Safe On-Policy Deep Reinforcement Learning-based Electric Vehicle Charging Management under Transformer Thermal Stress

Zhewei Zhang[1, 2, 4], Rémy Rigo-Mariani[1], Nouredine Hadjsaid[1] and Yan Xu[3]
1. Univ. Grenoble Alpes, CNRS, Grenoble INP, G2Elab, 38000, Grenoble, France
2. CNRS@CREATE,1 Create Way, #08-01 Create Tower, Singapore
3. Centre for Power Engineering (CPE), School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
4. Energy Research Institute @ NTU, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore
{zhewei.zhang, remy.rigo-mariani, nouredine.hadjsaid}@g2elab.grenoble-inp.fr, xuyan@ntu.edu.sg

***Abstract*—With the increasing penetration of Electric Vehicles (EVs), distribution grids face challenges in accommodating charging demand within limited infrastructure capacity. Efficient control strategies are required to maximize energy delivery to EV fleets while ensuring grid safety. However, these strategies are hindered by uncertainties in EV information, heterogeneous charger types, and non-negligible transformer thermal stress. To address these challenges, this study proposes a Safe Deep Reinforcement Learning (SDRL)-based charging management framework that considers unknown EV information and transformer thermal loading. The approach is evaluated under different charger types and ambient temperatures, and multiple categories of SDRL algorithms are compared. Results demonstrate that Constrained Optimization (CO)-based algorithms achieve the best overall performance, delivering more than 90% of total demand, reducing constraint violations by up to 90% compared to uncontrolled charging, and performing 31.7%–62.5% closer to the theoretical optimum than other SDRL categories.**

Keywords—Safe Deep Reinforcement Learning, Electric Vehicle, Transformer, Thermal

I. Introduction

In recent years, the Electric Vehicle (EV) has been taking an increasing share in the transportation sector [1]. This brings unneglectable challenges to the distribution grid. While contrasted with the rapid growth of charging demand, the upgrading of the grid infrastructure is relatively more time-consuming [2]. For distribution grid operators, the key challenge is accommodating EV charging within limited grid capacity—such as power transformer capacity—while maximizing energy delivery and ensuring grid safety.

One critical aspect of EV charging management lies in addressing the uncertainty associated with future charging sessions. To address the problem above, many researchers have proposed Safe Deep Reinforcement Learning (SDRL)-based methods, owing to SDRL's inherent capability to handle stochastic and uncertain environments while ensuring safety. In the formulation of SDRL, the control objective is treated as a reward, and the constraints are interpreted as costs. The safety is ensured by limiting the cost to a certain tolerance [3]. There are also some applications of SDRL in EV charging management. In [4], researchers adopted a Lagrangian-based SDRL to ensure the microgrid's battery capacity during EV charging. In [5], researcher adopted Constrained Policy Optimization (CPO) to ensure the EV battery charging capacity. In [6], researchers adopted Lagrangian-based SDRL to ensure the grid voltage constraint and line loading constraints. In [7], researchers adopted Lagrangian-based SDRL to ensure EV battery constraints during the charging.

However, the studies mentioned above still present several limitations that need to be addressed. First, the uncertainty in the departure time and energy demand of connecting EVs is often neglected. Most existing studies assume that EV users provide accurate information about their departure and energy requirements upon connection, which is unrealistic in practice [8]. Second, the thermal characteristics of critical grid infrastructure, such as transformer temperature rise due to increased loading, are not fully considered. However, temperature is a key indicator of safe and reliable operation. Moreover, the impact of EV charging on the distribution grid naturally depends on the type of charging station (e.g., AC Level 2 chargers and DC fast chargers). However, this aspect has not been comprehensively investigated. Therefore, to address the research gaps outlined above, we proposed an SDRL-based EV charging management scheme that considers transformer thermal loading. The objective is to maximize the total energy delivered to the EV fleet while ensuring grid safety by preventing both power and thermal overload of the transformer. The main contributions are summarized below:

- We proposed an SDRL-based EV charging management framework that accounts for unavailable EV charging information (i.e., unknown departure time and charging demand) and transformer thermal stress. The proposed method was tested under different ambient temperature scenarios.
- We investigated the proposed methods with different charger types to evaluate the effectiveness of the SDRL-based management strategy.
- We analyzed the performance of different categories of on-policy SDRL algorithms, and their performance was benchmarked with the theoretical optimum.

The paper is organized as following. In section II, we present our problem formulation. In section III, we introduce different categories of SDRL algorithms. In section IV, we present and analyze the results and the conclusion is draw in section V.

II. Problem Formulation

A. Case Study & Modeling

We consider the charging station is connected to the grid with a transformer, as Fig. 1 shows. As mentioned above, the control objective is to maximize energy delivery to EV while minimize the grid impact by EV charging. The charging station contains N chargers and control time resolution is Δt . The control horizon T is with daily basis.

979-8-3315-8138-1

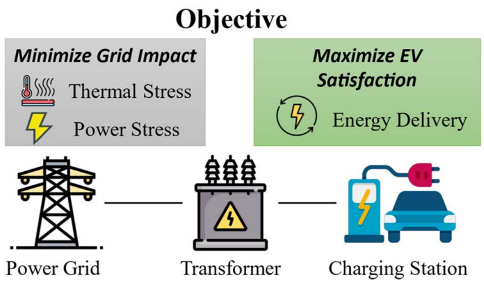

Fig. 1 Case Study & Control Objective

The EV charging model can be formulated as (1), where $p_{n,t}$ is the charging power of the port n at time t, $\overline{p}_n$ is the maximum charging power of the port, $T_n^{arr/dep}$ are arrival and departure time of the EV, $\Omega_{n,t}$ is a binary variable that denote the occupancy of the port, $\Delta E_{n,t}$ is the total delivered energy until time t and $\overline{E}_n$ is the demand of the EV. In the reality, T_n^{dep} and $\overline{E}_n$ are not available for the controller.

$$\begin{cases} 0 \le p_{n,t} \le \overline{p}_n \times \Omega_{n,t} \\ \Omega_{n,t} = 1, \forall t \in \left[T_n^{arr}, T_n^{dep}\right], else\ \Omega_{n,t} = 0 \\ \Delta E_{n,t} = \Delta E_{n,t-1} + p_{n,t} \times \Delta t \\ 0 \le \Delta E_{n,t} \le \overline{E}_n \end{cases} \tag{1}$$

In this work, the transformer thermal model is defined by IEC standard [9]. The transformer thermal model depends on the nominal loading p_t^{tr} which is normalized by transformer apparent capacity $\overline{P}^{tr}$, and is characterized with oil temperature θ_t^o which is calculated with (3) and hotspot temperature θ_t^h which is calculated by (4), where $\left[x, y, r, \kappa_{11}, \kappa_{21}, \kappa_{22}, \tau_0, \tau_w, \Delta\theta^{or}, \Delta\theta^{hr}\right]$ are the transformer parameters, δt is the calculation time resolution (1 minute resolution), while $\Delta\theta_t^{h1}$ and $\Delta\theta_t^{h2}$ are two auxiliary variables and are not observable (measurable) by the controller. By the definition of [9], the safety requirement for the transformer is summarized with (5).

$$p_t^{tr} = \frac{\sum_{n \in N} p_{n,t}}{\overline{P}^{tr}} \tag{2}$$

$$\begin{cases} \Delta\theta_t^o = \dfrac{1}{\kappa_{11}\tau_0}\left[\theta_t^a + \Delta\theta^{or}\left(\dfrac{1 + r\left(p_t^{tr}\right)^2}{1+r}\right)^x - \theta_{t-1}^o\right] \\ \theta_t^o = \theta_{t-1}^o + \Delta\theta_t^o \times \delta t \end{cases} \tag{3}$$

$$\begin{cases} \Delta\theta_t^{h1} = \dfrac{1}{\kappa_{22}\tau_w}\left[\kappa_{21}\Delta\theta^{hr}\left(p_t^{tr}\right)^y - \Delta\theta_{t-1}^{h1}\right] \\ \Delta\theta_t^{h2} = \dfrac{\tau_0}{\kappa_{22}}\left[(\kappa_{21}-1)\Delta\theta^{hr}\left(p_t^{tr}\right)^y - \Delta\theta_{t-1}^{h2}\right] \\ \theta_t^h = \theta_t^o + \left(\Delta\theta_t^{h1} - \Delta\theta_t^{h2}\right) \times \delta t \end{cases} \tag{4}$$

$$p_t^{tr} \le 1.5, \theta_t^o \le \overline{\theta}^o, \theta_t^h \le \overline{\theta}^h \tag{5}$$

Therefore, the control problem can be summarized as (6), which maximizing the energy delivery while ensure the safe operation.

$$\max \sum_{n,t} \Delta E_{n,t},\ s.t.\ (1) \sim (5) \tag{6}$$

B. Constrained Markov Decision Process Formulation

In order to address the unavailable EV information and the transformer safety requirement (5), we model the above problem by Partially Observable Constrained Markov Decision Process (PO-CMDP). In PO-CMDP, the agent takes action based on observations o_t rather than states s_t, and the constraints violation are modeled as cost c_t. The formulation is shown as followings:

States: The states is defined as (7), which contains all variables of system dynamics.

$$s_t = \begin{bmatrix} \left(\Omega_{n,t}, \Delta E_{n,t}, p_{n,t}, \overline{E}_n, T_n^{dep}\right)|_{n \in N}, \\ \theta_t^a, \theta_t^o, \theta_t^h, \Delta\theta_t^{h1}, \Delta\theta_t^{h2}, t \end{bmatrix} \tag{7}$$

Observations: The observations is defined as (8), which contains only observable variables from EV and transformer.

$$o_t = \left[\left(\Omega_{n,t}, \Delta E_{n,t}, p_{n,t}\right)|_{n \in N}, \theta_t^a, \theta_t^o, \theta_t^h, t\right] \tag{8}$$

Actions: The action is defined as (9), which are charging power at each port.

$$a_t = [p_{n,t}|_{n \in N}] \tag{9}$$

Rewards: The reward is defined as (10), as the objective is to maximize the energy delivery to the EV fleet.

$$r_t = \sum_n \frac{\Delta E_n}{\overline{E}_n}\Big|_{t = T_n^{dep}} \tag{10}$$

Costs: The cost is defined as (11), which denotes the constraint violation of transformer operation in (5).

$$c_t = \sum_{\chi_t \in \{p_t^{tr}, \theta_t^o, \theta_t^h\}} \max\left(\frac{\chi_t - \overline{\chi}_t}{\overline{\chi}_t}, 0\right) \tag{11}$$

State Transitions: The state transitions are governed by the EV charging and transformer thermal model (1) ~ (4).

III. On Policy Safe Deep Reinforcement Learning

The generic formulation of on-policy SDRL is (12). The idea is maximize the reward advantage A_π^R (against previous policy) meanwhile ensures the cost tolerance d, meanwhile the distance between new and previous policy is governed by KL divergence D_{KL} and δ is KL Divergence tolerance. Generally speaking, there are three major categories to achieve safety in on-policy SDRL [10]. The detailed explanation is in the followings.

$$\begin{aligned} &\pi_{k+1} = \arg\max \mathrm{E}_{\substack{o \sim \pi_k \\ a \sim \pi}}[A_\pi^R] \\ &s.t. \begin{cases} J^C(\pi_k) + \mathrm{E}_{\substack{o \sim \pi_k \\ a \sim \pi}}[A_\pi^C] \le d \\ D_{KL}(\pi \| \pi_k) \le \delta \end{cases} \end{aligned} \tag{12}$$

A. Constrained Optimization Methods (CO)

The idea of constrained optimization methods (CO) is, through Taylor Expansion, (12) is approximated into convex form. And in each iteration, by solving approximated convex optimization problem of (12), the policy is updated. Typical algorithms including Constrained Policy Optimization (CPO) [11] and Projection-based Constrained Policy Optimization

979-8-3315-8138-1

(PCPO) [12]. For example, the CPO update the policy parameter ϑ with (13), where g is the gradient of the reward advantage function, a is the gradient of the cost advantage function, $b := J^C(\pi_k) - d$ which is cost violation and H is the Hessian matrix of KL divergence.

$$\max_{\vartheta} g^T(\vartheta - \vartheta_k)$$
$$s.t. \begin{cases} a^T(\vartheta - \vartheta_k) + b \le 0 \\ \frac{1}{2}(\vartheta - \vartheta_k)^T H(\vartheta - \vartheta_k) \le \delta \end{cases} \tag{13}$$

B. Lagrangian-based Methods (Lag)

The idea of Lagrangian-based (Lag) methods is, through primal-dual decomposition, the cost constraints of (12) is adaptively aggregated by a Lagrangian multiplier, and the aggregated loss is used to update policy through DRL algorithms. After the policy updating, the Lagrangian multiplier is updated according to the resulted constraints violations. Typical algorithms include Lagrangian-based PPO (PPOLag) and Lagrangian-based Policy Gradient (PDO) [3]. For example, the PPOLag first calculate the aggregated advantage by (14) and update policy with A_{π_k} through PPO. Then update the Lagrangian multiplier λ_k by (15), when the constraint violation is greater than tolerance, i.e. $b > 0$, the λ increase and consequently, A_{π_k} is reduced.

$$A_{\pi_k} = A_{\pi_k}^R - \lambda_k A_{\pi_k}^C \tag{14}$$

$$\lambda_{k+1} = [\lambda_k + \eta b]_+ \tag{15}$$

C. Penalty Function Methods (PenF)

The idea of penalty function (PenF) methods is, using a barrier function such as log function, the cost $J^C(\pi)$ is aggregated with $J^R(\pi)$ and then the policy is updated with the aggregated loss function. Typical algorithms include Interior Point Policy Optimization (IPO) [13] and Penalized Proximal Policy Optimization (P3O) [14]. For example, the IPO adopts (16) to update the policy, where L^{PPO} is the clip loss function by PPO and $I(\square)$ is the barrier function ($I(b) \to -\infty$ when $b > 0$ and $I(b) = 0$ when $b \le 0$).

$$L^{IPO}(\vartheta) = L^{PPO}(\vartheta) + I(b) \tag{16}$$

IV. Results Analysis

A. Experiment Configuration

In this work, we consider two cases of charging stations: case 1 is AC level 2 charger with $\overline{p}_n$ 22 kW and case 2 is DC fast charger with $\overline{p}_n$ 50 kW. The charging station is connected with the grid by a transformer with $\overline{P}^{tr}$ 120 MVA, as we want to analysis charging management under limit grid infrastructure capacity. The parameter of the transformer is shown in Table I. According to [9], $\overline{\theta}^o$ is 105 °C and $\overline{\theta}^h$ is 120 °C. The ambient temperature are from [15] and the EV charging records are from [16]. On average, there are around 120 EV charging records on each day. We compared different value of cost tolerance d, as it controls the size of feasible action space. All agents are trained with historical data from Jan. ~ Jun. 2019, with in total 3,000 training epochs and each epochs contains 1,000 timesteps. The training and testing processes are run on our own PC. The configuration of our CPU is 13th Gen Intel(R) Core(TM) i7-13700H 2.40 GHz with 32 GB RAM.

Table I Transformer Parameters

Parameter	x	y	k_{11}	k_{21}	k_{22}
Value	0.8	1.6	1.0	1.0	2.0
Parameter	τ_o	τ_w	$\Delta\theta^{or}$	$\Delta\theta^{hr}$	r
Value	180 min	4 min	55 K	23 K	5

From the transformer thermal model (3) and (4), obviously, the ambient temperature θ_t^a has impact on the θ_t^o and θ_t^h. Therefore, to evaluate the effectiveness of the proposed strategy, the SDRL agents are tested under three scenarios with different ambient temperature: scenario "Hot" with entire Aug. 2019, scenario "Med" with entire Oct. 2019 and scenario "Cold" with entire Dec.2019.

B. Metrics

In order to compare the SDRL agents performance, we consider a theoretical optimum as the reference and an immediate charging that the EV is charged with $\overline{p}_n$ during connection. The theoretical optimum is obtained with the assumption that T_n^{dep} and $\overline{E}_n$ are available, and the transformer thermal model is piecewise linearized such as [17].

The agents' performance is evaluated by the distance to the results of theoretical optimum, as (17) shows, where r_a and c_a is the total reward and cost of each testing day by the SDRL agent, r_{opt} and c_{opt} is the total reward and cost of each testing day of the optimum. Obviously, smaller l indicates closer to the optimum and thus the corresponding agent is better.

$$l = \sqrt{\frac{(r_a - r_{opt})^2 + (c_a - c_{opt})^2}{r_{opt}^2 + c_{opt}^2}} \tag{17}$$

C. Numerical Results

The daily average energy delivery results of each scenario are shown Fig. 2 ~ Fig. 4. From the results, the CO-based SDRL can efficiently charge the EV fleet under each scenario, by delivering more than 90% energy with an AC level 2 charger, and around 80%~90% energy with a DC fast charger in every scenario. With the increase in the cost limit, the energy delivery increases as the feasible action space increases. Compared to the two types of chargers, the SDRL agents generally deliver more energy with the AC level 2 charger, indicating that the SDRL-based EV charging management strategy is more effective for the AC level 2 charger.

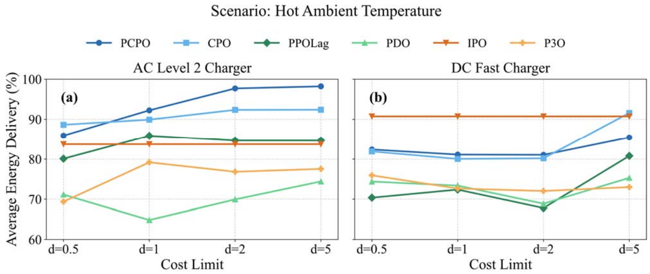

Fig. 2 Average Energy Delivery on Hot Temperature Scenario

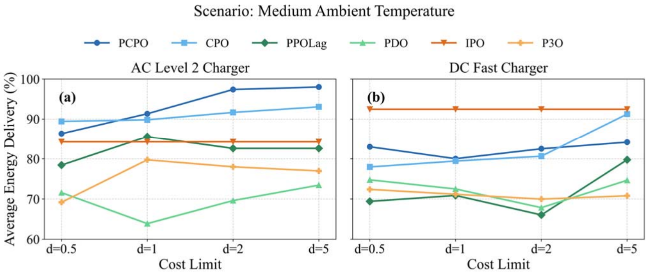

Fig. 3 Average Energy Delivery on Medium Temperature Scenario

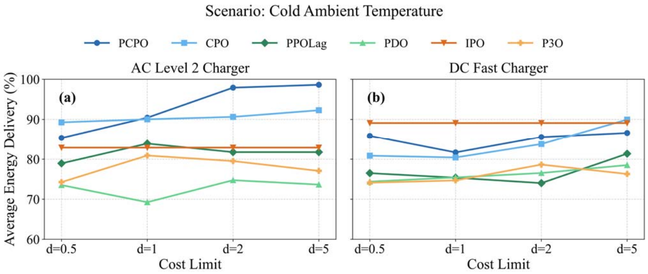

Fig. 4 Average Energy Delivery on Cold Temperature Scenario

To illustrate the capability to ensure safety, we present the hotspot temperatures resulting from SDRL-based management (PCPO) and immediate charging across the testing horizon in three scenarios in Fig. 5 and Fig. 6. From the results, the SDRL-based management can reduce hotspot temperature violations. Compared with immediate charging, the SDRL (PCPO)-based strategy can reduce constraint violations by 90.5%-96.6%.

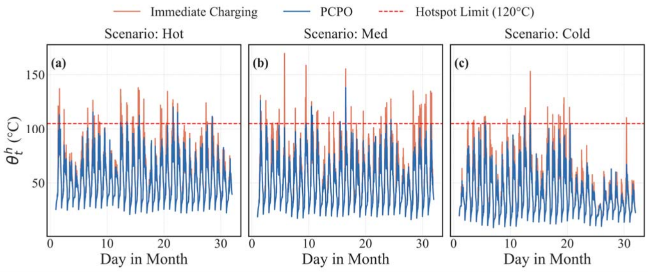

Fig. 5 Hotspot Temperature with AC Level 2 Chargers

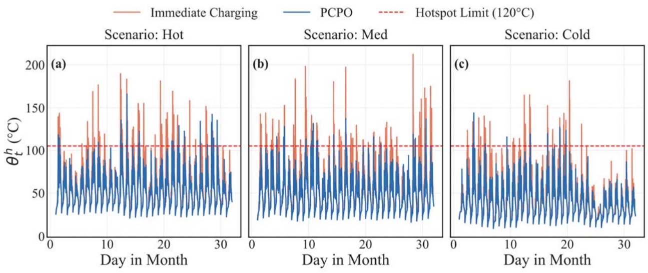

Fig. 6 Hotspot Temperature with DC Fast Chargers

The daily distribution of l for each algorithm throughout the testing horizon of each scenario are shown in Fig. 7 ~ Fig. 9, where (a) shows the results of AC Level 2 Charger and (b) shows the results of DC Fast Chargers. In every scenario, comparing both charging station configurations, the SDRL agents generally perform better with AC level 2 chargers. This is because the action space of the AC level 2 charger case is smaller (0~22 kW compared with 0~50 kW). Therefore, the agent can better explore the action space. Among the three categories of SDRL agents, CO-based agents perform the best over the other two kinds of SDRL. As cost tolerance increases, CO-based agents perform better. This is due to an increase in the feasible action space, so that the agent can find a better policy.

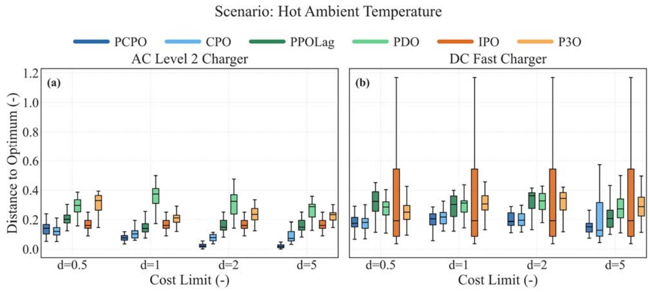

Fig. 7 Agent Performance Distribution Under Hot Scenario

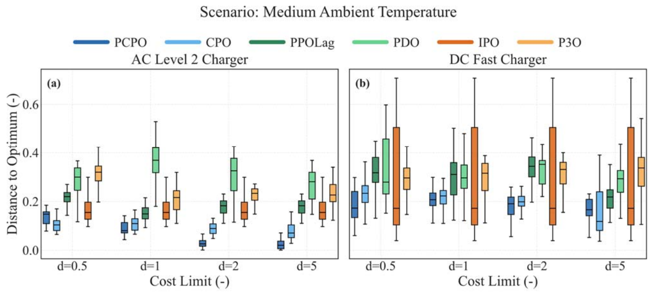

Fig. 8 Agent Performance Distribution Under Med Scenario

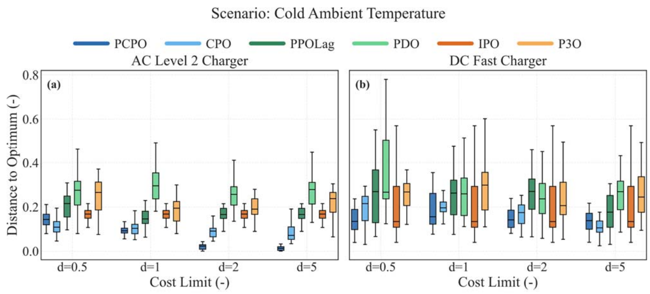

Fig. 9 Agent Performance Distribution Under Cold Scenario

In Table II, we present the average l to optimum (the smaller l indicates the better performance), the l value is averaged along the different cost tolerance d. Compared with three categories of on-policy SDRL algorithms, CO-based SDRL (PCPO, CPO) achieve approximately 55.0%~62.5% closer to the optimum compared with the Lag- and PenF-based SDRL in the AC level 2 charger case, and achieves 31.7%~41.2% closer to the optimum in the DC fast charger case. Especially among the CO-based SDRL, the PCPO performs better than CPO by 9.5%–36.4% across all scenarios and station charger types. This is mainly because the PCPO adopts a "correction" step during policy update to ensure safety [12], unlike the CPO, and therefore performs better.

Table II Average Distance to Optimum

	Station	CO		Lag		PenF	
		PCPO	CPO	PPOLag	PDO	IPO	P3O
Hot	Level 2	**0.07**	0.11	0.17	0.31	0.16	0.24
	DC Fast	**0.19**	0.22	0.29	0.31	0.36	0.33
Med	Level 2	**0.07**	0.11	0.18	0.30	0.16	0.24
	DC Fast	**0.19**	0.21	0.30	0.36	0.36	0.32
Cold	Level 2	**0.07**	0.10	0.19	0.20	0.17	0.23
	DC Fast	**0.16**	0.18	0.25	0.32	0.25	0.27

To evaluate the impact of measurement noise, we simulate the measurement noise from 3% to 10% and show its impact on reward and costs of PCPO agent in Fig. 10. From the results, the measurement noise has more evident impact on the cost (constraint violation) For the AC charger, the measurement noise can bring at most 6% more violations and for the DC charger, the measurement noise can bring at most 10% more violations.

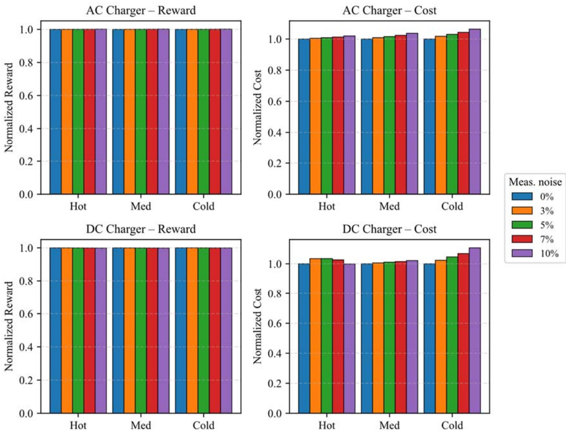

Fig. 10 Measurement Noise Impact

V. Conclustion

In this paper, we presented an SDRL-based EV charging management system that considers unknown EV information and transformer thermal stress. To evaluate the strategy's robustness, we tested scenarios with different ambient temperatures. The two types of chargers, AC Level 2 chargers and DC fast chargers, are also investigated. From the results, the CO-based SDRL EV charging management strategy is the most effective under each scenario and for both types of chargers. Compared with other categories of SDRL algorithms, CO-based algorithms perform 31.7%~62.5% better (closer to the optimum). What's more, the results show that the proposed strategy is more effective with the AC level 2 charger than DC fast chargers, as the action space is smaller and agents can explore more effectively. The significance of this work is that it provides an insight into the effectiveness of SDRL-based charging management based on available EV information, under different charger types and ambient temperatures. The limitation is that, as the SDRL ensures safety by setting constraints violations within a tolerance range, there are still constraints violations in the final results. Therefore, in future research, the methods of eliminating such violations, for example, with the help of physics knowledge, will be interesting and worthy to investigation.

Acknowledgment

This work was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme, of the grant DESCARTES.

References

[1] "Trends in electric cars – Global EV Outlook 2024 – Analysis," IEA. Accessed: Sep. 23, 2024. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024/trends-in-electric-cars

[2] "Electric Vehicle Integration into Power Grids," ENTSO-E, Mar. 2021. Accessed: Feb. 21, 2025. [Online]. Available: https://www.entsoe.eu/2021/04/02/electric-vehicle-integration-into-power-grids/

[3] J. Ji *et al.*, "OmniSafe: An Infrastructure for Accelerating Safe Reinforcement Learning Research," *Journal of Machine Learning Research*, vol. 25, no. 285, pp. 1–6, 2024, doi: 10.5555/3722577.3722862.

[4] 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, Oct. 2023, doi: 10.1016/j.apenergy.2023.121490.

[5] H. Li, Z. Wan, and H. He, "Constrained EV Charging Scheduling Based on Safe Deep Reinforcement Learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427–2439, May 2020, doi: 10.1109/TSG.2019.2955437.

[6] H. Yang, Y. Xu, H. Sun, Q. Guo, and Q. Liu, "Electric Vehicles Management in Distribution Network: A Data-Efficient Bi-level Safe Deep Reinforcement Learning Method," *IEEE Transactions on Power Systems*, pp. 1–15, 2024, doi: 10.1109/TPWRS.2024.3394398.

[7] G. Chen, L. Yang, and X. Cao, "A deep reinforcement learning-based charging scheduling approach with augmented Lagrangian for electric vehicles," *Applied Energy*, vol. 378, p. 124706, Jan. 2025, doi: 10.1016/j.apenergy.2024.124706.

[8] Z. Zhang, R. Rigo-Mariani, N. Hadjsaid, and Y. Xu, "Model-free safe deep reinforcement learning for grid-to-vehicle management considering grid constraints and transformer thermal stress," *Engineering Applications of Artificial Intelligence*, vol. 162, p. 112529, Dec. 2025, doi: 10.1016/j.engappai.2025.112529.

[9] IEC, *Power transformers – Part 7: Loading guide for mineral-oil-immersed power transformers*, IEC 60076-7:2018, Jan. 12, 2018. Accessed: Dec. 09, 2024. [Online]. Available: https://webstore.iec.ch/en/publication/34351

[10] T. Su, T. Wu, J. Zhao, A. Scaglione, and L. Xie, "A Review of Safe Reinforcement Learning Methods for Modern Power Systems," *Proceedings of the IEEE*, vol. 113, no. 3, pp. 213–255, Mar. 2025, doi: 10.1109/JPROC.2025.3584656.

[11] J. Achiam, D. Held, A. Tamar, and P. Abbeel, "Constrained Policy Optimization," in *Proceedings of the 34th International Conference on Machine Learning*, PMLR, Jul. 2017, pp. 22–31. doi: 10.48550/arXiv.1705.10528.

[12] T.-Y. Yang, J. Rosca, K. Narasimhan, and P. J. Ramadge, "Projection-Based Constrained Policy Optimization," presented at the Eighth International Conference on Learning Representations, Apr. 2020. doi: 10.48550/arXiv.2010.03152.

[13] Y. Liu, J. Ding, and X. Liu, "IPO: Interior-Point Policy Optimization under Constraints," *AAAI*, vol. 34, no. 04, pp. 4940–4947, Apr. 2020, doi: 10.1609/aaai.v34i04.5932.

[14] L. Zhang *et al.*, "Penalized Proximal Policy Optimization for Safe Reinforcement Learning," presented at the Thirty-First International Joint Conference on Artificial Intelligence, Jul. 2022, pp. 3744–3750. doi: 10.24963/ijcai.2022/520.

[15] F. Palacios, "Electric Power Consumption." Aug. 2022. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/electric-power-consumption

[16] E. Wong, "Electric Vehicle Charging Station Usage (July 2011 - Dec 2020)." Feb. 08, 2021. [Online]. Available: https://data.cityofpaloalto.org/dataviews/257812/electric-vehicle-charging-station-usage-july-2011-dec-2020/

[17] I. Daminov, R. Rigo-Mariani, R. Caire, A. Prokhorov, and M.-C. Alvarez-Hérault, "Demand Response Coupled with Dynamic Thermal Rating for Increased Transformer Reserve and Lifetime," *Energies*, vol. 14, no. 5, Art. no. 5, Jan. 2021, doi: 10.3390/en14051378.

979-8-3315-8138-1