ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Agile Collision Avoidance for Deformable-Tethered Multi-Robot Systems via Zone-Aware Hierarchical Learning and VLM-Guided Control

Zeyu Zhou[1,2], Jingwei Zhang[3], Hui Zhi[1], Yun Hao[1], Wei Tang[2], David Navarro-Alarcon[1]
[1]The Hong Kong Polytechnic University, Hong Kong [2]Northwestern Polytechnical University, Xi'an, China
[3]University of Chinese Academy of Sciences, Beijing, China Email: zeyu-zhou.zhou@connect.polyu.hk

***Abstract*—Navigating Linked Multi-Component Robotic Systems (L-MCRS)—robot pairs tethered by passive flexible hoses—through dynamic pedestrian environments is fundamentally harder than rigid multi-robot coordination, as the uncontrollable hose creates a variable-geometry collision footprint spanning 118 pairwise combinations. We propose H-SEPID, unifying zone-aware Hierarchical Reinforcement Learning grounded in Kinematic Flow Theory with VLM-guided cascaded optimization. A phase-aware dual attention value network performs C^0-continuous topological policy switching, while a Vision-Language Model infers strategic intent and quantifies action-space constraints governing hose geometry. A seven-category safety shield with ORCA fallback and a threading reward band produce emergent gap-threading maneuvers. H-SEPID achieves 94% success and 4% collision rate in an 8-robot, 5-pedestrian, 4-hose scenario, outperforming five baselines, and is validated on real e-puck2 robots across 12 configurations.**

I. Introduction

Linked Multi-Component Robotic Systems (L-MCRS)—robot pairs tethered by flexible hoses for material/energy transport [1]—face a unique challenge: the passive hose is entirely governed by inter-robot positioning, yet creates a continuously varying collision footprint. In the full configuration of 8 robots, 5 pedestrians, and 4 hoses, the pairwise collision combinations reach $\binom{8}{2}+8\times4+8\times5+4\times5+\binom{4}{2}=118$. Classical planners (DMPC [2], ECBS [3]) ignore hose coupling; reactive methods like ORCA [4] lack long-horizon reasoning; and flat MARL approaches (MAPPO [5], HAC [6]) have no structural priors for deformable-tether physics.

We propose **H-SEPID** with three contributions: (1) zone-aware HRL grounded in Kinematic Flow Theory—feasible trajectories define Flow Tubes whose intersections form an Interaction Potential Manifold approximated by a Hierarchical Simplicial Cover, with a phase-aware dual attention value network for C^0-continuous policy switching; (2) VLM-RL cascaded optimization where a VLM infers strategic intent (conservative, aggressive, thread-the-needle) and quantifies it into action-space constraints including partner distance and formation angle; (3) emergent gap-threading via a seven-category safety shield, threading reward band $[0.20, 0.50]$ m, and TTC anticipatory avoidance with an analytical hose model (Fig. 1).

(a) Single-pair gap-threading (b) Multi-pair coordination

Fig. 1. Tethered L-MCRS in dynamic pedestrian environments. (a) Single pair reshapes hose topology to thread through gaps. (b) Multiple pairs coordinate to prevent inter-pair hose entanglement.

II. Method

Framework overview. H-SEPID operates as a four-layer pipeline: egocentric perception, VLM strategic reasoning via cross-modal attention fusion $\mathbf{h} = \text{VLM}(\mathbf{z}_V \oplus \mathbf{z}_L)$ classifying three behavioral modes, zone-aware HRL with C^0-continuous policy blending, and formation-coordinated action execution. Training follows CTDE with full parameter sharing.

Zone-aware hierarchical RL. The workspace is modeled as a manifold where interaction zones emerge from topological intersections of Feasible Transition Tubes $\mathcal{T}_{i\to j}$. Their swept volumes form the Interaction Potential Manifold $\mathcal{I}_{\text{int}} = \text{cl}(\bigcup \Omega_{i\to j} \cap \Omega_{p\to q})$, approximated by a Hierarchical Simplicial Cover $\mathcal{K}_\Delta$. A phase-aware dual attention value network (Fig. 2) realizes topological switching:

$$V(s_t) = \alpha_h(\phi_t)V_{\text{HA}}(s_t) + \alpha_r(\phi_t)V_{\text{RA}}(s_t), \qquad (1)$$

where V_{HA} prioritizes pedestrian avoidance and V_{RA} prioritizes inter-robot coordination. Routing weights are zone-dependent with C^0 blending via $k(x) = (1+\exp(-\alpha \cdot \text{dist}(x, \partial\mathcal{K}_\Delta)))^{-1}$. Each sub-network (64→32→16 layers) employs self-attention with numerically stable masked softmax over variable-length entity sets.

VLM-RL cascaded optimization. The VLM infers intent $\xi \in \{\xi_{\text{cons}}, \xi_{\text{agg}}, \xi_{\text{ttn}}\}$ and generates a quantified de-

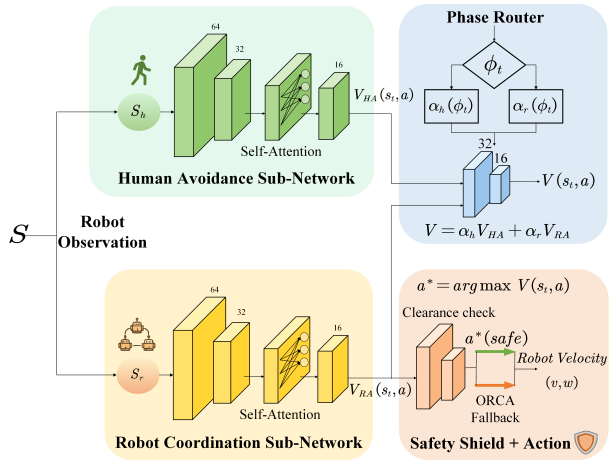


Fig. 2. Phase-aware dual attention value network with Phase Router and Safety Shield (ORCA fallback).

cision vector $\mathbf{q} = [\tau_{\text{cons}}, v_{\text{sug}}, d_{\text{target}}, \theta_{\text{form}}]^\top$ constraining the RL action space. Action selection follows $a^* = \arg\max_{a \in \mathcal{A}^t_{\text{VLM}}} Q_\theta(s_t, a)$.

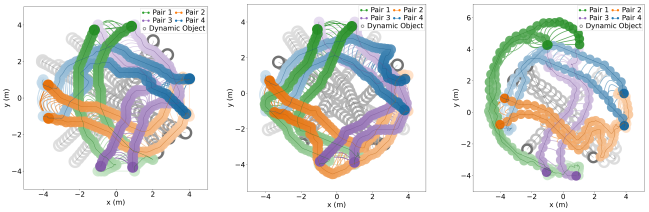
Safety shield & gap-threading. A multi-step shield checks seven collision categories (robot–human, robot–robot, hose–human, hose–robot, partner overlap, other-pair hose, out-of-bounds) with a four-stage ORCA fallback. The shield *enables* aggression by preserving "risky but safe" actions. A threading bonus $r_{\text{thread}}{=}{+}0.03$ for clearances in $[0.20, 0.50]$ m with positive goal progress, TTC penalties, and an analytical hose model with side-flip detection jointly produce emergent gap-threading.

III. Experiments

We evaluate under `circle_crossing`: 8 robots (4 hose pairs, $L{=}2.0$ m) on $R{=}4$ m with 5 ORCA pedestrians, 500 episodes$\times$3 seeds. Table I shows H-SEPID achieves 94% success rate with only 4% collision rate, substantially outperforming all baselines. Classical planners (DMPC, ECBS) fail to model hose coupling, achieving only 28–32% success; ORCA suffers deadlocks in 42% of its failure cases due to myopic decision-making; MAPPO and HAC lack structural priors for multi-phase tethered navigation, plateauing at 62–66% success.

Emergent strategies. The VLM-guided mode selection produces qualitatively distinct strategies across scales (Fig. 3). With 4 robots, aggressive mode (ξ_{agg}) yields direct gap-threading with hose clearances averaging 0.28 m; thread-the-needle mode (ξ_{ttn}) achieves precision hose reshaping through inter-pedestrian corridors; conservative mode (ξ_{cons}) prefers circumnavigation. Scaling to 8 robots reveals cooperative queuing and sequential gap entry patterns arising spontaneously from zone-aware policy switching.

Sim-to-real transfer. Deployment on 13 e-puck2 robots (8 tethered + 5 pedestrian proxies) via overhead camera at 30 Hz achieved 95% success under localization occlusion and 100% under position swap, across 12 obstacle configurations, with $<$5% performance drop versus simulation.



(a) Aggressive (b) Thread-the-needle (c) Conservative

Fig. 3. Emergent trajectories of 4 L-MCRS pairs (8 robots) vs. 5 pedestrians. (a) Aggressive: tight-orbit circumnavigation. (b) Thread-the-needle: gap exploitation. (c) Conservative: wide-orbit peripheral routing.

TABLE I
Performance comparison (500 episodes $\times$ 3 seeds).

Method	Type	SR↑	CR↓	FR↓	NT↓
DMPC [2]	Classical	0.32	0.45	0.68	6.82 s
ECBS [3]	Classical	0.28	0.38	0.72	7.14 s
ORCA [4]	Reactive	0.51	0.22	0.49	5.87 s
MAPPO [5]	Learning	0.62	0.18	0.38	5.24 s
HAC [6]	Learning	0.66	0.16	0.34	5.05 s
H-SEPID	Learning	**0.94**	**0.04**	**0.06**	**4.12 s**

IV. Conclusion

H-SEPID is the first framework enabling deformable tethered multi-robot systems to perform agile gap-threading among dynamic obstacles, achieving 94% success versus 66% for the best baseline with 75% collision reduction. The emergent thread-the-needle capability arises without explicit gap-threading programming. Sim-to-real transfer on e-puck2 robots confirms viability. Future work targets heterogeneous L-MCRS chains and formal CBF guarantees.

References

[1] J. Alonso-Mora, R. A. Knepper, R. Siegwart, and D. Rus, "Local motion planning for collaborative multi-robot manipulation of deformable objects," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5495–5502.

[2] C. E. Luis, M. Vukosavljev, and A. P. Schoellig, "Online trajectory generation with distributed model predictive control for multi-robot motion planning," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 604–611, 2020.

[3] M. Barer, G. Sharon, R. Stern, and A. Felner, "Suboptimal variants of the conflict-based search algorithm for the multi-agent pathfinding problem," in *Proceedings of the international symposium on combinatorial Search*, vol. 5, no. 1, 2014, pp. 19–27.

[4] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha, "Optimal reciprocal collision avoidance for multi-agent navigation," in *Proc. of the IEEE International Conference on Robotics and Automation, Anchorage (AK), USA*, 2010.

[5] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Baez, B. Bhatt, D. Fritsch, A. Bhatt, O. Ott *et al.*, "The surprising effectiveness of PPO in cooperative multi-agent games," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 611–24 624, 2022.

[6] A. Levy, G. Konidaris, R. Platt, and K. Saenko, "Learning multi-level hierarchies with hindsight," in *International Conference on Learning Representations*, 2019.