Long-Run Multi-Robot Planning With Uncertain Task Durations

Extended Abstract

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

Bruno Lacerda
Oxford Robotics Institute, University of Oxford, UK
bruno@robots.ox.ac.uk

Nick Hawes
Oxford Robotics Institute, University of Oxford, UK
nickh@robots.ox.ac.uk

Pedro Lima
Institute For Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Portugal
pedro.lima@tecnico.ulisboa.pt

ABSTRACT

This paper presents a multi-robot long-term planning approach under uncertainty on the duration of tasks. The proposed methodology takes advantage of generalized stochastic Petri nets to model multi-robot teams. It allows for unified modeling of action selection and uncertainty on duration of action execution. Goals are specified through the use of transition rewards and rewards per time unit. Our approach exploits the semantics provided by Markov reward automata in order to synthesize policies that optimize the long-run average reward. We provide an empirical evaluation on a simulated multi-robot monitoring problem, showing that the synthesized policy outperforms a carefully hand-crafted policy.

KEYWORDS

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

ACM Reference Format:
Carlos Azevedo, Bruno Lacerda, Nick Hawes, and Pedro Lima. 2020. Long-Run Multi-Robot Planning With 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

In recent years, there has been a growing interest on the use of multi-robot systems for disaster prevention applications, such as flooding and forest fires detection and traffic monitoring [5]. These applications require the team of robots to perform efficiently for long periods of time. Furthermore, they necessitate the use of approaches that handle the uncertainty inherent to teams of robots executing actions in real 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. Currently deployed solutions for multi-robot planning [8] 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, one cannot provide any formal guarantee on the deployed controllers. As the scale of the deployments grows, rule based approaches tend to be harder to define and become more suboptimal.

We consider the problem of synthesizing optimal policies to coordinate teams of robots over long periods of time in these monitoring applications. We propose a model-based methodology based on *generalized stochastic Petri net with rewards* (GSPNR), an extension of GSPNs [1] that includes rewards. We take advantage of the model checking algorithm proposed by [2], in order to synthesize policies that optimize a long-run average reward property.

2 GSPNRS FOR MULTI-ROBOT TEAMS

A recurrent problem when solving multi-robot problems with model-based approaches is that as the number of robots increases the state-space grows exponentially. To mitigate this state-space explosion while maintaining an intuitive representation of the multi-robot system, we extend the GSPN-based modeling approach presented in [6]. 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 rewards. There are two types of rewards that can be assigned to the model: place rewards and transition rewards. The place reward, $r_P : P \to \mathbb{R}_{\geq 0}$, 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

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. © 2020 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

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.

The interpretation of the marking process of the GSPNR model as a Markov reward automaton (MRA) [3] allows the use of the method proposed by [2] to compute the optimal long-run average reward. However, this method only allows for a straightforward extraction of the corresponding policy when the produced MRA is *unichain* [9]. To guarantee that we only produce unichain MRAs, we restrict our GSPNR models to be *reversible* [7], i.e. such that every marking is reachable from all the other markings.

3 EXPERIMENTS

We applied our method to a 4 robot setup implemented on the Stage simulator [4] and ROS. Figure 1 (a) shows the simulated environment where the map is discretized into 6 locations plus a charging station. The goal is to monitor locations *L4*, *L5* and *L6* with an increasing level of priority. To do that each robot can choose if it monitors the current location or if it moves to a different location. The robots can only move between locations that are connected through edges, depicted as arrows in Figure 1 (a), but are not restricted to a predefined path plan. The multi-robot team is initialized with random battery levels and all the robots start in the charging station. The simulation was kept as realistic as possible, by adding noise to the measurements and by using standard ROS packages for localization and navigation. Therefore, there is uncertainty associated with the charge/discharge time of the batteries, the time to traverse two locations and the time to gather enough monitoring data. This uncertainty, the choices that each robot can take and the goal is captured in one single GSPNR model. Figure 1 (b) shows the building blocks used to model this problem.

To assess the optimal policy obtained through our approach, we compare it against two baselines: a hand coded policy and a random policy. The hand coded policy sends the robots to the locations *L6*, *L5* and *L4* with decreasing order of priority. Since one robot monitoring a location is enough to get the maximum reward for that location, the hand coded policy only sends a robot to each location if it is unoccupied or no other robot is navigating towards there. In the case, where all three priority locations are occupied the fourth robot is sent to location *L2*, since is the one closest to location *L6*. The random policy selects a location for each robot according to a uniform distribution.

Figure 2 shows the results obtained after 50 runs of 1 hour each. These demonstrate that the random baseline is considerably outperformed by the hand coded policy, showing that it is sufficiently hard to outperform. Furthermore, given that we optimize for long-run average, the policy obtained using our approach outperforms all the others policies on the long term.

4 CONCLUSIONS

In this paper, we presented an overview of a multi-robot planning approach for long-term monitoring scenarios, that provides robust policies where uncertainty on the duration of tasks are taken into consideration. We show that these policies can outperform a carefully hand-crafted policy. In the future, we intend to extend this method to arbitrary GSPNR models. Additionally, we will compare it against other state-of-the-art methods, such as ones that optimize the discounted expected reward.

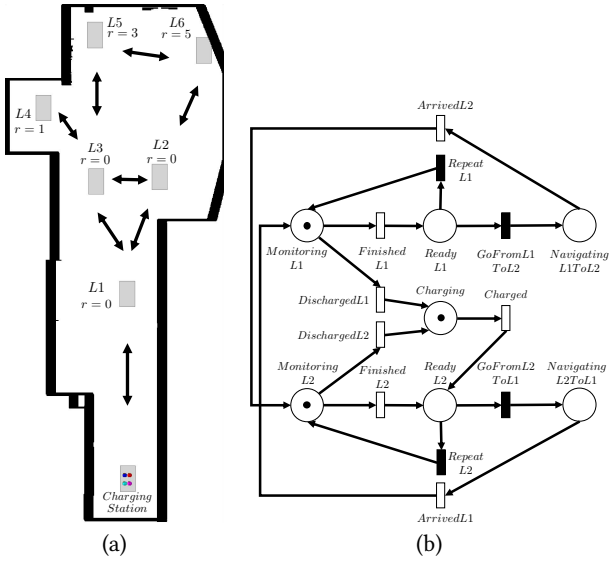

Figure 1: (a) Stage map. The setup consists in 4 robots, 6 locations and 1 charging station. Above each location is shown the reward assigned to it. (b) Building blocks of the GSPNR model that captures the monitoring problem.

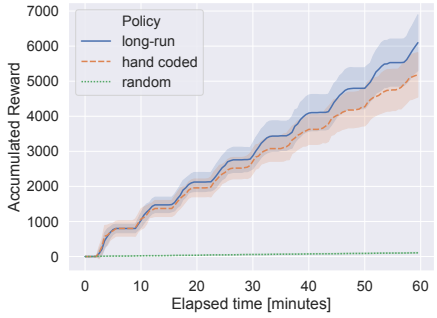

Figure 2: Accumulated reward while executing each policy in the Stage simulation. The lines represent the mean and the shadows the standard deviation.

ACKNOWLEDGMENTS

This work was supported by the Portuguese Fundação para a Ciência e Tecnologia (FCT) under grant SFRH/BD/135014/2017, UK Research and Innovation and EPSRC through the Robotics and Artificial Intelligence for Nuclear (RAIN) research hub [EP/R026084/1], and the European Union Horizon 2020 research and innovation programme under grant agreement No 821988 (ADE).

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] 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.

[4] Brian Gerkey, Richard T Vaughan, and Andrew Howard. 2003. The player/stage project: Tools for multi-robot and distributed sensor systems. In *Proceedings of the 11th International Conference on Advanced Robotics (ICAR)*. 317–323.

[5] Erez Hartuv, Noa Agmon, and Sarit Kraus. 2018. Scheduling Spare Drones for Persistent Task Performance under Energy Constraints. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, 532–540.

[6] 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.

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

[8] 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)*.

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