is bounded uniformly. If the model basis is orthogonal, then the dimension of the model can be increased arbitrarily without affecting either the stability conditions i)–iv) or the state performance transient bound (the case $k = 0$), because $\underline{\lambda}(R_m) = 1$.

V. Summary and Further Research

We have given LQ-type performance bounds for a class of approximate-model–based adaptive designs. The results can be easily extended to multi-input systems with matched uncertainty and other simple variants on the control designs. Due to the dependance of the proofs largely only on the existence of quadratic Lyapunov functions, extensions are also possible to the case of unmatched uncertainty via backstepping designs; see [3].

References

[1] F.-C. Chen and C.-C. Liu, "Adaptively controlling nonlinear continuous time systems using multilayer neural networks," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 1306–1310, June 1994.

[2] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides*, 1st ed. Amsterdam, The Netherlands: Kluwer, 1988.

[3] M. French, "Adaptive control of functionally uncertain systems," Ph.D. dissertation, Univ. Southampton, Southampton, U.K., 1998.

[4] M. French and E. Rogers, "Approximate models for adaptive feedback linearization," *Int. J. Contr.*, vol. 68, pp. 1305–1321, 1997.

[5] M. Krstić, I. Kanellakopoulos, and P. V. Kokotović, *Nonlinear and Adaptive Control Design*, 1st ed. New York: Wiley, 1995.

[6] Z. H. Li and M. Krstić, "Optimal design of adaptive tracking controllers for nonlinear systems," *Automatica*, vol. 33, pp. 1459–1473, 1997.

[7] M. M. Polycarpou and P. A. Ioannou, "On the existence and uniqueness of solutions in adaptive control systems," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 474–479, Mar. 1993.

[8] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1987.

[9] R. M. Sanner and J. J. E. Slotine, "Gaussian networks for direct adaptive control," *IEEE Trans. Neural Networks*, vol. 3, pp. 837–863, June 1992.

[10] S. Sastry and A. Isidori, "Adaptive control of linearizable systems," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 1123–1131, Nov. 1989.

[11] E. Tzirkel-Hancock, "Stable control of nonlinear systems using neural networks," Ph.D. dissertation, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 1992.

[12] B. Yao and M. Tomizuka, "Adaptive robust control of SISO nonlinear systems in a semi-strict feedback form," *Automatica*, vol. 33, pp. 893–901, 1997.

Smith's Principle for Congestion Control in High-Speed Data Networks

Saverio Mascolo

***Abstract*—In high-speed communication networks, large propagation delays could have an adverse impact on the stability of feedback control algorithms. In this paper, classical control theory and Smith's principle are exploited to design an algorithm for controlling "best effort" traffic in high-speed asynchronous transfer mode (ATM) networks. The designed algorithm guarantees the stability of network queues, along with the fair and full utilization of network links, in a realistic traffic scenario in which multiple available bit rate (ABR) connections, with different propagation delays, share the network with high priority traffic.**

***Index Terms*—Classical control, congestion control, data networks, Smith's principle.**

I. Introduction

In recent years, intense research efforts have been devoted to the issue of transmitting multimedia traffic over a fully integrated universal network [1]. To this purpose, broadband integrated service digital networks (B-ISDN's) have been introduced and the emerging asynchronous transfer mode (ATM) technology has been retained as the transport technology to be used in B-ISDN's. ATM networks seek to provide the end-to-end transfer of fixed size cells and with specified quality of service (QoS). The fixed size of the cells reduces the variance of transmission delay, making the networks suitable for integrated traffic consisting of voice, video, and data [1]–[3].

ATM is a class of *virtual circuit* networks, which has been conceived to merge the advantages of circuit-switched technology (telephone networks), with those of packet-switched technology (computer networks) [1]. In particular, ATM networks are connection-oriented in the sense that, before two systems on the network can communicate, they should inform all intermediate switches about their service requirements and traffic parameters by establishing a *virtual circuit* (VC). This network is similar to the telephone networks, in which an exclusive circuit is set up from the calling party to the called party, with the important difference that, in the case of ATM, many VC's can share network resources via *store-and-forward packet switching and statistical multiplexing* [1]. The sharing of network resources allows communication costs be drastically reduced and requires sophisticated mechanisms of flow and congestion control to avoid congestion phenomena [1], [2]. Moreover, an increasing amount of research has been devoted to different control issues (see [1]–[6], [23], [25], and their references). These research efforts mostly deal with the issue of ensuring that users get their desired QoS.

The ATM Forum Traffic Management Group defines the following four service classes to support multimedia traffic.

1) The *constant bit rate* (CBR) class, which is conceived for applications such as telephone, video conferencing, and television.
2) The *variable bit rate* (VBR) class, which allows users to send at a variable rate. This category is subdivided into two categories: real-time VBR (RT–VBR), and nonreal-time VBR (NRT–VBR). An example of RT–VBR is interactive compressed video or industrial control (you would like a command sent to a robot arm

Manuscript received December 5, 1997; revised October 1, 1998 and March 26, 1999. Recommended by Associate Editor, L. Dai.

The author is with the Dipartimento di Elettrotecnica ed Elettronica, Politecnico di Bari, 70125 Bari, Italy (e-mail: mascolo@poliba.it).

Publisher Item Identifier S 0018-9286(00)00816-3.

to reach it before the arm crashes into something), and that of NRT–VBR is multimedia email.

3) The *unspecified bit rate* (UBR) class, which is designed for those data applications, such as email, file transfer, etc., that want to use any leftover capacity and are not sensitive to cell loss or delay; because this class does not require service guarantee, cell losses may result in retransmissions, which further increase congestion.
4) The *available bit rate* (ABR) class which is a "best effort" class designed for normal data traffic such as file transfer and email. This class does not require cell transfer delay to be guaranteed. The source, however, is required to control its rate by means of a feedback control mechanism that takes into account the congestion status of the network. In this way cell loss and retransmissions are minimized, and the network utilization is improved [1]–[3]. To operate the closed-loop control, the ATM Forum prescribes that an ABR source must send one control cell, called the resource management (RM) cell, every $\mathrm{NRM} = 32$ data cells. At the destination, RM cells of a flow carry the minimum input rate computed by all switches encountered along the VC path and are relayed back to the source conveying this minimum value.

Congestion control is critical in both ATM and non-ATM networks [23], and it is the most essential aspect of traffic management. Nowadays, the interest in a control-theory–based approach to this issue is ever increasing.

To briefly summarize the algorithms proposed for ABR traffic control, we start by recalling the binary feedback schemes that were first introduced because of their easy implementation [2], [7], [18], [19], [27]. In these schemes, if the queue level in a switch is greater than a threshold, then a binary digit is set in the control management cell. Because of the binary feedback information, problems of stability and performance develop. In [7] and [8], an analytic method for the design of a congestion controller, which ensures good dynamic performance along with fairness in bandwidth allocation, has been proposed. This algorithm, however, requires a complex online tuning of control parameters to ensure stability and to damp oscillations under different network conditions; moreover, the authors were unable to prove its global stability because of the complexity of the control strategy. In [9], a dual proportional-derivative (PD) controller has been suggested to make easier the implementation of the algorithm presented in [7]. In [26], the algorithm proposed in [7] has been implemented using per-VC first in first out (FIFO) queuing. In [10], Smith's principle has been exploited to design a control law in case a per-VC FIFO queuing is maintained at switches. In [11], two linear feedback control algorithms have been proposed for the case of a *single connection* with a *constant service rate*. In [12], these algorithms have been extended to the case of multiple connections with the same round trip delay sharing the bottleneck queue, and the robustness of these algorithms for the nonstationary service rate has been analyzed. In [13] and [14], proportional-integral (PI) type controllers have been described. In [15], a single-controlled traffic source, sharing a bottleneck node with other sources, is considered. The traffic is modeled by an ARMA process, whereas the H-infinity approach is used for designing the controller. In [20] and [28], the flow control problem is solved within the framework of decentralized linear quadratic (Gaussian) team theory. The (CBR + VBR) traffic is modeled by an autoregressive process. The algorithms proposed in [16] and [17] compute the input rates dividing the measured available bandwidth by the number of active connections. The measurement of the ABR bandwidth, however, which is bursty, is a hard task. In fact, the sampled control mechanism via RM cells makes it difficult to supply the sampling rate required by the Shannon theorem to reconstruct a bursty signal, such as the ABR bandwidth.

In this paper, Smith's principle is proposed as a key tool for designing a control law for ABR input rates that is effective over a connection path with any bandwidth-delay product. In particular, first, the dynamic behavior of each network queue in response to data input rates is modeled using transfer functions as the cascade of an integrator with a sum of time delays. Then, a controller is designed following the Smith principle. The proposed control algorithm ensures: 1) queues stability, i.e., no cell losses; 2) full and fair utilization of network links in presence of multiple connections, with different round-trip propagation delays, sharing the network; 3) exponential convergence of queue levels to stationary values without oscillations or overshoots; and 4) "efficient coexistence" of quality-constrained services (CBR +VBR) with "best effort" service (ABR).

In contrast to [11] and [12], where links with constant available bandwidth have been assumed, herein links with time-varying available bandwidths are considered in order to model the interaction of ABR with (CBR + VBR) traffic. Moreover, because it is difficult to measure the available bandwidth, this bandwidth is modeled as a disturbance input.

The paper is organized as follows. Section II describes the data network model. In Section III, the control law is designed and its performance is evaluated via mathematical analysis and computer simulations. Finally, Section IV draws the conclusions.

II. The Data Network Model

The network employs a store-and-forward service; that is, cells enter the network from the source edge nodes and are then stored and forwarded along a sequence of intermediate nodes and communication links, finally reaching their destination nodes [1], [2], [7], [25]. The network can be considered as a graph consisting of a set $N = \{1, \cdots, n\}$ of nodes (properly switches) connected by a set $L = \{1, \cdots, l\}$ of communication links. For each node $i \in N$, let $O(i) \subset L$ denote the set of its outgoing links and $I(i) \subset L$ denote the set of its incoming links. Each node maintains a queue for each outgoing link, where cells to be transmitted are temporarily stored. Each link i is characterized by its transmission capacity $c_i = 1/t_i$ (cells/s), where t_i is the transmission time of a cell, and by its propagation delay of t_{di} s. Each node has a processing capacity of $1/t_{\mathrm{pri}}$ cell/s, where t_{pri} is the time between the moment a cell is received by the node and the moment it is placed in the queue of its outgoing link. It is assumed that the processing capacity of each node is larger than the total transmission capacity of its incoming links so that congestion is caused by transmission capacity only. In high-speed communication networks, the *bandwidth* $\times$ *delay* product $c_i t_{di}$ is an important parameter that represents a large number of cells "in flight" on the transmission link. These cells are also called *in pipe cells*.

The network traffic is contributed by source/destination pairs $(S, D) \in N \times N$ forming the set of active connections C. The number of active connections n_c is the cardinality of C. To each (S, D) connection is associated a VC mapped on the path $p(S, D)$ [1], [7], [25]. The path i is specified by the sequence of links $e_i^i e_2^i \cdots e_l^i$ that the VC_i traverses as it goes through the network.

We assume a deterministic fluid model approximation of cell flow; that is, sources transmission rates are described by the continuous variable $u(t)$, measured in cells/s. An ABR source is expected to declare only its peak cell rate, that is, its maximum transmission speed $c_s = 1/t_s$. Moreover, we assume that ABR sources always have a cell to send; that is, they are *persistent* sources [2].

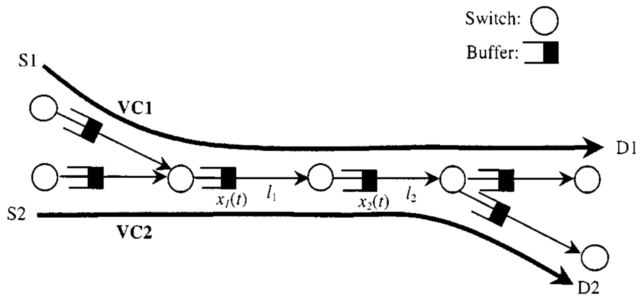

Fig. 1. Scheme of two *virtual circuits* (S_1, D_1) and (S_2, D_2) sharing links l_1 and l_2.

We assume that each output link maintains an FIFO queue that is shared by all VC's flowing through the link. Fig. 1 shows two VC connections sharing links l_1 and l_2. Let $x_j(t)$ be the queue level associated with the link l_j. By writing the flow conservation equations, the queue level $x_j(t)$, starting at $t = 0$ with $x_j(0) = 0$, is the time *integral* of the input rates minus the output rate

$$x_j(t) = \int_0^t \left(-d_j(\tau) + \sum_{i=1}^{n_j} u_{ij}(\tau - T_{ij}) \right) \cdot d\tau \qquad (1)$$

where

- n_j cardinality of the set C_j of VC's sharing the queue associated with link l_j;
- $u_{ij}(t) \geq 0$ inflow rate caused by the ith VC;
- T_{ij} propagation delay from the ith source to the jth queue;
- $d_j(t) \geq 0$ rate of packets leaving the jth queue.

Letting $b_{av,j}(t) \geq 0$ be the ABR bandwidth at link l_j, the relationship with the depletion rate is $d_j(t) = b_{av,j}(t) \cdot h(x_j)$, where $h(x_j)$ is shown at the bottom of the page.

Thus, it results $x_j(t) \geq 0$. Notice that (1) is linear because it neglects the saturation effect because the buffer capacity is finite. Therefore, it holds only if the queue level never exceeds the buffer capacity. As will be shown in Section III, this condition will be always satisfied by the controlled queue dynamics.

III. The Control Law

In this section, we first develop a general model of the controlled ABR flows and then we design the congestion control law. To design the controller and analyze the performance of the controlled system, we use the standard Laplace transform technique; that is, transfer functions are used to model and analyze the input–output dynamics of the controlled system.

A. Model of the Flow-Controlled Data Network

We assume that each switch-node i has a congestion controller associated with each outgoing link. This controller computes a unique admissible transmission rate for all VC's sharing the same outgoing link. Thus, the controlled system results to be a single-input–single-output (SISO) linear system with a disturbance. More precisely, Fig. 2 shows the block diagram of the controlled system consisting of the following functions.

1) Bottleneck FIFO queue $x_j(t)$, which is modeled, in the Laplace domain, by the integrator $1/s$.
2) Link service rate $d_j(t)$, which is modeled as an *unmeasured disturbance*; in fact, the ABR bandwidth is left over by the high priority (VBR + CBR) traffic loading the link, and it may change drastically. As a consequence, the measurement of ABR bandwidth is a hard task.
3) Round-trip delay T_i of each ABR connection i, which shares the bottleneck queue $(i = 1, n)$.
4) Controller transfer function $G_j(s)$, which computes the admissible input rate $u_j(t) = u_{ij}(t)$ for any i such that $\mathrm{VC}_i \in C_j$.
5) Set point $r_j(t)$, which sets a threshold for the queue level.

The feedback control scheme works as follows: the controller compares the reference signal $r_j(t)$ with the bottleneck queue level $x_j(t)$ and then inputs the difference into the controller $G_j(s)$, which computes the rate $u_j(t)$. The input rate $u_j(t)$ of the connection VC_i reaches the bottleneck queue as an effective inflow rate after the round-trip time T_i. It is assumed that propagation delays are dominant compared with other delays (processing, queuing, etc.).[1] As a consequence, round-trip delays are assumed to be constant and measured when a new VC is established.

B. Design of the Control Law Using the Smith Principle

The objective of the control law is to guarantee that the input rates of the sources promptly utilize all "best effort" bandwidth. At the same time, buffer overflow must be avoided. This goal can be formally stated with the following conditions.

1) *Stability:*

$$x_j(t) \leq r^o \qquad \text{for } t \geq 0 \text{ and for any } j \text{ so that } l_j \in L \qquad (2)$$

where r^0 is the queue capacity. This condition guarantees that network queues are bounded, i.e., no cell loss.

2) *Full Link Utilization:*

$$x_j(t) > 0 \qquad \text{for } t > T_{tr} \text{ and for any } j \text{ so that } l_j \in L \qquad (3)$$

which guarantees full utilization of network links, i.e., $b_{\mathrm{av},j}(t) = d_j(t)$, because a link has always data to send.

The time T_{tr} in (3) takes into account the round-trip propagation time inside the control loop and the transient time of the dynamics.

Now, letting the set point $r_j(t)$ be the bottleneck queue capacity r^o, the control variable $u_j(t)$ seeks to fill in the bottleneck queue at its full capacity, whereas the disturbance $d_j(t)$ seeks to empty the queue. Because of the possible large propagation delays inside the feedback loop, queue level dynamics might exhibit oscillations, and even become unstable. Because we have seen that the model of the communication system is known, a controller can be successfully designed following the Smith principle [21], [22].

Smith's principle is well known as an effective dead-time compensator for a stable process with large time delay. The main advantage of this technique is that the time delay is eliminated from the characteristic equation of the closed-loop system. Thus, the design problem for the process with delay can be transformed to the one without delay.

[1]ATM switches can give priority to control cells over data cells. Thus, the queuing time of control cells is zero and the round-trip time is constant and equal to the propagation time.

$$h(x_j) = \begin{cases} 1, & \text{if } x_j(t) > 0 \\ \alpha(t)/b_{av,j}(t), & \text{if } x_j(t) = 0, \text{ where } \alpha(t) = \min\left(\sum_{i=1}^{n_j} u_{ij}(t - T_{ij}), b_{av,j}(t) \right) \end{cases}$$

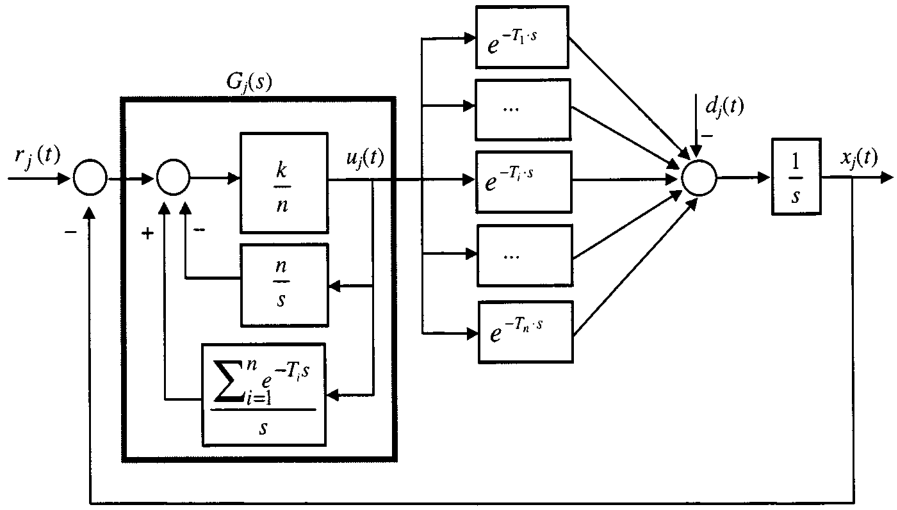

Fig. 2. Block diagram of n controlled VC's sharing a *first in first out* bottleneck queue.

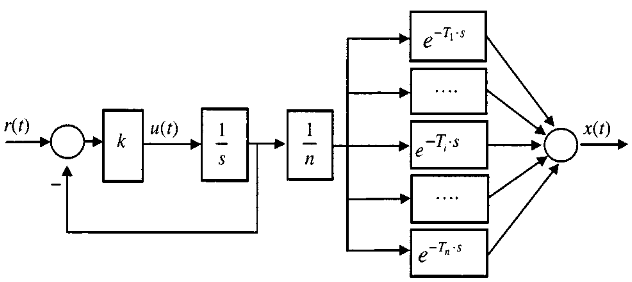

Fig. 3. Block diagram of the desired input–output dynamics.

Herein, following the Smith principle, we look for a controller $G(s)$[2] so that the input–output dynamics of the system reported in Fig. 2 becomes equal to the input–output dynamics of the system shown in Fig. 3. The target system in Fig. 3 has been carefully chosen to satisfy all of the following points.

1) The closed loop part of the target system is delay free; that is, delays are pushed out of the feedback loop (Smith's principle).
2) Delays are in cascade connection with the delay free system, which is a simple first-order dynamic system described by the transfer function $k/(k+s)$. This system is asymptotically stable for any $k > 0$.
3) A controller $G(s)$ exists that renders the input–output dynamics of the system reported in Fig. 2 equal to the input–output dynamics of the desired system reported in Fig. 3.

Notice that a nice feature of the chosen target system is that, letting the set point $r(t)$ be the step function $r^o \cdot 1(t)$, the output exponentially converges to the steady value r^o without oscillations or overshoots. In fact, the output is the sum of several delayed responses of a first-order system to a step function.

Proposition 1: The transfer functions $X(s)/R(s)$ of the systems reported in Figs. 2 and 3 can be made equal by using the controller described by the transfer function

$$G(s) = \frac{k/n}{1 + \dfrac{k/n}{s}\left(n - \displaystyle\sum_{i=1}^{n} e^{-T_i s}\right)}. \tag{4}$$

[2]From now on, the subscript j in $x_j(t)$, $d_j(t)$, $G_j(t)$, etc., which is used to refer the generic output link j, is dropped.

Proof: By equating the transfer functions of the systems reported in Figs. 2 and 3

$$\frac{\dfrac{G(s)}{s}\displaystyle\sum_{i=1}^{n} e^{-T_i s}}{1 + \dfrac{G(s)}{s}\displaystyle\sum_{i=1}^{n} e^{-T_i s}} = \frac{\dfrac{k}{s}}{1 + \dfrac{k}{s}} \frac{1}{n}\sum_{i=1}^{n} e^{-T_i s}$$

the controller (4) is derived after a little algebra. □

By looking at the controller (4) shown in Fig. 2, it is easy to write the rate control equation that is

$$\begin{aligned} u(t) &= \frac{k}{n}\left(r^o - x(t) - n\int_0^t u(\tau)\cdot d\tau + \sum_{i=1}^{n}\int_0^{t-T_i} u(\tau)\cdot d\tau\right) \\ &= \frac{k}{n}\left(r^o - x(t) - \sum_{i=1}^{n}\int_{t-T_i}^{t} u(\tau)\cdot d\tau\right). \end{aligned} \tag{5}$$

This equation can be intuitively interpreted as follows: the computed input rate is proportional, through the coefficient k/n, to the available room in the bottleneck queue $r^o - x(t)$ decreased by the number of cells released by each VC i during the last corresponding round-trip time T_i. Notice that (5) requires that the switch keeps track of the number of active sources.

C. The Reference Signal and the Disturbance

The controlled system reported in Fig. 2 is an SISO system with a disturbance. The system input $r(t)$ sets a threshold r^o for the bottleneck queue length. The disturbance $d(t)$ models the bottleneck link service rate. We normalize to unity the link transmission capacity, so that, if all link bandwidth is suddenly available for ABR connections at $t = t_o$, then the available bandwidth $b_{av}(t)$ is equal to the step function $1(t - t_o)$. Letting $b(t) < 1$ be the bandwidth consumed by coexisting (VBR + CBR) traffic, it results

$$0 \le b_{av}(t) \le 1(t) - b_m \cdot 1(t) = a \cdot 1(t)$$

where $b_m = \min_t\{b(t)\}$, $0 \le b_m \le 1$, and $a = (1 - b_m)$, with $0 \le a \le 1$.

In other words, the ABR bandwidth is bounded by the function $a \cdot 1(t)$, with $a \in [0, 1]$. This function represents a *worst-case disturbance* that models a bandwidth that is suddenly available at $t = 0$.

In the next section, we will show that the proposed control law ensures queue stability and full link utilization even in the presence of the worst-case disturbance $a \cdot 1(t)$.

D. Performance Evaluation of the Control Law Via Mathematical Analysis

Classical control theory provides an established set of tools that enables us to design algorithms whose performance can be predicted analytically rather than relying on simulations. To analyze the performance of the proposed algorithm, it is sufficient to use Laplace transform technique. The important advantage of mathematical analysis is that it allows us to demonstrate the properties of the proposed control law in a general setting, whereas the validation via computer simulations is inevitably restricted to the simulated scenarios.

Proposition 2: Considering the reference signal $r(t) = r^o \cdot 1(t)$, the worst-case disturbance $d(t) = a \cdot 1(t)$, with $0 \le a \le 1$, and the controller (4), the output queue level satisfies the *stability condition* $x(t) \le r^o$ for any $t \ge 0$.

Proof: From Proposition 1, the input–output dynamics of the systems reported in Figs. 2 and 3 are equal. Therefore, assuming the initial condition $x(0) = 0$, we can consider the Laplace transform of the output of the system in Fig. 3 in response to the set point $r^o \cdot 1(t)$; that is

$$X_r(s) = \frac{r^o}{s}\frac{1}{(1+s/k)} \cdot \frac{1}{n}\sum_{i=1}^{n} e^{-T_i s}$$

and, by antitransforming back to time domain, it results in

$$x_r(t) = \frac{r^o}{n}\sum_{i=1}^{n}\left(1 - e^{-k(t-T_i)}\right)\cdot 1(t-T_i).$$

Because $\sum_{i=1}^{n}(1 - e^{-k(t-T_i)}) \cdot 1(t-T_i) \le n$, it turns out that

$$x_r(t) \le r^o, \qquad \text{for } t \ge 0.$$

Moreover, the transfer function from $d(t)$ to the queue level $x_d(t)$ of the system reported in Fig. 2 is given by

$$\frac{X_d(s)}{D(s)} = -\frac{1}{s} + \frac{k}{n}\cdot\frac{1}{s(s+k)}\sum_{i=1}^{n} e^{-T_i s}.$$

Assuming the worst-case disturbance $D(s) = a/s$ and by antitransforming $X_d(s)$ back to time domain, it results in

$$x_d(t) = a\left(-t\cdot 1(t) + \frac{1}{n}\sum_{i=1}^{n}\left((t-T_i)\cdot 1(t-T_i) - \frac{1}{k}\left(1 - e^{-k(t-T_i)}\right)\cdot 1(t-T_i)\right)\right).$$

Because

$$\sum_{i=1}^{n}\left((t-T_i)\cdot 1(t-T_i) - \frac{1}{k}\left(1 - e^{-k(t-T_i)}\right)\cdot 1(t-T_i)\right) \le nt\cdot 1(t)$$

it follows that

$$x_d(t) \le 0, \qquad \text{for } t \ge 0.$$

Thus, it can be concluded that

$$x(t) = x_r(t) + x_d(t) \le x_r(t) \le r^o, \qquad \text{for } t \ge 0;$$

that is, the queue is upper bounded by r^o for any worst case disturbance $a \cdot 1(t)$. □

Remark 1: The queue dynamics $x(t)$ is characterized by the time constant $\tau = 1/k$. Therefore, the transient can be considered exhausted after the time $T_{tr} = \max_i(T_i) + 4\tau$.

The following proposition states a condition on the buffer capacity that guarantees the full utilization of link bandwidth even in the presence of bandwidth that is suddenly available.

Proposition 3: Considering the worst case disturbance $d(t) = a \cdot 1(t)$, with $a \in [0, 1]$, the controller (4) guarantees the full utilization of network links, that is $x(t) > 0$ for $t > T_{tr}$, if the capacity r^o of the network queues satisfies the following condition:

$$r^o > a\left(\tau + \frac{1}{n}\sum_{i=1}^{n} T_i\right). \tag{6}$$

Proof: Using the controller (4), the queue dynamics of the system in Fig. 2 in response to $r(t)$ and $d(t)$ is

$$\begin{aligned} x(t) &= x_r(t) + x_d(t) \\ &= \frac{r^o}{n}\sum_{i=1}^{n}\left(1 - e^{-k(t-T_i)}\right)\cdot 1(t-T_i) \\ &\quad + a\left(-t\cdot 1(t) + \frac{1}{n}\sum_{i=1}^{n}\left((t-T_i)\cdot 1(t-T_i) - \frac{1}{k}\left(1 - e^{-k(t-T_i)}\right)\cdot 1(t-T_i)\right)\right). \end{aligned}$$

For $t > \max_i(T_i) + 4\tau = T_{tr}$, the exponential terms can be considered exhausted, and $x(t)$ reaches the steady-state value

$$x(t) = x_s = r^o - \frac{a}{n}\sum_{i=1}^{n} T_i - \frac{a}{k}. \tag{7}$$

Thus, Proposition 3 is derived by requiring that $x_s > 0$ for $t > T_{tr}$. □

Remark 2: Proposition 3 guarantees full utilization of network links if each queue capacity is at least equal to the number of "in flight" cells contained in "a pipe" with propagation delay $\tau + (\sum_{i=1}^{n} T_i)/n$.

Finally, the following proposition shows that the "best effort" bandwidth is equally shared by the VC's.

Proposition 4: In the steady-state condition, the VC's capture all available ABR bandwidth in a fair way.

Proof: Considering (5), the steady-state values $x_s = x(t \to \infty)$ and $u_s = u(t \to \infty)$ satisfy the relation

$$u_s = \frac{r^o - x_s}{n\tau + \sum_{i=1}^{n} T_i}. \tag{8}$$

Substituting in (8) the value of x_s obtainable from (7), it turns out

$$u_s = \frac{a}{n}$$

which means that the ABR bandwidth a is equally shared by the n VC's. □

Remark 3: A continuous-time model of the flow-controlled data network dynamics has been derived. In ATM networks, however, each ABR traffic source interleaves control cells (RM cells) with data cells in a periodic way. Each congestion controller, associated with each outgoing link, computes a unique, admissible transmission rate for all VC's sharing the same outgoing link and stamps this computed value on the RM cells only if it is less than the rate already stored. At the destination, the RM cell carries the minimum input rate over all switches belonging to the same VC and it comes back to the source conveying the minimum allowed rate. Upon receiving this minimum rate, the source sets the input rate to this value [3], which means that feedback information is relayed in RM cells, and thus not available in continuous time, but rather in sampled form. Therefore, the control equation (5) should be transformed into discrete time form; that is, the controller should update the input rate every T_s units of time, in which T_s is the sampling time. For the sake of brevity, this discrete time analysis is not reported [10], [14], [24].

E. Computer Simulation Results

In this section, computer simulations are carried out to confirm the validity of the proposed algorithm. Four ABR VC's, sharing an FIFO bottleneck queue with (VBR + CBR) traffic, are considered (Fig. 4).

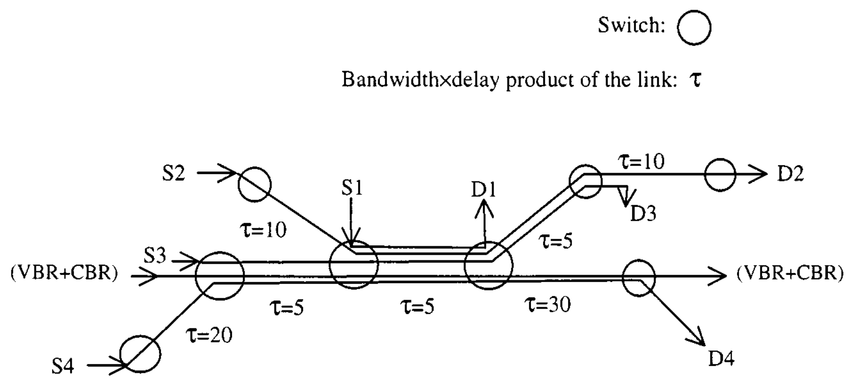

Fig. 4. Network topology and traffic scenario.

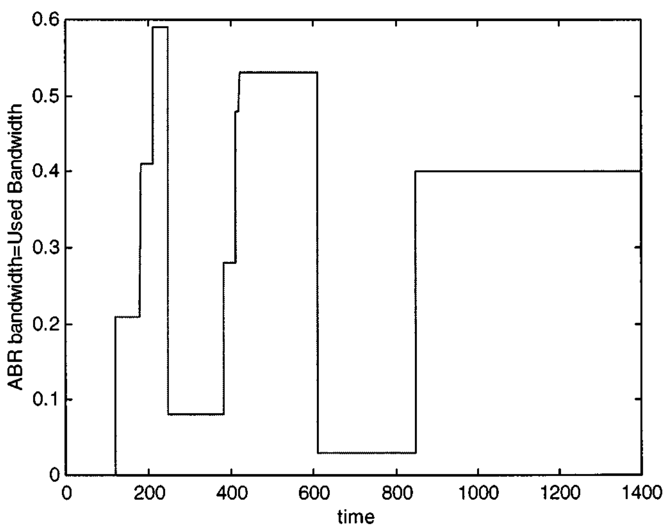

Fig. 5. ABR bandwidth $b_{\rm av}(t)$ and used bandwidth $d(t)$.

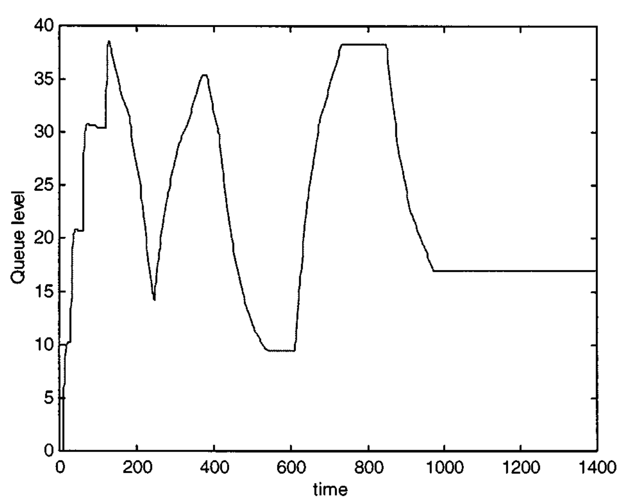

Fig. 6. Bottleneck queue level dynamics.

The connections are characterized by a bandwidth-delay product of ten, 30, 60, and 120 cells, respectively. Note that a bandwidth-delay product of ten is typical of a *local area network* (LAN), and one of 120 cells is typical of a metro or regional *wide area network* (WAN). Link bandwidth is normalized to unity. The interaction with quality constrained traffic (CBR + VBR) is taken into account by means of the disturbance function $d(t)$, which is shown in Fig. 5. This function is equal to the ABR bandwidth $b_{av}(t)$; that is, all "best effort" bandwidth is used by ABR traffic. A constant gain $k = 0.1/\text{s}$ and a buffer capacity $r^o = 40$ cells, which strictly satisfies the (6), are assumed. Fig. 6 shows that the bottleneck queue dynamics is upper bounded by r_o and lower bounded by zero; that is, no cell loss and full link utilization are guaranteed.

IV. Conclusion

Classical control theory and Smith's principle have been proposed as key tools for designing an effective congestion control algorithm for the class of ABR traffic in ATM networks. The suggested algorithm works in a realistic traffic scenario consisting of multiple ABR connections that share available bandwidth with VBR and CBR traffic. Mathematical analysis and simulation results show the validity of the algorithm. In particular, it has the following advantages:

1) it is a relatively simple algorithm;
2) it relaxes the too unrealistic hypothesis made in [11] and [12], in which connections with the same round-trip delay were assumed;
3) it relaxes the assumption made in [11] and [12] of links with constant service rate, because with this assumption it is not possible to consider a realistic traffic scenario with (CBR + VBR) traffic;
4) it allows us to analyze transient and steady-state behavior via mathematical analysis;
5) it ensures fast exponential convergence of queue levels to stationary values without oscillations or overshoots;
6) it allows us to prove the global stability of the network queues along with the full and fair utilization of network links;
7) it does not require the tuning of any control parameter to react to the changing traffic condition;
8) in contrast to [16] and [17], it does not require the measurement of the available bandwidth that is here modeled as a disturbance input.

Further research should focus on the issue of guaranteeing the required QoS in real-time communications.

Acknowledgment

The author thanks the anonymous reviewers for their constructive comments, which helped to improve the quality of the paper.

References

[1] P. Varaiya and J. Walrand, *High-Performance Communication Networks*. San Francisco, CA: Morgan Kaufmann, 1996.
[2] R. Jain, "Congestion control and traffic management in ATM networks: Recent advances and a survey," *Comput. Networks ISDN Syst.*, vol. 28, no. 13, pp. 1723–1738, 1996.

[3] (1996, Apr.) ATM Forum Traffic Management Specification Version 4.0. ATM Forum Technical Committee TMWG. [Online]. Available: http://www.atmform.com

[4] N. Akar and K. Sohraby, "System theoretic approach to teletraffic problems: A unifying framework," in *Proc. IEEE Globecom'96*, London, U.K., Nov. 1996.

[5] W. Ding, "Joint encoder and channel rate control of VBR video over ATM networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 266–278, Apr. 1997.

[6] J. Le Boudec, G. de Veciana, and L. Walrand, "QoS in ATM: Theory and practice," in *Proc. 35th Conf. Decision Contr.*, Kobe, Japan, Dec. 1996.

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

[8] ——, "Feedback control of congestion in packet switching networks: The case of multiple congested node," in *Proc. Am. Contr. Conf.*, Baltimore, MD, June 1994.

[9] A. Kolarov and G. Ramamurthy, "A control theoretic approach to the design of closed loop rate based flow control for high speed ATM networks," in *Proc. IEEE Infocom'97*, Kobe, Japan, Apr. 1997.

[10] S. Mascolo, D. Cavendish, and M. Gerla, "ATM rate based congestion control using a Smith predictor: An EPRCA implementation," in *Proc. IEEE Infocm'96*, San Francisco, CA, Mar. 1996.

[11] R. Izmailov, "Adaptive feedback control algorithms for large data transfer in high-speed networks," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1469–1471, Aug. 1995.

[12] ——, "Analysis and optimization of feedback control algorithms for data transfer in high-speed networks," *SIAM J. Contr. Optim.*, vol. 34, no. 5, pp. 1767–1780, 1996.

[13] C. E. Rohrs and R. A. Berry, "A linear control approach to explicit rate feedback in ATM networks," in *Proc. IEEE Infocom'97*, Kobe, Japan, Apr. 1997.

[14] P. Johansson and A. A. Nilsson, "Discrete time stability of an explicit rate algorithm for the ABR service," in *IEEE ATM'97 Workshop*, Lisboa, Portugal, May 1997.

[15] Z. Pan, E. Altman, and T. Basar, "Robust adaptive flow control in high speed telecommunication networks," in *Proc. 35th Conf. Decision Contr.*, Kobe, Japan, Dec. 1996.

[16] A. Charny, D. D. Clark, and R. Jain, "Congestion control with explicit rate indication," in *Proc. ICC'95*, Seattle, WA, June 1995.

[17] R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, and R. Viswanathan. (1995, Feb.) The ERICA Switch Algorithm: A Complete Description. ATM Forum Tech. Committee TMWG. [Online]. Available: http://www.atmform.com

[18] K. W. Fendick, M. A. Rodrigues, and A. Weiss, "Analysis of a rate-based feedback control strategy for long haul data transport," *Performance Eval.*, vol. 16, pp. 67–84, 1992.

[19] F. Bonomi, D. Mitra, and J. B. Seery, "Adaptive algorithms for feedback-based flow control in high-speed, wide-area ATM networks," *IEEE J. Selected Areas Commun.*, vol. 13, no. 7, pp. 1267–1283, 1995.

[20] E. Altman, T. Basar, and R. Srikant, "Multiuser rate-based flow control with action delays: A team-theoretic approach," in *Proc. 36th IEEE Conf. Decision Contr.*, San Diego, CA, Dec. 1997.

[21] O. Smith, "A controller to overcome dead time," *ISA J.*, vol. 6, no. 2, pp. 28–33, 1959.

[22] J. E. Marshall, *Control of Time-Delay Systems*. New York: Peter Peregrinus, 1979.

[23] V. Jacobson, "Congestion avoidance and control," in *Proc. SIGCOMM'88 Symp.*, Stanford, CA, 1998, pp. 314–329.

[24] E. Altman, F. Baccelli, and J. Bolot, "Discrete-time analysis of adaptive rate control mechanism," in *High Speed Networks and Their Performance*, H. G. Perros and Y. Viniotis, Eds. Amsterdam, The Netherlands: North-Holland, 1994, pp. 121–140.

[25] L. L. Peterson and B. S. Davie, *Computer Networks*. San Francisco, CA: Morgan Kaufmann, 1996.

[26] L. Benmohamed and Y. T. Wang, "A control-theoretic ABR explicit rate algorithm for ATM switches with per-VC queuing," in *Proc. IEEE Infocom*, San Francisco, CA, Mar. 1998.

[27] N. Yin and M. G. Hluchyj, "On closed-loop rate control for ATM cell relay networks," in *Proc. Infocom'94*, 1994.

[28] E. Altman, T. Basar, and R. Srikant, "Robust rate control for ABR sources," in *Proc. IEEE Infocom*, San Francisco, CA, Mar. 1998.

Frequency Domain System Identification with Missing Data

R. Pintelon and J. Schoukens

***Abstract*—This paper presents a frequency domain solution to the system identification problem with missing data in the input and output signals. No particular pattern for the missing data is assumed. The approach does not require a (parametric) model for the input signal, works for any model structure (ARX, ARMAX, OE, errors-in-variables, etc.), and can be applied to discrete time as well as continuous time models. The key idea is to treat the missing data as unknown parameters.**

***Index Terms*—Frequency domain, identification, missing data.**

I. Introduction

Because of temporary sensor failure or data transmission errors, data samples may be missing in the measured signals. The best thing to do then is to throw away the data set and to repeat the experiment, which is not always possible because, for example, the experiment is too expensive, or some of the data are collected in an irregular way using laboratory analysis. Sometimes the output is sampled at a lower rate than the input, which is a periodic missing output data problem [1], [2]. If not properly taken into account, the missing measurements can seriously deteriorate the quality (consistency, efficiency) of the estimates. Although the classical time domain approach assumes that the input is known [3], it is often measured. In that case, missing data can equally well occur at the input as at the output.

The problem of missing output data is well understood and has been studied extensively in time series analysis [4] and system identification (see [1], [2], and [5], and the references there in). Most solutions combine classical identification methods, for example, (recursive) least-squares algorithms in [1] and [2] and maximum likelihood in [5], with missing data predictors. All of these methods, except the one presented in [5], cannot handle the case of missing input data. In [5], this problem has been solved in the time domain for ARX models. The approach, however, assumes that the input signal can be modeled as an autoregressive (AR) process.

This paper presents an original frequency domain solution to the missing input or output data problem. The main advantages of the method are as follows:

- it requires no (parametric) model for the input signal;
- it is valid for any model structure (ARX, ARMAX, OE, errors-in-variables, etc.);
- it can be applied to discrete time as well as continuous time models.

The basic idea behind the solution is to treat the missing input/output data as unknown parameters in the identification problem.

Manuscript received February 26, 1998; revised August 18, 1998 and January 15, 1999. Recommended by Associate Editor, G. Bastin. This work was supported by the Fund for Scientific Research (FWO-Vlaanderen), the Flemish Government (GOA-IMMI), and the Belgian Program on the Interuniversity Poles of Attraction from the Belgian State Prime Minister's Office, Science Policy Programming (IUAP 4/2).

The authors are with the Department of Electrical Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium (e-mail: rik.pintelon@vub.ac.be).

Publisher Item Identifier S 0018-9286(00)00818-7.