Smith's Principle for Congestion Control in High Speed ATM Networks

Saverio Mascolo

Dipartimento di Elettrotecnica ed Elettronica, Politecnico di Bari, Via Orabona 4, 70125 Bari, Italy

e-mail: mascolo@poliba.it

Abstract

In high speed communication networks, long propagation delays are critical for the stability of closed loop congestion control algorithms. In this paper, Smith's principle is proposed as a key tool to design a feedback control law for congestion avoidance in high speed ATM networks. The proposed algorithm assures the stability of network queues and the full utilization of network links in a realistic scenario where many connections, with different propagation delays, share the network. Finally, Smith's principle is proposed to enhance the flow control algorithm of the TCP internet protocol.

1. Introduction

In recent years a lot of research has been developed in the field of packet switching communication networks. The objective is to transmit multimedia traffic over a "uniform" communication medium. The results of these efforts constitute the emerging Asynchronous Transfer Mode (ATM) technology, which is now coming to the market. This technology is conceived to merge the advantages of circuit switching technology (telephone networks), with those of packet switching technology (computer networks). Circuit switching technology, establishing a physical connection from the sender to the receiver, enables real time data transmission (e. g. voice traffic): the drawback is that the network is under utilized because communication links are hold by established connections even during idle periods. On the contrary, packet switching technology allows the sharing of network links among the users. This improves network utilization with the drawback that is difficult to ensure quality to real time data transmission. ATM technology is conceived to merge both the advantages of circuit switching and packet switching technology by means of the concept of *virtual circuit* [1], [2].

The goal of ATM networks is to support a broad range of services with distinct requirements for bandwidth, delay and cell loss. A key issue is the "efficient coexistence" of Constant Bit Rate (CBR) services, Variable Bit Rate (VBR) services and "best effort" services, also termed Available Bit Rate (ABR). ABR traffic is typically characterized by unspecified requirements for throughput and delay and it was conceived to rapidly "fill in" the bandwidth left unused by CBR and VBR traffic. Many algorithms dealing about congestion control have been proposed [3]-[12]. However none of these is completely satisfactory either for complexity or for lack of stability properties, as is well reported in the paper by Benmohamed and Meerkov [3]. In fact, due to propagation delay, most algorithms exhibit persistent oscillations and can even be unstable. In [3] and [4] an analytic method for the design of a congestion controller, which ensure good dynamic performance along with fairness in bandwidth allocation, has been proposed. However this algorithm requires a complex on-line tuning of control parameters in order to ensure stability and damping of oscillations under different network conditions. Moreover, it is difficult to prove the global stability, due to the complexity of the control strategy. In [5], a dual PD controller has been proposed to make easier the implementation of the algorithm presented in [3]. In [6] an algorithm using a Smith's predictor has been illustrated which uses per VC FIFO queuing.

The objective of this paper is to find a control law for ABR input rates so that network bandwidth is fully utilized without incurring network congestion. In particular, following a classical control approach, the dynamic behavior of each queue in response to network input rates is modeled as the cascade of an integrator with a sum of time delays. Then a controller based on Smith's principle is designed. The resulting algorithm assures no cells loss and full utilization of network links in presence of many connections, with different round trip delays, sharing the network. Moreover it does not require per VC queuing, but only a common FIFO queue per switch output link.

The control algorithm is developed starting from an accurate mathematical model. Therefore, its stability and efficiency are rigorously demonstrated. Unlike the algorithms proposed in [7] and [8], where links with constant available bandwidth have been assumed, the interaction of ABR traffic with other traffic is considered here by means of a time-varying available link bandwidth. Moreover, since from a practical point of view it is not meaningful to measure available bandwidth, this variable is modeled as a disturb input. In [8] it was assumed that all the connections sharing the bottleneck link were characterized by the same round trip delay. This restrictive assumption is here relaxed.

0-7803-4187-2 4595-4600

The paper is structured as follows: Section 2 describes the network model, the traffic model, the control architecture and the queue model; in Section 3 the proposed control law is developed and its properties are demonstrated; in Section 4 an application to the TCP internet protocol is described; finally, in Section 5 simulation results are reported.

2. The Model

2.1 The network model

The network employs a store-and forward service, i.e., cells enter the network from the source edge nodes, are then stored and forwarded along a sequence of intermediate nodes and communication links, finally reaching their destination nodes.

Mainly following the notation reported in [3], the network can be considered as a graph consisting of a set $N=\{1,..n\}$ of nodes (properly switches) connected by a set $L=\{1,...l\}$ of communication links. For each node $i \in N$, let $O(i) \subset L$ denote the set of its outgoing 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/sec), where t_i is the transmission time of a packet, and by its propagation delay of t_{di} sec. Each node has a processing capacity of $1/t_{pri}$ cell/sec, where t_{pri} is the time the switch i needs to take a packet from the input and place it on the output queue. 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. Finally it is worth noting that, in high speed wide area networks, the bandwidth delay product $c_i t_{di}$ (in pipe cells) represents a large number of cells “in flight” on the transmission link.

2.2 The traffic model

The network traffic is contributed by source/destination pairs $(S,D) \in N \times N$. To each (S,D) connection is associated a Virtual Circuit (VC) mapped on the path $p(S,D)$ [1], [3], [6]. The path contains one node for every network node and one directed link $e=(a,b)$ for every communication link from node a to node b. Therefore a virtual circuit i is specified by the sequence of links $e^i{}_1 e^i{}_2 ... e^i{}_n$ that traverses as it goes through the network.

A deterministic fluid model approximation of the cell flow is assumed, i.e., sources transmission rates are described by the function of time u(t) measured in cells/sec. An ABR source is expected to declare only its peak cell rate, that is, its maximum transmission speed $c_S = 1/t_S$. Moreover, it is assumed that ABR sources always have cells to send, that is they are *persistent* sources [2].

2.3 The control architecture

The feedback scheme recommended by ATM Forum is assumed [13]. This scheme requires that each ABR traffic source sends one control cell (RM cell) every N data cells. Each node i has a congestion controller which periodically computes, for each outgoing link $j \in O(i)$, an admissible transmission rate which is unique for all the *virtual circuits* sharing the same outgoing link. Each node encountered by the RM cell along the VC path, stamps the computed value for the input rate on the RM cell only if this value results to be less than the rate already stored. In this way, at the destination, the RM cell carries the minimum input rate over all the encountered switches and it comes back to the source conveying the minimum allowed rate. Upon receiving of this rate, the source sets the input rate to this value.

2.4 The queue model

In this subsection a dynamic model of each queue in response to input and output rate changes is developed. Fig. 1 shows two connections sharing one outgoing link.

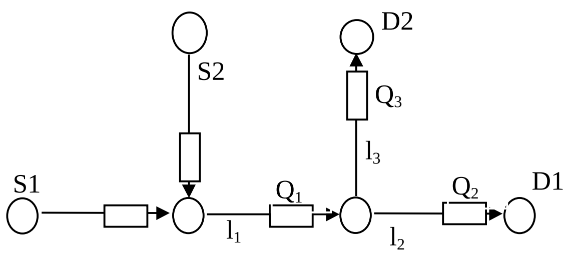

Fig. 1: Scheme of connections (S_1,D_1) and (S_2,D_2) sharing link l_1 and queue Q_1.

Each output link has a common FIFO queue for all virtual circuits sharing it. Let x_j(t) be the queue level associated with the link l_j. By writing flow conservation equations, the level of occupancy x_j(t), starting at $t=0$ with $x_j(0)=0$, is

$$x_j(t) = \int_0^t \sum_{i=1}^{n} u_{ij}(\tau - T_{ij}) \cdot d\tau - \int_0^t d_j(\tau) \cdot d\tau$$

where n is the number of connections sharing the queue, $u_{ij}(t)$ is the inflow rate due to the *i-th* connection, T_{ij} is the propagation delay from the *i-th* source to the *j-th* queue, and $d_j(t)$ is the rate of packets leaving the *j-th* queue, that is, the *ABR* available bandwidth. It is assumed that the propagation delay is dominant compared to other delays (processing, queuing, etc.). Consequently, the round trip time is assumed to be constant and measured when a new connection is established. Note that, since an output link is shared by ABR, VBR and CBR traffic, the available ABR bandwidth depends on the input traffic loading the link. Moreover, since it can be difficult to measure the available ABR bandwidth, $d(t)$ is here modeled as a disturb.

3. The control law

The aim of this section is to design a feedback control law $u(t)$ for the input rate of each ABR connection such that each network queue level $x(t)$ satisfies the following *stability condition*

$$x(t) \le r^o \tag{1}$$

where r^o is the queue capacity. Moreover, the control has to guarantee high utilization of network links. Formally, this can be expressed by the following *efficiency condition*

$$x(t) > 0 \quad \text{for } t>T \tag{2}$$

where T represents the transient time after the starting of network operation. In fact, this condition guarantees that any link has always data to send.

Fig. 2 shows the block diagram of the system model consisting of

1) the bottleneck FIFO queue modeled in the Laplace domain by the integrator $1/s$;
2) the available ABR bandwidth $d(t)$ modeled as a disturb;
3) the round trip delay T_i, ($i=1,n$) of each connection sharing the queue;
4) the controller transfer function $G(s)$;
5) the set point $r(t)$.

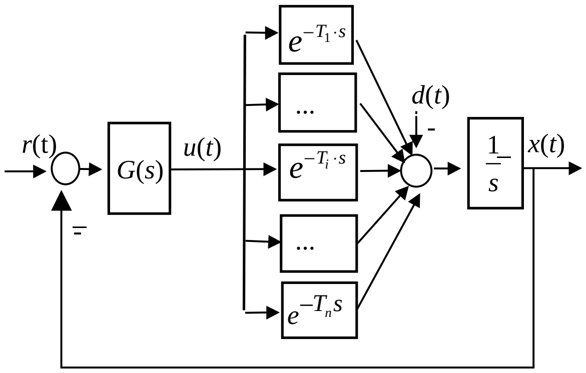

Fig. 2: Block diagram of n VC connections sharing a common FIFO queue

Due to the large delays inside the feedback loop, queue level dynamics might exhibit oscillations, and even become unstable. Since the model of the communication system is known without parameter uncertainty, a controller can be designed following Smith's principle [14], [15]. It is worth noting that, in wide area networks, round trip *delays* are mostly determined by propagation delay. To take into account the jitter of round trip time due to queuing time, a model containing time varying delays could be considered.

The key idea is to look for a controller $G(s)$ so that the input-output dynamic of the system reported in Fig. 2 becomes equivalent to the one of the system reported in Fig. 3.

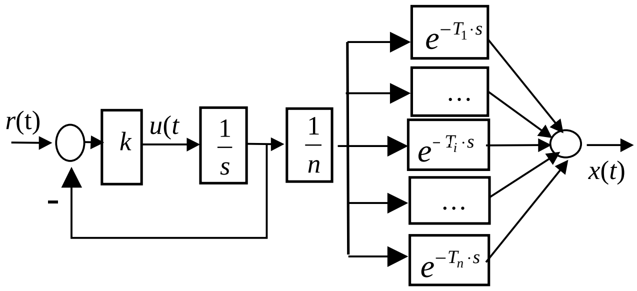

Fig. 3: Block diagram of the desired input-output dynamic

A major advantage of the system reported in Fig. 3 is that it is a first order system with a sum of delays in cascade. By letting the set point $r(t)$ be the step function $r^o \cdot 1(t)$, the output does not overshoot and the queue level is bounded by r^o.

Proposition 1: The system reported in Fig. 3, where the reference signal is the step function $r(t)= r^o \cdot 1(t)$, satisfies the stability condition $x(t) \le r^o$.

Proof:

The Laplace transform of the output $x(t)$ in response to the set point $r^o \cdot 1(t)$ is:

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

By anti-transforming $X(s)$, it follows:

$$x(t) = \frac{r^o}{n} \sum_{i=1}^{n} \left(1 - e^{-k(t-T_i)}\right) \cdot 1(t-T_i) \le r^o \tag{3}$$

This completes the proof.

Proposition 2: The transfer function $X(s)/R(s)$ of the systems reported in Fig. 2 and Fig. 3, respectively, can be made equivalent by using the controller described by the transfer function

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

Proof:

By equating the transfer functions of the systems in Fig. 2 and Fig. 3

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

the controller (4) is derived.

Fig. 4 shows the block diagram of the controller (4).

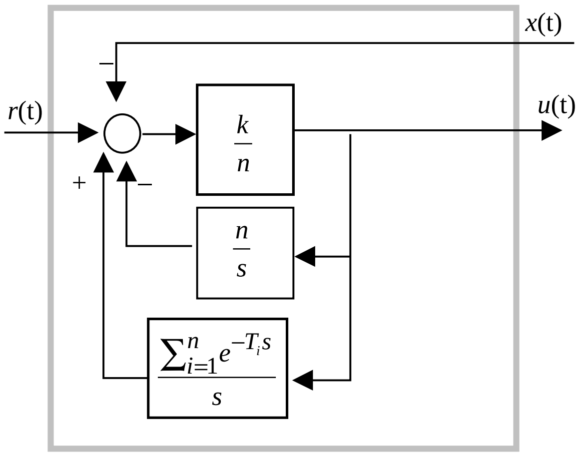

Fig. 4: Block diagram of the controller $G(s)$

By looking at this figure, it is easy to write the rate control equation in the time domain, that is[1]

$$u(t) = \frac{k}{n}\left(r^o - x(t) - n\int_0^t u_i(\tau)\cdot d\tau + \sum_{i=1}^{n}\int_0^{t-T_i} u_i(\tau)\cdot d\tau \right) =$$
$$= \frac{k}{n}\left(r^o - x(t) - \sum_{i=1}^{n}\int_{t-T_i}^{t} u(\tau)\cdot d\tau \right) \qquad (5)$$

This equation can be intuitively interpreted as follows: the computed input rate is proportional, through the coefficient k/n, to the available queue room $r^o - x(t)$ decreased by the number of cells released by each connection during the last corresponding round trip time T_i, that is, the sum of "in flight" cells of all connections sharing the queue.

Remark 1: The queue dynamics (3) is characterized by the time constant $\tau = 1/k$. Therefore the transient can be considered exhausted after the time $T_{tr} = \max_i (T_i) + 4\tau$.

To guarantee high link utilization in presence of the disturb $d(t)$, condition (2) has to be satisfied. A link transmission capacity normalized to unity is assumed, so that, if all bandwidth is available for ABR traffic, it results $d(t)=1(t)$. The coexistence of ABR with (VBR + CBR) traffic, which consumes the bandwidth $b(t)$, reduces the available ABR bandwidth to $d(t) = 1(t) - b(t) \geq 0$. By defining $b_m = \min_t \{b(t)\}$ it results

$$d(t) \leq 1(t) - b_m = a\cdot 1(t) \text{ where } a = (1-b_m) < 1.$$

Proposition 3: Control law (5) guarantees $x(t)>0$ for $t > \max_i (T_i) + 4\tau$ if the following condition is satisfied:

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

Proof:

By considering that the available ABR bandwidth $d(t)$ is an unknown function such that $d(t) \leq a\cdot 1(t)$, with $a<1$, the "worst case" disturb $a\cdot 1(t)$ is assumed. By using the controller (4), the transfer function from the available bandwidth $d(t)$ to the queue level $x_d(t)$ is

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

which, for $D(s) = \frac{a}{s}$, gives

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

Let $x_r(t)$ denote the system response to the set point r^o. The queue dynamics is

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

which, for $t > \max_i (T_i) + 4\tau$, becomes:

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

By requiring that $x(t)>0$, Proposition 3 follows.

Remark 2: Proposition 3 guarantees full utilization of network links if each queue capacity is at least equal to the

[1] Note that $u(t)=u_i(t)$ for $i=1,n$

number of “in flight” cells contained in a pipe with a round trip delay $\sum_{i=1}^{n} T_i / n + \tau$, that is, the mean of VCs round trip delays plus the system time constant $1/k$.

4. Smith’s principle for TCP internet protocol[2]

The TCP protocol for Internet was designed to operate reliably over almost any transmission medium regardless of transmission rate and propagation delay. The introduction of fiber optics is resulting in ever-higher transmission speeds and the fastest communication paths are moving out of the domain for which TCP was originally engineered [16]. Nowadays, active research is going on to extend the domain of TCP operability to high speed networks [16], [17]. In this section, again Smith’s principle is proposed as a key tool for designing an enhanced flow control algorithm for internet.

TCP flow control implements an end to end sliding window control [18], [19]. There are two buffers, one on the send side with capacity *MaxSendBuffer*, and one on the receive side with capacity *MaxRcvBuffer*. The size of the window sets the amount of data that can be sent without waiting for acknowledgment from the receiver. The TCP on the receive side must keep

$$LastByteReceived - NextByteRead \leq MaxRcvBuffer$$

to avoid overflow of its buffer. Then it advertises a window size of

$$AdvertisedWindow = MaxRcvBuffer - (LastByteReceive - NextByteRead)$$

which represents the amount of free space remaining in its buffer. Note that the *AdvertisedWindow* can be considered the equivalent of the quantity $(r^o - x(t))$ in eq. (5) of this paper. TCP on the send side must satisfy the advertised window it gets from the receiver

$$(LastByteSent - LastByteAck) \leq AdvertisedWindow \quad (6)$$

Following equation (5), the control equation (6) can be modified in

$$(LastByteSent - LastByteAck) \leq AdvertisedWindow - InPipeCells \quad (7)$$

This control equation guarantees no packet loss. Therefore, the traffic due to retransmissions is expected to reduce drastically, improving the *Goodput*, i.e. the *Throughput-RetransmissionTroughput.*

5. Simulation results

In Fig. 5, four ABR connections, which shares a FIFO bottleneck queue with (VBR+CBR) traffic, are considered.

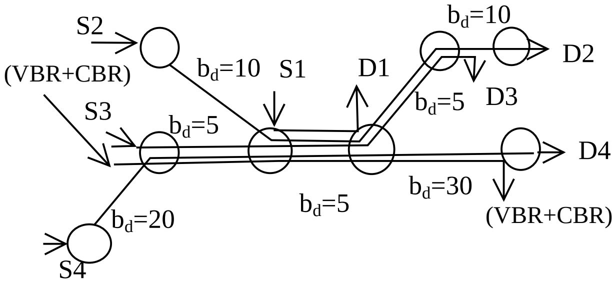

Fig. 5: Network topology and traffic scenario

The connections are characterized by a bandwidth-delay product of 10, 30, 60 and 120 cells, respectively. Note that a bandwidth-delay product of 10 is typical of a *local area network* (LAN), while one of 120 cells is typical of a metro or regional *wide area network* (WAN). For sake of simplicity, it is assumed that each ABR source has the same peak cell rate c_s normalized to unity. The interaction with quality constrained traffic (CBR+VBR) is considered by means of the time varying available bandwidth $d(t)$ whose time waveform is shown in Fig. 6. A buffer capacity r^o=40 *cells*, and a constant gain k=*0.1/sec* are assumed. Fig. 7 shows the sum of all ABR input rates at the bottleneck queue: it can be noted that the steady state value of u captures all available ABR bandwidth; Fig. 8 shows that the bottleneck queue dynamics is bounded by r^o, that is, cell losses are avoided.

Conclusions

Smith’s principle has been proposed as a key tool for designing congestion control algorithms for ABR traffic in ATM networks. The presented algorithm works in a realistic scenario consisting of many ABR connections which share available bandwidth with VBR and CBR traffic. Simulation results show the efficiency of the algorithm. Finally, Smith’s principle is proposed to improve the flow control of the TCP protocol for Internet.

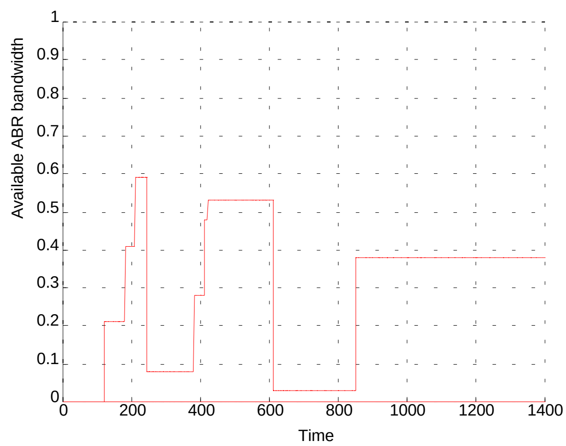

Fig. 6: Available ABR bandwidth

[2] An extended version of this section is work in progress.

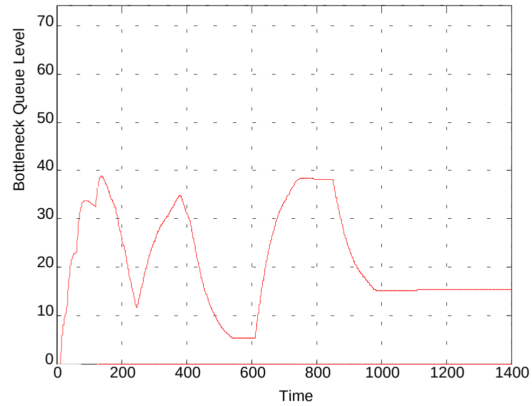

Fig. 7: Global input rate at the bottleneck queue

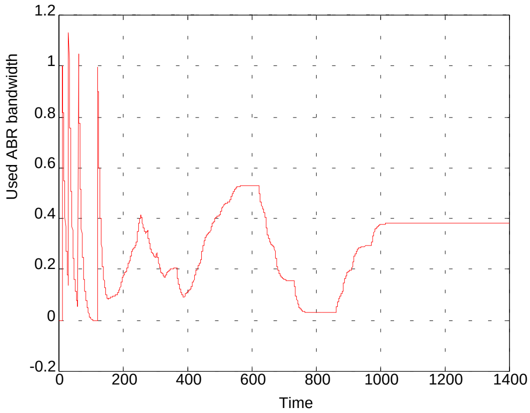

Fig. 8: Bottleneck queue level behavior

References

[1] P. Varaiya and J. Walrand, "High-Performance Communications Networks," *Morgan Kaufmann Publishers*, 1996.

[2] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey," *Computer Networks and ISDN Systems*, vol. 28, no. 13, pp. 1723-1738, 1996.

[3] L. Benmohamed, S. M. Meerkov, "Feedback Control of Congestion in Packet Switching Networks: The Case of a Single Congested Node," *IEEE/ACM Trans. on Networking*, vol.1, no.6, pp.693-708, 1993.

[4] L. Benmohamed, S. M. Meerkov, "Feedback Control of Congestion in Packet Switching Networks: The Case of Multiple Congested Node," *Proc. of the American Control Conference*, pp. 1104-1108, 1994.

[5] A. Kolarov, G. Ramamurthy, "A Control Theoretic Approach to the Design of Closed Loop Rate Based Flow Control for High Speed ATM Networks," *Proc. of IEEE Infocom97*, Kobe, Japan, 1997.

[6] S. Mascolo, D. Cavendish, M. Gerla, "ATM Rate Based Congestion Control Using a Smith Predictor: an EPRCA Implementation," *Proc. IEEE Infocm96*, S. Francisco, March 24-25, 1996. To appear on *Performance Evaluation*, Special Issue on "ATM Traffic Control", North-Holland, 1997.

[7] R. Izmailov, "Adaptive Feedback Control Algorithms for Large Data Transfer in High-Speed Networks," *IEEE Trans. on AC*, vol. 40, no. 8, pp. 1469-1471, 1995.

[8] R. Izmailov, "Analysis and optimization of feedback control algorithms for data transfer in high-speed networks," *SIAM J. Control and Optimization*, vol. 34, no. 5, pp. 1767-1780, 1996.

[9] V. Jacobson, "Congestion Avoidance and Control," *Proc. of the SIGCOMM'88 Symposium*, Stanford, CA, pp. 314-329, 1998.

[10] C. E. Rohrs, R. A. Berry, "A Linear Control Approach to explicit rate feedback in ATM networks", *Proc. of IEEE Infocom97*

[11] K. W. Fendick, M. A. Rodrigues and A. Weiss, "Analysis of a Rate-Based Feedback Control Strategy for Long Haul Data Transport," *Performance Evaluation*, vol. 16, pp. 67-84, 1992.

[12] F. Bonomi, D. Mitra, and J. B. Seery, "Adaptive Algorithms for Feedback-Based Flow Control in High-Speed, Wide-Area ATM Networks", *IEEE J. on Selected Areas in Communications*, vol. 13, no. 7, 1995.

[13] ATM Forum Technical Committee TMWG, "ATM Forum Traffic Management Specification Version 4.0", *ATM Forum/95-0013R11*.

[14] O. Smith, "A Controller to Overcome Dead Time", *ISA J.*, vol.6, no.2, ,pp.28-33, 1959.

[15] G. F. Franklin, J. D. Powell, A. Emami-Naeini, *Feedback Control of Dynamic systems*, Addison Wesley, 1994.

[16] V. Jacobson, "TCP Extensions for High Performance," *Internet Draft*, February 1997.

[17] D. Sisalem, H. Schulzrinne, "Congestion Control in TCP: Performance of Binary Congestion Notification Enhanced TCP Compared to Reno and Tahoe TCP," *Proc.of ICC96*.

[18] L. L. Peterson, B. S. Davie, *Computer Networks*, Morgan Kaufmann, 1996.

[19] D. E. Comer, *Internetworking with TCP/IP*, Prentice Hall, 1995.