Resilient Scheduling for Urban Infrastructure Maintenance and Inspection Fleets: A Potential-Guided MARL Approach

Shuyang Zhang[1]

[1]Shanghai Research Institute of Building Sciences Co., Ltd. China
zhangshuyang@sribs.com

Abstract –

Efficient fleet scheduling in urban infrastructure maintenance and inspection services requires a delicate balance between task productivity and operational safety. However, traditional centralized algorithms often struggle with the stochastic nature of dynamic urban environments, leading to communication latency and collision risks. To bridge the gap between global platform planning and local reactive execution, this paper proposes a Potential-Guided Multi-Agent Reinforcement Learning (PG-MARL) framework designed for resilient edge computing. By integrating global BFS-based topological guidance with decentralized reinforcement learning, the method utilizes a lightweight state representation to significantly accelerate training convergence and enable agents to navigate complex, non-convex environments autonomously. Experimental results in a discretized semantic grid environment demonstrate that PG-MARL achieves near-zero collision rates with an operational efficiency trade-off compared to optimal planning baselines. This study provides a resilient and safety-aware solution suitable for practical deployment as an autonomous edge intelligence module for city-level operation management platforms.

Keywords –

Multi-robot systems; Potential-Guided Multi-Agent Reinforcement Learning; Urban Infrastructure Maintenance and Inspection

1 Introduction

The rapid advancement of smart city infrastructures and construction automation has rendered Multi-Robot Systems indispensable in contemporary urban maintenance services, spanning infrastructure inspection, last-mile logistics, and environmental sanitation. As the scope of robot fleet management expands into increasingly complex domains [1], the fundamental challenge remains the achievement of efficient fleet scheduling and coordination within environments defined by intricate and unpredictable dynamic obstacles. In such urban ecosystems, safety and operational efficiency must be co-optimized through distributed intelligence to manage city-wide route planning and regional collision-free coordination [2].

Standard industrial solutions predominantly utilize centralized task allocation mechanisms, such as auction-based algorithms, coupled with deterministic path planning. Extensive research has sought to refine these methods, ranging from group-based distributed auctions for vehicle routing [3] to sequential single-item auctions that incorporate routing constraints to minimize congestion [4]. Other approaches have introduced resource abstraction [5] and multi-role goal assignment [6] to handle temporal tasks and high-dimensional state spaces. While these frameworks demonstrate superior theoretical efficiency in controlled settings, they manifest profound vulnerabilities during practical deployment. Even decentralized variants, such as tri-layered negotiation algorithms [7] or specialized goal distribution strategies for heterogeneous teams [8], often struggle with the inherent rigidity and computational overhead required to react to real-time disruptions.

The stochastic nature of urban dynamics—such as the sudden emergence of temporary maintenance zones or unexpected roadblocks—can instantly invalidate pre-computed rules. When preset logic fails to account for these disruptions, systems often descend into operational deadlocks or hazardous collisions. Multi-Agent Reinforcement Learning (MARL) has emerged as a principled alternative for decision-making in such uncertain environments [9], capable of capturing complex bidder behaviors and interaction equilibria that traditional models fail to reach [10]. Furthermore, a hierarchical needs-driven perspective suggests that satisfying lower-level safety requirements is a critical prerequisite for achieving higher-level teaming and learning objectives [11] [12]. However, standard MARL often suffers from training instability and the sparse reward trap in non-convex environments.

Given these limitations, recent research has explored hybrid architectures that combine the real-time

adaptability of RL with the global optimization of planning algorithms [9]. Specifically, potential-based reward shaping has been identified as a powerful mechanism to accelerate convergence and address credit assignment without altering the underlying optimal policy [12]. Ensuring safe motion in dense, obstacle-rich scenarios further requires the integration of mathematical constraints, such as those found in safe coverage control algorithms [13]. Consequently, there is an urgent need for more resilient, safety-aware control paradigms that can adapt to environmental flux without requiring constant centralized re-computation.

To address the latency limitations of centralized city-level operation management platforms, this paper proposes a Potential-Guided Multi-Agent Reinforcement Learning (PG-MARL) framework designed as a resilient edge-computing decision module for autonomous fleet scheduling. By integrating a BFS-based static map potential field into the learning process, we effectively accelerate the convergence of Q-learning, enabling agents to transcend the sparse reward trap while maintaining robust, autonomous obstacle avoidance capabilities. This lightweight architecture emphasizes practical deployment by reducing the computational burden on intelligent inspection equipment, providing a resilient alternative to traditional rule-bound systems under unified platform supervision. Through this approach, we demonstrate that a learning-based fleet can maintain high operational efficiency while ensuring the near-zero collision rates necessary for safe urban service integration.

The contributions of this research are summarized as follows:

- A hybrid control framework that integrates global BFS-based potential guidance with localized multi-agent reinforcement learning to enhance both convergence efficiency and decentralized adaptability.
- An empirical evaluation on a simplified urban scenario, which suggests that the proposed strategy can achieve near-zero collision rates with a trade-off in operational efficiency, demonstrating its potential for safety-critical practical deployments.

2 Methodology

The methodology of this research focuses on establishing a robust framework for autonomous fleet coordination within a constrained urban environment. We model the Urban infrastructure maintenance and inspection scheduling problem as a partially observable multi-agent system, integrating potential-based guidance to enhance the training efficiency of reinforcement learning.

2.1 Problem Formulation

The urban infrastructure maintenance and inspection environment is modeled as a discrete grid environment $\mathbf{G} \in \mathbb{R}^{10\times10}$, which represents a discretized district-level sector derived from City Information Modelling (CIM) data. We define a fleet of N autonomous agents ($N = 3$) tasked with visiting M distressed infrastructure points (e.g., road cracks, facility failures) ($M = 6$). The state space is characterized by the relative positions of agents and tasks, while the action space $\boldsymbol{A}$ consists of five discrete maneuvers: $\boldsymbol{A} = \{Up, Down, Left, Right, Stay\}$. This setup simulates a small-scale urban patrol scenario where temporary obstructions, including temporary road closures or maintenance equipment, act as continuous barriers $\boldsymbol{O}_{static}$. Crucially, these barriers transform the theoretically convex free space into a non-convex space, creating geometric constraints that pose challenges for simple heuristic planners. To mirror real-world logistical constraints, the system operates under a decentralized execution paradigm. Unlike traditional scenarios with perfect global maps, agents here rely on limited local observations. This partial observability, combined with the non-convex topology, creates dynamic local minima traps where standard algorithms typically oscillate or freeze without continuous global re-computation.

2.2 Modelling Baselines

To evaluate the efficacy of the proposed method two industrial-standard benchmarks are established based on distinct algorithmic logics:

- Greedy Heuristic: This baseline utilizes a local nearest-neighbour strategy. Each agent independently identifies the closest task based on the Manhattan distance $d_1 = |x_1 - x_2| + |y_1 - y_2|$and executes the action that reduces this distance. This strategy is used as the naive baseline.
- Auction+BFS Strategy: This method employs the Hungarian algorithm for global task allocation. The system constructs a global cost matrix based on distances between agents and tasks. It then utilizes the linear sum assignment method to solve for the global optimal assignment that minimizes the total fleet travel distance. Once a task is assigned, the agent utilizes a Breadth-First Search (BFS) algorithm to generate the shortest path that strictly avoids static obstacles. This strategy is used as a strong baseline. While theoretically efficient in static environments, this method is socially blind during execution.

2.3 Potential-Guided Multi-Agent Reinforcement Learning

2.3.1 Lightweight State Representation

To ensure the framework's adaptability to resource-constrained edge devices, we architect a lightweight state representation $\boldsymbol{s}$ (see Equation (1)) designed to decouple localized decision-making from taxing global localization systems. Unlike traditional navigation paradigms that rely on continuous global coordinate fusion, our state key is defined as a compact tuple that captures only the essential relative orientation and environmental occupancy:

$$\boldsymbol{s} = (sgn(\Delta x), sgn(\Delta y), \boldsymbol{O}_{local}) \quad (1)$$

Where $sgn(\Delta x), sgn(\Delta y) \in \{-1, 0, 1\}$ denote the sign of the relative displacement to the assigned target, providing a scale-invariant navigational gradient. The term $\boldsymbol{O}_{local}$ is a 4-bit binary vector derived from a hardware-agnostic perception layer. This layer abstracts the feedback from localized proximity sensing modules (e.g., ultrasonic, infrared, or LIDAR) into discrete occupancy signals for the immediate adjacent cells. If a sensor detects a static obstacle or boundary, it returns 1; otherwise, it returns 0.

This compact state space significantly reduces the dimensionality of the Q-table compared to global coordinate inputs, making the algorithm suitable for deployment on edge computing hardware with limited memory and computational power.

2.3.2 Exploration and Update Mechanism

The agents employ an ϵ-greedy strategy to balance exploration and exploitation. To ensure sufficient learning of the environment's topology, the exploration rate ϵ linearly decays from an initial value of $\epsilon_{start} = 0.9$ to a minimum of $\epsilon_{end} = 0.05$ over the course of training. The policy is updated using the standard Q-learning Bellman equation as Equation (2):

$$Q(s,a) \leftarrow Q(s,a) + \alpha[R_{total} + \gamma \max_{a'} Q(s',a') - Q(s,a)] \quad (2)$$

Where $Q(s,a)$ is the action-value function of the agent, with each agent maintaining an independent Q-table under the decentralized framework. The term s and a denote the agent's lightweight observation state and selected discrete action, respectively. While s' and a' denote the subsequent state after action execution and all optional actions in the new state, respectively. R_{total} is the total single-step reward including task reward, collision penalty, step cost, and BFS-based potential shaping reward; $\max_{a'} Q(s',a')$ represents the maximum expected Q-value of the agent in the subsequent state s' under the Bellman optimality principle. The coefficient α is the learning rate controlling the Q-value update amplitude, and γ is the discount factor balancing immediate and long-term rewards.

In our implementation, the learning rate is set to $\alpha = 0.1$ and the discount factor to $\gamma = 0.95$ to prioritize long-term mission success.

2.3.3 Reward Function and Safety Constraints

A carefully calibrated reward system structure is critical for balancing efficiency and safety. The total reward R_{total} is composed of task completion R_{task}, R_{coll} and R_{step}. R_{task} is the significant positive reward, which is granted upon target arrival to incentivize mission completion. R_{coll} is the collision penalty. To enforce strict safety standards, a heavy penalty is applied whenever an agent collides with a static obstacle or a teammate. This high penalty forces the agent to learn social yielding behaviours. R_{step}, the time step cost, is a small negative reward is set per step to encourage the agent to find the shortest path.

2.3.4 Potential-Based Reward Shaping

A primary challenge in applying reinforcement learning to urban navigation is the local minima trap in geometrically constrained environments. Standard Manhattan distance heuristics often fail in scenarios with continuous barriers because they do not account for the geodesic path around a wall. To overcome this, we introduce a BFS-based Reward Shaping mechanism. We define a potential function $\Phi(s,g)$ based on the BFS-calculated shortest distance to the goal g. The potential reward R_p is formulated as Equation (3):

$$\boldsymbol{R}_p = \eta \cdot \big(\Phi(s_t, g) - \Phi(s_{t+1}, g)\big) \quad (3)$$

Here, η represents the shaping weight. From a physical perspective, this potential field acts as a navigational gradient, logically guiding the agent through the environment's topology. When an agent moves around a barrier, the BFS distance decreases, providing an immediate positive signal even before the final task reward is reached. This mechanism significantly accelerates the convergence rate, transforming a sparse-reward problem into a dense-feedback engineering solution feasible for rapid training.

3 Experiments and Results

3.1 Experiment Setup

To validate the proposed methodology, we establish a physical-to-computational mapping as shown in Figure 1. A typical urban district is semantically abstracted into a 10×10 discrete grid world. As illustrated in Figure 1,

the physical context (left) involving six maintenance tasks and specific road closures is mapped into the computational grid (right). Blue dots represent autonomous agents, red stars indicate discretized task locations, and black regions represent continuous static barriers (e.g., zones closed to traffic).

The initial coordinates of agents and tasks are stochastically distributed to avoid obstacle-dense areas, creating a scenario specifically designed to test the ability of the algorithms to navigate around local minima traps. For the multi-agent task allocation and collision avoidance objectives, a reward shaping mechanism is calibrated to balance mission efficiency with stringent safety constraints.

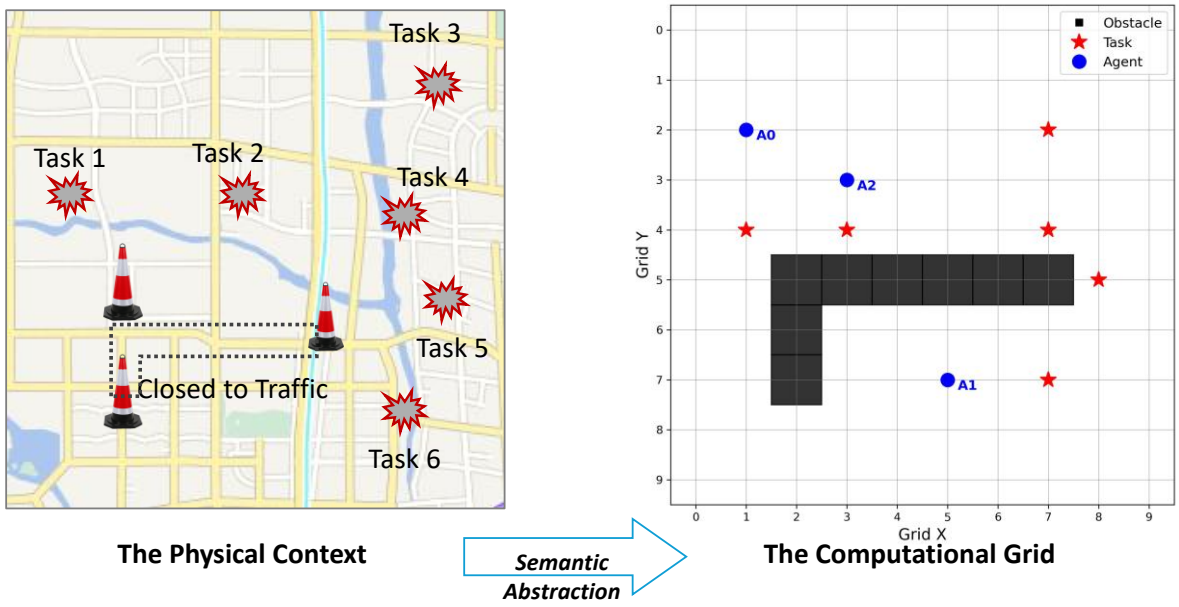

Figure 1. Illustration of the semantic abstraction workflow: mapping the physical urban infrastructure context (left) into a discretized computational grid (right) derived from CIM data.

The core reward weights are defined as follows: a task completion reward of +200.0 serves as the primary positive incentive, a collision penalty of -15.0 enforces strong safety requirements, and a time-step penalty of -0.5 provides slight temporal pressure. Furthermore, the BFS-based reward shaping weight is set at 1.5 to utilize pathfinding priors for accelerated convergence without inducing policy bias. This calibration ensures that the reward signals are well-suited to both the task logic and the convergence characteristics of reinforcement learning, preventing strategy deviations caused by weight imbalances.

3.2 Training Convergence

The iterative learning progression of the PG-MARL framework is demonstrated in Figure 2 and Figure 3.

As training progresses, the number of steps required for agents to complete each task shows a consistent downward trend, with a rapid reduction in task completion steps observed in the early training phase. By around Episode 1000, the step count has dropped to a relatively low range.

For safety performance, as shown in Figure 3, the collision count decreases rapidly in the early training phase, reaching single-digit values by Episode 1000, and continues to decline thereafter. After 1500 episodes, the collision frequency remains at an extremely low level. From Episode 2000 to the end of training, the algorithm maintains a consistently low collision rate.

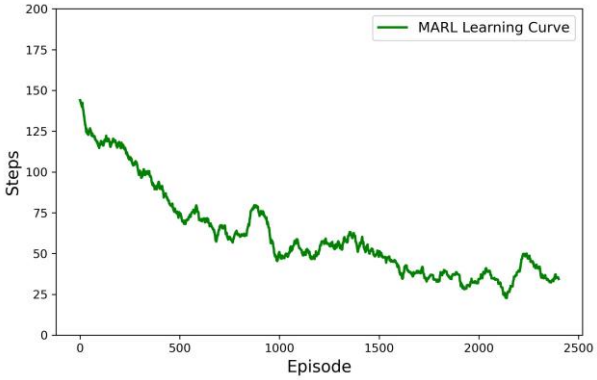

Figure 2. Convergence of total steps during the PG-MARL training phase

Integrating the safety and efficiency metrics, the PG-MARL algorithm exhibits stable performance in both step count and collision frequency around the 2000th

episode. This accelerated learning process and stabilized training dynamics are attributed to the BFS based potential guidance mechanism, which allows the RL agents to transcend the sparse reward problem and gradually master advanced obstacle avoidance and navigational maneuvers.

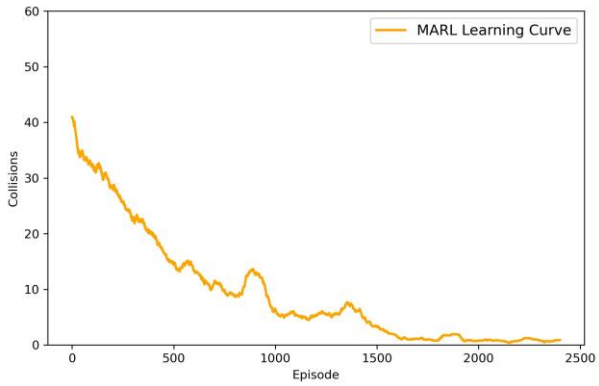

Figure 3. Evolution of obstacle avoidance performance during the PG-MARL training process

3.3 Operational Efficiency Analysis

The efficiency analysis based on the statistical results from Figure 4 and Figure 5 reveals the performance differences among the three strategies.

As illustrated in Figure 4, the Auction+BFS strategy achieves the theoretical minimum in total steps by utilizing a global optimal allocation coupled with static shortest-path planning. This combination eliminates the path redundancy caused by local decision-making and ensures minimal travel distance for the fleet in the static grid environment, making it the efficiency benchmark for the evaluated methods.

The PG-MARL approach demonstrates competitive efficiency, with its median total steps slightly higher than the Auction+BFS baseline. This marginal efficiency gap constitutes a deliberate trade-off for the algorithm's inherent dynamic collision avoidance capability, as the learned policy will occasionally adjust paths to avoid inter-agent conflicts in real time. Notably, PG-MARL demonstrates a far more concentrated box distribution than the Greedy heuristic, indicating significantly lower variability and superior stability in task execution performance.

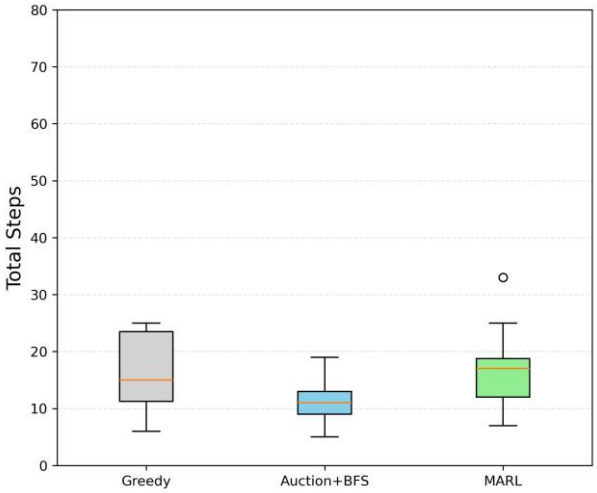

Figure 4. Statistical comparison of operational efficiency among Greedy, Auction+BFS, and PG-MARL strategies over 50 test iterations.

Computational time efficiency across the three strategies is further quantified in Figure 5, where PG-MARL is evaluated by decoupling training time and execution time to reflect practical deployment performance. Greedy yields the lowest time cost due to its simplistic local decision-making without pre-computation, yet at the cost of poor safety. Auction+BFS incurs moderate overhead from repeated Hungarian allocation and BFS planning, while PG-MARL exhibits the longest execution time, shown as the green bar in Figure 5.

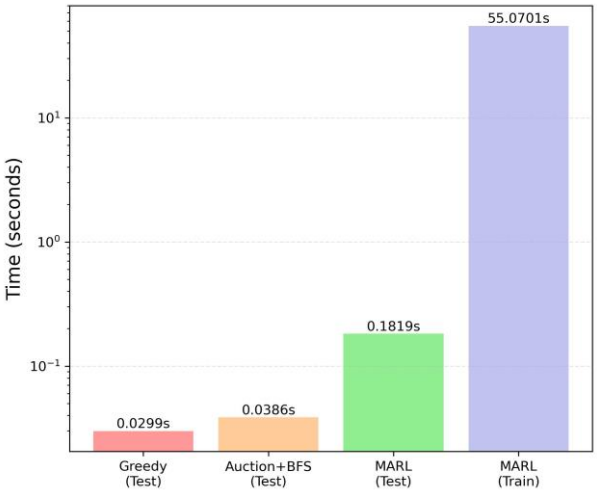

Figure 5. Computational overhead for Greedy, Auction+BFS, and PG-MARL during testing and training.

However, such time- and step-related metrics alone fail to present a comprehensive evaluation of the three

strategies, as none of them incorporate the substantial time and efficiency losses caused by collisions in the actual operation process. A holistic performance assessment must therefore consider the collision occurrences among agents under each strategy, and for this reason, a detailed statistical analysis of collision frequencies is conducted in Section 3.4 to further compare the safety performance of the three scheduling methods.

3.4 Safety Assessment

The critical advantage of the PG-MARL framework is elucidated in the safety performance analysis as shown in Figure 6.

The Greedy baseline records the highest collision frequency, as it lacks both static obstacle avoidance and dynamic coordination. Statistical results indicate that the Auction+BFS strategy significantly mitigates the frequency of collisions relative to the Greedy baseline. However, while this approach achieves the lowest mean total step count, the collision distribution exhibits several outliers, with incident counts ranging from 3 to 10 per episode. These data points demonstrate that despite its superior operational efficiency, the Auction+BFS strategy retains inherent collision risks. Collisions in the Auction+BFS baseline primarily stem from real-time dynamic inter-agent conflicts, which the static BFS planner fails to anticipate.

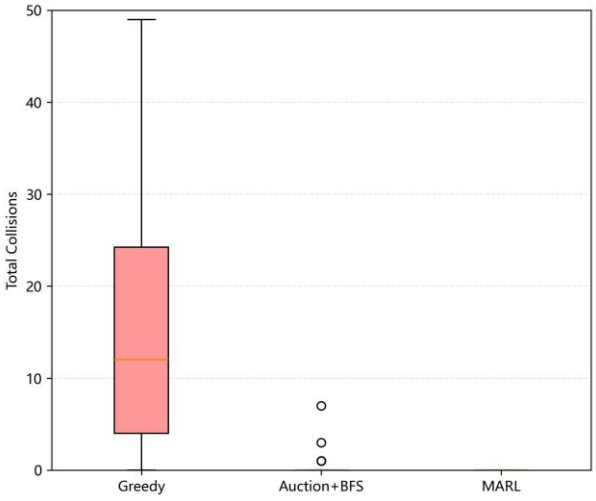

Figure 6. Total collisions during dynamic interactions for the three evaluated scheduling strategies

In contrast, the proposed PG-MARL strategy consistently achieves near-zero collision performance with negligible variance across all test iterations.

A comparison of Figure 5 and Figure 6 reveals that the raw execution time of Greedy is deceptive, as it fails to account for the catastrophic time and efficiency losses induced by collisions. In real-world urban maintenance, a collision necessitates hardware recovery, system re-initialization, or mission abortion—time penalties that far exceed the millisecond-level inference delay of PG-MARL. By ensuring zero-collision navigation, PG-MARL optimizes the net operational uptime, offering a more robust and ultimately more efficient solution for mission-critical infrastructure inspections.

These observations quantify the ability of the learned policy to maintain a stable safety profile in constrained urban infrastructure maintenance and inspection scenarios.

4 Discussion

4.1 Social Awareness in Dynamic Interaction

The experimental results reveal a profound distinction between the rigid coordination of traditional algorithms and the socially compliant behavior of the proposed PG-MARL framework. While the Auction+BFS strategy achieves high theoretical efficiency through global task allocation and static path planning, it lacks the necessary social awareness to handle real-time dynamic conflicts. As observed in the safety assessment, the deterministic nature of BFS paths often leads agents into unavoidable collisions when their trajectories intersect in narrow corridors or task vicinities. In contrast, the PG-MARL policy, reinforced by a significant collision penalty, internalizes the presence of teammates as dynamic constraints. By penalizing overlapping positions during the training phase, agents effectively learn yielding and negotiation behaviors. This emergent social navigation capability allows the fleet to resolve complex deadlocks autonomously, proving that learned adaptability is more resilient than rigid planning in high-density urban infrastructure maintenance and inspection scenarios.

4.2 Impact of Potential-Guided Reward Shaping

A critical technical contribution of this research is the utilization of BFS-based potential fields to overcome the "local minima" inherent in non-convex urban environments. Traditional heuristics, such as the Manhattan distance $d_1 = |x_1 - x_2| + |y_1 - y_2|$, frequently fail when agents encounter non-convex scenarios, as the distance gradient does not align with the traversable geodesic path. By integrating a BFS-based shaping weight ($\eta = 1.5$), we provide a continuous and logically sound navigational gradient that guides agents around physical barriers. This mechanism transforms a sparse-reward task into dense-feedback engineering

problem, enabling the system to achieve stable convergence within a practical timeframe of 2500 training episodes. The rapid stabilization of both step counts and collision rates suggests that the potential-guided approach is not only mathematically robust but also highly efficient for training complex multi-agent policies in geometrically constrained maps.

4.3 Insights into Practical Deployment and Scalability

From the perspective of deployment within a city-level operation management platform, the proposed framework emphasizes computational efficiency and decentralized robustness. While centralized platforms often suffer from communication latency when controlling large-scale fleets, our PG-MARL agents utilize a lightweight state key consisting of relative goal signs and data from four basic proximity sensors. Due to this sign-based relative state encoding, the acquired navigation policy exhibits inherent scale-invariance, allowing the BFS-guided gradient to remain effective across varying spatial dimensions without increasing the computational complexity of the edge-intelligence module. Crucially, the utilization of a discretized grid in our experiment is not merely a simulation constraint, but reflects the semantic abstraction derived from City Information Modeling (CIM) data, designed to minimize the computational burden on edge devices. Furthermore, the grid-based action space naturally aligns with the Manhattan-style topology of modern urban road networks, ensuring that the learned navigational maneuvers are directly transferable to real-world street operations.

This lightweight design minimizes the computational burden on edge devices with minimal memory overhead, serving as the resilient execution layer for smart inspection equipment. Furthermore, the observed trade-off—where PG-MARL achieves near zero-collision rates at the cost of an increase in total steps—carries significant economic implications for urban maintenance. In safety-critical environments, the operational cost associated with a single collision far exceeds the minor energy expenditure of a slightly sub-optimal path. By prioritizing safety-aware navigation, the proposed method offers a resilient and scalable solution for real-world autonomous fleet operations in the increasingly complex landscapes of smart cities.

5 Conclusion

This research presented a Potential-Guided Multi-Agent Reinforcement Learning (PG-MARL) framework designed for autonomous fleet scheduling in urban infrastructure maintenance and inspection missions. By embedding BFS-derived potential fields into a decentralized learning architecture, we effectively resolve navigation bottlenecks within complex urban environments characterized by non-convex obstacles. The empirical results demonstrate that while the industry-standard Auction algorithm coupled with BFS pathfinding achieves theoretical optimality in path length, the proposed PG-MARL approach provides a qualitative leap in operational safety. By achieving near-zero collision rates with an operational efficiency trade-off, the proposed framework demonstrates superior operational resilience for safety-critical maintenance missions.

From the perspective of practical deployment, the decentralized execution paradigm significantly reduces the city-level platform's reliance on high-frequency global communication, as agents derive decisions from localized proximity sensors and relative goal directions. The lightweight state-action mapping ensures that the algorithm is highly compatible with low-power embedded chips on autonomous edge devices, facilitating rapid real-time response. Furthermore, the safety-aware policy acquired through training prioritizes operational robustness and social compliance, which are more critical than pure temporal efficiency in modern human-robot co-existence environments. Future work will involve validating the proposed framework within real-world physical robot clusters to evaluate its performance under actual sensor noise and communication latencies in diverse urban landscapes.

Acknowledgement

This work was supported by the Program of National Key R&D Program of China (2023YFC3807600).

References

[1] P. H. C. Morais, K. C. T. Vivaldini, E. R. R. Kato, and R. S. Inoue. Review of Robot Fleet Management. *IEEE Access*, 13:118975–119003, 2025.

[2] Z. Wu, and H. Zhang. Review on Autonomous and Sustainable Urban Mobility Systems: Challenges and Future Directions. *Authorea Preprints*, 2025.

[3] X. Bai, A. Fielbaum, and M. Kronmüller, et al. Group-Based Distributed Auction Algorithms for Multi-Robot Task Assignment. *IEEE Transactions on Automation Science and Engineering*, 20(2):1292–1303, 2023.

[4] M. De Ryck, D. Pissoort, T. Holvoet, and E. Demeester. Decentral task allocation for industrial AGV-systems with routing constraints. *Journal of Manufacturing Systems*, 62:135–144, 2022.

[5] A. Hertle, and B. Nebel. Efficient auction based

coordination for distributed multi-agent planning in temporal domains using resource abstraction. In *Proceedings of Joint German/Austrian Conference on Artificial Intelligence*, pages 86-98, 2018.

[6] Y. Carreno, È. Pairet, Y. Petillot, and R. P. Petrick. Task allocation strategy for heterogeneous robot teams in offshore missions. In *Proceedings of the 19th International Conference on Autonomous Agents and Multi Agent Systems*, pages 222-230, 2020.

[7] M. Mapes, and Y. Xu. Row Allocation Negotiation for a Fleet of Strawberry Harvesting Robots. *ASME Letters in Dynamic Systems and Control*, 2(3), 2022.

[8] Y. Carreno, È. Pairet, and Y. Petillot, et al. A decentralised strategy for heterogeneous auv missions via goal distribution and temporal planning. In *Proceedings of the international conference on automated planning and scheduling*, 30: 431-439, 2020.

[9] J. Han, and H. Kim. MARL-Driven Decentralized Crowdsourcing Logistics for Time-Critical Multi-UAV Networks. *Electronics*, 15(2), 2026.

[10] G. d'Eon, N. Newman, and K. Leyton-Brown. Understanding Iterative Combinatorial Auction Designs via Multi-Agent Reinforcement Learning. in *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 1102–1130, 2024.

[11] Q. Yang. Hierarchical Needs-Driven Self-Adaptive Multi-Agent Systems: From Individual Behaviors to Swarm Intelligence. *Available at SSRN 4334527*, 2023.

[12] S. Devlin, L. Yliniemi, D. Kudenko, and K. Tumer. Potential-based difference rewards for multiagent reinforcement learning. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 165–172, 2014.

[13] S. F. Chaitra. Collision-Free Coverage Control of Multi-Agent Systems. 2025.