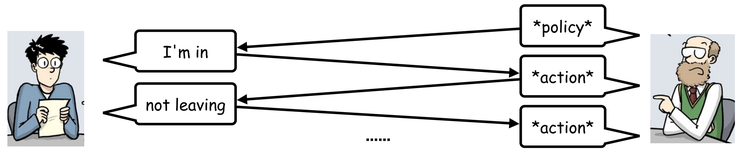
Planning with Participation Constraints

Hanrui Zhang **Yu Cheng** **Vincent Conitzer**

CMU **UIC -> Brown** **Duke -> CMU**

Problem setup

- computing optimal policies in MDPs **subject to participation constraints**
  - principal's policy: chooses action based on history and state
  - agent: decides whether to **leave** in response to principal's policy
  - problem: **maximize utility** while **keeping agent from leaving**



Applications

- service providers trying to keep customers
  - cable networks, banks, internet streaming companies, phone carriers
  - common strategy: discount in first several months + early termination fee; increase to normal rate when user becomes attached to service
  - key difficulty: user may **leave at any time** during service

xfinity CHASE NETFLIX AT&T

Our results

- polynomial-time algorithm for computing the exact optimal policy in finite-horizon environments
- approximation algorithm that is additively suboptimal by ε in polynomial time in $\log(1/\varepsilon)$ in infinite-horizon discounted environments

Why is the problem hard?

planning in classical MDPs

- wlog optimal policies are deterministic and history-independent
- computing optimal policies: simple backward induction procedure (and many other efficient algorithms)

planning with participation constraints

- optimal policies need to be randomized and history-dependent
- computing optimal policies: ???
  - "flat" representation of optimal policy is exponentially large

Overview of algorithms

- high-level strategy: evaluate Pareto-frontier curves, find points that encode optimal policy
- Pareto frontier curve f_s: maps each x to principal's maximum onward utility $f_s(x)$ when agent's onward utility is precisely x, subject to participation constraints in the future
  - value of optimal policy = $\max_{x \geq 0} f_{s_{\text{init}}}(x)$, where $s_{\text{init}} = 1$ is the initial state
- key difficulty: curves are huge
  - must schedule recursive evaluations very carefully

Related results

- what if agent has private information and can misreport?
  - in other words: principal's policy must also be incentive-compatible
- related paper: Automated Dynamic Mechanism Design (on arXiv!)
  - long horizon: problem is APX-hard, short horizon: LP-based polynomial-time algorithm