IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 17, 2020, accepted June 30, 2020, date of publication July 15, 2020, date of current version July 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009419

An Online Reinforcement Learning Approach for Dynamic Pricing of Electric Vehicle Charging Stations

VALEH MOGHADDAM[1], (Member, IEEE), AMIRMEHDI YAZDANI[2], (Member, IEEE), HAI WANG[2], (Senior Member, IEEE), DAVID PARLEVLIET[2], AND FARHAD SHAHNIA[2], (Senior Member, IEEE)

[1]School of Information Technology, Deakin University, Geelong, VIC 3220, Australia
[2]College of Science, Health, Engineering and Education, Murdoch University, Perth, WA 6150, Australia

Corresponding author: Amirmehdi Yazdani (amirmehdi.yazdani@murdoch.edu.au)

ABSTRACT The global market share of electric vehicles (EVs) is on the rise, resulting in a rapid increase in their charging demand in both spatial and temporal domains. A remedy to shift the extra charging loads at peak hours to off-peak hours, caused by charging EVs at public charging stations, is an online pricing strategy. This paper presents a novel combinatorial online pricing strategy that has been established upon a reward-based model to prevent network instability and power outages. In the proposed solution, the utility provides incentives to the charging stations for their contributions in the EVs charging load shifting. Then, a constraint optimization problem is developed to minimize the total charging demand of the EVs during peak hours. To control the EVs charging demands in supporting utility's stability and increasing the total revenue of the charging stations, treated as a multi-agent framework, an online reinforcement learning model is developed which is based on the combination of an adaptive heuristic critic and recursive least square algorithm. The effective performance of the proposed model is validated through extensive simulation studies such as qualitative, numerical, and robustness performance assessment tests. The simulation results indicate significant improvement in the robustness and effectiveness of the proposed solution in terms of utility's power saving and charging stations' profit.

INDEX TERMS Electric vehicles, charging stations, pricing strategy, reinforcement learning.

NOMENCLATURE

Abbreviations

ACO	Ant colony optimization.
AHC	Adaptive heuristic critic.
BSS	Battery storage system.
CDP	Coordinated dynamic pricing.
CS	Charging station.
EV	Electric vehicle.
MDP	Markov decision process.
RL	Reinforcement learning.
RLS	Recursive least square.

Parameters

α	EVs load index.
Δt	Time slot.
$\gamma_{(\Delta t-1)}$	Discount factor.
$\mathscr{B}_l$	Lower bound of the coefficient of partial charging.
$\mathscr{B}_u$	Upper bound of the coefficient of partial charging.
σ	Coefficient of partial charging.
$\tilde{E}_{\Delta t}$	Predicted EV load.
φ_x	Profit coefficient of pricing strategy x.
ζ_{m+s}	Service and maintenance cost of CS_j.
C_i	Charging cost of EV_i.
D	Total demand of EVs.
E^j_{EV}	Charging demand of EVs at CS_j.
E_j, Δt	Charging demand of EVs at CS_j.
i	Index of EVs.
j	Index of charging stations.
P^{max}	Maximum power delivery to CSs.
P^{min}	Minimum power delivery to CSs.
$Pr^j_{\Delta t}$	Charging demand of EVs at CS_j.

The associate editor coordinating the review of this manuscript and approving it for publication was Arash Asrari.

$R_{j,\Delta x}$	Reward coefficient of CS_j at Δt.
$Rev_{j,\Delta t}$	Revenue function of CS_j at Δt.
t	time [min].
t_s	Start of a time interval.
t_e	End of a time interval.

I. INTRODUCTION

This With the fast growing of energy demand and greenhouse gas emission concerns, adopting electric vehicles (EVs) could be a great option. With having more EVs on roads, there is a big need to focus more on establishing fast charging infrastructures. Many countries around the world have developed a network of EV charging stations (CSs), often called EV networks [1]. This infrastructure indeed facilitates the commuter and driver daily driving life and has a positive impact on decreasing the driving-based anxieties. However, simultaneous charging of high number of EVs with the uncoordinated charging demands at public CSs may change significantly the demand profile of a utility. As more EVs join the grid to charge their batteries, the more waiting time is added to the actual road traffic, and hence it will create another challenge [2], [3]. For addressing this challenge, there is a need to accommodate smart energy management and control strategies for CSs to control the EV charging demand during peak hours. One of the critical roles of such EV networks is to provide the possibility of charging EVs at off-peak hours which will prevent the negative impacts of extra EVs being charged at demand peak hours. Considering the different research on the impact of EV charging loads on the power grid [4]–[7] various charging control methods and smart scheduling of EVs have been investigated [8], [9]. The main rational behind all these control strategies is to minimize the peak hour's demand based on different scheduling techniques to charge EVs before or after the peak hours. In addition to these control mechanisms, there have been increasing research on designing proper demand response techniques to improve the overall system efficiency [10], [11]. One of the effective solutions in demand response mechanisms, is controlling demand with the price at different times of the day. In other words, by using online pricing methods at CSs, EVs can adjust their charging demands. As an example, [12] has introduced a price strategy for CSs which aims at minimizing the total latency of EV users and electricity cost. However, the charging rate has not been considered in conjunction with any load management of EV loads during peak hours. Research documented in [13], [14] have proposed new infrastructures for CSs with battery storage system (BSS) which considers hourly electricity price and estimates the EVs' demand. However, changing the price at CSs and controlling the EVs' demand at peak hours were not discussed. Power market is an important factor in establishing CSs in a city. Similar to petrol stations, multiple CSs in the same area may belong to different owners. Therefore, competition between different CSs is highly probable and should also be considered. In [15], [16], a competition system has been proposed based on the game theory in order to maximize the CSs' profit. However, the impact of EVs demand on the existing traffic of the grid at peak hours was not discussed.

The lack of price coordination within an EV network can lead to a nonuniform distribution of charging loads in hotspot areas across different CSs [17]. This problem was addressed in [18] by introducing a coordinated dynamic pricing (CDP) method to reduce the overlap between the PEV and residential loads during peak hours. The proposed dynamic pricing model was considered as a constrained optimization problem which estimates the EVs demand in response to the hourly price, distributed by the CSs. The performance of the proposed model in [18] is compared with the existing models in the literature and shows a significance improvement in controlling EVs charging demand at peak hours; however, the profit management of CSs is not incorporated. It should be noted that, majority of the existing dynamic pricing approaches are offline and assume perfect knowledge of charging demands during the specific planning time. Thus, for a reliable and practical approach, dynamic pricing algorithms must be robust against the uncertainties in future charging patterns and the users' preferences [19]. In [20], [21], reinforcement learning approaches have been proposed to find an optimal decision pricing for the energy trading and to predict the pricing of CSs. Likewise, [22] has implemented an optimized demand response framework for EV aggregators with an assumption that the future hourly electricity prices are known in a non-causal manner. Markov decision process (MDP)-based algorithms have been proposed in [21], [23] for the stochastic distributions of future events impacted by changing prices were proposed. However, such an approach is not cost-effective. Learning-based approaches that evolve by learning from data observed in the previous steps of the evaluation, are potential candidates to deal with this issue. For example, [24] has adopted a reinforcement learning (RL) algorithm for EV charging control without any a priori knowledge about the next EVs' arrivals. Likewise, the RL approach has also been adopted in [25] for a heuristic day-ahead planning of EV fleet charging. In [26], a heuristic solution based on a RL approach has proposed for the real-time fuel saving optimization of EVs. They introduced a model free algorithm which presented a better performance with its equivalent fuel economy and computational speed. Authors of [27] developed an intelligent optimization approach based on Multi-Modal approximate dynamic programming for charging/ discharging of EVs at a grid-connected charging station. They considered continuous state/ action spaces to represent a continuous charge/ discharge process of EVs. However, these studies of RL implementation, haven't considered a control strategy for EVs demands in a network of charging stations.

The RL algorithm is employed in [24], [25] to maximize the CSs' profit and control the demand; however, a reward-base model was not considered for pricing strategy of CSs in order to contribute in the temporal load shifting of the EV.

In this study, a novel combinatorial online pricing strategy has been proposed which is based on the RL and Fast adaptive

heuristic critic (AHC) techniques to control the EVs charging demands for supporting the utility's stability and increasing the CSs' total revenue. By employing a reward-based pricing control, EVs charging load can be controlled during peak period. Moreover, the RL-Fast AHC technique enables the system to implement a unified adaptive exponential tracking which can control and filter the updated rewards for different number of EVs charging in different periods. The main contribution of this paper to the research field can be summarized as:

- Proposing an online multi-agent framework to control the EVs' charging demands for supporting the utility's stability and increasing the CS's total revenue, using a reward-based model;
- Developing an RL framework, implemented by combining an adaptive filtering method (RLS) and the AHC-Fast heuristic algorithm to combine the benefits of the techniques; and
- Formulating an objective function which aims at reducing the EVs' charging demand at peak hours, as a constrained optimization problem..

The remainder of the paper is organized as follows: Section 2 introduces the proposed on-line RL model and the formulated problem. The mathematical model of the proposed AHC-RLS algorithm is provided in Section 3 followed by introducing the hourly-updated reward coefficients. The performance of the developed technique is evaluated by numerical simulation studies in Section 4 and this performance is compared against other similar existing approaches. Finally, the main findings of the research are summarized and highlighted in the last Section.

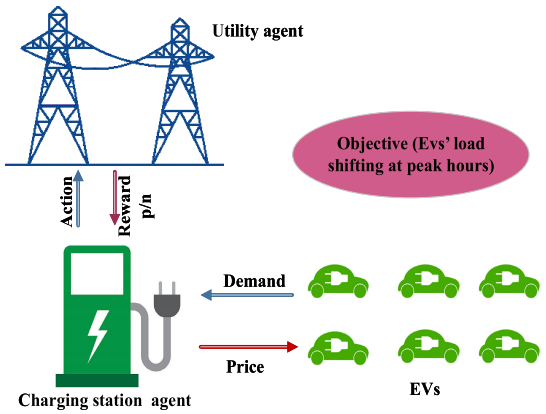


FIGURE 1. Pictorial representation of the multi-agent EV network model.

II. PROBLEM FORMULATION

A. EV NETWORK SYSTEM MODEL

Figure 1 depicts a pictorial representation of the considered EV network system model in the context of a multi-agent framework. In this model, a utility agent is responsible to distribute electricity to the CSs and to control their charging demands by following a load-shifting policy, contributed in this paper. Another agent, related to CSs, is responsible to propose hourly updated price vectors to EVs.

B. OPTIMIZATION MODEL FOR EVs LOAD SHIFTING

The main goal of this study is to implement a mechanism for minimizing the EVs' charging loads at peak hours and load-shifting to off-peak periods. This problem is formulated as a constrained optimization problem in the form of:

$$min \sum_{\forall j=1,\ldots,n,\forall \Delta t\in t_s,\ldots,t_e} \sigma_i^j E_{EV}^j \Delta t \quad (1)$$

$$subject\ to: \mathscr{B}_l \le \sigma_i^j \le \mathscr{B}_u \quad (2)$$

$$\sum_{\forall \Delta t\in t_s,\ldots,t_e} Pr_{\Delta} t^j \quad (3)$$

$$t_s \le \Delta t \le t_e \quad (4)$$

$$P^{min} \le P_{\Delta t} \le P^{max} \quad (5)$$

Equation 1 denotes the main objective (i.e., minimizing the charging demand of EVs at the CS (EEV) during the period of Δt while Equation 2 shows the lower and upper boundaries of the coefficient of partial charging, that represents the recharging quantity of EV's battery at CSs. Equation 3 defines the price vector of each period at CSs which is the main decision variable for the objective function (obtained by the proposed online RL model, introduced in Section 3). The starting and ending of each interval is shown in 4, while 5 specifies the boundaries of power delivery at each period for the CSs.

III. ONLINE RL METHOD

The RL is categorized as a machine learning algorithm in computer science and engineering [28]. The RL focuses on theories and algorithms of learning to solve the optimal control problem of MDPs, adopted as sequential decision processes of real-world applications [29]. Learning prediction and learning control are two required processes in RL. Learning control estimates the optimal value of a model free sequential process; and learning prediction can be considered as a sub-problem of learning control which aims to solve the policy evaluation problem of a sequential process [20], [30]. In an online RL method, an agent interacts with its environment in discrete time steps, and can update itself incrementally with each newly time step. The agent can choose an action from the set of available actions, which is subsequently sent to the environment [31]. The environment moves to a new state based on a reward that each agent receives at each iteration of evaluation, and then the optimal control strategy is evaluated by the amount of the received rewards. Thus, the RL enables the design of adaptive controllers that learn online and propose a solution to users for optimal control problems in real-time [32].

In this paper, as Figure 2 presents, the utility and CS agents perform the control strategy on EVs demand at the demand peak hours. The utility agent as a critic of the system, receives the hourly feedback from the environment which is defined as EVs' demand vector at each interval in the model. The utility agent is also responsible to determine the hourly reward

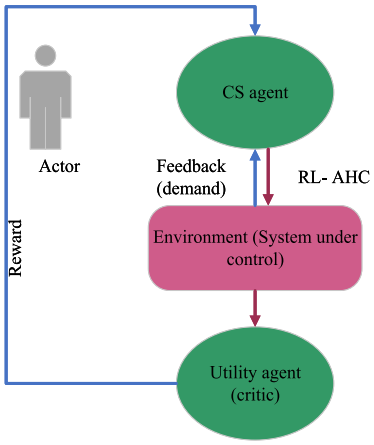


FIGURE 2. Multi-agent control system of the EV network operating under the RL-AHC algorithm.

coefficients for CSs. In addition to the RLS, the AHC-Fast critic is used for an actor-critic interaction, in order to improve the learning-prediction efficiency in the critic [29]. Using AHC filtering algorithm in the proposed model, the states of the EV loads can be evaluated at CSs in real-time. This further enables the calculation of an hourly reward for each CS. Followed by hourly rewards coefficients from the utility agent, the CS agent can update the price coefficients for all CSs using the given reward and by the help of the RLS algorithm, as seen from Figure 2.

A. AGENT PROPERTIES

The utility agent is responsible to provide power for CSs with a specified price and monitoring the charging load of the whole power system; it also calculates power flow or optimal power flow and power loss of the power system. In the model of Figure 1, the utility agent plays the controller role, as it evaluates the expected charging loads of EVs at CSs at each interval, and sends a reward/ punishment coefficient to the CS agents to update their price vector. When the CS agent receives the reward/punishment coefficient, it employs the CDP and the online reinforcement models to calculate the price vector for each Δt and then distributes it between all CSs. The interaction between different agents and EVs charging request is shown in Figure 3. As seen from this figure, the CS agent estimates the charging requests from the EVs in different intervals, and asks utility agent for the required power. The utility agent sells the electricity based on an evaluation, which would be a benchmark for controlling the demand during the peak hours. Then, the CS agent receives a reward from the utility agent for each interval. It also updates the price vectors for each Δt using the AHC-RLS algorithm.

B. REWARD COEFFICIENT UPDATE

Considering the EVs' demand curve at each Δt, the reward function broadcasts a proper coefficient for the proposed price vector of CSs. Therefore, the reward function of finding

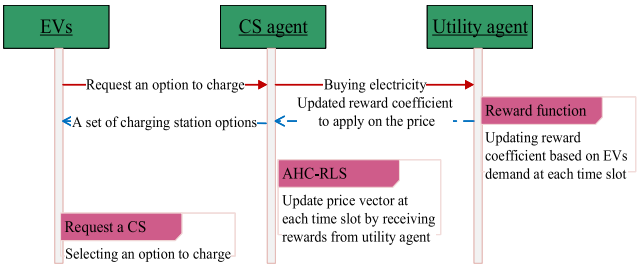


FIGURE 3. Sequence diagram of the proposed model.

the price coefficients is determined by evaluating three input variables for each Δt, the output of the RLS algorithm in the previous interval in terms of price coefficient of $\Delta t - 1$, the demand index of EVs, and the base price of the provider for each Δt of a day. The proposed critic network examines an eligibility function $E_{j,\Delta t}$ at each interval, which corresponds to the prediction of cumulative future rewards for the CS agent. In addition, it provides an internal reinforcement signal $\hat{R}(t)$ to the action network. The internal reinforcement signal is obtained based on the difference between the predicted and actual EVs loads. The eligibility function of the utility agent at each state is defined as

$$E_{j,\Delta t} = \alpha E_{j,(\Delta t-1)} - \tilde{E_{\Delta}} \tag{6}$$

where α is the EVs load coefficient at the interval while $\tilde{E}_{\Delta}$ is the predicted EVs' demand for the current interval. Considering the result of the function in the utility agent, the proposed reward coefficients for the CS agent is formulated as

$$R_{j,\Delta t} = \sum_{\forall \Delta t = t_s,\ldots,te} [\gamma_{(\Delta t-1)} E_{j,\Delta t} \tilde{E_{\Delta}}] \tag{7}$$

where $\gamma_{(\Delta t-1)}$ is the discount factor of the critic system and $0 < \gamma < 1$.

C. PROFIT FUNCTION OF CSs IN RESPONSE TO THE REWARD COEFFICIENT

The proposed online RL-based pricing policy increases the capability of the CSs to accommodate larger arrival rates at off-peak hours. In this section, the profit function of CSs based on the following parameters is presented. Each CS obtains the revenue from each charged EV with a specific price at a specific interval. Also, the maintenance and service costs for each charging process of the EVs needs to be added. As a result, the CSs profit function can be introduced as

$$Rev_{j,\Delta t} = \sum_{\forall \Delta t t_s,\ldots,te}^{i} [\sigma C_i - \zeta_{m+s}] \tag{8}$$

where at each Δt, σ is the coefficient of partial charging, C_i indicates the charging cost of each EVi and ζ_{m+s} is the maintenance and service cost of each CS.

Algorithm 1 summarizes the proposed online RL learning based on AHC. In this algorithm, the function of the reward

Algorithm 1 Proposed Online RL Model Using AHC

Require Updated price vector Pr^{j}_{Δ}
Require Learning rate $E_{(j,\Delta t)}$
Require Reward coefficient vector $R_{(j,\Delta t)}$
Require Number of EV at CS E^{j}_{EV}
Ensure: Minimum charging demand at peak time (Min E^{j}_{EV})

for E^{j}_{EV} **do**
  $R_{j,\Delta t} = \sum_{\forall \Delta t=t_s,\ldots,te}[\gamma_{(\Delta t-1)}E_{j,\Delta t}\tilde{E_{\Delta}}]$
  $Rev_{j,\Delta t} = \sum^{i}_{\forall \Delta tt_s,\ldots,te}[\sigma C_i - \zeta_{m+s}]$
  Sending rewards to each CS: $R_{(j,\Delta t)}$
  Call Equation 1
  if E^{j}_{EV} doesn't meet the criteria **then**
    while $E_{j,\Delta t} == 1$ **do**
      Call Pr^{j}_{Δ}
    end while
  end if
end for

coefficient update is iteratively called up to provide up-to-date reward information for each CS, in order to calculate the updated price vector at each Δt.

TABLE 1. Simulation parameters.

Description	Parameter
Number of CSs (N)	15∼20
Number of EVs (K)	500
Arrival rate at CS_j (λ_j)	3 to 10 EVs/hour
Rate of charge for ac charging	22 kW [34]
Rate of charge for dc charging	50 kW [33]

D. PERFORMANCE EVALUATION AND DISCUSSION

To investigate the effectiveness and robustness of the proposed Online RL model-based Fast AHC algorithm, the Washington City EV network [28], as shown in Figure 4, is considered as the simulation benchmark model. As provided in Table 1, a maximum of 20 CSs and 500 EVs are used in this study. The arrival rate of the EVs at each CS is modelled as a Poisson distribution with a rate of 3∼10 EVs/hour. Two different rates of charge (dc and ac) are used as representative charging options. The hourly charging prices at the CSs is generated by the discussed model of Section 3. The purchased electricity price from the utility without any changes by the proposed model is assumed as 22 to 46 cents/kWh [33], [34]. All computations are performed on a desktop PC with an Intel i7 3.20 GHz quad-core processor in MATLAB 2019a. For the qualitative assessment of the online RL model-based Fast AHC, the average electricity price is analyzed during a day for CSs and the amount of electricity saving on the utility side is discussed. Furthermore, the profit maximization of CSs as well as the probability of overlap between PEV and residential loads are elaborated. In the sequel, the convergence assessment of the AHC-RLS algorithm is also presented and analyzed.

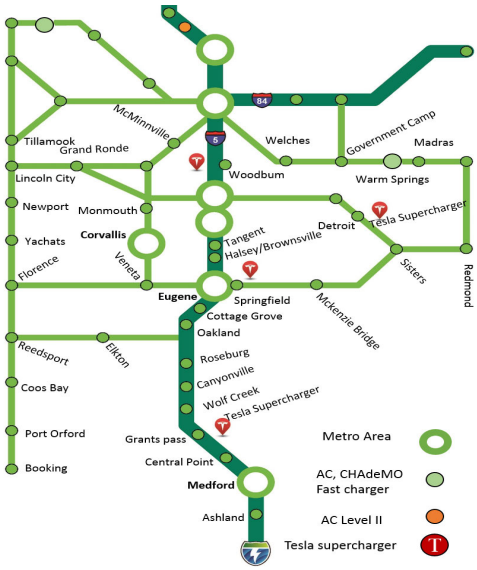


FIGURE 4. EV network, Washington Green Highway [28].

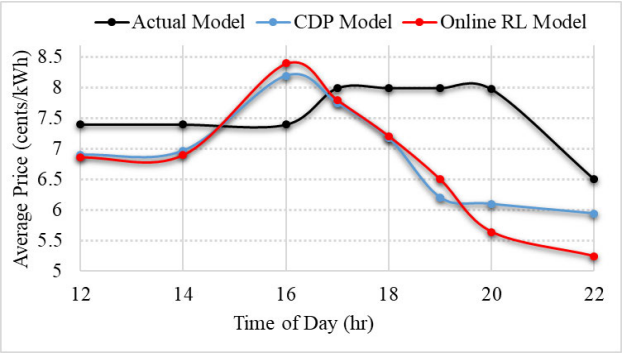


FIGURE 5. Average EV charging price during 12:00 to 22:00.

Figure 5 shows the average price in the period of 12:00 to 22:00 based on the developed on-line RL model. As the figure suggests, by using the new prices, the amount of electricity price grows up between 15:00 and 17:00. This trend of pricing proposes a promising controllability of the EVs' charging process during peak hours, as the costumers prefer to charge their EVs before the price increasing and after peak hours. As the number of EVs in the CSs decreases at peak hours, the amount of extra stress on the power grid decreases as well. The figure also indicates the superiority of the proposed online RL model in the price management against the well-known CDP model. Figure 6 illustrates the effect of the developed on-line RL model on the electricity saving of utility during peak hours. In this figure, the percentage of electricity saving increases if the CSs follow the strategy proposed by the developed model. The price of electricity at the CSs is one of the important decision factors for the EVs drivers, to recharge their batteries at public CSs. Hence, using the developed model, the EV demand appeared increases in the CSs before the demand peak hours (as the electricity price is low), and consequently, the extra loads at peak hours decreases. As a result, the utility will benefit

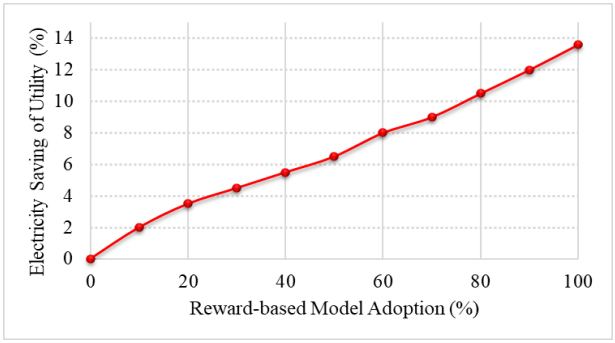


FIGURE 6. Electricity saving of utility by adopting the developed on-line RL model.

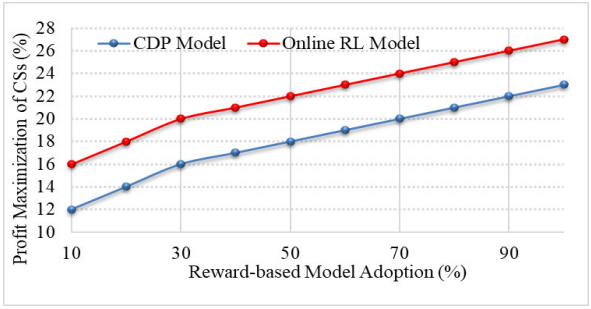


FIGURE 7. Profit maximization of CSs using the online RL model.

from this strategy in both sides of minimizing the cost of energy and maximizing the stability of the grid. As illustrated in Figure 7, as more CSs adopt the developed on-line RL model for their pricing strategy, their profit increases. This is due to the fact that the developed model updates their pricing vector based on the received rewards from the utility at off-peak hours, with cheaper electricity prices for charging. This consequently encourages more EVs to charge at CSs during that period. Fig. 8 shows that the online RL model provides an admissible reliability for the network by minimizing the overlap between EV and residential loads during different hours of a day. This robust performance is even more superior than the CDP counterpart. In essence, as more CSs employ the developed on-line RL model to update their price vector, the reduced demand peaks are more (because of moving from uncontrolled charging to a more coordinated charging system).

Figure 9 shows fluctuations of EV loads index during 12:00 and 22:00. As inferred from the figure, using the developed online RL model, the EV demand decreases during peak hours and shifts to the off-peak periods. As the proposed price of electricity by the CSs between 14:00 and 16:00 is low, the EVs prefers to charge their batteries before peak hours. Alternatively, they can postpone the charging process to the evening, after the peak hours, to minimize their cost.

Figure 10 illustrates the average of reward coefficients for 1,000 training episodes. As can be seen from the figure, with the increase in the number of EVs, the reward coefficients change significantly and provide a better performance in terms of controlling the EVs' demand.

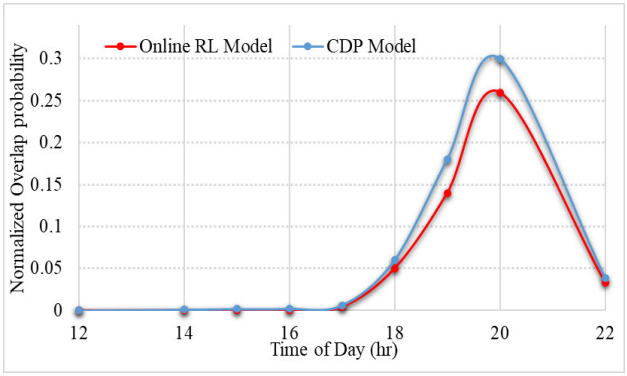


FIGURE 8. Probability of overlap between EV and residential loads at peak hours.

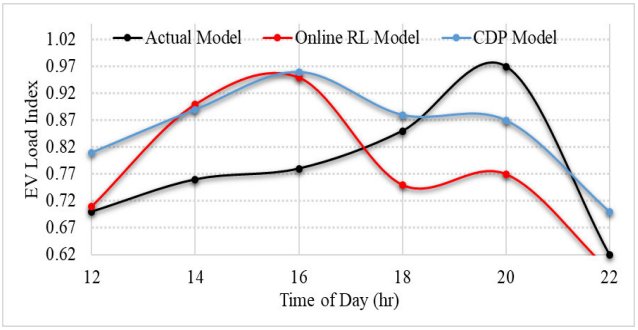


FIGURE 9. EV load index for a given time interval.

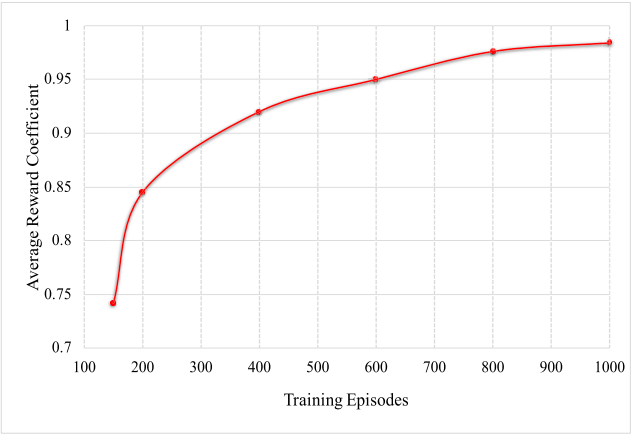


FIGURE 10. Average reward coefficient for 1000 training episodes.

Figure 11 provides an illustrative representation of convergence rate of the Fast-AHC critic used in this study. As discussed in Section 3, the Fast-AHC technique was developed by utilizing the RLS algorithm to estimate the updated price vectors at each interval using the hourly rewards coefficients. Compared to the RLS method itself, the convergence rate of the proposed Fast-AHC critic is quicker, which is an indication of the effectiveness for the practical system management.

E. EFFICIENCY IMPROVEMENT

One of the main challenges for the conventional RL methods is their slow convergence rate, particularly in the cases that the learning data set is not rich enough or hard to be

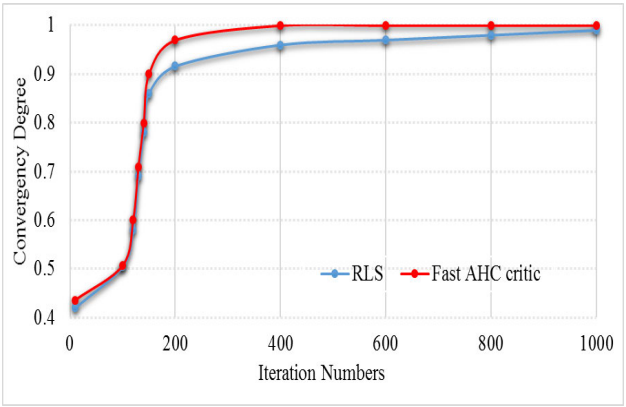


FIGURE 11. Convergence rate of Fast-AHC in load shifting process.

generated [22]. This drawback, directly impacting the learning prediction, can be addressed by developing an adaptive critic signal. For the proposed online RL model, at each iteration of updating the price vector in the CS agent and obtaining the reward coefficient from the utility agent, the policy of the actor will change adaptively with respect to the number of EVs in the system. Table 2 summarizes the improvement of using the online RL model over the existing CDP model which could significantly reduce the overlap between the EV and residential loads during peak hours. As Table 2 indicates, after comparing the results between the developed on-line RL model and the CDP a better performance is achieved in terms of decreasing overlap probability and EVs load index during peak hours (indicated by Improvement (%) in Table 2).

TABLE 2. Summarized simulation results.

Evaluation Parameters	CDP Model [17]	Developed on-line RL Model	Improvement (%)
Overlap probability	0.030	0.021	42.85%
EVs load index	0.91	0.78	16.6

F. ROBUSTNESS ASSESSMENT

To examine the robustness of the proposed online RL model, its performance is evaluated against a different benchmark problem introduced in [17]. The benchmark problem is a mixed-integer nonlinear programming problem and aims to minimize the electricity demand of CSs, as given by (6)-(9) [17]. Figure 12 compares the computational time of the developed online RL and CDP model in reaching an optimal solution under the benchmark problem of [17]. As seen from this figure, at the initial training stage, the developed model shows a slightly weaker performance compared to the CDP, as it is undergoing trials and errors. However, after experiencing more iterations, the developed model adapts to the learning environment. This adaptation not only considers the current reward but also updates the future rewards. This results in improving the performance by learning from the

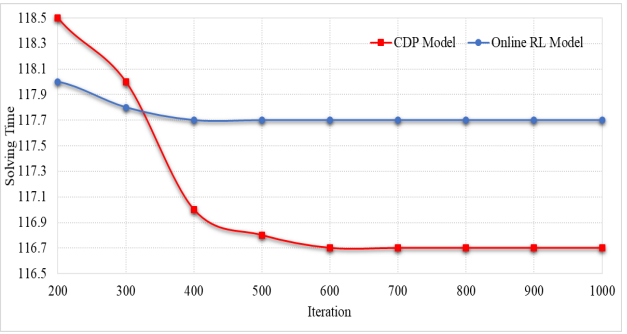


FIGURE 12. Robustness assessment based on the computational time and model complexity.

environment. In contrast, the CDP model has a low learning capability and an increase in the complexity of the problem requires more iterations before converging to an optimal solution, and thus, presents a higher computational time as well. It should be noted that the developed model has advantages in computation and is more suitable for online learning compared to the existing reinforcement learning algorithms. The effectiveness of the developed model is analyzed and verified by learning prediction of Markov chains with a wide range of parameter settings [22]. Although it requires more computation at each iteration for the online updates of the proposed system, the developed model is more efficient than the AHC itself. Assuming, K is the number of states for updating the reward coefficient in the proposed multi-agent control system of Figure 2, the computation time decreases to $O(K^2)$ using the developed model, that is another indication of the robustness and effectiveness of the proposed algorithm in practice.

TABLE 3. Numerical comparison of profit performance of different pricing strategies.

Days	Static pricing ϕ_s [8]	CDP ϕ_{CDP} [17]	Online RL model ϕ_{RL}
1	0.2	0.34	0.46
2	0.24	0.41	0.47
3	0.25	0.46	0.39
4	0.2	0.4	0.51
5	0.26	0.44	0.59
6	0.31	0.65	0.76
7	0.33	0.63	0.79

G. PROFIT PERFORMANCE IN DIFFERENT PRICING STRATEGIES

Table 3 compares the profit coefficients of three different pricing strategies, namely static pricing [8], CDP [17], and the developed online RL model. The static pricing strategy follows a fixed-pattern for the adaptation of EVs charging at CSs. Fixed-pricing or static pricing strategy adopts a competition strategy between different CSs with different cost parameters. Therefore, EV demands don't change with the price information as it follows a fix pattern at different time slots of a day. On the other hand, in the CDP model, the EVs demand estimated in response to charging prices at

various CSs uses a rule-based heuristic solution to address the dynamic pricing challenge in real-time. By doing so, the CDP model motivates more EVs before peak hours to charge at CSs, and it consequently increases the revenue of CSs as well. Finally, a significant increase is observed in the profit coefficient after a time elapse because the profit function of the proposed model updates their pricing vector based on the received rewards from the utility at off-peak hours with cheaper electricity prices. This consequently results in encouraging more EVs to charge at CSs during that time period.

IV. CONCLUSION

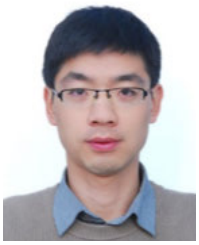
This paper introduced an EV load-shifting mechanism based on an online RL model in a multi-agent system framework. The new online RL model is developed based on the AHC-Fast critic and RLS as adaptive filtering algorithms. The proposed model, first evaluates and monitors the EVs' charging demand at the CSs from the utility agent; then the CS agent received rewards, calculates the updated price vector of charging stations at each interval. This enables the network to propose cheaper prices before peak hours to encourage more EVs to recharge their battery at those intervals, instead of demand peak hours. The effectiveness and robustness of the proposed online RL model were verified through extensive simulation studies. The results of simulations indicate the significant contribution of the proposed model in in decreasing extra demand of EVs charging at peak hours, and hence, increasing the profit of the CSs.

REFERENCES

[1] T. Franke and J. F. Krems, "What drives range preferences in electric vehicle users?" *Transp. Policy*, vol. 30, pp. 56–62, Nov. 2013.
[2] K. Bhavnagri. (May 2019). *How do I Use Electricity Throughout the day, the Load Curve*. [Online]. Available: https://www.solarchoice.net.au/blog/how-do-i-use-electricity-throughout-the-day-the-load-curve
[3] N. I. Guide. (Jul. 2018). *Daily Trafffic Counts*. [Online]. Available: https://www.service.nsw.gov.au/service-centre/roads-maritime-services
[4] A. Baniasadi, D. Habibi, O. Bass, and M. A. S. Masoum, "Optimal real-time residential thermal energy management for peak-load shifting with experimental verification," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5587–5599, Sep. 2019.
[5] A. S. Masoum, S. Deilami, M. A. S. Masoum, A. Abu-Siada, and S. Islam, "Online coordination of plug-in electric vehicle charging in smart grid with distributed wind power generation systems," in *Proc. IEEE PES Gen. Meeting Conf. Expo.*, Jul. 2014, pp. 1–5.
[6] Z. Moghaddam, I. Ahmad, D. Habibi, and Q. V. Phung, "Smart charging strategy for electric vehicle charging stations," *IEEE Trans. Transport. Electrific.*, vol. 4, no. 1, pp. 76–88, Feb. 2017.
[7] Y. Hu, L. Yang, B. Yan, T. Yan, and P. Ma, "An online rolling optimal control strategy for commuter hybrid electric vehicles based on driving condition learning and prediction," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4312–4327, Jun. 2016.
[8] M. H. Amini and A. Islam, "Allocation of electric vehicles' parking lots in distribution network," in *Proc. ISGT*, Feb. 2014, pp. 1–5.
[9] H. Xu, H. K. Nguyen, X. Zhou, and Z. Han, "Charging control of electric vehicles in smart grid: A stackelberg differential game based approach," *Mobile Netw. Appl.*, vol. 2018, pp. 1–9, Sep. 2018.
[10] A. Ghavami and K. Kar, "Nonlinear pricing for social optimality of pev charging under uncertain user preferences," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Oct. 2014, pp. 1–6.
[11] P. Wong and M. Alizadeh, "Congestion control and pricing in a network of electric vehicle public charging stations," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2017, pp. 762–769.
[12] S. Bai, D. Yu, and S. Lukic, "Optimum design of an EV/PHEV charging station with DC bus and storage system," in *Proc. IEEE Energy Convers. Congr. Expo.*, Sep. 2010, pp. 1178–1184.
[13] S. Negarestani, M. Fotuhi-Firuzabad, M. Rastegar, and A. Rajabi-Ghahnavieh, "Optimal sizing of storage system in a fast charging station for plug-in hybrid electric vehicles," *IEEE Trans. Transport. Electrific.*, vol. 2, no. 4, pp. 443–453, Dec. 2016.
[14] J. Tan and L. Wang, "Real-time charging navigation of electric vehicles to fast charging stations: A hierarchical game approach," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 846–856, Mar. 2017.
[15] W. Lee, L. Xiang, R. Schober, and V. W. S. Wong, "Electric vehicle charging stations with renewable power generators: A game theoretical analysis," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 608–617, Mar. 2015.
[16] S. Rasoul Etesami, W. Saad, N. Mandayam, and H. V. Poor, "Smart routing of electric vehicles for load balancing in smart grids," 2017, *arXiv:1705.03805*. [Online]. Available: http://arxiv.org/abs/1705.03805
[17] Z. Moghaddam, I. Ahmad, D. Habibi, and M. A. S. Masoum, "A coordinated dynamic pricing model for electric vehicle charging stations," *IEEE Trans. Transport. Electrific.*, vol. 5, no. 1, pp. 226–238, Mar. 2019.
[18] S. Limmer, "Dynamic pricing for electric vehicle charging—A literature review," *Energies*, vol. 12, no. 18, p. 3574, 2019.
[19] S. Zhou, Z. Hu, W. Gu, M. Jiang, and X.-P. Zhang, "Artificial intelligence based smart energy community management: A reinforcement learning approach," *CSEE J. Power Energy Syst.*, vol. 2, pp. 1–10, Oct. 2019.
[20] A. Chiä, J. Lundén, and V. Koivunen, "Reinforcement learning-based plug-in electric vehicle charging with forecasted price," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 3674–3684, May 2017.
[21] X. Xu, H. He, and D. Hu, "Efficient reinforcement learning using recursive least-squares methods," *J. Artif. Intell. Res.*, vol. 16, pp. 259–292, Apr. 2002.
[22] Q. Chen, F. Wang, B.-M. Hodge, J. Zhang, Z. Li, M. Shafie-Khah, and J. P. S. Catalao, "Dynamic price vector formation model-based automatic demand response strategy for PV-assisted EV charging stations," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2903–2915, Nov. 2017.
[23] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Syst.*, vol. 32, no. 6, pp. 76–105, Dec. 2012.
[24] R. Lu, S. H. Hong, and X. Zhang, "A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Appl. Energy*, vol. 220, pp. 220–230, Jun. 2018.
[25] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[26] T. Liu, X. Hu, W. Hu, and Y. Zou, "A heuristic planning reinforcement learning-based energy management for power-split plug-in hybrid electric vehicles," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6436–6445, Dec. 2019.
[27] C. D. Korkas, S. Baldi, S. Yuan, and E. B. Kosmatopoulos, "An adaptive learning-based approach for nearly optimal dynamic charging of electric vehicle fleets," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2066–2075, Jul. 2018.
[28] (2018). *Alternative Fuels Data Center, Laws Incentives, Electric Vehicle Infrastructure Definitions*. [Online]. Available: https://afdc.energy.gov/laws/6534
[29] E. Kuznetsova, Y.-F. Li, C. Ruiz, E. Zio, G. Ault, and K. Bell, "Reinforcement learning for microgrid energy management," *Energy*, vol. 59, pp. 133–146, Sep. 2013.
[30] S. Vandael, B. Claessens, D. Ernst, T. Holvoet, and G. Deconinck, "Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1795–1805, Jul. 2015.
[31] P. Kofinas, G. Vouros, and A. I. Dounis, "Energy management in solar microgrid via reinforcement learning using fuzzy reward," *Adv. Building Energy Res.*, vol. 12, no. 1, pp. 97–115, Jan. 2018.
[32] K. Doya, H. Kimura, and M. Kawato, "Neural mechanisms of learning and control," *IEEE Control Syst.*, vol. 21, no. 4, pp. 42–54, Aug. 2001.
[33] (Jul. 2017). *CHAdeMO User Manual*, [Online]. Available: http://store.evtv.me/proddetail.php?prod=20kwchademo&cat=23
[34] P. Morrissey, P. Weldon, and M. O'Mahony, "Future standard and fast charging infrastructure planning: An analysis of electric vehicle charging behaviour," *Energy Policy*, vol. 89, pp. 257–270, Feb. 2016.

VALEH MOGHADDAM (Member, IEEE) received the master's degree in embedded system design from the University of Lugano, Switzerland, in 2012, and the Ph.D. degree in computer systems from Edith Cowan University, Joondalup, WA, Australia, in 2019. She is currently a Lecturer with the School of Information Technology, Deakin University, Geelong, VIC, Australia. Her current research interests include management of smart grid and renewable energy systems with a focus on electric vehicles applications.

AMIRMEHDI YAZDANI (Member, IEEE) received the master's degree in mechatronics and automatic control from the Universiti Teknologi Malaysia, Malaysia, in 2012, and the Ph.D. degree in electrical-control engineering from Flinders University, Adelaide, SA, Australia, in 2017. From 2017 to 2018, he was employed as a Postdoctoral Research Associate with Flinders University. He is currently working as a Lecturer of electrical engineering with the College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia. He is also an Academic Chair of engineering technology, electrical power engineering, and renewable energy engineering with Murdoch University. His research interests include guidance and control of robotic, autonomous, and mechatronic systems, optimal control and state estimation theory, and intelligent control applications. He is also the Vice-Chair of the IEEE Industrial Electronic Society, WA Chapter.



HAI WANG (Senior Member, IEEE) received the B.E. degree in electrical and electronic engineering from Hebei Polytechnic University, China, in 2007, the M.E. degree in electrical and electronic engineering from Guizhou University, China, in 2010, and the Ph.D. degree in electrical and electronic engineering from the Swinburne University of Technology (SUT), Australia, in 2013. From 2014 to 2015, he was a Postdoctoral Research Fellow with the Faculty of Science, Engineering, and Technology, SUT. From 2015 to 2019, he was with the School of Electrical and Automation Engineering, Hefei University of Technology, China, where he served as a Professor (Huangshan Young Scholar) and the Deputy Discipline Head of Automation. He is currently a Senior Lecturer of electrical engineering and an Academic Chair of instrumentation and control engineering and industrial computer systems engineering with the College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia. His research interests include sliding mode control, adaptive control, robotics and mechatronics, neural networks, nonlinear systems, and vehicle dynamics and control.

DAVID PARLEVLIET received the Bachelor of Science degree in applied computational physics and computer science from Murdoch University, Perth, WA, Australia, in 2004, and the Ph.D. degree in physics from Flinders University, in 2008, focusing on silicon nanowires for photovoltaic applications. He is currently a Senior Lecturer and a Chief Remote Pilot in discipline of engineering and energy with the College of Science, Health, Education and Engineering, Murdoch University. He has a strong background in engineering, physics, and computer science. He is also an Expert in RPAS flight operations. His research interests include performance analysis and applications of renewable energy systems, autonomous technologies, performance and reliability of PV systems, and drone-based field monitoring of PV systems.

FARHAD SHAHNIA (Senior Member, IEEE) received the B.Sc. (Hons.) and M.Sc. (Hons.) degrees in electrical power engineering from the University of Tabriz, Tabriz, Iran, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from the Queensland University of Technology, Brisbane, QLD, Australia, in 2011. He was a Research Fellow with the Queensland University of Technology. From 2012 to 2015, he was a Lecturer with the Department of Electrical and Computer Engineering, Curtin University. He is currently an Associate Professor with Murdoch University, Perth, WA, Australia. His professional experience includes three years with Research Office-Eastern Azerbaijan Electric Power Distribution Company, Tabriz. His research interests include distribution networks, power quality, and application of power electronic in power systems.