Citation information: DOI: 10.1109/UPEC61344.2024.10892464, 2024 59th International Universities Power Engineering Conference (UPEC)

Active Voltage Control and Demand Peak-Clipping using Soft Actor-Critic Method

Thet Paing Tun
Department of Electronic and Electrical Engineering
Brunel University London
London, United Kingdom
thetpaing.tun@brunel.ac.uk

Ioana Pisica
Department of Electronic and Electrical Engineering
Brunel University London
London, United Kingdom
ioana.pisica@brunel.ac.uk

***Abstract*— This study presents a proprietary Soft Actor-Critic (SAC) algorithm for power distribution network active voltage control and demand peak-clipping, including on-load-tap-changing and direct load control. Voltage violation and transformer loads, determine the system's state, while actions comprise transformer tap changing, heat pump setting temperature, and Vehicle-to-Grid at 30% demand flexibility and rewards are designed to prevent voltage violations and transformer overload. After testing with various load levels, the trained agent can optimally select actions and minimise violations, however the model will be tuned with different parameters and demand shift and energy storage are suggested to be added in future work.**

Keywords—Deep Reinforcement Learning, Soft Actor-Critic, Vehicle-to-Grid, Electric Vehicle, Heat Pump

I. Introduction

The evolving landscape of voltage control and energy management in active distribution grids is currently experiencing notable progress. This progress is primarily driven by the incorporation of cutting-edge technologies and methodologies. The driving force behind this transition is the imperative to optimise grid operations, improve energy efficiency, and maintain grid stability in the face of growing demands and the ever-changing landscape of energy sources.

The significance of Demand Side Management (DSM) strategies, encompassing demand response and smart loads, has been thoroughly examined by Pálenský & Dietrich (2011), who underscored the pivotal role of intelligent energy systems in optimising grid operations [1]. Building upon this notion, the investigation conducted by Schellenberg et al. (2019) delved into the realm of grid-edge technologies, shedding light on the promising capabilities of heat pumps and thermal energy storage systems in bolstering the flexibility of the grid [2]. In addition, Fischer & Madani (2017) studied the effectiveness of heat pumps in improving energy efficiency and grid stability by emphasizing the deployment of heat pumps and their ability to enhance energy efficiency and grid stability, which is further supported by the implementation of predictive control mechanisms [3]. Moreover, the scholarly work conducted by Crawley et al. (2022) highlighted the projected surge in electricity consumption resulting from the extensive implementation of heat pumps and electric heaters and the outcome of the study further accentuated the significance of employing effective demand response strategies to ensure optimal energy usage [4]. .

The integration of renewable energy sources into the existing grid infrastructure and net-zero electrification in building heating systems and the transportation sector have emerged as a topic of great interest and importance in recent years. This phenomenon presents a unique set of opportunities and challenges for grid management. The study conducted by Schellenberg et al. (2018) delved into the topic of operational optimisation of heat pump systems emphasising the significance of employing genetic algorithms as a means to enhance energy efficiency highlighting the value of utilising advanced optimisation techniques to maximise the performance of heat pump systems [5].

The proposed study utilised the novel Soft Actor Critic (SAC), a sophisticated reinforcement learning method, to improve electrical distribution network efficiency and dependability to resolve the potential challenges incurred by complicated modern distribution networks due to the net zero electrification in domestic heating and transportation.

The scope of the proposed work covers the fundamental voltage management and peak-clipping of hiking electricity demand in end-user distribution network by means of the optimal operation of 33/11kV on-load-tap-changer and strategical direct load control of setting temperature values of heat pumps and vehicle-to-grid discharging power of electric vehicles owned by consumers participating in demand flexibility with the constraints of fixed participation rate of 30% but the future analysis shall cover the scenarios with varying participation rates.

II. Related Work

Deep Reinforcement Learning (DRL) has emerged as a powerful instrument for enhancing power system efficacy, offering novel resolutions to the complexities encountered within the energy industry. Scholars have conducted investigations into several applications of Deep Reinforcement Learning (DRL) in the optimisation of power systems, encompassing maintenance planning, real-time control, and dispatching techniques.

In their study, Yang and Yao (2021) proposed an optimisation approach for decision-making in power equipment maintenance plans [6]. The authors focused on the use of Deep Reinforcement Learning (DRL) in the context of intelligent power generation control and grid management. Shi et al. (2022) emphasised the significance of Deep Reinforcement Learning (DRL) in effectively optimising power system operations by focusing on day-ahead optimal dispatching of hybrid power systems [7].

The application of Deep Reinforcement Learning (DRL) in optimising power system operations has been expanded to include emergency control strategies. In their study, Huang et al. (2020) devised adaptive emergency control strategies utilising Deep Reinforcement Learning (DRL) to address intricate power system difficulties and their findings

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/UPEC61344.2024.10892464, 2024 59th International Universities Power Engineering Conference (UPEC)

showcased the efficacy of DRL in augmenting grid stability [8]. In addition, Zhu et al. (2022) utilised Deep Reinforcement Learning (DRL) to attain the most efficient functioning of a microgrid equipped with hydrogen storage [9]. This demonstrates the adaptability of DRL in tackling diverse energy system optimisation objectives.

Furthermore, the utilisation of DRL techniques has greatly enhanced the dynamic optimisation of integrated energy systems. In their study, Tang et al. (2022) examined the dynamic optimisation of integrated energy systems on the grid, taking into account peak-shaving demand. They emphasised the efficacy of Deep Reinforcement Learning (DRL) in enhancing energy efficiency and demand response tactics. Dong et al. (2022) introduced an energy scheduling technique utilising Deep Reinforcement Learning (DRL) to reduce operational expenses, carbon emissions, and improve the dependability of power supply [10].

The utilisation of Deep Reinforcement Learning (DRL) in the optimisation of power system operations also encompasses the integration of renewable energy sources and the management of grids. Futakuchi et al. (2021) examined the planned functioning of a wind farm equipped with a battery system using Deep Reinforcement Learning (DRL), showcasing the capacity of DRL to enhance the efficiency of renewable energy systems [11].

Overall, the application of Deep Reinforcement Learning (DRL) to optimise power system operations has the potential to be a beneficial approach for enhancing grid efficiency, stability, and sustainability. Scientists and professionals have used advanced learning algorithms like Deep Reinforcement Learning (DRL) to develop intelligent solutions for dealing with the complex and dynamic aspects of modern power systems.

III. Methodology

Python is the main programming language used for both power network creation and Deep Reinforcement Learning model (DRL). The Pandapower library package is utilised for the implementation of the power network in the environment, while pytorch serves as the primary tool for implementing DRL.

A. Power Network Modelling

The base power network is referred to the topology of a 11kV feeder of Hillingdon substation provided by Scottish Southern Energy Networks (SSEN) in Fig. 1. Among various outlets of Hillingdon substation, the feeder with longest length is selected to explore the effectiveness of the proposed DRL model. As the original line and transformer parameters are based on the existing loads without significant penetration of electric vehicle chargers and heat pumps, the transformer rating (100% rated capacity) was determined by summing the domestic load, heat pump, and electric vehicle (EV) consumption. To ensure operational flexibility and accommodate future enhancements by the trained voltage and energy management model, the transformer's threshold rating was set at 90% of its rated capacity.

The line impedance values were adjusted to ensure a voltage of around 0.92 per unit (pu) at the terminal buses during periods of high electricity consumption, taking into account the complete integration of heat pumps and electric vehicles. The improvements were made to fit the new loads in the network, with the purpose of allowing for subsequent refinements. This was necessary because the initial line impedance caused low voltage violations at roughly 0.7 pu. This approach allows the voltage and energy management model to effectively resolve any remaining difficulties and significantly enhance network performance by achieving a minimum voltage value greater than 0.94 per unit.

The EV charging pattern and the domestic demand were based on the Lower Carbon London project database [12][13] while non-domestic data were based on [14]. The allowed voltage limits were (0.94 - 1.06pu) for 11kV according to [15] and (0.94 - 1.1pu) for 0.4kV based on [16]. According to [17], transformer loading must be between 70% and 100% to maintain duty factor 1. However EV and Heat Pump integration may make it difficult to keep transformer loading below 70% to keep duty factor less than 1. To compensate all of those constraints, the maximum loading limit was selected as 90% for this study.

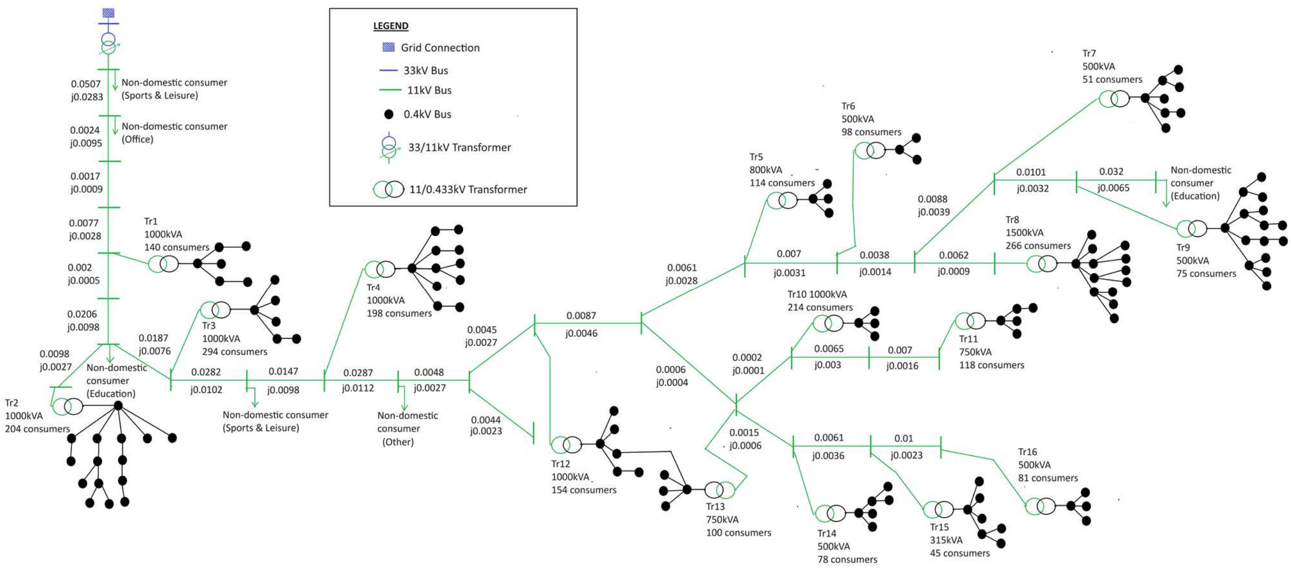


Fig. 1. Reference distribution network

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/UPEC61344.2024.10892464, 2024 59th International Universities Power Engineering Conference (UPEC)

B. Deep Reinforcement Learning Development

Soft Actor-Critic (SAC) algorithm is an off-policy methodology that integrates the advantages of actor-critic techniques with entropy regularisation in order to promote exploration in reinforcement learning problems. SAC, in contrast to conventional actor-critic algorithms, optimises both the anticipated return and the entropy of the policy, leading to a more varied and resilient exploration.

The agent aims to find the optimal policy in purpose of augmenting the achieved reward.

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma . \mathbf{E}_{s_{t+1} \sim p}[V_{target}(s_t, a_t)] \tag{1}$$

In the context of updating the Q-function and policy, the Soft Bellman Backup Operator assumes a vital role.

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma . \mathbf{E}_{s_{t+1} \sim p}\Big[Q(s_{t+1}, a_{t+1}) - \alpha \log\Big(\pi_{\emptyset}(a_{t+1}|s_{t+1})\Big)\Big] \tag{2}$$

and action at time step 't':

$$a_t = \mu_\theta(s_t) + \epsilon_t . \sigma_\theta(s_t) \tag{3}$$

where $Q(s_t, a_t) =$ Target Q value for given state − action pair at timestep t
$s_t, a_t =$ state and action at timestep t
$r =$ reward
$V_{target}(s_t, a_t) =$ Target state − value function for state (s_t)
$\mu_\theta =$ mean of action distribution output
$\epsilon_t =$ random noise sampled from normal distribution
$\mathbf{E} =$ expected long − term reward

SAC algorithm is pictorially represented by Fig. 2. In SAC algorithm, the actor network establishes a mapping between states and actions, which serves as the guiding principle for the agent's decision-making process. Critic networks play a crucial role in guiding the learning process of the actor network by providing feedback on the adequacy of the policy. SAC utilises twin critic networks to autonomously assess the values of state-action, hence improving the stability and resilience of the learning process and the loss function of critic network can be expressed as [18]:

$$\mathcal{L}(\theta) = \mathbf{E}_{s_t \sim D}\Big[\mathbf{E}_{a_t \sim \pi_\theta}[\alpha . \log(\pi_\theta(a_t|s_t)) - Q_\pi(s_t, a_t)]\Big] \tag{4}$$

The policy loss is:

$$\mathcal{L}(\emptyset) = \mathbf{E}_{(s_t, a_t, r_t, s_{t+1}) \sim D}\left[\frac{1}{2}\Big(Q_\emptyset(s_t, a_t) - Q_{target}(s_t, a_t)\Big)^2\right] \tag{5}$$

where $\mathcal{L}(\emptyset) =$ Loss with respect to Q function network parameter
$\mathcal{L}(\theta) =$ Loss with respect to policy network parameter
$D =$ Replay buffer containing experience tuples
$\pi_\theta(a_t|s_t) =$ Probability Density function of action and state under policy π_θ
$(s_t, a_t, r_t, s_{t+1}) =$ Experience tuples from Replay Buffer

In this case, the action space, observation space and reward are as described from Equation 6 to 13.

$$\mathcal{A} = a \in Z^3 | 0 \le a_i \le 16, \forall i \in 0,1,2 \tag{6}$$

where $\mathcal{A} =$ action space
$a[0] =$ Transformer tap position ($0 = -8$ tap, $16 = +8$ tap)
$a[1] =$ Vehicle − to − grid discharging (discharged power $= (a[1]/16) * 3$kW)
$a[2] =$ Setting temperature difference for heat pump (Reduced amount $= 23 - (a[2]/8)$

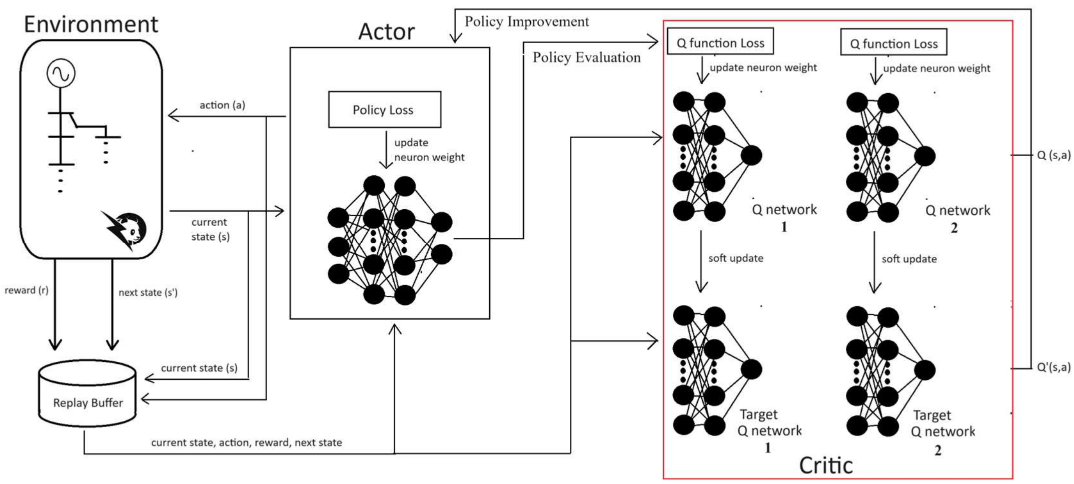


Fig. 2. Representation of Soft Actor-Critic Model

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/UPEC61344.2024.10892464, 2024 59th International Universities Power Engineering Conference (UPEC)

$$\mathcal{O} = \{o \in R^5\} \tag{7}$$

where $\mathcal{O} =$ observation space
$o[0] = 0.94 -$ lowest 11kV voltage within network
$o[1] =$ highest 11kV voltage within network $- 1.06$
$o[2] = 0.94 -$ lowest 0.4kV voltage within network
$o[3] =$ highest 0.4kV voltage within network $- 1.1$
$o[4] =$ Burden on highest loaded transformer (%) / 100%

Reward

$$p\_11\_1 = \begin{cases} -0.2 \times o[0], & o[0] < 0 \\ o[0], & o[0] \geq 0 \end{cases} \tag{8}$$

$$p\_11\_2 = \begin{cases} -0.2 \times o[1], & o[0] < 0 \\ o[1], & o[0] \geq 0 \end{cases} \tag{9}$$

$$p\_0.4\_1 = \begin{cases} -0.2 \times o[2], & o[0] < 0 \\ o[2], & o[0] \geq 0 \end{cases} \tag{10}$$

$$p\_0.4\_2 = \begin{cases} -0.2 \times o[3], & o[0] < 0 \\ o[3], & o[0] \geq 0 \end{cases} \tag{11}$$

$$p\_tr = \begin{cases} 0, & o[4] < 0 \\ o[4] - 0.9, & o[4] \geq 0 \end{cases} \tag{12}$$

$$reward = -0.5\,(p\_11\_1 + p\_11\_2 + p\_0.4\_1 + p\_0.4\_2) - 0.5\,(p\_tr) \tag{13}$$

In that case, p_11_1 and p_11_2 are the voltage penalties for 11kV level while $p\_0.4\_1$ and $p\_0.4\_2$ are those for 0.4kV level and the process run per step while training or running episode is as described in Table 1.

TABLE I. STEP () FUNCTION

No.	Process
1	Randomly select dates to retrieve domestic demand, EV demand and ambient temperature profiles.
2	Obtain step-ahead forecast data (actual data incorporated with random noise (+/-20%), for domestic demand, ambient temperature, and EV).
3	Apply actions (tr tapping, V2G discharged power, set temp) and calculate Heat Pump output based on the ambient temperature of the selected day using Pout = Q(Tset-Tamb)/COP (where Q = average heat demand per Kelvin of houses in designated area (400 taken in this case), Tset = setting temperature by action, Tamb = outdoor temp, COP = coffiecient of performance of heat pump))
4	Update the network with forecasted data and applied actions.
5	Increment the current step, update the network with actual data, and execute the power flow.
6	Collect observations and compute the reward.
7	If it's the end of the episode, reset the environment and initiate the process as new. Otherwise, proceed to the next step

In this study, the forecasted data are simply adding the random noise varying from 80% to 120% of the actual values. The simulation parameters are as described in Table 2.

The model used 0.0001 to update the policy network (actor) steadily and balances learning speed and stability, reducing policy update overshooting.

TABLE II. SIMULATION PARAMETERS

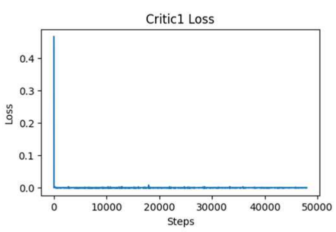
Parameter	Value
Actor Learning Rate	0.0001
Critic Learning Rate	0.0002
Target Entropy	-3.3
Steps per episode	48 (half-hour per step)
Soft Target Update Rate	0.005
Initial Temperature parameter	0.2
Discount Factor	0.85
Horizon of Direct Load control	30% of consumers

Critic network (value function) learns quickly to deliver correct value estimates to guide actor policy modifications and slightly higher rate helps critics react to policy changes faster. To balance exploration and exploitation, the target entropy was set to be more conservative than $-\dim(A)$, aiming for a more deterministic approach while yet allowing for necessary exploration during learning. Target network updates were seamless and stable with a soft target update rate of 0.005. This modest rate prevents dramatic goal value changes, stabilising training. The temperature parameter controls policy stochasticity. Starting at 0.2 encourages fair inquiry letting the agent try multiple actions early in training, which is essential for policy learning. In circumstances where future rewards are uncertain or less important, a discount factor of 0.85 favours immediate rewards over long-term ones. This component helps the agent focus on near-term results, which may be useful in dynamic, variable contexts.

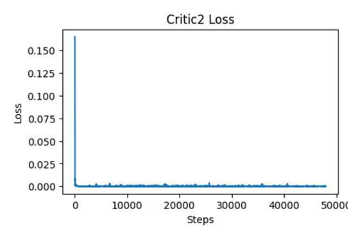
IV. RESULTS

After training the developed model with various agent and environment interactions, the training outcomes are resulted as described below.

Fig. 3 illustrates the study of training outcomes, which highlights specific attributes of the developed Soft Actor-Critic (SAC) model. These findings suggested that the model demonstrates good learning dynamics and convergence. Significantly, the losses incurred by critics, which are essential for assessing the values of state-action, exhibit a remarkable tendency to approach zero at an early stage of the training process. The model's ability to rapidly improve and accurately estimate state-action values is highlighted by this early convergence, indicating extremely efficient learning dynamics. To maintain the resilience and application of the acquired policies and to limit overfitting and hyperparameter sensitivity, it is crucial to do thorough validation and analysis, notwithstanding the hopeful stabilisation of critic losses. Furthermore, it is important to note that the stability of critic losses plays a crucial role in shaping the training dynamics of the actor network. This stability ensures that policy modifications and performance

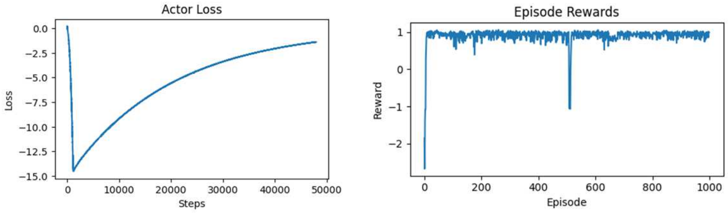


(a)



(b)

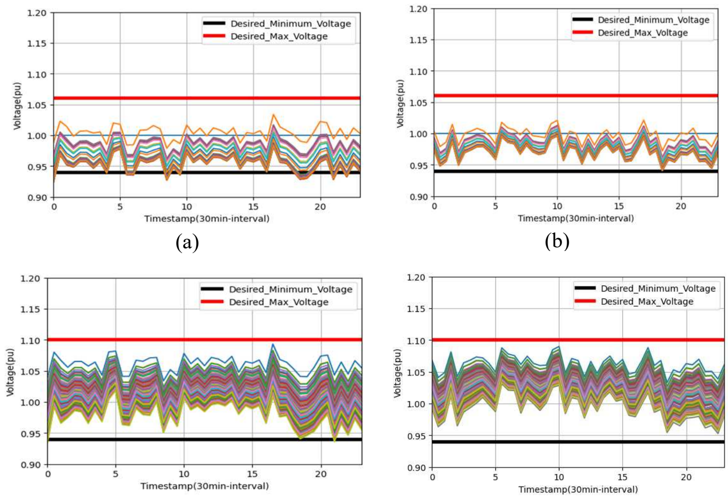
Citation information: DOI: 10.1109/UPEC61344.2024.10892464, 2024 59th International Universities Power Engineering Conference (UPEC)



(c) (d)

Fig. 3. Training Outcome: (a) Critic 1 Loss, (b) Critic 2 Loss, (c) Actor Loss and (d) Reward gained per episode (from episode 1 to 1000)

refinement receive constant feedback. In complex situations, the efficiency of policy exploration and optimisation is enhanced by the interaction between the critic and actor components within the SAC model. Overall, the findings highlight the effectiveness of the SAC approach in attaining rapid convergence and enhancement. In addition to this, the trained agent model is tested at the coldest winter day with full penetration of heat pumps together with penetration of electric vehicles varying from 25%, 50%, 75% to 100%. Originally, the situation at full penetration of heat pumps and electric vehicles without any control of tap changer and direct load controlled demand response incurred the violation of bus voltages especially at evening peak period from (5PM to 11PM).



(a) (b)
(c) (d)

Fig. 4. Voltage Profile at (a) 11kV without control (b) 11kV under control (c) 0.4kV without control and (d) 0.4kV under control with 100% Heat Pump and 100% EV penetration

By means of the controlled signal by the trained agent, the voltage violation is significantly compensated. However, it is worth noting that the voltage of 11kV at around 6PM and 11PM remained within the lower voltage margin of 0.94 per unit (pu) according to Fig. 4. In addition to this, Fig. 5 depicted that the transformer with the highest load density had a much lower loading % during morning peak (6-10AM) and evening peak (6-11PM). The heat pump power drawn due to the lower temperature at early morning and existing EV became lowered dye to regulated set temperature and V2G.

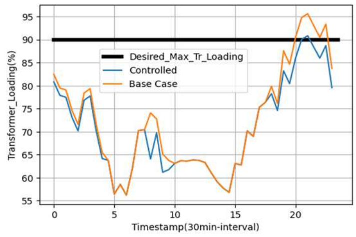
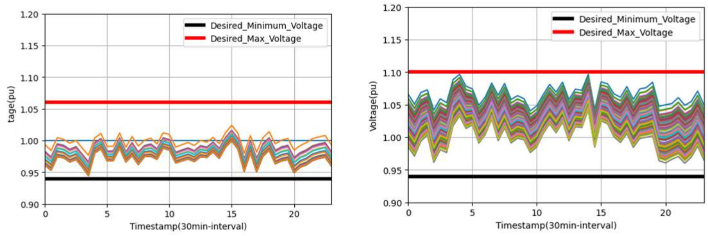


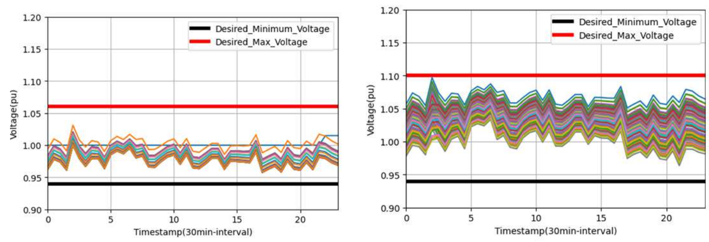
Fig. 5. Loading (%) on Transformer with highest loading at base case and under control at 100% Heat Pump and 100% EV penetration penetration

At EV penetration rates of 75%, 50%, and 25%, in conjunction with 100% Heat Pumps, both 11kV and 0.4kV voltages of all buses are within each of the threshold ranges according to Fig. 6, 7 and 8 and only minor voltage compensation was done by direct load control at evening peak as evident in Fig. 9, 10 and 11. Furthermore, the transformer loadings were first observed to be below the established threshold limit at 75% EV penetration.



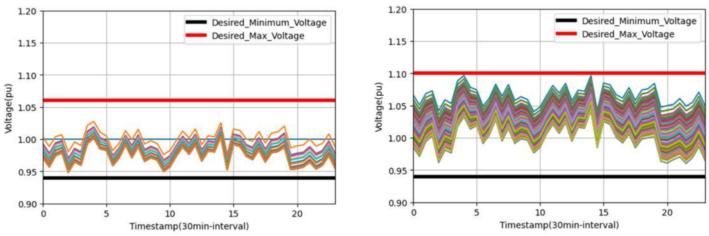
(a) (b)

Fig. 6. Voltage Profile at (a) 11kV buses and (b) 0.4kV buses controlled by trained agent with 100% Heat Pump and 75% EV penetration



(a) (b)

Fig. 7. Voltage Profile at (a) 11kV buses and (b) 0.4kV buses controlled by trained agent with 100% Heat Pump and 50% EV penetration



(a) (b)

Fig. 8. Voltage Profile at (a) 11kV buses and (b) 0.4kV buses controlled by trained agent at 100% Heat Pump and 25% EV penetration

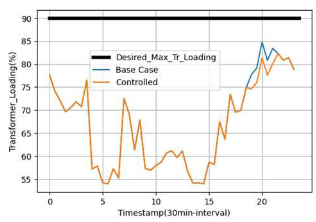


Fig. 9. Loading (%) on Transformer with highest loading at base case and under control at 100% Heat Pump and 75% EV penetration penetration

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/UPEC61344.2024.10892464, 2024 59th International Universities Power Engineering Conference (UPEC)

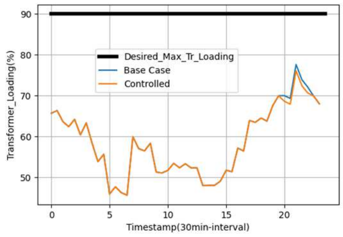


Fig. 10. Loading (%) on Transformer with highest loading at base case and under control at 100% Heat Pump and 50% EV penetration penetration

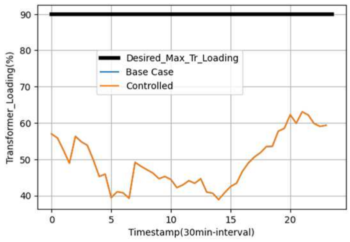


Fig. 11. Loading (%) on Transformer with highest loading at base case and under control at 100% Heat Pump and 25% EV penetration penetration

In scenarios with 50% and 75% electric vehicle (EV) penetration, the direct load control technique worked during evening peaks. This was important to address that raising transformer tap could cause overvoltage at buses closer to the grid in that situation.

At 50% EV penetration, there was no transformer overloading originally but only direct load control was operated to adjust the slight voltage variation. As the number of electric vehicles to be discharged declined, the load reduction due to V2G became inconsequential while load compensation is insignifcant because of minimum actions due to lacking violation at 25% EV penetration. In the base case of 25% EV penetration, the model did not return false decision signals when the environment does not need any corrective actions due to lacking violation in both voltage profile and transformer loading. Thus, the results ensured the reliable grid operation and the system's ability to self-regulate and performance optimization were significant when the minimum intervention was required at lower loading.

V. Discussion and Conclusion

Overall, mitigation measures improve system efficiency and reduce operating limitations. Analysis of transformer loading fluctuations in various electric vehicle penetration scenarios emphasises the need for adaptive management strategies to effectively manage voltage fluctuations, especially during high demand, using direct load control based on heat pump setting temperature and vehicle-to-grid operation. In any case, the model must also be trained and evaluated by tweaking and configuring hyper-parameters to retain dependable results in diverse settings, and the performances shall be compared among different deep reinforcement learning models. Demand flexibility rate fluctuation, demand shift mechanism and energy storage for environment model will be added in the future operation.

Acknowledgment

We extend our sincere gratitude to Scottish & Southern Energy Networks for generously providing the network assets and topology data at https://creativecommons.org/licenses/by/4.0/ .

References

[1] P. Pálenský and D. Dietrich, "Demand side management: demand response, intelligent energy systems, and smart loads," IEEE Transactions on Industrial Informatics, vol. 7, no. 3, pp. 381–388, Aug. 2011.

[2] C. Schellenberg, L. Dimache, and J. Lohan, "Grid-edge technology - Exploring the flexibility potential of a heat pump and thermal energy storage system," E3S Web of Conferences, vol. 111, p. 06002, Jan. 2019.

[3] D. Fischer and H. Madani, "On heat pumps in smart grids: A review," Renewable & Sustainable Energy Reviews, vol. 70, pp. 342–357, Apr. 2017.

[4] J. Crawley, A. Martin-Vilaseca, J. Wingfield, Z. Gill, M. Shipworth, and C. A. Elwell, "Demand response with heat pumps: Practical implementation of three different control options," Building Services Engineering Research and Technology, vol. 44, no. 2, pp. 211–228, Dec. 2022.

[5] C. Schellenberg, J. Lohan, and L. Dimache, "Operational optimisation of a heat pump system with sensible thermal energy storage using genetic algorithm," Thermal Science, vol. 22, no. 5, pp. 2189–2202, Jan. 2018.

[6] Y. Yang and L. Yao, "Optimization Method of Power Equipment Maintenance Plan Decision-Making based on deep reinforcement learning," Mathematical Problems in Engineering, vol. 2021, pp. 1–8, Mar. 2021.

[7] Y. Shi, C. Mu, Y. Hao, S. Ma, N. Xu, and Z. Chong, "Day - ahead optimal dispatching of hybrid power system based on deep reinforcement learning," Cognitive Computation and Systems, vol. 4, no. 4, pp. 351−361, Aug. 2022.

[8] Q. Huang, R. Huang, W. Hao, J. Tan, R. Fan, and Z. Huang, "Adaptive power system emergency control using deep reinforcement learning," IEEE Transactions on Smart Grid, vol. 11, no. 2, pp. 1171–1182, Mar. 2020.

[9] Z. Zhu, Z. Weng, and H. Zheng, "Optimal Operation of a Microgrid with Hydrogen Storage Based on Deep Reinforcement Learning," Electronics, vol. 11, no. 2, p. 196, Jan. 2022.

[10] J. Dong, H. Wang, J. Yang, L. Gao, K. Wang, and X. Zhou, "Low Carbon Economic Dispatch of integrated energy system considering power supply reliability and integrated demand response," Cmes-computer Modeling in Engineering & Sciences, vol. 132, no. 1, pp. 319–340, Jan. 2022.

[11] M. Futakuchi, S. Takayama, and A. Ishigame, "Scheduled Operation of Wind Farm with Battery System Using Deep Reinforcement Learning," IEEJ Transactions on Electrical and Electronic Engineering, vol. 16, no. 5, pp. 687–695, Apr. 2021.

[12] "SmartMeter Energy Consumption Data in London Households – London Datastore." https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

[13] "Low Carbon London Electric Vehicle Load Profiles – London Datastore." https://data.london.gov.uk/dataset/low-carbon-london-electric-vehicle-load-profiles

[14] "Electricity Demand Profiles." https://www.ofgem.gov.uk/sites/default/files/docs/2012/06/electricity-demand-profiles.xlsx

[15] Western Power Distribution, "Project NETWORK EQUILIBRIUM: Voltage Limits Assessment Discussion Paper," Nation Grid, Jan. 2016. Accessed: Dec. 12, 2023. [Online]. Available: https://www.nationalgrid.co.uk/downloads-view-reciteme/2503

[16] "Voltage changes," SPEnergyNetworks. https://www.spenergynetworks.co.uk/pages/voltage_changes.aspx#:~:text=In%20the%20UK%2C%20the%20declared,216.2%20volts%20to%20253.0%20volts

[17] B. Wells et al., "DNO COMMON NETWORK ASSET INDICES METHODOLOGY," Ofgem, Apr. 2021. Accessed: Jan. 05, 2024. [Online]. Available: https://www.ofgem.gov.uk/sites/default/files/docs/2021/04/dno_common_network_asset_indices_methodology_v2.1_final_01-04-2021.pdf

[18] Z. Huang, W. Zhong, D. Li, and H. Liu, "Delay constrained SFC orchestration for Edge Intelligence-Enabled IIOT: a DRL approach," Journal of Network and Systems Management, vol. 31, no. 3, Jun. 2023.