Memoryless Adversaries in Imperfect Information Games

Extended Abstract

Dhananjay Raju
The University of Texas at Austin
Austin, United States

Georgios Bakirtzis
The University of Texas at Austin
Austin, United States

Ufuk Topcu
The University of Texas at Austin
Austin, United States

ABSTRACT

Given an agent with limited sensing capabilities, we analyze whether it is possible to deploy a new agent in the operational space of the preexisting agent in a *safe* manner. One approach for modeling the interaction of the introduced agent with its environment, which contains the preexisting agent, is through a two-player game of imperfect information. However, the computational cost of solving this game is prohibitive. Restricting the preexisting agent's strategy to just *memoryless* strategies and assuming that the introduced agent has perfect information alleviates the computational cost while still modeling realistic environments. The proposed algorithm for solving the game finds a winning strategy for the introduced agent by solving a quantified Boolean formula (QBF) for the game. We justify this approach by establishing a matching PSPACE lower bound. We also show that this result holds even when the preexisting agent uses *bounded history* to condition its play.

KEYWORDS

Imperfect information games; memoryless adversaries; synthesis

ACM Reference Format:
Dhananjay Raju, Georgios Bakirtzis, and Ufuk Topcu. 2023. Memoryless Adversaries in Imperfect Information Games: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023,* IFAAMAS, 3 pages.

1 INTRODUCTION

Two-player games on graphs model agent interactions with their operational environment, which may include other agents. Most results about two-player games on graphs hypothesize that the agent has perfect information. Imperfect-information games in which an imperfect-information player plays against a perfect-information player relax the perfect-information hypothesis in an attempt to model scenarios in which agents rely on sensors to make the control decisions. The imperfect information player models a robotic agent, and the perfect-information player models the environment. To date, these games are well-understood from the perspective of the agent players' perspective but lack semantics from the perspective of the environment.

Motivating Example. Consider a storage warehouse in which a restocking robot P initially located at $(0, 0)$ uses directional markers (i.e., observational strategies) to navigate a warehouse in order to restock one of red, blue or yellow shelves (figure 1). We assume that this robot P does not have the capability to distinguish dynamic obstacles. These directional markers do not fix the policy for this robot P. We introduce a new robot F with perfect information, meaning it knows (1) the observation function of P, (2) the objective of P, and (3) current state of P. We are particularly interested in analyzing whether an introduced robot F can satisfy a *reachability objective* without violating the *safety property* of colliding with the old restocking robot P as it operates in the same warehouse. Due to the assumption that the old robot P cannot distinguish dynamic obstacles, it has imperfect information in the joint-operation scenario. Therefore, the onus is on the newly deployed robot F to avoid colliding with the old robot while navigating the shared environment.

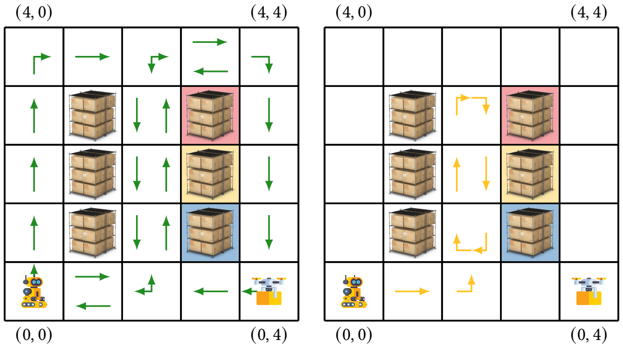

Figure 1: (Left) A restocking robot F (at $(0, 4)$) uses directional markers to navigate a warehouse with a preexisting agent P (at $(0, 0)$). (Right) A possible forgetful strategy for the drone (blue and yellow respectively).

The fundamental insight is that while the existing agent P is not cooperative concerning agent F, it is also not a *true adversary* to F. Agent P may coincidentally block agent F, but it is not the objective of agent P to interfere with the agent's F path.

2 MEMORYLESS IMPERFECT-INFORMATION PLAYER

The analysis of the game from the perspective of memoryless adversaries is motivated by preexisting agents that are restricted to observational strategies. Formally, the imperfect-information player is restricted to using strategies of the form $S_P : V^{\text{public}} \to \Sigma_P$. Therefore, the number of strategies for P is $m = |V^{\text{public}}|^{|\Sigma_P|}$. In the naive approach, F maintains a subset of all the possible strategies that are consistent with the play (so far). Agent F now simply plays in a manner such that it addresses all the remaining strategies. Intuitively, F learns the strategy (rules of the strategies) of the imperfect-information player. However, the use of a subset construction on the possible strategies of P leads to a prohibitive

---

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

Algorithm 1: Strategy synthesis against bounded P.

```
Input: Game $G$ on arena $A$,
       objective $\psi$, and
       length of history $k$.
Output: Strategy tree $T$ for player $F$.
$n \leftarrow \left|V^{\text{public}}\right|^k$
$m \leftarrow n\left(n\left|V^{\text{private}}\right| + 1\right)$ ▷ bound on number of rounds for $F$
$T \leftarrow$ solve-qbf($\psi_{A^{\text{forgetful}}}$) ▷ construct QBF (theorem 2.2)
if $T \neq \emptyset$ then
    return $(T, \top)$
end
return $\bot$
```

computational cost (2-EXPTIME). But, this is not the true complexity of this problem.

Solving the game by encoding it as a QBF

When playing against an agent with limited memory, player F has more ways to defeat it. However, finding a winning strategy is computationally harder. We bound the maximum number of rounds needed by the perfect-information player F to hit a winning loop.

Lemma 2.1. *If the perfect-information player can win, then in*

$$O\left(\left(\left|V^{private}\right|\left|V^{public}\right| + 1\right)\left|V^{public}\right|\right)$$

rounds she can force a loop whose minimal parity is the parity of the perfect-information player.

It is possible to encode perfect-information reachability (safety) games that end after a fixed number of rounds as a QBF with the number of quantifier alternations equal to the number of rounds [1]. We exploit that player P is forced to use memoryless strategies to construct a similar QBF encoding for parity objectives.

Theorem 2.2. *If player P is memoryless, then WIN_F is contained in PSPACE.*

In the scenario where the imperfect-information player P can use histories to condition its next action, we propose a new arena $A_{\text{forgetful}}$ derived from a given arena A where k moves are explicitly wasted before making the new move. Consequently, we derive a new upper bound on the number of rounds.

3 EVALUATION OF SYNTHESIS WITH IMPERFECT INFORMATION GAMES

Under a limit on the opponent's ability to condition on the history, we can use a bound on the game's length to symbolically encode the game into a QBF. Our use of QBFs is partially supported by the recently investigated efficacy of QBFs for solving problems in model-checking, synthesis, and planning [13]. The result of solving the QBF encoding for the game (theorem 2.2) is a decision tree that serves as a strategy (extensive form) for the full information player (algorithm 1).

We used our algorithm to find the winning strategy for the newly introduced drone (figure 1) to reach a cell in $\{(2,1),(2,2),(2,3)\}$ without colliding with P. The strategy obtained by solving the QBF requires F to move to $(3,1)$ and wait for a turn; if P comes back to $(2,1)$, F moves along column 0 to reach the objective $(2,3)$. Otherwise, F moves to $(2,1)$.

Table 1: Evaluation of approach on a 10x10 grid against a memoryless P.

distinct observations	size of arena due to subset construction	bound on number of rounds	synthesis time (sec)
4	10^4	16	159
5	10^{10}	25	234
6	10^{19}	36	333
7	10^{38}	49	359
8	10^{77}	64	468
9	10^{154}	81	602
10	∞	100	–

More generally, in a square grid world like the one depicted in figure 1, there are static obstacles in the middle rows that create narrow corridors. At any time, the existing agent P can move to a neighboring cell if the directional markers allow it. On the other hand, the newly introduced drone F can move to an adjacent cell if it is unblocked. Let $\left(x^P, y^P\right)$ and $\left(x^F, y^F\right)$ denote the current positions of the agents P and F, respectively. Further, let $SO = \{(x_1, y_1), \cdots (x_s, y_s)\}$ denote the locations of the static obstacles. The drone F must satisfy the following two objectives a) collision avoidance ϕ_1: The agent F must avoid collision with P and other static obstacles, specified as $\phi_1 = \phi_{SO} \wedge \phi_P$, where

$$\phi_{SO} = \bigvee_{(x_o, y_o) \in SO} \square\left(\neg\left((x^F = x^o) \wedge (y^F = y^o)\right)\right) \text{ and}$$

$$\phi_P = \square\left(\neg\left((x^F = x^P) \wedge (y^F = y^P)\right)\right).$$

and b) reachability ϕ_2: Agent F must reach objective state $R = (x^R, y^R)$, specified as $\phi_2 = \lozenge\left(\left((x^F = x^R) \wedge (y^F = y^R)\right)\right)$.

The hardness of solving QBFs increases with the number of alternations in the formula. Experiments (table 1) show that the proposed symbolic approach works when the inferred bound on the number of rounds is less than 100. Even for problems where the number of rounds is small, the size of the game arena is huge, making the problem intractable (if solved by solving the perfect information game). The experiments are performed on an Intel core i7 2.4 GHz machine with 16 GB memory.

4 CONCLUSION

Our work explores a new type of information asymmetry between the two players of a game: the preexisting player has limited information on the game's current state. The introduced player needs to learn from the environment's (preexisting agent) behavior *while* avoiding situations in which the unknown behavior of the environment player could cause undesired interactions (e.g., colliding with the preexisting robot) leading to the loss of the game [5]. As such, we believe that our work prepares the ground for more formal approaches to the design of *self-adaptive* systems.

ACKNOWLEDGMENTS

This work was supported in part by DARPA D19AP00004 and ARL ACC-APG-RTP W911NF1920333.

REFERENCES

[1] Rajeev Alur, P. Madhusudan, and Wonhong Nam. 2005. Symbolic computational techniques for solving games. *Int. J. Software Tools Technol. Trans.* 7, 2 (Apr 2005), 118–128. https://doi.org/10.1007/s10009-004-0179-0

[2] Rajeev Alur, Salar Moarref, and Ufuk Topcu. 2018. Compositional and symbolic synthesis of reactive controllers for multi-agent systems. *Inf. Comput.* (2018). https://doi.org/10.1016/j.ic.2018.02.021

[3] Yackolley Amoussou-Guenou, Souheib Baarir, Maria Potop-Butucaru, Nathalie Sznajder, Léo Tible, and Sébastien Tixeuil. 2020. On the Encoding and Solving of Partial Information Games. In *Networked Systems - 8th International Conference, NETYS 2020, Marrakech, Morocco, June 3-5, 2020, Proceedings (Lecture Notes in Computer Science)*. Springer. https://doi.org/10.1007/978-3-030-67087-0_5

[4] Béatrice Bérard, Pascal Lafourcade, Laure Millet, Maria Potop-Butucaru, Yann Thierry-Mieg, and Sébastien Tixeuil. 2016. Formal verification of mobile robot protocols. *Distributed Comput.* (2016). https://doi.org/10.1007/s00446-016-0271-1

[5] Doran Chakraborty and Peter Stone. 2014. Multiagent learning in the presence of memory-bounded agents. *Autonomous Agents and Multi-Agent Systems* 28, 2 (Mar 2014), 182–213. https://doi.org/10.1007/s10458-013-9222-4

[6] M. Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. 2014. Clingo = ASP + Control: Preliminary Report. *ArXiv* abs/1405.3694 (2014).

[7] Marius Kloetzer and Calin Belta. 2010. Automatic Deployment of Distributed Teams of Robots From Temporal Logic Motion Specifications. *IEEE Trans. Robotics* (2010). https://doi.org/10.1109/TRO.2009.2035776

[8] Orna Kupferman, Yoad Lustig, Moshe Vardi, and Mihalis Yannakakis. 2011. Temporal Synthesis for Bounded Systems and Environments. *Symposium on Theoretical Aspects of Computer Science (STACS2011)* 9 (03 2011). https://doi.org/10.4230/LIPIcs.STACS.2011.615

[9] Thanh Hai Nguyen, Mason Wright, Michael P. Wellman, and Satinder Singh. 2017. Multi-Stage Attack Graph Security Games: Heuristic Strategies, with Empirical Game-Theoretic Analysis. In *Proceedings of the 2017 Workshop on Moving Target Defense, MTD@CCS 2017*. ACM. https://doi.org/10.1145/3140549.3140562

[10] Christos H. Papadimitriou and Mihalis Yannakakis. 1994. On complexity as bounded rationality (extended abstract). In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*. ACM. https://doi.org/10.1145/195058.195445

[11] John H. Reif. 1979. Universal Games of Incomplete Information. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing (STOC '79)*. https://doi.org/10.1145/800135.804422

[12] Sasha Rubin, Florian Zuleger, Aniello Murano, and Benjamin Aminof. 2015. Verification of Asynchronous Mobile-Robots in Partially-Known Environments. In *PRIMA 2015: Principles and Practice of Multi-Agent Systems*. https://doi.org/10.1007/978-3-319-25524-8_12

[13] Ankit Shukla, Armin Biere, Luca Pulina, and Martina Seidl. 2019. A Survey on Applications of Quantified Boolean Formulas. In *31st IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2019, Portland, OR, USA, November 4-6, 2019*. https://doi.org/10.1109/ICTAI.2019.00020

[14] Ichiro Suzuki and Masafumi Yamashita. 1999. Distributed Anonymous Mobile Robots: Formation of Geometric Patterns. *SIAM J. Comput.* (1999). https://doi.org/10.1137/S009753979628292X

[15] Christopher J. C. H. Watkins and Peter Dayan. 1992. Q-learning. *Mach. Learn.* (1992). https://doi.org/10.1007/BF00992698

[16] Tichakorn Wongpiromsarn, Alphan Ulusoy, Calin Belta, Emilio Frazzoli, and Daniela Rus. 2013. Incremental synthesis of control policies for heterogeneous multi-agent systems with linear temporal logic specifications. In *2013 IEEE International Conference on Robotics and Automation*. IEEE. https://doi.org/10.1109/ICRA.2013.6631293