Bandwidth Adaptation for Vertical QoS Mapping in Protocol Stacks for Wireless Links

F. Davoli, M. Marchese, M. Mongelli

DIST - Department of Communications, Computer and Systems Science
University of Genoa, Via Opera Pia 13, 16145, Genova, Italy
{ franco, dama, mopa }@dist.unige.it

Abstract—**Telecommunications networks are composed of functional layers acting in cascade. Quality of Service (QoS) derives from the action of each layer that must assure a specific level of quality to the upper layer in terms of performance parameters. The action is called vertical QoS mapping and is provided through algorithms that compute the bandwidth necessary so to assure the requested QoS when information is transferred from one layer to the next one below. This paper proposes a scheme that adapts the bandwidth to be allocated to a buffer which conveys heterogeneous traffic (both concerning traffic sources and QoS requirements) in a layer-in-cascade model. The proposed algorithm is based only on measures and does not use closed-form expressions, a-priori information about traffic statistical properties, and assumptions about buffer dimension. It is called WI-RCBC (Wireless Interface – Reference Chaser Bandwidth Control).**

Keywords—QoS mapping, bandwidth control

I. INTRODUCTION

MODERN telecommunication networks are composed of devices which act through layered protocol stacks. If a specific Quality of Service (QoS) is required, the interface among the layers must be able to transport the request and possibly the answer so to create a dialogue between the two layers. The overall QoS result depends on the QoS achieved at each layer of the network and it is based on the services offered at the layer interfaces. The vertical interaction between layers in cascade is defined as "Vertical QoS Mapping" [1]. The paper takes the TI-SAP (Technology Independent – Service Access Point) approach as reference [1, 2]. The original protocol architecture has been proposed by ETSI [3] for the access points to a Broadband Satellite Multimedia (BSM) network portion and specified in [4], [5], and [6]. The idea proposed in [1] and developed in [2] is to extend the concept of functional independence between physical interfaces and upper layers through the separation of Technology Dependent (TD) and Technology Independent (TI) layers and the definition of a generic interface called TI-SAP (Technology Independent – Service Access Point). The aims of the TI-SAP are: 1) to get a formal separation between the functional layers that use specific hardware/software solutions, defined as TD layers, and often covered by patents and the layers that implement upper layers, such as IP, defined as TI layers; 2) to establish a common interface through which TI and TD layers can communicate without affecting the specific TD layers implementation. In this generalized framework, it is important to get a model to describe the action of each layer. The proposal of this paper is to model each layer as a series of buffers so that the communication between adjacent functional layers may be described through a cascade of groups of queues. The queue model allows to describe the problems of vertical QoS mapping and to formally introduce a bandwidth allocation adaptation scheme called WI-RCBC (Wireless Interface – Reference Chaser Bandwidth Control).

The remainder of the paper is organized as follows. Section II sets the application framework for the algorithm and formalizes vertical QoS mapping as a cascade of buffers. Section III describes a reference framework for dynamic bandwidth adaptation for layers in cascade. Section IV introduces the WI-RCBC bandwidth adaptation scheme and Section V outlines some possible alternatives to the algorithm proposed. Section VI shows the simulation results. Section VII contains conclusions and some ideas for future work.

II. CASCADE-OF-QUEUES MODEL FOR ADJACENT LAYERS

The idea is to model each layer through blocks of queues, similarly as done in [4]. The number of queues must be large enough to support the desired QoS model. In this framework, there are three problems arising from the action of layers in cascade [1]. 1) Change of information unit, which implies additional informational (overhead) and bandwidth update, when information passes from upper to lower layer. 2) Aggregation of heterogeneous traffic: as outlined in [7], typically the number of queues decreases from the upper (TI) to the lower (TD) layers for efficiency and speed needs. It means that traffic may need to be aggregated when it flows down from a layer to the adjacent one. The bandwidth at lower (TD) layer must be adapted consequently. 3) Fading affect: many transmission environments, such as satellite and wireless links, need to tackle time varying channel conditions due to fading. The three problems presented above can be seen jointly. The overall cascade-of-queues model is shown in Figure 1 through N buffers at upper layer (TI in figure 1) identified through the index $i=1,...,N$ and one at lower (TD) layer to simplify the presentation. The bandwidth assigned to each buffer so to provide a given quality of service to the flows entering the buffer is identified as $R_{id}^{TI}(t)$ at TI layer (where $id=h,i,j$ identifies the buffer, referencing to Fig. 1

and as R^{TD} at TD layer. From the mathematical viewpoint, the fading effect may be modelled as a reduction of the bandwidth actually "seen" by the TD buffer. The reduction is represented by a stochastic process $\phi(t)$. At time t, the "real" service rate $R^{TD}_{real}(t)$ (available for data transfer) is $R^{TD}_{real}(t)=R^{TD}(t)\cdot\phi(t),\ \phi(t)\in[0,1]$, where time dependency is explicitly indicated to enforce the concept of time varying channel conditions. There are N traffic classes. One for each buffer. $a_i(t)$ is the *input rate* process of the i-th traffic class and $a(t)$ the aggregate process of all $a_i(t),\ i=1,...,N$. Bandwidth measure unit is [packet/s]. The key point is bandwidth adaptation, which is very challenging both from theoretical and practical viewpoint. Referencing to Fig. 1, it means to dimension the bandwidth $R^{TD}(t)$ at TD layer so that the service is transparently guaranteed to the upper layer.

Bandwidth allocation is a widely treated subject in the literature. Most of the schemes are based on the concept of equivalent bandwidth (EqB), which is defined as the minimum service rate to be provided to a traffic buffer to guarantee a certain degree of QoS in terms of objective parameters (e.g. packet loss, delay, jitter). EqB techniques are usually obtained analytically for homogeneous traffic trunks, with respect to a single QoS constraint, and are heavily based on the knowledge of traffic features, which are mathematically modelled. The complexity of the overall input flow process $a(t)$ entering the TD layer in the vertical QoS mapping model described above makes hardly applicable the bandwidth allocation algorithms that use mathematical models of the flow process. The flow accessing the TD queue comes from the actions of format change, traffic aggregation and fading affect, which modify the original features of the flows $a_i(t)$ that enter the TI layer. The resulting flow is so complex that can hardly be analytically modelled.

III. DYNAMIC QOS MAPPING FOR LAYERS IN CASCADE

The basic idea is to allocate the bandwidth periodically at TD layer after receiving the QoS constraints through TI-SAP primitives. It may be generically applied to any decisional scheme. Time variable t_k identifies the reallocation instants. Index k is a progressive integer.

The bandwidth $R^{TD}(t_k)$ allocated at the instant t_k may depend on the bandwidth allocated at previous instants $\{t_{k-1},t_{k-2},t_{k-3},...,t_{k-D}\}$, where D is the depth of the allocation scheme memory, and on an information vector $I=\{i(t_k),i(t_{k-1}),i(t_{k-2}),i(t_{k-3}),...,i(t_{k-D})\}$. The latter may be composed of information about the TD buffer and/or, simply by the error $e(t_k)$, which is defined as the difference between the minimum bandwidth that guarantees the QoS constraints in the interval $T_k=[t_{k-1},t_k]$, which is known at t_k, and the bandwidth allocated at t_{k-1}, which has given origin to the performance in the interval T_k.

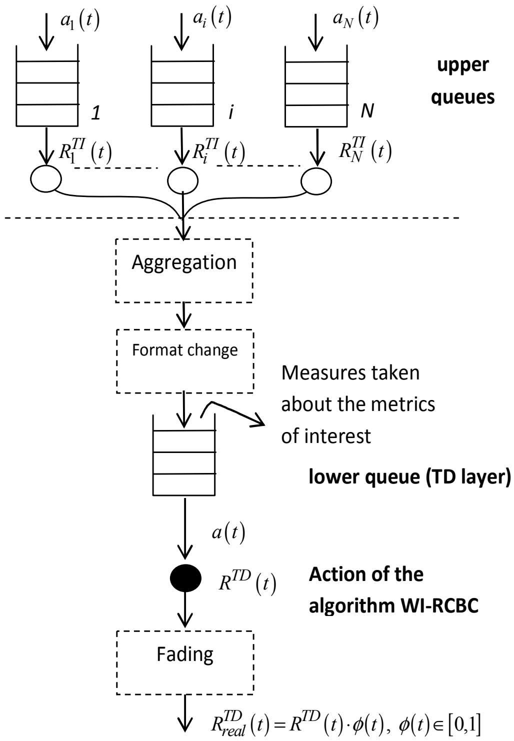

Figure 1. Cascade of queues model for adjacent layer.

More formally, if $R^{TD}_{thr}(t_k)$ is the bandwidth, computed at t_k, which would have been needed to assure the QoS constraints in T_k, the error in t_k is defined as $e(t_k)=R^{TD}_{thr}(t_k)-R^{TD}(t_{k-1})$. The multiplicative fading parameter $\phi(t)$ shown in the previous section is not included here only to simplify the notation, but all the comments are still valid including fading, as done in the results in the following.

Being $F(\cdot)$ a generic function, a possible generic representation of the allocated bandwidth is:

$$R^{TD}(t_k)=F\left(R^{TD}(t_{k-1}),R^{TD}(t_{k-2}),...,R^{TD}(t_{k-D}),i(t_k),i(t_{k-1}),i(t_{k-2}),...,i(t_{k-D})\right) \quad (1)$$

Where, as said above, $i(t_k)$ may be simply $e(t_k),\forall k$.

An interesting sub-class of bandwidth allocations algorithms may be described through the allocation in (2).

$$R^{TD}(t_k)=F\left(R^{TD}(t_{k-1}),i(t_k),i(t_{k-1}),i(t_{k-2})\right) \quad (2)$$

It includes the bandwidth allocations based on conventional discrete PID controller, which may be generically written as [8]:

$$R^{TD}(t_k) = R^{TD}(t_{k-1}) + w_k(t_k)\cdot e(t_k) + w_{k-1}(t_k)\cdot e(t_{k-1}) + w_{k-2}(t_k)\cdot e(t_{k-2}) \quad (3)$$

The details of the weights $w_k(t_k), w_{k-1}(t_k), w_{k-2}(t_k)$ and their possible computation may be found in [8] and other references related to discrete PID. To deal with nonlinear time-varying processes, also the weights may be time-varying and dependent on the information vector I. A more restricted algorithm sub-class is represented by the schemes where $D=1$:

$$R^{TD}(t_k) = F\left(R^{TD}(t_{k-1}), i(t_k)\right) \quad (4)$$

A corresponding bandwidth allocation update is reported in (5).

$$R^{TD}(t_k) = R^{TD}(t_{k-1}) + w_k(t_k)\cdot e(t_k) \quad (5)$$

If the requirement is that the bandwidth allocation algorithm does not use any a-priori information about traffic statistical properties, any assumption about buffer dimensions, and any closed-form expression of the involved variables, a possible solution is to use only measures of the ongoing processes. The weight $w_k(t_k)$ acts either as a reducer or as an amplifier of the bandwidth need estimation and may be dynamic over time.

IV. WI-RCBC

The aim here is to dynamically dimension the weight $w_k(t_k)$ every t_k so to chase the given performance thresholds. The reference environment has been described in section II and shown in Fig. 1. The quantities $R_i^{TI}(t)$, $R^{TD}(t)$, $a(t)$, $a_i(t)$, $i=1,...,N$ are defined in section II. So, there are N traffic classes. Traffic conveyed towards a single buffer is modelled through a *Stochastic Fluid Model* [9]. $a(t)$ is supposed ergodic for now, so that a single realization is representative of the entire process. This assumption will be relaxed later. There is no knowledge of $a_i(t)$ processes, as well as of the aggregate process $a(t)$. The only information about $a_i(t)$ and $a(t)$ may be got through real measures. The metrics used here are loss and delay. The following additional definitions are necessary. They are all applied at the TD buffer.

$R_i^{loss}\left(R^{TD}(t),t\right)$ is the loss rate process of the i-th traffic class in [packet/s]. $R_i^{delay}\left(R^{TD}(t),t\right)$ is the delayed packets rate process of the i-th traffic class, i.e. the rate of the packets which arrive with a delay over a given threshold d_{thr} [s]. $Loss_{i,thr}(t)$, which can also vary over time, is a probability that represents the performance threshold on the loss rate for class i; it is the performance reference received by the TI layer concerning losses.

$Delay_{i,thr}(t)$, which can also vary over time, is a probability that represents the performance threshold on the delay rate for class i; it is the performance reference received by the TI layer concerning delays. $R_{i,thr}^{loss}(t) = a_i(t)\cdot Loss_{i,thr}(t)$ is the loss rate process that can be tolerated (the loss rate threshold) of the i-th traffic class [packet/s]. $R_{i,thr}^{delay}(t) = a_i(t)\cdot Delay_{i,thr}(t)$ is the delayed packet rate process that can be tolerated (the delayed packet rate threshold) of the i-th traffic class [packet/s]

The average values of $R_i^{loss}\left(R^{TD}(t),t\right)$, $R_i^{delay}\left(R^{TD}(t),t\right)$, $R_{i,thr}^{loss}(t)$ and $R_{i,thr}^{delay}(t)$ are contained in (6) , (7), (8) and (9), respectively.

$$\bar{R}_i^{loss} = \lim_{\tau\to\infty}\frac{1}{\tau}\int_\tau R_i^{loss}(R^{TD}(t),t)\,dt;\ i=1,...,N \quad (6)$$

$$\bar{R}_i^{delay} = \lim_{\tau\to\infty}\frac{1}{\tau}\int_\tau R_i^{delay}(R^{TD}(t),t)\,dt;\ i=1,...,N \quad (7)$$

$$\bar{R}_{i,thr}^{loss} = \lim_{\tau\to\infty}\frac{1}{\tau}\int_\tau R_{i,thr}^{loss}(t)\,dt;\ i=1,...,N \quad (8)$$

$$\bar{R}_{i,thr}^{delay} = \lim_{\tau\to\infty}\frac{1}{\tau}\int_\tau R_{i,thr}^{delay}(t)\,dt;\ i=1,...,N \quad (9)$$

The aim is to provide the minimum TD buffer service bandwidth R_{opt}^{TD} so that the maximum quadratic distance between $\bar{R}_i^{loss}$ and $\bar{R}_{i,thr}^{loss}$ and between $\bar{R}_i^{delay}$ and is $\bar{R}_{i,thr}^{delay}$ minimized. It corresponds to define the following optimization problem.

$$R_{opt}^{TD,loss} = \arg\min_{R^{TD}} R_\Delta^{loss}\left(R^{TD}\right),\ R_\Delta^{loss}\left(R^{TD}\right) = \underset{i}{Max}\left[\bar{R}_i^{loss} - \bar{R}_{i,thr}^{loss}\right]^2 \quad (10)$$

$$R_{opt}^{TD,delay} = \arg\min_{R^{TD}} R_\Delta^{delay}\left(R^{TD}\right),\quad R_\Delta^{delay}\left(R^{TD}\right) = \underset{i}{Max}\left[\bar{R}_i^{delay} - \bar{R}_{i,thr}^{delay}\right]^2 \quad (11)$$

$$R_{opt}^{TD} = \max\left\{R_{opt}^{TD,loss}, R_{opt}^{TD,delay}\right\} \quad (12)$$

Being the involved stochastic processes unknown, problem (10)-(12) is solved by taking measures over the given $k-th$ *observation horizon* (OH), $T_k = [t_{k-1}, t_k]$, $k=1,2,..$, and performing a sequence of bandwidth reallocations, $R^{TD}(t_k)$, $k=1,2,...$, each T_k, as introduced in section V, based on the gradient method so that $R^{TD}(t_k)\xrightarrow{k\to\infty}R_{opt}^{TD}$. The quantities $\bar{R}_i^{loss}$, $\bar{R}_i^{delay}$, $\bar{R}_{i,thr}^{loss}$ and $\bar{R}_{i,thr}^{delay}$ are averaged over each OH, giving origin to the quantities from (13) to (16). Being used to solve the optimization problem (10)-(12),

$\hat{R}_i^{loss,k}$ and $\hat{R}_{i,thr}^{loss,k}$ must be representative of the average values $\bar{R}_i^{loss}$ and $\bar{R}_{i,thr}^{loss}$ $\forall\ i=1,...,N$ and $\forall k$, as well as $\hat{R}_i^{delay,k}$ and $\hat{R}_{i,thr}^{delay,k}$ must be representative of the average values $\bar{R}_i^{delay}$ and $\bar{R}_{i,thr}^{delay}$ $\forall\ i=1,...,N$ and $\forall k$

$$\hat{R}_i^{loss,k} = \frac{1}{T_k}\int_{T_k} R_i^{loss}(R^{TD}(t),t)\,dt;\ i=1,...,N \tag{13}$$

$$\hat{R}_i^{delay,k} = \frac{1}{T_k}\int_{T_k} R_i^{delay}(R^{TD}(t),t)\,dt;\ i=1,...,N \tag{14}$$

$$\hat{R}_{i,thr}^{loss,k} = \frac{1}{T_k}\int_{T_k} R_{i,thr}^{loss}(t)\,dt;\ i=1,...,N \tag{15}$$

$$\hat{R}_{i,thr}^{delay,k} = \frac{1}{T_k}\int_{T_k} R_{i,thr}^{delay}(t)\,dt;\ i=1,...,N \tag{16}$$

Bandwidth is adapted through the algorithm in Fig. 2, called Wireless Interface – Reference Chaser Bandwidth Control (WI-RCBC). It increases the bandwidth of the weighted needs sum in case there is at least one traffic class demanding bandwidth and decreases the bandwidth of the minimum weighted excess in case all classes show they have too much bandwidth. $step_k$ is the gradient stepsize. Modifications to WI-RCBC are possible by using the maximum bandwidth need and bandwidth excess as well as the sum of estimated bandwidth excesses or combinations of them but, on the one hand, the performance differences among them (measured through ad-hoc simulations not shown here) are not outstanding, on the other hand using the sum of bandwidth needs when adding and the minimum bandwidth excess when dropping is more conservative and safer than the alternatives.

Conditions $\left(\hat{R}_i^{loss,k}-\hat{R}_{i,thr}^{loss,k}\right)\geq 0$ and $\left(\hat{R}_i^{delay,k}-\hat{R}_{i,thr}^{delay,k}\right)\geq 0$ mean that the allocated bandwidth needs to be increased. Condition $\left(\hat{R}_i^{loss,k}-\hat{R}_{i,thr}^{loss,k}\right)<0$ and $\left(\hat{R}_i^{delay,k}-\hat{R}_{i,thr}^{delay,k}\right)<0$ state the opposite.

Derivatives $\frac{\partial\hat{R}_i^{loss,k}}{\partial R^{TD}}$ and $\frac{\partial\hat{R}_i^{delay,k}}{\partial R^{TD}}$, both negatives (this is the motivation for the negative sign before the quantities in Fig. 2), represent the sensitivity of loss and delay to infinitesimal variations of the rate serving the buffer. Intuitively they depend on the speed with which the system passes from an empty to a full state. They can be obtained by observing the buffer state evolution (as introduced in [9]) within each OH, which is divided into N_{T_k} busy periods (where the buffer is not empty) identified by the variable bp.

if $\left(\hat{R}_i^{loss,k}-\hat{R}_{i,thr}^{loss,k}\right)>0$ OR $\left(\hat{R}_i^{delay,k}-\hat{R}_{i,thr}^{delay,k}\right)>0$ for at least one i then{

$$\Delta_i^{loss}(t_k)=\begin{cases}-2\cdot\left.\frac{\partial\hat{R}_i^{loss,k}}{\partial R^{TD}}\right|_{R^{TD}=R^{TD}(t_{k-1})}\cdot\left[\hat{R}_i^{loss,k}-\hat{R}_{i,thr}^{loss,k}\right], & \text{if}\left[\hat{R}_i^{loss,k}-\hat{R}_{i,thr}^{loss,k}\right]\geq 0\\ 0, & \text{otherwise}\end{cases}$$

$$\Delta_i^{delay}(t_k)=\begin{cases}-2\cdot\left.\frac{\partial\hat{R}_i^{delay,k}}{\partial R^{TD}}\right|_{R^{TD}=R^{TD}(t_{k-1})}\cdot\left[\hat{R}_i^{delay,k}-\hat{R}_{i,thr}^{delay,k}\right], & \text{if}\left[\hat{R}_i^{delay,k}-\hat{R}_{i,thr}^{delay,k}\right]\geq 0\\ 0, & \text{otherwise}\end{cases}$$

$$R^{TD}(t_k)=R^{TD}(t_{k-1})+step_k\cdot\sum_{i=1}^{N}\left[\Delta_i^{loss}(t_k)+\Delta_i^{delay}(t_k)\right]$$

}

else { // $\left(\hat{R}_i^{loss,k}-\hat{R}_{i,thr}^{loss,k}\right)<0$ AND $\left(\hat{R}_i^{delay,k}-\hat{R}_{i,thr}^{delay,k}\right)<0\ \forall$i

$$\Delta_i^{loss}(t_k)=-2\cdot\left.\frac{\partial\hat{R}_i^{loss,k}}{\partial R^{TD}}\right|_{R^{TD}=R^{TD}(t_{k-1})}\cdot\left[\hat{R}_i^{loss,k}-\hat{R}_{i,thr}^{loss,k}\right]$$

$$\Delta_i^{delay}(t_k)=-2\cdot\left.\frac{\partial\hat{R}_i^{delay,k}}{\partial R^{TD}}\right|_{R^{TD}=R^{TD}(t_{k-1})}\cdot\left[\hat{R}_i^{delay,k}-\hat{R}_{i,thr}^{delay,k}\right]$$

$$Min\Delta^{-}(t_k)=\Delta_j^{\lambda}(t_k),\ j=\arg\min_i\left\{\left|\Delta_i^{\lambda}(t_k)\right|,\lambda=\{loss,delay\}\right\}$$

$$R^{TD}(t_k)=R^{TD}(t_{k-1})+step_k\cdot Min\Delta^{-}(t_k)$$

}

Figure 2. WI-RCBC.

$\frac{\partial\hat{R}_i^{loss,k}}{\partial R^{TD}}$ and $\frac{\partial\hat{R}_i^{delay,k}}{\partial R^{TD}}$ are approximated as in (17) and (18), respectively.

$$\frac{\partial\hat{R}_i^{loss,k}}{\partial R^{TD}}\cong-\frac{1}{T_k}\sum_{bp=1}^{N_{T_k}}\left[{}^{i}at_{T_k}^{bp}(R^{TD}(t_{k-1}))-{}^{i}ll_{T_k}^{bp}(R^{TD}(t_{k-1})\right] \tag{17}$$

$$\frac{\partial\hat{R}_i^{delay,k}}{\partial R^{RD}}\cong-\frac{1}{T_k}\sum_{bp=1}^{N_{T_k}}\left[{}^{i}at_{T_k}^{bp}(R^{TD}(t_{k-1}))-{}^{i}ld_{T_k}^{bp}(R^{TD}(t_{k-1})\right] \tag{18}$$

$\left[{}^{i}at_{T_k}^{bp}(R(t_{k-1}))-{}^{i}ll_{T_k}^{bp}(R(t_{k-1})\right]$ is the contribution to information loss of the $i-th$ traffic class for the busy period bp within T_k, $k=1,2,...$ ${}^{i}at_{T_k}^{bp}$ is the arrival time of the first packet of service class i within the busy period bp. ${}^{i}ll_{T_k}^{bp}$ is the time when the last loss of class i occurs during bp. If there is one traffic class, (17) is equality as proved in [10]. It is an approximation, introduced in [11], in case of aggregation. $\left[{}^{i}at_{T_k}^{bp}(R(t_{k-1}))-{}^{i}ld_{T_k}^{bp}(R(t_{k-1})\right]$ is the contribution to information delay of the $i-th$ traffic class for the busy period bp within T_k, $k=1,2,...$ ${}^{i}ld_{T_k}^{bp}$ is the time when the last delayed packet of class i arrives during bp. The idea is the same as for the loss. Approximation (18) is introduced in this paper. In practice bandwidth update WI-RCBC in Fig. 2 is in the form of equation (5). The derivatives multiplied by $step_k$ may be considered a form of the weight $w_k(t_k)$. The two differences $\left[\hat{R}_i^{loss,k} - \hat{R}_{i,thr}^{loss,k}\right]$ and $\left[\hat{R}_i^{delay,k} - \hat{R}_{i,thr}^{delay,k}\right]$ are the missing (or the excess of) bandwidth, i.e. a representation of the error $e(t_k)$.

V. OTHER TECHNIQUES

The following techniques are used for performance comparison with WI-RCBC. The aim is to highlight WI-RCBC control reliability with respect to other mechanisms taken from the literature and incapable to exactly tune bandwidth values in the presence of dynamic and heterogeneous conditions.

Proportional Integrative Derivative (PID) controller. The majority of industrial processes nowadays are still regulated by PID controllers. This does not just indicate the cautious attitude of the practicing engineer towards new techniques; it reveals the rich potential of this simple control strategy for meeting various specifications for a vast variety of practical applications. The PID equation applied is a slightly modified version of (3); the only difference relies on the error function $e(\cdot)$, which must be referred to the difference between the measured level of QoS and the threshold one. The weights in (3) depend on K_p, K_i, K_d (via basic algebra [8]), which are known as the proportional gain, the integral time constant and the derivative time constant. They are set 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.

Equivalent bandwidth (EqB). Due to the complexity of the *overall input rate process* $a(t)$, equivalent bandwidth approaches which use complex mathematical descriptors may be hardly applied in real time. The approach in [13] is applicable in this context for *Packet Loss Probability* (PLP) control. Being $m_a(t_k)$ and $\sigma_a(t_k)$ the measured *mean* and *standard deviation* of $a(t)$ over the OH, bandwidth is assigned at time t_k, $k = 1,2,...$ as in (23). PLP_{EqB}^* is the PLP upper bound and is defined as the most stringent PLP requirement out of N TI PLPs.

$$R(t_k) = m_a(t_k) + \sqrt{-2\ln(PLP_{EqB}^*) - \ln(2\pi)} \cdot \sigma_a(t_k) \quad (19)$$

Ideal Allocation for PLP (Ideal). An ideal allocation technique can be 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 perfect calculation of the exact bandwidth to assure a given PLP threshold.

VI. PERFORMANCE EVALUATION AND DISCUSSION

For the sake of synthesis, PLP metric is considered here, thus disregarding the delay metric, with respect to a single traffic class. VoIP traffic is taken as reference. According to 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 in DVB, thus generating the process $a(t)$ as output of the "Change of Format" box in Fig. 1. $a(t)$ enters the DVB buffer (62 DVB cells), where the VoIP loss rate in VoIP packets (of 80 bytes each) is measured every OH. PLP_{VoIP}^* is set to $1 \cdot 10^{-2}$; OH to 1 minute; WI-RCBC gradient stepsize to 1.0 (no optimization of the gradient stepsize is provided, for now). The number of VoIP sources is increased of 10 from 70 to 110 each 2124 s. WI-RCBC gradient descent is intialized by the VoIP average bandwidth of 70 sources, multiplied by the "cell tax" introduced by the DVB overhead. Fig. 3 shows the resulting PLP at the end of each OH for all the techniques and Fig. 4 the corresponding bandwidth allocations. From Fig. 3 it is evident that all the techniques periodically produce PLPs higher than the threshold, a part from the EqB which is always much below the threshold. The Ideal seems assuring the best performance. It must be noted that the Ideal was optimized by accurate simulation inspection: it was necessary to reduce its OH to 57s (in place of the original 1 minute) because the Ideal computation with respect to the 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. From Fig. 4 one significant achievement arises: the accuracy of the WI-RCBC computation. After two significant reallocation steps, just at the beginning of the simulation, WI-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. 2 and 3 suggests that WI-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 (and noted by: "Average_PLP", "StDev_PLP", "Average_Bw" and "StDev_Bw"), together with the percentage of the OH periods where PLP is over threshold ("OverTrh") and the average difference between measured PLP and the target ("AverageDiffOTrh"), for all the techniques considered. Though PID exactly matches the target on average, it produces over-threshold PLPs for the largest portion of time. It means that the corresponding average bandwidth (1.05 Mbps) is not sufficient to assure the target performance. EqB, on the other hand, overestimates the bandwidth need. The Ideal is almost perfect (only 12% of over-threshold PLPs), but only WI-RCBC minimizes the bandwidth effort while assuring the average PLP closest to the target and minimizing bandwidth oscillations. In other words, it means that the correct bandwidth need is exactly computed by WI-RCBC (1.06). Similar results can be obtained with respect to other working conditions (such as introducing other traffic categories, e.g., video streaming, or changing buffer dimension), and to the delay metric as well.

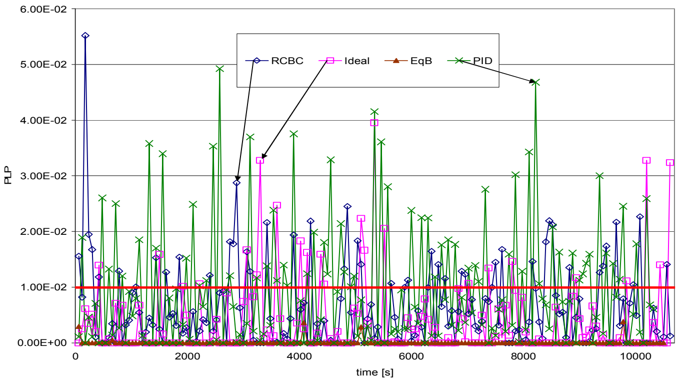

Figure 3. VoIP scenario: PLP.

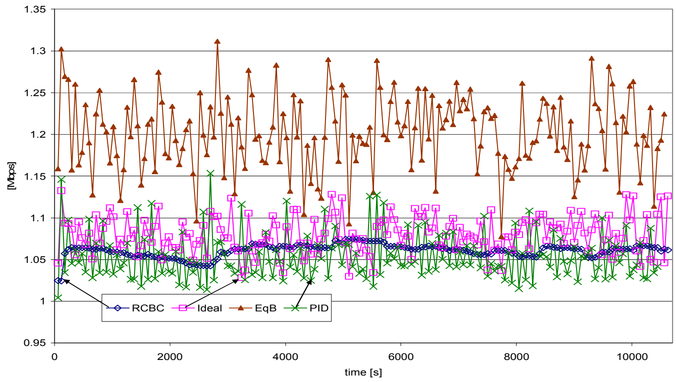

Figure 4. VoIP scenario: allocations.

VII. Conclusions and Future Work

The paper presents a novel control scheme that adapts the bandwidth to be allocated to a buffer which conveys heterogeneous traffic (both concerning traffic sources and QoS requirements) in a layer-in-cascade model. The proposed algorithm is based only on measures and does not use closed-form expressions, a-priori information about traffic statistical properties, and assumptions about buffer dimension. The reliability of the algorithm proposed, shown in the simulations, opens the door to future investigation involving the joint control of loss and delay together with other possible traffic categories.

References

[1] M. Marchese, M. Mongelli, "Vertical QoS Mapping over Wireless Interfaces", Wireless Communications Magazine, 2008, to appear.

[2] M. Marchese, "QoS over Heterogeneous Networks", John Wiley & Sons, Chichester, UK, 2007.

[3] ETSI. Satellite Earth Stations and Systems (SES). Broadband Satellite Multimedia, IP over Satellite. ETSI Technical Report, TR 101 985, V1.1.2, November 2002.

[4] ETSI. Satellite Earth Stations and Systems (SES), Broadband Satellite Multimedia (BSM) Services and Architectures, QoS Functional Architecture, TS 102 462 V0.4.2, 2006-01.

[5] ETSI. Satellite Earth Stations and Systems (SES), Broadband Satellite Multimedia (BSM) Services and Architectures, Interworking with RSVP-based QoS (IntServ), TS 102 463 V0.4.2, 2006-10.

[6] ETSI. Satellite Earth Stations and Systems (SES), Broadband Satellite Multimedia (BSM) Services and Architectures, Interworking with DiffServ QoS, TS 102 464 V0.4.1, 2006-09.

[7] ETSI. Satellite Earth Stations and Systems (SES). Broadband Satellite Multimedia. Services and Architectures; BSM Traffic Classes. ETSI Technical Specification, TS 102 295 V1.1.1, Feb. 2004.

[8] M. Xu, S. Li, C. Qi, W. Cai, "Auto-tuning of PID controller parameters with supervised receding horizon optimization", ISA transactions, vol. 44, n. 4, October 2005, pp. 491-500.

[9] C. G. Cassandras, G. Sun, C. G. Panayiotou, Y. Wardi, "Perturbation Analysis and Control of Two-Class Stochastic Fluid Models for Communication Networks," IEEE Trans. Automat. Contr., vol. 48, n. 5, May 2003, pp. 23-32.

[10] Y. Wardi, B. Melamed, C.G. Cassandras, C.G. Panayiotou, "Online IPA Gradient Estimators in Stochastic Continuous Fluid Models," J. of Optimization Theory and Applic., vol. 115, no. 2, pp. 369-405, Nov. 2002.

[11] M. Marchese, M. Mongelli, "Measurement-based Computation of Generalized Equivalent Bandwidth for Loss Constraints", IEEE Communications Letters, vol. 11, no. 12, December 2007.

[12] N. Celandroni, F. Davoli, E. Ferro, "Static and Dynamic Resource Allocation in a Multiservice Satellite Network with Fading," Internat. J. of Satellite Commun. and Networking, vol. 21, no. 4-5, July-Oct. 2003, pp. 469-487.

[13] R. Guérin, H. Ahmadi, M. Naghshineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks," IEEE J. Select. Areas Commun., vol. 9, no. 7, Sept. 1991, pp. 968–981.

	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**

Table 1. VoIP scenario: Average performance.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.