ABR TRAFFIC CONTROL OVER MULTI-SOURCE SINGLE-BOTTLENECK ATM NETWORKS

ANDRZEJ BARTOSZEWICZ*, TOMASZ MOLIK*

* Institute of Automatic Control, Technical University of Łódź
ul. Stefanowskiego 18/22, 90-924 Łódź, Poland
e-mail: andpbart@mail.p.lodz.pl

The problem of flow control in fast, connection-oriented communication networks supporting the traffic generated by multiple sources is considered. A novel sampled time strategy governing the behaviour of the sources is proposed. The strategy combines the Smith principle with the conventional sampled time proportional controller. It guarantees an equal resource allocation between various users, full bottleneck link utilisation and no cell loss in the controlled network. Consequently, the need for cell retransmission is eliminated and a high throughput is ensured. Furthermore, transmission rates generated by the algorithm are limited. This property permits a direct implementation of the proposed strategy in the network environment. A simulation example confirms favourable performances of the proposed control scheme.

Keywords: congestion control, ATM networks, time delay systems, Smith principle

1. Introduction

The asynchronous transfer mode (ATM) technology plays a crucial role in the design and implementation of broadband integrated services digital networks (B-ISDN). This technology is well suited for video, voice and data transmission through high speed telecommunication networks. The ATM networks are connection oriented, i.e. a virtual circuit (VC) is established between each source and destination for the connection lifetime. After setting up a VC, data are sent in relatively short, fixed size packets, usually called cells. Each data cell is 53 bytes long and consists of 48 bytes of transmitted information and a 5 byte long header. The small fixed cell size reduces delay variation, which could be particularly harmful for multimedia traffic.

In order to properly serve diverse needs of different users, the ATM Forum defines five service categories:

1. *The constant bit rate (CBR) service category* provides the bandwidth which is always available to its user. This category is used by a real time service. Television is the typical example of the service category.

2. *The variable bit rate (VBR)* is designed for both real and non-real time applications. An example of such a real time application is video conferencing, and multimedia e-mail is an example of a non-real time VBR service.

3. *The available bit rate (ABR)* is a service category whose rate depends on the available bandwidth. Users should adjust their flow rates according to the feedback information received from the network. Electronic mail is an example of this service category.

4. *The unspecified bit rate (UBR) category* is used to send data on the first in first out (FIFO) basis, using the capacity which has not been consumed by other services. No initial commitment is made to a UBR source and no feedback concerning congestion is provided. This type of service can be used for background file transfer.

5. *The guaranteed frame rate (GFR)* is a service intended for non-real time applications with minimum rate requirements. No feedback control protocol is applied in this service category. An example of this service is frame relay interworking.

As stated above, the ABR is the only service category using feedback information to control the source flow rate. Therefore, ABR control is particularly important for congestion avoidance and full resource utilisation. This problem, i.e. the ABR flow control in ATM networks, is considered in the paper.

A difficulty in the ABR flow control is mainly caused by long propagation delays in the network. If congestion occurs at a specific node, information about this condition must be conveyed to all the sources transmitting data cells through the node. Transferring this information involves feedback propagation delays. After this information has been received by a particular source, it can be used to adjust the flow rate of this source. However, the adjusted flow rate will start to affect the congested node only after a forward propagation delay.

ABR flow rate control has recently been studied in several papers (Chong *et al.*, 1998; Gómez-Stern *et al.*, 2002; Imer *et al.*, 2001; Izmailov, 1995; Jagannathan and Talluri, 2002; Jain, 1996; Laberteaux *et al.*, 2002; Lengliz and Kamoun, 2000; Mascolo, 1997; 1999; 2000; Quet, 2002). A valuable survey of earlier congestion control mechanisms is given in (Jain, 1996). Furthermore, Izmailov (1995) considered a single connection controlled by a linear regulator whose output signal is generated according to several states of the buffer measured at different time instants. The asymptotic stability, the nonoscillatory system behaviour and a locally optimal rate of convergence have been proved. Chong *et al*. (1998) proposed and thoroughly studied the performance of a simple queue length based flow control algorithm with dynamic queue threshold adjustment. Lengliz and Kamoun (2000) introduced a proportional plus derivative controller which is computationally efficient and can be easily implemented in ATM networks. Imer *et al.* (2001) gave a brief, excellent tutorial exposition of the ABR control problem and presented new stochastic and deterministic control algorithms. Another interesting approach to the problem of flow rate control in communication networks was proposed by Quet *et al.* In the recent paper (Quet, 2002) the authors considered a single bottleneck multi-source ATM network and applied minimisation of an H_∞ norm to the design of a flow rate controller. The proposed controller guarantees stability robustness to uncertain and time-varying propagation delays in various channels. Adaptive control strategies for ABR flow regulation were proposed by Laberteaux and Rohrs (2002). Their strategies reduce the convergence time and improve the queue length management. Also a neural network controller for the ABR service in ATM networks has recently been proposed. Jagannathan and Talluri (2002) showed that their neural network controller can guarantee the stability of the closed loop system and the desired quality of the service (QoS).

Due to the significant propagation delays which are critical for the closed loop performance, several researchers applied the Smith principle to control the ABR flow in communication networks (Gómez-Stern *et al.*, 2002; Mascolo, 1997; 1999; 2000). Mascolo (1999) considered the single connection congestion control problem in a general packet switching network. He used the deterministic fluid model approximation of the packet flow and applied transfer functions to describe the network dynamics. The designed continuous time controller was applied to the ABR traffic control in an ATM network and compared with the ERICA standard. Furthermore, Mascolo showed that the Transmission Control Protocol/Internet Protocol (TCP/IP) implements a Smith predictor to control network congestion. In the next paper (Mascolo, 2000) the same author applied the Smith principle to the network supporting multiple ABR connections with different propagation delays. The proposed control algorithm guarantees no cell loss, full and equal network utilisation, and ensures exponential convergence of queue levels to stationary values without oscillations or overshoots. Gomez-Stern *et al.* further studied the ABR flow control using the Smith principle (Gómez-Stern *et al.*, 2002). They proposed a continuous time proportional-integral (PI) controller which helps to reduce the average queue level and its sensitivity to the available bandwidth. Saturation issues in the system were handled using anti-wind up techniques.

In this paper we consider the ABR flow control in ATM networks. Our approach is similar to that introduced in (Gómez-Stern *et al.*, 2002; Mascolo, 1997; 1999; 2000); however, in contrast to those papers we propose a sampled time strategy. The strategy combines the Smith principle with the conventional proportional controller. It guarantees equal resource allocation between various users, full bottleneck node link utilisation and no cell loss in the network. As a result, the need for cell retransmission is eliminated and the maximum throughput is achieved. Furthermore, transmission rates generated by the algorithm are non-negative and limited. These properties permit a direct implementation of the proposed strategy in the network environment.

The remainder of this paper is organised as follows: The model of the network used throughout the paper is introduced in Section 2. Then the main result, i.e. the new sampled time control algorithm, is described in Section 3. Section 4 presents a simulation example, and Section 5 comprises conclusions.

2. Network Model

In this paper the ABR flow control in an ATM network supporting multiple sources is considered. The network consists of data sources, nodes and destinations—all of them interconnected via bi-directional links. Each node of the network (i.e. a switch) maintains one queue per output port. When a new data cell arrives at an input port of the node, it is directed to the appropriate output buffer, stored and forwarded to the next node on the first in first out (FIFO) basis. In much the same way as in the papers (Chong *et al.*, 1998; Gómez-Stern *et al.*, 2002; Jagannathan and Talluri, 2002; Kulkarni and Li, 1998; Laberteaux *et al*., 2002; Lengliz and Kamoun, 2000; Mascolo, 2000; Quet, 2002), the case of a single bottleneck link shared by n ideal sources (i.e. the sources having unlimited data in store to transmit at each time instant) is considered. Our purpose is to design a controller, to be implemented at the node, which will assure full bottleneck link utilisation and no buffer overflow.

The rate of the cell outflow from the bottleneck buffer depends on the available bandwidth modelled as an *a-priori* unknown, bounded function of time $d(t)$, where

$$0 \leq d(t) \leq d_{\max}. \tag{1}$$

This is motivated by the fact that the ABR service dynamically uses the bandwidth temporarily left non-consumed by rt-VBR and nrt-VBR, which typically support bursty and unpredictable traffic.

The sources send data cells (at the rate determined by the controller) and resource management (RM) cells. The RM cells are processed by the nodes on the priority basis, i.e. they are not queued but sent to the next node without a delay. These cells carry information about network conditions. After reaching the destination they are immediately sent back to the source, along the same path they arrived. The information carried by the RM cells is used to adjust the source rates.

Further in this paper the following notation is used: t denotes time, T represents the control period, and RTT_j stands for the round trip time of the j-th virtual circuit contributing to the bottleneck queue under control $(j = 1, 2, \ldots, n)$. This time is equal to the sum of forward and backward propagation delays denoted by T_{fj} and T_{bj}, respectively,

$$RTT_j = T_{fj} + T_{bj}. \tag{2}$$

Furthermore, $x(t)$ denotes the bottleneck queue length at time t, and x_d stands for the demand value of $x(t)$. The virtual connections are numbered in such a way that

$$RTT_1 \leq RTT_2 \leq \cdots \leq RTT_{n-1} \leq RTT_n. \tag{3}$$

Before setting up a connection, the bottleneck buffer is empty, i.e.

$$x(t < 0) = 0. \tag{4}$$

On the other hand, for $t \leq 0$,

$$x(t) = \sum_{j=1}^{n} \int_0^t a_j(\tau - T_{fj})\,\mathrm{d}\tau - \int_0^t h(\tau)\,\mathrm{d}\tau, \tag{5}$$

where $a_j(t)$ is the j-th source rate at time t, and $h(t)$ represents the bandwidth which is actually consumed by the bottleneck link at time t. It is assumed that we have $a_j(t < 0) = 0$ for any $j = 1, 2, \ldots, n$. Furthermore, if the queue length at the bottleneck link $x(t)$ is greater than zero, then the entire available bandwidth is consumed, $h(t) = d(t)$. Otherwise, i.e. if $x(t) = 0$, then $h(t)$ is determined by the rate of the data arrival at the node. In this case the available bandwidth may not be fully utilised. Consequently, for any time moment t,

$$0 \leq h(t) \leq d(t) \leq d_{\max}. \tag{6}$$

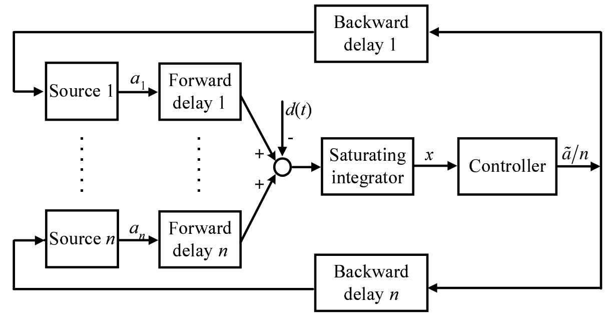

Fig. 1. Control system.

The block diagram of the flow control system considered in this paper is shown in Fig. 1.

The total source rate $\tilde{a}(t)$ is generated by the controller implemented at the switch, and evenly allocated between all the sources contributing to the bottleneck queue. In other words, all the sources are given an equal share of the available ABR bandwidth. Thus the j-th source rate is

$$a_j(t) = \frac{1}{n}\tilde{a}(t - T_{bj}). \tag{7}$$

In this paper a sampled time controller is designed, so the total computed source rate $\tilde{a}(t)$ is changed only at regularly spaced time instants $t = kT$, where k is a non-negative integer. Consequently,

$$\forall t \in \big[kT, (k+1)T\big), \quad \tilde{a}(t) = \tilde{a}(kT) = \text{const} \tag{8}$$

and

$$\begin{aligned} &\forall t \in \big[kT + T_{bj}, (k+1)T + T_{bj}\big), \\ &\quad a_j(t) = a_j(kT + T_{bj}) = \frac{1}{n}\tilde{a}(kT) = \text{const.} \end{aligned} \tag{9}$$

Further in the paper, it is assumed that each round trip time RTT_j is a multiple of the control period, i.e.

$$\forall\, j = 1, 2, \ldots, n\ \exists\, m_j \in \{1, 2, 3, \ldots\}\colon\ RTT_j = m_j T. \tag{10}$$

This assumption can be easily satisfied, since the backward delay T_{bj} can always be appropriately augmented by a source with an extra delay. Even in the worst case, this delay is shorter than the control period T.

The source rate should be determined in such a way that the bottleneck buffer overflow is avoided and the node has always enough data to send. The former condition implies that data cells are not lost and there is no need for their retransmission, while the latter assures full bottleneck link utilisation, which is highly desirable for economic reasons. In the next section a controller which ensures that the two conditions are satisfied is proposed.

3. Proposed Controller

In this section we propose the following control algorithm: The j-th source rate is determined by (7). For any time moment $t \geq 0$, the total source rate signal $\tilde{a}(t)$ in this equation is generated by the controller (placed at the bottleneck node) according to

$$\begin{aligned}\tilde{a}(t) &= \tilde{a}(kT)\\ &= K\left[x_d - x(kT) - \sum_{j=1}^{n}\frac{1}{n}\int\limits_{kT-RTT_j}^{kT}\tilde{a}(\tau)\,\mathrm{d}\tau\right],\end{aligned} \quad (11)$$

where k is the largest integer such that $kT \leq t$, and K denotes the control gain. It is assumed that for any time instant $t < 0$, the total source rate signal $\tilde{a}(t) = 0$. Notice that when the proposed strategy is applied, the j-th source does not send any data for a time less than T_{bj}, i.e.

$$\forall j \quad \forall t < T_{bj}, \quad a_j(t) = 0. \quad (12)$$

Consequently, for any time less than or equal to RTT_1 the queue length $x(t) = 0$.

The strategy proposed in this section combines a generalised continuous time Smith predictor with the sampled time proportional controller. In order to achieve the desired properties of the strategy this gain must be positive and bounded from above by T^{-1},

$$0 < K \leq T^{-1}. \quad (13)$$

In the sequel two theorems presenting important properties of the proposed control strategy are introduced.

Theorem 1. *If the proposed strategy is applied, then the bottleneck link queue length is always bounded from above by its demand value, i.e.* $x(t) \leq x_d$.

Proof. As has already been mentioned, for any time less than or equal to RTT_1 the queue length $x(t) = 0$. Therefore, in order to prove the theorem, it is necessary to show that the queue length does not exceed its demand value x_d at any time t greater than RTT_1.

Let RTT_p be the largest value of RTT_j $(j = 1, 2, \ldots, n)$ less than or equal to the analysed time instant t. In other words, p is defined as the following function of time $t \in [RTT_1, \infty)$:

$$p(t) = \max_{RTT_j \leq t}(j). \quad (14)$$

The particular value $p(t)$ of this function shows that at the time instant $t \in [RTT_1, \infty)$ first cells sent by p sources have already reached the bottleneck link buffer. Therefore, this value is a natural number which belongs to the interval $[1, n]$. Furthermore, for any time instant $t \geq RTT_n$, $p(t) = n$ and for any two time instants t_1 and t_2 satisfying $t_2 \geq t_1$ we have $p(t_2) \geq p(t_1)$ as well.

From (5) and (7) it follows that for any $k \geq 0$ the queue length

$$\begin{aligned}x(kT) &= \sum_{j=1}^{n}\int\limits_{0}^{kT}\frac{1}{n}\tilde{a}(\tau - RTT_j)\,\mathrm{d}\tau - \int\limits_{0}^{kT}h(\tau)\,\mathrm{d}\tau\\ &= \sum_{j=1}^{n}\int\limits_{0}^{kT-RTT_j}\frac{1}{n}\tilde{a}(\tau)\,d\tau - \int\limits_{0}^{kT}h(\tau)\,\mathrm{d}\tau.\end{aligned} \quad (15)$$

Since the total source rate signal $\tilde{a}(t < 0) = 0$, only the sources whose round trip time is less than or equal to kT contribute to the queue length at the time instant kT. Therefore, (15) can be rewritten as

$$x(kT) = \sum_{j=1}^{p(kT)}\int\limits_{0}^{kT-RTT_j}\frac{1}{n}\tilde{a}(\tau)\,\mathrm{d}\tau - \int\limits_{0}^{kT}h(\tau)\,\mathrm{d}\tau. \quad (16)$$

Substituting (15) into (11) for any time moment $t \geq 0$ and $t \in [kT, kT + T)$ the control signal can be expressed as

$$\begin{aligned}\tilde{a}(t) &= \tilde{a}(kT)\\ &= K\left[x_d - \sum_{j=1}^{n}\frac{1}{n}\int\limits_{0}^{kT-RTT_j}\tilde{a}(\tau)\,\mathrm{d}\tau + \int\limits_{0}^{kT}h(\tau)\,\mathrm{d}\tau\right.\\ &\qquad\left. - \sum_{j=1}^{n}\frac{1}{n}\int\limits_{kT-RTT_j}^{kT}\tilde{a}(\tau)\,\mathrm{d}\tau\right]\\ &= K\left[x_d - \sum_{j=1}^{n}\frac{1}{n}\int\limits_{0}^{kT}\tilde{a}(\tau)\,\mathrm{d}\tau + \int\limits_{0}^{kT}h(\tau)\,\mathrm{d}\tau\right]\\ &= K\left[x_d - \int\limits_{0}^{kT}\tilde{a}(\tau)\,\mathrm{d}\tau + \int\limits_{0}^{kT}h(\tau)\,\mathrm{d}\tau\right].\end{aligned} \quad (17)$$

The length of the bottleneck link queue is

$$\begin{aligned}x(t) &= x(kT + \delta)\\ &= x(kT) + \sum_{j=1}^{n}\int\limits_{kT}^{kT+\delta}a_j(\tau - T_{fj})\,\mathrm{d}\tau\\ &\qquad - \int\limits_{kT}^{kT+\delta}h(\tau)\,\mathrm{d}\tau\end{aligned}$$

$$= x(kT) + \sum_{j=1}^{n} \frac{1}{n} \int_{kT}^{kT+\delta} \tilde{a}(\tau - RTT_j)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau$$

$$= x(kT) + \sum_{j=1}^{n} \frac{1}{n} \int_{kT-RTT_j}^{kT+\delta-RTT_j} \tilde{a}(\tau)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau, \tag{18}$$

where k is the largest integer such that $kT < t$ and $\delta = t - kT$ is a positive real number less than or equal to T. Furthermore, taking into account the conditions (8) and (10), we get

$$x(t) = x(kT) + \sum_{j=1}^{n} \frac{1}{n} \tilde{a}(kT - RTT_j)\,\delta - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau. \tag{19}$$

Since the total source rate signal $\tilde{a}(t < 0) = 0$, (19) may be rewritten as

$$x(t) = x(kT) + \sum_{j=1}^{p(kT)} \frac{1}{n} \tilde{a}(kT - RTT_j)\,\delta - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau. \tag{20}$$

Substituting (17) into (20), for any time $t > RTT_1$ we get

$$x(t) = x(kT) + \sum_{j=1}^{p(kT)} \frac{K}{n} \Bigg[x_d - \int_{0}^{kT-RTT_j} \tilde{a}(\tau)\,\mathrm{d}\tau + \int_{0}^{kT-RTT_j} h(\tau)\,\mathrm{d}\tau \Bigg] \delta - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau. \tag{21}$$

On the other hand,

$$\int_{0}^{kT-RTT_j} h(\tau)\,\mathrm{d}\tau = \int_{0}^{kT} h(\tau)\,\mathrm{d}\tau - \int_{kT-RTT_j}^{kT} h(\tau)\,\mathrm{d}\tau. \tag{22}$$

By substituting (22) into (21) the following relation is obtained for any time moment $t > RTT_1$:

$$\begin{aligned} x(t) &= x(kT) + \sum_{j=1}^{p(kT)} \frac{K}{n} \Bigg[x_d - \int_{0}^{kT-RTT_j} \tilde{a}(\tau)\,\mathrm{d}\tau + \int_{0}^{kT} h(\tau)\,\mathrm{d}\tau - \int_{kT-RTT_j}^{kT} h(\tau)\,\mathrm{d}\tau \Bigg] \delta - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau \\ &= x(kT) + \frac{p(kT)}{n} K\,\delta\,x_d - K\,\delta \Bigg[\sum_{j=1}^{p(kT)} \frac{1}{n} \int_{0}^{kT-RTT_j} \tilde{a}(\tau)\,\mathrm{d}\tau - \frac{p(kT)}{n} \int_{0}^{kT} h(\tau)\,\mathrm{d}\tau \Bigg] \\ &\quad - K\,\delta \sum_{j=1}^{p(kT)} \int_{kT-RTT_j}^{kT} \frac{1}{n} h(\tau)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau \end{aligned} \tag{23}$$

and, consequently, using (16), the following inequality can be formulated:

$$x(t) \le x(kT) + K\,\delta\,x_d - K\,\delta\,x(kT) - K\,\delta \sum_{j=1}^{p(kT)} \int_{kT-RTT_j}^{kT} \frac{1}{n} h(\tau)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau. \tag{24}$$

After further calculations, we have

$$\begin{aligned} x(t) &\le x_d - x_d(1 - K\,\delta) + x(kT)(1 - K\,\delta) - K\,\delta \sum_{j=1}^{p(kT)} \int_{kT-RTT_j}^{kT} \frac{1}{n} h(\tau)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau \\ &= x_d - \big[x_d - x(kT)\big](1 - K\,\delta) - K\,\delta \sum_{j=1}^{p(kT)} \int_{kT-RTT_j}^{kT} \frac{1}{n} h(\tau)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta} h(\tau)\,\mathrm{d}\tau. \end{aligned} \tag{25}$$

Equation (6) and inequality (25) directly imply that, for any time moment $t > RTT_1$, the following inequality holds:

$$x(t) \le x_d - [x_d - x(kT)](1 - K\delta). \tag{26}$$

Furthermore, since $0 < \delta \le T$, taking into account (13) it is easy to conclude that

$$0 < K\delta \le 1. \tag{27}$$

Since for any time less than or equal to RTT_1 the queue length $x(t) = 0$, the relations (26) and (27) show that for any time t the queue length is actually always less than its demand value

$$x(t) \le x_d. \tag{28}$$

This conclusion completes the proof. ■

Theorem 1 gives the network designer significant practical information. It shows that using the buffer of a capacity greater than or equal to x_d one assures no buffer overflow and therefore eliminates a cell loss at the bottleneck node.

Another desired property of a suitably designed flow control system is full link utilisation. If the queue length is greater than zero, then the link bandwidth is fully used. The next theorem shows how the buffer capacity should be chosen in order to ensure the strictly positive queue length and, as a consequence, full bottleneck link bandwidth utilisation.

Theorem 2. *If the bottleneck link buffer capacity is greater than or equal to the demand value of the queue length x_d and the following inequality holds:*

$$x_d > \left(\frac{1}{K} + \sum_{j=1}^{n} \frac{1}{n} RTT_j\right) d_{\max}, \tag{29}$$

then for any $t > RTT_n$ the queue length is greater than zero.

Proof. First recall that for any time instant $t > RTT_n$, we have $p(t) = n$. Consequently, from (23) it follows that for any $t > RTT_n$ the queue length

$$\begin{aligned}
x(t) &= x(kT) + K\delta x_d \\
&\quad - K\delta\left[\sum_{j=1}^{n}\frac{1}{n}\int_0^{kT-RTT_j}\tilde{a}(\tau)\,\mathrm{d}\tau - \frac{p(kT)}{n}\int_0^{kT}h(\tau)\,\mathrm{d}\tau\right] \\
&\quad - K\delta\sum_{j=1}^{n}\int_{kT-RTT_j}^{kT}\frac{1}{n}h(\tau)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta}h(\tau)\,\mathrm{d}\tau \\
&= x(kT)(1-K\delta) + K\delta x_d \\
&\quad - K\delta\sum_{j=1}^{n}\int_{kT-RTT_j}^{kT}\frac{1}{n}h(\tau)\,\mathrm{d}\tau - \int_{kT}^{kT+\delta}h(\tau)\,\mathrm{d}\tau.
\end{aligned} \tag{30}$$

Furthermore, since $x(kT)$ is always non-negative, the relations (6), (27) and (30) imply

$$\begin{aligned}
x(t) &\ge K\delta x_d - K\delta\sum_{j=1}^{n}\int_{kT-RTT_j}^{kT}\frac{1}{n}h(\tau)\,\mathrm{d}\tau \\
&\quad - \int_{kT}^{kT+\delta}h(\tau)\,\mathrm{d}\tau \\
&\ge K\delta x_d - K\delta\sum_{j=1}^{n}\frac{1}{n}RTT_j\, d_{\max} - \delta d_{\max} \\
&= K\delta\left[x_d - \left(\sum_{j=1}^{n}\frac{1}{n}RTT_j + \frac{1}{K}\right)d_{\max}\right].
\end{aligned} \tag{31}$$

From (27) and (29) it follows that for any $t > RTT_n$, the product on the right-hand side of (31) is strictly positive, i.e.

$$x(t) > 0. \tag{32}$$

This completes the proof. ■

Theorem 2 shows that using the strategy proposed in this paper we can always assure full link utilisation, provided that the bottleneck node buffer capacity is greater than $(K^{-1} + \sum_{j=1}^{n} RTT_j/n)\, d_{\max}$.

Theorems 1 and 2 highlight important properties of the proposed strategy. However, the implementation of any network control algorithm is possible only if the source rates generated by the algorithm are always bounded and non-negative. In the sequel, another theorem showing those crucial properties of the proposed strategy is formulated and proved.

Theorem 3. *For any time $t \ge 0$ the total source rate signal $\tilde{a}(t)$ generated in accordance with (11) and the rates of all sources are non-negative and bounded.*

Proof. During the initial sampling period $t \in [0, T)$ the total source rate satisfies

$$\tilde{a}(t) = \tilde{a}(0) = K x_d \ge 0. \tag{33}$$

On the other hand, from (17) it follows that for any time instant $k\,T$, where $k \geq 0$, we have

$$\begin{aligned}\tilde{a}\left[(k+1)\,T\right] &= K\left[x_d - \int_0^{(k+1)T} \tilde{a}\,(\tau)\,\mathrm{d}\tau + \int_0^{(k+1)T} h\,(\tau)\,\mathrm{d}\tau\right] \\ &= K\left[x_d - \int_0^{kT} \tilde{a}\,(\tau)\,\mathrm{d}\tau + \int_0^{kT} h\,(\tau)\,\mathrm{d}\tau\right] \\ &\quad + K\left[-\int_{kT}^{(k+1)T} \tilde{a}\,(\tau)\,\mathrm{d}\tau + \int_{kT}^{(k+1)T} h\,(\tau)\,\mathrm{d}\tau\right] \\ &= \tilde{a}\,(k\,T) - K\,T\,\tilde{a}\,(k\,T) + K\int_{kT}^{(k+1)T} h\,(\tau)\,\mathrm{d}\tau \\ &= \tilde{a}\,(k\,T)\,(1 - K\,T) + K\int_{kT}^{(k+1)T} h\,(\tau)\,\mathrm{d}\tau. \quad (34)\end{aligned}$$

Taking into account the inequalities (6), (13) and (33), we conclude that $\tilde{a}(kT)$ is non-negative for any $k \geq 0$. Consequently, for any time $t \geq 0$ the total source rate $\tilde{a}(t)$ and the rates of all sources are also non-negative. Furthermore, from (6) it follows that

$$\begin{aligned}\tilde{a}\big[(k+1)\,T\big] &= \tilde{a}\,(k\,T)\,(1 - K\,T) + K\int_{kT}^{(k+1)T} h\,(\tau)\,\mathrm{d}\tau \\ &\leq \tilde{a}\,(k\,T)\,(1 - K\,T) + K\,T\,d_{\max} \\ &= \tilde{a}\,(k\,T) - K\,T\,\tilde{a}\,(k\,T) + K\,T\,d_{\max} \\ &= \tilde{a}\,(k\,T) + K\,T\,\big[d_{\max} - \tilde{a}\,(k\,T)\big]. \quad (35)\end{aligned}$$

Therefore, the total source rate $\tilde{a}[(k+1)T] \leq \tilde{a}(kT)$, unless $\tilde{a}(kT) < d_{\max}$. This implies that the total source rate $\tilde{a}(t)$ is always bounded by the maximum of the values $\tilde{a}(0) = K\,x_d$ and $d_{\max}$. Consequently, the rates of all sources are also bounded. This conclusion completes the proof. ■

One can easily notice that if the assumptions of Theorem 2 are satisfied, i.e. the inequality (29) holds, then the product $K\,x_d$ is greater than $d_{\max}$, and this product determines the upper bound of the total source rate $\tilde{a}(t)$. Furthermore, the steady state value of the queue length, that is, the queue length when the bandwidth available for the controlled connection $d(t) = d_{ss}$ is constant, can be expressed as follows:

$$x_{ss} = x_d - \left(\frac{1}{K} + \frac{1}{n}\sum_{j=1}^{n} RTT_j\right) d_{ss}. \quad (36)$$

From (36) it can be seen that decreasing the demand value x_d helps to obtain a better quality of the service, i.e. a smaller cell delay time and a smaller variation in this time. Therefore, this relation can be used to attain an appropriate trade-off between the required quality of the service and the degree of network utilisation.

Let us finally remark that when $K = T^{-1}$, the strategy proposed in the paper becomes a dead-bit control algorithm combined with the Smith predictor. In that case the flow rate $\tilde{a}(t)$ during the $(k+1)$-th control period is equal to the average bottleneck buffer link depletion rate on the k-th control interval,

$$\tilde{a}\big[(k+1)\,T\big] = \frac{1}{T}\int_{kT}^{(k+1)T} h\,(\tau)\,\mathrm{d}\tau. \quad (37)$$

4. Simulation Example

In order to verify the performance of the algorithm presented in this paper, computer simulations of a wide area ATM network with $n = 3$ sources were performed. The connections of the network are characterised by the following parameters: $RTT_1 = 20$ ms, $RTT_2 = 30$ ms, and $RTT_3 = 70$ ms. The maximum bandwidth was $d_{\max} = 4$ Mb/s, and the bandwidth actually available for the controlled connections (measured in cells per second) is shown in Fig. 2. The sampling period of the controller was $T = 10$ ms, and the gain $K = 0.03$ ms^{-1}. In order to assure a full link utilisation, the demand value of the queue length must be greater than 725.43 cells. To satisfy

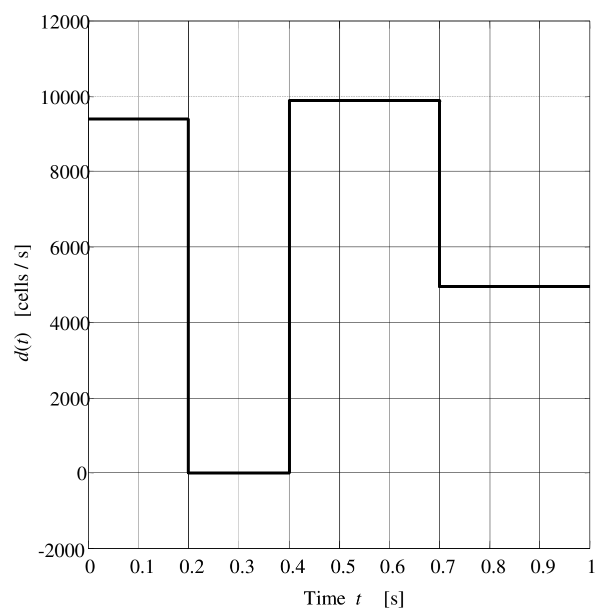

Fig. 2. Available bandwidth.

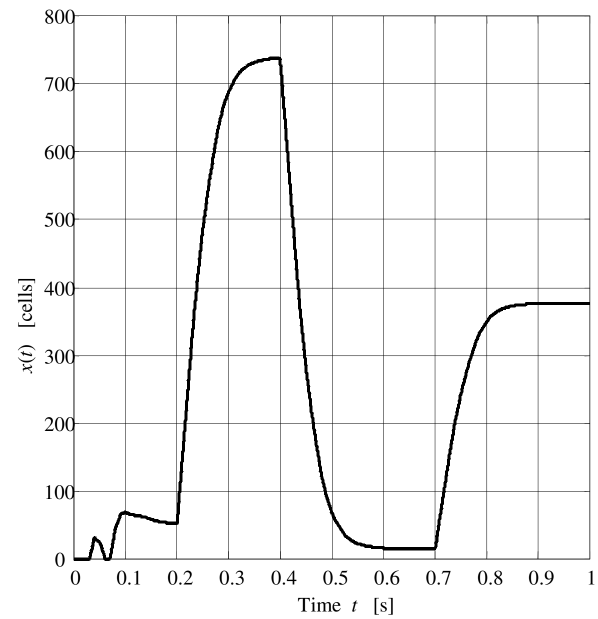

Fig. 3. Queue length.

this condition, $x_d = 740$ cells were chosen. The queue length in the buffer of the bottleneck link is presented in Fig. 3. From this figure it can be seen that the length never exceeds its demand value. This implies no cell loss in the bottleneck node. Furthermore, for any time greater than $T_f = 70$ ms the length is positive, which leads to the conclusion that the bottleneck link bandwidth is fully utilised. Finally, Fig. 4 shows that the total source rate $\tilde{a}(t)$ is non-negative and bounded from above by its initial value.

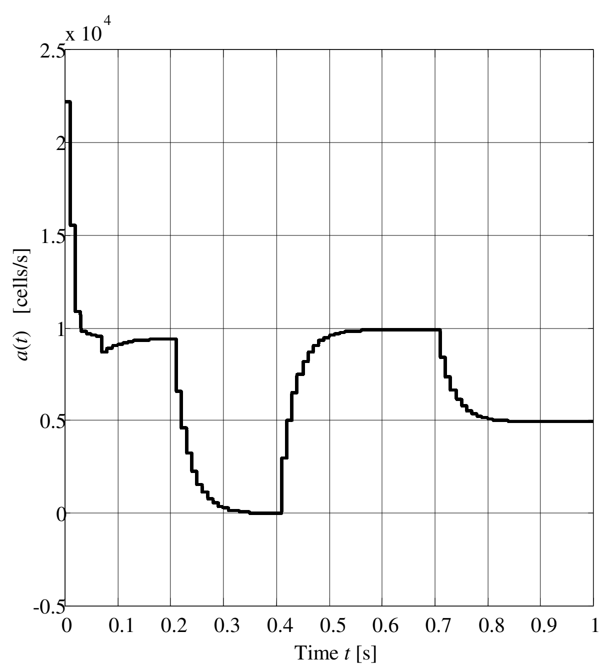

Fig. 4. Total source rate $\tilde{a}(t)$.

5. Conclusions

In this paper a new control strategy for flow control in multi-source ATM networks has been proposed. The strategy combines the Smith predictor with the sampled time proportional controller. The strategy guarantees even resource allocation, no cell loss and full bottleneck link utilisation. Furthermore, it was demonstrated that the transmission rates of all controlled sources, generated by the proposed strategy are always bounded and non-negative. This permits direct implementation of the strategy in a network environment.

Finally, let us remark that the control scheme proposed in this paper is particularly well suited for loss sensitive applications. However, for the traffic which is delay-sensitive, a modification of the scheme decreasing the average queuing time and its variation, possibly at the expense of some cell loss and slightly worse bottleneck link utilisation, might be a feasible option.

Acknowledgements

This work has been partly financed by the State Committee for Scientific Research in Poland under the research grant 3 T11C 043 26.

References

Chong S., Nagarajan R. and Wang Y.T. (1998): *First-order rate-based flow control with dynamic queue threshold for high-speed wide-area ATM networks.* — Comp. Netw. ISDN Syst., Vol. 29, No. 17–18, pp. 2201–2212.

Gómez-Stern F., Fornés J.M. and Rubio F.R. (2002): *Dead-time compensation for ABR traffic control over ATM networks.* — Contr. Eng. Prac., Vol. 10, No. 5, pp. 481–491.

Imer O.C., Compans S., Basar T. and Srikant R. (2001): *Available bit rate congestion control in ATM networks.* — IEEE Contr. Syst. Mag., pp. 38–56.

Izmailov R. (1995): *Adaptive feedback control algorithms for large data transfers in high-speed networks.* — IEEE Trans. Automat. Contr., Vol. 40, No. 8, pp. 1469–1471.

Jagannathan S. and Talluri J. (2002): *Predictive congestion control of ATM networks: Multiple sources/single buffer scenario.* — Automatica, Vol. 38, No. 5, pp. 815–820.

Jain R. (1996): *Congestion control and traffic management in ATM networks: Recent advances and a survey.* — Comput. Netw. ISDN Syst., Vol. 28, No. 13, pp. 1723–1738.

Kulkarni L.A. and Li S. (1998): *Performance analysis of a rate-based feedback control scheme.* — IEEE/ACM Trans. Networking, Vol. 6, No. 6, pp. 797–810.

Laberteaux K.P., Rohrs Ch.E. and Antsaklis P.J. (2002): *A practical controller for explicit rate congestion control.* — IEEE Trans. Automat. Contr., Vol. 47, No. 6, pp. 960–978.

Lengliz I. and Kamoun F. (2000): *A rate-based flow control method for ABR service in ATM networks*. — Comput. Networks, Vol. 34, No. 1, pp. 129–138.

Mascolo S. (1997): *Smith's principle for congestion control in high speed ATM networks*. — Proc. 36th IEEE Conf. *Decision and Control*, San Diego, pp. 4595–4600.

Mascolo S. (1999): *Congestion control in high-speed communication networks using the Smith principle*. — Automatica, Vol. 35, No. 12, pp. 1921–1935.

Mascolo S. (2000): *Smith's principle for congestion control in high-speed data networks*. — IEEE Trans. Automat. Contr., Vol. 45, No. 2, pp. 358–364.

Quet P.F., Ataslar B., Iftar A., Özbay H., Kalyanaraman S. and Kang T. (2002): *Rate-based flow controllers for communication networks in the presence of uncertain time-varying multiple time-delays*. — Automatica, Vol. 38, No. 6, pp. 917–928.

Received: 6 June 2003
Revised: 10 November 2003