OPTIMAL RATE CONTROL FOR HIGH SPEED TELECOMMUNICATION NETWORKS

Eitan Altman

INRIA B.P. 93
06902 Sophia Antipolis Cedex
FRANCE

Tamer Başar

Coordinated Science Lab.
University of Illinois
Urbana, IL 61801, U.S.A.

ABSTRACT

We present in this paper two approaches for designing controllers that dynamically regulate the rate of data flow into a network based on feedback state information. They result in controllers that are similar to ones that have been advocated for both end-to-end and hop-by-hop congestion control in high-speed networks. Many existing control protocols have been developed on growing available experience, using ad-hoc techniques that did not come as a result of a control-theoretical study. This is due to the high complexity of the controlled systems, that are typically decentralized, have non-linear dynamics, and may only use partial noisy delayed information. Some attempts have been made in recent years to use control theory to design flow controllers with, however, no explicit objective functions to be minimized; moreover, the class of control policies in existing theoretical work is quite restricted. In this paper we formulate explicitly some cost criteria to be minimized, related to performance measures such as delays, throughputs and loss probabilities. Using a linearized model, we then view the design problem as an optimal control problem. We follow two approaches to model interfering traffic and other unknown data: the H^∞ approach, and the LQG one; and determine for both cases the optimal controllers. Some simulation results complete the study.

1. Introduction

Adaptive flow control mechanisms are being used and developed extensively in High Speed Networks, in order to guarantee high quality of services on the one hand, and allow for the efficient use of the network, on the other. The controller has to ensure that when other sources transmit, its own transmission rate is adjusted adaptively so as to avoid congestion in the network, since congestion might result in low throughputs, high delays, and high rate of losses of packets. When the interfering traffic of other sources is low, the controller is expected to allow a high rate of transmission of information so as to make the best use of the bandwidth available.

In many existing flow control protocols, some information on the state of the network is available to sources. For example, in many protocols, when a packet arrives at the destination, an acknowledgement is sent back to the source (the TCP/IP protocol in the Internet [8]). Other types of information could also be made available to the flow controller, such as information on queue length in a bottleneck node, [6], and the effective service rate available to that source in a bottleneck node [9].

Supported by Grants NSF ECS 92-20632 and NSF ECS 93-12807 from the National Science Foundation.

In this paper we assume that the performance measures (throughputs, delays, loss probabilities etc...) are determined essentially by a bottleneck node. This assumption has both theoretical and experimental [4] justification, and is often made in the literature in order to analyze or design controllers [1, 5, 7].

We consider a communication network with a linearized dynamics (see [1, 5, 9] for its derivation and validation). The control problem is solved in Sections 3 and 4. The design of the controller is based on some model of the disturbance by exogenous traffic. In Section 5 we present simulation results, and discuss their interpretation.

2. The model

We consider a discrete time model, where a time unit corresponds to a round-trip delay. Let q_n denote the queue length at a bottleneck link, and μ_n denote the effective service rate available for traffic of the given source in that link at the beginning of the nth time slot. Let x_n denote the source rate during the nth time slot. We thus consider a rate-based flow control, where based on current and previous noisy information on μ_n, q_n, and x_n (to be made precise below), the controller updates the transmission rate. Hence, the queue length evolves according to

$$q_{n+1} = q_n + x_n - \mu_n. \tag{1}$$

Since several sources with varying transmission rates may share the same bottleneck link, the service rate μ_n available to the controlled source may change over time in an unpredictable way. The other sources may be represented as some interference, which we model as a stable ARMA process

$$\mu_n = \mu + \xi_n, \quad \xi_n = \sum_{i=1}^{d} \ell_i \xi_{n-i} + k\phi_n. \tag{2}$$

We assume that ℓ_i, which will typically be obtained by some estimation/identification procedure, are such that (2) describes a stable system. The variable ϕ_n in (2) stands for disturbance, the nature of which will be described later.

We assume that, at the end of step n, the source obtains a noisy estimate of the value of μ_n, and a noisy estimate of the value of q_n. We use $\hat{\mu}_n$ to denote a measurement for the bottleneck rate, and $\hat{q}_n$ to denote a measurement for the bottleneck queue size. For our analysis, we assume that these measurements can be written as

$$\hat{\mu}_n = \mu_n + cv_n, \quad \hat{q}_n = q_n + aw_n, \tag{3}$$

where v_n and w_n are disturbances whose nature will be described shortly, and a and c are some constants.

In [1], the starting point was a given controller of the form $x_{n+1} = (1-\alpha)x_n + \alpha\hat{\mu}_n - \beta(\hat{q}_n - Q)$, and the design was restricted in the choice of the parameters α and β. In particular, performance measures such as stability, expected transient (and steady state) queue lengths, the variance in steady state and the rate of convergence to steady state were studied as a function of these parameters. Here, Q stands for some desired level of the queue length; it will appear again later in the cost function we adopt.

Define the shifted versions of the variables $x_n, q_n, \hat{\mu}_n, \hat{q}_n$: $\tilde{x}_n := x_n - \mu$, $\tilde{q}_n := q_n - Q$, $\hat{\tilde{\mu}}_n := \hat{\mu}_n - \mu$, $\hat{\tilde{q}}_n := \hat{q}_n - Q$. We further define the (scaled) variation in the input rate by bu_n, where u_n will be called the control, and b is some constant. In terms of the new variables, we now have the following dynamics:

$$\tilde{q}_{n+1} = \tilde{q}_n + \tilde{x}_n - \xi_n \tag{4}$$

$$\xi_n = \sum_{i=1}^{d} \ell_i \xi_{n-i} + k\phi_n \tag{5}$$

$$\tilde{x}_{n+1} = \tilde{x}_n + bu_n \tag{6}$$

and the observations

$$\hat{\tilde{\mu}}_n = \xi_n + cv_n, \qquad \hat{\tilde{q}}_n = \tilde{q}_n + aw_n. \tag{7}$$

We define the following quadratic cost:

$$L = g\,\|\tilde{x} - \xi\|^2 + \|\tilde{q}\|^2 + \|u\|^2 \tag{8}$$

where $g > 0, \|u\|^2 := \sum_{n=0}^{\infty} |u_n|^2$, etc... This definition of the cost allows us to quantify optimization criteria that have been used in previous control models [1]:

(i) The first term represents the quality with which the input rate tracks the available service rate, where g is a positive weighting term. A high rate of tracking is known to be desirable; see the discussion in [5].

(ii) The second term is a penalty for deviating from a desirable queue length. Setting some desired level of queue reflects the fact that we do not wish to have a large queue, so as to avoid losses, and on the other hand, we do not wish to have a low level of queue, since if the queue is empty and the input rate is lower than the service rate then there is a waste of potential throughput. We have not included any additional weighting on this term, as any such positive weight can be absorbed into the other variables.

(iii) The last term stands for a penalty for high jitter, i.e. for high variability of (input) transmission rate, which is known to be undesirable; see the discussion in [5]. In many proposed telecommunication networks, the source will pay more for higher burstiness (variability) of its input rate, since highly variable input rate has typically a bad influence on other traffic. Again, as in the case of the second term, the weighting here has been normalized to 1.

We allow the controller to be a (measurable) function of all past and present observation and past actions: $u_n = f(\tilde{x}_m, \hat{\tilde{\mu}}_m, \hat{\tilde{q}}_m, u_{m-1}, m = 0, ..., n)$. We consider two types of models for the disturbances:

(M1) The Gaussian model: v_n, w_n, ϕ_n are taken as i.i.d. Gaussian random variables with zero mean and unit variance. It then follows that the measurements $\hat{\mu}_n$ and $\hat{q}_n$ are unbiased. The cost to be minimized in this case is the expected value of L, written as $E[L]$.

(M2) The H^∞ approach: In this case v_n, w_n and ϕ_n are taken as unknown (deterministic) disturbances. The cost to be minimized is the ℓ_2 gain from the disturbances to the cost L, i.e.,

$$\sup_{\{\phi_n, v_n, w_n\}_{n=0}^{\infty}} \frac{L}{\|\phi, v, w\|^2} \tag{9}$$

where $\|\phi, v, w\|^2 := \|\phi\|^2 + \|v\|^2 + \|w\|^2 = \sum_{n=0}^{\infty} |\phi_n|^2 + |v_n|^2 + |w_n|^2$. Let us denote the square-root of the infimum of this cost function over all admissible controllers $\{f_n\}$ by γ^*, i.e.

$$(\gamma^*)^2 = \inf_{\{f_n\}_{n=0}^{\infty}} \sup_{\{\phi_n, v_n, w_n\}_{n=0}^{\infty}} \frac{L(f; \phi, v, w)}{\|\phi, v, w\|^2}. \tag{10}$$

If there exists an actual minimizing control, say f^*, to the optimization problem above, then it can be shown [3] that, defining a soft-constrained cost function $L_\gamma(u; \phi, v, w) := L - \gamma^2 \|\phi, v, w\|^2$ to be viewed as the kernel of a two player game, to be minimized by Player 1 (controlling u) and maximized by Player 2 (controlling (ϕ, v, w)), f^* has the property:

$$\sup_{\phi, v, w} L_{\gamma^*}(f^*; \phi, v, w) = \inf_f \sup_{\phi, v, w} L_{\gamma^*}(f; \phi, v, w).$$

The quantity above is the upper value of the zero-sum dynamic game with kernel L_{γ^*}, which is in fact equal to zero. It can actually be shown that for any $\gamma \geq \gamma^*$, the upper value of the game with parametrized kernel L_γ is zero, and for $\gamma < \gamma^*$, its upper value is infinite. Hence, γ^* is the smallest positive scalar γ for which the zero-sum game with kernel L_γ has a finite upper value.

Instead of obtaining just f^* defined above, we will in fact solve a parametrized class of controllers, $\{f^\gamma, \gamma > \gamma^*\}$, where f^γ is obtained from

$$\sup_{\phi, v, w} L_\gamma(f^\gamma; \phi, v, w) = \inf_f \sup_{\phi, v, w} L_\gamma(f; \phi, v, w).$$

The controller f^γ will clearly have the property that it ensures a performance level γ^2 for the index adopted for M2, i.e. the attenuation is bounded by

$$\mathcal{A}(f^\gamma; \phi, v, w) := \frac{\{L(f^\gamma; \phi, v, w)\}^{1/2}}{\|\phi, v, w\|} \leq \gamma \quad \text{for all } \phi, v, w. \tag{11}$$

It will turn out that the limit $\lim_{\gamma \to \infty} f^\gamma =: f^\infty$ is a well-defined controller, and solves uniquely the control problem with the Gaussian model, M1.

3. Complete solution to the problem with Model M2 with additional perturbation

The optimal control problem formulated above can be solved by a suitable modification of the theory of discrete-time H^∞-control developed in [3]. Toward this end, we first write (4)-(6) in standard state variable form. Introduce the d-dimensional vector $\eta_n := (\xi_{n-d+1}, ..., \xi_n)'$ in terms of which (5) can be re-written as a first-order difference equation:

$$\eta_{n+1} = C\eta_n + k\,(0, 0, ..., 1)'\,\phi_n \tag{12}$$

where

$$C := \begin{pmatrix} 0 & 1 & 0 & 0 & .... & 0 \\ 0 & 0 & 1 & 0 & .... & 0 \\ . & . & . & . & .... & \\ \ell_d & \ell_{d-1} & \ell_{d-2} & & .... & \ell_1 \end{pmatrix}$$

Further introducing the $(d+2)$-dimensional vector $z_n := (\tilde{x}_n, \tilde{q}_n, \eta'_n)'$ we can write (4)-(6) in the compact form:

$$z_{n+1} = Az_n + D\phi_n + Bu_n \tag{13}$$

where

$$A := \left(\begin{array}{cc|ccccc} 1 & 0 & 0 & ... & 0 & 0 \\ 1 & 1 & 0 & ... & 0 & -1 \\ \hline & & & & & \\ \multicolumn{2}{c|}{0} & \multicolumn{4}{c}{\mathrm{C}} \\ & & & & & \end{array}\right)$$

$$D := (0, ..., 0, k)' \quad B := (b, 0, 0, ..., 0)'$$

Now, the stability of the ARMA process (2) is equivalent to the stability of the matrix C in (12), which is captured in the following assumption required to hold throughout the remaining analysis:

- **<u>A1.</u>** All eigenvalues of C are in the unit circle in the complex plane.

Under A1, we have the following fact, which will be useful in the statement of our main result.

- **<u>F1.</u>** Under A1, the pair (A,B) is <u>stabilizable</u>.

As far as the measurements available to the controller go, we have two disturbance corrupted state measurements, given by (7), and one perfect measurement of the state, $\tilde{x}_n$. This is a "mixed" structure, that leads to a *singular* measurement equation, which is not allowed in standard theory. To circumvent this difficulty, we will first perturb $\tilde{x}_n$ with a small independent disturbance, to arrive at the measurement equation $\hat{\tilde{x}}_n = \tilde{x}_n + \epsilon\tilde{\tilde{v}}_n$, $0 < \epsilon << 1$ and assume that now $\hat{\tilde{x}}_n$ is available for control purposes, instead of $\tilde{x}_n$. In the above, $\tilde{\tilde{v}}_n$ is the auxiliary disturbance introduced, and ϵ is a small positive quantity. This modification is only for mathematical convenience, which will in the end be made to vanish asymptotically by letting $\epsilon \to 0$, so as to recover the original information structure.

To introduce the modified measurement equation, let $y_n := (\hat{\tilde{x}}_n, \hat{\tilde{q}}_n, \hat{\tilde{\mu}}_n)'$. Then $y_n = Hz_n + E_\epsilon v_n$ where

$$H := \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix} \quad E_\epsilon := \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & c \end{pmatrix}$$

and $v_n = (\tilde{\tilde{v}}_n, w_n, \tilde{v}_n)'$. Since $E_\epsilon E'_\epsilon > 0$ for all $\epsilon > 0$, this is a nonsingular measurement equation. Compatible with this modification, we also modify the state equation (13), to write it as $z_{n+1} = Az_n + Bu_n + D_\epsilon\zeta_n$ where

$$D_\epsilon := \begin{pmatrix} 0 & 0 & ... & 0 & k \\ \epsilon & 0 & ... & 0 & 0 \end{pmatrix}' \qquad \zeta_n := (\tilde{\phi}_n, \phi_n)$$

with $\tilde{\phi}_n$ being the new disturbance introduced.

To complete the description, we have to write the cost function (8) in terms of these new variables:

$L = \sum_{n=0}^{\infty} \left\{|z_n|_N^2 + |u_n|^2\right\}$ where

$$N := \left(\begin{array}{cc|cccccc} g & 0 & 0 & . & . & . & 0 & -g \\ 0 & 1 & 0 & . & . & . & 0 & 0 \\ \hline 0 & 0 & & & & & & 0 \\ . & . & & & & & & . \\ . & . & & & 0 & & & . \\ . & . & & & & & & 0 \\ -g & 0 & 0 & . & . & . & 0 & g \end{array}\right)$$

and the H^∞ criterion (to be minimized) now becomes

$$\sup_{v,\zeta} \frac{L}{\|v,\zeta\|^2}, \tag{14}$$

the square-root of whose infimum over all admissible controllers we now denote by $\gamma^*(\epsilon)$. It turns out that $\gamma^*(\epsilon)$ is actually continuous at $\epsilon = 0$, and the quantity $\gamma^*(0)$ is equal to the γ^* defined earlier (for the unmodified problem).

Three additional properties, related to the system matrices, are now worth noting, which we state below as facts:

- **<u>F2.</u>** The pair (A, H) is observable.
- **<u>F3.</u>** The pair (A, N) is observable.
- **<u>F4.</u>** For each $\epsilon > 0$, the pair (A, D_ϵ) is controllable.

We are now in a position to state the solution to the modified H^∞ control problem formulated above, directly from Başar and Bernhard [3], Chap. 6.

Proposition 1 *Let $\epsilon > 0$ be fixed, and assumption A1 hold. Then, $\gamma^*(\epsilon)$ defined above is finite, and for all $\gamma > \gamma^*(\epsilon)$:*
1) There exists a minimal positive definite solution, M_ϵ, to

$$M = N + A'M\Lambda_\epsilon^{-1}A \tag{15}$$

such that

$$\gamma^2 I - D'_\epsilon M_\epsilon D_\epsilon > 0, \tag{16}$$

where

$$\Lambda_\epsilon := I + (BB' - \gamma^{-2}D_\epsilon D'_\epsilon)M. \tag{17}$$

2) There exists a minimal positive definite solution, Σ_ϵ, to

$$\Sigma = D_\epsilon D'_\epsilon + A\Sigma R_\epsilon^{-1}A' \tag{18}$$

such that

$$\gamma^2 I - N^{1/2}\Sigma_\epsilon N^{1/2} > 0, \tag{19}$$

where

$$R_\epsilon := I + \left(H'(E_\epsilon E'_\epsilon)^{-1}H - \gamma^{-2}N\right)\Sigma. \tag{20}$$

3)

$$\Sigma_\epsilon^{1/2} M_\epsilon \Sigma_\epsilon^{1/2} < \gamma^2 I. \tag{21}$$

4) An H^∞-controller that ensures the bound γ^2 in the index (14) is

$$u_{n,\epsilon}^{\gamma} = f_\epsilon^{\gamma}(\hat{z}_{n|n}) = -B'M_\epsilon\Lambda_\epsilon^{-1}A\hat{z}_{n|n}, \qquad n = 0, 1, ... \tag{22}$$

where $\hat{z}_{n|n}$ is generated by

$$\hat{z}_{n|n} = \left(I + \Sigma_\epsilon H'(E_\epsilon E'_\epsilon)^{-1}H - \gamma^{-2}\Sigma_\epsilon M_\epsilon\right)^{-1} \times \left(\check{z}_n + \Sigma_\epsilon H'(E_\epsilon E'\epsilon)^{-1}y_n\right) \tag{23}$$

$$\check{z}_{n+1} = A\check{z}_n + Bu_n + A\Sigma_\epsilon R_\epsilon^{-1} \times \left[\gamma^{-2} N\check{z}_n + H'(E_\epsilon E_\epsilon')^{-1}(y_n - H\check{z}_n)\right] \tag{24}$$

5) The controller (22) leads to a bounded input-bounded state stable system, i.e. for all bounded disturbances v, ζ, the system state z_n and the filter state $\hat{z}_{n|n}$ remain bounded. Equivalently, the two matrices $(I - BB'M_\epsilon \Lambda_\epsilon^{-1})A$ and $\left(I - H'(E_\epsilon E_\epsilon')^{-1} H\Sigma_\epsilon R_\epsilon^{-1}\right) A'$ have all their eigenvalues in the unit circle.

If any one of the three properties (1)-(3) above does not hold, then $\gamma < \gamma^*(\epsilon)$. ◇

4. The limiting solution as $\epsilon \downarrow 0$

We next study the limit of the solution given in Proposition 1 as $\epsilon \downarrow 0$. For the limit to be well defined, it will be sufficient to show that equations and relationships (15)-(21) are well-defined in the limit as $\epsilon \downarrow 0$.

The first three depend continuously on $\epsilon \geq 0$, and hence the limit (15) will be replaced by

$$M = N + A'M\Lambda_0^{-1}A \tag{25}$$

where

$$\Lambda_0 := I + \mathrm{diag}(b^2, 0, ..., 0, -\gamma^2 k^2)M \tag{26}$$

and (25) will have to be solved under the scalar condition (as counterpart of (16)):

$$k^2(0, ..., 0, 1)M(0, ..., 0, 1)' < \gamma^2 \tag{27}$$

which guarantees invertibility of (26). Again, for each $\gamma > \gamma^*(0)$, (25) admits a unique minimal positive definite solution, which also satisfies (27).

The limiting analysis of (18)-(20) is somewhat more complicated, since it involves the inverse of a matrix that becomes singular at $\epsilon = 0$ (which is $E_\epsilon E_\epsilon'$). However, this isolated singularity does not lead to any singularity for the solution of (18) as to be shown below. We will in fact show that in the limit as $\epsilon \downarrow 0$, the solution of (18) will be in the form

$$\Sigma_0 = \begin{pmatrix} 0 & . & . & . & 0 \\ . & & & & \\ . & & P & & \\ . & & & & \\ 0 & & & & \end{pmatrix} \tag{28}$$

where $P > 0$ is of dimension $(d+1) \times (d+1)$. To see this, let us first rewrite ΣR_ϵ^{-1} as

$$\Sigma R_\epsilon^{-1} = \Sigma^{1/2} \left[I + Z_0 +\right]^{-1} \Sigma^{1/2}$$

where $Z_0 := \Sigma^{1/2} H'(E_\epsilon E_\epsilon')^{-1} H \Sigma^{1/2} - \gamma^{-2}\Sigma^{1/2} N \Sigma^{1/2}$ and $\Sigma^{1/2}$ denotes the unique nonnegative definite square root of Σ. Let the $(d+1)^2$-dimensional lower block of N be denoted by $\tilde{N}$, which is

$$\tilde{N} = \mathrm{diag}(1, 0, ..., 0, g) \tag{29}$$

and suppose that Σ is given by (28), where P is arbitrary.

$$\Sigma^{1/2} N \Sigma^{1/2} = \text{block-diag}\left(0,\ P^{1/2}\tilde{N}P^{1/2}\right)$$

Then, straightforward manipulations lead to

$$\Sigma^{1/2} H'(E_\epsilon E_\epsilon')^{-1} H \Sigma^{1/2} = \text{block-diag}\left(0,\ P^{1/2}\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H}P^{1/2}\right)$$

where

$$\tilde{H} := \underbrace{\begin{pmatrix} 1 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 1 \end{pmatrix}}_{2\times(d+1)} \qquad \tilde{E} := \underbrace{\begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix}}_{2\times 2} \tag{30}$$

Denote

$$Z_1 = I + P^{1/2}\left(\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}\tilde{N}\right)P^{1/2},$$

$$\tilde{S} = P^{1/2}\left[I + P^{1/2}\left(\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}\tilde{N}\right)P^{1/2}\right]^{-1}P^{1/2}$$

Then,

$$\begin{aligned} \Sigma R_\epsilon^{-1} &= \Sigma^{1/2}\left[\text{block-diag}(1, Z_1)\right]^{-1}\Sigma^{1/2} \\ &= \text{block-diag}(0, \tilde{S}) \end{aligned} \tag{31}$$

which is independent of ϵ. Pre- and post-multiplying by A and A', we obtain

$$A\Sigma R_\epsilon^{-1}A' = \text{block-diag}\left(0,\ \tilde{A}\tilde{S}\tilde{A}'\right)$$

where $\tilde{A}$ is the $(d+1)^2$ dimensional lower block of A, i.e.

$$\tilde{A} := \left(\begin{array}{c|c} 1 & 0\ .\ .\ .\ 0\ \text{-}1 \\ \hline 0 & \\ . & \\ . & C \\ . & \\ 0 & \end{array}\right) \tag{32}$$

Furthermore, since $\lim_{\epsilon\to 0} D_\epsilon D_\epsilon' = \mathrm{diag}(0, ..., 0, k^2)$, it follows that the structure (28) is consistent with (18). Introducing a $(d+1)$-dimensional vector $\tilde{D} := (0, ..., 0, k)'$, the equation to be satisfied by P can now be written as

$$P = \tilde{D}\tilde{D}' + \tilde{A}P\left[I + \left(\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}\tilde{N}\right)P\right]^{-1}\tilde{A}', \tag{33}$$

and (19) becomes equivalent to

$$\gamma^2 I - \tilde{N}^{1/2}P\tilde{N}^{1/2} > 0. \tag{34}$$

Finally, (21) simplifies to

$$P^{1/2}\tilde{M}P^{1/2} < \gamma^2 I \tag{35}$$

where $\tilde{M}$ is the $(d+1)^2$-dimensional lower block of M. Hence, (15)-(17) are well-defined at $\epsilon = 0$, and (18)-(21) admit well-defined limits in the space of dimension $(d+1)$ instead of $(d+2)$.

To complete the limiting analysis, we still have to find the limiting form of the controller (22)-(24). Toward this end, write z, as $\epsilon \downarrow 0$, as $z = (\alpha, \beta')'$ where α is of dimension 1, and β of dimension $d+1$. Compatibly, write (again as $\epsilon \downarrow 0$):

$$\hat{z}_{n|n} = (\hat{\alpha}_{n|n},\ \hat{\beta}_{n|n}')'; \qquad \check{z}_n = (\check{\alpha}_n,\ \check{\beta}_n')'. \tag{36}$$

Then, using the earlier manipulations, and some straightforward extensions, it can be shown that $\hat{\alpha}_{n|n}, \hat{\beta}_{n|n}, \check{\alpha}_n, \check{\beta}_n$ are generated by (from (23)-(24), and using the notation introduced by (29)-(32)):

$$\hat{\alpha}_{n|n} = \check{\alpha}_n \tag{37}$$

$$\check{\alpha}_{n+1} = \check{\alpha}_n + bu_n \tag{38}$$

$$\hat{\beta}_{n|n} = \left(I + P\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}P\tilde{M}\right)^{-1} \times \left(\check{\beta}_n + P\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{y}_n\right) \tag{39}$$

$$\check{\beta}_{n+1} = \tilde{A}\check{\beta}_n + \tilde{A}\tilde{S}\left[\gamma^{-2}\tilde{N}\check{\beta}_n + \tilde{H}'(\tilde{E}\tilde{E}')^{-1}(\tilde{y}_n - \tilde{H}\check{\beta}_n)\right] \tag{40}$$

$$\tilde{y}_n := (\hat{\tilde{q}}_n, \hat{\tilde{\mu}}_n)' \tag{41}$$

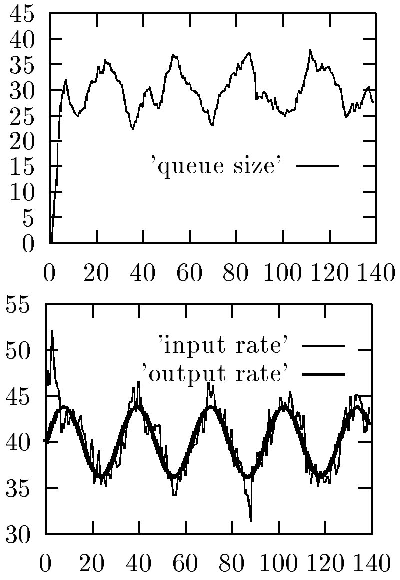

Figure 1: $\gamma = 6$, equally weighted costs $g = b = 1$, sinusoidal perturbations

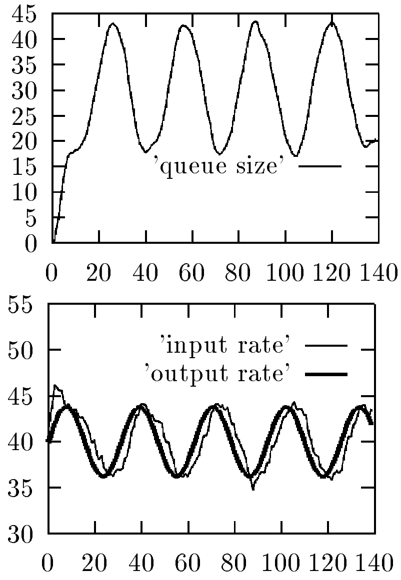

Figure 2: $\gamma = 13.8$, unequally weighted costs $g = b = 0.1$, sinusoidal perturbations

Clearly, (38) generates $\tilde{x}_n$ whose perfect value is already available to the controller. Hence, the limiting controller is of dimension $d+1$, with the worst-case observer given by (39)-(40). The counterpart of (22) is then

$$u_n^\gamma = f^\gamma(x_n, \hat{\beta}_{n|n}) = -B' M_0 \Lambda_0^{-1} A \begin{pmatrix} \tilde{x}_n \\ \hat{\beta}_{n|n} \end{pmatrix}, \qquad (42)$$

$n = 0, 1, \ldots$ This is summarized in the following:

Theorem 1 *Consider the original control problem formulated in Section 2, under the H^∞ model* (M2) *of the disturbance. Let Assumption A1 hold. Then γ^* defined by (10) is finite, and for all $\gamma > \gamma^*$:*
1) There exists a minimal positive definite solution, M_0, to (25) such that (27) holds.
2) There exists a minimal positive definite solution, P, to (33), such that (34) holds.
3) Condition (35) holds.
4) An H^∞-controller that ensures the bound γ^2 in the index (9) is given by (42).
5) The controller (42) leads to a bounded input-bounded state stable system (see Proposition 1 (5)).
For a given $\gamma > 0$, if any one of the properties (1)-(3) above does not hold, then $\gamma < \gamma^*$. $\diamond$

By letting $\gamma \to \infty$, we now solve the original problem under the Gaussian interpretation (M1) of the disturbances.

Corollary 1 *Under A1, the optimal control problem with the disturbance model M1 admits the unique solution:*

$$u_n^\infty = f^\infty(x_n, \hat{\beta}_{n|n}) = -B'\overline{M}(I + BB'\overline{M})^{-1} A \begin{pmatrix} \tilde{x}_n \\ \hat{\beta}_{n|n} \end{pmatrix}, \qquad (43)$$

$n = 0, 1, \ldots$ *where $\overline{M}$ is the unique positive definite solution of $\overline{M} = N + A'\overline{M}(I + BB'\overline{M})^{-1}A$ in the class of nonnegative-definite matrices, and*

$$\begin{aligned} \hat{\beta}_{n|n} &= \left(I + \overline{P}\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H}\right)^{-1} \times \\ &\quad \left(\breve{\beta}_n + \overline{P}\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{y}_n\right) \qquad (44) \\ \breve{\beta}_{n+1} &= \tilde{A}\breve{\beta}_n + \tilde{A}\overline{S}\tilde{H}'(\tilde{E}\tilde{E}')^{-1}(\tilde{y}_n - \tilde{H}\breve{\beta}_n) \qquad (45) \end{aligned}$$

where

$$\overline{S} := \overline{P}\left[I + \tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H}\overline{P}\right]^{-1} \qquad (46)$$

and $\overline{P}$ is the unique positive definite solution of

$$\overline{P} = \tilde{D}\tilde{D}' + \tilde{A}\overline{S}\tilde{A}' \qquad (47)$$

in the class of nonnegative definite matrices.
Furthermore, the controller (43) leads to a stable system.

5. Simulation Results

We present a set of selective simulation results from [2]. The observation noises $v(n)$ and $w(n)$ are chosen as standard i.i.d. Gaussian sequences. In order to illustrate the usefulness of the H^∞ (worst case) controller in the context of telecommunications, we present results obtained under different choices of perturbations (interfering traffic) ϕ_n (in (2)):
(i) i.i.d. uniformly distributed perturbations,
(ii) highly correlated perturbations: $\phi_n = \sin(0.2n)$.

We illustrate how the design parameters b (eq. (6)) and g (eq. (8)) influence the behavior of the system.

The parameter g will influence the tracking between input and output rates. This is seen from equations (4) and (8). The larger g is, the more costly will deviations between the input and output rates be. Indirectly, this may result in reducing the variations of the queue size (see (4)).

The parameter b is expected to influence the jitter, i.e. the variability of the transmission rate $\tilde{x}$. It is easily seen from (6) and (8) that choosing lower b results in a higher weight on deviations of u from zero, and thus a higher weight on the variability of $\tilde{x}$. Thus, we expect that smaller values for b will be useful when designing controllers for video or for voice traffic, where the transmission rate is required typically to be regular (i.e. low jitter).

Finally, if g is small and b large, then the second term in (8) becomes important, which means that we pay more for deviations of the queue length from its target Q. In practice this part will be responsible for decreasing loss probabilities (that occur when queue sizes are high) and increasing throughputs (since we may lose potential throughput when the queue is empty, if, at that time, the input rate is lower than the output rate).

We shall examine two situations in the simulations below:
Case 1: The three different criteria are weighted equally in the cost L to be minimized (see (6) and (8)): $g = b = 1$.
Case 2: $b = g = 0.1$. We thus place relatively less emphasis on the tracking between input and output rates (and thus, indirectly, allow for more variability of the queue length).

For Case 1, the optimal (smallest) value of γ (see eq. (10)) for the H^∞ controller is $\gamma^* = 5.9$. We chose $\gamma = 6$ (slightly larger than γ^*, so as to ensure stability). For Case 2 we obtained $\gamma^* = 13.5$ and chose $\gamma = 13.8$.

The interfering traffic uses an ARMA model of order two with $\ell_1 = 0.7, \ell_2 = -0.3$, $k = \sqrt{5}$ (same as in [1]). a and c (in (3)) were chosen $\sqrt{2}$, $Q = 30$, and $\mu = 40$ units (packets per round trip delay). The duration of each simulation is 140 time units. The queues were initially empty. A steady state was reached after 20 time units.

Figures 1 and 2 (which correspond to Cases 1 and 2, resp.) are concerned with the case of sinusoidal interfering traffic. They illustrate the influence of the parameters b and g on the queue sizes, on the variability of the (controlled) input rate and the way it tracks the output rate. We see that the queue size indeed has a much higher variability in Case 2, where it oscillates between the values of 18 and 43. In Case 1, the queue length remains always between 26 and 37 (except for the initial transient period). Comparing the input rates $\tilde{x}$ with the output rates ξ, we see that the input has a high "inertia" in Case 2: it does not hurry to track the output rate, and a tracking delay of 6 time units occurs. In Case 1, the tracking is quite instantaneous, but at the same time, there are higher bursts around the tracked signal.

The attenuation (at steady state) obtained by simulations was 3.40 for Case 1, and 7.01 for Case 2 (thus eq. (11) indeed holds), for sinusoidal ϕ. We also performed simulations for very high values of γ, which resulted in the controller that is optimal for the Gaussian noise. When using that Gaussian controller and feeding it by the sinusoidal perturbation, the attenuation obtained was 2.78 for Case 1 and 9.17 in Case 2. Comparing this with the H^∞ controller we see that for sinusoidal ϕ, it sometimes performs better and sometimes worse.

In Fig. 3 we zoom on some typical short time interval to show how b and g influence the way the input rate tracks the output rate. The upper figure illustrates Case 1, for which we get quick tracking with a bursty behavior (more overshoots). In the lower figure, corresponding to Case 2, we have a slow tracking and a smoother behavior of the input rate. The interfering traffic here is generated by i.i.d. uniformly distributed ϕ_n.

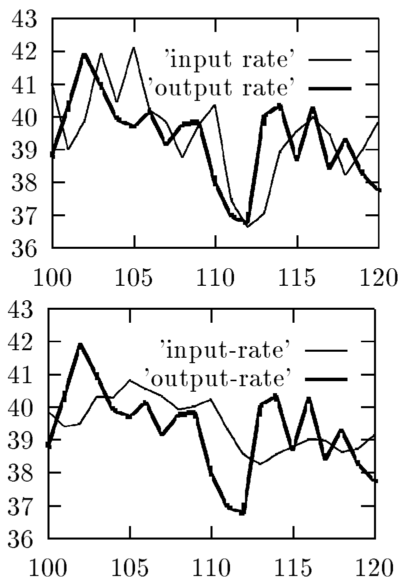

Figure 3: i.i.d. uniformly distributed ϕ_n. Input versus output rates for equal weighting (upper figure) and unequal weighting (lower figure)

6. References

[1] E. Altman, F. Baccelli and J. C. Bolot, "Discrete-time analysis of adaptive rate control mechanisms", *High Speed Networks and their performance*, Perros and Viniotis Eds., North Holland, pp. 121-140, 1994.

[2] E. Altman and T. Başar, "Optimal rate control for high speed telecommunication networks", preprint.

[3] T. Başar and P. Bernhard, *H^∞-Optimal Control and Relaxed Minimax Design Problems: A Dynamic Game Approach*, Birkhäuser, Boston, MA, 1991 (2nd edition, 1995).

[4] J.-C. Bolot, "End-to-end delay and loss behavior in the Internet," *Proc. ACM Sigcomm '93*, pp. 289-298, San Fransisco, CA, Sept.1993.

[5] J-C. Bolot and A. U. Shankar, "Analysis of a fluid approximation to flow control dynamics," *Proc. IEEE INFOCOM '92*, Florence, Italy, May 1992, pp. 2398-2407.

[6] T. Faber, L. H. Landweber and A. Mukherjee, "Dynamic time windows: Packet admission control with feedback," *Proc. ACM SIGCOMM '92*, Baltimore MD, Aug. 1992, pp. 124-135.

[7] K. Fendick, M. A. Rodrigues and A. Weiss, "Analysis of a rate-based control strategy with delayed feedback," *Proc. ACM SIGCOMM '92*, Baltimore, MD, August 1992, pp. 136-148.

[8] V. Jacobson, "Congestion avoidance and control," *Proc. ACM SIGCOMM '88*, Stanford, CA, pp. 314-329, August 1988.

[9] S. Keshav, "A control-theoretic approach to flow control," *Proc. ACM SIGCOMM '91*, Zurich, Switzerland, September 1991.