https://doi.org/10.1016/j.rser.2023.113466

Analysis of Strategic Interactions among Distributed Virtual Alliances in Electricity and Carbon Emission Auction Markets Using Risk-Averse Multi-Agent Reinforcement Learning

Ziqing Zhu[a,b], Ka Wing Chan[a,*], Siqi Bu[a], Siu Wing Or[a,b], Shiwei Xia[c]

[a] *Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

[b] *Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center, Hong Kong*

[c] *School of Electrical and Electronic Engineering, North China Electric Power University, Beijing, 102206, China*

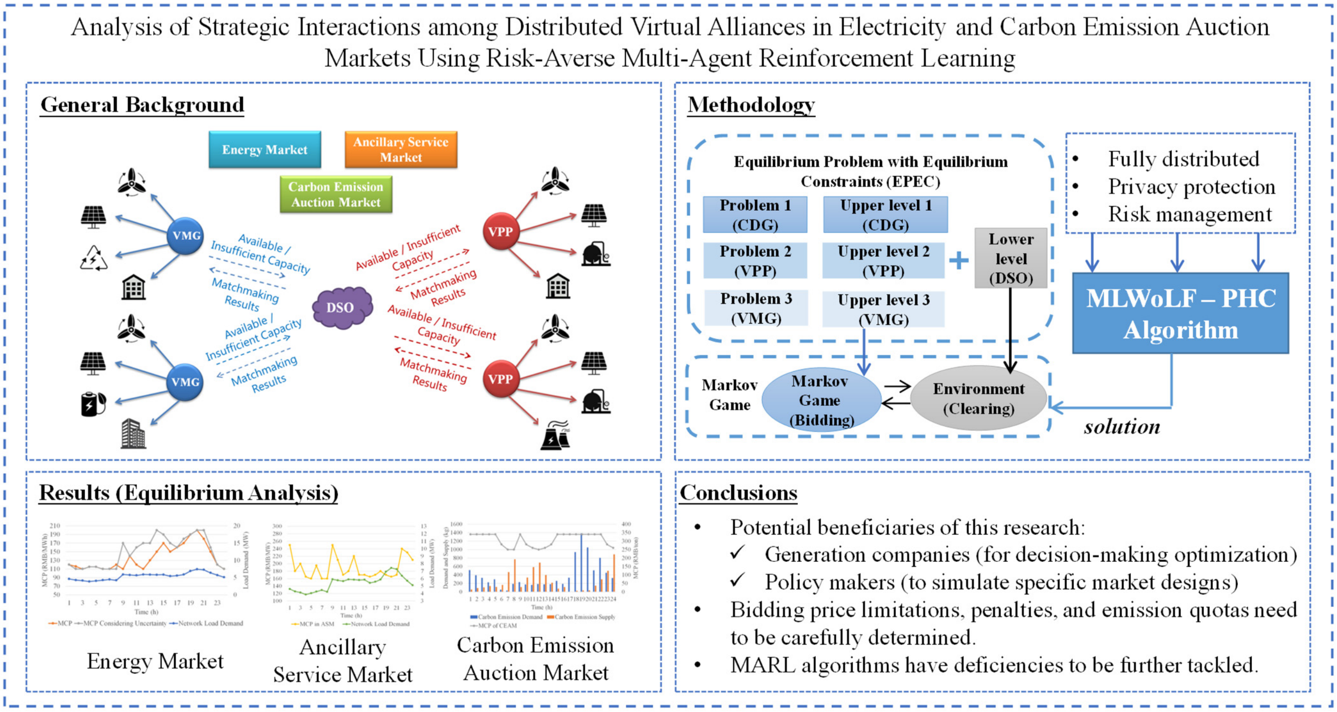
* *Corresponding Author: Ka Wing Chan, email: eekwchan@polyu.edu.hk*

Abstract—The incorporation of carbon emission auction market (CEAM) and ancillary service market (ASM) is an emerging trading paradigm in active distribution network (ADN). Such regime not only promotes the elimination of carbon emission, but also facilitates the secure operation of power network, especially considering the participation of distributed virtual alliances (DVAs) consisting of renewable distributed generators (RDGs) with uncertain output. In this research, a bi-level bidding and market clearing dynamic programming model is developed for in-depth analysis of market participants' bidding strategies and market equilibrium. This model allows DVAs to modify their bidding strategies in the energy market (EM), ASM and CEAM based on the market clearing results and uncertainty of RDG output. Also, a new Meta-Learning based Win-or-Learn-Fast (MLWoLF-PHC) algorithm, which not only enables the fully distributed bidding strategy modification, but also performs well considering uncertainty as a risk-averse method, is proposed to solve this model. Its computational performance, the market equilibrium analysis, and the impact of CEAM on the converged market clearing price of EM and ASM would be thoroughly investigated and examined in the case studies.

Index Terms—distributed network market, distributed virtual alliances, ancillary service market, carbon emission auction market, multi-agent reinforcement learning

(Words count: 9365)

Graphical Abstract


Analysis of Strategic Interactions among Distributed Virtual Alliances in Electricity and Carbon Emission Auction Markets Using Risk-Averse Multi-Agent Reinforcement Learning

General Background

Energy Market
Ancillary Service Market
Carbon Emission Auction Market
VMG
VPP
DSO
VMG
VPP
Available / Insufficient Capacity
Matchmaking Results
Available / Insufficient Capacity
Matchmaking Results
Available / Insufficient Capacity
Matchmaking Results
Available / Insufficient Capacity
Matchmaking Results

Methodology

- Fully distributed
- Privacy protection
- Risk management

Results (Equilibrium Analysis)

Energy Market

Ancillary Service Market

Carbon Emission Auction Market

Conclusions

- Potential beneficiaries of this research:
  - ✓ Generation companies (for decision-making optimization)
  - ✓ Policy makers (to simulate specific market designs)
- Bidding price limitations, penalties, and emission quotas need to be carefully determined.
- MARL algorithms have deficiencies to be further tackled.

Highlights

- The distribution level electricity and carbon emission market design is proposed.
- The optimal bidding strategy of distributed virtual alliances (DVAs) is formulated.
- The formulated optimal bidding strategy is converted to a Markov Decision Process.
- A novel MLWoLF-PHC algorithm is proposed to obtain the optimal bidding strategy.
- Impacts of carbon emission market on DVAs' bidding strategies are analyzed.

List of Abbreviations

ADN	Active Distribution Network
APF	Aggressive Policy Function
ASM	Ancillary Service Market
CEAM	Carbon Emission Auction Market
CEQ	Carbon Emission Quotas
CDG	Controllable Distributed Generators
CPF	Conservative Policy Function
DA	Day-Ahead
DER	Distributed Energy Resources
DVA	Distributed Virtual Alliances
DSO	Distribution System Operator
EM	Energy Market
EPEC	Equilibrium Programming with Equilibrium Constraint
ESS	Energy Storage Systems
ISO	Independent System Operator
MADDPG	Multi-Agent Deep Deterministic Policy Gradient
MARL	Multi-Agent Reinforcement Learning
MCP	Market Clearing Price
ML	Meta-learning
MDP	Markov Decision Process
MG	Microgrid
NEP	Nash Equilibrium Point
LA	Load Aggregator
RDG	Renewable Distributed Generators
RL	Reinforcement Learning
RT	Real-Time
VPP	Virtual Power Plant
VMG	Virtual Microgrids
WoLF-PHC	Win-or-Learn-Fast Policy-Hill-Climbing

Nomenclatures

A. Indices, Sets and Subscripts

$t \in T$	Index and set of time
$d \in D$	Index and set of days
$i \in I$	Index and set of controllable distributed generators (CDGs)
$j \in J$	Index and set of virtual microgrids (VMGs)
$n \in N$	Index and set of virtual power plants (VPPs)
$g \in G$	Index and set of DVAs
$\sigma \in \Omega$	Index and set of distribution lines

$s \in \mathcal{S}$	Index and set of state
$a \in \mathcal{A}$	Index and set of action
$r \in \mathcal{R}$	Index and set of reward
$p \in \mathcal{P}$	Index and set of transition probability
$k \in \mathcal{K}$	Index and set of agents
m	Index of dimensionality of action
h	Index of dimensionality of state
am	Subscript for ASM
av	Subscript for availability of capacity
$allo$	Subscript for allocated capacity for DVAs
ca	Subscript for CEAM
em	Subscript for EM

B. Variables

$C_{DSO,em}^{t,d}$	The total cost of distribution system operator (DSO) at the hour t of d-day
$C_{VMG,re}^{t,d}$	Cost of VMG for purchasing energy from DSO
$C_{ca}^{t,d}$	Cost of DVAs purchasing carbon emission quota
$C_{em}^{t,d}$	Generation cost of DVAs
$C_{pen,em}^{t,d}$	The penalty of DVAs for failing to provide energy or ancillary service to the DSO
$E_{av}^{t,d}$	Devoted quota of DVAs in CEAM
$E_{de}^{t,d}$	Demand quota of DVAs in CEAM
$P_{LA,re}^{t,d}$	The energy demand of load aggregator (LA)
$P_{VMG,re}^{t,d}$	The energy demand of VMG
$P_{am,req}^{t,d}$	The total capacity requirement for ancillary service
$P_{em,allo}^{t,d}$	Allocated capacity for the DVAs in the EM
$P_{em}^{t,d}$	Devoted capacity of DVAs in EM
$P_{err,em}^{t,d}$	The amount of capacity of DVAs failing to provide energy or ancillary service to the DSO
$P_{line,\sigma}^{t,d}$	The rated capacity of the σ^{th} line in the ADN
$R_{DSO,re}^{t,d}$	Revenue of DSO for selling energy to VMG/loads
$R_{em}^{t,d}$	Revenue of DVAs in EM
$\delta_{VMG}^{t,d}$	Binary variable indicating VMG purchase/sell energy from/to DSO
$\delta_{ca}^{t,d}$	Binary variable indicating purchase or sell carbon emission quota
$\lambda_{em}^{t,d}$	Submitted bidding price of DVAs
$\lambda_{mar,em}^{t,d}$	Previous market clearing price (MCP) in EM

C. Parameters

Δ_{sa}	The updating rate of policy function
$P_{line,\sigma}$	The capacity limitation of the σ^{th} line in the ADN
$P_{ramp,i}$	The ramping capacity limitation of the i^{th} DVA
ρ_k	The emission factor of the k^{th} DVA
α	The learning rate
γ	The discount rate

δ	The modification rate
μ	The risk coefficient

1. Introduction

With the decentralization and deregulation of ADN, distributed energy resources (DERs) invested and developed by independent stakeholders are becoming more autonomous and cannot be directly dispatched by the DSO [1]. Instead, distribution-level electricity market (DEM) would enable energy transactions between DERs and the ADN [2]. In emerging DEMs with DERs participation, in order to enlarge the market share to facilitate the participation of energy transaction, DERs would spontaneously form DVAs, including VPP and VMG [3]. VPP refers to the virtual alliance of DERs located at different places, while VMG additionally incorporates LAs.

Meanwhile, CEAM is currently introduced in numerous countries and pilot projects to encourage the reduction of carbon emission for mitigation of climate crisis [4]. Such paradigms aim to incentivize the involved DVAs to increase the penetration of RDGs, by restricting their carbon emission quantities within the assigned quota, while they are required to pay for the exceeding emission in CEAM [5]. However, how the incorporation of CEAM will affect the operation of existing electricity energy market is still unknown, especially in the aspect of MCP. In addition, the increasing penetration of RDGs with uncertain and intermit characteristics will adversely contribute to the secure operation of ADN [6]. Considering that conventional controllable generators in DVAs [7] with fast response capability can provide additional reserved capacity for ADN to achieve real-time balancing, the development of ancillary service market (ASM) in ADN is also under a pressing need to address problems caused by high penetration of RDGs [8].

In summary, the coordinated electricity market (including EM and ASM) and CEAM in ADN is an emerging market paradigm, and the involvement of multiple DVAs would further complicate the operation of such market. Developing new market paradigms raises critical issues for both distributed virtual aggregators (DVAs) and distribution system operators (DSOs), particularly concerning bidding procedure simulations and equilibrium computations. For DVAs, accurately estimating the market clearing price is essential to maximize profits and succeed in the bidding process. In addition, DSOs need to identify potential behaviors that may compromise fair market transactions, such as market power abuses and arbitrage, by simulating possible trading behaviors before implementing new market rules. This enables market regulations to be adjusted based on the incentive compatibility principle, ensuring fair trading practices. In light of these new market paradigms, which could significantly impact bidding strategy preferences, there is an urgent need to develop methodologies for analyzing and simulating DVAs' bidding strategies and converged market equilibrium in the coordinated market.

A considerable amount of studies has been published considering the interactions between downstream stakeholders including microgrids (MGs), VPPs and VMGs. In [9], the risk-averse optimal dispatching approach for DSO considering multiple MGs is proposed and solved by applying the Karush-Kuhn-Tucker conditions and the duality theory. In [10], a bi-level programming model is proposed to formulate the Stackelberg interaction of VPPs and ADN with solution of Karush-Kuhn-Tucker conditions and Fortuny-Amat transformation. These works have simply considered DVAs as dispatchable units without the incorporation of market trading. For the strategic bidding of DVAs, a stochastic bidding strategy for MGs in embedded energy hubs is introduced in [11] considering uncertainties of loads and energy price. A robust optimization model is formulated in [12] to develop the bidding strategy for MGs to minimize the expected net cost. These works involve the static bidding only, with solutions of mathematical programming tools, but without consideration of dynamic bidding procedure. In [13], the dynamic trading is formulated based on the blockchain technique and solved by the Particle Swarm Algorithm which is a collaborative heuristic method, and therefore is not suitable for use in the competitive bidding procedure.

The advancement of game theory, in particular, the equilibrium analysis, has given rise to techniques that hold the promise to transform numerous industries, including the electricity market [14]. For instance, the analysis of electricity market equilibrium is crucial for understanding the interactions among market participants, including power producers, consumers, and intermediaries [15]. This analysis provides insights into market dynamics, pricing mechanisms, and the overall efficiency of the electricity market. Specifically, the Equilibrium Problem with Equilibrium Constraints (EPEC) is a powerful tool for such analysis, as it models the strategic behavior of market participants under various constraints, reflecting the complex nature of the electricity market [16]. EPEC is particularly suited for electricity market analysis due to its ability to model multiple interacting agents, each seeking to optimize their own objectives while considering the actions of others. It captures the strategic interactions among market participants, including competition and cooperation, and the impact of various market rules and regulations. For example, [17] proposes an EPEC model to find the equilibrium of the interactions between distribution and transmission wholesale market clearing. The interaction among system operator and retailers in the day-ahead wholesale (DAW) and local power exchange (LPE) markets is formulated as the EPEC in [18], which is solved by the diagonalization algorithm. To accelerate the computation of equilibrium, a single-level equilibrium model that describes the equilibrium in general electricity market is developed in [19] as an equivalence of conventional bi-level model. In addition, [20] proposes a column-and-constraint generation algorithm to solve the EPEC model in an efficient manner. While EPEC provides valuable insights into market equilibrium analysis, caution should be exercised when interpreting its results in the context of real-world electricity markets due to its potential limitations, as the use of EPEC-derived equilibria for analyzing the electricity market can result in significant discrepancies between the analytical results and the actual market operations. This is primarily due to the stringent assumptions inherent in the EPEC model. Specifically, the model presumes that all market participants are perfectly rational, strictly adhere to their optimization models, and possess full observation of the market environment [21]. In reality, these conditions are rarely met in the electricity market. Market participants may not always act rationally due to various factors such as imperfect information, cognitive biases, and strategic behavior [22]. Furthermore, the assumption of full observation is often unrealistic, as market participants may not have complete information about the actions and strategies of others, or about future market conditions [23].

The emergence of multi-agent reinforcement learning (MARL) has revolutionized the field of electricity market operation by offering a cutting-edge methodology. By simulating how a group of "smart agents" interact with each other and optimize their strategies, MARL ensures maximum consistency with real-world scenarios [24]. The framework allows users to customize agent characteristics [25], such as decision-making and risk management, as well as the environment, which determines how profit is allocated to each agent based on their decisions. In comparison with conventional optimization models and EPEC method, the MARL-based methodology merits the following superiorities. First, for the privacy protection in the competitive electricity markets, MARL enables agents to keep their confidential objective (reward) function and policy function, without any communication with their rivals, and only need to submit their bidding price and available capacity at each round of bidding – this is exactly consistent with the real-world market operation [26]. Second, MARL enables agents to dynamically modify their bidding strategies at each round of bidding, and finally reach the equilibrium. Such a simulation is more consistent with the real-world market than EPEC with fixed objective function and constraints. Moreover, MARL algorithms can be implemented without strict assumptions of full observation and the perfect rationality of market participants. Finally, due to the fully-distributed training of MARL, the required computational resource will be effectively alleviated, and the computation speed will also be accelerated [25].

There are some initial attempts to deploy MARL in the distribution-level electricity markets with DVAs' participation. The optimal bidding strategy of market participants is computed in [25] and [26] using the multi-agent deep deterministic policy gradient (MADDPG) as a multi-agent reinforcement learning (MARL) method. This method is widely adopted as the most state-of-the-art agent-based method, because of its outstanding performance in terms of dealing with large dimensionality of continuous state and action space. However, this algorithm has two main limitations: firstly, the availability of other rivals' bidding strategies is required, and therefore the privacy of participants cannot be protected; secondly, this algorithm is not risk-averse, i.e., the obtained policy function cannot perform well under scenarios with risks, which are necessary to be taken into consideration of DVAs' decision making in an uncertain environment. While works [24] – [26] focused only on the energy market, few attempts have been made to investigate the impacts of incorporating multiple markets on both the bidding strategy of DVAs. In [27], an optimal bidding strategy is proposed for a coordinated EM and ASM. In [28], a decision-making method is developed for DVAs participating in EM and CEAM. The impact of CEAM incorporation in the aspects of EM equilibrium is analyzed in [29] and [30].

This research aims to provide a comprehensive investigation on the coordinated EM, ASM and CEAM with incorporation of DVAs, especially emerging DVAs including both VPP and VMG, in the aspects of optimal and dynamic bidding strategy of each DVA, and market equilibrium computation and analysis. The significance of this research is threefold. First, DVAs (as well as other kinds of generation companies) can benefit from understanding the implications of carbon emission auctions on their operational and strategic decisions. Therefore, they can make more informed decisions about fuel choices, investment in cleaner technologies, and bidding strategies in the electricity market. Second, Policymakers can use insights from the analysis of joint market to design more effective policies and regulations. This could include setting appropriate carbon prices to achieve emission reduction targets, designing mechanisms to mitigate potential adverse impacts on energy prices, and promoting investment in low-carbon technologies. Finally, the ultimate beneficiary of integrating the CEAM with the electricity market is the environment. By internalizing the cost of carbon emissions, the electricity market can drive a shift towards cleaner energy sources, leading to a reduction in greenhouse gas emissions.

Specifically, main contributions of this research are outlined as follows:

1) Optimal and Dynamic Bi-level Bidding Strategy and Market Clearing in Multiple Markets: This work develops a re-formulation of the EPEC model by using the Markov game model, to describe the DVA's decision making, considering the potential financial loss to DVAs caused by the uncertainty of DVAs' net load. This model innovatively formulates the DVAs' decision-making process as a dynamic (i.e., trial-and-error) and agent-based (i.e., distributed) manner, while enabling the consideration of risk management in their decision making. The Markov game model also lays the theoretical foundation of implementing the MARL algorithm to simulate how DVAs optimize their decisions.
2) Bidding Procedure Simulation and Market Equilibrium Calculus: A new Meta-Learning based WoLF-PHC (MLWoLF-PHC) algorithm is proposed to solve the proposed Markov game. In comparison with EPEC in a centralized computation manner, this algorithm enables the fully-distributed optimization of DVAs' bidding strategies without availability of other rival DVAs' private information. In addition, another superiority of this algorithm compared with existing methods is the risk-averse method considering the risk mitigation to prevent potential financial losses due to risks. Hence, the obtained equilibrium will be more consistent with the real-world electricity markets.
3) Computational Performance and Market Equilibrium Analysis: The computational performance of the proposed algorithm is firstly analyzed in terms of convergence speed and sensitivity to hyper-parameters. Meanwhile, the impacts of CEAM and the uncertainty of DVAs' net load on the bidding strategy of DVAs in the coordinated market are revealed.

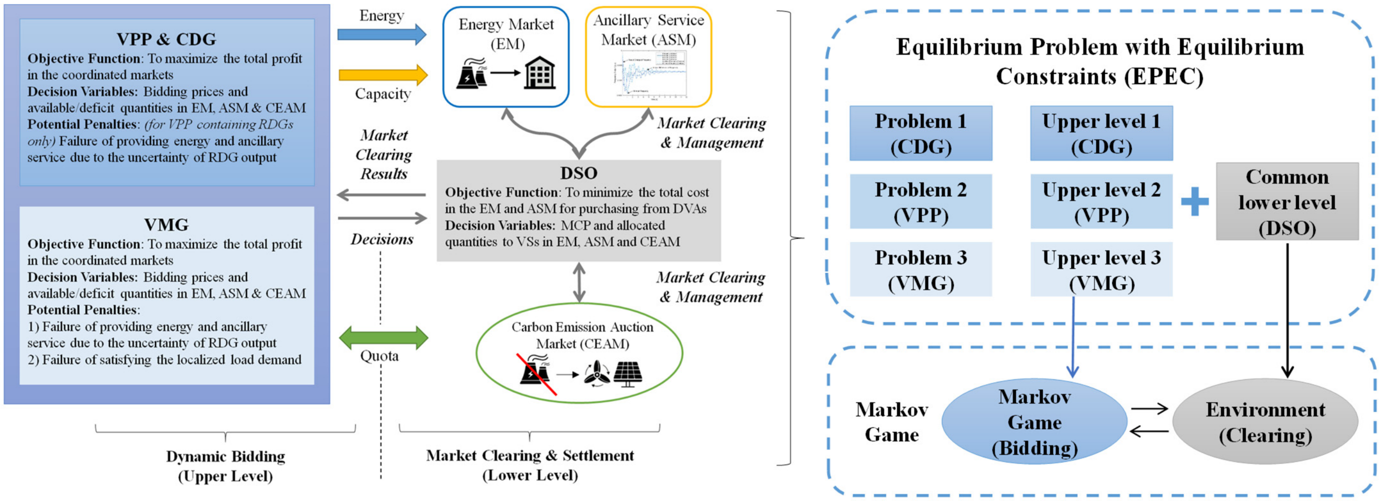
The remaining parts of this research are organized as follows. In Section 2, the EPEC model which describes the interaction among DVAs and the DSO is formulated and further re-written as an Markov Game model. In Section 3, the MLWoLF-PHC algorithm is proposed to solve the Markov Game and to compute the market equilibrium. In Section 4, a case study is conducted to demonstrate the impact of CEAM incorporation and DVAs participation on the MCP of both EM and ASM. Conclusions and future work are summarized in Section 5.

2. Problem Formulation

In this section, the market paradigm of the joint EM, ASM and CEAM is firstly introduced to clarify the general context. Then, the behavior of DVAs and DSO is formulated as a set of bi-level dynamic programming problems, i.e., the EPEC model, in which the decision variables and observable factors affecting the decision making are explicitly identified. Finally, this EPEC model is re-written as a Markov Game model for further processing using the proposed algorithm.

2.1 Market Paradigm and Model Description

Like conventional transmission-level wholesale market, the timeframe of distribution market constitutes day-ahead (DA) and real-time (RT), while the proposed EPEC model is based on DA trading. The market architecture, including: 1) market participants with their objectives and decision variables to be considered; 2) energy and information change among different participants and different markets, is illustrated in Fig.1. The product transacted in the EM and ASM are electricity and reserved capacity for frequency regulation response respectively, while the time-frame and coordination of this joint market is presented in Fig.2. Specifically, the novelties of this model are elaborated as follows:



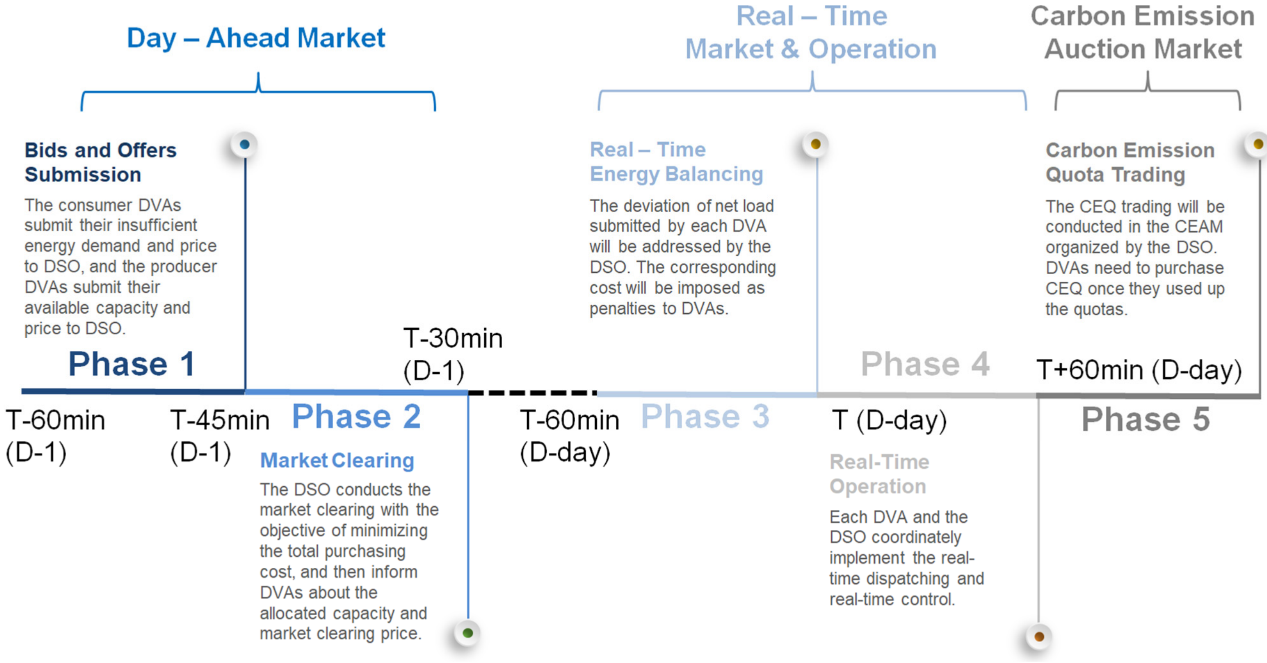
VPP & CDG
Objective Function: To maximize the total profit in the coordinated markets
Decision Variables: Bidding prices and available/deficit quantities in EM, ASM & CEAM
Potential Penalties: *(for VPP containing RDGs only)* Failure of providing energy and ancillary service due to the uncertainty of RDG output

VMG
Objective Function: To maximize the total profit in the coordinated markets
Decision Variables: Bidding prices and available/deficit quantities in EM, ASM & CEAM
Potential Penalties:
1) Failure of providing energy and ancillary service due to the uncertainty of RDG output
2) Failure of satisfying the localized load demand

DSO
Objective Function: To minimize the total cost in the EM and ASM for purchasing from DVAs
Decision Variables: MCP and allocated quantities to VSs in EM, ASM and CEAM

Energy
Capacity
Market Clearing Results
Decisions
Quota
Energy Market (EM)
Ancillary Service Market (ASM)
Market Clearing & Management
Market Clearing & Management
Carbon Emission Auction Market (CEAM)
Dynamic Bidding (Upper Level)
Market Clearing & Settlement (Lower Level)
Equilibrium Problem with Equilibrium Constraints (EPEC)
Problem 1 (CDG)
Problem 2 (VPP)
Problem 3 (VMG)
Upper level 1 (CDG)
Upper level 2 (VPP)
Upper level 3 (VMG)
Common lower level (DSO)
Markov Game
Markov Game (Bidding)
Environment (Clearing)

Fig.1 Coordinated EM, ASM and CEAM Architecture



Bids and Offers Submission
The consumer DVAs submit their insufficient energy demand and price to DSO, and the producer DVAs submit their available capacity and price to DSO.

Market Clearing
The DSO conducts the market clearing with the objective of minimizing the total purchasing cost, and then inform DVAs about the allocated capacity and market clearing price.

Real – Time Energy Balancing
The deviation of net load submitted by each DVA will be addressed by the DSO. The corresponding cost will be imposed as penalties to DVAs.

Real-Time Operation
Each DVA and the DSO coordinately implement the real-time dispatching and real-time control.

Carbon Emission Quota Trading
The CEQ trading will be conducted in the CEAM organized by the DSO. DVAs need to purchase CEQ once they used up the quotas.

Fig.2 Coordinated Market Design with Detailed Time-frame

In this model, the optimal bidding strategy of DVAs with dynamic modification in both EM and ASM is considered. DVAs are required to determine the available capacity devoted to each market along with bidding prices, which are submitted to the DSO, who would subsequently implement the simultaneous EM and ASM clearing, to minimize the total cost subjected to all necessary constraints. DVAs would be penalized if they fail to provide appointed energy or ancillary service to DSO, due to the uncertainty of net load. Also note that VMGs are required to purchase energy from DSO if the localized demand cannot be satisfied. These factors will be taken into DVAs' consideration when they make their bidding decisions. In addition, it is assumed that the hourly load demand for each single day is approximately the same. Hence, it is applicable for them to modify their bidding decisions in each hour based on the market clearing results and the reward obtained by the same hour in the previous day. Also note that decision making of each hour is independent with each other. This assumption is consistent with the real-world situations, and how this assumption is indicated in the model formulation and algorithm design will be revealed in the subsequent sections. The RT operation is designated to address deviations between DA scheduling and RT procurement due to the uncertainty of DVAs' net load. The deviation submitted by each DVA at RT will be addressed by the DSO via re-dispatch of DVAs, dispatch of DSO-owned generators and Energy Storage Systems (ESSs), as well as trading with the main grid. The resulted costs will be imposed as penalties for DVAs who resulted in such a deviation. Thereafter, the final generation schedule of each DVA is confirmed, and then each DVA will implement the self-dispatch based on the methodology mentioned in [31]. The voltage limitation and reactive power compensation is also considered in RT operation. Firstly, DVAs will control their Static Var Compensators and switching capacitors to constraint the voltage at each node to be within its limits while conducting the self-dispatching. Then, the DSO will determine the status of On-line Tap Changer and their owned Static Var Compensators and switching capacitors to further control the voltage in the whole network and especially the point of common coupling. However, the control of voltage will not affect the active power generated by each DVA, and therefore it will not have impact on the decision making of bidding and dispatching in both DA and RT, and the reactive power optimization (to minimize the total reactive power loss) would be out of the scope of this research.

The CEAM is interactively incorporated with electricity market, in order to further investigate the impact of CEAM on bidding strategies of DVAs in EM and ASM. DVAs with conventional controllable generators are allocated with carbon emission allowance quotas. After the RT operation, DVAs will compute their remained carbon emission allowance. If the allowance is insufficient, DVAs are required to submit the deficiency to the DSO, who will thereafter organize the trading for CEAM, in which DVAs with availability can submit their devoted quotas and bidding price. The market clearing and settlement will be conducted using the merit-order and uniform-price scheme in an hourly basis after the EM and ASM are cleared [32].

2.2 Model Formulation

In this subsection, the objective function and constraints of each DVA and the DSO are formulated, in which the decision variables and factors affecting the decisions are explicitly presented in the objective function. The model is developed as shown in (1)-(4), and the detailed description of objective function and constraints are summarized in Table I.

As indicated in Fig.1, the EPEC problem constitutes a set of bi-level problems, including 3 upper-level problems (referring to the optimal bidding of DVAs) and 1 common lower-level problem (referring to the market clearing of the DSO.) For each bi-level problem, the market clearing result in the lower-level is determined based on the decision made in the upper-level, in which the DVAs will make their decisions based on the market clearing results and the reward obtained by the same hour in the previous day. The result of the bidding and clearing processes is a set of bidding price plus generation/reserve capacities that simultaneously solve these 3 bi-level problems. In other words, the *Nash Equilibrium* of the DVA agent interactions are the solutions of this EPEC problem. The detailed numerical relationships between the total reward, decision variables and those factors affecting the decisions are neglected, while such relationship will be "learned" by the implementation of proposed MLWoLF-PHC algorithm, as elaborated in Section 3. Instead, the notation $R(A^{t,d}|B^{t,d-1})$ is used to indicate such relationships between decision $A^{t,d}$, observation $B^{t,d-1}$ and reward R, in which R is directly determined by $A^{t,d}$, and $A^{t,d}$ is determined based on $B^{t,d-1}$ (observation of the same hour in the previous day).

DVAs are required to purchase carbon emission quotas (CEQ) if the remained allowance is not sufficient. The allowance of CEQ is allocated based on the capacity and types of generation [33]. The deficiency of CEQ can be estimated by $E^{t,d}_{k,de} = E^{t,d}_{k,allo} - \rho_k P^{t,d}_{k}$, in which $E^{t,d}_{k,allo}$ denotes the allocated CEQ to the kth DVA, ρ_k is the emission factor, and $P^{t,d}_{k}$ is the actual generation in hour t.

Upper Level:

1) CDG:

$$\text{Max } \sum\nolimits_{t\in T}\left\{\begin{array}{c} \underbrace{R^{t,d}_{CDG,em}\left(P^{t,d}_{CDG,em},\lambda^{t,d}_{CDG,em}\middle|\lambda^{t,d-1}_{mar,em},P^{t,d-1}_{CDG,em,allo}\right)}_{\text{①}\ CDG\ Revenue\ in\ EM} \\ +\underbrace{R^{t,d}_{CDG,am}\left(P^{t,d}_{CDG,am},\lambda^{t,d}_{CDG,am}\middle|\lambda^{t,d-1}_{mar,am},P^{t,d-1}_{CDG,am,allo}\right)}_{\text{②}\ CDG\ Revenue\ in\ ASM} \\ +\underbrace{\delta^{t,d}_{CDG,ca}R^{t,d}_{CDG,ca}\left(E^{t,d}_{CDG,av},\lambda^{t,d}_{CDG,ca}\middle|\lambda^{t,d-1}_{mar,ca},E^{t,d-1}_{CDG,ca,allo}\right)}_{\text{③}\ CDG\ Revenue\ in\ CEAM\ (if\ available)} \\ -\underbrace{C^{t,d}_{CDG,em}\left(P^{t,d}_{CDG,em}\right)}_{\text{④}\ Generation\ Cost}-\underbrace{\left(1-\delta^{t,d}_{CDG,ca}\right)C^{t,d}_{CDG,ca}\left(E^{t,d}_{CDG,de}\right)}_{\text{⑤}\ Carbon\ Emission\ Cost} \end{array}\right\} \tag{1}$$

s.t.

$$P_{CDG,i,min} \le P^{t,d}_{CDG,i,dm} + P^{t,d}_{CDG,i,am} \le P_{CDG,i,max}\ ,\forall t,\ \forall i \tag{1-a}$$

$$\lambda_{CDG,min} \le \lambda^{t,d}_{CDG,i} \le \lambda_{CDG,max}\ ,\forall t,\ \forall i \tag{1-b}$$

$$P_{ramp,i,min} \le P^{t,d}_{CDG,i,em} - P^{t-1,d}_{CDG,i,em} \le P_{ramp,i,max}\ ,\forall t,\ \forall i \tag{1-c}$$

2) VPP:

$$\text{Max } \sum\nolimits_{t\in T}\left\{\begin{array}{c} \underbrace{R^{t,d}_{VPP,em}\left(P^{t,d}_{VPP,em},\lambda^{t,d}_{VPP,em}\middle|\lambda^{t,d-1}_{mar,em},P^{t,d-1}_{em,allo},P^{t,d-1}_{VPP,err,em}\right)}_{\text{①}\ VPP\ Revenue\ in\ EM} \\ +\underbrace{R^{t,d}_{VPP,am}\left(P^{t,d}_{VPP,am},\lambda^{t,d}_{VPP,am}\middle|\lambda^{t,d-1}_{mar,am},P^{t,d-1}_{VPP,am,allo}\right)}_{\text{②}\ VPP\ Revenue\ in\ AM} \\ +\underbrace{\delta^{t,d}_{VPP,ca}R^{t,d}_{VPP,ca}\left(E^{t,d}_{VPP,av},\lambda^{t,d}_{VPP,ca}\middle|\lambda^{t,d-1}_{mar,ca},E^{t,d-1}_{VPP,ca,allo}\right)}_{\text{③}\ VPP\ Revenue\ in\ CEAM\ (if\ available)} \\ -\underbrace{C^{t,d}_{VPP,em}\left(P^{t,d}_{VPP,em}\right)}_{\text{④}\ Generation\ Cost}-\underbrace{C^{t,d}_{VPP,pen,em}\left(P^{t,d}_{VPP,RDG,err}\right)}_{\text{⑤}\ Penalty} \\ -\underbrace{\left(1-\delta^{t,d}_{VPP,ca}\right)C^{t,d}_{VPP,ca}\left(E^{t,d}_{VPP,de}\right)}_{\text{⑥}Carbon\ Emission\ Cost} \end{array}\right\} \tag{2}$$

s.t.

$$P_{VPP,n,min} \le P^{t,d}_{VPP,n,em} + P^{t,d}_{VPP,n,am} \le P_{VPP,n,max}\ ,\forall t,\ \forall n \tag{2-a}$$

$$\lambda_{VPP,min} \le \lambda^{t,d}_{VPP,n} \le \lambda_{VPP,max}\ ,\forall t,\ \forall n \tag{2-b}$$

$$P_{ramp,n,min} \le P^{t,d}_{VPP,n,em} - P^{t-1,d}_{VPP,n,em} \le P_{ramp,n,max}\ ,\forall t,\ \forall n \tag{2-c}$$

3) VMG:

$$\text{Max } \sum_{t\in T}\left\{\begin{array}{l}\delta_{VMG}^{t,d}\left[\begin{array}{c}\underbrace{R_{VMG,em}^{t,d}\left(P_{VMG,j,em}^{t,d},\lambda_{VMG,em}^{t,d}\middle|\lambda_{mar,em}^{t,d-1},P_{em,allo}^{t,d-1},P_{VMG,err,em}^{t,d-1}\right)}_{\textcircled{1}\ VMG\ Revenue\ in\ EM}\\ +\underbrace{R_{VMG,am}^{t,d}\left(P_{VMG,am}^{t,d},\lambda_{VMG,am}^{t,d}\middle|\lambda_{mar,am}^{t,d-1},P_{VMG,am,allo}^{t,d-1}\right)}_{\textcircled{2}\ VMG\ Revenue\ in\ AM}\\ -\underbrace{C_{CDG,em}^{t,d}\left(P_{CDG,j,em}^{t,d}\right)}_{\textcircled{3}\ Generation\ Cost}-\underbrace{C_{pen,em}^{t,d}\left(P_{RDG,err}^{t,d}\right)}_{\textcircled{4}\ Penalty}\\ +\underbrace{\delta_{VMG,ca}^{t,d}R_{VMG,ca}^{t,d}\left(E_{VMG,av}^{t,d},\lambda_{VMG,ca}^{t,d}\middle|\lambda_{mar,ca}^{t,d-1},E_{VMG,ca,allo}^{t,d-1}\right)}_{\textcircled{5}\ VMG\ Revenue\ in\ CEAM\ (if\ available)}\end{array}\right]\\ -\left(1-\delta_{VMG}^{t,d}\right)\left[\underbrace{C_{VMG,re}^{t,d}\left(P_{VMG,re}^{t,d}\right)}_{\textcircled{6}Cost\ in\ EM}\right]-\underbrace{\left(1-\delta_{VMG,ca}^{t,d}\right)C_{VMG,ca}^{t,d}\left(E_{VMG,de}^{t,d}\right)}_{\textcircled{7}Carbon\ Emission\ Cost}\end{array}\right\} \quad (3)$$

s.t.

$$P_{VMG,j,min} \le P_{VMG,j}^{t,d} \le P_{VMG,j,max}\ ,\forall t,\ \forall j \quad (3\text{-a})$$

$$\lambda_{VMG,min} \le \lambda_{VMG,j}^{t,d} \le \lambda_{VMG,max}\ ,\forall t \quad (3\text{-b})$$

$$P_{ramp,j,min} \le P_{VMG,j,em}^{t,d} - P_{VMG,j,em}^{t-1,d} \le P_{ramp,j,max}\ ,\forall t,\ \forall j \quad (3\text{-c})$$

$$P_{Load,j}^{t,d} + P_{Load,j,cur}^{t,d} = P_{Load,j,total}^{t,d} \quad (3\text{-d})$$

$$P_{Load,j,cur}^{t,d} \le P_{Load,cur,max}^{t,d} \quad (3\text{-e})$$

Lower Level:

4) DSO:

$$\text{Min } \sum_{t\in T}\left\{\begin{array}{c}\underbrace{C_{DSO,em}^{t,d}\left(\lambda_{mar,em}^{t,d},P_{em,allo,g}^{t,d}\middle|P_{em,g}^{t,d},\lambda_{em,g}^{t,d}\right)}_{\textcircled{1}\ Cost\ of\ Purchasing\ Energy}\\ +\underbrace{C_{DSO,am}^{t,d}\left(\lambda_{mar,am}^{t,d},P_{am,allo,g}^{t,d}\middle|P_{am,g}^{t,d},\lambda_{am,g}^{t,d}\right)}_{\textcircled{2}\ Cost\ of\ Purchasing\ Ancillary\ Service}\\ -\underbrace{R_{DSO,re}^{t,d}\left(\lambda_{re}^{t,d}\middle|P_{VMG,re}^{t,d},P_{LA,re}^{t,d}\right)}_{\textcircled{3}\ Revenue\ of\ selling\ on\ EM}\end{array}\right\} \quad (4)$$

s.t.

$$\sum_{g\in G} P_{em,allo,g}^{t,d} - P_{VMG,re}^{t,d} - P_{LA,re}^{t,d} = 0,\ \forall t,\forall g \quad (4\text{-a})$$

$$\sum_{g\in G} P_{am,allo,g}^{t,d} \ge P_{am,req}^{t,d},\ \forall t \quad (4\text{-b})$$

$$P_{em,allo,g}^{t,d} \le P_{em,g}^{t,d}, \forall t,\forall g \quad (4\text{-c})$$

$$P_{am,allo,g}^{t,d} \le P_{am,g}^{t,d}, \forall t,\forall g \quad (4\text{-d})$$

$$P_{line,\sigma,min} \le P_{line,\sigma}^{t,d} \le P_{line,\sigma,max}, \forall t,\forall \sigma \quad (4\text{-e})$$

As indicated in (1) - (3), the objective functions of DVAs (i.e., the reward that can be pursued in the joint market) inherits several terms that are uncertain, i.e., without explicit numerical expression and cannot be predicted. These terms are specified in details as follows:

1) Uncertainties of DVAs' net load will lead to deviations between DA scheduling and RT procurement, and such deviations will result in penalties imposed by the DSO for the purpose of RT balancing, as shown in term ⑤ of (2) and term ④ of (3).

2) Uncertainties of CEQ price fluctuation in the CEAM will lead to additional costs for DVAs with insufficient CEQ allowance, as shown in term ⑤ of (1), term ⑥ of (2), and term ⑦ of (3).

TABLE I
MVACM Model Description

Types of Stakeholders	Model	Equation	Description
CDG	Objective function	(1)	✓ To maximize the total reward in consecutive $T = 24$ hours, by selling energy in the EM, selling reserved capacity in the ASM, and being as a prosumer in the CEAM (selling/purchasing the CEQ)
	Decision Variables	/	1) The available capacity and bidding price of each hour t in the d day's DA market (EM and ASM) 2) The available/insufficient CEQ and the bidding price of each hour t in the d day's CEAM.
	Constraints	(1-a) (1-b) (1-c)	Capacity limitation Bidding price limitation Ramp capacity limitation
VPP	Objective function	(2)	✓ To maximize the total reward in consecutive $T = 24$ hours, by being as a prosumer in the EM, ASM and CEAM considering potential risks
	Decision Variables	/	Same as CDGs'.
	Constraints	(2-a) (2-b) (2-c)	Capacity limitation Bidding price limitation Ramp capacity limitation
VMG	Objective function	(3)	✓ To maximize the total reward in consecutive $T = 24$ hours, by being as a prosumer in the EM, ASM and CEAM considering potential risks, while ensuring the localized load demand to be satisfied
	Decision Variables	/	1) The available/insufficient capacity and bidding price of each hour t in the d day's DA market (EM and ASM) 2) The available/insufficient CEQ and the bidding price of each hour t in the d day's CEAM.
	Constraints	(3-a) (3-b) (3-c) (3-d) (3-e)	3-a) Capacity limitation 3-b) Bidding price limitation 3-c) Ramp capacity limitation 3-d) Localized power balancing 3-e) Load curtailment limitation
DSO	Objective function	(4)	✓ To minimize the total cost of purchasing energy in the energy market, considering the revenue of selling the energy in retail market to VMGs and Load Aggregators.
	Decision Variables	/	Allocated capacity and CEQ of each DVA as successful bids, and the market clearing price in each market.
	Constraints	(4-a) (4-b) (4-c) (4-d) (4-e)	4-a) Network power balancing 4-b) The reserved capacity requirement 4-c) Limitation of available energy to be procured 4-d) Limitation of available reserved capacity to be procured 4-e) Line capacity limitation

2.3 Markov Game Model Formulation

As indicated in (1)-(3), the DVAs' decision making in the hour t of day d is based on the market clearing results of the same hour in the previous day. Such decision making therefore inherits the Markov Property, i.e., decisions are affected by observations in the previous time slot. Hence, the EPEC developed in (1)-(4) can be re-formulated as a Markov Game [25] to fit the implementation of MARL better, plus (4) indicating the non-strategic market clearing of DSO as an external environment, as presented in Fig.1. Specifically, the following modifications have been made: 1) The "strategic" indicates that DVAs can randomly modify their decisions to maximize total profits subject to the constraints, while the "non-strategic" indicates the DSO can only implement the market clearing based on the merit-order principle. 2) The market clearing of DSO is considered as the "external environment" with fixed rules, as indicated in (4).

The Markov Game can be modeled in the following compact form:

$$\tau =< \mathcal{S}_{k,h}^{t,d}, \mathcal{A}_{k,m}^{t,d}, \mathcal{P}^{t,d}, \mathcal{R}_{k}^{t,d} > \tag{5}$$

where $\mathcal{S}_{k,h}^{t,d}, \mathcal{A}_{k,m}^{t,d}, \mathcal{R}_{k}^{t,d}$ denote the set of *state* (with h dimensionalities), *action* (with m dimensionalities) and *reward* of the k th agent. $\mathcal{P}^{t}$ refers to the set of *transition probability*: $P(\mathcal{S}_{k,h}^{t,d} | \mathcal{S}_{k,h}^{t,d-1} = s_{k,h}^{t,d-1}, \mathcal{A}_{k,m}^{t,d-1} = a_{k,m}^{t,d-1})$ and $P(\mathcal{R}_{k}^{t,d} | \mathcal{S}_{k,h}^{t,d} = s_{k,h}^{t,d}, \mathcal{A}_{k,m}^{t,d} = a_{k,m}^{t,d})$, in which the state and reward in the hour t of day d are based on the state and action of the same hour in the previous day, being consistent with the Markov Property.

Generally, the state refers to the observation of each participant at the tth time step, the action refers to the submitted bidding decision at the tth time step, including the bidding price and capacity, and the reward refers to remuneration obtained by each participant at the tth time step. Here, the state is specified as the load demand announced by the DSO in the hour t of day d, as well as the decision made in the same hour of the previous day. The action refers to the modification of bidding decisions in the hour t of day d compared with the previous day. The reward refers to the total obtained profit after the RT operation and CEQ trading in the CEAM, considering all the risks mentioned in the previous subsection.

At time-step t, each DVA firstly observes some necessary information for their decision-making, i.e., the "states", including the load demand L_t, its available capacity, and the remuneration in the previous time-step, etc. Based on these observations, each DVA makes its decisions on how to participate in the market, i.e., the "actions". These decisions include whether to participate in markets, the bidding price and capacities in the EM, ASM and CEAM. After the market clearing and settlement, DVAs adjust their "strategies" in order to receive more profits in the coming rounds of bidding, i.e., the "rewards". At time-step $t+1$, each DVA observes the updated information, and makes decisions based on the updated strategies. Such a procedure repeats at $t+2, t+3, \dots$

However, even though the transition probability with initial values of state and action is given, the subsequent states and actions are still not determined, as the actions of each agent are subject to their own *policy*. The optimization of policy can be achieved by Reinforcement Learning (RL), to maximize the expected accumulated reward $\mathbb{E}\left[\sum_{t=1}^{T} \mathcal{R}_{k}^{t,d}\right]$ by finding the optimal policy (to be discussed in Section 3). Hence, the policy optimization procedure in the Markov game can be formulated as:

$$\text{Max} \sum_{t \in T} \mathcal{R}_{k}^{t,d} \left(\mathcal{A}_{k,m}^{t,d} = \{a_{k,m}^{t,d}\} | \mathcal{S}_{k,h}^{t,d} = \{s_{k,h}^{t,d}\}\right) \tag{6}$$

$$s.t. \quad a_{k,m,min}^{t,d} \leq a_{k,m}^{t,d} \leq a_{k,m,max}^{t,d}, \forall t, \forall m \tag{6-a}$$

where $\mathcal{H} = 1, \dots, h$ and $\mathcal{M} = 1, \dots, m$ are the dimensionality of state and action. The DSO's market clearing is considered as the environment of this Markov Game, as DVAs will receive rewards from the market clearing based on their actions.

3. Meta-Learning Based Modified WoLF-PHC (MLWoLF-PHC) Algorithm

In this section, the MLWoLF-PHC algorithm is proposed to solve the Markov Game model formulated in (6). Compared with conventional RL methods, this algorithm is risk-averse, i.e., the obtained policy can be considered as a dilemma between pursuing the maximum reward and risk mitigation. Firstly, some fundamental definitions are reviewed to facilitate subsequent elaborations. Then, the framework of proposed MLWoLF-PHC algorithm is described as a combination of conventional WoLF-PHC and a novel Meta-learning module for risk mitigation, followed by detailed implementations and justifications.

3.1 Basic Definitions

Definition 1 (Reinforcement Learning, RL): The concept of RL refers to the process of "agent" learning from interactions with the "environment", i.e. the transition probability, to optimize its "action" at each "state", which is the so-called the policy function, to maximize the expected accumulated reward $\mathbb{E}\left[\sum_{t=0}^{T} \mathcal{R}_k^{t,d}\right]$.

Definition 2 (Policy function and Q-function): The objective of RL is to obtain the optimal "policy", i.e. the action that should be taken at the given state, to maximize the expected cumulative reward, which is formulated as:

$$\mathcal{R}_k^d = \mathbb{E}_{a_{k,m}^{t,d} \sim \pi_k\left(a_{k,m}^{t,d} \middle| \mathcal{S}_{k,h}^{t,d}\right)}\left[\sum_{t=1}^{T} \gamma_t \, r_k^{t,d}\right] \tag{7}$$

where γ_t denotes the discount factor, $r_k^{t,d}$ is the immediate reward obtained by the agent at time t. The policy function $\pi_k\left(a_{k,m}^{t,d} \middle| \mathcal{S}_{k,h}^{t,d}\right)$ is parameterized by Θ^t, which is a probability density function of actions at each state. Once the policy function is determined, the expected reward can be expressed as the function of the state-action pair:

$$Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) = \mathbb{E}\left(\mathcal{R}_k^{t,d} \middle| a_{k,m}^{t,d} \sim \pi_k\left(a_{k,m}^{t,d} \middle| s_{k,h}^{t,d}\right)\right) \tag{8}$$

which is the so-called the Q-value.

Definition 3 (Multi-Agent Reinforcement Learning, MARL): Consider the Markov Game with multiple DVAs formulated in (6). In MARL, the reward $\mathcal{R}_{\mathrm{k}}^{\mathrm{t,d}}$ and the state $s_{k,h}^{t,d}$ are determined by the joint action with dimensions of $|a_{1,m}^{t,d}| \times |a_{2,m}^{t,d}| \times \ldots \times |a_{k,m}^{t,d}|$ and the joint policy, instead of the individual action and policy of each agent.

3.2 Overview of MLWoLF-PHC

To solve the Markov Game formulated in (6), the following concerns should firstly be considered. Firstly, in competitive electricity market, the bidding strategy of each DVA is confidential. However, for most of existing MARL algorithms, it is assumed that agent has information of others' policies. Secondly, the potential risks mentioned in Section 2.2 will further complicate the reward function, resulting in different rewards for a same state-action pair. However, conventional RL algorithms can only deal with Markov Game with certain reward function, i.e., for each state-action pair, the reward is deterministic. Hence, the obtained optimal policy may lead to the maximized reward in some scenarios, but it may also lead to huge losses.

In order to tackle the first problem, the proposed algorithm is based on the architecture of WoLF-PHC, to facilitate the fully-distributed training of each DVA's aggressive policy function (APF), which is intended to pursue the maximum reward without considering risks. For the second problem, a Meta-Learning based module is incorporated to compute the conservative policy function (CPF) considering the losses due to uncertainty. While selecting actions, the final implemented policy is the combination of obtained aggressive policy and conservative policy, while a risk coefficient is introduced to measure the preference of undertaking the risk.

3.3 Procedure of WoLF-PHC Module

The aim of this module is to update the Q-value and APF of each DVA agent. The Q-value $Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)$ serves as an indicator for the performance of policy $\pi_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)$. In conventional WoLF-PHC algorithm, the updating rule of Q-values is formulated as follows:

$$Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) \leftarrow Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) + \alpha\left[\mathcal{R}_k^t + \gamma \mathcal{V}_k^{t,d}\left(s_{k,h}^{t,d+1}, (a_{k,m}^{t,d+1})'\right) - Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)\right] \quad (10)$$

$$\mathcal{V}_k^{t,d}\left(s_{k,h}^{t,d+1}, (a_{k,m}^{t,d+1})'\right) = \text{Max}\, Q_k^{t,d}\left(s_{k,h}^{t,d}, (a_{k,m}^{t,d})'\right) \quad (11)$$

where $\mathcal{V}_k^t$ represents the maximum Q-value of the state-action pair $(s_{k,h}^{t,d}, a_{k,m}^{t,d})$, which is a threshold value for increasing or decreasing $Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)$. If the updated Q-value is larger than $\mathcal{V}_k^t$, which means the action $(a_{k,m}^{t,d})'$ is the current optimal action, then the Q-value needs to be increased indicating a better performance, and vice versa.

The modification of $\pi_{k,APF}$ is based on the rule of "win or learn fast". The status of win or lose is determined by the comparison of performance of the current policy and the so-called average policy $\bar{\pi}_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)$, which is defined as:

$$\bar{\pi}_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) \leftarrow \bar{\pi}_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) + \frac{1}{C(s_{k,h}^{t,d})}\left[\pi_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) - \bar{\pi}_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)\right] \quad (12)$$

where $C(s_{k,h}^t)$ is the number of times that state $S_{k,h}^t$ appears. The "performance" is therefore written as:

$$p\left(\pi_{k,APF}\right) = \sum_{a_{k,m}^{t,d} \in \mathcal{A}} \pi_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) \quad (13)$$

Agents will tend to "slowly" increase the policy functions with a small step (i.e., modification rate δ) when the performance of $\pi_{k,APF}$ is larger than $\bar{\pi}_{k,APF}$ (the "win" state), and otherwise to "quickly" decrease it (the "lose" state) for a faster improvement.

3.4 Procedure of Meta-Learning Module

The aim of this module is to compute the conservative policy function $\pi_{k,CPF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)$ for DVAs to mitigate the potential penalties due to uncertainty. This module sequentially constitutes the following steps.

Firstly, the Maximin method is implemented herein to guess the estimated CPF. The CPF specifically refers to the policy for maximizing the reward in the worst scenario, i.e., the minimum reward among all possible rewards:

$$x_k^*\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) = \text{argmax}_{x_k}\left\{\text{Min}_{a_{k,m}^{t,d}}\left[\mathcal{R}_k^{t,d}\left(x_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)\right)\right]\right\} \quad (14)$$

where $\mathcal{R}_k^{t,d}\left(x_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)\right)$ denotes the set of all the possible rewards of playing action $a_{k,m}^t$ considering risks. Similar to the WoLF-PHC module, the CPF is updated as:

$$\pi_{k,CPF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) \leftarrow x_k^*\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) + \Delta_{sa} \quad (15)$$

The final implemented policy is a convex combination of APF and CPF. Considering the accuracy of guessing the CPF will be increased, the updating rate of CPF can be determined as a variable which increases linearly with time, and the updating speed will be relatively slow at the beginning and be much faster at the end of training, as shown in (16):

$$\pi_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) \leftarrow \left[1 - \max\left(\mu, \frac{1}{w_{epi}}\right)\right]\pi_{k,CPF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) + \max\left(\mu, \frac{1}{w_{epi}}\right)\pi_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) \quad (16)$$

where μ denotes the risk coefficient, indicating the willingness of each DVA to undertake the risk, and w_{epi} denotes the number of episodes that have been implemented. The workflow of MLWoLF-PHC algorithm is summarized in Algorithm 1.

Algorithm 1 Workflow of MLWoLF-PHC Algorithm

```
Input:
    1) The network configuration parameters and line capacity limitations;
    2) The rated capacity (upper bound of the capacity limitation) of controllable generators;
    3) The bidding price limitation;
    4) The daily generation profile of each DER, the load profile of each DVA, with uncertainty set indicating the probabilistic of possible scenarios;
    5) The carbon emission factors and quota allocation of each DVA;
    6) The set of hyper-parameters for training.
Output:
    The converged policy (state-action pair) of bidding decisions of each DVA agent.
1:  Initialize $Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) = 0, \pi_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) = \frac{1}{\left|a_{k,m}^{t,d}\right|}$, $\bar{\pi}_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) = \frac{1}{\left|a_{k,m}^{t,d}\right|}$, $C\left(s_{k,h}^{t,d}\right) = 0$
2:  Each DVA randomly generates the initial state and initial action.
For iterations repeat
3:  DSO clears the market using (4) and releases the result to each DVA.
4:  Each DVA implements the RT dispatching, while the DSO organizes the RT balancing and the penalty execution using method elaborated in [36].
5:  Each DVA computes the remained CEQ and make bidding decisions in the CEAM.
6:  Each DVA obtains the final reward $\mathcal{R}_k^{t,d}$.
 (Proceeding to the WoLF-PHC module)
7:  Each DVA updates the Q-value as shown in (10).
8:  Each DVA updates the average policy as shown in (12).
9:  Each DVA updates the APF:
        $\pi_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) = \pi_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) + \Delta_{sa}$
        $\Delta_{sa} = \begin{cases} -\Delta_{sa}, & if \;\; a_{k,m}^{t} \neq \text{argmax}\, Q_k^{t,d}(s_{k,h}^{t,d}, (a_{k,m}^{t,d})') \\ & \Delta_{sa}, o.w. \end{cases}$
    where   $\Delta_{sa} = min\left(\pi_i\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right), \frac{\delta}{\left|a_{k,m}^{t,d}\right|-1}\right)$
    if
        $\sum_{a_{k,m}^{t,d} \in \mathcal{A}} \pi_{k,APF}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) > \sum_{a_{k,m}^{t,d} \in \mathcal{A}} \bar{\pi}_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right) Q_k^{t,d}\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)$
            $\delta = \delta_{win}$
    else
            $\delta = \delta_{lose}$
    end if
(Proceeding to the Meta-Learning module)
10: Each DVA computes the current CPF using (14) and (17).
11: Each DVA updates the final implemented policy function $\pi_k\left(s_{k,h}^{t,d}, a_{k,m}^{t,d}\right)$ using (16).
12: Each DVA updates the dataset of possible reward considering penalties.
(Proceeding to the bidding in the same hour in $d+1$ day)
13: Each DVA generates the action $a_{k,m}^{t,d+1}$ using the $\epsilon$-greedy method introduced in [16].
14: $d \leftarrow d+1$ and return to step 3 until convergence.
end for
```

4. Case Study

4.1 Test System and Input Data

In this case study, a test market consisting of 5 distributed DVAs is adopted, in which the DVAs are aggregations or virtual alliances of generators in a modified IEEE 33-Bus Network, as shown in Fig.3. The network configuration parameters and line capacity limitations are detailed in [34]. The rated capacity and generation cost of controllable generators are available in Table II. The renewable generation output and load profile with an hourly resolution, as well as the generated scenarios with corresponding probabilities for stochastic simulation, are referenced from [35], in which the scenario generation and reduction techniques are utilized considering the probabilistic correlations of different types of renewable generations and loads. The reserved capacity for ancillary service is estimated as 110% of the total net load variation in the network [36]. The carbon emission factors [32] and quota allocation [33] of each DVA are presented in Table III. The limitation of submitted bidding price (RMB) in the EM, ASM and CEAM are (110, 200), (160, 250) and (250, 340) respectively. The real-time dispatching methods of DVAs and penalties for failure to provide energy or ancillary service are referenced to [31]. The results presented in Fig.4 and Fig.6-Fig.8 are average values of 100 times of repetitive experiments for better consistency and reliability.

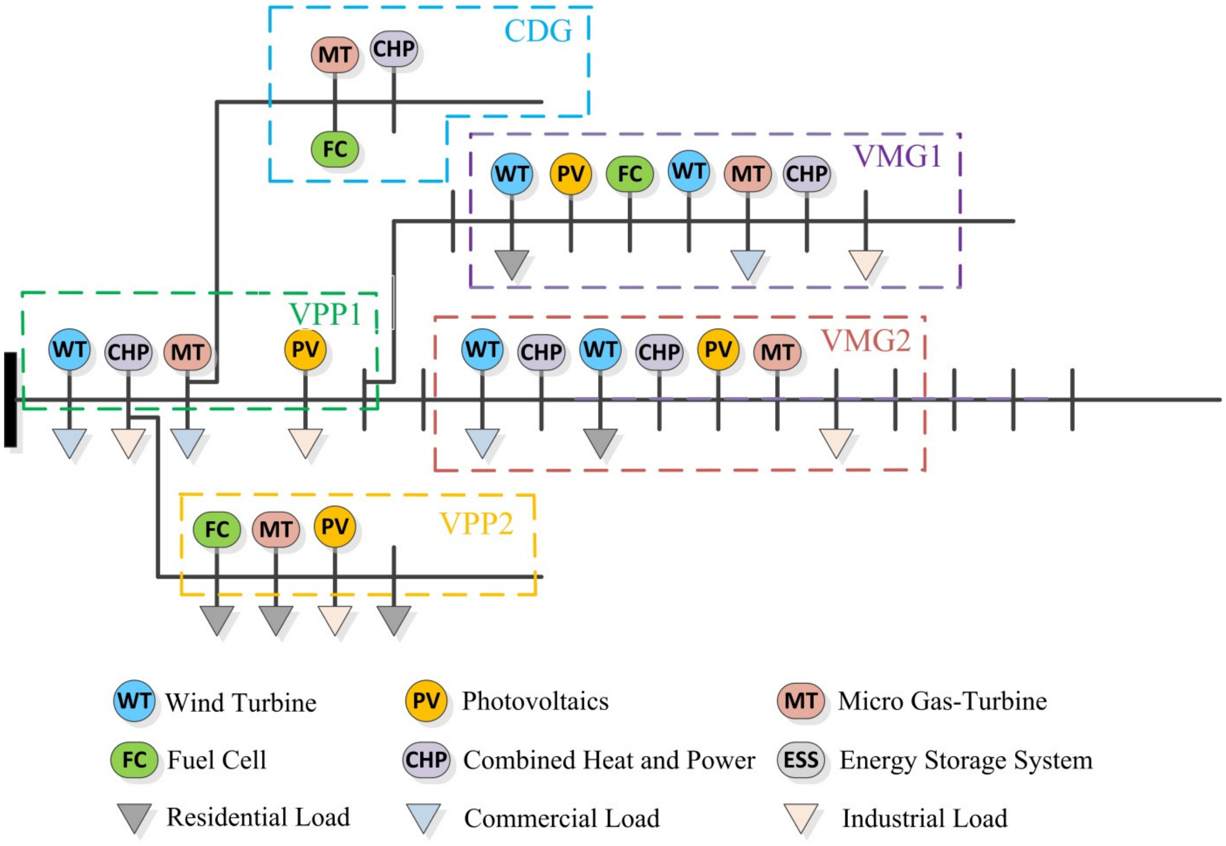


Fig.3 Configuration of Modified IEEE 33-Node Bus Network

TABLE II
Characteristics of Controllable Generators

Types of Generators	Capacity (MW)	Generation Cost (RMB/kWh)
CHP	0.8	$C_{CHP} = 0.0005(P_{CHP})^2 + 0.0495P_{CHP} + 0.001$
FC	0.6	$C_{MT} = 0.1976$
MT	0.5	$C_{MT} = 0.02(P_{MT})^2 + 0.01P_{MT} + 0.02$

TABLE III
Carbon Emission Factors and Quota Allocations

Types of Generators	Emission Factor (kg/KWh$_e$)	Allocated Emission Quotas (kg/h)	
CHP	0.667	CDG: 550.32	VPP1: 389.68
FC	0.543	VMG1: 550.32	VPP2: 550.32
MT	0.407	VMG2: 516.56	

4.2 Computational Performance of Different MARL Algorithms

In this section, the computational performance of the proposed MLWoLF-PHC algorithm, the conventional WoLF-PHC algorithm and Q-learning algorithm are compared in terms of their convergence speed with and without consideration of uncertainties. For simplicity, these algorithms are run to simulate the bidding procedure at 12.00AM in the EM only, in which the total energy demand is 5.66MW. Due to the necessity of discretization of state and action space, decisions of bidding price are discretized into 10 integer values, while states of allocated quantities are discretized into 8 integer values. For the hyper-parameter settings, Fig.4 shows the iterations for convergence with different hyper-parameters, and the optimal values are listed in Table IV.

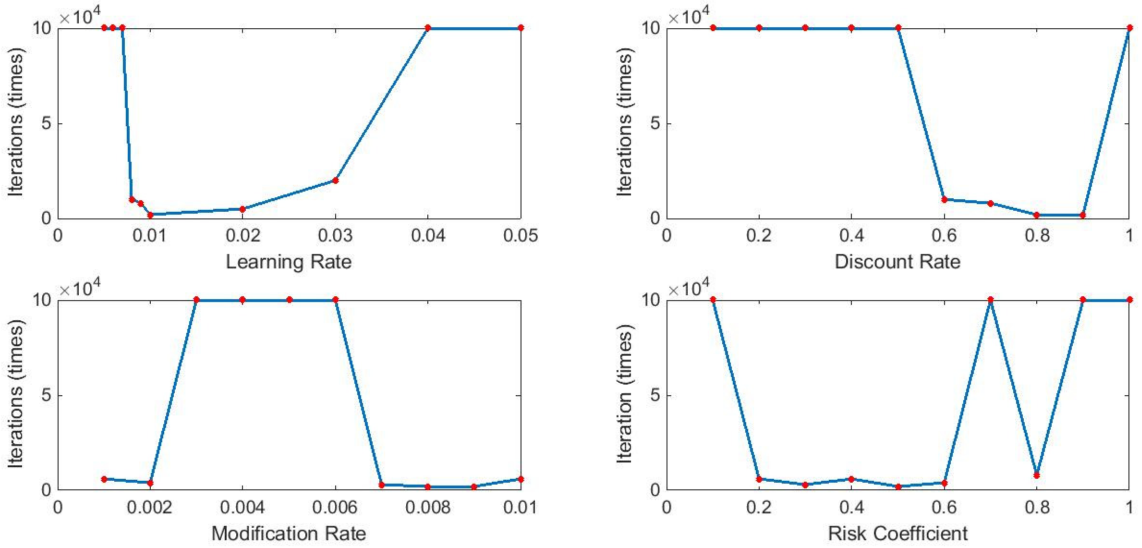


Fig.4 Iterations for Convergence with Different Hyper-Parameters

TABLE IV
Hyper-Parameter Settings of MLWoLF-PHC

Types of Hyper-parameters	Values	Types of Hyper-parameters	Values
Learning Rate	0.01	Modification Rate	0.008
Discount Rate	0.8	Risk Coefficient	0.5

In Case 1 (a1, a2, a3), the condition without considering uncertainties, i.e. all the DVAs will not be penalized due to the load loss or failure to provide energy to the main grid, is examined. The training procedure is plotted in Fig.5. It is clear that the bidding prices of these five energy producers have reached to the convergence in this condition, while using the MLWoLF-PHC and conventional WoLF-PHC. The reason of converging to the lowest limitation will be discussed in Section 4.3. However, when the Q-Learning algorithm is adopted, the bidding price failed to reach to the convergence within 2×10^4 iterations.

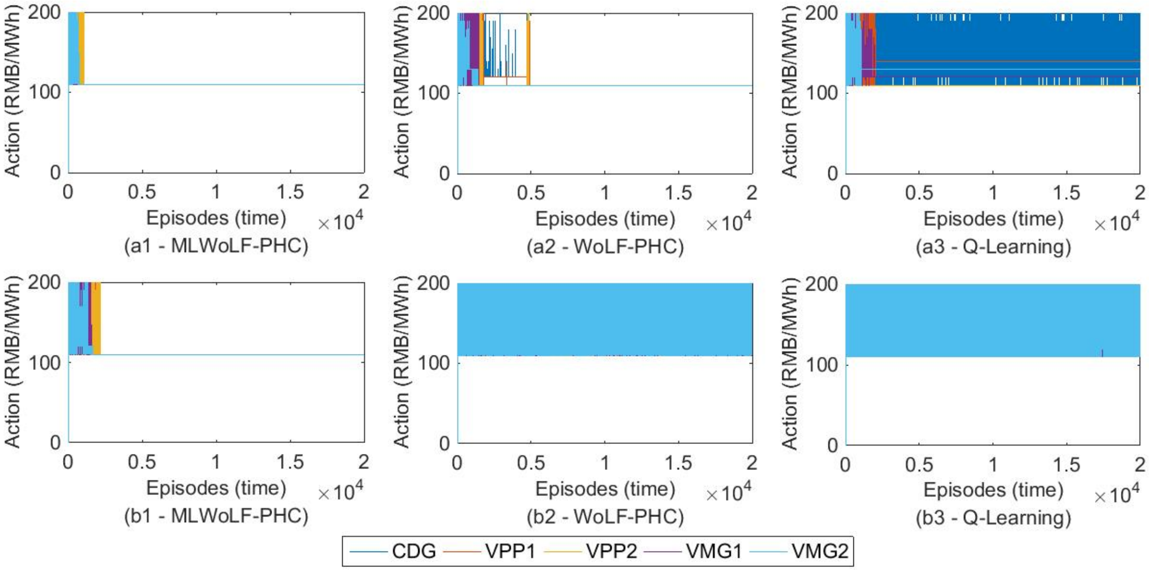


Fig.5 Computational Performance of Different Algorithms

In Case 2 (b1, b2, b3), the performance of these three algorithms with penalties considered is examined. As shown in Fig.5, the proposed MLWoLF-PHC demonstrates a better performance, reaching to the convergence within 2000 iterations, while both the conventional WoLF-PHC and Q-Learning fail to converge. Due to the uncertainty which leads to penalties, the reward of each agent under the same state-action pair may be different, and therefore the convergence performance will be adversely affected. By adopting the proposed MLWoLF-PHC, the fluctuation of Q-value is mitigated due to the intervention of CPF, and therefore the speed of convergence is accelerated. Hence, this proposed method would be the better approach under conditions with uncertainty.

4.3 Equilibrium Analysis in Electricity Market

In this section, the market equilibrium, i.e. the converged results of MCP in both EM and ASM, will be presented and analyzed without considering the CEAM trading. The market equilibrium without consideration of penalty due to uncertainty is presented in Fig.6 and Fig.7.

In the EM, from 1.00AM to 9.00AM, due to the lack of net load which varies from 3.87MW to 4.61MW, the MCP in EM is relatively low, fluctuating between 110RMB/MWh and 120RMB/MWh. Such a low energy price may attribute to the conservative bidding strategy adopted by the DVAs, i.e. all the DVAs prefer to lower their bidding price as a price-taker to secure the successful bidding. With the net load fluctuating between 5.11MW and 7.44MW from 9.00AM to 22.00PM, the MCP also varies from 110RMB/MWh to 200RMB/MWh. It can be observed that, the MCP is strongly correlated with the net load, and the DVAs will prefer the aggressive price-maker bidding strategy when the load demand is high, to pursue more benefit considering the successful bidding can be easily secured due to high demand. However, when the penalty caused by uncertainty is involved, the MCP is apparently increased, especially when the output of RDGs reaches to the peak value from 11.00AM to 16.00PM. As the devotion of RDG generation will result in overwhelming penalty, the willingness of VPPs and VMGs to provide RDG output will be decreased, and therefore the deducted supply will result in the rise of MCP.

Considering that the demand in ASM is much less than that of EM, the incorporation of ASM will not result in the lack of supply in the EM, and thus the MCP will not be significantly affected. For the MCP in ASM, it correlates with the fluctuation of net load, reaching the peak at 12.00PM and 9.00AM when the net load fluctuates sharply.

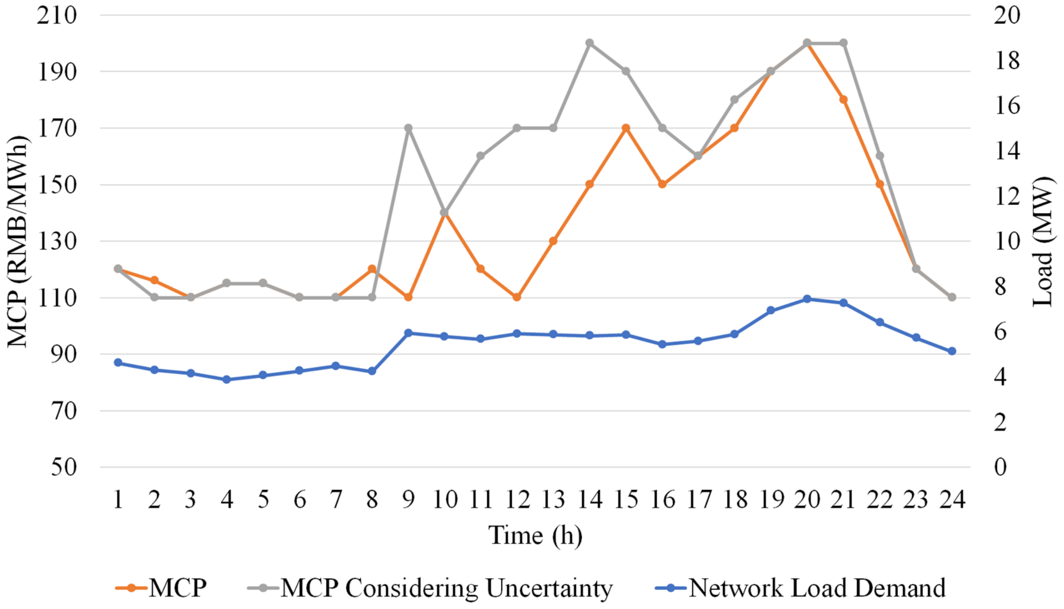


Fig.6 MCP of EM with and without Consideration of Uncertainty

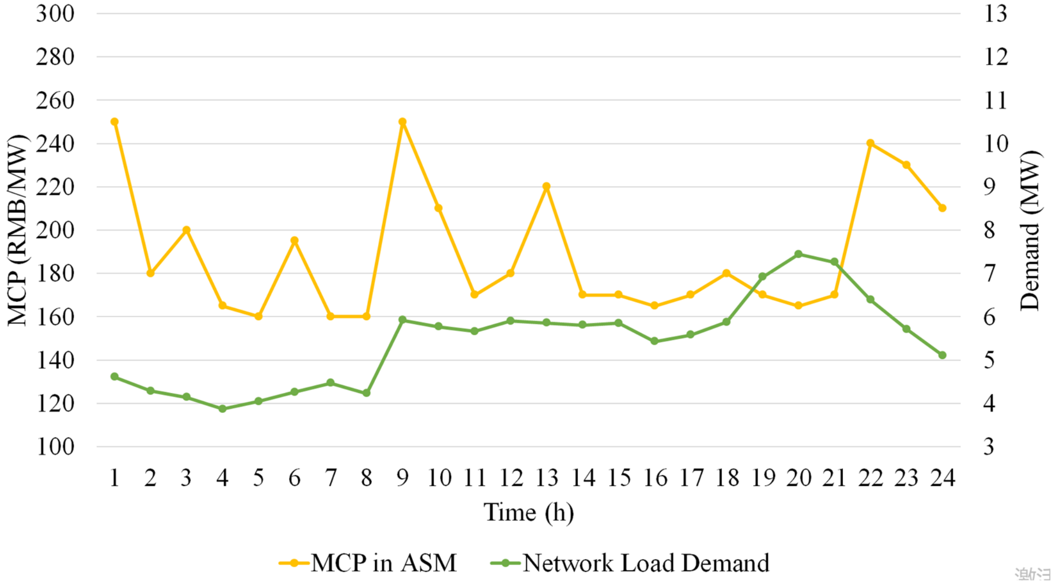


Fig.7 MCP of ASM without Consideration of Uncertainty

4.4 Impacts of CEAM Incorporation on MCP of Electricity Market

This section examines how the incorporation of CEAM would affect the MCP of both EM and ASM. It is assumed that each VA will perform the optimal dispatching using the methodology mentioned in [26]. Hence, the total carbon emission demand and available supply can be obtained based on the converged market clearing results and emission factors in Table III. The MCP in EM and ASM considering the incorporation of CEAM trading can be subsequently simulated using the MLWoLF-PHC algorithm. For the sake of simplicity, the uncertainty of RDG output and load demand is not considered here.

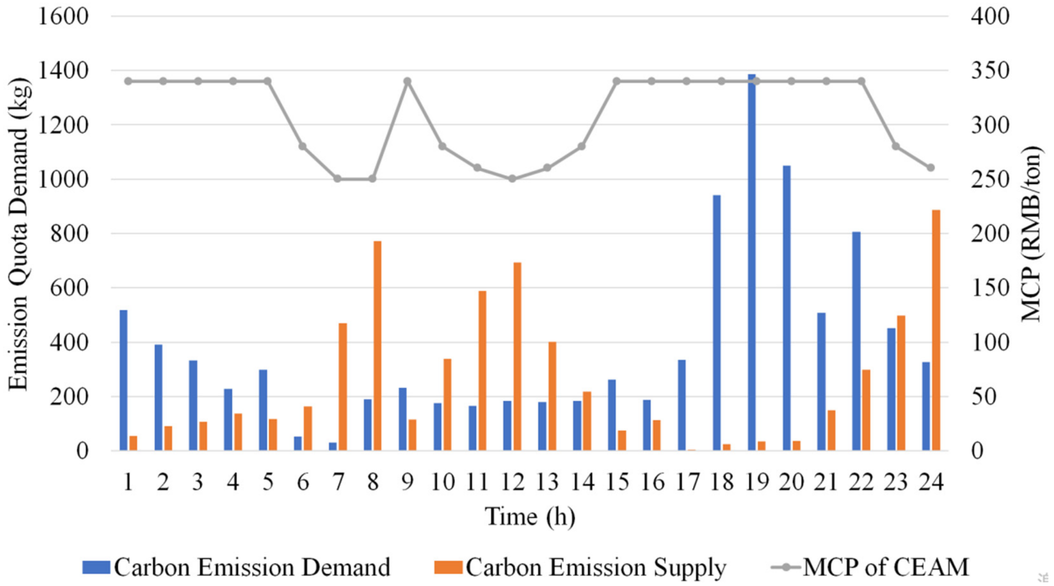


Fig.8 CEAM Demand, Supply and MCP

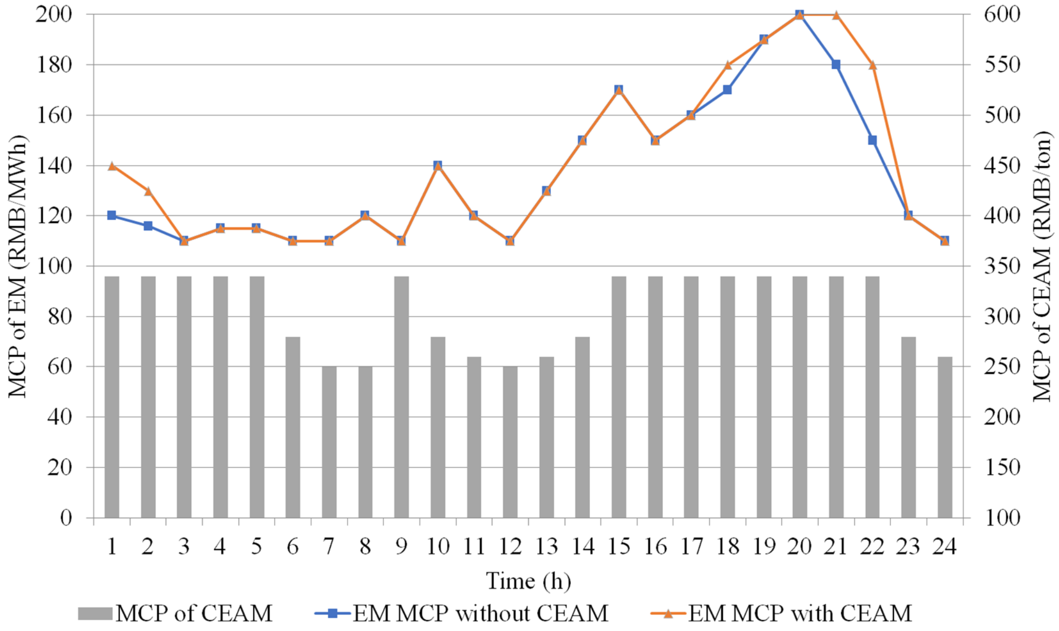


Fig.9 Impacts of CEAM on the MCP of EM

Fig.8 shows the converged MCP of CEAM with the variation of total supply and demand. From 1.00AM to 6.00AM, as well as 15.00PM to 22.00PM, the demand of quota obviously exceeds the supply, and the MCP raises to the maximum limitation, i.e. 340RMB/ton. During these periods, the DVAs may receive less profit because they need to pay more for excessive emissions. Hence, they tend to take two actions to avoid losses: intentionally uplift the bidding price, or claim less available capacity. No matter what actions they take, the MCP of EM will be increased, as shown in Fig.9. As the MCP of AM is only correlated to the load fluctuation, the involvement of CEAM will not have such impacts in AM.

4.5 Discussions

Although MARL algorithms have shown promise in simulating complex systems such as bidding games in electricity markets. However, their performance and reliability can be significantly influenced by factors such as hyper-parameter selection and the non-stationarity of the environment.

Hyper-parameters in MARL algorithms, such as learning rate and discount factor, play a crucial role in the learning process. The learning rate determines how quickly the algorithm updates its knowledge, the discount factor influences the importance of future rewards, and the exploration rate affects the balance between exploration and exploitation. Inappropriate selection of these hyper-parameters can lead to unstable learning processes, slow convergence, or even divergence. For instance, a high learning rate might cause the algorithm to overreact to recent changes, leading to instability, while a low learning rate might make the learning process too slow. In the context of electricity market bidding games, different settings of hyper parameters could result in inconsistent bidding strategies and unpredictable market dynamics.

Meanwhile, the environment in MARL is often non-stationary from the perspective of individual agents, as the policies of other agents are constantly evolving. This non-stationarity can pose significant challenges for the stability and consistency of MARL algorithms. Traditional reinforcement learning algorithms assume a stationary environment, which is not the case in multi-agent settings. As agents continually update their policies, the environment seen by each agent changes, making it difficult for the agents to converge to a stable policy. In the context of electricity market bidding games, this could lead to fluctuating bidding strategies and market prices, and the market might never reach a stable equilibrium.

Another important problem is how to practically use the MARL-based market simulations to facilitate the market design. Generally, the following steps shall be implemented. First, the controversial market design options should be clarified, as the general research question will be “to pick the best combination of market design options with maximum performance”. Second, the market operation that can be simulated by the MARL algorithms, such as the algorithm proposed in this research, should be identified. Finally, some market operation performance indicators should be proposed to assess the performance of different market design options. This is an interesting topic that could be considered a potential future research perspective.

5. Conclusion

The distribution level energy trading consisting of several markets (EM, ASM and CEAM) with participation of multiple stakeholders (DVAs) is an emerging paradigm to be investigated. In this research, the conventional EPEC model is re-written as a Markov Game, is developed to formulate the bidding among DVAs in coordinated markets. This model is solved by the proposed MLWoLF-PHC algorithm, which is a fully-distributed and risk-averse approach considering the potential penalty caused by failure of providing energy or ancillary service due to uncertainty of RDG output in DVAs. Case studies firstly demonstrate the feasibility of the proposed algorithm in solving the dynamic bidding problems. The market equilibrium calculus and analysis are also conducted, showing the strong correlations between MCP in EM and the system net load demand, as well as MCP in AM and the net load fluctuation, while the consideration of potential penalty apparently uplifts the MCP in EM. Finally, impacts of CEAM incorporation on the MCP are revealed. It can be concluded that the involvement of CEAM will increase the MCP of EM when the demand of carbon emission quota apparently exceeds the supply, but the CEAM incorporation will not have impact on the MCP of ASM. In summary, investigating the joint electricity market with consideration of the CEAM can provide valuable insights for a wide range of stakeholders, contributing to more informed decision-making, more effective policy design, and ultimately, a more sustainable energy system.

Acknowledgements

This work is jointly supported by the Research Grants Council of the HKSAR Government (Grant No. R5020-18), the Innovation and Technology Commission of the HKSAR Government to the Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center (Grant No. K-BBY1), The Hong Kong Polytechnic University and the National Natural Science Foundation of China (52077075).

References

[1] T. Lv and Q. Ai, “Interactive energy management of networked microgrids-based active distribution system considering large-scale integration of renewable energy resources,” *Applied Energy*, vol. 163, pp. 408-422, 2016.

[2] M. Xie, X. Ji, X. Hu, P. Cheng, Y. Du, and M. Liu, “Autonomous optimized economic dispatch of active distribution system with multi-microgrids,” *Energy*, vol. 153, pp. 479-489, 2018.

[3] S. M. Nosratabadi, R.-A. Hooshmand, and E. Gholipour, “A comprehensive review on microgrid and virtual power plant concepts employed for distributed energy resources scheduling in power systems,” *Renewable & sustainable energy reviews*, vol. 67, pp. 341-363, 2017

[4] M. Kara *et al.*, “The impacts of EU CO2 emissions trading on electricity markets and electricity consumers in Finland,” *Energy economics*, vol. 30, no. 2, pp. 193-211, 2008, doi: 10.1016/j.eneco.2006.04.001.

[5] W. Zhang, J. Li, G. Li, and S. Gao, “Emission reduction effect and carbon market efficiency of carbon emissions trading policy in China,” in *Energy (Oxford)*, vol. 196, p. 117117, 2020

[6] A. Ehsan and Q. Yang, “State-of-the-art techniques for modelling of uncertainties in active distribution network planning: A review,” in *Applied energy*, vol. 239, pp. 1509-1523, 2019

[7] K. Oikonomou, M. Parvania and R. Khatami, "Deliverable Energy Flexibility Scheduling for Active Distribution Networks," in *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 655-664, Jan. 2020

[8] M. Hu, F. Xiao, and S. Wang, “Neighborhood-level coordination and negotiation techniques for managing demand-side flexibility in residential microgrids,” *Renewable & sustainable energy reviews*, vol. 135, p. 110248, 2021

[9] S. Bahramara, P. Sheikhahmadi, A. Mazza, G. Chicco, M. Shafie-khah and J. P. S. Catalão, "A Risk-Based Decision Framework for the Distribution Company in Mutual Interaction With the Wholesale Day-Ahead Market and Microgrids," in *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 764-778, Feb. 2020

[10] Z. Yi, Y. Xu, J. Zhou, W. Wu and H. Sun, "Bi-Level Programming for Optimal Operation of an Active Distribution Network With Multiple Virtual Power Plants," in *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2855-2869, Oct. 2020, doi: 10.1109/TSTE.2020.2980317.

[11] T. Zhao, X. Pan, S. Yao, C. Ju and L. Li, "Strategic Bidding of Hybrid AC/DC Microgrid Embedded Energy Hubs: A Two-Stage Chance Constrained Stochastic Programming Approach," in *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 116-125, Jan. 2020

[12] J. Wang *et al.*, “Optimal bidding strategy for microgrids in joint energy and ancillary service markets considering flexible ramping products,” *Applied energy*, vol. 205, pp. 294-303, 2017, doi: 10.1016/j.apenergy.2017.07.047.

[13] B. Liu, M. Wang, J. Men and D. Yang, "Microgrid Trading Game Model Based on Blockchain Technology and Optimized Particle Swarm Algorithm," in *IEEE Access*, vol. 8, pp. 225602-225612, 2020.

[14] S. Wogrin, J. Barquin, and E. Centeno, “Capacity Expansion Equilibria in Liberalized Electricity Markets: An EPEC Approach,” *IEEE transactions on power systems*, vol. 28, no. 2, pp. 1531–1539, 2013, doi: 10.1109/TPWRS.2012.2217510.

[15] L. Guo, G.-H. Lin, D. Zhang, and D. Zhu, “An MPEC reformulation of an EPEC model for electricity markets,” *Operations research letters*, vol. 43, no. 3, pp. 262–267, 2015, doi: 10.1016/j.orl.2015.03.001.

[16] T. Dai and W. Qiao, “Finding Equilibria in the Pool-Based Electricity Market With Strategic Wind Power Producers and Network Constraints,” *IEEE transactions on power systems*, vol. 32, no. 1, pp. 389–399, 2017, doi: 10.1109/TPWRS.2016.2549003.

[17] H. Chen *et al.*, “Distribution Market-Clearing and Pricing Considering Coordination of DSOs and ISO: An EPEC Approach,” *IEEE transactions on smart grid*, vol. 12, no. 4, pp. 3150–3162, 2021, doi: 10.1109/TSG.2021.3061282.

[18] Q. Hong, F. Meng, J. Liu, and R. Bo, “A bilevel game-theoretic decision-making framework for strategic retailers in both local and wholesale electricity markets,” *Applied energy*, vol. 330, p. 120311–, 2023, doi: 10.1016/j.apenergy.2022.120311.

[19] X. Liu and A. J. Conejo, “Single-Level Electricity Market Equilibrium With Offers and Bids in Energy and Price,” *IEEE transactions on power systems*, vol. 36, no. 5, pp. 4185–4193, 2021, doi: 10.1109/TPWRS.2021.3054936.

[20] B. Fanzeres, A. Street, and D. Pozo, “A Column-and-Constraint Generation Algorithm to Find Nash Equilibrium in Pool-Based Electricity Markets,” *Electric power systems research*, vol. 189, p. 106806–, 2020, doi: 10.1016/j.epsr.2020.106806.

[21] C. N. Dimitriadis, E. G. Tsimopoulos, and M. C. Georgiadis, “A review on the complementarity modelling in competitive electricity markets,” *Energies (Basel)*, vol. 14, no. 21, p. 7133–, 2021, doi: 10.3390/en14217133.

[22] D. Pozo and J. Contreras, “Finding Multiple Nash Equilibria in Pool-Based Markets: A Stochastic EPEC Approach,” *IEEE transactions on power systems*, vol. 26, no. 3, pp. 1744–1752, 2011, doi: 10.1109/TPWRS.2010.2098425.

[23] C. Ruiz, A. J. Conejo, and Y. Smeers, “Equilibria in an Oligopolistic Electricity Pool With Stepwise Offer Curves,” *IEEE transactions on power systems*, vol. 27, no. 2, pp. 752–761, 2012, doi: 10.1109/TPWRS.2011.2170439.

[24] H. S. V. S. K. Nunna, A. Sesetti, A. K. Rathore and S. Doolla, "Multiagent-Based Energy Trading Platform for Energy Storage Systems in Distribution Systems With Interconnected Microgrids," in *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 3207-3217, May-June 2020, doi: 10.1109/TIA.2020.2979782.

[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, p. 107714, 2022, doi: 10.1016/j.epsr.2021.107714.

[26] X. Wang, Y. Liu, J. Zhao, C. Liu, J. Liu, and J. Yan, “Surrogate model enabled deep reinforcement learning for hybrid energy community operation,” *Applied Energy*, vol. 289, p. 116722, 2021, doi: 10.1016/j.apenergy.2021.116722.

[27] M. Parastegari, R. A. Hooshmand, A. Khodabakhshian, and A. H. Zare, “Joint operation of wind farm, photovoltaic, pump-storage and energy storage devices in energy and reserve markets,” *International journal of electrical power & energy systems*, vol. 64, pp. 275-284, 2015, doi: 10.1016/j.ijepes.2014.06.074.

[28] X. R. Li, C. W. Yu, Z. Xu, F. J. Luo, Z. Y. Dong and K. P. Wong, "A Multimarket Decision-Making Framework for GENCO Considering Emission Trading Scheme," in *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4099-4108, Nov. 2013

[29] B. Lin and Z. Jia, “What will China's carbon emission trading market affect with only electricity sector involvement? A CGE based study,” in *Energy economics*, vol. 78, pp. 301-311, 2019

[30] D. Liu et al., “Comprehensive effectiveness assessment of renewable energy generation policy: A partial equilibrium analysis in China,” Energy policy, vol. 115, pp. 330–341, 2018, doi: 10.1016/j.enpol.2018.01.018.

[31] Z. Zhu, K. Wing Chan, S. Bu, B. Zhou, and S. Xia, “Real-Time interaction of active distribution network and virtual microgrids: Market paradigm and data-driven stakeholder behavior analysis,” *Applied energy*, vol. 297, p. 117107, 2021, doi: 10.1016/j.apenergy.2021.117107.

[32] B. Lin and Z. Jia, “Impacts of carbon price level in carbon emission trading market,” in *Applied energy*, vol. 239, pp. 157-170, 2019, doi: 10.1016/j.apenergy.2019.01.194.

[33] X. Fan, X. Lv, J. Yin, L. Tian, and J. Liang, “Multifractality and market efficiency of carbon emission trading market: Analysis using the multifractal de-trended fluctuation technique,” in *Applied energy*, vol. 251, p. 113333, 2019, doi: 10.1016/j.apenergy.2019.113333.

[34] M. E. Baran and F. Wu, "Optimal capacitor placement on radial distribution systems," in *IEEE Transactions on Power Delivery*, vol. 4, no. 1, pp. 725-734, Jan. 1989, doi: 10.1109/61.19265.

[35] *Ajoulabadi*, S. N. Ravadanegh, and Behnam Mohammadi-Ivatloo, “Flexible scheduling of reconfigurable microgrid-based distribution networks considering demand response program,” in *Energy*, vol. 196, 2020.

[36] California Independent System Operator, “Flexible Ramping Products: Revised Draft Final Proposal,” pp. 1-18, 2015.