Control-Theoretic MAX-MIN Flow Control with Minimum Rate Guarantee *

Song Chong
Dept. of EECS
KAIST
Taejon 305-701, Korea
song@ee.kaist.ac.kr

Sang-Ho Lee
Telecom R&D Center
Samsung Electronics
Sungnam 463-050, Korea
sangho@telecom.samsung.co.kr

Sungho Kang
Dept. of EE
Yonsei University
Seoul 120-749, Korea
shkang@yonsei.ac.kr

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 context of the ATM 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-VC (virtual circuit) 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 not only in the context of ATM ABR service but also in general elastic traffic services.**

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.

The ATM Forum has adopted the rate-based closed-loop control approach for the flow control of ABR service [1]. Our primary concern in this paper is Explicit Rate (ER) allocation algorithm for the flow control. We construct a new ER allocation algorithm on a solid analytical basis and the novelty of our proposed algorithm is an explicit control of both rate and queue dynamics.

Benmohamed and Meerkov in their pioneering work [2] 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.

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

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 τ_{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 τ_{max}, and requires a large number of floating point multiplications every discrete time slot. Therefore, its practical use is limited as the round-trip delay increases [3].

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.

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

where A and B are positive 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

*This work was supported in part by Korea Science and Engineering Foundation under grant 98-0101-07-01-3 and by Samsung Electronics Co., Korea.

0-7803-7206-9/01/$17.00 © 2001 IEEE

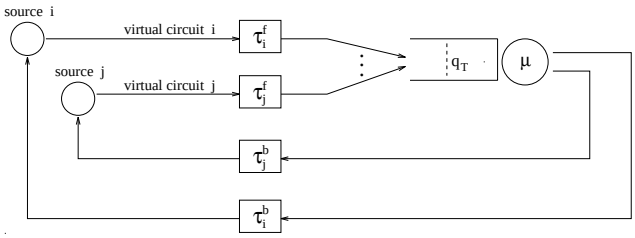

Fig. 1. Network model with a node of interest.

Q. Note that the proposed algorithm is in fact a special case of Benmohamed and Meerkov's algorithm (1) with $I = 1$, $\alpha_1 = -\frac{A}{|Q|}$, $\alpha_0 + \alpha_1 = \frac{BT}{|Q|}$ and $\beta_j = 0,\ \forall\ j$.

By the term an acceptable level of control, we mean that by properly choosing the controller gains for a 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 the equation (2). Another notable feature of the proposed algorithm is the normalization of the controller gains by the number of locally-bottlenecked VCs $|Q|$. We show that this normalization is indeed beneficial such a way that it makes the closed-loop performance to be virtually independent of the number of locally-bottlenecked VCs on a link.

II. Explicit Rate Flow Control - A Fluid Model

Consider a network model in Figure 1 where we model a single node explicitly and the other nodes implicitly to simplify the analysis. We then use fluid flow analysis which is fairly standard [4]. Assume that the round-trip delay, τ_i, of a VC i, which is the sum of forward-path delay τ_i^f and backward-path delay τ_i^b, is constant and the sources are *persistent* until the system reaches steady state. We also assume that the available bandwidth μ at the link is constant until the system reaches steady state and the buffer size at the link is 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

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

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

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

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.

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

and

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

where $A,\ B > 0$ and m_i denotes 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.

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

and the set of all the remotely-bottlenecked VCs, $N-Q$, at the link of interest is given by $N-Q = \{i | i \in N \text{ and } a_{is} = \min[b_{is}, p_i]\}$.

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

$$a_{is} = \min[r_{is}, b_{is}, p_i], \quad r_{is} = r_s + m_i, \qquad \forall\ i \in N, \tag{7}$$

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

$$\sum_{i\in N} a_{is} = \mu. \tag{8}$$

By combining the equations (7), (8) and the definitions of Q and $N-Q$, we obtain

$$\sum_{i\in Q} r_s + \sum_{i\in Q} m_i + \sum_{i\in N-Q} \min[b_{is}, p_i] = \mu \tag{9}$$

which implies that

$$r_s = \frac{\mu - \sum_{i\in N-Q} \min[b_{is}, p_i] - \sum_{i\in Q} m_i}{|Q|}. \tag{10}$$

The following proposition states the result.

Proposition III.1: For $\sum_{i\in N} m_i < \mu$ and $\min[b_{is}, p_i] \geq m_i$, there exists a unique steady state solution (equilibrium point) at which (i) the queue length is equal to the target queue length ($q_s = q_T$), (ii) the available bandwidth at the link is fully utilized ($\sum_{i\in N} a_{is} = \mu$), and (iii) 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,

$$a_{is} = \begin{cases} \frac{\mu - \sum_{i\in N-Q} \min[b_{is}, p_i] - \sum_{i\in Q} m_i}{|Q|} + m_i, & i \in Q \\ \min[b_{is}, p_i], & i \in N-Q. \end{cases} \tag{11}$$

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 as Benmohamed and Meerkov's algorithm (1) has [2].

IV. Asymptotic Stability

Suppose that there exists a neighborhood of the equilibrium point in which the followings are satisfied: a) $b_i(t) = b_{is},\ \forall\ i \in N$, i.e., the dynamics of the other nodes are in steady state; b) $\{i | i \in N \text{ and } a_i(t) = r_i(t - \tau_i^b)\} = Q$ 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 $\min[b_{is}, p_i]$; c) the saturation nonlinearities in (4) and (6) are not activated, i.e., both $q(t)$ and $r(t)$ are positive-valued; d) the $|Q|$-estimation process is in steady state, i.e., $|\hat{Q}|$ is constant. Then, in this neighborhood, we can simplify the dynamic equations (3), (4) and (6) as follows.

$$a_i(t) = \begin{cases} r_i(t - \tau_i^b), & i \in Q \\ \min[b_{is}, p_i], & i \in N - Q, \end{cases} \tag{12}$$

$$\dot{q}(t) = \sum_{i\in N} a_i(t - \tau_i^f) - \mu \tag{13}$$

and

$$\dot{r}(t) = -\frac{A}{|\hat{Q}|}\dot{q}(t) - \frac{B}{|\hat{Q}|}(\ q(t) - q_T\ ). \tag{14}$$

By combining (12) and (13), we obtain

$$\dot{q}(t) = \sum_{i\in N-Q} \min[b_{is}, p_i] + \sum_{i\in Q} r_i(t - \tau_i) - \mu. \tag{15}$$

Define an error function by $e(t) = q(t) - q_T$. By combining (14), the differentiation of (15) and the differentiation of (5), we obtain the following closed-loop equation

$$\ddot{e}(t) + \frac{A}{|\hat{Q}|}\sum_{i\in Q} \dot{e}(t - \tau_i) + \frac{B}{|\hat{Q}|}\sum_{i\in Q} e(t - \tau_i) = 0 \tag{16}$$

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

$$D(s) = s^2 + \frac{A}{|\hat{Q}|}\sum_{i\in Q} s e^{-s\tau_i} + \frac{B}{|\hat{Q}|}\sum_{i\in Q} e^{-s\tau_i} = 0 \tag{17}$$

which has infinite number of roots. For the asymptotic stability of the closed-loop equation (16), all the roots of the characteristic equation (17) must have negative real parts [9].

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 (16) becomes

$$\ddot{e}(t) + \frac{|Q|}{|\hat{Q}|}A\dot{e}(t - \tau) + \frac{|Q|}{|\hat{Q}|}Be(t - \tau) = 0. \tag{18}$$

This equation is normalized so that the time lag τ becomes unity. Let $t = \tau\xi$. In terms of the new variable ξ, (18) becomes

$$\ddot{e}(\xi) + U\dot{e}(\xi - 1) + Ve(\xi - 1) = 0 \tag{19}$$

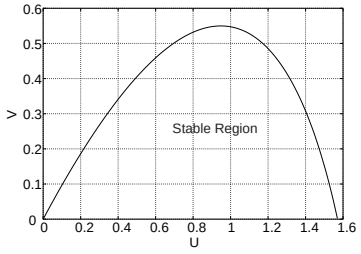

Fig. 2. Stable region with respect to U and V.

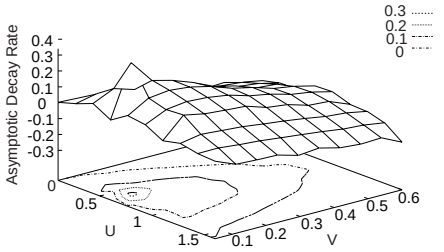

Fig. 3. The asymptotic decay rate α as a function of U and V.

where $U = \frac{|Q|}{|\hat{Q}|}A\tau$ and $V = \frac{|Q|}{|\hat{Q}|}B\tau^2$. The characteristic equation of (19) is

$$H(z) = z^2e^z + Uz + V = 0. \tag{20}$$

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 [9], [10] which yields the following result.

Proposition IV.1: Let

$$U = \frac{|Q|}{|\hat{Q}|}A\tau \quad \textit{and} \quad V = \frac{|Q|}{|\hat{Q}|}B\tau^2. \tag{21}$$

The closed-loop equation (18) is asymptotically stable if and only if

$$0 < U < \frac{\pi}{2} \quad \textit{and} \quad 0 < V < {\omega_1}^2\sqrt{1-(\frac{U}{\omega_1})^2} \tag{22}$$

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

For better presentation, we plot this stability condition in Figure 2. On the other hand, 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 (17). 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 τ_{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 (19) 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 $V = B\tau^2$.

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 (19) can be represented by a series [9]

$$e(\xi) = \sum_{n=1}^{\infty} p_n(\xi)e^{z_n\xi} \tag{23}$$

where $p_n(\xi)$ is a suitable polynomial and $z_n,\ \forall\ n$, are the roots of the corresponding characteristic equation (20). 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 (23) that for the original variable $t(= \tau\xi)$,

$$\|e(t)\| \leq ce^{-\frac{\alpha}{\tau}t} \quad \text{for large } t. \tag{24}$$

where $\|\cdot\|$ denotes the Euclidean norm and c is a constant depending on the initial conditions of (18). Note that $\frac{\alpha}{\tau}$ is the asymptotic decay rate at which the original system tends to the equilibrium point. Hence the inverse of it, $\frac{\tau}{\alpha}$, 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 [11] 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 (20) to

$$(\psi^2 - 2\sigma\psi + \sigma^2)e^{\psi} + Ue^{\sigma}\psi + (V - U\sigma)e^{\sigma} = 0. \tag{25}$$

For a given (U, V) pair satisfying the stability condition (22), if we choose σ to be the supremum of positive real numbers for which the transformed characteristic equation (25) 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 (25)

$$\ddot{e}(\varsigma)-2\sigma\dot{e}(\varsigma)+\sigma^2 e(\varsigma)+Ue^{\sigma}\dot{e}(\varsigma-1)+(V-U\sigma)e^{\sigma}e(\varsigma-1)=0. \tag{26}$$

For a given (U,V) pair satisfying the stability condition (22), 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. Figure 3 shows the result obtained by this numerical approach. We found that the asymptotic decay 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 Figure 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 (20) and equating the real and imaginary parts.

VI. Discrete-Time Implementation

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

$$r[k+1]=[\ r[k]-\frac{A}{|\hat{Q}|}(\tilde{q}[k]-\tilde{q}[k-1])-\frac{BT}{|\hat{Q}|}(\tilde{q}[k]-q_T)\ ]^+ \tag{27}$$

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 $\tilde{q}[k]=\frac{1}{T}\int_{(k-1)T}^{kT} q(t')dt'$. Note that (27) 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. Refer to [6] for the implementation details.

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 (22), i.e., reside inside the stable region depicted in Figure 2. Suppose that we choose stable A and B such that $A=\frac{U^*}{\tau}$, $B=\frac{V^*}{\tau^2}$ by assuming an ideal $|Q|$ estimator in (21). The actual system is then governed by the normalized closed-loop equation (19) with the coefficients being $U=\frac{|Q|}{|\hat{Q}|}U^*$ and $V=\frac{|Q|}{|\hat{Q}|}V^*$. Consider the stable region depicted in Figure 2. If $\frac{|Q|}{|\hat{Q}|}$ 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 $\frac{|Q|}{|\hat{Q}|}$ 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 [5], [7], [8]. Each varies in the degree of implementation complexity. The basic idea in Su, de Veciana and Walrand's algorithm [4], 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 [5]. Suppose that the j^{th} RM cell arrives at a switch at the switch time t^j. According to the ABR specification [1], if the j^{th} 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 fixed-length intervals of W seconds. For the l^{th} interval, the number of locally-bottlenecked VCs can be approximated by

$$|Q|_l=\sum_{t^j\in((l-1)W,\ lW]}\frac{NRM+1}{W\cdot CCR(t^j)}1\{CCR(t^j)-MCR(t^j)\geq\delta r(t^j)\} \tag{28}$$

where $0<\delta<1$, $1\{\cdot\}$ is the indicator function, $CCR(t^j)$ and $MCR(t^j)$ respectively denote the value in the CCR field and the value in the MCR field of the j^{th} RM cell, and $r(t^j)$ is the latest value of the common ER at time t^j. Upon arrival of the j^{th} 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 j^{th} 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 ap-

proaches 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, $\frac{W \cdot CCR}{NRM+1}$, 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.

$$|\hat{Q}|(lW) = \text{sat}_1^{|N|}[\,\lambda|\hat{Q}|((l-1)W) + (1-\lambda)|Q|_l\,], \quad 0 < \lambda < 1 \tag{29}$$

where λ is an averaging factor and the saturation function ensures that $1 \le |\hat{Q}|(t) \le |N|$ for all t. 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.

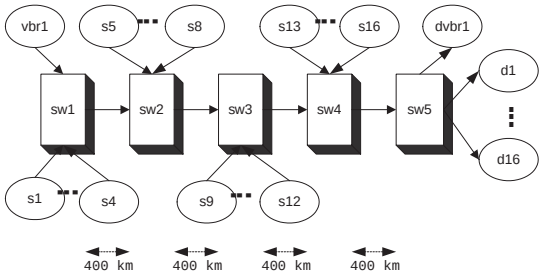

Fig. 4. The parking lot configuration.

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

TABLE I

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 Mbps).

VIII. Simulation Results

In the simulation studies we let $A = \frac{0.6}{\tau_{max}}$, $B = \frac{0.1}{\tau_{max}^2}$, $q_T = 800$ cells, $T = 32\Delta$, $W = 320\Delta$, $\delta = 0.9$ and $\lambda = 0.98$ where Δ is one cell transmission time. 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. We generate 32 data cells between two adjacent forwards RM cells, i.e., $NRM = 32$.

We consider a parking lot configuration, shown in Figure 4, to study the case of multiple bottleneck nodes and VCs with different round-trip delays. 16 ABR VCs with different source locations are contained and the capacity of the links is set equally at 600 Mbps except that the link between SW 3 and SW 4 is 300 Mbps. The VC models used in this simulation configuration are summarized in Table I 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 I. 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.

Figure 5 shows the simulation results with no VBR background traffic. Observe from Figures 5 a, b that the actual source transmission rates in the steady state perfectly agree with the theoretical fair rates given in Table I, irrespective of their round-trip delays and the bottleneck locations. The initial transient behavior is due to our initial condition that $r(0) = 0$ at all the switches, which is 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 Figures 5 c, d. Figures 5 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 I.

Next we study the effect of VBR background traffic in the parking lot configuration. The VBR background traffic was generated by superimposing multiple 10 sec video clips which were randomly selected parts of 6 different MPEG-1 encoded entertainment videos [12]. Its average rate is approximately 60 Mbps (superposition of 156 video clips) in [0, 4) sec and [7, ∞) sec and 30 Mbps (superposition of 74 video clips) in [4, 7) sec, where the transitions from 60 Mbps to 30 Mbps and back to 60 Mbps respectively represent the simultaneous leave and join of multiple video streams. See the trace in Figure 6 a. Upon transition of video traffic from 60 Mbps to 30 Mbps, the rates of s1-s3, s5-s7 and s9-s11 equally increase while the rates of s13 and s14 remain unchanged, as shown in Figures 6 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

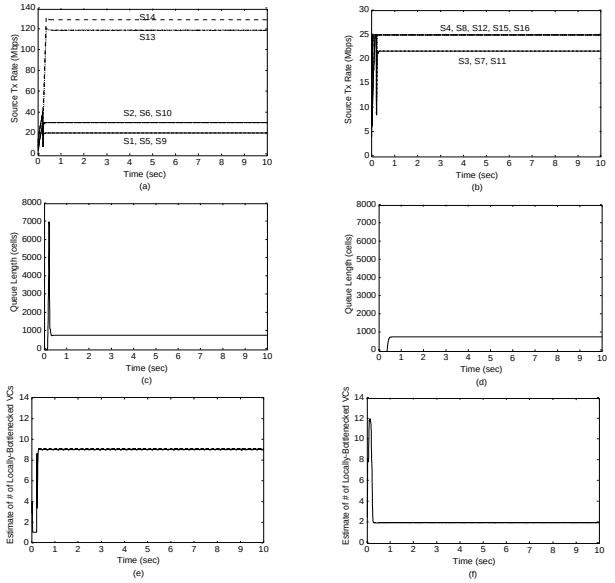

Fig. 5. Performance in the parking lot configuration (no VBR background traffic): (a) source transmission rate $a_i(t)$ of the VCs with PCR= 150 (Mbps), (b) source transmission rate $a_i(t)$ of the VCs with PCR= 25 (Mbps), (c) queue length at the SW 3, (d) queue length at the SW 4, (e) estimate of the 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.

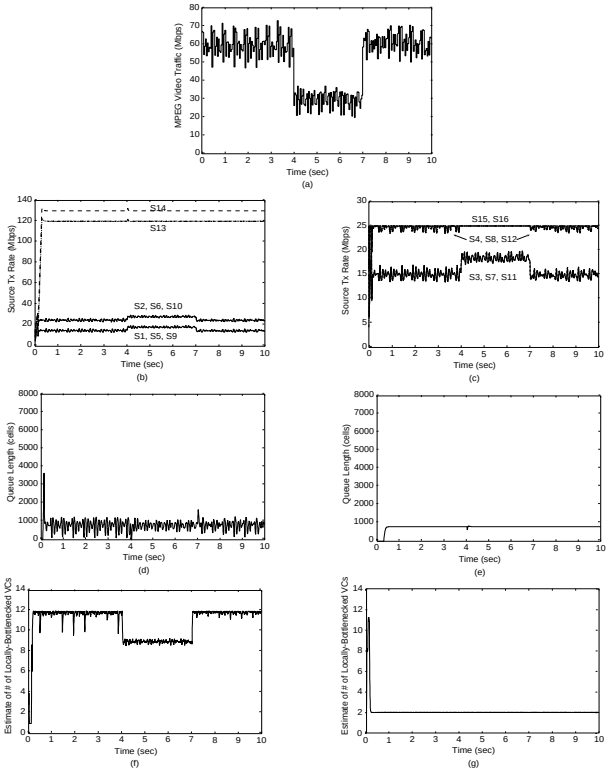

Fig. 6. Performance in the parking lot configuration (MPEG VBR background traffic): (a) trace of MPEG VBR video traffic loaded, (b) source transmission rate $a_i(t)$ of the VCs with PCR= 150 (Mbps), (c) source transmission rate $a_i(t)$ of the VCs with PCR= 25 (Mbps), (d) queue length at the SW 3, (e) 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.

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) sec and [7, ∞) sec, in Figure 6 c. This fluctuation implies that s4, s8 and s12 are no longer constantly PCR-constrained at 25 Mbps. 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 Figure 6 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 Figure 5 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 Figure 6 d).

References

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

[2] L. Benmohamed and S. M. Meerkov, Feedback Control of Congestion in Packet Switching Networks: The Case of Single Congested Node, *IEEE/ACM Trans. on Networking* **1**(6)(1993) 693-708.

[3] A. Kolarov and G. Ramamurthy, A Control-Theoretic Approach to the Design of an Explicit Rate Controller for ABR Service, *IEEE/ACM Trans. on Networking* **7**(5)(1999) 741-753.

[4] C. F. Su, G. de Veciana and J. Walrand, Explicit Rate Flow Control for ABR Services in ATM Networks, *IEEE/ACM Trans. on Networking* **8**(3)(2000) 350-361.

[5] M. K. Wong and F. Bonomi, A Novel Explicit Rate Congestion Control Algorithm, *Proc. IEEE GLOBECOM'98* **4** (1998) 2432-2439.

[6] Y. Choi, S. Kang and S. Chong, An Efficient ABR Service Engine for ATM Switch, preprint, 2000.

[7] L. Kalampoukas, A. Varma and K. K. Ramakrishnan, An Efficient Rate Allocation Algorithm for ATM Networks Providing Max-Min Fairness, Technical Report UCSC-CRL-95-29, Computer Engineering Dept., University of California, Santa Cruz, June 1995.

[8] 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. on Networking* **8** (1) (2000) 87-98.

[9] R. Bellman and K. L. Cooke, *Differential-Difference Equations* (Academic Press, New York, 1963).

[10] S. J. Bhatt and C. S. Hsu, Stability Criteria for Second-Order Dynamical Systems with Time Lag, *Journal of Applied Mechanics* (1966) 113-118.

[11] F. Brauer, Decay Rates for Solutions of a Class of Differential-Difference Equations, *SIAM J. Math. Anal.* **10** (4) (1979) 783-788.

[12] http://www-info3.informatik.uni-wuerzburg.de/rose.