Decentralized Stochastic Control with Partial History Sharing: A Common Information Approach

Ashutosh Nayyar, Aditya Mahajan and Demosthenis Teneketzis

Abstract

A general model of decentralized stochastic control called *partial history sharing* information structure is presented. In this model, at each step the controllers share part of their observation and control history with each other. This general model subsumes several existing models of information sharing as special cases. Based on the information commonly known to all the controllers, the decentralized problem is reformulated as an equivalent centralized problem from the perspective of a coordinator. The coordinator knows the common information and select prescriptions that map each controller's local information to its control actions. The optimal control problem at the coordinator is shown to be a partially observable Markov decision process (POMDP) which is solved using techniques from Markov decision theory. This approach provides (a) structural results for optimal strategies, and (b) a dynamic program for obtaining optimal strategies for all controllers in the original decentralized problem. Thus, this approach unifies the various ad-hoc approaches taken in the literature. In addition, the structural results on optimal control strategies obtained by the proposed approach cannot be obtained by the existing generic approach (the person-by-person approach) for obtaining structural results in decentralized problems; and the dynamic program obtained by the proposed approach is simpler than that obtained by the existing generic approach (the designer's approach) for obtaining dynamic programs in decentralized problems.

Index Terms

Decentralized Control, Stochastic Control, Information Structures, Markov Decision Theory, Team Theory

I. Introduction

Stochastic control theory provides analytic and computational techniques for centralized decision making in stochastic systems with noisy observations. For specific models such as Markov decision processes and linear quadratic and Gaussian systems, stochastic control gives

Preliminary version of this paper appeared in the proceedings of the 46th Allerton conference on communication, control, and computation, 2008 (see [1]).

results that are intuitively appealing and computationally tractable. However, these results are derived under the assumption that all decisions are made by a centralized decision maker who sees all observations and perfectly recalls past observations and actions. This assumption of a centralized decision maker is not true in a number of modern control applications such as networked control systems, communication and queuing networks, sensor networks, and smart grids. In such applications, decisions are made by multiple decision makers who have access to different information. In this paper, we investigate such problems of *decentralized stochastic control.*

The techniques from centralized stochastic control cannot be directly applied to decentralized control problems. Nonetheless, two general solution approaches that indirectly use techniques from centralized stochastic control have been used in the literature: (i) *the person-by-person approach* which takes the viewpoint of an individual decision maker (DM); and (ii) *the designer's approach* which takes the viewpoint of the collective team of DMs.

The person-by-person approach investigates the decentralized control problem from the viewpoint of one DM, say DM i and proceeds as follows: (i) arbitrarily fix the strategy of all DMs except DM i; and (ii) use centralized stochastic control to derive structural properties for the optimal best-response strategy of DM i. If such a structural property does not depend on the choice of the strategy of other DMs, then it also holds for globally optimal strategy of DM i. By cyclically using this approach for all DMs, we can identify the *structure* of globally optimal strategies for all DMs.

A variation of this approach may be used to identify person-by-person optimal strategies. The variation proceeds iteratively as follows. Start with an initial guess for the strategies of all DMs. At each iteration, select one DM (say DM i), and change its strategy to the best response strategy given the strategy of all other DMs. Repeat the process until a fixed point is reached, i.e., when no DM can improve performance by unilaterally changing its strategy. The resulting strategies are person-by-person optimal [2], and in general, not globally optimal.

In summary, the person-by-person approach identifies structural properties of globally optimal strategies and provides an iterative method to obtain person-by-person optimal strategies. This method has been successfully used to identify structural properties of globally optimal strategies for various applications including real-time communication [3]–[7], decentralized hypothesis testing and quickest change detection [8]–[16], and networked control systems [17]–[19]. Under

certain conditions, the person-by-person optimal strategies found by this approach are globally optimal [2], [20], [21].

The designer's approach, which is developed in [22], [23], investigates the decentralized control problem from the viewpoint of the collective team of DMs or, equivalently, from the viewpoint of a system designer who knows the system model and probability distribution of the primitive random variables and chooses control strategies for all DMs. Effectively, the designer is solving a centralized planning problem. The designer's approach proceeds by: (i) modeling this centralized planning problem as a multi-stage, *open-loop* stochastic control problem in which the designer's decision at each time is the control *law* for that time for all DMs; and (ii) using centralized stochastic control to obtain a dynamic programming decomposition. Each step of the resulting dynamic program is a functional optimization problem (in contrast to centralized dynamic programming where each step is a parameter optimization problem).

The designer approach is often used in tandem with the person-by-person approach as follows. First, the person-by-person approach is used to identify structural properties of globally optimal strategies. Then, restricting attention to strategies with the identified structural property, the designer's approach is used to obtain a dynamic programming decomposition for selecting the globally optimal strategy. Such a tandem approach has been used in various applications including real-time communication [4], [24], [25], decentralized hypothesis testing [13], and networked control systems [17], [18].

In addition to the above general approaches, other specialized approaches have been developed to address specific problems in decentralized systems. Decentralized problems with partially nested information structure were defined and studied in [26]. Decentralized linear quadratic Gaussian (LQG) control problems with two controllers and partially nested information structure were studied in [27], [28]. Partially nested decentralized LQG problems with controllers connected via a graph were studied in [29], [30]. A generalization of partial nestedness called stochastic nestedness was defined and studied in [31]. An important property of LQG control problems with partially nested information structure is that there exists an affine control strategy which is globally optimal. In general, the problem of finding the best affine control strategies may not be a convex optimization problem. Conditions under which the problem of determining optimal control strategies within the class of affine control strategies becomes a convex optimization problem were identified in [32], [33].

Decentralized stochastic control problems with specific models of information sharing among controllers have also been studied in the literature. Examples include systems with delayed sharing information structures [34]–[36], systems with periodic sharing information structure [37], control sharing information structure [38], [39], systems with broadcast information structure [19], and systems with common and private observations [1].

In this paper, we present a new general model of decentralized stochastic control called partial history sharing information structure. In this model, we assume that: (i) controllers sequentially share part of their past data (past observations and control) with each other by means of a shared memory; and (ii) all controllers have perfect recall of commonly available data (common information). This model subsumes a large class of decentralized control models in which information is shared among the controllers.

For this model, we present a general solution methodology that reformulates the original decentralized problem into an equivalent centralized problem from the perspective of a coordinator. The coordinator knows the common information and selects prescriptions that map each controller's local information to its control actions. The optimal control problem at the coordinator is shown to be a partially observable Markov decision process (POMDP) which is solved using techniques from Markov decision theory. This approach provides (a) structural results for optimal strategies, and (b) a dynamic program for obtaining optimal strategies for all controllers in the original decentralized problem. Thus, this approach unifies the various ad-hoc approaches taken in the literature.

A similar solution approach is used in [36] for a model that is a special case of the model presented in this paper. We present an information state (Eq. (51)) for the model of [36] that is simpler than that presented in [36, Theorem 2]. A preliminary version of the general solution approach presented here was presented in [1] for a model that had features (e.g., direct but noisy communication links between controllers) that are not necessary for partial history sharing. However, it can be shown that by suitable redefinition of variables, the model in [1] can be recast as an instance of the model in this paper and vice versa (see Appendix C). The information state for partial history sharing that is presented in this paper (see Thereom 4) is simpler than that presented in [1, Eq. (39)].

A. Common Information Approach for a Static Team Problem

We first illustrate how common information can be used in a static team problem with two controllers. Let X denote the state of nature and Y^*, Y^1, Y^2 be three correlated random variables that depend on X. Assume that the joint distribution of (X, Y^*, Y^1, Y^2) is given.

Controller i, $i = 1, 2$, observes (Y^*, Y^i) and chooses a control action $U^i = g^i(Y^*, Y^i)$. The system incurs a cost $l(X, U^1, U^2)$. The control objective is to choose (g^1, g^2) to minimize

$$J(g^1, g^2) \coloneqq \mathbb{E}^{(g^1, g^2)}[l(X, U^1, U^2)]$$

If all the system variables are finite valued, we can solve the above optimization problem by a brute force search over all control strategies (g^1, g^2). For example, if all variables are binary valued, we need to compute the performance of $2^4 \times 2^4 = 256$ control strategies and choose the one with the best performance.

In this example, both controllers have a common observation Y^*. One of the main ideas of this paper is to use such common information among the controllers to simplify the search process as follows. Instead of specifying the control strategies (g^1, g^2) directly, we consider a coordinated system in which a *coordinator* observes the common information Y^* and chooses *prescriptions* (Γ^1, Γ^2) where Γ^i is a mapping from Y^i to U^i, $i = 1, 2$. Hence, $(\Gamma^1, \Gamma^2) = d(Y^*)$, where d is called the *coordination strategy*. The coordinator then communicates these prescriptions to the controllers who simply use them to choose $U^i = \Gamma^i(Y^i)$, $i = 1, 2$.

It is easy to verify (see Proposition 3 for a formal proof) that choosing the control strategies (g^1, g^2) in the original system is equivalent to choosing a coordination strategy d in the coordinated system. The problem of choosing the best coordination strategy, however, is a centralized problem in which the coordinator is the only decision-maker.

For example, consider the case when all system variables are binary valued. For any coordination strategy d, let $(\gamma_0^1, \gamma_0^2) = d(0)$ and $(\gamma_1^1, \gamma_1^2) = d(1)$. Then, the cost associated with this coordination strategy is given as:

$$\begin{aligned} J(d) \coloneqq \mathbb{E}^{(d)}[l(X, U^1, U^2)] = \mathbb{P}(Y^* = 0)\mathbb{E}[l(X, \gamma_0^1(Y^1), \gamma_0^2(Y^2))|Y^* = 0] \\ + \mathbb{P}(Y^* = 1)\mathbb{E}[l(X, \gamma_1^1(Y^1), \gamma_1^2(Y^2))|Y^* = 1] \end{aligned}$$

To minimize the above cost, we can minimize the two terms separately. Therefore, to find the best coordination strategy d, we can search for optimal prescriptions for the cases $Y^* = 0$ and $Y^* = 1$

separately. Searching for the best prescriptions for each of these cases involves computing the performance of $2^2 \times 2^2 = 16$ prescription pairs and choosing the one with the best performance. Thus, to find the best coordination strategy, we need to evaluate the performance of $16 + 16 = 32$ prescription pairs. Contrast this with the 256 strategies whose costs we need to evaluate to solve the original problem by brute force.

The above example described a static system and illustrates that common information can be exploited to convert the decentralized optimization problem into a centralized optimization problem involving a coordinator. In this paper, we build upon this basic idea and present a solution approach based on common information that works for dynamical decentralized systems as well. Our approach converts the decentralized problem into a centralized stochastic control problem (in particular, a partially observable Markov decision process), identifies structure of optimal control strategies, and provides a dynamic program like decomposition for the decentralized problem.

B. Contributions of the Paper

We introduce a general model of decentralized stochastic control problem in which multiple controllers share part of their information with each other. We call this model the *partial history sharing information structure*. This model subsumes several existing models of information sharing in decentralized stochastic control as special cases (see Section II-B). We establish two results for our model. Firstly, we establish a structural property of optimal control strategies. Secondly, we provide a dynamic programming decomposition of the problem of finding optimal control strategies. As in [1], [36], our results are derived using a common information based approach (see Section III). This approach differs from the person-by-person approach and the designer's approach mentioned earlier. In particular, the structural properties found in this paper cannot be found by the person-by-person approach described earlier. Moreover, the dynamic programming decomposition found in this paper is distinct from —and simpler than— the dynamic programming decomposition based on the designer's approach. For a general framework for using common information in sequential decision making problems, see [40].

C. Notation

Random variables are denoted by upper case letters; their realization by the corresponding lower case letter. For integers $a \leq b$ and $c \leq d$, $X_{a:b}$ is a short hand for the vector $(X_a, X_{a+1}, \dots, X_b)$

while $X^{c:d}$ is a short hand for the vector $(X^c, X^{c+1}, \dots, X^d)$. When $a > b$, $X_{a:b}$ equals the empty set. The combined notation $X^{c:d}_{a:b}$ is a short hand for the vector $(X^j_i : i = a, a+1, \dots, b,$ $j = c, c+1, \dots, d)$. In general, subscripts are used as time index while superscripts are used to index controllers. Bold letters $\mathbf{X}$ are used as a short hand for the vector $(X^{1:n})$. $\mathbb{P}(\cdot)$ is the probability of an event, $\mathbb{E}(\cdot)$ is the expectation of a random variable. For a collection of functions $\boldsymbol{g}$, we use $\mathbb{P}^{\boldsymbol{g}}(\cdot)$ and $\mathbb{E}^{\boldsymbol{g}}(\cdot)$ to denote that the probability measure/expectation depends on the choice of functions in $\boldsymbol{g}$. $\mathbb{1}_A(\cdot)$ is the indicator function of a set A. For singleton sets $\{a\}$, we also denote $\mathbb{1}_{\{a\}}(\cdot)$ by $\mathbb{1}_a(\cdot)$.

For a singleton a and a set B, $\{a, B\}$ denotes the set $\{a\} \cup B$. For two sets A and B, $\{A, B\}$ denotes the set $A \cup B$. For two finite sets $\mathcal{A}, \mathcal{B}$, $F(\mathcal{A}, \mathcal{B})$ is the set of all functions from $\mathcal{A}$ to $\mathcal{B}$. Also, if $\mathcal{A} = \emptyset$, $F(\mathcal{A}, \mathcal{B}) := \mathcal{B}$. For a finite set $\mathcal{A}$, $\Delta(\mathcal{A})$ is the set of all probability mass functions over $\mathcal{A}$. For the ease of exposition, we assume that all state, observation and control variables take values in finite sets.

For two random variables (or random vectors) X and Y taking values in $\mathcal{X}$ and $\mathcal{Y}$, $\mathbb{P}(X = x|Y)$ denotes the conditional probability of the event $\{X = x\}$ given Y and $\mathbb{P}(X|Y)$ denotes the conditional PMF (probability mass function) of X given Y, that is, it denotes the collection of conditional probabilities $\mathbb{P}(X = x|Y), x \in \mathcal{X}$. Finally, all equalities involving random variables are to be interpreted as almost sure equalities (that is, they hold with probability one).

D. Organization

The rest of this paper is organized as follows. We present our model of a decentralized stochastic control problem in Section II. We also present several special cases of our model in this section. We prove our main results in Section III. We apply our result to some special cases in Section III-B. We present a simplification of our result and a generalization of our model in Section IV. We consider the infinite time-horizon discounted cost analogue of our problem in Section V. Finally, we conclude in Section VI.

II. Problem Formulation

A. Basic model: Partial History Sharing Information Structure

1) The Dynamic System: Consider a dynamic system with n controllers. The system operates in discrete time for a horizon T. Let $X_t \in \mathcal{X}_t$ denote the state of the system at time t, $U^i_t \in \mathcal{U}^i_t$

denote the control action of controller i, $i = 1, \ldots, n$ at time t, and $\mathbf{U}_t$ denote the vector $(U_t^1, \ldots, U_t^n)$.

The initial state X_1 has a probability distribution Q_1 and evolves according to

$$X_{t+1} = f_t(X_t, \mathbf{U}_t, W_t^0), \tag{1}$$

where $\{W_t^0\}_{t=1}^T$ is a sequence of i.i.d. random variables with probability distribution Q_W^0.

2) Data available at the controller: At any time t, each controller has access to three types of data: current observation, local memory, and shared memory.

(i) ***Current local observation***: Each controller makes a local observation $Y_t^i \in \mathcal{Y}_t^i$ on the state of the system at time t,

$$Y_t^i = h_t^i(X_t, W_t^i), \tag{2}$$

where $\{W_t^i\}_{t=1}^T$ is a sequence of i.i.d. random variables with probability distribution Q_W^i. We assume that the random variables in the collection $\{X_1, W_t^j, t = 1, \ldots, T, j = 0, 1, \ldots, n\}$, called *primitive random variables*, are mutually independent.

(ii) ***Local memory*** : Each controller stores a subset M_t^i of its past local observations and its past actions in a local memory:

$$M_t^i \subset \{Y_{1:t-1}^i, U_{1:t-1}^i\}. \tag{3}$$

At $t = 1$, the local memory is empty, $M_1^i = \emptyset$.

(iii) ***Shared memory***: In addition to its local memory, each controller has access to a shared memory. The contents C_t of the shared memory at time t are a subset of the past local observations and control actions of all controllers:

$$C_t \subset \{\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}\} \tag{4}$$

where $\mathbf{Y}_t$ and $\mathbf{U}_t$ denote the vectors $(Y_t^1, \ldots, Y_t^n)$ and $(U_t^1, \ldots, U_t^n)$ respectively. At $t = 1$, the shared memory is empty, $C_1 = \emptyset$.

Controller i chooses action U_t^i as a function of the total data (Y_t^i, M_t^i, C_t) available to it. Specifically, for every controller i, $i = 1, \ldots, n$,

$$U_t^i = g_t^i(Y_t^i, M_t^i, C_t), \tag{5}$$

where g_t^i is called the *control law* of controller i. The collection $\mathbf{g}^i = (g_1^i, \ldots, g_T^i)$ is called the *control strategy* of controller i. The collection $\mathbf{g}^{1:n} = (\mathbf{g}^1, \ldots, \mathbf{g}^n)$ is called the *control strategy* of the system.

3) Update of local and shared memories:

(i) *Shared memory update*: After taking the control action at time t, the local information at controller i consists of the contents M_t^i of its local memory, its local observation Y_t^i and its control action U_t^i. Controller i sends a subset Z_t^i of this local information $\{M_t^i, Y_t^i, U_t^i\}$ to the shared memory. The subset Z_t^i is chosen according to a pre-specified protocol. The contents of shared memory are nested in time, that is, the contents C_{t+1} of the shared memory at time $t+1$ are the contents C_t at time t augmented with the new data $\mathbf{Z}_t = (Z_t^1, Z_t^2, \dots, Z_t^n)$ sent by all the controllers at time t:

$$C_{t+1} = \{C_t, \mathbf{Z}_t\}. \tag{6}$$

(ii) *Local memory update*: After taking the control action and sending data to the shared memory at time t, controller i updates its local memory according to a pre-specified protocol. The content M_{t+1}^i of the local memory can at most equal the total local information $\{M_t^i, Y_t^i, U_t^i\}$ at the controller. However, to ensure that the local and shared memories at time $t+1$ don't overlap, we assume that

$$M_{t+1}^i \subset \{M_t^i, Y_t^i, U_t^i\} \setminus Z_t^i. \tag{7}$$

Figure 1 shows the time order of observations, actions and memory updates. We refer to the

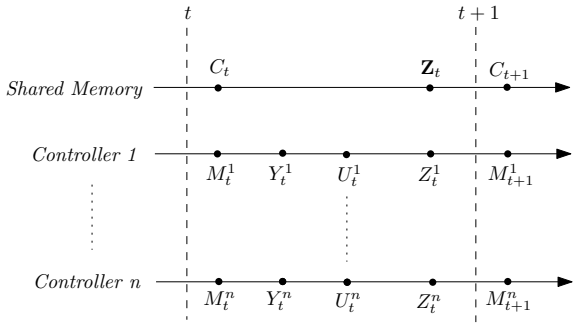

Fig. 1. Time ordering of Observations, Actions and Memory Updates

above model as the *partial history sharing information structure*.

4) The optimization problem: At time t, the system incurs a cost $l(X_t, \mathbf{U}_t)$. The performance of the control strategy of the system is measured by the expected total cost

$$J(\mathbf{g}^{1:n}) := \mathbb{E}^{\mathbf{g}^{1:n}}\Big[\sum_{t=1}^{T} l(X_t, \mathbf{U}_t)\Big], \tag{8}$$

where the expectation is with respect to the joint probability measure on $(X_{1:T}, \mathbf{U}_{1:T})$ induced by the choice of $\mathbf{g}^{1:n}$.

We are interested in the following optimization problem.

Problem 1 *For the model described above, given the state evolution functions f_t, the observation functions h_t^i, the protocols for updating local and share memory, the cost function l, the distributions Q_1, Q_W^i, $i = 0, 1, \ldots, n$, and the horizon T, find a control strategy $\mathbf{g}^{1:n}$ for the system that minimizes the expected total cost given by* (8)*.*

B. Special Cases: The Models

In the above model, although we have not specified the exact protocols by which controllers update the local and shared memories, we assume that pre-specified protocols are being used. Different choices of this protocol result in different information structures for the system. In this section, we describe several models of decentralized control systems that can be viewed as special cases of our model by assuming a particular choice of protocol for local and shared memory updates.

1) Delayed Sharing Information Structure: Consider the following special case of the model of Section II-A.

(i) The shared memory at the beginning of time t is $C_t = \{\mathbf{Y}_{1:t-s}, \mathbf{U}_{1:t-s}\}$, where $s \geq 1$ is a fixed number. The local memory at the beginning of time t is $M_t^i = \{Y_{t-s+1:t-1}^i, U_{t-s+1:t-1}^i\}$.

(ii) At each time t, after taking the action U_t^i, controller i sends $Z_t^i = \{Y_{t-s+1}^i, U_{t-s+1}^i\}$ to the shared memory and the shared memory at $t+1$ becomes $C_{t+1} = \{\mathbf{Y}_{1:t-s+1}, \mathbf{U}_{1:t-s+1}\}$.

(iii) After sending $Z_t^i = \{Y_{t-s+1}^i, U_{t-s+1}^i\}$ to the shared memory, controller i updates the local memory to $M_{t+1}^i = \{Y_{t-s+2:t}^i, U_{t-s+2:t}^i\}$.

In this spacial case, the observations and control actions of each controller are shared with every other controller after a delay of s time steps. Hence, the above special case corresponds to the delayed sharing information structure considered in [34], [36], [41].

2) Delayed State Sharing Information Structure: A special case of the delayed sharing information structure (which itself is a special case of our basic model) is the *delayed state sharing* information structure [35]. This information structure can be obtained from the delayed sharing information structure by making the following assumptions:

(i) The state of the system at time t is a n-dimensional vector $X_t = (X_t^1, X_t^2, \ldots, X_t^n)$.

(ii) At each time t, the current local observation of controller i is $Y_t^i = X_t^i$, for $i = 1, 2, \ldots, n$.

In this spacial case, the complete state vector X_t is available to all controllers after a delay of s time steps.

3) Periodic Sharing Information Structure: Consider the following special case of the model of Section II-A where controllers update the shared memory periodically with period $s \geq 1$:

(i) For time $ks < t \leq (k+1)s$, where $k = 0, 1, 2, \ldots$, the shared memory at the beginning of time t is $C_t = \{\mathbf{Y}_{1:ks}, \mathbf{U}_{1:ks}\}$. The local memory at the beginning of time t is $M_t^i = \{Y_{ks+1:t-1}^i, U_{ks+1:t-1}^i\}$.

(ii) At each time $t = (k+1)s$, $k = 1, 2, \ldots$, after taking the action U_t^i, controller i sends $Z_t^i = \{Y_{ks+1:(k+1)s}^i, U_{ks+1:(k+1)s}^i\}$ to the shared memory. At other times, each controller does not send anything (thus $Z_t^i = \emptyset$).

(iii) After sending Z_t^i to the shared memory, controller i updates the local memory to $M_{t+1}^i = \{M_t^i, Y_t^i, U_t^i\} \setminus Z_t^i$.

In this spacial case, the entire history of observations and control actions are shared periodically between controllers with period s. Hence, the above special case corresponds to the periodic sharing information structure considered in [37].

4) Control Sharing Information Structure: Consider the following special case of the model of Section II-A.

(i) The shared memory at the beginning of time t is $C_t = \{\mathbf{U}_{1:t-1}\}$. The local memory at the beginning of time t is $M_t^i = \{Y_{1:t-1}^i\}$.

(ii) At each time t, after taking the action U_t^i, controller i sends $Z_t^i = \{U_t^i\}$ to the shared memory.

(iii) After sending $Z_t^i = U_t^i$ to the shared memory, controller i updates the local memory to $M_{t+1}^i = Y_{1:t}^i$.

In this spacial case, the control actions of each controller are shared with every other controller after a delay of 1 time step. Hence, the above special case corresponds to the control sharing information structure considered in [38].

5) No Shared Memory with or without finite local memory: Consider the following special case of the model of Section II-A.

(i) The shared memory at each time is empty, $C_t = \emptyset$ and the local memory at the beginning of time t is $M_t^i = \{Y_{t-s:t-1}^i, U_{t-s:t-1}^i\}$, where $s \geq 1$ is a fixed number.

(ii) Controllers do not send any data to shared memory, $Z_t^i = \emptyset$.

(iii) At the end of time t, controllers update their local memories to $M_{t+1}^i = \{Y_{t-s+1:t}^i, U_{t-s+1:t}^i\}$.

In this special case, the controllers don't share any data. The above model is related to the finite-memory controller model of [42]. A related special case is the situation where the local memory at each controller consists of all of its past local observations and its past actions, that is, $M_t^i = \{Y_{1:t-1}^i, U_{1:t-1}^i\}$.

Remark 1 All the special cases considered above are examples of *symmetric sharing*. That is, different controllers update their local memories according to identical protocols and the data sent by a controller to the shared memory is selected according to identical protocols. However, this symmetry is not required for our model. Consider for example, the delayed sharing information structure where at the end of time t, controller i sends $Y_{t-s_i}^i, U_{t-s_i}^i$ to the shared memory, with $s_i, i = 1, 2, \ldots, n$, being fixed, but not necessarily identical, numbers. This kind of *asymmetric sharing* is also a special case of our model. □

III. Main Results

For centralized systems, stochastic control theory provides two important analytical results. Firstly, it provides a *structural result*. This result states that there is an optimal control strategy which selects control actions as a function only of the controller's posterior belief on the state of the system conditioned on all its observations and actions till the current time. The controller's posterior belief is called its *information state*. Secondly, stochastic control theory provides a *dynamic programming decomposition* of the problem of finding optimal control strategies in centralized systems. This dynamic programming decomposition allows one to evaluate the optimal action for each realization of the controller's information state in a backward inductive manner.

In this section, we provide a structural result and a dynamic programming decomposition for the decentralized stochastic control problem with partial information sharing formulated above (Problem 1). The main idea of the proof is to formulate an equivalent centralized stochastic control problem; solve the equivalent problem using classical stochastic-control techniques; and translate the results back to the basic model. For that matter, we proceed as follows:

1) Formulate a centralized *coordinated system* from the point of view of a *coordinator* that observes only the common information among the controllers in the basic model, i.e., the coordinator observes the shared memory C_t but not the local memories $(M_t^i, i = 1, \dots, n)$ or local observations $(Y_t^i, i = 1, \dots, n)$. The coordinator chooses prescriptions $\mathbf{\Gamma}_t = (\Gamma_t^1, \dots, \Gamma_t^n)$, where Γ_t^i is a mapping from (Y_t^i, M_t^i) to U_t^i, $i = 1, \dots, n$.
2) Show that the coordinated system is a POMDP (partially observable Markov decision process).
3) For the coordinated system, determine the structure of an optimal coordination strategy and a dynamic program to find an optimal coordination strategy.
4) Show that any strategy of the coordinated system is implementable in the basic model with the same value of the total expected cost. Conversely, any strategy of the basic model is implementable in the coordinated system with the same value of the total expected cost. Hence, the two systems are equivalent.
5) Translate the structural results and dynamic programming decomposition of the coordinated system (obtained in stage 3) to the basic model.

Stage 1: The coordinated system

Consider a *coordinated system* that consists of a coordinator and n passive controllers. The coordinator knows the shared memory C_t at time t, but not the local memories $(M_t^i, i = 1, \dots, n)$ or local observations $(Y_t^i, i = 1, \dots, n)$. At each time t, the coordinator chooses mappings $\Gamma_t^i : \mathcal{Y}_t^i \times \mathcal{M}_t^i \mapsto \mathcal{U}_t^i$, $i = 1, 2, \dots, n$, according to

$$\mathbf{\Gamma_t} = d_t(C_t, \mathbf{\Gamma_{1:t-1}}), \tag{9}$$

where $\mathbf{\Gamma_t} = (\Gamma_t^1, \Gamma_t^2, \dots, \Gamma_t^n)$. The function d_t is called the *coordination rule* at time t and the collection of functions $\mathbf{d} \coloneqq (d_1, \dots, d_T)$ is called the *coordination strategy*. The selected Γ_t^i is communicated to controller i at time t.

The function Γ_t^i tells controller i how to process its current local observation and its local memory at time t; for that reason, we call Γ_t^i the *coordinator's prescription* to controller i. Controller i generates an action using its prescription as follows:

$$U_t^i = \Gamma_t^i(Y_t^i, M_t^i). \tag{10}$$

For this coordinated system, the system dynamics, the observation model and the cost are the same as the basic model of Section II-A: the system dynamics are given by (1), each controller's current observation is given by (2) and the instantaneous cost at time t is $l(X_t, \mathbf{U}_t)$. As before, the performance of a coordination strategy is measured by the expected total cost

$$\hat{J}(\mathbf{d}) = \mathbb{E}\Big[\sum_{t=1}^{T} l(X_t, \mathbf{U}_t)\Big], \tag{11}$$

where the expectation is with respect to a joint measure on $(X_{1:T}, \mathbf{U}_{1:T})$ induced by the choice of $\mathbf{d}$.

In this coordinated system, we are interested in the following optimization problem:

Problem 2 *For the model of the coordinated system described above, find a coordination strategy* $\mathbf{d}$ *that minimizes the total expected cost given by* (11)*.*

Stage 2: The coordinated system as a POMDP

We will now show that the coordinated system is a partially observed Markov decision process. For that matter, we first describe the model of POMDPs [43].

POMDP Model: A partially observable Markov decision process consists of a state process $S_t \in \mathcal{S}$, an observation process $O_t \in \mathcal{O}$, an action process $A_t \in \mathcal{A}$, $t = 1, 2, \ldots, T$, and a single decision-maker where

1) The action at time t is chosen by the decision-maker as a function of observation and action history, that is,

$$A_t = d_t(O_{1:t}, A_{1:t-1}), \tag{12}$$

d_t is the decision rule at time t.

2) After the action at time t is taken, the new state and new observation are generated according to the transition probability rule

$$\mathbb{P}(S_{t+1}, O_{t+1} | S_{1:t}, O_{1:t}, A_{1:t}) = \mathbb{P}(S_{t+1}, O_{t+1} | S_t, A_t). \tag{13}$$

3) At each time, an instantaneous cost $\tilde{l}(S_t, A_t)$ is incurred.

4) The optimization problem for the decision-maker is to choose a decision strategy $\mathbf{d} := (d_1, \ldots, d_T)$ to minimize a total cost given as

$$\mathbb{E}[\sum_{t=1}^{T} \tilde{l}(S_t, A_t)]. \tag{14}$$

The following well-known results provides the structure of optimal strategies and a dynamic program for POMDPs. For details, see [43].

Theorem 1 (POMDP Result) *Let Θ_t be the conditional probability distribution of the state S_t at time t given the observations $O_{1:t}$ and actions $A_{1:t-1}$,*

$$\Theta_t(s) = \mathbb{P}(S_t = s | O_{1:t}, A_{1:t-1}), \quad s \in \mathcal{S}.$$

Then,

1) $\Theta_{t+1} = \eta_t(\Theta_t, A_t, O_{t+1})$*, where η_t is the standard non-linear filter: If θ_t, a_t, o_{t+1} are the realizations of Θ_t, A_t and O_{t+1}, then the realization of s^{th} element of the vector Θ_{t+1} is*

$$\begin{aligned}\theta_{t+1}(s) &= \frac{\sum_{s'} \theta_t(s') \mathbb{P}(S_{t+1} = s, O_{t+1} = o_{t+1} | S_t = s', A_t = a_t)}{\sum_{\hat{s},\tilde{s}} \theta_t(\hat{s}) \mathbb{P}(S_{t+1} = \tilde{s}, O_{t+1} = o_{t+1} | S_t = \hat{s}, A_t = a_t)} \\ &=: \eta_t^s(\theta_t, a_t, o_{t+1}) \end{aligned} \tag{15}$$

   and $\eta_t(\theta_t, a_t, o_{t+1})$ is the vector $(\eta_t^s(\theta_t, a_t, o_{t+1}))_{s \in \mathcal{S}}$.

2) *There exists an optimal decision strategy of the form*

$$A_t = \hat{d}_t(\Theta_t).$$

□

   Further, such a strategy can be found by the following dynamic program:

$$V_T(\theta) = \inf_a \mathbb{E}\{\tilde{l}(S_T, a) | \Theta_T = \theta\}, \tag{16}$$

   and for $1 \le t \le T-1$,

$$V_t(\theta) = \inf_a \mathbb{E}\big\{\tilde{l}(S_t, a) + V_{t+1}(\eta_t(\theta, a, O_{t+1})) \big| \Theta_t = \theta, A_t = a\big\}. \tag{17}$$

We will now show that the coordinated system can be viewed as an instance of the above POMDP model by defining the state process as $S_t := \{X_t, \mathbf{Y}_t, \mathbf{M}_t\}$, the observation process as $O_t := \mathbf{Z}_{t-1}$, and the action process $A_t := \mathbf{\Gamma}_t$.

Lemma 1 *For the coordinated system of Problem 2,*

1) *There exist functions $\tilde{f}_t$ and $\tilde{h}_t$, $t = 1, \dots, T$, such that*

$$S_{t+1} = \tilde{f}_t(S_t, \mathbf{\Gamma}_t, W_t^0, \mathbf{W}_{t+1}), \tag{18}$$

   and

$$\mathbf{Z}_t = \tilde{h}_t(S_t, \mathbf{\Gamma}_t). \tag{19}$$

In particular, we have that

$$\mathbb{P}(S_{t+1}, \mathbf{Z}_t | S_{1:t}, \mathbf{Z}_{1:t-1}, \mathbf{\Gamma}_{1:t}) = \mathbb{P}(S_{t+1}, \mathbf{Z}_t | S_t, \mathbf{\Gamma}_t). \tag{20}$$

2) *Furthermore, there exists a function* $\tilde{l}$ *such that*

$$l(X_t, \mathbf{U}_t) = \tilde{l}(S_t, \mathbf{\Gamma}_t). \tag{21}$$

Thus, the objective of minimizing (11) *is same as minimizing*

$$\hat{J}(\mathbf{d}) = \mathbb{E}\Big[\sum_{t=1}^{T} \tilde{l}(S_t, \mathbf{\Gamma}_t)\Big]. \tag{22}$$

Proof: The existence of $\tilde{f}_t$ follows from (1), (2), (10), (7) and the definition of S_t. The existence of $\tilde{h}_t$ follows from the fact that Z_t^i is a fixed subset of $\{M_t^i, Y_t^i, U_t^i\}$, equation (10) and the definition of S_t. Equation (20) follows from (18) and the independence of $W_t^0, \mathbf{W}_{t+1}$ from all random variables in the conditioning in the left hand side of (20). The existence of $\tilde{l}$ follows from the definition of S_t and (10). ■

Recall that the coordinator is choosing its actions according to a coordination strategy of the form

$$\mathbf{\Gamma_t} = d_t(C_t, \mathbf{\Gamma_{1:t-1}}) = d_t(\mathbf{Z}_{1:t-1}, \mathbf{\Gamma_{1:t-1}}). \tag{23}$$

Equation (23) and Lemma 1 imply that the coordinated system is an instance of the POMDP model described above.

Stage 3: Structural result and dynamic program for the coordinated system

Since the coordinated system is a POMDP, Theorem 1 gives the structure of the optimal coordination strategies. For that matter, define coordinator's information state

$$\Pi_t := \mathbb{P}(S_t \mid \mathbf{Z}_{1:t-1}, \mathbf{\Gamma}_{1:t-1}) = \mathbb{P}(S_t \mid C_t, \mathbf{\Gamma}_{1:t-1}). \tag{24}$$

Then, we have the following:

Proposition 1 *For Problem 2, there is no loss of optimality in restricting attention to coordination rules of the form*

$$\mathbf{\Gamma}_t = \hat{d}_t(\Pi_t). \tag{25}$$

Furthermore, an optimal coordination strategy of the above form can be found using a dynamic program. For that matter, observe that we can write

$$\Pi_{t+1} = \eta_t(\Pi_t, \mathbf{Z}_t, \mathbf{\Gamma}_t) \tag{26}$$

where η_t is the standard non-linear filtering update function (see Appendix A). We denote by $\mathcal{B}_t$ the space of possible realizations of Π_t. Thus,

$$\mathcal{B}_t := \Delta(\mathcal{X}_t \times \mathcal{Y}_t^1 \times \mathcal{M}_t^1 \times \ldots \times \mathcal{Y}_t^n \times \mathcal{M}_t^n). \tag{27}$$

Recall that $F(\mathcal{Y}_t^i \times \mathcal{M}_t^i, \mathcal{U}_t^i)$ is the set of all functions from $\mathcal{Y}_t^i \times \mathcal{M}_t^i$ to $\mathcal{U}_t^i$ (see Section I-C). Then, we have the following result.

Proposition 2 *For all π_t in $\mathcal{B}_t$, define*

$$V_T(\pi) = \inf_{\{\tilde{\gamma}_T^i \in F(\mathcal{Y}_T^i \times \mathcal{M}_T^i, \mathcal{U}_T^i), 1 \le i \le n\}} \mathbb{E}[\tilde{l}(S_t, \mathbf{\Gamma}_T) \mid \Pi_t = \pi, \mathbf{\Gamma}_T = (\gamma_T^1, \ldots, \gamma_T^n)], \tag{28}$$

and for $1 \le t \le T-1$,

$$V_t(\pi) = \inf_{\{\tilde{\gamma}^i \in F(\mathcal{Y}_t^i \times \mathcal{M}_t^i, \mathcal{U}_t^i), 1 \le i \le n\}} \mathbb{E}[\tilde{l}(S_t, \mathbf{\Gamma}_t) + V_{t+1}(\eta_t(\Pi_t, \mathbf{\Gamma}_t, \mathbf{Z}_t) \mid \Pi_t = \pi, \mathbf{\Gamma}_t = (\gamma_t^1, \ldots, \gamma_t^n)]. \tag{29}$$

Then the $\arg\inf$ *at each time step gives the coordinator's optimal prescriptions for the controllers when the coordinator's information state is π.* □

Proposition 2 gives a dynamic program for the coordinator's problem (Problem 2). Since the coordinated system is a POMDP, it implies that computational algorithms for POMDPs can be used to solve the dynamic program for the coordinator's problem as well. We refer the reader to [44] and references therein for a review of algorithms to solve POMDPs.

Stage 4: Equivalence between the two models

We first observe that since $C_s \subset C_t$, for all $s < t$, under any given coordination strategy $\mathbf{d}$, we can use C_t to evaluate the past prescriptions by recursive substitution. For example, for $t = 2, 3$, the past prescriptions can be evaluated as functions of C_2, C_3 as follows:

$$\mathbf{\Gamma}_1 = d_1(C_1) =: \tilde{d}_1(C_2),$$

$$\mathbf{\Gamma}_2 = d_2(C_2, \mathbf{\Gamma}_1) = d_2(C_2, \tilde{d}_1(C_2)) =: \tilde{d}_2(C_3)$$

We can now state the following result.

Proposition 3 *The basic model of Section II-A and the coordinated system are equivalent. More precisely:*

(a) *Given any control strategy* $\mathbf{g}^{1:n}$ *for the basic model, choose a coordination strategy* $\mathbf{d}$ *for the coordinated system of stage 1 as*

$$d_t(C_t) = \big(g_t^1(\cdot,\cdot,C_t), \ldots, g_t^n(\cdot,\cdot,C_t)\big).$$

Then $\hat{J}(\mathbf{d}) = J(\mathbf{g}^{1:n})$.

(b) *Conversely, for any coordination strategy for the coordinated system, choose a control strategy* $\mathbf{g}^{1:n}$ *for the basic model as*

$$g_1^i(\cdot,\cdot,C_1) = d_1^i(C_1),$$

and

$$g_t^i(\cdot,\cdot,C_t) = d_t^i(C_t, \mathbf{\Gamma}_{1:t-1}),$$

where $\mathbf{\Gamma}_k = d_k(C_k, \mathbf{\Gamma}_{1:k-1})$, $k = 1, 2, \ldots, t-1$ *and* $d_t^i(\cdot)$ *is the* i*-th component of* $d_t(\cdot)$ *(that is,* $d_t^i(\cdot)$ *gives the coordinator's prescription for the* i*-th controller). Then,* $J(\mathbf{g}^{1:n}) = \hat{J}(\mathbf{d})$.□

Proof: See Appendix B. ■

Stage 5: Structural result and dynamic program for the basic model

Combining Proposition 1 with Proposition 3, we get the following structural result for Problem 1.

Theorem 2 (Structural Result for Optimal Control Strategies) *In Problem 1, there exist optimal control strategies of the form*

$$U_t^i = \hat{g}_t^i(Y_t^i, M_t^i, \Pi_t), \quad i = 1, 2, \ldots, n, \tag{30}$$

where Π_t *is the conditional distribution on* $X_t, \mathbf{Y}_t, \mathbf{M}_t$ *given* C_t*, defined as*

$$\Pi_t(x, \mathbf{y}, \mathbf{m}) := \mathbb{P}^{\hat{g}_{1:t-1}^{1:n}}(X_t = x, \mathbf{Y}_t = \mathbf{y}, \mathbf{M}_t = \mathbf{m} | C_t), \tag{31}$$

for all possible realizations $(x, \mathbf{y}, \mathbf{m})$ *of* $(X_t, \mathbf{Y}_t, \mathbf{M}_t)$*.* □

We call Π_t the *common information state*. Recall that Π_t takes values in the set $\mathcal{B}_t$ defined in (27).

Consider a control strategy $\hat{\mathbf{g}}^{\mathbf{i}}$ for controller i of the form specified in Theorem 2. The control law $\hat{g}_t^i$ at time t is a function from the space $\mathcal{Y}_t^i \times \mathcal{M}_t^i \times \mathcal{B}_t$ to the space of decisions $\mathcal{U}_t^i$. Equivalently, the control law $\hat{g}_t^i$ can be represented as a collection of functions $\{\hat{g}_t^i(\cdot,\cdot,\pi)\}_{\pi\in\mathcal{B}_t}$, where each element of this collection is a function from $\mathcal{Y}_t^i \times \mathcal{M}_t^i$ to $\mathcal{U}_t^i$. An element $\hat{g}_t^i(\cdot,\cdot,\pi)$ of this collection specifies a control action for each possible realization of Y_t^i, M_t^i and a fixed realization π of Π_t. We call $\hat{g}_t^i(\cdot,\cdot,\pi)$ the *partial control law* of controller i at time t for the given realization π of the common information state Π_t.

We now use Proposition 2 to describe a dynamic programming decomposition of the problem of finding optimal control strategies. This dynamic programming decomposition allows us to evaluate optimal *partial control laws* for each realization π of the common information state in a backward inductive manner. Recall that $\mathcal{B}_t$ is the space of all possible realizations of Π_t (see (27)) and $F(\mathcal{Y}_t^i \times \mathcal{M}_t^i, \mathcal{U}_t^i)$ is the set of all functions from $\mathcal{Y}_t^i \times \mathcal{M}_t^i$ to $\mathcal{U}_t^i$ (see Section I-C).

Theorem 3 (Dynamic Programming Decomposition) *Define the functions $V_t : \mathcal{B}_t \mapsto \mathbb{R}$, for $t = 1, \dots, T$ as follows:*

$$V_T(\pi) = \inf_{\{\tilde{\gamma}_T^i \in F(\mathcal{Y}_T^i \times \mathcal{M}_T^i, \mathcal{U}_T^i), 1\le i\le n\}} \mathbb{E}\{l(X_T, \tilde{\gamma}_T^1(Y_T^1, M_T^1), \dots, \tilde{\gamma}_T^n(Y_T^n, M_T^n)) | \Pi_T = \pi\}, \tag{32}$$

and for $1 \le t \le T-1$,

$$V_t(\pi) = \inf_{\{\tilde{\gamma}_t^i \in F(\mathcal{Y}_t^i \times \mathcal{M}_t^i, \mathcal{U}_t^i), 1\le i\le n\}} \mathbb{E}\big\{l(X_t, \tilde{\gamma}_t^1(Y_t^1, M_t^i), \dots, \tilde{\gamma}_t^n(Y_t^n, M_t^n)) + V_{t+1}(\eta_t(\pi, \tilde{\gamma}_t^1, \dots, \tilde{\gamma}_t^n, \mathbf{Z}_t)) \,\big|\, \Pi_t = \pi\big\}, \tag{33}$$

where η_t is a $\mathcal{B}_{t+1}$-valued function defined in (26) *and Appendix A.*

For $t = 1, \dots, T$ and for each $\pi \in \mathcal{B}_t$, an optimal partial control law for controller i is the minimizing choice of $\tilde{\gamma}^i$ in the definition of $V_t(\pi)$. Let $\Psi_t(\pi)$ denote the $\arg\inf$ *of the right hand side of $V_t(\pi)$, and Ψ_t^i denote its i-th component. Then, an optimal control stategy is given by:*

$$\hat{g}_t^i(\cdot,\cdot,\pi) = \Psi_t^i(\pi). \tag{34}$$

A. Comparison with Person by Person and Designer Approaches

The common information based approach adopted above differs from the person-by-person approach and the designer's approach mentioned in the introduction. In particular, the structural

result of Theorem 2 cannot be found by the person-by- person approach. If we fix strategies of all but the ith controller to an arbitrary choice, then it is *not necessarily optimal* for controller i to use a strategy of the form in Theorem 2. This is because if controller j's strategy uses the entire common information C_t, then controller i, in general, would need to consider the entire common information to better predict controller j's actions and hence controller i's optimal choice of action may too depend on the entire common information. The use of common information based approach allowed us to prove that *all controllers can jointly use strategies of the form in Theorem 2 without loss of optimality*.

The dynamic programming decomposition of Theorem 3 is simpler than any dynamic programming decomposition obtained using the designer's approach. As described earlier, the designer's approach models the decentralized control problem as an *open-loop* centralized planning problem in which a designer at each stage chooses control laws g_t^i that map (Y_t^i, M_t^i, C_t) to U_t^i, $i = 1, \dots, n$. On the other hand, the common-information approach developed in this paper models the decentralized control problem as a *closed-loop* centralized planning problem in which a coordinator at each stage chooses the *partial* control laws γ_t^i that map (Y_t^i, M_t^i) to U_t^i, $i = 1, \dots, n$. The space of partial control laws is always smaller than the space of full control laws; if the common information is non-empty, then they are strictly smaller. Thus, the dynamic programming decomposition of Theorem 3 is simpler than that obtained by the designer's approach. This simplification is best illustrated by the example of Section IV-C1 where all controllers receive a common observation Y_t^{com}. For this example, we show that our information state (and hence our dynamic program) reduce to $\mathbb{P}(X_t|Y_{1:t}^{com})$, which is identical to the information state of centralized stochastic control. In contrast, the information state $\mathbb{P}(X_t, Y_{1:t}^{com})$ obtained by the designer's approach is much more complicated.

B. Special Cases: The Results

In Section II-B, we described several models of decentralized control problems that are special cases of the model described in Section II-A. In this section, we state the results of Theorems 2 and 3 for these models.

1) Delayed Sharing Information Structure:

Corollary 1 *In the delayed sharing information structure of section II-B1, there exist optimal*

control strategies of the form

$$U_t^i = \hat{g}_t^i(Y_{t-s+1:t}^i, U_{t-s+1:t-1}^i, \Pi_t), \quad i = 1, 2, \dots, n, \tag{35}$$

where

$$\Pi_t := \mathbb{P}^{\hat{g}_{1:t-1}^{1:n}}(X_t, \mathbf{Y}_{t-s+1:t}, \mathbf{U}_{t-s+1:t-1} | C_t). \tag{36}$$

Moreover, optimal control strategies can be obtained by a dynamic program similar to that of Theorem 3. □

The above result is analogous to the result in [36].

2) Delayed State Sharing Information Structure:

Corollary 2 *In the delayed state sharing information structure of section II-B2, there exist optimal control strategies of the form*

$$U_t^i = \hat{g}_t^i(X_{t-s+1:t}^i, U_{t-s+1:t-1}^i, \Pi_t), \quad i = 1, 2, \dots, n, \tag{37}$$

where

$$\Pi_t := \mathbb{P}^{\hat{g}_{1:t-1}^{1:n}}(X_{t-s+1:t}, \mathbf{U}_{t-s+1:t-1} | C_t). \tag{38}$$

Moreover, optimal control strategies can be obtained by a dynamic program similar to that of Theorem 3. □

The above result is analogous to the result in [36].

3) Periodic Sharing Information Structure:

Corollary 3 *In the periodic sharing information structure of section II-B3, there exist optimal control strategies of the form*

$$U_t^i = \hat{g}_t^i(Y_{ks+1:t}^i, U_{ks+1:t-1}^i, \Pi_t), \quad i = 1, 2, \dots, n, \quad ks < t \leq (k+1)s, \tag{39}$$

where

$$\Pi_t := \mathbb{P}^{\hat{g}_{1:t-1}^{1:n}}(X_t, \mathbf{Y}_{ks+1:t}, \mathbf{U}_{ks+1:t-1} | C_t), \quad ks < t \leq (k+1)s. \tag{40}$$

Moreover, optimal control strategies can be obtained by a dynamic program similar to that of Theorem 3. □

The above result gives a finer dynamic programming decomposition that [37]. In [37], the dynamic programming decomposition is only carried out at the times of information sharing, $t = ks$, $s = 1, 2, \dots$; and at each step the partial control laws until the next sharing instant are chosen. In contrast, in the above dynamic program, the partial control laws of each step are chosen sequentially.

4) Control Sharing Information Structure:

Corollary 4 *In the control sharing information structure of section II-B4, there exist optimal control strategies of the form*

$$U_t^i = \hat{g}_t^i(Y_{1:t}^i, \Pi_t), \quad i = 1, 2, \dots, n, \tag{41}$$

where

$$\Pi_t := \mathbb{P}^{\hat{g}_{1:t-1}^{1:n}}(X_t, \mathbf{Y}_{1:t} | C_t). \tag{42}$$

Moreover, optimal control strategies can be obtained by a dynamic program similar to that of Theorem 3. □

5) No Shared Memory with or without finite local memory:

Corollary 5 *In the information structure of Section II-B5, there exist optimal control strategies of the form*

$$U_t^i = \hat{g}_t^i(Y_t^i, M_t^i, \Pi_t) \tag{43}$$

where

$$\Pi_t = \mathbb{P}^{\hat{g}_{1:t-1}^{1:n}}(X_t, \mathbf{Y}_t, \mathbf{M}_t) \tag{44}$$

Moreover, optimal control strategies can be obtained by a dynamic program similar to that of Theorem 3. □

Note that, since the common information is empty, the common information state Π_t is now an *unconditional* probability. In particular, Π_t is a constant random variable and takes a fixed value that depends only on the choice of past control laws. Therefore, we can define an appropriate control law $\tilde{g}_t^i$ such that $\hat{g}_t^i(Y_t^i, M_t^i, \Pi_t) = \tilde{g}_t^i(Y_t^i, M_t^i)$, with probability 1. Hence, the structural result of (43) may be simplified to

$$U_t^i = \hat{g}_t^i(Y_t^i, M_t^i, \Pi_t) = \tilde{g}_t^i(Y_t^i, M_t^i).$$

This result is redundant since all control laws are of the above form. Nonetheless, Corollary 5 gives a procedure of finding such control laws using the dynamic program of Theorem 3.

The above result is similar to the results in [42] for the case of one controller with finite memory and to those in [23] for the case of two controllers with finite memories.

IV. Simplifications and Generalizations

A. Simplification of the Common Information State

Theorems 2 and 3 identify the conditional probability distribution on $(X_t, \mathbf{Y}_t, \mathbf{M}_t)$ given C_t as the common information state for our problem. In the following lemma, we make the simple observation that in our model the conditional distribution on $(X_t, \mathbf{Y}_t, \mathbf{M}_t)$ given C_t is completely determined by the conditional distribution on $(X_t, \mathbf{M}_t)$ given C_t.

Lemma 2 *For any choice of control laws $\hat{g}^{1:n}_{1:t-1}$, define the conditional distribution on $X_t, \mathbf{M}_t$ given C_t as*

$$\Pi^{new}_t(x, \mathbf{m}) := \mathbb{P}^{\hat{g}^{1:n}_{1:t-1}}(X_t = x, \mathbf{M}_t = \mathbf{m} | C_t),$$

for all possible realizations $(x, \mathbf{m})$ of $(X_t, \mathbf{M}_t)$. Also define $\mathcal{B}^{new}_t := \Delta(\mathcal{X}_t \times \mathcal{M}^i_t \times \ldots \times \mathcal{M}^n_t)$. Then,

$$\Pi^{new}_t(x, \mathbf{m}) = \sum_{\mathbf{y}} \Pi_t(x, \mathbf{y}, \mathbf{m}). \tag{45}$$

Therefore, $\Pi^{new}_t = \chi_t(\Pi_t)$, where each component of the $\mathcal{B}^{new}_t$- valued function χ_t is determined by the right hand side of (45)*. Also,*

$$\Pi_t(x, \mathbf{y}, \mathbf{m}) = \Pi^{new}_t(x, \mathbf{m}) \mathbb{P}(\mathbf{Y}_t = \mathbf{y} | X_t = x), \tag{46}$$

where the second term on right hand side of (46) *is determined by the fixed distribution of the observations noises. Therefore, $\Pi_t = \zeta_t(\Pi^{new}_t)$, where each component of the $\mathcal{B}_t$- valued function ζ_t is determined by the right hand side of* (46)*.* □

Lemma 2 implies that the results of Theorems 2 and 3 can be written in terms of Π^{new}_t.

Theorem 4 (Alternative Common Information State) *In Problem 1, there exist optimal control strategies of the form*

$$U^i_t = \hat{\hat{g}}^i_t(Y^i_t, M^i_t, \Pi^{new}_t), \quad i = 1, 2, \ldots, n, \tag{47}$$

where

$$\Pi_t^{new} := \mathbb{P}^{\hat{g}^{1:n}_{1:t-1}}(X_t, \mathbf{M}_t | C_t). \tag{48}$$

Further, define the functions $V_t^{new} : \mathcal{B}_t^{new} \mapsto \mathbb{R}$ *, for* $t = 1, \dots, T$ *as follows:*

$$V_T^{new}(\pi^{new}) = \inf_{\{\tilde{\gamma}_T^i \in F(\mathcal{Y}_T^i \times \mathcal{M}_T^i, \mathcal{U}_T^i), 1 \le i \le n\}} \mathbb{E}\{l(X_T, \tilde{\gamma}_T^1(Y_T^1, M_T^1), \dots, \tilde{\gamma}_T^n(Y_T^n, M_T^n)) | \Pi_T = \zeta_T(\pi^{new})\}, \tag{49}$$

and for $1 \le t \le T-1$*,*

$$V_t^{new}(\pi^{new}) = \inf_{\{\tilde{\gamma}_t^i \in F(\mathcal{Y}_t^i \times \mathcal{M}_t^i, \mathcal{U}_t^i), 1 \le i \le n\}} \mathbb{E}\big\{ l(X_t, \tilde{\gamma}_t^1(Y_t^1, M_t^i), \dots, \tilde{\gamma}_t^n(Y_t^n, M_t^n)) + V_{t+1}^{new}(\chi_t(\eta_t(\Pi_t, \tilde{\gamma}_t^1, \dots, \tilde{\gamma}_t^n, \mathbf{Z}_t))) \,\big|\, \Pi_t = \zeta_t(\pi^{new}) \big\}, \tag{50}$$

where ζ_t, χ_t *are defined in Lemma 2, and* η_t *is defined in* (26) *and Appendix A.*

For $1 \le t \le T$ *and for each* π^{new}*, an optimal partial control law for controller* i *is the minimizing choice of* $\tilde{\gamma}^i$ *in the definition of* $V_t^{new}(\pi^{new})$*.* □

Proof: For any $\pi^{new} \in \mathcal{B}_t^{new}$ and any $\pi \in \mathcal{B}_t$, it is straightforward to establish using a backward induction argument that $V_t^{new}(\pi^{new}) = V_t(\zeta_t(\pi^{new}))$ and $V_t(\pi) = V_t^{new}(\chi_t(\pi))$, where $V_t(\cdot)$ is the value function from the dynamic program in Theorem 3. The optimality of the new dynamic program then follows from the optimality of the dynamic program in Theorem 3. ■

The result of Theorem 4 is conceptually the same as the results in Theorems 2 and 3. Theorem 4 implies that the Corollaries of Section III-B can be restated in terms of new information states by simply removing Y_t from the definition of original information states. For example, the result of Corollary 1 for delayed sharing information structure is also true when Π_t is replaced by

$$\Pi_t^{new} := \mathbb{P}^{\hat{g}^{1:n}_{1:t-1}}(X_t, \mathbf{Y}_{t-s+1:t-1}, \mathbf{U}_{t-s+1:t-1} | C_t). \tag{51}$$

This result is simpler than that of [36, Theorem 2].

B. Generalization of the Model

The methodology described in Section III relies on the fact that the shared memory is *common information* among all controllers. Since the coordinator in the coordinated system knows only the common information, any coordination strategy can be mapped to an equivalent control strategy in the basic model (see Stage 4 of Section III). In some cases, in addition to the shared

memory, the current observation (or if the current observation is a vector, some components of it) may also be commonly available to all controllers. The general methodology of Section 2 can be easily modified to include such cases as well.

Consider the model of Section II-A with the following modifications:

1) In addition to their current local observation, all controllers have a *common observation* at time t.

$$Y_t^{com} = h_t^{com}(X_t, V_t) \tag{52}$$

where $\{V_t, t = 1, \dots, T\}$ is a sequence of i.i.d. random variables with probability distribution Q_V which is independent of all other primitive random variables.

2) The shared memory C_t at time t is a subset of $\{Y_{1:t-1}^{com}, \mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}\}$.

3) Each controller selects its action using a control law of the form

$$U_t^i = g_t^i(Y_t^i, M_t^i, C_t, Y_t^{com}). \tag{53}$$

4) After taking the control action at time t, controller i sends a subset Z_t^i of $\{M_t^i, Y_t^i, U_t^i, Y_t^{com}\}$ that necessarily includes Y_t^{com}. That is,

$$Y_t^{com} \subset Z_t^i \subset \{M_t^i, Y_t^i, U_t^i, Y_t^{com}\}.$$

This implies that the history of common observations is necessarily a part of the shared memory, that is, $Y_{1:t-1}^{com} \subset C_t$.

The rest of the model is same as in Section II-A. In particular, the local memory update satisfies (7), so the local memory and shared memory at time $t+1$ don't overlap. The instantaneous cost is given by $l(X_t, U_t)$ and the objective is to minimize an expected total cost given by (8).

The arguments of Section III are also valid for this model. The observation process in Lemma 1 is now defined as $R_{t+1} = \{\mathbf{Z}_t, Y_{t+1}^{com}\}$. The analysis of Section III leads to structural results and dynamic programming decompositions analogous to Theorems 2 and 3 with Π_t now defined as

$$\Pi_t := \mathbb{P}^{g_{1:t-1}^{1:n}}(X_t, \mathbf{Y}_t, \mathbf{M}_t | C_t, Y_t^{com}). \tag{54}$$

Using an argument similar to Lemma 2, we can show that the result of Theorem 4 is true for the above model with Π_t^{new} defined as

$$\Pi_t^{new} := \mathbb{P}^{\hat{g}_{1:t-1}^{1:n}}(X_t, \mathbf{M}_t | C_t, Y_t^{com}). \tag{55}$$

C. Examples of the Generalized Model

1) Controllers with Identical Information: Consider the following special case of the above generalized model.

1) All controllers only make the common observation Y_t^{com}; controllers have no local observation or local memory.
2) The shared memory at time t is $C_t = Y_{1:t-1}^{com}$. Thus, at time t, all controllers have identical information given as $\{C_t, Y_t^{com}\} = Y_{1:t}^{com}$.
3) After taking the action at time t, each controller sends $Z_t^i = Y_t^{com}$ to the shared memory.

Recall that the coordinator's prescription Γ_t^i in Section III are chosen from the set of functions from $\mathcal{Y}_t^i \times \mathcal{M}_t^i$ to $\mathcal{U}_t^i$. Since, in this case $\mathcal{Y}_t^i = \mathcal{M}_t^i = \emptyset$, we interpret the coordinator's prescription as prescribed actions. That is, $\Gamma_t^i \equiv U_t^i$. With this interpretation, the common information state becomes

$$\Pi_t := \mathbb{P}^{g_{1:t-1}^{1:n}}(X_t | Y_{1:t}^{com}) \tag{56}$$

and the dynamic program of Theorem 3 becomes

$$V_T(\pi) = \inf_{\{u_T^i \in \mathcal{U}_T^i), 1 \le i \le n\}} \mathbb{E}\{l(X_T, u_T^1, \dots, u_T^n) | \Pi_T = \pi\}, \tag{57}$$

and for $1 \le t \le T-1$,

$$V_t(\pi) = \inf_{\{u_t^i \in \mathcal{U}_T^i), 1 \le i \le n\}} \mathbb{E}\big\{ l(X_t, u_t^1, \dots, u_t^n) + V_{t+1}(\eta_t(\pi, u_t^1, \dots, u_t^n, Y_{t+1}^{com})) \,\big|\, \Pi_t = \pi \big\}. \tag{58}$$

Since all the controllers have identical information, the above results correspond to the centralized dynamic program of Theorem 1 with a single controller choosing all the actions.

2) Coupled subsystems with control sharing information structure: Consider the following special case of the above generalized model.

1) The state of the system at time t is a $(n+1)$-dimensional vector $X_t = (X_t^1, X_t^2, \dots, X_t^n, X_t^*)$, where X_t^i, $i = 1, \dots, n$ corresponds to the local state of subsystem i, and X_t^* is a global state of the system.
2) The state update function is such that the global state evolves according to

$$X_{t+1}^* = f_t^*(X_t^*, \mathbf{U}_t, N_t^0),$$

while the local state of subsystem i evolves according to

$$X_{t+1}^i = f_t^i(X_t^i, X_t^*, \mathbf{U}_t, N_t^i),$$

where $\{N_t^0, t = 1, \dots T\}, \dots, \{N_t^n, t = 1, \dots T\}$ are mutually independent i.i.d noise processes that are independent of the initial state, $X_1 = (X_1^1, X_1^2, \dots, X_1^n, X_1^*)$.

3) At time t, the common observation of all controllers is given by $Y_t^{com} = X_t^*$.
4) At time t, the local observation of controller i is given by $Y_t^i = X_t^i$, $i = 1, \dots, n$.
5) The shared memory at time t is $C_t = \{X_{1:t-1}^*, \mathbf{U}_{1:t-1}\}$. At each time t, after taking the action U_t^i, controller i sends $Z_t^i = \{X_t^*, U_t^i\}$ to the shared memory.

The above special case corresponds to the model of coupled subsystems with control sharing considered in [39], where several applications of this model are also presented. It is shown in [39] that there is no loss of optimality in restricting attention to controllers with no local memory, i.e., $M_t = \emptyset$. With this additional restriction, the result of Theorems 1 and 2 apply for this model with Π_t defined as

$$\Pi_t := \mathbb{P}^{g_{1:t-1}^{1:n}}(X_t^*, X_t^1, \dots, X_t^n | X_{1:t}^*, \mathbf{U}_{1:t-1}).$$

Note that Π_t can be evaluated from X_t^* and $\mathbb{P}^{g_{1:t-1}^{1:n}}(X_t^1, \dots, X_t^n | X_{1:t}^*, \mathbf{U}_{1:t-1})$. It is shown in [39] that $X_t^1, X_t^2, \dots, X_t^n$ are conditionally independent given $X_{1:t}^*, \mathbf{U}_{1:t-1}$, hence the joint distribution $\mathbb{P}^{g_{1:t-1}^{1:n}}(X_t^1, \dots, X_t^n | X_{1:t}^*, \mathbf{U}_{1:t-1})$ is a product of its marginal distributions.

3) Broadcast information structure: Consider the following special case of the above generalized model.

1) The state of the system at time t is a n-dimensional vector $X_t = (X_t^1, X_t^2, \dots, X_t^n)$, where X_t^i, $i = 1, \dots, n$ corresponds to the local state of subsystem i. The first component $i = 1$ is special and called the *central node*. Other components, $i = 2, \dots, n$, are called *peripheral nodes*.
2) The state update function is such that the state of the central node evolves according to

$$X_{t+1}^1 = f_t^1(X_t^1, U_t^1, N_t^1)$$

while the state of the peripheral nodes evolves according to

$$X_{t+1}^i = f_t^i(X_t^i, X_t^1, U_t^i, U_t^1, N_t^i)$$

where $\{N_t^i, i = 1, 2, \dots n; t = 1, \dots\}$ are noise processes that are independent across time and independent of each other.
3) At time t, the common observation of all controllers is given by $Y_t^{com} = X_t^1$.

4) At time t, the local observation of controller i, $i > 2$, is given by $Y_t^i = X_t^i$. Controller 1 does not have any local observations.
5) No controller sends any additional data to the shared memory. Thus, the shared memory consists of just the history of common observations, *i.e.*, $C_t = Y_{1:t-1}^{com} = X_{1:t-1}^1$.

The above special case corresponds to the model of decentralized systems with broadcast structure considered in [19]. It is shown in [19] that there is no loss of optimality in restricting attention to controllers with no local memory, i.e., $M_t = \emptyset$. With this additional restriction, the result of Theorems 1 and 2 apply for this model with Π_t defined as

$$\Pi_t := \mathbb{P}^{g_{1:t-1}^{1:n}}(X_t^1, \dots, X_t^n | X_{1:t}^1).$$

Note that Π_t can be evaluated from X_t^1 and $\mathbb{P}^{g_{1:t-1}^{1:n}}(X_t^2, \dots, X_t^n | X_{1:t}^1)$. It is shown in [19] that $X_t^2, \dots, X_t^n$ are conditionally independent given $X_{1:t}^1$, hence the joint distribution $\mathbb{P}^{g_{1:t-1}^{1:n}}(X_t^2, \dots, X_t^n | X_{1:t}^1)$ is a product of its marginal distributions.

V. Extension to infinite horizon

In this Section, we consider the basic model of Section II-A with an infinite time horizon. Assume that

(i) The state of the system, the observations and the control actions take value in time-invariant sets $\mathcal{X}, \mathcal{Y}^i, \mathcal{U}^i$, respectively.
(ii) The local memories M_t^i and the updates to the shared memory Z_t^i take values in time-invariant sets $\mathcal{M}^i$ and $\mathcal{Z}^i$ respectively.
(iii) The dynamics of the system (equation (1)) and the observation model (equation (2)) are time-homogeneous. That is, the functions f_t and h_t in equations (1) and (2) do not vary with time.

Let the cost of using a strategy $\mathbf{g}^{1:n}$ be defined as

$$J(\mathbf{g}^{1:n}) := \mathbb{E}^{\mathbf{g}^{1:n}}\Big[\sum_{t=1}^{\infty} \beta^{t-1} l(X_t, \mathbf{U}_t)\Big], \tag{59}$$

where $\beta \in [0, 1)$ is a discount factor. We can follow the arguments of Section III to formulate the problem of the coordinated system with an infinite time horizon. As in Section III, the coordinated system is equivalent to a POMDP. The time-homogeneous nature of the coordinated

system and its equivalence to a POMDP allows us to use known POMDP results (see [43]) to conclude the following theorem for the infinite time horizon problem.

Theorem 5 *Consider Problem 1 with infinite time horizon and the objective of minimizing the expected cost given by equation* (59)*. Then, there exists an optimal time-invariant control strategy of the form:*

$$U_t^i = g^i(Y_t^i, M_t^i, \Pi_t), \quad i = 1, 2, \dots, n, \tag{60}$$

Furthermore, consider the fixed point equation,

$$V(\pi) = \inf_{\{\tilde{\gamma}^i \in F(\mathcal{Y}^i \times \mathcal{M}^i, \mathcal{U}^i), 1 \le i \le n\}} \mathbb{E}\big\{ l(X_t, \tilde{\gamma}_t^1(Y_t^1, M_t^i), \dots, \tilde{\gamma}_t^n(Y_t^n, M_t^n)) + \beta V(\eta_t(\pi, \tilde{\gamma}^1, \dots, \tilde{\gamma}^n, \mathbf{Z}_t)) \,\big|\, \Pi_t = \pi \big\}. \tag{61}$$

Then, for any realization π of Π_t, the optimal partial control laws are the choices of γ^i that achieve the infimum in the right hand side of (61)*.* □

All the special cases of our information structure considered in Sections II-B and IV-C can be extended to infinite horizon problems if the state, observation and actions spaces are time-invariant and the systems dynamics and observation equations are time homogeneous. The only exception is the control sharing information structure of section II-B4 where the local memory takes values in sets that are increasing with time.

The Case of No Shared Memory: As discussed in Section III-B, if the shared memory is always empty then the common information state defined in Theorem 2 is the *unconditional* probability $\Pi_t = \mathbb{P}^{g_{1:t-1}^{1:n}}(X_t, \mathbf{Y}_t, \mathbf{M}_t)$. In particular, Π_t is a random variable that takes a fixed (constant) value which depends only on the choice of past control laws. Therefore, for any function g_t^i of Y_t^i, M_t^i, Π_t, there exists a function $\tilde{g}_t^i$ of Y_t^i, M_t^i such that $\tilde{g}_t^i(Y_t^i, M_t^i) = g_t^i(Y_t^i, M_t^i, \Pi_t)$ with probability 1. While Theorem 5 establishes optimality of a time-invariant g_t^i, such time-invariance may not hold for the corresponding $\tilde{g}_t^i$. Similar observations were reported in [25].

VI. Discussion and Conclusions

In centralized stochastic control, the controller's belief on the current state of the system plays a fundamental role for predicting future costs. If the control strategy for the future is fixed as a function of future beliefs, then the current belief is a sufficient statistic for future

costs under any choice of current action. Hence, the optimal action at the current time is only a function of current belief on the state. In decentralized problems where different controllers have different information, using a controller's belief on the state of the system presents two main difficulties: (i) Since the costs depend both on system state as well as other controllers' actions any prediction of future costs must involve a belief on system state as well as some means of predicting other controllers' actions. (ii) Secondly, since different controllers have different information, the beliefs formed by each controller and their predictions of future costs cannot be expected to be consistent.

The approach we adopted in this paper tries to address these difficulties by using the fact that sharing of data among controllers creates common knowledge among the controllers. Beliefs based on this common knowledge are necessarily consistent among all controllers and can serve as a consistent sufficient statistic. Moreover, while controllers cannot accurately predict each other's control actions, they can know, for the observed realization of common information, the exact mapping used by each controller to map its local information to control action. These considerations suggest that common information based beliefs and partial control laws should play an important role in a general theory of decentralized stochastic control problems. The use of a fictitious coordinator allows us to make these considerations mathematically precise. Indeed, the coordinator's beliefs are based on common information and the coordinator's decision are the partial control laws. The results of the paper then follow by observing that the coordinator's problem can be viewed as a POMDP by identifying a new state that includes both the state of the dynamic system as well as the local information of the controllers.

The specific model of shared and local memory update that we assumed is crucial for connecting the coordinator's problem to POMDPs and centralized stochastic control. A key assumption in centralized stochastic control is perfect recall, that is, the information obtained at any time is remembered at all future times. This is essential for the update of the beliefs in POMDPs. Our assumption that the shared memory is increasing in time ensures that the perfect recall property is true for the coordinator's problem. If the shared memory did not have perfect recall (that is, if some past contents were lost over time), then the update of common information state in (26) would not hold and the results of Theorems 2 and 3 would not be true.

Another key factor in our result is that $S_t := \{X_t, \mathbf{Y}_t, \mathbf{M}_t\}$ serves as a state for the coordinator's problem. If the system state, observations and local memories take value in a time-invariant

space, we have a state for the coordinator's problem which takes value in a time-invariant space. Hence, the common information state is a belief on a time-invariant space. The local memory update in (7) ensures that S_t is a state. If local memory update depended on shared memory as well, that is, if (7) were replaced by

$$M^i_{t+1} \subset \{C_t, M^i_t, Y^i_t, U^i_t\},$$

then S_t would no longer suffice as a state for the coordinator. In particular, the state update equations in Lemma 1 would no longer hold. The only recourse then would be to include C_t as a part of the state which would necessarily mean that the state space keeps increasing with time. This is undesirable not only because large state spaces imply increased complexity, but the increasing size of state spaces also makes extensions of finite horizon results to infinite horizon problems conceptually difficult.

The connection between the coordinator's problem and POMDPs can be used for computational purposes as well. The dynamic program of Theorem 3 is essentially a POMDP dynamic program. In particular, just as in POMDP, the value-functions are piecewise linear and concave in π_t. This characterization of value functions is utilized to find computationally efficient algorithms for POMDPs. Such algorithmic solutions to general POMDPs are well-studied and can be employed here. We refer the reader to [44] and references therein for a review of algorithms to solve POMDPs.

While our results apply to a broad class of models, it would be worthwhile to identify special cases where the specific model features can be exploited to simplify our structural result. Examples of such simplification appear in [19], [39]. A common theme in many centralized dynamic programming solutions is to identify a key property of the value functions and use it to characterize the optimal decisions. Since our results also provide a dynamic program, an important avenue for future work would be to identify cases where properties of value functions can be analyzed to deduce a solution or to reduce the computational burden of finding the solution.

Our approach in this paper illustrates that common information provides a common conceptual framework for several decentralized stochastic control problems. In our model, we explicitly included a shared memory which naturally served the purpose of common information among the controllers. More generally, we can *define* common information for any sequential decision-

making problem and then address the problem from the perspective of a coordinator who knows the common information. Such a common information based approach for general sequential decision-making problems is presented in [40].

VII. Acknowledgments

This work was supported by NSERC through the grant NSERC-RGPIN 402753-11, by NSF through the grant CCF-1111061 and by NASA through the grant NNX09AE91G.

References

[1] 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.

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

[3] H. S. Witsenhausen, "On the structure of real-time source coders," *Bell System Technical Journal*, vol. 58, no. 6, pp. 1437–1451, July-August 1979.

[4] J. C. Walrand and P. Varaiya, "Optimal causal coding—decoding problems," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 814–820, Nov. 1983.

[5] D. Teneketzis, "On the structure of optimal real-time encoders and decoders in noisy communication," *IEEE Trans. Inf. Theory*, pp. 4017–4035, Sep. 2006.

[6] A. Nayyar and D. Teneketzis, "On the structure of real-time encoders and decoders in a multi-terminal communication system," *IEEE Trans. Info. Theory*, vol. 57, no. 9, pp. 6196–6214, September 2011.

[7] Y. Kaspi and N. Merhav, "Structure theorem for real-time variable-rate lossy source encoders and memory-limited decoders with side information," in *International Sympnosium on Information Theory*, 2010.

[8] R. R. Tenney and N. R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aerospace Electron. Systems*, vol. AES-17, no. 4, pp. 501–510, July 1981.

[9] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.

[10] D. Teneketzis and Y. C. Ho, "The Decentralized Wald problem," *Information and Computation, 73*, pp. 23–44, 1987.

[11] V. V. Veeravalli, T. Basar, and H. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, pp. 433–442, Mar. 1993.

[12] ——, "Decentralized sequential detection with sensors performing sequential tests," *Mathematics of Control, Signals and Systems*, vol. 7, no. 4, pp. 292–305, 1994.

[13] A. Nayyar and D. Teneketzis, "Sequential problems in decentralized detection with communication," *IEEE Trans. Info. Theory*, vol. 57, no. 8, pp. 5410–5435, August 2011.

[14] ——, "Decentralized detection with signaling," in *Proceeding of the Workshop on the Mathematical Theory of Networks and Systems (MTNS)*, 2010.

[15] D. Teneketzis and P. Varaiya, "The decentralized quickest detection problem," *IEEE Trans. on Automatic Control*, vol. AC-29, no. 7, pp. 641–644, July 1984.

[16] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1657–1665, 2001.

[17] J. C. Walrand and P. Varaiya, "Causal coding and control of Markov chains," *System and Control Letters*, vol. 3, pp. 189–192, 1983.

[18] A. Mahajan and D. Teneketzis, "Optimal performance of networked control systems with non-classical infromation structures," *SIAM Journal of Control and Optimization*, vol. 48, no. 3, pp. 1377–1404, May 2009.

[19] J. Wu and S. Lall, "A dynamic programming algorithm for decentralized markov decision processes with a broadcast structure," in *Proceedings of 49th IEEE Conference on Decision and Control*, Dec. 2010, pp. 6143–6148.

[20] R. Radner, "Team decision problems," *Annals of Mathmatical Statistics*, vol. 33, pp. 857–881, 1962.

[21] J. Krainak, J. Speyer, and S. Marcus, "Static team problems–Part I: Sufficient conditions and the exponential cost criterion," *IEEE Transactions on Automatic Control*, vol. 27, no. 4, pp. 839 – 848, Aug 1982.

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

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

[24] A. Mahajan and D. Teneketzis, "On the design of globally optimal communication strategies for real-time communcation systems with noisy feedback," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 4, pp. 580–595, May 2008.

[25] ——, "Optimal design of sequential real-time communication systems," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5317–5337, 2009.

[26] Y.-C. Ho and K.-C. Chu, "Team decision theory and information structures in optimal control problems–Part I," *IEEE Trans. Autom. Control*, vol. 17, no. 1, pp. 15–22, 1972.

[27] J. Kim and S. Lall, "A unifying condition for separable two player optimal control problems," in *Proceedings of 50th IEEE Conference on Decision and Control*, Dec. 2011.

[28] L. Lessard and S. Lall, "A state-space solution to the two-player decentralized optimal control problem," in *Proceedings of 49th Annual Allerton Conference on Communication, Control and Computing*, 2011.

[29] A. Rantzer, "Linear quadratic team theory revisited," in *Proceedings of the American control conference*, 2006.

[30] A. Gattami, "Control and estimation problems under partially nested information pattern," in *Proceedings of 48th IEEE Conference on Decision and Control*, Dec. 2009, pp. 5415–5419.

[31] S. Yuksel, "Stochastic nestedness and the belief sharing information pattern," *IEEE Trans. on Automatic Control*, vol. 54, no. 12, pp. 2773–2786, December 2009.

[32] B. Bamieh and P. G. Voulgaris, "A convexcharacterization of distributed control problems in spatially invariant systems with communication constraints," *System and Control Letters*, pp. 575–583, 2005.

[33] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control." *IEEE Trans. on Automatic Control*, vol. 51, no. 2, pp. 274–286, February 2006.

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

[35] M. Aicardi, F. Davoli, and R. Minciardi, "Decentralized optimal control of markov chains with a common past information set," *IEEE Transactions on Automatic Control*, vol. 32, no. 11, Nov. 1987.

[36] A. Nayyar, A. Mahajan, and D. Teneketzis, "Optimal control strategies in delayed sharing information structures," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1606–1620, July 2011.

[37] 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.

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

[39] A. Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," in *Proceedings of 50th IEEE Conference on Decision and Control*, Dec. 2011.

[40] A. Nayyar, "Sequential decision-making in decentralized systems," Ph.D. dissertation, University of Michigan, 2011.

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

[42] N. R. Sandell, Jr., "Control of finite-state, finite-memory stochastic systems," Ph.D. Thesis, Massachussets Institute of Technology, Cambridge, MA, 1974.

[43] P. Whittle, *Optimization Over Time*, ser. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, 1983, vol. 2.

[44] H. Zhang, "Partially observable Markov decision processes: A geometric technique and analysis," *Operations Research*, 2009.

APPENDIX A
THE UPDATE FUNCTION η_t OF THE COORDINATOR'S INFORMATION STATE

Consider a realization c_{t+1} of the shared memory C_{t+1} at time $t+1$. Let $(\gamma_{1:t})$ be the corresponding realization of the coordinator's prescriptions until time t. We assume the realization $(c_{t+1}, \pi_{1:t}, \gamma_{1:t})$ to be of non-zero probability. Then, the realization π_{t+1} of Π_{t+1} is given by

$$\pi_{t+1}(s) = \mathbb{P}\{S_{t+1} = s | c_{t+1}, \gamma_{1:t}\}. \tag{62}$$

Use Lemma 1 to simplify the above expression as

$$\sum_{s_t, w_t^0, \mathbf{w}_{t+1}} \mathbb{1}_s(\tilde{f}_t(s_t, \gamma_t, w_t^0, \mathbf{w}_{t+1})) \cdot \mathbb{P}\{W_t^0 = w_t^0\} \cdot \mathbb{P}\{\mathbf{W}_{t+1} = \mathbf{w}_{t+1}\} \cdot \mathbb{P}\{S_t = s_t | c_{t+1}, \gamma_{1:t}\}. \tag{63}$$

Since $c_{t+1} = (c_t, \mathbf{z}_t)$, write the last term of (63) as

$$\mathbb{P}\{S_t = s_t | c_t, \mathbf{z}_t, \gamma_{1:t}\} = \frac{\mathbb{P}\{S_t = s_t, \mathbf{Z}_t = \mathbf{z}_t | c_t, \gamma_{1:t}\}}{\sum_{s'} \mathbb{P}\{S_t = s', \mathbf{Z}_t = \mathbf{z}_t | c_t, \gamma_{1:t}\}}. \tag{64}$$

Use Lemma 1 and the sequential order in which the system variables are generated to write the numerator as

$$\mathbb{P}\{S_t = s_t, \mathbf{Z}_t = \mathbf{z}_t | c_t, \gamma_{1:t}\} = \mathbb{1}_{\tilde{h}_t(s_t, \gamma_t)}(\mathbf{z}_t) \cdot \mathbb{P}\{S_t = s_t | c_t, \gamma_{1:t}\} \tag{65}$$

$$= \mathbb{1}_{\tilde{h}_t(s_t, \gamma_t)}(\mathbf{z}_t) \cdot \pi_t(s_t). \tag{66}$$

where we dropped γ_t from conditioning in (65) since under the given coordinator's strategy, it is a function of the rest of the terms in the conditioning. Substitute (66), (64), and (63) into (62), to get

$$\pi_{t+1}(s) = \eta_t^s(\pi_t, \gamma_t, z_t),$$

where $\eta_t^s(\cdot)$ is given by (62), (63), (64), and (66). $\eta_t(\cdot)$ is the vector $(\eta_t^s(\cdot))_{s\in\mathcal{S}}$.

APPENDIX B
PROOF OF PROPOSITION 3

(a) For any given control strategy $\mathbf{g}^{1:n}$ in the basic model, define a coordinated strategy $\mathbf{d}$ for the coordinated system as

$$d_t(C_t) = \big(g_t^1(\cdot,\cdot,C_t), \dots, g^n(\cdot,\cdot,C_t)\big). \tag{67}$$

Consider Problems 1 and 2. Use control strategy $\mathbf{g}^{1:n}$ in Problem 1 and coordination strategy $\mathbf{d}$ given by (67) in Problem 2. Fix a specific realization of the primitive random variables $\{X_1, W_t^j, t = 1,\dots,T, j = 0,1,\dots,n\}$ in the two problems. Equation (2) implies that the realization of $\mathbf{Y}_1$ will be the same in the two problems. Then, the choice of $\mathbf{d}$ according to (67) implies that the realization of the control actions $\mathbf{U}_1$ will be the same in the two problems. This implies that the realization of the next state X_2 and the memories $\mathbf{M}_2$, C_2 will be the same in the two problems. Proceeding in a similar manner, it is clear that the choice of $\mathbf{d}$ according to (67) implies that the realization of the state $\{X_t; t = 1,\dots,T\}$, the observations $\{\mathbf{Y}_t; t = 1,\dots,T\}$, the control actions $\{\mathbf{U}_t; t = 1,\dots,T\}$ and the memories $\{\mathbf{M}_t; t = 1,\dots,T\}$ and $\{C_t; t = 1,\dots,T\}$ are all identical in Problem 1 and 2. Thus, the total expected cost under $\mathbf{g}^{1:n}$ in Problem 1 is same as the total expected cost under the coordination strategy given by (67) in Problem 2. That is, $J(\mathbf{g}^{1:n}) = \hat{J}(\mathbf{d})$.

(b) The second part of Proposition 3 follows from similar arguments as above.

APPENDIX C
EQUIVALENCE BETWEEN THE MODEL OF THIS PAPER AND THE MODEL OF [1]

We refer to the model of this paper as the PHS (partial history sharing) model and the model of [1] as the CO (common observation) model. First, we describe the CO model and then show

the both models are equivalent by showing that the PHS model is a special case of CO model and vice versa.

The CO Model

The following model was presented in [1]; we use a slightly different notation so that the notation matches with that of our paper.

Consider a system with n controllers. Let X_t denote the state of the system, Z_t denote the common observation of all controllers, Y_t^i denote the private observation of controller i, M_t^i the contents of the memory of controller i, and U_t^i the control action of controller i, $i = 1, \dots, n$.

The system dynamics and observation equations are given by

$$X_{t+1} = f_t(X_t, U_t^{1:n}, W_t^0), \tag{68}$$

$$Y_t^i = h_t^i(X_t, U_t^{1:i-1}, W_t^i), \quad i = 1, \dots, n, \tag{69}$$

$$Z_t = c_t(X_t, U_{t-1}^{1:n}, Q_t), \tag{70}$$

where $\{X_1, Q_t, W_t^i, i = 0, \dots, n, t = 1, \dots, T\}$ are independent random variables.

At time t, controller i generates a control action and updates its memory as follows:

$$U_t^i = g_t^i(Z_{1:t}, Y_t^i, M_{t-1}^i), \tag{71}$$

$$M_t^i = r_t^i(Z_{1:t}, Y_t^i, M_{t-1}^i). \tag{72}$$

At each time an instantaneous cost $l_t(X_t, U_t^{1:n})$ is incurred. The system objective is to choose a control strategy $g_{1:T}^{1:n}$ and a memory update strategy $r_{1:T}^{1:n}$ to minimize a total expected cost.

The PHS model is a special case of CO model

Consider the PHS model described in Sec II-A of the paper and define

$$\tilde{X}_t = (X_t, Y_t^{1:n}, M_t^{1:n}, Z_{t-1}^{1:n}),$$

$$\tilde{U}_t^i = U_t^i, \quad i = 1, \dots, n$$

$$\tilde{Y}_t^i = (Y_t^i, M_t^i), \quad i = 1, \dots, n$$

$$\tilde{Z}_t = Z_{t-1}^{1:n},$$

$$\tilde{M}_t^i = \emptyset, \quad i = 1, \dots, n.$$

Define the cost function

$$\tilde{l}_t(\tilde{X}_t, \tilde{U}_t^{1:n}) = l_t(X_t, U_t^{1:n}).$$

It is easy to verify that the model $(\tilde{X}_t, \tilde{U}_t^{1:n}, \tilde{Y}_t^{1:n}, \tilde{M}_t^{1:n}, \tilde{Z}_t)$ defined above is a special case of CO model.

The CO model is a special case of PHS model

In the CO model, the local observations Y_t^i of controller i depends on the control action $U_t^{1:i-1}$. This feature is not present in PHS model. Nonetheless, we can show that CO model is a special case of the PHS model by splitting time and assuming that in the PHS model only one controller acts at each time.

Define the following system variables for $\tau = 1, \dots, nT$. For ease of notation, when $tn < \tau \le (t+1)n$, we will write τ as $tn + i$. Thus, the system variables are defined for $t = 1, \dots, T$ and $i = 1, \dots, n$:

$$\tilde{X}_{tn+1} = (X_t, U_{t-1}^{1:n}, M_{t-1}^{1:n}), \qquad \tilde{X}_{tn+i} = (X_t, U_t^{i:i-1}, M_t^{1:i-1}, M_{t-1}^{i:n}), \quad i = 2, \dots, n,$$
$$\tilde{Z}_{tn+1} = Z_t, \qquad \tilde{Z}_{tn+i} = \emptyset, \quad i = 2, \dots, n,$$

$$\tilde{Y}_{tn+j}^i = \begin{cases} (Y_t^1, M_{t-1}^1, Z_t), & \text{if } i = j = 1, \\ (Y_t^i, M_{t-1}^i) & \text{if } i = j \neq 1, \\ \emptyset, & \text{otherwise;} \end{cases} \quad i, j = 1, \dots, n$$

$$\tilde{U}_{tn+j}^i = \begin{cases} (U_t^i, M_t^i), & \text{if } i = j, \\ \emptyset, & \text{otherwise;} \end{cases} \quad i, j = 1, \dots, n$$

$$\tilde{M}_{tn+j}^i = \emptyset, \quad j = 1, \dots, n.$$

Define the cost function as:

$$\tilde{l}_{tn+i}(\tilde{X}_{tn+i}, \tilde{U}_{tn+i}^{1:n}) = \begin{cases} l_t(X_t, U_t^{1:n}), & \text{if } i = n, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to verify that the model $(\tilde{X}_\tau, \tilde{U}_\tau^{1:n}, \tilde{Y}_\tau^{1:n}, \tilde{M}_\tau^{1:n}, \tilde{Z}_\tau)$ defined above is a special case of PHS model.