Long-Run Multi-Robot Planning Under Uncertain Task Durations

Doctoral Consortium

Carlos Azevedo
Institute For Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Portugal
cguerraazevedo@tecnico.ulisboa.pt

ABSTRACT

This paper presents part of the work developed so far within the scope of my PhD and suggests possible future research directions. My thesis tackles the problem of multi-robot coordination under uncertainty over the long-term. We present a preliminary approach that tackles multi-robot monitoring problems under uncertain task durations. We propose a methodology that takes advantage of a modeling formalism for robot teams: generalized stochastic Petri nets with rewards (GSPNR). A GSPNR allows for unified modeling of action selection and uncertainty on duration of action execution. At the same time, it allows for goal specification through the use of transition rewards and rewards per time unit. The proposed approach exploits the well-defined semantics provided by Markov reward automata in order to synthesize policies.

KEYWORDS

Multi-robot systems, Planning under uncertainty, Long-run average optimization

ACM Reference Format:
Carlos Azevedo. 2020. Long-Run Multi-Robot Planning Under Uncertain Task Durations. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Long-term multi-robot coordination is important in many real-world robotics applications, such as disaster prevention, mining, traffic-monitoring, and ware-house automation. We focus on the problem of synthesizing policies to coordinate teams of robots in these applications. In such scenarios, it is crucial to deploy robust and efficient behaviors. There are many factors that influence the team's performance, and consequently, approaches that account for as many factors as possible usually perform better. An important factor is uncertainty that emerges from the execution in non-controlled environments. Common sources of uncertainty are present in the outcome of the actions, in the duration of each task, or in the battery autonomy of each robot. Moreover, since these applications are inherently long term missions, approaches that take into account infinite horizons usually perform more efficiently. Many of the currently deployed solutions for multi-robot planning [7] make use of hand-crafted behaviors, not reasoning explicitly about uncertainty. These ad-hoc approaches can be dependable, but every new scenario requires a significant engineering effort. Furthermore, as the scale of the deployments grows, rule based approaches tend to be harder to define and become more suboptimal.

In recent works, including one awaiting publication at AAMAS, we proposed a methodology to solve multi-robot persistent monitoring problems. We proposed a model-based methodology based on an *generalized stochastic Petri net with rewards* (GSPNR), an extension of generalized stochastic Petri nets (GSPNs) [1] to include rewards. We take advantage of the model checking algorithm proposed by [2], in order to synthesize policies that optimize the long-run average reward criterion.

2 GSPNRS FOR MULTI-ROBOT TEAMS

Example 2.1. Consider the problem where an homogeneous team of robots can move between a set of locations. The goal is to persistently monitor some of these locations. In order to achieve that, each robot can perform some tasks such as navigate or gather data. However, each robot has a limited battery autonomy and, can only execute tasks for a certain amount of time. The time that it takes to execute each task as well as the time that it takes to discharge or recharge is not deterministic since it is influenced by many uncontrollable factors.

We extend the GSPN-based modeling approach presented in [5] to capture in one single model multi-robot persistent monitoring problems including the objective to be optimized. The key point to achieve this is representing each robot as a token, which is possible under the assumption that our team of robots is homogeneous. A GSPNR for multi-robot teams is defined as $GR = \langle P, T, W^+, W^-, F, m_0, r_P, r_T \rangle$. P is a finite set of places that represent tasks that each robot can execute, or external processes that each robot must wait to be accomplished, such as battery charging. $T = T_I \cup T_E$ is a finite set of transitions partitioned into two subsets, where T_I contains all immediate transitions and T_E contains all exponential transitions. The exponential transitions, model the time uncertainty associated with these uncontrollable events, such as the time that a robot takes to move from one location to another. The immediate transitions represent the controllable actions that the team of robots has available. $W^- : P \times T \to \mathbb{N}$ and $W^+ : T \times P \to \mathbb{N}$ are input and output arc weight functions, respectively. Input arcs assign to tasks uncontrollable events or the choice of deciding among multiple controllable actions. Output arcs assign to each event the outcome states to where the system is lead after selecting a particular action or after the conclusion of an uncontrollable event. The goal is specified through the use of

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

rewards. There are two types of rewards that can be assigned to the model: place rewards and transition rewards. The place reward, $r_P : P \rightarrow \mathbb{R}_{\geq 0}$, is awarded, per time unit, while at least one token is present in the assigned place. Therefore, the reward obtained by the multi-robot system is proportional to the amount of time in the assigned places, encouraging the team of robots to remain or avoid those places, depending on whether it is posed as a maximization or minimization problem. The transition rewards are given by $r_T : T_I \rightarrow \mathbb{R}_{\geq 0}$, with reward $r_T(tn)$ being awarded every time $t_n \in T_I$ is fired, i.e. whenever a robot takes the action corresponding to t_i. These rewards are useful when specifying a goal where some particular actions should be promoted or discouraged.

Figure 1 shows the GSPNR model that captures an instance of the monitoring problem stated in Example 2.1. This problem consists in a team of 3 robots that must monitor 2 locations. The robots, denoted as tokens, are able to monitor, navigate and recharge. These tasks are captured by the places labeled as *Monitoring*, *Navigating* and *Charging*, respectively. The uncertainty associated with the duration of each of these tasks is captured by the exponential transitions labeled as *Finished*, *Arrived* and *Charged*, respectively. The exponential transitions labeled as *Discharged* capture the stochasticity of the multi-robot team battery autonomy. Moreover, the places labeled as *Ready* represent states where the robots can decide between monitoring again the same location or moving to another location, by choosing between the immediate transitions *Repeat* and *GoFromTo*, respectively.

The interpretation of the marking process of the GSPNR model as a Markov reward automaton (MRA) [4] allows the use of the method proposed by [2] to compute the optimal long-run average reward and extract the policy that maximizes this criterion. Figure 2 exemplifies how to represent a GSPNR as an MRA. Each reachable marking is interpreted as a state of the MRA, and the initial marking corresponds to the initial state. Immediate and exponential transitions have the same meaning in GSPNR and MRA. The set of actions is formed by the set of immediate transitions, plus an action to characterize all exponential transitions, denoted as $\perp$. The state rewards correspond to the sum of all place rewards with at least one token while the transition rewards have a one-to-one correspondence.

However, this method only allows for a straightforward extraction of the corresponding policy when the produced MRA is *unichain* [8]. To guarantee that we only produce unichain MRAs, we restrict our GSPNR models to be *reversible* [6], i.e. such that every marking is reachable from all the other markings.

3 FUTURE WORK

We split the future work into three research directions: more expressive models, scalable methods and approaches that allow capturing the trade-off between multiple objectives.

Currently the presented model only captures the multi-robot team behavior and its interaction with the environment. We intend to extend the model to also capture resources such as the time elapsed since the last time a location was visited, or the battery level of each robot. Since the MRA also accounts for uncertainty in the action outcomes we plan to take advantage of this and incorporate in our GSPNR model the possibility to obtain policies that

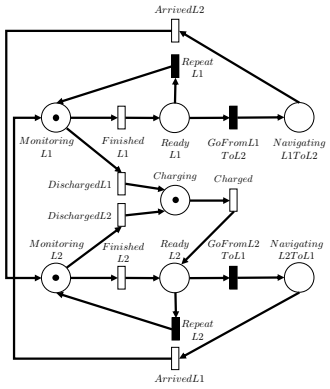

Figure 1: GSPNR model of a simple monitoring example with 3 robots and 2 locations.

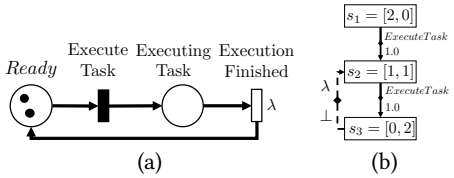

Figure 2: (a) Example GSPNR, where circles, solid rectangles and white rectangles represent places, immediate transitions and exponential transitions, respectively. (b) The MRA obtained from the GSPNR. The dashed lines correspond to exponential transitions and the solid lines to immediate transitions.

take that into consideration. Furthermore, restricting the approach only to exponential distributions is suitable for a large number of scenarios, but still a significant restriction. Consequently, we would like to explore other approaches in order to model uncertainty with other distributions, by for example incorporating phase-type distributions.

Despite synthesizing policies that perform very well on the long-run, this method lacks scalability due to the slow convergence of the relative value iteration algorithm. Therefore, we are currently investigating methods that exploit sub-optimal policies in order to scale for larger teams. One direction that we are currently exploring is the adaptation of heuristic search algorithms for MDPs, such as labeled real time dynamic programming. Moreover, we plan to extend this method to arbitrary GSPNR models, since currently the method is limited to GSPNRs that produce unichain MRAs.

Additionally, in many scenarios, it is very hard to characterize the solution of a problem in terms of a unique utility function. These cases require the definition of trade-offs between multiple objectives. For this reason we will also investigate the extension of the current method to handle multi-objective problems, possibly extending ideas from [3] to the context of continuous-time models.

ACKNOWLEDGMENTS

This work was supported by the Portuguese Fundação para a Ciência e Tecnologia (FCT) under grant SFRH/BD/135014/2017.

REFERENCES

[1] Gianfranco Balbo. 2007. Introduction to generalized stochastic Petri nets. In *International School on Formal Methods for the Design of Computer, Communication and Software Systems*. Springer, 83–131.

[2] Yuliya Butkova, Ralf Wimmer, and Holger Hermanns. 2017. Long-run rewards for Markov automata. In *Proceedings of the 19th International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer, 188–203.

[3] Krishnendu Chatterjee. 2007. Markov decision processes with multiple long-run average objectives. In *International Conference on Foundations of Software Technology and Theoretical Computer Science*. Springer, 473–484.

[4] Christian Eisentraut, Holger Hermanns, Joost-Pieter Katoen, and Lijun Zhang. 2013. A semantics for every GSPN. In *Proceedings of the 34th International Conference on Applications and Theory of Petri Nets and Concurrency (Petri Nets)*. Springer, 90–109.

[5] Masoumeh Mansouri, Bruno Lacerda, Nick Hawes, and Federico Pecora. 2019. Multi-robot planning under uncertain travel times and safety constraints. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*. 478–484.

[6] Tadao Murata. 1989. Petri nets: Properties, analysis and applications. *Proc. IEEE* 77, 4 (1989), 541–580.

[7] Federico Pecora, Henrik Andreasson, Masoumeh Mansouri, and Vilian Petkov. 2018. A Loosely-Coupled Approach for Multi-Robot Coordination, Motion Planning and Control. In *In Proceedings of the 28th International Conference on Automated Planning and Scheduling (ICAPS)*.

[8] Martin L Puterman. 2014. *Markov Decision Processes.: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.