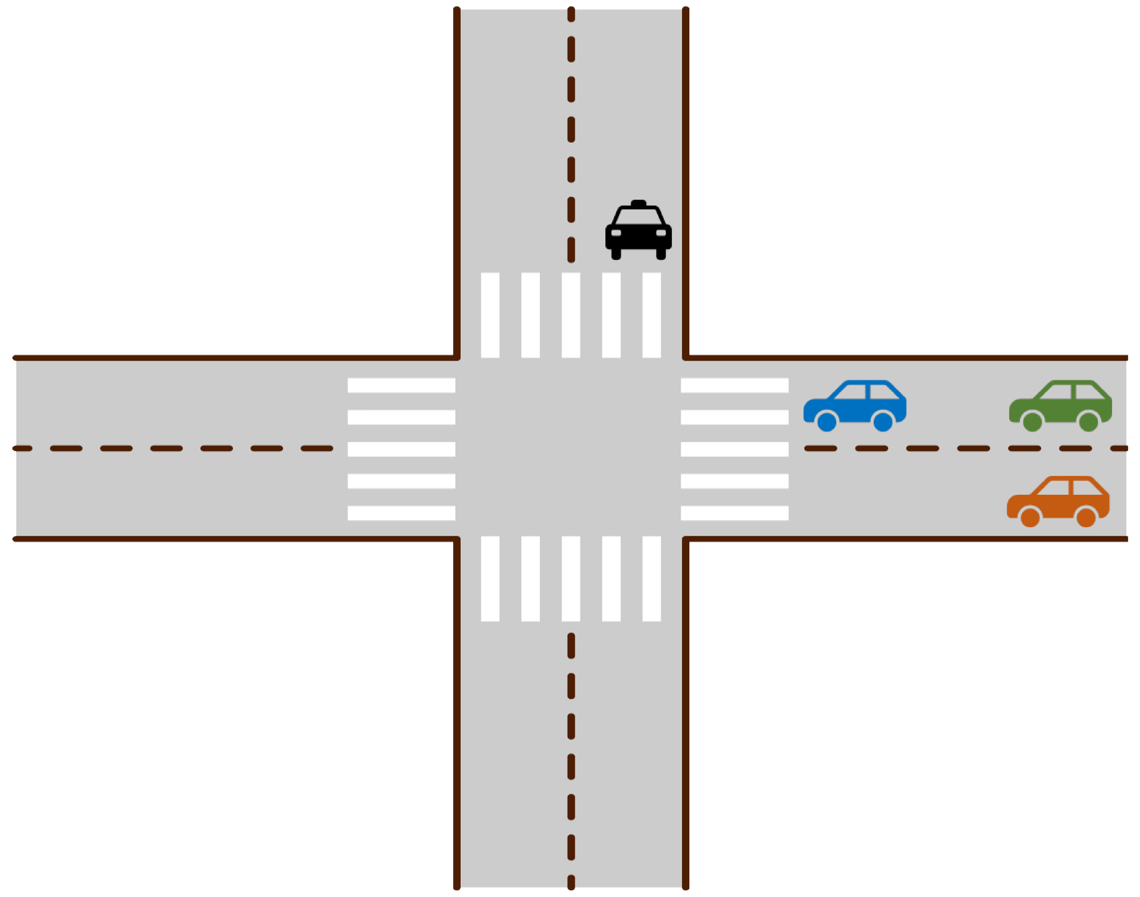
Specification-Guided Learning of Nash Equilibria with High Social Welfare

Kishor Jothimurugan, Suguman Bansal, Osbert Bastani and Rajeev Alur

Penn University of Pennsylvania



Multi-agent System

- n agents and finite state space S.
- Action space A_i for agent i.
- Transition probability $P(s' \mid s, a)$ for $s, s' \in S$ and $a \in \prod_i A_i$.

User Input

- A specification ϕ_i for each agent i.
- A method to sample from $P(\cdot \mid s, a)$.

Problem Statement

Given a joint policy π, score of agent i is $\mathrm{J_i}(\pi) = \Pr_{\zeta \sim \mathrm{D}(\pi)}[\zeta \vDash \phi_i]$. Solve

$$\arg\max_{\pi} \sum_i J_i(\pi)$$

s.t. π is ϵ-Nash equilibrium

Our Framework

Phase I: Prioritized Enumeration

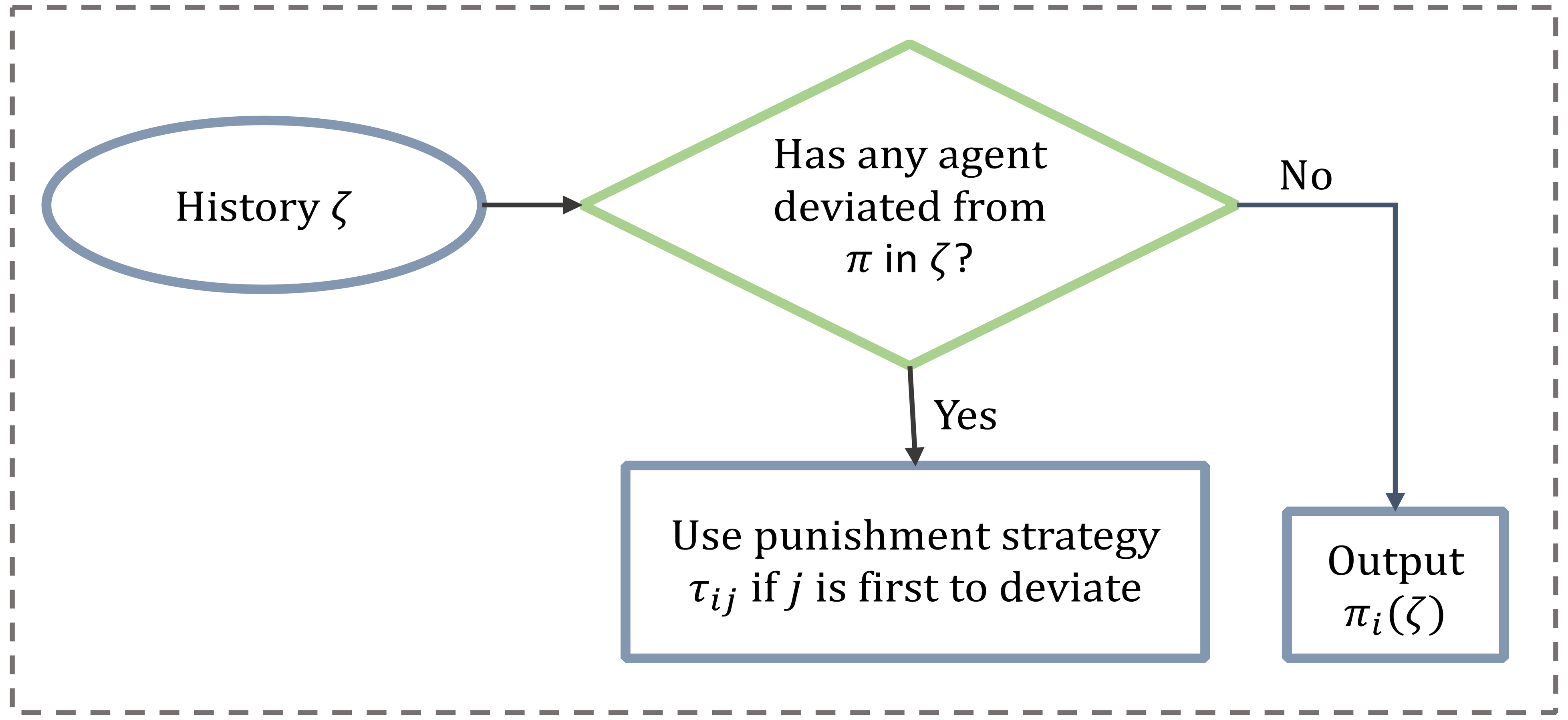
Use the specification of each agent to enumerate ***finite-state deterministic*** joint policies.

- Uses the specifications to construct multiple ***abstract graphs*** whose edges denote joint subtasks.
- Uses ***single-agent RL*** to learn joint policies for edges in the abstract graphs.
- Each ***path*** in the abstract graphs corresponds to a finite state joint policy (applies the edge policies in order).
- The policies are ordered in decreasing value of social welfare.

Phase II: Nash Verification

Checks if a joint policy can be ***modified*** to get an ϵ-Nash equilibrium ***without affecting social welfare***.

- Modifications are restricted to ***adding punishment strategies*** which trigger when some agent deviates.
- Uses a ***self-play RL algorithm*** to compute the best punishment strategies.
- Return the first joint policy (from Phase I) which can be converted to ϵ-Nash equilibrium this way.



Experiments on Intersection Environment

Spec.	Num. of agents	Algorithm	welfare(π) (avg ± std)	$\epsilon_{\min}(\pi)$ (avg ± std)	Num. of runs terminated	Avg. num. of sample steps (in millions)
ϕ^1	3	HighNash	**0.33 ± 0.00**	**0.00 ± 0.00**	10	1.78
		NVI	0.32 ± 0.00	**0.00 ± 0.00**	10	1.92
		MAQRM	0.18 ± 0.01	0.51 ± 0.01	10	2.00
ϕ^2	4	HighNash	**0.55 ± 0.10**	**0.01 ± 0.02**	10	11.53
		NVI	0.04 ± 0.01	0.02 ± 0.01	10	12.60
		MAQRM	0.12 ± 0.01	0.20 ± 0.03	10	15.00
ϕ^3	4	HighNash	**0.49 ± 0.01**	**0.00 ± 0.01**	10	11.26
		NVI	0.45 ± 0.01	**0.00 ± 0.01**	10	12.60
		MAQRM	0.11 ± 0.01	0.22 ± 0.02	10	15.00
ϕ^4	3	HighNash	0.90 ± 0.15	**0.00 ± 0.00**	10	2.16
		NVI	**0.98 ± 0.00**	**0.00 ± 0.00**	4	2.18
		MAQRM	0.23 ± 0.01	0.39 ± 0.04	10	2.00
ϕ^5	5	HighNash	**0.58 ± 0.02**	**0.00 ± 0.00**	10	62.17
		NVI	0.05 ± 0.01	0.01 ± 0.01	7	80.64
		MAQRM	Timeout	Timeout	0	Timeout

Code

Paper