Optimal Control of MDPs with Temporal Logic Constraints

Mária Svoreňová, Ivana Černá, Calin Belta

To be presented on The 52nd IEEE Conference on Decision and Control

December 10 – 13, Florence, Italy

Control Theory vs. Theoretical Computer Science

control theory (CT) = engineering and mathematics

- control of dynamic (continuous/discrete) systems with inputs

theoretical computer science (TCS) = CS and mathematics

- emphasis on mathematical technique and rigor

this work

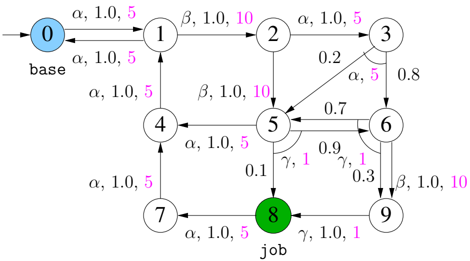
- belongs in CT (process control, optimal control, automatic control)
- applies results from TCS (verification, automata, game theory)

provide an optimal solution to a control problem that has only been solved sub-optimally using control approaches

Markov decision process (MDP)

- labeled states
- actions
- probabilistic transitions
- costs of applying an action in a state

example



$$\mathbf{GF}\bigcirc \wedge \mathbf{GF}\bullet \wedge \mathbf{G}(\bigcirc \Rightarrow \mathbf{X}(\neg \bigcirc \,\mathbf{U}\, \bullet))$$

Linear Temporal Logic (LTL)

- properties of runs
- formulae with persistent surveillance (= LTL):

$$\varphi \wedge \mathbf{GF}\pi_{\mathsf{sur}}$$

Problem Formulation

strategy – a function mapping finite runs onto actions
– induces a (possibly infinite) Markov chain

given

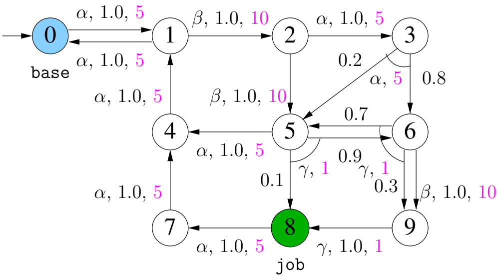
- an MDP
- an LTL formula with surveillance

find a strategy that

- guarantees satisfaction of the formula with probability 1
- provides the minimum average expected cumulative cost between consecutive visits of a surveillance state (ACPC, i.e. average cost per cycle) among all strategies satisfying the formula

(combination of correctness and optimization)

Ding, Smith, Belta, Rus: *MDP Optimal Control under Temporal Logic Constraints.* CDC-ECC, 2011.



strategy that

- satisfies the mission
- at the same time, minimizes the average cost between jobs

$$\mathbf{GF}\,{\color{cyan}\bullet} \wedge \mathbf{GF}\,{\color{green}\bullet} \wedge \mathbf{G}({\color{cyan}\bullet} \Rightarrow \mathbf{X}(\neg {\color{cyan}\bullet}\ \mathbf{U}\ {\color{green}\bullet}))$$

the problem as a game:

MDP	$\Rightarrow$	$1^{1}/_{2}$-player game
LTL constraint	$\Rightarrow$	parity winning condition
ACPC optimization	$\Rightarrow$	generalized mean-payoff winning condition

Chatterjee, Doyen: *Energy and Mean-Payoff Parity Markov Decision Processes.* MFCS, 2011.

correctness + optimization

$\varphi \wedge \mathbf{GF}\pi_{\mathsf{sur}}$ ACPC

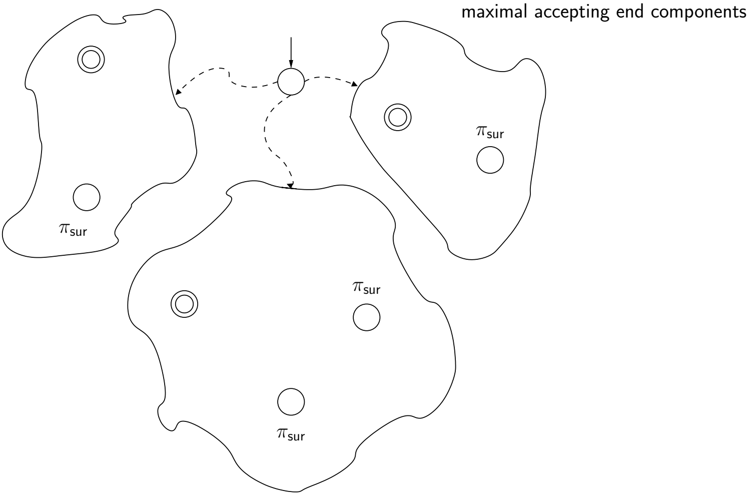
LTL model checking
and control theory

control theory

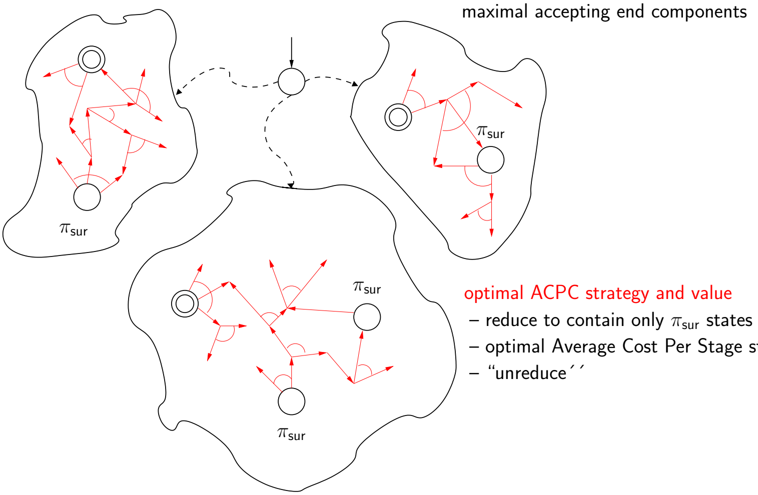
combine resulting strategies
using game techniques

Chatterjee, Doyen: *Energy and Mean-Payoff Parity Markov Decision Processes.* MFCS, 2011.

MDP × automaton for the formula



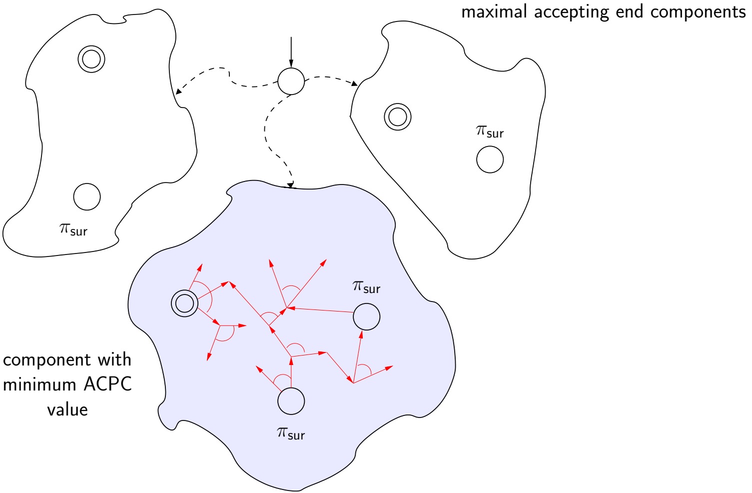
MDP $\times$ automaton for the formula



optimal ACPC strategy and value

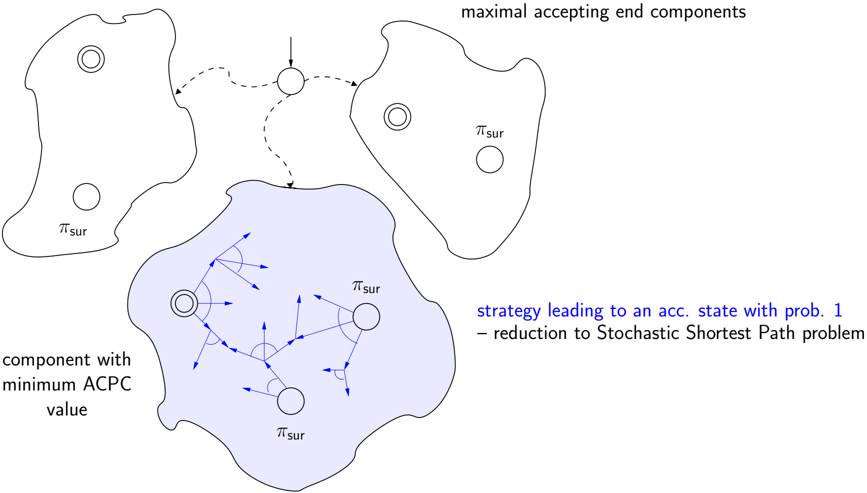
- reduce to contain only π_{sur} states (Kleene-like)
- optimal Average Cost Per Stage strategy and value
- "unreduce´´

MDP × automaton for the formula



Synthesis Algorithm

MDP × automaton for the formula



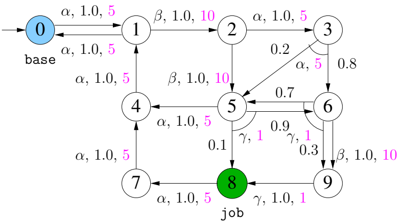
Synthesized Strategy

strategy for the product:

1. reach the chosen component
2. rounds:
   - phase 1: reach an accepting state
   - phase 2: optimize ACPC value (predefined number of visits of surv. states for each round, to ensure convergence)

project onto MDP $\Longrightarrow$ **optimal solution to the original problem**

Example – Solution and Related Work



$\mathbf{GF}$ ○ $\wedge$ $\mathbf{GF}$ ● $\wedge$ $\mathbf{G}($ ○ $\Rightarrow \mathbf{X}(\neg$ ○ $\mathbf{U}$ ● $))$

	Condition	0	1	2	3	4	5	6	7	8	9
C_{init}		α	–	–	–	–	–	–	–	–	–
C_{p1}	before job	α	β	α	α	α	γ	γ	α	α	γ
	after job	α	α	α	α	α	γ	γ	α	α	γ
C_{p2}		α	β	α	α	α	γ	γ	α	α	γ

Ding, Smith, Belta, Rus: *MDP Optimal Control under Temporal Logic Constraints.* CDC-ECC, 2011.