A Scalable Network-Aware MARL Framework for Distributed Converter-based MG Voltage Control

Han Xu, Tsinghua University

Guannan Qu, Carnegie Mellon University

NEURAL INFORMATION PROCESSING SYSTEMS

1. MOTIVATION & GOAL

Background:

- A large number of Distributed Generations (DGs) in Microgrids (MGs)
- Distributed voltage control is required to avoid communication overheads while maintaining the safe operation of MGs

Motivation:

- Multi-agent reinforcement learning (MARL) has great potential for realizing distributed voltage control
- Traditional Centralized Training and Decentralized Execution (CTDE) framework suffers from the curse of dimensionality (e.g., poor scalability)

Goals:

- Propose a **scalable** MARL framework that can realize the distributed voltage control of the **large-scale** MGs

2. PROBLEM FORMULATION

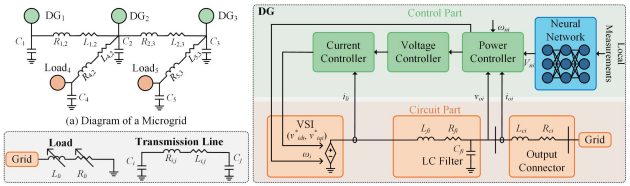


Diagram of DGs, loads and transmission lines

Key Components in MGs:

- **Loads**, time-varying impedances
- **DGs**, inverter-based sources
- **Transmission lines**, linear impedances

Operation Challenges:

- **Stochasticity**: The power consumed by loads and generated by renewable energy resources varies stochastically
- **Voltage fluctuation**: Unbalanced power between loads and DGs result in voltage fluctuation

Control Task:

- DGs, without communications, collaborate together to balance the power of DGs and loads to maintain the voltage within the safe range

3. METHOD

Main Contribution:

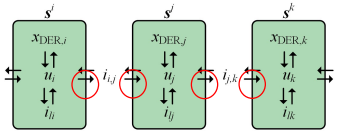
- An innovative formulation of the voltage control problem that preserves the **network structure**
- An innovative actor-critic-based MARL framework that utilizes the **network structure** to achieve scalability
- Extensive experiments to validate the effectiveness and scalability of the proposed approach.

- **Networked Markov Decision Process (MDP):** The MG voltage control problem is formulated as a networked MDP as in [1]. Consider a MG with N DGs (e.g., agents), then we claim the state transition of DGs follows a network structure:

$$P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t) = \prod_{i=1}^{N} P(\mathbf{s}_{i,t+1}|\mathbf{s}_{\mathcal{N}_i,t}, \mathbf{a}_{i,t}), \quad (1)$$

where N_i denotes the neighborhood of i (including i itself). This relation means the next individual state is independently generated and is only dependent on neighbors.

- **Justification:** After examining the state variables relations within the system, the third-order discretized results can satisfy (1).



Relations of state variables

- **The (c,ρ)-exponential decay property:** For networked MDP, for a specific agent i, the impact of other agents' observations and actions on its individual Q-function lessens with increasing topological distance which can be described by the following inequality

$$|Q_i^{\pi}(\mathbf{o}_{\mathcal{N}_i^\kappa}, \mathbf{o}_{\mathcal{N}_i^\kappa \setminus i}, \mathbf{a}_{\mathcal{N}_i^\kappa}, \mathbf{a}_{\mathcal{N}_i^\kappa \setminus i}) - Q_i^{\pi}(\mathbf{o}_{\mathcal{N}_i^\kappa}, \bar{\mathbf{o}}_{\mathcal{N}_i^\kappa \setminus i}, \mathbf{a}_{\mathcal{N}_i^\kappa}, \bar{\mathbf{a}}_{\mathcal{N}_i^\kappa \setminus i})| \leq c\rho^{\kappa+1} \quad (2)$$

where N_i^k denotes the k-hop neighborhood of i (including i itself).

- **Actor-Critic-based Scalable Network-Aware framework:**
  - Critics that take only observations and actions of k-hop neighbors:

$$\min J_{Q_i}(\theta_i) = \mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[\frac{1}{2}(\hat{Q}_{\theta_i}(\mathbf{o}_{\mathcal{N}_i^\kappa,t}, \mathbf{a}_{\mathcal{N}_i^\kappa,t}) - \hat{y})^2], \quad (3)$$

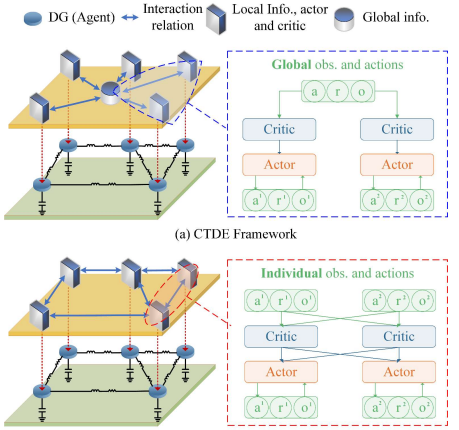
where

$$\hat{y} = r_{i,t} + \gamma[\bar{Q}_{\bar{\theta}_i}(\mathbf{o}_{\mathcal{N}_i^\kappa,t+1}, \mathbf{a'}_{\mathcal{N}_i^\kappa,t+1}) - \alpha \log(\pi_{\phi_i}(\mathbf{a'}_{i,t+1}|\mathbf{o}_{i,t+1}))], \quad (4)$$

  - Actors that are guided by critics of its k-hop neighbors:

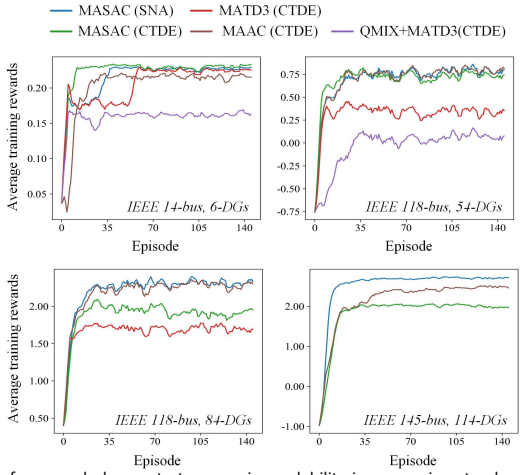
$$\min \mathbb{E}_{(\mathbf{o}_t)\sim\mathcal{D}}[\mathrm{D}_{\mathrm{KL}}(\pi_{\phi_i}(\cdot|\mathbf{o}_t^i)||\frac{\exp(\frac{1}{N}\sum_{j\in\mathcal{N}_i^\kappa}\hat{Q}_j^{\pi}(\mathbf{o}_t^{\mathcal{N}_j/i}, \cdot, \mathbf{a}_t^{\mathcal{N}_j/i}))}{Z_i(\frac{1}{N}\sum_{j\in\mathcal{N}_i^\kappa}\hat{Q}_j^{\pi}(\mathbf{o}_t^{\mathcal{N}_j/i}, \cdot, \mathbf{a}_t^{\mathcal{N}_j/i}))})] \quad (5)$$

  - Rationale: Based on exponential decay property, truncated Q-functions can accurately estimate the expected individual rewards without taking all infos.



Centralized Training Decentralized Execution (CTDE) vs the proposed Scalable Network Aware (SNA) framework

4. RESULT



- Our framework demonstrates superior scalability in comparison to algorithms based on Centralized Training Decentralized Execution (CTDE), effectively achieving voltage control in a Microgrid (MG) with **114** DGs. (The SATO algorithm, as reported in [3], achieved voltage control in a Microgrid with **40** DGs.)

References

[1] G. Qu, A. Wierman, and N. Li, "Scalable Reinforcement Learning for Multi-Agent Networked Systems," Oct. 2021.

[2] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," Aug. 2018.

[3] D. Chen et al., "PowerNet: Multi-Agent Deep Reinforcement Learning for Scalable Powergrid Control," IEEE Trans. Power Syst., vol. 37, no. 2, pp. 1007–1017, Mar. 2022, doi: 10.1109/TPWRS.2021.3100898.