2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

An Explicit Solution for Optimal Two-Player Decentralized Control over TCP Erasure Channels with State Feedback

Chung-Ching Chang[1] Sanjay Lall[2]

Abstract

We develop an optimal controller synthesis algorithm for decentralized control problems where control actions are transmitted through TCP-like erasure channels. We consider a simple two-player interconnected linear system and Bernoulli distributed erasure channels. We recast the problem to a centralized Partially Observed Markov Decision Process (POMDP) under the fictitious player framework, in which we construct an optimal controller using belief states and value function recursions. Finally, we provide explicit state space formulae for the optimal decentralized controller.

1 Introduction

Decentralized control arises in a variety of engineering branches, such as communication systems, sensor networks, vehicle coordinations, and flight formations. In decentralized control, multiple controllers are cooperatively actuating a system to minimize a certain cost. Contrary to centralized control, where only one controller takes all measurements and decides all actions to actuate a system, in decentralized control, each controller measures only partial information and decides partial actions. Furthermore, because these subsystems are connected over networks, it is critical to resolve issues of communication delay [1], data loss [2], and synchronization. In this paper, we focus on data loss.

We consider the case when the control communication channels of a interconnected system are vulnerable to some Bernoulli distributed packet drops. At every time instance, control actions generated by controllers are sent to subsystems through control communication channels suffering from packet drops. We consider a TCP-like protocol, where link conditions are known to the controllers upon the next observation through acknowledgements. This problem was addressed by [2] who proposed an explicit state space solution to a centralized TCP-like LQG problem

[1]Chung-Ching Chang is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. bobbyc@stanford.edu

[2]S. Lall is with the Department of Electrical Engineering and Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA. lall@stanford.edu

The design of a decentralized controller is dramatically different from that of a centralized controller. Consider a general linear dynamical system such as Linear Quadratic Gaussian (LQG) control, which is known to have linear solutions for centralized controllers. However, in the decentralized case, the problem is generally nonlinear or intractable [9], unless there are some particular information structures such as partial nestedness [3] or quadratic invariance [6]. The problem under consideration in this paper is neither partially nested nor quadratically invariant, nor is the solution linear.

Without packet drops, our problem has been solved in [8] using spectral factorization and in [7] using dynamic programming. In this paper, we take an alternative approach to [7] in dynamic programming by using the fictitious player framework. Instead of solving the decentralized problem directly, we first recast the problem to a centralized problem from the perspective of a fictitious player as suggested by Mahajan et al. [5], and simplify it according to [10]. Given that the problem becomes centralized, we can utilize the standard Markov decision theory [4] to solve the problem using belief states and value function recursions. Finally, we derive the corresponding state space solution for the original decentralized problem.

2 Problem Formulation

We consider two interconnected subsystems where the dynamics of subsystem 1 may affect the dynamics of subsystems 2, but not vice versa. The system dynamics are as follows:

$$\begin{bmatrix} z_{t+1}^1 \\ z_{t+1}^2 \end{bmatrix} = \begin{bmatrix} A^{11} & 0 \\ A^{21} & A^{22} \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} + \begin{bmatrix} B^{11} & 0 \\ B^{21} & B^{22} \end{bmatrix} \begin{bmatrix} N_t^1 & 0 \\ 0 & N_t^2 \end{bmatrix} \begin{bmatrix} u_t^1 \\ u_t^2 \end{bmatrix} + \begin{bmatrix} v_t^1 \\ v_t^2 \end{bmatrix} \quad (1)$$

for $t = 0, 1, \ldots, T-1$. Let $\mathbb{R}$ denote real numbers and $\mathbb{E}$ denote expectations. For all $i \in \{1, 2\}$, $z_t^i \in \mathbb{R}^{\eta_i}$ is the state of the subsystem i, $u_t^i \in \mathbb{R}^{\lambda_i}$ is the action of player i, and $v_t^i \sim N(0, \Sigma_v^i)$ is independent Gaussian noise. The initial condition $z_0 \sim N(\mu_s, \Sigma_s)$ is independent of noises. We focus on the case when $\mu_s = 0$ and $\Sigma_s = \begin{bmatrix} \Sigma_s^1 & \\ & \Sigma_s^2 \end{bmatrix}$. The link condition of the actuator channel i is modeled by $N_t^i = \mathrm{diag}(n_t^i)$, where $n_t^i = (n_t^{i1}, n_t^{i2}, \ldots, n_t^{i\lambda_i})$.

$n_t^{ij} \sim \text{Bernoulli}(\bar{n}^{ij})$ denotes the Independent and Identically distributed (I.I.D.) Bernoulli random binary variable modeling the random information drop on j-th actuator of channel i at time t. The average transmission successful rate on the actuator channel of player i is therefore $\bar{N}^i = \mathbb{E}N_t^i = \text{diag}(\mathbb{E}(n_t^i))$.

For convenience, we let

$$z_t = \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix}, \quad u_t = \begin{bmatrix} u_t^1 \\ u_t^2 \end{bmatrix}, \quad v_t = \begin{bmatrix} v_t^1 \\ v_t^2 \end{bmatrix},$$
$$A = \begin{bmatrix} A^{11} & 0 \\ A^{21} & A^{22} \end{bmatrix}, \quad B = \begin{bmatrix} B^{11} & 0 \\ B^{21} & B^{22} \end{bmatrix},$$
$$N_t = \begin{bmatrix} N_t^1 & 0 \\ 0 & N_t^2 \end{bmatrix}, \quad \bar{N} = \begin{bmatrix} \bar{N}^1 & 0 \\ 0 & \bar{N}^2 \end{bmatrix}.$$

We use the notation $x_{0:t}^1 = (x_0^1, \ldots, x_t^1)$ to refer to the list of variables corresponding to the subsystem 1 from time 0 to t. Similarly, $x_{0:t}^{1:2} = (x_0^1, \ldots, x_t^1, x_0^2, \ldots, x_t^2)$.

The objective of the problem is to minimize $\mathcal{J}(K) =$

$$\min_K \mathbb{E}\left\{ \sum_{t=0}^{T-1} \left(z_t^{\mathsf{T}} Q z_t + u_t^{\mathsf{T}} N_t R N_t u_t \right) + z_T^{\mathsf{T}} Q_T z_T \right\}, \quad (2)$$

for some $Q \geq 0$, $Q_T \geq 0$, and $R > 0$. $K = (g^1, g^2)$ with $g^1 = (g_{0:T-1}^1)$ and $g^2 = (g_{0:T-1}^2)$ is a set of the control policies of player 1 and player 2 such that the control actions are defined by

$$\begin{aligned} u_t^1 &= g_t^1(z_{0:t}^1, u_{0:t-1}^1, n_{0:t-1}^1), \\ u_t^2 &= g_t^2(z_{0:t}, u_{0:t-1}^2, n_{0:t-1}). \end{aligned} \quad (3)$$

That is, player 1 observes only current and past states of subsystem 1, all his past actions, and all past channel 1 link conditions, while player 2 observes current and past states of both subsystems, his past actions, and the past link conditions of both channels.

3 Main Results

The main theorem of this paper, which gives a state space solution for the optimal controller (3) that minimizes the objective (2) for the system (1).

Theorem 1. *Let $P_t \in \mathbb{R}^{(\eta_1+\eta_2)\times(\eta_1+\eta_2)}$, $Y_t \in \mathbb{R}^{\eta_2 \times \eta_2}$, and $r_t \in \mathbb{R}$ satisfy the following recursions*

$$\begin{aligned} P_t &= Q + A^{\mathsf{T}} P_{t+1} A - A^{\mathsf{T}} P_{t+1} B \bar{N} \\ &\quad \times \left(\mathbb{E}(N_t (R + B^{\mathsf{T}} P_{t+1} B) N_t) \right)^{-1} \bar{N} B^{\mathsf{T}} P_{t+1} A, \quad (4) \\ Y_t &= Q^{22} + A^{22\mathsf{T}} Y_{t+1} A^{22} \\ &\quad - A^{22\mathsf{T}} Y_{t+1} B^{22} \bar{N}^2 \left(\mathbb{E}(N_t^2 (R^{22} + B^{22,\mathsf{T}} Y_{t+1} B^{22}) N_t^2) \right)^{-1} \\ &\quad \times \bar{N}^2 B^{22\mathsf{T}} Y_{t+1} A^{22}, \quad (5) \\ r_t &= r_{t+1} + \text{trace}\left(\begin{bmatrix} P_{t+1}^{11} & P_{t+1}^{12} \\ P_{t+1}^{21} & Y_{t+1} \end{bmatrix} \Sigma_v \right), \quad (6) \end{aligned}$$

with $P_T = Q_T$, $Y_T = P_T^{22}$, and $r_T = 0$. Define J_t and K_t to be

$$\begin{aligned} K_t &= \left(\mathbb{E}(N_t (R + B^{\mathsf{T}} P_{t+1} B) N_t) \right)^{-1} \bar{N} B^{\mathsf{T}} P_{t+1} A, \quad (7) \\ J_t &= \left(\mathbb{E}(N_t^2 (R^{22} + B^{22,\mathsf{T}} Y_{t+1} B^{22}) N_t^2) \right)^{-1} \\ &\quad \times \bar{N}^2 B^{22\mathsf{T}} Y_{t+1} A^{22}. \quad (8) \end{aligned}$$

Let

$$\begin{aligned} A_t^K(N_t^1) &= A^{22} - B^{21} N_t^1 K_t^{12} - B^{22} \bar{N}^2 K_t^{22}, \quad (9) \\ B_t^K(N_t^1) &= A^{21} - B^{21} N_t^1 K_t^{11} - B^{22} \bar{N}^2 K_t^{21}, \quad (10) \end{aligned}$$

where N_t^1 is known to controllers from observations. The optimal controllers are

- *Controller 1 has realization*

$$\begin{aligned} \xi_{t+1} &= A_t^K(N_t^1)\xi_t + B_t^K(N_t^1) z_t^1 \\ u_t^1 &= -K_t^{11} z_t^1 - K_t^{12} \xi_t \end{aligned}$$

- *Controller 2 has realization*

$$\begin{aligned} \xi_{t+1} &= A_t^K(N_t^1)\xi_t + B_t^K(N_t^1) z_t^1 \\ u_t^1 &= -K_t^{21} z_t^1 - K_t^{22} \xi_t - J(z_t^2 - \xi_t) \end{aligned}$$

where $\xi_0 = 0$. The optimal cost $\min_K \mathcal{J}(K)$ is

$$\sum_{t=1}^{T} \text{trace}\left(\begin{bmatrix} P_t^{11} & P_t^{12} \\ P_t^{21} & Y_t \end{bmatrix} \Sigma_v \right) + \text{trace}\left(\begin{bmatrix} P_0^{11} & P_0^{12} \\ P_0^{21} & Y_0 \end{bmatrix} \Sigma_s \right).$$

We will develop the proof in the following sections. The philosophy behind the proof is to first recast the two player decentralized problem described in Section 2 into a centralized POMDP using the idea in Section 4.2. With the problem being centralized, we are able to solve the problem through the value function recursion as in Lemma 2. In calculating the cost function and optimal policy in the fictitious player problem, we solve the state space formulae and the optimal cost of the original decentralized problem.

4 Preliminaries

In this paper, we solve the decentralized two-player TCP-like LQG problem with state feedback through the fictitious player framework in [5]. In order to explain our results, we briefly present the idea of the fictitious player framework (Model A) in [5] in Section 4.2. We simplified the idea for the case of only two players with state feedback according to Theorem 1 in [10].

4.1 Notations

In most cases, we use subscript to denote time index and superscript $i \in \{1, 2\}$ to denote subsystems. For example, x_t^i, u_t^i, and y_t^i denote the state, action, and observation of subsystem i at time t, respectively. Conventionally, we let y_t^0 denote the common observation at time

t. For other cases, we use superscript $k \in \{1,2,3\}$ to denote the sub-time index in the fictitious player framework. For example, s_t^k denotes the fictitious state at time t^k. Conventionally, we use dummy superscript i to denote the subsystem index and k to denote the sub-time index.

We use ϕ to denote the empty set and calligraphy capital letters, like $\mathcal{X}$ and $\mathcal{S}_t^0$, to denote sets. In POMDP, we use the Sans-serif font to denote sample paths; For example, we take x_t as any sample path of x_t. We use $f_\Sigma(x-\mu)$ to denote the Probability Density Function (PDF) for the Gaussian random vector x with mean μ and covariance Σ. For a set $\mathcal{Y}^1$, although the superscript denotes the subsystem, we use the superscript over the parenthesis $(\mathcal{Y}^1)^n$ or $\prod_{t=0}^{n-1}(\mathcal{Y}^1)$ to denote the n-fold Cartesian product of the set, that is, $(\mathcal{Y}^1)^n = \mathcal{Y}^1 \times \cdots \times \mathcal{Y}^1$ n times. The notations $(a_1, \ldots, a_t)$ and $\begin{bmatrix} a_1^\mathsf{T} & \ldots & a_t^\mathsf{T} \end{bmatrix}^\mathsf{T}$ are interchangeable throughout the text. Finally, let

$$\delta(a,b) = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{elsewhere.} \end{cases}$$

and $\Gamma(a,b,c) = \delta(a,b(c))$ when b is a map. When $a = (a^1, \ldots, a^n)$ and $b = (b^1, \ldots, b^n)$ are two sets of equal size, we overload the function $\delta(a,b)$ with $\delta(a,b) = \delta(a^1,b^1)\cdots\delta(a^n,b^n)$. Conventionally, we define $\delta(\phi,\phi) = 1$. Furthermore, we use da to denote $da_1 \cdots da_n$ compactly in an integral.

4.2 The Fictitious Player Framework

Consider a discrete time system consists of a plant and two players. Let $x_t \in \mathcal{X}$ denote the state of the plant and $u_t^i \in \mathcal{U}^i$ denote the control action of player i. The plant follows the dynamic

$$x_{t+1} = l_t(x_t, u_t^1, u_t^2, w_t^3),$$

where $l_t(\cdot)$ is the plant function and $w_t^3 \in \mathcal{W}^3$ the process noise. Let $y_t^0 \in \mathcal{Y}^0$ denote the common observation and $y_t^i \in \mathcal{Y}^i$ the private observations.

The observations are generated according to

$$y_t^i = h_t^i(x_t, w_t^i) \quad \forall i \in \{0,1,2\},$$

where $h_t^i(\cdot)$ is the observation function and $w_t^i \in \mathcal{W}^i$ is the observation noise. Let $m_t^i = (y_{0:t}^i, u_{0:t}^i) \in \mathcal{M}_t^i$ denote the set of private memory for player i, where $\mathcal{M}_t^i = (\mathcal{Y}^i)^{t+1} \times (\mathcal{U}^i)^{t+1}$. Conventionally, we take $m_{-1}^i = \phi$.

At each time t, after player i observes common observation y_t^0 and private observation y_t^i, he generates control action u_t^i and updates his private memory according to

$$\begin{aligned} u_t^i &= g_t^i(m_{t-1}^i, y_t^i, y_{0:t}^0), \\ m_t^i &= (y_{0:t}^i, u_{0:t}^i) = (m_{t-1}^i, y_t^i, g_t^i(m_{t-1}^i, y_t^i, y_{0:t}^0)), \end{aligned}$$

where $g_t^i \in \mathcal{G}_t^i$ is any control policy.

The objective is to select a set of control policies $K = (g^1, g^2)$ with $g^i = (g_{0:T-1}^i)$ such that it minimizes the finite horizon cost $\mathcal{J}(K) = \mathbb{E}\left(\sum_{t=0}^T \rho_t(x_t, u_t^1, u_t^2)\right)$. This problem is a non-classical POMDP since the observations are different for each player.

Mahajan et al. suggested a framework to transform a decentralized control problem into a centralized control problem from the perspective of a fictitious player, with respect to a fictitious plant, through sequential decomposition in [5]. Consider a fictitious player who observes common observations and determines maps $\bar{g}_t^i \in \bar{\mathcal{G}}_t^i$ for each player i such that $\bar{g}_t^i(m_{t-1}^i, y_t^i) = g_t^i(m_{t-1}^i, y_t^i, y_{0:t}^0)$. Each player i then generates his private action with $u_t^i = \bar{g}_t^i(m_{t-1}^i, y_t^i)$ and updates his private memory upon receiving private observation y_t^i.

The controller is centralized from the perspective of the fictitious player. We now reformulate the (fictitious) plant from the perspective of the fictitious player, where all real players are part of the fictitious plant. Let s_t^k be the state of the fictitious plant, where

$$\begin{aligned} s_t^1 &= (x_t, m_{t-1}^1, m_{t-1}^2) \in \mathcal{S}_t^1, \\ s_t^2 &= (x_t, m_t^1, m_{t-1}^2) \in \mathcal{S}_t^2, \\ s_t^3 &= (x_t, m_t^1, m_t^2) \in \mathcal{S}_t^3, \end{aligned} \tag{11}$$

and $\mathcal{S}_t^1 = \mathcal{X} \times \mathcal{M}_{t-1}^1 \times \mathcal{M}_{t-1}^2$, $\mathcal{S}_t^2 = \mathcal{X} \times \mathcal{M}_t^1 \times \mathcal{M}_{t-1}^2$, and $\mathcal{S}_t^3 = \mathcal{X} \times \mathcal{M}_t^1 \times \mathcal{M}_t^2$. According to the sequential decomposition, one time step t is decomposed into several subtime steps t^k with $k \in \{1,2,3\}$, and the states evolve in the order of $s_t^1, s_t^2, s_t^3, s_{t+1}^1$ and so on. At time t^1, the state evolves according to $x_t = l_{t-1}(x_{t-1}, u_{t-1}^1, u_{t-1}^2, w_{t-1}^3)$, and the fictitious player measures a new common observation y_t^0. At time t^k, the fictitious player determines optimal map $\bar{g}_t^k$ to be his control action. At time t^3, cost $\bar{\rho}_t^3(s_t^3) = \rho_t(x_t, u_t^1, u_t^2)$ is incurred. The objective of the fictitious player problem is to find optimal policy $\bar{K} = (\bar{g}^1, \bar{g}^2)$ with $\bar{g}^i = (\bar{g}_{0:T-1}^i)$ such that $\bar{\mathcal{J}}(\bar{K}) = \mathbb{E}\left(\sum_{t=0}^T \bar{\rho}_t(s_t^3)\right)$ is minimized.

The fictitious player framework is, in fact, a centralized POMDP problem. From the perspective of the fictitious player, he measures common observations and determines control actions $\bar{g}_t^k$ without further constraints. This centralized framework allows us to define belief states and value function recursions as a classical POMDP. The information state given to the fictitious player at time t^k is the following:

$$I_t^k = \begin{cases} (y_{0:t}^0, \bar{g}_{0:t-1}^1, \bar{g}_{0:t-1}^2) & k = 1, \\ (y_{0:t}^0, \bar{g}_{0:t}^1, \bar{g}_{0:t-1}^2) & k = 2, \\ (y_{0:t}^0, \bar{g}_{0:t}^1, \bar{g}_{0:t}^2) & k = 3, \end{cases} \tag{12}$$

and the belief state is $\pi_t^k = \Pr(s_t^k | I_t^k)$. Furthermore, we have the value function recursions as follows:

Lemma 2. *The Value function recursions for the centralized fictitious player problem of the two-player decentralized state feedback problem are as follows:*

$$V_{T+1}^1(\tilde{\pi}_{T+1}^1) = 0,$$

and for $t = 0, \dots, T$

$$V_t^3(\tilde{\pi}_t^3) = \mathbb{E}\{\bar{\rho}_t(s_t^3) + V_{t+1}^1(\pi_{t+1}^1) \mid \pi_t^3 = \tilde{\pi}_t^3\}, \quad (13)$$

$$V_t^2(\tilde{\pi}_t^2) = \inf_{\bar{g}_t^2} \left\{\mathbb{E}(V_t^3(\pi_t^3) \mid \pi_t^2 = \tilde{\pi}_t^2, \bar{g}_t^2)\right\}, \quad (14)$$

$$V_t^1(\tilde{\pi}_t^1) = \inf_{\bar{g}_t^1} \left\{\mathbb{E}(V_t^2(\pi_t^2) \mid \pi_t^1 = \tilde{\pi}_t^1, \bar{g}_t^1)\right\}. \quad (15)$$

Furthermore, the optimal cost $\min_{\bar{K}} \bar{\mathcal{J}}(\bar{K}) = \mathbb{E}(V_0^1(\pi_0^1))$.

Just like the framework for a classical POMDP, the set of value function recursions provide an algorithmic procedure to solve the optimal policy. However, this procedure does not guarantee that the optimal policy is linear.

For every policy K in the decentralized problem, there exists a unique control policy $\bar{K}$ such that $\bar{g}_t^i(m_{t-1}^i, y_t^i) = g_t^i(m_{t-1}^i, y_t^i, y_{0:t}^0)$, and they achieve the same cost, and vice versa. To prove this, we can list the POMDP tuples for the centralized and decentralized problems, and show that there is a bijection between the centralized and the decentralized policies. Therefore, if there exists a policy K that minimizes cost $\mathcal{J}(K)$, then the corresponding $\bar{K}$ must also achieve the same optimal cost in the centralized framework, and vice versa.

We simplify the fictitious states above according to two facts. First, note that the map between $(y_{0:t}^i, u_{0:t}^i)$ and $(y_{0:t}^i, \bar{g}_{0:t}^i)$ is bijective. Second, Theorem 1 in [10] suggests that u_t^2 is a function of only $(y_{0:t}^0, y_t^2)$ for our particular decentralized MDP problem, where $y_t^1 = 0$ for all t. Thus, we can define $m_t^1 = (\bar{g}_{0:t}^1)$ and $m_t^2 = (y_t^2, \bar{g}_t^2)$ and change $\mathcal{M}_t^i$, $\mathcal{G}_t^i$, and $\bar{\mathcal{G}}_t^i$ accordingly. All statements above now hold with $u_t^1 = \bar{g}_t^1(\bar{g}_{0:t-1}^1) = g_t^1(\bar{g}_{0:t-1}^1, y_{0:t}^0)$ and $u_t^2 = \bar{g}_t^2(y_t^2) = g_t^2(y_t^2, y_{0:t}^0)$. In fact, we will use this definition for m_t^i throughout the paper.

5 POMDP Formulation

We first define the states, the observations, and the actions for the decentralized POMDP. We incorporate n_t into the observations y_{t+1} so that the controller can utilize n_t. Thus, we incorporate n_t into the states as follows:

$$x_t = \begin{cases} z_0 & t = 0, \\ (z_t^1, z_t^2, n_{t-1}^1, n_{t-1}^2) & 0 < t \le T. \end{cases} \quad (16)$$

We define the common observations to be

$$y_t^0 = \begin{cases} z_0^1 & t = 0, \\ (z_t^1, n_{t-1}^1) & 0 < t \le T. \end{cases} \quad (17)$$

where we denote $y_t^{01} = z_t^1$ and $y_t^{02} = n_{t-1}^1$. We define the private observations to be $y_t^1 = \phi$ and

$$y_t^2 = \begin{cases} z_0^2 & t = 0, \\ (z_t^2, n_{t-1}^2) & 0 < t \le T. \end{cases} \quad (18)$$

where we denote $y_t^{21} = z_t^2$ and $y_t^{22} = n_{t-1}^2$. By convention, we let $y_0^{02} = n_{-1}^1 = \phi$ and $y_0^{22} = n_{-1}^2 = \phi$.

We now define the centralized POMDP. Let centralized state s_t^k and information state I_t^k be as defined in (11) and (12).

Definition 3. *(Centralized POMDP) Define the centralized fictitious POMDP tuple* $(\bar{A}, \bar{C}, \bar{\rho})$ *as follows:*

1. $\bar{A}$ *is a sequence* $\bar{A}_0^1$, $\bar{A}_0^2$, $\bar{A}_0^3$, $\bar{A}_1^1$, $\bar{A}_1^2$, $\bar{A}_1^3$, …, $\bar{A}_T^1$, $\bar{A}_T^2$, $\bar{A}_T^3$ *with* $\bar{A}_0^1 \in M_{\mathcal{S}_0^1}$, *and for* $t \ge 1$, $\bar{A}_t^1 : \mathcal{S}_t^1 \times \mathcal{S}_{t-1}^3 \mapsto [0,1]$ *such that* $\bar{A}_t^1(\cdot, s) \in M_{\mathcal{S}_t^1}$ *for all* $s \in \mathcal{S}_{t-1}^3$. *For all* $t \ge 0$, $\bar{A}_t^k : \mathcal{S}_t^k \times \mathcal{S}_t^{k-1} \times \bar{\mathcal{G}}_t^{k-1} \mapsto [0,1]$ *such that* $\bar{A}_t^k(\cdot, s, \bar{g}_t^{k-1}) \in M_{\mathcal{S}_t^k}$ *for all* $s \in \mathcal{S}_t^{k-1}$ *and* $\bar{g}_t^{k-1} \in \bar{\mathcal{G}}_t^{k-1}$ *with* $k \in \{2, 3\}$.
2. $\bar{C}$ *is a sequence* $\bar{C}_0, \dots, \bar{C}_T$ *with* $\bar{C}_t : \mathcal{Y}^0 \times \mathcal{S}_t^1 \mapsto [0,1]$ *such that* $\bar{C}_t(\cdot, s) \in M_{\mathcal{Y}^0}$ *for all* $t \ge 0$ *and* $s \in \mathcal{S}_t^1$.
3. $\bar{\rho}$ *is a sequence* $\bar{\rho}_0, \dots, \bar{\rho}_T$ *with* $\bar{\rho}_t : \mathcal{S}_t^3 \mapsto \mathbb{R}$ *for all* $t \ge 0$.

Definition 4. *(Centralized Policy) A POMDP policy for the centralized fictitious player problem is a sequence* $\bar{K} = (\bar{K}_0^1, \bar{K}_0^2, \bar{K}_1^1, \bar{K}_1^2, \dots, \bar{K}_T^1, \bar{K}_T^2)$, $\bar{K}_t^1 : (\mathcal{Y}^0)^{t+1} \times \prod_{\tau=0}^{t-1}(\bar{\mathcal{G}}_\tau^1 \times \bar{\mathcal{G}}_\tau^2) \mapsto [0,1]$ *and* $\bar{K}_t^2 : (\mathcal{Y}^0)^{t+1} \times \prod_{\tau=0}^{t-1}(\bar{\mathcal{G}}_\tau^1 \times \bar{\mathcal{G}}_\tau^2) \times \bar{\mathcal{G}}_t^1 \mapsto [0,1]$ *such that for all* $t \ge 0$, $\bar{K}_t^1(\cdot, y_{0:t}^0, \bar{g}_{0:t-1}^{1:2}) \in M_{\bar{\mathcal{G}}_t^1}$ *and* $\bar{K}_t^2(\cdot, y_{0:t}^0, \bar{g}_{0:t-1}^{1:2}, \bar{g}_t^1) \in M_{\bar{\mathcal{G}}_t^2}$ *for all* $y_\tau^0 \in \mathcal{Y}^0$ *and* $\bar{g}_\tau^i \in \bar{\mathcal{G}}_\tau^i$ *with* $\tau \in \{0, \dots, t\}$.

We label the entries of m_t^i as $m_t^1 = (q_{0:t}^1)$ and $m_t^2 = (y_t^2, q_t^2)$ to avoid confusion between the state variables q_t^i and the actions $\bar{\mathbf{g}}_t^i$. Let $b_t(n_t)$ be the PDF of n_t, then $b(n_t) = \det(N_t \bar{N} + (I - N_t)(I - \bar{N}))$ with the convention that $b(\phi) = 1$. Similarly, we define $b^i(n_t^i) = \det(N_t^i \bar{N}^i + (I - N_t^i)(I - \bar{N}^i))$ and $b^i(\phi) = 1$.

Definition 5. *The POMDP tuple* $(\bar{A}, \bar{C}, \bar{\rho})$ *defined in Definition 3 has the explicit form for the problem defined in Section 2:*

1. *The state transition* $\bar{A}$

$$\begin{aligned}
&\bar{A}_0^1(z_0) = f_{\Sigma_s}(z_0 - \mu_s) \quad \text{with } z_0 = x_0 = s_0^1, \\
&\bar{A}_t^1(s_t^1, \hat{s}_{t-1}^3) = \delta(m_{t-1}^{1:2}, \hat{m}_{t-1}^{1:2})\, b(n_{t-1}) \\
&\qquad \times f_{\Sigma_v}\left(z_t - A\hat{z}_{t-1} - BN_{t-1}\begin{bmatrix} \hat{q}_{t-1}^1(\hat{q}_{0:t-2}^1) \\ \hat{q}_{t-1}^2(\hat{y}_{t-1}^2) \end{bmatrix}\right), \\
&\bar{A}_t^2(s_t^2, \hat{s}_t^1, \bar{g}_t^1) = \delta(x_t, \hat{x}_t)\, \delta(m_{t-1}^{1:2}, \hat{m}_{t-1}^{1:2}) \delta(q_t^1, \bar{g}_t^1), \\
&\bar{A}_t^3(s_t^3, \hat{s}_t^2, \bar{g}_t^2) = \delta(x_t, \hat{x}_t)\, \delta(m_t^1, \hat{m}_t^1) \\
&\qquad \times \delta(q_t^2, \bar{g}_t^2)\, \delta(y_t^{21}, z_t^2)\, \delta(y_t^{22}, n_{t-1}^2).
\end{aligned}$$

2. *The observation* $\bar{C}$

$$\bar{C}_t(y_t^0, s_t^1) = \delta(z_t^1, y_t^{01})\, \delta(n_{t-1}^1, y_t^{02}).$$

3. *The cost* $\bar{\rho}$

$$\begin{aligned}
&\bar{\rho}_T(s_T^3) = z_T^\mathsf{T} Q_T z_T \quad \text{at } t = T\text{; otherwise} \\
&\bar{\rho}_t(s_t^3) = z_t^\mathsf{T} Q z_t \\
&\qquad + \begin{bmatrix} q_t^1(q_{0:t-1}^1) \\ q_t^2(y_t^2) \end{bmatrix}^\mathsf{T} \mathbb{E}(N_t R N_t) \begin{bmatrix} q_t^1(q_{0:t-1}^1) \\ q_t^2(y_t^2) \end{bmatrix}.
\end{aligned}$$

where x_t, y_t^0, y_t^2 are given by (16), (17), (18), and

$$\begin{aligned} s_t^1 &= (x_t, m_{t-1}^1, m_{t-1}^2), & \hat{s}_t^1 &= (\hat{x}_t, \hat{m}_{t-1}^1, \hat{m}_{t-1}^2),\\ s_t^2 &= (x_t, m_t^1, m_{t-1}^2), & \hat{s}_t^2 &= (\hat{x}_t, \hat{m}_t^1, \hat{m}_{t-1}^2),\\ s_t^3 &= (x_t, m_t^1, m_t^2), & \hat{s}_t^3 &= (\hat{x}_t, \hat{m}_t^1, \hat{m}_t^2). \end{aligned}$$

6 Estimation and the Belief States

Since controllers are separable in centralized problems, the optimal controller can be calculated through belief states and value function recursions.

Definition 6. *Define the belief states*

$$\pi_t^k(s_t^k) = \Pr(s_t^k \,|\, \mathsf{I}_t^k) \quad \textit{for all } k \in \{1,2,3\}, t \geq 0,$$

where I_t^k is given by (12).

In a centralized POMDP, we can derive the explicit formulations for the belief states through Kalman filter.

Theorem 7. *(Belief States of the Fictitious Player) Consider the centralized fictitious player POMDP model as given in Definition 5, the belief states are*

$$\begin{aligned} \pi_0^1(s_0^1) &= \delta(z_0^1, \mathsf{y}_0^{01})\, \zeta_0(z_0^2),\\ \pi_0^2(s_0^2) &= \delta(z_0^1, \mathsf{y}_0^{01})\, \zeta_0(z_0^2)\, \delta(q_0^1, \bar{\mathsf{g}}_0^1),\\ \pi_0^3(s_0^3) &= \delta(z_0^1, \mathsf{y}_0^{01})\, \delta(z_0^2, y_0^{21})\, \delta(q_0^1, \bar{\mathsf{g}}_0^1)\, \delta(q_0^2, \bar{\mathsf{g}}_0^2)\, \zeta_0(y_0^{21}), \end{aligned}$$

and for $t > 0$,

$$\begin{aligned} \pi_t^1(s_t^1) &= \delta(q_{0:t-1}^1, \bar{\mathsf{g}}_{0:t-1}^1)\, \delta(q_{t-1}^2, \bar{\mathsf{g}}_{t-1}^2)\, \zeta_{t-1}(y_{t-1}^{21})\, b^2(y_{t-1}^{22})\\ &\quad \times \delta(z_t^1, \mathsf{y}_t^{01})\, \zeta_t(z_t^2)\, \delta(n_{t-1}^1, \mathsf{y}_t^{02})\, b^2(n_{t-1}^2),\\ \pi_t^2(s_t^2) &= \delta(q_{0:t}^1, \bar{\mathsf{g}}_{0:t}^1)\, \delta(q_{t-1}^2, \bar{\mathsf{g}}_{t-1}^2)\, \zeta_{t-1}(y_{t-1}^{21})\, b^2(y_{t-1}^{22})\\ &\quad \times \delta(z_t^1, \mathsf{y}_t^{01})\, \zeta_t(z_t^2)\, \delta(n_{t-1}^1, \mathsf{y}_t^{02})\, b^2(n_{t-1}^2),\\ \pi_t^3(s_t^3) &= \delta(q_{0:t}^1, \bar{\mathsf{g}}_{0:t}^1)\, \delta(q_t^2, \bar{\mathsf{g}}_t^2)\, \zeta_t(y_t^{21})\, b^2(y_t^{22})\\ &\quad \times \delta(z_t^1, \mathsf{y}_t^{01})\, \delta(z_t^2, y_t^{21})\, \delta(n_{t-1}^1, \mathsf{y}_t^{02})\, \delta(n_{t-1}^2, y_t^{22}), \end{aligned}$$

where $\zeta_t(z)$ is a PDF on z such that $\mathbb{E}(z) = \mu_t$ and $\mathrm{cov}(z) = \Sigma_t$ with $\zeta_0(z) = f_{\Sigma_s^2}(z - \mu_s^2)$, i.e. $\mu_0 = \mu_s^2$, $\Sigma_0 = \Sigma_s^2$, and

$$\begin{aligned} \mu_{t+1} &= A^{21}\mathsf{y}_t^{01} + A^{22}\mu_t\\ &\quad + B^{21}\,\mathrm{diag}(\mathsf{y}_{t+1}^{02})\bar{\mathsf{g}}_t^1(\bar{\mathsf{g}}_{0:t-1}^1) + B^{22}\bar{N}^2\bar{\mathsf{u}}_t^2, \qquad (19)\\ \Sigma_{t+1} &= \Sigma_v^2 + A^{22}\Sigma_t A^{22\mathsf{T}} + B^{22}\,\mathrm{cov}(N_t^2\mathsf{u}_t^2 \,|\, \mathsf{I}_{t+1}^1)B^{22\mathsf{T}}, \end{aligned}$$

where $\bar{\mathsf{u}}_t^2 = \mathbb{E}_{y_t^2}(\bar{\mathsf{g}}_t^2(y_t^2) \,|\, \mathsf{I}_{t+1}^1)$.

Proof. Prove by Induction using the POMDP formula in Corollary 5. The full formulation for $\zeta_{t+1}(z_{t+1}^2)$ is

$$\begin{aligned} \zeta_{t+1}(z_{t+1}^2) &= \int \zeta_t(y_t^{21}) b^2(y_t^{22}) f_{\Sigma_v^2}(z_{t+1}^2 - A^{21}\mathsf{y}_t^{01}\\ &- A^{22}y_t^{21} - B^{21}\,\mathrm{diag}(\mathsf{y}_{t+1}^{02})\bar{\mathsf{g}}_t^1(\bar{\mathsf{g}}_{0:t-1}^1) - B^{22}\bar{N}^2\bar{\mathsf{g}}_t^2(y_t^2))dy_t^2 \end{aligned}$$

The proof is omitted here due to space constraints. ∎

7 Controller and the Value Function Recursion

In a centralized POMDP, there is a standard procedure for value function recursions through dynamic programming. We will repeat the procedure to our centralized fictitious player problem. While the procedure to the recursions is standard, the difficulties lies in exploiting the structure of the value functions so that the recursions are remain tractable as we progress recursively.

Theorem 8. *(Fictitious Controller) Consider the centralized fictitious player POMDP tuple as given in Definition 5 and the belief states as given in Theorem 7, let V_t^k be as given in Lemma 2, then $V_t^1(\pi_t^1) = \int \sigma_t^1(s_t^1)\pi_t^1(s_t^1)ds_t^1$ where*

$$\sigma_t^1(s_t^1) = \begin{bmatrix} z_t \\ z_t^2 - \mu_t \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} P_t & \\ & Y_t - P_t^{22} \end{bmatrix} \begin{bmatrix} z_t \\ z_t^2 - \mu_t \end{bmatrix} + r_t,$$

and P_t, Y_t, and r_t are defined by recursions (4), (5), and (6), respectively, with $P_T = Q_T$, $Y_T = Q_T^{22}$, and $r_T = 0$. The optimal cost $\min_{\bar{K}} \bar{\mathcal{J}}(\bar{K}) =$

$$\sum_{t=1}^{T} \mathrm{trace}\left(\begin{bmatrix} P_t^{11} & P_t^{12} \\ P_t^{21} & Y_t \end{bmatrix}\Sigma_v\right) + \mathrm{trace}\left(\begin{bmatrix} P_0^{11} & P_0^{12} \\ P_0^{21} & Y_0 \end{bmatrix}\Sigma_s\right),$$

and the optimal controllers are

$$\begin{aligned} \bar{g}_t^1(\bar{g}_{0:t-1}^1) &= -K_t^{11}z_t^1 - K_t^{12}\mu_t, \qquad (20)\\ \bar{g}_t^2(z_t^2) &= -K_t^{21}z_t^1 - K_t^{22}\mu_t - J_t(z_t^2 - \mu_t), \qquad (21) \end{aligned}$$

with constant matrices K_t and J_t as given in (7) and (8). μ_t in (19) then becomes a function of $(z_{0:t}^1, n_{0:t-1}^1)$ by

$$\mu_t = A_{t-1}^K(N_{t-1}^1)\mu_{t-1} + B_{t-1}^K(N_{t-1}^1)z_{t-1}^1, \qquad (22)$$

where $A_t^K(N_t^1)$ and $B_t^K(N_t^1)$ are functions of N_t^1 as given in (9) and (10).

Proof. Prove by backward induction by using Lemma 2. Suppose $V_{t+1}^1(\tilde{\pi}_{t+1}^1) = \int \sigma_{t+1}^1(s_{t+1}^1)\tilde{\pi}_{t+1}^1(s_{t+1}^1)ds_{t+1}^1$. Let $W_t = Y_t - P_t^{22}$ for all t. By (13), we have $V_t^3(\tilde{\pi}_t^3) = \int \sigma_t^3(s_t^3)\tilde{\pi}_t^3(s_t^3)ds_t^3$ where $\sigma_t^3(s_t^3)$ equals

$$\int \left(\overbrace{\begin{bmatrix} z_t \\ u_t \\ z_t^2 - \mu_t \\ u_t^2 - \bar{u}_t^2 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \alpha & \\ & \beta \end{bmatrix} \begin{bmatrix} z_t \\ u_t \\ z_t^2 - \mu_t \\ u_t^2 - \bar{u}_t^2 \end{bmatrix}}^{\hat{\sigma}_t^3(z_t, u_t)} + r_t \right) \times \delta(u_t^1, \bar{g}_t^1(\bar{g}_{0:t-1}^1))\delta(u_t^2, \bar{g}_t^2(y_t^2))du_t,$$

with $r_t = \mathrm{trace}(P_{t+1}\Sigma_v) + \mathrm{trace}(W_{t+1}\Sigma_v^2) + r_{t+1}$ and

$$\alpha = \begin{bmatrix} Q + A^{\mathsf{T}}P_{t+1}A & A^{\mathsf{T}}P_{t+1}B\bar{N} \\ \bar{N}B^{\mathsf{T}}P_{t+1}A & \mathbb{E}(N_t(R + B^{\mathsf{T}}P_{t+1}B)N_t) \end{bmatrix},$$

$$\beta = \begin{bmatrix} A^{22\mathsf{T}}W_{t+1}A^{22} & A^{22\mathsf{T}}W_{t+1}B^{22}\bar{N}^2 \\ \bar{N}^2B^{22\mathsf{T}}W_{t+1}A^{22} & \mathbb{E}(N_t^2(B^{22,\mathsf{T}}W_{t+1}B^{22})N_t^2) \end{bmatrix}.$$

By (14) and after some calculations, we have

$$V_t^2(\tilde{\pi}_t^2) = \int \inf_{u_t^2} \left(\hat{\sigma}_t^3(z_t, u_t)|_{z_t^1 = y_t^{01}, z_t^2 = y_t^{21}, u_t^1 = \bar{g}_t^1(\bar{g}_{0:t-1}^1)} \right) \times \zeta_t(y_t^{21})\, b^2(y_t^{22}) dy_t^2.$$

By taking derivative with respect to u_t^2 inside $\inf_{u_t^2}(\cdot)$ and set to zero, we have the following

$$\bar{u}_t^2 = -(F^{22})^{-1}(F^{21}u_t^1 + H^{21}z_t^1 + H^{22}\mu_t), \quad (23)$$

$$u_t^2 = -(F^{22})^{-1}F^{21}u_t^1 - E^{-1}(\beta^{22}(F^{22})^{-1}H^{22} - \beta^{21})\mu_t - (F^{22})^{-1}H^{21}z_t^1 - E^{-1}(\beta^{21} + H^{22})z_t^2, \quad (24)$$

where $\bar{u}_t^2 = \mathbb{E}_{y_t^2}(u_t^2)$ and

$$F = \alpha^{22} = \mathbb{E}(N_t(R + B^{\mathsf{T}}P_{t+1}B)N_t),$$
$$H = \alpha^{21} = \bar{N}B^{\mathsf{T}}P_{t+1}A,$$
$$E = \mathbb{E}(N_t^2(R^{22} + B^{22^{\mathsf{T}}}Y_{t+1}B^{22})N_t^2) = \beta^{22} + F^{22}.$$

Then, $V_t^2(\tilde{\pi}_t^2) = \int \sigma_t^2(s_t^2)\tilde{\pi}_t^2(s_t^2)ds_t^2$ where $\sigma_t^2(s_t^2) = \int \hat{\sigma}_t^3(z_t, u_t)|_{(24)}\delta(u_t^1, \bar{g}_t^1(\bar{g}_{0:t-1}^1))du_t^1$. Similarly, by (15),

$$V_t^1(\tilde{\pi}_t^1) = \inf_{u_t^1} \left\{ \int \hat{\sigma}_t^3(z_t, u_t)|_{(24)}\delta(z_t^1, y_t^{01})\zeta_t(z_t^2)dz_t \right\}.$$

By taking derivative with respect to u_t^1 inside $\inf_{u_t^1}\{\cdot\}$ and set to zero, we have u_t^1 as in (20) where

$$K_t = F^{-1}H = \begin{bmatrix} K_t^{11} & K_t^{12} \\ K_t^{21} & K_t^{22} \end{bmatrix}.$$

By plugging (20) back to (24) and let $J_t = E^{-1}(\beta^{21} + H^{22})$ be as given in (8), we have u_t^2 as in (21). Finally, by plugging (20) and (21) back to V_t^1, $V_t^1(\tilde{\pi}_t^1) = \int \sigma_t^1(s_t^1)\tilde{\pi}_t^1(s_t^1)ds_t^1$ with $\sigma_t^1(s_t^1)$ as given in the theorem.

The optimal cost is straightforward by calculating $\min_{\bar{K}} \bar{\mathcal{J}}(\bar{K}) = \mathbb{E}_{y_0^0}(V_0^1(\pi_0^1))$ with $\mu_s = 0$. By plugging (20), (23), and (17) into (19), we have the recursion (22). ■

We are now ready to prove the main Theorem.

Proof. (Theorem 1) Given the optimal centralized fictitious player policy as in Theorem 8, the rest is to show the corresponding decentralized cost and policy. According to Section 4.2, we know that for any policy $\bar{K}$ in the fictitious centralized problem, there exists a decentralized policy K with the same cost. Given that (20) and (21) are optimal policies with the optimal cost

$$\sum_{t=1}^{T} \operatorname{trace}\left(\begin{bmatrix} P_t^{11} & P_t^{12} \\ P_t^{21} & Y_t \end{bmatrix} \Sigma_v \right) + \operatorname{trace}\left(\begin{bmatrix} P_0^{11} & P_0^{12} \\ P_0^{21} & Y_0 \end{bmatrix} \Sigma_s \right)$$

for the centralized problem, the optimal cost of the decentralized problem must be the same, and the correspond decentralized policies must be $u_t^1 = g_t^1(\bar{g}_{0:t-1}^1, y_{0:t}^0) = \bar{g}_t^1(\bar{g}_{0:t-1}^1)$ and $u_t^2 = g_t^2(z_t^2, y_{0:t}^0)$ as in (20) and (21). Also, note that the recursion for μ_t and ξ_t are exactly the same. This completes the proof. ■

8 Conclusion

In this paper, we derived the explicit state space formulae for a decentralized two-player problem under TCP-like erasure channels with state feedback. We first characterize the problem as a decentralized POMDP and recast it as an centralized POMDP in the fictitious players framework. By calculating the belief states and the value function recursions, we solved the estimator and the optimal controller for the decentralized problem.

The main Theorem of this paper is a generalization of the main results in [7] and a decentralization of the main results in [2]. We showed that the optimal decentralized controllers for both players require an estimator of the state of the subsystem 2 conditioned on the information given to player 1.

References

[1] Sachin Adlakha, Sanjay Lall, and Andrea Goldsmith. Networked markov decision process with delays. To appear, IEEE Transactions on Automatic Control, 2009.

[2] E. Garone, B. Sinopoli, A. Goldsmith, and A. Casavola. Lqg control for distributed systems over tcp-like erasure channels. In *Proceedings of IEEE Conference on Decision and Control*, 2007.

[3] Y-C. Ho and K. C. Chu. Team decision theory and information structures in optimal control problems – Part I. *IEEE Transactions on Automatic Control*, 17(1):15–22, 1972.

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

[5] Aditya Mahajan, Ashutosh Nayyar, and Demosthenis Teneketzis. Identifying tractable decentralized control problems on the basis of information structure. In *Proceedings of the Allerton Conference*, 2008.

[6] Michael Rotkowitz and Sanjay Lall. A characterization of convex problems in decentralized control. *IEEE Transactions on Automatic Control*, 51(2):247–286, 2002.

[7] John Swigart and Sanjay Lall. An explicit dynamic programming solution for a decentralized two-player optimal linear-quadratic regulator. In *Proceedings of the International Symposium on Mathematical Theory of Networks and Systems*, 2010.

[8] John Swigart and Sanjay Lall. An explicit state-space solution for a decentralized two-player optimal linear-quadratic regulator. In *Proceedings of the American Control Conference*, 2010.

[9] H. S. Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal of Control*, 6(1):131–147, 1968.

[10] Jeff Wu and Sanjay Lall. A dynamic programming algorithm for decentralized markov decision processes with a broadcast structure. In *Proceedings of IEEE Conference on Decision and Control*, 2010.