frontiers | Research Topics

# Future electricity system based on energy internet: Energy storage system design, optimal scheduling, security, attack model and countermeasures

**Edited by**
Dou An, Huan Xi, Hanlin Zhang, Jianhua Yang and Lei Chai

**Published in**
Frontiers in Energy Research

ISSN 1664-8714
ISBN 978-2-8325-2984-3
DOI 10.3389/978-2-8325-2984-3

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Future electricity system based on energy internet: Energy storage system design, optimal scheduling, security, attack model and countermeasures

**Topic editors**

Dou An — Xi'an Jiaotong University, China
Huan Xi — Xi'an Jiaotong University, China
Hanlin Zhang — Qingdao University, China
Jianhua Yang — Columbus State University, United States
Lei Chai — Brunel University London, United Kingdom

**Citation**

An, D., Xi, H., Zhang, H., Yang, J., Chai, L., eds. (2023). *Future electricity system based on energy internet: Energy storage system design, optimal scheduling, security, attack model and countermeasures*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2984-3

# Table of contents

frontiers | Frontiers in Energy Research

TYPE Editorial
PUBLISHED 14 August 2023
DOI 10.3389/fenrg.2023.1261340



OPEN ACCESS

EDITED AND REVIEWED BY
ZhaoYang Dong,
Nanyang Technological University,
Singapore

*CORRESPONDENCE
Huan Xi,
huanxi@xjtu.edu.cn

RECEIVED 19 July 2023
ACCEPTED 07 August 2023
PUBLISHED 14 August 2023

CITATION
An D, Xi H, Yang J and Zhang H (2023), Editorial: Future electricity system based on energy internet: energy storage system design, optimal scheduling, security, attack model and countermeasures.
*Front. Energy Res.* 11:1261340.
doi: 10.3389/fenrg.2023.1261340



# Editorial: Future electricity system based on energy internet: energy storage system design, optimal scheduling, security, attack model and countermeasures

Dou An[1], Huan Xi[2]*, Jianhua Yang[3] and Hanlin Zhang[4]

[1]MOE Key Laboratory for Intelligent Networks and Network Security, Xi'an, China, [2]School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an, China, [3]School of Computer Science, Columbus State University, Columbus, IN, United States, [4]College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China

KEYWORDS

**energy internet, energy storage system design, optimal scheduling, security design, data integrity attack**

Editorial on the Research Topic
Future electricity system based on energy internet: energy storage system design, optimal scheduling, security, attack model and countermeasures

## 1 Introduction

Energy Internet, a futuristic evolution of electricity system, is conceptualized as an energy sharing network. The energy internet integrates advanced sensors, efficient measurement technologies, advanced control methods, efficient energy utilization/conversion/storage system to achieve economical, efficient, and environmentally friendly operation of the power grid system. The energy internet also contains a large amount of heterogeneous information, which requires the more support of information technology in the system design than traditional power systems. Moreover, due to the open network environment of the energy internet, any anomaly or malicious attack in the system can bring unpredictable and significant losses to the overall grid operation.

The Research Topic entitled "Future Electricity System Based on Energy Internet: Energy storage system design, Optimal Scheduling, Security, Attack Model and Countermeasures" aims to investigate energy storage system design, optimal scheduling, attack detection model and the state restoration strategy from the perspective of the energy internet. Moreover, the Research Topic also includes efficient energy utilization, conversion and storage technologies, and cyber-physical attacks against the smart grid from the adversaries' perspective. The researches of this Research Topic are helpful in improving the security and the operation efficiency of power grid system and can be conveniently applied to the real-world security management system of the energy internet. There are in total 13 articles accepted for this Research Topic after careful peer-to-peer review, and they cover the following four categories.

## 1.1 Data integrity attacks against the dynamic state estimation and the interactive energy information

Zhang and Yang (2022) proposed a deep learning-based detection approach against false data injection attacks for dynamic state estimation. In this study, the Kalman filter was used to dynamically estimate the state values from IEEE standard bus systems. A long short-term memory network was utilized to extract the sequential observations from states at multiple time steps. Simulation results in multiple IEEE standard bus systems demonstrated that the proposed detection approach outperforms benchmarks in improving the detection accuracy of malicious attacks. Zhang et al. (2023) developed a detection model of scaling attacks in smart grid considering consumption pattern diversity (*SA2CPD*) to ensure that scaling attacks can be effectively detected when users have multiple consumption patterns. The proposed detection approach leveraged K-means method to distinguish different consumption patterns, divided time periods in every day into two categories based on the binarization values, and used one of them with the greatest information gain to construct a decision tree for judgment. Both theoretical and simulation results based on the GEFCom2012 dataset show that the *SA2CPD* model has a higher *F1 score* than the decision tree model without considering consumption pattern diversity, the KNN model and the Naive Bayes model. Li et al. (2023) formalized the bidding decision problem of EVs into a Markov Decision Process, designed a local Fast Gradient Sign Method which affects the environment and the results of reinforcement learning by changing its own bidding form the perspective of attackers and designed a reinforcement learning training network containing an attack identifier based on the deep neural network. Comprehensive simulation results shown that the proposed attack method will reduce the auction profit by influencing reinforcement learning algorithm, and the protection method will be able to completely resist such attacks.

## 1.2 Artificial intelligence technology-based optimal scheduling of power grid

Cui et al. (2023) modeled charging scheduling problem as a Markov decision process (MDP) and utilized the twin delayed deep deterministic policy gradient algorithm (TD3) to ensure the maximum benefit of the electric vehicle aggregator (EVA), while maintaining minimal fluctuation in the microgrid exchange power. To verify the effectiveness of the proposed method, this paper set up two comparative experiments, using the disorder charging method and deep deterministic policy gradient (DDPG) method, respectively. Results shown that the strategy obtained by TD3 is optimal, which can reduce power purchase cost by 10.9% and reduce power fluctuations by 69.4%. Li et al. (2023) formalized the charging and discharging sequential decision problem of the parking lot into the Markov process and used a Deep Q-Network (DQN)-based reinforcement learning architecture to solve the MDP model. Simulation results with real-world power usage data shown that the proposed method will reduce the peak load by 10% without affecting the travel plan of all electric vehicles. Besides, compared with random charging and discharging scenarios, the proposed method achieved better performance in terms of state-of-charge (SoC) achievement rate and peak load shifting effect. Lv et al. (2023) proposed an optimization method for determining the capacity of energy storage system for smoothing the power output of renewable energy. In this paper, the energy storage configuration model was built according to the objective function and constraints and the genetic algorithm was used to solve the optimization model, obtain the corresponding parameters, and complete the configuration of energy storage capacity. Simulation results shown that at 1 and 10 min, the flattened volatility is about 2% and 5%, while the actual penetration volatility is about 20% and 30%.

## 1.3 Security design for renewable energy utilization

Zhang et al. (2022) proposed several secure multi-party computation (MPC) protocols that enable deep learning training and inference for electricity consumer characteristics identification while keeping the retailer's raw data confidential. Comprehensive experiments based on the Irish Commission for Energy Regulation dataset to verified that the proposed MPC-based protocols have comparable performance in multiple neural network models and optimization strategies. Zhai et al. (2023) put forward a lightweight and dynamic authenticated key agreement and management protocol based on identity cryptosystem and elliptic curve cryptography. The proposed protocol can significantly reduce the computation overhead of the resource-constrained smart meters. Systematic proof of this paper showed that the designed protocol not only guaranteed the confidentiality and integrity of transmitted messages, but also resisted various attacks from an adversary. Gai et al. (2023) proposed a certificateless public auditing scheme for cloud-based smart grid data, which can avoid complicated certificate management and inherent key escrow problems. In order to prevent the disclosure of the private data collected by the smart grid during the phase of auditing, the proposed method used the random masking technology to protect data privacy. The security analysis and the performance evaluation shown that the proposed scheme is secure and efficient. Deng et al. (2023) investigated the problem of system line failures caused by AC or DC blockages from the attacker's perspective and utilized the multiple-feed short-circuit ratio constraint method, output adjustment measures of the energy storage system, sensitivity control, and distance third-segment protection adjustment to reduce system losses from the perspective of dispatch-side defense. Besides, a deep reinforcement learning algorithm was proposed to obtain the Nash equilibrium of the game model. Simulation results verify the appropriateness of the two-stage dynamic zero-sum game model to schedule online defense strategies and the effectiveness and superiority of the energy storage system participating in defense adjustment.

## 1.4 Information technology for the energy internet system

Zhang et al. (2022) addressed the state estimation problem of linear dynamic systems with high-order autoregressive moving

average non-Gaussian noise and proposed a new filter based on correntropy instead of the commonly used minimum mean square error (MMSE) to deal with non-Gaussian noise. Simulation results verify the effectiveness of the proposed algorithm. Xia et al. (2022) designed a new discrete harmonic extractor called quadrature sine wave extractor (QSE), which used the idea of the observer to extract multiple harmonic components at the same time. Compared to the widely used proportional multi-resonant controller, the proposed QSE can reduce current harmonics and improve system stable performance by using it in the current control of grid-connected inverters. Comparative experiments on a three-phase grid-connected inverter verified the effectiveness of the proposed method. Han et al. (2022) proposed a novel automatic modulation classification (AMC) method for low SNR signals. First, the sampled I/Q data is converted to constellation diagram, smoothed pseudo Wigner-Ville distribution (SPWVD), and contour diagram of the spectral correlation function (SCF). Second, convolution auto-encoder (Conv-AE) is used to denoise and extract image feature vectors. Finally, multi-layer perceptron (MLP) is employed to fuse multimodal features to classify signals. Simulation results on RadioML 2016.10A public dataset proved that AMC-MLP provides significantly better classification accuracy of signals in low SNR range than that of other latest deep-learning AMC methods.

## 2 Conclusion

This Research Topic aims to collect and encourage research related to the exploitation and implementation of data integrity attacks, optimal scheduling and security design from the perspective of the energy internet, which aims to improve the security and the operation efficiency of power grid system. Fortunately, this Research Topic has received widespread interests and submissions from the researchers, which published 13 articles in total until its close date. The published articles cover following four categories: data integrity attacks against the dynamic state estimation and the interactive energy information, artificial intelligence technology-based optimal scheduling of power grid, security design for renewable energy utilization and information technology for the energy internet system. The published articles in this Research Topic can be conveniently applied to the real-world security management system of the energy internet.

## Author contributions

DA: Writing-original draft, Writing-review and editing. HX: Writing-original draft, Writing-review and editing. JY: Writing-original draft. HZ: Writing-review and editing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 05 September 2022
DOI 10.3389/fenrg.2022.990267



OPEN ACCESS

EDITED BY
Hanlin Zhang,
Qingdao University, China

REVIEWED BY
Wenting Shen,
Qingdao University Qingdao, China
Junjun Guo,
KMUST, China

*CORRESPONDENCE
Feng Lian,
lianfeng1981@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 09 July 2022
ACCEPTED 25 July 2022
PUBLISHED 05 September 2022

CITATION
Zhang G, Zeng L, Lian F, Liu X, Fu N and Dai S (2022), State estimation for dynamic systems with higher-order autoregressive moving average non-Gaussian noise. *Front. Energy Res.* 10:990267.
doi: 10.3389/fenrg.2022.990267



# State estimation for dynamic systems with higher-order autoregressive moving average non-Gaussian noise

Guanghua Zhang[1], Linghao Zeng[2], Feng Lian[1]*, Xinqiang Liu[3], Na Fu[4] and Shasha Dai[5]

[1]School of Automation Science and Engineering, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, [2]School of Economics and Management, Chang'an University, Xi'an, China, [3]Beijing Institute of Electronic System Engineering, Beijing, China, [4]State Key Laboratory of Astronautic Dynamics, Xi'an Satellite Control Center, Xi'an, China, [5]Xi'an Satellite Control Center, Xi'an, China

The classical Kalman filter is a very important state estimation approach, which has been widely used in many engineering applications. The Kalman filter is optimal for linear dynamic systems with independent Gaussian noises. However, the independence and Gaussian assumptions may not be satisfied in practice. On the one hand, modeling physical systems usually results in discrete-time state-space models with correlated process and measurement noises. On the other hand, the noise is non-Gaussian when the system is disturbed by heavy-tailed noise. In this case, the performance of the Kalman filter will deteriorate, or even diverge. This paper is devoted to addressing the state estimation problem of linear dynamic systems with high-order autoregressive moving average (ARMA) non-Gaussian noise. First, a triplet Markov model is introduced to model the system with high-order ARMA noise, since this model relaxes the independence assumption of the hidden Markov model. Then, a new filter is derived based on correntropy, instead of the commonly used minimum mean square error (MMSE), to deal with non-Gaussian noise. Unlike the MMSE, which uses only second-order statistics of error, correntropy can capture second-order and higher-order statistics. Finally, simulation results verify the effectiveness of the proposed algorithm.

KEYWORDS

**kalman filter, higher-order autoregressive moving average, non-Gaussian, triplet markov model, correntropy**

## 1 Introduction

State estimation is a very important problem in many engineering applications, such as energy internet, system control, tracking, and so on [Zandavi and Chung (2019); Zhang et al. (2022)]. These engineering applications are essentially a dynamic system, which is usually described as a state-space model. The hidden Markov model (HMM) is

the one of the most commonly used state-space models (Zhang et al. (2018)). For linear case, the state estimation problem is generally solved by the well-known Kalman filter(KF) (Kalman (1960)), which is an optimal filter in the minimum mean square error (MMSE) sense. MMSE is one of the most commonly used cost function in the case of Gaussian noise, and MMSE approach is an estimator which minimizes the mean square error. In addition, a large number of nonlinear filters have been proposed to solve nonlinear estimation problems, such as extended Kalman filter, unscented Kalman filter, cubature Kalman filter, particle filter, to name but a few (Anderson and Moore (2012)).

Although the KF in general performs well, it has rigorous requirements, that is, the process and measurement noises of dynamic systems are independent and Gaussian. However, the independence and Gaussian assumptions do not always hold in practice (Zhang G. et al. (2021)). On one hand, in fact, the noise of most dynamic systems is correlated. Research has shown that modeling physical systems usually results in discrete-time state-space models with correlated process and measurement noises, and some practical applications are explained in (Saha and Gustafsson (2012)). In addition, the dynamic and measurement noise may even high-order (i.e., multi-step) correlated in some severe environments (Zhang D. et al. (2021)). On the other hand, the main reason for using the Gaussian assumption is that it is mathematically simple, but in fact, dynamic systems are usually disturbed by some heavy-tailed impulse noise (Roth et al. (2013)). When the independence and Gaussian assumptions are not satisfied, the KF may fail to output reliable estimation results.

To deal with correlated noise, the traditional method is to reconstruct an HMM by prewhitening processing, and then the classical KF can be used to estimate the state (Bar-Shalom et al. (2001)). Another solution is to characterize dynamic systems with correlated noise through more flexible state-space models, such as the pairwise Markov model (PMM) (Pieczynski and Desbouvries (2003)) and the triplet Markov model (TMM) (Ait-El-Fquih and Desbouvries (2006)). In the PMM, the state and measurement as a whole are regarded as a Markov process, which improves the modeling ability of complex dynamic systems. In the TMM, an auxiliary variable is introduced to completely describe the dynamic systems. This auxiliary variable can play a very significant role in some engineering applications. For example, it can characterize the uncertainty of parameters, non-stationarity and error sources. It has been proved that the TMM is more general than the HMM and the PMM, and it structural advantages make it more preferable in addressing some real-world applications, such as image segmentation (Derrode and Pieczynski (2004)), speech processing (Ait El Fquih and Desbouvries (2005)), target tracking (Zhang et al. (2017); Lehmann and Pieczynski (2020, 2021)), and so on.

For non-Gaussian noise, several approaches have been proposed, which are mainly divided into three categories (Izanloo et al. (2016)). The first is to replace the Gaussian distribution with a more extensive heavy-tailed distribution (Huang et al. (2017)). For example, the Student's t distribution is one of the most commonly used heavy-tailed distribution. The main disadvantage of heavy-tailed distributions is that they are usually analytically difficult, which brings about that related estimation approaches have no closed form solution. The second is the multiple model technique. In this approach, non-Gaussian noise is represented as a finite sum of Gaussian distribution (Shan et al. (2021)). The main difficulty of this approach is how to design the model set reasonably, and the disadvantage is that the amount of calculation will increase sharply with the increase of the number of models. The third is the Monte Carlo approach, in which a set of weighted random particles are employed to characterize the state (Liu et al. (2018)). Generally, sampling-based algorithms can be categorized into deterministic sampling method and random sampling method. Particularly, in the random sampling method, enough particles can approximate the real state with arbitrary precision, at the cost of expensive computation.

In the past few years, the correntropy-based filtering technology has become an important orientation to solve the state estimation of dynamic systems with non-Gaussian noise (Kulikova (2017); Chen et al. (2017)). In information theory, correntropy is a significant mathematical tool to measure the similarity of two random variables. Unlike the commonly utilized MMSE cost function, which uses only second-order statistic of error, the correntropy captures second-order and higher-order information, and is more suitable for non-Gaussian noise, such as heavy-tailed impulsive noise. Several filtering algorithms based on correntropy have been designed in the framework of HMMs, and they are more robust to non-Gaussian noise than the KF and its variants.

In this paper, we are devoted to addressing the state estimation problem of linear dynamic systems with high-order autoregressive moving average (ARMA) non-Gaussian noise. A new Kalman-like filter is developed in the framework of the TMM based on correntropy. First, we resort to a linear TMM to describe dynamic systems with high-order ARMA noise, since the TMM is more general than the HMM. Second, based on the model, a new Kalman-like filter is derived by using correntropy cost function, instead of the commonly used MMSE cost function. Because correntropy can capture not only second-order but also higher-order statistics of error, the proposed algorithm is more robust to non-Gaussian noise than the traditional filter. Finally, simulation results show the effectiveness of the proposed algorithm.

The rest of the paper is organized as follows. **Section 2** is the modeling of linear dynamic systems with high-order ARMA noise. **Section 3** derives a new Kalman filter by using correntropy

cost function in the framework of the TMM. In **Section 4**, we validate the proposed algorithm *via* simulations. Finally, conclusion is provided in **Section 5**.

# 2 Modeling of linear dynamic systems with high-order ARMA noise

## 2.1 Linear hidden markov model

Consider the following linear dynamic system

$$\begin{cases} x_{k+1} = F_k x_k + G_k w_k \\ z_k = H_k x_k + v_k \end{cases} \tag{1}$$

where $k$ is the time index, $x_k \in \mathbb{R}^{n_x}$ is the state vector of dimension $n_x$, $F_k$ the transition matrix, $G_k$ is the process noise matrix, $w_k \sim \mathcal{N}(0, Q_k)$ is the process noise, $z_k \in \mathbb{R}^{n_z}$ is the measurement vector of dimension $n_z$, $H_k$ is the measurement matrix, and $v_k \sim \mathcal{N}(0, R_k)$ is the measurement noise. $\mathcal{N}(m, P)$ denotes a Gaussian distribution with mean vector $m$ and covariance matrix $P$.

In general, noise sequences $w = \{w_k\}_{k\in\mathrm{IN}}$ and $v = \{v_k\}_{k\in\mathrm{IN}}$ are assumed independent, jointly independent and independent of the initial state $x_0 \sim \mathcal{N}(\hat{x}_0, P_0)$. Then the state estimate can be obtained by the classical KF, which is an optimal filter in the MMSE sense. However, the independence and Gaussian assumptions that are typically assumed in the HMM do not always hold in practice, such as dynamic systems with high-order ARMA non-Gaussian noise. In this case, the KF may not output reliable estimation results.

## 2.2 Linear triplet markov model for dynamic systems with high-order ARMA noise

### 2.2.1 Linear triplet markov chain model

We resort to a linear TMM to describe a linear HMM with correlated noise. Let $x_k \in \mathbb{R}^{n_x}$ is the state vector, $z_k \in \mathbb{R}^{n_z}$ is the measurement vector, $r_k \in \mathbb{R}^{n_r}$ is an auxiliary variable, and $\zeta_k = \left[x_k^{\mathrm{T}}, r_k^{\mathrm{T}}, z_{k-1}^{\mathrm{T}}\right]^{\mathrm{T}}$. If $\zeta = \{\zeta_k\}_{k\in\mathrm{IN}}$ is a Markov process, the following system is called a linear TMM (Ait-El-Fquih and Desbouvries (2006)):

$$\underbrace{\begin{bmatrix} x_{k+1} \\ r_{k+1} \\ z_k \end{bmatrix}}_{\zeta_{k+1}} = \underbrace{\begin{bmatrix} \mathcal{F}_k^{xx} & \mathcal{F}_k^{xr} & \mathcal{F}_k^{xz} \\ \mathcal{F}_k^{rx} & \mathcal{F}_k^{rr} & \mathcal{F}_k^{rz} \\ \mathcal{F}_k^{zx} & \mathcal{F}_k^{zr} & \mathcal{F}_k^{zz} \end{bmatrix}}_{\mathcal{F}_k} \begin{bmatrix} x_k \\ r_k \\ z_{k-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \xi_k^x \\ \xi_k^r \\ \xi_k^z \end{bmatrix}}_{\xi_k} \tag{2}$$

where $\xi = \{\xi_k\}_{k\in\mathrm{IN}}$ is zero mean white noise and independent of the initial state $\zeta_0$.

### 2.2.2 Modeling high-order ARMA noise using TMM

In this section, we utilize a linear TMM to model dynamic systems with high-order ARMA noise (Zhang D. et al. (2021)). The TMM provides a general framework for these typical stochastic systems.

(1) High-Order ARMA Process Noise.

For high-order ARMA process noise, it can usually be written in the form of the following difference equation:

$$w_k = -\sum_{i=1}^{p^w} \alpha_i^w w_{k-i} + \sum_{i=0}^{q^w} \beta_i^w \xi_{k-i}^w \tag{3}$$

where $\xi_k^w$ is white noise. Model (3) is a typical high-order ARMA model, in which $p^w$ is the autoregressive order, $q^w$ is the moving average order, and coefficient parameters $\alpha_i^w$ and $\beta_i^w$ are determined by the spectral factor $H_k^w(z)$ of the power spectral density $\Phi_k^w(z)$ of the process noise $w_k$.

Suppose $\Phi_k^w(z)$ is a rational spectrum. According to the spectral decomposition theorem, there is a spectral factor satisfying

$$\Phi_k^w(z) = H_k^w(z) H_k^w(z)^* \tag{4}$$

where $(\cdot)^*$ represents complex conjugate transpose operation, and $H_k^w(z)$ can be written by

$$H_k^w(z) = C_k^w \left(zI - A_k^w\right)^{-1} B_k^w + D_k^w \tag{5}$$

Then, high-order ARMA process noise can be formulated by

$$\begin{cases} x_{k+1}^w = A_k^w x_k^w + B_k^w \xi_k^w \\ w_k = C_k^w x_k^w + D_k^w \xi_k^w \end{cases} \tag{6}$$

If the process noise in model 1) is high-order ARMA noise, it can be described by model (6). In this case, $\{x_k\}$ is no longer a Markov process, but $\left\{\left(x_k, x_k^w\right)\right\}$ is a Markov process. Let $\zeta_k = \left(x_k, r_k = x_k^w, z_{k-1}\right)$. Model 1) with (6) can be written in the form of linear TMM (2), i.e.,

$$\begin{bmatrix} x_{k+1} \\ x_{k+1}^w \\ z_k \end{bmatrix} = \begin{bmatrix} F_k & G_k C_k^w & 0 \\ 0 & A_k^w & 0 \\ H_k & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_k^w \\ z_{k-1} \end{bmatrix} + \begin{bmatrix} G_k D_k^w \xi_k^w \\ B_k^w \xi_k^w \\ v_k \end{bmatrix} \tag{7}$$

Assuming white noise $\xi_k^w \sim \mathcal{N}(0, Q_k)$, the noise covariance matrix of model 7) is

$$\mathcal{Q}_k = \begin{bmatrix} G_k D_k^w Q_k \left(D_k^w\right)^{\mathrm{T}} G_k^{\mathrm{T}} & G_k D_k^w Q_k \left(B_k^w\right)^{\mathrm{T}} & 0 \\ B_k^w Q_k \left(D_k^w\right)^{\mathrm{T}} G_k^{\mathrm{T}} & B_k^w Q_k \left(B_k^w\right)^{\mathrm{T}} & 0 \\ 0 & 0 & R_k \end{bmatrix} \tag{8}$$

(2) High-order ARMA Measurement Noise.

For high-order ARMA measurement noise, it can also be written in the following form of difference equation

$$v_k = -\sum_{i=1}^{p^v} a_i^v v_{k-i} + \sum_{i=0}^{q^v} b_i^v \xi_{k-i}^v \tag{9}$$

where $\xi_k^v$ is white noise, $p^v$ and $q^v$ are autoregressive order and moving average order, respectively. If the spectral density of high-order ARMA measurement noise is $\Phi_k^v(z)$ and the corresponding spectral factor is $H_k^v(z)$, similar to model (6), $v_k$ can be modeled as follows

$$\begin{cases} x_{k+1}^v = A_k^v x_k^v + B_k^v \xi_k^v \\ v_k = C_k^v x_k^v + D_k^v \xi_k^v \end{cases} \tag{10}$$

If the measurement noise in model 1) is high-order ARMA noise, it can be described by model (10). Let $\zeta_k = \left(x_k, r_k = x_k^v, z_{k-1}\right)$. Model 1) with (10) can be written in the form of linear TMM (2), i.e.,

$$\begin{bmatrix} x_{k+1} \\ x_{k+1}^v \\ z_k \end{bmatrix} = \begin{bmatrix} F_k & 0 & 0 \\ 0 & A_k^v & 0 \\ H_k & C_k^v & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_k^v \\ z_{k-1} \end{bmatrix} + \begin{bmatrix} G_k w_k \\ B_k^v \xi_k^v \\ D_k^v \xi_k^v \end{bmatrix} \tag{11}$$

Assuming white noise $\xi_k^v \sim \mathcal{N}(0, R_k)$, the noise covariance matrix of model 11) is

$$\mathcal{Q}_k = \begin{bmatrix} G_k Q_k G_k^{\mathrm{T}} & 0 & 0 \\ 0 & B_k^v R_k \left(B_k^v\right)^{\mathrm{T}} & B_k^v R_k \left(D_k^v\right)^{\mathrm{T}} \\ 0 & D_k^v R_k \left(B_k^v\right)^{\mathrm{T}} & D_k^v R_k \left(D_k^v\right)^{\mathrm{T}} \end{bmatrix} \tag{12}$$

(3) High-Order ARMA Process and Measurement Noises.

If the process noise and measurement noise are high-order ARMA noises, they can be described by model 6) and model (10), respectively. Let $\zeta_k = \left(x_k, r_k = \left(x_k^w, x_k^v\right), z_{k-1}\right)$. Model 1) with (6) and (10) can be written in the form of linear TMM (2), i.e.,

$$\begin{bmatrix} x_{k+1} \\ x_{k+1}^w \\ x_{k+1}^v \\ z_k \end{bmatrix} = \begin{bmatrix} F_k & G_k C_k^w & 0 & 0 \\ 0 & A_k^w & 0 & 0 \\ 0 & 0 & A_k^v & 0 \\ H_k & 0 & C_k^v & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_k^w \\ x_k^v \\ z_{k-1} \end{bmatrix} + \begin{bmatrix} G_k D_k^w \xi_k^w \\ B_k^w \xi_k^w \\ B_k^v \xi_k^v \\ D_k^v \xi_k^v \end{bmatrix} \tag{13}$$

Assuming white noises $\xi_k^w \sim \mathcal{N}(0, Q_k)$ and $\xi_k^v \sim \mathcal{N}(0, R_k)$, the noise covariance matrix of model 13) is

$$\mathcal{Q}_k = \begin{bmatrix} G_k D_k^w Q_k \left(D_k^w\right)^{\mathrm{T}} G_k^{\mathrm{T}} & G_k D_k^w Q_k \left(B_k^w\right)^{\mathrm{T}} & 0 & 0 \\ B_k^w Q_k \left(D_k^w\right)^{\mathrm{T}} G_k^{\mathrm{T}} & B_k^w Q_k \left(B_k^w\right)^{\mathrm{T}} & 0 & 0 \\ 0 & 0 & B_k^v R_k \left(B_k^v\right)^{\mathrm{T}} & B_k^v R_k \left(D_k^v\right)^{\mathrm{T}} \\ 0 & 0 & D_k^v R_k \left(B_k^v\right)^{\mathrm{T}} & D_k^v R_k \left(D_k^v\right)^{\mathrm{T}} \end{bmatrix} \tag{14}$$

## 2.3 Restoration algorithm

Let $x_k^* = (x_k, r_k)$. Then model 2) can be written as

$$\begin{bmatrix} x_{k+1}^* \\ z_k \end{bmatrix} = \begin{bmatrix} \mathcal{F}_k^{x^* x^*} & \mathcal{F}_k^{x^* z} \\ \mathcal{F}_k^{z x^*} & \mathcal{F}_k^{zz} \end{bmatrix} \begin{bmatrix} x_k^* \\ z_{k-1} \end{bmatrix} + \begin{bmatrix} \xi_k^{x^*} \\ \xi_k^z \end{bmatrix} \tag{15}$$

where the initial state $x_0^*$ and noise $\xi_k$ are

$$x_0^* \sim \mathcal{N}\left(\hat{x}_0^*, P_0^*\right), \ \xi_k \sim \mathcal{N}\left(0, \underbrace{\begin{bmatrix} \mathcal{Q}_k^{x^* x^*} & \mathcal{Q}_k^{x^* z} \\ \mathcal{Q}_k^{z x^*} & \mathcal{Q}_k^{zz} \end{bmatrix}}_{\mathcal{Q}_k}\right) \tag{16}$$

For model 15) with (16), a Kalman-like filter, called triplet Kalman filter (TKF), has been derived to estimate the state $x_k^*$. For convenience, the recursive equations are summarized as follows (Ait-El-Fquih and Desbouvries (2006)).

Initialization:

$$\hat{x}_{0|0}^* = \hat{x}_0^*, \ P_{0|0} = P_0^* \tag{17}$$

$$\hat{\mathcal{F}}_k^{x^* x^*} = \mathcal{F}_k^{x^* x^*} - \mathcal{Q}_k^{x^* z}\left(\mathcal{Q}_k^{zz}\right)^{-1} \mathcal{F}_k^{z x^*} \tag{18}$$

$$\hat{\mathcal{F}}_k^{x^* z} = \mathcal{F}_k^{x^* z} - \mathcal{Q}_k^{x^* z}\left(\mathcal{Q}_k^{zz}\right)^{-1} \mathcal{F}_k^{zz} \tag{19}$$

$$\hat{\mathcal{Q}}_k^{x^* x^*} = \mathcal{Q}_k^{x^* x^*} - \mathcal{Q}_k^{x^* z}\left(\mathcal{Q}_k^{zz}\right)^{-1} \mathcal{Q}_k^{z x^*} \tag{20}$$

Prediction:

$$\hat{x}_{k|k-1}^* = \hat{\mathcal{F}}_{k-1}^{x^* x^*} \hat{x}_{k-1|k-1}^* + \mathcal{Q}_{k-1}^{x^* z}\left(\mathcal{Q}_{k-1}^{zz}\right)^{-1} z_{k-1} + \hat{\mathcal{F}}_{k-1}^{x^* z} z_{k-2} \tag{21}$$

$$P_{k|k-1}^* = \hat{\mathcal{F}}_{k-1}^{x^* x^*} P_{k-1|k-1}^* \left(\hat{\mathcal{F}}_{k-1}^{x^* x^*}\right)^{\mathrm{T}} + \hat{\mathcal{Q}}_{k-1}^{x^* x^*} \tag{22}$$

Update:

$$e_k = z_k - \mathcal{F}_{k-1}^{z x^*} \hat{x}_{k|k-1}^* - \mathcal{F}_{k-1}^{zz} z_{k-1} \tag{23}$$

$$R_{e,k} = \mathcal{F}_{k-1}^{z x^*} P_{k|k-1}^* \left(\mathcal{F}_{k-1}^{z x^*}\right)^{\mathrm{T}} + \mathcal{Q}_{k-1}^{zz} \tag{24}$$

$$K_k = P_{k|k-1}^* \left(\mathcal{F}_{k-1}^{z x^*}\right)^{\mathrm{T}} R_{e,k}^{-1} \tag{25}$$

$$\hat{x}_{k|k}^* = \hat{x}_{k|k-1}^* + K_k e_k \tag{26}$$

$$P_{k|k}^* = \left(I - K_k \mathcal{F}_k^{z x^*}\right) P_{k|k-1}^* \tag{27}$$

The TKF is also an optimal filter in the MMSE sense. It in general performs well in Gaussian noise. However, it performance will deteriorate or even diverge when applied to non-Gaussian systems, since the TKF is derived under MMSE criterion, which only uses second-order statistics of error. To solve this problem, in the next section, a new filter is developed by using correntropy cost function, which utilizes not only second-order but also higher-order statistics information.

# 3 Correntropy-based triplet kalman filter

## 3.1 Correntropy

Correntropy is a very useful metric tool to measure the similarity of two random variables in information theory

(Chen et al. (2017)). For variables $X$ and $Y$, the correntropy is defined by

$$C(X,Y) = \mathrm{E}\left[\kappa(X,Y)\right] = \int \kappa(x,y)\, df_{XY}(x,y) \tag{28}$$

where $\mathrm{E}[\cdot]$ is an expectation operator, $\kappa(\cdot,\cdot)$ is a kernel function, and $f_{XY}(x,y)$ is the joint probability density function of $X$ and $Y$. Generally, $f_{XY}(x,y)$ is unknown, and only a limited amount of data is provided. Thus, the correntropy can be computed by

$$\hat{C}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}\kappa(x,y). \tag{29}$$

There are many options for kernel function. In this paper, we choose the Gaussian kernel function

$$\kappa(x,y) = \mathrm{G}_\sigma(x_i - y_i), \tag{30}$$

where $\sigma$ is the kernel size, and $\mathrm{G}_\sigma(x_i - y_i) = \exp\left(-\frac{\|x_i - y_i\|^2}{2\sigma^2}\right)$. The Gaussian kernel function is positive definite and bounded. When $X = Y$, it takes the maximum value.

For the Gaussian kernel function, its Taylor series expansion can be written as

$$C(X,Y) = \sum_{n=0}^{\infty}\frac{(-1)^n}{2^n\sigma^{2n}n!}\mathrm{E}\left[(X-Y)^{2n}\right]. \tag{31}$$

It can be seen that the correntropy is in essence the weighted sum of all even-order moments of error. Compared with the MMSE, which uses only the second-order statistics of error, correntropy captures the second-order and higher-order statistics.

## 3.2 Main results

In this section, a new filter, called correntropy-based TKF (CTKF), is derived by using correntropy under TMM. For clarity, we first provide the main results, and then give the mathematical derivation.

The initialization and prediction steps of the CTKF are the same as those of the TKF, and its update step is summarized as follows:

$$e_k = z_k - \mathcal{F}_{k-1}^{zx^*}\hat{x}_{k|k-1}^* - \mathcal{F}_{k-1}^{zz}z_{k-1} \tag{32}$$

$$\lambda_k = \mathrm{G}_\sigma\left(\|e_k\|_{(Q_k^{zz})^{-1}}\right) \tag{33}$$

$$R_{e,k} = \mathcal{F}_{k-1}^{zx^*}P_{k|k-1}^*\left(\mathcal{F}_{k-1}^{zx^*}\right)^{\mathrm{T}} + \mathcal{Q}_{k-1}^{zz} \tag{34}$$

$$K_k^\lambda = \lambda_k P_{k|k-1}^*\left(\mathcal{F}_{k-1}^{zx^*}\right)^{\mathrm{T}}R_{e,k}^{-1} \tag{35}$$

$$\hat{x}_{k|k}^* = \hat{x}_{k|k-1}^* + K_k^\lambda e_k \tag{36}$$

$$P_{k|k}^* = \left(I - K_k^\lambda\mathcal{F}_k^{zx^*}\right)P_{k|k-1}^* \tag{37}$$

Proof. For the linear TMM (2), we have

$$\begin{bmatrix}\hat{x}_{k|k-1}^*\\ z_k\end{bmatrix} = \begin{bmatrix}I\\ \mathcal{F}_k^{zx^*}\end{bmatrix}x_k^* + \begin{bmatrix}0\\ \mathcal{F}_k^{zz}\end{bmatrix}z_{k-1} + \eta_k \tag{38}$$

where I and 0 are identity and zeros matrices, and

$$\eta_k = \begin{bmatrix}-\left(x_k^* - \hat{x}_{k|k-1}^*\right)\\ w_k^z\end{bmatrix} \text{ with } \mathrm{E}\left[\eta_k\eta_k^{\mathrm{T}}\right] = \begin{bmatrix}P_{k|k-1}^* & 0\\ 0 & \mathcal{Q}_k^{zz}\end{bmatrix} \tag{39}$$

To address non-Gaussian noise, we use correntropy instead of MMSE to derive update equations. The cost function based on correntropy is established by

$$\begin{aligned}J\left(x_k^*\right) = {} & \mathrm{G}_\sigma\left(\left\|z_k - \mathcal{F}_k^{zx^*}x_k^* - \mathcal{F}_k^{zz}z_{k-1}\right\|_{(Q_k^{zz})^{-1}}\right)\\ & + \mathrm{G}_\sigma\left(\left\|x_k^* - \hat{x}_{k|k-1}^*\right\|_{(P_{k|k-1}^*)^{-1}}\right)\end{aligned} \tag{40}$$

Then the optimal estimation of $x_k^*$ is $\hat{x}_k^* = \arg\max_{x_k^*} J\left(x_k^*\right)$, which can be obtained by

$$\begin{aligned}\frac{\partial J\left(x_k^*\right)}{\partial x_k^*} = {} & \frac{1}{\sigma^2}\mathrm{G}_\sigma\left(\left\|z_k - \mathcal{F}_k^{zx^*}x_k^* - \mathcal{F}_k^{zz}z_{k-1}\right\|_{(Q_k^{zz})^{-1}}\right)\\ & \left(\mathcal{F}_k^{zx^*}\right)^{\mathrm{T}}\left(Q_k^{zz}\right)^{-1}\left(z_k - \mathcal{F}_k^{zx^*}x_k^* - \mathcal{F}_k^{zz}z_{k-1}\right)\\ & - \frac{1}{\sigma^2}\mathrm{G}_\sigma\left(\left\|x_k - \hat{x}_{k|k-1}^*\right\|_{(P_{k|k-1}^*)^{-1}}\right)\left(P_{k|k-1}^*\right)^{-1}\\ & \left(x_k - \hat{x}_{k|k-1}^*\right)\\ = {} & 0.\end{aligned} \tag{41}$$

**Equation 41** can be written by

$$\begin{aligned}\Psi_k x_k^* = {} & \left(P_{k|k-1}^*\right)^{-1}\hat{x}_{k|k-1}^* + \lambda_k\left(\mathcal{F}_k^{zx^*}\right)^{\mathrm{T}}\left(Q_k^{zz}\right)^{-1}\\ & \left(z_k - \mathcal{F}_k^{zz}z_{k-1}\right)\end{aligned} \tag{42}$$

where

$$\Psi_k = \left(P_{k|k-1}^*\right)^{-1} + \lambda_k\left(\mathcal{F}_k^{zx^*}\right)^{\mathrm{T}}\left(Q_k^{zz}\right)^{-1}\mathcal{F}_k^{zx^*}, \tag{43}$$

$$\lambda_k = \frac{\mathrm{G}_\sigma\left(\left\|z_k - \mathcal{F}_k^{zx^*}x_k^* - \mathcal{F}_k^{zz}z_{k-1}\right\|_{(Q_k^{zz})^{-1}}\right)}{\mathrm{G}_\sigma\left(\left\|x_k^* - \hat{x}_{k|k-1}^*\right\|_{(P_{k|k-1}^*)^{-1}}\right)}. \tag{44}$$

Adding and subtracting a term $\lambda_k\left(\mathcal{F}_k^{zx^*}\right)^{\mathrm{T}}\left(Q_k^{zz}\right)^{-1}\mathcal{F}_k^{zx^*}\hat{x}_{k|k-1}^*$ on the right-hand side of **(42)**, we have

$$\begin{aligned}\Psi_k x_k^* = {} & \Psi_k\hat{x}_{k|k-1}^* + \lambda_k\left(\mathcal{F}_k^{zx^*}\right)^{\mathrm{T}}\left(Q_k^{zz}\right)^{-1}\\ & \left(z_k - \mathcal{F}_k^{zx^*}\hat{x}_{k|k-1}^* - \mathcal{F}_k^{zz}z_{k-1}\right).\end{aligned} \tag{45}$$

Thus, the estimation of $x_k^*$ can be computed by

$$\hat{x}_{k|k}^* = \hat{x}_{k|k-1}^* + K_k^\lambda\left(z_k - \mathcal{F}_k^{zx^*}\hat{x}_{k|k-1}^* - \mathcal{F}_k^{zz}z_{k-1}\right) \tag{46}$$

where

$$K_k^\lambda = \Psi_k^{-1}\lambda_k\left(\mathcal{F}_k^{zx^*}\right)^{\mathrm{T}}\left(Q_k^{zz}\right)^{-1} \tag{47}$$

Note that the parameter $\lambda_k$ is function of $x_k^*$. For simplicity, let $x_k^* \approx \hat{x}_{k|k-1}^*$ in (44), and $\lambda_k$ can be obtained by

$$\lambda_k = \mathrm{G}_\sigma\left(\left\|z_k - \mathcal{F}_k^{zx^*}\hat{x}_{k|k-1}^* - \mathcal{F}_k^{zz}z_{k-1}\right\|_{\left(Q_k^{zz}\right)^{-1}}\right) \tag{48}$$

In addition, parameter σ plays an important role in correntropy-based filters. Inspired by (Kulikova (2017)), this paper adopts an adaptive method to choose σ, i.e.,

$$\sigma = \left\|z_k - \mathcal{F}_k^{zx^*}\hat{x}_{k|k-1}^* - \mathcal{F}_k^{zz}z_{k-1}\right\|_{\left(Q_k^{zz}\right)^{-1}} \tag{49}$$

In this section, a CTKF is developed to address the estimation problem of dynamic systems with high-order ARMA non-Gaussian noise. Instead of the commonly used MMSE criterion, which uses only second-order statistics of error, correntropy is employed to derive the filter, since it can captures second-order and higher-order statistics of error. It can be seen that the structure of CTKF is similar to that of TKF, except that an extra scale parameter $\lambda_k$ is involved. The scale parameter is computed according to correntropy criterion to control the gain matrix $K_k^\lambda$, which results in that the CTKF in general performs well for non-Gaussian noise. In addition, the CTKF has a simple form, which facilitates its practical application.

## 4 Numerical simulations

In this section, two scenarios, i.e., dynamic system with high-order ARMA Gaussian and non-Gaussian noise, are taken into account to verify the effectiveness of the TMM and CTKF. In model (1), the state is $x_k = \left[p_{x,k}, v_{x,k}, p_{y,k}, v_{y,k}\right]^{\mathrm{T}}$, and relevant

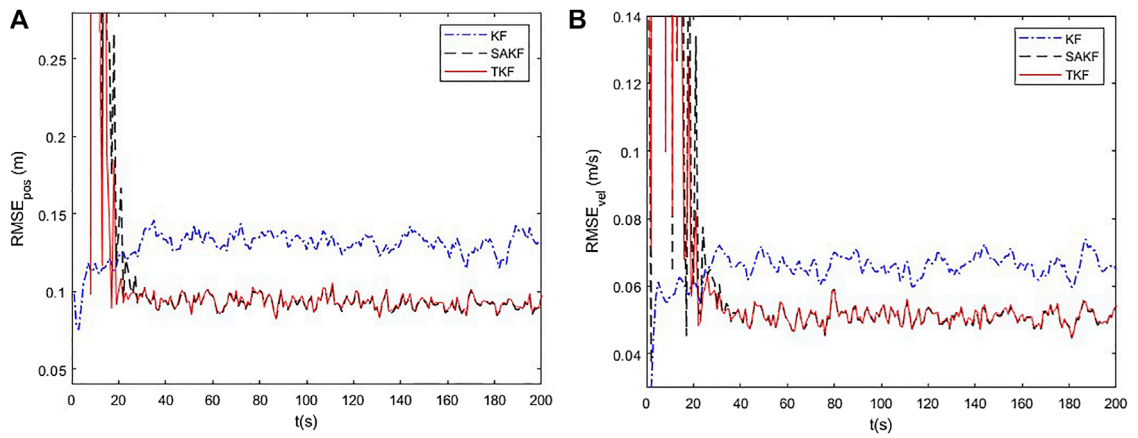



FIGURE 1
RMSE results for different filters. **(A)** Position RMSE. **(B)** Velocity RMSE.

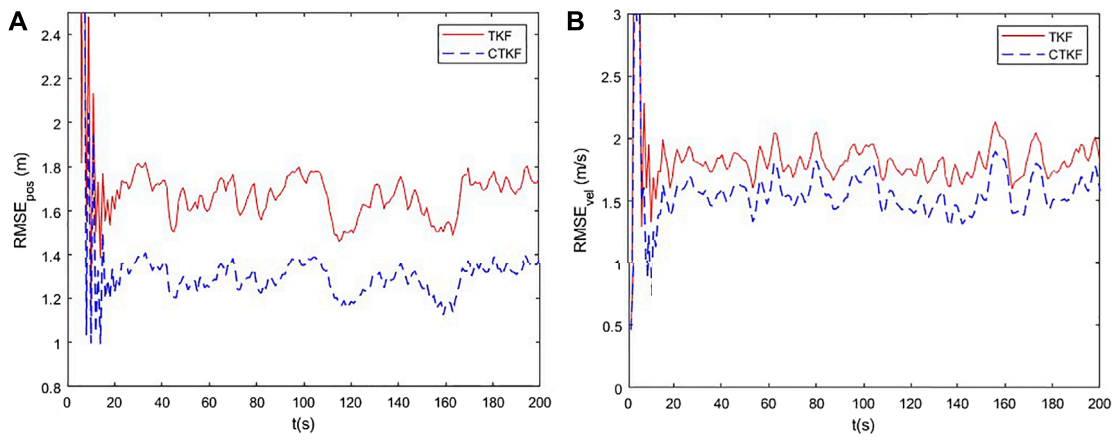



FIGURE 2
RMSE results for different filters. **(A)** Position RMSE. **(B)** Velocity RMSE.

matrices are

$$F_k = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}, G_k = \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 0 \\ 0 & \frac{T^2}{2} \\ 0 & T \end{bmatrix},$$
$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{50}$$

where $T = 1$ is the sampling period. The spectral factors of process and measurement noises are

$$H^w(z) = \frac{z^4 - 0.4z^3 + 0.9z^2 - 0.1z - 0.3}{z^4 - z^3 + 0.5z^2 + 0.2z - 0.4} \tag{51}$$

$$H^v(z) = \frac{z^6 + 0.6z^5 + 0.4z^4 + 0.3z^3 - 0.08z^2 + 0.05z + 0.01}{z^6 + 0.8z^5 + 0.6z^4 + 0.2z^3 - 0.09z^2 - 0.08z + 0.01} \tag{52}$$

Case 1: $\xi_k^w \sim \mathcal{N}(0, Q_k)$ and $\xi_k^v \sim \mathcal{N}(0, R_k)$ are Gaussian noises, where $Q_k = \mathrm{diag}(0.01^2, 0.01^2)$ and $R_k = \mathrm{diag}(0.1^2, 0.1^2)$. For comparison, the standard Kalman filter (KF), the traditional state augmented Kalman filter (SAKF) (Bar-Shalom et al. (2001)), and the triplet Kalman filter (TKF) are tested. Besides, the root mean square error (RMSE) is used to evaluate estimation performance, which is computed by

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\left(x_k^i - \hat{x}_k^i\right)^2 + \left(y_k^i - \hat{y}_k^i\right)^2\right)} \tag{53}$$

where $x_k^i$ and $y_k^i$ are the true values at time k in the ith Monte Carlo trail, and $\hat{x}_k^i$ and $\hat{y}_k^i$ are the corresponding estimation values. The number of Monte Carlo trails is $M = 200$.

Position and velocity RMSE results are provided in **Figure 1**. It can be seen that the TKF and SAKF have similar estimation performance, and are better than the standard KF. High-order ARMA process and measurement noises do not meet the independence assumption, resulting in poor estimation performance of the KF. The TKF and SAKF are essentially equivalent, and they are optimal in the MMSE sense. The former models dynamic system with high-order ARMA noise through TMM, and the latter deals with high-order ARMA noise through prewhitening technique. Simulation results show that TMM can accurately model dynamic systems with high-order ARMA noise.

Case 2: $\xi_k^w \sim \mathcal{N}(0, Q_k)$ and $\xi_k^v \sim \mathcal{N}(0, R_k)$ are Gaussian noise disturbed by shot noise with probability of 0.2, where $Q_k$ and $R_k$ are the same as those in case 1, and the shot noise is generated by $0.1 \times \mathrm{randi}([5, 10])$. Symbol randi([a, b]) denotes that an integer is returned from the uniform distribution of [a, b].

For comparison, the TKF and the proposed CTKF are tested. Position and velocity RMSE results are provided in **Figure 2**. It can be seen that the CTKF performs better than the TKF. Non-Gaussian noise results in the poor estimation performance of the TKF, since it adopts the MMSE criterion, which uses only second-order statistic of error. The CTKF shows stronger robustness to non-Gaussian noise, because the adopted correntropy cost function can capture second-order and higher-order statistics of error. Simulation results show that the CTKF is an effective state estimation method for dynamic systems with high-order ARMA non-Gaussian noise.

## 5 Conclusion

In this paper, a new filter is designed to solve the state estimation problem of dynamic systems with high-order ARMA non-Gaussian noise. In this filter, high-order ARMA process and measurement noises are modeled in the TMM framework, and then the recursive algorithm is derived by using corretropy cost function. On the one hand, the TMM is more general than the HMM, and it can directly model dynamic systems with high-order ARMA noise. On the other hand, correntropy can capture second-order and higher-order statistics of error, and is more suitable for non-Gaussian noise than the MMSE cost function, which uses only second-order statistics of error. In addition, the CTKF has a simple form, which facilitates its practical application.

## Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## Author contributions

GZ provided the idea of the work and organized the manuscript, LZ performed the experiment, FL provided revisions to this paper, XL designed experimental scenarios, and NF and SD conducted data analysis.

## Funding

This work is supported by National Natural Science Foundation of China under Grants 62103318 and 62173266.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2022.990267/full#supplementary-material

## References


Ait El Fquih, B., and Desbouvries, F. (2005). "Kalman filtering for triplet Markov chains : Applications and extensions," in IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Philadelphia, PA, USA, 685–688. vol. IV. doi:10.1109/ICASSP.2005.1416101

Ait-El-Fquih, B., and Desbouvries, F. (2006). Kalman filtering in triplet Markov chains. *IEEE Trans. Signal Process.* 54, 2957–2963. doi:10.1109/TSP.2006.877651

Anderson, B. D., and Moore, J. B. (2012). *Optimal filtering*. New York: Courier Corporation.

Bar-Shalom, Y., Li, X. R., and Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation: Theory, algorthims and software*. New York: Wiley.

Chen, B., Liu, X., Zhao, H., and Principe, J. C. (2017). Maximum correntropy kalman filter. *Automatica* 76, 70–77. doi:10.1016/j.automatica.2016.10.004

Derrode, S., and Pieczynski, W. (2004). Signal and image segmentation using pairwise Markov chains. *IEEE Trans. Signal Process.* 52, 2477–2489. doi:10.1109/TSP.2004.832015

Huang, Y., Zhang, Y., Li, N., Wu, Z., and Chambers, J. A. (2017). A novel robust Student's t-based kalman filter. *IEEE Trans. Aerosp. Electron. Syst.* 53, 1545–1554. doi:10.1109/TAES.2017.2651684

Izanloo, R., Fakoorian, S. A., Yazdi, H. S., and Simon, D. (2016). "Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise," in 2016 Annual Conference on Information Science and Systems, CISS, Princeton, NJ, USA, 500–505. doi:10.1109/CISS.2016.7460553

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45. doi:10.1115/1.3662552

Kulikova, M. V. (2017). Square-root algorithms for maximum correntropy estimation of linear discrete-time systems in presence of non-Gaussian noise. *Syst. Control Lett.* 108, 8–15. doi:10.1016/j.sysconle.2017.07.016

Lehmann, F., and Pieczynski, W. (2021). Reduced-dimension filtering in triplet markov models. *IEEE Trans. Autom. Contr.* 67, 605–617. doi:10.1109/TAC.2021.3050721

Lehmann, F., and Pieczynski, W. (2020). Suboptimal kalman filtering in triplet markov models using model order reduction. *IEEE Signal Process. Lett.* 27, 1100–1104. doi:10.1109/LSP.2020.3002420

Liu, X., Qu, H., Zhao, J., and Yue, P. (2018). Maximum correntropy square-root cubature kalman filter with application to SINS/GPS integrated systems. *ISA Trans.* 80, 195–202. doi:10.1016/j.isatra.2018.05.001

Pieczynski, W., and Desbouvries, F. (2003). "Kalman filtering using pairwise Gaussian models," in IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2003, Hong Kong, China, 57–60. doi:10.1109/ICASSP.2003.1201617

Roth, M., özkan, E., and Gustafsson, F. (2013). "A Student's t filter for heavy tailed process and measurement noise," in IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 5770–5774. doi:10.1109/ICASSP.2013.6638770

Saha, S., and Gustafsson, F. (2012). Particle filtering with dependent noise processes. *IEEE Trans. Signal Process.* 60, 4497–4508. doi:10.1109/TSP.2012.2202653

Shan, C., Zhou, W., Jiang, Z., and Shan, H. (2021). A new Gaussian approximate filter with colored non-stationary heavy-tailed measurement noise. *Digit. Signal Process.* 122, 103358. doi:10.1016/j.dsp.2021.103358

Zandavi, S. M., and Chung, V. (2019). State estimation of nonlinear dynamic system using novel heuristic filter based on genetic algorithm. *Soft Comput.* 23, 5559–5570. doi:10.1007/s00500-018-3213-y

Zhang, D., Duan, Z., Wang, P., and Zhang, Y. (2021a). "Spacecraft state estimation with multichannel higher-order ARMA colored noises," in 2021 International Conference on Control, Automation and Information Sciences (ICCAIS) (Xi'an, China: IEEE), 602–607. doi:10.1109/ICCAIS52680.2021.9624490

Zhang, G. H., Han, C. Z., Lian, F., and Zeng, L. H. (2017). Cardinality balanced multi-target multi-Bernoulli filter for pairwise markov model. *Acta Autom. Sin.* 43, 2100–2108. doi:10.16383/j.aas.2017.c160430

Zhang, G., Lan, J., Zhang, L., He, F., and Li, S. (2021b). Filtering in pairwise markov model with Student's t non-stationary noise with application to target tracking. *IEEE Trans. Signal Process.* 69, 1627–1641. doi:10.1109/TSP.2021.3062170

Zhang, G., Lian, F., Han, C., Chen, H., and Fu, N. (2018). Two novel sensor control schemes for multi-target tracking via delta generalised labelled multi-Bernoulli filtering. *IET signal Process.* 12, 1131–1139. doi:10.1049/iet-spr.2018.5124

Zhang, X., Liang, H., Feng, J., and Tan, H. (2022). Kalman filter based high precision temperature data processing method. *Front. Energy Res.* 10, 832346. doi:10.3389/fenrg.2022.832346

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fenrg.2022.1005660



OPEN ACCESS

EDITED BY
Hanlin Zhang,
Qingdao University, China

REVIEWED BY
Chao Li,
Tianjin Normal University, China
Xialei Zhang,
Shanxi University, China

*CORRESPONDENCE
Qingyu Yang,
yangqingyu@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 28 July 2022
ACCEPTED 12 August 2022
PUBLISHED 15 September 2022

CITATION
Zhang F and Yang Q (2022), False data injection attack detection in dynamic power grid: A recurrent neural network-based method.
*Front. Energy Res.* 10:1005660.
doi: 10.3389/fenrg.2022.1005660



# False data injection attack detection in dynamic power grid: A recurrent neural network-based method

Feiye Zhang and Qingyu Yang*

School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China

The smart grid greatly facilitates the transmission of power and information by integrating precise measurement technology and efficient decision support systems. However, deep integration of cyber and physical information entails multiple challenges to grid operation. False data injection attacks can directly interfere with the results of state estimation, which can cause the grid regulator to make wrong decisions and thus poses a huge threat to the stability and security of grid operation. To address this issue, we propose a detection approach against false data injection attacks for dynamic state estimation. The Kalman filter is used to dynamically estimate the state values from IEEE standard bus systems. A long short-term memory (LSTM) network is utilized to extract the sequential observations from states at multiple time steps. In addition, we transform the attack detection problem into supervised learning problem and propose a deep neural network-based detection approach to identify attacks. We evaluate the effectiveness of the proposed detection approach in multiple IEEE standard bus systems. The simulation results demonstrate that the proposed detection approach outperforms benchmarks in improving the detection accuracy of malicious attacks.

KEYWORDS

**smart grid, false data injection attack, dynamic state estimation, Kalman filter, deep learning**

## 1 Introduction

Recently, precise measurement technology and decision support systems have been increasingly widely used in the smart grid, including grid monitoring, information sharing, and attack detection, which have significantly improved the safety and efficiency of grid operation (Lee and Lee, 2015). The integration of IoT systems greatly facilitates the transmission efficiency of smart grids through bilateral flow of power and information. A large number of advanced measurement devices have significantly improved the automation and management level of the smart grid. However, the smart grid is more fragile to attacks than a typical grid because of its open network environment (Connolly et al., 2019; Yang et al., 2017).

As a new type of malicious attack against grid monitoring systems, false data injection attacks are extremely threatening because it is difficult for the defender to identify

the attack behaviors (Deng et al., 2015;Cintuglu et al., 2016). The adversary injects a malicious attack based on current state estimation and adjusts the attack vector, which can bypass existing detection methods, resulting a deviation in the results of state estimation results and posing a huge threat to grid operation (Rahman and Mohsenian-Rad, 2013). Unlike traditional power grid attacks, the objects and methods of false data injection attacks have diverse characteristics, thus, it is difficult to detect them with traditional detection methods (e.g., residue-based bad data detection and measurement mutation detection) (Giani et al., 2011; Sandberg et al., 2010).

In recent years, false data injection attacks have received growing attention, and extensive detection approaches have been investigated. Most recent research efforts focus on estimating the state in a static scenario (Guan and Ge, 2017;Rahman and Mohsenian-Rad, 2012). For instance, Guan and Ge (2017) constructed a resilient attack detection approach to detect the presence of false data injection. James et al. (2018) proposed an online deep neural network-based detection approach to oppose false data injection attacks in AC systems. Li et al. (2014) introduced the use of the generalized likelihood ratio to address the attack detection problem with unknown parameters. Rahman and Mohsenian-Rad (2012) proposed a novel measurement to rank smart grid topologies to detect malicious attacks.

Detection approaches based on dynamic state estimation have attracted growing attention in the recent years (Karimipour and Dinavahi, 2017;Kurt et al., 2018b). For instance, Karimipour and Dinavahi (2017) proposed a robust attack detection method based on Euclidean distance metric and Markov decision progress. Taha et al. (2016) presented a dynamic attack detection strategy to mitigate the impact of unknown cyber-attacks. Chakhchoukh et al. (2019) proposed a statistical outlier detection algorithm based on successive batch regression representations of the Kalman filter. Ünal et al. (2021) developed a novel detection approach that employs machine learning, deep learning, and parallel computing techniques. Dayaratne et al. (2022) reported a data-driven unsupervised anomaly detection approach that is based on the k-means clustering method and the Spectral Residual method to detect false data injection attacks in smart grid demand response.

By contrast with existing studies, in this study, we first analyze the basic principles of false data injection attacks from the attackers' perspective and then present a detection approach for false data injection attacks with dynamic state estimation using a recurrent neural network and a Kalman filter. The main contributions of this study are outlined below:

- We first review the dynamic state estimation of a grid system and briefly analyze the basic principles of false data injection attacks on a smart grid from the attacker's perspective.
- We then transform the detection problem of injection attacks into a binary classification problem, and we propose a LSTM-based malicious attack detection approach of smart grid.
- Finally, we demonstrate the effectiveness of the proposed attack detection method in multiple IEEE standard bus systems. The experimental results show that the proposed detection method greatly outperforms benchmarks in terms of accuracy.

# 2 Background

In this section, we present the basic operating principles of the dynamic state estimation of power system. Then, we introduce a conventional bad data detection mechanism. Finally, we briefly show a false data injection attack model from the attacker's perspective.

## 2.1 Dynamic state estimation

State estimation refers to the obtaining of network topology and real-time measurement data through a supervisory control and data acquisition (SCADA) system. The SCADA system estimates the state of grid operations to perform a power system analysis, safety monitoring, etc. Dynamic state estimation then obtains the estimated state value based on measurement data instead of directly calculating the state value at the current moment, such as in static state estimation. As an example of typical dynamic state estimation approach, a Kalman filter uses discrete measurement sequences $\{z_1, z_2, \ldots, z_n\}$ to estimate discrete state sequences $\{x_1, x_2, \ldots, x_n\}$. For a discrete state sequence, there are two ways to estimate the state value $x_{t+1}$ at time $t+1$ from the state value at time $t$: 1) estimate the state value $x_{t+1}$ with measurement data $z_{t+1}$ obtained at $t+1$; 2) predict the state value $x_{t+1}$ through the system state $x_t$ at $t$.

Notice that part of the state quantity $x_{t+1}$ is calculated from the indirect estimation of the measurement data $z_{t+1}$, and the other part of the state quantity $x_{t+1}$ is calculated from the system state $x_t$ through state transition prediction. In a power system that contains $m$ measurement data and $n+1$ nodes, the state prediction equation and measurement equation for a discrete system are expressed as:

$$\boldsymbol{x}_{t+1} = \boldsymbol{F}_t \boldsymbol{x}_t + \boldsymbol{W}_t, \tag{1}$$

$$\boldsymbol{z}_t = \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{V}_t, \tag{2}$$

where $\boldsymbol{x}_t$ denotes the system state value with $n \times 1$ dimensions at time $t$, $\boldsymbol{F}_{t-1}$ is the transition function with $n \times n$ dimensions, $\boldsymbol{W}_t$ is an $n \times 1$ dimensional noise with mean zero. $\boldsymbol{z}_t$ is a measurement vector with $m \times 1$ dimensions, $\boldsymbol{H}_t$ is the measurement matrix of a system with $m \times n$ dimensions. $\boldsymbol{V}_t$ is $m \times 1$ dimensional noise with a mean of zero.

We aim seek the estimation of state $\hat{\boldsymbol{x}}_t$ with the known measurement sequence $\{z_0, z_1, \ldots, z_t\}$ with minimal error $\boldsymbol{e} = \boldsymbol{x}_t - \hat{\boldsymbol{x}}_t$:

$$E\left[\boldsymbol{e}\boldsymbol{e}^T\right] = min. \tag{3}$$

The basic principle of Kalman filtering techniques includes two components: state prediction and state update (Fan and Li, 2009).

State prediction:

$$\begin{aligned} \hat{\boldsymbol{x}}_t^- &= \boldsymbol{F}_t\hat{\boldsymbol{x}}_{t-1} + \boldsymbol{W}_t, \\ \boldsymbol{P}_t^- &= \boldsymbol{F}_t\boldsymbol{P}_{t-1}\boldsymbol{F}_t^T + \boldsymbol{V}_t. \end{aligned} \tag{4}$$

State update:

$$\begin{aligned} \boldsymbol{K}_t &= \boldsymbol{P}_t^-\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{P}_t^-\boldsymbol{H}^T + \boldsymbol{V}\right)^-, \\ \hat{\boldsymbol{x}}_t &= \boldsymbol{x}_t^- + \boldsymbol{K}_t\left(z_t - \boldsymbol{H}\boldsymbol{t}_t^-\right), \\ \boldsymbol{P}_t &= \left(\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H}\right)\boldsymbol{P}_t^-, \end{aligned} \tag{5}$$

where $\hat{\boldsymbol{x}}_t^-$ indicates the estimated state value of time step $t$ conditioned on the optimal estimated state $\hat{\boldsymbol{x}}_{t-1}^-$ at time step $t-1$. $\hat{\boldsymbol{x}}_t$ is the optimal estimated state value at time step $t$. $\boldsymbol{K}_t$ is the Kalman matrix, and $\boldsymbol{P}_t$ is the covariance matrix of error. Notice that from the above procedure, the optimal estimated state $\hat{\boldsymbol{x}}_t$ at time step $t$ can be formulated as the predicated value $\hat{\boldsymbol{x}}_{t-1}^-$ of time $t$ to add to the deviation with the Kalman matrix weight.

Kalman filter technology uses a recursive method to dynamically estimate the state of system. It only needs current measurement data $z_{t+1}$ and the estimation data $x_t$ from the previous period to estimate the optimal state $\hat{x}_{t+1}$. It does not require much storage space, which is suitable for combining artificial intelligence approaches.

## 2.2 Bad data detection mechanism

At time step $k$, the error vector in the dynamic state estimation process of power grid, denoted as $\boldsymbol{e}_t$, can be formulated as:

$$\boldsymbol{e}_t = \boldsymbol{z}_t - \boldsymbol{H}\hat{\boldsymbol{x}}_t, \tag{6}$$

where $\boldsymbol{e}_t$ is the error vector at time step $t$, and $\boldsymbol{z}_t$ denotes the measurement vector. Notice that $\boldsymbol{e}_t$ follows a Gaussian distribution.

The normalized error vector $\lambda_t$ is derived as:

$$\lambda_{t,i} = e_{t,i}/V_{t,i}, \tag{7}$$

where $e_{t,i}$ is the $i-th$ component of error vector, and $V_{t,I}$ is the $i-th$ component of measurements error covariance matrix. As shown in **Table 1**, a traditional bad data detection mechanism judges the system state according to the value $\lambda_{t,i}$:Note that we believe that the system does not encounter false data injection attacks if (8) is satisfied:

TABLE 1 Bad data detection mechanism.

| System State | Range of $\lambda_{t,i}$ |
|---|---|
| Normal state | $\lambda_{t,i} < \tau$ |
| Abnormal state | $\lambda_{t,i} > \tau$ |
| Critical state | $\lambda_{t,i} = \tau$ |

$$|\lambda_{t,i}| \le \tau, \forall i (i \in m), \tag{8}$$

where $\tau$ is the detection threshold.

## 2.3 Attack strategy

The basic principle of a bad data detection mechanism is to identify whether the normalized error vector $\lambda_t$ in (**Eq. 8**) surpasses $\tau$. However, the adversary's objective is to manipulate the attack vector to bypass the detection approach. We thus briefly present the attack model from the attacker's perspective (Ding and Liu, 2017; Hu et al., 2015).

State measurement data $z'_{t,i}$ after being attacked is expressed as:

$$z'_{t,i} = z_{t,i} + a_{t,i}, \tag{9}$$

where $z_{t,i}$ is the measurement data, and $a_{t,I}$ is the malicious attack vector injected by the attacker.

According to the bad data detection mechanism $|\lambda_{t,i}| \le \tau$ presented in **Section 2.2**, the following equation can be derived:

$$\left|\left(z'_{t,i} - H_i\hat{x}_{t,i}\right)/V_{t,i}\right| \le \tau.$$

By bringing (9) into (2.3), we have,

$$|z_{t,i} + a_{t,i} - H_i\hat{x}_{t,i}| \le V_{t,i}\tau. \tag{10}$$

Finally, we can derive the safety range of the attack vector:

$$H_i\hat{x}^- - V_{t,i}\tau - z_{t,i} \le a_{t,i} \le V_{t,i}\tau + H_i\hat{x}_{t,i}^- - z_{t,i}. \tag{11}$$

Obviously, we can see that the data integrity attack can bypass traditional detection if the attack $a_{t,i}$ is in the interval indicated in (11).

# 3 Proposed solution

In this section, we propose a neural network-based approach to detect data integrity attacks against the dynamic state estimation of a smart grid. From **Section 2.2** we know that

traditional bad data detection mechanisms determine whether a system is abnormal by comparing the normalized error vector $\lambda_k$ against a specific threshold. However, the detection accuracy is greatly affected by the value of the threshold. When the threshold is high, the detection accuracy decreases, and when the threshold is low, the amount of false detections increase. Thus, in this section we propose an LSTM network that can draw information from observations in previous $m$ episodes to determine whether the system is currently under a data injection attack to implicitly and automatically analyze changes in the threshold when detecting malicious attacks.

## 3.1 Observations of the system

Before introducing the proposed approach, we present observations of the system regarded as the metrics for determining whether the power grid system is under attack.

According to **Section 2.2**, it is difficult to directly check the presence of data injection attacks by the state vector of the system. In this way, we define the computable observation $o_t$ of the system state at time $t$:

$$o_t = \frac{\|z_t - H\hat{x}_t\|}{\|\omega\|}, \tag{12}$$

where $\|z_t - H\hat{x}_t\|$ denotes the size of the error vector at time $t$, and $\|\omega\|$ is the size of the noise. From **Section 2.2**, we can see that in the situation where the system is in normal operation, the size of the error is small, and the value of $o_t$ is also small. On the other hand, when the system is in an abnormal state, the size of the error and related observation $o_t$ is large. Therefore, it is reasonable to utilize these observations to reflect the presence of data integrity attack. Furthermore, the presence of system noise greatly interferes with the judgment of whether the system is under attack, so we introduce the parameter $\|\omega\|$ to reduce the impact of system noise on detecting a malicious attack.

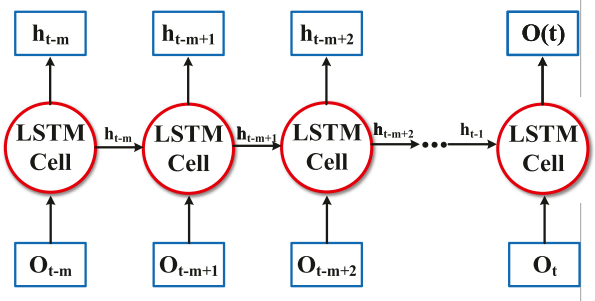



FIGURE 1
LSTM network structure.

## 3.2 LSTM-based feature extraction

From **Section 2.1**, we know that the measurement data for the dynamic state estimation are correlated in time, so it is reasonable to utilize the previous measurement data to judge the presence of data injection attacks. Moreover, because the measurement data of the power grid always contains system noise, only using measurement data at a single time step without considering the sequential information in the system is not a good choice for determining the presence of a false data injection attack. Note that a recurrent neural network (RNN) has shown excellent performance in processing sequential data, and it can extract sequential measurement features to improve the accuracy of false data injection attacks (Sutskever et al., 2014). An LSTM network is a type of RNN that is designed to model temporal sequences, and its prediction of long-term dependencies is more accurate than typical RNN (Gers et al., 2000). Thus, we utilize the LSTM network to extract measurement data over multiple time periods to check the presence of data injection attacks in the power grid system.

**Figure 1** presents the basic structure of the LSTM network utilized in this study. From **Figure 1**, we can see that there are $m$ LSTM cells that are used to store the observations of the power grid system in previous time steps. Specifically, the input of the first LSTM cell is the observation $o_{t-m}$ of the system at time $t-m$. Then, the first LSTM cell utilizes $o_{t-m}$ to calculate the hidden state $h_{t-m}$, which contains information on the previous observation. After that, the second LSTM cell calculates its hidden state $h_{t-m+1}$ by $h_{t-m}$ and its current observation of the system $o_{t-m+1}$. This calculation is then repeated in all LSTM cells, and we utilize the final output of the last LSTM cell as the aggregated observation, denoted as $o(t)$, which contains not only current information on the system observation but the representation of the observation over the past $m$ time steps.

## 3.3 Attack detection algorithm

In the following, the proposed attack-detection method is described. It includes three main procedures: data preprocessing, neural network training, and detection accuracy testing. The overall structure of the attack detection algorithm is illustrated in **Figure 2**, in which, the system first preprocesses the data from IEEE standard bus system and divides the data into two parts: a training set and a testing set, where the training set is used to update the parameters of the neural networks, and the testing set is used to evaluate the accuracy of thee attack detection approach. Then, we train the proposed deep neural network using the training data. Finally, the trained network is used to detect whether the system is under attack. The details of these three parts are presented as follows.

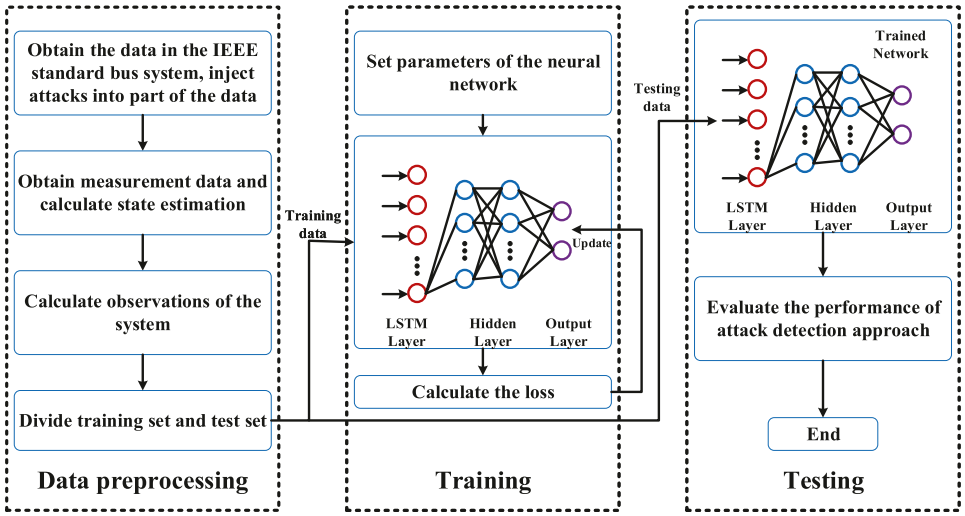



FIGURE 2
Structure of an attack detection algorithm.

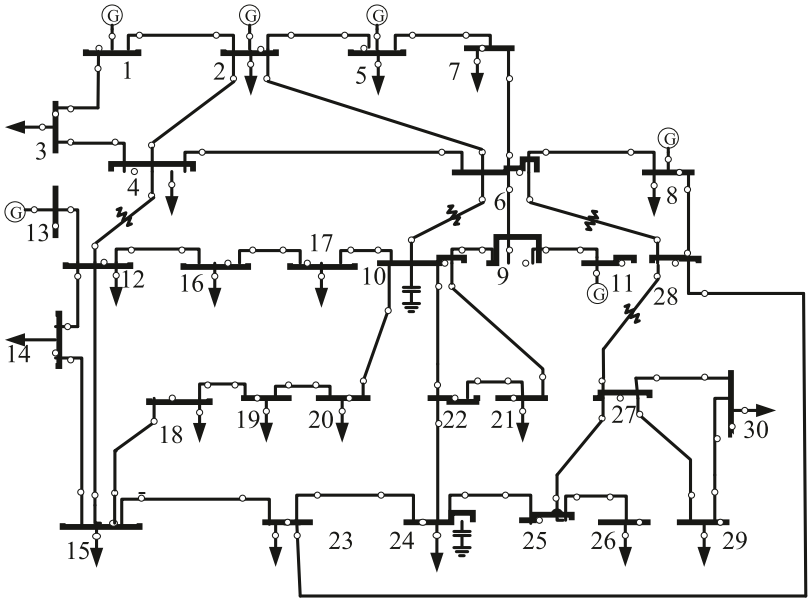



FIGURE 3
IEEE—30 standard bus system.

### 3.3.1 Data preprocessing

We define the state of the system with $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$, where each state $x_i$ contains information on phase angles and voltage magnitudes. The number of episodes in the data-preprocessing process is $E$, and each episode lasts for $T$ time steps. For each time step $t$, we obtain the state of the system $\boldsymbol{x}$ from the IEEE standard bus system, as illustrated in Figure 3. The measurement of the system is expressed as $z = \{z_1, z_2, \ldots, z_m\}$, and it is calculated by Eq. (2). After this, Kalman filter technology is employed to estimate the system state $\hat{\boldsymbol{x}}$ with Eqs. (4, 5). To check the presence of data injection attacks in the power grid system, we calculate the observations of the system with Eq. (12).

The attacker adjusts the attack vector to bypass the traditional bad data detection approach. To generate training data, we inject the attack vector, denoted as $\boldsymbol{a}$, into the measurement

**for** $e = 1$ *to* $E$ **do**
  **for** $t = 1$ *to* $T$ **do**
    Obtain the state of the system $x$ from IEEE standard bus system.
    Calculate the measurement of the system by (2).
    Estimate the system state $\hat{x}$ by (4) and (5).
    Inject the attack into part of the measurement data by (9):
  **end**
  **for** *all measurement data* $\boldsymbol{z}_i$ **do**
    **if** $a \neq 0$ **then**
      $\boldsymbol{z}_i = [\boldsymbol{z}_i, [0, 1]]$.
    **else**
      $\boldsymbol{z}_i = [\boldsymbol{z}_i, [1, 0]]$.
    **end**
  **end**
**end**
Calculate the system observations by (12).
Divide all the observations into training set and testing set.

Algorithm 1. Data preprocessing of the attack detection approach.

data and formulate the process of the data injection attack, as follows:

$$\boldsymbol{z}' = \boldsymbol{z} + \boldsymbol{a}, \tag{13}$$

where $\boldsymbol{z}$ is the original measurement data and $\boldsymbol{z}'$ is the measurement data after the system is attacked. Then we label the measurement data, such that $label = [1, 0]$ means that the data are under attack, and $label = [0, 1]$ means that they are not being attacked. In this way, we transform the attack detection problem into a supervised learning problem, and we utilize a deep neural network to classify the labeled observations. Finally, we divide the labeled observations into a training set and a testing set, using the training set to update the parameters of the neural networks and the testing set to evaluate the accuracy of the attack detection approach. The pseudocode for the data preprocessing is presented in Algorithm 1, in which we obtain the system states in the IEEE standard bus system and inject attacks to the measurements data. Then we label the observations and utilize the deep neural network for classification.

### 3.3.2 Neural network training

As shown in **Figure 2**, we utilize the training data divided from the labeled observations to update the parameters of the neural network. The neural network has an excellent ability to model nonlinear functions Nielsen (2015). The proposed classifier is a deep neural network that consists of three layers. We randomly initialize the parameters $\theta$ of the neural network including weights $\omega$ and bias $b$ at the beginning of the training. In each training episode $e$, we sample a mini-batch of the training data with size $M$. We regard the LSTM layer as the input layer to make full use of the impact of previous observations on the current state, and we feed the LSTM layer with the time series of observations $[o_{t-m}, o_{t-m+1}, \ldots, o_t]$ with length $m$. Therefore, the output of the LSTM layer is given as follows:

$$o(t) = f_1\left(o_{t-m}, o_{t-m+1}, \ldots, o_t\right), \tag{14}$$

where $f_1(\cdot)$ represents the calculation function of the LSTM layer. Then the output of the LSTM layer $o(t)$ is fed into the hidden layer, including two fully-connected neural networks. The output

Randomly initialize the neural network parameters $\theta$.
**for** $e = 1$ *to* $E$ **do**
  Sample the mini-batch with size $|M|$ from the training data.
  Calculate the output of the LSTM layer by (14).
  Obtain the output of the neural network $P$.
  Calculate the loss function by (16).
  Update the parameters of neural network by (17).
**end**

Algorithm 2. Neural network training of the attack detection approach.

is represented by:

$$h(t) = f_3\left(f_2\left(o_t\right)\right). \tag{15}$$

where $f_2(\cdot)$ and $f_3(\cdot)$ are the calculation functions of each fully-connected layer. Finally, the output layer contains two neurons that generate the judgment of the system state based on the system observations in the current parameters of the neural network. Specifically, the output of the neural network is $P = [p_0, p_1]$, where $p_0 > p_1$ indicates that the detector believes that the system is operating normally, and $p_0 < p_1$ denotes that the system is under attack.

The loss function of the entire neural network is indicated as the square of the difference between the outputs and labels:

$$\mathcal{L}(e) = \sum_{i=1}^{M} \left(P_i - label_i\right)^2, \tag{16}$$

where $P_i$ is the output of the neural network from feeding the $i-th$ observations and $label_i$ is the label for the $i-th$ observations in the mini-batch. The gradient of loss function $\nabla_\theta \mathcal{L}(e)$ is back-propagated, and the neural network parameters $\theta$ are updated as follows:

$$\theta = \theta - \alpha \nabla_\theta \mathcal{L}_\theta, \tag{17}$$

where $\alpha$ is the learning rate, which determines neural network training speed.

The pseudocode for the neural network training is given in Algorithm 2, which mainly describes the training process for the neural network, using the sampled mini-batch from the training data. The parameters of the neural networks are updated by the back-propagation of the gradient of loss function.

### 3.3.3 Detection accuracy testing

After the training process, the trained neural network is used to evaluate detection accuracy against malicious attacks. We utilize testing data with size $N_2$ from the labeled observations to determine whether the system can correctly capture false data injection attacks. First, sequential observations are regarded as the input of the trained neural network. Then the network generates the output $P = [p_0, p_1]$. If $p_0 > p_1$, the detector believes that the system is operating normally, and if $p_0 < p_1$, the detector thinks the system is under attack. Finally, we compare the output of the detector with the label of the observations to indicate the correctness of detection and define the number of correctly

Load the parameters of trained neural network.
Initialize the sum of the detection error $n = 0$.
**for** $e = 1$ *to* $N_2$ **do**
  Feed the observation $o_e$ into the trained neural network.
  Obtain the output of the neural network $P = [p_0, p_1]$.
  **if** $p_0 > p_1$ **then**
    $s_e = s_n$.
  **else**
    $s_e = s_a$.
  **end**
  **if** $s_e \neq label_e$ **then**
    $n = n + 1$.
  **end**
**end**
Calculate the detection accuracy as $1 - \frac{n}{N_2}$

Algorithm 3. Detection accuracy testing.

classified samples, divided by the total number of samples as the metric. The procedure for detection accuracy testing is outlined in Algorithm 3, in which it can be seen that the trained neural network is utilized to evaluate the accuracy of detecting the malicious data injection attacks.

# 4 Experiments

In this section, we evaluate the detection accuracy of the proposed detection approach in IEEE standard bus systems. We first describe the experimental settings. Then, the detection accuracy of our proposed detection approach is presented, compared with benchmarks. After that, we evaluate the impact of different attack amplitudes and different model parameters on the performance of detection accuracy.

## 4.1 Experimental settings

### 4.1.1 Parameters

We investigate the performance of our proposed attack detection approach in IEEE-9, 14, 30, 118, and 300 standard bus systems. The initial state value of the system and measurement matrix are obtained from MATPOWER Zimmerman et al. (2010). The main parameters of the proposed detection approach are presented in **Table 2**. Specifically, we set the size of the observation sequence of the LSTM layer as 4. We set the number of training data accounting for 90% of the total number of labeled observations, such that the testing process utilizes 10% of the observations to evaluate detection performance. The amplitude of the attack is 1% of the measurement data. The size of the mini-batch is set as 40. The number of episodes $E$ is set to 100, and each episode lasts for 50 time steps. In addition, the learning rate $\alpha$ is set to 0.001, and the structure of hidden layer is set as (36, 64, 64).

TABLE 2 Parameters of proposed detection approach.

| Parameter | Value |
|---|---|
| Size of the observation sequence $m$ | 4 |
| The proportion of the training set | 90% |
| The proportion of the testing set | 10% |
| Attack amplitude | 1% |
| Size of the mini-batch $M$ | 40 |
| Number of episodes $E$ | 100 |
| Number of time steps $T$ | 50 |
| Learning rate $\alpha$ | 0.001 |
| Structure of hidden layers | (36, 64, 64) |

### 4.1.2 Benchmarks

To evaluate its effectiveness, we compare the proposed attack-detection method with the following two benchmarks:

- BPNN: Back propagation neural network-based (BPNN) detection approach utilizes a fully connected neural network to detect whether the grid system is under attack. The BPNN-based detection approach does not utilize the LSTM layer to extract the previous observations to estimate the current state. The remaining settings are the same as those of the proposed approach.
- BPNN-imp: BPNN-imp is an enhanced detection approach based on the BPNN. It utilizes the concept of a sliding window to input multiple observations into the neural network at once Kurt et al. (2018a). We set the size of the sliding window to the same length as $m$. The remaining settings are the same as a BPNN-based approach.

### 4.1.3 Attack scenario

To demonstrate the effectiveness of the proposed detection mechanism in improving the detection accuracy, we introduce two types of attack scenarios: continuous attack and discontinuous attack. The details of these attack scenarios are given in the following:

- continuous attack: in a continuous attack scenario, the attack is launched at the half time step of the episode, i.e., $t = \frac{T}{2}$, and the attack is sustained until the end of the episode.
- discontinuous attack: in the discontinuous attack scenario, the attack is launched at any time step after the half time step of the episode, i.e., $t = \frac{T}{2}$. Each time steps after $t = \frac{T}{2}$, the system has 50% probability of being attacked.

## 4.2 Results of attack detection

We utilize detection accuracy as the evaluation metric to identify the effectiveness of the proposed detection method. The detection accuracy is defined as the number of correctly classified samples, divided by the total number of samples.

We first conduct experiments to compare the detection accuracy of the proposed attack detection approach with BPNN-based approach and a BPNN-imp-based approach against a

TABLE 3 Detection accuracy in continuous attack scenario.

| Systems | IEEE-9 | IEEE-14 | IEEE-30 | IEEE-118 | IEEE-300 |
|---|---|---|---|---|---|
| Proposed | 0.9422 | 0.9452 | 0.9424 | 0.9442 | 0.9368 |
| BPNN-imp | 0.875 | 0.8692 | 0.8626 | 0.8728 | 0.8628 |
| BPNN | 0.7362 | 0.7462 | 0.7424 | 0.7424 | 0.7442 |

TABLE 4 Detection accuracy in discontinuous attack scenario.

| Systems | IEEE-9 | IEEE-14 | IEEE-30 | IEEE-118 | IEEE-300 |
|---|---|---|---|---|---|
| Proposed | 0.7634 | 0.7598 | 0.757 | 0.7642 | 0.7582 |
| BPNN-imp | 0.6794 | 0.6792 | 0.6754 | 0.6784 | 0.674 |
| BPNN | 0.5918 | 0.5972 | 0.6004 | 0.6096 | 0.5944 |

continuous false data injection attack under the IEEE-9, 18, 30, 118, and 300 bus standard systems. Table 3 shows the simulation results:

Table 3 proves that our proposed detection approach outperforms BPNN-based approach and the BPNN-imp-based approach on the continuous attack scenario in terms of attack detection accuracy. Specifically, the average detection accuracy of the proposed approach reaches about 0.9422, 0.9452, 0.9424, 0.9442, and 0.9368 in the IEEE-9, 14, 30, 118, and 300 systems, respectively. Obviously, the proposed detection approach significantly outperforms the benchmarks in detecting the continuous false data injection attacks, which achieves a 7.7, 8.7, 9.3, 8.2, and 8.6% higher detection accuracy than the BPNN-imp-based approach and a 28.0, 26.7, 26.9, 27.2, and 25.9% higher detection accuracy than the BPNN-based approach in IEEE-9, 14, 30, 118, and 300 systems, respectively. In addition, the accuracies of the three detection approaches are basically unchanged under different systems, which demonstrates that the complexity of the system has no impact on the performance of detection accuracy against a continuous attack.

We then compare the detection accuracy of our proposed detection approach with benchmarks for the discontinuous attack model. Table 4 shows the detection accuracy for detection approaches on a discontinuous attack scenario. The results in Table 4 are in general same as those for the continuous attacks. The average detection accuracy against the discontinuous attack of the proposed approach reaches 0.7634, 0.7598, 0.757, 0.7642, and 0.7582 in IEEE-9, 14, 30, 118, and 300 systems, respectively. Clearly, the proposed detection approach achieves a 12.4, 11.9, 12.1, 12.6,and 12.5% higher detection accuracy than the BPNN-imp-based approach and a 29.0, 27.2, 26.1, 25.4, and 27.6% higher detection accuracy than the BPNN-based approach in IEEE-9, 14, 30, 118, and 300 systems, respectively. Furthermore, as can be seen in Tables 3, 4, discontinuous attacks are more difficult to detect than continuous ones, which have lower detection accuracy with the same approaches and the same testing systems.

## 4.3 Training time

Next, we investigate the training time of the proposed detection approach in different systems in Figure 4, where we can see that, although the detection accuracy of the proposed detection approach under different systems is substantially equal, there is a large difference in the training time. Specifically, as the complexity of the system gradually increases, the running time of the detection approach increases significantly. The running time of the detection approach for the IEEE-9 bus system is only 6.57 s, and the running time of the detection approach on IEEE-300 bus system increases by nearly 118 times, to 777.32 s.

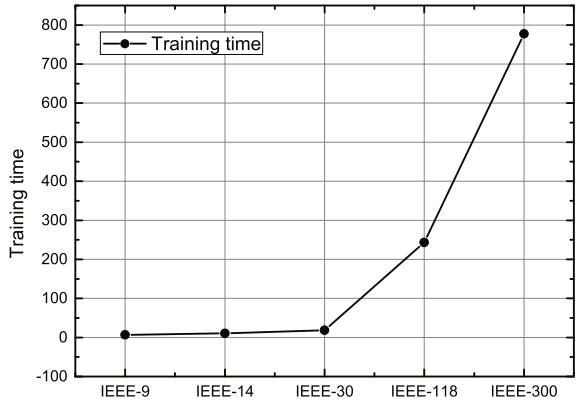



FIGURE 4
Training time of different systems.

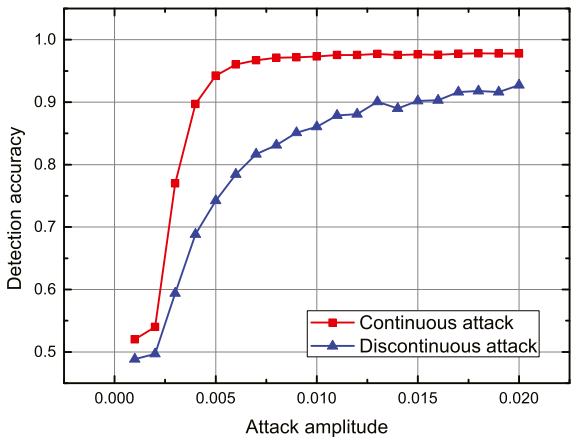



FIGURE 5
Detection delay under different attack amplitudes.

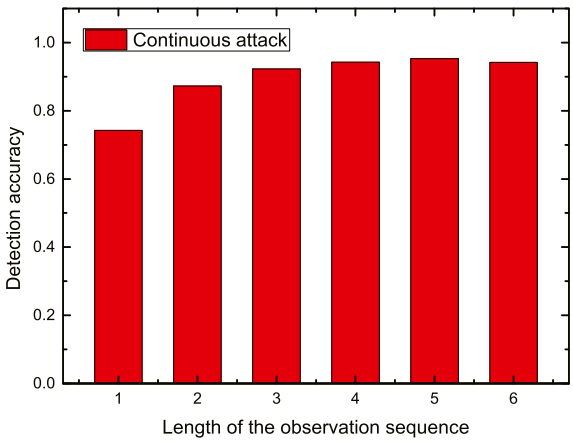



FIGURE 6
Detection accuracy at different *m* under continuous attacks.

## 4.4 Discussion of attack amplitude

We now consider the impacts of different attack amplitudes on the performance of detection accuracy on an IEEE-30 bus standard system. The results are shown in Figure 5, in which we set the attack amplitude to increase from 0.1 to 2% at a step of 0.1%. In Figure 5, the red line with square marks represents detection accuracy of the continuous attacks. The blue line with triangle marks illustrates the detection accuracy of discontinuous attacks. From Figure 5 we can see that, with increasing attack amplitude, detection accuracy also increases, which demonstrates that the attacks with larger amplitudes are easier to identify by the detection approach. We can also see from Figure 5 that the increment of the detection accuracy is more obvious with increasing attack amplitude when the attack amplitude is small (e.g., the attack amplitude is larger than 0.1% and less than 0.5%). As the attack amplitude gradually increases, the growth rate of the detection accuracy also gradually slows. When the attack amplitude reaches a certain level, the detection accuracy tends to be stable. Moreover, the increment of detection accuracy with the increase in attack amplitude in a continuous attack scenario is faster than that in a discontinuous attack scenario, and the detection accuracy in a continuous attack scenario is always higher than that in a discontinuous attack.

## 4.5 Impact of training parameters

Finally, we evaluate the impact of the observation sequence length *m* on the performance of the proposed detection approach in the IEEE-30 bus standard system. The results are shown in Figures 6 and 7, where Figure 6 represents the results of detection accuracy at different *m* under continuous attack, and Figure 7 represents the results of detection accuracy at different

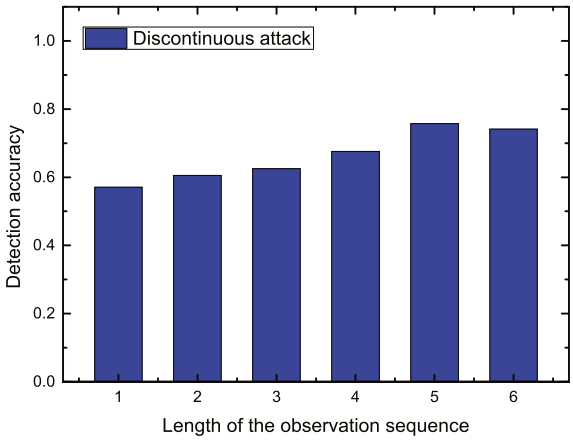



FIGURE 7
Detection accuracy at different *m* under discontinuous attacks.

*m* under discontinuous attack. From Figures 6 and 7, we can see that, when the observation sequence length *m* is larger than 1 and smaller than 4, the increase of *m* results in the improvement of the detection accuracy. However, when the observation sequence length *m* is larger than 4, the increase of *m* has little impact on the accuracy of detection, and the detection accuracy for both continuous and discontinuous attacks tends to converge to a certain value.

# 5 Conclusion

In this study, we propose an LSTM-based false data injection attack detection approach for dynamic state estimation in a smart grid. We propose a neural network model that utilizes

the LSTM network to extract the previous observations to determine the current state estimation. We transform a malicious data injection attack detection into supervised learning and train the proposed deep neural network for classification. We conduct extensive experiments to illustrate the effectiveness of proposed detection method and investigate the impact of attack amplitudes and model parameters on detection accuracy. The simulation results demonstrate that the proposed detection approach outperforms BPNN-imp-based approach and BPNN-based approach in detection accuracy.

## Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: https://www.mathworks.com/matlabcentral/fileexchange/72085-matpower.

## Author contributions

FZ: conceptualization, methodology, software, investigation, formal analysis, writing—original draft. QY (corresponding author): conceptualization, funding acquisition, resources, supervision, writing—review and editing.

## Funding

The work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1704103, in part by the National Science Foundation of China under Grants 61973247, 61673315 and 62173268, in part by the Fundamental Research Funds for the Central Universities under Grant xzy022021050.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Chakhchoukh, Y., Lei, H., and Johnson, B. K. (2019). Diagnosis of outliers and cyber attacks in dynamic pmu-based power state estimation. *IEEE Trans. Power Syst.* 35, 1188–1197. doi:10.1109/tpwrs.2019.2939192

Cintuglu, M. H., Mohammed, O. A., Akkaya, K., and Uluagac, A. S. (2016). A survey on smart grid cyber-physical system testbeds. *IEEE Commun. Surv. Tutorials* 19, 446–464. doi:10.1109/comst.2016.2627399

Connolly, L. Y., Lang, M., and Wall, D. S. (2019). Information security behavior: A cross-cultural comparison of Irish and us employees. *Inf. Syst. Manag.* 36, 306–322. doi:10.1080/10580530.2019.1651113

Dayaratne, T., Salehi, M., Rudolph, C., and Liebman, A. (2022). False data injection attack detection for secure distributed demand response in smart grids," in 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Baltimore, MD, USA, June 27–30, 2022, 367–380. doi:10.1109/DSN53405.2022.00045

Deng, R., Xiao, G., and Lu, R. (2015). Defending against false data injection attacks on power system state estimation. *IEEE Trans. Ind. Inf.* 13, 198–207. doi:10.1109/tii.2015.2470218

Ding, Y., and Liu, J. (2017). "Real-time false data injection attack detection in energy internet using online robust principal component analysis," in 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, November 26–28, 2017, 1–6. doi:10.1109/EI2.2017.8245663

Fan, W., and Li, Y. (2009). "Accuracy analysis of sigma-point kalman filters," in *2009 Chinese control and decision conference* (IEEE), 2883.

Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural Comput.* 12, 2451–2471. doi:10.1162/089976600300015015

Giani, A., Bitar, E., Garcia, M., McQueen, M., Khargonekar, P., and Poolla, K. (2011). "Smart grid data integrity attacks: Characterizations and counter measures," in 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium, October 17–20, 2011, 232–237. doi:10.1109/SmartGridComm.2011.6102324

Guan, Y., and Ge, X. (2017). Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks. *IEEE Trans. Signal Inf. Process. Netw.* 4, 48–59. doi:10.1109/tsipn.2017.2749959

Hu, Z., Wang, Y., Tian, X., Yang, X., Meng, D., and Fan, R. (2015). "False data injection attacks identification for smart grids," in 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), Beirut, Lebanon, April 26–May 1, 2015, 139–143. doi:10.1109/TAEECE.2015.7113615

James, J., Hou, Y., and Li, V. O. (2018). Online false data injection attack detection with wavelet transform and deep neural networks. *IEEE Trans. Ind. Inf.* 14, 3271–3280. doi:10.1109/tii.2018.2825243

Karimipour, H., and Dinavahi, V. (2017). Robust massively parallel dynamic state estimation of power systems against cyber-attack. *IEEE Access* 6, 2984–2995. doi:10.1109/access.2017.2786584

Kurt, M. N., Ogundijo, O., Li, C., and Wang, X. (2018a). Online cyber-attack detection in smart grid: A reinforcement learning approach. *IEEE Trans. Smart Grid* 10, 5174–5185. doi:10.1109/tsg.2018.2878570

Kurt, M. N., Yılmaz, Y., and Wang, X. (2018b). Distributed quickest detection of cyber-attacks in smart grid. *IEEE Trans. Inf. Forensic. Secur.* 13, 2015–2030. doi:10.1109/tifs.2018.2800908

Lee, I., and Lee, K. (2015). The internet of things (iot): Applications, investments, and challenges for enterprises. *Bus. horizons* 58, 431–440. doi:10.1016/j.bushor.2015.03.008

Li, S., Yılmaz, Y., and Wang, X. (2014). Quickest detection of false data injection attack in wide-area smart grids. *IEEE Trans. Smart Grid* 6, 2725–2735. doi:10.1109/tsg.2014.2374577

Nielsen, M. A. (2015). *Neural networks and deep learning*, 25. San Francisco, CA, USA: Determination press.

Rahman, M. A., and Mohsenian-Rad, H. (2013). False data injection attacks against nonlinear state estimation in smart power grids in 2013 IEEE Power & Energy Society General Meeting (IEEE), Vancouver, BC, July 21–25, 2013, 1–5. doi:10.1109/PESMG.2013.6672638

Rahman, M. A., and Mohsenian-Rad, H. (2012). "False data injection attacks with incomplete information against smart power grids," in 2012 IEEE Global Communications Conference (GLOBECOM) (IEEE), Anaheim, CA, USA, December 3–7, 2012, 3153–3158. doi:10.1109/GLOCOM.2012.6503599

Sandberg, H., Teixeira, A., and Johansson, K. H. (2010). "On security indices for state estimators in power networks," in *First workshop on secure control systems (SCS)* Stockholm.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). "Sequence to sequence learning with neural networks," in 27th Advances in Neural Information Processing Systems (NIPS 2014), Montreal, Canada, December 7–14, 2014. doi:10.5555/2969033.2969173

Taha, A. F., Qi, J., Wang, J., and Panchal, J. H. (2016). Risk mitigation for dynamic state estimation against cyber attacks and unknown inputs. *IEEE Trans. Smart Grid* 9, 886–899. doi:10.1109/tsg.2016.2570546

Ünal, F., Almalaq, A., Ekici, S., and Glauner, P. (2021). Big data-driven detection of false data injection attacks in smart meters. *IEEE Access* 9, 144313–144326. doi:10.1109/ACCESS.2021.3122009

Yang, Q., An, D., Min, R., Yu, W., Yang, X., and Zhao, W. (2017). On optimal pmu placement-based defense against data integrity attacks in smart grid. *IEEE Trans. Inf. Forensic. Secur.* 12, 1–1750. doi:10.1109/tifs.2017.2686367

Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. (2010). Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. Power Syst.* 26, 12–19. doi:10.1109/tpwrs.2010.2051168

frontiers | Frontiers in Energy Research

TYPE Methods
PUBLISHED 20 September 2022
DOI 10.3389/fenrg.2022.1020676



OPEN ACCESS

EDITED BY
Dou An,
MOE Key Laboratory for Intelligent Networks and Network Security, China

REVIEWED BY
Shengquan Li,
Yangzhou University, China
Dezhi Xu,
Jiangnan University, China

*CORRESPONDENCE
Zetao Feng,
fengzetao118@126.com

SPECIALTY SECTION
This article was submitted to Smart Grids,
a section of the journal
Frontiers in Energy Research

RECEIVED 16 August 2022
ACCEPTED 26 August 2022
PUBLISHED 20 September 2022

CITATION
Xia Y, Feng Z, Chen R, Wu J and Huang Q (2022), A harmonic suppression strategy for grid-connected inverters based on quadrature sinewave extractor.
*Front. Energy Res.* 10:1020676.
doi: 10.3389/fenrg.2022.1020676



# A harmonic suppression strategy for grid-connected inverters based on quadrature sinewave extractor

Yan Xia[1,2], Zetao Feng[1,3]*, Renzhao Chen[4], Jie Wu[3] and Qinyuan Huang[1,2]

[1]School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yinbin, China, [2]Intelligent Electric Power Grid Key Laboratory of Sichuan Province, Sichuan University and State Grid Sichuan Electric Power Company, Chengdu, China, [3]Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yinbin, China, [4]Zonergy Co., Ltd., Zigong, China

Grid-connected inverters need to reduce current harmonics as much as possible. After introducing the input signal's fundamental and main harmonic quadrature components, a discrete state model is created, and the discrete observer design method is used to propose a harmonic extraction algorithm called quadrature sinewave extractor (QSE). The QSE is a stable recursive operator with the advantages of no phase displacement and the ability to eliminate mutual influence between harmonic components. Compared to the widely used proportional multi-resonant controller, QSE can reduce current harmonics and improve system stable performance by using it in the current control of grid-connected inverters. Finally, comparative experiments on a three-phase grid-connected inverter are used to verify the proposed method's effectiveness.

KEYWORDS

**grid-connected inverters, harmonic suppression, quadrature sinewave extractor, phase displacement, observer**

## Introduction

The exhaustion and pollution of traditional fossil energy promote the development and utilization of renewable energy such as solar and wind energy, and the proportion of new energy power generation in the power system increases greatly. The grid-connected inverter converts DC energy into AC energy, and its performance directly affects the power grid. One key indicator of new energy power quality is the inverter output current harmonics. The requirements for outputting current harmonic content are clearly stated in IEEE 1547–2003 and other specifications, so lowering the harmonic content is critical (Lee and Cho, 2020).

A variety of factors can cause the output current harmonics of the inverter. The modulation dead time of the inverter's upper and lower bridge arms, the voltage drop of power switching devices, the grid voltage distortion, the fluctuation of DC power supply,

the voltage fluctuation of the DC bus capacitor, and the nonlinearity of the system transformer or reactor are the main causes of inverter output current harmonics. The harmonic suppression method studied in this paper is to detect the harmonic in the current and realize harmonic suppression through closed-loop feedback. Therefore, this paper aims at the harmonic generated by any of the above reasons.

The LCL filter can filter out the output current's high-frequency harmonics, and the current closed-loop controller is primarily responsible for suppressing the low-frequency current harmonics (Chen et al., 2013; Yang et al., 2015). The current closed-loop controller is an important factor affecting output current harmonics' performance (Cárdenas et al., 2012; Tuan and Santoso, 2016). There are many effective current closed-loop control algorithms for suppressing current harmonics, and the main difference between them is the harmonic extraction method (Castilla et al., 2013; Kulkarni and John, 2013; Zhao et al., 2013; Bourguiba et al., 2016; Wang et al., 2016; Qian et al., 2017; Zhaoyang et al., 2017; Busarello et al., 2018; Hu et al., 2018; McDonald and Li, 2018; Chen et al., 2019; Choi and Sarlioglu, 2019; Fei et al., 2019; Song et al., 2020).

The repetitive controller (RC) is derived from the internal model control principle (Busarello et al., 2018; McDonald and Li, 2018), which takes advantage of harmonics' periodic characteristics. A positive feedback loop with periodic delay is chosen as the internal model, assuming that the harmonic signal repeatedly occurs in each fundamental cycle. As long as the input signal contains a harmonic signal that is an integral multiple of the fundamental frequency, the internal model's output can continuously accumulate the harmonic signal and suppress the harmonic by multiplying the output harmonic signal with a certain coefficient as the control voltage. The advantage of RC is that all harmonics can be extracted with simple delayed positive feedback. In comparison, the main disadvantage of RC is that it has a resonance effect on all harmonics, rather than focusing on suppressing some major harmonics. In addition, RC needs to save the period data of the fundamental wave, which cannot extract non-integer harmonics. So the control frequency of RC is required to be an integral multiple of the fundamental frequency, and it is not suitable for frequency variation. These shortcomings limit its application in practice (McDonald and Li, 2018).

Proportional multi-resonance control uses multiple quasi-resonance controllers (MQR) to suppress some major low-order harmonics (Castilla et al., 2013; Bourguiba et al., 2016; Wang et al., 2016; Qian et al., 2017; Zhaoyang et al., 2017; Hu et al., 2018; Choi and Sarlioglu, 2019; Song et al., 2020). It takes advantage of the fact that the amplitudes of some lower harmonics, especially odd harmonics, are usually larger than that of the higher harmonics (Qian et al., 2017). A resonant controller is used as its internal mode for harmonic extraction and suppression of the main harmonics to be suppressed. In practice, a quasi-resonant controller is generally formed by adding appropriate damping to the system to deal with a slight deviation of the fundamental frequency. Introducing a resonance controller can significantly improve the open-loop gain of the transfer function at the dominant harmonics frequency and reduces the harmonic content. However, the defect of resonant controller is that while extracting the desired frequency harmonics, it will also pass through other harmonics. The phase shift of other harmonics will increase the amplitude of corresponding harmonics and even make the system diverge (Zhao et al., 2013).

There are also some intelligent harmonic extraction methods, such as least mean square filter (Kulkarni and John, 2013), adaptive neural network filter (Fei et al., 2019), and discrete Fourier transform based on extended Kalman filter (Chen et al., 2019). However, their design process is complex, and the calculation time is long, so they are not widely used. The most widely harmonic suppression algorithms used are proportional multi resonance and repetitive control at present.

In view of the shortcomings of MQR in extracting harmonics (Kulkarni and John, 2013), this paper designs a new discrete harmonic extractor called quadrature sine wave extractor (QSE), which uses the idea of the observer (Quan et al., 2016; Wang, 2016) to extract multiple harmonic components at the same time, which is an improvement of MQR. QSE will not cause the phase shift of the frequency component to be extracted, which avoids the phase shift problem caused by the discretization of controller design in the traditional continuous domain. QSE extracts each frequency component simultaneously, avoiding the coupling problem caused by the traditional MQR to extract each frequency component separately. Experiments show that the high-performance harmonic extraction method is conducive to reducing output current harmonics of the grid-connected inverter.

# Quadrature sinewave extractor principle

## Orthogonal sine wave model

According to the Fourier decomposition theory, the periodic signal can be decomposed into a linear combination of multiple cosine waves:

$$u(n) = \sum_{k\epsilon\Theta} u_{ck} = M_k \cos(nk\omega T + \delta_k) \quad (1)$$

Where, $u$ is the input signal; $n$ is the $n$th sampling time; $u_{ck}(n) = M_k\cos(nk\omega T + \delta_k)$. $u_{ck}(n)$ is the $k$th cosine wave. When k is zero, $u_{ck}(n)$ is the DC component; when k is 1, it is the fundamental wave; when k is greater than 1, it is the $k$th harmonic. In this paper, $u_{ck}(n)$ is collectively referred to as the $k$th cosine wave. $M_k$ is the $k$th cosine wave amplitude; $\delta_k$ is the initial phase of the $k$th cosine wave; $T$ is the sampling time interval; $\omega$ is the fundamental

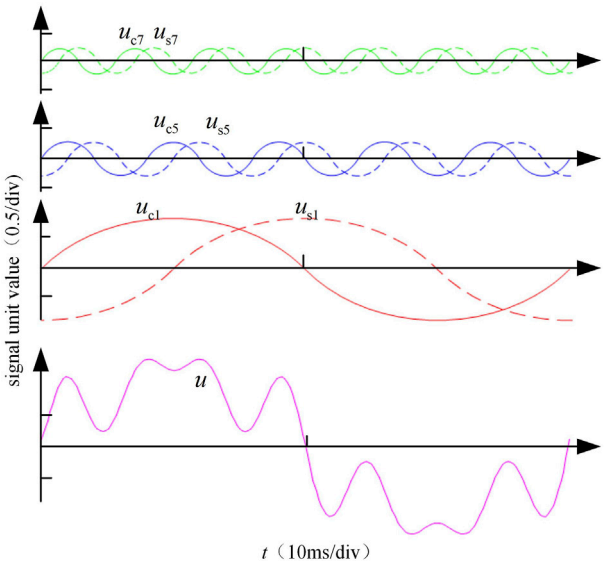



FIGURE 1
Typical periodic signal.

angular frequency; Θ is the value set of k. Considering the main low-order harmonics of the actual three-phase grid-connected inverter are 5 and 7 times in practice (Bourguiba et al., 2016), and it is also necessary to control the fundamental wave, the Θ of QSE is set as {1,5,7}.

In order to establish the mathematical model, the sinusoidal signal delaying one-quarter of the period of each cosine wave is introduced. That is $u_{sk}(n) = M_k \sin(nk\omega T + \delta_k)$, $u_{ck}$, and $u_{sk}$ form a set of orthogonal sine waves called cosine and sine wave, respectively. Taking the amplitude and initial phase of the orthogonal sine wave as constants, the state equations of $u_{ck}$ and $u_{sk}$ can be expressed:

$$\begin{bmatrix} u_{ck}(n) \\ u_{sk}(n) \end{bmatrix} = \mathbf{R}_k \begin{bmatrix} u_{ck}(n-1) \\ u_{sk}(n-1) \end{bmatrix} = \begin{bmatrix} \cos(k\omega T) & -\sin(k\omega T) \\ \sin(k\omega T) & \cos(k\omega T) \end{bmatrix} \begin{bmatrix} u_{ck}(n-1) \\ u_{sk}(n-1) \end{bmatrix} \quad (2)$$

Where, $\mathbf{R}_k = \begin{bmatrix} \cos(k\omega T) & -\sin(k\omega T) \\ \sin(k\omega T) & \cos(k\omega T) \end{bmatrix}$ is a counterclockwise rotation transformation matrix, which means that the current orthogonal sine wave value is transformed counterclockwise by a small angle $k\omega T$ from the orthogonal sine wave value of the last time. Figure 1 shows the waveform of $u$ in a fundamental period, which is a typical periodic signal. The QSE can also extract other signals such as current signal and voltage signal.

The cosine wave in Figure 1 includes fundamental wave $u_{c1}$, fifth harmonic $u_{c5}$ and seventh harmonic $u_{c7}$, that is $u = u_{c1} + u_{c5} + u_{c7}$. The dotted lines in Figure 1 are sine waves, namely $u_{s1}$, $u_{s5}$ and $u_{s7}$.

## Quadrature sinewave extractor design

The basic idea of QSE is to inversely deduce each orthogonal sine wave according to the sampled signal based on state Eq. 2 and output Eq. 1. Considering the variables $u_{ck}$ and $u_{sk}$ as the state of the system and $u$ as the measured value of the system, then the harmonic extraction problem is transformed into a typical state observer design problem. Based on this idea, the QSE proposed in this paper is designed as a discrete closed-loop block diagram, as shown in Figure 2.

Figure 2 shows the unit negative feedback structure. If the number of elements in the set Θ is N, the forward channel contains N oscillators in parallel. QSE design includes three steps: prediction, calculation of prediction error, and correction.

The first step is prediction, as shown in the red part of Figure 2. According to the prediction of model Eq. 2, the prediction formula of the $k$th orthogonal sine wave is:

$$\begin{bmatrix} x'_{ck}(n) \\ x'_{sk}(n) \end{bmatrix} = \mathbf{R}_k \begin{bmatrix} x_{ck}(n-1) \\ x_{sk}(n-1) \end{bmatrix} \quad (3)$$

Where, $x_{ck}(n\text{-}1)$ and $x_{sk}(n\text{-}1)$ are the estimated value of the $k$th orthogonal sine wave at the last time, and the initial value of its iteration can be taken as zero, which is reflected in the output of the $z^{-1}$ module in Figure 1 $x'_{ck}(n)$ and $x'_{sk}(n)$ are the predicted value of the $k$th orthogonal sine wave at the current time.

The second step is calculating the prediction error, as shown in the blue part of Figure 2. The calculation equation of prediction error is:

$$e'(n) = u(n) - x'_c(n) \quad (4)$$

Where $e'(n)$ is the prediction error, $x'_c(n) = \sum_{k\in\Theta} x'_{ck}(n)$ is the total prediction.

The third step is correction, as shown in the green part in Figure 2. The correction formula of the $k$th orthogonal sine wave is:

$$\begin{cases} x_{ck}(n) = x'_{ck}(n) + \rho e'(n) \\ x_{sk}(n) = x'_{sk}(n) \end{cases} \quad (5)$$

Where $x_{ck}(n)$ and $x_{sk}(n)$ are the current extraction results of the orthogonal sine wave, $\rho$ is the update coefficient. The total estimated value after correction is $x_c(n) = \sum_{k\in\Theta} x_{ck}(n)$. In each sampling period, (Eqs. 3–5) form QSE in turn.

## Performance analysis and comparison

**Property 1.** When the update coefficient $\rho$ meets $0 < \rho < 2 / N$, QSE is convergent.

**Property 2.** The k times fundamental frequency components contained in $x_{ck}$ and $x_{sk}$ of QSE steady-state output are equal to

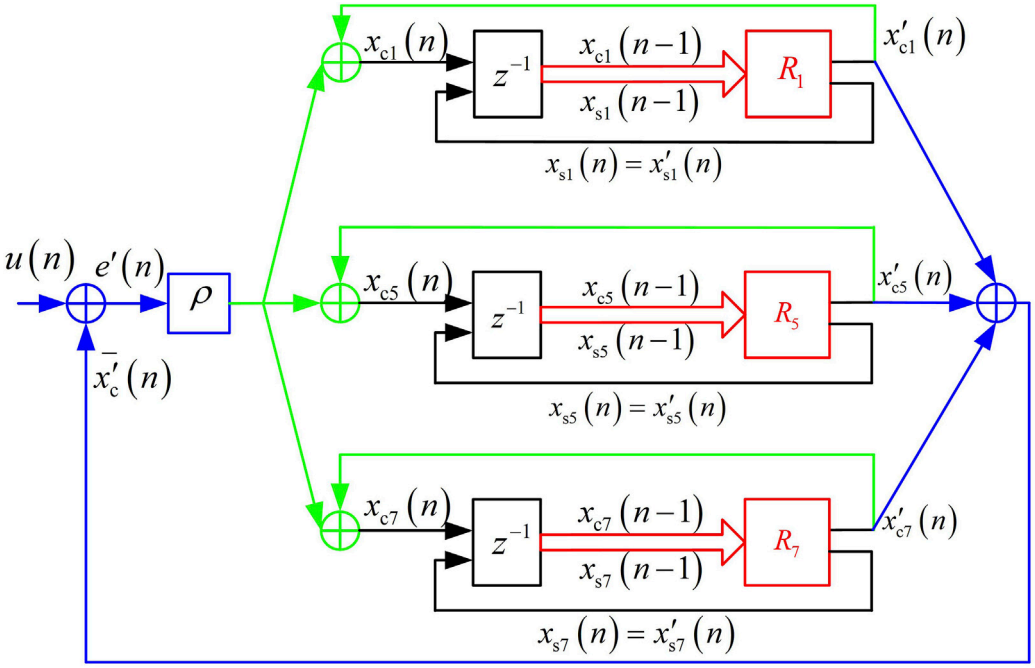



FIGURE 2
Block diagram of QSE.

$u_{ck}$ and $u_{sk}$ respectively, and there is no phase offset and amplitude change between $x_{ck}/x_{sk}$ and $u_{ck}/u_{sk}$.

**Property 3.** $x_{ck}$ and $x_{sk}$ of QSE steady-state output do not contain any other components in sets Θ except k times fundamental frequency components.

The above three properties are called the stability, zero-error and decoupling of QSE, respectively. The process of proving the three properties is shown in Appendix A.

The discrete transfer function of the $k$th oscillator can be obtained from (Eqs. 3–5):

$$G_k(z) = \frac{x'_{ck}(z)}{e'(z)} = \rho \frac{\cos(k\omega T)z - 1}{z^2 - 2\cos(k\omega T)z + 1} \tag{6}$$

Substituting $z = e^{jk\omega T}$ into Eq. 6 at frequency $k\omega$, the following equation can be written:

$$\lim_{z \to e^{jk\omega T}} |G_k(z)| = \infty \tag{7}$$

Eq. 7 shows that the open-loop gain of the system is infinite, so the steady-state error of that is zero. Therefore, the oscillator has the effect of resonance to the $k$th cosine wave or sine wave. Since QSE is stable, there must be no $k$th cosine or sine wave in the steady-state error. Otherwise, the oscillator's output will become larger, eventually leading to QSE output divergence. The zero-error property of QSE is further explained based on the above discussions.

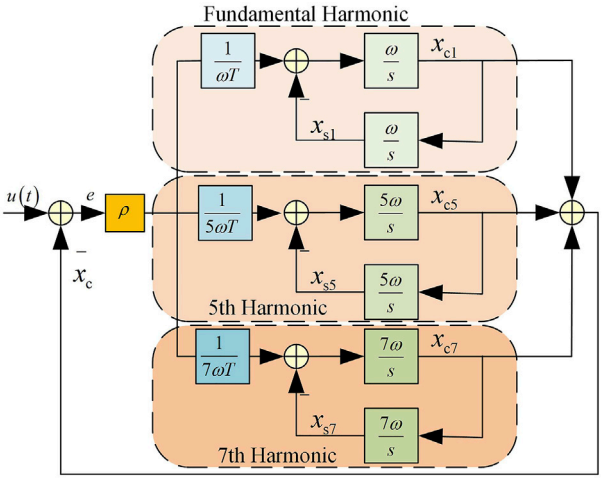



FIGURE 3
Corresponding harmonic observer in continuous domain of QSE.

Further, the transfer function of QSE can be deduced as follow

$$G_{QSE}(z) = \frac{x_c(z)}{u(z)} = \frac{\rho + \sum_{k\in\Theta} G_k(z)}{1 + \sum_{k\in\Theta} G_k(z)} \tag{8}$$

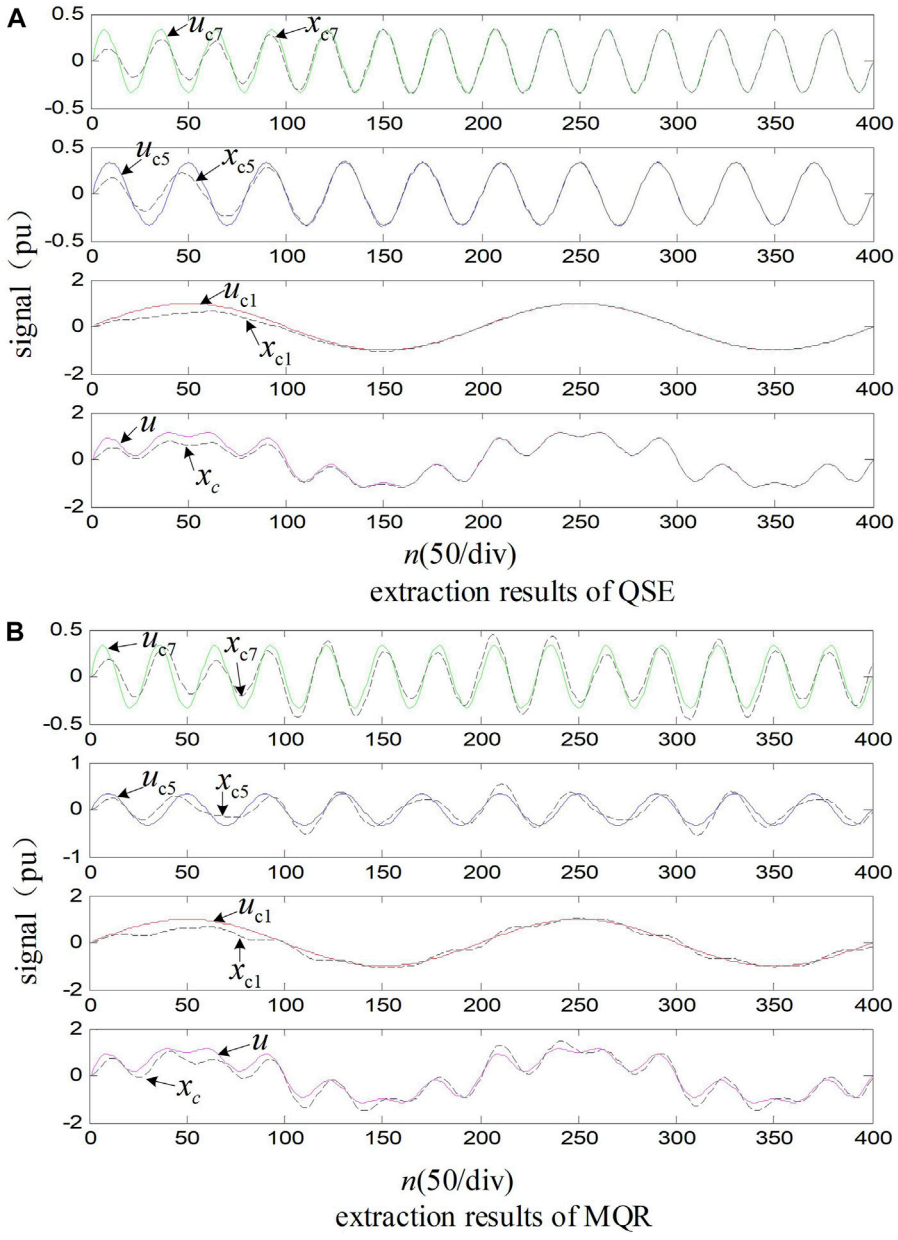



FIGURE 4
Comparation of extraction results of QSE and MQR, "signal" is the signal to be extracted and "n" is the sampling period.

According to the transformation relationship between the *s*-domain and *z*-domain transfer function, the continuous domain observer of QSE is shown in Figure 3.

If there is only one oscillator in the forward channel in the QSE, the QSE degenerates into a band-pass filter, that is, a quasi-resonant controller. The continuous transfer function of the *k*th band-pass filter is:

$$\frac{x_{ck}(s)}{u(s)} = \frac{Bs}{s^2 + Bs + (k\omega)^2} \tag{9}$$

Where, $B$ is the filter bandwidth, and its relationship with the update coefficient $\rho$ of QSE is $BT = \rho$. This formula establishes the relationship between the update coefficient and filter bandwidth, and provides a reference for selecting the update coefficient $\rho$.

MQR uses multiple band-pass filters to extract harmonics, respectively. The output of each band-pass filter must contain other extracted harmonics. For example, the output of the fundamental band pass filter not only contains certain fifth and seventh harmonics components, but also has a certain

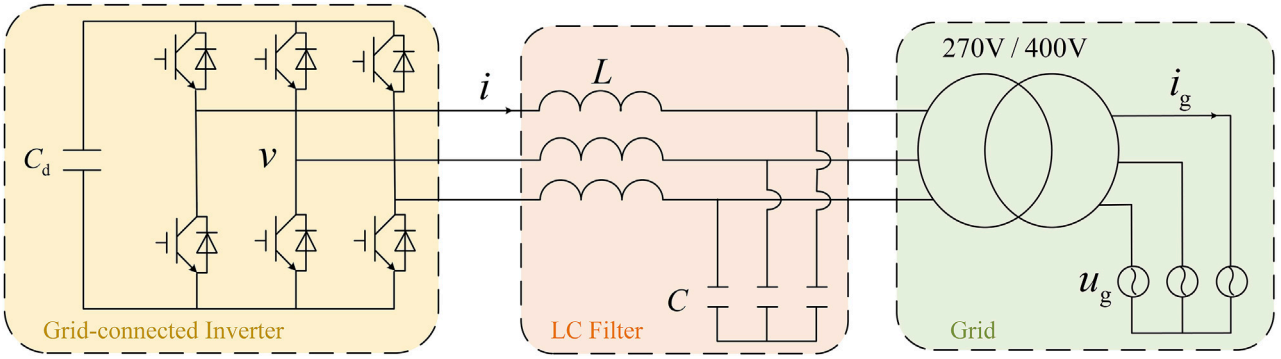



FIGURE 5
Three-phase grid-connected inverter.

phase shift to the fifth and seventh harmonics. Correspondingly, the output of the fifth or seventh harmonic bandpass filter also contains the fundamental component and has a certain phase shift for the fundamental wave, reflecting the interaction between the components extracted with MQR.

The input signal shown in Figure 1 is extracted and simulated by QSE and MQR. Their output comparison results are shown in Figure 4.

In Figure 4, the initial value of each orthogonal sine wave is taken as zero, the update coefficient as 0.05, and the sampling period as 100 μs? The number that one fundamental cycle can be updated is 20 m/100 μs= 200. Figure 4A shows the QSE extraction results, it can be seen that in less than half a fundamental cycle, each output sine wave basically coincides with each component of the input signal, and the total output $x_c$ (sum of cosine waves) also coincides with the input signal, achieving an ideal extraction effect. The extraction effect of MQR on each component of the input signal under the same parameters and initialization conditions as QSE is shown in Figure 4B. It can be seen that in the steady state, the total output and input signals cannot coincide. The comparison in Figure 4 fully reflects the advantage that QSE can eliminate the interaction between the harmonics components to be extracted.

## Grid-connected inverter current control based on quadrature sinewave extractor

As a harmonic extraction algorithm, QSE has a very wide application prospect. This paper only takes the current loop of a three-phase grid-connected inverter as an example to illustrate the application of QSE in harmonic elimination. The circuit topology of the three-phase grid-connected inverter is shown in Figure 5.

In Figure 5, $L$ is the filter inductance of the inverter, $C$ is the AC filter capacitor, and $C_d$ is the DC Bus support capacitor. $u_g$ is the grid voltage, $v$ is the output modulation voltage of the inverter, $i$ is the inverter output current, and $i_g$ is the grid current.

The three-phase system in Figure 5 has no neutral line, so only the two-phase current can be controlled independently. The coordinate transformation from three-phase (abc) to two-phase (αβ)can reduce the number of control loops. The transformation matrix is:

$$\mathbf{T}_{3/2} = \frac{2}{3}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \tag{10}$$

In the two-phase stationary coordinate system, the current loop based on QSE is shown in Figure 6. The frequency locked loop (FLL) is used to obtain the angular frequency and fundamental positive sequence component of grid voltage. The specific implementation of FLL can refer to references (Zhao et al., 2013) and (Kulkarni and John, 2013). Due to space limitations, this article will not repeat it.

Eq. 11 in Figure 6 is the given current calculation formula of the current loop, which can be written as:

$$\begin{bmatrix} i_\alpha^* \\ i_\beta^* \end{bmatrix} = \frac{1}{\left(u_{g\alpha}^+\right)^2 + \left(u_{g\beta}^+\right)^2} \begin{bmatrix} u_{g\alpha}^+ & -u_{g\beta}^+ \\ u_{g\beta}^+ & u_{g\alpha}^+ \end{bmatrix} \begin{bmatrix} P^* \\ Q^* \end{bmatrix} \tag{11}$$

Where, $i^*$ α, $i^*$β are the current given values in the two-phase stationary coordinate system, respectively; $p^*$, $Q^*$ are the given values of active power and reactive power to be sent to the power grid, respectively; $u$ + gα and $u$ + gβ are the positive sequence components of the grid voltage in the two-phase stationary coordinate system output by FLL respectively.

The current control based on QSE consists of three parts: current error proportional control, grid voltage feedforward element, and QSE. Proportional control is the basic control

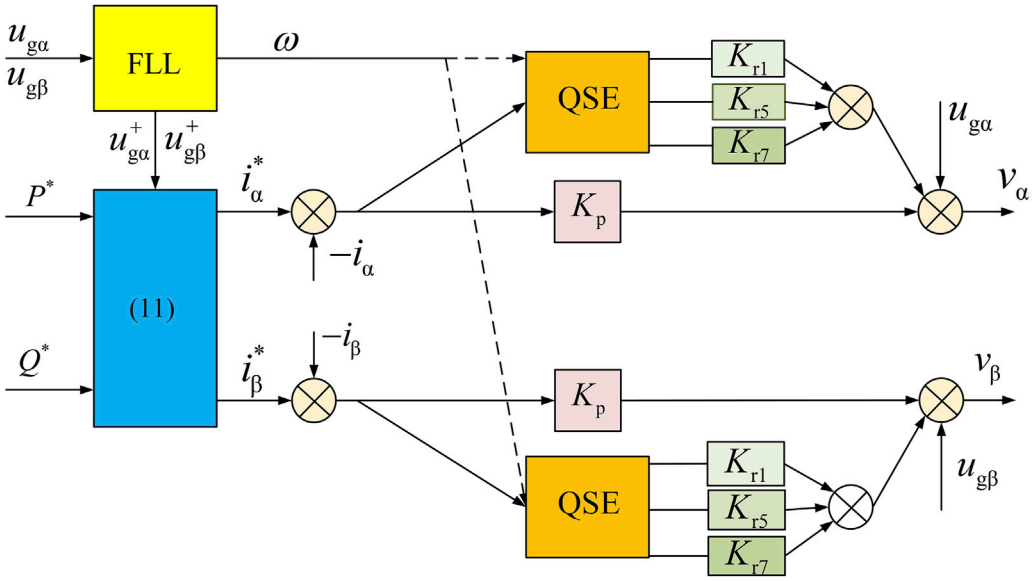



FIGURE 6
Current loop based on QSE.

that makes the system form negative feedback and closed-loop stability. The feedforward elements $u_{g\alpha}$ and $u_{g\beta}$ are to counteract dynamic effects of grid voltage disturbance.

This paper mainly introduces the specific design methods of QSE, which are implemented in the α and β control loop respectively. The input of QSE is the error between the given current and feedback current. Theoretically, QSE should extract the current harmonics of the grid current, that is the current of transformer secondary side. In this paper, considering that the current sensor is installed at the transformer primary side, the inverter output current is selected as the feedback current. The effect of the two current feedback modes is consistent when the grid voltage distortion is small enough to be negligible. On the contrary, if the influence of grid distortion is to be taken into account, QSE needs to extract the harmonic of the grid current.

The output of FLL is the angular frequency ω required for QSE design. If the grid voltage frequency fluctuation is very small, the ω can also be regarded as a constant, which reduces the complexity of the FLL and avoids the real-time trigonometric function operation of QSE. The fundamental, fifth and seventh cosine waves output by QSE are multiplied by the corresponding control coefficients $K_{r1}$, $K_{r5}$ and $K_{r7}$ respectively, and the results are used as one of the control voltages $v$ of the inverter.

The recommended values of the proportional control coefficient and each harmonic control coefficient are: $K_p = Lf_c$, $K_{r1} = K_{r5} = K_{r7} = 5K_p$, where $f_c$ is switching frequency of the inverter power switch devices, which needs to be corrected according to the transfer function of the system. This paper will discuss the selection of these control parameters in combination with specific experimental parameters in the following section.

TABLE 1 Parameters of experiments.

| Parameter | Value |
|---|---|
| AC filter inductance $L$ | 0.4 mH |
| AC filter capacitor $C$ | 100 μF |
| DC support capacitor $C_d$ | 1360 μF |
| sampling and control period $T$ | 100 μs |
| switching frequency $f_c$ | 5 kHz |
| dead time of power devices | 4 μs |
| proportional control coefficient $K_p$ | 2 |
| fundamental wave control coefficient $K_{r1}$ | 10 |
| 5th harmonic control coefficient $K_{r5}$ | 10 |
| 7th harmonic control coefficient $K_{r7}$ | 10 |
| QSE update factor $\rho$ | 0.001 |

# Experimental verifications

## Experimental platform and parameter design

In order to verify the adaptability of the proposed QSE algorithm to the grid frequency fluctuation, an experimental setup is built up in the laboratory. The main circuit topology of

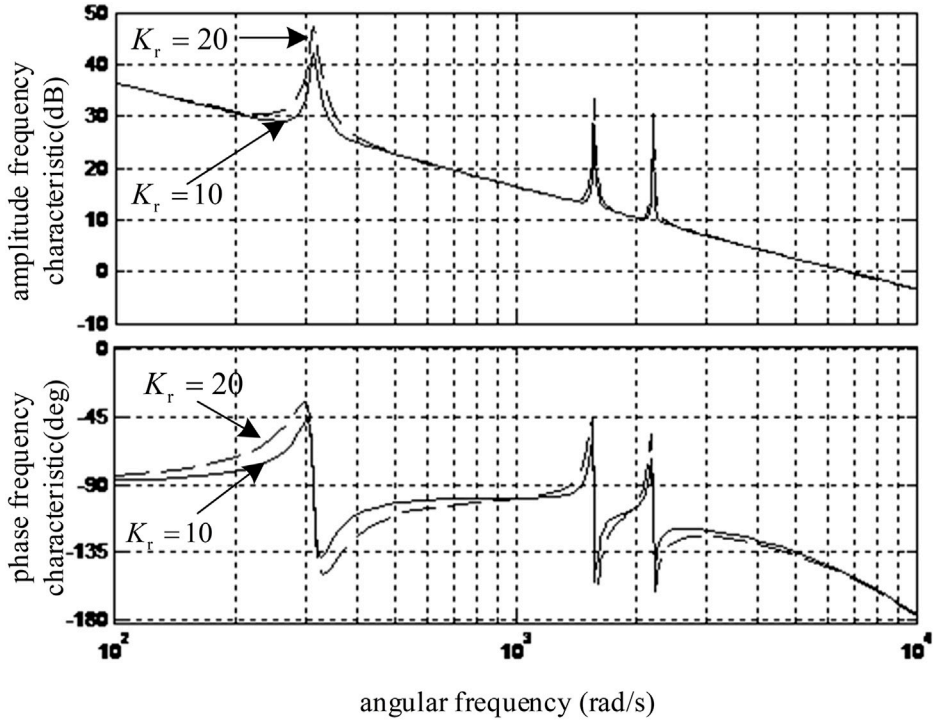



FIGURE 7
The open loop Bode graph of the system.

the experimental platform is shown in Figure 5. The inverter rated power is 50kW, the rated voltage (line voltage) is 400 V, and the rated current is 72A. The transformer capacity is 50 kV A, the transformation ratio is 270 V/400 V, the connection method is △/y, and the transformer leakage inductance is about 5% of the rated impedance. This paper adopts the inverter side current feedback control to avoid LCL resonance. The LC filter parameters are shown in Table 1. SVPWM modulation is adopted, and the control platform is based on DSP28335. The switching frequency is 5 kHz. In one SVPWM carrier cycle, current sampling and PWM duty cycle are both updated twice, that is, the control cycle $T$ = 100 μs? The update coefficient $\rho$ of QSE is 0.001, which is equivalent to the filtering bandwidth of 10 rad/s. Using FLL to obtain the angular frequency required by QSE can make it adapt to the frequency change in a larger range. Based on those parameters selected above, the experimental platform and control parameters are listed in Table 1.

Correcting the selected parameters in combination with the system transfer function is necessary. Because of structure consistency between the two control loops $\alpha$ and $\beta$, it is only necessary to design the parameters of one loop. For the sake of simplification, the system model only considers the inductance in the design of control parameters, and ignores the influence of filter capacitance and transformer leakage inductance. According to this simplified model, a certain stability margin is reserved in the design, which has lower complexity and facilitates the application in practice. Therefore, the system model can be described as that the modulation voltage $v$ outputted by the inverter acts on the inductor L and generates feedback current after a control period $T$ delay. The discrete transfer function of feedback current and modulation voltage can be obtained as follows:

$$G_m(z) = \frac{i(z)}{v(z)} = \frac{T}{L(z-1)}z^{-1} \tag{12}$$

$$G_o(z) = \left[K_p + K_r G_{QSE}(z)\right] G_m(z) \tag{13}$$

Where $K_r = K_{r1} = K_{r5} = K_{r7}$.

According to the parameters in Table 1, the open-loop Bode diagram is drawn as the solid line in Figure 7. If the harmonic control coefficient $K_r$ is doubled, the open-loop Bode diagram is shown as the dotted line in the figure. The condition for the stability of the system is that the phases of the phase frequency characteristics on the left of the crossing frequency are above −180 degrees. Comparing the dotted with the solid line, it can be seen that excessive harmonic control coefficient will reduce the stability margin of the system, and the variation of harmonic coefficient within a certain range will increase the amplitude-frequency characteristics at the harmonic frequency without reducing the stability of the whole system. It can also be seen from Bode diagram that if harmonic control is added, the higher the

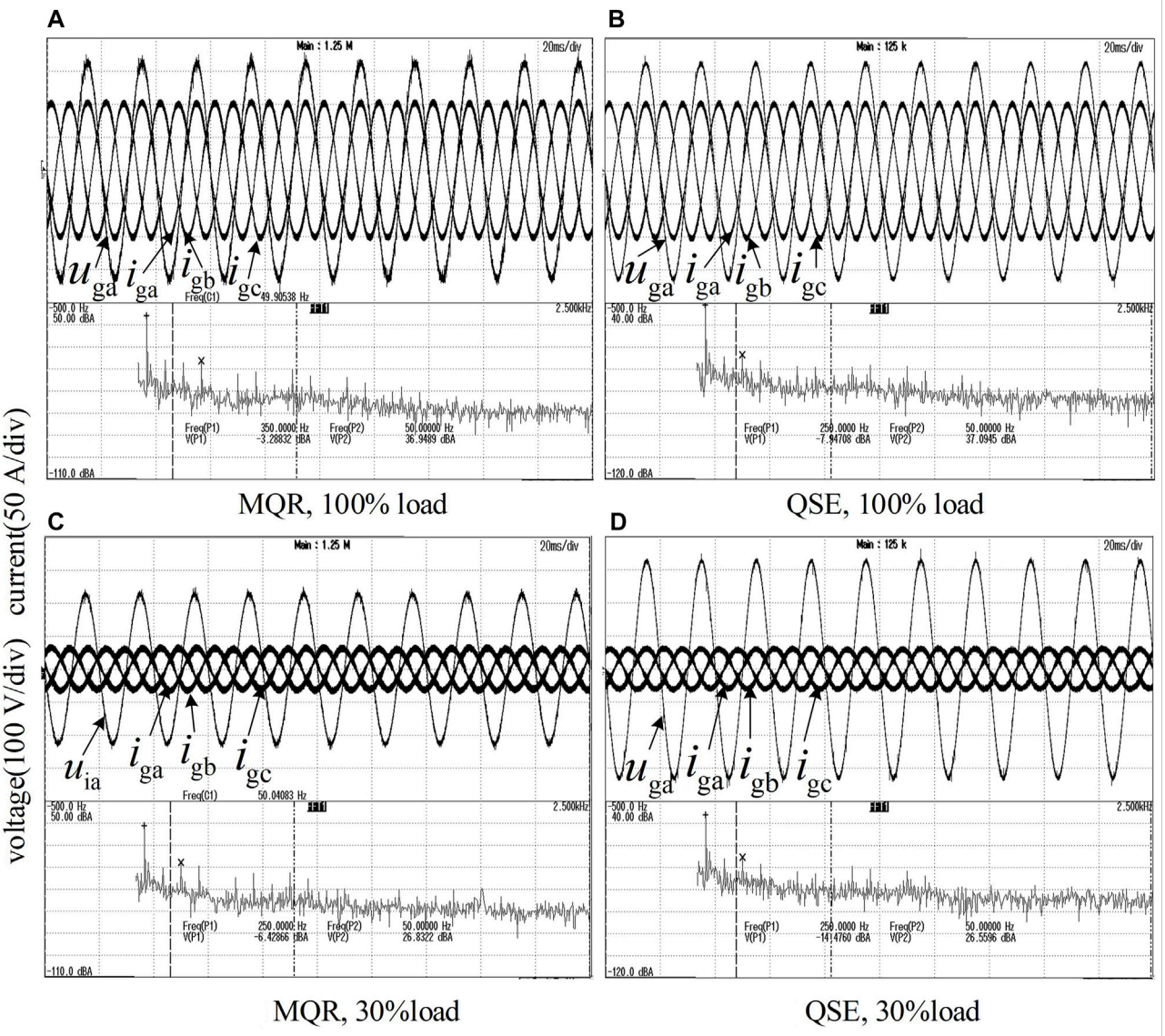



FIGURE 8
Experimental results at 50 Hz

frequency, the lower the phase margin. Therefore, it is not recommended to suppress high-frequency harmonic components.

## Steady-state experiment

With an adjustable frequency AC source, experiments are carried out at 50 and 55 Hz, respectively. The QSE-based and MQR-based algorithms are compared when the load ratio is 100 and 30%, and the experimental environment of the two control methods is consistent. The measured waveforms are shown in Figure 8 and Figure 9 $i_{ga}$, $i_{gb}$ and $i_{gc}$ in the figure are the three-phase grid current, respectively. $u_{ga}$ is the voltage on the high-voltage side of the transformer, that is, the grid side. $u_{ia}$ is the voltage on the low-voltage side of the transformer, that is, the inverter side.

Because the harmonic amplitude is very small compared with the fundamental amplitude, the oscilloscope's fast Fourier transform (FFT) function is enabled in the experiment, and each waveform shows the FFT result of A phase current. For example, in Figure 8A, the decibel at 350 Hz is −3.288 dB, which means that the amplitude of the seventh harmonic current $I_7$ meets 20l g ($I_7$) = −3.288, that is, the RMS value of $I_7$ is 0.68 A. The decibel of the fundamental current is 36.94 dB, corresponding to 70.3 A. Therefore, it can be calculated that the content of seventh harmonic is 0.97%. The oscilloscope can automatically locate and display the maximum value of each harmonic. According to each current waveform

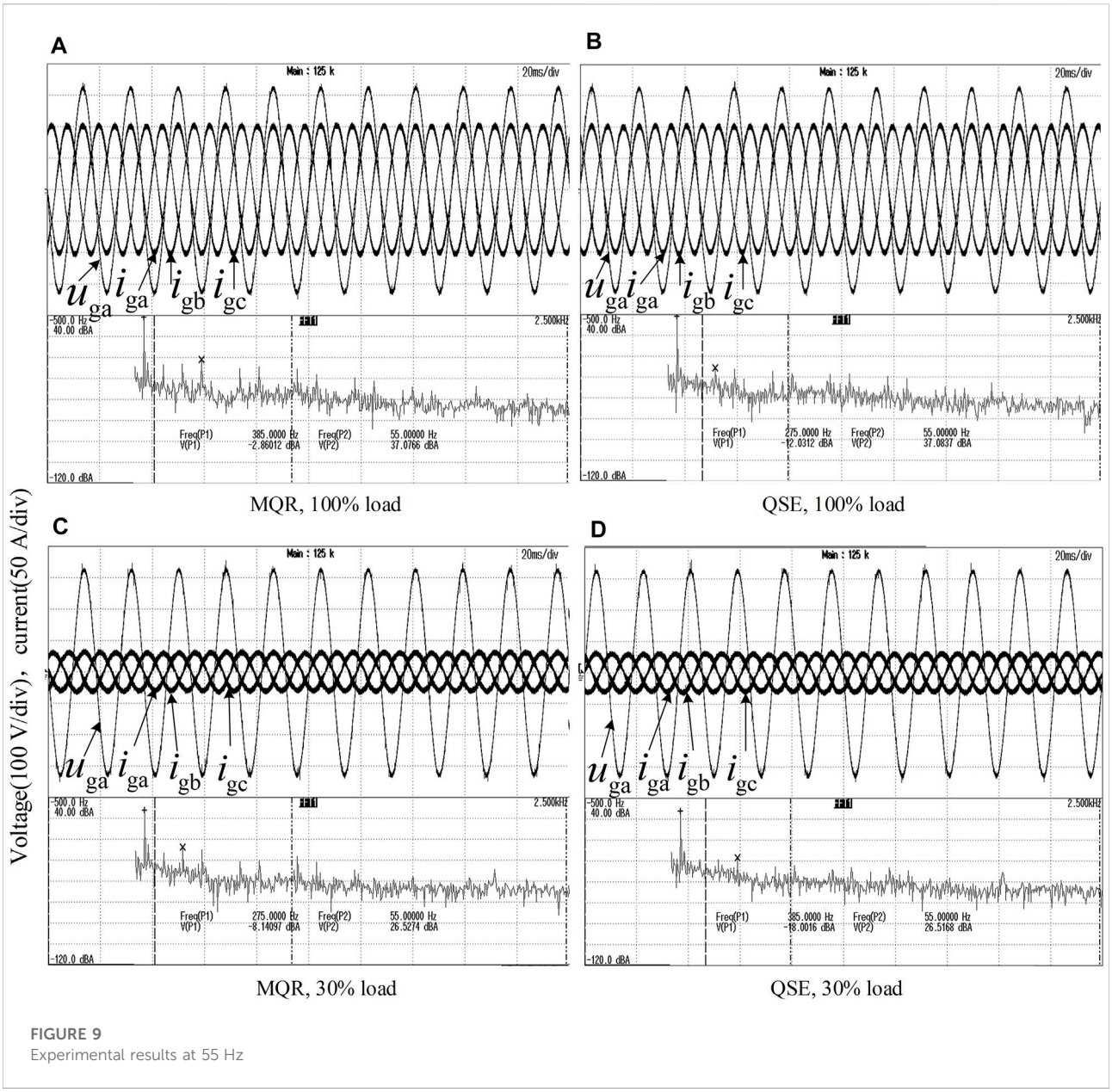



FIGURE 9
Experimental results at 55 Hz

TABLE 2 Maximum of harmonic content.

| Frequency (Hz) | Load ratio (%) | Maximum harmonic of MQR (%) | Maximum harmonic of QSE (%) |
|---|---|---|---|
| 50 | 100 | 0.97 | 0.56 |
| 50 | 30 | 2.10 | 0.85 |
| 55 | 100 | 1.01 | 0.35 |
| 55 | 30 | 1.85 | 0.58 |

TABLE 3 THD comparison.

| Frequency (Hz) | Load ratio (%) | Maximum harmonic of MQR (%) | Maximum harmonic of QSE (%) |
|---|---|---|---|
| 50 | 100 | 1.41 | 0.86 |
| 50 | 30 | 3.32 | 1.35 |
| 55 | 100 | 1.58 | 0.35 |
| 55 | 30 | 3.63 | 1.28 |

data, oscilloscope record the maximum harmonic value under each working condition, as shown in Table 2, and the recorded total harmonic distortion rate (THD) is shown in Table 3. It can be seen from Table 2 and Table 3 that compared with the MQR, because QSE extracts harmonics more accurately, the control based on QSE can further reduce the harmonic content.

## Conclusion

According to the above theoretical and experimental results, the following conclusions can be drawn:

1) QSE is designed directly in the discrete domain, avoiding the error caused by the discretization of the continuous domain transfer function so that it can extract the fundamental wave, main harmonic and their orthogonal components without static error.
2) QSE processes all components to be extracted simultaneously. The error used for updating is the difference between the combination of input signal and each component to be extracted, which can eliminate the interaction among those components.
3) QSE can be used in the current control loop of grid-connected inverter, which can greatly reduce current harmonics while maintaining system stability.

In the paper, the study is carried out for steady-state conditions. When considering the grid frequency mutation, we will use the grid simulator to complete the frequency mutation experiments for the QSE. The grid simulator can change frequency, amplitude and other parameters of the grid. Meanwhile, we will conduct QSE extraction experiments under different conditions such as high switching frequency, high-frequency noise and inter-harmonics in the subsequent study.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

## Author contributions

YX contributed for analysis of the work and wrote the first draft of the manuscript. ZF is the corresponding author. All authors contributed to manuscript revision, read and approved the submitted version.

## Funding

This work is partially supported by the Intelligent Electric Power Grid Key Laboratory of Sichuan Province (2022-IEPGKLSP-KFYB05)and Artificial Intelligence Key Laboratory of Sichuan Province (2021RZJ02). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

## Conflict of interest

Author YX and QH were employed by State Grid Sichuan Electric Power Company. Author RC was employed by Zonergy Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Bourguiba, I., Houari, A., Belloumi, H., and Kourda, F. (2016). "Control of single-phase grid connected photovoltaic inverter," in Proceeding International Conference on Control Engineering & Information Technology, Hammamet Tunisia, 16-18 December 2016 (IEEE), 1–6. doi:10.1109/CEIT.2016.7929116

Busarello, T. D. C., Pomilio, J. A., and Simoes, M. G. (2018). "Design procedure for a digital proportional-resonant current controller in a grid connected inverter," in Proceeding IEEE 4th Southern Power Electronics Conference (SPEC), Singapore, 10-13 December 2018 (IEEE), 1–8. doi:10.1109/SPEC.2018.8636052

Cárdenas, R., Juri, C., Peña, R., Wheeler, P., and Clare, J. (2012). The application of resonant controllers to four-leg matrix converters feeding unbalanced or nonlinear loads. *IEEE Trans. Power Electron.* 27 (3), 1120–1129. doi:10.1109/TPEL.2011.2128889

Castilla, M., Miret, J., Camacho, A., Matas, J., and Vicuna, L. G. d. (2013). Reduction of current harmonic distortion in three-phase grid-connected photovoltaic inverters via resonant current control. *IEEE Trans. Ind. Electron.* 60 (4), 1464–1472. doi:10.1109/TIE.2011.2167734

Chen, D., Zhang, J., and Qian, Z. (2013). An improved repetitive control scheme for grid-connected inverter with frequency-adaptive capability. *IEEE Trans. Ind. Electron.* 60 (2), 814–823. doi:10.1109/TIE.2012.2205364

Chen, Z. S., Rhee, S. H., and Liu, G. L. (2019). Empirical mode decomposition based on Fourier transform and band-pass filter. *Int. J. Nav. Archit. Ocean Eng.* 11 (2), 939–951. doi:10.1016/j.ijnaoe.2019.04.004

Choi, W., and Sarlioglu, B. (2019). "Comparative analysis on performance of power quality improvement of grid-connected inverters," in Proceeding IEEE Energy Conversion Congress and Exposition (ECCE), Baltimore MD USA, 29 September 2019 - 03 October 2019 (IEEE), 4281–4286. doi:10.1109/ECCE.2019.8912488

Fei, J., Liu, N., and Hua, M. (2019). "Adaptive backstepping neural control of active power filter using complementary sliding mode approach," in Proceeding International Russian Automation Conference, Sochi, Russia, 08-14 September 2019 (IEEE), 2090–2094. doi:10.1109/RUSAUTOCON.2019.8867729

Hu, J., Li, D., and Lin, Z. (2018). "Reduced design of PR controller in selected harmonic elimination APF based on magnetic flux compensation," in Proceeding 21st International Conference on Electrical Machines and Systems (ICEMS), 27 November 2018 (ACM Digital Library), 2090–2094. doi:10.23919/ICEMS.2018.8549000

Kulkarni, A., and John, V. (2013). Mitigation of lower order harmonics in a grid-connected single-phase PV inverter. *IEEE Trans. Power Electron.* 28 (11), 5024–5037. doi:10.1109/tpel.2013.2238557

Lee, N., and Cho, Y. (2020). "A PLL-based repetitive controller for a single-phase grid-connected NPC inverter," in Proceeding IEEE PELS Workshop on Emerging Technologies: Wireless Power Transfer (WoW), Seoul, Korea, 15-19 November 2020 (IEEE), 206–209. doi:10.1109/WoW47795.2020.9291334

McDonald, B., and Li, Y. (2018). "A novel LLC resonant controller with best-in-class transient performance and low standby power consumption," in Proceeding IEEE Applied Power Electronics Conference and Exposition (APEC), North Northlake Way, 1 March 2018 (Semantic Scholar), 489–493. doi:10.1109/APEC.2018.8341056

Qian, Q., Xie, S., Huang, L., Xu, J., Zhang, Z., and Zhang, B. (2017). Harmonic suppression and stability enhancement for parallel multiple grid-connected inverters based on passive inverter output impedance. *IEEE Trans. Ind. Electron.* 64 (9), 7587–7598. doi:10.1109/TIE.2017.2711526

Quan, X., Dou, X., Wu, Z., Hu, M., and Ni, C. (2016). The design for digital adaptive frequency-locked-loop based on discrete resonators. *Proc. CSEE* 2016, 99.doi:10.1109/ACCESS.2020.2963993

Song, T., Wang, P., Zhang, Y., Gao, F., Tang, Y., and Pholboon, S. (2020). Suppression method of current harmonic for three-phase PWM rectifier in EV charging system. *IEEE Trans. Veh. Technol.* 69 (9), 9634–9642. doi:10.1109/TVT.2020.3005173

Tuan, N., and Santoso, S. (2016). "Improving proportional-resonant controller for unbalanced voltage and frequency variation grid," in Proceeding IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Dallas, TX, USA, 03-05 May 2016 (IEEE), 1–5. doi:10.1109/TDC.2016.7520071

Wang, J. (2016). *Modern control theory*. Hebei, China: Journal of Hebei Radio & TV University.

Wang, X., Blaabjerg, F., and Loh, P. C. (2016). Grid-current-feedback active damping for LCL resonance in grid-connected voltage-source converters. *IEEE Trans. Power Electron.* 31 (1), 213–223. doi:10.1109/TPEL.2015.2411851

Yang, Y., Zhou, K., Wang, H., Blaabjerg, F., Wang, D., and Zhang, B. (2015). Frequency adaptive selective harmonic control for grid-connected inverters. *IEEE Trans. Power Electron.* 30 (7), 3912–3924. doi:10.1109/TPEL.2014.2344049

Zhao, X., Jin, X., Zhou, F., and Li, G. (2013). "Unbalanced control of grid-connected inverters based on proportion integraland reduced order resonant controllers," in Proceedings of the CSEE, North Northlake Way (Semantic Scholar).

Zhaoyang, Y., Da, L., Qingshan, Z., Chenyang, L., Lijun, Y., and Jianxia, L. (2017). "Full current harmonic detection method based on sinusoidal amplitude integrator," in Proceeding IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, 29, October - 01, November, 2017, 1172–1179. doi:10.1109/IECON.2017.8216200

## Appendix A: Proof of the nature of QSE

From Eq. A3 ~ (5), the state equation of QSE is:

$$\mathbf{x}(n) = \mathbf{G}\mathbf{x}(n-1) + \rho\mathbf{b}u(n) \tag{A1}$$

Where, both $\mathbf{x} = [x_{c1}, x_{s1}, x_{c5}, x_{s5}, x_{c7}, x_{s7}, \ldots]^T$ and $\mathbf{b} = [1, 0, 1, 0, 1, 0, \ldots]^T$ is a 2N dimensional column vector. Total state transition matrix $\mathbf{G} = \mathbf{G}_0 - \rho\mathbf{b}\mathbf{r}^T$ is a 2N × 2N dimensional square matrix, here $\mathbf{G}_0 = \begin{bmatrix} \mathbf{R}_1 & \mathbf{O}_2 & \mathbf{O}_2 & \\ \mathbf{O}_2 & \mathbf{R}_5 & \mathbf{O}_2 & \cdots \\ \mathbf{O}_2 & \mathbf{O}_2 & \mathbf{R}_7 & \\ & \vdots & & \ddots \end{bmatrix}$ is a square matrix composed of the diagonal array of each harmonic transfer matrixes, where $\mathbf{O}_2$ is a 2 × 2 dimensional zero matrix; $\mathbf{r} = [\cos(\omega T), -\sin(\omega T), \cos(5\omega T), -\sin(5\omega T), \cos(7\omega T), -\sin(7\omega T), \ldots]^T$ is the 2N dimensional column vector composed of the first row elements in the discrete transfer matrix of each harmonic. Writing $\boldsymbol{u} = [u_{c1}, u_{s1}, u_{c5}, u_{s5}, u_{c7}, u_{s7}, \ldots]^T$, the error vector $\boldsymbol{e} = \boldsymbol{u}\text{-}\boldsymbol{x}$, then the recurrence equation of the error vector $\mathbf{e}(n)$ can be obtained by combining (Eqs 1, 2, A1):

$$\boldsymbol{e}(n) = \mathbf{G}\mathbf{e}(n-1) \tag{A2}$$

It can be calculated that when $0<\rho < 2/N$:

$$\begin{aligned} &\mathbf{e}(n)^T\mathbf{e}(n) = \mathbf{e}(n-1)^T\mathbf{G}^T\mathbf{G}\mathbf{e}(n-1) \\ &= \mathbf{e}(n-1)^T\mathbf{e}(n-1) + \left(-2\rho + N\rho^2\right)\left(\mathbf{r}^T\mathbf{e}(n-1)\right)^2 \\ &\le \mathbf{e}(n-1)^T\mathbf{e}(n-1) \end{aligned} \tag{A3}$$

That is, the module of $\mathbf{e}(n)$ is less than or equal to the module of $\mathbf{e}(n\text{-}1)$ until $\mathbf{e}(n)$ converges to the zero vector.

The above derivation proves that when the update coefficient meets $0< \rho < 2/N$, QSE is convergent, and the cosine and sine waves of each frequency in the set Θ of input signals are accurately extracted in the steady state, that is, there is no phase offset and amplitude change for each sine wave extraction in the input signal. The above analysis proves the stability and zero-error of QSE.

In the set Θ, since the error $\mathbf{e}(n)$ does not contain various harmonics, the output of $k$th orthogonal sine wave $u_{ck}$ and $u_{sk}$ does not contain other frequency harmonics except the fundamental frequency component, which proves the decoupling of QSE.

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 26 September 2022
DOI 10.3389/fenrg.2022.992117



OPEN ACCESS

EDITED BY
Dou An,
MOE Key Laboratory for Intelligent Networks and Network Security, China

REVIEWED BY
Kuangyu Zheng,
Beihang University, China
Mingjian Jiang,
Qingdao University of Technology, China

*CORRESPONDENCE
Qian Lu,
luqian@qdu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 12 July 2022
ACCEPTED 25 July 2022
PUBLISHED 26 September 2022

CITATION
Zhang Z, Lu Q, Xu H, Xu G, Kong F and Yu Y (2022), Privacy-preserving deep learning for electricity consumer characteristics identification.
*Front. Energy Res.* 10:992117.
doi: 10.3389/fenrg.2022.992117



# Privacy-preserving deep learning for electricity consumer characteristics identification

Zhixiang Zhang[1], Qian Lu[1]*, Hansong Xu[2], Guobin Xu[3], Fanyu Kong[4] and You Yu[5]

[1]College of Computer Science and Technology, Qingdao University, Qingdao, China, [2]School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai, China, [3]Department of Computer Science, Morgan State University, Baltimore, MD, United States, [4]School of Software, Shandong University, Jinan, China, [5]Beijing Jingyi City Science and Industry Co., Ltd, Beijing, China

Deep learning models trained from smart meter data have proven to be effective in predicting socio-demographic characteristics of electricity consumers, which can help retailers provide personalized service to electricity customers. Traditionally, deep learning models are trained in a centralized manner to gather large amounts of data to ensure effectiveness and efficiency. However, gathering smart meter data in plaintext may raise privacy concerns since the data is privately owned by different retailers. This indicates an imminent need for privacy-preserving deep learning. This paper proposes several secure multi-party computation (MPC) protocols that enable deep learning training and inference for electricity consumer characteristics identification while keeping the retailer's raw data confidential. In our protocols, the retailers secret-share their raw data to three computational servers, which implement deep learning training and inference through lightweight replicated secret sharing techniques. We implement and benchmark multiple neural network models and optimization strategies. Comprehensive experiments are conducted on the Irish Commission for Energy Regulation (CER) dataset to verify that our MPC-based protocols have comparable performance.

KEYWORDS

**machine learning, secure multi-party computation, replicated secret sharing, smart meter, characteristics identification**

## 1 Introduction

Nowadays, smart meters are widely applied in residential households, which allow both customers and retailers to learn a large amount of accurate electricity consumption data (Wang et al., 2015; Mallapuram et al., 2017). In general, these fine-grained data are closely bound up with electricity consumption behavior of customers (Liang et al., 2019). Data analytics can extract the deeper insights from smart meter data, which can be used to enhance efficiency, save energy and improve smart grid systems. A vast amount of studies in machine learning algorithms have

been applied to smart meter data, including classification, regression, clustering, and sparse coding (Chicco, 2016). Applications include non-technical loss detection (Jokar et al., 2015; Júnior et al., 2016), price strategy (Chen et al., 2016; Li et al., 2016), demand response program enrollment (Wang et al., 2016a; Chen and Liu, 2017), load forecasting (Taieb et al., 2016) and the electricity consumer characteristics identification (Beckel et al., 2014).

Understanding the relationship between electricity consumer characteristics and smart meter data benefits most participants in the electricity market. Through the estimated electricity consumer characteristics, retailers can infer consumer consumption patterns and thus improve demand response programs, provide more personalized services and promote energy efficiency. There is no doubt that this will significantly enhance the competitiveness of retailers who are proficient in this capability. On the other hand, customers will enjoy better services and save energy due to the technological advances.

In the literature, several data analysis methods are applied to extract mathematical models that enable the identification of electricity consumer characteristics from smart meter data. Generally, these methods consist of three phases: feature extraction, feature selection and classification or regression. In order to infer the socio-demographic characteristics of electricity consumers from smart meter data, Beckel et al. (2013) propose a automatic classification system called CLASS, and the characteristic prediction accuracies of this system are higher than 70%. Viegas et al. (2016) estimate the characteristics of consumers by transparent fuzzy models. Wang et al. (2016b) utilize non-negative sparse coding to extract hidden consumption patterns and implement classification using support vector machine (SVM). Zhong and Tam (2014) achieve the classification of customers by discrete Fourier transform. The majority of these works rely on manually extracting features, while the manually extracted features may not effectively model the high variability and nonlinearity of individual load profiles. To solve this problem, the emerging deep learning techniques (LeCun et al., 2015) are applied to electricity consumer characteristics identification. Wang et al. (2018) leverage convolutional neural networks (CNN) to extract the highly nonlinear features from massive load profiles, and demonstrate the effectiveness by experiments on the Irish CER dataset. Lin et al. (2021) combine CNN and long short-term memory (LSTM) to predict the household characteristics.

Training an accurate deep learning model requires a large amount of available data. However, smart meter data is privately held by different retailers. In order to solve this problem, the previous works assume that there is a server having access to the raw data of retailers so that it can provide machine learning services in a centralized manner. Note that smart meter data and socio-demographic characteristics are sensitive information for consumers. Information leakage may lead to dissatisfaction from customers and public opinion attacks from competitors. As a result, retailers may not reveal raw data to the server due to privacy concerns and potential business risks. In addition, governments are also pushing for strict regulation of data privacy. For instance, the General Data Protection Regulation (GDPR) is already in effect in the European Union.

Secure multi-party computation (MPC) (Yao, 1986; Goldreich et al., 2019) provide a solution to these privacy-preserving issues, which is an important cryptographic technique that is commonly employed in previous studies for privacy-preserving machine learning, such as SecureML (Mohassel and Zhang, 2017), ABY3 (Mohassel and Rindal, 2018), SecureNN (Wagh et al., 2019) and Falcon (Wagh et al., 2021). Secure multi-party computation enables multiple parties $P_1, \ldots, P_n$ to collectively compute a function $f$ with their private input $x_1, \ldots, x_n$ and without revealing any information except the output. In this paper, we leverage replicated secret sharing techniques to construct MPC protocols for deep learning. The secret sharing based protocols require that all computing parties stay online during the execution process and have sufficient computing power. Consider that some retailers may not be able to meet both requirements, we assume that retailers distribute smart meter data in the form of secret shares to three servers, which provide the deep learning training and inference services. Such an outsourced computation pattern has proven to be very practical (Zhang et al., 2020; Zhang et al., 2021; Lu et al., 2022).

## 1.1 Our contributions

We summarize our contributions as follows.

- We design several MPC protocols that enable privacy-preserving deep learning training and inference for electricity consumer characteristics identification while keeping the retailer's raw data confidential. In our protocols, we implement two deep neural network models and multiple optimization strategies.
- To relieve the burden on retailers, we propose a system architecture that allows retailers to not have to engage in online computation. Retailers only need to upload secret shares of their smart meter data.
- To demonstrate the practicality, we implement our protocols based on the MP-SPDZ framework (Keller, 2020) and conduct a series of experiments on the Irish Commission for Energy Regulation (CER) dataset (Commission for Energy Regulation, 2012).

## 1.2 MPC frameworks

In recent years, MPC has evolved from theoretical research to provide practical privacy-preserving protocols for many

TABLE 1 Notations used in this paper.

| Symbols | Descriptions |
|---|---|
| $\alpha$ | The learning rate |
| $B$ | The mini-batch size |
| $N$ | The number of retailers |
| $D_n$ | The smart meter dataset of the $n$th retailer |
| $D$ | The data set consisting of all $D_n$ |
| $l$ | The bit length of the arithmetic circuit |
| x | Lowercase bold letter denotes vector |
| $x^{(i)}$ | The $i$th element of x |
| $[x]_M$ | A arithmetic secret sharing of $x \in M$ |
| $[x]_2$ | A binary secret sharing of $\in \mathbb{Z}_2$ |
| $[x]_B$ | A vector of $l$ binary secret sharing which encodes $x \in \mathbb{Z}_{2^l}$ |

machine learning tasks, such as training and evaluation of linear regression, logistic regression and neural networks (Mohassel and Zhang, 2017; Mohassel and Rindal, 2018; Wagh et al., 2019, 2021). Here, we give a brief overview on works related to our protocols. ABY3(Mohassel and Rindal, 2018) follow the same blueprint as ABY(Demmler et al., 2015) by mixing replicated secret sharing with garbled circuits. Eerikson et al. (2019) introduce an optimization that can leverage pseudo-random generator (PRG) to reduce the communication costs of input sharing in replicated secret sharing. Keller and Sun, (2021) implement purely training of neural network in MPC with 99% accuracy. Furthermore, Keller and Sun, (2021) discuss in detail how to implement various building block of secure computation with replicated secret sharing.

### 1.3 Road map

The rest of the paper is organized as follows: we present the problem statement in **Section 2**. In **Section 3**, we introduce basic three-party protocol. We introduce in detail how to construct the required secure computation building blocks in **Section 4**. In **Section 5**, we discuss the building blocks for deep learning. We report the experimental results in **Section 6**. Finally, we conclude this paper in **Section 7**.

## 2 Problem statement

### 2.1 Notation

We summarize the notations used in this paper in **Table 1**.

### 2.2 Privacy-preserving deep learning

Deep learning is broadly applied in many domains, such as language translation and image classification, often leading to breakthroughs in each domain. The model used for deep learning is a deep neural network, which consists of linear layers and nonlinear layers. Linear layers, including fully connected layers and convolutional layers, can be reduced to arithmetic operations as multiplications and vector dot products. While the activation functions required for the nonlinear layers, such as ReLU functions and max-pooling functions, can be efficiently implemented on binary circuits. Privacy-preserving deep learning is very challenging due to it involves the "mixed" evaluation of arithmetic and binary circuits. The previous works have proposed two main cryptographic approaches that can implement privacy-preserving deep learning: homomorphic encryption (Paillier, 1999; Gentry, 2009) and MPC. The (fully) homomorphic encryption is mainly used for computing linear layers in two-party (client-server model) secure neural network inference. The nonlinear layers in two-party secure neural network inference are usually implemented *via* oblivious transfer (OT) (Asharov et al., 2013) or garbled circuit (GC) (Yao, 1986), which are important cryptographic primitives of MPC. The studies on secure neural network training mainly focus on two types of three-party MPC protocols (Mohassel and Rindal, 2018; Wagh et al., 2019, 2021) for efficiency. The first have the two computing parties performing the secure neural network training by two-party additive secret sharing and the remaining party generating the materials required by the two computing parties The second utilize the three-party replicated secret sharing to accomplish the secure neural network training. In both scenarios, the participants who own the data are not directly involved in the computation, but instead distribute the raw data to the three computing parties in the form of secret shares. In this paper, we investigate how to use the three-party replicated secret sharing technique to construct privacy-preserving deep learning protocols.

### 2.3 System architecture

This paper targets to privacy-preserving deep learning (PPDL) for electricity consumer characteristics identification. Our system architecture is shown in **Figure 1**. At the core, there are two types of entities: the retailer and the server. Retailers are the owner of smart meter data and wish to accomplish the training of the deep neural network models. Since deep learning is a data-driven analytics approach, different retailers wish to work together to ensure the effectiveness of the models. Let N be the number of retailers. The smart meter dataset of the $n$th retailer is denoted by $D_n$ ($n \in \{1, 2, \ldots, N\}$) and the dataset consisting of all $D_n$ is denoted by $D$. Traditionally, this can be achieved through the Machine-Learning-as-a-Service (MLaaS) architecture, which leverages the power of the computational servers. Many major companies such as Amazon, Google, or Microsoft all provide computational services. Since smart meter data is privacy sensitive, retailers want to ensure confidentiality

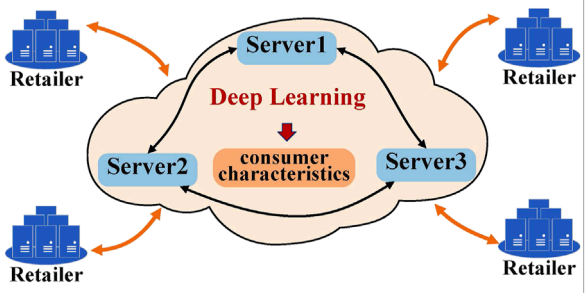



FIGURE 1
System architecture.

while enjoying the benefits of the computational servers. As a result, retailers are reluctant to supply the smart meter data in plain text, but the ciphertexts of the smart meter data are supplied instead. In addition, it may be impractical to keep all retailers online at the same time to perform the active data interactions required by the MPC protocols. Hence, our system should allow retailers to stay offline after uploading the shares of smart meter data.

In our system, three servers $S_0$, $S_1$, $S_2$ play the role of computing parties to ease the burden on retailers. We assume that government or other social deterrents are sufficient to make the servers strictly execute the protocol and not collude with each other. Similar to the popular security designs in recent years (Mohassel and Rindal, 2018; Wagh et al., 2021), servers use the lightweight replicated secret sharing techniques to collaboratively accomplish the secure deep neural network training. The workflow of our system is as follows. Before deep neural network training, retailers encrypt their smart meter data by splitting it into three secret shares, which can form three unique pairs. Retailers distribute each pair to a server. With shared smart meter data, servers perform the training of deep learning models by invoking various secure computation building blocks, such as dot product, secure comparison and oblivious selection. The trained model parameters are stored on the server in the form of secret shares. Retailers or other users can query the system or download the model parameters directly.

## 2.4 Security model

Security definition. Our protocol works under a three-party honest-majority setting in which an adversary $\mathcal{A}$ can corrupt at most one party. We assume $\mathcal{A}$ takes static corruption strategy; it decides which party to corrupt before executing the protocol. The adversary is semi-honest; it faithfully follows the protocol and attempts to learn sensitive information from protocol execution.

We use the simulation-based security definition (Canetti, 2001; Goldreich, 2009; Araki et al., 2016) for three-party computation (3PC). A 3PC protocol $\Pi$ computes a functionality $f:(\{0,1\}^*)^3 \to (\{0,1\}^*)^3$. For an input tuple $\vec{x} = (x_0, x_1, x_2)$ where party $P_i$ provides $x_i$, the output is $f(\vec{x}) = (f_0(\vec{x}), f_1(\vec{x}), f_2(\vec{x}))$, and $P_i$ receives $f_i(\vec{x})$. Intuitively, $\Pi$ is secure if for any corrupted party $P_i$, there exists a probabilistic polynomial-time simulator $Sim_i$ who can generate a view that is indistinguishable from the one from real-world execution. Formally, let $\text{View}_i^{\Pi}(1^{\lambda}, \vec{x})$ be the view of $P_i$, security is defined as follows:

Definition 1. A protocol $\Pi$ securely computes a deterministic functionality $f$ in the presence of static semi-honest adversaries if there exist a probabilistic polynomial time simulator $\text{Sim}_i$ ($i \in \{0,1,2\}$) generating computationally indistinguishable view:

$$\left\{\text{Sim}_i\left(1^{\lambda}, x_i, f_i(\vec{x})\right)\right\} \overset{c}{\equiv} \left\{\text{View}_i^{\Pi}\left(1^{\lambda}, \vec{x}\right)\right\},$$

by only taking $P_i$'s input $x_i$, output $f_i(\vec{x})$, and other allowed public information (e.g., bitlength of inputs, the size of each $D_n$).

Three-party decision tree evaluation. Our PPDL protocols are special cases of three-party secure computation. In our protocols, there are three servers $S_0$, $S_1$ and $S_2$ hold the secret shares of $D$. The three servers perform privacy-preserving deep learning (training and inference) using secret-sharing. Our protocols allow the servers to learn some public information during PPDL protocols. In particular, we allow $S_i$ to learn some public information about each $D$ (*e.g*, the number of retailers, the size of each $D_n$), for which we denote as $\mathcal{L}_i(D)$. Formally, let $\mathcal{F}_{\text{DL}}(D) \to (\mathcal{F}_0(D), \mathcal{F}_1(D), \mathcal{F}_2(D))$ be the ideal deep learning functionality, where $\mathcal{F}_i(D)$ contains the shares of model parameters and public information $\mathcal{L}_i(D)$. A PPDL protocol $\Pi$ securely computes $\mathcal{F}_{\text{DL}}$ if there exist a PPT simulator $Sim_i$ such that for any $D$:

$$\left\{\text{Sim}_i\left(1^{\lambda}, x_i, \mathcal{F}_i(D)\right)\right\} \overset{c}{\equiv} \left\{\text{View}_i^{\Pi}\left(1^{\lambda}, \mathbf{x}\right)\right\}.$$

Security in hybrid model. In this paper, we rely on necessary secure computation protocols (*e.g.*, random bit generation, domain conversion, secure comparison and dot product) to design our PPDL protocols. Since security of these secure components has already been proven secure, we will directly use their corresponding ideal functionalities in our design. This approach is known as the hybrid model (Canetti, 2001; Hazay and Lindell, 2010) and is commonly used in existing works.

# 3 Three-party MPC protocol

In this paper, we use replicated secret sharing techniques (Mohassel and Rindal, 2018; Eerikson et al., 2019; Keller and Sun, 2021) to construct MPC protocols for deep learning, which can be traced back to Benaloh and Leichter, (1988). We begin by introducing the basic secret sharing framework and then move on to high-level building blocks.

## 3.1 Secret sharing scheme

Replicated secret sharing is a variant of additive secret sharing by appending redundant shares. As we mentioned, three servers $S_0, S_1$ and $S_2$ play the role of computing parties to perform three-party MPC protocols. We denote the next and previous servers of $S_i$ as $S_{i-1}$, $S_{i+1}$, *i.e.*, the indices are computed modulo three. The secret value $x$ is represented as the sum of the three secret shares: $x = x_0 + x_1 + x_2 \pmod M$, where $x_{i-1}$ and $x_{i+1}$ are sent to $S_i$. Such a 2-out-of-3 replicated secret sharing is denoted as $[x]_M$. In this paper, we set $M = 2^l$ to utilize the properties of the ring. In general, the computation with $M = 2$ is known as binary circuits, while computing with larger moduli is called arithmetic circuits. In addition, if $M$ is clear from context, we will omit this from the sharing notation.

## 3.2 Generating randomness

Throughout this paper, we require to generate randomness using pseudo-random generators (PRG). In the initialization phase, $S_i$ and $S_{i+1}$ share a key of PRG so that they can generate the same random number $r_{i,i+1}$. That is, each server will hold two PRG keys during the protocol. To generate a 3-out-of-3 additive secret sharing of zero, each server $S_i$ compute $r_{i-1,i}$ and $r_{i,i+1}$ and set $r_{i,i+1} - r_{i-1,i}$ as its share. To generate a 2-out-of-3 replicated secret sharing of a random number, each server $S_i$ compute $r_{i-1,i}$ and $r_{i,i+1}$ and set $(r_{i-1,i}, r_{i,i+1})$ as its share.

## 3.3 Input and open secret values

There are two types of inputting parties in our protocols. The first type of inputting parties sample and distribute secret shares from external to the servers, *e,g,* retailers. The second are computing parties, *i.e.* Servers, who need to share secret values for some building blocks. In our protocols, servers sample and distribute secret shares based on the method of Eerikson et al. (2019). If $S_i$ wish to share a secret value $x$, then $x_i$ is set to zero and $x_{i-1}$ is generated by $S_i$ and $S_{i+1}$ using PRG. With $x_i$ and $x_{i-1}$, $S_i$ can compute $x_{i+1}$ and send it to $S_{i-1}$.

Open secrets also have two types of situations. In order to open a secret value $x$ to retailers or other entities, each server sends one share to the receiver, who can reconstruct $x$ by computing $x = x_0 + x_1 + x_2 \pmod M$. To open a secret value $x$ to all servers, $S_i$ send $x_{i+1}$ to $S_{i+1}$. We emphasize that the values revealed to the servers are independent of the dataset or model parameters. Hence, it does not leak sensitive information.

### 3.3.1 Linear operations

The additive property of the secret sharing scheme implies that linear operations can be computed locally. Let $c$ be a public constant and $[x]$ $[y]$ be shared values. The addition of $[x]$ and $[y]$ can be computed as $[x] + [y] = [x + y] \coloneqq (x_1 + y_1, x_2 + y_2, x_3 + y_3)$. The same applies to subtraction. In addition, we define $[x \pm c]$ as $(x_1 \pm c, x_2, x_3)$ to add or subtract a shared value with a public constant. As for the scalar multiplication $c[x]$, we define as $c[x] = [cx] \coloneqq (cx_0, cx_1, cx_2)$.

### 3.3.2 Multiplication

The multiplication of two secret values $[x]$ and $[y]$ is shown below:

$$\begin{aligned} x \cdot y &= (x_0 + x_1 + x_2) \cdot (y_0 + y_1 + y_2) \\ &= (x_0 y_0 + x_0 y_1 + x_1 y_0) + (x_1 y_1 + x_1 y_2 + x_2 y_1) \\ &\quad + (x_2 y_2 + x_2 y_0 + x_0 y_2) \end{aligned} \tag{1}$$

We can obsever that each server can compute one summand using its own share. Let $z = xy = z_0 + z_1 + z_2$ and $z_0 = x_1 y_1 + x_1 y_2 + x_2 y_1$, $z_1 = x_2 y_2 + x_2 y_0 + x_0 y_2$, $z_2 = x_0 y_0 + x_0 y_1 + x_1 y_0$, where $z_i$ can be locally computed by $S_i$. Then, servers perform the operation called *re-sharing* to hold two shares as defined. To this end, each server $S_i$ need to sends $z_i$ to another server. However, since $z_0$, $z_1$ and $z_2$ are not entirely randomized, servers need to generate a 3-out-of-3 sharing of zero to mask them. Let $(\alpha_0, \alpha_1, \alpha_2)$ be a 3-out-of-3 sharing of zero and $S_i$ hold $\alpha_i$. $S_i$ computes $z_i' = z_i + \alpha_i$ and sends $z_i'$ to $P_{i+1}$ to generate 2-out-of-3 sharing $((xy)_{i-1}, (xy)_{i+1}) = (z_i', z_{i-1}')$.

# 4 Building blocks for secure computation

In this section, we will describe the building blocks for secure computation in the RSS setting. To the best of our knowledge, we are the first to apply these techniques to privacy-preserving deep learning for electricity consumer characteristics identification.

## 4.1 Multiplication of fixed-point values

Since computing with floating-point number is extremely expensive (Aliasgari et al., 2013), decimals are usually represented as fixed-point numbers in the MPC protocols, *e.g.* Catrina and Saxena, (2010). A decimal $x$ is represented as $x = \lfloor x \cdot 2^p \rfloor$, where $p$ is a positive integer used to specify the precision. For the case of addition or subtraction, the precision of the results will not change. However, the multiplication of two fixed-point numbers doubles the precision $(x \cdot 2^p) \cdot (y \cdot 2^p) = xy \cdot 2^{2p}$, which causes the precision to accumulate until it overflows $M$. To address this problem, the previous works have proposed a method known as truncation. There are three ways to implement truncation.

- The easiest way is to multiply the result of each multiplication by $2^{-p}$. However, this method can lead to errors with a certain probability and the absolute value of the errors is 1. For more details, we refer the readers to Mohassel and Zhang, (2017).
- The most effective way to reduce the negative impact of errors is the nearest truncation, which requires to shift the result by $p$ bits after adding $2^{p-1}$ to the integer representation The nearest truncation can be instantiated by mixed-circuit computation (Dalskov et al., 2021).
- Catrina and Saxena, (2010) present a solution called probabilistic truncation that can effectively balance cost and accuracy. This method utilizes the uniformly selected random numbers. Let $x$ be the truncated secret value and $r$ be a random number. Servers first compute $[x+r] = [x] + [r]$ and then perform truncation. Finally, servers remove the mask $r$ to obtain $[x]$. For instance, if $x = 0.6$, then $x$ will round to 1 with 60% probability. In this work, we mainly use probabilistic truncation.

## 4.2 Dot product

The dot product is the core building block of the linear layer. Let $\mathbf{x}$ and $\mathbf{y}$ be two $m$-dimensional vectors. The dot product of. $\mathbf{x}$ and $\mathbf{y}$ is shown below:

$$\begin{aligned}\mathrm{x}\cdot\mathrm{y} &= \sum_{i=1}^{m} x^{(i)}\cdot y^{(i)} = \sum_{i=1}^{m}\left(x_0^{(i)}+x_1^{(i)}+x_2^{(i)}\right)\cdot\left(y_0^{(i)}+x_1^{(i)}+y_2^{(i)}\right)\\ &= \sum_{i=1}^{m}\left(x_0^{(i)}y_0^{(i)}+x_0^{(i)}y_1^{(i)}+x_1^{(i)}y_0^{(i)}\right)\\ &\quad+\sum_{i=1}^{m}\left(x_1^{(i)}y_1^{(i)}+x_1^{(i)}y_2^{(i)}+x_2^{(i)}y_1^{(i)}\right)\\ &\quad+\sum_{i=1}^{m}\left(x_2^{(i)}y_2^{(i)}+x_2^{(i)}y_0^{(i)}+x_0^{(i)}y_2^{(i)}\right)\end{aligned}\tag{2}$$

Intuitively, the dot product of x and y should be reduced to $m$ parallel multiplications and one summation, which requires $m$ re-sharing operations and $m$ truncations. However, it is feasible to reduce the usage of truncation and resharing by delaying them to after the summation. Each server can first compute one of the three sums in the last term locally. Then, all servers perform re-sharing and truncation on the sums. In this way, the communication cost of one dot product is the same as a single multiplication.

## 4.3 Domain conversion

Recall that we use two different versions of replicated secret sharing techniques. The first is the arithmetic sharing with $M = 2^l$, which is more suitable for arithmetic operations such as addition, multiplication and dot product. The second is the binary sharing with $M = 2$, which is more suitable for binary operations and non-linear operations that need to access the individual bits directly, such as comparison. For situations that require both versions, the ideal solution is to construct efficient building blocks that allow the secret sharing of two versions to convert to each other. Especially in deep learning, domain conversion is the bridge between linear layers and nonlinear layers. For brevity, we use Bit2A and A2B to represent the conversions in two directions, respectively.

### 4.3.1 Random bit generation

An efficient solution of Bit2A is to leverage XOR operation, which can be defined as the function $f(x,y) = x + y - 2\cdot x\cdot y$ for $x,y \in \mathbb{F}_p$. As we can see, an XOR operation require to compute one multiplication of secret values, while using the pre-processed random bits can reduce the XOR operation to linear operations. Note that these random bits need to be secret-shared in the arithmetic circuit and no server is aware of their true value. In order to obtain such random bits, two servers sample and share a random bit respectively. Let $r_0$ and $r_1$ be the sample random bits. Then, all servers jointly compute $[r]$ by $[r] = [r_0] + [r_1] - 2\cdot[r_0]\cdot[r_1]$.

### 4.3.2 Bit2A

With the arithmetic sharing of random bits, Bit2A can be implemented by the idea of "daBits" (Rotaru and Wood, 2019). A daBit is a random bit that is shared in both arithmetic and Boolean circuits. Let $r$ be a random bit and $[r]$ $[r]_2$ be available. Servers can mask a secret bit $b$ with $r$ and open $b \oplus r$ without leaking any information. Then, servers can remove the mask $r$ in arithmetic circuits to obtain $[b]$.

To construct a daBit, servers need to invoke one random bit generation to obtain $[r]$. On the other hand, as introduced by Escudero et al. (2020) $[r]_2$ can be locally generated for powers of two are compatible. Observe that

$$\begin{aligned}\sum_{i=0}^{i=2}(r_i \bmod 2) \bmod 2 &= \left(\sum_{i=0}^{i=2} r_i \bmod 2^l\right) \bmod 2\\ &= r \bmod 2\end{aligned}\tag{3}$$

Hence, servers can locally generate $[r]_2$ by extracting the least significant bit of $[r]$.

### 4.3.3 A2B

A2B can be considered as a special case of bit decomposition. In this paper, we adopt the method proposed by Araki et al. (2016); Mohassel and Rindal, (2018) to perform A2B. Recall that the arithmetic sharing of $x$ is $[x] \coloneqq (x_0, x_1, x_2)$. We use $[x]_B$ to denote a vector of $l$ binary secret sharing which encodes $x \in \mathbb{Z}_{2^l}$. We can observe that $[x_0]_B \coloneqq (x_0, 0, 0)$, $[x_1]_B \coloneqq (0, x_1, 0)$ and $[x_2]_B \coloneqq (0, 0, x_2)$ are valid but not random binary sharings.

Then, servers can compute $[x]_B = [x_0]_B + [x_1]_B + [x_2]_B$ in the binary addition circuit.

## 4.4 Secure comparison

The comparison is essential for the implementation of a lot of activation functions, such as ReLU functions, max-pooling functions and approximate sigmoid functions. The comparison is defined as $b = x \overset{?}{<} y$ for $x, y \in \mathbb{Z}_{2^l}$. As introduced by Mohassel and Rindal, (2018); Keller and Sun, (2021), the most significant bit (MSB) denotes the sign of a ring element, which implies that the comparison can be reduced to the MSB extraction of the difference between the two operands. Given the secret sharings of $x$ and $y$, servers first compute the difference $a$ locally by $[a] = [y] - [x]$ and then convert the arithmetic sharing of $a$ to its binary sharing. It remains to extract the MSB of $a$ and convert it to the arithmetic sharing by invoking Bit2A for subsequent operations.

## 4.5 Oblivious selection

Oblivious Selection is an essential building block for segmentation functions. It can avoid participants learning which branch is selected. Oblivious selection can be reduced to a polynomial. Taking the 1-out-of-2 oblivious selection as an example, let $x$ and $y$ represent the branches and $b \in \{0,1\}$ represent the condition. The oblivious selection can be done by $x + b \cdot (y - x)$. And so on, the oblivious selection with more branches can be implemented by polynomials with higher orders.

## 4.6 Division

Since arithmetic operations are performed on the ring $\mathbb{Z}_{2^l}$, we cannot compute the division directly. There are two main ways to solve this problem: sequential comparison and numerical methods. The specific method we use is the numerical method by Catrina and Saxena, (2010), which instantiates the algorithm of Goldschmidt, (1964) in MPC. This method iteratively approximates the results by multiplication. Therefore, the error of the results mainly depends on the usage of iterations.

## 4.7 Logarithm and exponentiation

Similar to division, logarithm and exponentiation are implemented by numerical methods (Aly and Smart, 2019). We use $\log_a x$ and $x^y$ to represent the instances of logarithm and exponentiation, respectively, where $x$ and $y$ are two secret values and $a$ is an arbitrary public base.

$\mathrm{Log}_a x$ can be reduced to $\log_a 2 \cdot \log_2 x$. Then, $x$ is represent as $x = b \cdot 2^c$ such that $\log_2 x$ can be computed by $\log_2 x = \log_2 b + c$, where $b \in [\,0.5, 1\,)$ and $c \in \mathbb{Z}$ $\log_2 b$ can be computed by Padé approximation (Hart, 1978), which is achieved by a division of polynomials.

$X^y$ can be reduced to $x^y = 2^{y \log_2 x}$. Computation exponentiation with base two can be done by $2^a = 2^{\lfloor a \rfloor} \cdot 2^{a - \lfloor a \rfloor}$, where the former is achieved by polynomial approximation and the latter by bit decomposition and multiplication. Let $b = \sum_{k \geq 0} b_k 2^k$ is an integer with $b_k \in \{0,1\}$, the integer power of 2 can be computed as follows

$$2^b = 2^{b = \sum_{k \geq 0} b_k 2^k} = \prod_{k \geq 0} 2^{b_k 2^k} = \prod_{k \geq 0} \left(1 + b_k \cdot \left(2^{2^k - 1}\right)\right) \tag{4}$$

The above three operations are approximated by numerical methods, the accuracies of which depend on the number of iterations or the truncation method used for multiplication.

# 5 Building blocks for deep learning

In this section, we will introduce how to construct the building blocks for deep learning.

## 5.1 Fully connected layers

A fully connected layer is also called a dense layer, which is a linear transformation parameterized by the weight $W$ and the bias $b$. Let $x$ be the input to a fully connected layer. The output $u$ can be computed by $u = W \cdot x + b$. The matrix multiplications are implemented by dot products. To save communication rounds, all dot products of a matrix multiplication are computed in parallel.

## 5.2 Convolution layers

Convolutional layers are the main layers for feature extraction. Each convolutional layer has a certain number of kernels (also known as filters). These kernels are represented as vectors so that the convolution can be performed using only dot products. Furthermore, these dot products are also computed in parallel to reduce communication rounds.

## 5.3 ReLU

ReLU functions (Nair and Hinton, 2010) enhance the nonlinear relationship between the layers of the neural network, which can be mathematically defined as follows

$$\mathrm{ReLU}(x) \coloneqq \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

ReLU functions can be reduced to one comparison and one oblivious selection. The comparison results of forward propagation are reused in backward propagation to reduce the invocations of comparison.

## 5.4 Softmax

The objective of softmax functions is to present the results of multi-classification in the form of probabilities. The probability of the $i$th class can be computed as follows, and the classification result is the class corresponding to the maximum probability.

$$\text{Softmax}(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}. \tag{6}$$

## 5.5 Sigmoid

Sigmoid is one of the most widely used activation functions. However, since sigmoid requires costly exponential operations, previous works usually use segmentation functions to approximate it. There are two method to approximately compute sigmoid: 3-piece approximation (Mohassel and Zhang, 2017) and 5-piece approximation (Hong et al., 2020). The two piecewise functions are shown below:

$$3-\text{piece Sigmoid}(x) \coloneqq \begin{cases} 0 & x < -0.5 \\ x + 0.5 & -0.5 \le x < 0.5 \\ 1 & x \ge 0.5 \end{cases} \tag{7}$$

$$5-\text{piece Sigmoid}(x) \coloneqq \begin{cases} 10^{-4} & x < -5 \\ 0.02776 \cdot x + 0.145 & -5 < x \le -2.5 \\ 0.17 \cdot x + 0.5 & -2.5 < x \le 2.5 \\ 0.02776 \cdot x + 0.5 & 2.5 < x \le 5 \\ 1 - 10^{-4} & x \ge 5 \end{cases} \tag{8}$$

In this way, sigmoid functions can be implemented by comparison and oblivious selection In the experiments for binary classifications, we use the 5-piece sigmoid function.

## 5.6 Max-pooling

Pooling layers can effectively reduce the size of the parameter matrix and thus reducing parameters in the final connection layer. Therefore, adding pooling layers can speed up the computation and prevent overfitting. In this paper, we mainly use max-pooling, the functionality of which is to return the maximum value of a small window. To reduce communication rounds, the input secret shared values are grouped in the form of a balanced tree to allow multiple comparisons to be computed in parallel.

## 5.7 Stochastic gradient descent (SGD)

SGD is an efficient approximation algorithm for gradually searching for a local optimum of a problem. As a widely used optimization function, SGD has proven to converge to a global minimum and is usually very fast in practice. In addition, how to securely compute SGD with MPC has been explored by a series of studies, which only involves basic arithmetic operations. As a result, we mainly focus on SGD in this paper. The workflow of SGD algorithm is as follows: the coefficients are initialized to random values or all zeros. In each iteration, a coefficient $w_j$ is updated as

$$w_j \coloneqq w_j - \frac{\alpha}{B} \sum_{i=1}^{B} \frac{\partial l_i}{\partial w_j}. \tag{9}$$

where $\alpha$ is the learning rate, $B$ is the mini-batch size and $l_i$ is the loss regarding the $i$th sample in the mini-batch.

# 6 Case studies

In this section, we conduct a series of experiments based on the Irish CER dataset to demonstrate that our protocols not only efficiently maintain the confidentiality of the raw data, but also ensure the accuracy of the models.

## 6.1 Dataset description

We conduct experiments on a public dataset provided by Commission for Energy Regulation. (2012), which is the regulator for the electricity and natural gas sectors in Ireland. The CER dataset contains raw smart meter data of 4,232 residential consumers. The smart meter data is recorded at an interval of 30 min over a total of 75 weeks. In the data cleansing process, if the measurements for one of the weeks have missing data, we will delete the load profiles of this week. Besides, we limit each week starting on Friday. We select a total of 20,000 weeks of smart meter data, where the measurements of 17,000 weeks are used to train the models, and the rest are used to test the model performance.

In addition to smart meter data, the CER dataset also contains the characteristics information of the participants, which is privately collected through the questionnaire. The surveyed issues are mainly in three categories: the occupant socio-demographic information (*e.g.*, employment, social class), consumption habits (*e.g.*, the number of energy-efficient light bulbs), home appliances (*e.g.*, cooking facility type). We select

TABLE 2 The characteristics to be studied.

| Question No. | Consumer Characteristic Question | Class Labels | Number |
|---|---|---|---|
| 300 | Age of chief income earner | Young (<35) | 436 |
| | | Medium (3565) | 2819 |
| | | Old (>65) | 953 |
| 310 | Chief income earner has retired or not | Yes | 1285 |
| | | No | 2947 |
| 401 | Social class of chief income earner | A or B | 642 |
| | | C1 or C2 | 1840 |
| | | D or E | 1593 |
| 410 | Have children or not | Yes | 1229 |
| | | No | 3003 |
| 450 | House type | Detached or bunglow | 2189 |
| | | Semi-detached or terraced | 1964 |
| 453 | Age of the house | Old (>30) | 2151 |
| | | New(<30) | 2077 |
| 460 | Number of bedrooms | Very low (<3) | 404 |
| | | low (=3) | 1884 |
| | | High (4) | 1470 |
| | | Very high (>4) | 474 |
| 4704 | Cooking facility type | Electrical | 1272 |
| | | Not electrical | 2960 |
| 4905 | Energy-efficient light bulb proportion | Up to half | 2041 |
| | | Three quarters or more | 2191 |

TABLE 3 Hyperparameters of network A.

| Layer | Layer Type | Hyperparameters | Activation function |
|---|---|---|---|
| FC1 | Fully Connected | Input size: 7 × 48<br>Neuron number: 128 | ReLU |
| FC2 | Fully Connected | Input size: 128<br>Neuron number: 128 | ReLU |
| FC3 | Fully Connected | Input size: 128<br>Neuron number: 1 | - |
| O1 | Output | - | Sigmoid |

TABLE 4 Hyperparameters of network B.

| Layer | Layer Type | Hyperparameters | Activation function |
|---|---|---|---|
| C1 | Convolution | Input size: 7 × 48<br>Kernel size: 3 × 3<br>Kernel number: 16 | ReLU |
| C2 | Convolution | Input size: 5 × 46<br>Kernel size: 3 × 3<br>Kernel number: 16 | ReLU |
| P1 | Max-Pooling | Window size: 2 × 2 | - |
| FC1 | Fully Connected | Neuron number: 32 | ReLU |
| O1 | Output | - | Softmax |

nine survey questions for benchmarking, which are listed in Table 2.

## 6.2 Setup

We implement privacy-preserving deep learning for electricity consumer characteristics identification using the MP-SPDZ framework (Keller, 2020). The framework enables benchmarking the secure program with a series of generic MPC protocols. All experiments are run on a commodity desktop equipped with Intel (R) Core i7-11700K CPU at 3.60 GHz × 16 running Ubuntu 20.04 on VMware Workstation allocated with 32 GB memory, ignoring network latework. We set the batch size to $B = 128$ and the bit length to $l = 64$. The learning rate $\alpha$ is settled for 0.01. The fixed-point values are set to 16-bit precision with probabilistic truncation. We mainly use the two neural networks shown in Table 3, 4. Network A is used to train

TABLE 5 Performance of communication (MB/epoch), computation (s/epoch) and accuracy (%).

| Question No. | Communication | Computation | Accuracy |
|---|---|---|---|
| 300 | 63432 | 360 | 70.09 |
| 310 | 10441 | 24 | 71.83 |
| 401 | 63432 | 348 | 57.33 |
| 410 | 10441 | 24 | 74.67 |
| 450 | 10441 | 23 | 62.40 |
| 453 | 63432 | 353 | 66.53 |
| 460 | 63432 | 356 | 54.43 |
| 4704 | 10441 | 24 | 66.87 |
| 4905 | 10441 | 24 | 63.40 |

the binary-class classifiers for #310, #410, #450, #4704, #4905, while Network B is used to train the multi-class classifiers for #300, #401, #453, #460. The computation costs and accuracies reported are averaged over ten runs. The accuracies are recorded at 10 epochs.

## 6.3 Performance evaluation

Table 5 details the performance of the two deep neural network models we tested. Network A consists of three fully connected layers, where the first and second fully connected layers use the ReLU activation function. For the output layer of Network A, we set the sigmoid function as activation function. The computation cost required for Network A is desirable, only 24 s for each epoch. While the communication cost is 10,441 MB for each epoch. Network B contains two convolutional layers and one fully connected layer, all of which use the ReLU activation function. After the second convolution layer, we set a max-pooling layer with a window size of $2 \times 2$. For the output layer of Network B, we set the softmax function as activation function. Compared with Network A, Network B needs to invoke more secure comparisons and multiple costly building blocks, including division, logarithm and exponentiation. So, it requires more communication and computation costs. The computation cost required for Network B is around 354 s for each epoch, while the communication cost is 63,432 MB for each epoch. The communication and computation costs required are practically affordable for the resource-rich servers. In addition, the random bit generation can be performed in the preprocessing phase when servers are idle, so as to reduce the burden on servers to provide privacy-preserving deep learning services.

Now, we report the average accuracy of the survey questions. One third of the survey questions have accuracies higher than 70%, which are #300, #310 and #410. The classifiers for these three survey questions are all trained using network A. The survey question #410 has the highest accuracy of 74.67%. Only two survey questions have accuracies less than 60%, which are #401 and #460. The accuracy of the remaining survey questions is 60%～70%. In summary, the accuracy of Network A is comparable, while Network B needs to be adjusted to improve the accuracy.

# 7 Conclusion

We implement privacy-preserving deep learning for electricity consumer characteristics identification by lightweight replicated secret sharing techniques, which not only enable to protect the retailer's sensitive raw data but also achieve favorable performance. Our system allows retailers to stay offline after uploading the shares of smart meter data, and the burden of computation is transferred to three powerfully equipped servers. After the training of the models, retailers can enjoy the inference service provided by servers or download the model parameters directly. Future work might consider improving the accuracy of the deep neural network models.

# Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

# Author contributions

ZZ: Conceptualization, Methodology, Resources, Validation, Visualization, Writing—Original Draft, Writing—Review and Editing. QL: Funding Acquisition, Writing—Review and Editing, Project Administration. HX: Resources, Validation. GX: Validation, Visualization. FK: Software, Supervision. YY: Software, Validation, Language Modification.

# Funding

This work is supported by Project funded by China Postdoctoral Science Foundation under Grant 2022T150416 and sponsored by Shanghai Pujiang program under Grant 21PJ1407000.

# Conflict of interest

Author YY was employed by the company Beijing Jingyi city science and Industry Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Aliasgari, M., Blanton, M., Zhang, Y., and Steele, A. (2013). *Secure computation on floating point numbers*. (San Diego, CA, United States: computer arithmetic).

Aly, A., and Smart, N. P. (2019). "Benchmarking privacy preserving scientific operations," in International Conference on Applied Cryptography and Network Security (Springer), 509–529.

Araki, T., Furukawa, J., Lindell, Y., Nof, A., and Ohara, K. (2016). "High-throughput semi-honest secure three-party computation with an honest majority," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 805–817.

Asharov, G., Lindell, Y., Schneider, T., and Zohner, M. (2013). "More efficient oblivious transfer and extensions for faster secure computation," in Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security,535–548.

Beckel, C., Sadamori, L., and Santini, S. (2013). "Automatic socio-economic classification of households using electricity consumption data," in Proceedings of the fourth international conference on Future energy systems, 75–86.

Beckel, C., Sadamori, L., Staake, T., and Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy* 78, 397–410. doi:10.1016/j.energy.2014.10.025

Benaloh, J., and Leichter, J. (1988). "Generalized secret sharing and monotone functions," in Conference on the Theory and Application of Cryptography (Springer), 27–35.

Canetti, R. (2001). "Universally composable security: A new paradigm for cryptographic protocols," in Proceedings 42nd IEEE Symposium on Foundations of Computer Science (IEEE), 136–145.

Catrina, O., and Saxena, A. (2010). "Secure computation with fixed-point numbers," in Financial Cryptography and Data Security, 14th International Conference, FC 2010, Tenerife, Canary Islands, January 25-28, 2010, Revised Selected Papers.

Chen, S., and Liu, C.-C. (2017). From demand response to transactive energy: State of the art. *J. Mod. Power Syst. Clean. Energy* 5, 10–19. doi:10.1007/s40565-016-0256-x

Chen, S., Love, H. A., and Liu, C.-C. (2016). Optimal opt-in residential time-of-use contract based on principal-agent theory. *IEEE Trans. Power Syst.* 31, 4415–4426. doi:10.1109/tpwrs.2016.2518020

Chicco, G. (2016). "Customer behaviour and data analytics," in 2016 International Conference and Exposition on Electrical and Power Engineering (EPE) (IEEE), 771.

Commission for Energy Regulation (CER) (2012). CER smart metering Project - electricity customer behaviour trial, 2009-2010. [Online]. Available at: https://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

Dalskov, A. P. K., Escudero, D., and Keller, M. (2021). "Fantastic four: Honest-majority four-party secure computation with malicious security," in USENIX Security Symposium.

Demmler, D., Schneider, T., and Zohner, M. (2015). "Aby-a framework for efficient mixed-protocol secure two-party computation," in NDSS.

Eerikson, H., Keller, M., Orlandi, C., Pullonen, P., Puura, J., and Simkin, M. (2019). *Use your brain! arithmetic 3pc for any modulus with active security* In 5:1–5:24. ePrint Arch. doi:10.4230/LIPIcs.ITC.2020.5

Escudero, D., Ghosh, S., Keller, M., Rachuri, R., and Scholl, P. (2020). "Improved primitives for mpc over mixed arithmetic-binary circuits," in Annual International Cryptology Conference.

Gentry, C. (2009). "Fully homomorphic encryption using ideal lattices," in Proceedings of the forty-first annual ACM symposium on Theory of computing, 169–178.

Goldreich, O. (2009). *Foundations of cryptography: Volume 2, basic applications*. (Cambridge, United Kingdom: Cambridge University Press).

Goldreich, O., Micali, S., and Wigderson, A. (2019). "How to play any mental game, or a completeness theorem for protocols with honest majority," in Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali, 307–328.

Goldschmidt, R. E. (1964). *Applications of division by conPrivacy-preserving household characteristic identification with federatedvergence*. Ph.D. thesis. (Cambridge, MA, United States: Massachusetts Institute of Technology).

Hart, J. F. (1978). *Computer approximations*. (Malabar, FL, United States: Krieger Publishing Co., Inc).

Hazay, C., and Lindell, Y. (2010). *Efficient secure two-party protocols: Techniques and constructions*. (Berlin, Germany: Springer Science and Business Media).

Hong, C., Huang, Z., Lu, W.-j., Qu, H., Ma, L., Dahl, M., et al. (2020). "Privacy-preserving collaborative machine learning on genomic data using tensorflow," in Proceedings of the ACM Turing Celebration Conference-China, 39–44.

Jokar, P., Arianpoo, N., and Leung, V. C. (2015). Electricity theft detection in ami using customers' consumption patterns. *IEEE Trans. Smart Grid* 7, 216–226. doi:10.1109/tsg.2015.2425222

Júnior, L. A. P., Ramos, C. C. O., Rodrigues, D., Pereira, D. R., de Souza, A. N., da Costa, K. A. P., et al. (2016). Unsupervised non-technical losses identification through optimum-path forest. *Electr. Power Syst. Res.* 140, 413–423. doi:10.1016/j.epsr.2016.05.036

Keller, M. (2020). "Mp-spdz: A versatile framework for multi-party computation," in CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security.

Keller, M., and Sun, K. (2021). Secure quantized training for deep learning. *arXiv preprint arXiv:2107.00501*

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539

Li, R., Wang, Z., Gu, C., Li, F., and Wu, H. (2016). A novel time-of-use tariff design based on Gaussian mixture model. *Appl. energy* 162, 1530–1536. doi:10.1016/j.apenergy.2015.02.063

Liang, H., Ma, J., Sun, R., and Du, Y. (2019). A data-driven approach for targeting residential customers for energy efficiency programs. *IEEE Trans. Smart Grid* 11, 1229–1238. doi:10.1109/tsg.2019.2933704

Lin, J., Ma, J., and Zhu, J. (2021). Privacy-preserving household characteristic identification with federated learning method. *IEEE Trans. Smart Grid* 13, 1088–1099. doi:10.1109/tsg.2021.3125677

Lu, Q., Li, S., Zhang, J., and Jiang, R. (2022). Pedr: Exploiting phase error drift range to detect full-model rogue access point attacks. *Comput. Secur.* 114, 102581. doi:10.1016/j.cose.2021.102581

Mallapuram, S., Ngwum, N., Yuan, F., Lu, C., and Yu, W. (2017). "Smart city: The state of the art, datasets, and evaluation platforms," in 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS) (IEEE), 447–452.

Mohassel, P., and Rindal, P. (2018). "Aby3: A mixed protocol framework for machine learning," in Proceedings of the 2018 ACM SIGSAC conference on computer and communications security, 35–52.

Mohassel, P., and Zhang, Y. (2017). "Secureml: A system for scalable privacy-preserving machine learning," in 2017 IEEE symposium on security and privacy (SP) (IEEE), 19.

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines vinod nair," in Proceedings of the 27th International Conference on Machine Learning (ICML-10).

Paillier, P. (1999). "Public-key cryptosystems based on composite degree residuosity classes," in International conference on the theory and applications of cryptographic techniques (Springer), 223–238.

Rotaru, D., and Wood, T. (2019). "Marbled circuits: Mixing arithmetic and boolean circuits with active security," in International Conference on Cryptology in India (Springer), 227–249.

Taieb, S. B., Huser, R., Hyndman, R. J., and Genton, M. G. (2016). Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Trans. Smart Grid* 7, 2448–2455. doi:10.1109/tsg.2016.2527820

Viegas, J. L., Vieira, S. M., and Sousa, J. (2016). "Mining consumer characteristics from smart metering data through fuzzy modelling," in International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (Springer), 562–573.

Wagh, S., Gupta, D., and Chandran, N. (2019). Securenn: 3-party secure computation for neural network training. *Proc. Priv. Enhancing Technol.*, 26–49. doi:10.2478/popets-2019-0035

Wagh, S., Tople, S., Benhamouda, F., Kushilevitz, E., Mittal, P., and Rabin, T. (2021). Falcon: Honest-majority maliciously secure framework for private deep learning. *Proc. Priv. Enhancing Technol.* 1, 188–208. doi:10.2478/popets-2021-0011

Wang, Y., Chen, Q., Gan, D., Yang, J., Kirschen, D. S., and Kang, C. (2018). Deep learning-based socio-demographic information identification from smart meter data. *IEEE Trans. Smart Grid* 10, 2593–2602. doi:10.1109/tsg.2018.2805723

Wang, Y., Chen, Q., Kang, C., and Xia, Q. (2016a). Clustering of electricity consumption behavior dynamics toward big data applications. *IEEE Trans. Smart Grid* 7, 2437–2447. doi:10.1109/tsg.2016.2548565

Wang, Y., Chen, Q., Kang, C., Xia, Q., and Luo, M. (2016b). Sparse and redundant representation-based smart meter data compression and pattern extraction. *IEEE Trans. Power Syst.* 32, 2142–2151. doi:10.1109/tpwrs.2016.2604389

Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., and Zhao, Y. (2015). Load profiling and its application to demand response: A review. *Tinshhua. Sci. Technol.* 20, 117–129. doi:10.1109/tst.2015.7085625

Yao, A. C.-C. (1986). "How to generate and exchange secrets," in 27th Annual Symposium on Foundations of Computer Science (sfcs 1986) (IEEE), 162–167.

Zhang, H., Gao, P., Yu, J., Lin, J., and Xiong, N. N. (2021). Machine learning on cloud with blockchain: A secure, verifiable and fair approach to outsource the linear regression. *arXiv preprint arXiv:2101.02334*

Zhang, H., Yu, J., Tian, C., Xu, G., Gao, P., and Lin, J. (2020). Practical and secure outsourcing algorithms for solving quadratic congruences in internet of things. *IEEE Internet Things J.* 7, 2968–2981. doi:10.1109/jiot.2020.2964015

Zhong, S., and Tam, K.-S. (2014). Hierarchical classification of load profiles based on their characteristic attributes in frequency domain. *IEEE Trans. Power Syst.* 30, 2434–2441. doi:10.1109/tpwrs.2014.2362492

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 16 December 2022
DOI 10.3389/fenrg.2022.1041862



OPEN ACCESS

EDITED BY
Dou An,
MOE Key Laboratory for Intelligent Networks and Network Security, China

REVIEWED BY
Youcef Belkhier,
Maynooth University, Ireland
Jing Tian,
Shenyang Aerospace University, China
Hui Liu,
Xi'an University of Posts and Telecommunications, China

*CORRESPONDENCE
Jia Han,
hanjia27@126.com

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 11 September 2022
ACCEPTED 21 October 2022
PUBLISHED 16 December 2022

CITATION
Han J, Yu Z and Yang J (2022), Multimodal attention-based deep learning for automatic modulation classification. *Front. Energy Res.* 10:1041862. doi: 10.3389/fenrg.2022.1041862



# Multimodal attention-based deep learning for automatic modulation classification

Jia Han[1]*, Zhiyong Yu[1] and Jian Yang[2]

[1]Department of Computer, Rocket Force University of Engineering, Xi'an, Shaanxi, China, [2]Department of Engineering, Rocket Force University of Engineering, Xi'an, Shaanxi, China

Wireless Internet of Things (IoT) is widely accepted in data collection and transmission of power system, with the prerequisite that the base station of wireless IoT be compatible with a variety of digital modulation types to meet data transmission requirements of terminals with different modulation modes. As a key technology in wireless IoT communication, Automatic Modulation Classification (AMC) manages resource shortage and improves spectrum utilization efficiency. And for better accuracy and efficiency in the classification of wireless signal modulation, Deep learning (DL) is frequently exploited. It is found in real cases that the signal-to-noise ratio (SNR) of wireless signals received by base station remains low due to complex electromagnetic interference from power equipment, increasing difficulties for accurate AMC. Therefore, inspired by attention mechanism of multi-layer perceptron (MLP), AMC-MLP is introduced herein as a novel AMC method for low SNR signals. Firstly, the sampled I/Q data is converted to constellation diagram, smoothed pseudo Wigner-Ville distribution (SPWVD), and contour diagram of the spectral correlation function (SCF). Secondly, convolution auto-encoder (Conv-AE) is used to denoise and extract image feature vectors. Finally, MLP is employed to fuse multimodal features to classify signals. AMC-MLP model utilizes the characterization advantages of feature images in different modulation modes and boosts the classification accuracy of low SNR signals. Results of simulations on RadioML 2016.10A public dataset prove as well that AMC-MLP provides significantly better classification accuracy of signals in low SNR range than that of other latest deep-learning AMC methods.

KEYWORDS

**Internet of things, automatic modulation classification, auto-encoder, deep learning, spectrum sensing**

## Introduction

AMC refers to the automatic and fast classification of unknown signal modulation types by algorithms. AMC has been widely used in military and civil wireless communications, which can efficiently manage spectrum resources. In the power wireless Internet of Things (IoT), there are many types of wireless communication terminals, diverse modulation methods, and complex electromagnetic environment of wireless channels, which render AMC operation extremely difficult. AMC algorithm can

not only be compatible with a variety of wireless communication terminals, but also reduce the price of the system (Abdel-Moneim et al., 2021). Traditional AMC methods can be sorted into two categories, namely likelihood based (LB) and feature based (FB) methods. As the name implies, LB algorithm is based on likelihood, where different types of likelihood functions are used to improve the classification accuracy. There are four common likelihood functions: maximum likelihood (ML) (Wen and Mendel, 2000), average likelihood ratio test (ALRT) (Huan and Polydoros, 1995; Hong and Ho, 2003), generalized likelihood ratio test (GLRT) (Panagiotou et al., 2000) and hybrid likelihood ratio test (HLRT) (Hong et al., 2001). Due to the high space complexity and time complexity of LB algorithm in condition of too many modulation types and unknown parameters, the classification accuracy is low when faced with the new modulation mode, which cannot meet the requirements. Many researchers are committed to the research of FB algorithm: Nandi and Azzouz (1998) and Shen and Gao (2014) proposed spectrum as the main classification feature, but the classification accuracy drops sharply when it comes to intra class modulation; Orlic and Dukic (2009), Mirarab and Sobhani (2007) proposed a method to classify signals based on statistical distribution features. However, there are also problems such as computational complexity and dependence on prior knowledge, and only a few modulation types with obvious features can be identified; Yu et al. (2003), Zhou et al. (2017), and Satija et al. (2015) proposed a method to classify using signal transformation domain features, which has better classification accuracy at high SNR and lower classification accuracy at low SNR; Mobasseri (2000) proposed to use the constellation diagram for classification, which achieved good results under high SNR, but failed to classify new modulation methods. To sum up, the traditional AMC classification method cannot to meet the classification task of new modulation methods, and encounters a problem of excessive space-time consumption.

With the development of artificial intelligence technology, DL has been used in data processing and analysis, and is being applied in the field of AMC. In recent years, DL has been widely used to solve AMC problems. Convolutional neural network (CNN) was first used to directly perform AMC on I/Q raw data. Experiments show that its performance is significantly much better than the classification method based on cyclic spectral features (O'Shea et al., 2016). Long-Short Term Memory (LSTM) neural networks are used to establish the characteristics of the relationship between amplitude and phase of sequential I/Q data. When using a fully connected network for classification, the average classification accuracy of the proposed model is close to 90% under various SNR of 0–20 dB, and good experimental results have been obtained (Rajendran et al., 2018). Some researchers proposed to use the SCF to generate two-dimensional profiles of modulated signals, and then use CNN network for classification, which also achieved good classification results under low SNR (Zhang et al., 2021). Hou et al. (2021) transformed one-dimensional I/Q signals into SPWVD, and then used CNN to extract features for AMC, which also achieved good classification accuracy. Qiao et al. (2022) aimed to solve the problem of low classification accuracy with low SNR. A denoizing and a classification network were used for synchronous learning, which effectively improved the classification accuracy and performed better than the existing classification methods. In the 0 dB SNR environment, the proposed multi-task CNN method outperforms the traditional CNN method by 20%. Ke and Vikalo (2022) designed a learning framework for LSTM denoising encoder, which can automatically extract stable robustness features from noisy signals according to amplitude and phase, and use the learned robustness features for modulation classification. This model is structurally compact, easy to implement on low-cost embedded platforms, and can effectively classify received wireless signals. Mao et al. (2021) designed a multi constellation AMC framework, used CNN network to extract deep features, weighted the attention of feature vectors, and finally implemented AMC, which achieved good classification results on the open dataset. Xu and Darwazeh (2020) used Software Defined Radio (SDR) to test the real environment and evaluate various performances, providing specific test contents. Although DL method can quickly and accurately classify modulation modes under high SNR, due to the existence of electromagnetic interference of power equipment, he low SNR of wireless channel results in the low classification accuracy of DL method. The existing digital modulation modes cannot be classified accurately using I/Q data or using a single features map. It is necessary to study a DL model that can resist noise and intra class modulation methods.

In recent years, self-attention mechanism has gradually shifted from natural language processing to computer vision. Vaswani et al. (2017) used the self-attention mechanism under the transformer architecture to process natural language sequences in parallel, significantly improving the processing speed and accuracy, and obtained good experimental results. Dosovitskiy et al. (2020) introduced the self-attention mechanism into the field of computer vision and achieved excellent performance on several benchmark datasets, such as ImageNet, COCO and ADE20k. Compared with the traditional CNN algorithm, self-attention can establish the global relationship, which is different from the local relationship established by CNN, and it has a great improvement in visual application. Liu et al. (2021) and Tolstikhin et al. (2021), likewise achieved good classification accuracy on public datasets using MLP, proving that in addition to the transformer, MLP also efficiently achieves image data classification tasks. The AMC is qualified to classify signals in different spaces such as space-time features, statistical features and time-frequency features and unsuitable to classify existing modulated signals with a single feature. Therefore, it is necessary to extract and fuse features of multi-dimensional spatial information, and further use DNN using self-attention mechanism for AMC.

In the face of strong electromagnetic interference, in order to improve the classification accuracy and robustness of AMC algorithm, the modulated signal is characterized in multimodal in this paper, which avoids the lack of representation ability of a single feature. The modulated signal is characterized from the space-time characteristics, time-frequency characteristics and statistical characteristics respectively. Therefore, the constellation diagram, SPWVD and contour diagram of the SCF are used to represent the modulation characteristics of the signal. Constellation diagram can be used to classify the modulated signal in the space-time domain. SPWVD can classify modulation types in time-frequency domain. The SCF reflects the statistical characteristics of the signal and is insensitive to interference. It has good noise resistance and can keep the classification effect in the low SNR range. To reduce the computational complexity, we use the contour diagram of the SCF, and we use the above three images in the production of the dataset.

Unlike the existing research, most DL modulation classification schemes mainly select the characteristics of a single as the input of the network or optimize the network structure for high-dimensional mapping to improve modulation recognition performance, ignoring the complementarity between features in different transformation domains and different classifiers. Unlike the existing research, instead of inputting the signal into the classifier, we preprocess the data, including dataset construction of three feature diagram, image synchronization denoising using Conv-AE, feature vector extraction, construct global relationship construction using self-attention, implement AMC after multimodal feature fusion and verify the classification accuracy.

In summary, the main contributions of this study are as follows:

1. In the space-time domain, time-frequency domain and statistical domain, use the multi-modal characteristics of the constellation diagram, SPWVD and contour diagram of SCF as the network input.
2. The design uses Conv-AE for synchronous denoising and low dimensional feature extraction of feature maps, which is helpful to improve the robustness of the model and simplify the model parameters, thus simplifying the MLP model, accelerating the model training and reasoning.
3. Use multimodal feature fusion method, use the complementarities between feature maps, enhance the communication between different transform domains, improve the feature expression ability. Use MLP of self-attention mechanism for classification
4. Study the classification accuracy changes of different types of modulated signals in different additive white Gaussian noise (AWGN) channels and compared with the reference method.

The rest of this paper is organized as follows. This paper proposes a multimodal modulation classification modal based on MLP self-attention mechanism, which is composed of constellation diagram, SPWVD, SCF contour diagram data generation module, Conv-AE feature denoising and extraction module, and MLP self-attention classification module. We provide the architecture model of the system, and then complete the algorithm analysis and dataset generation of constellation diagram, SPWVD and SCF contour diagram, is presented in detail in Section 2. Then, in Section 3, we analyze the experimental process, simulation test and result analysis to prove the effectiveness of our algorithm and its superiority over the benchmark algorithm. Finally, a brief conclusion is given in Section 4.

## Materials and methods

In this section, we introduce the proposed AMC system design, including the feature map generation module, the feature extraction module based on Conv-AE, and the classifier structure of MLP attention. The data set from RadioML 2016.10A (O'Shea and West, 2016) is used in our experiment to generate feature maps.

### System model

The proposed AMC model, shown in Figure 1, classifies the modulation types of the Conv-AE eigenvectors of the constellation diagram, SPWVD, and the contour diagram of SCF. To reduce the noise influence and accurately distinguish intra-and inter-class modulation modes, we first sampled the unknown signal, and then generated the constellation diagram, SPWVD, and the contour diagram of the SCF, respectively. The purpose of this method is to improve the representation of signal in different fields and resist the influence of channel noise. Subsequently, CNN is used to extract the feature vectors of the three feature maps, after which the feature vectors are input into the MLP network of the self-attention mechanism for classification. The following four modulation types are most commonly used in digital communication: the binary phase-shift keying (BPSK), binary frequency-shift keying, Gauss frequency-shift keying (GFSK), quadrature phase-shift keying (QPSK), and 16 quadrature amplitude modulation (16QAM).

### Constellation diagram

Generally, the received signal is expressed as Eq. 1.

$$x(t) = s(t) * c(t) + n(t) \tag{1}$$

$s(t)$ is the transmitted signal without noise, $c(t)$ is the time-varying pulse of the transmission wireless channel, and $n(t)$ is the AWGN of zero mean and variance $\sigma_n^2$ in the wireless signal. $x(t)$ is the received signal, because of the mathematical and physical circuit

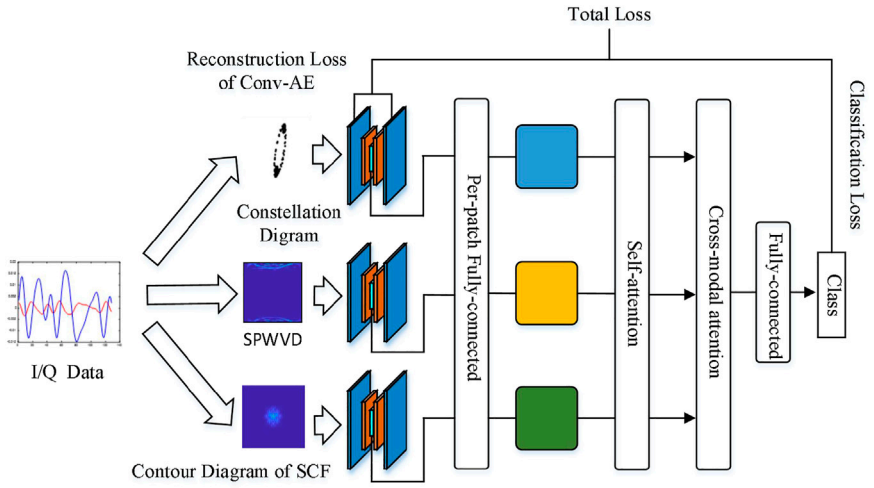



FIGURE 1
Schematic diagram of AMC system. Generation of three different feature maps was completed using the original baseband I/Q data. Conv-AE was used for feature extraction, and the modulated signal classification at different SNR was completed using multimodal attention.

design requires that we commonly use the I/Q format to represent for the in-phase component and quadrature component, the received signal samples $x_i = (I_i, Q_i)$, including $I_i = A_i \cos(\phi_i)$ and $Q_i = A_i \sin(\phi_i)$, where $A_i$ and $\phi_i$ are the instantaneous amplitude and phase Angle of the received signal $x(t)$, as shown in Eq. 2.

$$A_i = \sqrt{I_i^2 + Q_i^2}$$
$$\phi_i = arctan\left(\frac{Q_i}{I_i}\right) \tag{2}$$

The constellation diagram is a 2-D image representation of scatterers drawn from baseband I/Q sampled data in the I/Q coordinate system. The generated image is shown in Figure 2. It is often used for modulation signal classification, as it can efficiently characterize the modulation type and data order. There is a good mapping relationship between the constellation diagram and modulation type, especially at classification accuracy. When the SNR is high, the modulation types are efficiently classified; however, due to interference of noise in the channel, it is difficult to identify high-order modulation signals at low SNR. Therefore, the modulation classification method using the constellation diagram is a difficult task in low SNR environments.

## SPWVD

For the modulated signal $x(t)$, its Wigner–Ville Distribution (WVD) is the Fourier transform of the instantaneous correlation function of $x(t)$, which is defined as Eq. 3:

$$W_x(t,f) = \int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right)x^*\left(t+\frac{\tau}{2}\right)e^{-j2\pi f\tau}d\tau = \int_{-\infty}^{+\infty} R_x(t,\tau)e^{-j2\pi f\tau}d\tau \tag{3}$$

where $\tau$ is the delay variable, $t$ and $f$ are the time and frequency variables, respectively, and $R_x(t,\tau)$ is the instantaneous correlation function of the signal $x(t)$. WVD represents the joint energy distribution of a signal in the time-frequency domain, and has two important properties, namely, time- and frequency-shift invariance.

To suppress the influence of cross terms, the pseudo Wigner–Ville distribution (PWVD) is obtained by time-domain windowing based on WVD. The windowed method not only retains the excellent performance (better resolution) of the original algorithm WVD but also eliminates some cross-term interference.

$$PWVD_x(t,f) = \int_{-\infty}^{+\infty} h(\tau)x\left(t+\frac{1}{2}\tau\right)x^*\left(t-\frac{1}{2}\tau\right)e^{-j2\pi f\tau}d\tau \tag{4}$$

$h(\tau)$ is a window function added to the time domain, which is equivalent to a low-pass filter. It plays a smooth role in the frequency domain, to reduce the cross-term interference of multi-component signals in the frequency direction. However, it also destroys the edge distribution and instantaneous frequency characteristics of WVD.

SPWVD a one-time windowing process in the frequency domain based on PWVD, and its definition is given by Eq. 5.

$$SPWVD_x(t,f) = \iint h(\tau)g(v)x\left(t-v+\frac{\tau}{2}\right)x^*\left(t-v-\frac{\tau}{2}\right)e^{-j2\pi f\tau}dvd\tau \tag{5}$$

Herein, $h(\tau)g(\mathrm{v})$ is the two window functions of Winger–Wiley distribution in frequency domain and time domain, which

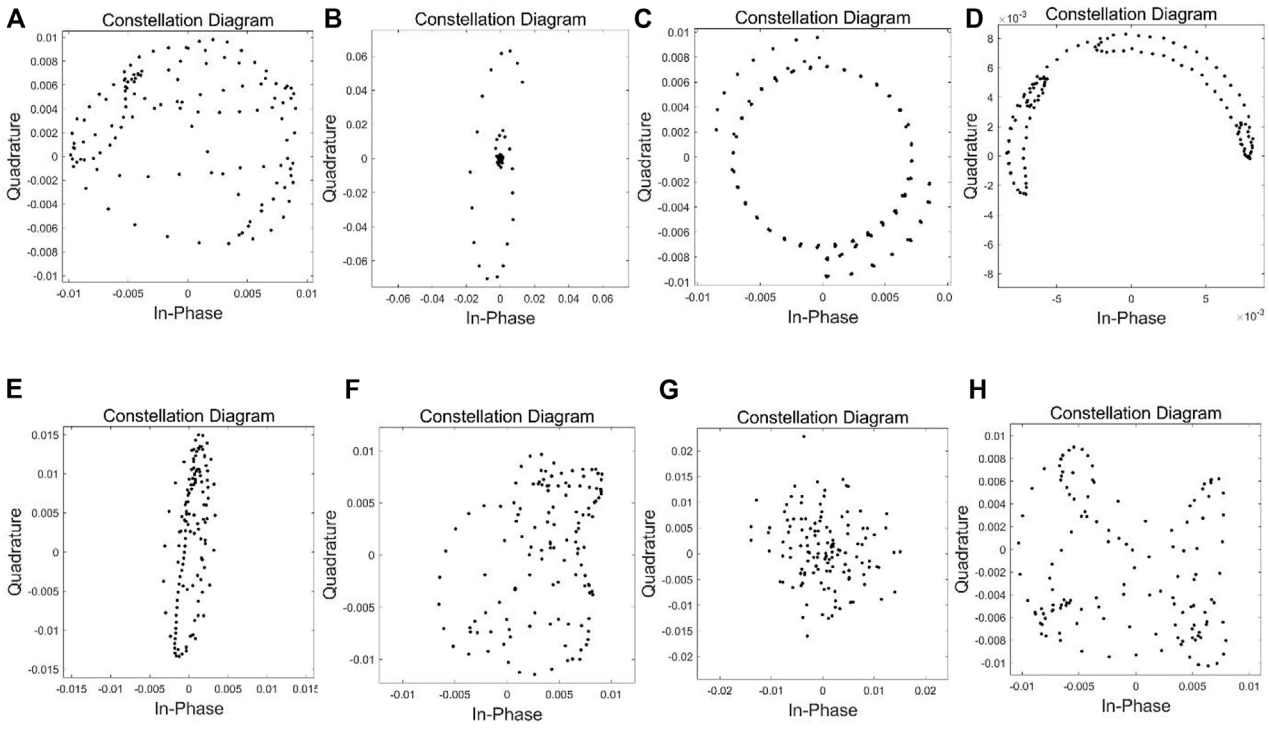



FIGURE 2
Constellation diagrams of modulated signals at SNR = 18 dB in RadioML 2016.10A dataset. **(A)** 8PSK, **(B)** BPSK, **(C)** CPFSK, **(D)** GFSK, **(E)** PAM4, **(F)** QAM16, **(G)** QAM64, **(H)** QPSK.

realizes the double smoothing effect in time-frequency domain, and the two window functions are both real even functions. Compared with PWVD, the two window functions in SPWVD definition are windowed in the time and frequency domain, respectively, that is, filtering is carried out in time and frequency domain at the same time to achieve the elimination of cross-term interference to a large extent. $x(t)$ is the analytic signal of $r(t)$ as given by Eq. 6:

$$x(t) = r(t) + jH[r(t)] \tag{6}$$

where $H[\cdot]$ represents the Hilbert transformation. The generated image is shown in Figure 3.

## Contour diagram of spectral correlation function

Because the autocorrelation function $R_x(t,\tau)$ is periodic, its Fourier series expansion is performed as Eq. 7.

$$R_x(t,\tau) = \sum R_x^{\alpha}(\tau) e^{j2\pi\alpha t} \tag{7}$$

where $R_x(t,\tau)$ is called the cyclic autocorrelation function and represents the cyclic autocorrelation strength of random process $x(t)$ at frequency $\alpha$, which is defined as Eq. 8:

$$R_x^{\alpha}(\tau) \triangleq \lim_{\tau\to\infty} \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\alpha t} dt \tag{8}$$

where $\alpha$ is the cycle frequency. When $\alpha = 0$, $R_x^{\alpha}(\tau)$ is a conventional autocorrelation function. Taking the Fourier transform of the cyclic autocorrelation function $R_x^{\alpha}(\tau)$:

$$S_x^{\alpha}(f) = \int_{-\infty}^{\infty} R_x^{\alpha}(\tau) \mathrm{e}^{-j2\pi f\tau} \mathrm{d}\tau \tag{9}$$

$S_x^{\alpha}(f)$ is the cyclic spectrum density function substituting Eq. 8 into Eq. 9, the cyclic spectral density function can be expressed as:

$$S_x^{\alpha}(f) = \lim_{T_0\to\infty} \frac{1}{T_0} X_{T_0}\left(t, f+\frac{\alpha}{2}\right) X_{T_0}^*\left(t, f-\frac{\alpha}{2}\right) \tag{10}$$

where $X_{T_0}(t,f)$ is the short-time Fourier transform of the stochastic process $x(t)$:

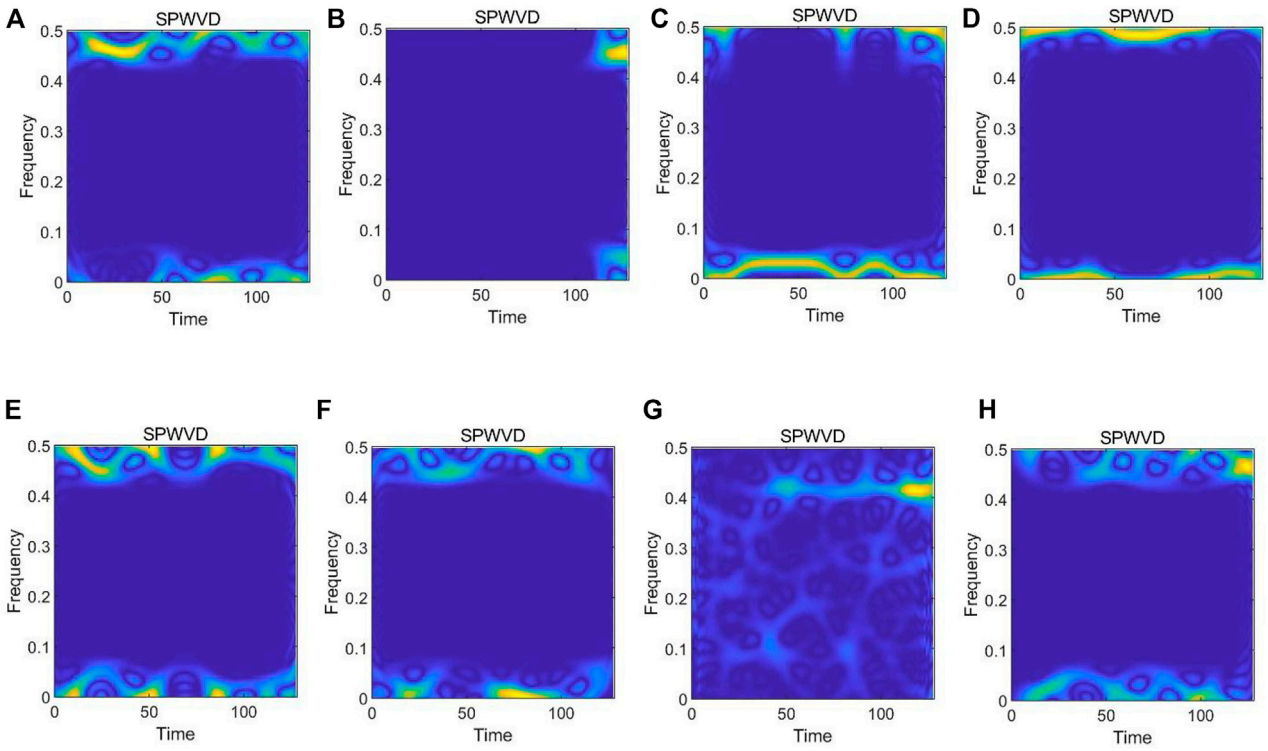



FIGURE 3
Smoothed pseudo Wigner–Ville distribution of modulated signals at SNR = 18 dB in RadioML 2016.10A dataset. **(A)** 8PSK, **(B)** BPSK, **(C)** CPFSK, **(D)** GFSK, **(E)** PAM4, **(F)** QAM16, **(G)** QAM64, **(H)** QPSK.

$$X_{T_0}(t,f) = \int_{t-T_0}^{t+T_0} x(u)e^{-j2\pi fu}du \tag{11}$$

Eq. 10 shows that the cyclic spectral density value at a frequency $f$ in the spectrum of the stochastic process $x(t)$ can be obtained by the cross-correlation of two short-time Fourier transform components above and below $f$ with a spacing of $\alpha/2$. Therefore, the cyclic spectral density function is also known as the spectral correlation function (SCF).

The FFT accumulation method (FAM) employed by Roberts et al. (1991) is used, where the discrete smoothed cycle period plot in the time domain is expressed as Eq. 12.

$$S^{\alpha}_{x_{N'}}(n,f) = \frac{1}{N}\sum_{n=0}^{N-1}\left[\frac{1}{N'}X_{N'}\left(n,f+\frac{\alpha}{2}\right)X^{*}_{N}\left(n,f-\frac{\alpha}{2}\right)\right] \tag{12}$$

In Eq. 12, $N$ represents the total length of data, $X_N(n,f)$ is the discrete short-time Fourier transform of random process $x(t)$, Eq. 13.

$$X_{N'}(n,f) = \sum_{n=0}^{N-1} w(n)x(n)e^{-\frac{j2\pi nf}{N'}} \tag{13}$$

where $w(n)$ is the window function used to truncate data (such as Hamming window). FAM consists of three basic steps: windowing the input sequence and applying $N'$ point short-time Fourier transform to obtain spectral components with frequency $f$, frequency shifting the output of short-time Fourier transform to obtain two spectral components with an interval of $\alpha/2$ above and below $f$, and replacing the average calculation in smoothing with P point Fourier transform. The generated image is shown in Figure 4.

## Conv-AE

To obtain the low-dimensional features of different feature maps under various modulation modes, we use a multi-layer Conv-AE, including a learnable convolution kernel and activation function, to extract the low-dimensional features of images. The structure of Conv-AE is summarized in Table 1.

For three different feature maps, the same Conv-AE is used for feature extraction. First, the feature map is input and

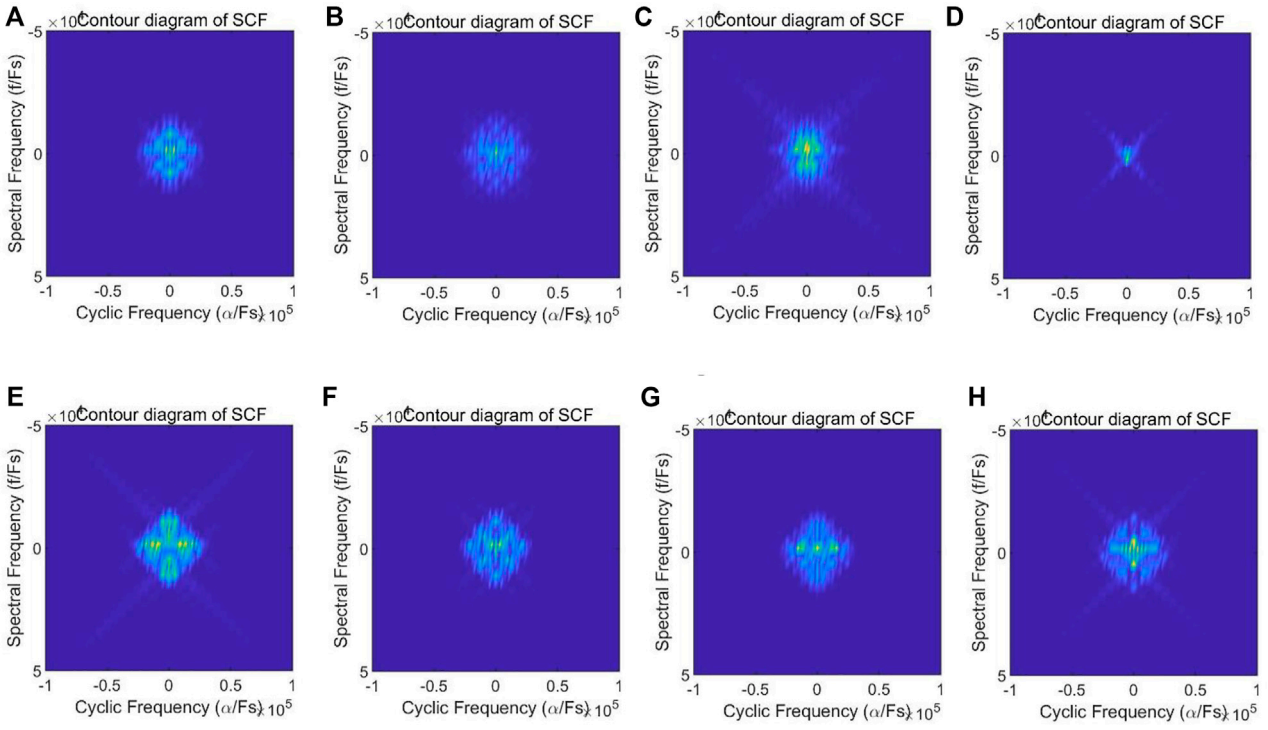



FIGURE 4
Contour diagram of spectral correlation function of modulated signals at SNR = 18 dB in RadioML 2016.10A dataset. **(A)** 8PSK, **(B)** BPSK, **(C)** CPFSK, **(D)** GFSK, **(E)** PAM4, **(F)** QAM16, **(G)** QAM64, **(H)** QPSK.

TABLE 1 Configuration of Conv-AE.

| Stage | Layer | Output | Kernel size |
|---|---|---|---|
| 1 | Input | 3,256,256 | Feature map |
| 2 | Conv2D-1 | 16,256,256 | Number of filters: 16 Kernel Size (2 × 2) |
| 3 | Pool | 16,128,128 | Maxpooling2D: (2 × 2) |
| 4 | Conv2D-2 | 32,128,128 | Number of filters:32 Kernel Size: (2 × 2) |
| 5 | Pool | 32,64,64 | Maxpooling2D: (2 × 2) |
| 6 | Conv2D-3 | 1,64,64 | Number of filters: 1 Kernel Size: (1 × 1) |
| 7 | Flatten | 1,64,64 | Encoder Output |
| 8 | Conv2D-4 | 8,64,64 | Number of filters: 8 Kernel Size: (2 × 2) |
| 9 | Up-Sampling | 8,128,128 | UpSampling2D: (2 × 2) |
| 10 | Conv2D-5 | 16,128,128 | Number of filters: 16 Kernel Size: (2 × 2) |
| 11 | Up-Sampling | 16,256,256 | UpSampling2D (2 × 2) |
| 12 | Conv2D-6 | 32,256,256 | Number of filters:32 Kernel Size: (2 × 2) |
| 13 | Conv2D-7 | 1,256,256 | Number of filters: 1 Kernel Size: (2 × 2) |
| 14 | Output | 3,256,256 | Consistent with Input |

the convolution kernel is used to extract the features. Maximum pooling is used to extract the evident features. After multiple convolution and pooling operations, the feature vectors are obtained. During the entire training process, the input image and the reconstruction loss are calculated.

TABLE 2 Description of RadioML 2016.10A parameters.

| Parameter | Value |
|---|---|
| Sampling frequency | 200 kHz |
| Sampling rate offset standard deviation | 0.01 Hz |
| Maximum sampling rate offset | 50 Hz |
| Carrier frequency offset standard deviation | 0.01 Hz |
| Maximum carrier frequency offset | 500 Hz |
| Sample length | 128 |
| SNR Range | −20 to 18 dB |
| Modulations | BPSK, QPSK, BPSK, GFSK, CPFSK, PAM4, QAM16, QAM64 |

## MLP classifier

To ensure the accuracy of the AMC method, we use the self-attention mechanism in the model, the main idea of which was derived by Liu et al. (2021). We improve the previous single feature classification method and use the attention mechanism to enhance the interaction of features between modules. In the subsequent experimental process, MLP is compared with the traditional methods. Vaswani et al. (2017) reported that the self-attention mechanism is good at capturing the direct relationship of long-distance features in the process of natural language processing, which is different from the CNN method for capturing local features. The self-attention mechanism takes the $d_k$, and $d_v$, values of each patch as the Query and Key, respectively. The Key is the label of each patch, which is used to distinguish the features among them. The Query is used to find all the keys and determine the best matching one. In the self-attention mechanism, we must calculate the dot product of the key of each patch and the Query of the remaining patch, and use the *Softmax* function for classification, essentially converting the number vector into a probability vector, and finally obtaining the weight.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{14}$$

Multi-head Self-attention $(I_1 \rightarrow I_1^{'}, I_1^{''}, I_1^{'''})$
Multi-head Self-attention $(I_2 \rightarrow I_2^{'}, I_2^{''}, I_2^{'''})$
Multi-head Self-attention $(I_3 \rightarrow I_3^{'}, I_3^{''}, I_3^{'''})$

The common multi-head self-attention mechanism in the transformer allows the model to represent relationship information between subspaces of different locations. In the self-attention mechanism, the Query, Key, and Value are generated according to different patches. The self-attention mechanism can learn the interaction between itself and other parts, and predict the correlation between the input and output. This feature can be used for modulation classification. In the example used in this study, the feature vectors of the graph are respectively used as input to generate self-attention, including generating the Key, Query, and Value, after which the cross-attention mechanism (Tan and Bansal, 2019; Golovanevsky et al., 2022) is used to generate the relationship between the three feature graphs. Bi-directional, cross-modal attention can be performed in each multi-head attention module. Attention output between the three groups of feature maps is based on the attention layers that characterize the features, which can be manipulated as the following modules:

Concat (cross-attention $(I_1^{'}, I_1^{''}, I_1^{'''} \rightarrow I_1^{'}, I_1^{''}, I_1^{'''})$, cross-attention $(I_1^{'}, I_1^{''}, I_1^{'''} \rightarrow I_3^{'}, I_3^{''}, I_3^{'''})$)

Concat (cross-attention $(I_2^{'}, I_2^{''}, I_2^{'''} \rightarrow I_2^{'}, I_2^{''}, I_2^{'''})$, cross-attention $(I_2^{'}, I_2^{''}, I_2^{'''} \rightarrow I_3^{'}, I_3^{''}, I_3^{'''})$)

Concat (cross-attention $(I_3^{'}, I_3^{''}, I_3^{'''} \rightarrow I_3^{'}, I_3^{''}, I_3^{'''})$, cross-attention $(I_3^{'}, I_3^{''}, I_3^{'''} \rightarrow I_2^{'}, I_2^{''}, I_2^{'''})$)

Finally, the full connection without attention mechanism of network connection is used to produce a dense layer of output. This method uses the attention mechanism to complete the most advanced algorithm of AMC in MLP, meets the requirements of multi-type, fast and accurate classification of wireless IoT base station, and has a very high practical significance for improving the communication capability in low SNR environment.

## Loss function

The loss function of this model is mainly composed of two parts, namely, the reconstruction loss of Convolution-AE and the classification loss of MLP. The total loss of the algorithm is expressed as the weighted combination of two terms as follows:

$$L_{\text{total}} = (1-\mu_1)L_{\text{classification}} + \mu_1 L_{\text{Conv-AF}} \tag{15}$$

where $\mu_1$ is the hyperparameter controlling the weight loss of classification and auto-encoder reconstruction. In engineering practice, occasionally large $\mu_1$ severely interfere with the classification accuracy and convergence speed of the model. In general, the mean-squared error is used to calculate the reconstruction loss of the autoencoder, and the cross entropy loss is used to calculate the classification loss.

$$L_{\text{Conv-AE}} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f(x_i)\right)^2 \tag{16}$$

In the above Eq. 16, $y_i$ and $f(x_i)$ represent the true and predicted values of the $i$ sample respectively, and n represents the number of samples.

$$L_{\text{classification}} = \frac{1}{N}\sum_{i} L_i = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \tag{17}$$

In the above Eq. 17, $N$ represents the size of batch, $M$ represents the number of classes, $y_{ic}$ is a sign function, taking the value of 0 or 1, if the true class of the sample is equal to 1, 0 otherwise, $p_{ic}$

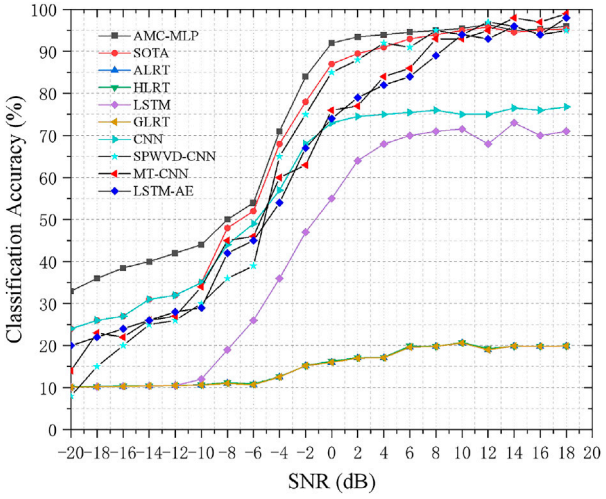



FIGURE 5
Classification accuracy of our proposed model vs existing models on RadioML 2016.10A dataset.

is the predicted probability that the observed sample $i$ belongs to category $c$ type.

## Simulations and discussions

In this section, we test the performance of the model. We use the public dataset to construct the feature image dataset, and further conduct parameter adjustment and performance test of the algorithm to obtain the final test results. The model is further compared with the existing algorithms, and finally the experimental results are compared and analyzed.

### Description of experimental dataset

We use the RadioML 2016.10A public dataset for model performance measurement. It includes 11 modulation types. The software defined radio is used for I/Q dual-channel sampling, and the length of a single data is 128. The SNR level of the signal ranges from −20 dB to 18 dB, where the step size is 2 dB, and there are 220,000 samples in total. The noise added by the channel is white Gaussian noise, and the specific data are shown in the following Table 2.

### Experimental procedure

First, the open dataset of RadioML 2016.10A is read, and eight kinds of digital signals are selected: 8BPSK, QPSK, BPSK, GFSK, CPFSK, PAM4, QAM16, and QAM64 were classified in different SNR environments. Contour diagrams were generated using the baseband data of I/Q, SPWVD, and the SCF, based on 20 different SNR samples, ranging from −20 dB to 18 dB, each with 128 samples length. Each signal has 1000 samples and a total of 160,000 I/Q sampling data. By generating three different feature maps, the total number of samples is 480,000. We use 70% of the samples as the training set, and the remaining samples as the test set. After the data set was made, three different feature maps were input into Conv-AE for feature extraction, and the three Conv-AE shared parameters. The optimal feature vector was found through the input and reconstructed loss function, and the following step was input into attention-MLP for classification. The classification loss was calculated, and the parameters were optimized by back propagation. The optimization of the whole model was achieved by the overall optimization of Conv-AE and attention-MLP. The dropout rate of the fully connected layer is set to 0.2, and the hyperparameter $\mu_1$ is set to 0.1 under supervised conditions. The training data set epochs is 128, and the learning rate is 0.001. We used 70%, 20% and 10% of the dataset for training, validation and testing, with an Adam Optimizer applied.

### Model testing

We used the RadioML 2016.10A dataset to compare and verify the algorithms mentioned in the reference ALRT (Hong and Ho, 2003), GLRT (Panagiotou et al., 2000), HLRT (Hong and Ho, 2003), CNN (O'Shea et al., 2016), MT-CNN (Qiao et al., 2022), SPWVD -CNN (Hou et al., 2021), LSTM-AE (Ke and Vikalo, 2022), LSTM (Rajendran et al., 2018) and SOTA (Zhang et al., 2021). The precision curves of various algorithms are shown in Figure 5.

Figure 5 shows the modulation classification of several likelihood functions. Due to the uncertain calculation caused by too many unknown parameters and modulation types, GLRT, HLRT and ALRT, the three likelihood modulation classification methods, cannot perform good modulation classification even when 10 dB–18 dB. When the SNR ranges from −20 dB to 10 dB, the modulation classification can hardly be carried out. The LSTM modulation classification method, whose algorithm only focuses on the relationship between one-way data, cannot identify the features of high capture dimension, and the single feature faces difficulty to identify the spatial representation of the approximate modulation mode, especially when the SNR is less than −10 dB, showing almost no difference in the classification accuracy with the likelihood estimation method. Due to the innate local feature extraction, the CNN modulation classification method cannot extract global features. Compared with the LSTM method, it exhibits great improvement when SNR is −10 dB and below. Moreover, when SNR ranges from −10 dB to 2 dB, the classification accuracy is gradually improved with the increase in SNR. When SNR is more than 2 dB, the classification accuracy does not improve significantly.

The classification accuracy of MT-CNN and LSTM-AE has little difference within the SNR from −20 dB to 10 dB range, but

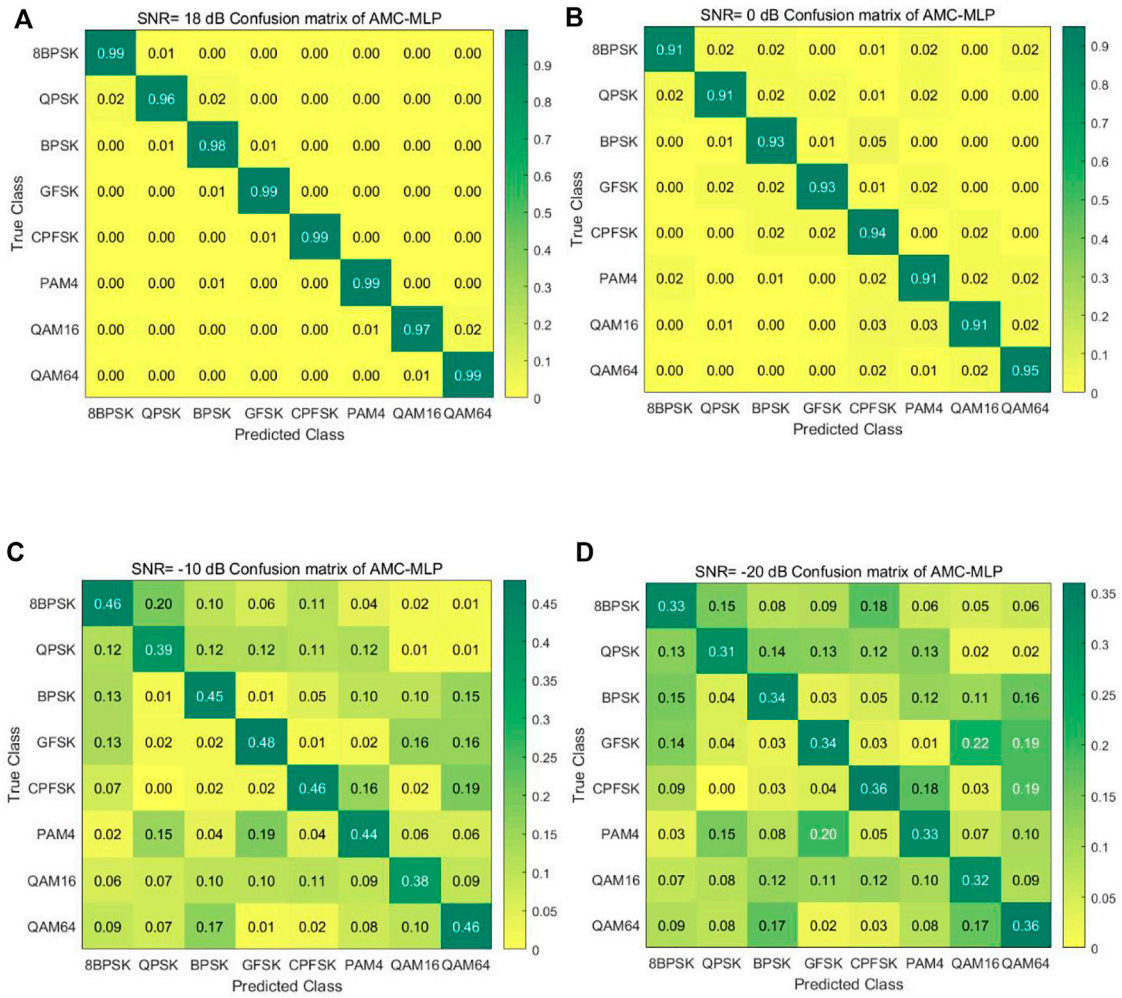



FIGURE 6
Accuracy confusion matrix of AMC-MLP algorithm with different SNR. **(A)** SNR = 18 dB, **(B)** SNR = 0 dB, **(C)** SNR = −10 dB, **(D)** SNR = −20 dB.

LSTM-AE is more compact than MT-CNN's model, with fewer parameters and lower computational complexity, which is conducive to deployment on low-cost platforms. The classification accuracy of SPWVD-CNN model has little difference with other models in the SNR from−20 dB to −4 dB range, but it is significantly higher than the MT-CNN model in the SNR from −4 dB to 18 dB range, but lower than AMC-MLP. The essential reason is that the representation ability of a single feature is insufficient.

The SOTA method uses a 2-D section graph of SCF for modulation and generation, and employs CNN for noise reduction. When deep neural networks is use for classification, it shows a considerably high classification accuracy. Especially when SNR is above −6 dB, the statistical characteristics of SCF itself show evident classification accuracy when fighting noise interference. After the CNN noise reduction, the features are extremely evident, and the overall energy absorption is relatively excellent. However, this also demonstrates the single use of features, which still has evident shortcomings at low SNR.

Conv-AE has strong data reconstruction and feature extraction abilities. We use the approximation between input and output to compare the reconstruction ability of AE, and use the middle low-dimensional feature vector to represent the features of the original image. The higher the approximation between input and output, the stronger the coding ability of AE. This experiment also showed that CNN in Conv-AE has a strong feature grabbing ability, which has a natural advantage compared with other AE.

This experiment also proves that MLP has the same capability as the transformer, which can not only be used for computer vision but also for radio frequency signal classification after optimization (Figure 5). This is because using the low-dimensional feature vectors extracted by AE as the input of MLP can significantly reduce the network scale, training time, and future inference time on the edge compared with directly using MLP. The setting of the loss rate of the fully connected layer further

reduces the scale of the network and preliminarily realizes the compression of the network model.

## Analysis of results

In this experiment, as Figure 6 shows, we compared with the latest AMC algorithm (Zhang et al., 2021) to study the overall classification accuracy variation trend of the model composed of Conv-AE and attention-MLP under the supervised condition of different SNR environments in the range of −20 dB–18 dB. AMC-MLP of the overall classification accuracy is higher than the existing classification method, especially in −20 dB–8 dB range, has obvious advantages, highlight the model of classification accuracy in low SNR environment. Through the classification confusion matrix of AMC-MLP under different SNR, the classification advantages of the new model can be clearly seen. In the SNR environment of 0 dB–18 dB, the AMC-MLP model can maintain good classification accuracy and robustness. In the SNR environment of −20 dB–0 dB, the classification accuracy of AMC-MLP model is greatly reduced by noise interference, but it has been greatly improved compared with SOTA model. The reason is that using the use of a variety of characteristics of attention ability greatly improve obviously against noise, better solve the complex electromagnetic environment in power energy system environment, AMC-MLP meets the requirements of fast and diverse modulation classification methods for base stations. It is very suitable for deployment on Xilinx Zynq UltraScale+™MPSoC. The multi-core architecture has significant advantages. The FPGA core uses a two-stage pipeline for baseband I/Q data sampling and feature map conversion. Then, the internal bus is used to transfer the low-dimensional feature data to Mali-400MP2 GPU for MLP acceleration, and the internal quad-core ARM is used to manage the model.

## Conclusion and future work

We used the AMC method combining Conv-AE and attention-MLP. We employed Conv-AE for low-dimensional feature extraction of multi-feature maps and attention-MLP for AMC classification under attention. The method was verified by experiments and compared with the traditional AMC method. Under the condition of high SNR, AMC-MLP can not only obtain better classification performance, but also obtain higher classification accuracy under the condition of low SNR. The model has simple structure, few parameters, high robustness, and can maintain high classification accuracy and real-time performance when reasoning, which meets the requirements of power wireless Internet of things.

The following are suggestions for future researches. First, novel modulation type classification methods, such as orthogonal frequency division multiplexing, should be investigated to improve the generalization and robustness of the model. Second, hard and soft compute of resources in IoT systems are limited; hence, it is necessary to focus on lightweight and low-power classification methods used in IoT terminals. Finally, most existing algorithms are trained based on supervised learning, which requires a large amount of labeled data as the basis. Therefore, it is necessary to propose semi-supervised or few-shot samples modulation classification methods in AMC research.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.deepsig.ai/datasets.

## Author contributions

JH completed the theoretical research, technical method design, experiment and analysis, and the writing and revision of the paper. ZY and JY conducted research guidance, technical method demonstration, and writing and review of the paper.

## Funding

This project was supported by grants from the National Natural Science Foundation of China (62071481).

## Acknowledgments

The authors would like to thank all of the people who participated in the studies.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References

Abdel-Moneim, M. A., El-Shafai, W., Abdel-Salam, N., El-Rabaie, E. S. M., and El-Samie, F. E. A. (2021). A survey of traditional and advanced automatic modulation classification techniques, challenges, and some novel trends. *Int. J. Commun. Syst.* 34, e4762. doi:10.1002/dac.4762

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Golovanevsky, M., Eickhoff, C., and Singh, R. (2022). Multimodal attention-based deep learning for alzheimer's disease diagnosis. Priprint arXiv:220608826G.

Hong, L., and Ho, K. C. (2003). "Classification of BPSK and QPSK signals with unknown signal level using the Bayes technique," in Proceedings of the 2003 IEEE International Symposium on Circuits and Systems (ISCAS), Bangkok, Thailand, 25-28 May, 2003, IV.

Hong, L., and Ho, K. C. (2001). "Modulation classification of BPSK and QPSK signals using a two element antenna array receiver," in Proceedings of the 2001 MILCOM Proceedings Communications for Network-Centric Operations: Creating the Information Force (Cat. No.01CH37277), McLean, VA, USA, 28-31 Oct. 2001, 118–122.

Hou, C., Li, Y., Chen, X., and Zhang, J. (2021). Automatic modulation classification using KELM with joint features of CNN and LBP. *Phys. Commun.* 45, 101259. doi:10.1016/j.phycom.2020.101259

Huan, C. Y., and Polydoros, A. (1995). Likelihood methods for MPSK modulation classification. *IEEE Trans. Commun.* 43, 1493–1504. doi:10.1109/26.380199

Ke, Z., and Vikalo, H. (2022). Real-time radio technology and modulation classification via an LSTM auto-encoder. *IEEE Trans. Wirel. Commun.* 21, 370–382. doi:10.1109/twc.2021.3095855

Liu, H., Dai, Z., So, D. R., and Le, Q. V. (2021). Pay attention to MLPs. Priprint arXiv:2105.08050.

Mao, Y., Dong, Y. Y., Sun, T., Rao, X., and Dong, C. X. (2021). Attentive siamese networks for automatic modulation classification based on multitiming constellation diagrams. *IEEE Trans. Neural Netw. Learn. Syst.* 46, 1–15. doi:10.1109/TNNLS.2021.3132341

Mirarab, M. R., and Sobhani, M. A. (2007). "Robust modulation classification for PSK/QAM/ASK using higher-order cumulants," in Proceedings of the 2007 6th International Conference on Information, Communications & Signal Processing, Singapore, 10-13 Dec. 2007, 1–4.

Mobasseri, B. G. (2000). Digital modulation classification using constellation shape. *Signal Process.* 80, 251–277. doi:10.1016/s0165-1684(99)00127-9

Nandi, A. K., and Azzouz, E. E. (1998). Algorithms for automatic modulation recognition of communication signals. *IEEE Trans. Commun.* 46, 431–436. doi:10.1109/26.664294

Orlic, V. D., and Dukic, M. L. (2009). Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions. *IEEE Commun. Lett.* 13, 917–919. doi:10.1109/LCOMM.2009.12.091711

O'Shea, T. J., Corgan, J., and Charles, C. T. (2016). Convolutional radio modulation classification networks. Priprint arXiv:1602.04105.

O'Shea, T. J., and West, N. (2016). Radio machine learning dataset generation with GNU radio. *Proc. GNU Radio Conf.* 1, 1–6.

Panagiotou, P., Anastasopoulos, A., and Polydoros, A. (2000). "Likelihood ratio tests for modulation classification," in Proceedings of the MILCOM 2000 Proceedings. 21st Century Military Communications. Architectures and Technologies for Information Superiority (Cat. No.00CH37155), Los Angeles, CA, USA, 22-25 Oct. 2007, 670–674.

Qiao, J., Chen, W., Chen, J., and Ai, B. (2022). Blind modulation classification under uncertain noise conditions: A multitask learning approach. *IEEE Commun. Lett.* 26, 1027–1031. doi:10.1109/LCOMM.2022.3149284

Rajendran, S., Meert, W., Giustiniano, D., Lenders, V., and Pollin, S. (2018). Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans. Cogn. Commun. Netw.* 4, 433–445. doi:10.1109/TCCN.2018.2835460

Roberts, R. S., Brown, W. A., and Loomis, H. H. (1991). Computationally efficient algorithms for cyclic spectral analysis. *IEEE Signal Process. Mag.* 8, 38–49. doi:10.1109/79.81008

Satija, U., Mohanty, M., and Ramkumar, B. (2015). "Automatic modulation classification using S-transform based features," in Proceedings of the 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 19-20 Feb. 2015, 708–712.

Shen, W. G., and Gao, Q. X. (2014). "Automatic digital modulation recognition based on locality preserved projection," in Proceedings of the 2014 International Conference on Wireless Communication and Sensor Network, Wuhan, China, 13-14 Dec. 2014, 348–352.

Tan, H., and Bansal, M. (2019). Lxmert: Learning cross modality encoder representations from transformers. Priprint arXiv:1908.07490.

Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., et al. (2021). MLP-Mixer: An all MLP architecture for vision. Priprint arXiv:2105.01601.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. Priprint arXiv:1706.03762.

Wen, W., and Mendel, J. M. (2000). Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Trans. Commun.* 48, 189–193. doi:10.1109/26.823550

Xu, T., and Darwazeh, I. (2020). "Deep learning for over the air NonOrthogonal signal classification," in Proceedings of the 2020 IEEE 91st Vehicular Technology Conference, Antwerp, Belgium, 25-28 May 2020, 1–5. doi:10.1109/VTC2020Spring48590.2020.9128869

Yu, Z., Shi, Y. Q., and Su, W. (2003). "M-ary frequency shift keying signal classification based-on discrete Fourier transform," in Proceedings of the 2003 IEEE Conference on Military Communications - Volume II, Boston, MA, 13-16 Oct. 2003, 1167–1172.

Zhang, L., Liu, H., Yang, X., Jiang, Y., and Wu, Z. (2021). Intelligent denoising-aided deep learning modulation recognition with cyclic spectrum features for higher accuracy. *IEEE Trans. Aerosp. Electron. Syst.* 57, 3749–3757. doi:10.1109/TAES.2021.3083406

Zhou, L., Sun, Z., and Wang, W. (2017). Learning to short-time Fourier transform in spectrum sensing. *Phys. Commun.* 25, 420–425. doi:10.1016/j.phycom.2017.08.007

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 10 January 2023
DOI 10.3389/fenrg.2022.1071948



OPEN ACCESS

EDITED BY
Hanlin Zhang,
Qingdao University, China

REVIEWED BY
Guobin Xu,
Morgan State University, United States
Youcef Belkhier,
Maynooth University, Ireland

*CORRESPONDENCE
Qingyu Yang,
yangqingyu@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 17 October 2022
ACCEPTED 07 December 2022
PUBLISHED 10 January 2023

CITATION
Li D, Yang Q, Ma L, Wang Y, Zhang Y and Liao X (2023), An electrical vehicle-assisted demand response management system: A reinforcement learning method.
*Front. Energy Res.* 10:1071948.
doi: 10.3389/fenrg.2022.1071948



# An electrical vehicle-assisted demand response management system: A reinforcement learning method

Donghe Li[1], Qingyu Yang[1,2]*, Linyue Ma[3], Yiran Wang[1], Yang Zhang[1] and Xiao Liao[3]

[1]School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China, [2]State Key Laboratory Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China, [3]State Grid Information & Telecommunication Group Co.,LTD., Beijing, China

With the continuous progress of urbanization, determining the charging and discharging strategy for randomly parked electric vehicles to help the peak load shifting without affecting users' travel is a key problem. This paper design a reinforcement learning-based method for the electric vehicle-assisted demand response management system. Specifically, we formalize the charging and discharging sequential decision problem of the parking lot into the Markov process, in which the state space is composed of the state of parking spaces, electric vehicles, and the total load. The charging and discharging decision of each parking space acts as the action space. The reward comprises the penalty term that guarantees the user's travel and the sliding average value of the load representing peak load shifting. After that, we use a Deep Q-Network (DQN)-based reinforcement learning architecture to solve this problem. Finally, we conduct a comprehensive evaluation with real-world power usage data. The results show that our proposed method will reduce the peak load by 10% without affecting the travel plan of all electric vehicles. Compared with random charging and discharging scenarios, we have better performance in terms of state-of-charge (SoC) achievement rate and peak load shifting effect.

KEYWORDS

**demand response, electric vehicle, reinforcement learning method, MDP, DQN**

# 1 Introduction

With the continuous progress of urbanization, the population is constantly funneled into large- and medium-sized cities, and it brings great power load pressure to most cities (Lin and Zhu, 2020). Considering the fact that the speed of power infrastructure construction cannot keep up with the speed of load growth, time-sharing power supply is the most commonly used method to deal with the electricity outage. For example, from 2020 to the present, major cities in China have experienced limited electricity consumption due to excessive load, such as Guangzhou, Shenyang, Xi'an, and Chengdu. Evidently, this coercive method will seriously affect people's daily life. Therefore, when the power infrastructure cannot be rapidly deployed to improve the power generation/supply capacity, it becomes urgent to find a more effective way to relieve the power consumption pressure.

Smart grid (Fang et al., 2012; Gungor et al., 2011), which supports bi-directional information and energy transformation, can attract users to adjust their electricity consumption habits and actively participate in the dispatch of the power grid, that is, demand response (Medina et al., 2010; Palensky and Dietrich, 2011). Therefore, the wide application of smart grid technology can ensure the safe, reliable, and efficient operation of the power grid by introducing distributed energy storage equipment and distributed power users into the demand response process when the power grid load supply cannot be rapidly increased. Great research effort has been devoted to leveraging energy storage equipment to assist demand response management (Cui et al., 2017; Tang et al., 2019). For instance, it has to be mentioned that the unrestricted deployment of electrical energy storage devices will bring great economic burden, which is impractical.

Recently, with the increasingly severe global energy shortage and environmental pollution, the automobile industry has been undergoing major changes, and electric vehicles (EVs) have become a new direction for the development of various automobile companies (Emadi et al., 2008; Lopes et al., 2011). The EV industry has been developing rapidly in recent years, with the total value of the global EV market growing from $18 billion in 2018 to $22.42 billion in 2019 and an annual growth rate of more than 7.5%. Statistical studies show that most electric vehicles are in shutdown state 90% of the time, during which the on-board battery of electric vehicles can be regarded as a distributed energy storage device to participate in the demand side management of the microgrid, which is called V2G technology (vehicle-to-grid) (Madawala and Thrimawithana, 2011; Ota et al., 2012).

Electric vehicles have been widely used as a load-balancing tool in academic circles because of their good power storage capacity and flexibility. However, due to the uncertainty of vehicle owners' commuting behavior, electricity power demand and load, etc., reasonably planning the charging and discharging strategy of electric vehicles to help the power grid carry out peak load filling and demand response is a key problem. Scholars have made great efforts in designing an EV charging/discharging strategy, which can be mainly divided into two categories: optimization scheduling (Karapetyan et al., 2021; Zhang et al., 2022) and trading-based method (Li et al., 2019; Yang et al., 2020). However, these two mainstream approaches have their drawbacks. The scheduling-based method is suitable for the offline environment, that is, decision-makers need to obtain some prior knowledge, such as the EV owners' travel plan and the future electricity supply/demand. At the same time, this method will ignore the needs of EV owners in pursuit of higher power conversion efficiency. Regarding the trading-based method, it can be used in an online environment and fully meets the needs of EV owners, but it has limited help for demand response because it is a completely free market with limited incentives for EV owners.

As introduced earlier, it is necessary to find a novel optimization algorithm to satisfy the aforementioned requirements (online, demand response, and EV owners' travel plan and enthusiasm). Reinforcement learning (RL) (Kaelbling et al., 1996) is a new artificial intelligence method to obtain optimal strategies for sequential decision problems. In the energy trading market, each participant can be regarded as an agent, while the trading market is formed as a multi-agent cooperation model (Wu et al., 2007). In such a multi-agent model, the purpose of each agent is to improve its own utility and meet its own needs, which causes great difficulty in constructing and solving the decision-making model of such a multi-agent-based energy trading market. Since RL can formulate effective coordination strategies for the agent without explicitly building a complete decision model, it can adapt the agent's behavior to the uncertain and changing dynamic environment and improve the agent's performance through interaction. Thus, RL can often achieve good results in the scenario of multi-agent cooperation, such as energy trading, which has aroused extensive research by scholars (Liu et al., 2017; Hua et al., 2019). Nowadays, there exist many RL-based energy management methods; for example, Qian et al. (2020) proposed a reinforcement learning-based EV charging strategy focusing on the intelligent transportation system, and it can minimize the total travel distance and charging cost. Zhang et al. (2022) proposed a multi-agent reinforcement learning method to make an optimal energy purchase schedule for charging stations and a long short-term memory (LSTM) neural network to predict the EV's charging demand. Although the existing research studies are focused on power scheduling, different scenarios have different problems, and the existing reinforcement learning method cannot be applied to the EV-assisted demand response scenario studied in our study. Specifically, in this scenario, there are electric vehicles with uncertain quantities and uncertain charging/discharging

requirements. This scenario is evidently a multi-agent scenario, considering that the policy trained by multi-agent reinforcement learning is only for one agent. However, in this scenario, EV entry and exit are not restricted, and the strategy for an agent will not be practical after the EV leaves. So in the EV-assisted demand response system with uncertain EVs, determining the strategy of EVs' charging/discharging behavior is a challenge.

To this end, in this paper, we will study an EV-assisted demand response management system to relieve the power consumption pressure in urban peak hours by planning EV charging/discharging behaviors. Considering the efficiency of decision-making, we aim to design a reinforcement learning method for an EV-assisted demand response management system. The main contribution of this paper is as follows: we first formalize the EV charging/discharging strategy as an MDP model. Electric vehicles can enter and exit at any time; we focus on making decisions for parking spaces, and the action space is the charging and discharging strategy of each parking space. Considering that too many electric vehicles lead to too much action space, we classify electric vehicles, and similar electric vehicles share one action. The state space includes the state of parking spaces, EVs, and total demand. Because our system is a multi-objective optimization problem, we use a penalty item to ensure that the departure SoC will be enough for the next travel, and we use a moving average reward to ensure the peaking load shifting effect. After that, we design a DQN reinforcement learning architecture to solve the MDP model. Finally, comprehensive evaluations are conducted with real-world data to verify the effectiveness of our method.

The remainder of this paper is organized as follows: in **Section 2**, we briefly review the research efforts related to EV charging/discharging strategy and reinforcement learning method. In **Section 3**, we introduce the models of our EV-assisted demand response management system and build the EV charging and discharging scheduling optimization model. In **Section 4**, the background of deep reinforcement learning is proposed, and the MDP process of the EV charging/discharging behavior is modeled. In **Section 5**, the DQN reinforcement learning method is introduced, and we propose a DQN-based EV charging/discharging strategy algorithm. In **Section 6**, we evaluate the performances of the proposed method and compare our method with other methods, concluding the effectiveness of the proposed method in peak load shifting. Finally, we conclude this paper in **Section 7**.

## 2 Related work

With the rapid growth of electric users, the imbalance between power supply and demand in the power grid becomes more prominent. Many scholars have focused on alleviating the imbalance of power supply through demand response technology without increasing power infrastructure (Althaher et al., 2015; Eksin et al., 2015; Wang et al., 2018). For example, Jeddi et al. (2021) proposed a coordinated load scheduling method for each home customer in order to optimize their energy consumption at the neighborhood level. Under such a load scheduling method, the home customers will be rewarded, and demand response will be implemented. Similarly, facing the residential demand response problem, Liu et al. (2019) proposed an energy trading method based on game theory, in which the householder with renewable resources will transfer energy through a peer-to-peer (P2P) trading market. It can be seen that the demand response mechanism has been widely used in electric energy scheduling, especially in residential areas.

Moreover, electric vehicles (EVs) equipped with large capacity batteries can be used as distributed energy storage devices to participate in demand response. Therefore, scholars have also carried out research on demand response methods involving EVs by vehicle-to-grid technology. Kikusato et al. (2019) proposed an EV charge–discharge management framework, in which the home energy management system (HEMS) decides the EV charge–discharge plan with information from the grid energy management system, aiming to reduce the residential operation cost. Li et al. (2020) proposed an auction market that allows electric vehicles with surplus energy to sell their energy to those with insufficient energy. (Li et al., 2019). Yang et al. (2020) proposed the auction-based EV energy trading market, in which EVs with insufficient energy act as buyers and EVs with surplus energy act as sellers. This mechanism can help peak-load shifting to a certain extent. Thus, using EVs as energy storage equipment for demand response has become a new direction in the academic world. However, the traditional scheduling and transaction methods are not enough in terms of efficiency and user satisfaction for flexible energy storage devices such as EVs.

Reinforcement learning, as an optimal strategy solution, has been widely used in energy scheduling and trading. Compared with previous non-artificial intelligence scheduling or trading methods, the RL method has a good effect in coping with environmental changes, so it has a better performance in the scenario involving EVs. For example, Wan et al. (2019) proposed a model-free approach to determine the real-time optimal schedules based on deep reinforcement learning. The approach contains a representation network to extract discriminative features from the electricity prices and a Q network to approximate the optimal action-value function. Zhang et al. (2020) proposed a DQN-based method to manage the EV charging behavior in order to improve the income of EV owners and to reduce the pressure on the power grid as much as possible.

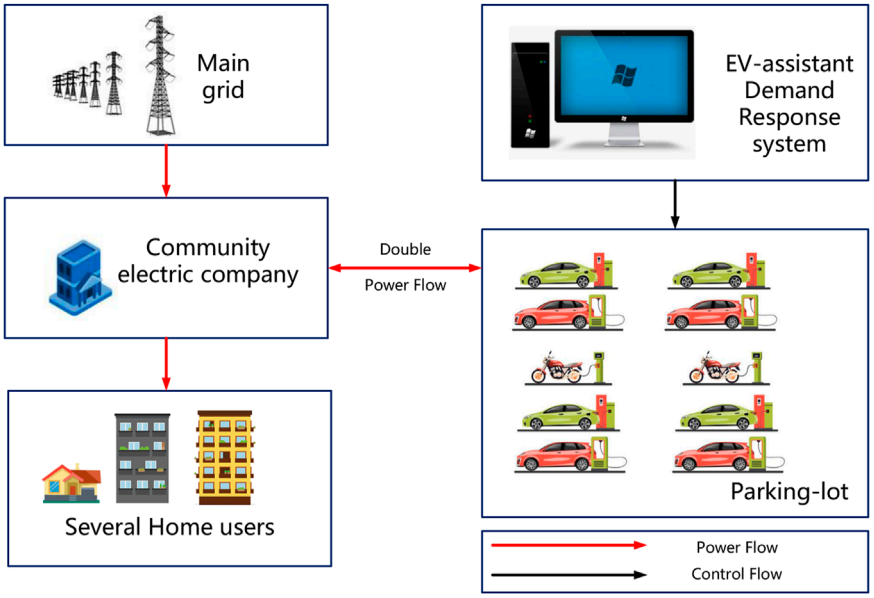



FIGURE 1
Market structure.

## 3 EV-assisted demand response management system

In this section, we will introduce the EV-assisted demand response management system in detail. First, we will give the system model, and then we will introduce the EVs' model and the optimization formulation.

### 3.1 System model

The proposed EV-assisted demand response management system will be considered to be deployed in a small area which often has a lot of electricity users, such as a residential area and shopping mall. In such an area, there exists a parking lot for electric vehicles (EVs) to charge and discharge. Notably, EVs with high state-of-charge (SoC) can be regarded as the energy storage unit and supply electric power to the electric users within this area *via* V2G technology. So, the V2G EV-assisted demand response management system will greatly help to reduce the load pressure on the grid and enable more power users to use electricity. Meanwhile, since the position between the EVs and the electricity user is very close, the power transfer will not undergo multiple voltage changes, so the power loss is assumed to be 0. The system model is shown in Figure 1, and the important notations are shown in Table 1. The symbolic expression and some assumptions are as follows:

The EVs are saved into set $P$, and the $m$ parking spaces are saved into set $G$. Furthermore, the EVs with insufficient energy are saved into set $P_b$, and the EVs with extra energy are saved into set $P_s$. The time is slotted and is denoted by an integer set $T = [1, 2, 3, \dots]$, and 1 day is divided into 24 slots. For an EV $i \in P$, when he/she enters the parking lot, he/she will upload some status and requirements information to the platform, including the arrival and departure time, arrival SoC, and departure SoC. The arrival and departure times are denoted as $ta_i$ and $td_i$, respectively. The arrival SoC is denoted as $SoC_i^a$ and represents the state-of-charge when the EV $i$ arrives into the system at $ta_i$. The departure SoC is denoted as $SoC_i^d$ and represents the minimum state-of-charge when the EV $i$ departs at $td_i$. Notably, the departure SoC is determined by their travel plans. The management system is responsible for making decisions about the charging and discharging behavior of EVs. In each time slot $t$, the platform will make a decision to determine the charging and discharging behavior of each EV in this time slot and charge or pay according to the real-time electricity bill at that time. The system makes full use of the capability of EVs for scheduling while ensuring the departure SOC.

### 3.2 Electric vehicle model

As introduced before, in our paper, EVs act as both load and supply. Generally speaking, an electric vehicle can be considered as a mobile chemical energy storage unit. When it is parked in the parking lot and connected to the EV-assisted demand response management system, it is no different from the conventional chemical energy storage unit. However, electric vehicles need

**TABLE 1** Key notations.

| Symbols | Descriptions |
|---|---|
| $P, P_b, P_s$ | Set of EVs, EVs with insufficient energy, and EVs with extra energy |
| $T, G$ | Sets of time slots and parking spaces |
| $m$ | Number of parking spaces |
| $ta_i, td_i$ | Arrival and departure time of EV $i$ |
| $SoC_i^a, SoC_i^d$ | Arrival and departure SoC of EV $i$ |
| $SoC_i^t$ | SoC of EV $i$ at time slot $t$ |
| $q_i$ | Charging/discharging speed of EV $i$ |
| $C_i$ | Battery capacity of EV $i$ |
| $x_{ij}^t$ | The connecting status between EV $i$ and EV $j$ at time slot $t$ |
| $s^t$ | The system state at time slot $t$ |
| $l_i^t$ | The state of parking space $i$ at time slot $t$ |
| $Q^t$ | The state of total demand at time slot $t$ |
| $a^t$ | The system action at time slot $t$ |
| $a_i^t$ | The action of parking space $i$ at time slot $t$ |
| $r^t$ | The system reward at time slot $t$ |
| $r_i^t$ | The penalty term of parking space $i$ at time slot $t$ |
| $r_{load}^t$ | The reward of peak load shifting at time slot $t$ |

to assume the responsibility of vehicles and cannot always be parked in the parking lot as a power dispatching tool. So, the EVs can only be dispatched to discharge during their parking time, especially for the EVs that need to be charged. Specifically, the EV $i$'s parking time period is denoted as $T_i = [ta_i, td_i]$. Notably, the EV must be parked in a parking space, so $i$ also denotes the ID of parking space $i \in G$. At the same time, different models of electric vehicles also have differences in battery capacity and charging and discharging speed. We use the symbols $C_i$ and $q_i$ to represent the battery capacity and charging/discharging speed of EV $i$. Notably, for convenience of calculation, we assume that the charging/discharging speed ($q_i$) of EV $i$ is determined by the parking pile that they park. Then, we construct the following constraints to model electric vehicles:

$$SoC_i^t = SoC_i^{t-1} + \frac{q_i}{C_i}, t \in T_i = [ta_i, td_i], \tag{1}$$

where $SoC_i^t$ represents the SoC of EV $i$ at time slot $t$. This equation reflects the linear transfer formula of battery state when energy loss is ignored.

$$SoC_i^{min} \le SoC_t^i \le SoC_i^{max}, t \in T_i = [ta_i, td_i], \tag{2}$$

where $SoC_i^{min}$ and $SoC_i^{max}$ represent the lower and upper bounds of SoC. This equation constrains the state of the battery according to its physical properties so that it can maintain a better performance.

$$q_i = 0, t \notin T_i. \tag{3}$$

This equation means that charging and discharging scheduling cannot be carried out when electric vehicles are not in the parking lot.

## 3.3 EV charging/discharging scheduling problem formulation

In the traditional demand response management system, the EV charging/discharging scheduling problem is always formulated as an optimization problem. The system collects the information of all EVs in the future for a period of time to make a comprehensive charging and discharging decision so that a certain index can reach the optimum. In our paper, we consider that the optimization goal is maximization of peak load shifting, that is, to help the power grid to cut peak and fill valley as much as possible. Then, the EV charging/discharging scheduling problem can be expressed as a mixed-integer linear program (MILP) model as follows:

$$max \sum_{t=1}^{t_n} \| \sum_{i=1}^{m} x_i^t SoC_i^t - \frac{\sum_{t=1}^{t_n} Q^t}{t_n} \|, \tag{4}$$

subject to:

$$SoC_i^t = SoC_i^{t-1} + \frac{q_i}{C_i}, t \in [ta_i + 1, ta_i], i \in G, \tag{5}$$

$$x_i^t \left( SoC_i^{t-1} - SoC_i^t \right) = q_i, t \in [ta_i + 1, ta_i], i \in G, \tag{6}$$

$$SoC_i^{td_i} \ge SoC_i^d, i \in G, \tag{7}$$

$$SoC_i^{min} \le SoC_t^i \le SoC_i^{max}, t \in T_i = [ta_i, td_i], i \in P, \tag{8}$$

$$x_i^t = 0, 1, -1 i \in G, \tag{9}$$

where $x_i^t$ is the optimization parameters, and it represents the charging/discharging status of the EV which is parked at parking space $i$(we call it EV $i$ for convenience). When $x_i^t = 1, -1, 0$, they mean that at time slot $t$, EV $i$ will charge, discharge, or stop, respectively.

Here, we will briefly introduce the aforementioned optimization model. First, the objective function represents that we want to maximize the effect of peak load shifting for a period of time. Constrain 1 shows the calculation rules of power transmission and SoC. Constraint 2 specifies that the charging and discharging speed shall satisfy the physical limits of parking piles. Constraint 3 specifies that the SoC of all EVs will have

enough SoC when the EVs leave the parking lot. Constraint 4 specifies the constraint of optimization parameters.

To solve such an optimization equation, it is evident that an optimal charging and discharging strategy will be obtained, but considering many practical factors, it cannot be really used in practice. First, solving such a model requires the system to obtain accurate future information in advance, which is evidently impossible in practice. Second, even if the system obtains future information, the optimal matching strategy is obtained by solving the optimization problem. But once the future environment changes, the strategy will no longer be optimal. Therefore, in order to deal with this uncertain electric vehicle charging and discharging problem, it is necessary to design a method that can obtain the optimal strategy according to the current conditions.

# 4 MDP model of EV-assisted demand response strategy

In this section, we will introduce the Markov decision process (MDP) model of our proposed EV-assisted demand response system. First, we will introduce the rational of introducing the MDP model. Then, the background of deep reinforcement learning is proposed. Finally, the MDP model of charging/discharging strategy is given.

## 4.1 Rational

As introduced in **Section 3.3**, we have formulated the EV charging/discharging management as an MILP problem. However, this method has a high demand for future state and is too sensitive to environmental changes, which makes it unable to be deployed in actual scenarios. Reinforcement learning can obtain the best strategies in different environments through continuous exploration, and it has natural advantages in the face of such complex and changeable scenes.

But in our proposed EV-assisted demand response system, there still exists the following challenges: 1) there exist multiple EVs with different states and targets: different EVs have different SoC and different charging and discharging requirements, so it is necessary to learn different strategies for them. 2) EVs enter and exit the parking lot at any time; therefore, if each specific EV is given a learning strategy, the learned strategy will not be used after it leaves. 3) Considering that parking spaces are fixed, we can set strategies based on them. However, with the increase in the number of parking spaces, the dimension of action space is too large, which often leads to failure to learn useful strategies. To address the aforementioned problem, we divide EVs into several categories according to the SoC and charging and discharging requirements and provide learning strategies for each, respectively. In this way, we only need to classify the new EVs to get the related strategy.

## 4.2 Deep reinforcement learning

Deep reinforcement learning (DRL) combines the perceptual capability of deep learning (DL) with the decision-making capability of reinforcement learning (RL), where the agent perceives information through a higher dimensional space and applies the obtained information to make decisions for complex scenarios. Deep reinforcement learning is widely used because it can achieve direct control from original input to output through end-to-end learning. Existing research mainly classifies deep reinforcement learning algorithms into three main categories: one based on value functions, one based on policy gradients, and one based on multiple agents.

Mnih of DeepMind proposed Deep Q-Networks (DQNs) (Mnih et al., 2013), and people gradually started to study them at a deeper level while applying them to a wider range of fields. In recent years, research in deep reinforcement learning has focused on DQN, which combines convolutional neural networks with Q-learning and introduces an experience replay mechanism that allows algorithms to learn control policies by directly sensing high-dimensional inputs. The Deep Q-Network uses a Q-value function $Q(s,a,\theta)$ with parameters $\theta$ to approximate the value function. Under environment $\epsilon$, when the number of iterations is $i$, the definition of the loss function $L_i(\theta_i)$ is expressed as follows:

$$L_i(\theta_i) = E_{s,a\ \rho(.)}\left[\left(y_i - Q(s,a,\theta_i)\right)^2\right], \tag{10}$$

where $\rho(.)$ denotes the probability distribution of $s$ choosing action $a$ in a given environment, and $y_i$ denotes the objective of the $i$th iteration Q-value function, which is defined as follows:

$$y_i = E_{s'\ \epsilon}\left[r + \gamma max Q\left(s', a', \theta_{i-1} | s, a\right)\right], \tag{11}$$

where $r$ is the reward value fed to the agent by the environment, and $\gamma$ is the discount factor. The goal of learning depends on the network weights, and the update formula of network weights is

$$\nabla_{\theta_i} L_i(\theta_i) = E\left[\left(r + \gamma max Q\left(s', a', \theta_{i-1}\right)\right.\right. \\ \left.\left. - Q(s,a,\theta_i)\right)\nabla Q(s,a,\theta_i)\right]. \tag{12}$$

Although DQN based on the Q-learning algorithm has achieved good results in many fields, DQN is no longer applicable when facing continuous action space. Therefore, policy gradient methods have been introduced to deep reinforcement learning. Lillicrap et al. (2015) proposed the deep deterministic policy gradient (DDPG) algorithm in 2015. DDPG is an algorithm for deep reinforcement learning applied to continuous action space, which combines a deterministic policy gradient (DPG) algorithm with an actor–critic framework. In the DDPG, the

objective function is defined as the sum of the awards with discounts:

$$J(\theta^{\mu}) = E_{\theta^{\mu}}\left[r_1 + \gamma r_2 + \gamma^2 r_3 + \cdots\right]. \tag{13}$$

Then, the stochastic gradient descent method is used for end-to-end optimization of the objective function. Through a series of experiments, it is shown that DDPG not only performs stably in the continuous action space but is also much faster than DQN in terms of solution speed.

A multi-agent system (MAS) is a collection of multiple agents whose goal is to build complex systems into easily manageable systems. Multi-agent reinforcement learning (MARL) is the application of reinforcement learning ideas and algorithms to multi-agent systems. In the 1990s, Littman (1994) proposed MARL with a Markov decision process (MDP) as the environmental framework, which provided a template for solving most reinforcement learning problems. The environment of MARL is an MDP-based casuistic game framework with the following tuple:

$$< S, A_1, \ldots A_n, R_1, \ldots, R_n, P >, \tag{14}$$

where $n$ is the number of agents and $A$ is the set of joint action spaces of all agents:

$$A = A_1 \times \cdots \times A_n, \tag{15}$$

where $R_i$ is the reward function for each agent:

$$R_i{:}S \times A \times S \rightarrow R, \tag{16}$$

where $P$ is the state transfer function:

$$P{:}S \times A \times S \rightarrow [0,1]. \tag{17}$$

In the case of multiple agents, the state transfer is the result of all agents acting together, so the reward of the agents depends on the joint policy. The policy $H$ is defined as the joint policy of agents, and accordingly, the reward of each agent is

$$R_i^H = E\left[R_{t+1} | S_t = s, A_t = a, H\right]. \tag{18}$$

Its Bellman equation is

$$v_i^H(s) = E_i^H\left[R_{t+1} + \gamma V_i^H(S_{t+1}) | S_t = s\right], \tag{19}$$

$$Q_i^H(s,a) = E_i^H\left[R_{t+1} + \gamma Q_i^H(S_{t+1}, A_{t+1}) | S_t = s, A_t = a\right]. \tag{20}$$

Depending on the type of task, MARL can be classified as fully cooperative, fully competitive, or hybrid, using different algorithms for different problems.

## 4.3 MDP for EV-assisted demand response management

First, to address the EV-assisted demand response problem *via* reinforcement learning, the main goal is to design a central agent to achieve peak and valley filling in the area, taking into account the individual economic benefits of EVs. Considering that the charging and discharging behavior of EVs are implemented at regular intervals, so the EV-assisted demand response problem can be regarded as a sequential decision problem. Therefore, we introduce an MDP model with discrete time steps to establish the charging and discharging behavior of EVs in the EV-assisted demand response system. Briefly, the agent represents a community electric company, and it observes the environmental status $s_t$, including the electricity demand in the current region and the battery status of each electric vehicle. Then, the charging/discharging action $a_t$ is selected for the EVs, and the environment provides a corresponding reward $r_t$ for the replacement. After that, the aforementioned process will be repeated in the time series. Finally, we can obtain the best execution strategy $\pi^*$ by repeating the aforementioned process (training process) many times. Furthermore, the transition relationship between states is no longer internal but is determined by both states and actions. In the following, we will introduce the elements of the proposed Markov model in detail, including agent, state space, action space, observation space, transition, and reward.

In the model, the agent is the parking lot provider, and the responsibility of the agent can be expressed as follows: they give charging or discharging instructions to EVs in each parking space according to the state at each moment. It can help the power grid to cut peak load and fill valley load and, at the same, time try to satisfy the power demand of EVs.

We denote $S$ as the state space in our MDP model and $s^t$ is the state at time slot $t$. Specifically, $s^t$ can be expressed as

$$s^t = \left\{l_1^t, l_2^t, \ldots, l_m^t, Q^{t-1}\right\}, \tag{21}$$

where $l_i^t$ represents the state of parking space $i \in G$, $m$ represents the number of total parking spaces, and $Q^{t-1}$ represents the total demand of this area at time slot $t-1$. Furthermore, $l_i^t$ can be expressed as

$$l_i^t = \left\{speed_i, work, classify^t, td_i, SoC_i^t, SoC_i^d, C_i\right\}, \tag{22}$$

where $speed_i$ represents the charging/discharging speed of charging pile $i$. *work* represents the dispatch demand of charging pile $i$, and when there is no EV in this parking space or it does not need to be dispatched, the value is 0; otherwise, it is 1. $classify^t$, $td_i$, $SoC_i^t$, $SoC_i^d$, and $C_i$ represent the category, departure time, current SoC, departure SoC, and battery capacity of the EV in parking space $i$, and when there is no car in the parking space, these values are all 0.

We denote $A$ as the action space in our MDP model and $a^t$ as the action at time slot $t$. As introduced before, we will classify

EVs into five categories to help the agent to learn strategies more effectively and reduce the dimension of the action space:

- Case A: The EV *i*' SoC at time slot *t* is within the following range:

$$SoC_i^t - SoC_i^d \leq 5\%. \tag{23}$$

These EVs will not give action at this moment.

- Case B: The EV *i*' SoC at time slot *t* is within the following range, and the charging piles are the DC model (7 KW in our paper):

$$SoC_i^t - SoC_i^d > 5\%. \tag{24}$$

- Case C: The EV *i*' SoC at time slot *t* is within the following range, and the charging piles are AC model (30 KW in our paper):

$$SoC_i^t - SoC_i^d > 5\%. \tag{25}$$

- Case D: The EV *i*' SoC at time slot *t* is within the following range, and the charging piles are DC model (7 KW in our paper):

$$SoC_i^t < SoC_i^d. \tag{26}$$

- Case E: The EV *i*' SoC at time slot *t* is within the following range, and the charging piles are AC model (30 KW in our paper):

$$SoC_i^t < SoC_i^d. \tag{27}$$

Specifically, $a^t$ can be expressed as

$$a^t = \{a_1^t, a_2^t, a_3^t, a_4^t\}, \tag{28}$$

where $m$ represents the number of charging piles. $a_1^t$ to $a_4^t$ represents the action of the aforementioned categories from Case B to E at time slot *t*. When the value is 0, it means that the EV in the parking space will not be charged or discharged; when the value is 1, it means that the EV in the parking space will be charged; and when the value is −1, it means that the EV in the parking space will be discharged.

Considering the state $s^t$ and action $a^t$, at time slot $t+1$, each parking space will update its status according to the charging/discharging decision at the last time slot and read the new status if a new EV enters.

Finally, considering the two goals of peak load reduction and satisfying the SoC demand of EV as much as possible, the reward will consist of two parts: 1) the reward represents the peak shifting and 2) the penalty item represents the SoC demand of EV. Specifically, the reward space $R$ can be expressed as

$$R^t = \left\{r_1^t, r_2^t, \ldots, r_m^t, r_{load}^t\right\}, \tag{29}$$

where $r_i^t$ represents the penalty item which is calculation at each time slot.

$$r_i^t = \begin{cases} 0 & SoC_i^t \geq SoC_i^d \\ -1 & SoC_i^t < SoC_i^d \end{cases} \tag{30}$$

While $r_{load}$ represents the reward of peak shifting, and we express it in the form of moving average:

$$r_{load}^t = 1 - \| \frac{Ave_{power}^t - Q^t}{Ave_{power}^t} \|, \tag{31}$$

where $Ave_{power}^t$ represents the total power of the last $o$ time slot before time slot *t*:

$$Ave_{power}^t = \frac{Q^{t-o+1} + \cdots + Q^t}{o}. \tag{32}$$

Notably, in our paper, $o$ is set as 4.

Overall, the total reward at time slot *t* can be expressed as

$$r^t = \frac{1}{m}\sum_{i=1}^{m} r_i^t + r_{load}^t. \tag{33}$$

# 5 Solutions *via* deep reinforcement learning

In this section, we design a value-based reinforcement learning method to adaptively learn the policy of the agent, which can obtain the algorithm performance while effectively lightening the attack success rate. The diagram of the proposed method is illustrated in **Figure 2**.

## 5.1 Network structure

Two neural networks are introduced for different objectives in this paper: 1) a value evaluation network $Q(s^t, a^t; \theta)$ for evaluating the performance of employed action policy under state given and 2) a target network $Q'(s^t, a^t; \theta')$ for stabilizing the policy training process.

Specifically, the output of the value evaluation network is an estimation of cumulative reward function $E\left[\sum_{\tau=t}^{T} \gamma^{\tau-1} r^t | s^t, a^t\right]$. The estimation methodology using neural networks prevents the reinforcement learning method from the curse of dimensionality that traditional tabular reinforcement learning methods face.

Recalling the Bellman equation in Eq. **20**, the update target of the value evaluation network includes the evaluation network

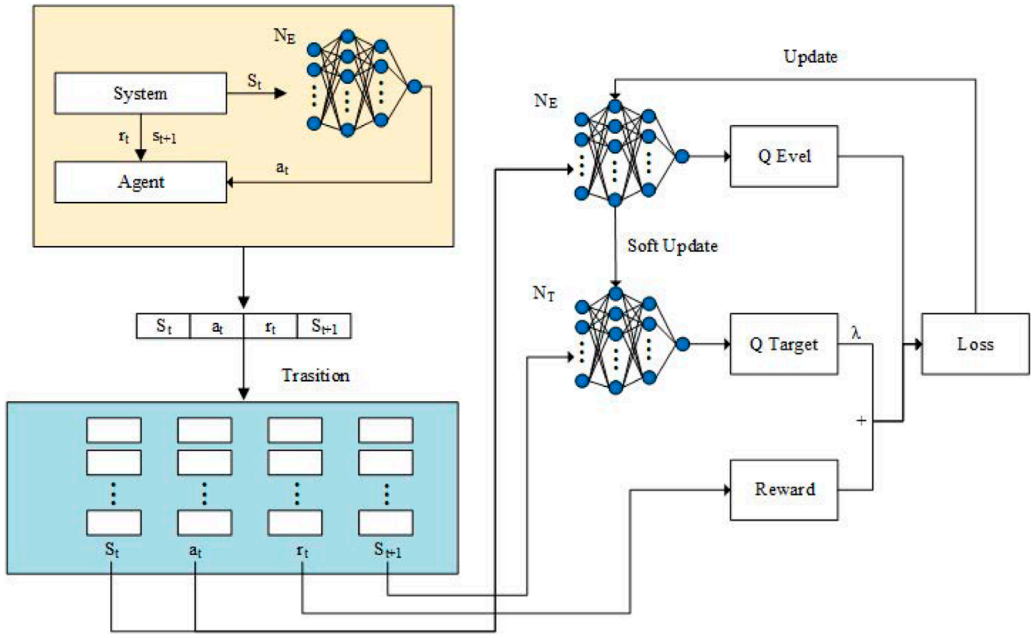



FIGURE 2
Deep reinforcement learning.

itself. It leads to the problem of instability when updating the evaluation network. To address this problem, the target network $Q^{'}(s^t, a^t;\ \theta^{'})$ is proposed. When updating the evaluation network, the fixed target network is used to replace the Q-value estimation on the right side of Eq. (20). Furthermore, after certain times of evaluation network update, the parameters of the target network will be reset as the parameters of the value evaluation network. The details of the update mechanism will be introduced in Section 5.3.

## 5.2 Action selection

During each time step, the agent needs to first observe the state of the environment, and based on this, the agent selects the action with the aim to maximize the future cumulative reward. The critical part is to balance the relationship between the exploration and exploitation in action selection. If the agent explores the environment more, the convergence speed of the policy learning process will be inevitably reduced. Nevertheless, if the agent chooses to exploit existing knowledge deeply, it may be trapped in the sub-optimal policy.

To balance between the exploration and exploitation well, an annealing $\epsilon$-greedy is used in this paper. At each time step $t$, the action chosen is determined based on a parameter $\epsilon$, which varies from (0,1). The agent will choose a random action from the action space with probability $\epsilon$ to explore the environment. Otherwise, the agent selects the action $a^t = \underset{a}{argmax}\, Q(s^t, a; \theta)$ with probability $1-\epsilon$ for exploiting existing knowledge. During the early stage of optimal policy learning, the agent has relatively less knowledge about the environment, so the agent should explore the environment more than exploiting. Thus, the value of $\epsilon$ is set as a high value during the early stage. As the agent possesses more knowledge about the environmental dynamics, the weight of exploitation should be enlarged when selecting an action, and the value of $\epsilon$ should be reduced gradually. In implementation, the value of $\epsilon$ is initialized as $\epsilon_{ini}$ which is a relatively high value before the policy learning. At each training step, the value of $\epsilon$ minus an annealing parameter $\epsilon_{dec}$ until the value of $\epsilon$ is not larger than a small value $\epsilon_{\min}$.

## 5.3 Policy iteration mechanism

At each time step $t$, the interaction information between the agent and environment $[s^t, a^t, r^t, s^{t+1}]$ is stored in an experience replay memory with size $E_r$. When updating, to ensure the property of independently identically distribution (i.i.d) of training data, a mini-batch of interaction data $[s^\tau, a^\tau, r^\tau, s^{\tau+1}]_{\tau=1}^{N_p}$ is randomly selected from the experience replay memory as training data to update the network. $N_p$ is the size of the mini-batch.

**TABLE 2** EV models.

| EV | Battery capacity (kWh) | Market share |
|---|---|---|
| Tesla Model 3 | 55 | .210 |
| Tesla Model Y | 60 | .350 |
| Tesla Model S/X | 100 | .025 |
| BYD Han EV | 85 | .100 |
| Zeeker 001 | 86 | .080 |
| Xiaopeng P7 | 60 | .130 |
| Porsche Taycan | 79.2 | .005 |
| BMW iX3 | 80 | .100 |

The value evaluation network $Q(s^t, a^t; \theta)$ is updated according to loss function as follows:

$$L = \frac{1}{N_p}\sum_{i=1}^{N_p}\left[\left(y - Q(s_\tau, a_t; \theta)\right)^2\right], \tag{34}$$

$$y = r_\tau + \gamma \max_a Q(s_\tau, a_\tau; \theta). \tag{35}$$

In order to keep the stability of the policy learning process, the target network is updated *via* tracking the value evaluation network slowly. Specifically, the parameters of the target network are reset as the parameters of the value evaluation network at every $D$ time step.

The training process of the DQN-based load hiding algorithm is summarized in **Algorithm 1**.

**Input**: EV charging/discharging environment, Env; single training step length, t; maximum number of training sessions, N; exploring mechanisms and strategies, e; window length, m; and network structure parameters, hyperparameters of DQN
**Output**: Optimal execution strategy $\pi^*$

Initialize the target network ^Q and the evaluation network Q according to the network structure;
Get the EV charging/discharging environment *Env*;
Give a state space *S* and an action space *A*;
**for** *i* = 1 *to N* **do**
Initialization of the load hiding environment *Env*;
**for** *j* = 1 *to t* **do**
Get the current state *s* from the environment;
Calculate $l_t$ based on window length;
Select an action a from the action space according to the pre-defined exploration mechanism and strategy *e*;
Get the reward for the current action from the environment observation *r*;
Get the state $s'$ after executing the current action from the environment observation;
**if** *The experience pool is not full* **then**
Store data (*s*, *a*, *r*, $s'$) to the experience pool
**else**
Let go of old experiences and deposit new ones
**end**
Randomly sample data from the experience pool (^*s*, ^*a*, ^*r*, ^$s'$);
Calculate the target value using the target network ^ *Q*;
Update the parameters of the evaluation network Q using the target value;
Assign the parameters of the evaluation network *Q* to the target network ^*Q* every *k* times;
**end**
Update the optimal action $A = a_1, a_2, \ldots a_t$;

Algorithm 1. DQN-based EV charging/discharging strategy algorithm.

## 6 Performance evaluation

In this section, we conduct several comprehensive evaluations to verify the performance of our proposed method. In the following, at first, the evaluation settings are given. Then, the results of our proposed method are introduced. Finally, the comparison results are shown.

### 6.1 Evaluation settings

In the following evaluations, the EV-assisted demand response environment follows the following settings and assumptions. First, we assume that the method is deployed in a community [randomly selected 40 electricity users in the REDD dataset (Kelly and Knottenbelt, 2015)] in 1 day (24 time slots). In addition, there exists a parking lot to help with the demand

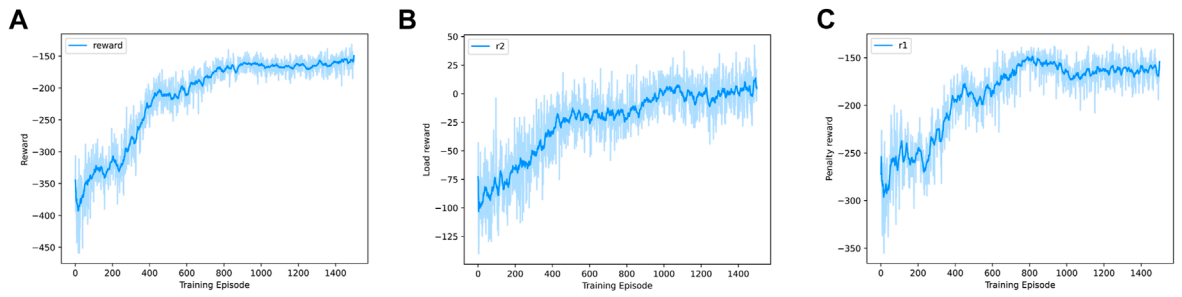



FIGURE 3
**(A)** Training process, **(B)** peak load shifting reward, and **(C)** penalty term.

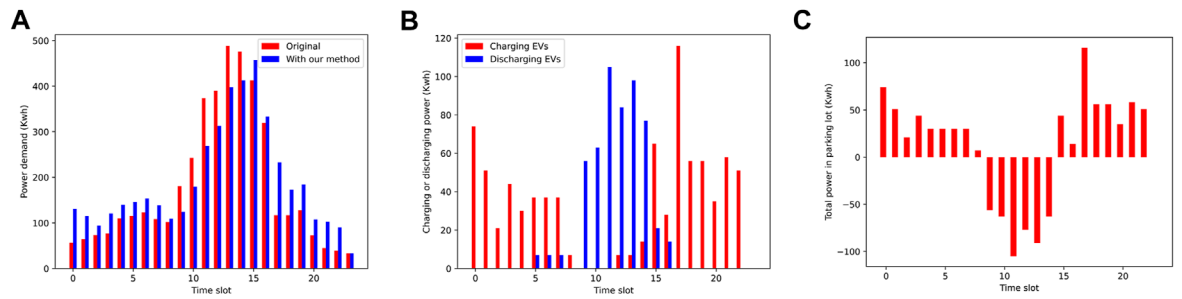



FIGURE 4
**(A)** Load curve, **(B)** charging/discharging volume, **(C)** and the difference between charging and discharging.

response, which contains 20 parking spaces. The parking lot is equipped with V2G charging piles to satisfy the charging and discharging between EVs. There exist twelve 7-kWh charging piles and eight 30-kWh charging piles. The electric vehicle models and their proportions are shown in Table 2. Their electric power varies from 55 kWh to 100 kWh. The proportion is also reasonably assumed according to the sales of electric vehicles. Then, when a parking space is free, there is the probability of $\varepsilon = .85$ that an EV will enter the parking lot and park at that location at the next time slot. The model of EVs will follow the assumption of Table 2, and its arrival SoC follows the uniform distribution from 0 to 100, while the departure SoC follows the normal distribution from 15% to 85%.

## 6.2 Reinforcement learning performance

First, we will show the performance of the reinforcement learning training process. As shown in Figure 3, we can see that when the training episode reaches about 1,000 rounds, the reward will converge quickly. Regarding Figures 3B, C, these two figures show the changes in two main components of reward. Similarly, we can see that the convergence speed is very fast. The reason for the fast convergence speed is that we simplify the action space so that the training process becomes simpler. In fact, we have tried to give each parking space an action. In this large-scale action space, there is no trend of convergence after 10,000 times of training. Meanwhile, after convergence, there still exists little penalty term. The reason behind this is the setting of the trade-off coefficient between the two awards. In order to completely eliminate the behaviors prohibited by the penalty term, we can appropriately increase the coefficient of the penalty term. From these results, we can see that our MDP model and reinforcement learning method can effectively solve the charging/discharging decision problem of EVs.

## 6.3 Power grid performance

After verifying the feasibility and effectiveness of the proposed method in training, we will verify the actual performance of the proposed method in the power grid. In Figure 4A, we can see that the proposed method is likely to allow the electric vehicle to charge at a low load and discharge at a high load. Therefore, the effect of peak load shifting can be achieved. Specifically, in the area introduced before, our proposed method can reduce 10% of the peak load and improve over 50% of the valley load. Moreover, Figures 4B, C show the

**TABLE 3** Satisfaction ratio and SoC achievement rate.

| Number of charging piles | Satisfaction ratio (%) | SoC achievement rate (%) |
|---|---|---|
| 10 | 88.6 | 90.9 |
| 20 | 85.0 | 87.5 |
| 30 | 82.1 | 78.0 |
| 40 | 89.1 | 86.8 |
| 50 | 88 | 86.9 |

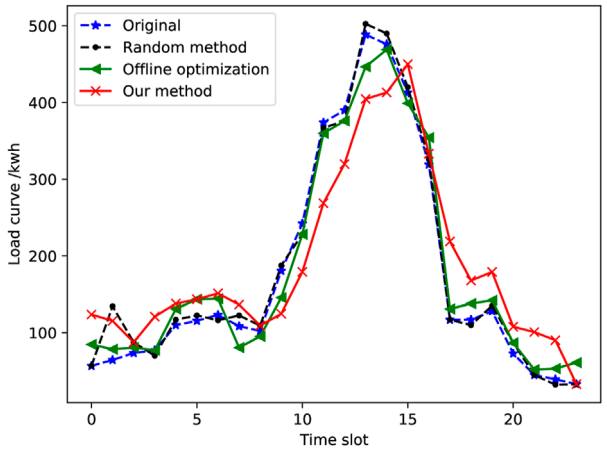



FIGURE 5
Comparison: load curve.

charging/discharging power in each time slot. The results show that in daily time, the charging behavior will be strictly limited. While during the night, in order to increase the valley load, the discharging behavior will be completely prohibited. Therefore, our method can effectively learn effective strategies to cut peak and fill valley.

The aforementioned results show that at the overall level of the power grid, our method is conducive to achieving peak load shifting. Next, we will discuss the performance of our method in ensuring the future travel of individual EV owners. As shown in Table 3, we have verified the proportion of EVs that can leave the parking lot with enough SoC (satisfaction ratio) and the final SoC achievement rate of EVs that need to be charged under different numbers of parking spaces. It can be seen that almost all EVs can leave the parking lot with the target SoC, and almost all EVs that need to be recharged can satisfy their charging requirements. In addition, our method has similar efficiency in dealing with parking spaces of different sizes because we have classified and simplified the action space, thus reducing the coupling between each action and increasing the effectiveness of the strategy.

In conclusion, our method achieves the effect of peak load shifting while ensuring individual demand.

## 6.4 Comparison

Finally, we will compare our method with other methods, such as the offline optimization method and the random method (i.e., freely charging and discharging). Regarding Figure 5, it can be demonstrated that the offline optimization method has a certain peak shaving effect but does not perform as well as our proposed method. For the random method, it will not greatly change the demand response of the grid because there are electric vehicles that need to be charged or discharged at every moment, and their loads will offset each other. While regarding the satisfaction ratio, the result of the offline optimization method is similar to that of the proposed method. The random method has no restriction on the user's behavior, so it will be equal to 1. It is not very different from the 90% achieved by our method, and it is completely acceptable.

All in all, our proposed method has a good effect in terms of convergence speed, load-shifting performance, and EV satisfaction ratio, and it also performs better than other methods.

# 7 Conclusion

In our paper, addressing the problem of demand response in a small area, we proposed a reinforcement learning-based method for an EV-assisted demand response management system to determine the best charging/discharging strategy. Specifically, we formalized the EV charging/discharging strategy determination problem as a Markov decision process (MDP), and the MDP model is constructed as follows: the state space mainly consists of occupation and charging speed of charging piles, current SoC, departure SoC, battery capacity, departure time of EVs, *etc*. The action refers to the EV charging/discharging behavior in each charging pile. We use a sliding average load method to represent the reward about the peak load shifting effect, and we set a series of penalty terms to ensure the departure SoC is enough for the next travel. Then, we proposed a DQN-based reinforcement learning architecture to solve this problem. Finally, the evaluation based on the real world shows that our method can effectively help regional peak load shifting and

has better performance than the random scheduling and offline optimization methods.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

## Author contributions

DL (first author): conceptualization, methodology, investigation, results analysis, and writing—original draft; QY (corresponding author): conceptualization, supervision, and writing—review and editing; YW: survey of methods and simulation; YZ: simulation; XL: analysis of results; LM: data processing and writing—review and editing.

## Funding

The work was supported in part by the Key Research and Development Program of Shaanxi under Grant 2022GY-033, in part by the National Science Foundation of China under Grants 61973247, 62203350, and 61673315, in part by China Postdoctoral Science Foundation 2021M692566, and in part by the operation expenses for universities' basic scientific research of central authorities xzy012021027.

## Conflict of interest

Authors LM and XL were employed by State Grid Information & Telecommunication Group Co., LTD., China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Althaher, S., Mancarella, P., and Mutale, J. (2015). Automated demand response from home energy management system under dynamic pricing and power and comfort constraints. *IEEE Trans. Smart Grid* 6, 1874–1883. doi:10.1109/TSG.2014.2388357

Cui, B., Gao, D.-c., Xiao, F., and Wang, S. (2017). Model-based optimal design of active cool thermal energy storage for maximal life-cycle cost saving from demand management in commercial buildings. *Appl. Energy* 201, 382–396. doi:10.1016/j.apenergy.2016.12.035

Eksin, C., Deliç, H., and Ribeiro, A. (2015). Demand response management in smart grids with heterogeneous consumer preferences. *IEEE Trans. Smart Grid* 6, 3082–3094. doi:10.1109/TSG.2015.2422711

Emadi, A., Lee, Y. J., and Rajashekara, K. (2008). Power electronics and motor drives in electric, hybrid electric, and plug-in hybrid electric vehicles. *IEEE Trans. Ind. Electron.* 55, 2237–2245. doi:10.1109/TIE.2008.922768

Fang, X., Misra, S., Xue, G., and Yang, D. (2012). Smart grid — the new and improved power grid: A survey. *IEEE Commun. Surv. Tutorials* 14, 944–980. doi:10.1109/SURV.2011.101911.00087

Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., et al. (2011). Smart grid technologies: Communication technologies and standards. *IEEE Trans. Ind. Inf.* 7, 529–539. doi:10.1109/TII.2011.2166794

Hua, H., Qin, Y., Hao, C., and Cao, J. (2019). Optimal energy management strategies for energy internet via deep reinforcement learning approach. *Appl. energy* 239, 598–609. doi:10.1016/j.apenergy.2019.01.145

Jeddi, B., Mishra, Y., and Ledwich, G. (2021). Distributed load scheduling in residential neighborhoods for coordinated operation of multiple home energy management systems. *Appl. Energy* 300, 117353. doi:10.1016/j.apenergy.2021.117353

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: A survey. *J. Artif. Intell. Res.* 4, 237–285. doi:10.1613/jair.301

Karapetyan, A., Khonji, M., Chau, S. C.-K., Elbassioni, K., Zeineldin, H., El-Fouly, T. H. M., et al. (2021). A competitive scheduling algorithm for online demand response in islanded microgrids. *IEEE Trans. Power Syst.* 36, 3430–3440. doi:10.1109/TPWRS.2020.3046144

Kelly, J., and Knottenbelt, W. (2015). "Neural nilm: Deep neural networks applied to energy disaggregation," in Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments, 55–64.

Kikusato, H., Mori, K., Yoshizawa, S., Fujimoto, Y., Asano, H., Hayashi, Y., et al. (2019). Electric vehicle charge–discharge management for utilization of photovoltaic by coordination between home and grid energy management systems. *IEEE Trans. Smart Grid* 10, 3186–3197. doi:10.1109/TSG.2018.2820026

Lange, S., and Riedmiller, M. (2010). "Deep auto-encoder neural networks in reinforcement learning," in The 2010 international joint conference on neural networks (IJCNN) (IEEE), 1–8.

Li, D., Yang, Q., Yu, W., An, D., Zhang, Y., and Zhao, W. (2019). Towards differential privacy-based online double auction for smart grid. *IEEE Trans. Inf. Forensic. Secur.* 15, 971–986. doi:10.1109/tifs.2019.2932911

Li, D., Yang, Q., Yu, W., An, D., Zhang, Y., and Zhao, W. (2020). Towards differential privacy-based online double auction for smart grid. *IEEE Trans. Inf. Forensic. Secur.* 15, 971–986. doi:10.1109/TIFS.2019.2932911

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., et al. (2015). *Continuous control with deep reinforcement learning*. arXiv preprint arXiv:1509.02971.

Lin, B., and Zhu, J. (2020). Chinese electricity demand and electricity consumption efficiency: Do the structural changes matter? *Appl. Energy* 262, 114505. doi:10.1016/j.apenergy.2020.114505

Littman, M. L. (1994). "Markov games as a framework for multi-agent reinforcement learning," in *Machine learning proceedings 1994*. Editors W. W. Cohen, and H. Hirsh (San Francisco (CA): Morgan Kaufmann), 157–163. doi:10.1016/B978-1-55860-335-6.50027-1

Liu, T., Hu, X., Li, S. E., and Cao, D. (2017). Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle. *Ieee. ASME. Trans. Mechatron.* 22, 1497–1507. doi:10.1109/tmech.2017.2707338

Liu, W., Qi, D., and Wen, F. (2019). Intraday residential demand response scheme based on peer-to-peer energy trading. *IEEE Trans. Ind. Inf.* 16, 1823–1835. doi:10.1109/tii.2019.2929498

Lopes, J. A. P., Soares, F. J., and Almeida, P. M. R. (2011). Integration of electric vehicles in the electric power system. *Proc. IEEE* 99, 168–183. doi:10.1109/JPROC.2010.2066250

Madawala, U. K., and Thrimawithana, D. J. (2011). A bidirectional inductive power interface for electric vehicles in v2g systems. *IEEE Trans. Ind. Electron.* 58, 4789–4796. doi:10.1109/TIE.2011.2114312

Medina, J., Muller, N., and Roytelman, I. (2010). Demand response and distribution grid operations: Opportunities and challenges. *IEEE Trans. Smart Grid* 1, 193–198. doi:10.1109/TSG.2010.2050156

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). *Playing atari with deep reinforcement learning*. arXiv preprint arXiv:1312.5602.

Ota, Y., Taniguchi, H., Nakajima, T., Liyanage, K. M., Baba, J., and Yokoyama, A. (2012). Autonomous distributed v2g (vehicle-to-grid) satisfying scheduled charging. *IEEE Trans. Smart Grid* 3, 559–564. doi:10.1109/TSG.2011.2167993

Palensky, P., and Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Trans. Ind. Inf.* 7, 381–388. doi:10.1109/TII.2011.2158841

Qian, T., Shao, C., Wang, X., and Shahidehpour, M. (2020). Deep reinforcement learning for ev charging navigation by coordinating smart grid and intelligent transportation system. *IEEE Trans. Smart Grid* 11, 1714–1723. doi:10.1109/TSG.2019.2942593

Tang, R., Li, H., and Wang, S. (2019). A game theory-based decentralized control strategy for power demand management of building cluster using thermal mass and energy storage. *Appl. Energy* 242, 809–820. doi:10.1016/j.apenergy.2019.03.152

Wan, Z., Li, H., He, H., and Prokhorov, D. (2019). Model-free real-time ev charging scheduling based on deep reinforcement learning. *IEEE Trans. Smart Grid* 10, 5246–5257. doi:10.1109/TSG.2018.2879572

Wang, S., Bi, S., and Zhang, Y.-J. A. (2018). Demand response management for profit maximizing energy loads in real-time electricity market. *IEEE Trans. Power Syst.* 33, 6387–6396. doi:10.1109/TPWRS.2018.2827401

Wu, Q., Liu, W., Yang, Y., Zhao, C., and Yong, L. (2007). "Intelligent decision support system for power grid dispatching based on multi-agent system," in International Conference on Power System Technology.

Yang, Q., Li, D., An, D., Yu, W., Fu, X., Yang, X., et al. (2020). Towards incentive for electrical vehicles demand response with location privacy guaranteeing in microgrids. *IEEE Trans. Dependable Secure Comput.* 19, 131–148. doi:10.1109/tdsc.2020.2975157

Zhang, F., Yang, Q., and An, D. (2020). Cddpg: A deep-reinforcement-learning-based approach for electric vehicle charging control. *IEEE Internet Things J.* 8, 3075–3087. doi:10.1109/jiot.2020.3015204

Zhang, D., Zhu, H., Zhang, H., Goh, H. H., Liu, H., and Wu, T. (2022). Multi-objective optimization for smart integrated energy system considering demand responses and dynamic prices. *IEEE Trans. Smart Grid* 13, 1100–1112. doi:10.1109/TSG.2021.3128547

Zhang, Y., Yang, Q., An, D., Li, D., and Wu, Z. (2022). "Multistep multiagent reinforcement learning for optimal energy schedule strategy of charging stations in smart grid," in IEEE Transactions on Cybernetics, 1–14. doi:10.1109/TCYB.2022.3165074

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 10 January 2023
DOI 10.3389/fenrg.2022.1000828



OPEN ACCESS

EDITED BY
Hanlin Zhang,
Qingdao University, China

REVIEWED BY
Wei Gao,
Ludong University, China
Bin Li,
North China Electric Power University, China

*CORRESPONDENCE
Ting Yang,
yangting@tju.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 22 July 2022
ACCEPTED 26 August 2022
PUBLISHED 10 January 2023

CITATION
Zhai F, Yang T, Sun W and Fang X (2023), Lightweight and dynamic authenticated key agreement and management protocol for smart grid.
*Front. Energy Res.* 10:1000828.
doi: 10.3389/fenrg.2022.1000828



# Lightweight and dynamic authenticated key agreement and management protocol for smart grid

Feng Zhai[1,2], Ting Yang[1]*, Wei Sun[3] and Xu Fang[4]

[1]School of Electrical Engineering and Automation, Tianjin University, Tianjin, China, [2]China Electric Power Research Institute, State Grid, Beijing, China, [3]State Grid Corporation of China, Beijing, China, [4]Henan Xj Metering Co, Ltd., Henan, China

With the development of IoT and 5G, the smart grid, as one of the key component for the smart city, can provide the uninterrupted and reliable electricity service by properly adjusting the electricity supply according to the consumption of users. The advanced metering infrastructure (AMI), as an important part of smart grid system, is a complete network and system for measuring, collecting, storing and analyzing the electricity consumption information of users. The security of AMI plays a vital role in the smooth operation of smart grid. In this paper, we study how to establish the secure communication between two entities in AMI, namely the smart meter and the electricity service provider. Although, there are many authentication and key management protocols for AMI, the high complexity and computation overhead of these protocols hinder their application in the smart grid environment. Based on identity cryptosystem and elliptic curve cryptography (ECC), we put forward a lightweight and dynamic authenticated key agreement and management protocol, which can significantly reduce the computation overhead of the resource-constrained smart meters. In addition, we utilize a one-way key tree technique to efficiently generate and update the group key in the multicast communication. We give a systematic proof to show that our designed protocol not only guarantees the confidentiality and integrity of transmitted messages, but also resists various attacks from an adversary. Finally, we carry out some simulated experiments to demonstrate the high efficiency of our designed protocol.

KEYWORDS

**key management, identity-based cryptosystem, mutual authentication, elliptic curve, key update**

## 1 Introduction

As the next generation electricity supply network, the smart grid (Song et al., 2022; Verma et al., 2022) plays an indispensable role in the progress of society and the improvement of life quality. With the development of Industrial Internet of Things (IIoT) (Ge et al., 2021), the research about smart grid has gradually become a hot topic. The

smart grid combines communication technology (Liu L. et al., 2022; Mensi et al., 2022), grid technology and computer software to complete the production, distribution and transmission of electricity. AMI, as an important part of smart grid system, generally consists of two entities: one is the electricity service provider and the other is the smart meter device. The smart meter device is usually composed of communication module and sensor module, which can collect and transmit the user's electricity consumption information in real time. The electricity service provider is usually composed of communication module and control module, which can store and analyze data. On the one hand, the electricity service provider can analyze these data detected by the smart meter in real time to formulate the more reasonable electricity supply strategy, which can effectively improve efficiency, reliability and security of smart grid. On the other hand, the smart meter device can adjust some parameters, such as unit price, based on these messages sent from an electricity service provider.

Although the smart grid has brought great convenience to the people's lives, it still faces a series of challenges and attacks (He et al., 2017; Kumar et al., 2019b; Peng et al., 2019). The smart gird is vulnerable to various attacks, such as replay attack, impersonation attack and desynchronization attack, which may cause some serious damage to the security of smart grid and the interest of users. Communications between the smart meter and the electricity service provider are carried out *via* the wired and wireless links, which are easily eavesdropped, modified and intercepted by a malicious adversary. In addition to the external adversary's attack, the secure problems brought by the insiders are also non-negligible. The transmitted messages between the smart meter and the electricity service provider often contain some confidential and sensitive data. Once these data are obtained by a malicious adversary, it will cause the serious damage to the interest of users. For example, an adversary can analyze the user's electricity consumption to determine whether the user is at home at the current time, which seriously violates the privacy of users. Therefore, how to ensure the confidentiality of transmitted messages is the first challenge in the smart grid. According to the received messages, the electricity service provider or smart meter device will formulate a electricity distribution strategy or adjust the corresponding parameters, such as updating electricity price or deciding whether to cut electricity. Once a malicious adversary modifies the messages, the electricity service provider or the smart meter may make some inappropriate modifications, decisions, and adjustments based on the modified messages, which will affect the security and stability of entire smart grid. Therefore, how to ensure the integrity of transmitted messages is the second challenge in the smart grid. The encrypted transmission of messages can be carried out by using the secret key generated through the key agreement protocol. In the smart grid, not only the privacy of messages, but also the legitimacy of messages must be ensured. Therefore, before the key agreement, both the service provider and the smart meter should authenticate each other's identity. However, most of existing authentication protocols contain some complex cryptographic operations, which are not suitable for the resource-constrained smart meter devices. How to reduce the computation overhead of smart meter during the execution phase of protocol is the third challenge in the smart grid.

The authenticated key agreement, as a key establishment method, can not only complete the key agreement, but also authenticate the identity of both parties. There are two ways to implement the authenticated key agreement: the public-key infrastructure and the identity-based cryptography. The method based on the public-key infrastructure needs a certification authority (CA) to generate and manage all certificates for users, where the certificate contains the user's public key information and other information. By verifying the validity of received certificate sent from its peer, the two communication parties can authenticate each other, which will increase the burden of certificate management and does not apply in the smart grid environment. The identity-based cryptosystem is a more sensible approach to design an authenticated key agreement protocol. However, some previous identity-based authenticated key agreement protocols usually involve some complex cryptographic operations such as bilinear pairing, which are not suitable for the smart meter devices with the limited computation and storage resources. With the rapid development of cloud computing (Gao et al., 2021; Liu Y. et al., 2022), although the resource-constrained smart meter devices can outsource the complex cryptographic operations to the cloud servers with the powerful computation resources (Li H. et al., 2022, 2021), this method not only increases the communication cost and monetary cost of smart meter devices, but also requires adjusting the architecture of entire smart grid. Therefore, one of the most straightforward ways is to design a lightweight authenticated key agreement protocol.

A smart grid system may contain millions of smart meter devices, which will result in the electricity service provider needing to manage millions of session keys at the same time. The electricity service provider not only needs to communicate with a single smart meter, but also potentially needs to carry out the multicast communication with thousands of smart meters. How to generate the group key from session key of multiple smart meters is worth investigating. In addition, once a new smart meter device is added or an old smart meter device is deleted, the group key must be modified accordingly. It is a challenge to design a protocol in which each group member can efficiently compute and update the group key, and the newly added or deleted members do not obtain the updated group key. Therefore, scalability is very important for a key management protocol.

In this paper, in order to solve the above problems, we put forward a lightweight and dynamic authenticated key agreement

and management protocol based on identity cryptosystem. Our designed protocol combines the symmetric encryption with the public key encryption, and utilizes ECC and one-way key tree structure (Sherman and McGrew, 2003) to realize authentication, key agreement and group key management and update. The main contributions of this paper can be summarized in three aspects:

- For the resource-constrained smart meter devices, we design a lightweight authenticated key agreement and management protocol based on identity cryptography and ECC. In the execution of designed protocol, each smart meter device only needs to perform several times scalar multiplication and does not need to perform other complex cryptographic operations. The designed protocol not only realizes the mutual authentication between the smart meter and the service provider, but also ensures the confidentiality and integrity of messages transmitted between the two entities.
- For the different communication methods between the smart meter and service provider, including unicast communication and multicast communication, we design a group key generation and update protocol. The service provider can generate a group key based on the session key of each smart meter and update the group key in real time according to the join and exit of smart meter. The service provider and smart meter can efficiently update the group key with the low computation cost. In addition, the designed group key update protocol can realize the forward security and backward security.
- We conduct a comprehensive security analysis to prove that our designed protocol can achieve secure authentication and message transmission, and resist to various attacks. In addition, we carry out some experiments to show that our designed protocol is efficient and lightweight.

The remainder of this paper is organized as follows. **Section 2** reviews some related work about the authentication and key management. **Section 3** gives a detailed description about system model and security requirements. The background knowledge about ECC is given in **Section 4**. In **Section 5**, we describe the designed key agreement and update protocol in detail. A formal security analysis of designed protocol is provided in **Section 6**. In **Section 7**, we evaluate the proposed protocol through the numerical analysis and experiments. Finally, we give a conclusion in **Section 8**.

# 2 Related work

In this section, we will review some previous authentication and key management protocols in the smart grid. These authentication and key management protocols are constantly modified to achieve a specific security goal and defend against various attacks.

## 2.1 Some attacks on key management protocols

At first, we introduce some attacks on the previous key management protocols. Wu and Zhou (2011) put forward a novel protocol to solve the secure key management problem in the smart grid, which combined the symmetric encryption technology based on Needham-Schroeder authentication and public key cryptosystem to realize the simplicity and scalability of key management as well as other desirable properties. Their designed protocol not only could resist some common attacks in the smart grid, such as the man-in-the-middle attack and the replay attack, but also could solve the issue of additional vulnerabilities on the session key by utilizing a strict one-time use rule and the fly key generation. However, Xia and Wang (2012) found that the adversary could utilize the man-in-the-middle attack to easily break the Wu's key management protocol. Based on the previous communication model, the authors designed a new key distribution protocol for the smart grid with the high efficiency as well as the high security, which could resist the impersonation attack, the replay attack and the man-in-the-middle attack. On the one hand, their protocol defined a lightweight directory access protocol (LDAP) server as a third-party, which could significantly reduce operation overhead. On the other hand, when revoking the user's key, their protocol only needed to remove the related entries of user. Afterwards, Park et al. (2013) pointed out that the Xia's protocol could not resist the impersonation attack. This meant that the adversary was able to impersonate the responder to the initiator.

## 2.2 Key management protocols based on ECC

Then, we introduce ECC-based key management protocols. Wan et al. (2014) designed a new scalable key management protocol, which combined the identity-based cryptosystem and the efficient key tree technique to manage the group key and take full advantage of heterogeneity of AMI system. Their protocol could significantly improve the efficiency of key management and resist the desynchronization attack, which was a problem that the previous protocol (Liu et al., 2013) did not solve. Wazid et al. (2017) put forward a three-factor authentication protocol for the remote users in the renewable energy based smart grid environment. The proposed protocol utilized the lightweight cryptographic operations such as one-way hash function, bitwise XOR operation and ECC, which could support the smart meter's dynamic addition, the

TABLE 1 Comparison with previous protocols.

| | Technology | Dynamic | Replay | Impersonation | Desynchronization |
|---|---|---|---|---|---|
| Xia and Wang (2012) | *SKC*, *PRF* | × | ✓ | × | × |
| Wan et al. (2014) | *BP*, *ECC* | ✓ | ✓ | ✓ | ✓ |
| Mahmood et al. (2018) | *ECC* | × | ✓ | ✓ | ✓ |
| our | *ECC* | ✓ | ✓ | ✓ | ✓ |

*SKC*, symmetric key cryptography; *PRF*, pseudorandom function; *BP*, bilinear pairing; *ECC*, elliptic curve cryptography.

flexibility of password and biometric update, the anonymity and untraceability of user. However, this protocol could not flexibly remove the malicious or faulty smart meters. Mahmood et al. (2018) put forward a lightweight authentication protocol based on ECC. The authors used the automated verification tool named ProVerif to analyze the security of proposed protocol and adopted the Burrows-Abadi-Needham (BAN) logic to prove the integrity and completeness of proposed protocol. Although their protocol provided the mutual authentication between the two parties, it didn't support the anonymity of smart meter. Kumar et al. (2019a) proposed a lightweight authentication and key agreement protocol, which could realize trust, anonymity, integrity and adequate security in the domain of smart energy network. The designed protocol was based on ECC, symmetric encryption, hash function and message authentication code, which could ensure the desired security with the lower computation cost. By utilizing the AVISPA (automated verification of Internet security protocol and application) tool, the authors proved that the designed protocol was semantically secure.

## 2.3 Key management protocols based on other technologies

Finally, we introduce some key management protocols based on other novel technologies, such as lattice encryption, blockchain and attribute encryption. Chaudhary et al. (2018) designed a lattice-based key exchange protocol to generate the secret session key between the two communication entities. In their protocol, a third party could securely authenticate all entities in network. The encryption algorithm was defined over the quotient ring by using the polynomial vector and simple arithmetic operations, which could ensure the confidentiality and integrity of data. In addition, the authors designed a temporary key-based protocol for detection of suspicious activity to provide the enhanced security. Based on the blockchain technology, Wang et al. (2020) put forward a mutual authentication and key agreement protocol for the smart grid system based on the edge computing, which could support the efficient conditional anonymity and key management, and didn't need other complex cryptographic primitives. Their designed protocol not only could provide the basic security properties, such as mutual authentication, secure key agreement and resisting replay attack, but also could support the efficient key update and revocation, and the conditional identity anonymity with the low computational overhead and communication overhead. Tomar and Tripathi (2022) designed a mutual authentication and key agreement protocol based on blockchain and fog computing in the smart grid environment, which could overcome some disadvantages of relying on a single trusted authority by creating a blockchain-based distributed environment assisted by cloud servers and fog nodes. The proposed protocol could achieve the default goals and was proven secure under the Real or Random (RoR) model. Based on blockchain and attribute encryption, Li J. et al. (2022) put forward an asymmetric group key agreement protocol for IIoT, which can achieve the efficient access control of participants. The proposed protocol not only realized the automation of access control, but also ensured the tamper resistance and the non-repudiation of agreement process.

**Table 1** shows the differences between our proposed protocol and some previous protocols.

# 3 System model and security model

In this section, we will give a detailed description about the system model and security model.

## 3.1 System model

As shown in **Figure 1**, the system model in the designed protocol consists of three entities as follows:

- Trusted Third Party (TTP) is trusted by all entities in this system, and responsible to produce and publish some system parameters and generate the secret key for each entity based on their identities.
- Service Provider (SP) has the sufficient computation resources and storage resources. The SP will perform mutual authentication with multiple smart meters and negotiate a session key with each smart meter. The SP stores all

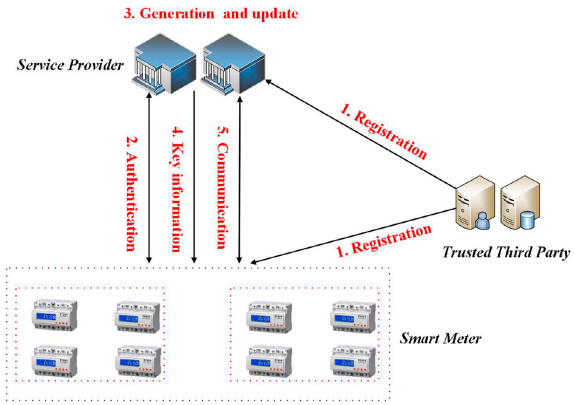



FIGURE 1
System model.

session keys to generate and update the group key by using a one-way key tree. The SP uses the session key and group key to carry out the unicast communication and multicast communication with the smart meters, respectively.

- Smart Meter (SM) has the limited computation resources and storage resources. Each SM has a session key and a group key. Each SM can communicate with the SP by the session key and use the group key to decrypt the message broadcast by the SP.

The overall execution flow of system is as follows:

Once the system is initialized by the TTP, 1) any newly added device SM or SP will register in system with submitting her/his identity to the TTP to obtain a secret key; 2) each SM and SP carry out the mutual authentication and negotiate a session key by using their respective secret keys and other information; 3) the SP divides all SMs into several groups, and uses the session key of each group member and one-way key tree technology to generate the group key of each group. In addition, the SP can update the group key according to the changes of group member; 4) the SP sends some related and necessary key information to the corresponding group members to let them generate and update the group key; 5) SP and SM choose the different communication methods (unicast or multicast) according to the different scenarios.

## 3.2 Security model and requirements

In this paper, we use the security model adopted by many previous papers (Mahmood et al., 2018). In the designed protocol, we assume that TTP is fully trusted and the secret key of TTP will not be disclosed to the adversary. The adversary can pretend to be any SM or SP during the execution of protocol. We assume that the adversary knows the identity of any SM and SP. The adversary can eavesdrop on these information transmitted on the public channels. In addition, the adversary can retrieve, modify, replay, inject new messages and discard any messages.

The designed authenticated key agreement and management protocol needs to meet the following requirements including confidentiality, integrity, availability (resilience to various attacks) and privacy:

- Confidentiality: In the smart grid, the messages transmitted between the SM and SP usually contain some confidential and sensitive information, which cannot be leaked to the adversary. Once leaked, it will cause serious damage to the interest of users and the security of smart grid. So, the designed protocol should protect the confidentiality of transmitted messages between the SM and SP.
- Integrity: The integrity of transmitted messages is an indispensable attribute of a secure authentication and key management protocol. The SP will make the important decision according to the received information or the SM will make the corresponding operation according to the received information. So, the designed protocol should protect the integrity of transmitted messages between the SM and SP.
- Availability: In the practical applications, various attacks have a serious impact on the security of smart grid. A robust authentication and key management protocol needs to keep availability under various attacks, such as replay attack, impersonation attack and desynchronization attack. The designed protocol should restrict the ability of internal or external users to launch various attacks against other components or networks.
- Privacy: During the process of updating the group key, the newly added or deleted SM may obtain some information about the group key, which cannot be leaked to them. The designed protocol should maintain both forward and backward security. This means that the newly added SM cannot obtain the previous group key and the deleted SM cannot obtain the after group key.

# 4 Background knowledge

Compared with other public key cryptography algorithms, such as RSA and Elgaml, ECC has some obvious advantages. ECC can achieve the same level of security as other schemes with the smaller scale of secret key. An elliptic curve on a finite field $F_q$ can be represented as: $y^2 = x^3 + ax + b \bmod q$, where $q$ is a large prime and $a, b \in Z_q$, $4a^3 + 27b^2 \bmod q \neq 0$. We define $E/F_q$

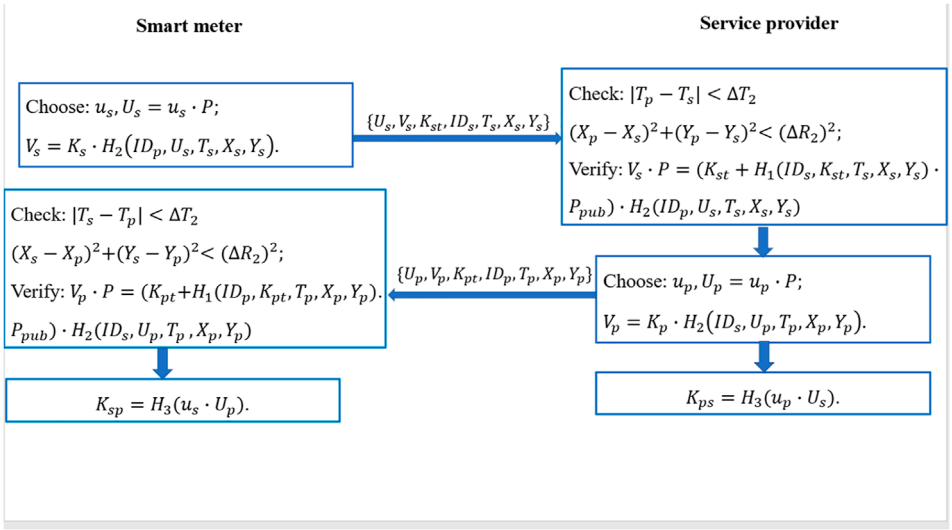



FIGURE 2
Authentication phase of designed protocol.

**Input:**
$n \in Z_q$ and $P \in E/F_q$.
**Output:**
$Q = n \cdot P$.
Set $n = \sum_{i=1}^{m} n_{i-1} 2^{i-1}$ ($n_{i-1}$ is 0 or 1).
Set $Q \leftarrow O$.
**for** $i = 1$ to $m$ **do**
  **if** $n_{i-1} = 1$ **then**
    $Q = Q + P$.
  **end if**
  $P = 2P$.
**end for**
**return** $Q$.

Algorithm 1. Scalar multiplication.

as the set of point. Given a point $P$ and an integer $n \in Z_q$, the scalar multiplication can be defined as $Q = n \cdot P$. Double-and-add algorithm is an efficient way to compute scalar multiplication, which contains two basic blocks: point addition and point doubling.

Point addition: let $P$ and $Q$ be two points on the elliptic curve, point addition describes the addition of $P$ and $Q$. There is a straight line between the point $P$ and $Q$. The line intersects the elliptic curve at another point $-F$. The output of the addition of $P$ and $Q$ is the point $F$, where the point $F$ is the reflection of the point $-F$ with respect to the $x$-axis.

Point doubling: let $P$ be a point on the elliptic curve, point doubling describes the double of the point $P$. There is one tangent line to the elliptic curve at the point $P$. The tangent line intersects the elliptic curve at another point $-F$. The output of the double of the point $P$ is the point $F$, where the point $F$ is the reflection of the point $-F$ with respect to the $x$-axis.

Algorithm 1 describes how to compute scalar multiplication, in which the point $O$ is the torsion point.

## 4.1 Definition 1 (DDH assumption)

Assume that $P$ is a random point selected from $E/F_q$ and $a, b, c$ are randomly selected from $Z_q$, the Decisional Diffie-Hellman problem is to distinguish $(P, aP, bP, abP)$ from $(P, aP, bP, cP)$. For any PPT distinguisher $\mathbb{D}$, the advantage is defined as:

$$|Pr[\mathbb{D}(P, aP, bP, abP)] - Pr[\mathbb{D}(P, aP, bP, cP)]| < negli(\lambda).$$

where $negli()$ is a negligible function of security parameter $\lambda$.

# 5 Protocol

In this section, we will introduce our proposed protocol in detail. As shown in Table 2, we define the mainly used notations in this paper. The designed protocol mainly contain three phases: initialization phase, registration phase and authentication phase. Details of each phase are described as follows:

- Initialization phase: Given a security parameter $\lambda$, TTP generates and publishes some system parameters. At first, TTP chooses a $\lambda$ bits prime $q$ and constructs $\{F_q, E/F_q, P\}$, where $P$ is a generator of group $E/F_q$. Then, TTP randomly

TABLE 2 Notations.

| Notation | Description | Notation | Description |
|---|---|---|---|
| $\lambda$ | security parameter | $H_1, H_2, H_3$ | hash function |
| $q$ | a large prime | $s, r_{st}, r_{pt}, u_p, u_s$ | element in $Z_q$ |
| $P$ | generator of group | $T_s/T_p$ | timestamp |
| $P_{pub}$ | public key | $(X_t, Y_t)/(X_p, Y_p)$ | location information |
| $ID_s/ID_p$ | user's identity | $GK^i$ | group key |
| $K_s/K_p$ | user's key | $K_{sp}/K_{ps}$ | session key |
| $K_{st}/K_{pt}$, $U_s/U_p$ | random point | $K_{ts}/K_{tp}$, $V_s/V_p$ | random number |

chooses a number $s \in Z_q$ as the master key and computes $P_{pub} = s \cdot P$. In addition, TTP chooses three hash functions $H_1 : 0,1^* \times Z_q \to Z_q$, $H_2 : 0,1^* \times Z_q \to Z_q$ and $H_3 : 0,1^* \to Z_q$. Finally, TTP publishes $\{F_q, E/F_q, P, P_{pub}, H_1, H_2, H_3\}$ as the system parameters and keeps the master key $s$ secret for itself.

- Registration phase:
  - The SM firstly chooses a random number $r_{st} \in Z_q$ and computes $K_{st} = r_{st} \cdot P$. Then, the SM sends $K_{st}$ to TTP along with its identification $ID_s$, the current timestamp $T_s$ and the current location $(X_s, Y_s)$ *via* a secure channel.
  - TTP firstly checks whether the two inequalities $|T_t - T_s| < \triangle T_1$ and $(X_t - X_s)^2 + (Y_t - Y_s)^2 < (\triangle R_1)^2$ hold. If the two inequalities hold, TTP computes $K_{ts} = s \cdot H_1(ID_s, K_{st}, T_s, X_s, Y_s)$ and sends it to the SM *via* a secure channel; otherwise, the registration process is aborted.
  - The SM verifies the validity of $K_{ts}$ by checking whether the equation $H_1(ID_s, K_{st}, T_s, X_s, Y_s) \cdot P_{pub} = K_{ts} \cdot P$ holds.
  - If the verification passes successfully, the SM computes its key $K_s = r_{st} + K_{ts}$.

For the SP, it can utilize the similar method to randomly choose a number $r_{pt} \in Z_q$ and compute $K_{pt}$. Then, the SP sends $K_{pt}$ along with its identification $ID_p$, the current timestamp $T_p$ and the current location $(X_p, Y_p)$. TTP can check and return $K_{tp}$ to the SP. The SP verifies the validity of $K_{tp}$ by checking whether the equation $H_1(ID_p, K_{pt}, T_p, X_p, Y_p) \cdot P_{pub} = K_{tp} \cdot P$ holds. Finally, the SP computes its key $K_p = r_{pt} + K_{tp}$.

- Authentication phase:
  - At first, the SM chooses a random number $u_s \in Z_q$ and computes $U_s = u_s \cdot P$. In addition, the SM computes $V_s = K_s \cdot H_2(ID_p, U_s, T_s, X_s, Y_s)$. Then, the SM sends these parameters $\{U_s, V_s, K_{st}, ID_s, T_s, X_s, Y_s\}$ to the SP.
  - The SP firstly checks whether the current time and location of the SM meet the preset conditions by the inequalities $|T_p - T_s| < \triangle T_2$ and $(X_p - X_s)^2 + (Y_p - Y_s)^2 < (\triangle R_2)^2$. If all conditions are met, the SP will verify whether the equation $V_s \cdot P = (K_{st} + H_1(ID_s, K_{st}, T_s, X_s, Y_s) \cdot P_{pub}) \cdot H_2(ID_p, U_s, T_s, X_s, Y_s)$ holds.
  - If the verification passes successfully, the SP chooses a random number $u_p \in Z_q$ and computes $U_p = u_p \cdot P$. In addition, the SP computes $V_p = K_p \cdot H_2(ID_s, U_p, T_p, X_p, Y_p)$. Then, the SP sends these parameters $\{U_p, V_p, K_{pt}, ID_p, T_p, X_p, Y_p\}$ to the SM. Finally, the SP computes $K_{ps} = H_3(u_p \cdot U_s)$.
  - The SM firstly checks whether the current time and location of the SP meet the preset conditions by the inequalities $|T_s - T_p| < \triangle T_2$ and $(X_s - X_p)^2 + (Y_s - Y_p)^2 < (\triangle R_2)^2$. If all conditions are met, the SM will verify whether the equation $V_p \cdot P = (K_{pt} + H_1(ID_p, K_{pt}, T_p, X_p, Y_p) \cdot P_{pub}) \cdot H_2(ID_s, U_p, T_p, X_p, Y_p)$ holds.
  - If the verification passes successfully, the SM computes $K_{sp} = H_3(u_s \cdot U_p)$.

Figure 2 shows the entire implementation of the proposed protocol. When the authentication process is completed, the SM and SP negotiate a session key $K_i = K_{sp} = K_{ps}$. The SM and SP can encrypt and transmit messages through the session key. Communications between the SM and SP can be divided into unicast communication and multicast communication according to the number of SMs. When the two parties conduct the secure unicast communication, they only need to use the negotiated session key between the two parties. The detailed process is as follows: Suppose there is a SP and a SM whose identity is $ID_i$. The session key negotiated by the two parties is $K_i$. When a message $m$ needs to be transmitted, the SM (SP) utilizes the session key $K_i$ and a symmetric encryption algorithm $Enc()$ such as DES or AES to encrypt the message $m$ into $M = Enc(m, K_i)$. In order to ensure the integrity of message $m$, we adopt the Hash-based Message Authentication Code (HMAC) to realize it. The SP needs to send $\{ID_i, M = Enc(m, K_i), HMAC(m, K_i)\}$ to the SM. After receiving the ciphertext of message, the SM firstly utilizes the session key $K_i$ and the corresponding decryption algorithm $Dec()$ to decrypt $M$ to obtain the message $m$. Then, the SM will recalculate the HMAC of message $m$ and compares it with the received HMAC. If the two HMACs are consistent, the message is complete and has not been tampered with.

When a SP needs to multicast with multiple SMs, it is necessary to generate a group key for these SMs. Then, we will introduce how to generate the group key and how to update the group key.

We adopt a method called One-Way Function Tree (OFT) to construct the key tree and generate the multicast key. The OFT is a particular type of binary tree in which each interior node has exactly two children. The value of each leaf node in the OFT is associated with a group member. The value of root node in the OFT is the group key (multicast key). The SP can utilize the group key to securely communicate with all members of this group. The SP can use the session key of all group members to generate the OFT as follows: The value of each leaf node in the OFT is the previously negotiated session key for each SM. For the value of

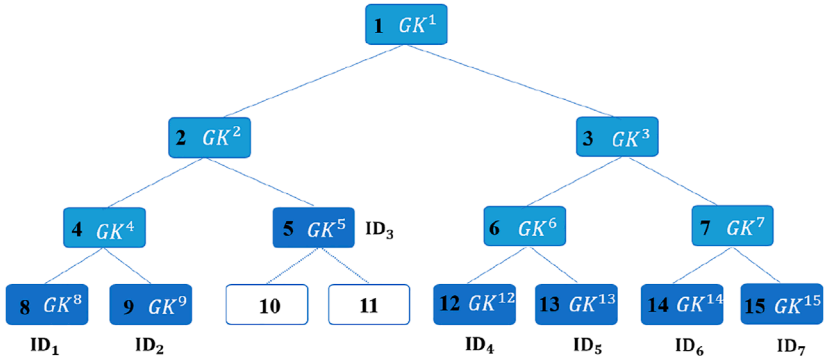



FIGURE 3
An example of OFT key tree.

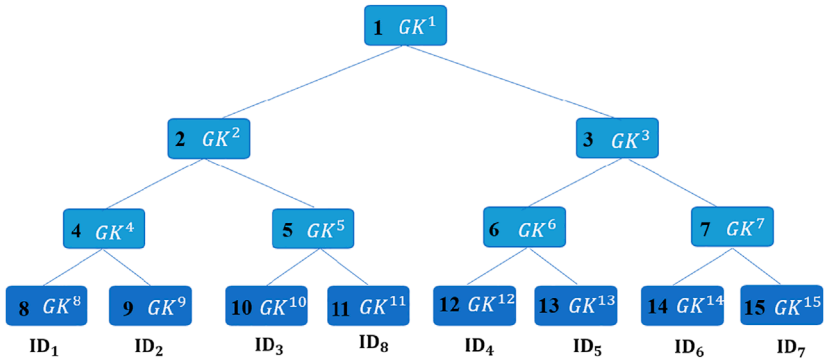



FIGURE 4
SM joining.

any interior node in the OFT key tree can be generated from the value of its two child nodes. For an interior node $v$. the value $K_v$ of node $v$ can be defined as $K_v = f(K_l) \oplus f(K_r)$. $f()$ is a special one-way function, $K_l$ and $K_r$, respectively represent the value of left child node and the value of right child node, and $\oplus$ is bitwise exclusive-or.

Each group member not only maintains the value of leaf node, but also stores a list of blinded values for all siblings of nodes along the path from this node to the root. The SP can send these blinded values to the corresponding group members, which enables the corresponding group members to compute the values of node along its path to the root, including the root key and the keys of node along this path. Once a group member (SM) is added or removed, the SP will send the necessary update information to the corresponding group members. According to the received information and locally stored information, each group member will recompute the values of node on its path to the root and obtain a new group key.

For convenience of presentation, we need to number each node in the OFT. When numbering nodes, we view the OFT as a complete tree. In other words, there are some unoccupied leaves for the future group members. Figure 3 shows the overall structure of the OFT. As shown in Figure 3, the value of each leaf node is the session key by running the above authentication protocol. The node with number 5 is a special leaf node that contains two virtual leaf nodes. For each non-leaf node $i$. the value $GK^i$ can be compute as $GK^i = f(GK^{2i}) \oplus f(GK^{2i+1})$. The value of root node $GK^1$ is the group key.

The OFT is construct by the SP. Then, the SP will broadcasts the blinded value of each sibling node along the path from the member to the root. Each SM can compute the group key according to blinded values. Each blinded value is encrypted by the value of the sibling node, so that only members in the sibling subtree can learn the blinded value. For example, in order to the SM with $ID_1$ can obtain the group key, the SP needs to send the blinded values $\{f(GK^9), f(GK^5), f(GK^3)\}$ to it. To preserve security and privacy, the SP should encrypt $f(GK^9), f(GK^5), f(GK^3)$ with $GK^8$. To preserve integrity, the SP also utilizes HMAC. So, the SP needs to send $\{Enc(f(GK^9), GK^8), Enc(f(GK^5), GK^8),$

$Enc\,(f\,(GK^3),\ GK^8)$, $HMAC\,(f\,(GK^9 \| f\,(GK^5) \| f\,(GK^3),\ GK^8)\}$ to the SM with $ID_1$.

The OFT is dynamic and updatable. When a new SM adds to this group or an existing SM leaves this group, the SP will re-compute the group key and send the updated blinded values to the corresponding SMs. Each SM can update the group key according to the received information. Let's take the OFT in Figure 3 as an example to show the changes in the OFT after adding a new SM with $ID_8$. As shown in Figure 4, the two new nodes with numbers 10 and 11 are added to the original OFT. The SM with $ID_3$ is associated with a leaf node with number 10 and the SM with $ID_8$ is associated with a leaf node with number 11. The value of node with number 5 is generated from $f(GK^{10})$ (the blinded value of $ID_3$'s session key) and $f(GK^{11})$ (the blinded value of $ID_8$'s session key). The original leaf node with number 5 becomes an interior node, which contains two leaf nodes with numbers 10 and 11. Then, the SP needs to send $\{Enc\,(f\,(GK^3), GK^{11}), Enc\,(f\,(GK^4), GK^{11}), Enc\,(f\,(GK^{10}), GK^{11}),\ HMAC\,(f\,(GK^3 \| f\,(GK^4) \| f\,(GK^{10})),\ GK^{11})\}$ to the SM with $ID_8$ to compute the group key. The SP needs to send $\{Enc(f(GK^{11}), GK^{10}), HMAC(f(GK^{11}), GK^{10})\}$ to the SM with $ID_3$ to update the group key. The SP needs to send $\{Enc(f(GK^5), GK^4), HMAC(f(GK^5), GK^4)\}$ to the SM with $ID_1$ and $ID_2$ to update the group key. The SP needs to send $\{Enc(f(GK^2), GK^3), HMAC(f(GK^2), GK^3)\}$ to the SM with $ID_4$, $ID_5$, $ID_6$ and $ID_7$ to update the group key. Each SM can update the group key with the blinded values of corresponding nodes according to the equation $GK^i = f(GK^{2i}) \oplus f(GK^{2i+1})$.

When an existing SM leaves this group, the SP can use a similar method to update the group key.

The SP can multicast with multiple SMs by the group key $GK$. Similar to the unicast communication, the SP utilizes the group key $GK$, the symmetric encryption algorithm $Enc()$ and authentication code $HMAC$ to broadcast a message $m$. The SP broadcasts $\{GID, M = Enc(m, GK), HMAC(m, GK)\}$ to all group members, where $GID$ is the group identity. On receiving the above message, each SM will decrypt the ciphertext to obtain the message $m$ and verify the integrity of $m$ by computing HMAC.

# 6 Security analysis

In this section, we will conduct the security analysis about the authentication phase and the group key update phase under the security model defined in Section 3. The security analysis about the registration phase is similar to the authentication phase.

## 6.1 Replay attack

A replay attack means that the adversary can eavesdrop on the exchanged messages and resend some messages at the adversary's will. In the authentication phase, the SM and the SP can challenge each other. Note that the exchanged messages contain the current timestamp $T_s$ or $T_p$. In the communications between the two parties, $T_s$ or $T_p$ is not only transmitted in the form of plaintext, but also hidden in $V_s = K_s \cdot H_2(ID_p, U_s, T_s, X_s, Y_s)$ or $V_p = K_p \cdot H_2(ID_s, U_p, T_p, X_p, Y_p)$. For example, the adversary generates and sends the fresh timestamp $t_s$. The adversary expects the SP to return something that matches its secret key in the next message. However, the SP fails to verify the equation $V_s \cdot P = (K_{st} + H_1(ID_s, K_{st}, T_s, X_s, Y_s) \cdot P_{pub}) \cdot H_2(ID_p, U_s, T_s, X_s, Y_s)$ in our designed protocol. Therefore, the replay attack is thwarted. When the SP sends some messages to the SM, in a similar way, we can prove that this process is also resistant to the replay attack due to $T_p$. This is because the SM fails to verify the equation $V_p \cdot P = (K_{pt} + H_1(ID_p, K_{pt}, T_p, X_p, Y_p) \cdot P_{pub}) \cdot H_2(ID_s, U_p, T_p, X_p, Y_p)$ if the adversary generate a fresh timestamp $t_p$.

## 6.2 Impersonation attack

The impersonation attack means that the adversary can be authenticated and communicate with the other parties. That is to say, the adversary can pretend to be the SM (SP) and communicate with the SP (SM). In our designed protocol, the SM and the SP need to carry out the mutual authentication by utilizing the secret key generated by the TTP. If the adversary wants to impersonate the SM, he should generate a valid request $\{U_s, V_s, K_{st}, ID_s, T_s, X_s, Y_s\}$ to the SP. However, the process of generating $U_s$ and $V_s$ involves the SM's private key $K_s$. Based on the DDH assumption, the adversary cannot recover the private key $K_s$ from the intercepted messages. Similarly, if the adversary wants to impersonate the SP, he should generate a valid response $\{U_p, V_p, K_{pt}, ID_p, T_p, X_p, Y_p\}$ to the SM. The process of generating $U_p$ and $V_p$ involves the SP's private key $K_p$. Therefore, our designed protocol is resistant to the impersonation attack.

## 6.3 Desynchronization attack

The desynchronization attack means that the adversary can block message transmission between the SM and the SP to make them lose key synchronization permanently. Once this desynchronization attack is successful, the SM and the SP will no longer communicate with each other. In our designed protocol, the session key is constructed by the random number and timestamp. There is no connection between the newly generated session key and the previously generated session key. Therefore, our designed protocol is resistant to the desynchronization attack. Even if the message is blocked, we can run the designed protocol again to synchronize.

## 6.4 Unicast and multicast communications security

On the one hand, the unicast key is the session key negotiated between the SM and the SP. The multicast key is generated by using the OFT. The security of session key depends on the security of designed authentication protocol, which we have proven through various attacks. As for the multicast key, according to the construct of OFT, we can know that only the corresponding group members can obtain the group key. Other entities cannot obtain the group key without knowing the relevant key material. On the other hand, it is obvious that our designed protocol can both protect the confidentiality and integrity of messages. By encrypting the content of messages with the session key or the group key, our designed protocol can protect the confidentiality of messages. By computing the HMAC of messages with the session key or the group key, our designed protocol can guarantee the integrity of messages.

## 6.5 Backward security

Backward security means that when a SM joins the group, it will not be able to calculate the previous group key, even if multiple newly joined SMs collude. When a new SM (leaf node) joins the group, as shown in **Figure 4**, all values of node on the path from this node to the root in the OFT key tree will be updated. The key tree will add two leaf nodes. The original leaf node will become the parent node of the two leaf nodes, which is an interior node. The newly added node can only receive a blinded value of the original leaf node. All values of node on the path from the leaf node to the root is based on the real value of the original leaf node. However, after updating, all values of node on the path from the leaf node to the root is based on the blinded value of the original leaf node. Without knowing the real value of the original leaf node, the newly joined SM will not recover the previous group key. Even though multiple newly joined SMs collude, they cannot recover the previous group key. This is because they only receive the blinded values of their sibling nodes (the leaf node in the original key tree). The previous group key is computed based on the real values of their sibling nodes. Therefore, no matter how many newly joined smart members collude together, they cannot recover the previous group key.

## 6.6 Forward security

Forward security means that when a SM is removed from the group, it will not be able to compute the new group key, even if multiple removed SMs collude. Similar to backward security, we can prove that our designed protocol compliant with forward security by the same way. The previous group key is computed based on the blinded values of their sibling nodes. After removing some SMs, the new group key is compute based on the real values of their sibling nodes. Therefore, without knowing the real values of their sibling nodes, no matter how many removed SMs collude together, they cannot obtain the new group key.

TABLE 3 Analysis about protocol.

| | Computation cost | Communication cost |
|---|---|---|
| Registration (SM/SP) | $3 \cdot sm + 1 \cdot hash$ | $2\lambda$ bits |
| Registration (TTP) | $1 \cdot hash$ | $\lambda$ bits |
| Authentication (SM/SP) | $5 \cdot sm + 4 \cdot hash$ | $5\lambda$ bits |

# 7 Evaluation

## 7.1 Numerical evaluation

We give some numerical analysis about computation cost and communication cost. In the computation cost analysis, we only focus on the number of each entity performs the scalar multiplication and hash algorithm. We ignore other lightweight operations. We denote *sm* as once scalar multiplication and *hash* as once hash. At first, in the registration phase, SM/SP needs once *sm* to compute $K_{st}/K_{pt}$, and once *hash* and twice *sm* to complete verification. TTP needs once *hash* to compute $K_{ts}/K_{tp}$. In terms of communication cost, SM/SP needs to send $K_{st}/K_{pt}$ to TTP and TTP needs to return $K_{ts}/K_{tp}$ to SM/SP. The bit length of $K_{st}/K_{pt}$ is $2\lambda$ and the bit length of $K_{ts}/K_{tp}$ is $\lambda$. In the authentication phase, SM/SP needs five times *sm* and four times *hash* to complete authentication and key agreement. SM/SP needs to send $\{U_s, V_s, K_{st}\}/\{U_p, V_p, K_{pt}\}$ to another entity. The bit length of $\{U_s, V_s, K_{st}\}/\{U_p, V_p, K_{pt}\}$ is $5\lambda$. We ignore other transmitted data. **Table 3** shows the analysis about computation cost and communication cost.

## 7.2 Experiment evaluation

In this section, we carry out some experiments to show that our designed protocol is lightweight and efficient. In our experiments, we use a computer with Linux Ubuntu 20.04.2 LTS operating system and Intel Core i5 processors with 2.4 GMz and 2G memory to simulate all entities in the designed system, including SM, SP and TTP. Our experiments utilize the C++ programming language to implement our designed protocol and adopt the PBC library to perform scalar multiplication on elliptic curve. The hash function in our experiment is SHA-256.

In the first experiment, as the bit length of modulus $q$ increases, we count the time cost in the different stages of each

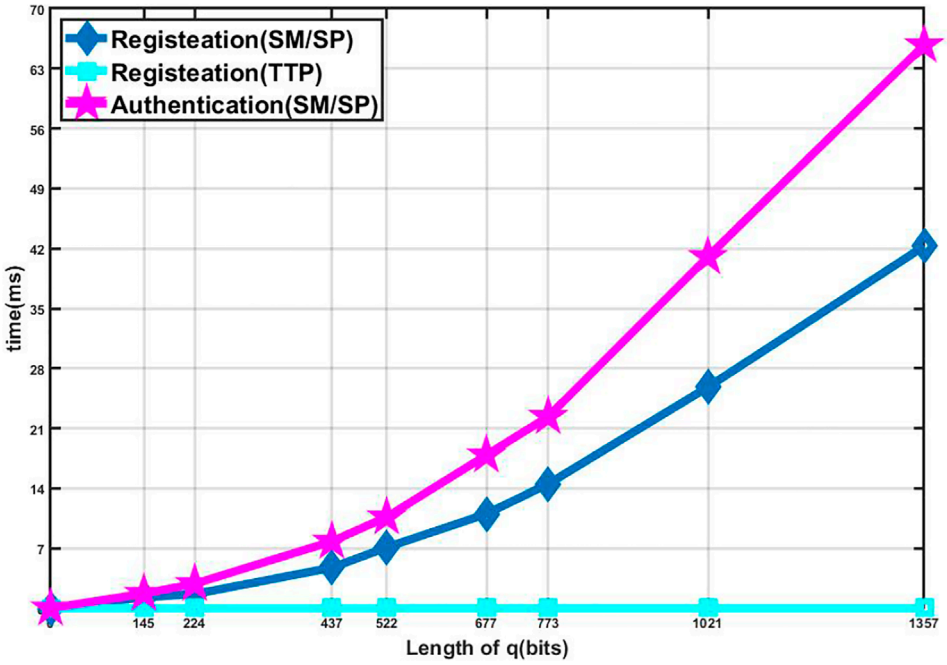



FIGURE 5
Time cost under different bit lengths of *q*.

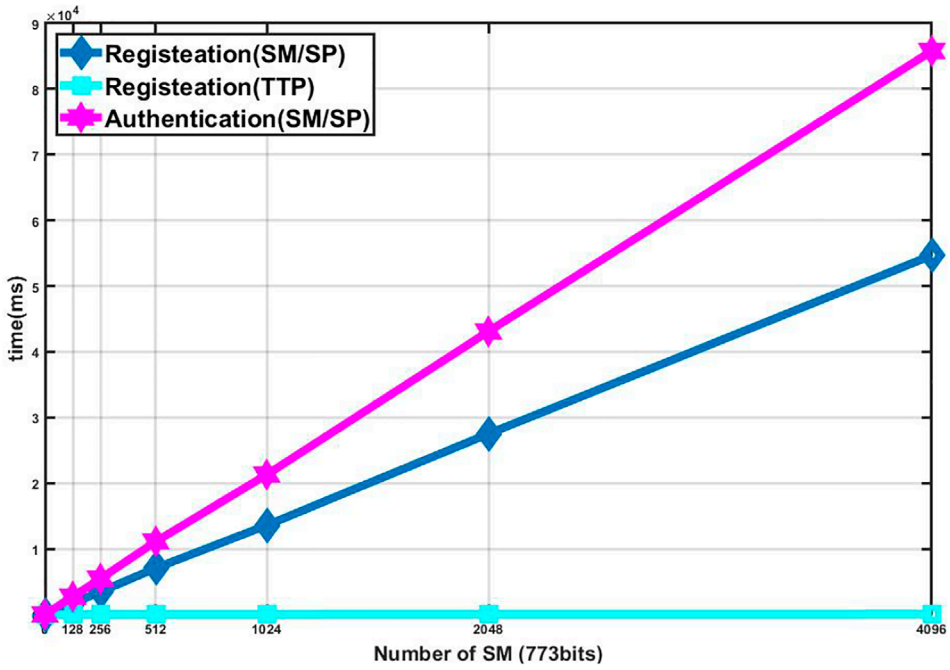



FIGURE 6
Time cost under different numbers of SM.

entity. As shown in Figure 5, as the bit length of $q$ increases, the computation overhead of each entity in each stage will also increase accordingly, which also means that the designed protocol has higher security. From Figure 5, we can find that the time cost of TTP is much smaller than that of SM and SP. This is because TTP only communicates with SM or SP and generates the secret key during the registration phase, which only contains once hash algorithm and some lightweight operations in this phase, such as computing the product of two numbers. SM or SP needs to perform multiple scalar multiplications on elliptic curve during the registration and authentication phase. In addition, because the registration phase requires three times scalar multiplications and the authentication phase requires five times scalar multiplications, the time cost of authentication phase is higher than that of the registration phase.

In a second experiment, we show the time cost of each entity when the number of SMs increases. As shown in Figure 6, we can find that, when the number of SMs increases, the time cost of TTP does not change significantly. It shows that TTP can efficiently generate the secret key for each SM in a large-scale smart grid environment.

## 8 Conclusion

In this paper, we design a lightweight authenticated key agreement and management protocol based on the identity cryptosystem and scalar multiplication on elliptic curve. The designed protocol takes time and geographical factors into account, and can quickly realize the mutual authentication and key negotiation between the two parties in the smart grid. In addition, we design a group key generation and update protocol, which enables the SP and SM to efficiently generate and update the group key in the multicast communication by utilizing a one-way key tree structure. Then, we give an analysis to show that our designed protocol satisfies our given design goals including confidentiality, integrity, and availability. We also prove that the forward and backward security of group key can be guaranteed in the update of group key. Finally, we show the efficiency of proposed protocol through experiments. Our proposed protocol may be not perfect and has some shortcomings. On the one hand, in the current protocol, TTP needs to send the necessary key information to the corresponding SM whenever the group membership changes. If the SM changes frequently, this greatly increases the communication complexity between the two parties. On the other hand, the designed protocol does not take quantum attacks into account, which may have an impact on the security of protocol. In future research, we will explore how to reduce the communication complexity in key update and how to improve the security.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Author contributions

FZ, TY, and WS contributed to conception and design of the study. FZ and XF organized the database. FZ and TY performed the statistical analysis. FZ wrote the first draft of the manuscript. FZ and TY wrote sections of the manuscript. FZ, TY, WS, and XF contributed to manuscript revision, read, and approved the submitted version.

## Funding

This work was supported in part by the National Key Research and Development Program of China (2017YFE0132100) and the National Natural Science Foundation of China (61971305).

## Conflict of interest

FZ was employed by China Electric Power Research Institute, State Grid. WS was employed by State Grid Corporation of China. XF was employed by Henan Xj Metering Co, Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References


Chaudhary, R., Aujla, G. S., Kumar, N., Das, A. K., Saxena, N., and Rodrigues, J. J. P. C. (2018). "Lacsys: Lattice-based cryptosystem for secure communication in smart grid environment," in 2018 IEEE International Conference on Communications (ICC), 1–6.

Gao, X., Yu, J., Chang, Y., Wang, H., and Fan, J. (2021). "Checking only when it is necessary: Enabling integrity auditing based on the keyword with sensitive information privacy for encrypted cloud data," in IEEE Transactions on Dependable and Secure Computing, 1.

Ge, X., Yu, J., Zhang, H., Bai, J., Fan, J., and Xiong, N. N. (2021). "SPPS: A search pattern privacy system for approximate shortest distance query of encrypted graphs in IIoT," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, 1–15.

He, D., Kumar, N., Zeadally, S., Vinel, A., and Yang, L. T. (2017). Efficient and privacy-preserving data aggregation scheme for smart grid against internal adversaries. *IEEE Trans. Smart Grid* 8, 2411–2419. doi:10.1109/tsg.2017.2720159

Kumar, P., Gurtov, A., Sain, M., Martin, A., and Ha, P. H. (2019a). Lightweight authentication and key agreement for smart metering in smart energy networks. *IEEE Trans. Smart Grid* 10, 4349–4359. doi:10.1109/tsg.2018.2857558

Kumar, P., Lin, Y., Bai, G., Paverd, A., Dong, J., and Martin, A. (2019b). Smart grid metering networks: A survey on security, privacy and open research issues. *IEEE Commun. Surv. Tutorials* 21, 2886–2927. doi:10.1109/comst.2019.2899354

Li, H., Yu, J., Fan, J., and Pi, Y. (2022a). "DSOS: A distributed secure outsourcing system for edge computing service in iot," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, 1–13.

Li, H., Yu, J., Yang, M., and Kong, F. (2021). Secure outsourcing of large-scale convex optimization problem in internet of things. *IEEE Internet Things J.* 9, 8737–8748. doi:10.1109/jiot.2021.3116127

Li, J., Qiao, Z., and Peng, J. (2022b). "Asymmetric group key agreement protocol based on blockchain and attribute for industrial internet of things," in IEEE Transactions on Industrial Informatics, 1.

Liu, L., Zhang, Z., Wang, N., Zhang, H., and Zhang, Y. (2022a). "Online resource management of heterogeneous cellular networks powered by grid-connected smart micro grids," in IEEE Transactions on Wireless Communications, 1.

Liu, N., Chen, J., Zhu, L., Zhang, J., and He, Y. (2013). A key management scheme for secure communications of advanced metering infrastructure in smart grid. *IEEE Trans. Ind. Electron.* 60, 4746–4756. doi:10.1109/tie.2012.2216237

Liu, Y., Yu, J., Fan, J., Vijayakumar, P., and Chang, V. (2022b). Achieving privacy-preserving dsse for intelligent iot healthcare system. *IEEE Trans. Ind. Inf.* 18, 2010–2020. doi:10.1109/tii.2021.3100873

Mahmood, K., Chaudhry, S. A., Naqvi, H., Kumari, S., Li, X., and Sangaiah, A. K. (2018). An elliptic curve cryptography based lightweight authentication scheme for smart grid communication. *Future Gener. Comput. Syst.* 81, 557–565. doi:10.1016/j.future.2017.05.002

Mensi, N., Rawat, D. B., and Balti, E. (2022). Gradient ascent algorithm for enhancing secrecy rate in wireless communications for smart grid. *IEEE Trans. Green Commun. Netw.* 6, 107–116. doi:10.1109/TGCN.2021.3093821

Park, J. H., Kim, M., and Kwon, D. (2013). Security weakness in the smart grid key distribution scheme proposed by xia and wang. *IEEE Trans. Smart Grid* 4, 1613–1614. doi:10.1109/tsg.2013.2258823

Peng, C., Sun, H., Yang, M., and Wang, Y. (2019). A survey on security communication and control for smart grids under malicious cyber attacks. *IEEE Trans. Syst. Man. Cybern. Syst.* 49, 1554–1569. doi:10.1109/tsmc.2018.2884952

Sherman, A. T., and McGrew, D. A. (2003). Key establishment in large dynamic groups using one-way function trees. *IIEEE. Trans. Softw. Eng.* 29, 444–458. doi:10.1109/tse.2003.1199073

Song, E. Y., FitzPatrick, G. J., Lee, K. B., and Griffor, E. (2022). A methodology for modeling interoperability of smart sensors in smart grids. *IEEE Trans. Smart Grid* 13, 555–563. doi:10.1109/tsg.2021.3124490

Tomar, A., and Tripathi, S. (2022). Blockchain-assisted authentication and key agreement scheme for fog-based smart grid. *Clust. Comput.* 25, 451–468. doi:10.1007/s10586-021-03420-2

Verma, G. K., Gope, P., and Kumar, N. (2022). PF-DA: Pairing free and secure data aggregation for energy internet-based smart meter-to-grid communication. *IEEE Trans. Smart Grid* 13, 2294–2304. doi:10.1109/tsg.2021.3138393

Wan, Z., Wang, G., Yang, Y., and Shi, S. (2014). Skm: Scalable key management for advanced metering infrastructure in smart grids. *IEEE Trans. Ind. Electron.* 61, 7055–7066. doi:10.1109/tie.2014.2331014

Wang, J., Wu, L., Choo, K. K. R., and He, D. (2020). Blockchain-based anonymous authentication with key management for smart grid edge computing infrastructure. *IEEE Trans. Ind. Inf.* 16, 1984–1992. doi:10.1109/tii.2019.2936278

Wazid, M., Das, A. K., Kumar, N., and Rodrigues, J. J. P. C. (2017). Secure three-factor user authentication scheme for renewable-energy-based smart grid environment. *IEEE Trans. Ind. Inf.* 13, 3144–3153. doi:10.1109/tii.2017.2732999

Wu, D., and Zhou, C. (2011). Fault-tolerant and scalable key management for smart grid. *IEEE Trans. Smart Grid* 2, 375–381. doi:10.1109/tsg.2011.2120634

Xia, J., and Wang, Y. (2012). Secure key distribution for the smart grid. *IEEE Trans. Smart Grid* 3, 1437–1443. doi:10.1109/tsg.2012.2199141

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fenrg.2022.1071973



OPEN ACCESS

EDITED BY
Hanlin Zhang,
Qingdao University, China

REVIEWED BY
Linqiang Ge,
Columbus State University, United States
Yalong Wu,
University of Houston–Clear Lake, United States

*CORRESPONDENCE
Qingyu Yang,
yangqingyu@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 17 October 2022
ACCEPTED 21 December 2022
PUBLISHED 12 January 2023

CITATION
Li D, Yang Q, Ma L, Peng Z and Liao X (2023), Offense and defence against adversarial sample: A reinforcement learning method in energy trading market. *Front. Energy Res.* 10:1071973.
doi: 10.3389/fenrg.2022.1071973



# Offense and defence against adversarial sample: A reinforcement learning method in energy trading market

Donghe Li[1], Qingyu Yang[1,2]*, Linyue Ma[3], Zhenhua Peng[1] and Xiao Liao[3]

[1]School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China, [2]State Key Laboratory Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an, China, [3]State Grid Information and Telecommunication Group Co., LTD, Beijing, China

The energy trading market that can support free bidding among electricity users is currently the key method in smart grid demand response. Reinforcement learning is used to formulate optimal strategies for them to obtain optimal strategies. Non-etheless, the security problem raised by artificial intelligence technology has been paid more and more attention. For example, the neural network has been proved to be able to resist adversarial example attacks, thus affecting its training results. Considering that reinforcement learning is also widely used for training by neural networks, the security problem can not be ignored, especially in scenarios with high security requirements such as smart grids. To this end, we study the security issues in reinforcement learning-based bidding strategy method facing by the adversarial example. First of all, regarding to the electric vehicle double auction market, we formalize the bidding decision problem of EVs into a Markov Decision Process, so that reinforcement learning is used to solve this problem. Secondly, from the perspective of attackers, we have designed a local Fast Gradient Sign Method which affects the environment and the results of reinforcement learning by changing its own bidding. Then, from the perspective of the defender, we designed a reinforcement learning training network containing an attack identifier based on the deep neural network, so as to identify malicious injection attacks to resist against adversarial attacks. Finally, comprehensive simulations are conducted to verify our proposed method. The results shows that, our proposed attack method will reduce the auction profit by influencing reinforcement learning algorithm, and the protect method will be able to completely resist such attacks.

KEYWORDS

**double auction, markov decision process, reinforcement learning, adversarial example, fast gradient sign method, adversarial example detection**

# 1 Introduction

With the application of more and more Internet of Things equipment and information technology, the traditional purely physical power grid has gradually transformed into the Cyber Physic System-based (CPS) Smart Grid (SG) Zhang et al. (2016); Hong et al. (2019); Bandyszak et al. (2020); Zhao et al. (2021); An et al. (2022). Smart grid provides bi-direction information flow and power flow through advanced information technology, and realize effective interconnection of power generation, transmission, distribution and others Grigsby (2007); Haller et al. (2012). The most important function of smart grid is to plan and guide users to actively adjust their power load by taking advantage of the bi-direction transmission of information between the grid and users, so as to achieve the effect of peak load shifting, which is called Demand Response (DR) Croce et al. (2017); Albadi and El-Saadany (2007); Huang et al. (2019).

With the development of science, technology and society, almost all equipment depends on electric power transportation. People are increasingly dependent on electricity, which brings great pressure to the stable operation of the power grid. It is urgent to use demand response methods to alleviate the pressure. The mainstream demand response methods are divided into two categories, one is price-based DR and the other is incentive-based Hahn and Stavins (1991); Pyka (2002); Liu et al. (2005) DR. The price-based DR method guides users to adjust the load actively by setting the price, such as Time of Use Price (TOU), Real Time Pricing (RTP), and so on Ding et al. (2016); Cheng et al. (2018); Samadi et al. (2010). The incentive-based DR method realizes load migration by directly managing the user's load, such as Direct Load Control (DLC), Energy Trading Market, and so on Wu et al. (2015); Ruiz et al. (2009); Ng and Sheble (1998).

Due to the continuous increase of renewable energy Hosseini et al. (2021); Giaconi et al. (2018) and the popularity of flexible load and energy storage Miao et al. (2015); Liu et al. (2018) equipment such as electric vehicles, the energy trading market, which allows users to freely bid and transmit electric energy, has received extensive attention Kim et al. (2019); Esmat et al. (2021). Generally speaking, in typical energy trading market, the electricity users (or electric energy company) with surplus energy will act as sellers, the electricity uses with insufficient energy will act as buyers. Regarding to the winner decision mechanism in energy trading market, considering that the market has strong uncertainty, and the market needs to ensure the benefits of participants, fairness and other properties to attract more participants, the auction mechanism has better performance than the optimization algorithm, which is the mainstream of current research. In recent years, most scholars have devoted themselves to studying a more efficient auction mechanism from the perspective of auction platform. For example, two example of auction mechanism.

With further research, scholars found that determining optimal bidding strategy from the perspective of participants also affects the performance of the energy trading market. Reinforcement learning is a branch of machine learning, which focuses on interactive goal oriented learning Mohan and Laird (2014); Erhel and Jamet (2016). It can independently explore the environment and constantly optimize its own strategies driven by rewards. Deep reinforcement learning combines the independent exploration ability of reinforcement learning with the strong fitting ability of neural network, and has been widely studied Yu et al. (2022); Zhang et al. (2019). Deep reinforcement learning technology is widely used in the optimal decision-making of smart grid due to its strong perception and understanding ability and sequential decision-making ability Barto et al. (1989); Roijers et al. (2013). For example, two example of RL bidding.

In recent years, deep learning technology has made unprecedented development and has been widely used in many fields. However, its security problems have become increasingly prominent. Szegedy et al. (2013) found that the deep neural network is extremely vulnerable to the attack of adding disturbance to the confrontation sample image. This attack will cause the neural network to classify the image with high confidence, and the human can hardly distinguish the confrontation sample from the original image with their eyes. For instance, in **Figure 1**, the original panda image is judged as a panda by the depth learning image classification model with 57.7% confidence, but after adding small random noise, the model will misjudge the image as a gibbon with high confidence Goodfellow et al. (2014). The sample, which is carefully created or generated and leads to the wrong prediction of the deep learning model, is called Adversarial Example (AE) Szegedy et al. (2013). The training process of deep reinforcement learning also relies on neural networks, so theoretically, there is also a risk of being attacked by adversarial example. Moreover, the smart grid system, which requires high reliability, will have a great impact once the reinforcement learning algorithm is attacked by the adversarial example.

As introduced above, it is urgent to study the security problems of reinforcement learning algorithm applied in smart grid. In our paper, we mainly focus on the attack and defense of the bidding strategy algorithm based on reinforcement learning of double energy trading mechanism. At present, the research on counter attack has been carried out for several years, but the following problems still exist. 1) The application scenarios of reinforcement learning algorithms are mostly game environments. The research on adversarial attack is mainly carried out on images, and the effectiveness of scenes other than images is hardly explored. 2) It is worth exploring the adversarial attack and defense effects when the state observation is very limited information.

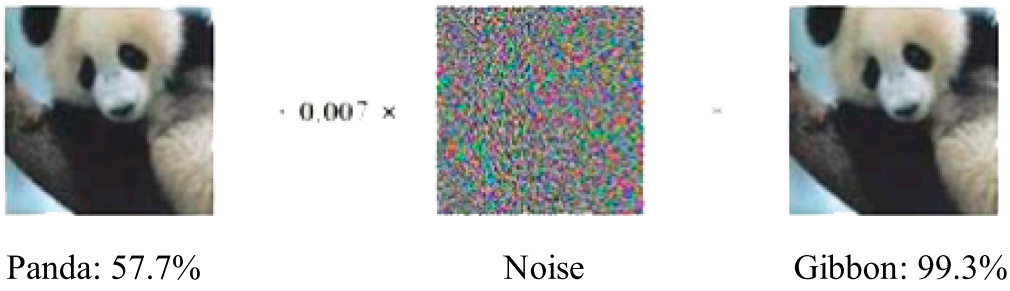



FIGURE 1
Examples of adversarial attack.

To this end, in this paper we will study the security issue aiming at the reinforcement learning-based bidding strategy method in Electric Vehicle double energy trading market. Specifically, we first conduct a double auction model/mechanism of EV double energy trading market. And we formalize the EVs' bidding strategy as a Markov Decision Process model so as to solve it by deep reinforcement learning. After that, we studied a method of generating Adversarial Example based on fast gradient sign method from the adversary's point of view, and explore the impact of AE on deep reinforcement learning algorithm. Then, we designed a deep reinforcement learning network that contains a deep neural network-based adversarial example discriminator to resist such attacks from the perspective of the defender. Finally, comprehensive simulations are conducted to verify our methods.

The remainder of this paper is organized as follows. In **Section 2**, we briefly review the research efforts related to energy trading market, reinforcement learning method and the adversarial example. In **Section 3**, we introduce the preliminaries of this paper. In **Section 4**, a local-fast gradient adversarial example generating method is proposed. In **Section 5**, the deep neural network-based adversarial example discriminator is proposed to protect the reinforcement learning method. In **Section 6**, the simulations are conducted. Finally, we conclude this paper in **Section 7**.

## 2 Related work

With the development of distributed energy and energy storage equipment, the electricity trading market between users has become an important research content in smart grid demand response. For example, Jin et al. (2013) studied the electric vehicle charging scheduling problem from the perspective of energy market, and proposed a mixed integer linear programming (MILP) model and a simple polynomial time heuristic algorithm to provide the best solution. Zeng et al. (2015) introduced a group sales mechanism for electric vehicle demand response management in the vehicle to grid (V2G) system and designed a group auction transaction mechanism to realize the bidding decision of electric vehicle users. The results show that this mechanism can reduce the system cost. Zhou et al. (2015) proposed an online auction mechanism to solve the demand response in smart grid, expressed the problem of maximizing social welfare as an online optimization problem in the form of natural integer linear programming, and obtained the optimal solution.

Reinforcement learning, as a powerful artificial intelligence tool in sequential decision-making problems, has been increasingly applied to the scheduling, decision-making and energy trading strategies in smart grid. For instance, Zhang et al. (2018) summarized the application research work of deep learning, reinforcement learning and deep reinforcement learning in smart grid. Wan et al. (2018) proposed a deep reinforcement learning real-time scheduling method considering the randomness of EV users' behavior and the uncertainty of real-time electricity price for a single EV user, designed a representation network to extract identification features from electricity prices and a deep Q network to approximate the optimal action value function to determine the optimal strategy.

As introduced above, the research and application of reinforcement learning in smart grid has been very extensive, so its security must be guaranteed. While as the application potential of deep reinforcement learning algorithm is gradually exploited, the adversarial attack and defense against deep reinforcement learning has gradually attracted the attention of scholars. For example, Huang et al. (2017) proved the effectiveness of adversarial attack against the neural network strategy in reinforcement learning. Lin et al. (2017) proposed two adversarial attack methods against the reinforcement learning neural network, and verified the effectiveness of the attack in a typical reinforcement learning environment. Qu et al. (2020) proposed a "minimalist attack" method for the deep reinforcement learning strategy network, and formulated countermeasures by defining three key settings and verified the effect of the attack. Although the above research is aimed at

reinforcement learning, in fact, the adversarial examples are aimed at environmental information mainly based on pictures. In the application of smart grid reinforcement learning, most of the environmental information is digital, so the research in this area needs to be carried out urgently.

# 3 Preliminaries

In this section, we will first introduce the Electric Vehicle double auction model, and then introduce the definition of deep reinforcement learning and adversarial attack.

## 3.1 Electric vehicle double auction energy Trading Market

System Model: In this paper, we consider a Electric Vehicle energy trading market which is shown in **Figure 2**. Specifically, the EVs which need to charge act as buyers, and the EVs with surplus energy and want to discharge to get some profits act as sellers. They are allowed to submit their charing/discharing request freely. The bidding information always including the valuation, demand/supply volume, arriving/departuring time. These bidding information would submitted to the auctioneer, which is acted by microgrid control center. The auctioneer will make a fair, effective determination within these information. In general, the auctioneer is always assumed as a trust platform, which means auctioneer will not steal or tamper the bidding information to threat the auction market. Note that, in our paper, the auction determination rule is considered as the typical double auction mechanism: McAfee mechanism. Due to the limit of the space, we will not introduce the work flow in detail.

In the above auction market, the bidding strategy of EVs is the key issue affecting their profits in the market. However, in such a game market, the information of competitors and environment is constantly changing, and it is impossible to obtain an optimal bidding strategy through traditional methods. And reinforcement learning can get an optimal strategy to adapt to different environments in the future by constantly exploring the environment. Therefore, at present, using reinforcement learning to find the optimal strategy is the mainstream to solve the bidding strategy problem.

Threat Model: Non-etheless, reinforcement learning is an artificial intelligence method, and neural networks are often used in the solution process. The neural network has been proved to be vulnerable to attacks against samples, that is, by adding a little noise to the samples, the training results of the neural network are affected. In our EV double auction market, the reinforcement learning bidding strategy will be attacked by this attack. So in our paper, we assume the adversary is one participant in the auction market. He/she modifies his/her own bidding information, thereby affecting the reading of the environment by reinforcement learning, and thus affecting the bidding strategy of other users. Specific attack methods will be given in **Section 4**.

## 3.2 Deep reinforcement learning

Deep reinforcement learning (DRL) combines the perceptual capability of deep learning (DL) with the decision-making capability of reinforcement learning (RL), where agent perceives information through a higher dimensional space and applies the obtained information to make decisions for complex scenarios. Deep reinforcement learning is widely used because it can achieve direct control from original input to output through end-to-end learning. Initially, due to the lack of training data and computational power, scholars mainly used deep neural networks to downscale high-latitude data, which were later used in traditional reinforcement learning algorithms Lange and Riedmiller (2010). Then Mnih of DeepMind proposed Deep Q-networks (DQN) Mnih et al. (2013), and people gradually started to study them in a deeper level while applying them to a wider range of fields. In recent years, research in deep reinforcement learning has focused on DQN, which combines convolutional neural networks with Q-learning and introduces an experience replay mechanism that allows algorithms to learn control policies by directly sensing high-dimensional inputs. As the most basic reinforcement learning algorithm, because of its good training speed and effect, it is widely used in various practical scenes.

# 4 Adversarial attack method against reinforcement learning -based trading market

## 4.1 Adversarial attack

Deep learning algorithms have been widely used in many fields, but the ensuing security issues deserve attention. Adversarial attack is an important risk. Since the input of deep neural network is a numerical vector, the attacker can maliciously design a targeted numerical vector (called adversarial sample) to make the deep neural network make a misjudgment. In the field of deep learning, we assume that $x$ is the input and $f$ represents a deep neural network, the production of adversarial samples can be represented as:

$$\min_{\delta} d(x, x+\delta) \tag{1}$$

subject to

$$f(x) \neq f(x+\delta) \tag{2}$$

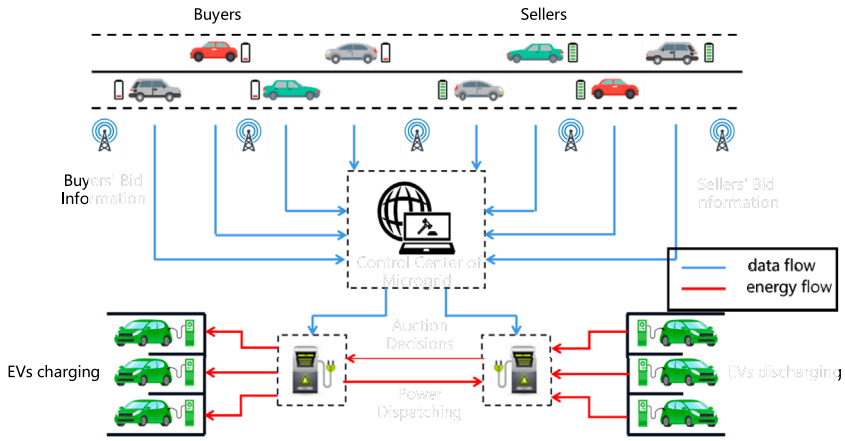



FIGURE 2
Ev double auction model.

where $d$ represents the distance metric, which is calculated by $l$ – norm.

The above equation also shows that the attacker tries to find the minimal perturbation $\delta$ that can make the deep neural network output wrong results.

Deep reinforcement learning (DRL) algorithms integrate deep neural networks based on the theory of reinforcement learning, which also leads to the risk of suffering from adversarial attacks. In value-based RL algorithms, adversarial samples can make the neural network misestimate the value of a specific action at a specific state and guide the agent to choose the wrong action. In policy-based RL algorithms, the attacker can make the agent unable to use the policy gradient to select the optimal policy through the adversarial sample.

## 4.2 Double auction and bidding strategy formalization

In the double auction scene model of electric vehicles, there are mainly the following three participants: auctioneer, buyer and seller. Among them, the microgrid control center serves as the auctioneer of the trading market, the electric vehicle with insufficient electric energy serves as the buyer, and the electric vehicle with surplus electric energy serves as the seller. In the electricity trading market, there are multiple buyers and sellers who can participate in the auction using their mobile devices or Internet of vehicles systems. The winning bidder trades the electric energy through the charging pile, avoiding the transmission loss of electric energy in the traditional power grid system.

According to the characteristics of the auction process, this paper discretizes the transaction process and adopts the integer set $T = \{1, 2, \cdots\}$ to represent the time series in the transaction process. $B$ is the set of buyers, and the number of buyers is $|B|$. $S$ is the set of sellers, and the number of sellers is $|S|$.

At time slot $t$, the actual power demand of the $i$th buyer is $d_{i,t}$, and its bidding information is denoted as a triplet:

$$\chi_{b,i,t} = \left\{i, p_{b,i,t}, q_{b,i,t}\right\}, \quad i \in B \tag{3}$$

where $i$ represents the buyer ID, $p_{b,i,t}$ represents the valuation of one unit electricity submitted by $i$th buyer, and $q_{b,i,t}$ represents the submitted volume.

Similarly, at time slot $t$, the actual power supply of the $j$th seller is $u_{j,t}$, and its bidding information is:

$$\chi_{s,j,t} = \left\{j, p_{s,j,t}, q_{s,j,t}\right\}, \quad j \in S \tag{4}$$

where $j$ represents the seller ID, $p_{s,j,t}$ represents the valuation of one unit electricity submitted by $j$th seller, and $q_{s,j,t}$ represents the submitted volume.

The actual power supply/demand and the submitted volume:

$$q_{b,i,t} \le d_{i,t}, t \in T, \quad i \in B \tag{5}$$

$$q_{s,j,t} \le u_{j,t}, t \in T, \quad j \in S \tag{6}$$

In the energy trading market, electric vehicle users with insufficient and excessive electric energy report bidding information according to their own wishes. The microgrid control center, as the auctioneer, organizes a double auction to determine the winning buyer and seller, and then determine the transaction price and volume of each buyer and seller. Subsequently, the auction results (including the winning buyer/seller) are released to all participants in the system to ensure the fairness and verifiability of the auction.

## 4.3 Rational

Research on adversarial attack theory in deep learning has made some progress. At present, deep learning plays an important role in the field of computer vision. Most of the adversarial attack methods for deep learning are based on the image-based system. The latest research on adversarial attack methods in deep reinforcement learning algorithms is also mainly oriented at game scenarios, and the observations of agents are also images. Note that the application of deep reinforcement learning algorithm is also likely to face the threat of adversarial samples in the scenario of electric vehicle energy trading.

Theoretically, the observation of agents in smart grid is mainly the digital data of electric energy, electricity price and so on. Similarly with the image data, these digital data is also the numerical vectors, and it has fewer input features. This makes it possible to produce adversarial samples for deep reinforcement learning algorithms in energy trading market theoretically. Once affected by the attacker's malicious interference, it may have a negative impact on the benefits of users in the power grid.

In the process of participating in smart grid energy trading, electric vehicle users need to continuously submit their bidding information to participate in double auction, and achieve their optimal benefits through multi-step decision-making. There is correlation between continuous decisions. Deep reinforcement learning algorithm can exactly give full play to its unique advantages in this process. Considering the current situation of actual power grid charging and discharging, it is appropriate to discretize the bidding price and trading volume of energy trading. Deep-Q-network (DQN) is a good choice in power trading scenarios. This paper considers the adversarial attack research on deep Q learning algorithm in power transaction of smart grid.

## 4.4 Adversarial attack method against reinforcement learning-based bidding strategy

In the double auction process, the attacker can affect the benefits of the other auctioneer by maliciously modifying his real demand/supply, making the average cost of the buyer group rise or making the average profit of the seller group decrease. Because each participant in the double auction has limited observations, and some state quantities cannot be explicitly modified, once changed, they are easy to be screened out and lose the attack effect. Therefore, this paper considers that attackers can change the state of agents in the system by submitting false bidding information. As a result, the deep-Q-network selects non-optimal bidding strategies to reduce the average reward.

To be specific, in the bilateral auction market, it is considered that there is an attacker in the buyer group. His purpose is to influence the state observation of other buyers by maliciously modifying his quantity demanded, so other buyers will make non-optimal bidding strategy. Ultimately, it affects the utility of the buyer group. Similarly, for the seller group, this paper also considers the existence of an attacker and studies the impact of the generated adversarial samples on the seller group's revenue. Adversarial attacks in electrical energy trading are shown in **Figure 3**.

At present, most of the adversarial sample production method for deep reinforcement learning borrows from the methods in deep learning. The Fast Gradient Sign Method (FGSM) make adversarial perturbations and add them to the observations, so as to attack the DRL agent. The core idea is to add perturbations along the direction where the deep neural network model gradient changes the most to induce the model output error results. Formally, adversarial samples generated by FGSM can be expressed as follows:

$$x' = x + \varepsilon \cdot sign\left(\nabla_x J(\theta, x, y)\right) \tag{7}$$

where $\varepsilon$ is the size of the disturbance, $J$ represents the cross-entropy loss function, $\theta$ is the parameter of the neural network, $x$ represents the model input, and $y$ represents the sample label (here refers to the optimal action term). The cross-entropy loss function here measures the difference between the distribution of the label y and the distribution that puts all the weight on the optimal action.

Inspired by the original FGSM, to address the problem that the attacker can only modify some observations to avoid being detected by the system, in this paper, a local-FGSM is proposed to make adversarial samples by modifying some components of the agent state vector, which can be expressed as follows:

$$x' = x + \varepsilon \cdot sign\left(\nabla_x J(\theta, x, y)\right) \cdot \mu \tag{8}$$

where $\mu$ is a vector whose dimension is equal to the dimension of input $x$, the value of the dimension corresponding to the component to be modified by the agent state vector is 1, and the rest is 0.

The attack process is shown in **Algorithm 1**.

# 5 Adversarial sample recognition-based reinforcement learning-based energy Trading Mechanism

In this section, we propose a adversarial sample recognition-based reinforcement learning method for the above double auction.

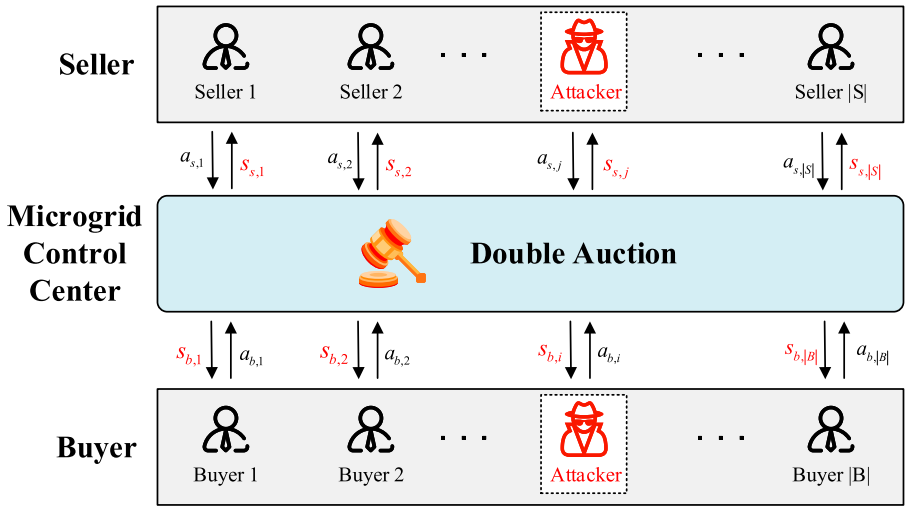



FIGURE 3
Adversarial attacks in electricity trading.

Input the number of buyers $|B|$ and the number of sellers $|S|$.
Select a well-trained buyer Q-network and seller Q-network.
**for** *epoch= 1 to* $E_1$ **do**
  Initialize the states of buyers and sellers.
  **for** *step= 1 to* $v$ **do**
    Make bidding decisions according to the buyer Q-network and get reward $r_{step}$.
    Prepare the cross-entropy loss function J.
    Randomly select a buyer as the attacker and make adversarial sample $x' = x + \varepsilon \cdot sign\left(\nabla_x J\left(\theta, x, y\right)\right) \cdot \mu$
    Make bidding decisions according to the buyer Q-network and get reward $r'_{step}$ again.
    Update the states of buyers and sellers.
  **end**
  Compare the buyers' average cumulative rewards with and without adversarial attacks.
  Initialize the states of buyers and sellers.
  **for** *step= 1 to* $v$ **do**
    Make bidding decisions according to the buyer Q-network and get reward $r_{step}$.
    Prepare the cross-entropy loss function J.
    Randomly select a seller as the attacker and make adversarial sample $x' = x + \varepsilon \cdot sign\left(\nabla_x J\left(\theta, x, y\right)\right) \cdot \mu$
    Make bidding decisions according to the seller Q-network and get reward $r'_{step}$ again.
    Update the states of buyers and sellers.
  **end**
  Compare the sellers' average cumulative rewards with and without adversarial attacks.
**end**

Algorithm 1. The process of local-FGSM adversarial attack.

## 5.1 Markov Decision Process model

We construct the EV electric energy trading double auction scenario as a Markov Decision Process (MDP) with discrete time steps, which can be expressed as a quadruple $\{S, A, P, R\}$.

$S$ stands for the state space. $S_B$ and $S_S$ denote the state space sets of buyers and sellers, respectively. Electric vehicle users can be informed of the total demand and total supply during the current period. Assuming that each participant conducts $v$ auctions in the trading market, the variable $\sigma$ is introduced to indicate whether the participant currently participated in the last auction or not. In time slot $t$, the states of the *ith* buyer and the *jth* seller are denoted as:

$$s(b,i,t) = \{d_{i,t}, D_t, U_t, \sigma\} \tag{9}$$

$$s(s,j,t) = \{u_{j,t}, D_t, U_t, \sigma\} \tag{10}$$

$A$ stands for action space. After acquiring observations at each time slot, buyers and sellers need to submit bidding information to participate in the bilateral auction, and the decision of bidding price and bidding volume will have an impact on their respective profits. In this system, the bidding information submitted by buyers and sellers is regarded as their respective actions. In time slot t, the actions of buyer EV users and seller EV users are denoted as

$$a(b,i,t) = \{p_{b,i,t}, q_{b,i,t}\} \tag{11}$$

$$a(s,j,t) = \{p_{s,j,t}, q_{s,j,t}\} \tag{12}$$

$R$ is the reward function. The immediate reward of the buyer and seller of time slot $t$ is denoted as $r(t)$. For the buyer, if he wins the bid in the bilateral auction, the cost is $p_{b,i,t} \cdot q_{b,i,t}$. The buyer's goal is to keep the cost as low as possible, but win the auction as much as possible. For the seller, if he succeeds in winning the bid in the bilateral auction, then his profit from selling electric energy is $p_{s,j,t} \cdot q_{s,j,t}$; otherwise, if he fails to win the bid, his profit $p_{s,j,t} \cdot q_{s,j,t}$ is 0. Then, the reward function of the buyer in time slot $t$ is denoted by:

$$r(b,i,t) = \begin{cases} -p_{b,i,t} q_{b,i,t}, & i \in M_B \\ -2p_{b,i,t} q_{b,i,t}, & i \notin M_B \end{cases} \tag{13}$$

Setting the buyer's reward function as negative can make the optimal strategy for buyers and sellers using deep reinforcement learning algorithms formally consistent, and the goal is to

Input the number of buyers $|B|$ and the number of sellers $|S|$.
Randomly initialize the parameters of the deep Q-network $\theta$.
Initialize the parameters of the target network $\theta'$.
Initialize the capacity of replay buffer $N$.
Initialize the greedy coefficient $\varepsilon$.
**for** *epoch= 1 to* $E_2$ **do**
  Obtain the initial states of the buyers and sellers.
  **for** *t= 1 to T* **do**
    **if** a random number $x \leq \varepsilon$ **then**
      Randomly choose the action $a_t$ from action space.
    **else**
      Select $a_t = \arg\max_a Q(s_t, a, \theta)$.
    **end if**
    Perform double auction and get the reward $r$ and next state $s_{t+1}$.
    Store the transition $[s_t, a_t, r_t, s_{t+1}]$ in experience replay.
    Randomly choose mini-batch $B$ from the replay buffer.
    Calculate $Q_{evel} = Q(S_t, A_t, \theta)$.
    Calculate $Q_{target} = r_t + \gamma \cdot Q\left(s_{t+1}, \arg_{a'} Q(s_{t+1}, a', \theta), \theta'\right)$.
    Perform the gradient descent on $\sum_{i=1}^{B} (Q_{target} - Q_{evel})^2$ with respect to the Q-network parameters.
    Update the greedy coefficient $\varepsilon$.
    Every C steps update $\theta' = \theta$.
  **end**
**end**

Algorithm 2. The training process of deep Q-learning algorithm in double auction.

maximize their own long-term benefits. The buyer's cost is the absolute value of the reward. For the buyer who fails to win the bid, it may need to spend more money to buy the much-needed power, so a large penalty coefficient is added to it to encourage the buyer to avoid the failure as much as possible.

The seller's reward function is denoted as:

$$r(s,j,t) = p_{s,j,t} \cdot q_{s,j,t} \tag{14}$$

$P$ represents the state transition function. Function $p_t$ is defined as a transition function. The state transition probability from state $s_t$ to state $s_{t+1}$ is expressed as:

$$p_t : s_t \times a_t \rightarrow s_{t+1} \tag{15}$$

## 5.2 Solution *via* reinforcement learning

In the electric energy trading market model designed in this paper, deep Q-learning algorithm is used to learn the optimal bidding strategy for buyers and sellers in microgrid bilateral auction respectively. In order to estimate the state action value function, this paper defines a multi-layer perceptron as a deep-Q-network for buyers and sellers respectively, taking the state as input and the state action value $Q(s,a) \approx Q(s,a,\theta)$ as output, where theta is the neural network parameter. Deep-Q-network is a fully connected neural network with two hidden layers.

In the process of training deep-Q-network, the state $s_t$, action $a_t$, reward $r_t$ and next state $s_{t+1}$ obtained from each interaction with the system environment can form an empirical tuple, denoted as $[s_t, a_t, r_t, s_{t+1}]$. For buyers and sellers, a experience replay is set to store the corresponding experience tuples respectively, and its capacity is $N$.

In addition, a target network with the same structure as the deep-Q-network is defined to solve the correlation and stability problems. Both the deep-Q-network and the target network initially have the same parameters. In the training process, the target network's parameter $\theta'$ is updated to the deep Q network's parameter $\theta$ every $C$ steps. At each training session, a mini-batch sample of size $B$ is sampled from the experience replay and used as input to the main network, and the output is selected to calculate the Q-value:

$$Q_{evel} = Q(S_t, A_t, \theta) \tag{16}$$

The target Q-value is:

$$Q_{target} = r_t + \gamma \cdot Q\left(s_{t+1}, \arg_{a'} Q(s_{t+1}, a', \theta), \theta'\right) \tag{17}$$

The $\gamma$ is a discount factor, indicating the extent to which the future reward affects the current reward. The smaller the $\gamma$, the more the agent focuses on the current reward, and *vice versa*.

The loss function is calculated from the difference between the target Q-value and the estimated Q-value, and the parameters of the main network $\theta$ are updated by gradient descent. The loss function is:

$$L(t) = \sum_{i=1}^{B} \left(Q_{target} - Q_{evel}\right)^2 \tag{18}$$

The training process of deep Q-learning algorithm is shown in **Algorithm 2**.

## 5.3 Defense Strategy Architecture

At present, deep learning mainly achieves defense effect by modifying network structure, objective function or training process, but most defense methods cannot meet the practical application scenarios of DRL. From the perspective of data security and reliability, this paper considers the use of additional network to preprocess the data of the perturbed observation vector and screen out the adversarial samples to ensure the system security.

When EV users participate in the electric energy trading market, they obtain their current state according to the data published by the microgrid control center. Based on the deep Q learning algorithm proposed above, deep Q network is used to help EV users make optimal bidding decisions. In order to avoid adversarial samples that may appear in the process of electric energy trading, the state information of all EV users is screened out by an adversarial sample discriminator before bilateral auction. In this way, only real samples can be allowed to participate in the auction, and then bidding decisions can be made based on the deep-Q-network, and further transaction decisions and scheduling optimization can be made by the microgrid control center. **Figure 4** shows the architecture of the adversarial defense model.

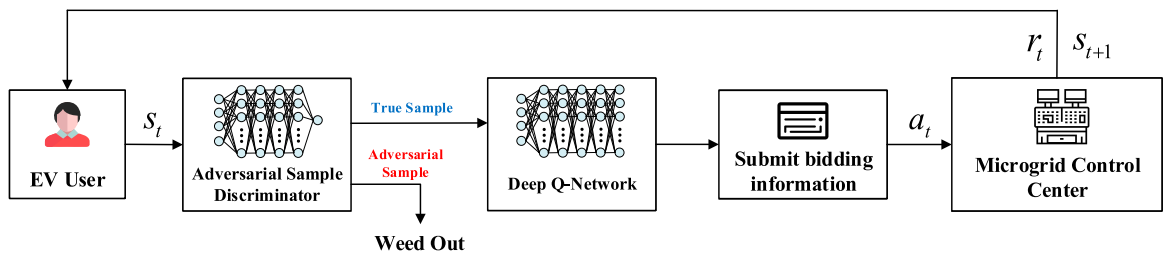



FIGURE 4
Defense strategy architecture.

The adversarial sample discriminator is designed by a fully connected deep neural network with four hidden layers. The specific network structure is shown in the following table. The input layer consists of four neurons corresponding to the four elements of the electric vehicle user's state. The output is 0 or 1, representing the input EV states as adversarial and real samples, respectively.

The adversarial sample discriminator is essentially a binary classifier, which is used to judge whether the input EV state sample is an adversarial sample, and its training process is a supervised learning process. Firstly, select a buyer deep-Q-network and train it well, so that users can make the optimal bidding decision according to it. Then the data set is collected and made. In each episode of electric energy trading, after the user state is initialized, 10 bilateral auctions are conducted successively, and the next state of the user will be obtained after each auction. The local-FGSM is used to make adversarial samples, and the real next time state and adversarial samples are stored with labels being made. After that, by using the collected adversarial samples and real samples, the training set and test set are divided to train the adversarial sample discriminator, and the weight is updated by using the back propagation to reduce the loss function value. Finally, the effectiveness of the adversarial sample discriminator is verified by the test set. Similarly, the adversarial sample discriminator of sellers' Q-network is trained.

The training process of the adversarial sample discriminator is shown in **Algorithm 3**.

Select a well-trained Q-network.
Randomly initialize the parameters of the adversarial sample discriminator $\theta$.
**for** *epoch= 1 to* $E_3$ **do**
  Obtain the initial states of the buyers and sellers.
  **for** $t$*= 1 to* $T$ **do**
    Make bidding decisions according to the Q-network.
    Perform double auction and get the reward $r$ and next state $s_{t+1}$.
    Store real samples $[s_t, a_t, s_{t+1}]$ and make labels.
    Make the adversarial sample $s'_{t+1}$ according to Equation (8).
    Store fake samples $[s_t, a_t, s'_{t+1}]$ and make labels.
  **end**
**end**
Divide the training set and test set.
**for** *epoch= 1 to* $E_4$ **do**
  Sample the mini-batch from the training data.
  Obtain the output of the adversarial sample discriminator.
  Calculate the loss function and update the adversarial sample discriminator.
**end**

Algorithm 3. The training process of the adversarial sample discriminator.

**TABLE 1** Buyer's average cost per round.

| Number of buyers | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| Buyers' average cost (DQN) | 6.5649 | 6.4973 | 5.0431 | 4.8707 |
| Buyers' average cost (random) | 29.7003 | 31.5305 | 32.2234 | 32.5878 |

**TABLE 2** Seller's average profit per round.

| Number of sellers | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| Sellers' average profit (DQN) | 11.0802 | 11.2976 | 11.4352 | 11.5323 |
| Sellers' average profit (random) | 8.1247 | 8.0215 | 7.9866 | 7.9864 |

# 6 Performance evaluation

In this section, we conduct several comprehensive evaluations to verify the performance of our proposed method. In the following, first the evaluation settings are given. Then the results of our proposed method is introduced. Finally, the comparison results are shown.

In this section, we conduct several comprehensive evaluations to verify the performance of our proposed method. In the following, first the evaluation settings are given. Then the results of our proposed method is introduced. Finally, the comparison results are shown.

## 6.1 Evaluation settings

### 6.1.1 Environment settings

Consider a microgrid in which energy trading is performed 10 times per round, that is, each round of bilateral auction is divided into 10 time slots, and 8,000 rounds of bilateral auction are conducted to train the deep Q-learning algorithm. In order

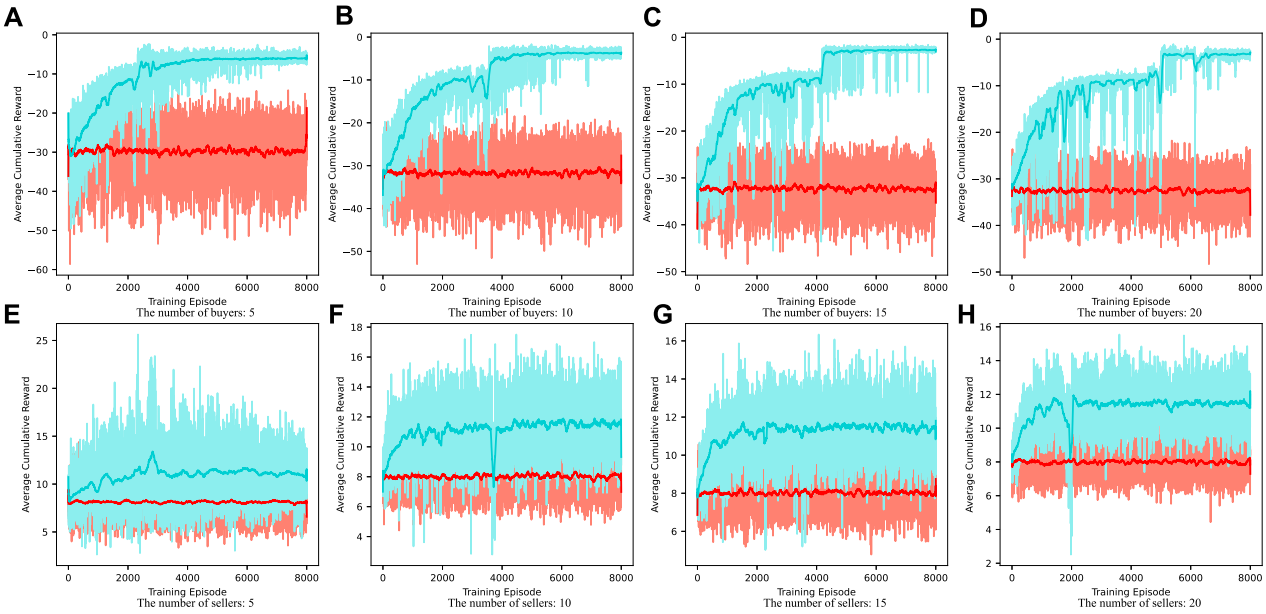



FIGURE 5
Convergence process of deep-Q-network with the numbers of buyers are **(A)** 5; **(B)** 10; **(C)** 15; **(D)** 20, and the numbers of sellers are **(E)** 5; **(F)** 10; **(G)** 15; **(H)** 20.

**TABLE 3** Success rate of adversarial attack.

| Magnitude of perturbation | | 0.1 (%) | 0.2 (%) | 0.3 (%) | 0.4 (%) | 0.5 (%) | 0.6 (%) | 0.7 (%) | 0.8 (%) | 0.9 (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Buyers' Q-network | 5 buyer | 43.3 | 48.1 | 47.3 | 47.7 | 46 | 47.7 | 47.8 | 46.7 | 45.6 |
| | 10 buyer | 34.6 | 33.8 | 33.6 | 34.5 | 32.5 | 36.3 | 31.8 | 32.1 | 32.3 |
| | 15 buyer | 47.1 | 49.6 | 49.6 | 48.5 | 50 | 48.5 | 48.7 | 46.5 | 48.2 |
| | 20 buyer | 49.3 | 52.1 | 47.5 | 48.1 | 48.8 | 50.3 | 49 | 46.9 | 49 |
| Sellers' Q-network | 5 seller | 23.5 | 36 | 44.4 | 48.2 | 50.1 | 52.5 | 51.9 | 49.5 | 47.8 |
| | 10 seller | 39.5 | 48 | 56.7 | 59.7 | 57.1 | 62.2 | 60.3 | 60.4 | 62.3 |
| | 15 seller | 39.9 | 56.4 | 61.1 | 63.5 | 67.5 | 66.8 | 66.2 | 66.2 | 66.7 |
| | 20 seller | 45.4 | 59.4 | 61.7 | 66.7 | 65.8 | 66.4 | 68.8 | 66.7 | 68.8 |

to make the simulation fit the actual transaction as much as possible and avoid the dimension explosion problem, this paper discretizes the bid price and bid volume of the buyer and seller. The bid price is selected from [0.6, 1.5] with a spacing of .1, a total of 10 bid price schemes, and the bid quantity is selected from [0.5, 5] with a spacing of .5, a total of 10 bid volume schemes. In order to facilitate the simulation, the number of EVs of the buyer is assumed to be equal to the number of EVs of the seller in each training process, and the number of the two parties is considered to be 5, 10, 15 and 20 respectively. In each round of 10 auctions, the emerging demand or supply generated by each participant is a discrete number chosen from the set (0.5, 1.5]. Assuming that the unmet demand or supply from the previous step will be inherited to the next auction with an inheritance rate of .9, then

$$d_{i,t+1} = 0.9\left(d_{i,t} - w_{b,i,t}\right) + \varphi, \quad i \in B \tag{19}$$

$$u_{j,t+1} = 0.9\left(u_{j,t} - w_{s,j,t}\right) + \varphi, \quad j \in S \tag{20}$$

where, $w_{b,i,t}$ and $w_{s,j,t}$ respectively represent the transaction volume of the *ith* buyer and the *jth* seller in time slot $t$, and the value of $\varphi$ satisfies the uniform distribution on (0.5, 1.5].

### 6.1.2 Reinforcement learning settings

For the deep Q-learning algorithm, the learning rate is set to .001, the buyer discount rate is set to .99, the seller discount rate is set to .7, and the time interval for replacing the target

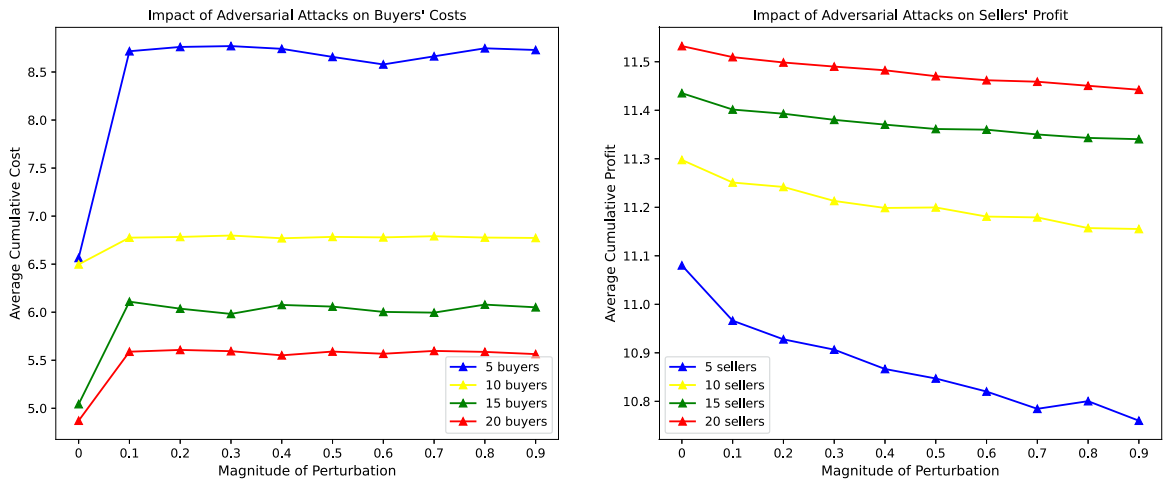



FIGURE 6
Effect of adversarial attacks.

network parameter $\theta'$ with the deep-Q-network parameter $\theta$ is set to 5. The size of the experience replay buffer is set to 3,000 for both buyer and seller, and the mini-batch size $B$ sampled from it during training is set to 32. The input layer of the deep-Q-network is set to four neurons, the output layer is set to 100 neurons, and the number of neurons in the four hidden layers is 20, 512, 256 and 128, respectively. The greedy coefficient satisfies the following relation:

$$\varepsilon = \varepsilon_2 + (\varepsilon_1 - \varepsilon_2)\, e^{-\frac{t}{8000}} \tag{21}$$

where $\varepsilon_1$ and $\varepsilon_2$ are the values of .99 at the beginning of training and 0 at the end of training, respectively.

## 6.2 Effectiveness analysis of deep Q-Learning algorithms

In the case of different number of participants, deep Q-learning algorithm and random strategy are respectively used to compare the average cost per round of buyers and the average profit per round of sellers in the last 1,000 rounds. The results are shown in Table 1 and Table 2.

It can be seen from Table 1 and Table 2 that buyers and sellers can obtain more significant benefits when making bidding decisions based on deep-Q-network compared with random strategy. The average cost of buyers is the negative value of the cumulative reward in each round. It can be seen from Table 1 that under the deep Q-learning algorithm, with the increase of the number of buyers, the average cost of buyers participating in electric energy trading will also decrease. The average profit of sellers is the cumulative reward in each round. It can be seen from Table 2 that under the deep Q-learning algorithm, with the increase of the number of sellers, the average profit of sellers participating in electric energy trading will also rise. This shows the effectiveness of the algorithm and fully considers the willingness and interests of the participants. It can also encourage EV users to participate in the electric energy trading market and contribute to the peak regulation of the power grid. The convergence process of deep-Q-network training is shown in Figure 5.

## 6.3 Effectiveness evaluation of adversarial attacks

The effectiveness evaluation of adversarial attacks can be considered from two aspects. One is the success rate, and the other is the extent to which adversarial attacks affect participants' utilities. For the setting of user states in this paper, the attacker affects other users' state observations mainly by maliciously modifying its own demand/supply. Therefore, in local-FGSM for buyer-Q-network, the values in the first two dimensions of vector $u$ are one and the rest are 0. The value of the first and third dimensions of the vector $u$ of local-FGSM for the seller-Q-network is 1.

When attacking the buyer Q-network, a buyer is selected as the attacker in each auction, and its state is modified to affect the state observation of other buyers, then a non-optimal bidding strategy is selected to participate in the bilateral auction. Similarly, the seller Q-network is also attacked. If the buyer's average cumulative cost per turn increases or the seller's average cumulative profit per turn decreases, the attack is successful. The success rate against the attack is shown in Table 3.

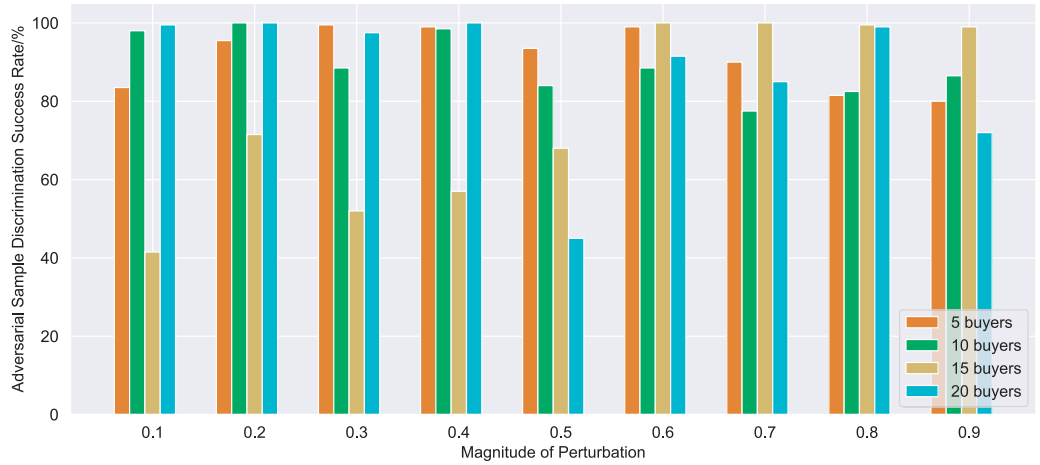



FIGURE 7
Effect of adversarial sample discriminator for buyer Q-Network.

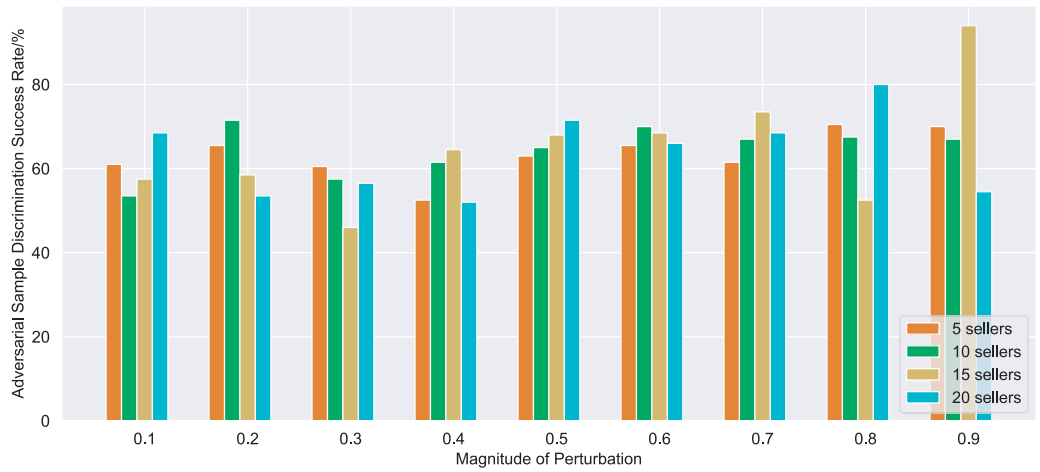



FIGURE 8
Effect of adversarial sample discriminator for seller Q-Network.

The impact of adversarial attacks on user benefits is shown in Figure 6. As can be seen from Figure 5, when adversarial samples are added, the average cumulative cost of buyers per round increases, especially when the number of buyers is small, the impact is greater. When adversarial samples are added, the average cumulative profit of sellers in each round decreases, and with the increase of disturbance size, the profit becomes lower and lower. When the number of sellers is small, the profit decreases more significantly.

## 6.4 Effectiveness evaluation of defense strategy

The adversarial sample discriminator is trained by supervised learning, and the well-trained buyer Q-network and seller Q-network are selected to collect 20,000 real samples and 4,000 adversarial samples in the process of adversarial attack for 2000 training times. The learning rate is set to .001. The final defense effect of the adversarial sample discriminator is shown in Figure 7 and Figure 8.

It can be seen from Figure 7 and Figure 8 that the adversarial defense method proposed in this paper can achieve defense effect in most cases. The buyer adversarial sample discriminator has a good screening effect on adversarial samples with different disturbance sizes, and can basically achieve a screening success rate of more than 80% in trading scenarios with different number of buyers. Compared with buyer adversarial sample discriminator, seller adversarial sample discriminator has a poor performance, but the success rate of adversarial sample screening generally reaches more than 60%, and it can also play a good adversarial defense effect in most cases.

## 7 Conclusion

In this paper, focusing the EV double auction market, we study the security issue of bidding strategy based on reinforcement learning raised by adversarial example. First, we construct a Markov Decision Process for EV energy trading, and use DQN to solve this problem. Second, we design a local-fast gradient sign method to try to counter attacks on DQN from the perspective of attackers. Third, from the perspective of defenders, we choose the method of adding additional network, and use the deep neural network to build the adversarial example discriminator to screen the adversarial example. Finally, the simulation results shows that adversarial example would have an impact on the deep reinforcement learning algorithm, and different disturbance sizes will have different degrees of negative impact on market profits. While after adding the discriminant network, it can almost completely resist such attacks.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

DL: Conceptualization, Methodology, Investigation, Results Analysis, Writing—Original Draft; QY: Conceptualization, Supervision, Writing—Review and Editing; ZP: Survey of Methods, Simulation; XL: Results Analysis; LM: Data Processing, Writing—Review and Editing.

## Funding

The work was supported in part by Key Research and Development Program of Shaanxi under Grants 2022GY-033, in part by the National Science Foundation of China under Grants 61973247 and 61673315, in part by China Postdoctoral Science Foundation 2021M692566, in part by the operation expenses for universities' basic scientific research of central authorities xzy012021027.

## Conflict of interest

Authors LM and XL were employed by the company State Grid Information and Telecommunication Group Co., LTD, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Albadi, M. H., and El-Saadany, E. F. (2007). "Demand response in electricity markets: An overview," in 2007 IEEE power engineering society general meeting (IEEE), Tampa, FL, USA, 24-28 June 2007, 1–5. doi:10.1109/PES.2007.385728

An, D., Zhang, F., Yang, Q., and Zhang, C. (2022). Data integrity attack in dynamic state estimation of smart grid: Attack model and countermeasures. *IEEE Trans. Automation Sci. Eng.* 19, 1631–1644. doi:10.1109/TASE.2022.3149764

Bandyszak, T., Daun, M., Tenbergen, B., Kuhs, P., Wolf, S., and Weyer, T. (2020). Orthogonal uncertainty modeling in the engineering of cyber-physical systems. *IEEE Trans. Automation Sci. Eng.* 17, 1–16. doi:10.1109/TASE.2020.2980726

Barto, A. G., Sutton, R. S., and Watkins, C. (1989). *Learning and sequential decision making*. Amherst, MA: University of Massachusetts.

Cheng, J., Chu, F., and Zhou, M. (2018). An improved model for parallel machine scheduling under time-of-use electricity price. *IEEE Trans. Automation Sci. Eng.* 15, 896–899. doi:10.1109/TASE.2016.2631491

Croce, D., Giuliano, F., Tinnirello, I., Galatioto, A., Bonomolo, M., Beccali, M., et al. (2017). Overgrid: A fully distributed demand response architecture based on overlay networks. *IEEE Trans. Automation Sci. Eng.* 14, 471–481. doi:10.1109/TASE.2016.2621890

Ding, J.-Y., Song, S., Zhang, R., Chiong, R., and Wu, C. (2016). Parallel machine scheduling under time-of-use electricity prices: New models and optimization approaches. *IEEE Trans. Automation Sci. Eng.* 13, 1138–1154. doi:10.1109/TASE.2015.2495328

Erhel, S., and Jamet, E. (2016). The effects of goal-oriented instructions in digital game-based learning. *Interact. Learn. Environ.* 24, 1744–1757. doi:10.1080/10494820.2015.1041409

Esmat, A., de Vos, M., Ghiassi-Farrokhfal, Y., Palensky, P., and Epema, D. (2021). A novel decentralized platform for peer-to-peer energy trading market with blockchain technology. *Appl. Energy* 282, 116123. doi:10.1016/j.apenergy.2020.116123

Giaconi, G., Gündüz, D., and Poor, H. V. (2018). Smart meter privacy with renewable energy and an energy storage device. *IEEE Trans. Inf. Forensics Secur.* 13, 129–142. doi:10.1109/TIFS.2017.2744601

Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). *Explaining and harnessing adversarial examples*. arXiv preprint arXiv:1412.6572.

Grigsby, L. L. (2007). *Electric power generation, transmission, and distribution*. Boca Raton: CRC Press.

Hahn, R. W., and Stavins, R. N. (1991). Incentive-based environmental regulation: A new era from an old idea. *Ecol. LQ* 18, 1.

Haller, M., Ludig, S., and Bauer, N. (2012). Bridging the scales: A conceptual model for coordinated expansion of renewable power generation, transmission and storage. *Renew. Sustain. Energy Rev.* 16, 2687–2695. doi:10.1016/j.rser.2012.01.080

Hong, Z., Wang, R., Ji, S., and Beyah, R. (2019). Attacker location evaluation-based fake source scheduling for source location privacy in cyber-physical systems. *IEEE Trans. Inf. Forensics Secur.* 14, 1337–1350. doi:10.1109/TIFS.2018.2876839

Hosseini, S. M., Carli, R., and Dotoli, M. (2021). Robust optimal energy management of a residential microgrid under uncertainties on demand and renewable power generation. *IEEE Trans. Automation Sci. Eng.* 18, 618–637. doi:10.1109/TASE.2020.2986269

Huang, S., Papernot, N., Goodfellow, I., Duan, Y., and Abbeel, P. (2017). *Adversarial attacks on neural network policies*. arXiv preprint arXiv:1702.02284.

Huang, W., Zhang, N., Kang, C., Li, M., and Huo, M. (2019). From demand response to integrated demand response: Review and prospect of research and application. *Prot. Control Mod. Power Syst.* 4, 12–13. doi:10.1186/s41601-019-0126-4

Jin, C., Tang, J., and Ghosh, P. (2013). Optimizing electric vehicle charging with energy storage in the electricity market. *IEEE Trans. Smart Grid* 4, 311–320. doi:10.1109/tsg.2012.2218834

Kim, H., Lee, J., Bahrami, S., and Wong, V. W. (2019). Direct energy trading of microgrids in distribution energy market. *IEEE Trans. Power Syst.* 35, 639–651. doi:10.1109/tpwrs.2019.2926305

Lange, S., and Riedmiller, M. (2010). "Deep auto-encoder neural networks in reinforcement learning," in The 2010 international joint conference on neural networks (IJCNN), Barcelona, Spain, 18-23 July 2010 (IEEE), 1–8. doi:10.1109/IJCNN.2010.5596468

Lin, Y.-C., Hong, Z.-W., Liao, Y.-H., Shih, M.-L., Liu, M.-Y., and Sun, M. (2017). *Tactics of adversarial attack on deep reinforcement learning agents*. arXiv preprint arXiv:1703.06748.

Liu, P., Zang, W., and Yu, M. (2005). Incentive-based modeling and inference of attacker intent, objectives, and strategies. *ACM Trans. Inf. Syst. Secur. (TISSEC)* 8, 78–118. doi:10.1145/1053283.1053288

Liu, Y., Duan, J., He, X., and Wang, Y. (2018). Experimental investigation on the heat transfer enhancement in a novel latent heat thermal storage equipment. *Appl. Therm. Eng.* 142, 361–370. doi:10.1016/j.applthermaleng.2018.07.009

Miao, L., Wen, J., Xie, H., Yue, C., and Lee, W.-J. (2015). Coordinated control strategy of wind turbine generator and energy storage equipment for frequency support. *IEEE Trans. Industry Appl.* 51, 2732–2742. doi:10.1109/tia.2015.2394435

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). *Playing atari with deep reinforcement learning*. arXiv preprint arXiv:1312.5602.

Mohan, S., and Laird, J. (2014). "Learning goal-oriented hierarchical tasks from situated interactive instruction," in Proceedings of the AAAI Conference on Artificial Intelligence. doi:10.1609/aaai.v28i1.8756

Ng, K.-H., and Sheble, G. B. (1998). Direct load control-a profit-based load management using linear programming. *IEEE Trans. Power Syst.* 13, 688–694. doi:10.1109/59.667401

Pyka, A. (2002). Innovation networks in economics: From the incentive-based to the knowledge-based approaches. *Eur. J. Innovation Manag.* 5, 152–163. doi:10.1108/14601060210436727

Qu, X., Sun, Z., Ong, Y.-S., Gupta, A., and Wei, P. (2020). Minimalistic attacks: How little it takes to fool deep reinforcement learning policies. *IEEE Trans. Cognitive Dev. Syst.* 13, 806–817. doi:10.1109/tcds.2020.2974509

Roijers, D. M., Vamplew, P., Whiteson, S., and Dazeley, R. (2013). A survey of multi-objective sequential decision-making. *J. Artif. Intell. Res.* 48, 67–113. doi:10.1613/jair.3987

Ruiz, N., Cobelo, I., and Oyarzabal, J. (2009). A direct load control model for virtual power plant management. *IEEE Trans. Power Syst.* 24, 959–966. doi:10.1109/tpwrs.2009.2016607

Samadi, P., Mohsenian-Rad, A.-H., Schober, R., Wong, V. W., and Jatskevich, J. (2010). "Optimal real-time pricing algorithm based on utility maximization for smart grid," in 2010 First IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, USA, 04-06 October 2010 (IEEE), 415–420. doi:10.1109/SMARTGRID.2010.5622077

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., et al. (2013). *Intriguing properties of neural networks*. arXiv preprint arXiv:1312.6199.

Wan, Z., Li, H., He, H., and Prokhorov, D. (2018). Model-free real-time ev charging scheduling based on deep reinforcement learning. *IEEE Trans. Smart Grid* 10, 5246–5257. doi:10.1109/tsg.2018.2879572

Wu, Y., Tan, X., Qian, L., Tsang, D. H., Song, W.-Z., and Yu, L. (2015). Optimal pricing and energy scheduling for hybrid energy trading market in future smart grid. *Ieee Trans. industrial Inf.* 11, 1585–1596. doi:10.1109/tii.2015.2426052

Yu, B., Lu, J., Li, X., and Zhou, J. (2022). Salience-aware face presentation attack detection via deep reinforcement learning. *IEEE Trans. Inf. Forensics Secur.* 17, 413–427. doi:10.1109/TIFS.2021.3135748

Zeng, M., Leng, S., Maharjan, S., Gjessing, S., and He, J. (2015). An incentivized auction-based group-selling approach for demand response management in v2g systems. *IEEE Trans. Industrial Inf.* 11, 1554–1563. doi:10.1109/tii.2015.2482948

Zhang, D., Han, X., and Deng, C.Taiyuan University of Technology, and China Electric Power Research Institute (2018). Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst.* 4, 362–370. doi:10.17775/cseejpes.2018.00520

Zhang, K., Sprinkle, J., and Sanfelice, R. G. (2016). Computationally aware switching criteria for hybrid model predictive control of cyber-physical systems. *IEEE Trans. Automation Sci. Eng.* 13, 479–490. doi:10.1109/TASE.2016.2523341

Zhang, W., Song, K., Rong, X., and Li, Y. (2019). Coarse-to-fine uav target tracking with deep reinforcement learning. *IEEE Trans. Automation Sci. Eng.* 16, 1522–1530. doi:10.1109/TASE.2018.2877499

Zhao, S., Li, F., Li, H., Lu, R., Ren, S., Bao, H., et al. (2021). Smart and practical privacy-preserving data aggregation for fog-based smart grids. *IEEE Trans. Inf. Forensics Secur.* 16, 521–536. doi:10.1109/TIFS.2020.3014487

Zhou, R., Li, Z., and Wu, C. (2015). "An online procurement auction for power demand response in storage-assisted smart grids," in 2015 IEEE Conference on Computer Communications (INFOCOM), Hong Kong, China, 26 April 2015 - 01 May 2015 (IEEE), 2641–2649. doi:10.1109/INFOCOM.2015.7218655

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fenrg.2022.1046756



OPEN ACCESS

EDITED BY
Dou An,
MOE Key Laboratory for Intelligent Networks and Network Security, China

REVIEWED BY
Neeraj Kumar Singh,
Larsen & Toubro, India
Hanlin Zhang,
Qingdao University, China

*CORRESPONDENCE
Xialei Zhang,
xl.zhang@sxu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 17 September 2022
ACCEPTED 31 October 2022
PUBLISHED 13 January 2023

CITATION
Zhang X, Chang D and Liao X (2023), A detection model of scaling attacks considering consumption pattern diversity in AMI.
*Front. Energy Res.* 10:1046756.
doi: 10.3389/fenrg.2022.1046756



# A detection model of scaling attacks considering consumption pattern diversity in AMI

Xialei Zhang[1]*, Da Chang[1] and Xuening Liao[2,3]

[1]School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi, China, [2]School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi, China, [3]Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an, Shaanxi, China

As an important branch of the Internet of Things, the smart grid has become a crucial field of modern information technology. It realizes the two-way information flow and power flow by integrating the advanced metering infrastructure (AMI) and distributed energy resources, which greatly improves users' participation. However, owing to smart meters, the most critical components of AMI, are deployed in an open network environment, AMI is a potential target for data integrity attacks. Among various attack types, the scaling attack is the most typical one, because it can be used as a general expression for most of other ones. By launching a scaling attack, adversaries can randomly reduce hourly reported values in smart meters, thereby causing economic losses. A number of research efforts have been devoted to detecting data integrity attacks. Nonetheless, most of the existing investigations focus on all attack types, and little attention has been paid to a detection strategy specially designed for scaling attacks. Our contribution addresses this issue in this paper and hence, developing a detection model of scaling attacks considering consumption pattern diversity (*SA2CPD*), to ensure that scaling attacks can be effectively detected when users have multiple consumption patterns. To be specific, we leverage Kmeans to distinguish different consumption patterns, and then the consumption intervals can be extracted to binarize the data. We divide time periods in every day into two categories based on the binarization values, and use one of them with the greatest information gain to construct a decision tree for judgment. Both theoretical and simulation results based on the GEFCom2012 dataset show that our *SA2CPD* model has a higher *F1 score* than the decision tree model without considering consumption pattern diversity, the KNN model and the Naive Bayes model.

KEYWORDS

**smart grid, smart meter, advanced metering infrastructure (AMI), scaling attack detection, consumption pattern diversity, binarize**

## 1 Introduction

The traditional power grid has a history of more than 100 years. Owing to its disadvantages of one-way information flow, low user participation, etc., it has gradually been unable to adapt to the modern society. As a consequence, the smart grid emerges as the times require, which not only incorporates renewable energy resources such as solar energy and wind energy to support multiple energy supply, but also integrates the advanced metering infrastructure (AMI) to control the power layer, realizing the two-way flow of information and power (Zanetti et al., 2019; Rouzbahani et al., 2020; Choi et al., 2021; Sarenche et al., 2021; Chaudhry et al., 2022; Huang et al., 2022; Park et al., 2022). Specifically, smart meters which play a vital role in AMI are deployed in demand sides, i.e., users, to collect and upload information about power consumption and supply to the utility. The utility then makes decisions on real-time pricing, and energy scheduling, among others, based on the uploaded information, and then feeds back the decisions to guide users supply and consume electricity smartly (Singh et al., 2017; Zheng et al., 2018; Choi et al., 2021; Chaudhry et al., 2022; Huang et al., 2022; Park et al., 2022; Verma et al., 2022). However, as smart meters are deployed in an open network environment, they are vulnerable to data integrity attacks, by launching which an adversary can seriously endanger the safe operation of the smart grid through tampering with the information in smart meters (Jokar et al., 2016; Hu et al., 2019; Jakaria et al., 2019; Yao et al., 2019; Zheng et al., 2019; Rouzbahani et al., 2020; Tehrani et al., 2020; Bhattacharjee and Das, 2021; Singh and Mahajan, 2021; Sun et al., 2021; Yan and Wen, 2021; Chaudhry et al., 2022; Mudgal et al., 2022; Verma et al., 2022). Therefore, the research on data integrity attacks detection is of significant importance and has become a research hotspot in the field of the smart grid (Jokar et al., 2016; Zheng et al., 2019; Tehrani et al., 2020; Ibrahem et al., 2021).

Recently, much work has been conducted on the detection for data integrity attacks in AMI, which is mainly divided into three categories (Jiang et al., 2014; Jokar et al., 2016; Yao et al., 2019), including state-based (Huang et al., 2013; Salinas et al., 2014; Leite and Mantovani, 2018; Lo and Ansari, 2013; McLaughlin et al., 2013; Aziz et al., 2020; Bhattacharjee et al., 2021b,a), game theory-based (Cardenas et al., 2012; Yang et al., 2016; Wei et al., 2018, 2017; Paul et al., 2020) and classification-based (Jokar et al., 2016; Singh et al., 2017; Ismail et al., 2018; Yeckle and Tang, 2018; Zheng et al., 2018; Fernandes et al., 2019; Jakaria et al., 2019; Punmiya and Choe, 2019; Zheng et al., 2019; Rouzbahani et al., 2020; Tehrani et al., 2020; Yan and Wen, 2021). As a result of the popularity of artificial intelligence technologies, the feasibility of machine learning to detect attacks in AMI has attracted much attention of a large number of researchers. Therefore, classification-based detection has gradually become a mainstream technology. For example, Jokar et al. (Jokar et al., 2016) proposed a data integrity attacks detection model based on SVM. They compared the reported total consumption value with the actual total consumption value to find out the suspicious area, and then used the historical data and synthetic attack data to train SVM. Tehrani et al. (Tehrani et al., 2020) took sampling values of 24 h and their mean, standard deviation, minimum and maximum values as features. Firstly, they used Kmeans for clustering, and then generated false data according to the synthetic attack method proposed in the literature (Jokar et al., 2016) to construct a complete dataset for training and testing the decision tree, random forest and gradient boosting. Nevertheless, the existing studies all have the problem of dealing with different attack types indiscriminately, but different attacks have different characteristics, and there is currently no algorithm that can contrapuntally detect scaling attacks. Thus, it is vital to design a detection model specially for scaling attacks.

To fill this gap, in this paper we propose a detection model of scaling attacks considering consumption pattern diversity in AMI (*SA2CPD*). Compared with existing schemes which deal with all attack types indiscriminately, our *SA2CPD* model focuses on the scaling attack only, as the scaling attack is a typical data integrity attack. The reason is that the scaling attack can not be easily judged by manual methods, and can be used as a generalization of several other attack types. In addition, we also consider consumption pattern diversity of users caused by living conditions, work and rest habits, etc. Specifically, we first leverage the clustering technology to differentiate different consumption patterns and extract consumption intervals. Then the data are discretized by binarization on the basis of consumption intervals, which can distinguish normal data from false data. Finally, the discretized data are used as the input of the decision tree. In this step, we divide the 24 time periods of a day into two categories, and the decision tree makes judgement in accordance with one of the two corresponding to the time periods with the greatest information gain, to successfully detect the false data injected by scaling attacks.

To further validate the effectiveness and efficiency of our *SA2CPD* model, we conduct a performance simulation based on the GEFCom2012 dataset (Hong, 2014). The consumer in our experimental scenario has three different consumption patterns, and each pattern has 1,586 data. We use the widely adopted criteria as comparison metrics including the False Positive Rate (*FPR*), False Negative Rate (*FNR*) and *F1 score*, which can comprehensively measure the recall and the precision. We design two experiments. In the first experiment, we test the performance of our model when the proportion of false data in the test set is varied from 10% to 80%. The result verifies the effectiveness of our detection model and is accord with our theoretical analysis. In the second experiment, through the comparative experiments

with the decision tree model without considering consumption pattern diversity, the KNN model and the Naive Bayes model, the results show that our model is more efficient. For example, when the attack proportion is 50%, our *FPR* and *FNR* are 0.2% and 6.78%, and the *F1 score* is 96.38%, while those of the Naive Bayes model are 0.18%, 11% and 94% respectively, and those of the KNN model are 0.02%, 13% and 92.96%.

The remainder of the paper is organized as follows: In **Section 2**, we present the network and threat models, and then briefly describe the related machine learning algorithms. In **Section 3**, we present the detailed design of our *SA2DCP* model. In **Section 4**, we describe the metrics and conduct performance analysis in comparison with the decision tree model without considering consumption pattern diversity, the KNN model and the Naive Bayes model. In **Section 5**, we show experimental results to validate the effectiveness and efficiency of *SA2CPD* model. In **Section 6**, we discuss other related issues. Related literature is reviewed in **Section 7**. Finally, we conclude the paper in **Section 8**.

# 2 Preliminary

In this section, we first present the network and threat models and then briefly introduce the Kmeans and decision tree model used in *SA2CPD*.

## 2.1 Network models

AMI plays a crucial role in the smart grid and greatly promotes the intelligence of the power grid. As shown in **Figure 1**, AMI consists of smart meters, i.e., SM1-SM4, data concentrators (DC), the utility and communication networks between them (Jiang et al., 2014; Huang et al., 2022). The communication networks in AMI enable the smart grid to realize the two-way flow of information. Specifically, the smart meter, domestic appliances and distributed renewable equipments in a user's home form a home area network (HAN). The smart meter is responsible for collecting the consumption and supply information of domestic appliances and renewable equipments. A neighborhood area network (NAN) consists of a data concentrator and adjacent smart meters. The DC collects the information from all smart meters in the NAN over wireless networks, and then forwards it to the utility through wired networks such as optic fiber in the wide area networks (WAN). Based on the received information, the utility makes decisions such as the time-of-use price, the optimal electricity plan which are conducive to the operation of the smart grid, and finally feeds back the decisions to users. Users can view the feedback information through smart meters and conduct corresponding power supply or consumption. For example, a supply-user determines his optimal power supply according to the decision information and a demand-user decides when to use electricity to save money according to the real-time price.

## 2.2 Threat models

Data integrity attacks in AMI mainly include six types (Jokar et al., 2016; Hu et al., 2019; Zanetti et al., 2019; Zheng et al., 2019; Yan and Wen, 2021) from $h_1$ to $h_6$ formalized as

$$\begin{cases} h_1(x_t) = \alpha x_t, \alpha = random(0.1, 0.8) \\ h_2(x_t) = \beta_t x_t \\ \beta_t = \begin{cases} 0 & start\_time < t < end\_time \\ 1 & else \end{cases} \\ start\_time = random(0, 24) \\ end\_time = random(start\_time, 24) \\ h_3(x_t) = \gamma_t x_t, \gamma_t = random(0.1, 0.8) \\ h_4(x_t) = \gamma_t mean(x), \gamma_t = random(0.1, 0.8) \\ h_5(x_t) = mean(x) \\ h_6(x_t) = x_{24-t} \end{cases}. \quad (1)$$

$h_1$ represents contaminating the hourly reported value of meters through multiplying by a same random number. $h_2$ represents that adversaries control a smart meter to report its measured values as 0 for a certain duration. $h_3$ represents manipulating the hourly reported value of meters through multiplying by a different random number. $h_5$ expresses reporting values of meters as the mean value of a day. $h_4$ multiplies each reported value by a different random number on the basis of $h_5$. $h_6$ reverses the order of reported values of the meter in a day. We divide these six attack types into two categories, Category 1 and Category 2. Category 1 is that the damage is caused by the reduction of the reported total consumption including $h_1 - h_4$, and Category 2 is that the total amount remains unchanged including $h_5$ and $h_6$. Because the majority of the attack types are caused by changing the reported total consumption, we focus on Category 1, namely $h_1 - h_4$. Furthermore, the damage of $h_2$ is extremely obvious and the effects of $h_1$ and $h_4$ can be represented by $h_3$. Therefore, we focus on $h_3$ and name it the scaling attacks.

## 2.3 Kmeans

Kmeans is one of the most commonly used methods in clustering, which can achieve the best distinction between classes based on the similarity of distances between points (Jain, 2010). The goal of the Kmeans is to divide a dataset into $k$ classes, so that each point is closest to the center of the class it belonged to. After all points are divided once, the class center is recalculated

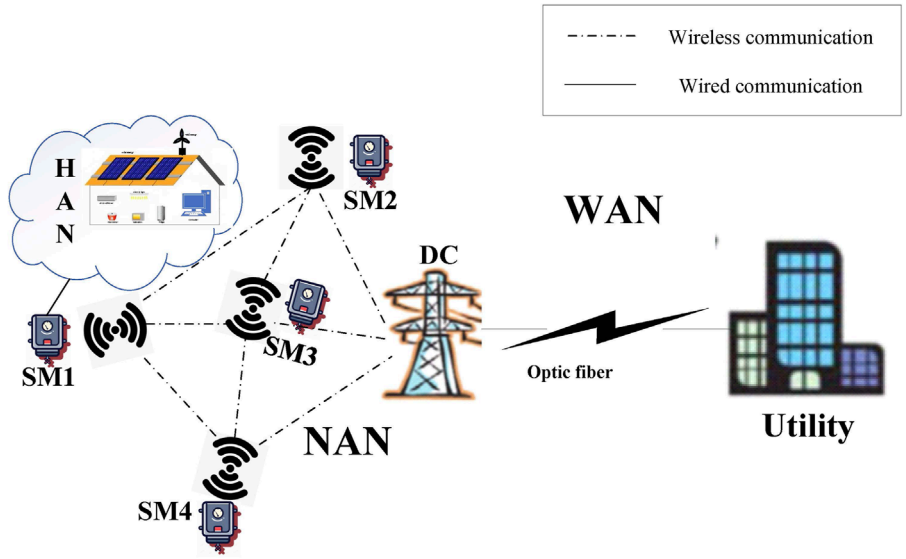



FIGURE 1
The architecture of AMI.

according to points within each class, and then iteratively assign points and update the class center until it no longer changes.

## 2.4 Decision tree

The decision tree is a popular algorithm often used to classify or regress data, which learns a large number of training samples to construct a tree and judges the selected features in the tree in turn, so as to determine the label of samples (Safavian and Landgrebe, 1991). A decision tree consists of a root node, internal nodes and leaf nodes. The predictions for all samples are judged sequentially from the root node. After a series of judgments in internal nodes, the marked results can be obtained at the leaf node. The judgement from the root node to the leaf node is a process in which the uncertainty of information is continuously reduced.

It is how to select the most appropriate features to make use of the least judgment to draw a conclusion, so as to avoid the decision tree being too large, that is the most important thing in the process of constructing a decision tree. Decision trees use information gain to minimize the uncertainty of information (i.e., information entropy) in each judgment, which can be formalized as

$$g(D,A) = H(D) - H(D|A). \tag{2}$$

In Eq 2, $g(D,A)$ is the information gain of feature $A$ to dataset $D$, $H(D)$ is the information entropy of dataset $D$ before judgment, and $H(D|A)$ is the empirical conditional entropy of $D$ when the feature $A$ is given. All notations used in this paper are defined in Table 1. It is worth noting that the $t^{th}$ time period represents the $(t-1)^{th}$ hour to the $t^{th}$ hour.

# 3 The detection model of scaling attacks considering consumption pattern diversity in AMI

In this section, we first present the basic idea of the proposed detection model of scaling attacks considering consumption pattern diversity in AMI (*SA2CPD*). Then, we show details of the *SA2CPD*.

## 3.1 Basic idea

Recall that adversaries will randomly inject reduced values into original data when launching scaling attacks, so we can distinguish normal data from false data as long as we find power consumption intervals of the original normal data and use these intervals as the boundary. Moreover, due to the differences in living conditions, work and rest habits, etc., each user has different electricity consumption patterns, so clustering need to be performed before classification to find multiple normal intervals of the original data. After that, we use intervals to binarize the data so that all values are 0 or 1. Finally, the binarized data is involved in training and classification judgment. Based on the above statement, our *SA2CPD* consists of the following three steps. First, find out $k$ consumption patterns by clustering and then extract consumption intervals. Second, generate false

TABLE 1 Notations.

| | |
|---|---|
| $h(\cdot)$: | Type of data integrity attacks. |
| $\alpha$: | The attack parameter, which is a random fixed number from 0.1 to 0.8. |
| $\beta_t$: | The flag to represent whether $h_2$ attack is launched or not in the $t^{th}$ time period. If launched, the value is 0, otherwise it is 1. |
| $\lambda_t$: | The scaling attack parameter in the $t^{th}$ time period, which is a random number from 0.1 to 0.8. |
| $g(D,A)$: | The information gain of feature $A$ to dataset $D$. |
| $H(D)$: | The information entropy of dataset $D$. |
| $H(D\|A)$: | The empirical conditional entropy of dataset $D$ when feature $A$ is given. |
| $c_j$: | The power consumption vector on the $j^{th}$ day. |
| $c_{j-h}$: | Power consumption in the $h^{th}$ time period on the $j^{th}$ day. |
| $sum$: | Total collection days of user data. |
| $K/k$: | The number of consumption patterns/the order number of consumption patterns. |
| $n_{mv}$: | The number of missing values in a consumption vector. |
| $c_{center-k}$: | The center vector of the $k^{th}$ consumption pattern. |
| $c^h{}_{center-k}$: | Power consumption in the $h^{th}$ time period in the $c_{center-k}$. |
| $l_{jk}$: | The distance between $c_j$ and $c_{center-k}$. |
| $l_{c_{j1}c_{j2}}$: | The distance between the power consumption vectors $c_{j1}$ and $c_{j2}$. |
| $C^k$: | The set of power consumption data corresponding to the $k^{th}$ consumption pattern. |
| $I_k$: | The consumption interval of the $k^{th}$ consumption pattern. |
| $min_k$: | The minimum power consumption per unit time period in the $k^{th}$ consumption pattern. |
| $max_k$: | The maximum power consumption per unit time period in the $k^{th}$ consumption pattern. |
| $\varphi$: | The set of time period features of consumption data. |
| $T_h$: | The $h^{th}$ time period of consumption data. |
| $T_{in}/T_{out}$: | The set of time periods in which values of most consumption data in this time period are within or outside the normal interval after the attack is launched. |
| $T_{h-in}/T_{h-out}$: | The consumption data set in which consumption values in the $h^{th}$ time period are within or outside the normal interval. |

data on the basis of the scaling attack model, $h_3$ ($x_t$), described in **Section 2.2**, and then discretize the data. Finally, use the discretized data as the input of the classifier for detection.

In order to achieve better performance, we leverage Kmeans for clustering and the decision tree for classification, and propose a detection model of scaling attacks considering consumption pattern diversity in AMI (*SA2CPD*). **Algorithm 1** shows the specific implementation process, which consists of four steps: i) data preprocessing; ii) distinguishing different consumption patterns and extracting consumption intervals; iii) binarization and iv) classification.

## 3.2 Our method

### 3.2.1 Data preprocessing

This step corresponds to lines 1–12 in **Algorithm 1**. Power consumption collected by the smart meter deployed on the user side can be represented as a matrix $c = [c_1, c_2 \ldots c_j \ldots c_{sum}]^T$, where $sum$ indicates total collection days and $c_j$ represents the power consumption vector on the $j^{th}$ ($j \in [1, sum]$) day. $c_j = [c_{j-1}, c_{j-2} c_{j-h} c_{j-24}]$, in which $c_{j-h}$ represents power consumption in the $h^{th}$ time period on the $j^{th}$ day. Assume the number of missing values is $n_{mv}$ in a consumption vector. When the number of missing values is no more than 6 ($n_{mv} \leq 6$), if missing values are not consecutive, we take the mean of power consumption of the previous time period and the next time period to fill each missing value (Jakaria et al., 2019), and the average value of the consumption vector instead to fill them if there are consecutive missing values. When the number of missing values exceeds a quarter ($n_{mv} > 6$), the consumption vector is denoted as unavailable (Tehrani et al., 2020).

### 3.2.2 Distinguish consumption patterns

This step corresponds to lines 13–14 in **Algorithm 1**. Affected by personal habits, holidays and other factors, each user has different power consumption patterns, and the power consumption patterns of different users are also different from each other. Therefore, it is necessary to cluster the power consumption data before classification to reduce the false negative rate. There are various methods that can be used to distinguish power consumption patterns, here we use the Kmeans method, which is the most commonly used in clustering (Jokar et al., 2016; Tehrani et al., 2020).

The implementation of Kmeans clustering is an iterative process including three steps. First step, $K$ vectors are randomly selected from $c$ as centers of initial consumption pattern sets $c_{center} = [c_{center-1}, c_{center-2}\ c_{center-k}\ c_{center-K}]$. Second step, for each $c_j$, calculate the distance between it and each $c_{center-k}$ as

$$l_{jk} = \|c_j - c_{center-k}\|_2^2 = \sqrt{\left(c_{j-1} - c^1{}_{center-k}\right)^2 + \cdots + \left(c_{j-h} - c^h{}_{center-k}\right)^2}, \quad (3)$$

where $c_{center-k}$ represents the center of the $k^{th}$ consumption pattern, $c^h{}_{center-k}$ represents power consumption in the $h^{th}$ time period in $c_{center-k}$, and $l_{jk}$ represents the distance between $c_j$ and

**Input:** User's consumption data $c=\begin{bmatrix} c_{1\_1} & c_{1\_h} & \cdots & c_{1\_24} \\ c_{2\_1} & c_{2\_h} & \cdots & c_{2\_24} \\ \cdots & \cdots & \cdots & \cdots \\ c_{j\_1} & c_{j\_h} & \cdots & c_{j\_24} \end{bmatrix}$, the time periods vector $\varphi=[T_1 T_2 \ldots T_h \ldots T_{24}]$, the set of time periods $T_{in}$ and $T_{out}$
**Output:** The label for new data $c_j$

1: **for** $j=1$ to $sum$ **do**
2: **if** the number of missing values in the $c_j$ is $n_{mv} \leq 6$ **then**
3: **if** there are consecutive missing values in the $c_j$ **then**
4: Each missing value $c_{j-h}$ is expressed as the average of the $c_j$
5: **else**
6: For each missing value $c_{j-h}=\frac{c_{j-(h-1)\%24}+c_{j-(h+1)\%24}}{2}$
7: **end if**
8: **end if**
9: **if** $n_{mv}>6$ **then**
10: The $c_j$ is denoted as unavailable
11: **end if**
12: **end for**
13: Cluster the dataset into $K$ sets as $C=[C^1, C^2 \ldots C^k \ldots C^K]$
14: Extract consumption intervals $I_k=[min_k, max_k]$ from each $C^k$
15: Generate false data for each group of normal data
16: **for** $j=1$ to $sum$ **do**
17: **for** $h=1$ to 24 **do**
18: **if** $c_{j-h} \in I_k$ **then**
19: $c_{j-h}=0$
20: $T_{j-h} \in T_{h-in}$
21: $|T_{h-in}|=|T_{h-in}|+1$
22: **else**
23: $c_{j-h}=1$
24: $T_{j-h} \in T_{h-out}$
25: $|T_{h-out}|=|T_{h-out}|+1$
26: **end if**
27: **end for**
28: **end for**
29: **if** $|T_{h-in}| \gg |T_{h-out}|$ **then**
30: $T_h \in T_{in}$
31: Calculate the empirical conditional entropy of this time period to get $H(D|T_h) \approx 1$
32: **else**
33: $T_h \in T_{out}$
34: Calculate the empirical conditional entropy of this time period to get $H(D|T_h) \approx 0$
35: **end if**
36: **while** new user's consumption data $c_j$ is collected **do**
37: The decision tree preferentially selects the time period $T_h \in T_{out}$ as the judgment condition, and then judges
38: **if** any $T_h \in T_{out}$, the value is 1 **then**
39: The label of $c_j$ is normal data
40: **else**
41: The label of $c_j$ is false data
42: **end if**
43: **end while**
44: Return the label of $c_j$

Algorithm 1. The detection model of scaling attacks considering consumption pattern diversity in AMI(*SA2CPD*).

$c_{center-k}$. Third step, $c_j$ is classified into the $C^k$ corresponding to the smallest $l_{jk}$, where $C^k$ represents the $k^{th}$ consumption pattern set. Then recalculate the new center of $C^k$ as

$$c_{center-k}=\frac{\sum_{c_j \in C^k} c_j}{\left|C^k\right|}, \tag{4}$$

where $\left|C^k\right|$ represents the number of power consumption vectors in the $k^{th}$ consumption pattern set. The iteration stops until centers do not change, meaning that the clustering is finished, and we can obtain $K$ consumption patterns set $C=[C^1, C^2 \ldots C^k \ldots C^K]$, where

$$C^k=\begin{bmatrix} c^k{}_{1\_1} & c^k{}_{1\_h} & \cdots & c^k{}_{1\_24} \\ c^k{}_{2\_1} & c^k{}_{2\_h} & \cdots & c^k{}_{2\_24} \\ \cdots & \cdots & \cdots & \cdots \\ c^k{}_{\mathrm{d}\_1} & c^k{}_{d\_h} & \cdots & c^k{}_{d\_24} \end{bmatrix}, \tag{5}$$

$c^k{}_{d\_h}$ represents power consumption of the $h^{th}$ time period on the $d^{th}$ ($d \leq sum$) day in the $k^{th}$ power consumption pattern.

Notice that the random selection of initial centers may result in a local optimal solution rather than a global optimal solution. Therefore, we take advantage of the characteristic that there is no intersection between different consumption patterns to set filter conditions to exclude local optimal solutions, which can be formalized as

$$\begin{aligned} &\text{if} \\ &\quad max_s < max_t \\ &\text{then} \\ &\quad max_s < min_t, (s,t \in [1,K], s \neq t) \end{aligned}. \tag{6}$$

Here, $min_i$ ($i=s,t$) and $max_i$ ($i=s,t$) represent the minimum and maximum values of power consumption per unit time period in the $i^{th}$ consumption pattern.

### 3.2.3 Binarization

This step corresponds to lines 15–28 in **Algorithm 1**. When detecting data integrity attacks, it is necessary to analyze the difference between normal data and false data. Here, we leverage the binarization method to transform fine granularities into coarse granularities to make the difference of features larger to improve detection efficiency.

We can extract $K$ corresponding intervals of $K$ consumption patterns as $I=[I_1, I_2 \ldots I_k \ldots I_K]$, where

$$I_k=[min_k, max_k]. \tag{7}$$

In **Eq. 7**, $I_k$ represents the interval of the $k^{th}$ consumption pattern set.

After extracting consumption intervals, false data is generated through multiplying each $c^k{}_{d\_h}$ by a $\lambda_t \in [0.1, 0.8]$. Then, use the interval $I_k$ in $I$ for binarization after mixing normal data and generated false data as

$$\begin{cases} c^k{}_{d\_h}=0, c^k{}_{d\_h} \in I \\ c^k{}_{d\_h}=1, c^k{}_{d\_h} \notin I \end{cases}. \tag{8}$$

For each $c^k{}_{d\_h}$, if it is within $I$, it is binarized to 0. Otherwise it is binarized to 1.

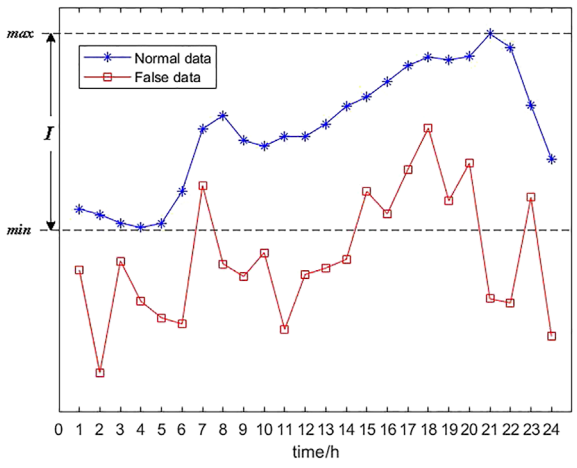



FIGURE 2
False data vs. normal data.

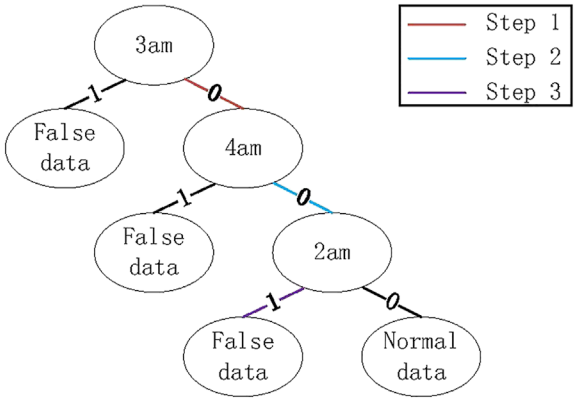



FIGURE 3
Decision tree generated by consumption data.

## 3.2.4 Classification

This step corresponds to lines 29–43 in **Algorithm 1**. Let $\varphi = [T_1, T_2 \ldots T_h \ldots T_{24}]$ represents 24 time periods of a power consumption vector. For normal data, values of all 24 time periods are within $I$, so all values are binarized to 0. For false data, although there are some values within $I$, most of values are outside $I$, as shown in **Figure 2**.

**Figure 2** shows the comparison between normal data and false data. It can be seen that only values in eight time periods in false data including 7, 15, 16, 17, 18, 19, 20 and 23 are greater than the minimum value of normal data. Let $T_{in}$ and $T_{out}$ represent the set of time periods in which most values in this period in power consumption vectors are within or without $I$ after the attack is launched. For the decision tree, the empirical conditional entropy of time period $T_h$ to dataset $D$ is denoted as

$$\begin{aligned} H(D|T_h) &= \sum_{i=0}^{1} \frac{|D_i|}{|D|} H(D_i) \\ &= -\sum_{i=0}^{1} \frac{|D_i|}{|D|} \sum_{l=0}^{1} \frac{|D_{il}|}{|D_i|} \log_2 \frac{|D_{il}|}{|D_i|} \end{aligned}, \tag{9}$$

where $i$ indicates the value of power consumption after binarization in the time period $T$ which is 0 or 1, $l$ is a flag representing whether the power consumption vector is normal (denoted as 0) or false (denoted as 1), $D_i$ is the set of power consumption vector when its value in the time period $T$ is $i$, $D_{il}$ is the set of power consumption vector when its value in the time period $T$ is $i$ and the flag is $l$, $|.|$ represents the quantity of power consumption vectors in a set. The object of constructing a decision tree is to find the time period with the greatest information gain, which can be formalized as

$$\max g(D, T_h) = H(D) - H(D|T_h). \tag{10}$$

It is also equivalent to

$$\min \{H(D|T_h)\}. \tag{11}$$

For the time period in $T_{in}$, as the number of power consumption data (vectors) increases, we have

$$\begin{aligned} |D_{i=1}| &\to 0 \\ |D_{i=0}| &\to |D| \end{aligned}. \tag{12}$$

If the dataset is balanced, we can derive that

$$|D_{i=0,\ \ l=0}| \approx |D_{i=0,\ \ l=1}| \approx \frac{1}{2} |D_{i=0}|. \tag{13}$$

Hence, on the basis of **Eqs 9**, **12**, **13**, when $T_h \in T_{in}$, we can obtain that

$$H(D|T_h) \approx 1. \tag{14}$$

For the time period in $T_{out}$, as the number of power consumption data (vectors) increases, we have

$$|D_{i=0}| \approx |D_{i=1}| \approx \frac{1}{2} |D|. \tag{15}$$

If the dataset is balanced, we can derive that

$$\begin{aligned} |D_{i=0,\ \ l=0}| &\approx |D_{i=1,\ \ l=1}| \approx |D_{i=0}| \approx |D_{i=1}| \\ |D_{i=0,\ \ l=1}| &\approx |D_{i=1,\ \ l=0}| \approx 0 \end{aligned}. \tag{16}$$

Hence, on the basis of **Eqs 9**, **15**, **16**, when $T_h \in T_{out}$, we can obtain that

$$H(D|T_h) \approx 0. \tag{17}$$

Therefore, from **Eqs 11**, **14**, **17**, we know that a decision tree should be constructed based on power consumption during time periods in $T_{out}$ and those of time periods in $T_{in}$ will not be adopted, which can maximize information gain and avoid the decision tree being too large. For example, after the scaling

attack is launched, power consumption in time periods 1–6 are outside $I$ with the greatest probability, because these time periods usually belong to valley time periods for many users. Therefore, time periods 1–6 belong to $T_{out}$, based on which the decision tree shown in **Figure 3** can be constructed. When a new power consumption vector is collected, the judgement will be made from the root node to a leaf node. For example, if binarization values in time periods 1–6 are [0,1,0,0,1,0], it will be detected as false data after judgements in Step 1, Step 2 and Step 3.

# 4 Detection performance analysis

In this section, we first introduce metrics of detection performance, and then show comparison analysis with other models.

## 4.1 Metrics

We use the *FPR*, the *FNR* and the *F1 score* as metrics to compare with other algorithms (Amara korba and El Islem karabadji, 2019; Jakaria et al., 2019; Rouzbahani et al., 2020). The higher the *F1 score* is, the lower the *FPR* and *FNR* are, the better the performance is. Relevant notations are given below.

(1) *TP/TN/FP/FN*: False data is detected as false data/normal data is detected as normal data/normal data is detected as false data/false data is detected as normal data.

(2) *Recall (Rec)*: The ratio of the number of false data being detected as false data versus the total number of false data, meaning that

$$Rec = \frac{TP}{TP + FN}. \tag{18}$$

(3) *FNR*: The ratio of the number of false data being detected as normal data versus the total number of false data, meaning that

$$FNR = \frac{FN}{TP + FN} = 1 - Rec. \tag{19}$$

(4) *Precision (Pre)*: The ratio of the number of false data being detected as false data versus the total number of data being detected as false data, meaning that

$$Pre = \frac{TP}{TP + FP}. \tag{20}$$

(5) *FPR*: The ratio of the number of normal data being detected as false data versus the total number of normal data, meaning that

$$FPR = \frac{FP}{FP + TN}. \tag{21}$$

(6) *F1 Score*: The harmonic average of *Precision* and *Recall*, which is

$$F_1 = 2 \cdot \frac{Pre \cdot Rec}{Pre + Rec}. \tag{22}$$

## 4.2 Comparison with other models

### 4.2.1 *SA2CPD* VS. models without considering consumption pattern diversity

Different from our *SA2CPD* model in **Section 3**, in which clustering is performed first to obtain multiple power consumption patterns, and then consumption intervals can be extracted as the basis for binarization. When consumption pattern diversity is not considered, all power consumption data of users are regarded as belonging to a single pattern, in which case only a known value can be selected as the threshold. Here we discuss two models with the minimum value and the mean value as thresholds and we call them as B-MIN Model and B-MEAN Model.

#### 4.2.1.1 Binarization based on the minimum value (B-MIN model)

Compared with the *B-MIN* model, performance of our *SA2CPD* model is better in terms of the *FNR* and the *F1 score*, as shown in Theorem 1.

Theorem 1. *The F1 score of our SA2CPD model is greater than that of the B-MIN model. The FNR of our SA2CPD model is smaller than that of the B-MIN model. The FPR of our SA2CPD model is larger than that of the B-MIN model.*

**Proof:** When using the *B-MIN* model, only false data in the pattern the minimum value belonged to among all of the power consumption patterns can be effectively detected, while hourly collected values of the false data in the other patterns are likely to remain greater than the minimum value so that these false data will be detected as normal data, which will result in a lower *Recall* and a higher *FNR*. Similarly, since only normal data in the pattern the minimum value belonged to may be detected as false data, we can obtain that $FP_{\min} < FP_{our}$ so that the *FPR* of the *B-MIN* is smaller than ours. Furthermore, there are few numbers of $FP_{\min}$, so the *Precision* of *B-MIN* is higher than ours. However, since the number of $FP_{our}$ is also very small, the *Precision* of *SA2CPD* is about equal to that of *B-MIN*. Take both *Recall* and *Precision* into consideration, the *F1 score* of the *B-MIN* will be lower than our *SA2CPD* model. The above analysis process can be formalized as

$$\begin{aligned}
&FN_{min} > FN_{our} \Rightarrow FNR_{min} > FNR_{our}\\
&FP_{min} < FP_{our} \Rightarrow FPR_{min} < FPR_{our}\\
&\begin{cases} Recall_{min} < Recall_{our} \\ Precision_{min} > Precision_{our} \\ \vartriangle Recall \gg \vartriangle Precision \end{cases}\\
&\Rightarrow F_{1_{min}} < F_{1_{our}}
\end{aligned} \tag{23}$$

#### 4.2.1.2 Binarization based on the minimum value (B-MIN model)

Compared with the *B-MEAN* model, performance of our *SA2CPD* model is better in terms of the *FPR*, the *FNR* and the *F1 score*, as shown in Theorem 2.

Theorem 2. The *F1 score of our SA2CPD model is greater than that of the B-MIN model. Either one or both of the FNR and the FPR of our SA2CPD model are smaller than those of the B-MEAN model.*

**Proof:** When using the *B-MEAN* model, values greater than the mean value are binarized to 0, otherwise 1. When a scaling attack is launched, false data greater than the mean value will be falsely detected as normal data and normal data less than the mean value will be falsely regarded as false data. Therefore, $FN_{mean}$ and $FP_{mean}$ are influenced by the mean value. Specifically, there are three cases. When the mean value is small, there will be more false data being detected as normal data, resulting in a greater $FN_{mean}$. When the mean value is large, there will be more normal data being detected as false data, resulting in a greater $FP_{mean}$. When the mean value is the median, both $FN_{mean}$ and $FP_{mean}$ will be larger. Thus, we can conclude that either one or both of the *FPR* and *FNR* are larger, which can be formalized as

$$\begin{cases} FN_{mean} \gg FN_{\text{our}} \\ FP_{mean} < FP_{\text{our}} \end{cases} \Rightarrow \begin{cases} FNR_{mean} \gg FNR_{\text{our}} \\ FPR_{mean} < FPR_{\text{our}} \end{cases}$$
$$or \begin{cases} FN_{mean} > FN_{\text{our}} \\ FP_{mean} > FP_{\text{our}} \end{cases} \Rightarrow \begin{cases} FNR_{mean} > FNR_{\text{our}} \\ FPR_{mean} > FPR_{\text{our}} \end{cases}. \quad (24)$$
$$or \begin{cases} FN_{mean} < FN_{\text{our}} \\ FP_{mean} \gg FP_{\text{our}} \end{cases} \Rightarrow \begin{cases} FNR_{mean} < FNR_{\text{our}} \\ FPR_{mean} \gg FPR_{\text{our}} \end{cases}$$

The derivation process of corresponding *recall*, *precision* and *F1 score* is

$$\begin{cases} Recall_{mean} \ll Recall_{our} \\ Precision_{mean} > Precision_{our} \\ \vartriangle Recall > \vartriangle Precision \end{cases}$$
$$or \begin{cases} Recall_{mean} < Recall_{our} \\ Precision_{mean} < Precision_{our} \end{cases} \Rightarrow F_{1_{mean}} < F_{1_{our}}. \quad (25)$$
$$or \begin{cases} Recall_{mean} > Recall_{our} \\ Precision_{mean} \ll Precision_{our} \\ \vartriangle Recall < \vartriangle Precision \end{cases}$$

### 4.2.2 Decision tree VS. Naive Bayes

The reason why we choose the decision tree as the classifier is that the binarization can help the decision tree discretize continuous values and can make a great difference between normal data and false data. Similarly, Naive Bayes can also use the binarization method to improve the detection efficiency. However, it is only suitable to the situation where the distribution of power consumption data of each pattern is concentrated. When the distribution of power consumption data of each pattern is scattered, compared with the Naive Bayes model, performance of our *SA2CPD* model is better in terms of the *FNR* and the *F1 score*, as shown in Theorem 3.

Theorem 3. T*he F1 score of our SA2CPD model is greater than that of the Naive Bayes model. The FNR of our SA2CPD model is smaller than that of the Naive Bayes model. The FPR of our SA2CPD model is higher than that of the Naive Bayes model.*

**Proof:** Different from *SA2CPD* in which only power consumption during time periods in $T_{out}$ is used for detection, power consumption in all time periods need be considered in the Naive Bayes model. When the distribution of power consumption data of each pattern is concentrated, most of the values of false data will be outside normal intervals and binarized to 1 in the training set, so that newly collected false data can be correctly detected. Nevertheless, when the distribution of power consumption data of each pattern is scattered, some values of false data will be within normal intervals and then binarized to 0 in the training set, which will have an impact on the judgment of newly collected data. Under these circumstances, the probability of correctly detecting false data will be reduced, resulting in a decrease of $TP_{Bayes}$ and an increase of $FN_{Bayes}$. The more dispersed the distribution is, the greater the impact is. As a result, the *FNR* will be larger and the *Recall* will be smaller. Furthermore, since all normal data can be binarized to 0, the number of *FP* will be small so that the *FPR* will be slightly lower than ours and the *Precision* will be approximately equal to ours. Take both *Recall* and *Precision* into consideration, the *F1 score* of the Naive Bayes model is smaller than that of our *SA2CPD* model, which can be formalized as

$$\begin{cases} FN_{Bayes} > FN_{our} \\ FP_{Bayes} < FP_{our} \end{cases} \Rightarrow \begin{cases} FNR_{Bayes} > FNR_{our} \\ FPR_{Bayes} < FPR_{our} \end{cases}$$
$$\Rightarrow \begin{cases} Recall_{\text{Bayes}} < Recall_{our} \\ Precision_{\text{Bayes}} \approx Precision_{our} \end{cases}. \quad (26)$$
$$\Rightarrow F_{1_{\text{Bayes}}} < F_{1_{our}}$$

### 4.2.3 Decision tree VS. KNN

After the adversary launches the scaling attack, values in the power consumption vector will be reduced, resulting in a distance between false data and normal data. Hence, KNN can be used to detect scaling attacks whose classification is according to the distance. However, compared with the KNN model, performance of our *SA2CPD* model is better in terms of the *FNR* and the *F1 score*, as shown in Theorem 4.

Theorem 4. T*he F1 score of our SA2CPD model is greater than that of the KNN model. The FNR of our SA2CPD model is smaller than that of the KNN model. The FPR of our SA2CPD model is higher than that of the KNN model.*

**Proof:** In some cases, scaling attacks may cause the distance between false data and false data at the same feature, i.e., power

consumption in the same time period, to be greater than that between false data and normal data. Assume the values of two false data $c'_1$ and $c'_2$ in time period $h$ are obtained through multiplying two similar normal data by $\lambda_1$ and $\lambda_2$ respectively and if $\lambda_1$ and $\lambda_2$ satisfy

$$|\lambda_1 - \lambda_2| > |1 - \lambda_i|, \quad i = 1 \ or \ 2. \tag{27}$$

When $i$ is 1, the distance between $c'_1$ and the original power consumption vector $c$ is

$$\begin{aligned} l_{c'_1 c} &= \sqrt{(c'_{1-h} - c_h)^2 + \cdots} \\ &= \sqrt{(1-\lambda_1)^2 {c_{1-h}}^2 + \cdots} \end{aligned}, \tag{28}$$

the distance between $c'_1$ and $c'_2$ is

$$\begin{aligned} l_{c'_1 c'_{2-h}} &= \sqrt{(c'_{1-h} - c'_{2-h})^2 + \cdots} \\ &= \sqrt{(\lambda_2 - \lambda_1)^2 {c_{1-h}}^2 + \cdots} \end{aligned}. \tag{29}$$

When this situation also exists in many other time periods, we can obtain that

$$l_{c'_1 c} < l_{c'_1 c'_2}. \tag{30}$$

Thus, $c'_1$ will be detected as normal data and *FN* will be greater, resulting in a larger *FNR* and a lower *Recall*. When $i$ is 2, The analysis process is the same. In terms of *Precision*, the majority consumption data closest to normal data is normal data although false data may exist, so there is almost no *FP* and the *Precision* will be almost unaffected, and the *FPR* is lower than ours. Take both *Recall* and *Precision* into consideration, the *F1 score* of the KNN model is smaller than that of our *SA2CPD* model, which can be formalized as

$$\begin{aligned} &\begin{cases} FN_{KNN} \gg FN_{our} \\ FP_{KNN} < FP_{our} \end{cases} \\ &\Rightarrow \begin{cases} FNR_{KNN} \gg FNR_{our} \\ FPR_{KNN} < FPR_{our} \end{cases} \\ &\Rightarrow \begin{cases} \mathrm{Rec}_{KNN} \ll Rec_{our} \\ Pre_{KNN} > Pre_{our} \end{cases} . \\ &\because \because \triangle FN \gg \triangle FP \\ &\Rightarrow \triangle Recall \gg \triangle Precision \\ &\Rightarrow F_{1_{KNN}} < F_{1_{our}} \end{aligned} \tag{31}$$

# 5 Performance evaluation

In this section, we first introduce the simulation setup. We then show experimental results to validate the effectiveness and efficiency of the *SA2CPD*.

## 5.1 Evaluation setup

In our evaluation, the GEFCom2012 dataset (Hong, 2014) from the global energy forecasting competition was used to carry out the performance validation of our *SA2CPD* model. The dataset includes historical records of hourly collected power consumption in 20 zones from 1 January 2004 to 30 June 2008. Each record includes 28 columns. The first column is zone_id, the second to fourth columns are year, month and day, and the fifth to 28*th* columns are 24 hourly collected power consumption values. There are no missing values in the dataset. By extracting three zones with large power consumption differences, we simulate a user with three power consumption patterns and there are 1,586 power consumption vectors in each consumption pattern. We set the ratio of training set to test set as 7: 3 and 854 power consumption vectors are taken from each consumption pattern to generate false data for training. Hence, the size of the training set is 5,124 and the size of the testing set is 2,196.

Based on the above simulation settings, we conduct two experiments, each of which includes 500 random evaluation cases. In the first experiment, we evaluate performance of our *SA2CPD* model under different attack proportions. In the second experiment, we conduct two groups of comparative experiments. Firstly, our *SA2CPD* model is compared with the decision tree without considering consumption pattern diversity. Then we compare our *SA2CPD* model with the KNN model and the Naive Bayes model mentioned in **Section 4**.

## 5.2 Effectiveness of our *SA2CPD* model

**Figure 4** illustrates the *FPR*, the *FNR* and the *F1* score of our *SA2CPD* model under different attack proportions. As shown in this figure, when the attack proportion in the testing set increases

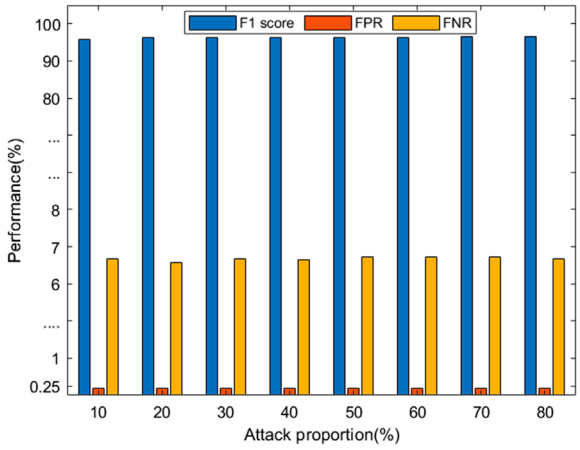



FIGURE 4
Performance of *SA2CPD* under different attack proportions.

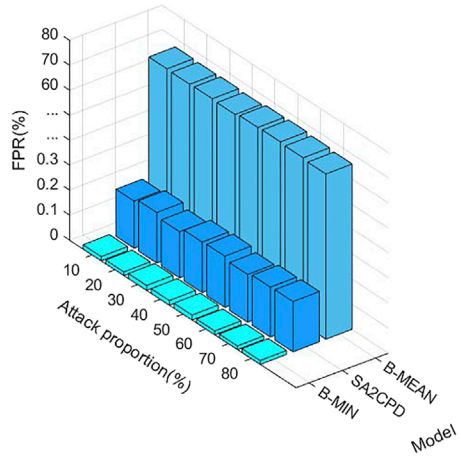



FIGURE 5
The *FPR* of our *SA2CPD* model, the B-MIN model and the B-MEAN model under different attack proportions.

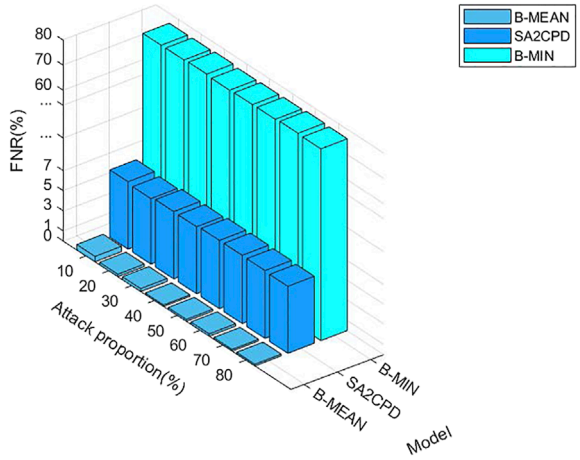



FIGURE 6
The *FNR* of our *SA2CPD* model, the B-MIN model and the B-MEAN model under different attack proportions.

from 10% to 80%, the *F1* score are [ 95.7%, 96.23%, 96.32%, 96.4%, 96.38%, 96.41%, 96.46%, 96.52% ], the *FPR* are [ 0.18%, 0.20%, 0.18%, 0.20%, 0.20%, 0.19%, 0.19%, 0.20% ], and the *FNR* are [ 6.69%, 6.51%, 6.69%, 6.63%, 6.78%, 6.79%, 6.75%, 6.67% ]. Hence, regardless of the attack proportion, our *SA2CPD* model has a high *F1* score, a low *FPR* and a low *FNR*, validating the effectiveness of our *SA2CPD* model.

## 5.3 Comparison with the B-MIN model and the B-MEAN model

**Figure 5** depicts the *FPR* of our *SA2CPD* model, the B-MIN model and the B-MEAN model. It can be seen from the figure that the *FPR* of the *B-MIN* is the lowest, followed by our method. Both of them are lower than 0.5% and the difference between them is very small. However, the *FPR* of the *B-MEAN* is over 60%. **Figure 6** depicts the *FNR* of our *SA2CPD* model, the B-MIN model and the B-MEAN model. It can be seen that the *B-MEAN* has the lowest *FNR*, followed by our method. Similarly, the difference between them is small. However, the *FNR* of the *B-MIN* is over 80%. The above experimental results are consistent with our analysis in 4.2.1.

**Figure 7** shows *F1 score* of our *SA2CPD* model, the B-MIN model and the B-MEAN model. As can be seen from figure, the *F1 score* of the *B-MIN* model is less than 50%, and it is almost unchanged with the increase of attack proportions. The *F1 score* of the *B-MEAN* model under different attack ratios are [24.8%, 42.83%, 56.10%, 66.72%, 74.82%, 81.58%, 87.41%, 92.25%], which increases significantly with the increase of attack proportions. The *F1 score* of our method are all above 95% and are

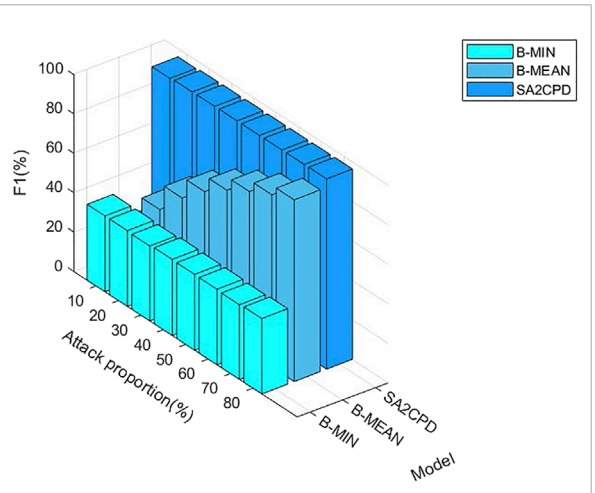



FIGURE 7
The *F1 score* of our *SA2CPD* model, the B-MIN model and the B-MEAN model under different attack proportions.

always greater than those of the B-MIN model and the B-MEAN model. The above experimental results verify the theoretical analysis in 4.2.1.

## 5.4 Comparison with the Naive Bayes model and the KNN model

**Figure 8** displays the *FPR* of our *SA2CPD* model, the Naive Bayes model and the KNN model. From the figure, performance of the KNN model is the best and the highest *FPR* is only 0.06% when the attack proportion is 60%. The performance of the Naive Bayes model takes the second place, and its *FPR* does not exceed

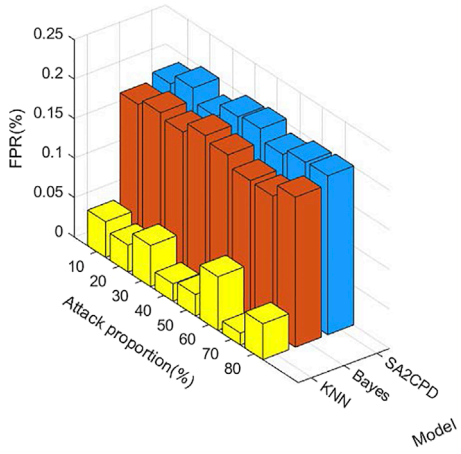



FIGURE 8
The *FPR* of our *SA2CPD* model, the Naive Bayes model and the KNN model under different attack proportions.

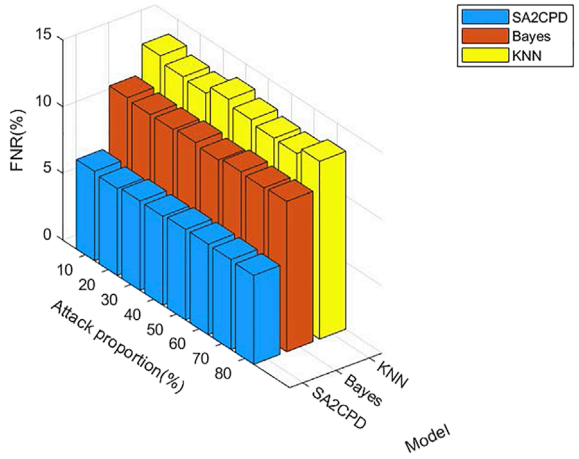



FIGURE 9
The *FNR* of our *SA2CPD* model, the Naive Bayes model and the KNN model under different attack proportions.

0.2%. Our *SA2CPD* model shows the worst performance, but all *FPR* values are maintained at about 0.2%. Figure 9 displays the *FNR* of our *SA2CPD* model, the Naive Bayes model and the KNN model. It can be seen that the decision tree has the best performance and all *FNR* values are slightly higher than 5%. Nonetheless, the *FNR* of the Naive Bayes model is more than 10%, and the *FNR* of the KNN model is close to 15%. The above experimental results verify the theoretical analysis in 4.2.2 and 4.2.3.

Figure 10 displays the *F1 score* of our *SA2CPD* model, the Naive Bayes model and the KNN model. Although the performance of our *SA2CPD* model on the *FNR* is much better than the other two, the *FPR* is slightly higher, it can not be

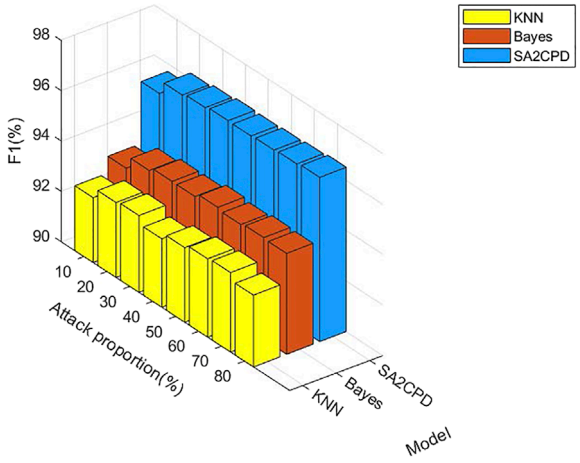



FIGURE 10
The *F1 score* of our *SA2CPD* model, the Naive Bayes model and the KNN model under different attack proportions.

concluded that the performance of our *SA2CPD* model is the best. Hence, we further show the performance of these three models in terms of *F1 score* in Figure 10. As we can see, no matter what the attack proportion is, the *F1 score* of our *SA2CPD* model are always greater than those of the Naive Bayes model and the KNN model. Specifically, the *F1 score* of the KNN model is slightly higher than 92%, and that of the Naive Bayes model is slightly higher than 94%, while that of our *SA2CPD* model is always higher than 96%. The above experimental results are consistent with our theoretical analysis in 4.2.2 and 4.2.3.

# 6 Discussion

We now discuss the following problems and extensions related to this paper: detection of scaling attacks by injecting enlarged values, and detection of the *h*4 attack mentioned in Section 2.

## 6.1 Dedection of scaling attacks by injecting enlarged values

In this paper we investigate scaling attacks tampering with reduced reported values in smart meters, which can be represented as $h_3(x_t) = \gamma_t x_t, \gamma_t = random(0.1, 0.8)$. In fact, injecting enlarged data into smart meters also belongs to the category of scaling attacks, which can be formalized as $h_3(x_t) = \gamma_t x_t, \gamma_t = random(1, +\infty)$. Classical detection methods compare the total power supply and the total power consumption of all users to detect scaling attacks (Jokar et al., 2016). If these two values are close to

each other, it is considered that no attack has occurred (Bhattacharjee et al., 2021a,b). If the total power consumption is less than the power supply, it is considered that an attack has occurred. Hence, if an adversary launches scaling attacks by injecting both reduced values and enlarged values and ensure that the total power supply and the total power consumption are close to each other, he can easily escape from these detection methods. In contrast, *SA2CPD* can still detect this adversary effectively. This is because, no matter whether the data is reduced or enlarged, as long as the false data falls outside the extracted consumption intervals, it will be binarized to 1, which can be detected by the decision tree.

## 6.2 Detection of the $h4$ attack

In the **Section 2**, we mentioned that the $h4$ attack can be represented by the $h3$ attack. It is because $h_4$ represents manipulating the hourly reported value of smart meter as the mean value of a day multiplying by a different random number. Therefore, $h_4 = \lambda_t(h4) \cdot mean(x) = \left(\lambda_t(h4) \cdot \frac{mean(x)}{x_t}\right) \cdot x_t$ so that the $h4$ attack can be transformed to the $h3$ attack. Hence, we can first transform $h4$ attacks into $h3$ attacks and then use *SA2CPD* for detection. It is worth noting that *SA2CPD* can also be directly used to detect $h4$ attacks. Compared to attacking $x_t$ directly, the fake data value after attacking the mean may be larger or smaller than $x_t$. *SA2CPD* can effectively detect false data in either case, the only difference is that the features in $T_{out}$ are different. When $\left(\lambda_t(h4) \cdot \frac{mean(x)}{x_t}\right) \leq 1$, including: i) $\frac{mean(x)}{x_t} \leq 1 \Rightarrow \left(\lambda_t(h4) \cdot \frac{mean(x)}{x_t}\right) \leq 1$; ii) $\frac{mean(x)}{x_t} > 1$ and $\left(\lambda_t(h4) \cdot \frac{mean(x)}{x_t}\right) \leq 1$, time period features in $T_{out}$ increase or remain unchanged. When $\frac{mean(x)}{x_t} > 1$ and $\left(\lambda_t(h4) \cdot \frac{mean(x)}{x_t}\right) > 1$, time period features in $T_{out}$ decrease.

# 7 Related work

As we described in **Section 1**, existing research on data integrity attacks detection in AMI falls into three main categories (Jiang et al., 2014; Jokar et al., 2016; Yao et al., 2019). The first type is state-based. For example, Huang et al. (Huang et al., 2013) used state estimation and analysis of variance (ANOVA) based on customer metering data aggregated at distribution transformers to detect contaminated meters and estimate the actual usage. Salinas et al. (Salinas et al., 2014) studied data integrity attacks in microgrids. They took values of stolen electricity as the measurement bias and used the least square method to make the optimal estimation, then the honest meter will show a zero bias and the compromised meter will show a non-zero bias. Leite et al. (Leite and Mantovani, 2018) used the data of meters to detect the power loss based on the multivariate procedure of monitoring and control. Through the combination with GIS program, the geographical location of fraud can also be found. Using power information and sensors placement, Lo et al. (Lo and Ansari, 2013) developed a hybrid detection framework to detect data integrity attacks by detecting abnormal activities in the power grid. McLaughlin et al. (McLaughlin et al., 2013) proposed a system that combines multiple technologies to detect data integrity attacks. The system collects relevant evidence of attacks from three different information sources to minimize the number of false positives. Aziz et al. (Aziz et al., 2020) rely on the results of state estimation in centralised aggregators, located between smart meters and the control center, to aid in false data detection. Bhattacharjee et al. (Bhattacharjee et al., 2021b) embedded the appropriate unbiased mean, the median absolute deviation, etc. to produce trust scores for smart meters to classify compromised smart meters from normal ones.

The second type is based on game theory, and the goal of game theory-based methods is to find a balance between the utility and adversaries (Cardenas et al., 2012; Yang et al., 2016; Wei et al., 2018, 2017; Paul et al., 2020). For example, Yang et al. (Yang et al., 2016) proposed a game theory model to deal with the situation that multiple adversaries jointly launch attacks. They introduced a penalty factor to represent the punishment for adversaries when they were detected. When an adversary decides to participate in a joint attack and succeed, the gain will be distributed to each adversary. When the attack is detected and fails, the adversaries participating in will be punished. The more adversaries involved, the greater the probability of failure. Based on the operational cost model of the utility, Cardenas et al. (Cardenas et al., 2012) expressed the problem of attack detection as a game between adversaries and the utility. Adversaries aim to achieve the best benefit and not be found, while the utility want to detect attacks as much as possible at a lower cost. Paul et al. (Paul et al., 2020) formulated interactions between defenders and adversaries as a repeated game, of which the solution is designed based on the reinforcement learning algorithm. Wei et al. (Wei et al., 2018) modeled interactions between defenders and attackers as a resource allocation stochastic game and introduce a novel learning algorithm to enable players to reach their equilibrium. Wei et al. (Wei et al., 2017) leveraged the Stackelberg game-theoretic model to model interactions between a single defender and multiple attackers, and then conduct a Likelihood Ratio Test (LRT) to detect malicious meters.

The third category is based on classification (Jokar et al., 2016; Singh et al., 2017; Ismail et al., 2018; Yeckle and Tang, 2018; Zheng et al., 2018; Fernandes et al., 2019; Jakaria et al., 2019; Punmiya and Choe, 2019; Zheng et al., 2019; Rouzbahani et al., 2020; Tehrani et al., 2020; Yan and Wen, 2021). For example, Singh et al. (Singh et al., 2017) proposed a detection scheme based on the principal component analysis (PCA) technology. The PCA was used to convert

high-dimensional data to low-dimensional data, after which anomaly scores were calculated and compared with predefined thresholds to find out attacks. Yan et al. (Yan and Wen, 2021) proposed a detection model based on the extreme gradient boosting algorithm including two phases. In the training phase, normal samples can be obtained after preprocessing, and then malicious samples were generated from normal data according to the attack type. After that, the normal and malicious samples were jointly trained in the classification model. In the application phase, the trained classifier is used to determine whether the new collected sample is normal or malicious. Jokar et al. (Jokar et al., 2016) proposed a SVM-based data integrity attack detection model. They compared the reported total consumption value with the actual total consumption value to find out the suspicious area. Then, the historical data and synthetic attack data were used to train a multiclass SVM to detect malicious data. Tehrani et al. (Tehrani et al., 2020) took collected consumption values of 24 h and their mean, standard deviation, minimum and maximum values as characteristics. Firstly, they used Kmeans for clustering, and then generated false data according to the synthetic attack method proposed in the literature (Jokar et al., 2016) to construct a complete dataset for training and testing the decision tree, random forest and gradient boosting. Zheng et al. (Zheng et al., 2019) developed a scheme combined the maximum information coefficient and CFSFDP for detecting malicious behaviors. Yeckle et al. (Yeckle and Tang, 2018) used seven outlier detection algorithms to detect anomalies. The same as literature (Tehrani et al., 2020), they preprocessed the data by using kmeans, and conduct simulation based on consumption of five customers, including seven different attack types. The comprehensive experiment results validate the effectiveness of data integrity attacks detection.

## 8 Conclusion

In this paper, we investigate a scaling attack detection model in AMI that considers consumption pattern diversity (*SA2CPD*), which can effectively distinguish normal data from false data. Specifically, we first perform Kmeans clustering to find out different power consumption patterns to avoid low detection efficiency. After the clustering is completed, the interval of each consumption pattern is used to binarize the power consumption data, so that most values of false data are 1, and all values in normal data are 0. Finally, 24 time periods are divided into two categories, that is $T_{in}$ and $T_{out}$. The decision tree is constructed based on time periods in $T_{out}$ and used as a classifier. Experimental results show that our proposed *SA2CPD* model can effectively detect false data. Compared with detection schemes that do not consider power consumption pattern diversity and other machine learning algorithms including the KNN model and the Naive Bayes model, the evaluation results show that our model has a higher *F1 score*, indicating that our approach is more efficient.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://blog.drhongtao.com/2016/07/gefcom2012-load-forecasting-data.html.

## Author contributions

XZ: Conceptualization, Methodology, Investigation, Formal Analysis, Validation, Writing—Original Draft and Review and Editing, Funding Acquisition; DC: Investigation, Data Curation, Formal Analysis, Validation, Writing—Original Draft; XL: Investigation, Validation, Writing—Original Draft.

## Funding

The work is supported in part by the National Science Foundation of China (NSFC) under grants: 62002210 and 62001273, and in part by the Key R and D program (international science and technology cooperation project) of Shanxi Province, China (No. 201903D421003), and in part by the Open Project Program of the Shaanxi Key Laboratory for Network Computing and Security Technology (NCST2021YB-02).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References


Amara korba, A., and El Islem karabadji, N. (2019). "Smart grid energy fraud detection using svm," in *2019 international conference on networking and advanced systems (ICNAS)*, 1–6. doi:10.1109/ICNAS.2019.8807832

Aziz, I. T., Abdulqadder, I. H., Alturfi, S. M., Imran, R. M., and Flaih, F. M. F. (2020). "A secured and authenticated state estimation approach to protect measurements in smart grids," in *2020 international conference on innovation and intelligence for informatics, computing and technologies (3ICT)*, 1–5. doi:10.1109/3ICT51146.2020.9311984

Bhattacharjee, S., and Das, S. K. (2021). Detection and forensics against stealthy data falsification in smart metering infrastructure. *IEEE Trans. Dependable Secure Comput.* 18, 356–371. doi:10.1109/TDSC.2018.2889729

Bhattacharjee, S., Madhavarapu, P., and Das, S. K. (2021a). A diversity index based scoring framework for identifying smart meters launching stealthy data falsification attacks. *Proc. 2021 ACM Asia Conf. Comput. Commun. Secur.*, 26–39. doi:10.1145/3433210.3437527

Bhattacharjee, S., Madhavarapu, V. P. K., Silvestri, S., and Das, S. K. (2021b). Attack context embedded data driven trust diagnostics in smart metering infrastructure. *ACM Trans. Priv. Secur.* 24, 1–36. doi:10.1145/3426739

Cardenas, A. A., Amin, S., Schwartz, G., Dong, R., and Sastry, S. (2012). "A game theory model for electricity theft detection and privacy-aware control in ami systems," in *2012 50th annual allerton conference on communication, control, and computing* (Allerton: IEEE), 1830–1837. doi:10.1109/Allerton.2012.6483444

Chaudhry, S. A., Yahya, K., Garg, S., Kaddoum, G., Hassan, M., and Zikria, Y. B. (2022). Las-sg: An elliptic curve based lightweight authentication scheme for smart grid environments. *IEEE Trans. Ind. Inf.* 1, 1–8. doi:10.1109/TII.2022.3158663

Choi, J. S., Lee, S., and Chun, S. J. (2021). A queueing network analysis of a hierarchical communication architecture for advanced metering infrastructure. *IEEE Trans. Smart Grid* 12, 4318–4326. doi:10.1109/TSG.2021.3088879

Fernandes, S. E. N., Pereira, D. R., Ramos, C. C. O., Souza, A. N., Gastaldello, D. S., and Papa, J. P. (2019). A probabilistic optimum-path forest classifier for non-technical losses detection. *IEEE Trans. Smart Grid* 10, 3226–3235. doi:10.1109/TSG.2018.2821765

Hong, T. (2014). *Global energy forecasting competition 2012.* Available at: http://blog.drhongtao.com/2016/07/gefcom2012-load-forecasting-data.html.

Hu, T., Guo, Q., Shen, X., Sun, H., Wu, R., and Xi, H. (2019). Utilizing unlabeled data to detect electricity fraud in ami: A semisupervised deep learning approach. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 3287–3299. doi:10.1109/TNNLS.2018.2890663

Huang, C., Sun, C.-C., Duan, N., Jiang, Y., Applegate, C., Barnes, P. D., et al. (2022). Smart meter pinging and reading through ami two-way communication networks to monitor grid edge devices and ders. *IEEE Trans. Smart Grid* 13, 4144–4153. doi:10.1109/TSG.2021.3133952

Huang, S.-C., Lo, Y.-L., and Lu, C.-N. (2013). Non-technical loss detection using state estimation and analysis of variance. *IEEE Trans. Power Syst.* 28, 2959–2966. doi:10.1109/TPWRS.2012.2224891

Ibrahem, M. I., Nabil, M., Fouda, M. M., Mahmoud, M. M. E. A., Alasmary, W., and Alsolami, F. (2021). Efficient privacy-preserving electricity theft detection with dynamic billing and load monitoring for ami networks. *IEEE Internet Things J.* 8, 1243–1258. doi:10.1109/JIOT.2020.3026692

Ismail, M., Shahin, M., Shaaban, M. F., Serpedin, E., and Qaraqe, K. (2018). "Efficient detection of electricity theft cyber attacks in ami networks," in *2018 IEEE wireless communications and networking conference (WCNC)*, 1–6. doi:10.1109/WCNC.2018.8377010

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognit. Lett.* 31, 651–666. doi:10.1016/j.patrec.2009.09.011

Jakaria, A. H. M., Rahman, M. A., and Moula Mehedi Hasan, M. G. (2019). "Safety analysis of ami networks through smart fraud detection," in *2019 IEEE conference on communications and network security (CNS)*, 1–7. doi:10.1109/CNS.2019.8802845

Jiang, R., Lu, R., Wang, Y., Luo, J., Shen, C., and Shen, X. (2014). Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Sci. Technol.* 19, 105–120. doi:10.1109/TST.2014.6787363

Jokar, P., Arianpoo, N., and Leung, V. C. M. (2016). Electricity theft detection in ami using customers' consumption patterns. *IEEE Trans. Smart Grid* 7, 216–226. doi:10.1109/TSG.2015.2425222

Leite, J. B., and Mantovani, J. R. S. (2018). Detecting and locating non-technical losses in modern distribution networks. *IEEE Trans. Smart Grid* 9, 1023–1032. doi:10.1109/TSG.2016.2574714

Lo, C.-H., and Ansari, N. (2013). Consumer: A novel hybrid intrusion detection system for distribution networks in smart grid. *IEEE Trans. Emerg. Top. Comput.* 1, 33–44. doi:10.1109/TETC.2013.2274043

McLaughlin, S., Holbert, B., Fawaz, A., Berthier, R., and Zonouz, S. (2013). A multi-sensor energy theft detection framework for advanced metering infrastructures. *IEEE J. Sel. Areas Commun.* 31, 1319–1330. doi:10.1109/JSAC.2013.130714

Mudgal, S., Pranjale, S., Balaji, T., Ahmed, S. A. A., Singh, N. K., Gupta, P. K., et al. (2022). Impact of cyber-attacks on economy of smart grid and their prevention. *UPjeng.* 8, 51–64. doi:10.24840/2183-6493_008.002_0005

Park, K., Lee, J., Das, A. K., and Park, Y. (2022). Bpps:blockchain-enabled privacy-preserving scheme for demand-response management in smart grid environments. *IEEE Trans. Dependable Secure Comput.* 1, 1–12. doi:10.1109/TDSC.2022.3163138

Paul, S., Ni, Z., and Mu, C. (2020). A learning-based solution for an adversarial repeated game in cyber-physical power systems. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 4512–4523. doi:10.1109/TNNLS.2019.2955857

Punmiya, R., and Choe, S. (2019). Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing. *IEEE Trans. Smart Grid* 10, 2326–2329. doi:10.1109/TSG.2019.2892595

Rouzbahani, H. M., Karimipour, H., and Lei, L. (2020). "An ensemble deep convolutional neural network model for electricity theft detection in smart grids," in *2020 IEEE international conference on systems, man, and cybernetics (SMC)*, 3637–3642. doi:10.1109/SMC42975.2020.9282837

Safavian, S., and Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man. Cybern.* 21, 660–674. doi:10.1109/21.97458

Salinas, S., Luo, C., Liao, W., and Li, P. (2014). "State estimation for energy theft detection in microgrids," in *9th international conference on communications and networking in China*, 96–101. doi:10.1109/CHINACOM.2014.7054266

Sarenche, R., Salmasizadeh, M., Ameri, M. H., and Aref, M. R. (2021). A secure and privacy-preserving protocol for holding double auctions in smart grid. *Inf. Sci.* 557, 108–129. doi:10.1016/j.ins.2020.12.038

Singh, N. K., and Mahajan, V. (2021). End-user privacy protection scheme from cyber intrusion in smart grid advanced metering infrastructure. *Int. J. Crit. Infrastructure Prot.* 34, 100410. doi:10.1016/j.ijcip.2021.100410

Singh, S. K., Bose, R., and Joshi, A. (2017). "Pca based electricity theft detection in advanced metering infrastructure," in *2017 7th international conference on power systems (ICPS)*, 441–445. doi:10.1109/ICPES.2017.8387334

Sun, C.-C., Sebastian Cardenas, D. J., Hahn, A., and Liu, C.-C. (2021). Intrusion detection for cybersecurity of smart meters. *IEEE Trans. Smart Grid* 12, 612–622. doi:10.1109/TSG.2020.3010230

Tehrani, S. O., Moghaddam, M. H. Y., and Asadi, M. (2020). "Decision tree based electricity theft detection in smart grid," in *2020 4th international conference on smart city, Internet of things and applications (SCIOT)*, 46–51. doi:10.1109/SCIOT50840.2020.9250194

Verma, G., Gope, P., Saxena, N., and Kumar, N. (2022). Cb-da: Lightweight and escrow-free certificate-based data aggregation for smart grid. *IEEE Trans. Dependable Secure Comput.* 1, 1–13. doi:10.1109/TDSC.2022.3169952

Wei, L., Sarwat, A. I., Saad, W., and Biswas, S. (2018). Stochastic games for power grid protection against coordinated cyber-physical attacks. *IEEE Trans. Smart Grid* 9, 684–694. doi:10.1109/TSG.2016.2561266

Wei, L., Sundararajan, A., Sarwat, A. I., Biswas, S., and Ibrahim, E. (2017). "A distributed intelligent framework for electricity theft detection using benford's law and stackelberg game," in *2017 resilience week (RWS)*, 5–11. doi:10.1109/RWEEK.2017.8088640

Yan, Z., and Wen, H. (2021). Electricity theft detection base on extreme gradient boosting in ami. *IEEE Trans. Instrum. Meas.* 70, 1–9. doi:10.1109/TIM.2020.3048784

Yang, X., He, X., Lin, J., Yu, W., and Yang, Q. (2016). "A game-theoretic model on coalitional attacks in smart grid," in *2016 IEEE trustcom/BigDataSE/ISPA*, 435–442. doi:10.1109/TrustCom.2016.0094

Yao, D., Wen, M., Liang, X., Fu, Z., Zhang, K., and Yang, B. (2019). Energy theft detection with energy privacy preservation in the smart grid. *IEEE Internet Things J.* 6, 7659–7669. doi:10.1109/JIOT.2019.2903312

Yeckle, J., and Tang, B. (2018). "Detection of electricity theft in customer consumption using outlier detection algorithms," in *2018 1st international conference on data intelligence and security (ICDIS)*, 135–140. doi:10.1109/ICDIS.2018.00029

Zanetti, M., Jamhour, E., Pellenz, M., Penna, M., Zambenedetti, V., and Chueiri, I. (2019). A tunable fraud detection system for advanced metering infrastructure using short-lived patterns. *IEEE Trans. Smart Grid* 10, 830–840. doi:10.1109/TSG.2017.2753738

Zheng, K., Chen, Q., Wang, Y., Kang, C., and Xia, Q. (2019). A novel combined data-driven approach for electricity theft detection. *IEEE Trans. Ind. Inf.* 15, 1809–1819. doi:10.1109/TII.2018.2873814

Zheng, Z., Yang, Y., Niu, X., Dai, H.-N., and Zhou, Y. (2018). Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Trans. Ind. Inf.* 14, 1606–1615. doi:10.1109/TII.2017.2785963

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 16 January 2023
DOI 10.3389/fenrg.2022.1058125



OPEN ACCESS

EDITED BY
Dou An,
MOE Key Laboratory for Intelligent Networks and Network Security, China

REVIEWED BY
Chengyu Hu,
Shandong University, China
Yalong Wu,
University of Houston—Clear Lake, United States

*CORRESPONDENCE
Wenting Shen,
shenwentingmath@163.com

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 30 September 2022
ACCEPTED 31 October 2022
PUBLISHED 16 January 2023

CITATION
Gai C, Shen W, Yang M and Su Y (2023), Certificateless public auditing with data privacy preserving for cloud-based smart grid data.
*Front. Energy Res.* 10:1058125.
doi: 10.3389/fenrg.2022.1058125



# Certificateless public auditing with data privacy preserving for cloud-based smart grid data

Chao Gai[1], Wenting Shen[1,2]*, Ming Yang[2] and Ye Su[3]

[1]College of Computer Science and Technology, Qingdao University, Qingdao, China, [2]Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, [3]School of Information Science and Engineering, Shandong Normal University, Jinan, China

As the promising next generation power system, smart grid can collect and analyze the grid information in real time, which greatly improves the reliability and efficiency of the grid. However, as smart grid coverage expands, more and more data is being collected. To store and manage the massive amount of smart grid data, the data owners choose to upload the grid data to the cloud for storage and regularly check the integrity of their data. However, traditional public auditing schemes are mostly based on Public Key Infrastructure (PKI) or Identity Based Cryptography (IBC) system, which will lead to complicated certificate management and inherent key escrow problems. We propose a certificateless public auditing scheme for cloud-based smart grid data, which can avoid the above two problems. In order to prevent the disclosure of the private data collected by the smart grid during the phase of auditing, we use the random masking technology to protect data privacy. The security analysis and the performance evaluation show that the proposed scheme is secure and efficient.

KEYWORDS

**smart grid, certificateless public auditing, cloud computing, cloud storage, privacy—preserving**

## 1 Introduction

With the development of information technology, the smart grid becomes a new promising power system, which is allowed to collect and analyze smart grid data and provides more reliable, cost-effetive and efficient power management compared to traditional power grids (Chen et al., 2014; He et al., 2018a; Zhang et al., 2021c; Peng et al., 2021). A large amount of data are colleted with the expansion of smart grid coverage. Nevertheless, the traditional smart grid data management system without large storage space is unable to meet the data owners' storage requirements. Thus, more and more data owners choose to store smart grid data on the cloud.

Although the cloud provides a large amount of storage and computing resources for the data owners, there are some security issues that cannot be ignored (Zhang et al., 2021a; Yang et al., 2020; Lu et al., 2022; Shao et al., 2022). For example, the smart grid data stored in the cloud might be corrupted by hacker attacks, administrator's error operation, and damaged devices. Once the data is uploaded to the cloud, the data

owner will lose the physical control of the smart grid data stored in the cloud and cannot directly determine whether the data is intact or not. In order to ensure the integrity of cloud data, plenty of public auditing schemes are proposed (Ji et al., 2022; Li et al., 2021; Zhou et al., 2021; Liu et al., 2022). In public auditing, the data owner can delegate the data integrity auditing tasks to a Third Party Auditor (TPA) with abundant computation resources. In practice, the data collected by the smart grid might contain sensitive data, such as regional electricity consumption habits, residential electricity consumption patterns, etc (McDaniel and McLaughlin, 2009; Liu et al., 2021; Zhang et al., 2021b). Once the TPA is delegated to audit the data integrity, the data owner's data will inevitably be exposed to the TPA. The TPA is able to obtain sensitive information during the auditing phase. Therefore, it is critical to protect data privacy from the TPA in public auditing.

Nevertheless, most of the existing public auditing schemes are based on the traditional Public Key Infrastructure (PKI), which can lead to complex certificate management issue. In order to solve this problem, Identity Based Cryptography (IBC) had been proposed. In IBC system, there is a key generation center (KGC) which uses the data owner's identity to generate a private key for the data owner. The data owner can use his own identity as his public key. IBC eliminates the certificate management problem of PKI. However, the KGC holds the user's private key, the security of the user's private key will completely depend on the KGC, which leads to the inherent key escrow problem (Al-Riyami and Paterson, 2003; Wang et al., 2013). Therefore, in order to obtain better efficiency and higher security, the certificateless public key cryptography is proposed (Zhou et al., 2022; Zhang J et al., 2020; Xu et al., 2021). In certificateless public key cryptography systems, the data owner's private key is jointly generated by the KGC and the data owner. Therefore, the KGC does not know the data owner's complete private key. Certificateless public key cryptography can solve the inherent key escrow problem of IBC.

The contribution of our scheme can be summarized as follows:

1) Based on the certificateless public key cryptography, we proposed a certificateless public auditing scheme. Different from the existing public auditing schemes based on PKI or IBC, our scheme can avoid complex certificate management problem and key escrow problem.
2) To achieve data privacy preserving, we utilize the novel random masking technology in the phase of auditing. The TPA cannot obtain the sensitive data from the proof generated by the cloud.
3) We give the security proof of the proposed scheme. Furthermore, the theoretical analysis and experimental results show that the proposed scheme is efficient.

## 1.1 Related work

Ateniese et al. (2007) proposed the first "Provable Data Possession" (PDP) scheme, in which the integrity of the remote data can be checked by the client. In this scheme, the homomorphic authenticators and the random sampling technique are employed to achieve the data integrity checking. Juels and Kaliski (2007) constructed a "Proofs of Retrievability" (PoR) scheme, which guarantees the data integrity and data retrievability on the cloud. However, in this scheme, the verifier can only perform a finite number of data integrity verification. In 2008, Shacham and Waters (2008) designed an improved PoR scheme, which is provably secure.

To support data dynamic, Ateniese et al. (2008) constructed a PDP scheme supporting data dynamic operations. Guo et al. (2020) designed a dynamic proof of data possession and replication scheme. In this scheme, the multiple replicas share a single authenticated tree. Erway et al. (2015) designed a rank-based skip list and constructed the PDP scheme supporting full data dynamic operations. Wang et al. (2019) proposed a blockchain-based private data integrity verification scheme by using RSA signature. Wang et al. (2017b) designed a cloud storage auditing scheme based on the online/offline signature, in which the data owner can reduce the burden of authenticator generation in the online phase. To improve the auditing efficiency, Gao et al. (2021) designed a data integrity checking scheme based on the keyword. This scheme allows the TPA to verify the integrity of files containing the specified keyword. To preserve the data privacy, Li et al. (2018) designed a privacy-preserving data integrity verification scheme with zero-knowledge proof. In addition, there are many researches devoted to the key exposure problem (Yu et al., 2016; Yu and Wang, 2017; Xu et al., 2020).

To solve complex certificate management, in 1984, Identity Based Cryptography (IBC) is proposed by Shamir (1985). In IBC system, the data owner's private key is calculated by a trusted Key Generation Center (KGC) with the data owner's identity. The data owner uses his identity as the public key, which eliminates the complex certificate management. Wang et al. (2014) proposed the first identity-based data integrity checking scheme by using the Schnorr signature. To support efficient user revocation, Zhang Y et al. (2020) designed an identity-based data integrity verification scheme for shared data. Shen et al. (2019) proposed a data integrity checking scheme supporting data sharing and sensitive data hiding. In their scheme, the data owner's sensitive data can be protected under the assistance of the sanitizer. Wang et al. (2017a) designed an identity-based comprehensive data integrity checking scheme, in which the authenticators can be generated with the help of the proxy. To protect the data owner's identity privacy, Zhang et al. (2019) utilized the anonymous identity to replace the data owner's real identity, and constructed a

conditinal identity privacy-preserving data integrity checking scheme.

Unfortunately, although the IBC system can avoid the certificate management problem caused by PKI, it still has the inherent key escrow issue. Zhang et al. (2015) designed a secure certificateless public data integrity verification scheme, which can resist the malicious TPA. He et al. (2018b) proposed a certificateless data integrity auditing scheme which can resist the attacks of two types of adversaries in certificateless crytography (The adversary is able to replace the public keys of the users and the adversary is able to access the master key of the KGC). To eliminate the problem of key escrow in IBC, Wu et al. (2019) proposed a certificateless public auditing scheme which supports identity privacy protection. Zhou et al. (2022) applied the certificateless technology to the multi-replica environment. This scheme can realize the efficient data dynamic in the multi-replica environment by using the new Merkle Hash Tree structure.

## 1.2 Organization

The remainder of this paper is organized as follows: In **Section 2**, we introduce the system model and design goals of our scheme; In **Section 3**, we describe the preliminaries and definition; We give a detailed algorithm of our scheme in **Section 4**; In **Section 5**, we analyzed the security of our scheme; We show the performance analysis of our scheme in **Section 6**; We make a conclusion in **Section 7**.

# 2 System model and design goals

We give the system model and the design goals in this section.

## 2.1 System model

As shown in **Figure 1**, the system model of our scheme contains four entities: the cloud, the data owner, the Key Generation Center (KGC) and the Third Party Auditor (TPA).

1) The cloud: The cloud is an entity which has enormous storage and computation resources. It is responsible for storing and managing the smart grid data for the data owner. After receiving the TPA's challenge, the cloud needs to send the corresponding auditing proof to the TPA.
2) The data owner: The data owner is an entity with limited storage space and computation resources. He outsources his smart grid data to the cloud for storage and delegates the TPA to verify the integrity of cloud data.

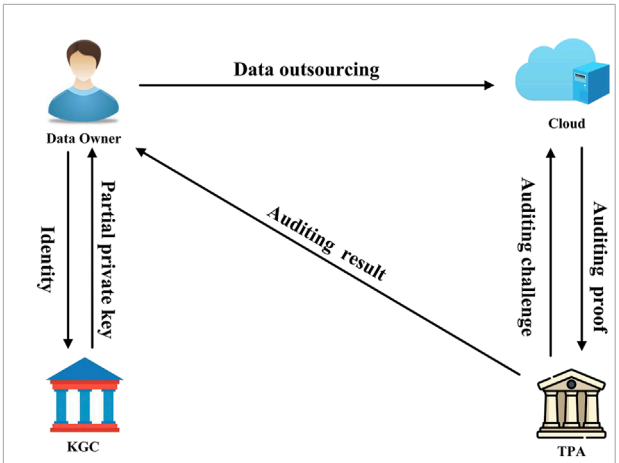



FIGURE 1
System model.

3) KGC: The KGC is an entity which takes charge of producing the system parameters and calculating the partial private key for the data owner based on the data owner's identity.
4) TPA: The TPA is an entity with powerful computing capabilities, which needs to generate and deliver the auditing challenge to the cloud and audit the integrity of the cloud data.

## 2.2 Design goals

In order to achieve privacy preserving in certificateless public auditing for cloud-based smart grid data, our scheme needs to meet the following design goals:

1) Correctness: If the KGC generates the partial private key for the data owner honestly, the partial private key can pass the data owner's checking. If the cloud generates the auditing proof honestly, the auditing proof can pass the TPA's checking.
2) Soundness: The cloud cannot pass the TPA's verification if the data has been corrupted.
3) Privacy protection: In the phase of data integrity auditing, the TPA cannot obtain the smart grid data from the cloud's auditing proof.

# 3 Preliminaries and definition

In this section, we present the preliminaries applied in our scheme. Then, we give the definitions of our scheme.

## 3.1 Preliminaries

### 3.1.1 Bilinear pairing

Suppose that there are two different multiplicative cyclic groups $G_1$ and $G_T$ with the same prime order $p$. The generator of the group $G_1$ is $g$. If $e$:$G_1 \times G_1 \rightarrow G_T$ is a bilinear pairing, it satisfies (Boneh et al., 2001):

i) Bilinearity: For $\forall u, v \in G_1$ and $\forall \alpha, \beta \in Z_p^*$, we have $e(u^\alpha, v^\beta) = e(u,v)^{\alpha\beta}$.
ii) Non-degeneracy: $\exists u, v \in G_1$ and $e(u,v) \neq 1_{G_T}$.
iii) Computability: For $\forall u, v \in G_1$, $e(u,v)$ is able to be computed efficiently.

### 3.1.2 Computational diffie-hellman problem

Given $g, g^\alpha, g^\beta \in G_1$, where g is the generator of $G_1$ and $\alpha, \beta \in Z_p^*$, compute $g^{\alpha\beta} \in G_1$. The CDH assumption in $G_1$ holds if it is hard to solve the CDH problem in $G_1$ (Bao et al., 2003).

### 3.1.3 Discrete logarithm problem

Given $g, g^\alpha \in G_1$, where g is the generator of $G_1$ and $\alpha \in Z_p^*$, compute $\alpha$. The DL assumption in $G_1$ holds if it is hard to solve the DL problem in $G_1$ (McCurley, 1990).

## 3.2 Definition

Definition 1: Our scheme includes seven algorithms: Setup, PartialKeyGen, PrivateKeyGen, AuthenticatorGen, ChallengeGen, ProofGen and ProofVerify.

1) Setup $(1^\lambda) \rightarrow (sk_K, params)$: This algorithm is run by the KGC. Taking $\lambda$ as input, it outputs the KGC's master secret key $sk_K$ and the system parameters *params*.
2) PartialKeyGen $(ID_O, sk_K, params) \rightarrow (\sigma_O)$: This algorithm is run by the KGC and the data owner. Inputting the data owner's identify $ID_O$, the KGC's master secret key $sk_K$ and the system parameters *params*, it outputs the data owner's partial private key $\sigma_O$. The data owner can check whether the partial private key $\sigma_O$ is valid or not.
3) PrivateKeyGen $(\sigma_O, params) \rightarrow (sk_O, pk_O)$: This algorithm is run by the data owner. Taking the partial private key $\sigma_O$ and the system parameters *params* as input, it outputs the data owner's private key $sk_O$ and the corresponding public key $pk_O$.
4) AuthenticatorGen $(sk_O, F, ID_F) \rightarrow (T, tag)$: This algorithm is run by the data owner. Inputting the data owner's private key $sk_O$, the file $F$ and the file's identifier $ID_F$, it generates the authenticator set $T$ and the file tag *tag*.
5) ChallengeGen $(s, K_1, K_2) \rightarrow (chal)$: This algorithm is run by the TPA. Taking three random values $s$, $K_1$ and $K_2$ as input, it produces the auditing challenge *chal*.
6) ProofGen $(chal, F, T) \rightarrow (proof)$: This algorithm is run by the cloud. Taking the challenge *chal*, the file $F$ and the authenticator set $T$ as input, it outputs the auditing proof *proof*.
7) ProofVerify $(tag, proof, pk_K, pk_O) \rightarrow 0,1$: This algorithm is run by the TPA. Inputting the file tag *tag*, the auditing proof *proof*, the KGC's public key $pk_K$ and the data owner's public key $pk_O$, it outputs the auditing result. If the proof is valid, the result is "1"; otherwise, the result is "0."

# 4 The proposed scheme

In this section, we give the detailed algorithms of our scheme.

1) Setup $(1^\lambda) \rightarrow (sk_K, params)$
   a) Let $e$ be a bilinear pairing $e$:$G_1 \times G_1 \rightarrow G_T$, where $G_1$ and $G_T$ are two different cyclic multiplicative groups with the same prime order $p$. The KGC selects two independent generators $g$ and $u$ of the group $G_1$ and sets two different hash functions: $H$:$\{0.1\}^* \rightarrow G_1$ and $H_1$:$\{0.1\}^* \rightarrow G_1$.
   b) The KGC randomly picks an element $sk_K \in Z_p^*$ as its master secret key, generates $pk_K = g^{sk_K}$ as its system public key, and publishes the system parameter $params = (G_1, G_T, e, g, u, pk_K, H, H_1)$.
2) PartialKeyGen $(ID_O, sk_K, params) \rightarrow (\sigma_O)$
   a) The data owner transmits his identify $ID_O \in \{0,1\}^*$ to the KGC.
   b) The KGC computes $\sigma_O = H(ID_O)^{sk_K}$, and transmits $\sigma_O$ to the data owner as his partial private key.
   c) After receiving $\sigma_O$, the data owner checks whether the partial private key is correct or not according to the following equation

$$e(\sigma_O, g) = e(H(ID_O), pk_K) \tag{1}$$

If **Equation 1** holds, the data owner accepts the partial private key $\sigma_O$.

3) PrivateKeyGen $(\sigma_O, params) \rightarrow (sk_O, pk_O)$

The data owner picks a random value $x \in Z_p^*$ and sets $sk_O = \{x, \sigma_O\}$ as his private key. The data owner calculates $pk_O = g^x$ as his public key.

4) AuthenticatorGen $(sk_O, F, ID_F) \rightarrow (T, tag)$
   a) The data owner divides the file $F$ into $n$ data blocks $d_i$ $(i \in [1,n])$. The data owner generates the corresponding authenticator $t_i = \sigma_O \cdot (H_1(ID_F \| i \| n) u^{d_i})^x$ for each data block $d_i$ $(i \in [1,n])$, where $ID_F$ is the identifier of the file $F$. The file $F$'s authenticator set is denoted as $T = \{t_i\}_{1 \le i \le n}$.
   b) The data owner produces a file tag $tag = ID_F \| n \| SSig_{ssk}(ID_F \| n)$ using the signature *SSig*, where *ssk* is the private key of the signature *SSig*.
   c) The data owner uploads the file $F$, the authenticator set $T$ and the file tag *tag* to the cloud.

5) ChallengeGen $(s, K_1, K_2) \rightarrow (chal)$
   a) For each challenge, the TPA randomly picks three values $s$ ($s \in [1, n]$) and $K_1, K_2 \in Z_p^*$, where $K_1$ is the key of pseudo-random permutation $\pi_{K_1}(\cdot)$ and $K_2$ is the key of pseudo-random function $\phi_{K_2}(\cdot)$.
   b) The TPA sends the challenge $chal = \{s, K_1, K_2\}$ to the cloud.
6) ProofGen $(chal, F, T) \rightarrow (proof)$
   a) According to the challenge *chal*, the cloud generates the challenged block's index $l_j = \pi_{K_1}(j)$ for each $1 \le j \le s$, where $l_j \in [1, n]$.
   b) The cloud calculates a random value $v_j = \phi_{K_2}(j)$ for each $1 \le j \le s$, in which $v_j \in Z_p^*$.
   c) The cloud computes $\Gamma = \prod_{j=1}^{s} t_{l_j}{}^{v_j}$, $\mu' = \sum_{j=1}^{s} d_{l_j} v_j$.
   d) In order to protect the data privacy, the cloud chooses a random element $r \in Z_p^*$ and computes $\mu = \mu' - r$ to blind $\mu'$. The cloud calculates $R = u^r$.
   e) The cloud transmits the proof $proof = (\mu, \Gamma, R)$ and the file tag *tag* to the TPA.
7) ProofVerify $(tag, proof, pk_K, pk_O) \rightarrow 0.1$
   a) The TPA verifies the validity of the file tag *tag*. If *tag* is valid, the TPA parses the file's identifier $ID_F$ and the number of data blocks $n$.
   b) For each $1 \le j \le s$, the TPA calculates $l_j = \pi_{K_1}(j)$ and $v_j = \phi_{K_2}(j)$.
   c) The TPA verifies whether the auditing proof*proof* is valid or not according to the following equation

$$e(\Gamma, g) = e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu} \cdot R, pk_O\right) \quad (2)$$

If the above equation holds, the TPA returns "1", which means that the remote data is intact; otherwise, it returns "0", which means that the remote data is broken.

# 5 Security analysis

In this section, we give the security proof of our scheme from the aspects of correctness, soundness and data privacy protection.

## 5.1 Theorem 1 (Correctness)

In our scheme, if the KGC, the TPA, and the cloud honestly perform the specified procedures, the partial private key and the auditing proof are able to pass the verification.

### 5.1.1 Proof

The derivation process for the data owner to verify whether the partial key is correct is as follows:

$$\begin{aligned} e(\sigma_O, g) &= e\left(H(ID_O)^{sk_K}, g\right) \\ &= e\left(H(ID_O), g^{sk_K}\right) \\ &= e\left(H(ID_O), pk_K\right) \end{aligned}$$

The derivation process for the TPA to verify whether the auditing proof is valid is as follows:

$$\begin{aligned} & e(\Gamma, g) \\ &= e\left(\prod_{j=1}^{s} t_{l_j}{}^{v_j}, g\right) \\ &= e\left(\prod_{j=1}^{s} \left(\sigma_O \cdot \left(H_1\left(ID_F \| l_j \| n\right) u^{d_{l_j}}\right)^{x}\right)^{v_j}, g\right) \\ &= e\left(\prod_{j=1}^{s} \sigma_O{}^{v_j} \cdot \prod_{j=1}^{s} \left(H_1\left(ID_F \| l_j \| n\right) u^{d_{l_j}}\right)^{x v_j}, g\right) \\ &= e\left(\prod_{j=1}^{s} \sigma_O{}^{v_j}, g\right) \cdot e\left(\prod_{j=1}^{s} \left(H_1\left(ID_F \| l_j \| n\right) u^{d_{l_j}}\right)^{x v_j}, g\right) \\ &= e\left(\sigma_O{}^{\sum_{j=1}^{s} v_j}, g\right) \cdot e\left(\prod_{j=1}^{s} \left(H_1\left(ID_F \| l_j \| n\right) u^{d_{l_j}}\right)^{v_j}, g^{x}\right) \\ &= e\left(H(ID_O)^{sk_K \cdot \sum_{j=1}^{s} v_j}, g\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot \prod_{j=1}^{s} u^{d_{l_j} v_j}, pk_O\right) \\ &= e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, g^{sk_K}\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\sum_{j=1}^{s} d_{l_j} v_j}, pk_O\right) \\ &= e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu'}, pk_O\right) \\ &= e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu + r}, pk_O\right) \\ &= e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu} \cdot u^{r}, pk_O\right) \end{aligned}$$

$$= e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu} \cdot R, pk_O\right)$$

## 5.2 Theorem 2 (Soundness)

Suppose the CDH assumption holds in $G_1$ and the signature scheme used for generating tag is existentially unforgeable. In our scheme, for an adversary, it is computationally infeasible to generate a bogus proof that is able to pass the TPA's checking if the cloud data has been damaged.

*Proof.* We will prove this theorem with the method of knowledge proof. The malicious cloud is viewed as adversary and the user plays the role on the challenger.

**Game 0**. If the adversary submits one tag, the challenger will abort if the tag is a valid *SSig* signature but not signed by the challenger.

**Analysis**. If the challenger aborts in Game 0 with non-negligible probability, the adversary is able to forge a valid *SSig* signature. This contradicts the assumption that *SSig* is an unforgeable signature. Therefore, the file identifier and the number of data blocks in the interactions with the adversary are all valid and generated by the challenger.

**Game 1**. Game 1 is the same as Game 0, with only one difference. The challenger keeps a list of his responses to the queries from the adversary. If the adversary wins the game 1 but the aggregated authenticator $\Gamma^*$ in the proof is different from $\Gamma = \prod_{j=1}^{s} t_{l_j}^{v_j}$, then the challenger will abort.

**Analysis**. Assume $proof = (\mu, \Gamma, R)$ is a valid proof. We have:

$$e(\Gamma, g) = e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu} \cdot R, pk_O\right) \quad (3)$$

Suppose $proof^* = (\mu^*, \Gamma^*, R)$ is a forged auditing proof, where $\Gamma^*$ is different from $\Gamma$. Because the forgery is successful, $proof^*$ can pass the verification of the following equation:

$$e\left(\Gamma^*, g\right) = e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu^*} \cdot R, pk_O\right) \quad (4)$$

It is obviously that $\mu \neq \mu^*$; otherwise $\Gamma = \Gamma^*$, which contradicts the above assumption. Let $\Delta\mu = \mu^* - \mu$. The adversary can win the game 1 with a non-negligible probability only if there is a simulator can solve the CDH problem.

Given $g, g^{\varepsilon}, h \in G_1$, the simulator needs to generate $h^{\varepsilon}$. The simulator picks two random values $\beta, \theta \in Z_p^*$ and sets $u = g^{\beta} h^{\theta}$ and $pk_O = g^{\varepsilon}$. For each $l_j$, the simulator selects a random value $v_j$ and programs a random oracle at $l_j$ as $H_1(ID_F \| l_j \| n) = g^{v_j} / (g^{\beta} h^{\theta})^{d_{l_j}}$. So, we can obtain $H_1(ID_F \| l_j \| n) u^{d_{l_j}} = g^{v_j} / (g^{\beta} h^{\theta})^{d_{l_j}} \cdot (g^{\beta} h^{\theta})^{d_{l_j}} = g^{v_j}$.

Dividing **Eq. 4** by **Eq. 3**, we have

$$\begin{aligned} & e\left(\Gamma^* / \Gamma, g\right) \\ &= e\left(u^{\mu^* - \mu}, pk_O\right) \\ &= e\left(\left(g^{\beta} h^{\theta}\right)^{\Delta\mu}, pk_O\right) \\ &= e\left(g^{\beta\Delta\mu}, pk_O\right) \cdot e\left(h^{\theta\Delta\mu}, pk_O\right) \\ &= e\left(g^{\beta\Delta\mu}, g^{\varepsilon}\right) \cdot e\left(h^{\theta\Delta\mu}, g^{\varepsilon}\right) \\ &= e(g,g)^{\beta\Delta\mu\varepsilon} \cdot e(h^{\varepsilon}, g)^{\theta\Delta\mu} \end{aligned}$$

According to the above equation, we can obtain $e(\Gamma^* / \Gamma \cdot g^{-\beta\Delta\mu\varepsilon}, g) = e(h^{\varepsilon}, g)^{\theta\Delta\mu}$. So, we have $h^{\varepsilon} = (\Gamma^* / \Gamma \cdot g^{-\beta\Delta\mu\varepsilon})^{\frac{1}{\theta\Delta\mu}}$.

The probability of $\theta\Delta\mu \neq 0$ is $1 - \frac{1}{p}$, which is non-negligible. So, we can solve the CDH problem with the probability $1 - \frac{1}{p}$, which is contradiction with the assumption that the CDH problem in $G_1$ is computationally infeasible.

**Game 2**. Game 2 is similar to Game 1, with one difference. The challenger records all interactions with the adversary.

If the adversary wins the game 2 but the aggregated data block $\mu^*$ in the proof is different from the expected one $\mu$, then the challenger will abort.

**Analysis**. Suppose $proof = (\mu, \Gamma, R)$ is a valid proof. We get:

$$e(\Gamma, g) = e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu} \cdot R, pk_O\right)$$

Assume $proof^* = (\mu^*, \Gamma^*, R)$ is a forged auditing proof. Because the forgery is successful, we get:

$$e\left(\Gamma^*, g\right) = e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu^*} \cdot R, pk_O\right)$$

Based on Game 1, we have $\Gamma = \Gamma^*$. Set $\Delta\mu = \mu^* - \mu$, we can design a simulator to solve the DL problem.

Inputting $g, h \in G_1$, the simulator needs to output $\varepsilon$ satisfying $h = g^{\varepsilon}$. The simulator selects two random values $\beta, \theta \in Z_p^*$ and

sets $u = g^{\beta}h^{\theta}$. Based on $\Gamma = \Gamma^{*}$, we obtain

$$\begin{aligned} & e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \\ & \quad \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu} \cdot R, pk_O\right) \\ & = e(\Gamma, g) \\ & = e\left(\Gamma^{*}, g\right) \\ & = e\left(H(ID_O)^{\sum_{j=1}^{s} v_j}, pk_K\right) \\ & \quad \cdot e\left(\prod_{j=1}^{s} H_1\left(ID_F \| l_j \| n\right)^{v_j} \cdot u^{\mu^{*}} \cdot R, pk_O\right) \end{aligned}$$

Further, we obtain that $u^{\mu} = u^{\mu^{*}}$ and $1 = u^{\Delta\mu} = (g^{\beta}h^{\theta})^{\Delta\mu} = g^{\beta\Delta\mu}h^{\theta\Delta\mu}$. Hence, we can solve the DL problem as follow:

$$h = g^{-\frac{\beta\Delta\mu}{\theta\Delta\mu}} = g^{-\frac{\beta}{\theta}}$$

The probability of $\theta \neq 0$ is $1 - \frac{1}{p}$, and it is non-negligible. So, we can solve the DL problem with the non-negligible probability $1 - \frac{1}{p}$, which is contradiction with the assumption that the DL problem in $G_1$ is computationally infeasible. Therefore, if the cloud can pass the TPA's verification with non-negligible probability, it means that the cloud correctly stores the smart grid data.

## 5.3 Theorem 3 (data privacy protection)

In our scheme, the TPA cannot extract the real data from the cloud's auditing proof.

*Proof.* On the one hand, in the auditing proof $= (\mu, \Gamma, R)$, the original aggregated data block $\mu' = \sum_{j=1}^{s} d_{l_j} v_j$ is blinded as $\mu$ by the random value $r$, where $\mu = \mu' - r$. Because the DL problem in $G_1$ is hard, the TPA cannot extract the value $r$ from $R$, where $R = u^{r}$. Thus, the TPA cannot obtain the original aggregated data block $\mu'$ from $\mu$. On the other hand, we get that

$$\begin{aligned} \Gamma &= \prod_{j=1}^{s} t_{l_j}{}^{v_j} = \prod_{j=1}^{s} \left(\sigma_O \cdot \left(H_1\left(ID_F \| l_j \| n\right) u^{d_{l_j}}\right)^{x}\right)^{v_j} \\ &= \prod_{j=1}^{s} \left(\sigma_O \cdot H_1\left(ID_F \| l_j \| n\right)^{x}\right)^{v_j} \cdot \left(u^{\mu'}\right)^{x} \end{aligned}$$

From the above equation, we know that $(u^{\mu'})^{x}$ is blinded by $\prod_{j=1}^{s} (\sigma_O \cdot H_1(ID_F \| l_j \| n)^{x})^{v_j}$. It is computational infeasible to compute $\prod_{j=1}^{s} (\sigma_O \cdot H_1(ID_F \| l_j \| n)^{x})^{v_j}$ from $\prod_{j=1}^{s} (\sigma_O \cdot H_1(ID_F \| l_j \| n))^{v_j}$ and $g^{x}$ because the CDH problem in $G_1$ is hard. So, the TPA cannot get $(u^{\mu'})^{x}$ from $\Gamma$. Consequently, the TPA cannot obtain the real smart grid data during the auditing phase.

TABLE 1 Computation overhead in different phases.

| Phase | Computation overhead |
|---|---|
| Partial key generation | $(2P + 2H + E)$ |
| Authenticator generation | $n(H + 2M + 2E)$ |
| Proof generation | $(2s - 1)M + sE + (s - 1)A$ |
| Proof verification | $3P + (s + 1)(H + M) + (s + 2)E + (s - 1)A$ |

TABLE 2 Communication overhead.

| Phase | Communication overhead |
|---|---|
| Partial key generation | $\|q\| + \|p\|$ |
| Data outsourcing | $n\|q\| + (n + 1)\|p\|$ |
| Data integrity auditing | $\|n\| + 2\|q\| + 3\|p\|$ |

# 6 Performance analysis

In this section, we systematically analyze the performance of our scheme from both theoretical analysis and experimental results.

## 6.1 Theoretical analysis

We respectively use $P, H, M, E$, and $A$ to denote one pairing operation, one hash operation, one multiplication operation, one exponentiation operation and one addition operation. Suppose that the file is divide into $n$ data blocks, and the TPA challenges $s$ data blocks. In Table 1, we describe the computation overhead of our scheme in different phases. In the phase of partial key generation, the computation overhead is $(2P + 2H + E)$. In the phase of authenticator generation, the user requires $n$ $(H + 2M + 2E)$ computation overhead to generate the authenticators. In the phase of proof generation, the computation overhead on the cloud side is $(2s - 1)M + sE + (s - 1)A$. In the phase of proof verification, the TPA needs to cost $3P + (s + 1)(H + M) + (s + 2)E + (s - 1)$ to verify the auditing proof.

Let $|q|$, $|p|$ and $|n|$ be the size of an element in $G_1$, $Z_p^{*}$ and set $[1, n]$ respectively. We present the communication overhead of our scheme in Table 2. The communication overhead of partial key generation is $|q| + |p|$. The communication overhead of data outsourcing is $n|q| + (n + 1)|p|$. In the data integrity auditing phase, the communication overhead is $|n| + 2|q| + 3|p|$.

## 6.2 Experimental results

In order to show the performance of our scheme, we design a series of experiments to simulate our scheme. We utilize C programming language with the GNU Multiple


FIGURE 2
The computation overhead of authenticator generation.

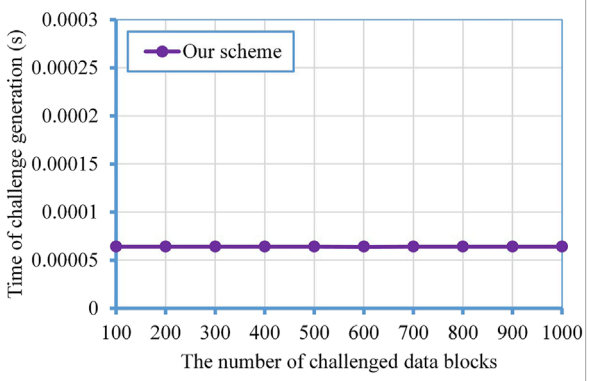



FIGURE 3
The computation overhead of challenge generation.

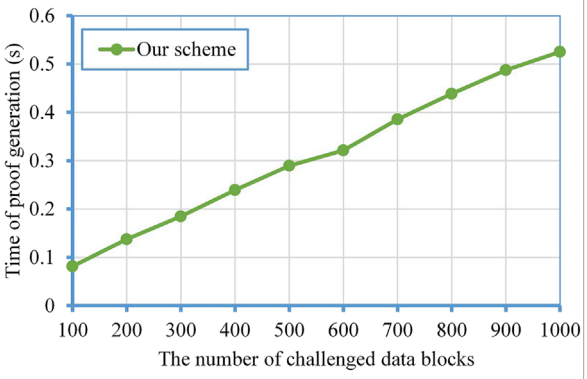



FIGURE 4
The computation overhead of proof generation.

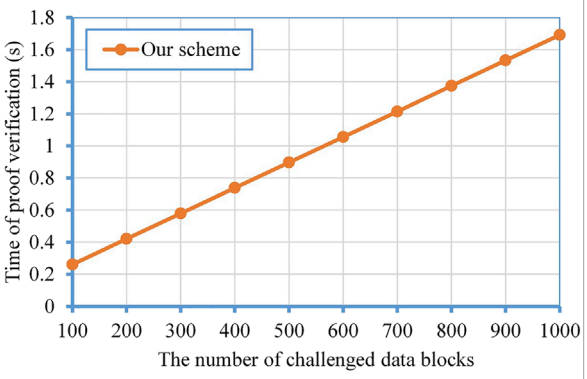



FIGURE 5
The computation overhead of proof verification.

Precision Arithmetic (GMP) Library (GMP-6.2.1) (GMP, 1991) and the Pairing Based Cryptography (PBC) Library (PBC-0.5.14) (Lynn, 2015) to implement the experiments. The experiments are conducted on the Ubuntu 20.04 (4 GB memory) VMware 16.1 pro in a desktop running Windows 10 with Intel(R) Core (TM) i7-9700T @ 2.0 GHZ and 16 GB RAM. In the following experiments, we set the base field size to be 512 bits, the size of an element in $Z_p^*$ to be $|p| = 160$ bits.

### 6.2.1 Authenticator generation

In the proposed scheme, we test the time of authenticator generation for different numbers of data blocks, ranging from 100 to 1,000. The experimental result is represented in **Figure 2**. When the data owner outsources 100 data blocks and calculates the authenticators for these data blocks, the time to generate the authenticators is 0.390975s. When the number of data blocks is 1,000, the time of authenticator generation is 3.951747s. We can find that the time of authenticator generation is related to the number of data blocks.

### 6.2.2 Challenge generation

In the following experiment, we set the number of data blocks to 1,000, and the number of queried data blocks ranges from 100 to 1,000. As shown in **Figure 3**. When the number of challenged data blocks is 100, the time of challenge generation is 0.000064s. And if the number of challenged data blocks is 1,000, the challenge generation time is 0.000063s. Obviously, the overhead of challenge generation is independent of the number of challenged data blocks.

### 6.2.3 Proof generation

In **Figure 4**, we can obtain that the computation overhead of proof generation increases linearly with the number of challenged data blocks. When the number of queried data blocks is 100, the proof generation time is 0.081322s. When

the number of challenged data blocks is 1,000, the time cost is 0.525229s.

### 6.2.4 Proof verification

**Figure 5** shows that there is a proportional relationship between the computation cost of proof verification and the number of challenged data blocks. As the number of challenged data blocks increases from 100 to 100, the time of proof verification increases from 0.261052s to 1.691584s.

## 7 Conclusion

In this paper, we proposed a certificateless public auditing scheme for cloud-based smart grid data, which supports data privacy preserving. Compare with the traditional public auditing schemes based on PKI or IBC, our scheme can avoid the complex certificate management issue and key escrow isssue. In addition, the TPA cannot obtain the original smart grid data during the data integrity auditing phase. We give the security proof of the scheme, and the results show that our scheme is secure. We also evaluate the efficiency of our scheme through a series of experiments.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

CG organized the manuscript and wrote the first draft of the manuscript, WS provided revisions to this paper, MY and YS designed experimental scenarios.

## Funding

This research is supported by National Natural Science Foundation of China (62102211), Shandong Provincial Natural Science Foundation, China (ZR2021QF018), the Open Research Fund from Shandong Key Laboratory of Computer Network (SDKLCN-2020-02).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Al-Riyami, S. S., and Paterson, K. G. (2003). "Certificateless public key cryptography," in *Advances in cryptology - asiacrypt 2003*. Editor C.-S. Laih (Berlin, Heidelberg: Springer Berlin Heidelberg), 452–473.

Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., et al. (2007). "Provable data possession at untrusted stores," in Proceedings of the 14th ACM Conference on Computer and Communications Security, Alexandria, Virginia, USA, 28-31 Oct. 2007, 598–609. doi:10.1145/1315245.1315318

Ateniese, G., Di Pietro, R., Mancini, L. V., and Tsudik, G. (2008). "Scalable and efficient provable data possession," in Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks, New York, NY, USA, September 22, 2008, 1–10.

Bao, F., Deng, R. H., and Zhu, H. (2003). "Variations of diffie-hellman problem," in Proceedings of the International conference on information and communications security (Springer), Berlin, Heidelberg, October 10, 2003, 301–312.

Boneh, D., Lynn, B., and Shacham, H. (2001). "Short signatures from the weil pairing," in Proceedings of the International conference on the theory and application of cryptology and information security (Springer), Berlin, Heidelberg, November 20, 2001, 514–532.

Chen, X., Li, J., Ma, J., Tang, Q., and Lou, W. (2014). New algorithms for secure outsourcing of modular exponentiations. *IEEE Trans. Parallel Distrib. Syst.* 25, 2386–2396. doi:10.1109/tpds.2013.180

Erway, C. C., Küpçü, A., Papamanthou, C., and Tamassia, R. (2015). Dynamic provable data possession. *ACM Trans. Inf. Syst. Secur.* 17, 1–29. doi:10.1145/2699909

Gao, X., Yu, J., Chang, Y., Wang, H., and Fan, J. (2021). Checking only when it is necessary: Enabling integrity auditing based on the keyword with sensitive information privacy for encrypted cloud data. *IEEE Trans. Dependable Secure Comput.*, 1–1. doi:10.1109/TDSC.2021.3106780

GMP (1991). The gnu multiple precision arithmetic library (gmp). Available at: http://gmplib.org. (Accessed September 1, 2022).

Guo, W., Qin, S., Gao, F., Zhang, H., Li, W., Jin, Z., et al. (2020). Dynamic proof of data possession and replication with tree sharing and batch verification in the cloud. *IEEE Trans. Serv. Comput.* 15, 1813–1824. doi:10.1109/TSC.2020.3022812

He, D., Kumar, N., Zeadally, S., and Wang, H. (2018a). Certificateless provable data possession scheme for cloud-based smart grid data management systems. *IEEE Trans. Ind. Inf.* 14, 1232–1241. doi:10.1109/TII.2017.2761806

He, D., Zeadally, S., and Wu, L. (2018b). Certificateless public auditing scheme for cloud-assisted wireless body area networks. *IEEE Syst. J.* 12, 64–73. doi:10.1109/JSYST.2015.2428620

Ji, Y., Shao, B., Chang, J., Xu, M., and Xue, R. (2022). Identity-based remote data checking with a designated verifier. *J. Cloud Comput. (Heidelb).* 11, 7–14. doi:10.1186/s13677-022-00279-5

Juels, A., and Kaliski, B. S. (2007). "Pors: Proofs of retrievability for large files," in Proceedings of the 14th ACM Conference on Computer and Communications Security, New York, NY, USA, October 28, 2007, 584–597. doi:10.1145/1315245.1315317

Li, B., He, Q., Chen, F., Dai, H., Jin, H., Xiang, Y., et al. (2021). Cooperative assurance of cache data integrity for mobile edge computing. *IEEE Trans. Inf. Forensic. Secur.* 16, 4648–4662. doi:10.1109/tifs.2021.3111747

Li, Y., Yu, Y., Yang, B., Min, G., and Wu, H. (2018). Privacy preserving cloud data auditing with efficient key update. *Future Gener. Comput. Syst.* 78, 789–798. doi:10.1016/j.future.2016.09.003

Liu, Y., Yu, J., Fan, J., Vijayakumar, P., and Chang, V. (2021). Achieving privacy-preserving dsse for intelligent iot healthcare system. *IEEE Trans. Ind. Inf.* 18, 2010–2020. doi:10.1109/tii.2021.3100873

Liu, Z., Ren, L., Li, R., Liu, Q., and Zhao, Y. (2022). Id-based sanitizable signature data integrity auditing scheme with privacy-preserving. *Comput. Secur.* 121, 102858. doi:10.1016/j.cose.2022.102858

Lu, Q., Li, S., Zhang, J., and Jiang, R. (2022). Pedr: Exploiting phase error drift range to detect full-model rogue access point attacks. *Comput. Secur.* 114, 102581. doi:10.1016/j.cose.2021.102581

Lynn, B. (2015). The pairing-based cryptographic library. Available at: https://crypto.stanford.edu/pbc/. (Accessed September 1, 2022).

McCurley, K. S. (1990). The discrete logarithm problem. *Proc. Symp. Appl. Math (USA)* 42, 49–74.

McDaniel, P., and McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *IEEE Secur. Priv. Mag.* 7, 75–77. doi:10.1109/MSP.2009.76

Peng, X., Xian, H., Lu, Q., and Lu, X. (2021). Semantics aware adversarial malware examples generation for black-box attacks. *Appl. Soft Comput.* 109, 107506. doi:10.1016/j.asoc.2021.107506

Shacham, H., and Waters, B. (2008). "Compact proofs of retrievability," in *Advances in cryptology - asiacrypt 2008*. Editor J. Pieprzyk (Berlin, Heidelberg: Springer Berlin Heidelberg), 90–107.

Shamir, A. (1985). "Identity-based cryptosystems and signature schemes," in *Advances in cryptology*. Editors G. R. Blakley, and D. Chaum (Berlin, Heidelberg: Springer Berlin Heidelberg), 47–53.

Shao, Y., Tian, C., Han, L., Xian, H., and Yu, J. (2022). Privacy-preserving and verifiable cloud-aided disease diagnosis and prediction with hyperplane decision-based classifier. *IEEE Internet Things J.*, 9, 21648–21661. doi:10.1109/JIOT.2022.3181734

Shen, W., Qin, J., Yu, J., Hao, R., and Hu, J. (2019). Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage. *IEEE Trans. Inf. Forensic. Secur.* 14, 331–346. doi:10.1109/tifs.2018.2850312

Wang, B., Li, B., Li, H., and Li, F. (2013). "Certificateless public auditing for data integrity in the cloud," in Proceedings of the 2013 IEEE Conference on Communications and Network Security (CNS)., National Harbor, MD, USA, October 14–16, 2013 18, 136–144. doi:10.1109/CNS.2013.6682701

Wang, H., Wang, Q., and He, D. (2019). Blockchain-based private provable data possession. *IEEE Trans. Dependable Secure Comput.*, 1–10. doi:10.1109/TDSC.2019.2949809

Wang, H., Wu, Q., Qin, B., and Domingo Ferrer, J. (2014). Identity-based remote data possession checking in public clouds. *IET Inf. Secur.* 8, 114–121. doi:10.1049/iet-ifs.2012.0271

Wang, Y., Wu, Q., Qin, B., Shi, W., Deng, R. H., and Hu, J. (2017a). Identity-based data outsourcing with comprehensive auditing in clouds. *IEEE Trans. Inf. Forensic. Secur.* 12, 940–952. doi:10.1109/tifs.2016.2646913

Wang, Y., Wu, Q., Qin, B., Tang, S., and Susilo, W. (2017b). Online/offline provable data possession. *IEEE Trans. Inf. Forensic. Secur.* 12, 1182–1194. doi:10.1109/TIFS.2017.2656461

Wu, G., Mu, Y., Susilo, W., Guo, F., and Zhang, F. (2019). Privacy-preserving certificateless cloud auditing with multiple users. *Wirel. Pers. Commun.* 106, 1161–1182. doi:10.1007/s11277-019-06208-1

Xu, Y., Sun, S., Cui, J., and Zhong, H. (2020). Intrusion-resilient public cloud auditing scheme with authenticator update. *Inf. Sci.* 512, 616–628. doi:10.1016/j.ins.2019.09.080

Xu, Z., He, D., Vijayakumar, P., Gupta, B., and Shen, J. (2021). Certificateless public auditing scheme with data privacy and dynamics in group user model of cloud-assisted medical wsns. *IEEE J. Biomed. Health Inf.*, 1–1. doi:10.1109/jbhi.2021.3128775

Yang, H., Su, Y., Qin, J., and Wang, H. (2020). Privacy-preserving outsourced inner product computation on encrypted database. *IEEE Trans. Dependable Secure Comput.* 19, 1. doi:10.1109/tdsc.2020.3001345

Yu, J., Ren, K., and Wang, C. (2016). Enabling cloud storage auditing with verifiable outsourcing of key updates. *IEEE Trans. Inf. Forensic. Secur.* 11, 1362–1375. doi:10.1109/tifs.2016.2528500

Yu, J., and Wang, H. (2017). Strong key-exposure resilient auditing for secure cloud storage. *IEEE Trans. Inf. Forensic. Secur.* 12, 1931–1940. doi:10.1109/tifs.2017.2695449

Zhang, H., Gao, P., Yu, J., Lin, J., and Xiong, N. N. (2021a). Machine learning on cloud with blockchain: A secure, verifiable and fair approach to outsource the linear regression. arXiv preprint arXiv:2101.02334.

Zhang, H., Tong, L., Yu, J., and Lin, J. (2021b). Blockchain-aided privacy-preserving outsourcing algorithms of bilinear pairings for internet of things devices. *IEEE Internet Things J.* 8, 15596–15607. doi:10.1109/jiot.2021.3073500

Zhang J, J., Li, Z., Wang, B., Wang, X. A., and Ogiela, U. (2020). Enhanced certificateless auditing protocols for cloud data management and transformative computation. *Inf. Process. Manag.* 57, 102287. doi:10.1016/j.ipm.2020.102287

Zhang, X., Huang, C., Zhang, Y., and Cao, S. (2021c). Enabling verifiable privacy-preserving multi-type data aggregation in smart grids. *IEEE Trans. Dependable Secure Comput.*, 1–1. doi:10.1109/TDSC.2021.3124546

Zhang, X., Zhao, J., Xu, C., Li, H., Wang, H., and Zhang, Y. (2019). Cipppa: Conditional identity privacy-preserving public auditing for cloud-based wbans against malicious auditors. *IEEE Trans. Cloud Comput.* 9, 1362–1375. doi:10.1109/TCC.2019.2927219

Zhang Y, Y., Yu, J., Hao, R., Wang, C., Ren, K., Jia, X., et al. (2020). Towards identification of molecular mechanism in which the overexpression of wheat cytosolic and plastid glutamine synthetases in tobacco enhanced drought tolerance. *Plant Physiol. biochem.* 17, 608–620. doi:10.1016/j.plaphy.2020.04.013

Zhang, Y., Xu, C., Yu, S., Li, H., and Zhang, X. (2015). Sclpv: Secure certificateless public verification for cloud-based cyber-physical-social systems against malicious auditors. *IEEE Trans. Comput. Soc. Syst.* 2, 159–170. doi:10.1109/TCSS.2016.2517205

Zhou, L., Fu, A., Mu, Y., Wang, H., Yu, S., and Sun, Y. (2021). Multicopy provable data possession scheme supporting data dynamics for cloud-based electronic medical record system. *Inf. Sci.* 545, 254–276. doi:10.1016/j.ins.2020.08.031

Zhou, L., Fu, A., Yang, G., Wang, H., and Zhang, Y. (2022). Efficient certificateless multi-copy integrity auditing scheme supporting data dynamics. *IEEE Trans. Dependable Secure Comput.* 19, 1–1132. doi:10.1109/TDSC.2020.3013927

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fenrg.2022.1042882



OPEN ACCESS

EDITED BY
Hanlin Zhang,
Qingdao University, China

REVIEWED BY
Yu Ma,
Chang'an University, China
Youjun Deng,
Tianjin University, China

*CORRESPONDENCE
Qingyu Yang,
yangqingyu@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 13 September 2022
ACCEPTED 24 October 2022
PUBLISHED 12 January 2023

CITATION
Cui F, Lin X, Zhang R and Yang Q (2023), Multi-objective optimal scheduling of charging stations based on deep reinforcement learning.
*Front. Energy Res.* 10:1042882.
doi: 10.3389/fenrg.2022.1042882



# Multi-objective optimal scheduling of charging stations based on deep reinforcement learning

Feifei Cui[1], Xixiang Lin[1], Ruining Zhang[2] and Qingyu Yang[1]*

[1]School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China, [2]School of Artificial Intelligence, Nanjing Agricultural University, Nanjing, China

With the green-oriented transition of energy, electric vehicles (EVs) are being developed rapidly to replace fuel vehicles. In the face of large-scale EV access to the grid, real-time and effective charging management has become a key problem. Considering the charging characteristics of different EVs, we propose a real-time scheduling framework for charging stations with an electric vehicle aggregator (EVA) as the decision-making body. However, with multiple optimization objectives, it is challenging to formulate a real-time strategy to ensure each participant's interests. Moreover, the uncertainty of renewable energy generation and user demand makes it difficult to establish the optimization model. In this paper, we model charging scheduling as a Markov decision process (MDP) based on deep reinforcement learning (DRL) to avoid the afore-mentioned problems. With a continuous action space, the MDP model is solved by the twin delayed deep deterministic policy gradient algorithm (TD3). While ensuring the maximum benefit of the EVA, we also ensure minimal fluctuation in the microgrid exchange power. To verify the effectiveness of the proposed method, we set up two comparative experiments, using the disorder charging method and deep deterministic policy gradient (DDPG) method, respectively. The results show that the strategy obtained by TD3 is optimal, which can reduce power purchase cost by 10.9% and reduce power fluctuations by 69.4%.

KEYWORDS

**electric vehicle, microgrid, multi-objective optimization, charging scheduling, deep reinforcement learning**

## 1 Introduction

In recent years, the global energy structure (Tian et al., 2018; Peng et al., 2021) is transforming into clean energy (Fu et al., 2018; Rajendran et al., 2022), which provides an incentive for the development of EVs. According to research, the exhaust gas emitted by fuel vehicles is one of the main causes of global warming (Purushotham Reddy et al., 2021). Against the background of carbon neutrality (Duan, 2021), some countries have introduced relevant policies promoting EVs

(Yang et al., 2020) to replace traditional fuel vehicles. A smart grid is a new type of modern grid that is stable, efficient, and economical. However, with the large-scale charging demand of EVs, the smart grid faces many challenges (Choi et al., 2017; Brenna et al., 2018), such as increasing exchange power fluctuations and degrading power quality. Therefore, the stable access of EVs to the smart grid is a key issue that must be solved.

EVs have the advantages of flexibility and adjustability due to the power battery. Taking advantage of these features, people can control the charging or discharging of EVs to realize grid stability. At present, various optimization approaches have been proposed to manage the charge or discharge of EVs, that is, convex-optimum methods, programming-based methods (Hu et al., 2013; Ordoudis et al., 2019), and heuristic-based methods (Megantoro et al., 2017; Li et al., 2019). Shi et al. (2017) optimized the day-ahead scheduling of EVs by Lyapunov optimization, which can realize real-time management but relies on precise objective functions. Based on mixed integer programming, Koufakis et al. (2020) minimized EV charging costs and load fluctuations, which also relies on accurate predictions of environmental information. Combining genetic algorithms and dynamic programming algorithms, Ravey et al. (2012) formulated energy management strategies for EVs, but the method shows poor robustness. Therefore, the uncertainty of renewable energy generation and user demand makes it difficult to establish the optimization model based on traditional methods.

There are two ways to deal with the uncertainty in charge and discharge management of EVs. One is to predict uncertain values before optimization through physical models or probability distributions (Kabir et al., 2020). However, this method is only suitable for scenarios with low accuracy requirements, such as day-ahead prediction. Another solution benefits from the development of DRL (Franaois-Lavet et al., 2018). DRL includes two types of methods, model-based and model-free. Model-free DRL (Wan et al., 2019) has attracted great attention in this field due to the following two advantages: 1) neural network as a function approximator (Zhang et al., 2021) can extract more data features based on data history. The data features are input into the policy network to learn the optimal policy. This process does not rely on the predicted values. 2) This method makes decisions according to the current state, so it is suitable for real-time decision scenarios with high precision requirements.

Based on DRL, Chis et al. (2017) and Li et al. (2022) combined neural networks and DRL, which effectively reduced the charging cost. Zhao and Lee (2022) and Su et al. (2020) proposed a dynamic pricing mechanism based on DRL to minimize charging costs. Abdalrahman and Zhuang (2022) improved user satisfaction by maximizing the quality of the charging service. Qian et al. (2022) proposed a pricing mechanism based on multi-agent reinforcement learning and reduced the cost of charging stations. However, the afore-mentioned works in the literature only consider the benefits of the demand side while ignoring the benefits of the supply side. In the electricity market, EVA (Okur et al., 2020; Kong et al., 2021) plays an important role in integrating demand, participating in bidding, and purchasing resources. Qiu et al. (2020) and Tao et al. (2022) studied the efficient pricing problem from the perspective of EVA. Considering the operation cost of microgrids and the purchasing cost of EVs, Zhaoxia et al. (2019) reduced overall costs through day-ahead optimized scheduling. Kandpal and Verma (2021) and Mahmud et al. (2019) considered the microgrid benefits by minimizing grid peaks, but they still used inefficient traditional methods. As an important part of the electricity market, the role of EVA participating in electricity ancillary services (Yang et al., 2017; Yuan et al., 2021) is neglected. Meanwhile, there are few works in the literature (Wang and Cui, 2020; Zhou et al., 2021) that consider the behavioral characteristics of different cars such as taxis, buses, and private cars. Most of them are about path planning or pricing issues.

To sum up, at present, EV charging scheduling based on DRL is used and the following problems still exist: 1) As one of the most effective mechanisms to integrate the market, the benefits and the electric ancillary service functions of EVA are neglected. 2) When applying DRL, it is difficult to learn a strategy that can balance multiple optimization objectives. 3) In the charging model, the characteristics of different types of EVs are not taken into account. Aiming to fill the research gaps, this paper proposes a real-time charging scheduling framework with EVA as the decision-making body. We build a scheduling model with continuous action space, which is solved based on TD3. Our optimization objectives are set to minimize the cost of EVA and minimize the fluctuations of microgrids. The main contributions of this paper are as follows:

- A real-time charging scheduling framework with EVA as a decision-making body is proposed. Considering multiple optimization objectives, the EVA cost and the microgrid exchange power fluctuations are minimized.
- Considering the charging characteristics of taxis and private cars, the charging scheduling process is established as an MDP model. EVA is an agent that interacts with the environment to maximize accumulated rewards.
- The MDP model is solved by TD3. Compared with the disorder charging method and DDPG, the TD3 achieves lower EVA costs and lower microgrid fluctuations.

The remainder of this paper is organized as follows. In **Section 2**, we introduce the system model, constraints, and optimization objectives in detail. In **Section 3**, we introduce the main elements of the design of the MDP model. In **Section 4**, we briefly introduce the TD3 method. In **Section 5**, we perform

groups of experiments and analyze the results. Finally, in Section 6, we conclude this paper.

# 2 System model

## 2.1 System framework

In this paper, the framework of charging station scheduling is shown in Figure 1. First, the power supplier microgrid is composed of DER, an energy storage system (ESS), a micro-power dispatching center (MDC), and a load. Microgrid stability is affected by output power $P_{DER}$ and load power $P_L$. DER is a electricity-generating unit, and its excess energy is stored or sent to the main grid. We assume that $P_{DER}$ includes two types of output sources, photovoltaic power and wind force. EVA mainly sends charging strategy $\pi$ to the charging station according to the physical information and economic information. The physical information includes the microgrid exchange power and the state of charge (SOC) of the charging stations. Economic information is determined by the market side, including resource buyers and market operators. The market operator acts as a middleman, matching tenders and resource buyers. On the market side, EVA minimizes the purchase cost of resources. On the grid side, EVA minimizes power fluctuations.

When the EVs arrive at the station, EVA will obtain their maximum charging power and charging demand. According to the day-ahead exchange power $P_{mg}$ and the electricity price from the market side, EVA formulates the charging strategy $\pi$. According to the strategy, DER supplies energy for EVs in charging stations. Meanwhile, the charging station feeds back the SOC to EVA, which provides a reference for its decision-making.

## 2.2 Constraint model

This paper considers two types of vehicles, taxis and private cars. We divided the 24-h scheduling time into $T$ time steps, that is, $t = \{1, 2, \cdots T\}$. One time step is denoted as $\tau$.

At time $t$, the number of taxis and private cars is $X_t$ and $Y_t$, respectively. When updating the number of cars at the next moment, we need to remove the cars that meet the charging expectations and add new cars. We assumed that the number of newly added taxis and private cars at each moment is $M_t$ and $N_t$, and the proportion with fast-charging demand is $\sigma_1$ and $\sigma_2$. In the scheduling process, the charging or discharging power is limited to the following two conditions.

### 2.2.1 Power limitations of charging station

$$P_t^{min} \le P_t \le P_t^{max}, \tag{1}$$

where $P_t^{min}$ and $P_t^{max}$ are the maximum and minimum charging power of the charging station at time $t$. The positive $P_t$ represents the charging power, while the negative $P_t$ represents

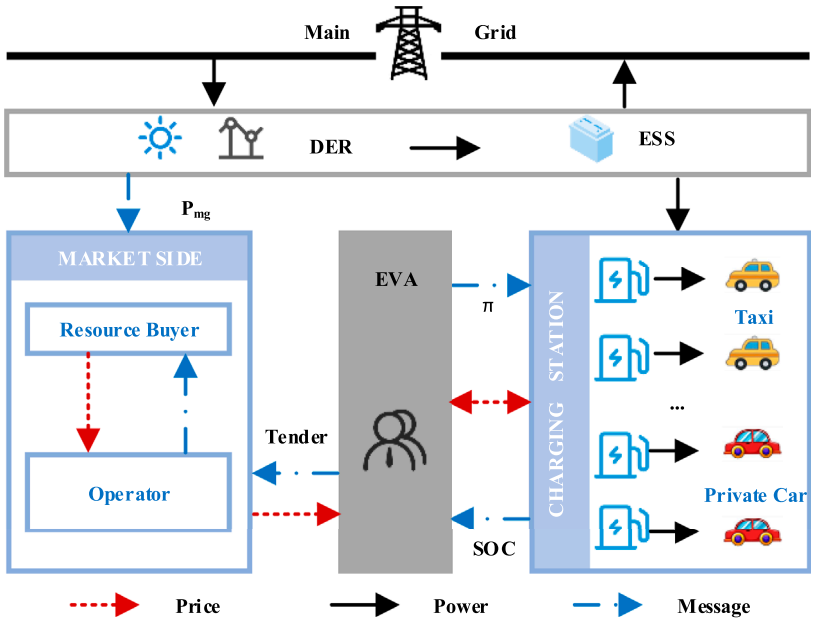



FIGURE 1
Scheduling framework of the EV charging station in a microgrid.

the discharging power of EVs. $P_t^{min}$ and $P_t^{max}$ are limited by two conditions, which can be expressed as

$$P_t^{min} = \max\left\{-P_t^{station}, -\eta_{dis}\frac{S_t^{SOC}c}{\tau}\right\}, \tag{2}$$

$$P_t^{max} = \min\left\{P_t^{station}, \frac{1}{\eta_{dis}}\cdot\frac{\left(X_t + Y_t - S_t^{SOC}\right)c}{\tau}\right\}, \tag{3}$$

where the first term represents the maximum or minimum power of the charging station during time $\tau$. The second term represents the remaining available capacity of the battery. $P_t^{station}$ is the maximum charging power of the charging station. c, $\eta_{dis}$, and $\eta_{ch}$ are the battery capacity, discharging efficiency, and charging efficiency, respectively. $S_t^{SOC}$ is the sum SOC of the EVs at time $t$.

### 2.2.2 State of charge limitations of electric vehicles

$$0 \le S_t^{SOC} \le X_t + Y_t, \tag{4}$$

$$\left|S_t^{SOC} - E_t^{SOC}\right| \le \delta, \tag{5}$$

where $\delta$ is the allowable difference factor between the SOC expected value $E_t^{SOC}$ and the actual value $S_t^{SOC}$. Eq. 4 expresses the total SOC range of EVs. Eq. 5 is the judgment condition for whether EVs reach the expected values.

At time $t$, the maximum power of the charging station is

$$P_t^{station} = P_t^{f,max} + P_t^{s,max}, \tag{6}$$

where $P_t^{f,max}$ and $P_t^{s,max}$ are the maximum power for fast charging and slow charging at time $t$.

$$P_t^{f,max} = P_f \cdot n_t^{fast} \cdot \tau, \tag{7}$$

$$P_t^{s,max} = P_s \cdot n_t^{slow} \cdot \tau, \tag{8}$$

where $P_f$ and $P_s$ are a fast power and slow power of the charging station, which are fixed values. $n_t^{fast}$ and $n_t^{slow}$ are the number of fast-charging vehicles and the slow-charging vehicles at time $t$, respectively.

For the total power $P_t$ allocated to the charging station, the power distributed to each fast-charging and slow-charging vehicle is

$$P_{t,i}^{fast} = \frac{P_t^{f,max}}{P_t^{station}\cdot n_t^{fast}}\cdot P_t, \tag{9}$$

$$P_{t,i}^{slow} = \frac{P_t^{s,max}}{P_t^{station}\cdot n_t^{slow}}\cdot P_t. \tag{10}$$

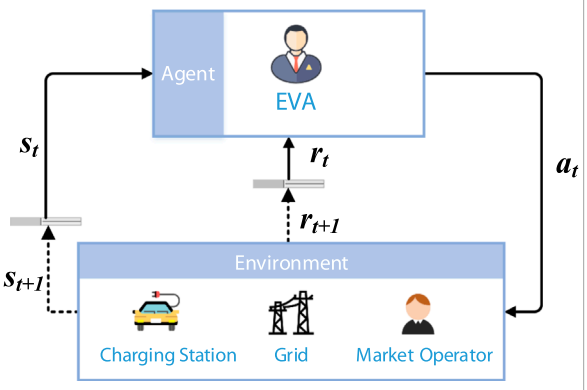



FIGURE 2
RL structure for charging scheduling.

The SOC update satisfies the following equation:

$$S_{t+1,i}^{SOC} = \begin{cases} S_{t,i}^{SOC} + \frac{1}{\eta_{dis}}\cdot\frac{P_{t,i}\cdot\tau}{c}, P_{t,k,i} \le 0 \\ S_{t,i}^{SOC} + \eta_{ch}\cdot\frac{P_{t,i}\cdot\tau}{c}, P_{t,k,i} > 0 \end{cases}. \tag{11}$$

Note that for each time step, we need to remove the cars that have reached the expected values and add the new cars. Therefore, the total value of SOC at the next moment can be expressed as

$$S_{t+1}^{SOC} = \sum S_{t+1,i}^{SOC} + \sum S_{t+1,i}^{SOC,M+N} - \sum S_{t,i}^{ESOC}, \tag{12}$$

where the first item is the sum of the updated SOC of all vehicles at time $t$. The second term is the sum of the SOC of newly added $M_{t+1} + N_{t+1}$ vehicles at time $t+1$. The third term is the sum of the SOC that reaches the expected values at time $t$.

## 2.3 Optimization objective

According to the system model in Section 2.1, we set two optimization objectives, namely, maximizing the benefits of EVA and minimizing power fluctuations.

Assuming that the service cost of EVA is a fixed value, reducing the power purchase cost can maximize EVA's profit. The first optimization objective can be set as follows:

$$\min F_a = \sum_{t=1}^{T}\lambda_t P_t \cdot \tau, \tag{13}$$

where $\lambda_t$ is the time-of-use electricity price at time $t$. $P_t$ is limited by Eq. 1.

We define $F_P$ as the exchange power fluctuation of the microgrid. The second optimization objective can be set as follows:

$$\min F_P = \sum_{t=1}^{T}\left(P_{MG,t} - \frac{1}{T}\sum_{t=1}^{T}\hat{P}_{MG,t}\right), \tag{14}$$

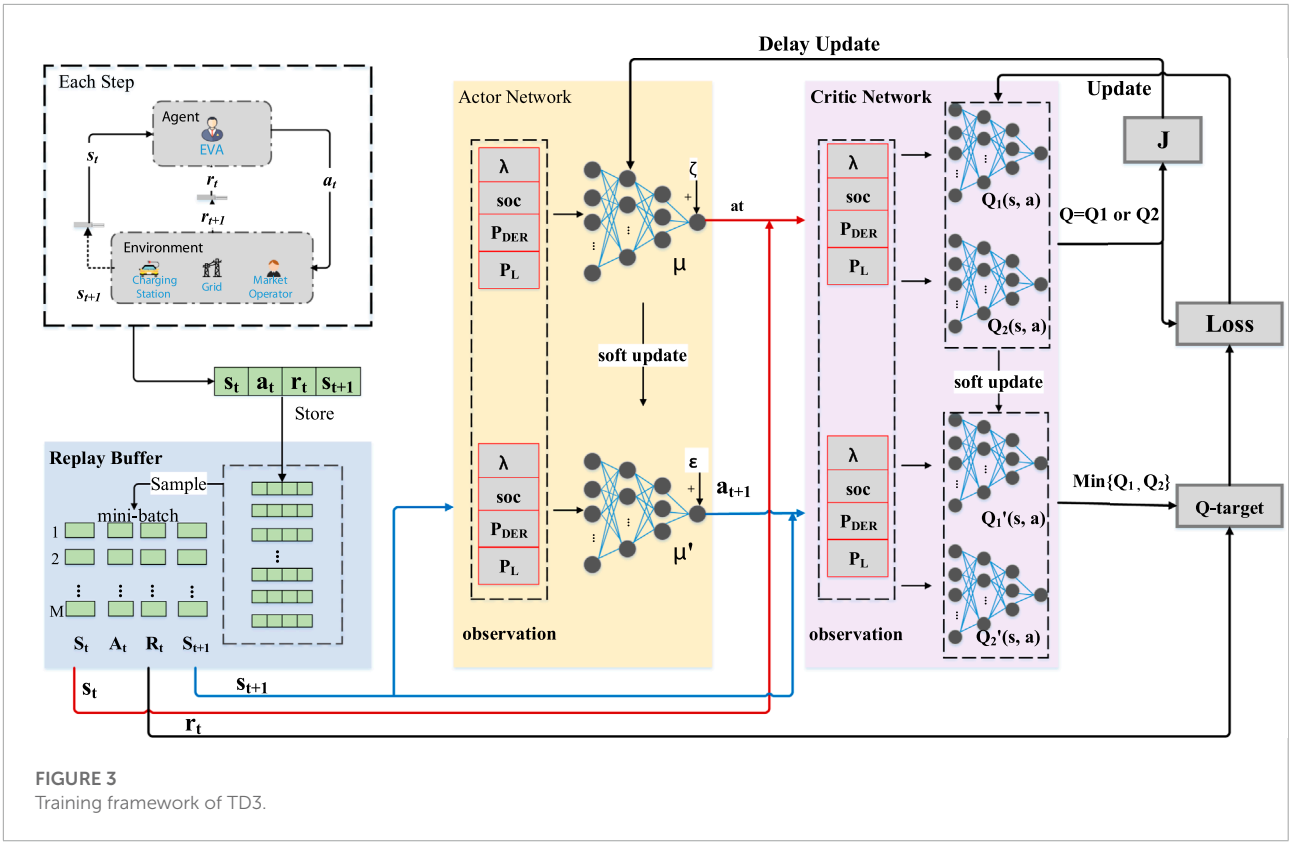



FIGURE 3
Training framework of TD3.

where $P_{MG,t}$ is the microgrid real-time power including EVs at time $t$. $\hat{P}_{MG,t}$ is the day-ahead forecasting value of the microgrid power without EVs at time $t$.

$$P_{MG,t} = P_{EV,t} + P_{L,t} - P_{DER,t}, \tag{15}$$

$$\hat{P}_{MG,t} = \hat{P}_{L,t} - \hat{P}_{DER,t}, \tag{16}$$

where $P_{EV,t}$, $P_{L,t}$, and $P_{DER,t}$ are the real-time values of EVs, other loads, and DER, respectively. $\hat{P}_{L,t}$ and $\hat{P}_{DER,t}$ are the day-head predicted values of other loads and DER, respectively.

# 3 Model design of Markov decision process

## 3.1 Markov decision process

For reinforcement learning (RL), the agent and environment are two main interacting objects, as shown in Figure 2. The agent perceives the state and reward from the environment to learn and make decisions, while the environment updates the state and reward at the next moment based on the current action from the agent. The purpose of this process is to learn a strategy that satisfies the optimization objectives through continuous interactions.

The learning process of RL is usually described by MDP. We set the EVA as an agent and information such as price and power as the environment. At time $t$, the agent interacts with the environment to give a policy $\pi$ and implement action $a_t$ within the action range. The environment reacts to $a_t$ and updates the state $s_{t+1}$. The state transition function $P$ determines the update from $s_t$ to $s_{t+1}$. The environment feedbacks to the agent a reward $r_t = R(s_t, a_t)$ to guide the agent to achieve the optimization objectives. To express this process, we need four elements, state, action, state transition function, and reward function, which are denoted as a tuple $\langle S, A, P, R \rangle$.

## 3.2 Model design

State space $S$ is the set of state values. $S$ is a description of the current situation and should not contain redundant information. Therefore, in this paper, $s_t \in S$ contains four variables, time-of-use electricity price, the sum of charging station SOC, the output power of DER, and the power of other loads, denoted as $s_t = \{\lambda_t, S_t^{SOC}, P_{DER,t}, P_{L,t}\}$.

Action space $A$ is the set of action values. We set the total charging or discharging power of the charging station as the action. Limited by the maximum and minimum charging power,

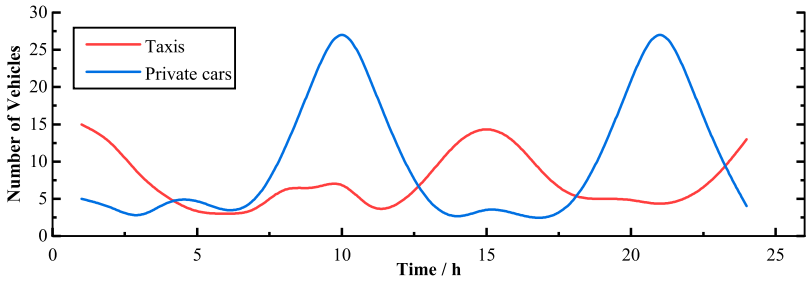



FIGURE 4
Number distributions of taxis and private cars.

TABLE 1 Time-of-use electricity price.

| Time | Electricity price (yuan/kW·h) |
|---|---|
| 0:00–6:00 | 0.385 |
| 6:00–8:00, 11:00–18:00 | 0.555 |
| 8:00–11:00, 18:00–23:00 | 0.725 |

$a_t$ can be expressed as

$$a_t = \begin{cases} P_t^{min} & ,P_t < P_t^{min} \\ P_t^{max} & ,P_t > P_t^{max} \\ P_t & ,others \end{cases}, \tag{17}$$

where $P_t^{min}$ and $P_t^{max}$ are determined by Eqs. 2, 3, respectively.

State transition function $P$ is the rule for state update, denoted as

$$\mathrm{P}: s_t \times a_t \rightarrow s_{t+1}. \tag{18}$$

In Eq. 18, it can be seen that $s_{t+1}$ is determined by the action and state $s_t$. The probability of taking action $a_t$ in state $s_t$ is denoted as $p\left(s_{t+1}|s_t, a_t\right)$.

Reward function $R\left(s_t, a_t\right)$ represents the optimization objectives of the model. In order to maximize the benefits of EVA, the reward function can be designed as

$$r_{t,1} = -\lambda_t a_t \cdot \tau. \tag{19}$$

At time $t$, in order to minimize the exchange power fluctuations of the microgrid, the reward function can be designed as

$$r_{t,2} = -\left|P_{MG,t} - \hat{P}_{MG,t}\right|. \tag{20}$$

At time $t$, in order to encourage the agent to charge and satisfy the needs of users, the reward function can be designed as

$$r_{t,3} = \begin{cases} 1 & \left|S_t^{SOC} - E_t^{SOC}\right| < \delta \\ 0 & \left|S_t^{SOC} - E_t^{SOC}\right| > \delta \end{cases}, \tag{21}$$

where $E_t^{SOC}$ is the sum of the expected SOC values at time $t$.

In order to balance these three rewards, the total reward function can be expressed as

$$r_t = \beta_1 r_{t,1} + \beta_2 r_{t,2} + \beta_3 r_{t,3}, \tag{22}$$

where $\beta_1$, $\beta_2$, and $\beta_3$ are the balance coefficients of three rewards, respectively.

# 4 Proposed approach

TD3 is a type of deterministic strategy gradient algorithm, which is a relatively advanced method. In Section 3, the action is continuously adjustable. Therefore, it is necessary to select a type of RL method with continuous action space. Compared with traditional RL methods, such as Q-learning, TD3 can handle decision problems with continuous action space and continuous state space. The training process has fast convergence speed and good stability. The following is the principle and training process of TD3.

## 4.1 TD3 algorithm

TD3 is an optimization method of DDPG, which is based on the actor-critic framework. Methods based on the actor-critic framework consist of critic networks and actor networks. The purpose of the actor networks is to establish a relational mapping of $s_t$ and $a_t$, while the purpose of the critic networks is to evaluate this mapping relationship and output the value function $Q$. Its mapping relationship can be described as

$$\begin{aligned} &Actor: s_t \rightarrow a_t \\ &Critic: \left[s_t, a_t\right] \rightarrow Q \end{aligned}. \tag{23}$$

DDPG uses the experience replay of Deep Q-learning (Gao and Jin, 2022), and adds two target networks, namely, the target-actor network and the target-critic network. The loss function $L$

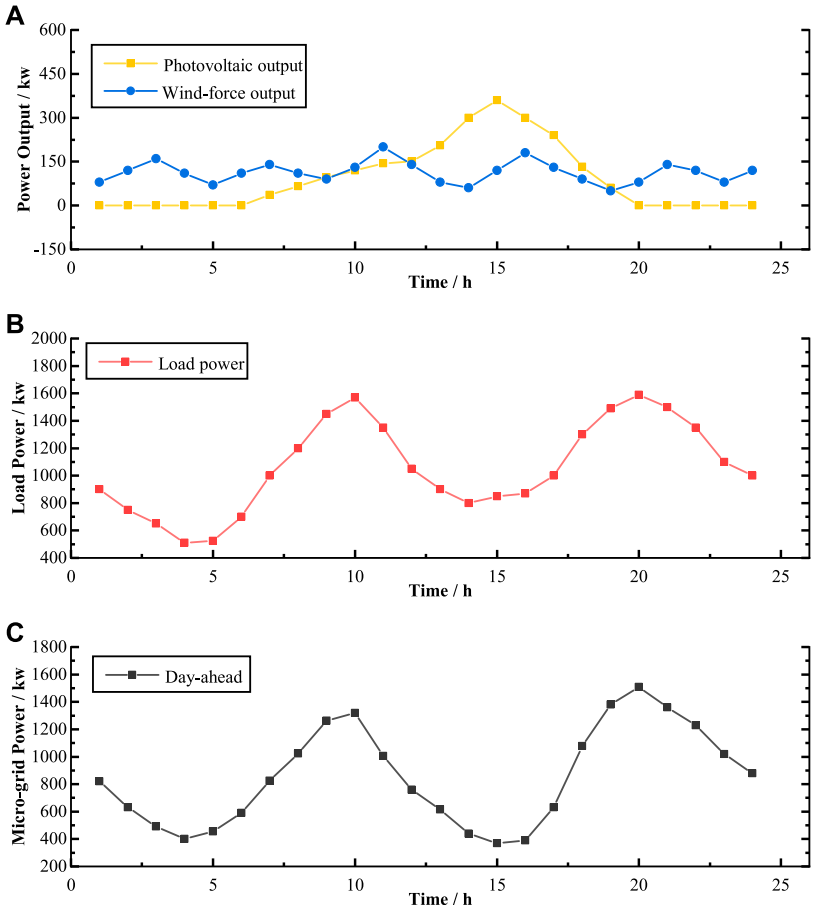



FIGURE 5
Day-ahead power curves for microgrid: **(A)** output power; **(B)** load power; **(C)** microgrid exchange power.

of the critic network is defined as

$$L(\theta_Q) = \frac{1}{M}\sum_{i=1}^{M}\left[Q^{target} - Q(s_t, a_t, \theta_Q)\right]^2, \tag{24}$$

where $\theta_Q$ is the critic network parameter. $M$ is the number of learning samples selected from the experience replay buffer. $Q_t^{target}$ is the value function of the target-critic network, which is calculated as follows:

$$Q_t^{target} = r_t + \gamma Q'\left[s_{t+1}, \mu'(s_{t+1}, \theta_{\mu'}), \theta_{Q'}\right], \tag{25}$$

where $\gamma$ is the discount factor. $\mu'$ and $\theta_{\mu'}$ represent the target-actor network and its parameter, respectively. $Q'$ and $\theta_{Q'}$ represent the target-critic network and its parameter, respectively.

The current state is mapped to action by the function $\mu(s_t, \theta_\mu)$. The actor-network parameter is updated through the gradient back-propagation algorithm. Its loss gradient is

$$\nabla_{\theta_\mu} J \approx \frac{1}{M}\sum_{i=1}^{M}\left[\nabla_a Q(s_t, a, \theta_Q)|_{a=\mu(s_t,\theta_\mu)} \cdot \nabla_{\theta_\mu}\mu(s_t, \theta_\mu)\right], \tag{26}$$

where $\nabla$ is the gradient. $\mu$ and $\theta_\mu$ are the output value and parameters of the actor-network.

The target network parameters $\theta_{Q'}$ and $\theta_{\mu'}$ can be updated by smoothing exponentials

$$\theta_{Q'} = \tau\theta_Q + (1-\tau)\theta_{Q'}, \tag{27}$$

$$\theta_{\mu'} = \tau\theta_\mu + (1-\tau)\theta_{\mu'}, \tag{28}$$

TABLE 2 Training parameters of TD3.

| Parameter | Value |
|---|---|
| Number of training episodes | 60,000 |
| Batch size $M$ | 256 |
| Discount factor $\gamma$ | 0.99 |
| Soft update factor $\tau$ | 0.005 |
| Policy noise $\varepsilon$ | 0.2 |
| Noise clip $d$ | 0.5 |
| Greedy coefficient $\varsigma$ increment | 0.00002 |
| $\varsigma_{max}$ | 0.95 |
| Policy frequency | 2 |

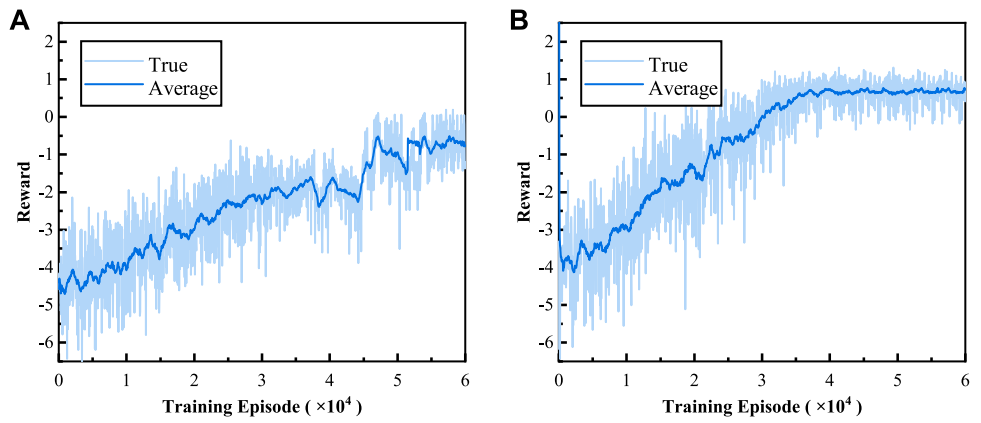



FIGURE 6
Reward curves in the training process: **(A)** DDPG and **(B)** TD3.

where $\tau$ is the update factor.

When updating the Q value of the critic network, it can be expressed as

$$Q = r + \gamma \max Q(s_{t+1}, a_{t+1}). \tag{29}$$

However, if the value function is estimated by maximum, the DDPG method will overestimate, which causes slow convergence and a suboptimal solution. In order to overcome these shortcomings, TD3 has been improved in the following three aspects.

First, in order to overcome the over-estimation problem, TD3 establishes two independent critic networks and two target-critic networks. The target network Q value is updated by the minimum Q value, as follows:

$$Q_t^{target} = r_t + \gamma \min_{k=1,2} Q_k{}' \left[ s_{t+1}, \mu' \left( s_{t+1}, \theta_{\mu'} \right), \theta_{Q_k'} \right], \tag{30}$$

where $Q_k{}'$ and $\theta_{Q_k'}$ $(k = 1,2)$ represent two target-critic networks and their parameters.

The loss function can be improved as

$$L\left(\theta_{Q_k}\right) = \frac{1}{M} \sum_{i=1}^{M} \left[ Q^{target} - Q_k\left(s_t, a_t, \theta_{Q_k}\right) \right]^2, \quad k = 1,2, \tag{31}$$

where $Q_k$ and $\theta_{Q_k}$ $(k = 1,2)$ represent two critic networks and their parameters, respectively.

Second, if we update the actor-network $\mu$ and critic networks $Q_k$ in each loop, the training process will be unstable. Fixing $\mu$ and only training the Q-function can converge faster and get better results. Therefore, TD3 adds the concept of actor-network training frequency, which is less than the update frequency of the critic network. That is also the meaning of "delay."

Finally, to avoid overfitting, TD3 adds a target-actor smoothing step. The action output by the actor-network is improved as

$$\tilde{a}_{t+1} = \mu' \left( s_{t+1}, \theta_{\mu'} \right) + \varepsilon, \varepsilon \sim clip\left(0, \sigma, -d, d\right), d > 0, \tag{32}$$

where $\varepsilon$ is the noise obeying the truncated normal distribution. $\sigma$ is the variance, and $d$ is the truncated amplitude.

## 4.2 Training process

Based on TD3, the training framework for optimal scheduling of charging stations is shown in Figure 3. The detailed training steps of the agent are as follows.

First, as shown by the red line in Figure 3, the agent interacts with the environment to get $s_t$ and uses the actor-network to get $\mu(s_t, \theta_\mu)$. To increase the exploratory effect, we add random noise to the action, which is $\tilde{a}_t = \mu(s_t, \theta_\mu) + (1 - \zeta) \cdot N(0,1)$. In the environment, get the next moment state $s_{t+1}$ and reward $r_t$. The tuple $[s_t, a_t, r_t, s_{t+1}]$ is stored in the experience replay buffer for sampling. When the data in the buffer reach a certain amount, $M$ samples for training are randomly selected.

Then, as shown by the blue line in Figure 3, the target networks get target action by Eq. 32. Through the critic networks, value functions $Q_1(s_t, a_t)$ and $Q_2(s_t, a_t)$ are calculated. Through the target-critic networks, the target value functions $Q_1{}'(s_t, a_t)$ and $Q_2{}'(s_t, a_t)$ are calculated. Then, the target value function $Q_t^{target}$ is obtained by Eq. 30.

Finally, as shown by three gray squares on the right in Figure 3, the critic network parameters $\theta_{Q_1}$ and $\theta_{Q_2}$, are updated, which are determined by Eq. 31. The actor-network parameter $\theta_\mu$ is delay updated, which is determined by Eq. 26. Three target network parameters $\theta_{\mu'}$, $\theta_{Q_1'}$, and $\theta_{Q_2'}$, are soft updated, which are

TABLE 3 Experimental results of charging strategies under different methods.

| Experiment | $\bar{F}_a$ (yuan/kwh) | Price effect (%) | $F_p$(kw) | Fluctuation effect (%) | Charging power (kw) | $\xi$ (%) | Convergence point | Reward |
|---|---|---|---|---|---|---|---|---|
| Disorder | 0.586 | +5.60 | 8,456 | +14.80 | 11,499 | 100 | None | None |
| DDPG | 0.515 | −7.20 | 4,592 | −37.70 | 10,651 | 92.62 | 50,000 | −0.77 |
| TD3 | 0.494 | −10.90 | 2,249 | −69.40 | 8,048 | 70.00 | 36,000 | 0.73 |

determined by Eqs 27, 28. The next training loop is continued until the reward curve converges.

Algorithm 1 outlines the process of learning the optimal policy based on TD3.

**Algorithm 1:** Training procedure of the TD3

Initialize critic networks and actor network with random parameters $\theta_\mu$, $\theta_{Q_1}$, $\theta_{Q_2}$.
Initialize three target networks $\theta_{\mu'} \leftarrow \theta_\mu$, $\theta_{Q_1'} \leftarrow \theta_{Q_1}$, $\theta_{Q_2'} \leftarrow \theta_{Q_2}$.
Initialize replay buffer.
**for** $i = 1$ *to episode* **do**
  Initialize state $s_0$.
  **for** $t = 1$ *to* $T$ **do**
    Select action with exploration noise $\varepsilon$.
    Update next state by (12) and store transition tuple $[s_t, a_t, r_t, s_{t+1}]$ in replay buffer.
    Sample mini-batch of $M$ transitions from buffer.
    Update actor network $\theta_\mu$ by (26) and critic network $\theta_{Q_1}$, $\theta_{Q_2}$ by (31).
    Soft update three target networks by (27) and (28).
  **end**
  Update the greedy coefficient by $\zeta = \zeta + \Delta\zeta$.
**end**

# 5 Experiment

## 5.1 Experimental settings

In this paper, we simulate charging station scheduling in a complex park during a working day. This type of park includes offices, residences, and commercial shops. The microgrid contains wind force, photovoltaic outputs, charging stations, and other household loads. The detailed parameter settings are as follows.

### 5.1.1 Environmental parameters

We consider two types of vehicles, taxis and private cars. Figure 4 shows the number distributions of taxis and private cars. Taxis usually use a two-shift system, with shifts at 6:00 and 18:00, respectively. Therefore, we assumed that the taxi charging peak occurs at 1:00 and 15:00. The private car charging peaks are affected by two groups of people, employees and residents. Therefore, it is assumed that the charging peaks occur at 10:00 and 21:00, respectively. The fast-charging ratios of taxis and private cars are set to $\sigma_1 = 0.7$ and $\sigma_2 = 0.1$, respectively. The initial SOC distributions of taxis and private cars are $N(0.35, 1)$ and $N(0.45, 1)$, respectively, and the expected SOC distributions are $N(0.95, 1)$ and $N(0.90, 1)$, respectively.

The charging station power is divided into four gears: −30, −7, +7, and +30 kw. When each vehicle leaves the charging pile, the deviation of SOC from the expected value is less than the tolerance factor $\delta = 0.05$. The time step is set to $\tau = 1h$ and the time-of-use electricity price is listed in Table 1.

Figure 5 shows the day-ahead power curves that are output forecast curves (Figure 5A) and household load forecast curves (Figure 5B). Assuming that this working day is an ordinary sunny day, the photovoltaic output power reaches the peak about at 12:00, and the wind-force output power fluctuates randomly. According to Eq. 16, we can obtain the forecast curve of microgrid exchange power (Figure 5C), which shows that

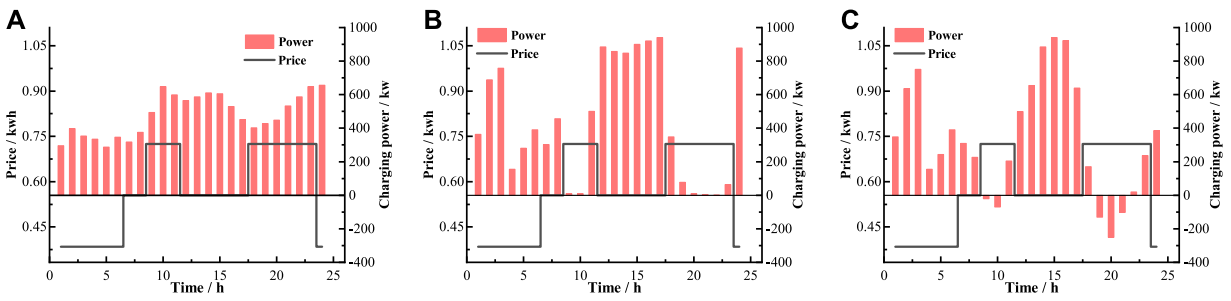



FIGURE 7
Charging or discharging strategies obtained by **(A)**disorder charging, **(B)** DDPG, and **(C)** TD3.

there are evident peaks and valleys in some periods. To rule out the possibility that the method depends on the distribution, we set the uncertain power to be ±10% of the forecast power.

### 5.1.2 Algorithm settings

To evaluate the performance of the TD3, we set up two different methods, the disorder charging method and DDPG. The parameters and rule settings of these three methods are as follows:

- TD3: The detailed parameters of TD3 are shown in **Table 2**. In **Eq. 24**, we set the number $M$ of learning samples from the experience replay buffer to 256. When updating the $Q$ value of two critic networks, the discount factor $\gamma$ set to 0.99 (Zhang et al., 2021) works best. In **Eqs 27**, **28**, the soft update factor $\tau$ is set to 0.005. In **Eq. 32**, the added noise is set to $\varepsilon \sim clip\,(0.2, 1, -0.5, 0.5)$. For better exploration, set th initial value of greedy coefficient $\varsigma$ is set to zero and its increment for each episode to 0.00002. Note that when setting the reward functions, each reward is normalized by its maximum values.
- DDPG: Compared with TD3, DDPG uses one critic network and does not add noise $\varepsilon$ when the actor-target network updates the action. In order to compare the performance of different algorithms, other training parameters of DDPG keep the same as that of TD3.
- Disorder charging: To provide a quantitative reference for the performance of the DRL methods, we set up a disordered charging experiment. When one EV arrives at the station, the charging station starts to continuously supply power with $P_f = 30kw$ or $P_f = 7kw$ until its SOC reaches the expected value.

### 5.1.3 Metrics

To quantitatively evaluate the performance of the three methods, we set the following three metrics.

- Average price: $\bar{F}_a = \sum_{t=1}^{24} \lambda_t a_t / \sum_{t=1}^{24} a_t$ represents the average cost of EVA in one day. The lower $\bar{F}_a$ is, the greater the benefits EVA can get.
- Fluctuation: $F_P = \sum_{t=1}^{24} (P_{MG,t} - \sum_{t=1}^{24} P_{MG,t}/24)$ represents the total fluctuations of microgrid changing power in one day. The lower the $F_P$, the better the charging strategy is in reducing the fluctuations of the microgrid.
- Satisfaction: $\xi = \sum_{t=1}^{24} a_t / A_{max}$, where $A_{\max}$ represents the maximum charging demand in one day. In order to ensure the user's experience, we set the satisfaction coefficient $\xi$. The higher the $\xi$, the better the charging strategy performs in improving the users' experience.

## 5.2 Training results

We evaluate three groups of experiment results and training processes by the three metrics in **Section 5.1.3**. The following are the analysis results.

**Figure 6** shows the training process of two DRL methods. To be more intuitive, we average the rewards every 30 episodes, the results of which are shown as the dark blue curve in **Figure 6**. It can be seen that at the beginning, the reward of both methods is low. When the reward curve tends to stabilize, it means that the agent has explored the optimal strategy. Compared with DDPG (**Figure 6A**), the convergence point of TD3 (**Figure 6B**) is 28% earlier and the reward value is 1.5 (**Table 3**). Therefore, TD3 has notable advantages of better stability, faster convergence, and higher reward in solving the model proposed in this paper.

**Figure 7** shows the results of charging or discharging strategies obtained by three experiments. The gray curves represent the time-of-use price, and the red columns represent the charging or discharging power in each time step. Positive values represent electricity purchased and negative values represent electricity sold by EVA. In **Figure 7A**, it is evident

TABLE 4 Experimental results of charging strategies under different balance coefficient $\beta$.

| $\beta$ | $\bar{F}_a$ (yuan/kw·h) | Price effect (%) | $F_p$(kw) | Fluctuation effect (%) | Charging power (kw) | $\xi$ (%) | Convergence point | Reward |
|---|---|---|---|---|---|---|---|---|
| [0.6, 0.2, 0.2] | −53.320 | 9,707 | 10,968 | +48.92 | 38 | 0.33 | 45,000 | None |
| [0.2, 0.6, 0.2] | 0.494 | −10.90 | 2,249 | −69.40 | 8,048 | 70.00 | 36,000 | 0.494 |
| [0.2, 0.2, 0.6] | 0.590 | +6.3 | 9,844 | +33.65 | 10,801 | 93.92 | 42,000 | 0.73 |

that the disorder charging method does not respond to price. Its charging strategy is to meet the maximum charging demand in each time step. In Figure 7B, the scheduling strategy obtained by DDPG is to charge less during the high-price hours and charge more during other hours. In Figure 7C, the TD3-based strategy has evident discharge behaviors during high-price hours, which indicates a more adequate response to price. From the perspective of the overall benefit, the TD3-based strategy can reduce the electricity price by 10.90% (Table 3), which performs better than DDPG.

Figure 8 shows the results of the microgrid exchanging power. Compared with the day-ahead values, it is evident that the average exchanging power of all experiments increases, which is the result of balancing the charging satisfaction. As shown by the red line in Figure 8, if the charging behavior of EVs is not managed, a large number of EV loads will increase the peak-to-valley difference. In addition, compared with DDPG, the strategy obtained by TD3 can drastically reduce the power fluctuation by 69.40% (Table 3), which is almost twice that of DDPG. Note that in the hours of 3:00–7:00, there is a valley for both RL methods. Combined with Figure 7, during this low-price period, the agent sacrifices certain power fluctuations, which can not only reduce the charging cost but also improve the charging satisfaction.

Table 3 summarizes the results of the three groups of experiments. In terms of charging satisfaction, compared with the other two methods, TD3 sacrifices a certain degree of satisfaction. However, compared to DDPG's results, it is worth sacrificing 24% satisfaction to reduce 51% cost and 84% power fluctuations. Therefore, for the charging model in this paper, the strategy based on TD3 is optimal, which can obtain the real-time scheduling strategy faster and higher overall benefits.

## 5.3 Impact of model parameters

In the training process of TD3, the balance coefficient $\beta = [\beta_1, \beta_2, \beta_3]$ has an important influence on the exploration of optimal strategy. Figure 9 shows the training curves for three different groups of balance coefficients. In order to explore the influence on strategy formulation, experiments are conducted with $\beta_1, \beta_2, \beta_3$ as the dominant factors, respectively. From Figure 9, it can be seen that the reward dominated by $\beta_2$ is the largest. Table 4 summarizes the experiment results, from which we can see that the strategies dominated by $\beta_1$ and $\beta_3$ are two extreme cases. The former reduces costs with maximum discharging, while the latter improves satisfaction with maximum charging. On the whole, when dominated by $\beta_2$, the strategy can guarantee both low cost and low power fluctuations. Therefore, for the training in Section 5.2, the balance coefficient is set to be [0.2, 0.6, 0.2] dominated by $\beta_2$.

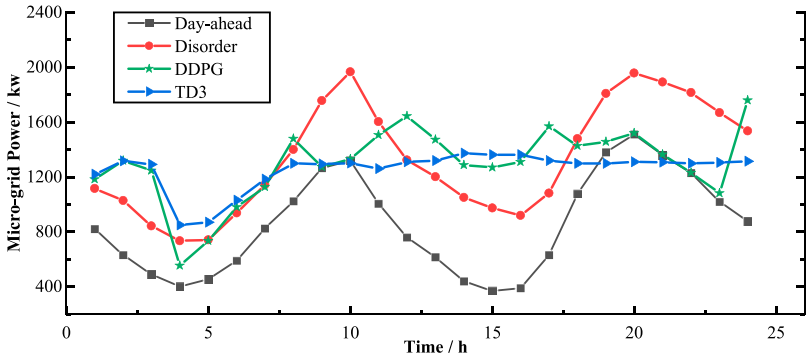



FIGURE 8
Exchange power of the microgrid in one day obtained by day-ahead prediction; disorder charging, DDPG; TD3.

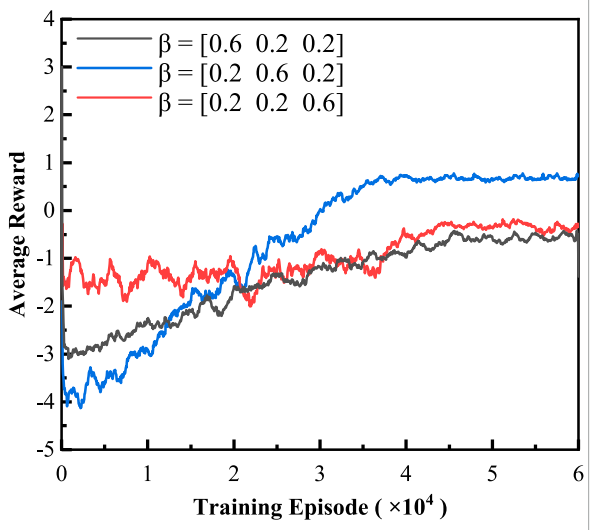



FIGURE 9
Training curves under different balance coefficient $\beta$ based on TD3.

## 6 Conclusion

In the current EV charging management market, balancing the interests of each participant will be an important part in improving the market structure. Therefore, it is necessary to formulate a charging management strategy that considers the interests of each participant. Considering the participation of EVA, microgrids, and users, this paper provides a reference for solving this problem.

Based on DRL, we propose a charging scheduling framework with EVA as the decision-making body. Considering the charging characteristics of electric taxis and private cars, we formulate a charging strategy for charging stations based on TD3. Compared with the disorder charging method and DDPG, TD3 can reduce power purchase costs by 10.9% and reduce power fluctuations by 69.4% on the basis of ensuring certain user satisfaction.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

## Author contributions

FC performed the experiment and wrote the manuscript, XL contributed to the analysis and manuscript preparation, RZ helped perform the analysis with constructive discussions, and QY contributed to the conception of the study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References


Abdalrahman, A., and Zhuang, W. (2022). Dynamic pricing for differentiated pev charging services using deep reinforcement learning. *IEEE Trans. Intell. Transp. Syst.* 23, 1415–1427. doi:10.1109/tits.2020.3025832

Brenna, M., Foiadelli, F., Soccini, A., and Volpi, L. (2018). "Charging strategies for electric vehicles with vehicle to grid implementation for photovoltaic dispatchability," in 2018 International Conference of Electrical and Electronic Technologies for Automotive, Milan, Italy, July 09–11, 2018 (IEEE), 1–6.

Chis, A., Lund'en, J., and Koivunen, V. (2017). Reinforcement learning-based plug-in electric vehicle charging with forecasted price. *IEEE Trans. Veh. Technol.* 66, 3674–3684. doi:10.1109/TVT.2016.2603536

Choi, W., Wu, Y., Han, D., Gorman, J., Palavicino, P. C., Lee, W., et al. (2017). "Reviews on grid-connected inverter, utility-scaled battery energy storage system, and vehicle-to-grid application - challenges and opportunities," in 2017 IEEE Transportation Electrification Conference and Expo (ITEC), Chicago, IL, June 22–24, 2017 (IEEE), 203–210.

Duan, P. (2021). "Research on the transaction and settlement mechanism of yunnan clean energy's participation in the west to east power transmission for the goal of "carbon peak" and "carbon neutral," in 2021 IEEE Sustainable Power and Energy Conference (iSPEC), Nanjing, China, December 23–25, 2021 (IEEE), 1843–1850.

Franaois-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., and Pineau, J. (2018). *An introduction to deep reinforcement learning*. Hanover, United States: Now Foundations and Trends.

Fu, G., Liu, J., Liu, J., and Liu, R. (2018). "Quantitative analysis of the feasibility of realizing the transformation to clean energy for China's energy increment by 2035," in 2018 International Conference on Power System Technology (POWERCON), Guangzhou, China, November 06–08, 2018 (IEEE), 510–515.

Gao, G., and Jin, R. (2022). "An end-to-end flow control method based on dqn," in 2022 International Conference on Big Data, Information and Computer Network, Sanya, China, January 20–22, 2022 (IEEE), 504–507.

Hu, W., Su, C., Chen, Z., and Bak-Jensen, B. (2013). Optimal operation of plug-in electric vehicles in power systems with high wind power penetrations. *IEEE Trans. Sustain. Energy* 4, 577–585. doi:10.1109/tste.2012.2229304

Kabir, M. E., Assi, C., Tushar, M. H. K., and Yan, J. (2020). Optimal scheduling of ev charging at a solar power-based charging station. *IEEE Syst. J.* 14, 4221–4231. doi:10.1109/jsyst.2020.2968270

Kandpal, B., and Verma, A. (2021). Demand peak reduction of smart buildings using feedback-based real-time scheduling of evs. *IEEE Syst. J.* 16, 1–12. doi:10.1109/JSYST.2021.3113977

Kong, W., Luo, F., Jia, Y., Dong, Z. Y., and Liu, J. (2021). Benefits of home energy storage utilization: An Australian case study of demand charge practices in residential sector. *IEEE Trans. Smart Grid* 12, 3086–3096. doi:10.1109/tsg.2021.3054126

Koufakis, A.-M., Rigas, E. S., Bassiliades, N., and Ramchurn, S. D. (2020). Offline and online electric vehicle charging scheduling with v2v energy transfer. *IEEE Trans. Intell. Transp. Syst.* 21, 2128–2138. doi:10.1109/tits.2019.2914087

Li, H., Yang, D., Su, W., Lv, J., and Yu, X. (2019). An overall distribution particle swarm optimization mppt algorithm for photovoltaic system under partial shading. *IEEE Trans. Ind. Electron.* 66, 265–275. doi:10.1109/tie.2018.2829668

Li, S., Hu, W., Cao, D., Dragicevic, T., Huang, Q., Chen, Z., et al. (2022). Electric vehicle charging management based on deep reinforcement learning. *J. Mod. Power Syst. Clean Energy* 10, 719–730. doi:10.35833/mpce.2020.000460

Mahmud, K., Hossain, M. J., and Ravishankar, J. (2019). Peak-load management in commercial systems with electric vehicles. *IEEE Syst. J.* 13, 1872–1882. doi:10.1109/jsyst.2018.2850887

Megantoro, P., Danang Wijaya, F., and Firmansyah, E. (2017). "Analyze and optimization of genetic algorithm implemented on maximum power point tracking technique for pv system," in *2017 international seminar on application for technology of information and communication (iSemantic)* (New Jersey, United States: IEEE), 79–84.

Okur, O., Heijnen, P., and Lukszo, Z. (2020). "Aggregator's business models: Challenges faced by different roles," in *2020 IEEE PES innovative smart grid technologies europe (ISGT-Europe)* (New Jersey, United States: IEEE), 484–488.

Ordoudis, C., Pinson, P., and Morales, J. M. (2019). An integrated market for electricity and natural gas systems with stochastic power producers. *Eur. J. Operational Res.* 272, 642–654. doi:10.1016/j.ejor.2018.06.036

Peng, L., Jinyu, X., Jiawei, W., Zhengxi, C., and Shining, Z. (2021). "Development of global wind and solar resource to cope with global climate change," in 2021 IEEE Sustainable Power and Energy Conference (iSPEC), Nanjing, China, December 23–25, 2021 (IEEE), 986–996.

Purushotham Reddy, M., Aneesh, A., Praneetha, K., and Vijay, S. (2021). "Global warming analysis and prediction using data science," in 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, November 11–13, 2021 (IEEE), 1055–1059.

Qian, T., Shao, C., Li, X., Wang, X., Chen, Z., and Shahidehpour, M. (2022). Multi-agent deep reinforcement learning method for ev charging station game. *IEEE Trans. Power Syst.* 37, 1682–1694. doi:10.1109/tpwrs.2021.3111014

Qiu, D., Ye, Y., Papadaskalopoulos, D., and Strbac, G. (2020). A deep reinforcement learning method for pricing electric vehicles with discrete charging levels. *IEEE Trans. Ind. Appl.* 56, 5901–5912. doi:10.1109/tia.2020.2984614

Rajendran, A., Jayan, P. P., Mohammed Ajlif, A., Daniel, J., Joseph, A., and Surendran, A. (2022). "Energy performance improvement in house boat tourism through clean energy route interfaced with energy efficient power conversion techniques and energy storage," in 2022 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy (PESGRE), Trivandrum, India, January 02–05, 2022 (IEEE), 1–7.

Ravey, A., Roche, R., Blunier, B., and Miraoui, A. (2012). "Combined optimal sizing and energy management of hybrid electric vehicles," in 2012 IEEE Transportation Electrification Conference and Expo (ITEC), Dearborn, MI, June 18–20, 2012 (IEEE), 1–6.

Shi, W., Li, N., Chu, C.-C., and Gadh, R. (2017). Real-time energy management in microgrids. *IEEE Trans. Smart Grid* 8, 228–238. doi:10.1109/tsg.2015.2462294

Su, Z., Lin, T., Xu, Q., Chen, N., Yu, S., and Guo, S. (2020). "An online pricing strategy of ev charging and data caching in highway service stations," in 2020 16th International Conference on Mobility, Sensing and Networking (MSN), Tokyo, Japan, December 17–19, 2020 (IEEE), 81–85.

Tao, Y., Qiu, J., and Lai, S. (2022). Deep reinforcement learning based bidding strategy for evas in local energy market considering information asymmetry. *IEEE Trans. Ind. Inf.* 18, 3831–3842. doi:10.1109/tii.2021.3116275

Tian, Y., Yu, Z., Zhao, N., Zhu, Y., and Xia, R. (2018). "Optimized operation of multiple energy interconnection network based on energy utilization rate and global energy consumption ratio," in 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, October 20–22, 2018 (IEEE), 1–6.

Wan, Z., Li, H., He, H., and Prokhorov, D. (2019). Model-free real-time ev charging scheduling based on deep reinforcement learning. *IEEE Trans. Smart Grid* 10, 5246–5257. doi:10.1109/tsg.2018.2879572

Wang, G., and Cui, D. (2020). "Research on vehicle routing branch pricing algorithm for multi-model electric vehicles based on board testing," in 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, July 28–30, 2020 (IEEE), 954–959.

Yang, J., Fei, F., Xiao, M., Pang, A., Zeng, Z., Lv, L., et al. (2017). "A noval bidding strategy of electric vehicles participation in ancillary service market," in 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, November 11–13, 2017 (IEEE), 306–311.

Yang, Y., Zhang, B., Wang, W., Wang, M., and Peng, X. (2020). "Development pathway and practices for integration of electric vehicles and internet of energy," in 2020 IEEE Sustainable Power and Energy Conference (iSPEC), Chengdu, China, November 23–25, 2020 (IEEE), 2128–2134.

Yuan, H., Lai, X., Wang, Y., and Hu, J. (2021). "Reserve capacity prediction of electric vehicles for ancillary service market participation," in 2021 IEEE 2nd China International Youth Conference on Electrical Engineering (CIYCEE), Chengdu, China, December 15–17, 2021 (IEEE), 1–7.

Zhang, F., Yang, Q., and An, D. (2021). Cddpg: A deep-reinforcement-learning-based approach for electric vehicle charging control. *IEEE Internet Things J.* 8, 3075–3087. doi:10.1109/jiot.2020.3015204

Zhao, Z., and Lee, C. K. M. (2022). Dynamic pricing for ev charging stations: A deep reinforcement learning approach. *IEEE Trans. Transp. Electrific.* 8, 2456–2468. doi:10.1109/tte.2021.3139674

Zhaoxia, X., Hui, L., Tianli, Z., and Huaimin, L. (2019). "Day-ahead optimal scheduling strategy of microgrid with evs charging station," in 2019 IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Xi'an, China, June 03–06, 2019 (IEEE), 774–780.

Zhou, J., Zhao, Y., Li, Y., Kong, J., Yang, C., and Tian, Z. (2021). "A heterogeneous network for electric vehicle charging station communication," in 2021 6th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, September 17–20, 2021 (IEEE), 1204–1208.

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 13 March 2023
DOI 10.3389/fenrg.2023.1094970



OPEN ACCESS

EDITED BY
Huan Xi,
Xi'an Jiaotong University, China

REVIEWED BY
Youcef Belkhier,
Maynooth University, Ireland
Peng Tan,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
You Lv,
you.lv@hotmail.com
Ruijun Qin,
regena@ncepu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 10 November 2022
ACCEPTED 27 February 2023
PUBLISHED 13 March 2023

CITATION
Lv Y, Qin R, Sun H, Guo Z, Fang F and Niu Y (2023), Research on energy storage allocation strategy considering smoothing the fluctuation of renewable energy.
*Front. Energy Res.* 11:1094970.
doi: 10.3389/fenrg.2023.1094970



# Research on energy storage allocation strategy considering smoothing the fluctuation of renewable energy

You Lv[1,2]*, Ruijun Qin[2]*, Hao Sun[2], Ziming Guo[2], Fang Fang[2] and Yuguang Niu[2]

[1]State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, China, [2]School of Control and Computer Engineering, North China Electric Power University, Beijing, China

Energy storage technology can effectively solve the problems caused by large-scale grid connection of renewable energy with volatility and uncertainty. Due to the high cost of the energy storage system, the research on capacity allocation of energy storage system has important theoretical and application value. In this paper, an optimization method for determining the capacity of energy storage system for smoothing the power output of renewable energy is proposed. First, based on the actual data of Ulanqab, the output characteristics of wind power and photovoltaic power generation are studied, and the K-means algorithm is used to select typical days. Then, the energy storage configuration model is built according to the objective function and constraints. Finally, genetic algorithm is used to solve the optimization model, obtain the corresponding parameters, and complete the configuration of energy storage capacity. Based on the results of renewable energy spectrum analysis, the minimum capacity of the energy storage system that meets the constraint of target power output volatility after compensation by the energy storage system can be optimized. The simulation results show that at 1 and 10 min, the flattened volatility is about 2% and 5%, while the actual penetration volatility is about 20% and 30%. The volatility of the optimized model is greatly reduced, which proves the effectiveness of the proposed strategy.

KEYWORDS

**wind-solar power generation system, energy storage system, capacity optimization, K-means algorithm, genetic algorithm**

## 1 Introduction

Due to the rapid use and pollution of fossil fuels, how to achieve carbon neutrality has become an issue of global engagement. Wind and solar energy are renewable resources that currently contain large amounts of energy. The networked operation of wind power and photovoltaic power generation is an effective way to achieve large-scale development of renewable energy. However, the fluctuation and intermittency of wind power have a significant adverse effect on its connection to the grid, which limits the ability of large-scale renewable energy to connect to the grid. Therefore, on the premise of ensuring the safe and stable operation of the power grid, how to absorb

renewable energy on a large scale has become an urgent problem to be solved (Li et al., 2021).

The energy storage system can charge and discharge by itself, which provides an effective means to smooth the output fluctuation of renewable energy and improve the ability to connect to the grid. Ref. (Ramseebaluck, 2019) proposed a hybrid model of rural sites in sub-Saharan Africa (Neeru and Boipuso) using HOMER. Six configurations of locally available energy at each site were considered in the analysis, including solar photovoltaics, wind turbines, batteries for storage, and diesel generators for backup. Ref. (Mishra et al., 2023) proposed model consists of local renewable sources i.e., solar and biomass along with batteries for storage purposes. The major objectives of the study are to optimize the sizing of PV panels and batteries with power sales or purchases from the grid. Ref. (Lawal, 2015) proposed a hydro energy based hybrid system for the rural area, of which the optimization was carried out using HOMER. The Pico hydro system was combined with a solar photovoltaic system for providing the supply. Ref. (Soliman et al., 2021) provides a new energy management control technique for a smart DC-microgrid based on a combined fuzzy logic controller and high-order sliding mode methods. The hybrid energy provider integrated into the DC microgrid is made up of a battery bank, wind energy, photovoltaic energy, and tidal energy sources. Ref. (Alahmadi et al., 2021) proposes an intelligent energy management controller based on combined fuzzy logic and fractional-order proportional-integral-derivative controller methods for a smart DC microgrid. The configurations consist of solar PV, wind turbine, and battery for storage. However, current energy storage systems are expensive, so the research on the capacity configuration of energy storage systems has important theoretical and applied value (Kabeyi and Olanrewaju, 2022).

There has been much research on the capacity allocation strategy of energy storage with renewable energy at home and abroad. Most of the current studies focus on the capacity configuration of energy storage systems from an economic perspective. Ref. (Feng, 2019) evaluates the economy of energy storage projects based on the actual operation data of photovoltaic power stations. A whole life cycle cost model for energy storage was developed in Ref. (Meiqin et al., 2021). A risk management approach is considered in Ref. (Sadeghian et al., 2020), which leads to a study of the optimal capacity of the energy storage system in terms of cost and reliability. In Ref. (Ghaffari et al., 2022), in addition to power losses and energy storage system cost, flicker emission and voltage deviation are also minimized in the objective function. Ref. (Zhang et al., 2022) considers the energy storage configuration scheme of comprehensive demand response according to the real-time market tariff scheme. Ref. (Sun et al., 2023) aims to achieve the lowest operating cost of the distribution network and construct the allocation model by considering the generalized demand-side resources. In Ref. (Paliwal, 2021), the energy storage configuration model is established based on butterfly particle swarm optimization algorithm. The goal is to minimize the economic parameter called the equilibrium cost of energy. The expected unserved energy is used as a reliability constraint. Such a configuration would achieve maximum cost savings, but it does not consider the risks of large-scale access to renewable energy, which is volatile and intermittent.

There are also many ways to consider smoothing the volatility of renewable energy for the current configuration of energy storage capacity, mainly by decomposing renewable energy output by frequency and predicting renewable energy output. In Ref. (Makarov et al., 2012), the sliding average method is used to determine the grid-connected power of wind power and extract the hybrid energy storage output, and the low-frequency and high-frequency fluctuation components of the battery and supercapacitor are reasonably distributed by wavelet packet decomposition. In Ref. (Shen et al., 2023), the wind power output frequency is decomposed to build the configuration model of energy storage capacity by combining the decomposition methods of integrated empirical modal decomposition and empirical modal decomposition. In Ref. (Lamsal et al., 2018), the discrete Kalman filter method is used to predict the renewable energy output, so as to evaluate the capacity of the required energy storage system. In Ref. (Wang et al., 2021), the high-dimensional prediction error of multiple wind farms is considered based on the copula theory, which leads to the capacity allocation of the energy storage system. In Ref. (Zhang et al., 2016), Monte Carlo simulation was used to model the uncertainty of wind power output and system load, so as to optimize and determine the location and capacity of BESS while ensuring the level of wind power utilization. Ref. (Ahmad et al., 2020) discretized the continuous joint power distribution of wind farms and combined it with multi-objective hybrid particle swarm optimization and non-dominated sorting genetic algorithm to optimize the optimal position and capacity of the energy storage system. Ref. (Jamroen et al., 2019) presents a simulation analysis of the PV power smoothing method based on hull enhanced linear exponential smoothing technique using an energy storage system. However, the time-frequency analysis methods, such as low-pass filtering, EMD decomposition, discrete Fourier transform and so on, commonly used in the above research still have shortcomings in maintaining the security, reliability and economy of the power grid, and there are mode aliasing and other conditions, which need to be further improved.

Specifically, the contribution of this paper is to establish an energy storage allocation strategy to limit the volatility of renewable energy, which can effectively improve the economy and sustainable performance of the system. Energy storage technology can effectively solve the problems caused by large-scale grid connection of renewable energy with volatility and uncertainty. Due to the high cost of the energy storage system, the research on capacity allocation of energy storage system has important theoretical and application value. In this paper, an optimization method for determining the capacity of energy storage system for smoothing the power output of renewable energy is proposed. Based on the results of renewable energy spectrum analysis, the minimum capacity of the energy storage system that meets the constraint of target power output volatility after compensation by the energy storage system can be optimized. The remainder of the paper is organized as follows. Models for the wind-solar-storage combined system and renewable energy power generation characteristics are discussed in Section 2. The proposed optimization problem

formulation for selecting typical days and sizing energy storage capacity allocation is given in Section 3. Case studies are presented in Section 4, and concluding remarks are discussed in Section 5.

# 2 System and characteristic analysis

## 2.1 Wind-solar-storage combined system model

When the energy storage device realizes the stabilization effect of the new energy output, it can operate with a single unit or coordinate control with the entire field group. The change of position can cause the stabilization effect or the configuration power and configuration cost. However, due to the energy storage system, compared with a single wind turbine or a single wind farm, the total capacity required for coordinating with the farm group is smaller, the effect is more obvious and the overall maintenance and control of the system are convenient (Dirin et al., 2023). This simulation combines the energy storage system with wind power and photovoltaic field groups are connected to the same bus, as shown in Figure 1.

This paper is based on the establishment of a model in a certain area in Ulanqab City, in which the total installed capacity of wind farms is 149 MW and of photovoltaic field groups is 100 MW.

## 2.2 Wind power output characteristics

Ulanqab, a prefectural-level city in Inner Mongolia Autonomous Region, is located in the north of China, in the middle of Inner Mongolia Autonomous Region, with a total area of about 54,500 square kilometers. Its climate has four distinct seasons, cold winter with less snow, dry and windy spring, cool summer, and cool frosty autumn.

The annual distribution curve of daily power generation based on the 2020 wind farms in the Ulanqab area is shown in Figure 2, and the typical sunrise power curve in each season is shown in Figure 3. According to Figure 1, wind power output in a certain area of Ulanqab City is fluctuating, uncertain, and intermittent. The daily minimum total power generation is 4985 MW and the maximum total power generation is 84,230 MW, which is a difference of nearly 20 times. Its daily total power generation is also indescribable.

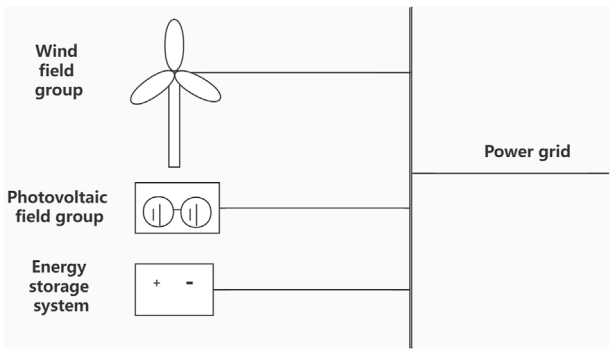



FIGURE 1
Diagram of wind-solar-storage system structure.

According to Figure 3, the typical daily wind power output varies from season to season. There are mainly two periods of high wind power output in spring, from 9:00 am to 12 noon and 3:00 pm to evening. The period of small wind power output is mainly from night to early morning. In summer, the wind power output period is from the evening, and the output during the rest period is close and small. During the autumn and winter seasons, the wind power output is relatively large during the daytime, and the trend of change is also roughly the same. In comparison, the power generation in spring is the largest among the four seasons, summer and autumn are roughly similar, and winter is the season with the least power generation.

## 2.3 Photovoltaic output characteristics

According to the survey, the annual average total solar radiation of Ulanqab is 5,500–6200 MJ/m$^2$, the accumulated temperature greater than or equal to 10°C is 2,228–3,033°C, and the annual average sunshine hours are 2,775–3,080 h.

The annual distribution curve of daily power generation based on the PV field group in the Ulanqab area in 2020 is shown in Figure 4, and the typical sunrise force curve in each season is shown in Figure 5. According to Figure 4, PV output in a certain area of Ulanqab city also has volatility and uncertainty. The minimum total daily power generation is zero, and the maximum total daily power generation is 6477 MW, with a large difference. However, compared with wind power output, photovoltaic power generation is smaller but more stable, with daily power generation fluctuating between 4,000 MW and 7,000 MW except in rainy or extreme weather conditions.

According to Figure 5, the typical daily PV output of the four seasons is concentrated in the period from sunrise in the morning to sunset, and the maximum power is generated at noon. Because high temperatures will lead to component power loss, although summer has the longest illumination time, the most abundant illumination, and the largest intensity, the instantaneous power and generation are far less than spring and autumn, and even less than winter.

# 3 Mathematical model

## 3.1 Typical day selection

Different eras have different methods for how to classify and select typical. In the past, people often classified input data according to the actual meaning of parameters or life and production experience. Today, with the continuous increase of data, when the correlation between input and output is difficult to directly see, clustering technology emerges as the times require.

Clustering analysis technology belongs to the unsupervised learning in artificial intelligence technology, that is, the data of unlabeled samples are automatically grouped to discover their spatial distribution characteristics, laws, and other natural structures. Clustering refers to

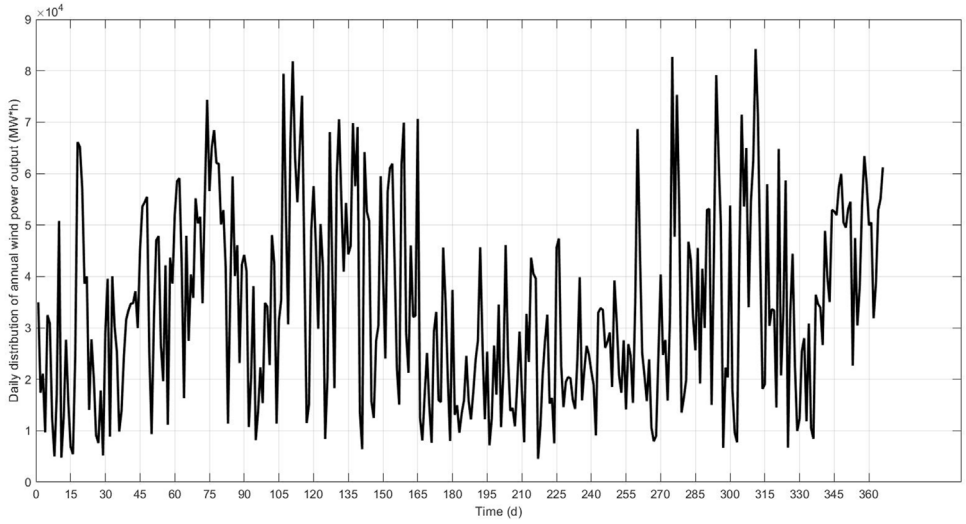



FIGURE 2
Daily distribution of wind power along a year.

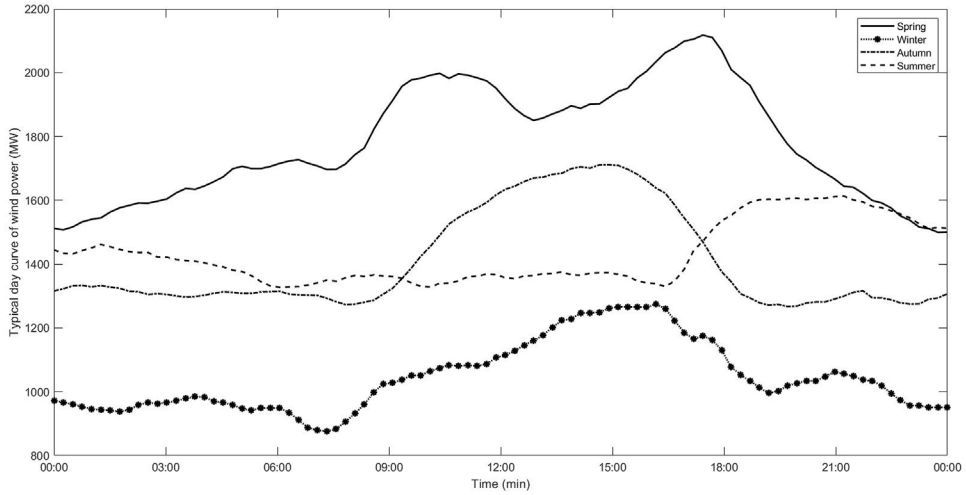



FIGURE 3
The typical curve for each season of wind power.

dividing the total sample set into different sets, which can be divided into different clusters (classes) according to the distance (Euclidean distance, Manhattan distance, etc.) or similarity between the data, and each element in each cluster. There are obvious similar characteristics between each other, and elements in different categories have obvious different characteristics from each other.

At present, clustering has been widely used in the field of data classification and processing. The research of clustering algorithms mainly includes three aspects: technology, data, and derivative problems. Technical research refers to how to use algorithms to divide data. Data research refers to the research on the differences between measures and rules due to the differences in data types. The research on derived problems mainly includes whether the data can be clustered, how to select cluster features, how to visualize, and how to perform cluster verification.

In this typical day selection, the output of wind turbines and the output of photovoltaic units in each season are divided by clustering algorithms. Among the clustering algorithms, the K-Means algorithm has the advantages of being relatively simple and with high efficiency for large-scale data sets (Mingoti and Lima, 2006; Zhao et al., 2022). The K-Means algorithm is a hard clustering algorithm, and the number of values is only 0 or 1. The algorithm mainly uses the error sum of squares criterion function as the core rule. First, set K (the number of clusters) arbitrarily from the N data feature vector sets as the initial center, and

10.3389/fenrg.2023.1094970

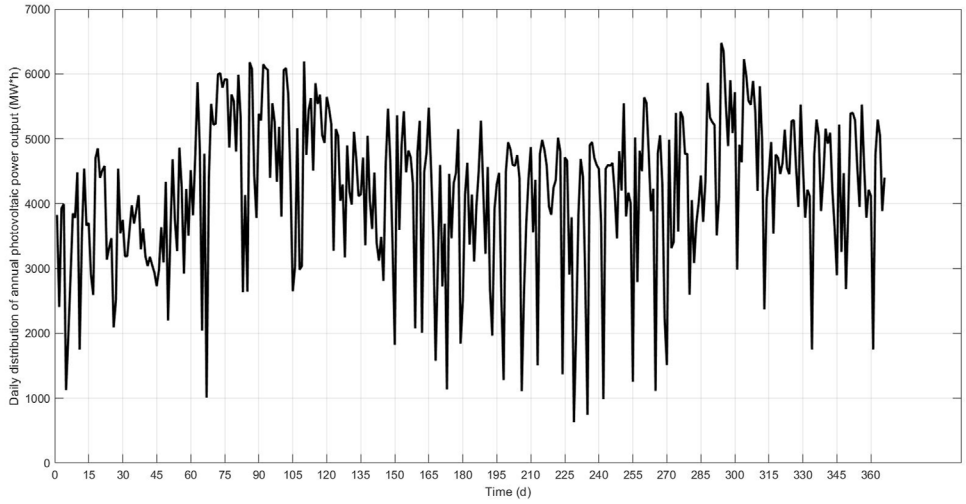



FIGURE 4
Daily distribution of photovoltaic power along a year.

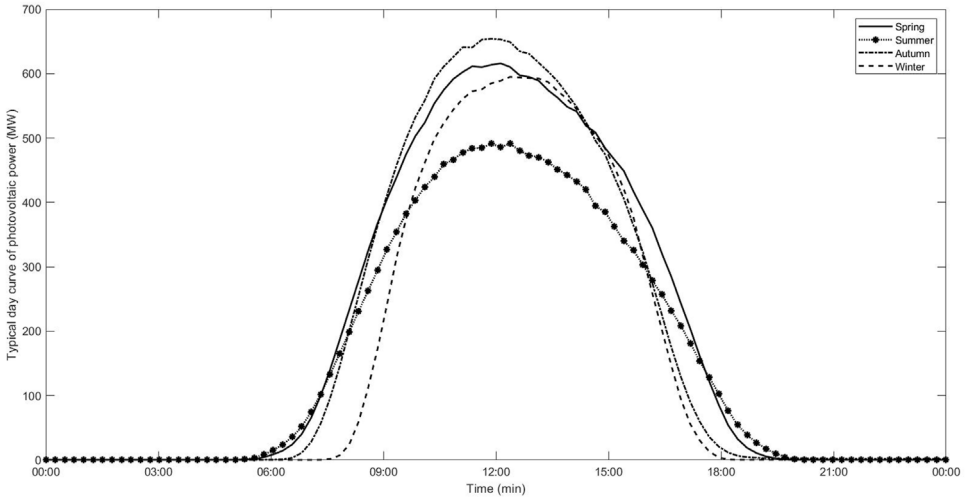



FIGURE 5
Typical day curve of photovoltaic power.

then calculate the distance between the center and the data according to the rule and use this as a basis to re-classify and update the center, and iterate repeatedly until the evaluation index constant. Its flow chart is shown in Figure 6.

The specific implementation steps are as follows:

(1) Initialization.

K points $C_i$ are arbitrarily set from N data ($i = 1, 2, 3, ..., k$)as the initial center.

(2) Calculation and judgment.

Calculate the interval from each data feature point to the cluster center. According to the rule, determine which cluster each point belongs to. When the data point is included in the ith class, the weight $w_{ji} = 1$, and is 0 when it does not belong.

$$\boldsymbol{d_{ij}} = \left\|\boldsymbol{X_j}-\boldsymbol{C_i}\right\|, \boldsymbol{i} = \boldsymbol{1}, \ldots \boldsymbol{k} \tag{1}$$

$$\boldsymbol{X_j} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{1} \end{bmatrix} \tag{2}$$

$$\boldsymbol{W} = \boldsymbol{X_j} \cdot \boldsymbol{C_i} \tag{3}$$

$$\boldsymbol{w_{ji}} = \begin{cases} \boldsymbol{1}, & \left\|\boldsymbol{X_j}-\boldsymbol{C_i}\right\| \leq \left\|\boldsymbol{X_j}-\boldsymbol{C_m}\right\|, \forall \boldsymbol{m} \neq \boldsymbol{j} \\ \boldsymbol{0}, & \boldsymbol{others} \end{cases} \tag{4}$$

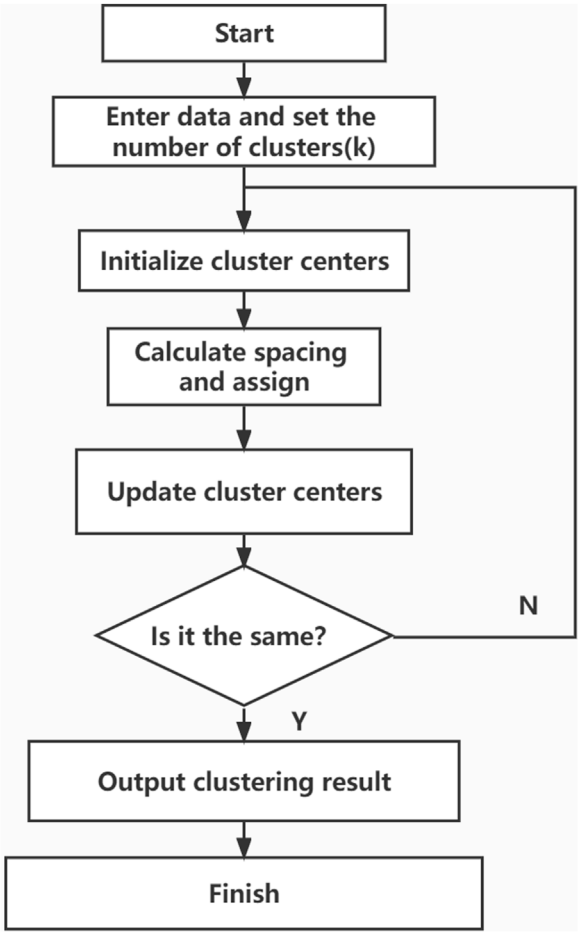



FIGURE 6
Flowchart of K-means algorithm.

where, $X_j$ is the data point, $C_i$ is the cluster point, and $w_{ji}$ is the weight value. And also can satisfy the following formulas:

$$\sum_{i=1}^{K} w_{ji} = 1, \forall j = 1, \ldots, N; \sum_{i=1}^{K}\sum_{j=1}^{N} w_{ji} = N \tag{5}$$

(3) Update cluster midpoints.

$$C_i = \frac{\sum_{j=1}^{n} w_{ji} X_j}{\sum_{j=1}^{n} w_{ji}} \tag{6}$$

(4) Calculate the evaluation index J.

If J remains unchanged, the clustering results have stably converged to this category, and the iteration can be ended; if the evaluation index has a significant change, continue the iteration and update the cluster center point.

$$J = \sum_{i=1}^{K} J_i = \sum_{i=1}^{K}\sum_{j=1}^{N} w_{ji} \left\| X_j - C_i \right\|^2 \tag{7}$$

Therefore, the relevant data of Ulanqab City can be selected for typical days and the relevant output curves of the typical days described in Section 2 can be obtained.

## 3.2 Objective function

As described in Section 2, the power system structure connected to the energy storage system is a wind farm cluster, photovoltaic field cluster, and energy storage system are connected to the same bus and then sent to the power grid.

$$P_{ess,i} = P_{grid,i} - P_{ne,i} \tag{8}$$

$$P_{ne,i} = \sum_{1}^{n_1} P_{w,i} + \sum_{1}^{n_2} P_{pv,i} \tag{9}$$

where, $i$ is the sampling time, $P_{ess}$ is the energy storage system output, $P_{grid}$ is the target grid-connection output after leveling, $P_{ne}$ is the new energy system output, $P_w$ is the wind turbine output, $P_{pv}$ is the photovoltaic array output, $n_1$ and $n_2$ are the number of wind farms and photovoltaic farms, respectively.

According to research, the fluctuation coefficient can be defined as the ratio between the change value of grid-connected power and the change value of new energy output power at a certain moment, namely:

$$N_i = \frac{K_{grid,i}}{K_{ne,i}} \tag{10}$$

$$K_{grid,i} = \frac{P_{grid,i+\Delta i} - P_{grid,i}}{\Delta i} \tag{11}$$

$$K_{ne,i} = \frac{P_{ne,i+\Delta i} - P_{ne,i}}{\Delta i} \tag{12}$$

where, $N_i$ is the fluctuation coefficient, which ranges from 0 to 1. This variable describes the relationship between grid-connected power fluctuation and new energy output power fluctuation. $K_{grid}$ is the change of grid-connected power per unit moment. $K_{ne}$ is the change of new energy power generation.

**TABLE 1 Grid-connected power fluctuation index of China.**

| The installed capacity of new energy power stations (MW) | Maximum volatility of 1 min (MW) | Maximum volatility of 10 min (MW) |
|---|---|---|
| <30 | 3 | 10 |
| 30–150 | One-tenth of installed capacity | One-third of installed capacity |
| >150 | 15 | 50 |

In this paper, the genetic algorithm is adopted to select the optimal fluctuation coefficient corresponding to each moment, to achieve the goal of reducing the rated capacity of the required energy storage equipment under the condition of meeting the national new energy grid power fluctuation index, to reduce the energy storage cost and improve the economic benefits. Therefore, the main objective function of this model is to obtain the minimum variance between grid-connected output and the actual output of new energy, namely:

$$f_1 = min\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[P_{grid,i} - P_{ne,i}\right]^2} \tag{13}$$

where, $f_1$ is the value of the objective function, $n$ is the number of sampling points in the calculation time scale, and the value of this model is 1,440.

In addition, it is necessary to ensure the sustainability of the energy storage system. Therefore, a secondary objective function is established to make the daily initial electric quantity equal to the daily final electric quantity as much as possible, namely:

$$f_2 = min\left|\sum_{i=1}^{n}\left(P_{grid,i} - P_{ne,i}\right)\right| \tag{14}$$

where, $f_2$ is the value of the objective function, which should be as close to zero as possible.

## 3.3 Constraints

New energy power generation is different from conventional energy power generation due to the resource characteristics of scenery. As described in Section 2, wind power has strong randomness and obvious intermittence, and the electric energy incorporated into the power grid has a large fluctuation range and irregular fluctuation frequency, which will lead to voltage and frequency fluctuation of the power grid. A large proportion of new energy connected to the grid will have a certain degree of negative impact on the safe and stable operation of the grid. Based on this issue, China has clarified the constraints on the fluctuation index of grid-connected power of generation, as shown in Table 1.

According to the index constraints on volatility in China, this study puts forward the following two limiting conditions:

(1) Volatility per minute shall not exceed 2% of system installed capacity.
(2) Volatility should not exceed 15% of installed capacity every 10 minutes.

$$\begin{cases} \left|P_{grid}(i+1) - P_{grid}(i)\right| \le w_1 \cdot P_{NE} \\ \left|\max(P) - min(P)\right| \le w_2 \cdot P_{NE} \\ P = P_{grid}(i \sim i+9) \end{cases} \tag{15}$$

where, $w_1$ and $w_2$ is the maximum value of fluctuation rate every 1 min and every 10 min respectively, $P_{NE}$ is the rated installed capacity of new energy, and $P$ is the grid-connected power array with a unit of one ten-step length.

# 4 Case studies

## 4.1 Solution process

Genetic algorithms simulate the phenomena of replication, natural selection and crossover, and variation in heredity. It starts from each original population and creates a group of individuals that are more suitable for the environment through random selection, crossover, and mutation operation, so that the population develops into a better and better area in the search room to obtain high-quality solutions to the problem. The process is shown in Figure 7.

## 4.2 Results and discussion

The new energy power generation system in a certain area of Ulanqab is selected as the analysis object. The installed capacity of the wind farm cluster and photovoltaic field cluster of this system are 149MW and 100 MW respectively. The scheduling cycle is 96 periods a day, 15 min each period. Based on the form of generating burrs from the original data, the model was expanded to 1,440 sections, and the model was

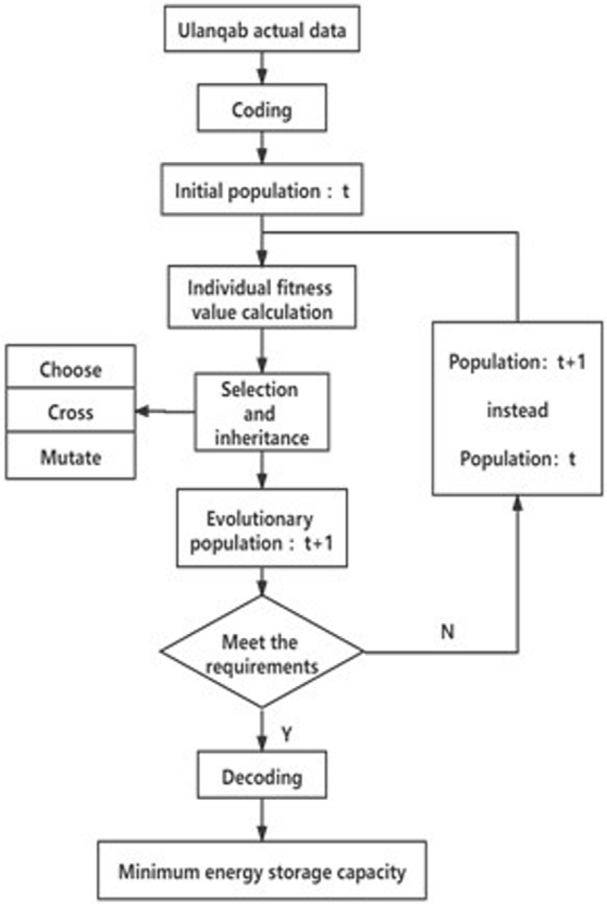



FIGURE 7
Genetic algorithm flow chart.

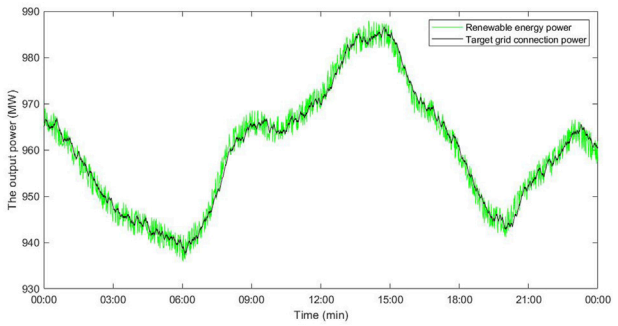



FIGURE 8
Renewable energy grid-connected target output curve.

According to Figure 9, after suppression, the volatility is about 2% and 15% at 1 min and 10 min respectively, while the actual network entry volatility is about 20% and 30%. After optimization of this model, the volatility is greatly reduced, so this model has certain effectiveness. Charge and discharge power of the energy storage system is shown in Figure 10.

The genetic algorithm was used to search iteratively for optimization, and the convergence diagram of the optimal solution of the objective function was shown in Figure 11. The optimal solution was found by traversing 51 times.

As shown in Table 2, the energy storage system solved by the model has the following configuration parameters. The results show that the rated capacity of the required energy storage system is 6.6706 MW*h. The required energy storage system has a rated power of 5.084 MW.

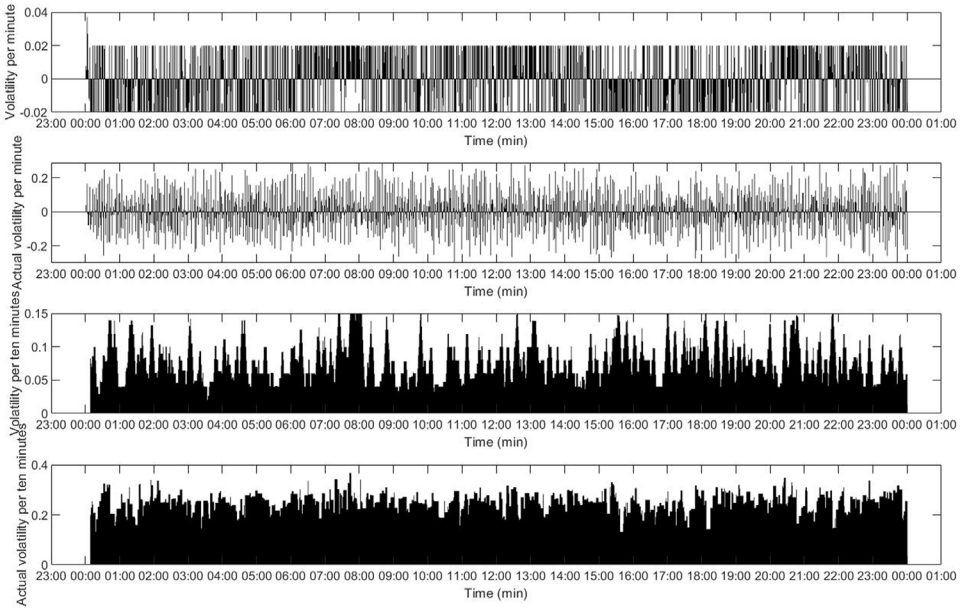



FIGURE 9
Volatility comparison chart.

solved according to Figure 6. The population size is set to 30, the number of iterations is set to 200, the code is binary code, the length is 10, the crossover probability is 60%, and the mutation probability is 10%. The solution results are given as follows.

As shown in Figure 8, curves in different colors represent the actual output of new energy and the target grid-connected output respectively. The grid-connection curve after leveling is obviously smoother with smaller fluctuation. In order to better reflect the leveling effect of this strategy, the optimized grid-connection fluctuation rate and the actual new energy output fluctuation rate in 1 min and 10 min are respectively compared, as shown in Figure 9.

Based on the results of renewable energy spectrum analysis, the minimum capacity of the energy storage system that meets the constraint of target power output volatility after compensation by the energy storage system can be optimized. The volatility of the optimized model is greatly reduced, which proves the effectiveness of the proposed strategy. Although this paper has made some progress in the research of energy storage configuration strategy, control strategy is also an important part of the operation of energy storage system. Only by combining the two can energy storage system achieve optimal utilization. The operation mode of energy storage system will be further studied in the future.

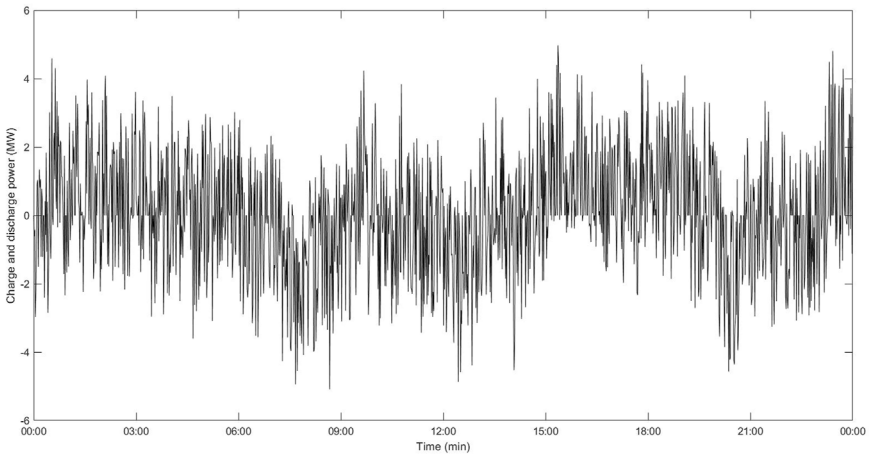



FIGURE 10
Charging and discharging power curve.

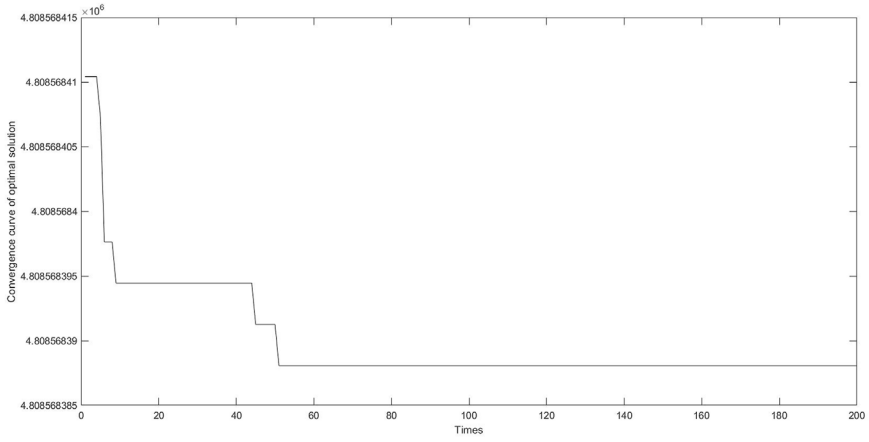



FIGURE 11
Optimal solution convergence graph.

**TABLE 2 Energy storage system configuration parameters.**

| Energy storage system parameters | Parameter value and unit |
|---|---|
| Total charge and discharge | 6.6706 MWh |
| Rated power | 5.084 MW |
| Nominal capacity | 6.6706 MWh |
| Difference in electric quantity at the beginning and end of each day | 1.848 kWh |

## 5 Conclusion

Based on the measured data of wind and solar output in a certain area of Ulanqab City, this paper proposes a new energy storage allocation strategy by analyzing the characteristics of the total output of wind farms and photovoltaic farms and the typical daily output curve of each season, which can stabilize the fluctuation of new energy grid-connected output.

After analyzing the characteristics of wind power output and photovoltaic output in a certain area of Ulanqab in 2020, it is found that there are fluctuations, uncertainty and intermittency. The minimum total daily power generation of wind power is 4985 MW, and the maximum total power generation is 84,230 MW, fluctuating wildly. In addition to the analysis of the overall output characteristics, the wind power output and photovoltaic output of a typical day in each season are also analyzed. The algorithm chosen for a typical day is the K-means algorithm. The analysis shows that the wind power output is the largest in spring and smaller in summer. The photovoltaic output is larger in spring and autumn and smaller in summer. For wind power output, photovoltaic output is generally small and relatively regular.

While achieving the smallest configuration energy storage capacity, the new energy storage allocation strategy has the ability to work continuously for a long time, and make the wind and solar active power output fluctuation as small as possible and meet the national new energy grid-connected fluctuation index. After the model is established, it is solved and simulated by genetic algorithm to obtain the required parameters. The effectiveness of the proposed strategy is proved by comparing the grid-connected power curves before and after optimization. In order to further verify the validity of the model, the optimized grid-connected volatility and the actual new energy output volatility within 1 min and 10 min were compared. The results show that the rated capacity of the required energy storage system is 6.6706 MW*h. The required energy storage system has a rated power of 5.084 MW. After stabilization, the volatility is about 2% and 15% at 1 min and 10 min, respectively. After the optimization of the model, the volatility is greatly reduced, which provides new ideas for energy storage to stabilize the volatility of renewable energy.

Through retrospective analysis, this work basically provides a new method for optimal configuration of energy storage to smooth out the volatility of wind power and photovoltaic active power grid connection. In future work, transmission system models will also be considered to ensure the feasibility of energy storage capacity allocation and energy storage power allocation. In addition, a stochastic approach that takes into account future demand and power generation forecasts will be adopted to better face the complex and changing power generation environment.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding authors.

## Author contributions

YL, RQ, HS, ZG, FF, and YN contributed to conception and design of the study. HS and ZG processed data. YL and YN performed the statistical analysis. RQ wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## Funding

The paper is supported by the Headquarters Science and Technology Project of State Grid Corporation of China (Project No. 52060021N00P). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Ahmad, A. A., Sirjani, R., and Daneshvar, S. (2020). New hybrid probabilistic optimisation algorithm for optimal allocation of energy storage systems considering correlated wind farms. *J. Energy Storage* 29, 101335. doi:10.1016/j.est.2020.101335

Alahmadi, A. A. A., Belkhier, Y., Ullah, N., Abeida, H., Soliman, M. S., Khraisat, Y. S. H., et al. (2021). Hybrid wind/PV/battery energy management-based intelligent non-integer control for smart DC-microgrid of smart university. *IEEE Access* 9, 98948–98961. doi:10.1109/access.2021.3095973

Ramseebaluck, A., and Chowdhury, S., Hybrid renewable-based reliable and cost-effective rural electrification in southern Africa. Proceedings of the IEEE Access, 2019, Sousse, Tunisia, March 2019: p. 1–6. doi:10.1109/IREC.2019.8754568

Dirin, S., Rahimiyan, M., and Baringo, L. (2023). Optimal offering strategy for wind-storage systems under correlated wind production. *Appl. Energy* 333, 120552. doi:10.1016/j.apenergy.2022.120552

Feng, X. (2019). Economical analysis of photovoltaic power station with battery energy storage system. *Adv. Technol. Electr. Eng. Energy* 38 (9), 52–58.

Ghaffari, A., Askarzadeh, A., and Fadaeinedjad, R. (2022). Optimal allocation of energy storage systems, wind turbines and photovoltaic systems in distribution network considering flicker mitigation. *Energy* 319, 119253. doi:10.1016/j.apenergy.2022.119253

Jamroen, C., Usaratniwart, E., and Sirisukprasert, S. (2019). PV power smoothing strategy based on HELES using energy storage system application: A simulation analysis in microgrids. *IET Renew. Power Gener.* 13 (13), 2298–2308. doi:10.1049/iet-rpg.2018.6165

Kabeyi, M. J. B., and Olanrewaju, O. A. (2022). Sustainable energy transition for renewable and low carbon grid electricity generation and supply. *Front. Energy Res.* 9. doi:10.3389/fenrg.2021.743114

Lamsal, D., Sreeram, V., Mishra, Y., and Kumar, D. (2018). Kalman filter approach for dispatching and attenuating the power fluctuation of wind and photovoltaic power generating systems. *IET Generation, Transm. Distribution* 12 (7), 1501–1508. doi:10.1049/iet-gtd.2017.0663

Lawal, K. O. (2015) IEEE. doi:10.1109/PSC.2015.7101691Hydro-based, renewable hybrid energy sytem for rural/remote electrification in Nigeria, Proceedings of the 2015 Clemson University Power Systems Conference (PSC), Clemson, SC, USA, March 2015.

Li, Q., Zhou, F., Guo, F., Fan, F., and Huang, Z. (2021). Optimized energy storage system configuration for voltage regulation of distribution network with PV access. *Front. Energy Res.* 9. doi:10.3389/fenrg.2021.641518

Makarov, Y. V., Du, P., Kintner-Meyer, M. C. W., Jin, C., and Illian, H. F. (2012). Sizing energy storage to accommodate high penetration of variable energy resources. *IEEE Trans. Sustain. Energy* 3 (1), 34–40. doi:10.1109/tste.2011.2164101

Meiqin, M., Jialing, H., and Liuchen, Z. (2021). Energy storage optimization configuration method considering conditional forecast error correction. *Acta Energiae Solaris Sin.* 42 (2), 410–416.

Mingoti, S. A., and Lima, J. O. (2006). Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *Eur. J. Operational Res.* 174 (3), 1742–1759. doi:10.1016/j.ejor.2005.03.039

Mishra, S., Saini, G., Chauhan, A., Upadhyay, S., and Balakrishnan, D. (2023). Optimal sizing and assessment of grid-tied hybrid renewable energy system for electrification of rural site. *Renew. Energy Focus* 44, 259–276. doi:10.1016/j.ref.2022.12.009

Paliwal, P. (2021). Reliability-based optimal sizing for an isolated wind–battery hybrid power system using butterfly PSO. *Adv. Energy Technol.* 766, 163–172.

Sadeghian, O., Oshnoei, A., Khezri, R., and Muyeen, S. (2020). Risk-constrained stochastic optimal allocation of energy storage system in virtual power plants. *J. Energy Storage* 31, 101732. doi:10.1016/j.est.2020.101732

Shen, J., Huang, S., Liu, C., Li, S., and Wu, J. (2023). Optimal configuration method of wind farm hybrid energy storage based on EEMD-EMD and grey relational degree analysis. *Front. Energy Res.* 10. doi:10.3389/fenrg.2022.1021189

Soliman, M. S., Belkhier, Y., Ullah, N., Achour, A., Alharbi, Y. M., Al Alahmadi, A. A., et al. (2021). Supervisory energy management of a hybrid battery/PV/tidal/wind sources integrated in DC-microgrid energy storage system. *Energy Rep.* 7, 7728–7740. doi:10.1016/j.egyr.2021.11.056

Sun, W., Gong, Y., and Luo, J. (2023). Energy storage configuration of distribution networks considering uncertainties of generalized demand-side resources and renewable energies. *Sustainability* 15 (2), 1097. doi:10.3390/su15021097

Wang, J., Du, W., Yang, D., He, G., and Zhang, X. (2021). Optimal configuration of multi-type energy storage for integrated energy system considering multi-energy trading with load substitution. *E3S Web Conf.* 237, 02016. doi:10.1051/e3sconf/202123702016

Zhang, L., Chen, L., Zhu, W., Lyu, L., Cai, G., and Hai, K. L. (2022). Research on the optimal allocation method of source and storage capacity of integrated energy system considering integrated demand response. *Energy Rep.* 8, 10434–10448. doi:10.1016/j.egyr.2022.08.184

Zhang, Y., Dong, Z. Y., Luo, F., Zheng, Y., Meng, K., and Wong, K. P. (2016). Optimal allocation of battery energy storage systems in distribution networks with high wind power penetration. *IET Renew. Power Gener.* 10 (8), 1105–1113. doi:10.1049/iet-rpg.2015.0542

Zhao, M., Wang, Y., Wang, X., Chang, J., Zhou, Y., and Liu, T. (2022). Modeling and simulation of large-scale wind power base output considering the clustering characteristics and correlation of wind farms. *Front. Energy Res.* 10. doi:10.3389/fenrg.2022.810082

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 31 March 2023
DOI 10.3389/fenrg.2023.1167316



OPEN ACCESS

EDITED BY
Huan Xi,
Xi'an Jiaotong University, China

REVIEWED BY
Yushuai Li,
University of Oslo, Norway
Hong Fan,
Shanghai University of Electric Power, China

*CORRESPONDENCE
Shirui Wang,
283126345@qq.com

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 16 February 2023
ACCEPTED 20 March 2023
PUBLISHED 31 March 2023

CITATION
Deng X, Wang S, Wang W, Yu P and Xiong X (2023), Optimal defense strategy for AC/DC hybrid power grid cascading failures based on game theory and deep reinforcement learning.
*Front. Energy Res.* 11:1167316.
doi: 10.3389/fenrg.2023.1167316



# Optimal defense strategy for AC/DC hybrid power grid cascading failures based on game theory and deep reinforcement learning

Xiangli Deng[1], Shirui Wang[1]*, Wei Wang[2], Pengfei Yu[3] and Xiaofu Xiong[3]

[1]School of Electric Power Engineering,Shanghai University of Electric Power, Shanghai, China, [2]State Grid Wenzhou Power Supply Company, Wenzhou, Zhejiang Province, China, [3]School of Electric Power Engineering, Chongqing University, Chongqing, China

This paper proposes a two-person multi-stage zero-sum game model considering the confrontation between cascading failures and control strategies in an AC/DC hybrid system to solve the blocking problem of DC systems caused by successive failures at the receiving end of an AC/DC system. A game model is established between an attacker (power grid failure) and a defender (dispatch side). From the attacker's perspective, this study mainly investigates the problem of system line failures caused by AC or DC blockages. From the perspective of dispatch-side defense, the multiple-feed short-circuit ratio constraint method, output adjustment measures of the energy storage system, sensitivity control, and distance third-segment protection adjustment are used as strategies to reduce system losses. Using as many line return data as possible as samples, the deep Q-network (DQN), a deep reinforcement learning algorithm, is used to obtain the Nash equilibrium of the game model. The corresponding optimal dispatch and defense strategies are also obtained while obtaining the optimal sequence of tripping failures for AC/DC hybrid system cascading failures. Using the improved IEEE 39-node system as an example, the simulation results verify the appropriateness of the two-stage dynamic zero-sum game model to schedule online defense strategies and the effectiveness and superiority of the energy storage system participating in defense adjustment.

KEYWORDS

**AC/DC hybrid power system, cascading failure, energy storage system, deep reinforcement learning, multi-stage dynamic zero-sum game**

## 1 Introduction

Recently, there have been major outages caused by interlocking faults around the world. To analyze these incidents, the principles of interlocking fault propagation have been studied extensively (Ding et al., 2017; Fang., 2014). Currently, related research is still mainly focused on the conventional AC grid, for example, complex systems theory (Cao et al., 2012; Cao et al., 2011), particularly complex network theory (Fan et al., 2018; Dey et al., 2016; Xu et al., 2010), has been used to study the chain fault dynamics and evolution form in terms of complex network topology, but without considering the specific system fault risk. Moreover, based on the research perspective of tidal current calculation and stability analysis, the chain fault development process has been expressed. In the paper (Wang et al., 2019), the residual load rate and chain fault propagation distance metrics are used to quantify the impact of

vulnerable lines on the depth and breadth of chain fault propagation; in the paper (Zhang et al., 2017), a branch fault percolation probability model was constructed to identify vulnerable branches of the grid under normal operation and predict the faulty branches of the grid after a fault occurs.

When a chain fault occurs in an AC system, it will easily lead to a DC lockout owing to a decrease in the support capacity of the AC system to the DC system, which will in turn lead to a larger-scale tidal shift and increase the probability of a major power outage in the grid. The existing N-k fault scheduling strategy requires a large amount of data calculation to determine chain faults, which consumes considerable time and cannot meet the actual needs. Therefore, studying a fast identification method for fault-tripping sequences and an optimal regulation strategy for tripping sequences in the chain fault evolution of hybrid AC/DC systems is necessary.

The chain of faults in the evolution process will continuously trigger the tripping and decommissioning of grid lines. The regulation strategy can simultaneously control the system to cut off the propagation path of the chain of faults, which can be considered as a multi-stage dynamic game between two people. Hence, game theory can be applied to the chain fault sequence search process. For example, the paper (Zhang et al., 2020) studied the contribution of various fault chains to grid losses by constructing a cooperative game framework for grid branches. Paper (Ding et al., 2016) analyzed the coordination between preventive control and blocking control of cascading failures, and a coordinated control model based on risk assessment is proposed for power system cascading failures in terms of reducing the risk of blackout.

Deep reinforcement learning algorithms have a broad application scenario for solving the Nash equilibrium of a two-person multi-stage game model and can be solved quickly and accurately. Since Minh et al. proposed the concept of deep Q-networks (DQN) in 2015, the application scenario and scope of the DQN algorithm have been continuously expanded. DQN is a novel deep reinforcement learning algorithm combining deep learning with reinforcement learning (Mnih et al., 2015; Van Hasselt et al., 2016); in particular, it combines the Q-learning reinforcement learning algorithm and convolutional neural network. This causes shorter convergence time and training time than the Q-learning algorithm and more convenient processing for increasing the data dimension of the AC/DC hybrid system trend. The introduction of deep learning in reinforcement learning strengthens the generalization ability of the algorithm.

As the grid is added to the DC transmission line, the corresponding novel energy-generating units will also be integrated into the system operation, along with the continuous development and progress of energy storage technology. The excellent power characteristics of the storage system can be triggered in the event of a fault on the AC side of the grid initiating DC side voltage fluctuations by quickly releasing or absorbing the stored power of the storage system to maintain the system's normal operation. There has been further improvement in the fault ride-through capability of new energy units (Li et al., 2022). In the paper (Duan et al., 2019), a reinforcement-learning-based online optimal (RL-OPT) control method is proposed for the hybrid energy storage system (HESS) in ac–dc microgrids involving photovoltaic systems and diesel generators (DGs). The paper (Ying et al., 2023) proposes an online energy management strategy (OEMS) based on long short-term memory (LSTM) network and deep deterministic policy gradient (DDPG) algorithm to counteract the effects of these real-time fluctuations, and the proposed OEMS has the advantages of small tracking error, model-free control, and continuous action control. This paper (Yang et al., 2022) combined with the deep reinforcement learning algorithm, the Markov multi-energy interaction model is established with distributed structure, and the problem of continuous action in the model is solved, and finally the energy profit of the local energy market (LEM) in Energy Internet (EI) is maximized.

Therefore, energy storage system adjustment can be used as one of the control strategies employed in AC/DC hybrid systems to cope with chain failures; therefore, the capacity configuration of the storage system needs to be studied. Paper (Liu et al., 2016) proposed a control strategy using energy storage devices to improve the injection current characteristics of wind farms to ensure the smooth operation of the system; the capacity requirements of the storage system were studied by simulation. Studies (Yan et al., 2020; Dai et al., 2016; Song et al., 2018) have proposed an energy storage power control strategy to adjust the output based on the battery's state of charge (SOC); they used simulation analysis to obtain the battery capacity requirements. In paper (Liu et al., 2022), an optimal configuration model of the energy storage double layer was established based on the effective use of energy storage for the load margin of the integrated energy system, thus achieving an effective balance of the load margin in the integrated energy system.

In summary, this study first constructs a two-person multi-stage zero-sum game model to consider the process of mutual confrontation between interlocking faults and regulation strategies in AC/DC hybrid systems. From the attacker (power grid failure)'s point of view, this study studied the problem of continuous fault decommissioning caused by AC disturbance or DC blocking from the defender (dispatch side)'s point of view. This study used sensitivity control, distance III protection adjustment, the multi-feeder short-circuits ratio constraint method, and energy storage system capacity adjustment measures as defense strategies to reduce system losses. With as many lines decommissioning data as possible as samples, the DQN deep reinforcement learning algorithm was used to find the Nash equilibrium of the game model and obtain the optimal fault tripping sequence of the AC/DC hybrid system chain fault while obtaining the corresponding optimal dispatching defense strategy.

# 2 AC/DC system interlocking fault and its defense model

## 2.1 AC/DC system fault evaluation index

### 2.1.1 Line return risk

AC/DC interlocking faults are analyzed in two main aspects: first, the tidal current transfer and hidden faults of protection are the main factors; second, the phase change failure of the DC system is the main factor triggering the DC system. Therefore, this study uses

the risk factor as an evaluation index to further assess the impact of the subsequent decommissioning of the line.

The probability of grid line decommissioning is affected by the state of the grid after the occurrence of the previous fault; the corresponding Markov chain fault probability model is shown in Equation 1.

$$P_n(t) = \begin{cases} 1 & F \geq F_{\max} \\ \dfrac{(1-\mu_1)F + \mu_1 F_{\max} - F^n_{\max}}{F_{\max} - F^n_{\max}} & F^n_{\max} \leq F < F_{\max} \\ \mu_1 & F^n_{\min} < F < F^n_{\max} \end{cases} \quad (1)$$

where $\mu_1$ is the historical outage statistical probability of the branch, and $F$ is the tide on the branch after the last fault removal. $F^n_{\min}, F^n_{\max}$ is the lowest and highest tide value for the normal operation of the branch, and $F_{\max}$ is the tide limit of the branch. According to the definition of risk, the AC system risk indicator can be obtained as shown in Equation 2.

$$\delta_1 = P_n(t) \times L_{loss}(t) \quad (2)$$

where $P_n(t)$ is the probability of decommissioning of the branch $n$ and $L_{loss}(t)$ is the load loss rate at $t$ , i.e., the ratio of the load loss to the total system load. Therefore, the risk indicator $\delta_1$ can be used to assess the risk of AC system decommissioning.

### 2.1.2 AC bus multi-feeder voltage support capability

In mixed-connection systems, the main focus is on assessing the voltage support capability of the AC system and the phase-change bus voltage. To evaluate the voltage stability of the hybrid system, a multi-feeder short-circuit ratio was used to reflect the system's grid strength and voltage support capacity (Lin et al., 2008).

The multi-feeder short-circuit ratio indicator is defined as shown in Equation 3.

$$M_{ISCRi} = \frac{S_{aci}}{P_{deqi}} = \frac{S_{aci}}{P_{dNi} + \sum_{j=1, j\neq i}^{n} M_{IIFji} \cdot P_{dNj}} \quad (3)$$

where is the short-circuit capacity of the converter bus; is the equivalent DC power after considering other DC effects, is the rated DC power of the DC, respectively, and is the multi-feed influence factor between branches. Therefore, the short-circuit ratio variation is established as an indicator to assess the voltage support capability of the receiving system, as shown in Equation 4.

$$\delta_2 = \sum_{i=1}^{n} \left| M_{ISCRi,s+1} - M_{ISCRi,s} \right| \quad (4)$$

Short-circuit capacity decline is mainly triggered by the fault line opening, which leads to changes in the system structure, causing the system impedance to become larger, and the AC to DC system support capacity is reduced, increasing the possibility of system voltage fluctuations. Thus, the multi-feed short-circuit ratio index can effectively reflect the impact of line opening on the system voltage support capacity.

### 2.1.3 Risk of DC phase change failure

The action criterion of phase-change failure protection is that the DC line's bus voltage on the inverter side is lower than the threshold voltage and exceeds a certain time; then, the protection will be activated, and the DC line will be blocked. When the overrun arc extinguishing angle $\gamma$ is smaller than the limit arc extinguishing angle $\gamma_{\min}$ , AC disturbance will occur on the inverter side, which will cause a DC phase-change failure fault. Thus, the phase change failure is evaluated by determining the limit arc extinguishing angle when the phase change fails. The minimum arc-extinguishing angle at phase-change failure is obtained, as shown in Equation 5.

$$\gamma_{\min} = \arccos\left( \frac{\sqrt{2}\, k L_c I_d^*}{U_L^*} + \cos\beta^* \right) \quad (5)$$

where $I_d^*$ is the DC at the time of phase-change failure, $U_L^*$ is the voltage at the time of phase-change failure, and $\beta^*$ is the inverter override trigger angle at the time of phase-change failure. The commutation bus voltage evaluation index can be established from this, as shown in Equation 6.

$$\delta_3 = \omega_1 \frac{U_L^*}{U_L} + \omega_2 \frac{I_d^*}{I_d} + \omega_3 \frac{\beta^*}{\beta} \quad (6)$$

where $U_L$ denotes the rated voltage of the line; $I_d$ denotes the rated current of the DC line, and $\beta$ denotes the rated override trigger angle of the inverter.

## 2.2 Evaluation of interlocking faults in AC/DC transmission systems

Based on the basic structure of the AC/DC hybrid system and the possible risk of safety failure, this study compiles the evolution form of the chain failure of the AC/DC hybrid system, as shown in Figure 1.

The AC system risk indicator $\delta_1$ was used to assess the possible overload decommissioning of AC system lines owing to frequency and power angle problems in the AC system. The short-circuit ratio variation in $\delta_2$ was used to assess the grid's support capability. The converter bus voltage assessment indicator $\delta_3$ was used to assess the converter bus low-voltage situation, reflecting the voltage support capability of the converter side.

The analysis of the AC/DC hybrid system chain fault characteristics and assessment indexes shows that the evolution form of chain faults mainly lies in their mutual coupling on the AC/DC side of the development of changes and then continuously expands the scale and coverage of chain faults. Combined with the above chain fault mechanism analysis, the final establishment of line disconnection risk assessment indicators is shown in Eq. 7.

$$R = \lambda_1 \delta_1 + \lambda_2 \delta_2 + \lambda_3 \delta_3 \quad (7)$$

where $\lambda_1, \lambda_2, \lambda_3$ is the scale factor corresponding to each evaluation index.

## 2.3 Chain fault regulation strategy

### 2.3.1 Response to chain failure power adjustment strategy

(1) Generator and load sensitivity control strategies

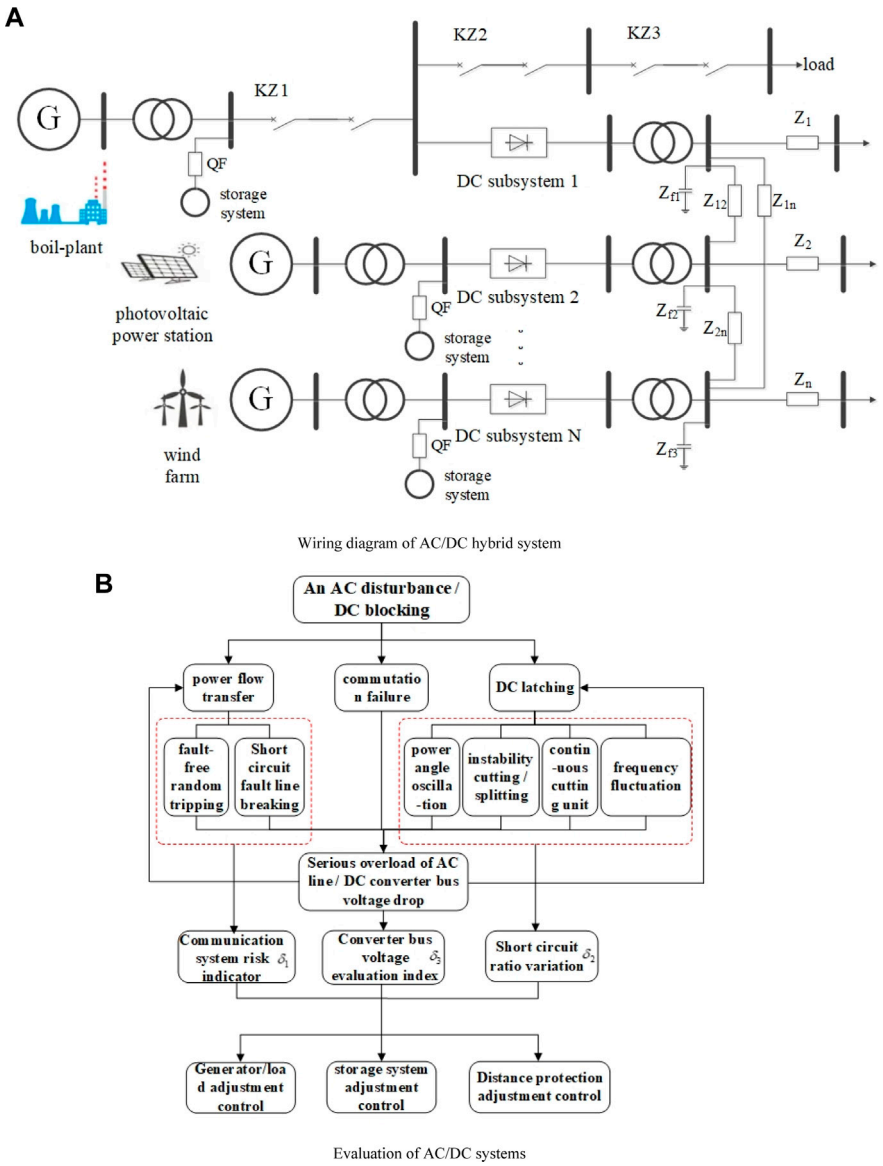



FIGURE 1
Evolution form of cascading failures in AC/DC hybrid system.

The branch's overload due to the branch's disconnection can be adjusted by the method of generator output and load control using sensitivity coefficient pairing. The sensitivity pairing method comprising generator and load pairing using a correlation matrix is simpler and quicker to control. Therefore, priority is given to controlling nodes with large power sensitivity to minimize system losses. Eq. 8 shows the power sensitivity $\eta_{i-j}$ between node pairs $i$, $j$.

$$\eta_{i-j} = \beta_{n,i} - \beta_{n,j} \tag{8}$$

Therefore, the sensitivity control strategy is used to obtain the branch generator set output and the load power is to be adjusted, as shown in Eq. 9.

$$\begin{cases} \Delta P_{L1} = \sum_{n=1}^{m} (P_{Ln} - P_{Ln\,\max}) \big/ \eta_{Li-Lj} \\ \Delta P_{G1} = \sum_{n=1}^{m} (P_{Gn} - P_{Gn\max}) \big/ \eta_{Gi-Gj} \end{cases} \tag{9}$$

where $P_n$ is the actual power of branch $n$; $P_{n\max}$ is the power limit of branch; and $m$ is the total number of branches.

(2) Multi-feed short-circuit ratio constraint

For a grid containing multi-feeder DC systems, each DC system's multi-feeder short-circuit ratio index needs to be controlled within the normal range to ensure that the AC system strength can match the transmission capacity of the DC system constrained, as shown in Eq. 10.

$$K_{MISCRi} \geq K_{MISCR,\min} \quad i = 1, 2, \cdots N_{dc} \tag{10}$$

where $K_{MISCRi}$ represents the multi-feed-in short-circuit ratio of the inverter side of the DC system $i$ ; $K_{MISCR,\min}$ is the multi-feed-in short-circuit ratio limit used in the hybrid system.

Therefore, the corresponding multi-feed-in short-circuit ratio parameters are calculated from the generator unit output at node $\Delta P_{G2}$ and the load power at node $\Delta P_{L2}$. The above constraints are applied to these multi-feed-in short-circuit ratio parameters to control the voltage support capability of the AC to DC system, which can ensure the avoidance of phase-change failure of the DC system.

(3) Prevention of phase-change failure commutation bus voltage adjustment strategy

In the AC/DC hybrid system, the converter transformer ratio, DC operating current, converter phase reactance, converter bus voltage, and overrun trigger angle affect the magnitude of the arc extinguishing angle. The reactive power adjustment value of the inverter-side converter to be adjusted in the DC transmission system is obtained using the control strategy, as shown in Eq. 11.

$$\Delta Q_{\mathrm{dc}} = \frac{2\mu + \sin(2\gamma) - \sin(2(\gamma+\mu))}{2(\cos\gamma - \cos(\gamma+\mu))} I_{\mathrm{d}} U_{\mathrm{d}} \tag{11}$$

where μ is the phase change angle; $\gamma$ is the arc extinguishing angle; $I_d$ is the DC-side current, and $U_d$ is the ideal no-load DC voltage.

Therefore, the corresponding reactive power adjustment is calculated using the converter bus voltage evaluation index $\delta_3$, and the adjustment is incorporated into the subsequent overall adjustment strategy of the hybrid grid to realize the adjustment of the system.

### 2.3.2 Scheduling and adjustment methods of storage, source, network, and load of the AC/DC hybrid grid

By adjusting the storage source and network load scheduling for each phase of the interlocking fault, the interlocking fault is cut off before expanding further. Here, the defensive measures for chain faults are mainly the multi-feeder short-circuit ratio constraint method, energy storage system adjustment, matching distance III protection action adjustment strategy (Yang et al., 2011; Lin et al., 2011) and sensitivity control (Xu et al., 2017) for auxiliary control of the AC side. The line adjustments for the tidal overload are as follows.

1) When the line is overloaded, the distance protection section Ⅲ is adjusted so that it does not misfire. The action characteristic angle of the distance protection section Ⅲ is first adjusted, and the energy storage system output near the branch is adjusted.
2) After ensuring that the distance protection section III does not misfire and that the energy storage system is involved in the adjustment, sensitivity control and multi-feeder short-circuit ratio constraint are used to complete the control of the line tide.
3) After ensuring that the system tide can operate normally, the power output of each energy storage system of the entire AC/DC system is calculated to restore the system's balance.

(1) Energy-storage regulation strategy

When the system is in normal operation, the tidal shift or fault causes changes in the power output of the grid generating units and load power, which further affects the grid's tidal fluctuation. The energy storage regulation strategy is initiated, charging the energy storage system when the active power output increases, and discharging the energy storage system to maintain the normal operation of the system when the active power output decreases or the load power is lost. Therefore, when a chain fault occurs in the hybrid system, the energy storage system can be used to adjust the power output of the storage unit to achieve tidal control of the grid; the control strategy requires a high response speed of the energy storage system. Therefore, this study mainly uses power-type energy storage devices such as supercapacitors in the storage unit.

The DC-side power variation during the dynamic process is given by Equation 12.

$$\begin{cases} \Delta P_{\mathrm{C}} = P_{\mathrm{S}} - P_{\mathrm{G}} - P_{\mathrm{sc}} - P_{Lg} \\ \Delta P_{\mathrm{C}} \cdot \Delta t = \frac{1}{2} C (u_{\mathrm{dc}} + \Delta u_{\mathrm{dc}})^2 - \frac{1}{2} C u_{\mathrm{dc}}^2 \end{cases} \tag{12}$$

where $P_s, P_G, P_{Lg}$ is the power generated by the generator side, grid side, and reactor of the generator set; $u_{dc}$ is the DC-side voltage value of the generation system during stable operation, and $\Delta u_{dc}$ is the DC-side voltage variation.

Let the energy flowing to the energy storage system during the failure time $\Delta t$ be $W_{SC}$. From Equation 12, we have $\Delta P_{\mathrm{SC}} = P_{\mathrm{S}} - P_{\mathrm{G}} - P_{\mathrm{L}}$ , and we obtain Equation 13 as follows:

$$W_{\mathrm{SC}} = \Delta P_{\mathrm{SC}} \cdot \Delta t \tag{13}$$

Eq. 14 can also be obtained as follows.

$$W_{\mathrm{SC}} = \frac{1}{2} C \left(U_{\mathrm{SC}} - I_{\mathrm{SC}} R_{eq}\right)^2 \tag{14}$$

Substituting Eq. 14 into Equation 13 yields the formula for calculating the capacity of the energy storage unit, as shown in Eq. 15.

$$C = \frac{2\Delta P_{\mathrm{SC}} \cdot \Delta t}{\left(U_{\mathrm{SC}} - I_{\mathrm{SC}} \cdot R_{eq}\right)^2} \tag{15}$$

where $\Delta P_{SC}$ is the power input to the energy storage system; $\Delta t$ is the fault duration; $U_{SC}$ is the upper voltage limit of the supercapacitor; $I_{SC}$ is the charging current, and $R_{eq}$ is the equivalent resistance of the energy storage system.

Alternatively, after determining the power of the energy storage system based on the tidal short-circuit calculation, the supercapacitor capacity value (Tian et al., 2016) can be obtained to absorb and store all the power passing through the generator-side alternator to maintain the system voltage stability, as shown in Eq. 16.

$$C_{SC} = \frac{s_T P_T t}{U_{\mathrm{scmax}}^2 - U_{\mathrm{scnorm}}^2} \tag{16}$$

where $C_{SC}$ is the capacitance value; $P_T$ is the output power of the generator set under normal grid operation; $s_T$ is the speed difference when the output power is $P_T$; $t$ is the fault duration; $U_{sc\,\max}$ and $U_{scnorm}$ are the maximum operating voltage and normal operating voltage allowed for the energy storage unit, respectively. When

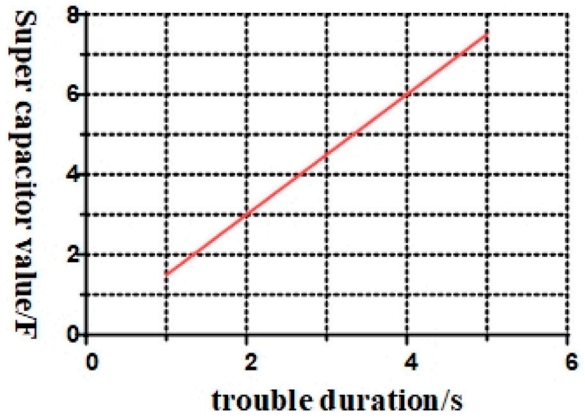



FIGURE 2
Relationship between super capacitance value and failure duration.

considering the worst-case condition, the rated power of the unit can be substituted in the calculation, as shown in Eq. 17.

$$\begin{cases} P_{\mathrm{T}} = P_{\mathrm{N}} \\ s_{\mathrm{T}} = s_{\mathrm{N}} \end{cases} \tag{17}$$

The energy storage unit must store the excess power generated by the system because of the adjustment when the fault occurs, but not to exceed the upper limit of the storage unit. According to Eq. 16, Eq. 17, a schematic of the supercapacitor value *versus* the duration of the failure can be plotted, as shown in Figure 2.

Figure 2 shows that the length of the fault duration is proportional to the supercapacitor value. According to the calculation of Eq. 16 and Eq. 17, if the fault duration of the generator set is 2 s, the required supercapacitor value is 3.06 F.

(2) Generator and load regulation strategies

According to the required adjustment amount of power output of all generating units and the required adjustment amount of load nodes obtained from the scheduling defense strategy in the previous section, mainly including the total power adjustment amount of each generating unit node side $\Delta S_G = \Delta P_{G1} + \Delta P_{G2} + \Delta Q_{dc}$ and the load loss amount of load side $\Delta S_L = \Delta P_{L1} + \Delta P_{L2}$, the above adjustment amounts are calculated using Eq. 16 and Eq. 17 to obtain the corresponding required adjustment of energy storage system power output, to realize the system generators and the load control strategy through the energy storage system power output control adjustment. The above adjustment amount is calculated using Eq. 16 and Eq. 17 to obtain the corresponding required adjustment of the energy storage system output.

(3) Grid section III distance protection setting adjustment strategy.

To prevent the distance protection from being triggered by the tidal current transfer when the chain fault occurs, distance protection control measures are used; that is, the action characteristic angle of section III distance protection is adjusted.

Whether the distance protection takes action is determined by identifying whether the measured impedance of the protection position falls into the action characteristics generated by the rectified impedance. Assuming that the measured impedance fully exhibits the resistance characteristics, the measured impedance is given by Eq. 18.

$$Z_m = \frac{U_m^2}{P_m} \tag{18}$$

where $U_m$ is the measured voltage at the line distance protection position and $P_m$ is the tidal power of the line where the line distance protection position is located.

When the line is identified as having a tidal shift, the action characteristic of section III distance protection is adjusted to make $Z_m$ avoid the action range of section III of distance protection by reducing the range of action characteristics. The adjusted action-angle characteristic is given by Eq. 19.

$$90° + \theta < \arg\frac{Z_{set} - Z_m}{Z_m} < 270° - \theta \tag{19}$$

where θ is the action characteristic adjustment angle, according to Eq. 19, can be obtained from the adjustment angle θ, as shown in Eq. 20.

$$\theta = \arg\frac{Z_{set} - Z_m}{Z_m} - 90° \tag{20}$$

# 3 Optimal defense strategy for chain failures based on game DQN model

## 3.1 Multi-stage zero-sum game-based chain failure model

The above demonstrates that the goal of the chain fault is to cause damage to the power system, whereas the goal of the scheduler is to interrupt the development of the chain fault and reduce the loss of the power system. Therefore, the chain fault and scheduling adjustment can be regarded as an attacker and defender against the power system, respectively, and the interaction between them can be expressed as a game between them.

When a chain fault occurs, the state of the grid under each stage evolves; hence, the state of the grid at stage $t$ is defined as $s_t = \{s_{1,t}, s_{2,t}, \cdots s_{N,t}\}$ ; where $s_{n,t}$ is the state of the branch line $n$ and takes the value $s_{n,t} \in \{0, 1\}$ to indicate whether the branch line $n$ operates normally at stage $t$ . Then, the attacker's strategy $a_t^1 \in \{1, 2, \cdots, N\}$ is defined, mainly selecting the branch number to take an attack action to decommission the line. The defender's strategy $a_t^2$ is defined to represent the above-mentioned scheduling adjustment measures based on the decommissioned line, and the set of actions taken by both games is defined as $\pi$.

Using the risk factor as a function of the gains of the attacker and defender at each stage of the fault development process subsequently allows for evaluating the losses triggered by the attacker. Therefore,

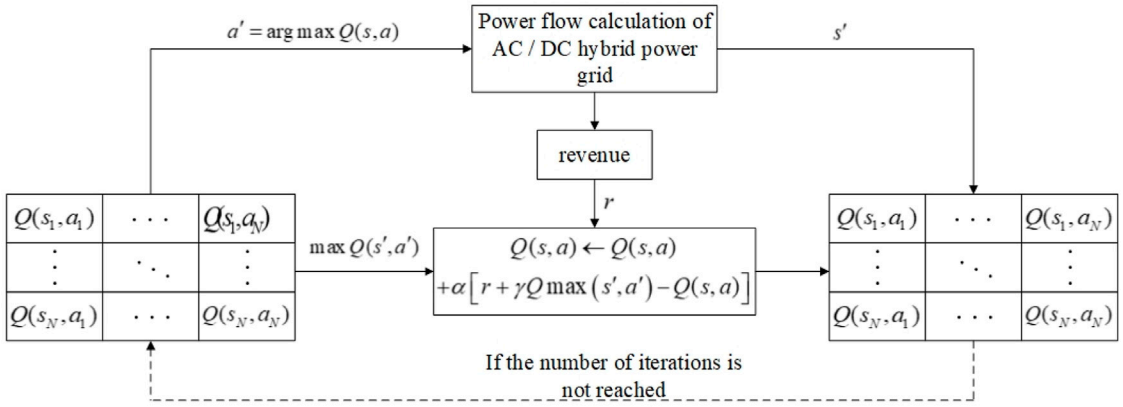



FIGURE 3
Q-Learning algorithm calculation flowchart.

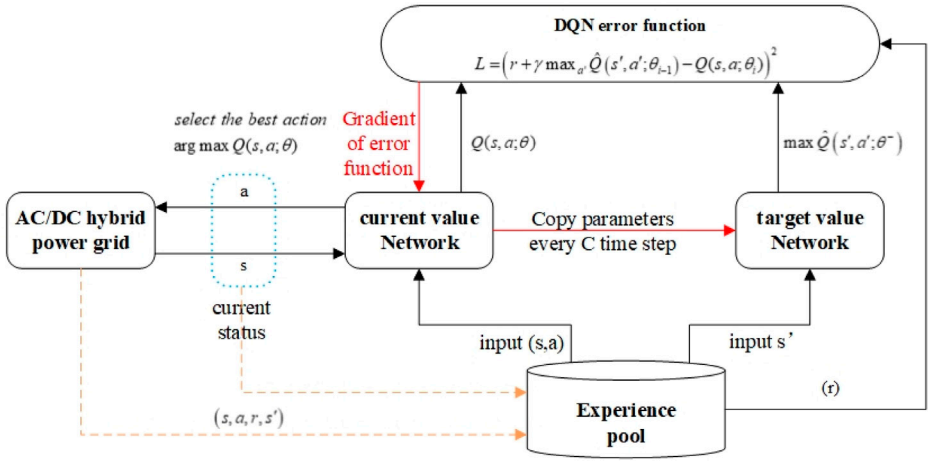



FIGURE 4
Structural diagram of Q value calculation for DQN algorithm.

the gains for both sides are given by Eq. 21 based on the assumptions above.

$$\begin{cases} V_1 = \sum_{t=0}^{T} \gamma^{t-1} r_t^1 = \sum_{t=0}^{T} \Psi_n(t) \\ V_2 = -V_1 \end{cases} \tag{21}$$

where $\gamma$ is the discount factor; $T$ is the number of gaming stages.

By analyzing the payoff function, we find that the payoff of the multi-stage zero-sum game pays more attention to the overall payoff generated by one game considering the single-stage payoff. When the defender considers the optimal adjustment strategy, the optimal attack strategy of the attacker can be sought as the Nash equilibrium of the game model, i.e., only the optimal gain of the attacker needs to be considered.

Therefore, both attackers and defenders must adopt the optimal strategy of $\pi^*$ to maximize their gains, which is expressed as shown in Eq. 22.

$$\begin{cases} V_1\left(a_1^{1\pi^*}, a_2^{1\pi^*}, \cdots a_T^{1\pi^*}\right) \geq V_1\left(a_1^{1\pi}, a_2^{1\pi}, \cdots, a_T^{1\pi}\right) \\ V_2\left(a_1^{2\pi^*}, a_2^{2\pi^*}, \cdots a_T^{2\pi^*}\right) \geq V_2\left(a_1^{2\pi}, a_2^{2\pi}, \cdots, a_T^{2\pi}\right) \end{cases} \tag{22}$$

where $V_1$ is the gain for the attacker; $V_2$ is the gain for the defender, and $a_t^{1\pi^*}, a_t^{2\pi^*}$ represents the action strategies of both sides of the game if the optimal strategy $\pi^*$ is used in the $t$ phase.

## 3.2 Nash equilibrium solution method for chain failure game model based on DQN algorithm

### 3.2.1 Q-learning reinforcement learning algorithm

A Markov decision process is typically used to solve a problem using reinforcement learning. It is mainly represented by $<S, A, P, R>$, which contains a set of grid states $S$, a set of trip sequences $A$, state transfer probabilities $P(s,a,s,s^*)$, and a reward function $R(s,a,s,s^*)$ in the search model for the optimal tripping

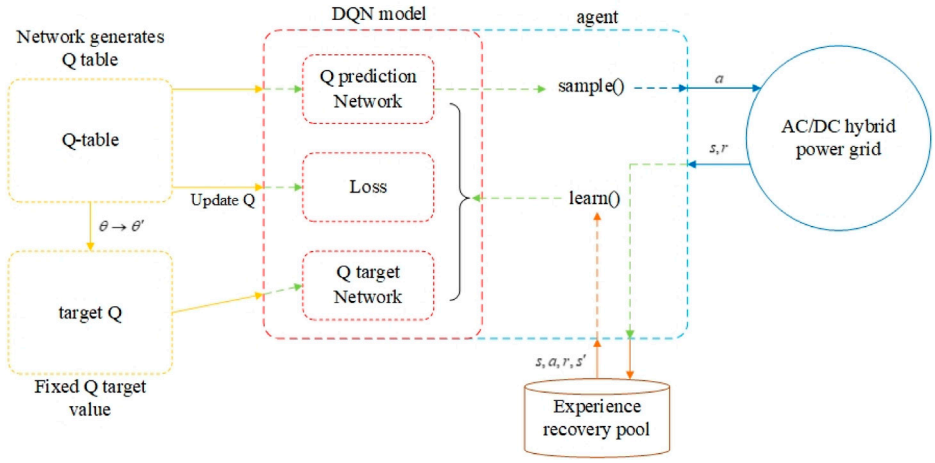



FIGURE 5
Flowchart of DQN algorithm calculation.

sequence of interlocking faults in AC/DC hybrid systems. Q-learning algorithm, as a common reinforcement learning algorithm, has $Q(s,a)$ as the expectation that an action $a$ can be taken at a state $s$ at a certain time to obtain a gain, and then the environment is based on the agent's action $r$. Finally, the algorithm constructs the state $s$ and the action $a$ into a Q-table to store the Q-value, and selects the action that can obtain the maximum benefit according to the Q-value. The main advantage of the Q-learning algorithm is the use of the Bellman equation to determine the optimal policy for the Markov process. The Bellman equation used by the algorithm is shown in Eq. 23 and Eq. 24:

$$V^{\pi}(s) = \sum_{a} \pi(s,a) \sum_{s'} P^{a}_{ss'} \left[ R^{a}_{ss'} + \gamma V^{e}(s') \right] \tag{23}$$

$$Q^{\pi}(s,a) = \sum_{s} P^{a} \left[ R^{a}_{ss'} + \gamma \sum_{a'} Q^{\pi}(s',a') \right] \tag{24}$$

where $Q^{\pi}(s,a)$ denotes the cumulative return obtained when state $s$ and action $a$ both adopt the optimal strategy $\pi$ . $V^{\pi}(s)$ denotes the cumulative return obtained when the state $s$ adopts the optimal strategy $\pi$ .

The Q-learning algorithm is updated as shown in Eq. 25.

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[ r + \gamma Q \max(s',a') - Q(s,a) \right] \tag{25}$$

The formation process of the Q-table and its parameters during the calculation of the Q-learning algorithm are shown in Figure 3.

### 3.2.2 Q-learning algorithm based on deep learning Q-function

However, maintaining and updating Q-table tables in the Q-learning algorithm requires a lot of computing resources and computing time, and there is a dimensional explosion problem. Therefore, a non-linear function approximator can be used to approximate Q. Neural network is a commonly used non-linear function approximator, and a Q-learning algorithm that uses a deep learning network as a Q function approximator is the DQN algorithm.

The DQN algorithm focuses on two main aspects: constructing the target network and introducing an experience-replay mechanism.

(1) Construction of the target network

The DQN algorithm continues to consider the task of agent-environment interaction in sequences of actions, observations, and rewards. In each stage, the agent selects an action $a_t$ from the action set $A = \{1, \ldots, K\}$, after which the environment modifies its state and receives a reward.

The agent aims to interact with the network by selecting actions that maximize future returns. Similarly, the depreciation factor $\gamma$ needs to be set to define the future depreciation return at time $t$, as shown in Eq. 26.

$$R_t = \sum_{t'=t}^{T} \gamma^{t'-t} r_{t'} \tag{26}$$

For the original Q-learning algorithm, the Bellman equation, i.e., linear function approximator is used as an iterative update to estimate the action-value function. While the DQN algorithm uses a non-linear function approximator, i.e., a neural network for estimation, we refer to the neural network function approximator with weights $L_i(\theta_i) = E_{s,a\sim\rho(\cdot)}[(y_i - Q(s,a;\theta_i))^2]$ as Q-network. The Q-network can be trained by minimizing the loss function, as shown in Eq. 27.

$$L_i(\theta_i) = E_{s,a\sim\rho(\cdot)}\left[ (y_i - Q(s,a;\theta_i))^2 \right] \tag{27}$$

where $y_i = E_{s'\sim\varepsilon}[r + \gamma \max_{a'} Q(s',a';\theta_{i-1}) | s,a]$ .

For this neural network, we can use the stochastic gradient descent to minimize the loss function such that the parameters of the neural network can be updated to the maximum extent, and the gradient of the loss function is shown in Eq. 28.

$$\nabla_{\theta_i} L_i(\theta_i) = E_{s,a\sim\rho(\cdot);s'\sim\varepsilon}\left[ (r + \gamma \max_{a'} Q(s',a';\theta_{i-1}) - Q(s,a;\theta_i)) \nabla_{\theta_i} Q(s,a;\theta_i) \right] \tag{28}$$

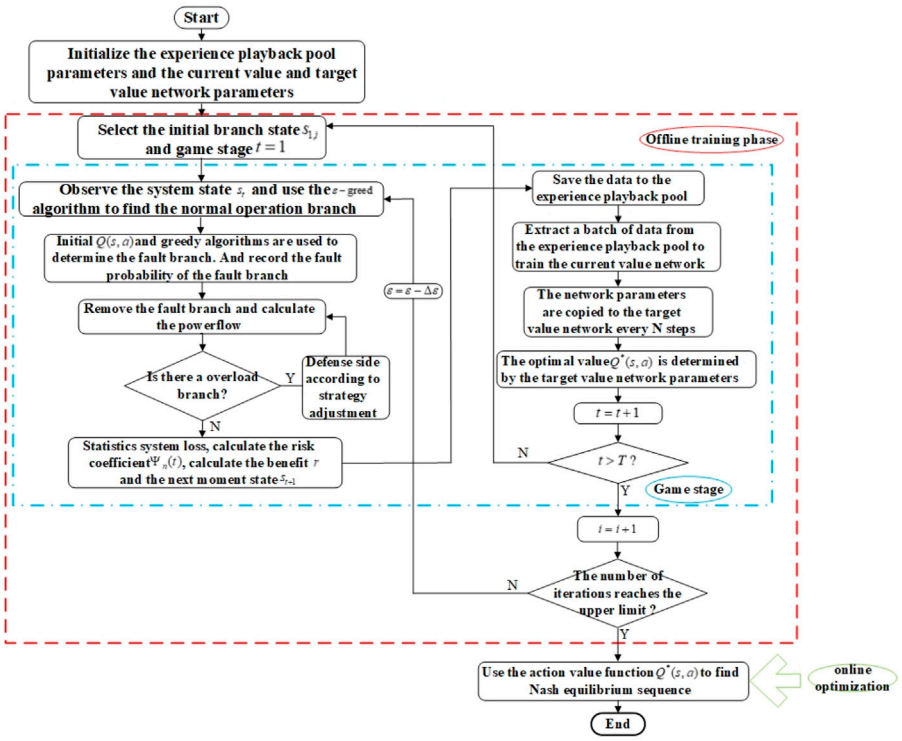



FIGURE 6
Flowchart for determining optimal branch trip sequence of cascading failures.

Hence, the DQN algorithm updates the formula as shown in Eq. 29.

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left[r_t + \gamma \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)\right] \tag{29}$$

(2) Experience-replay mechanism

The interaction information between the agent and environment at each decision moment can be represented as one experience $= (s_t, a_t, r_t, s_{t+1})$ . All experiences are stored in the sequence $D = \{e_1, e_2, \ldots, e_N\}$ to establish the experience recall mechanism. DQN modifies the Q-learning algorithm in two main aspects: DQN uses a deep convolutional neural network to approximate the Q-value function, and DQN utilizes the experience playback mechanism to train the learning process of reinforcement learning. The operation structure of the specific algorithm is shown in Figure 4.

Here, we adopted the $\varepsilon - greed$ algorithm to modify the algorithm action probability, as shown in Eq. 30.

$$\begin{cases} P_{ss'}\left(a_t = a^*|s_t\right) = 1 - \varepsilon \\ P_{ss'}\left(a_t \neq a^*|s_t\right) = \dfrac{\varepsilon}{|A| - 1} \end{cases} \tag{30}$$

where $|A|$ indicates the number of actions that can be selected.

As the algorithm interacts with the environment, it is possible to make $\varepsilon$ decrease over time using the step size $\Delta\varepsilon$ , all the way down to the initial set value of the algorithm.

(3) DQN algorithm training process.

Thus, the training process of the DQN algorithm is as follows.

1) First, initialize the current value network $Q(s,a)$ and the target value network $\hat{Q}(s,a)$;
2) Obtain the grid state based on the parameters of the AC/DC hybrid system $s_t = \{s_{1,t}, s_{2,t}, \cdots s_{N,t}\}$.
3) During the algorithm's training, the action network is responsible for interacting with the environment to obtain the action $a_t$ under the state $s_t$ according to policy selection.
4) During the learning process, after selecting the action $a_t = \{a_t^1, a_t^2\}$, the state of the grid changes, i.e., $s_t \rightarrow s_{t+1}$ and gains are made $r_t^1$.
5) Save the reward $r_t$ and system status $s_t$ to the experience replay pool and train the current value network by extracting a batch of data from the experience replay pool. Whenever the training reaches N steps, the parameters of the current value network data are copied to the target value network to update the target value network parameters.
6) At this time, to increase the number of games while judging whether the number of attacks reaches the maximum number of games, if the maximum number of games, stop iteration, initialize the grid state, return to step 2), and continue the algorithm training learning.

When the algorithm reaches the initially set maximum number of iterations, the algorithm training stops representing the end of

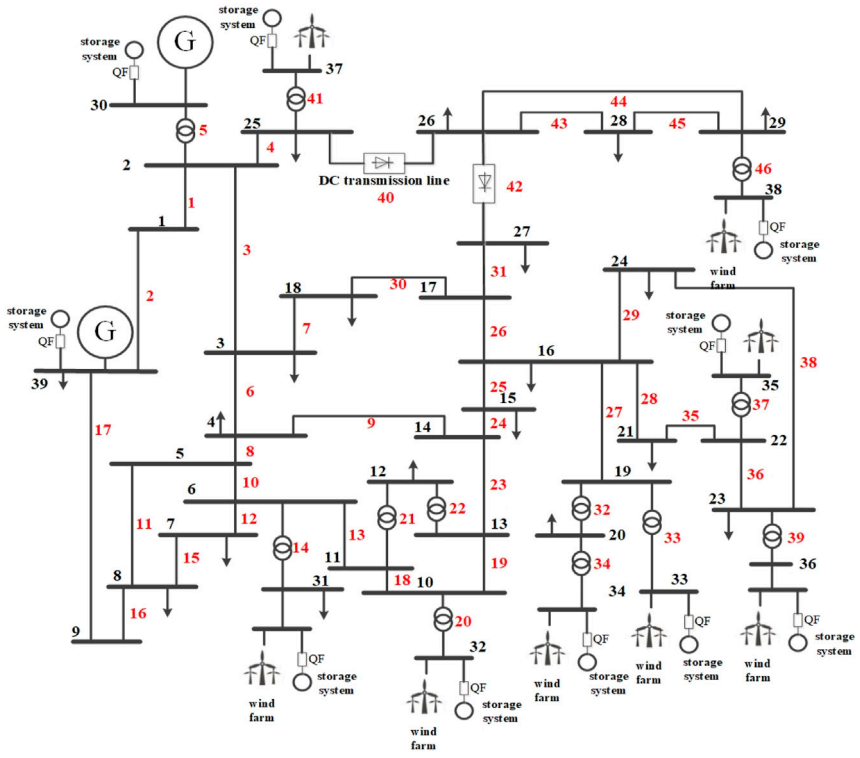



FIGURE 7
Example system wiring diagram.

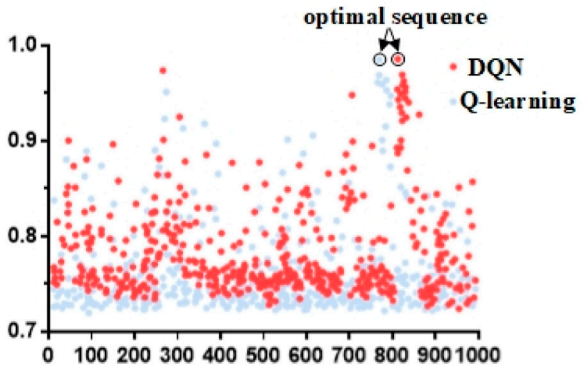



FIGURE 8
Q-table comparison of two algorithms.

learning. Through algorithm training and learning, the target value network parameters are continuously updated, and the optimal action value function $Q^*(s,a)$ is determined based on the final network parameters, as shown in Figure 5. The agent uses the optimal action value function $Q^*(s,a)$ to select the optimal strategy to obtain the maximum benefit.

Through the reinforcement learning algorithm, the action value function $Q(s,a)$ will gradually converge to the action value function $Q^*(s,a)$ under the optimal policy. When the training of the algorithm is completed, the agent can obtain the optimal action value as the next action when the grid is in any state $s$, i.e., the optimal strategy can be obtained to achieve the Nash equilibrium of the game model.

## 3.3 Optimal defense strategy for AC/DC grid interlocking faults

The algorithm of the deep reinforcement learning game model for optimal defense strategy in chain failures comprises learning training using the DQN algorithm to obtain the optimal action value function $Q^*(s,a)$ to obtain the optimal line tripping sequence, and online optimization search using the optimal action value function $Q^*(s,a)$ to obtain the optimal regulation strategy. First, the initial parameters of the network associated with the DQN algorithm and the state of the hybrid system are initialized, and the target value network parameters are updated after a training phase by mutual gaming between the attackers and defenders. At the final end of the training, the optimal action value function $Q^*(s,a)$ is determined, and the attacker takes the maximum action $a_t$ as the attack target to form the optimal line-tripping sequence for the attacker. The procedure for determining the optimal tripping sequence is shown in Figure 6.

In the specific optimal branch trip sequence finding process, the attacker uses the $\varepsilon - greed$ algorithm to select the action after the grid tide calculation, obtains the relevant indicators mentioned above based on the tide calculation results, and uses these indicators as the relevant reference basis to make corresponding

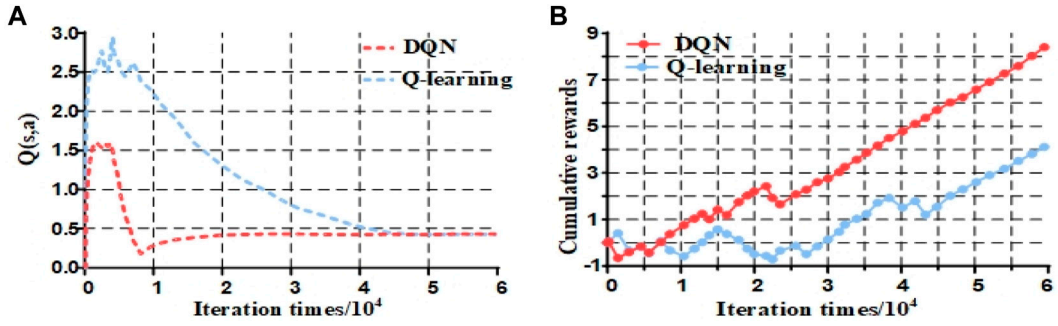



FIGURE 9
Comparison diagram of intelligent algorithms.

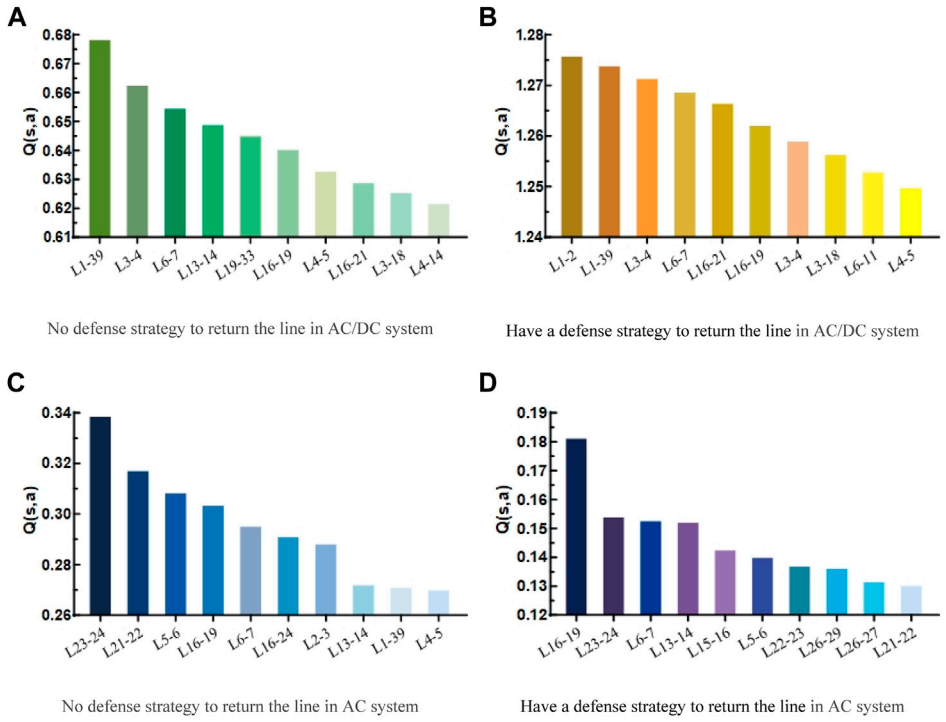



FIGURE 10
Return circuit diagram of system when T = 2.

adjustments according to the established storage source network load regulation strategy to ensure stable grid operation, and then obtains the risk-benefit function $r_t^1 = \Psi_n(t)$, and uses the function to calculate the benefit. Subsequently, the state is used as the empirical replay pool data to update the current value network and target value network parameters. The next stage of the algorithm training is judged according to the initial set number of gaming stages, and the parameters of the $\varepsilon - greed$ algorithm are updated.

After all the training is completed, we enter the online optimization phase, in which we quickly complete the online optimization process of the optimal tripping sequence and optimal regulation strategy.

In summary, this study adopts the AC system risk index $\delta_1$, short-circuit ratio variation $\delta_2$, and commutation bus voltage assessment index $\delta_3$ to establish the interlocking fault risk assessment coefficient R of the hybrid AC/DC system. It uses the deep reinforcement learning DQN algorithm to solve for the line with the largest risk assessment coefficient in the hybrid grid, that is, the most hazardous line in the case of interlocking faults, to establish the optimal defense strategy for interlocking faults in the hybrid AC/DC system. Owing to the characteristics of the DQN algorithm, the optimal action value function $Q^*(s,a)$ is positively correlated with the risk assessment coefficient R of interlocking faults in the hybrid AC/DC system; therefore, the optimal defense strategy is established

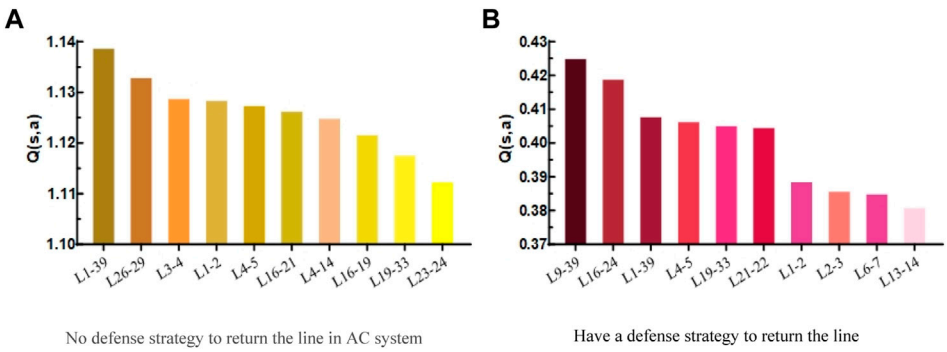



FIGURE 11
Circuit diagram of interlocking fault return when T = 3 in AC/DC system.

in the subsequent simulation by determining the value of the optimal action value function $Q^*(s,a)$.

# 4 Simulation example verification

## 4.1 Introduction to the simulation system

The above theory was simulated and analyzed using a modified IEEE39 node system with the specific wiring diagram shown in Figure 7. Among them, based on the original standard IEEE39 node AC system, the generating units on Buses 31–38 are changed to wind farm units with adjustable wind power output, and the adjacent transmission lines between Buses 25 and 26 and Buses 26–27 are modified as DC transmission systems, while the corresponding capacity of energy storage systems are configured at Buses 30–39, such that the original conventional AC system is changed to a hybrid AC/DC system with new energy access.

## 4.2 DQN algorithm training results

When using the DQN algorithm for training, the relevant algorithm parameters were initialized as follows: = 0.9 , the total experience pool was 10,000; the initial setting of $\varepsilon$ was 0.9; the initial termination value of $\varepsilon$ was 0.1, and the step size of $\Delta\varepsilon$ was 0.0001. Because it is necessary to simulate the complete process of system destabilization caused by a chain fault in a hybrid AC/DC system, the initial setting of two-line decommissioning, and the corresponding settings of two- and three-game phases, to verify the effect of the game model.

First, the DQN algorithm is trained. The two algorithm models with or without defense strategy are analyzed separately for comparison, while two or three game stages are taken for learning training. At the end of the algorithm learning training, the optimal action value function $Q^*(s,a)$ is determined to form the corresponding Q-table. The Q-table obtained from the training of the DQN algorithm and the Q-learning algorithm is compared, as shown in Figure 8.

As shown in Figure 8, the Q-learning algorithm produces cumulative Q values mostly in the lower position. In contrast, the DQN algorithm produces a larger fraction of Q values closer to the optimal sequence, indicating that the DQN algorithm exhibits better convergence.

To verify the performance of the DQN algorithm proposed here, it was compared and analyzed with the Q-learning algorithm. First, the convergence of the Q-value change curves of the DQN algorithm and the Q-learning algorithm are compared, and the two are compared in terms of Q-value estimation. The comparison in terms of Q-value change trends shows the advantages of the DQN algorithm proposed here in the offline training process. Figure 9A shows a comparison of the change trends of the Q-value under the two algorithms. Figure 9A shows that the Q-learning algorithm estimates the Q-value from a higher starting point under the same number of iterations. The DQN algorithm improves the overestimation of the Q-value caused by the problem of increasing the dimensionality of the data obtained from the tide calculation after the AC/DC hybrid system is added to the DC system owing to the optimization of the objective function.

The algorithm's convergence was verified by storing the cumulative gains obtained after each gaming phase; the results are shown in Figure 9B. The Q-learning algorithm cannot determine the correct action at the beginning of training. It only starts to find the correct action after the number of training iterations reaches 30,000. However, there are still some fluctuations between, and only after approximately 45,000 iterations are fully determined and continue to increase. In contrast, the DQN algorithm kept fluctuating and rising at the beginning of training, even though it was fumbling to find the correct action, and then selected the correct action to obtain a positive reward and kept rising linearly for approximately 25,000 iterations, indicating that the algorithm found a suitable control strategy to complete the convergence of the algorithm.

## 4.3 Analysis of online optimization search results

### 4.3.1 Comparison of AC/DC hybrid system and conventional AC system

After causing a chain failure in the system according to the initially set attack sequence $L_{8-9}$, $L_{9-39}$ , the next fault-line sequence

**TABLE 1 Defense strategy with energy storage adjustment.**

| Gaming phase | Attacker action | Defensive side action | *Q* value |
|---|---|---|---|
| 1 | $L_{8-9}, L_{9-39}$ | The system is not overloaded, and no policy is taken | - |
| 2 | $L_{16-24}$ | An overload occurs on line $L_{16-17}$, whose distance protection section III action angle is reduced by 9.67°, raising the stored energy output at node 36 by 357 F | 0.4189 |
| 3 | $L_{16-17}$ | Line $L_{17-27}$, $L_{17-18}$ overload, line $L_{16-17}$ distance protection section III action angle is reduced by 10.52°. Simultaneously, the output of energy storage system at node 30 is reduced by 102 F; the output of energy storage system at node 31 is increased by 69.36 F, the output of energy storage system at node 32 is increased by 76.5 F; the output of energy storage system at node 33 is increased by 20.4 F; the output of energy storage system at node 37 is increased by 24.48 F; the output of energy storage system at node 38 is increased by 173.4 F, and the output of energy storage system at node 39 is increased by output by 102 F | 0.5667 |

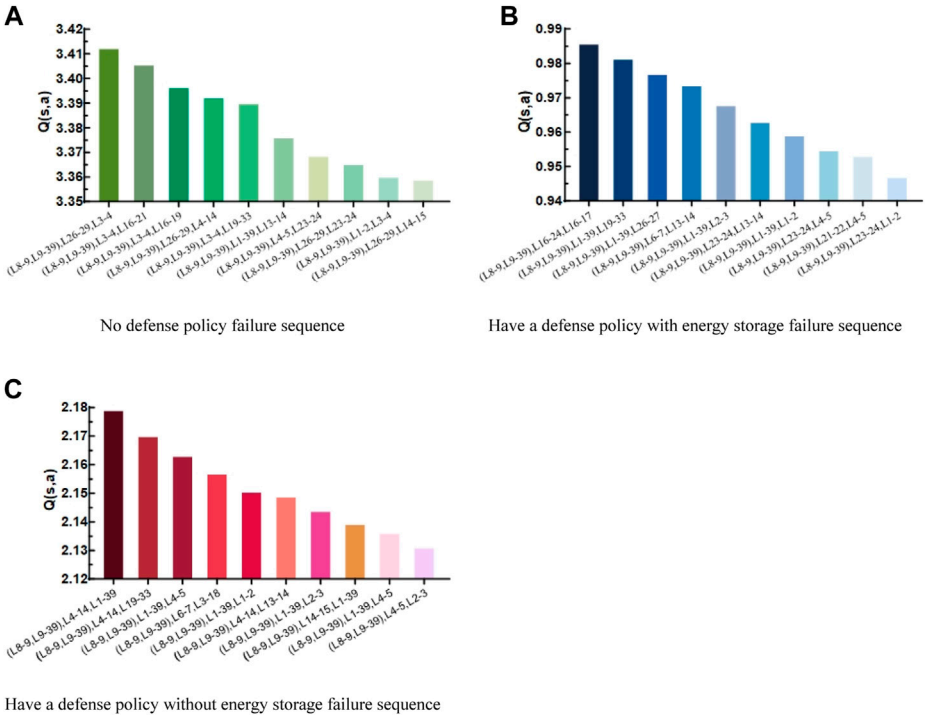



FIGURE 12
Interlocking fault trip sequence diagram with or without energy storage adjustment defense strategy.

is searched according to the Q-table determined by the optimal action-value function $Q^*(s,a)$ . Only one line failure is considered when there are only two gaming phases because the initial failure has already occurred; therefore, the subsequent fault lines are sorted by risk size, as shown in Figures 10A,B.

The analysis of the faulty lines in Figure 10 shows a large part of duplication in the lines obtained by the search with or without the defines strategy. According to Figure 7, the fault lines mentioned above are lines around $L_{8-9}, L_{9-39}$ and contact lines. Lines such as $L_{6-7}, L_{1-39}$ are affected by lines, $L_{8-9}L_{9-39}$ , *etc.*, and lines such as $L_{16-19}, L_{13-14}, L_{3-4}$ are important liaison lines that connect the upper and lower systems. As a comparison, the risk-ranking diagram of the subsequent fault lines of the AC system is drawn, as shown in Figures 10C,D.

The lines found are roughly the same whether in the AC system or the hybrid AC/DC system with the addition of DC lines and new energy generating units. However, owing to the influence of the new energy generating units and DC lines, the impact of the AC/DC system produces a larger Q value when a line closer to the DC system is decommissioned. Similarly, owing to the influence of the DC system, the risk caused by the failure of the system to produce a decommissioned line was greater than that of a pure AC system.

When three game phases are used, the attacker will cause multiple line failures in the hybrid system after attacking multiple phases, thus posing a significant threat to the grid. The risk ranking of the top ten ranked subsequent failed lines with and without the defense strategy is plotted in Figure 11.

**TABLE 2 Defense strategy without energy storage adjustment.**

| Gaming phase | Attacker action | Defensive side action | $Q$ value |
|---|---|---|---|
| 1 | $L_{8-9}, L_{9-39}$ | The system is not overloaded, and no policy is taken | - |
| 2 | $L_{4-14}$ | Line $L_{4-5}$ is overloaded, and its distance protection section III action angle is reduced by 14.84°, reducing the output of the generating unit at node 31 by 620 MW and increasing the output of the generating unit at node 35 by 620 MW. | 0.8596 |
| 3 | $L_{1-39}$ | Line $L_{2-3}$, $L_{3-4}$ overload, line $L_{4-5}$ distance protection section III action angle is reduced by 18.78°. Simultaneously, the output of generator set at node 30 is reduced by 300 MW; the output of generator set at node 31 is increased by 184 MW; the output of generator set at node 32 is increased by 225 MW; the output of generator set at node 33 is increased by 60 MW; the output of generator set at node 37 is increased by 72 MW; the output of generator set at node 38 is increased by 300MW; the output of generator set at node 39 is increased by 240 MW, and reduce the load power at node 3 by 775 MW. | 1.3193 |

**TABLE 3 Methods comparison.**

| | Fault sequence | Load loss (MW) | Risk coefficient income |
|---|---|---|---|
| risk coefficient ranking method | $,(L_{8-9}, L_{9-39}), L_{26-29}, L_{4-14}$ | 1768.6 | 0.6445 |
| The algorithm in the manuscript - no defense | $(L_{8-9}, L_{9-39}), L_{26-29}, L_{3-4}$ | 2098 | 0.7641 |
| The algorithm in the manuscript - with defense | $(L_{8-9}, L_{9-39}), L_{16-24}, L_{16-17}$ | 435 | 0.001942 |

Because the optimal trip sequence needs to be searched by the fault gain function and the cumulative action value $Q(s,a)$, the plot of the chain fault trip sequence with and without the defense strategy with cumulative $Q(s,a)$ ranking in the top 10 is shown in Figures 12A,B.

An analysis of Figure 12 shows a significant difference between the cumulative $Q(s,a)$ with and without defensive strategy, that is, the cumulative $Q(s,a)$ without defensive strategy is significantly higher than the cumulative $Q(s,a)$ with defensive strategy, which shows that the defensive strategy may help reduce the risk of the grid. In addition, the optimal attack sequence of the attacker can be determined by searching for sequences with a higher cumulative $(s,a)$. The analysis of the above figure shows a significant difference between the line fault sequences with and without the defense strategy, mainly because the regulation strategy adjusts the power of each node to change the tide, which in turn changes the high-risk fault sequence, thus creating the difference between the two.

The cumulative benefit analysis shows that the attacker's benefit is significantly higher without the defender's participation than with the defender's participation because the defender mainly aims to reduce grid losses. However, even if the defender adopts the optimal regulation strategy, the attacker still poses a greater risk to the system because the attackers are attacking the main system contact lines. Therefore, a large amount of energy storage system output must be regulated to mitigate the risk.

Using the above fault sequence $(L_{8-9}, L_{9-39}), L_{16-24}, L_{16-17}$ as the attacker's strategy, the optimal regulation strategy for the defender is prepared, as shown in Table 1. Analyzing the data in the table, when the initial fault occurs, there is no tidal overload in the system. However, because $L_{16-24}L_{16-17}$ is the central contact line of the hybrid system and is mainly responsible for connecting the DC transmission system with most of the generating units, the attacker prefers to attack these two lines. At the same time, the defender also adopts strategies such as distance protection and output adjustment of the energy storage system at each node to ensure the normal operation of the grid.

### 4.3.2 Comparison of energy storage system out of power participation or not

To consider the case of system failure when energy storage system capacity adjustment is used as the defense strategy here, the energy storage system output adjustment part of the original overall defense strategy is removed, and only the original AC side conventional defense strategy is retained before and after comparison. The chain fault trip sequence without energy storage adjustment defense strategy is plotted with the cumulative $Q(s,a)$ ranking of the top 10, as shown in Figure 12C.

Through the overall comparison of Figures 12B,C, it is first found that owing to the addition of new energy-generating units and DC lines in the hybrid AC/DC system, the original conventional AC steady-state defense strategy still plays a role in the chain fault of the AC/DC system. However, although the main role is still focused on the original direct connection line of each sub-grid and the adjustment at the level of conventional generating units, the DC lines and new energy units are not. Hence, it is impossible to take more defensive measures to reduce the final Q value, but there is still a greater risk threat. Second, when the energy storage system output adjustment comes into play, the cumulative Q value decreases significantly, and the risk of chain failures is significantly reduced by the joint action of the AC and DC sides, thus proving the feasibility of the defense strategy of increasing the energy storage system. Finally, comparing specific fault sequences reveals that after the addition of the energy storage system, the attacker of the system focuses more on its attack strategy of dealing with the connection between the nodes where the energy storage system is located after the second stage of the game. In contrast, the overall DC-side

defense strategy makes it possible for each node to adjust individually, thus better targeting the attack strategy and proving the importance of the defense strategy to increase the energy storage system adjustment.

Using the fault sequence ($L_{8-9}, L_{9-39}$), $L_{4-14}$, and $L_{1-39}$ in Figure 12C as the attacker's strategy, the specific defense strategy of the defender is prepared, as shown in Table 2. When the energy storage adjustment is no longer part of the defense strategy, the attacker's action focuses more on the contact lines between the nodes where each unit is located. The main role of the conventional AC defense strategy is still focused on the original direct connection lines of each sub-grid and the adjustment at the level of conventional generating units. No more defense measures can be taken for DC lines and new energy units, and only the conventional measures of increasing generator output and load shedding can be taken to cope with them. However, owing to the lack of defensive measures on the DC side, the chain failure will further increase the damage to the system when the next game phase is initiated, resulting in a greater adjustment of generator output and more load loss compared to the impact of energy storage system adjustment, which is extremely harmful to the grid. This also proves the importance of increasing the defense strategy of energy storage system adjustment.

The above simulation results verify that the energy storage system adjustment as a defense strategy can be fast and efficient for interlocking faults in the hybrid AC/DC system. Because the energy storage system has a millisecond power response speed, it can effectively improve the resilience and flexibility of the hybrid AC/DC system, which can in turn minimize the damage caused and avoid the sudden load-cutting action of the system that would have otherwise caused greater losses on the user side. Similarly, the energy storage system can also smooth out the intermittent and fluctuating power generated by new energy-generating units, such as wind power and photovoltaics, during normal operation, which is conducive to new energy consumption.

### 4.3.3 Comparison of different methods

To show the advantages of the chain failure game model based on the deep reinforcement learning DQN algorithm proposed in this paper for obtaining the chain failure trip sequence and the optimal defense strategy, the simulation system in this paper is taken as an example, and compared with the traditional risk coefficient ranking method, and the results are shown in Table 3.

It can be found from the table that the fault sequence found by the traditional risk ranking method is not the optimal sequence, which is only the fourth in the previous Figure 12A. The main reason is that the traditional risk ranking is to select the line with the highest risk as the attack target at each stage of cascading fault, which only considers the loss caused by each stage to the power grid, but does not consider the impact of the fault sequence on the power grid as a whole, It is easy to fall into local optimum; The model algorithm proposed in this paper focuses on the impact of a fault sequence on the power grid as a whole, and focuses on the global optimal defense strategy.

From the perspective of loss, the load loss is caused by the fault sequence found by the risk ranking method and the risk return is lower than the fault sequence in this paper, which shows that the algorithm in this paper can find the sequence that makes the maximum return. In the game with defense, the development direction of the fault sequence is changed, and the load loss and risk income caused by the fault sequence are far lower than those without defense, but it can find the fault sequence that makes the power system lose, so the model in this paper has certain advantages over the traditional risk ranking method.

## 5 Conclusion

The algorithm here considers the impact of interlocking faults on the AC/DC hybrid system from the perspectives of both steady-state and transient systems, establishes a multi-stage dynamic zero-sum game interlocking fault model by finding the fault sequence through the DQN algorithm, and proposes a corresponding defense strategy to provide reference to grid operation and dispatchers. In summary.

(1) This study proposes a method for searching chain fault-tripping sequences and finding optimal regulation and control strategies for hybrid AC/DC systems based on game deep reinforcement learning algorithms. The method applies the theory of multi-stage zero-sum game to the scheduling and control adjustment of the hybrid AC/DC system. It uses the DQN algorithm to train the optimal action value function to find the most threatening fault line in the complex hybrid grid and the optimal regulation and control defense strategy to reduce the risk of the grid.
(2) Here, a multi-stage zero-sum game chain fault model for an AC/DC hybrid system is proposed, which can completely describe the dynamic process after the chain fault occurs in an AC/DC hybrid grid and the involvement of a regulation strategy. Furthermore, a novel deep reinforcement learning algorithm was used to solve the Nash equilibrium of the game model, which improved the convergence and accuracy of the algorithm.
(3) The multi-feeder short-circuit ratio constraint method and the energy storage system adjustment strategy were used as defense strategies to cope with the faults occurring in the DC system in the hybrid AC/DC system. The energy storage system is fully utilized to improve the fault ride-through capability of new energy units and the rapidity and economy of fault handling. Considering the scheduling of the energy storage system at the whole grid level, it is superior as a defense strategy to cope with interlocking faults in complex AC/DC systems.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Author contributions

The XD wrote the original draft. SW, WW, PY and XX provided the supervision, review, and editing of the draft. All authors contributed to the article and approved the submitted version.

## Funding

This work was Funded by the National Nature Fund (51777119).

## Conflict of interest

WW was employed by the State Grid Wenzhou Power Supply Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Reviewer HF declared a shared affiliation with the author XD, SW at the time of the review.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Cao, Y. J., Chen, Y. R ., Cao, L. H., and Tan, Y. D. (2012). Prospects of studies on application of complex system theory in power systems[J]. *Proc. CSEE* 32 (19), 1–9+178. doi:10.13334/j.0258-8013.pcsee.2012.19.001

Cao, Y. J., Wang, G. Z., Cao, L. H., and Ding, L. J. (2011). An identification model for self-organized criticality of power grids based on power flow entropy[J]. *Automation Electr. Power Syst.* 35 (07), 1–6. CNKI:SUN:DLXT.0.2011-07-002.

Dai, W. C., Dong, Y., Zhao, X. F., Shang, X. J., Gao, K., and Jin, P. (2016). Allocated method for capacity of energy storage based on adjustment of SOC[J]. *Acta Energiae Solaris Sin.* 37 (2), 261–268. doi:10.3969/j.issn.0254-0096.2016.02.001

Dey, P., Mehra, R., Kazi, F., Wagh, S., and Singh, N. M. (2016). Impact of topology on the propagation of cascading failure in power grid. *IEEE Trans. Smart Grid* 7 (4), 1970–1978. doi:10.1109/TSG.2016.2558465

Ding, M., Qian, Y. C., and Zhang, J. J. (2017). Multi-timescale cascading failure evolution and risk assessment model[J]. *Proc. CSEE* 37 (20), 5902–5912. doi:10.13334/j.0258-8013.pcsee.162612

Ding, M., Qian, Y. C., Zhang, J. J., He, J., and Yi, J. (2016). Coordinated control model of power system cascading failures based on risk assessment[J]. *Automation Electr. Power Syst.* 40 (07), 1–8. doi:10.7500/AEPS20150522005

Duan, J., Yi, Z., Shi, D., Lin, C., Lu, X., and Wang, Z. (2019). Reinforcement-learning-based optimal control of hybrid energy storage systems in hybrid AC–DC microgrids. *IEEE Trans. Industrial Inf.* 15 (9), 5355–5364. doi:10.1109/TII.2019.2896618

Fan, W. L., Zhang, X. M., Mei, S. W., and Huang, S. W. (2018). Vulnerable transmission line identification considering depth of *K* -shell decomposition in complex grids. *IET Generation, Transm. Distribution* 12 (5), 1137–1144. doi:10.1049/iet-gtd.2017.0767

Fang, Y. (2014). Reflections on stability technology for reducing risk of system collapse due to cascading outages. *J. Mod. Power Syst. Clean Energy* 2 (3), 264–271. doi:10.1007/s40565-014-0067-x

Li, X. J., Ma, H. M., and Jiang, Q. (2022). Review of energy storage configuration technology on renewable energy side[J]. *Electr. Power* 55 (01), 13–25. doi:10.11930/j.issn.1004-9649.202109032

Lin, W. F., Tang, Y., and Bu, G. Q. (2008). Study on voltage stability of multi-infeed HVDC power transmission system[J]. *Power Syst. Technol.* 32 (11), 7–12.

Lin, X. N., Xia, W. L., Xiong, W., Li, Z. T., Liu, X. C., and Liu, J. Z. (2011). Study of adaptive adjustment of operation characteristics of distance backup protection immune to the impact of power flow transferring[J]. *Proc. CSEE* 31 (S1), 83–87. doi:10.13334/j.0258-8013.pcsee.2011.s1.036

Liu, J., Yao, W., Hou, Y. H., Wen, J. Y., and Chen, X. (2016). Stability control for improving the characteristic of wind farm injection current during low voltage ride-through using energy storage system[J]. *Trans. China Electrotech. Soc.*, 31(14):93–103. doi:10.19595/j.cnki.1000-6753.tces.2016.14.011

Liu, S. Q., Gu, J., Lai, B. X., and Jin, Z. J. (2022). Bi-level optimal allocation of energy storage in regional integrated energy system considering load margin[J]. *Electr. Power Autom. Equip.*, 42(07):150–158. doi:10.16081/j.epae.202202022

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 58 (7540), 529–533. doi:10.1038/nature14236

Song, F. L., Wu, Z. Y., and Zhang, Y. (2018). Fuzzy scene clustering based grid-energy storage coordinated planning method with large-scale wind power[J]. *Electr. Power Autom. Equip.* 38 (02), 74–80. doi:10.16081/j.issn.1006-6047.2018.02.010

Tian, L., Zhang, Y., and Li, D. X. (2016). Power-type storage system capacity configuration and control strategy for enhancing LVRT ability[J]. *Electr. Power Constr.* 37 (8), 84–89. doi:10.3969/j.issn.1000-7229.2016.08.013

Van Hasselt, H., Guez, A., and Silver, D. (2016). "Deep reinforcement learning with double Q-learning[C]," in Proceedings of the Thirtieth AAAI conference on artificial intelligence, 12 February 2016, 2094–2100. doi:10.48550/arXiv.1509.06461

Wang, T., Liu, Y. M., Gu, X. P., and Qin, X. H. (2019). Vulnerable lines identification of power grid based on cascading fault space-time graph[J]. *Proc. CSEE* 39 (20), 5962–5972+6176. doi:10.13334/j.0258-8013.pcsee.181730

Xu, L., Wang, X. L., and Wang, X. F. (2010). Electric betweenness and its application in vulnerable line identification in power system[J]. *Proc. CSEE* 30 (01), 33–39. doi:10.13334/j.0258-8013.pcsee.2010.01.007

Xu, Y., Zhi, J., and Fan, S. T. (2017). Line overload emergency control based on power sensitivity and minimized economic compensation[J]. *Electr. Power Autom. Equip.* 37 (1), 118–123. doi:10.16081/j.issn.1006-6047.2017.01.019

Yan, G. G., Zhu, W., and Duan, S. M. (2020). Power control strategy of energy storage system considering consistency of lead carbon battery pack[J]. *Automation Electr. Power Syst.* 44 (11), 61–67. doi:10.7500/AEPS20190911004

Yang, L., Sun, Q., Zhang, N., and Li, Y. (2022). Indirect multi-energy transactions of energy Internet with deep reinforcement learning approach. *IEEE Trans. Power Syst.* 37 (5), 4067–4077. doi:10.1109/TPWRS.2022.3142969

Yang, W. H., Bi, T. S., Xue, A. C., Huang, S. F., and Yang, Q. X. (2011). Adaptive strategy for back up protections within power transferring area against cascading trips [J]. *Automation Electr. Power Syst.* 35 (18), 1–6.

Ying, Y., Liu, Q., Wu, M., and Zhai, Y. (2023). Online energy management strategy of the flexible Smart traction power supply system. *IEEE Trans. Transp. Electrification* 9 (1), 981–994. doi:10.1109/TTE.2022.3192141

Zhang, J., Tong, X. Y., and Jiang, J. W. (2017). Analysis on power system cascading failure based on percolation and risk theory[J]. *Automation Electr. Power Syst.* 41 (5), 46–52. doi:10.7500/AEPS20160515018

Zhang, Z. M., Huang, S. W., Mei, S. W., Zhang, X. M., and Jiang, Y. F. (2020). Vulnerability assessment method of branch lines in power grid based on cooperative game[J]. *Automation Electr. Power Syst.* 44 (06), 9–16. doi:10.7500/AEPS20190626005

# Frontiers in Energy Research

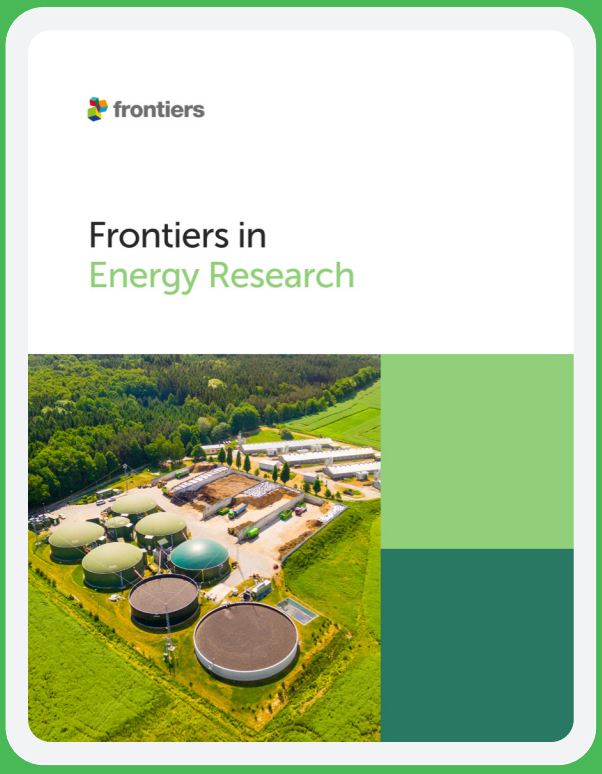
**Advances and innovation in sustainable, reliable and affordable energy**

Explores sustainable and environmental developments in energy. It focuses on technological advances supporting Sustainable Development Goal 7: access to affordable, reliable, sustainable and modern energy for all.

---

## Discover the latest Research Topics

See more →




**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact