Tree-Based Solution Methods for Multiagent POMDPs with Delayed Communication

Matthijs T.J. Spaan [a] Frans A. Oliehoek [b]

[a] *Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands*
[b] *Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*

Abstract

Multiagent Partially Observable Markov Decision Processes (MPOMDPs) provide a powerful framework for optimal decision making under the assumption of instantaneous communication. We focus on a delayed communication setting (MPOMDP-DC), in which broadcasted information is delayed by at most one time step. In this paper, we show that computation of the MPOMDP-DC backup can be structured as a tree and we introduce two novel *tree-based pruning* techniques that exploit this structure in an effective way.

The full version of this paper has appeared in *Proceedings of the Twenty-Sixth Conference on Artificial Intelligence (AAAI-12)*.

1 Introduction

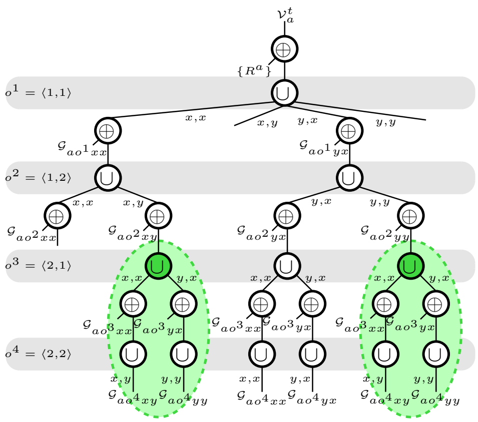
Planning under uncertainty in multiagent systems can be neatly formalized as a *decentralized partially observable Markov decision process (Dec-POMDP)*, but solving a Dec-POMDP is a complex (NEXP-complete) task. Communication can mitigate some of these complexities; by allowing agents to share their individual observations the problem reduces to a so-called *multiagent POMDP (MPOMDP)*, a special instance of the standard POMDP [2] which is in PSPACE. However, this model requires the agents to perform full synchronization of their knowledge before selecting a next action, which is inappropriate in domains in which agents may need to act fast in response to their individual observations.

A prime example is decentralized protection control in electricity distribution networks by so-called Intelligent Electronic Devices (IED). As power grids move towards integrating more distributed generation capability (e.g., provided by solar panels or fuel cells), more intricate protection schemes have to be developed as power flow is no longer unidirectional. In response, modern IEDs not only decide based on locally available sensor readings, but can receive information from other IEDs through a communication network with deterministic delays. When extreme faults such as circuit or cable failures occur, however, no time can be wasted waiting for information from other IEDs to arrive.

2 Tree-based pruning methods for MPOMDP-DC models

In this paper we focus on a class of problems where agents share their individual observations with a one step delay. That is, agents act using a *one step delayed sharing pattern*, resulting in an *MPOMDP with delayed communication (MPOMDP-DC)*. Solutions for such settings are also useful under longer delays [4]. Moreover, this class is particularly interesting, because it avoids the delay in action selection due to synchronization, while it is very similar to the standard POMDP. However, even though dynamic programming algorithms date back to the eighties, computational difficulties have limited the model's applicability.

The MPOMDP-DC value function is piecewise-linear and convex over the joint belief space, which is a property exploited by many regular POMDP solvers. However, *incremental pruning (IP)* [1], that performs a key operation, the so-called *cross-sum*, more efficiently, is not directly able to achieve the same improvements under delayed communication. A naive application of this technique (NAIVE IP) needs to loop over a number of decentralized control laws that is exponential both in the number of agents and in the number of observations. We target this additional complexity by proposing two novel methods that operate over a tree structure [3], illustrated in Figure 1a. These methods prune exactly the same vectors as NAIVE IP, but they iterate over the set of candidate vectors in a different way: NAIVE IP loops over all decentralized control laws β, while TBP methods exploit the similar parts in different β. The first method, called TBP-M for tree-based pruning with memoization, avoids *duplicate* work by



(a) The computation tree of $\mathcal{V}_a^t$.

Problem(h)	TBP-M	TBP-BB	NAIVE IP
Dec-Tiger(5)	0.13	0.09	0.23
Dec-Tiger(10)	0.31	0.43	0.73
Dec-Tiger(15)	0.98	1.44	2.54
OneDoor(3)	53.64	1546.73	304.72
GridSmall(2)	3.93	125.45	64.03
MG2x2(2)	171.07	2689.35	382093.00
MG2x2(3)	640.70	11370.40	
MG2x2(4)	1115.06	24125.30	
D-T Creaks(2)	63.14	93.16	109.27
D-T Creaks(3)	149.06	172.79	1595.17
D-T Creaks(4)	203.44	292.67	4030.47
D-T Creaks(5)	286.53	619.25	8277.32
Box Push.(2)	132.13	6663.04	1832.98

(b) Experimental results.

Figure 1: (a) Example of the identified tree structure. (b) Timing results (in s).

caching the result of computations at internal nodes and thus accelerates computation at the expense of memory. The second algorithm, branch and bound (TBP-BB), tries to avoid *unnecessary* computation by making use of upper and lower bounds to prune parts of the tree, providing a different space/time tradeoff.

The empirical evaluation of the proposed methods on a number of test problems shows a clear improvement over NAIVE IP, as presented in Table 1b. TBP-M provides speedups of up to 3 orders of magnitude. TBP-BB does not consistently outperform the baseline, but is still able to provide large speedups on a number of test problems, while using little memory.

3 Conclusions

In our work we considered multiagent planning under uncertainty formalized as a multiagent POMDP with delayed communication (MPOMDP-DC). A key feature of this model is that it allows a fast response to certain local observations, relevant in time-critical applications such as intelligent power grid control. We proposed two novel methods for computing optimal control policies in this model, and the experimental results show that we have successfully mitigated the additional complexity that the MPOMDP-DC backup exhibits over the MPOMDP one, allowing for the solution of larger problems.

Acknowledgments

M.S. is funded by the FP7 Marie Curie Actions Individual Fellowship #275217 (FP7-PEOPLE-2010-IEF). Research supported by AFOSR MURI project #FA9550-09-1-0538 and NWO CATCH project #640.005.003.

References

[1] Anthony Cassandra, Michael L. Littman, and Nevin L. Zhang. Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes. In *UAI*, pages 54–61. Morgan Kaufmann, 1997.

[2] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, 1998.

[3] Frans A. Oliehoek and Matthijs T. J. Spaan. Tree-based pruning for multiagent POMDPs with delayed communication. In *AAAI*, 2012.

[4] Matthijs T. J. Spaan, Frans A. Oliehoek, and Nikos Vlassis. Multiagent planning under uncertainty with stochastic communication delays. In *ICAPS*, pages 338–345, 2008.