Reinforcement learning based charging management for electric vehicle fleet to reduce transformer overloading in distribution networks

Yuzhuo Fu*, Detlef Schulz
Electrical Power Systems
Helmut Schmidt University / University of the Federal Armed Forces Hamburg
Hamburg, Germany
*yuzhuo.fu@hsu-hh.de

Dennis Salvador Versen*, Marcus Stiemer
Theory of Electrical Engineering and Numerical Simulations
Helmut Schmidt University / University of the Federal Armed Forces Hamburg
Hamburg, Germany
*versend@hsu-hh.de

***Abstract* – With the widespread integration of electric vehicles (EV) in distribution networks (DN), the distribution network operator faces new challenges. Uncontrolled charging processes of a large amount of EVs can result in many problems in power grids such as power quality issues and transformer overloading. Recently, many model-free methods based on Reinforcement Learning (RL) are proposed for EV charging management, which finds the optimal policy through the interaction between agent and environment. This paper investigates the performance of the RL-based EV charging management on relieving transformer overloading with finer measurement data (1-minute resolution) comparing the previous work with 15-minutes resolution. In addition, the states of the system in the past 5 minutes are sent together to the agent for a better perception of the future dynamics. The simulation results show that the proposed charging strategy works well with finer time resolution.**

Keyword* – *charging management, reinforcement learning, electric vehicle, time resolution

NOMENCLATUR

DN	Distribution Network
DQN	Deep-Q-Network
EV	Electric Vehicle
MDP	Markov Decision Process
RES	Renewable Energy Resource
RL	Reinforcement Learning
SOC	State of Charge
V2G	Vehicle-to-Grid
PV	Photovoltaic
s_t	state of the system at the time t
Λ_t	transformer load ratio
$P_{\mathrm{H},t}$	cumulated load of all households
$P_{\mathrm{EV},t}^{\max}$	cumulated charging demand of EVs
$E_{\mathrm{EV},t}$	cumulated energy requirement of EVs
T_{id}	intraday time index
a_t	charging factor decided by the agent
$P_{\mathrm{EV},t}$	cumulated charging power of EVs
r_t	reward at the time t
$r_{\Lambda,t}$	penalty term in the reward function
$\Lambda_{\max}$	threshold of the transformer loading
$\overline{A_{\mathrm{EV}}}$	average charging factor
$\bar{\Lambda}$	average transformer loading
N_{Λ}	number of threshold violation of the transformer limit
E_{fail}	unsatisfied energy demand of EVs

I. INTRODUCTION

The use of fossil fuels creates many environmental problems. The Intergovernmental Panel on Climate Change has found that emissions from fossil fuels are the dominant cause of global warming [1]. Germany produced 762 million tons of carbon dioxide in 2021, a share of roughly 30 % comes from power sector and 20 % is caused by transport sector [2]. From the perspective of electricity generation, the integration of RES like wind and photovoltaic systems produced about 225 TWh electricity energy in 2021 and the share of renewables in the public net electricity generation is about 45.7 % [3]. In contrast to the electricity generation, the change on electricity consumption side is much slower, especially in the transportation sector. The share of renewables rose to 6.8 % in regards to transport in Germany but the main contribution comes from biofuels not the directly use of clear electricity from renewable energy resources (RES) [4]. Nowadays, the share of electric and hybrid cars reaches nearly 50 % (full-EV 14 %, hybrid-EV 29 %) in all new registrations [5]. However, it will take a long time to retire the large number of traditional cars and increase the overall share. The widespread integration of EV can affect the power grid in a negative way, such as the significant reduction in the system's power quality and overloads in the distribution lines and transformers. Van der Burgt [6] found that a transformer overload can be observed at a low EV penetration level of 25 % with simultaneous charging. For this purpose, a charging strategy to cooperate the charging processes of EV fleet is needed.

Some optimization methods have been proposed for the cooperative charging in the literature such as dynamic and stochastic programming. The biggest disadvantage of optimization-based methods is that accurate models are required to describe charging behaviour, energy demand and available flexibility [7]. However, the user behaviour is a random process, whose patterns are affected by a variety of uncertain factors such as family structure, income status, location and weather. Exact modelling of each individual user is nearly impossible. Nowadays, model-free approaches based on reinforcement learning have achieved great success in complex decision-making applications. The advantage of RL in Sequential Decision-Making problem can be summarized as the follows. The ability to learn the pattern and dynamics of stochastic processes through historical data avoids complex modelling of uncertain factors and makes the training process in a model-free way. In addition, a RL-agent after training can response to dynamic environment in real-time, which would be difficult for traditional optimization approaches in a large and complex system. Recently, many works related to RL-based charging management for EV fleets can be found in literature. In [8], a coordinated EV charging based on a policy gradient algorithm aiming to smooth out the load profile of a parking lot is proposed. In [9], a soft-actor-critic based method in combination with nodal multi-target characterization is proposed to schedule of large-scale EV in DN.

In this work, an intelligent charging management based on reinforcement learning, which is first presented in the previous paper [10] will be further developed. In order to manage the charging processes of EVs in a residential area, the objective is to satisfy the charging demands of end users as quickly as possible without exceeding the threshold given for transformer loading. Concretely, the RL-agent receives data from meters installed on the transformer and charging points and then sends a control signal to limit charging power. The system is dynamic due to the uncertainty of user power consumption and EV mobility behaviour. Then, the agent tries to take the optimal charging control in order to maximize the benefits of the entire system. The contribution of this work can be summarized as the follow. At first, finer time interval in the RL-environment is considered for the purpose to better modelling of dynamic of user behaviours. In order to investigate the impact of time interval on the performance of RL-agent, the simulation results for 1-minute and 15-minutes time steps are compared with each other. In addition, inspired by some applications of RL like Atari Breakout [11], the state of the past few time steps are also considered aiming to better reflect the trend of the system. The effects of these two improvements will be shown by comparison between different configurations and scenarios.

This paper is organized as follows: the whole system with assumptions and the mathematical description of the EV charging problem are presented in Section II. The RL algorithm and training process based on real world data is introduced in Section III. Then, a case study and the simulation results including comparison are shown in Sections IV and V. Finally, Section VI summarizes the paper.

II. MATHEMATICAL DESCRIPTION OF THE EV CHARGING PROBLEM

In this paper, we consider a DN in low voltage level, in which EVs belonging to private households need to be charged in a residential area. All EVs can only be charged at this area, the multi-location charging is not considered yet. Bidirectional charging (V2G) takes advantage of the storage capacity of the EV battery to relieve the imbalance between local electricity production and electricity consumption. The aim of this paper is firstly to avoid the overloading or reduce the loading of transformers in order to integrate more DERs or EVs in DN. So only the unidirectional charging is allowed in our assumptions, which is possible to be implemented within a shorter amount of time, than V2G applications. Furthermore, some typical DERs like photovoltaics are not considered in the simulation even through more and more private PV systems can be found in DNs. The reason is, the uncertainty of the system contains the load profiles of households and the mobility behaviours of EVs. If a RL-agent handles the dynamic and uncertainty of this system well, it can also in principle response to the additional randomness of photovoltaics.

The EV charging management developed in this work is a centralized control system, in which a communication among distribution network operator, EV users and RL-agent is necessary. The communication network is bases on a naive assumption regardless of privacy protection and availability of metering equipment. Firstly, the RL-agent is able to receive the measurement data at the level of transformer such as the current loading of the transformer. In addition, the usage data at the customer level like the load profile of households and the charging data of EVs are also send to the RL-agent. This is worth to noting that the information at customer level is at first aggregated and then send to the agent, which means the charging management only knows the cumulative value of the loads but not the measurement data of each household and EV. The advantage of such a communication system is that the dimensionality of input data is significantly reduced to an aggregated level, which makes the training process faster. The major disadvantage is that the centralized charging management is not able to control the charging process at an individual level in regard to the preference of each user.

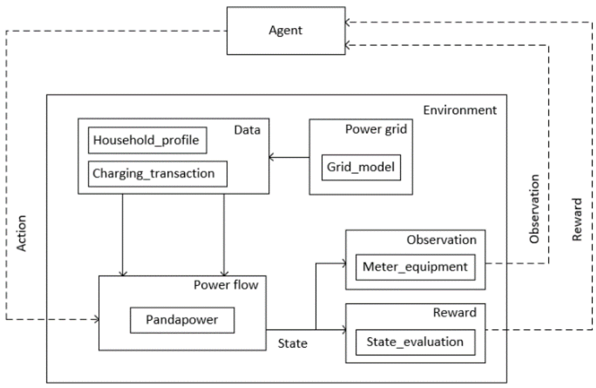

FIGURE 1: SIMULATION ENVIRONMENT [10].

The aim of the RL-agent is to satisfy the charging demands of end users as quickly as possible without exceeding the threshold given for transformer loading. The agent can receive the current information on the transformer and usage data at the aggregated customer level. The biggest challenge is that the agent is disable to predict the future user behaviour so a learning process is needed to capture the pattern of electricity consumption and mobility mode. For this reason, a reinforcement learning based EV charging management is used, which aims to maximize the future cumulative rewards through taking an optimal action according to the current state of system. The RL-agent developed in another work [10] is the basis of this paper, in which the definition of a Markov Decision Process (MDP) and the mathematical description of the entire system are introduced in detail. The environment built to train the RL-agent for EV charging management is shown in FIGURE 1. In following part, only the most important components of the MDP will be introduced.

The state space is the set of values obtained from measurements or provided actively by end-users. In the environment, the state s_t at time step t is described:

$$s_t = \{\Lambda_t, P_{\mathrm{H},t}, P_{\mathrm{EV},t}^{\max}, E_{\mathrm{EV},t}, T_{\mathrm{id}}\} \tag{1}$$

Λ_t defines the transformer load ratio caused by the load of all feeders and cable losses. $P_{\mathrm{H},t}$ presents the cumulated load of all households. T_{id} is the intraday time index used to reflect temporal relevance of the environment. $E_{\mathrm{EV},t}$ is the cumulated energy requirement of all EVs and $P_{\mathrm{EV},t}^{\max}$ is the cumulated charging power demand of all EVs which is dependent on the individual state of charge (SOC) of the EV or the customer preference.

The action space consists of all possible control signals of the RL-agent. In this MDP, the action a_t at time step t is modeled as a factor that determines the acceptable EV charging power by multiplying it with the desired charging demand $P_{\mathrm{EV},t}^{\max}$:

$$P_{\mathrm{EV},t+1} = a_t P_{\mathrm{EV},t}^{\max} = \sum_{i=1}^{N_{\mathrm{EV}}} a_t P_{\mathrm{EV},i,t}^{\max} \tag{2}$$

In order to keep the dimension of the action space within a reasonable range, we discretized the charging factor a_t into 6 stages: 0, 0.2, 0.4, 0.6, 0.8, 1. The reward function r_t is used to evaluate the effectiveness of a_t in the state s_t in MDP, it is formulated as follows:

$$r_t = a_t \cdot 5 + r_{\Lambda,t} \tag{3}$$

$$r_{\Lambda,t} = \begin{cases} -200 & \text{if } \Lambda_t > \Lambda_{\max} \\ 0 & \text{else} \end{cases} \tag{4}$$

$a_t \cdot 5$ indicates that the RL-agent seeks to provide the maximum charging power at all time steps so that the energy demand of EVs can be satisfied as soon as possible. $r_{\Lambda,t}$ is defined as a penalty term, once the prescribed threshold of the transformer loading $\Lambda_{\max}$ is exceeded. In this work, $\Lambda_{\max}$ is set to be 70 % according to the base-load, which will be explained in chapter V.

III. REINFORCEMENT LEARNING ALGORITHM

The state transition probability of the above environment is difficult to be accurately modelled due to the high uncertainty of many factors such as the load profiles and mobility behaviour of end users. On this account, a well-known model-free reinforcement learning algorithm deep Q-network (DQN) is used to solve the build MDP. The configuration of the DQN and the correspondent training process can be found in the work [10]. DQN is easy to implement and can show a high efficiency in small observed spaces with discrete action spaces even though they converge slow. Instead of using a Q-Table like in Q-Networks the DQN-Algorithm uses a neural network to predict the action outcome in specific states, which causes the best reward gain in the current step just as in the future steps. Similar to the MDP an adapted Bellman-Equation is used to update the neural network. Instead of the discounted retracted future reward, the discounted next step which recursively contains the assumable future reward is used go estimate the score in the future.

As mentioned above a realistic estimation of the behaviour of the load curve is hard to give due to many volatile factors. For this reason, mathematic models often fail or give just a vague proximation of the reality. RL-Algorithms are able to see patterns in the history of load profiles. Thus, one idea is to take the past observation spaces into account. The information through the feedback of the past observed states can help to derive the optimal charging factor for the next timestep without knowing the future state. One advantage of this method is that the agent can react without knowing the income of new loading inquiries. That makes it easier to implement the agent to the real world. According to that, our state is not only a depiction of the status quo but rather a time series of the last observed states.

$$\boldsymbol{S_t} = \{s_t, s_{t-1}, s_{t-2}, \ldots, s_{t-n}\} \tag{5}$$

The factor n will define how far in the past the states will be considered. One object for this work is to regard different n´s to see how this will make an impact on the performance of the agent.

IV. CASE STUDY

A. Grid Model

The benchmark network CIGRE Testbench LV is used in this paper as a test system, which was specially developed for the investigation of integration of DERs in DNs [12]. This network model in a residential area consists of 18 buses, 85 household loads and 17 lines, which is fed by a single 400 kVA, 20 kV / 400 V transformer. The aim of the simulation is not to test the EV charging management system with many different scenarios regards to the EV penetration level but to investigate the performance of charging management with finer time step and past information, so we consider a fixed number of charging points in this area. A total of 60 private charging points is allocated to the 85 households, which means about 70 % EV penetration level.

B. Data Preparation

The RL-agent is trained on a real-world power usage dataset consisting of household load profiles and EV charging transactions. The load data of households is taken from the UMass Smart Dataset, which collected power consumption of 114 single-family apartments for three years [13]. The home EV charging transactions are collected from Electric Nation with nearly 700 EV owners taking part in an 18-month trial, which records the charging station ID, arrival time, departure time and consumed energy during the session [14]. Both data banks collected usage profiles in 1-minute interval. The finer measurement data makes the performance comparison of RL-agent with different time intervals possible.

C. Performance metrics

This paper uses the following four performance metrics, which are proposed in our previous work [10] , to evaluate the RL-based charging management system: average charging factor $\overline{A_{\mathrm{EV}}}$, average transformer loading $\bar{\Lambda}$, N_Λ is the number of threshold violation of the transformer limit $\Lambda_{\max}$ and E_{fail} represents the amount of unsatisfied energy demand.

D. Simulation Setup

In order to investigate the impact of time interval on the performance of charging management, the RL-agent is trained with two different time resolution: 15-minute and 1-minute. The original configuration of the simulation (called the first scenario) depends on 15-minute time interval with feedback of only the current state, which corresponds to the configuration of our work [10]. Theoretically, the finer measurement data show more dynamic information, which is useful for the agent to learn the user's power consumption pattern and even better predict the trend. Considering the fineness of available data, both the second and the third scenario of the simulation setup are based on the 1-minute resolution while the third scenario observes all states of the past 5 minutes. As a result, there are a total of three different configurations of the simulation setup, which can be seen in TABLE Ⅰ.

TABLE Ⅰ: THREE SCENARIOS OF THE SIMULATION SETUP.

Scenario	Simulation Setup	
	Time Resolution	***Feedback of the past information***
1	15-minute	No
2	1-minute	No
3	1-minute	Yes

Although the full year of usage profiles are available in the data bank, the training data is defined only as the power consumption of the households and EV charging transaction for the month January. The main reason is the calculation burden especially for the simulation with 1-minute time resolution, which results in 525.600 time steps for a year-long simulation. The power consumption data in February is chosen as the test data to evaluate the performance of the charging algorithm. The environment is built on basis of OpenAI Gym and the RL-agent is implemented using TensorFlow.

V. SIMULATION RESULTS

E. *Uncontrolled charging*

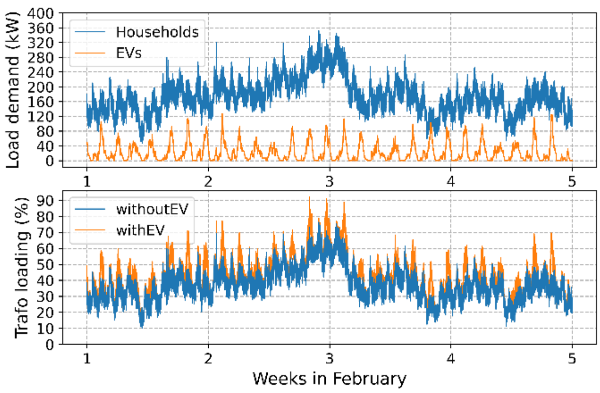

FIGURE 2: LOAD DEMAND AND TRANSFORMER LOADING WITH UNCONTROLLED CHARGING IN FEBRUARY.

It is worth to show the transformer loading and charging profiles of EVs without any management at first. Then, the effectiveness of the RL-based charging management can be better evaluated. The impact of 60 EVs can be seen in FIGURE 2 where the simulation results for February are displayed and charging coincides with peak household consumption at many time steps. The accentuated peak load leads to around 10 %-20 % increase in transformer peak loading compared to the base load without EVs. In contrast to our previous work [10], the usage pattern is more difficult to be captured because of a large load variance. In addition, the peak load caused by households at some time steps exceeds the defined threshold of the transformer loading $\Lambda_{\max}$. That will result in the agent being penalized for any action it took, which brings great challenge to the training of the agent.

F. *The impact of finer measurent data on the performance*

Limited by computational power, we only train 5 agents for each setting in the TABLE Ⅰ. Then the agent with the best performance among them is chosen to compare with each other. Before comparing the performance of agents with different time resolution, the performance improvements compared to uncontrolled charging are first evaluated (see TABLE Ⅱ). The impact of the two trained agents with 15-minute (the first scenario) and 1-minute (the second scenario) time resolution respectively on the transformer loading is shown in FIGURE 3. It shows that some load peaks in the both scenarios are reduced but the behaviours of both agents are not same.

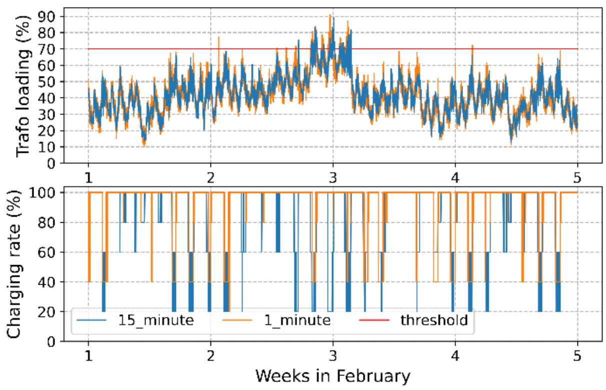

FIGURE 3: IMPACT OF AGENTS WITH DIFFERENT CONFIGURATIONS REGARDING TO THE TIME RESOLUTION ON TRANSFORMER LOADING AND THE CHARGING RATE.

It is noticeable that the load variance of the scenario with 1-minute resolution is much larger than the scenario with 15-minute resolution. The reason would not be the agent's policy but more likely the finer measurement data itself. It can be confirmed that the agent of the second scenario fails at some time steps in regard to the threshold of the transformer loading such as the load peak at the start of the second week. In contrast, the agent of the first scenario seems like to have a better performance in relation to the number of threshold violation of the transformer limit if we ignore that there are more load peaks in the finer measurement data. The both agents can sometimes even limit the transformer load to just around the threshold, which shows that they find the best compromise between satisfying the power demand of EVs and avoiding the transformer overloads. However, if we take the simulation results for the last two weeks with relatively low power consumption into account, the policy of the second agent is closer to optimal. From the FIGURE 3, it can be seen that during the last two weeks the total power consumption is low enough that the EVs are able to be charged with the maximum rate at any time without causing many overloading. For this reason, the behaviour of the second agent is more aggressive and tries to keep the charging rate at a higher value. The policy of the first agent is confusing during this time, in which it always tries to reduce the EV charging rate to a low level even through the current transformer loading is far from the threshold.

TABLE II compares the performance of the two RL-agents using the performance metrics defined in Section IV. Only small differences between the two agents can be observed regards to the average transformer loading $\overline{\Lambda}$ and the number of threshold violation N_{Λ}. Due to the more conservative policy, the average charging factor $\overline{A_{\mathrm{EV}}}$ of the first scenario with 15-minute time resolution deceases by around 4 %. The number of threshold violation of the transformer limit can't be directly compared because of the different time resolution. Therefore, we calculate the mean value of each time block in 15-minute interval in order to unify the time resolution. It is worth noting that significant more failed working points (exceed the limit value 70 %) of the second scenario can be seen in the FIGURE 3 than the number in the TABLE II because of the calculate of mean value for each 15-minute time interval. In terms of the amount of unsatisfied EV energy demand, the index E_{fail} shows that a total of 116 kWh is not successfully charged into all EVs during a month, which accounts for only about 1.8 % of total energy demand of about 14.743 kWh. With almost the same performance on reducing transformer loading, only 83 kWh energy is not successfully charged into the EVs by the second scenario.

TABLE II : COMPARING THE PERFORMANCE OF THE AGENTS WITH DIFFERENT TIME RESOLUTION.

Scenario	Performance Metrics				
	$\overline{A_{EV}}$	$\overline{\Lambda}$	N_{Λ} - 70%	N_{Λ} - 80%	E_{fail}
uncontrolled	100%	42.03%	69	15	0 kWh
1	88.73%	41.98%	48	8	116.1 kWh
2	92.92%	41.85%	46	8	83.1 kWh

In summary, the finer measurement data with more detailed and dynamic information seems like to result in a better performance but it might be caused by the finer measurement data rather than the policy. Due to more frequent interventions, it becomes easier for the agent with 1-minute time resolution to find the compromise between transformer overloading and insufficient energy of EVs. However, the simulation results indicate that both RL-agents converge to a local optimum but not the global optimum.

G. The impact of the past information on the performance

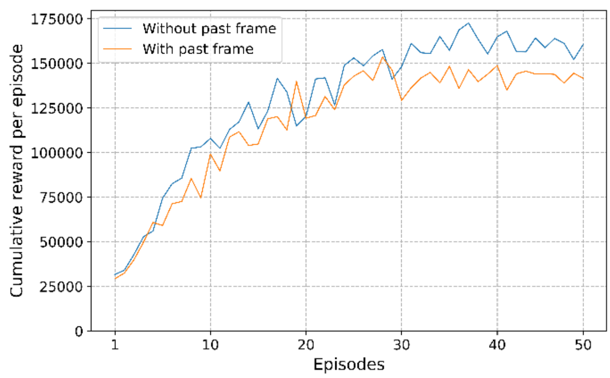

FIGURE 4: CUMULATED REWARD PER EPISODE OF THE TRAINING PROCESSES OF THE SECOND UND THE THIRD SCENARIO.

The states of the past few time intervals contain more dynamic information of the system that principally enables the agent to control the charging rate with more consideration on the possible future states. The training processes of the two scenarios with and without the feedback of states in the past 5 minutes are shown in FIGURE 4. For both scenarios, 5 agents are trained with the same initial conditions and the average cumulative reward per episode is calculated.

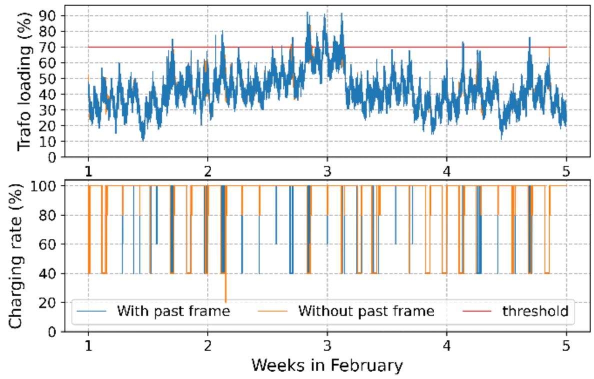

FIGURE 5: IMPACT OF THE AGENTS WITH THE SECOND AND THE THIRD SCENARIO ON THE TRANSFORMER LOADING AND CHARGING RATE.

Unexpectedly, the agent considering the past states do not have a better performance than the agent just observing the current state. In most episodes, the cumulative rewards of the third scenario are lower than the second scenario. In addition, the gap is expanded at the beginning and then stable at a certain value towards the end of the training. It is noticeable that the training curve of the agent with feedback of the past states seems to be stretched along the x-axis, which means it could coverage to a better policy if we increase the training episodes. Similar to the previous chapter, the curve of transformer loading, the charging rate and the corresponding performance metrics are shown in the FIGURE 5 and TABLE III respectively. There is no significant difference between the two scenarios regarding to the average charging factor $\overline{A_{\mathrm{EV}}}$ and the average transformer loading $\overline{\Lambda}$, which means both of them try to keep the charging rate at a high level. However, the more frequent thresholds exceeding N_{Λ} and the larger amount of energy shortage E_{fail} in EVs indicates that the third Agent reduces the charging power not at the optimal moments.

TABLE III: COMPARING THE PERFORMANCE OF THE AGENTS WITH AND WITHOUT THE PAST INFORMATION.

Scenario	Performance Metrics				
	$\overline{A_{EV}}$	$\overline{\Lambda}$	N_{Λ} 70%	N_{Λ} - 80%	E_{fail}
2	92.92%	41.85%	46	8	83.1 kWh
3	92.48%	41.85%	53	10	101.9 kWh

By comparing the two scenarios, it can be shown that the agent is unable to utilize the past information properly and then track the system dynamics for a better perception. In addition, the past information even negatively affects the agent's decision-making, which makes the training process more difficult and ineffective.

VI. CONCLUSIONS

In this paper, the RL-based EV charging management is further developed, which is proposed in our previous work. A slight improvement on the performance can be seen once the measurement data becomes finer (from 15-minutes resolution to 1-minute resolution). However, the agent can't extract the dynamic information from the past few time intervals for a better performance. The reason might be the definition of the state space (the input layer of neural network). Some external information such as the intraday time index T_{id} restricts the agent's experience to be applied only at that time step, which can't be generalized for the other scenarios. For the future work, a wider range of sensitivity analysis is needed.

From the perspective of the optimization objective, avoid overloading of transformers is not a suitable application to maximize the potential of RL-algorithms. Theoretically, a simple controller is also able to limit the transformer loading, which only needs to calculate the power flow of the grid according to charging power requirement of EVs for the next time step and then reduce the charging rate if needed. The advantage of RL is that it can adjust the current action in a random and unpredictable environment to maximize total revenue over a certain time. For the future work, new challenges such as collective self-consumption and energy management regarding to cost will be considered. In addition, a single control signal applying on all charging stations is not efficient. Each EV customer has its own mobility- and charging behaviour, which needs a customized individual charging management. For this reason, a distributed energy management with multi-agent-system will be developed in the future.

Acknowledgement

This research paper is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr which we gratefully acknowledge dtec emob [UT 7001 – Digitalization and E-Mobility].

References

[1] IPCC, „Climate Change 2022-Mitigation of Climate Change," 2022.

[2] „Emissionsquellen," Umwelt Bundesamt, 2021. [Online]. Available: https://www.umweltbundesamt.de/themen/klima-energie/treibhausgas-emissionen/emissionsquellen#energie-stationar.

[3] „Public Net Electricity Generation in Germany in 2021," Frauenhofer, 2022. [Online]. Available: https://www.ise.fraunhofer.de/en/press-media/news/2022/public-net-electricity-in-germany-in-2021-renewables-weaker-due-to-weather.html.

[4] Umwelt Bundesamt, „Erneuerbare Energie im Verkehr," 2022. [Online]. Available: https://www.umweltbundesamt.de/daten/energie/erneuerbare-energie-im-verkehr#:~:text=%E2%80%9EAnteil%20erneuerbarer%20Energien%20am%20%E2%81%A0,wieder%20auf%206%2C8%20%25..

[5] Clear Energy Wire, „Germany reaches one-million electric vehicles target with half a year delay," 2021. [Online]. Available: https://www.cleanenergywire.org/news/germany-reaches-one-million-electric-vehicles-target-half-year-delay.

[6] J. v. d. Burgt, B. Wille-Haussmann und A. N. Andersen, „Grid Impact of Charging Electric Vehicles; Study Cases in Denmark, Germany and The Netherlands," *IEEE Eindhoven PowerTech,* 2015.

[7] N. Sadeghianpourhamami, J. Deleu und C. Develder, „Definition and Evaluation of Model-Free Coordination of Electrical Vehicle Charging With Reinforcement Learning," *IEEE Transactions on Smart Grid,* 2020.

[8] Y. Jiang, Q. Ye und B. Sun, „Data-Driven Coordinated Charging for Electric Vehicles with Continuous Charging Rates: A Deep Policy Gradient Approach," *IEEE INTERNET OF THINGS JOURNAL,* 2021.

[9] J. Jin und Y. Xu, „Optimal Policy Characterization Enhanced Actor-Critic Approach for Electric Vehicle Charging Scheduling in a Power Distribution Network," *IEEE TRANSACTIONS ON SMART GRID,* 2021.

[10] Y. Fu, D. S. Versen und A. Stadler, „AI-based charging management for the integration of electric vehicles using a reference low voltage grid in Hamburg," in *NEIS*, Hamburg, 2022.

[11] Keras, „Deep Q-Learning for Atari Breakout," 2020. [Online]. Available: https://keras.io/examples/rl/deep_q_network_breakout/.

[12] International Council on Large Electric Systems, „Benchmark Systems for Network Integration of Renewable and Distributed Energy Resources," CIGRE, 2014.

[13] UMassTraceRepository, „Smart* Data Set for Sustainability," 2017. [Online]. Available: https://traces.cs.umass.edu/index.php/smart/smart.

[14] Western Power Distribution, „Electric Nation Data," 2022. [Online]. Available: https://www.westernpower.co.uk/electric-nation-data.