Multi-Agent Reinforcement Learning Based Electric Vehicle Charging Control for Grid-Level Services

Md Golam Dastgir, Xiang Huo, and Mingxi Liu
Department of Electrical and Computer Engineering
University of Utah
Salt Lake City, UT, USA
{golam.dastgir, xiang.huo, mingxi.liu}@utah.edu

IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society | 978-1-6654-8025-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/IECON49645.2022.9968587

***Abstract*—Coordinating the charging process of a large population of electric vehicles (EV) is promising in increasing power grid flexibility from the demand side, yet requires highly scalable control protocols. In contrast to classical decentralized optimization based methods that require approximated distribution network models, this paper frames the EV charging control problem into a multi-agent reinforcement learning (MARL) framework. The MARL-based framework is trained through an actor-critic network and adopts the structure of centralized training and decentralized execution with partial observations. Comparing with model-based approaches, the developed MARL-based approach better captures the attributes of the distribution network, improves grid-level service performance, achieves better network constraints control, reduces the communication load, and achieves a faster response. The efficacy and efficiency of the developed method are verified by simulations on the IEEE 13-bus test feeder.**

***Index Terms*—Multi-agent reinforcement learning, electric vehicle charging control, smart grid, distribution grid service, distributed control**

I. Introduction

The market share of electric vehicles (EV) continues to grow because of EVs's reduced emissions and governmental incentives. Besides, the aggregated energy capacity of a large population of EVs is proven to be effective in increasing power system flexibility, through providing grid-level services at the demand side, to facilitate the integration of renewable energy resources [1], [2]. These features have led to recent research on EV charging control framework design for valley filling [3], power loss minimization [4], etc.

Historical EV charging control schemes rely on centralized control and optimization [4], [5]. As the number of EVs participating in grid services increases and the service coverage area grows, centralized approaches are strained by their inherently impaired scalability, preventing their use in large-scale implementations [6]. To alleviate the scalability issues, recent research focuses on distributed and decentralized EV charging control schemes. In [7], a distributed EV charging control scheme was developed for valley filling based on the alternating direction method of multipliers (ADMM). In [8], the authors proposed a distributed model predictive control (MPC) scheme to control a large population of EVs. However, the peer-to-peer communication overhead resulting from ADMM and distributed MPC impairs the practicality of those methods. To resolve this issue, Liu *et al.* [9] proposed the shrunken primal dual subgradient (SPDS) algorithm to realize decentralized EV charging control, eliminating the peer-to-peer communication load. This method was later extended in [10] as shrunken primal multi-dual subgradient (SPMDS) to achieve improved scalability *w.r.t.* distribution network dimension.

Despite the promising scalability, the aforementioned decentralized/distributed approaches must rely on well-defined distribution network models. In nature, the distribution network is a very complex network which is highly nonlinear and non-convex. Therefore, the distribution network model must be linearized or convexified for use in any optimization-based EV charging control approaches. In [9], [10], LinDistFlow model [11] was used to obtain single-phase linear distribution network models, where power loss was omitted. Sankur *et al.* [12] proposed using semi-definite programming to linearize the unbalanced distribution network. However, this approach must rely on assumptions of ratios of angles across three phases are fixed and zero line losses. Nick *et al.* [13] proposed the use of relaxed-optimal power flow by relaxing power losses to convexify the distribution network. However, this method is only applicable for balanced networks. Though linearizing and convexifying the network model enable optimization-based EV charging control, many nonlinear, nonconvex, but critical characters of distribution networks are omitted, which is expected to result in service performance discrepancies, network constraint violations, and power loss increase in practical applications.

To circumvent the model-simplification-induced performance deviations, model-free, especially reinforcement learning (RL) based, EV charging control and general distributed energy resource control approaches are attracting growing attentions. Dang *et al.* [14] introduced a Q-learning based single-agent RL technique to realize EV charging control for peak load reduction. Paraskevas *et al.* [15] introduced a deep Q-learning based single-agent RL approach for EV charging control, however, without consideration of individual objectives such as battery state-of-health (SOH) protection. More importantly, the action space defined in this approach is discrete, which is not suitable for grid services. Chiş *et al.* [16] proposed a demand response program for decreasing the long-

This work has been supported in part by NSF Award: ECCS-2145408.

term battery charging cost for EVs using RL-based techniques. Though the above work validated the possibility of using RL in EV charging control, the single-agent nature makes them unscalable for large-scale implementation. Considering the scalability, ease of implementation, and flexibility, multi-agent RL (MARL), which inherently considers the interactions of multiple agents in the same environment, has emerged in recent EV charging control and distributed energy resource coordination. MARL model is normally trained by using an actor-critic network in a dual network mechanism. The trained model provides each agent an individual policy that allows it to, based on local observations and inherently considering the interactions with other agents, generate its optimal actions to maximize the reward. This enables MARL to be more robust, computationally feasible, and practical solutions for large-scale EV charging control. Jiang *et al.* [17] developed a MARL approach based on multi-step Q learning to realize EV charging control in a completely cooperative setting. Cao *et al.* in [18] used multi-agent deep deterministic policy gradient (MADDPG) to train a MARL model in a cooperative setting to leverage solar photovoltaic for distribution voltage regulation. Huang *et al.* [19] also adopted MADDPG to train a MARL model in minimizing the charging cost of EV and avoiding peak demand. Though the above methods provide a promising direction of using MARL in EV charging control, the completely cooperative setting made them fail to consider each agent's customized local objectives. Considering both global and local objectives in MARL requires a mixed cooperative-competitive setting which is the most challenging setting in MARL algorithm design. Marinescu *et al.* [20] made an attempt by proposing a prediction-based MARL (P-MARL) EV charging control approach in a mixed setting to encourage charging during low-demand hours. However, the application of this method is not general enough to realize various grid services. Therefore, a generic scalable MARL-based EV charging control framework is yet to be developed.

This paper will construct an MARL based EV charging control framework to achieve decentralized EV charging control for the provision of various grid-level services while maintaining grid constraints. The contributions of this paper include: (1) the EV charging control problem is, for the first time, framed into MARL framework under a mixed cooperative-competitive setting to concurrently consider grid services and individual charging objectives, and (2) it validates that directly applying model-based approaches in real distribution networks causes service performance degradation and violations of network constraints.

The rest of the paper is structured as follows. Section II presents the problem formulations of the contemporary optimization-based EV charging control framework. The developed MARL-based EV charging control is presented in Section III. Section IV presents the simulation results which compare the model-based and model-free approaches. Finally, the paper is concluded in Section V with some future research directions.

II. Model-Based Approach

In this work, we consider controlling the charging process of a large population of EVs to provide distribution grid-level services while fulfilling local charging objectives. The general EV charging control can be formulated as

$$\begin{aligned} \min\ & F(\mathcal{U}) + \sum_{i=1}^{n} f_i(\mathcal{U}_i) \\ \text{s.t. } & \boldsymbol{g}(\mathcal{U}) \le \mathbf{0} \\ & \mathcal{U}_i \in \mathbb{U}_i,\ \forall\ i = 1, \dots, n, \end{aligned} \tag{1}$$

where $\mathcal{U}_i \in \mathbb{R}^K$ denotes the charging schedule, along the concerned period of length K, of the ith EV where each element $u_i(k)$ is the charging rate, ranging between 0 and 1, at time k, $\mathcal{U}$ is the collection of all EVs' charging schedules, $F(\cdot)$ represents the grid service objective, $f_i(\cdot)$ represents the ith EV's local objective function, $\boldsymbol{g}(\cdot)$ denotes the distribution network operational constraint function that involves power flow equations, and $\mathbb{U}_i$ is the local charging constraint set of the ith EV.

Solving the problem in (1) by decentralized optimizations requires a well-defined, more practically convex, distribution network model to facilitate the construction of objective functions and network constraint functions. LinDistFlow [11] has been widely used to characterize a linear network model. For a distribution network with h nodes, the LinDistFlow model at time k is written as

$$\boldsymbol{V}(k) = \boldsymbol{V}_0 - 2\boldsymbol{R}\boldsymbol{p}(k) - 2\boldsymbol{X}\boldsymbol{q}(k), \tag{2}$$

where $\boldsymbol{V}(k) = [V_1(k) \cdots V_h(k)]^\mathsf{T} \in \mathbb{R}^h$ with $V_\imath(k)$ being the squared voltage magnitude at node $\imath$, $\boldsymbol{V}_0 = [V_0 \cdots V_0]^\mathsf{T} \in \mathbb{R}^h$ and V_0 denotes the squared nominal voltage magnitude at the feeder head, $\boldsymbol{p}(k) = [p_1(k) \cdots p_h(k)]^\mathsf{T} \in \mathbb{R}^h$ with $p_\imath(k)$ being node $\imath$'s aggregated real power consumption, $\boldsymbol{q}(k) = [q_1(k) \cdots q_h(k)]^\mathsf{T} \in \mathbb{R}^h$ with $q_\imath(k)$ being node $\imath$'s aggregated reactive power consumption, and $\boldsymbol{R} \in \mathbb{R}^{h\times h}$ and $\boldsymbol{X} \in \mathbb{R}^{h\times h}$ are linear mappings from nodal real power consumption and reactive power consumption, respectively, to squared voltage magnitude drops. Details of the definitions of $\boldsymbol{R}$ and $\boldsymbol{X}$ can be found in [9], [21].

At the $\imath$th node, the aggregated real power consumption consists of the uncontrollable baseline load $p_{\imath,b}(k)$ and controllable EV charging load $p_{\imath,EV}(k)$, indicating

$$p_\imath(k) = p_{\imath,b}(k) + p_{\imath,EV}(k). \tag{3}$$

Assuming EVs do not consume a significant amount of reactive power, we have

$$q_\imath(k) = q_{\imath,b}(k). \tag{4}$$

Consequently, (2) is rewritten as

$$\boldsymbol{V}(k) = \boldsymbol{V}_0 - \boldsymbol{V}_b(k) - 2\boldsymbol{R}\boldsymbol{p}_{EV}(k), \tag{5}$$

where $\boldsymbol{V}_b(k)$ is the squared voltage drop due to the baseline load, and $\boldsymbol{p}_{EV}(k) = [p_{1,EV}(k)\ \cdots\ p_{h,EV}(k)]^\mathsf{T}$.

In this paper, we consider two types of distribution services for the global objective $F(\cdot)$, i.e., valley filling and power loss minimization. The former can be formulated as

$$F(\mathcal{U}) = \left\| \boldsymbol{P}_b + \sum_{i=1}^{n} \bar{P}_i \mathcal{U}_i \right\|_2^2, \tag{6}$$

where $\boldsymbol{P}_b \in \mathbb{R}^K$ is the aggregated baseline load profile across the valley filling period of the distribution feeder and $\bar{P}_i$ is the maximum charging power of the ith EV.

For power loss minimization, at any time instant k, define

$$\tilde{\boldsymbol{P}}(k) = [\bar{P}_1 u_1(k)\ \bar{P}_2 u_2(k)\ \cdots\ \bar{P}_n u_n(k)]^{\mathsf{T}}. \tag{7}$$

Assuming each node has the ideal voltage magnitude, the total active power loss can be calculated as [22]

$$F(\mathcal{U}) = \sum_{k=1}^{K} \tilde{\boldsymbol{P}}(k)^{\mathsf{T}} \tilde{\boldsymbol{R}} \tilde{\boldsymbol{P}}(k), \tag{8}$$

where, $\tilde{\boldsymbol{R}} \in \mathbb{R}^{n \times n}$ denotes a matrix evaluated as

$$\tilde{\boldsymbol{R}} = \boldsymbol{D}^{\mathsf{T}} \boldsymbol{A}^{\mathsf{T}} \hat{\boldsymbol{R}} \boldsymbol{A} \boldsymbol{D}, \tag{9}$$

where, $\boldsymbol{D} = \mathrm{diag}\{\boldsymbol{D}_\imath\} \in \mathbb{R}^{h \times n}$, $\imath = 1, 2, \ldots, n$ with $\boldsymbol{D}_\imath = \mathbf{1}_{n_\imath}^{\mathsf{T}}$ is the nodal aggregation vector, $n_\imath$ is the number of EVs connected to node $\imath$, $\boldsymbol{A} \in \mathbb{R}^{h \times h}$ denotes the connectivity matrix with $\boldsymbol{A}_{\imath,\jmath} = 1$ if the line segment $(\imath, \jmath) \in \mathbb{S}_\imath$, where $\mathbb{S}_\imath$ denotes all downward line segments from node $\imath$, otherwise $\boldsymbol{A}_{\imath,\jmath} = 0$, and $\hat{\boldsymbol{R}} = \mathrm{diag}\{r_{\imath\jmath}\} \in \mathbb{R}^{h \times h}$, for $\imath, \jmath = 1, \cdots, h$, where, $r_{\imath\jmath}$ denotes the resistance for the line $(\imath, \jmath)$.

Each EV would have its own charging objective. Herein, we consider the battery SOH protection for illustration, which can be represented as

$$f_i(\mathcal{U}_i) = \|\mathcal{U}_i\|_2^2. \tag{10}$$

For the ith EV, let η_i be the charging efficiency, Δk be the sampling time interval, $x_i(0)$ be the initial charging demand, $E_i(k)$ be the remaining energy needed to charge at time k, then the charging requirement is expressed as

$$E_i(K) = x_i(0) - \sum_{k=1}^{K} \eta_i \bar{P}_i u_i(k) \Delta k = 0. \tag{11}$$

Assuming no reactive power supply or other distributed generation in the network and considering the lower bound of the distribution network operating voltage, the global constraint $g(\mathcal{U})$ can be expressed as

$$g(\mathcal{U}) = \mathcal{Y}_b - \sum_{i=1}^{n} \mathcal{D}_i \mathcal{U}_i \leq 0, \tag{12}$$

where, $\mathcal{Y}_b \in \mathbb{R}^{hK}$ denotes the squared baseline voltage magnitude profile across the service period and $\mathcal{D}_i \in \mathbb{R}^{hK \times K}$ denotes mapping from EV charging schedules to the voltage drop profile. Detailed derivations and definitions can be found in [9].

To realize decentralized control, problem in (1) can be solved using ADMM [7], regularized primal-dual subgradient (RPDS) [23], or shrunken-primal-dual subgradient (SPDS) [9]. In this paper, we chose SPDS as the representative to present the results from model-free approaches. Detailed SPDS algorithm can be found in [9].

III. Model-Free Approach

In this section, we adopt MADDPG to develop a MARL-based EV charging control framework. The MARL-based framework used in this work is formulated into a mixed cooperative-competitive setting to incorporate both global and local objectives. To achieve this purpose, we extend the Markov Decision Process (MDP) to a multi-agent optimization problem with EVs' charging objectives. The MDP in (1) is represented by the tuple $(\mathcal{N}, \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{R})$, where $\mathcal{N}$ represents the set of n agents (EVs), $\mathcal{S}$ represents the joint state space, $\mathcal{A}$ represents the joint action space, $\mathcal{O}$ represents the joint observation space, $\mathcal{R}$ represents the joint reward achieved in transition of states. Each state $s = \{s_1, s_2, \ldots, s_K\} \in \mathcal{S} := \mathcal{S}_1 \times \cdots \times \mathcal{S}_K$ consists of the network states across the charging scheduling duration K. Each action $a = \{a_1, a_2, \ldots, a_n\} \in \mathcal{A} := \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ consists of the actions across the charging schedule of all n EVs. Then each EV constructs its own policy as $\pi_i(a_i|o_i) : \mathcal{A} \times \mathcal{O} \to [0, 1]$. The objective of the training is to learn a joint policy $\pi := [\pi_i]_{i=1}^n$ based on the joint action a so that the reward is maximized. The probability to move from state $s_1 \in S$ to $s_2 \in S$ after executing action a_1 is denoted by $\mathcal{P}(s_2|s_1; a_1) : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [1]$. This transition results in each EV receiving two types of rewards, i.e., the local reward and the global reward, for the transition from s to s' after executing a.

1) State Space: Assume each EV is connected to an associated building. Then the state space $\mathcal{S}$ includes all buildings' baseline loads, charging demands of all EVs, and the baseline nodal voltage magnitudes, i.e., at time k, $s_k = [p_{k,b}\ \boldsymbol{E}_k^{\mathsf{T}}\ \boldsymbol{v}_k^{\mathsf{T}}]^{\mathsf{T}}$, where $p_{k,b}$ is the aggregated baseline load, $\boldsymbol{E}_k = [E_{1,k}\ \ E_{2,k} \ldots E_{n,k}]^{\mathsf{T}} \in \mathbb{R}^n$ represents the energy remained to be charged of all n EVs, and $\boldsymbol{v}_k = [v_{1,k}\ v_{2,k} \ldots v_{h,k}]^{\mathsf{T}} \in \mathbb{R}^h$ represents the baseline nodal voltage magnitudes of all nodes.

2) Action Space: In the action space, $a_i = \mathcal{U}_i$ denotes the charging schedule of the ith EV. The joint action, $a \in \mathcal{A}$, is therefore represented as

$$a = \mathcal{U} = [\mathcal{U}_1^{\mathsf{T}}\ \cdots\ \mathcal{U}_n^{\mathsf{T}}]^{\mathsf{T}} \tag{13}$$

3) Reward: The local reward of the ith EV is constructed to meet its local charging demand and comply with the nodal voltage constraints. Let, during each state s, the local reward EV i will receive after the joint action a is taken be denoted by r_i, then the reward can be calculated as

$$r_i = r_i^E + r_i^v + r_i^B, \tag{14}$$

where

$$r_i^E = \begin{cases} 0 & \text{if } E'_{K,i} = 0 \\ -10^8 & \text{if } E'_{K,i} \neq 0, \end{cases} \tag{15}$$

and

$$r_i^v = \begin{cases} 0 & \text{if } v'_{k,j} \geq \underline{v}_0 \\ -10^8 & \text{if } v'_{k,j} < \underline{v}_0, \end{cases} \tag{16}$$

and

$$r_i^B = -\|a_i\|_2^2 \tag{17}$$

$\forall k = 1, \ldots, K$ and j is the node where EV i is connected.

Considering the grid service objectives, the total reward EV i will receive is calculated as

$$\mathcal{R}_i = r_g + r_i,$$

where

$$r_g = \begin{cases} -F(\mathcal{U}) & \text{for valley filling} \\ -\text{ Power loss} & \text{for power loss minimization.} \end{cases} \tag{18}$$

4) Observation Space: For the ith EV, the observation space includes its charging demand and the baseline nodal voltage profile of the node $\imath$ it is connected to. Mathematically, this can be represented by $o_i = [\boldsymbol{v}_i^{\mathsf{T}}\ x_i(0)]^{\mathsf{T}}$.

5) Actor-Critic Network & MADDPG Algorithm: In this paper, we adopt MADDPG to realize the training of the MARL-based EV charging control strategy. The structural overview of the training and execution of the MARL framework is shown in Fig. 1. The MADDPG algorithm works in

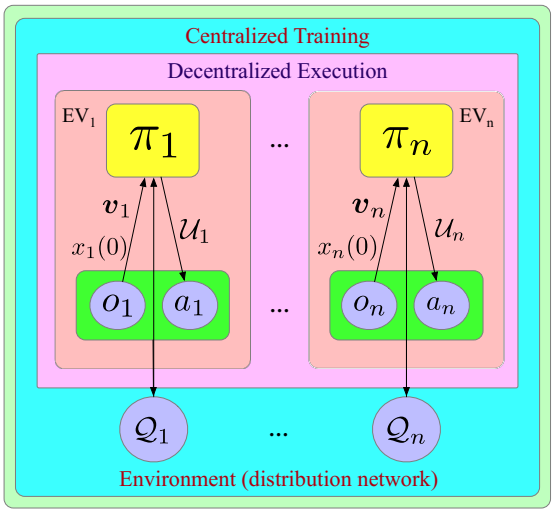

Fig. 1. Actor-critic network for centralized training and decentralized execution

such a way that during the training phase, the critic network has information of all the parameters including the global and local variables, while the actor network only has access to local information. We use a centralized critic to provide information about the optimality of the policy's actions for the entire system. The action space determined from the joint optimal policy sets the charging schedule for each EV. After the training is completed, the actors act in a decentralized way based on the local information only. Since each agent has its own critic function, the agents are allowed to possess different actions and rewards. Finally, the reward is calculated based on the actions from each agent. During the training process, the actor network takes action $a_i \in \mathcal{A}$ in the training data set; the critic network maps the state-action pair to the global and local rewards. For the actor network, the policy gradient for thewhich maps the observation to action is denoted by

$$\begin{aligned} \nabla_{\theta_i}\pi_i = \mathbb{E}_{s,a\sim\mathcal{B}}[\ & \nabla_{\theta_i}\pi_i(a_i|o_i) \\ & \nabla_{a_i}\mathcal{Q}_i^{\pi}(s,a)|_{a_i=\pi_i(o_i)}] \end{aligned} \tag{19}$$

where $\mathcal{B}$ is the replay buffer, $\mathcal{Q}_i^{\pi}(\cdot)$ denotes the critic function, θ_i denotes the weight given to the policy. The critic function used for the critic network is represented as

$$\mathcal{Q}_i^{\pi}(s,a) = \mathbb{E}_{s,s'\sim M}\left[\mathcal{R}_i(s,a)\right] \tag{20}$$

where M is the environment and $\mathcal{R}_i(s,a)$ denotes the total reward agent i receives. The critic network estimates the value of the policy followed by the actor after transitioning from state s to s'. The goal of the critic is to maximize the total reward received at each iteration.

IV. Simulation Results

A. Environment Setup

To compare the performance of the optimization-based and MARL-based approaches, a single-phase IEEE-13 bus test feeder is used to simulate the distribution network environment. For the optimization-based approach, the environment is formulated by using LinDistFlow in Matlab. For the MARL-based model-free approach, the environment is constructed in GridLab-D for distribution network simulation. The MARL model is trained by Spyder using the *keras* library and *tensorflow* library. Once the model is trained, same test data sets are used in both optimization-based and MARL-based approaches to compare the performance, efficiency, and efficacy.

B. Data Collection

In total 540 EVs are configured to connect to the single-phase IEEE-13 bus test feeder, i.e., 45 EVs are connected to each node. The EVs' battery capacities vary from 18 kWh to 20 kWh. The initial and desired state of charge (SOC) of all EVs uniformly distribute within $[0.3,\ 0.5]$ and $[0.7,\ 0.9]$, respectively. The baseline load of each house is collected and scaled from Southern California Edison. Each baseline load data for each time instant is randomized at each data point within 1% to 5% and in total 10400 data points are generated. The service period is set to from 7 pm to 8 am next day with a 15-minute time resolution, and the lower voltage bound is set to $\underline{v} = 0.954$. Three testing aggregated baseline load profiles are shown in Fig. 2.

C. Simulation Results

By using the MADDPG algorithm presented in Section III, the MARL-based EV charging control framework is trained by using the training data generated in Section IV-B, providing each EV an optimal policy. By using the optimal policies, each EV generates its own charging profiles. Adopting the same three testing data sets as in Section IV-B, the charging profiles result in the evolution of the energy remained to be charged of all 540 EVs shown in Fig. 3, which readily validates that all EVs' charging requests can be fulfilled at the end of the service period.

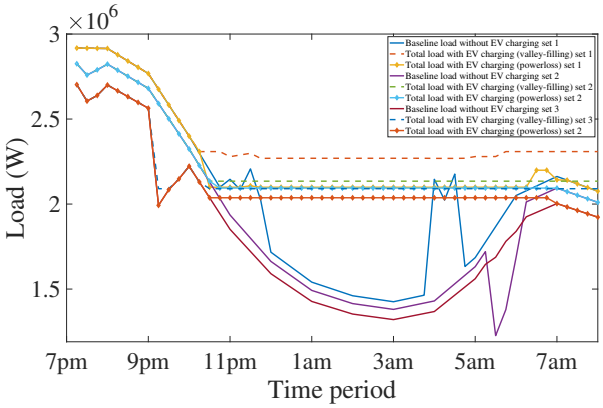

Fig. 2. Load curves for valley filling and power loss minimization under three data sets.

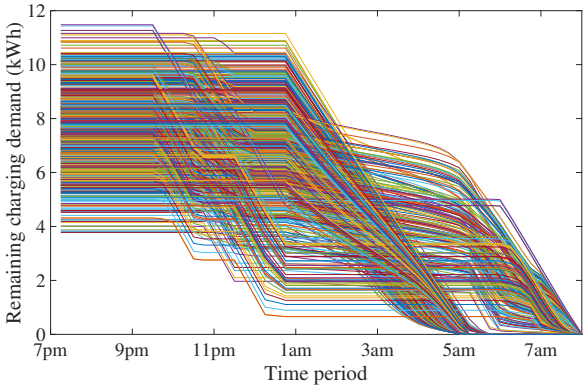

Fig. 3. Evolution of energy remained to be charged.

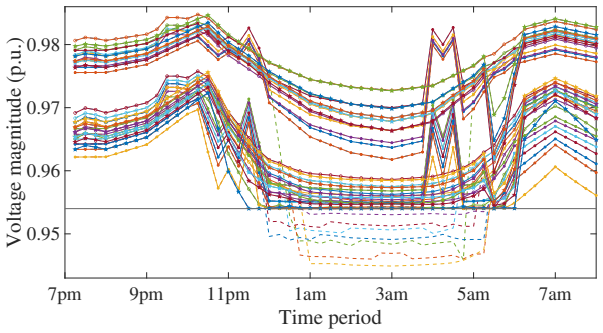

Fig. 4. Voltage magnitudes at 12 nodes: circle, asterisk, and diamond markers are from MARL-based approach for test sets 1, 2, and 3, respectively; dashed lines are from optimization-based approach for test set 1.

To compare the performance in terms of total power loss

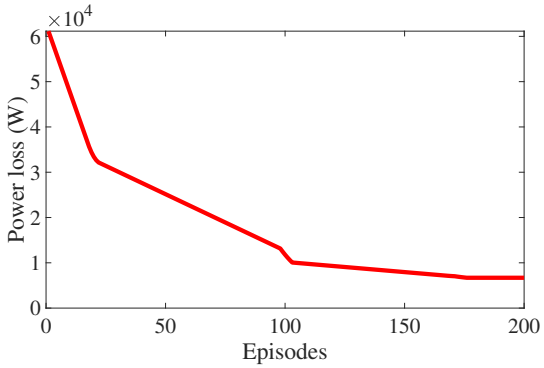

Fig. 5. Convergence of power loss along episodes

To compare the network constraint control performance of optimization-based and MARL-based approaches, charging profiles generated from those two methods are implemented in the same GridLab-D environment and result in the nodal voltage profiles as shown in Fig. 4. For the clarity of presentation, for the optimization-based approach only voltage magnitude profile of test data 1 is shown. It can be readily seen that the MARL-based approach can well maintain all nodes' voltage magnitudes above the designated lower bound, while the optimization-based approach fails to do so.

Fig. 2 demonstrates the feeder total load profiles under the objectives of valley-filling and power loss minimization using the MARL-based approach. In either case, the trained control framework tends to smooth the total load profile which fulfills the valley-filling objective.

Fig. 5 shows the evolution of power loss as the MARL model training proceeds in the power loss minimization case. of the optimization-based and MARL-based approaches, Fig. 6 presents the power loss comparisons by the same three testing data sets. It can be seen that the optimization-based approach has an average 50% more power loss than that of the MARL-based approach. This is owing to the fact that the optimization-based approach has omitted many critical nonlinear and nonconvex features of the distribution network, resulting in the power-loss-"minimizing" charging schedules not truly minimizing the power loss. In contrast, the MARL-based approach considers the full set of features of the distribution network, resulting in better charging schedules.

V. Conclusions

In this paper, a model-free EV charging control framework was developed by being framed into MARL under mixed cooperative-competitive setting and was trained by using the one-stage MADDPG algorithm. Simulations were conducted

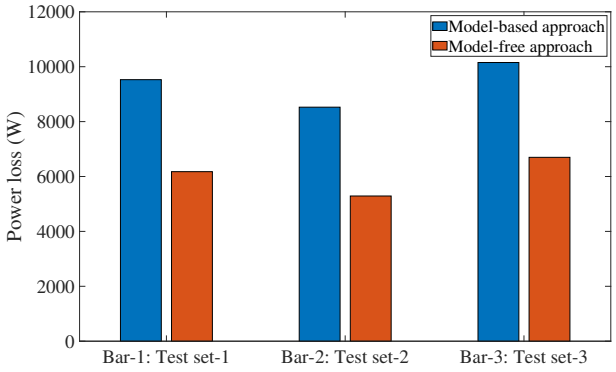

Fig. 6. Power loss comparison of model-based and model-free approaches

to compare the performance between the developed model-free approach and traditional decentralized optimization-based approaches. It has been validated that the MARL-based EV charging control outperforms the optimization-based control framework in terms of: (1) the model-free approach relaxes the control design from requiring an exact model of the network, which is usually impossible; (2) the adopted model-free approach captures the full features of the distribution network, and has a better control on power loss minimization; and (3) the model-based approach failed to satisfy the global network constraints, which has been resolved by the developed model-free approach. As a future research direction, the work can be extended to include the effect of distributed energy resources (DERs), and reactive power supplies in the network. Moreover, we may also improve the framework of the algorithm to achieve a more reduced power loss.

REFERENCES

[1] M. Gilleran, E. Bonnema, J. Woods, P. Mishra, I. Doebber, C. Hunter, M. Mitchell, and M. Mann, "Impact of electric vehicle charging on the power demand of retail buildings," *Advances in Applied Energy*, vol. 4, p. 100062, 2021.

[2] M. Blonsky, P. Munankarmi, and S. P. Balamurugan, "Incorporating residential smart electric vehicle charging in home energy management systems," in *Proceedings of the IEEE Green Technologies Conference*, pp. 187–194, Held Virtually, April 07-09, 2021.

[3] M. Liu, Y. Shi, and H. Gao, "Aggregation and charging control of PHEVs in smart grid: A cyber–physical perspective," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1071–1085, 2016.

[4] X. Luo and K. W. Chan, "Real-time scheduling of electric vehicles charging in low-voltage residential distribution systems to minimise power losses and improve voltage profile," *IET Generation, Transmission & Distribution*, vol. 8, no. 3, pp. 516–529, 2014.

[5] I. Sharma, C. Canizares, and K. Bhattacharya, "Smart charging of PEVs penetrating into residential distribution systems," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1196–1209, 2014.

[6] L. P. Fernandez, T. G. San Román, R. Cossent, C. M. Domingo, and P. Frias, "Assessment of the impact of plug-in electric vehicles on distribution networks," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 206–213, 2010.

[7] J. Rivera, P. Wolfrum, S. Hirche, C. Goebel, and H.-A. Jacobsen, "Alternating direction method of multipliers for decentralized electric vehicle charging control," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 6960–6965, Firenze, Italy, Dec 10-13, 2013.

[8] Y. Zheng, Y. Song, D. J. Hill, and K. Meng, "Online distributed MPC-based optimal scheduling for EV charging stations in distribution systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 638–649, 2019.

[9] M. Liu, P. K. Phanivong, Y. Shi, and D. S. Callaway, "Decentralized charging control of electric vehicles in residential distribution networks," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 1, pp. 266–281, 2019.

[10] X. Huo and M. Liu, "Two-facet scalable cooperative optimization of multi-agent systems in the networked environment," *IEEE Transactions on Control Systems Technology*, pp. 1–16, in press, 2022.

[11] M. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *IEEE Transactions on Power Delivery*, vol. 4, no. 1, pp. 735–743, 1989.

[12] M. D. Sankur, R. Dobbe, E. Stewart, D. S. Callaway, and D. B. Arnold, "A linearized power flow model for optimization in unbalanced distribution systems," *arXiv preprint arXiv:1606.04492*, 2016.

[13] M. Nick, R. Cherkaoui, J.-Y. Le Boudec, and M. Paolone, "An exact convex formulation of the optimal power flow in radial distribution networks including transverse components," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 682–697, 2017.

[14] Q. Dang, D. Wu, and B. Boulet, "A Q-learning based charging scheduling scheme for electric vehicles," in *Proceedings of the IEEE Transportation Electrification Conference and Expo*, pp. 1–5, Novi, MI, USA, June 19-21, 2019.

[15] A. Paraskevas, D. Aletras, A. Chrysopoulos, A. Marinopoulos, and D. I. Doukas, "Optimal management for EV charging stations: A win–win strategy for different stakeholders using constrained Deep Q-learning," *Energies*, vol. 15, no. 7, p. 2323, 2022.

[16] A. Chiş, J. Lundén, and V. Koivunen, "Reinforcement learning-based plug-in electric vehicle charging with forecasted price," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3674–3684, 2016.

[17] C. Jiang, Z. Jing, X. Cui, T. Ji, and Q. Wu, "Multiple agents and reinforcement learning for modelling charging loads of electric taxis," *Applied Energy*, vol. 222, pp. 158–168, 2018.

[18] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, and Z. Chen, "Distributed voltage regulation of active distribution system based on enhanced multi-agent deep reinforcement learning," *arXiv preprint arXiv:2006.00546*, 2020.

[19] S. Huang, M. Yang, J. Yun, P. Li, Q. Zhang, and G. Xiang, "A data-driven multi-agent PHEVs collaborative charging scheme based on deep reinforcement learning," in *Proceedings of the IEEE/IAS Industrial and Commercial Power System Asia*, pp. 326–331, Chengdu, China, Jul 18-21, 2021.

[20] A. Marinescu, I. Dusparic, A. Taylor, V. Cahill, and S. Clarke, "P-MARL: Prediction-based multi-agent reinforcement learning for non-stationary environments.," in *Proceedings of the Autonomous Agents and Multiagent Systems*, pp. 1897–1898, Istanbul, Turkey, May 4-8, 2015.

[21] M. Farivar, L. Chen, and S. Low, "Equilibrium and dynamics of local voltage control in distribution systems," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 4329–4334, Firenze, Italy, Dec 10-13, 2013.

[22] M. Liu, "Chance-constrained SPDS-based decentralized control of distributed energy resources," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 3272–3278, Nice, France, Dec 11-13, 2019.

[23] D. Yuan, D. W. C. Ho, and S. Xu, "Regularized primal–dual subgradient method for distributed constrained optimization," *IEEE Transactions on Cybernetics*, vol. 46, no. 9, pp. 2109–2118, 2016.