Towards Addressing Model Uncertainty: Robust Execution-time Coordination for Teamwork (Short Paper)

Jun-young Kwak*, Rong Yang*, Zhengyu Yin*, Matthew E. Taylor† and Milind Tambe*
*University of Southern California, Los Angeles, CA 90089, {junyounk,yangrong,zhengyuy,tambe}@usc.edu
†Lafayette College, Easton, PA 18042, taylorm@lafayette.edu

***Abstract*—Despite their worst-case NEXP-complete planning complexity, DEC-POMDPs remain a popular framework for multiagent teamwork. This paper introduces effective teamwork under model uncertainty (i.e., potentially inaccurate transition and observation functions) as a novel challenge for DEC-POMDPs and presents MODERN, the first execution-centric framework for DEC-POMDPs explicitly motivated by addressing such model uncertainty. MODERN's shift of coordination reasoning from planning-time to execution-time avoids the high cost of computing optimal plans whose promised quality may not be realized in practice. There are three key ideas in MODERN: (i) it maintains an exponentially smaller model of other agents' beliefs and actions than in previous work and then further reduces the computation-time and space expense of this model via bounded pruning; (ii) it reduces execution-time computation by exploiting BDI theories of teamwork, and limits communication to key trigger points; and (iii) it limits its decision-theoretic reasoning about communication to trigger points and uses a systematic markup to encourage extra communication at these points — thus reducing uncertainty among team members at trigger points. We empirically show that MODERN is substantially faster than existing DEC-POMDP execution-centric methods while achieving significantly higher reward.**

***Keywords*-DEC-POMDPs; Model Uncertainty; Teamwork**

I. Introduction

Despite their NEXP-complete policy generation complexity [1], *Distributed Partially Observable Markov Decision Problems* (DEC-POMDPs) have become a popular paradigm for multiagent teamwork [2], [3], [4], [5]. DEC-POMDPs quantitatively express observational and action uncertainty, and yet optimally plan communications and domain actions.

This paper focuses on teamwork under *model uncertainty* (i.e., potentially inaccurate transition and observation functions) in DEC-POMDPs. In many domains, we only have an approximate model of agent observation or transition functions. To address this challenge we rely on execution-centric frameworks [6], [7], [8], which simplify planning in DEC-POMDPs (e.g., by assuming cost-free communication at plan-time), and shift coordination reasoning to execution time. Execution-centric frameworks appear better-suited to address model uncertainty as they (i) lead to provably exponential improvement in worst-case complexity [4], [6]; (ii) avoid paying a high planning cost for a "high-quality" DEC-POMDP policy that cannot be realized in practice; and (iii) allow for coordination reasoning at execution-time to mitigate model uncertainty.

Unfortunately, past work in execution-centric approaches such as ACE-PJB-COMM (APC) [6] and MAOP-COMM (MAOP) [7] assumes a correct world model, and the presence of model uncertainty exposes three key weaknesses in that work. First, they maintain the entire team's belief states for execution-time reasoning, a costly undertaking that is not well-justified given model uncertainty. Second, they reason at execution-time about the right action and communication before each decision step, leading to inefficient computation. Third, their detailed and expensive reasoning about communication and action is based on the assumption of an accurate model — again, given model uncertainty, such precise computation is wasteful due to its inaccuracy.

This paper provides two sets of contributions. The first is a new execution-centric framework for DEC-POMDPs called MODERN (MOdel uncertainty in Dec-pomdp Execution-time ReasoNing). MODERN's key insight is that given model uncertainty, it is wasteful to maintain a very detailed model of other agents or the team's belief states (it will be inaccurate anyway) and reason in depth with such a detailed model — the inferences will also be inaccurate. Instead, model uncertainty drives MODERN to simplify modeling and reasoning of other agents and boost communication at key junctures instead.

MODERN is the first execution-centric framework for DEC-POMDPs explicitly motivated by model uncertainty. It is based on three key ideas. First, MODERN reasons with an exponentially smaller model of other agents' beliefs and actions than the entire set of joint beliefs as done in previous work [6], [7], [8]; then it further reduces the computation time and space expense of this model via bounded pruning. Second, MODERN reduces execution-time computation by: (i) engaging in decision-theoretic reasoning about communication only at *Trigger Points* — instead of every agent reasoning about communication at every step, only agents encountering trigger points perform such reasoning; and (ii) utilizing a pre-planned policy for actions that do not involve interactions, avoiding on-line planning at every step. Our approach has significant advantages in domains with interaction-sparseness. Third, MODERN increases communication at trigger points by doing a markup of the expected

utility gain under the assumption that communication has significant value in reducing uncertainty. We justify our design decisions in MODERN through a systematic empirical evaluation. As our evaluation shows, MODERN outperforms competing algorithms in terms of its run-time performance while finding higher quality solutions.

This paper's second set of contributions is in opening up model uncertainty as a new research direction for DEC-POMDPs and emphasizing the similarity of this problem to the *Belief-Desire-Intention* (BDI) model for teamwork [9], [10]. In particular, BDI teamwork models also assume inaccurate mapping between real-world problems and domain models. As a result, they emphasize robustness via execution-time reasoning about coordination [10]. Given some of the successes of prior BDI research in teamwork, we leverage insights from BDI in designing MODERN.

II. Design Decisions

During planning, MODERN[1] has a standard single-agent POMDP planner [11] plan a policy for the team of agents by assuming zero-cost communication. The resulting policy, provided to each agent, would be optimal if agents fully communicated at each time step (zero cost communication); but given non-zero-cost communication, agents must selectively communicate. Thus, at execution-time, agents model other agents' beliefs and actions so as to reason about when to communicate with teammates, reason about what action to take if not communicating, etc.

MODERN's design is driven by the model uncertainty, and thus MODERN simplifies modeling and reasoning about other agents by maintaining a bounded approximate model (compared to previous work [6], [7] which maintains a very detailed model). Instead, MODERN boosts communication at trigger points. As shown in our experimental results, it is precisely due to this aggressive reliance on communication rather than detailed reasoning that MODERN outperforms its competitors who are much more reliant on their models of other agents. We describe these ideas in the following.

Modeling Other Agents: In contrast with the complete tree of joint beliefs in [6], [7], MODERN maintains an approximate and exponentially smaller set of beliefs to model other agents via (i) *Individual estimate of joint Beliefs (IB)* and (ii) *Bounded Pruning*.

IB is a concept used in MODERN to decide whether or not communication would be beneficial and to choose a joint action when not communicating. IB can be conceptualized as a subset of team beliefs that depends on an agent's local history. IB^t describe the set of nodes of the possible belief trees of depth t. Each node θ in IB^t has a tuple consisting of $\langle \mathbf{b}(\theta), \mathbf{h}(\theta), \mathbf{a}(\theta), p(\theta) \rangle$, where $\mathbf{b}(\theta)$ is a probability distribution over the set of joint states given that $\mathbf{h}(\theta)$, $\mathbf{h}(\theta)$ is the joint observation history, $\mathbf{a}(\theta)$ is the joint action obtained from a given policy tree, and $p(\theta)$ is the likelihood of observing $\mathbf{h}(\theta)$.

Although IB is exponentially smaller than previous work in [6], [7], the number of possible beliefs in IB grows rapidly, particularly when agents choose not to communicate for long time periods. Hence, we propose a new pruning algorithm that provides further savings. In particular, it keeps a fixed number of *most likely* beliefs per time step in IB. Our pruning method first expands beliefs using the Bayes update rule and then selects the most likely belief at each time step until the selected number of beliefs reaches a pre-defined upper-limit. This reduced belief set is used to detect trigger points and reason about communication in MODERN. Note that MODERN uses a *sync* action in communication (discussed below) that is useful to ensure that all agents create an identical belief. This provides a way to ascertain the team's joint status and avoid miscoordination.

Using IB to Detect Trigger Points: The policy provided to each agent from MODERN's planning maps the agent team's joint observation to joint actions. Unlike [6] or [7], MODERN does not require agents to reason from scratch about what action or communication to execute at every time step. Instead, agents follow the provided policy, mapping their own observation in the policy to their own action, except at "trigger points." Trigger points include any situation involving ambiguity in mapping an agent's observation to its action in the joint policy. The key idea is that in sparse interaction domains, agents will not have to reason about coordination at every time step and only infrequently encounter trigger points, thus significantly reducing the burden of execution-time reasoning.

Using trigger points to reason about communication is similar to the use of joint commitments in BDI teamwork to reason about communication [10]. Indeed, ask and tell in MODERN share some similarity to the initiation and termination (respectively) of joint commitments to trigger communication in BDI teamwork [9], [10].

How to Reason — The Markup Function: MODERN's reasoning about communication is governed by the following formula: $f(\kappa, t) \cdot (U_{\text{C}}(i) - U_{\text{NC}}(i)) > \sigma$, where κ is a markup rate, t is a time step, $U_{\text{C}}(i)$ is the expected utility of agent i if agents were to communicate, $U_{\text{NC}}(i)$ is the expected utility of agent i when it does not communicate, and σ is a given communication cost. The two novelties in MODERN's reasoning are how it computes $U_{\text{C}}(i)$ and $U_{\text{NC}}(i)$, and how it uses the markup function $f(\kappa, t)$. Both of these are motivated by model uncertainty.

In MODERN, agents reason if communication would be beneficial. If they communicate, all agents synchronize local observation histories. Thus, all agents reach a specific belief node, θ, and can choose a joint action for the team. Otherwise, if no agent chooses to communicate, each agent chooses the best locally optimal action based on estimated most likely actions of other agents. Computation

[1]More details about MODERN and experiments can be found in [12].

of $U_{\mathrm{C}}(i) - U_{\mathrm{NC}}(i)$ is performed as following:

$$
\begin{aligned}
U_{\mathrm{C}}(i) = & \sum_{\theta \in \mathrm{IB}_i} p(\theta) \cdot V\left(\mathbf{b}(\theta), \mathbf{a}(\theta)\right), \\
U_{\mathrm{NC}}(i) = & \max_{a_i \in A_i} U_{\mathrm{IB}_i}(\langle a_i, a^*_{-i} \rangle), \\
a^*_{-i} = & \; a_{-i}(\theta^*) \;\; \text{s.t.} \;\; \theta^* \in \Theta, \\
U_{\mathrm{IB}_i}(\mathbf{a}) = & \sum_{\theta \in \mathrm{IB}_i} p(\theta) \cdot V(\mathbf{b}(\theta), \mathbf{a}).
\end{aligned}
$$

$V(\mathbf{b}, \mathbf{a})$ is the expected utility when an action $\mathbf{a}$ is taken at belief state $\mathbf{b}$. a^*_{-i} is agent i's estimate of the most likely action of all other agents. This is greedily selected using the most likely observation sequence for all other agents, Θ, at every time step. $U_{\mathrm{C}}(i)$ is calculated by considering two-way synchronization, which emphasizes the benefits from communication. $U_{\mathrm{NC}}(i)$ is computed based on the individual evaluation of heuristically estimated actions of other agents.

The markup function, $f(\kappa, t)$, helps agents to reduce uncertainty among team members by marking up the expected utility gain from communication rather than perform precise local computation over erroneous models — the markup in essence selectively boosts communication. In this work, we use an exponential markup rate, $f(\kappa, t) = \kappa^t$. Because uncertainty among team members increases as time passes, the markup rate should increase according to the time step.

III. Empirical Validation

We evaluate the performance of MODERN and show some preliminary results compared to two previous techniques: APC [6] and MAOP [7]. The *planning* time for all algorithms is identical and thus we only measure the average execution-reasoning time per agent. Noise in transition matrix and observation matrix follow a Dirichlet distribution (which is not known by the planner or the agents). The level of model error is represented by a parameter α ($\sum_i^L \beta_i$) in Dirichlet distribution: error *increases* as α *decreases*. We evaluate MODERN under four different amounts of error by varying α from 10 to 10000. The experiments were run on Intel Core2 Quadcore 2.4GHz CPU with 3GB main memory. All techniques were evaluated for 600 independent trials throughout this section. We report the average rewards.

Comparison — Solution Quality: We compared the average rewards achieved by all algorithms for three different communication costs in two small grid domains and the dec-tiger domain[2]. The communication costs are selected proportional to the expected value of the policies: $5\%, 20\%$, and 50%. The time horizon was set to 3 in this set of experiments.

In Table I, σ in column 1 displays the different communication cost and α in column 2 represents the level of model error. Columns 3–6 display the average reward achieved by each algorithm in the 1×5 grid domain. Columns 7–10 show the results in the 2×3 grid domain. Columns 11–14 are for the multi-agent tiger domain. For the markup function in MODERN (MD in Table I), κ_1=1.0 and κ_2=1.25 were used. We performed experiments with a belief bound of 10 nodes per time-step for our algorithm.

Table I shows that MODERN (columns 3–4, and 7–8) significantly outperformed APC (columns 5 and 9) and MAOP (columns 6 and 10) in the grid domains that have sparse interactions. MODERN received statistically significant improvements (via t-tests, $p < 0.01$), relative to other algorithms. In the highly-coupled tiger domain, APC (column 13) had slightly higher reward than MODERN (columns 11–12) when communication cost (σ) was low (5%, rows 3–6) or medium (20%, rows 7–10), but the difference was only about 10% in reward. However, when σ was high (50%, rows 11–14), MODERN outperformed APC. In particular, even at this high σ, MODERN selectively utilized communication to successfully perform a joint task, and thus it achieved higher reward. MAOP (column 14) showed the worst results regardless of α and σ.

In these small domains, the average reward in MODERN was similar regardless of the markup rate (columns 3–4, 7–8, and 11–12). *Indeed, without carefully tuning κ (e.g., κ=1.0), MODERN's rewards were still statistically significantly higher than others.*

Comparison — Runtime: Here, we compare the average (execution) runtime per agent of the algorithms. MODERN used 10 belief nodes for the bounded pruning (for small domains, this limit was never reached). Communication cost was 5% of the expected utility. The maximum runtime per trial was set to 1,800 seconds. In the tiger domain, all algorithms showed similar results. In small 1×5 and 2×3 grid domains, MODERN and APC took similar amounts of time. The runtime of MAOP was 1.39–1.89 times that of MODERN's runtime in both domains, where this difference was statistically significant (via t-tests, $p < 0.01$). In a scaled-up 2×3 grid domain with longer time horizon (T=5), MAOP was not able to finish running within the time limit. APC uses a particle filtering technique to improve speed, but even with only one particle, APC exceeded the time limit to finish a trial, whereas MODERN took less than 125 seconds.

We then ran experiments in the larger grid domain with increased time horizons. We tested the algorithm under two different communication costs: 5% and 50%. MAOP and APC (with 1 particle) could not solve the problem within the given time limit for even the shortest time horizon — while MODERN took significantly less time than other algorithms. In particular, when T=8, MODERN takes less than 150 seconds with σ=5% and about 1500 seconds with σ=50%, which are still lower than the time limit. As the time horizon increased, MODERN obtained higher rewards (9.8–15.7), since there was more time for agents to recover from any failed actions. With σ=50%, MODERN took more time than

[2]The domain details are described in [12].

		1×5 Grid				2×3 Grid				Dec-Tiger			
σ	α	MD(κ_1)	MD(κ_2)	APC	MAOP	MD(κ_1)	MD(κ_2)	APC	MAOP	MD(κ_1)	MD(κ_2)	APC	MAOP
5%	10	5.36	5.38	-1.20	1.52	5.28	5.30	-2.25	-0.36	11.45	11.36	12.56	-3.09
	50	5.24	5.11	-1.20	1.49	5.28	5.33	-2.04	-0.68	10.95	10.94	11.92	-3.44
	100	5.16	5.20	-1.20	1.47	5.02	5.03	-1.85	-0.63	11.18	11.23	12.33	-3.37
	10000	4.46	4.38	-1.20	1.13	4.62	4.61	-1.80	-0.78	10.92	10.96	11.92	-3.51
20%	10	4.70	4.65	-1.20	0.38	4.62	4.68	-1.20	-1.47	8.35	8.41	10.70	-5.69
	50	4.58	4.71	-1.20	0.28	4.62	4.66	-1.20	-1.72	7.59	7.56	10.64	-6.13
	100	4.50	4.46	-1.20	0.28	4.36	4.40	-1.20	-1.68	8.09	8.11	10.55	-6.04
	10000	3.80	3.71	-1.20	-0.12	3.96	3.91	-1.20	-1.86	7.77	7.73	10.31	-6.21
50%	10	3.38	3.39	-1.20	-1.90	3.30	3.29	-1.20	-3.69	0.24	0.17	-6.0	-11.78
	50	3.26	3.25	-1.20	-2.15	3.30	3.34	-1.20	-3.80	-0.81	-0.80	-6.0	-12.40
	100	3.18	3.16	-1.20	-2.12	3.04	3.06	-1.20	-3.79	-1.42	-1.39	-6.0	-12.27
	10000	2.48	2.52	-1.20	-2.61	2.64	2.62	-1.20	-4.01	-1.18	-1.26	-6.0	-12.51

Table I
COMPARISON OF MODERN (MD) WITH APC AND MAOP: AVERAGE PERFORMANCE IN SMALL DOMAINS

with σ=5%, although still scaling linearly with time horizon.

IV. CONCLUSIONS AND RELATED WORK

This paper aims to open a new area of research for DEC-POMDPs: in many real-world domains, we will not have a perfect model of the world, and hence DEC-POMDPs must address model uncertainty. To combat model uncertainty, we presented a new framework called MODERN that simplifies DEC-POMDP planning (significantly reducing its complexity) and instead relies on agents' execution-time reasoning. There are three major new ideas in MODERN's execution time reasoning: (i) it avoids excessive reliance on a complete model by maintaining an approximate model of other agents by bounded pruning, resulting in exponentially smaller beliefs compared to previous work, (ii) it reduces computational burden by exploiting BDI teamwork and sparse interactions between agents to limit reasoning about communication, and (iii) it marks up the expected gain in utility to reduce uncertainty among team members by boosting communication. We justified our design decisions in MODERN via an empirical evaluation that considers several factors including communication costs and markup rates in different domains. We showed that not only is MODERN faster than existing algorithms, it also achieves significantly superior solution quality.

We have discussed related work throughout the paper and specifically in terms of comparing the performance of MODERN to other execution-centric approaches, specifically [6] and [7], illustrating MODERN's superior performance. Indeed, MODERN maintains exponentially smaller models of other agents than [6], [7], performs significantly less computation because it uses trigger points that are absent in [6], [7], and uses markup functions to further boost its performance that is absent in other previous work. Other execution-centric approaches include Xuan and Lesser's work [8]; however, that focuses on DEC-MDPs rather than DEC-POMDPs and also does not handle model uncertainty. BDI teamwork approaches [9], [10] did focus on execution-centric reasoning, but they lacked the explicit representation of costs and uncertainties, the main strength of the DEC-POMDP approach. Other DEC-POMDP approaches have focused on reasoning about optimal communication at planning time [2], [4], but again there is no explicit discussion of model uncertainty in that work; and as a result it remains focused on planning-centric (which has a NEXP-complete complexity) rather than an execution-centric approach. Indeed, we expect handling model uncertainty via execution-centric approaches will be more and more critical that as we transition DEC-POMDPs to the real-world.

ACKNOWLEDGMENT

We thank Perceptronics Solutions, Inc. for their support of this research, and Maayan Roth for providing us with the source code for ACE-PJB-COMM.

REFERENCES

[1] D. S. Bernstein, S. Zilberstein, and N. Immerman, "The complexity of decentralized control of markov decision processes," in *UAI*, 2000.
[2] C. V. Goldman and S. Zilberstein, "Optimizing information exchange in cooperative multi-agent systems," in *AAMAS*, 2003.
[3] R. Nair, M. Yokoo, M. Roth, and M. Tambe, "Communication for improving policy computation in distributed POMDPs," in *AAMAS*, 2004.
[4] D. V. Pynadath and M. Tambe, "The communicative multiagent team decision problem: Analyzing teamwork theories and models," *JAIR*, vol. 16, pp. 389–423, 2002.
[5] S. Seuken and S. Zilberstein, "Formal models and algorithms for decentralized decision making under uncertainty," in *AAMAS*, 2008.
[6] M. Roth, R. Simmons, and M. Veloso, "Reasoning about joint beliefs for execution-time communication decisions," in *AAMAS*, 2005.
[7] F. Wu, S. Zilberstein, and X. Chen, "Multi-agent online planning with communication," in *ICAPS*, 2009.
[8] P. Xuan and V. Lesser, "Multi-agent policies: from centralized ones to decentralized ones," in *AAMAS*, 2002.
[9] H. J. Levesque, P. R. Cohen, and J. H. T. Nunes, "On acting together," in *AAAI*, 1990.
[10] M. Tambe, "Towards flexible teamwork," *JAIR*, vol. 7, pp. 83–124, 1997.
[11] L. Kaelbling, M. Littman, and A. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, pp. 99–134, 1998.
[12] J. Kwak, R. Yang, Z. Yin, M. E. Taylor, and M. Tambe, "Robust execution-time coordination in DEC-POMDPs under model uncertainty," in *the MSDM workshop at AAMAS*, 2011.