Decentralized Multi-Robot Task Allocation and Navigation in Complex Environments (DC-MRTA)

Sharmad Kalpande (22166), Anirudha Patil (22234), Saurabh Shirke (22297), Rushikesh Bhosale (22078)

System Model

- n holonomic robots in $W \subset \mathbb{R}^2$
- Each agent A_i has:
  - Position $\mathbf{p}_i$, velocity $\mathbf{v}_i$
  - Neighborhood set $\mathcal{N}_i = \{j \mid \|\mathbf{p}_j - \mathbf{p}_i\| \leq r_{\text{sense}}\}$
- Tasks defined as tuples:

$$\mathcal{T}_i = (\mathbf{o}_i, \mathbf{d}_i, k_i, l_i)$$

  where:
  - $\mathbf{o}_i$: origin (pickup)
  - $\mathbf{d}_i$: destination (dropoff)
  - k_i: distance to robot
  - l_i: task length

Optimization Objective

Primary Objective

Minimize total travel distance:

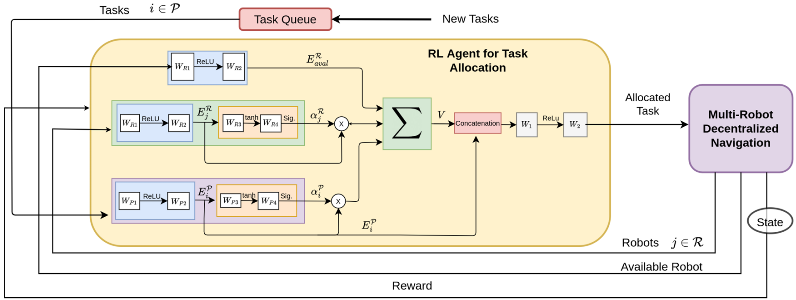
$$\min \sum_{i=1}^{n} \|\mathbf{g}_i - \mathbf{p}_i\|$$

where $\mathbf{g}_i \in \{\mathbf{o}_i, \mathbf{d}_i\}$ is current goal

Constraints

$$\begin{aligned} A_i \cap A_j &= \emptyset, && \forall j \in \mathcal{N}_i \\ A_i \cap O_k &= \emptyset, && \forall k \in \{1, ..., m\} \end{aligned}$$

System Architecture Overview



- **High-Level**: RL-based task allocator
- **Low-Level**: ORCA navigation controller
- **Coupling**: Reward feedback from navigation to allocation

Velocity Obstacle Formulation

VO Definition

$$VO_{i|j}^{\tau} = \{\mathbf{v} \mid \exists t \in [0, \tau], t\mathbf{v} \in D(\mathbf{p}_j - \mathbf{p}_i, r_i + r_j)\}$$

where:

- τ: time horizon
- $D(\cdot)$: disk of radius $r_i + r_j$

Collision Condition

$$\mathbf{v}_i^{opt} - \mathbf{v}_j^{opt} \in VO_{i|j}^{\tau} \Rightarrow \text{Collision}$$

ORCA Formulation

ORCA Half-Plane

$$ORCA^{\tau}_{i|j} = \left\{ \mathbf{v} \mid \left(\mathbf{v} - \left(\mathbf{v}_i^{opt} + \frac{1}{2}\mathbf{u} \right) \right) \cdot \mathbf{n} \geq 0 \right\}$$

where:

- $\mathbf{u}$: minimum avoidance velocity
- $\mathbf{n}$: normal to VO boundary

Velocity Selection

$$\mathbf{v}_i^{new} = \underset{\mathbf{v} \in ORCA^{\tau}_{i|j}}{\operatorname{argmin}} \|\mathbf{v} - \mathbf{v}_i^{pref}\|$$

MDP Formulation

MDP Components

$$\mathcal{M} = (S, A, R, P, \gamma)$$

- State space S:

$$\mathbf{S} = \{(\mathbf{p}_j, r_j)_{\forall j \in \mathcal{R}}, (\mathbf{o}_i, \mathbf{d}_i, k_i, l_i)_{\forall i \in \mathcal{P}}, j_{sel}\}$$

- Action space A: Task selection
- Reward R: $-\text{Time}(\mathbf{o}_i, \mathbf{p}_i)$

Policy Architecture

Neural Network Policy

$$\pi_\theta(a|s) = \text{softmax}(f_\theta(s))$$

where f_θ is attention-based DNN with:

- Input: Full state $\mathbf{S}$
- Output: Task selection probabilities

Attention Mechanism

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

applied to:

- Q: Current robot query
- K: Task/Robot keys
- V: Task values

Reward Formulation

Decoupled Reward

$$R_{\text{dec}} = \begin{cases} 1 & \text{task done} \\ 0 & \text{otherwise} \end{cases}$$

Coupled Reward

$$R_{\text{coup}} = -\text{Time}(\mathbf{o}_i, \mathbf{p}_i) - \lambda c$$

where:

- c: collision penalty
- λ: weighting factor

Performance Metrics

Key Results

- Makespan improvement:

$$\frac{T_{\text{baseline}} - T_{\text{ours}}}{T_{\text{baseline}}} = 14\%$$

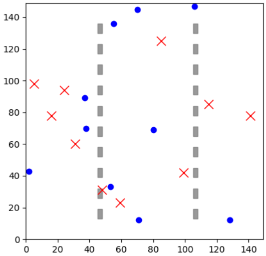
- Collision reduction:

$$\frac{C_{\text{baseline}} - C_{\text{ours}}}{C_{\text{baseline}}} = 40\%$$

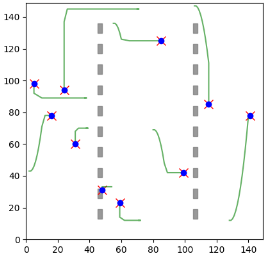
- Scalability:

$$\text{Time} \sim O(n^{1.2}) \text{ for } n \leq 1000 \text{ robots}$$

Visual Results



(a) Initial configuration



(b) Task completion

Figure: 10-robot scenario

Key Equations Summary

- **Navigation**:

$$ORCA^{\tau}_{i|j} = \{\mathbf{v} \mid (\mathbf{v} - \mathbf{v}^*_i) \cdot \mathbf{n} \geq 0\}$$

- **Task Allocation**:

$$\pi_\theta(a|s) = \text{softmax}(f_\theta(s))$$

- **Reward**:

$$R = -\text{Time}(\mathbf{o}_i, \mathbf{p}_i)$$

- **Performance**:

$$\Delta T = 14\%, \Delta C = 40\%$$