Constructive Solutions to a Decentralized Team Problem with Application in Flow Control for Networks*

Orhan Çağrı İmer and Tamer Başar†
Department of Electrical and Computer Engineering
and
Coordinated Science Laboratory
University of Illinois
1308 W. Main Street
Urbana, Illinois 61801-2307, USA

Abstract

The paper considers a decentralized stochastic team problem with a *partially-nested* information pattern, that arises in the context of flow control. Basically, we consider a network with a number of users having access to differently delayed versions of the same information, with each one deciding on his own rate of transmission, but participating in a common cost quantifying the outcome of their joint actions. This leads to a Linear-Quadratic-Gaussian (LQG) team problem, with partially nested information. We study the derivation of the optimal solution in a two user network and show that the solution exists in both finite and infinite-horizon cases. The controller, which turns out to be certainty-equivalent, is constructed recursively using a dynamic programming type approach. We also present an algorithm to construct the optimal solution for the most general case with multiple (more than two) users. Finally, we present various simulation results to illustrate the performance of the optimal controller under different scenarios.

1 Introduction

It is a well-known and well-acknowledged fact that in stochastic teams the information pattern plays an important role in the existence as well as the

*Research that led to this paper was supported in part by the AFOSR MURI Grant AF DC 5-36128

†Address for correspondence: Prof. Tamer Başar, Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois, 1308 W. Main Street, Urbana, Illinois 61801-2307, E-mail: tbasar@decision.csl.uiuc.edu; FAX: (217) 244-1653; Tel: (217) 333-3607.

derivation of optimal team solutions. The simplest information pattern one can think of is the so-called *classical* one, in which all members of a team receive the same information and have perfect recall. The next level is the *partially nested* one, which has the property that if a team member's information depends on the control variable of some other member, then the former has access to all the information accessible to the latter. It is well-known [7] that a stochastic team problem defined on a finite horizon and with partially nested information structure can be transformed to a static team problem (albeit of a much higher dimension), with the equivalence being in the sense that the solution of one can be obtained from the solution of the other. This equivalence readily leads to existence results, such as the one of the partially nested linear-quadratic-Gaussian team, where the solution exists, is unique, and is affine in the available information [7] – a result that follows from an existence and uniqueness result for quadratic Gaussian teams [9].

The equivalence alluded to above does not lead, however, to any constructive methods for obtaining the team-optimal solution for a general partially nested information structure, nor to closed-form solutions[1]. It also does not say much about the solution of infinite-horizon team problems with partially nested information. It is the latter class of problems that we will be addressing in this paper, for a specific model motivated by an application involving congestion control in communication networks.

Hence, we consider here a network of a number of users, viewed as members of a team, having access to differently delayed versions of the same information. Each user controls his own rate of transmission and the state of the network is described by a difference equation. The expected cost to be minimized is quadratic in state and the decision variables. The motivation behind this formulation is justified in a real network with a bottleneck node that plays a major role in determining the performance of a number of users. Although in a real network environment, users may be interconnected in several ways, the single bottleneck node assumption admits theoretical as well as experimental justification [6]. The available service rate is modeled as an autoregressive (AR) process driven by a white noise process, as in [4]. The state of the network is nothing but the queue length at the bottleneck node. To decide on the rate of transmission of each user, we assume that both queue length and total service rate information is available to the node, but the transmission of these decisions taken by a node to the sources (users) incur different (propagation) delays, depending on the distance between the node and each user. Hence, even though it may appear that this problem is a centralized one with the decisions made only by the node, but with *action delays*, it can in fact be shown that it is equivalent to one where the decisions are made at the sources (by the users), but with information delay (see, [2]).

The information pattern here is hence partially nested, since each user

[1] If the partially nested information is of a special type, such as *one-step-delay information sharing* pattern, then a closed-form solution can be obtained by recursive decomposition [10], [5].

simply has a subset of the information available to the users with smaller delays. This special structure with a different information delay for each user poses a fundamental difficulty in the actual computation of the solution. One way to circumvent this difficulty is first to solve the problem by ignoring the delays. In this case, it can be shown that the problem can be reduced to a standard discrete-time linear-quadratic-regulator problem, which is known to admit a unique linear solution. Then, one can incorporate the delays into the controllers, using the certainty equivalence approach. An earlier paper [2] has employed this method to obtain the solution to the flow control problem and has presented two sub-optimal certainty-equivalent controllers, called *Controller 1* and *Controller 2*. The controller constructed here being optimal, outperforms both of these controllers, as the simulation results also corroborate; it is however more complex than either Controller 1 or Controller 2 of [2].

As can be deduced from the preceding discussion, the problem actually admits multiple certainty-equivalent solutions, all of which however not being optimal. Here we obtain the optimal certainty-equivalent solution, and show that this is actually globally team optimal. The derivation of the team-optimal controller has been carried out rigorously, and the controller is expressed in closed form, which is linear in the two user case. For the general multiple-user case, we present an algorithm for the computation of the optimal solution, and present various simulation results to support and illustrate the approach taken.

This paper is organized as follows. In Section 2, we introduce and motivate the mathematical model. The derivation of the optimal controller for the two user case is presented in Section 3, and Section 4 extends this derivation to the multiple user case. Section 5 is devoted to presentation of the simulation results. The paper ends with the concluding remarks of Section 6.

2 Mathematical Model

The mathematical formulation of the problem follows along the lines of [1], [2]. We consider a set $\mathcal{M} = \{1, \ldots, M\}$ of users that share a common bottleneck node in a network. In our model, the time unit corresponds to the round trip delay, which is the time it takes for a packet to reach its destination and come back. Let q_n denote the queue length at the bottleneck link, and μ_n denote the effective service rate available in that link at the beginning of the nth time slot. Let $r_{m,n}$ denote the effective rate of user $m \in \mathcal{M}$ during the nth time slot. This rate may actually be the outcome of an action taken by the user m several time steps earlier. In our formulation, we will not recognize this delay explicitly, but instead include a delay factor in the information available to each user for the construction of transmission rates for that user. As shown in [2], these two formulations are equivalent.

Now, in terms of the notation introduced, the queue length evolves according to

$$q_{n+1} = q_n + \sum_{m=1}^{M} r_{m,n} - \mu_n \tag{1}$$

The above equation corresponds to a linearized version of the actual queue dynamics. Specifically, we ignore the fact that the queue length is restricted to be positive. Also, we assume no upper bound on the queue length. Simulations in [2], [3] show that these are valid assumptions. The service rate μ_n available to the M sources may change over time in an unpredictable way. We model this by a p-dimensional stable AR process:

$$\mu_n = \mu + \xi_n, \;\; \xi_n = \sum_{i=1}^{p} \alpha_i \xi_{n-i} + \phi_{n-1}$$

where μ is the constant nominal service rate (known to all sources), and α_i, $i = 1, \dots, p$, are known parameters. $\{\phi_n\}_{n \geq 1}$ is a zero mean *i.i.d* sequence with finite variance σ_ϕ^2.

The objective function to be minimized (collectively by all M users) is

$$J = \limsup_{N \to \infty} \frac{1}{N} E \left\{ \sum_{n=1}^{N} \left[(q_n - Q)^2 + \sum_{m=1}^{M} c_m^2 \left(r_{m,n} - a_m \mu_n \right)^2 \right] \right\}$$

where Q is the target queue length, $\sum_{m=1}^{M} a_m = 1$, and c_m's are positive constants. The first additive term above represents a penalty for deviating from a desirable queue length. The second additive term is a measure of the quality with which the input rate for each user tracks a given fraction of the available service rate, where c_m's are weighting factors that serve to prioritize relative importance of these. For example, if we desire *fair* sharing of the available bandwidth, we would choose

$$a_1 = a_2 = \ldots = a_M = \frac{1}{M}$$

assuming that everything else is symmetric for the sources.

The information available to user m at time n is I_{n-D_m}, where

$$I_n := \{q_1, q_2, \dots, q_n; \mu_1, \mu_2, \dots, \mu_n\}$$

and D_m's denote delays in the acquisition of queue length and service rate information. Without any loss of generality, we take D_m's to be ordered in the following way:

$$0 \leq D_1 \leq D_2 \leq \ldots \leq D_M \tag{2}$$

Hence,

$$r_{m,n} = \gamma_{m,n} \left(I_{n-D_m} \right), \; n = 1, 2, \dots ; \; m = 1, 2, \dots, M$$

where $\gamma_{m,n}$ are some measurable functions, with respect to which J will be minimized. For convenience, we introduce the new (appropriately shifted) variables

$$\begin{aligned} x_n &:= q_n - Q \\ u_{m,n} &:= r_{m,n} - a_m \mu \end{aligned} \tag{3}$$

which will serve as the state and control, respectively. The queue dynamics (1) can be re-written in terms of these quantities as:

$$x_{n+1} = x_n + \sum_{m=1}^{M} u_{m,n} - \xi_n \tag{4}$$

$$\xi_{n+1} = \sum_{i=1}^{p} \alpha_i \xi_{n+1-i} + \phi_n \tag{5}$$

and the cost function J as:

$$J = \limsup_{N \to \infty} \frac{1}{N} E \left\{ \sum_{n=1}^{N} \left[(x_n)^2 + \sum_{m=1}^{M} c_m^2 \left(u_{m,n} - a_m \xi_n \right)^2 \right] \right\}$$

What we have here is a decentralized optimal control problem with a partially nested information structure, and we know that the optimal solution to any finite horizon ($N < \infty$) version of this problem is linear in the available queue length and rate information [7]. If we ignore the delays in the system and solve the corresponding standard regulator problem, when it comes to incorporate the delays into the controllers again, we face the problem of non-uniqueness of the representation of the full-information (no delay) optimal controller, as discussed in [1]. Therefore, there are indeed many ways to define certainty-equivalent controllers and each such controller could lead to a different value of the performance index [1], [4].

In the next section, we first derive the finite-horizon optimal controller for the two user case. We further show that the infinite horizon version of the two user problem admits a solution, which is linear in information variables. Section 4 discusses the extension of this derivation to M users. In particular, we present an easily implementable algorithm to calculate the optimal control action of each user for the finite horizon problem. Our simulation results strongly suggest that the infinite horizon version of the M user problem admits a solution, but we do not yet have a proof that the solution indeed exists in this case.

3 Derivation of Optimal Decentralized Flow Controllers for the Two User Case

3.1 Finite Horizon Optimal Controller

In this section, we consider the two user, N-stage problem where delay of user 1 is D_1 units, and that of user 2 is D_2 units. Define the relative delay of information between the users as

$$D := D_2 - D_1$$

Here, without any loss of generality we can assume that $D_2 > D_1$, and we do not consider the trivial case $D_2 = D_1$, because in this case the problem can be reduced to a standard decentralized optimum control problem whose solution can be found easily. Our objective is to minimize the finite horizon cost

$$J^N = E\left\{ \sum_{n=D_2+1}^{N} \left[x_{n+1}^2 + c_1^2 \left(u_{1,n} - a_1\xi_n\right)^2 + c_2^2 \left(u_{2,n} - a_2\xi_n\right)^2 \right] + \sum_{n=D_1+1}^{D_2} \left[x_{n+1}^2 + c_1^2 \left(u_{1,n} - \xi_n\right)^2 \right] \right\}$$

where $a_1 + a_2 = 1$ and c_1, c_2 are positive weights.

3.1.1 Derivation of the Controller

We use a dynamic programming approach to obtain the solution to the above problem. More precisely, we start at time N and keep minimizing the expected cost backwards in time. In determining the control actions one can use completion of squares to see how the cost accumulates. Clearly, for $n = N, N-1, \dots, N-D+1$ the minimization will be over a single variable, namely $u_{1,n}$. The reason for this is that the information field of user 1 at time n will be equivalent to that of user 2 at time $n + D$. Thus, until we reach the time $N - D$, there will be a sort of transient minimization process in which only the control actions of user 1 will be determined. At every step down to $N-D$, after we complete the squares in $u_{1,n}$ and pick the appropriate control, some terms remain which contribute to the cost of the succeeding step. If we study the problem carefully, it follows that these terms have a certain structure. To this end, we first minimize J with respect to $u_{1,N}$ at stage N. If we denote the cost to be minimized at stage n by J_n, we have, for $n = N$:

$$\begin{aligned} J_N &= x_{N+1}^2 + c_1^2 \left(u_{1,N} - a_1\xi_N\right)^2 \\ &= \left(x_N + u_{1,N} + u_{2,N} - \xi_N\right)^2 + c_1^2 \left(u_{1,N} - a_1\xi_N\right)^2 \end{aligned}$$

Completion of squares yields

$$J_N = \left(1+c_1^2\right)\left[u_{1,N} + \frac{1}{1+c_1^2}\left(x_N + u_{2,N}\right) - \left(1 + \frac{c_1^2\left(a_1-1\right)}{1+c_1^2}\right)\xi_N\right]^2 + \frac{c_1^2}{1+c_1^2}\left[u_{2,N} + x_N + \left(a_1-1\right)\xi_N\right]^2 \tag{6}$$

where $u_{1,N}$ must be selected such that the expected value of the first additive term in (6) is minimized. In fact, what we are interested in is the term that remains after the completion of squares, because, as we will see shortly, these terms determine the structure of the controllers when a steady state is reached. Let R_n denote these remaining terms at stage n. Thus the second additive term in (6) is R_N. In picking the control $u_{1,N}$, we do not need to consider the second term in J_N, since it will be taken care of at the next stage. Note that, user 1 at time N knows the control $u_{2,N}$, because $u_{1,N}$ has access to the information field of $u_{2,N}$. Therefore, the only difficulty in this minimization arises from the unknowns x_N and ξ_N. Since ϕ_n's are independent random variables, the best $u_{1,N}$ can do is to replace ξ_n's for $n = N, N-1, \ldots, N-D_1+1$ with their best estimates conditioned on its own information field. That is,

$$\hat{\xi}^N_{1,n|N-D_1} = E\left\{\xi_n \mid I_{N-D_1}\right\},\ n = N, N-1, \ldots, N-D_1+1 \tag{7}$$

The reason for this can be seen if we substitute for x_N from the state equation (4). The result is

$$J_N = \left(1+c_1^2\right)\left[u_{1,N} - \left(1 + \frac{c_1^2\left(a_1-1\right)}{1+c_1^2}\right)\xi_N + \frac{1}{1+c_1^2}\left(x_{N-D_1} + u_{2,N} + \sum_{n=1}^{D_1}\left(u_{1,N-n} + u_{2,N-n} - \xi_{N-n}\right)\right)\right]^2 \tag{8}$$

In (8) the only unknowns to $u_{1,N}$ are $\xi_N, \xi_{N-1}, \ldots, \xi_{N-D_1+1}$. We already know that ξ_n's have a linear dynamics given by (5). Thus, (8) can be rewritten as

$$J_N = \left(1+c_1^2\right)\left[u_{1,N} - \frac{1}{1+c_1^2}\sum_{n=1}^{D_1-1}\xi_{N-n} - \left(1 + \frac{c_1^2\left(a_1-1\right)}{1+c_1^2}\right)\xi_N + \frac{1}{1+c_1^2}\left(x_{N-D_1} + u_{2,N} + \xi_{N-D_1} + \sum_{n=1}^{D_1}\left(u_{1,N-n} + u_{2,N-n}\right)\right)\right]^2$$

One can use (5) to express the ξ_n's in terms of ξ_{N-D_1} plus some zero mean random variable with a known variance. Thus, the best decision function $u_{1,N}$ is solely determined by the first term in (6) with unknown states replaced from the state equation (4) and unknown ξ_n's replaced with their best estimates given by (7). Our task at stage N ends with stating what R_N is

$$R_N = \gamma_0 \left(u_{2,N} + \rho_{0,1} x_N + (a_1 - 1)\, \xi_N\right)^2$$

where $\gamma_0 = \frac{c_1^2}{1+c_1^2}$ and $\rho_{0,1} = 1$.

The next stage is $n = N-1$. Assuming the relative delay of information between the users is larger than one time unit, we complete the squares in $u_{1,N-1}$ only. The cost functional to be minimized is

$$J_{N-1} = x_N^2 + c_1^2 \left(u_{1,N-1} - a_1 \xi_N\right)^2 + R_N$$

Completion of squares results in three additive terms and two of them are transferred to the next stage. The first additive term determines the control $u_{1,N-1}$ at this stage. As in the previous step, the best $u_{1,N-1}$ can do is to replace ξ_n's for $n = N-1, N-2, \ldots, N-D_1$ with their best estimates conditioned on its information field. These estimates are given by

$$\hat{\xi}_{1,n|N-D_1-1}^{N-1} = E\left\{\xi_n \mid I_{N-D_1-1}\right\}, \;\; n = N-1, N-2, \ldots, N-D_1$$

and R_{N-1} can be found to be

$$\begin{aligned} R_{N-1} = \gamma_1 (&u_{2,N} + \rho_{1,1} x_{N-1} + \rho_{1,1} u_{2,N-1} + \rho_{1,1} (a_1 - 1)\, \xi_{N-1} \\ &+ (a_1 - 1)\, \xi_N)^2 + \gamma_0 \left(u_{2,N-1} + \rho_{0,1} x_{N-1} + (a_1 - 1)\, \xi_{N-1}\right)^2 \end{aligned}$$

We proceed in this manner until we reach the time $N - D + 1$. From this point on, we start minimizing J_n's over two independent variables, namely $u_{1,n}$ and $u_{2,n+D}$. It is easy to calculate R_n's for the so-called transient part of this minimization process extending from time N to $N - D + 1$. For example, R_{N-2} is

$$\begin{aligned} R_{N-2} = \gamma_2 (&u_{2,N} + \rho_{2,1} x_{N-2} + \rho_{2,1} u_{2,N-2} + \rho_{2,1} (a_1 - 1)\, \xi_{N-2} \\ &+\rho_{2,2} u_{2,N-1} + \rho_{2,2} (a_1 - 1)\, \xi_{N-1} + (a_1 - 1)\, \xi_N)^2 \\ &+\gamma_1 (u_{2,N-1} + \rho_{1,1} x_{N-2} + \rho_{1,1} u_{2,N-2} + \rho_{1,1} (a_1 - 1)\, \xi_{N-2} \\ &\quad + (a_1 - 1)\, \xi_{N-1})^2 \\ &+\gamma_0 \left(u_{2,N-2} + \rho_{0,1} x_{N-2} + (a_1 - 1)\, \xi_{N-2}\right)^2 \end{aligned}$$

and R_{N-3} equals

$$\begin{aligned} R_{N-3} = \gamma_3 (&u_{2,N} + \rho_{3,1} x_{N-3} + \rho_{3,1} u_{2,N-3} + \rho_{3,1} (a_1 - 1)\, \xi_{N-3} \\ &+\rho_{3,2} u_{2,N-2} + \rho_{3,2} (a_1 - 1)\, \xi_{N-2} + \rho_{3,3} u_{2,N-1} \\ &+\rho_{3,3} (a_1 - 1)\, \xi_{N-1} + (a_1 - 1)\, \xi_N)^2 \end{aligned}$$

$$+\gamma_2 \left(u_{2,N-1} + \rho_{2,1} x_{N-3} + \rho_{2,1} u_{2,N-1} + \rho_{2,1} (a_1 - 1) \xi_{N-3} \right.$$
$$\left. + \rho_{2,2} u_{2,N-2} + \rho_{2,2} (a_1 - 1) \xi_{N-2} + (a_1 - 1) \xi_{N-1}\right)^2$$
$$+\gamma_1 \left(u_{2,N-2} + \rho_{1,1} x_{N-3} + \rho_{1,1} u_{2,N-3} + \rho_{1,1} (a_1 - 1) \xi_{N-3} \right.$$
$$\left. + (a_1 - 1) \xi_{N-2}\right)^2$$
$$+\gamma_0 (u_{2,N-3} + \rho_{0,1} x_{N-3} + (a_1 - 1) \xi_{N-3})^2$$

Note that the evolution of R_n obeys a certain structure. One can exploit this structure to determine the constants γ_d's and ρ_d's:

$$\tau_{-1} = 1 + c_1^2,\ \tau_d = 1 + c_1^2 + \sum_{n=0, D\geq 2}^{d} \gamma_n \rho_{n,1}^2, \gamma_{-1} = 0,\ \gamma_0 = \frac{c_1^2}{1 + c_1^2}$$

$$\rho_{0,1} = 1,\ \gamma_{d+1} = \gamma_d - \frac{\gamma_d^2 \rho_{d,1}^2}{\tau_d},\ d = 0, 1, \ldots, D-2,\ D \geq 2$$

$$\rho_{d+1,1} = \frac{\gamma_d \rho_{d,1}}{\gamma_{d+1}} \left(1 - \frac{1 + \gamma_0 \rho_{0,1} + \gamma_1 \rho_{1,1}^2 + \ldots + \gamma_d \rho_{d,1}^2}{\tau_d}\right)$$
$$,\ d = 0, 1, \ldots, D-2,\ D \geq 2$$

$$\rho_{d+1,2} = \frac{\gamma_d \rho_{d,1}}{\gamma_{d+1}} \left(1 - \frac{\gamma_0 \rho_{0,1} + \gamma_1 \rho_{1,1}^2 + \ldots + \gamma_d \rho_{d,1}^2}{\tau_d}\right)$$
$$,\ d = 1, 2, \ldots, D-2,\ D \geq 3$$

$$\rho_{d+1,3} = \frac{\gamma_d \rho_{d,2}}{\gamma_{d+1}} - \frac{\gamma_d \rho_{d,1}}{\gamma_{d+1}} \left(\frac{\gamma_1 \rho_{1,1} + \gamma_2 \rho_{2,1} \rho_{2,2} + \ldots + \gamma_d \rho_{d,1} \rho_{d,2}}{\tau_d}\right)$$
$$,\ d = 2, 3, \ldots, D-2,\ D \geq 4$$

$$\vdots$$

$$\rho_{d+1,k+1} = \frac{\gamma_d \rho_{d,k}}{\gamma_{d+1}} - \frac{\gamma_d \rho_{d,1}}{\gamma_{d+1}} \left(\frac{\gamma_{k-1} \rho_{k-1,1} + \gamma_k \rho_{k,1} \rho_{k,k} + \ldots + \gamma_d \rho_{d,1} \rho_{d,k}}{\tau_d}\right)$$
$$,\ d = k, k+1, \ldots, D-2,\ D \geq k+2,\ k = 4, \ldots, D-2$$

For $n = N, N-1, \ldots, N-D+1$, the cost functionals we minimize have the form

$$J_n = x_{n+1}^2 + c_1^2 (u_{1,n} - a_1 \xi_n)^2 + R_{n+1} \tag{9}$$

with $R_{N+1} = 0$. And the optimal control for this time interval can be written as

$$u^*_{1,N-d} = -\sigma_{0,d}\left(\hat{x}^{N-d}_{1,N-d|N-D_1-d} + u_{2,N-d}\right)$$
$$- \sum_{k=1,d\neq 0}^{d} \sigma_{k,d}\left(u_{2,N-d+k} + (a_1-1)\,\hat{\xi}^{N-d}_{1,N-d+k|N-D_1-d}\right)$$
$$+ (a_1 - \sigma_{0,d}(a_1-1))\,\hat{\xi}^{N-d}_{1,N-d|N-D_1-d},\ d = 0,\dots,D-1 \quad (10)$$

where

$$\sigma_{0,d} = \frac{\tau_{d-1} - c_1^2}{\tau_{d-1}},\ d \geq 0$$
$$\sigma_{1,d} = \frac{\gamma_0\rho_{0,1} + \gamma_1\rho_{1,1}^2 + \dots + \gamma_{d-1}\rho_{d-1,1}^2}{\tau_{d-1}},\ d \geq 1$$
$$\sigma_{2,d} = \frac{\gamma_1\rho_{1,1} + \gamma_2\rho_{2,1}\rho_{2,2} + \dots + \gamma_{d-1}\rho_{d-1,1}\rho_{d-1,2}}{\tau_{d-1}},\ d \geq 2$$
$$\vdots$$
$$\sigma_{k,d} = \frac{\gamma_{k-1}\rho_{k-1,1} + \gamma_k\rho_{k,1}\rho_{k,k} + \dots + \gamma_{d-1}\rho_{d-1,1}\rho_{d-1,k}}{\tau_{d-1}},\ d \geq k$$
$$,\ k = 4,\dots,D-1$$

In the above set of equations, $\hat{\xi}^l_{1,n|k}$ denotes source 1's estimate at time l of the value of ξ at time n based on I_k. A similar interpretation holds for $\hat{x}^l_{1,n|k}$.

Now, the next step is to calculate the optimal control laws for $D_2 + 1 \leq n \leq N - D$. First, we observe that the cost functional to be minimized at any stage after D steps back in time is

$$J_n = [x_n + u_{1,n} + u_{2,n} - \xi_n]^2 + c_1^2[u_{1,n} - a_1\xi_n]^2 + c_2^2[u_{2,n+D} - a_2\xi_{n+D}]^2$$
$$+\eta_n[u_{2,n+D} + \rho_{D-1,1}x_{n+1} + \rho_{D-1,1}u_{2,n+1} + \rho_{D-1,1}(a_1-1)\xi_{n+1}$$
$$+\rho_{D-1,2}u_{2,n+2} + \rho_{D-1,2}(a_1-1)\xi_{n+2}$$
$$+\dots+\rho_{D-1,D-1}u_{2,n+D-1} + \rho_{D-1,D-1}(a_1-1)\xi_{n+D-1}$$
$$+(a_1-1)\xi_{n+D}]^2$$

$$
\begin{aligned}
&+\gamma_{D-2}\left[u_{2,n+D-1}+\rho_{D-2,1}x_{n+1}+\rho_{D-2,1}u_{2,n+1}\right.\\
&\qquad+\rho_{D-2,1}\left(a_1-1\right)\xi_{n+1}+\rho_{D-2,2}u_{2,n+2}\\
&\qquad+\rho_{D-2,2}\left(a_1-1\right)\xi_{n+2}+\ldots+\rho_{D-2,D-2}u_{2,n+D-2}\\
&\qquad\left.+\rho_{D-2,D-2}\left(a_1-1\right)\xi_{2,n+D-2}+\left(a_1-1\right)\xi_{n+D-1}\right]^2\\
&+\ldots+\gamma_1\left[u_{2,n+2}+\rho_{1,1}x_{n+1}+\rho_{1,1}u_{2,n+1}+\rho_{1,1}\left(a_1-1\right)\xi_{n+1}\right.\\
&\qquad\left.+\left(a_1-1\right)\xi_{n+2}\right]^2\\
&+\gamma_0\left[u_{2,n+1}+\rho_{0,1}x_{n+1}+\left(a_1-1\right)\xi_{n+1}\right]^2 \qquad (11)
\end{aligned}
$$

where

$$
\eta_{N-D}=\gamma_{D-1} \qquad (12)
$$

A recursive formula to calculate η_n's can be found after some algebraic manipulations:

$$
\eta_{n-1}=\frac{A_1\eta_n+A_0}{B_1\eta_n+B_0},\ n=N-D,N-D-1,\ldots,D_2+1 \qquad (13)
$$

where

$$
\begin{aligned}
A_1 &= \left(\gamma_{D-2}+c_2^2\rho_{D-1,D-1}^2\right)\left(1+c_1^2+\sum_{k=3,D\geq 3}^{D}\gamma_{D-k}\rho_{D-k,1}^2\right)\\
&+ c_2^2\gamma_{D-2}\left(\rho_{D-1,1}-\rho_{D-2,1}\rho_{D-1,D-1}\right)^2\\
A_0 &= c_2^2\gamma_{D-2}\left(1+c_1^2+\sum_{k=3,D\geq 3}^{D}\gamma_{D-k}\rho_{D-k,1}^2\right)\\
B_1 &= \left(1+c_1^2+c_2^2\rho_{D-1,1}^2+\sum_{k=2,D\geq 2}^{D}\gamma_{D-k}\rho_{D-k,1}^2\right)\\
B_0 &= c_2^2\left(1+c_1^2+\sum_{k=2,D\geq 2}^{D}\gamma_{D-k}\rho_{D-k,1}^2\right) \qquad (14)
\end{aligned}
$$

Note that all these four quantities are positive. Now, define $\psi_{n|D}$ for a given relative delay D as

$$
\psi_{n|D}:=1+c_1^2+\eta_n\rho_{D-1,1}^2+\gamma_{D-2}\rho_{D-2,1}^2+\ldots+\gamma_0\rho_{0,1}^2
$$

The optimal controls can be found by minimizing (11) over $u_{1,n}$ and $u_{2,n+D}$. Before writing down the complete solution, let us introduce some definitions. For given n and D, let $\pi_{k|n,D}$ be defined by

$$\pi_{0|n,D} := \frac{\eta_n \rho_{D-1,1}}{\psi_{n|D}}$$

$$\pi_{1|n,D} := \frac{\eta_n \rho_{D-1,1}\rho_{D-1,D-1} + \gamma_{D-2}\rho_{D-2,1}}{\psi_{n|D}}$$

$$\pi_{k|n,D} := \frac{\eta_n \rho_{D-1,1}\rho_{D-1,D-k} + \sum_{j=2}^{k}\left[\gamma_{D-j}\rho_{D-j,1}\rho_{D-j,D-k}\right]}{\psi_{n|D}} + \frac{\gamma_{D-k-1}\rho_{D-k-1,1}}{\psi_{n|D}},\ k = 2, 3, \dots, D-1$$

$$\pi_{D|n,D} := \pi_{D-1|n,D} + \frac{1}{\psi_{n|D}},$$

and let $\lambda_{n|D}$ be defined as

$$\lambda_{n|D} := c_2^2 + \eta_n - \frac{\eta_n^2 \rho_{D-1,1}^2}{\psi_{n|D}}$$

Then, the optimum control $u^*_{2,n+D}$ becomes

$$u^*_{2,n+D} = -\frac{\eta_n \rho_{D-1,1} - \psi_{n|D}\pi_{0|n,D}\pi_{D|n,D}}{\lambda_{n|D}} \Big(\hat{x}^{n+D}_{2,n|n-D_1} + u_{2,n} + (a_1 - 1)\,\hat{\xi}^{n+D}_{2,n|n-D_1}\Big) - (a_1 - 1)\,\hat{\xi}^{n+D}_{2,n+D|n-D_1} - \sum_{k=1, D\geq 1}^{D-1} \Bigg[\frac{\eta_n \rho_{D-1,k} - \psi_{n|D}\pi_{0|n,D}\pi_{D-k|n,D}}{\lambda_{n|D}} \Big(u_{2,n+k} + (a_1 - 1)\,\hat{\xi}^{n+D}_{2,n+k|n-D_1}\Big)\Bigg],\ n = N-D, \dots, D_2 + 1 \qquad (15)$$

Similarly, one can write down the control $u^*_{1,n}$ as

$$u^*_{1,n} = -\pi_{0|n,D}(u^*_{2,n+D} + (a_1 - 1)\hat{\xi}^n_{1,n+D|n-D_1}) - \sum_{k=1}^{D-1}\Big[\pi_{k|n,D}\Big(u_{2,n+D-k} + (a_1 - 1)\hat{\xi}^n_{1,n+D-k|n-D_1}\Big)\Big] - \pi_{D|n,D}(\hat{x}^n_{1,n|n-D_1} + u_{2,n}) + (1 + \frac{c_1^2(a_1 - 1)}{\psi_{n|D}})\hat{\xi}^n_{1,n|n-D_1},\ n = N-D, \dots, D_2 + 1 \qquad (16)$$

Note that, $u^*_{2,n+D}$ appears in the expression for $u^*_{1,n}$, but this is not necessary, as one can substitute (15) into (16) and have a formula for $u^*_{1,n}$ expressed only in terms of its own information variables. The calculation of the controllers for the last D steps of the problem is omitted, but it can be carried out without much effort.

3.1.2 Certainty Equivalence

The optimal controller derived above is certainty-equivalent in the sense that if users 1 and 2 had access to perfect state information, they would simply replace the estimated variables in (15) and (16) with their actual values. In fact, this follows from the construction of the controller. In the derivations, at each stage we replaced the unknown states with their best estimates conditioned on users' own information fields at that stage. Hence, if we assume that these unknown states are known to the users, the best they can do is to use this new information in their expressions of control. As we mentioned earlier, there are many ways in which a certainty-equivalent controller can be defined and each such controller leads to a different value for the cost [1], [4]. Of course, among these different controllers (each one being a particular *representation* of the perfect state information controller), the one we constructed above has the lowest possible cost due to its optimality.

3.2 Infinite Horizon Optimal Controller

We now return to the original infinite-horizon problem with two users. Our objective was actually to minimize the infinite-horizon cost

$$\begin{aligned} J \quad &= \quad \limsup_{N\to\infty} \frac{1}{N} E\left\{ \sum_{n=D_2+1}^{N} \left[x_{n+1}^2 + c_1^2 \left(u_{1,n} - a_1\xi_n\right)^2 + c_2^2 \left(u_{2,n} - a_2\xi_n\right)^2 \right] \right. \\ &+ \left. \sum_{n=D_1+1}^{D_2} \left[x_{n+1}^2 + c_1^2 \left(u_{1,n} - \xi_n\right)^2 \right] \right\} \qquad (17) \end{aligned}$$

Here we prefer to write lim sup instead of just lim, because we do not know yet whether the limit exists. To be able to prove the existence of the infinite-horizon optimal controller, we first show that the controller dynamics given by (12) and (13) converge as N tends to ∞. Next, we use this result to show that the optimal finite-horizon controller is stabilizing as $N \to \infty$. And finally, we prove that the optimum value of (17), denoted by J^*, equals to the limit of the optimum finite horizon average cost. Thus, the linear stationary policies which are given as the limit of the optimal policies (15) and (16) are optimal for the infinite-horizon problem.

3.2.1 Convergence of Controller Dynamics

In an infinite horizon problem, one usually wants to know whether the recursively defined sequence (13) has a limit point for arbitrary values of system parameters. And if so, we have to make sure that this limit point is unique. In what follows we address these questions. First of all, note that, A_1, A_0, B_1 and B_0 are all positive constants. The initial condition of (13) being a positive constant, η_n's remain positive for all n. To investigate the existence and uniqueness of the limit we formulate the problem as follows. The sequence of points given by (13) is computed in reverse time by a formula of the form

$$\eta_{n-1} = F\left(\eta_n\right), \; n = N-D, N-D-1, \dots, D_2+1$$

where

$$F\left(\eta\right) = \frac{A_1\eta + A_0}{B_1\eta + B_0} \tag{18}$$

Note that, F is positive and continuous for $\eta \geq 0$, since A_1, A_0, B_1, B_0 are all positive. For convenience, as $N \to \infty$, we take the sequence $\{\eta_n\}$ to be defined in forward time as

$$\eta_{n+1} = F\left(\eta_n\right) \tag{19}$$

with the initial value $\eta_0 = \gamma_{D-1}$. This new definition does not change anything about the existence and uniqueness of the limit. Now, assume that the sequence $\{\eta_n\}$ has at least one limit point denoted by η_∞. If so, we must have η_n's converging to this number, and hence to a solution of the equation:

$$\eta_\infty = \frac{A_1\eta_\infty + A_0}{B_1\eta_\infty + B_0}$$

which is equivalent to the following quadratic equation in η_∞:

$$B_1\eta_\infty^2 + \left(B_0 - A_1\right)\eta_\infty - A_0 = 0 \tag{20}$$

whose discriminant is

$$\Delta = \left(B_0 - A_1\right)^2 + 4B_1A_0 > 0 \tag{21}$$

Thus, we have two real roots. The signs of the roots are opposite, because their product, given by $-\frac{A_0}{B_1}$, is negative. As a result, we can conclude that we have a single positive real root of (20), designated by η_∞^+. Note that, η_∞^+ is a *fixed point* of (18), and is unique on the interval $[0, \infty)$. By the *contraction mapping theorem* [8], this point is indeed the unique limit of every sequence obtained from (19) with any nonnegative starting point if F is a contraction mapping on $[0, \infty) \cup \{\infty\}$. We proceed by showing that F is indeed a contraction on this closed interval. To this end we seek a number $\beta < 1$, such that

$$\left|F\left(x\right) - F\left(y\right)\right| \leq \beta\left|x - y\right|, \; \forall x, y \geq 0 \tag{22}$$

Plugging (18) into (22) yields

$$\frac{|x-y|\,|A_0B_1-A_1B_0|}{(B_1x+B_0)\,(B_1y+B_0)} \leq \beta\,|x-y|$$

Cancelling $|x-y|$'s and using the facts that the minima of (B_1x+B_0) and (B_1y+B_0) over $[0,\infty)$ occur at $x=0$ and $y=0$ respectively, and B_0^2 is positive, we can rewrite the condition of contraction as

$$|A_0B_1-A_1B_0| - B_0^2 < 0$$

Next, we substitute for A_0, B_0, A_1 and B_1 from (14) and after some algebraic manipulations we get

$$\begin{aligned}
&\Bigg[(\rho_{D-1,D-1}-1)\left(\gamma_{D-2}\rho_{D-2,1}^2+1+c_1^2+\sum_{k=3,D\geq 3}^{D}\gamma_{D-k}\rho_{D-k,1}^2\right)\\
&-\gamma_{D-2}\rho_{D-1,1}\rho_{D-2,1}\Bigg]\times\\
&\Bigg[\rho_{D-1,D-1}\left(\gamma_{D-2}\rho_{D-2,1}^2+1+c_1^2+\sum_{k=3,D\geq 3}^{D}\gamma_{D-k}\rho_{D-k,1}^2\right)\\
&+\left(\gamma_{D-2}\rho_{D-2,1}^2+1+c_1^2+\sum_{k=3,D\geq 3}^{D}\gamma_{D-k}\rho_{D-k,1}^2\right)\\
&-\gamma_{D-2}\rho_{D-1,1}\rho_{D-2,1}\Bigg] < 0
\end{aligned} \tag{23}$$

The first multiplicative term in (23) is negative, because $(\rho_{D-1,D-1}-1) < 0$ for any D, and the second multiplicative term is positive, since we have $\gamma_{D-2}\rho_{D-1,1}\rho_{D-2,1} < 1$ by construction. As a result, the inequality in (23) is satisfied for arbitrary values of system parameters, but (23) is equivalent to (22), and thus F is a contraction mapping. Hence, the contraction mapping theorem applies and the fixed point η_∞^+ turns out to be the unique limit point of the sequence $\{\eta_n\}$. The value of η_∞^+ can be calculated as the positive root of (20):

$$\eta_\infty^+ = \frac{(A_1-B_0)+\sqrt{(B_0-A_1)^2+4B_1A_0}}{2B_1}$$

Now, the stationary optimal policies can be found by plugging the limiting value of the sequence $\{\eta_n\}$ into (15) and (16). We will see shortly that these stationary policies indeed constitute the solution of the infinite-horizon version of the problem, but first we prove another useful result which is also used in showing the existence of a solution to the infinite-horizon problem.

3.2.2 Stabilizing Property of the Optimal Controller

In this subsection, we investigate the stabilizing property of the optimal controller and show that as N goes to infinity, the average cost remains bounded. An immediate consequence of this result is that neither the shifted queue dynamics nor the user rates can blow up when the optimal controller is applied. In [1] it was shown that the average cost is bounded for two suboptimal certainty-equivalent controllers, referred to as *Controller 1* and *Controller 2*. For instance, for Controller 1 we have

$$J^* \leq \lim_{N\to\infty} \frac{J^N}{N} \leq C_1$$

where C_1 is the scalar given in Section 5 of [1]. The controller here being optimal leads to an average cost that is no larger. Therefore,

$$J^* \leq \lim_{N\to\infty} \frac{J^N}{N} \leq C_{opt} \leq C_1$$

which proves the stabilizability of the optimal controller.

Here we present a more direct proof of this fact by calculating the exact value of the cost per stage. Starting at time N, until we reach the time $N-D$, the control is given by (10), and for any given D the cost of these first D stages, denoted by L_∞, can be found by plugging (10) into (9) and it will depend on the parameters of the AR process as well as the control (10). It is clear that L_∞ is finite, because we have only a finite number of steps until the time $N-D+1$. We also note that the optimal control laws for $D_2+1 \leq n \leq N-D$ were found by minimizing J_n's over $u_{1,n}$ and $u_{2,n+D}$. If we complete J_n to squares in $u_{1,n}$ and $u_{2,n+D}$ and substitute for the optimal controls (15) and (16), the expected value of the resulting expression gives us a cost that we cannot avoid due to the delays in the system, plus some remaining terms that are transferred to the next stage of the minimization process. These remaining terms are taken care of in minimizing J_{n-1}. Thus, the cost of stage n essentially equals to the expected value of the following quantity

$$\begin{aligned}\Omega_n = \psi_{n|D} \Bigg[& u^*_{1,n} + \sum_{k=1}^{D-1} \left[\pi_{k|n,D}\left(u_{2,n+D-k} + (a_1-1)\,\xi_{n+D-k}\right)\right] \\ & + \pi_{0|n,D}\left(u^*_{2,n+D} + (a_1-1)\,\xi_{n+D}\right) + \pi_{D|n,D}\left(x_n + u_{2,n}\right) \\ & - \left(1 + \frac{c_1^2\,(a_1-1)}{\psi_{n|D}}\right)\xi_n \Bigg]^2\end{aligned}$$

$$+\lambda_{n|D}\left[u^*_{2,n+D}+\frac{\eta_n\rho_{D-1,1}-\psi_{n|D}\pi_{0|n,D}\pi_{D|n,D}}{\lambda_{n|D}}\left(x_n+u_{2,n}\right.\right.$$

$$+\left(a_1-1\right)\xi_n\big)+\left(a_1-1\right)\xi_{n+D}$$

$$+\sum_{k=1,D\geq 1}^{D-1}\left[\frac{\eta_n\rho_{D-1,k}-\psi_{n|D}\pi_{0|n,D}\pi_{D-k|n,D}}{\lambda_{n|D}}\left(u_{2,n+k}\right.\right.$$

$$\left.\left.\left.+\left(a_1-1\right)\xi_{n+k}\right)\right]\right]^2$$

If we substitute for the optimal controls $u^*_{1,n}$ and $u^*_{2,n+D}$, we get

$$\Omega_n=\psi_{n|D}\left[\pi_{0|n,D}\left(a_1-1\right)\left(\xi_{n+D}-\hat{\xi}^n_{1,n+D|n-D_1}\right)\right.$$

$$+\pi_{D|n,D}\left(x_n-\hat{x}^n_{1,n|n-D-1}\right)+\sum_{k=1}^{D-1}\pi_{k|n,D}\left(a_1-1\right)\left(\xi_{n+D-k}\right.$$

$$\left.\left.-\hat{\xi}^n_{1,n+D-k|n-D_1}\right)-\left(1+\frac{c_1^2\left(a_1-1\right)}{\psi_{n|D}}\right)\left(\xi_n-\hat{\xi}^n_{1,n|n-D-1}\right)\right]^2$$

$$+\lambda_{n|D}\left[\frac{\eta_n\rho_{D-1,1}-\psi_{n|D}\pi_{0|n,D}\pi_{D|n,D}}{\lambda_{n|D}}\left(\left(x_n-\hat{x}^{n+D}_{2,n|n-D_1}\right)\right.\right.$$

$$\left.+\left(a_1-1\right)\left(\xi_n-\hat{\xi}^{n+D}_{2,n|n-D_1}\right)\right)+\left(a_1-1\right)\left(\xi_{n+D}\right.$$

$$\left.-\hat{\xi}^{n+D}_{2,n+D|n-D_1}\right)+\sum_{k=1,D\geq 1}^{D-1}\frac{\eta_n\rho_{D-1,k}-\psi_{n|D}\pi_{0|n,D}\pi_{D-k|n,D}}{\lambda_{n|D}}$$

$$\left.\left(a_1-1\right)\left(\xi_{n+k}-\hat{\xi}^{n+D}_{2,n+k|n-D_1}\right)\right]^2$$

Hence, to calculate the cost at stage n, first we need to express the terms involving the difference between estimated variables and their actual values in a convenient form. Following along the lines of Section 5 of [1], one can show that the expected value of Ω_n can equivalently be written as

$$E\left\{\Omega_n\right\}=\varphi_n E\left\{\phi^2\right\}=\varphi_n\sigma^2_\phi \tag{24}$$

where φ_n is the sequence obtained from Ω_n by using the fact that the estimation errors are linear in $\{\phi_n\}$'s which are *i.i.d* with variance σ^2_ϕ. For a more detailed reasoning behind this idea see [1], Section 5. Now, as n tends to D_2+1, while N tending to ∞, we know that η_n's converge to the unique number η^+_∞. To investigate the limiting behavior of φ_n, we need to consider that of $\pi_{l|n,D}, l=0,\ldots,D$, $\psi_{n|D}$ and $\lambda_{n|D}$, because φ_n for each n is just a continuous function of these sequences. Hence, if one can show that these three sequences converge, then by continuity φ_n's will converge as well. To this end, we first recall that

$$\psi_{n|D} = 1 + c_1^2 + \eta_n \rho_{D-1,1}^2 + \gamma_{D-2}\rho_{D-2,1}^2 + \ldots + \gamma_0 \rho_{0,1}^2 \tag{25}$$

$$\begin{aligned}
\pi_{0|n,D} &:= \frac{\eta_n \rho_{D-1,1}}{\psi_{n|D}} \\
\pi_{1|n,D} &:= \frac{\eta_n \rho_{D-1,1}\rho_{D-1,D-1} + \gamma_{D-2}\rho_{D-2,1}}{\psi_{n|D}} \\
\pi_{k|n,D} &:= \frac{\eta_n \rho_{D-1,1}\rho_{D-1,D-k} + \sum_{j=2}^{k}\left[\gamma_{D-j}\rho_{D-j,1}\rho_{D-j,D-k}\right]}{\psi_{n|D}} \\
&\quad + \frac{\gamma_{D-k-1}\rho_{D-k-1,1}}{\psi_{n|D}},\ k = 2, 3, \ldots, D-1 \\
\pi_{D|n,D} &:= \pi_{D-1|n,D} + \frac{1}{\psi_{n|D}} \qquad (26)
\end{aligned}$$

and

$$\lambda_{n|D} = c_2^2 + \eta_n - \frac{\eta_n^2 \rho_{D-1,1}^2}{\psi_{n|D}} \tag{27}$$

Taking the limit of both sides of (25), (26) and (27) we see that

$$\psi_{\infty|D} = 1 + c_1^2 + \eta_\infty^+ \rho_{D-1,1}^2 + \gamma_{D-2}\rho_{D-2,1}^2 + \ldots + \gamma_0 \rho_{0,1}^2$$

$$\begin{aligned}
\pi_{0|\infty,D} &:= \frac{\eta_\infty^+ \rho_{D-1,1}}{\psi_{\infty|D}} \\
\pi_{1|\infty,D} &:= \frac{\eta_\infty^+ \rho_{D-1,1}\rho_{D-1,D-1} + \gamma_{D-2}\rho_{D-2,1}}{\psi_{\infty|D}} \\
\pi_{k|\infty,D} &:= \frac{\eta_\infty^+ \rho_{D-1,1}\rho_{D-1,D-k} + \sum_{j=2}^{k}\left[\gamma_{D-j}\rho_{D-j,1}\rho_{D-j,D-k}\right]}{\psi_{\infty|D}} \\
&\quad + \frac{\gamma_{D-k-1}\rho_{D-k-1,1}}{\psi_{\infty|D}},\ k = 2, 3, \ldots, D-1 \\
\pi_{D|\infty,D} &:= \pi_{D-1|\infty,D} + \frac{1}{\psi_{\infty|D}}
\end{aligned}$$

and

$$\lambda_{\infty|D} = c_2^2 + \eta_\infty^+ - \frac{{\eta_\infty^+}^2 \rho_{D-1,1}^2}{\psi_{\infty|D}}$$

Since all three limits exist, the sequence $\{\varphi_n\}$ converges to a real number say, φ_∞. Using this result we can write the cost J^N as

$$J^N = \sum_{n=D_1+1}^{N-D} E\{\Omega_n\} + L_\infty = L_\infty + \sigma_\phi^2 \sum_{n=D_1+1}^{N-D} \varphi_n$$

Hence, the average cost as $N \to \infty$ is equal to

$$\lim_{N\to\infty} \frac{J^N}{N} = \lim_{N\to\infty} \frac{1}{N} \left(L_\infty + \sigma_\phi^2 \sum_{n=D_1+1}^{N-D} \varphi_n \right)$$

We can actually evaluate this limit by using the following fact.

Fact: Let $\{a_n\}$ be a convergent sequence with limit a, and let $m < \infty$ be an arbitrary scalar. Then, the following infinite sum

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=m}^{N} a_n$$

converges to the real number a.

Therefore, we must have

$$\lim_{N\to\infty} \frac{J^N}{N} = \sigma_\phi^2 \varphi_\infty$$

which proves that the average cost is bounded if the limit of the finite-horizon optimal controller is applied to the system. This does not necessarily mean that this limiting stationary controller minimize the average infinite horizon cost. In the next subsection, however, we will show that this is indeed the case.

3.2.3 Infinite Horizon Controller

Having shown that the controller dynamics converge and the limiting case of the finite-horizon optimal controller is stabilizing, in this section we prove a stronger result, which enables us to calculate the infinite-horizon optimal controller as the limit of the finite-horizon one. Let M be a positive integer and let $\Pi = \{\Gamma_{D_1+1}, \Gamma_{D_1+2}, \dots, \Gamma_M, \Gamma_{M+1}, \dots\}$ be any admissible policy where each Γ_m consists of two elements: $u_{1,m}$ and $u_{2,m+D}$. We shall denote the optimal policy (15)-(16) by $\Pi^* = \left\{\Gamma^*_{D_1+1}, \Gamma^*_{D_1+2}, \dots, \Gamma^*_M, \Gamma^*_{M+1}, \dots\right\}$. Without any loss of generality, we assume that M is such that the transient part of the dynamic programming is over. First, recall from Section 2 that

$$\min_{u_{1,M}, u_{2,M+D}} E\{S_M + H_M\} = E\{\Omega_M\} + H_{M-1} \tag{28}$$

where

$$S_n := x_{n+1}^2 + c_1^2 \left[u_{1,n} - a_1 \xi_n\right]^2 + c_2^2 \left[u_{2,n+D} - a_2 \xi_{n+D}\right]^2$$

and

$$
\begin{aligned}
H_n := \eta_n & \left[u_{2,n+D} + \rho_{D-1,1} x_{n+1} + \rho_{D-1,1} u_{2,n+1} + \rho_{D-1,1} (a_1 - 1) \xi_{n+1} \right. \\
& + \rho_{D-1,2} u_{2,n+2} + \rho_{D-1,2} (a_1 - 1) \xi_{n+2} \\
& + \ldots + \rho_{D-1,D-1} u_{2,n+D-1} + \rho_{D-1,D-1} (a_1 - 1) \xi_{n+D-1} \\
& \left. + (a_1 - 1) \xi_{n+D}\right]^2 \\
& + \gamma_{D-2} \left[u_{2,n+D-1} + \rho_{D-2,1} x_{n+1} + \rho_{D-2,1} u_{2,n+1} \right. \\
& \quad + \rho_{D-2,1} (a_1 - 1) \xi_{n+1} + \rho_{D-2,2} u_{2,n+2} \\
& \quad + \rho_{D-2,2} (a_1 - 1) \xi_{n+2} + \ldots + \rho_{D-2,D-2} u_{2,n+D-2} \\
& \quad \left. + \rho_{D-2,D-2} (a_1 - 1) \xi_{2,n+D-2} + (a_1 - 1) \xi_{n+D-1}\right]^2 \\
& + \ldots + \gamma_1 \left[u_{2,n+2} + \rho_{1,1} x_{n+1} + \rho_{1,1} u_{2,n+1} + \rho_{1,1} (a_1 - 1) \xi_{n+1} \right. \\
& \quad \left. + (a_1 - 1) \xi_{n+2}\right]^2 \\
& + \gamma_0 \left[u_{2,n+1} + \rho_{0,1} x_{n+1} + (a_1 - 1) \xi_{n+1}\right]^2
\end{aligned}
$$

Using (24), we can rewrite (28) as

$$
\min_{u_{1,M}, u_{2,M+D}} E\left\{S_M + H_M\right\} = \sigma_\phi^2 \varphi_M + H_{M-1} \tag{29}
$$

At this point, for convenience, we introduce a mapping $T_{\Gamma_n}(\cdot)$ which simply evaluates the expected value of its argument when the control policy is given by Γ_n. We have, from (29)

$$
T_{\Gamma_M}(S_M + H_M) \geq \sigma_\phi^2 \varphi_M + h(M-1) \tag{30}
$$

By applying $T_{\Gamma_{M-1}}$ to both sides of (30) we see that

$$
\begin{aligned}
T_{\Gamma_{M-1}}\left(T_{\Gamma_M}(S_M + H_M) + S_{M-1}\right) & \geq T_{\Gamma_{M-1}}\left(\sigma_\phi^2 \varphi_M + H_{M-1} + S_{M-1}\right) \\
& = \sigma_\phi^2 \varphi_M + T_{\Gamma_{M-1}}(H_{M-1} + S_{M-1}) \\
& \geq \sigma_\phi^2 (\varphi_M + \varphi_{M-1}) + H_{M-2}
\end{aligned}
$$

where the first inequality follows from a simple principle of optimality argument. Proceeding in the same manner, we finally obtain

$$
\begin{aligned}
& T_{\Gamma_{D_1+1}}\left(T_{\Gamma_{D_1+2}}\left(\ldots\left(T_{\Gamma_M}(S_M + H_M) + \ldots\right) + S_{D_1+2}\right) + S_{D_1+1}\right) \\
\geq \ & \sigma_\phi^2 \sum_{n=D_1+1}^{M} \varphi_n
\end{aligned} \tag{31}
$$

with equality if each Γ_n, $n = D_1+1, D_1+2, \dots, M$ equals Γ_n^*. Now, the left-hand side of (31) is equal to the M-stage cost corresponding to the policy $\Pi = \{\Gamma_{D_1+1}, \Gamma_{D_1+2}, \dots, \Gamma_M\}$. In other words,

$$\begin{aligned}
&T_{\Gamma_{D_1+1}}\left(T_{\Gamma_{D_1+2}}\left(\dots\left(T_{\Gamma_M}\left(S_M+H_M\right)+\dots\right)+S_{D_1+2}\right)+S_{D_1+1}\right) \\
= \quad & E\left\{\sum_{n=D_2+1}^{M}\left[x_{n+1}^2+c_1^2\left(u_{1,n}-a_1\xi_n\right)^2+c_2^2\left(u_{2,n}-a_2\xi_n\right)^2\right]\right. \\
&\left.\quad + \sum_{n=D_1+1}^{D_2}\left[x_{n+1}^2+c_1^2\left(u_{1,n}-\xi_n\right)^2\right]\middle|\,\Pi\right\}
\end{aligned}$$

One can use this relation in (31) to get

$$\begin{aligned}
& E\left\{\sum_{n=D_2+1}^{M}\left[x_{n+1}^2+c_1^2\left(u_{1,n}-a_1\xi_n\right)^2+c_2^2\left(u_{2,n}-a_2\xi_n\right)^2\right]\right. \\
&\left.\quad + \sum_{n=D_1+1}^{D_2}\left[x_{n+1}^2+c_1^2\left(u_{1,n}-\xi_n\right)^2\right]\middle|\,\Pi\right\} \\
&\quad \geq \sigma_\phi^2 \sum_{n=D_1+1}^{M} \varphi_n
\end{aligned} \tag{32}$$

If we divide both sides of (32) by M and take the limit as M goes to ∞, we obtain the inequality

$$\begin{aligned}
&\lim_{M\to\infty}\frac{1}{M}E\left\{\sum_{n=D_2+1}^{M}\left[x_{n+1}^2+c_1^2\left(u_{1,n}-a_1\xi_n\right)^2+c_2^2\left(u_{2,n}-a_2\xi_n\right)^2\right]\right. \\
&\left.\quad + \sum_{n=D_1+1}^{D_2}\left[x_{n+1}^2+c_1^2\left(u_{1,n}-\xi_n\right)^2\right]\middle|\,\Pi\right\} \\
&\quad \geq \lim_{M\to\infty}\frac{1}{M}\sigma_\phi^2\sum_{n=D_1+1}^{M}\varphi_n = \sigma_\phi^2\varphi_\infty
\end{aligned}$$

with equality if $\Pi = \Pi^*$. Thus, the infinite horizon cost J is lower bounded by the real number $\sigma_\phi^2\varphi_\infty$ and this bound can be achieved if the policy is that of the limit of the optimal finite-horizon controller. Hence, we established the result that the infinite horizon optimal controller can be obtained as the limit of the finite-horizon one.

4 Derivation of Optimal Decentralized Flow Controllers for M Users

The results of the previous section can be extended to the most general case of M users to a certain extent. In derivations to follow, for ease of notation we will assume that one of the users has no delay in acquiring the queue length and effective service rate information from the switch. In other words, we let $D_1 = 0$ and delays of the other users can be taken to be ordered as in (2). As in the two user case, we use a dynamic programming approach to calculate the controls, but this time the expressions become too cumbersome to write down analytically. However, the main idea of our calculations remains essentially the same: completion of squares. Starting at stage N again, down to stage $N - D_2$, we minimize J_n's over the control of the user that has the least information delay, namely the user 1. The terms that remain after this minimization have the form

$$R_d = \sum_{n=0}^{N-d} \gamma_n \left[\sum_{l=2}^{M} (u_{l,N-n-D_2+1} - a_2 \xi_{N-n-D_2+1}) + \rho_{n,1} x_{N-D_2+1} \right.$$
$$\left. + \sum_{k=1,n\geq 1}^{n} \sum_{l=2}^{M} \rho_{n,k} \left(u_{l,N-D_2+k} - a_l \xi_{N-D_2+k} \right) \right]^2$$
$$, \; d = N, \ldots, N - D_2 + 1$$

where γ_n's and $\rho_{n,k}$'s can be precomputed, and they depend on D_2 and the weights c_i's.

Between the stages $N - D_2$ and $N - D_3 + 1$, we minimize the objective functionals over two controls simultaneously. This brings in additional remaining terms on top of the slightly modified R_{N-D_2+1}. The new terms have the following structure:

$$A_d = \sum_{n=D_2}^{N-d} \gamma_n \left[\sum_{l=3}^{M} (u_{l,N-n-D_3+1} - a_2 \xi_{N-n-D_3+1}) + \rho_{n,1} x_{N-D_3+1} \right.$$
$$\left. + \sum_{k=1,n\geq 1}^{n} \sum_{l=3}^{M} \rho_{n,k} \left(u_{l,N-D_3+k} - a_l \xi_{N-D_3+k} \right) \right]^2$$
$$, \; d = N - D_2, \ldots, N - D_3 + 1 \qquad (33)$$

and the modified R_{N-D_2+1} equals

$$\tilde{R}_{N-D_2+1} = \sum_{n=0}^{D_2-1} \gamma_n \left[\sum_{l=2}^{M} (u_{l,N-n-D_3+1} - a_2 \xi_{N-n-D_3+1}) \right.$$

$$+\rho_{n,1}x_{N-D_3+1}$$

$$\left. + \sum_{k=1,n\geq 1}^{n} \sum_{l=2}^{M} \rho_{n,k} \left(u_{l,N-D_3+k} - a_l \xi_{N-D_3+k} \right) \right]^2 \quad (34)$$

Now, (33) and (34) can be combined to find all of the terms that are transferred to the next stage at time $N - D_3 + 1$. The result is

$$R_{N-D_3+1} = A_{N-D_3+1} + \tilde{R}_{N-D_2+1}$$

From time $N - D_3$ until $N - D_4 + 1$, additional terms accumulate. These terms have the same structure as in (33) and are given by

$$A_d = \sum_{n=D_3}^{N-d} \gamma_n \left[\sum_{l=4}^{M} \left(u_{l,N-n-D_4+1} - a_2 \xi_{N-n-D_4+1} \right) + \rho_{n,1} x_{N-D_4+1} \right.$$

$$\left. + \sum_{k=1,n\geq 1}^{n} \sum_{l=4}^{M} \rho_{n,k} \left(u_{l,N-D_4+k} - a_l \xi_{N-D_4+k} \right) \right]^2$$

$$,\ d = N - D_3, \ldots, N - D_4 + 1$$

Fortunately, the old terms that have already accumulated preserve their structure, but their time indices re-shift. Thus, $\tilde{R}_{N-D_2+1}$ is re-modified to

$$\hat{R}_{N-D_2+1} = \sum_{n=0}^{D_2-1} \gamma_n \left[\sum_{l=2}^{M} \left(u_{l,N-n-D_4+1} - a_2 \xi_{N-n-D_4+1} \right) \right.$$

$$+\rho_{n,1}x_{N-D_4+1}$$

$$\left. + \sum_{k=1,n\geq 1}^{n} \sum_{l=2}^{M} \rho_{n,k} \left(u_{l,N-D_4+k} - a_l \xi_{N-D_4+k} \right) \right]^2$$

and R_{N-D_3+1} becomes

$$\tilde{R}_{N-D_3+1} = \sum_{n=D_2}^{D_3-1} \gamma_n \left[\sum_{l=3}^{M} \left(u_{l,N-n-D_4+1} - a_2 \xi_{N-n-D_4+1} \right) \right.$$

$$+\rho_{n,1}x_{N-D_4+1}$$

$$\left. + \sum_{k=1,n\geq 1}^{n} \sum_{l=3}^{M} \rho_{n,k} \left(u_{l,N-D_4+k} - a_l \xi_{N-D_4+k} \right) \right]^2$$

We again combine all the terms to get

$$R_{N-D_4+1} = A_{N-D_4+1} + \hat{R}_{N-D_2+1} + \tilde{R}_{N-D_3+1}$$

This process can be repeated several times till the stage $N - D_M$ is reached. Eventually, we find R_{N-D_M+1}. From that point on, this so-called transient minimization process ends, and we start minimizing J_n's over the controls of all of the M users. This task can be reduced to finding a recursive relation such as (13) in the two user case, but this time the algebra is too complicated to carry out the manipulations analytically by hand. However, a computer can do these calculations easily for us once we know that the structure of the remaining terms are as in the above expressions. This is actually what is done in the next section. Note that, the finite-horizon controller for the M user case is a certainty equivalent controller as in the two user case. This fact simply follows from the construction of the controller. Let us make a final remark on the infinite-horizon version of the M user problem. The only practical difficulty of the M user case is the complication of algebra, which prevents us from finding a recursive formula for the controller dynamics. We know that such a relation exists, but we do not know yet whether it is convergent or not. If we had to assume that it converges, the extension to the infinite-horizon case would be exactly the same as in the two user case. To put it another way, the infinite horizon optimal controller would be the limit of the finite horizon one.

5 Simulation Results

We performed an extensive simulation study to investigate the performance of the optimal controller. In all simulations, we considered a network of $M = 3$ users, with the following values of system parameters

$$a_1 = a_2 = a_3 = 1/3,\ c_1 = c_2 = c_3 = c,\ p = 2,\ \alpha_1 = \alpha_2 = 0.4,\ \sigma_\phi^2 = 1$$

where ϕ_n's are assumed to be *i.i.d* Gaussian. And the run length of all simulations is $N = 1000$.

First, for comparison purposes, in Table 1, Table 2 and Table 3 we present the performance of the optimal controller along with two previously proposed certainty equivalent controllers referred to as Controller 1 and Controller 2 in [2]. As expected, optimal controller does a better job in regulating the queue length with comparable control effort and the resulting average cost is smaller in all cases. In [1] and [2], the simulation results indicated that the Controller 2 performs better than Controller 1, which is the case here, too. Next, we consider three scenarios to investigate the performance of the optimal controller only. In scenario 1, we decrease the delay of user 2 holding that of users 1 and 3 at their previous values. As depicted in Table 4, in this case the optimal overall cost decreases, because user 2 has access to more recent information about the queue length and available service rate. Next, we increase the delay of user 2 and study its effect on the optimal cost. Table 5 summarizes our simulation results in this case. Clearly, this time the optimal cost increases, because user 2 has worse knowledge on the system variables. Finally, we decrease the delay of user 3, while holding the delays of users 1 and 2 fixed at their base values. The performance of optimal controller in this case is depicted in Table 6. As the results indicate, the optimal average cost is smaller as compared to the case in which $D_3 = 10$. The decrease in the delay of user 3 results in a better performance of the system, because user 3 can make better estimates of the queue length and available service rate which are used in calculating his control actions and the overall cost.

Another issue we investigated in simulations is the rate of convergence of system dynamics. In all cases, the finite-horizon optimal control policies converged to stationary policies at very fast rates. This result suggests the existence of a stationary infinite horizon optimal controller for the general M user case, a result that we did not prove here.

c	$\sqrt{E(x^2)}$	$\sqrt{E(u_1^2)}$	$\sqrt{E(u_2^2)}$	$\sqrt{E(u_3^2)}$	J
0.1	0.0045	1.2830	0.2667	0.0898	0.0116
1	0.2178	1.3057	0.3271	0.1081	1.1545
10	3.1474	0.7896	0.3802	0.1773	66.7900

Table 1: Optimal Controller with $D_1 = 0, D_2 = 5, D_3 = 10$

c	$\sqrt{E(x^2)}$	$\sqrt{E(u_1^2)}$	$\sqrt{E(u_2^2)}$	$\sqrt{E(u_3^2)}$	J
0.1	2.69	0.89	0.63	0.49	7.68
1	3.18	0.84	0.60	0.49	11.44
10	7.19	0.38	0.48	0.50	114.66

Table 2: Controller 1 with $D_1 = 0, D_2 = 5, D_3 = 10$

c	$\sqrt{E(x^2)}$	$\sqrt{E(u_1^2)}$	$\sqrt{E(u_2^2)}$	$\sqrt{E(u_3^2)}$	J
0.1	0.01	1.48	1.18	0.49	0.04
1	0.80	1.01	0.82	0.50	2.46
10	6.00	0.38	0.47	0.50	85.37

Table 3: Controller 2 with $D_1 = 0, D_2 = 5, D_3 = 10$

c	$\sqrt{E(x^2)}$	$\sqrt{E(u_1^2)}$	$\sqrt{E(u_2^2)}$	$\sqrt{E(u_3^2)}$	J
0.1	0.0042	1.1634	0.4776	0.0990	0.0085
1	0.2603	1.1012	0.5411	0.1049	0.8587
10	2.1791	0.6745	0.4920	0.1503	47.1239

Table 4: Optimal Controller with $D_1 = 0, D_2 = 2, D_3 = 10$

c	$\sqrt{E(x^2)}$	$\sqrt{E(u_1^2)}$	$\sqrt{E(u_2^2)}$	$\sqrt{E(u_3^2)}$	J
0.1	0.0045	1.4276	0.1649	0.0978	0.0137
1	0.2013	1.4193	0.1919	0.1082	1.4320
10	3.9501	0.9362	0.3049	0.2215	90.6461

Table 5: Optimal Controller with $D_1 = 0, D_2 = 8, D_3 = 10$

c	$\sqrt{E(x^2)}$	$\sqrt{E(u_1^2)}$	$\sqrt{E(u_2^2)}$	$\sqrt{E(u_3^2)}$	J
0.1	0.0048	1.4776	0.2902	0.1779	0.0140
1	0.2225	1.2763	0.2816	0.1571	1.1583
10	3.3677	0.8089	0.3747	0.2743	66.1541

Table 6: Optimal Controller with $D_1 = 0, D_2 = 5, D_3 = 7$

6 Conclusions

In this paper, we have presented a method to calculate the optimal solution to a class of team problems with a partially nested information pattern, which particularly arises in the context of designing flow controllers in communication networks. As a basic approach towards obtaining the solution we adopted dynamic programming with which we first obtained the solution in a finite time horizon setting. Next, we exploited the special structure of the two user problem to extend this result to the infinite horizon case. We proved the existence of the infinite horizon solution and showed that it can be calculated as the limit of the finite horizon one. Further extension of this result to the M user case is still an open problem that we are currently working on. Yet another future direction of research would be to look into the continuous-time version of the same problem.

References

[1] E. Altman, T. Başar, and R. Srikant. "A Team-Theoretic Approach to Congestion Control", submitted to *Automatica*, June 1998.

[2] E. Altman, T. Başar, and R. Srikant. "Robust rate control for ABR sources", In *Proceedings of IEEE INFOCOM*, San Francisco, CA, March 1998.

[3] E. Altman, T. Başar, and R. Srikant. "Multi-user rate-based flow control with action delays: A team-theoretic approach.", In *Proceedings of the 36th IEEE Conference on Decision and Control*, pp. 2387-2392, San Diego, CA, December 1997.

[4] E. Altman and T. Başar. "Optimal rate control for high speed telecommunication networks", In *Proceedings of the 34th IEEE Conference on Decision and Control*, pp. 1389-1394, New Orleans, LA, December 1995.

[5] T. Başar and J. B. Cruz, Jr. "Concepts and methods in multi-person coordination and control", In S. G. Tzafestas, editor,*Optimization and Control of Dynamic Operational Research Models*, chapter 11, pp. 351-387, North-Holland Publ. Co., 1982.

[6] J-C. Bolot. "End-to-end packet delay and loss behavior in the Internet", *Proceedings of ACM Sigcomm '93*, pp. 289-298, San Francisco, CA, September 1993.

[7] Y. C. Ho and K. C. Chu. "Team decision theory and information structures in optimal control problems: Part I", *IEEE Transactions on Automatic Control*, 17(1):15-21, February 1972.

[8] D. Kincaid and W. Cheney. "Numerical Analysis", *Brooks/Cole Publ. Co.*, pp. 80-82, 1991. "Team decision theory and information structures in optimal control problems-Part I", *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 15-22, Feb. 1972.

[9] R. Radner. "Team decision problems", *Ann. Math. Statist.*, vol. 33, no. 3, pp. 857-881, 1962.

[10] N. Sandell and M. Athans. "Solution of some non-classical LQG stochastic decision problems", *IEEE Transactions on Automatic Control*, AC-19:108-116, 1974.