Bounds for Independent Regulated Inputs Multiplexed in a Service Curve Network Element

Milan Vojnović and Jean-Yves Le Boudec

Abstract—**We consider the problem of bounding the probability of buffer overflow in a network node receiving independent inputs that are each constrained by arrival curves, but that are served as an aggregate. Existing results (for example [1] and [2]) assume that the node is a constant rate server. However, in practice, one finds various types of schedulers that do not provide a constant service rate, and thus to which the existing bounds do not apply. Now many schedulers can be adequately abstracted by a service curve property. We extend the results in [1] and [2] to such cases. As a by-product, we also provide a slight improvement to the bound in [2]. Our bounds are valid for both discrete and continuous time models.**

Keywords— **Statistical multiplexing, scheduling, queuing analysis, quality-of-service**

I. Introduction

BOUNDS on the probability of buffer overflow in a network node receiving independent inputs that are each constrained by arrival curves are obtained in [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11] under various assumptions. We say that a flow is regulated, or constrained, by an arrival curve $\alpha(\cdot)$ if the number of bits observed on the flow during any time interval of duration t is at most $\alpha(t)$. Leaky bucket regulation corresponds to an affine function $\alpha(\cdot)$. Existing results focus on work conserving queuing systems offering a constant service rate. However, in practice, one finds implementations with various types of schedulers that do not provide a constant service rate. It turns out that many such schedulers satisfy a service curve property [12], [13], [14], [15], [16]. A service curve property, with service curve β, means that at any time t, the total output traffic observed in $[0, t]$ is at least equal to $R(s) + \beta(t - s)$ for some s, where $R(s)$ is the total input traffic in $[0, s]$. Thus, it is of a practical importance to derive performance bounds for a service curve network element. We extend here the results by Kesidis and Konstantopoulos [1] on one hand, Chang, Song, and Chiu [2], on the other hand, to such cases. As a by-product, we also slightly improve the bound in [2], even for the case of a constant rate server.

Kesidis and Konstantopoulos [1] consider a constant rate server, and also assume that arrival curves are the combination of two leaky buckets (as is commonplace with ATM and in the Internet). In Section III (Theorem 1), we extend their results to a network node offering any arbitrary service curve, and to any arrival curve constraints. For this, we use a different proof, based on Little's formula and Hoeffding's bounds; it is simpler, even for the original case considered in [1].

Chang, Song, and Chiu [2] consider the same problem as Kesidis and Konstantopoulos, but allow for any arbitrary arrival curve. In Section IV (Theorem 3), we extend their result to a node offering a *super-additive* service curve. A function β is super-additive if $\beta(s+t) \geq \beta(s) + \beta(t)$ for all $s, t \geq 0$; convex functions such that $\beta(0) = 0$ are super-additive. Service curves used in the Internet usually have the form $\beta(t) = c \max[t-e, 0]$ ("rate-latency" service curves) and are super-additive, but some other service curves are not [17]. Extending [2] to a super-additive service curve is essentially a simple modification of the original proof; however, we also show how the proof can be linked to Hoeffding's bounds [18]. This allows us to derive other bounds for the heterogeneous case, as explained later. We also slightly improve the bound in [2] (even for the original case), using an *under-sampling* argument. Incidentally, this makes the bound valid in continuous time, whereas [2] considers the discrete time case.

Both [1] and [2] give explicit results for the homogeneous case (all arrival curves are identical) and leave the heterogeneous case as a non-trivial optimization problem. For both cases, we also give simple formulas that apply to the heterogeneous case (Theorems 2 and 4). Of course, the bounds for the heterogeneous case also apply to the homogeneous case, but they are not as tight; this feature is inherited from Hoeffding's bounds.

We also derive a variant for the heterogeneous case (Theorem 5), by exploiting the nature of the proof of Theorem 4. The bound in Theorem 4 (as Theorem 2) requires to know the arrival curves of all flows. In contrast, Theorem 5 requires only a limited knowledge about the arrival curves. For instance, for dual leaky-bucket constrained processes, it suffices to know the aggregate burstiness and sustainable rate. The bound is less tight than Theorem 4, but may be more useful in a context of differentiated services, where only aggregate information is available.

Chang, Song, and Chiu showed numerically that their bound is tighter than Kesidis and Konstantopoulos' bound. We confirm this, also by numerical tests, for our extensions: Theorems 3 and 4 seem to provide tighter bounds than Theorems 1 and 2, and should thus be preferred in practice. Section V shows a sample of numerical results. Another aspect would be to compare the bounds with simulations, but this goes beyond the scope of such a short paper, as it involves the issue of rare events and selection of arrival processes.

The proofs of two lemmas are given in Appendix.

II. Notation and Assumptions

Time is either continuous or discrete. Consider I arrival processes to a network element. Define $R_i(t)$, for $t > 0$, as the number of bits observed on input flow i in the interval $(0, t]$ and for $t \leq 0$, as the opposite of the number of bits observed on input flow i in the interval $(t, 0]$. Similarly, define $R_i^*(t)$ for the output of the ith flow. Let $R(t) := \sum_{i=1}^{I} R_i(t)$ and $R^*(t) := \sum_{i=1}^{I} R_i^*(t)$. We make the following assumptions:

(A1) $R_1, R_2, \ldots, R_I$ are independent.

The authors are with the Communication Systems Department, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. E-mail: {milan.vojnovic, jean-yves.leboudec}@epfl.ch.

(A2) R_i has α_i as an arrival curve, i.e.,

$$R_i(t) - R_i(s) \le \alpha_i(t-s), \text{ for all } s \le t$$

where α_i is a non-negative, wide-sense increasing function[1]. We assume, without loss of generality, that α_i is sub-additive, i.e. $\alpha(s+t) \ge \alpha(s) + \alpha(t)$ for all $s, t \ge 0$ [13], [14], [15], [16].
(A3) R_i and R_i^* have stationary increments and are ergodic.
(A4) The network element offers the service curve β to the aggregate of all flows:

$$\forall t \in \mathbb{R},\ \exists s \le t \text{ such that } R^*(t) - R(s) \ge \beta(t-s)$$

where β is a non-negative wide-sense increasing function.

Define the sustainable rate of flow i by

$$\rho_i = \lim_{t\to\infty} \frac{\alpha_i(t)}{t} = \inf_{t>0} \frac{\alpha_i(t)}{t}$$

The last equality comes from the sub-additivity of α_i [19]. It can readily be seen from (A2) and (A3) that

$$\mathsf{E}[R_i(1) - R_i(0)] \le \rho_i \quad (1)$$

Let $\rho = \sum_{i=1}^I \rho_i$ and $\alpha(t) = \sum_{i=1}^I \alpha_i(t)$.

For two functions f and g, we define the *vertical* and *horizontal deviations* by [16]

$$v(f,g) = \sup_{t\ge 0}\{f(t) - g(t)\}$$

and

$$h(f,g) = \sup_{t \ge 0}\{\inf\{u \ge 0 | f(t) \le g(t+u)\}\}.$$

Note that $v(f,g)$ is the worst case backlog for a network element offering the service curve g to an input flow that has f as an arrival curve. Similarly, $h(f,g)$ is the worst case virtual delay (equal to the worst case delay if the node would be FIFO).

We also define function λ_a by $\lambda_a(t) = at$ for $t \ge 0$ and $\lambda_a(t) = 0$ for $t < 0$.

Call $Q(t)$, the backlog of the network element at time t; $Q(t) = R(t) - R^*(t)$. From (A4), it follows that

$$Q(t) \le \sup_{s \le t}\{R(t) - R(s) - \beta(t-s)\}. \quad (2)$$

By (A3), the distribution of $Q(t)$ is independent of t. In the next two sections we give upper bounds to $\mathsf{P}(Q(0) > b)$.

III. EXTENDING KESIDIS AND KONSTANTOPOULOS' BOUND

We extend [1] in the following two theorems, the proofs of which are given at the end of this section.

Theorem 1 (Homogeneous Case) Suppose (A1)–(A4) and $\alpha_i = \alpha_1$, for all $i = 1$ to I. Then, for $\rho h(\alpha, \beta) < b < v(\alpha,\beta)$,

$$\mathsf{P}(Q(0) > b) \le \exp\left(-I\frac{b}{v}\ln\frac{b}{\rho h} + I\left(1 - \frac{b}{v}\right)\ln\frac{v - \rho h}{v - b}\right)$$

where for brevity $v = v(\alpha,\beta)$ and $h = h(\alpha,\beta)$.

[1] We say that function $\alpha(\cdot)$ is wide-sense increasing if $s \le t$ always implies $\alpha(s) \le \alpha(t)$. This is also called “non-decreasing”.

Note that $v(\alpha,\beta)$ is the required buffer size ensuring a loss-free operation, while $\rho h(\alpha,\beta)$ is the product of the utilization and the maximum virtual delay. For $b \le \rho h(\alpha,\beta)$ we have the trivial bound $\mathsf{P}(Q(0) > b) \le 1$. For $b \ge v(\alpha,\beta)$, $\mathsf{P}(Q(0) > b) = 0$.

We can apply Theorem 1 to the original case in [1] by letting $\alpha_1(t) = \min(\pi_1 t, \rho_1 t + \sigma_1)$ and $\beta(t) = ct$. The bound in Theorem 1 is obtained by computing $\sup_{\theta > 0} F(\theta)$, where

$$F(\theta) = \theta b - I\ln\left(1 - \frac{\rho_1}{c} + \frac{\rho_1}{c}e^{\theta\frac{\pi_1 - c}{\pi_1 - \rho_1}\sigma_1}\right)$$

which is exactly the result in Theorem 1 of [1]; this shows that we do have an extension of [1].

Next, we provide a looser bound than in Theorem 1, but which holds for the heterogeneous case.

Theorem 2 (Heterogeneous Case) Suppose (A1)–(A4). Then, for $\rho h(\alpha,\beta) < b < v(\alpha,\beta)$,

$$\mathsf{P}(Q(0) > b) \le \exp\left(-\frac{2(b - \rho h(\alpha,\beta))^2}{\inf_{\underline{\gamma}\in\mathcal{C}}\sum_{i=1}^I v(\alpha_i, \gamma_i\beta)^2}\right) \quad (3)$$

where $\underline{\gamma} = (\gamma_1, \ldots, \gamma_I)$, $\gamma_i \ge 0$, for all $i = 1$ to I, and $\mathcal{C} = \{\underline{\gamma} | \sum_{i=1}^I \gamma_i \le 1\}$.

Proof: [Theorem 1] Define

$$Q_i(t) = \sup_{s \le t}\{R_i(t) - R_i(t-s) - \gamma_i\beta(t-s)\}$$

for any $\gamma_i \ge 0$ such that $\sum_{i=1}^I \gamma_i \le 1$. From (2)

$$Q(t) \le \sum_{i=1}^I Q_i(t).$$

thus

$$\mathsf{P}(Q(0) > b) \le \mathsf{P}(\sum_{i=1}^I Q_i(0) > b). \quad (4)$$

First, note that (A1) implies

$$Q_1(0), Q_2(0), \ldots, Q_I(0) \text{ are independent}. \quad (5)$$

Second,

$$0 \le Q_i(0) \le v(\alpha_i, \gamma_i\beta). \quad (6)$$

Third, by Little's law $\mathsf{E}[Q(0)] \le \mathsf{E}[R(1) - R(0)]\mathsf{E}^0[V_0]$, where $\mathsf{E}^0(V_0)$ is the expected sojourn time seen by any arbitrary bit if the system would be FIFO. Obviously $\mathsf{E}[V_0] \le h(\alpha,\beta)$; by (1):

$$\mathsf{E}[Q(0)] \le \rho h(\alpha,\beta). \quad (7)$$

Let $\gamma_i = \frac{1}{I}$. By (4)-(6) and using (4.5) in the proof of Hoeffding's inequality ([18], Theorem 1), we obtain that for any $\theta > 0$

$$\mathsf{P}(Q(0) > b) \le$$
$$\le e^{-\theta b}\left(1 - \frac{\mathsf{E}[Q(0)]}{v(\alpha,\beta)} + \frac{\mathsf{E}[Q(0)]}{v(\alpha,\beta)}e^{\theta v(\alpha_1, \beta/I)}\right)^I$$

This bound is increasing with $\mathsf{E}[Q(0)]$. Thus, by (7)

$$\mathsf{P}(Q(0) > b) \le \exp\left(-\sup_{\theta > 0} F(\theta)\right)$$

where

$$F(\theta) = b\theta - I \ln \left(1 - \rho \frac{h(\alpha,\beta)}{v(\alpha,\beta)} + \rho \frac{h(\alpha,\beta)}{v(\alpha,\beta)} e^{\theta \frac{v(\alpha,\beta)}{I}}\right). \quad (8)$$

Computing $\sup_{\theta>0} F(\theta)$ yields the desired result. ■

Note that we could immediately apply Hoeffding's inequality ([18], Theorem 1) to (4)-(6), and then use (7). However, the last part of the proof is given for the sake of comparison with [1].

Proof: [Theorem 2] The proof builds upon the proof of Theorem 1. Given (4)-(6), the problem is equivalent to deriving an upper-bound to the complementary distribution (4) of a summation of independent *non-uniformly* bounded random variables. From Hoeffding's inequality ([18], Theorem 2) it follows that

$$\mathsf{P}(Q(0) > b) \leq \exp\left(-\sup_{\underline{\gamma}\in\mathcal{C}} \frac{2(b - \mathsf{E}[Q(0)])^2}{\sum_{i=1}^{I} v(\alpha_i, \gamma_i\beta)^2}\right).$$

The latter bound, for $b > \mathsf{E}[Q(0)]$, is wide-sense increasing in $\mathsf{E}[Q(0)]$. Given (7), for $b > \rho h(\alpha,\beta)$, the inequality in (3) holds, which completes the proof. ■

IV. Extending Chang, Song, and Chiu's bound

We extend [2] in three theorems, the proofs of which are given at the end of this section.

Assume in addition to (A1)–(A4) that

(A5) β is super-additive

For any integer K and for any $t \geq 0$, let $\mathcal{S}_K(t)$ be the set of partitions of $[0,t]$ in K intervals, in other words

$$\mathcal{S}_K(t) = \{(s_0, s_1, ..., s_K) | 0 = s_0 \leq s_1 \leq ... \leq s_K = t\}$$

(if time is discrete, we require that s_k is integer for all k). Also define

$$\tau = \inf\{t \geq 0 | \alpha(t) \leq \beta(t)\}. \quad (9)$$

Theorem 3 (Homogeneous Case) Suppose (A1)–(A5) and $\alpha_i = \alpha_1$, for all $i = 1$ to I. Then,

$$\mathsf{P}(Q(0) \geq b) \leq \inf_{K\in\mathbb{N}, \underline{s}\in\mathcal{S}_K(\tau)} \sum_{k=0}^{K-1} \exp(-I g_k(\underline{s})) \quad (10)$$

where

$$g_k(\underline{s}) = \begin{cases} +\infty, \; b > \alpha(s_{k+1}) - \beta(s_k) \\ 0, \; b < \rho s_{k+1} - \beta(s_k) \\ \frac{\beta(s_k)+b}{\alpha(s_{k+1})} \ln \frac{\beta(s_k)+b}{\rho s_{k+1}} + \\ + \left(1 - \frac{\beta(s_k)+b}{\alpha(s_{k+1})}\right) \ln \frac{\alpha(s_{k+1})-\beta(s_k)-b}{\alpha(s_{k+1})-\rho s_{k+1}}, \; \text{otherwise} \end{cases}$$

If time is discrete, and we let $\beta(s) = cs$, $K = t$, $s_k = k$, then Theorem 3 gives the same bound as [2]. However, even for the original scenario in [2], we have a slight improvement: if τ is large (which may happen simply because our time unit is very small), we expect the bound in [2] to be large, because it relies on the union bound. We expect to have a better bound by allowing K to be smaller than τ (under-sampling). This is verified in Section V. Note that the theorem implies that for any K and $\underline{s} \in \mathcal{S}_K(\tau)$, the right hand-side in (10) *is* a bound.

Next, we provide a looser bound than in Theorem 1, but which holds for the heterogeneous case.

Theorem 4 (Heterogeneous Case) Suppose (A1)–(A5). Then, for $b < v(\alpha,\beta)$,

$$\mathsf{P}(Q(0) > b) \leq \inf_{K\in\mathbb{N}, \underline{s}\in\mathcal{S}_K(\tau)} \sum_{k=0}^{K-1} \exp(-g_k(\underline{s})) \quad (11)$$

where

$$g_k(\underline{s}) = \frac{2(b + \beta(s_k) - \rho s_{k+1})^2}{\sum_{i=1}^{I} \alpha_i(s_{k+1})^2}$$

for $b > \rho s_{k+1} - \beta(s_k)$, and $g_k(\underline{s}) = 0$, otherwise.

We can exploit the proof of the above theorems and derive an additional bound for the heterogeneous case that requires only aggregate information about the arrival curves. We obtain this by using a convenient majorization (similar to [5] for leaky-bucket constrained processes).

Theorem 5 (Heterogeneous Case) Suppose (A1)–(A6). Then, for $b < v(\alpha,\beta)$,

$$\mathsf{P}(Q(0) > b) \leq \inf_{K\in\mathbb{N}, \underline{s}\in\mathcal{S}_K(\tau)} \sum_{k=0}^{K-1} \exp(-g_k(\underline{s})) \quad (12)$$

where

$$g_k(\underline{s}) = \frac{(b + \beta(s_k) - \rho s_{k+1})^2}{2\sum_{i=1}^{I} v(\alpha_i, \lambda_{\rho_i})^2}$$

for $b > \rho s_{k+1} - \beta(s_k)$, and $g_k(\underline{s}) = 0$, otherwise.

The proofs of the above theorems requires two lemmas, proved in appendix. The proof of Lemma 1 extends a thought in [2].

Lemma 1: Under (A2), (A4), and (A5), it holds

$$Q(0) \leq \sup_{0\leq s\leq\tau} \{R(0) - R(-s) - \beta(s)\}.$$

Lemma 2: $\mathsf{P}(Q(0) > b) \leq$

$$\inf_{K\in\mathbb{N}, \underline{s}\in\mathcal{S}_K(\tau)} \{\sum_{k=0}^{K-1} \mathsf{P}(R(s_{k+1}) - R(0) > b + \beta(s_k))\}. \quad (13)$$

Proof: [Theorem 3] Note that the term in the summation in (13) is the complementary distribution of a sum of independent uniformly bounded random variables $0 \leq R_i(s_{k+1}) - R_i(0) \leq \alpha_1(s_{k+1})$. By Hoeffding's inequality ([18], Theorem 1), and $\mathsf{E}[R_i(s_{k+1})] \leq \rho_i s_{k+1}$, it is upper-bounded by $\exp(-I g_k(\underline{s}))$, for $b > \rho s_{k+1} - \beta(s_k)$. Combining with Lemma 2 completes the proof. ■

Proof: [Theorem 4] By Hoeffding's inequality ([18], Theorem 2), the term in the summation in (13) is upper-bounded with

$$\exp\left(-\frac{2(b + \beta(s) - \mathsf{E}[R(s) - R(0)])^2}{\sum_{i=1}^{I} \alpha_i^2(s)}\right)$$

For $b > \mathsf{E}[R(s) - R(0)] - \beta(s)$, the latter bound is wide-sense increasing with $\mathsf{E}[R(s) - R(0)]$. Thus, given $\mathsf{E}[R(s) - R(0)] \leq \rho s$, it is sufficient that $b > \rho s_{k+1} - \beta(s_k)$ for (11) to hold. Combining with Lemma 2 completes the proof. ■

Proof: [Theorem 5] Define

$$\tilde{Q}_i(t) = \sup_{s\leq t}\{R_i(t) - R_i(s) - \rho_i(t-s)\}$$

and $Z_i(t) = \tilde{Q}_i(t) - \tilde{Q}_i(0)$. For $s \le t$,

$$\tilde{Q}_i(t) - \tilde{Q}_i(s) \ge R_i(t) - R_i(s) - \rho_i(t-s).$$

Thus, $Z_i(t) \ge R_i(t) - R_i(0) - \rho_i t$. By a stochastic majorization

$$\begin{aligned}&\mathsf{P}(\textstyle\sum_{i=1}^{I} R_i(s) - R_i(0) - \rho_i s > r) \le \mathsf{P}(\sum_{i=1}^{I} Z_i(s) > r) \le \\ &\le \exp\left(-\frac{r^2}{2\sum_{i=1}^{I} v(\alpha_i, \lambda_{\rho_i})^2}\right)\end{aligned} \tag{14}$$

where the latter inequality is by applying Hoeffding's inequality ([18], Theorem 2) for a summation of independent zero-mean non-uniformly bounded random variables $-v(\alpha_i, \lambda_{\rho_i}) \le Z_i(t) \le v(\alpha_i, \lambda_{\rho_i})$. Combining (14) with Lemma 2 and a simple substitution complete the proof. ■

V. Numerical Comparison of Bounds

We consider arrival processes constrained with dual leaky-buckets, $\alpha_i(t) = \min(\pi_i t, \rho_i t + \sigma_i)$, traversing a network element offering the rate-latency service curve $\beta(t) = c\max(t - e, 0)$, with $c = 150$ Mbps, $e =$MTU$/c$ and MTU=1500 Bytes. We compute numerical results for both homogeneous and heterogeneous cases. For the bounds in Section IV, we take a uniform time partition $s_k = \frac{k}{K}\tau$, for $k = 0$ to K, and take the minimum over all $K \in \mathbb{N}$.

For the homogeneous case, we set $\rho_1 = \alpha c/I$, $0 < \alpha < 1$, $\pi_1 = 10\rho_1$, and $\sigma_1 = 8$ MTU. The numerical results are shown in Fig. 1 for $I = 100$, $\alpha = 0.2$ and 0.8. For the heterogeneous case, we suppose two classes of the arrival processes each of which consisting of I_1 and I_2 arrival processes, respectively. Consider i and j, a class-1 and class-2 arrival process, respectively. Then, we suppose $\rho_i = 2\rho_j$, $I_1\rho_i + I_2\rho_j = \alpha c$, $0 < \alpha < 1$, $\pi_i = 10\rho_i$, $\pi_j = 20\rho_j$, $\sigma_i = 8$ MTU, and $\sigma_j = 5$ MTU. The numerical results are shown in Fig. 2 for $I_1 = I_2 = 50$, $\alpha = 0.2$ and 0.8.

We did many numerical experiments similar to those reported above. We find first that the extensions to Chang, Song, and Chiu's bound (excluding Theorem 5, which should be handled separately), is substantially tighter than the extensions to Kesidis and Konstantopoulos' bound. This confirms a similar observation in [2]. Second, the bound in Theorem 3 becomes tighter as we optimize with respect to K; this slightly improves upon [2]. Lastly, the bound in Theorem 5 is loose in some cases; it becomes tighter at heavy load.

VI. Concluding Remark

Note that all the bounds in this paper, and thus the original bounds in [1] and [2], are applications of Hoeffding's inequalities [18].

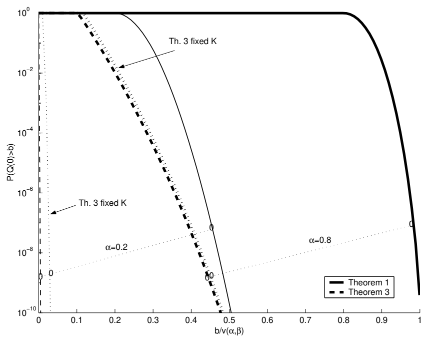

Fig. 1. Bounds for the homogeneous case with $I = 100$, $v(\alpha,\beta) = 445.44$ MTU for $\alpha = 0.2$ (thin lines), and $v(\alpha,\beta) = 778.78$ MTU for $\alpha = 0.8$ (thick lines). The dotted curves show the bound of Theorem 3 with fixed $K = \lceil \tau/e \rceil$.

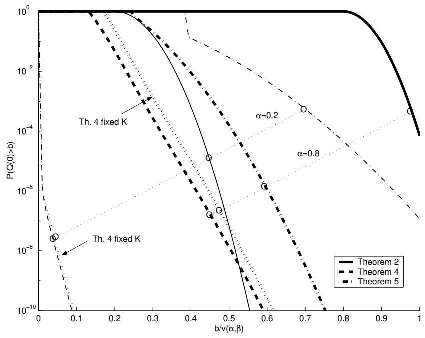

Fig. 2. Bounds for the heterogeneous case with $I_1 = I_2 = 50$, $v(\alpha,\beta) = 384.33$ MTU for $\alpha = 0.2$ (thin lines), and $v(\alpha,\beta) = 634.33$ MTU for $\alpha = 0.8$ (thick lines). The dotted curves show the bound of Theorem 4 with fixed $K = \lceil \tau/e \rceil$.

References

[1] George Kesidis and Takis Konstantopoulos, "Worst-case performance of a buffer with independent shaped arrival processes," *IEEE Communication Letters*, 2000.

[2] Cheng-Shang Chang, Wheyming Song, and Yuh ming Chiu, "On the performance of multiplexing independent regulated inputs," in *Proc. of Sigmetrics 2001*, Massachusetts, USA, May 2001.

[3] George Kesidis and Takis Konstantopoulos, "Extremal shape-controlled traffic patterns in high-speed networks," *IEEE Trans. on Communications*, vol. 48, no. 5, pp. 813–819, May 2000.

[4] Cheng-Shang Chang, *Performance Guarantees in communication networks*, Springer-Verlag, 2000.

[5] Laurent Massoulie and Anthony Busson, "Stochastic majorization of aggregates of leaky bucket-constrained traffic streams," submitted, http://www.research.microsoft.com/users/lmassoul/, 2000.

[6] A. Elwalid, D. Mitra, and R. H. Wentworth, "A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an atm node," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1115–1127, 1995.

[7] F. Lo Presti, Z. Zhang, D. Towsley, and J. Kurose, "Source time scale optimal buffer/bandwidth trade-off for regulated traffic in an atm node," *IEEE Trans. on Networking*, vol. 7, no. 4, pp. 490–501, August 1999.

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

[9] Srini Rajagopal, Martin Reisslein, and Keith W. Ross, "Packet multiplexers with adversarial regulated traffic," in *Proc. of IEEE INFOCOM 1998*, 1998, pp. 347–355.

[10] Robert Boorstyn, Almut Burchard, Jorg Liebeherr, and Chaiwat Oottamakorn, "Statistical service assurances for traffic scheduling algorithms," *Journal of Selected Areas in Communications*, vol. 18, no. 12, pp. 2651–2664, December 2000.

[11] Krishnan Kumaran and Michel Mandjes, "Multiplexing regulated traffic streams: Design and performance," in *Proc. of IEEE INFOCOM 2001*, March 2001.

[12] 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. on Networking*, vol. 1-3, pp. 344–357, June 1993.

[13] Jean-Yves Le Boudec, "Application of network calculus to guaranteed service network," *IEEE Trans. on Information Theory*, vol. 44, pp. 1087–1096, May 1998.

[14] C. S. Chang, "On deterministic traffic regulation and service guarantee: A systematic approach by filtering," *IEEE/ACM Trans. on Networking*, vol. 44, pp. 1096–1107, August 1998.

[15] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan, "Performance bounds

for flow control protocols," *IEEE/ACM Trans. on Networking*, vol. 7, no. 3, pp. 310–323, June 1999.
[16] J.-Y. Le Boudec and P. Thiran, *Network Calculus*, Springer-Verlag, 2001.
[17] H. Sariowan, *A Service Curve Approach to Performance Guarantees in Integrated Service Networks*, Ph.D. thesis, UCSD, 1996.
[18] Wassily Hoeffding, "Probability inequalities for sums of bounded random variables," *American Statistical Association Journal*, pp. 13–30, March 1963.
[19] J. F. C. Kingman, "Subadditive processes," in *Ecole d'été de probabilité de Saint-Flour*, vol. Lecture Notes in Mathematics (539), pp. 165–223. Springer Verlag, 1976.

APPENDIX

I. PROOF OF LEMMA 1

Proof: Fix $t \geq \tau$. Let $s \leq t$ and $u = \lfloor \frac{t-s}{\tau} \rfloor \tau + s$. Then,

$$\alpha(u-s) \leq \frac{u-s}{\tau}\alpha(\tau) \leq \beta(u-s) \tag{15}$$

where the former inequality is due to (A5), and the latter inequality comes from $\alpha(\tau) \leq \beta(\tau)$ (9), and (A6).

Next, observe, for $s < t - \tau$

$$\begin{aligned}
&R(t) - R(s) - \beta(t-s) \\
&\leq R(t) - R(u) + R(u) - R(s) - \beta(t-u) - \beta(u-s) \\
&\leq R(t) - R(u) - \beta(t-u) \\
&\leq \sup_{t-\tau \leq u \leq t}\{R(t) - R(u) - \beta(t-u)\}
\end{aligned}$$

where the first inequality is due to (A6), and the second holds since from (A2) and (15) $A(u) - A(s) - \beta(u-s) \leq 0$. Finally, it follows that (2) degenerates to

$$Q(t) \leq \sup_{t-\tau \leq s \leq t}\{R(t) - R(s) - \beta(t-s)\}$$

and a simple variable substitution completes the proof. ■

II. PROOF OF LEMMA 2

Proof: Fix $K \in \mathbb{N}$ and $0 = s_0 \leq s_1 \leq \ldots \leq s_K = \tau$. Note, for $s_k \leq s < s_{k+1}$

$$R(-s) \geq R(-s_{k+1}) \text{ and } \beta(s) \geq \beta(s_k)$$

Thus

$$\begin{aligned}
&\sup_{0 \leq s \leq \tau}\{R(0) - R(-s) - \beta(s)\} = \\
&= \sup_{k \in \{0,\ldots,K-1\}} \sup_{s_k \leq s \leq s_{k+1}}\{R(0) - R(-s) - \beta(s)\} \\
&\leq \sup_{k \in \{0,\ldots,K-1\}}\{R(0) - R(-s_{k+1}) - \beta(s_k)\}
\end{aligned}$$

Combining the latter with Lemma 1 and the sub-σ-additivity property we obtain

$$\mathsf{P}(Q(0) > b) \leq \sum_{k=0}^{K-1} \mathsf{P}(R(s_{k+1}) - R(0) > b + \beta(s_k))$$

Since the latter inequality holds for any partition $0 \leq s_1 \leq \ldots \leq s_K = \tau$, we obtain (13), which completes the proof. ■