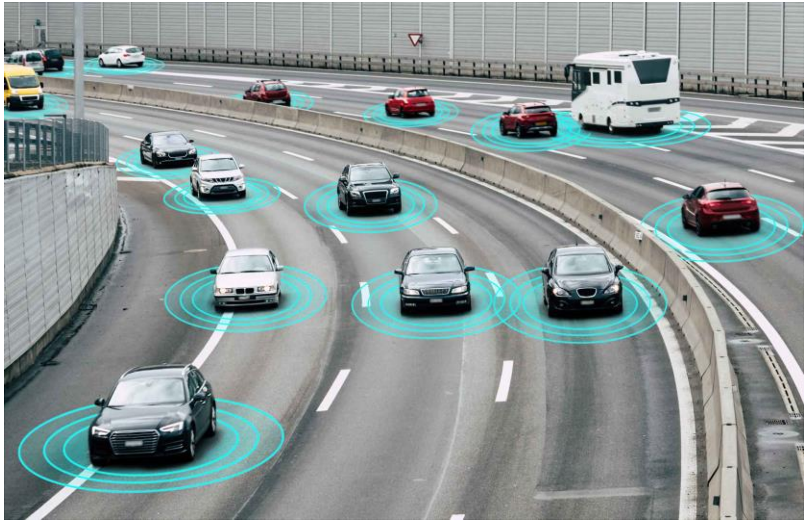
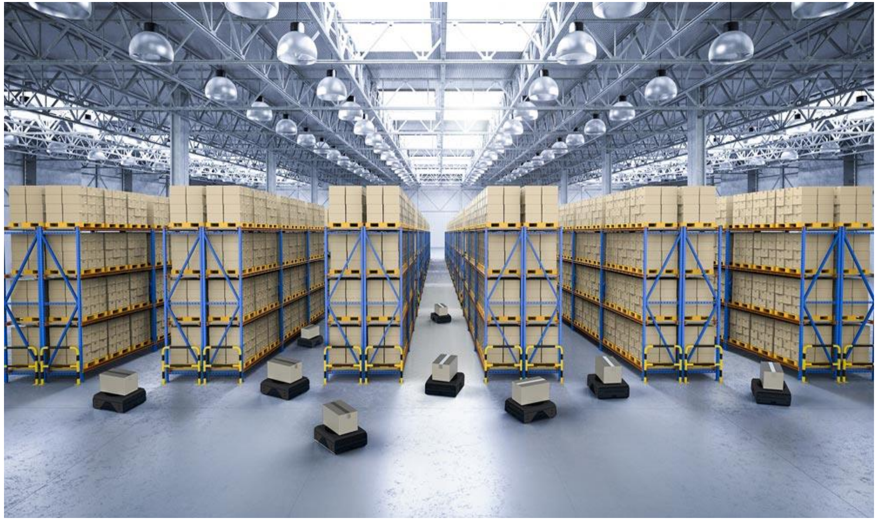
A Safety Guaranteed Hierarchical Multi-Agent Reinforcement Learning Approach Based on Control Barrier Functions for Safety-Critical Systems

Authors: H M Sabbir Ahmad, Ehsan Sabouni, Alexander Wasilkoff, Param Budhraja, Zijian Guo, Songyuan Zhang, Chuchu Fan, Christos G. Cassandras, Wenchao Li
Date: 11/06/2025

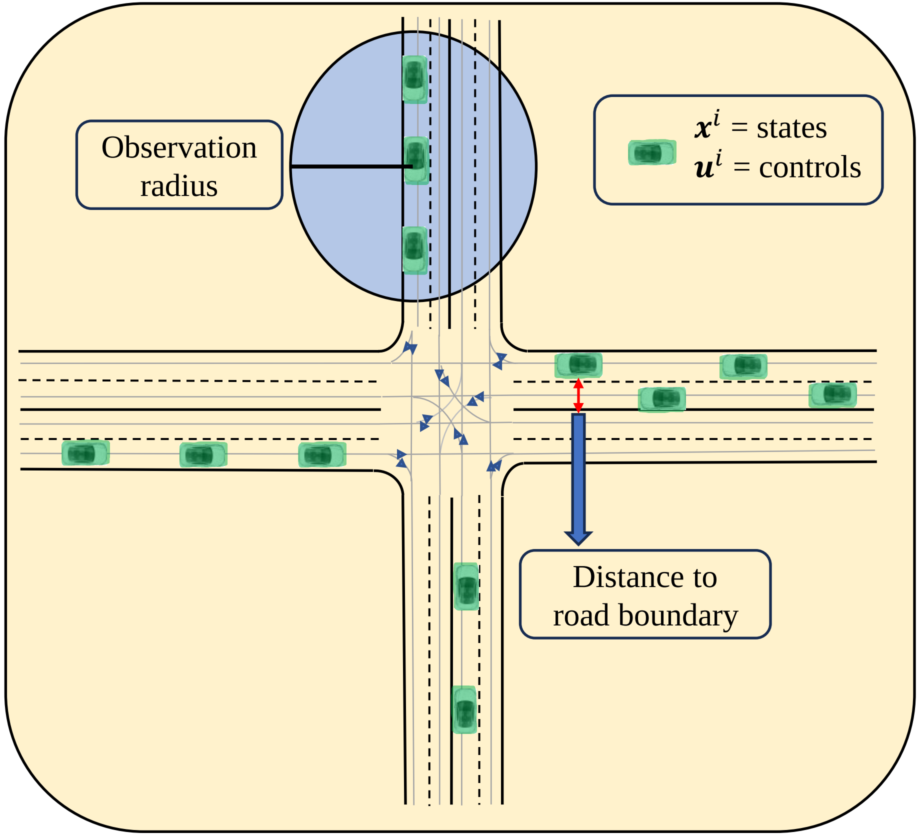
Motivation





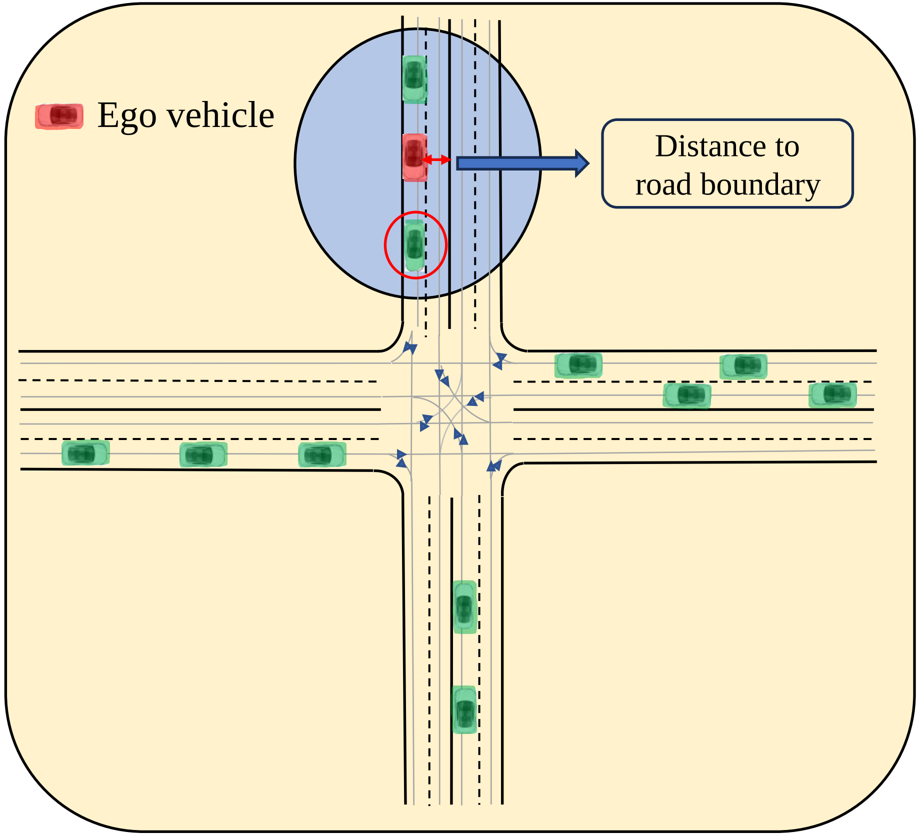
- Safety-critical autonomous systems are growing with advances in machine learning, computing, communication and sensing technologies.
- Examples: robotic systems e.g. ground robots, unmanned aerial vehicles (UAVs), autonomous underwater vehicles (AUVs), and self-driving cars to name a few.
- Safety – critical systems: i. inter agent safety and ii. environment safety.

Problem Formulation - I



- A fully cooperative multi-agent safety-critical system setting.
- A finite but arbitrary number of agents.
- Agent dynamics: $\boldsymbol{x}_{t+1}^i = f\left(\boldsymbol{x}_t^i, \boldsymbol{u}_t^i\right), \boldsymbol{x}^i \in \mathcal{X}^i$ and $\boldsymbol{u}^i \in \mathcal{U}^i$ are states and controls.
- The environment state is denoted by $\boldsymbol{x}^e \in \mathcal{X}^e$.
- Stage cost agent of agent i is denoted as $l^i\left(\boldsymbol{x}^i, \boldsymbol{u}^i\right)$ and the total stage cost is: $l(\boldsymbol{x}, \boldsymbol{u}) = \sum_{i=1,\ldots,\mathcal{N}} l^i\left(\boldsymbol{x}^i, \boldsymbol{u}^i\right)$.
- Partially observability: $\boldsymbol{o}^i \in \mathcal{O}^i \subseteq \mathcal{X}$.

Problem Formulation - II



- Agent specific safety functions $\{b_j^i(\boldsymbol{o}^i)\}, j \in \mathcal{N}^i(\boldsymbol{o}^i)$, $\mathcal{N}^i(\boldsymbol{o}^i)$ is a finite set.

- Thus, our safety-constrained cooperative game is expressed as follows:

$$\min_{\pi_1,\dots\pi_N} \mathbb{E}_{\tau\sim(\pi_1,\dots,\pi_N)}\left[\sum_{t=0}^{\infty} \gamma^t l(\boldsymbol{x}_t, \boldsymbol{u}_t)\right]$$

$$\text{Subject to:} b_j^i(\boldsymbol{o}_t^i) \geq 0, \forall i, \forall j \in \mathcal{N}^i(\boldsymbol{o}_t^i), \forall t \geq 0$$

Existing Works

- Model Predictive Control [1], [2].
- Multi-Agent Reinforcement Learning [3].
- Multi-Agent Constrained Policy Optimization [4].
- RL with Safe Control [5].

Our approach: A Hierarchical Multi-Agent Reinforcement Learning Approach using CBF based Skills (HMARL-CBF).

1. A. Carron, D. Saccani, L. Fagiano and M. N. Zeilinger, "Multi-agent Distributed Model Predictive Control with Connectivity Constraint," *IFAC PapersOnLine*, vol. 56, no. 2, pp. 3806-3811, 2023.
2. P. Wang and B. Ding, "A synthesis approach of distributed model predictive control for homogeneous multi-agent system with collision avoidance," *International Journal Control*, vol. 87, no. 1, pp. 52-63, 2014.
3. R. Lowe, Y. Wu, A. Tamar, et al., "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," *Advances in Neural Information Processing Systems (NeurIPS),* 2017.
4. Y. Zhao, Y. Yang, Z. Lu, W. Zhou and H. Li, "Multi-Agent First Order Constrained Optimization in Policy Space," in *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, 2023.
5. Y. Cheng, P. Zhao and N. Hovakimyan, "Safe and efficient reinforcement learning using disturbance-observer-based control barrier functions," in *Proc. 5th Annual Conference on Learning for Dynamics and Control (L4DC)*, 2023, pp. 104-115

HMARL-CBF

Safety-constrained cooperative game:

$$\min_{\pi_1,\ldots\pi_N} \mathbb{E}_{\tau\sim(\pi_1,\ldots,\pi_N)}\left[\sum_{t=0}^{\infty} \gamma^t l(\boldsymbol{x}_t, \boldsymbol{u}_t)\right]$$

Subject to: $\mathbb{E}\left[\min\left\{b_j^i\left(\boldsymbol{o^i}\right), \boldsymbol{0}\right\}\right] \geq 0, \forall i \in \mathcal{N},$
$j \in \mathcal{N}^i(\boldsymbol{o^i}), \forall t \geq 0$

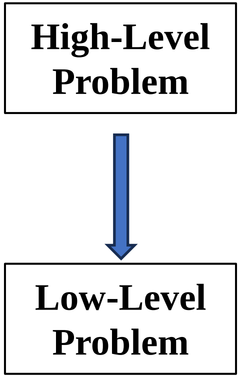
Hierarchical Policy Learning

High-Level Problem
Learning joint cooperative behavior using skills.

Low-Level Problem
Learning the agent skills.

HMARL-CBF

Hierarchical Policy Learning



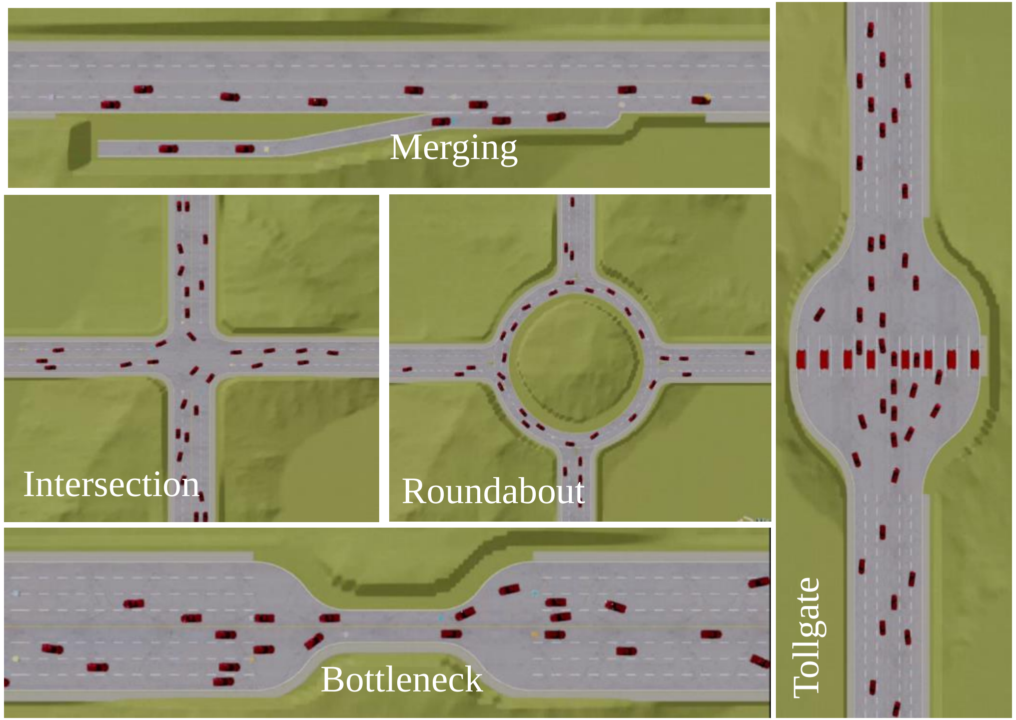
- Solve the Unconstrained Problem – minimize the total discounted cost
- Learns a feedback policy that maps to skills of agents.

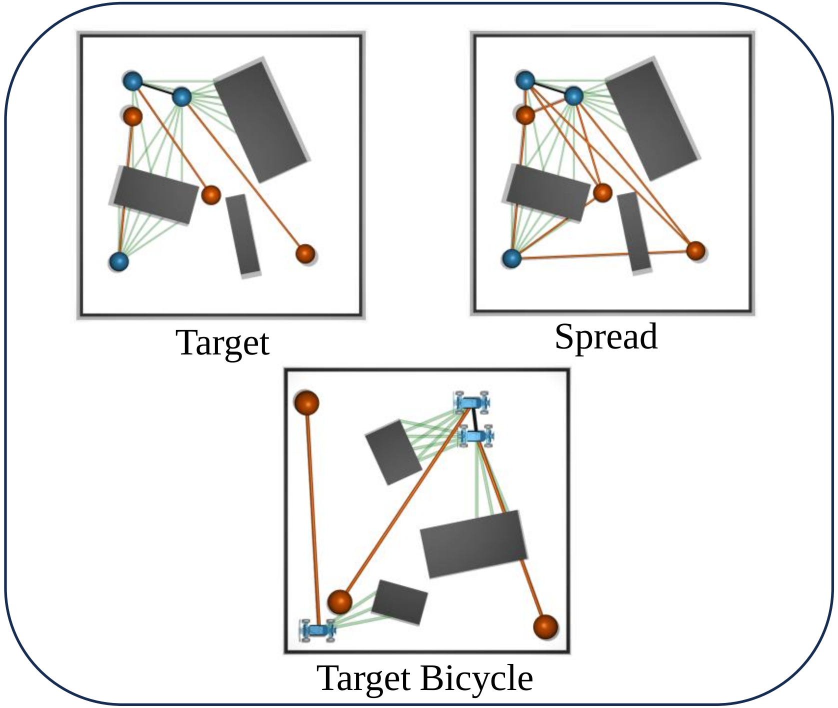
- Agent-Wise Constrained Problem solved using CBFs to learn the skill policy.
- The constraints correspond to safety execution.

Remarks

- High level policy optimization focuses on coordination.
- High level problem is unconstrained and solved in lower dimensional latent space – reduced sample complexity.
- Safety is enforced locally through agent-wise constrained problems.
- Shared low-level Agent-wise problems assuming homogeneity – scalability, transferability and generalizability.

Results - Environment (METADRIVE)



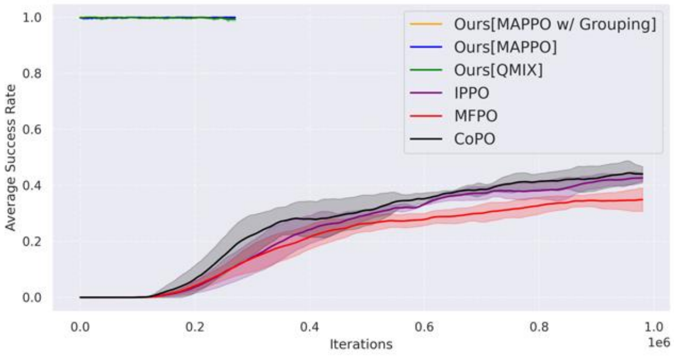


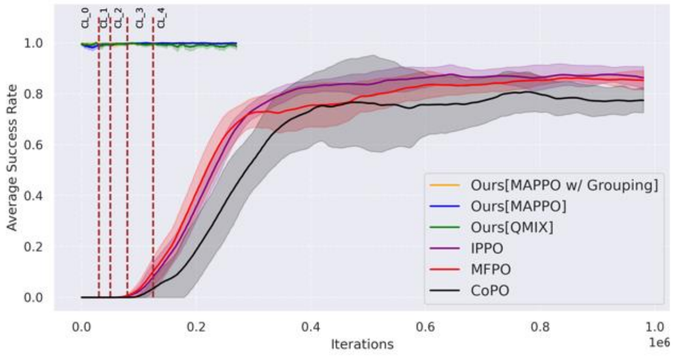
Results - Training

Merging Environment

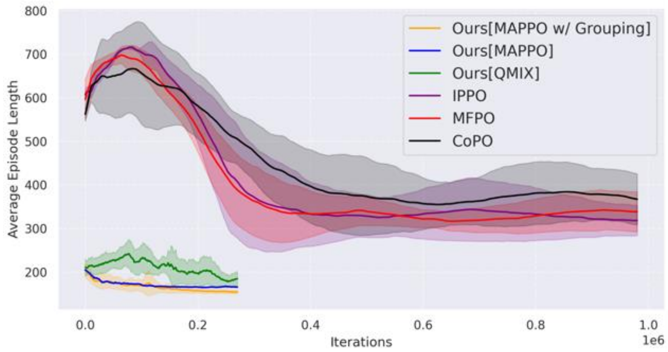
Roundabout Environment

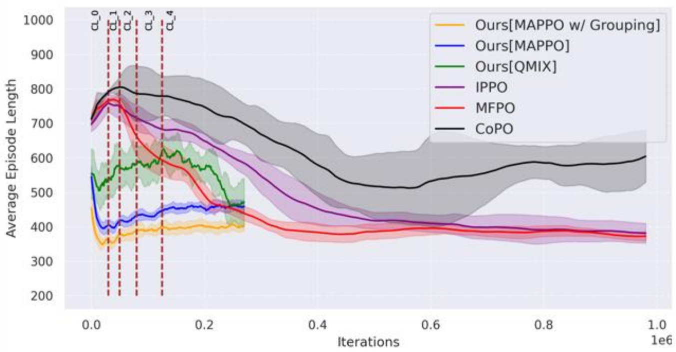
Success Rate





Average Travel Time





Conclusion

- Propose HMARL-CBF, a hierarchical MARL method combining high-level skill selection with low-level safe execution via CBFs, learnt jointly without posing additional sample complexity.
- Our approach guarantees safety during training and evaluation.
- The framework is scalable to a large number of agents.

Project Page

Depend Lab

Realm Lab

Codes Lab