Agency*. The up-link (30 GHz) and down-link (20 GHz) samples considered were 1-s averages, expressed in dB, of the signal power attenuation with respect to clear sky conditions. The *Carrier/Noise Power* ( $C/N_0$ ) factor is monitored at each station and, on the basis of its values, different bit and coding rates are applied, in order to limit the BER below a chosen threshold of  $10^{-7}$ . Six different fading classes are defined, corresponding to combinations of channel bit rate and coding rate that give rise to redundancy factors  $\Gamma_{\text{level}}^i, i = 1, 2, \text{level} = 1, \dots, 6 (\Gamma_{\text{level}}^i \geq 1.0)$ .  $\Gamma_{\text{level}}^i$  represents the ratio between the information bit rate

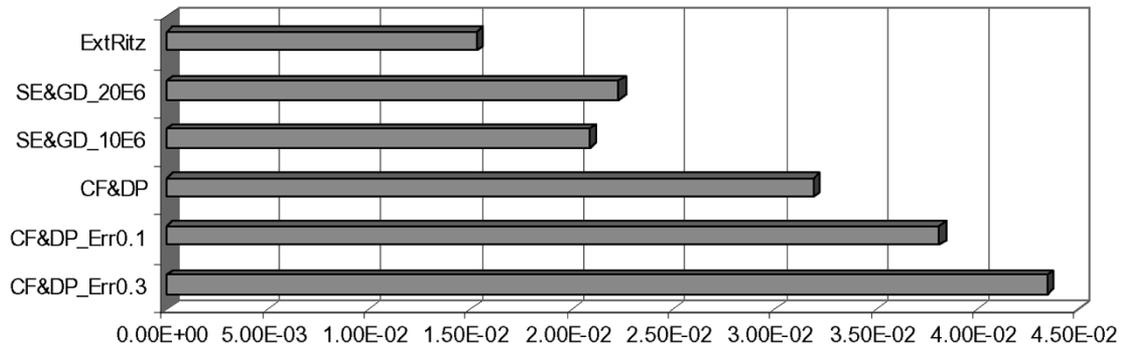


Fig. 8. Traffic load changes. Loss performance.

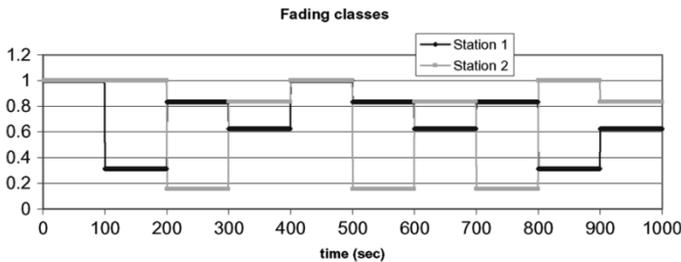


Fig. 9. Fading changes.

(IBR) in clear sky and the IBR in the specific working condition. This gives rise to corresponding bandwidth reduction factors  $\phi_i(t) = (1/\Gamma_{\text{level}}^i(t))$ . With the data adopted in [10] we have

$\phi_i(t) \in \{0.0, 0.15625, 0.3125, 0.625, 0.8333, 1.0\}$ ,  $i = 1, 2$  and, consequently, the bandwidth reduction can be computed as  $\theta_i(t) = \phi_i(t) \cdot \theta_i^d(t)$ ;  $\phi_i(t) = \frac{1}{\Gamma_{\text{level}}^i(t)}$  (the value  $\phi_i(t) = 0$  corresponds to an outage condition).

As is shown in Fig. 9, the employed fading processes determine strong peaks of channel degradation, especially for the second station.

The depth  $\Xi$  of the time horizon of the information vector  $\mathbf{I}(\cdot)$  in the ExtRitz technique was set to 3, namely, the neural bandwidth allocation strategy at time  $k\hat{t}$  depends on the values assumed by the fading processes  $\varphi(t)$  at the time  $(k-1)\hat{t}$  (i.e., the time of the previous bandwidth allocation) and at the time instants  $(k-2)\hat{t}$ ,  $(k-3)\hat{t}$ . 300 independent replications of the proposed simulation scenario have been necessary to achieve a steady state in the ExtRitz's bandwidth allocation. The simulation time of the training phase was around 10 h (on the same AMD Athlon at 1.8 GHz).

In Figs. 10 and 11, the bandwidth allocations of the SE&GD and the CF&DP techniques are depicted for each satellite station. This time, the application of the CF&DP technique guarantees more sensible differences in the bandwidth allocations than SE&GD. On the other hand, SE&GD reacts to fading variations only when the first strong attenuation at station 2 arises (after around 200 s of simulation). As is clear from Fig. 12, in which the bandwidth allocations obtained with the SE&GD and the CF&DP techniques are compared for the second station, SE&GD is not able to follow the variable fading conditions, due to their fast variations. However, it yields, for station 2, bandwidth allocations higher than the CF&DP's ones, especially during the time periods of lowest fading levels (time

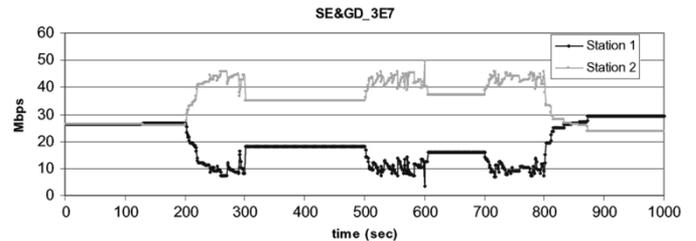


Fig. 10. Fading changes. SE&amp;GD\_3E7 bandwidth allocation.

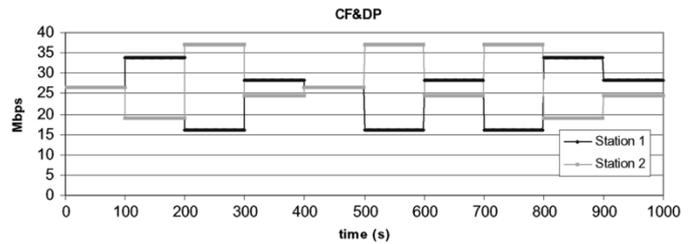


Fig. 11. Fading changes. CF&amp;DP bandwidth allocation.

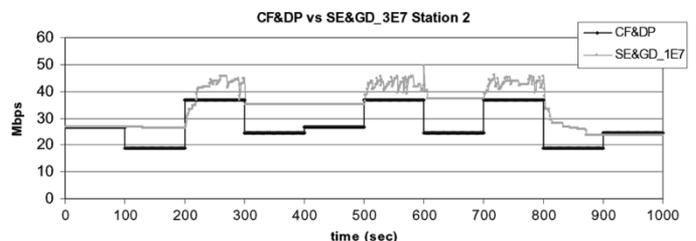


Fig. 12. Fading changes. CF&amp;DP versus SE&amp;GD\_3E7.

intervals [200, 300], [500, 600], [700, 800]). For this reason, after optimizing the gradient stepsize, as is shown in Fig. 13, the SE&GD outperforms the CF&DP's performance.

On the other hand, the ExtRitz strategy (Fig. 14) yields the best reallocation decisions, as it leads to a sequence of bandwidth reallocations sensibly different from the CF&DP's ones (Figs. 15 and 16) without any transient period.

In Fig. 17, the presence of estimation errors over the state of the traffic load is taken into account. As previously, "CF&DP\_ErrX" means that the feedback over the state of the first station underestimates the real traffic load with a percentage error of X%. This time, the CF&DP technique reveals a performance more insensitive with respect to estimation errors over the traffic state, because the major impact is due to the fading effect (Fig. 13). It is worth noting that the estimation

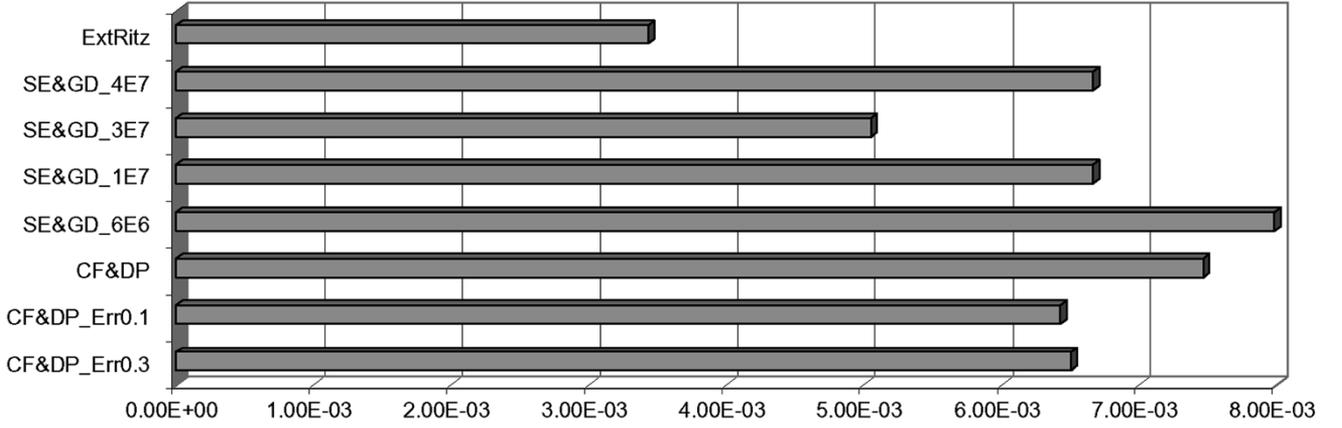


Fig. 13. Fading changes. Loss performance.

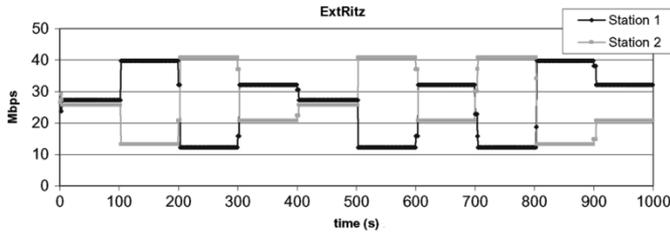


Fig. 14. Fading changes. ExtRitz allocations.

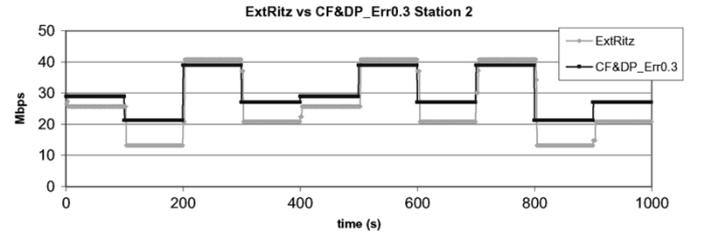


Fig. 17. Fading changes. ExtRitz versus CF&amp;DP\_Err0.3 at station 2.

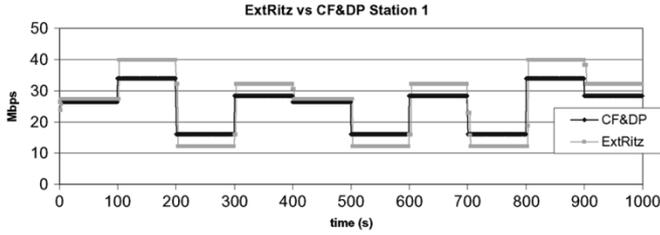


Fig. 15. Fading changes. CF&amp;DP versus ExtRitz.

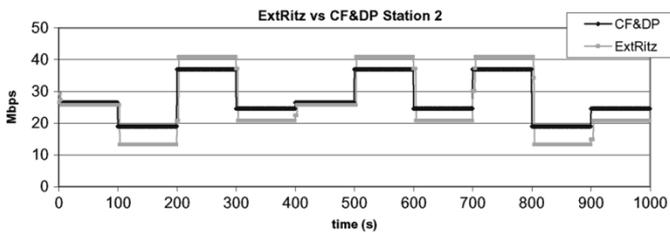


Fig. 16. Fading changes. CF&amp;DP versus ExtRitz.

errors over the real traffic state on station 1 allow the CF&DP technique to achieve better performance, as they lead to a bandwidth allocation closer to the optimal one in the time interval of the lowest fading levels for station 2 (Fig. 17).

#### D. Scalability of the Neural Control

The last topic addressed in the performance evaluation regards the scalability of the neural control. Our major concern relies on the computational time needed to complete the learning descent (24). Actually, in the presence of a high number of stations in heavy traffic conditions (either with a high number of sources or with high peak rates), the channel capacity should

be properly dimensioned to assure reasonable loss probabilities (usually  $10^{-2}$ ,  $10^{-3}$  are the performance targets). As a result, the involved high channel capacity implies a fine granularity of the simulation that severely affects the duration of the training phase. For instance, even a small cycle of training steps would require several hours to complete. Consequently, we need to find out a way to accelerate the ExtRitz training, without penalizing the good performance shown in previous subsections.

A possible way of facing this problem can be attempted with the following heuristic. We define a modified information vector, which depends on the arrival rate differences among the stations. Let  $m$  be the index of the station that measures the lowest arrival rate at a given time instant. We redefine the information vector as ruled by (28). It does not depend on the absolute values of arrival rates as in (15), whereas it exploits the levels of congestion, computed as the relative increase in the inflow rates with respect to the  $m$ th station (i.e., the one affected by the lightest traffic load)

$$\mathbf{I}(k\hat{t}) = \text{col}\{\theta^d((k - \Xi)\hat{t}), \dots, \theta^d((k - 1)\hat{t}), \\ \varphi((k - \Xi)\hat{t}), \dots, \varphi((k - 1)\hat{t}), \\ \tilde{\alpha}((k - \Xi)\hat{t}), \dots, \tilde{\alpha}((k - 1)\hat{t})\}$$

with

:

$$\varphi(k\hat{t}) = \text{col}\{\phi_1(k\hat{t}), \dots, \phi_N(k\hat{t})\} \\ \tilde{\alpha}(k\hat{t}) = \text{col}\left\{\frac{\alpha_1(k\hat{t}) - \alpha_m(k\hat{t})}{\alpha_m(k\hat{t})}, \dots, \frac{\alpha_N(k\hat{t}) - \alpha_m(k\hat{t})}{\alpha_m(k\hat{t})}\right\} \\ m = \arg \min_{i=1, \dots, N} [\alpha_1(h\hat{t}), \dots, \alpha_i(h\hat{t}), \dots, \alpha_N(h\hat{t})] \\ h = k - \Xi, \dots, k - 1. \quad (28)$$

The rationale behind this procedure stems from the fact that the neural networks produce at their output allocation values be-

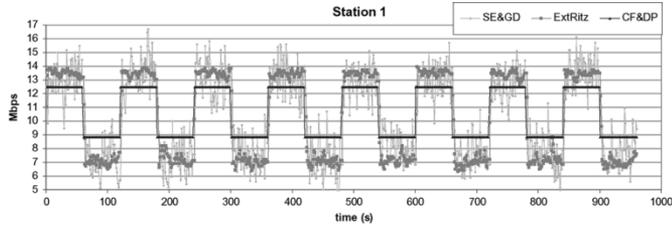


Fig. 18. Scalability. Allocations of the control techniques.

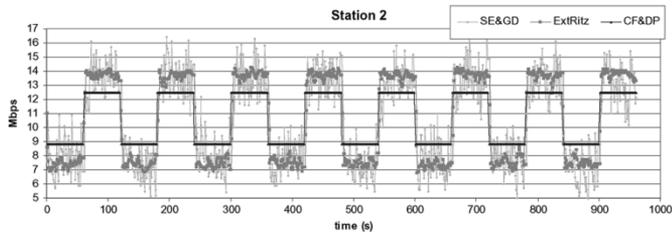


Fig. 19. Scalability. Allocations of the control techniques.

fore normalization and discretization; in such “unconstrained” allocations, relative congestion differences can be taken into account, which are then rescaled by the subsequent operations. This choice allows us to train the neural bandwidth allocation function with a small number of sources, with small source peak rates and with a low channel capacity. Therefore, the time computation of the involved training is significantly reduced. After training, the ExtRitz scales up its performance independently of the working conditions (in terms of traffic load and channel capacity). We validated this approach as detailed in the following.

We performed several tests with 5 stations in the system, by using the previous method. We took as target performance a loss probability around 2% (being, as for ITU-T.54 recommendation, the regular performance requirement of a *Voice over IP* service). The channel capacity is adapted in dependence of the number of active sources and of the given peak rates to guarantee the mentioned loss probability. The training phase is relative to 50 sources in each station producing 120 kb/s of peak rate each. The channel capacity is 12.0 Mb/s. The buffer size is 200 ATM cells. The depth of the information horizon is 3. The training parameters are fixed as:  $c_1 = 1$ ,  $c_2 = 2$ ,  $\rho = 0.3$ . Two different burst arrival rates have been considered as in Table I. The training procedure took around 1 h to complete. After training, we verified a stable performance increase by using ExtRitz. The increase, expressed in percentage, is 23% with respect to the CF&DP approach and 20% with respect to SE&GD. It means that when ExtRitz achieves 2% of loss probability, the losses obtained with CF&DP and SE&GD are 2.71% and 2.61%, respectively. The same performance is maintained when increasing the sources’ peak rate up to 500 kb/s and the channel capacity to 55.0 Mb/s, without repeating the training phase. The corresponding allocations are reported in Figs. 18–22. Note that the SE&GD allocations, owing to their strong oscillations, are impractical. The optimal steady states, achieved by ExtRitz, are sensibly different from the CF&DP’s ones. The mentioned SE&GD’s performance is relative to the best value of the employed gradient step size. However, the computation of the optimal gradient stepsize requires an extensive simulation inspection.

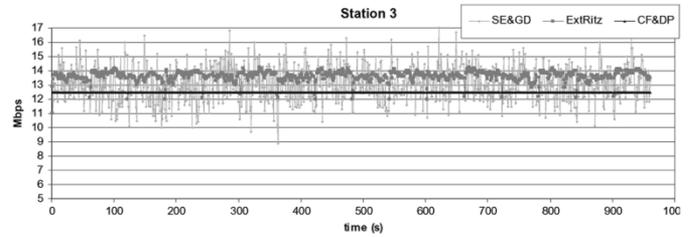


Fig. 20. Scalability. Allocations of the control techniques.

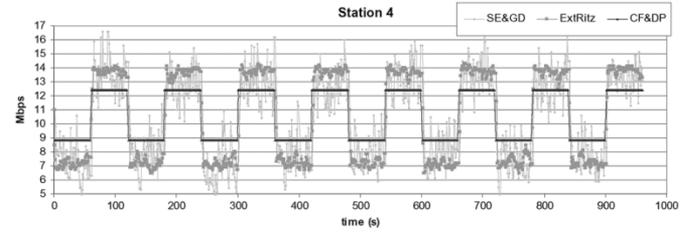


Fig. 21. Scalability. Allocations of the control techniques.

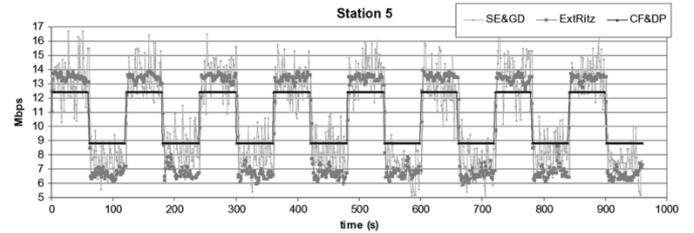


Fig. 22. Scalability. Allocations of the control techniques.

### E. Final Remarks

In the proposed simulation scenarios, clear steady states in the optimal allocations arise, with respect to each combination of fading classes and burst arrival rates. For instance, in the first simulation scenario, it is easily observable from Figs. 2–4 (SE&GD) and Fig. 5 (ExtRitz) that a bandwidth allocation around  $\theta_1^{d*} = 53.0$  Mb/s,  $\theta_2^{d*} = 27.0$  Mb/s reveals to be optimal with respect to the burst arrival rates  $\lambda_{\text{burst}}^1 = 50.0$  bursts/s,  $\lambda_{\text{burst}}^2 = 25.0$  bursts/s (and viceversa when the role of the station in higher traffic load is inverted). In spite of the convergence capability of the SE&GD technique, one could choose to adopt it off line, in order to compute, for each possible combination of fading classes and burst arrival rates, the optimal allocations; then, to store them in a lookup table and to reapply them in dependence of the current state of the network. In this way, we avoid the suboptimal transient periods introduced by the SE&GD approach. Thus, the investigation of more complicated control techniques (as the ExtRitz proposed in this work) would be useless. However, such heuristic assumes again a perfect knowledge about the traffic sources. As we outlined before, the maintenance of this knowledge is a very hard task. The adoption of adaptive strategies, capable to “learn” the best resource through real time measurements reveals to be topical. In particular, it constitutes a powerful tool for the optimization of DESes in which no closed-form formulas of the system dynamics are available. The SE&GD strategy catches the optimal resource allocation at the end of its gradient descent. On the other hand, after the training phase, the

ExtRitz yields the optimal resource allocation in dependence of the data collected in the information vector.

## VII. CONCLUSION AND FUTURE WORK

Two novel optimization algorithms have been proposed to react to fading and traffic load changes over a satellite network, based on an online surrogate optimization methodology and on a neural network-based OLFC approach, respectively. They allow the application of online control decisions in real operating conditions and with a small computational effort. The simulation results have shown that they guarantee good performance, even better than the one obtained by employing a strategy that maintains a perfect knowledge about the statistical properties of the traffic sources. A proper training algorithm has been used, aimed at the offline numerical approximation of the optimal OLFC law. Such training algorithm, differing from [25], [26], and [28], does not need any analytical expression of both the functional cost and the system dynamics.

The proposed OLFC approach shows the best performance, as it is capable to avoid the suboptimal transient periods of the online surrogate optimization technique. In fact, the optimal control law can be numerically approximated during the offline training phase, and then it can be employed online without further optimization steps. Such optimization technique reveals to be applicable and is currently being investigated in other network scenarios (for example in the contexts of terrestrial wireless or QoS networks) and with respect to different QoS requirements (such as delay and delay jitter).

## REFERENCES

- [1] K. Gokbayrak and C. G. Cassandras, "Online surrogate problem methodology for stochastic discrete resource allocation problems," *J. Opt. Theory Appl.*, vol. 108, no. 2, pp. 349–376, Feb. 2001.
- [2] —, "Generalized surrogate problem methodology for online stochastic discrete optimization," *J. Opt. Theory Appl.*, vol. 114, no. 1, pp. 97–132, Jul. 2002.
- [3] H. J. Chao and X. Guo, *Quality of Service Control in High-Speed Networks*. New York: Wiley, 2002.
- [4] N. Celandroni, E. Ferro, N. James, and F. Potortí, "FODA/IBEA: A flexible fade countermeasure system in user oriented networks," *Int. J. Satellite Commun.*, vol. 10, no. 6, pp. 309–323, Nov.–Dec. 1992.
- [5] R. Bolla, N. Celandroni, F. Davoli, E. Ferro, and M. Marchese, "Bandwidth allocation in a multiservice satellite network based on long-term weather forecast scenarios," *Comput. Commun.*, vol. 25, pp. 1037–1046, Jul. 2002.
- [6] K. Gokbayrak and C. Cassandras, "Adaptive call admission control in circuit-switched networks," *IEEE Trans. Autom. Control*, vol. 47, no. 6, pp. 1234–1248, Jun. 2002.
- [7] P. Marbach, O. Milhatsch, and J. Tsitsiklis, "Call admission control and routing in integrated services networks using neuro-dynamic programming," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 2, pp. 197–208, Feb. 2000.
- [8] C. M. Barnhart, J. E. Wieselthier, and A. Ephremides, "Admission control policies for multi-hop wireless networks," *Wireless Netw.*, vol. 1, no. 4, pp. 373–387, 1995.
- [9] G. Apostolopoulos, R. Guerin, S. Kamat, A. Orda, and S. K. Tripathi, "Intradomain QoS routing in IP networks: A feasibility and cost/benefit analysis," *IEEE Network*, vol. 13, no. 5, pp. 42–54, Sep.–Oct. 1999.
- [10] N. Celandroni, F. Davoli, and E. Ferro, "Static and dynamic resource allocation in a multiservice satellite network with fading," *Int. J. Satellite Commun. Netw.*, vol. 21, no. 4–5, pp. 469–487, July–Oct. 2003.
- [11] N. J. Keon and G. Anandalingam, "Optimal pricing for multiple services in telecommunications networks offering quality-of-service guarantees," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 66–80, Feb. 2003.
- [12] A. Iera, A. Molinaro, and S. Marano, "Call admission control and resource management issues for real-time VBR traffic in ATM-satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 2393–2403, Nov. 2000.
- [13] J. M. Pitts and J. A. Schormans, *Introduction to IP and ATM Design and Performance*, 2nd ed. New York: Wiley, 2000.
- [14] R. Bolla, F. Davoli, and M. Marchese, "Bandwidth allocation and admission control in ATM networks with service separation," *IEEE Commun. Mag.*, vol. 35, no. 5, pp. 130–137, May 1997.
- [15] T. Ibaraki and N. Katoh, *Resource Allocation Problems: Algorithmic Approaches*. Cambridge, MA: MIT Press, 1988.
- [16] R. Bolla, F. Davoli, M. Marchese, and M. Perrando, "Call admission control and routing of QoS-aware and best effort flows in an IP-over-ATM networking environment," in *Lecture Notes in Computer Science*. New York: Springer-Verlag, Jan. 2001, pp. 1989, pp. 33–49.
- [17] B. Tsybakov and N. D. Georganas, "Self-similar traffic and upper bounds to buffer-overflow probability in an ATM queue," *Perform. Eval.*, vol. 32, pp. 57–80, 1998.
- [18] D. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Belmont, MA: Athena Scientific, 2001.
- [19] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Norwell, MA: Kluwer, 1999.
- [20] F. J. Vazquez-Abad, C. G. Cassandras, and V. Julka, "Centralized and decentralized asynchronous optimization of stochastic discrete-event systems," *IEEE Trans. Autom. Control*, vol. 43, no. 5, pp. 631–655, May 1998.
- [21] Y. Wardi, B. Melamed, C. G. Cassandras, and C. G. Panayiotou, "Online IPA gradient estimators in stochastic continuous fluid models," *J. Opt. Theory Appl.*, vol. 115, no. 2, pp. 369–405, Nov. 2002.
- [22] C. G. Cassandras, G. Sun, C. G. Panayiotou, and Y. Wardi, "Perturbation analysis and control of two-class stochastic fluid models for communication networks," *IEEE Trans. Autom. Control*, vol. 48, no. 5, pp. 23–32, May 2003.
- [23] M. Marchese and M. Mongelli, "On-line bandwidth control for quality of service mapping over satellite independent service access points," *Comput. Netw.* Special Issue on Network Modeling and Simulation, to be published.
- [24] F. Davoli, M. Marchese, and M. Mongelli, "Optimal resource allocation in satellite networks: Certainty equivalent approach versus sensitivity estimation algorithms," *Int. J. Commun. Syst.*, vol. 17, no. 10, pp. 935–962, Dec. 2004.
- [25] R. Zoppoli, M. Sanguineti, and T. Parisini, "Approximating networks and extended Ritz method for solution of functional optimization problems," *J. Opt. Theory Appl.*, vol. 112, no. 2, pp. 403–439, Feb. 2002.
- [26] T. Parisini and R. Zoppoli, "Neural networks for feedback feedforward nonlinear control systems," *IEEE Trans. Neural Netw.*, vol. 5, pp. 436–449, 1994.
- [27] V. Kurkova and M. Sanguineti, "Comparison of worst case errors in linear and neural network approximation," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 264–275, Jan. 2002.
- [28] M. Baglietto, T. Parisini, and R. Zoppoli, "Distributed-information neural control: The case of dynamic routing in traffic networks," *IEEE Trans. Neural Netw.*, vol. 12, no. 3, pp. 485–502, May 2001.
- [29] F. Alagoz, B. Vojcic, D. Walters, A. Alrustamani, and R. Pickholtz, "Fixed versus adaptive admission control in direct broadcast satellite networks with return channel systems," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 2, pp. 238–249, Feb. 2004.
- [30] A. El Gamal, E. Uysal, and B. Prabhkar, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. Infocom*, vol. 1, Anchorage, AK, Apr. 22–26, 2001, pp. 386–394.
- [31] A. Fu, E. Modiano, and J. N. Tsitsiklis, "Optimal energy allocations and transmission control for communications satellites," in *Proc. Infocom*, vol. 2, New York, Jun. 2002, pp. 648–656.
- [32] B. Tsybakov, "File transmission over wireless fast fading downlink," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2323–2337, Aug. 2002.
- [33] P. M. Krishna, M. V. Gadre, and B. D. Uday, *Multifractal Based Network Traffic Modeling*. Norwell, MA: Kluwer, 2003.
- [34] K. W. Ross, *Multiservice Loss Models for Broad-band Telecommunication Networks*. London, U.K.: Springer-Verlag, 1995.
- [35] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.
- [36] S. Haykin, *Neural Networks, a Comprehensive Foundation*. New York: MacMillan Publishing, 1994.
- [37] P. Marbach and J. N. Tsitsiklis, "Simulation-based optimization of Markov reward processes," *IEEE Trans. Autom. Control*, vol. 46, no. 2, pp. 191–209, Feb. 2001.

- [38] P. Marbach and J. Tsitsiklis, "Approximate gradient methods in policy-space optimization of Markov reward processes," *J. Discrete Event Dyn. Syst.*, vol. 13, no. 1–2, pp. 111–148, Jan.–Apr. 2003.
- [39] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, 1994.
- [40] V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, 1995.
- [41] M. W. Garret and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," in *Proc. SIGCOMM*, London, U.K., 1994, pp. 269–280.
- [42] A. Jamalipour, *The Wireless Mobile Internet: Architectures, Protocols and Services*. New York: Wiley, 2003.
- [43] J. Ho and Y. Zhu, "Wireless packet data traffic profile," Nortel Networks Res. Rep., Feb. 1999.



**Marco Baglietto** (M'04) was born in Savona, Italy, in 1970. He received the Laurea degree in electronic engineering and the Ph.D. degree in electronic engineering and computer science from the University of Genoa, Genoa, Italy, in 1995 and 1999, respectively.

Since 1999, he has been an Assistant Professor of Automatic Control in the Department of Communications, Computer and Systems Science (DIST), University of Genoa. His research interests include neural approximations, linear and nonlinear estimation, distributed-information control systems.



**Franco Davoli** (M'90–SM'99) received the Laurea degree in electronic engineering from the University of Genoa, Genoa, Italy, in 1975.

Since 1990, he has been a Full Professor of Telecommunication Networks at the University of Genoa, where he is with the Department of Communications, Computer and Systems Science (DIST). From 1989 to 1991 and from 1993 to 1996, he was also with the University of Parma, Parma, Italy. His past research activities have included adaptive and decentralized control, large scale systems, routing

and multiple access in packet-switched communication networks, packet radio networks. His current research interests are in bandwidth allocation, admission control and routing in multiservice networks, wireless mobile and satellite networks and multimedia communications and services. He has co-authored over 210 scientific publications in international journals, book chapters and conference proceedings. He is a member of the Editorial Board of the *International Journal of Communication Systems* and of the international journal *Studies in Informatics and Control*, and an Associate Editor of *Simulation—Transactions of the SCS*. He has been a guest co-editor of two Special Issues of the *European Transactions on Telecommunications* and of a Special Issue of the *International Journal of Communication Systems*. In 2004, he was the recipient of an Erskine Fellowship from the University of Canterbury, Christchurch, New Zealand, as a Visiting Professor. He has been a Principal Investigator in a large number of projects and has served in several positions in the Italian National Consortium for Telecommunications (CNIT). He was the Head of the CNIT National Laboratory for Multimedia Communications in Naples in the years 2003–2004, and he is currently Vice-President of the CNIT Management Board.



**Mario Marchese** (S'94–M'97–SM'04) was born in Genoa, Italy in 1967. He received the Laurea degree (*cum laude*) from the University of Genoa, Genoa, Italy, in 1992, the Qualification as a Professional Engineer in 1992, and the Ph.D. degree in telecommunications from the University of Genoa in 1996.

From 1999 to 2004, he worked with the the University of Genoa Research Unit, Italian Consortium of Telecommunications (CNIT), where he was Head of Research. Since 2005, he has been an Associate Professor in the Department of Communications, Computer and Systems Science (DIST), University of Genoa. He is the Founder and still the technical responsible of CNIT/DIST Satellite Communication and Networking Laboratory (SCNL), University of Genoa, which contains high value devices and tools and implies the management of different units of specialized scientific and technical personnel. He is Vice-Chair of the IEEE Satellite and Space Communications Technical Committee. He is author and coauthor of more than 80 scientific works, including international journals, international conferences and book chapters. His main research interests include satellite networks, transport layer over satellite and wireless networks, quality of service over ATM, IP, and MPLS, and data transport over heterogeneous networks.



**Maurizio Mongelli** (S'04) was born in Savona, Italy in 1975. He received the Laurea (*cum laude*) and Ph.D. degrees from the University of Genoa, Genoa, Italy, in 2000 and 2004, respectively.

He worked for both Selenia S.p.A and the Italian Consortium of Telecommunications (CNIT), University of Genoa Research Unit, from 2000 to 2004. He is currently a Member of the Research Staff of the Telecommunication Networking Research Group, University of Genoa, with a postdoctoral scholarship founded by Selenia S.p.A. His main research interests include QoS protocol architectures and management functions, QoS interworking, resource allocation, optimization algorithms, and pricing for telecommunication systems.