2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 979-8-3503-7770-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IROS58592.2024.10801372

Multi-Robot Communication-Aware Cooperative Belief Space Planning with Inconsistent Beliefs: An Action-Consistent Approach

Tanmoy Kundu[1], Moshe Rafaeli[2] and Vadim Indelman[3]

***Abstract*— Multi-robot belief space planning (MR-BSP) is essential for reliable and safe autonomy. While planning, each robot maintains a belief over the state of the environment and reasons how the belief would evolve in the future for different candidate actions. Yet, existing MR-BSP works have a common assumption that the beliefs of different robots are consistent at planning time. Such an assumption is often highly unrealistic, as it requires prohibitively extensive and frequent communication capabilities. In practice, each robot may have a different belief about the state of the environment. Crucially, when the beliefs of different robots are inconsistent, state-of-the-art MR-BSP approaches could result in a lack of coordination between the robots, and in general, could yield dangerous, unsafe and sub-optimal decisions. In this paper, we tackle this crucial gap. We develop a novel decentralized algorithm that is guaranteed to find a consistent joint action. For a given robot, our algorithm reasons for action preferences about 1) its local information, 2) what it perceives about the reasoning of the other robot, and 3) what it perceives about the reasoning of itself perceived by the other robot. This algorithm finds a consistent joint action whenever these steps yield the same best joint action obtained by reasoning about action preferences; otherwise, it self-triggers communication between the robots. Experimental results show efficacy of our algorithm in comparison with two baseline algorithms.**

I. Introduction

Multi-robot decision making under uncertainty in partially observable domains is a fundamental problem in robotics and AI, with numerous applications where multiple agents operate in the same environment. Examples include autonomous driving, environmental monitoring, and search and rescue missions. Such problems are often formulated within the framework of Decentralized Partially Observed Markov Decision Process (Dec-POMDP) or multi-robot Belief Space Planning (MR-BSP).

Multi-robot decision making under uncertainty and belief space planning have been investigated in recent years under different perspectives. Calculating a globally optimal solution of the underlying problem is computationally intractable [6]. Nevertheless, progress has been made considering a decentralized general-purpose paradigm, Dec-POMDP, and macro-actions (e.g. [1], [2], [3], [7], [15], [16]).

[1]Tanmoy Kundu is with the department of Computer Science and Engineering, IIIT Delhi, India. This work was carried out when he was with the department of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel. tanmoy.kundu@iiitd.ac.in.

[2]Moshe Rafaeli is with the Technion Autonomous Systems Program (TASP), Technion - Israel Institute of Technology, Haifa 32000, Israel. mosh305@campus.technion.ac.il.

[3]Vadim Indelman is with the department of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel. vadim.indelman@technion.ac.il.

This research was supported by the Israel Science Foundation (ISF).

Approaches that consider a non-cooperative setting, where each robot has its own reward function (representing a different task), typically tackle the problem within the framework of dynamic games by reasoning about the Nash equilibrium of the multi-robot system (e.g. [14], [18], [20]), or by leveraging multi-objective optimization (e.g. [12]). Multi-Robot BSP (MR-BSP) approaches have been also investigated considering a cooperative setting, i.e. all robots have the same reward function (same task). For instance, the works [4], [10], [17] consider MR-BSP with Gaussian high-dimensional distributions in the context of cooperative active SLAM and active inference.

Yet, a prevailing assumption in existing approaches, being explicit or implicit, is that the beliefs of different robots at planning time are *consistent*, i.e. conditioned on the *same* information. Such an assumption requires all the data (observations) captured by different robots to be available to each robot, such that the beliefs of individual robots can be conditioned on the same data. This, in turn, requires prohibitively extensive and frequent communication capabilities. However, in numerous problems and scenarios, extensive data sharing between the robots cannot be made on a regular basis. Moreover, it is often the case that only a partial or compressed version of the data can be communicated in practice. As a result, each robot may have access to different data, which would correspond to a different belief about the state of the environment, i.e. to inconsistent beliefs.

Crucially, when the beliefs of different robots are inconsistent, the state-of-the-art MR-BSP approaches could result in a lack of coordination between the robots, and in general, could yield dangerous, unsafe, and sub-optimal decisions. For instance, consider the toy example shown in Figure 1, where two robots aim to reach a common goal point without colliding with each other. Due to lack of communication, the robots beliefs become inconsistent; MR-BSP in such a setting may result in each robot calculating a different joint action, which can lead to a collision (Figure 1(c)).

Despite this, to our knowledge, multi-robot planning with inconsistent beliefs has not been explicitly addressed thus far. Arguably, the closest work to our paper is [22], where the authors explicitly consider the beliefs of different robots may be inconsistent. However, that work requires communication whenever the beliefs of different robots are detected to be inconsistent, so that, eventually, planning is done with consistent beliefs.

In this paper we address this crucial gap of MR-BSP with inconsistent beliefs. To our knowledge, this work is the first to address this gap. Specifically, we develop a novel

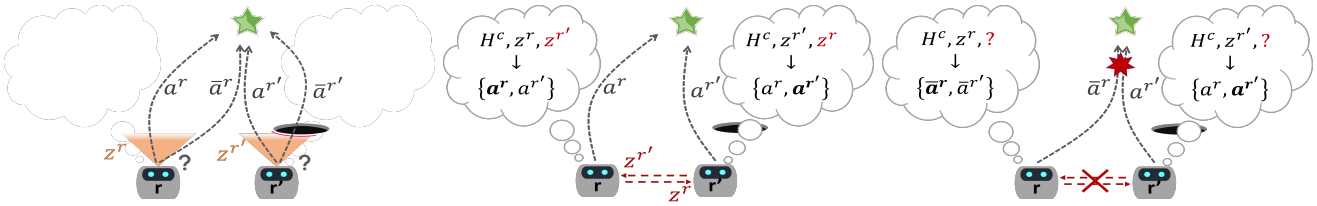


Fig. 1. (a) Two robots, r and r', acquire separate observations (z^r, $z^{r'}$) and begin a planning session. The robots aim to cooperatively reach the green star while satisfying a safety property of avoiding obstacles and collisions; each agent has two candidate actions ($\{a^r, \bar{a}^r\}$ for robot r, and $\{a^{r'}, \bar{a}^{r'}\}$ for robot r'). (b) The robot communicate their observations to each other (red color), after which their histories, and beliefs, become consistent. Decentralized MR-BSP in such case yields the same best joint action for both robots. Here, H^c represents the common history between the robots prior to acquiring the observations z^r and $z^{r'}$. (c) The robots do not communicate their observations, and as a result, the robots' beliefs become inconsistent. This causes the robots to conclude inconsistent joint-actions, which leads to a collision.

decentralized MR-BSP framework that explicitly accounts for inconsistent beliefs and self-triggers communication only when the same joint action selection by different robots cannot be guaranteed.

At the core of our proposed approach is the notion of *action consistency* [8], [9], [11], which captures the observation that decision making involves identifying which action is preferable over other actions: If in two decision making problems a certain action is preferred over other candidate actions, then this action would be identified as the best action in both problems, regardless of the actual objective function values. In such case, the two problems are action-consistent. Thus far, this concept has been considered for *simplification* with performance guarantees of *single-robot* POMDP and BSP problems (see, e.g., [5], [13], [19], [21], [23]).

Our key observation is that when the robots perform MR-BSP with inconsistent beliefs, each robot solves a *different* planning problem; nevertheless, these problems can still yield the *same* best joint action if they are action-consistent. Leveraging this key observation, in this paper we develop an approach that detects if a consistent decision making among robots can be guaranteed albeit the inconsistent beliefs. This involves reasoning about the missing information (observations) and the corresponding beliefs of the other robot(s). In the case where an action-consistent decision-making cannot be guaranteed, our approach self-triggers communication of the information that will eventually lead to an action-consistent decision-making. The communications are self-triggered because each robot reasons about when it has to initiate a communication without being triggered by any other robot.

To summarize, our main contributions in this paper are: (a) we introduce a formulation of a new problem, i.e. MR-BSP with inconsistent beliefs; (b) we develop a novel approach to address this problem by leveraging the concept of action consistency and extending it to the multi-robot setting. A key innovation here is that we can often have the same joint action selection calculated by different robots, despite having inconsistent beliefs, even without any communication. Otherwise, communications are self-triggered to ensure action consistency. We provide a theoretical guarantee that our approach will eventually identify a consistent joint action for the robots. (c) we benchmark our approach in simulation and compare it against two baseline approaches.

II. Preliminaries & Problem Formulation

A. Preliminaries

Consider a team of u robots $\Gamma = \{r_1, r_2, \ldots, r_u\}$ performing some task(s). We pick a robot r arbitrarily from Γ, and denote by $-r$ the rest of the robots in the group. From the perspective of robot r, we define a *decentralized* multi-robot POMDP as a 7-tuple: $\langle \mathcal{X}, \mathcal{Z}, \mathcal{A}, T, O, \rho^r, b_k^r \rangle$. Here, $\mathcal{X}$ is the application-dependent joint state space, and $\mathcal{Z}$ and $\mathcal{A}$ are the joint observation and joint action spaces. $T(x' \mid x, a)$ and $O(z \mid x)$ are the joint transition and observation models, where $x \in \mathcal{X}$, $a \in \mathcal{A}$ and $z \in \mathcal{Z}$ are, respectively, the joint state, action and observation. Furthermore, we assume that the observations of different robots are independent conditioned on the state, i.e. $O(z \mid x) = \prod_{r \in \Gamma} O^r(z^r \mid x)$ where $z^r \in \mathcal{Z}^r$ is the local observation of robot r, and $\mathcal{Z}^r$ and $O^r(.)$ are the corresponding observation space and model. ρ^r is a general belief-dependent reward function $\rho^r : \mathcal{B} \times \mathcal{A} \mapsto \mathbb{R}$, where $\mathcal{B}$ is the belief space. In this work we consider a *cooperative* setting, i.e. each robot has the same reward function ρ that describes a joint task allocated to the group (e.g. information gathering). Therefore, $\rho^r(b, a) = \rho^{r'}(b, a)$ for any $r, r' \in \Gamma$.

We denote by b_k^r the belief of robot r at time k over the state $x_k \in \mathcal{X}$,

$$b_k^r[x_k] \triangleq \mathbb{P}(x_k \mid H_k^r), \tag{1}$$

where $H_k^r \triangleq \{a_{0:k-1}, z_{1:k}^r, z_{1:k}^{-r}\}$ is the history available to robot r at time k, which includes its own actions and observations, as well as those of other robots in the group.

In a *collaborative* setting robot r reasons over the joint actions of the robots, instead of its individual actions. The joint action a_ℓ at any time ℓ is defined as $a_\ell \triangleq (a_\ell^{r_1}, \ldots, a_\ell^{r_u}) = (a_\ell^r, a_\ell^{-r}) \in \mathcal{A}^{r_1} \times \cdots \times \mathcal{A}^{r_u} \equiv \mathcal{A}$, where $\mathcal{A}^r$ is the individual action space of robot $r \in \Gamma$.

For ease of exposition we shall consider an open loop setting, although this is not a limitation of our proposed concept. Let $\mathcal{A}_{k+} = \mathcal{A}_{k:k+L-1} = \{a_{k:k+L-1}\}$ denote a set of L-step joint action sequences formed from the joint action space $\mathcal{A}_{k+}$. Under these assumptions, the objective function of robot r for a horizon of L time steps and a candidate joint

action sequence $a_{k+} \triangleq a_{k:k+L-1} \in \mathcal{A}_{k+}$ is defined as

$$J(b_k^r, a_{k+}) = \mathop{\mathbb{E}}_{z_{k+1:k+L}}[\sum_{l=0}^{L-1} \rho(b_{k+l}^r, a_{k+l}) + \rho(b_{k+L}^r)], \quad (2)$$

where the expectation is over future observations $z_{k+1:k+L}$ of all robots in the group with respect to the distribution $\mathbb{P}(z_{k+1:k+L} \mid b_k^r, a_{k+})$. The optimal joint action sequence is:

$$a_{k+}^* = \mathop{\arg\max}_{a_{k+} \in \mathcal{A}_{k+}} J(b_k^r, a_{k+}). \quad (3)$$

In this paper we use the terms "action sequence" and "action" interchangeably.

B. Problem Formulation

A typical assumption in existing multi-robot belief space planning approaches is that of *consistent histories* across all the robots in Γ at any planning time instant k, i.e.

$$\forall r, r' \in \Gamma, \quad H_k^r \equiv H_k^{r'}, \quad (4)$$

which corresponds to the assumption that each robot has access to the observations of all other robots. Yet, in numerous real world problems and scenarios, such an assumption is clearly unrealistic (see Section I).

We shall use the term *inconsistent beliefs* whenever (4) is not satisfied[1]. If any two robots r and r' have inconsistent beliefs, b_k^r and $b_k^{r'}$, their theoretical objective function values (2) for the same joint action a_{k+} are not necessarily the same. There are two reasons for this. First, the expectation in (2) is taken with respect to two different distributions i.e. $\mathbb{P}(z_{k+1:k+L} \mid b_k^r, a_{k+})$ and $\mathbb{P}(z_{k+1:k+L} \mid b_k^{r'}, a_{k+})$. Second, even when conditioned on the same realization of a future observation sequence $z_{k+1:k+l}$ for any time step $l \in [1, L]$, the theoretical posterior future beliefs b_{k+l}^r and $b_{k+l}^{r'}$ are still inconsistent, and hence their corresponding rewards are different.

As a result, it is no longer guaranteed that different robots will indeed be coordinated on the theoretical level as the optimal joint action to be identified by robots r and r' are no longer necessarily identical. In other words, generally, $\arg\max_{a_{k+}} J(b_k^r, a_{k+}) \neq \arg\max_{a_{k+}} J(b_k^{r'}, a_{k+})$. Such a situation is clearly undesired as it may lead to sub-optimal planning performance, and to dangerous, unsafe decision making.

Specifically, consider any two robots $r, r' \in \Gamma$. In a limited communication setting, at time k, consider that the last time instant when the beliefs of these two robots were consistent is $p \in [1, k]$ time steps behind k. In other words, at time instant $k-p$, robots r and r' communicated with each other, resulting in $b_{k-p}^r = \mathbb{P}(x_{k-p} \mid H_{k-p}^r)$ and $b_{k-p}^{r'} = \mathbb{P}(x_{k-p} \mid H_{k-p}^{r'})$ with $H_{k-p}^r = \{a_{0:k-p-1}, z_{1:k-p}^r, z_{1:k-p}^{-r}\} \equiv \{a_{0:k-p-1}, z_{1:k-p}^{r'}, z_{1:k-p}^{-r'}\} = H_{k-p}^{r'}$. $H_{k-p} \triangleq H_{k-p}^r = H_{k-p}^{r'}$ for any $r, r' \in \Gamma$. In particular, if there are only two robots in the group, then $H_{k-p} = \{a_{0:k-p-1}, z_{1:k-p}^r, z_{1:k-p}^{r'}\}$.

During time period $[k-p+1, k]$, there was no communication and any robots $r, r' \in \Gamma$ do not have access to the non-local observations from these time instances. In other words, their beliefs

$$b_k^r = \mathbb{P}(x_k \mid H_k^r), \quad b_k^{r'} = \mathbb{P}(x_k \mid H_k^{r'}), \quad (5)$$

are inconsistent since robot r does not have access to $z_{k-p+1:k}^{-r}$, and robot r' does not have access to $z_{k-p+1:k}^{-r'}$. Currently we assume the actions performed by each robot by time instant k are known. Thus, $H_k^r \neq H_k^{r'}$ for any two robots $r, r' \in \Gamma$, where

$$H_k^r = H_{k-p}^r \cup \{a_{k-p:k-1}, z_{k-p+1:k}^r\}, \quad (6)$$

$$H_k^{r'} = H_{k-p}^{r'} \cup \{a_{k-p:k-1}, z_{k-p+1:k}^{r'}\}. \quad (7)$$

The challenge addressed in this paper is to select the same (consistent) joint action sequence a_{k+}^* for all the robots in the group even though their beliefs are inconsistent.

III. Approach

Our key objective is to guarantee action consistency for multiple robots with inconsistent beliefs. With inconsistent beliefs of the robots, the J-values for a given joint action evaluated by different robots, each with its own belief, will generally be different. If we can guarantee the *same preference ordering* of the candidate joint actions derived by all the robots, then we yield a consistent best joint action regardless of the magnitude of the corresponding J-values. This is in striking contrast to existing approaches that implicitly ensure multi-robot action consistency (MR-AC) by requiring the robots to have consistent beliefs, i.e. assuming all the data between robots are communicated.

Specifically, we propose to utilize the concept of action consistency to address multi-robot decision making problems with inconsistent beliefs. We extend the definition of action consistency considering a multi-robot setting.

Definition 3.1 (Multi-Robot Action Consistency (MR-AC)): Consider two robots $r, r' \in \Gamma$ where $r \neq r'$. At time k, the joint actions selected by r and r' are $a \in \mathcal{A}_{k+}$ and $a' \in \mathcal{A}_{k+}$ respectively. Robots r and r' are *action consistent* at time k if and only if $a = a'$. If, at time k, action consistency is satisfied for any two robots $r, r' \in \Gamma$, then the system of robots Γ is action consistent at that time.

A. Action preferences with different beliefs

We use the notion of comparing J-values where the *order* of the values matters, and not the magnitude. We define *action preference* as a binary relation $\succcurlyeq$. Consider two joint actions $a, a' \in \mathcal{A}_{k+}$. The joint action a is preferred over a' w.r.t. a given set of beliefs $\mathcal{B}_Z$ when action a dominates a' for all the beliefs in $\mathcal{B}_Z$:

$$\forall b \in \mathcal{B}_Z \quad a \succcurlyeq a' \iff J(b, a) \geq J(b, a'). \quad (8)$$

While this is valid for any set of beliefs $\mathcal{B}_Z$, in this paper we shall consider *a given set of observations* Z, and each

[1]Note that in nonparametric inference methods beliefs are generally inconsistent also given consistent histories (as given the same history the belief could be represented by different sets of particles). We leave the extension of our approach to such a setting to future work, and consider herein deterministic inference methods.

belief $b \in \mathcal{B}_Z$ results from Bayesian inference considering a particular observation $z \in Z$. For m number of joint actions in $\mathcal{A}$, the joint action $a^* \in \mathcal{A}$ is most preferred if $a^* \succcurlyeq a$ holds for all $a \in \mathcal{A}$.

Definition 3.2 (Consistent observations): Consider a set of observations Z, a set of joint actions $\mathcal{A}_{k+}$ and an objective function $J(.)$. If there exists an $a^* \in \mathcal{A}_{k+}$ satisfying $a^* \succcurlyeq a$ for all $a \in \mathcal{A}_{k+}$, we call Z to be consistent in favor of a^*. We denote the consistency of Z favoring action a^* as $\texttt{con}_{a^*}(Z) = \texttt{true}$.

Define $\texttt{cobs}_{a^*}(Z)$ as the consistent set of observations in Z favoring action a^*:

$$\texttt{cobs}_{a^*}(Z) = \{z \in Z' |\, Z' \subseteq Z \ \wedge \ \texttt{con}_{a^*}(Z') = \texttt{true}\}. \quad (9)$$

When $\texttt{con}_{a^*}(Z) = \texttt{true}$, $\texttt{cobs}_{a^*}(Z)$ contains the entire Z because all the observations in Z are consistent in favor of a^*. When $\texttt{con}_{a^*}(Z) = \texttt{false}$, $\texttt{cobs}_{a^*}(Z)$ contains a proper subset of observations $Z' \subset Z$ (instead of the entire Z) consistent in favor of a^*.

Next, we present our approach to check if MR-AC exists despite the robots having inconsistent beliefs, and then describe a mechanism for self-triggered communications until MR-AC is achieved.

B. MR-AC for robots with inconsistent beliefs

Consider a group of two robots $\Gamma = \{r, r'\}$. In a decentralized setting, we propose a mechanism to ensure MR-AC for Γ from the perspective of an arbitrarily chosen robot $r \in \Gamma$. The aim of r is to select, at planning time k, a joint action which is necessarily the same with the one chosen by robot r', though r and r' have inconsistent beliefs b_k^r and $b_k^{r'}$ from (5).

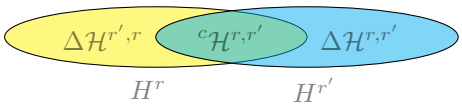


Fig. 2. Illustration of H^r, $H^{r'}$, ${}^c\mathcal{H}$, $\Delta\mathcal{H}^{r,r'}$, and $\Delta\mathcal{H}^{r',r}$. See text for details.

Though robots r and r' have inconsistent histories at planning time k, they have a common part of history that we shall denote as ${}^c\mathcal{H}_k^{r,r'} \triangleq H_k^r \cap H_k^{r'}$. Accordingly, we define by $\Delta\mathcal{H}_k^{r,r'} \triangleq H_k^{r'} \setminus {}^c\mathcal{H}_k^{r',r}$ the part in history of robot r', i.e. an observation sequence, that is *unavailable* to robot r. As discussed below, robot r will have to reason about these missing observations of robot r'. Similarly we define $\Delta\mathcal{H}_k^{r',r}$. Therefore, $H_k^r = \{{}^c\mathcal{H}_k^{r,r'}, \Delta\mathcal{H}_k^{r',r}\}$ and $H_k^{r'} = \{{}^c\mathcal{H}_k^{r,r'}, \Delta\mathcal{H}_k^{r,r'}\}$ as illustrated in Figure 2.

The beliefs (5) can then be expressed as:

$$\begin{aligned} b_k^r &= \mathbb{P}(x_k \mid {}^c\mathcal{H}_k^{r,r'}, \Delta\mathcal{H}_k^{r',r}) \\ b_k^{r'} &= \mathbb{P}(x_k \mid {}^c\mathcal{H}_k^{r',r}, \Delta\mathcal{H}_k^{r,r'}). \end{aligned} \quad (10)$$

Recall we assumed that the robots have consistent histories until time $k-p$. Initially, as no communication was triggered at planning time k, according to (6) and (7), ${}^c\mathcal{H}_k^{r,r'} = \mathcal{H}_{k-p} \cup \{a_{k-p:k-1}\}$, $\Delta\mathcal{H}_k^{r,r'} = \{z_{k-p+1:k}^{r'}\}$ and $\Delta\mathcal{H}_k^{r',r} = \{z_{k-p+1:k}^{r}\}$.

We propose the following steps to identify MR-AC by robot r despite inconsistent beliefs b_k^r and $b_k^{r'}$ (10) of the robots. Conceptually, robot r needs to analyze the joint action preferences from different perspectives: i) its own perspective, ii) perspective of the other robot r', and iii) its own perspective reasoned by the other robot r'. If r finds the same best joint action from all the above perspectives, then it can be assured that the other robot has also calculated the same best joint action; hence, in such case, both robots are action-consistent, i.e. choose the same joint action.

Since robot r does not have access to $\Delta\mathcal{H}_k^{r,r'}$, and it is aware that robot r' does not have access to $\Delta\mathcal{H}_k^{r',r}$, these steps involve reasoning about all possible values of these missing observations. We denote the corresponding joint observation spaces, that represent these possible values, by $\Delta\mathcal{Z}_k^{r,r'}$ and $\Delta\mathcal{Z}_k^{r',r}$. Thus, $\Delta\mathcal{H}_k^{r,r'} \in \Delta\mathcal{Z}_k^{r,r'}$ and $\Delta\mathcal{H}_k^{r',r} \in \Delta\mathcal{Z}_k^{r',r}$. Also, define $\Delta\mathcal{Z}_k$ as

$$\Delta\mathcal{Z}_k = \Delta\mathcal{Z}_k^{r,r'} \cup \Delta\mathcal{Z}_k^{r',r}. \quad (11)$$

Prior to any communication self-triggered by our algorithm, since $\Delta\mathcal{H}_k^{r,r'} = \{z_{k-p+1:k}^{r'}\}$ and $\Delta\mathcal{H}_k^{r',r} = \{z_{k-p+1:k}^{r}\}$,

$$\Delta\mathcal{Z}_k^{r,r'} = \mathcal{Z}_{k-p+1}^{r'} \times \mathcal{Z}_{k-p+2}^{r'} \times \ldots \times \mathcal{Z}_k^{r'} \quad (12)$$

$$\Delta\mathcal{Z}_k^{r',r} = \mathcal{Z}_{k-p+1}^{r} \times \mathcal{Z}_{k-p+2}^{r} \times \ldots \times \mathcal{Z}_k^{r}. \quad (13)$$

In this paper we assume these observation spaces to be discrete. We leave the extension to continuous observation spaces to future work.

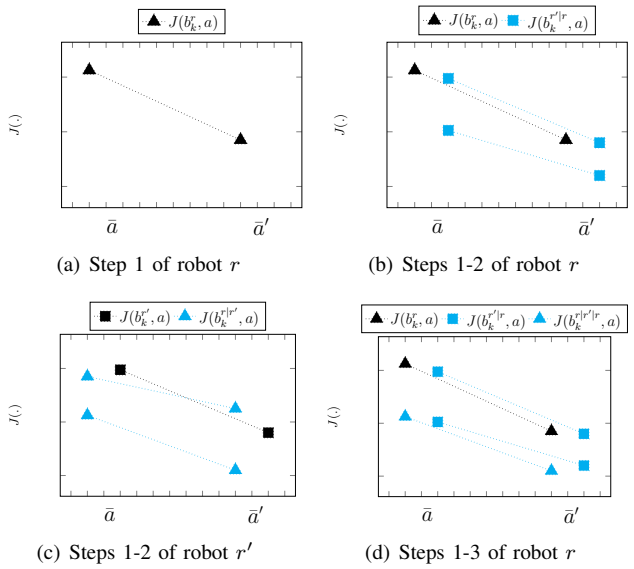
C. Algorithm VERIFYAC

We present an algorithm named VERIFYAC that verifies MR-AC from the perspective of robot r. The algorithm captures the above concept and we now present it in detail. Figure 3 illustrates *conceptually* the mentioned steps in a toy example that has only two possible joint actions and two possible observations for each robot. The steps of VERIFYAC are described below.

Step 1: *Robot r calculates the best joint action given its own belief b_k^r via (3):* This involves evaluation of the objective function $J(b_k^r, \bar{a})$ for different candidate joint actions in $\mathcal{A}_{k+}$. In other words, this involves evaluation of the objective function considering the belief b_k^r from (10) which is conditioned on the consistent history ${}^c\mathcal{H}_k^{r,r'}$ and on the actual local observation(s) $\Delta\mathcal{H}_k^{r',r}$ of robot r. Finally, robot r selects the best action $\bar{a} \in \mathcal{A}_{k+}$ such that $J(b_k^r, \bar{a}) > J(b_k^r, \bar{a}') \ \ \forall \bar{a}' \in \mathcal{A}_{k+}$ and $\bar{a} \neq \bar{a}'$. The concept is illustrated in Figure 3(a) by black triangles.

However, since robots r and r' have inconsistent beliefs, it is not guaranteed that the joint action chosen by robot r will be the same as chosen by r'. So, we move to Step 2.

Step 2: *Robot r mimics the reasoning done by robot r':* The belief $b_k^{r'}$ (10) of the other robot r' is conditioned on $\Delta\mathcal{H}_k^{r,r'}$ (initially, $\Delta\mathcal{H}_k^{r,r'} = \{z_{k-p+1:k}^{r'}\}$) which is unavailable to robot r. Moreover, $b_k^{r'}$ is not conditioned on the actual observation of robot r, i.e. $\Delta\mathcal{H}_k^{r',r}$ (initially $z_{k-p+1:k}^{r}$), which



(a) Step 1 of robot r (b) Steps 1-2 of robot r

(c) Steps 1-2 of robot r' (d) Steps 1-3 of robot r

Fig. 3. Illustration of VERIFYAC from the perspective of robot r. Robots r and r' have inconsistent beliefs b_k^r and $b_k^{r'}$ at time k. Candidate joint actions are $\bar{a}$ and $\bar{a}'$. Triangles and squares denote objective function ($J(.)$) evaluations for r and r' respectively. (a) Step 1 of r: Robot r computes its belief for its actual observation. Chooses $\bar{a}$ as the best action. (b) Step 1-2 of r: In Step 2, r computes $J(.)$ for each possible observation of r'. All the observations are consistent in favor of $\bar{a}$. (c) Step 1-2 of r': Similarly, robot r' computes Step 1 for its actual observation, and Step 2 for all possible observations of r. (d) Step 3 of r: Combines (a)-(c) and verifies that the observations at each step are consistent in favor of action $\bar{a}$. Hence, r can be assured that r' also has chosen $\bar{a}$. Thus r chooses action $\bar{a}$ at time k.

is unavailable to robot r'. Explicitly, the two beliefs are given by (10).

As $\Delta\mathcal{H}_k^{r,r'}$ is unavailable to robot r, it has now to reason over all the possible observation realizations in $\Delta\mathcal{Z}_k^{r,r'}$ of robot r'. For instance, initially, prior to self-triggered communication, due to (13), this corresponds to all the possible realizations of observation sequences of robot r' between time instances $k-p+1$ and k (one of which is the actually captured sequence of observations, $z_{k-p+1:k}^r$).

Robot r verifies the consistency of observations $\Delta\mathcal{Z}_k^{r,r'}$, i.e. $\texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k^{r,r'}) = \texttt{true}$, in favor of the action $\bar{a}$ derived in Step 1. For each such possible realization, denoted abstractly by $\tilde{z}^{r'} \in \Delta\mathcal{Z}_k^{r,r'}$, robot r first constructs a plausible corresponding belief of robot r', denoted as

$$b_k^{r'|r}(\tilde{z}^{r'}) \triangleq \mathbb{P}(x_k \mid {}^c\mathcal{H}_k^{r,r'}, \tilde{z}^{r'}). \tag{14}$$

Note the belief is still over the state x_k, and it varies for different values of $\tilde{z}^{r'}$ it is conditioned upon.

In practice, the belief (14) can be calculated in a Bayesian manner either by down-dating the observations $\Delta\mathcal{H}_k^{r',r}$ (initially $z_{k-p+1:k}^r$) from b_k^r (10) and updating with $\tilde{z}^{r'}$, or equivalently, directly from $\mathbb{P}(x_k \mid {}^c\mathcal{H}_k^{r,r'})$ which would have to be maintained. For instance, for $p=1$, we get $b_k^{r'|r}(\tilde{z}^{r'}) = b_k^r \frac{\mathbb{P}(z_k^r|{}^c\mathcal{H}_k^{r,r'})\mathbb{P}(\tilde{z}_k^{r'}|x_k)}{\mathbb{P}(z_k^r|x_k)\mathbb{P}(\tilde{z}_k^{r'}|{}^c\mathcal{H}_k^{r,r'})} = \mathbb{P}(x_k \mid {}^c\mathcal{H}_k^{r,r'}) \frac{\mathbb{P}(\tilde{z}_k^{r'}|x_k)}{\mathbb{P}(\tilde{z}_k^{r'}|{}^c\mathcal{H}_k^{r,r'})}$.

Then, for each $\tilde{z}^{r'} \in \Delta\mathcal{Z}_k^{r,r'}$ of r', robot r evaluates the objective function $J(b_k^{r'|r}(\tilde{z}^{r'}), a)$ for different candidate joint actions $a \in \mathcal{A}_{k+}$. This is illustrated in Figure 3(b) using blue squares. Generally, each $\tilde{z}^{r'} \in \Delta\mathcal{Z}_k^{r,r'}$ yields its own J-values. Moreover, we do not necessarily expect either of these values to match the objective values $J(b_k^r, \bar{a})$ calculated by robot r in Step 1 (black triangles), since generally the observation models and spaces of different robots could vary. Importantly, with this formulation, the actual observation that robot r' captured will be considered, since $\Delta\mathcal{H}_k^{r,r'} \in \Delta\mathcal{Z}_k^{r,r'}$.

Regardless of the magnitude of J-values, it may happen that for all $\tilde{z}^{r'}$ the same joint action is chosen, and that action is identical to the one chosen in Step 1. Such a situation is depicted in Figure 3(b) where $\bar{a}$ is the best joint action in both steps 1 and 2. In other words, in this scenario, regardless of what the actual observation of r' is, robot r can be assured that when r' performs its own decision making, i.e. step 1, it will necessarily choose the same joint action as the one chosen by r. Therefore, r checks if the best action selected for each $\tilde{z}^{r'} \in \Delta\mathcal{Z}_k^{r,r'}$ is the action $\bar{a}$ derived in Step 1, i.e. if $\texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k^{r,r'}) = \texttt{true}$ holds. Thus, r captures the reasoning about the action selection by r'.

Yet, at this point, robot r cannot guarantee that robot r' will also reach the same conclusion, i.e. regardless of the actual observation of robot r (that is unavailable to robot r'), the same joint action will be selected by robots r' and r. Therefore, Steps 1 and 2 are insufficient to guarantee MR-AC between the two robots, which brings us to Step 3.

Step 3: *Robot r mimics the reasoning done by robot r' that mimics the reasoning done by robot r:* Robot r', on its side, similarly performs Steps 1 and 2. In Step 1, r' evaluates the objective function for different candidate joint actions based on its own belief $b_k^{r'}$ conditioned on ${}^c\mathcal{H}_k^{r',r}$ and $\Delta\mathcal{H}_k^{r,r'}$ (see (10)). Refer to Figure 3(c). Since $\Delta\mathcal{H}_k^{r,r'} \in \Delta\mathcal{Z}_k^{r,r'}$, this calculation will be considered by robot r as a part of its Step 2.

Robot r' also performs Step 2, on its side, in which r' reasons about all possible observations of r, i.e. $\Delta\mathcal{Z}_k^{r',r}$. Robot r' thus calculates $b_k^{r|r'}(\tilde{z}^r)\ \forall \tilde{z}^r \in \Delta\mathcal{Z}_k^{r',r}$. This is illustrated in Figure 3(c) by blue triangles. Now, if robot r, on its side, mimics this reasoning done by r', robot r can perceive what r' thinks about the reasoning done by r.

Put formally, robot r verifies if all observations in $\Delta\mathcal{Z}_k^{r',r}$ are in favor of the action $\bar{a}$ derived in Step 1 of r, i.e. $\texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k^{r',r}) = \texttt{true}$. So, r computes $b_k^{r|r'|r}(\tilde{z}^r)$,

$$b_k^{r|r'|r}(\tilde{z}^r) \triangleq \mathbb{P}(x_k \mid {}^c\mathcal{H}_k^{r',r}, \tilde{z}^r), \tag{15}$$

and evaluates $J(b_k^{r|r'|r}(\tilde{z}^r), a)$ for each $\tilde{z}^r \in \Delta\mathcal{Z}_k^{r',r}$ and for all candidate joint actions $a \in \mathcal{A}_{k+}$. Thus, r captures the reasoning about the action selection by itself reasoned by r'.

Combining Steps 1-3, robot r checks for MR-AC by reasoning about selecting a consistent joint action by r and r', which involves considering the observations in $\Delta\mathcal{Z}_k$ (defined in (11)). When the same joint action $\bar{a}$ is chosen in Steps 1-3, as illustrated in Figure 3(d), MR-AC is identified by robot r using VERIFYAC. Thus, despite having inconsistent beliefs, robots r and r' are action consistent, i.e. identify the same joint action chosen by both the robots at time k.

Theorem 3.3: Steps 1-3 of VERIFYAC are necessary and sufficient for any robot r to find MR-AC, if MR-AC exists,

with the robots in $\Gamma = \{r, r'\}$ having inconsistent beliefs.

Proof: Steps 1-3 are sufficient: In Steps 1-3 of VERIFYAC, robot r analyzes the observation spaces of the robots $\Delta\mathcal{Z}_k$ (defined in (11)) as per the algorithm. Among the observations in $\Delta\mathcal{Z}_k$ we have the actual local observations $\Delta\mathcal{H}_k^{r,r'}$ and $\Delta\mathcal{H}_k^{r',r}$. When MR-AC exists, $\texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k)$ becomes $\texttt{true}$ in favor of some joint action $\bar{a}$. This implies that the J-values corresponding to the actual observation values are also consistent in favor of $\bar{a}$. We know that the actual observations give the joint action preferences with zero uncertainty. Therefore, $\texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k) = \texttt{true}$ in Steps 1-3 implies MR-AC for robots r and r'.

Steps 1-3 are necessary: Without communication, robot r is not aware of the actual observation $\Delta\mathcal{H}_k^{r,r'}$ of the other robot r'. In Steps 1-3 of our algorithm, robot r exhaustively considers all observations in $\Delta\mathcal{Z}_k$ that include the actual observations of both the robots in Γ. The actual observations give the joint action preferences of the robots with zero uncertainty. If we remove a randomly selected observation $z \in \Delta\mathcal{Z}_k$ in any of the steps in VERIFYAC, there is a possibility of z being the actual observation of a robot. This does not guarantee selecting the joint action preference correctly. Hence, Steps 1-3 are *necessary* to verify MR-AC. ∎

However, it may often happen that for given beliefs b_k^r and $b_k^{r'}$, an MR-AC does not exist, i.e. after performing Steps 1-3 of VERIFYAC we cannot find a joint action $\bar{a}$ which is chosen by all the steps of VERIFYAC. Mathematically,

$$\nexists \bar{a} \in \mathcal{A}_{k+} \quad \texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k) = \texttt{true}. \tag{16}$$

In the next section, we discuss such scenarios when MR-AC is not satisfied, and describe our approach to initiate different communications (COMMs) until MR-AC is enforced.

D. Self-triggered decision for communication

We present an algorithm ENFORCEAC that enforces MR-AC via COMMs when VERIFYAC fails to find MR-AC. Each robot assesses the requirement of a COMM by its own reasoning and it *self-triggers* a COMM whenever needed.

When (16) holds, a COMM is required. A COMM can send a local observation, either from r to r', from r' to r, or in both directions, based on the conditions specified below.

Let $\bar{a}$ be the best joint action calculated by robot r in Step 1. Robot r reasons about necessity of a COMM from itself to robot r' and self-triggers the COMM if

- From the perspective of r, the observations in Step 3 of VERIFYAC are *not consistent* in favor of the same action $\bar{a}$, i.e. $\texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k^{r',r}) = \texttt{false}$.

In such a case, robot r deduces that robot r', which reasons about observations of robot r by considering the observation space $\Delta\mathcal{Z}_k^{r',r}$ (as part of Step 2 of r'), will find some inconsistent observation realizations of r in $\Delta\mathcal{Z}_k^{r',r}$. Therefore, r sends its local observation(s) from $\Delta\mathcal{H}_k^{r',r}$ to r'.

Additionally, a COMM from r to r' will be triggered if:

- Step 2 of r gives $\texttt{con}_{a'}(\Delta\mathcal{Z}_k^{r',r}) = \texttt{true}$ where $a' \neq \bar{a}$ ($\bar{a}$ is the action chosen in Step 1 of r).

In this case, r perceives that the observations in Step 2 are consistent in favor of a', though a' does not match with the best action $\bar{a}$ in Step 1 of r. Also, r does not know whether r' detects the same inconsistency on its side (comparing Step 1 and Step 3 of r'). However, it is required to modify the action preferences in Step 2 of r via COMM. So, r sends its local observation to r'. This COMM modifies ${}^c\mathcal{H}_k^{r,r'}$ which, in turn, modifies the action preferences in Step 2 of r.

Similarly, robot r reasons about a COMM from r' to r if:

- From the perspective of r, the observations in step 2 of VERIFYAC are not consistent in favor of action $\bar{a}$, i.e. $\texttt{con}_{\bar{a}}(\Delta\mathcal{Z}_k^{r,r'}) = \texttt{false}$.

Due to the inconsistent observations in Step 2, robot r needs access to more observations of r'; in other words, robot r understands the necessity of a COMM from r'. For robot r, one possibility is to ask r' to initiate a COMM. However, r knows that the COMM from r' to r will happen automatically without any intervention by r. This is because robot r deduces that robot r', on its side, analyzes the necessity of the same COMM. That is, when r' will execute its Step 3 it will find inconsistent observations of r and then r' will trigger a COMM from itself to r.

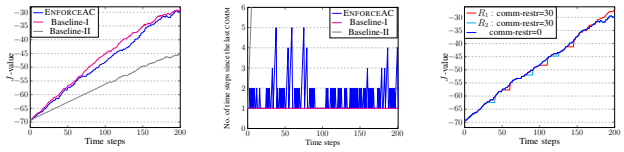
After a COMM, we update ${}^c\mathcal{H}_k^{r,r'}$, $\Delta\mathcal{H}_k^{r',r}$, $\Delta\mathcal{H}_k^{r,r'}$, $\Delta\mathcal{Z}_k^{r',r}$ and $\Delta\mathcal{Z}_k^{r,r'}$ with the transmitted observations. As a result, at least one of the histories H_k^r and $H_k^{r'}$ gets updated.

For instance, if robot r communicated some observation $z_j^r \in \Delta\mathcal{H}_k^{r',r}$ to robot r', then, from the perspective of robot r, it updates ${}^c\mathcal{H}_k^{r,r'} \leftarrow {}^c\mathcal{H}_k^{r,r'} \cup \{z_j^r\}$, $\Delta\mathcal{H}_k^{r,r'} \leftarrow \Delta\mathcal{H}_k^{r,r'} \setminus \{z_j^r\}$. In this case, $H_k^r = \{{}^c\mathcal{H}_k^{r,r'}, \Delta\mathcal{H}_k^{r',r}\}$ remains the same as robot r did not receive any new observation(s). Robot r also updates the joint observation space $\Delta\mathcal{Z}_k^{r',r}$ from (13) to exclude the corresponding observation space $\mathcal{Z}_j^r$ of the actual observation z_j^r. Robot r', upon receiving the observation z_j^r, does a similar update to ${}^c\mathcal{H}_k^{r,r'}$, $\Delta\mathcal{H}_k^{r,r'}$ and $\Delta\mathcal{Z}_k^{r',r}$. Consequently, $H_k^{r'} = \{{}^c\mathcal{H}_k^{r,r'}, \Delta\mathcal{H}_k^{r,r'}\}$ is updated.

Given the updated histories H_k^r and $H_k^{r'}$, we update the beliefs b_k^r and $b_k^{r'}$ according to (10), and execute VERIFYAC to check if MR-AC exists with the updated beliefs. If MR-AC is not satisfied, again COMMs are triggered. This continues until MR-AC is achieved. Thus, ENFORCEAC enforces MR-AC via COMMs even though MR-AC is not satisfied initially.

Time complexity of ENFORCEAC*:* The worst case time complexity of ENFORCEAC is $\mathcal{O}(pn)$, where p is the number of previous time points before which the robots had consistent history and $n \triangleq |\Delta\mathcal{Z}_k|$.

Proof: ENFORCEAC calls VERIFYAC. First we analyze the time complexity of VERIFYAC, which iterates over all possible observations in $\Delta\mathcal{Z}_k^{r,r'}$ and $\Delta\mathcal{Z}_k^{r',r}$ in Steps 2 and 3, respectively. For each such observation VERIFYAC calculates and compares between the corresponding J-values for different candidate joint actions. Whenever VERIFYAC finds an inconsistent observation, a COMM is triggered. Overall, this incurs a runtime of $\mathcal{O}(|\Delta\mathcal{Z}_k^{r',r}| + |\Delta\mathcal{Z}_k^{r,r'}|) = \mathcal{O}(|\Delta\mathcal{Z}_k|) = \mathcal{O}(n)$ of VERIFYAC. Note that calculation and comparison of J-values for each observation takes a constant



(a) *comm-restr* = 0 (b) *comm-restr* = 0 (c) *comm-restr* = 30

Fig. 4. (a) J values and (b) number of time steps since the last communication, with no restriction in COMM considering *MaxEntropy-Init*, 4 motion primitives, and epoch $E = 200$. Planning is done at each time step. (c) Performance of ENFORCEAC for two different realizations (R_1 and R_2) each with 30 time steps at which the robots cannot communicate.

amount of time.

Inconsistent observations in VERIFYAC lead to triggering at most $2p$ number of COMMs by ENFORCEAC. This happens when during every COMM each of the two robots sends its unshared local observation at every single time point in the time range $[k-p+1, k]$. After each such COMM, VERIFYAC is invoked again by ENFORCEAC to check for MR-AC. So, at most $2pn$ number of comparisons of J-values in total and hence complexity of ENFORCEAC is $\mathcal{O}(pn)$. ■

Theorem 3.4: ENFORCEAC converges to MR-AC in a finite amount of time, even if MR-AC does not exist initially.

Proof: During a COMM, the unknown observation sequences can be shared partially, i.e. some of the observations in the time range $[k-p+1:k]$ can be shared to the other robot. In the worst case, to ensure MR-AC, robot r may have to send its entire actual observation in time range $[k-p+1, k]$ to r'. Sharing all the observations takes finite amount of time. Hence, ENFORCEAC achieves MR-AC in finite time. ■

IV. IMPLEMENTATION AND RESULTS

We demonstrate the applicability and performance of our approach by simulating a *search and rescue* application in a disaster ravaged area. We compare our algorithm ENFORCEAC with two baseline algorithms – Baseline-I and Baseline-II. In Baseline-I robots do two-way COMMs at each time point, and that leads to consistent beliefs of the robots at each time point. In Baseline-II robots do not communicate at all, and end up having inconsistent beliefs at every time step. Simulations were carried out in an Intel Core i7-6500U with 2.5 GHz clock. The algorithms are implemented in `Julia`.

A. Search and rescue in a disaster affected region

Consider that a team of two robots $R = \{r, r'\}$ are engaged in finding an *unknown* number of targets (for example, victims) in a disaster affected region. Task of the robot team is to collaboratively find targets with high confidence (reduced uncertainty) which is facilitated by having different observations at different locations. Due to poor bandwidth and other connectivity constraints, the robots have limited scope of communication between them. The simulations are done in a 2-D occupancy grid sub-divided into discrete cells with unique identifiers: $\{s_i\}$ for $i = [1 \dots X]$ where X is the maximum index of the cells. A target is either *present* or *not present* in a given cell s_i and the presence of a target in s_i is denoted as $x_i = 1$, else $x_i = 0$.

We assume the following for ease of implementation and our algorithm is not limited to these assumptions: (a) each robot precisely knows the locations of all the robots in Γ; thus the belief is the probability distribution over the joint state $x \triangleq \{x_i\}$. Denote the known pose of any robot $r \in \Gamma$ at time instant k by ξ^r_k; (b) the initial beliefs b^r_0 and $b^{r'}_0$ of both robots are given and consistent, i.e. $b^r_0 = b^{r'}_0 = b_0$. The cells are initially independent of each other, i.e. $b_0 = p(x \mid p_0) = \prod_i p_0(x_i)$; (c) at any time instant, each robot observes a single cell where it is located. Based on these assumptions, it is not difficult to show that the cells remain independent of each other at any time instant k, i.e. $b_k = p(x \mid \mathcal{H}_k, \xi^r_{0:k}, \xi^{r'}_{0:k}) = \prod_i p(x_i \mid \mathcal{H}_k, \xi^r_{0:k}, \xi^{r'}_{0:k}) \triangleq \prod_i b_k[x_i]$ for any history $\mathcal{H}_k$.

To reduce the uncertainty of target occurrences in the locations of the workspace, we choose an information-theoretic reward function. Specifically, we consider (minus) entropy, which measures uncertainty over presence of targets at different locations. Recalling that according to the assumptions above, the cells are independent for any time instant k, we get $\rho(b_k) \triangleq -H[x] = \sum_i \sum_{j \in \{0,1\}} b_k[x_i = j] \log b_k[x_i = j]$.

For a given epoch $\langle 1, 2, \dots, E \rangle$, we do planning using our algorithm ENFORCEAC at every time step with planning horizon $L = 1$. Each robot has four (`N`, `S`, `E`, `W`) or eight (additionally `NE`, `NW`, `SW`, `SE`) motion primitives each of which moves the robot to a unit distance in the respective direction. Initially $\mathbb{P}(x_i) = 0.5$ for all cells (MaxEntropy-Init), or initially $\mathbb{P}(x_i) = 0.7$ if the cell is occupied and $\mathbb{P}(x_i) = 0.3$ otherwise (Entropy-Init). At any time $k \in [1, E]$, for robot r, beliefs b^r_k, $b^{r'|r}_k$ and $b^{r|r'|r}_k$ are updated as formulated in Section III-C for the joint state x.

Also, at some time instances in E there can be a restriction in communication even if the algorithm decides to communicate. We denote a scenario with m such time instances by *comm-restr*=m, while *comm-restr*=0 corresponds to a scenario with no communication restrictions. This restriction is not available, in advance, to any of the algorithms.

We run ENFORCEAC with the above setting for epoch $E = 200$ without communication restrictions, i.e. *comm-*

TABLE I

NOT-AC (ACTION INCONSISTENCY), COMMS AND TIME FOR $E = 200$.

Input	Algorithm	Not-AC	COMM	Time
comm-restr = 0	Baseline-II	181	0	1.3s
Motion prim. = 4	Baseline-I	0	400	1.3s
MaxEntropy-Init	ENFORCEAC	0	238	12.4s
comm-restr = 0	Baseline-II	185	0	1.3s
Motion prim. = 4	Baseline-I	0	400	1.4s
Entropy-Init	ENFORCEAC	0	268	8.7s
comm-restr = 0	Baseline-II	194	0	3.6s
Motion prim. = 8	Baseline-I	0	400	3.5s
MaxEntropy-Init	ENFORCEAC	0	248	36.4s
comm-restr = 0	Baseline-II	188	0	3.6s
Motion prim. = 8	Baseline-I	0	400	3.6s
Entropy-Init	ENFORCEAC	0	278	31.1s
comm-restr = 20	Baseline-II	194	0	3.3s
Motion prim. = 8	Baseline-I	14	360	4.3s
MaxEntropy-Init	ENFORCEAC	13	224	94.9s
comm-restr = 20	Baseline-II	188	0	3.2s
Motion prim. = 8	Baseline-I	14	360	3.6s
Entropy-Init	ENFORCEAC	10	251	31.2s
comm-restr = 30	Baseline-II	188	0	3.4s
Motion prim. = 8	Baseline-I	22	340	4.0s
MaxEntropy-Init	ENFORCEAC	20	238	46.9s

restr = 0. We show the objective function values in Fig. 4(a), and the number of time steps since the last communication in Fig. 4(b). With EnforceAC, the robots did not communicate for up to five consecutive time steps; nevertheless, as shown in Table I, in all planning sessions there were no inconsistent actions despite the robots having inconsistent beliefs (*Not-AC* is zero for this problem instance). On the other hand, Baseline-I communicated two-way at each planning session, resulting in no action inconsistency, and Baseline-II did not communicate at all resulting in numerous inconsistent actions.

Table I summarizes results for additional scenarios, which vary according to the number of action primitives and the prior belief (MaxEntropy or Entropy-Init). As seen, considering no communication restrictions (*comm-restr* = 0), EnforceAC reduces the number of one-way COMMs by 30-40% compared to Baseline-I, and in all cases ensures consistent decision making between the robots, despite having inconsistent beliefs.

However, compared to Baseline-I, EnforceAC provides sub-optimal action selection: As shown in Fig. 4(a), the J values computed by EnforceAC are slightly worse compared to Baseline-I, though the values are well above the values for Baseline-II. Thus EnforceAC ensures MR-AC with inconsistent beliefs at the expense of quality of the selected action and higher computational complexity. We leave further investigation of these aspects to future research.

Dynamic arrival of communication restrictions: In some scenarios, COMM restrictions may arise dynamically, barring COMMs between the robots, even though COMM is suggested by the algorithm. Table I shows that action inconsistencies (Not-AC) have occurred due to *comm-restr* > 0 for both EnforceAC and Baseline-I. However, the number of Not-ACs is less than the number of restricted time steps (*comm-restr*) in E. For EnforceAC, the reason being: (a) COMM was not required in some of the *comm-restr* time steps as EnforceAC ensures MR-AC even without COMM, and (b) the action selections may be same coincidentally. In fact, the number of Not-ACs is less for EnforceAC compared to Baseline-I (Table I). This is because, in contrast to EnforceAC, Baseline-I reduces Not-ACs by means of only reason (b). Action selections, and therefore, objective values, may differ for two realizations of communication restrictions (Fig. 4(c)), due to different degrees of belief inconsistencies.

V. Conclusion

We have addressed an open problem of ensuring consistent action selection even if the robots have inconsistent beliefs. Our algorithm EnforceAC verifies multi-robot action consistency. On successful verification, action consistency is identified without communication. Otherwise, communication is self-triggered. We provide guarantee that our algorithm provides action consistency eventually. Experimental results show a reduction of $30-40\%$ in the number of communications, though we do not guarantee optimal action selection. Future scope is to further reduce the number of communications and improve the quality of action selection.

References

[1] C. Amato, G. Konidaris, A. Anders, G. Cruz, J. P. How, and L. P. Kaelbling. Policy search for multi-robot coordination under uncertainty. *Intl. J. of Robotics Research*, 35(14):1760–1778, 2016.

[2] C. Amato, G. Konidaris, G. Cruz, C. A. Maynor, J. P. How, and L. P. Kaelbling. Planning for decentralized control of multiple robots under uncertainty. *arXiv preprint arXiv:1402.2871*, 2014.

[3] C. Amato and S. Zilberstein. Achieving goals in decentralized pomdps. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 593–600, 2009.

[4] N. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas. Decentralized active information acquisition: Theory and application to multi-robot SLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 4775–4782, 2015.

[5] Moran Barenboim, Idan Lev-Yehudi, and Vadim Indelman. Data association aware pomdp planning with hypothesis pruning performance guarantees. *IEEE Robotics and Automation Letters (RA-L)*, 2023.

[6] D. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of operations research*, 27(4):819–840, 2002.

[7] Jesus Capitan, Matthijs TJ Spaan, Luis Merino, and Anibal Ollero. Decentralized multi-robot cooperation with auctioned pomdps. *Intl. J. of Robotics Research*, 32(6):650–671, 2013.

[8] Khen Elimelech and Vadim Indelman. Simplified decision making in the belief space using belief sparsification. *The International Journal of Robotics Research*, 41(5):470–496, 2022.

[9] V. Indelman. No correlations involved: Decision making under uncertainty in a conservative sparse information space. *IEEE Robotics and Automation Letters (RA-L)*, 1(1):407–414, 2016.

[10] V. Indelman. Cooperative multi-robot belief space planning for autonomous navigation in unknown environments. *Autonomous Robots*, pages 1–21, 2017.

[11] Andrej Kitanov and Vadim Indelman. Topological belief space planning for active slam with pairwise gaussian potentials and performance guarantees. *Intl. J. of Robotics Research*, 43(1):69–97, 2024.

[12] T. Kundu and I. Saha. Mobile recharger path planning and recharge scheduling in a multi-robot environment. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 3635–3642, 2021.

[13] I. Lev-Yehudi, M. Barenboim, and V. Indelman. Simplifying complex observation models in continuous pomdp planning with probabilistic guarantees and practice. In *AAAI Conf. on Artificial Intelligence*, February 2024.

[14] Negar Mehr, Mingyu Wang, Maulik Bhatt, and Mac Schwager. Maximum-entropy multi-agent dynamic games: Forward and inverse solutions. *IEEE Trans. Robotics*, 2023.

[15] Frans A Oliehoek. Decentralized pomdps. In *Reinforcement Learning*, pages 471–503. Springer, 2012.

[16] Shayegan Omidshafiei, Ali-akbar Agha-mohammadi, Christopher Amato, and Jonathan P How. Decentralized control of partially observable markov decision processes using belief space macro-actions. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2015.

[17] T. Regev and V. Indelman. Decentralized multi-robot belief space planning in unknown environments via efficient re-evaluation of impacted paths. *Autonomous Robots*, 2017. Special Issue on Online Decision Making in Multi-Robot Coordination.

[18] Wilko Schwarting, Alyssa Pierson, Sertac Karaman, and Daniela Rus. Stochastic dynamic games in belief space. *IEEE Trans. Robotics*, 37(6):2157–2172, 2021.

[19] M. Shienman and V. Indelman. Nonmyopic distilled data association belief space planning under budget constraints. In *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, 2022.

[20] Oswin So, Paul Drews, Thomas Balch, Velin Dimitrov, Guy Rosman, and Evangelos A Theodorou. Mpogames: Efficient multimodal partially observable dynamic games. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 3189–3196. IEEE, 2023.

[21] Ori Sztyglic and Vadim Indelman. Speeding up online pomdp planning via simplification. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.

[22] Feng Wu, Shlomo Zilberstein, and Xiaoping Chen. Online planning for multi-agent systems with bounded communication. *Artificial Intelligence*, 175(2):487–511, 2011.

[23] A. Zhitnikov and V. Indelman. Simplified risk aware decision making with belief dependent rewards in partially observable domains. *Artificial Intelligence, Special Issue on "Risk-Aware Autonomous Systems: Theory and Practice"*, 2022.