Channel Sharing by Multi-class Rate Adaptive Streams: Performance Region and Optimization.

N. Argiriou, L. Georgiadis
Aristotle Univ. of Thessaloniki
Faculty of Engineering
School. of Electrical and Computer Engineering
Telecommunications Dept.
Thessaloniki, 54124, GREECE
e-mails: narg@engatia.ee.auth.gr, leonid@auth.gr

Abstract

We consider the problem of channel sharing by rate adaptive streams belonging to various classes. The performance measure per class is the average scaled bandwidth allocated to connections in the class. We first provide a bandwidth adaptation policy that maximizes a linear combination of class performance measures; then, we use this result to characterize the region where the class performance measures lies under any bandwidth adaptation policy. Finally, based on the results above we use stochastic approximation techniques to provide a policy that optimizes a combination of concave rewards associated with class performance measures.

Keywords: Bandwidth Sharing, Rate-adaptive Streams, Wireless Channel Sharing, Performance Region, Stochastic Approximation.

1 Introduction

We consider a communication channel whose bandwidth is shared by randomly arriving connections belonging to a number of classes. Connection bandwidth may be adapted within a given range, hence giving the opportunity for bandwidth management in order to achieve good reception quality with low blocking probability. Connection bandwidth adaptation can be achieved by various coding techniques such as layered coding [1], [2] and adaptation of compression parameters [3], [4], [5]. Depending on the technique, rate adaptation can take one of a number of discrete values, or it can take any value within a specific range.

The channel may be wireless or wired; each of these channels has certain capacity characteristics, e.g., fixed capacity in wired versus variable in wireless, and may require different handling of incoming connections. For example, in a wireless cellular environment preferential treatment may be given to connections arriving to a cell during a hand-off so that already admitted connections at another cell are not rejected. This may be achieved by a properly designed connection admission control policy. The results we present in this work are applicable to either of these channels under reasonable assumptions on the connection admission policy.

In recent years, several works addressed the problem of bandwidth adaptation management under various assumptions on channel characteristics and the bandwidth adaptation policies [6], [7], [8], [9], [10], [11], [12], [13]. Our approach follows [10], [12], where for single class systems, bandwidth management policies optimizing the average connection scaled bandwidth were presented. For the single class system we show in Section 3 that the policy presented in [10], [12], is optimal under quite general conditions. We then consider multiclass systems (Section 4). We provide a policy for maximizing a linear combination of class performance measures in Section 4.1; using this, we characterize the performance region of the system in Section 4.2. These results permit us to apply the general framework developed in [14], [15], to provide a bandwidth adaptation policy that optimizes a combination of concave rewards associated with class performance measures (Section 5). Finally, in Section 6 we discuss some extensions of our results.

Due to lack of space proofs and several discussions are omitted. The interested reader can refer to an extended version of the paper at

http://users.auth.gr/˜leonid/public/het-netFull.pdf.

2 System Model and Notation

Throughout the paper, sets are denoted by calligraphic capital letters. The number of elements in a set $\mathcal{X}$ is denoted by the corresponding normal capital letter, i.e., $X = |\mathcal{X}|$. Vectors are denoted by boldface letters, e.g., $\mathbf{x} = \{x_i\}_{i=1}^N$, $\mathbf{x} \in \mathbb{R}^N$ ($\mathbb{R}$ is the set of real numbers). The Euclidean norm of a vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|$.

We consider a communication channel of bandwidth B bps. Extensions to the case where the bandwidth of the link may vary over time are discussed in Section 6. Connections arrive for transmission over the link. Let $\mathcal{A}(t)$ be the set of connections that arrived and have been accepted by the system up to time t. Let also $\mathcal{N}(t)$ be the set of connections that are present in the system at time t.

Connection i arrives at time a_i and departs at time $d_i > a_i$. The connection holding time $h_i = d_i - a_i > 0$, is assumed known (e.g., connections may correspond to transmission of prerecorded movies). Through appropriate compression techniques, transmission of connection i may take place at rates belonging to a set $\mathcal{B}_i$. We call these rates "bandwidth levels". We denote by $\underline{B}_i$ and $\overline{B}_i$ the minimum and maximum bandwidth levels in $\mathcal{B}_i$. Transmission rates of connection i may be adapted over time, but must belong to $\mathcal{B}_i$ for acceptable reception quality. Hence, if $b_i(t)$ is the bandwidth allocated to connection i at time t, $a_i \leq t < d_i$, it must hold for any t,

$$\sum_{i \in \mathcal{N}(t)} b_i(t) \leq B, \tag{1a}$$

$$b_i(t) \in \mathcal{B}_i, \ i \in \mathcal{N}(t). \tag{1b}$$

In order to operate the system, two policies must be defined: the "Connection Admission Policy" and the "Bandwidth Adaptation Policy". The Connection Admission Policy decides whether to accept or reject a newly arriving connection, while the Bandwidth Adaptation Policy adjusts at any time t the bandwidth of the connections that are currently in the system.

Regarding the Connection Admission Policy, we assume initially that an arriving connection is admitted in the system whenever this can be done without degrading any of the already admitted connections below their acceptable levels ($\underline{B}_i$). Formally, the Connection Admission Policy is the following.

Connection Admission Policy. A connection that arrives at time t is admitted in the system if it is possible to rearrange the bandwidths of the connections already in the system, so that constraints (1) are satisfied if $\mathcal{N}(t)$ includes the newly arrived connection.

This type of Connection Admission Policy have been proposed in [7], [10], [11], [12]. In Section 6 we discuss other variants for which the results presented in this work are applicable. As will be seen, for our purposes the important property of the Connection Admission Policy is that $\mathcal{N}(t)$ and hence $\mathcal{A}(t)$ are independent of the employed Bandwidth Adaptation Policy. This may be desirable since as a result parameters affecting connection blocking probability can be designed independently of the bandwidth adaptation mechanism. From a performance point of view, the Connection Admission Policy accepts a new connection as long as there is a feasible solution to (1), while the Bandwidth Adaptation Policy rearranges the connection bandwidths so that certain performance optimization criteria are satisfied.

Regarding the Bandwidth Adaptation Policy, its objective is to allocate bandwidth to connections in such a manner that certain connection performance measures are satisfied. There are various such measures that may be defined and the issue which of them or combination thereof is appropriate, is still an open research problem [16], [17], [18], [19]. Among the most relevant ones is the "scaled mean connection bandwidth" [12], on which we concentrate in this paper. The (scaled) mean bandwidth allocated to connection i when Bandwidth Adaptation Policy π is employed, is defined as

$$\widehat{b}_i^\pi = \frac{\int_{a_i}^{d_i} b_i^\pi(t)dt}{h_i \overline{B}_i}, \tag{2}$$

that is, $\widehat{b}_i^\pi$ represents the quality of the average bandwidth received by the connection relative to the best possible, $\overline{B}_i$. For our purposes, it will be convenient to extend this definition for all times $t \geq 0$, as follows.

$$\widehat{b}_i^\pi(t) = \begin{cases} 0 & t < a_i \\ \frac{1}{h_i \overline{B}_i} \int_{a_i}^{t} b_i^\pi(s)ds & a_i \leq t < d_i \\ \widehat{b}_i^\pi & t \geq d_i \end{cases}. \tag{3}$$

Let $V_i = h_i \overline{B}_i$ and set $b_i(t) = 0$ if either $t < a_i$, or $t \geq d_i$. We can then rewrite (3) in the following form which will be useful in the sequel.

$$\widehat{b}_i^\pi(t) = \frac{\int_0^t b_i^\pi(s)ds}{V_i}. \tag{4}$$

Since $\underline{B}_i \leq b_i^\pi(t) \leq \overline{B}_i$, it holds for $i \in \mathcal{A}(t)$,

$$0 \leq \widehat{b}_i^\pi(t) \leq 1. \tag{5}$$

The performance of the system at time t is defined as the average of the performance measures of all connections that have been admitted by the system up to time t, that is,

$$\widehat{B}^\pi(t) = \frac{\sum_{i \in \mathcal{A}(t)} \widehat{b}_i^\pi(t)}{A(t)}. \tag{6}$$

For a single-class system, we are interested in defining Bandwidth Adaptation Policies that maximize (6) either at any time t, or asymptotically as $t \to \infty$ (the precise meaning will be given in Section 4.1).

In this paper we will mainly consider multiclass systems where arriving connections belong to one of the classes in a set $\mathcal{C}$. In this case we define the sets $\mathcal{A}_c(t)$ and $\mathcal{N}_c(t)$ as well as the class performance measure $\widehat{B}_c^\pi(t)$, in a manner analogous to the definition of $\mathcal{A}(t)$, $\mathcal{N}(t)$ and $\widehat{B}^\pi(t)$ respectively. Hence we have, $\mathcal{A}(t) = \cup_{c\in\mathcal{C}}\mathcal{A}_c(t)$ and $\mathcal{N}(t) = \cup_{c\in\mathcal{C}}\mathcal{N}_c(t)$. We also denote by $\mathcal{C}(t)$ the set of classes for which $\mathcal{N}_c(t) \neq \varnothing$, i.e., for any class in $\mathcal{C}(t)$, say class c, there is a least one class c connection present in the system at time t. Notice that again, due to the Connection Admission Policy, $\mathcal{N}_c(t)$, $\mathcal{A}_c(t)$, and $\mathcal{C}(t)$ are independent of the employed Bandwidth Adaptation Policy. We will be interested in optimizing measures that are functions of class performance measures $\widehat{B}_c^\pi(t)$, $c \in \mathcal{C}$.

In the rest of the paper we will use the following conventions regarding summations, $\sum_{i\in\mathcal{X}} x_i = 0$, if $\mathcal{X} = \varnothing$, $\sum_{i=a}^{b} x_i = 0$, if $a > b$.

3 Single Class System

In [10], [12], a policy π with maximum $\widehat{B}^\pi \triangleq \lim_{t\to\infty} \widehat{B}^\pi(t)$ has been proposed for the single class system assuming Poisson arrivals and i.i.d connection holding times. As will be seen next, this policy optimizes $\widehat{B}^\pi(t)$ at any time t, under arbitrary arrival patterns and connection holding times.

The following identity, obtained by interchanging summation and integration, is important for the subsequent development.

$$\sum_{i\in\mathcal{A}(t)} \widehat{b}_i^\pi(t) = \sum_{i\in\mathcal{A}(t)} \frac{\int_0^t b_i^\pi(s)ds}{V_i} = \int_0^t \sum_{i\in\mathcal{A}(t)} \frac{b_i^\pi(s)}{V_i}ds. = \int_0^t \sum_{i\in\mathcal{N}(s)} \frac{b_i^\pi(s)}{V_i}ds. \tag{7}$$

The last equality follows from the fact that by definition $b_i^\pi(s) = 0$ if a connection is not present in the system at time t. Note that since $\mathcal{A}(t)$ is independent of the employed Bandwidth Adaptation Policy, in order to maximize $\widehat{B}^\pi(t)$ it suffices to maximize $\sum_{i\in\mathcal{A}(t)} \widehat{b}_i(t)$, or according to (7), to maximize, $\int_0^t \sum_{i\in\mathcal{N}(s)} \frac{b_i^\pi(s)}{V_i}ds$.

Observe next that a) the set $\mathcal{N}(t)$, is independent of the employed policy and b) constraints (1), which must be satisfied at any time s, depend only on the choice of connection bandwidths at time s, hence they neither affect, nor are affected by the choice of the bandwidths at other times. Hence in order to maximize $\int_0^t \sum_{i\in\mathcal{N}(s)} \frac{b_i^\pi(s)}{V_i}ds$, it suffices to maximize $\sum_{i\in\mathcal{N}(t)} \frac{b_i^\pi(s)}{V_i}$ at any time s. We summarize the discussion above in the following lemma.

Lemma 1 *Under arbitrary connection arrival pattern, connection holding times and connection bandwidth levels, the policy π^* that at any time $s \geq 0$ allocates to connection $i \in \mathcal{N}(s)$ bandwidth $b_i^{\pi^*}(s) = b_i^*$, where b_i^* is the solution to the following optimization problem,*

$$\max\left\{\sum_{i\in\mathcal{N}(s)} \frac{b_i}{V_i}\right\} \text{ where } \sum_{i\in\mathcal{N}(s)} b_i \leq B \text{ and } b_i \in \mathcal{B}_i,\ i \in \mathcal{N}(s),$$

maximizes $\widehat{B}^\pi(t)$ for all $t \geq 0$.

Note: The solution to the optimization problem of Theorem 1 depends on s only through the set $\mathcal{N}(s)$. So, as long as $\mathcal{N}(s)$ remains the same the bandwidth allocated to connections by π^* need not change. Therefore, the Bandwidth Adaptation Policy may be invoked only at connection arrival and departure instances.

4 Multiple Class System

In this section we deal with a multiclass system. We first consider the problem of optimizing a linear combination of class performance measures, under quite general conditions. Next, with additional statistical assumptions on the arrival rates and holding times, we derive the performance region of the system, i.e., the region in $\mathbb{R}^C$ where the vector of class performances takes values. Finally, we consider the case where concave (instead of linear) rewards are associated with each class and provide an optimal Bandwidth Adaptation Policy.

4.1 Linear Rewards

In this section we assume that connections belonging to class c are accepted by the system at a long-term rate λ_c. Specifically,

$$\lim_{t\to\infty} \frac{A(t)}{t} = \lambda_c,\ 0 < \lambda_c < \infty,\ c \in \mathcal{C}(t). \tag{8}$$

Since connections belong to multiple classes, it may be desirable to provide discriminatory service depending on the class to which a connection belongs. The simplest such class discriminatory service is to associate with class c a reward $r_c \geq 0$ per unit of received performance. Let $\mathbf{r} = \{r_c\}_{c\in\mathcal{C}}$. Then, the total system reward at time t under Bandwidth Adaptation Policy π becomes,

$$\widehat{B}^\pi_{\mathbf{r}}(t) = \sum_{c\in\mathcal{C}} r_c \widehat{B}^\pi_c(t) = \sum_{c\in\mathcal{C}} r_c \frac{\int_0^t \sum_{i\in\mathcal{N}_c(s)} \frac{b_i^\pi(s)}{V_i} ds}{A_c(t)} = \frac{1}{t}\int_0^t \sum_{c\in\mathcal{C}(s)} \sum_{i\in\mathcal{N}_c(s)} \frac{t r_c b_i^\pi(s)}{A_c(t) V_i} ds. \tag{9}$$

The next theorem provides a policy that optimizes the system performance measure $\widehat{B}^\pi_{\mathbf{r}}(t)$ asymptotically, as $t \to \infty$. Since we do not make any assumptions on the connection holding times, bandwidth levels and on the manner a Bandwidth Adaptation Policy operates, the limit $\lim_{t\to\infty} \widehat{B}^\pi_{\mathbf{r}}(t)$ may not exist. Therefore, regarding the asymptotic measures of performance, we consider either $\limsup_{t\to\infty} \widehat{B}^\pi_{\mathbf{r}}(t)$ or $\liminf_{t\to\infty} \widehat{B}^\pi_{\mathbf{r}}(t)$.

Theorem 2 *Let the connection acceptance rate be $\lambda_c,\ c \in \mathcal{C},\ 0 < \lambda_c < \infty$, and the connection holding times and bandwidth levels arbitrary. Let π^* be the Bandwidth Adaptation Policy that at any time t allocates to connection $i \in \mathcal{N}(t)$ bandwidth $b_i^{\pi^*}(t) = b_i^*$, where b_i^* is the solution to the following optimization problem.*

$$\max \left\{ \sum_{c\in\mathcal{C}(t)} \sum_{i\in\mathcal{N}_c(t)} \frac{t}{A_c(t)} \frac{r_c}{V_i} b_i \right\} \tag{10}$$

$$\sum_{c\in\mathcal{C}(t)} \sum_{i\in\mathcal{N}_c(t)} b_i \leq B \tag{11}$$

$$b_i \in \mathcal{B}_i,\ i \in \mathcal{N}_c(t),\ c \in \mathcal{C}(t). \tag{12}$$

Then, it holds for any Bandwidth Adaptation Policy π.

$$\limsup_{t\to\infty} \widehat{B}^\pi_{\mathbf{r}}(t) \leq \limsup_{t\to\infty} \widehat{B}^{\pi^*}_{\mathbf{r}}(t),$$
$$\liminf_{t\to\infty} \widehat{B}^\pi_{\mathbf{r}}(t) \leq \liminf_{t\to\infty} \widehat{B}^{\pi^*}_{\mathbf{r}}(t).$$

Note: It can be seen from the proof that if λ_c is known apriori, the same conclusions hold if in (10), we replace $t/A_c(t)$ with $1/\lambda_c$. This observation will be useful in Section 4.2.

As with the single-class system, since $\mathcal{C}(t)$ and $\mathcal{N}_c(t)$ are changing only at connection arrival and departure instants, the bandwidth allocated to connections by policy π^* may change only at those instants.

4.2 System Performance Region

In this section we derive the performance region of a multiclass system. For this, we will need to make further statistical assumptions on the system parameters. These assumptions are the following.

Assumptions

1. The arrival processes $A_c(t)$, $c \in \mathcal{C}$, are independent and Poisson. The arrival rate of class c connections is λ_c^I, $0 < \lambda_c^I < \infty$. We set $\lambda^I = \sum_{c \in \mathcal{C}} \lambda_c^I$.
2. The connection holding times are i.i.d. per class and independent among classes.
3. Each class c has associated bandwidth levels $\mathcal{B}_c$, that are common for all connections belonging to the class. By $\underline{B}_c$ and $\overline{B}_c$ we denote the minimum and maximum level in $\mathcal{B}_c$, respectively. Without loss of generality we assume that $\underline{B}_c \leq B$, $c \in \mathcal{C}$ (if $\underline{B}_c > B$, connections from class c are never admitted in the system and can therefore be excluded from further consideration).

Taking into account the Connection Admission Policy, we conclude that under any Bandwidth Adaptation Policy, a class e connection arriving at time t is admitted by the system if,

$$\sum_{c \in \mathcal{C}(t)} \underline{B}_c N_c(t^-) + \underline{B}_e \leq B, \tag{13}$$

where $N_c(t^-)$ is the number of class c connections in the system before the decision to accept of reject the newly arriving connection is made. This, and Assumptions 1 and 2 imply through the Insensitivity Property [20, Theorem 5.3] that the stationary distribution $P_S(\mathbf{n})$ of the process $\mathbf{N}(t) = \{N_c(t)\}_{c \in \mathcal{C}}$ exists and that $P_S(\mathbf{0}) > \mathbf{0}$. This in turn, combined with the assumptions stated above permit us to use results from the theory of regenerative processes [21] to establish existence of limits of interest in the following development. Specifically, the processes we will be interested in are regenerative with respect to the i.i.d. process $\{T_k\}_{k=0}^{\infty}$, where $T_0 = 0$ and T_k is the kth time that the system empties, i.e., $\mathbf{N}(t^-) > 0$ and $\mathbf{N}(t) = 0$.

Given a vector $\mathbf{r} = \{r_c\}_{c \in C}$, $r_c \geq 0$, consider the policy $\pi^{\mathbf{r}}$ that operates as follows:

Policy $\pi^{\mathbf{r}}$: At time t allocate to connection $i \in \mathcal{N}(t)$ bandwidth $b_i^{\pi^*}(t) = b_i^*$, where b_i^* is the solution to the following optimization problem.

$$\max \left\{ \sum_{c \in \mathcal{C}(t)} \sum_{i \in \mathcal{N}_c(t)} \frac{r_c}{\lambda_c} \frac{b_i}{V_i} \right\} \tag{14a}$$

$$\sum_{c \in \mathcal{C}(t)} \sum_{i \in \mathcal{N}_c(t)} b_i \leq B \tag{14b}$$

$$b_i \in \mathcal{B}_c, \; i \in \mathcal{N}_c(t), \; c \in \mathcal{C}(t). \tag{14c}$$

Let

$$\widetilde{b}_c^{\pi^{\mathbf{r}}}(t) = \sum_{i \in \mathcal{N}_c(t)} \frac{b_i^{\pi^{\mathbf{r}}}(t)}{V_i}.$$

Under the stated assumptions, and since policy $\pi^{\mathbf{r}}$ depends only on the current state of the network, we conclude that the process $\left\{\widetilde{b}_c^{\pi^{\mathbf{r}}}(t)\right\}_{c\in\mathcal{C}}$ is regenerative with respect to T_k. Based on this, we derive,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \sum_{i\in\mathcal{N}_c(s)} \frac{b_i^{\pi^{\mathbf{r}}}(s)}{V_i} ds = \frac{\mathrm{E}\left[\int_0^{T_1} \widetilde{b}_c^{\pi^{\mathbf{r}}}(s) ds\right]}{\mathrm{E}\left[T_1\right]} = \widetilde{B}_{\mathbf{r},c} \ ,$$

$$\lim_{t\to\infty} \widehat{B}_c^{\pi^{\mathbf{r}}}(t) = \lim_{t\to\infty} \frac{t}{A(t)} \frac{1}{t}\int_0^t \sum_{i\in\mathcal{N}_c(s)} \frac{b_i^{\pi^{\mathbf{r}}}(s)}{V_i} ds = \frac{\widetilde{B}_{\mathbf{r},c}}{\lambda_c} \triangleq \widehat{B}_{\mathbf{r},c} \ ,$$

and,

$$\lim_{t\to\infty} \widehat{B}_{\mathbf{r}}^{\pi^{\mathbf{r}}}(t) \triangleq \lim_{t\to\infty} \frac{1}{t}\int_0^t \sum_{c\in\mathcal{C}} \sum_{i\in\mathcal{N}_c(s)} \frac{t}{A_c(t)} r_c \frac{b_i^{\pi}(s)}{V_i} ds = \sum_{c\in\mathcal{C}} r_c \widehat{B}_{\mathbf{r},c} \triangleq f(\mathbf{r}).$$

According to the discussion in Section 4.1 (see note at the end of that section) we have that under any Bandwidth Adaptation Policy π,

$$\limsup_{t\to\infty} \sum_{c\in\mathcal{C}} r_c \widehat{B}_c^{\pi}(t) \le f(\mathbf{r}). \tag{15}$$

Let $\mathcal{Q}$ be the convex hull of the set of points

$$\left\{\mathbf{b} : \ \mathbf{b} = \left\{\widehat{B}_{\mathbf{r},c}\right\}_{c\in\mathcal{C}}, \mathbf{r} = \{r_c\}_{c\in\mathcal{C}}, r_c \ge 0\right\}.$$

Let also $\mathcal{P}$ be the set of all points that are coordinate-wise smaller than some point in $\mathcal{Q}$, that is,

$$\mathcal{P} = \left\{\mathbf{b} = \{b_c\}_{c\in\mathcal{C}} : b_c \le b_c^q, \text{ for some } \mathbf{b}^q = \{b_c^q\}_{c\in\mathcal{C}} \in \mathcal{Q}\right\}.$$

The following lemma states that the system performance lies asymptotically in the region $\mathcal{P}$. We denote $\widehat{\mathbf{B}}^{\pi}(t) = \left\{\widehat{B}_c^{\pi}(t)\right\}_{c\in\mathcal{C}}$.

Lemma 3 *Under any Bandwidth Adaptation Policy* π*,*

$$\lim_{t\to\infty} \left(\inf\left\{\left\|\widehat{\mathbf{B}}^{\pi}(t) - \mathbf{b}\right\| : \mathbf{b} \in \mathcal{P}\right\}\right) = 0.$$

The region $\mathcal{P}$ is characterized implicitly through the vectors $\mathbf{r}$ with nonnegative coordinates. Under additional assumptions on class holding times and bandwidth levels we show in the extended version of the paper that the region is a polymatroid. However, for the purposes of developing optimal policies in Section 5, the implicit characterization will suffice. We will also need the next lemma which follows from the definitions above.

Lemma 4 *It holds for any vector* $\mathbf{x}\in\mathcal{P}$ *and any vector* $\mathbf{r} = \{r_c\}_{c\in\mathcal{C}}$ *with* $r_c \ge 0$*,*

$$\sum_{c\in\mathcal{C}} r_c x_c \le \sum_{c\in\mathcal{C}} r_c \widehat{B}_{\mathbf{r},c}.$$

5 An Adaptive Policy for Separable Convex Optimization

In this section we assume that with each class $c \in \mathcal{C}$ there is an associated reward function $f_c(b)$ so that, under Bandwidth Adaptation Policy π, the reward of class c for receiving performance $\widehat{B}_c^\pi(t)$ by time t, is $f_c\left(\widehat{B}_c^\pi(t)\right)$. The function $f_c(b)$ is assumed to be concave, twice continuously differentiable. Then, the total system reward by time t is,

$$\sum_{i \in \mathcal{C}} f_c\left(\widehat{B}_c^\pi(t)\right).$$

We will define a Bandwidth Adaptation Policy π^* that maximizes the long-term system reward, i.e. for any other policy π,

$$\limsup_{t \to \infty} \sum_{i \in \mathcal{C}} f_c\left(\widehat{B}_c^\pi(t)\right) \leq \lim_{t \to \infty} \sum_{i \in \mathcal{C}} f_c\left(\widehat{B}_c^{\pi^*}(t)\right).$$

By the proper choice of the class reward functions, various fairness criteria may by satisfied [22].

The class c performance measure, $\widehat{B}_c^\pi(t)$, is in effect a time-average of scaled bandwidths allocated to class c connections. In Section 4.1 we presented a simple optimal policy for the case of linear rewards, i.e., $f_c(b) = r_c b$ and in Section 4.2 we presented the performance region of the system under any Bandwidth Adaptation Policy. We therefore have all the ingredients to apply the framework proposed in [15] in order to develop an optimal adaptive policy π^*.

Next we present the resulting policy. In the extended version of the paper we present the arguments justifying its optimality.

Optimal Policy π^*. Let t_n be an increasing sequence of update times such that $t_n \to \infty$. At time t_n, $n = 1, 2, ...$, set

$$r_c(t_n) = \left. \frac{df_c(b)}{db} \right|_{b = \widehat{B}_c^{\pi^*}(t_n))}.$$

At time $t \in [t_n, t_{n+1})$, allocate to connection $i \in \mathcal{N}(t)$ bandwidth $b_i^{\pi^*}(t) = b_i^*$, where b_i^* is the solution to the following optimization problem.

$$\max \left\{ \sum_{c \in \mathcal{C}(t)} \sum_{i \in \mathcal{N}_c(t)} r_c(t_n) \frac{t}{A_c(t)} \frac{b_i}{V_i} \right\}$$
$$\sum_{c \in \mathcal{C}(t)} \sum_{i \in \mathcal{N}_c(t)} b_i \leq B$$
$$b_i \in \mathcal{B}_c, \; i \in \mathcal{N}_c(t), \; c \in \mathcal{C}(t).$$

The times t_n need to satisfy the following mild technical condition.

Condition on Update Times: For some $L \geq 1$, for each $m \geq 0$, there is $n \geq 1$ such that

$$T_{mL} \leq t_n < T_{(m+1)L}.$$

Consequently, successive updates are no more than $2L$ regeneration periods apart.

6 Discussion and Extensions

We considered the problem of channel sharing by rate-adaptive multi-class streams. We presented policies maximizing a linear combination of average scaled connection bandwidths, under quite general conditions. Under additional but reasonable statistical assumptions, we described the performance region of the system and applied a general methodology to provide a Bandwidth Adaptation Policy for maximizing a combination of convex class performance rewards.

In Section 2 we assumed that the channel bandwidth is fixed. If the channel bandwidth is a function of time as may be the case in wireless systems, then the results in Sections 3 and 4.1 still hold. Regarding the results in Sections 4.2 and 5, extensions are still possible by adding statistical assumptions on the channel bandwidth fluctuation. While these assumptions may not be severe, several technical questions need to be answered in order to completely justify the optimality of the proposed policy.

Another assumption in Section 2 concerns the Connection Admission Policy adopted in this paper. This is a reasonable policy that decouples in effect connection admission from bandwidth adaptation and has been proposed by several works. We note, however, that the only property of the policy we used in this paper is that $\mathcal{N}_c(t),\ c \in \mathcal{C}$, is independent of the employed Bandwidth Adaptation Policy. Other connection admission policies that satisfy this property may also be employed, e.g. those proposed in [8], [9].

In this work, we concentrated on connection bandwidth as the basic performance measure. The results can be easily applied if as measure for the perceptual quality of reception is considered a function $r(b(t))$ of the bandwidth received by a connection at time t [7], [9]. The main difference is that now one has to solve nonlinear optimization problems at the decision instants.

Our final comment regards the adaptation capabilities of the proposed policy. Since the policy is based on the observation of past history (through the quantities $\widehat{B}_c^{\pi^*}(t)$), in order to adapt faster to changes in statistical parameters, it is important to introduce techniques that weigh recent history more. A standard way of doing this, is to work with the weighted averages of the quantities involved. It can be shown that this way the adaptivity of the policy can be improved at the expense of small deviation from optimality.

References

[1] A. Eleftheriadis, D. Anastasiou, Optimal data partitioning of MPEG-2 coded video, in: First IEEE International Conference on Image Processing, 1994.

[2] P. Pancha, M. Zarki, Prioritized transmission of variable bit rate MPEG video, in: IEEE GLOBECOM, 1992.

[3] N. G. Duffield, K. K. Ramakrishnan, A. R. Reibman, SAVE: An algorithms for smoothed adaptive video over explicit rate networks, in: IEEE INFOCOM, 1998.

[4] J. Bolot, T. Turletti, A rate control mechanism for packet video in the internet, in: IEEE INFOCOM, 1994.

[5] D. Taubman, A. Zakhor, A common framework for rate and distortion based scaling of highly scalable compressed video, IEEE Transactions on Circuits and Systems for Video Technology 6 (4) (1996) 329–354.

[6] A. K. Talukdar, B. R. Badrinath, A. Acharya, Rate adaptation schemes in networks with mobile hosts, in: Proceedings of ACM/IEEE MobiCom '98, 1998, pp. 169–180.

[7] V. Bharghavan, K. Lee, S. Lu, S. Ha, D. Dwyer, The TIMELY adaptive resource management architecture, IEEE Personal Communications Magazine 5 (4) (1998) 20–31.

[8] T. Kwon, S. Kim, Y. Choi, N. Naghshineh, Threshold-type call admission control in Wireless/Mobile multimedia networks using prioritized adaptive framework, Electronics Letters 36 (9) (2000) 852–854.

[9] T. Kwon, Y. Choi, S. K. Das, Bandwidth adaptation algorithms for adaptive multimedia services in mobile cellular networks, Wireless Personal Communications 22 (2002) 337–357.

[10] S. Weber, G. de Veciana, Telecommunications Network Design, Academic Publisher, 2002, Ch. Asymptotic Analysis of Rate Adaptive Multimedia Streams.

[11] T. Kwon, Y. Choi, C. Bisdikian, M. Naghshineh, QoS provisining in Wireless/Mobile multimedia networks using and adaptive framework, Wireless Networks 9 (2003) 51–59.

[12] S. Weber, G. Veciana, Rate adaptivre multimedia streams: Optimization, admission control, and distributed algorithms, IEEE Transactions on Networking 1 (1) (2003) 1–16.

[13] N. Argiriou, L. Georgiadis, Channel sharing by rate-adaptive streaming applications, Performance Evaluation 55 (3-4) (2004) 211–229.

[14] P. P. Bhattacharya, L. Georgiadis, P. Tsoucas, Problems of adaptive optimization in multiclass M/GI/1 queues with bernoulli feedback, Mathematics of Operations Research 20 (1995) 356–380.

[15] V. Tsibonis, L. Georgiadis, An adaptive framework for addressing fairness issues in wireless networks, Computer Communications 28 (10) (2005) 1167–1178.

[16] ITU-T, Objective quality measurements of telephone band and (300-3400 hz) speech codecs, in: Series P: Telepnoe Transmission Quality, Telephone Installation, Local Line Networks, 1998.

[17] ITU-500-R, ITU-500-r recommendation BT.500-8 - methodology for the subjective accessment of the quality of television pictures, 1998.

[18] S. Winkler, A perceptual distortion metric for digital color video, in: Proceedings of SPIE Human Vision and Electronic Imaging, Vol. 3644, 1999, pp. 175–184.

[19] A. R. Prasad, R. Esmailzadeh, S. Winkler, T. Ihara, B. Rohani, B. Pinguet, M. Capel, Perceptual quality measurement and control: Definition, application and performance, in: Poceedings 4th International Symposium on Wireless Personal Multimedia Communications, Aarborg, Denmark, 2001, pp. 547–552.

[20] K. W. Ross, Multiservice Loss Models for Broadband Telecommunication Networks, Springer, 1995.

[21] R. W. Wolff, Stochastic Modeling and the Theory of Queues, Prentice Hall, 1989.

[22] J. Y. L. Boudec, Rate adaptation, c ongestion control and fairness: A tutorial, available at: http://people.epfl.ch/jean-yves.leboudec/.