Minimal Disclosure in Partially Observable Markov Decision Processes

Nathalie Bertrand[1] and Blaise Genest [2]

[1]INRIA Rennes, France

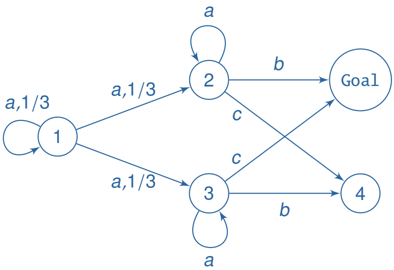
[2]CNRS UMI IPAL, Singapore

FSTTCS
December 14th 2011

Outline

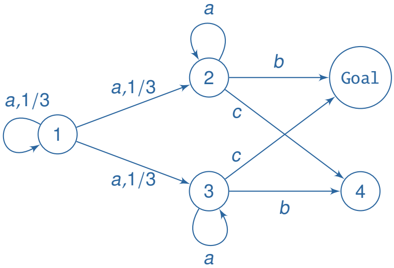
Markov Decision Processes (MDP)

States: Q; Actions: Act; Probabilistic transition function: Δ



Markov Decision Processes (MDP)

States: Q; Actions: Act; Probabilistic transition function: Δ

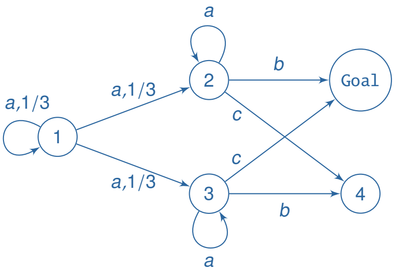


Strategy for the controller: based on actions and states

$$\sigma : Q \cdot (Act \cdot Q)^* \rightarrow Dist(Act)$$

Markov Decision Processes (MDP)

States: Q; Actions: Act; Probabilistic transition function: Δ



Strategy for the controller: based on actions and states

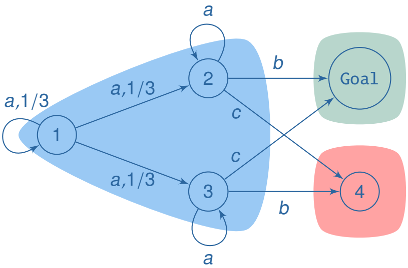
$$\sigma : Q \cdot (Act \cdot Q)^* \to Dist(Act)$$

Memoryless pure strategy to reach `Goal` almost-surely:

$$\sigma(1) = a,\ \sigma(2) = b,\ \sigma(3) = c$$

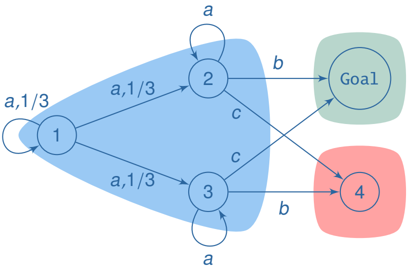
Partially Observable MDP (POMDP)

Partial observation: induced by partition $\mathcal{O}$



Partially Observable MDP (POMDP)

Partial observation: induced by partition $\mathcal{O}$

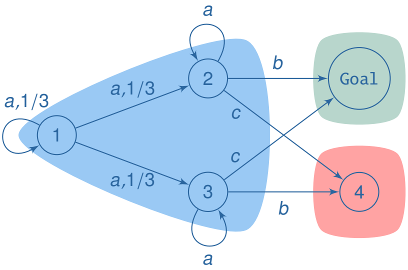


Strategy for the controller: based on actions and observations

$$\sigma : \mathcal{O} \cdot (Act \cdot \mathcal{O})^* \rightarrow Dist(Act)$$

Partially Observable MDP (POMDP)

Partial observation: induced by partition $\mathcal{O}$



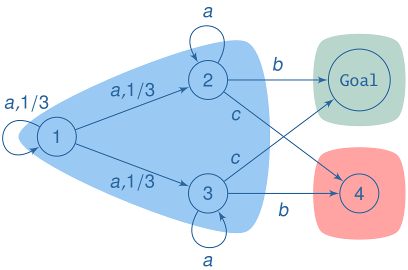
Strategy for the controller: based on actions and observations

$$\sigma : \mathcal{O} \cdot (Act \cdot \mathcal{O})^* \to Dist(Act)$$

No strategy to reach `Goal` almost-surely.

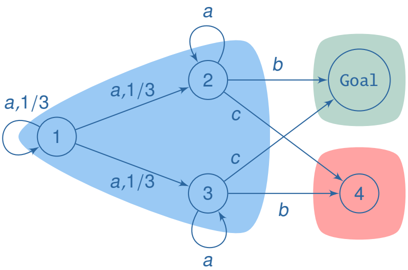
POMDP with disclosure

Additional request action to reveal the precise state of system.
Observations: partition + individual states



POMDP with disclosure

Additional request action to reveal the precise state of system.
Observations: partition + individual states

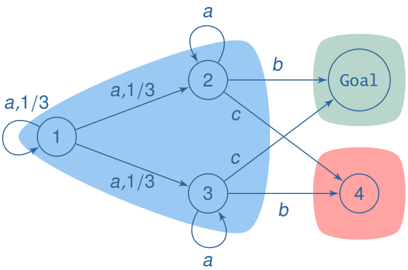


Strat. for the controller: based on extended actions and observations

$$\sigma : \mathcal{O}' \cdot (Act' \cdot \mathcal{O}')^* \to Dist(Act')$$

POMDP with disclosure

Additional request action to reveal the precise state of system.
Observations: partition + individual states



Strat. for the controller: based on extended actions and observations

$$\sigma : \mathcal{O}' \cdot (Act' \cdot \mathcal{O}')^* \rightarrow Dist(Act')$$

Cheap strategy to reach `Goal` almost-surely?

Problem statement

cost of a path = number of requests for disclosure
cost of a strategy σ =

- worst-case cost along σ-paths (max number of requests)
- average cost along σ-paths (expected number of requests)

Problem statement

cost of a path = number of requests for disclosure
cost of a strategy σ =

- worst-case cost along σ-paths (max number of requests)
- average cost along σ-paths (expected number of requests)

Problem statement

Finding almost-surely winning strategies that minimize:

- the worst-case cost, or
- the average cost

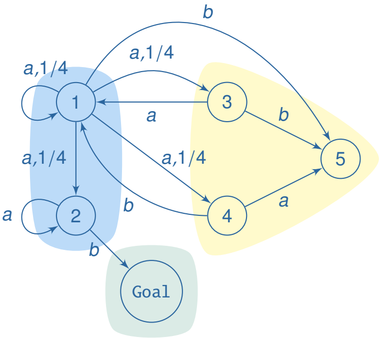
Outline

Belief

Belief: (distribution over) states the system can be in

Belief

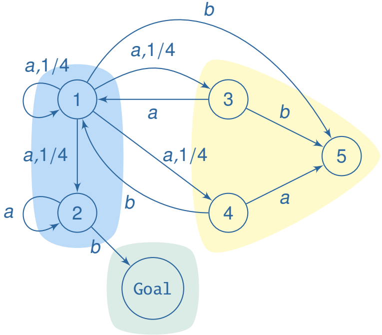
Belief: (distribution over) states the system can be in



$\{1\}$

Belief

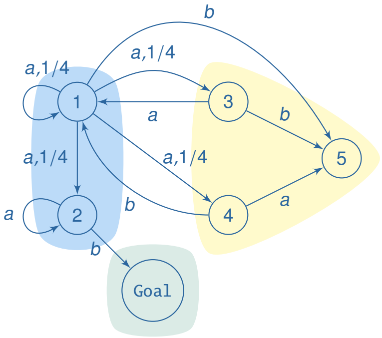
Belief: (distribution over) states the system can be in



$\{1\}$ $(a, \bullet)$

Belief

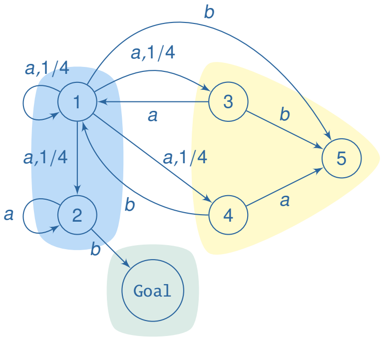
Belief: (distribution over) states the system can be in



$\{1\}\ (a, \bullet)\ \{1, 2\}$

Belief

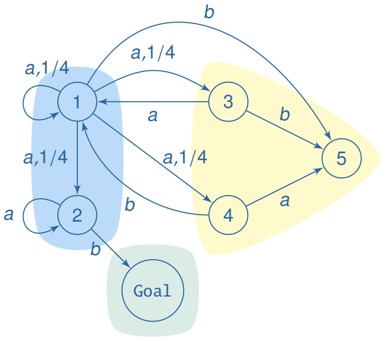
Belief: (distribution over) states the system can be in



$\{1\}$ $(a, \bullet)$ $\{1,2\}$ $(a, \bullet)$

Belief

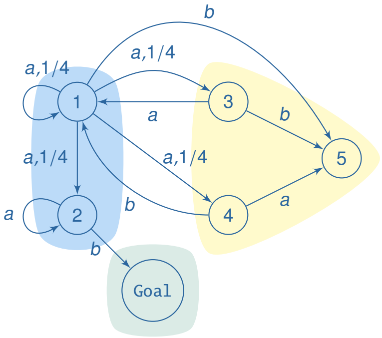
Belief: (distribution over) states the system can be in



$\{1\}\ (a, \bullet)\ \{1,2\}\ (a, \circ)\ \{3,4\}$

Belief

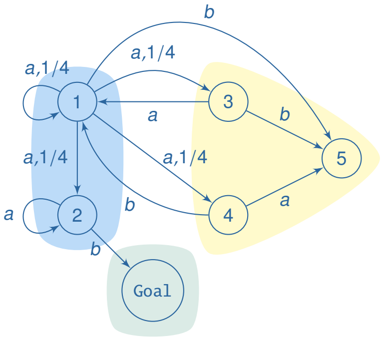
Belief: (distribution over) states the system can be in



$\{1\}$ $(a, \bullet)$ $\{1, 2\}$ $(a, \bullet)$ $\{3, 4\}$ $(b, \bullet)$

Belief

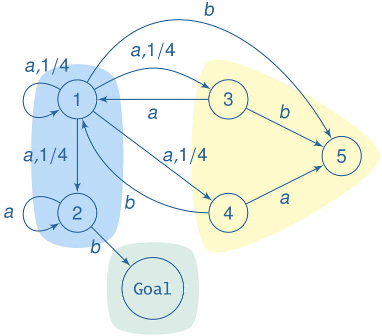
Belief: (distribution over) states the system can be in



$\{1\}\ (a, \bullet)\ \{1,2\}\ (a, \circ)\ \{3,4\}\ (b, \circ)\ \{5\}$

Belief

Belief: (distribution over) states the system can be in

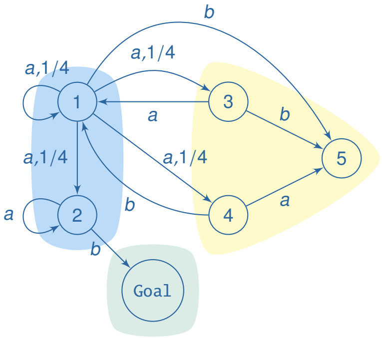


$\{1\}\ (a,●)\ \{1,2\}\ (a,●)\ \{3,4\}\ (b,●)\ \{5\}$

$\{1\}\ (a,●)\ \{3,4\}\ (\textit{req},\{4\})\ \{4\}\ (b,●)\ \{1\}$

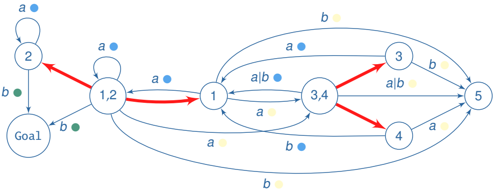
Belief

Belief: (distribution over) states the system can be in



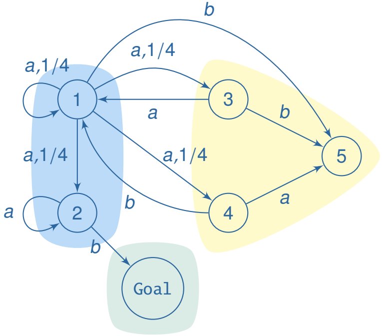
$\{1\}\ (a, ●)\ \{1,2\}\ (a, ●)\ \{3,4\}\ (b, ●)\ \{5\}$

$\{1\}\ (a, ●)\ \{3,4\}\ (req, \{4\})\ \{4\}\ (b, ●)\ \{1\}$



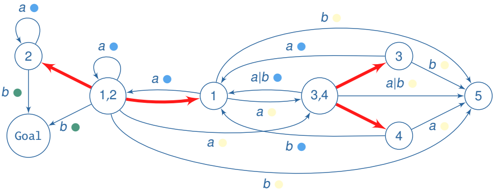
Belief

Belief: (distribution over) states the system can be in



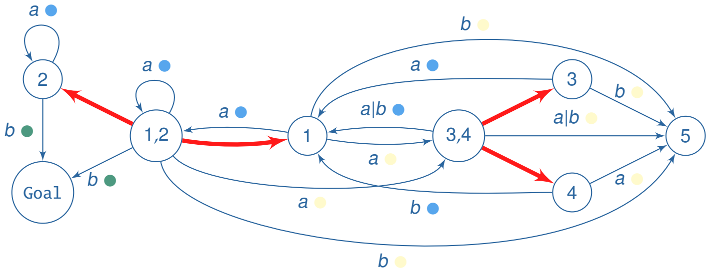
$\{1\}$ (a, ●) $\{1,2\}$ (a, ●) $\{3,4\}$ (b, ●) $\{5\}$

$\{1\}$ (a, ●) $\{3,4\}$ (req, $\{4\}$) $\{4\}$ (b, ●) $\{1\}$



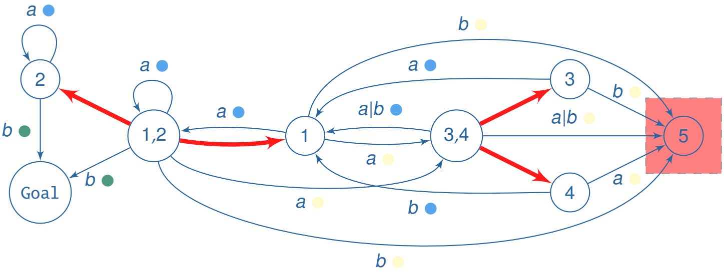
$up(S, a, O)$: belief update from S, after action a and observation O

Reaching the goal almost-surely



Reaching the goal almost-surely

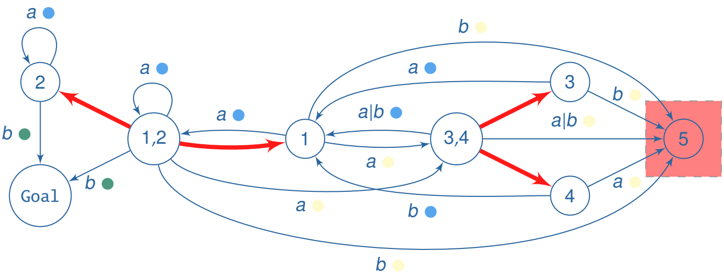
Lose: beliefs that contain a state losing in the (fully-observable) MDP



Reaching the goal almost-surely

Lose: beliefs that contain a state losing in the (fully-observable) MDP

$Win = \mathcal{B} \setminus Lose = W_{ok} \sqcup W_{req} \sqcup W_{safe}$

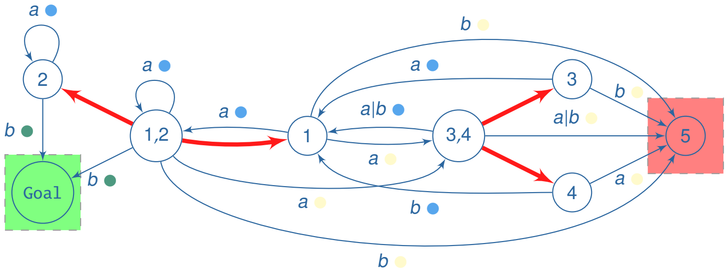


Reaching the goal almost-surely

Lose: beliefs that contain a state losing in the (fully-observable) MDP

$Win = \mathcal{B} \setminus Lose = W_{ok} \sqcup W_{req} \sqcup W_{safe}$

- $W_{ok} = \{S \mid S \subseteq \texttt{Goal}\}$

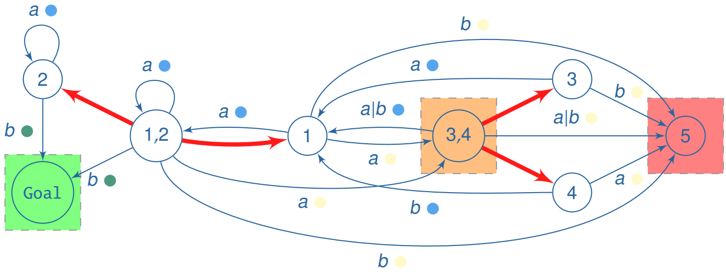


Reaching the goal almost-surely

Lose: beliefs that contain a state losing in the (fully-observable) MDP

$Win = \mathcal{B} \setminus Lose = W_{ok} \sqcup W_{req} \sqcup W_{safe}$

- $W_{ok} = \{S \mid S \subseteq \texttt{Goal}\}$
- $W_{req} = \{S \mid \forall a \in Act\ \exists O \in \mathcal{O},\ up(S, a, O) \in Lose\}$



Reaching the goal almost-surely

Lose: beliefs that contain a state losing in the (fully-observable) MDP

$Win = \mathcal{B} \setminus Lose = W_{ok} \sqcup W_{req} \sqcup W_{safe}$

- $W_{ok} = \{S \mid S \subseteq \mathtt{Goal}\}$
- $W_{req} = \{S \mid \forall a \in Act\ \exists O \in \mathcal{O},\ up(S, a, O) \in Lose\}$

Canonical family of strategies $(\sigma_n)_{n \in \mathbb{N}}$:

- In W_{req}, play *req*, and
- in W_{safe}, play *req* with prob. $1/n$ and unif. prob. on safe actions.

a is safe from S if $\forall O$, $up(S, a, O) \notin Lose$

Lemma

σ_n is almost-surely winning from Win.

Optimizing the worst-case cost

Iterative computation of S_k: set of beliefs where k req are sufficient.

$$S_0 \subseteq S_1 \subseteq S_2 \cdots \subseteq Win$$

Optimizing the worst-case cost

Iterative computation of S_k: set of beliefs where k req are sufficient.

$$S_0 \subseteq S_1 \subseteq S_2 \cdots \subseteq Win$$

Optimizing the worst-case cost

Iterative computation of S_k: set of beliefs where k req are sufficient.

$$S_0 \subseteq S_1 \subseteq S_2 \cdots \subseteq Win$$

Computation of S_0: $S_0 = \text{reach}_{=1}(W_{ok})$

almost-sure reachability question for the belief-MDP without requests

Optimized strategy: no request from $S \in S_0$

Optimizing the worst-case cost

Iterative computation of S_k: set of beliefs where k req are sufficient.

$$S_0 \subseteq S_1 \subseteq S_2 \cdots \subseteq Win$$

Computation of S_1

- $L_1 = \{S \mid \forall s \in S, \{s\} \in S_0\}$
- $S_1 = \text{reach}_{=1}(L_1 \cup S_0)$

Optimized Strategy:

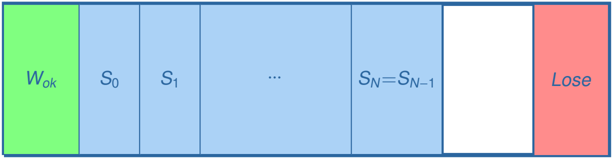
request from $S \in L_1 \setminus S_0$

uniform distribution on actions ensuring to stay in S_1, othw

Optimizing the worst-case cost

Iterative computation of S_k: set of beliefs where k req are sufficient.

$$S_0 \subseteq S_1 \subseteq S_2 \cdots \subseteq Win$$

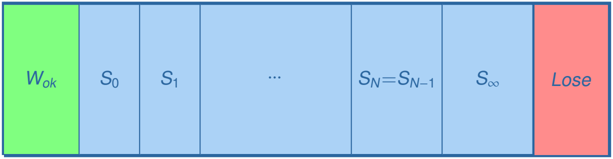


Stabilisation for $N \leq |\mathcal{B}|$

Optimizing the worst-case cost

Iterative computation of S_k: set of beliefs where k req are sufficient.

$$S_0 \subseteq S_1 \subseteq S_2 \cdots \subseteq Win$$



Stabilisation for $N \leq |\mathcal{B}|$

$S_\infty = Win \setminus S_N$

Optimizing the worst-case cost

Iterative computation of S_k: set of beliefs where k req are sufficient.

$$S_0 \subseteq S_1 \subseteq S_2 \cdots \subseteq Win$$

W_{ok}	S_0	S_1	...	$S_N = S_{N-1}$	S_∞	*Lose*

Proposition

The minimum worst-case cost can be computed in EXPTIME, together with a finite-memory strategy.

Outline

Undecidability

Value: infimum of average cost over almost-surely winning strategies

$$\textit{val}(G) = \textit{inf}\{\textit{av\_cost}(\sigma) \mid \sigma \text{ almost-surely winning}\}$$

Undecidability

Value: infimum of average cost over almost-surely winning strategies

$$val(G) = inf\{av\_cost(\sigma) \mid \sigma \text{ almost-surely winning}\}$$

The value cannot be computed

For all $K > 0$, one cannot decide whether $val(G) \leq K$.

Undecidability

Value: infimum of average cost over almost-surely winning strategies

$$val(G) = inf\{av\_cost(\sigma) \mid \sigma \text{ almost-surely winning}\}$$

The value cannot be computed

For all $K > 0$, one cannot decide whether $val(G) \leq K$.

Not too surprising: optimizing cost functions for POMDP is undecidable

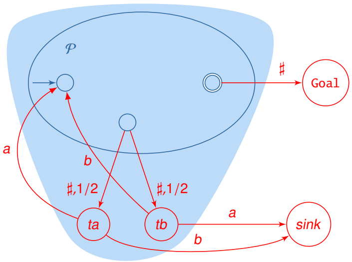
Skip proof

Proof idea

$\mathcal{P}$ PFA s.t. either all words have probability $\leq \varepsilon$, or some word has probability $> 1 - \varepsilon$. Which holds is undecidable. [Madani Hanks Condon 03]

Proof idea

$\mathcal{P}$ PFA s.t. either all words have probability $\leq \varepsilon$, or some word has probability $> 1 - \varepsilon$. Which holds is undecidable. [Madani Hanks Condon 03]



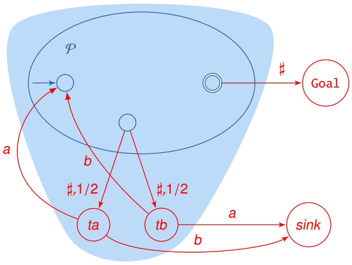
$\mathcal{P}$ accepts a word with probability greater than $1 - \varepsilon$

iff

$val(G) < \frac{\varepsilon}{1-\varepsilon}$

Proof idea

$\mathcal{P}$ PFA s.t. either all words have probability $\leq \varepsilon$, or some word has probability $> 1 - \varepsilon$. Which holds is undecidable. [Madani Hanks Condon 03]



$\mathcal{P}$ accepts a word with probability greater than $1 - \varepsilon$

iff

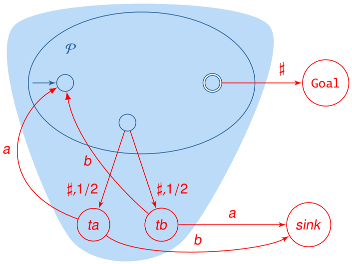
$val(G) < \frac{\varepsilon}{1-\varepsilon}$

($\Rightarrow$) σ plays $(w \sharp\ req\ a|b)^*$ for w with $\mathbb{P}(w) > 1 - \varepsilon$

$val(\sigma) < 0 \times (1-\varepsilon) + 1 \times \varepsilon(1-\varepsilon) + 2 \times \varepsilon^2(1-\varepsilon) \cdots = \varepsilon/(1-\varepsilon)$

Proof idea

$\mathcal{P}$ PFA s.t. either all words have probability $\leq \varepsilon$, or some word has probability $> 1 - \varepsilon$. Which holds is undecidable. [Madani Hanks Condon 03]



$\mathcal{P}$ accepts a word with probability greater than $1 - \varepsilon$

iff

$val(G) < \frac{\varepsilon}{1-\varepsilon}$

($\Rightarrow$) σ plays $(w \sharp\, req\ a|b)^*$ for w with $\mathbb{P}(w) > 1 - \varepsilon$

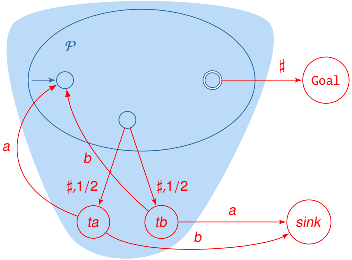
$val(\sigma) < 0 \times (1-\varepsilon) + 1 \times \varepsilon(1-\varepsilon) + 2 \times \varepsilon^2(1-\varepsilon) \cdots = \varepsilon/(1-\varepsilon)$

($\Leftarrow$) p probability in σ to have $\sharp$ before req

$val(\sigma) > (1-p) + p(1-\varepsilon) \geq 1 - \varepsilon$

Proof idea

$\mathcal{P}$ PFA s.t. either all words have probability $\leq \varepsilon$, or some word has probability $> 1 - \varepsilon$. Which holds is undecidable. [Madani Hanks Condon 03]



$\mathcal{P}$ accepts a word with probability greater than $1 - \varepsilon$

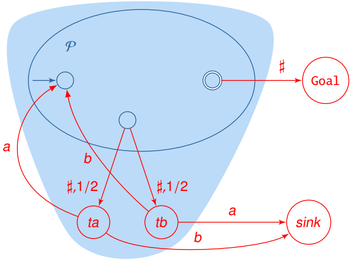
iff

$val(G) < \frac{\varepsilon}{1-\varepsilon}$

best approximation: $|v - val(G)| = \big(\varepsilon/(1-\varepsilon) + (1-\varepsilon)\big)/2$

Proof idea

$\mathcal{P}$ PFA s.t. either all words have probability $\leq \varepsilon$, or some word has probability $> 1 - \varepsilon$. Which holds is undecidable. [Madani Hanks Condon 03]



$\mathcal{P}$ accepts a word with probability greater than $1 - \varepsilon$

iff

$$val(G) < \frac{\varepsilon}{1-\varepsilon}$$

best approximation: $|v - val(G)| = \big(\varepsilon/(1-\varepsilon) + (1-\varepsilon)\big)/2$

approximation factor: $\frac{|v-val(G)|}{val(G)} = \frac{(1-\varepsilon)(1/(1-\varepsilon)-\varepsilon)}{2\varepsilon} \xrightarrow{\varepsilon \to 0} \infty$

Non-approximability

Corollary: For every δ it is undecidable to approximate $val(G)$ within δ.

NB: bigger δ need bigger POMDP

Non-approximability

Corollary: For every δ it is undecidable to approximate $val(G)$ within δ.

NB: bigger δ need bigger POMDP

NP-hardness of good approximations

Assuming $P \neq NP$ there is a POMDP G with
few reachable belief states (quadratic in n) s.t.
any polynomial time algorithm $\mathcal{A}$ returns for G a value v with

approximation factor: $\frac{|v-val(G)|}{val(G)} \geq 2^{n-1}/n^2$, and

absolute approximation error: $|v - val(G)| \geq 2^{n-1}/n$.

Proof idea

φ 3-SAT instance with m clauses and k variables; $n = mk$

φ is satisfiable if for each clause C_i, one can choose a literal l_i and the choices do not conflict

POMDP behaviour:

- random choice of variable to monitor
- conflicts force a request not to lose

Proof idea

φ 3-SAT instance with m clauses and k variables; $n = mk$

φ is satisfiable if for each clause C_i, one can choose a literal l_i and the choices do not conflict

POMDP behaviour:

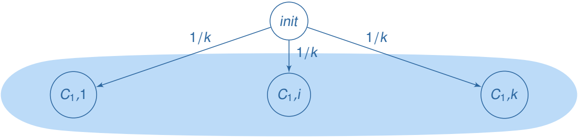
- random choice of variable to monitor
- conflicts force a request not to lose

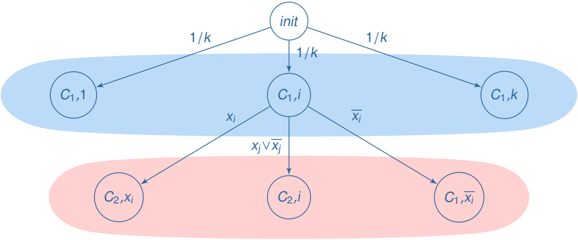
Properties of the reduction

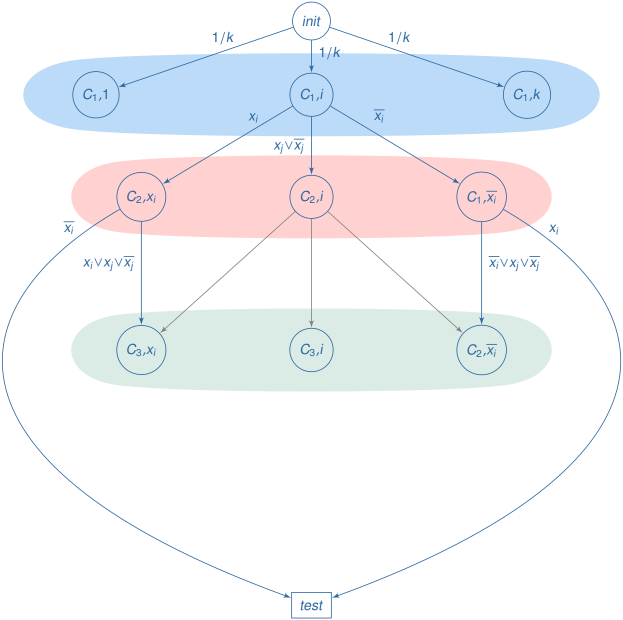
- φ satisfiable $\Rightarrow val(G) < n$
- φ not satisfiable $\Rightarrow val(G) > 2^n/n - 2$

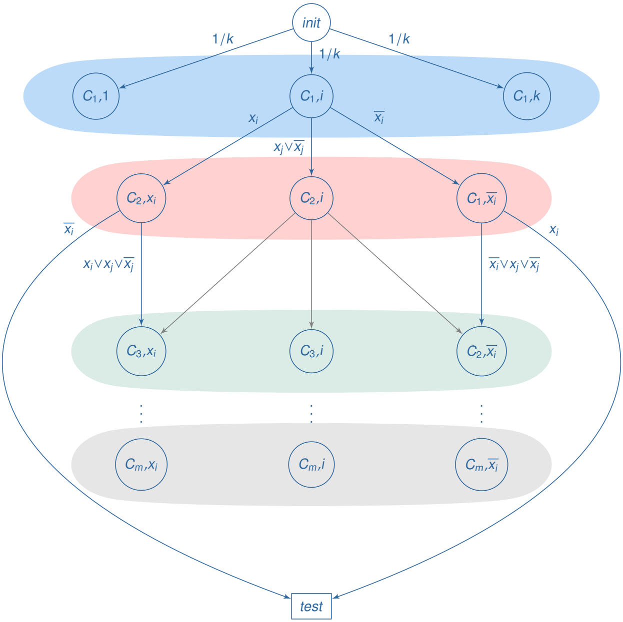
Skip details

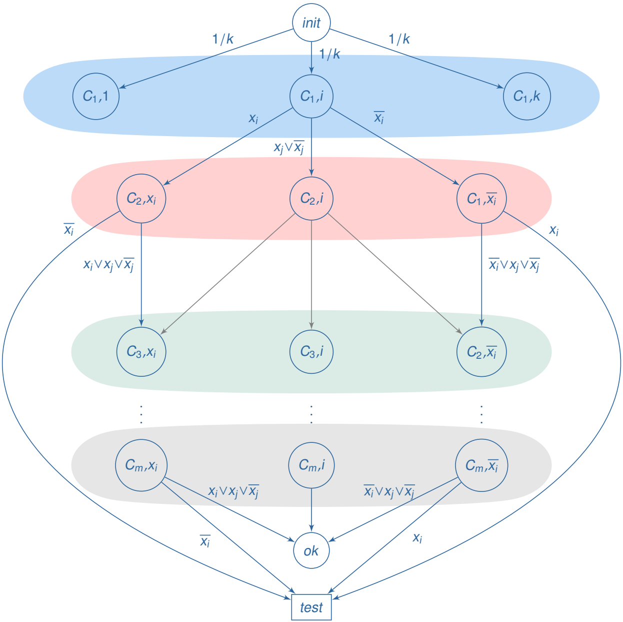

init


init
1/k
1/k
1/k
C1,1
C1,i
C1,k

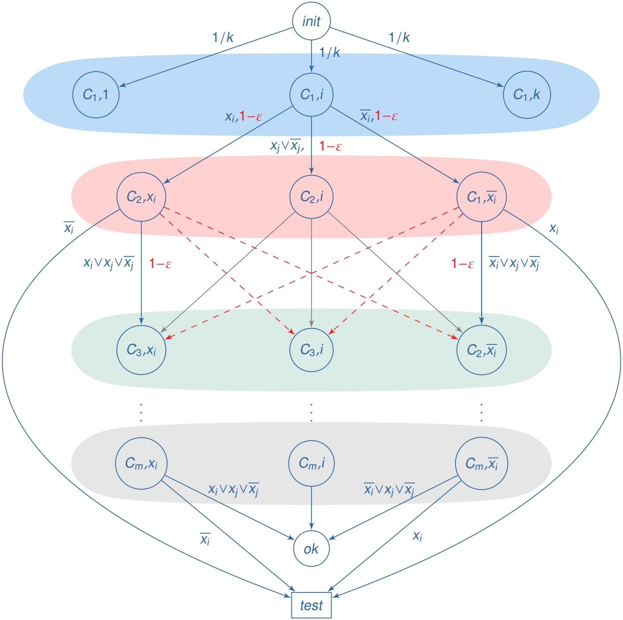

init
1/k
1/k
1/k
C1,1
C1,i
C1,k
xi
xj∨x̄j
x̄i
C2,xi
C2,i
C1,x̄i


init
1/k
1/k
1/k
C_1,1
C_1,i
C_1,k
x_i
$\overline{x_i}$
$x_j \vee \overline{x_j}$
C_2,x_i
C_2,i
C_1,$\overline{x_i}$
$\overline{x_i}$
x_i
$x_i \vee x_j \vee \overline{x_j}$
$\overline{x_i} \vee x_j \vee \overline{x_j}$
C_3,x_i
C_3,i
C_2,$\overline{x_i}$
test


init
1/k
1/k
1/k
$C_1,1$
C_1,i
C_1,k
x_i
$\overline{x_i}$
$x_j \vee \overline{x_j}$
C_2,x_i
C_2,i
$C_1,\overline{x_i}$
$\overline{x_i}$
x_i
$x_i \vee x_j \vee \overline{x_j}$
$\overline{x_i} \vee x_j \vee \overline{x_j}$
C_3,x_i
C_3,i
$C_2,\overline{x_i}$
C_m,x_i
C_m,i
$C_m,\overline{x_i}$
test



init
1/k
1/k
1/k
$C_1,1$
C_1,i
C_1,k
x_i
$\overline{x_i}$
$x_j \vee \overline{x_j}$
C_2,x_i
C_2,i
$C_1,\overline{x_i}$
$\overline{x_i}$
x_i
$x_i \vee x_j \vee \overline{x_j}$
$\overline{x_i} \vee x_j \vee \overline{x_j}$
C_3,x_i
C_3,i
$C_2,\overline{x_i}$
C_m,x_i
C_m,i
$C_m,\overline{x_i}$
$x_i \vee x_j \vee \overline{x_j}$
$\overline{x_i} \vee x_j \vee \overline{x_j}$
$\overline{x_i}$
x_i
ok
test


init
1/k
1/k
1/k
C1,1
C1,i
C1,k
xi,1−ε
x̄i,1−ε
xj∨x̄j, 1−ε
C2,xi
C2,i
C1,x̄i
x̄i
xi
xi∨xj∨x̄j 1−ε
1−ε x̄i∨xj∨x̄j
C3,xi
C3,i
C2,x̄i
Cm,xi
Cm,i
Cm,x̄i
xi∨xj∨x̄j
x̄i∨xj∨x̄j
x̄i
xi
ok
test

Outline

Conclusion

Contribution

Minimize requests for full-information in POMDP
under an almost-sure reachability objective.

- Worst-case cost
  - computation in EXPTIME, together with an optimal strategy
- Average cost
  - computation undecidable
  - approximation unfeasible
  - large least approximation factors for polytime algorithms

Conclusion

Contribution

Minimize requests for full-information in POMDP
under an almost-sure reachability objective.

- Worst-case cost
  - computation in EXPTIME, together with an optimal strategy
- Average cost
  - computation undecidable
  - approximation unfeasible
  - large least approximation factors for polytime algorithms

Future work

- extend to several information levels
  - succesive partition refinement
- tradeoff between objective (reachability probability) and cost