Assembly and Competition for Virtual Power Plants with Multiple ESPs through A "Recruitment – Participation" Approach

Xin Lu, *Student Member, IEEE*, Jing Qiu, *Senior Member, IEEE*, Cuo Zhang, *Member, IEEE*, Gang Lei, *Senior Member, IEEE*, Jianguo Zhu, *Senior Member, IEEE*

***Abstract*— Distributed energy resource (DER) owners have many concerns when participating in virtual power plants (VPPs), including low returns and a limited variety of plans. Therefore, it lacks incentives for them to participate in VPPs. This paper proposes a flexible VPP assembly approach to recruit DERs by energy service providers (ESPs). The contributions of each ESP and DER to VPPs are evaluated separately as the basis for a fair payoff allocation. Rebates are adopted to incentivize DER owners to participate in the VPPs. Based on the above fundamentals, a multi-ESP VPP assembly and competition method are proposed and simulated, considering diverse preference behaviors and electricity price fluctuations. The existence of the solution to the proposed multi-ESP VPP rebate competition problem is discussed. Case studies are conducted to simulate a 35-VPP assembly and disassembly process and a 6-ESP rebate competition process. The results indicate that DER owners are motivated to participate in VPPs with high payoffs, and the payoffs of ESPs increase by attracting more customers.**

***Index Terms*—Virtual Power Plant (VPP), VPP Assembly, Recruitment–Participation, Sharpley Value, Rebate Competition.**

I. Introduction

The number of household distributed energy resources (DERs), including rooftop photovoltaics (PVs) and batteries, has increased rapidly in recent years. They operate mainly for self-consumption and are not allowed to participate directly in energy trading. Nowadays, DERs can be aggregated by energy service providers (ESPs) in the form of virtual power plants (VPPs) [1] and can be regarded as prosumers to participate in several markets [2, 3] to gain extra profits. However, DER owners are currently not enthusiastic about participating in VPPs, since VPPs are usually formed by electricity retailers acting as ESPs. They can only passively participate in the VPPs with the price offered by the ESPs [4, 5]. There are two keys to promoting DERs' participation in VPPs. One is to enhance the voice of DERs and change the role of DERs from passive participants to active ones, and the other is to increase the profit of DERs through fair profit distribution, which means that the current ESP-centered approach for VPP construction needs to be revisited.

X. Lu, J. Qiu, C. Zhang and J. Zhu are with the School of Electrical and Information Engineering, The University of Sydney, Darlington 2008, New South Wales, Australia (e-mail: xin.lu@sydney.edu.au; jeremy.qiu@sydney.edu.au; cuo.zhang@sydney.edu.au; jianguo.zhu@sydney.edu.au)

G. Lei is with School of Electrical and Data Engineering, University of Technology Sydney, Ultimo 2007, New South Wales, Australia (e-mail: gang.lei@uts.edu.au)

Previous studies have focused on the aggregation and profitability of VPPs in the energy market with optimized strategies. Liu et al. [6] proposed a risk-averse bidding strategy with a data-driven distributional robust model. A deep reinforcement learning (DRL) method was proposed to model the bidding action and achieve a higher profit for a VPP [7]. A competitive framework was designed to model the participation of VPPs in the electricity market with multiple objectives, including profit maximization and cost minimization, through a multi-leader-follower game theory and Karush–Kuhn–Tucker conditions [8]. These studies are typically based on the assumption of an existing VPP and ignore how to incentivize DER owners to participate in the VPP.

Several research works have been conducted on how to increase the profits of DER participants in VPP. The dynamic pricing mechanism is the most common approach. Nguyen et al. [9] designed an optimal dynamic pricing model for demand response integration. The load-serving entity was considered the leader, and the load aggregators were followers in a Stackelberg game. A dynamic retail pricing scheme was studied in [10] with load demand and market price information based on a binary genetic algorithm. However, these VPPs are constructed by ESPs, who are commonly the profit allocation rule makers. Although ESPs provide DERs with dynamic prices, they occupy the interest space of DERs [11].

A commonly used approach to achieve fair profit allocation is to base it on the contribution of each VPP participant to the whole. The Sharpley Value (SV) method is one of the most common solutions and was introduced in [12, 13] to allocate benefits among PVs, wind power plants, gas turbines, and batteries. However, the ESP's contribution was not evaluated, and no payoff was allocated to the ESP. In [14], Gougheri et al. proposed an SV-based fair allocation method to divide profit among PVs, batteries, and other components. However, there are multiple PVs and batteries in a VPP, and the allocation method in [16] only evaluates the contributions of the PV and battery groups but cannot allocate precise profits to each battery or PV within the groups. Therefore, a fair method for profit allocation to every DER individual instead of DER groups is required.

In addition to fair benefit allocation, providing DER participants with certain incentives is also a commonly used method in VPP profit distribution. Customized prices or rebates are the most common ways. Chen et al. in [15] proposed a customized pricing mechanism to motivate VPP prosumers.

The interaction between a retailer and prosumers is modelled as a Stackelberg game and solved by a Salp Swarm Algorithm. A customized rebate pricing mechanism was proposed in [16] to reward prosumers for supporting the power system operation, considering the heterogeneous prosumers' dynamic selecting process based on an evolutionary game. Therefore, a profit allocation method should combine fairness and incentives.

The studies mentioned above focused solely on the increase of individual DER profits, mainly from the perspective of ESPs, without modeling the behaviors of DER owners. Moreover, it remains unclear whether these studies provide incentives for the participation rate of DER participants. Currently, there is a lack of literature on modeling the willingness of DERs to participate in VPPs or to choose ESPs. However, some studies have proposed nonlinear models to represent the willingness to install solar panels [17], the behavior of electric vehicles (EVs) to accept dispatch [18], and the behavior of consumers in selecting electricity retailers [19]. Nevertheless, these models do not entirely address the issue, as consumers primarily aim to reduce their bills. Selecting ESPs for DERs' participation in VPPs has many similarities to mutual fund investment. Professionals manage VPPs and mutual funds to achieve asset appreciation through an optimal portfolio. Therefore, the mutual fund selection model [20-22] can be introduced to represent DERs' behavior for ESP selection, making the VPP assembly process more realistic.

DER owners consider more than just profit when selecting ESPs, as their demands are diverse. Some prioritize high short-term returns while others seek stable long-term returns. Therefore, differentiated VPP plans are necessary to accommodate the varying demands of different DER owners. Since introducing the multi-retailer model has brought dynamism and increased efficiency to the electricity market [23], introducing ESP competition in VPPs can ensure their economic operation with full competitiveness and provide differentiated services to attract participants. Xu et al. modelled competitive behaviors between two VPPs based on a non-cooperative pricing game in [1] with a limited number of DERs. They partitioned the original optimization problem into two sub-problems. An interactive dispatch model of VPPs was proposed in [24]. The aforementioned studies focused on the competition between two or three VPPs, but more complex competition among a larger number of VPPs still needs to be investigated for practical implementation.

The rebate competition problem among multiple ESP is difficult to solve using conventional mathematical methods. In recent years, DRL algorithms, combining conventional reinforcement learning and deep learning, have provided a new opportunity to address such problems. The deep deterministic policy gradient (DDPG) algorithm and the twin delay deep deterministic policy gradient (TD3) algorithm [25] have been used to solve the profit game between EV charging stations and VPPs. Ma et al. designed a DDPG-based incentive determination agent to simultaneously maximize the profits of both operators and prosumers [26]. Multi-agent DRL was proposed in [27] to realize coordinated control and improve the performance of VPPs in the frequency regulation market. Multi-agent DDPG (MADDPG) was deployed to obtain the optimal dispatching strategy of the main network and microgrids in a bi-level energy management framework [27]. As a model-free algorithm, MADDPG is also suitable for competition between multiple VPPs.

As identified above, the following knowledge gaps should be addressed:

1) Previous studies ignore how to incentivize DER owners to participate in the VPP, and it remains uncertain whether providing payoff incentives would yield a higher participation rate among DER participants.
2) Existing VPPs are ESP-centered, which means ESPs are the profit allocation rule makers. Therefore, the profits of ESPs and individual DERs cannot be fairly distributed.
3) A single VPP cannot cater to the diverse needs of DER owners. It is unknown whether the assembly and competition of multiple VPPs will affect individual DER returns and overall participation rates.

To address these issues, this paper designs a "recruitment–participation" approach to VPP assembly and proposes a multi-ESP VPP competition process. The main contributions of this work are as follows:

1) The paper proposes a recruitment-participation approach to replace the ESP-centered VPP construction model, where ESPs offer different risk-return strategies and recruit DERs to build a VPP together.
2) A payoff allocation method based on fairness and incentives is introduced. A fair profit allocation method is developed for individual DERs and ESPs based on the SV method. The incentive approach aims to encourage DERs to participate in VPPs through rebates.
3) To meet the diverse needs of DERs, the paper proposes a multi-ESP VPP model that considers DERs' behaviors and allows for ESP selection. DRL is employed to solve the rebate competition problem among ESPs, achieving a stable solution among ESPs.

II. Approach for VPP Assembly and Competition through A Recruitment – Participation Approach

In this paper, VPPs are no longer ESP-centered. ESPs can recruit any DERs, and DERs can participate in any ESP at will. If the ESPs and DERs are considered modules, the VPP construction process, jointly carried out by ESPs and DERs, can be called VPP assembly. The following assumptions on the VPP assembly are made in this study.

Assumption 1: DERs can choose ESPs according to their preferences. There is no cost for DERs to switch ESPs.

Assumption 2: The operation cycle (OC) is defined as the smallest operation period for a VPP assembly. Within each OC, DERs cannot exit VPPs, and ESPs cannot change the published strategy.

Assumption 3: An ESP recruits DERs through a prospectus. When the parties sign a contract, control of the DER is placed under the control of the ESP.

Fig.1 shows the timeline of the entire process of 'Recruitment - Participation' approach. Initially, there is no VPP, only DER owners who are willing to participate in VPPs and potential ESPs as shown in Fig. 1(a). Then, the VPPs are assembled (Fig. 1(b)), followed by possible challenges of disassembly (Fig. 1(c)), reassembly (Fig. 1(d)), and competition. In this process, the profit-risk profiles are also changed, as shown in Figs. 1(e-g). The details are introduced and explained in the next subsection.

A. Recruitment – Participation Approach

1) VPP Assembly

Assume that there currently exists an area with several PV and battery owners. Some of whom are not satisfied to use PV and battery only for their own consumption and wish to obtain more profit. Therefore, they are highly willing to participate in VPP; however, based on their own preferences, their preferences for risk and return are diverse, which means that one single VPP plan cannot satisfy everyone's needs. Therefore, multiple VPP ESPs need to be introduced to provide differentiated plans with competition (Fig. 1 (a)).

ESP: Multiple ESPs join as participants, and through prediction, simulation and calculation, each ESP provides a prospectus to attract DERs before an OC.

DER: Partial DERs choose different ESPs to assemble VPPs according to the published prospectus and their own preferences, while the remaining DERs are in a wait-and-see situation, as shown in Fig. 1(b). During the initial stage of VPP assembly, DERs are not allowed to exit the VPP for some cycles, named closed OCs.

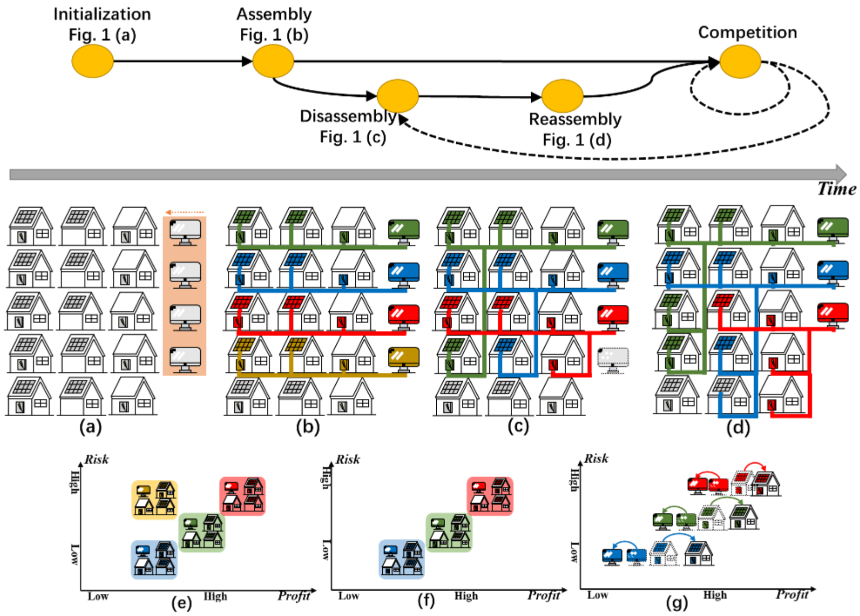

Fig. 1 Proposed 'Recruitment – Participation' approach for VPP. (a) Initialization, (b) Assembly, (c) Disassembly, (d) Reassembly, (e-g) Changes in risk-profit in VPP competition

2) VPP Disassembly and Reassembly

After the closed OCs, actual profit and expected profit of VPP operation may be different. Therefore, DERs' confidence may change since the actual and expected payoffs differ. Some DERs will leave the current VPP and choose a new ESP to assemble VPP.

ESP: With the loss of members, the high-risk and low-profit VPP ESPs (the yellow one in Fig. 1(e)) will eventually disassemble the VPP. Only the low-risk and high-profit VPPs (the green, blue and red ones in Figs. 1(e) and 1(f)) can be retained, and the ESPs will gradually become oligopolies by offering differentiated VPP plans.

DER: The DERs of the disassembled VPPs can choose other ESPs, as shown in Fig. 1(c). And the participants of the other VPPs can also re-select different VPP plans with the existing ESPs according to their preferences.

3) VPP Competition

After a period of operation, the number of DER participants in each VPP cannot be further increased. ESPs give DER owners rebates to attract non-participating DER owners, while ESPs compete to attract DERs from other VPPs.

ESP: ESPs provide rebates to DER participants to reward them [16]. In Fig. 1(g), the arrows represent the profit change of ESPs and other participants before and after the rebates. The dotted icons are profits before the rebates, and the solid ones are with the rebates.

DER: The rebates promote ESP reselection from PV and battery participants [28]. In addition, as shown in Fig. 1 (d), the high profits with rebates can attract wait-and-see DERs.

B. Operation Cycle

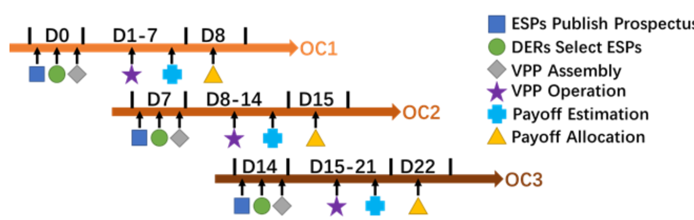

Fig. 2 VPP assembly flow for OCs

The OC is the smallest unit of time for VPP assembly. A short OC will lead to frequent changes in the number of DERs, which is a challenge for ESPs to participate in the energy market. However, if OC is very long, it is not conducive for DER participants to regularly re-select ESP based on actual benefits and risks. Note that OC can be set to any number of days. This paper considers the weekly periodicity of the electricity spot price market, so OC is set to seven days. Fig. 2

illustrates the VPP assembly process for OCs.

Before the VPP assembly, Day0 (D0) in Fig. 2, ESPs publish a prospectus. Potential DERs can select the appropriate ESPs based on their preferences and prospectus disclosures, and then they are assembled into VPPs. Starting from D1, the first day of OC1, ESPs perform VPP operations according to the established optimization strategy while giving the daily estimated payoffs for each DER until D7. Meanwhile, each ESP is required to publish a prospectus for OC2 on the last day of OC1, and DERs can re-select ESPs. On D8, the ESPs calculate the actual payoff of each DER based on the SV and make the allocation. D8 also serves as the first day of OC2, on which ESPs start a new VPP operation round and then repeat the above process.

C. *Prospectus*

The ESPs attract DERs by publishing a prospectus, which includes the following information:

- A summary of the ESP's background,
- Basic information on the VPP, including
  - A brief description of the strategy,
  - Predicted profits,
  - Nominal risks, and
  - Rebate coefficients;
- Historical information of VPP (if available), including
  - Actual VPP profits and fluctuations,
  - Actual long and short term returns for DERs, and
  - Actual size of DERs.

III. Behaviors in ESPs Competition and Proposed Models

Prospectus directly influences DERs' selection on ESPs. The choice of participants, in turn, promotes the ESPs' rebate competition. The participation, ESP selection, and rebate competition models are proposed to simulate the participants' behaviors in the VPP competition process.

A. *Participation or non-participation*

Different DERs have various elasticities of willingness under different payoffs, and linear models are insufficient to portray the behavior of DERs participating in VPP. However, there is a lack of research on the relationship between payoffs and VPP participation rate for DERs. Based on the PV installation and EV dispatching models in [17, 18], a nonlinear model is assumed to describe the payoff increment and VPP participation rate as follows:

$$\mu = -\frac{1}{\left((\Theta^{w/}-\Theta^{w/o})/\Theta^{w/o}\right)*a+b} + 1 \tag{1}$$

where μ represents the overall participation rate for DERs; $\Theta^{w/}$ and $\Theta^{w/o}$ represent the average profits with and without participation in the VPPs with different ESPs; a and b are coefficients used to control for the effect of incremental marginal payoffs on the participation ratio, $a \geq 1$ and $b \geq 1$, when $\Theta^{w/} = \Theta^{w/o}$, and the initial participation rate $\mu = 1 - 1/b$. Therefore, the marginal incremental payoff diminishes the increase in the DER participation rate, meaning most DERs' willingness significantly increases with a slight payoff rise. In contrast, a small number of DERs' willingness remains low even after a significant payoff increase.

B. *ESP Selection by Participants*

DER participants should carefully consider various factors when selecting ESPs to assemble VPPs. Since the multi-ESP competition model is proposed for the first time in this paper, there is a lack of research on how DER participants choose ESPs. The model for electricity retailer selection [19] does not apply here, as customers have a single purpose, reducing their bill. However, the participation of DER owners in VPP by selecting ESPs can be considered as an investment and it has many similarities to mutual fund investment. For example: 1) Fund investment requires investors to invest cash to obtain returns, while VPP participants need to provide control over PV and battery to obtain returns. 2) Both funds and VPPs offer diversification in risk and return and require professionals to manage risks. 3) Funds and VPPs can recruit investors and DER participants, respectively, through the disclosure of prospectus, including historical returns, risk, and scale. 4) Fund managers receive management fees, while VPP ESPs also receive payoff for bringing return growth to DER participants. Therefore, based on the mutual fund selection model [20-22], an ESP selection model is proposed as follows:

$$\Gamma = \xi_1\Theta_{\mathrm{ReturnL}} + \xi_2\Theta_{\mathrm{ReturnS}} + \xi_3(1-\Theta_{\mathrm{Risk}}) + \xi_4(1-\Theta_{\mathrm{Fluc}}) + \xi_5\Theta_{\mathrm{Scale}} + \xi_6\Theta_{\mathrm{Other}} \tag{2}$$

where Γ denotes the potential for an ESP to be selected by DERs; coefficients $\xi_1, \xi_2, \cdots, \xi_6$ indicate the weights for investment preference of DER participants; $\Theta_{\mathrm{ReturnL}}$ and $\Theta_{\mathrm{ReturnS}}$ represent the historical average long- and short-term payoffs which can be obtained from the disclosed prospectus. Due to the lack of data on historical returns for the new assembly VPP, the expected returns can be used instead for the historical payoffs; Θ_{Risk} represents the nominal risk rating; Θ_{Fluc} represents the fluctuation of historical returns. of disclosed historical returns; and Θ_{Scale} represents the average number of DER owners recruited by one ESP, respectively. The five factors can be obtained from the prospectus of VPP and should be normalized to 0 to 1. Apart from the deterministic factors mentioned above, investment decisions often involve some uncertain factors. In [29], Dimitrios et. al points out that individual investors may rely on newspapers, media and noise in the market when making their investment decisions. Reputation, brand effects, and other factors may also influence investors' choices [30]. However, these factors are often difficult to quantify, so we use a random number between 0 and 1 (denoted as Θ_{Other}) to describe the uncertainty of individual choices.

Similar to fund investors [20, 31], DER owners also have different preferences when choosing an ESP to assemble VPPs. We categorize DER owners into different groups based on their investment preferences, including professional, ambitious, moderate, conservative, and cautious [32]. Their risk profiles and profit expectations are shown in Table I. Additionally, according to Table 1, we assumed weights $\xi_1, \xi_2, \cdots, \xi_6$ for different preference participants, which are detailed in the experiment setting of Section VI.

TABLE I. Classification of Participant Groups

Types	Risk Profile	Expectation

Professional	Takes Necessary Risks	Maximum Return
Ambitious	Highly Risk Taking	High Short-Term Return
Moderate	Comfortable Levels of Risk	Good, Steady Return
Conservative	Risk Averse	Regular Return
Cautious	Extremely Risk Averse	Minimum Return

Algorithm 1 below simulates the ESP selection process for each DER. Firstly, the participant's investment type and the factors' coefficients are determined. The relevant variables $[\Theta_{\text{ReturnL}}, \Theta_{\text{ReturnS}}, \cdots, \Theta_{\text{Scale}}]$ can be extracted from the prospectus published by each ESP. A random variable Θ_{Other} is generated, and the participant's probability of participation for each ESP, Γ, can be calculated. The Γ of all ESPs are then collected and composed into multiple probability intervals, such as $[0, \Gamma_1]$ $[\Gamma_1, \Gamma_1+\Gamma_2]$, $\cdots$, $[\sum_{j=1}^{J-1}\Gamma_j, \sum_{j=1}^{J}\Gamma_j]$. The participant generates a random value between $[0, \sum_{j=1}^{J}\Gamma_j]$ and determines in which ESP's probability interval the random value lies to simulate the participant's willingness. Repeat the above steps until all participants have made their decisions.

C. ESP Rebate Competition

ESPs that adopt homogeneous competition are highly susceptible to disassembly due to the loss of participants once it becomes less profitable than other ESPs offering the same risk plan. The ESPs offer differentiated plans to achieve differentiated competition to attract participants with different preferences. For further profit, ESPs give rebates to attract more participants. The ESP selection based on (2) is related to historical profits Θ_{ReturnL} and Θ_{ReturnS}, profit fluctuations Θ_{Fluc}, which are related to rebate coefficient λ. In addition, real-world ESPs tend to be bounded by rational, which means that their decisions do not immediately lead to optimal solutions.

Algorithm 1 Simulation of ESP Selection for DER Participants

For DER Participant $j=1:J$:
  Obtain related coefficients $\xi_1, \xi_2, \cdots, \xi_7$ for DER j.
  For ESP $m=1:M$:
    Obtain related factors $\Theta_{\text{ReturnL}}, \Theta_{\text{ReturnS}}, \cdots, \Theta_{\text{Scale}}$ from prospectus published by ESP m.
    Generate random factor Θ_{Other}.
    Calculate potential Γ for ESP m selected by DER j.
  End
  Collect all potentials $[\Gamma_1, \Gamma_2, \cdots, \Gamma_M]$ with different ESPs for DER j.
  DER j determines the selected ESP.
End

Therefore, the rebate coefficient λ_c^m for DER participant for each ESP is defined as follows:

$$\lambda_c^m = \begin{cases} [\lambda_{c,PV}^m, \lambda_{c,battery}^m], if\ c>1 \\ \quad [0,0] \quad\quad , if\ c=1 \end{cases} \tag{3}$$

where $\lambda_{c,PV}^m$ and $\lambda_{c,battery}^m$ denote the rebate coefficients of the *m-th* VPP for OC c for PV and battery, respectively, and $\lambda_c^m \in [0,1]$. The rebate competition process involves several ESPs and time-varying variables, such as electricity price and solar power generation, leading to a more intense ESP rebate. Therefore, we use DRL to simulate the multi-ESP competition.

IV. System Models for VPP Profit and Allocation

The ESPs recruit participants by publishing a complete prospectus and assemble the VPP with DER participants so that the VPP plan can be considered an expression of the will of all participating members. This section provides generalized system models for VPPs participating in the energy market.

A. Profit Seeking and Risk Control

The ESP makes profits by predicting electricity prices, solar power generation, and battery dispatching to buy low and sell high for electricity. Conventional VPP profit-seeking models are given by (4) to (10). The profit $\varphi_{t,\omega}$ of the VPP for the energy market at time t with predicted scenario ω is calculated as follows:

$$\varphi_{t,\omega} = \rho_{t,\omega_\rho}^{RT} P_t^{VPP} \tag{4}$$

$$P_t^{VPP} = P_{t,\omega_{PV}}^{PV} - \mu_t^{ch} P_t^{ch} + \mu_t^{dch} P_t^{dch} \tag{5}$$

$$\mu_t^{ch} + \mu_t^{dch} \le 1 \tag{6}$$

$$P_{t,\omega_{PV}}^{PV}, P_t^{ch}, P_t^{dch} \ge 0 \tag{7}$$

where $\rho_{t,\omega_\rho}^{RT}$ represents the probabilistic real-time electricity price scenario, and P_t^{VPP} the output of VPP; $P_{t,\omega_{PV}}^{PV}$, P_t^{ch}, and P_t^{dch} denote the PV generation power, battery charging, and discharging power, respectively; P_t^{ch} and P_t^{dch} are less than the rated power capacity P^m. These two battery components are non-negative values and are limited by binary variables μ_t^{ch} and μ_t^{dch}. The battery state of charge (SoC) constraints are shown below:

$$SoC_t = SOC_{t-1} + \left(\mu_t^{ch} P_t^{ch} \eta^{ch} - \frac{\mu_t^{dch} P_t^{dch}}{\eta^{dch}}\right) \Delta t / E^m \tag{8}$$

$$SoC_{min} \le SoC_t \tag{9}$$

$$SoC_0 = SoC_T \tag{10}$$

where Δt denotes the time duration of each time interval, which is 30 minutes in this paper. η^{ch} and η^{dch} are the battery charging and discharging efficiencies, respectively. There is a lower limit regarding the battery SoC_t, which is set at 10%; the initial and final SoCs are assumed to be the same and equal to 20%, as in (10).

ESPs attract different DER owners by offering differentiated VPP strategies, including risk-neutral, risk-averse and risk-seeking. ESPs seek to maximize the expected profit with different scenarios ω by scheduling the battery operations Ξ, which can be expressed as:

$$\underset{\Xi}{Max} \sum_\omega \sum_t \pi_\omega \varphi_{t,\omega} \tag{11}$$

where π_ω denotes the probability of scenario ω. The strategy in (11) can consider all possibilities and is thus called the risk-neutral strategy. However, conservative DER participants prefer stable profits and avoid risk. The risk-averse VPP plan, which can manage its risks in the worst-case scenarios, can be described as the follows:

$$\underset{\Xi}{Max} (1-\beta_r) \sum_\omega \sum_t \pi_\omega \varphi_{t,\omega} + \beta_r \eta_r \tag{12}$$

where β_r denotes the risk-averse degree parameter, and a larger β_r indicates that the participant is more risk-averse. η_r represents the conditional value at risk, a risk-averse measurement [6]. Aggressive participants are willing to take higher risks to obtain higher profits, which is called the risk-seeking VPP plan and can be described as follows:

$$\underset{\Xi}{Max} (1-\beta_s) \sum_\omega \sum_t \pi_\omega \varphi_{t,\omega} + \beta_s \eta_s \tag{13}$$

where β_s denotes the risk-seeking degree parameter, and η_s represents the value at the best scenarios, a risk-seeking measurement [33].

The η_r and η_s involve a confidence parameter, which is set to 0.90. It is worth noting that the above three risk control methods are just examples for the following simulations in this paper, and the methods used for real-world VPP optimization scheduling are diverse.

B. Fair Profit Allocation

ESPs and DERs assemble VPPs in the recruitment–participation approach, where there is no profit allocation rule maker. ESPs and all DERs need to agree on an allocation method, and the best allocation method is based on their respective contributions [14]. Therefore, a fair profit distribution based on DER individual contribution instead of DER group is proposed, which contains two steps. The first step is profit allocation among ESPs and PV and battery groups with SV, and the second is the internal reallocation for individuals in PV and battery groups. Defining $X/\langle i\rangle\}$ to be the set of all parts after removing part i, the marginal contribution of part i to a coalition S, $S \subseteq X/\langle i\rangle$, is

$$\epsilon_i(S) = v(S \cup \{i\}) - v(S) \tag{14}$$

where $v(S)$ represents the total payoffs that the members of S can obtain by cooperation. The SV for each part in VPP can be calculated as follows:

$$\phi_i = \sum_{S\subseteq X/\langle i\rangle} \frac{|S|!(n-|S|-1)!}{n!}\epsilon_i(S) \tag{15}$$

where n represents the total number of parts and is set to 3 since the VPP is divided into three parts, PVs, batteries and ESP. The existing studies generally allocate profits to PV and battery groups rather than to real individual owners. To motivate DER individuals to participate in VPP, a reasonable internal group profit allocation method for PV and battery owners, δ_{PV-j} and $\delta_{battery-k}$, based on actual contributions, is proposed as follows:

$$\delta_{PV-j} = \phi_{PV}\frac{\sum_{t=1}^{T}\rho_t^{RT}P_{t,j}^{PV}}{\sum_{j=1}^{J}\sum_{t=1}^{T}\rho_t^{RT}P_{t,j}^{PV}} \tag{16}$$

$$\delta_{battery-k} = \phi_{battery}\frac{E_k}{\sum_{k=1}^{K}E_k} \tag{17}$$

$$\delta_{ESP} = \phi_{ESP} \tag{18}$$

where ρ_t^{RT} represents the wholesale real-time electricity price at time t; $P_{t,j}^{PV}$ is the available output of the *j-th* PV owner at time t, and J represents the total number of PV participants; E_k denotes the actual capacity of the *k-th* available for VPP, and K the total number of battery participants. For PV individuals, the payoff is related to the amount of generation available for VPP and the corresponding real-time electricity price, as in (16). In contrast, for batteries, the payoff is related to the capacity available for VPP, as in (17).

In addition to the SV-based profit allocation method with contributions, other allocation methods, such as investment shares - based, may lead to a fixed aggregation of VPP, which can impede the flexibility of assembly. Furthermore, non-contribution-based profit allocation can give rise to the emergence of profit distribution rule makers, which conflicts with the proposed non-ESP-centered VPP assembly.

C. DER Participants' Payoffs including Rebates

The ESP adopts the rebates to DER owners to reward them. The rebate coefficients are defined as $\lambda_{c,PV}$ and $\lambda_{c,battery}$ for PVs and batteries in OC c, respectively. The total payoff for individual PV and battery can be calculated as follows:

$$\zeta_{PV-j} = \delta_{PV-j} + \lambda_{c,PV}\delta_{ESP}^{PV-j} \tag{19}$$

$$\zeta_{battery-k} = \delta_{battery-k} + \lambda_{c,battery}\delta_{ESP}^{battery-k} \tag{20}$$

where ζ_{PV-j} and $\zeta_{battery-k}$ denote the final PV and battery payoffs, respectively, including the allocated payoff and rebate; δ_{ESP}^{PV-j} and $\delta_{ESP}^{battery-k}$ denote the ESP payoffs exploited from the *j-th* PV and *k-th* battery, respectively, which can be calculated with SV.

D. ESP's Payoff excluding Rebates

The proposed VPP assembly approach involves multi-ESPs. δ_{ESP-m} is proposed to represent the actual source of the total payoff of ESP before rebate competition, as the follows:

$$\delta_{ESP-m} = \sum_{j=1}^{J}\delta_{ESP-m}^{PV-j} + \sum_{k=1}^{K}\delta_{ESP-m}^{battery-k} \tag{21}$$

After the *m-th* ESP provides the rebate λ_c^m in OC c, the number of DER participants changes, and the total payoff of ESP changes to ζ_{ESP-m} as follows:

$$\zeta_{ESP-m} = \underbrace{\left(1-\lambda_{c,PV}^{m}\right)\sum_{j=1}^{J^{Rebate}}\delta_{ESP-m}^{PV-j}}_{\zeta_{ESP-m}^{PV}} + \underbrace{\left(1-\lambda_{c,battery}^{m}\right)\sum_{k=1}^{K^{Rebate}}\delta_{ESP-m}^{battery-k}}_{\zeta_{ESP-m}^{battery}} \tag{22}$$

where J^{Rebate} and K^{Rebate} represent the numbers of PV and battery participants after the rebate competition, respectively. The first part ζ_{ESP-m}^{PV} represents the payoffs exploited from PVs, and the second part $\zeta_{ESP-m}^{battery}$ represents the payoffs from batteries. J^{Rebate} is defined as the follows

$$J^{Rebate} = \mu\frac{\Gamma^m}{\sum_{m=1}^{M}\Gamma^m}N^{PV} = \frac{\Gamma^m}{\sum_{m=1}^{M}\Gamma^m}N^{PV}\left(-\frac{1}{\frac{\delta_{PV-j}+\lambda_{c,PV}^{m}\delta_{ESP}^{PV-j}-\Theta_{PV}^{w/o}}{\Theta_{PV}^{w/o}}*a^{PV}+b^{PV}}+1\right) \tag{23}$$

where K^{Rebate} has the same structure as J^{Rebate}.

The best rebate response function of each ESP can be drawn by taking the first-order derivative of ζ_{ESP-m} with respect to $\lambda_{c,PV}^{m}$ and $\lambda_{c,battery}^{m}$. Since $\zeta_{ESP-m}^{battery}$ has the same structure with ζ_{ESP-m}^{PV}, only the first-order derivative of ζ_{ESP-m}^{PV} is presented with respect to $\lambda_{c,PV}^{m}$ for brevity as follows:

$$\frac{\partial\zeta_{ESP-m}^{PV}}{\partial\lambda_{c,PV}^{m}} = \frac{\partial\left(\frac{\left(1-\lambda_{c,PV}^{m}\right)\Gamma^m N^{PV}\delta_{ESP-m}^{PV-j}}{\sum_{m=1}^{M}\Gamma^m}\left(-\frac{1}{\frac{\delta_{PV-j}+\lambda_{c,PV}^{m}\delta_{ESP}^{PV-j}-\Theta_{PV}^{w/o}}{\Theta_{PV}^{w/o}}*a^{PV}+b^{PV}}+1\right)\right)}{\partial\lambda_{c,PV}^{m}} \tag{24}$$

Let $A = \frac{\Gamma^m}{\sum_{m=1}^{M}\Gamma^m}$, $B = N^{PV}\delta_{ESP-m}^{PV-j}$, $C = a^{PV}\delta_{PV-j} + (b^{PV} - a^{PV})\Theta_{PV}^{w/o}$, $D = a^{PV}\delta_{ESP}^{PV-j}$, and $E = a^{PV}\delta_{PV-j} + (b^{PV} - a^{PV} - 1)\Theta_{PV}^{w/o}$. Equation (24) can be rewritten as follows:

$$\frac{\partial\zeta_{ESP-m}^{PV}}{\partial\lambda_{c,PV}^{m}} = AB\frac{(D\lambda_{c,PV}^{m})^2+2CD\lambda_{c,PV}^{m}+EC+ED-CD}{(D\lambda_{c,PV}^{m}+C)^2} = 0 \tag{25}$$

By setting the first-order derivative to zero, we can get the best rebate $\lambda_{c,PV}^{m}$ as follows:

$$(\lambda_{c,PV}^{m})^* = \frac{-\left(\begin{array}{c}a^{PV}\delta_{PV-j}+\\(b^{PV}-a^{PV})\Theta_{PV}^{w/o}\end{array}\right)+\sqrt{\Theta_{PV}^{w/o}\left(\begin{array}{c}a^{PV}\left(\delta_{PV-j}+\delta_{ESP}^{PV-j}-\Theta_{PV}^{w/o}\right)\\+b^{PV}\Theta_{PV}^{w/o}\end{array}\right)}}{a^{PV}\delta_{ESP}^{PV-j}} \tag{26}$$

Since $\delta_{PV-j} + \delta_{ESP}^{PV-j}$ denote the all ESP payoff, which must be larger than $\Theta_{PV}^{w/o}$, $a^{PV}\left(\delta_{PV-j} + \delta_{ESP}^{PV-j} - \Theta_{PV}^{w/o}\right) > 0$ and $b^{PV}\Theta_{PV}^{w/o} > 0$ can be readily obtained, which means $(\lambda_{c,PV}^{m})^*$

must exist. Since $\zeta_{ESP-m}^{battery}$ has the same structure as ζ_{ESP-m}^{PV}, $(\lambda_{c,battery}^{m})^{*}$ must exist too.

V. Rebate Competition with MADDPG Algorithm

A. Markov Game Model for Multi-ESP Competition

In a multi-agent environment, the rebate decision making by each ESP is influenced by the joint actions of all the ESPs. A Markov decision process (MDP) model [27, 34, 35] for I ESPs can be formalized as the follows:

$$\tau =< \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P} > \quad (27)$$

where $\mathcal{S}$, $\mathcal{A}$, $\mathcal{R}$ and $\mathcal{P}$ denote the state space, action space, reward space, and transition probability, respectively. At time t, ESP m observes a state $s_m^t \in \mathcal{S}$ and choose an action $a_m^t \in \mathcal{A}$ based on a certain policy, which produces a new state s_m^{t+1}. The transition probability from $s_m^t, a_1^t, a_2^t, \cdots, a_I^t$ to s_m^{t+1} can be represented as $\mathcal{P}(s_m^{t+1} | s_m^t, a_1^t, a_2^t, \cdots, a_I^t)$. In this work, $\mathcal{S}$ contains the numbers and returns of DER participants for each ESP, VPP profits for each ESP, historical rebate coefficients of each ESP, historical VPP size of each ESP, other public information can be obtained from prospectus. In addition, historical half-hourly electricity prices are included in $\mathcal{S}$. $\mathcal{A}$ represents rebate coefficients $\lambda_{t,PV}^{m}$ and $\lambda_{t,battery}^{m}$ for each ESP at time t. For our work, the objective for each ESP is to maximize the expected accumulated payoffs by finding the optimal policy. The payoff model for each model in (21) can be re-formulated as:

$$Max \sum_{t\in T} \gamma^{T-t} \zeta_{ESP-m_t}(\lambda_{t,PV}^{m}, \lambda_{t,battery}^{m}) \quad (28)$$

where γ represents the reward discount factor, T represents the total number of OC.

B. Multi-Agent Deep Deterministic Policy Gradient

In the MADDPG algorithm, multiple agents are involved, and the centralized training and distributed implementation mode [36] are used. For agents, ESPs in our work, the objective of reinforcement learning to maximize the expected reward with policy gradient algorithm by adjusting the parameter θ of policy $\boldsymbol{\pi}$ at the direction of $\nabla_\theta J(\theta)$, as:

$$\nabla_\theta J(\theta) = \mathbb{E}_{s\sim\rho^\pi, a\sim\pi_\theta}[\nabla_\theta log\pi_\theta(a|s) Q^\pi(s,a)] \quad (29)$$

where the policy π_θ is stochastic, which may cause large computational consumption. In [35], authors proposed a deterministic policy gradient, which considers a deterministic policy $\boldsymbol{\mu}_\theta: \mathcal{S} \rightarrow \mathcal{A}$, to integrate over the state space. The deterministic policy gradient can be derived as:

$$\nabla_\theta J(\theta) = \mathbb{E}_{s\sim p^\mu}[\nabla_\theta \mu_\theta(s) \nabla_a Q^\mu(s,a)|_{a=\mu_\theta(s)}] \quad (30)$$

Compared with policy gradient, the deterministic policy gradient can reduce the computational complexity and improve the learning efficiency. For a game with I agents, $\boldsymbol{\mu}_\theta = \{\mu_{\theta_1}, \mu_{\theta_2}, \cdots, \mu_{\theta_I}\}$ is parameterized by $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots, \theta_L\}$, which is a deterministic policy set for all agents. The gradient of the expected return for each agent can be obtained by:

$$\nabla_{\theta_m} J(\mu_{\theta_m}) = \mathbb{E}_{\boldsymbol{o},a\sim\mathcal{D}}\left[\nabla_{\theta_m}\mu_{\theta_m}(a_m|s_m)\nabla_{a_m}Q_m^{\boldsymbol{\mu}}(\boldsymbol{o}, a_1, a_2, \cdots, a_I)|_{a_m=\mu_{\theta_m}(s_m)}\right] \quad (31)$$

where $\boldsymbol{o}$ denotes the observation information, which generally includes the states of all agents; $Q_m^{\boldsymbol{\mu}}(\boldsymbol{o}, a_1, a_2, \cdots, a_I)$ denotes a centralized action-value function; $\mathcal{D}$ denotes the experience replay buffer, which stores the experience of all agents, expressed as $(\boldsymbol{o}, \boldsymbol{o}', a_1, a_2, \cdots, a_I, r_1, r_2, \cdots, r_I)$. $\mathcal{D}$ can be transmitted to the critic network for centralized updating with the back-propagation approach by minimizing the loss function:

$$L(\theta_m) = \mathbb{E}_{\boldsymbol{o},\boldsymbol{o}',a,r}[(Q_m^{\boldsymbol{\mu}}(\boldsymbol{o}, a_1, a_2, \cdots, a_I) - y)^2] \quad (32)$$

$$y = r_m + \gamma Q_m^{\boldsymbol{\mu}'}(\boldsymbol{o}', a'_1, a'_2, \cdots, a'_I)|_{a'_m=\boldsymbol{\mu}'(o'_m)} \quad (33)$$

The parameters θ_m^μ and θ_m^Q are manipulated by the soft-update approach as below:

$$\theta_m^{\mu\prime} = \varrho\theta_m^\mu + (1-\varrho)\theta_m^{\mu\prime} \quad (34)$$

$$\theta_m^{Q\prime} = \varrho\theta_m^Q + (1-\varrho)\,\theta_m^{Q\prime} \quad (35)$$

C. MADDPG-based Multi-ESPs Rebate Competition

The MADDPG-based multi-ESPs rebate competition with the key models from Section III is given in Algorithm 2.

Algorithm 2 Simulation of Multi-ESP Competition with MADDPG

Input: Structures of actor $\boldsymbol{\mu}$, critic $\boldsymbol{Q}$ networks, and initialized weights $\boldsymbol{\theta^\mu}$ and $\boldsymbol{\theta^Q}$.

Output: Actor networks' weights $\boldsymbol{\theta}_m^{\boldsymbol{\mu}\prime}$ for each ESP

```
For each episode:
    For each training step:
        For each OC:
            For each ESP agent m:
                Actor of each ESP agent generates rebate action for current OC: a_m^ESP=[λ_PV^m, λ_battery^m].
            End
            ESPs publish prospectus which contains all rebate actions a^ESP = [a_1^ESP, a_2^ESP, ..., a_I^ESP].
            DERs select ESPs using (1-2) and Algorithm 1. Each ESP obtain actual total numbers of PV and battery participants.
            For each day in OC:
                Obtain predicted PV generation and electricity price with the CTSGAN in [37].
                VPPs operated by different ESPs with different risk control strategies using (11)-(13).
                Payoffs allocation for each DERs and ESPs using (19)-(21) for one day.
            End
            ESPs obtain the reward R for OC using (22) and new states o_new.
            Store transition sample (o, a, R, o_new) into D.
            Sample a random minibatch of samples (o^d, a^d, R^d, o_new^d) from D.
            Set y^d = R^d + γ Q_m^μ'(o_new^d, a'_1, a'_2, ..., a'_I)|_{a'_m=μ'(o_m)}.
            Update critic by minimizing the loss function:
                L(θ_m) = 1/N Σ_{d=1}^N (y^d − Q_m^μ(o_new^d, a_1, a_2, ..., a_I)^2)
            Update actor network using the sampled policy gradient:
                ∇_{θ_m} J = 1/N Σ_{d=1}^N (∇_{θ_m} μ_{θ_m}(o_m^d) ∇_{a_m} Q_m^μ(o^d, a_1, a_2, ..., a_I)|_{a_m=μ_{θ_m}(o_m^d)})
            Update target network parameters for each agent m:
                θ_m^μ' = ϱθ_m^μ + (1 − ϱ)θ_m^μ'
                θ_m^Q' = ϱθ_m^Q + (1 − ϱ) θ_m^Q'
        End
    End
End
```

VI. Case Study

A. Experiment Settings

It is assumed that there are 5,600 households in an area, of which 4,200 are installed with PVs and 1,400 with batteries. In addition, 35 experienced ESPs are involved in this area and wish to assemble VPPs for participation in the energy market. ESPs publish prospectus for the relevant plans. The PV generation data are from Ausgrid (a major Australian power company) [38], and the energy market electricity prices for VPP participation are from the Australian Energy Market Operator (AEMO) from July 2021 to June 2022 [39]. Table II lists the parameters for the 2,800 household batteries.

TABLE II PARAMETERS OF BATTERIES

E^m	η^{ch}	η^{dch}	P^m
7kWh	0.95	0.95	3.5kW

In this section, several cases are studied to simulate the proposed approach for VPP assembly, disassembly, and competition as follows:

VPPs Assembly. It is assumed that only 20% of DER initially participate in assembling VPPs. Table. III assumes seven common risk control strategies, including three risk-averse, three risk-seeking, and one risk-neutral, the corresponding risk degree parameters β_r and β_s, and the number of ESPs, PVs and batteries. The first 10 OCs are set as closed periods, and the PV and battery participants cannot exit.

TABLE III INITIAL VPP ASSEMBLY SETTINGS

	Risk-Averse			Risk-Neutral	Risk-Seeking		
Risk Degree	0.3	0.2	0.1	N.A.	0.1	0.2	0.3
ESP Number	5	5	5	5	5	5	5
PV Number	120	120	120	120	120	120	120
Battery Number	40	40	40	40	40	40	40

VPPs Competition. ESP attract more DER participants through rebates. The willingness of PV and battery owners to participate in VPP is based on (1). The involved coefficients a^{PV}, b^{PV}, $a^{battery}$, and $b^{battery}$ are assumed as 20, 1.25, 10, and 1.25, respectively [17]. The coefficients of the participants' preferences are assumed as listed in Table IV. In addition, we assume a standard normal distribution for all types of DERs, with the proportions of professional, ambitious, moderate, conservative, and cautious participants in order of 7%, 24%, 38%, 24%, and 7%.

TABLE IV COEFFICIENTS OF PARTICIPANT GROUPS

Types	$[\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6]$
Professional	[0.30, 0.25, 0.15, 0.15, 0.10, 0.05]
Ambitious	[0.10, 0.50, 0.15, 0.10, 0.10, 0.05]
Moderate	[0.30, 0.15, 0.20, 0.20, 0.10, 0.05]
Conservative	[0.25, 0.10, 0.25, 0.25, 0.10, 0.05]
Cautious	[0.25, 0.05, 0.30, 0.25, 0.10, 0.05]

It is noted that the coefficients are assumed based on the description in Table 1, which may not reflect the actual situation. Even for the same investment preference, the weights may be influenced by factors such as social development, education level, and local laws and regulations, leading to changes in the weights. Therefore, to ensure the effectiveness of the VPP competition model presented in this work, the relevant coefficients should be well designed based on actual data before applying the model.

In this work, two base cases are employed for VPP assembly stage (Case 1) and VPP competition stage (Case 2).

Case 1: VPP Assembly

Case 1.a: ESPs initially assemble VPPs with recruited DER owners. VPPs with poor strategies will be disassembled.

Case 1.b: VPP profits are allocated to each DER participants.

Case 1.c: ESPs earn payoffs by helping DER participants increase their payoffs.

Case 2: VPP Competition

Case 2.a: ESPs provide rebates to attract DER participants, and the final state of competition can be stable with MADDPG.

Case 2.b: Changes in rebate competition with large electricity prices fluctuations.

Case 2.c: Changes in rebate competition with different participation related coefficients.

B. Simulation Results

1) VPP Initial Assembly

a) Comparison of VPP Profits with Multi-ESPs

Thirty-five ESPs assemble 35 VPPs with 840 PVs and 280 batteries with multiple risk degree levels. The first 10 OC periods, from July 5 to September 12, 2021, is set as closed period. The multi-ESPs first use day-ahead prediction of electricity prices and PV generation and then apply an optimization algorithm to optimize the battery scheduling for profit maximization based on the chosen risk level. For simplicity, the prediction method used is conditional time series generative adversarial network (CTSGAN) [37], and the optimization method is the particle swarm optimization (PSO) algorithm. We set different hyperparameters for CTSGAN and PSO to simulate the performance of different ESPs. Fig. 3 illustrates the average profit, nominal risk degree level, and actual profit fluctuations for each ESP in 10 OCs. The sign A$ on the coordinate axis represents the Australian dollar. Dots of the same color represent the same risk degree level.

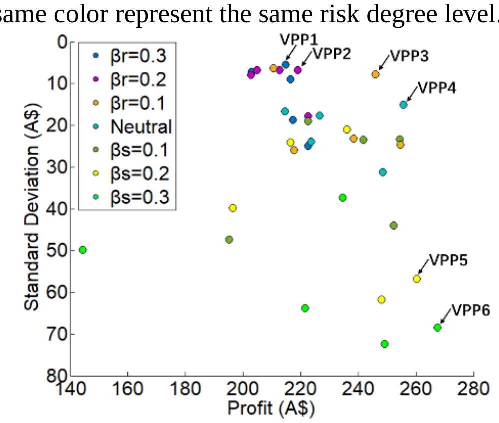

Fig. 3 Profit and standard deviation for 35 VPPs

One can find that all ESPs have different performances; only 6 VPPs lie on the Pareto frontier of standard deviation and total profit, including three risk-averse VPPs with β_r=0.3 (VPP1 in Fig. 3), β_r=0.2 (VPP2), and β_r=0.1 (VPP3), one risk-neutral VPP (VPP4), and two risk-seeking VPPs with β_s=0.2 (VPP5) and β_r=0.3 (VPP6), respectively. The average daily profits for risk-averse VPPs are approximately between A$215 and A$245 and are stable with a standard deviation of less than A$7. The profits for risk-seeking VPPs are slightly higher and more volatile, with a large standard deviation of more than A$55. The remaining VPPs have poor performance, which means that their DERs will choose other ESPs at the end of the closed period, which leads to the disassembly of the VPP. Note that the performance of the five risk-seeking VPPs with β_s =0.1 simulated in this paper is poor, but this does not mean that

β_s=0.1 is inappropriate for any VPP.

b) Comparison of VPP Profits Allocation for Each Participant

Taking VPP1-VPP6 in Fig. 3 as an example, the total profit is allocated to ESPs and PV and battery groups, using SV based on the contribution. Fig. 4 illustrates the average payoff for ESP, PV and battery groups. The PV group's payoff is stable at around A\$138, while the payoff of the battery group increases from A\$61.09 to A\$89.93 as the risk increases. The ESP payoff is similar to that of the battery group, meaning that the payoff comes mainly from the arbitrage of battery charging and discharging.

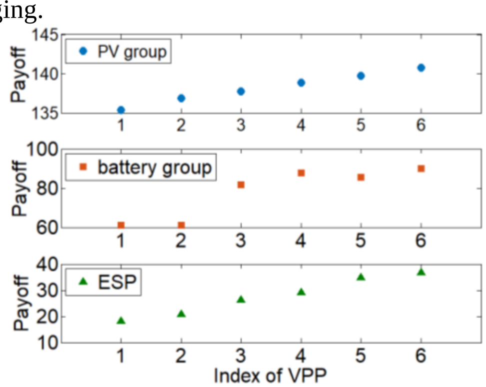

Fig. 4 Average payoff for ESP, PV and battery groups based on SV

TABLE V AVERAGE DAILY PAYOFF FOR EACH PARTICIPANT (A\$/DAY)

	VPP1	VPP2	VPP3	VPP4	VPP5	VPP6	Retailer -led
PV	1.128	1.140	1.156	1.157	1.173	1.189	1.080
Battery	1.527	1.530	2.043	2.191	2.142	2.249	1.456
ESP	18.221	20.828	26.403	29.094	34.999	36.809	NA.

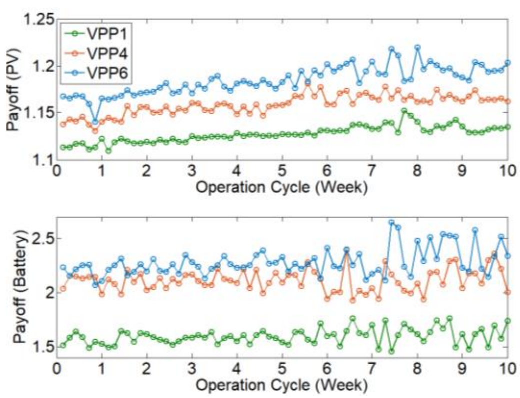

Fig. 5 Payoff for PV and battery individuals for VPP1, VPP4 and VPP 6

Based on the above DER group payoff, the payoff allocation of the individuals in PV and battery groups can be calculated based on (16)-(17). Table V provides the average daily payoffs for PV and battery participants in the first ten cycles with the proposed VPP assembly model and those of current Australian existing retailer-led VPPs [40]. It can be seen that the daily payoff of individual PV participants is stable between A\$1.12 and A\$1.19, while battery payoff varies from A\$1.5 to A\$2.2 for different VPPs. In addition, the payoff of individual PV is improved by at least 4.44% (VPP1) and at most 10.09% (VPP 6) compared to participation in the existing VPP. The battery participant's payoff increment is more obvious, increasing by 54.46% for VPP6. The payoff received by ESPs according to their contributions differs widely from A\$18.221 to A\$36.809. Fig. 5 presents the actual payoff of PV and battery participants for 70 consecutive days. It can be seen that the daily payoff for PV fluctuates very little, even for the risk-seeking VPP, such as VPP 6, with fluctuations ranging between A\$1.18 and A\$1.22 in the 7th OC. In contrast, battery payoff is significantly more volatile for the same VPP 6 and OC, ranging from A\$2.11 to A\$2.65.

c) Payoff Sources for VPP ESPs

Although ESPs do not make money directly by generating electricity or providing energy storage, ESPs get returns through dispatch, which are coming from each PV and battery participant. Table VI lists the average daily payoff for ESP earned from each PV and battery individual for the first 10 OCs. One can find that the payoffs for ESPs provided by PVs are small and stable, with less than A\$0.05 per PV per day. Payoffs from batteries account for a significant portion of ESPs and increase significantly with the increase in total VPP profits, from A\$0.372 to A\$0.833 per battery per day.

TABLE VI ESPS' PAYOFFS EARNED FROM EACH PV AND BATTERY (A\$/DAY,10E-2)

	VPP1	VPP2	VPP3	VPP4	VPP5	VPP6
PV	4.764	4.794	5.828	5.831	5.869	5.910
Battery	37.273	43.700	57.524	64.245	78.891	83.293

2) VPP Rebate Competition

By the end of the closed period, all rational participants will leave those ESPs with high volatility and low profit, resulting in the majority of the total VPPs being disassembled and only 6 VPPs remaining. This section will explore the rebate competition process within the remaining VPPs as examples. Although new ESPs or eliminated ESPs can reassemble VPPs, we focus on simulating rebate competition and therefore do not consider the case of assembling new VPPs.

a) Simulation for VPP Rebate Competition

The MADDPG method is used to simulate the rebate competitive process for 20 OCs between September 13, 2021 and January 31, 2022. Fig. 6 presents the training process of rebate competition for each ESP.

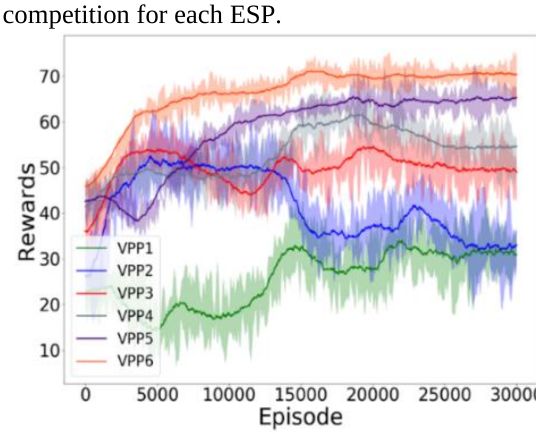

Fig. 6 Training process of the rebate competition

TABLE VII ESPS REBATE COMPETITION RESULTS (%)

	VPP1	VPP2	VPP3	VPP4	VPP5	VPP6
λ_{PV}	26.96	25.90	23.28	18.14	17.68	13.42
$\lambda_{battery}$	38.01	36.64	34.75	32.46	26.50	24.85

One can find that the ESP of VPP6 is the first to reach a stable payoff, then followed by the ESPs for VPP5, VPP4, and VPP3. The competition between ESP1 and ESP2 is more intense; therefore, the reward does not stabilize until around episode 27,000. Table VII tabulates the rebate competition results. One can find that ESPs give lower rebates to PVs than batteries, partly because PVs bring lower profits to ESPs, and partly because PV participants are more likely to be incentivized to participate in VPPs by low rebates. In addition, the high-risk-return VPP gives a low rebate coefficient, with values of 13.42% (λ_{PV}^{6}) and 24.85% ($\lambda_{battery}^{6}$) for VPP6, while those of the low-risk ones are high, with values of 36.64% and 38.01% for battery groups in VPP2 and VPP1. The reason is mainly the higher profitability of ESP for VPP6, where a small rebate coefficient can lead to a large actual rebate. Therefore, due to the increased risk and unstable profitability, providing a large rebate coefficient would result in extremely low profitability for ESP in VPP6 in specific cases.

With the rebates provided by ESPs, DER participants can re-select the VPP operation before each cycle since the change of return, payoff stability, and VPP scale with (2). In addition, the total number of DERs increases due to payoff increment obtained from rebates. Fig. 7 shows the change in PV and battery participants for each VPP. The number of participants in each VPP increases. The average number of PV owners who participate VPPs with rebates is 1.23 times higher than that without rebates; in contrast, the number of battery participants with rebates is 2.11 times higher than that without rebates, from 46.83 to 98.23.

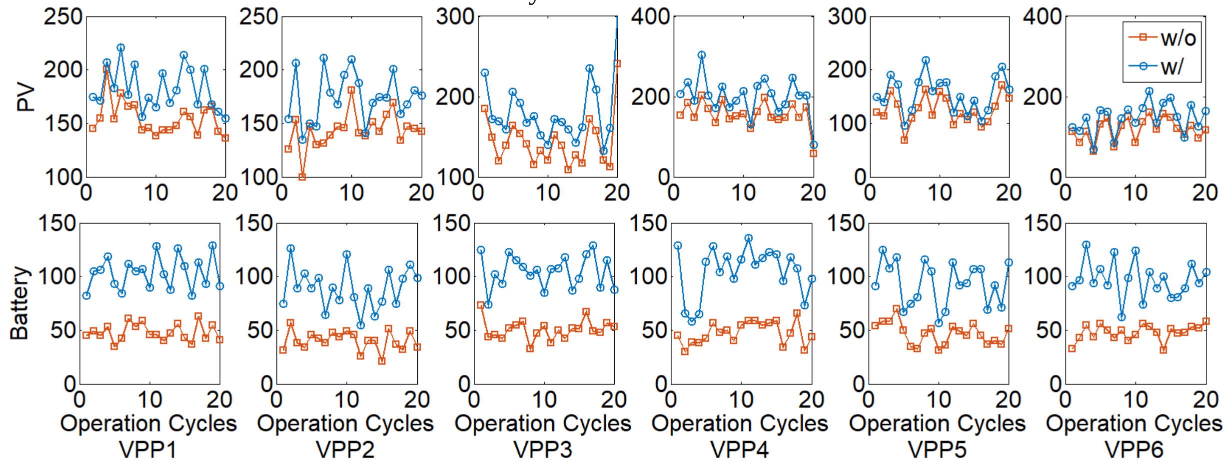

Fig. 7 Change in the number of PV and battery participants for 20 OCs

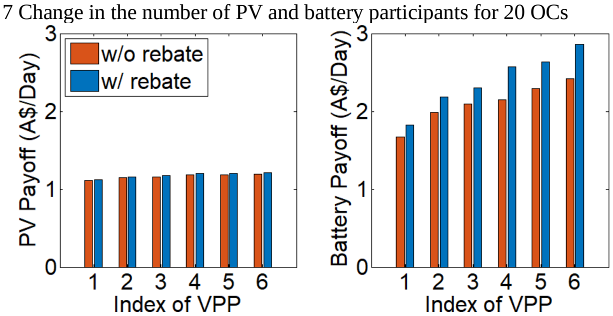

Fig. 8 Daily payoff for PV and battery participants

TABLE VIII COMPARISON OF ESPS' DAILY PAYOFF WITH AND WITHOUT REBATE COMPETITION (A$/DAY)

	ESP1	ESP2	ESP3	ESP4	ESP5	ESP6
W/O Rebate	26.861	26.619	36.377	40.196	45.989	49.629
W/ Rebate	31.451	32.774	46.737	53.870	63.283	71.491
Improve	0.171	0.231	0.285	0.340	0.376	0.440

Fig. 8 presents the daily payoffs for DERs for the 20 OCs with or without rebates. It can be seen that the payoffs of PV participants rise slightly with rebates compared to those without rebates, by approximately 0.62% to 1.79%. In contrast, the payoffs of battery participants increase more significantly, with a minimum of 9.28% in VPP1, from A$1.672 to A$1.826, and a maximum of 18.25% in VPP6, from A$2.428 to A$2.854. Table VIII lists the average payoff for ESPs. Although ESPs offer a portion of their payoffs in the form of rebates to attract more participants, each ESP's daily payoff is increased significantly. The low-risk VPP1, with a small rebate, attracts only a limited number of participants, so that the daily payoff increases from A$26.861 to A$31.451. In contrast, the high-risk VPP6 sees a more significant increment in daily payoff, with an improvement of 44.0% from A$49.629 to A$71.491.

b) Impact of Electricity Price Fluctuations on Rebate Competition

Price fluctuations have a direct relationship with VPP profitability. In this section, the prices between April 4 and June 12, 2022, are used to study the impact of electricity price fluctuations on the multi-ESP VPP competition. During this period, significant unplanned outages, challenges in the coal and natural gas supply chains, and the early onset of winter have combined to cause significant volatility in electricity prices [41], leading AEMO to temporarily suspend Wholesale Market on June 15 [42]. Table IX lists the daily profits of the participants before the rebate competition and the final rebate coefficients for each VPP. One can find that fluctuating electricity prices can bring higher payoffs to DER participants than those with normal electricity prices, where the payoff increment of batteries is significantly higher than that of PVs. For example, in VPP6, the average payoff for the battery individual increases from A$2.249 (normal) to A$6.392 (fluctuating), while that for PV individuals increases from A$1.189 (normal) to A$1.272 (fluctuating). In addition, there is no significant change in the battery rebate coefficients given by the high-risk-return ESPs compared to those with normal electricity prices ($\lambda_{PV}, \lambda_{battery}$ in Table VII), while that given by VPP1 increases significantly, from 38.01% to 52.16%. The ESP selection of participants with different preferences also changes with price fluctuation. Fig. 9 presents the changes in the number of participants with different preferences for normal and fluctuating electricity prices.

TABLE IX PARTICIPANT'S AVERAGE DAILY PAYOFF (A$/DAY) AND ESP'S REBATE COEFFICIENTS (%)

	VPP1	VPP2	VPP3	VPP4	VPP5	VPP6

PV	1.195	1.217	1.233	1.249	1.257	1.272
Battery	3.621	4.033	4.165	4.630	5.031	6.392
λ_{PV}	27.93	24.91	23.04	19.70	16.92	13.53
$\lambda_{battery}$	52.16	43.84	36.25	33.96	25.83	26.73

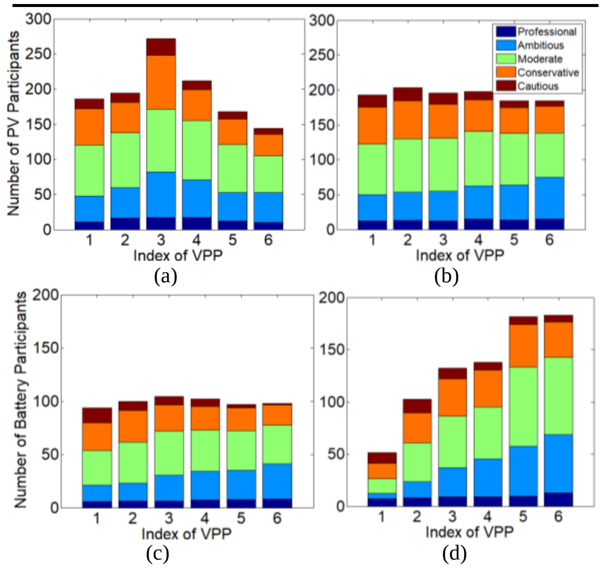

Fig. 9 Changes of operation selection for different participant groups for normal (a, c) and fluctuating (b, d) electricity price

For PV participants, it can be seen that with normal prices, ESP3 recruits the highest number of participants and is about 100 more than ESP6. When the prices are volatile, ESP5 and ESP6 have significantly more participants, around 180. Ambitious participants prefer ESP3 with the normal electricity prices, while ambitious ones are more inclined to ESP6 with high fluctuating prices. In addition, the change in the number of battery participants is more pronounced. The number of battery participants for each VPP is consistent with the normal electricity prices; however, that of VPP1 participants decreases significantly to 50 with the fluctuating prices, while the numbers of VPP5 and VPP6 participants are almost doubled. Conservative and cautious participants, driven by high returns, see a significant increase in participation in VPP5 and VPP6.

c) Impact of VPP Participation Rate Related Coefficients on Rebate Competition

In the above section, the VPP participation rate related coefficients a^{PV}, $a^{battery}$, b^{PV} and $b^{battery}$ are assumed to be 20, 10, 1.25, and 1.25 respectively. As there is currently a lack of literature describing the accurate relationship between VPP participation rates and payoff improvement, to further explore the response of different DER owners to payoff improvement, we assume the participation rates related coefficient group as shown in Table X, which includes different $b^{PV}, b^{battery}$ ranging from 1.11 to 1.43, a^{PV} ranging from 10 to 40, and $a^{battery}$ ranging from 5 to 20.

TABLE. X ASSUMED COEFFICIENTS IN THE RELATIONSHIP BETWEEN VPP PARTICIPATION RATE AND PAYOFF INCREMENT

Coefficients Group	1	2	3	4	5	6	7	8	9
Initial Participation		10%			20%			30%	
$b^{PV}, b^{battery}$		1.11			1.25			1.43	
a^{PV}	10	20	40	10	20	40	10	20	40
$a^{battery}$	5	10	20	5	10	20	5	10	20

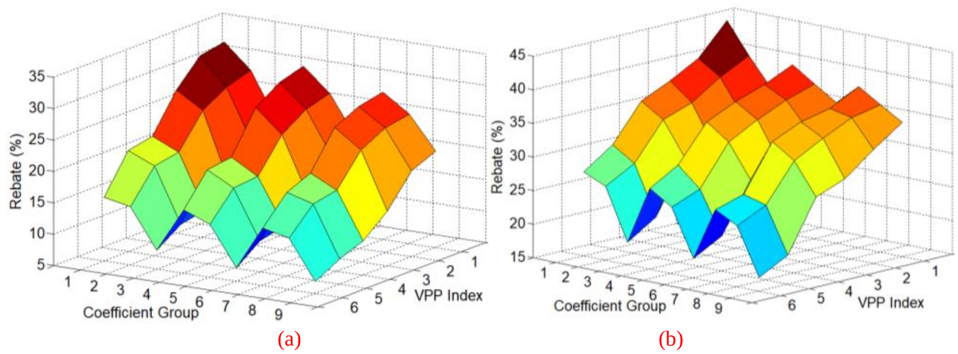

Fig. 10 Changes of rebate for PV (a) and battery (b) owners with assumed participation rate related coefficient group

Fig. 10 provides the changes of rebate coefficients for PV and battery with assumed coefficient groups in Table X using MADDPG. It can be observed that in (a), with the same initial participation rate, the rebate coefficients with larger a^{PV} values significantly decrease compared to those with smaller a^{PV} values. For instance, in VPP1, λ_{PV} is 32.48% (coefficient group 1), 28.73% (coefficient group 2), and 21.21% (coefficient group 3). This phenomenon can be observed in different VPPs and also in other coefficient groups, such as groups 4, 5, and 6. Moreover, it is found that for the same VPP, when the initial participation rate is high (groups 7, 8, and 9 with high b^{PV}), the difference in rebate coefficients is smaller than that when the b^{PV} is low, with 28.45% (group 7) and 20.34% (group 9) compared to 32.48% (group 1) and 21.21% (group 3). Due to the low initial participation rate and the high remaining number of potential participants, ESP needs to increase the rebate coefficients to attract DER owners to participate, thereby increasing the number of participants and profits. These phenomena also exist in VPP competition for battery rebate. Compared with PV rebate competition, the fluctuation of battery rebates is more affected by VPP participation-related coefficients. For instance, in VPP1 with group 1, $\lambda_{battery}$ is 44.82%, while in VPP6, group 9, the $\lambda_{battery}$ is only 17.04%. Fig. 11 presents the average daily payoffs for ESPs with the above λ_{PV} and $\lambda_{battery}$.

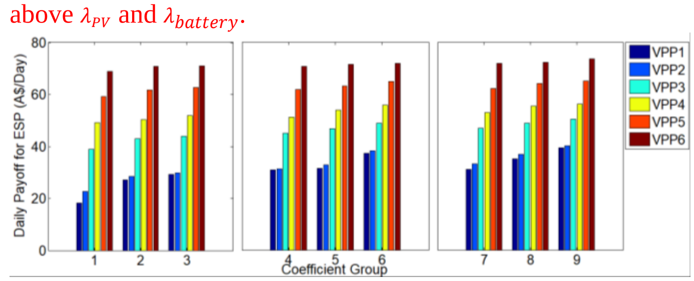

Fig. 11 Average daily payoffs for ESPs with assumed participation rate related coefficient group

Compared to the payoffs for 6 ESPs with group 6 (the payoffs in Table XIII), when the ESPs face DER owners who are easily incentivized (group 7 with higher a^{PV} and $a^{battery}$), the payoffs of each ESP increase, but the payoff for ESP1 increases the most, from A$31.45 to A$37.32, while that for ESP6 increases the least, just from A$71.49 to A$71.98. In addition, it can be seen that ESP6 is consistently able to maintain a higher and more consistent payoff when facing DER owners with coefficient group 1 to 9, ranging from A$68.93 to A$73.66, while the return of ESP1 fluctuates greatly from A$18.21 to A$39.42. This means that ESP6 is always in a proactive position of concession in VPP competition, while ESP1 often needs to be forced to make concessions.

VII. Conclusion

In this paper, the ESPs were separated for the first time from the existing VPP models, and their payoffs can be allocated fairly based on their contributions to the whole VPP with SV. The recruitment–participation approach for VPP assembly was proposed. The ESPs that meet DER preferences can be remained, while the ESPs that DERs are not satisfied with will be eliminated. Moreover, our work introduced multiple ESPs into VPPs for the first time with reference to the existing multi-retailer competition model in the electricity market. Finally, the rebate competition of multi-ESPs was simulated for these DERs with diverse preferences and different electricity price fluctuations.

The case studies demonstrated that the payoffs of ESPs were mainly from the battery individual rather than the PV, and the payoffs of DERs were significantly higher compared to those with existing VPPs in Australia. Thus, DERs had better motivations to participate in VPPs.

Rebate competition was also simulated. The following conclusions are drawn: 1) high risk-return ESPs can provide smaller rebate coefficients than low risk-return ones; 2) ESPs can offer higher rebate coefficients to batteries compared to PVs; 3) both ESPs and DERs can improve payoffs after rebates; 4) DERs with different investment preferences has shown a larger difference under different scenarios of electricity price fluctuations. It is noted that the coefficients applied in investment preference may be influenced by factors such as social development, education level, and local regulations, leading to changes in the weights. Current research in relevant fields generally assumes that participants can be incentivized through an increase in profits, without conducting empirical investigations into the attitudes and consumption habits of actual participants, limiting the transferability of these models to real-world applications. Future works can adopt a human-centered approach by utilizing advanced research methods, such as online ethnography, to gain a better understanding of participants' perspectives and preferences, and identify the factors beyond pricing that can motivate individuals to participate in VPPs.

References

[1] Z. Xu, Y. Guo, and H. Sun, "Competitive Pricing Game of Virtual Power Plants: Models, Strategies, and Equilibria," *IEEE Transactions on Smart Grid*, 2022.

[2] W. Chen, J. Qiu, J. Zhao, Q. Chai, and Z. Y. Dong, "Bargaining game-based profit allocation of virtual power plant in frequency regulation market considering battery cycle life," *IEEE Transactions on Smart Grid,* vol. 12, no. 4, pp. 2913-2928, 2021.

[3] K. N. Singh, N. Das, M. Maurya, A. K. Goswami, and N. B. D. Chudhury, "An Intelligent Bidding Strategy based on Social Welfare of Virtual Power Plant considering Carbon Trading." pp. 1-5.

[4] F.-f. Yang, and X.-g. Zhao, "Policies and economic efficiency of China's distributed photovoltaic and energy storage industry," *Energy,* vol. 154, pp. 221-230, 2018.

[5] C. A. Hill, M. C. Such, D. Chen, J. Gonzalez, and W. M. Grady, "Battery energy storage for enabling integration of distributed solar power generation," *IEEE Transactions on smart grid,* vol. 3, no. 2, pp. 850-857, 2012.

[6] H. Liu, J. Qiu, and J. Zhao, "A data-driven scheduling model of virtual power plant using Wasserstein distributionally robust optimization," *International Journal of Electrical Power & Energy Systems,* vol. 137, pp. 107801, 2022.

[7] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, and G. Strbac, "Deep reinforcement learning for strategic bidding in electricity markets," *IEEE Transactions on Smart Grid,* vol. 11, no. 2, pp. 1343-1355, 2019.

[8] A. Zangeneh, A. Shayegan-Rad, and F. Nazari, "Multi-leader–follower game theory for modelling interaction between virtual power plants and distribution company," *IET generation, transmission & distribution,* vol. 12, no. 21, pp. 5747-5752, 2018.

[9] D. T. Nguyen, H. T. Nguyen, and L. B. Le, "Dynamic pricing design for demand response integration in power distribution networks," *IEEE Transactions on power systems,* vol. 31, no. 5, pp. 3457-3472, 2016.

[10] M. B. Rasheed, M. A. Qureshi, N. Javaid, and T. Alquthami, "Dynamic pricing mechanism with the integration of renewable energy source in smart grid," *IEEE Access,* vol. 8, pp. 16876-16892, 2020.

[11] S. Yin, Q. Ai, J. Li, D. Li, and Q. Guo, "Trading mode design for a virtual power plant based on main-side consortium blockchains," *Applied Energy,* vol. 325, pp. 119932, 2022.

[12] F. Fang, S. Yu, and M. Liu, "An improved Shapley value-based profit allocation method for CHP-VPP," *Energy,* vol. 213, pp. 118805, 2020.

[13] S. Tian, L. Ying, G. Zhu, and X. Xie, "Study on the Profit Distribution Under the Participation of Virtual Power Plants in Market Bidding." pp. 198-202.

[14] S. S. Gougheri, M. Dehghani, A. Nikoofard, H. Jahangir, and M. A. Golkar, "Economic assessment of multi-operator virtual power plants in electricity market: A game theory-based approach," *Sustainable Energy Technologies and Assessments,* vol. 53, pp. 102733, 2022.

[15] W. Chen, J. Qiu, and Q. Chai, "Customized Critical Peak Rebate Pricing Mechanism for Virtual Power Plants," *IEEE Transactions on Sustainable Energy,* vol. 12, no. 4, pp. 2169-2183, 2021.

[16] W. Chen, J. Qiu, J. Zhao, Q. Chai, and Z. Y. Dong, "Customized Rebate Pricing Mechanism for Virtual Power Plants Using a Hierarchical Game and Reinforcement Learning Approach," *IEEE Transactions on Smart Grid*, 2022.

[17] C.-W. Hsu, "Using a system dynamics model to assess the effects of capital subsidies and feed-in tariffs on solar PV installations," *Applied energy,* vol. 100, pp. 205-217, 2012.

[18] J. Hu, W. Ma, Y. Xue, L. Yao, and D. Xie, "Quantification of Reserve Capacity Provided by Electric Vehicle Aggregator Based on Framework of Cyber-Physical-Social System in Energy," *Automation of Electric Power System,* vol. 46, no. 18, pp. 46-54, 2022.

[19] E. F. Bompard, G. Abrate, R. Napoli, and B. Wan, "Multi-agent models for consumer choice and retailer strategies in the competitive electricity market," *International Journal of Emerging Electric Power Systems,* vol. 8, no. 2, 2007.

[20] L. Zheng, "Is money smart? A study of mutual fund investors' fund selection ability," *the Journal of Finance,* vol. 54, no. 3, pp. 901-933, 1999.

[21] V. S. Mutswenje, "A survey of the factors influencing investment decisions: the case of individual investors at the NSE," University of Nairobi, 2009.

[22] S. Dhar, S. K. Salema, and A. Saha, "Factors affecting individual investor behavior: empirical evidence from mutual fund investors in Dhaka city," *Management,* vol. 31, no. 3&4, 2017.

[23] P. Cramton, “Electricity market design,” *Oxford Review of Economic Policy,* vol. 33, no. 4, 2017.

[24] Y. Wang, X. Ai, Z. Tan, L. Yan, and S. Liu, “Interactive dispatch modes and bidding strategy of multiple virtual power plants based on demand response and game theory,” *IEEE Transactions on Smart Grid,* vol. 7, no. 1, pp. 510-519, 2015.

[25] J. Wang, C. Guo, C. Yu, and Y. Liang, “Virtual power plant containing electric vehicles scheduling strategies based on deep reinforcement learning,” *Electric Power Systems Research,* vol. 205, pp. 107714, 2022.

[26] R. Ma, Z. Yi, Y. Xiang, D. Shi, C. Xu, and H. Wu, “A Blockchain-Enabled Demand Management and Control Framework Driven by Deep Reinforcement Learning,” *IEEE Transactions on Industrial Electronics*, 2022.

[27] J. Li, T. Yu, and X. Zhang, “Coordinated load frequency control of multi-area integrated energy system using multi-agent deep reinforcement learning,” *Applied Energy,* vol. 306, pp. 117900, 2022.

[28] T. J. Gerpott, and I. Mahmudova, “Determinants of price mark-up tolerance for green electricity–lessons for environmental marketing strategies from a study of residential electricity customers in Germany,” *Business strategy and the environment,* vol. 19, no. 5, pp. 304-318, 2010.

[29] D. I. Maditinos, Ž. Šević, and N. G. Theriou, “Investors' behaviour in the Athens Stock Exchange (ASE),” *Studies in economics and Finance,* vol. 24, no. 1, pp. 32-50, 2007.

[30] T. Lal, “Investigating the Factors Affecting Investment Behaviour of Mutual Fund Investors.”

[31] M. Tamiz, R. A. Azmi, and D. F. Jones, “On selecting portfolio of international mutual funds using goal programming with extended factors,” *European Journal of Operational Research,* vol. 226, no. 3, pp. 560-576, 2013.

[32] H. Amiril, and A. M. Gil Lafuente, “Studying of the factors affecting on the mutual fund by individual investor in Iran, Malaysia, Turkey and US,” *Modern Applied Science, 2016, vol. 10, num. 9, p. 192-204*, 2016.

[33] D. Xiao, H. Chen, C. Wei, and X. Bai, “Statistical Measure for Risk-Seeking Stochastic Wind Power Offering Strategies in Electricity Markets,” *Journal of Modern Power Systems and Clean Energy*, 2021.

[34] Z. Zhu, K. W. Chan, S. Xia, and S. Bu, “Optimal Bi-Level Bidding and Dispatching Strategy between Active Distribution Network and Virtual Alliances Using Distributed Robust Multi-Agent Deep Reinforcement Learning,” *IEEE Transactions on Smart Grid*, 2022.

[35] Y. Zhang, Z. Mou, F. Gao, J. Jiang, R. Ding, and Z. Han, “UAV-enabled secure communications by multi-agent deep reinforcement learning,” *IEEE Transactions on Vehicular Technology,* vol. 69, no. 10, pp. 11599-11611, 2020.

[36] H. Liu, C. Zhang, Q. Chai, K. Meng, Q. Guo, and Z. Y. Dong, “Robust regional coordination of inverter-based Volt/Var control via multi-agent deep reinforcement learning,” *IEEE Transactions on Smart Grid,* vol. 12, no. 6, pp. 5420-5433, 2021.

[37] X. Lu, J. Qiu, G. Lei, and J. Zhu, “Scenarios modelling for forecasting day-ahead electricity prices: Case studies in Australia,” *Applied Energy,* vol. 308, pp. 118296, 2022.

[38] AusGrid. "Solar home electricity data," https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data.

[39] AEMO. "Aggregated price and demand data," https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data.

[40] "VPP Offers Available, 2020, Australian Energy Market Commission (AEMC) ".

[41] AEMO. "AEMO suspends NEM Wholesale Market," https://aemo.com.au/newsroom/media-release/aemo-suspends-nem-wholesale-market.

[42] AEMO. "Guide to Market Suspension in the NEM," https://aemo.com.au/en/energy-systems/electricity/emergency-management/guide-to-market-suspension-in-the-nem.