

# Adaptive Call Admission and Bandwidth Control in DVB-RCS Systems

Mario Marchese and Maurizio Mongelli

**Abstract:** The paper presents a control architecture aimed at implementing bandwidth optimization combined with call admission control (CAC) over a digital video broadcasting (DVB) return channel satellite terminal (RCST) under quality of service (QoS) constraints. The approach can be applied in all cases where traffic flows, coming from a terrestrial portion of the network, are merged together within a single DVB flow, which is then forwarded over the satellite channel. The paper introduces the architecture of data and control plane of the RCST at layer 2. The data plane is composed of a set of traffic buffers served with a given bandwidth. The control plane proposed in this paper includes a layer 2 resource manager (L2RM), which is structured into decision makers (DM), one for each traffic buffer of the data plane. Each DM contains a virtual queue, which exactly duplicates the corresponding traffic buffer and performs the actions to compute the minimum bandwidth need to assure the QoS constraints. After computing the minimum bandwidth through a given algorithm (in this view the paper reports some schemes taken in the literature which may be applied), each DM communicates this bandwidth value to the L2RM, which allocates bandwidth to traffic buffers at the data plane. Real bandwidth allocations are driven by the information provided by the DMs. Bandwidth control is linked to a CAC scheme, which uses current bandwidth allocations and peak bandwidth of the call entering the network to decide admission. The performance evaluation is dedicated to show the efficiency of the proposed combined bandwidth allocation and CAC.

**Index Terms:** Call admission control (CAC), digital video broadcasting (DVB)-return channel via satellite (RCS)/second generation (S2), measurement-based equivalent bandwidth, quality of service (QoS) mapping.

## I. INTRODUCTION

### A. Technological Scenario

This paper deals with a satellite network, based on the digital video broadcasting (DVB)-return channel via satellite (RCS) standard [1], [2], composed of a geostationary orbit (GEO) stationary bent-pipe satellite, return channel satellite terminals (RCSTs), and a network control center (NCC) that is connected to the Internet (Fig. 1). Local area networks (LANs) may be connected to RCSTs. RCSTs are fixed and use the return channel via satellite (RCS). The NCC provides control and monitoring functions and manages network resources allocation to RCSTs. DVB-satellite (DVB-S) is used for the forward link (from NCC to RCSTs, with a data rate higher than 10 Mbps) and DVB-RCS is employed for the return link (from RCSTs to NCC, with a data

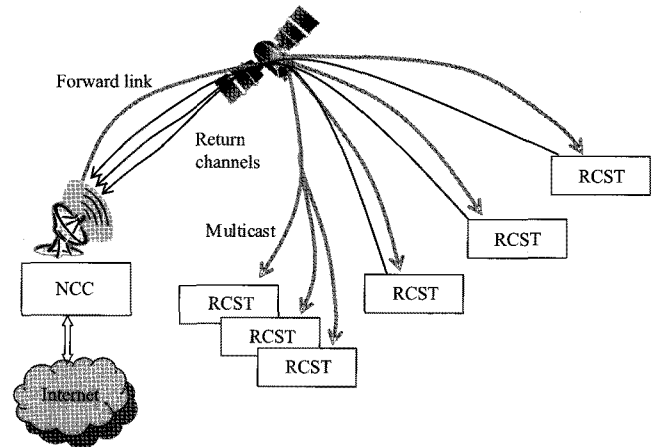


Fig. 1. DVB-RCS-second generation (S2) communication system.

rate around 2 Mbps). In DVB-S, the multi protocol encapsulation (MPE) provides segmentation and reassembly of IP packets, thus providing the motion picture experts group (MPEG2) stream whose packets size amounts of 188 bytes. The transmission is performed through channel coding and modulation. The DVB-RCS air interface is based on multi frequency-time division multiple access. The NCC assigns a group of slots, characterized by frequency, bandwidth, start time, and duration, to each RCST, and communicates resource allocations to RCSTs through the terminal burst time plan. DVB-RCS makes use of 5 DVB classes, each implemented through a dedicated queue at DVB layer, and 5 corresponding resource allocation types.

- i) Continuous rate assignment (CRA),
- ii) Rate-based dynamic capacity (RBDC),
- iii) Volume-based dynamic capacity (VBDC), where an RCST dynamically and cumulatively requests the total number of slots needed to idle its queue,
- iv) Absolute volume-based dynamic capacity (AVBDC), where an RCST dynamically requests the number of slots, but requests are not cumulative,
- v) Free capacity assignment (FCA).

VBDC and AVBDC, which imply time-varying bandwidth requests, and CRA, whose requirements affect the overall available bandwidth that can be dynamically allocated, are meaningful for this paper.

### B. State of the Art and Motivations

The problem of bandwidth allocation among the RCSTs to satisfy quality of service (QoS) levels naturally arises in DVB environments, especially for the return channel where bandwidth is a scarce resource. This, in turn, leads to the problem of QoS mapping of IP flows over the DVB classes (see [3],

Manuscript received April 30, 2010.

The authors are with the Department of Communications, Computer, and Systems Science, University of Genova, Via Opera Pia 13, 16145, Genova, Italy, email: {mario.marchese, maurizio.mongelli}@unige.it.

[4], and references therein). This topic received the attention of the satellite community in the last years for what concerns research projects (e.g., [5]–[7]) and scientific literature (an excellent overview can be found in [8]). The nature of the problem recalls the principles of “adaptive feedback control”, as DVB classes dynamically ask and release bandwidth resources on the basis of measures of the traffic flows. The overall aim of dynamic bandwidth control is to avoid wasting resources. Over provisioning of a-priori allocations is highly inefficient in time-varying conditions. Several works address the dynamic bandwidth control problem in DVB networks (see [9] and references therein), by exploiting optimal control methodologies, e.g., [10], and transmission control protocol (TCP) adaptations to the satellite, e.g., [11]. A peculiarity of the DVB technology, not addressed by the mentioned literature, is that bandwidth allocation must be implemented when IP flows with different statistics and QoS constraints are merged into a smaller set of DVB streams at layer 2 [4], [6], [7]. This action is called vertical QoS mapping.

### C. Aim of This Paper

The underlying idea of this paper is to design control blocks and actions of a combined call admission control (CAC)-bandwidth allocation scheme, implemented within the RCST layer 2 resource manager (L2RM). The design is addressed by investigating theoretical aspects and protocol rules, and it is heavily based on the concept of virtual queues. Given a real traffic buffer within each RCST, a virtual queue is another queue, mirroring the real traffic one, which receives data and can perform measures and actions without interfering with routine traffic forwarding. The words “buffer” and “queue” are used indifferently in this paper. The specific algorithm to compute the minimum bandwidth, commonly called “equivalent bandwidth” in the scientific literature, is performed through schemes taken in the literature and based on traffic measures. The idea of driving the CAC with measurement-based equivalent bandwidth is derived from [12] and [13]; the computation is extended to heterogeneous conditions in [14] and to the presence of noisy wireless channels [4]. A huge quantity of literature addresses the concept of bandwidth and CAC control, see, e.g., [15] and [16]. Additionally, the solution proposed in this paper is aimed at practical implementation and use in the field. For this motivation not only scientific literature is referenced but also active patents such as [17], which uses the concept of virtual queue to drive bandwidth allocation as in this paper, and [18], which may be applicable as a possible control law for the computation of the bandwidth need. Bandwidth availability over satellites is often time variant both because of fading and of terminal mobility. This aspect is mathematically modeled in this paper through a multiplicative factor that decreases the bandwidth available for data transmission. The paper is structured as follows: The next section contains the proposed architecture of the L2RM, which includes virtual queues and decision makers (DMs). Section III introduces the action of virtual queue linked to bandwidth allocation and describes the role of DMs. Section IV presents the bandwidth allocation action at L2RM and the CAC rule, and discusses the possible impact of time variant bandwidth availability over bandwidth allocation. Section V reports some algorithms taken in the literature and used within DMs. Section VI extends

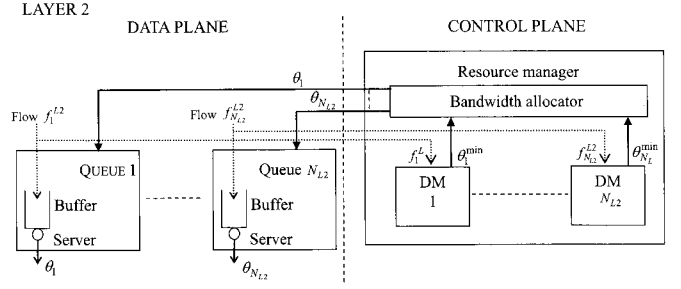


Fig. 2. RCST Layer 2.

the proposed architecture and control solution when more than one flow is multiplexed over one DVB queue. Section VII shows the performance evaluation and Section VIII the conclusions.

## II. ARCHITECTURE OF RCST LAYER 2 RESOURCE MANAGER (L2RM)

Fig. 2 shows L2 data and control planes of the RCST. Given  $N_{L2}$  traffic flows at layer 2, the data plane is modeled as a set of traffic buffers, one for each flow,  $(f_1^{L2}, \dots, f_{N_{L2}}^{L2})$ , served with rate  $\theta_1, \dots, \theta_{N_{L2}}$ , respectively. The layer 2 receives the traffic flows from the layer 3, which is not the object of this paper. A proposal concerning formal communication between layer 3 and layer 2 is reported in [19]: The IP protocol stack is divided into lower layers (layer 2 and 1), called satellite dependent (SD) layers, and upper layers (layer 3-IP and above layers), called satellite independent (SI) layers. The interface between SI and SD layers (in practice the interface between IP and layer 2) is defined through an interface called SI-service access point (SI-SAP), which provides a set of communication primitives for IP-layer 2 communication and assures the separation of SI, independent of satellite technology, and SD layers, strictly dependent on the used satellite technology. In this formal context, IP traffic flows, as well as IP flows performance requirements, if any, can flow through the SI-SAP and must be mapped at layer 2 (DVB, in this paper). The control plane contains the L2RM, which is composed of  $N_{L2}$  DMs, whose role, specified in detail in the following, is to compute an estimation  $(\theta_1^{\min}, \dots, \theta_{N_{L2}}^{\min})$  of the minimum bandwidth necessary at traffic buffers to provide a given quality of service  $(QoS_1, \dots, QoS_{N_{L2}})$ . After computing the minimum bandwidths, the vector  $(\theta_1^{\min}, \dots, \theta_{N_{L2}}^{\min})$  is forwarded to the bandwidth allocator block that decides and communicates to traffic buffers real bandwidth allocations  $(\theta_1, \dots, \theta_{N_{L2}})$ . Real bandwidth allocations should depend on the values of the minimum bandwidth. Mathematically,  $\theta_1 = f(\theta_1^{\min}), \dots, \theta_{N_{L2}} = f(\theta_{N_{L2}}^{\min})$ . A proposal concerning function  $f(\cdot)$  is reported in the following.

## III. DECISION MAKERS: VIRTUAL QUEUES AND MINIMUM BANDWIDTH COMPUTATION

The architecture of generic  $i$ th DM is shown in Fig. 3. Each traffic flow  $(f_1^{L2}, \dots, f_{N_{L2}}^{L2})$  is divided into two parts. One is directed towards the corresponding traffic buffer to be forwarded to the physical layer as shown in Fig. 2. The other part is sent to another buffer, the “virtual queue”, which is an exact mirror of the traffic one but it does not interfere with routinary forwarding

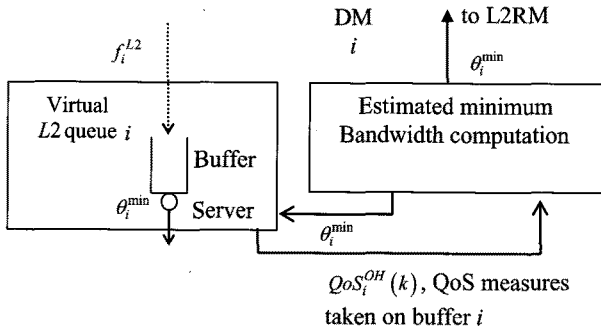


Fig. 3. DMi architecture.

operations. “Virtual queues” are one of the most important concepts within the block schemes introduced in this paper. Traffic buffers and virtual queues work in parallel but virtual queues are served with the estimated minimum rates  $(\theta_1^{\min}, \dots, \theta_{N_{L2}}^{\min})$ , computed by the “estimated minimum bandwidth computation” block in each DM. DMi computes  $\theta_i^{\min}$  by following a set of steps. A sequence  $k = 1, 2, \dots$  of observation horizons (OHs) for DMi is defined  $(OH_i(k))$ , during which the virtual queue is monitored in appropriate instants of time. These instants and monitoring information compose an information vector  $I_i(k)$  for each  $OH_i(k)$ .  $I_i(k)$  drives the service rate computation of queue  $i$  at time  $k + 1$ , together with the previous allocations up until a time depth  $d$ , thus generating  $\theta_i^{\min}(k + 1)$ .

$$\theta_i^{\min}(k + 1) = F(\theta_i^{\min}(k), \dots, \theta_i^{\min}(k - d), I_i(k), \dots, I_i(k - d)). \quad (1)$$

The actions are repeated by DMi during each  $OH_i(k)$ ,  $k = 1, 2, \dots$ . An example of information vector may be represented by the vector of differences  $e_1, \dots, e_{N_{L2}}$  between the QoS levels measured over the virtual queues during the observation horizons  $OH_i(k)$  and denoted by  $QoS_i^{OH}(k)$ , and the QoS threshold values  $QoS_i$ :  $e_i(\cdot, k) = (QoS_i - QoS_i^{OH}(k))^2$ . A possible simplification of (1) is contained in (2), where  $\theta_i^{\min}(k + 1)$  is computed by using only information at instant  $k$ .

$$\theta_i^{\min}(k + 1) = F(\theta_i^{\min}(k), e_i(\cdot, k)). \quad (2)$$

Fig. 4 shows the steps to get  $\theta_i^{\min}(k + 1)$  taking (2) as a reference. Usually, the temporal dimension of the  $OH_i(\cdot)$  is in the range  $[30, 360]$  s, depending on the specific applications to be monitored within queue  $i$ . Examples of algorithms to compute  $\theta_i^{\min}$  are reported in Section V. Actually, the specific minimum bandwidth computation is not part of the novelty of this paper but it may be interesting to have an idea of possible solutions leaving full details to specific papers in the literature. After getting  $\theta_i^{\min}$  each DMi implements a scheme to evaluate the stabilization of  $\theta_i^{\min}$  computation. The steady state of  $\theta_i^{\min}$ ,  $\forall i \in [1, \dots, N_{L2}]$  is captured under the following condition:  $|\theta_i^{\min}(k + 1) - \theta_i^{\min}(k)| \leq \epsilon_i$ , where  $\epsilon_i$  is the stabilization threshold for queue  $i$ . A reasonable practical value of  $\epsilon_i$  might be:  $\epsilon_i = \theta_i^{\min}(k)/10$ . This means  $\theta_i^{\min}$  is in steady state if it shows small oscillations between two consecutive observation horizons. More refined stabilization conditions may be used and

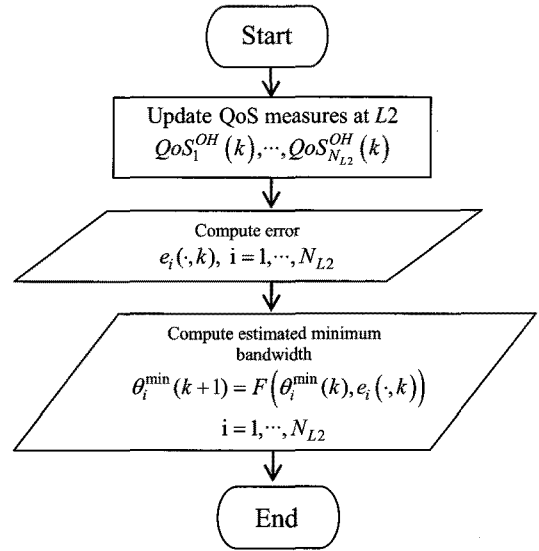


Fig. 4. Estimated minimum bandwidth computation by DMi.

reasonably applied in the context of the paper. The replication process of the packets towards virtual queue (together with the computations to obtain  $\theta_i^{\min}$ ) may require the application of a dedicated chip in case of computational limitation of the native hardware structure of the RCST. After evaluating the stabilization of  $\theta_i^{\min}$ , each DMi transmits the minimum bandwidth values to L2RM bandwidth allocator (see Fig. 2), which acts as indicated in the next section.

#### IV. BANDWIDTH ALLOCATION AT L2RM AND CAC

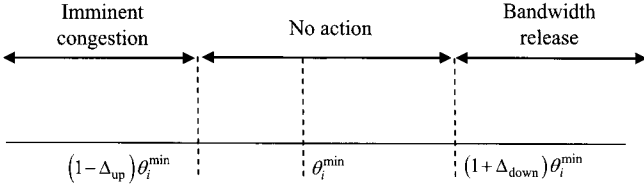
##### A. L2RM Bandwidth Allocator

The real bandwidth  $\theta_i$ , allocated to queue  $i$ , is changed over time by following indications coming from DMi and depending on the value of  $\theta_i^{\min}$ . This paper proposes a possible action but other solutions may be applied. L2RM, on the basis of the value of  $\theta_i^{\min}$ , assumes three possible states for queue  $i$ .

- “No action required”,
  - “Imminent congestion”,
  - “Bandwidth release”.
- Ideally, the 3 states can be defined by using only the value of  $\theta_i^{\min}$ .
- If  $\theta_i < \theta_i^{\min} \Rightarrow$  not enough allocated bandwidth–imminent congestion,
  - If  $\theta_i = \theta_i^{\min} \Rightarrow$  minimum allocated bandwidth–no action required,
  - If  $\theta_i > \theta_i^{\min} \Rightarrow$  too much allocated bandwidth–bandwidth release.

Operatively, the proposal is too hard and can generated unexpected bandwidth oscillations with consequent impact on real traffic. A smoother solution is preferred in practice: The mentioned states are defined by 2 thresholds  $(1 - \Delta_{up})$  and  $(1 + \Delta_{down})$  as follows

- If  $\theta_i < (1 - \Delta_{up})\theta_i^{\min} \Rightarrow$  not enough allocated bandwidth–imminent congestion,
- If  $(1 - \Delta_{up})\theta_i^{\min} \leq \theta_i \leq (1 + \Delta_{down})\theta_i^{\min} \Rightarrow$  minimum allocated bandwidth–no action required,


 Fig. 5. Queue  $i$  possible states.

- If  $\theta_i > (1 + \Delta_{\text{down}})\theta_i^{\min} \Rightarrow$  too much allocated bandwidth—bandwidth release.

Fig. 5 graphically shows the concepts introduced above. If the allocated bandwidth  $\theta_i$  is “close enough” to  $\theta_i^{\min}$ , no action is required; if  $\theta_i$  value is too much below  $\theta_i^{\min}$ , bandwidth is underprovisioned and congestion may happen in the next future; if  $\theta_i$  value is too much above  $\theta_i^{\min}$ , bandwidth is overprovisioned and may be released. Reasonable practical values of  $\Delta_{\text{up}}$  and  $\Delta_{\text{down}}$ , not necessarily equal, may be in the range  $[0.05, 0.3]$ .

### B. Call Admission Control (CAC)

The bandwidth update provided by the L2RM bandwidth allocator described above is limited by the maximum available bandwidth  $C$  at a specific RCST. This value is provided by the NCC, shown in Fig. 1. In this case, referring to the specific RCST in Fig. 2, which implements  $N_{L2}$  queues, there is the following constraint

$$\sum_{i=1}^{N_{L2}} \theta_i \leq C.$$

The maximum bandwidth value allocated by the NCC may be seen also divided for single queue within the RCST, if needed for control reasons. For example, referring again to the specific RCST in Fig. 2, a maximum bandwidth  $C_i$  might be forecast for queue  $i$ . The constraint is  $\sum_{i=1}^{N_{L2}} C_i = C$ . Obviously, in this case,  $\theta_i \leq C_i, \forall i = 1, \dots, N_{L2}$ . CAC is implemented within L2RM and must consider both the currently used bandwidth  $\theta_i$  and the maximum available bandwidth, either  $C$  or  $C_i$ , as explained above. The scheme used in this paper is very simple but effective if applied together with the bandwidth allocation scheme proposed in the previous subsection. Without making any per queue maximum bandwidth allocation, so referring to constraint  $\sum_{i=1}^{N_{L2}} \theta_i \leq C$ , an incoming connection with peak bandwidth  $p$  is accepted at queue  $i$  if  $\sum_{i=1}^{N_{L2}} \theta_i + p \leq C$ .

Alternatively, assigning a maximum bandwidth  $C_i$  at each queue  $i$ ,  $\forall i = 1, \dots, N_{L2}$ , so referring to the constraint set  $\theta_i \leq C_i, \forall i = 1, \dots, N_{L2}$ , a new connection, with peak bandwidth  $p$ , is accepted at queue  $i$  if  $\theta_i + p \leq C_i$ . Even if CAC is performed on the basis of the peak bandwidth, overprovisioning is only temporary because  $\theta_i$  is updated by L2RM as shown before.

### C. Time Variant Bandwidth Availability: Tackling Fading and Terminal Mobility

Satellite bandwidth availability may be time variant for different motivations. One of them is represented by fading. Obviously, bandwidth variability is not the only effect of fading

but this paper limits its scope to it. The concept may be also seen from another viewpoint. This paper models fading as a bandwidth reduction. In practice, different fading classes are defined, corresponding to combinations of channel bit and coding rate that give rise to redundancy factors  $\xi_{\text{level}}(t)$ , level = 1, 2,  $\dots$  ( $\xi_{\text{level}}(t) \geq 1.0$ ).  $\xi_{\text{level}}(t)$  represents the ratio between the information bit rate (IBR) in clear sky and the IBR in specific working conditions. The corresponding bandwidth reduction factor is defined as:  $\phi = 1/\xi_{\text{level}}(t)$ . The bandwidth reduction at the L2 queue, denoted by  $\theta_i^{\text{real}}(t)$ , can be computed as  $\theta_i^{\text{real}}(t) = \phi\theta_i(t), \forall i$ . Concerning terminal mobility, it is a recent research topic, of main interest for industry. One of the main problems is handover, which occurs when a mobile RCST changes several satellite beams over time. Research projects have been recently launched concerning mobility over satellites [20], [21]. During the movements, physical effects, such as doppler and multipath fading, may degrade the performance at the physical level [22]. Critical communication periods imply either signal blockages, which generate complete outage and no bandwidth availability, or due to shadowing. The countermeasures, for both the forward and return links, are: “Link layer-forward error correction (FEC)” (such as raptor codes), mainly to tackle shadowing, and proactive retransmission (PR), to tackle total outage. The former may be modeled through generic multiplicative bandwidth reduction factors, similarly as done for fading. The latter consists of “freezing” the transmitter, thus buffering the incoming data during outage periods. The approach proposed in this paper can be involved in this mechanism as follows. When an outage is foreseen (“smart mode” proposed by [23] suggests that a mobile RCST can estimate its position over time by using global positioning system (GPS) and so can try foreseeing outage periods) over a given period of time, the minimum bandwidth estimation algorithm, together with data transmission, should be “frozen” in order to avoid useless bandwidth computations during data buffering. After the outage period, the estimation of the minimum bandwidth should be reinitialized with the bandwidth value before the outage.

## V. ALGORITHMS TO COMPUTE THE MINIMUM BANDWIDTH $\theta_i^{\min}$

Different forms of the control law  $F(\cdot)$ , appearing in (1) and (2), may be reasonably applied in the context of the paper to get  $\theta_i^{\min}$ .

### A. Reference Chaser Bandwidth Controller (RCBC)

If the QoS of interest is the packet loss probability (PLP) or the delay one could use infinitesimal perturbation analysis to derive a gradient-based formulation of the control law  $F(\cdot)$  as follows [4]

$$\theta_i^{\min}(k+1) = \theta_i^{\min}(k) + \eta_k \left. \frac{\partial e_i(\cdot, k)}{\partial \theta_i} \right|_{\theta_i = \theta_i^{\min}(k)} \quad (3)$$

where  $\eta_k$  is the gradient step size; more specifically, for the PLP case

$$\left. \frac{\partial e_i(\cdot, k)}{\partial \theta_i} \right|_{\theta_i = \theta_i^{\min}(k)} = 2 \left. \frac{\partial \hat{l}_i(\theta_i)}{\partial \theta_i} \right|_{\theta_i = \theta_i^{\min}(k)} [\hat{l}_i^*(\cdot, k) - \hat{l}_i(\cdot, k)] \quad (4)$$

where

$$\left. \frac{\partial \hat{l}_i(\theta_i)}{\partial \theta_i} \right|_{\theta_i = \theta_i^{\min}(k)} = -\frac{1}{T_k} \sum_{bp=1}^{N_{T_k}} [at_{T_k}^{bp}(\theta_i^{\min}(k)) - ll_{T_k}^{bp}(\theta_i^{\min}(k))] \quad (5)$$

where  $\hat{l}_i(\cdot, k)$  is the measured loss rate of queue  $i$  over the  $OH_i(k)$ ,  $\hat{l}_i^*(\cdot, k)$  is the target loss rate coming from the required PLP value for queue  $i$  ( $PLP_i^*$ ):  $\hat{l}_i^*(\cdot, k) = \int_{OH_i(k)} PLP_i^* a_i(t) dt$ ,  $a_i(t)$  is the measured input rate of traffic class  $i$  over the  $OH_i(k)$ , and  $T_k$  is the size of  $OH_i(k)$ . A busy period ( $bp$ ), in (5), is a period of time in which the buffer is not empty. The quantity in brackets in (5) is the difference between the last loss during the busy period  $bp$  and the starting time of  $bp$ . A comparable gradient-based formulation can be obtained for the control of the delay performance (details can be found in [24], and are not reported here for the sake of synthesis). This technique is indicated here as reference chaser bandwidth controller (RCBC) for loss control.

### B. PID Control

Other, more traditional, approaches for the control law  $F(\cdot)$  are possible. For example, proportional integrative derivative (PID) control laws may be applied

$$\theta_i^{\min}(k+1) = \theta_i^{\min}(k) + w_{k+1}(k+1)e_i(\cdot, k+1) + w_k(k)e_i(\cdot, k) + w_{k-1}(k-1)e_i(\cdot, k-1) \quad (6)$$

where the weights are the tuning parameters used to optimize the PID temporal behavior in dependence of the specific application of interest. A huge amount of scientific literature exists on PID applications and parameters optimization, also for the bandwidth allocation case. The majority of industrial processes nowadays are still regulated by PID controllers. This reveals the rich potential of this simple control strategy for meeting various specifications for a vast variety of practical applications. The PID choice is mandatory for complicated metrics, such as jitter, for which no gradient formulations are available. For both RCBC and PID cases, it is well known that if the input rate processes of the buffer are ergodic and the quality constraints do not vary over time at least within convergence times, and other simple conditions are met (e.g., the decreasing behavior of the gradient steps size  $\eta_k$  for the RCBC case), the above control laws converge to the exact value of  $\theta_i^{\min}$ ; it means that the required QoS levels are satisfied with the minimum amount of bandwidth allocation (i.e.,  $\theta_i^{\min}$ ).

### C. Equivalent Bandwidth

Other control laws, not directly dependent on error  $e_i(\cdot, k)$ , are applicable. An example is

$$\theta_i^{\min}(k+1) = m_i(k) + \sqrt{-2 \ln(PLP_i^*) - \ln(2\pi)\sigma_i(k)}. \quad (7)$$

The equation above reports an equivalent bandwidth (EqB) technique applicable for the PLP case [12], [13];  $m_i(k)$  and  $\sigma_i(k)$  are the average and standard deviation, respectively, of the input rate process of queue  $i$  over  $OH_i(k)$  and  $PLP_i^*$  is the PLP requirement at queue  $i$ . The joint control of PLP and delay with EqB require some heuristic adaptations, such as the limitation of the maximum delay by properly setting a-priori the queue size to small values, see, e.g., [25].

### D. Other Approaches

Other approaches are possible for the choice of the control law  $F(\cdot)$  and to get  $\theta_i^{\min}$ ; for example, some neural or fuzzy techniques capable to support self-learning adaptation of the  $\theta_i^{\min}$  estimation. Anyway, every measurement-based algorithm dedicated to support precise estimation of  $\theta_i^{\min}$  can be applied within the framework reported in Fig. 4. Specific attention should be also devoted to the computational effort required for the computation. The effort of control laws mentioned above is low, particularly for the RCBC and PID approaches. Concerning EqB, the computational burden depends on the specific algorithm chosen to estimate mean and standard deviation parameters.

## VI. EXTENSION TO THE MULTICLASS CASE

The section is aimed at extending the control scheme presented in sections from II to V to the case where more than one flow (coming from layer 3) is conveyed within one layer 2 flow ( $f_i^{L2}$ , taking the flow as a reference) and, consequently, one layer 2 queue ( $i$ , following the same  $i$ th layer 2 reference). In practice,  $f_i^{L2}$  is composed by more than one “component” flow, called traffic class in the following where necessary to avoid confusion. An example may be two IP flows, e.g., a voice over IP (VoIP) and a video flow, which are multiplied over a single DVB flow entering a single DVB buffer. The DVB technology allows mapping different IP traffic flows over different DVB flows, thus maintaining the distinction among traffic flows even if conveyed to a single DVB buffer. The tools are either the packet identifier field (13 bit) or the asynchronous transfer mode (ATM) virtual channels identifiers, carried in the header of each DVB transport stream, in the forward or return link, respectively. Again, using the example above: VoIP and video flows are conveyed towards the same DVB buffer but they can still be differentiated. This distinction may be reflected mathematically by using an additional index  $j$ , which identifies a single traffic class, within a layer 2 flow. Fig. 6 contains the representation of the RCST layer 2 of Fig. 2 in the multiclass case. For each layer 2 flow (and corresponding buffer)  $i$ , the group of  $M_i$  traffic classes  $f_{i,1}^{L2}, \dots, f_{i,j}^{L2}, \dots, f_{i,M_i}^{L2}$  is multiplexed over the flow  $f_i^{L2}$ . Some quantities defined up to now must be redefined by considering also the index  $j$ . The quality of service threshold of each buffer  $i$  are redefined per traffic class:  $(QoS_{i,1}, \dots, QoS_{i,j}, \dots, QoS_{i,M_i})$ , as well as the quality of service measures in the observation horizon  $OH_i(k)$ ,  $k = 1, 2, \dots$ :  $(QoS_{i,1}^{OH}, \dots, QoS_{i,j}^{OH}, \dots, QoS_{i,M_i}^{OH})$  and the errors  $(e_{i,1}(\cdot, k), \dots, e_{i,j}(\cdot, k), \dots, e_{i,M_i}(\cdot, k))$ . In consequence, (1) and its simplification (2), related to generic buffer

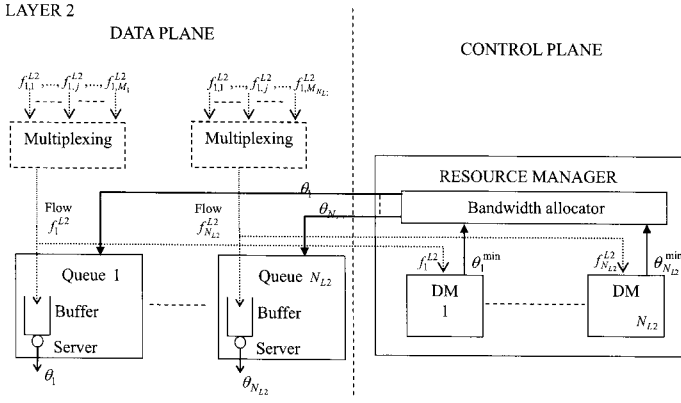


Fig. 6. RCST layer 2-multiclass case.

$i$  may be rewritten as

$$\theta_{i,j}^{\min}(k+1) = F(\theta_{i,j}^{\min}(k), \dots, \theta_{i,j}^{\min}(k-d), I_{i,j}(k), \dots, I_{i,j}(k-d)), \quad (8)$$

$$\theta_{i,j}^{\min}(k+1) = F(\theta_{i,j}^{\min}(k), e_{i,j}(\cdot, k)). \quad (9)$$

The estimated minimum bandwidth  $\theta_i^{\min}(k+1)$  related to buffer  $i$  should be allocated as

$$\theta_i^{\min}(k+1) = \max_j [\theta_{i,1}^{\min}(k+1), \dots, \theta_{i,j}^{\min}(k+1), \dots, \theta_{i,M_i}^{\min}(k+1)]. \quad (10)$$

Consequently, the DMi actions in Fig. 4 should be modified as shown in Fig. 7. The CAC is not modified by the introduction of traffic classes except for the definition of the peak rate, which may be defined per class. The algorithms to compute the estimated minimum bandwidth shown in the previous section are still applicable. From the formal viewpoint, it is sufficient adding the index  $j$  to the equations in Section V. The only substantial modification should be done to (5) where, in the multi-class case, the equality is no longer true. The equation must be rewritten as

$$\left. \frac{\partial \hat{l}_{i,j}(\theta_i)}{\partial \theta_{i,j}} \right|_{\theta_{i,j}=\theta_{i,j}^{\min}(k)} \cong -\frac{1}{T_k} \sum_{bp=1}^{N_{T_k}} [at_{T_k}^{bp}(\theta_{i,j}^{\min}(k)) - ll_{T_k}^{bp}(\theta_{i,j}^{\min}(k))] \quad (11)$$

and represents an approximation introduced in [14].

## VII. PERFORMANCE EVALUATION AND DISCUSSION

The aims of the performance evaluation are twofold. First, a performance analysis is made for the proposed bandwidth computation algorithms; in order to stress the working conditions, no CAC is applied and  $\theta_i = \theta_i^{\min}$  (i.e., the real queue is ideally an exact replication of the virtual one). Second, the proposed CAC is validated under the proposed bandwidth allocation scheme; a real fading trace is also considered in this case.

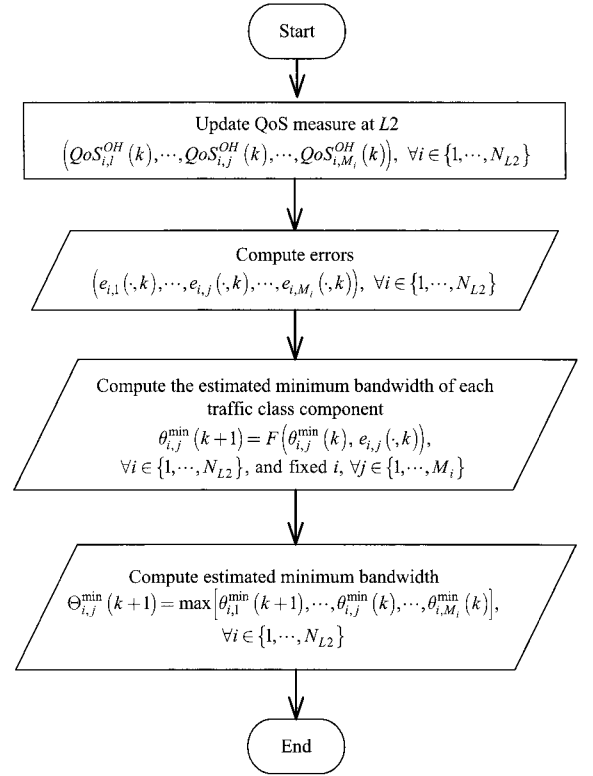


Fig. 7. Estimated minimum bandwidth computation by DMi-multiclass case.

### A. Bandwidth Control without CAC

The bandwidth control algorithms under investigation are: RCBC, PID, and EqB. An ideal allocation technique (Ideal) is also considered for PLP control, which exploits a perfect knowledge of future packet arrivals (though it is not realistic, it can be done easily via simulation). This knowledge offers support to a computation of the exact bandwidth to assure a given quality of service threshold. For the sake of synthesis, PLP metric is considered here, thus disregarding the delay metric. There is just one traffic: VoIP. According to international telecommunication union telecommunication standardization sector recommendation P series (ITU-T P).59, each source is an on-off process with exponentially distributed on and off time durations (mean 1.008 s and 1.587 s, respectively) and peak bandwidth of 16 kbps. VoIP traffic enters an IP buffer whose length and service rate (set by the traffic peak bandwidth) guarantee no packet loss rate. IP traffic is encapsulated into DVB frames, thus generating the process  $f_i^{L2}$ ,  $i$  representing VoIP, shown in Figs. 2 and 3;  $f_i^{L2}$  enters the DVB buffer (62 DVB cells), where the VoIP loss rate is measured every OH.  $PLP_i^* = PLP_{\text{VoIP}}^*$  is set to  $10^{-2}$ ; OH to 1 minute. The number of VoIP sources is increased of 10 from 70 to 110 each 2124 s. The PID weights are set here to 3.00, 1.50, and 1.25, respectively. These values guarantee the best PID performance in the following scenario and were found through accurate simulation inspection via brute force analysis. RCBC gradient stepsize is set to 1.0 (no optimization of the gradient stepsize is provided). RCBC gradient descent is initialized by the VoIP average bandwidth of 70 sources and by considering the introduced DVB overhead. Fig. 8 shows the resulting PLP at the

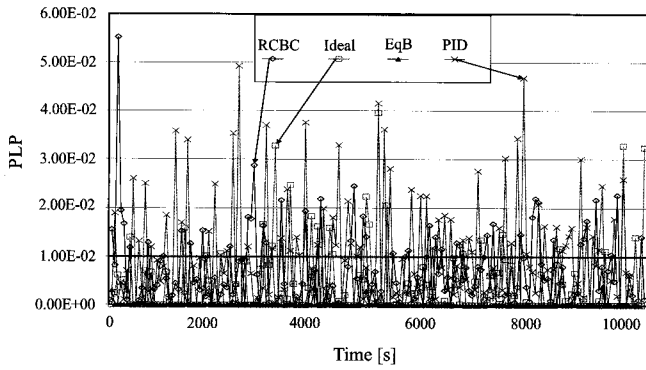


Fig. 8. VoIP scenario: PLP.

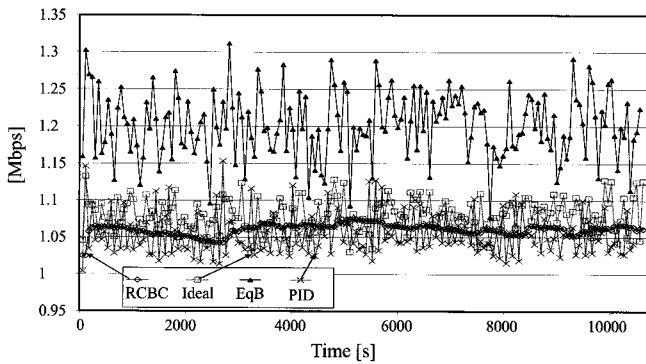


Fig. 9. VoIP scenario: Allocations.

end of each OH for all the techniques and Fig. 9 the corresponding bandwidth allocations. It is evident from Fig. 8 that all the techniques sometimes produce PLPs higher than the threshold, except for EqB which is always much below the threshold. Ideal assures the best performance. It must be noted that Ideal was optimized by accurate simulation inspection: It was necessary to reduce its OH to 57 s (in place of the original 1 minute) because the ideal computation with respect to future arrivals (registered over OHs of 1 minute each) underestimates the necessary bandwidth to meet the target. The rationale behind this behavior relies on the intrinsic burstiness of the sources, which no approximation of the ideal bandwidth (averaged over finite time periods) can match perfectly. The accuracy of the RCBC computation is clear from Fig. 9. After two reallocation steps, just at the beginning of the simulation, RCBC rate is smoothly changed over time with much higher precision in comparison with the considerable oscillations provided by ideal and PID. The simple observation of Figs. 8 and 9 suggests that RCBC reacts quickly to traffic changes also minimizing bandwidth oscillations. This has an impact on the overall performance over the entire simulation horizon. Quantitative metrics may help the interpretation of this qualitative behavior.

Table 1 represents the average and standard deviation of PLP and bandwidth over the simulation period. Table 1 also shows the percentage of the OH periods where PLP is over threshold ("OverTrh") and the average difference between measured PLP and target ("AverageDiffOTrh"), for all the techniques considered. Though PID exactly matches the target on average, it produces over-threshold PLPs for a relevant portion of time. It

Table 1. VoIP scenario: Average performance.

	RCBC	Ideal	EqB	PID
Average PLP	6.33E-03	3.73E-03	1.66E-04	1.01E-02
StDev PLP	7.56E-03	6.74E-03	1.27E-03	1.17E-02
OverTrh [%]	25.4	12.9	0.56	39.5
AverageDiffOTrh	1.73E-03	1.06E-03	3.14E-05	3.49E-03
Average BW [Mbps]	1.06	1.07	1.2	1.05
StDev BW [Mbps]	0.007	0.02	0.06	0.03

means that its bandwidth allocation is not sufficient to assure the target performance over time. EqB, on the other hand, overestimates the bandwidth need. The Ideal algorithm is almost perfect (only 12% of over-threshold PLPs). RCBC minimizes the bandwidth effort while assuring the average PLP closest to the target and minimizing bandwidth oscillations. According to these results, a practical application of the decision schemes presented in Sections III and IV can be ruled as follows. RCBC is used for computing  $\theta_i^{\min}$  (the RCBC convergence needs less than 5 OHs, see Fig. 9). At beginning,  $\theta_i$  is overprovisioned and later set to  $\theta_i^{\min}$  after 5 OHs.  $\Delta_{\text{up}}$  and  $\Delta_{\text{down}}$  in subsection IV.A are both set to 0.1 because the RCBC allocation variance is very low (0.007, see Table 1).

### B. Integration with CAC

The application of CAC is now considered together with bandwidth allocation. A channel fading condition is also considered. There is just one buffer at layer 2. The maximum available rate of the L2 queue, introduced in the CAC subsection of Section IV is  $C = 2.0$  Mbps; its buffer size amounts of 150 DVB cells. All the results given in this section are based on averages of 10 independent simulations, each of them of 8500 s. The fading trace is applied after the first 500 s period. Together with the VoIP traffic used in the previous subsection, video sources are also used. Both traffic classes are conveyed towards the single layer 2 buffer, so the multiclass case introduced in Section VI must be considered. Taking [25] as a reference, each video source is modeled as an exponentially modulated on-off process, with mean rate 0.25 Mbps and peak rate 0.7 Mbps and an average burst of 10.0 s. The active time of VoIP and video sources are log-normally distributed with an average of 200 s. The activation processes are Poisson arrival processes. Following [12], varying load conditions ranging from 0.5 to 5 are considered, where the value of 1.0 corresponds to a mean source activation rate equal to 240 sources/hour (4 calls per minute); an incoming call is randomly set to be VoIP or video with identical probabilities. The performance metrics under investigation are the achieved PLP for each traffic class and link utilization under a given load. Again, the delay metric is not considered for the sake of simplicity; the maximum achievable delay with a full L2 buffer (of 150 DVB cells with rate of 2.0 Mbps) is 112 ms. It is a rare event with a PLP lower than  $10^{-2}$ . The target PLPs are  $10^{-2}$  for VoIP and  $10^{-2}$  for video. The employed fading process, modeled through the  $\phi$  factor introduced in Section IV, has been taken from [26], where real attenuation samples are extracted from an experimental data set carried out in the Ka band on the Olympus satellite by the Centro Studi sulle Telecomunicazioni Spaziali (CSTS) Institute (Milan, Italy), on behalf of the Italian Space Agency. The carrier/noise

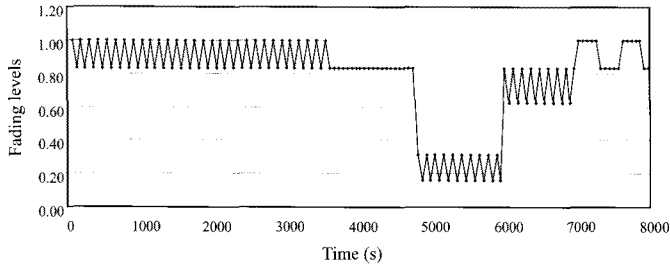


Fig. 10. Heterogeneous aggregation (VoIP and video) with fading:  $\phi(t)$  fading levels over time [26].

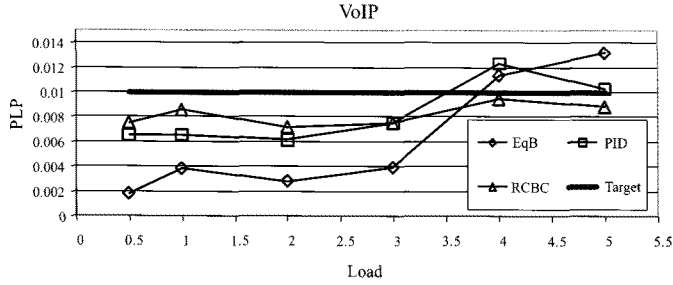


Fig. 11. Heterogeneous aggregation (VoIP and video) with fading: VoIP PLP (PLP target=0.01).

power ( $C/N_0$ ) factor is monitored at each station and, on the basis of its values, different bit and coding rates are applied to limit the bit error rate (BER) below a chosen threshold of  $10^{-7}$ . Six different fading classes are defined in this case corresponding to  $\phi(t) \in \{0.0, 0.15625, 0.3125, 0.625, 0.8333, 1.0\}$ . As said in Section IV, the bandwidth reduction at the L2 queue is computed as  $\theta_i^{\text{real}}(t) = \phi \theta_i(t), \forall i$ . As  $\theta_i^{\text{real}}(t)$  is just the bandwidth available for data traffic,  $\theta_i(t)$  has to be tuned over time in order to maintain the required QoS. As RCBC and PID set automatically the bandwidth under the measured losses, they do not need to know the exact value of  $\phi(t)$ . On the other hand, EqB follows the bandwidth computation rule in (7), not distinguishing between the two traffic classes (in short, not using the extension to the multiclass case to simplify the measure of  $m_i(k)$  and  $\sigma_i(k)$ ), but including the bandwidth reduction factor  $\phi(t)$ . The applied minimum bandwidth estimation is therefore

$$\theta_i^{\min}(k+1) = \left( \frac{1}{\phi(k+1)} \right) m_i(k) + \sqrt{-2 \ln(PLP_i^*) - \ln(2\pi) \sigma_i(k)}. \quad (12)$$

Having just one DVB queue, the index  $i$  would not be necessary here, but it is maintained for coherence with (7).  $PLP_i^*$  is the most restrictive requirement between the two involved traffic classes. The  $1/\phi(k+1)$  quantity corrects the bandwidth computation in dependence of the current information rate dedicated to protection codes. EqB, used without multiclass extension, is not able to tune the bandwidth in proportion to the actual presence of ongoing VoIP or video calls. RCBC and PID, on the other hand, automatically distinguish the bandwidth allocation between VoIP (requiring  $10^{-2}$  of PLP) and video (requiring  $10^{-3}$  of PLP). Fig. 10 shows the value of  $\phi(t)$  over time. Figs. 11 and 12 show the achieved PLP for each service under

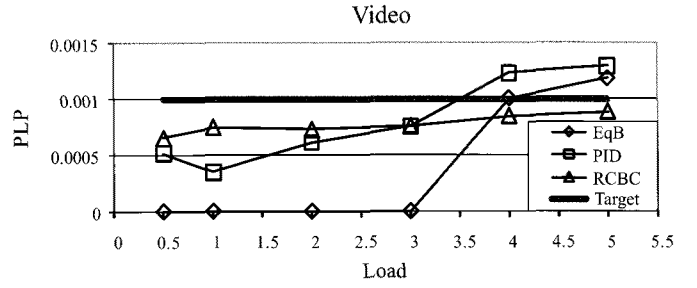


Fig. 12. Heterogeneous aggregation (VoIP and video) with fading: Video PLP (PLP target=0.001).

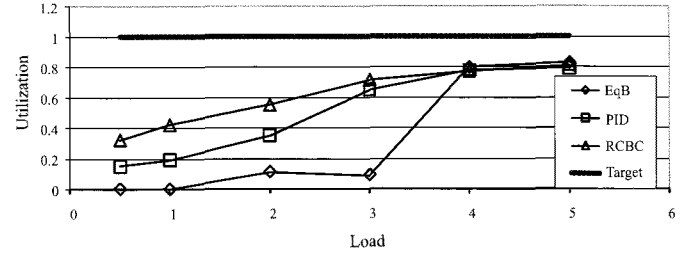


Fig. 13. Heterogeneous aggregation (VoIP and video) with fading: Utilization of the available bandwidth.

different traffic loads; Fig. 13 deals with the overall utilization of the available bandwidth. The relation between CAC and bandwidth allocation is clear. RCBC reveals to be more precise in tuning the bandwidth; the PLP curves are almost flat and safely stay a little below the corresponding targets. This has clearly an impact on CAC, thus obtaining the highest levels of utilization for all the considered traffic. The PLP performance of PID and EqB is good, but fails to stay below the targets with the highest loads. This can be solved by setting the  $\Delta_{\text{up}}$  value to about 0.17 (value validated by other simulation results, not reported here), but this also affects the utilization. The maximum achievable utilization is around 0.8 (surprisingly, obtained by all the techniques) under a load of 5. It is finally worth noting that the EqB has been applied without any optimization of the OH size. Anyway, even if the performance of EqB can be improved by using the concept of “dominant time scale” [12] and by applying the multiclass case so taking distinguished measures of  $m_{i,j}(k)$  and  $\sigma_{i,j}(k)$  depending on traffic class  $j$ , EqB seems not so efficient in the bandwidth allocation/CAC context presented in the paper.

## VIII. CONCLUSION AND FUTURE WORK

The paper proposes a new approach for bandwidth allocation and CAC over the DVB return channel of a satellite system. It is suitable for the optimization of the bandwidth under QoS constraints. The problem is addressed by outlining the architectures and the protocols of the entities involved in the decision process. In more detail, the RCST control plane proposed in this paper acts at layer 2 and includes a resource manager structured into DMs. DMs implement virtual queues, which are mirrors of real traffic buffers and represent the core of the control scheme presented in this paper. DMs compute the minimum bandwidth to support a given quality and communicate this bandwidth value to the L2RM, which allocates bandwidth to traffic buffers at the

data plane. Bandwidth control is joined to a CAC action so composing an overall control scheme within RCSTs. The results reported in the paper, validated under different conditions including a real fading trace, show promising performance and open the door to future evolutions such as the delay and jitter control and the implementation of the control scheme within a real network node [27].

## REFERENCES

- [1] "Digital video broadcasting (DVB); interaction channel for satellite distribution systems; guidelines for the use of EN 301 790," ETSI TR 101 790, V1.2.1, Tech. Rep., 2003.
- [2] "Digital video broadcasting (DVB); interaction channel for satellite distribution systems," ETSI EN 301 790, V1.5.1, 2009.
- [3] "Satellite earth stations and systems (SES), broadband satellite multimedia (BSM); broadband satellite multimedia services and architectures: QoS functional architecture," ETSI TS 102 462, V0.4.2, 2006.
- [4] M. Marchese and M. Mongelli, "On-line bandwidth control for quality of service mapping over satellite independent service access points," *Computer Netw.*, vol. 50, no. 12, Aug. 2006.
- [5] (Oct. 2007.). Applications layer QoS in DVB-RCS systems. Research Project. European Space Agency. [Online]. Available: <http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=27802>
- [6] European Satellite Communications Network of Excellence (SatNEX) I&II. [Online]. Available: <http://www.satnexus.org>
- [7] SatNEXIII. [Online]. Available: <http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=30263>
- [8] H. Skinnemoen, A. Vermesan, A. Luoras, G. Adams, and X. Lobao, "VoIP over DVB-RCS with QoS and bandwidth on demand," *IEEE Wireless Commun. Mag.*, vol. 12, pp. 46–53, Oct. 2005.
- [9] G. Giambene, *Adaptive Resource Management and Optimization in Satellite Networks*. Ed., Springer Verlag, Berlin, Dec. 2006.
- [10] S. Bracciali, R. Fantacci, T. Pecorella, L. Chisci, and M. A. V. Castro, "Proactive vs. reactive DVB-RCS terminal using ACM techniques," in *Proc. IEEE ICC*, Beijing, China, May 2008, pp. 1881–1885.
- [11] M. Luglio, C. Roseti, and F. Zampognaro, "Improving performance of TCP-based applications over DVB-RCS links," in *Proc. IEEE ICC*, Dresden, Germany, June 2009, pp. 1–6.
- [12] S. Georgoulas, P. Trimintzios, G. Pavlou, and K. Ho, "Measurement-based admission control for real-time traffic in IP differentiated services networks," in *Proc. IEEE ICT*, Capetown, South Africa, May 2005, pp. 1–5.
- [13] S. Georgoulas, P. Trimintzios, G. Pavlou, and K. Ho, "An integrated bandwidth allocation and admission control framework for the support of heterogeneous real-time traffic in class-based IP networks," *Computer Commun.*, vol. 31, no. 1, Jan. 2008.
- [14] M. Marchese and M. Mongelli, "Measurement-based computation of generalized equivalent bandwidth for loss constraints," *IEEE Commun. Lett.*, vol. 11, no. 12, Dec. 2007.
- [15] M. Grossglauser and D. Tse, "A framework for robust measurement-based admission control," *IEEE/ACM Trans. Netw.*, vol. 7, no. 3, June 1999.
- [16] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 7, Sept. 1991.
- [17] J. F. Whitehead, "Global packet dynamic resource allocation for wireless networks," US Patent 6295285, Sept. 2001.
- [18] M. C. Chuah, "Method for admitting new connections based on measured quantities in a multiple access system for communications networks," US Patent 6377548, Apr. 2002.
- [19] "Satellite earth stations and systems (SES), broadband satellite multimedia (BSM); IP over satellite," ETSI TR 101 985, V0.2.0, Tech. Rep. 2002.
- [20] ESA Project. Feasibility study of a mobile Ku-band terminal. Contract No. 15593/01/NL/DS. Final Report. [Online]. Available: [http://telecom.esa.int/feasibility\\_ndsatcom](http://telecom.esa.int/feasibility_ndsatcom)
- [21] H. Skinnemoen and P. T. Thompson, "Overview of DVB-RCS+M and its development," *Int. J. Sat. Commun. Netw.*, vol. 28, pp. 119–135, 2010.
- [22] "Digital video broadcasting (DVB); interaction channel for satellite distribution systems; guidelines for the use of EN 301 790 in mobile scenarios," ETSI TR 102 768, V1.1.1, Tech. Rep. 2009.
- [23] N. Celandroni, F. Davoli, E. Ferro, and A. Gotta. (2009). Video streaming transfer in a smart satellite mobile environments. *Int. J. Dig. Multimed. Broadcast*. [Online]. Available: <http://www.hindawi.com/journals/ijdmbl/2009/369216.htm>
- [24] F. Davoli, M. Marchese, and M. Mongelli, "Bandwidth adaptation for vertical QoS mapping in protocol stacks for wireless links," in *Proc. IEEE GLOBECOM*, Honolulu, Hawaii, 2009, pp. 1–6.
- [25] 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, Feb. 2003.
- [26] N. Celandroni, F. Davoli, and E. Ferro, "Static and dynamic resource allocation in a multiservice satellite network with fading," *Int. J. Sat. Commun. Netw.*, vol. 21, no. 4–5, July–Oct. 2003.
- [27] Emulator for an ETSI BSM-Compliant SI-SAP Interface. Research Project-ARTES5 Workplan 2007. European Space Agency. June 2009. [Online]. Available: <http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=29973>



**Mario Marchese** (S'94-M'97-SM'04) was born in Genoa, Italy in 1967. He got his "Laurea" degree cum laude at the University of Genoa, Italy in 1992 and the Qualification as Professional Engineer in April 1992. He obtained his Ph.D. (Italian Dottorato di Ricerca) degree in Telecommunications at the University of Genoa in 1996. From 1999 to 2004, he worked with the Italian Consortium of Telecommunications (CNIT), by the University of Genoa Research Unit, where he was Head of Research. From February 2005 he has been Associate Professor at the University of Genoa, Department of Communication, Computer and Systems Science (DIST). He is the founder and still the technical responsible of CNIT/DIST Satellite Communications and Networking Laboratory (SCNL) by the University of Genoa. He chaired the IEEE Satellite and Space Communications Technical Committee from 2006 to 2008. He is author and co-author of about 200 scientific works, including international magazines, international conferences, and book chapters and of the book "Quality of Service over Heterogeneous Networks", John Wiley & Sons, Chichester, 2007. His main research activity concerns satellite and radio networks, transport layer over satellite and wireless networks, quality of service and data transport over heterogeneous networks, emulation and simulation of telecommunication networks, and satellite components.



**Maurizio Mongelli** got his "Laurea" degree cum laude at the University of Genoa, Italy in 2000 and the Qualification as Professional Engineer in April 2002. He obtained his Ph.D. (Italian Dottorato di Ricerca) degree in Electronic and Computer Engineering at the University of Genoa in 2004. His Ph.D. was funded by Selex Communications S.p.A (Selex). He worked for both Selex and the Italian Consortium of Telecommunications (CNIT), by the University of Genoa Research Unit from 2000 to 2009. He is now a Member of the research staff of the Telecommunication Networking Research Group by the University of Genoa, with a CNIT research position. His main research activity concerns QoS architectures, resource allocation, and optimization algorithms for telecommunication systems.