energies

MDPI

Article

Two-Stage Optimization of Virtual Power Plant Operation Considering Substantial Quantity of EVs Participation Using Reinforcement Learning and Gradient-Based Programming

Rong Zhu [1], Jiwen Qi [1], Jiatong Wang [2] and Li Li [1,*]

[1] School of Electrical and Data Engineering, University of Technology Sydney, 15 Broadway, Ultimo 2007, Australia; rong.zhu@student.uts.edu.au (R.Z.); jiwen.qi@student.uts.edu.au (J.Q.)
[2] School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Northfields Avenue, Wollongong 2522, Australia; jiatongw@uow.edu.au
* Correspondence: li.li@uts.edu.au

Abstract

Modern electrical vehicles (EVs) are equipped with sizable batteries that possess significant potential as energy prosumers. EVs are poised to be transformative assets and pivotal contributors to the virtual power plant (VPP), enhancing the performance and profitability of VPPs. The number of household EVs is increasing yearly, and this poses new challenges to the optimization of VPP operations. The computational cost increases exponentially as the number of decision variables rises with the increasing participation of EVs. This paper explores the role of a large number of EVs as prosumers, interacting with a VPP consisting of a photovoltaic system and battery energy storage system. To accommodate the large quantity of EVs in the modeling, this research adopts the decentralized control structure. It optimizes EV operations by regulating their charging and discharging behavior in response to pricing signals from the VPP. A two-stage optimization framework is proposed for VPP-EV operation using a reinforcement algorithm and gradient-based programming. Action masking for reinforcement learning is explored to eliminate invalid actions, reducing ineffective exploration, thereby accelerating the convergence of the algorithm. The proposed approach is capable of handling a substantial number of EVs and addressing the stochastic characteristics of EV charging and discharging behaviors. Simulation results demonstrate that the VPP-EV operation optimization increases the revenue of the VPP and significantly reduces the electricity costs for EV owners. Through the optimization of EV operations, the charging cost of 1000 EVs participating in the V2G services is reduced by 26.38% compared to those that opt out of the scheme, and VPP revenue increases by 27.83% accordingly.

Keywords: reinforcement learning; virtual power plant; electrical vehicle (EV); vehicle-to-grid (V2G); gradient-based programming; two-stage optimization

Academic Editor: Giovanni Lutzemberger

Received: 28 September 2025
Revised: 3 November 2025
Accepted: 7 November 2025
Published: 10 November 2025

Citation: Zhu, R.; Qi, J.; Wang, J.; Li, L. Two-Stage Optimization of Virtual Power Plant Operation Considering Substantial Quantity of EVs Participation Using Reinforcement Learning and Gradient-Based Programming. *Energies* **2025**, *18*, 5898. https://doi.org/10.3390/en18225898

1. Introduction

In the context of the increasing demand for electricity, distributed energy resources (DERs) can effectively enhance the existing centralized energy network, offering greener energy and greater flexibility and reliability. The virtual power plant (VPP) presents a highly efficient and popular solution for seamlessly integrating distributed energy resources into the distribution network [1,2].

EVs provide flexible, mobile energy storage solutions [3–5], and their widespread adoption is increasingly recognized as a crucial element of the energy transition. As sales

of EVs flourish, operations are facing the challenge of an ever-increasing number of EVs integrated into the VPP. The intermittent nature of renewable generation and the uncontrolled charging/discharging of EVs can threaten the security and reliability of power system operations. The optimal operation of VPPs with EVs is a complex problem due to the inherent intermittency of DERs and the stochastic nature of EV charging and discharging behaviors. Modelling a realistic number of EVs interacting with VPPs incurs a significant computational burden for traditional optimization methods, for instance, mixed integer programming [6,7] and dynamic programming [8]. With increasing EV participation, the number of decision variables expands, resulting in an exponential rise in computational costs.

Voluminous historical studies have focused on VPP scheduling and operation [9] while excluding EVs from the modeling entirely or treating them merely as standard loads [9,10] without considering bidirectional interaction between VPPs and EVs. Ref. [11] introduces a new VPP that integrates an EV fleet to mitigate the variability of wind power output. In [12], the energy-saving potential of a VPP comprising PV and energy storage systems (ESS) is examined using historical data. Ref. [3] utilizes EVs as a storage medium to overcome the variability of wind generation. Ref. [4] proposes a two-stage robust optimization model for a VPP that aggregates EV energy storage; it analyzes the distribution characteristics of EVs over time, along with the responsiveness of EV users, to create a model for energy storage capacity. The deployment of VPPs with EVs to enhance renewable energy integration and manage EV charging and discharging efficiently is focused on in [13]. Ref. [14] researches the technical challenges associated with integrating EVs and renewable energy sources into the electric power system.

Ref. [15] has explored pricing strategies for VPP operators that benefits both the operators and EV users. The proposed model involves a Stackelberg game where the VPP acts as a power sales operator, guiding EV users to charge orderly by setting an appropriate power sales price. Ref. [16] optimizes the scheduling and operation of a VPP comprising charging stations for EVs, stationary batteries, and renewable energy sources. The potential of using bidirectional chargers to turn EV battery packs into a VPP that supports the power grid is explored in [17]. Ref. [18] discusses the impact of uncoordinated EV charging behavior on the power grid and the potential for coordinated operation to provide grid flexibility through VPPs. Ref. [15] offers valuable insights for VPPs to efficiently manage EV charging behavior and enhance their operating revenue. The need for detailed charging models of EVs is highlighted in [10], and their impact on VPP operations is explored by considering four different types of EVs. Ref. [19] presents a method to enhance the reliability of a multi-type load power supply, reduce disorder in EV charging, and ensure the low-carbon economic operation of a VPP. Ref. [5] discusses the potential for EV charging to function as a VPP to support distribution system operators.

VPPs are crucial for coordinating EV participation in the power market as aggregators of renewable energies and diverse loads. Ref. [20] proposes a vehicle-to-vehicle (V2V) market mechanism as a supplement to vehicle-to-grid (V2G) operations, aiming to maximize the revenue of each EV owner and create a distributed electricity market. The V2V market allows for trade among EV owners within a local distribution grid, leveraging charging points such as charging stations and auto parks. Ref. [21] introduces an agent-based control system for coordinating the charging of EVs in distribution networks. Its objective is to charge EVs during low electricity price periods while adhering to technical constraints. Ref. [22] considers multiple EV parking lots that are controlled by the VPP and its competitors, vying to attract EVs through competitive offering strategies. In [23], the power optimization of PQ and PV nodes in the power grid was addressed using the SARSA method based on the convergence of power flow calculations. Ref. [24] uses SARSA

methods to address Q-learning's overestimation issue in the automatic generation control strategy for interconnected power grids. Ref. [18] introduces discomfort function for scheduling EV charging that accounts for EV drivers' reluctance to change their initial charging patterns.

V2G [25] technology allows not only for the charging of the EVs but also for the discharge of energy back into the grid. Through V2G services, EV groups can play a dual-role as both energy consumer and provider. To tap into the battery capacity of EVs in a parking lot, Ref. [26] proposes V2G services and a dynamic charging price to optimally control the charging and discharging of EVs in a parking lot. Ref. [27] designs an energy management system to optimize the energy distribution between a workplace's PV system, grid, and battery electric vehicles (BEVs), leading to reductions in charging costs and decreased grid energy consumption.

There is increasing research on V2G services in the context of VPPs, but the research on the integration of V2G services provided by a large quantity of EVs into VPPs is still not sufficiently explored. Although the price signal from the VPP to the EV group is well-understood and has been extensively studied [21,26], the potential for V2G services offered by the EV group to the VPP remains under-explored in existing research. Refs. [20,26,27] did not discuss what impact the arrival of a significant number of EVs at a charging parking lot would have on the aggregator, and how to optimize and mitigate these impacts without compromising the interests of the power provider.

Although various studies [6,18,25,26,28–31] have explored VPP–EV coordination using different optimization approaches, scalability remains a major challenge. As the number of EVs increases, ensuring computational efficiency and solution quality becomes increasingly difficult. Game-theoretic [32,33] methods often perform well in small systems but face exponential growth in complexity with more agents, leading to convergence issues. Similarly, neural network-based reinforcement learning (RL) models can be slow to converge and yield suboptimal results compared to lighter, more interpretable tabular RL algorithms in small-scale settings [34].

Compared to the existing literature, the core motivation for this research is addressing the challenges posed by the substantial quantity of EVs integrated into VPP operations. To overcome the challenges, special considerations are required to ensure optimal solutions are obtained within reasonable time budget. Interactive optimization frameworks in the form of centralized control [15–18] structures demand substantial computational resources and time. Consequently, this study adopts a decentralized [35,36] approach to decouple the modeling of the VPP and EVs. Additionally, RL algorithms exhibit robustness in dynamic and uncertain environments, enabling adaptive decision-making when the EV model is trained iteratively using the Monte Carlo (MC) simulation. For VPP optimization, a gradient-based optimization algorithm combined with a custom loss function is novelly employed to leverage the high computational capacity of GPUs.

This paper presents a two-stage optimization framework for managing the operation of a VPP that integrates EVs, PVs, and a battery energy storage system (BESS). The VPP aims to maximize the revenue from both the EV group and the electricity wholesale market while offering reduced charging costs to the EV group to incentivize their participation in V2G discharge activities. To study the impact of a realistic number of EVs, a MC-SARSA is proposed to train an EV model on the electricity price provided by the VPP.

The contributions of this paper are threefold:

(1). A two-stage optimization framework for the coordinated operation of a VPP that integrates PVs, and a BESS, serving a substantial quantity of EVs that act as prosumers.
(2). A MC-SARSA algorithm is used to train the EV model based on the electricity price provided by the VPP, with training accelerated through action masking for RL.

(3). A gradient-base optimization algorithm with custom loss function to achieve optimal solution for VPP operation, also an exemplary pricing strategy is proposed for the analytic purposes of VPP profitability and EV cost reduction.

2. Problem Formulation

EVs parked at workplace charging facilities typically remain for extended durations, providing sufficient time to charge their batteries to the expected energy level before departure. However, the energy capacity stored in the EV batteries during this period after being fully charged is wasted. The capacity of EV batteries can be leveraged to increase the financial benefits for both the VPP and EV owners. EV owners will obtain a better charging electricity price by participating in the V2G scheme and offer the battery capacity to VPP as long as the battery is charged to the expected energy level at departure. The desired SOC level before departure is modeled as a variable set by the EV owner upon arrival. This design introduces flexibility into the EV model, enabling it to accommodate diverse driving scenarios and individual user preferences. For example, drivers with mileage anxiety may choose to set their desired SOC to the maximum limit. By specifying the target SOC, the model captures realistic behavioral variations among EV owners. This flexibility ensures that charging strategies are better aligned with individual needs and driving habits. At the aggregate level, the VPP can tap from this EV battery capacity at disposal through trading in the wholesale electricity market.

This research does not adopt a centralized control structure, as centralized control relies on physical communication infrastructure, requiring sufficient speed, stability, and bandwidth. Moreover, it places the entire computational and decision-making burden on a single node, creating a critical single point of failure that compromises the system's robustness and scalability. In such a setup, decisions are made centrally without considering the specific preferences of individual participants, and the lack of personalized incentives may discourage user participation. Therefore, in this research, the VPP does not directly manage EV charging and discharging; rather, the EV model functions as a self-interested agent. The final product of the EV optimization is a self-governing agent that requires no additional tuning during runtime. This EV model is capable of directing the charging/discharging behavior of EVs as they enter the car park. Additionally, since the EV model can simulate a large number of EV charging and discharging behaviors, it can also be used for price strategy analysis.

2.1. Two-Stage Optimization Framework

MC simulation is a robust statistical technique employed to model and analyze the impact of risk and uncertainty in prediction and forecasting models [37]. This method relies on the generation of a large number of random samples to investigate the potential outcomes of a process or system. The inputs to the simulation are typically characterized by probability distributions (normal, uniform…), which delineate the range and likelihood of various input values. The process entails conducting numerous iterations, with each simulation using different sets of random inputs to generate a distribution of possible outcomes.

The two-stage optimization framework is a strategic approach utilized in decision-making processes, characterized by the sequential execution of decisions across two distinct stages. This framework proves particularly advantageous for addressing problems where uncertainty or variability plays a significant role.

2.1.1. First Stage: EV Model Training

In the first stage, the EV drive-in time, departure time, state of charge (SOC) at drive-in, and minimum charging hours are mathematically modeled using stochastic distribution. Based on this stochastic distribution, an MC simulation is employed to randomly generate an EV customer's entry into the system, optimizing the EV model until achieving optimal convergence.

Direct optimization methods such as dynamic programming (DP) and mixed integer linear programming (MILP) are not considered in this study due to their inherent limitations. DP requires full knowledge of system dynamics and suffers from the curse of dimensionality [34], making it computationally inefficient for large state or action spaces. This limitation is particularly relevant to our problem, which involves optimizing the behavior of a large number of EVs. While MILP is effective for problems with a moderate number of decision variables, it struggles with non-convex constraints and objectives, making it less suitable for complex, uncertain, or nonlinear scenarios involving interdependent EV participants.

For EV modeling, Monte Carlo SARSA (MC-SARSA) was selected over deep RL methods due to its superior convergence stability, computational efficiency, and interpretability. In mathematical optimization, deep RL often encounters convergence instability, leading to suboptimal outcomes. In contrast, tabular MC-SARSA provides more reliable learning for EV charging scheduling. It also requires far less computational power while achieving comparable performance, making it suitable for edge deployment on EV terminals. Moreover, its tabular structure offers better interpretability of learned policies compared with the black-box nature of DNN.

The motivation for modeling EV behaviors using RL also lies in its efficiency and adaptability. Once trained, an RL model requires relatively low computational resources, making it well-suited for deployment in resource-constrained environments such as the charging terminals in our case. Accordingly, the operational behavior of the EV model is formulated using an MC-SARSA RL algorithm, enhanced with action masking to accelerate convergence.

Upon completion of the EV model training, 1000 EVs are randomly generated, preparing for the second stage, which involves VPP operational optimization. In this stage, the 1000 EVs act as contracted prosumers, and their operational behavior is guided by the optimal EV model trained in this stage.

2.1.2. Second Stage: VPP Operational Optimization

In the second stage, a BESS and a PV system are mathematically modeled, and the base electricity price of wholesale market is determined. PV generation is highly sensitive to the weather conditions of a given area. To develop a model that is robust to the stochastic nature of this renewable resource, traditional direct optimization methods, such as linear programming and dynamic programming, often struggle due to the issue of variable explosion when large volumes of PV data are considered.

In the context of VPP modeling, convergence time is less critical compared to EV modeling, which must accommodate a substantially large population of EVs. Neural networks, owing to their high representational capacity, are particularly well-suited for optimizing problems characterized by stochastic PV generation. Deep neural network (DNN)-powered gradient optimization offers distinct advantages, including inherent differentiability and GPU-accelerated parallelization, handling computationally intensive optimization. Consequently, the formulation of objective functions in VPP modeling can be extended to include general differentiable functions, thereby relaxing the constraints traditionally imposed by convexity [38].

As a result, a DNN is designed for the gradient-based optimization of the VPP model. A custom loss function is designed as the objective function of the VPP modeling. As the DNN converges, the stabilized loss value converges to the optimal results of the VPP model. Figure 1 illustrates the two-stage optimization framework for managing the operation of a VPP that integrates EVs, PVs, and a BESS.

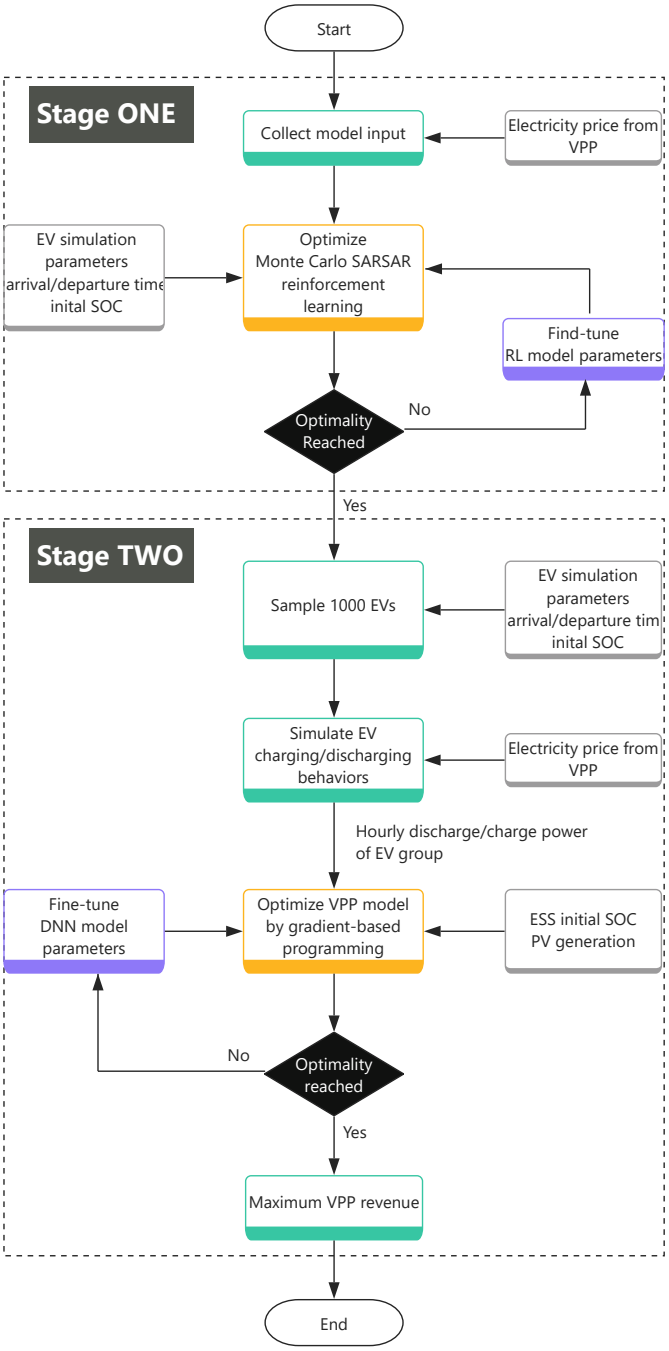

Figure 1. Two-stage optimization framework.

2.2. EV Model

EV enters the workplace car park in a stochastic manner, and leaves at around 5 pm in the afternoon. During the period of stay, the EV owner expects the vehicle to be charged to the required hours set upon entering. During the span of around 9 h, the vehicle is assumedly contracted to the VPP owner to offer the battery capacity to the VPP, and, in return, VPP allows the EV owner to charge the vehicle at a favorable hour and price. The EV will always charge during the hours with the lowest price and discharge during

the hours with the highest price, based on the assumption that the EV acts as a rational agent prioritizing its own benefits. EV owners participating in V2G schemes benefit from reduced prices in exchange for granting the VPP the right to use their EV batteries during their parking periods. As this paper focuses on optimizing the benefits of the VPP, EV battery degradation is not directly modeled.

The objective of the first stage optimization is to minimize the cost of charging the EVs to the minimum charging requirements which are formulated as follows:

$$min(C_{pur}^{EV} - R_{V2G}^{EV}) \tag{1}$$

The total cost of charging the EV group is the sum of the individual charging costs for each EV, minus the V2G services revenue generated by each EV. R_{V2G}^{EV} is the revenue by selling electricity to the VPP through V2G service. C_{pur}^{EV} is the charging cost incurred by purchasing electricity from the VPP.

Upon arrival, an EV_i has an initial amount of energy stored in the battery, the SOC of the EV_i is set to the initial SOC level, $S_{SOC-init}^{EV_i}$. In this research, the initial SOC level of the EV_i follows a normal distribution.

$$S^{EV_i}(t_{arr}^{EV_i}) = S_{SOC-init}^{EV_i} \tag{2}$$

where $t_{arr}^{EV_i}$ is the arrival time of the i^{th} EV, $S_{SOC-init}^{EV_i}$ is the initial drive-in SOC of the i^{th} EV.

The V2G revenue from the VPP to the EV group is the accumulating value of the V2G revenue at each time step t for each EV_i in the group:

$$R_{V2G}^{EV} = \sum_{i=1}^{N} \left(\sum_{t=0}^{T} \lambda_{V2G}^{EV}(t) \times P_{disch}^{EV_i}(t) \times \Delta t - \mu_{V2G}^{EV_i} \right) \tag{3}$$

where N is the total number of EVs in the study, T is the total number of time steps considered in the optimization period, and Δt is the duration of each time step. $P_{disch}^{EV_i}(t)$ is the discharged power from i^{th} EV to the VPP at the time t, $\lambda_{V2G}^{EV}(t)$ is the V2G electricity price offered to EV group by the VPP at the time t. μ_{V2G}^{EV} is the V2G service fee charged by the VPP to the EV owner participating in the scheme.

The cost of purchasing electricity from the VPP follows (4). The total cost of charging the EV group is the sum of the costs of purchasing electricity at each time step t for each EV_i in the group:

$$C_{pur}^{EV} = \sum_{i=1}^{N} \sum_{t=0}^{T} \lambda_{pur}^{EV}(t) \times P_{chg}^{EV_i}(t) \times \Delta t \tag{4}$$

where $P_{chg}^{EV_i}(t)$ is the charging power of the i^{th} EV from the VPP at the time t, $\lambda_{pur}^{EV}(t)$ is the charging electricity price offered to EV group by the VPP at the time t.

Before arrival and after departure, the SOC of EV_i is set to zero:

$$S^{EV_i}(t) = 0 \qquad t < t_{arr}^{EV_i} \quad or \quad t > t_{dep}^{EV_i} \tag{5}$$

where $t_{dep}^{EV_i}$ is the departure time of the i^{th} EV.

The SOC of the EV_i at time $t+1$ during the scheduling is calculated as follows:

$$S^{EV_i}(t+1) = S^{EV_i}(t) + \left(\eta_{chg}^{EV_i} \times \frac{P_{chg}^{EV_i}(t)}{C^{EV_i}} - \frac{1}{\eta_{disch}^{EV_i}} \times \frac{P_{disch}^{EV_i}(t)}{C^{EV_i}} \right) \times \Delta t \quad t_{arr}^{EV_i} \leqslant t \leqslant t_{dep}^{EV_i} - 1 \tag{6}$$

where $S^{EV_i}(t)$ is the SOC of the i^{th} EV at time t, $\eta_{chg}^{EV_i}$ is the charging efficiency of the i^{th} EV, $\eta_{disch}^{EV_i}$ is the discharging efficiency of the i^{th} EV, and C^{EV_i} is the battery capacity of the i^{th} EV.

Outside the parking period (from arrival to departure), the charging and discharging power of EV_i should be zero:

$$P_{disch}^{EV_i}(t) = P_{chg}^{EV_i}(t) = 0 \quad t < t_{arr}^{EV_i} \quad or \quad t > t_{dep}^{EV_i} \tag{7}$$

During the MC simulation, the departure time for EV_i should always be later than the arrival time at all times. Both the departure time and the arrival time should be a positive integer following a normal distribution, with a different mean value.

$$0 \leqslant t_{arr}^{EV_i} < t_{dep}^{EV_i} \tag{8}$$

The charging and discharging power of the EV_i at any time during the scheduling must not exceed the maximum and minimum charging/discharging power limits, as described by the following equation:

$$P_{disch-min}^{EV_i} \leqslant P_{disch}^{EV_i}(t) \leqslant P_{disch-max}^{EV_i} \quad P_{disch}^{EV_i}(t) \neq 0, \ t_{arr}^{EV_i} \leqslant t \leqslant t_{dep}^{EV_i} \tag{9}$$

$$P_{chg-min}^{EV_i} \leqslant P_{chg}^{EV_i}(t) \leqslant P_{chg-max}^{EV_i} \quad P_{chg}^{EV_i}(t) \neq 0, \ t_{arr}^{EV_i} \leqslant t \leqslant t_{dep}^{EV_i} \tag{10}$$

where $P_{disch-min}^{EV_i}$ and $P_{disch-max}^{EV_i}$ are the minimum and maximum discharging power of the i^{th} EV respectively; Likewise, $P_{chg-min}^{EV_i}$ and $P_{chg-max}^{EV_i}$ are the minimum and maximum charging power of the i^{th} EV.

To protect the battery life of the EV, the SOC of the EV_i is constrained by the following conditions:

$$S_{min}^{EV_i} \leqslant S^{EV_i}(t) \leqslant S_{max}^{EV_i} \quad t_{arr}^{EV_i} \leqslant t \leqslant t_{dep}^{EV_i} \tag{11}$$

where $S_{min}^{EV_i}$ and $S_{max}^{EV_i}$ are the minimum and maximum SOC of the i^{th} EV, respectively.

The power exchange between the EV group and the VPP adheres to the following balance equation:

$$P_{EV}^{VPP}(t) = \sum_{i=1}^{N} P_{chg}^{EV_i}(t) - \sum_{i=1}^{N} P_{disch}^{EV_i}(t) \qquad 0 \leqslant t \leqslant T \tag{12}$$

To prevent simultaneous charging and discharging within any operational time horizon, the power exchange between the i^{th} EV and the VPP is regulated by the following condition:

$$P_{chg}^{EV_i}(t) \times P_{disch}^{EV_i}(t) = 0 \qquad 0 \leqslant t \leqslant T \tag{13}$$

Upon departure, the SOC of the i^{th} EV should be charged to the required SOC level.

$$S^{EV_i}(t_{dep}^{EV_i}) = S^{EV_i}(t_{arr}^{EV_i}) + S_{reqd}^{EV_i} \tag{14}$$

where $S_{reqd}^{EV_i}$ is the required energy that need to be charged to the i^{th} EV at departure in terms of SOC.

2.3. VPP Model

The VPP in this study owns a PV system, a BESS. Additionally, a group of EVs that randomly enter and leave the workplace parking lot during working hours are the contracted customers of the VPP. The purpose of the BESS is to smooth the intermittent PV power generation and store the surplus energy. PV energy generation involves converting sunlight into electricity using photovoltaic cells, which convert the light into direct current (DC), then transformed into alternating current (AC) for use in homes and businesses or for feeding into the electrical grid. The VPP generates revenue by selling electricity on the wholesale market and providing energy to meet the charging demands of the EV group as shown in Figure 2. The scale of the VPP business is modeled to accommodate N number of EVs entering daily, where N represents any realistic figure that makes business sense to the VPP owner. Once the RL-based EV model is trained, it enables efficient scalability to 100–3000 EVs with minimal computational overhead, due to its decentralized decision-making policy. Although the experimental results in this paper focus on 1000 EVs, the model can scale to 5000 or more without notable computation time increase. The trained EV model is capable of analyzing business strategies for a VPP at a fine-grained scale, factoring in the number of EVs included in the modeling.

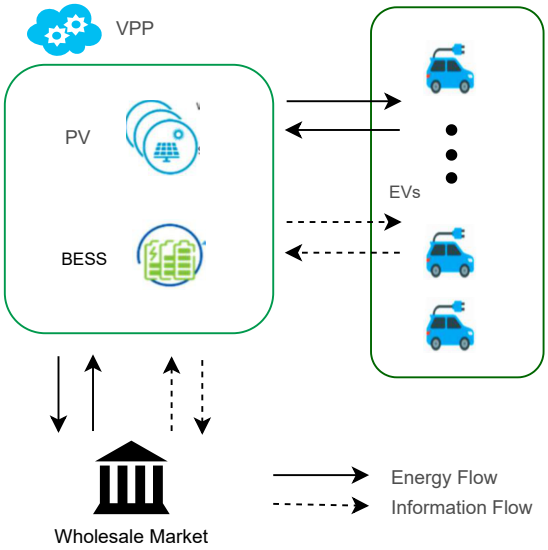

Figure 2. Relationship diagram of VPP and EV group.

The objective of this VPP study, which is the second stage of the optimization framework, is to maximize the revenue of the VPP which is calculated as follows:

$$max(R_{feed-in}^{VPP} - C_{pur}^{VPP} + C_{pur}^{EV} - R_{V2G}^{EV}) \tag{15}$$

where $R_{feed-in}^{VPP}$ is the revenue gained by selling electricity to the wholesale market. C_{pur}^{VPP} is the cost incurred by purchasing electricity from the wholesale market during energy deficiency.

The revenue of the VPP is the surplus from the revenue gained by selling electricity to the wholesale market and the cost incurred by purchasing electricity from the wholesale market during energy deficiency. Another stream of revenue comes from the energy transactions between EVs and the VPP. This includes the energy purchased by EVs for charging during the scheduled time period, offset by the energy sold back to the VPP through V2G services.

The VPP feed-in revenue is calculated by accumulating all the VPP feed-in revenue during each time step t:

$$R_{feed-in}^{VPP} = \sum_{t=0}^{T} \lambda_{feed-in}^{VPP}(t) \times P_{feed-in}^{VPP}(t) \times \Delta t \tag{16}$$

where $\lambda_{feed-in}^{VPP}(t)$ is the feed-in price at the time t, $P_{feed-in}^{VPP}(t)$ is the power fed into the grid by the VPP at time step t.

The cost of purchasing electricity from the wholesale market is calculated as follows:

$$C_{pur}^{VPP} = \sum_{t=0}^{T} \lambda_{pur}^{VPP}(t) \times P_{grid}^{VPP}(t) \times \Delta t \tag{17}$$

C_{pur}^{VPP} is the cost incurred from purchasing electricity from the wholesale market. $\lambda_{pur}^{VPP}(t)$ is the wholesale market electricity price at the time t. $P_{grid}^{VPP}(t)$ is the required power from the grid for VPP at time step t.

The feed-in electricity price sold to the wholesale market at each time step t is marked down from the purchase electricity price from the wholesale market by a discount factor α, where $0 < \alpha < 1$.

$$\lambda_{feed-in}^{VPP}(t) = \alpha \times \lambda_{pur}^{VPP}(t) \qquad 0 \leqslant t \leqslant T \tag{18}$$

2.3.1. PV Generation

For the purpose of mathematical modeling, the PV panels are aggregated into a single unit, calculated using the following formula [26]:

$$P_{PV}^{VPP}(t) = I(t) \times A \times \eta^{PV} \tag{19}$$

where P_{PV}^{VPP} is the PV power generation of the VPP, $I(t)$ is the global horizontal irradiance at time t, A is the area of the PV panels, and η^{PV} is the efficiency of the PV panels.

2.3.2. Battery Storage

During the period of VPP scheduling, the discharging and charging of the ESS is constrained by the following condition:

$$S^{ESS}(T+1) = S^{ESS}(t) + \left(\eta_{chg}^{ESS} \times \frac{P_{chg}^{ESS}(t)}{C^{ESS}} - \frac{1}{\eta_{disch}^{ESS}} \times \frac{P_{disch}^{ESS}(t)}{C^{ESS}} \right) \times \Delta t \qquad 0 \leqslant t \leqslant T \tag{20}$$

where $S^{ESS}(t)$ is the SOC of the ESS at time t, $P_{chg}^{ESS}(t)$ is charging power of the ESS at time t, $P_{disch}^{ESS}(t)$ is the discharging power of the ESS at time t, η_{chg}^{ESS} is the charging efficiency of the ESS, η_{disch}^{ESS} is the discharging efficiency of the ESS, and C^{ESS} is the battery capacity of the ESS.

The charging and discharging power of the ESS at any time during the scheduling can not exceed the maximum and minimum power limits, as described by the following constraints:

$$\begin{aligned} P_{disch-min}^{ESS} \leqslant P_{disch}^{ESS}(t) \leqslant P_{disch-max}^{ESS} \\ P_{disch}^{ESS}(t) \neq 0, \ \ 0 \leqslant t \leqslant T \end{aligned} \tag{21}$$

$$\begin{aligned} P_{chg-min}^{ESS} \leqslant P_{chg}^{ESS}(t) \leqslant P_{chg-max}^{ESS} \\ P_{chg}^{ESS}(t) \neq 0, \ \ 0 \leqslant t \leqslant T \end{aligned} \tag{22}$$

where $P_{chg-min}^{ESS}$ and $P_{chg-max}^{ESS}$ are the minimum and maximum charging power of the ESS, respectively; likewise, $P_{disch-min}^{ESS}$ and $P_{disch-max}^{ESS}$ are the minimum and maximum discharging power of the ESS, respectively.

The ESS cannot be charged and discharged simultaneously at any time during the scheduling period, and this is ensured by the following condition:

$$P_{chg}^{ESS}(t) \times P_{disch}^{ESS}(t) = 0 \qquad 0 \leqslant t \leqslant T \tag{23}$$

At the end of the scheduling, the SOC of the ESS should be the same as the initial SOC S_{init}^{ESS}.

$$S^{ESS}(T+1) = S^{ESS}(0) = S_{init}^{ESS} \tag{24}$$

For the safe operations of the ESS, the SOC of the ESS is constrained by (25) at any time t:

$$S_{min}^{ESS} \leqslant S^{ESS}(t) \leqslant S_{max}^{ESS} \qquad 0 \leqslant t \leqslant T \tag{25}$$

where S_{min}^{ESS} and S_{max}^{ESS} are the minimum and maximum SOC, respectively.

2.3.3. Power Equation

During the scheduling period, the VPP system is constrained by the following balance equation:

$$\begin{aligned} P_{PV}^{VPP}(t) + P_{disch}^{ESS}(t) - P_{chg}^{ESS}(t) - P_{EV}^{VPP}(t) \\ = P_{feed-in}^{VPP}(t) - P_{grid}^{VPP}(t) \\ 0 \leqslant t \leqslant T \end{aligned} \tag{26}$$

where $P_{EV}^{VPP}(t)$ is defined in (12). $P_{EV}^{VPP}(t) > 0$ when EVs are charging, $P_{EV}^{VPP}(t) < 0$ when EVs are discharging.

Power exchange between the VPP and the grid during the operational time horizon is governed by the following condition:

$$P_{grid}^{VPP}(t) \times P_{feed-in}^{VPP}(t) = 0 \qquad 0 \leqslant t \leqslant T \tag{27}$$

2.3.4. Modeling and Simulation

The VPP model is a three-layer neural network with a neuron configuration of 24-50-100-50-24, which converges stably to the optimality within 3 min and 47 s. The input vector is divided into two segments: the first segment consists of the hourly electricity prices, and the second segment includes 24 inputs representing the solar power generation for each hour. The model outputs 24 values, which correspond to the hourly discharge or charging schedule operations of the BESS.

The output decision vector of the neural network is constrained to the range $[-1, 1]$ using the hyperbolic tangent activation function, where negative values indicate VPP battery discharging, positive values represent charging, and a value of zero signifies an idle state with no charging or discharging activity. Positive raw values from the neural network

are scaled to the range $[P^{ESS}_{chg-min}, P^{ESS}_{chg-max}]$, while negative raw values are similarly scaled to $[P^{ESS}_{disch-min}, P^{ESS}_{disch-max}]$. As a result, constrait (23) is upheld. Furthermore, $S^{ESS}(t)$ is cliped within the bounds $[S^{ESS}_{min}, S^{ESS}_{max}]$ to ensure compliance with the predefined operational limits, if the output decision vector of the neural network results in a violation of the lower or upper bounds of $S^{ESS}(t)$ during training.

The MC sampling method is employed to simulate 1000 EVs driving at various times with diverse charge demands, computing the hourly power exchange between EVs and the VPP, leveraging the optimal EV model previously trained. On average, the MC-SARSA model with action masking converges within approximately 4 min 18 s. This convergence time reflects the model's ability to efficiently explore and exploit the action space while maintaining stability in the learning process. Note that while the choice of 1000 vehicles reflects the scale of the VPP business and is not constrained by computational capabilities.

2.4. Pricing Strategy

The electricity price offered by VPP is vital to both parties in terms of charging costs and revenue. The pricing strategy itself is a research direction that attracts many researchers. To prove the concept, this paper only includes an exemplary price strategy, formulated as (28), to demonstrate the significant impact of pricing on VPP-EV operations. The choice of pricing strategy can result in varying outcomes, potentially resulting in a win–win, win–lose, or lose–lose situation for both stakeholders—the VPP and the EVs. The exemplary pricing strategy is that the VPP proprietor gets the wholesale market electricity price for 24 h covering the whole operational period by prediction or other means. For each individual time step t, we know the price $p(t)$. The price will be shuffled in terms of the time stamp. The sum of the 24 prices will be unchanged but for individual time step $\tilde{t}$, the new price $\tilde{p}(\tilde{t})$ is more likely from a different time-step of $p(t)$. Once a price $p(t)$ is selected, it won't be chosen anymore in the process. Consider 24 prices as 24 balls in a bag, each ball has a price $p(t)$ attached to it. According to our pricing strategy, for each time $\tilde{t}$ ranging from 1 to 24, the pricing strategy picks a ball from the bag, and the price $p(t)$ attached will be selected for the current time step $\tilde{t}$. The ball is subsequently removed from the bag. At the end of the process, no ball should be left in the bag.

$$\tilde{p}(t) = \mathcal{P}_{strtg}(\lambda_{whsle}, t) \tag{28}$$

$$\lambda^{EV}_{base}(t) = \mu \times \tilde{p}(t) \tag{29}$$

$$\lambda^{EV}_{pur}(t) = \tau_{mu} \times \lambda^{EV}_{base}(t) \tag{30}$$

$$\lambda^{EV}_{V2G}(t) = \tau_{disc} \times \lambda^{EV}_{base}(t) \tag{31}$$

where λ_{whsle} is the wholesale market price, and $\mathcal{P}_{strtg}$ is the price strategy that accepts the wholesale price and time step t and outputs a price for the time t. $\tilde{p}(t)$ is the base electricity price derived from the pricing strategy at time t. μ is the price multiplier which plays a significant role in revenue allocation for a VPP; $\lambda^{EV}_{base}(t)$ is the base EV electricity price, τ_{disc} is the discount of $\lambda^{EV}_{base}(t)$ for V2G, a positive decimal which is less than one, and τ_{mu} is the markup of $\lambda^{EV}_{base}(t)$ for charging, a positive decimal which is larger than one. Figure 3 shows the process of deriving the EV charging/discharging price from the wholesale market price.

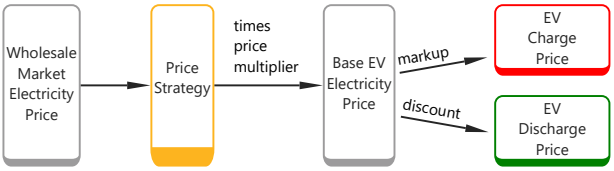

Figure 3. Process of deriving EV charging/discharging price.

3. Solution Methods

3.1. Gradient-Based Programming

Gradient-based programming is used in optimization where gradients of a function are utilized to find the optimal solution. Gradient-based programming fundamentally revolves around optimizing an objective function [39]. This function could represent various metrics, such as error or cost, that the program seeks to optimize. The gradient of the objective function is calculated for each parameter, pointing toward the direction of maximum growth of the function. By taking the negative of the gradient, the direction that most steeply decreases the function is obtained. Parameters are iteratively updated using the computed gradients. The updates are usually performed using a method like gradient descent, stochastic gradient descent, mini-batch gradient descent. These methods adjust the parameters in the direction of the negative gradient. The update step for parameters $\mathbf{x}$ at iteration t is as follows:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \eta \nabla f(\mathbf{x}_t) \tag{32}$$

where η is the learning rate and $\nabla f(\mathbf{x}_t)$ is the gradient of the loss function, $f(\cdot)$, at the current point. The process is repeated until the parameters converge to a minimum of the function. During backpropagation in neural network training, the gradient is calculated by recursively applying the chain rule to propagate derivatives backward through the layers of the network.

3.2. Custom Loss Function

A custom loss function is a user-defined function designed to evaluate the performance of a machine learning model by minimizing the expected loss from the input data. Unlike standard loss functions, such as Mean Squared Error (MSE) or Cross-Entropy Loss, a custom loss function is tailored to address the specific needs of a particular problem. This is especially useful when standard loss functions fail to adequately capture the application-specific performance metrics that are crucial for the given application [40]. For deep neural networks (DNNs), it is essential that the custom loss function be differentiable to facilitate backpropagation.

In this study, a custom loss function is crafted to optimize the VPP model. The nonlinear constraints are addressed through the design of this custom loss function, tailored to penalize constraint violations during the training process. Any violation of the constraint results in a significantly large penalty, effectively discouraging the neural network from exploring infeasible regions of the solution space. The loss function integrates the feed-in revenue, purchase cost, and revenue generated from the EV group. The primary goal of the VPP model is to maximize its revenue. The custom loss function is formulated as in (33) and (34), where L^{VPP} is the loss function of the VPP model, α^{VPP} is the penalty coefficient, and ρ^{VPP} is the penalty exponent for the VPP modeling. As training progresses, the DNNs are iteratively updated to minimize the loss function, thereby maximizing the revenue of the VPP. Minimizing the loss function also guides the model toward solutions

that respect the system constraints. In this way, constraint compliance is achieved as a result of loss minimization.

$$L^{VPP} = -R^{VPP}_{feed-in} + C^{VPP}_{pur} - C^{EV}_{pur} + R^{EV}_{V2G} + \psi^{VPP}_{pen} \tag{33}$$

$$\psi^{VPP}_{pen} = \left(|S^{ESS}(T+1) - S^{ESS}_{init}| \times \alpha^{VPP}\right)^{\rho^{VPP}} \tag{34}$$

3.3. Monte Carlo SARSA

This research trains the EV model using MC-SARSA RL algorithm. MC-SARSA is a type of RL algorithm that combines the Monte Carlo method with the SARSA algorithm. It is particularly useful in environments where episodes are clearly defined and end in a finite number of steps, and it can be applied when the model of the environment is unknown or too complex to model accurately. Unlike standard SARSA which updates the value function after every step within an episode, MC-SARSA waits until the end of the episode to perform the update [41], using the entire sequence of states, actions, and rewards to make its calculations. This can be advantageous in environments where immediate rewards are sparse or delayed, as it allows the agent to learn from the full sequence of interactions with the environment.

For an episode of length T, starting at time step t, the return G_t is defined as the total discounted reward from time step $t+1$ to the end of the episode:

$$G_t = \sum_{k=t+1}^{T} \gamma^{k-(t+1)} R_k \tag{35}$$

where γ is the discount factor, and R_k is the reward received at time step k. MC-SARSA differs from regular SARSA in that SARSA updates after each step using bootstrapping, while MC-SARSA uses the full return G_t. After an episode, for each state-action pair encountered in the episode, the action-value function is updated using the following formula:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha\big(G_t - Q(S_t, A_t)\big) \tag{36}$$

where α is the learning rate, S_t is the state at time step t, A_t is the action taken at that time, and $Q(S_t, A_t)$ denotes the action-value function for the state-action pair (S_t, A_t).

The reward function of the EV model is defined as in (37) and (38), where $R^{EV_i}_{rwd}$ is the reward function of the i^{th} EV, $\mu^{EV_i}_{V2G}$ represents the V2G participation fee of the i^{th} EV, $\psi^{EV_i}_{pen}$ is the penalty term for the i^{th} EV.

The penalty term $\psi^{EV_i}_{pen}$ is added to the reward function to penalize the EV model for not meeting the EV's minimum charging requirements upon departure. The penalty term is calculated based on the value difference between the SOC of the EV at departure and the required SOC, where α^{EV} is the penalty coefficient, and ρ^{EV} is the penalty exponent for EV modeling.

$$R^{EV_i}_{rwd} = -\sum_{t=0}^{T}\left(\lambda^{EV}_{pur}(t) \times P^{EV_i}_{chg}(t) \times \Delta t - \lambda^{EV}_{V2G}(t) \times P^{EV_i}_{disch}(t) \times \Delta t\right) - \mu^{EV_i}_{V2G} - \psi^{EV_i}_{pen} \tag{37}$$

$$\psi^{EV_i}_{pen} = \left(|S^{EV_i}(t^{EV_i}_{dep}) - S^{EV_i}_{reqd}| \times \alpha^{EV}\right)^{\rho^{EV}} \tag{38}$$

3.4. Action Masking

Action masking in RL is a technique that restricts an agent's available actions to only valid ones at a given state, preventing the selection of actions that are nonsensical

or invalid. This is particularly important in environments with rules or constraints that limit the agent's choices at certain times [42]. By using action masking, the agent can focus on learning from valid actions, improving its efficiency and avoiding penalties associated with exploring invalid ones. This technique helps speed up the learning process, reduces computational waste, and is particularly useful in environments with large and dynamic action spaces, such as complex games, robotics, or tasks with specific constraints.

The action mask is defined as $m(s) \in \{0,1\}^{|\mathcal{A}|}$ for state s, where $\mathcal{A}$ is the set of all possible actions. The action mask is a binary vector of size $|\mathcal{A}|$, where $|\mathcal{A}|$ is the number of actions. $m_a(s) = 1$ indicates that action a is valid in state s, while $m_a(s) = 0$ marks it as invalid. To ensure the algorithm only selects valid actions, the Q-values of invalid actions are replaced with a large negative number. The modified Q-value function can be expressed as follows:

$$Q_{\text{masked}}(s,a) = \begin{cases} Q(s,a), & \text{if } m_a(s) = 1 \\ -10^9, & \text{if } m_a(s) = 0 \end{cases} \tag{39}$$

where $Q_{\text{masked}}(s,a)$ is the modified Q-value function, $Q(s,a)$ is the original Q-value function. This results in a masked action selection rule:

$$a^* = \arg\max_{a \in \mathcal{A}} Q_{\text{masked}}(s,a) \tag{40}$$

where a^* is the selected action.

Leveraging domain knowledge, the action mask is designed to eliminate invalid charging or discharging actions, ensuring that the EV only considers sensible options at the given state. This approach enhances the efficiency of the EV's learning process and prevents the exploration of actions that leads to violations of operational conditions.

4. Case Studies

4.1. Parameters and Case Settings

The battery capacity of an EV in this study is assumed to be 80 kWh. The maximum charging and discharging power of an EV is set as 7.4 kW. For safety considerations, overcharging a battery can pose risks such as overheating, potential fire or explosion hazards. On the other hand, surpassing the maximum SOC during EV battery charging typically yields diminishing returns in energy efficiency. In light of this, in this study, the EV's minimum SOC is constrained to 0.2, while its maximum is set at 0.8 [43].

Both the charging and discharging efficiencies are configured to be 0.95. For revenue generation, the VPP sets the base electricity price for EVs at 2.0 times the rate sourced from the Australian Energy Market Operator (AEMO) [44]. Also, the charge price is set at a 1.1 markup of the base electricity price offered by the VPP, while the discharge price is discounted by 0.9 of the base electricity price. According to [32], without loss of generality, in this modeling, the individual EV arrival times follow a normal distribution with a mean of 9 and a standard deviation of 1. Departure times are also normally distributed with a mean of 17 and a standard deviation of 1. The initial SOC for each individual is modeled with a mean of 0.34 and a standard deviation of 0.1. The minimum charge time for each individual is drawn from a discrete uniform distribution ranging from 2 to 5 (Table 1) [26].

To use the SARSA algorithm, which is a tabular reinforcement learning method, discretization is required. Electricity prices and SOC levels are both rounded to two decimal places for efficiency in memory storage. The customer-facing digits with two decimal places is an acceptable compromise in the context of business sense. The penalty coefficient, α^{EV} is set to 100, while ρ^{EV} is assigned a value of 3.

Table 1. Stochastic parameters for EV simulation.

Arrival Time (Hour)	N(9, 1)
Departure Time (Hour)	N(17, 1)
Initial SOC	N(0.34, 0.1)
Minimum Charge Time (Hour)	U(2, 5)

The VPP contains a BESS with a capacity of 8 MWh, the charging/discharging efficiency of the BESS is 0.95, and its maximum charging/discharging power is 5 MWh. The penalty coefficient α^{VPP} is set as 100, while ρ^{VPP} is set to 3. The price of electricity from the VPP to the grid is discounted by 10% compared to the grid price. Figure 4 is an example of VPP charge/discharge electricity price from/to the wholesale market [44]. The hourly PV generation [45] for a typical operational day is shown in Table 2.

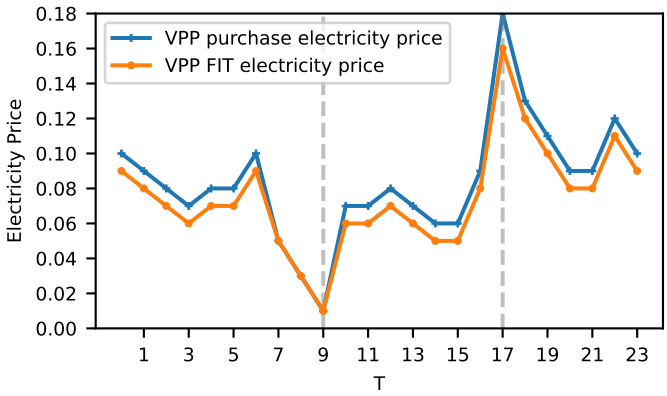

Figure 4. VPP electricity price from the wholesale market.

Table 2. Hourly PV generation.

t	kW	t	kW	t	kW
1	0	9	3340	17	4495
2	0	10	4955	18	2420
3	0	11	6655	19	615
4	0	12	7830	20	15
5	0	13	8315	21	0
6	255	14	7560	22	0
7	1085	15	4455	23	0
8	1760	16	3815	24	0

The EV operational model is trained using the MC-SARSA RL algorithm to minimize the charging cost of EVs while ensuring the EVs are charged to the minimum charging requirement before departure. The accumulated rewards of the EV RL model climb up as the training progresses, indicating the model is learning to optimally direct the charging and discharging operation to minimize the charging cost of EVs. In one scenario, the MC-SARSA EV model stabilizes and converges to a total cost of AUD 5385 for 1000 EVs after 500,000 training episodes (Figure 5). The reached minimum cost is the stochastic mean value of the total cost of charging 1000 randomly entered EVs. The Brownian reward fluctuation after convergence is caused by the stochastic sampling of EVs with varying charging needs and parking durations during the training process.

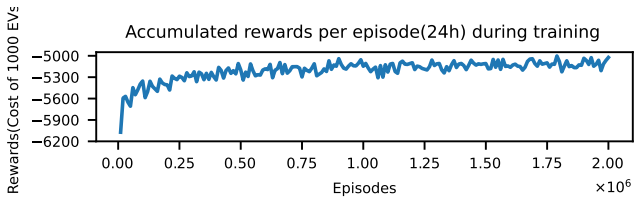

Figure 5. Accumulated rewards per episode (over 24 h) during the training of the EV scheduling model.

Take one EV sample of the Monte Carlo simulation, and examine the charging and discharging behavior of the EV during the parking period. The EV charging and discharging operation is optimized by the EV model trained using MC-SARSA algorithm. Figure 6 shows time steps 7, 8, and 9 presents the cheapest electricity price, and the EV is charging during these hours. Time steps 6 and 12 presents the highest electricity price, and the EV is discharging during these hours.

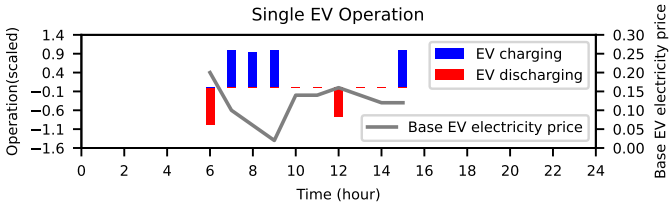

Figure 6. Operations of a single EV over time T.

The EV enters the carpark at hour 6, with an initial SOC of 0.24, and it is charged to the minimum SOC requirement (0.42) before departure at hour 16. Figure 7 presents the changes in the SOC of a randomly selected EV from the 1000 EVs in the simulation over the time span T, following the optimized charging and discharging operational behaviors. From Figure 7, it is observed that there is a SOC declination at hours 6 and 12, and a SOC increment at hours 7, 8, 9 and 15. Hours 6 and 12 are the time with the highest electricity price, and thus the EV is discharging during these hours. As hours 7, 8, 9, and 15 are the times with the lowest electricity prices, the EV is charging during these hours.

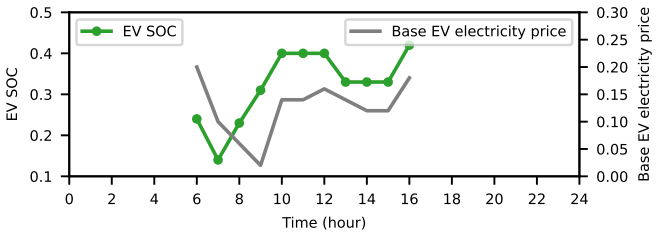

Figure 7. Changes in the SOC of a single EV over time T.

4.1.1. Charging Strategies

Based on the preferred charging strategy, EV owners are classified into three types: T1: those who charge immediately until their energy needs are met, despite being less sensitive to higher charging prices; T2: those who optimize their charging without participating in the V2G scheme and pay a moderate charging price; and T3: those who optimize their charging while participating in the V2G scheme, thereby benefiting from the lowest charging prices.

To facilitate comparison, it is advisable to assess the total cost for all 1000 EVs assuming they follow charging strategies T1, T2, or T3. Additionally, understanding the revenue generated by the VPP and V2G under these different strategies will enable a clear comparison of the impacts on both the VPP and the EVs. In the case study, three categories

of EV groups (C1, C2, and C3), each adopting a different charging strategy, are analyzed. All three categories use the same EV electricity price, P0, to facilitate comparison, with the details illustrated in Table 3.

Table 3. Categories of EV charging groups (C1–C3).

Category	Charging Strategy	EV Electricity Price
C1	T1	P0
C2	T2	P0
C3	T3 (V2G)	P0

Executing the price strategy depicted in Figure 3 will yield varying results for the VPP and EVs. Among the prices generated by this strategy, three specific prices (P0, P1, and P2), as seen in Figure 8, have been selected for the purpose of analysis.

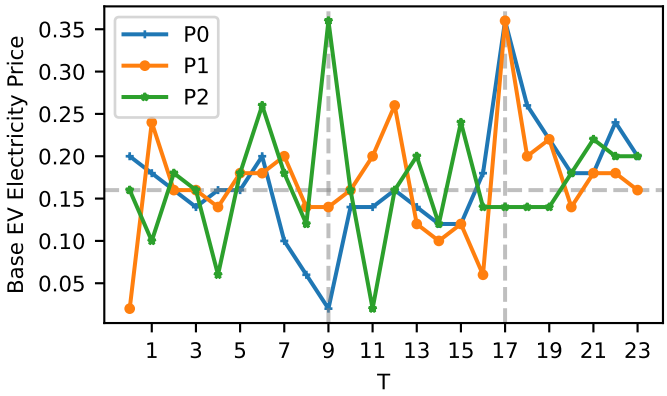

Figure 8. Three base EV electricity price for case studies.

4.1.2. Battery Degradation

Battery charging and discharging cycles contribute to battery degradation over time. Frequent deep discharges and high-rate charging can accelerate capacity loss. To account for the battery degradation of BESS in VPP, a case study is designed to include the battery degradation cost in the VPP optimization model. Including the battery degradation for BESS in VPP will economically effect the outcome of the optimization. The battery degradation cost is modeled following the approach in [26] as follows:

$$B = \frac{R}{L \times \eta^{ESS}} \tag{41}$$

$$C_{deg}^{ESS} = \sum_{t=0}^{T} P_{disch}^{ESS}(t) \times B \tag{42}$$

where B is the battery degradation cost per kWh discharged, R is the battery replacement cost, L is the total lifetime energy throughput. η^{ESS} is the squared root of the roundtrip efficiency of the BESS, and C_{deg}^{ESS} is the total battery degradation cost of the BESS over the scheduling period T. In this study, the battery degradation coefficient B is set to AUD 0.038 per kWh [43], this value reflects the amortized cost of capacity loss over the expected lifetime of the battery.

The EV battery modeling enforces upper and lower bounds on the SOC to ensure safe operation within limits. However, additional battery degradation effects are not explicitly considered in this study. The economic benefits obtained from reduced EV charging costs are inherently designed to offset the potential degradation costs associated with V2G

discharging. Since this paper primarily focuses on the economic operation of the VPP, EV battery degradation is not explicitly modeled in this context.

4.1.3. Computational Environment

The simulation and optimization were conducted on a laptop equipped with a 12th Gen Intel Core i7-1265U CPU (1.80 GHz), 16 GB of RAM, and an Nvidia GeForce MX550 GPU. The optimization of the VPP model was implemented using the PyTorch 2.1 Python package. The MC-SARSAR algorithm was self-implemented in the C# programming language, running on the .NET 7.0 platform.

4.2. Results and Analysis

The case study evaluates various EV charging categories and their impact on both EV charging costs and VPP revenue. The analysis also examines pricing models and their implications for both the VPP and the EVs. The results are presented in the following sections.

4.2.1. EV Charging Strategy Analysis

Table 4 presents the total cost of charging 1000 EVs across three distinct charging categories: C1, C2, and C3. The base electricity rate offered by the VPP for this analysis is P0.

Table 4. Comparison of charging costs for 1000 EVs (C1, C2 and C3).

Categories	EV Charging Cost	VPP Revenue
C1	AUD 6517.55	**AUD 8125.62**
C2	AUD 2719.82	AUD 5872.81
C3	**AUD 2565.79**	AUD 5889.29

Charging Category C1

In the case of C1, EV will start charging upon arrival until the required energy is met, the base electricity offered by VPP is P0. Among the 1000 sample EVs, the minimum charging cost for an EV is AUD 2.59, the maximum charging cost for an EV is AUD 14.95, the average charging cost for an EV is AUD 6.52, and the total charging cost for 1000 EVs is AUD 6517.55. EV customers in C1 pay the most expensive charging price as a result of disorderly charging behaviors.

Charging Category C2

In the case of C2, where the EV will charge at the lowest price until the charging requirement is met during the scheduling period, EV in C2 does not participate in the V2G scheme. For C2, among the 1000 sample EVs, the minimum charging cost for an EV is AUD 0.27, the maximum charging cost for an EV is AUD 5.42, the average charging cost for an EV is AUD 2.72, and the total charging cost for 1000 EVs is AUD 2719.82. Compared with the charging cost of EVs in the case of C1, the total charging cost for all the EVs in case C2 is reduced by 58.27%.

Charging Category C3

EV customers in C3 charge at the lowest price until the charging requirement is met during the scheduling period; furthermore, they also participate in the V2G scheme. For C3, among the 1000 sample EVs, the minimum charging cost for an EV is AUD −1.10, where the negative value indicates that the customer gains revenue while charging the EV to the required SOC, the maximum charging cost for an EV is AUD 5.55, the average charging cost for an EV is AUD 2.57, and the total charging cost for 1000 EVs is AUD 2565.79. Compared

to the charging cost of EVs in the case of C2, the total charging cost for all the EVs in case C3 is reduced by AUD 154.03, which is 5.66% cost reduction. It should be noted that out of the AUD 2565.79 charging cost, the V2G service fee in (3) is AUD 150, that is, 15 cents per EV customer.

The C3 case with V2G emerges as the most advantageous charging approach among the three analyzed categories (C1, C2, and C3). The implementation of V2G not only facilitates cost reductions for EV owners but also enhances the overall efficiency of energy usage guided by the VPP. The cumulative effect of these financial benefits, coupled with the potential for improved energy management and reduced strain on the grid, positions V2G as the superior strategy.

The simulation of 1000 EVs driving at various times with diverse charge demands is analyzed by computing the hourly power exchange between EVs and the VPP, as presented in Figure 9 for category C3. V2G services only accounts for a small portion of the total power exchange between EVs and the VPP. This is because the electricity prices offered by the VPP vary by a very small amount at different hours during the charging period, leaving little room to incentivize EVs to discharge to the grid.

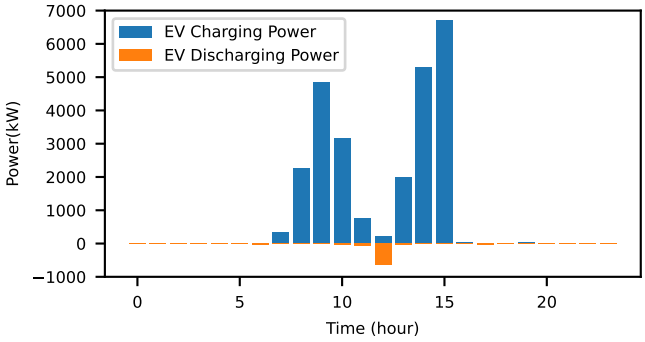

Figure 9. Hourly power exchange (in kW) between the VPP and 1000 EVs in category C3.

The VPP revenue is AUD 5872.81 in the case of C2, and AUD 5889.29 in C3, which is reduced by about AUD 2200 compared with C1. In the case of C2, EV customers pay a hefty additional cost of AUD 3700 for VPP to have the AUD 2200 extra revenue, compared with C1. Considering that the price scheme and price multiplier can be leveraged to readjust VPP revenue, C2 and C3 are superior business models than C1.

4.2.2. Sensitivity Analysis

EV Electricity Price

The charge price is 10% markup over the base electricity price, the discharge price is the 90% discount on the base electricity price for V2G. Three examplery base EV electricity prices (Figure 8) are used for the case studies. To design the cases, for each base EV electricity, the charging strategies T2 and T3 with V2G are examined, as seen in Table 5. The total cost of charging 1000 EVs across six distinct charging categories: C2 to C7, are presented in Table 6.

When comparing the charging categories C2 and C3, the EV charging cost in C3, which incorporates V2G services, is AUD 154.03 lower than that of C2, resulting in a cost reduction of 5.66%. Although C3 provides only a slight advantage over C2, the difference is largely driven by the pricing strategy.

In the comparison between C4 and C5, the EV charging cost in C5, also with V2G services, is AUD 333.76 less than that of C4, representing an 11.33% reduction. However, it is important to note that the revenue generated by the VPP in C5 is AUD 333.89 lower than that in C4. Consequently, there is no incentive for the VPP to provide V2G services in

C5, as the reduction in EV costs merely offsets the revenue loss, creating an unfavorable win-loss scenario. This pricing strategy should therefore be avoided.

Table 5. Categories of EV charging groups (C2–C7).

Category	Charging Strategy	EV Electricity Price
C2	T2	P0
C3	T3 (V2G)	P0
C4	T2	P1
C5	T3 (V2G)	P1
C6	T2	P2
C7	T3 (V2G)	P2

Table 6. Comparison of charging costs for 1000 EVs and VPP revenue (C2–C7).

Categories	EV Charging Cost	VPP Revenue
C2	AUD 2719.82	AUD 5872.81
C3	**AUD 2565.79**	**AUD 5889.29**
C4	AUD 2944.86	**AUD 5666.96**
C5	**AUD 2611.10**	AUD 5333.07
C6	AUD 2807.16	AUD 4205.66
C7	**AUD 2066.58**	**AUD 5376.62**

In contrast, the comparison between the charging categories C6 and C7 reveals that the EV charging cost in C7, which includes V2G services, is AUD 740.58 lower than that in C6, translating to a substantial 26.38% cost reduction. Additionally, the VPP revenue in C7 exceeds that of C6 by AUD 1170.96, representing a 27.83% increase. Thus, the V2G services in C7 represent the most cost-effective and mutually beneficial charging strategy among all the cases considered from C2 to C7.

Overall, the cases from C2 to C7 demonstrate that with an appropriate pricing strategy for EVs, there exists significant potential for both EV owners and VPPs to benefit from V2G services, resulting in a favorable win–win situation.

Battery Capacity

Battery capacity is a key factor influencing the total charging cost of EVs. Table 7 illustrates how varying battery sizes affect overall charging expenses. A smaller capacity, such as 60 kWh, results in a higher total cost of AUD 3090.84 for 1000 EVs, as these vehicles require more frequent charging sessions, often coinciding with peak electricity price periods. In contrast, larger capacities like 100 kWh provide greater flexibility to charge during off-peak hours, lowering the total cost further to AUD 2065.29. The base case of 80 kWh yields an intermediate cost of AUD 2565.79. This analysis underscores the economic advantage of larger battery capacities in reducing EV charging costs.

Table 7. Impact of battery capacity on total EV charging cost.

Battery Capacity	EV Charging Cost
60 kWh	AUD 3090.84
80 kWh (BASE)	AUD 2565.79
100 kWh	**AUD 2065.29**

Rated Charging Power

The rated charging power has a substantial impact on the charging cost of EVs. Table 8 shows how varying the rated charging power influences overall charging expenses. A lower charging rate, such as 7.4 kW, results in a higher total cost of AUD 2565.79 for 1000 EVs because slower chargers limit flexibility, forcing more charging during peak-price periods. In contrast, higher power levels—22 kW and 40 kW—enable faster charging, allowing EVs to better exploit off-peak electricity prices and reducing costs to AUD 2123.16 and AUD 1549.11, respectively. This analysis underscores the economic advantage of higher-power chargers in minimizing charging costs for EV owners.

Table 8. Effect of rated charger power on total EV charging cost.

Rated Charger Power	EV Charging Cost
7.4 kW (BASE)	AUD 2565.79
22 kW	AUD 2123.16
40 kW	**AUD 1549.11**

Max SOC Limit (EV)

Table 9 demonstrates that the maximum SOC limit of EVs positively affects total EV charging costs. Higher SOC thresholds (0.90 and 0.95) enable greater flexibility to charge during off-peak hours, reducing expenses to AUD 2507.31 and AUD 2506.89 for the EV group. In contrast, a lower limit of 0.80 limits energy storage and increases the frequency of charging during costly periods, raising overall costs. Beyond 0.90, further EV charging cost savings diminish, and battery degradation increases. Thus, setting an optimal SOC limit is essential to balance economic efficiency and battery health.

Table 9. Effect of maximum EV charging SOC limit.

Max SOC Limit	EV Charging Cost
0.80 (BASE)	AUD 2565.79
0.90	AUD 2507.31
0.95	**AUD 2506.89**

EV Charger Efficiency

EV charger efficiency significantly impacts the total charging cost for EVs. Table 10 illustrates the effect of varying charger efficiencies on the overall charging expenses. A lower charger efficiency results in higher energy losses during the charging process, leading to increased costs. For instance, with a charger efficiency of 0.80, the total charging cost for 1000 EVs is AUD 3252.84. As the efficiency improves to 0.90, the cost decreases to AUD 2763.79. The base case with a charger efficiency of 0.95 yields the lowest cost of AUD 2565.79. This analysis underscores the importance of utilizing high-efficiency chargers to minimize energy losses and reduce charging costs for EV owners.

Table 10. Impact of EV charger efficiency on total charging cost.

EV Charger Efficiency	EV Charging Cost
0.80	AUD 3252.84
0.90	AUD 2763.79
0.95 (BASE)	**AUD 2565.79**

Battery Degradation

Including battery degradation costs in the VPP optimization model affects the VPP revenue. Table 11 compares the VPP revenue with and without considering battery degradation for the BESS. When battery degradation is not included (BASE case), the VPP revenue is AUD 5889.29. However, when battery degradation costs are accounted for, the VPP revenue decreases to AUD 5612.33, with a degradation cost of AUD 173.28. This reduction in revenue highlights the economic impact of battery wear and tear on the VPP's profitability. Incorporating battery degradation into the optimization model encourages more sustainable operation strategies that balance immediate revenue generation with long-term asset preservation.

Table 11. Impact of including BESS degradation on VPP revenue.

Include BESS Degradation	VPP Revenue	Degradation Cost
No (BASE)	AUD 5889.29	AUD 0
Yes	AUD 5612.33	AUD 173.28

4.2.3. Scalability

Once the RL model is trained, it can be efficiently applied to larger populations of EVs without retraining. In particular, the trained MC-SARSA model was successfully tested to direct the operation of 100,000 EVs, costing only 380.13 s on the researcher's laptop (see configuration above). This highlights the framework's capability to handle large-scale implementations with minimal computational overhead. The near-linear scalability in inference time between 1000 and 100,000 EVs illustrates the efficiency of the decentralized, policy-based design, which allows each EV agent to operate independently using the shared learned policy.

4.2.4. Comparison

The solution method, MC-SARSA, adopted in this research, is compared with the SARSA algorithm, which is widely used in the literature [23,24]. Figure 10 presents the individual charging costs for 50 EVs randomly chosen from the 1000 EVs under the two methods in the case of category C2. From the figure, it is obvious that the MC-SARSA algorithm is dominantly more cost-effective than the model trained using the SARSA algorithm. Both models, SARSA and MC-SARSA, were trained for 120 million epochs. Using the MC-SARSA method, the total charging cost for the 50 EVs is AUD 129.64, compared to AUD 219.28 with the SARSA optimization method. These results demonstrate that the MC-SARSA method achieves a better solution, reducing the EV charging costs by AUD 89.64, which represents a 40.87% decrease. The theoretical advantage of MC-SARSA in producing higher-quality solutions lies in its use of the true return from a complete, real episode. In contrast, SARSA updates the current state-action pair based on an estimated value of the next state-action, which is typically less accurate and introduces bias [41].

To demonstrate the motivation for choosing a RL-based optimization method over traditional MILP approaches [6,7] for large-scale EV dispatching, two comparative experiments were conducted under a V2G setting. In both experiments, the goal is to optimize the charging and discharging behaviors during runtime, given a baseline EV electricity price $P0$. The results, summarized in Table 12, show that MILP models—solved using Gurobi—require 125 s to compute charging/discharging actions for 1000 EVs, whereas the MC-SARSA algorithm completes the same task in just 2.7 s. Once trained, the MC-SARSA model can leverage learned policies to make quick decisions in real time, whereas MILP approach requires solving complex optimization problems from scratch each time. This

paradigm shift from traditional optimization-based methods to pre-trained, learning-based approaches accounts for the substantial reduction in inference time.

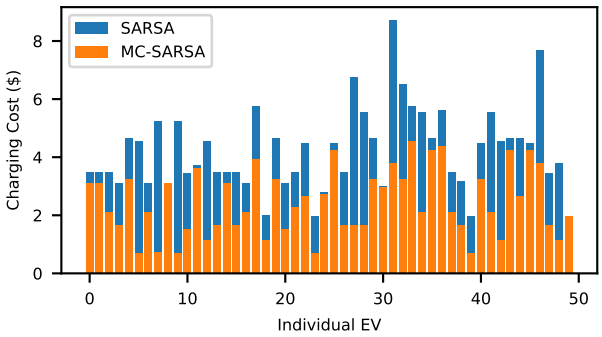

Figure 10. Charging cost comparison of SARSA and MC-SARSA.

In addition to the substantial reduction in inference time, MC-SARSA yields a lower EV charging cost of AUD 2565.79 compared to AUD 2603.92 achieved by MILP. While the VPP revenue obtained using MILP is AUD 5930.42—slightly higher than the AUD 5889.29 from MC-SARSA—the increase is mainly contributed by the inflated EV charging costs. These findings highlight the advantage of the RL-based model in terms of computational time and solution quality, making it better suited for real-time and large-scale applications.

Table 12. Comparison of solution quality and inference time between MILP and MC-SARSA.

Solution Method	VPP Revenue	EV Charging Cost	Inference Time
MC-SARSA	AUD 5889.29	**AUD 2565.79**	**2.7 s**
MILP	**AUD 5930.42**	AUD 2603.92	125.0 s

4.2.5. Limitation

In a decentralized implementation, several potential communication and coordination challenges may arise. Since the VPP does not exert direct control over individual EVs, coordination relies on price signals and local decision-making by autonomous agents. This distributed structure can lead to short-term power fluctuations or collective overreactions in the aggregated load. A well-designed pricing strategy can help mitigate these effects by guiding user behavior away from grid peak hours. Heterogeneous communication standards among charging stations and varying network reliability can introduce further uncertainty in coordination.

The model assumes deterministic market prices and renewable generation forecasts, whereas real-world systems are affected by considerable uncertainty in both market dynamics and renewable generation. Moreover, the RL model's generalization capability may be limited when applied to different price structures or environmental variables, as the trained policy reflects the statistical properties of its training data.

The current study focuses primarily on optimizing the operational and economic performance of the VPP and EVs under idealized grid conditions. Consequently, it does not explicitly consider the physical or operational limitations of the power grid, such as transformer capacity, or voltage stability constraints that may arise when a large number of EVs engage in simultaneous V2G discharging.

5. Conclusions

This study introduces a two-stage optimization framework designed to manage the operation of a VPP that integrates EVs, PVs, and a BESS. The VPP aims to maximize

revenue from both the EV group and the electricity wholesale market, while offering reduced charging costs to the EV group to incentivize their participation in V2G discharge activities. By modeling 1000 EVs in the optimization framework, the VPP model determines the optimal solution by scheduling BESS and trading in the wholesale electricity market. Including a large number of EVs in the VPP modeling enables the business to capture more accurate and realistic data, results, and insights, thereby aiding in better business decision-making. While optimizing VPP operational modeling, gradient-based programming leverages neural networks as a powerful tool for fast and efficient optimality searching in VPP modeling. Custom loss functions for large neural networks overcome the limitations of traditional programming methods, such as strict linearity, convexity, and limited scales of decision variables. MC-SARSA increases the convergence speed for optimizing EV charging and discharging operations, delivering more optimal results within a reasonable time limit. The simulation results demonstrate that, under some charging strategy and EV electricity price selection, optimizing the operations of EVs can lead to a 26.38% reduction in the charging costs for 1000 EVs and the VPP revenue increases by 27.83% with the implementation of V2G services.

Modeling under idealized grid conditions limits the applicability of the findings to real-world networks, as grid constraints can influence feasible power flows and economic outcomes. Future research can extend the model to incorporate detailed grid-level constraints and network-aware optimization. Incorporating a more comprehensive benchmarking and discussion of computational efficiency and convergence would further strengthen the methodological foundation.

The integration of stochastic or robust optimization and transfer learning to mitigate the uncertainties identified in the limitation section can further improve the adaptability and reliability of the proposed method. Moreover, investigating the iterative optimization of EV electricity pricing within a bi-level framework represents a promising direction. In addition, studying the impacts of limited charging infrastructure on VPP profitability and the charging/discharging behaviors of a large number of EVs remains an important area for further exploration.

Author Contributions: Conceptualization, R.Z. and J.Q.; methodology, R.Z.; formal analysis, R.Z., J.Q., J.W. and L.L.; writing—original draft preparation, R.Z.; writing—review and editing, R.Z. and L.L. supervision, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available on request from the authors

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

List of symbols used in this paper

Symbol	Description
EV Model Variables	
N	Total number of EVs
T	Total number of time steps in scheduling period
t	Time step index
$t_{arr}^{EV_i}$	Arrival time of i-th EV
$t_{dep}^{EV_i}$	Departure time of i-th EV
$S^{EV_i}(t)$	State of charge (SOC) of i-th EV at time t
$S_{SOC-init}^{EV_i}$	Initial SOC of i-th EV upon arrival
$S_{reqd}^{EV_i}$	Required SOC increment before departure

$S_{min}^{EV_i}, S_{max}^{EV_i}$	Minimum and maximum SOC limits of i-th EV
C^{EV_i}	Battery capacity of i-th EV
$P_{chg}^{EV_i}(t)$	Charging power of i-th EV at time t
$P_{disch}^{EV_i}(t)$	Discharging power of i-th EV at time t
$P_{chg-min}^{EV_i}, P_{chg-max}^{EV_i}$	Minimum and maximum charging power
$P_{disch-min}^{EV_i}, P_{disch-max}^{EV_i}$	Minimum and maximum discharging power
$\eta_{chg}^{EV_i}, \eta_{disch}^{EV_i}$	Charging and discharging efficiency of i-th EV
$\lambda_{pur}^{EV}(t)$	Electricity purchase price from VPP to EV
$\lambda_{V2G}^{EV}(t)$	V2G electricity selling price from EV to VPP
$\mu_{V2G}^{EV_i}$	V2G participation fee charged to i-th EV owner
C_{pur}^{EV}	Total EV charging cost
R_{V2G}^{EV}	Total V2G revenue from EVs to VPP
Δt	Duration of each time step
VPP Model Variables	
$P_{EV}^{VPP}(t)$	Net power exchanged between EV group and VPP
$P_{grid}^{VPP}(t)$	Power purchased from the wholesale grid
$P_{feed-in}^{VPP}(t)$	Power sold by VPP to the grid
$\lambda_{pur}^{VPP}(t)$	Wholesale electricity purchase price
$\lambda_{feed-in}^{VPP}(t)$	Electricity feed-in price to the wholesale market
C_{pur}^{VPP}	Cost of electricity purchased from the market
$R_{feed-in}^{VPP}$	Revenue from selling electricity to the grid
α	Discount factor for feed-in electricity price
BESS Variables	
$S^{ESS}(t)$	SOC of BESS at time t
S_{init}^{ESS}	Initial SOC of the BESS
$S_{min}^{ESS}, S_{max}^{ESS}$	Minimum and maximum SOC limits
$P_{chg}^{ESS}(t)$	Charging power of BESS at time t
$P_{disch}^{ESS}(t)$	Discharging power of BESS at time t
$P_{chg-min}^{ESS}, P_{chg-max}^{ESS}$	Minimum and maximum charging power
$P_{disch-min}^{ESS}, P_{disch-max}^{ESS}$	Minimum and maximum discharging power
$\eta_{chg}^{ESS}, \eta_{disch}^{ESS}$	Charging and discharging efficiency of BESS
C^{ESS}	Energy capacity of the BESS
PV System Variables	
$P_{PV}^{VPP}(t)$	PV generation power at time t
$I(t)$	Global horizontal irradiance at time t
A	Total PV panel area
η^{PV}	PV conversion efficiency
Pricing Strategy Parameters	
$\tilde{p}(t)$	Base electricity price at time t
$\lambda_{base}^{EV}(t)$	Base EV electricity price
μ	Price multiplier for base EV price
τ_{mu}	Markup factor for charging price
τ_{disc}	Discount factor for V2G price
$P_{strtg}(\lambda_{whsle}, t)$	Pricing strategy function
λ_{whsle}	Wholesale electricity price vector
Algorithm Parameters	
α_{EV}, ρ_{EV}	Penalty coefficient and exponent for EV model
α_{VPP}, ρ_{VPP}	Penalty coefficient and exponent for VPP model
L_{VPP}	Custom loss function for VPP optimization
$\psi_{pen}^{EV_i}$	Penalty terms for i-th EV for constraint violation
ψ_{pen}^{VPP}	Penalty terms for VPP model for constraint violation
$R_{rwd}^{EV_i}$	Reward function for i-th EV

γ	Discount factor in reinforcement learning
$Q(s,a)$	Action-value function for state s and action a
$Q_{masked}(s,a)$	Masked Q-value excluding invalid actions
$m(s)$	Action mask vector for valid actions
$\mathcal{A}$	Set of possible actions
a^*	Optimal action selection
R_k	Reward at step k
G_t	Return (cumulative discounted reward)
η	Learning rate for gradient descent
$\nabla f(x_t)$	Gradient of loss function

List of abbreviations used in the paper.

Abbreviation	**Full Form**
AEMO	Australian Energy Market Operator
BESS	Battery Energy Storage System
DER	Distributed Energy Resource
DNN	Deep Neural Network
DP	Dynamic Programming
ESS	Energy Storage System
EV	Electric Vehicle
MC	Monte Carlo
MC-SARSA	Monte Carlo SARSA (episodic reinforcement learning method)
MILP	Mixed-Integer Linear Programming
MSE	Mean Squared Error
PV	Photovoltaic
RL	Reinforcement Learning
SARSA	State-Action-Reward-State-Action (RL algorithm)
SOC	State of Charge
V2G	Vehicle-to-Grid
VPP	Virtual Power Plant

References

1. Liu, H.; Zhao, Y.; Gu, C.; Ge, S.; Yang, Z. Adjustable capability of the distributed energy system: Definition, framework, and evaluation model. *Energy* **2021**, *222*, 119674. [CrossRef]
2. Mashhour, E.; Moghaddas-Tafreshi, S. A review on operation of micro grids and virtual power plants in the power markets. In Proceedings of the 2009 2nd International Conference on Adaptive Science & Technology (ICAST), Accra, Ghana, 14–16 January 2009; IEEE: New York, NY, USA, 2009; pp. 273–277.
3. Vasirani, M.; Kota, R.; Cavalcante, R.L.; Ossowski, S.; Jennings, N.R. An agent-based approach to virtual power plants of wind power generators and electric vehicles. *IEEE Trans. Smart Grid* **2013**, *4*, 1314–1322. [CrossRef]
4. Chen, Y.; Niu, Y.; Du, M.; Wang, J. A two-stage robust optimization model for a virtual power plant considering responsiveness-based electric vehicle aggregation. *J. Clean. Prod.* **2023**, *405*, 136690. [CrossRef]
5. Argade, S.G.; Aravinthan, V.; Esra Büyüktahtakın, I.; Joseph, S. Performance and consumer satisfaction-based bi-level tariff scheme for EV charging as a VPP. *IET Gener. Transm. Distrib.* **2019**, *13*, 2112–2122. [CrossRef]
6. Ding, Z.; Lu, Y.; Lai, K.; Yang, M.; Lee, W.J. Optimal coordinated operation scheduling for electric vehicle aggregator and charging stations in an integrated electricity-transportation system. *Int. J. Electr. Power Energy Syst.* **2020**, *121*, 106040. [CrossRef]
7. Alanazi, M.; Alanazi, A.; Alruwaili, M.; Salem, M.; Ueda, S.; Senjyu, T.; Mohamed, F.A. Developing a Transactive Charging Control Framework for EV Parking Lots Equipped With Battery and Photovoltaic Panels: A MILP Approach. *IEEE Access* **2024**, *12*, 108731–108743. [CrossRef]
8. Bellman, R. Dynamic programming. *Science* **1966**, *153*, 34–37. [CrossRef]
9. Tan, Z.; Wang, G.; Ju, L.; Tan, Q.; Yang, W. Application of CVaR risk aversion approach in the dynamical scheduling optimization model for virtual power plant connected with wind-photovoltaic-energy storage system with uncertainties and demand response. *Energy* **2017**, *124*, 198–213. [CrossRef]
10. Yang, D.; He, S.; Wang, M.; Pandžić, H. Bidding strategy for virtual power plant considering the large-scale integrations of electric vehicles. *IEEE Trans. Ind. Appl.* **2020**, *56*, 5890–5900. [CrossRef]

11. Wang, W.; Chen, P.; Zeng, D.; Liu, J. Electric vehicle fleet integration in a virtual power plant with large-scale wind power. *IEEE Trans. Ind. Appl.* **2020**, *56*, 5924–5931. [CrossRef]
12. Wang, Y.; Gao, W.; Qian, F.; Li, Y. Evaluation of economic benefits of virtual power plant between demand and plant sides based on cooperative game theory. *Energy Convers. Manag.* **2021**, *238*, 114180. [CrossRef]
13. Feng, B.; Liu, Z.; Huang, G.; Guo, C. Robust federated deep reinforcement learning for optimal control in multiple virtual power plants with electric vehicles. *Appl. Energy* **2023**, *349*, 121615. [CrossRef]
14. Honarmand, M.; Zakariazadeh, A.; Jadid, S. Integrated scheduling of renewable generation and electric vehicles parking lot in a smart microgrid. *Energy Convers. Manag.* **2014**, *86*, 745–755. [CrossRef]
15. Liu, Q.; Tian, J.; Zhang, K.; Yan, Q. Pricing Strategy for a Virtual Power Plant Operator with Electric Vehicle Users Based on the Stackelberg Game. *World Electr. Veh. J.* **2023**, *14*, 72. [CrossRef]
16. Falabretti, D.; Gulotta, F.; Siface, D. Scheduling and operation of RES-based virtual power plants with e-mobility: A novel integrated stochastic model. *Int. J. Electr. Power Energy Syst.* **2023**, *144*, 108604. [CrossRef]
17. Rahman, S.; Sales-Ortiz, J.; Ardakanian, O. Making a Virtual Power Plant out of Privately Owned Electric Vehicles: From Contract Design to Scheduling. In Proceedings of the 14th ACM International Conference on Future Energy Systems, Orlando, FL, USA, 20–23 June 2023; pp. 459–472.
18. Cui, J.; Wu, J.; Wu, C.; Moura, S. Electric vehicles embedded virtual power plants dispatch mechanism design considering charging efficiencies. *Appl. Energy* **2023**, *352*, 121984. [CrossRef]
19. Wang, Y.; Dou, W.; Tong, Y.; Yang, B.; Zhu, H.; Xu, R.; Yan, N. Optimal configuration method of electric vehicle's participating in Load Aggregator's VPP low-carbon economy. *Energy Rep.* **2023**, *9*, 1093–1100. [CrossRef]
20. Li, Z.; Lei, X.; Shang, Y.; Jia, Y.; Jian, L. A genuine V2V market mechanism aiming for maximum revenue of each EV owner based on non-cooperative game model. *J. Clean. Prod.* **2023**, *414*, 137586. [CrossRef]
21. Papadopoulos, P.; Jenkins, N.; Cipcigan, L.M.; Grau, I.; Zabala, E. Coordination of the charging of electric vehicles using a multi-agent system. *IEEE Trans. Smart Grid* **2013**, *4*, 1802–1809. [CrossRef]
22. Rashidizadeh-Kermani, H.; Vahedipour-Dahraie, M.; Shafie-Khah, M.; Siano, P. A stochastic short-term scheduling of virtual power plants with electric vehicles under competitive markets. *Int. J. Electr. Power Energy Syst.* **2021**, *124*, 106343. [CrossRef]
23. Wu, S.; Hu, W.; Lu, Z.; Gu, Y.; Tian, B.; Li, H. Power system flow adjustment and sample generation based on deep reinforcement learning. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1115–1127. [CrossRef]
24. Xi, L.; Zhou, L.; Xu, Y.; Chen, X. A multi-step unified reinforcement learning method for automatic generation control in multi-area interconnected power grid. *IEEE Trans. Sustain. Energy* **2020**, *12*, 1406–1415. [CrossRef]
25. Inci, M.; Savrun, M.M.; Çelik, Ö. Integrating electric vehicles as virtual power plants: A comprehensive review on vehicle-to-grid (V2G) concepts, interface topologies, marketing and future prospects. *J. Energy Storage* **2022**, *55*, 105579. [CrossRef]
26. Qi, J.; Li, L. Economic Operation Strategy of an EV Parking Lot with Vehicle-to-Grid and Renewable Energy Integration. *Energies* **2023**, *16*, 1793. [CrossRef]
27. Van der Meer, D.; Mouli, G.R.C.; Mouli, G.M.E.; Elizondo, L.R.; Bauer, P. Energy management system with PV power forecast to optimally charge EVs at the workplace. *IEEE Trans. Ind. Inform.* **2016**, *14*, 311–320. [CrossRef]
28. Rouzbahani, H.M.; Karimipour, H.; Lei, L. A review on virtual power plant for energy management. *Sustain. Energy Technol. Assess.* **2021**, *47*, 101370. [CrossRef]
29. Song, J.; Yang, Y.; Xu, Q. Two-stage robust optimal scheduling method for virtual power plants considering the controllability of electric vehicles. *Electr. Power Syst. Res.* **2023**, *225*, 109785. [CrossRef]
30. Mei, S.; Tan, Q.; Liu, Y.; Trivedi, A.; Srinivasan, D. Optimal bidding strategy for virtual power plant participating in combined electricity and ancillary services market considering dynamic demand response price and integrated consumption satisfaction. *Energy* **2023**, *284*, 128592. [CrossRef]
31. Shi, Y.; Tuan, H.D.; Savkin, A.V.; Duong, T.Q.; Poor, H.V. Model predictive control for smart grids with multiple electric-vehicle charging stations. *IEEE Trans. Smart Grid* **2018**, *10*, 2127–2136. [CrossRef]
32. Zeng, H.; Huang, Y.; Zhan, K.; Yu, Z.; Zhu, H.; Li, F. Multi-Agent DDPG-Based Multi-Device Charging Scheduling for IIoT Smart Grids. *Sensors* **2025**, *25*, 5226. [CrossRef] [PubMed]
33. Chen, P.; Han, L.; Xin, G.; Zhang, A.; Ren, H.; Wang, F. Game theory based optimal pricing strategy for V2G participating in demand response. *IEEE Trans. Ind. Appl.* **2023**, *59*, 4673–4683. [CrossRef]
34. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
35. Wang, L.; Dubey, A.; Gebremedhin, A.H.; Srivastava, A.K.; Schulz, N. MPC-based decentralized voltage control in power distribution systems with EV and PV coordination. *IEEE Trans. Smart Grid* **2022**, *13*, 2908–2919. [CrossRef]
36. Yang, Q.; Wang, H.; Wang, T.; Zhang, S.; Wu, X.; Wang, H. Blockchain-based decentralized energy management platform for residential distributed energy resources in a virtual power plant. *Appl. Energy* **2021**, *294*, 117026. [CrossRef]

37. Li, Z.; Yang, Y.; Xia, R.; Xia, H.; Su, Y.; Li, Y. Multi-objective optimization of indoor comfort and carbon emissions using a feature-fusion CNN-MLP model for morphological analysis of mixed residential areas. *Energy Rep.* **2025**, *13*, 6291–6327. [CrossRef]
38. Nesterov, Y. *Lectures on Convex Optimization*; Springer: Cham, Switzerland, 2018; Volume 137.
39. Murphy, K.P. *Probabilistic Machine Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2022.
40. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2020.
41. Szepesvári, C. *Algorithms for Reinforcement Learning*; Springer Nature: Cham, Switzerland, 2022.
42. Wang, Z.; Li, X.; Sun, L.; Zhang, H.; Liu, H.; Wang, J. Learning state-specific action masks for reinforcement learning. *Algorithms* **2024**, *17*, 60. [CrossRef]
43. Qi, J.; Li, L.; Lei, G. Economic Operation of a Workplace EV Parking Lot under Different Operation Modes. In Proceedings of the 2021 31st Australasian Universities Power Engineering Conference (AUPEC), Virtual, 26–30 September 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
44. Australian Energy Market Operator. Wholesale Electricity Market (WEM), Aggregated Price and Demand Data. 2024. Available online: https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data : (accessed on 2 November 2024).
45. NSRDB: National Solar Radiation Database . Available online: https://nsrdb.nrel.gov/data-viewer (accessed on 5 May 2024).