A Simple, Scalable, and Stable Explicit Rate Allocation Algorithm for MAX–MIN Flow Control with Minimum Rate Guarantee

Song Chong, *Member, IEEE*, Sangho Lee, and Sungho Kang, *Member, IEEE*

***Abstract*—In this paper, we present a novel control-theoretic explicit rate (ER) allocation algorithm for the MAX–MIN flow control of elastic traffic services with minimum rate guarantee in the setting of the ATM available bit rate (ABR) service. The proposed ER algorithm is *simple* in that the number of operations required to compute it at a switch is minimized, *scalable* in that per-virtual-circuit (VC) operations including per-VC queueing, per-VC accounting, and per-VC state management are virtually removed, and *stable* in that by employing it, the user transmission rates and the network queues are asymptotically stabilized at a unique equilibrium point at which MAX–MIN fairness with minimum rate guarantee and target queue lengths are achieved, respectively. To improve the speed of convergence, we normalize the controller gains of the algorithm by the estimate of the number of locally bottlenecked VCs. The estimation scheme is also computationally simple and scalable since it does not require per-VC accounting either. We analyze the theoretical performance of the proposed algorithm and verify its agreement with the practical performance through simulations in the case of multiple bottleneck nodes. We believe that the proposed algorithm will serve as an encouraging solution to the MAX–MIN flow control of elastic traffic services, the deployment of which has been debated long due to their lack of theoretical foundation and implementation complexity.**

***Index Terms*—Asymptotic decay rate, elastic traffic services, max–min flow rate, scalibility, stability.**

I. INTRODUCTION

MANY data applications are highly bursty and have no way of predicting data traffic requirements in advance, but have well-defined packet loss requirements and can tolerate time-varying and unpredictable packet delays. More importantly, they are able to modify their data transfer rates according to network loading. Thus the notion of *elastic traffic* services was introduced, by which the data transfer rates (or flows) are adjusted at the source depending on the available bandwidth at the network. A representative example of the elastic traffic services is the available bit rate (ABR) service in ATM networks.

Manuscript received April 23, 2000; revised August 22, 2000 and January 18, 2001; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor J.-Y. LeBoudec. This work was supported in part by Korea Science and Engineering Foundation under Grant 98-0101-07-01-3 and by Samsung Electronics, Korea.

S. Chong is with the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Taejon 305-701, Korea (e-mail: song@ee.kaist.ac.kr).

S. Lee is with the Telecom Research and Development Center, Samsung Electronics, Sungnam 463-050, Korea (e-mail: sangho@metro.telecom.samsung.co.kr).

S. Kang is with the Department of Electrical Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: shkang@yonsei.ac.kr).

Publisher Item Identifier S 1063-6692(01)04732-X.

The ATM Forum has adopted the rate-based closed-loop control approach for the flow control of ABR service [1], [2]. The rate-based closed-loop control, as its name implies, uses feedback information from the network to control the rate at which each source can transmit cells into the network. The feedback information is carried by special control cells called resource management (RM) cells. There are three mechanisms for a switch to write its congestion status onto RM cells: explicit forward congestion indication (EFCI) marking, relative rate (RR) marking, and explicit rate (ER) marking, at least one of which has to be implemented on a switch for the rate-based flow control. Our primary concern in this paper is the ER allocation algorithm for the ER marking.

We construct a new ER allocation algorithm on a solid analytical basis with the following performance and complexity objectives and show that our algorithm indeed achieves these goals for a variety of network scenarios.

- Existence of an asymptotically stable (oscillation-free) equilibrium point at which high utilization, low cell loss, and MAX–MIN fair rate allocation are achieved.
- High responsiveness to the changes of available bandwidth due to the variation of high-priority variable bit rate (VBR) traffic and the activity of VBR and ABR virtual circuits (VCs).
- Low and scalable degree of implementation complexity for which not only the number of operations required to compute the algorithm is minimized but also per-VC operations including per-VC queueing, per-VC accounting, and per-VC state management are virtually removed.

The major difficulty in the ABR flow control design arises from the long and heterogeneous round-trip delays involved in the closed-loop control. In particular, if only a binary feedback mechanism (either EFCI or RR marking, or both) is employed, ABR queues in the network inevitably exhibit a persistent oscillation, namely a limit cycle, in the steady state due to the stale feedback information, with its magnitude and period being an increasing function of the delay-bandwidth product [3], [18], [19]. It is obvious that such an oscillatory behavior of ABR queues will increase the likelihood of cell loss and link underutilization, respectively, due to repeated buffer overflow and underflow as the speed and size of the network become higher and larger as it is today. The major premise of the ER marking is its potential ability to realize asymptotic stability of ABR queues and hence overcome this drawback of binary feedback mechanisms. However, designing an asymptotically

stable ER allocation algorithm, particularly in a simple form, is not an easy problem either.

Benmohamed and Meerkov in their pioneering work [4], [5] formulated the rate-based flow control problem as a discrete-time feedback control problem with delays. Based on this formulation, they derived a control-theoretic ER allocation algorithm which not only achieves asymptotic stability of the closed-loop system but also allows for arbitrary control of the closed-loop performance. Their ER allocation algorithm is as follows.

$$r[k+1] = r[k] - \sum_{i=0}^{I} \alpha_i (q[k-i] - q_T) - \sum_{j=0}^{\tau_{\max}} \beta_j r[k-j] \tag{1}$$

where $r[k]$, $q[k]$, and q_T are the ER computed by the switch at discrete time k, the per-class ABR queue length at time k and the target queue length, respectively. α_i and β_j are the controller gains and $\tau_{\max}$ and I are the largest round-trip delay of ABR VCs on this link and an arbitrary integer greater than 0, respectively. Its complete controllability of the closed-loop performance, however, comes at a high cost. That is, it requires long memory of the queue lengths and the ER values at present and in the past up to time lags I and $\tau_{\max}$, and requires a large number of floating point multiplications every discrete time slot. Therefore, its practical use is limited by the high degree of implementation complexity particularly as the round-trip delay increases [6].

We take a different approach. We aim to design an ER allocation algorithm which allows for low degree of implementation complexity but with an *acceptable* level of control rather than arbitrary control for the closed-loop performance. More specifically, we trade off the capability of arbitrary control of the closed-loop performance for low degree of implementation complexity by removing the long memory of past queue lengths and ER values. Our proposed discrete-time algorithm is as follows.

$$r[k+1] = r[k] - \frac{A}{|Q|}(q[k] - q[k-1]) - \frac{BT}{|Q|}(q[k] - q_T), \quad A, B > 0 \tag{2}$$

where A and B are the controller gains, T is the duration of update interval, Q denotes the set of locally bottlenecked VCs at the link, and $|Q|$ is the cardinality of Q. Note that the proposed algorithm is in fact a special case of Benmohamed and Meerkov's algorithm (1) with $I = 1, \alpha_1 = -A/|Q|, \alpha_0 + \alpha_1 = (BT)/|Q|$ and $\beta_j = 0,\ \forall j$.

By the term "an acceptable level of control," we mean that by properly choosing the controller gains for given round-trip delays, one can completely control the *asymptotic* behavior of the closed-loop system. An explicit condition to achieve this level of control is given in the paper. On the other hand, the available link bandwidth has to be allocated in the MAX–MIN fair sense to the individual sources. It is shown that this happens automatically in the steady state by virtue of (2). Another notable feature of the proposed algorithm is the normalization of the controller gains by the number of locally bottlenecked VC's $|Q|$. We show that this normalization is indeed beneficial in such a way that it makes the closed-loop performance to be virtually independent of the number of locally bottlenecked VCs on a link.

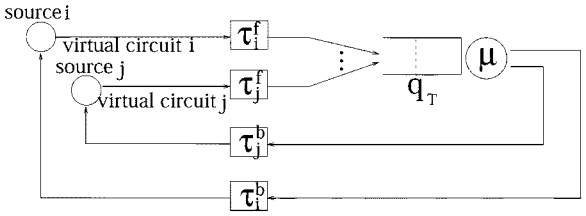

Fig. 1. Network model with a node of interest.

Estimation of the number of locally bottlenecked VCs is a challenging research subject [8], [13]–[15]. The difficulty lies in that the dynamics of any $|Q|$-estimation process interacts with those of the ER allocation process until the closed-loop system reaches the steady state so that it can make the closed-loop system unstable. In this paper, we also present a stable yet scalable $|Q|$-estimation algorithm in relation to the proposed ER allocation algorithm.

So far, we have discussed control-theoretic ER allocation algorithms where queue-length control is the primary concern and fair rate allocation is achieved automatically by virtue of the queue-length control. There is another class of ER allocation algorithms [10]–[17] where fair rate allocation with input-output rate matching is the primary concern and queue-length control, if any, is supplementary. The algorithms in [13]–[16] require that each switch collects the information on the available bandwidth of the outgoing link, the fair rate of remotely bottlenecked VCs which is determined at remote links and the number of locally bottlenecked VCs. A common drawback of these algorithms is that they suffer from not only high-degree of implementation complexity due to the per-VC operations involved but also loosely controlled queue length. The algorithms in [10]–[12], [17] do not need to track bottleneck location of VCs but require that each switch measures incoming traffic rate as well as available bandwidth of outgoing link. These algorithms also suffer from either uncontrolled or loosely controlled queue length. In this respect, the novelty of our proposed ER algorithm is an explicit control of both rate and queue dynamics.

II. Explicit Rate Flow Control—A Fluid Model

Consider a network model in Fig. 1 where we model a single node explicitly and the other nodes implicitly to simplify the analysis. The assumptions employed for the analysis of the network model are as follows and are fairly standard [7], [18].

- A.1. The traffic is viewed as a deterministic fluid flow and the network queueing process and the flow control mechanism are continuous in time. This assumption enables us to model the closed-loop system by a differential equation.
- A.2. The round-trip delay τ_i of a VC i is the sum of the forward-path delay τ_i^f and the backward-path delay τ_i^b, which includes propagation, queueing, and transmission and processing times. We assume that the round-trip delay is constant.
- A.3. The sources are *persistent* until the system reaches steady state. By persistent, we mean that the source always has enough data to transmit at the allocated rate.

A.4. There are no arrivals and departures of VCs until the system reaches steady state.

A.5. The available bandwidth μ at the link is constant until the system reaches steady state. Also, the buffer size at the link is assumed infinite.

Let $a_i(t)$ and $r_i(t)$ respectively denote the rate at which source i transmits data at the source time t and the explicit rate of VC i computed by the node of interest at the node time t. Also, let $b_i(t)$ and p_i respectively denote the latest minimum value of the explicit rates allocated to VC i by the nodes along the VC i's path except the one allocated by the node of interest and the peak rate constraint of VC i (i.e., PCR of VC i in the ATM ABR terminology).

The source behavior can be modeled by

$$a_i(t) = \min\left[r_i\left(t-\tau_i^b\right), b_i(t), p_i\right], \quad \forall i \in N \tag{3}$$

where N denotes the set of all the VCs whose route includes the node of interest. This model implies that a source transmits data at the smallest value among the ERs allocated by the nodes along the route and the PCR of the VC.

The dynamics of the per-class ABR queue of interest are given by

$$\dot{q}(t) = \begin{cases} \sum_{i\in N} a_i\left(t-\tau_i^f\right) - \mu, & q(t) > 0 \\ \left[\sum_{i\in N} a_i\left(t-\tau_i^f\right) - \mu\right]^+, & q(t) = 0 \end{cases} \tag{4}$$

where $[\cdot]^+ = \max[\cdot, 0]$.

The proposed ER allocation algorithm is a distributed algorithm which runs independently and identically at each switch based on the current network state including the queue length $q(t)$, the derivative of the queue length $\dot{q}(t)$, and the estimate of the number of locally bottlenecked VCs, $|\hat{Q}|$. The algorithm is given by the following equations in continuous time.

$$r_i(t) = r(t) + m_i, \quad \forall i \in N \tag{5}$$

and

$$\dot{r}(t) = \begin{cases} -\frac{A}{|\hat{Q}|}\dot{q}(t) - \frac{B}{|\hat{Q}|}(q(t) - q_T), & r(t) > 0 \\ \left[-\frac{A}{|\hat{Q}|}\dot{q}(t) - \frac{B}{|\hat{Q}|}(q(t) - q_T)\right]^+, & r(t) = 0 \end{cases} \tag{6}$$

where $A, B > 0$ and m_i denote the minimum data rate which the node is required to guarantee during the entire holding time of VC i (i.e., MCR in the ATM ABR terminology). We assume that $m_i \le p_i$, $\forall i \in N$ and there exists a call admission control which guarantees $\sum_{i\in N} m_i < \mu$. Note that $r(t)$ is the common part of per-VC ER allocations, $r_i(t)$, $\forall i$, and the only per-VC computation required is the addition of m_i to the common part of ER, $r(t)$. This is why this algorithm is scalable in terms of computational complexity with increasing number of VCs. Recall that in the current ABR protocol each RM cell of VC i carries the value m_i. Hence, a good implementation of the algorithm which does not require per-VC MCR table is that the node periodically updates $r(t)$ based on (6) in the background, irrespective of the arrivals and departures of RM cells, reads m_i from incoming VC i's RM cells, and writes the sum of the latest $r(t)$ and m_i on the outgoing RM cells. Refer to [9] for the implementation details.

A notable feature of the proposed algorithm is the normalization of the controller gains, A and B, by the estimate of the number of locally bottlenecked VCs, $|\hat{Q}|$. This normalization is optional, i.e., it is not absolutely necessary but it is recommended since, as will be discussed in Section V, it makes the closed-loop dynamics to be virtually independent of the number of locally bottlenecked VCs on the link.

The terms "remotely bottlenecked VC" and "locally bottlenecked VC" are defined in the steady state for a given network loading. Locally bottlenecked VCs at a link are defined to be those VCs whose fair share is determined at this link. In the same way, remotely bottlenecked VCs at a link are defined to be those VCs whose fair share is determined at other places because either their data transfer rate is limited by their PCR or they are bottlenecked at some other link in the path. Let $a_{is} = \lim_{t\to\infty} a_i(t)$, $r_{is} = \lim_{t\to\infty} r_i(t)$ and $b_{is} = \lim_{t\to\infty} b_i(t)$. Then, more formally, the set of all the locally bottlenecked VCs, Q, at the link of interest is given by

$$Q = \{i \,|\, i \in N \text{ and } a_{is} = r_{is}\} \tag{7}$$

and the set of all the remotely bottlenecked VCs, $N - Q$, at the link of interest is given by

$$N - Q = \{i \,|\, i \in N \text{ and } a_{is} = \min[b_{is}, p_i]\}. \tag{8}$$

III. Steady State and Fairness

In this section, we study the steady-state characteristics of the closed-loop dynamics when our ER allocation algorithm is applied. Suppose that the closed-loop dynamics have an equilibrium point at which the derivatives of the system variables are zero, i.e., $\lim_{t\to\infty} \dot{q}(t) = 0$ and $\lim_{t\to\infty} \dot{r}(t) = 0$. Let $r_s = \lim_{t\to\infty} r(t) > 0$. Then, from (3), (5), and (6), we have

$$a_{is} = \min[r_{is}, b_{is}, p_i], \quad r_{is} = r_s + m_i, \quad \forall i \in N \tag{9}$$

and $q_s = q_T$ where $q_s = \lim_{t\to\infty} q(t)$. Since $q_s = q_T > 0$, the buffer equation (4) implies that

$$\sum_{i\in N} a_{is} = \mu. \tag{10}$$

By combining (9), (10), and the definitions (7) and (8), we obtain

$$\sum_{i\in Q} r_s + \sum_{i\in Q} m_i + \sum_{i\in N-Q} \min[b_{is}, p_i] = \mu \tag{11}$$

which implies that

$$r_s = \frac{\mu - \sum_{i\in N-Q} \min[b_{is}, p_i] - \sum_{i\in Q} m_i}{|Q|}. \tag{12}$$

The following proposition states the result.

Proposition 3.1: For $\sum_{i\in N} m_i < \mu$ and $\min[b_{is}, p_i] \ge m_i$, there exists a unique steady state solution (equilibrium point) at which 1) the queue length is equal to the target queue length ($q_s = q_T$), 2) the available bandwidth at the link is fully utilized ($\sum_{i\in N} a_{is} = \mu$), and 3) individual MCRs are guaranteed at the link and the bandwidth subtracted by the sum of MCRs,

$\mu - \sum_{i\in N} m_i$, is allocated in the MAX–MIN fair sense to the individual sources. That is

$$a_{is} = \begin{cases} \frac{\mu - \sum_{i\in N-Q} \min[b_{is},p_i] - \sum_{i\in Q} m_i}{|Q|} + m_i, & i \in Q \\ \min[b_{is},p_i], & i \in N-Q. \end{cases} \tag{13}$$

This proposition implies that when our ER allocation algorithm is applied, the closed-loop system has a unique equilibrium point at which the MAX–MIN fairness with MCR guarantee is achieved and the queue length is equal to the target value q_T, no matter what the network loading is. This is exactly the same property that Benmohamed and Meerkov's algorithm (1) has [4], [5].

IV. Asymptotic Stability

In general, the stability of the equilibrium point given in Proposition 3.1 could be investigated in the case of multiple nodes. However, due to the complex nature of the coupled dynamics between nodes, the analysis could be so involved that in this paper we do not attempt to solve the global stability problem in the coupled multinode setting, and we only investigate it by simulations in Section VIII. On the other hand, in [5] Benmohamed and Meerkov showed that under a special service discipline there exists a neighborhood of the equilibrium point in which node-to-node dynamics are decoupled. Moreover, they showed through simulations that the local stability condition derived in the neighborhood works well for the first come first serve (FCFS) service discipline as well. By appealing to this result, we suppose that such a neighborhood, say R, exists in our case as well.

Consider a subset of the neighborhood R in which the following are satisfied: 1) $b_i(t) = b_{is}$, $\forall i \in N$, i.e., the dynamics of the other nodes are in steady state; 2) $\{i \,|\, i \in N \text{ and } a_i(t) = r_i(t - \tau_i^b)\} = Q$ and $\{i \,|\, i \in N \text{ and } a_i(t) = \min[b_{is}, p_i]\} = N - Q$, i.e., the locally bottlenecked VCs transmit data at $r_i(t - \tau_i^b)$ and the remotely bottlenecked VCs transmit data at $\min[b_{is}, p_i]$; 3) the saturation nonlinearities in (4) and (6) are not activated, i.e., both $q(t)$ and $r(t)$ are positive valued; and 4) the $|Q|$-estimation process is in steady state, i.e., $|\hat{Q}|$ is constant.

In this neighborhood of the equilibrium point, we can simplify the dynamic equations (3), (4), and (6) as follows.

$$a_i(t) = \begin{cases} r_i\left(t - \tau_i^b\right), & i \in Q \\ \min[b_{is}, p_i], & i \in N - Q \end{cases} \tag{14}$$

$$\dot{q}(t) = \sum_{i\in N} a_i\left(t - \tau_i^f\right) - \mu \tag{15}$$

and

$$\dot{r}(t) = -\frac{A}{|\hat{Q}|}\dot{q}(t) - \frac{B}{|\hat{Q}|}(q(t) - q_T). \tag{16}$$

By combining (14) and (15), we obtain

$$\dot{q}(t) = \sum_{i\in N-Q} \min[b_{is}, p_i] + \sum_{i\in Q} r_i(t - \tau_i) - \mu. \tag{17}$$

Define an error function by $e(t) = q(t) - q_T$. By combining (16), the differentiation of (17), and the differentiation of (5), we obtain the following closed-loop equation:

$$\ddot{e}(t) + \frac{A}{|\hat{Q}|}\sum_{i\in Q} \dot{e}(t - \tau_i) + \frac{B}{|\hat{Q}|}\sum_{i\in Q} e(t - \tau_i) = 0 \tag{18}$$

which is a second-order retarded differential equation. The characteristic equation of the closed-loop equation is given by

$$D(s) = s^2 + \frac{A}{|\hat{Q}|}\sum_{i\in Q} s e^{-s\tau_i} + \frac{B}{|\hat{Q}|}\sum_{i\in Q} e^{-s\tau_i} = 0 \tag{19}$$

which has infinite number of roots. For the asymptotic stability of the closed-loop equation (18), all the roots of the characteristic equation (19) must have negative real parts [20], [21].

To find the necessary and sufficient condition for exponential polynomials to have stable roots, one can appeal to Pontryagin's criterion [20], [22] assuming discrete delays of rational ratios. For general cases with continuous delays or discrete delays of irrational ratios, Stépán's criterion [21] provides a way to construct the necessary and sufficient condition. However, constructing such a condition in an explicit form is extremely complicated particularly for the case with a large number of heterogeneous round-trip delays.

Instead, we derive the necessary and sufficient condition for the asymptotic stability in the case that all the round-trip delays are identical. Let $\tau_i = \tau$, $\forall i$. Then, the closed-loop equation (18) becomes

$$\ddot{e}(t) + \frac{|Q|}{|\hat{Q}|}A\dot{e}(t - \tau) + \frac{|Q|}{|\hat{Q}|}Be(t - \tau) = 0. \tag{20}$$

This equation is normalized so that the time lag τ becomes unity. Let $t = \tau\xi$. In terms of the new variable ξ, (20) becomes

$$\ddot{e}(\xi) + U\dot{e}(\xi - 1) + Ve(\xi - 1) = 0 \tag{21}$$

where $U = (|Q|/|\hat{Q}|)A\tau$ and $V = (|Q|/|\hat{Q}|)B\tau^2$. The characteristic equation of (21) is

$$H(z) = z^2 e^z + Uz + V = 0. \tag{22}$$

To find the necessary and sufficient condition that all the roots of this exponential polynomial have negative real parts, we appeal to Pontryagin's criterion which yields the following result. The proof is given in the Appendix.

Proposition 4.1: Let

$$U = \frac{|Q|}{|\hat{Q}|}A\tau \quad \text{and} \quad V = \frac{|Q|}{|\hat{Q}|}B\tau^2. \tag{23}$$

The closed-loop equation (20) is asymptotically stable if and only if

$$0 < U < \frac{\pi}{2} \quad \text{and} \quad 0 < V < \omega_1^2\sqrt{1 - \left(\frac{U}{\omega_1}\right)^2} \tag{24}$$

where ω_1 is the unique solution of $U = \omega \sin\omega$ in the interval $(0, \pi/2)$.

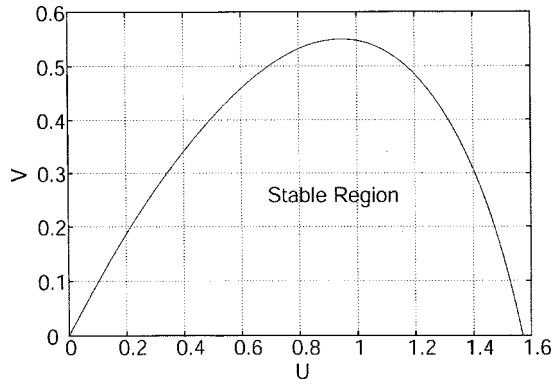

Fig. 2. Stable region with respect to U and V.

For better presentation, we plot this stability condition in Fig. 2.

The stability analysis in the case of heterogeneous round-trip delays would be much more involved due to the complex nature of the characteristic equation (19). Moreover, the stability condition, even if found in an explicit form, could be the particular solution of a given network scenario, which means that we have to repeatedly estimate individual round-trip delays and check the stability condition as the network scenario changes. Obviously, it is impractical to implement. Instead, we conjecture that the stability condition we derived for the case of homogeneous delays will work for the case of heterogeneous delays as well if we set $\tau = \tau_{\max}$ where $\tau_{\max} = \max\{\tau_i, i \in N\}$. We verify this conjecture by simulations in Section VIII. Estimation of $\tau_{\max}$ will be much easier than that of individual round-trip delays in reality.

If the controller gains were not normalized by the estimate of $|Q|$, the coefficients of the closed-loop equation in (21) would vary largely according to the changes of $|Q|$ since $U = |Q|A\tau$ and $V = |Q|B\tau^2$ in that case. The proposed normalization resolves this problem as long as $|\hat{Q}| \approx |Q|$ by making the coefficients of the closed-loop equation and thus the closed-loop dynamics to be virtually independent of $|Q|$. Therefore, with the normalization, one can choose stable (or optimal) A and B independently of $|Q|$ by letting $U = A\tau$ and $V = B\tau^2$. In contrast, without the normalization, the coefficients depend on $|Q|$, i.e., $U = |Q|A\tau$ and $V = |Q|B\tau^2$, so that the stable gains should be selected to cope with the worst case that $|Q| = N$. Note that N could be as large as tens (or hundreds) of thousands so that the gains selected from the worst case are typically very small. These small gains are problematic since they would significantly lower the speed of system convergence for the ordinary cases that $|Q| \ll N$.

V. Principal Root and Asymptotic Decay Rate

In this section, we determine the rate at which the stable closed-loop system approaches steady state. Any solution to the normalized closed-loop equation (21) can be represented by a series [20]

$$e(\xi) = \sum_{n=1}^{\infty} p_n(\xi) e^{z_n \xi} \tag{25}$$

where $p_n(\xi)$ is a suitable polynomial and z_n, $\forall n$, are the roots of the corresponding characteristic equation (22). Consider the principal root, denoted by z^*, which is the root having the largest real part. Let $z^* = -\alpha \pm j\beta$, $\alpha > 0$, $\beta \geq 0$. It follows from (25) that

$$e(\xi) \sim Ce^{z^*\xi} \tag{26}$$

where C is a constant depending on the initial conditions of (21), and by $x(\xi) \sim y(\xi)$ we mean that $x(\xi)/y(\xi)$ asymptotically approaches 1. Also, from (25)

$$\|e(\xi)\| \leq ce^{-\alpha\xi} \quad \text{for large } \xi \tag{27}$$

where $\|\cdot\|$ denotes the Euclidean norm and c is a constant depending on the initial conditions of (21). In terms of the original variable t $(=\tau\xi)$, (27) can be rewritten by

$$\|e(t)\| \leq ce^{-\frac{\alpha}{\tau}t} \quad \text{for large } t. \tag{28}$$

Note that α/τ is the asymptotic decay rate at which the original system tends to the equilibrium point. Hence the inverse of it, τ/α, is the time constant of the original closed-loop system, i.e., the time it takes for a small perturbation around the equilibrium point to decrease by a factor of e^{-1}. Similarly, α and α^{-1} are the asymptotic decay rate and the time constant of the normalized system.

The difficulty of computing the principal root is notorious for the general class of characteristic equations. In [23], the principal root for a first-order delay-differential equation is found. We take the similar approach here for our second-order delay-differential equation. We first determine α and then determine β. The change of variable $z = \psi - \sigma$, $\sigma > 0$, transforms the characteristic equation (22) to

$$(\psi^2 - 2\sigma\psi + \sigma^2)e^{\psi} + Ue^{\sigma}\psi + (V - U\sigma)e^{\sigma} = 0. \tag{29}$$

For a given (U, V) pair satisfying the stability condition (24), if we choose σ to be the supremum of positive real numbers for which the transformed characteristic equation (29) has all roots in the left half plane, then $\alpha = \sigma$. In principle, the necessary and sufficient conditions that all roots of the above characteristic equation be in the left half plane can be obtained from the Pontryagin's criterion and thus one can find the supremum value of σ subject to those conditions. This maximization problem, however, turns out to be too complicated to obtain the solution in an explicit form. Thus, we take a numerical approach as follows. Consider the following second-order delay-differential equation whose characteristic equation is identical to the transformed characteristic equation in (29)

$$\ddot{e}(\varsigma) - 2\sigma\dot{e}(\varsigma) + \sigma^2 e(\varsigma) + Ue^{\sigma}\dot{e}(\varsigma - 1) + (V - U\sigma)e^{\sigma}e(\varsigma - 1) = 0. \tag{30}$$

For a given (U, V) pair satisfying the stability condition (24), we repeatedly solve this differential equation by increasing σ from zero until the solution begins to diverge. The value of σ right before the divergence is then taken as the supremum of σ, i.e., α, for that given (U, V) pair. Fig. 3 shows the result obtained by this numerical approach. We found that the asymptotic decay

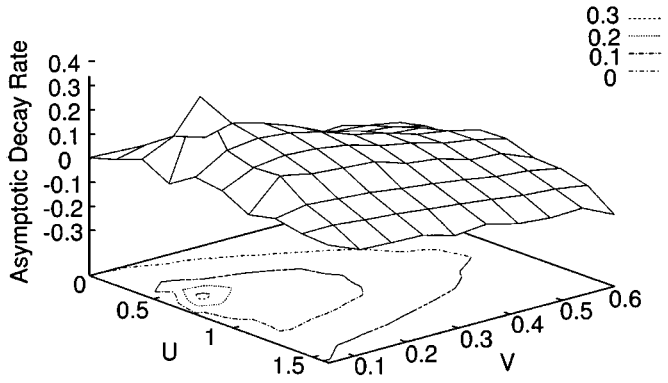

Fig. 3. Asymptotic decay rate α as a function of U and V.

rate α is a concave function with respect to both U and V with its maximum being approximately 0.3 at $(U,V) = (0.6, 0.1)$. The contour line at $\alpha = 0$ corresponds to the boundary of the stable region shown in Fig. 2. Once α is determined for a given (U,V) pair, one can readily determine β by substituting $z = -\alpha + j\beta$ in the characteristic equation (22) and equating the real and imaginary parts.

VI. Discrete-Time Implementation

So far, we have dealt with a continuous-time model of the MAX–MIN flow control problem. In reality, however, feedback information is relayed in RM cells, and thus not available in continuous time, but rather in sampled form. A recommended discrete-time implementation of the proposed ER allocation algorithm, (5) and (6), at a switch is as follows. Update the common part of ER periodically with an update interval T by

$$r[k+1] = \left[r[k] - \frac{A}{|\hat{Q}|}(\tilde{q}[k] - \tilde{q}[k-1]) - \frac{BT}{|\hat{Q}|}(\tilde{q}[k] - q_T) \right]^+, \quad A, B > 0 \qquad (31)$$

where $\tilde{q}[k]$ denotes the low-pass filtered queue length. Particularly in our simulation studies in Section VIII, we use a periodic-averaging filter such that $\tilde{q}[k] = (1/T)\int_{(k-1)T}^{kT} q(t')\,dt'$. Note that (31) corresponds to (6) as $T \to 0$ if $\tilde{q}[k] \approx q[k]$. In contrast to the periodic computation of the common part of ER, we recommend that per-VC ER allocation be performed aperiodically upon arrival of the corresponding RM cells in either forward path or backward path depending on the implementation. That is, upon arrival of VC i's RM cell at time t, the switch computes $r_i(t) = r(t) + m_i$ and writes the result on that RM cell where $r(t)$ is the present value of $r[k]$ and the value of m_i is available from either the RM cell being arrived or the MCR table being maintained in the switch, depending on the implementation. Therefore, the only per-VC operation required in our discrete-time ER allocation algorithm is single addition unless we use the MCR table.

Let us assume that the interarrival time between two adjacent feedback RM cells arriving at the source i is nearly constant in the asymptotic region and denote the constant by g_i. According to the ATM ABR specification [1], $g_i = (1+\mathrm{NRM})/f_i$ where NRM is the maximum number of data cells that source may send for each forward RM cell and f_i is the MAX–MIN fair rate of VC i. Note that g_i increases as NRM increases or f_i decreases. The question is how one can achieve the continuous-like performance under this discrete control. From Nyquist sampling theorem and from control theory it is known that, in order to have a continuous-like performance of the system, the ratio of the rise time of the system over the sampling time must fall into the interval (2, 4) [24]. Consider the principal root, $z^* = -\alpha \pm j\beta$, $\alpha > 0$, $\beta \geq 0$, of the normalized system (22). If $\beta = 0$, the system in the asymptotic region is nonoscillatory and hence the rise time of the system, denoted by T_r, is equal to the time constant α^{-1}. Otherwise, the system in the asymptotic region is oscillatory and the rise time of the system is given by $T_r = e^{\zeta/\tan\theta}/\omega_n$ where ω_n and ζ, satisfying $\alpha = \omega_n\zeta$ and $\beta = \omega_n\sqrt{\zeta^2 - 1}$, are the natural frequency and the damping ratio, respectively, and $\zeta = \cos\theta$ [24]. Therefore, in order for the closed-loop dynamics with discrete-time control to have the continuous-like performance, $(\tau T_r)/(g_i) \in (2,4)$, $\forall i$ must be satisfied.

VII. $|Q|$ Estimation

In Section IV, we showed that normalization of the controller gains by $|\hat{Q}|$ makes the closed-loop dynamics to be virtually independent of $|Q|$. However, underestimation of Q can cause system instability due to the following reason. Let a pair (U^*, V^*) satisfy the stability condition (24), i.e., reside inside the stable region depicted in Fig. 2. Suppose that we choose stable A and B such that $A = U^*/\tau$, $B = V^*/\tau^2$ by assuming an ideal $|Q|$ estimator in (23). The actual system is then governed by the normalized closed-loop equation (21) with the coefficients being $U = (|Q|/|\hat{Q}|)U^*$ and $V = (|Q|/|\hat{Q}|)V^*$. Consider the stable region depicted in Fig. 2. If $|Q|/|\hat{Q}|$ is less than 1, (U,V) is also stable since it resides somewhere on the straight line connecting the point (U^*, V^*) and the origin (0, 0). In contrast, as $|Q|/|\hat{Q}|$ increases beyond 1, the point (U,V) moves upward to the right along the straight line including the origin (0, 0) and the point (U^*, V^*) and eventually gets out of the stable region. In short, overestimation of $|Q|$ is tolerable since it does not affect the stability of the system (it only changes the asymptotic decay rate) whereas underestimation of $|Q|$ should be avoided since it can make the system unstable. This is why we introduce a certain margin in the $|Q|$ estimator design in the next.

Many schemes have been proposed to estimate the number of locally bottlenecked VCs [8], [13]–[15]. Each varies in the degree of implementation complexity. The basic idea in Su, de Veciana, and Walrand's algorithm [7], which estimates the number of ON sources sharing a link, is attractive since it does not require per-VC accounting. We modify the algorithm to estimate the number of locally bottlenecked VCs without doing per-VC accounting. A similar approach has been reported in [8]. Suppose that the jth RM cell arrives at a switch at the switch time t^j. According to the ABR specification [1], if the jth RM cell happens to be a RM cell of VC i, it carries the value $a_i(t^j - \tau_i^f)$ in the CCR field and the value m_i in the MCR field. The switch monitors the RM cell arrivals in a synchronous fashion over

TABLE I
RECOMMENDED VALUES FOR THE DESIGN PARAMETERS IN OUR SWITCH ALGORITHM

ER Allocation Algorithm				$\lvert Q\rvert$-Estimation Algorithm		
A	B	q_T	T	W	δ	λ
$\frac{0.6}{\tau_{max}}$	$\frac{0.1}{\tau_{max}^2}$	800 cells	32Δ	320Δ	0.9	0.98

($\tau_{max} = \max\{\tau_i,\ i \in N\}$, Δ=one cell transmission time)

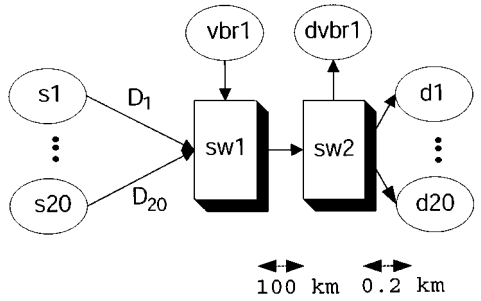

Fig. 4. Single-link configuration.

fixed-length intervals of W seconds. For the lth interval, the number of locally bottlenecked VCs can be approximated by

$$|Q|_l = \sum_{t^j \in ((l-1)W, lW]} \frac{\mathrm{NRM}+1}{W \cdot \mathrm{CCR}(t^j)} 1\{\mathrm{CCR}(t^j) - \mathrm{MCR}(t^j) \geq \delta \cdot r(t^j)\}, \quad 0 < \delta < 1 \qquad (32)$$

where $1\{\cdot\}$ is the indicator function, $\mathrm{CCR}(t^j)$ and $\mathrm{MCR}(t^j)$ respectively denote the value in the CCR field and the value in the MCR field of the jth RM cell, and $r(t^j)$ is the latest value of the common ER at time t^j. Upon arrival of the jth RM cell, if the current cell rate subtracted by the MCR is greater than or equal to the latest value of the common ER at the switch, the VC to which the jth RM cell belongs is counted as a locally bottlenecked VC. Otherwise, it is treated as a remotely bottlenecked VC. Here δ is the margin to avoid the underestimation of the number of locally bottlenecked VCs particularly near the steady state. As the system approaches the steady state, the current cell rate of a locally bottlenecked VC stays around the sum of the MCR and the common ER. Thus without the margin δ the VC could be counted wrongly as a remotely bottlenecked VC even for small perturbation in the current cell rate. By having this margin, however, one can effectively avoid this type of underestimation. Through simulations, we found that $\delta = 0.9$ is the recommended choice. Also note that the value of the indicator function is normalized by the expected number of RM cell arrivals of the VC within W seconds, $(W \cdot \mathrm{CCR})/(\mathrm{NRM}+1)$, so that the summation of these values over a W-second interval gives a correct estimate of the number of locally bottlenecked VCs. Based on this estimate for each interval, the recursive estimate is computed at the end of every interval as follows.

$$|\hat{Q}|(lW) = \mathrm{sat}_1^{|N|}[\lambda|\hat{Q}|((l-1)W) + (1-\lambda)|Q|_l], \quad 0 < \lambda < 1 \qquad (33)$$

where λ is an averaging factor and the saturation function ensures that $1 \leq |\hat{Q}|(t) \leq |N|$ for all t. The actual number of locally bottlenecked VCs can be zero for some network loading, but we intentionally lower-bound the value of its estimate by 1 to avoid the division by near-zero value in (31).

TABLE II
VC MODELS USED AND THE FAIR RATES SATISFYING THE MAX–MIN FAIRNESS WITH MCR GUARANTEE IN THE SINGLE-LINK CONFIGURATION. (THE UNITS OF PCR, MCR, ICR, AND THE FAIR RATES ARE IN Mb/s AND THE UNITS OF ARRIVAL AND DEPARTURE TIMES ARE IN SECONDS)

Source	PCR	MCR	ICR	Arr.	Dept.	Fair Rate			
						$0 \sim 2$	$2 \sim 4$	$4 \sim 6$	$6 \sim \infty$
s1 - s4	150	0	10	0	∞	41.1	38.9	35	37
s5 - s9	150	10	10	0	∞	51.1	48.9	45	47
s10	150	0	10	4	∞			35	37
s11 - s14	20	0	15	0	∞	20	20	20	20
s15 - s19	20	10	15	0	∞	20	20	20	20
s20	20	0	15	2	6		20	20	

TABLE III
VC MODELS USED AND THE FAIR RATES SATISFYING THE MAX–MIN FAIRNESS WITH MCR GUARANTEE IN THE PARKING LOT CONFIGURATION. (THE UNITS OF PCR, MCR, ICR, AND THE FAIR RATES ARE IN Mb/s AND THE UNITS OF ARRIVAL AND DEPARTURE TIMES ARE IN SECONDS)

Source	PCR	MCR	ICR	Arr.	Dept.	Fair Rate	Bottl.
s1, s5, s9	150	0	10	0	∞	21.67	SW 3
s13	150	0	10	0	∞	120	SW 4
s2, s6, s10	150	10	10	0	∞	31.67	SW 3
s14	150	10	10	0	∞	130	SW 4
s3, s7, s11	25	0	25	0	∞	21.67	SW 3
s15	25	0	25	0	∞	25	PCR
s4, s8, s12, s16	25	10	25	0	∞	25	PCR

The question is how to choose the interval W and the averaging factor λ. As the number of VCs sharing a link increases or the available bandwidth decreases for a given W, the interarrival time of RM cells of a VC increases so that the switch begins to see smaller number of RM cells of the VC for the interval and thus the estimate $|Q|_l$ gets to fluctuate largely. To solve this problem, one could possibly adjust W according to the changes of the number of VCs sharing the link and the available bandwidth, but this is not easy to implement. Instead, one can choose large λ at a value close to 1 in the hope that the averaging operation in (33) will effectively filter out the variability of $|Q|_l$. Through simulations we found that $\lambda = 0.98$ yields stable and effective estimation of $|Q|$ for a wide range of number of VCs sharing a link and the available bandwidth, irrespective of the choice of W. In Section VIII, through simulations we verify the performance of the proposed $|Q|$-estimation algorithm in conjunction with our ER allocation algorithm.

VIII. SIMULATION RESULTS

In this section, we present simulation results to verify the analysis given in the previous sections and demonstrate the excellent performance of our algorithm. The simulation model is developed on the NIST ATM simulator platform [25]. We consider two different network configurations, the single-link configuration and the parking lot configuration with multiple bot-

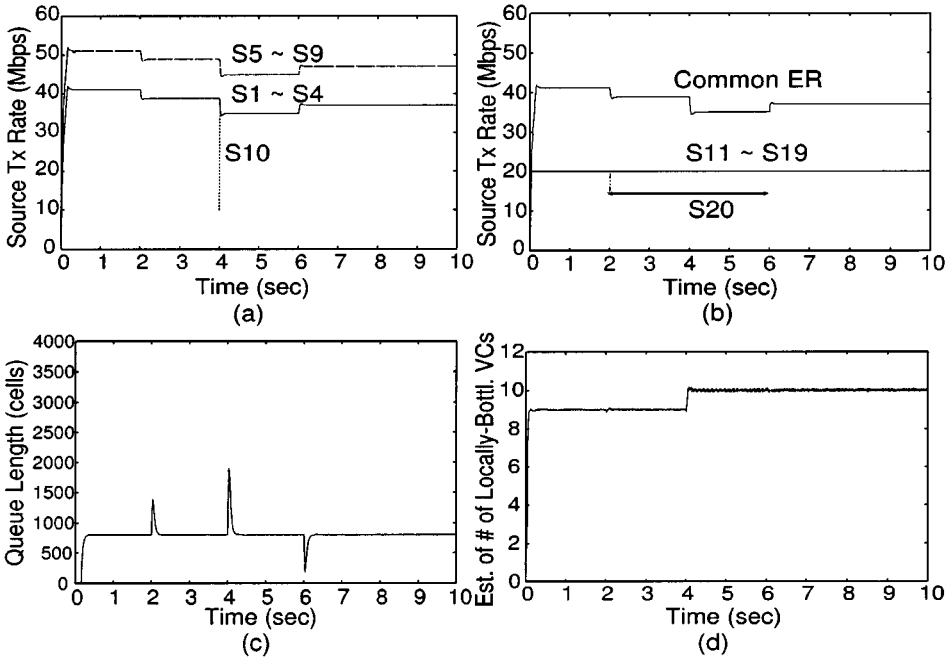

Fig. 5. Performance in the single-link configuration ($D_1, D_2 = 100$ km, $D_3, D_4 = 200$ km, $\ldots, D_{19}, D_{20} = 1000$ km, no VBR background traffic). (a) Source transmission rate $a_i(t)$ of the VCs with $\mathrm{PCR} = 150$ (Mb/s). (b) Source transmission rate $a_i(t)$ of the VCs with $\mathrm{PCR} = 20$ (Mb/s). (c) Queue length at the SW 1. (d) Estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 1.

tleneck links, which are fairly standard. In Table I, we summarize the recommended values for the design parameters in our proposed switch algorithm and use these values in the following simulation studies. We neglect additive increase and multiplicative decrease operation at each source to study the performance of the proposed ER allocation algorithm only separate from that of the binary feedback control mechanism.

First, we consider the single-link configuration, shown in Fig. 4, where 20 ABR VCs with different round-trip delays are contained and the capacity of all links is set equally at 600 Mb/s. To represent a WAN environment, we set the distances between individual sources and SW 1, denoted by D_i's, with the maximum being 1000 km. If we assume that the signal propagation speed is 2.0×10^5 km/s and that the queueing times are negligible, $\tau_{\max}$ is roughly 10 ms. The VC models used in this simulation configuration are summarized in Table II and all the sources are assumed to be persistent. Note that we vary the PCR value, the MCR value, the ICR value, and the arrival and departure times of the VCs in order to investigate the impact of the PCR-constrained sources, the differences in MCR and ICR, and the call activities on the network performance. For comparison purpose, we have computed the theoretical fair rates satisfying the MAX–MIN fairness with MCR guarantee for the given simulation scenario based on Proposition 3.1, and include the results in Table II. Observe that the fair rate of each VC varies in time according to the arrivals and departures of the other VCs and that the sources s1–s10 are bottlenecked at the SW 1 and the sources s11–s20 are bottlenecked at the access point by its PCR constraint. For example, the ABR sources s1–s4 should transmit data at 41.1 Mb/s in $[0, 2)$ s, 38.9 Mb/s in $[2, 4)$ s, 35 Mb/s in $[4, 6)$ s, and 37 Mb/s in $[6, \infty)$ s to be MAX–MIN fair. We generate 32 data cells between two adjacent forward RM cells, i.e., $\mathrm{NRM} = 32$.

Fig. 5 shows the simulation results when $D_1, D_2 = 100$ km, $D_3, D_4 = 200$ km, $\ldots, D_{19}, D_{20} = 1000$ km, and no VBR background traffic is applied. Observe from Fig. 5(a), (b) that the actual source transmission rates perfectly agree with the theoretical fair rates given in Table II. The transmission rates of s1–s4, s10 are equal to the common ER, $r(t)$, computed by the SW 1 since their MCR is 0 Mb/s, the transmission rates of s5–s9 are greater than the common ER by 10 Mb/s since their MCR is 10 Mb/s and s11–s20 are PCR constrained irrespective of their MCR values. The initial transient behavior is due to our initial condition that $r(0) = 0$ at both SW 1 and SW 2, i.e., it takes a time for the common ER value to ramp up to the operating point, which is, however, a phenomenon that hardly occurs during the normal operation. The queue length at the bottleneck node, SW 1, is shown in Fig. 5(c). The joins of s20 at 2 s and s10 at 4 s result in the surge of the queue length and the leave of s20 at 6 s results in the sudden drop of the queue length. The proposed algorithm, however, rapidly recovers the queue length to the target value q_T (=800 cells) and restabilizes it at the value. Fig. 5(d) shows the estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 1. Note that except the initial transient period this estimate perfectly agrees with the true value, $|Q|(t)$, which is shown to be 9 in $[0, 4)$ s and 10 in $[4, \infty)$ s in Table II. In conclusion, this result serves as evidence that the controller gains selected from the worst-case round-trip delay (in this example, $\tau_{\max} = 10$ ms) indeed achieve the system stability as in the theory with identical round-trip delays. We show more evidence in the next examples with change of round-trip delays and addition of VBR background traffic.

Fig. 6 summarizes the simulation results for the same single-link configuration as in Fig. 5, with the difference that we either add VBR background traffic (MPEG VBR video and periodic VBR traffic) or change the round-trip delays. First, Fig. 6(a), (b),

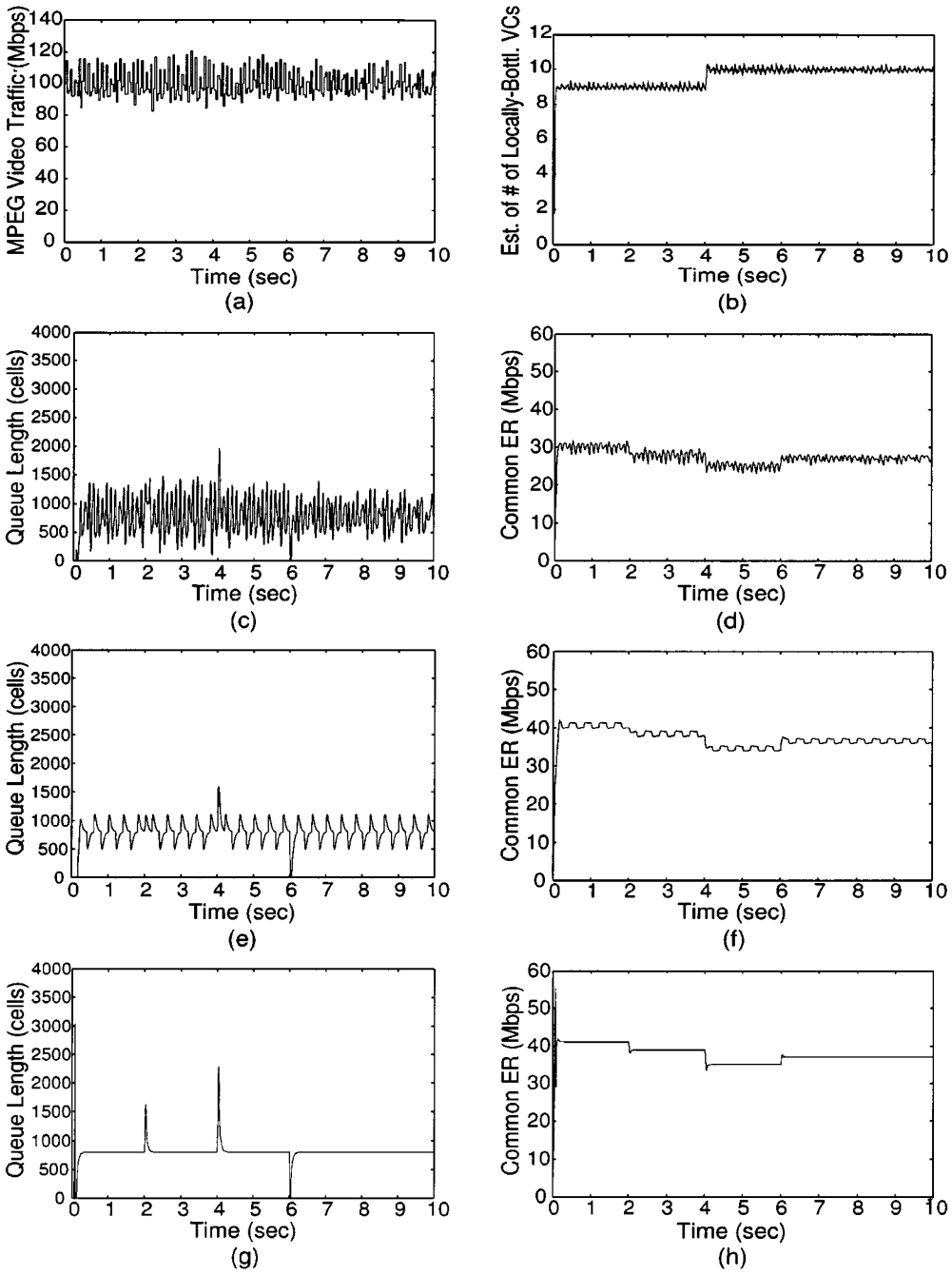

Fig. 6. Performance in the single-link configuration. I. Case of MPEG VBR background traffic: (a) Trace of MPEG VBR video traffic loaded. (b) Estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 1. (c) Queue length at the SW 1. (d) Common part of ER at the SW 1. II. Case of deterministic on/off background traffic: (e) Queue length at the SW 1. (f) Common part of ER at the SW 1. III. Case of $D_1 \sim D_{10} = 1000$ km and $D_{11} \sim D_{20} = 300$ km with no VBR background traffic: (g) Queue length at the SW 1. (h) Common part of ER at the SW 1.

(c), (d) show the control performance when MPEG VBR video traffic is applied while keeping the round-trip delays as identical as in the previous example. The MPEG video traffic was generated by superimposing 255 different 10-s video clips which were randomly selected parts of six different MPEG-1 encoded entertainment videos [26], and its average rate is approximately 100 Mb/s. See the trace in Fig. 6(a). As compare Fig. 6(b), (c), (d) with Fig. 5, we find that the macroscopic (time-average) behavior of the system is virtually identical to that of the case with no VBR traffic. The only differences are the reduced common ER value and the high-frequency fluctuation around the equilib-

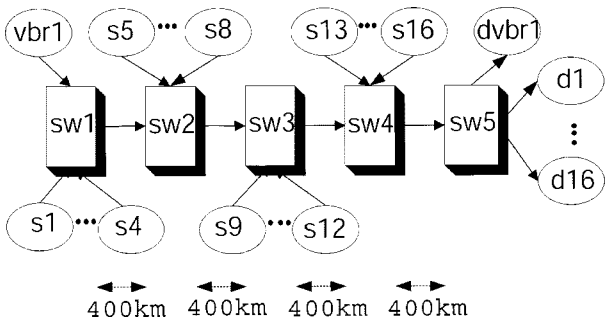

Fig. 7. Parking lot configuration.

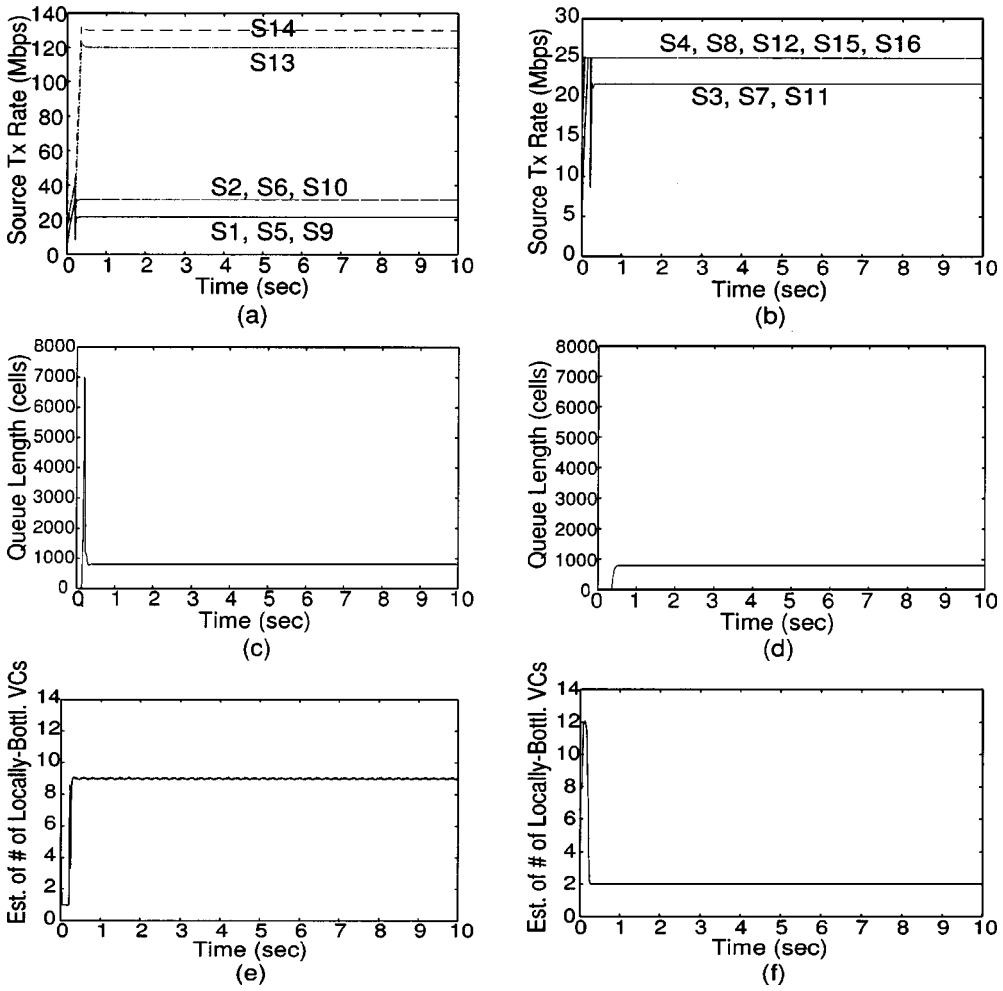

Fig. 8. Performance in the parking lot configuration (no VBR background traffic). (a) Source transmission rate $a_i(t)$ of the VCs with $\mathrm{PCR} = 150$ (Mb/s). (b) Source transmission rate $a_i(t)$ of the VCs with $\mathrm{PCR} = 25$ (Mb/s). (c) Queue length at the SW 3. (d) Queue length at the SW 4. (e) Estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 3. (f) Estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 4.

rium, which are obviously the consequence of the MPEG video traffic loaded in the background. From this example, we can readily conclude that the system remains stable with bounded transient performance and the time-average performance behaves as desired even in the presence of unpredictable high-frequency fluctuation of the MPEG background traffic. More specifically, the queue length at the bottleneck node, SW 1, stays around the target value q_T (=800 cells) yielding almost 100% link utilization and the $|Q|$ estimation remains stable and accurate. Second, we load a deterministic on/off source as background VBR traffic with the peak rate and the lengths of on and off periods being, respectively, 10 Mb/s, $20\tau_{\max}$ s, and $20\tau_{\max}$ s. Observe from Fig. 6(e) that the on/off behavior of the VBR background traffic causes the repeated surges and drops of the ABR queue length. The proposed ER allocation algorithm, however, rapidly recovers the queue length to the target value, thereby yielding stable queueing performance. Fig. 6(f) shows that the common ER value almost perfectly follows the on/off pattern of the VBR background traffic. Finally, we change the round-trip delays such that $D_1 \sim D_{10} = 1000$ km and $D_{11} \sim D_{20} = 300$ km with no VBR background traffic. As shown in Fig. 6(g), (h), both the queue length at the bottleneck node and the common ER behave almost identically as the previous example (Fig. 5), which consequently serves as an additional evidence that the controller gains selected from the worst-case round-trip delay, $\tau_{\max}$, indeed achieve the system stability as if the round-trip delays were all identical at $\tau_{\max}$.

Next, we consider the parking lot configuration, shown in Fig. 7, to study the case of multiple bottleneck nodes and VCs with different round-trip delays. Sixteen ABR VCs with different source locations are contained and the capacity of the links is set equally at 600 Mb/s, except that the link between SW 3 and SW 4 is 300 Mb/s. The VC models used in this simulation configuration are summarized in Table III and all the sources are assumed to be persistent. For comparison purpose, we also computed the theoretical fair rates satisfying the MAX–MIN fairness with MCR guarantee for the given simulation scenario, and include the results in Table III. To further clarify the scenario, we also include the theoretical bottleneck location of each VC in the table, which is the location at which each individual fair rate is determined. Fig. 8 shows the simulation results with no VBR background traffic. Observe from Fig. 8(a), (b) that the actual source transmission rates in the steady state perfectly agree with the theoretical fair rates given in Table III, irrespective of their round-trip delays and the bottleneck locations. The initial transient behavior is due to our initial condition that

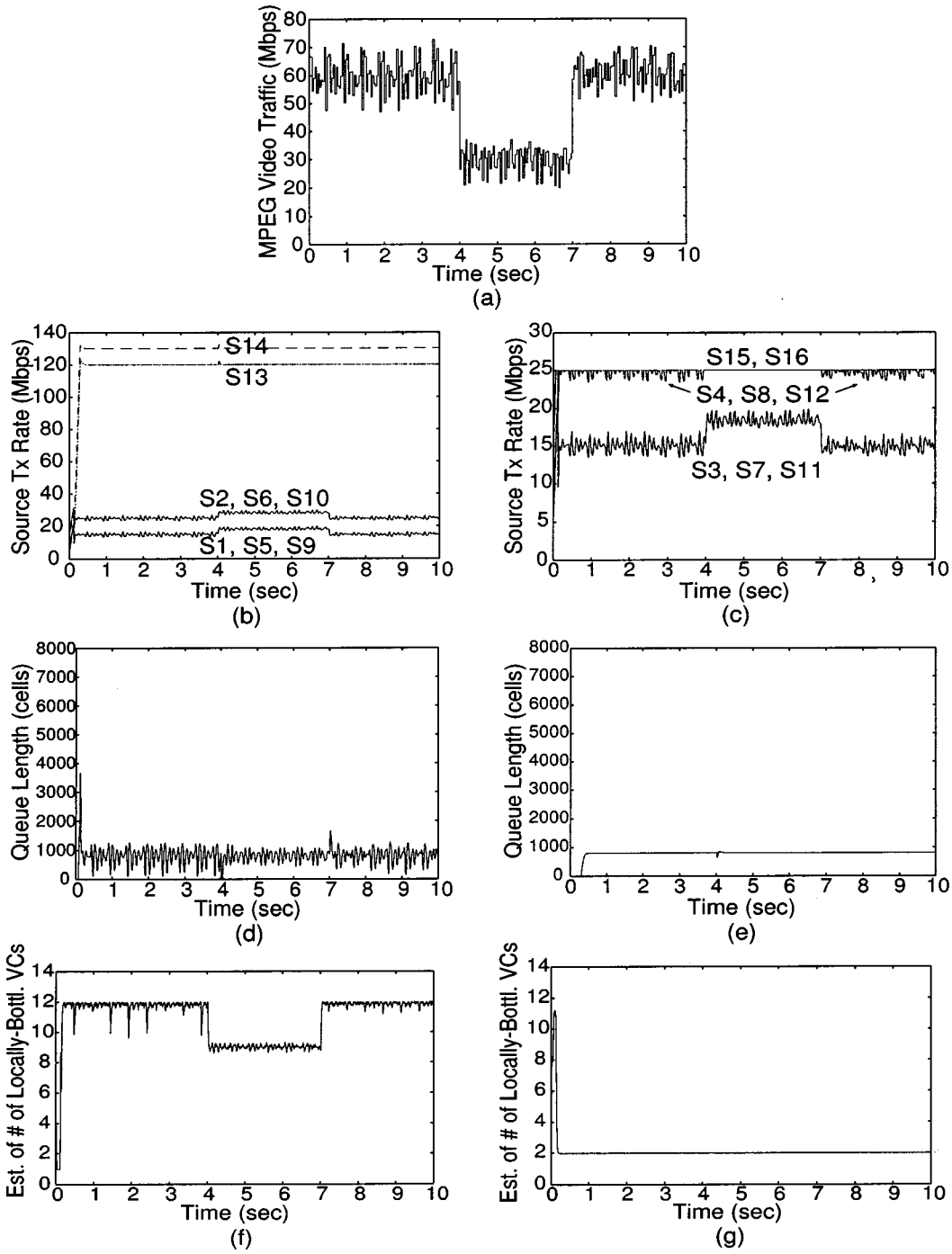

Fig. 9. Performance in the parking lot configuration (MPEG VBR background traffic). (a) Trace of MPEG VBR video traffic loaded. (b) Source transmission rate $a_i(t)$ of the VCs with $\mathrm{PCR} = 150$ (Mb/s). (c) Source transmission rate $a_i(t)$ of the VCs with $\mathrm{PCR} = 25$ (Mb/s). (d) Queue length at the SW 3. (e) Queue length at the SW 4. (f) Estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 3. (g) Estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 4.

$r(0) = 0$ at all the switches, which is again a phenomenon that hardly occurs during the normal operation. In the given scenario, there are two congested nodes, SW 3 and SW 4. As expected, the queue length at these congested nodes converges to the target value, 800 cells, which is shown in Fig. 8(c), (d). Fig. 8(e), (f) show the estimate of the number of locally bottlenecked VCs, $|\hat{Q}|(t)$, at the SW 3 and SW 4, respectively. We see that in the steady state the estimates stay around 9 and 2 at SW 3 and SW 4, respectively, which agrees with the data in Table III.

Finally, we study the effect of VBR background traffic in the parking lot configuration. Again, the VBR background traffic was generated by superimposing multiple 10-s video clips which were randomly selected parts of six different MPEG-1 encoded entertainment videos. Its average rate is approximately 60 Mb/s (superposition of 156 video clips) in $[0, 4)$ s and $[7, \infty)$ s, and 30 Mb/s (superposition of 74 video clips) in $[4, 7)$ s, where the transitions from 60 to 30 Mb/s and back to 60 Mb/s respectively represent the simultaneous leave

and join of multiple video streams. See the trace in Fig. 9(a). Upon transition of video traffic from 60 to 30 Mb/s, the rates of s1–s3, s5–s7, and s9–s11 equally increase while the rates of s13 and s14 remain unchanged, as shown in Fig. 9(b), (c). This is because the rates allocated to s13 and s14 are sufficiently greater than the rate allocated to s1–s3, s5–s7, and s9–s11, so that increasing the rates of s13 and s14 would not be MAX–MIN fair. Another thing to note is the high-frequency fluctuation of s4, s8, and s12's transmission rates in the periods, $[0, 4)$ s and $[7, \infty)$ s, in Fig. 9(c). This fluctuation implies that s4, s8, and s12 are no longer constantly PCR-constrained at 25 Mb/s. In fact, their bottleneck location is alternating between SW 3 and the source due to the underlying fluctuation of video traffic in these periods. This is also why in Fig. 9(f) the estimate of number of locally bottlenecked VCs at SW 3 varies between 9 and 12 in these periods rather than constantly indicating 9 as in Fig. 8(e). Note, however, that the estimate tends to be closer to 12 rather than 9 due to the margin δ in the $|Q|$ estimator, which implies that the $|Q|$ estimator would improve the system stability against high-frequency perturbation. The high-frequency dynamics of the MPEG video traffic yields the high-frequency oscillation of the queue length at SW 3 around its target value but never causes the system instability, which means the transient performance is well bounded under our control [see Fig. 9(d)].

APPENDIX

Pontryagin's criterion [20], [22], which we state below for convenience, yields Proposition 4.1.

Consider the exponential polynomial

$$H(z) = \sum_{l=0}^{L} \sum_{m=0}^{M} b_{lm} z^l (e^z)^m = 0 \tag{34}$$

with the principal term $b_{LM} z^L (e^z)^M$. Let $F(\omega) = \operatorname{Re} H(i\omega)$, $G(\omega) = \operatorname{Im} H(i\omega)$. In order for all the zeros of $H(z)$ to have negative real parts, it is sufficient and necessary that

1) $G(\omega)$ [or $F(\omega)$] has exactly $4kM + L$ real zeros in the interval $-2k\pi + \epsilon \le \omega \le 2k\pi + \epsilon$ starting with sufficiently large k, ϵ being some appropriate constant.
2) For each zero of $G(\omega)$ [or $F(\omega)$], to be denoted by ω_i

$$F(\omega_i)\dot{G}(\omega_i) > 0 \quad [\text{or } -\dot{F}(\omega_i)G(\omega_i) > 0]. \tag{35}$$

Proof of Proposition 4.1: From (22), we have $L = 2$, $M = 1$

$$F(\omega) = -\omega^2 \cos\omega + V, \quad G(\omega) = -\omega^2 \sin\omega + U\omega \tag{36}$$

and

$$F(\omega)\dot{G}(\omega) = (\omega^2 \cos\omega - V)(\omega^2 \cos\omega + 2\omega \sin\omega - U). \tag{37}$$

First, we show that if $0 < U < \pi/2$, $G(\omega)$ satisfies the condition 1) and, moreover, there exists no nonzero zero of $G(\omega)$, say ω_i, for which the following two inequalities

$$\omega_i^2 \cos\omega_i \le 0 \quad \text{and} \quad \omega_i^2 \cos\omega_i + \omega_i \sin\omega_i > 0 \tag{38}$$

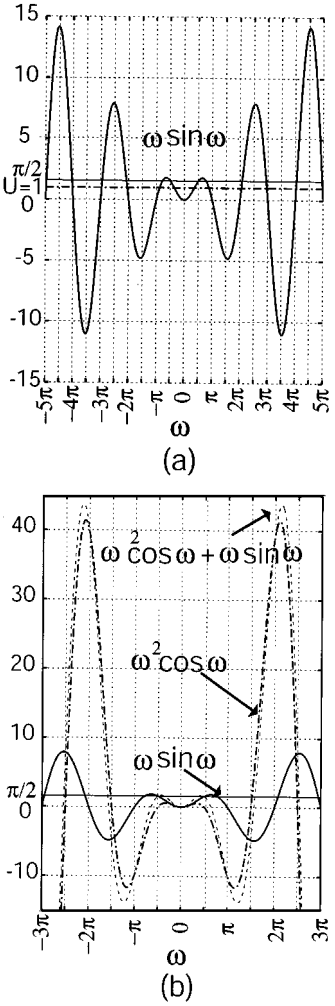

Fig. 10. Figures used in the proof of Proposition 4.1.

hold simultaneously. If we rewrite $G(\omega) = 0$ by

$$\omega(\omega \sin\omega - U) = 0 \tag{39}$$

we know that the zeros of $G(\omega)$ consist of $\omega = 0$ and the roots of $\omega \sin\omega = U$. Let us investigate the intersections between the function $\omega \sin\omega$ and the line U for $0 < U < \pi/2$, in Fig. 10(a). Take $\epsilon = \pi/2$. Then, for $k = 1$, the number of intersections between the function $\omega \sin\omega$ and the line U is five and hence $G(\omega)$ has six real zeros including $\omega = 0$ in the interval $[-(3/2)\pi, (5/2)\pi]$, as required. If we increase k from 1 to 2, the total number of intersections between the function $\omega \sin\omega$ and the line U increases by four and hence $G(\omega)$ has ten real zeros including $\omega = 0$ in the interval $[-(7/2)\pi, (9/2)\pi]$. In the same manner, the total number of real zeros of $G(\omega)$ in the corresponding interval continues to increase by four as k increases by one. Therefore, $G(\omega)$ has the appropriate number of real zeros as required if $0 < U < \pi/2$. On the other hand, by carefully examining Fig. 10(b), we can readily verify that there exists no nonzero zero of $G(\omega)$, say ω_i, for which the two inequalities in (38) hold simultaneously. Moreover, we see from Fig. 10(b) that the intervals $(-2\pi j - \pi, -2\pi j - (\pi/2))$, $(-2\pi j - (\pi/2), -2\pi j)$, $(2\pi j, 2\pi j + (\pi/2))$, and $(2\pi j + (\pi/2), 2\pi j + \pi)$ for $j = 0, 1, 2, \ldots$, include one and only one nonzero zero of $G(\omega)$, respectively.

Next we find the condition on V to satisfy the condition 2), i.e., each zero of $G(\omega)$ must satisfy

$$F(\omega)\dot{G}(\omega) = (\omega^2 \cos\omega - V)(\omega^2 \cos\omega + 2\omega \sin\omega - U) > 0 \tag{40}$$

when $0 < U < \pi/2$. Since $\omega = 0$ is a zero of $G(\omega)$, by substituting $\omega = 0$ in (40), we get $UV > 0$ which implies $V > 0$ since $U > 0$. Since $U = \omega_i \sin\omega_i$ for each nonzero zero ω_i of $G(\omega)$, if we substitute this in (40), we get the following conditions to be satisfied: for each ω_i,

$$\left(\omega_i^2 \cos\omega_i - V\right)\left(\omega_i^2 \cos\omega_i + \omega_i \sin\omega_i\right) > 0. \tag{41}$$

Consider ω_i's positioned in the intervals, $(-2\pi j - (\pi/2), -2\pi j)$ and $(2\pi j, 2\pi j + (\pi/2))$, $j = 0, 1, 2, \ldots$. For these zeros, we know from Fig. 10(b) that $\omega_i^2 \cos\omega_i + \omega_i \sin\omega_i > 0$ and $\omega_i^2 \cos\omega_i > 0$. Thus in order for (41) to hold for these zeros, we need conditions on V that $\omega_i^2 \cos\omega_i > V$ for each of these zeros. These conditions can be further simplified since we know from Fig. 10(b) that $\omega_i^2 \cos\omega_i$ has the smallest value for the zero positioned in the interval $(0, \pi/2)$, say ω_1, which is the unique solution of $U = \omega \sin\omega$ in the interval $(0, \pi/2)$. Therefore, $\omega_1^2 \cos\omega_1 > V$ is a necessary and sufficient condition. Next, consider ω_i's positioned in the intervals, $(-2\pi j - \pi, -2\pi j - \pi/2)$ and $(2\pi j + (\pi/2), 2\pi j + \pi)$, $j = 0, 1, 2, \ldots$. For these zeros, we know from Fig. 10(b) that $\omega_i^2 \cos\omega_i + \omega_i \sin\omega_i < 0$ and $\omega_i^2 \cos\omega_i < 0$ since none of them satisfies the two inequalities in (38) simultaneously. Thus (41) holds for these zeros since $V > 0$.

In conclusion, the condition on V to satisfy the condition 2) when $0 < U < \pi/2$ is $0 < V < \omega_1^2 \cos\omega_1$ which becomes the second condition in (24) by using the relationships $U = \omega_1 \sin\omega_1$ and $\sin^2\omega_1 + \cos^2\omega_1 = 1$. Q.E.D.

REFERENCES

[1] ATM Forum Traffic Management Specification, Version 4.0 (1996, Apr.). [Online]. Available: ftp://ftp.atmforum.com/pub/approved-specs/af-tm-0056.00.ps

[2] F. Bonomi and K. W. Fendick, "The rate-based flow control framework for the available-bit-rate ATM service," *IEEE Network*, vol. 9, pp. 25–39, Feb. 1995.

[3] E. Hernandez-Valencia, L. Benmohamed, R. Nagarajan, and S. Chong, "Rate-control algorithms for the ATM ABR service," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 7–20, 1997.

[4] L. Benmohamed and S. M. Meerkov, "Feedback control of congestion in packet switching networks: The case of single congested node," *IEEE/ACM Trans. Networking*, vol. 1, pp. 693–708, Dec. 1993.

[5] ——, "Feedback control of congestion in packet-switching networks: The case of multiple congested nodes," *Int. J. Commun. Syst.*, vol. 10, no. 5, pp. 227–246, 1997.

[6] A. Kolarov and G. Ramamurthy, "A control-theoretic approach to the design of an explicit rate controller for ABR service," *IEEE/ACM Trans. Networking*, vol. 7, pp. 741–753, Aug. 1999.

[7] C. F. Su, G. de Veciana, and J. Walrand, "Explicit rate flow control for ABR services in ATM networks," *IEEE/ACM Trans. Networking*, vol. 8, pp. 350–361, June 2000.

[8] M. K. Wong and F. Bonomi, "A novel explicit rate congestion control algorithm," in *Proc. IEEE GLOBECOM*, vol. 4, 1998, pp. 2432–2439.

[9] Y. Choi, S. Kang, and S. Chong, An efficient ABR service engine for ATM switch, 2000. preprint.

[10] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice Hall, 1992.

[11] S. P. Abraham and A. Kumar, "MAX–MIN fair rate control of ABR connections with nonzero MCRs," in *Proc. IEEE GLOBECOM*, 1997, pp. 498–502.

[12] ——, "A Stochastic approximation approach for MAX–MIN fair adaptive rate control of ABR sessions with MCRs," in *Proc. IEEE INFOCOM*, 1998, pp. 1358–1365.

[13] L. Kalampoukas, A. Varma, and K. K. Ramakrishnan, "An efficient rate allocation algorithm for ATM networks providing MAX–MIN fairness," Computer Engineering Dept., Univ. of California, Santa Cruz, CA, Tech. Rep. UCSC-CRL-95-29, June 1995.

[14] A. Charny, K. K. Ramakrishnan, and A. Lauck, "Time scale analysis and scalability issue for explicit rate allocation in ATM networks," *IEEE/ACM Trans. Networking*, vol. 4, pp. 569–581, Aug. 1996.

[15] S. Kalyanaraman, R. Jain, S. Fahmy, R. Goyal, and B. Vandalore, "The ERICA switch algorithm for ABR traffic management in ATM networks," *IEEE/ACM Trans. Networking*, vol. 8, pp. 87–98, Feb. 2000.

[16] N. Ghani and J. W. Mark, "Enhanced distributed explicit rate allocation for ABR services in ATM networks," *IEEE/ACM Trans. Networking*, vol. 8, pp. 71–86, Jan. 2000.

[17] C. Fulton, S. Q. Li, and C. S. Lim, "An ABR feedback control scheme with tracking," in *Proc. IEEE INFOCOM*, vol. 2, 1997, pp. 806–815.

[18] F. Bonomi, D. Mitra, and J. B. Seery, "Adaptive algorithms for feedback-based flow control in high-speed, wide-area ATM networks," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1267–1283, July 1995.

[19] K. K. Ramakrishnan and R. Jain, "A binary feedback scheme for congestion avoidance in computer networks with a connectionless network layer," in *Proc. ACM SIGCOMM*, 1988, pp. 303–313.

[20] R. Bellman and K. L. Cooke, *Differential-Difference Equations*. New York: Academic, 1963.

[21] G. Stépán, *Retarded Dynamical Systems: Stability and Characteristic Functions*. White Plains, NY: Longman, 1989.

[22] S. J. Bhatt and C. S. Hsu, "Stability criteria for second-order dynamical systems with time lag," *J. Appl. Mech.*, pp. 113–118, 1966.

[23] F. Brauer, "Decay rates for solutions of a class of differential-difference equations," *SIAM J. Math. Anal.*, vol. 10, no. 4, pp. 783–788, 1979.

[24] K. J. Åström and B. Wittenmark, *Computer Controlled Systems: Theory and Design*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[25] *The NIST ATM Network Simulator*, NIST, 1995.

[26] (2001). [Online]. Available: http://www-info3.informatik.uni-wuerzburg.de/rose

Song Chong (M'93) received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Korea, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin.

From 1994 to 1996, he was a Member of Technical Staff at the Performance Analysis Department, AT&T Bell Laboratories, Holmdel, NJ, where he worked on design and analysis of high-performance broadband/ATM switches. From 1996 to 2000, he was with the Department of Electronic Engineering, Sogang University, Korea, as an Assistant Professor. He is currently an Assistant Professor of the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Korea. His current research interests are in the areas of high-speed communication networks, high-performance switching/routing systems, multimedia networking, and performance evaluation. He has published more than 20 papers in international journals and conferences and holds two U.S. patents with several pending in these areas.

Dr. Chong is a member of the Association for Computing Machinery. He served as a member of the Technical Program Committee for IEEE INFOCOM from 1997 to 1999.

Sangho Lee received the B.S. and M.S. degrees in electronic engineering from Sogang University, Korea.

He is an R&D Engineer with Samsung Electronics, Sungnam, Korea, where he is currently developing software for ATM core switches.

Sungho Kang (M'89) received the B.S. degree from Seoul National University, Seoul, Korea, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Texas at Austin.

He was a Post Doctoral Fellow at the University of Texas at Austin, a Research Scientist at Schlumberger Laboratory for Computer Science, Schlumberger Inc., and a Senior Staff Engineer at the Semiconductor Systems Design Technology, Motorola Inc. Since 1994, he has been an Associate Professor of Department of Electrical and Electronic Engineering, Yonsei University, Korea. His current research interests include VLSI design, VLSI CAD, VLSI testing, and design for testability.