Predictive Congestion Control of ATM Networks: Multiple Sources/Single Buffer Scenario *

S. *Jagannathan and J. Talluri*
Intelligent Systems Laboratory
Department of Electrical Engineering
The University of Texas
6900 North Loop 1604 West
San Antonio, Texas 78249

August 31, 2000

Abstract

This paper proposes an predictive congestion control methodology for the Available Bit Rate (ABR) service class in an ATM network for the case of multiple node with single buffer scenario. Adaptive controller is developed to control traffic where sources adjust their transmission rates in response to the feedback information from the network nodes. Specifically, the dynamics of the ABR sources with the buffer is modeled as a nonlinear system and an autoregressive moving average based (ARMAX) adaptive controller is designed to predict the explicit values of the transmission rates of the sources so as to prevent network congestion. Stability analysis of the closed-loop system is presented. Simulation results are provided to justify the theoretical conclusions.

1 Introduction

Asynchronous transfer mode (ATM) is a key technology for integrating broad-band multimedia services (B-ISDN) in heterogeneous networks, where data, video and voice sources transmit information. Due to the uncertainties of broad band traffic patterns, unpredictable statistical fluctuations of traffic flows can cause congestion in the network switches, concentrators, transmission links and so on.

In a B-ISDN, congestion is defined as a condition of an ATM network, where the network does not meet a stated performance objective or refereed to as Quality of Service (QoS). In order to prevent the QoS from severely degrading during short-term congestion, an appropriate congestion control scheme is required. Since the ATM forum decided to use a closed-loop rate based congestion control scheme as the standard for the available bit rate (ABR) service [7], several feedback congestion control schemes were proposed in the literature [3-9]. In terms of feedback control, the congestion control is viewed as changing the source rates and regulating the traffic submitted by these sources onto the network connections. Most of these schemes are based upon a linear model of the ABR buffer. Each ABR buffer has then a corresponding congestion controller, which sends congestion notification cells (CNC) back to the sources. These proposed feedback control schemes [5, 8-9] were based on buffer length, buffer change rate, cell losses, buffer thresholds and so on. Most of the papers [3-7,9] report simulation results only. Mathematical analysis is not included to demonstrate the performance of these controllers in terms of QoS.

In this paper, the traffic rate on the network connections is considered to be nonlinear and the ABR buffer availability at each node element of an ATM network is modeled as a nonlinear system. A novel congestion controller scheme is proposed using an adaptive methodology, where an explicit rate for each source is derived. Autoregressive moving average (ARMAX) based approach is employed to approximate the unknown network traffic that is being accumulated at the buffer. This estimated traffic value is used in the adaptive controller to modify the source rate. Tuning laws are provided for the coefficients of the ARMAX traffic model to improve accuracy. Closed-loop stability of the adaptive congestion controller is proven using a Lyapunov-based analysis. It is shown that the controller guarantees the desired performance, which is specified in terms of buffer length, even though

*This research is supported through the NSF CAREER Award Number 9985739

network traffic uncertainties exists. Simulation results are given to justify the theoretical conclusions for multiple source and single node scenario.

2 Background

2.1 Autoregressive Moving Average Approach

A general function $f(x) \in C^{(S)}$ can be approximated using an ARMAX model as

$$f(x) = \theta^T \varphi(x(k)) + \epsilon(k) \tag{1}$$

where θ^T are constant parameters and $\varphi(k)$ denotes the regression matrix at the instant k, with $\epsilon(k)$ a reconstruction error vector. The output of the ARMAX model is defined as

$$\hat{y}(k) = \hat{\theta}^T(k)\varphi(x(k)). \tag{2}$$

It is important to note that, the unknown parameters in an ARMAX approach enter in a linear fashion as given in 2.

2.2 Stability of Systems

To formulate the discrete-time controller, the following stability notion is needed. Consider the nonlinear system given by

$$\begin{aligned} x(k+1) &= f(x(k), u(k)) \\ y(k) &= h(x(k), \end{aligned} \tag{3}$$

where $x(k)$ is a state vector, $u(k)$ is the input vector and $y(k)$ is the output vector. The solution is said to be *uniformly ultimately bounded (UUB)* if for all $x(k_0) = x_0$, there exists an $\delta \geq 0$ and a number $N(\delta, x_0)$ such that $\|x(k)\| \leq \delta$ for all $k \geq k_0 + N$.

It is not possible to show the asymptotic stability of the closed-loop error system in the presence of bounded disturbances and approximation errors.

2.3 Traffic Rate Modeling on Network Connections

Figure 1 shows a multiple source single node/buffer scenario for an ATM network. It is modeled as a multi-input and multi-output (MIMO) discrete-time nonlinear system, to be controlled, as follows

$$x(k+1) = f(x(k), x(k-1), ,..) + Tu(k) + d(k) \tag{4}$$

with state $x(k) \in \Re^n$ being the actual buffer length at time instant k, and $u(k) \in \Re^n$ being the feedback

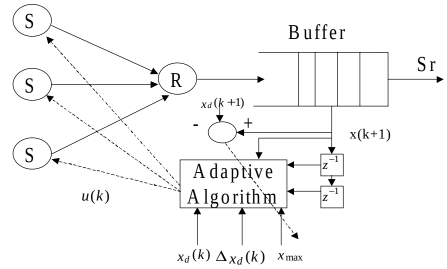

Figure 1: Multiple Source/Single Buffer Model of an ATM Network.

signal or referred to as source rate correction. The nonlinear function, $f(\cdot)$, which is assumed to be unknown,includes the cell arrival rate from the source to the buffer, buffer service rate and the current queue length. The disturbance vector acting on the buffer at the instant k is $d(k) \in \Re^n$, which we assume unknown but bounded so that $\| d(k) \| \leq d_M$ a known constant. Figure 1 shows the case of multiple sources with one buffer scenario where the nonlinear function, $f(\cdot)$, describes the dynamics of the source/buffer system. It is well known in the literature that by monitoring mean cell arrival rate (MSR), peak cell rate (PCR) and so on for a ABR source does not indicate the network traffic [3-9]. Further, the net dynamics is not linear due to bursty traffic conditions. The disturbance vector, $d(k)$, can be considered as an unexpected traffic burst or a change in available bandwidth.

The state vector $x(k)$ is scalar variable for a single network node scenario and it is a vector for the case of multiple nodes. Similarly, the feedback source rate is a scalar variable for one source/one buffer and multiple sources/one buffer scenarios, whereas it is a vector otherwise. In the case of multiple sources with single buffer scenario, the source rate obtained through feedback is either used to regulate equally the transmission rates of the sources or the transmission rates for each source is calculated by dividing the feedback source rate with the individual PCRs.

Given a desired buffer length, $x_d(k)$, define the performance criterion in terms of buffer length error as

$$e(k) = x(k) - x_d(k). \tag{5}$$

Equation (5) can be expressed as

$$e(k+1) = x(k+1) - x_d(k+1). \tag{6}$$

Using (4) in (6), the dynamics of the buffer (4) can be written in terms of the buffer length error as

$$e(k+1) = f(x(\cdot)) - x_d(k+1) + Tu(k) + d(k) \quad (7)$$

In an ATM network, the goal of a traffic rate controller then is to make the network available bandwidth to be fully utilized while maintaining good QoS. In this paper, the objective of selecting a suitable feedback traffic rate input, $u(k)$, is to minimize the difference between the actual or desired buffer length so as to prevent congestion while maintaining maximum link utilization and minimizing cell losses. The congestion level is monitored through the buffer cell losses, buffer queue errors and a change in buffer length.

Define the feedback source rate u(k) as

$$u(k) = (1/T)(x_d(k+1) - \hat{f}(x(\cdot)) + k_v e(k)), \quad (8)$$

with a diagonal gain matrix k_v, and $\hat{f}(x(k))$ an estimate of the dynamics of the source/buffer, f(x(k)). Then, the closed-loop queue error dynamics becomes

$$e(k+1) = k_v e(k) + \tilde{f}(x(k)) + d(k) \quad (9)$$

where the estimation error in the network dynamics is given by

$$\tilde{f}(\cdot) = f(\cdot) - \hat{f}(\cdot). \quad (10)$$

Note that the queue error system is driven by the traffic estimation error and unknown disturbances.

In this paper, a ARMAX approach is used in discrete-time to provide the estimate of the cell arrival rate, $\hat{f}(\cdot)$ to each buffer. The error system (9) is used to focus on selecting parameter tuning algorithms that guarantee the QoS, which is defined here as cell losses $e(k)$.

3 Traffic Rate Controller Design

In this section, an ARMAX based adaptive controller is considered. In this case, the coefficients of the ARMAX model enter in a linear fashion. A novel parameter tuning method is derived to provide an accurate estimate of the network dynamics, which is then used in the adaptive controller. Stability analysis by Lyapunov's direct method is presented to show the convergence of the buffer error or cell losses. Assume, therefore, that there exist some constant ideal parameters θ for the ARMAX network traffic rate model so that the nonlinear traffic function in (4) can be written as

$$f(x) = \theta^T \varphi(x(k)) + \epsilon(k) \quad (11)$$

where $\varphi(x(k))$ is a vector of past values of buffer length or else it is referred to as regression matrix.

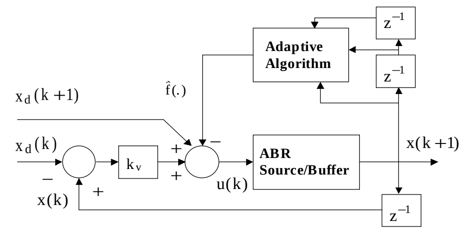

Figure 2: Adaptive ABR traffic rate controller structure.

3.1 Source/Buffer Error System Dynamics

Defining the estimate for the network dynamics in the controller (8) by

$$\hat{f}(x(k)) = \hat{\theta}^T(k)\varphi(x(k)) \quad (12)$$

with $\hat{\theta}(k)$ be the current value of the parameters, yields the controller structure shown in Figure 2. The output of the buffer system is processed through a series of delays to obtain the past values of the output,and fed as inputs to the ARMAX model so that the nonlinear traffic in (4)can be suitably approximated. The next step is to determine the parameter updates so that the performance of the closed-loop error dynamics of the buffer is guaranteed.

Let θ be the unknown but ideal parameters required for the approximation to hold in (12) and assume they are bounded by known values so that

$$\| \theta \| \leq \theta_{max}. \quad (13)$$

Then the error in the parameters during estimation is given by

$$\tilde{\theta}(k) = \theta - \hat{\theta}(k) \quad (14)$$

The traffic rate input u(k) is

$$u(k) = (1/T)(x_d(k+1) - \hat{\theta}^T(k)\varphi(x(k)) + k_v e(k)), \quad (15)$$

and the closed-loop dynamics become

$$e(k+1) = k_v e(k) + \bar{e}_i(k) + \epsilon(k) + d(k), \quad (16)$$

where the *identification error* for the network dynamics is defined by

$$\bar{e}_i(k) = \tilde{\theta}^T(k)\varphi(x(k)). \quad (17)$$

3.2 Parameter Tuning

A parameter tuning paradigm [2] based on the error in buffer length that guarantees the stability of the closed-loop system (16) is presented in this section. This algorithm guarantees that both the buffer length and error in parameter estimates are bounded.

Theorem 3.1 *Let the desired trajectory, $x_d(k)$, be finite and the network dynamics reconstruction error bound ϵ_N and the disturbance bound d_M be known constants. Consider the parameter tuning provided by*

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \alpha\varphi(x(k))e^T(k+1) \\ -\Gamma \parallel I - \alpha\varphi(x(k))\varphi^T(x(k)) \parallel \hat{\theta}(k) \quad (18)$$

with $\Gamma > 0$ a design parameter. Then the error in buffer length $e(k)$ and the parameter estimates $\hat{\theta}(k)$ are UUB, with the bounds specifically given by (29) and (31) provided:

$$(1) \quad \alpha \parallel \varphi(x(k)) \parallel^2 < 1, \quad (19)$$

$$(2) \quad 0 < \Gamma < 1, \quad (20)$$

with

$$(3) \quad k_{vmax} < \frac{1}{\sqrt{\bar{\sigma}}} \quad (21)$$

where $\bar{\sigma}$ is given by

$$\bar{\sigma} = \eta + \frac{1}{(1-\alpha \parallel \varphi(x(k)) \parallel^2)} [\Gamma^2(1-\alpha \parallel \varphi(x(k)) \parallel^2)^2 + 2\alpha\Gamma \parallel \varphi(x(k)) \parallel^2 (1-\alpha \parallel \varphi(x(k)) \parallel^2)], \quad (22)$$

with η being

$$\eta = \frac{1}{(1-\alpha \parallel \varphi(x(k)) \parallel^2)}. \quad (23)$$

Proof:
Define the Lyapunov function candidate

$$J = e^T(k)e(k) + \frac{1}{\alpha} tr(\tilde{\theta}^T(k)\tilde{\theta}(k)). \quad (24)$$

The first difference is given by

$$\triangle J = e^T(k+1)e(k+1) - e^T(k)e(k) + \frac{1}{\alpha} tr(\tilde{\theta}^T(k+1)\tilde{\theta}(k+1) - \tilde{\theta}^T(k)\tilde{\theta}(k)) \quad (25)$$

Use the buffer length error dynamics (16) and tuning mechanism (18) to obtain

$$\triangle J \leq -[1-\bar{\sigma}k_{vmax}^2] \parallel e(k) \parallel^2 \\ -[1-\alpha\varphi^T(x(k))\varphi(x(k))] \parallel \bar{e}_i(k) \\ -\frac{1}{(1-\alpha\varphi^T(x(k))\varphi(x(k))} (\alpha\varphi^T(x(k))\varphi(x(k)) + 2\Gamma \parallel I - \alpha\varphi(x(k))\varphi^T(x(k)) \parallel) (k_v e(k) + \epsilon(k) + d(k)) \parallel^2 + 2\gamma k_{vmax} \parallel e(k) \parallel + \rho - \frac{1}{\alpha} \parallel I - \alpha\varphi(x(k))\varphi^T(x(k)) \parallel^2 [\Gamma(2-\Gamma) \parallel \tilde{\theta(k)} \parallel \theta_{max} - \Gamma^2\theta_{max}^2], \quad (26)$$

where

$$\gamma = [\eta(\epsilon_N + d_M) + \Gamma(1-\alpha \parallel \varphi(x(k)) \parallel^2) \parallel \varphi(x(k)) \parallel \theta_{max}] \quad (27)$$

and

$$\rho = [\eta(\epsilon_N + d_M)^2 + 2\Gamma(1-\alpha \parallel \varphi(x(k)) \parallel^2) \parallel \varphi(x(k)) \parallel \theta_{max}(\epsilon_N + d_M)]. \quad (28)$$

Completing the squares for $\parallel \tilde{\theta}(k) \parallel$ in (26) results in $\triangle J \leq 0$ as long as the conditions in (19) through (21) are satisfied and with the upper bound on the buffer length error given by

$$\parallel e(k) \parallel > \frac{1}{(1-\bar{\sigma}k_{vmax}^2)}[\gamma k_{vmax} + \sqrt{\rho_1(1-\bar{\sigma}k_{vmax}^2)}] \quad (29)$$

where

$$\rho_1 = \rho + \frac{1}{\alpha}\frac{\Gamma}{(2-\Gamma)}(1-\alpha \parallel \varphi(x(k)) \parallel^2)^2\theta_{max}^2. \quad (30)$$

On the other hand, completing the squares for $\parallel e(k) \parallel$ in (26) results in $\triangle J \leq 0$ as long as the conditions(19)-(21) are satisfied and

$$\parallel \tilde{\theta}(k) \parallel > \frac{\Gamma(1-\Gamma)\theta_{max} + \sqrt{\Gamma^2(1-\Gamma)^2\theta_{max}^2 + \Gamma(2-\Gamma)\psi}}{\Gamma(2-\Gamma)}, \quad (31)$$

where

$$\psi = [\Gamma^2\theta_{max}^2 + \frac{\alpha\rho_1}{(1-\alpha \parallel \varphi(x(k)) \parallel^2)^2}], \quad (32)$$

and

$$\rho_1 = \rho + \frac{\gamma^2 k_{vmax}^2}{(1-\bar{\sigma}k_{vmax}^2)} \quad (33)$$

In general $\triangle J \leq 0$ as long as (19) and (21) are satisfied and either (29) or(31) holds. According to a standard Lyapunov extension theorem [2], this demonstrates that the buffer length or ABR cell arrival rate error and the error in parameter estimates are UUB.

■

Note that for practical purposes, (29) with (31) can be considered as bounds for $\| e(k) \|$ and $\| \tilde{\theta}(k) \|$. The next result discusses the behavior of the closed-loop system in the idealized case of no estimation errors in the network dynamics with no disturbances present in the dynamics of the system.

Theorem 3.2 *Let the desired buffer length $x_d(k)$ be finite and the network dynamic estimation error bound ϵ_N and the disturbance bound d_M be equal to zero. Let the source rate for (4) be given by (15) with the parameter tuning provided by*

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \alpha\varphi(x(k))e^T(k+1), \qquad (34)$$

where $\alpha > 0$ is a constant learning rate parameter or adaptation gain. Then the buffer length error e(k) approaches asymptotically to zero and the parameter estimates are bounded provided (19) with

$$(1) \qquad k_{vmax} < \frac{1}{\sqrt{\eta}} \qquad (35)$$

where η is given by

$$\eta = \frac{1}{(1 - \alpha \| \varphi(x(k)) \|^2)}. \qquad (36)$$

Proof: Since the functional reconstruction error and the disturbances are all zero, these new assumptions yield the error system

$$r(k+1) = k_v e(k) + \bar{e}_i(k) \qquad (37)$$

For the case of the parameter tuning mechanism given in (34), select the Lyapunov function candidate as (24), and use the new assumptions as well as the update law (34) to obtain

$$\begin{aligned} \triangle J \leq & -[1 - \eta k_{vmax}^2] \| e(k) \|^2 - \\ & [1 - \alpha \| \varphi(x(k)) \|^2] \\ & \| \bar{e}_i(k) - \frac{(\alpha \| \varphi(x(k)) \|^2)}{(1 - \alpha \| \varphi(x(k)) \|^2)} k_v e(k) \|^2 \end{aligned} \qquad (38)$$

where η is given in (36). Since $J > 0$ and $\triangle J \leq 0$, this shows stability in the sense of Lyapunov, provided the conditions in (19) and (35) hold so that $e(k)$ and $\tilde{\theta}(k)$ (and $\hat{\theta}(k)$) are bounded if $e(k_0)$ and $\tilde{\theta}(k_0)$ are bounded. Now using standard techniques [1-2], sum both sides of (38) to note that as $k \to \infty$, the tracking error $\| e(k) \| \to 0$. ∎

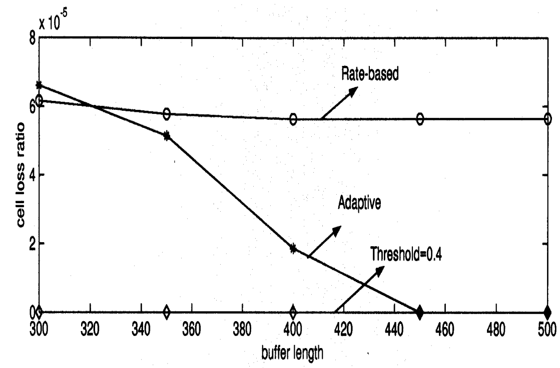

Figure 3: Cell loss ratio with no feedback delay.

4 Simulation Example

In our queue management simulation example, we drive four ABR sources with MPEG data, which cannot be adequately modeled using a conventional traffic model. Though MPEG is of VBR type traffic, here it is used as a bencmarking test data for our congestion controller as it is bursty and not well behaved. The PCRs for the sources are given by 6527, 8941, 9154, 4648 cells/sec respectively and the combined PCR rate for all the sources is 11,812 cells/sec with a MSR of 7317 cells/sec. The desired buffer length, x_d is selected to be 300, 350, 400 and so on. Cell losses are defined as $x(k) - x_d$. The cell loss ratio is defined as the total number of cells discarded at the receiver due to buffer overflows divided by the total number of cells generated at the source. Here transmission delay from the source to the receiver due to the connections and feedback delays from the controller to the source are assumed to be zero initially. Cells are removed from the buffer at a constant rate of 11,200 cells/second and the simulation is performed with a sampling interval of $1msec$. Source rates are adjusted from feedback $(u(k))$by assuming that individual sources have their own buffers and their transmission rates can be calculated accurately. The gain for the rate based control, k_v, is selected as 0.5. The adaptation gain is selected using a projection algorithm as $\alpha = \frac{0.5}{(0.1+\varphi(k)^T\varphi(k))}$. Past six values of the buffer length were used and hence the size of the unknown parameter vector is taken to be 6×1.

Figure 3 shows the cell loss ratio due to rate based, adaptive controller and thresholding method for different buffer length. From the simulation results, it is seen that the adaptive controller guarantees the performance with a cell loss ratio of 10^-5. As expected, as the buffer length is increased, the cell loss ratio decreases. Though thresholding approach (threshold of

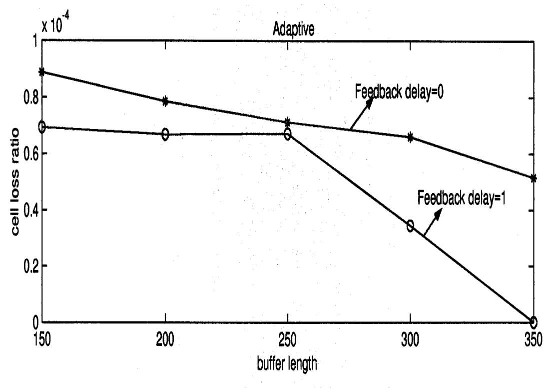

Figure 4: Cell loss ratio with feedback delays.

0.4) provides minimum cell losses, the performance cannot be guaranteed. From the results, it is clear that the adaptive controller outperforms the conventional rate based controller for multiple source scenario.

Figure 4 shows the cell loss ratio due to this adaptive controller when the feedback delays are injected. From the simulation results a cell loss ratio of 10^-4 was observed even when the delay is nonzero. Further, the closed-loop tracking error system is stable and the adaptive controller is robust against feedback delays equal to atleast 1 times the sampling interval.

5 Conclusions

This paper proposes an auto-regressive moving average (ARMAX)-based adaptive traffic controller for ATM networks. The ATM network traffic flow is modeled as a nonlinear function at a given switching node/buffer and an adaptive controller is designed to prevent network congestion. This approach estimates the traffic accumulated at the node based on current and past values of buffer length. A rigorous mathematical analysis is provided for the closed-loop system unlike in other available congestion control techniques. In fact, the proposed adaptive controller guarantees performance as shown through the Lyapunov analysis. In addition, the performance in terms of buffer length error can be reduced to arbitrarily small values by choosing appropriately selecting the gains, which in turn dictates the desirable source rates.

Simulation example is given to justify the theoretical conclusions. From the results, it was found that the cell loss ratio decreases with an increase in the buffer size. Further, the developed adaptive controller guarantees the closed-loop stability of the error system in the presence of feedback delays. Finally, the adaptive methodology provides an efficient queue management schemes for ATM networks. Future work will involve extending the adaptive controller for multiple source/multiple node scenario of an ATM network with propagation delays and in the presence of network disturbances.

References

[1] K. J. Åström and B. Wittenmark, *Adaptive control*, Addison-Wesley Company, Reading, Massachusetts, 1989.

[2] S. Jagannathan and F. L.Lewis, "Robust implicit self tuning regulator/MRAC convergence and Stability," *Automatica*, vol.32, December 1996.

[3] B. Chen, Y. Zhang, J. Yen, and W. Zhao, "Fuzzy adaptive connection admission control for real-time applications in ATM-based heterogeneous networks," *Journal of Intelligent and Fuzzy Systems*, vol.7, no.2, 1999.

[4] Steve Fuhrman, Yaakov Kogan, and Rodolfo A. Milito, "An adaptive autonomous network congestion controller," *Proc. of the IEEE Conference of decision and Control*, pp. 301-306, 1996.

[5] Ray-Guang Cheng, Chung-Ju Chang, and Li-Fong Lin, "A QoS-provisioning neural fuzzy connection admission controller for multimedia high-speed networks," *IEEE/ACM Transaction on Networking*, vol.7, no.1, pp. 111-121, 1999.

[6] A. Hiramatsu, "ATM communications network control by neural networks," *IEEE Transaction on Neural Networks*, vol.1, no.1, pp. 122-130, 1990.

[7] F. Bonomi and K. Fendick, "The rate-based flow control framework for the ABR ATM service," *IEEE Network*, vol.9, no.2, pp. 25-39, 1995.

[8] L. Benmohamed and S. M. Meerkov, "Feedback control of congestion in packet switching networks: The case of a single congested node," *IEEE/ACM Transaction on Networking*, vol.1, no.6, pp. 693-707, 1993.

[9] Yao-Ching Liu and C. Douligeris, "Static vs. adaptive feedback congestion controller for ATM networks," *IEEE INFOCOM*, pp. 291-295, 1995.