V. Conclusion

This paper presents a sufficient condition for the *DHCD* problem for nonlinear systems. The resulting controller guarantees local asymptotic stability and provides a predetermined L_2-gain bound on the closed-loop system. Two design methods of the local observers are given: one is based on the centralized observer gain and another one is related to the solution of the matrix inequalities. The results are extensions of those in [8] and [13] for the case of linear systems.

References

[1] J. Ball, J. W. Helton, and M. L. Walker, "H_∞ control for nonlinear systems via output feedback," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 546–559, 1993.

[2] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis, "State space solutions to standard H_2 and H_∞ control problems," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 831–847, 1989.

[3] A. Isidori and A. Astolfi, "Disturbance attenuation and H_∞ control via measurement feedback in nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 1283–1293, 1992.

[4] A. Isidori, "H_∞ control via measurement feedback for affine nonlinear systems," *Int. J. Robust and Nonlinear Contr.*, vol. 4, pp. 553–574, 1994.

[5] ——, "A necessary condition for nonlinear H_∞ control via measurement feedback," *Syst. Contr. Lett.*, vol. 23, pp. 169–177, 1994.

[6] W. M. Lu and J. C. Doyle, "H_∞ control of nonlinear systems via output feedback: Controller parameterization," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 2517–2521, 1994.

[7] R. A. Paz and J. V. Medanić, "Decentralized H_∞ norm bounding control for discrete-time systems," in *Proc. Amer. Contr. Conf.*, Chicago, IL, 1992, pp. 3274–3278.

[8] R. A. Paz, "Decentralized H_∞ control," in *Proc. Amer. Control Conf.*, San Francisco, CA, 1993, pp. 2381–2385.

[9] N. R. Sandell, P. Varaiya, M. Athans, and M. G. Safonov, "Survey of decentralized control methods for large scale systems," *IEEE Trans. Automat. Contr.*, vol. 23, pp. 108–128, 1978.

[10] M. H. Shor, W. R. Perkins, and J. V. Medanić, "Design of reliable decentralized controllers: A unified continuous/discrete formulation," *Int. J. Contr.*, vol. 56, pp. 943–965, 1992.

[11] D. D. Siljak, *Decentralized Control of Complex Systems*. San Diego, CA: Academic, 1991.

[12] A. J. Van der Schaft, "L_2-gain analysis on nonlinear systems and nonlinear state feedback H_∞ control," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 770–784, 1992.

[13] R. J. Veillette, J. V. Medanić, and W. R. Perkins, "Design of reliable control systems," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 290–304, 1992.

A Methodology for the Design of Optimal Traffic Shapers in Communication Networks

Venkat Anantharam and Takis Konstantopoulos

***Abstract*—The authors consider the problem of optimally regulating the source traffic in a communication network to simultaneously satisfy a finite number of affine burstiness constraints. They prove that an optimal solution is a series connection of correspondingly dimensioned "leaky buckets." They propose a simple "fork-join" implementation of the optimal solution and study extensions to the problem of optimally shaping the traffic flow to meet a burstiness constraint specified by a concave increasing function. A consequence of their optimality results is that permutations of leaky buckets in a series connection are input–output equivalent.**

***Index Terms*—Communication networks, flow control, Skorokhod reflection mapping.**

I. Preliminaries

In this paper we consider the problem of designing flow control schemes in a communication network. Flow control is necessary for the regulation and shaping of a source traffic stream, which must interact and share network resources with other traffic streams after it is admitted. Therefore, one normally requires the admitted flow to satisfy certain "burstiness" or "shaping" constraints. It is also desirable that the controller be optimal, in that the offered traffic is transmitted as quickly as possible.

A general model for a traffic process is a nonnegative sigma-finite Borel measure A on the time axis $\mathbb{R}_+$. This is represented by an increasing right-continuous process $\{A_t, t \geq 0\}$; the interpretation is that for $0 \leq s \leq t$, $A_t - A_s$ gives the volume of traffic (in cells) on the time interval $(s, t]$. Write $\mathcal{M}$ for the collection of such processes. We write A^S for the restriction of A on $S \subseteq \mathbb{R}_+$, defined by $A_t^S := \int_{S\cap[0,t]} dA_s$. We also define a partial ordering on $\mathcal{M}$ by $A \leq B \iff A_t \leq B_t$, for all $t \geq 0$. We say that $A \in \mathcal{M}$ is $(\sigma_{i0}, \sigma_i, \rho_i)_{i=1,\cdots,n}$ *constrained* iff, for all $0 \leq s \leq t$

$$A_t \leq \min_{1\leq i\leq n}\{\sigma_{i0} + \rho_i t\},\ A_t - A_s \leq \min_{1\leq i\leq n}\{\sigma_i + \rho_i(t-s)\}. \quad (1)$$

Here, $\sigma_i \geq \sigma_{i0} \geq 0$, $\rho_i \geq 0$, for all i. For $n = 1$ we simply say that A is (σ_0, σ, ρ) constrained. The above definitions are discussed in Anantharam [1] and Cruz [4], [5], and they also closely match the standard shaping descriptors that have been adopted in practice for high-speed networks. More generally, for f_0, f arbitrary concave increasing functions from $\mathbb{R}_+$ into $\mathbb{R}_+$, we say that A is (f_0, f) *constrained* iff

$$A_t \leq f_0(t), \quad A_t - A_s \leq f(t-s). \quad (2)$$

Of course, (1) is a convenient special case of (2). A *traffic regulator*, or *flow controller* is simply a map $\varphi: \mathcal{M} \to \mathcal{M}$. Some properties that such a map may possess are as follows.

Manuscript received January 6, 1998. Recommended by Associate Editor, W.-B. Gong. This work was supported in part by NSF under Grant NCR 88-57731, NSF Research Initiation Award NCR-9211343, NSF Faculty Career Development Award NCR-9502582, and the Texas Higher Education Coordinating Board under Grant ARP-224.

V. Anantharam is with the Department of EECS, University of California, Berkeley, CA 94720 USA (e-mail: ananth@eecs.berkeley.edu).

T. Konstantopoulos is with the ECE Department, University of Texas, Austin, TX 78712 USA (e-mail: takis@alea.ece.utexas.edu).

Publisher Item Identifier S 0018-9286(99)01280-5.

- F1) φ is *causal* iff for any $A_1, A_2 \in \mathcal{M}$, $t \geq 0$, $\varphi(A_1^{[0,t)} + A_2^{[t,\infty)})^{[0,t)} = \varphi(A_1)^{[0,t)}$. Intuitively, the decisions of a causal φ are based only on past arrival information.
- F2) φ is *realizable* iff $\varphi(A) \leq A$ for all $A \in \mathcal{M}$. Intuitively, a realizable φ cannot borrow flow from the future.[1]
- F3) φ is $(\sigma_{i0}, \sigma_i, \rho_i)_{i=1,\cdots,n}$ *constrained* iff, for all $A \in \mathcal{M}$, the output process $\varphi(A)$ is $(\sigma_{i0}, \sigma_i, \rho_i)_{i=1,\cdots,n}$ constrained.
- F4) φ is (f_0, f) *constrained* iff, for all $A \in \mathcal{M}$, the output process $\varphi(A)$ is (f_0, f) constrained.

We say that φ is *optimal* iff it satisfies (F1), (F2), (F3) [respectively, (F4)], and $\varphi(A) \geq \psi(A)$, for all $A \in \mathcal{M}$ and all flow controllers ψ satisfying (F1), (F2), (F3) [respectively, (F4)]. The construction (with proofs) of an optimal flow controller, is given in Section III (respectively, Section V). Section IV contains a simple "fork-join " or "parallel" implementation of the optimal flow controller. A corollary of our optimality theorem is that the effect of a tandem of leaky buckets on an arbitrary offered traffic stream is the same, whatever the order in which the leaky buckets are met. See also Cruz [5] for related results. Our methodology allows for arbitrary arrival processes which, in special cases, could represent continuous, piecewise constant, or slotted time arrivals. We avoid induction-type proofs by using the unifying concept of reflection mapping, explored in our earlier papers [2], [7], [8]: it is summarized in Section II and explained for the special case of a single constraint; the single-constraint optimal flow controller coincides with the popular "leaky bucket" scheme. Finally, the form of the optimal flow controller satisfying (F4) is surprisingly simple: in Section V it is seen that the total amount of backlogged traffic for the flow controller constructed in Sections III and IV evolves in time in a fashion similar to a continuous version of Lindley's equation of queueing theory but with a "concave server."

II. The Single Constraint Case

Recall [8] that a function $x \in D[0, \infty)$ (viz., right continuous with left limits), can be reflected at zero in the sense that there is a unique increasing $\ell \in D[0, \infty)$, with $\ell_{0-} = 0$, $q_t := x_t + \ell_t \geq 0$, for all $t \geq 0$, and $\int_{[0,\infty)} 1(q_t > 0) d\ell_t = 0$. Write $q = \mathcal{R}(x)$ for the *reflected process* and $\ell = \mathcal{L}(x)$ for the *reflector*; see [3], [6], [8], and [10]. The pair $(\mathcal{R}, \mathcal{L})$ is the *reflection mapping*. It is seen that

$$\begin{aligned} \ell_t &= \mathcal{L}(x)_t = -\left(\inf_{0\leq s\leq t} x_s \wedge 0\right) \\ q_t &= \mathcal{R}(x)_t = x_t + \ell_t = \sup_{0\leq s\leq t}(x_t - x_s) \vee x_t. \end{aligned} \tag{3}$$

The operator $\mathcal{R}$ is *causal* because, for $0 < t_1 < t_2$, $q_{t_2-} = (q_{t_1-} + x_{t_2-} - x_{t_1-}) \vee \sup_{t_1 \leq s < t_2}(x_{t_2-} - x_s)$, while $\mathcal{L}$ is *minimal*, i.e., if $x \in D[0,\infty)$, $\ell = \mathcal{L}(x)$, and $\tilde{\ell} \in D[0, \infty)$ is any increasing nonnegative process such that $x_t + \tilde{\ell}_t \geq 0$, for all t, then $\ell_t \leq \tilde{\ell}_t$, for all t.

The key observation is the fact that {a process $B \in \mathcal{M}$ is (σ_0, σ, ρ) constrained} $\Longleftrightarrow$ {the reflection of $x_t := \sigma - \sigma_0 + B_t - \rho t$ at zero stays below σ for all t}.

A (σ_0, σ, ρ) *leaky bucket* flow controller φ: $\mathcal{M} \to \mathcal{M}$ is defined below by giving its action $B = \varphi(A)$ on an arbitrary $A \in \mathcal{M}$

$$\begin{aligned} &x_t := \sigma - \sigma_0 + A_t - \rho t, \quad q := \mathcal{R}(x), \quad \ell := \mathcal{L}(x) \\ &B_t := A_t - (q_t - \sigma)^+ = \sigma_0 + \rho t - \ell_t - (\sigma - q_t)^+. \end{aligned} \tag{4}$$

[1] From a physical point of view, φ uses an infinite buffer where past arrivals are stored; the decision that a cell is to be transmitted or rejected can be taken at any point after arrival. If we are only allowed to use a buffer of finite size K, then we have to add the requirement that $\varphi(A)_t - \varphi(A)_s \leq K + A_t - A_s$, for all $t \geq s \geq 0$ and all $A \in \mathcal{M}$; while this restriction is not considered here, see [8] for results on optimal flow control with finite buffer constraints.

We call $c_t := (q_t - \sigma)^+$, respectively, $r_t := (\sigma - q_t)^+$, the amount of *cells*, respectively, *tokens*, stored in the cell, respectively, token, buffer at time t. The parameters σ and ρ, are referred to as *token buffer size* and *token arrival rate*, respectively. The interpretation is that traffic can be transmitted by consuming an equal amount of tokens; if no tokens are available then arriving cells are stored in the cell buffer. If no cells are present when tokens arrive, they are stored in the token buffer. The cell buffer is infinite, the token buffer has size σ, the token arrival rate is ρ, and the initial amount of tokens is σ_0. In [8] we proved that a (σ_0, σ, ρ) leaky bucket is causal, realizable, (σ_0, σ, ρ) constrained [in the sense of (F1)–(F3)], and optimal.

III. Main Optimality Results: Multiple Constraints

We now pass on to the solution of the optimal multiply constrained flow control problem. We first fix a (σ_0, σ, ρ) leaky bucket and discuss some of its properties in the following lemmas.

Lemma 1—Monotonicity: Let φ be a (σ_0, σ, ρ) leaky bucket. If $A \geq \tilde{A}$, then $\varphi(A) \geq \varphi(\tilde{A})$.

Proof: Apply relations (4) defining the leaky bucket φ to A and $\tilde{A}$. Let $B = \varphi(A)$, $\tilde{B} = \varphi(\tilde{A})$, and let all quantities in (4) corresponding to $\tilde{A}$ be denoted by tildes. Observe that $x_t + \tilde{\ell}_t = \tilde{q}_t + A_t - \tilde{A}_t \geq 0$. By the minimality of $\mathcal{L}$ we obtain $\tilde{\ell}_t \geq \ell_t$ for all t. Suppose now that $q_t < \sigma$. Then, using (4), $B_t = A_t \geq \tilde{A}_t \geq \tilde{B}_t$. Suppose next that $q_t \geq \sigma$. Then, again from (4), $B_t = \sigma_0 + \rho t - \ell_t \geq \sigma_0 + \rho t - \tilde{\ell}_t \geq \sigma_0 + \rho t - \tilde{\ell}_t - (\sigma - \tilde{q}_t)^+ = \tilde{B}_t$, completing the proof. □

Lemma 2—Invariance: Let φ be a (σ_0, σ, ρ) leaky bucket and let $A \in \mathcal{M}$ be (σ_0, σ, ρ) constrained. Then $\varphi(A) = A$.

Proof: If A is (σ_0, σ, ρ) constrained, then from the key observation mentioned earlier, $q_t = \sigma - \sigma_0 + A_t - \rho t \leq \sigma$ for all t. But then (4) gives $\varphi(A)_t = A_t - (q_t - \sigma)^+ = A_t$. □

Lemma 3—Preservation of Burstiness Constraints: Let φ be a (σ_0, σ, ρ) leaky bucket and let $A \in \mathcal{M}$ be $(\tilde{\sigma}_0, \tilde{\sigma}, \tilde{\rho})$ constrained. Then $\varphi(A)$ is also $(\tilde{\sigma}_0, \tilde{\sigma}, \tilde{\rho})$ constrained, for all values of σ_0, $\tilde{\sigma}_0$, σ, $\tilde{\sigma}$, ρ, $\tilde{\rho}$.

Proof: Let $B = \varphi(A)$. Since $B_t \leq A_t$ (realizability), we immediately get $B_t \leq \tilde{\sigma}_0 + \tilde{\rho} t$ for all $t \geq 0$. It remains to show that $B_t - B_s \leq \tilde{\sigma} + \tilde{\rho}(t - s)$. We make use of the inequalities

$$B_t - B_s \leq A_t - A_s + (q_s - \sigma)^+ \tag{5}$$
$$B_t - B_s \leq \rho(t - s) + (\sigma - q_s)^+ \tag{6}$$

following directly from (4). We distinguish two cases. First, assume that $\rho \leq \tilde{\rho}$. Using the fact that A is $(\tilde{\sigma}_0, \tilde{\sigma}, \tilde{\rho})$ constrained, (5) and (6) become

$$\begin{aligned} B_t - B_s &\leq \tilde{\sigma} + \tilde{\rho}(t - s) + (q_s - \sigma)^+ \\ B_t - B_s &\leq \tilde{\sigma} + \tilde{\rho}(t - s) + (\sigma - q_s)^+ \end{aligned}$$

and, since $(q_s - \sigma)^+ \wedge (\sigma - q_s)^+$ is identically zero, we obtain $B_t - B_s \leq \tilde{\sigma} + \tilde{\rho}(t - s)$. Secondly, assume $\rho > \tilde{\rho}$. Recall (3) for the reflection mapping and use inequality (5)

$$\begin{aligned} B_t - B_s &\leq A_t - A_s + (q_s - \sigma)^+ \\ &\leq A_t - A_s + \left\{\sup_{0\leq u\leq s}[(A_s - \rho s) - (A_u - \rho u)] \right. \\ &\qquad \left. \vee [\sigma - \sigma_0 + A_s - \rho s] - \sigma \right\} \vee 0 \\ &\leq \sup_{0\leq u\leq s}[(A_t - A_u) - \rho(s - u)] \\ &\qquad \vee [A_t - \rho s] \vee [A_t - A_s]. \end{aligned}$$

At this point use the assumption that A is $(\tilde{\sigma}_0, \tilde{\sigma}, \tilde{\rho})$ constrained, as well as $\rho > \tilde{\rho}$, to continue the inequality in the obvious way

$$B_t - B_s \le \sup_{0 \le u \le s} [\tilde{\sigma} + \tilde{\rho}(t-u) - \rho(s-u)] \vee [\tilde{\sigma_0} + \tilde{\rho}t - \rho s] \le [\tilde{\sigma} + \tilde{\rho}(t-s)] \vee [\tilde{\sigma_0} + \tilde{\rho}(t-s)] = \tilde{\sigma} + \tilde{\rho}(t-s).$$

Theorem 1: Let $\varphi_i : \mathcal{M} \to \mathcal{M}$ be a $(\sigma_{i0}, \sigma_i, \rho_i)$ leaky bucket, for $i = 1, \cdots, n$. Let $\varphi := \varphi_n \circ \cdots \circ \varphi_1$ be the composition of $\varphi_1, \cdots, \varphi_n$. Then the map φ is causal, realizable, $(\sigma_{i0}, \sigma_i, \rho_i)_{i=1,\cdots,n}$ constrained, and, for any other realizable (but not necessarily causal) $(\sigma_{i0}, \sigma_i, \rho_i)_{i=1,\cdots,n}$ constrained map $\psi : \mathcal{M} \to \mathcal{M}$, we have $\varphi(A) \ge \psi(A)$, for all $A \in \mathcal{M}$.

Proof: Causality follows from the fact that each φ_i is causal; hence, so is the composition. Realizability is also immediate from the realizability of each φ_i. Lemma 3, applied to each φ_i, shows that the tandem configuration is $(\sigma_{i0}, \sigma_i, \rho_i)_{i=1,\cdots,n}$ constrained. Finally, let $\psi : \mathcal{M} \to \mathcal{M}$ be as in the statement of the theorem. By the optimality result for a single leaky bucket, we have $\varphi_1(A) \ge \psi(A)$. By Lemma 1, $\varphi_2(\varphi_1(A)) \ge \varphi_2(\psi(A))$. But $\varphi_2(\psi(A)) = \psi(A)$, by Lemma 2. Continuing this process n times gives the desired. □

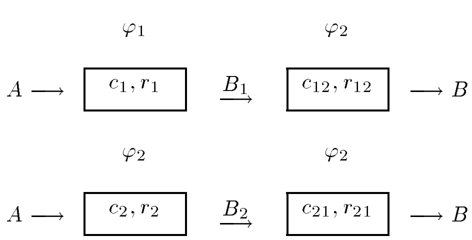
Corollary 1: Given n leaky buckets dimensioned by $(\sigma_{i0}, \sigma_i, \rho_i)$, $1 \le i \le n$, consider sending an arbitrary traffic process through these regulators in series. The overall output process is the same whatever the order in which these regulators are placed.

Proof: This is an immediate consequence of the theorem. Indeed, whatever order in which the leaky buckets are met the overall regulator is feasible and optimal. Thus the overall output process must be the same irrespective of the order of the regulators. □

IV. Parallel Implementation

An equivalent and more succinct description of the optimal flow controller is one that places the n leaky buckets in parallel and directs their outputs into a join (a simple synchronization mechanism). We show this below.

Let φ_i be a $(\sigma_{i0}, \sigma_i, \rho_i)$ leaky bucket, $i = 1, 2$. Consider the connections $\varphi_2 \circ \varphi_1$ and $\varphi_1 \circ \varphi_2$. For $A \in \mathcal{M}$, let $B_1 := \varphi_1(A)$, $B_2 := \varphi_2(A)$, $B := \varphi_2 \circ \varphi_1(A) = \varphi_1 \circ \varphi_2(A)$, the latter equality being a consequence of Theorem 1. Let c_i, r_i denote the amount of cells and tokens, respectively, in leaky bucket φ_i with input process A, $i = 1, 2$. Let c_{12}, r_{12} be the amount of cells and tokens, respectively, in leaky bucket φ_2 with input process $B_1 = \varphi_1(A)$. Similarly, c_{21}, r_{21} refer to leaky bucket φ_1 fed by $B_2 = \varphi_2(A)$. The following diagram should help as a reminder:



Lemma 4: For all $t \ge 0$, $c_{12,t}$ and $c_{21,t}$ cannot be both positive.

Proof: Assume first $\rho_1 \ne \rho_2$. If $c_{12,s} > 0$, $c_{21,s} > 0$, for some $s > 0$, then, by right continuity, there is $t > s$ such that $c_{12,u} > 0$, $c_{21,u} > 0$, for all $s \le u \le t$. Since cells and tokens cannot be simultaneously positive, we have, by balancing tokens in the downstream leaky buckets of each configuration

$$B_t - B_s = \rho_1(t-s), \quad B_t - B_s = \rho_2(t-s)$$

implying $\rho_1 = \rho_2$, contradicting our assumption. Next assume $\rho_1 = \rho_2 = \rho$ and, without loss of generality, $\sigma_1 \le \sigma_2$. If $\sigma_{10} \le \sigma_{20}$, then B_1, being $(\sigma_{10}, \sigma_1, \rho)$ constrained is also $(\sigma_{20}, \sigma_2, \rho)$ constrained and thus remains invariant after passing through the downstream leaky bucket. Hence there are never cells stored in that leaky bucket, i.e., $c_{12,t} = 0$ for all t. The only remaining case is $\sigma_{20} < \sigma_{10} \le \sigma_1 \le \sigma_2$, $\rho_1 = \rho_2 = \rho$. Initially the amount of tokens in the upstream leaky bucket of the second configuration is $\sigma_{20} < \sigma_1$. Let

$$\tau = \inf\{t > 0 : r_{2,t} = \sigma_1\}.$$

A token process does not have positive jumps, hence $\tau > 0$. Since $r_{2,t} < \sigma_1$ for all $0 \le t < \tau$, and since $\sigma_{20} < \sigma_{10}$, the process B_2 is $(\sigma_{10}, \sigma_1, \rho)$ constrained on the interval $[0, \tau)$, implying that it passes invariant through the downstream leaky bucket, on the same interval. Thus $c_{21,t} = 0$ for all $0 \le t < \tau$. If $\tau = +\infty$ there is nothing else to add. Otherwise, if $\tau < +\infty$, we claim that the trajectories of the tokens in the φ_2 leaky bucket of each configuration are identical on $[0, \tau)$. Indeed, they are initially the same and equal to σ_{20}; the amount of tokens arriving on $[0, \tau)$ is $\rho\tau$ in both cases; the token departures on $[0, \tau)$ are also identical since $B_t = B_{2,t}$ for all $t \in [0, \tau)$.[2] We now argue on the interval $[\tau, \infty)$, using τ as the new origin of time. Since the amount of tokens in the downstream leaky bucket of the first configuration at time τ equals that of the upstream leaky bucket of the second configuration, which is equal to σ_1, we obtain that the downstream leaky bucket of the first configuration is $(\sigma_1, \sigma_2, \rho)$ constrained on $[\tau, \infty)$. But B_1 is $(\sigma_{10}, \sigma_1, \rho)$ constrained. Since $\sigma_{10} \le \sigma_1$ and $\sigma_1 \le \sigma_2$, the flow B_1 passes invariant through this leaky bucket on $[\tau, \infty)$. Hence $c_{12,t} = 0$ for all $t \ge \tau$. □

For two flow controllers φ, ψ, the symbol $\varphi \wedge \psi$ stands for the flow controller defined by $(\varphi \wedge \psi)(A) := \min\{\varphi(A), \psi(A)\}$.

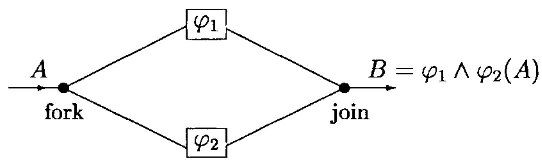
Theorem 2: Let φ_i be a $(\sigma_{i0}, \sigma_i, \rho_i)$ leaky bucket, $i = 1, \cdots, n$. Then, for each permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$, $\varphi_{i_1} \circ \cdots \circ \varphi_{i_n} = \varphi_1 \wedge \cdots \wedge \varphi_n$.

We prove the statement for $n = 2$ first. By Lemma 4, the amount of cells $\varphi_1(A)_t - \varphi_2(\varphi_1(A))_t$ and $\varphi_2(A)_t - \varphi_1(\varphi_2(A))_t$ cannot be simultaneously positive. Thus their minimum is equal to zero. But $\varphi_2(\varphi_1(A))_t = \varphi_1(\varphi_2(A))_t$, and so $\varphi_2(\varphi_1(A))_t = \varphi_1(A)_t \wedge \varphi_2(A)_t$. The general case follows by induction. □

Interpretation: The operation of taking the minimum of two traffic processes can be interpreted as a "join," an operation frequently encountered in manufacturing networks. Let the tokens in leaky bucket i be of "color" i. When a cell arrives into the flow controller it is split into n copies which are simultaneously sent to each leaky bucket. Thus the ith copy picks up a token of color i if it is available and departs instantaneously or waits in the cell buffer. The ith copy eventually departs from the ith leaky bucket by carrying a token of type i. It is then directed toward another buffer where it waits until the first moment of time that the other copies, carrying tokens of different colors, depart from their respective leaky buckets. At that moment, all copies depart simultaneously, and physically the flow controller triggers the transmission of a cell into the network. The final stage of the system is a "join," performing the synchronization operation just described. The join provides a buffer for storing copies of cells carrying colored tokens. Note that at all points of time at least

[2] Another way of expressing this is by recalling that $\{\sigma_2 - r_{2,t}, 0 \le t < \tau\}$ is the reflection of $\{\sigma_2 - \sigma_{20} + B_{2,t} - \rho t, 0 \le t < \tau\}$, while $\{\sigma_2 - r_{12,t}, 0 \le t < \tau\}$ is the reflection of $\{\sigma_2 - \sigma_{20} + B_t - \rho t, 0 \le t < \tau\}$. Since $B_{2,t}$ is identical to B_t for $0 \le t < \tau$, both $\{\sigma_2 - r_{2,t}, 0 \le t < \tau\}$ and $\{\sigma_2 - r_{12,t}, 0 \le t < \tau\}$ are reflections of the same process; hence, they are identical.

one of the colors is missing from the buffer.



The algorithm just described can be implemented simply.

V. Optimal Traffic Shaping via Concave Constraints

We deal with the design of the optimal flow controller satisfying (F1), (F2), and (F4). The key observation is that (F4) is equivalent to

$$\varphi(A)_t \le f_0^*(\rho) + \rho t, \qquad \varphi(A)_t - \varphi(A)_s \le f^*(\rho) + \rho(t-s)$$

for all $t \ge s \ge 0$, $\rho \ge 0$, $A \in \mathcal{M}$. Here, f^* (respectively, f_0^*) denotes the *dual* of f (respectively, f_0) defined by (c.f., Rockafellar [9]) $f^*(\rho) := \sup_{t\ge 0}[f(t) - \rho t]$, $\rho \ge 0$. The function $f^*\colon \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$ is convex, being the supremum of affine functions. Furthermore, $f(t) = \inf_{\rho \ge 0}[f^*(\rho) + \rho t]$, $t \ge 0$. The guess that an optimal flow controller is formed by "parallelizing" a collection of leaky buckets, one for each ρ, turns out to be correct.

Theorem 3: Define the map $\varphi\colon \mathcal{M} \to \mathcal{M}$ by the following equations:

$$x_t^\rho := f^*(\rho) - f_0^*(\rho) + A_t - \rho t, \quad \ell^\rho := \mathcal{L}(x^\rho) \tag{7}$$

$$\varphi(A)_t := A_t \wedge \inf_{\rho \ge 0}[f_0^*(\rho) + \rho t - \ell_t^\rho]. \tag{8}$$

Then φ is causal, realizable, (f_0, f) constrained, and optimal.

Proof: Notice that the map $\varphi_\rho(A)_t := A_t \wedge [f_0^*(\rho) + \rho t - \ell_t^\rho]$ is a $(f_0^*(\rho), f^*(\rho), \rho)$ leaky bucket. Consider a (f_0, f) constrained map ψ such that $\psi(A) \le A$ for all A. By the discussion preceding the theorem, such a map also satisfies a single $(f_0^*(\rho), f^*(\rho), \rho)$ constraint and hence, by the optimality of the leaky bucket, $\varphi_\rho(A) \ge \psi(A)$, for all A. Since $\varphi(A) = \inf_{\rho\ge 0} \varphi_\rho(A)$, we also have $\varphi(A) \ge \psi(A)$, for all A. Causality follows from the fact that each φ_ρ is causal, while realizability is due to the obvious inequality $\varphi(A) \le A$. Finally, we show that φ is (f_0, f) constrained. Notice that for any *finite set* $J \subseteq \mathbb{R}_+$ the map $\varphi_J := \inf_{\rho \in J} \varphi_\rho$ is $(f_0^*(\rho), f^*(\rho), \rho)_{\rho \in J}$ constrained, by Theorems 2 and 1. Fix now $t \ge s \ge 0$, $\varepsilon > 0$, and choose $J = \{\rho_1, \rho_2\}$ to be such that $\varphi_{\rho_1}(A)_s \le \varphi(A)_s + \varepsilon$, and $f^*(\rho_2) + \rho_2(t-s) \le f(t-s) + \varepsilon$. Thus, $\varphi_J(A)_s \le \varphi_{\rho_1}(A)_s \le \varphi(A)_s + \varepsilon$, while $\varphi(A)_t \le \varphi_J(A)_t$. Subtracting, we obtain $\varphi(A)_t - \varphi(A)_s \le \varphi_J(A)_t - \varphi_J(A)_s + \varepsilon \le \min_{\rho \in J}[f^*(\rho) + \rho(t-s)] + \varepsilon \le f(t-s) + 2\varepsilon$. Since ε is arbitrary, we conclude that $\varphi(A)_t - \varphi(A)_s \le f(t-s)$. Similarly, $\varphi(A)_t \le f_0(t)$. Thus φ satisfies (f_0, f) constraints and the theorem is proved. □

It is interesting to find an expression for the amount of cells stored in the optimal flow controller. By (7) and (3), $\ell_t^\rho = \inf_{0 \le s \le t}[f^*(\rho) - f_0^*(\rho) + A_s - \rho s] \wedge 0$. Inserting this in (8) we obtain

$$\varphi(A)_t = A_t \wedge \inf_{\rho \ge 0} \inf_{0 \le s \le t}[f^*(\rho) + A_s + \rho(t-s)] \wedge [f_0^*(\rho) + \rho t].$$

Interchanging the order of the infima, and using the representations of f and f_0 in terms of their duals, we obtain

$$\varphi(A)_t = A_t \wedge \inf_{0 \le s \le t}[A_s + f(t-s)] \wedge f_0(t). \tag{9}$$

Finally, the formula for c_t, the amount of cells stored in the cell buffer at time t, is obtained by a simple subtraction: $c_t = A_t - \varphi(A)_t$, and so

$$c_t = \sup_{0 \le s \le t}[(A_t - A_s) - f(t-s)] \vee [A_t - f_0(t)] \vee 0. \tag{10}$$

The map $A \mapsto c$ can be thought of as having been obtained through some kind of "reflection mapping." The ordinary reflection mapping reveals itself only upon choosing $f(t) = f_0(t) = \rho t$. In other words, (10) is a generalization of "Lindley's equation" of classical queueing theory.

It is easy to see that a tandem of optimal $(f_{0i}, f_i)_{i=1,\cdots,n}$ constrained flow controllers is input/output equivalent to a single $(\min_i f_{0i}, \min_i f_i)$ constrained flow controller. The map φ of (9), although causal, is realizable by a finite-dimensional algorithm if and only if both f and f_0 are piecewise linear functions with finitely many pieces.

One can start from (9) and show its optimality by using various inequalities, but the approach presented here seems more systematic and insightful. It should also be pointed out that the existence of an explicit formula such as (9) is quite unexpected. For instance, if we add the requirement that the flow controller cannot store more than K cells (see footnote 1), then, just as in the single-constraint case [8], we do not expect closed-form expressions for an optimal flow controller. However, we do expect that an optimal flow controller is expressible in terms of reflection mappings.

An (f_0, f) constrained flow controller can be used to implement fairly sophisticated traffic shaping. As mentioned earlier, the need for traffic shaping is imposed by the fact that the network is a shared resource. The problem of how to select the shaping functions will be influenced by what are considered appropriate quality of service (QoS) metrics, and by the kinds of guaranteed QoS services the network will choose to offer, which in turn will be influenced by economic factors. This is a problem for future research.

Finally, we should mention that it is desirable to devise methods for joint flow control of interacting inputs, as for instance when one is simultaneously controlling different substreams generated by a multimedia application. This leads to multidimensional reflection mapping problems of a related nature to the one considered here and appears to be quite challenging.

References

[1] V. Anantharam, "An approach to the design of high speed networks for bursty traffic," in *Proc. Int. Teletraffic Congr. Sponsored Seminar on Teletraffic Methods for Current and Future Telecom Networks*, Indian Institute of Science, Bangalore, India, Nov. 15–19, 1993, pp. 163–168.

[2] V. Anantharam and T. Konstantopoulos, "Optimality and interchangeability of leaky buckets in tandem," in *Proc. 32nd Allerton Conf.*, Monticello, IL, Oct. 1994, pp. 235–244; also Univ. Texas at Austin, Tech. Rep. SCC-95-01.

[3] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration.* Boston, MA: Birkhäuser, 1983.

[4] R. L. Cruz, "A calculus for network delay—Part I: Network elements in isolation," *IEEE Trans. Info. Th.*, vol. 37, pp. 114–131, 1991.

[5] ——, "A calculus for network delay—Part II: Network analysis," *IEEE Trans. Info. Th.*, vol. 37, pp. 132–141, 1991.

[6] J. M. Harrison, *Brownian Motion and Stochastic Flow Systems.* New York: Wiley, 1985.

[7] T. Konstantopoulos and V. Anantharam, "An optimal flow control scheme that regulates the burstiness of traffic subject to delay constraints," in *Proc. 32nd IEEE Conf. Decision and Control*, 1993.

[8] ——, "Optimal flow control schemes that regulate the burstiness of traffic," *IEEE/ACM Trans. Networking*, vol. 3, pp. 423–432, 1995.

[9] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton Univ. Press, 1970.

[10] A. V. Skorokhod, "Stochastic equations for diffusion processes with boundaries," *Th. Prob. Appl.*, vol. 6, pp. 287–298, 1961.