Partially Observable MDPs

Revealing ones, and other decidable classes thereof

Guillermo A. Pérez + Pierre Vandenhove

Formal methods & verification seminar @ ULB, 2026

Building bridges: RL–Automata theory

Paper under consideration

We first present here

Revelations: A Decidable Class of POMDPs with Omega-Regular Objectives
by Marius Belly, Nathanaël Fijalkow, Hugo Gimbert,
Florian Horn, Guillermo A. Pérez, Pierre Vandenhove

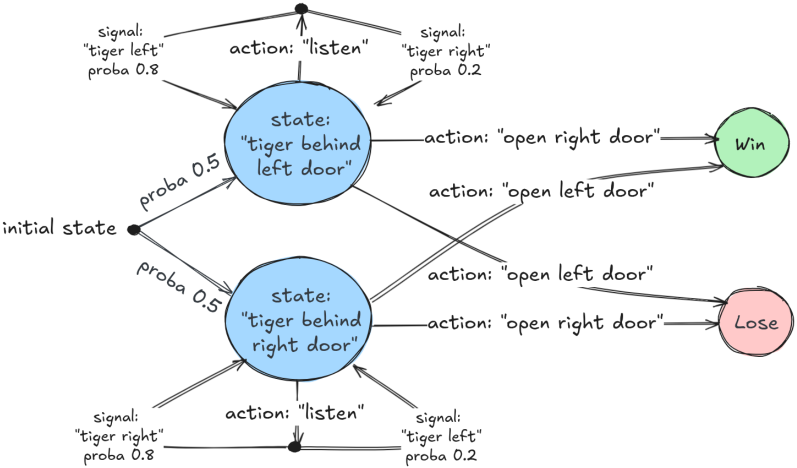
published at AAAI '25.

Implementation at `https://github.com/gaperez64/pomdps-reveal`.

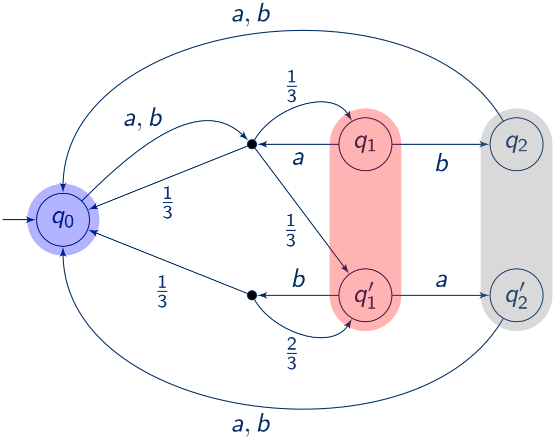
- Code for RL solutions + comparison with our algo.
- Benchmarks by M. Chmelik, and Cassandra `https://www.pomdp.org/`

An example: “Tiger”

generated with ChatGPT



Partially observable MDPs



States Q, **initial state** q_0, **actions** Act, **observations** Obs.
Strategies are functions $(\mathsf{Act} \times \mathsf{Obs})^* \to \mathcal{D}(\mathsf{Act})$.

Beyond immediate observations? Belief!

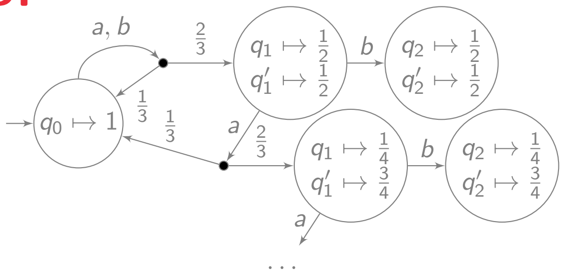
From $b \in \mathcal{D}(Q)$, we play a and receive observation o.
Then, we believe we are in q' with probability...

$$b'(q') = \frac{obs(o \mid q', a) \sum_{q \in Q} P(q' \mid q, a) b(q)}{\sum_{q' \in Q} obs(o \mid q', a) \sum_{q \in Q} P(q' \mid q, a) b(q)}$$

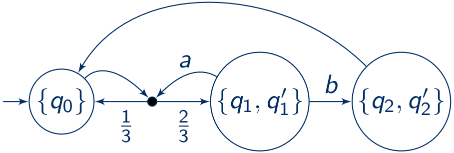
This assumes **observations** depend on the action played and the target state of the transition...
they could just as well be deterministic and target-state-dependent only.

Belief (support) MDP

POMDPs induce **infinite**
belief MDPs:



Finite: only keep
belief **supports**:



When does the analysis of the belief **support** MDP suffice?

In general, neither sound nor complete. . .

What can we solve in POMDPs?

Unbounded horizon problems

- Expected discounted reward optimization
- Expected limit-average reward optimization
- Omega-regular objective sat-probability optimization
- ...

Bad news [Madani, Hanks, Condon '99]

Already asking whether there exists a finite-memory strategy σ such that:

$$\Pr_{\sigma}(\mathrm{Reach}(T)) \geq \frac{1}{2}$$

is undecidable. Most by reduction from (the gap version of) the emptiness problem for probabilistic automata.

Lots of bad news

8 *O. Madani et al. / Artificial Intelligence 147 (2003) 5–34*

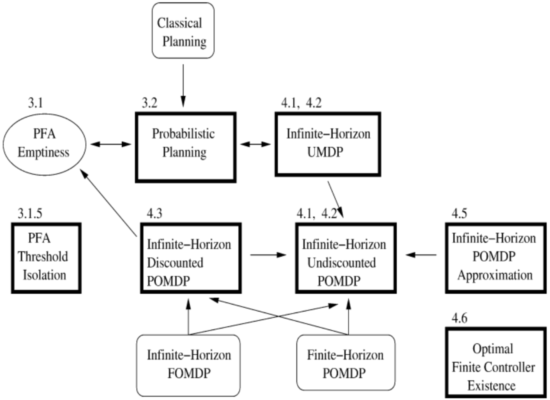


Fig. 1. Summary of Undecidability Results. Problems in bold rectangles are those established as undecidable in this paper, with the proofs starting from the result in the oval. In the rounded rectangles are related problems with previously known complexity results. Arrows point from "easier" to "harder" problems. Above each problem is the section number where the problem is addressed.

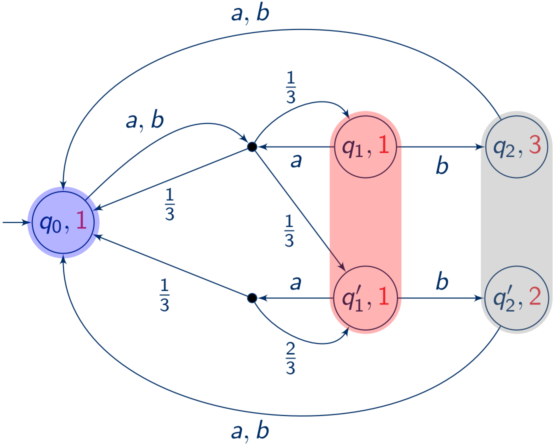
Good news for omega-regular objectives

- Common **objectives**:
  - **Reachability**: a good state is eventually visited,
  - **Büchi**: $p\colon Q \to \{1, 2\}$; good states (2) are visited infinitely often,
  - **coBüchi**: $p\colon Q \to \{0, 1\}$; bad states (1) are visited finitely often.
- More generally: function $p\colon Q \to \{0, \ldots, d\}$ assigning **priorities** to **states**.
- **Parity objective**: the **maximal** priority seen infinitely often is **even**.
- **Question**: does there exist an **almost-sure** strategy?

Decidability in POMDPs [Baier et al. '12; Chatterjee et al. '16]

- Almost-sure **reachability**, **safety**, and **Büchi** are **EXPTIME-complete**.
- Almost-sure **coBüchi** (and therefore **parity**) are **undecidable**.

Example of a difficult POMDP

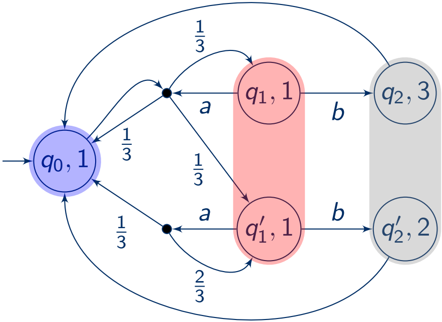


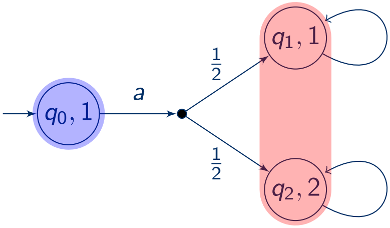
Almost-sure strategy? Move to q_2/q'_2 when increasingly high probability to be in q'_1.

Weak revelations

Weak revelations

A POMDP is **weakly revealing** if for all strategies, almost surely, the **current state is known** infinitely often.





Not weakly revealing

Weakly revealing: q_0 is visited infinitely often

Weak revelations

Weak revelations

A POMDP is **weakly revealing** if for all strategies,
almost surely, the **current state is known** infinitely often.

When a revealing history happens, the finite belief **support** MDP contains **as much information** as the infinite belief MDP.

$$\{q_0\} \approx q_0 \mapsto 1$$

Weak revelations: results

"Weakly revealing" is a semantic property, but is **decidable**.

Priorities $\{0, 1, 2\}$ (encompassing Büchi and coBüchi)

There exists an almost-sure strategy...

in a **weakly revealing POMDP** $\mathcal{P}$ $\iff$ in the **belief support MDP** of $\mathcal{P}$.

Decidability

Almost-sure **parity** $\{0, 1, 2\}$ for **weakly revealing** POMDPs is EXPTIME-complete.

Algorithm: solve the **belief support MDP** $\rightsquigarrow$ in EXPTIME.

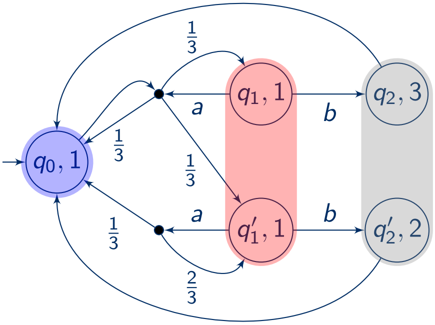
Why restrict to parity $\{0, 1, 2\}$? Unfortunately...

Full parity remains undecidable

Undecidability

Almost-sure **parity** $\{1, 2, 3\}$ is **undecidable** for **weakly revealing** POMDPs.

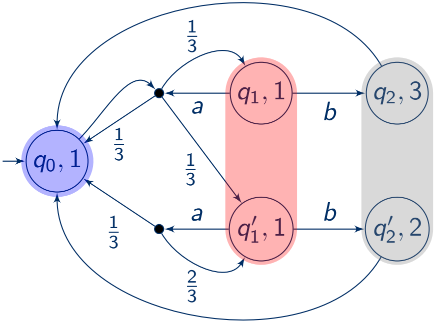
Belief support MDP does not help for this **weakly revealing** POMDP with priorities 1, 2, 3.



Strong revelations

Strong revelations

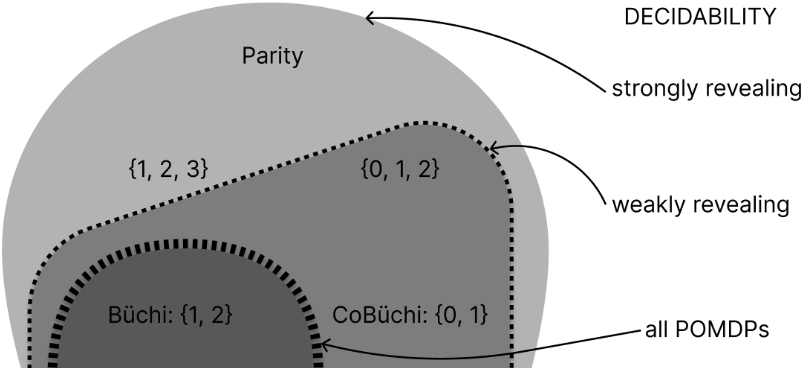
A POMDP is **strongly revealing** if for every transition $q \xrightarrow{a} q'$, there is a non-zero probability of **"revealing"** q'.



Not strongly revealing: $q_1 \xrightarrow{a} q_1'$ is a possible transition, but nothing can reveal q_1' with certainty.

Almost sure parity is decidable for strongly revealing POMDPs!

Decidability boundary



Revelations make POMDPs easier and allow for simpler algorithms/policies.

Follow-up paper

Our results on **strongly** revealing POMDPs were very recently complemented by

Revealing POMDPs: Qualitative and Quantitative Analysis for Parity Objectives
by Ali Asadi, Krishnendu Chatterjee, David Lurie, and Raimundo Saona

showing that **strongly revealing** POMDPs are even more decidable than we thought:

- the limit-sure (also called *value*-1) problem for parity objectives is decidable, and
- the value of parity objectives can be approximated.

Everything is in EXPTIME.

Strongly revealing games

- Everything ω-**regular becomes decidable** in strongly revealing POMDPs.
- Is "strongly revealing" a sufficient property to **make all partially observable "stuff" with ω-regular objectives decidable**?
- **NO**: the existence of almost-sure strategies in partially observable zero-sum **games** with coBüchi objectives, even under strong revelations, are still undecidable.

Interesting open problems

Expected limit average and discounted reward

Does there exist a strategy σ such that $E_\sigma[Val] \geq t$?

- Undecidable in general [Madani, Hanks, Condon '99]
- For revealing POMDPs this may still be (un)decidable

Probabilistic constraints

Does there exist a (finite-memory) strategy σ such that:

$$\Pr_{\sigma}(\mathrm{Val} \geq t) = 1$$

in a given (revealing) POMDP?

Beware of the competition! 5+ revealing POMDP submissions to AAAI'26

BONUS: second paper

We now (briefly) turn to

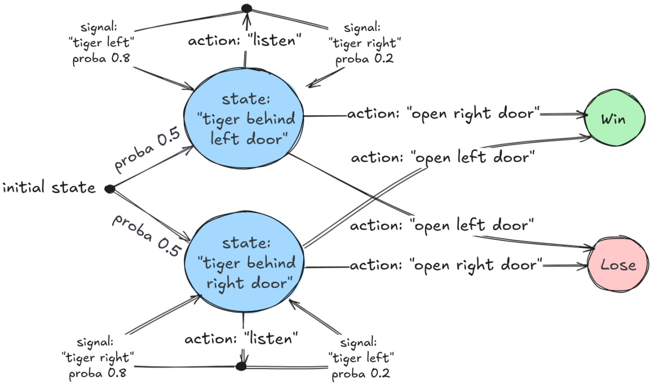
Computing the Reachability Value of Posterior-Deterministic POMDPs
by N. Fijalkow, A. Ghosh, R. Kniazev, G.A. Pérez, P. Vandenhove

submitted to ICALP '26.

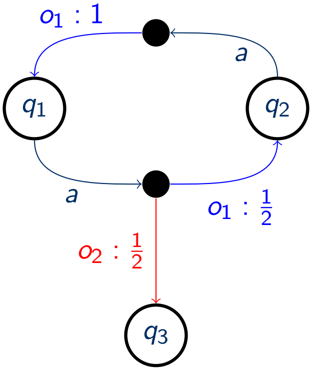
Posterior-deterministic POMDPs

Posterior determinism

A POMDP is **posterior-deterministic** if for all states q and all action-observation pairs (a, o) we have at most one transition $q \xrightarrow{a,o} q'$.



Retroactive revelations



Our results

Main theorem (reachability value)

For any posterior-deterministic POMDP $\mathcal{P}$, initial belief $\mathbf{b}$, and a tolerance $\epsilon > 0$, one can compute a value $v \in [0,1]$ such that $|\Pr^{\mathcal{P}}_{\mathrm{Reach}}(\mathbf{b}) - v| \leq \varepsilon$ in triply-exponential time.

Indistinguishability. We look at end components with states p, q such that, for all histories $h = a_1 o_1 \ldots a_n$, the observation-distributions $\delta(p,h)$ and $\delta(q,h)$ are the same.

Tiger has a distinguishing blue pair of states! But not if the listening error is 0.5.

Conjecture/Theorem (WiP) (value-1 reachability)

The value-1 problem can be decided in single-exponential time for posterior-deterministic POMDPs.

Thanks!