

Safe Decentralized Operation of EV Virtual Power Plant with Limited Network Visibility via Multi-Agent Reinforcement Learning

Chenghao Huang¹, Jiarong Fan¹, Weiqing Wang¹, Hao Wang^{1,2*}

¹Department of Data Science and AI, Faculty of IT, Monash University, Australia

²Monash Energy Institute, Monash University, Australia

Abstract—As power systems advance toward net-zero targets, behind-the-meter renewables are driving rapid growth in distributed energy resources (DERs). Virtual power plants (VPPs) increasingly coordinate these resources to support power distribution network (PDN) operation, with EV charging stations (EVCSs) emerging as a key asset due to their strong impact on local voltages. However, in practice, VPPs must make operational decisions with only partial visibility of PDN states, relying on limited, aggregated information shared by the distribution system operator. This work proposes a safety-enhanced VPP framework for coordinating multiple EVCSs under such realistic information constraints to ensure voltage security while maintaining economic operation. We develop Transformer-assisted Lagrangian Multi-Agent Proximal Policy Optimization (TL-MAPPO), in which EVCS agents learn decentralized charging policies via centralized training with Lagrangian regularization to enforce voltage and demand-satisfaction constraints. A transformer-based embedding layer deployed on each EVCS agent captures temporal correlations among prices, loads, and charging demand to improve decision quality. Experiments on a realistic 33-bus PDN show that the proposed framework reduces voltage violations by approximately 45% and operational costs by approximately 10% compared to representative multi-agent DRL baselines, highlighting its potential for practical VPP deployment.

Index Terms—Virtual power plant, distribution network, electric vehicle charging station, voltage safety, multi-agent reinforcement learning.

I. INTRODUCTION

Decarbonization commitments and the global push toward net-zero energy systems are accelerating the integration of renewable resources into distribution networks [1]. As behind-the-meter PV adoption increases, distributed energy resources (DERs), such as rooftop PV, residential batteries, and electric vehicles (EVs) [2], are increasingly coordinated through virtual power plants (VPPs) to provide flexibility services and support grid stability [3].

However, the rapid growth of DERs introduces significant operational challenges for power distribution networks (PDNs). High penetrations of distributed PV can cause reverse power flows and voltage rise, while uncoordinated EV charging can create peak load surges and voltage violations

[4]. Among various DERs, EV charging stations (EVCSs) stand out as a priority for VPP coordination because their high-throughput, long-dwell, and highly flexible yet spatially concentrated charging demand can induce substantial voltage deviations in PDNs.

Recent research on EV charging coordination has ranged from centralized rule-based control and optimization [5], [6] to multi-agent reinforcement learning (MARL) approaches that improve scalability and adaptability under uncertainty [7], [8]. Ensuring grid safety, however, remains challenging. Existing efforts include embedding operational limits in the action space [9], reward shaping to penalize constraint violations [10], [11], and Lagrangian regularization for principled constraint handling [12], [13]. Despite the progress, safe mechanisms within MARL remain relatively underexplored in power system applications.

Furthermore, VPPs seldom have full access to PDN states or topology due to privacy, regulatory, and cybersecurity constraints. This limited visibility makes it difficult to anticipate how local charging actions propagate through electrical coupling to affect voltages and network constraints elsewhere, leading to potentially unsafe or inefficient decisions. Therefore, two key challenges remain insufficiently addressed in existing MARL-based EVCS coordination: limited safety guarantees during learning and deployment, which hinder principled voltage-limit enforcement, and the common assumption of full grid-state visibility, whereas real VPP-DSO interactions offer only partial, privacy-preserving indicators. These limitations restrict the practical deployment of MARL in real distribution networks.

In this work, we study the problem of coordinating multiple EVCSs under the realistic constraint that a VPP only receives partial and privacy-preserving grid indicators from the DSO. We integrate transformer-based temporal capturing with multi-agent reinforcement learning and principled safety regularization, forming a framework tailored for EVCS coordination under partial PDN visibility. The main contributions of this work are summarized as follows:

- We formalize a realistic VPP-DSO coordination setting where multiple EVCSs operate under partial PDN visibility, capturing core challenges in achieving economic

*Corresponding author: Hao Wang (hao.wang2@monash.edu).

This work was supported in part by the Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) under Grant DE230100046.

charging and voltage safety amid uncertainties in EV demand, PV output, and electricity prices.

- We propose TL-MAPPO, a safety-enhanced multi-agent RL framework that learns decentralized EVCS policies under centralized training. It integrates Lagrangian regularization for enforcing safety constraints and a transformer-based embedding layer for improved temporal context under limited grid visibility.
- Experiments on the IEEE 33-bus system show that TL-MAPPO reduces voltage violations by about 45% and operational costs by about 10% compared with representative state-of-the-art MARL baselines, demonstrating its potential for safe and practical EVCS coordination.

II. SYSTEM MODEL

In a PDN operated by a DSO, a VPP coordinates multiple EVCSs deployed at different electric buses. The VPP manages EV charging activities and rooftop photovoltaic (PV) usage to minimize operational costs while maintaining voltage safety and fulfilling user demand requirements. An overview of the system configuration is shown in Fig. 1. Detailed PDN information (e.g., voltage and power at nodes) is generally available to the VPP, based on pilot projects [14], and we **assume** that the DSO grants each VPP-managed EVCS access to partial PDN information corresponding to its local neighborhood, including the voltage magnitudes and aggregated loads of its hosting bus and adjacent buses within one hop. This limited visibility reflects realistic communication and privacy constraints between the DSO and the VPP.

A. PDN

The PDN consists of N buses connected through distribution lines and operates over a discrete time horizon $\mathcal{T} = 1, 2, \dots, T$. At each bus $i \in \mathcal{N}$ and time step $t \in \mathcal{T}$, the active power injection is denoted by $p_{i,t}$, and the voltage magnitude by $v_{i,t}$. Give the allowable voltage magnitude range, the voltage of node i is constrained as

$$V^{\min} \leq v_{i,t} \leq V^{\max}, \quad \forall i \in \mathcal{N}, t \in \mathcal{T}. \quad (1)$$

B. EVCS and Neighborhood Nodes

The VPP manages a set of EVCSs $\mathcal{K} = 1, \dots, K$, each equipped with rooftop PV generation and a set of chargers $\mathcal{C}_k = 1, \dots, C_k$. Each EVCS k is physically connected to one bus of the PDN, represented by a binary deployment matrix $\mathbf{K} \in \{0, 1\}^{N \times K}$, where $K_{ik} = 1$ if EVCS k is located at bus i , and $K_{ik} = 0$ otherwise,

$$\sum_{k \in \mathcal{K}} K_{ik} \leq 1, \quad \forall i \in \mathcal{N}. \quad (2)$$

For each bus $i \in \mathcal{N}$, let $\mathcal{N}_i^{(1)}$ denote its 1-hop neighbor set, which contains all buses directly connected to bus i through distribution lines:

$$\mathcal{N}_i^{(1)} = \{j \in \mathcal{N} \setminus \{i\} \mid (i, j) \in \mathcal{L}\}, \quad (3)$$

where $\mathcal{L}$ is the set of distribution lines connecting pairs of buses.

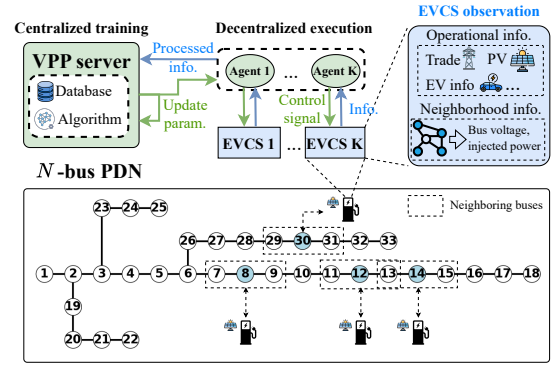


Fig. 1. System model of decentralized EVCS coordination under a VPP's centralized training in a PDN.

Given the EVCS deployment matrix $\mathbf{K} \in \{0, 1\}^{N \times K}$, the 1-hop neighbor bus set associated with EVCS k (located at bus i where $K_{ik} = 1$) is defined as

$$\mathcal{N}_k^{(1)} = \mathcal{N}_i^{(1)} = \{j \in \mathcal{N} \setminus \{i\} \mid (i, j) \in \mathcal{L}, K_{ik} = 1\}. \quad (4)$$

This set represents all the buses physically adjacent to the bus hosting EVCS k , whose voltage and injection information can be partially accessed under the DSO-VPP collaboration assumption.

On the EV side, each charger must also respect the physical charging requirements imposed by individual EVs, characterized by their arrival-departure windows and target SoC trajectories. Specifically, $\text{SoC}_{c_k}^{\text{target}}(t)$ is the expected SoC trajectory between the arrival and departure times of each EV:

$$\text{SoC}_{c_k}^{\text{target}}(t) = \text{SoC}_{c_k}^{\text{arr}} + \frac{t - T_{c_k}^{\text{arr}}}{T_{c_k}^{\text{dep}} - T_{c_k}^{\text{arr}}} (\text{SoC}_{c_k}^{\text{dep}} - \text{SoC}_{c_k}^{\text{arr}}). \quad (5)$$

Each charger satisfies the operational constraints below:

$$0 \leq p_{c_k}^{\text{ch}}(t) \leq P_{c_k}^{\text{ch,max}}, \quad 0 \leq p_{c_k}^{\text{dis}}(t) \leq P_{c_k}^{\text{dis,max}}, \quad (6a)$$

$$p_{c_k}^{\text{ch}}(t)p_{c_k}^{\text{dis}}(t) = 0, \quad (6b)$$

$$\text{SoC}_{c_k}^{\min} \leq \text{SoC}_{c_k}(t) \leq \text{SoC}_{c_k}^{\max}, \quad (6c)$$

$$\Delta \text{SoC}_{c_k}(t) = \eta^{\text{ch}} p_{c_k}^{\text{ch}}(t) - \frac{1}{\eta^{\text{dis}}} p_{c_k}^{\text{dis}}(t). \quad (6d)$$

C. Objective

The objective of the VPP is to coordinate EVCS operations to minimize overall costs while satisfying physical and operational constraints. At each time t , the k -th EVCS controls the charging and discharging power of its chargers, $\{p_{c_k}^{\text{ch}}(t), p_{c_k}^{\text{dis}}(t)\}_{c_k=1}^{C_k}$, subject to charger and battery limits. The optimization objective is formulated as:

$$\min_{\mathbf{p}^{\text{ch}}, \mathbf{p}^{\text{dis}}} \sum_{t \in \mathcal{T}} \left[\sum_{k \in \mathcal{K}} [\beta_1 f_k^{\text{TR}}(t) + \beta_2 f_k^{\text{DG}}(t) + \beta_3 f_k^{\text{DS}}(t)] + \beta_4 f^{\text{VT}}(t) \right], \quad (7)$$

where $f_k^{\text{TD}}(t)$ and $f_k^{\text{DG}}(t)$ denote the energy trading and battery degradation costs (also called cycling overhead in Section IV), $f^{\text{VT}}(t)$ quantifies total voltage violations, $f_k^{\text{DS}}(t)$ represents the charging dissatisfaction penalty of EVCS k , and $\beta_1 - \beta_4$ are

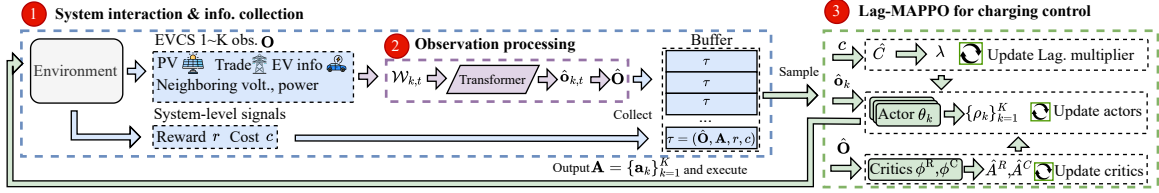


Fig. 2. The developed TL-MAPPO for safe EVCS coordination under partial PDN visibility, primarily built on MARL and Transformer.

the coefficients of each term. Specifically, the cost components are defined as

$$f_k^{\text{TD}}(t) = \begin{cases} \lambda_t^{\text{buy}} p_k^{\text{TD}}(t), & p_k^{\text{TD}}(t) > 0, \\ \lambda_t^{\text{sell}} p_k^{\text{TD}}(t), & \text{otherwise,} \end{cases} \quad (8)$$

$$f_k^{\text{DG}}(t) = \alpha_e \sum_{c_k \in \mathcal{C}_k} ([p_{c_k}^{\text{ch}}(t)]^2 + [p_{c_k}^{\text{dis}}(t)]^2), \quad (9)$$

$$f_k^{\text{VT}}(t) = \sum_{i \in \mathcal{N}} ([v_{i,t} - V^{\text{max}}]^+ + [V^{\text{min}} - v_{i,t}]^+), \quad (10)$$

$$f_k^{\text{DS}}(t) = \sum_{c_k \in \mathcal{C}_k} [\text{SoC}_{c_k}^{\text{target}}(t) - \text{SoC}_{c_k}(t)]^+, \quad (11)$$

where $[x]^+ = \max(0, x)$, and $\text{SoC}_{c_k}^{\text{target}}(t)$ is the EV c_k 's target SoC at time t , linearly interpolated from arrival to departure. The power traded between the EVCS k and the grid is

$$p_k^{\text{TD}}(t) = \sum_{c_k \in \mathcal{C}_k} [p_{c_k}^{\text{ch}}(t) - p_{c_k}^{\text{dis}}(t)] - p_{k,t}^{\text{PV}}. \quad (12)$$

III. METHODOLOGY

This section presents the overall framework of the proposed TL-MAPPO method, illustrated in Fig. 2 which integrates three tightly coupled methodological components for safe EVCS coordination under partial network visibility. First, the EVCS coordination problem is formulated as a partially observable constrained Markov decision process (PO-CMDP), which captures long-horizon decision making with limited local observations and safety constraints. Based on this formulation, each EVCS agent collects local measurements and operational information to construct a temporal observation sequence. Second, a Transformer-based observation encoder processes this sequence to extract compact temporal representations, which serve as the input to decentralized control policies. Third, these encoded observations are used by a Lagrangian MAPPO algorithm, where decentralized actors are trained with centralized critics to jointly optimize economic performance and enforce voltage and demand-satisfaction constraints. During execution, each EVCS operates independently using its local policy, while safety is ensured through the learned Lagrangian regularization.

A. PO-CMDP Formulation

The coordination of EVCSs under partial visibility and safety constraints is modeled as a PO-CMDP $\mathcal{M} = \langle \mathcal{K}, \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathbb{P}, \mathbb{R}, \mathbb{C}, \gamma \rangle$, where $\mathcal{K}$ is the set of EVCS agents, $\mathcal{S}$ the global state, $\mathcal{O} = \prod_k \mathcal{O}_k$ the joint observations, $\mathcal{A} = \prod_k \mathcal{A}_k$ the joint actions, $\mathbb{P}$ the transition, $\mathbb{R}$ the reward,

$\mathbb{C}$ the safety cost, and γ the discount factor. At time t , EVCS k observes

$$\mathbf{o}_k(t) = \left[\underbrace{[p_{i,t}, v_{i,t}]_{i \in \mathcal{N}_k^{(1)}}}_{\text{Neighborhood}}, \underbrace{[p_{k,t}^{\text{PV}}, \lambda_t^{\text{buy}}, \lambda_t^{\text{sell}}]}_{\text{PV and prices}}, \underbrace{[T_{c_k}^{\text{dep}} - t, \text{SoC}_{c_k}^{\text{dep}} - \text{SoC}_{c_k}(t)]_{c_k \in \mathcal{C}_k}}_{\text{EV info}} \right]. \quad (13)$$

The action vector $\mathbf{a}_k(t) = [a_{c_k}(t)]_{c_k \in \mathcal{C}_k}^T$ denotes charging/discharging power, bounded by $-P_{c_k}^{\text{dis}, \text{max}} \leq a_{c_k}(t) \leq P_{c_k}^{\text{ch}, \text{max}}$ and subject to SoC/exclusivity constraints. The reward follows the economic objective:

$$r(t) = \sum_{k \in \mathcal{K}} [\beta_1 f_k^{\text{TD}}(t) + \beta_2 f_k^{\text{DG}}(t)]. \quad (14)$$

The safety cost aggregates voltage violations and demand dissatisfaction:

$$c(t) = \beta_3 \sum_{k \in \mathcal{K}} f_k^{\text{DS}}(t) + \beta_4 f_k^{\text{VT}}(t). \quad (15)$$

B. Transformer-based Observation Processing

We adopt a Transformer [15] to capture temporal correlations among the input features. For each EVCS k , an observation window $\mathcal{W}_k(t) = [\mathbf{o}_k(t - t' + 1); \dots; \mathbf{o}_k(t)]$ is encoded by a multi-head self-attention encoder:

$$\mathbf{h}_k^{\text{Trans}}(t) = g_k^{\text{Trans}}(\mathcal{W}_{k,t}), \quad (16)$$

where $g_k^{\text{Trans}}(\cdot)$ denotes the Transformer encoder that extracts temporal representations from the observation window $\mathcal{W}_{k,t}$. Next, the processed observation fed to MARL is $\hat{\mathbf{o}}_k(t) = [\mathbf{h}_k^{\text{Trans}}(t)]$, with $\hat{\mathbf{O}}(t) = [\hat{\mathbf{o}}_1(t), \dots, \hat{\mathbf{o}}_K(t)]$.

C. Lag-MAPPO for EVCS Coordination

We adopt a Lag-MAPPO framework, where decentralized actors $\{\pi_{\theta_k}\}$ are trained with two centralized critics V_{ϕ^R} and V_{ϕ^C} . Each EVCS agent samples its control action from

$$\mathbf{a}_k(t) \sim \pi_{\theta_k}(\mathbf{a}_k(t) | \hat{\mathbf{o}}_k(t)). \quad (17)$$

Given trajectories $\tau(t) = [\hat{\mathbf{O}}(t), \mathbf{A}(t), r(t), c(t)]$, the normalized advantages for reward and cost are

$$\hat{A}^R(t) = \sum_{l=0}^{T-t} \gamma^l r(t+l) - V_{\phi^R}(\hat{\mathbf{O}}(t)), \quad (18a)$$

$$\hat{A}^C(t) = \sum_{l=0}^{T-t} \gamma^l c(t+l) - V_{\phi^C}(\hat{\mathbf{O}}(t)), \quad (18b)$$

where V_{ϕ^R} and V_{ϕ^C} estimate the expected reward and cost under current policy, respectively.

The actor optimizes the clipped PPO objective with Lagrangian shaping:

$$\hat{A}^{\text{Lag}}(t) = \hat{A}^R(t) - \lambda \hat{A}^C(t), \quad (19a)$$

$$L_k^{\text{Act}} = \mathbb{E} \left[-\min\{\rho_k(t), \sigma[\rho_k(t), 1 - \epsilon, 1 + \epsilon]\} \cdot \hat{A}^{\text{Lag}}(t) \right], \quad (19b)$$

$$\rho_k(t) = \frac{\pi_{\theta_k}(\mathbf{a}_k(t)|\hat{\mathbf{o}}_k(t))}{\pi_{\theta'_k}(\mathbf{a}_k(t)|\hat{\mathbf{o}}_k(t))}, \quad (19c)$$

where $\rho_k(t)$ is the importance ratio between current and old policies, σ is clipping operation, ϵ is the PPO clipping threshold, and λ is the Lagrangian multiplier.

The critic minimizes the squared error of the Lagrangian-shaped return:

$$L^{\text{Critic}} = \mathbb{E} \left[\left(V_{\phi^C}(\hat{\mathbf{O}}(t)) - [r(t) - \lambda c(t)] \right)^2 \right]. \quad (20)$$

The multiplier λ is updated by projected dual ascent:

$$\lambda \leftarrow [\lambda + \eta^{\text{Lag}}(\hat{C}(t) - \bar{C})]^+, \quad (21)$$

where η^{Lag} is the dual learning rate, $\hat{C}(t)$ is the observed cost, $\bar{C}$ is the desired constraint threshold, and $[\cdot]^+$ ensures non-negativity. Finally, parameters θ_k , ϕ^R , and ϕ^C are updated by gradient descent on L_k^{Act} and L^{Critic} , respectively. This yields stable policy improvement while prioritizing voltage and demand safety under partial visibility.

IV. SIMULATIONS

A. Experimental Setup

Four EVCSs, each equipped with ten chargers, are deployed at buses 8, 12, 14, and 30, as shown in Fig. 1. The simulation adopts a 33-bus PDN operated over a one-day horizon (288 steps of 5 minutes each), with EV SoC constrained within [0,80] kWh, charging/discharging rates within [-22,22] kW, and efficiency of 0.95. The network voltage is maintained within the safety range of [0.95,1.05] p.u.

For the numerical simulation, we use daily EV data collected by Caltech [16] and rooftop PV generation data of Solar Home Electricity Dataset [17] provided by Ausgrid's electricity network from July 1, 2010, to June 30, 2013. The wholesale price of AEMO is used for energy transaction [18], and the TOU price is set by the DSO, taking Melbourne as an example. In terms of the EV data, the features of over 900,000 EVs with charging duration shorter than one day (1440 minutes) are included, where most of the EV demands are less than 20 kWh in their charging durations primarily ranged from 10 to 800 minutes. All EVs follow a First-In-First-Out principle.

In terms of the comparison baselines, we evaluate our developed TL-MAPPO against three DRL baselines: ① MAPPO - A centralized-training, decentralized-execution extension of PPO, with good stability and strong performance. ② MATD3 - A multi-agent version of TD3 that mitigates overestimation bias and is effective for continuous control under deterministic policies. ③ MASAC - A multi-agent extension of SAC that

TABLE I
TEST PERFORMANCE OVER 100 INDEPENDENT EPISODES.

Method	Energy cost (\$)	Cycling overhead	Avg. volt. (p.u./5 min)	vio. (p.u./5 min)	Avg. demand dissat. (kWh/EV)
MATD3	144.2 ± 3.8	118.7 ± 2.5%	7.8 ± 0.6 × 10 ⁻³	0.88 ± 0.07	
MAPPO	140.6 ± 4.1	122.4 ± 3.2%	6.1 ± 0.5 × 10 ⁻³	0.76 ± 0.06	
MASAC	148.9 ± 3.9	126.5 ± 2.9%	9.3 ± 0.8 × 10 ⁻³	1.12 ± 0.09	
TL-MAPPO	133.5 ± 3.4	110.2 ± 2.1%	4.2 ± 0.4 × 10⁻³	0.58 ± 0.05	

leverages entropy regularization to enhance exploration and improve learning robustness.

To comprehensively evaluate the performance of the experimented method, we report four key metrics that reflect economic operation, constraint satisfaction, and service quality: ① **Energy cost** (\$): Total operational cost, equal to the cumulative reward over one day. ② **Cycling overhead** (unitless): Ratio of total energy throughput (charge + discharge) to charging demand. ③ **Average voltage violation** (p.u./5 min): Mean per-step voltage violation over all buses. ④ **Average demand dissatisfaction** (kWh/EV): Mean unmet charging demand per EV over one episode.

B. Result Analysis

1) *Overall Performance*: The training curves in Fig. 3 and the results in Table I together show that TL-MAPPO achieves superior learning efficiency, stability, and constraint satisfaction. It yields the lowest daily energy cost (about 125 AUD), whereas baselines remain between 130-150 AUD. TL-MAPPO also exhibits the smallest voltage violations, lowest unmet EV demand, and the most stable cycling overhead, with consistently narrow confidence intervals. Convergence is faster and smoother across all metrics. Table I further confirms its advantage over 100 episodes: energy cost is reduced by up to 10.7%, voltage violations by approximately 45%, and demand dissatisfaction by up to 35%, with the lowest variance among all methods.

Overall, TL-MAPPO strikes a balanced trade-off between economic operation and grid safety, outperforming MARL baselines under partial visibility. These gains stem from adaptive constraint handling via Lagrangian multipliers that balance reward and safety objectives, and from transformer-assisted observation encoding that captures temporal dependencies for improved long-term control.

2) *Voltage Analysis*: Fig. 4 shows the voltage distributions across all buses over a representative day. TL-MAPPO maintains a much tighter and safer voltage range than the baselines, keeping almost all downstream buses above the 0.95 p.u. limit, whereas the other methods exhibit frequent undervoltage, especially at buses 14-33.

Fig. 5(a) further compares charging behaviors. TL-MAPPO provides the most stable and grid-aware profile, tracking demand while suppressing charging during peak-load periods. In contrast, MAPPO and MATD3 charge aggressively whenever EVs are present, and MASAC displays strong oscillations. As shown in Fig. 5(b), TL-MAPPO consistently preserves voltages within the safe band, while the other controllers

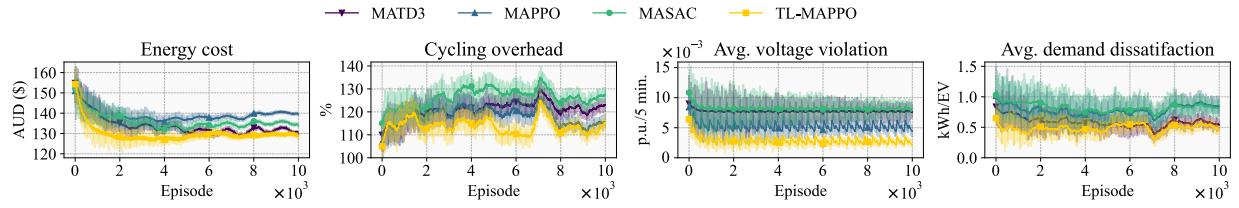


Fig. 3. Learning curves of the developed TL-MAPPO and three baselines.

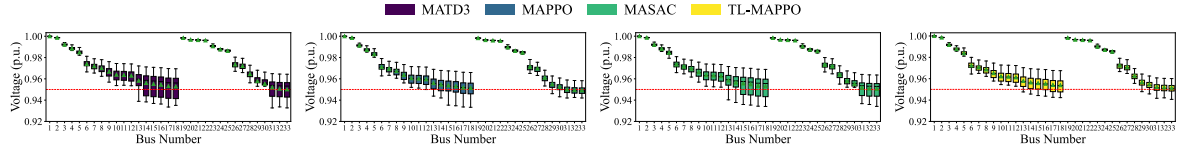


Fig. 4. Voltage statistical distributions of buses in one day of the compared methods.

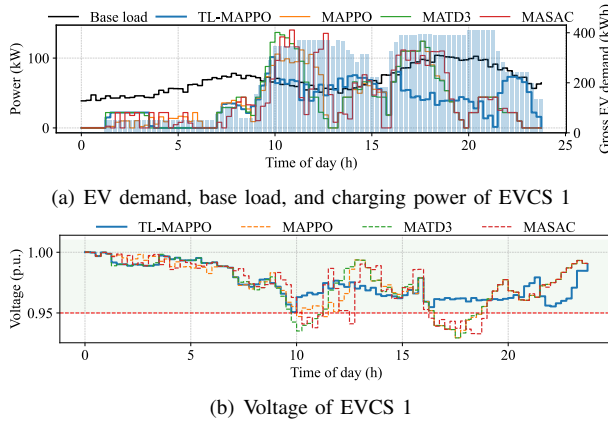


Fig. 5. EVCS-level comparison of (a) charging power rate and (b) voltage.

often cause voltages to drop below 0.95 p.u. Overall, TL-MAPPO achieves the most reliable voltage regulation and robust performance across network locations.

V. CONCLUSION

In this paper, we investigated the coordinated EV charging from a VPP perspective under partial distribution network visibility. To address voltage safety, limited network information, and decentralized decision-making, we proposed TL-MAPPO, which integrates Lagrangian-based safety regulation with Transformer-assisted temporal encoding in a multi-agent reinforcement learning framework. Simulation results on a realistic distribution network show that the proposed method reduces voltage violations and operational costs compared with representative MARL baselines while maintaining charging demand satisfaction. Future work will investigate larger-scale VPP deployments and more communication-efficient coordination mechanisms.

REFERENCES

- [1] International Energy Agency (IEA), "Net zero by 2050: A roadmap for the global energy sector." [Online]. Available: <https://www.iea.org/reports/net-zero-by-2050>
- [2] —, "IEA (2024), global EV outlook 2024." [Online]. Available: <https://www.iea.org/reports/global-ev-outlook-2024>
- [3] D. Pudjianto, C. Ramsay, and G. Strbac, "Virtual power plant and system integration of distributed energy resources," *IET Renewable power generation*, vol. 1, no. 1, pp. 10–16, 2007.
- [4] M. R. Khalid, I. A. Khan, S. Hameed, M. S. J. Asghar, and J. Ro, "A comprehensive review on structural topologies, power levels, energy storage systems, and standards for electric vehicle charging stations and their impacts on grid," *IEEE Access*, vol. 9, pp. 128 069–128 094, 2021.
- [5] C. Jiang, A. Liebman, and H. Wang, "Network-aware electric vehicle coordination for vehicle-to-everything value stacking considering uncertainties," in *2023 IEEE/IAS 59th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, 2023, pp. 1–9.
- [6] J. Zhang, L. Che, X. Wan, and M. Shahidehpour, "Distributed hierarchical coordination of networked charging stations based on peer-to-peer trading and EV charging flexibility quantification," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 2961–2975, 2022.
- [7] J. Fan, H. Wang, and A. Liebman, "MARL for decentralized electric vehicle charging coordination with V2V energy exchange," in *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, 2023, pp. 1–6.
- [8] R. Seppehrzad, M. J. Faraji, A. Al-Durra, and M. S. Sadabadi, "Enhancing cyber-resilience in electric vehicle charging stations: A multi-agent deep reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 11, pp. 18 049–18 062, 2024.
- [9] J. Zhang, Y. Guan, L. Che, and M. Shahidehpour, "EV charging command fast allocation approach based on deep reinforcement learning with safety modules," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 757–769, 2024.
- [10] S. Lee and D.-H. Choi, "Three-stage deep reinforcement learning for privacy-and safety-aware smart electric vehicle charging station scheduling and volt/var control," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8578–8589, 2024.
- [11] F. Rossi, C. Diaz-Londono, Y. Li, C. Zou, and G. Gruosso, "Smart electric vehicle charging algorithm to reduce the impact on power grids: a reinforcement learning based methodology," *IEEE Open Journal of Vehicular Technology*, pp. 1–13, 2025.
- [12] A. Stooke, J. Achiam, and P. Abbeel, "Responsive safety in reinforcement learning by pid lagrangian methods," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9133–9143.
- [13] S. Zhang, R. Jia, H. Pan, and Y. Cao, "A safe reinforcement learning-based charging strategy for electric vehicles in residential microgrid," *Applied Energy*, vol. 348, p. 121490, 2023.
- [14] Australian Renewable Energy Agency (ARENA), "Advanced VPP grid integration project," 2021, <https://arena.gov.au/assets/2021/05/advanced-vpp-grid-integration-final-report.pdf>.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, E. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [16] Z. J. Lee, T. Li, and S. H. Low, "ACN-data: Analysis and applications of an open EV charging dataset," in *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, 2019, pp. 139–149.
- [17] E. L. Ratnam, S. R. Weller, C. M. Kellett, and A. T. Murray, "Residential load and rooftop pv generation: an australian distribution network dataset," *International Journal of Sustainable Energy*, vol. 36, no. 8, pp. 787–806, 2017.
- [18] Australian Energy Market Operator (AEMO), "NEM data dashboard," 2023, <https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/data-dashboard-nem>.