International Journal of Power and Energy Systems, Vol. 46, No. 10, 2026
An Open Access Paper

INTELLIGENT METER DATA PROTECTION VIA DIFFERENTIAL PRIVACY AND TOP-K QUERY ALGORITHM

Peiyu Chen,* Zhaohui Hu,** and Qianjun Tu***

Abstract

To address the privacy risks of high-frequency and fine-grained electricity data collected by smart meters, which may expose user behavior patterns and lifestyle habits, this study proposes a smart meter data protection model that integrates K-modes clustering–based shuffled differential privacy with a time auto-wave neural network (TANN) for TOP-K queries. In the data perturbation stage, the model allocates privacy budgets efficiently through frequency prediction and a gradient random response mechanism. In the query stage, it employs temporal dependency modeling within the TANN structure to enhance the real-time capability and accuracy of TOP-K queries. Experimental results show that the proposed model achieves a normalized intra-cluster variance of 0.23, an F1 score of 0.976, and a root mean square error of 0.155, indicating superior clustering performance. The TOP-K query time is only 4.6 seconds, with a mean absolute error of 4.5% and a re-identification rate of 7.2%, both significantly lower than those of the three comparison models. These results demonstrate that the proposed approach effectively enhances both the privacy and availability of smart meter data while maintaining high query accuracy and strong resistance to inference attacks, offering a practical solution for smart power data privacy protection.

Key Words

Differential privacy; TOP-K query; Time automatic wave neural network; Frequency prediction; K-modes clustering

* Zhejiang College of Security Technology; e-mail: PeiyuCh@outlook.com
** Zhejinag Oumeilong Meter Co, Ltd., Yueqing, 325600, China; e-mail: huzhaohuidz@163.com
*** Zhejinag Oumeilong Meter Co, Ltd., Yueqing, 325600, China; e-mail: tqjn@163.com
Corresponding author: Peiyu Chen

Recommended by: B Rajanarayan Prusty
(DOI:10.2316/J.2026.203-0637)

1. Introduction

With the widespread application of smart grids, smart meters play a crucial role in detailed energy management and real-time data collection [1]. However, the high-frequency, fine-grained data they collect can leak user behavior patterns and living habits, leading to serious privacy risks [2]-[4]. Therefore, how to ensure effective privacy protection while maintaining data usability has become a key issue in the field of smart meter data management. In recent years, Differential Privacy (DP) has received widespread attention as a core technology for ensuring data release security. By perturbing the original data, it effectively prevents individual information from being reverse-engineered [5]. However, traditional DP methods often degrade the utility of data when dealing with high-dimensional, strong temporal data from smart meters, and can lead to accuracy loss in tasks such as TOP-K queries [6]. Existing studies are mainly based on the Laplace mechanism and the exponential mechanism. While these methods have achieved certain results with static data or simple queries, they still face issues such as low accuracy and privacy budget waste when processing multidimensional temporal data, complex clustering, or TOP-K retrieval tasks [7]. The Shuffled Differential Privacy (SDP) strategy of the K-modes clustering algorithm can enhance data indistinguishability and improve privacy protection strength. When combined with the Temporal Auto-Wave Neural Network (TANN) for TOP-K queries, it exhibits excellent temporal data awareness, effectively mitigating privacy leakage risks while ensuring query accuracy. Therefore, this study proposes a smart meter data protection framework that combines K-modes clustering-based DP perturbation with neural network-based TOP-K queries. The goal is to balance privacy protection and data usability, creating an efficient data protection model for smart meter environments. The main contributions of the study are as follows:

(1) A smart meter data protection framework that integrates K-modes clustering, shuffle differential privacy and TANNTOP-K is proposed. The framework

achieves joint optimization of data perturbation and query mechanism through algorithm-level collaborative fusion, taking into account both privacy protection strength and data availability.

(2) A shuffle differential privacy strategy based on frequency prediction and gradient random response is designed. By introducing frequency aggregation and random response mechanism in the perturbation stage, the utilization efficiency of the privacy budget is improved, and the stability and interpretability of the clustering results are maintained.

(3) A TANN model with time perception and fluctuation propagation characteristics is constructed. Through time window adjustment and fluctuation propagation mechanism, the model effectively captures the temporal dependence characteristics of high-frequency meter data to improve the real-time performance and accuracy of TOP-K query, providing new ideas for the design and application of subsequent privacy protection technologies.

2. Related Works

DP has become one of the key directions in current data privacy research, as it protects against individual information leakage. Scholars both domestically and internationally have conducted extensive studies on DP. For example, Hu and others proposed a protection method based on federated learning and differential privacy to safeguard sensitive information in training data of federated learning models. They applied sparse perturbation for local sparsification and then used Gaussian noise to further increase the perturbation, improving model confidentiality and accuracy [8]. Zhang et al., addressing the issue that traditional privacy algorithms overlook the freshness of data in privacy protection, proposed an age-related differential privacy protection framework. This framework characterizes data obsolescence and the relationship between time-sensitive data, using aging data as a new strategy for data privacy protection [9]. Huang et al., aiming to solve the problem that direct transmission of gradient information increases the risk of privacy leakage, proposed a privacy protection method based on gradient tracking. They added noise to the transmitted information and analyzed the convergence performance of the step size sequence to optimize the privacy protection algorithm [10]. TOP-K queries, which select the top K data items based on scores, are widely used in querying large-scale sensitive data. For instance, Zhu and others, addressing the challenge of data leakage in network management where shared data is vulnerable, proposed a differential privacy-based local protection mechanism. This mechanism searches for TOP-K flows across multiple independent clients and uses iterative approximation methods to reduce computational costs, ensuring the efficiency and practicality of the query method [11]. Xu et al., in response to the challenge of achieving efficient multi-user encrypted searches in the Internet of Things data interaction, proposed a privacy-preserving dynamic multi-keyword search scheme based on encrypted cloud data. They queried TOP-K using a specific search structure and employed a greedy breadth-first search algorithm to achieve sub-linear search, ensuring privacy protection in the search mode [12]. Kara and Eyüpoğlu proposed an improved data anonymization algorithm that integrates an anomaly detection mechanism to solve the problem that the existing k-anonymity algorithm is difficult to deal with abnormal data interference, resulting in an imbalance between privacy protection and data utility. This method introduces an anomaly factor algorithm based on connectivity to identify outliers in high-dimensional complex data sets, and optimizes the partitioning strategy based on this to generate a more balanced equivalence class structure [13]. In addition, for privacy enhancement mechanisms in distributed environments, Li et al. proposed a decentralized privacy-enhancing federated learning framework that is resistant to poisoning attacks. By combining local differential perturbations with cluster shuffling strategies, they achieved data security sharing and aggregation optimization under privacy budget constraints [14]. This "shuffling + clustering" mechanism provides direct theoretical inspiration for the K-modes cluster shuffling differential privacy proposed in the study. At the same time, Guo et al. proposed a contextual knowledge-enhanced neural network model in the study of speech recognition in air traffic control communications, which significantly improved the recognition accuracy by using dynamic propagation mechanisms and temporal dependency modeling [15]. This idea provides an important reference for the time-series modeling and TOP-K query optimization of high-frequency meter data using the time-series automatic wave neural network (TANN) designed by the institute.

As a critical terminal in smart grids, smart meters have enhanced energy management but also brought significant privacy risks. In response, many scholars worldwide have conducted in-depth research. For example, Yan et al., addressing the privacy leakage problem when aggregating data from multiple smart meters, proposed a differential privacy-based encryption scheme. This scheme uses an improved homomorphic encryption method for data aggregation and employs a dual-noise distributed technique to prevent data theft, thus protecting electricity usage data privacy [16]. Singhal et al., addressing the risk of data leakage in large amounts of consumer data accumulated on smart meters, proposed a smart meter data verification method based on blockchain technology. They enhanced user-side security with blockchain diversity and improved blockchain storage performance through data pruning techniques, achieving privacy protection [17]. Wang et al., facing privacy protection challenges in electricity theft detection by distribution system operators, proposed a decentralized federated learning framework. This framework uses threshold homomorphic encryption for serverless parameter aggregation and employs a decentralized federated extreme gradient boosting model to enhance performance and ensure privacy protection [18]. Singh and Kumar, addressing security and privacy issues in smart meter data, proposed a security and privacy-preserving data aggregation and classification model based on fog and cloud architecture. They completed data aggregation using fog nodes and outsourced classification using three machine learning classifiers in the cloud, achieving privacy protection [19].

In conclusion, existing research has made significant progress in the field of smart meter data privacy protection, covering multiple directions. However, for high-frequency, fine-grained time-series meter data, balancing usability and privacy under complex query tasks remains a challenge. Therefore, this study proposes a framework that integrates K-modes clustering-based SDP with TANN and TOP-K queries. The aim is to address the shortcomings of existing methods in balancing precise queries and privacy protection, providing more efficient and practical technical support for data privacy protection in smart meter environments.

The research's main contributions and framework are divided into four parts. The first part introduces the research background and relevant literature, analyzes the current status of smart meter data privacy protection, and points out that traditional differential privacy algorithms suffer from reduced accuracy and privacy budget waste when processing high-dimensional, strongly time-series data. The paper then proposes a smart meter data protection framework that integrates the K-modes clustering shuffle differential privacy algorithm with the time-autonomous wave neural network top-K query, balancing privacy protection and data availability. The second part designs the K-modes clustering shuffle differential privacy algorithm and constructs the TANNTOP-K algorithm, elaborating on its core mechanisms and advantages. The K-modes clustering shuffle differential privacy algorithm reduces the noise amplification effect through cluster compression and shuffle randomization, enhancing data anonymity and perturbation stability. TANNTOP-K utilizes time series feature learning to optimize top-K queries for high-frequency meter data. The third part conducts experimental verification on a typical meter dataset, examining cluster consistency, privacy budget sensitivity, and top-K query performance. The results demonstrate that the proposed model effectively balances privacy strength and query accuracy, and the overall framework outperforms existing algorithms in terms of accuracy, robustness, and scalability. The fourth part discusses and summarizes the experimental results, points out the shortcomings of the model in parameter adaptability, and looks forward to combining federated learning with an adaptive privacy budget mechanism in the future to further improve the generalization and practicality of the model in dynamic power data scenarios.

3. Data Protection Research Combining DP and TOP-K Query Algorithms

3.1 K-Modes Clustering SDP Algorithm for Meter Data Protection

With the accelerated deployment of smart grids, smart meters accumulate massive electricity usage data in billing, scheduling, optimization, and value-added services, supporting the intelligent transformation of the grid and enhancing user experience [20]. However, the increase in data value also brings privacy leakage risks. Compared to traditional DP mechanisms that rely on centralized processing and trusted third parties, SDP weakens the risk of single-point leakage by introducing a shuffling step, improving the robustness of privacy protection [21]. The data protection architecture for smart grid data based on this mechanism is shown in Figure 1.

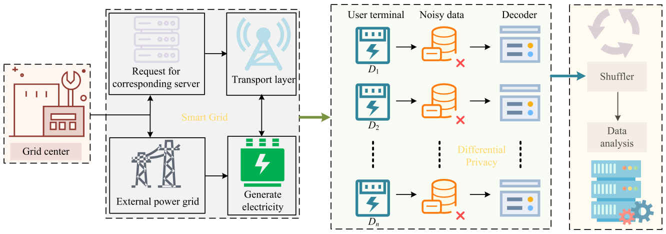


Figure 1. Data Protection Architecture Based on SDP for Smart Grids

As shown in Figure 1, the grid center first aggregates regional electricity data and coordinates resource scheduling. The distribution network and communication layer work together to achieve efficient resource allocation between nodes. After the energy is transmitted to the user side, smart meters collect electricity usage information in real time. To enhance privacy, user data is locally processed with differential perturbation before being uploaded to the shuffling node, where random rearrangement breaks the data correlation, thereby meeting the centralized differential privacy requirements. This shuffling step, that is, the core process of the SDP mechanism, mainly includes local perturbation, random shuffling, and aggregation calculation. First, the user side locally perturbs the original meter data x and adds a noise term η where the noise obeys the Laplace distribution $Lap\left(\frac{\Delta f}{\varepsilon}\right)$ and the amplitude is determined by the function sensitivity Δf and the privacy budget ε. Secondly, the shuffling stage performs a random scrambling operation on the perturbed data set, making it difficult to distinguish between any two adjacent user data in the output space, thereby reducing the risk of single-point leakage. Finally, the aggregation stage uses frequency prediction and a gradient random response mechanism to calculate the global statistics of the perturbed data to achieve data reconstruction and feature aggregation under differential privacy conditions. Specifically, A mechanism M is preset to satisfy (ε, δ) -SDP. For any two adjacent user data sets D and D', the output results satisfy the differential privacy condition $S \subseteq \text{Range(M)}$, as shown in Equation (1).

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr\left[M\left(D'\right) \in S\right] + \delta \tag{1}$$

In Equation (1), S represents the set of output results after random shuffling of n users, ε is the privacy budget, and δ is the probability that the output results of adjacent data sets D and D' do not satisfy DP. Then, noise is added to the function $f(D)$, with the noise perturbation shown in

Equation (2) [22].

$$M(D) = f(D) + \operatorname{Lap}\left(\frac{\Delta f}{\varepsilon}\right) \tag{2}$$

In Equation (2), Δf represents the sensitivity of the function (the maximum variation between adjacent data sets). However, due to the rough design of the perturbation probability in the current shuffling differential mechanism, the algorithm results may fluctuate. Therefore, the study introduces a Frequency-based Shuffling Differential Privacy (FSDP), which generates noise data using gradient-driven stochastic response and achieves fast convergence with K-modes clustering, effectively reducing computational complexity. The process of the K-modes clustering SDP algorithm is shown in Figure 2.

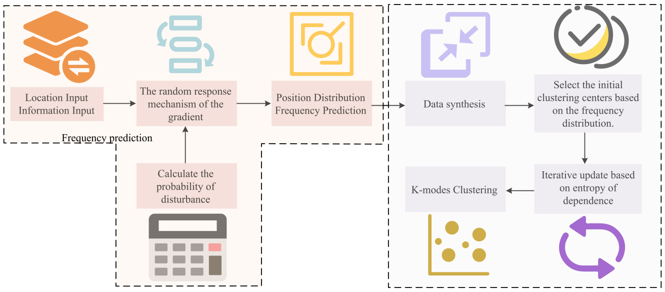

Location Input Information Input
The random response mechanism of the gradient
Position Distribution Frequency Prediction
Frequency prediction
Calculate the probability of disturbance
Data synthesis
Select the initial clustering centers based on the frequency distribution.
Iterative update based on entropy of dependence
K-modes Clustering

Figure 2. Process of the K-Modes Clustering SDP Algorithm

As shown in Figure 2, the algorithm first groups data samples based on their distance relationships and calculates the perturbation probability for each group. Then, it introduces a Gradient-driven Stochastic Response Strategy (GSRS) to add noise and perform frequency estimation. Next, the K-modes clustering method aggregates the perturbed data, and the synthetic data set generated maintains the overall distribution characteristics while effectively preventing the leakage of individual features, thus achieving differential privacy protection.Among them, the distance from the output y to the input value x is used to partition into $m = [G_1(x), \cdots, G_m(x)]$ groups, with the partition $G_j(x)$ depending on the input value. The perturbation probabilities for all groups are calculated as shown in Equation (3).

$$\forall x \in \Omega,\ P(y \mid x) = \begin{cases} \alpha_1(x), & y \in G_1(x) \\ & \vdots \\ \alpha_m(x), & y \in G_m(x) \end{cases} \tag{3}$$

In Equation (3), Ω represents the given domain, perturbation probability $\alpha_1(x), \cdots, \alpha_m(x)$ denotes the probability distribution in the m groups based on distance partitioning. The perturbation probability change between adjacent groups remains linear, i.e., the difference in perturbation probability between the groups is constant. $\alpha_{\max}\ (x)$ and $\alpha_{\min}\ (x)$ are the maximum and minimum values of the perturbation probability, with the exact maximum perturbation probability shown in Equation (4).

$$\begin{cases} \alpha_{\min}\ (x) = \frac{m-1}{(m-1)\omega \cdot c-(c-1)\sum_{j=2}^{m}[(j-1)\cdot|G_j(x)|]} \\ \alpha_{\max}\ (x) = \alpha_{\min}\ (x) \cdot c \end{cases} \tag{4}$$

In Equation (4), c is a constant, ω is the domain size, $|G_j(x)|$ is the group size of $G_j(x)$, and $\forall j \in [1, m]$ is the perturbation value. The privacy budget ε upper limit is then calculated, as shown in Equation (5).

$$\varepsilon = \max_{x,x' \in \Omega} \log\left(c \cdot \frac{(m-1)d \cdot c-(c-1)\sum_{j=2}^{m-1}[(j-1)\cdot|G_j(x)|]}{(m-1)d \cdot c-(c-1)\sum_{j=2}^{m-1}[(j-1)\cdot|G_j(x')|]}\right) \tag{5}$$

In Equation (5), for any input value x and x' belonging to the given domain Ω, the optimal perturbation probability c is calculated using the parameters ε, Ω, and m. After processing with GSRS, the output y in the candidate data set C_x is derived to obtain the sampling frequency, as shown in Equation (6).

$$\vartheta\left(C_x\right) = \vartheta(x) \sum_{y \in C_x} P(y \mid x) + \sum_{x_i' \neq 1} \vartheta\left(x'\right) \cdot \left[\sum_{y \in C_x} P\left(y \mid x'\right) + \sum_{y \in C_x \cap C_{x'}} P\left(y \mid x'\right)\right] \tag{6}$$

In Equation (6), $\vartheta(x)$ is the frequency probability distribution of x, and $\vartheta\left(C_x\right)$ is the sampling probability of $y \in C_x$. Finally, the FSDP algorithm perturbs the electricity data set and uses the K-modes algorithm to improve the aggregated expression of the data structure, establishing a K-modes clustering SDP-based meter data protection method, named KmFSDP. The architecture of this method is shown in Figure 3.

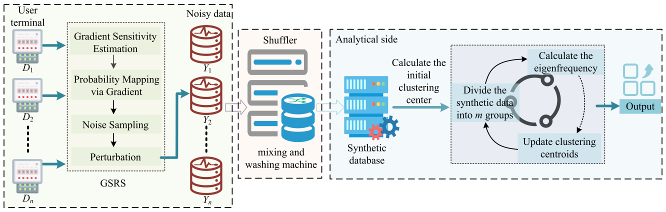

User terminal
D_1
D_2
D_n
Gradient Sensitivity Estimation
Probability Mapping via Gradient
Noise Sampling
Perturbation
GSRS
Noisy data
Y_1
Y_2
Y_n
Shuffler
mixing and washing machine
Analytical side
Synthetic database
Calculate the initial clustering center
Divide the synthetic data into m groups
Calculate the eigenfrequency
Update clustering centroids
Output

Figure 3. KmFSDP Operational Framework Diagram

As shown in Figure 3, the user electricity data is first input into the GSRS module, where the gradient-driven stochastic response mechanism performs differential perturbation, generating a privacy-protected data set. The data is then reshuffled to eliminate potential temporal or identity associations. After receiving the perturbed data, the data processing side performs feature vector-level aggregation estimation, approximating the original data and generating a synthetic data set that matches the original data distribution. Finally, the K-modes clustering algorithm is applied to classify the synthetic data, providing support for subsequent analysis. The calculation of the cluster center μ_k for the synthetic data set D'' is shown in Equation (7) [23].

$$\mu_k = \arg \max_{v \in \mathrm{Dom}(x_k)} \sum_{l=1}^{n} \gamma\left(x_{kl} = v\right) \tag{7}$$

In Equation (7), x_{kl} represents the value of the k-th data sample on the l-th feature, $\mathrm{Dom}\left(x_k\right)$ is the value range of the k-th feature, $\gamma\left(x_{kl} = v\right)$ is an indicator function, and v represents the frequency of a certain value in the l-th feature across all samples. The most common v is selected as the value of the clustering center μ_k for that feature, avoiding direct exposure of the original user features and strengthening privacy protection.

3.2 Meter Data Protection Model Combining KmFSDP and TANN-Based TOP-K Queries

Although the optimized data protection mechanism can effectively meet privacy needs, it often sacrifices query efficiency when processing complex time-series data. Therefore, to balance privacy protection of smart meter data with efficient key node identification, this research further proposes a query architecture that integrates neural network-based TOP-K query and SDP. This architecture protects user behavior data from leakage while accurately identifying key nodes in the power grid and improving query efficiency. TOP-K queries in smart grids help identify critical smart meter nodes, enhancing real-time monitoring of electric energy flow, load trends, and energy consumption anomalies [24]. Given the high frequency, high dimensionality, and strong temporal correlation of smart meter data, this method uses a time-automated wave neural network structure controlled by differential privacy to both improve query efficiency and effectively avoid data leakage risks. In the neural network component implementation, the TANN structure primarily consists of an input layer, a wave propagation layer, and an output layer. The input layer receives frequency feature vectors perturbed and clustered by the

KmFSDP algorithm, and converts them into a time series input matrix. In the neural network component implementation, the TANN structure primarily consists of an input layer, a wave propagation layer, and an output layer. The input layer receives frequency feature vectors perturbed and clustered by KmFSDP and converts them into a time series input matrix, providing a privacy-preserving data foundation for subsequent time series modeling. The wave propagation layer calculates the propagation strength and arrival time between neurons based on the automatic wave propagation mechanism. The propagation strength $P_{\lambda\eta}^{k}$ is determined by the frequency change rate and the time interval, while the arrival time function $TS_{\lambda\eta}(t)$ describes the temporal relationship between signals from different neurons. The activation threshold $L_{\lambda\eta}^{k}$ determines the wave triggering condition. When the accumulated wave strength exceeds the threshold, the wave firing process is triggered, and energy transfer and information diffusion are completed based on the spatial position parameter U_{gw}^{e} and the position update amount $U_{\lambda\eta}^{k}$. The output layer performs temporal sorting and screening of the top-K nodes based on the propagation path's time window and spatial position parameter update results, thereby enabling accurate querying of high-frequency meter data and identification of key nodes under differential privacy constraints. The general neural network structure in the time-automated wave neural network is shown in Figure 4.

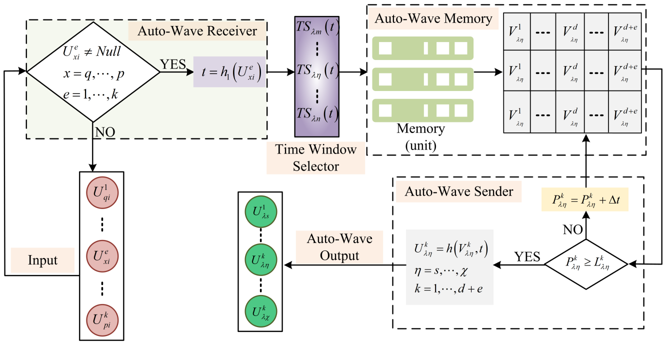


Figure 4. General Neuron Structure Diagram

As shown in Figure 4, each neuron g in this structure receives a series of input wave signals U_{gw}^{e} from preceding neurons, processes them according to interpretation logic, and extracts effective time windows to determine the time segments that participate in the current activation judgment. Then, according to the corresponding time storage window, spatiotemporal location information is mapped. The cumulative fluctuation intensity is then checked to see if it meets the threshold, determining whether it enters the wave generation phase. Finally, neurons that meet the conditions will output an automatic wave at a certain time. The automatic wave from neuron g to the current neuron w is expressed as shown in Equation (8).

$$U_{gw}^{e} = \left\{ \begin{array}{l} \text{Null , Valid signal exists} \\ \text{non } - \text{ Null , No valid signal exists} \end{array} \right. \tag{8}$$

In Equation (8), U_{gw}^{e} represents the initial excitation unit of the automatic wave, and its path spans from neuron v_g to v_w. Based on this judgment logic, the time required for the neuron to reach is given in Equation (9).

$$t = h_1\left(U_{gw}^{e}\right) = t_{gw}^{e} \tag{9}$$

In Equation (9), t_{gw}^{e} represents the arrival time of U_{gw}^{e}. Another form of this automatic wave is $U_{gw}^{e} = \left[V_{gw}^{e}, t_{gw}^{e}\right]$, where V_{gw}^{e} represents the e-th element in the set of feasible paths between neurons v_g and $v_w.h_1(\cdot)$ is the fluctuation intensity extraction function, affecting the time response judgment of the signal in the time window. Then, according to the time window function $TS_{\lambda\eta}(t)$, the best time window is selected, as shown in Equation (10).

$$TS_{\lambda\eta}(t) = c_{\lambda\eta}^{t} \tag{10}$$

In Equation (10), $c_{\lambda\eta}^{t}$ represents the effective time window span between neurons v_x and v_i at time t. Based on the dynamic adjustment mechanism of the time window, the TANN exhibits good temporal adaptability, enabling it to effectively capture the temporal characteristics of information propagation between nodes [25], [26]. Based on these characteristics, TANN is applied to tasks such as time-varying network shortest path and TOP-K key node queries, with the specific process shown in Figure 5.

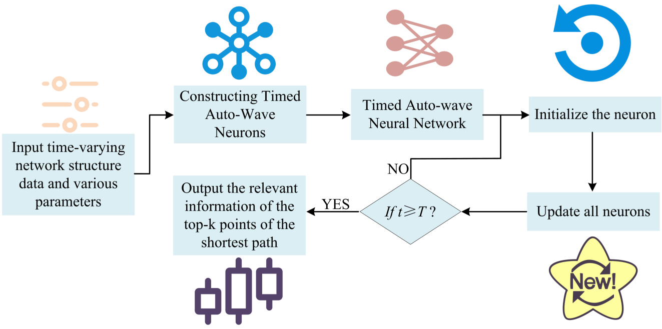

Input time-varying network structure data and various parameters
Constructing Timed Auto-Wave Neurons
Timed Auto-wave Neural Network
Initialize the neuron
Update all neurons
$If\ t \geqslant T\ ?$
NO
YES
Output the relevant information of the top-k points of the shortest path
New!

Figure 5. Shortest Path Search and Key Node Query Diagram

As shown in Figure 5, this algorithm, based on the automatic wave principle, finds the optimal path from the source node to the target node and retrieves and integrates all feasible solutions within the best time window, forming a candidate path set. Then, based on evaluation metrics for the nodes in the set, the TOP-K optimal solutions for key nodes are determined. To support fast access and dynamic updates of path information, the algorithm uses a two-dimensional matrix to efficiently store and manage automatic waves, where each column corresponds to a complete automatic wave record. The expression is given in Equation (11).

$$\begin{cases} V_{\lambda\eta}^{e} = h_3\left(U_{gw}^{e}\right) = \left\langle v_s^e, \cdots, v_\eta^e \right\rangle \\ L_{\lambda\eta}^{e} = TS_{\lambda\eta}(t) \\ P_{\lambda\eta}^{e} = h_4\left(TS_{\lambda\eta}(t)\right) = 0 \end{cases} \tag{11}$$

In Equation (11), $L_{\lambda\eta}^{e}$ represents the distance between neurons v_λ^e and v_η^e, and $P_{\lambda\eta}^{e}$ is the spatiotemporal location parameter. Then, by detecting whether the cumulative wave strength $P_{\lambda\eta}^{k}$ reaches the threshold $L_{\lambda\eta}^{k}$, the wave generation is activated, as shown in Equation (12).

$$P_{\lambda\eta}^{k} \geq L_{\lambda\eta}^{k} \tag{12}$$

After satisfying the activation conditions, the neuron enters the wave generation phase and synchronously updates its corresponding spatial location parameters. The specific update expression is given in Equation (13).

$$\tilde{P}_{\lambda\eta}^{k} = P_{\lambda\eta}^{k} + \Delta t \tag{13}$$

In Equation (13), Δt represents the dynamic update increment of the neuron position parameters. Then, under the previous constraint conditions, wave signal calculations are performed based on the temporal parameters between neural nodes, promoting the formation of new waveforms and storing them in the automatic wave storage. The new automatic wave is shown in Equation (14).

$$U_{\lambda\eta}^{k} = h\left(V_{\lambda\eta}^{k}, t\right) \tag{14}$$

In Equation (14), t represents the current time. Finally, the automatic wave emitter sends the automatic wave to subsequent neurons, completing the transmission of fluctuations and generating output.” to ”Ultimately, the automatic wave transmitter transmits wave signals to subsequent neurons based on the energy coupling relationship between neurons, realizing the layer-by-layer propagation of information and response output, and using the stable and convergent wave intensity as the sorting basis for TOP-K queries [27]. This serves as the basis for TOP-K query sorting. In summary, this research proposes a model that combines the KmFSDP algorithm with the TANN-based TOP-K query algorithm, creating a smart meter data protection model integrating differential privacy and TOP-K queries, named KSDP-TTK. The specific process architecture is shown in Figure 6.

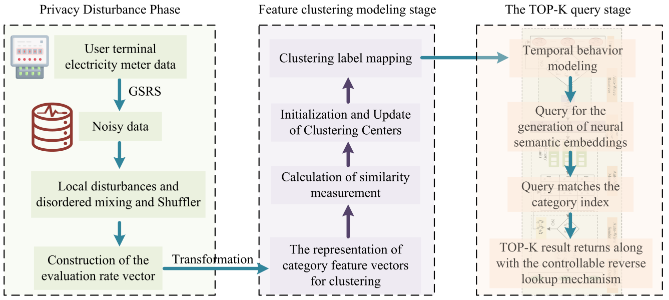

Privacy Disturbance Phase
Feature clustering modeling stage
The TOP-K query stage
User terminal electricity meter data
GSRS
Noisy data
Local disturbances and disordered mixing and Shuffler
Construction of the evaluation rate vector
Transformation
The representation of category feature vectors for clustering
Calculation of similarity measurement
Initialization and Update of Clustering Centers
Clustering label mapping
Temporal behavior modeling
Query for the generation of neural semantic embeddings
Query matches the category index
TOP-K result returns along with the controllable reverse lookup mechanism

Figure 6. KSDP Process Flow Diagram

As shown in Figure 6, this model first introduces the FSDP algorithm, perturbing the original smart meter data by combining frequency prediction and the gradient-stochastic response mechanism, enhancing privacy protection through local shuffling. Then, the K-modes algorithm is used to cluster the perturbed categorical frequency vectors, constructing the clustering results with optimal perturbation probabilities, compressing the query space, and generating efficient indexes. Finally, the TANN-based TOP-K algorithm dynamically models temporal features, integrates neuron automatic wave queries, and uses neural networks for high correlation matching and fast sorting, accurately locating the shortest path TOP-K target data. This architecture optimizes privacy protection and query performance in layers through perturbation, clustering compression, and intelligent querying.

4. Verification of the Meter Data Protection Model Combining KmFSDP and TANNTOP-K

4.1 Performance Verification of the K-modes Clustering SDP Algorithm

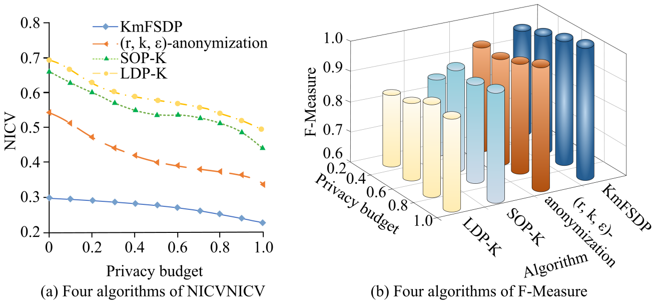
In smart meter data, user electricity consumption data was grouped by time period or usage pattern through cluster-

Table 1
Experimental Equipment Environment and Specific Configuration

Item	Disposition or Experimental parameters
Data set	REDD
Number of power consumption by users	6
Power signal frequency	15 kHz
The number of electrical equipment or circuits	24
Recorded days	18
Operating system	Windows 10 64-bit
Experiment condition	Intel Core i7-7500UCPU@16G, NVIDIA GeForce RTX 4060X
Simulation tool	Python 3.7.4

ing, generating data clusters rather than individual user-specific data, achieving a certain level of anonymization. To verify the superior performance of the KmFSDP algorithm in meter data protection, the study compared it with three other algorithms: Privacy-Preserving Data Publishing Algorithm ((r, k, ε)-anonymization), Local Differential Privacy and K-means (LDP-K), and Shuffler-based Differential Privacy and K-means (LDP-K), and Shuffler-based Differential Privacy and K-means (SDP-K) [28]. The dataset used was the Reference Energy Disaggregation Data Set (REDD) from MIT, and the experimental equipment environment and specific configuration parameters are shown in Table 1.

Based on the experimental setup, to verify the clustering performance of the KmFSDP algorithm under different privacy budgets, the study compared the normalized intra-cluster variance (NICV) and the weighted harmonic mean of precision and recall (F-Measure) for the four algorithms. The experimental results are shown in Figure 7.



(a) Four algorithms of NICVNICV

(b) Four algorithms of F-Measure

Figure 7. NICV and F-Measure Experimental Results

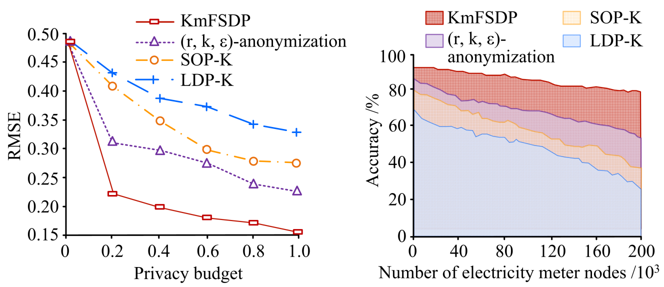
As shown in Figure 7(a), as the privacy budget increased, the KmFSDP algorithm's HICV value gradually decreased to 0.23, significantly outperforming the other three comparison algorithms. This indicated that under the shuffling and differential disturbance mechanisms, KmFSDP still maintained strong clustering consistency, with more robust clustering performance. Figure 7(b) further shows that the KmFSDP algorithm's F-Measure value remained above 0.9 across various budgets, reaching a maximum of 0.976, demonstrating a clear performance advantage over other methods. This indicated that the algorithm achieved efficient anonymous protection without relying on third-party servers, significantly improving data privacy. At the same time, the improvement in clustering quality provided more accurate indexing support for subsequent TOP-K queries. To further evaluate the query accuracy of each algorithm under different privacy budget conditions, the study compared the precision and recall of the four algorithms. The results are shown in Table 2.

Table 2
Performance Comparison of Four Algorithms on the REDD Dataset Under Different Privacy Budgets

Privacy Budget ε	Algorithm	Precision (%)	Recall (%)
0.5	KmFSDP (Proposed)	95.41	94.26
	SDP-K	86.12	82.45
	LDP-K	88.54	85.09
	(r, k, ε)-Anonymization	90.63	87.74
1.0	KmFSDP (Proposed)	97.84	96.73
	SDP-K	89.42	86.57
	LDP-K	92.18	90.11
	(r, k, ε)-Anonymization	94.65	92.79
2.0	KmFSDP (Proposed)	98.73	97.58
	SDP-K	91.08	88.34
	LDP-K	93.47	91.82
	(r, k, ε)-Anonymization	95.62	94.03
3.0	KmFSDP (Proposed)	99.05	98.12
	SDP-K	91.92	89.55
	LDP-K	94.22	92.68
	(r, k, ε)-Anonymization	96.07	94.56

Table 2 shows that the KmFSDP algorithm exhibits optimal performance under various privacy budget conditions. When $\varepsilon = 0.5$, its precision and recall reach 95.41% and 94.26%, respectively, maintaining high query stability under strong privacy constraints. As ε increases to 3.0 , precision and recall further improve to 99.05% and 98.12%, consistently significantly outperforming algorithms such as SDP-K, LDP-K, and ($\mathrm{r}, \mathrm{k}, \varepsilon$)-Anonymization. This demonstrates that KmFSDP achieves an optimal balance between privacy budget utilization, cluster consistency, and query accuracy, effectively balancing data availability and privacy protection strength, and demonstrating superior robustness and practicality. To further demonstrate the excellent performance of the KmFSDP algorithm in handling data while ensuring privacy security, the study compared the Root Mean Squared Error (RMSE) and frequency prediction accuracy of the four algorithms' clustering performance. The comparison results are shown in Figure 8.

Figure 8 RMSE and prediction accuracy results for clustering performance As shown in Figure 8(a), as the privacy budget increased, the KmFSDP algorithm exhibited a better convergence trend in clustering error, with its RMSE value decreasing the most, reaching a minimum of 0.155 when the budget was 1 , significantly outperforming the other three comparison methods. Figure 8(b) further demonstrates that under expanding data scales, the KmFSDP algorithm maintained a high prediction accuracy, ultimately reaching 83%, while the other algorithms had accuracy rates below 60% when the sample size reached 200,000 . These results indicated that KmFSDP



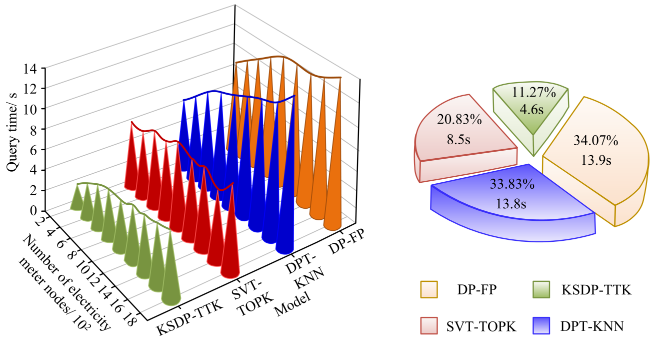
(a) RMSE comparison (b) Accuracy comparison

Figure 8. RMSE and Prediction Accuracy Results for Clustering Performance

effectively suppressed errors caused by privacy disturbance while accurately capturing frequency distribution characteristics, ensuring the stability and reliability of meter data predictions.

4.2 Verification of the Meter Data Protection Model Combining DP and TANNTOP-K

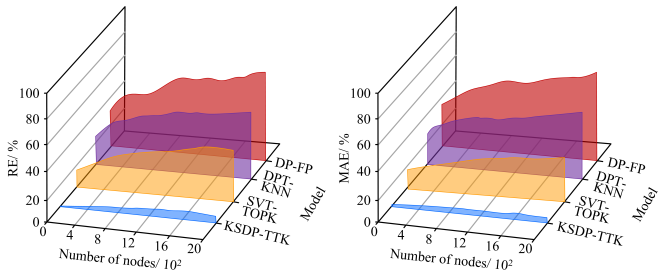
The study applied the KmFSDP method to perturb meter data for privacy protection but also aimed to meet specific query efficiency requirements to address the practical demands of smart meter data in sensitive privacy scenarios. Therefore, to further evaluate the performance of the KSDP-TTK model in TOP-K queries and privacy protection, the study compared it with three other models: Sparse Vector Technique for TOP-K Query (SVT-TOP-K), Differentially Private Time-series KNN (DPT-KNN), and Differential Privacy-based Frequency Prediction (DP-FP). The dataset used was the smart energy dataset, and the comparison of the TOP-K query times for the four algorithms under different data scales is shown in Figure 9.



(a) Comparison of query times of four algorithms on different nodes (b) The query time for the four algorithms in terms of the maximum number of nodes

Figure 9. Query Time Comparison Under Different Data Scales

As shown in Figure 9(a), as the number of meter nodes increased, the KSDP-TTK model's query time remained low, ultimately taking 4.6s, while the other three comparison algorithms' query times were 8.5s, 13.9s, and 13.8s, respectively. Figure 9(b) shows that when the number of nodes was 1800, the total query time for the four models was summed, and the time proportions for each model, from smallest to largest, were 11.27%, 20.83%, 33.83%, and 34.07%. The KSDP-TTK model's query time was significantly lower than the comparison algorithms, demonstrating its significant TOP-K vertex importance query efficiency advantage in large-scale scenarios. To further verify the strong privacy protection capabilities of the proposed method while ensuring query accuracy, the study analyzed and discussed the Mean Absolute Error (MAE) and Relative Error (RE) between the original frequency and the perturbed predicted frequency. The experimental results are shown in Figure 10.

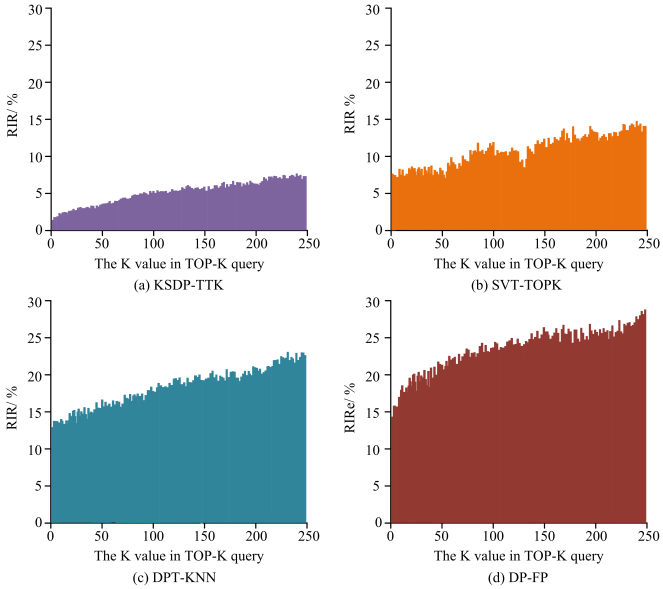


(a) RER comparison (b) MAE comparison

Figure 10. MAE and RE Results Between Original and Perturbed Frequencies

As shown in Figure 10(a), as the TOP-K query scale increased, the RE values for all models rose, but the KSDP-TTK model consistently maintained the lowest relative error level, with an RE value of 6.9%, significantly lower than the comparison models, indicating better frequency fidelity. This showed that the model effectively reduced the query bias introduced by the privacy mechanism, ensuring higher query usability. Figure 10(b) further illustrates the change trend of MAE values with increasing query scale. It could be seen that as the query scale increased, the MAE of all models increased to varying degrees. The KSDP-TTK model's MAE value remained the lowest at 4.5%, far lower than the 36.9%, 54.7%, and 63.2% values for the comparison algorithms. In conclusion, the proposed model achieved a good balance between accuracy and privacy through clustered shuffling differential perturbation, effectively supporting high-quality TOP-K queries under the privacy protection of smart meter data. To further verify the performance of the KSDP-TTK model in terms of information leakage and protection capabilities, the study compared the Re-Identification Rate (RIR) of the four models under large-scale data. The experimental results are shown in Figure 11.

As shown in Figure 11, as the TOP-K query value increased, the RIR values for the four algorithms showed varying degrees of increase. Among them, the KSDP-TTK model consistently maintained the lowest re-identification rate. When K was 100 , its RIR value was only 5.1%, far lower than 12.5% for SVT-TOP-K, 18.2% for DPT-KNN, and 23.7% for DP-FP. When K increased to 25, although the re-identification risk increased for all algorithms, the model still controlled the risk at below 10%, with the lowest RIR value at 7.2%, demonstrating good scalability stability. These results showed that during the TOP-K query



(a) KSDP-TTK

(b) SVT-TOPK

(c) DPT-KNN

(d) DP-FP

Figure 11. RIR Experimental Results Under Large-Scale Data

process, the KSDP-TTK model effectively suppressed the privacy leakage risks caused by an increase in the number of query targets, providing stronger anonymity and resistance to inference.

5. Conclusion

Due to the continuous collection of high-frequency, fine-grained electricity consumption information by smart meters, users' behavior patterns and lifestyle habits can easily be inferred, leading to significant privacy leakage risks. Therefore, the study proposed a smart meter data protection model that integrates the K-modes clustering shuffling differential privacy algorithm and TOP-K query based on TANN, aiming to achieve an efficient data processing method that balances both privacy protection and query performance. Experimental results demonstrated that the proposed KmFSDP algorithm performed excellently in terms of clustering consistency, with the lowest HICV value of 0.23 and the highest F-Measure value of 0.976. In terms of clustering performance, its RMSE value exhibited the largest reduction, dropping to the lowest value of 0.155 when the budget was 1, and the prediction accuracy ultimately reached 83%. At the same time, the KSDP-TTK model's TOP-K query time was only 4.6 s, and in terms of query error, its RE value was 6.9% and its MAE value was 4.5%, achieving a good balance between accuracy and privacy while effectively supporting high-quality TOP-K query demands under smart meter data privacy protection. The model's RIR value was the lowest at 7.2%, significantly lower than the three comparison models, demonstrating stronger anonymity and resistance to inference. In summary, the KSDP-TTK model successfully balanced privacy protection strength and query accuracy, possessing strong data perturbation control capabilities and efficient query performance, making it well-suited to meet the practical query needs of smart meter data in privacy-sensitive scenarios. However, some parameters of the proposed model depend on manual settings and lack adaptive optimization. In the future, a federated learning framework could be introduced to further enhance the model's scalability and practicality in dynamic privacy scenarios.

Funding

The research is supported by 2025 University-Level Key Scientific Research Project: Research on Key Technologies of Cloud-Edge-Terminal Collaborative LoRa Fire Monitoring System and Its Application in Public Security Field (AF2025Z08); Wenzhou City Basic Industrial Science and Technology Project: Project Title: Research on Anti-Reset Device for Preventing Theft Status of Smart Electric Meters (G2020029).

References

[1] M. Irfan, A. Niaz, M. Q. Habib, M. U. Shoukat, S. H. Atta, and A. Ali, "Digital Twin Concept, Method and Technical Framework for Smart Meters," *European Journal of Theoretical and Applied Sciences*, vol. 1, no. 3, pp. 105–117, 2023.

[2] S. Zidi, A. Mihoub, S. M. Qaisar, M. Krichen, and Q. A. Al-Haija, "Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 1, pp. 13–25, 2023.

[3] C. Chen, "Regional differences and spatial correlation network characteristics of power consumption in China under dual-carbon targets," *Pollution and Environment*, vol. 7, no. 1, pp. 10–14, 2023.

[4] X. Li, M. Zhuang, W. Yang, X. Zhu, and Q. Wang, "Neural Network Based Data Quality Monitoring and Real-Time Analysis Method for Energy Storage Power Plants," *International Journal of Power and Energy Systems*, vol. 44, no. 10, pp. 1–15, 2024.

[5] B. Ning, X. Zhang, S. Gao, and G. Li, "Dp-agm: a differential privacy preserving method for binary relationship in mobile networks," *Mobile Networks and Applications*, vol. 28, no. 5, pp. 1597–1616, 2023.

[6] C. Xu, P. Zhang, L. Mei, Y. Zhao, and L. Xu, "Ranked searchable encryption based on differential privacy and blockchain," *Wireless Networks*, vol. 30, no. 6, pp. 4735–4748, 2024.

[7] X. Sun, Q. Ye, H. Hu, *et al.*, "Synthesizing realistic trajectory data with differential privacy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 5502–5515, 2023.

[8] R. Hu, Y. Guo, and Y. Gong, "Federated learning with sparsified model perturbation: Improving accuracy under client-level differential privacy," *IEEE Transactions on Mobile Computing*, vol. 23, no. 8, pp. 8242–8255, 2023.

[9] M. Zhang, E. Wei, R. Berry, and J. Huang, "Age-dependent differential privacy," *IEEE Transactions on Information Theory*, vol. 70, no. 2, pp. 1300–1319, 2023.

[10] L. Huang, J. Wu, D. Shi, S. Dey, and L. Shi, "Differential privacy in distributed optimization with gradient tracking," *IEEE Transactions on Automatic Control*, vol. 69, no. 9, pp. 5727–5742, 2024.

[11] Y. Zhu, Q. Song, and Y. Luo, "Differentially Private Top-k Flows Estimation Mechanism in Network Traffic," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 3, pp. 2462–2472, 2023.

[12] D. Xu, C. Peng, W. Wang, H. Liu, S. A. Shaikh, and Y. Tian, “Privacy-preserving dynamic multi-keyword ranked search scheme in multi-user settings,” *IEEE Transactions on Consumer Electronics*, vol. 69, no. 4, pp. 890–901, 2023.

[13] B. C. Kara and C. Eyüpoğlu, “A New Privacy-Preserving Data Publishing Algorithm Utilizing Connectivity-Based Outlier Factor and Mondrian Techniques,” *Computers, Materials & Continua*, vol. 76, no. 2, pp. 1515–1535, 2023.

[14] X. Li, H. Zhao, J. Xu, G. Zhu, and W. Deng, “APDPFL: Anti-poisoning attack decentralized privacy enhanced federated learning scheme for flight operation data sharing,” *IEEE Transactions on Wireless Communications*, vol. 23, no. 12, pp. 19098–19109, 2024.

[15] D. Guo, S. Zhang, J. Zhang, B. Yang, and Y. Lin, “Exploring Contextual Knowledge-Enhanced Speech Recognition in Air Traffic Control Communication: A Comparative Study,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 9, pp. 16085–16099, 2025.

[16] R. Yan, Y. Zheng, N. Yu, and C. Liang, “Multi-smart meter data encryption scheme based on distributed differential privacy,” *Big Data Mining and Analytics*, vol. 7, no. 1, pp. 131–141, 2023.

[17] D. Singhal, L. Ahuja, and A. Seth, “POSMETER: proof-of-stake blockchain for enhanced smart meter data security,” *International Journal of Information Technology*, vol. 16, no. 2, pp. 1171–1184, 2024.

[18] X. Wang, H. Xie, L. Tang, C. Chen, and Z. Bie, “Decentralized privacy-preserving electricity theft detection for distribution system operators,” *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 2179–2190, 2023.

[19] A. K. Singh and J. Kumar, “A secure and privacy-preserving data aggregation and classification model for smart grid,” *Multimedia Tools and Applications*, vol. 82, no. 15, pp. 22997–23015, 2023.

[20] B. Ouvrard, R. Préget, A. Reynaud, and L. Tuffery, “Nudging and subsidising farmers to foster smart water meter adoption,” *European Review of Agricultural Economics*, vol. 50, no. 3, pp. 1178–1226, 2023.

[21] Z. Ju and Y. Li, “V2V-ESP: Vehicle-to-Vehicle Energy Sharing Privacy Protection Scheme Based on SDP Algorithm,” *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 1, pp. 1093–1105, 2023.

[22] P. Zhao, Z. Yang, and G. Zhang, “Personalized and Differential Privacy-Aware Video Stream Offloading in Mobile Edge Computing,” *IEEE Transactions on Cloud Computing*, vol. 12, no. 1, pp. 347–358, 2024.

[23] Y. Buchana, “Identifying Clusters of Innovation Barriers in Farming Enterprises: A K-modes Clustering Approach,” *Agrekon*, vol. 64, no. 1, pp. 32–49, 2025.

[24] S. Souror, M. Badawy, and N. El-Fishawy, “Secure Query Processing for Smart Grid Data Using Searchable Symmetric Encryption,” *The Journal of Supercomputing*, vol. 80, no. 16, pp. 24173–24211, 2024.

[25] G. Zhao, J. Tan, and N. Sun, “Research on key technology of condition monitoring system of substation equipment,” *International Journal of Power and Energy Systems*, vol. 45, no. 10, 2025.

[26] M. Rezaali, R. Fouladi-Fard, and A. Karimi, “Performance of TANN, NARX, and GMDHT Models for Urban Water Demand Forecasting: A Case Study in a Residential Complex in Qom, Iran,” *Avicenna Journal of Environmental Health Engineering*, vol. 10, no. 2, pp. 85–97, 2023.

[27] J. Li, Y. Peng, Z. Yang, *et al.*, “Virtual power plant economic dispatch model based on parallel molecular differential evolution algorithm,” *International Journal of Power and Energy Systems*, vol. 45, no. 2, pp. 64–77, 2025.

[28] B. C. Kara, C. Eyupoglu, and O. Karakus, “(r, k, ε)-Anonymization: Privacy-Preserving Data Publishing Algorithm Based on Multi-Dimensional Outlier Detection, k-Anonymity and ε-Differential Privacy,” *IEEE Access*, vol. 15, no. 13, pp. 70422–70435, 2025.

Biographies

Peiyu Chen graduated from Zhejiang University of Technology in 2015 with a Master’s degree in Mechanical Engineering. He currently serves as a full-time teacher in the Internet of Things Application Technology major at the School of Artificial Intelligence, Zhejiang College of Security Technology, and his main research directions focus on embedded software development and intelligent electric meter design and development.

Zhaohui Hu graduated from Xiangtan Institute of Technology (now renamed ”Hunan University of Science and Technology”), majoring in Industrial Automation, and obtained a bachelor’s degree. He is currently serving as Chief Engineer of Haosheng (Zhejiang) Intelligent Electrical Technology Co., Ltd. He is mainly engaged in the research and development of smart electric energy meters and the forward-looking development of the energy meter market.

Qianjun Tu graduated from Huanggang Polytechnic College in 2008, majoring in Mechatronics Technology. He is currently serving as General Manager of Haosheng (Zhejiang) Intelligent Electrical Technology Co., Ltd. In addition to overseeing the overall operation of the company, he is also engaged in the development of smart electric energy meters and the forward-looking control of the energy meter market.