Enabling Symbiosis in Multi-Robot Systems through Multi-Agent Reinforcement Learning

Xuezhi Niu, Natalia Calvo Barajas & Didem Gürdür Broo

Cyber-Physical Systems Lab, Department of Information Technology,
Uppsala University

May 13, 2025

IEEE Industrial Electronics Society

UPPSALA UNIVERSITET

CPS-LAB

Agenda

Introduction & Motivation

Autonomous Mobile Robots (AMRs) working in a warehouse. Image source: Wetuc

- **Cyber-Physical System (CPS)** are increasingly deployed across domains; multi-robot systems are a clear instance of this trend [1].
- Autonomous robots are increasingly deployed in warehouse logistics, offering a concrete and scalable example of **multi-robot systems**.
- **Extended types of robots** these systems are often designed in isolation, hindering their ability to act as holistic ecosystem [2].

In multi-robot systems, robots acting in isolation lead to:

- Idle robots
- Resource conflicts (e.g., charging stations)
- Inefficient energy use and delays

Challenges

- **Warehouse robots operate autonomously** Each robot makes local decisions based on its own task queue and battery.
- **But pure independence causes inefficiencies**
  - Charging contention: multiple robots queue for a limited number of stations.
  - Imbalance: some robots do most of the work; others conserve too much energy.
- **Centralized MARL suffers from the curse of dimensionality** High training cost and poor execution-time adaptability under partial observability.

We ask: Can minimal cooperation improve coordination in decentralized multi-robot systems?

Contributions

- **Symbiotic MARL with bio-inspired coordination:** We develop a novel symbiotic MARL architecture that enhances multi-robot collaboration by incorporating ecological principles into the learning process
- **Case study in warehouse logistics:** Simulated warehouse experiments demonstrate measurable gains in system performance (10.7%) and resource utilization (13.81%) compared to non-symbiotic baselines.

Symbiosis!

What is Symbiosis?

A biological relationship where two or more organisms interact for continuous existance, including mutualism, commensalism, and parasitism.

- Mycorrhizal networks between trees and fungi — sharing resources and information to support collective survival [3]

Mycorrhizal networks between trees and fungi. Image source: Rainbo

Focus on **mutualism**: Agents shares critical information to support collective behaviors [3].

Symbiosis into MARL

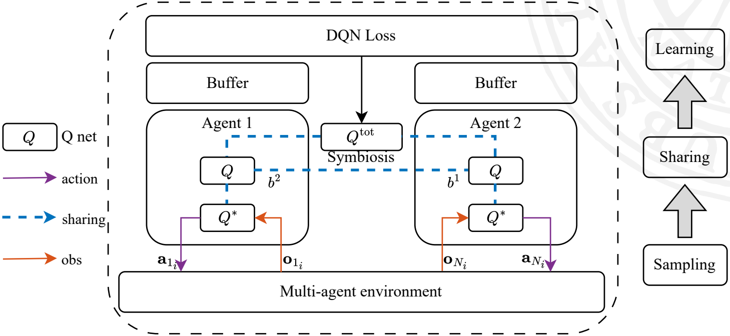

Figure 1: Agents share battery information through symbiosis connections (blue dashed lines) while maintaining individual Q-networks for local decision making. The framework integrates sampling from the environment (orange arrows), sharing of symbiotic information, and learning through DQN loss computation. Q and Q* represent online and target networks respectively, with individual buffers for experience replay.

Algorithm

Algorithm 1 MARL Training with Battery Symbiosis

Require: Replay buffer $\mathcal{D}$, learning rate α, discount factor γ, soft update parameter τ

```
Initialize online Q-networks $Q^i$ and target networks $Q^{*i}$ for all agents $i = 1, \ldots, N$
while training not converged do
    Sample batch $(s_t, a_t, r_t, s_{t+1}, a_{t+1}, b_t)$ from $\mathcal{D}$
    for each agent $i = 1 \ldots N$ do
        Augment local state:
        $\tilde{s}_t^i \leftarrow [s_t^i, b_t^1, \ldots, b_t^{i-1}, b_t^{i+1}, b_t^n]$
        Compute online Q-value:
        $Q^i(\tilde{s}_t^i, a_t^i) \leftarrow Q_t^i(\tilde{s}_t^i, a_t^i) + \alpha(r_t + \gamma \max_{a_t^i} Q_t^{*i}(s_t^i, a_t^i) - Q_t^i(\tilde{s}_t^i, a_t^i))$
    end for
    Compute total Q-value:
    $Q^{\text{tot}}(\tilde{s}_t, a_t) \leftarrow \sum_{i=1}^{N} Q^i(\tilde{s}_t^i, a_t^i)$
    Compute TD target:
    $y_t^{\text{tot}} \leftarrow r_t + \gamma \max_{\mathbf{a}_{t+1}} Q^{\text{tot}}(s_{t+1}, \mathbf{a}_{t+1})$
    Compute loss: $L \leftarrow (y_t^{\text{tot}} - Q^{\text{tot}}(s_t, \mathbf{a}_t))^2$
    Backpropagate loss and update $Q^i$ parameters for all agents
    Perform soft update of target networks:
    for each agent $i = 1 \ldots N$ do
        $Q^{*i} \leftarrow \tau Q^i + (1 - \tau) Q^{*i}$
    end for
end while
```

Training Results

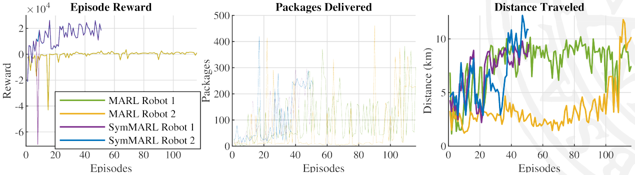

Figure 2: Training performance across three metrics: (a) Cumulative reward per episode, (b) number of packages delivered, and (c) distance traveled. Symbiotic MARL shows faster convergence and more stable learning compared to the non-symbiotic baseline.

- Non-symbiotic MARL struggles to converge due to limited context and poor coordination.
- Symbiotic MARL, enabled by minimal battery state sharing, converges faster and yields better task and energy metrics.

$$r^{i}_{\text{local}}(t) = \epsilon_1 \cdot p_i(t) - \epsilon_2 \cdot e^{20\cdot(0.1-b_i(t))} \cdot \mathbb{1}(b_i(t) < 0.1)$$

$$r_{\text{global}}(t) = T_{\text{total}} - \frac{1}{N}\sum_{i=1}^{N}(T_i - \bar{T}) - 10 \cdot \mathbb{1}(n_a(a_i(t)))$$

Results

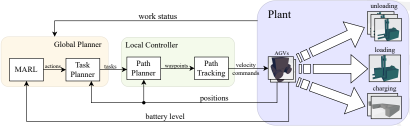

Figure 3: System architecture of the warehouse setup, showing the plant (robot dynamics and energy model), local controllers (path planning and execution), and a global controller integrating MARL for coordination, task allocation, and collision avoidance.

10.7% system performance improvement and 13.81% resource utilization efficiency

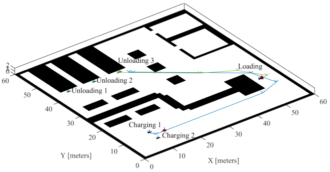

Figure 4: Layout of the simulated warehouse environment (60m × 60m).

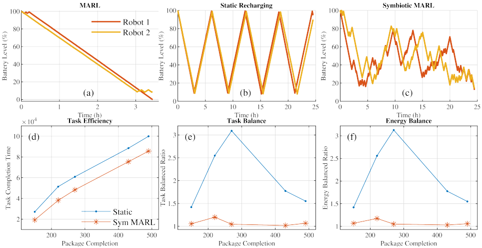

Figure 5: Evaluation of static recharging and MARL with and without symbiosis.

Conclusions & Future Works

Conclusions

- Introduced a bio-inspired Symbiotic MARL framework using battery-state sharing to improve coordination in multi-robot systems.
- Demonstrated performance gains in task completion time (**10.7%**) and energy balance (**13.8%**) in a warehouse simulation.
- Validated that ecological symbiosis principles can address coordination and interoperability challenges in CPS.

Future Works

- Scale to larger heterogeneous fleets and more complex environments.
- Compare against additional MARL baselines using VDN, QMIX, and actor-critic methods (e.g., MADDPG).
- Apply actor-critic methods (e.g., MADDPG) to support group-level symbiosis with decentralized execution.

References

Back to start

[1] V. Lesch, M. Züfle, A. Bauer, L. Iffländer, C. Krupitzer, and S. Kounev, "A literature review of iot and cps—what they are, and what they are not," *Journal of Systems and Software*, vol. 200, p. 111631, 2023.

[2] D. Gürdür and F. Asplund, "A systematic review to merge discourses: Interoperability, integration and cyber-physical systems," *Journal of Industrial information integration*, vol. 9, pp. 14–23, 2018.

[3] S. W. Simard, K. J. Beiler, M. A. Bingham, J. R. Deslippe, L. J. Philip, and F. P. Teste, "Mycorrhizal networks: mechanisms, ecology and modelling," *Fungal Biology Reviews*, vol. 26, no. 1, pp. 39–60, 2012.