2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Optimal decentralized control of coupled subsystems with control sharing

Aditya Mahajan

Abstract— **The optimal decentralized control of coupled subsystems with control sharing is investigated. The system consists of n-coupled subsystems, each with a local control station. The evolution of a subsystem is controlled by the actions of all control stations. However, each control station observes only the state of its subsystem and the one-step delayed actions of all control stations. At each time, a cost that depends on the state of all subsystems and the actions of all control stations is incurred. The system has non-classical information structure; since each control station observes the delayed control actions of all other control stations, the system is said to have *control-sharing* information structure. We use the approach of Mahajan et al. (2008), to obtain the structure of optimal control stations and a dynamic programming decomposition, which is similar to the dynamic program for centralized partially observed systems. The structure of optimal control stations is simpler than the general structure proposed in Mahajan et al. (2008), and, consequently, so is the dynamic programming decomposition.**

I. Introduction

A. Motivation

In this paper, we investigate one of the simplest architectures for networked control systems—a collection of dynamically coupled subsystems, each with a local control station. A local control station directly observes the state of its subsystem, but does not observe the state of other subsystems. However, the control actions of any control station are observed by all control stations with one-step delay. Such a *control sharing* happens naturally in applications like queueing networks and multi-terminal communication, or when control actions are communicated over a broadcast medium like the Internet.

The above model provides a modular architecture for networked control systems. In this paper, we investigate the optimal design of such a decentralized control system. The system has a non-classical information structure. In general, the optimal design of decentralized control systems with non-classical information structure is notoriously difficult. Nonetheless, we show that the salient features of the model—each local control observes the state of its subsystem; the dynamics of a subsystem does not depend on the state of other subsystems; and all control actions are shared between the control stations—simplify the design of such a system.

B. Notation

We denote random variables with upper case letters, their realization with lower case letters, and their space of realizations by script letters. For example, for a random variable X, x denotes its realization and $\mathcal{X}$ denotes its space of realizations. Subscripts denote time and superscripts denote the subsystem. For example, X^i_t denotes the state of subsystem i at time t. The short hand notation $X^i_{1:t}$ denotes the vector $(X^i_1, X^i_2, \dots, X^i_t)$. Bold face letters denotes the collection of variables at all subsystems. For example, $\mathbf{X}_t$ denotes $(X^1_t, X^2_t, \dots X^n_t)$. The notation $\mathbf{X}^{-i}_t$ denotes the vector $(X^1_t, \dots, X^{i-1}_t, X^{i+1}_t, \dots, X^n_t)$.

$\Delta(\mathcal{X})$ denotes the probability simplex on the space $\mathcal{X}$. $\mathbb{P}(A)$ denotes the probability of an event A, and $\mathbb{E}[X]$ denotes the expectation of a random variable X. Let $\mathbb{N}$ denote the set of natural numbers.

C. Model and Problem Formulation

Consider a discrete-time networked control system with n subsystems. Let $Z_t \in \mathcal{Z}$ denote the *global state* of the system and $X^i_t \in \mathcal{X}^i$, $i = 1, \dots, n$, denote the *local state* of subsystem i at time t. The initial global state Z_1 has a distribution P_Z. Conditioned on the initial global state Z_1, the initial local state of all subsystems are independent; initial local state X^i_1 is distributed according to $P_{X^i|Z}$, $i = 1, \dots, n$. Let $\mathbf{X}_t := (X^1_t, \dots, X^n_t)$ denote the local state of all subsystems.

A control station is co-located with each subsystem. Let $U^i_t \in \mathcal{U}^i$ denote the control action of control station i and $\mathbf{U}_t := (U^1_t, U^2_t, \dots, U^n_t)$ denote the collection of all control actions.

At time t, control station i, $i = 1, \dots, n$, perfectly observes the global state Z_t, the local state X^i_t of subsystem i, and the one-step delayed control actions $\mathbf{U}_{t-1}$ of all control stations—thus, the system has a *control sharing information structure*.

Control station i, $i = 1, \dots, n$, chooses a control action $U^i_t \in \mathcal{U}^i_t$ based on all the data available to it. Thus,

$$U^i_t = g^i_t(Z_{1:t}, X^i_{1:t}, \mathbf{U}_{1:t-1}) \tag{1}$$

where $Z_{1:t} := (Z_1, \dots, Z_t)$, $X^i_{1:t} := (X^i_1, X^i_2, \dots, X^i_t)$ and $\mathbf{U}_{1:t-1} := (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_{t-1})$. The function g^i_t is called the *control law* of control station i.

The global state and the local state of each subsystems are coupled through the control actions; the global state evolves according to

$$Z_{t+1} = f^0_t(Z_t, \mathbf{U}_t, W^0_t) \tag{2}$$

while the local state of subsystem i, $i = 1, \dots, n$, evolves according to:

$$X^i_{t+1} = f^i_t(Z_t, X^i_t, \mathbf{U}_t, W^i_t) \tag{3}$$

Aditya Mahajan is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 2A7, Canada. aditya.mahajan@mcgill.ca

978-1-61284-799-3/11/$26.00 ©2011 IEEE

$W_t^i \in \mathcal{W}^i$, $i = 0, 1, \dots, n$, is the plant disturbance with distribution P_{W^i}. The processes $\{W_t^i, t = 1, \dots\}$, $i = 0, 1, \dots, n$, are assumed to be independent of each other and also independent of the initial state $(Z_1, \mathbf{X}_1)$ of the system.

Note that the updated local state of subsystem i depend only the previous local state of subsystem i and previous global state but is controlled by all control stations.

The subsystems are also coupled through cost. At time t, the system incurs a cost $c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t)$ that depends on the global state, the local state of all subsystems, and the actions of all control stations. The system runs for a time horizon T. The collection $\mathbf{g}^i := (g_1^i, g_2^i, \dots, g_T^i)$ of control laws at control station i is called the *control strategy of control station i*. The collection $\mathbf{g} := (\mathbf{g}^1, \mathbf{g}^2, \dots, \mathbf{g}^n)$ of control strategies of all control stations is called the *control strategy of the system.* The performance of a control strategy $\mathbf{g}$ is measured by the expected total cost incurred by that strategy, which is given by

$$J(\mathbf{g}) := \mathbb{E}\Big[\sum_{t=1}^{T} c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t)\Big] \tag{4}$$

where the expectation is with respect to a joint measure of $(Z_{1:T}, \mathbf{X}_{1:T}, \mathbf{U}_{1:T})$ induced by the choice of the control strategy $\mathbf{g}$.

We are interested in the following optimal control problem:

Problem 1: Given the distributions P_Z, $P_{X^i|Z}$ and P_{W^i} of the initial global state, initial local state, and plant disturbance of subsystem i, $i = 1, \dots, n$, a horizon T, and the cost functions c_t, $t = 1, \dots, T$, find a control strategy $\mathbf{g}$ that minimizes the expected total cost given by (4).

D. Literature overview

The model described above has a non-classical information structure [1], [2] because no control station knows the information available to all other control stations. There are a few general methods to obtain a dynamic programming decomposition of systems with non-classical information structure: for finite horizon systems, a framework was presented by Witsenhausen [3]; for two-agent finite and infinite horizon systems, a framework was presented by Mahajan [4]. We are interested in a solution framework that works for multiple control stations and extends to infinite horizon systems.

Given the difficulty of a general framework for dynamic programming for systems with non-classical information structures, researchers have focused attention on specific non-classical information structures. One common theme has been sharing of information between the control stations. Examples include:

1) Systems in which the state of the plant is observed by all control stations after a delay. Such systems are said to have a *delayed state observation* information structure and were investigated in [5], who obtained the structure of optimal control strategies and a dynamic programming decomposition for such systems.
2) Systems in which the (possibly noisy) observations and control actions of a control station are observed by all control stations with a delay. Such systems are called *delayed (observation) sharing* information structure. For such systems, the structure of optimal control strategies and a dynamic programming decomposition were obtained in [6] (for one-step delay) and in [7] (for general delay).
3) Systems in which the control action of a control station is observed by all control stations with a delay. Such systems are said to have a *control sharing* information structure. One-step delayed control sharing with continuous valued control actions was considered in [8], [9], who exploited the continuous nature of the control actions by embedding the observations densely in the controls. This information embedding transforms the systems to a one-step delayed observation sharing information structures and incurs an arbitrarily small loss in performance. However, the resulting control laws are not continuous.
4) System in which the state of the plant is observed periodically by all control stations. Such systems are said to have a *periodic sharing* information structure and were investigated in [10], who obtained the structure of optimal control strategies and a dynamic programming decomposition.
5) Systems in which the belief of each control station on the state of the plant is shared between all control stations after a delay. Such systems are said to have a *belief sharing* information structure. Systems in which the sharing delay is one were considered in [11], who obtained the structure of optimal control strategies and a dynamic programming decomposition for the system.
6) Systems in which the observations of the control stations is split between common observations and private observations in such a way that the size of the private observations does not increase. Such systems were investigated in [12], who obtained the structure of the optimal control stations and a dynamic programming decomposition for the system.

The model considered in this paper has a one-step delayed control sharing information structure. We want a solution approach that will also work when the control actions are finite valued (as is the case in network controlled systems). So, the technique proposed by Bismut [8] to embed the observations in the control actions does not necessarily work.

E. Main result

In the model of Section I-C, the data available at control station i increases with time. Consequently, the domain of the control laws of the form (1) increases with time, which makes it difficult to implement the control laws. In this paper, we show that without loss of optimality we can restrict attention to control laws whose domain does not increase with time.

For simplicity of exposition, in the sequel we will assume that the alphabets $\mathcal{Z}$, $\mathcal{X}^i$, $\mathcal{U}^i$, and $\mathcal{W}^i$, $i = 1, \dots, n$, are finite. The results extend to general alphabets under suitable technical conditions.

Definition 1: Let Π^i_t, $i = 1, \dots, n$, $t = 1, \dots, T$, denote the posterior probability of the local state of subsystem i given the past history of global state and control actions of all the control stations, i.e., for any $x \in \mathcal{X}^i$, the component x of Π^i_t is given by

$$\Pi^i_t(x) := \mathbb{P}(X^i_t = x | Z_{1:t}, \mathbf{U}_{1:t-1}; \mathbf{g})$$

Π^i_t is a random variable taking values in $\Delta(\mathcal{X}^i)$. π^i_t denotes the realization of Π^i_t and $\mathbf{\Pi}_t$ denotes $(\Pi^1_t, \Pi^2_t, \dots, \Pi^n_t)$.

Theorem 1 (Structure of control laws): In Problem 1, restricting attention to control stations of the form

$$U^i_t = g^i_t(X^i_t, Z_t, \mathbf{\Pi}_t) \tag{5}$$

is without any loss of optimality.

In centralized stochastic control problems, the dynamic program consists of a sequence of nested optimality equations—one for each time step. The optimality equation at time t finds the best control action for the current (information) state.

In contrast, the dynamic programming decomposition in decentralized stochastic control is coarser. The dynamic program still consists of a sequence of nested optimality equations. However, as different control stations have different information, the optimality equations cannot find the best control action for all control stations. To over this limitations, we exploit the structure of optimal control laws derived in Theorem 1. We split the control law at control station i into two parts: first a *coordinator* chooses function sections

$$D^i_t(\cdot) = g^i_t(\cdot, Z_t, \mathbf{\Pi}_t)$$

based on *common data* $(Z_t, \mathbf{\Pi}_t)$ for all control stations $i = 1, \dots, n$. Then, each control station uses the prescription D^i_t and its *local data* X^i_t to generate a control action

$$X^i_t = D^i_t(X^i_t).$$

In the dynamic programming decomposition, the optimality equation at time t finds the best function sections $\mathbf{D}_t = (D^1_t, \dots, D^n_t)$. Such a dynamic programming decomposition is given below.

Theorem 2 (Dynamic programming decomposition): For a particular realization z_t of Z_t and $\boldsymbol{\pi}_t$ of $\mathbf{\Pi}_t$, an optimal choice $\mathbf{d}_t = (d^1_t, \dots, d^n_t)$ of function sections $\mathbf{D}_t = (D^1_t, \dots, D^n_t)$ is given by the solution of the following nested optimality equations

$$V_T(z_T, \boldsymbol{\pi}_T) = \min_{\mathbf{d}_T} \mathbb{E}\Big[c_T(\mathbf{X}_T, \mathbf{U}_T) \Big| Z_T = z_T, \mathbf{\Pi}_T = \boldsymbol{\pi}_T, \mathbf{D}_T = \mathbf{d}_T\Big] \tag{6}$$

and for $t = T-1, T-2, \dots, 1$,

$$V_t(z_t, \boldsymbol{\pi}_t) = \min_{\mathbf{d}_t} \mathbb{E}\Big[c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(F_t(\boldsymbol{\pi}_t, Z_{t+1}, \mathbf{U}_t, \mathbf{d}_t) \Big| Z_t = z_t, \mathbf{\Pi}_t = \boldsymbol{\pi}_t, \mathbf{D}_t = \mathbf{d}_t\Big] \tag{7}$$

where F_t is a function that will be defined later in Lemma 7. The arg min at each step in (6) and (7) gives an optimal choice for the function section D_t. Denote the arg min at $(z_t, \boldsymbol{\pi}_t)$ by $\mathbf{d}^*_t(z_t, \boldsymbol{\pi}_t)$. Then, the optimal control law $g^{*,i}_t$ at time t is given by

$$g^{*,i}_t(x^i_t, z_t, \boldsymbol{\pi}_t) = d^{*,i}_t(z_t, \boldsymbol{\pi}_t)(x^i_t). \tag{8}$$

The rest of this paper is organized as follows. We prove Theorems 1 and 2 in Sections II and III. We argue how to extend the results to infinite horizon in Section IV and conclude in Section V.

II. Proof of structural result

The proof of Theorem 1 proceeds in two stages. First, we show that the past values of the local state $X^i_{1:t-1}$ are irrelevant at control station i at time t. Thus, shedding this irrelevant information at each control station does not entail any loss of optimality. Second, we show that the common data $(Z_{1:t}, \mathbf{U}_{1:t-1})$ observed by all control stations may be replaced by an appropriate sufficient statistic. This replacement results in the structural result of Theorem 1.

A. Shedding of irrelevant information

The result of this section depends on the following result.

Lemma 3: Consider the model of Section I-C for an arbitrary but fixed choice of control strategy $\mathbf{g}$. Then, conditioned on the history of global state and control actions, the local states of all subsystems are independent. Specifically, for any realization $z_t \in \mathcal{Z}$, $x^i_t \in \mathcal{X}^i$ and $u^i_t \in \mathcal{U}^i$ of X^i_t and U^i_t, $i = 1, \dots, n$, $t = 1, \dots, T$, we have

$$\mathbb{P}(\mathbf{X}_{1:t} = \mathbf{x}_{1:t} \,|\, Z_{1:t} = z_{1:t}, \mathbf{U}_{1:t} = \mathbf{u}_{1:t}) = \prod_{i=1}^{n} \mathbb{P}(X^i_{1:t} = x^i_{1:t} \,|\, Z_{1:t} = z_{1:t}, \mathbf{U}_{1:t} = \mathbf{u}_{1:t}) \tag{9}$$

This is proved in Appendix I.

An immediate consequence of the above result is the following:

Lemma 4: Consider the model of Section I-C for an arbitrary but fixed choice of control strategy $\mathbf{g}$. Define $R^i_t = (X^i_t, Z_{1:t}, \mathbf{U}_{1:t-1})$. Then,

1) The process $\{R^i_t, t = 1, \dots, T\}$ is a controlled Markov process with control action U^i_t, i.e., for any $x^i_t, \tilde{x}^i_t \in \mathcal{X}^i$, $z_t, \tilde{z}_t \in \mathcal{Z}$, $u^i_t, \tilde{u}^i_t \in \mathcal{U}^i$, $r^i_t = (x^i_t, z_{1:t}, \mathbf{u}_{1:t-1})$, $\tilde{r}^i_t = (\tilde{x}^i_t, \tilde{z}_{1:t}, \tilde{\mathbf{u}}_{1:t-1})$, $i = 1, \dots, n$, and $t = 1, \dots, T$,

$$\mathbb{P}(R^i_{t+1} = \tilde{r}^i_{t+1} \,|\, R^i_{1:t} = r^i_{1:t}, U^i_{1:t} = u^i_{1:t}) = \mathbb{P}(R^i_{t+1} = \tilde{r}^i_{t+1} \,|\, R^i_t = r^i_t, U^i_t = u^i_t)$$

2) The instantaneous conditional cost simplifies as follows:

$$\mathbb{E}[c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t) \,|\, R^i_{1:t} = r^i_{1:t}, U^i_{1:t} = u^i_{1:t}] = \mathbb{E}[c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t) \,|\, R^i_t = r^i_t, U^i_t = u^i_t]$$

The proof is omitted due to lack of space.

In light of Lemma 4, pick any control station i, $i = 1, \dots, n$, arbitrarily fix the choice of control strategy $\mathbf{g}^i$ for all other control stations, and consider the subproblem of finding an optimal strategy for control station i in Problem 1. In this subproblem, control station i has access to $R^i_{1:t}$, chooses U^i_t, and incurs an expected instantaneous cost $\mathbb{E}[c_t(\mathbf{X}_t, \mathbf{U}_t) \mid R^i_{1:t}, U^i_{1:t}]$. Lemma 4 implies that the optimal choice of control strategy $\mathbf{g}^i$ is a Markov decision process. Thus, using Markov decision theory [13], we get the following (recall that $R^i_t = (X^i_t, Z_{1:t}, \mathbf{U}_{1:t-1})$):

Proposition 5: In Problem 1, restricting attention to control stations of the form

$$U^i_t = g^i_t(X^i_t, Z_{1:t}, \mathbf{U}_{1:t-1}) \tag{10}$$

is without loss of optimality.

B. Sufficient statistic for common data

Now consider Problem 1 with control strategies of the form (10). Split the data at each control station into two parts: the common data $(z_{1:t}, \mathbf{u}_{1:t-1})$ that is observed by all control stations and the local (or private) data x^i_t that is observed by only control station i. Note that the size of the local data does not increase with time. Mahajan *et al.* [12] showed that this particular subclass of non-classical information structures is tractable. Thus, Proposition 5 transforms Problem 1 to a form for which a solution technique is known.

The solution proposed in [12] proceeds in the following steps:

1) Formulate a stochastic control problem from the point of view of a *coordinator* that observes the common data $(Z_{1:t}, \mathbf{U}_{1:t-1})$. We call this system the *coordinated* system.
2) Show that the coordinated system is equivalent to the original model. That is, any strategy in the coordinated system is implementable in the original model and vice versa.
3) Show that by suitable expansion of the state-space, the coordinator's problem is a MDP (Markov decision process). Then, use results from Markov decision theory to find the structure of optimal control strategy and a dynamic programming decomposition for the coordinated system.

For completeness, we briefly describe these steps below. See [12] for complete details.

Step 1: The coordinated system

Consider a *coordinated system* that consists of a *coordinator* and the n control stations. The coordinator observes the common data $(Z_{1:t}, \mathbf{U}_{1:t-1})$ and chooses function sections $D^i_t : \mathcal{X}^i \mapsto \mathcal{U}^i$, $i = 1, \dots, n$ according to

$$\mathbf{D}_t = h_t(Z_{1:t}, \mathbf{U}_{1:t-1}) \tag{11}$$

where $\mathbf{D}_t := (D^1_t, \dots, D^n_t)$. The function $h_t(\cdot)$ is called the *coordination law*.

All control stations i, $i = 1, \dots, n$, are passive. They use the prescription D^i_t of the coordinator and act as follows:

$$U^i_t = D^i_t(X^i_t) \tag{12}$$

The system dynamics and the cost remain unchanged. The system dynamics are given by (2)–(3) and the instantaneous cost at time t is $c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t)$.

The collection $\mathbf{h} = (h_1, \dots, h_T)$ is called a *coordination strategy*. The performance of a coordination strategy is measured by the expected total cost incurred by that strategy, which is given by

$$\hat{J}(\mathbf{h}) = \mathbb{E}\Big[\sum_{t=1}^{T} c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t)\Big] \tag{13}$$

where the expectation is with respect to a joint measure of $(Z_{1:T}, \mathbf{X}_{1:T}, \mathbf{U}_{1:T})$ induced by the choice of the coordination strategy $\mathbf{h}$.

In the coordinated system, we are interested in the following optimal control problem.

Problem 2: Given the distributions P_Z, $P_{X^i|Z}$ and P_{W^i} of the initial global state, initial local state, and plant disturbance of subsystem i, $i = 1, \dots, n$, a horizon T, and the cost functions c_t, $t = 1, \dots, T$, find a coordination strategy $\mathbf{h}$ that minimizes the total cost given by (13).

Step 2: Equivalence between the two models

Proposition 6: Problem 1 with control stations of the form (10) is equivalent to Problem 2. Specifically, for any control strategy $\mathbf{g}$ of the form (10) for Problem 1 there is a coordination strategy $\mathbf{h}$ for Problem 2 such that $\hat{J}(\mathbf{h}) = J(\mathbf{g})$. Conversely, for any coordination strategy $\mathbf{h}$ for Problem 2, there is a control strategy $\mathbf{g}$ for Problem 1 such that $J(\mathbf{g}) = \hat{J}(\mathbf{h})$.

Proof: Given a control strategy $\mathbf{g}$ of the form (10) for Problem 1, pick the coordination strategy $\mathbf{h}$ according to:

$$h^i_t(z_{1:t}, \mathbf{u}_{1:t-1})(\cdot) = g^i_t(\cdot, z_{1:t}, \mathbf{u}_{1:t-1}) \tag{14}$$

where h^i_t denotes the i-th component of h_t. Then, for any realization of the primitive random variables Z_1, $\mathbf{X}_1$, $W^i_{1:T}$, $i = 0, 1, \dots, n$, the system variables $(Z_{1:T}, \mathbf{X}_{1:T}, \mathbf{U}_{1:T})$ have the same realizations in Problem 1 and Problem 2. Hence, $\hat{J}(\mathbf{h}) = J(\mathbf{g})$.

Conversely, given a control strategy $\mathbf{h}$ of Problem 2, pick a control strategy $\mathbf{g}$ for Problem 1 according to

$$g^i_t(x^i_t, z_{1:t}, \mathbf{u}_{1:t-1}) = h^i_t(z_{1:t}, \mathbf{u}_{1:t-1})(x^i_t) \tag{15}$$

By a similar argument as before, we can show that $J(\mathbf{g}) = \hat{J}(\mathbf{h})$. ■

Step 3: The coordinated system as a MDP

In this section, we show that the optimization problem at the coordinator is a MDP (Markov decision process). First, recall the definition of Π_t given in Definition 1. The dependence of Π_t on the control strategy $\mathbf{g}$, or equivalently the dependence on the coordination strategy $\mathbf{h}$, is only though the function sections $\mathbf{D}_{1:t-1}$. Thus, for any $x \in \mathcal{X}^i$, the component x of Π^i_t is given by

$$\Pi^i_t(x) := \mathbb{P}(X^i_t = x | Z_{1:t}, \mathbf{U}_{1:t-1}; \mathbf{D}_{1:t-1}).$$

Let π^i_t denote the realization of Π^i_t and $\mathbf{\Pi}_t$ denote $(\Pi^1_t, \Pi^2_t, \dots, \Pi^n_t)$.

Lemma 7: There exists a deterministic function F_t such that

$$\mathbf{\Pi}_{t+1} = F_t(\mathbf{\Pi}_t, Z_{t+1}, \mathbf{U}_t, \mathbf{D}_t) \tag{16}$$

The proof follows from the law of total probability and Bayes rule.

Lemma 8: Consider the coordinated system for an arbitrary but fixed coordination strategy $\mathbf{h}$. Then

1) The process $\{(Z_t, \mathbf{\Pi}_t),\ t = 1, \dots, T\}$, is a controlled Markov process with control action D_t, *i.e.*, for any $z_t \in \mathcal{Z}$, $\pi^i_t \in \Delta(\mathcal{X})$, $B_{t+1} \subset \Delta(\mathcal{X}^1) \times \cdots \times \Delta(\mathcal{X}^n)$, and any choice d^i_t of D^i_t, for $i = 1, \dots, n$ and $t = 1, \dots, T$, we have that

$$\begin{aligned}\mathbb{P}(Z_{t+1} = z_{t+1}, \mathbf{\Pi}_{t+1} \in B_{t+1} \mid Z_{1:t} = z_{1:t},\\ \mathbf{\Pi}_{1:t} = \boldsymbol{\pi}_{1:t}, \mathbf{D}_{1:t} = \mathbf{d}_{1:t})\\ = \mathbb{P}(Z_{t+1} = z_{t+1}, \mathbf{\Pi}_{t+1} \in B_{t+1} \mid Z_t = z_t,\\ \mathbf{\Pi}_t = \boldsymbol{\pi}_t, \mathbf{D}_t = \mathbf{d}_t)\end{aligned} \tag{17}$$

2) The instantaneous conditional cost simplifies as follows:

$$\begin{aligned}\mathbb{E}[c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t) \mid Z_{1:t} = z_{1:t}, \mathbf{\Pi}_{1:t} = \boldsymbol{\pi}_{1:t},\\ \mathbf{D}_{1:t} = \mathbf{d}_{1:t}]\\ = \mathbb{E}[c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t) \mid Z_t = z_t, \mathbf{\Pi}_t = \boldsymbol{\pi}_t, \mathbf{D}_t = \mathbf{d}_t]\end{aligned} \tag{18}$$

Proof: Part 1) follows from the update equation (2) for the global state Z_t, the behavior (12) of the control stations in the coordinated system, and the update (16) of the information state $\mathbf{\Pi}_t$. Part 2) follows from the definition of $\mathbf{\Pi}_t$ and the behavior (12) of the control stations in the coordinated system. ■

Lemma 8 shows that the choice of optimal function sections D_t is a Markov decision process with state $(Z_t, \mathbf{\Pi}_t)$. Thus, using Markov decision theory [13], we get the following:

Proposition 9: In Problem 2, restricting attention to coordination strategies of the form

$$\mathbf{D}_t = h_t(Z_t, \mathbf{\Pi}_t) \tag{19}$$

is without loss of optimality. Due to the equivalence with Problem 1 (see Proposition 6), we get that in Problem 1, restricting attention to control strategies of the form

$$U^i_t = g^i_t(X^i_t, Z_t, \mathbf{\Pi}_t) \tag{20}$$

is without loss of optimality.

The second part of Proposition 9 proves Theorem 1.

III. Proof of Dynamic Programming Decomposition

Lemma 8 shows that the choice of optimal function sections D_t is a Markov decision process with state $(Z_t, \mathbf{\Pi}_t)$. Thus, using Markov decision theory [13], we get the following dynamic programming decomposition

Proposition 10: Define $V_t : \mathcal{Z} \times \Delta(\mathcal{X}^1) \times \cdots \times \Delta(\mathcal{X}^n) \mapsto \mathbb{R}$ as follows: for any $z \in \mathcal{Z}$ and $\pi^i \in \Delta(\mathcal{X}^i)$, define

$$\begin{aligned}V_T(z, \boldsymbol{\pi}) = \min_{\mathbf{d}} \mathbb{E}\Big[c_t(\mathbf{X}_T, \mathbf{U}_T) \,\Big|\, Z_T = z,\\ \mathbf{\Pi}_T = \boldsymbol{\pi}, \mathbf{D}_T = \mathbf{d}\Big]\end{aligned} \tag{21}$$

and for $t = T-1, T-2, \dots, 1$,

$$\begin{aligned}V_t(z, \boldsymbol{\pi}) = \min_{\mathbf{d}} \mathbb{E}\Big[c_t(Z_t, \mathbf{X}_t, \mathbf{U}_t)\\ + V_{t+1}(F_t(\boldsymbol{\pi}, Z_{t+1}, \mathbf{U}_t, \mathbf{d}) \,\Big|\, Z_t = z, \mathbf{\Pi}_t = \boldsymbol{\pi}, \mathbf{D}_t = \mathbf{d}\Big]\end{aligned} \tag{22}$$

where F_t is defined as in Lemma 7. The arg min at each stage in (21) and (22) gives the optimal coordination strategy $h_t(\boldsymbol{\pi})$.

Theorem 2 follows from the equivalence of Proposition 6 and Proposition 10.

IV. Extension to infinite horizon

The results of Theorems 1 and 2 can be easily extended to infinite horizon expected discounted cost setup: Assuming that the plant function f_t and the instantaneous cost c_t are time-invariant, choose a strategy $\mathbf{g} := (g_1, g_2, \dots,)$ to minimize

$$\sum_{t=1}^{\infty} \beta^{t-1} c(\mathbf{X}_t, \mathbf{U}_t)$$

where $\beta \in (0, 1)$.

The results of Theorems 1 and 2 rely on Proposition 6—the equivalence between the original and coordinated systems—which remains valid even for infinite horizon. The process $\{(Z_t, \mathbf{\Pi}_t), t = 1, 2, \dots\}$ remains a controlled Markov process. So, the results of Propositions 9 and 10 extend to infinite horizon setup in the standard manner. These extensions can then be translated back to the original system along the lines of the translations presented in this paper for finite horizon system. This process will yield the following dynamic programming decomposition for infinite horizon: The choice of the function section $\mathbf{d}_t$ as a function of $\boldsymbol{\pi}_t$ does not depend on time as is given by the solution to the following fixed point equation: for any $\boldsymbol{\pi} \in \Delta(\mathcal{X}^1 \times \cdots \times \mathcal{X}^n)$

$$\begin{aligned}V(\boldsymbol{\pi}) = \min_{\mathbf{d}} \mathbb{E}\Big[c(\mathbf{X}, \mathbf{U})\\ + \beta V(F(\boldsymbol{\pi}, \mathbf{d}, \mathbf{U})) \,\Big|\, \mathbf{\Pi} = \boldsymbol{\pi}, \mathbf{D} = \mathbf{d}\Big]\end{aligned}$$

where $F(\cdot)$ is the time-homogeneous version of $F_t(\cdot)$.

V. Conclusion

We investigate the optimal decentralized control of coupled subsystems with control sharing. The evolution of each subsystem is controlled by the action of all control stations; each control station observes the state of its subsystem and the one-step delayed state of all control stations. The subsystems are further coupled by the cost.

First, we show that each control station can discard the past values of the state of its subsystem. Next, we consider a coordinated system in which a coordinator observes the one-step delayed actions of all control stations and prescribes a partially evaluated section of the control law to each control station. The control stations use this prescription to compute the corresponding control action. The coordinated system is a centralized system. We show that the original and the coordinated systems are equivalent. We analyze the coordinated system using standard tools from Markov decision theory and translate the results back to the original system by exploiting the equivalence between the two systems.

Appendix I
Proof of Lemma 3

For simplicity of notation, we use $\mathbb{P}(z_{1:t}, \mathbf{x}_{1:t}, \mathbf{u}_{1:t})$ to denote $\mathbb{P}(Z_{1:t} = z_{1:t}, \mathbf{X}_{1:t} = \mathbf{x}_{1:t}, \mathbf{U}_{1:t} = \mathbf{u}_{1:t})$ and a similar notation for conditional probability. Define

$$\begin{aligned} \alpha^i_t &:= \mathbb{P}(u^i_t \,|\, z_{1:t}, x^i_{1:t}, \mathbf{u}_{1:t-1}), \\ \beta^i_t &:= \mathbb{P}(x^i_t \,|\, z_{t-1}, x^i_{t-1}, \mathbf{u}_{t-1}), \\ \gamma^i_t &:= \mathbb{P}(z_t \,|\, z_{t-1}, \mathbf{u}_{t-1}) \end{aligned}$$

and

$$A^i_t := \prod_{s=1}^{t} \alpha^i_s, \quad B^i_t := \prod_{s=1}^{t} \beta^i_s, \quad \Gamma_t := \prod_{s=1}^{t} \gamma_s.$$

From law of total probability it follows that:

$$\mathbb{P}(z_{1:t}, \mathbf{x}_{1:t}, \mathbf{u}_{1:t}) = \left(\prod_{i=1}^{n} A^i_t B^i_t \right) \Gamma_t.$$

Summing over all realizations of $\mathbf{x}_{1:t}$ and observing that A^i_t and B^i_t depends only on $(z_{1:t}, x^i_{1:t}, \mathbf{u}_{1:t})$, we get

$$\begin{aligned} \mathbb{P}(z_{1:t}, \mathbf{u}_{1:t}) &= \sum_{x^1_{1:t}} \sum_{x^2_{1:t}} \cdots \sum_{x^n_{1:t}} \left(\prod_{i=1}^{n} A^i_t B^i_t \right) \Gamma_t \\ &= \left(\prod_{i=1}^{n} \left(\sum_{x^i_{1:t}} A^i_t B^i_t \right) \right) \Gamma_t. \end{aligned}$$

Thus, using Bayes rule we get

$$\mathbb{P}(\mathbf{x}_{1:t} \,|\, z_{1:t}, \mathbf{u}_{1:t}) = \prod_{i=1}^{n} \frac{A^i_t B^i_t}{\left(\sum_{x^i_{1:t}} A^i_t B^i_t \right)} \tag{23}$$

Summing both sides over $x^i_{1:t}$, $i \neq j$, we get

$$\mathbb{P}(x^j_{1:t} \,|\, z_{1:t}, \mathbf{u}_{1:t}) = \frac{A^j_t B^j_t}{\left(\sum_{x^j_{1:t}} A^j_t B^j_t \right)} \tag{24}$$

The result follows from combining (23) and (24).

References

[1] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proc. IEEE*, vol. 59, no. 11, pp. 1557–1566, Nov. 1971.

[2] Y.-C. Ho, "Team decision theory and information structures," *Proc. IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[3] H. S. Witsenhausen, "A standard form for sequential stochastic control," *Mathematical Systems Theory*, vol. 7, no. 1, pp. 5–11, 1973.

[4] A. Mahajan, "Sequential decomposition of sequential dynamic teams: applications to real-time communication and networked control systems," Ph.D. dissertation, University of Michigan, Ann Arbor, MI, 2008.

[5] M. Aicardi, F. Davoli, and R. Minciardi, "Decentralized optimal control of Markov chains with a common past information set," *IEEE Trans. Autom. Control*, vol. 32, no. 11, pp. 1028–1031, 1987.

[6] P. Varaiya and J. Walrand, "On delayed sharing patterns," *IEEE Trans. Autom. Control*, vol. 23, no. 3, pp. 443–445, 1978.

[7] A. Nayyar, A. Mahajan, and D. Teneketzis, "Optimal control strategies in delayed sharing information structures," *accepted for publication inIEEE Trans. Autom. Control*, 2011.

[8] J.-M. Bismut, "An example of interaction between information and control: The transparency of a game," *IEEE Trans. Autom. Control*, vol. 18, no. 5, pp. 518–522, Oct. 1972.

[9] N. Sandell and M. Athans, "Solution of some nonclassical lqg stochastic decision problems," *IEEE Trans. Autom. Control*, vol. 19, pp. 108–116, 1974.

[10] J. M. Ooi, S. M. Verbout, J. T. Ludwig, and G. W. Wornell, "A separation theorem for periodic sharing information patterns in decentralized control," *IEEE Trans. Autom. Control*, vol. 42, no. 11, pp. 1546–1550, Nov. 1997.

[11] S. Yüksel, "Stochastic nestedness and the belief sharing information pattern," *IEEE Trans. Autom. Control*, pp. 2773–2786, Dec. 2009.

[12] A. Mahajan, A. Nayyar, and D. Teneketzis, "Identifying tractable decentralized control problems on the basis of information structures," in *proceedings of the 46th Allerton conference on communication, control and computation*, Sep. 2008, pp. 1440–1449.

[13] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation Identification and Adaptive Control*. Prentice Hall, 1986.

[14] B. Hajek, K. Mitzel, and S. Yang, "Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm," *IEEE Trans. Inf. Theory*, vol. 64, pp. 608–622, Feb. 2008.

[15] B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Autom. Control*, vol. AC-27, no. 3, pp. 559–569, Jun. 1982.

[16] M. G. Hluchyj and R. G. Gallager, "Multiacces of a slotted channel by finitely many users," in *Proceedings of National Telecommunication Conference*, 1981, pp. D.4.2.1–D.4.2.7.

[17] J. M. Ooi and G. W. Wornell, "Decentralized control of a mutliple access broadcast channel: performance bounds," in *Proccedings of the 35th Conference on Decision and Control*, Kobe, Japan, 1996, pp. 293–298.

[18] F. C. Schoute, "Decentralized control in packet switched satellite communication," *IEEE Trans. Autom. Control*, vol. AC-23, no. 2, pp. 362–271, Apr. 1976.

[19] P. Varaiya and J. Walrand, "Decentralized control in packet switched satellite communication," *IEEE Trans. Autom. Control*, vol. AC-24, no. 5, pp. 794–796, Oct. 1979.

[20] A. Mahajan, "Optimal transmission policies for two-user multiple access broadcast using dynamic team theory," in *proceedings of the 48th Allerton conference on communication, control control and computation*, 2010.