2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Optimal Communication and Control Strategies for a Multi-Agent System in the Presence of an Adversary

Dhruva Kartik, Sagar Sudhakara, Rahul Jain and Ashutosh Nayyar

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9992871

Abstract— **We consider a multi-agent system in which a decentralized team of agents controls a stochastic system in the presence of an adversary. Instead of committing to a fixed information sharing protocol, the agents can strategically decide at each time whether to share their private information with each other or not. The agents incur a cost whenever they communicate with each other and the adversary may eavesdrop on their communication. Thus, the agents in the team must effectively coordinate with each other while being robust to the adversary's malicious actions. We model this interaction between the team and the adversary as a stochastic zero-sum game where the team aims to minimize a cost while the adversary aims to maximize it. Under some assumptions on the adversary's capabilities, we characterize a min-max control and communication strategy for the team. We supplement this characterization with several structural results that can make the computation of the min-max strategy more tractable.**

I. Introduction

In multi-agent systems, the agents may not be able to fully observe the system state and the actions of other agents. A multi-agent system is said to have an *asymmetric* information structure when different agents have access to different information. Each agent must select its actions based only on the limited information available to it. Decision-making scenarios with information asymmetry arise in a range of domains such as autonomous driving, power grids, transportation networks, cyber-security of networked computing and communication systems, and competitive markets and geopolitical interactions (see, for example, [1]–[5]).

Based on the nature of interactions between the agents, multi-agent systems can broadly be classified into three types: (i) teams, (ii) games and (iii) team-games. In teams, all the agents act in a cooperative manner to achieve a shared objective. In games, each agent has its own objective and is self-interested. In team-games, agents within a team are cooperative but the team as a whole is non-cooperative with respect to other teams. For agents in the same team, sharing information with each other aids coordination and improves performance. Various information sharing mechanisms [4] arise depending on the underlying communication environment. For instance, if the agents have access to a perfect, costless communication channel, they can share their entire information with each other. On the other hand, if communication is too expensive, the agents may never share their information. In this paper, instead of fixing the information sharing mechanism for agents in a team, we consider a model in which the agents can strategically decide whether to share their information with other agents or not. By doing so, the agents in the team can balance the trade-off between the control cost and the communication cost. This joint design of control and communication strategies was considered in [6] and a team-optimal solution was provided using the common information approach [4].

In some scenarios (e.g. a battlefield), the team of agents may be susceptible to adversarial attacks. Also, the adversary may have the capability to intercept the communication among the agents. This makes the information sharing mechanism substantially more complicated. While sharing information with teammates may be beneficial for intra-team coordination, it can reveal sensitive information to the adversary. The adversary may exploit this information to inflict severe damage on the system. Such interactions between a team of cooperative agents and an adversary can be modeled as a zero-sum team-game [7].

In this paper, our focus is on a zero-sum game between a team of two agents and an adversary in which the team aims to minimize the control and communication cost while the adversary aims to maximize it. The system state in this game has three components: a local state for each agent in the team and global state. The adversary controls the global state and each of the agents control their respective local states. We restrict our attention to models in which the agents in the team are more informed than the adversary. Our model allows us to capture several scenarios of interest. For example, the adversary in our model can affect the quality of and the cost associated with the agents' communication channel and the agents can perfectly or imperfectly encrypt their communication. We analyze a family of such zero-sum team vs. adversary game and provide a characterization of an optimal (min-max) control and communication strategy for the team. This characterization is based on common information belief based min-max dynamic program for team vs. team games discussed in [7].

a) Related Works: There is a large body of prior work on decision-making in multi-agent systems. In this section, we discuss related works on cooperative teams and team-games. In decentralized stochastic control literature, a variety of information structures (obtained from different information sharing protocols) have been considered [3]–[6]. Another well-studied class of multi-agent teams with asymmetric information is the class of Decentralized Partially Ob-

D. Kartik, S. Sudhakara, R. Jain and A. Nayyar are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 (e-mail: mokhasun@usc.edu; sagarsud@usc.edu; rahul.jain@usc.edu and ashutosn@usc.edu).

This work was supported by National Science Foundation (NSF) grants ECCS 2025732, ECCS 1750041, CMMI-1839842, ECCS-1810447 and ONR award N00014-20-1-2258.

servable Markov Decision Processes (Dec-POMDPs). Several methods for solving such generic Dec-POMDPs exist in the literature [8]–[13].

Dynamic games among teams have received some attention over the past few years. Two closely related works are [14] and [7]. In [14], a model of games among teams where players in a team internally share their information with some delay was investigated. The authors of [14] characterize Team-Nash equilibria under certain existence assumptions. In [7], a general model of zero-sum games between two teams was considered. For this general model, the authors provide bounds on the upper and lower values of the zero-game. A relatively specialized model was also studied in [7] and for this model, a min-max strategy for one of the teams was characterized in addition to the min-max value. In [15], the authors formulate and solve a particular malicious intrusion game between two teams of mobile agents.

The works that are most closely related to our work are [6] and [7]. In [6], the authors consider a *team* problem in which the agents can strategically decide when to communicate with each other. While our model is inspired by the model in [6], our model is substantially more general and complicated because of the presence of an adversary. In team problems, the agents can use deterministic strategies without loss of optimality, whereas in games, the agents can benefit with randomization. Due to the randomness in agents' strategies and the need to solve a min-max problem as opposed to a simpler minimization problem, different techniques are required for analyzing and solving the team-game. Our game model is a special case of one of the models studied in [7] and hence, we can use the results in [7] to characterize a min-max strategy for the team. While we borrow some results from [7], our results on private information reduction in this paper are novel.

b) Notation: Random variables are denoted by upper case letters, their realizations by the corresponding lower case letters. In general, subscripts are used as time index while superscripts are used to index decision-making agents. For time indices $t_1 \leq t_2$, $X_{t_1:t_2}$ is the short hand notation for the variables $(X_{t_1}, X_{t_1+1}, ..., X_{t_2})$. Similarly, $X^{1:2}$ is the short hand notation for the collection of variables (X^1, X^2). Operators $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ denote the probability of an event, and the expectation of a random variable respectively. For random variables/vectors X and Y, $\mathbb{P}(\cdot|Y=y)$, $\mathbb{E}[X|Y=y]$ and $\mathbb{P}(X=x \mid Y=y)$ are denoted by $\mathbb{P}(\cdot|y)$, $\mathbb{E}[X|y]$ and $\mathbb{P}(x \mid y)$, respectively. For a strategy g, we use $\mathbb{P}^g(\cdot)$ (resp. $\mathbb{E}^g[\cdot]$) to indicate that the probability (resp. expectation) depends on the choice of g. For any finite set $\mathcal{A}$, $\Delta\mathcal{A}$ denotes the probability simplex over the set $\mathcal{A}$. For any two sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{F}(\mathcal{A}, \mathcal{B})$ denotes the set of all functions from $\mathcal{A}$ to $\mathcal{B}$. We define RAND to be mechanism that given (i) a finite set $\mathcal{A}$, (ii) a distribution d over $\mathcal{A}$ and a random variable K uniformly distributed over the interval $(0,1]$, produces a random variable $X \in \mathcal{A}$ with distribution d, i.e.,

$$X = \text{RAND}(\mathcal{A}, d, K) \sim d. \tag{1}$$

II. Problem Formulation

Consider a discrete-time control system with a team of two agents (agent 1 and agent 2) and an adversary. The system comprises of a global state and local states for each agent in the team. Let $X_t^0 \in \mathcal{X}^0$ denote the global state and let $X_t^i \in \mathcal{X}^i$ denote the local state of agent i. $X_t := (X_t^1, X_t^2)$ represents the local state of both agents in the team. The initial global state and the initial local states of both agents are independent random variables with state X_1^i having the probability distribution $P_{X_1^i}$, $i = 0, 1, 2$. Each agent perfectly observes its own local state and the global state is perfectly observed by all agents (including the adversary). Let $U_t^i \in \mathcal{U}^i$ denote the control action of agent i at time t. $U_t := (U_t^1, U_t^2)$ denotes the control actions of both agents at time t. Further, let $U_t^a \in \mathcal{U}^a$ denote the control action of the adversary at time t. The global and local states of the system evolve according to

$$X_{t+1}^0 = k_t^0(X_t^0, U_t^a, W_t^0), \tag{2}$$

$$X_{t+1}^i = k_t^i(X_t^0, X_t^i, U_t^i, W_t^i), \quad i = 1, 2, \tag{3}$$

where $W_t^i \in \mathcal{W}^i$, $i = 0, 1, 2$ is the disturbance in dynamics with probability distribution P_{W^i}. The initial states X_1^0, X_1^1, X_1^2 and the disturbances $\{W_t^i\}_{t=1}^{\infty}$, $i = 0, 1, 2$, are independent random variables. Note that the next local state of agent i depends on the current local state and control action of agent i *and* the global state. The next global state depends on the current global state and the adversary's action.

In addition to deciding the control actions at each time, the two agents in the team need to decide whether or not to initiate communication at each time. We use the binary variable M_t^i to denote the communication decision taken by agent i. Let $M_t^{or} := \max(M_t^1, M_t^2)$ and let Z_t^{er} represent the information exchanged between the agents at time t. In this model when global state $X_t^0 = x$, agents lose packets or fail to communicate with probability $p_e(x)$ even when one (or both) of the agents decides to communicate, i.e. when $M_t^{or} = 1$. Here, $p_e : \mathcal{X}^0 \to [0,1]$ maps the global state to a failure probability. Based on the communication model described above we can define variable Z_t^{er} given that $X_t^0 = x$ as:

$$Z_t^{er} = \begin{cases} X_t^{1,2}, \ w.p.\ 1 - p_e(x) & \text{if } M_t^{or} = 1. \\ \phi, \quad w.p. \quad p_e(x) & \text{if } M_t^{or} = 1. \\ \phi, & \text{if } M_t^{or} = 0. \end{cases} \tag{4}$$

At time t^+, the adversary observes a noisy version Y_t of the variable Z_t^{er} given by

$$Y_t = l_t(Z_t^{er}, M_t, X_t^0, W_t^y), \tag{5}$$

where W_t^y is the observation noise.

Information structure and decision strategies: At the beginning of the t-th time step, the information available to agent i is given by (i) history of global states and its local states, (ii) its control actions, (iii) communication actions and

messages and (iv) adversary's action and observation history:

$$I_t^i = \{X_{1:t}^0, X_{1:t}^i, U_{1:t-1}^i, M_{1:t-1}^{1,2}, Z_{1:t-1}^{er}, U_{1:t-1}^a, Y_{1:t-1}\}. \quad (6)$$

Agent i can use this information to make its communication decision at time t. We allow the agent to randomize its decision. Thus, agent i first selects a distribution δM_t^i over $\{0,1\}$ based n its information and then it randomly picks M_t^i according to the chosen distribution:

$$\delta M_t^i = f_t^i(I_t^i); \qquad M_t^i = \text{RAND}(\{0,1\}, \delta M_t^i, K_t^i) \quad (7)$$

where $K_t^i, i = 1, 2, t \geq 1$, are independent random variables uniformly distributed over the interval $(0, 1]$ that are used for randomization (these variables are also independent of initial states and all noises/disturbances). The function f_t^i is referred to as the communication strategy of agent i at time t. At this point, the adversary does not take any action. After the communication decisions are made and the resulting communication (if any) takes place, the information available to agent i is $I_{t^+}^i = \{I_t^i, Z_t^{er}, M_t^{1,2}, Y_t\}$.

I_t^a denotes the adversary's information just before the communication at time t and $I_{t^+}^a$ denotes the adversary's information after communication at time t^+. Our model allows for different scenarios of adversary's information which will be described later.

Agent i and the adversary choose their control actions based on their post-communication information according to

$$\delta U_t^i = g_t^i(I_{t^+}^i) \quad (8)$$

$$U_t^i = \text{RAND}(\mathcal{U}_t^i, \delta U_t^i, K_{t^+}^i), \quad i = 1, 2, a, \quad (9)$$

where $K_{t^+}^i, i = 1, 2, t \geq 1$, are independent random variables uniformly distributed over the interval $(0, 1]$ that are used for randomization (these variables are also independent of all other randomization variables, initial states and all noises/disturbances). The functions g_t^i and g_t^a are referred to as the control strategy of agent i and the adversary at time t. The tuples $f^i := (f_1^i, f_2^i, ..., f_T^i)$ and $g^i := (g_1^i, g_2^i, ..., g_T^i)$ are called the communication and control strategy of agent i respectively. The collection $f := (f^1, f^2)$, $g := (g^1, g^2)$ of communication and control strategies of both agents are called the communication and control strategy of the team. Similarly, $g^a := (g_1^a, g_2^a, ..., g_T^a)$ is called the control strategy of the adversary.

We can split the information available to the agents into two parts – common information and private information. Common information at a given time is the information available to all the decision-makers (including the adversary) at the given time. Private information of an agent includes all of its information at the given time except the common information.

1) At the beginning of time step t, before the communication decisions are made, the common (C_t) and private information (P_t^i) is defined as

$$C_t := I_t^1 \cap I_t^2 \cap I_t^a \quad (10)$$

$$P_t^i := I_t^i \setminus C_t \quad \forall i \in \{1, 2, a\}. \quad (11)$$

2) After the communication decisions are made and the resulting communication (if any) takes place, the common and private information is defined as

$$C_{t^+} := I_{t^+}^1 \cap I_{t^+}^2 \cap I_{t^+}^a \quad (12)$$

$$P_{t^+}^i := I_{t^+}^i \setminus C_{t^+} \quad \forall i \in \{1, 2, a\}. \quad (13)$$

Assumption 1. *We assume that the following conditions are satisfied:*

1) Monotonicity: *The adversary's information grows with time. Thus, $I_t^a \subseteq I_{t^+}^a \subseteq I_{t+1}^a$ for every t.*

2) Nestedness: *The adversary's information is common information and each agent in the team has access to adversary's information, i.e.,*

$$C_t = I_t^a \subseteq I_t^1 \cap I_t^2 =: C_t^{\text{TEAM}}$$
$$C_{t^+} = I_{t^+}^a \subseteq I_{t^+}^1 \cap I_{t^+}^2 =: C_{t^+}^{\text{TEAM}}.$$

Therefore, $P_t^a = P_{t^+}^a = \varnothing$.

3) Common Information Evolution: *(i) Let $Z_{t^+} \doteq C_{t^+} \setminus C_t$ and $Z_{t+1} \doteq C_{t+1} \setminus C_{t^+}$ be the increments in common information at times t^+ and $t+1$, respectively . Thus, $C_{t^+} = \{C_t, Z_{t^+}\}$ and $C_{t+1} = \{C_{t^+}, Z_{t+1}\}$. The common information evolves as*

$$Z_{t^+} = \zeta_{t^+}(P_t^{1:2}, M_t^{1:2}, Z_t^{er}, Y_t), \quad (14)$$

$$Z_{t+1} = \zeta_{t+1}(P_{t^+}^{1:2}, U_t^{1:2}, X_{t+1}^{0:2}) \quad (15)$$

where ζ_{t+1} and ζ_{t^+} are fixed transformations.

4) Private Information Evolution: *The private information evolves as*

$$P_{t^+}^i = \xi_{t^+}^i(P_t^{1:2}, M_t^{1:2}, Z_t^{er}, Y_t) \quad (16)$$

$$P_{t+1}^i = \xi_{t+1}^i(P_{t^+}^{1:2}, U_t^{1:2}, X_{t+1}^{0:2}) \quad (17)$$

where $\xi_{t^+}^i$ and ξ_{t+1}^i are fixed transformations and $i = 1, 2$.

Due to the nestedness condition in Assumption 1, the team is always more-informed than the adversary. Scenarios where the adversary has some private information are beyond the scope of this paper. The third and fourth conditions in Assumption 1 on the evolution of common and private information are very mild [4], [16] and most information structures of interest satisfy these conditions.

Strategy optimization problem: At time t, the system incurs a cost $c_t(X_t^0, X_t, U_t, U_t^a)$ that depends on the global state, the team's state, control actions of both agents and the adversary's action. Whenever agents decide to share their states with each other, they incur a state-dependent cost $\rho(X_t^0, X_t)$. The system runs for a time horizon T. The total expected cost over the time horizon T associated with a strategy profile $((f, g), g^a)$ is:

$$J((f, g), g^a) = \mathbb{E}^{((f,g),g^a)}\Big[\sum_{t=1}^{T} c_t(X_t^0, X_t, U_t, U_t^a) + \rho(X_t^0, X_t)\mathbb{1}_{\{M_t^{er}=1\}}\Big]. \quad (18)$$

The objective of the team is to find communication and control strategies (f, g) for the team in order to minimize the

worst-case expected total expected cost $\max_{g^a} J((f,g),g^a)$. This min-max optimization problem can be viewed as a zero-sum game between the team and the adversary. We denote this zero-sum game with Game $\mathscr{G}$. We denote the min-max value of this game $\mathscr{G}$ with $S^u(\mathscr{G})$, i.e.,

$$S^u(\mathscr{G}) = \min_{(f,g)} \max_{g^a} J((f,g),g^a). \tag{19}$$

III. Preliminary Results and Simplified Game $\mathscr{G}_s$

In this section we show that agents in the team can ignore parts of their information without losing optimality. This removal of information narrows the search for optimal strategies to a class of simpler strategies and is a key step in our approach for finding optimal strategies.

Let us define the team's *common private information* D_t before communication at time t and D_{t^+} after communication at time t^+ as

$$D_t := P_t^1 \cap P_t^2; \qquad D_{t^+} := P_{t^+}^1 \cap P_{t^+}^2. \tag{20}$$

The variables C_t, D_t (resp. C_{t^+}, D_{t^+}) constitute the *team's common information* at time t (resp. t^+), i.e.,

$$C_t \cup D_t = I_t^1 \cap I_t^2 = C_t^{\text{TEAM}} \tag{21}$$

$$C_{t^+} \cup D_{t^+} = I_{t^+}^1 \cap I_{t^+}^2 = C_{t^+}^{\text{TEAM}}. \tag{22}$$

Notice that C_t and D_t depend on the adversary's information structure. However, since the team's information structure is fixed, C_t, D_t combined do not depend on the adversary's information structure. The following lemma establishes a key conditional independence property that will be critical for our analysis.

Lemma 1 (Conditional independence property)**.** *At any time t, the two agents' local states and control actions are conditionally independent given the team's common information (C_t, D_t) (before communication) or C_{t^+}, D_{t^+} (after communication). That is, if $c_t, d_t, c_{t^+}, d_{t^+}$ are the realizations of the common information and common private information before and after communication respectively, then for any realization $x_{1:t}, u_{1:t-1}$ of states and actions, we have*

$$\mathrm{P}(x_{1:t}, u_{1:t-1} | c_t, d_t) = \prod_{i=1}^{2} \mathrm{P}(x_{1:t}^i, u_{1:t-1}^i | c_t, d_t), \tag{23}$$

$$\mathrm{P}(x_{1:t}, u_{1:t} | c_{t^+}, d_{t^+}) = \prod_{i=1}^{2} \mathrm{P}(x_{1:t}^i, u_{1:t}^i | c_{t^+}, d_{t^+}). \tag{24}$$

Further, $\mathrm{P}(x_{1:t}^i, u_{1:t-1}^i | c_t, d_t)$ and $\mathrm{P}(x_{1:t}^i, u_{1:t}^i | c_{t^+}, d_{t^+})$ depends on only on agent i' strategy.

Proof. The proof of this lemma is very similar to the proof of Lemma 1 in [6]. For a detailed proof, see Appendix I in [17]. □

The following proposition shows that agent i at time t and t^+ can ignore its past states and actions, i.e. $X_{1:t-1}^i$ and $U_{1:t-1}^i$, without losing optimality. This allows agents in the team to use simpler strategies where the communication and control decisions are functions only of the current state and the team's common information.

Proposition 1. *Agent i, $i = 1, 2$, can restrict itself to strategies of the form below*

$$M_t^i \sim \bar{f}_t^i(X_t^i, C_t, D_t) \tag{25}$$

$$U_t^i \sim \bar{g}_t^i(X_t^i, C_{t^+}, D_{t^+}) \tag{26}$$

without loss of optimality. In other words, at time t and t^+, agent i does not need the past local states and actions, $X_{1:t-1}^i, U_{t-1}^i$, for making optimal decisions.

Proof. See Appendix II in [17]. □

Proposition 1 leads to a simplified game in which the information used by the players in the team is substantially reduced. We will refer to this game as Game $\mathscr{G}_s$. Game $\mathscr{G}_s$ has the same dynamics and cost model as Game $\mathscr{G}$. The key difference between these two games lies in the team's information structure and strategy spaces. In Game $\mathscr{G}_s$, the information used by player i in the team at time t and t^+ respectively is

$$I_t^i = \{X_t^i\} \cup D_t \cup C_t \tag{27}$$

$$I_{t^+}^i = \{X_t^i\} \cup D_{t^+} \cup C_{t^+}. \tag{28}$$

Therefore, the common information in the simplified game $\mathscr{G}_s$ is the same as in the original game $\mathscr{G}$. In the simplified game $\mathscr{G}_s$, the private information[1] $P_t^i = X_t^i \cup D_t$.

Corollary 1. *If (f^*, g^*) is a min-max strategy in Game $\mathscr{G}_s$, then it is a min-max strategy in Game $\mathscr{G}$. Further, the min-max values of games $\mathscr{G}$ and $\mathscr{G}_s$ are identical.*

Henceforth, we make the following mild assumption on the information structure of agents in the simplified game $\mathscr{G}_s$. Several examples that satisfy this assumption are provided in [17].

Assumption 2. *The information structure in the simplified game $\mathscr{G}_s$ with reduced private information satisfies Assumption 1.*

Remark 1. *The reduced information in equations* (27) *and* (28) *is* unilaterally sufficient information *(see Definition 2.4 in [18]) for each player in the team. Proposition 1 can alternatively be shown using the concept of unilaterally sufficient information and Theorem 2.6 in [18].*

IV. Dynamic Program Characterization of a Min-max Strategy

It was shown in [7] that for certain zero-sum game models with a special structure, a virtual game $\mathscr{G}_e$ can be constructed based on the simplified Game $\mathscr{G}_s$, and this virtual game can be used to obtain the min-max value and a min-max strategy for the minimizing team. In our game model described in Section II, the adversary does not have any private information at any given time and hence, this model

[1] With a slight abuse of notation, we use the same letter for denoting private information in both games $\mathscr{G}$ and $\mathscr{G}_s$.

can be viewed as a special case of the game model described in paragraph (a), Section IV-A of [7]. Therefore, we can use the result in [7] to obtain the min-max value and a min-max strategy for our original Game $\mathscr{G}$. The virtual game $\mathscr{G}_e$ involves the same underlying system model as in game $\mathscr{G}_s$. The main differences among games $\mathscr{G}_s$ and $\mathscr{G}_e$ lie in the manner in which the actions used to control the system are chosen. In the virtual game $\mathscr{G}_e$, all the players in the team of game $\mathscr{G}_s$ are replaced by a virtual player (referred to as virtual player b) and the adversary is replaced by a virtual player (referred to as virtual player a). These virtual players in Game $\mathscr{G}_e$ operate as described in the following sub-section.

A. Virtual Game $\mathscr{G}_e$

Consider virtual player a associated with the adversary. At each time t^+, virtual player a selects a distribution Γ^a_t over the space $\mathcal{U}^a_t$. The set of all such mappings is denoted by $\mathcal{B}^a_t \doteq \Delta\mathcal{U}^a_t$. Consider virtual player b associated with the team. At each time t and for each $i = 1, 2$, virtual player b selects a function Γ^i_t that maps private information P^i_t to a distribution δM^i_t over the space $\{0, 1\}$. Thus, $\delta M^i_t = \Gamma^i_t(P^i_t)$. The set of all such mappings is denoted by $\mathcal{B}^i_t \doteq \mathcal{F}(\mathcal{P}^i_t, \Delta\{0, 1\})$. We refer to the tuple $\Gamma_t \doteq (\Gamma^1_t, \Gamma^2_t)$ as virtual player b's *prescription* at time t. The set of all possible prescriptions for virtual player b at time t is denoted by $\mathcal{B}_t \doteq \mathcal{B}^1_t \times \mathcal{B}^2_t$. At each time t^+ and for each $i = 1, 2$, virtual player b selects a function Λ^i_t that maps private information $P^i_{t^+}$ to a distribution $\delta U^i_{t^+}$ over the space $\mathcal{U}^i_t$. Thus, $\delta U^i_t = \Lambda^i_t(P^i_{t+})$. The set of all such mappings is denoted by $\mathcal{B}^i_{t^+} \doteq \mathcal{F}(\mathcal{P}^i_{t^+}, \Delta\mathcal{U}^i_t)$. We refer to the tuple $\Lambda_t \doteq (\Lambda^1_t, \Lambda^2_t)$ as virtual player b's *prescription* at time t^+. The set of all possible prescriptions for virtual player b at time t^+ is denoted by $\mathcal{B}_{t^+} \doteq \mathcal{B}^1_{t^+} \times \mathcal{B}^2_{t^+}$. Once virtual players select their prescriptions at times t and t^+, the corresponding actions are generated as

$$M^i_t = \textsc{rand}(\{0, 1\}, \Gamma^i_t(P^i_t), K^i_t) \tag{29}$$

$$U^i_t = \textsc{rand}(\mathcal{U}^i_t, \Lambda^i_t(P^i_{t^+}), K^i_{t^+}) \tag{30}$$

$$U^a_t = \textsc{rand}(\mathcal{U}^a_t, \Gamma^a_t, K^a_{t^+}). \tag{31}$$

In virtual game $\mathscr{G}_e$, virtual players' information I^v_t at time t comprises of the common information C_t and the past prescriptions of both players $\Gamma_{1:t-1}, \Gamma^a_{1:t-1}, \Lambda_{1:t-1}$. At time t, Virtual player b selects its prescription according to a control law χ^b_t, i.e. $\Gamma_t = \chi^b_t(I^v_t)$. Note that at time t, Virtual player a does not take any action. At time t^+, the virtual players information $I^v_{t^+}$ comprises of C_{t^+} and all the past prescriptions of both players $\Gamma_{1:t}, \Gamma^a_{1:t-1}, \Lambda_{1:t-1}$. Virtual player a selects its prescription according to a control law χ^a_t, i.e., $\Gamma^a_t = \chi^a_t(I^v_{t^+})$ and virtual player b selects its prescription according to a control law $\chi^b_{t^+}$, i.e. $\Lambda_t = \chi^b_t(I^v_{t^+})$. For virtual player a, the collection of control laws over the entire time horizon $\chi^a = (\chi^a_1, \dots, \chi^a_T)$ is referred to as its control strategy. Similarly for virtual player b. Let $\mathcal{H}^a_t$ be the set of all possible control laws for virtual player a at time t and let $\mathcal{H}^a$ be the set of all possible control strategies for virtual player a, i.e. $\mathcal{H}^a = \mathcal{H}^a_1 \times \cdots \times \mathcal{H}^a_T$. For virtual player b, the collection of control laws over the entire time horizon $\chi^b = (\chi^b_1, \chi^b_{1^+} \dots, \chi^b_T, \chi^b_{T^+})$ is referred to as its control strategy. Let $\mathcal{H}^b_t$ (resp. $\mathcal{H}^b_{t^+}$) be the set of all possible control laws for virtual player b at time t (resp. t^+) and let $\mathcal{H}^b$ be the set of all possible control strategies for virtual player b. The total cost associated with the game for a strategy profile (χ^a, χ^b) is

$$\mathcal{J}(\chi^a, \chi^b) = \tag{32}$$
$$\mathbb{E}^{(\chi^a, \chi^b)} \Big[\sum_{t=1}^{T} c_t(X^0_t, X_t, U_t, U^a_t) + \rho(X^0_t, X_t) \mathbb{1}_{\{M^{er}_t = 1\}} \Big].$$

where the functions c_t and ρ are the same as in games $\mathscr{G}$ and $\mathscr{G}_s$. In this virtual game, virtual player a aims to maximize the cost while virtual player b aims to minimize the cost. The upper value of Game $\mathscr{G}_e$ is denoted by $S^u(\mathscr{G}_e)$.

B. Common Information Belief and the Dynamic Program

1) Common Information Belief: Before communication at time t, the CIB is given as:

$$\Pi_t(x^0, x, d) = \mathrm{P}[X^0_t = x^0, X_t = x, D_t = d | I^v_t]. \tag{33}$$

After the communication decisions are made and Z^{er}_t is realized, the CIB is given as:

$$\Pi_{t^+}(x^0, x, d) = \mathrm{P}[X^0_t = x^0, X_t = x, D_{t^+} = d | I^v_{t^+}]. \tag{34}$$

The CIB satisfies two key properties: (i) the CIB can be computed without using the virtual players' strategies χ^a and χ^b; (ii) since the adversary does not have any private information at any given time, the CIB does not depend on the adversary's prescriptions (see Section IV-A and Appendix VI of [7]). This can be stated formally as the following lemma.

Lemma 2. *$\Pi_1(x^0_1, x_1, d_1)$ is the belief $\mathrm{P}(X^0_1 = x^0_1, X_1 = x_1, D_1 = d_1)$ and for each $t \geq 1$,*

$$\Pi_{t^+} = \eta_t(\Pi_t, \Gamma_t, Z_{t^+}); \quad \Pi_{t+1} = \beta_t(\Pi_{t^+}, \Lambda_t, Z_{t+1}),$$

where η_t, β_t are fixed transformations derived from the system model using Bayes' rule (see Appendix VI of [7]).

We now describe the dynamic program that provides us with the value of the game $\mathscr{G}$ and an algorithm to compute a min-max strategy for the team.

2) Dynamic Program: Define the value function $V_{T+1}(\pi) := 0$ for all π at time $T + 1$. The cost-to-go functions w_t (resp. w_{t^+}) and value functions V_t (resp. V_{t^+}) for $t = T, \dots, 2, 1$, are defined as follows:

$$w_{t^+}(\pi, \lambda, \gamma^a) :=$$
$$\mathbb{E}\Big[c_t(X^0_t, X_t, U_t, U^a_t) + V_{t+1}(\beta_t(\pi, \lambda, Z_{t+1})) \mid \pi, \lambda, \gamma\Big],$$
$$V_{t^+}(\pi) := \min_{\lambda} \max_{\gamma^a} w_{t^+}(\pi), \tag{35}$$
$$w_t(\pi, \gamma) :=$$
$$\mathbb{E}\Big[\rho(X^0_t, X_t)\mathbb{1}_{\{M^{er}_t = 1\}} + V_{t^+}(\eta_t(\pi^{1,2}, \gamma, Z_{t^+})) \mid \pi, \gamma\Big],$$
$$V_t(\pi) := \min_{\gamma} w_t(\pi, \gamma). \tag{36}$$

Algorithm 1 Strategies f^{i*}, g^{i*} for Player i in the Team

Input: $\Xi_t(\pi), \Xi_{t^+}(\pi)$ obtained from DP for all t and all π
for $t = 1$ **to** T **do**
Before communication:
Current information: C_t, P_t^i {where $C_t = \{C_{(t-1)^+}, Z_t\}$}
Update CIB $\Pi_t = \beta_{t-1}(\Pi_{(t-1)^+}, \Xi^1_{(t-1)^+}(\Pi_{t-1^+}), Z_t)$
{If $t = 1$, Initialize CIB Π_t using C_1}
Get prescription $\Gamma_t = (\Gamma_t^1, \Gamma_t^2) = \Xi_t(\Pi_t)$
Get distribution $\delta M_t^i = \Gamma_t^i(P_t^i)$ and select action $M_t^i = \text{RAND}(\{0, 1\}, \delta M_t^i, K_t^i)$
After communication decisions are made:
Current information: $C_{t^+}, P_{t^+}^i$ {where $C_{t^+} = \{C_t, Z_{t^+}\}$}
Update CIB $\Pi_{t^+} = \eta_t(\Pi_t, \Xi_t^1(\Pi_t), Z_{t^+})$
Get prescription $\Lambda_t = (\Lambda_t^1, \Lambda_t^2) = \Xi_{t^+}(\Pi_{t^+})$
Get distribution $\delta U_t^i = \Lambda_t^i(P_t^i)$ and select action $U_t^i = \text{RAND}(\mathcal{U}_t^i, \delta U_t^i, K_{t^+}^i)$
end for

Let $\Xi_t(\pi)$ (resp. $\Xi_{t^+}(\pi)$) be a minimizer (resp. minmaximizer) of the cost-to-go function in (36) (resp. (35)).

Theorem 1. *The min-max value of games $\mathscr{G}$, $\mathscr{G}_s$ and $\mathscr{G}_e$ are identical, i.e., we have $S^u(\mathscr{G}) = S^u(\mathscr{G}_e) = \mathbb{E}[V_1(\Pi_1)]$. Further, the strategy pair f^*, g^* described by Algorithm 1 is a min-max strategy for the team in the original game $\mathscr{G}$.*

Proof. Because of our assumption on the information structure of Game $\mathscr{G}_s$ (Assumption 2), the evolution of CIB in Game $\mathscr{G}_e$ does not depend virtual player a's prescription. This property allows us to use Theorems 4 and 5 in [7] and obtain our result. □

The dynamic program is helpful for characterizing the min-max value and a min-max strategy in a general setting. However, solving the dynamic program involves computational challenges. The main cause of these challenges is that the private information $(X_t^i \cup D_t)$ space can be very large even after the private information reduction in the simplified game $\mathscr{G}_s$. In [17], we discuss some special cases in which the private information is small or can be reduced further to a manageable size. Once the private information has been reduced sufficiently, one can use the computational methodology discussed in Appendix X of [7] to solve the dynamic program.

V. Conclusions

We considered a zero-sum game between a team of two agents and a malicious agent. The agents can strategically decide at each time whether to share their private information with each other or not. The agents incur a cost whenever they communicate with each other and the adversary may eavesdrop on their communication. Under certain assumptions on the system dynamics and the information structure of the adversary, we characterized a min-max control and communication strategy for the team using a common information belief based min-max dynamic program. For certain specialized information structures, we proved that the agents in the team can ignore a large part of their private information without losing optimality. This reduction in private information substantially simplifies the dynamic program and hence, improves computational tractability.

References

[1] A. Washburn and K. Wood, "Two-person zero-sum games for network interdiction," *Operations research*, vol. 43, no. 2, pp. 243–251, 1995.

[2] R. J. Aumann, M. Maschler, and R. E. Stearns, *Repeated games with incomplete information*. MIT press, 1995.

[3] A. Nayyar, A. Mahajan, and D. Teneketzis, "Optimal control strategies in delayed sharing information structures," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1606–1620, 2010.

[4] ——, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1644–1658, 2013.

[5] A. Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2377–2382, 2013.

[6] S. Sudhakara, D. Kartik, R. Jain, and A. Nayyar, "Optimal communication and control strategies in a multi-agent mdp problem," *arXiv preprint arXiv:2104.10923*, 2021.

[7] D. Kartik, A. Nayyar, and U. Mitra, "Common information belief based dynamic programs for stochastic zero-sum games with competing teams," *arXiv preprint arXiv:2102.05838*, 2021.

[8] D. Szer, F. Charpillet, and S. Zilberstein, "Maa*: A heuristic search algorithm for solving decentralized pomdps," *arXiv preprint arXiv:1207.1359*, 2012.

[9] S. Seuken and S. Zilberstein, "Formal models and algorithms for decentralized decision making under uncertainty," *Autonomous Agents and Multi-Agent Systems*, vol. 17, no. 2, pp. 190–250, 2008.

[10] A. Kumar, S. Zilberstein, and M. Toussaint, "Probabilistic inference techniques for scalable multiagent decision making," *Journal of Artificial Intelligence Research*, vol. 53, pp. 223–270, 2015.

[11] J. S. Dibangoye, C. Amato, O. Buffet, and F. Charpillet, "Optimally solving dec-pomdps as continuous-state mdps," *Journal of Artificial Intelligence Research*, vol. 55, pp. 443–497, 2016.

[12] T. Rashid, M. Samvelyan, C. Schroeder, G. Farquhar, J. Foerster, and S. Whiteson, "Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 4295–4304.

[13] H. Hu and J. N. Foerster, "Simplified action decoder for deep multi-agent reinforcement learning," in *International Conference on Learning Representations*, 2019.

[14] D. Tang, H. Tavafoghi, V. Subramanian, A. Nayyar, and D. Teneketzis, "Dynamic games among teams with delayed intra-team information sharing," *arXiv preprint arXiv:2102.11920*, 2021.

[15] S. Bhattacharya and T. Başar, "Multi-layer hierarchical approach to double sided jamming games among teams of mobile agents," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 5774–5779.

[16] A. Nayyar, A. Gupta, C. Langbort, and T. Başar, "Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 555–570, 2014.

[17] D. Kartik, S. Sudhakara, R. Jain, and A. Nayyar, "Optimal communication and control strategies for a multi-agent system in the presence of an adversary," *arXiv preprint arXiv:2209.03888*, 2022.

[18] D. Tang, "Games in multi-agent dynamic systems: Decision-making with compressed information," Ph.D. dissertation, 2021.

[19] F. Gensbittel, M. Oliu-Barton, and X. Venel, "Existence of the uniform value in zero-sum repeated games with a more informed controller," *Journal of Dynamics and Games (JDG)*, vol. 1, no. 3, pp. 411–445, 2014.