# Influence-Optimistic Local Values for Multiagent Planning

# (Extended Abstract)

Frans A. Oliehoek
University of Amsterdam
University of Liverpool
fao@liverpool.ac.uk

Matthijs T. J. Spaan
Delft University of Technology
The Netherlands
m.t.j.spaan@tudelft.nl

Stefan J. Witwicki
Swiss Federal Institute of
Technology (EPFL)
stefan.witwicki@epfl.ch

## ABSTRACT

Over the last decade, methods for multiagent planning under uncertainty have increased in scalability. However, many methods assume value factorization or are not able to provide quality guarantees. We propose a novel family of *influence-optimistic* upper bounds on the optimal value for problems with 100s of agents that do not exhibit value factorization.

## Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent Systems*

## General Terms

Algorithms

## Keywords

Multi-agent planning, factored Dec-POMDPs

## 1. INTRODUCTION

Recent years have seen the development of methods for multiagent planning under uncertainty that scale to tens or even hundreds of agents [7, 9, 2, 5, 11]. Many methods, however, provide approximate solutions *without* any guarantees on quality, leaving the user to wonder how good the results really are. Other methods provide guarantees, but only for restricted sub-classes of problems. For instance, heuristic search methods leverage structure in problems where the value function can be additively factored into local components (involving small subsets of agents) [7]. Unfortunately, there currently is no known method of computing upper bounds for problems that do not exhibit strict value factorization, thereby precluding the leverage of heuristic search techniques.

Our work addresses this problem by introducing *influence-optimistic upper bounds (IO-UBs)* for factored Dec-POMDPs (fDec-POMDPs) that do not admit value factorization. In particular, we consider problems (e.g., the FIREFIGHTINGGRAPH problem [5], illustrated in Fig. 1a) in which the state is factored and the *reward* function *is* additively factored into local components, but the *value* function *is not* due to the fact that the effect of actions of each agent can propagate through the system (e.g., since fire can spread to neighboring houses). Such locally-connected systems can be found in applications as traffic control [11] or communication networks [1, 3].

We derive bounds on very large fDec-POMDPs by subdividing them in *sub-problems (SPs)*, and by making optimistic assumptions with respect to the *influence* [6] that will be exerted by the rest of the system on each of these SPs. In numerically evaluating our bounds, we demonstrate how, for the first time ever, we can achieve a non-trivial guarantee that the solution found by a heuristic method for problems with hundreds of agents is close to optimal.

**Appears in:** *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*


## 2. INFLUENCE OPTIMISTIC BOUNDS

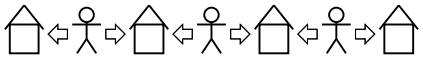
The basic idea is to decompose an fDec-POMDP into a set of *sub-problems* $c \in \mathcal{C}$ that contain a subset of agents, state factors and local reward functions, as shown in Fig. 1c. In particular, we apply a non-overlapping decomposition $\mathcal{C}$ (i.e., a partitioning) of the reward functions $\{R^l\}$ of the original factored Dec-POMDP into SPs $c \in \mathcal{C}$, and compute an IO upper bound $\hat{V}_c^{IO}$ for each $c$. Our *global influence-optimistic upper bound* is then given by:

$$\hat{V}^{IO} \triangleq \sum_{c \in \mathcal{C}} \hat{V}_c^{IO}.$$

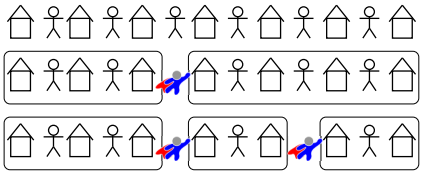
In Fig. 1b we illustrate the construction of global upper bounds $\hat{V}$ for the 6-agent FFG problem. Shown are the original problem (top row) and two possible decompositions into SPs. The second row specifies a decomposition into two SPs, while the third row uses three SPs. As illustrated, our decomposition eliminates certain agents completely and replaces them with idealized superhero-agents that provide optimistic influences. For instance, in the second row, the computation of $\hat{V}_c^{IO}$ for *both* SPs ($c$ = 1,2) assumes that agent 3 will always fight fire at $c$; in effect, our bounds assume that agent 3 fights fire at both house 3 and house 4 simultaneously.

## 3. LOCAL BOUNDS

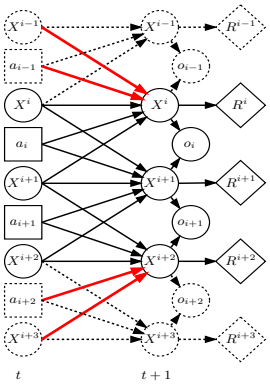
In order to compute *local* IO UBs $\hat{V}_c^{IO}$ for an SP $c$, we introduce three techniques. All three make optimistic assumptions with respect to the *influences* exerted on $c$ by the rest of the system: they assume that the highlighted *influence links* in Fig. 1c will lead to the most favorable transitions. The techniques differ in additional optimistic assumptions that they make: *IO-Q-MMDP* assumes that



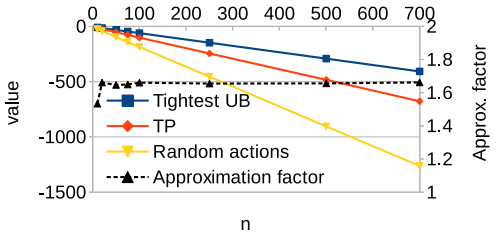
(a) FireFightingGraph.



(b) Global upper bounds.



(c) A 2-agent sub-problem within 4-agent FFG.




(d) The global IO-Q-Dec-POMDP upper bound on large FFG instances.

Figure 1: The FFG benchmark, its decompositions into sub-problems and results.

the SP is fully observable, *IO-Q-MPOMDP* assumes that the agents can communicate freely, and *IO-Q-Dec-POMDP* makes no additional assumptions. For more details, please see [4].

## 4. RESULTS

We report on the ability to provide informative *global* upper bounds. For other experiments that compare the different types of IO-UBs, investigate how they are affected by the *influence strength,* and demonstrate their potential in heuristic influence search [10], please see the long version of this paper [4].

Figure 1d (left y-axis) plots the value of the tightest upper bound we could find among different SP partitions (SP sizes ranging from $n = 2$–$5$) to investigate the guarantees that it can provide for transfer planning (TP) [5], which is one of the methods capable of providing solutions for large factored Dec-POMDPs. The value of TP, $V^{TP}$, is determined using 10.000 simulations of the found joint policy leading to accurate estimates. To put the results into context, we also show the value of a random policy. Finally, we show (right y-axis in Fig. 1d) what we call the *empirical approximation factor (EAF)*: $EAF = \max\{\frac{\hat{V}^{IO}}{V^{TP}}, \frac{V^{TP}}{\hat{V}^{IO}}\}$, a number comparable to the approximation factors of *approximation algorithms* [8].

As shown, the upper bound is relatively tight: the solutions found by TP lie typically within an EAF of 1.4–1.7, thus providing firm guarantees for solutions of factored Dec-POMDPs with up to 700 agents. Moreover, we see that the EAF stays roughly constant for the larger problem instances indicating that *relative* guarantees do not degrade as the number of agents increases. A similar approach for the Aloha benchmark also indicates that IO-UBs can be very tight [4]. In particular, we find EAFs $\leq 1.06$ for up to 250 agents, and for $n = 50$ agents we find an EAF of 1.00, essentially guaranteeing that the heuristic TP solution is optimal.

## 5. CONCLUSIONS

Here, we have introduced the first general techniques for computing upper bounds for large Dec-POMDPs that, despite containing factored structure, do not exhibit value factorization. We have demonstrated their usefulness in bounding existing approximate methods on problems involving 100s of agents. This paper focused on the finite-horizon case, but the principle of influence optimism can be applied in infinite-horizon settings too. Our techniques can also be modified to compute 'pessimistic' influence (i.e., lower) bounds. In an extended version of this paper, we provide evidence that the upper bounds may also be useful in improving the effectiveness of heuristic influence search, and discuss further potential applications to multiagent planning [4].

## Acknowledgments

F.O. is supported by NWO Innovational Research Incentives Scheme Veni #639.021.336.

## REFERENCES

[1] E. A. Hansen, D. S. Bernstein, and S. Zilberstein. Dynamic programming for partially observable stochastic games. In *AAAI*, 2004.

[2] A. Kumar, S. Zilberstein, and M. Toussaint. Scalable multiagent planning using probabilistic inference. In *IJCAI*, 2011.

[3] A. Mahajan and M. Mannan. Decentralized stochastic control. *Annals of OR*, 2014.

[4] F. A. Oliehoek, M. T. J. Spaan, and S. Witwicki. Influence-optimistic local values for multiagent planning — extended version. *ArXiv e-prints*, arXiv:1502.05443, 2015.

[5] F. A. Oliehoek, S. Whiteson, and M. T. J. Spaan. Approximate solutions for factored Dec-POMDPs with many agents. In *AAMAS*, 2013.

[6] F. A. Oliehoek, S. Witwicki, and L. P. Kaelbling. Influence-based abstraction for multiagent systems. In *AAAI*, 2012.

[7] P. Varakantham, J. Marecki, Y. Yabu, M. Tambe, and M. Yokoo. Letting loose a SPIDER on a network of POMDPs: Generating quality guaranteed policies. In *AAMAS*, 2007.

[8] V. V. Vazirani. *Approximation Algorithms*. Springer-Verlag, 2001.

[9] P. Velagapudi, P. Varakantham, P. Scerri, and K. Sycara. Distributed model shaping for scaling to decentralized POMDPs with hundreds of agents. In *AAMAS*, 2011.

[10] S. Witwicki, F. A. Oliehoek, and L. P. Kaelbling. Heuristic search of multiagent influence space. In *AAMAS*, 2012.

[11] F. Wu, S. Zilberstein, and N. R. Jennings. Monte-Carlo expectation maximization for decentralized POMDPs. In *IJCAI*, 2013.