

# Application of Network Calculus to Guaranteed Service Networks

Jean-Yves Le Boudec, *Member, IEEE*

**Abstract**—We use recent network calculus results to study some properties of lossless multiplexing as it may be used in guaranteed service networks. We call network calculus a set of results that apply min-plus algebra to packet networks. We provide a simple proof that shaping a traffic stream to conform to a burstiness constraint preserves the original constraints satisfied by the traffic stream. We show how all rate-based packet schedulers can be modeled with a simple rate latency service curve. Then we define a general form of deterministic effective bandwidth and equivalent capacity. We find that call acceptance regions based on deterministic criteria (loss or delay) are convex, in contrast to statistical cases where it is the complement of the region which is convex. We thus find that, in general, the limit of the call acceptance region based on statistical multiplexing when the loss probability target tends to 0 may be strictly larger than the call acceptance region based on lossless multiplexing. Finally, we consider the problem of determining the optimal parameters of a variable bit rate (VBR) connection when it is used as a trunk, or tunnel, given that the input traffic is known. We find that there is an optimal peak rate for the VBR trunk, essentially insensitive to the optimization criteria. For a linear cost function, we find an explicit algorithm for the optimal remaining parameters of the VBR trunk.

**Index Terms**—Arrival curves, asynchronous transfer mode (ATM), effective bandwidth, equivalent capacity, guaranteed quality of service, max-plus algebra, min-plus algebra, network calculus, queueing systems, service curves.

## I. INTRODUCTION

WE call network calculus a set of rules and results that can be used for computing tight bounds on delays, backlogs, and arrival envelopes in a lossless setting applicable to packet networks. Fundamental work has been pioneered by Parekh and Gallager [1], [2] and Cruz [3]–[5], where general bounds based on the concepts of arrival and service curves are derived. Other fundamental work for specific or general scheduling policies is described in [6] and [7]. Recently, this work has been extended and simplified independently and simultaneously under equivalent forms by Sariowan [8] who gives a formal and general treatment of the concepts of arrival and service curves, two fundamental network calculus tools. The mathematics involved in network calculus uses min-plus algebra, as described in [9]. In this paper, we show how these results can be applied to derive general concepts for packet networks with guaranteed service. In Section II we recall

the main network calculus results we need in this paper. In Section III we give results on shapers. These results were found independently in [10] and [11]. We also provide a simple proof that shaping a traffic stream to conform with a burstiness constraint preserves the original constraints satisfied by the traffic stream.

The Internet Engineering Task Force (IETF) is developing a framework for packet networks offering guaranteed quality of service. We show how the network calculus results mentioned above can be applied to derive general concepts for packet networks with guaranteed service. First, in Section IV, we show how rate-based schedulers can be modeled simply in a way that fits with the IETF framework [7], [12], [13].

Secondly, in Section V, we introduce the general concept of deterministic effective bandwidth, which was introduced in a narrower context in [14, pp. 270–273]. We give a simple, general definition, and show that it is a convex function of the arrival curve. This enables us to determine that call acceptance regions based on deterministic delay constraints are convex. This is in contrast to call acceptance regions based on statistical multiplexing with large deviation asymptotics, in which case it is the complement in the positive orthant which is convex [15]. This also shows that, in general, the limit of the call acceptance region based on statistical multiplexing when the loss probability target tends to 0 may be strictly larger than the call acceptance region based on lossless multiplexing. We also define similarly the deterministic equivalent capacity, by analogy to the work in [16].

Third, in Section VI (and this was our initial motivation), we consider a connection admission control (CAC) method, for the case where connections are admitted onto an Asynchronous Transfer Mode (ATM) variable bit rate (VBR) trunk, or tunnel [17]. In such a case, the CAC method can be split into two subproblems: 1) predict the traffic for the next prediction period and 2) given a predicted traffic, find the optimal VBR trunk parameters that can carry the traffic. Point 2) is not as simple as for a conventional, constant bit rate (CBR) trunk, because, for a VBR trunk, there are three parameters to set, instead of one for CBR. It is point 2) that we solve in this paper, using network calculus. The VBR trunk is defined by a peak rate, a sustainable rate, and a burst tolerance. We find that, if a deterministic delay constraint is used, then there is always an optimal peak rate, which is the effective bandwidth of the arrival traffic. If the cost function for the VBR trunk is linear, then we find an explicit algorithm for computing the optimal VBR trunk characteristics.

Manuscript received January 9, 1997; revised September 18, 1997.

The author is with the Laboratoire de Réseaux de Communication, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (<http://lrcwww.epfl.ch>).

Publisher Item Identifier S 0018-9448(98)02345-1.

## II. BACKGROUND: NETWORK CALCULUS

In this section we recall a few definitions and results, which we collectively call “network calculus” [8], [10], [11], [18], [19].

We consider wide-sense increasing functions of time, with nonnegative, possibly infinite values (we say that function  $\gamma(\cdot)$  is wide-sense increasing, also called “nondecreasing,” when  $\gamma(s) \leq \gamma(t)$  for all  $s \leq t$ ). For two wide-sense increasing functions  $\gamma_1$  and  $\gamma_2$ , define  $\gamma_1 \otimes \gamma_2$  by

$$\gamma_1 \otimes \gamma_2(t) = \inf_{u \text{ such that } 0 \leq u \leq t} \{\gamma_1(u) + \gamma_2(t - u)\}. \quad (1)$$

This operation is called “min-plus convolution.” It is the correspondent of standard convolution when we move from standard algebra to min-plus algebra [9]. Min-plus convolution has a number of nice properties: it is associative and commutative; if  $\gamma_1(0) = \gamma_2(0) = 0$  then  $\gamma_1 \otimes \gamma_2 \leq \min(\gamma_1, \gamma_2)$  with equality if  $\gamma_1$  and  $\gamma_2$  are concave; if  $\gamma_1$  and  $\gamma_2$  are convex and piecewise-linear, then  $\gamma_1 \otimes \gamma_2$  is obtained by putting end-to-end the different linear pieces of the individual service curves, sorted by increasing slopes [2].

Consider a data flow, described by its arrival function  $R(t)$ , which is equal to the number of bits seen on the flow in time interval  $[0, t]$ . Given a wide-sense increasing function  $\alpha$ , we say that the flow is constrained by  $\alpha$  if and only if for all  $s \leq t$ :  $R(t) - R(s) \leq \alpha(t - s)$ . This is equivalent to requiring that [10], [11]

$$R \leq R \otimes \alpha. \quad (2)$$

Function  $\alpha$  is called an *arrival curve* for the flow. For example, a flow controlled by a leaky bucket has an arrival curve of the form  $\alpha(t) = b + rt$ . Similarly, an ATM flow, constrained by the GCR algorithm with parameters  $(T, \tau)$  has an arrival curve  $\alpha(t) = B + Pt$ , with  $P$  in bits per second (bits/s) and  $B$  in bits given by  $B = \tau P + \delta$ ,  $P = \delta/T$ . In the formulas,  $\delta$  is the cell size in bits. A flow conforming to the draft IETF specification for integrated service [12], with maximum packet size  $M$ , peak rate  $p$ , sustainable rate  $r$ , and burst tolerance  $b$ , has an arrival curve defined by  $\alpha(t) = \min(M + pt, b + rt)$ . In many references, such as [13], it is the number of bits arriving in time interval  $[s, t]$  (instead of  $(s, t]$ ) for which a bound is required; in that case, a bound is  $\alpha(t - s) + l$ , where  $l$  is the maximum packet size, which is also the maximum number of bits that can arrive instantly.

The function  $\alpha$  may be any nonnegative, wide-sense-increasing function, but it defines a meaningful constraint only if it is subadditive, which means that  $\alpha(s + t) \leq \alpha(s) + \alpha(t)$  for all  $s, t \geq 0$ . If  $\alpha$  is not subadditive, it can be replaced by its subadditive closure [10]. If  $\alpha$  is concave, with  $\alpha(0) = 0$ , then  $\alpha$  is subadditive. As a consequence, all curves defined by the minimum of affine functions are subadditive.

Consider now a system  $\mathcal{S}$ , which we view as a blackbox;  $\mathcal{S}$  takes data in and outputs data after a variable delay. Call  $R(t)$  the input function, namely, the number of bits seen on the input flow in time interval  $[0, t]$ , and  $R^*(t)$  the output function. The *backlog* at time  $t$  is  $R(t) - R^*(t)$ ; it is the amount of bits that are held inside the system. Similarly, the

*virtual delay* at time  $t$  is

$$d(t) = \inf\{T: T \geq 0 \text{ and } R(t) \leq R^*(t + T)\}.$$

It is the delay that would be experienced by a bit arriving at time  $t$  if all bits received before it were served before it. If the output function is continuous (no batch departure) then  $R^*[t + d(t)] = R(t)$ .

Network calculus results give computational rules for bounding virtual delays and backlog for arbitrary systems that represent networks. We first need the definition of *service curve*. We say that  $\mathcal{S}$  offers to the flow a service curve  $\beta$  if and only if [10], [18]

$$R^* \geq R \otimes \beta. \quad (3)$$

This is equivalent in practice to saying that for all  $t \geq 0$ , there exists some  $t_0 \geq 0$ , with  $t_0 \leq t$ , such that

$$R(t) - R^*(t_0) \geq \beta(t - t_0).$$

The definition of a service curve is an abstraction of the “strict service curve” concept defined by Cruz in [5]. A GPS scheduler [1], [2] with rate guarantee  $r$  to a flow offers this flow a service curve  $\beta(t) = rt$ . Note that the universal service curve property in [1] and [2] is a stronger guarantee than (3); it assumes that a service rate equal to at least  $r$  is offered during *any* period of time where the backlog for the flow is not empty.

A delay-based scheduler, which guarantees a delay constraint  $T$ , offers a service curve  $\delta_T$ , where  $\delta_T$  is the impulse function defined by  $\delta_T(t) = 0$  if  $0 \leq t \leq T$  and  $\delta_T(t) = +\infty$  if  $t > T$ .

A first series of results gives tight bounds given arrival and service curve constraints [18].

- Assume a flow, constrained by arrival curve  $\alpha$ , traverses a system that offers a service curve  $\beta$ . The backlog  $R(t) - R^*(t)$  for all  $t$  is bounded by

$$R(t) - R^*(t) \leq \sup_{s \geq 0} \{\alpha(s) - \beta(s)\}.$$

The virtual delay  $d(t)$  is bounded by

$$d(t) \leq h(\alpha, \beta) \quad (4)$$

where  $h(\alpha, \beta)$  is the horizontal deviation between the two functions, namely,

$$h(\alpha, \beta) = \sup_{s \geq 0} [\inf\{T: T \geq 0 \text{ and } \alpha(s) \leq \beta(s + T)\}].$$

- Assume a flow traverses systems  $\mathcal{S}_1$  and  $\mathcal{S}_2$  in sequence. Assume that  $\mathcal{S}_i$  offers a service curve of  $\beta_i$ ,  $i = 1, 2$  to the flow. Then the concatenation of the two systems offers a service curve of  $\beta_1 \otimes \beta_2$  to the flow.

In the next section we give a second family of results which we will need in this paper.

### III. SHAPERS

Most results on shapers were also discovered independently in [8], [10], and [11].

A shaper, with shaping curve  $\sigma$ , is a bit-processing device that forces its output to have  $\sigma$  as arrival curve, by delaying bits in a buffer whenever sending a bit would violate the constraint for the output traffic. A buffered leaky bucket is a shaper with  $\sigma$  of the form  $\sigma(t) = b + rt$ . The curve  $\sigma$  can be any nonnegative, wide-sense-increasing function, but as for any arrival curve, we may assume that it is subadditive (otherwise, replace  $\sigma$  by its subadditive closure). In Section VI, we consider network nodes that act as shapers. We first recall a fundamental result on shapers.

*Theorem 1.* [8], [10], [11], [19]: A shaper with subadditive shaping curve  $\sigma$  such that  $\sigma(0) = 0$  offers  $\sigma$  as a service curve.

The simplest proof for this result is based on the linear characterization of shapers which has been established independently in [8], [10], and [11]. They find that shapers are simple linear systems, in terms of min-plus algebra. Indeed, the output  $R^*$  of a shaper with input  $R$  is given by

$$R^* = R \otimes \sigma \quad (5)$$

provided that  $\sigma$  is subadditive and  $\sigma(0) = 0$  (otherwise, replace  $\sigma$  by its closure [10]). This result uses a general method of max-plus or min-plus algebra [9]. The theorem follows directly from (5). An alternative proof is given in [19], which applies only to the discrete time case and is based on what is called in [20] the time method.

A corollary follows. Consider a case where reshapers are introduced along a path. The shapers act as additional buffers, that could increase the end-to-end delay. However, we can derive easily that they “come for free.”

*Corollary 1. Shaping Does Not Increase the Delay Bound:* Assume a flow, constrained by arrival curve  $\alpha$ , is input to networks  $\mathcal{S}_1$  and  $\mathcal{S}_2$  in sequence. Assume a shaper, with curve  $\sigma$  is added between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ . Assume that  $\alpha \leq \sigma$ . Then the delay bound in (4) for the system without the shaper is also valid for the system with the shaper.

Note that the condition  $\alpha \leq \sigma$  means that the shaper may be less restrictive than the original constraint.

*Proof:* We can assume without loss of generality that  $\sigma$  is subadditive and  $\sigma(0) = 0$ , otherwise, we replace  $\sigma$  by its subadditive closure [10] and the assumptions still hold.

Call  $\beta_i$  the service curve of network  $\mathcal{S}_i$ . From Section II, the delay bound for the system with shaper is  $h(\alpha, \beta_1 \otimes \sigma \otimes \beta_2)$  which, by associativity and commutativity, is also equal to  $h(\alpha, \sigma \otimes \beta_1 \otimes \beta_2)$ . Since the bounds are tight [11], the latter is the worst case delay for the system where the shaper is put immediately after the source; in that system, the shaper introduces a zero delay, therefore,

$$h(\alpha, \sigma \otimes \beta_1 \otimes \beta_2) = h(\alpha, \beta_1 \otimes \beta_2). \quad \square$$

Finally, we can derive the following proposition, which is a generalization of original results in [5] and [19].

#### *Proposition 1. Shaping Conserves Arrival Constraints:*

Assume a flow with arrival curve  $\alpha$  is input to a shaper with shaping curve  $\sigma$ . Then the output of the shaper is still constrained by the original arrival curve  $\alpha$ .

*Proof:* We can assume without loss of generality that  $\sigma$  is subadditive and  $\sigma(0) = 0$ . Call  $R$  the input flow and  $R^*$  the output. By the characterization of arrival curves in [10] recalled in Section II, we have  $R \leq R \otimes \alpha$ . Combining with (5) we have

$$R^* = R \otimes \sigma \leq (R \otimes \alpha) \otimes \sigma = (R \otimes \sigma) \otimes \alpha = R^* \otimes \alpha. \quad \square$$

### IV. MODELING GUARANTEED QUALITY OF SERVICE NODES

In this and the following section we now apply network calculus results to integrated services packet networks. The IETF implicitly uses a generic service curve model; it assumes that every node offers a service of the form

$$\beta(t) = R(t - T)^+ \quad (6)$$

for some delay  $T$  and rate  $R$ . Following [11], we call it a “rate latency” curve. We show in this section that there are a number of scheduling policies that offer such service curves. The IETF further assumes [12] that the delay parameter  $T$  depends on the rate  $R$  according to  $T = \frac{C}{R} + D$  for some constants  $C$  and  $D$ . The values of  $C$  and  $D$  are computed during reservation setup, with a protocol such as the Resource Reservation Protocol (RSVP) [21].

A number of scheduling policies have been proposed in the literature (see, for example, [1], [7], [13], [22], [23]). Consider the general form of scheduling proposed in [13] under the name of Guaranteed Rate (GR) scheduling. It is shown in [13] that Guaranteed Rate scheduling includes as particular cases: virtual clock scheduling [23], packet-by-packet generalized processor sharing [1], and self-clocked fair queuing [22]. Essentially identical results can be found in [24] using the concept of “minimum bandwidth property.”

Following [13], we say that a scheduling policy is of the *guaranteed rate* type, with rate  $R$  and delay  $v$  for a given flow if it guarantees that packet  $j$  of the flow is served by time  $GRC(j) + v$ , with  $GRC$  defined by  $GRC(0) = 0$  and

$$GRC(j) = \max\{A(j), GRC(j-1)\} + \frac{l(j)}{R}. \quad (7)$$

In the formula,  $l(j)$  is the length in bits of packet  $j$  and  $A(j)$  is the arrival time of packet  $j$ .

*Theorem 2. Modeling GR Nodes:* A node with guaranteed rate scheduling policy offers a service curve defined by

$$\beta(t) = R \left( t - \frac{l}{R} - v \right)^+$$

where  $l = \sup_j l(j)$  is the maximum packet size in bits.

*Proof:* Call  $D(i)$  the departure time for the last bit of packet  $i$ . If  $\sup_i D(i)$  is finite, then if  $t > \sup_i D(i)$  then  $R(t) = R^*(t)$  since any packet is guaranteed to leave the system after a finite time. The service curve property is trivially

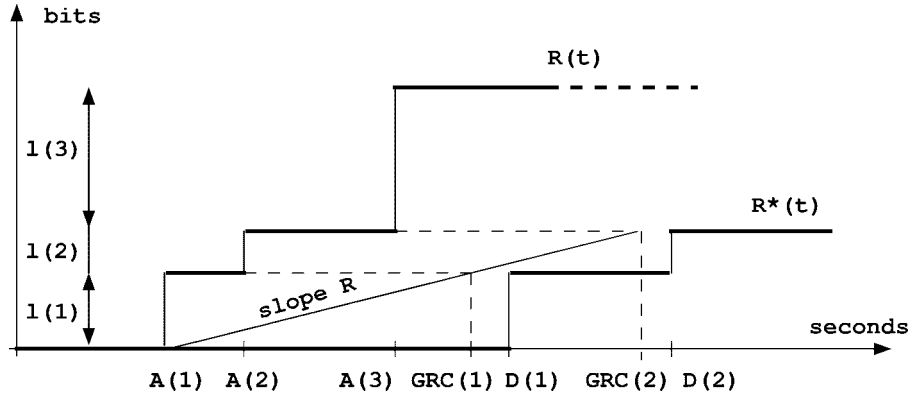


Fig. 1. Input and output functions for GR scheduling.

true in that case. We can thus assume now that there is a packet index  $j$  such that  $D(j-1) \leq t < D(j)$ . Thus

$$R^*(t) = \sum_{i=1}^{j-1} l(i). \quad (8)$$

Define

$$i_0 = \max\{i \text{ such that } 1 \leq i \leq j \text{ and } GRC(i-1) \leq A(i)\}.$$

Fig. 1 illustrates a case with  $j = 2$  and  $i_0 = 1$ . We first consider the case where  $A(i_0) \leq D(j-1)$ .

Now

$$GRC(i_0) = A(i_0) + \frac{l(i_0)}{R}$$

and for all  $i$  such that  $i_0 < i \leq j$  we have

$$GRC(i) = GRC(i-1) + \frac{l(i)}{R}$$

thus

$$GRC(j) = A(i_0) + \frac{1}{R} \sum_{i=i_0}^j l(i). \quad (9)$$

Define  $t_0 = A(i_0) - \epsilon$ , where  $\epsilon$  is small enough for  $t + \epsilon \leq D(j)$  and  $A(i_0) - \epsilon > A(i_0 - 1)$ . We have

$$R(t_0) = \sum_{i=1}^{i_0-1} l(i) \quad (10)$$

and

$$t - t_0 \leq D(j) - A(i_0) \leq GRC(j) + v - A(i_0)$$

and from (9), it follows that

$$t - t_0 \leq v + \frac{1}{R} \sum_{i=i_0}^j l(i) \leq v + \frac{l}{R} + \frac{1}{R} \sum_{i=i_0}^{j-1} l(i).$$

Thus by definition of  $\beta$

$$\beta(t - t_0) \leq \sum_{i=i_0}^{j-1} l(i), \quad (11)$$

Assume now that  $A(i_0) \leq D(j-1)$ , which implies that  $t - t_0 \geq 0$ . From (8) and (10), the last term in (11) is precisely  $R^*(t) - R(t_0)$ , which proves the service curve property in that case.

Now if  $A(i_0) > D(j-1)$  then necessarily  $i_0 = j$ . If  $t \geq A(i_0)$ , then the above reasoning applies and shows the service curve property. Otherwise,  $D(i_0 - 1) \leq t < A(i_0)$ , the queue is empty at  $t$ , and the service curve property is true with  $t_0 = t$ .  $\square$

The result in this section is compatible with the common node model of IETF, who chose to restrict the modeling to rate latency curves. Together with the network calculus results of Section II, they can be used to derive tight upper bounds on delays and backlogs, which gives as a particular case the results in [13]. Remember that, in performing the comparison, one should be careful to map the “arrival curve” condition in [13] to  $\alpha(t-s) + l$ , where  $l$  is the maximum packet length.

There are known forms of scheduling that do not belong to the GR type. One such form is the delay-based scheduler [25]. As mentioned in Section II, a delay-based scheduler offers a service curve of  $\delta_T$  for some  $T$ , which is thus a particular case of a rate latency service curve. There are, however, other forms of scheduling which cannot be modeled well with a rate-latency curve. For example, the VBR trunk discussed in Section VI does not fit in that framework. A more general discussion of this point leads to the concept of service-curve scheduling, which is defined and analyzed in [8] and [26]. The restriction by IETF to rate latency curves is thus at the cost of losing some flexibility.

In practice, we use network calculus to compute bounds on delay variation and backlogs. Constant propagation delays between systems can be ignored for the computation of delay bounds, as they simply transform an input function into the functions shifted in time by a fixed amount. Variable delays can be modeled simply by the addition of  $\delta_T$  functions to the service curves. For the computation of the backlog inside a system  $\mathcal{S}$ , constant delays inside  $\mathcal{S}$  cannot be ignored and should be modeled with a  $\delta_T$  service curve. As a consequence of the commutativity of  $\otimes$ , such delays can be inserted in any order along a sequence of buffers, without altering the delay bounds.

## V. DETERMINISTIC EFFECTIVE BANDWIDTH AND EQUIVALENT CAPACITY

In this section we introduce the concepts of effective bandwidth and equivalent capacity in a deterministic context.

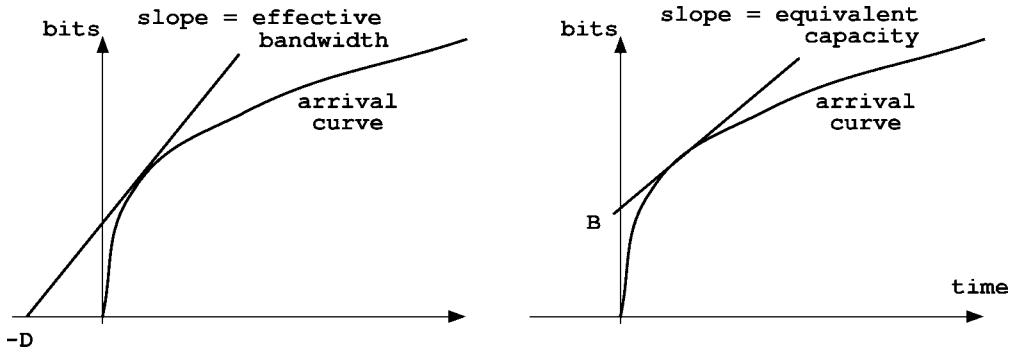


Fig. 2. Effective bandwidth for a delay constraint  $D$  and equivalent capacity for a buffer size  $B$ .

### A. Effective Bandwidth and Equivalent Capacity of a Flow

Consider a trunk system that serves a flow in a work-conserving manner, at a constant rate  $C$ . We assume that the arrival flow is constrained by an arrival curve  $\alpha$  and would like to characterize the minimum value of  $C$  that is required for a given  $\alpha$  and a given constraint on delay or buffer capacity. This problem has been studied in [14, pp. 270–273], in the specific case of a flow constrained by one leaky bucket. The authors in [14] find that, if we impose a fixed-delay constraint  $D$  to the flow, then the condition on  $C$  is that  $C \geq C_D$ , where  $C_D$  depends on the leaky bucket parameters and the delay constraint.  $C_D$  is called the (deterministic) effective bandwidth of the flow, for a delay constraint of  $D$ . If  $N$  identical flows are superimposed, the effective bandwidth of the aggregate flow is  $NC_D$ ; in contrast, for a heterogeneous mix of flows, there is no such additive property.

We show now how these results derive from a more general concept. Back to the general case, for a given arrival curve  $\alpha$ , we wish to find a rate  $C$  such that the horizontal deviation  $h(\alpha, \lambda_C)$  is not more than  $D$ , where  $\lambda_C$  is the function defined by  $\lambda_C(t) = Ct$ . This is equivalent to requiring that  $\alpha(s) \leq C(s+D)$  for all  $s \geq 0$ , which in turn can be expressed as

$$C \geq \sup_{s \geq 0} \frac{\alpha(s)}{s+D}.$$

We have thus shown the following:

**Proposition 2. Effective Bandwidth:** The queue with constant rate  $C$  guarantees a delay bound of  $D$  to a flow with arrival curve  $\alpha$  if  $C \geq e_D(\alpha)$ , with

$$e_D(\alpha) = \sup_{s \geq 0} \frac{\alpha(s)}{s+D}. \quad (12)$$

We call  $e_D(\alpha)$  the *effective bit rate*, or *deterministic effective bandwidth* corresponding to the arrival curve  $\alpha$ , for a delay constraint  $D$ . If  $\alpha$  is differentiable,  $e(D)$  is the slope of the tangent to the arrival curve, drawn from the time axis at  $t = -D$  (Fig. 2).

Assume  $\alpha$  is subadditive. We define the sustainable rate  $m$  as

$$m = \inf_{s \rightarrow +\infty} \alpha(s)/s$$

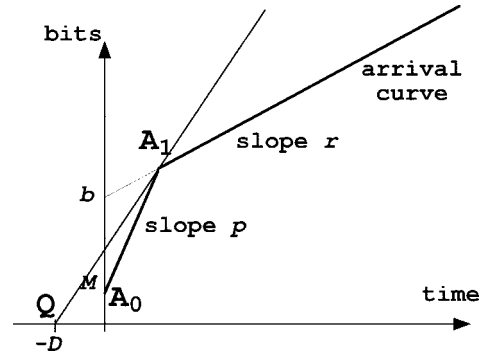


Fig. 3. Effective bandwidth for an arrival curve according to the IETF specification.

and the peak rate by

$$p = \sup_{s \geq 0} \alpha(s)/s.$$

Then  $m \leq e_D(\alpha) \leq p$  for all  $D$ . Moreover, if  $\alpha$  is concave, then

$$\lim_{D \rightarrow +\infty} e_D(\alpha) = m.$$

For example, for a flow constrained according to the IETF specification, with maximum packet size  $M$ , peak rate  $p$ , sustainable rate  $r$ , and burst tolerance  $b$ , the effective bandwidth is the maximum of  $r$  and the slopes of lines  $(QA_0)$  and  $(QA_1)$  in Fig. 3; it is thus equal to

$$e_D = \max \left\{ \frac{M}{D}, r, p \left( 1 - \frac{D - \frac{M}{p}}{x + D} \right) \right\} \quad (13)$$

with  $x = \frac{b-M}{p-r}$ .

It also follows directly from the definition in (12) that

$$e_D \left( \sum_i \alpha_i \right) \leq \sum_i e_D(\alpha_i). \quad (14)$$

In other words, the effective bandwidth for an aggregate flow is less than or equal to the sum of effective bandwidths, assuming the delay constraints are identical for all flows. If the flows have all *identical* arrival curve, then the aggregate effective

bandwidth is simply  $I \times e_D(\alpha_1)$ . It is this latter relation which is the origin of the term “effective bandwidth.” The difference

$$\sum_i e_D(\alpha_i) - e_D\left(\sum_i \alpha_i\right)$$

is a (nonstatistical) multiplexing gain; it tells us how much capacity is saved by sharing a buffer between the flows.

Similar results hold if we replace delay constraints by the requirement that a fixed buffer size is not exceeded. Indeed, the queue with constant rate  $C$ , guarantees a maximum backlog of  $B$  (in bits) for a flow with arrival curve  $\alpha$  if  $C \geq f_B(\alpha)$ , with

$$f_B(\alpha) = \sup_{s \geq 0} \frac{\alpha(s) - B}{s}. \quad (15)$$

We call  $f_B(\alpha)$  the *equivalent capacity*, by analogy to [16] (Fig. 2 gives a graphical interpretation). Similar to effective bandwidth, the equivalent capacity of a heterogeneous mix of flows is less than or equal to the sum of equivalent capacities of the flows, provided that the buffers are also added up; in other words,

$$f_B(\alpha) \leq \sum_i f_{B_i}(\alpha_i)$$

with  $\alpha = \sum_i \alpha_i$  and  $B = \sum_i B_i$ .

Note that (12) or (15), or both, can be used to estimate the capacity required for a flow, based on a measured arrival curve.

### B. Call Acceptance Regions

Coming back to (14), we can state in more general terms that the effective bandwidth is a convex function of function  $\alpha$ , namely,

$$e_D(a\alpha_1 + (1-a)\alpha_2) \leq ae_D(\alpha_1) + (1-a)e_D(\alpha_2) \quad (16)$$

for all  $a \in [0, 1]$ . The same is true for the equivalent capacity function.

Consider now a call acceptance criterion based solely on a delay bound, or based on a maximum buffer constraint, or both. Consider further that there are  $I$  types of connections; connections within type  $i$  are characterized by a common arrival curve  $\alpha_i$ . The connections are multiplexed on the same trunk, and all connections have an identical delay constraint  $D$ . Define the acceptance region  $\mathcal{A}$  as the set of values  $(n_1, \dots, n_I)$  that satisfy the call acceptance criterion, where  $n_i$  is the number of connections of class  $i$ . From (16) we can extend the acceptance region to noninteger values of  $n_i$ ; we do this in order to simplify the following discussion on convexity. From the convexity of the effective bandwidth and equivalent capacity functions, it follows that the acceptance region  $\mathcal{A}$  is *convex*.

This contrasts with acceptance regions based on statistical multiplexing with large deviation asymptotics. In such cases, a set of values  $(n_1, \dots, n_I)$  is acceptable if a given upper bound on loss probability is less than  $\epsilon$ , where  $\epsilon$  is a small, fixed value. In a broad family of cases, and especially for bufferless models, it is the complement in the positive orthant which is convex [15]. Call  $\mathcal{C}_\epsilon$  the acceptance region for a given

value of  $\epsilon$ .  $\mathcal{C}_\epsilon$  becomes smaller as  $\epsilon$  decreases to 0, and the limit is

$$\mathcal{C} = \bigcap_{\epsilon > 0} \mathcal{C}_\epsilon.$$

Under the assumptions in [15], the complement of  $\mathcal{C}$  in the positive orthant is also convex, since the union of an increasing set of convex sets is convex.

Of course, we have  $\mathcal{A} \subset \mathcal{C}_\epsilon$  for all  $\epsilon$  and thus  $\mathcal{A} \subset \mathcal{C}$ . Ideally, we would like to have equality, but this may in general not be possible because  $\mathcal{A}$  is usually strictly convex (see [27] for an example in the case of a call acceptance criterion based on a buffer requirement only). Thus the limit of acceptance regions based on statistical multiplexing, as the loss probability tends to 0, may be strictly larger than the acceptance region based on a lossless criterion.

## VI. OPTIMAL PARAMETERS FOR AN ATM VARIABLE BIT-RATE TRUNK

### A. Optimal Peak Rate

As mentioned earlier, our initial motivation is a measurement-based call admission control (CAC) method, for the case where connections are admitted onto an ATM variable bit rate (VBR) trunk, or tunnel [17]. Tunneling refers to the multiplexing of several flows into a larger flow, which is handled in subsequent nodes as a single entity (called a virtual trunk in [17]). Tunneling occurs when a number of RSVP flows are multiplexed onto one single ATM connection, or over one RSVP flow itself. We expect tunneling to play an important role in the scalability of integrated services networks. Using VBR trunks rather than constant bit rate trunks is advantageous in particular for trunks carrying medium amounts of traffic, that are thus best able to benefit from statistical multiplexing in a larger network [28].

More specifically, in this paper, we focus on one issue, namely, given a predicted traffic, find the optimal VBR trunk parameters that can carry the traffic. The same problem arises in prerecorded video scenarios, where prediction is not based on traffic measurements but on the contents of a stored file.

We assume in this section that a number of flows, with an aggregate arrival curve  $\alpha$ , is multiplexed into a VBR trunk. The VBR trunk is viewed as a single flow by downstream nodes; as such, it is constrained by an arrival curve  $\sigma$ . We are interested in finding, for a given arrival curve  $\alpha$ , the optimal values of  $\sigma$ , for a criterion that we define later. If we assume a work-conserving scheduling policy, then the multiplexing node acts as a general shaper. From Section III, we can conclude that it offers a service curve  $\sigma$  to the aggregate flow. Assume that the constraint at the multiplexor node is to guarantee a maximum delay  $D$ . From Section II, the requirement on the virtual trunk is that, for all  $s \geq 0$ , we have

$$\sigma(s+D) \geq \alpha(s). \quad (17)$$

We have assumed that the virtual trunk is an ATM Variable Bit Rate (VBR) connection. The shaping curve for a VBR

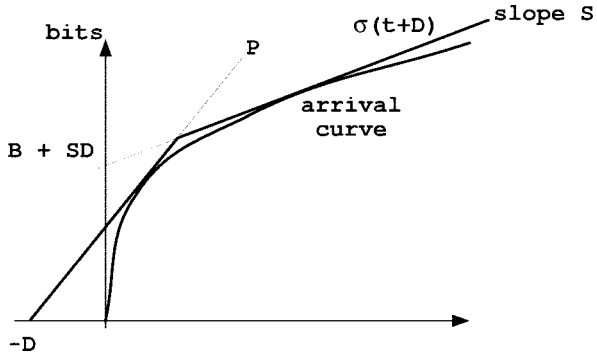


Fig. 4. Virtual trunk with parameters  $(P, S, B)$  satisfies the delay constraint  $D$  for traffic with arrival curve drawn on the picture.

connection has the form  $\sigma(t) = \min(Pt, St + B)$  where  $P$  is the peak rate,  $S$  the sustainable rate, and  $B$  a burst tolerance parameter. Equation (17) is illustrated on Fig. 4.

Equation (17) becomes

$$\text{for all } s \geq 0: \begin{cases} (s+D)P \geq \alpha(s) \\ (s+D)S + B \geq \alpha(s). \end{cases} \quad (18)$$

The first condition in (18) implies that  $P \geq e_D(\alpha)$ , which is thus a necessary and sufficient condition on  $P$ . In other words, we have shown that for a virtual trunk of the ATM VBR type, *there is a minimum peak rate  $P_0$ , which is the effective bandwidth of the arrival stream* and that this minimum peak rate is also optimal. More precisely, the latter statement means the following. We say that the parameter set  $(P, S, B)$  of the virtual trunk is feasible if the virtual trunk is able to carry the traffic with a delay less than the delay constraint  $D$ . The result is that, if  $(P, S, B)$  is feasible, then on the one hand, necessarily  $P \geq P_0 = e_D(\alpha)$ , and on the other hand,  $(P_0, S, B)$  is also feasible. Another aspect of this result is that, from a bandwidth point of view, using a VBR trunk rather than a constant bit rate (CBR) trunk is all benefit since, by definition of the effective bandwidth, the CBR trunk would have a rate of at least  $P_0$ .

### B. Optimal VBR Trunk

If we wish to solve the problem of finding a complete optimal parameter set for  $(S, B)$ , then we need an optimality criterion. As an example, we assume that we wish to minimize a cost function  $c(P, S, B)$  which is affine and wide-sense increasing in each of its variables. The value of  $P$  should thus be set to  $P_0$  and we wish to minimize

$$c(P_0, S, B) - c(P_0, 0, 0) = uS + vB$$

for fixed values of  $u$  and  $v$ . Without loss of generality, we can set  $v = 1$ .

The values of  $S$  and  $B$  are also normally limited not to exceed maximum values  $S_{\max}$ , respectively,  $B_{\max}$ . Without loss of generality, we can further assume that

$$S_{\max} \leq P_0. \quad (19)$$

We thus have to minimize  $uS + B$  in the acceptable region, defined by

$$0 \leq S \leq S_{\max} \quad (20)$$

$$0 \leq B \leq B_{\max} \quad (21)$$

for all

$$s \geq 0: B + (s+D)S - \alpha(s) \geq 0, \quad (22)$$

First note that the acceptance region is not empty if and only if

$$B_{\max} \geq \sup_{s \geq 0} [\alpha(s) - (s+D)S_{\max}]. \quad (23)$$

In the remainder of this section we assume that (23) holds.

*Reduction to Another Optimization Problem:* It is suggested by Fig. 4, and is confirmed in the rest of the derivations, that it is convenient to perform the change of variable defined by

$$\begin{cases} x = S \\ y = B + SD. \end{cases} \quad (24)$$

Our optimization problem becomes

Minimize  $(u - D)x + y$  on the set defined by

$$\begin{cases} 0 \leq x \leq S_{\max} \\ 0 \leq y - Dx \leq B_{\max} \\ y \geq -\check{\alpha}(x). \end{cases} \quad (25)$$

In the above,  $\check{\alpha}$  is the concave conjugate of  $\alpha$ , defined by

$$\check{\alpha}(x) = \inf_{s \geq 0} \{xs - \alpha(s)\}.$$

We call  $\mathcal{R}$  the region of the  $(x, y)$  plane defined in (25).

*Optimization Region:* We now study the optimization region  $\mathcal{R}$ . The region is convex, as it is the intersection of half-planes. Define  $\mathcal{R}_1$  as the set of points  $(x, y)$  such that

$$\begin{cases} 0 \leq x \leq S_{\max} \\ 0 \leq y - Dx \leq B_{\max} \end{cases}$$

and  $\mathcal{R}_2$  as the set of points  $(x, y)$  such that

$$y \geq -\check{\alpha}(x)$$

so that  $\mathcal{R}$  is the intersection of  $\mathcal{R}_1$  and  $\mathcal{R}_2$ . We assume that  $\alpha$  is subadditive and define the sustainable rate  $m$  as

$$m = \inf_{s \rightarrow +\infty} \frac{\alpha(s)}{s}$$

and the peak rate by

$$p = \sup_{s \geq 0} \frac{\alpha(s)}{s}.$$

Then we have

$$\begin{cases} \text{if } x < m, & \text{then } -\check{\alpha}(x) = +\infty \\ \text{if } x \geq p, & \text{then } -\check{\alpha}(x) = \alpha(0). \end{cases}$$

Since  $-\check{\alpha}$  is convex, the latter equation implies that  $-\check{\alpha}$  is wide-sense decreasing. Define the points  $Q = (x_Q, y_Q)$ ,  $R = (x_R, y_R)$ , and  $W = (x_W, y_W)$  in the  $(x, y)$  plane as the intersection of the border  $\partial\mathcal{R}_2$  of  $\mathcal{R}_2$  and the line defined by  $y = Dx + B_{\max}$  (respectively,  $y = Dx$ ,  $x = S_{\max}$ ) and  $A$  as

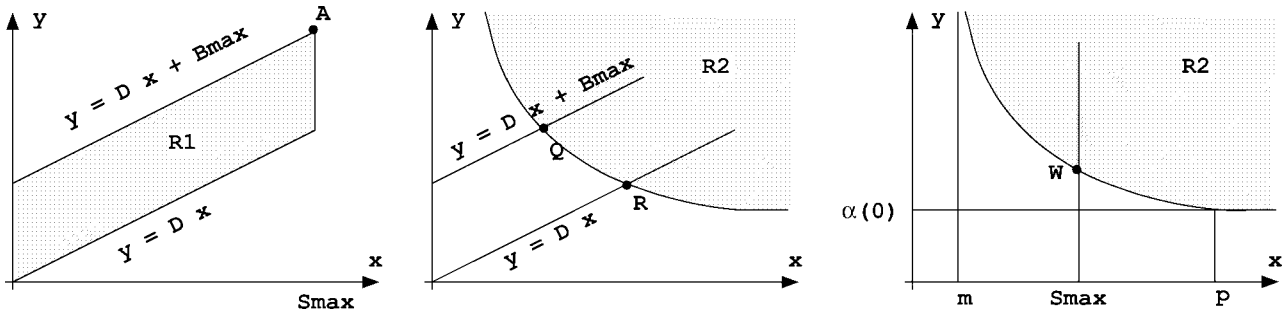


Fig. 5. The optimization region  $\mathcal{R}$  is the intersection of  $\mathcal{R}_1$  and  $\mathcal{R}_2$ . The figure shows the points  $A$ ,  $Q$ ,  $R$ , and  $W$ .

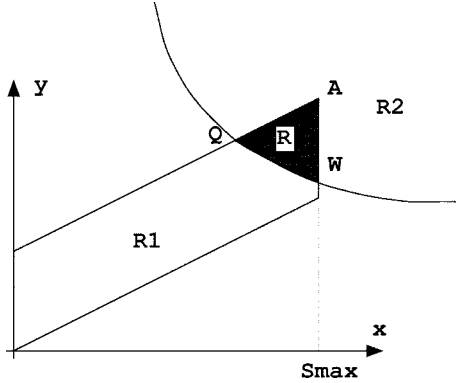


Fig. 6. Relative placements of  $\mathcal{R}_1$  and  $\mathcal{R}_2$  under the conditions in this section.

the intersection of the lines  $y = Dx + B_{\max}$  and  $x = S_{\max}$  (see Fig. 5).

We now study the relative positions of points  $A$ ,  $Q$ ,  $R$ , and  $W$ . Point  $Q$  is defined by

$$\begin{cases} \sup_{s \geq 0} (\alpha(s) - x_Q s) = y_Q \\ y_Q = Dx_Q + B_{\max} \end{cases}$$

which, after some algebra and using the fact that  $D > 0$ , gives

$$\begin{cases} x_Q = \sup_{s \geq 0} \frac{\alpha(s) - B_{\max}}{s + D} \\ y_Q = Dx_Q + B_{\max}. \end{cases}$$

Similarly, point  $R$  is defined by

$$\begin{cases} x_R = \sup_{s \geq 0} \frac{\alpha(s)}{s + D} = P_0 \\ y_R = Dx_R. \end{cases}$$

Combining with (19) and (23), we obtain

$$x_Q \leq S_{\max} \leq x_R.$$

Finally, we have

$$\begin{cases} x_W = S_{\max} \\ y_W = -\alpha(S_{\max}). \end{cases}$$

Again from (23), we have

$$y_W \leq DS_{\max} + B_{\max} = y_A$$

(see Fig. 5).

Alltogether, this shows that  $\mathcal{R}_2$  intersects the border of  $\mathcal{R}_1$  at points  $Q$  and  $W$ . This is illustrated in Fig. 6.

*Computation of the Optimum:* Since the cost function is linear, its minimum is attained on the border of  $\mathcal{R}$ , and it can be only on a point of the border  $\partial\mathcal{R}_2$  of  $\mathcal{R}_2$  or at point  $A$  [29]. It is straightforward to see that in all cases, the cost at  $A$  is superior than or equal to the costs at  $Q$  and  $R$ , thus the cost is minimum for a point on  $\partial\mathcal{R}_2$  which lies between  $Q$  and  $R$ . Thus the optimization problem is reduced to the following one:

Minimize  $(u - D)x - \check{\alpha}(x)$  on the set defined by  $x_Q \leq x \leq S_{\max}$ . (26)

Call  $g$  the function  $x \rightarrow (u - D)x - \check{\alpha}(x)$ . If  $u \leq D$  then  $g$  is wide-sense decreasing thus its minimum is for  $x = S_{\max}$ .

Assume now that  $u > D$ . Function  $g$  is convex, thus its minimum is attained on an interval  $[x_1, x_2]$ . In addition,  $g$  is wide-sense decreasing on the set  $x < x_1$  and wide-sense increasing on  $x > x_2$  (some of these sets may be empty). In order to determine the minimum of  $g$  on the set  $[x_Q, S_{\max}]$ , we determine the relative positions of all these intervals. This gives the following result. Let  $x_0$  be one point which minimizes  $g(x) = (u - D)x - \check{\alpha}(x)$ .

- If  $x_Q \leq x_0 \leq S_{\max}$  then the minimum of our cost function is obtained for  $x = x_0$ .
- Else if  $x_0 > S_{\max}$  then the minimum is obtained for  $x = S_{\max}$ .
- Else (namely, if  $x_0 < x_Q$ ) then the minimum is obtained for  $x = x_Q$ .

In general we need to obtain the value of one  $x_0$  by a numerical procedure.

*Special Cases:* In some cases we can assume that  $\alpha$  is two times differentiable. In such cases, we can parameterize  $\partial\mathcal{R}_2$  by  $s$  instead of  $x$ , using the fact that

$$x = \alpha'(s)$$

for some  $s$  that attains the maximum in  $\alpha(s) - xs$ . We have thus

$$g(x) = (u - D - s)\alpha'(s) + \alpha(s) = G(s).$$

Function  $G$  is now differentiable and

$$G'(s) = (u - D - s)\alpha''(s),$$

Thus one point  $x_0$  where the minimum of  $g$  is attained is for  $s = u - D$  and is  $x_0 = \alpha'(u - D)$ .



Define  $\tilde{\alpha}(x) = \inf_{s \geq 0} (xs - \alpha(s))$ ;  
 $P_0 = \sup_{s \geq 0} \frac{\alpha(s)}{s+D}$ ;  
 $S_{\max} = \min \{P_0, S_{\max}\}$ ;  
 if  $B_{\max} < \sup_{s \geq 0} \{\alpha(s) - (s+D)S_{\max}\}$  there is no solution;  
 else {  
   if  $(u < D)S_0 = S_{\max}$ ;  
   else {  
      $\bar{S} =$  one value of  $x$  that minimizes  $(u-D)x - \tilde{\alpha}(x)$ ; (1)  
      $S_e = \sup_{s \geq 0} \frac{\alpha(s) - B_{\max}}{s+D}$ ;  
     if  $(\bar{S} > S_{\max})S_0 = S_{\max}$ ;  
     else if  $(\bar{S} < S_e)S_0 = S_e$ ;  
     else  $S_0 = \bar{S}$  (2);  
   }  
    $B_0 = \sup_{s \geq 0} \{\alpha(s) - (s+D)S_0\}$  (3)  
   }  
 }  
 If  $\alpha$  is twice differentiable, then replace (1) by  
 $\bar{S} = \alpha'(u-D)$   
 If moreover  $\alpha$  is concave then in case (2) replace (3) by  
 $B_0 = \alpha(u-D) - uS_0$

Fig. 7. The algorithm for the optimal VBR virtual trunk parameters  $(P_0, S_0, B_0)$  for a given arrival curve  $\alpha$ , a delay constraint  $D$ , a cost parameter  $u$ , and maximum values  $S_{\max}$  and  $B_{\max}$ .

If we further assume that  $\alpha$  is concave, then, since it is continuous (because it is differentiable), it is equal to the concave conjugate of  $\tilde{\alpha}$ , thus

$$\inf_{x \geq 0} [sx - \tilde{\alpha}(x)] = \alpha(s)$$

and, in that case, the value of  $y_0$  corresponding to the minimum cost function is given by

$$y_0 = \alpha(u-D) - (u-D)x_0.$$

Note that the assumption that  $\alpha$  is concave is too strong in the cases where  $\alpha$  is obtained by measurements.

Now if we put together these results and perform the reverse mapping from (24), we obtain the optimal value  $(P_0, S_0, B_0)$  according to the algorithm in Fig. 7.

The algorithm shows the relation between  $u$  and  $D$ . We can interpret  $u$  as the ratio of the monetary costs attributed to the rate  $S$  and to the burst tolerance  $B$ . If  $u$  is less than the time constant  $D$  of the node, then the optimum rate  $S_0$  is equal to the maximum value  $S_{\max}$ .

The above result generalizes to the case where the virtual trunk arrival curve is defined by any number of affine constraints (not just two as in this example). Similar results hold if a maximum buffer requirement is considered.

The algorithm in Fig. 7 can be used to estimate the optimal characteristics of the virtual trunk, based on a measured arrival curve  $\alpha$ .

## VII. CONCLUSION

We have shown how network calculus can be used to provide a unifying treatment to a variety of problems arising in packet networks offering guaranteed service. We have shown that all rate-based schedulers fit in the framework of the IETF. Then we have applied the results to defining the concepts

of deterministic effective bandwidth and equivalent capacity. We have shown that the limit of the acceptance region based on statistical multiplexing, as the loss probability tends to 0, may be in general strictly larger than the acceptance region based on a lossless criterion. Finally, we have given an explicit algorithm for determining the optimal characteristics of a VBR trunk, assuming that the input traffic is known, and for a strict delay constraint. We have shown that the peak rate for the VBR trunk is the very rate of the CBR trunk that would carry the traffic with the same strict delay constraint, thus in this sense, a VBR trunk is the benefit.

The network calculus results derived in this paper have been developed independently and simultaneously by Chang [10], Agrawal and Rajan [11], and Cruz and Okino [18], who have developed applications to window flow control.

## ACKNOWLEDGMENT

The author wishes to thank R. L. Cruz for very constructive comments.

## REFERENCES

- [1] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single node case," *IEEE/ACM Trans. Networking*, vol. 1, pp. 344–357, June 1993.
- [2] ———, "A generalized processor sharing approach to flow control in integrated services networks: The multiple node case," *IEEE/ACM Trans. Networking*, vol. 2, pp. 137–150, Apr. 1994.
- [3] R. L. Cruz, "A calculus for network delay, part I: Network elements in isolation," *IEEE Trans. Inform. Theory*, vol. 37, pp. 114–131, Jan. 1991.
- [4] ———, "A calculus for network delay, part II: Network analysis," *IEEE Trans. Inform. Theory*, vol. 37, pp. 132–141, Jan. 1991.
- [5] ———, "Quality of service guarantees in virtual circuit switched networks," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1048–1056, Aug. 1995.
- [6] J. Kurose, "On computing per-session performance bounds in high-speed multi-hop computer networks," in *Proc. ACM Sigmetrics/Performance '92*, 1992.
- [7] L. Georgiadis, R. Guérin, V. Peris, and R. Rajan, "Efficient support of delay and rate guarantees in an Internet," in *Proc. Sigcomm '96*, Aug. 1996, pp. 106–116.
- [8] H. Sariowan, "A service curve approach to performance guarantees in integrated service networks," Ph.D. dissertation, Univ. Calif. San Diego, 1996.
- [9] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and Linearity, An Algebra for Discrete Event Systems*. New York: Wiley, 1992.
- [10] C. S. Chang, "On deterministic traffic regulation and service guarantee: A systematic approach by filtering," in *Proc. Infocom 1997*, 1997.
- [11] R. Agrawal and R. Rajan, "Performance bounds for guaranteed and adaptive services," IBM Tech. Rep. RC 20649, Dec. 1996.
- [12] S. Shenker, C. Partridge, and R. Guérin, "Specification of guaranteed quality of service," Aug. 1996, Internet draft. Available FTP: ftp://ds.internic.net/internet-drafts/draft-ietf-intserv-guaranteed-svc-06.txt (Internet drafts can be obtained by FTP from your local Internet documentation sites).
- [13] P. Goyal, S. S. Lam, and H. Vin, "Determining end-to-end delay bounds in heterogeneous networks," in *5th Int. Workshop on Network and Operational System Support for Digital Audio and Video* (Durham, NH, Apr. 1995).
- [14] J. Walrand and P. Varayia, *High Performance Communication Networks*. San Francisco, CA: Kaufmann, 1996.
- [15] F. Kelly, "Effective bandwidths of multiclass queues," *Queueing Syst.*, vol. 9, no. 1, pp. 5–15, 1991.
- [16] L. Gun and R. Guérin, "Bandwidth management and congestion control framework of the broadband network architecture," *Bandwidth Manag. Congestion Contr. Framework Broadband Network Architecture*, vol. 26, pp. 61–78, 1993.
- [17] S. Giordano, J.-Y. Le Boudec, P. Oechslin, and S. Robert, "Vbr over vbr: The homogeneous, loss-free case," in *Infocom '97*, 1997.

- [18] R. L. Cruz and C. M. Okino, "Service guarantees for window flow control," in *34th Allerton Conf. Communication, Control, and Computing* (Monticello, IL, Oct. 1996).
- [19] J.-Y. Le Boudec, "Network calculus made easy," 1996, Tech. Rep. EPFL-DI 96/218. Available on line at [http://lrcwww.epfl.ch/PS\\_files/d4paper.ps](http://lrcwww.epfl.ch/PS_files/d4paper.ps).
- [20] J.-Y. Le Boudec and P. Thiran, "A note on time and space methods in network calculus," Apr. 1997, EPFL Tech. Rep. DI 97/224. Available on line at [http://lrcwww.epfl.ch/PS\\_files/TimeSpaceMthdsv1.ps](http://lrcwww.epfl.ch/PS_files/TimeSpaceMthdsv1.ps).
- [21] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource reservation protocol (rsvp)—Version 1 functional specification," July 1996. Internet draft. Available on line at <ftp://ds.internic.net/internet-drafts/draft-ietf-rsvp-spec-16.txt>. (Internet drafts can be obtained by FTP from your local Internet documentation sites).
- [22] S. J. Golestani, "A self clocked fair queuing scheme for high speed applications," in *Proc. Infocom'94*, 1994.
- [23] L. Zhang, "A new traffic control algorithm for packet switching networks," in *Proc. ACM Sigcomm'90*, 1990.
- [24] H. Hung and G. Kesidis, "Bandwidth scheduling for wide-area atm networks using virtual finishing times," *IEEE/ACM Trans. Networking* vol. 4, pp. 49–54, Feb. 1996.
- [25] H. Zhang, "Service disciplines for guaranteed performance service in packet-switching networks," *Proc. IEEE*, vol. 84, pp. 1374–1396, Oct. 1996.
- [26] H. Sariowan, R. L. Cruz, and G. C. Polyzos, "Scheduling for quality of service guarantees via service curves," in *Proc. ICCCN'95*, Sept. 1995, pp. 512–520.
- [27] F. Lo Presti, Z. L. Zhang, D. Towsley, and J. Kurose, "Source time scale and optimal buffer/bandwidth tradeoff for regulated traffic in atm node," Tech. Rep. UMAS TR UM-CS-96-38, 1996.
- [28] J.-Y. Le Boudec and A. Ziedins, "A cac algorithm for vbr connections over a vbr trunk," in *Proc. ITC'97*, June 1997.
- [29] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.