Day-ahead Pricing Strategy for Virtual Power Plants Considering Privacy Protection of User Data

Xiangyu Kong, Bohan Yang, Peirong Zhang, Wenqi Lu, and Gaohua Liu

Abstract—**As an advanced aggregation and control technology for multi-type source-load resources, virtual power plant (VPP) enables intelligent interconnection and coordinated optimization of distributed load resources. This paper proposes a day-ahead pricing strategy for VPPs considering privacy protection of user data. First, a day-ahead pricing model for VPP is constructed, which considers historical demand response (DR) contributions of users to enhance fairness in the pricing process. The asynchronous advantage actor-critic (A3C) algorithm is adopted to solve this model and optimize pricing decisions. Furthermore, an encrypted horizontal federated learning (HFL) method is proposed for model training, which is based on improved weighted federated averaging algorithm and aims to mitigate data silos and protect privacy of user data across different regions. Instead of directly transmitting user data during training, partially homomorphic encryption (PHE) scheme is applied to encrypt exchanged model parameters, thereby reducing the risk of privacy leakage. Case studies validate that the proposed pricing strategy can enhance the fairness of multi-user regulation decisions and improve the performance of VPPs in day-ahead DR while ensuring the security of user data.**

Index Terms—**Virtual power plant (VPP), demand response (DR), day-ahead pricing, privacy protection, horizontal federated learning (HFL), partially homomorphic encryption (PHE), asynchronous advantage actor-critic (A3C).**

I. Introduction

With the increasing penetration of renewable energy, the new-type power system poses challenges to both source-load security regulation and electricity market reform [1]. On the one hand, the output of renewable energy is highly volatile and stochastic, and its large-scale integration into the power system increases the risk of supply-demand imbalance [2]. On the other hand, the growing presence of diverse demand-side resources and distributed energy resources (DERs) is reshaping the electricity market structure, making it difficult for traditional market mechanisms to effectively stimulate the participation of new market entities. As an advanced technology for resource aggregation and optimal allocation, the virtual power plant (VPP) has attracted widespread attention for enabling flexible participation of DERs in new-type power system [3].

VPPs are commonly employed to aggregate diverse large-scale demand-side resources, enabling organized participation of small- and medium-scale users in demand response (DR) markets [4]. During grid interaction, VPPs must accurately assess the DR potential of users, estimate the load adjustment capacity under given electricity prices, and formulate optimal incentive pricing strategies based on these assessments [5]. Excessively high incentive prices erode the arbitrage margins of VPPs in electricity markets, while insufficient incentives fail to effectively motivate the willingness and DR potential of users, thereby increasing the risk of non-compliance with DR commitments to the grid. Therefore, formulating internal incentive pricing strategies based on the DR potential of large-scale users is crucial for VPPs to balance economic efficiency and DR reliability.

However, DR pricing for VPPs remains challenging issues.

1) Pricing decision challenges in dynamic uncertain environments: the operation of VPPs confronts high-dimensional uncertainties from both internal and external sources. Internally, the DR behavior of aggregated users is stochastic. While some users demonstrate strong willingness and high reliability, others may engage in speculative behavior or lack the capacity for fulfillment, making precise prediction inherently difficult. Externally, the volatility of incentive price signals renders conventional deterministic optimization or rule-based pricing approaches inadequate for dynamically evolving market conditions.

2) Privacy protection challenges of user data: the electricity consumers aggregated within VPPs originate from diverse organizations and geographical locations, creating data silos during joint training. Centralized data collection by VPP operators not only imposes excessive computational burdens but also fails to guarantee robust privacy protection. Traditional privacy-preserving methods such as differential privacy typically rely on noise injection, which inevitably degrades the precision of internal pricing signals and leads to economic sub-optimization. Furthermore, the lack of a high-security framework discourages the user participation, solidifying data silos that hinder the ability of VPPs to aggregate DR resources at scale. Addressing this challenge requires ad-

Manuscript received: October 9, 2025; revised: November 13, 2025; accepted: April 14, 2026. Date of CrossCheck: April 14, 2026. Date of online publication: June 19, 2026.

This work was supported by the National Key R&D Program of China (No. 2021YFB2401200).

X. Kong, B. Yang, P. Zhang, and G. Liu (corresponding author) are with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China (e-mail: eekongxy@tju.edu.cn; ayangbohan@tju.edu.cn; przhang@tju.edu.cn; suppig@126.com).

W. Lu is with the State Grid Tianjin Electric Power Company, Tianjin, China (e-mail: wenqi.lu@tj.sgcc.com.cn).

DOI: 10.35833/MPCE.2025.000971

vanced cryptographic techniques that protect privacy of user data while preserving optimization performance.

Existing studies addressing these challenges can be broadly categorized into two dimensions: the exploration of internal pricing strategies and the development of privacy protection methods for user data. Traditional pricing strategies for VPPs such as flat-rate and time-of-use (TOU) pricing are widely adopted [6], while their limited flexibility restricts applicability in modern power systems with high uncertainty and heterogeneous users [7], particularly in scenarios requiring rapid user-side responses to support renewable energy integration [8]. Existing research on DR pricing can be grouped into three main approaches: optimization-based pricing, game-theoretic pricing, and reinforcement learning (RL)-based dynamic pricing [9].

Optimization-based pricing approaches construct mathematical models to determine the optimal prices. For example, the mixed-integer nonlinear programming models [10], Lagrangian multiplier methods combined with alternating direction method of multipliers (ADMM) [11], and approaches based on Karush-Kuhn-Tucker conditions [12] have been proposed to improve solution efficiency, while piecewise linear approximations are often used to handle nonlinear utility functions [13]. Multi-objective optimization frameworks have also been developed to balance the economic and operational objectives [14]. However, as the scale and heterogeneity of VPP users increase, the computational burden becomes significant, and the reliance on simplified assumptions about user DR behavior limits practical applicability.

Game-theoretic pricing approaches model VPP pricing as strategic interactions between energy suppliers and users, typically formulated as cooperative games aiming to maximize collective welfare [15]. Although these approaches are effective in achieving global optimality, benefit allocation among participants remains a key challenge. Studies have proposed fairness-oriented mechanisms to evaluate individual contributions of users to DR [16], which remains a critical issue for VPP operation [17]. Moreover, the assumption of fully rational users often contradicts real-world behavior characterized by uncertainty and bounded rationality, reducing the effectiveness of such approaches in practice [18].

By contrast, RL-based dynamic pricing approaches can be formulated as a Markov decision process (MDP), where the VPP determines incentive prices, users adjust loads accordingly, and the resulting economic rewards are observed. RL is well suited for modeling such sequential interactions [19]. Early studies mainly rely on dynamic programming or Monte Carlo methods, which suffer from low learning efficiency in stochastic environments [20]. Temporal difference (TD) learning enables continuous decision-making without explicit environment models [21]. Reference [22] integrates long short-term memory (LSTM) networks with RL to enhance DR potential assessment and decision-making, while [23] utilizes a cloud-edge architecture to alleviate computational pressure on centralized RL decision-making. Nevertheless, because prices and user responses are continuous variables, traditional RL based on discretization faces inherent trade-offs between accuracy and efficiency [24]. Deep reinforcement learning (DRL) addresses this limitation by leveraging neural networks to learn the optimal policies in continuous spaces [25]. Reference [26] proposes a DRL method based on prioritized experience replay, which selectively emphasizes high-value samples, thereby significantly improving exploration efficiency and decision performance in continuous domains. Reference [27] proposes a three-stage, distributed, and privacy-oriented optimization method based on DRL that enhances the dynamic interplay between electricity pricing and consumer behavior. However, these methods rely heavily on large-scale user data during training, raising significant concerns regarding data privacy and security.

To resolve the aforementioned conflict between the massive data requirements of DRL and the strict privacy constraints of dispersed users, decentralized architectures have become a research focal point. Federated learning (FL), a decentralized machine learning, enables collaborative model training across multiple parties without requiring the sharing of raw data, making it an ideal solution for addressing privacy preservation and data collaboration challenges in VPPs [28]. Reference [29] proposes a multi-agent game framework combined with federated RL algorithms to prevent privacy leakage during coordinated optimization among agents. Reference [30] employs FL to exchange model training parameters among agents, enabling confidential data interaction. However, standard FL still faces several challenges including communication efficiency, model fairness, and data heterogeneity. By analyzing the model parameters transmitted during the interaction, attackers may still reverse-engineer partial sensitive raw user data [31]. Thus, existing research has introduced privacy-enhancing technologies.

For DR optimization in the power sector, [32] proposes a pricing incentive mechanism based on FL and differential privacy. It selects users with different characteristics for DR participation via Stackelberg game theory while applying differential privacy to protect user data during federated updates. However, the core mechanism of differential privacy makes it challenging to balance efficiency and security, as stronger privacy protection typically results in a significant decline in model accuracy. Reference [33] introduces a federated competitive deep Q-network learning algorithm based on the sine-cosine algorithm, achieving reduced communication latency and superior convergence. Additionally, cryptographic techniques such as homomorphic encryption are often integrated with FL to enhance privacy protection. Reference [34] highlights that homomorphic encryption enables direct aggregation of encrypted model parameters, allowing weight summation without exposing raw information and effectively defending against parameter reconstruction attacks. Reference [35] proposes a resilient distributed algorithm that leverages homomorphic encryption to mitigate privacy leakage and malicious data injection caused by adversarial agents. In summary, standard FL alone cannot fully meet the privacy requirements for user participation in DR within VPP. Compared to decentralized methods that sacrifice global optimality for local privacy, integrating partially homomorphic encryption (PHE) with FL enables precise aggregation of local gradients. This integration addresses the dual challenge of protecting sensitive behavioral patterns while providing the DRL agent with the exact information needed for optimal pricing decisions.

To address the pricing decision-making issues within VPPs, this paper proposes a day-ahead pricing strategy for VPPs considering privacy protection of user data. The main contributions of this paper are as follows.

1) A day-ahead pricing model considering historical DR contributions of users is constructed to address the limitations of conventional pricing schemes that assume user homogeneity and perfect rationality. A contribution factor is constructed by integrating DR quantity, DR participation rate, and compliance and is evaluated using the Mahalanobis distance (MD). The day-ahead pricing model enables accurate identification of highly reliable users and facilitates differentiated incentive strategies, thereby reducing the external compensation risk faced by VPPs due to user non-compliance.

2) An asynchronous advantage actor-critic (A3C) algorithm is proposed to address pricing challenges arising from complex interactions between user DR behavior and market dynamics. The actor network explores pricing strategies within a continuous action space, while the critic network evaluates and guides long-term economic returns following policy execution. This algorithm enables robust and adaptive pricing decisions under stochastic user responses and volatile market conditions.

3) An encrypted horizontal federated learning (HFL) method for model training is proposed to address data silos in distributed load data and security risks associated with parameter transmission in FL. The improved weighted federated averaging algorithm is adopted to adaptively adjust the contribution weights of local models, ensuring reliable pricing decisions while effectively protecting privacy of user data. In addition, gradient parameters are encrypted using PHE scheme to enhance the security and reliability of local model further.

The remainder of this paper is organized as follows. Section II details the day-ahead pricing model considering historical DR contributions of users. Section III elaborates on the encrypted HFL method for model training. Section IV presents the day-ahead pricing process for VPPs. Section V provides the case study. Section VI concludes this paper.

II. Day-ahead Pricing Model Considering Historical DR Contributions of Users

A. Scenario Description of Day-ahead Pricing for VPPs

This paper considers a typical scenario, in which a VPP uses day-ahead incentive price signals to incentivize users to participate in DR, thereby achieving peak shaving in the power system. In this process, particular attention is given to dynamic pricing decisions under uncertain conditions and to privacy protection of user data.

The typical scenario of day-ahead pricing for VPP is illustrated in Fig. 1. During the operating period, the VPP receives peak-shaving requirements from the upper-level trading center. Concurrently, on the demand side, local data are utilized to perform FL model training, thereby securely modeling the user DR behaviors. The VPP then performs day-ahead pricing to formulate internal incentive pricing strategies, which directly support the centralized bidding. Subsequently, during the trading period, the VPP proceeds to broadcast incentive price. This drives user DR to accomplish the required peak shaving, ultimately generating VPP profit through the market differential and concluding with market clearing.

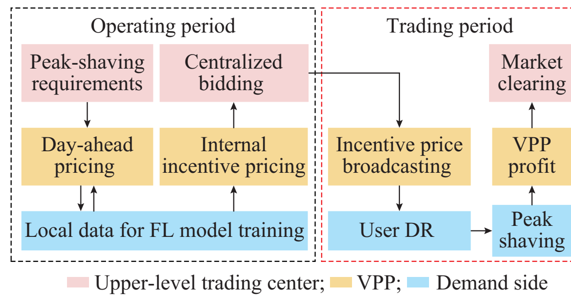

Fig. 1. Typical scenario of day-ahead pricing for VPP.

In this paper, a decision-making method is proposed that differentiates between offline training and online deployment of the day-ahead pricing model. During the offline training phase, the VPP management center does not directly collect user data; instead, it adopts the FL paradigm. During the online deployment phase, the well-trained global model is deployed at the VPP management center to determine and publish optimal day-ahead incentive prices to users based on market signals, without accessing any raw user data.

B. Modeling of User DR Behavior and Contribution Assessment

Unlike the recurrent structures of traditional recurrent neural network (RNN) and LSTM, the Transformer model introduces a self-attention mechanism that directly computes the relational weights between any elements in a sequence and processes the entire sequence in parallel. This mechanism enables the Transformer model to capture long-range dependencies and global contextual information more effectively, thereby enabling more accurate extraction of behavioral features from historical data. It facilitates learning patterns in user DR behavior and simulates user DR under specific pricing inputs.

User DR behaviors exhibit significant differences across time periods. To better match the training process with the practical requirements of day-ahead pricing, this paper divides a day into T time intervals, indexed by t. To predict the load adjustment at target time interval t, the input data $\boldsymbol{I}_t$ can be constructed using the incentive prices and load adjustments over the previous m time intervals, as represented by (1).

$$\boldsymbol{I}_t=[(\lambda_{\mathrm{dr},t-m},d_{t-m}) \quad \dots \quad (\lambda_{\mathrm{dr},t-1},d_{t-1}) \quad (\lambda_{\mathrm{dr},t},0)] \tag{1}$$

where $\lambda_{\mathrm{dr},t}$ is the incentive price at time interval t; d_t is the load adjustment at time interval t; and m is the number of previous time intervals that affect the price of current time interval.

By inputting $\boldsymbol{I}_t$, the load adjustment d_t can be obtained, as shown in Fig. 2.

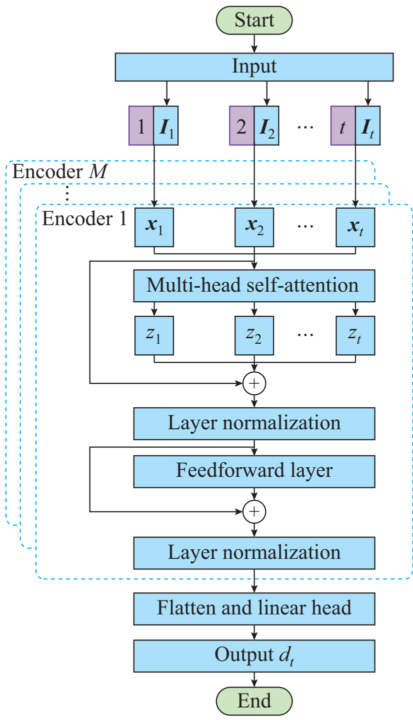

Fig. 2. Transformer-based prediction model of user DR behavior.

The input data are transformed into model-compatible representations through learnable positional encoding:

$$\boldsymbol{x}_t = \boldsymbol{I}_t + PE(Position(\boldsymbol{I}_t)) \tag{2}$$

where $\boldsymbol{x}_t$ is the model-compatible representation of input data at time interval t; $Position(\cdot)$ is the function embedding a relative time-step index to each element within the input sequence, as illustrated by purple blocks in Fig. 2; and $PE(\cdot)$ is the linear projection function for position encoding.

The core of the Transformer model is the self-attention mechanism, which computes similarities between features to capture their correlations, effectively addressing long-range dependencies in time-series data. This mechanism applies weighted attention to the input data, enabling the Transformer model to focus on key information at various positions. The self-attention mechanism $Attention(\cdot)$ is computed as:

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)\boldsymbol{V} \tag{3}$$

where $\boldsymbol{Q}$ is the query vector; $\boldsymbol{K}$ is the key vector; $\boldsymbol{V}$ is the value vector; softmax$(\cdot)$ is a normalization function that converts the scaled similarity scores into attention weights; and d_{k} is the dimensionality of key vector.

The multi-head self-attention mechanism $Multihead(\cdot)$ performs multiple attention operations in parallel, enabling the Transformer model to capture distinct features across multiple subspaces. This helps the Transformer model focus on the influence of information at different spatiotemporal locations in the input sequence, enhancing its ability to model complex dependencies:

$$\boldsymbol{z} = Multihead(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = Concat(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_h)\boldsymbol{W}_O \tag{4}$$

$$\boldsymbol{h}_i = Attention(\boldsymbol{Q}_i, \boldsymbol{K}_i, \boldsymbol{V}_i) \tag{5}$$

where i is the index of attention head; $\boldsymbol{z} = [z_1, z_2, \ldots, z_t]$ is the contextual feature representation obtained from the multi-head self-attention mechanism; $Concat(\cdot)$ is an operation that concatenates the outputs of all attention heads into a single unified representation; $\boldsymbol{h}_i$ is the output of the attention head i; h is the number of attention heads; and $\boldsymbol{W}_O$ is the output weight matrix for multi-head self-attention mechanism.

The multi-head self-attention mechanism captures diverse features of the input sequence in different subspaces and integrates the attention information to provide a more comprehensive feature representation for the Transformer model.

The root mean square error (RMSE) is used as the loss function for the Transformer model during training, expressed as:

$$RMSE = \sqrt{\frac{1}{n'}\sum_{i'=1}^{n'}(d_{i'} - \hat{d}_{i'})^2} \tag{6}$$

where $d_{i'}$ and $\hat{d}_{i'}$ are the actual and predicted DR values of the sample i', respectively; and n' is the total number of samples.

This paper proposes a contribution factor that characterizes DR contributions of users to the VPP. The contribution factor integrates the DR quantity, DR participation rate, and compliance. DR quantity refers to the actual load reduction achieved by the user during the required peak response periods. DR participation rate denotes the ratio of the number of times that a user participates in DR to the total number of required responses within a quarter. Compliance measures the deviation between actual load reduction of user and the committed reduction. The DR participation rate and compliance are calculated as:

$$\rho_a = \frac{f_a}{\sum\limits_{a=1}^{N} f_a} \tag{7}$$

$$\Delta d_a = \sum_{t=1}^{T}\frac{k}{(d_{a,t} - d_{\mathrm{Ac},a,t} + d_{\mathrm{avg},t})^2} \tag{8}$$

where a is the index of users; ρ_a is the DR participation rate of user a within a quarter; f_a is the number of times that user a participates in DR; Δd_a is the compliance of user a; $d_{a,t}$ is the predicted DR quantity of user a at time interval t; $d_{\mathrm{Ac},a,t}$ is the actual load reduction of user a at time interval t; $d_{\mathrm{avg},t}$ is the average DR of all users at time interval t; k is the penalty parameter for non-compliance; and N is the total number of users.

The contribution factor requires a comprehensive evaluation across DR quantity, DR participation rate, and compliance. A simple weighted summation not only makes it difficult to determine weights objectively but also fails to account for interdependencies among these three indicators. To overcome these limitations, the MD is adopted. The MD is scale-invariant, effectively eliminating disturbances introduced by heterogeneous data units. Moreover, it incorporates variable covariance, thereby enabling a more robust characterization of individual sample properties. Consequently, the MD provides a more objective and equitable assessment of actual DR contributions of users. The calculation process is shown as:

$$D_{\mathrm{M}a}=\sqrt{(\boldsymbol{X}_a-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_a-\boldsymbol{\mu})} \tag{9}$$

$$\boldsymbol{X}_a=[d_a \quad \rho_a \quad \Delta d_a] \tag{10}$$

where $D_{\mathrm{M}a}$ is the contribution factor of user a within a quarter computed using MD; $\boldsymbol{X}_a$ is the feature vector of user a; $\boldsymbol{\mu}$ is the mean feature vector of all users; and $\boldsymbol{\Sigma}^{-1}$ is the covariance matrix of the multivariate random variables.

C. Day-ahead Pricing Model for VPPs

The goal of day-ahead controllable load management in a VPP is to encourage users to participate in DR, thereby helping flatten peak loads and reducing operational pressure of the grid during high-demand periods. At the same time, as a profit-seeking entity, the VPP participates in DR bidding with upper-level power operators using a bidding price $\lambda_{\mathrm{vpp},t}$ and sets an internal incentive price $\lambda_{\mathrm{dr},t}$ for its downstream users. The VPP earns profits through the margin between these two prices. The optimization objective is expressed as:

$$\max F=\sum_{t=1}^{T}f(\lambda_{\mathrm{dr},t})=\sum_{t=1}^{T}(f_{\mathrm{b}}(\lambda_{\mathrm{dr},t})-f_{\mathrm{p}}(\lambda_{\mathrm{dr},t})) \tag{11}$$

$$f_{\mathrm{b}}(\lambda_{\mathrm{dr},t})=(\lambda_{\mathrm{vpp},t}-\lambda_{\mathrm{dr},t})\sum_{a=1}^{N}d_{a,t}\eta D_{\mathrm{M}a} \tag{12}$$

$$f_{\mathrm{p}}(\lambda_{\mathrm{dr},t})=\lambda_{\mathrm{p}}\sum_{a=1}^{N}|d_{a,t}-d_{\mathrm{Ac},a,t}| \tag{13}$$

where F is the total profit function of VPP; $f(\lambda_{\mathrm{dr},t})$ is the net profit function of VPP; $f_{\mathrm{b}}(\lambda_{\mathrm{dr},t})$ is the profit function from selling electricity; $f_{\mathrm{p}}(\lambda_{\mathrm{dr},t})$ is the penalty function incurred when the VPP fails to meet the planned load reduction; λ_{p} is the penalty price for non-compliance; and η is a coefficient used to control the influence of user contribution factors.

Given that load adjustment capabilities of users are limited, their adjustments should not exceed their maximum consumption capacity or the load reduction target to avoid energy supply-demand imbalances. Moreover, the internal incentive price set by the VPP must fluctuate within a suitable range. If it is too high, the incentive effectiveness loses; if it is too low, the economic return of VPP is reduced. Therefore, the objective must satisfy the following constraint:

$$\begin{cases}\lambda_{\min,t}\leq\lambda_{\mathrm{dr},t}\leq\lambda_{\max,t}\\ 0\leq d_{a,t}\leq\min(\delta' d_{\max,a,t},d_{\mathrm{target},t})\end{cases} \tag{14}$$

where $\lambda_{\min,t}$ and $\lambda_{\max,t}$ are the minimum and maximum bounds of internal incentive price at time interval t, respectively; δ' is the maximum load adjustment coefficient; $d_{\max,a,t}$ is the maximum load adjustment capacity of user a at time interval t; and $d_{\mathrm{target},t}$ is the load reduction target for VPP at time interval t.

D. Solution of Day-ahead Pricing Model Based on A3C Algorithm

The A3C algorithm involves four key elements: state space $S=\{s_x\}$, action space $A=\{a_x\}$, reward $R=\{r_x\}$, and state transition probability $P(s_{x+1}|s_x,a_x)$, where the state s_x in the x^{th} iteration includes information such as the DR target and bidding outcomes; the action a_x in the x^{th} iteration refers to the internal incentive price set by the VPP; the reward r_x in the x^{th} iteration is the profit obtained by the VPP through DR; and the state transition probability $P(s_{x+1}|s_x,a_x)$ is the probability of transitioning from states s_x to s_{x+1} after taking action a_x [36].

The pricing decision in the A3C algorithm is made by a policy network consisting of an actor network and a critic network. The actor network performs the pricing decision, outputting a probability distribution over the possible pricing actions. The critic network evaluates the expected reward $V(s_x)$ that can be obtained by following the policy of the actor network, and guides the actor network toward more rewarding decisions [37]. The performance of pricing decision is evaluated as:

$$V(s_x)=E[R_x|s_x] \tag{15}$$

$$R_x=\sum_{t=1}^{T}\gamma^t f(\lambda_{\mathrm{dr},t}) \tag{16}$$

where $E[\cdot]$ is the expectation function; and γ^t is the discount factor at time interval t.

Since directly computing the true value of $V(s_x)$ is difficult, the A3C algorithm instead uses the TD error to approximate the difference between the output of critic network and the actual return. The goal is to minimize the TD error δ:

$$\delta=r_x+\gamma V(s_{x+1})-V(s_x) \tag{17}$$

After obtaining the internal incentive price $\lambda_{\mathrm{dr},t}$ within the VPP through the actor network, the Transformer-based prediction model of user DR behavior is employed to simulate the corresponding load adjustment d_t. The reward r_x in the x^{th} iteration is then computed, and the system transitions to the subsequent state s_{x+1}. Based on s_x and s_{x+1}, the critic network calculates $V(s_x)$, $V(s_{x+1})$, and δ. The expected reward $V(s_x)$ directly influences the future pricing actions of the actor network, guiding its training. If a pricing action results in a reward higher than expected, it is considered a positive experience for training.

The loss function and gradient calculations of the critic network are defined as:

$$loss_{\mathrm{critic}}=\delta^2 \tag{18}$$

$$\mathrm{d}\varepsilon\leftarrow\mathrm{d}\varepsilon+\beta\frac{\partial(\delta^2)}{\partial\varepsilon} \tag{19}$$

where $loss_{\mathrm{critic}}$ is the loss function of the critic network; ε is the gradient parameter of the critic network; and β is the learning rate of the critic network. The training objective of the critic network is to minimize the TD error.

The loss function and gradient calculations of the actor network are defined as:

$$loss_{\mathrm{actor}}=\delta\cdot\ln P(a_x|s_x,\theta)+cH(P(s_x,\theta)) \tag{20}$$

$$\mathrm{d}\theta\leftarrow\mathrm{d}\theta+\alpha\delta\nabla_\theta\ln P(a_x|s_x,\theta)+c\nabla_\theta H(P(s_x,\theta)) \tag{21}$$

where $loss_{\text{actor}}$ is the loss function of the actor network; θ is the gradient parameter of the actor network; α is the learning rate of the actor network; H is the entropy of action distribution; c is the weighting coefficient; $P(a_x|s_x,\theta)$ is the policy probability generated by actor network, representing the probability of selecting action a_x under state s_x; and $P(s_x,\theta)$ is the action probability generated by the actor network under state s_x.

The pricing flowchart for VPPs based on the A3C algorithm is illustrated in Fig. 3, where s_{terminal} is the terminal state of the interaction process; X is the total number of iterations; and Q is the cumulative reward used in the training process of A3C algorithm. This pricing process primarily consists of two key steps: price selection by the actor network and policy evaluation by the critic network. Note that the superscript t is introduced to each variable to indicate the inner time interval within a single iteration.

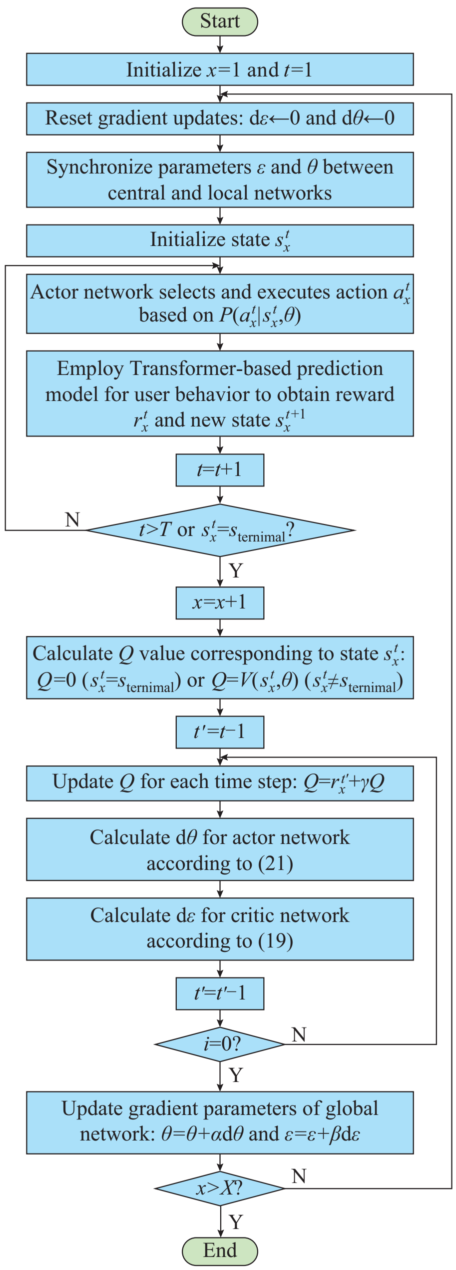

Fig. 3. Pricing flowchart for VPPs based on A3C algorithm.

III. Encrypted HFL Method for Model Training

This paper proposes an encrypted HFL method for model training, enhancing security in both data and model aspects during the VPP pricing process. For privacy protection of user data, the improved weighted federated averaging algorithm is adopted; for model encryption, the PHE scheme is adopted.

A. Improved Weighted Federated Averaging Algorithm

This paper focuses on a VPP that manages users within the same category or industry, where different units managed by VPP exhibit similar user DR patterns despite differing users. This aligns with the HFL application context.

This paper proposes an improved weighted federated averaging algorithm for aggregating model gradients, which is an extension of the traditional federated averaging algorithm. The core enhancement lies in dividing the training data into testing and training samples. The original training samples are used to train the day-ahead pricing model on the initial global model, employing the traditional weighted averaging algorithm. The testing samples are used to evaluate the accuracy of the previous training step, serving as a measure of both data quality and model training effectiveness. The training accuracy q_n is calculated as:

$$q_n=\frac{C_n}{X_n} \tag{22}$$

where n is the index of units managed by the VPP, and each unit is equipped with a local server (LS) that aggregates the data from its users; C_n is the number of correctly predicted samples by the local model of the LS for the n^{th} unit when compared with the testing samples; and X_n is the total number of testing samples for the n^{th} unit.

Clients with higher-quality data are consistently prioritized, potentially marginalizing those with smaller datasets or higher noise levels. However, in the context of VPP pricing, prioritizing more reliable data sources is essential to ensure stable and effective DR. The primary objective is to enhance the overall accuracy and robustness of the global model by mitigating the impact of contributions that could negatively affect performance.

Assume the VPP manages N' units, with each unit containing $k_n\,(n=1,2,\ldots,N')$ users participating in DR. After obtaining the model training accuracy from each LS, the center server (CS) aggregates the local model parameters using the improved weighted federated averaging algorithm:

$$\bar{\omega}_{x+1}\leftarrow\sum_{n=1}^{N'}\alpha_{n,x+1}\omega_{n,x+1} \tag{23}$$

$$\alpha_{n,x+1}=\frac{k_n q_{n,x+1}^{-L}}{\sum\limits_{n=1}^{N'}k_n q_{n,x+1}^{-L}} \tag{24}$$

where $\bar{\omega}_{x+1}$ is the global model parameter updated in the $(x+1)^{\text{th}}$ iteration, which is also broadcasted to all local models after the x^{th} iteration; $\omega_{n,x}$ is the local model parameter of LS for the n^{th} unit in the x^{th} iteration; $\alpha_{n,x+1}$ is the training quality factor of the n^{th} unit in the $(x+1)^{\text{th}}$ iteration; $q_{n,x+1}$ is the training accuracy of the n^{th} unit in the $(x+1)^{\text{th}}$ iteration obtained from testing; and L is an adjustable hyperparameter used to modulate the influence of the training quality factor.

B. PHE Scheme

Although the improved weighted federated averaging algorithm does not directly transmit DR data of users, it still exposes intermediate results of the computation process, providing no additional structural security for data. If data containing model gradients or other parameters are leaked or stolen, sensitive user information could still be compromised. Thus, this paper enhances the improved weighted federated averaging algorithm using the PHE scheme. The key feature of this scheme is its support for additive homomorphism and scalar multiplicative homomorphism [38]. Compared to fully homomorphic encryption (FHE) scheme, PHE scheme has significantly lower computational overhead and simpler key management, making it more suitable for deployment in VPP environments. Then, under the PHE scheme, the following properties hold:

$$Dec_{\text{sk}}([[\boldsymbol{u}]]\oplus[[\boldsymbol{v}]])=Dec_{\text{sk}}([[\boldsymbol{u}+\boldsymbol{v}]]) \tag{25}$$

$$Dec_{\text{sk}}([[\boldsymbol{u}]]\odot l)=Dec_{\text{sk}}([[\boldsymbol{u}\cdot l]]) \tag{26}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are the model gradient parameters; $[[\cdot]]$ is the symbol of homomorphic encryption; Dec_{sk} is the decryption function using secret key; $\oplus$ denotes the multiplication operation that can be performed on ciphertexts; $\odot$ denotes the scalar multiplication operation of ciphertexts; and l is a scalar.

C. Model Training Considering Data and Model Encryption

Based on the improved weighted federated averaging algorithm and the PHE scheme described above, this paper proposes an encrypted HFL method for model training. The process consists of four stages: parameter distribution, local training, gradient update, and parameter aggregation.

First, during the parameter distribution stage, the CS initializes the global model parameters and transmits them to each LS. During global model updates, the CS broadcasts the updated global parameters of the pricing model. Before the parameter distribution, the CS encrypts these parameters using the PHE scheme.

After decrypting the received global parameters, each LS uses the DR data from users within its local unit to directly train the day-ahead pricing model for the VPP. In this step, no data are shared between LSs; instead, each LS conducts local training independently, effectively preventing the privacy leakage of energy consumption.

During the local training stage, gradients and other parameters are continuously updated and computed. Once the training is completed, the updated model parameters are obtained as:

$$\omega_{n,x}=\omega_{n,x-1}-\xi\nabla f(\omega_{n,x-1}) \tag{27}$$

where ξ is the learning rate of algorithm iteration; and $\nabla f(\cdot)$ is the gradient descent computation function.

When the local training reaches the maximum number of iterations, each LS applies PHE scheme to the trained gradient and parameter information before uploading it to the CS.

During the parameter aggregation stage, CS aggregates the local model parameters using the improved weighted federated averaging algorithm, as shown in (23) and (24), to generate the updated global model parameters. These are then encrypted and sent back to the LSs. Each LS decrypts the received global model parameters, updates its own DR pricing model, and continues training using local user data.

The above process is repeated until the predefined maximum number of communication rounds is reached, completing the model training process using encrypted HFL method.

IV. Day-ahead Pricing Process for DR

The day-ahead pricing process for DR is shown in Fig. 4, which systematically progresses from the day-ahead pricing model and user contribution evaluation, followed by the encrypted HFL method for model training and application of day-ahead pricing model, and ends with the day-ahead DR execution and result settlement.

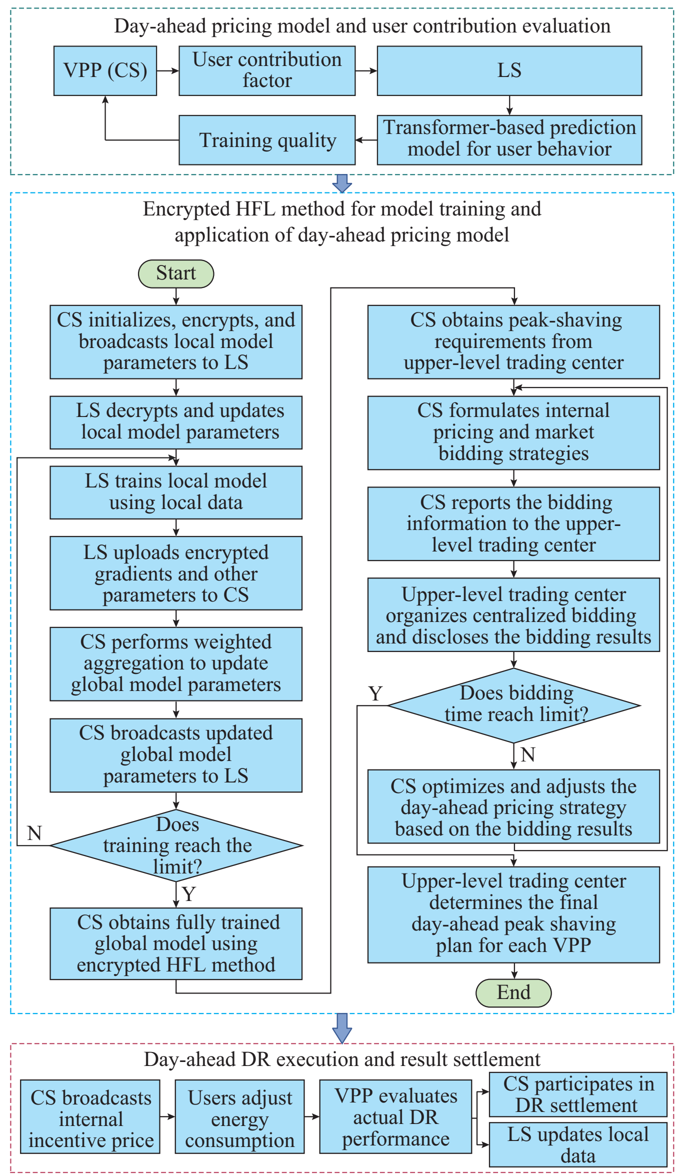

Fig. 4. Day-ahead pricing process for DR.

V. Case Study

A. Scenario and Parameter Settings

1) Scenario Setup

To verify the effectiveness of the proposed day-ahead pricing strategy, this paper conducts a case study using a DR dataset from May 17, 2020 to May 16, 2021 in a certain city. This dataset is derived from a VPP demonstration project that includes 900 users. Based on the energy usage and DR behavior characteristics, users are divided into Groups A, B, and C, each directly managed by different LSs. Group A comprises large-scale commercial users with high and predictable daytime loads. Group B includes small- and medium-scale industrial users exhibiting significant load variability. Group C represents residential users distinguished by a pronounced evening peak. This dataset is split into training, validation, and testing sets at a ratio of 7:1:2. The testing is carried out in the following three scenarios.

Scenario 1: the VPP does not participate in DR, and the energy consumption is priced at 0.75 ¥/kWh.

Scenario 2: the VPP adopts a TOU pricing scheme. The electricity prices for the peak, normal, and off-peak periods are 0.95 ¥/kWh, 0.70 ¥/kWh, and 0.65 ¥/kWh, respectively.

Scenario 3: the VPP performs day-ahead pricing. The pricing period is 24 hours, with a time resolution of 24 points per day. The day-ahead pricing range is $[0.5, 1.25]$¥/kWh.

During the testing, the time periods are categorized as follows: peak periods are 09:00-14:00 and 19:00-23:00; mid-peak periods are 07:00-09:00 and 14:00-19:00; and off-peak periods are 23:00-07:00. For day-ahead DR aimed at peak shaving, the upper limit of peak load is set to be 4200 kW, and the peak-to-valley load difference threshold is defined as 1300 kW.

2) Parameter Setup

This paper employs a Transformer-based prediction model of user DR behavior, and its basic parameter settings are detailed in Table I. Since user data within the VPP are divided into three groups, the number of edge trainers is set to be 3. The parameter settings of the A3C algorithm are provided in Table II. The parameter settings of the encrypted HFL method are shown in Table III.

TABLE I
Basic Parameter Settings of Transformer-based Prediction Model

Parameter	Setting value
Batch size	32
Optimizer	Adam
Learning rate	0.0005
Number of encoder layers	2
Number of attention heads	4
Model dimension	128
Feed-forward network dimension	512
Training iterations	200
Dropout rate	0.1

TABLE II
Parameter Settings of A3C Algorithm

Parameter	Setting value
Discount factor γ	0.90
Optimizer	Adam
Entropy coefficient c	0.015
Learning rate of actor network	0.005
Learning rate of critic network	0.05
Number of neurons (actor network)	100, 100, 50
Number of neurons (critic network)	100, 50, 1
Activation function (actor network)	ReLU function or Softmax function
Activation function (critic network)	ReLU function

TABLE III
Parameter Settings of Encrypted HFL Method

Parameter	Setting value
Training environment	Pysyft 0.2
Number of clients per round	3
Number of local training epochs	5
Number of global aggregation rounds	200
Learning rate	0.001
Decay rate	0.2

The proposed day-ahead pricing strategy is implemented in Python 3.7 using the TensorFlow 2.2 framework and Keras 2.3.1. The experiments are conducted on a computer equipped with an AMD Ryzen 7 5800H CPU at 3.20 GHz and 16 GB of RAM. The encrypted HFL method is deployed in a single-machine setup.

B. DR Behavior and Pricing Analysis

Figure 5 illustrates the load consumption profiles in different scenarios. In Scenario 1, the load consumption profile shows two distinct peaks (one from 11:00 to 15:00 and another from 19:00 to 21:00) and a notable valley between 00:00 and 05:00. The peak-to-valley load difference reaches 2130 kW. On this day, the VPP management center received seven peak-shaving requests during high-load intervals. These requests form the foundation for day-ahead pricing optimization in Scenario 3. Based on this, the RL agent is configured with an allowable pricing range of $[0.5, 1.25]$¥/kWh, with a step size of 0.05 ¥/kWh for adjustments.

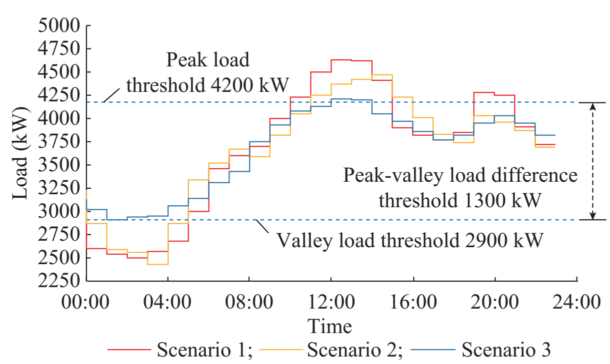

Fig. 5. Load consumption profiles in different scenarios.

In Scenarios 2 and 3, TOU pricing and day-ahead pricing are applied, respectively, to encourage DR and load shifting among users. After the pricing adjustments, the electricity prices and the load adjustments in different scenarios are shown in Figs. 6 and 7, respectively.

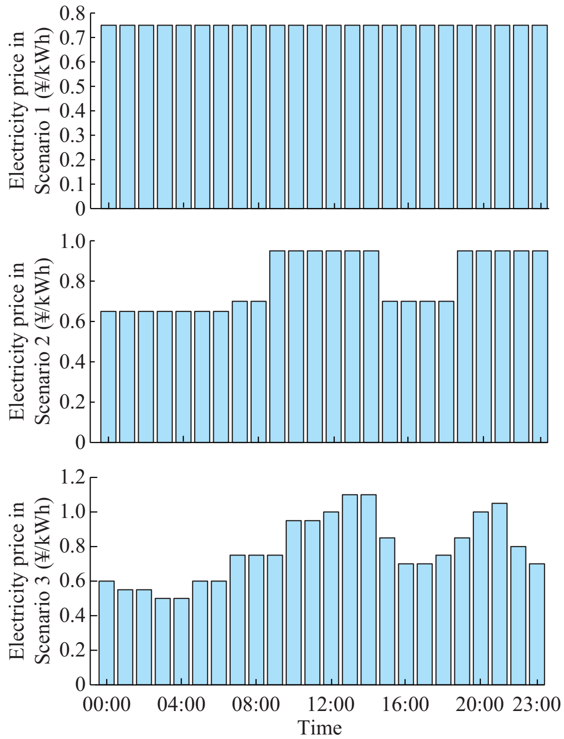

Fig. 6. Electricity prices in different scenarios.

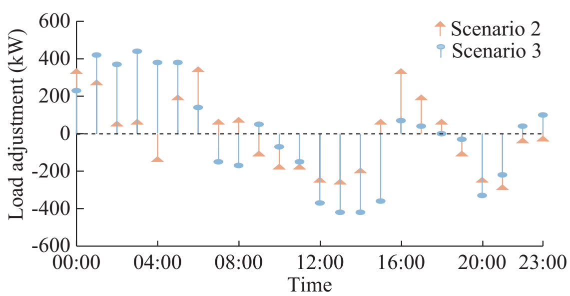

Fig. 7. Load adjustments in different scenarios.

The results indicate that although TOU pricing (Scenario 2) differentiates incentive prices across peak, mid-peak, and off-peak periods, the fixed nature of these time segments prevents adaptation to the actual daily load pattern. Consequently, TOU pricing achieves only a partial reduction of peak loads. On the one hand, the peak load is not fully shifted, and some periods still experience capacity violations. On the other hand, the peak-to-valley load difference in Scenario 2 is 2040 kW, revealing substantial remaining flexibility for load reshaping.

The proposed day-ahead pricing strategy implemented in Scenario 3 performs day-ahead pricing optimization tailored to user DR behavior characteristics across different periods. This strategy sets varying incentive prices and communicates them to users one day in advance to promote coordinated energy consumption. Given the significant price increase during peak periods, users are incentivized to shift their energy consumption to adjacent or off-peak periods. This shift enables users to fulfill their electricity demands at a lower cost without reducing overall usage. Although this strategy is more computationally complex, it offers superior pricing flexibility across time slots. As shown in Fig. 7, Scenario 3 enables more flexible load shifting and reduces the peak-to-valley load difference to 1290 kW, demonstrating effective peak shaving.

To assess the economic viability of the proposed day-ahead pricing strategy, Table IV presents the economic outcomes in different scenarios from the perspectives of both VPP and users.

TABLE IV
ECONOMIC OUTCOMES IN DIFFERENT SCENARIOS

Scenario	Total electricity cost of users (¥)	Settlement profit of VPP (¥)	Penalty cost of VPP (¥)	Net profit of VPP (¥)
1	69178.0	0	0	0
2	68914.0	1885.4	137.6	1747.8
3	68167.5	2722.8	184.0	2538.8

From the perspective of VPP, Scenario 3 achieves a net profit of ¥2538.8, approximately a 45.3% increase compared with ¥1747.8 in Scenario 2, indicating a substantial improvement in profitability. This improvement is primarily attributable to the dynamic pricing, which enables more precise extraction and coordination of DR potential, leading to higher settlement profits in the electricity market. The profit gains considerably outweigh the minor penalty costs incurred due to dense scheduling. From the perspective of users, in Scenario 3, the total electricity cost decreases to ¥68167.5, representing savings of ¥1010.5 compared with that in Scenario 1, more than twice the savings achieved in Scenario 2. Accordingly, the economic feasibility and superiority of the proposed day-ahead pricing strategy in practical electricity market environments are proven.

C. Analysis of Training Methods

A control parameter is adjusted to vary the influence of historical DR contribution of users during pricing optimization. The comparison of user DR accuracy between decentralized and FL training methods under different η is shown in Fig. 8. The decentralized training method exhibits relatively larger prediction errors with average deviations of 7%-9%, corresponding to user DR accuracies of approximately 90%-93%. In contrast, the FL training method achieves higher accuracy, concentrated in the range of 95%-97%, with significantly lower overall deviations. This improvement stems from the limitations of decentralized training method: each group possesses a relatively small dataset and represents a single user type, which constrains the ability to capture the stochastic nature of user DR behavior. As a result, the decentralized training method lacks adaptability to users with diverse consumption patterns. FL, by aggregating parameters through an improved weighted federated averaging algorithm, enables global model optimization. This facilitates more accurate day-ahead pricing strategy tailored to user DR behavior, thereby enhancing the overall effectiveness of DR.

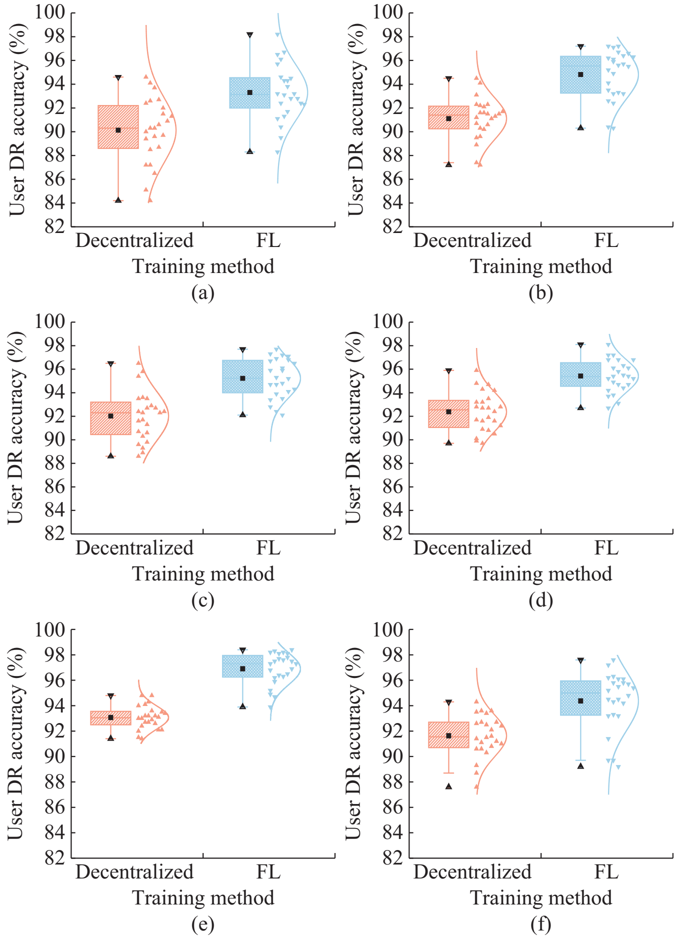

Fig. 8. Comparison of user DR accuracy between decentralized training and FL training methods under different η. (a) $\eta=0$. (b) $\eta=0.25$. (c) $\eta=0.5$. (d) $\eta=0.75$. (e) $\eta=0.8$. (f) $\eta=0.95$.

Furthermore, by analyzing the user DR accuracy associated with varying values of η, the influence of the weighting of user contribution factors on the DR outcome can be observed. A higher value of η increases the emphasis on rewarding users with a strong record of DR participation during the pricing optimization process. As illustrated in Fig. 8, when $\eta=0$ (i.e., historical DR contributions are not considered), the DR performance is comparatively inferior. Gradually increasing η improves both user DR accuracy and the stability of user DR behavior, as reflected in a more concentrated distribution of prediction errors.

However, when η approaches 1, the DR performance deteriorates. This degradation stems from an overreliance on historical behavior, which can disproportionately reduce incentives for users who previously performed poorly, potentially discouraging their future participation and creating a negative feedback cycle. Empirical tests indicate that setting η around 0.8 yields the best balance among high DR accuracy, stable deviation, and optimal overall system performance.

D. Performance Evaluation of Encrypted HFL Method

The convergence comparison of cumulative rewards under different training methods is analyzed in Fig. 9.

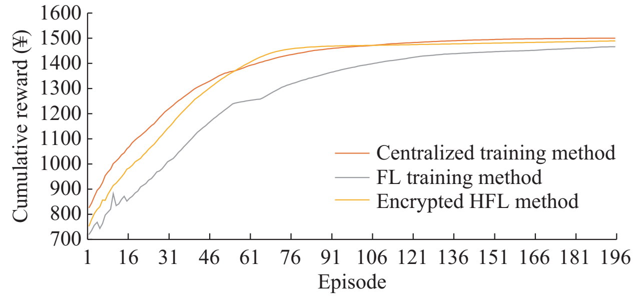

Fig. 9. Convergence comparison of cumulative rewards under different training methods.

Centralized training method converges the fastest and achieves the highest cumulative reward due to full access to global data. In contrast, FL training method exhibits noticeable oscillations between 4 and 20 communication rounds and converges to a lower cumulative reward because of data heterogeneity among LSs. The encrypted HFL method introduces a weighted aggregation mechanism guided by test accuracy, enabling high-quality local updates to contribute more to the global model. Consequently, the convergence becomes more stable during intermediate training, and the final cumulative reward approaches that of centralized training method. These results indicate that the encrypted HFL method preserves privacy without compromising decision performance.

The hyperparameter L controls the influence of local training quality on global parameter aggregation. Sensitivity analysis of hyperparameter L on load prediction accuracy is evaluated using the RMSE metric, as illustrated in Fig. 10. As L increases, the RMSE first decreases and then increases. When $L=1$, the global model degenerates into conventional weighted averaging and is therefore susceptible to noise in local data. When $L=5$, the global model attains its global optimum in terms of load prediction accuracy, effectively filtering high-reliability DR patterns from heterogeneous user data. In contrast, an excessively large L causes the global model to overfit a limited number of high-quality sites, resulting in inadequate representation of behavioral characteristics from other user groups. The results indicate that an appropriate quality-aware weighting setting can effectively alleviate the impact of low-quality local data and provide guidance for selecting the optimal hyperparameter under different levels of data heterogeneity.

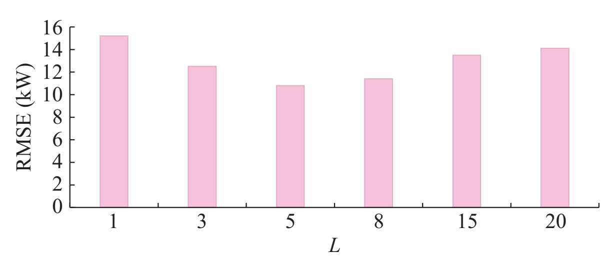

Fig. 10. Sensitivity analysis of hyperparameter L on load prediction accuracy

To validate the performance of the proposed encrypted HFL method, the following comparative training methods are designed.

Method 1: centralized training method. The CS collects all historical load data of users for deep RL training.

Method 2: FL training method. Users perform local training and upload plain-text model gradients to the CS for aggregation.

Method 3: differential privacy FL. Based on Method 2, local models upload gradients with differential privacy applied.

Method 4: local training method. Each user trains their own prediction model solely using their historical load data.

Table V compares the performance of different training methods. Method 1 requires transmission of all raw data, resulting in the largest data transmission volume (DTV) and significant privacy risks, although it achieves the highest aggregate profit. In contrast, Method 4 ensures complete data security by avoiding data transmission but leads to substantially lower economic performance.

TABLE V
PERFORMANCE COMPARISON OF DIFFERENT TRAINING METHODS

Method	PPL	Model structure	DTV (GB)	Time consumption (hour)	Aggregate profit (¥)
1	None	Single global model	1620	1.3	8777.8
2	Moderate	1 global model and 3 local models	1080	2.1	8602.3
3	High	1 global model and 3 local models	1080	2.2	8075.6
4	Complete	900 independent local models	0	1.2	7461.2
Proposed	High	1 global model and 3 local models	8640	8.0	8558.4

Note: PPL is short for privacy protection level.

Both Method 2 and Method 3 transmit model gradients instead of raw data, significantly reducing DTV but requiring more training rounds. Method 3 enhances privacy protection at the cost of reduced profit, which is approximately 5.64% lower than the proposed encrypted HFL method. Method 2 maintains higher economic performance but remains vulnerable to membership inference attacks due to plain-text gradient transmission.

The proposed encrypted HFL method adopts homomorphic encryption for gradient aggregation. Although ciphertext transmission increases communication and computational overhead, it does not affect model accuracy. As a result, its economic performance is comparable to Method 2 and significantly higher than Method 3. These results indicate that the proposed encrypted HFL method effectively balances privacy protection and economic performance, making it suitable for day-ahead pricing strategy for VPP.

VI. CONCLUSION

To ensure the security and privacy of user data for participants in DR within the VPP, this paper proposes a day-ahead pricing strategy considering privacy protection of user data. This strategy leverages the proposed encrypted HFL method to enable global optimization of the day-ahead pricing model within the VPP for users from different regions and entities that exhibit similar DR behaviors, without requiring user data to leave their local environments.

A contribution factor reflecting historical DR contributions of users is incorporated into the pricing decision. This factor assesses the actual performance of each user from DR quantity, DR participation rate, and compliance. An A3C algorithm is then employed to optimize the day-ahead pricing and formulate pricing strategies. This enhances fairness in the pricing process, facilitating load peak-shaving and flexible load shifting through day-ahead DR, thereby effectively improving the long-term economic benefits of day-ahead DR management.

Despite the effectiveness of the proposed day-ahead pricing strategy in privacy protection, several limitations remain.

1) Extending this strategy to real-time balancing or ancillary services markets requires further research.

2) The scalability of FL under larger and more heterogeneous user populations has not been fully evaluated.

3) User comfort is indirectly represented through price incentives without an explicit comfort model, which could be incorporated into future optimization objectives.

REFERENCES

[1] T. Peirelinck, C. Hermans, F. Spiessens *et al*., "Domain randomization for demand response of an electric water heater," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1370-1379, Mar. 2021.

[2] Z. Tan, H. Zhong, Q. Xia *et al*., "Estimating the robust p-q capability of a technical virtual power plant under uncertainties," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4285-4296, Nov. 2020.

[3] S. Yu, F. Fang, Y. Liu *et al*., "Uncertainties of virtual power plant: problems and counter measures," *Applied Energy*, vol. 239, pp. 454-470, Apr. 2019.

[4] J. Li, Q. Ai, and S. Yin, "Market mechanism and foreign experience of virtual power plant participating in peak regulation and frequency-regulation," *Proceedings of the CSEE*, vol. 42, no. 1, pp. 37-56, Jan. 2022.

[5] X. Ge, Z. Fu, F. Xu *et al*., "Business model and key technologies of virtual power plant for new power system," *Automation of Electric Power Systems*, vol. 46, no. 18, pp. 129-146, Sept. 2022.

[6] Z. Cen and J. Wang, "Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer," *Energy*, vol. 169, pp. 160-171, Feb. 2019.

[7] M. Afrasiabi, M. Mohammadi, M. Rastegar *et al*., "Multi-agent microgrid energy management based on deep learning forecaster," *Energy*, vol. 186, p. 115873, Nov. 2019.

[8] J. E, J. Ye, L. He *et al*., "Energy price prediction based on independent component analysis and gated recurrent unit neural network," *Energy*, vol. 189, p. 116278, Dec. 2019.

[9] X. Kong, Y. Ma, Q. Ai *et al*., "Review on power consumption characteristics modeling and load forecasting of power consumption for diverse users in new power system," *Automation of Electric Power Systems*, vol. 47, no. 13, pp. 2-17, Jul. 2023.

[10] C. Yang, X. Du, D. Xu *et al*., "Optimal bidding strategy of renewable-based virtual power plant in the day-ahead market," *International*

Journal of Electrical Power & Energy Systems, vol. 144, p. 108557, Jan. 2023.

[11] B. Stephen, P. Neal, C. Eric *et al*., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1-122, Jan. 2011.

[12] Q. Jin, "On convergence rates of proximal alternating direction method of multipliers," *Journal of Scientific Computing*, vol. 97, no. 3, p. 66, Nov. 2023.

[13] T. Hlalele, J. Zhang, R. Naidoo *et al*., "Multi-objective economic dispatch with residential demand response programme under renewable obligation," *Energy*, vol. 218, p. 119473, Mar. 2022.

[14] H. Zhang, D. Yue, C. Dou *et al*., "PBI based multi-objective optimization via deep reinforcement elite learning strategy for micro-grid dispatch with frequency dynamics," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 488-498, Mar. 2022.

[15] B. Gao, C. Chen, Y. Qin *et al*., "Evolutionary game theoretic analysis for residential users considering integrated demand response," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 6, pp. 1500-1509, Nov. 2021.

[16] Y. Yang, Y. Wang, and W. Wu, "Allocating ex-post deviation cost of virtual power plants in distribution networks," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 3, pp. 1014-1019, May 2023.

[17] W. Sun, X. Liu, W. Xiang *et al*., "Master-slave game based optimal pricing strategy for load aggregator in day-ahead electricity market," *Automation of Electric Power Systems*, vol. 45, no. 1, pp. 159-167, Jan. 2021.

[18] T. Zhang, W. Qiu, Z. Zhang *et al*., "Optimal bidding strategy and profit allocation method for shared energy storage-assisted VPP in joint energy and regulation markets," *Applied Energy*, vol. 329, p. 120158, Jan. 2023.

[19] J. Liu, H. Guo, Y. Li *et al*., "Reinforcement learning based bidding method with high-dimensional bids in electricity markets," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 4, pp. 1373-1382, Jul. 2025.

[20] H. Niu, K. Xu, and C. Liu, "A decomposition-ensemble model with regrouping method and attention-based gated recurrent unit network for energy price prediction," *Energy*, vol. 231, p. 120941, Sept. 2021.

[21] A. Demidovskij, "Exploring neural turing machines applicability in neural-symbolic decision support systems," in *Proceedings of 2021 International Conference on Electrical, Communication, and Computer Engineering*, Kuala Lumpur, Malaysia, Jun. 2021, pp. 1-5.

[22] X. Kong, D. Kong, J. Yao *et al*., "Online pricing of demand response based on long short-term memory and reinforcement learning," *Applied Energy*, vol. 271, p. 114945, Aug. 2020.

[23] Z. Ye, D. Zhang, Z. Wu *et al*., "A3C-based intelligent event triggering control of networked nonlinear unmanned marine vehicles subject to hybrid attacks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12921-12934, Aug. 2021.

[24] Y. Du and F. Li, "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1066-1076, Mar. 2020.

[25] Y. Jiang, J. Dong, and H. Huang, "Optimal bidding strategy for the price-maker virtual power plant in the day-ahead market based on multi-agent twin delayed deep deterministic policy gradient algorithm," *Energy*, vol. 306, p. 132388, Oct. 2024.

[26] Y. Ye, D. Qiu, M. Sun *et al*., "Deep reinforcement learning for strategic bidding in electricity markets," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1343-1355, Mar. 2020.

[27] H. Zhang, D. Yue, C. Dou *et al*., "A three-stage optimal operation strategy of interconnected microgrids with rule-based deep deterministic policy gradient algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 1773-1784, Feb. 2024.

[28] A. Grataloup, S. Jonas, and A. Meyer, "A review of federated learning in renewable energy applications: potential, challenges, and future directions," *Energy and AI*, vol. 17, p. 100375, Sept. 2024.

[29] W. Yang, X. Peng, H. Quan *et al*., "Multi-agent collaborative game optimization strategy for active distribution networks based on federated reinforcement learning," *Automation of Electric Power Systems*, vol. 49, no. 13, pp. 70-82, Jun. 2025.

[30] T. Pu, S. Du, Y. Li *et al*., "Collaborative optimization strategy of distributed generators based on federated reinforcement learning for privacy preservation," *Automation of Electric Power Systems*, vol. 47, no. 8, pp. 62-70, Apr. 2023.

[31] S. Agrawal, S. Sarkar, O. Aouedi *et al*., "Federated learning for intrusion detection system: concepts, challenges and future directions," *Computer Communications*, vol. 195, pp. 346-361, Nov. 2022.

[32] H. Cheng, T. Lu, R. Hao *et al*., "Incentive-based demand response optimization method based on federated learning with a focus on user privacy protection," *Applied Energy*, vol. 358, p. 122570, Mar. 2024.

[33] H. Li, Y. Li, Y. Yang *et al*., "Energy management strategy for multi-microgrid based on improved federated dueling deep Q network," *Automation of Electric Power Systems*, vol. 48, no. 8, pp. 174-184, Apr. 2024.

[34] R. Salavi, M. Math, and U. Kulkarni, "A survey of various cryptographic techniques: from traditional cryptography to fully homomorphic encryption," in *Proceedings of Innovations in Computer Science and Engineering*, Hyderabad, India, Aug. 2018, pp. 295-305.

[35] H. Zhang, C. Yu, M. Zeng *et al*., "Homomorphic encryption-based resilient distributed energy management under cyber-attack of micro-grid with event-triggered mechanism," *IEEE Transactions on Smart Grid*, vol. 15, no. 5, pp. 5115-5126, Sept. 2024.

[36] M. Loschenbrand, "Modeling competition of virtual power plants via deep learning," *Energy*, vol. 214, p. 118870, Jan. 2021.

[37] K. Olsen, Y. Zong, S. You *et al*., "Multi-timescale data-driven method identifying flexibility requirements for scenarios with high penetration of renewables," *Applied Energy*, vol. 264, p. 114702, Apr. 2020.

[38] M. Vahedipour, H. Rashidizadeh, M. Shafie *et al*., "Risk-averse optimal energy and reserve scheduling for virtual power plants incorporating demand response programs," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1405-1415, Mar. 2021.

Xiangyu Kong received the B.S., M.S. and Ph.D. degrees from Tianjin University, Tianjin, China, in 2001, 2004, and 2007, respectively. He is currently a Professor at Tianjin University. His research interests include electricity market and demand-side management, renewable energy generation, virtual power plant, and optimal operation of power system.

Bohan Yang received the B.S. degree from Taiyuan University of Technology, Taiyuan, China, in 2023. He is currently pursuing the M.S. degree in Tianjin University, Tianjin, China. His research interests include frequency support, demand response, and offshore wind power.

Peirong Zhang received the B.S. degree from Tianjin University, Tianjin, China, in 2023. She is currently pursuing the M.S. degree in Tianjin University. Her research interests include virtual power plant and demand-side management.

Wenqi Lu received the B.S. degree and M.S. degree from Tianjin University, Tianjin, China, in 2020 and 2023, respectively. She is currently working in Tianjin Electric Power Company, Tianjin, China. Her research interests include virtual power plant and demand-side management.

Gaohua Liu received the B.S. degree in communication engineering from Qingdao Science and Technology University, Qingdao, China, in 2010, and the M.S. degree in electromagnetic field and microwave technology from Tianjin University, Tianjin, China, in 2013, where she is currently pursuing the Ph.D. degree in information and communication engineering. Her main research interests include deep learning-based behavior analysis and interaction system development.