Optimal Control Strategies in Delayed Sharing Information Structures

Ashutosh Nayyar *Student Member, IEEE*, Aditya Mahajan *Member, IEEE*, and Demosthenis Teneketzis *Fellow, IEEE*

***Abstract*—The n-step delayed sharing information structure is investigated. This information structure comprises of K controllers that share their information with a delay of n time steps. This information structure is a link between the classical information structure, where information is shared perfectly between the controllers, and a non-classical information structure, where there is no "lateral" sharing of information among the controllers. Structural results for optimal control strategies for systems with such information structures are presented. A sequential methodology for finding the optimal strategies is also derived. The solution approach provides an insight for identifying structural results and sequential decomposition for general decentralized stochastic control problems.**

***Index Terms*—Decentralized control, Non-classical information structures, Team theory, Markov decision theory, Stochastic control**

I. Introduction

A. Motivation

One of the difficulties in optimal design of decentralized control systems is handling the increase of data at the control stations with time. This increase in data means that the domain of control laws increases with time which, in turn, creates two difficulties. Firstly, the number of control strategies increases doubly exponentially with time; this makes it harder to search for an optimal strategy. Secondly, even if an optimal strategy is found, implementing functions with time increasing domain is difficult.

In centralized stochastic control [1], these difficulties can be circumvented by using the conditional probability of the state given the data available at the control station as a sufficient statistic (where the data available to a control station comprises of all observations and control actions till the current time). This conditional probability, called *information state*, takes values in a time-invariant space. Consequently, we can restrict attention to control laws with time-invariant domain. Such results, where data that is increasing with time is "compressed" to a sufficient statistic taking values in a time-invariant space, are called *structural results*. While the information state and structural result for centralized stochastic control problems is well known, no general methodology to find such information states or structural results exists for decentralized stochastic control problems.

Ashutosh Nayyar and Demosthenis Teneketzis are with the department of EECS, University of Michigan, Ann Arbor, MI – 48109, USA. (email: {anayyar,teneket}@umich.edu). Their work was supported in part by NSF Grant CCR-0325571 and NASA Grant NNX06AD47G.

Aditya Mahajan is with the department of ECE, McGill University, Montreal, QC – H3A-2A7, Canada. (email: adityam@ieee.org)

The structural results in centralized stochastic control are related to the concept of separation. In centralized stochastic control, the information state, which is conditional probability of the state given all the available data, does not depend on the control strategy (which is the collection of control laws used at different time instants). This has been called a one-way separation between estimation and control. An important consequence of this separation is that for any given choice of control laws till time $t-1$ and a given realization of the system variables till time t, the information states at future times do not depend on the choice of the control law at time t but only on the realization of control action at time t. Thus, the future information states are *separated* from the choice of the current control law. This fact is crucial for the formulation of the classical dynamic program where at each step the optimization problem is to find the best control action for a given realization of the information state. No analogous separation results are known for general decentralized systems.

In this paper, we find structural results for decentralized control systems with delayed sharing information structures. In a system with n-step delayed sharing, every control station knows the n-step prior observations and control actions of all other control stations. This information structure, proposed by Witsenhausen in [2], is a link between the classical information structures, where information is shared perfectly among the controllers, and the non-classical information structures, where there is no "lateral" sharing of information among the controllers. Witsenhausen asserted a structural result for this model without any proof in his seminal paper [2]. Varaiya and Walrand [3] proved that Witsenhausen's assertion was true for $n = 1$ but false for $n > 1$. For $n > 1$, Kurtaran [4] proposed another structural result. However, Kurtaran proved his result only for the terminal time step (that is, the last time step in a finite horizon problem); for non-terminal time steps, he gave an abbreviated argument, which we believe is incomplete. (The details are given in Section V of the paper).

We prove two structural results of the optimal control laws for the delayed sharing information structure. We compare our results to those conjectured by Witsenhausen and show that our structural results for n-step delay sharing information structure simplify to that of Witsenhausen for $n = 1$; for $n > 1$, our results are different from the result proposed by Kurtaran.

We note that our structural results do not have the separated nature of centralized stochastic control. That is, for any given realization of the system variables till time t, the realization of information states at future times depend on the choice of the control law at time t. However, our second structural result shows that this dependence only propagates to the next $n-1$

time steps. Thus, the information states from time $t+n-1$ onwards are separated from the choice of control laws before time t. We call this a *delayed* separation between information states and control laws.

The absence of classical separation rules out the possibility of a classical dynamic program to find the optimum control laws. However, optimal control laws can still be found in a sequential manner. Based on the two structural results, we present two sequential methodologies to find optimal control laws. Unlike classical dynamic programs, each step in our sequential decomposition involves optimization over a space of functions instead of the space of control actions.

B. Notation

Random variables are denoted by upper case letters; their realization by the corresponding lower case letter. For some function valued random variables (specifically, $\gamma_t^k, r_{m,t}^k$ in the paper), a tilde ($\tilde{}$) denotes their realization (for example, $\tilde{\gamma}_t^k$). $X_{a:b}$ 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 $(X^c, X^{c+1}, \dots, X^d)$. The combined notation $X_{a:b}^{c:d}$ is a short hand for the vector $(X_i^j : i = a, a+1, \dots, b, j = c, c+1, \dots, d)$. $\mathbb{P}(\cdot)$ 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 and 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 finite set A, $\mathcal{P}\{A\}$ denotes the space of probability mass functions on A.

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

C. Model

Consider a system consisting of a plant and K controllers (control stations) with decentralized information. At time t, $t = 1, \dots, T$, the state of the plant X_t takes values in a finite set $\mathcal{X}$; the control action U_t^k at control station k, $k = 1, \dots, K$, takes values in a finite set $\mathcal{U}^k$. The initial state X_0 of the plant is a random variable taking value in $\mathcal{X}$. With time, the plant evolves according to

$$X_t = f_t(X_{t-1}, U_t^{1:K}, V_t) \tag{1}$$

where V_t is a random variable taking values in a finite set $\mathcal{V}$. $\{V_t; t = 1, \dots, T\}$ is a sequence of independent random variables that are also independent of X_0.

The system has K observation posts. At time t, $t = 1, \dots, T$, the observation Y_t^k of post k, $k = 1, \dots, K$, takes values in a finite set $\mathcal{Y}^k$. These observations are generated according to

$$Y_t^k = h_t^k(X_{t-1}, W_t^k) \tag{2}$$

TABLE I
SUMMARY OF THE CONTROL LAWS IN THE MODEL FOR $K = 2$.

	Controller 1	Controller 2
Observations (actual)	$\begin{pmatrix} Y_{1:t}^1 & Y_{1:t-n}^2 \\ U_{1:t-1}^1 & U_{1:t-n}^2 \end{pmatrix}$	$\begin{pmatrix} Y_{1:t-n}^1 & Y_{1:t}^2 \\ U_{1:t-n}^1 & U_{1:t-1}^2 \end{pmatrix}$
Observations (shorthand)	(Δ_t, Λ_t^1)	(Δ_t, Λ_t^2)
Control action	U_t^1	U_t^2
Control laws	g_t^1	g_t^2

where W_t^k are random variables taking values in a finite set $\mathcal{W}^k$. $\{W_t^k; t = 1, \dots, T; k = 1, \dots, K\}$ are independent random variables that are also independent of X_0 and $\{V_t; t = 1, \dots, T\}$.

The system has n-step delayed sharing. This means that at time t, control station k observes the current observation Y_t^k of observation post k, the n steps old observations $Y_{t-n}^{1:K}$ of all posts, and the n steps old actions $U_{t-n}^{1:K}$ of all stations. Each station has perfect recall; so, it remembers everything that it has seen and done in the past. Thus, at time t, data available at station k can be written as (Δ_t, Λ_t^k), where

$$\Delta_t \coloneqq (Y_{1:t-n}^{1:K}, U_{1:t-n}^{1:K})$$

is the data known to all stations and

$$\Lambda_t^k \coloneqq (Y_{t-n+1:t}^k, U_{t-n+1:t-1}^k)$$

is the additional data known at station k, $k = 1, \dots, K$. Let $\mathcal{D}_t$ be the space of all possible realizations of Δ_t; and $\mathcal{L}^k$ be the space of all possible realizations of Λ_t^k. Station k chooses action U_t^k according to a control law g_t^k, i.e.,

$$U_t^k = g_t^k(\Lambda_t^k, \Delta_t). \tag{3}$$

The choice of $\boldsymbol{g} = \{g_t^k; k = 1, \dots, K; t = 1, \dots, T\}$ is called a *design* or a *control strategy*. $\mathcal{G}$ denotes the class of all possible designs. At time t, a cost $c_t(X_t, U_t^1, \dots, U_t^K)$ is incurred. The performance $\mathcal{J}(\boldsymbol{g})$ of a design is given by the expected total cost under it, i.e.,

$$\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}\left\{\sum_{t=1}^{T} c_t(X_t, U_t^{1:K})\right\} \tag{4}$$

where the expectation is with respect to the joint measure on all the system variables induced by the choice of $\boldsymbol{g}$. For reference, we summarize the notation of this model in Table I. We consider the following problem.

Problem 1: Given the statistics of the primitive random variables X_0, $\{V_t; t = 1, \dots, T\}$, $\{W_t^k; k = 1, \dots, K; t = 1, \dots, T\}$, the plant functions $\{f_t; t = 1, \dots, T\}$, the observation functions $\{h_t^k; k = 1, \dots, K; t = 1, \dots, T\}$, and the cost functions $\{c_t; t = 1, \dots, T\}$ choose a design $\boldsymbol{g}^*$ from $\mathcal{G}$ that minimizes the expected cost given by (4).

Remarks on the Model:

1) We assumed that all primitive random variables and all control actions take values in finite sets for convenience of exposition. Similar results can be obtained with uncountable sets under suitable technical condtions.

2) In the standard stochastic control literature, the dynamics and observations equations are defined in a different manner than (1) and (2). The usual model is

$$X_{t+1} = f_t(X_t, U_t^{1:K}, V_t) \tag{5}$$
$$Y_t^k = h_t^k(X_t, W_t^k) \tag{6}$$

However, Witsenhausen [2] as well as Varaiya and Walrand [3] used the model of (1) and (2) in their papers. We use the same model so that our results can be directly compared with earlier conjectures and results. The arguments of this paper can be used for the dynamics and observation model of (5) and (6) with minor changes.

D. The structural results

Witsenhausen [2] asserted the following structural result for Problem 1.

Structural Result (Witsenhausen [2]): In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k(\Lambda_t^k, \mathbb{P}(X_{t-n} \,|\, \Delta_t)). \tag{7}$$

Witsenhausen's result claims that all control stations can "compress" the common information Δ_t to a sufficient statistic $\mathbb{P}(X_{t-n} \,|\, \Delta_t)$. Unlike Δ_t, the size of $\mathbb{P}(X_{t-n} \,|\, \Delta_t)$ does not increase with time.

As mentioned earlier, Witsenhausen asserted this result without a proof. Varaiya and Walrand [3] proved that the above separation result is true for $n = 1$ but false for $n > 1$. Kurtaran [4] proposed an alternate structural result for $n > 1$.

Structural Result (Kurtaran [4]): In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k\big(Y_{t-n+1:t}^k, \mathbb{P}^{g_{1:t-1}^{1:K}}\big(X_{t-n}, U_{t-n+1:t-1}^{1:K} \,\big|\, \Delta_t\big)\big). \tag{8}$$

Kurtaran used a different labeling of the time indices, so the statement of the result in his paper is slightly different from what we have stated above.

Kurtaran's result claims that all control stations can "compress" the common information Δ_t to a sufficient statistic $\mathbb{P}^{g_{1:t-1}^{1:K}}\big(X_{t-n}, U_{t-n+1:t-1}^{1:K} \,\big|\, \Delta_t\big)$, whose size does not increase with time.

Kurtaran proved his result for only the terminal time-step and gave an abbreviated argument for non-terminal time-steps. We believe that his proof is incomplete for reasons that we point out in Section V. In this paper, we prove two alternative structural results.

First Structural Result (this paper): In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k\big(\Lambda_t^k, \mathbb{P}^{g_{1:t-1}^{1:K}}\big(X_{t-1}, \Lambda_t^{1:K} \,\big|\, \Delta_t\big)\big). \tag{9}$$

This result claims that all control stations can "compress" the common information Δ_t to a sufficient statistic $\mathbb{P}^{g_{1:t-1}^{1:K}}\big(X_{t-1}, \Lambda_t^{1:K} \,\big|\, \Delta_t\big)$, whose size does not increase with time.

Second Structural Result (this paper): In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k\big(\Lambda_t^k, \mathbb{P}(X_{t-n} \,|\, \Delta_t), r_t^{1:K}\big). \tag{10}$$

where $r_t^{1:K}$ is a collection of partial functions of the previous $n-1$ control laws of each controller,

$$r_t^k := \{(g_m^k(\cdot, Y_{m-n+1:t-n}^k, U_{m-n+1:t-n}^k, \Delta_m), \quad m = t-n+1, t-n+2, \dots, t-1\},$$

for $k = 1, 2, \dots, K$. Observe that r_t^k depends only on the previous $n-1$ control laws ($g_{t-n+1:t-1}^k$) and the realization of Δ_t (which consists of $Y_{1:t-n}^{1:K}, U_{1:t-n}^{1:K}$). This result claims that the belief $\mathbb{P}(X_{t-n} \,|\, \Delta_t)$ and the realization of the partial functions $r_t^{1:K}$ form a sufficient representation of Δ_t in order to optimally select the control action at time t.

Our structural results cannot be derived from Kurtaran's result and vice-versa. At present, we are not sure of the correctness of Kurtaran's result. As we mentioned before, we believe that the proof given by Kurtaran is incomplete. We have not been able to complete Kurtaran's proof; neither have we been able to find a counterexample to his result.

Kurtaran's and our structural results differ from those asserted by Witsenhausen in a fundamental way. The sufficient statistic (also called information state) $\mathbb{P}(X_{t-n} \,|\, \Delta_t)$ of Witsenhausen's assertion does not depend on the control strategy. That is, for any realization δ_t of Δ_t, the knowledge of control laws is not required in evaluating the conditional probabilities $\mathbb{P}(X_{t-n} = x \,|\, \delta_t)$. The sufficient statistics $\mathbb{P}^{g_{1:t-1}^{1:K}}\big(X_{t-n}, U_{t-n+1:t-1}^{1:K} \,\big|\, \Delta_t\big)$ of Kurtaran's result and $\mathbb{P}^{g_{1:t-1}^{1:K}}\big(X_{t-1}, \Lambda_t^{1:K} \,\big|\, \Delta_t\big)$ of our first result *depend on the control laws used before time* t. Thus, for a given realization δ_t of Δ_t, the realization of information state depends on the choice of control laws before time t. On the other hand, in our second structural result, the belief $\mathbb{P}(X_{t-n} \,|\, \Delta_t)$ is indeed independent of the control strategy, however information about the previous $n-1$ control laws is still needed in the form of the partial functions $r_t^{1:K}$. Since the partial functions $r_t^{1:K}$ do not depend on control laws used before time $t-n+1$, we conclude that the information state at time t is separated from the choice of control laws before time $t-n+1$. We call this a delayed separation between information states and control laws.

The rest of this paper is organized as follows. We prove our first structural result in Section II. Then, in Section III we derive our second structural result. We discuss a special case of delayed sharing information structures in Section IV. We discuss Kurtaran's structural result in Section V and conclude in Section VI.

II. Proof of the first structural result

In this section, we prove the structural result (9) for optimal strategies of the K control stations. For the ease of notation, we first prove the result for $K = 2$, and then show how to extend it for general K.

A. Two Controller system ($K = 2$)

The proof for $K = 2$ proceeds as follows:

1) First, we formulate a centralized stochastic control problem from the point of view of a coordinator who observes the shared information Δ_t, but does not observe the private information $(\Lambda_t^1, \Lambda_t^2)$ of the two controllers.
2) Next, we argue that any strategy for the coordinator's problem can be implemented in the original problem and vice versa. Hence, the two problems are equivalent.
3) Then, we identify states sufficient for input-output mapping for the coordinator's problem.
4) Finally, we transform the coordinator's problem into a MDP (Markov decision process), and obtain a structural result for the coordinator's problem. This structural result is also a structural result for the delayed sharing information strucutres due to the equivalence between the two problems.

Below, we elaborate on each of these stages.

Stage 1

We consider the following modified problem. In the model described in Section I-C, in addition to the two controllers, a coordinator that knows the common (shared) information Δ_t available to both controllers at time t is present. At time t, the coordinator decides the *partial functions*

$$\gamma_t^k : \mathcal{L}^k \mapsto \mathcal{U}^k$$

for each controller k, $k = 1, 2$. The choice of the partial functions at time t is based on the realization of the common (shared) information and the partial functions selected before time t. These functions map each controller's *private information* Λ_t^k to its control action U_t^k at time t. The coordinator then informs all controllers of all the partial functions it selected at time t. Each controller then uses its assigned partial function to generate a control action as follows.

$$U_t^k = \gamma_t^k(\Lambda_t^k). \tag{11}$$

The system dynamics and the cost are same as in the original problem. At next time step, the coordinator observes the new common observation

$$Z_{t+1} \coloneqq \{Y_{t-n+1}^1, Y_{t-n+1}^2, U_{t-n+1}^1, U_{t-n+1}^2\}. \tag{12}$$

Thus at the next time, the coordinator knows $\Delta_{t+1} = (Z_{t+1}, \Delta_t)$ and its choice of all past partial functions and it selects the next partial functions for each controller. The system proceeds sequentially in this manner until time horizon T.

In the above formulation, the only decision maker is the coordinator: the individual controllers simply carry out the necessary evaluations prescribed by (11). At time t, the coordinator knows the common (shared) information Δ_t and all past partial functions $\gamma_{1:t-1}^1$ and $\gamma_{1:t-1}^2$. The coordinator uses a decision rule ψ_t to map this information to its decision, that is,

$$(\gamma_t^1, \gamma_t^2) = \psi_t(\Delta_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2), \tag{13}$$

TABLE II
SUMMARY OF THE MODEL WITH A COORDINATOR.

	Coordinator	Controller k (passive)
Observations (actual)	$(Y_{1:t-n}^{1:K}, U_{1:t-n}^{1:K})$	$(Y_{t-n+1:t}^k, U_{t-n+1:t-1}^k)$
Observations (shorthand)	Δ_t	Λ_t^k
Control action	$\gamma_t^{1:K}$	U_t^k
Control laws	ψ_t	γ_t^k

or equivalently,

$$\gamma_t^k = \psi_t^k(\Delta_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2), \quad k = 1, 2. \tag{14}$$

For reference, we summarize the notation of this model in Table II.

The choice of $\boldsymbol{\psi} = \{\psi_t; t = 1, \dots, T\}$ is called a *coordination strategy*. Ψ denotes the class of all possible coordination strategies. The performance of a coordinating strategy is given by the expected total cost under that strategy, that is,

$$\hat{\mathcal{J}}(\boldsymbol{\psi}) = \mathbb{E}^{\boldsymbol{\psi}} \left\{ \sum_{t=1}^{T} c_t(X_t, U_t^1, U_t^2) \right\} \tag{15}$$

where the expectation is with respect to the joint measure on all the system variables induced by the choice of $\boldsymbol{\psi}$. The coordinator has to solve the following optimization problem.

Problem 2 (The Coordinator's Optimization Problem): Given the system model of Problem 1, choose a coordination strategy $\boldsymbol{\psi}^*$ from Ψ that minimizes the expected cost given by (15).

Stage 2

We now show that the Problem 2 is equivalent to Problem 1. Specifically, we will show that any design $\boldsymbol{g}$ for Problem 1 can be implemented by the coordinator in Problem 2 with the same value of the problem objective. Conversely, any coordination strategy $\boldsymbol{\psi}$ in Problem 2 can be implemented in Problem 1 with the same value of the performance objective.

Any design $\boldsymbol{g}$ for Problem 1 can be implemented by the coordinator in Problem 2 as follows. At time t the coordinator selects partial functions (γ_t^1, γ_t^2) using the common (shared) information δ_t as follows.

$$\gamma_t^k(\cdot) = g_t^k(\cdot, \delta_t) =: \psi_t^k(\delta_t), \quad k = 1, 2. \tag{16}$$

Consider Problems 1 and 2. Use design $\boldsymbol{g}$ in Problem 1 and coordination strategy $\boldsymbol{\psi}$ given by (16) in Problem 2. Fix a specific realization of the initial state X_0, the plant disturbance $\{V_t; t = 1, \dots, T\}$, and the observation noise $\{W_t^1, W_t^2; t = 1, \dots, T\}$. Then, the choice of $\boldsymbol{\psi}$ according to (16) implies that the realization of the state $\{X_t; t = 1, \dots, T\}$, the observations $\{Y_t^1, Y_t^2; t = 1, \dots, T\}$, and the control actions $\{U_t^1, U_t^2; t = 1, \dots, T\}$ are identical in Problem 1 and 2. Thus, any design $\boldsymbol{g}$ for Problem 1 can be implemented by the coordinator in Problem 2 by using a coordination strategy given by (16) and the total expected cost under $\boldsymbol{g}$ in Problem 1 is same as the total expected cost under the coordination strategy given by (16) in Problem 2.

By a similar argument, any coordination strategy $\boldsymbol{\psi}$ for Problem 2 can be implemented by the control stations in Problem 1 as follows. At time 1, both stations know δ_1; so, all of them can compute $\gamma_1^1 = \psi_1^1(\delta_1)$, $\gamma_1^2 = \psi_1^2(\delta_1)$. Then station k chooses action $u_1^k = \gamma_1^k(\lambda_1^k)$. Thus,

$$g_1^k(\lambda_1^k, \delta_1) = \psi_1^k(\delta_1)(\lambda_1^k), \quad k = 1, 2. \tag{17a}$$

At time 2, both stations know δ_2 and γ_1^1, γ_1^2, so both of them can compute $\gamma_2^k = \psi_2^k(\delta_2, \gamma_1^1, \gamma_1^2)$, $k = 1, 2$. Then station k chooses action $u_2^k = \gamma_2^k(\lambda_2^k)$. Thus,

$$g_2^k(\lambda_2^k, \delta_2) = \psi_2^k(\delta_2, \gamma_1^1, \gamma_1^2)(\lambda_2^k), \quad k = 1, 2. \tag{17b}$$

Proceeding this way, at time t both stations know δ_t and $\gamma_{1:t-1}^1$ and $\gamma_{1:t-1}^2$, so both of them can compute $(\gamma_{1:t}^1, \gamma_{1:t}^2) = \psi_t(\delta_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)$. Then, station k chooses action $u_t^k = \gamma_t^k(\lambda_t^k)$. Thus,

$$g_t^k(\lambda_t^k, \delta_t) = \psi_t^k(\delta_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)(\lambda_t^k), \quad k = 1, 2. \tag{17c}$$

Now consider Problems 2 and 1. Use coordinator strategy $\boldsymbol{\psi}$ in Problem 2 and design $\boldsymbol{g}$ given by (17) in Problem 1. Fix a specific realization of the initial state X_0, the plant disturbance $\{V_t; t = 1, \dots, T\}$, and the observation noise $\{W_t^1, W_t^2; t = 1, \dots, T\}$. Then, the choice of $\boldsymbol{g}$ according to (17) implies that the realization of the state $\{X_t; t = 1, \dots, T\}$, the observations $\{Y_t^1, Y_t^2; t = 1, \dots, T\}$, and the control actions $\{U_t^1, U_t^2; t = 1, \dots, T\}$ are identical in Problem 2 and 1. Hence, any coordination strategy $\boldsymbol{\psi}$ for Problem 2 can be implemented by the stations in Problem 1 by using a design given by (17) and the total expected cost under $\boldsymbol{\psi}$ in Problem 2 is same as the total expected cost under the design given by (17) in Problem 1.

Since Problems 1 and 2 are equivalent, we derive structural results for the latter problem. Unlike, Problem 1, where we have multiple control stations, the coordinator is the only decision maker in Problem 2.

Stage 3

We now look at Problem 2 as a controlled input-output system from the point of view of the coordinator and identify a state sufficient for input-output mapping. From the coordinator's viewpoint, the input at time t has two components: a stochastic input that consists of the plant disturbance V_t and observation noises W_t^1, W_t^2; and a controlled input that consists of the partial functions γ_t^1, γ_t^2. The output is the observations Z_{t+1} given by (12). The cost is given by $c_t(X_t, U_t^1, U_t^2)$. We want to identify a state sufficient for input-output mapping for this system.

A variable is a state sufficient for input output mapping of a control system if it satisfies the following properties (see [5]).

P1) The next state is a function of the current state and the current inputs.
P2) The current output is function of the current state and the current inputs.
P3) The instantaneous cost is a function of the current state, the current control inputs, and the next state.

We claim that such a state for Problem 2 is the following.

Definition 1: For each t define

$$S_t := (X_{t-1}, \Lambda_t^1, \Lambda_t^2) \tag{18}$$

Next we show that S_t, $t = 1, 2, \dots, T+1$, satisfy properties (P1)–(P3). Specifically, we have the following.

Proposition 1:

1) There exist functions $\hat{f}_t$, $t = 2, \dots, T$ such that

$$S_{t+1} = \hat{f}_{t+1}(S_t, V_t, W_{t+1}^1, W_{t+1}^2, \gamma_t^1, \gamma_t^2). \tag{19}$$

2) There exist functions $\hat{h}_t$, $t = 2, \dots, T$ such that

$$Z_t = \hat{h}_t(S_{t-1}). \tag{20}$$

3) There exist functions $\hat{c}_t$, $t = 1, \dots, T$ such that

$$c_t(X_t, U_t^1, U_t^2) = \hat{c}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1}). \tag{21}$$

Proof: Part 1 is an immediate consequence of the definitions of S_t and Λ_t^k, the dynamics of the system given by (1), and the evaluations carried out by the control stations according to (11). Part 2 is an immediate consequence of the definitions of state S_t, observation Z_t, and private information Λ_t^k. Part 3 is an immediate consequence of the definition of state and the evaluations carried out by the control stations according to (11). ■

Stage 4

Proposition 1 establishes S_t as the state sufficient for input-output mapping for the coordinator's problem. We now define information states for the coordinator.

Definition 2 (Information States): For a coordination strategy $\boldsymbol{\psi}$, define *information states* Π_t as

$$\Pi_t(s_t) := \mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s_t \,\middle|\, \Delta_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2\right). \tag{22}$$

As shown in Proposition 1, the state evolution of S_t depends on the controlled inputs (γ_t^1, γ_t^2) and the random noise $(V_t, W_{t+1}^1, W_{t+1}^2)$. This random noise is independent across time. Consequently, Π_t evolves in a controlled Markovian manner as below.

Proposition 2: For $t = 1, \dots, T-1$, there exists functions F_t (which do not depend on the coordinator's strategy) such that

$$\Pi_{t+1} = F_{t+1}(\Pi_t, \gamma_t^1, \gamma_t^2, Z_{t+1}). \tag{23}$$

Proof: See Appendix A. ■

At $t = 1$, since there is no shared information, Π_1 is simply the unconditional probability $\mathbb{P}(S_1) = \mathbb{P}\left(X_0, Y_1^1, Y_1^2\right)$. Thus, Π_1 is fixed a priori from the joint distribution of the primitive random variables and does not depend on the choice of coordinator's strategy ψ. Proposition 2 shows that at $t = 2, \dots, T$, Π_t depends on the strategy $\boldsymbol{\psi}$ only through the choices of $\gamma_{1:t-1}^1$ and $\gamma_{1:t-1}^2$. Moreover, as shown in Proposition 1, the instantaneous cost at time t can be written in terms of the current and next states (S_t, S_{t+1}) and the control inputs (γ_t^1, γ_t^2). Combining the above two properties, we get the following:

Proposition 3: The process $\Pi_t,\ t = 1, 2, \ldots, T$ is a controlled Markov chain with γ_t^1, γ_t^2 as the control actions at time t, i.e.,

$$\begin{aligned}\mathbb{P}\left(\Pi_{t+1} \,\middle|\, \Delta_t, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2\right) &= \mathbb{P}\left(\Pi_{t+1} \,\middle|\, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2\right) \\ &= \mathbb{P}\left(\Pi_{t+1} \,\middle|\, \Pi_t, \gamma_t^1, \gamma_t^2\right). \end{aligned} \tag{24}$$

Furthermore, there exists a deterministic function C_t such that

$$\mathbb{E}\left\{\hat{c}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1}) \,\middle|\, \Delta_t, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2\right\} = C_t(\Pi_t, \gamma_t^1, \gamma_t^2). \tag{25}$$

Proof: See Appendix B. ∎

The controlled Markov property of the process $\{\Pi_t, t = 1, \ldots, T\}$ immediately gives rise to the following structural result.

Theorem 1: In Problem 2, without loss of optimality we can restrict attention to coordination strategies of the form

$$(\gamma_t^1, \gamma_t^2) = \psi_t(\Pi_t), \quad t = 1, \ldots, T. \tag{26}$$

Proof: From Proposition 3, we conclude that the optimization problem for the coordinator is to control the evolution of the controlled Markov process $\{\Pi_t,\ t = 1, 2, \ldots, T\}$ by selecting the partial functions $\{\gamma_t^1, \gamma_t^2,\ t = 1, 2, \ldots, T\}$ in order to minimize $\sum_{t=1}^T \mathbb{E}\left\{C_t(\Pi_t, \gamma_t^1, \gamma_t^2)\right\}$. This is an instance of the well-known Markov decision problems where it is known that the optimal strategy is a function of the current state. Thus, the structural result follows from Markov decision theory [1]. ∎

The above result can also be stated in terms of the original problem.

Theorem 2 (Structural Result): In Problem 1 with $K = 2$, without loss of optimality we can restrict attention to coordination strategies of the form

$$U_t^k = g_t^k(\Lambda_t^k, \Pi_t), \quad k = 1, 2. \tag{27}$$

where

$$\Pi_t = \mathbb{P}^{(g_{1:t-1}^1, g_{1:t-1}^2)}\left(X_{t-1}, \Lambda_t^1, \Lambda_t^2 \,\middle|\, \Delta_t\right) \tag{28}$$

where $\Pi_1 = \mathbb{P}\left(X_0, Y_1^1, Y_1^2\right)$ and for $t = 2, \ldots, T$, Π_t is evaluated as follows:

$$\Pi_{t+1} = F_{t+1}(\Pi_t, g_t^1(\cdot, \Pi_t), g_t^2(\cdot, \Pi_t), Z_{t+1}) \tag{29}$$

Proof: Theorem 1 established the structure of the optimal coordination strategy. As we argued in Stage 2, this optimal coordination strategy can be implemented in Problem 1 and is optimal for the objective (4). At $t = 1$, $\Pi_1 = \mathbb{P}\left(X_0, Y_1^1, Y_1^2\right)$ is known to both controllers and they can use the optimal coordination strategy to select partial functions according to:

$$(\gamma_1^1, \gamma_1^2) = \psi_1(\Pi_1)$$

Thus,

$$U_1^k = \gamma_1^k(\Lambda_1^k) = \psi_1^k(\Pi_1)(\Lambda_1^k) =: g_1^k(\Lambda_1^k, \Pi_1), \quad k = 1, 2. \tag{30}$$

At time instant $t + 1$, both controllers know Π_t and the common observations $Z_{t+1} = (Y_{t-n+1}^1, Y_{t-n+1}^2, U_{t-n+1}^1, U_{t-n+1}^2)$; they use the partial functions $(g_t^1(\cdot, \Pi_t), g_t^2(\cdot, \Pi_t))$ in equation (23) to evaluate Π_{t+1}. The control actions at time $t + 1$ are given as:

$$\begin{aligned}U_{t+1}^k = \gamma_{t+1}^k(\Lambda_{t+1}^k) &= \psi_{t+1}(\Pi_{t+1})(\Lambda_{t+1}^k) \\ &=: g_{t+1}^k(\Lambda_{t+1}^k, \Pi_{t+1}), \quad k = 1, 2. \end{aligned} \tag{31}$$

Moreover, using the design $\boldsymbol{g}$ defined according to (31), the coordinator's information state Π_t can also be written as:

$$\begin{aligned}\Pi_t &= \mathbb{P}^{\boldsymbol{\psi}}\left(X_{t-1}, \Lambda_t^1, \Lambda_t^2 \,\middle|\, \Delta_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2\right) \\ &= \mathbb{P}^{\boldsymbol{g}}\left(X_{t-1}, \Lambda_t^1, \Lambda_t^2 \,\middle|\, \Delta_t, g_1^{1:2}(\cdot, \Pi_1), \ldots, g_{t-1}^{1:2}(\cdot, \Pi_{t-1})\right) \\ &= \mathbb{P}^{(g_{1:t-1}^1, g_{1:t-1}^2)}\left(X_{t-1}, \Lambda_t^1, \Lambda_t^2 \,\middle|\, \Delta_t\right) \end{aligned} \tag{32}$$

where we dropped the partial functions from the conditioning terms in (32) because under the given control laws $(g_{1:t-1}^1, g_{1:t-1}^2)$, the partial functions used from time 1 to $t-1$ can be evaluated from Δ_t (by using Proposition 2 to evaluate $\Pi_{1:t-1}$). ∎

Theorem 2 establishes the first structural result stated in Section I-D for $K = 2$. In the next section, we show how to extend the result for general K.

B. Extension to General K

Theorem 2 for two controllers ($K = 2$) can be easily extended to general K by following the same sequence of arguments as in stages 1 to 4 above. Thus, at time t, the coordinator introduced in Stage 1 now selects partial functions $\gamma_t^k : \mathcal{L}^k \mapsto \mathcal{U}^k$, for $k = 1, 2, \ldots, K$. The state sufficient for input output mapping from the coordinator's perspective is given as $S_t := (X_{t-1}, \Lambda_t^{1:K})$ and the information state Π_t for the coordinator is

$$\Pi_t(s_t) := \mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s_t \,\middle|\, \Delta_t, \gamma_{1:t-1}^{1:K}\right). \tag{33}$$

Results analogous to Propositions 1–3 can now be used to conclude the structural result of Theorem 2 for general K.

C. Sequential Decomposition

In addition to obtaining the structural result of Theorem 2, the coordinator's problem also allows us to write a dynamic program for finding the optimal control strategies as shown below. We first focus on the two controller case ($K = 2$) and then extend the result to general K.

Theorem 3: The optimal coordination strategy can be found by the following dynamic program: For $t = 1, \ldots, T$, define the functions $J_t : \mathcal{P}\{S\} \mapsto \mathbb{R}$ as follows. For $\pi \in \mathcal{P}\{\mathcal{S}\}$ let

$$J_T(\pi) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} C_T(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2). \tag{34}$$

For $t = 1, \ldots, T-1$, and $\pi \in \mathcal{P}\{\mathcal{S}\}$ let

$$\begin{aligned}J_t(\pi) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \Big[&C_t(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2) \\ &+ \mathbb{E}\left\{J_{t+1}(\Pi_{t+1}) \,\middle|\, \Pi_t = \pi, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}\right\}\Big]. \end{aligned} \tag{35}$$

The arg inf $(\gamma_t^{*,1}, \gamma_t^{*,2})$ in the RHS of $J_t(\pi)$ is the optimal action for the coordinator at time t then $\Pi_t = \pi$. Thus,

$$(\gamma_t^{*,1}, \gamma_t^{*,2}) = \phi_t^*(\pi)$$

The corresponding control strategy for Problem 1, given by (17) is optimal for Problem 1.

Proof: As in Theorem 1, we use the fact that the coordinator's optimization problem can be viewed as a Markov decision problem with Π_t as the state of the Markov process. The dynamic program follows from standard results in Markov decision theory [1]. The optimality of the corresponding control strategy for Problem 1 follows from the equivalence between the two problems. ∎

The dynamic program of Theorem 3 can be extended to general K in a manner similar to Section II-B.

D. Computational Aspects

In the dynamic program for the coordinator in Theorem 3, the value functions at each time are functions defined on the continuous space $\mathcal{P}\{\mathcal{S}\}$, whereas the minimization at each time step is over the finite set of functions from the space of realizations of the private information of controllers ($\mathcal{L}^k$, $k = 1, 2$) to the space of control actions ($\mathcal{U}^k$, $k = 1, 2$). While dynamic programs with continuous state space can be hard to solve, we note that our dynamic program resembles the dynamic program for partially observable Markov decision problems (POMDP). In particular, just as in POMDP, the value-function at time T is piecewise linear in Π_T and by standard backward recursion, it can be shown that value-function at time t is piecewise linear and concave function of Π_t. (See Appendix C). Indeed, the coordinator's problem can be viewed as a POMDP, with S_t as the underlying partially observed state and the belief Π_t as the information state of the POMDP. The characterization of value functions as piecewise linear and concave 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 [6] and references therein for a review of algorithms to solve POMDPs.

E. One-step Delay

We now focus on the one-step delayed sharing information structure, i.e., when $n = 1$. For this case, the structural result (7) asserted by Witsenhausen is correct [3]. At first glance, that structural result looks different from our structural result (9) for $n = 1$. In this section, we show that for $n = 1$, these two structural results are equivalent.

As before, we consider the two-controller system ($K = 2$). When delay $n = 1$, we have

$$\Delta_t = (Y^1_{1:t-1}, Y^2_{1:t-1}, U^1_{1:t-1}, U^2_{1:t-1}),$$
$$\Lambda^1_t = (Y^1_t), \quad \Lambda^2_t = (Y^2_t),$$

and

$$Z_{t+1} = (Y^1_t, Y^2_t, U^1_t, U^2_t).$$

The result of Theorem 2 can now be restated for this case as follows:

Corollary 1: In Problem 1 with $K = 2$ and $n = 1$, without loss of optimality we can restrict attention to control strategies of the form:

$$U^k_t = g^k_t(Y^k_t, \Pi_t), \quad k = 1, 2. \tag{36}$$

where

$$\Pi_t := \mathbb{P}^{(g^1_{1:t-1}, g^2_{1:t-1})}\left(X_{t-1}, Y^1_t, Y^2_t \,\middle|\, \Delta_t\right) \tag{37}$$

We can now compare our result for one-step delay with the structural result (7), asserted in [2] and proved in [3]. For $n = 1$, this result states that without loss of optimality, we can restrict attention to control laws of the form:

$$U^k_t = g^k_t(Y^k_t, \mathbb{P}(X_{t-1} \,|\, \Delta_t)), \quad k = 1, 2. \tag{38}$$

The above structural result can be recovered from (37) by observing that there is a one-to-one correspondence between Π_t and the belief $\mathbb{P}(X_{t-1} \,|\, \Delta_t)$. We first note that

$$\begin{aligned}\Pi_t &= \mathbb{P}^{(g^1_{1:t-1}, g^2_{1:t-1})}\left(X_{t-1}, Y^1_t, Y^2_t \,\middle|\, \Delta_t\right) \\ &= \mathbb{P}\left(Y^1_t \,\middle|\, X_{t-1}\right) \cdot \mathbb{P}\left(Y^2_t \,\middle|\, X_{t-1}\right) \\ &\quad \cdot \mathbb{P}^{(g^1_{1:t-1}, g^2_{1:t-1})}(X_{t-1} \,|\, \Delta_t)\end{aligned} \tag{39}$$

As pointed out in [2], [3] (and proved later in this paper in Proposition 4), the last probability does not depend on the functions $(g^1_{1:t-1}, g^2_{1:t-1})$. Therefore,

$$\Pi_t = \mathbb{P}\left(Y^1_t \,\middle|\, X_{t-1}\right) \cdot \mathbb{P}\left(Y^2_t \,\middle|\, X_{t-1}\right) \cdot \mathbb{P}(X_{t-1} \,|\, \Delta_t) \tag{40}$$

Clearly, the belief $\mathbb{P}(X_{t-1} \,|\, \Delta_t)$ is a marginal of Π_t and therefore can be evaluated from Π_t. Moreover, given the belief $\mathbb{P}(X_{t-1} \,|\, \Delta_t)$, one can evaluate Π_t using equation (40). This one-to-one correspondence between Π_t and $\mathbb{P}(X_{t-1} \,|\, \Delta_t)$ means that the structural result proposed in this paper for $n = 1$ is effectively equivalent to the one proved in [3].

III. Proof of the Second Structural Result

In this section we prove the second structural result (10). As in Section II, we prove the result for $K = 2$ and then show how to extend it for general K. To prove the result, we reconsider the coordinator's problem at Stage 3 of Section II and present an alternative characterization for the coordinator's optimal strategy in Problem 2. The main idea in this section is to use the dynamics of the system evolution and the observation equations (equations (1) and (2)) to find an equivalent representation of the coordinator's information state. We also contrast this information state with that proposed by Witsenhausen.

A. Two controller system ($K = 2$)

Consider the coordinator's problem with $K = 2$. Recall that γ^1_t and γ^2_t are the coordinator's actions at time t. γ^k_t maps the private information of the k^{th} controller $(Y^k_{t-n+1:t}, U^k_{t-n+1:t-1})$ to its action U^k_t. In order to find an alternate characterization of coordinator's optimal strategy, we need the following definitions:

Definition 3: For a coordination strategy ψ, and for $t = 1, 2, \ldots, T$ we define the following:

1) $\Theta_t := \mathbb{P}(X_{t-n} \,|\, \Delta_t)$
2) For $k = 1, 2$, define the following partial functions of γ^k_m

$$r^k_{m,t}(\cdot) := \gamma^k_m(\cdot, Y^k_{m-n+1:t-n}, U^k_{m-n+1:t-n}),$$
$$m = t-n+1, t-n+2, \ldots, t-1 \tag{41}$$

Since γ_m^k is a function that maps $(Y_{m-n+1:m}^k, U_{m-n+1:m-1}^k)$ to U_m^k, $r_{m,t}^k(\cdot)$ is a function that maps $(Y_{t-n+1:m}^k, U_{t-n+1:m-1}^k)$ to U_m^k. We define a collection of these partial functions as follows:

$$r_t^k := (r_{m,t}^k(\cdot), m = t-n+1, t-n+2, \ldots, t-1) \quad (42)$$

Note that for $n=1$, r_t^k is empty.

We need the following results to address the coordinator's problem:

Proposition 4: 1) For $t = 1, \ldots, T-1$, there exists functions Q_t, Q_t^k, $k = 1, 2$, (which do not depend on the coordinator's strategy) such that

$$\begin{aligned} \Theta_{t+1} &= Q_t(\Theta_t, Z_{t+1}) \\ r_{t+1}^k &= Q_t^k(r_t^k, Z_{t+1}, \gamma_t^k) \end{aligned} \quad (43)$$

2) The coordinator's information state Π_t is a function of (Θ_t, r_t^1, r_t^2). Consequently, for $t = 1, \ldots, T$, there exist functions $\hat{C}_t$ (which do not depend on the coordinator's strategy) such that

$$\begin{aligned} &\mathbb{E}\left\{\hat{c}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1}) \,\middle|\, \Delta_t, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2\right\} \\ &\qquad = \hat{C}_t(\Theta_t, r_t^1, r_t^2, \gamma_t^1, \gamma_t^2) \end{aligned} \quad (44)$$

3) The process (Θ_t, r_t^1, r_t^2), $t = 1, 2, \ldots, T$ is a controlled Markov chain with γ_t^1, γ_t^2 as the control actions at time t, i.e.,

$$\begin{aligned} &\mathbb{P}\left(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2 \,\middle|\, \Delta_t, \Theta_{1:t}, r_{1:t}^1, r_{1:t}^2, \gamma_{1:t}^1, \gamma_{1:t}^2\right) \\ &\quad = \mathbb{P}\left(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2 \,\middle|\, \Theta_{1:t}, r_{1:t}^1, r_{1:t}^2, \gamma_{1:t}^1, \gamma_{1:t}^2\right) \\ &\quad = \mathbb{P}\left(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2 \,\middle|\, \Theta_t, r_t^1, r_t^2, \gamma_t^1, \gamma_t^2\right). \end{aligned} \quad (45)$$

Proof: See Appendix D. ■

At $t = 1$, since there is no sharing of information, Θ_1 is simply the unconditioned probability $\mathbb{P}(X_0)$. Thus, Θ_1 is fixed a priori from the joint distribution of the primitive random variables and does not depend on the choice of the coordinator's strategy ψ. Proposition 4 shows that the update of Θ_t depends only on Z_{t+1} and not on the coordinator's strategy. Consequently, the belief Θ_t depends only on the distribution of the primitive random variables and the realizations of $Z_{1:t}$. We can now show that the coordinator's optimization problem can be viewed as an MDP with (Θ_t, r_t^1, r_t^2), $t = 1, 2, \ldots, T$ as the underlying Markov process.

Theorem 4: (Θ_t, r_t^1, r_t^2) is an information state for the coordinator. That is, there is an optimal coordination strategy of the form:

$$(\gamma_t^1, \gamma_t^2) = \psi_t(\Theta_t, r_t^1, r_t^2), \quad t = 1, \ldots, T. \quad (46)$$

Moreover, this optimal coordination strategy can be found by the following dynamic program:

$$\begin{aligned} J_T(\theta, \tilde{r}^1, \tilde{r}^2) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} &\mathbb{E}\big\{\hat{C}_T(\Theta_T, r_T^1, r_T^2, \gamma_T^1, \gamma_T^2) \,\big|\, \\ &\Theta_T = \theta, r_T^{1:2} = \tilde{r}^{1:2}, \gamma_T^{1:2} = \tilde{\gamma}^{1:2}\big\}. \end{aligned} \quad (47)$$

For $t = 1, \ldots, T-1$, let

$$\begin{aligned} J_t(\theta, \tilde{r}^1, \tilde{r}^2) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} &\mathbb{E}\big\{\hat{C}_t(\Theta_t, r_t^1, r_t^2, \gamma_1^1, \gamma_t^2) \\ &+ J_{t+1}(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2) \,\big|\, \\ &\Theta_t, = \theta, r_t^{1:2} = \tilde{r}^{1:2}, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}\big\} \end{aligned} \quad (48)$$

where $\theta \in \mathcal{P}\{\mathcal{X}\}$, and $\tilde{r}^1, \tilde{r}^2$ are realizations of partial functions defined in (41) and (42). The arg inf $(\gamma_t^{*,1}, \gamma_t^{*,2})$ in the RHS of (48) is the optimal action for the coordinator at time t when $(\Theta_t, r_t^1, r_t^2) = (\theta, \tilde{r}^1, \tilde{r}^2)$. Thus,

$$(\gamma_t^{*,1}, \gamma_t^{*,2}) = \psi_t^*(\Theta_t, r_t^1, r_t^2)$$

The corresponding control strategy for Problem 1, given by (17) is optimal for Problem 1.

Proof: Proposition 4 implies that the coordinator's optimization problem can be viewed as an MDP with (Θ_t, r_t^1, r_t^2), $t = 1, 2, \ldots, T$ as the underlying Markov process and $\hat{C}_t(\Theta_t, r_t^1, r_t^2, \gamma_t^1, \gamma_t^2)$ as the instantaneous cost. The MDP formulation implies the result of the theorem. ■

The following result follows from Theorem 4.

Theorem 5 (Second Structural Result): In Problem 1 with $K = 2$, without loss of optimality we can restrict attention to coordination strategies of the form

$$U_t^k = g_t^k(\Lambda_t^k, \Theta_t, r_t^1, r_t^2), \quad k = 1, 2. \quad (49)$$

where

$$\Theta_t = \mathbb{P}(X_{t-n} \mid \Delta_t) \quad (50)$$

and

$$\begin{aligned} r_t^k = \{(g_m^k(\cdot, Y_{m-n+1:t-n}^k, U_{m-n+1:t-n}^k, \Delta_m), \\ t-n+1 \le m \le t-1\} \end{aligned} \quad (51)$$

Proof: As in Theorem 2, equations (17) can be used to identify an optimal control strategy for each controller from the optimal coordination strategy given in Theorem 4. ■

Theorem 4 and Theorem 5 can be easily extended for K controllers by identifying $(\Theta_t, r_t^{1:K})$ as the information state for the coordinator.

B. Comparison to Witsenhausen's Result

We now compare the result of Theorem 4 to Witsenhausen's conjecture which states that there exist optimal control strategies of the form:

$$U_t^k = g_t^k(\Lambda_t^k, \mathbb{P}(X_{t-n} \mid \Delta_t)). \quad (52)$$

Recall that Witsenhausen's conjecture is true for $n = 1$ but false for $n > 1$. Therefore, we consider the cases $n = 1$ and $n > 1$ separately:

Delay $n = 1$: For a two-controller system with $n = 1$, we have

$$\begin{aligned} \Delta_t = (Y_{1:t-1}^1, Y_{1:t-1}^2, U_{1:t-1}^1, U_{1:t-1}^2), \\ \Lambda_t^1 = (Y_t^1), \quad \Lambda_t^2 = (Y_t^2), \end{aligned}$$

and

$$r_t^1 = \emptyset, \quad r_t^2 = \emptyset$$

Therefore, for $n = 1$, Theorem 5 implies that there exist optimal control strategies of the form:

$$U_t^k = g_t^k(\Lambda_t^k, \mathbb{P}(X_{t-n} \,|\, \Delta_t)), \quad k = 1, 2. \tag{53}$$

Equation (53) is the same as equation (52) for $n = 1$. Thus, for $n = 1$, the result of Theorem 4 coincides with Witsenhausen's conjecture which was proved in [3].

Delay $n > 1$: Witsenhausen's conjecture implied that the controller k at time t can choose its action based only on the knowledge of Λ_t^k and $\mathbb{P}(X_{t-n} \,|\, \Delta_t)$, without any dependence on the choice of previous control laws ($g_{1:t-1}^{1:2}$). In other words, the argument of the control law g_t^k (that is, the information state at time t) is separated from $g_{1:t-1}^{1:2}$. However, as Theorem 5 shows, such a separation is not true because of the presence of the collection of partial functions r_t^1, r_t^2 in the argument of the optimal control law at time t. These partial functions depend on the choice of previous $n-1$ control laws. Thus, the argument of g_t^k depends on the choice of $g_{t-n+1:t-1}^{1:2}$. One may argue that Theorem 5 can be viewed as a *delayed or partial* separation since the information state for the control law g_t^k is separated from the choice of control laws before time $t - n + 1$.

Witsenhausen's conjecture implied that controllers employ common information only to form a belief on the state X_{t-n}; the controllers do not need to use the common information to guess each other's behavior from $t-n+1$ to the current time t. Our result disproves this statement. We show that in addition to forming the belief on X_{t-n}, each controller should use the common information to predict the actions of other controllers by means of the partial functions r_t^1, r_t^2.

IV. A Special Case of Delayed Sharing Information Structure

Many decentralized systems consist of coupled subsystems, where each subsystem has a controller that perfectly observes the state of the subsystem. If all controllers can exchange their observations and actions with a delay of n steps, then the system is a special case of the n-step delayed sharing information structure with the following assumptions:

1) *Assumption 1:* At time $t = 1, \dots, T$, the state of the system is given as the vector $X_t := (X_t^{1:K})$, where X_t^i is the state of subsystem i.
2) *Assumption 2:* The observation equation of the k^{th} controller is given as:

$$Y_t^k = X_{t-1}^k \tag{54}$$

This model is similar to the model considered in [7]. Clearly, the first structural result and the sequential decomposition of Section II apply here as well with the observations Y_t^k being replaced by X_t^k. Our second structural result simplifies when specialized to this model. Observe that in this model

$$\Delta_t = (Y_{1:t-n}^{1:K}, U_{1:t-n}^{1:K}) = (X_{1:t-n-1}, U_{1:t-n}^{1:K}) \tag{55}$$

and therefore the belief,

$$\Theta_t = \mathbb{P}(X_{t-n} \,|\, \Delta_t) = \mathbb{P}\left(X_{t-n} \,\middle|\, X_{t-n-1}, U_{t-n}^{1:K}\right) \tag{56}$$

where we used the controlled Markov nature of the system dynamics in the second equality in (56). Thus, Θ_t is a function only of $X_{t-n-1}, U_{t-n}^{1:K}$. The result of Theorem 4 can now be restated for this case as follows:

Corollary 2: In Problem 1 with assumptions 1 and 2, there is an optimal coordination strategy of the form:

$$(\gamma_t^1, \gamma_t^2) = \psi_t(X_{t-n-1}, U_{t-n}^1, U_{t-n}^2, r_t^1, r_t^2), \quad t = 1, \dots, T. \tag{57}$$

Moreover, this optimal coordination strategy can be found by the following dynamic program:

$$\begin{aligned} &J_T(x, u^1, u^2, \tilde{r}^1, \tilde{r}^2) \\ &\quad = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \mathbb{E}\big\{\hat{C}_T(X_{T-n}, r_T^1, r_T^2, \gamma_T^1, \gamma_T^2) \,\big|\, X_{T-n-1} = x, \\ &\qquad\qquad U_{T-n}^{1:2} = u^{1:2}, r_T^{1:2} = \tilde{r}^{1:2}, \gamma_T^{1:2} = \tilde{\gamma}^{1:2}\big\}. \end{aligned} \tag{58}$$

For $t = 1, \dots, T-1$, let

$$\begin{aligned} J_t(x, u^1, u^2, \tilde{r}^1, \tilde{r}^2) &= \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \mathbb{E}\{\hat{C}_t(X_{t-n}, r_t^1, r_t^2, \gamma_t^1, \gamma_t^2) \\ &\quad + J_{t+1}(X_{t-n+1}, r_{t+1}^1, r_{t+1}^2) \,\big|\, X_{t-n-1} = x, \\ &\quad U_{t-n}^{1:2} = u^{1:2}, r_t^{1:2} = \tilde{r}^{1:2}, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}\big\}. \end{aligned} \tag{59}$$

We note that the structural result and the sequential decomposition in the corollary above is analogous to [7, Theorem 1].

A. An Example

We consider a simple example of a delayed sharing information structure with two controllers ($K = 2$), a time horizon $T = 3$ and delay $n = 2$. Varaiya and Walrand [3] used this example to show that Witsenhausen's proposed structure was suboptimal.

The system dynamics are given by

$$\begin{aligned} X_0 &= (X_0^1, X_0^2) \\ X_1 &= (X_1^1, X_1^2) = (X_0^1 + X_0^2, 0) \\ X_2 &= (X_2^1, X_2^2) = (X_1^1, U_2^2) = (X_0^1 + X_0^2, U_2^2) \\ X_3 &= (X_3^1, X_3^2) = (X_2^1 - X_2^2 - U_3^1, 0) \\ &= (X_0^1 + X_0^2 - U_2^2 - U_3^1, 0) \end{aligned}$$

X_0^1, X_0^2 are zero-mean, jointly Gaussian random variables with variance 1 and covariance -0.5. The observation equations are:

$$Y_t^k = X_{t-1}^k$$

and the total cost function is

$$\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}\left\{(X_3^1)^2 + (U_3^1)^2\right\} \tag{60}$$

We can now specify the common and private informations.
Common Information:

$$\Delta_1 = \emptyset, \quad \Delta_2 = \emptyset,$$
$$\Delta_3 = \{Y_1^1, Y_1^2, U_1^1, U_1^2\} = \{X_0^1, X_0^2, U_1^1, U_1^2\}$$

Private Information for Controller 1:

$$\begin{aligned} \Lambda_1^1 &= \{Y_1^1\} = \{X_0^1\}, \\ \Lambda_2^1 &= \{Y_1^1, Y_2^1, U_1^1\} = \{X_0^1, (X_0^1 + X_0^2), U_1^1\} \\ \Lambda_3^1 &= \{Y_2^1, Y_3^1, U_2^1\} = \{(X_0^1 + X_0^2), (X_0^1 + X_0^2), U_2^1\} \end{aligned}$$

Private Information for Controller 2:

$$\begin{aligned}\Lambda_1^2 &= \{Y_1^2\} = \{X_0^2\},\\ \Lambda_2^2 &= \{Y_1^2, Y_2^2, U_1^2\} = \{X_0^2, 0, U_1^2\}\\ \Lambda_3^2 &= \{Y_2^2, Y_3^2, U_2^2\} = \{0, U_2^2, U_2^2\}\end{aligned}$$

The total cost can be written as:

$$\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}\left\{(X_0^1 + X_0^2 - U_2^2 - U_3^1)^2 + (U_3^1)^2\right\} \tag{61}$$

Thus, the only control actions that affect the cost are U_2^2 and U_3^1 . Hence, we can even assume that all other control laws are constant functions with value 0 and the performance of a design is completely characterized by control laws g_2^2 and g_3^1. Using the fact that all control actions other than U_2^2 and U_3^1 are 0, we get the following simplified control laws:

$$\begin{aligned}U_2^2 &= g_2^2(\Lambda_2^2, \Delta_2) = g_2^2(X_0^2)\\ U_3^1 &= g_3^1(\Lambda_3^1, \Delta_3) = g_3^1((X_0^1 + X_0^2), X_0^1, X_0^2)\\ &= g_3^1(X_0^1, X_0^2)\end{aligned}$$

Now consider control laws of the form given in Theorem 5 given by

$$U_t^k = g_t^k(\Lambda_t^k, \Theta_t, r_t^1, r_t^2) \tag{62}$$

For $k = 2$ and $t = 2$, Θ_2 is a fixed prior distribution of X_0, while r_2^1, r_2^2 are constant functions. Hence, Θ_t, r_t^1, r_t^2 provide no new information and the structure of equation (62) boils down to

$$U_2^2 = g_2^2(\Lambda_2^2) = g_2^2(X_0^2) \tag{63}$$

For $k = 1$ and $t = 3$,

$$\Theta_3 = \mathbb{P}\left(X_1 \,|\, \Delta_3\right) = \mathbb{P}\left((X_0^1 + X_0^2, 0) \,\middle|\, X_0^1, X_0^2\right)$$

and

$$\begin{aligned}r_3^2 &= \{(g_m^2(\cdot, Y_{m-1:1}^2, U_{m-1:1}^2, \Delta_m), 2 \le m \le 2\}\\ &= \{(g_2^2(\cdot, Y_1^2, U_1^2, \Delta_2)\}\\ &= \{(g_2^2(\cdot, X_0^2)\} = U_2^2,\end{aligned}$$

while r_3^1 are partial functions of constant functions. Therefore, equation (62) can now be written as:

$$\begin{aligned}U_3^1 &= g_3^1((X_0^1 + X_0^2), \mathbb{P}\left((X_0^1 + X_0^2, 0) \,\middle|\, X_0^1, X_0^2\right), U_2^2)\\ &= g_3^1((X_0^1 + X_0^2), (X_0^1 + X_0^2), U_2^2) \qquad (64)\\ &= g_3^1((X_0^1 + X_0^2), U_2^2) \qquad (65)\end{aligned}$$

where we used the fact that knowing $\mathbb{P}\left((X_0^1 + X_0^2, 0) \,\middle|\, X_0^1, X_0^2\right)$ is same as knowing the value of $(X_0^1 + X_0^2)$ in (64).

The optimal control laws can be obtained by solving the coordinator's dynamic program given in Theorem 4. Observe that $\Theta_3 = \mathbb{P}\left((X_0^1 + X_0^2, 0) \,\middle|\, X_0^1, X_0^2\right)$ is equivalent to $(X_0^1 + X_0^2)$ and that r_3^2 is equivalent to U_2^2. Thus, the dynamic program can be simplified to:

$$\begin{aligned}&J_3((x_0^1 + x_0^2), u_2^2)\\ &= \inf_{\tilde\gamma^1} \mathbb{E}\left\{(X_3^1)^2 + (U_3^1)^2 \,\middle|\, \begin{array}{l}(X_0^1 + X_0^2) = (x_0^1 + x_0^2),\\ U_2^2 = u_2^2, \gamma_3^1 = \tilde\gamma^1\end{array}\right\}\end{aligned}$$

where, for the given realization of $((x_0^1 + x_0^2), u_2^2)$, $\tilde\gamma^1$ maps $\Lambda_3^1 = (X_0^1 + X_0^2)$ to U_3^1. Further simplification yields:

$$\begin{aligned}&J_3((x_0^1 + x_0^2), u_2^2)\\ &= \inf_{\tilde\gamma^1} \mathbb{E}\Big\{(X_0^1 + X_0^2 - U_2^2 - U_3^1)^2 + (U_3^1)^2 \,\Big|\\ &\qquad (X_0^1 + X_0^2) = (x_0^1 + x_0^2), U_2^2 = u_2^2, \gamma_3^1 = \tilde\gamma^1\Big\}\\ &\ge (x_0^1 + x_0^2 - u_2^2)^2/2, \qquad (66)\end{aligned}$$

where the right hand side in (66) is the lower bound on the expression $(x_0^1 + x_0^2 - u_2^2 - u_3^1)^2 + (u_3^1)^2$ for any u_3^1. Given the fixed realization of $((x_0^1 + x_0^2), u_2^2)$, choosing γ^1 as a constant function with value $(x_0^1 + x_0^2 - u_2^2)/2$ achieves the lower bound in (66). For $t = 2$, the coordinator has no information and the value function at time $t = 2$ is

$$\begin{aligned}J_2 &= \inf_{\tilde\gamma^2} \mathbb{E}\left\{J_3((X_0^1 + X_0^2), U_2^2) \,\middle|\, \gamma_2^2 = \tilde\gamma^2\right\}\\ &= \inf_{\tilde\gamma^2} \mathbb{E}\left\{(X_0^1 + X_0^2 - U_2^2)^2/2 \,\middle|\, \gamma_2^2 = \tilde\gamma^2\right\} \qquad (67)\end{aligned}$$

where $\tilde\gamma^2$ maps $\Lambda_2^1 = (X_0^2)$ to U_2^2. The optimization problem in (67) is to choose, for each value of x_0^2, the best estimate (in a mean squared error sense) of $(X_0^1 + X_0^2)$. Given the Gaussian statistics, the optimal choice of $\tilde\gamma^2$ can be easily shown to be $\gamma^2(x_0^2) = x_0^2/2$.

Thus, the optimal strategy for the coordinator is to choose $\gamma_2^2(x_0^2) = x_0^2/2$ at time $t = 2$, and at $t = 3$, given the fixed realization of $((x_0^1 + x_0^2), u_2^2)$, choose $\gamma^1(\cdot) = (x_0^1 + x_0^2 - u_2^2)/2$. Thus, the optimal control laws are:

$$\begin{aligned}U_2^2 &= g_2^2(X_0^2) = X_0^2/2 \qquad (68)\\ U_3^1 &= g_3^1((X_0^1 + X_0^2), U_2^2)\\ &= (X_0^1 + X_0^2 - U_2^2)/2 \qquad (69)\end{aligned}$$

These are same as the unique optimal control laws identified in [3].

V. KURTARAN'S SEPARATION RESULT

In this section, we focus on the structural result proposed by Kurtaran [4]. We restrict to the two controller system ($K = 2$) and delay $n = 2$. For this case, we have

$$\begin{aligned}\Delta_t &= (Y_{1:t-2}^1, Y_{1:t-2}^2, U_{1:t-2}^1, U_{1:t-2}^2),\\ \Lambda_t^1 &= (Y_t^1, Y_{t-1}^1, U_{t-1}^1), \quad \Lambda_t^2 = (Y_t^2, Y_{t-1}^2, U_{t-1}^2),\end{aligned}$$

and

$$Z_{t+1} = (Y_{t-1}^1, Y_{t-1}^2, U_{t-1}^1, U_{t-1}^2).$$

Kurtaran's structural result for this case states that without loss of optimality we can restrict attention to control strategies of the form:

$$U_t^k = g_t^k(\Lambda_t^k, \Phi_t), \quad k = 1, 2, \tag{70}$$

where

$$\Phi_t := \mathbb{P}^{\boldsymbol{g}}\left(X_{t-2}, U_{t-1}^1, U_{t-1}^2 \,\middle|\, \Delta_t\right).$$

Kurtaran [4] proved this result for the terminal time-step T and simply stated that the result for $t = 1, \ldots, T-1$ can be established by the dynamic programming argument given in [8]. We believe that this is not the case.

In the dynamic programming argument in [8], a critical step is the update of the information state Φ_t, which is given by [8, Eq (30)]. For the result presented in [4], the corresponding equation is

$$\Phi_{t+1} = F_t(\Phi_t, Y^1_{t-1}, Y^2_{t-1}, U^1_{t-1}, U^2_{t-1}). \tag{71}$$

We believe that such an update equation cannot be established.

To see the difficulty in establishing (71), lets follow an argument similar to the proof of [8, Eq (30)] given in [8, Appendix B]. For a fixed strategy $\boldsymbol{g}$, and a realization δ_{t+1} of Δ_{t+1}, the realization φ_{t+1} of Φ_{t+1} is given by

$$\begin{aligned}\varphi_{t+1} &= \mathbb{P}\left(x_{t-1}, u^1_t, u^2_t \,\middle|\, \delta_{t+1}\right) \\ &= \mathbb{P}\left(x_{t-1}, u^1_t, u^2_t \,\middle|\, \delta_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1}\right) \\ &= \frac{\mathbb{P}\left(x_{t-1}, u^1_t, u^2_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1} \,\middle|\, \delta_t\right)}{\displaystyle\sum_{(x',a^1,a^2)\in\mathcal{X}\times\mathcal{U}^1\times\mathcal{U}^2} \begin{aligned}&\mathbb{P}(X_{t-1}=x', U^1_t=a^1, U^2_t=a^2, \\ &\quad y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1} \mid \delta_t)\end{aligned}}\end{aligned} \tag{72}$$

The numerator can be expressed as:

$$\begin{aligned}&\mathbb{P}\left(x_{t-1}, u^1_t, u^2_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1} \,\middle|\, \delta_t\right) \\ &\quad= \sum_{(x_{t-2},y^1_t,y^2_t)\in\mathcal{X}\cdot\mathcal{Y}^1\cdot\mathcal{Y}^2} \begin{aligned}&\mathbb{P}(x_{t-1}, u^1_t, u^2_t, y^1_{t-1}, y^2_{t-1}, \\ &\quad u^1_{t-1}, u^2_{t-1}, x_{t-2}, y^1_t, y^2_t|\delta_t)\end{aligned} \\ &\quad= \sum_{(x_{t-2},y^1_t,y^2_t)\in\mathcal{X}\cdot\mathcal{Y}^1\cdot\mathcal{Y}^2} \mathbb{1}_{g^1_t(\delta_t,u^1_{t-1},y^1_{t-1},y^1_t)}[u^1_t] \\ &\qquad \cdot \mathbb{1}_{g^2_t(\delta_t,u^2_{t-1},y^2_{t-1},y^2_t)}[u^2_t] \\ &\qquad \cdot \mathbb{P}\left(y^1_t \,\middle|\, x_{t-1}\right) \cdot \mathbb{P}\left(y^2_t \,\middle|\, x_{t-1}\right) \\ &\qquad \cdot \mathbb{P}\left(x_{t-1} \,\middle|\, x_{t-2}, u^1_{t-1}, u^2_{t-1}\right) \\ &\qquad \cdot \mathbb{1}_{g^1_{t-1}(\delta_{t-1},u^1_{t-2},y^1_{t-2},y^1_{t-1})}[u^1_{t-1}] \\ &\qquad \cdot \mathbb{1}_{g^2_t(\delta_{t-1},u^2_{t-2},y^2_{t-2},y^2_{t-1})}[u^2_{t-2}] \\ &\qquad \cdot \mathbb{P}\left(y^1_{t-1} \,\middle|\, x_{t-2}\right) \cdot \mathbb{P}\left(y^2_{t-1} \,\middle|\, x_{t-2}\right) \\ &\qquad \cdot \mathbb{P}\left(x_{t-2} \,\middle|\, \delta_t\right)\end{aligned} \tag{73}$$

If, in addition to φ_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, and u^2_{t-1}, each term of (73) depended only on terms that are being summed over (x_{t-2}, y^1_t, y^2_t), then (73) would prove (71). However, this is not the case: the first two terms also depend on δ_t. Therefore, the above calculation shows that φ_{t+1} is a function of $\varphi_t, Y^1_{t-1}, Y^2_{t-1}, U^1_{t-1}, U^2_{t-1}$ *and* δ_t. This dependence on δ_t is not an artifact of the order in which we decided to use the chain rule in (73) (we choose the natural sequential order in the system). No matter how we try to write φ_{t+1} in terms of φ_t, there will be a dependence on δ_t.

The above argument shows that it is not possible to establish (71). Consequently, the dynamic programming argument presented in [8] breaks down when working with the information state of [4], and, hence, the proof in [4] is incomplete. So far, we have not been able to correct the proof or find a counterexample to it.

VI. Conclusion

We studied the stochastic control problem with n-step delay sharing information structure and established two structural results for it. Both the results characterize optimal control laws with time-invariant domains. Our second result also establishes a partial separation result, that is, it shows that the information state at time t is separated from choice of laws before time $t-n+1$. Both the results agree with Witsenhausen's conjecture for $n = 1$. To derive our structural results, we formulated an alternative problem from the point of a coordinator of the system. We believe that this idea of formulating an alternative problem from the point of view of a coordinator which has access to information common to all controllers is also useful for general decentralized control problems, as is illustrated by [9] and [10].

Appendix A
Proof of Proposition 2

Fix a coordinator strategy $\boldsymbol{\psi}$. Consider a realization δ_{t+1} of the common information Δ_{t+1}. Let $(\tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})$ be the corresponding realization of partial functions until time t. Assume that the realization $(\delta_{t+1}, \pi_{1:t}, \gamma^1_{1:t}, \tilde{\gamma}^2_{1:t})$ has non-zero probability. Then, the realization π_{t+1} of Π_{t+1} is given by

$$\pi_{t+1}(s_{t+1}) = \mathbb{P}^{\boldsymbol{\psi}}\left(S_{t+1} = s_{t+1} \,\middle|\, \delta_{t+1}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right). \tag{74}$$

Using Proposition 1 , this can be written as

$$\begin{aligned}\sum_{s_t,v_t,w^1_{t+1},w^2_{t+1}} &\mathbb{1}_{s_{t+1}}(\hat{f}_{t+1}(s_t, v_t, w^1_{t+1}, w^2_{t+1}, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t)) \\ &\cdot \mathbb{P}\left(V_t = v_t\right) \cdot \mathbb{P}\left(W^1_{t+1} = w^1_{t+1}\right) \\ &\cdot \mathbb{P}\left(W^2_{t+1} = w^2_{t+1}\right) \cdot \mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s_t \,\middle|\, \delta_{t+1}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right).\end{aligned} \tag{75}$$

Since $\delta_{t+1} = (\delta_t, z_{t+1})$, the last term of (75) can be written as

$$\begin{aligned}&\mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s_t \,\middle|\, \delta_t, z_{t+1}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right) \\ &\quad= \frac{\mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s_t, Z_{t+1} = z_{t+1} \,\middle|\, \delta_t, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right)}{\sum_{s'} \mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s', Z_{t+1} = z_{t+1} \,\middle|\, \delta_t, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right)}.\end{aligned} \tag{76}$$

Use (20) and the sequential order in which the system variables are generated to write

$$\begin{aligned}&\mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s_t, Z_{t+1} = z_{t+1} \,\middle|\, \delta_t, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right) \\ &\quad= \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}^{\boldsymbol{\psi}}\left(S_t = s_t \,\middle|\, \delta_t, \tilde{\gamma}^1_{1:t-1}, \tilde{\gamma}^2_{1:t-1}\right) &(77) \\ &\quad= \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t). &(78)\end{aligned}$$

where $\tilde{\gamma}^1_t, \tilde{\gamma}^2_t$ are dropped from conditioning in (77) because for the given coordinator's strategy, they are functions of the rest of the terms in the conditioning. Substitute (78), (76), and (75) into (74), to get

$$\pi_{t+1}(s_{t+1}) = F_{t+1}(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, z_{t+1})(s_{t+1})$$

where $F_{t+1}(\cdot)$ is given by (74), (75), (76), and (78).

Appendix B
Proof of Proposition 3

Fix a coordinator strategy ψ. Consider a realization δ_{t+1} of the common information Δ_{t+1}. Let $\pi_{1:t}$ be the corresponding realization of $\Pi_{1:t}$ and $(\tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})$ the corresponding choice of partial functions until time t. Assume that the realization $(\delta_{t+1}, \pi_{1:t}, \gamma^1_{1:t}, \tilde{\gamma}^2_{1:t})$ has a non-zero probability. Then, for any Borel subset $A \subset \mathcal{P}\{\mathcal{S}\}$, where $\mathcal{P}\{\mathcal{S}\}$ is the space of probability mass functions over the finite set $\mathcal{S}$ (the space of realization of S_t), use Proposition 2 to write

$$\begin{aligned}\mathbb{P}\left(\Pi_{t+1} \in A \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right) &= \sum_{z_{t+1}} \mathbb{1}_A(F_{t+1}(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, z_{t+1})) \\ &\quad \cdot \mathbb{P}\left(Z_{t+1} = z_{t+1} \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right)\end{aligned} \tag{79}$$

Now, use (20), to obtain

$$\begin{aligned}\mathbb{P}\left(Z_{t+1} = z_{t+1} \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right) &= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}\left(S_t = s_t \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right) \\ &= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t)\end{aligned} \tag{80}$$

where we used the fact that for any realization $(\delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})$ of positive probability, the conditional probability $\mathbb{P}\left(S_t = s_t \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right)$ is same as $\pi_t(s_t)$. Substitute (80) back in (79), to get

$$\begin{aligned}\mathbb{P}\left(\Pi_{t+1} \in A \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right) &= \sum_{z_{t+1}} \sum_{s_t} \mathbb{1}_A(F_{t+1}(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, z_{t+1})) \\ &\quad \cdot \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t) \\ &= \mathbb{P}\left(\Pi_{t+1} \in A \,\middle|\, \pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t\right),\end{aligned} \tag{81}$$

thereby proving (24).

Now, use Proposition 1 to write,

$$\begin{aligned}&\mathbb{E}\left\{\hat{c}_t(S_t, \gamma^1_t, \gamma^2_t, S_{t+1}) \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right\} \\ &= \sum_{s_t, v_t, w^1_{t+1}, w^2_{t+1}} \hat{c}_t(s_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, \hat{f}_{t+1}(s_t, v_t, w^1_{t+1}, w^2_{t+1}, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t)) \\ &\quad \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}\left(W^1_{t+1} = w^1_{t+1}\right) \cdot \mathbb{P}\left(W^2_{t+1} = w^2_{t+1}\right) \\ &\quad \cdot \mathbb{P}\left(S_t = s_t \,\middle|\, \delta_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}\right) \\ &= \sum_{s_t, v_t, w^1_{t+1}, w^2_{t+1}} \hat{c}_t(s_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, \hat{f}_{t+1}(s_t, v_t, w^1_{t+1}, w^2_{t+1}, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t)) \\ &\quad \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}\left(W^1_{t+1} = w^1_{t+1}\right) \\ &\quad \cdot \mathbb{P}\left(W^2_{t+1} = w^2_{t+1}\right) \cdot \pi_t(s_t) \\ &=: C_t(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t).\end{aligned} \tag{82}$$

This proves (25).

Appendix C
Piecewise linearity and concavity of value function

Lemma 1: For any realization $\tilde{\gamma}^{1:2}_t$ of $\gamma^{1:2}_t$, the cost $C_t(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t)$ is linear in π_t.

Proof:

$$\begin{aligned}&C_t(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t) \\ &= \mathbb{E}\left\{\hat{c}_t(S_t, \gamma^1_t, \gamma^2_t, S_{t+1} \,\middle|\, \Pi_t = \pi_t, \gamma^{1:2}_t = \tilde{\gamma}^{1:2}_t\right\} \\ &= \sum \hat{c}_t(s_t, \tilde{\gamma}^1, \tilde{\gamma}^2, \hat{f}_{t+1}(s_t, v_t, w^1_{t+1}, w^2_{t+1}, \tilde{\gamma}^1, \tilde{\gamma}^2)) \\ &\quad \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}\left(W^{1:2}_{t+1} = w^{1:2}_{t+1}\right) \cdot \pi_t(s_t)\end{aligned}$$

where the summation is over all realizations of $(s_t, v_t, w^{1:2}_{t+1})$. Hence $C_t(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t)$ is linear in π_t. ■

We prove the piecewise linearity and concavity of the value function by induction. For $t = T$,

$$J_T(\pi) = \inf_{\tilde{\gamma}^{1:2}_t} C_T(\pi, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t).$$

Lemma 1 implies that $J_T(\pi)$ is the inifimum of finitely many linear functions of π. Thus, $J_T(\pi)$ is piecewise linear and concave in π. This forms the basis of induction. Now assume that $J_{t+1}(\pi)$ is piecewise linear and concave in π. Then, J_{t+1} can be written as the infimum of a finite family I of linear functions as

$$J_{t+1}(\pi) = \inf_{i \in I} \left\{ \sum_{s \in \mathcal{S}} a_i(s) \cdot \pi(s) + b_i \right\}, \tag{83}$$

where $a_i(s)$, b_i, $i \in I$, $s \in \mathcal{S}$ are real numbers. Using this, we will prove that the piecewise linearity and concavity of $J_t(\pi)$.

$$\begin{aligned}J_t(\pi) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \Big[&C_t(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2) \\ &+ \mathbb{E}\left\{J_{t+1}(\Pi_{t+1}) \,\middle|\, \Pi_t = \pi, \gamma^{1:2}_t = \tilde{\gamma}^{1:2}\right\} \Big].\end{aligned} \tag{84}$$

For a particular choice of $\tilde{\gamma}^{1:2}$, we concentrate on the terms inside the square brackets. By Lemma 1 the first term is linear in π. The second term can be written as

$$\mathbb{E}\left\{J_{t+1}(\Pi_{t+1}) \,\middle|\, \Pi_t = \pi, \gamma^{1:2}_t = \tilde{\gamma}^{1:2}\right\} \tag{85}$$

$$\begin{aligned}&= \sum_{z_{t+1}} J_{t+1}(F_{t+1}(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2, z_{t+1})) \\ &\quad \cdot \mathbb{P}\left(Z_{t+1} = z_{t+1} \,\middle|\, \Pi_t = \pi, \gamma^{1:2}_t = \tilde{\gamma}^{1:2}\right) \\ &= \sum_{z_{t+1}} \Big[\inf_{i \in I} \Big\{ \sum_s a_i(s) \cdot (F_{t+1}(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2, z_{t+1}))(s) + b_i \Big\} \\ &\quad \cdot \mathbb{P}\left(Z_{t+1} = z_{t+1} \,\middle|\, \Pi_t = \pi, \gamma^{1:2}_t = \tilde{\gamma}^{1:2}\right) \Big]\end{aligned} \tag{86}$$

where the last expression follows from (83).

Note that

$$\mathbb{P}\left(Z_{t+1} = z_{t+1} \,\middle|\, \Pi_t = \pi, \gamma^{1:2}_t = \tilde{\gamma}^{1:2}\right) = \sum_{s' \in \mathcal{S}} \mathbb{1}_{\hat{h}_t(s')}(z_{t+1}) \cdot \pi(s') \tag{87}$$

Focus on each term in the outer summation in (86). For each value of z_{t+1}, these terms can be written as:

$$\begin{aligned}\inf_{i \in I} \Big\{ &\sum_s a_i(s) \cdot (F_{t+1}(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2, z_{t+1}))(s) \\ &\cdot \sum_{s' \in \mathcal{S}} \mathbb{1}_{\hat{h}_t(s')}(z_{t+1}) \cdot \pi(s') \\ &+ b_i \cdot \sum_{s' \in \mathcal{S}} \mathbb{1}_{\hat{h}_t(s')}(z_{t+1}) \cdot \pi(s')\end{aligned} \tag{88}$$

The second summand is linear in π. Using the characterization of F_{t+1} from the proof of Proposition 2 (Appendix A), we can write the first summand as

$$a_i(s) \cdot \Bigg\{ \sum_{s_t, v_t, w^1_{t+1}, w^2_{t+1}} \mathbb{1}_s(\hat{f}_{t+1}(s_t, v_t, w^1_{t+1}, w^2_{t+1}, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t)) \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}\left(W^{1:2}_{t+1} = w^{1:2}_{t+1}\right) \cdot \mathbb{1}_{\hat{h}(s_t)}(z_{t+1}) \pi(s_t) \Bigg\} \tag{89}$$

which is also linear in π. Substituting (88) and (89) in (86), we get that for a given choice of $\tilde{\gamma}^1, \tilde{\gamma}^2$, the second expectation in (84) is concave in π. Thus, the value function $J_t(\pi)$ is the minimum of finitely many functions each of which is linear in π. This implies that J_t is piecewise linear and concave in π. This completes the induction argument.

Appendix D
Proof of Proposition 4

We prove the three parts separately.

Part 1)

We first prove that Θ_{t+1} is a function of Θ_t and Z_{t+1}. Recall that $Z_{t+1} = (Y^1_{t-n+1}, Y^2_{t-n+1}, U^1_{t-n+1}, U^2_{t-n+1})$ and $\Delta_{t+1} = (\Delta_t, Z_{t+1})$. Fix a coordination strategy $\boldsymbol{\psi}$ and consider a realization δ_{t+1} of Δ_{t+1}. Then,

$$\begin{aligned}
&\theta_{t+1}(x_{t-n+1}) \\
&:= \mathbb{P}(X_{t-n+1} = x_{t-n+1} | \delta_{t+1}) \\
&= \mathbb{P}(X_{t-n+1} = x_{t-n+1} | \delta_t, y^{1:2}_{t-n+1}, y^2_{t-n+1}, u^{1:2}_{t-n+1}) \\
&= \sum_{x \in \mathcal{X}} \mathbb{P}(X_{t-n+1} = x_{t-n+1} | X_{t-n} = x, u^{1:2}_{t-n+1}) \\
&\quad \cdot \mathbb{P}(X_{t-n} = x | \delta_t, y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1}) \\
&= \sum_{x \in \mathcal{X}} \mathbb{P}(X_{t-n+1} = x_{t-n+1} | X_{t-n} = x, u^{1:2}_{t-n+1}) \\
&\quad \cdot \frac{\mathbb{P}(X_{t-n} = x, y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1} | \delta_t)}{\sum_{x'} \mathbb{P}(X_{t-n} = x', y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1} | \delta_t)}
\end{aligned} \tag{90}$$

Consider the second term of (90), and note that under any coordination strategy $\boldsymbol{\psi}$, the variables $u^{1:2}_{t-n+1}$ are deterministic functions of $y^{1:2}_{t-n+1}$ and δ_t (which is same as $y^{1:2}_{1:t-n}, u^{1:2}_{1:t-n}$). Therefore, the numerator of the second term of (90) can be written as

$$\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}\left(x_{t-n}, y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1} \,\middle|\, \delta_t\right) \\
&= \mathbb{P}^{\boldsymbol{\psi}}\left(u^{1:2}_{t-n+1} \,\middle|\, x_{t-n}, y^{1:2}_{t-n+1}, \delta_t\right) \\
&\quad \cdot \mathbb{P}^{\boldsymbol{\psi}}\left(y^{1:2}_{t-n+1} \,\middle|\, x_{t-n}, \delta_t\right) \cdot \mathbb{P}^{\boldsymbol{\psi}}\left(x_{t-n} \,\middle|\, \delta_t\right) \\
&= \mathbb{P}^{\boldsymbol{\psi}}\left(u^{1:2}_{t-n+1} \,\middle|\, y^{1:2}_{t-n+1}, \delta_t\right) \\
&\quad \cdot \mathbb{P}\left(y^{1:2}_{t-n+1} \,\middle|\, x_{t-n}\right) \cdot \theta_t(x_{t-n})
\end{aligned} \tag{91}$$

Substitute (91) in (90) and cancel $\mathbb{P}^{\boldsymbol{\psi}}\left(u^{1:2}_{t-n+1} \,\middle|\, y^{1:2}_{t-n+1}, \delta_t\right)$ from the numerator and denominator. Thus, θ_{t+1} is a function of θ_t and z_{t+1}.

Next we prove that r^k_{t+1} is a function of r^k_t, Z_{t+1} and γ^k_t. Recall that

$$r^k_{t+1} := (r^k_{m,(t+1)}, t-n+2 \le m \le t).$$

We prove the result by showing that each component $r^k_{m,(t+1)}$, $t-n+2 \le m \le t$ is a function of r^k_t, Z_{t+1} and γ^k_t.

1) For $m = t$, we have $r^k_{t,(t+1)} := \gamma^k_t(\cdot, Y^k_{t-n+1})$. Since Y^k_{t-n+1} is a part of Z_{t+1}, $r^k_{t,(t+1)}$ is a function of γ^k_t and Z_{t+1}.
2) For $m = t-n+2, t-n+3, \ldots, t-1$,

$$\begin{aligned}
r^k_{m,t+1}(\cdot) &:= \gamma^k_m(\cdot, Y^k_{m-n+1:t+1-n}, U^k_{m-n+1:t+1-n}) \\
&= \gamma^k_m(\cdot, Y^k_{t-n+1}, U^k_{t-n+1}, Y^k_{m-n+1:t-n}, U^k_{m-n+1:t-n}) \\
&= r^k_{m,t}(\cdot, Y^k_{t-n+1}, U^k_{t-n+1})
\end{aligned} \tag{92}$$

Thus, for $m = t-n+2, t-n+3, \ldots, t-1$, $r^k_{m,t+1}$ is a function of $r^k_{m,t}$ and Z_{t+1}.

Part 2)

First, let us assume that the coordinator's belief Π_t defined in (22) is a function of (Θ_t, r^1_t, r^2_t), that is, there exist functions H_t, for $t = 1, 2, \ldots, T$, such that

$$\Pi_t = H_t(\Theta_t, r^1_t, r^2_t) \tag{93}$$

From (25) of Proposition 3, we have that

$$\begin{aligned}
&\mathbb{E}\left\{\hat{c}_t(S_t, \gamma^1_t, \gamma^2_t, S_{t+1}) \,\middle|\, \Delta_t, \Pi_{1:t}, \gamma^1_{1:t}, \gamma^2_{1:t}\right\} \\
&= C_t(\Pi_t, \gamma^1_1, \gamma^2_t) \\
&= \hat{C}_t(\Theta_t, r^1_t, r^2_t, \gamma^1_1, \gamma^2_t)
\end{aligned} \tag{94}$$

where the last equation uses (93). Thus, to prove this part of the proposition, only need to prove (93). For that matter, we need the following lemma.

Lemma 2: $S_t := (X_{t-1}, \Lambda^1_t, \Lambda^2_t)$ is a deterministic function of $(X_{t-n}, V_{t-n+1:t-1}, W^1_{t-n+1:t}, W^2_{t-n+1:t}, r^1_t, r^2_t)$, that is, there exists a fixed deterministic function D_t such that

$$\begin{aligned}
S_t &:= (X_{t-1}, \Lambda^1_t, \Lambda^2_t) \\
&= D_t(X_{t-n}, V_{t-n+1:t-1}, W^{1:2}_{t-n+1:t}, r^{1:2}_t)
\end{aligned} \tag{95}$$

Proof: We first prove a slightly weaker result: for $t-n+1 \le m \le t-1$, there exists a deterministic function $\hat{D}_{m,t}$ such that

$$\begin{aligned}
&(X_{t-n+1:m}, Y^{1:2}_{t-n+1:m}, U^{1:2}_{t-n+1:m}) \\
&= \hat{D}_{m,t}(X_{t-n}, V_{t-n+1:m}, W^{1:2}_{t-n+1:m}, r^{1:2}_{t-n+1:m,t})
\end{aligned} \tag{96}$$

using induction. First consider $m = t-n+1$. For this case, the LHS of (96) equals $(X_{t-n+1}, Y^{1:2}_{t-n+1}, U^{1:2}_{t-n+1})$. For $k = 1, 2$,

$$\begin{aligned}
Y^k_{t-n+1} &= h^k_{t-n+1}(X_{t-n}, W^k_{t-n+1}) \\
U^k_{t-n+1} &= r^k_{t-n+1,t}(Y^k_{t-n+1})
\end{aligned}$$

Furthermore, by the system dynamics,

$$X_{t-n+1} = f_t(X_{t-n}, U^{1:2}_{t-n+1}, V_{t-n+1})$$

Thus $(X_{t-n+1}, Y^{1:2}_{t-n+1}, U^{1:2}_{t-n+1})$ is a deterministic function of $(X_{t-n}, W^{1:2}_{t-n+1}, V_{t-n+1}, r^{1:2}_{t-n+1,t})$. This proves (96) for $m = t-n+1$. Now assume that (96) is true for some m, $t-n+1 \le m < t-1$. We show that this implies that (96) is also true for $m+1$. For $k = 1, 2$,

$$\begin{aligned}
Y^k_{m+1} &= h^k_{m+1}(X_m, W^k_{m+1}) \\
U^k_{m+1} &= r^k_{m+1,t}(Y^k_{t-n+1:m+1}, U^k_{t-n+1:m})
\end{aligned}$$

Furthermore, by the system dynamics,

$$X_{m+1} = f_t(X_m, U^{1:2}_{m+1}, V_{m+1})$$

Thus, $(X_{m+1}, Y^{1:2}_{m+1}, U^{1:2}_{m+1})$ is a deterministic function of

$$(X_m, Y^{1:2}_{t-n+1:m}, U^{1:2}_{t-n+1:m}, W^{1:2}_{m+1}, V_{m+1}, r^{1:2}_{m+1,t})$$

Combining this with the induction hypothesis, we conclude that $(X_{t-n+1:m+1}, Y^{1:2}_{t-n+1:m+1}, U^{1:2}_{t-n+1:m+1})$ is a function of $(X_{t-n}, W^{1:2}_{t-n+1:m+1}, V_{t-n+1:m+1}, r^{1:2}_{t-n+1:m+1,t})$. Thus, by induction (96) is true for $t-n+1 \le m \le t-1$.

Now we use (96) to prove the lemma. For $k = 1, 2$

$$Y^k_t = h^k_t(X_{t-1}, W^k_t)$$
$$r^k_t = r^k_{t-n+1:t-1,t}$$

Combining this with (96) for $m = t-1$ implies that there exists a deterministic function $\hat{D}_t$ such that

$$(X_{t-n+1:t-1}, Y^{1:2}_{t-n+1:t}, U^{1:2}_{t-n+1:t-1}) = \hat{D}_t(X_{t-n}, V_{t-n+1:t-1}, W^{1:2}_{t-n+1:t}, r^{1:2}_t)$$

This implies that there exists a function D_t such that Lemma 2 is true. ∎

Now consider

$$\begin{aligned}\Pi_t(s_t) &:= \mathbb{P}^\psi\left(S_t = s_t \,\middle|\, \Delta_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1}\right) \\ &= \sum \mathbb{1}_{s_t}\{D_t(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^{1:2}_t\} \\ &\quad \cdot \mathbb{P}\left(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^{1:2}_t \,\middle|\, \Delta_t, \gamma^{1:2}_{1:t-1}\right)\end{aligned} \tag{97}$$

where the summation is over all choices of $(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^{1:2}_t)$. The vectors $r^{1:2}_t$ are completely determined by Δ_t and $\gamma^{1:2}_{1:t-1}$; the noise random variables $v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}$ are independent of the conditioning terms and X_{t-n}. Therefore, we can write (97) as

$$\begin{aligned}&\sum \mathbb{1}_{s_t}\{D_t(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^1_t, \tilde{r}^2_t)\} \\ &\quad \cdot \mathbb{P}\left(v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}\right) \cdot \mathbb{1}_{\tilde{r}^1_t, \tilde{r}^2_t}(r^1_t, r^2_t) \\ &\quad \cdot \mathbb{P}\left(x_{t-n} \,\middle|\, \Delta_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1}\right)\end{aligned} \tag{98}$$

In the last term of (98), we dropped $\gamma^{1:2}_{1:t-1}$ from the conditioning terms because they are functions of Δ_t. The last term is therefore same as $\mathbb{P}(x_{t-n} \,|\, \Delta_t) = \Theta_t$. Thus, Π_t is a function of Θ_t and r^1_t, r^2_t, thereby proving (93).

Part 3)

Consider the LHS of (45)

$$\begin{aligned}&\mathbb{P}\left(\Theta_{t+1} = \theta_{t+1}, r^{1:2}_{t+1} = \tilde{r}^{1:2}_{t+1}, \,\middle|\, \delta_t, \theta_{1:t}, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^{1:2}_{1:t}\right) \\ &= \sum_{z_{t+1}} \mathbb{1}_{\theta_{t+1}}(Q_{t+1}(\theta_t, z_{t+1})) \cdot \mathbb{1}_{\tilde{r}^1_{t+1}}(Q^1_{t+1}(\tilde{r}^1_t, \tilde{\gamma}^1_t, z_{t+1})) \\ &\quad \cdot \mathbb{1}_{\tilde{r}^2_{t+1}}(Q^2_{t+1}(\tilde{r}^2_t, \tilde{\gamma}^2_t, z_{t+1})) \\ &\quad \cdot \mathbb{P}\left(Z_{t+1} = z_{t+1} \,\middle|\, \delta_t, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^1_{1:t}, \tilde{r}^2_{1:t}\right)\end{aligned} \tag{99}$$

The last term of (99) can be written as

$$\begin{aligned}&\mathbb{P}\left(Z_{t+1} = z_{t+1} \,\middle|\, \delta_t, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^1_{1:t}, \tilde{r}^2_{1:t}\right) \\ &= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}\left(S_t = s_t \,\middle|\, \delta_t, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^1_{1:t}, \tilde{r}^2_{1:t}\right) \\ &= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}\left(S_t = s_t \,\middle|\, \delta_t\right) \\ &= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t) \\ &= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot H_t(\theta_t, \tilde{r}^1_t, \tilde{r}^2_t)(s_t)\end{aligned} \tag{100}$$

Substituting (100) back in (99), we get

$$\begin{aligned}&\mathbb{P}\left(\Theta_{t+1} = \theta_{t+1}, r^{1:2}_{t+1} = \tilde{r}^{1:2}_{t+1} \,\middle|\, \delta_t, \theta_{1:t}, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^{1:2}_{1:t}\right) \\ &= \sum_{z_{t+1}, s_t} \mathbb{1}_{\theta_{t+1}}(Q_{t+1}(\theta_t, z_{t+1})) \\ &\quad \cdot \mathbb{1}_{\tilde{r}^1_{t+1}}(Q^1_{t+1}(\tilde{r}^1_t, \tilde{\gamma}^1_t, z_{t+1})) \\ &\quad \cdot \mathbb{1}_{\tilde{r}^2_{t+1}}(Q^2_{t+1}(\tilde{r}^2_t, \tilde{\gamma}^2_t, z_{t+1})) \\ &\quad \cdot \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot H_t(\theta_t, \tilde{r}^1_t, \tilde{r}^2_t)(s_t) \\ &= \mathbb{P}\left(\Theta_{t+1} = \theta_{t+1}, r^{1:2}_{t+1} = \tilde{r}^{1:2}_{t+1} \,\middle|\, \theta_t, \tilde{r}^{1:2}_t, \tilde{\gamma}^{1:2}_t\right)\end{aligned} \tag{101}$$

thereby proving (45).

References

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

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

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

[4] B. Kurtaran, “Corrections and extensions to ”decentralized stochastic control with delayed sharing information pattern”,” *IEEE Trans. Autom. Control*, vol. 24, no. 4, pp. 656–657, Aug. 1979.

[5] H. S. Witsenhausen, “Some remarks on the concept of state,” in *Directions in Large-Scale Systems*, Y. C. Ho and S. K. Mitter, Eds. Plenum, 1976, pp. 69–75.

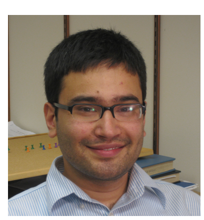
[6] H. Zhang, “Partially observable markov decision processes: A geometric technique and analysis,” *Operations Research*, 2009.

[7] 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.

[8] B. Kurtaran, “Decentralized stochastic control with delayed sharing information pattern,” *IEEE Trans. Autom. Control*, vol. 21, pp. 576–581, Aug. 1976.

[9] A. Nayyar and D. Teneketzis, “On the structure of real-time encoders and decoders in a multi-terminal communication system,” *IEEE Trans. Info. Theory*, 2009, submitted.

[10] 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.



Ashutosh Nayyar received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, Delhi, India in 2006 and the M.S. in Electrical Engineering and Computer Science from the University of Michigan, Ann Arbor in 2008. He is currently a Ph.D. student in Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. His research interests include decentralized stochastic control, stochastic scheduling and resource allocation, game theory and mechanism design.

Aditya Mahajan (S'06-M'09) received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India in 2003 and the MS and PhD degrees in Electrical Engineering and Computer Science from the University of Michigan, Ann Arbor, MI, USA in 2006 and 2008.

He is currently an Assistant Professor of Electrical and Computer Engineering, McGill University, QC, Canada. From 2008 to 2010, he was a Postdoctoral Researcher in the department of Electrical Engineering at Yale University, New Haven, CT. His research interests include decentralized stochastic control, team theory, real-time communication, information theory, and discrete event systems.

Demosthenis Teneketzis (M'87-SM'97-F'00) received the diploma in Electrical Engineering from the University of Patras, Patras, Greece, and the M.S., E.E., and Ph.D. degrees, all in Electrical Engineering, from the Massachusetts Institute of Technology, Cambridge, in 1974, 1976, 1977, and 1979, respectively.

He is currently Professor of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. In winter and spring 1992, he was a Visiting Professor at the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. Prior to joining the University of Michigan, he worked for Systems Control, Inc., Palo Alto, CA, and Alphatech, Inc., Burlington, MA. His research interests are in stochastic control, decentralized systems, queueing and communication networks, stochastic scheduling and resource allocation problems, mathematical economics, and discrete-event systems.