Multiplexing Regulated Traffic Streams: Design and Performance

Krishnan Kumaran and Michel Mandjes
Bell Laboratories/Lucent Technologies,
600 Mountain Ave., P.O. Box 636, Murray Hill, NJ 07974-0636, USA
{kumaran,michel}@research.bell-labs.com

Abstract— **The main network solutions for supporting QoS rely on traffic policing (conditioning, shaping). In particular, for IP networks the IETF has developed Intserv (individual flows regulated) and Diffserv (only aggregates regulated). The regulator proposed could be based on the (dual) leaky-bucket mechanism. This explains the interest in network element performance (loss, delay) for leaky-bucket regulated traffic.**

This paper describes a novel approach to the above problem. Explicitly using the correlation structure of the sources' traffic, we derive approximations for both small and large buffers. Importantly, for small (large) buffers the short-term (long-term) correlations are dominant. The large buffer result decomposes the traffic stream in a stream of constant rate and a periodic impulse stream, allowing direct application of the Brownian bridge approximation. Combining the small and large buffer results by a concave majorization, we propose a simple, fast and accurate technique to statistically multiplex homogeneous regulated sources.

To address heterogeneous inputs, we present similarly efficient techniques to evaluate the performance of multiple classes of traffic, each with distinct characteristics and QoS requirements. These techniques, applicable under more general conditions, are based on optimal resource (bandwidth and buffer) partitioning. They can also be directly applied to set GPS (Generalized Processor Sharing) weights and buffer thresholds in a shared resource system.

Key words—**packet networks, policing, admission control, queueing theory, buffer overflow, ATM, IP, Internet, Intserv, Diffserv.**

I. INTRODUCTION

We are currently faced with important changes in the nature of Internet applications. Roughly speaking, traditional applications like web browsing, file transfer, and email can be considered purely as data transfers, which do not have strict Quality of Service (QoS) requirements. The intended broader set of services, however, has an essentially different character, as it *does* need relatively hard guarantees (think for instance of real-time services and services that require a minimal throughput). Consequently, the currently employed service model of best effort service will not be appropriate to support the new services.

Initiatives within the IETF. The awareness of the 'QoS problem' led to a number of initiatives within the Internet Engineering Task Force, sse for instance [26]. The initially proposed architecture, *Intserv* [28], relied on QoS guarantees made on a per flow basis. With the so-called Reservation Protocol (RSVP) resources are explicitly allocated along a route through the network. Because QoS is guaranteed for individual flows, all these flows must be policed and subject to admission control. In other words, the concept proposed shows a strong similarity to ATM, a connection-oriented network technology (as opposed to the connectionless Internet Protocol).

The crucial disadvantage of Intserv is its lack of scalability. In the core of the network the number of flows simultaneously using a router is usually extremely large, whereas the flows themselves are active for a relatively short amount of time. This imposes severe requirements on the amount of state to be maintained by the routers. It was recognized that Intserv might be applied in the edge of the network (where the number of flows is relatively low), but not in the core.

For the core another mechanism has been proposed: *Diffserv* [2]. In Diffserv no micro-flows are identified, hence solving the scalability problems; instead agreements are made for *aggregates* of flows. The different ISPs have contracts with each other, thus 'guaranteeing' the end-users an end-to-end QoS. Any ISP has the right to condition the incoming aggregate flows on the basis of the agreement that it has with neighboring domains, usually enforced by policing. (It should be noted that there are two essentially different flavors within Diffserv: expedited forwarding (EF) and assured forwarding (AF). In this paper the exact differences between these are not crucial). Notice that there is no real-time admission control — ISPs keep track of the resources they allocated to the users of their network (for instance other ISPs or big 'aggregate' customers) — renegotiations are usually done by bandwidth-broker type of mechanisms.

Policing. Dual leaky bucket. As noted above, in both the edge mechanism and the core mechanism a crucial role is played by *policing,* albeit on different levels. In Intserv all flows are asked to declare a traffic contract, typically the three parameters of the dual leaky bucket, and those are policed. In Diffserv, at a peer router, there is the possibility of policing the aggregate of a number of microflows.

The dual leaky bucket (DLB) [24] is a policing device that has three parameters: peak rate p and sustainable rate r (which is a maximum on the average rate), as well as the maximum allowed burst period T (which is the maximum time that the source is allowed to send consecutively at peak rate). Formally, if $A(s,t)$ is the amount of traffic controlled by the policer in $[s,t)$, then $A(s,t) \leq \min\{p(t-s), r(t-s)+\alpha\}$, where $\alpha = (p-r)T$ (for any s, t with $s < t$).

Per node performance. One node in the network can be seen as a queue fed by a number of sources, where a source of type i is characterized by its 'triple' (r_i, p_i, T_i). We assume that the system (router) is characterized by a (constant) link rate C and a buffer B. A central question then is: suppose we have N_i sources of type i, $i = 1, \ldots, K$, is the combination $(N_1, \ldots, N_K)$ *admissible*, in that it offers all sources sufficient QoS? One could for instance aim for zero loss, which leads to solving a linear program [4], [15], [17]. Notice that an equivalent problem arises when the guarantee is specified as maximum delay D, where $D = B/C$.

However, in practice the guarantees required are mostly not so strict. Rather than deterministic guarantees (zero loss, delay threshold never exceeded) customers are usually satisfied with probabilistic criteria (the probability of exceeding buffer level B smaller than L, the probability that the delay is larger than D smaller than L'). The most significant advantage of these statistical guarantees is that the resulting admissible region is substantially larger than under deterministic guarantees: the number of admissible sources can be increased considerably.

Worst case traffic. Having decided to use statistical guarantees, the next question is: how to model regulated traffic, e.g. traffic that has been subject to a DLB policer? To construct a mechanism that reliably offers the agreed QoS, our calculations should be based on the 'worst' traffic that is compliant with the traffic contract, i.e., 'adversarial' traffic that in some sense maximizes the loss probability. A number of studies have tried to identify this 'worst case traffic'. Significant contributions were made in [7], [21], [5], and [19]. In this study we use the – non-exact, but justifiable – assumption that worst case traffic is modeled as deterministic on-off.

Deterministic on-off traffic is periodic: a source with descriptor (r, p, T) transmits at rate p during a fixed time T, and is silent during a fixed time $T' := (p/r - 1)T$, such that the average rate is r. We employ on the customary fluid model, thus neglecting the discrete nature of packet traffic. As long as the periods of the sources relate as rational numbers, the resulting queue is periodic. The only random effect is the phase of the sources; as said above, the durations of the activity periods and silences themselves are not random.

Deterministic on-off sources have an interesting correlation structure. On a short time scale there is positive correlation: the traffic rate is likely to be unchanged over small intervals. On the other hand, on somewhat longer timescales there is negative correlation: if the source is on, then it is certain that it will be off T units of time later.

It is noted that models with purely negative correlations (for instance the $N{\cdot}$D/D/1 queue [22, p. 397] where sources transmit their bursts instantaneously rather than gradually) have been solved (i.e., an explicit form for the buffer content distribution has been derived). On the other hand also queues with sources with purely positive correlations (for instance on-off sources with exponential on- and off-times [1]) have been analyzed successfully. However, the model with deterministic on-off traffic is known to be notoriously difficult, and no closed-form solution is known.

Contribution of this work. A significant contribution on the performance of a node with regulated input is by Elwalid, Mitra, and Wentworth [10]. They present an insightful procedure for calculating an upper bound to the loss probability. They translate a technique to estimate loss probabilities into a manageable admission control (for more results on admission control, see also [12] and survey paper [25]). However, the construction applied used there does not completely reflect the statistical sharing of resources and tends to be conservative. In the present paper we will explain in detail which effects are captured and which are not.

The contribution of our work lies in the fast and accurate methods for evaluation of the loss probability. The key idea behind them is the exploitation of the different types of correlation. For small buffers the positive correlations of the short time scale are dominant – hence we propose a method that captures this explicitly, related to the solution of [1]. For larger buffers the negative correlations essentially determine overflow behavior – also in this regime we come up with an explicitly tailored method, related to the $N{\cdot}$D/D/1 model and the Brownian bridge approximation.

This leads, for homogeneous input (i.i.d. deterministic on-off sources), to explicit formulae for the loss probability and the number of sources that can be admitted to achieve a given QoS criterion. It also allows us to tackle the problem of resource allocation for heterogeneous input, even for the case of service differentiation (per class a specific QoS criterion). The techniques for heterogeneous traffic are applicable under more general conditions. They can also be directly applied to set GPS (Generalized Processor Sharing) weights.

This paper is organized as follows. Section II presents preliminaries on the queue with deterministic on-off input. Section III and IV treat small buffers and large buffers, fed by homogeneous input, respectively. Section V illustrates the theory of the previous sections by means of numerical examples. In Section VI we focus on heterogeneous input; the corresponding numerical examples are provided in Section VII.

II. PRELIMINARIES

This section focuses on the preliminaries of our analysis. We will introduce the model (Section II-A), describe the state-of-the-art of the analysis of the model (Section II-B), and motivate our approach (Section II-C).

A. Model

We consider traffic from a number of determinsitic on-off sources arriving at a buffered resource. A source of type i has mean rate r_i, peak rate p_i and on-time T_i. There are N_i sources of type i. The resource is modeled as a queue with constant depletion rate C. An important case is when traffic is homogeneous: all sources have the same traffic descriptor (r, p, T). Notice that the only random effect is the 'phase' of the sources: the start of the on-time is uniformly distributed on the period of the source. We are interested in the probability of the buffer content exceeding level B, denoted by $L(B, C)$, or, equivalently, the probability that the delay exceeds $D = B/C$. To avoid trivialities, $\sum_i N_i r_i < C < \sum_i N_i p_i$.

B. Literature

This subsection reviews and evaluates the existing literature in the field of queues fed by adversarial traffic. We distinguish between three approaches: exact bounds, the virtual buffer-trunk construction by Elwalid *et al.* [10], and large deviations asymptotics.

Exact bounds. Kvols and Blaabjerg [16] consider the discrete-time version of our model – it is assumed that there are N homogeneous sources. Then the model is equivalently given, by on-time T, off-time T' and cell spacing D (i.e., during bursts every D time units a cell arrives). The well-known Beneš formula [22] then states (Q_t denoting the queue size at time t, assuming that at time t stationary behavior is reached) that $L(B, C)$ equals

$$\mathbb{P}(Q_t > B) = \sum_{s=1}^{\rho(T+T')-B} \mathbb{P}(N(t-s,t) = B+s)$$
$$\times \mathbb{P}(Q_{t-s} = 0 | N(t-s,t) = B+s).$$

Here ρ is the load of the multiplexer, and $N(s, t)$ is the number of arrivals from all N sources in time interval $[s, t]$. However, evaluating the above expression is hard, because no closed form expression for the conditional probability exists. Unfortunately, the trivial upper bound

$$L(B, C) \leq \sum_{s=1}^{\rho(T+T')-B} \mathbb{P}(N(t-s,t) = B+s)$$

is not very accurate. Therefore the authors also derive a (more accurate, but not necessarily conservative) approximation. However, even in the homogeneous case the method turns out to be quite complex (N-fold convolutions have to be determined explicitly). In Garcia, Barceló, and Casals [11] similar algorithms are found for the continuous-time (i.e., *fluid*) version of the model.

Virtual buffer-trunk construction. In Elwalid *et al.* [10], as a first step, a worst-case analysis of this model is treated assuming coincident bursts. Then it is easy to verify that if the solution of the following linear program is smaller than C there cannot be any loss:

$$\min \sum_{i=1}^{N} c_i, \quad \text{subject to} \quad \sum_{i=1}^{N} (p_i - c_i) T_i \leq B, \tag{1}$$

where $r_i \leq c_i \leq p_i$. This is an immediate consequence of the relation

$$\left\{ \sum_{i=1}^{N} c_i \leq C \right\} \Rightarrow \left\{ \text{Buffer contents} \leq \sum_{i=1}^{N} (p_i - c_i) T_i \right\} \tag{2}$$

Although the mathematical program (1) can be solved explicitly [4], [15], [17], the authors of [10] impose the additional requirement $b_i/c_i = B/C$, where $b_i = (p_i - c_i)T_i$. With this assumption, it is easy to show that there is no loss if the sum of the *effective bandwidths* is not larger than C, where the effective bandwidth c_i of a source with parameters (r_i, p_i, T_i) is given by

$$c_i := \max \left(r_i, p_i \left(1 + \frac{B}{CT_i} \right)^{-1} \right).$$

Since its solution satisfies the extra proportionality constraint, this gives an underestimate of the real admissible region. However, the crucial advantage is that the effective bandwidth of a source of a given class does not depend on the traffic mix, i.e., the characteristics of the other classes.

As argued in the introduction, in practice, it is not required that there is zero loss, but some small loss fraction ϵ is allowed, say 10^{-6}. In order to allow for such a small loss probability, Elwalid *et al.* [10] propose a statistical multiplexing routine. Clearly a source of type i requires c_i only a fraction r_i/c_i of time. So, at a random moment in time the required bandwidth by a type i source amounts to c_i with probability r_i/c_i and is zero else. For that reason, we have to investigate the probability

$$\mathbb{P}\left(\sum_{i=1}^{n} N_i C_i > C\right), \tag{3}$$

where C_i equals c_i with probability r_i/c_i, and is zero otherwise. The 'L-admissible region' then consists of all N_i combinations that keep the above probablity below L. In [10] it is proposed to use the Chernoff bound to approximate the above probability, thus deriving the corresponding admissible region.

The most important advantages of this method are its intuitive nature, its low complexity, and its applicability for the case of multiple types of traffic. However, the construction is sometimes overly pessimistic: $L(B, C)$ could be much smaller than (3). Two important reasons for that are: (i) During the period that the source needs bandwidth c_i, it does not need buffer space b_i all the time; the buffer usage has (as a function of time) a triangular shape. (ii) In the construction a source can never be served at a higher rate than the lossless effective bandwidth, even if there is enough bandwidth available to serve at higher rates, e.g., the peak rate. It is pointed out in [17] how to adapt the framework of [10] to overcome the former problem (in fact only partially, because there is still a union bound involved, that may turn out to be quite crude). However, as it does not solve the latter, the procedure of [17] is still conservative. The key point is that the construction of [10] loses detail by turning the dynamical queueing model into the static criterion (3).

Large deviations approach. A third approach is of an asymptotic nature, and relies on the theory of large deviations. First consider the homogeneous-input model We rescale the resources by the number of sources: $C \equiv Nc$ and $B \equiv Nb$. In the scaled model we define

$$L_N(b, c) := \begin{array}{l}\text{steady-state probability that the}\\ \text{buffer content exceeds level } Nb.\end{array}$$

From Simonian and Guibert [23], with $A(t)$ the amount of traffic generated by a single source in $[0, t)$,

$$I(b) := -\lim_{N\to\infty} \frac{1}{N} \log L_N(b, c) = \inf_{t>0} \sup_{\theta} \left(\theta(b + ct) - \log \mathbb{E}e^{\theta A(t)}\right). \tag{4}$$

We call this curve, i.e., the decay rate of the loss probability as function of b, the *loss curve,* see also [3], [6].

C. Approach

A general conjecture is that the loss curve is convex (concave) at a specific b, if the packet arrivals are negatively (positively) correlated on the time scale related to overflow. Empirical motivation for this conjecture can be found in Section 4.4 of Botvich and Duffield [3]. They consider a discrete-time queue fed by sources with geometric(q_1) on-times and geometric(q_2) of times. They found that depending on the correlation structure, the loss curve had a convex or concave shape. More precisely, for $q_1 + q_2 > 1$ (negative correlation) they showed convexity, for $q_1 + q_2 < 1$ concavity (positive correlation).

As said before, deterministic on-off fluid sources exhibit positive correlation on short time-scales and negative correlation on larger time-scales. These correlations can be visualized by examining the variance of $A(t)$. To calculate this variance, we first show how to compute the distribution of $A(t)$.

The distribution of $A(t)$ can be made explicit as follows. The period of the source is denoted by $S := T + T'$. We assume that $T < T'$, but an analogous reasoning applies to the case that $T \geq T'$. Because U (the epoch of the start of the burst) is uniformly distributed on $[0, S)$, three cases can be distinguished:

- $t < T$.
$$X(t) \stackrel{d}{=} \begin{cases} t - U & U \in [0, t] \\ 0 & U \in [t, T'] \\ U - T' & U \in [T', t + T'] \\ t & U \in [t + T', S) \end{cases}$$
- $T \leq t < T'$.
$$X(t) \stackrel{d}{=} \begin{cases} T & U \in [0, t - T] \\ t - U & U \in [t - T, t] \\ 0 & U \in [t, T'] \\ U - T' & U \in [T', S) \end{cases}$$
- $t \geq T'$.
$$X(t) \stackrel{d}{=} \begin{cases} T & U \in [0, t - T] \\ t - U & U \in [t - T, T'] \\ t - T' & U \in [T', t] \\ U - T' & U \in [t, S) \end{cases}$$

This enables the calculation of $\mathbb{E}\exp(\theta A(t))$, allowing us to evaluate (4), in principle. However, this is computationally involved, which motivates our search for fast and explicit approximations.

We are also in a position to calculate $\sigma^2(t) := p^2 \mathrm{Var} X(t)$:

$$\mathrm{Var} X(t) = \begin{cases} -\frac{t^3}{3S} + \frac{Tt^2}{S} - \frac{T^2t^2}{S^2}, & t \in [0, T) \\ -\frac{T^3}{3S} + \frac{T^2 t}{S} - \frac{T^2t^2}{S^2}, & t \in [T, T') \\ -\frac{T^3}{3S} + \frac{T^2 t}{S} + \frac{(t-T')^3}{3S} - \frac{T^2t^2}{S^2}, & t \in [T', S) \end{cases}$$

Between 0 and S, $\sigma^2(t)$ is first convex, then concave, and finally convex again; notice that concavity (convexity) indicates negative (positive) correlations.

The first part of the curve will be concave, due to the positive correlations on the short time scale, whereas the second part will be convex, because of the negative correlations on the somewhat longer time scale. The idea of this work is to develop specific approximations for both parts of the loss curve. For small values of b, we will use the 'square root' short-buffer approximation as given in Section III, in which explicitly the positive correlations are exploited. For larger values of b we will use an approximation based on the *Brownian bridge*, which naturally incorporates the negative correlations of the traffic offered (Section IV).

III. Approximation for small buffers, homogeneous input

In this section we review a result derived earlier, specifically for small buffers. As noted before, in this regime the positive correlations are dominant; in fact we will show that $I(b) \approx \alpha + \beta\sqrt{b}$. We will show that this approximation immediately leads to an expression of the number of sources to be admitted.

A. Approximation

First define

$$\alpha(c) := \frac{c}{p}\log\left(\frac{c}{r}\right) + \left(1-\frac{c}{p}\right)\log\left(\frac{p-c}{p-r}\right),$$

and $\beta(c) :=$

$$\frac{2}{p}\sqrt{\left(\frac{c}{T}+\frac{p-c}{T'}\right)\log\left(\frac{c}{p-c}\cdot\frac{T'}{T}\right) - 2\left(\frac{c}{T}-\frac{p-c}{T'}\right)}.$$

Mandjes and Kim [18] proved that $I(b)$ increases rapidly for small b as $\sqrt{b}$:

$$-I(b) = \lim_{N\to\infty}\frac{1}{N}\log L_N(b,c) = -\alpha(c) - \beta(c)\sqrt{b} + O(b). \tag{5}$$

Notice that this formula reflects the positive correlations: $I(b)$ is concave for small b. This leads to the loss approximation

$$L(B,C) \approx \exp\left(-N\cdot\alpha\left(\frac{C}{N}\right) - \beta\left(\frac{C}{N}\right)\sqrt{BN}\right). \tag{6}$$

B. Admissible region

For small values of B we expect that the number of admissible sources grows rapidly, suggested by (5). We will verify this property here. If $e^{-\delta}$ is the required loss fraction, due to (6) we have to solve

$$N_B\cdot\alpha\left(\frac{C}{N_B}\right) + \beta\left(\frac{C}{N_B}\right)\sqrt{BN_B} = -\delta;$$

here N_B is the number of admissible flows for a given buffer level B. Implicit differentiation with respect to B yields

$$\frac{\mathrm{d}N}{\mathrm{d}B}\left(\frac{C}{N}\alpha' - \alpha + \frac{C\sqrt{B}}{N\sqrt{N}}\beta' - \sqrt{\frac{B}{N}}\frac{\beta}{2}\right) = \sqrt{\frac{N}{B}}\frac{\beta}{2},$$

where $'$ denotes derivative with respect to the argument C/N. This yields $\sqrt{B}\cdot\mathrm{d}N/\mathrm{d}B \to K/2$, when $B \downarrow 0$, with

$$K := \sqrt{N_0}\cdot\beta\left(\frac{C}{N_0}\right)\left(\frac{C}{N_0}\cdot\alpha'\left(\frac{C}{N_0}\right) - \alpha\left(\frac{C}{N_0}\right)\right)^{-1}.$$

We notice that hence N_B grows quickly for small B: like $N_0 + K\sqrt{B}$; N_0 follows from $N_0\cdot\alpha(C/N_0) = -\delta$.

IV. Approximation for large buffers, homogeneous input

This sections deals with an approximation of the loss probability for large values of the buffer size.

A. Approximation

In the literature [8], [13], [20] it has been proposed to approximate the tail probabilities in the N·D/D/1 queue (i.e., the queue fed by N streams that emit a packet every, say, S units of time) by using a *Brownian bridge*. In this section we use a similar approximation for the queue fed by N deterministic on-off sources. We first explain the rationale behind the Brownian bridge approximation for the N·D/D/1 queue.

- Let $B(t)$ Brownian motion with drift $\mathbb{E}B(t) = \mu t$ and $\mathrm{Var}B(t) = \sigma^2 t$. It can be verified that, $t \in [0,S]$,

$$\mathbb{E}(B(t) \mid B(S) = \mu S) = \mu t;$$

$$\mathrm{Var}(B(t) \mid B(S) = \mu S) = \sigma^2\frac{t}{S}(S-t).$$

  We call $(B(t) \mid B(S) = \mu S)$ the Brownian bridge ($t \in [0,S]$).
- Now consider traffic generated by a single stream of the N·D/D/1 model. Let $\bar{A}(t)$ be the number of packets generated by such a stream in $[0,t)$. It is easy to check that $\mathbb{E}\bar{A}(t) = t/S$ and $\mathrm{Var}\bar{A}(t) = t(S-t)/S^2$. In other words, for $\mu = 1/S$ and $\sigma^2 = 1/S$, the first two moments of $\bar{A}(t)$ and the Brownian bridge coincide. In the central limit regime (i.e., when the load of the queue is relatively high), the mean and variance essentially determine the moment generating function of the input process:

$$\log\mathbb{E}[e^{\theta X(t)}] \approx \theta\mu t + \frac{1}{2}\theta^2\sigma^2(t) + O(\theta^3).$$

This justifies the use of the Brownian bridge as an approximation of the $N{\cdot}$D/D/1 queue. The probability distribution of the Brownian bridge is explicitly known [8], [13], [20].

In the case of deterministic on-off sources, $\mathrm{Var}A(t)$ is not proportional to $t(S-t)$, as we noted in Section II, but it *does* have the property that it drops to 0 at time S. This suggests that we could choose μ and σ^2 of the Brownian bridge to fit the mean and to conservatively bound the variance $\mathrm{Var}A(t)$ with a function proportional to $t(S-t)$. Then we apply the Brownian bridge on the corresponding process to get a conservative estimate.

It is not hard to see that we should choose $\mu = rS = pT$. The conservative (i.e., dominating) variance should satisfy $\sigma^2_{\mathrm{BB}}(t) \geq \sigma^2(t)$ for all $t \in [0,S]$. If $0 \leq T \leq S/2$, this may be done by choosing $\sigma^2_{\mathrm{BB}}(t) = \sigma^2 t(S-t)$ where

$$\sigma^2 = p^2\left(T^2 - \frac{4T^3}{3S}\right) =: p^2\tau^2,$$

having equality at $t = 0, S/2$, and T, and stricts inequality elsewhere (see Figure 1) . T has to be replaced by $S-T$ in the formula if $S/2 < T \leq S$.

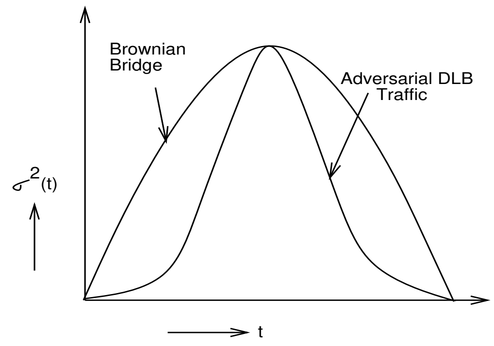

Fig. 1. Variance plots for deterministic on-off traffic and the closest Brownian bridge fit.

Let $B_i(\cdot)$ is a Brownian bridge with parameters $\mu_{\mathrm{BB}}(\cdot)$ and $\sigma^2_{\mathrm{BB}}(\cdot)$. With (i) because of the conservative choice of τ^2, and (ii) by direct application of the overflow probability formula for Brownian bridge [13], we obtain:

$$\begin{aligned} L(B,C) &= \mathbb{P}\left(\exists t \in [0,S] : \sum_{i=1}^{N} A_i(t) > B + Ct\right) \\ &\overset{\text{(i)}}{\lesssim} \mathbb{P}\left(\exists t \in [0,S] : \sum_{i=1}^{N} B_i(t) > B + Ct\right) \\ &\overset{\text{(ii)}}{=} \exp\left(-\frac{2B}{Np^2\tau^2}(B + CS - NpT)\right). \quad (7) \end{aligned}$$

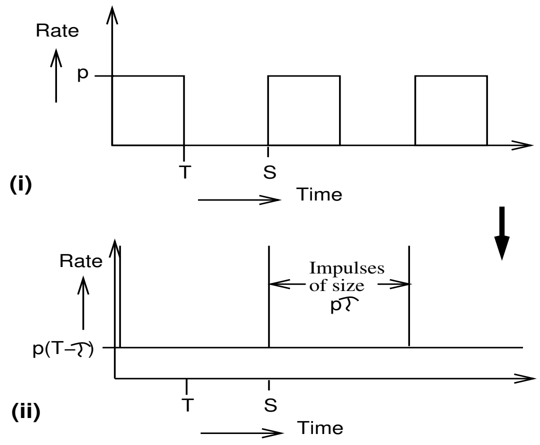

Fig. 2. (i) Single deterministic on-off source, which is (ii) conservatively replaced by a superposition of constant rate and periodic impulse.

B. Admissible region

For the $N{\cdot}$D/D/1 model with impulses of size $p\tau$ every S units of time, the Brownian bridge approximation would look like

$$\exp\left(-\frac{2B}{Np^2\tau^2}(B + CS - Np\tau)\right).$$

This means that approximation (7) has an interesting interpretation, as is illustrated in Figure 2. The procedure essentially bounds deterministic on-off source by an equivalent superposition of a constant rate and periodic impulse. The parameter $p\tau$, picked using the tightest variance fit to a Brownian bridge, determines an instantaneous burst that is generated every S units of time, whereas the difference $p(T-\tau)$ arrives at a constant rate. Evidently this is a significant improvement over replacing the entire square wave profile with impulses (corresponding to the choice $\tau = T$).

For a given loss probability $L = e^{-\delta}$, we obtain the maximum number of admissible sources N_B as

$$N_B = \frac{2B(B+CS)}{p^2\tau^2\delta + 2BpT} \qquad (8)$$

The results compare favorably with alternative methods, as will demonstrate numerically in the next section.

V. Numerical results for homogeneous traffic

We have described the computation of the overflow probability and admissible region for the regime of small buffers (Section III) and the regime of large buffers (Section IV). In this section both regimes are combined into a unified approximation.

In Figure 4, the number of admissible connections is shown, as a function of the buffer size. The traffic parameters of class 2 and 3 are taken from Figure 13 of [10]. The

Class	Mean(Mbps)	Peak(Mbps)	T(ms)	S(ms)	Loss
1	0.15	0.3	353	707	10^{-9}
2	0.15	1.5	70.7	707	10^{-9}
3	0.15	6.0	1.77	70.7	10^{-3}
4	0.15	20	0.53	70.7	10^{-3}

TABLE I
Traffic parameters for the experimental classes.

link rate is 150 Mbit/s. We chose the overflow probability of the last column; it is mentioned that other values give comparable graphs. Notice that the sources are ordered in increasing burstiness.

Exact and EMW. In both graphs we have plotted the number of admissible sources based on (4). As this approach is asymptotically exact (given that the number of sources is large), we call this curve *exact* and will serve as a benchmark throughout. Also, we show the acceptance curve based on [10] (denoted by *EMW* in the graphs). We see that this approximation performs quite well for small buffers, but loses its efficiency for larger buffer sizes. Apparently, the conservative elements identified in Section II-B play an important role.

Large and small buffer. Also the Brownian bridge-based large buffer approximation has been computed. If the buffer is large this approximation is close to the exact curve. Notice that there are parts of the curve in which the Brownian bridge is not conservative. Apparently, the conservative bound on the second moment of $A(t)$ is not enough here; higher moments play an important role (particularly when the load is relatively low: then the central limit regime does not apply). However, as seen from this graph and extensive other numerical experiments the difference between both curves is in general not very significant. On the other hand, we see that the small buffer approximation captures the exact loss curve quite well in the left hand side of the graphs.

Concave envelope. Having good approximations for large and small buffers, we consider a unified approximation: we merge the two curves by adding a linear segment such that we get a concave envelope. We see that this curve gives a good fit for all traffic classes, and is nearly always still conservative.

Algorithm by Guérin et al. [12]. In [12] traffic is modeled as on-off, with *exponentially distributed* on- and off-times. In other words, the model is different from ours, and hence the results are in principle not applicable to DLB regulated traffic (for instance, the burst size is unbounded). By fitting the mean on and off-times to T and T' we have compared their algorithm to our concave envelope and the exact curve. We found [12] too aggressive for small buffers, and significantly conservative for larger buffers.

VI. Heterogeneous traffic

We now consider the case of multiple classes of regulated sources sharing common resources C and B. The sources are homogeneous within each class, but may differ in traffic (leaky bucket) parameters across classes. The difference in leaky bucket parameters generally implies that the deterministic on-off profiles for each class may have different periods, on-times, and amplitudes. This fact makes it difficult to characterize the aggregate process, and hence also difficult to analyze performance. In particular, it is tempting to repeat the method of upper bounding the aggregate variance by a Brownian bridge. However, the Brownian bridge variance profile is a poor fit, since the aggregate variance is a superposition of periodic functions with possibly widely different periods.

As a conservative alternative, we consider the problem of optimally partitioning (B, C) into (B_i, C_i) among the classes, given a fixed number (N_i) of sources in each class. 'Optimal' here refers to maximizing the size of the admissible region by maximizing N for a given *connection mix vector* $\eta = \{\eta_1, ..., \eta_K\} := \{\frac{N_1}{N}, ..., \frac{N_K}{N}\}$, where $N := \sum_j N_j$. Partitioning of resources, which can support diverse QoS requirements by protecting individual classes, generally loses the multiplexing advantage obtained by sharing across classes. Thus for homogeneous QoS requirements, we get a conservative solution to the FIFO system. However, partitioning is still quite efficient in the practically interesting case of small number of classes K, each with large number of connections $N_i \gg K \; \forall \, i$.

We begin this section by formulating a general version of the partitioning problem, followed by a specialized solution for the Brownian bridge. We conclude by discussing the relevance of the partitions to GPS and shared buffer management.

A. Solution to the General Problem

Consider the admissible region specified by

$$R_i(C_i, B_i, N_i) = \delta_i := -\log L_i \tag{9}$$

for each class i with distinct traffic/QoS parameters. As can be verified, for given N_i, δ_i, equation (9) defines a convex B_i vs. C_i tradeoff curve, for the concave envelope introduced in Section V, as well as the small and large buffer approximations individually. To find the admissible region we then seek to solve

$$\begin{array}{rcl} \text{Maximize} & N & \text{subject to} \\ R_i(C_i, B_i, N\eta_i) & \geq & \delta_i \quad \forall\, i \\ \sum_i C_i \leq C, & & \sum_i B_i \leq B. \end{array} \tag{10}$$

We first observe that, at optimum, all the loss constraints in (10) would hold with equality, as otherwise we may reduce B_i and/or C_i for the corresponding class and admit more connections[1]. We hence suppose that, for fixed N (in particular the optimal value), we may invert (9) to obtain the buffer as a function of the other parameters, i.e., $B_i = B_i(C_i, N\eta_i, \delta_i)$. Now consider the intermediate problem:

$$\begin{aligned} \text{Minimize} \quad & \sum_i B_i(C_i, N\eta_i, \delta_i) \qquad \text{subject to} \\ \sum_i C_i \quad & \leq \quad C \end{aligned} \tag{11}$$

While we omit proofs and various details here due to limited space, readers familiar with optimization would recognize (11) as a standard convex minimization to which the *Strong Lagrangian Principles* (see, for example [27]) can be applied, which yields the following *Kuhn-Tucker* conditions:

$$-\frac{\partial B_i}{\partial C_i} = \frac{\partial R_i/\partial C_i}{\partial R_i/\partial B_i} = \lambda$$

for some global (class-independent) non-negative Lagrange multiplier λ. It hence follows that the maximum value of N retains feasibility of the following conditions for some λ, which represents the slope of each of the bandwidth-buffer tradeoff curves at the optimal operating point:

$$\begin{aligned} R_i(C_i, B_i, N\eta_i) \quad &= \quad \delta_i \\ -\frac{\partial B_i}{\partial C_i} \quad &= \quad \frac{\partial R_i/\partial C_i}{\partial R_i/\partial B_i} = \lambda \\ \sum_i C_i \leq C, \qquad & \qquad \sum_i B_i \leq B. \end{aligned} \tag{12}$$

Figure 3 pictorially illustrates the solution to the above optimization problem. The final pair of constraints on total buffer and bandwidth may be further augmented by a per-class delay constraint of the form $B_i \leq D_i C_i$ (illustrated in Figure 3(a)), or indeed any other jointly semi-concave constraints of the form $f(B_i, C_i) \geq 0$, without altering the solution procedure significantly. When the λ condition in (13) cannot be satisfied for all classes, which may happen when $\max -\frac{\partial B_i}{\partial C_i} < \lambda$ or $\min -\frac{\partial B_i}{\partial C_i} > \lambda$ within the operating region for some i (see Figure 3(b)), we may replace λ by the closest achieved value for class i, and choose the corresponding operating point (B_i, C_i) on the tradeoff curve. These modifications are justified by standard results from theory of convex programming. This implies that the optimal operating point for this class corresponds to either the mean rate (or the maximum delay limit when delay constrained) or the peak rate respectively. The operating point for each class i is hence uniquely specified for given λ when equation (9) satisfies the above described convexity property for each class.

[1] We ignore the fact that N needs to be integral, but this is of minor consequence when $N \gg 1$.

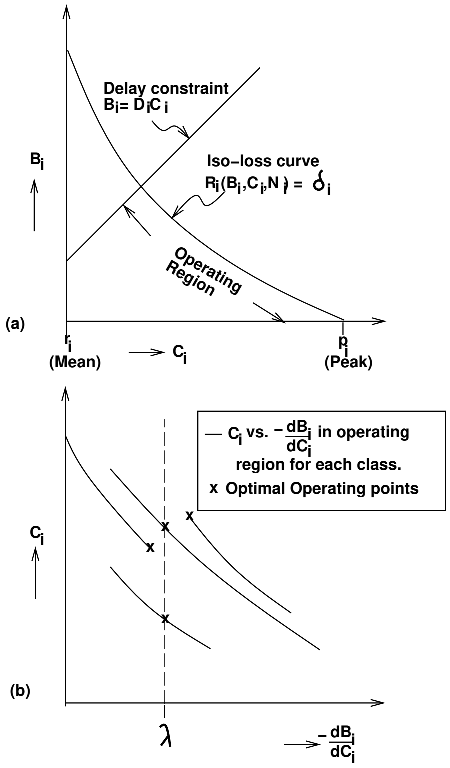

Fig. 3. Illustration of optimal partitioning solution.

We have thus reduced the multiclass problem to a two variable optimization on N, λ, which we solve using a simple bisection search procedure on λ for fixed N followed by an outer bisection on N. We use this general procedure to obtain the admissible region for the concave fit, but further simplification is possible if the Brownian bridge approximation is adequate, as explained in Section VI-B. Note that the partitioning solution presented here generalizes the lossless multiclass solution presented in [15], where the tradeoff curves $B_i(C_i)$ were linear for fixed N, η, δ_i.

General traffic types. It is evident from the above discussion that this procedure is applicable to all traffic types with convex bandwidth-buffer trade off curves, of which leaky bucket regulated traffic is an example. We believe that the convexity holds under more general circumstances, motivated by the supposition of 'diminishing returns' in bandwidth-buffer tradeoffs. This refers to the intuitive observation that, for a fixed number of connections and specified performance target, one obtains diminishing savings in bandwidth as buffer space is increased, and vice-versa. Note that the trade off curves could be obtained by measurements when traffic characteristics are not available explicitly.

Measurement-based resource allocation. The above observations can also be utilized to devise a control algorithm to achieve the optimal operating point without explicitly computing the admissible region. The per-class (B_i, C_i) trade-off curves, along with measurements of buffer and bandwidth utilizations, are the main drivers of this control. It is easy to see that the following simple proportional control would converge at the unique optimal value of λ, which necessarily exists if the problem is at all feasible: *Reduce (Increase) λ, if the buffer utilization is higher (lower) than bandwidth utilization, by an amount proportional to the absolute difference in the utilizations.* Reducing λ would decrease the B_i and increase C_i for every class, while increasing λ does the opposite. Hence the algorithm efficiently reallocates resources to achieve balanced buffer and bandwidth usage overall.

B. Brownian bridge

Applying the above described partitioning technique to this special case, we seek to maximize N subject to

$$\frac{2B_i}{N\eta_i p_i^2 \tau_i^2}(B_i + C_i S_i - p_i N \eta_i T_i) \geq \delta_i \quad \forall\, i$$
$$\sum_i C_i \leq C, \qquad \sum_i B_i \leq B. \tag{13}$$

The necessary convexity conditions are easily shown to hold here, and it turns out that these equations can be solved nearly in closed-form, (using techniques described in the Appendix) except for numerically solving for λ in

$$Cf(\lambda) - Bg(\lambda) = \frac{\alpha}{f(\lambda)}B^2 \tag{14}$$

where

$$\begin{aligned} f(\lambda) &= \sum_i \sqrt{\frac{\eta_i p_i^2 \tau_i^2 \delta_i}{2}} \sqrt{\frac{\lambda}{S_i - \lambda}} \\ g(\lambda) &= \sum_i \frac{1}{S_i}\sqrt{\frac{\eta_i p_i^2 \tau_i^2 \delta_i}{2}} \frac{S_i - 2\lambda}{\sqrt{\lambda(S_i-\lambda)}} \\ \alpha &= \sum_i \frac{\eta_i T_i P_i}{S_i}. \end{aligned}$$

The optimum value for N is given by $N = \frac{B}{f(\lambda)}$ for λ satisfying (14).

C. Relation to GPS and Buffer Management

We may use the optimal partitioning results to set GPS weights (see [9], [14] for GPS weight computation using the methods of [10]) and perform buffer management in an integrated fashion. Given the optimal (B_i, C_i) split, the GPS weight ϕ_i for each class i connection (assuming per connection queueing) can be set as $\phi_i = C_i/(N_i C)$. If the queueing is per-class, the weights would of course be $\Phi_i = C_i/C$ for class i. However, to guarantee performance requirements, the corresponding buffer space must be guaranteed as well. This can be accomplished using the technique of virtual partitioning as described in [15]. The per-VC (resp. per-class) *nominal* allocations described therein are easily recognized to be $b_i = \frac{B_i}{N_i}$ (resp. B_i) from the optimal partition. This approach hence naturally integrates GPS, buffer management and admission control in a heterogeneous setting, which has previously proved a challenge.

VII. Numerical Results for Heterogeneous Traffic

We now present the results of the optimal resource partitioning procedure for two traffic classes. The traffic parameters used in the results of Figure 5 correspond to classes 2 and 3 in Table I, with the loss requirements chosen as 10^{-9} and 10^{-3} respectively for classes 2 and 3. Figure 5 plots the admissible region using the different methods. The two bounding straight lines, provided for comparison, are obtained by joining corner points from [10] and the exact method (4), which do not admit diverse QoS. Note that the Brownian bridge and concave fit methods improve on EMW [10], are reasonable approximations for the 'exact' region, and are computationally simpler than these alternatives. The discrepancy between 'exact' and the concave fit observed in Figure 5(a) is mainly due to the fact that the buffer space assumed is not yet 'large' for class 3 connections, as seen by the improved fit in Figure 5(b) with a larger buffer. In particular, the Brownian bridge method is very fast and reasonably accurate when the cumulative buffer space is large and the delay constraint for individual classes is not too stringent. Our experiments also show close agreement between the concave fit and Brownian bridge methods in the optimal operating points (B_i, C_i) .

VIII. Conclusions

We have presented improved approximations for statistically multiplexing homogeneous regulated sources. These are used to develop techniques to optimally partition resources among heterogeneous classes. All the proposed procedures are fast, accurate, and simple.

In the intermediate buffer range, our concave fit differs appreciably from the exact results. Also our partitioning solution neglects possible inter-class sharing of resources, which could lead to significant inefficiency when the number of classes is large. These constitute directions for further research.

REFERENCES

[1] D. Anick, D. Mitra, and M. M. Sondhi. Stochastic Theory of a Data Handling System with Multiple Sources. *Bell Systems Technical Journal*, 61:1871–1894, 1982.

[2] Y. Bernet, J. Binder, S. Blake, M. Carlson, B. Carpenter, S. Keshav, E. Davies, B. Ohlman, D. Verma, Z. Wang, and W. Weiss. A framework for differentiated services. *Internet draft*, 1999.

[3] D. Botvich and N. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems*, 20:293–320, 1995.

[4] J.-Y. Le Boudec and A. Ziedins. A CAC Algorithm for VBR Connections over a VBR Trunk. In *Proc. ITC-15*, pages 59–70, 1997.

[5] C. Courcoubetis, F. Kelly, and R. Weber. Measurement-based charging in communications networks. *Preprint*, 1999.

[6] C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a buffer handling many traffic sources. *Journal of Applied Probability*, 33:886–903, 1996.

[7] B. Doshi. Deterministic rule based traffic descriptors for broadband ISDN: Worst case behavior and connection acceptance control. In *Proc. ITC-14*, pages 591–600, 1994.

[8] L. Dron, G. Ramamurthy, and B. Sengupta. Delay analysis of continuous bit rate traffic over an ATM network. *IEEE Journal on Selected Areas in Communications*, 9:402 – 407, 1991.

[9] A. Elwalid and D. Mitra. Design of Generalized Processor Sharing Schedulers Which Statistically Multiplex Heterogeneous QoS Classes. In *Proc. IEEE INFOCOM 1999*, pages 1220–1230, 1999.

[10] A. Elwalid, D. Mitra, and R. H. Wentworth. A New Approach for Allocating Buffers and Bandwidth to Heterogenous, Regulated Traffic in an ATM Node. *IEEE Journal on Selected Areas in Communications*, 13(6):1115–1127, Aug. 1995.

[11] J. Garcia, J. Barceló, and O. Casals. An exact model for the multiplexing of worst case traffic sources. In *Sixth International Conference on Data Communication Systems and their Performance*, 1995.

[12] R. Guérin, H. Ahmadi, and M. Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE Journal on Selected Areas in Communications*, 9:868–981, 1991.

[13] F. Kelly. Notes on Effective Bandwidths. In F. Kelly, S. Zachary, and I. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 141 – 168. Oxford University Press, 1996.

[14] K. Kumaran, G. Margrave, D. Mitra, and K. Stanley. Novel Techniques for the Design and Control of Generalized Processor Sharing Schedulers for Multiple QoS Classes. In *Proc. IEEE INFOCOM 2000*, 2000.

[15] K. Kumaran and D. Mitra. Performance and Fluid Simulations of a Novel Shared Buffer Management System. In *Proc. IEEE INFOCOM 1998*, pages 1449–1461, 1998.

[16] K. Kvols and S. Blaabjerg. Bounds and approximations for the periodic on-off queue with application to ATM traffic control. In *IEEE INFOCOM 1992*, 1992.

[17] F. LoPresti, Z. Zhang, D. Towsley, and J. Kurose. Source Time Scale Optimal Buffer/Bandwidth Trade-off for Regulated Traffic in an ATM Node. In *Proc. IEEE INFOCOM 1997*, pages 675–682, 1997.

[18] M. Mandjes and J.-H. Kim. Large deviations for small buffers: an insensitivity result. *To appear in: Queueing Systems*, 2000.

[19] L. Massoulié and A. Busson. Stochastic majorization of aggregates of leaky-bucket constrained traffic streams. *Preprint*, 1999.

[20] I. Norros, J. Roberts, A. Simonian, and J. Virtamo. The superposition of variable bit rate sources in an ATM multiplexer. *IEEE Journal on Selected Areas in Communications*, 9:378 – 387, 1991.

[21] S. Rajagopal, M. Reisslein, and K. Ross. Packet Multiplexers with Adversarial Regulated Traffic. *Preprint*, 1997.

[22] J. Roberts, U. Mocci, and J. Virtamo. *Broadband Network Teletraffic: Performance evaluation and design of broadband multiservice networks*. Springer, 1996.

[23] A. Simonian and J. Guibert. Large deviations approximation for fluid queues fed by a large number of on/off sources. *IEEE Journal of Selected Areas in Communications*, 13:1017–1027, 1995.

[24] J. Turner. New directions in communications (or which way to the information age). *IEEE Communications Magazine*, 1986.

[25] H. van den Berg and M. Mandjes. Admission Control in Integrated Networks: Overview, Evaluation and Research Perspectives. In *Proceedings 8th International Conference on Telecommunication Systems: Modelling and Analysis, Nashville TN*, 2000.

[26] W. Weiss. QoS with differentiated services. *Bell Labs Techical Journal*, 3:48–62, 1998.

[27] P. Whittle. *Optimization under Constraints*. John Wiley, 1971.

[28] J. Wroclawski. Specification of the controlled load network element service. *Internet Engineering Task Force RFC 2211*, 1997.

APPENDIX: SOLUTION TO THE MULTI-CLASS BROWNIAN BRIDGE PROBLEM

Following the results of section VI, we seek to maximize N subject to $\sum_i C_i \leq C, \quad \sum_i B_i \leq B$, and

$$\frac{2B_i}{N\eta_i p_i^2 \tau_i^2}(B_i + C_i S_i - N p_i \eta_i T_i) = \delta_i \quad \forall\, i. \qquad (15)$$

From equations (13) for fixed N,

$$\frac{dB_i}{dC_i} = -\frac{S_i}{1 + \frac{N\eta_i p_i^2 \tau_i^2 \delta_i}{2B_i^2}} = -\lambda,$$

for some global, class-independent parameter λ. Thus,

$$\begin{aligned} B_i &= \sqrt{\frac{N\eta_i p_i^2 \tau_i^2 \delta_i}{2}}\sqrt{\frac{\lambda}{S_i - \lambda}} \\ C_i &= \frac{1}{S_i}\left(N p_i \eta_i T_i - B_i + \frac{N\eta_i p_i^2 \tau_i^2 \delta_i}{2B_i}\right). \end{aligned}$$

Combining with $\sum_i C_i \leq C, \quad \sum_i B_i \leq B$, N and λ satisfy

$$N\sum_i \sqrt{\frac{\eta_i p_i^2 \tau_i^2 \delta_i}{2}}\sqrt{\frac{\lambda}{S_i - \lambda}} = B$$

$$N\sum_i \frac{\eta_i T_i P_i}{S_i} + \sqrt{N}\sum_i \sqrt{\frac{\eta_i p_i^2 \tau_i^2 \delta_i}{2}}\,\frac{1 - 2\lambda/S_i}{\sqrt{\lambda(S_i - \lambda)}} = C \quad (16)$$

Explicit elimination of N from the two equations, followed by some algebraic simplification, leads to the condition (14) for λ and the optimal solution by recovering N.

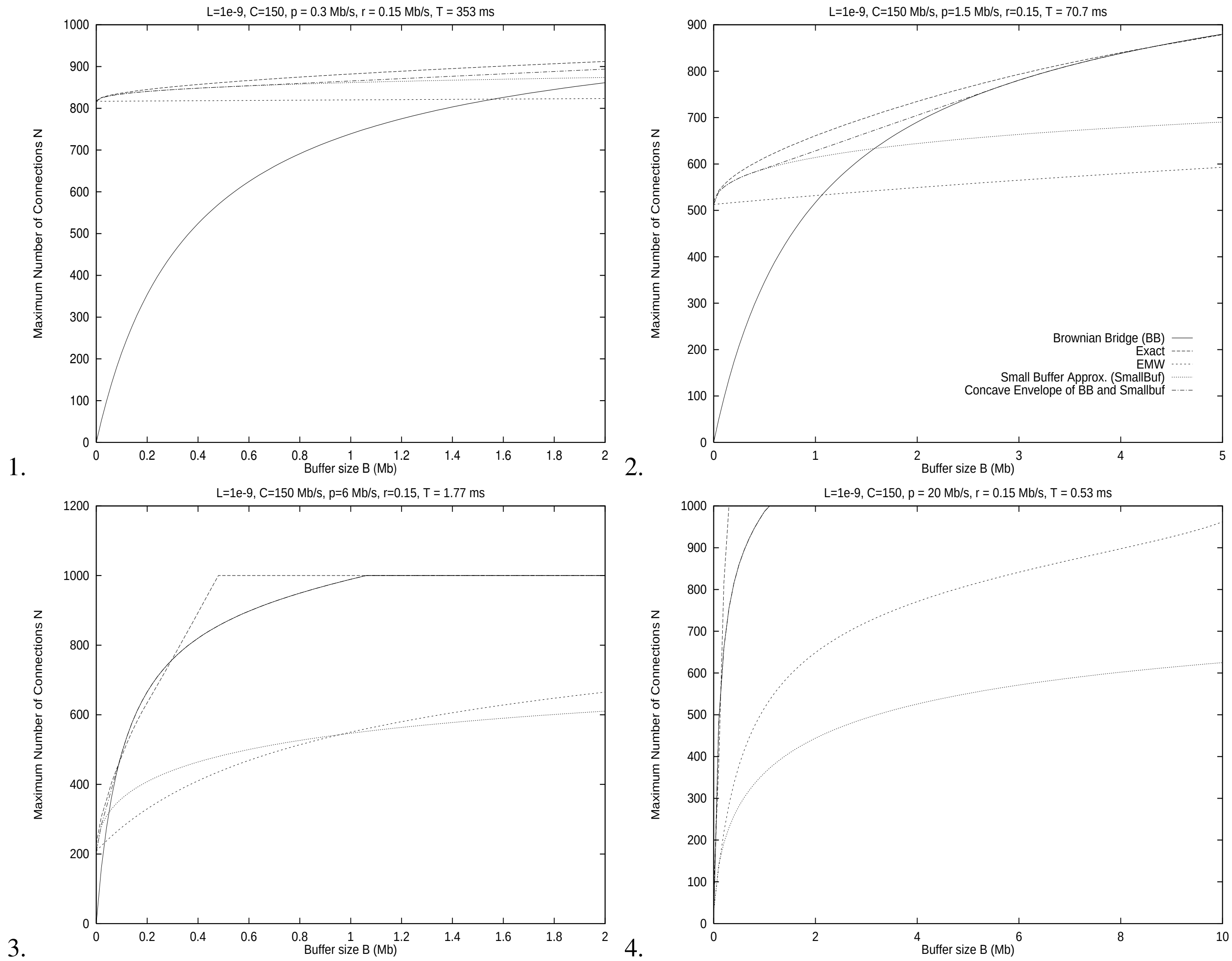

Fig. 4. Admissibility curves for the 4 connection types of Table I.

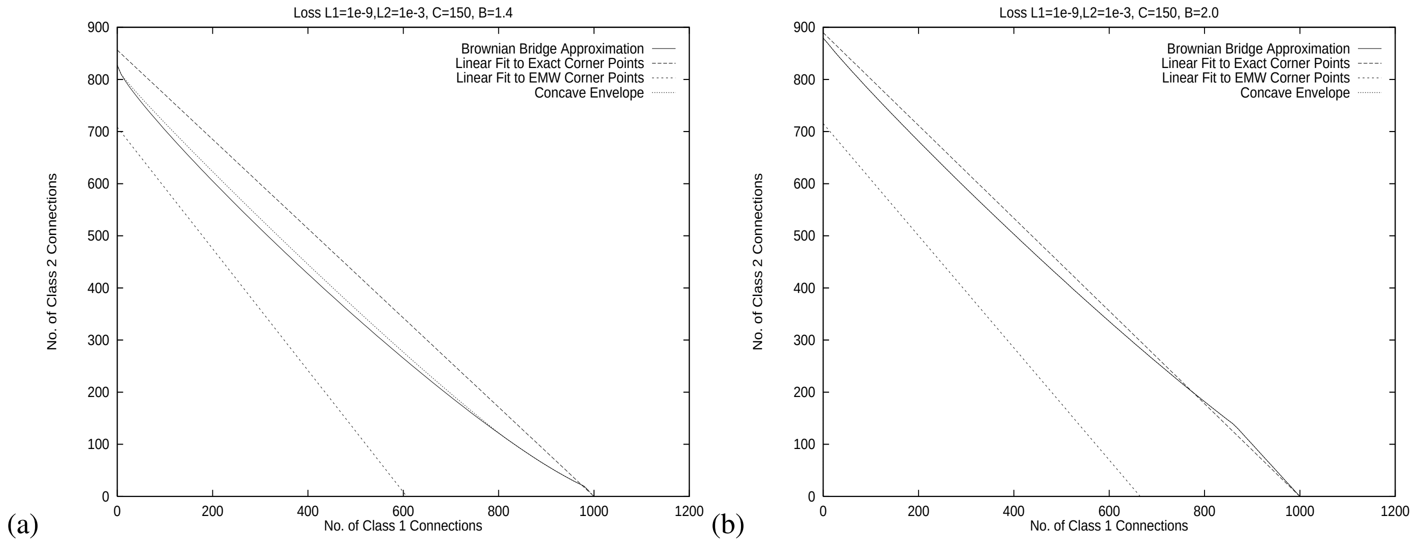

Fig. 5. Two-class admissible regions with (a) B=1.4 Mb (b) B=2.0 Mb.