On Rate-Based Congestion Control in High Speed Networks: Design of an $\mathcal{H}^\infty$ Based Flow Controller for Single Bottleneck

Hitay Özbay
Department of Electrical Engineering
The Ohio State University
2015 Neil Avenue, Columbus, OH 43210, U.S.A.

Shivkumar Kalyanaraman
Department of Electrical Computer and Systems Eng.
Rensselaer Polytechnic Institute
110, 8th Street, Troy, NY 12180-3590, U.S.A.

Altuğ İftar
Department of Electrical and Electronics Eng.
Anadolu University
Eskişehir 26470, Turkey

Abstract

Typical congestion control algorithms for high speed networks include local flow controllers at the bottleneck nodes. In this paper an $\mathcal{H}^\infty$ based controller is developed for rate feedback in a single bottleneck network. The rates can be assigned to the sources only after a certain transmission delay. Controller design specifications for this time delay system include "fairness" to multiple users, "usage" optimization, and minimization of the transients in the queue length. Stability robustness, against uncertainties in time delays, is also specified as a design goal. By a simple algebra the problem is transformed to an $\mathcal{H}^\infty$ control of a plant with a time delay, and it is solved by using an algorithm developed earlier for this class of problems.

Key Words:
Data flow control, high speed networks, $\mathcal{H}^\infty$ control, time delay systems, robust control.

1 Introduction

Congestion is the problem which occurs when demand for a resource outstrips the capacity. In computer networks, congestion may occur in routers and switches which are shared by several sources. Note that "flow control" is a term used for the end-to-end control of a single flow (or multiple flows on a per-flow basis). Congestion control has been studied widely in computer networks and communications literature mainly for high speed data networks, [2, 3, 21, 24], and more recently for Asynchronous Transfer Mode (ATM) networks, [4, 10, 20, 25, 30] (these references are only samples from a huge literature on the subject, they are by no means complete).

There are two popular methods of controlling congestion. The "rate-based" method is the one where sources send data at a particular rate and adjusts its rate based upon network feedback. The "window based" control (or "credit based" control) has the sources sending no more than a window of unacknowledged packets; sources adjust window sizes based on network feedback.

The Available Bit Rate (ABR) Service in ATM networks uses a rate-based, end-to-end traffic management framework [1]. This framework has three main components: source end-systems, switches (or network nodes) and destination end-systems. Once every N packets (called "cells") the sources send a control cell which can be used by switches to convey feedback. The control cells travel to the destination and are returned to the source in the same path. Feedback signal may be in the form of a single bit or an explicit rate value, and can be written in the forward or reverse direction of travel of the control cell. In this framework the key elements of congestion control in a network are the flow controllers, at the bottleneck nodes, which determine the feedback signal. Indeed many of the papers cited above deal with this basic problem, see for example [2, 9, 20, 25, 30] and their references. For a survey of different traffic management schemes in ATM networks see [12]; for details of a specific scheme we refer to [11].

In this paper we consider a single bottleneck which receives data from n sources. The data flow rates from these sources are assumed to be generated by a feedback controller. Due to physical restrictions there are time delays in data flow between the sources and the bottleneck node. A feedback controller is to be designed for this time delay system. There are several controller design methods for different classes of systems with time delays, see for example [7, 8, 13, 14, 16, 17, 18, 19, 22, 23, 26] and their references. The techniques developed in [7, 28] are used in this paper. One of the design goals considered here is "fairness" to multiple users trying to send information through the same bottleneck node. That is, the minimum rate allocated to individual sources should be maximized. When sources are unconstrained (i.e., the demand is always greater or equal to the allocation), this reduces to simply equalizing the rates of sources. Another design objective is to maximize the "utilization," i.e. total data flow into the node should be kept close to the full "capacity" (maximum allowable flow rate for the data leaving the node) and the queue size should be kept close to a certain desired size. However, the most important design specification is *stability* of the feedback system. *Stability robustness*, with respect to uncertainties in the values of time delays in each flow path, is also desired.

In Section 2 of this paper a mathematical description of the system model is given. Then in

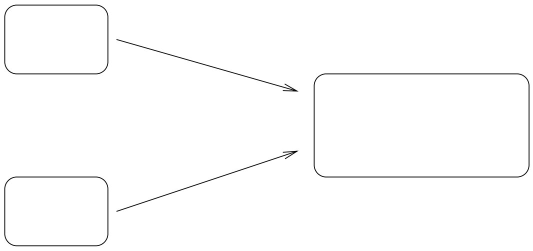

Figure 1: Flow control for n sources and a single bottleneck node

Section 3 feedback controller design problem is transformed to an $\mathcal{H}^\infty$ optimization problem for an unstable time delay system. The controller is derived from the procedure of [27] developed for this type of control problems. Structure of the controller is discussed in Section 4 via an example; implementation of the resulting controller is also illustrated. Concluding remarks are made in Section 5.

2 Mathematical Model

Consider a bottleneck node with n source connections, as shown in Figure 1. Let $q(t) \geq 0$ denote the queue length at the bottleneck node, and $r_i(t) \geq 0$ be the data flow rate at the ith source. The maximum rate at which the ith source can send data will be denoted by d_i. In other words, r_i is restricted to be $r_i(t) \leq d_i$ for all $i = 1, \ldots, n$. The rates $r_1(t), \ldots, r_n(t)$ will be assigned to the sources by a feedback controller which measures the queue length, $q(t)$, at the bottleneck node. The capacity is the rate at which data is sent out from the node; it is denoted by $c(t)$. A dynamical model for this system is given by

$$\dot{q}(t) = \sum_{i=1}^{n} r_i(t - \tau_i) - c(t) \tag{1}$$

where τ_i is the time delay from the ith source to the bottleneck node. This time delay is equal to the amount of time it takes for the feedback control signal to reach the source plus the amount of time it takes for the data to reach the node after it is sent from the source. A block diagram of the feedback control system is shown in Figure 2.

The utilization is defined to be the quantity

$$\rho(t) := \min \left\{ \frac{\sum_{i=1}^{n} r_i(t - \tau_i)}{c(t)} \; , \; 1 \right\}.$$

One of the objectives in high speed networks is to keep the utilization as close to 1 as possible. The controller should also achieve "fairness" in steady state, i.e. the rates allocated to different sources should be as close to each other as possible. A trivial choice is then to select equal rates for the sources, with a steady state value c/n, where $c = \lim_{t\to\infty} c(t)$ which is assumed to be non-zero.

In the frequency domain the feedback system can be represented by transfer functions as shown in Figure 3, where q_d represents the "desired" queue length at the bottleneck node, $P(s)$ is the underlying

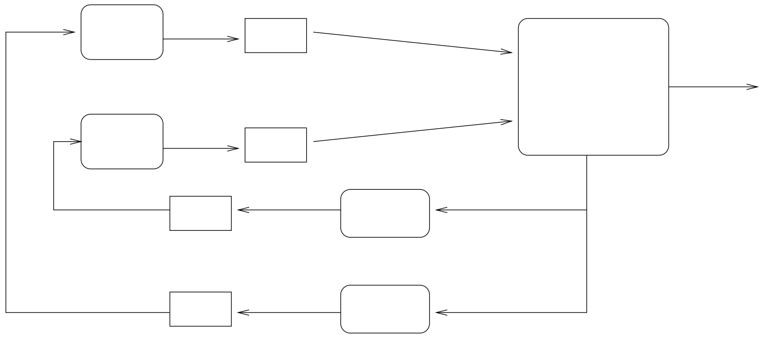

Figure 2: Feedback control system.

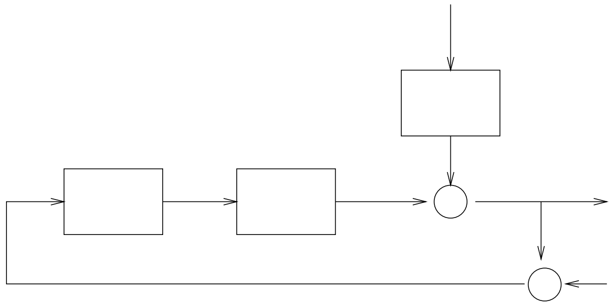

Figure 3: Feedback control system with MISO plant.

plant and $K(s)$ is the controller.

Clearly, the plant is a Multi Input Single Output (MISO) system whose transfer function is in the form

$$P(s) = \frac{1}{s}\,[e^{-\tau_1 s}\ ,\ \ldots\ ,\ e^{-\tau_n s}]\ .$$

Similarly, the controller is an $n \times 1$ system with transfer function

$$K(s) = [K_1(s)\ ,\ \ldots\ ,\ K_n(s)]^{\mathsf{T}}.$$

In general time delays $\tau_1, \ldots, \tau_n$ are unknown (in fact they might be time varying, but here they will be assumed to be fixed). Let $h_i \geq 0$ be a given lower bound of τ_i, and define

$$\delta_i := (\tau_i - h_i).$$

For controller design, let $\delta_i^+ \geq 0$ be a known estimate of the size of the uncertainty, i.e. $\delta_i \in [0\ ,\ \delta_i^+)$.

Define $P_o(s)$ to be the nominal plant

$$P_o(s) = \frac{1}{s}\,[e^{-h_1 s}\ ,\ \ldots\ ,\ e^{-h_n s}]\ .$$

The following is a coprime factorization in $\mathcal{H}^\infty$

$$P_o(s) = D(s)^{-1} N_o(s) = N_o(s) D(s)^{-1},$$

where $D(s) = s(s+\alpha)^{-1}$ and

$$N_o(s) = \frac{1}{s+\alpha} \; [e^{-h_1 s} \; , \; \ldots \; , \; e^{-h_n s}] \; ,$$

for any $\alpha > 0$. The uncertain plant can be represented as

$$P(s) = (N_o(s) + \Delta_N(s))D(s)^{-1}$$

with

$$\Delta_N(s) = \frac{1}{s+\alpha} \; [(e^{-\delta_1 s} - 1) \; , \; \ldots \; , \; (e^{-\delta_n s} - 1)] \; \mathrm{diag}(e^{-h_1 s}, \ldots, e^{-h_n s})$$

where $\mathrm{diag}(e^{-h_1 s}, \ldots, e^{-h_n s})$ denotes the diagonal matrix whose entries are $e^{-h_1 s}, \ldots, e^{-h_n s}$.

In the next section the simplified model developed above will be considered for feedback control. It should be mentioned that several different control schemes have been determined in [20, 25, 30] for similar models. In this paper robustness to uncertainty in the time delays will be taken into account by putting the problem into the framework of $\mathcal{H}^\infty$ control.

3 Feedback Controller Design

3.1 Stabilization

For the nominal plant, P_o, given above with the particular coprime factorization, all stabilizing controllers are parameterized as (see for example the recent books [7, 31] and their references)

$$K(s) = (X(s) + D(s)Q(s))(Y(s) - N_o(s)Q(s))^{-1} \tag{2}$$

where Q is the free $n \times 1$ vector of $\mathcal{H}^\infty$ functions, and stable X, Y satisfy

$$N_o X + DY = 1 \quad \text{ or } \quad Y = \frac{1 - N_o X}{D}.$$

Note that X is $n \times 1$ and Y is scalar. Moreover, since $D(s) = s/(s+\alpha)$, $X(s)$ must satisfy

$$\alpha = [1 \; \cdots \; 1] \; X(0)$$

which has infinitely many solutions for $X(0)$.

3.2 Robust stability

Using standard results of robust control theory it is a simple exercise to show that a controller $K(s)$ is robustly stabilizing (i.e. it stabilizes all plants of the form $P(s)$) if it stabilizes $P_o(s)$ and satisfies

$$\|WK(1 + P_o K)^{-1}\|_\infty \leq 1 \tag{3}$$

where $W = [w_1 \ , \ \ldots \ , \ w_n]$, with $w_i, w_i^{-1} \in \mathcal{H}^\infty$ and

$$|w_i(j\omega)| > \left| \frac{e^{-j\delta_i \omega} - 1}{j\omega} \right| \quad \text{for all} \ \ \omega, \quad \text{and} \quad \delta_i \in [0 \ , \ \delta_i^+). \tag{4}$$

For a given δ_i^+ it is easy to determine a low order weight $w_i(s)$ such that the conservatism introduced in (4) is very small, see e.g. [15].

In terms of the free parameter Q, stability robustness is guaranteed by

$$\|WD(X + DQ)\|_\infty \leq 1. \tag{5}$$

3.3 Performance issues

One of the performance objectives of rate-based congestion control is to keep the queue length, $q(t)$, as close to its desired value, $q_d(t)$, as possible. Consider

$$q_d(s) = \frac{1}{s} \widehat{q}(s)$$

where $\widehat{q}$ is an arbitrary bounded energy signal (for example if $\widehat{q}$ is a pulse of finite duration, then q_d is a saturating ramp signal). Then the tracking error, expressed in the frequency domain, is

$$E(s) = q_d(s) - q(s) = S_o(s) \ \frac{1}{s} \ (\widehat{q}(s) + c(s)),$$

where $S_o(s) = (1 + P_o(s)K(s))^{-1}$. Since K is a stabilizing controller, and P_o has a pole at $s = 0$, we have that $S_o(0) = 0$, and by the final value theorem the steady state value of the error is

$$e_{ss} = \lim_{s \to 0} sE(s) = \lim_{s \to 0} ([e^{-h_1 s}, \ldots, e^{-h_n s}]K(s))^{-1} s(\widehat{q}(s) + c(s)).$$

The signal $\widehat{q}(t)$ is assumed to have finite energy, so its final value is zero, and hence

$$e_{ss} = ([1, \ldots, 1]K(0))^{-1} c.$$

Thus, at least one of the entries of the controller must have a pole at $s = 0$, in order to have zero steady state error.

Also note that the rate feedback signals are given (in the frequency domain) by

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_n(s) \end{bmatrix} = K(s)(1 + P_o(s)K(s))^{-1} \frac{1}{s} \ (\widehat{q}(s) + c(s)),$$

and recall that for "fairness" it is desired to have the steady-state values of $r_1, \ldots, r_n$ equal to c/n. This means that the entries, $K_i(s)$, of the controller $K(s)$ must satisfy

$$\lim_{s \to 0} K_i(s)([e^{-h_1 s}, \ldots, e^{-h_n s}]K(s))^{-1} = \frac{1}{n} \quad \text{for all} \ \ i = 1, \ldots, n.$$

In conclusion, for good steady state performance the controller must be in the form $K(s) = K_o(s)D(s)^{-1}$ with $K_o(0) = [1,\dots,1]^{\mathrm{T}}K_{o\nu}(0)$ for some scalar $K_{o\nu}(0) \neq 0$.

Now considering the parameterization (2), a desired form of the controller is obtained if $Q \in \mathcal{H}^\infty$ is such that

$$Y(0) = N_o(0)Q(0) \Longleftrightarrow \lim_{s\to 0}\frac{\partial}{\partial s}((s+\alpha) - [e^{-h_1 s},\dots,e^{-h_n s}]X(s)) = \frac{1}{\alpha}\,[1,\dots,1]Q(0) \tag{6}$$

and

$$\begin{aligned} K_o(0) &= \lim_{s\to 0}\frac{D(s)}{(Y(s)-N_o(s)Q(s))}(X(s)+D(s)Q(s)) \\ &= \lim_{s\to 0}\frac{D(s)}{(Y(s)-N_o(s)Q(s))}X(0) = [1,\dots,1]^{\mathrm{T}}K_{o\nu}(0). \end{aligned}$$

By the above discussion $X \in \mathcal{H}^\infty$ must be chosen in such a way that

$$X(0) = [1,\dots,1]^{\mathrm{T}}\frac{\alpha}{n},$$

otherwise $X(s)$ is free. The following choice for $X(s)$ simplifies the analysis:

$$X(s) = \frac{\alpha}{n}F(s), \quad \text{where} \quad F(s) := [e^{-(h-h_1)s},\dots,e^{-(h-h_1)s}]^{\mathrm{T}} \quad h = \max\{h_1,\dots,h_n\}.$$

In this case $Y(s)$ becomes

$$Y(s) = 1 + \frac{\alpha}{s}(1-e^{-hs})$$

and $Y(0) = (1+\alpha h)$. Thus for (6) to hold the free parameter $Q(s)$ should satisfy

$$\alpha(1+\alpha h) = [1,\dots,1]Q(0)\ . \tag{7}$$

In summary, with the above choices of $X(s)$ and $Y(s)$, the steady state performance requirements are met if the controller is $K = (X+DQ)(Y-N_oQ)^{-1}$, where $Q(s)$ is a free stable $n\times 1$ vector satisfying (7). Clearly, a simple admissible Q is

$$Q_o(s) = F(s)\frac{\alpha(1+\alpha h)}{n}Q_1(s) \tag{8}$$

where $Q_1(s)$ is an arbitrary stable transfer function with $Q_1(0) = 1$. In this case the controller becomes

$$K_o(s) = F(s)\ \frac{1}{n}\ \frac{G_o(s)}{1-\frac{1}{s}e^{-hs}G_o(s)} \quad \text{where} \quad G_o(s) = \frac{\alpha s}{s+\alpha}\left(1+\frac{s}{s+\alpha}(1+\alpha h)Q_1(s)\right). \tag{9}$$

Selection of the free parameter $Q_1(s)$ will be discussed in the following sections.

Besides the steady-state behavior, it is also desired to control the transient response of the system. For this purpose the $\mathcal{H}^\infty$ norm of the weighted sensitivity function can be minimized (this corresponds to worst energy minimization for the tracking error, see for example [5] for a detailed discussion). More precisely, the problem is to

$$\text{minimize} \quad \|W_s(1+P_oK)^{-1}\|_\infty \tag{10}$$

over all controllers $K(s)$ stabilizing $P_o(s)$, where $W_s(s)$ is the sensitivity weighting filter. Since external signals $c(t)$ and $q_d(t)$ are "step-like functions" an appropriate weight is $W_s(s) = 1/s$. By the controller parameterization given above the problem (10) is equivalent to

$$\text{minimize} \quad \|V(Y - N_oQ)\|_\infty \quad : \quad Q \in \mathcal{H}^\infty. \tag{11}$$

where $V(s) := W_s(s)D(s) = (s+\alpha)^{-1}$.

3.4 Special optimization problems

For the solution of (11) under the condition (5) there is significant amount of degree of freedom in Q. There are several ways to pose optimal control problems which determine the free parameter Q (and hence the controller K) uniquely. Some of these problems are discussed below.

Problem 1. Time delays in the feedback paths between different sources and the bottleneck node can be equalized by the controller by choosing

$$K(s) = F(s)K_\nu(s)$$

where $K_\nu(s)$ is a scalar controller, and $h = \max\{h_1, \ldots, h_n\}$, (note that this is similar to the controller structure (9)). Then the sensitivity function is $S_o(s) = (1 + P_\nu(s)K_\nu(s))^{-1}$ where $P_\nu(s) = \frac{n}{s}e^{-hs}$, and the controller $K_\nu(s)$ must be a stabilizing controller for the new SISO plant $P_\nu(s)$. In the controller parameterization discussed in Section 3.1, this structural choice corresponds to having $X(s) = F(s)X_\nu(s)$ and $Q(s) = F(s)Q_\nu(s)$, with $X_\nu \in \mathcal{H}^\infty$ satisfying $X_\nu(0) = 1/N_\nu(0)$, where $N_\nu(s) = ne^{-hs}/(s+\alpha)$, and $Q_\nu \in \mathcal{H}^\infty$ is the free parameter.

Recall that by (7) for good steady state performance it is desired to have $Q_\nu(s) = \frac{\alpha(1+\alpha h)}{n}Q_1(s)$, where $Q_1 \in \mathcal{H}^\infty$ satisfies $Q_1(0) = 1$. Let us fix $Q_1(s) = 1$ and examine the stability robustness inequality for this particular choice of the free parameter:

$$\|WK(1+P_\nu K)^{-1}\|_\infty \le 1, \tag{12}$$

where

$$K(s) = F(s)\,\frac{1}{n}\,\left(\frac{G_o(s)}{1 - \frac{e^{-hs}}{s}G_o(s)}\right), \qquad G_o(s) = \frac{\alpha s}{s+\alpha}\left(1 + \frac{s}{s+\alpha}(1+\alpha h)\right).$$

For this controller structure, (12) is equivalent to

$$\|W_2(s)\frac{1}{s}G_o(s)\|_\infty \le 1 \tag{13}$$

where $W_2(s)$ is a new uncertainty weight determined from

$$|W_2(j\omega)| \geq \frac{1}{n}\left|\sum_{i=0}^{n}|e^{-j\delta_i\omega}-1|e^{-j(h-h_i)\omega}\right|, \tag{14}$$

for all ω, and all $\delta_i \in [0\ ,\ \delta_i^+)$, see e.g. [15]. Let δ_k^+ be the maximum of $\{\delta_1^+,\ldots,\delta_n^+\}$. Then a simple, yet conservative, form of W_2 is

$$W_2(s) = \frac{\varepsilon + (2+\varepsilon)\tau s}{1+\tau s}, \tag{15}$$

where $\varepsilon > 0$ and $\tau > 0$ are determined from

$$|W_2(j\omega)| > |e^{-j\delta_k\omega}-1|$$

for all ω, and all $\delta_k \in [0\ ,\ \delta_k^+)$. Clearly, (13) holds if

$$\left\|\varepsilon\left(\frac{1+2\tau(1+1/\varepsilon)s}{1+\tau s}\right)\left(\frac{1}{1+s/\alpha}\right)\left(1+\frac{(1+h\alpha)}{\alpha}\ \frac{s}{1+s/\alpha}\right)\right\|_\infty \leq 1. \tag{16}$$

If we choose α such that

$$\frac{1}{\alpha} = 2\tau\left(1+\frac{1}{\varepsilon}\right)$$

then (16) holds if

$$\varepsilon\left(2+\frac{h}{2\tau(1+1/\varepsilon)}\right) \leq 1.$$

Hence, if $\varepsilon > 0$ is sufficiently small, then this controller guarantees stability robustness in the presence of uncertainties in time delays.

In the above discussion Q_1 was fixed. Now let us leave Q_1 free, and define a performance optimization problem as follows: minimize $\gamma > 0$ such that K_o is stabilizing P_o and

$$\left\|\begin{bmatrix}\gamma^{-1}W_s(1+P_oK_o)^{-1}\\ WK_o(1+P_oK_o)^{-1}\end{bmatrix}\right\|_\infty \leq 1. \tag{17}$$

Considering the controller structure (9) the problem (17) can be expressed as minimizing $\gamma > 0$ in

$$\left\|\begin{bmatrix}\frac{1}{\gamma s}(1-e^{-hs}\frac{1}{s}G_o(s))\\ W_2(s)\ \frac{1}{s}\ G_o(s)\end{bmatrix}\right\|_\infty \leq 1, \tag{18}$$

over all stable $Q_1(s)$, with $Q_1(0) = 1$. For notational convenience define

$$Q_2(s) := \frac{1}{s}G_o(s) = \left(\frac{1}{1+s/\alpha}\right)\left(1+\frac{s}{s+\alpha}(1+\alpha h)\right). \tag{19}$$

So, in terms of Q_2, the free parameter Q_1 can be written as

$$Q_1(s) = \left(\left(\frac{1+s/\alpha}{1+\epsilon s} \right) Q_2(s) - 1 \right) \frac{(s+\alpha)}{(s+\epsilon)(1+\alpha h)} \qquad \epsilon \to 0.$$

If we relax the condition $Q_1(0) = 1$ (which corresponds to steady state performance objective), the solution of (18) can be approximated by the solution of the following problem: minimize $\gamma > 0$ over all $Q_2 \in \mathcal{H}^\infty$ satisfying

$$\left\| \begin{bmatrix} \frac{1}{\gamma(s+\epsilon)}(1 - e^{-hs} Q_2(s)) \\ W_2(s) Q_2(s) \end{bmatrix} \right\|_\infty \leq 1, \tag{20}$$

with $\epsilon > 0$ and $\epsilon \to 0$, and $W_2(s)$ is as defined above. Note that the problem (20) is equivalent to a mixed sensitivity minimization problem for the SISO plant

$$P_2(s) := e^{-hs} W_2^{-1}(s), \tag{21}$$

which can be solved using certain techniques from operator theory. See [6, 7, 28] and their references for explicit computations of the optimal controller. In Section 4 implementation of the controller obtained from (20) will be discussed in detail.

Problem 2. Let $D_\varepsilon(s) = \frac{s+\varepsilon}{s+\alpha}$, where $\varepsilon > 0$ is a small real number to be determined shortly. Using a change of variable

$$Q = -D_\varepsilon^{-1}(X + \widehat{Q}) \tag{22}$$

the condition (5) can be re-written as

$$\|WD(1 - DD_\varepsilon^{-1})X - WDDD_\varepsilon^{-1}\widehat{Q}\|_\infty \leq 1.$$

Note that $\varepsilon > 0$ can be chosen such that

$$\epsilon_x := \|WD(1 - DD_\varepsilon^{-1})X\|_\infty \ll 1.$$

Under these conditions stability robustness is guaranteed by

$$\|WDDD_\varepsilon^{-1}\widehat{Q}\|_\infty \leq (1 - \epsilon_x) \ . \tag{23}$$

Another design objective which reduces the problem to a scalar optimization is to equalize the contribution of each control channel to the robustness inequality (23). In this case the free parameter $\widehat{Q}$ is chosen as

$$\widehat{Q} = \begin{bmatrix} w_1^{-1} \\ \vdots \\ w_n^{-1} \end{bmatrix} \widetilde{Q} \tag{24}$$

where $\widetilde{Q} \in \mathcal{H}^\infty$ is the new scalar free parameter. In terms of the new variable, $\widetilde{Q}$, stability robustness is guaranteed if

$$\|D^2 D_\varepsilon^{-1} \widetilde{Q}\|_\infty \leq \frac{(1-\epsilon_x)}{\sqrt{n}}. \tag{25}$$

Then the following scalar optimization problem can be posed: minimize $\gamma > 0$ over all $\widetilde{Q} \in \mathcal{H}^\infty$ satisfying

$$\left\| \begin{bmatrix} \gamma^{-1} V(\frac{1-N_o X(1-DD_\varepsilon^{-1})}{D} + \widehat{N} D_\varepsilon^{-1} \widetilde{Q}) \\ \frac{\sqrt{n}}{(1-\epsilon_x)} D^2 D_\varepsilon^{-1} \widetilde{Q} \end{bmatrix} \right\|_\infty \leq 1, \tag{26}$$

where $\widehat{N} = N_o[w_1^{-1}, \ldots, w_n^{-1}]^{\mathrm{T}}$. The problem (26) can also be solved using results from operator theory. The key difficulty here is to find an inner-outer factorization of $\widehat{N}$, which can be seen as a parallel connection of n stable time delay systems. Clearly, if time delays are equal in each channel, then this problem is as easy as the first problem (17). When the delays are commensurate, i.e. $h_k = N_k h_o$ for some h_o and integers N_k, $k = 1, \ldots, n$, then finding inner outer factorizations amounts to finding right half plane zeros of a quasi-polynomial, which is a feasible problem. On the other hand, when delays are arbitrary the problem is difficult, see references in [28, 29]. Nevertheless, since h_k's are *nominal* values of time delays, they can be chosen as commensurate, without introducing too much conservatism.

Problem 3. In order to find an appropriate controller the following multivariable optimization problem can also be posed; though in this case the solution may not be unique. Recall that the rate feedback signals (in the frequency domain) are

$$\begin{bmatrix} r_1(s) \\ \vdots \\ r_n(s) \end{bmatrix} = K(s)(1 + P_o(s)K(s))^{-1} \frac{1}{s} (\widehat{q}(s) + c(s))$$

and it is assumed that $|r_k(t)| \leq d_k$, where d_k is the "demand" corresponding to the kth source. So, in order to put a penalty on the rates, another design objective can be specified: the controller $K(s)$ should be such that

$$\|W_3 K (1 + P_o K)^{-1}\|_\infty \leq 1,$$

where W_3 is the diagonal weighting matrix $W_3 = \mathrm{diag}(d_1^{-1}, \ldots, d_n^{-1})$. Then a new optimization problem can be posed: minimize $\gamma > 0$ over all $Q \in \mathcal{H}^\infty$ satisfying

$$\left\| \begin{bmatrix} \gamma^{-1} V(1 - N_o(X + DQ)) \\ WD(X + DQ) \\ W_3 D(X + DQ) \end{bmatrix} \right\|_\infty \leq 1.$$

The above problem is a multivariable two-block $\mathcal{H}^\infty$ problem. The key step in its solution is the inner-outer factorization of the $(n+2)\times n$ stable transfer matrix

$$\begin{bmatrix} \gamma^{-1}VN_o \\ WD \\ W_3D \end{bmatrix} .$$

In general it is not possible to factor out a common time delay term from the first row of this matrix, so the problem is non-trivial. However, once the inner-outer factorization is performed the methods of [7] (Chapter 8) can be used to solve the optimization problem.

4 Controller Structure and Implementation

In this section we discuss implementation of the controller obtained from (20). By applying the formulae given in [27], the optimal controller (denoted by $K_1(s)$ to distinguish it from $K_o(s)$ with $Q_1(s) = 1$), can be easily computed. It is determined from the problem data $P_o(s)$ (the plant), $W_s(s) = 1/s$ (the sensitivity weight), and $W_2(s)$, (the robustness weight, defined above). After several steps, involving algebraic simplifications and substitutions (which are ommited here for brevity), the controller can be found as

$$K_1(s) := F(s)\ \frac{1}{n}\ \frac{R_\gamma(s)}{1-e^{-hs}R_\gamma(s)}\left(\frac{1}{\gamma^2 s}+s\right) \tag{27}$$

where

$$R_\gamma(s) = \frac{\gamma}{\sqrt{1-\varepsilon^2}}\frac{s(1+\tau s)}{(1+b_\gamma s + a_\gamma s^2)}$$

$$a_\gamma = \frac{(2+\varepsilon)\gamma\tau}{\sqrt{1-\varepsilon^2}}\ , \qquad b_\gamma = \sqrt{\frac{\tau^2+\gamma^2\varepsilon^2+2(2+\varepsilon)\gamma\tau\sqrt{1-\varepsilon^2}-(2+\varepsilon)^2\tau^2}{1-\varepsilon^2}}$$

and $\gamma > 0$ is the smallest positive solution of

$$e^{jh/\gamma} = R_\gamma(j/\gamma).$$

Note that the term $(1+\gamma^2 s^2)$ appearing in the numerator of $K_1(s)$ gets cancelled by the roots of the denominator, $(1-e^{-hs}R_\gamma(s))$, at $\pm j\frac{1}{\gamma}$. It is a simple exercise to show that γ is the unique solution of

$$f(\gamma) = 0 \quad \text{where} \quad f(\gamma) = \frac{h}{\gamma} - \frac{\pi}{2} + \mathrm{Tan}^{-1}\left(\frac{\gamma b_\gamma}{\gamma^2 - a_\gamma}\right) - \mathrm{Tan}^{-1}\left(\frac{\tau}{\gamma}\right) \tag{28}$$

Therefore, if h increases then so does γ, which means that the performance gets worse. The feedback diagram illustrating a possible implementation of this controller is shown in Figure 4. Recall that in

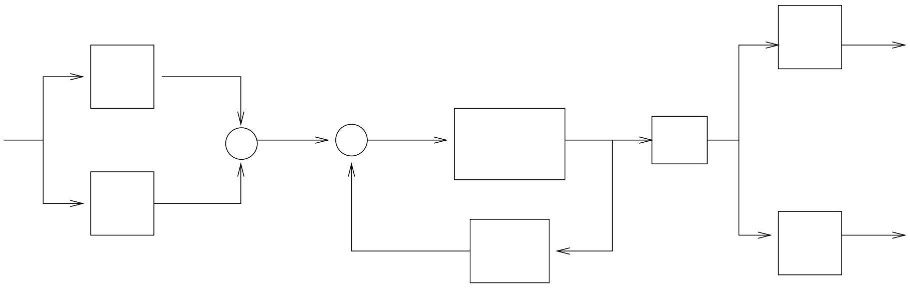

Figure 4: Controller implementation.

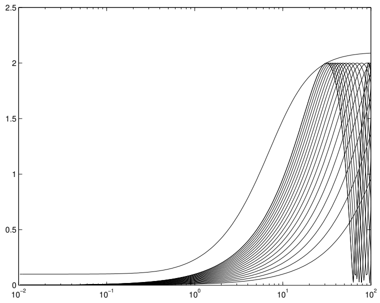

Figure 5: Uncertainty bound.

the above formulae ε and τ are the parameters of $W_2(s)$, which determines a bound on the time delay uncertainties, and h is the maximum nominal delay between the sources and the bottleneck node.

Numerical Example: Let us pick $h = 0.5$sec and $\delta^+ = 0.1$sec, i.e. there is 20% uncertainty in the time delay. Then, $W_2(s)$, with $\varepsilon = 0.1$ and $\tau = 0.1$, is a valid uncertainty weight as shown in Figure 5. The optimal performance level γ_o can be found from the plot of $f(\gamma)$ versus γ. This is shown in Figure 6, which indicates that $\gamma = 0.148$ is the optimal value. Once γ is determined, $R_\gamma(s)$ can be computed as

$$R_\gamma(s) = \frac{0.15s(1+0.1s)}{(1+0.17s+0.03s^2)}.$$

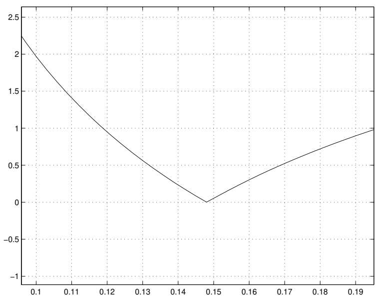

Figure 6: $|f(\gamma)|$ versus γ.

For the controller $K_o(s)$, (9), with any stable $Q_1(s)$ having $Q_1(0) = 1$, the DC gain is infinite ($K_o(s)$ contains a pole at $s = 0$). On the other hand, $K_1(0) = 1/\gamma$. So, unless $\gamma > 0$ can be made arbitrarily small, the controller $K_1(s)$ will not contain a pole at $s = 0$. If this is an undesirable situation, the sensitivity weight should be modified to $W_s(s) = 1/s^2$ to force the controller to have a pole at $s = 0$. Such a modification would slightly complicate the controller computations. Indeed, the above formula for $K_1(s)$ is obtained by hand calculations; when the weights are of high order one needs to use a computer. It is also not so easy to give an explicit "one line" formulae for the controllers that appear in the solution of Problems 2 and 3. However, once the problem data is specified, there are numerically feasible techniques, [7], that can be used to find the controller.

In the full version of the paper the controllers K_o and K_1 will be compared to the controllers obtained using Smith predictor, [20], and H^2 optimization, [30], via realtime simulations.

5 Conclusions

In this paper several different schemes have been proposed for $\mathcal{H}^\infty$ based flow controllers for a single bottleneck node. The controllers K_o and K_1, given explicitly in this paper, are designed to equalize the time delays in each feedback channel. This way non-uniqueness problem has been avoided. Another way to circumvent this problem is to assign a weight to each channel so that effects of the time delay uncertainties (on a measure of stability robustness) are equalized. Once an inner outer factorization of the resulting plant $\widehat{N}D^{-1}$ is determined, the controller can be found from the similar methods that are used to find $K_1(s)$. The third problem we have proposed is a multivariable $\mathcal{H}^\infty$ optimization.

In the full version of the paper, performances of the controllers determined here (and obtained elsewhere) will be compared by real time simulations on a typical traffic network.

References

[1] ATM Forum Traffic Management AF-TM-0056.000, "The ATM Forum Traffic Management Specification Version 4.0," April 1996.
Available as `ftp://ftp.atmforum.com/pub/approved-specs/af-tm-0056.000.ps`

[2] Benmohamed, L., and 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 (1993), pp. 693–707.

[3] Bolot, J-C., and A U. Shankar, "Dynamical behavior of rate-based flow control mechanisms," *Computer Communication Review*, vol. 20 (1990), No. 2, pp. 35–49.

[4] Bonomi, F., and K. W. Fendick, "The rate-based flow control framework for the available bit rate ATM service," *IEEE Network*, March/April 1995, pp. 25–39.

[5] Doyle, J., B. Francis and A. Tannenbaum, *Feedback Control Theory*, McMillan, New York, NY, 1992.

[6] Dym, H., T. T. Georgiou, and M. C. Smith, "Explicit formulas for optimally robust controllers for delay systems," *IEEE Transactions on Automatic Control*, vol. 40 (1995), pp. 656–669.

[7] Foias, C., H. Özbay, A. Tannenbaum, *Robust Control of Infinite Dimensional Systems: Frequency Domain Methods*, Lecture Notes in Control and Information Sciences, No. 209, Springer-Verlag, London, 1996.

[8] Fu, M., A. Olbrot, M. Polis, "Edge theorem and graphical tests for robust stability of neutral time delay systems," *Automatica*, vol. 27 (1991), pp. 739–741.

[9] Jaffe, J., "Bottleneck flow control," *IEEE Trans. on Communications*, vol. 29 (1981), pp. 954–962.

[10] Jain, R., "Congestion control and traffic management in ATM networks: recent advances and a survey," *Computer Networks and ISDN Systems*, vol. 28 (1996), pp. 1723–1738.

[11] Jain, R., S. Kalyanaraman, R. Goyal, S. Fahmy, and R. Viswanathan, "The ERICA scheme for traffic management in ATM networks" submitted for publication in *IEEE Transactions on Networking*, 1997. Available from `http://www.cis.ohio-state.edu/ jain`.

[12] Kalyanaraman, S., *Traffic Management for the Available Bit Rate (ABR) Service in Asynchronous Transfer Mode (ATM) Networks*, Ph.D. Dissertation, Department of Computer and Information Sciences, The Ohio State University, 1997.

[13] Kojima A., K. Uchida, and E. Shimemura, "Robust stabilization of uncertain time delay systems via combined internal-external approach," *IEEE Transactions on Automatic Control*, vol. 38 (1993), pp. 373–378.

[14] Kwon, W., and A. E. Pearson, "Feedback stabilization of linear systems with delayed control," *IEEE Trans. on Automatic Control*, vol. 25, no. 2, pp. 266-269, 1980.

[15] Lam, J., "Convergence of a class of Padé approximations for delay systems," *Int. J. Control*, vol. 52 (1990), pp. 989–1008.

[16] Lee, E. B., and S. Zak, "On spectrum placement for linear time-invariant delay systems," *IEEE Trans. on Automatic Control*, vol. 27, no. 2, pp. 446-449, 1982.

[17] Lehman, B., J. Bentsman, S. V. Lunel, E. I. Verriest, "Vibrational control of nonlinear time lag systems with bounded delay: Averaging theory, stabilizability, and transient behavior," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp.898–912.

[18] Li, X., and C. E. de Souza, "LMI approach to delay-dependent robust stability and stabilization of uncertain linear delay systems," *Proc. of 34th IEEE Conference on Decison and Control*, New Orleans LA, 1995, pp. 3614–3619.

[19] Marshall, J. E., *Time Delay Systems*, Ellis Horwood (series in mathematics and its applications), New York , 1992.

[20] Mascolo, S., D. Cavendish, and M. Gerla, "ATM rate based congestion control using a Smith predictor: an EPRCA implementation," *Proc. of IEEE INFOCOM '96*, San Fancisco CA, March 1996, vol. 5, pp. 569–576.

[21] Mitra, D., "Asymptotically optimal design of congestion control for high speed data networks," *IEEE Trans. on Communications*, vol. 40 (1992), pp. 301–311.

[22] Niculescu, S-I., J. M. Dion, L. Dugard, "Robust stabilization for uncertain time delay systems containing saturating actuators," *IEEE Transactions on Automatic Control*, vol. 41 (1996), pp. 742–746.

[23] Olbrot, A., "Control of equilibrium of linear delay-differential systems," *IEEE Trans. on Automatic Control*, vol. 28, no. 4, pp. 521-523, 1983.

[24] Parekh, A. K., and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the multiple node case," *IEEE/ACM Trans. on Networking*, vol. 2 (1994), pp. 137–150.

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

[26] Stepan, G., *Retarded dynamical systems : stability and characteristic functions*, Longman Scientific & Technical ; New York : Wiley, 1989.

[27] Toker, O., and H. Özbay, "$\mathcal{H}^\infty$ Optimal and suboptimal controllers for infinite dimensional SISO plants," *IEEE Transactions on Automatic Control*, vol. 40 (1995), pp. 751–755.

[28] Toker, O., and H. Özbay, "Complexity issues in robust stability of linear delay-differential systems," *Mathematics of Control, Signals, and Systems*, vol. 9, pp. 386–400, 1996.

[29] Ulus, C., *Control and Approximation of Time Delay Systems*, M.S. Thesis, Department of Electrical Eng., The Ohio State University, 1994.

[30] Zhao, Y., S.Q. Li, and S. Sigarto, "A linear dynamic model for design of stable explicit-rate ABR control schemes," *Proc. of INFOCOM '97*, Kobe Japan, 1997.

[31] Zhou, K., Doyle, J.C., Glover, K., *Robust and Optimal Control*, Prentice-Hall, 1996.