A Two-Tailed Pricing Scheme for Optimal EV Charging Scheduling Using Multiobjective Reinforcement Learning

Kayode E. Adetunji, Ivan W. Hofsajer, *Member, IEEE*, Adnan M. Abu-Mahfouz, *Senior Member, IEEE*, and Ling Cheng, *Senior Member, IEEE*

***Abstract*—Electric vehicles (EVs) are crucial to the reduction of carbon emissions. However, their charging poses a threat to power system networks. Hence, EV charging control strategies are developed to curb this challenge, using charging prices to incentivize EV drivers to choose EV charging stations (EVCS) favourable to the grid's stability. The challenge of this strategy is the likelihood of EV drivers accepting EVCS suggestions. To increase the probability of accepting EVCS suggestions, we introduce a two-tailed incentive pricing (TTIP) scheme in an EV charging coordination model, where incentives are offered as charging prices and parking time. We formalized the EV charging problem as a multiobjective Markov decision process and proposed a deep deterministic policy gradient (DDPG) to solve it. To tackle the challenge of continuous action space that leads to the dimensionality curse, the proposed DDPG models the action space using a metaheuristic-based technique. The proposed scheme implements a multiple reward system to generate Pareto optimal solutions and a decision-making technique to choose the compromise reward. Using real-world electricity prices and the IEEE 33-bus distribution network, numerical simulations show that our proposed TTIP scheme yields an average of 18% improvement in grid stability than the sustainable policy following, random, and price-greedy algorithms. It also improves the EV charging profit margins by an average of 28%.**

***Index Terms*—Charging price, distribution networks, electric vehicle (EV) charging stations (EVCSs), EVs, multiple rewards system, policy gradient, reinforcement learning (RL), transportation.**

Manuscript received 1 January 2023; revised 16 May 2023 and 8 June 2023; accepted 7 August 2023. Date of publication 31 August 2023; date of current version 23 February 2024. This work was supported in part by the Council for Scientific and Industrial Research, Pretoria, South Africa, through the Smart Networks collaboration initiative and IoT-Factory Programme (Funded by the Department of Science and Innovation (DSI), South Africa) and in part by the National Research Foundation of South Africa under Grant 148765, Grant 132651, and Grant 129311. Paper no. TII-22-5241. *(Corresponding author: Ling Cheng.)*

Kayode E. Adetunji, Ivan W. Hofsajer, and Ling Cheng are with the School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg 2050, South Africa (e-mail: kayvins@gmail.com; ivan.hofsajer@wits.ac.za; ling.cheng@wits.ac.za).

Adnan M. Abu-Mahfouz is with the Council for Scientific and Industrial Research, Pretoria 0184, South Africa (e-mail: a.abumahfouz@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2023.3305682.

Digital Object Identifier 10.1109/TII.2023.3305682

I. Introduction

The transportation sector is one of the largest emitters of greenhouse gas, owing to about 7.3 billion metric tons of carbon dioxide emissions [1]. Over 58% of fossil fuel usage is consumed by the transport sector [2]. While electric vehicles (EVs) are deemed the solution for mitigating carbon emissions from the transport sector, they can potentially cause damage to the utility grid. The EV charging coordination model is recently studied because of the potential to mitigate the adverse effects of uncoordinated charging on the utility grid. The model is equipped with a control strategy that uses charging price signals to encourage EV drivers to choose the suggested charging station. This approach is firmly based on the assumption that drivers usually choose the cheapest charging station on the go. However, recent studies have criticized this approach, positing that EV drivers are likely not to accept the suggested charging station but choose their preferred choice based on an impulse [3], [4]. This setback may cause the model to be impracticable. A two-stage incentive-based control strategy will be introduced to incentivize EV drivers to accept recommended charging stations. The strategy involves offering the driver multiple options to choose from. Furthermore, a mechanism is developed to handle situations where the driver declines, which resuggests charging stations according to the charging station the driver opted for instead of the previously recommended one.

Authors have applied different optimization techniques to solve the EV charging coordination problem. Liu et al. [5] developed a differential Evolution variant to find an optimal charging schedule and travel route while minimizing average time cost, charging cost, and EV's state of charge (SoC). Sun et al. [3] used a game-based technique to minimize waiting time, charging time, and driving time to EV charging stations (EVCSs) in EV charging coordination scheme. A correlated equilibrium strategy was implemented to cater for more than one potential EV charging service, considering neighboring EVs already charging, EVs that have submitted charging requests, and EVs that have not submitted a charging request. Ding et al. [6] modeled the interaction between the EV aggregator and the EVCS utility. The authors use a look-ahead price incentive greedy (LPIG) approach, where day-ahead electricity prices are considered to find the optimal charging price and operating profit in an EV charging coordination model. To improve the robustness of handling electricity prices, Korcas et al. [7] considered different electricity prices in their control strategy for buildings comprising HVACs, EVs, and battery systems. While these optimization techniques offer optimal EV charging decisions and energy management schemes, they are restricted to a static dataset and predetermined model information.

Reinforcement learning (RL) techniques, such as Q-learning, have been used to develop EV charging control strategies. However, these techniques suffer from the dimensionality curse obtained from large and dynamic Q-tables. Hence, authors in [8] and [9] used linearization techniques to approximate states and actions in an EV's scheduling problem. However, the major drawbacks to linearization techniques are the low tolerance to flawed parameters or feature states and the incapability to handle large Q-tables, which are mainly required for real-world scenarios. Authors have developed an approximate dynamic programming (ADP) approach to tackle this setback. Korkas et al. [10] developed a multimodal ADP to propose optimal charging/discharging schedules for EVs in a power system network setting, using continuous space and actions in the model. The technique could handle stochastic EV arrival and departure times, emphasizing its robustness in complex applications. While their study modeled the cost objective for a relatively long time horizon, the consideration of decision-making was barely considered. Decision-making in the RL scheme is important for generating correct optimal policies for applications like the EV charging coordination model; hence, Liang et al. [11] developed a comprehensive reward system to find optimal policies in an Internet of vehicles application, aiming to improve the travel experience, optimize traffic management efficiency, and improve road safety. The reward system accommodates the revenue, expenses, and additional costs, where each term varies dynamically according to the change in the offered services. Zhang et al. [12] developed a reward mechanism that selects the preference between the optimal travel distance and charging time to determine the desirability to transition from the current charging station to another. In [4], constraints are applied in the reward function to enforce the conditions that EVs are fully charged before a transition occurs in a deep RL (DRL)-based EV control strategy. Shin et al. [13] developed a comprehensive reward function that implements a summation equation using cost rewards, benefit rewards, shared rewards, and overcharging rewards, to determine the next battery power, charging price, and load state in a deep Q-network (DQN)-based energy management scheme. A similar reward function was also used in [7], where energy costs and user satisfaction are modeled as a multicriterion cost objective for performance index. However, these studies have only considered an a priori approach where multiple rewards are summed up before the learning process. This approach potentially introduces a bias in the learning process, making the final results unreliable [14], [15].

A key observation in the literature is that the price-based EV control strategies are, to some degree, trivial to the EV coordination model, not being enough incentive for EV drivers to accept charging station recommendations. EV drivers can reject recommended EVCSs, which may negatively affect the authenticity of the EV charging model. To enhance existing literature, this article proposes a charging incentive price scheme that offers price and parking time incentives to EV drivers to encourage the acceptance of EVCS suggestions. A stochastic learning technique is developed for our proposed pricing scheme, using a cooperative mechanism from the whale optimization algorithm (WOA) and RL to create Q-tables. Finally, a highlighted gap in previous studies is the trivially formulated multireward systems, which do not represent the multifaceted benefits of EV charging scheduling models. This article develops a multiple rewards system based on a multiobjective optimization (MOO) scheme, which considers charging cost minimization, profit maximization, grid stability improvement, and emission cost reduction.

The contributions of the research are as follows.

1) This article develops a novel two-tailed charging price incentive control scheme that facilitates the recommendation of charging stations to drivers. Unlike the conventional approach, where charging prices are offered only in the event of EV charging requests, this approach provides charging price incentives together with an allotted parking time for the current charge and a submechanism to instantaneously resuggest a charging station if a driver rejects the first recommendation.
2) To handle the continuous states and actions coupled with the time-varying dynamic Q-tables reported in the literature [8], [9], we develop a cooperative WOA-RL that maps the Q-table into a population-based matrix and uses the DDPG algorithm to find an optimal strategy for recommending EVCSs.
3) Finally, we develop a multiobjective RL decision-making model that generates Pareto optimal solutions that represent a balance among the technical, economic, and environmental objectives. Previously developed decision-making models [11], [12] have only considered one representative value for all objective functions. Our proposed model ensures the practicality of optimization results using the multiple reward system.

The rest of this article is organized as follows. Section II discusses the EV charging coordination system model while Section III discusses the problem formulation; Section IV discusses the proposed RL technique; Section V provides the simulation results to prove the effectiveness of the proposed framework. Finally, Section VI concludes this article.

II. System Model

In this article, we propose an EV charging-scheduling model based on the complementarity of charging price incentives and parking time to optimize the operational decisions for a distribution network operator (DNO) and an EV aggregator. The proposed model formulates the problem as the responsive EV aggregator actions to recommend a charging station while ensuring a balance between the distribution network performance and the maximum participation of EV drivers. The system model is related to an energy management system where energy usage can be controlled by shifting potentially destructive loads to less vulnerable buses on the distribution

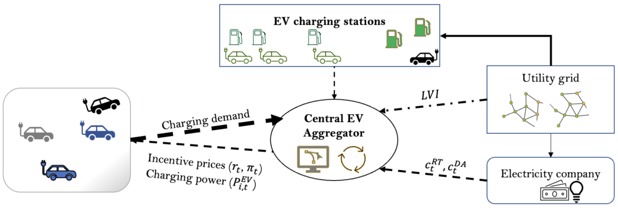


Fig. 1. Proposed EV charging coordination model.

network using price signals. The model consists of a central aggregator that handles the scheduling of EVs and provides an optimal pricing strategy while satisfying the EV drivers, EVCS operators, and the DNO. These goals are achieved by using an EVCS model and a distribution network model. Fig. 1 illustrates the working process of the model, where the EVs make a charging demand at different scenarios; hence, different shades. The central EV aggregator offers a charging price and an incentive for the next charge. The next charge is based on the assumption that the EV charges at the next horizon. The central aggregator tracks the following three entities: the utility grid, the electricity company, and charging stations.

A. EVCS Model

Consider the operation of a central EV charging control center over a time horizon T, where EV drivers request a charging slot t, at a charging station p connected to a bus k of a distribution network. The control center recommends charging stations for each EV by sending a list of prices for nearby charging stations. It is assumed that the EV driver will choose the charging station with the most inviting price. The transportation networks have been avoided in this study to avoid using sensitive information by the EVCS center. EV drivers will only decline when there is an obvious emergency, such as a case of a meager charge left, which is catered for while generating price lists. The SoCs are not considered, given the proportionality to charging demand. The SoC indicates the charging power needed from the charging station for a time duration. Therefore, our focus is shifted from the individual SoCs of EVs to the charging demand from the EVCS utility. In the proposed model, we categorize the EV into the following three:

1) EVs that are already charging;
2) EVs that are proceeding to their recommended charging stations;
3) EVs that have submitted a charging request.

Using notation, we denote every EV i has an estimated arrival time, parking time, and charging demand. The demand will be used to allocate the maximum parking time for charging. The parking time and the demand must be fulfilled before the EV leaves the charging station gives a large degree of freedom, making our EV charging model susceptible to high subjectivity. The EV serves as a function of the time remaining of complete charge; hence, the EVs that are already charging are defined as

$$\mathcal{C}_t^i = \mathrm{t}_{\mathrm{ass}}^i$$

where $\mathrm{t}_{\mathrm{ass}}^i$ is the time assigned along with the charging price to EV$_i$ and $\mathrm{t}_{\mathrm{arr}}^i$ is the estimated arrival time of EV i. $\mathrm{t}_{\mathrm{arr}}^i$ will increase at every time interval, thereby reducing the value if parking is left to zero. The EVs arriving at time t, are denoted as $\mathcal{A}_t^i = \mathrm{t}_{\mathrm{ass}}^i - \mathrm{t}_{\mathrm{arr}}^i$, where $T = 0$, hence, $\mathcal{A}_t^i$ will be a function of $\mathrm{t}_{\mathrm{ass}}$ until it has started the charge. Both $\mathcal{C}_i^t$ and $\mathcal{A}_i^t$ will reach zero to end charging. A penalty coefficient, ζ is introduced to constrain the model toward a policy that allows full charge. The penalty is multiplied by the final value of each EV in the set $\mathcal{C}_t^i$, Therefore, a successfully completed charge would have a zero penalty, while an incomplete charge would have a penalty value related to the time of "early" departure.

1) Charging Cost Minimization: This objective is an important aspect of a centralized EV charging control since it can closely determine the acceptance rate of such a model. A high charging cost may cause a high rejection rate, which can cause an uncoordinated charging system. The objective is a function of the incentive price offered to the EV driver, the regular charging price, and the charging demand, defined as

$$\mathrm{CC} = \min \sum_{i \in \mathfrak{I}} \sum_{t \in T} P_{i,t}^{\mathrm{EV}} \cdot c_t^{\mathrm{RT}} + \sum_{i \in \mathfrak{I}} \sum_{t \in H} P_{i,t}^{\mathrm{EV}} \cdot c_t^{\mathrm{DA}} \tag{1}$$

where c_t^{RT} and c_t^{DA} are the real-time (or current) and day-ahead (next charge) electricity prices, respectively; $P_{i,t}^{\mathrm{EV}}$ is the charging power demand of EV i at time t, either at the current time horizon T or the next horizon H.

2) Profit Maximization: This objective is considered for the EV aggregators whose business lies in making a profit while offering charging services to EV drivers. The objective is the difference between the payment collected from the EV driver after applying the current charging price, considering the current and day-ahead electricity price, defined as

$$\begin{aligned}\mathrm{PM} = \ &\max \sum_{t \in T} (r_t - \pi_t) - c_t^{\mathrm{RT}} P_{p,t}^{\mathrm{EVCS}}) \\ &+ \sum_{t \in H} \left[r_t - c_t^{\mathrm{DA}} P_{p,t}^{\mathrm{EVCS}} \right]\end{aligned} \tag{2}$$

where r_t and π_t are the current charging price and the incentive price for the next charge, respectively; $P_{p,t}^{\mathrm{EVCS}}$ is the EVCS charging capacity, which is the energy procured to the EVCS at different time horizon.

3) EVCS Constraints:

$$\pi^{\min}, r^{\min} \leqslant \pi, r_{i,t} \leqslant \pi^{\max}, r^{\max} \tag{3}$$

$$p_{i,t}^{\mathrm{grid}} + p_{i,t}^{\mathrm{PV}} = p_{i,t}^{\mathrm{EV}} + p_{i,t}^{\mathrm{EV.base}} \tag{4}$$

$$P_i^{\mathrm{EV.min}} \leqslant P_{i,t}^{\mathrm{EV}} \leqslant P_i^{\mathrm{EV.max}} \tag{5}$$

$$\sum_{i \in \mathcal{R}_t} x_t^i \leqslant e^{\max} \tag{6}$$

$$\sum_{T_{\mathrm{arr}}} x_t^i \geqslant E_i^d. \tag{7}$$

The charging incentive prices are limited using Constraint (3). Constraint (4) ensures that there is a power balance at every charging station at time t while Constraint (5) keeps the EVCS charging power $P_{i,t}^{\mathrm{EV}}$ within the set limit. Constraint (6) ensures that the total number of EVs to be charged does not exceed the maximum charging rate of the EVCS p. Lastly, Constraint (7) ensures that every EV's charging demand is satisfied before leaving the charging station.

B. Distribution Network Model

The considered technical objectives are to ensure grid performance while accepting charging at certain EVCSs. We monitor the grid performance by reducing the overloading of distribution network lines. To achieve this, we minimize the real peak power of a bus that is connected to a particular charging station, expressed as

$$\sum_{k=1}^{K} P_{k,t}^{\text{LOAD}} \tag{8}$$

where $P_{k,t}^{\text{LOAD}}$ represents the peak real power from bus k (which corresponds to CS k) at time t. With the monitoring of (8), the technical objectives can be calculated and are defined in the following sections.

1) Load Variance Index: Given that EVs act as a dynamic load, there is a need to minimize the fluctuations on the grid through a proper charging scheduling. An optimal cancelation of this objective can reduce power loss. A typical load variance index (LVI) objective is formulated as

$$\text{LVI} = \min \frac{1}{T} \sum_{t\in T} \left[\left(P_t^{\text{conv}} + \sum_{i\in \Im} P_{i,t}^{\text{EV}} - \beta \right) \right]^2 \tag{9}$$

where

$$\beta = \frac{1}{T} \sum_{t\in T} \left(P_t^{\text{conv}} + \sum_{i\in \Im} P_{i,t}^{\text{EV}} \right). \tag{10}$$

Here, P_t^{conv} is the conventional load without the EV load at time t and $P_{t,n}^{\text{EV}}$ is the charging power of the nth EV at time t. It is crucial to note that N is the maximum number of charging stations, which is assumed to be occupied by EVs at a predetermined time slot, t.

2) Emission Cost Reduction: The emission cost objective is formulated as the carbon emission value of real power from a substation [16], and can be defined as

$$\text{ECR} = \min \sum_{t\in T} P_{p,k,t}^{\text{EVCS}} \times E_f \tag{11}$$

where E_f is the emission in kg/kWh and $P_{p,k,t}^{\text{EVCS}}$ is the fossil fuel-based real power from the substation that feeds EVCS p from bus k at time t.

3) Distribution Network Constraints:

$$P_D^{\max} \geqslant P_t^D + P_{p,k,t}^{\text{EVCS}} \tag{12}$$

$$V_k^{\min} \leqslant V_{t,k} \leqslant V_k^{\max}. \tag{13}$$

Constraint in (12) shows the transformer capacity must not be exceeded and must be bounded to the sum or less the sum of the conventional load demand P_t^D and the EVCS charging rate $P_{p,k,t}^{\text{EVCS}}$ at time t. Constraint (13) sets the lower and upper voltage limits per unit, with $V_k^{\min}$ and $V_k^{\max}$ being 0.98 and 1.01, respectively.

C. Utility Maximization Model

This article's main objective is to maximise the EVCS and distribution network utility by deciding the scheduling and pricing policies. The optimization problem will be

$$F_t = \{\min \text{CC} | \max \text{PM} | \min \text{LVI} | \max \text{ECR}\} \tag{14}$$

where F_t is the utility value computed by generating Pareto optimal solutions and the technique for order of preference by similarity to ideal solution (TOPSIS) decision-making technique. The Pareto solutions are generated using the dynamic weight assignment.

D. Two-Tailed Incentive Pricing (TTIP) Scheme

The EV drivers are motivated to choose the recommended charging station by developing a TTIP scheme. Contrary to most pricing mechanisms, where EV drivers are offered exact charging prices for certain EVCSs, this article offers a price incentive to EV drivers, where the mechanism sets the charging price incentive together with an estimated parking time. The parking time is an additional motivator for drivers to choose the recommended charging station. The recommended charging station will have a charging price with a high parking time to avoid penalties.

The control center collates the payment $r_i^t (i \in \mathcal{I}_t)$ of all EVs, where $\mathcal{I}_t$ is a set of EVs requesting a charging slot at time t. The center then collects the electricity bill information c^t and c_h (where $t \in T$ and $h \in H$), and offers the price incentive π_t. T and H are time horizons, where T is the current horizon and H is the next horizon in which the day ahead electricity price is considered for profit calculation. The EV drivers that accept the incentive price π are moved from set $\mathcal{I}_t$ to set $\mathcal{A}_t$. It is to note that π_t is offered at the current time horizon only. As discussed in Section II-A2, the charging price is fixed for T and H horizons, but the price incentive is deducted from the charging price of the current time horizon, T.

The incentive-based pricing mechanism is implemented as a sequential decision problem, offering an incentive for EV drivers together with a corresponding parking time for a recommended charging station. A detailed approach is presented in Section III.

III. Problem Formulation Using Markov Decision Process (MDP)

A. Markov Decision Process

For the EV charging scheduling model, each time slot t comprises a charging request, which is returned with a charging price and parking time for a recommended charging station. The modeling of the state, action, and reward is discussed as follows.

1) State Space: The system state space at every time slot t is represented by the set of EVs charging at time t, the charging demand, and the parking time, denoted as a vector $S = \{C_t, D_t, p_t\}$. The parking time T_{ass} is assigned as a function of the estimated arrival time and the charging demand.

2) Action and Transition Function: The action taken, A_t at every time t is the incentive price and the scheduling power, denoted by a vector $A_t = \{\pi_t, p_i^t | \forall i \in R_t\}$. Note that action A_t will be for EVs that made a charging request, knowing the action taken at a particular state, the transition function from the current state S_t to the next state S_{t+1} will be determined by the set of EVs charging at time t and the set of EVs that made a charging request at time t, given as $\mathcal{C}_{t+1}$. Therefore, the next state is formed as

$$S_{t+1} := f(S_t, A_t, \mathcal{C}_{t+1}). \tag{15}$$

3) Reward Function: The reward function, $R : S \times A \to \mathbb{R}$, reflects the decision quality to transit from state S_t to state S_{t+1}. The decision quality is based on different reward forms, which is the objective function of equation (23). The immediate reward is the central EV aggregator assignment of an EV i to a charging station at time t, as shown in [16]. We introduce a multiobjective RL (MORL) approach to handle the maximization of the long-term utility of the central EV aggregator, given as

$$R_t(S_t, A_t) := \max \left[\sum F_t\right]. \tag{16}$$

The Q-learning-based algorithm is proposed to solve the problem as per (14). However, the action to solve the problem is continuous; therefore, it is assumed that action A_t follows a stochastic policy $\pi(A|S)$. The effect of the driver choosing action A_t in state S_t, following policy π is calculated as

$$Q^{\pi}(S_t, A_t) = \mathcal{E}\left[R_t + \gamma Q^{\pi}(S_{t+1}, A_{t+1})\right] \tag{17}$$

s.t. constraints in (3)–(7), (12)–(13).

The $Q^{\pi}(S_t, A_t)$ term represents the state-action value function and $\mathcal{E}[\cdot]$ is the expectation over a horizon T. The γ sign represents the discount factor that describes the balance between short and long-term rewards. Since the Q-learning algorithm is a model-free algorithm, which does not require information, such as future electricity prices, EV driving patterns, and road networks, The $Q^{\pi}(S_t, A_t)$ function learns the quality of choosing an action in a certain state and using the policy π to make decisions.

Finding the optimal policy that demonstrates, which action to take at a certain model state is the main task of an RL algorithm. However, the EV charging scheduling model consists of continuous state and action pairs that need to be discretized, yielding an extremely large table that becomes too complex. In addition, the table also varies because of the arrival and departure of EVs at different time intervals. These setbacks can be tackled using the deep neural network to approximate the state and action pairs. The RL algorithm can learn the optimal strategy through a recursive update of the state-action value function from (18), given as

$$\begin{aligned} Q(S_t, A_t) \Longleftarrow (1-\alpha)Q(S_t, A_t) + \alpha_t[R_t(S_t, A_t) \\ + \gamma[Q(S_{t+1}, A_{t+1})]]. \end{aligned} \tag{18}$$

The learning rate $\alpha \in [0, 1]$ is applied to control the preferential rate between the old value of $Q_t(S_t, A_t)$ and the term $\alpha_t[R_t(S_t, A_t) + \gamma[f(S_t, A_t, A_t)]]$, which also decreases to zero.

To maximise the reward, EV charging scheduling aims to find optimal policies π over all feasible policies. Hence, given as

$$Q^*(S, A) = \max_{\pi} Q^{\pi}(S_t, A_t). \tag{19}$$

Given that the pricing scheme is offered for two charging terms, the decisions are based on two scenarios; the current charging request and the next charging request. Hence, we implement two horizons to cater for the two price incentives.

B. Multiobjective MDP

This is the regular MDP but with a different reward function $R : S \times A \to \mathbb{R}^n$; n being the number of objectives. A multiobjective property $\phi = \{R_1(\max), \ldots, R_n(\min)\}$ aims to simultaneously optimize n number objectives of the expected total rewards. A multiobjective property ϕ in the MDP $\mathcal{M}$ will produce a set of Pareto optimal solutions that bounds a set of feasible solutions, $X = \{x \in \mathcal{R}^n\}$. An optimal solution x^* is Pareto optimal if another point $x \in X$ does not exist in feasible set X such that $x_i \leqslant x_i^*; \forall i$ The feasible of x for ϕ must also follow a convex structure.

Given the combination of decision makers' preferences, $\tilde{w}$ in a multiobjective $\mathcal{M}$, the expected total weighted reward sum is $\mathcal{E}^{\pi}_{\mathcal{M}}(w \cdot R) \overset{\text{def}}{=} \sum_{i=1}^{N} w_i \mathcal{E}^{\pi}_{\mathcal{M}}(R_i)$, where $\pi \in \mathcal{E}_{\mathcal{M}}$ is the policy. The optimal strategy π^* becomes truly optimal wrt ϕ and w if it corresponds to a unique Pareto optimal solution in a Pareto front.

A weighted sum approach will be used to solve the MORL problem, which stems from the utility function in (14), involving four objective functions. As discussed in [14], there can be more than one reward while attaining a certain policy, but a naive summation (i.e., scalarization) is not the appropriate approach to handle such cases, especially in our model where the objectives are not closely related to each other. The model interests a DNO and an EV aggregator. Hence, a weighted sum approach aggregates the preference for multiple rewards. Given this, the state-action value functions will be linearly assigned weights as

$$Q(S, A) = \sum_{i=1}^{N} w_i Q_i(S, A) \tag{20}$$

where w_i is the ith weight assigned to the ith $Q(S, A)$ value.

The final step of the MORL technique is to select an action based on the array of Q-values produced from (20). We adopt the TOPSIS decision-making method to select a compromise. TOPSIS, implemented in [17] and [18], chooses a solution that has the shortest geometric distance from the best positive solution while ensuring the same solution is furthest from the best negative solution. Contrary to conventional MDPs where the maximum Q-value is the indicator for selecting the next action, our MOMDP model generates a Pareto set of optimal Q-values. The TOPSIS technique is then used to identify the compromise Q-value from the Pareto set, deciding the next action.

IV. Proposed DRL Approach

The common DRL approach is the DQN, which uses value-based methods, such as Q-learning and SARSA, and can only handle discrete and low-dimension domains. Since applications like the EV charging model have continuous action spaces and high dimensions, the best alternative is to discretize these continuous domains and use the DQN's experience replay mechanism to handle the resultant dimensionality curse problem. However, the generated samples are correlated to a great extent and are nonstationary, causing a slow convergence rate. Even further, discretizing the action space causes loss of information and introduces quantization noise that may hinder convergence, making it difficult to find the genuine optimal policy [19]. Therefore, we adopt the DDPG approach to model the action space.

A. DDPG Approach

This article uses the actor–critic learning framework to learn the optimal policy and the value function of the EV charging scheduling model. The actor and critic networks are interactively infused in a deep neural network, using weights θ_a and θ_c. The actor-network defines the policy, generating actions to be taken in a particular state. At the same time, the critic agent evaluates and finds fault in the current policy by handling the corresponding reward from the environment. The actor agent then uses the decision from the critic to update its policy. Introducing the critic agent gives a mechanism that monitors and reduces the error over time, thereby making the actor–critic method have a good convergence rate when learning from continuous domains with stochastic policies [20].

To make the learning performance of the DDPG more stable, both actor and critic agent must possess a corresponding target network, $\theta^{\mathfrak{t}}_{\mathbb{A}}$ and $\theta^{\mathfrak{t}}_{\mathbb{C}}$, to update the Q-values, represented as

$$\theta^{\mathfrak{t}}_{\mathbb{A}} \leftarrow \tau\theta_{\mathbb{A}} + (1-\tau)\theta^{\mathfrak{t}}_{\mathbb{A}} \tag{21}$$

$$\theta^{\mathfrak{t}}_{\mathbb{C}} \leftarrow \tau\theta_{\mathbb{A}} + (1-\tau)\theta^{\mathfrak{t}}_{\mathbb{C}}. \tag{22}$$

Each action a_t from a state s_t generates a new tuple $\{S_t, A_t, R_t, S_{t+1}\}$, which is stored in a replay buffer $\mathcal{D}$. The replay buffer is then sampled in mini-batches to update the actor and critic networks. For the continuous and infinite domain, the critic network is trained to find a policy π, minimizing the loss function represented as

$$\mathcal{L} = \frac{1}{T}\sum_{t\in T}(y_t - \mathbb{C}(S_t, A_t|\theta_{\mathbb{C}})^2 \tag{23}$$

where

$$y_t = R_t + \gamma \max_{a_{t+1}\in\mathcal{A}} Q(s_{t+1}a_{t+1}|\theta_{\mathbb{A}}). \tag{24}$$

B. Cooperative WOA-RL

This article instantiates the cooperative WOA-based RL (WOA-RL) approach by combining the DDPG's mechanism with the WOA's, similar to neuroevolution strategies. Neuroevolution is a branch of machine learning that applies evolutionary algorithms to neural networks [21]. The main idea behind developing the cooperative WOA-RL is to incorporate the WOA's population-based approach for producing diverse experiences while leveraging the DDPG's powerful gradient-based mechanism to learn from them. The WOA is from a family of metaheuristics that uses a probabilistic wheel to generate a new population with a better fitness value, or in this article, the DDPG actor. The WOA's mechanism ensures that actors with the best solutions have a higher chance of being selected and passed on to the next population.

The generalized flow of the WOA-RL is to initialize a population of actor-networks, and one additional actor–critic network pair, all with random weights; the fitness for each actor-network in the initialized population is evaluated in an interactive episode, computed as the cumulative sum of reward received over the time intervals in that episode; the actors in the population are passed through the WOA update mechanism [from (WOA-1) and (WOA-2) in Algorithm (1)] to produce the next population of actors. The WOA-RL retains the fitness weights as an experience in the DDPG's replay buffer and learns from them repeatedly within those episodes. The fitness weights are evaluated using Algorithm (2). The algorithm evaluates each actor in the initialized population using the MOMDP structure, discussed in Section III. The WOA-RL for every episode at each time interval, stores each actor network's experience (expressed as the tuple $\{S_t, A_t, R_t, S_{t+1}\}$ in its replay buffer). The next steps are continued in the DDPG architecture, where the critic network uses gradient descent to train and update the small batches from the replay buffer, and the actor network is trained using the sampled policy gradient. The target networks are finally updated [from (21) and (22)] using the trained actor and critic weights.

V. Simulation Results

This section evaluates the proposed TTIP performance considering its optimality results, efficiency, and practicality. We investigate the framework's impact on the grid and observe the added benefits compared to previously proposed models. We used Python as the simulation software, running on an INTEL core-i5 10505 CPU at 3.20 GHz speed using 8 GB RAM.

A. Experimental Setup

We use historical data for electricity prices and EV arrivals from [22] and [23], respectively. The EV requests are categorized into three; urgent, normal, and opportunistic, with a variance of 5.1, 3.7, and 7.8, respectively. The model is simulated in a one-minute resolution, which can be adjusted to preference. From the learning process, the discount factor and learning rate are set to 0.9 and 0.95, respectively. Similar to [7], the electricity prices are modeled to change in different frequencies, emulating the market dynamics.

To evaluate the performance of our proposed pricing scheme, we compare the TTIP results with other related benchmarks, evaluating them on a modified IEEE 33-bus distribution network. The network is preloaded with EVCSs, PV plants, and capacitor banks [24], as shown in Fig. 2. For a fair comparison, we develop the actor/critic DDPG architecture for policy benchmarks. For example, we did not select the DQN as a benchmark because of the extra step of action space discretization, which performs relatively poorly. Our focus is on producing an optimal pricing policy based on an action selection, therefore, we adopted the benchmarks explained as follows.

a) Current Price Incentive-Based Greedy (CPIG) Scheme: Recommends the highest incentive to EV drivers, using the current real-time electricity price. This scheme is similar to a price-based greedy scheme in [8].

Algorithm 1: WOA-RL Training Process.

Initialize actor and critic network, $\mathbb{A}$ and $\mathbb{C}$, with parameters $\theta_{\mathbb{A}}$ and $\theta_{\mathbb{C}}$ respectively
Initialize $\mathbb{A}^{t}$ and $\mathbb{C}^{t}$ network, with parameters $\theta^{t}_{\mathbb{A}}$ and $\theta^{t}_{\mathbb{C}}$ respectively
Initialize a population of a actors $pop_{\mathbb{A}}$ and an empty cyclic replay buffer $\mathcal{D}$
for all iteration $\mathrm{t} = 1$ **to** $\approx$ **do**
  for actor $\mathbb{A} \in pop_{\mathbb{A}}$ **do**
    $U, \mathcal{D} \leftarrow eval(pop_{\mathbb{A}}, \mathcal{D})$
  end for
  Generate $a, A, \mathfrak{p}, \&C$
  for all $\mathbb{A} \in pop_{\mathbb{A}}$ **do**
    if $|\mathfrak{A}| \geq 1$ **then**
      Update next $\mathbb{A}$ using the WOA mechanism

$$\overrightarrow{\mathbb{A}}_{t+1} = \begin{cases} \overrightarrow{\mathbb{A}}^{*}_{t} - A \cdot \overrightarrow{D}, & \mathfrak{p} < 0.5 \\ \overrightarrow{D} \cdot \ e^{bl} \cdot \cos(2\pi l) + \overrightarrow{\mathbb{A}}^{*}_{t} & \mathfrak{p} \geqslant 0.5 \end{cases} \quad (\mathrm{WOA}-1)$$

    else
      Choose a random actor $\mathbb{A} \in pop_{\mathbb{A}}$ and update next $\mathbb{A}$ using

$$\overrightarrow{\mathbb{A}}_{t+1} = \overrightarrow{\mathbb{A}}_{rand} - \mathfrak{A} \cdot \overrightarrow{D} \qquad (\mathrm{WOA}-2)$$

    end if
    $-, \mathcal{D} \leftarrow eval(\mathbb{A}, \mathcal{D})$
    Generate initial state S_t
    for all $C = 1$ **to** N_{CR} **do**
      Randomly sample a batch of transitions, $B = \{< S_t, A_t, R_t, S_{t+1} >\}^{B}_{i}$ from $\mathcal{D}$
      Compute the target using (24)
      Update $\mathbb{C}$ by minimizing the loss using (23)
      Update $\mathbb{A}$ using the sampled policy gradient;

$$\nabla_{\mathbb{A}} J \approx \frac{1}{|\mathcal{B}|} \sum \nabla_{a} \mathbb{C}(s, a|\theta_{\mathbb{C}}) \nabla_{\theta_{\mathbb{C}}} \mathbb{A}(S|\theta_{\mathbb{C}})$$

      Update the target networks using (21) and (22)
    end for
  end for
end for

Algorithm 2: Function Eval.

Procedure
  eval$(pop_{\mathbb{A}}, \mathcal{D})$
  $U \leftarrow 0$
  for $i = 1$ **to** ε **do**
    Reset the environment, get the initial state, S_t
    while EV $i \nexists \mathfrak{I}_t$ **do**
      Select an action A_t
      Execute A_t, observe reward R_t, receive new data, and generate new state S_{t+1}
      Append transition tuple $< S_t, A_t, R_t, S_{t+1} >$ to $\mathcal{D}$
      $U \leftarrow R_t$, $S_t \leftarrow S_{t+1}$
    end while
  end for
  return $U, \mathcal{D}$
End Procedure

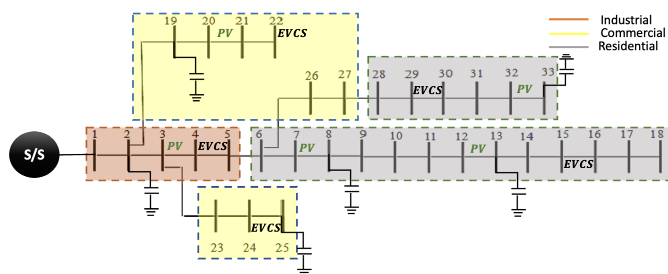


Fig. 2. Modified IEEE 33-bus distribution network for evaluation purposes.

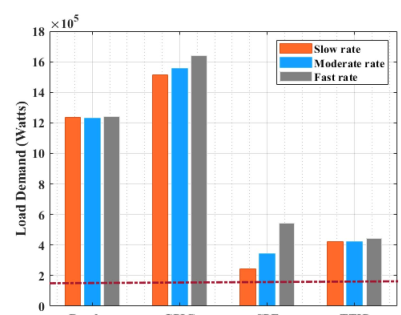


Fig. 3. Load variance of the grid at different electricity pricing frequencies.

b) LPIG scheme: Recommends the highest next-day incentive to attract EV drivers, using the day-ahead electricity price. The authors in [6] used this approach.

c) Random Policy: Randomly takes actions to attract EV drivers, guided by an epsilon value [3].

d) Sustainable Policy Following (SPF): Similar to the work in [9], the agent prioritizes using renewable energy for charging.

B. Optimality Result

The first set of experiments is carried out to test the results of the proposed TTIP framework.

1) Impact on the Utility Grid: The load variance is investigated to show the stability of the grid. While at operating conditions, frequent load variation can alter the system's frequency. How does our proposed framework affect the utility grid without additional control equipment? Fig. 3 illustrates the load variance level of the grid when different EV charging schemes are implemented. The SPF scheme depicts the most performing load variance with 37% off the base case. The TTIP followed with a 41% variance from the base case.

2) Incentive Offer Performance: We evaluate the performance of offering incentives per the EVCS capacity considering the moderate electricity price frequency. Incentives must not only be beneficial to the EV drivers but must practically follow the EVCS capacities. Fig. 4 illustrates the accepted incentives and the EVCS capacity. First, it is seen that all schemes offer incentives according to the change in electricity prices and do not exceed EVCS except for the random policy scheme. It is

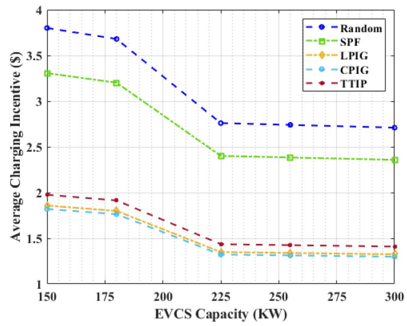
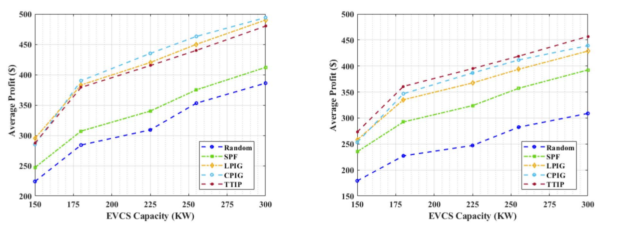


Fig. 4. Average charging incentive for different EVCS capacities.



(a) (b)

Fig. 5. Average profit for different EVCS capacities. (a) Regular electricity price frequency. (b) Fast-varying electricity price frequency.

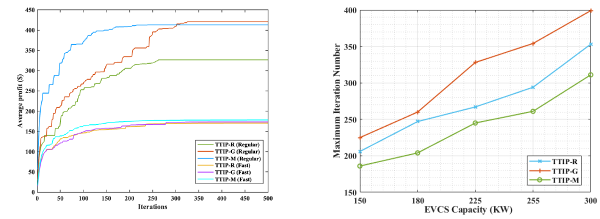
seen that the proposed scheme offers the highest incentive than other schemes.

3) Profit Margin Evaluation: While it is important to satisfy EV drivers enough to choose an EVCS location, the EV charging aggregator must also benefit from the EV charging business. The charging power is the commodity sold to EV drivers. Hence, we plot the profit according to the available commodity for EVCS facilities, as shown in Fig. 5(a). It is observed that the CPIG and LPIG schemes yield the best profit. However, it is to note that these schemes compensate for high profit with the utility grid performance. Another experiment is carried out to know the profit according to the fast electricity price. It is seen in Fig. 5(b) that the fast varying electricity price produces the least profit for all schemes, including TTIP. However, the TTIP maintains a relatively stable profit margin, making it perform better than other schemes across the board.

C. Efficiency Results

As part of the ablation study for our proposed scheme, we evaluate the efficiency of the different TTIP implementation schemes by investigating the time complexity and convergence characteristics. The schemes are differently implemented based on their action selection strategy. In addition to our proposed metaheuristic-based TTIP (TTIP-M), we developed TTIP-greedy (TTIP-G) and TTIP random (TTIP-R).

1) Convergence Analysis: We illustrate the convergence characteristics of the benchmarks on the 225 KW EVCS capacity in Fig. 6(a). The TTIP-M shows the average best performance for the regular and fast-varying electricity price frequency. It is seen that all the implementations perform relatively poorly in the fast-varying electricity price scenario.



(a) (b)

Fig. 6. Average profit for different EVCS capacities. (a) Convergence of TTIP for different price frequencies. (b) Plot of converging iteration number and EVCS capacities.

TABLE I
COMPUTATION TIME (IN SECONDS) FOR A 500 KW EVCS CAPACITY

Scheme	Slow	Regular	Fast
TTIP-R	438	451	514
TTIP-G	551	565	849
TTIP-M	355	382	621

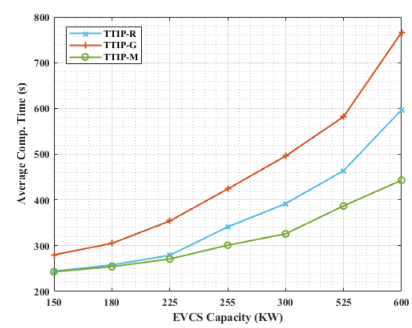
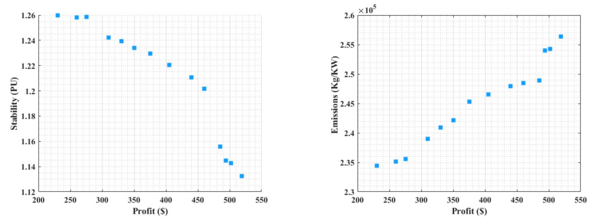


Fig. 7. Plot of computational time and EVCS capacities.

To show the performance of each implementation for different scales, the convergence graph is plotted in Fig. 6(b), showing the maximum iteration number for convergence at different EVCS capacities in a regular electricity price frequency scenario. It is observed that while the iteration increases concerning the increase in EVCS charging capacity, the TTIP is slower to increase iteration number than other benchmarks, showing a higher resilience in handling large-scale systems.

2) Computational Time Complexity Analysis: We evaluate the time complexity of the three TTIP implementations. Table I shows the average computational time for each technique to produce results for an EVCS capacity of 500 KW at different frequencies of electricity prices. For the slow and regular frequencies, the TTIP-M scheme has the lowest computational time, followed by the TTIP-R. However, for the fast-varying frequency, the TTIP-R scheme has the fastest computational time complexity, followed by the TTIP-M. To see the behavior of the schemes, we plot a graph in Fig. 7 illustrating computational time for different EVCS capacities at a regular electricity price frequency. It is seen that all schemes produce computational times that increase exponentially with the increase in EVCS capacity. However, it is observed that



(a) (b)

Fig. 8. Pareto front, (a) profit and grid stability plot, (b) profit and carbon emission plot.

TABLE II
PARETO OPTIMAL SOLUTIONS QUALITY EVALUATION OF THE MOMDP

Stats	SP-metric	IGD	Hypervolume
Best	0.065	0.034	0.842
Mean	0.072	0.053	0.849
Variance	1.67	2.49	2.03

the TTIP-M possesses the lowest exponential rate, which indicates a better performance than other implementations in large-scale systems. A few higher EVCS capacities may become very complex for other benchmarks, especially for the TTIP-R scheme.

D. Multiobjective Learning Evaluation

The feasibility of implementing simulation results in real-world scenarios lies in practicality. First, the model must have considered important benefits and handled them in a way that fairly represents all benefits (or rewards) without minimal bias. Second, there should be room for deliberation in choosing a compromise solution out of many potential good solutions. Formed as rewards, the variation of weights is applied to multiple objectives adopted in the framework to generate Pareto optimal solutions. A TOPSIS approach is used to select the compromise solution to decide the next state in transition in the learning process. Fig. 8(a) and 8(b) depicts the sample of Pareto optimal solutions between the profit and voltage stability and the profit and emission cost.

To evaluate the performance of generating Pareto solutions, we use the Spacing (SP-metric), inverted generational distance (IGD), and hypervolume metrics. Table II shows the statistical values of the metrics. The best, mean, worst, and standard deviation are computed for different MOO techniques. It is seen from Table II that the multireward system produces fairly good results, with extremely low values for the IGD and SP metrics. Since the hypervolume metric works with a reference point, which is set as $r = 0 \in \mathbb{R}_d^+$ where values in the Pareto set are positive numbers.

VI. CONCLUSION

In this article, we developed a TTIP scheme for suggesting EVCSs for EV charging requests considering load variation minimization, profit maximization, charging price minimization, and emission cost reduction. Given the potential rejection of EVCS recommendations, we modeled a mechanism that offers charging price discounts, accompanied by parking time. We formulated the EVCS recommendation framework as an MOMDP, and adopted the DDPG for handling continuous states and actions. To address the problem of action selection from a large search space, we implemented a metaheuristic-based selection technique, forming a population of actions at every iteration and selecting them according to the WOA mechanism. Furthermore, we applied a weighted approach to generate Pareto solutions of Q-values, and with a TOPSIS decision-making to select the compromise Q-value for the next action. Our proposed TTIP scheme was validated on the IEEE 33-bus distribution network using real-world electricity prices. By comparing with benchmark policies, simulation results show the significant performance of TTIP in reducing load variation at different electricity price frequencies. An optimal profit margin and charging price were also maintained at different EVCS capacities.

A random and greedy-based implementation—TTIP-R and TTIP-G, respectively, were developed to evaluate the efficiency. The proposed TTIP, dubbed TTIP-M, has the fastest convergence rate and the quickest computational time for regular varying electricity prices. The TTIP-R has the quickest computational time in the fast-varying electricity price condition; however, the TTIP is shown to perform better in grid stability and profit maximization. Finally, the generation of Pareto optimal values was evaluated. The statistical values show fairly good results. These values can, however, be improved using other forms of Pareto optimal solutions.

For future works, the proposed TTIP scheme can be improved to handle fast-varying electricity prices. Parameter tuning or a different metaheuristic algorithm can improve the efficiency of the model. Another direction is to improve the practicality of the model by implementing a noncooperative game among DNO, EV drivers, and EVCS operators. Finally, the MOMDP framework can be improved to produce more uniformly spaced Pareto solutions.

REFERENCES

[1] I. TISEO, "Global transport CO_2 emissions breakdown 2020," Dec. 2021. [Online]. Available: https://www.statista.com/statistics/1185535/transport-carbon-dioxide-emissions-breakdown/
[2] B. Salvi, K. Subramanian, and N. Panwar, "Alternative fuels for transportation vehicles: A technical review," *Renewable Sustain. Energy Rev.*, vol. 25, pp. 404–419, 2013.
[3] C. Sun, X. Wen, Z. Lu, J. Zhang, and X. Chen, "A graphical game approach to electrical vehicle charging scheduling: Correlated equilibrium and latency minimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 505–517, Jan. 2021.
[4] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2427–2439, May 2020.
[5] W.-L. Liu, Y.-J. Gong, W.-N. Chen, Z. Liu, H. Wang, and J. Zhang, "Coordinated charging scheduling of electric vehicles: A mixed-variable differential evolution approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5094–5109, Dec. 2020.
[6] Z. Ding, Y. Lu, K. Lai, M. Yang, and W.-J. Lee, "Optimal coordinated operation scheduling for electric vehicle aggregator and charging stations in an integrated electricity-transportation system," *Int. J. Elect. Power Energy Syst.*, vol. 121, 2020, Art. no. 106040.

[7] C. D. Korkas, M. Terzopoulos, C. Tsaknakis, and E. B. Kosmatopoulos, "Nearly optimal demand side management for energy, thermal, EV and storage loads: An approximate dynamic programming approach for smarter buildings," *Energy Buildings*, vol. 255, 2022, Art. no. 111676.
[8] S. Wang, S. Bi, and Y. A. Zhang, "Reinforcement learning for real-time pricing and scheduling control in EV charging stations," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 849–859, Feb. 2021.
[9] A. Chiş, 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.
[10] 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.
[11] H. Liang et al., "Reinforcement learning enabled dynamic resource allocation in the Internet of Vehicles," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4957–4967, Jul. 2021.
[12] C. Zhang, Y. Liu, F. Wu, B. Tang, and W. Fan, "Effective charging planning based on deep reinforcement learning for electric vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 542–554, Jan. 2021.
[13] M. Shin, D.-H. Choi, and J. Kim, "Cooperative management for PV/ESS-enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3493–3503, May 2020.
[14] M. Pirotta, S. Parisi, and M. Restelli, "Multi-objective reinforcement learning with continuous Pareto frontier approximation," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015.
[15] P. Mannion, S. Devlin, K. Mason, J. Duggan, and E. Howley, "Policy invariance under reward transformations for multi-objective reinforcement learning," *Neurocomputing*, vol. 263, pp. 60–73, 2017.
[16] S. Zeynali, N. Rostami, and M. Feyzi, "Multi-objective optimal short-term planning of renewable distributed generations and capacitor banks in power system considering different uncertainties including plug-in electric vehicles," *Int. J. Elect. Power Energy Syst.*, vol. 119, 2020, Art. no. 105885.
[17] K. E. Adetunji, I. W. Hofsajer, A. M. Abu-Mahfouz, and L. Cheng, "Category-based multiobjective approach for optimal integration of distributed generation and energy storage systems in distribution networks," *IEEE Access*, vol. 9, pp. 28237–28250, 2021.
[18] A. A. Al-Obaidi and H. E. Farag, "Decentralized quality of service based system for energy trading among electric vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6586–6595, Jul. 2022.
[19] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[20] Y. Wei, F. R. Yu, M. Song, and Z. Han, "User scheduling and resource allocation in HetNets with hybrid energy supply: An actor-critic reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 680–692, Jan. 2018.
[21] Y. Liu, Y. Sun, B. Xue, M. Zhang, G. G. Yen, and K. C. Tan, "A survey on evolutionary neural architecture search," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 2, pp. 550–570, Feb. 2023.
[22] CAISO, "Caiso: Day-ahead price—LCG consulting," Jun. 2017. [Online]. Available: http://www.energyonline.com/Data/GenericData.aspx?DataId=22&CAISO_Day-Ahead_Price
[23] UCDavis, "Richards ave station arrivals," May 2019. [Online]. Available: http://anson.ucdavis.edu/clarkf/richards.csv.gz
[24] K. E. Adetunji, I. W. Hofsajer, A. M. Abu-Mahfouz, and L. Cheng, "An optimization planning framework for allocating multiple distributed energy resources and electric vehicle charging stations in distribution networks," *Appl. Energy*, vol. 322, 2022, Art. no. 119513.

Kayode E. Adetunji received the master's degree in electrical and electronics engineering technology from the University of Johannesburg, Johannesburg, South Africa, in 2018, and the Ph.D. degree from in electrical and information engineering from the University of the Witwatersrand, Johannesburg, South Africa, in 2022.

His current research interests include optimization algorithms, decision theory and preference aggregation, and multiobjective optimization, intuitive machine learning models, health and energy, and smart grid systems.

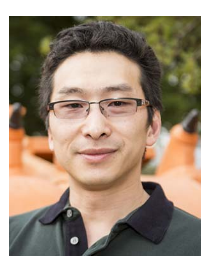
Ivan W. Hofsajer (Member, IEEE) was born in Johannesburg, South Africa. He received the B.Ing. and D.Ing. degrees in electrical engineering from the former Rand Afrikaans University, Johannesburg, South Africa, in 1991 and 1998, respectively.

He worked in the field of electromagnetic interference with the South African Atomic Energy Corporation before joining the Faculty of the Rand Afrikaans University. He is currently an Associate Professor with the School of Electrical and Information Engineering, University of Witwatersrand, Johannesburg, South Africa. His interests include power electronics and electromagnetics.

Adnan M. Abu-Mahfouz (Senior Member, IEEE) received the M.Eng. and Ph.D. degrees in computer engineering from the University of Pretoria, Pretoria, South Africa, in 2005 and 2011, respectively.

He is currently a Chief Researcher and the Centre Manager of the Emerging Digital Technologies for 4IR (EDT4IR) Research Centre, Council for Scientific and Industrial Research, an Extraordinary Professor with the University of Pretoria, and a Professor Extraordinaire with the Tshwane University of Technology, Pretoria, South Africa. His research interests are wireless sensor and actuator networks, low power wide area networks, software-defined wireless sensor networks, cognitive radio, network security, network management, and sensor/actuator node development.

Dr. Abu-Mahfouz is a Section Editor-in-Chief with the *Journal of Sensor and Actuator Networks*, an Associate Editor for IEEE INTERNET OF THINGS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON CYBERNETICS, and IEEE ACCESS, and a member of many IEEE technical communities.



Ling Cheng (Senior Member, IEEE) received the B.Eng. degree in electronics and information (*cum laude*) from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1995, the M.Ing. degree in electrical and electronics (*cum laude*) in 2005, and the D.Ing. degree in electrical and electronics from the University of Johannesburg (UJ), Johannesburg, South Africa, in 2011.

In 2010, he joined the University of the Witwatersrand, Johannesburg, South Africa, where he was promoted to Full Professor in 2019. He is currently the Associate Editor of three journals. He has authored or coauthored more than one hundred research papers in journals and conference proceedings. He has been a Visiting Professor with five universities and the Principal advisor for more than 50 full research postgraduate students. His research interests include telecommunications and artificial intelligence.

Dr. Cheng was the recipient of the Chancellor's medals in 2005 and 2019 and the National Research Foundation ratings in 2014 and 2020. The IEEE ISPLC 2015 Best Student Paper award was made to his Ph.D. student in Austin. He is the Vice-Chair of IEEE South African Information Theory Chapter.