Design of Generalized Processor Sharing Schedulers Which Statistically Multiplex Heterogeneous QoS Classes

Anwar Elwalid   *Debasis Mitra*

Bell Laboratories
Lucent Technologies
Murray Hill, New Jersey 07974

Abstract

Generalized Processor Sharing (GPS) is the basis for the packet scheduler of choice in IP routers and ATM switches of the future. The currently accepted approach for the design of GPS schedulers is based on deterministic QoS guarantees, which, it is generally accepted, is overly conservative and leads to limitations on capacity. To address this problem we develop a framework for GPS scheduling which is based on statistical QoS guarantees and statistical multiplexing. We give the design of GPS weights which maximize the coverage of operating points, and also the design of the connection admission control (CAC). The general framework is end-to-end, with two heterogeneous QoS classes coexisting with a third, best effort class. Each QoS class has a specified delay bound together with a bound on the probability of its violation. An important objective is to maximize the bandwidth available to best effort traffic while just satisfying the guarantees of the QoS classes. To this end, we consider output regulated GPS scheduling which has the additional feature of limiting each connection's share of the bandwidth to a specified value, a design parameter which is determined by our analysis. The sources are subject to standard dual leaky bucket regulation. For the design of the GPS weights we give procedures based on two key concepts, the realizable set and the critical weights. The realizable set is the union of all admissible sets of connections of both classes over all weights. One of the main contributions is a pragmatic design process by which most of the realizable set is realized by only two critical weights. In the benign case, the system is "Effectively Homogeneous" and a single GPS weight suffices, while in the complementary "Effectively Non-Homogeneous" case it is necessary to switch between the critical weights. The numerical results, which are for a single node with a wide range of traffic and QoS parameters, validate the design procedure and also show that there are substantial capacity gains from statistical multiplexing.

1. Introduction

Generalized Processor Sharing (GPS) provides the basis for the packet scheduler of choice in IP routers and ATM switches of the future [KES97, KES96, SVA98(a), SVA98(b)]. The currently accepted approach for the design of the GPS scheduler and control of networks which use it is based on deterministic QoS guarantees. In this connection the work and bounds of Parekh and Gallager [PAR92, PGA93, PGA94] form an important and original point of reference. However, it is generally accepted that the bounds are overly conservative, limiting capacity and the utility of the approach to guide the design and operations of real networks. To address this problem we develop in this paper a framework for GPS scheduling and network control which is based on statistical QoS guarantees and statistical multiplexing. We give the design of GPS weights which maximize coverage of operating points in the number of connections of heterogeneous QoS classes with only a very small repertoire of weights, together with the associated design of connection admission control (CAC). The gain in capacity is typically significant.

The results presented in this paper are in the framework of end-to-end QoS guarantees. We consider two heterogeneous QoS classes coexisting with a third, best effort class. The QoS classes have a specified end-to-end delay requirement together with a bound, typically quite small, on the probability of its violation, or loss. While not always in the forefront, the role of the best effort traffic is important in our conceptual framework, and a high level objective is to maximize the bandwidth available to best effort traffic while just satisfying the guarantees of the QoS classes. A similar goal is expressed by Cruz [CRU97], although the frameworks and enabling techniques are apparently quite different.

The main contributions of this paper are for a single node in which the service discipline is GPS. There are $\mathbf{K} = (K_1, K_2)$ connections or sources of the two QoS classes, and each connection's scheduling parameters are ϕ_j and $P_o^{(j)}$, for class j connections, where the weight ϕ_j, in relation to the weights of the other connections, determines the connection's share of the bandwidth, which, however, is not allowed to exceed the peak rate $P_o^{(j)}$. Our analysis for the single node is carried out first with peak rate regulation absent, and this analysis is then used to select $P_o^{(\cdot)}$ such that the analysis still holds when peak rate regulation is operative. The QoS parameters for class j are D_j, the delay bound, and L_j, the loss probability. The class dependence is important, for in one representative example, class 1 is voice-like for which D_1 is small and L_1 large, while class 2 is premium data-like for which D_2 is large and L_2 small.

Our source model is the fluid rate process which is adversarial while compliant with dual leaky bucket (r, B_T, P) regulation, where r and P are the mean and peak rate constraints, respectively, and B_T is the token buffer size [TUR86, CRU91, ROB92]. This is an on-off process, which we call the extremal regulated source process, that transmits at peak rate from the instant that the token buffer is full till it is empty, and then turns off and remains so till the token buffer is full. In various prior works [EMW95, MMO95, EMI97, LZT97, RRR98(a), RRR98(b)] this process has been used and its extremal adversarial properties noted (see, in particular, [RRR98(b)]).

The development of the single node analysis is in two phases, lossless multiplexing followed by statistical multiplexing. The former assumes that sources may collude and in this case the above on-off process is shown to be maximally adversarial. This phase parallels the standard analysis for deterministic guarantees, except for the additional element of the delay constraint. The analysis in the latter phase combines (i) the assumption of non-colluding sources, which gives statistical independence to the sources and uniformly distributed random phases to the individual sources' rate processes, and (ii) the use of the unutilized portions of the allocated resources to sources, which is not exploited in lossless multiplexing.

How should the GPS weights (ϕ_1, ϕ_2) or, more specifically, the ratio $\phi \triangleq \phi_1/\phi_2$ (which we shall also refer to as a weight) be chosen? Our first step is to determine the admissible set $\mathcal{A}(\phi)$, i.e., for fixed ϕ, the set of all (K_1, K_2) such that the statistical QoS of each connection is satisfied.

The size and shape of the admissible region depends strongly on the weight. It is easy to mismatch the GPS weight to the traffic and QoS parameters of the classes. For instance, if ϕ is excessively large then the admissible set is limited by the QoS considerations of the class 2 connections and is consequently small in size, with the relatively few class 1 connections receiving QoS in excess of their requirements. When the weight is not properly designed resources are misallocated and sharing is inhibited.

An important concept for GPS design is the realizable set $\mathcal{R}$. This is the set of all (K_1, K_2) which are contained in the admissible set of some weight ϕ. That is, the realizable set is the union of the admissible sets $\mathcal{A}(\phi)$ over all ϕ. The role of the realizable set is as follows: say exogeneous (traffic and market) conditions dictate a desirable operating neighborhood in connections around the point (K_1, K_2). If this neighborhood is contained in $\mathcal{R}$ then an inverse problem is solved to obtain the appropriate weight. As the exogeneous conditions change, typically on a slow time scale, the GPS weight changes too in correspondence. Now, on the face of it, this notion is only of theoretical value with possibly little practical relevance, since the realization of all of $\mathcal{R}$ may in principle require switching between a large, even possibly an uncountable, number of weights. One of our main contributions is to show by a pragmatic design approach that most of $\mathcal{R}$ may be realized by either one or two weights. These are called critical weights, $\phi_c^{(1)}$ and $\phi_c^{(2)}$, and simple procedures for their calculations are given in the paper.

An important, interesting dichotomy is at the root of whether one or two weights are needed. We classify the entire system, as characterized by the traffic and QoS parameters of both classes, to be either "Effectively Homogeneous" or "Effectively Non-Homogeneous", depending upon whether $\phi_c^{(1)} \leq \phi_c^{(2)}$ or otherwise. The former is the benign case and the mid-point between the critical weights is a suitable choice to realize almost all of $\mathcal{R}$. In the latter case realizing $\mathcal{R}$ requires switching between the critical weights, which, as argued above, would occur on a slow time scale to track exogeneous traffic conditions. In particular, switching at this rate would be compatible with the tacit assumption in the analysis that steady state exists. We should point out that in our numerical investigations we have encountered examples of effectively homogeneous systems in which there exists considerable disparity in traffic and QoS parameters between the classes.

Our numerical investigations consider a wide range of traffic and QoS parameters. The results reported here for the single node provide support for our pragmatic approach to the design problem, which is based on piece-wise linear approximations to the boundaries of the admissible and realizable regions, and, more specifically, for our claim that either one or two weights suffice to realize most of $\mathcal{R}$. Finally, these results show that there are substantial capacity gains from statistical over deterministic, or lossless, QoS guarantees.

Proceeding to the network-wide context, recall that our objective is to maximize carried best-effort traffic, while meeting the requirements of the QoS classes. This is the prime motivation for regulation of the peak rate of the output of the GPS schedulers. Output regulated GPS has the effect of allocating resources to the admitted QoS connections to just meet their requirements. There are further choices to be made regarding the delay budget, i.e., allocation of portions of the end-to-end delay to nodes in the path of the connection. An attractive option, we argue, is to allocate the entire delay budget to the first node in the path. A similar proposal in the FIFO scheduling framework has been made recently by Ross, Reisslein and Rajagopal [RRR98(b)]. In the GPS framework one attraction, among others, of this proposal is that the complexity of the GPS design is contained and the detailed analysis given here applies. Otherwise there is a proliferation of classes for which the GPS design problem is complexified, even though the basic framework of a single GPS node analysis of this paper remains intact.

We should point out that the current preference in the DiffServ working group of the IETF is to work with a small number, possibly just two, of QoS classes. Also, the dual leaky bucket regulator has been adopted by the ATM Forum [ATM93] and is reflected in various IETF protocol specifications, such as RSVP [WRO97(a)], the Controlled Load [WRO97(b)], and Guaranteed [SPG97] quality of service.

We note that there is a well-developed body of prior work on the analysis of GPS in the setting of stochastic source models, with the most complete results applying to two queues [YSI94, DVK96, LZT96]. As previously noted, the analysis in the deterministic setting is well developed and based on quite general source models, for instance Cruz's Linear Bounded Arrival Process [CRU91]. Finally, we note that the antecedents of the analytic techniques developed here are in techniques for FIFO scheduling in [EMW95]. Curiously the construct of the single source system while only 'virtual' in the earlier work is quite real here in the GPS setting.

Section 2 is on preliminaries and background facts. Section 3 is on lossless multiplexing with deterministic guarantees and Section 4 is on the single node analysis of statistical multiplexing and GPS design. Section 5 reports on numerical investigations. Section 6 is on end-to-end design and performance. Concluding remarks are in Section 7.

2. Preliminaries and Background Facts

2.1. Flows, Rates, Dual Leaky Bucket Regulator

Here we review some basic quantities and operations of the dual leaky bucket regulator (r, B_T, P), where r is the token (mean) rate, B_T is the token buffer size and P is the peak rate. Traffic streams are modelled as fluid flow here and throughout the paper. Let $U(t)$ and $R(t)$ be *rate processes*, where $U(t)$ is offered by the source and $R(t)$ is the conforming process which is passed to the network by the regulator. We let $A_U(t,\, t+\tau)$ and $A_R(t,\, t+\tau)$ denote the *total flows* during the interval $[t,\, t+\tau]$ for the U and R rate processes, respectively, where rates and total flows are generally related thus:

$$A_R(t,\, t+\tau) = \int_t^{t+\tau} R(s)\, ds \,. \tag{2.1}$$

The basic relation describing the conforming process $R(t)$ is

$$A_R(t,\, t+\tau) = \min(A_U(t,\, t+\tau),\ P\tau,\ z(t) + r\tau) \quad \forall t,\ \forall \tau > 0\ . \tag{2.2}$$

In (2.2) $z(t)$ is the token content of the token buffer at time t. The governing equations of $z(t)$ are

$$\left.\begin{aligned} \frac{d}{dt} z(t) &= r - R(t), && \text{if } \ 0 < z(t) < B_T \\ &= [r - R(t)]^{+}, && \text{if } \ z(t) = 0 \\ &= [r - R(t)]^{-}, && \text{if } \ z(t) = B_T \end{aligned}\right\} . \tag{2.3}$$

It may be verified that the regulated rate process $R(t)$ is constrained thus,

$$\begin{aligned} R(t) &= \min(U(t),\, P) && \text{if } \ z(t) > 0 \\ &= \min(U(t),\, r) && \text{if } \ z(t) = 0\ . \end{aligned} \tag{2.4}$$

We will find it convenient to equivalently characterize the dual leaky bucket regulator by the three-tuple, (r, Q, P), where Q is the maximum burst size and is given by $Q \triangleq PB_T/(P-r)$.

2.2. Generalized Processor Sharing with Output Regulation

Figure 2.1 shows a functional diagram of a generalized processing sharing (GPS) scheduler. The traffic of each connection is dual leaky bucket regulated prior to arrival at the GPS node. In the GPS node each connection is provided a buffer whose output is subject to peak rate regulation. There are two QoS classes of connections sharing bandwidth C. (This is the total link bandwidth minus the bandwidth allocated to the best effort traffic.) Each connection of class j (j=1,2) has two parameters: ϕ_j and $P_o^{(j)}$, which are used to schedule service to the connection. The weight ϕ_j, in relation to the weights associated with the other connections, determines the connection's share of the bandwidth, which, however, is not allowed to exceed $P_o^{(j)}$. Thus the output rate from the connection buffer (see Figure 2.1) does not exceed $P_o^{(j)}$. In our work, the buffer sizes, $\{\phi_j\}$ and $\{P_o^{(j)}\}$ are design parameters to be determined in this paper. The best effort traffic, which is not shown in the diagram, is allocated fixed bandwidth for its exclusive use and is allowed access to the bandwidth unused by the QoS connections. Hence, the QoS classes are oblivious to the best effort traffic. Section 6 deals with issues related to best effort traffic.

The GPS scheduler is work conserving, and in conjunction with the output peak rate regulation, the bandwidth to a connection of class j is at least

$$\min\left(P_o^{(j)},\, \frac{\phi_j C}{\phi_1 \sum_{i=1}^{K1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{K2} \xi_i^{(2)}}\right) \qquad (j = 1, 2)$$

where $\xi_i^{(j)}$ is an indicator variable for the ith connection of class j, which is 1 if the connection requires bandwidth and 0 otherwise.

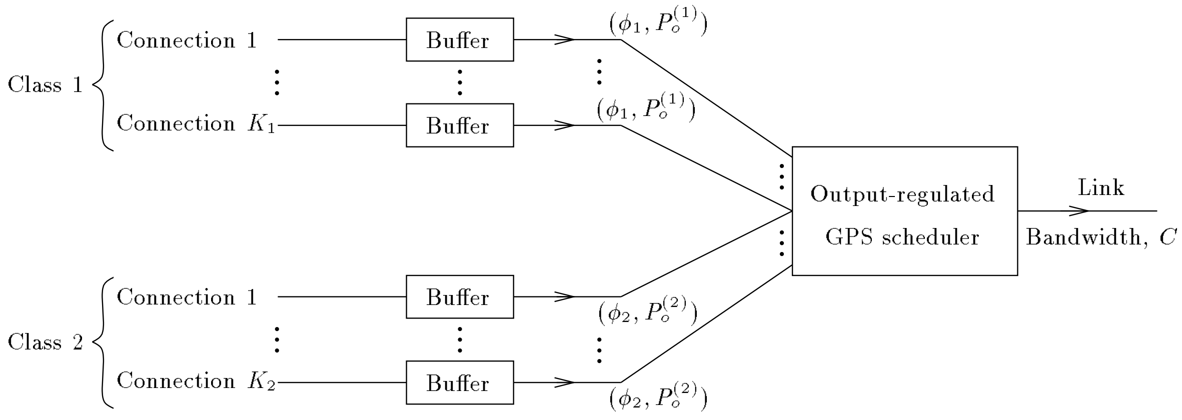

Figure 2.1: Functional diagram of GPS scheduling at a node.

2.3. Chernoff Approximation

The Chernoff large deviations approximation [BIL86, CSE93, PET65] is an effective technique for estimating loss in bufferless multiplexing systems. The procedure given here is due to Hui and Kelly [HUI90, KEL91]. Consider

$$P_{\text{loss}} = \Pr\,(U > C)\ , \tag{2.5}$$

where the total instantaneous demand $U = \sum_{j=1}^{J} \sum_{k=1}^{K_j} u_{jk}$, and $\{u_{jk}\}$ are independent random variables and $\{u_{jk}\}_k$ are nonnegative identically distributed for each j. Here there are J source classes and K_j sources of class j. When C is the nodal bandwidth, P_{loss} is the fraction of time that losses occur. Other performance measures related to P_{loss} are discussed in [ROB92].

Assume that the instantaneous loads u_{jk} have moment generating functions

$$M_j(s) = E[e^{su_{jk}}] \quad (j = 1, 2, \cdots, J)\ . \tag{2.6}$$

Also assume the stability condition $\sum_j K_j E(u_{jk}) < C$ and that $\lim_{s\to\infty} \sum_j K_j M_j'(s)/M_j(s) > C$, as otherwise there is no loss. Then the Chernoff bound augmented by a lower order refinement [PET65] gives the following approximation for the scaling $K_j = \gamma_j C$, where $\gamma_j = O(1)$ and $C \to \infty$.

FACT 1:

$$P_{\text{loss}} = \frac{\exp\{-F(s^*)\}}{s^*\sigma(s^*)\sqrt{2\pi}}\,[1+o(1)] \tag{2.7}$$

where $F(s) = sC - \sum_j K_j \log M_j(s)$, $F(s^*) = \sup_{s\geq 0} F(s)$, and $\sigma^2(s) = \partial^2 \log E(e^{sU})/\partial s^2$.

□

Later in the paper we will encounter the case where u_{jk} are independent on-off random variables which take values 0 and 1 and $E(u_{jk}) = \omega^{(j)}$, i.e.,

$$\Pr\,(u_{jk}=1) = 1 - \Pr\,(u_{jk}=0) = \omega^{(j)}\,. \tag{2.8}$$

Assume that $\Sigma K_j\omega^{(j)} < C$ and $\Sigma K_j > C$.

FACT 2: The Chernoff approximation for P_{loss} with multiple class is

$$P_{\text{loss}} \sim \frac{e^{-F(s^*)}}{s^*\sigma(s^*)\sqrt{2\pi}}\,, \tag{2.9}$$

where s^* is the unique positive real solution to the equation

$$e^s \sum_j \frac{K_j\omega^{(j)}}{1-\omega^{(j)}+\omega^{(j)}e^s} = C\,, \tag{2.10}$$

$$F(s) = sC - \sum_j K_j \log\left\{1-\omega^{(j)}+\omega^{(j)}e^s\right\}\,, \tag{2.11}$$

and

$$\sigma^2(s) = e^s \sum_j \frac{K_j\omega^{(j)}(1-\omega^{(j)})}{\left\{1-\omega^{(j)}+\omega^{(j)}e^s\right\}^2}\,. \tag{2.12}$$

□

Of considerable interest is the following explicit form of the Chernoff asymptotic approximation for the further specialization to a single source class. Here we drop the class index j.

FACT 3:

$$P_{\text{loss}} \sim \frac{e^{-F(s^*)}}{s^*\sigma(s^*)\sqrt{2\pi}}\,, \tag{2.13}$$

where

$$a = \frac{C}{K}, \quad s^* = \log\left[\frac{a}{1-a}\cdot\frac{1-\omega}{\omega}\right]\,,$$

$$F(s^*) = K\left[a\log\left(\frac{a}{\omega}\right) + (1-a)\log\left(\frac{1-a}{1-\omega}\right)\right]\,, \tag{2.14}$$

and

$$\sigma^2(s) = K\omega(1-\omega)e^s/\{1-\omega+\omega\ e^s\}^2\,. \tag{2.15}$$

□

The numerical results that we report in this paper use the expressions in Facts 2 and 3.

Although we have no specific need for the result, the reader may find it interesting to note a result [MMO95] in the multi-class framework concerning distributions of independent random variables which maximize the Chernoff approximation $P_{\text{loss}} \sim \exp(-F(s^*))$. Assume that the random variables u_{jk} are subject to the peak constraint of $e_0^{(j)}$ and their mean is bounded by $r^{(j)}$. The result states that the exponential Chernoff approximation is maximized by on-off random variables which take values 0 and $e_0^{(j)}$, and have mean $r^{(j)}$.

3. Lossless Multiplexing

In this section we consider the case in which there is a single node and the QoS guarantee is deterministic, i.e., no loss, or violation, is permitted of the delay requirement. The delay bound may depend on the class of the connection. The treatment in the next section on statistical multiplexing expands on the base built here, and the numerical results presented later compare the capacities obtained from lossless and statistical multiplexing.

Peak rate regulation of the output of each connection's buffer is ignored in the analysis of this section. The results will determine the value of the peak rate.

3.1. Single Source System

We begin by considering a single regulated traffic stream $R(t)$ which feeds a buffer of unlimited capacity and the buffer is serviced by a trunk of capacity, i.e., bandwidth, c ($c > r$). See Figure 3.1 and recall Section 2.1.

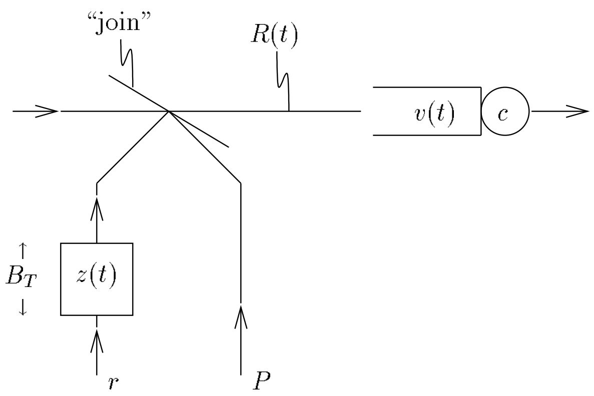

Figure 3.1: Single source system.

The content of the system buffer is denoted by $v(t)$. The main goal of this subsection is to derive a simple expression for the maximum of the system buffer content. For our purposes here the regulated stream $R(t)$ is such that the total flow in any interval $[t, t+\tau]$, $A_R(t, t+\tau)$, is bounded thus

$$A_R(t, t+\tau) \le \min(P\tau,\ z(t) + r\tau) \quad \forall t,\ \forall \tau > 0\ . \tag{3.1}$$

The "min" operation in (3.1) is performed by the "join" in the schematic of Figure 3.1.

Proposition 1: For all t

$$v(t) \le \frac{(P-c)B_T}{(P-r)} \triangleq b\ . \tag{3.2}$$

□

Proof: Clearly the maximum of $v(t)$ occurs during a busy period of the buffer. Let t_0 correspond to the onset of a busy period, i.e., $v(t_0) = 0$ and $v(t_0+\ ) > 0$. As Figure 3.2 shows, in the busy period the total flow into the buffer during the interval $[t_0, t_0 + \tau]$ is, from (3.1), bounded by

$$\begin{cases} P\tau & 0 \le \tau \le z(t_0)/(P-r) \\ z(t_0) + r\tau & z(t_0)/(P-r) \le \tau\ . \end{cases} \tag{3.3}$$

On the other hand, since by assumption the interval $[t_0, t_0 + \tau]$ is contained in the busy period, the total flow out of the buffer is at least $c\tau$. The difference, $\bar{v}(\tau)$, is a bound on the buffer content during the buffer's busy period. Hence, for $\tau \ge 0$,

$$\bar{v}(\tau) \le \frac{(P-c)\, z(t_0)}{(P-r)} \le \frac{(P-c)\, B_T}{(P-r)}\ ,$$

which completes the proof. □

The result in (3.2) is tight, i.e., the bound is realized by the *extremal regulated source process*, which we now specify. This rate process turns on when the token buffer is full

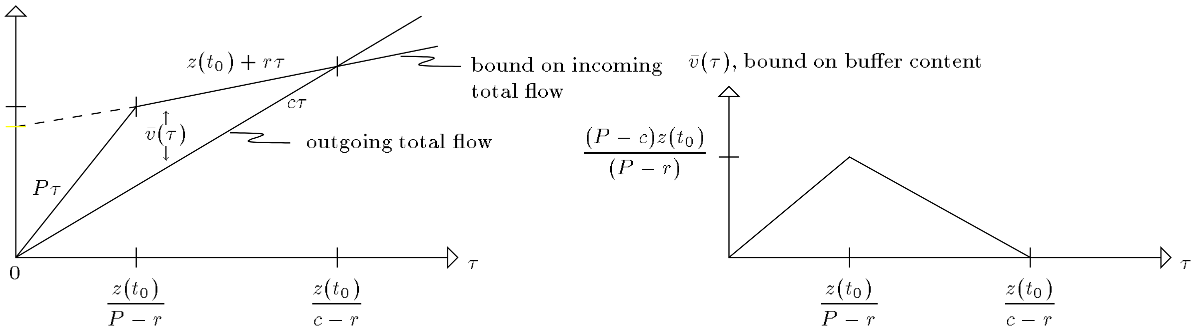

Figure 3.2: Bound on buffer content during busy period.

and transmits at peak rate P, it turns off when the token buffer first becomes empty, and resumes the cycle at the instant the token buffer again becomes full. We denote this extremal process by $R_\Omega(t)$:

$$\begin{aligned} R_\Omega(t) &= P \qquad && 0 \le t \le T_{\text{on}} \\ &= 0 && T_{\text{on}} < t \le T_{\text{on}} + T_{\text{off}} \end{aligned} \tag{3.4}$$

where $T_{\text{on}} \leqq B_T/(P-r)$ and $T_{\text{off}} = B_T/r$. Also let $T = T_{\text{on}} + T_{\text{off}}$ and the source activity factor $\omega_\Omega = r/P$.

As indicated in (3.4) the bound on the buffer content is given by b, which may be viewed as a function of the bandwidth variable c, i.e., $b = b(c)$. Note from Figure 3.2 that the buffer is non-empty for an interval which is bounded by D_{on}, where

$$D_{\text{on}} = T_{\text{on}} + b/c\,. \tag{3.5}$$

Similarly, the bandwidth c, when viewed as a resource, is utilized during the same period that the buffer is non-empty. When the input is the extremal regulated source process $R_\Omega(t)$, this period is followed by a period of length D_{off} during which the buffer and bandwidth is not utilized, where

$$D_{\text{off}} = T_{\text{off}} - b/c\,. \tag{3.6}$$

Hence, for this input process the fraction of time that the buffer and bandwidth is utilized is

$$\omega = \frac{D_{\text{on}}}{D_{\text{on}} + D_{\text{off}}} = \omega_\Omega + \frac{b}{c} \cdot \frac{1}{T} \tag{3.7}$$

3.2. Multiplexing and GPS Design

We consider lossless multiplexing in the GPS system, the design of weights (ϕ_1, ϕ_2) and its impact on the admissible and realizable sets. The QoS requirements are that (i) there are no losses, and (ii) the maximum delay is bounded by D_j for class j $(j = 1, 2)$. We will lead to the specification of the realizable set $\mathcal{R}_0$, which is the set of sources $\mathbf{K} = (K_1, K_2)$ for which there exists weights such that the QoS requirements are satisfied.

We have seen that, for the single source system, when the bandwidth allocated is c, the buffer content is at most $b(c)$, see (3.2). Hence the delay is at most b/c and consequently the QoS requirement that the delay be at most D is satisfied if

$$b/c \le D\,. \tag{3.8}$$

As indicated in Figure 3.3, the smallest value of the bandwidth c which satisfies (3.2) and (3.8) is e_0, and the corresponding buffer size for which no losses occur is b_0. We tacitly assume that $D \leq B_T/r$, since otherwise the selection $(e_0, b_0) = (r, B_T)$ suffices. Hence

$$e_0 = \frac{P}{1 + D/T_{\mathrm{on}}} \tag{3.9}$$

We call e_0 the effective bandwidth for lossless multiplexing (the subscript 0 is a mnemonic for zero loss).

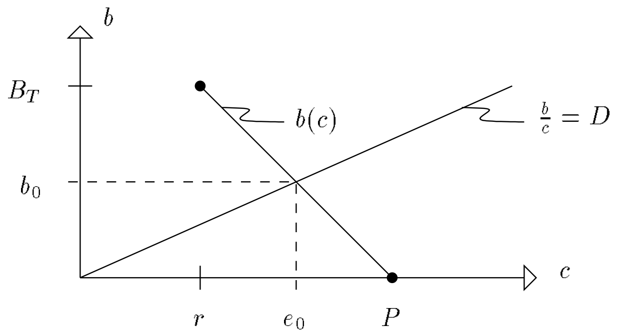

Figure 3.3: Effective bandwidth e_0 and corresponding buffer b_0 for lossless multiplexing.

Proceeding to the design of GPS weights, we begin by tagging a source of class 1, say source 1, and we assume that it is active, i.e., its bandwidth is utilized. Hence, the tagged source will satisfy its QoS requirements if it receives bandwidth $e_0^{(1)}$, the effective bandwidth of lossless multiplexing for class 1 sources. The fair share of the tagged source is at least

$$\frac{\phi_1 C}{\phi_1 + \phi_1 \sum_{i=1}^{K_1-1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{K_2} \xi_i^{(2)}}. \tag{3.10}$$

Here K_j is the number of sources of class j and $\xi_i^{(j)}$ is an indicator variable for source i of class j that takes value 1 if the source is active and 0 otherwise. Hence the bandwidth demand of the active tagged source is satisfied if

$$e_0^{(1)} \leq \frac{\phi_1 C}{\phi_1 + \phi_1 \sum_{i=1}^{K_1-1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{K_2} \xi_i^{(2)}}, \tag{3.11}$$

i.e.,

$$\phi_1 \sum_{i=1}^{K_1-1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{K_2} \xi_i^{(2)} \leq \phi_1 \left(\frac{C}{e_0^{(1)}} - 1 \right). \tag{3.12}$$

By an analogous argument for a tagged active source of class 2,

$$\phi_1 \sum_{i=1}^{K_1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{K_2-1} \xi_i^{(2)} \leq \phi_2 \left(\frac{C}{e_0^{(2)}} - 1 \right) . \tag{3.13}$$

Both of the above conditions have to hold for QoS to be satisfied.

We return to these two conditions in the next section when we consider statistical multiplexing. Here, where the topic is lossless multiplexing, we consider the rare but still possible case where the bursts of all sources are coincident, i.e., active at a common point in time. In this worst-case scenario $\xi_i^{(j)} = 1$ for all j and i in (3.12) and (3.13). This gives

$$K_1\phi_1 + K_2\phi_2 \leq C \left[\min \left(\frac{\phi_1}{e_0^{(1)}}, \frac{\phi_2}{e_0^{(2)}} \right) \right] . \tag{3.14}$$

The selection of weights

$$\phi_1 = e_0^{(1)}, \qquad \phi_2 = e_0^{(2)} \tag{3.15}$$

realizes the following admissible set of sources $\mathbf{K}$,

$$K_1 e_0^{(1)} + K_2 e_0^{(2)} \leq C \, , \tag{3.16}$$

which we denote by $\mathcal{A}_0$. It is easy to verify that this set cannot be improved upon. Hence the realizable set $\mathcal{R}_0$ coincides with admissible set $\mathcal{A}_0$ obtained from the weights in (3.15), a fortuitous fact.

Recall that the above derivation has relied on sufficiency for satisfying QoS, as in (3.10) and (3.11). Thus the possibility exists that the realizable set may be larger than $\mathcal{R}_0$ for certain, benign sources. Now recall that to satisfy QoS, the extremal regulated source process defined in Section 3.1 has a bandwidth demand process which is on-off and takes value $e_0^{(j)}$, $j = 1, 2$, while on. For this process the admissible set in (3.16) is maximal, as is easy to verify, and in this sense the characterization of the realizable set in (3.16) is tight for lossless multiplexing.

Note from (3.15) and (3.16) that the bandwidth allocation to class j sources is $e_0^{(j)}$, which is typically greater than the usual allocation of the mean rate $r^{(j)}$. This, of course, is because of the delay requirement in the QoS specifications, which is not usually considered.

Now if we consider reinstatement of peak rate regulation of the output of each connection's buffer, a natural choice for the peak rate is $e_0^{(j)}$, for class j connections, since it is the smallest value which will not modify the above analysis and the QoS.

4. Statistical Multiplexing and GPS Design: Single Node Analysis

Here we adopt the statistical point of view and allow small probabilities that the delay requirements are not satisfied. We call these the loss probabilities. This terminology, besides being ingrained, also reflects the scenario where buffers are sized in conjunction with bandwidth allocations, and insufficient bandwidth availability is responsible for both buffer overflow and violation of the delay requirement. The QoS specifications for class j are D_j and L_j, where D_j is, as before, the delay bound and L_j is the loss probability. In the statistical multiplexing framework the design of GPS weights, while obviously complex, has the simple goal of allocating resources to just match the resources required to satisfy QoS requirements. A final overriding concern is that the design rely on only a very small set of weights, since frequent switching of weights is not feasible in operations and also leads to violations of the tacit assumption in the analysis that steady state exists.

It will become apparent that the analysis of this section applies to output regulated GPS provided only that the peak rate of the output of each connection's buffer is not less than $e_0^{(j)}$, for class j connections. Our design ensures this, and as a consequence output regulation is transparent to the analysis and hence it is ignored in this section.

4.1. Statistical QoS

We consider the extremal regulated source process, for which it was shown in Section 3 that the bandwidth demand to satisfy QoS is an on-off process taking values $e_0^{(j)}$ and 0. Denote this process by $u_{ji}(t)$ for source i of class j. From the assumption of statistical independence of sources

$$u_{ji}(t) = u_j(t + \theta_{ji}) \tag{4.1}$$

where $u_j(t)$ is a fixed, periodic on-off function with period T_j, which takes values $e_0^{(j)}$ and 0 for fractions $\omega^{(j)}$ and $(1 - \omega^{(j)})$, respectively (see (3.7)), where $\omega^{(j)} = r^{(j)}/e_0^{(j)}$; the phases θ_{ji} are independent random variables uniformly distributed over the interval T_j.

As a consequence of the above model the activity indicators $\xi_i^{(j)}$ are independent, binomial random variables,

$$\Pr\ (\xi_i^{(j)} = 1) = 1 - \Pr\ (\xi_i^{(j)} = 0) = \omega_i^{(j)}\,. \tag{4.2}$$

We return to the QoS requirements in (3.12) and (3.13) and obtain their statistical counterparts:

$$\boxed{\Pr\left[\phi_1 \sum_{i=1}^{K_1-1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{K_2} \xi_i^{(2)} > \phi_1(C/e_0^{(1)} - 1)\right] \le L_1\,,} \tag{4.3}$$

$$\boxed{\Pr\left[\phi_1 \sum_{i=1}^{K_1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{K_2-1} \xi_i^{(2)} > \phi_2(C/e_0^{(2)} - 1)\right] \le L_2\,.} \tag{4.4}$$

Here L_1 and L_2 are small QoS parameters. In (4.3), which is concerned with QoS for class 1, it is assumed that $K_1 \ge 1$. Similarly, (4.4) is concerned with QoS for class 2 and hence $K_2 \ge 1$. Importantly, both conditions have to be simultaneously satisfied. Also note that the approach leading to (4.3) and (4.4) is based on the assumption made in (3.12) and (3.13) that the tagged source is active. There are alternative measures which may be readily incorporated, for instance where the probabilistic behavior of the tagged source's activity is included. We do not expect such a modification to make a significant difference to the admissible regions.

4.2. Maximum Admissible Sources, or Corner Points

Let $\bar{K}_1^{(1)}$ and $\bar{K}_2^{(2)}$ be called "pinned corner points" for class 1 and 2, respectively, where $\bar{K}_j^{(j)}$ is the maximum number of sources of class j such that QoS of class j is satisfied when there are no sources of the complementary class. That is, from (4.3),

$$\Pr\left[\sum_{i=1}^{\bar{K}_1^{(1)}-1} \xi_i^{(1)} > (C/e_0^{(1)} - 1)\right] \le L_1 < \Pr\left[\sum_{i=1}^{\bar{K}_1^{(1)}} \xi_1^{(1)} > (C/e_0^{(1)} - 1)\right]\,. \tag{4.5}$$

We adopt the following notation to abbreviate (4.5):

$$\Pr\left[\sum_{i=1}^{\bar{K}_1^{(1)}-1} \xi_i^{(1)} > (C/e_0^{(1)} - 1)\right] \lesssim L_1\,. \tag{4.6}$$

Similarly, from (4.4),

$$\Pr\left[\sum_{i=1}^{\bar{K}_2^{(2)}-1} \xi_i^{(2)} > (C/e_0^{(2)} - 1)\right] \lesssim L_2\,. \tag{4.7}$$

The reason we call these corner points pinned is because they do not depend on the design variables ϕ_1 and ϕ_2.

Obviously the design only depends on the ratio of weights, which we denote by $\phi \triangleq \phi_1/\phi_2$. Let $\bar{K}_2^{(1)}(\phi)$ and $\bar{K}_1^{(2)}(\phi)$ be "design corner points" for class 1 and 2, respectively, where $\bar{K}_i^{(j)}(\phi)$ is the maximum number of class i sources such that, for fixed ϕ, with a single class j $(j \neq i)$ source, i.e. $K_j = 1$, this source, while active, receives resources to satisfy its QoS. Hence, from (4.3),

$$\Pr\left[\sum_{i=1}^{\bar{K}_2^{(1)}(\phi)} \xi_i^{(2)} > \phi(C/e_0^{(1)} - 1)\right] \lesssim L_1 \tag{4.8}$$

and, from (4.4),

$$\Pr\left[\sum_{i=1}^{\bar{K}_1^{(2)}(\phi)} \xi_i^{(1)} > \frac{1}{\phi}(C/e_0^{(2)} - 1)\right] \lesssim L_2\,. \tag{4.9}$$

Note that the expressions to the left of (4.6)–(4.9) are basically simple since only single classes are involved and they may be evaluated most rapidly and conveniently by using the single class Chernoff approximation in Fact 3, Section 2.3. Note too that in our notational system for the corner points the superscript refers to the class whose QoS is under consideration, and the subscript refers to the class for which we seek the maximum number of sources.

The following monotonicity and range properties of $\bar{K}_2^{(1)}(\phi)$ and $\bar{K}_1^{(2)}(\phi)$, while self-evident from (4.8) and (4.9), are of crucial importance. As ϕ ranges from small to large values, $\bar{K}_2^{(1)}(\phi)$ is monotonic, non-decreasing and $\bar{K}_1^{(2)}(\phi)$ is monotonic, non-increasing. Also, from (4.8) as $\phi \to \infty$, $\bar{K}_2^{(1)}(\phi) \to \infty$ and as $\phi \to 0$, $\bar{K}_2^{(1)}(\phi) \to 0$. Similarly, from (4.9), $\bar{K}_1^{(2)}(\phi)$ can be made as large and as small by making ϕ sufficiently small and large, respectively. In summary, we have for the design corner points,

Proposition 2:

(i) $\bar{K}_2^{(1)}(\phi)$ *is monotonic, non-decreasing in* ϕ *with range* $[0, \infty]$.

(ii) $\bar{K}_1^{(2)}(\phi)$ *is monotonic, non-increasing in* ϕ *with range* $[0, \infty]$. □

The above behavior appears at first to be counter-intuitive. On further reflection it is understandable that as ϕ increases, $\bar{K}_2^{(1)}(\phi)$ also increases, since increasing ϕ, while reducing the relative weight of each class 2 source, does increase the priority and isolation accorded to the single class 1 source, and hence an increased number of class 2 sources is consistent with preservation of QoS of the single class 1 source. When ϕ is small and decreasing, the number of class 2 sources needs to be increasingly constrained to prevent bandwidth starvation for the single class 1 source, hence $\bar{K}_2^{(1)}(\phi)$ approaches 0.

4.3. Critical Weights

We introduce two special values of the weight ϕ, $\phi_c^{(1)}$ and $\phi_c^{(2)}$, which we call *critical weights*, that will play important roles in the subsequent development. The critical weight $\phi_c^{(j)}$ is that value of ϕ for which the QoS of a single active class j source is satisfied when the number of sources of the complementary class j_c is $\bar{K}_{j_c}^{(j_c)}$, and there is only one source of class j. For example, $\phi_c^{(1)}$ is obtained from the condition

$$\bar{K}_2^{(1)}(\phi_c^{(1)}) = \bar{K}_2^{(2)}\,, \tag{4.10}$$

where it is assumed that $\bar{K}_2^{(2)}$ has been obtained previously. Equivalently $\phi_c^{(1)}$ may be obtained from (4.8) thus,

$$\Pr\left[\sum_{i=1}^{\bar{K}_2^{(2)}} \xi_i^{(2)} > \phi_c^{(1)}(C/e_0^{(1)} - 1)\right] \lesssim L_1\,. \tag{4.11}$$

In general, the defining relation for $\phi_c^{(j)}$ is

$$\bar{K}_{j_c}^{(j)}(\phi_c^{(j)}) = \bar{K}_{j_c}^{(j_c)} \quad (j = 1, 2)\,. \tag{4.12}$$

Using the Chernoff approximation for a single class given in Fact 3, Section 2.3, it is straightforward to calculate the critical weights from (4.11).

The existence of the critical weights follow from Proposition 2. In the case of $\phi_c^{(1)}$, for instance, we know from (4.8) and Proposition 2 that as ϕ is increased from a small value, $\bar{K}_2^{(1)}(\phi)$ crosses the given level $\bar{K}_2^{(2)}$ from below and the value of ϕ at the crossing is $\phi_c^{(1)}$.

Note from the monotonicity properties in Proposition 2 that, if $\phi \geq \phi_c^{(1)}$ then $\bar{K}_2^{(1)} \geq \bar{K}_2^{(2)}$ and for ϕ sufficiently large the latter inequality is strict. Similarly, if $\phi \geq \phi_c^{(2)}$ then $\bar{K}_1^{(2)}(\phi) \leq \bar{K}_1^{(1)}$ and for ϕ sufficiently small the latter inequality is strict.

4.4. Admissible and Realizable Sets

These important sets highlight the role of the weights. Let $\mathcal{A}^{(j)}(\phi)$ denote the set of sources $\mathbf{K}$ which are admissible from the perspective of QoS for class j sources for given ϕ. That is,

$$\mathcal{A}^{(1)}(\phi) \triangleq \{\mathbf{K} : (4.3), \text{i.e. QoS for class 1, is satisfied, given } \phi\} \tag{4.13}$$

$$\mathcal{A}^{(2)}(\phi) \triangleq \{\mathbf{K} : (4.4), \text{i.e. QoS for class 2, is satisfied, given } \phi\} \tag{4.14}$$

The *admissible set* $\mathcal{A}(\phi)$ is defined thus:

$$\begin{aligned}\mathcal{A}(\phi) &\triangleq \mathcal{A}^{(1)}(\phi) \cap \mathcal{A}^{(2)}(\phi) \\ &= \{\mathbf{K} : (4.3) \text{ and } (4.4), \text{i.e. QoS for classes 1 and 2 are satisfied, given } \phi\}\end{aligned} \tag{4.15}$$

Finally, the *realizable set* $\mathcal{R}$ is defined thus:

$$\mathcal{R} \triangleq \bigcup_{\phi} \mathcal{A}(\phi) = \{\mathbf{K} : \exists \phi \text{ such that } \mathbf{K} \in \mathcal{A}(\phi)\} \ . \tag{4.16}$$

We let the boundary of each of the above sets be denoted by ∂, so that $\partial\mathcal{A}^{(1)}(\phi)$ denotes the boundary of $\mathcal{A}^{(1)}(\phi)$.

The following result gives a procedure for computing the realizable set $\mathcal{R}$.

Proposition 3: *The boundary $\partial\mathcal{R}$ parameterized by K_2 is obtained by solving for K_1 and ϕ from (4.3) and (4.4) in which the outer inequalities are replaced by equalities.* $\square$

The proof is simple and by contradiction. Suppose for fixed K_2, (K_1, ϕ) are obtained by the above procedure, and (K_1', ϕ') are such that equality holds in (4.3) and strict inequality holds in (4.4), and also $K_1' > K_1$. By comparing the two expressions corresponding to (4.4) it may be inferred that $\phi' < \phi$. This ordering when used in the two equations corresponding to (4.3) immediately gives a contradiction. To complete the proof we need to also consider the possibility that (K_1', ϕ') are such that strict inequality and equality hold in (4.3) and (4.4), respectively. A contradiction by a similar argument follows.

4.5. Linear Approximations to the Admissible Sets

The boundaries of the admissible regions $\mathcal{A}(\phi)$ defined above are typically nonlinear. However, our experience has been that typically good approximations are provided by simplexes, i.e., regions with boundaries composed of linear segments. Our design methodology is based on linear approximations of the admissible regions. Such approximations are completely determined by a small number (2–4) of corner points. We let $\mathcal{L}^{(j)}(\phi)$ be the linear approximation to $\mathcal{A}^{(j)}(\phi)$ ($j = 1, 2$) and $\mathcal{L}(\phi)$ be the linear approximation to $\mathcal{A}(\phi)$. Also, $\mathcal{R}_\ell$ is the approximation to the realizable set $\mathcal{R}$ obtained from the linear approximations to the admissible set. However $\mathcal{R}_\ell$ itself is not a simplex in general. In each of two cases to be described below, linear approximations to the realizable sets are also developed, but we defer further descriptions of these entities.

We construct the linear approximation $\mathcal{L}^{(j)}(\phi)$ by making its corner points be equal to the corner points of $\mathcal{A}^{(j)}(\phi)$, and then joining these points by linear segments. For instance, the boundaries of $\mathcal{L}^{(1)}(\phi)$ and $\mathcal{A}^{(1)}(\phi)$ have in common the corner points $(\bar{K}_1^{(1)}, 0)$, $(1, \bar{K}_2^{(1)}(\phi))$ and $(0, \infty)$. Similarly, the boundaries of $\mathcal{L}^{(2)}(\phi)$ and $\mathcal{A}^{(2)}(\phi)$ have $(0, \bar{K}_2^{(2)})$, $(\bar{K}_1^{(2)}(\phi), 1)$ and $(\infty, 0)$. (In our sketches we do not attempt to represent $(0, \infty)$ and $(\infty, 0)$.) The boundary of $\mathcal{L}^{(j)}(\phi)$, $\partial\mathcal{L}^{(j)}(\phi)$, is composed of two linear segments connecting the three aforementioned corner points, although we depict only one segment in the sketches. Finally,

$$\mathcal{L}(\phi) = \mathcal{L}^{(1)}(\phi) \cap \mathcal{L}^{(2)}(\phi) \tag{4.17}$$

The approximation to the realizable set based on the linear approximations to the admissible sets is $\mathcal{R}_\ell$, where

$$\mathcal{R}_\ell = \bigcup_\phi \mathcal{L}(\phi) = \{\mathbf{K} : \exists\phi \text{ such that } \mathbf{K} \in \mathcal{L}(\phi)\}\ . \tag{4.18}$$

The reader may examine sketches of the linear approximations in Figure 4.1.

4.6. Qualitative Discussion of the Influence of the Weights on the Admissible Set

When ϕ is large, i.e. the relative weight given to class 1 sources is large, an examination of (4.9) shows that $\bar{K}_1^{(2)}(\phi)$ is small. That is, in the presence of a single active class 2 source and with the QoS of this source at issue, the number of admissible class 1 sources is small, somewhat paradoxically. Moreover, this quantity increases as ϕ decreases. See Figure 4.1.

Why? Because, from the perspective of QoS for the class 2 source, the system is in effect hard-partitioned with only a small allocation of aggregate bandwidth to class 1 sources. That is, to satisfy this QoS, the unbalanced choice of weights has to be compensated by an oppositely unbalanced allocation of bandwidth. Hence, any expectation that very large relative weight given to class 1 sources will allow large numbers of admissible class 1 sources is thwarted by considerations of QoS for class 2.

When ϕ is large, decreasing ϕ, i.e., decreasing the relative weight given to class 1 sources, has the effect of increasing the number of admissible class 1 sources when there is an active

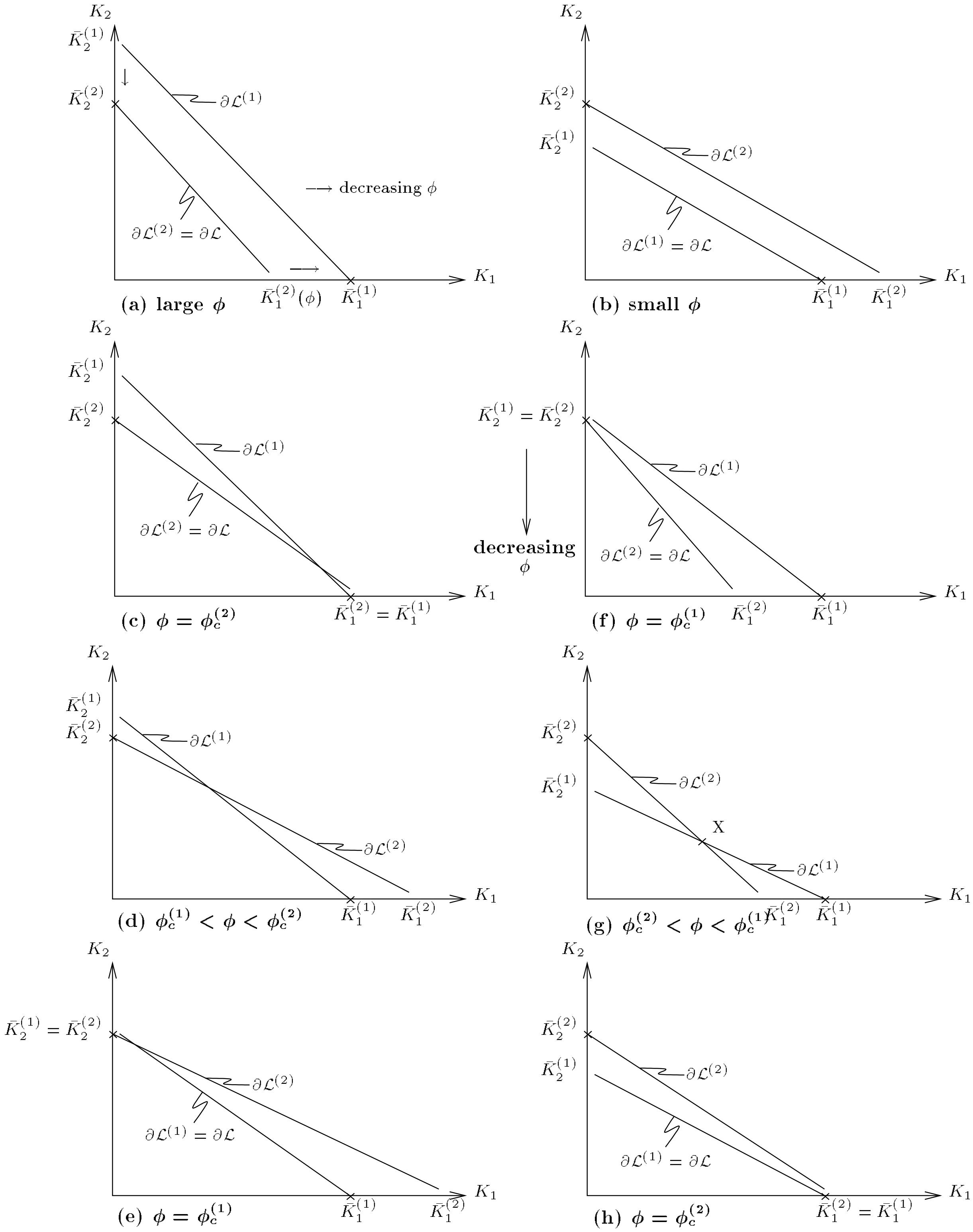

Figure 4.1: Influence of ϕ on admissible regions.

Effectively Homogeneous case, $\phi_c^{(1)} \leq \phi_c^{(2)}$: (c)–(e)

Effectively Non-Homogeneous case, $\phi_c^{(2)} < \phi_c^{(1)}$: (f)–(h) single class 2 source. The explanation is contained in the above discussion. First observe in Fig. 4.1, (a), that for ϕ large the boundary of the admissible region contains the points $(\bar{K}_1^{(2)}(\phi), 1)$ and $(\bar{K}_1^{(1)}, 0)$, and also $\bar{K}_1^{(2)}(\phi) \ll \bar{K}_1^{(1)}$. Hence the number of admissible class 1 sources in the presence of a single active class 2 source is $\bar{K}_1^{(2)}(\phi)$. As is clear from (4.9) and the above discussion, $\bar{K}_1^{(2)}(\phi)$ increases as ϕ decreases.

Note that in this case the misallocation of weights leads to a small number of class 1 sources benefiting by receiving QoS well in excess of requirements. Hence excessively large ϕ simultaneously inhibits sharing as well as misallocates resources.

We conclude by drawing attention to the feature in Fig. 4.1, (a), which follows from the symmetrical implication on $\bar{K}_2^{(1)}(\phi)$ of the argument given above. From the perspective of QoS of a single class 1 source, the number of admissible class 2 sources is large, for large ϕ, see (4.8). The bandwidth partitioning is such that the aggregate allocation to class 2 is large. Hence $\bar{K}_2^{(1)}(\phi)$ is large, and the excess over $\bar{K}_2^{(2)}$ which QoS for class 2 dictates, is a measure of the inefficiency due to misallocation of weights.

4.7. The "Effectively Homogeneous" Case

The justification for the above label will become apparent later and it will also become clear that this is a benign case from the design perspective. The case is depicted in Figure 4.1, (c)–(e), and it arises if any of the following conditions hold.

Proposition 4: *The following conditions are equivalent*

(i)

$$\phi_c^{(1)} \le \phi_c^{(2)} \tag{4.19}$$

(ii) *For some ϕ,*

$$\bar{K}_1^{(2)}(\phi) = \bar{K}_1^{(1)} \textit{ and } \bar{K}_2^{(2)} \le \bar{K}_2^{(1)}(\phi)\,. \tag{4.20}$$

(iii) *For some ϕ,*

$$\bar{K}_2^{(1)}(\phi) = \bar{K}_2^{(2)} \textit{ and } \bar{K}_1^{(1)} \le \bar{K}_1^{(2)}(\phi)\,. \tag{4.21}$$

□

Proof: Note that in (4.20), $\phi = \phi_c^{(2)}$, and in (4.21), $\phi = \phi_c^{(1)}$. This follows from (4.12). If (4.20) holds then $\bar{K}_2^{(2)} \le \bar{K}_2^{(1)}(\phi_c^{(2)})$, and since $\bar{K}_2^{(1)}(\phi)$ is monotonic, non-increasing with decreasing ϕ, as stated in Proposition 2, (4.19) follows. If (4.19) holds then from the above property of $\bar{K}_2^{(1)}(\phi)$ and the monotonic non-increasing behavior of $\bar{K}_1^{(2)}(\phi)$ with increasing ϕ, (4.21) follows. By a similar argument (4.20) follows from (4.21). □

From the key monotonicity properties of $\bar{K}_2^{(1)}(\phi)$ and $\bar{K}_1^{(2)}(\phi)$, which have been summarized in Proposition 2, the important fact follows that the linear approximation to the admissible set $\mathcal{L}(\phi)$ for $\phi < \phi_c^{(2)}$ is dominated by $\mathcal{L}(\phi_c^{(2)})$. That is, as shown in Figure 4.1, (a) and (c),

$$\mathcal{L}(\phi) \subseteq \mathcal{L}(\phi_c^{(2)}) \text{ for all } \phi > \phi_c^{(2)} . \tag{4.22}$$

Similarly (see Figure 4.1, (b) and (c)),

$$\mathcal{L}(\phi) \subseteq \mathcal{L}(\phi_c^{(1)}) \text{ for all } \phi < \phi_c^{(1)} . \tag{4.23}$$

Hence we have

Proposition 5: *To maximize the linear approximation to the admissible set, it is disadvantageous to select ϕ outside the interval $[\phi_c^{(1)}, \phi_c^{(2)}]$ in the effectively homogeneous case.* □

Next observe that in this case, for all ϕ in the interval $[\phi_c^{(1)}, \phi_c^{(2)}]$, the boundaries of $\mathcal{A}(\phi)$ and $\mathcal{L}(\phi)$ have the important and desirable property of containing the extreme points $(\bar{K}_1^{(1)}, 0)$ and $(0, \bar{K}_2^{(2)})$. Now let $\mathcal{L}_H$ denote the triangular region with the origin $(0,0)$ and these extreme points as vertices. From the construction of $\mathcal{L}(\phi)$ for $\phi \in [\phi_c^{(1)}, \phi_c^{(2)}]$ it should be clear that $\mathcal{L}_H$ is a conservative, but generally close, approximation to the union of $\mathcal{L}(\phi)$. That is,

$$\mathcal{L}_H \subseteq \bigcup_{\phi \in [\phi_c^{(1)}, \phi_c^{(2)}]} \mathcal{L}(\phi) . \tag{4.24}$$

From Proposition 5, the right hand set is $\mathcal{R}_\ell$. Hence, $\mathcal{L}_H$ is a conservative approximation to $\mathcal{R}_\ell$.

In summary, the set $\mathcal{L}_H$ has the important property of containing the extreme points $(K_1^{(1)}, 0)$ and $(0, \bar{K}_2^{(2)})$, and, since it is a triangle, it is simple, and also a conservative approximation to the realizable set in the effectively homogeneous case.

4.8. The "Effectively Non-Homogeneous" Case

This case arises when

$$\phi_c^{(2)} < \phi_c^{(1)} \tag{4.25}$$

The behavior of the admissible set for various values of ϕ in this case is depicted in Figure 4.1, (f)–(h). It follows from Proposition 4 that in this case

$$\bar{K}_1^{(2)}(\phi) = \bar{K}_1^{(1)} \Rightarrow \bar{K}_2^{(2)} > \bar{K}_2^{(1)}(\phi) \tag{4.26}$$

and

$$\bar{K}_2^{(1)}(\phi) = \bar{K}_2^{(2)} \Rightarrow \bar{K}_1^{(1)} > \bar{K}_1^{(2)}(\phi)\,. \tag{4.27}$$

One implication of these observations is that, unlike the effectively homogeneous case, the triangular region $\mathcal{L}_H$ cannot be realized by any fixed ϕ and, also, not even by a range of ϕ. This is a handicap which is addressed later when design procedures are considered.

Observe a similarity with the effectively homogeneous case in that it is also disadvantageous here to select ϕ outside the interval $[\phi_c^{(2)}, \phi_c^{(1)}]$. This is because

$$\mathcal{L}(\phi) \subseteq \mathcal{L}(\phi_c^{(1)}) \quad \text{for all } \phi > \phi_c^{(1)} \tag{4.28}$$

and

$$\mathcal{L}(\phi) \subseteq \mathcal{L}(\phi_c^{(2)}) \quad \text{for all } \phi > \phi_c^{(2)} \tag{4.29}$$

from the monotonicity properties in Proposition 2.

Next, we give for the benefit of the reader the corner points of the simplex $\mathcal{L}(\phi)$ for $\phi \in [\phi_c^{(2)}, \phi_c^{(1)}]$. See Figure 4.1, (g) for a sketch. These points are $(0, \bar{K}_2^{(2)})$, $(1, \bar{K}_2^{(1)}(\phi))$, $(\bar{K}_1^{(2)}(\phi), 1)$ and $(\bar{K}_1^{(1)}, 0)$. At the two extreme points of the interval for ϕ, i.e. at $\phi = \phi_c^{(2)}$ and $\phi = \phi_c^{(1)}$, the above four corner points reduce to three.

We define now a set $\mathcal{L}_{NH}$ which plays a role in the effectively non-homogeneous case comparable to $\mathcal{L}_H$ in the effectively homogeneous case.

$$\mathcal{L}_{NH} \triangleq \mathcal{L}(\phi_c^{(1)}) \cup \mathcal{L}(\phi_c^{(2)})\,. \tag{4.30}$$

In Figure 4.2 this region is the simplex with corner points 0, A_2, C and A_1. Note that $\mathcal{L}_{NH}$ is a conservative approximation to the realizable set $\mathcal{R}_\ell$, i.e., $\mathcal{L}_{NH} \subseteq \mathcal{R}_\ell$. An intuitive appreciation of the difference between $\mathcal{L}_{NH}$ and $\mathcal{R}_\ell$ may be obtained by superimposing on Figure 4.2 the locus of points $X(\phi)$, $\phi_c^{(2)} \le \phi \le \phi_c^{(1)}$, where $X(\phi)$, shown in Figure 4.1, (g), is the point of intersection of $\partial\mathcal{L}^{(1)}(\phi)$ and $\partial\mathcal{L}^{(2)}(\phi)$.

4.9. Proposed Design Procedures

The goal of the design procedures which we propose is to realize, by simple techniques for selecting ϕ, the set $\mathcal{L}_H$ in the effectively homogeneous case and the set $\mathcal{L}_{NH}$ in the

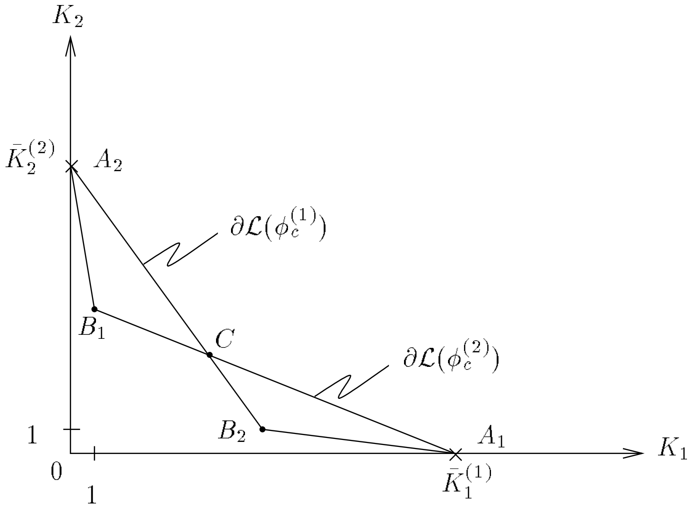

Figure 4.2: The region $\mathcal{L}_{NH}$ in the Effectively Non-Homogeneous case.

effectively non-homogeneous case. These sets are approximations to the realizable set $\mathcal{R}$. Our experience has been that $\mathcal{L}_H$ more closely approximates the realizable set than $\mathcal{L}_{NH}$. However even in the latter case the gap between $\mathcal{R}$ and $\mathcal{L}_{NH}$ is necessarily small and restricted to the interior since the extremal corner points $(\bar{K}_1^{(1)}, 0)$ and $(0, \bar{K}_2^{(2)})$ are common to both sets.

In the effectively homogeneous case the design procedure is straightforward since any fixed ϕ in the interval $[\phi_c^{(1)}, \phi_c^{(2)}]$ will give $\mathcal{L}_H$ as its admissible set. A refinement to this procedure is to select $\phi = (\phi_c^{(1)} + \phi_c^{(2)})/2$, which, because of the convexity of $\mathcal{L}(\phi)$, see Figure 4.1, (d), will yield an admissible set that dominates $\mathcal{L}_H$ and captures most of the gap between $\mathcal{L}_H$ and $\mathcal{R}$.

Clearly the design procedure is more complicated for the effectively non-homogeneous case since no single ϕ will realize all of $\mathcal{L}_{NH}$. Our procedure in this case is to realize $\mathcal{L}_{NH}$ with only two values of ϕ, namely, $\phi_c^{(1)}$ and $\phi_c^{(2)}$. Say the desired operating point (K_1, K_2) is in the simplex in Figure 4.2 with corner points 0, A_2, B_1, and A_1, which we denote by $S(0, A_2, B_1, A_1)$. In this case we use $\phi = \phi_c^{(2)}$. Now suppose that the desired operating point moves into the triangular region $S(A_2, B_1, C)$. We propose switching ϕ to $\phi_c^{(1)}$. Alternatively we may realize $S(0, A_1, B_2, A_2)$ by $\phi = \phi_c^{(1)}$ and realize the complementary $S(A_1, B_2, C)$ by $\phi = \phi_c^{(2)}$.

Besides the fact that two, rather than one, weights are required in the design procedure to handle the non-homogeneity, another difference is that the bulk of the gap between $\mathcal{R}$

	Case 1		Case 2		Case 3	
	class 1	class 2	class 1	class 2	class 1	class 2
Q (cells)	250	250	25	250	100	1000
P (Mbps)	1.5	1.5	1.5	1.5	0.032	1.5
r (Mbps)	0.15	0.15	0.15	0.15	0.016	0.15
L	10^{-3}	10^{-9}	10^{-3}	10^{-9}	10^{-3}	10^{-9}
D (sec)	0.001	0.6	0.01	0.01	0.01	0.6
$e_0^{(\cdot)}$ (Mbps)	1.48	0.158	0.62	1.31	0.0318	0.48

Table 5.1: Traffic and QoS parameters for the three cases in Figures 5.1, 5.2 and 5.3.

and $\mathcal{L}_{NH}$ cannot be captured by any simple strategy, i.e., a large number of values of ϕ with correspondingly large number of switching of weights is necessary. This is impractical, and, moreover, as has been observed above, not worth the effort since the gap cannot be substantial.

5. Numerical Results

The investigations reported in this section numerically illustrate the mathematical results of prior sections on statistical multiplexing, GPS design and call admission control. Specifically, the set of figures address the following issues: (i) the advantage of exploiting statistical multiplexing in GPS design in terms of capacity gain and comparison with traditional designs which are based on deterministic QoS guarantees; (ii) the admissible region as a function of GPS weights; (iii) optimum design of weights and the realizable admissible region; (iv) application to voice and data support.

In each of the three cases we consider there are two QoS source classes sharing a total bandwidth of 45 Mbps (this is the link bandwidth minus the bandwidth allocated to the best effort traffic.) Table 5.1 gives the dual leaky bucket and QoS parameters associated with each class in each case, as well as the lossless effective bandwidth $e_0^{(\cdot)}$. Case 1 is considered in Figure 5.1, where the two QoS classes have the same dual leaky bucket parameters, but different maximum delay D and maximum violation, i.e. loss, probability L. The two critical (ratios of) weights $\phi_c^{(1)}$ and $\phi_c^{(2)}$ in this case are found to be 6.4 and 9.2. The boundaries of the admissible regions obtained using these values are plotted. Note that these two source classes are effectively non-homogeneous, i.e. no single weight is sufficient for all traffic mixes. The boundary of the realizable region, which is obtained by varying ϕ, the ratio of weights, between its two critical values, is also plotted. As in Section 4.9, we observe that the realizable region may be approximately achieved by using only the two critical weights and switching between them on a slow time scale as the desired operating point in connections changes in correspondence to changes in traffic mix. Observe in Table 5.1 that $e_0^{(\cdot)}$ is close to the peak rate for class 1 and to the mean rate for class 2. Also observe that, not surprisingly, the gain in capacity from statistical multiplexing is more pronounced for class 1.

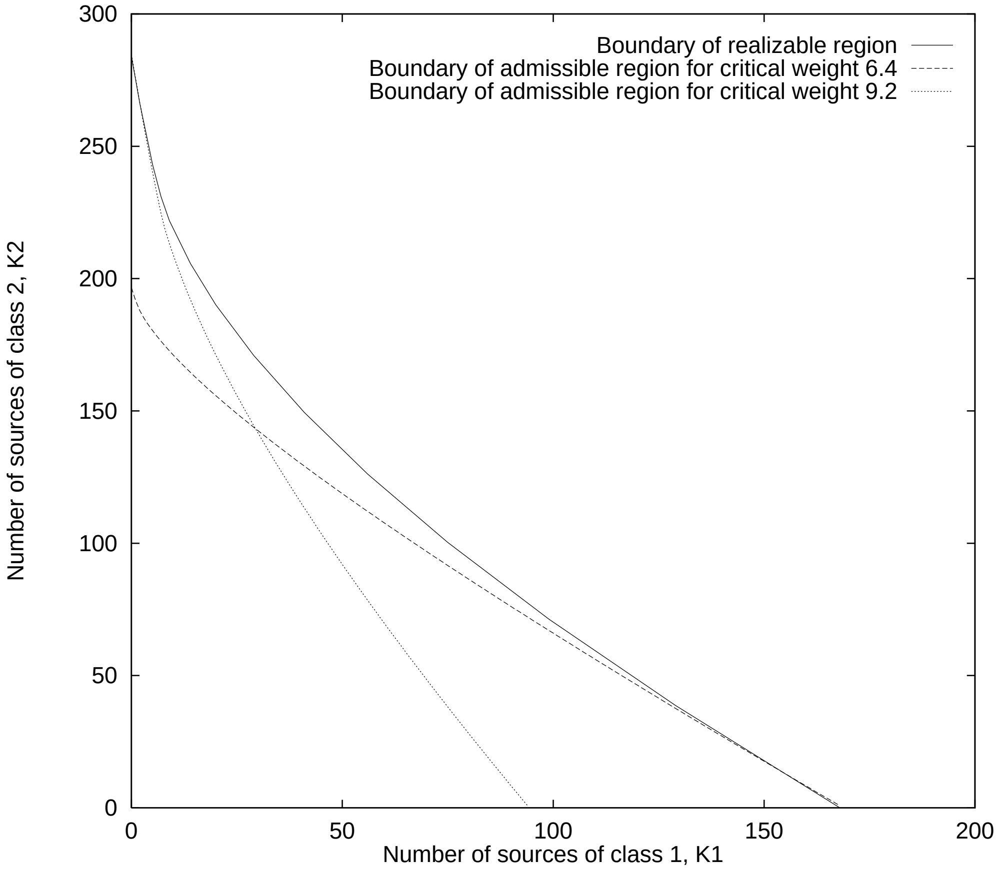

Figure 5.1: The admissible region corresponding to the critical weights, $\phi_c^{(1)}$ and $\phi_c^{(2)}$, and the realizable region for Case 1.

The results for Case 2 are shown in Figure 5.2, where the sources differ in their dual leaky bucket parameters and QoS requirements. We determine the critical weights to be 0.372 and 0.307, and find that the two corresponding admissible regions are equivalent, which implies that the two classes are effectively homogeneous. The realizable region in this case is, as expected, only slightly larger than that obtained with either fixed critical weight. We also plot the admissible region for the case of deterministic QoS guarantees, i.e., with $L_j = 0$ ($j = 1, 2$) and all other parameters unchanged. The advantage of statistical multiplexing is clearly demonstrated in this figure; there is a factor of 3 in multiplexing gain.

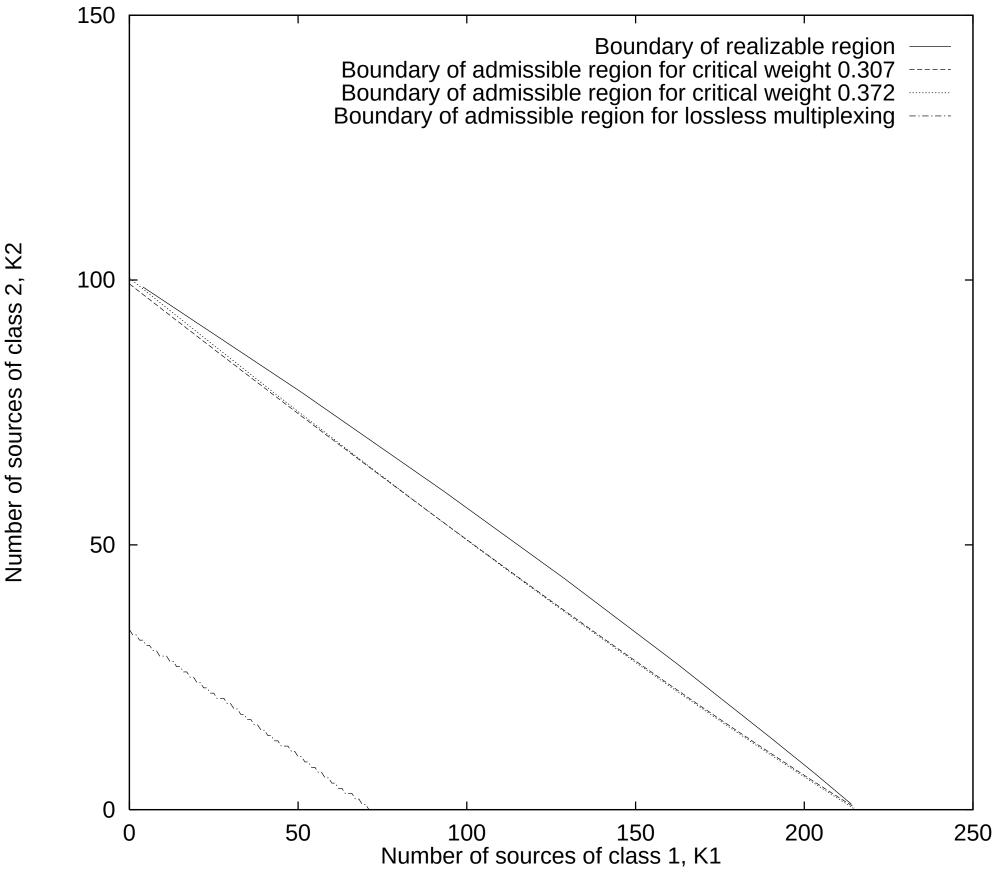

Figure 5.2: The admissible region corresponding to the critical weights, $\phi_c^{(1)}$ and $\phi_c^{(2)}$, and the realizable regions for lossless and statistical multiplexing for Case 2.

Figure 5.3 shows the results for Case 3 where we consider the support of real time and non-real time traffic classes. Class 1 is voice coded at 32Kb/s and employing silence-detection; its dual leaky bucket parameters are appropriately chosen, and its nodal maximum delay is 10 msec. Class 2 models LAN traffic on a T1 link, which is delay-tolerant but has stringent loss requirement. The critical weights are 0.0527 and 0.0628. Note that the two classes, while quite different in their dual leaky bucket and QoS parameters, are almost effectively homogeneous, i.e. any one of the critical weights provides almost all of the realizable region.

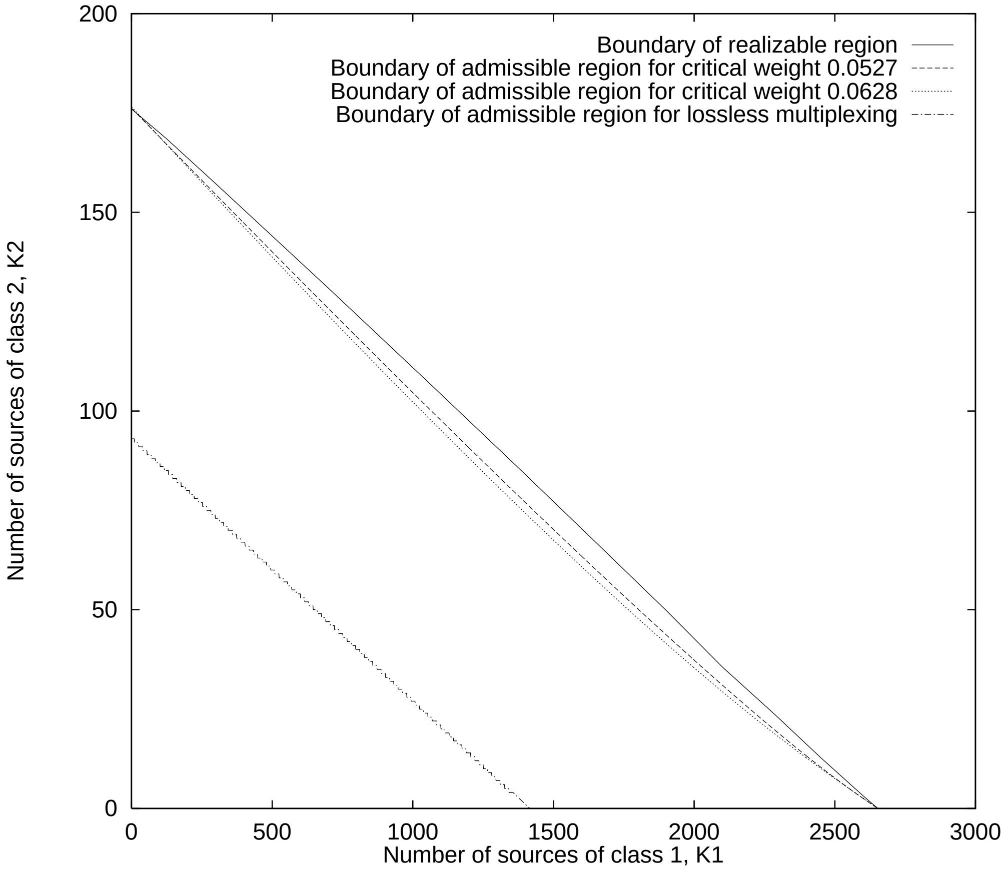

Figure 5.3: The admissible region corresponding to the critical weights, $\phi_c^{(1)}$ and $\phi_c^{(2)}$, and the realizable regions for lossless and statistical multiplexing for Case 3.

6. End-to-end Heterogeneous Statistical QoS Guarantees with GPS Service

In this section we consider the problem of supporting two QoS service classes and a best-effort class in a network-wide setting. Our solution approach is based on results from the single node analysis and the following two ideas: (i) The best-effort class is allocated part of the resources in each node for its exclusive use. The best-effort traffic typically consumes these resources; the unused resources, if any, may be too highly variable to be of use for the QoS classes. (ii) The output of each QoS source queue in each node is regulated, where the peak rate is not allowed to exceed a certain value – this value is not necessarily equal to the peak input rate; in fact it is typically smaller to account for the buffering and multiplexing at the node. In our work we limit the peak output rate to e_0, which is the lossless effective bandwidth defined in (3.7). The incorporation of output regulation in GPS is a novel feature, and the motivation behind it is twofold. First, since the admitted QoS sources are allocated resources to meet their contractual requirements, there is no need to allow them access to more resources; excess resources are made available to the best-effort class. Second, output regulation leads to a manageable and more accurate end-to-end resource management, since the output of each queue, which is the input to the queue in the next node, can be tightly characterized. As a consequence, in output regulated GPS our prior single node analysis applies to all nodes in the network. The total bandwidth available to best-effort traffic is a function of the number of admitted QoS sources and the activity of each source; it will be characterized in more detail below.

We now outline a scheme for analyzing end-to-end resource management of virtual connections, which are either originating at the node or transit (recall that in prior sections we used the term sources to describe all connections.) The bandwidth demand of each connection on each node is characterized by an appropriate pair (r, e_0), where r is the bound on the mean flow rate and e_0 is defined in (3.7). Thus, the single node CAC and GPS designs of Sections 3 and 4 apply. We assume that the end-to-end QoS requirement for class j is specified by (D_j^T, L_j^T), where

$$\Pr((\text{Total\_delay})_j > D_j^T) \leq L_j^T \qquad j = 1, 2$$

and $(\text{Total\_delay})_j$ is the maximum end-to-end delay. As before, each node implements output regulated GPS and buffering per connection. In each node, each originating connection of class j is characterized by the dual leaky bucket parameters $(r^{(j)}, Q^{(j)}, P^{(j)})$.

Let $(\text{Delay})_{ji}$ denote the maximum delay for class j at node i, so that $\sum_i (\text{Delay})_{ji} = (\text{Total\_delay})_j$, $(j = 1, 2)$, and let the QoS requirement for class j at node i be

$$\Pr((\text{Delay})_{ji} > D_{ji}) \leq L_j \qquad j = 1, 2;\ i = 1, \cdots, N \tag{6.1}$$

where D_{ji} is the maximum allocated delay for class j at node i. For instance, L_j may be set equal to L_j^T / N_j, where N_j is the maximum number of hops of all connections of class j. For typical applications where L_j^T is small and N_j is not large this approximation is conservative but reasonable.

We first consider the case where the entire delay budget for each class is allocated to the first node in the path of the connection, i.e., $D_{j1} = D_j^T$ and $D_{ji} = 0$ for $i > 1$. The motivation behind this is that buffering at the network edges may be cheaper and the traffic smoothing thus obtained alleviates congestion at intermediate and potential bottleneck nodes. In this case, specifically on account of output regulated GPS, the CAC and GPS design in the edge and all intermediate nodes are identical, since the demand of all connections of class j is described by a common $(r^{(j)}, e_0^{(j)})$. Note that the actual mean rate of the traffic at an intermediate node may be less than $r^{(j)}$ due to possible loss at the node; however, the probablity of loss due to full buffer is, by design, very small.

Next consider the case of the general allocation of the end-to-end delay budget among the nodes in the path of the connection. In this case, the CAC, GPS design and the value of the output rate are obtained from the prior single node analysis as if the incoming transit traffic is regulated by a dual leaky bucket with parameters $(r^{(\cdot)}, D_{on}^{(\cdot)} e_0^{(\cdot)}, e_0^{(\cdot)})$, where $e_0^{(\cdot)}$ is the peak output rate from the preceding (upstream) node, which together with $D_{on}^{(\cdot)}$, is obtained from the (lossless multiplexing) analysis of that node. Note that a consequence of the general allocation in the delay budget is a proliferation of classes, leading to considerable computational complexity, even though the solution framework remains unchanged. An alternative view is that the combination of edge-only buffering and the approximation of uniform losses across nodes in each class has the virtue of containing the explosion in complexity.

6.1. Characterization of the Best-Effort Channel

As pointed out above, for each node in the network, the bandwidth available to the best-effort traffic consists of a fixed provisioned component and a variable component which is a function of the carried QoS traffic. The variable bandwidth component has large fluctuations at the time-scale of QoS connections, i.e. the time-scale of connection arrivals and departures, and small superimposed fluctuations at the burst time-scale, i.e. the time-scale of on-off dynamics of the admitted connections. The bandwidth available to the best effort traffic, C_{be}, may be characterized by a modulated Brownian motion. The modulating process captures the slow-time scale component and models the number of admitted calls of each QoS source type and their dynamics (arrivals and departures.) For a given state of the modulating process $\mathbf{K} = (K_1, K_2, \ldots)$, where K_j is the number of admitted connections of QoS class j, the Brownian motion captures the fast time-scale component. Specifically, the mean and the variance of $C_{be}(\mathbf{K})$ are given as:

$$\mathrm{Mean}(C_{be}(\mathbf{K})) = C - \sum_j K_j r^{(j)}, \tag{6.2}$$

where C is the nodal bandwidth, and

$$\mathrm{Var}(C_{be}(\mathbf{K})) = \sum_j K_j (e_0^{(j)})^2 \omega^{(j)} (1 - \omega^{(j)}) \tag{6.3}$$

Recall that $r^{(j)} = e_0^{(j)} \omega^{(j)}$. In the simplest connection model, arrivals are Poisson and holding periods are exponentially distributed, with rates which may depend on the class. For such models, the modulating connection process has a steady state distribution which is product-form and easily computed. We remark that the modulated Brownian motion characterization of the best-effort channel is appropriate to the general expectations of relatively higher loss, which justifies the diffusion, rather than the large deviations asymptotic, regime. Techniques for analysis of such modulated Brownian motion may be found in [KKU95, ASS95].

7. Conclusions

In this paper we have considered the analysis and design of GPS schedulers under statistical QoS guarantees. While the basic results are for a single node, the general setting considered is that of a network of GPS schedulers supporting two QoS classes and a best effort class. The connections are regulated by dual leaky buckets and the output of each connection buffer is peak rate-controlled. The latter feature makes the bandwidth in excess of what is need to satisfy the demand of the QoS sources available to the best effort traffic; it also facilitates the end-to-end analysis since the output of each connection queue can be tightly characterized. We have derived the realizable admissible connection region, which is the largest possible admissible region obtained by varying the weights. We have derived the notion of critical weights, and shown that the realizable region can be found by varying the weights only between the two critical weights. We have shown further that the realizable region can be closely approximated by using one of the critical weights, or by switching between the two only once as the operating point shifts due to changes in the traffic mix. For the best effort traffic we have developed a characterization of its share of the link bandwidth which should be useful in its analysis and engineering. The validity of our overall analytical and design results as well as the considerable advantage of using statistical QoS guarantees over deterministic guarantees has been demonstrated by numerical examples.

This work can be extended in a number of interesting directions. The analysis may be extended to the general case of more than two source classes. Multiplicity of source classes in our framework arises not only from multiplicity of QoS requirements, but also from the existence of multiple traffic classes (defined by different dual leaky bucket parameters). Algorithms could be devised which systematically derive the realizable region and the critical weights, based on extensions of the results for the two-class case.

Another possible extension is to hierarchical GPS schedulers. These schedulers are attractive since they provide isolation between classes and flexibility in bandwidth allocation. It is also of interest to develop hierarchical analysis of such schedulers, and which may be used as an approximation the the analysis of to the non-hierarchical schedulers.

Another issue which is worth investigating further is how the results are influenced by the coexistence of "statistically multiplexable" and "non-statistically multiplexable" sources, as defined in [EMW95], and in particular, how the linearity of the admissible region is affected.

References

[ASS95] S. Assmussen, "Stationary distribution for fluid flow models with or without Brownian motion", *Stochastic Models* **11**, 1995, pp. 21–49.

[ATM93] ATM Forum, *ATM User-Network Interface Specification, Version* 3.0, Prentice Hall, 1993.

[BIL86] P. Billingsley, *Probability and Measure*, 2nd ed. (Wiley, New York) 1986.

[CRU91] R. L. Cruz, "A calculus for network delay, Part I: Network elements in isolation", *IEEE Trans. Inform Theory* **37**, 1991, pp. 114–131.

[CRU97] R. L. Cruz, "SCED+: Efficient management of quality of service guarantees", *Proc. IEEE INFOCOM'98*, 1998, pp. 625–634.

[CSE93] N. R. Chaganty and J. Sethuraman, "Strong large deviation and local limit theorems", *Ann. Prob.* **21**:(3), 1993, pp. 1671–1690.

[DVK96] G. de Veciana and G. Kesidis, "Bandwidth allocation for multiple qualities of service using generalized processor sharing", *IEEE Trans. Info. Theory* **42**, Jan. 1966, pp. 268–272.

[EMI97] A. Elwalid and D. Mitra, "Traffic shaping at a network node: Theory, optimum design, admission control", *Proc. IEEE INFOCOM'97*, 1997.

[EMW95] A. I. Elwalid, D. Mitra and R. H. Wentworth, "A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node", *IEEE J. Select. Areas Commun.* **13**:(6), 1995, pp. 1115–1127.

[HUI90] J. Y. Hui, "Resource allocation for broadband networks", *IEEE J. Select. Areas Commun.* **6**, 1988, pp. 1598–1608.

[KEL91] F. P. Kelly, "Effective bandwidths at multi-type queues", *Queueing Systems* **9**, 1991, pp. 5–15.

[KES97] S. Keshav, *An Engineering Approach to Computer Networking*, Addison-Wesley, Reading, MA, 1997.

[KES96] G. Kesidis, *ATM Network Performance*, Kluwer, Boston, MA, 1996.

[KKU95] R. L. Karandikar and V. G. Kulkarni, "Second order fluid flow models: reflected Brownian motion in a random environment", *Opns. Res.* **45**, 1995, pp. 77–88.

[LZT96] F. LoPresti, Z.-L. Zhang and D. Towsley, "Bounds, approximations and applications for a two-queue GPS system", *Proc. IEEE INFOCOM'96*, 1996, pp. 1310–1317.

[LZT97] F. LoPresti, Z. Zhang, D. Towsley and J. Kurose, "Source time scale and optimal buffer/bandwidth trade-off for regulated traffic in an ATM node", in *Proc. IEEE INFOCOM'97*, 1997.

[MMO95] D. Mitra and J. A. Morrison, "Multiple time scale regulation and worst case processes for ATM network control", in *Proc.* 34$^{\text{th}}$ *Conf. on Decision and Control*, 1995, pp. 353–358.

[PAR92] A. K. Parekh, "A generalized processor sharing approach to flow control in integrated services network", Ph.D. dissertation, LIDS-TH-2089, MIT, Feb. 1992.

[PET65] V. V. Petrov, "On the probabilities of large deviations for sums of independent random variables", *Theory Probab. Applicat.* **10**, 1965, pp. 287–298.

[PGA93] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control – The single node case", *IEEE/ACM Trans. Networking* **1**, June 1993, pp. 344–357.

[PGA94] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services network – The multiple node case", *IEEE/ACM Trans. Networking*, April 1994, pp. 137–150.

[RRR98(a)] S. Rajagopal, M. Reisslein and K. W. Ross, "Packet multiplexers with adversarial regulated traffic", *Proc. IEEE INFOCOM'98*, 1998, pp. 347–355.

[RRR98(b)] K. W. Ross, M. Reisslein and K. W. Ross, "A pragmatic approach for guaranteeing QoS to regulated traffic in networks", preprint, 1998.

[ROB92] *Performance Evaluation and Design of Multiservice Networks*, Ed. J. W. Roberts, Commission of the European Communities, Final Rep. COST 224 project, 1992.

[SPG97] S. Shanker, C. Partridge and R. Guerin, "Specification of guaranteed quality of service", RFC 2212, IETF, September 1997.

[SVA98(a)] D. Stiliadis and A. Varma, "Rate-proportional servers: A design methodology for fair queueing algorithms", *IEEE/ACM Trans. Networking* **6**, April 1998, pp. 164–174.

[SVA98(b)] D. Stiliadis and A. Varma, "Efficient fair queueing algorithms for packet-switched networks", *IEEE/ACM Trans. Networking* **6**, April 1998, pp. 175–185.

[TUR86] J. S. Turner, "New directions in communications (or which way to the information age?)", *IEEE Commun. Mag.*, Oct. 1986.

[WRO97(a)] J. Wroclawski, "The use of RSVP with IETF integrated services", RFC 2210, IETF, September 1997.

[WRO97(b)] J. Wroclawski, "Specification of the controlled load quality of service", RFC 2211, IETF, September 1997.

[YSI94] D. Yaron and M. Sidi, "Generalized processor sharing networks with exponentially bounded burstiness arrivals", *Proc. IEEE INFOCOM'94*, 1994, pp. 628–634.