Resolving Conflicting Constraints in Multi-Agent Reinforcement Learning with Layered Safety

Jason J. Choi*, Jasmine Jerry Aloor*, Jingqi Li*, Maria G. Mendoza, Hamsa Balakrishnan†, Claire J. Tomlin†

Webpage, Paper, Videos

Takeaways

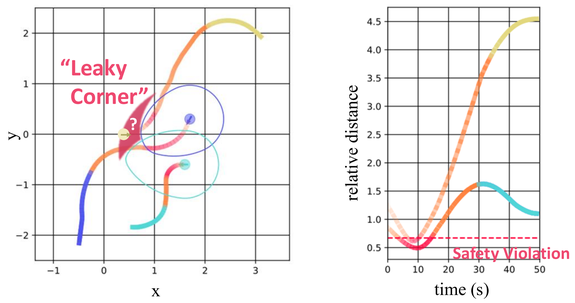
We proposed a ***layered architecture*** incorporating ***safety-informed reward design for MARL*** and ***CBVF-based safety filters*** during training via curriculum learning, enabling proactive conflict resolution by avoiding potential conflict zones and improved coordination efficiency.

Conflicting Constraints

Main challenge in composing safe sets in multi-agent interaction:

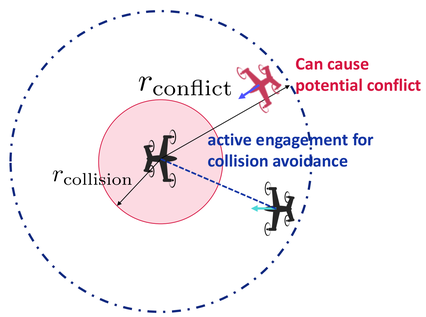
Leaky Corners arising from Conflicting Constraints

$$S(C_1) \cap S(C_2) \neq S(C_1 \cap C_2)$$



Potential Conflict Zone:

If **N>2** are within the potential conflict zone, **safety of all agents are not guaranteed anymore.**



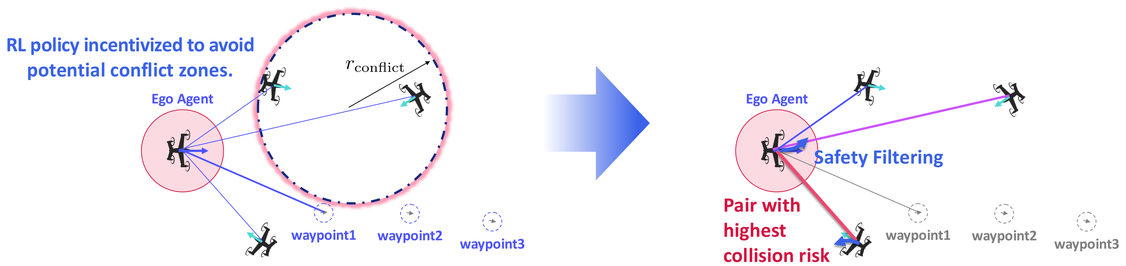
Layered Safety

1. Potential Conflict Zone as Soft Constraint for MARL

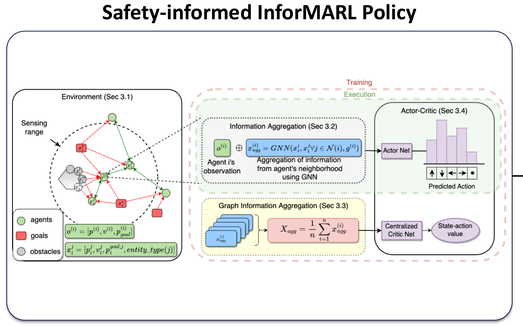
Based on local observations, GNN-based MARL Policy learns navigation **while incentivized to avoid potential conflict zones.**

2. Control Barrier-Value Function (CBVF)-based Safety Filtering

Active safety filtering against agent with worst CBVF value, based on the pairwise relative dynamics.

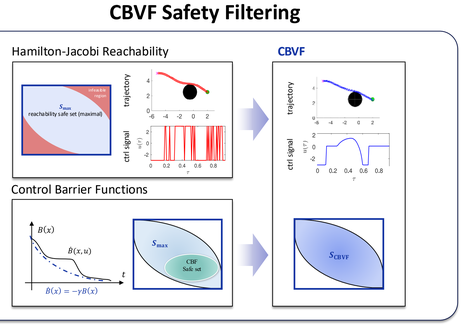


Safety-informed InforMARL Policy



Prioritization

CBVF Safety Filtering



Safety Informed Training

Curriculum learning

1. First train without any safety filter or reward penalty
2. Activate safety filter
3. Gradually increase safety radius and conflict radius

Safety-informed reward

Penalty for entering potential conflict zone:

$$\mathcal{C}_{\text{conflict}} := \sum_{j \in \{j \mid \text{dist}(s^{(ij)}) < r_{\text{conflict}}\}} \max\{0, r_{\text{conflict}} - \text{dist}(s^{(ij)})\} \times \max\left\{0, -\underbrace{\begin{bmatrix} x^{(ij)} & y^{(ij)} \end{bmatrix} \begin{bmatrix} v_x^{(ij)} \\ v_y^{(ij)} \end{bmatrix}}_{\text{relative distance change}}\right\},$$

Combined Reward:

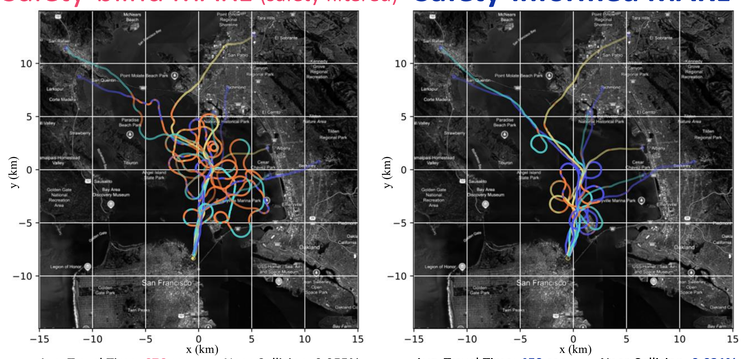
$$\mathcal{R}_{\text{total}}(o_k^{(i)}, a_k^{(i)}) = \mathcal{R}_{\text{tracking}}(o_k^{(i)}, a_k^{(i)}) + \rho_{\text{goal}}\mathcal{R}_{\text{goal}}(o_k^{(i)}, a_k^{(i)}) - \rho_{\text{conflict}}\mathcal{C}_{\text{conflict}}$$

Highlighted Results

1. Air Taxi Operation without Centralized Traffic Control

8 vehicles departing from various locations in Bay Area, merging to land in San Francisco

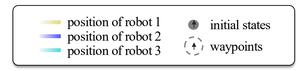
Safety-blind MARL (safety filtered) **Safety-informed MARL**



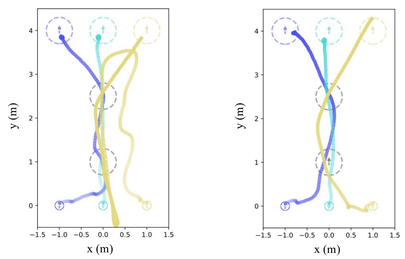
Avg. Travel Time: 676 sec. Near Collision: 0.055% **Avg. Travel Time: 450 sec. Near Collision: 0.021%**

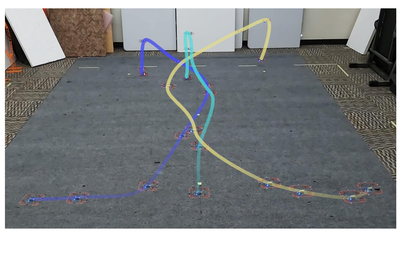
2. Hardware experiments with quadrotors

Three drones going through the same air corridor to get to their landing spots.



Safety-blind MARL **Safety-informed MARL**





3. An extensive empirical study of different safety-informing methods in MARL

Our method generalizes better than prior works in terms of number of agents. (N =number of agents, M =number of waypoints)

Safety-informed training vs no safety filter or safety-informed reward **(Airtaxi scenarios)**

Methods	Merging Scenario (N=8, M=5)		
	Travel t(s)(↓)	Near collision(%)(↓)	Conflict(%)(↓)
Safety-blind	675.6	0.055	2.4
No penalty	617.9	0.042	5.5
Proposed	**450.5**	**0.021**	**3.2**

Methods	Intersection Scenario (N=16, M=6)		
	Travel t(s)(↓)	Near collision(%)(↓)	Conflict(%)(↓)
Safety blind	987.4	0.058	2.1
No penalty	780.5	0.129	3.8
Proposed	**660.8**	**0.056**	**1.6**

Comparison to other model-based and model-free baselines: **(Crazyflie scenarios)**

N=4:

Methods	Goal reach(%)	Near collision(%)
DG-PPO	96 ± 11.8	0.04 ± 0.16
Exponential CBF	100 ± 0	0.0 ± 0.0
Our Method	100 ± 0	0.0 ± 0.0

N=8:

Methods	Goal reach(%)	Near collision(%)
DG-PPO	100 ± 0	9.1 ± 2.7
Exponential CBF	93 ± 8.9	8.8 ± 10.7
Our Method	**100 ± 0**	**0.0 ± 0.0**

References

- **Leaky Corners:** I. Mitchell. "Scalable calculation of reach sets and tubes for nonlinear systems with terminal integrators." HSCC 2011.
- **InforMARL:** S. Nayak, et al. "Scalable multi-agent reinforcement learning through intelligent information aggregation." ICML 2023.
- **CBVF:** J. Choi, et al. "Robust control barrier–value functions for safety-critical control.", CDC 2021.
- **DG-PPO:** S. Zhang, et al. "Discrete GCBF Proximal Policy Optimization for Multi-agent Safe Optimal Control." ICLR 2025.
- **Exponential CBF:** A. Agrawal and K.Sreenath. "Discrete CBFs for safety-critical control of discrete systems with application to bipedal robot navigation." RSS 2017.

Acknowledgement: This work is supported in part by the NASA ULI in Safe Aviation Autonomy, NASA Clean Sheet Airspace Operating Design project, DARPA Assured Autonomy, DARPA ANSR, and ONR Basic Research Challenge in Multibody Control Systems program.