Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.054377

ARTICLE

Hierarchical Privacy Protection Model in Advanced Metering Infrastructure Based on Cloud and Fog Assistance

Linghong Kuang[1,2], Wenlong Shi[1,2] and Jing Zhang[1,2,*]

[1]School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou, 350118, China

[2]Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fuzhou, 350118, China

*Corresponding Author: Jing Zhang. Email: jing165455@126.com

Received: 26 May 2024 Accepted: 18 July 2024 Published: 15 August 2024

ABSTRACT

The Advanced Metering Infrastructure (AMI), as a crucial subsystem in the smart grid, is responsible for measuring user electricity consumption and plays a vital role in communication between providers and consumers. However, with the advancement of information and communication technology, new security and privacy challenges have emerged for AMI. To address these challenges and enhance the security and privacy of user data in the smart grid, a Hierarchical Privacy Protection Model in Advanced Metering Infrastructure based on Cloud and Fog Assistance (HPPM-AMICFA) is proposed in this paper. The proposed model integrates cloud and fog computing with hierarchical threshold encryption, offering a flexible and efficient privacy protection solution that significantly enhances data security in the smart grid. The methodology involves setting user protection levels by processing missing data and utilizing fuzzy comprehensive analysis to evaluate user importance, thereby assigning appropriate protection levels. Furthermore, a hierarchical threshold encryption algorithm is developed to provide differentiated protection strategies for fog nodes based on user IDs, ensuring secure aggregation and encryption of user data. Experimental results demonstrate that HPPM-AMICFA effectively resists various attack strategies while minimizing time costs, thereby safeguarding user data in the smart grid.

KEYWORDS

AMI; cloud and fog assistance; fuzzy comprehensive analysis; hierarchical threshold encryption

1 Introduction

The Smart Grid (SG) is a modern infrastructure of Cyber-Physical Systems (CPS), encompassing multiple domains and facilitating services across seven key areas: power transformation, dispatching, consumption, distribution, transmission, generation, and communications [1]. As shown in Fig. 1, the SG consists of digital and electrical technologies that communicate and transfer information from one device to another [2]. The core unit of the smart grid is the AMI, facilitating bi-directional communication and interaction between power consumers and suppliers, while optimizing the organizational structure between power loads and supply systems. Given the characteristics of the AMI system, the primary distinction between smart grids and traditional power grids lies in the increasing integration

of power and information networks, leading to a broader spectrum of services. However, concurrently, smart grids face escalating privacy and security risks.

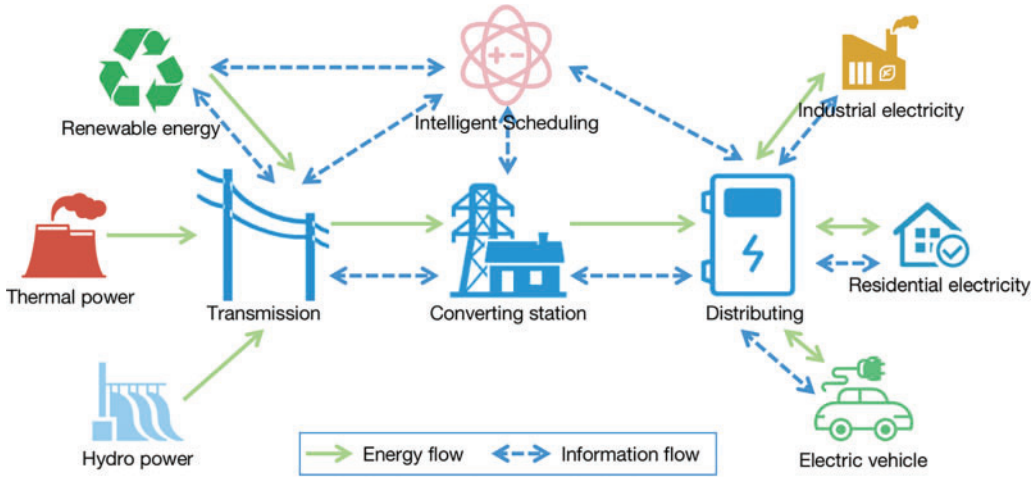

Figure 1: Main components of a smart grid system

In the smart grid, AMI can provide necessary technical and platform support for energy storage, demand response management, real-time interaction, etc. However, despite offering intelligent services, AMI also confronts substantial privacy and security challenges, including hacker attacks, energy fraud, and other criminal activities that may result in power outages, leading to significant economic losses for power companies and even posing serious impacts on people's lives [3]. To address the problem of privacy leakage in SG, researchers typically employ methods such as anonymity, encryption, homomorphic aggregation, and differential privacy. Wang et al. [4] propose a novel privacy-preserving data aggregation scheme (PDAM) for IoT-enabled smart grids. This scheme supports efficient data source authentication, integrity checking, and secure dynamic user management including join and exit procedures. Wang et al. [5] initially cluster raw user electricity data, and then enhance privacy protection by combining the clustered data with an adaptive k-anonymity algorithm for optimized clustering. This approach ensures comprehensive protection of users' personal privacy information and enables real-time collection of electricity consumption data. Hu et al. [6] propose an efficient and privacy-preserving data aggregation and trust management scheme (PATM) for IoT smart grid based on smart contracts. Various existing methods such as signature authentication, differential privacy, and encryption have been studied and designed to address privacy and security concerns in smart grids, effectively safeguarding users' personal information from potential attacks [7–9].

Unfortunately, there are still some problems in the privacy and security protection of smart grids, such as (1) During the analysis, collection, and storage of power consumption information within the AMI system, attackers may launch attacks on AMI nodes, leading to the theft of detailed personal power consumption data [10,11]. (2) Existing privacy security protection methods often overlook the diverse categories and complex characteristics of users. They tend to apply uniform privacy protection methods to a large number of users in the SG, resulting in exorbitant costs and inability to meet the real-time performance requirements of the SG [12]. (3) Simple privacy protection strategies struggle to achieve comprehensive coverage of the AMI system, while complex strategies often lack practicality. Given the immense volume of data and computational costs associated with AMI, lightweight algorithms are preferable for practical implementation in the SG [13–15].

In order to address the issues of privacy protection in smart grids, a Hierarchical Privacy Protection Model in Advanced Metering Infrastructure based on Cloud and Fog Assistance is proposed in this paper. The motivation of HPPM-AMICFA is shown intuitively in Fig. 2. Firstly, an Advanced Metering Infrastructure based on Cloud and Fog Assistance (AMICFA) is designed, and its existing privacy threats are elucidated to establish a foundation for subsequent privacy protection solutions. Secondly, a fuzzy comprehensive analysis algorithm with entropy weight method is designed to set user protection levels. Finally, a hierarchical threshold privacy protection algorithm is designed. This algorithm sets different threshold values for users of varying protection levels based on their user IDs, thereby ensuring tailored privacy protection for users. The main contributions of this study are summarized as follows:

1. To address the issue of attackers potentially launching attacks on AMI nodes, a specific AMICFA is proposed to better illustrate the privacy threat issue in AMI. By analyzing potential security attacks that may occur at various stages of AMICFA, a solid research foundation can be provided for the design of subsequent privacy protection methods.
2. To address the limitations of existing smart grid data protection methods, which ignore user diversity and complexity, and to manage the large volume and high computational costs of AMI privacy protection data, a dynamic threshold encryption model is proposed. This model first employs a fuzzy comprehensive analysis algorithm to set user protection levels, minimizing costs while providing tailored privacy protection. Subsequently, a hierarchical threshold encryption algorithm is used based on these protection levels, offering a more efficient and lightweight method for user privacy protection.
3. This article proposes a hierarchical privacy protection method with dynamic threshold. The analysis of the model's ability to resist attacks under different attack strategies and the model's runtime cost demonstrates the effectiveness of HPPM-AMICFA.

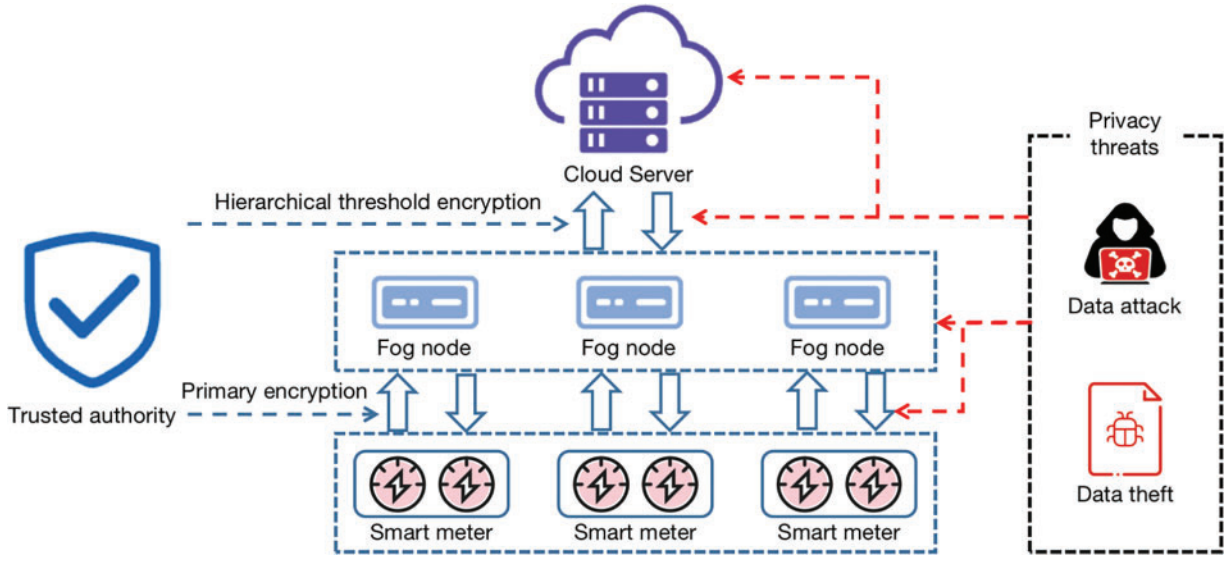

Figure 2: The motivation of HPPM-AMICFA. The AMICFA architecture consists of smart meters, fog nodes, and cloud server. The red dashed arrows indicate potential privacy threats to AMICFA, while the blue dashed arrows indicate encryption methods provided by Trusted authority for AMICFA

The rest of the paper is organized as follows: Section 2 reviews existing research related to our work. The preliminaries are given in Section 3. The privacy protection method is designed in Section 4. The simulation analysis is discussed to evaluate the effectiveness of HPPM-AMICFA in Section 5. Finally, the conclusions and outlooks are presented in Section 6.

2 Related Work

In the smart grid, achieving bidirectional communication between power flow and information flow relies heavily on the role of smart meters within the AMI system. Smart meters are pivotal as they not only collect and process information but also facilitate command scheduling and terminal data transmission. However, the electricity consumption data captured by smart meters, being both real and real-time, contains significant amounts of personal privacy information. In recent years, numerous scholars have explored privacy protection methods for smart grid data, including anonymity, differential privacy, and encryption algorithms.

Electric power companies can utilize various pieces of information such as house numbers, homeowner names, electricity usage behavior, and other data to potentially identify users' personal identities, as well as match their consumption patterns and lifestyles. Moreover, when combined with anonymous information, other user data can facilitate data matching. Srinivas et al. [16] design a novel anonymous signature-based authenticated key exchange scheme named AAS-IoTSG specifically tailored for smart grid environments supporting the Internet of Things (IoT). AAS-IoTSG offers enhanced security and functional features compared to existing state-of-the-art authentication mechanisms in smart grid systems. Wu et al. [17] propose a privacy-preserving data aggregation scheme tailored for smart grids, incorporating user anonymity and designated recipients. In this scheme, smart meters gather users' power consumption data, encrypt it using homomorphic re-encryption, and anonymously transmit the ciphertext. Subsequently, an agent employs distributed re-encryption to further encrypt the aggregated data, ensuring that only designated recipients can decrypt it. This proposed scheme offers a more secure and flexible solution for privacy-preserving data aggregation in smart grid environments. Chen et al. [18] propose a dual-blockchain-assisted secure anonymous data aggregation scheme for fog-enabled smart grids, named DA-SADA. This scheme combines Paillier encryption, batch aggregate signatures, and anonymous authentication to establish a secure anonymous data aggregation mechanism with minimal computational overhead. Adewole et al. [19] propose the DFTMicroagg algorithm, which provides dual perturbations to improve the anonymity and privacy of smart grid data. This algorithm leverages the benefits of Discrete Fourier Transform (DFT) and microaggregation to provide an additional layer of protection. In the context of identity anonymity within smart grids, it is crucial not only to ensure the traceability of identity information but also to safeguard the anonymity of user identities. However, achieving these goals in low-bandwidth and low-capacity electricity meters presents significant challenges.

Differential privacy protection technology primarily involves the addition of random noise to electricity data [20]. This approach helps to thwart attackers from obtaining centralized data. However, excessive differential noise can potentially impair the availability of aggregated data. Hence, effectively measuring both data availability and privacy is crucial in the context of differential protection technology. Dwork [21] first proposes a method of obtaining differential privacy by adding random noise, supported by rigorous mathematical proofs. Leveraging the properties of distributed differential noise, a distributed protocol is introduced to generate multiple instances of random noise, thereby fortifying defenses against malicious attacks, including those perpetrated by database administrators. Zhao et al. [22] conduct a study on the privacy of smart meters utilizing differential privacy principles.

They introduce a new random BLH algorithm aimed at ensuring differential privacy effectively. Furthermore, they propose the Multitasking BLH Exp3 algorithm, which dynamically updates the BLH algorithm based on context and constraints, thereby enhancing its adaptiveness. Gough et al. [23] develop an innovative differential privacy compatible algorithm to ensure the protection of data from consumer smart meters. This novel algorithm comprehensively examines the impact on the operation of distribution networks, considering perspectives such as consumer electricity bills and the power system. Gai et al. [24] propose a differential privacy aggregation scheme that operates without the need for a trusted authority, thereby supporting dynamic user joining and exiting. Additionally, they design a data discretization algorithm based on conditional probability, which effectively enhances the accuracy of aggregated data. Zheng et al. [25] propose a decentralized mechanism for privacy protection computation in smart grids, called DDP. This mechanism maintains differential privacy while extending data cleaning from the range to the time domain. However, the high degree of protection for differential privacy and the irreversible protection process may result in reduced data availability.

Nowadays, encryption technology stands as a crucial tool for bolstering data confidentiality and is widely employed for privacy protection both at the user and power supply ends of the smart grid. Encryption technology is divided into symmetric-key algorithm and asymmetric encryption algorithm [26]. Sarenche et al. [27] propose a protocol enabling the secure implementation of various double auction mechanisms in smart grids. In this scheme, to safeguard the anonymity and privacy of users, each participant is allocated a pseudo-identity, and bids/asks are encrypted using the Paillier cryptosystem. Chen et al. [28] propose a smart meter data aggregation scheme utilizing the Paillier homomorphic cryptographic system. This scheme allows utility suppliers to aggregate the total consumption of all smart meters, while preventing them from accessing the consumption data of individual smart meters. Xu et al. [29] propose a privacy protection framework that implements homomorphic encryption with trust boundaries for various Smart Meter System (SMS) scenarios as a system privacy protection solution for SMS. Since blockchain-based industrial wireless sensor networks can provide secure and resilient data transmission to promote intelligent integration, monitoring, and control of smart grids, Faheem et al. [30] propose a smart contracts framework in Solana called Advanced Solana Blockchain (ASB) for smart grids. This scheme can achieve resilient and secure real-time control and monitoring in smart grids. Wang et al. [31] propose two efficient pairing-free ciphertext-policy attribute-based schemes. These schemes eliminate computation-intensive bilinear pairing operations, thereby enhancing their deployment efficiency in cloud-assisted smart grids. Fan et al. [32] innovatively propose a searchable encryption scheme that supports multi keyword subset retrieval to share data and reduce local storage on Cloud Edge End Orchestrated (CEEO) securely. Wu et al. [33] design HTV-PRE, a homomorphic threshold proxy re-encryption scheme with re-encryption verifiability, proposing a robust, lightweight data aggregation scheme with strong privacy protection for smart grids.

Compared to previous network systems, the characteristics presented in the AMI system are very unique. For instance, deploying AMI components in the public domain renders simple privacy protection methods less effective due to weak security, while complex solutions may have certain limitations. Moreover, given the substantial volume of data and computational costs associated with AMI, it becomes imperative to explore more convenient and lightweight encryption methods based on the design of security models when addressing security issues in AMI.

3 Preliminaries

In this section, firstly, some Definitions related to AMICFA will be introduced. Secondly, some encryption algorithms will be introduced.

3.1 Definitions of AMICFA

The advanced metering infrastructure based on cloud and fog assistance consists of four types of entities as shown in Fig. 2: a trusted authority, smart meters, fog nodes, and a cloud server.

Definition 1 ***Grid area.*** In this article, the power grid area is assumed to be divided into several sub-regions. Each sub-region contains a group of meter users and is managed by a unique fog node. Additionally, there is a single cloud server responsible for overall control of the entire power grid area. The overall power grid area G can be represented as $G = \{(g_1, \ldots, g_m), C\}$, where g represents the sub-region, m represents the number of sub-regions, and C represents the cloud server.

Definition 2 ***Trusted authority.*** In this article, trusted authority is assumed to be a secure third-party platform primarily responsible for generating keys and distributing them to all entities.

Definition 3 ***Smart meter.*** Smart meter users can be represented by SM_{ij}, $i \leq m, j \leq n$, where n represents the number of smart meters managed by each fog node. Each smart meter SM_{ij} is capable of generating real-time electricity consumption information for users, encrypting and signing the data, and providing preliminary protection of personal information privacy. Subsequently, these encrypted and signed data reports are transmitted to the corresponding fog nodes for aggregation.

Definition 4 ***Fog node.*** Fog node can be represented by f_i, $i \leq m$, and is considered to exist in the middle of smart meters and cloud server, completing basic deployment operations at the edge of the network. Due to the possibility of criminals attacking through identity forgery, data at the fog nodes need to be verified through signature verification.

Definition 5 ***Cloud server.*** The cloud server C can decrypt and aggregate data, generate reports on the electricity consumption of each sub-region, and send them to the fog nodes to reduce the resource occupation of the cloud server. At the same time, since the fog nodes are located at the edge of the network, users can achieve real-time querying of electricity consumption data with lower latency.

3.2 Encryption Algorithm

Definition 6 ***Secret sharing*** [34]. Secret sharing refers to a secure way of sharing secrets among multiple participants, forming the basis of a threshold cryptography system. In a secret sharing scheme, a secret S is divided into several pieces. The secret S is represented as $S = \{s_1, s_2, \ldots, s_p\}$, and each piece s_i is sent by the executor to a corresponding user u_i from the user set $U = \{u_1, u_2, \ldots, u_p\}$. In this case, it is necessary to aggregate the secret pieces of p individuals together in order to decrypt the value of S. However, this scheme is relatively fragile, as the secret cannot be reconstructed if any participant is absent.

Definition 7 ***Threshold secret sharing*** [35]. Based on considerations of portability and security in secret sharing, Pang et al. propose a new (t, n) multi-secret sharing scheme. The basic idea of this scheme is that if there are p secret pieces, only k ($1 < K < n$) personal information needs to be gathered to obtain the original secret. Firstly, a polynomial of degree $t - 1$ is generated, as shown in Eq. (1).

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{t-1} x^{t-1} \tag{1}$$

where a_0 is the secret number to be shared, and $a_1, a_2, \ldots, a_{t-1}$ is the random number generated by the executor. Secondly, the secret a_0 needs to be shared and allocated to p users, as shown in Eq. (2).

$$s_i = f(i) = \sum_{j=0}^{t-1} a_j * i^j, i = 0, 1, \ldots, p \tag{2}$$

It can be seen that as long as any t of $\{s_1, s_2, \ldots, s_p\}$ are put together, the secret S can be solved, while any $t-1$ put together cannot obtain the exact solution of S, which meets the requirement of (t, n) threshold key sharing.

Definition 8 ***Threshold encryption*** [36]. The threshold encryption scheme is a distributed encryption and decryption protocol wherein any user can encrypt data using a public key. In this scheme, the data contributor specifies multiple secret holders, and decryption can only be achieved when a certain number of secret holders collaborate to aggregate the decryption pieces. Firstly, participants agree on a threshold value t and obtain their own private key pieces r. They jointly calculate the data encryption public key pk and make it public. Secondly, the data contributor encrypts the raw data R using the public key pk, generating encrypted data $E(R)$. Thirdly, participants decrypt $E(R)$ using their own private key pieces r to generate the decryption pieces $D_r(R)$. Finally, participants aggregate the decryption pieces, and only when the aggregated decryption pieces $D_r(R)$ from different participants are not less than the threshold t, can the decryption be completed, and the original data R be obtained.

4 HPPM-AMICFA

The specific process of HPPM-AMICFA proposed in this paper is shown in Fig. 3. Firstly, the AMICFA is introduced. Next, protection levels of different users are set, including using fuzzy data completion algorithms to complete missing data for users, and then using fuzzy comprehensive evaluation algorithms to score the importance of different users to determine the protection levels. Finally, user electricity consumption data privacy is protected using a hierarchical threshold encryption algorithm. The specific steps are as follows:

Step 1: The workflow of AMICFA is introduced and the privacy threats that exist within them are analyzed.

Step 2: Missing data processing based on fuzzy data completion algorithm: For users with missing data, firstly, take the user's incomplete power grid data as algorithm input. Then confirm the time period and time point of the lost data, use a fuzzy logic algorithm to complete the data. Finally, obtain the completed data as the algorithm output.

Step 3: Rating user importance based on fuzzy comprehensive evaluation algorithm: In order to address the challenge of quantifying smart grid user data, smart grid users are initially classified into four categories. The classification results and complete user data (including electricity consumption levels) are used as inputs for the algorithm. This analysis aims to rate the importance of different users and obtain their protection levels as the output of the algorithm.

Step 4: Protecting user privacy based on hierarchical threshold encryption algorithm: Firstly, the user electricity consumption data and user protection levels are used as inputs to the algorithm. Then, based on the user IDs and their corresponding different protection levels, different threshold encryption strategies are set for each fog node to achieve encrypted protection of user privacy. Finally, the algorithm will output privacy protected electricity consumption data.

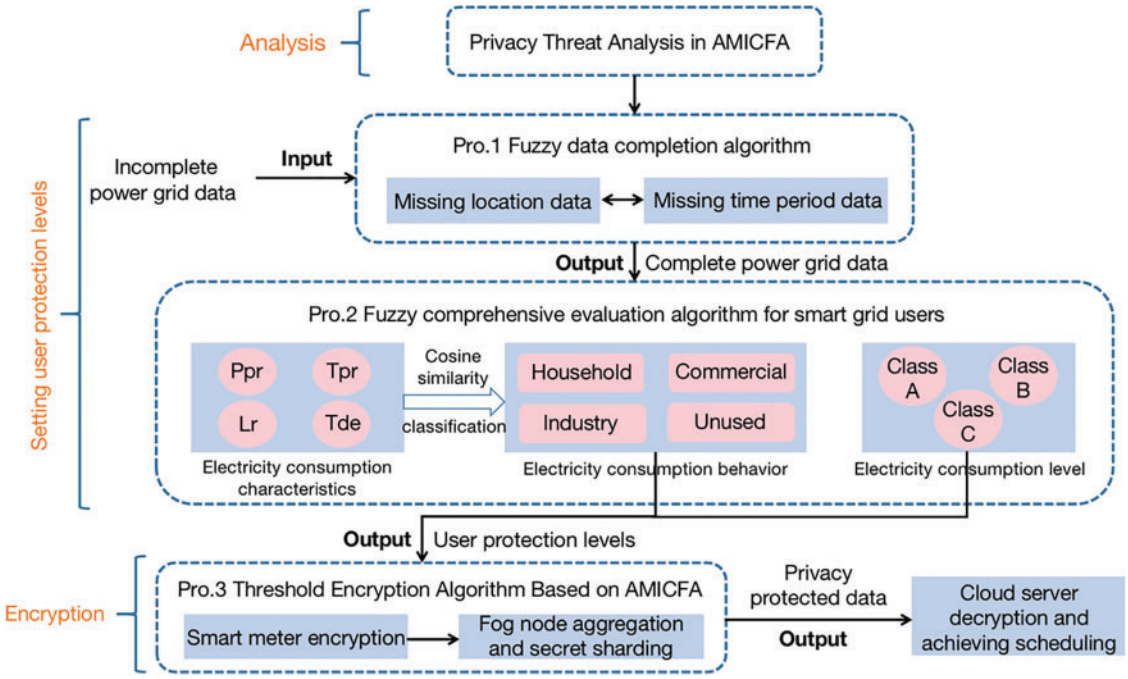

Figure 3: The workflow of HPPM-AMICFA

4.1 AMICFA

4.1.1 Composition of AMICFA

The model of AMICFA proposed in this paper consists of four entities: a trusted authority, smart meters, fog nodes, and a cloud server, as shown in Fig. 2. The trusted authority oversees user registration and is responsible for generating system parameters and keys for each user. The cloud server, a semi-trusted entity, provides various services including data storage, computation, and transmission. It can conduct threshold sharding on encrypted data from fog nodes, decrypt and retrieve aggregated data, and perform parsing to obtain electricity consumption data for each sub-region, enabling flexible regulation of electricity. Fog nodes aggregate user data within their respective regions and verify authenticity through signature verification. Smart meters generate user electricity consumption information, encrypt, and sign it before transmitting it to the corresponding fog nodes. Compared to traditional models, the AMICFA model utilizes fog nodes to store and process data at the network edge more efficiently, thereby reducing transmission costs and enhancing real-time data processing capabilities.

4.1.2 Security Threats in AMICFA

In the AMICFA scenario, based on existing research and analysis, as well as the characteristics of the cloud and fog assistance structure itself, attackers may obtain or tamper with the transmission information between nodes, posing a threat to the security of the smart grid. The threats that may occur in the AMICFA can be divided into internal and external attacks in this paper.

Internal attack: The first type of internal attacks consists of malicious node attacks, which occur during the data transmission process between users and the fog computing layer. For example, when a user transmits electricity data to a fog node, malicious nodes pretend to be legitimate nodes in the

network and initiate proactive attacks (such as identity forgery and data modification) to disrupt the authenticity and integrity of private data [37]. The second type of internal attack is described as honest-but-curious in terms of fog and cloud nodes, which perform tasks as required, but also access sensitive data as much as possible and may be caught by attackers. For example, fog nodes may be vulnerable to attacks from undetected malware, which can eavesdrop on data from devices. Therefore, it is necessary to ensure that fog nodes do not observe user privacy data throughout the data transmission process. Similarly, the model should ensure that the user's personal privacy data cannot be exported from the cloud server [18].

External attack: Attackers may attack link connections between various entities to obtain user privacy information. Therefore, the system must ensure that attackers are unable to successfully obtain private information communication links [38].

4.2 Setting User Protection Levels

4.2.1 Missing Data Processing Based On Fuzzy Data Completion Algorithm

In incomplete datasets, the algorithm first considers the importance of lost data by determining the time period in which the missing data is located. Specifically, data lost during the time interval from 23:00 pm to 7:00 am the next day is deemed less critical. Secondly, considering the characteristics of the time points where the missing data is located, if the time interval between consecutive missing data points is very short and falls outside peak or valley periods, it is also classified as less significant. However, other missing data points are regarded as important. For less important data, the linear regression method is employed to fill the missing values [39]. Linear regression is a fundamental statistical method that provides sufficient accuracy. For less important data, this method can quickly and effectively predict missing values, making it more computationally efficient.

Algorithm 1: Fuzzy data completion algorithm

Input: Incomplete smart grid dataset D, Fuzzy importance matrix A
Output: Complete dataset D'

for d_k **in** D **do**
  if $d_k = \varnothing$ **then**
    importantPer ←Determine the importance of time period (d_k);
    importantPoi ←Determine the importance of time point (d_k);
    if *importantPer* and *importantPoi* **then**
      $d_k \leftarrow \frac{\sum_{i=1}^{n} d_i \cdot a_{ik}^*}{\sum_{i=1}^{n} a_{ik}^*}$;
    else
      linear regression filling (d_k);
    end
  else
    continuous;
  end
end
D' ←Completing Data D;
Return D'

For important data, a fuzzy data completion algorithm is used to address the uncertainty and ambiguity inherent in the missing data, particularly when the data loss occurs non-randomly. This

method leverages a fuzzy importance matrix to prioritize the imputation process based on the time of day and data characteristics, which is particularly suited for datasets with patterned absences, such as those frequently encountered in electricity consumption data [40]. Specifically, a fuzzy importance matrix is designed as shown in Eq. (3).

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \tag{3}$$

where a_{ij} represents the importance of data d_i to data d_j, $a_{ij} \in (0, 1)$. Taking into account the communication environment based on the Internet of Things (IoT), the characteristics of the IoT are further considered when calculating the importance weights. The importance matrix element a_{ij} is calculated by Eq. (4).

$$a_{ij}^* = a_{ij} \times P_{ij} \times S_{ij} \times \exp\left(-\frac{E_{ij}}{E_{max}}\right) \tag{4}$$

where P_{ij} represents the probability of successful data transmission between d_i and d_j, E_{ij} represents the energy consumption of communication between d_i and d_j, E_{max} represents the maximum allowable energy consumption, and S_{ij} represents the strength of security measures during data transmission. Consider these variables as factors that affect the importance weight, and the final completion process of missing data is shown in Eq. (5).

$$d_k = \frac{\sum_{i=1}^{n} d_i \cdot a_{ik}^*}{\sum_{i=1}^{n} a_{ik}^*} \tag{5}$$

The missing data can be completed by using the known data and its importance weight to the missing data. Unlike traditional methods that may apply a one-size-fits-all approach, a resource optimization approach is proposed in this paper, which involves using efficient but relatively simple methods for less important data, while retaining more complex and costly methods for important data with greater impact. This not only ensures the efficiency of processing, but also ensures the accuracy and reliability of key information. By using this method of allocation, it is possible to balance processing costs and data quality, thereby achieving the best overall effect. The missing data completion process is shown in Algorithm 1. The line 2 to the line 4 calculate the importance of the missing data d_k. If the missing data is important data, the fuzzy data completion algorithm is used to complete the data (line 6). Otherwise, the linear regression method is used (line 8). Finally, after all the missing data is filled in, the complete dataset D' is obtained (line 14).

4.2.2 Setting User Protection Level Based on Fuzzy Comprehensive Evaluation Algorithm

The user electricity consumption data collected by smart meters is not only vast in quantity but also diverse in variety, but the data value density is low. Therefore, feature extraction of load-side user electricity consumption data can be more concise and effective in classifying the importance of users, thereby obtaining the corresponding protection level of users. The load curve features extracted in this article include:

Definition 9 ***Peak power rate, Ppr.*** Assuming that the daily peak electricity consumption is PeE and the total daily electricity consumption is TE, the Ppr can be calculated by Eq. (6).

$$Ppr = \frac{PeE}{TE} \tag{6}$$

Definition 10 ***Trough power rate, Tpr.*** Assuming that the daily trough electricity consumption is TrE and the total daily electricity consumption is TE, the Tpr can be calculated by Eq. (7).

$$Tpr = \frac{TrE}{TE} \tag{7}$$

Definition 11 ***Load rate, Lr.*** Assuming that the daily average power consumption is AvE and the total daily electricity consumption is TE, the Lr can be calculated by Eq. (8).

$$Lr = \frac{AvE}{TE} \tag{8}$$

It can be inferred that the feature vector $x_i = \{Ppr_i, Tpr_i, Lr_i, TE_i\}$ can be used to represent the electricity consumption characteristics of the i-th user. Due to the different modulus values of the four elements in the user feature vector, in order to avoid a large proportion of one element and prevent errors caused by distance classification, cosine similarity is used to classify user electricity consumption behavior in this paper, as shown in Eq. (9).

$$simcos\,(x_1, x_2) = \frac{x_1 * x_2}{\sqrt{x_1^2 + x_2^2}} \tag{9}$$

After classifying the user's electricity consumption data, the user's electricity consumption behavior classification can be obtained. This classification, together with the user's electricity consumption levels, serves as the evaluation index for the fuzzy comprehensive evaluation algorithm, allowing for the determination of the evaluation level and weight matrix. The specific steps of the fuzzy comprehensive evaluation algorithm are as follows:

Firstly, establishing a comprehensive evaluation factor set $I = \{EBC, EC\}$, where $EBC = \{Household, Commercial, Industry, Unused\}$ represents electricity consumption characteristics, and the classified user electricity consumption characteristics will belong to one of these. $EC = \{> 420\,\text{kWh}, 231 \sim 420\,\text{kWh}, 50 \sim 230\,\text{kWh}, < 50\,\text{kWh}\}$ represents electricity consumption levels.

Secondly, establishing an evaluation set $V = \{Level1, Level2, Level3, Level4\}$ to indicate the importance of users. For ease of calculation, set a score for each level, corresponding to $V = [100, 80, 60, 30]$.

Algorithm 2: Fuzzy comprehensive evaluation algorithm

Input: User electricity consumption characteristics data x, electricity payment status EP, weight matrix A, fuzzy membership matrix R and evaluation set V

Output: User protection level rating S

$EBC \leftarrow$Using cosine similarity to classify data x;
$I \leftarrow (EBC, EP)$;
for o_{ij} **in** $\boldsymbol{I}$ **do**
  $p_{ij} \leftarrow \left(o_{ij} - o_{min}\right) / \left(o_{max} - o_{min}\right)$;
end
for p_{ij} **in** $\boldsymbol{I}$ **do**
  $q_{ij} \leftarrow p_{ij} / \sum_{i=1}^{n} p_{ij}$;
end
for j **in** $\boldsymbol{m}$ **do**

(Continued)

Algorithm 2 (continued)

$e_j \leftarrow -\frac{1}{\ln n}\sum_{i=1}^{n} q_{ij} \ln q_{ij}$;

$A_j = \frac{1-e_j}{\sum_{j=1}^{m} e_j}$;

end

$B \leftarrow A * R$; //Establishing a fuzzy comprehensive evaluation matrix.

$S \leftarrow B * V$; //Calculate the protection level rating.

Return *S*

Thirdly, using entropy weight method to determine the weights of each factor. The steps to calculate the weight matrix $A = [A_1, A_2]$ using the entropy weight method are as follows:

1. The elements of matrix I are dimensionless, as shown in Eq. (10).

$$p_{ij} = \frac{o_{ij} - o_{min}}{o_{max} - o_{min}} \tag{10}$$

2. The weight of each factor for n samples is calculated, as shown in Eq. (11).

$$q_{ij} = \frac{p_{ij}}{\sum_{i=1}^{n} p_{ij}} \tag{11}$$

3. The entropy and weight of the j-th factor is calculated, as shown in Eqs. (12)–(13).

$$e_j = -\frac{1}{\ln n}\sum_{i=1}^{n} q_{ij} \ln q_{ij} \tag{12}$$

$$A_j = \frac{1-e_j}{\sum_{j=1}^{m} e_j} \tag{13}$$

Fourthly, setting membership values for fuzzy evaluation factors. Assuming the membership matrix of EBC is $R_1 = [D_1, D_2, D_3, D_4]$, and the membership matrix of EC is $R_2 = [M_1, M_2, M_3, M_4]$. The fuzzy membership matrix is shown in Eq. (14).

$$R = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \tag{14}$$

Fifthly, establishing a fuzzy comprehensive evaluation matrix. The fuzzy vector A on I is changed to the fuzzy vector B on V through fuzzy transformation as shown in Eq. (15).

$$B = A * R \tag{15}$$

where $*$ represents matrix multiplication.

$$S = B * V^{-1} \tag{16}$$

Finally, calculate the protection rating for each user as shown in Eq. (16). According to the ratings, users with scores in the top 10% are set to a high protection level, while the remaining users are set to a low protection level. The fuzzy comprehensive evaluation algorithm is shown in Algorithm 2. First,

use cosine similarity to classify the data x (line 1) and establish a comprehensive evaluation factor set I (line 2). Second, the lines 3–12 calculate the weight of each evaluation factor with the entropy weight method. Then the weight matrix A and the fuzzy membership matrix R are used to obtain the fuzzy comprehensive evaluation matrix B (line 13). Finally, B and the evaluation set V are used to calculate the protection level score of each user (line 14) and return them (line 15).

4.3 Protecting User Electricity Privacy Based on Hierarchical Threshold Encryption Algorithm

After obtaining the protection levels of electricity users, a hierarchical threshold encryption algorithm based on AMICFA is proposed with different electricity users as protection targets. The specific process of hierarchical threshold encryption algorithm is shown in Fig. 4.

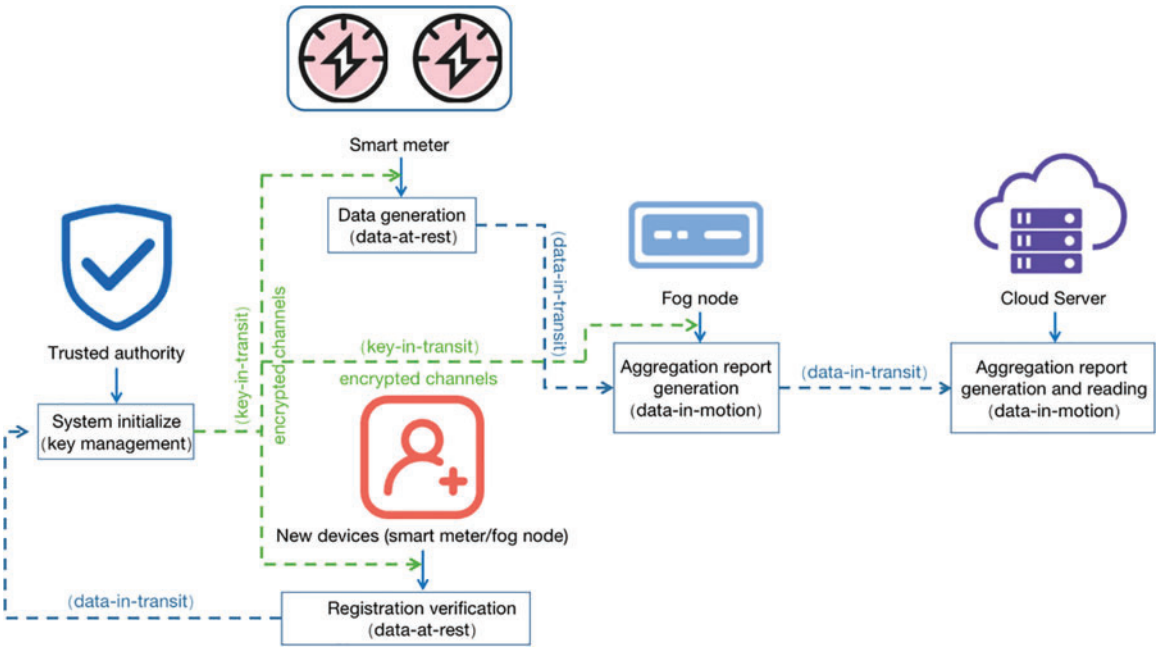

Figure 4: The process of hierarchical threshold encryption algorithm

4.3.1 System Parameter Initialization

At this stage, initializing the algorithm parameters with a robust key management strategy for data-at-rest encryption. Assuming that Trusted Authority (TA) has a global key pool $\{k_{ij}, k_j; 0 \leq i \leq w; 0 \leq j \leq m\}$, where keys are regularly rotated and updated to ensure the security of data-at-rest. Distinct keys from the pool are distributed to smart meters and fog computing devices by TA for secure registration and data-at-rest encryption. The specific generation process of each parameter is as follows:

(1) As part of the initialization, TA selects two secure prime numbers p and q. They are the basis for generating other encryption parameters. In the Paillier encryption system, these two prime numbers are used to calculate the modulus N, which is $N = pq$. Modulus N is the core part of subsequent encryption and decryption operations, and by ensuring the security and robustness of modulus N, it can effectively protect encrypted data from being cracked. The security of data-at-rest depends on the strength of these basic encryption parameters.

(2) λ is the least common multiple of $p-1$ and $q-1$, used to calculate the secret parameters in Paillier decryption process. The choice of λ is directly related to the security of the entire encryption system. If λ is leaked, the entire encryption system will be cracked. Therefore, it is possible to regularly replace the relevant prime numbers p and q to update λ, in order to prevent the risks caused by long-term use of the same key combination.

(3) The system randomly selects an integer g that satisfies $gcd\left(L\left(g^{\lambda} mod N^2\right), N\right) = 1$ to obtain Paillier public key (g, N). Due to the random selection of g, it becomes extremely difficult to attack encrypted data, ensuring the security of the data-at-rest encryption process.

(4) Individual keys α_{ij} for each smart meter and β_j for each fog node are generated. These keys can be used to encrypt data-in-transit or data-at-rest, they are regularly updated. Only entities with corresponding keys can access and interpret the data to prevent unauthorized access. This ensures that data is not tampered with during the transmission from the smart meters to the fog nodes and from the fog nodes to the server.

(5) Threshold pieces quantity c_1 and c_2 are generated, which can be used to define specific conditions that need to be met during the encryption and decryption process. c_1 represents the basic threshold condition to initiate the decryption process, while c_2 represents the threshold condition that needs to be met at a higher level of protection. The threshold encryption algorithm enhances the security of data-in-motion.

(6) Upon successful generation of the system parameters $\left(N, \lambda, g, \alpha_{ij}, \beta_j, c_1, c_2\right)$, TA will publish the public parameters (g, N). These parameters are necessary for encryption operations, and anyone can use these public parameters to encrypt data. For private parameters, encrypted channels are usually used to securely distribute these keys to the corresponding smart meters and fog nodes. This process ensures the security of the keys during transmission.

4.3.2 Registration Verification

After initialization, the newly added smart meters need to undergo registration verification. The specific steps are as follows:

(1) The smart meter will generate information RM_{ij} through its built-in algorithm as shown in Eq. (17), RM_{ij} mainly including the smart meter ID_{ij}, household information $Hold_{ij}$, location information Loc_{ij}. The registration information has unique identification.

$$RM_{ij} = \left\{ID_{ij}, Hold_{ij}, Loc_{ij}\right\} \tag{17}$$

(2) To safeguard the confidentiality of RM_{ij} during transmission, the smart meter encrypts this information using the public key (g, N) and an encryption parameter α_{ij} provided by TA. The resulting ciphertext, shown in Eq. (18), is securely transmitted by encrypted channel to TA, ensuring that registration data is protected in transit.

$$C_{ij} = g^{RM_{ij}} \alpha_{ij}^{N} mod N^2 \tag{18}$$

(3) To ensure the integrity and authenticity of the encrypted registration information while in transit, a Hash-based Message Authentication Code (HMAC) is generated. The HMAC algorithm is a key based verification method for message integrity, which generates a fixed length message digest as output by combining the key and message input. The communication parties confirm the legitimacy of the message by comparing this authentication code, and its security is based on the hash encryption algorithm used [41]. In order to achieve higher security, SHA-256 is used as the hash function in

this paper. Taking the private key α_{ij} and RM_{ij} as input for the hash function to generate a message verification code from the ciphertext, as shown in Eq. (19), and sent to TA.

$$MAC_{ij} = H\left(\left(\alpha_{ij} \oplus opad\right) | H\left(\left(\alpha_{ij} \oplus ipad\right) | RM_{ij}\right)\right) \tag{19}$$

where H represents SHA-256, *ipad* represents internal padding, and *opad* represents external padding.

(4) Upon receipt of the registration data and MAC, TA decrypts the information and verifies the MAC by recomputing it. If the received MAC matches the computed one, the registration is confirmed as legitimate, and a successful verification message is sent back to the smart meter. This two-fold verification process ensures that both the identity of the smart meters and the integrity of the transmitted data are secured, thus effectively protecting data-in-transit from interception or tampering. If the MAC does not match, the registration is denied. Similarly, the addition of new fog node devices also requires registration verification.

4.3.3 Electricity Data Encryption

After completing the verification, to prevent real-time electricity data exposure, it is necessary to encrypt the electricity data. The specific steps are as follows:

(1) Smart meters generate electricity consumption data for a given time period T_{ij}, as shown in Eq. (20). This data is immediately encrypted to protect it while in transit and in motion, as detailed in Eq. (21). The encrypted data, along with a digitally signed signature, is then uploaded to the fog nodes for authentication, ensuring that only verified data is processed and aggregated by the fog nodes.

$$M_{ij} = \left\{ID_{ij}, EC_{ij}, T_{ij}\right\} \tag{20}$$

$$C_{ij} = g^{EC_{ij}} \alpha_{ij}^{N} mod N^2 \tag{21}$$

(2) To mitigate the risk of replay attacks, a pseudo-random number R_{ij} is generated by combining the current timestamp T_{ij} with the smart meter's identity. This unique identifier ensures that each transmission is distinct and securely linked to its point of origin, further securing data-in-motion.

(3) At the fog nodes, further threshold encryption is employed based on the privacy of the data, a method that reinforces data-in-motion security by adapting the encryption depth according to the user protection levels determined by their *ID*. The protection level for each fog node (fog_{pl}) is elevated:

- If a user in the fog node is judged as a high protection level user, the value of fog_{pl} is increased by one.
- If a user in the fog node is judged as an ordinary user, the value of fog_{pl} remains unchanged.

Algorithm 3: Hierarchical threshold encryption algorithm

Input:	User electricity consumption behavior and electricity consumption information I_{user}, User protection level rating S
Output:	Privacy protected I'_{user}
1	// Generate system parameters.
2	$p, q \leftarrow$TA selects two secure prime numbers;
3	$N \leftarrow p * q$;
4	$\lambda \leftarrow lcm\,(p-1, q-1)$;
5	// Defined function.
6	$L\,(x) = (x-1)\,/N$;

(Continued)

Algorithm 3 (continued)

// Generate random integer.
$\alpha_{ij} \leftarrow random\,(Z)$;
$\beta_j \leftarrow random\,(Z)$;
$c_1, c_2 \leftarrow$Generate by TA;
Publish system parameters $(N, \lambda, g, \alpha_{ij}, \beta_j, c_1, c_2)$ to various entities;
// Registration verification.
// Registration information.
$RM_{ij} \leftarrow \{ID_{ij}, Hold_{ij}, Loc_{ij}\}$;
$C_{ij} \leftarrow g^{RM_{ij}} \alpha_{ij}^{N} mod N^2$;
$MAC_{ij} \leftarrow H\left(\left(\alpha_{ij} \oplus opad\right) | H\left(\left(\alpha_{ij} \oplus ipad\right) | RM_{ij}\right)\right)$;
if $MAC' == MAC$ **then**
  Agree to register;
else
  Refuse registration;
end
// Smart meter data encryption;
$M_{ij} \leftarrow \{ID_{ij}, EC_{ij}, T_{ij}\}$;
$C_{ij} \leftarrow g^{EC_{ij}} \alpha_{ij}^{N} mod N^2$;
$R_{ij} \leftarrow \{ID_{ij} \cup T_{ij}\}$;
$MAC_{ij} \leftarrow H\left(\left(R_{ij} \oplus opad\right) | H\left(\left(R_{ij} \oplus ipad\right) | M_{i,j}\right)\right)$;
Send user reports to fog nodes;
Signature authentication for fog nodes;
if ID is the 10% in $\boldsymbol{S}$ **then**
  $fog_{pl} = fog_{pl} + 1$;
else
  $fog_{pl} = fog_{pl} + 0$;
end
$Tss_j \leftarrow$According to different fog_{pl} values, different threshold encryption strategies are adopted for different fog nodes;
$I'_{user} \leftarrow$Protect user data on fog nodes based on different threshold encryption strategies Tss_j;
Return I'_{user}

(4) The protection levels of fog nodes are judged based on fog_{pl}. If it is a regular fog node, it only needs to meet the basic decryption criteria c_1 on the cloud server to decrypt, that is, it only needs to meet the minimum number $C\,(m-1, n)$ of key fragments, where m represents the total number of key fragments involved, n is the actual number of key fragments required during the decryption process. If it is a high protection level fog node device, it can only be decrypted if the cloud server meets the stringent decryption criteria c_2. Even if the basic number of decryption fragments is met, additional verification or key fragments from another important fog node of the same level are also required to finally decrypt the data. This threshold encryption strategy effectively improves the security of the system, especially in complex systems involving multiple nodes and different security levels, ensuring that sensitive data can only be accessed and decrypted when all set conditions are met. The privacy protection process based on hierarchical threshold encryption is shown in Algorithm 3. First, the TA generates system parameters and publishes them to corresponding entities (lines 1–11). Then, lines

12–21 complete the registration verification of the newly added smart meters and fog nodes. Next, lines 23–24 handle the encryption of electricity data on the smart meters, while lines 25–28 ensure data protection during transmission between the meters and fog nodes. Lines 29–33 determine the protection level of the fog nodes based on users' protection levels. Finally, lines 34–35 assign different threshold encryption strategies to different fog nodes according to their protection levels, and line 36 returns the encrypted user data.

5 Experiments

5.1 Experimental Environment

In this section, the method proposed in this paper is simulated and analyzed. The simulation platform utilized is PyCharm, employing Python 3.8. The code runs on the Windows 11 operating system with a hardware environment comprising an Intel Core i5-4200H processor with 2.80 GHz and 8 GB of memory. The experimental dataset used is based on the database called "Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States" [42]. This dataset encompasses the collection of smart meter data from diverse types of users in 95 cities across the United States, with a collection period of one year and a collection interval of one hour. The data format is shown in Table 1.

Table 1: Partial data from electricity load data from various states in the USA

UserID	Time	Facility	Fans	Cooling	Heating	...
0	1/2(01:00:00)	105.56575	23.04534406	11.6249027	8.65660857	...
0	1/2(02:00:00)	95.968336	19.27497843	9.215620549	9.411857538	...
0	1/2(03:00:00)	113.86157	24.09372304	10.62250021	14.1056054	...
0	1/2(04:00:00)	99.424518	19.27497843	7.599902891	12.68124854	...
0	1/2(05:00:00)	142.72785	28.48931737	9.572050659	14.89920319	...
0	1/2(06:00:00)	173.05479	28.07571151	12.45216031	43.1874417	...

5.2 Evaluation Indicators

5.2.1 Anti-Attack Success Rate

Due to its own characteristics, ensuring the safe and stable operation of smart grids is of utmost importance when confronted with various security threats. In order to verify the defense ability of the proposed model against network attacks, the Anti-attack success rate is defined in this paper as shown in Eqs. (22)–(24) to verify the effectiveness of HPPM-AMICFA.

$$P_{node} = 1 - \frac{1}{3}\left[\left(\prod_{i=1}^{n} \alpha_i \sigma_i\right)/n + \left(\prod_{j=1}^{m} \beta_j \sigma_i\right)/m + \gamma\right] \tag{22}$$

$$P_{link} = 1 - \frac{1}{2}\left[\left(\prod_{i=1}^{n} \varepsilon_i \sigma_i\right)/n + \left(\prod_{j=1}^{m} \delta_j \sigma_i\right)/m\right] \tag{23}$$

$$P_{success} = \frac{P_{node} + P_{link}}{2} \tag{24}$$

Assuming equal likelihood of attackers targeting the cloud server, fog nodes, and smart meters, let's denote the total number of smart meters as n and the total number of fog nodes in the power

grid as m. The probabilities of successfully attacking smart meters, fog nodes, and the cloud server are represented by α, β, and γ, respectively. When attackers target the links between entities, let ε denote the probability of successfully intercepting the links between smart meters and fog nodes, and δ denote the probability of successfully intercepting the links between fog nodes and the cloud server. Additionally, the probability of the encryption key being cracked is σ.

The Anti-attack success rate can be defined as the average of the success rates of nodes and links resisting attacks. A higher value indicates a stronger anti-attack capability of the smart grid privacy protection scheme, while a lower value suggests lower security.

5.2.2 Calculate Costs

The computational cost of privacy protection in smart grid applications is a critical consideration due to the large volume of data involved. Achieving privacy protection with minimal computational overhead while ensuring security is essential for practical deployment. The calculation of costs typically encompasses the following aspects: (1) In the encryption process, smart meters encrypt electricity consumption data, and after aggregation, the data is encrypted again at the fog nodes. (2) In the aggregation stage, operations are performed on encrypted data to aggregate ciphertexts. (3) In the decryption process, cloud server and fog nodes decrypt ciphertexts to analyze user data after obtaining encrypted data.

5.3 Baselines

Four baseline models are used to compare the performance of the HPPM-AMICFA on the above two evaluation indicators:

1. Paillier encryption algorithm [43]: It is an encryption algorithm based on the composite residue class hard problem hypothesis. It mainly consists of three parts: key generation, encryption, and decryption.
2. Privacy preserving Data Aggregation against False data (PDAF) [44]: It is based on Paillier homomorphic encryption scheme and uses blinding factors to design a privacy protection method to protect privacy in fog computing.
3. Security Enhanced Data Aggregation (SEDA) [45]: It is a data aggregation scheme for smart grid communication security enhancement based on homomorphic cryptosystem, trapdoor hash functions and homomorphic authenticators.
4. Threshold encryption [36]: It is a distributed cryptographic technology based on multi-party participation. It can divide a key into multiple parts, and each part can only restore the complete key when it reaches a certain threshold.

5.4 Performance Analysis

5.4.1 Cluster Analysis of Electricity Consumption Characteristics

Based on the real test data of electricity loads in various states in the United States, fuzzy comprehensive analysis algorithm is utilized to set the protection level of users in this paper. In order to obtain the evaluation index of the fuzzy comprehensive analysis algorithm, users need to be classified according to the electricity consumption behavior characteristics, so as to obtain the electricity consumption characteristics of each type of user as one of the inputs of the algorithm.

The actual load curves of the power grid are shown in Fig. 5. Using the K-means method to cluster the electricity consumption of users, as shown in Fig. 6, it can be seen that users are categorized into

four groups. Subsequently, features are extracted from the *Ppr*, *Tpr*, *Lr*, and *TE* of each user's power load curve, allowing each user to be classified into its most similar category.

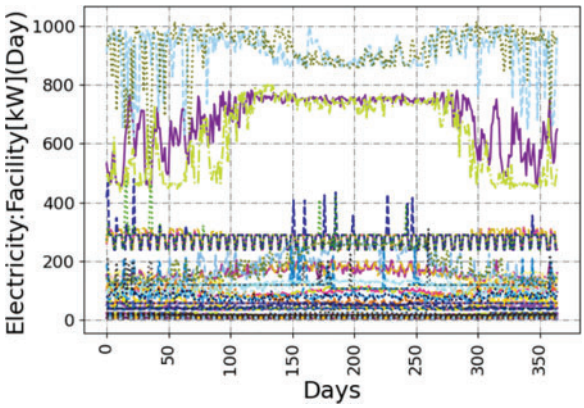

Figure 5: Partial actual electricity load curves in the dataset. The curves of different colors represent the electricity consumption characteristics of different users, with the horizontal axis representing the date and the vertical axis representing the total daily electricity load

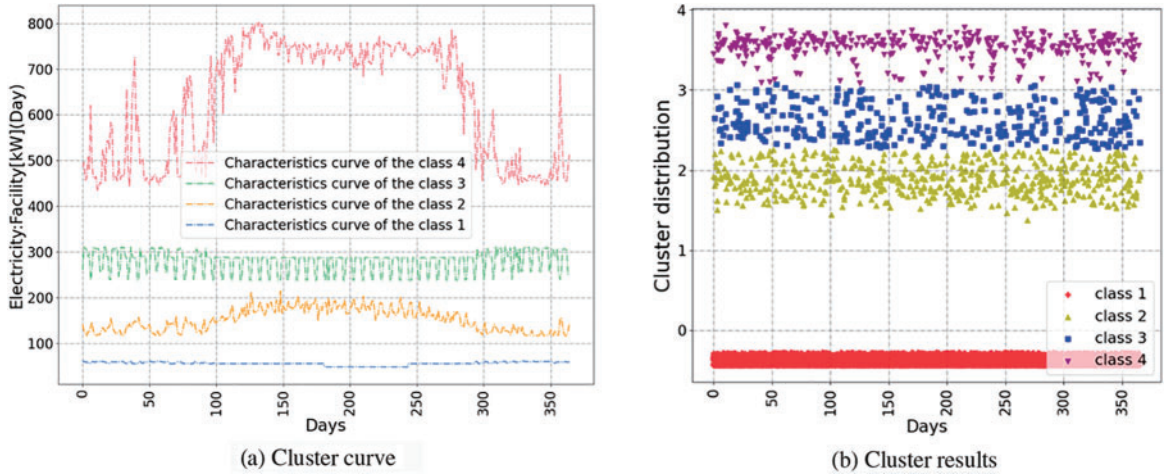

Figure 6: Clustering results of electricity consumption characteristics. (a) Indicating the daily electricity consumption characteristic curves of four types of users. (b) Indicating the results of clustering the daily electricity consumption characteristics of users

Each clustering result is optimized by weighted average, and different clustering results are categorized into four types of electricity consumption behaviors: Household electricity, Commercial electricity, Industrial electricity, and Unused electricity as shown in Fig. 7. These will be used as input to the fuzzy comprehensive analysis algorithm together with the user electricity consumption levels.

5.4.2 Security Analysis

In the AMICFA model, user electricity data is susceptible to attacks during transmission, and fog nodes and cloud server also face the risk of being compromised. Therefore, in order to protect user privacy data from leakage, the hierarchical threshold encryption algorithm is proposed in this paper.

The primary objective of implementing variable thresholds is to provide differentiated protection levels that are aligned with the risk profiles and privacy needs of various user categories. By design, this method does not merely adjust the number of parties needed for decryption but strategically enhances the security barriers based on the potential impact and vulnerability of the data involved. Before conducting a comparative analysis of the experimental results, a theoretical analysis of the security of the HPPM-AMICFA model is conducted and the advantages are listed below:

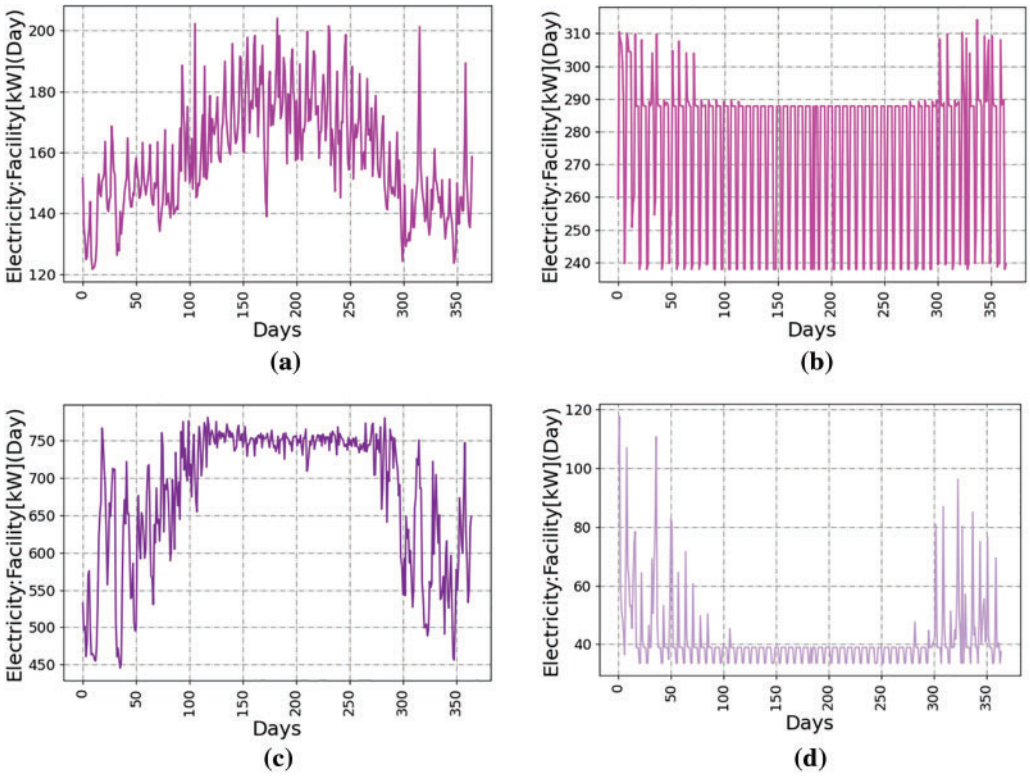

Figure 7: Four types of electricity consumption. (a) Household electricity. (b) Commercial electricity. (c) Industrial electricity. (d) Unused electricity

(1) Enhanced security for sensitive data. For users with higher risk profiles—such as commercial or industrial entities where data exposure could lead to significant financial or operational repercussions—the system requires a higher threshold. This means more shares must be present to decrypt, effectively increasing the difficulty for unauthorized access or malicious collusion.

(2) Flexibility and scalability. This approach allows the system to dynamically adjust the security measures based on real-time assessments of risk and privacy requirements. It provides a robust framework that can evolve as threats landscape changes or as different user needs emerge.

(3) Proportional to the data's sensitivity. Ensuring that more sensitive data has stronger protections inherently supports privacy by reducing the likelihood of unauthorized access.

(4) Adaptive to threat models. The method acknowledges that not all data is equally attractive to attackers, and not all users face the same level of threat. Adjusting the decryption threshold according to the assessed risk aligns the security measures with the actual needs and threats.

Therefore, while the hierarchical threshold mechanism may superficially appear as merely altering the collaborative requirements for decryption, it introduces an additional layer of security, which fundamentally enhances user data privacy and security. The advantages of HPPM-AMICFA will be verified through the following experimental results analysis.

5.4.3 Anti-Attack Success Rate Analysis

In the analysis of the anti-attack success rates, the robustness of the HPPM-AMICFA scheme against tampering attacks is focused on, where attackers aim to capture nodes or manipulate data links to access and alter user privacy data. Since HPPM-AMICFA can be regarded as an improved combination of Paillier and Threshold encryption, a detailed comparison is provided to illustrate its superior effectiveness in mitigating threats. This comparison includes the widely used Paillier and Threshold encryption algorithms under similar attack conditions. Additionally, HPPM-AMICFA is compared with PDAF, which also demonstrates good performance in cloud-fog assistance systems. These comparisons further highlight the security and efficiency of HPPM-AMICFA in privacy protection within cloud-fog assistance systems.

Firstly, assuming it is an attack by ordinary criminals who lacks knowledge about the power grid's background. These attackers randomly target nodes, posing risks to both smart meters and fog nodes. Various configurations of fog nodes and smart meters across different sub-regions are simulated, with the initial attack success probabilities α, β, and γ for each type of node set to 0.3, 0.2, and 0.1, respectively. Similarly, the initial attack success probabilities ε and δ for each type of link are set to 0.2 and 0.2, respectively. These settings are used to measure and compare the success rates of both the baseline algorithms and the HPPM-AMICFA scheme in thwarting these attacks. As shown in Fig. 8, a distinct advantage in using HPPM-AMICFA is revealed, particularly due to its threshold encryption mechanism at the fog nodes. This mechanism ensures that unless the number of compromised nodes reaches a critical threshold, the decryption of user data remains unfeasible, thus significantly enhancing data security.

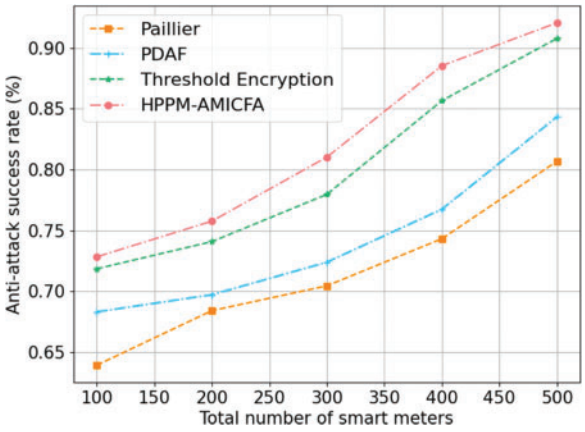

Figure 8: Success rate of resisting attacks without background knowledge

In simulations involving 100 to 500 smart meters, HPPM-AMICFA consistently outperformed the other baseline models in resisting attacks, with success rates ranging from 73% to 92%, showing an average improvement of about 14% compared to the Paillier algorithm. Specifically, when the total number of smart meters remains unchanged, the two algorithms based on threshold encryption consistently perform better than Paillier and PDAF, with the performance difference becoming more

pronounced as the number of smart meters increases. This is because, with the threshold encryption algorithm, even if some node keys are cracked, the attacker cannot obtain useful information as long as the decryption threshold is not reached. This method significantly reduces the probability of successful key cracking. Additionally, although the anti-attack success rates of HPPM-AMICFA are similar to those of traditional threshold encryption algorithms, HPPM-AMICFA consistently performs better. This is due to HPPM-AMICFA's more flexible threshold encryption mechanism, which employs different threshold encryption strategies for different nodes, thereby increasing the difficulty of decryption and ensuring the security of more important data. Furthermore, compared to PDAF, HPPM-AMICFA demonstrates superior effectiveness in protecting the privacy of cloud-fog assistance systems in smart grids.

Furthermore, when considering attacks by adversaries with background knowledge targeting specifically important users, the initial values of attack success probabilities α, β, and γ for each type of node involved in the experiment are set to 0.5, 0.4, and 0.3, respectively, and for each type of link, the probabilities ε and δ are set to 0.4 and 0.4, respectively. The Anti-attack success rates of the two protection methods are as shown in Fig. 9. It can be seen that, compared to the other three baselines, HPPM-AMICFA still demonstrates the best performance, with an overall average improvement of 18.5% compared to the Paillier algorithm. However, this time, the performance of traditional threshold encryption algorithms does not match that of HPPM-AMICFA, and the gap between them increases as the number of smart meters rises. This is because when attackers possess background knowledge, they can target specific important user data. Among these methods, only HPPM-AMICFA classifies user importance and applies more stringent threshold encryption strategies for more important users. Moreover, as the number of smart meters increases, the number of important users also grows, making the advantages of HPPM-AMICFA more evident.

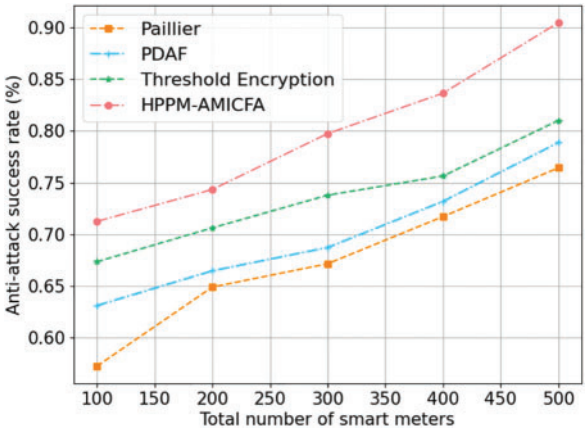

Figure 9: Success rate of resisting attacks with background knowledge

This result indicates that the enhanced ability of HPPM-AMICFA to resist complex attacks and the scalability of HPPM-AMICFA, suggesting its growing effectiveness as the network expands. This trend is crucial for smart grid applications where the number of connected devices consistently rises, requiring robust security mechanisms that can adapt to increasing complexity and potential vulnerabilities.

These findings clearly demonstrate the superior anti-attack capabilities of HPPM-AMICFA over traditional encryption methods, particularly in environments susceptible to both random and targeted

attacks. By incorporating hierarchical encryption thresholds and leveraging cloud and fog assistance architecture, HPPM-AMICFA not only enhances the security of user data but also ensures that the smart grid remains resilient against evolving cyber threats. This comparative analysis underscores the significant improvements HPPM-AMICFA offers over existing privacy protection methods, affirming its potential for widespread adoption in critical energy infrastructure protection.

5.4.4 Computational Costs Analysis

To demonstrate the efficiency of the HPPM-AMICFA scheme in handling computational costs, particularly in large-scale smart grid environments, it is compared with other baseline models. For a more comprehensive experimental comparison, Paillier is replaced with SEDA, which is based on a homomorphic cryptosystem. Thus, HPPM-AMICFA is compared with three baseline models based on different encryption algorithms in terms of computational costs. This comparison validates the effectiveness of HPPM-AMICFA in optimizing computational resources without compromising security.

As shown in Fig. 10, HPPM-AMICFA exhibits significantly lower computational overhead as the number of smart meters scales up, maintaining robustness and efficiency. This is evident when comparing the total time costs across varying scales of deployment. For instance, at 500 smart meters, HPPM-AMICFA demonstrates a 55.3% reduction in computational costs compared to SEDA, which records a time cost of 2550 ms, whereas HPPM-AMICFA only requires 1140 ms. Such efficiency is achieved through the optimized cryptographic processes employed by HPPM-AMICFA, which simplifies computational tasks without reducing cryptographic strength.

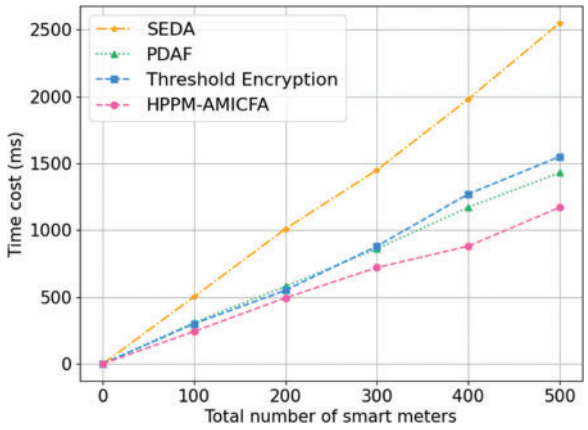

Figure 10: Comparison of computational costs of different algorithms

The time costs of the two algorithms, PDAF and Threshold encryption, are relatively close overall. However, as the number of smart meters increases, PDAF shows slightly better performance than Threshold encryption. This is because the PDAF algorithm uses a batch verification mechanism, which improves the efficiency of verifying multiple encrypted data points. Despite this, the time cost of these two algorithms is generally higher than that of HPPM-AMICFA for the following reasons.

Throughout the entire execution of the method, the computational cost of hash operations can be disregarded. For n smart meter users, utilizing the improved Paillier encryption algorithm (HPPM-AMICFA), each meter necessitates two multiplication operations and one exponential operation to generate its ciphertext and signature, respectively. Conversely, in the other two algorithms SEDA

and PDAF, each meter requires $2n$ exponential operations for encryption. In the fog nodes, data verification entails m multiplication and exponential operations, while secret sharing requires m multiplication operations. Following this, a signature is generated post an exponential operation, after which the data is uploaded to the cloud server. Upon the cloud server's attempt to perform decryption operations, it initially needs to aggregate fog nodes data, which, if the number meets the threshold value, requires t multiplication operations. The final decryption operation includes a power operation and a multiplication operation.

For the traditional threshold encryption algorithms, multiple key fragments need to be distributed to different nodes, which increases the communication overhead during the initial setup. Additionally, decryption necessitates the participation of multiple nodes to collect enough key fragments, which means each decryption requires extra communication. In contrast, HPPM-AMICFA achieves hierarchical protection and aggregated data transmission through the collaboration of cloud servers and fog nodes. This approach optimizes keys distribution and nodes participation, thereby reducing communication overhead.

In smart grid systems, where real-time data processing and frequent communication are essential, the higher computational costs of SEDA, PDAF, and Threshold encryption may not be justifiable in all scenarios, especially in systems where rapid data processing is crucial. HPPM-AMICFA offers an optimal balance between security and efficiency, making it particularly suitable for modern smart grids that face varied and sophisticated cyber threats.

6 Conclusions and Outlooks

In response to the potential exposure of user habits and infringement of user privacy during data transmission in the AMI system of smart grids, a secure advanced metering infrastructure based on cloud and fog assistance in smart grids is designed in this paper. Based on this model, a hierarchical privacy protection method based on threshold encryption is proposed. The effectiveness of this algorithm in protecting the privacy and security of smart grid data has been verified through security analyses and experimental simulations.

The HPPM-AMICFA model sets the protection levels of smart grid users using fuzzy comprehensive analysis with entropy weight method and provides privacy encryption protection with the assistance of fog nodes. While this approach is suitable for the user-side of the smart grid, more nuanced privacy protection strategies are required for the supply-side of the grid, where employees with varying roles are present. Therefore, a privacy protection strategy under a partial order structure for the power grid supply side will be focused on to achieve multi-level privacy protection in the future.

Acknowledgement: The authors grate fully acknowledge the helpful comments and suggestions of the editors and reviewers, which have improved the presentation.

Funding Statement: This research was funded by the National Natural Science Foundation of China (Grant Number 61902069), Natural Science Foundation of Fujian Province of China (Grant Number 2021J011068), Research Initiation Fund Program of Fujian University of Technology (GY-S24002, GY-Z21048), Fujian Provincial Department of Science and Technology Industrial Guidance Project (Grant Number 2022H0025).

Author Contributions: The authors confirm contribution to the paper as follows: Study conception and design: Linghong Kuang, Wenlong Shi; data collection: Linghong Kuang, Wenlong Shi, Jing

Zhang; analysis and interpretation of results: Wenlong Shi, Jing Zhang; draft manuscript preparation: Linghong Kuang, Wenlong Shi. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author upon request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. H. Ullah and J. D. Park, "Distributed energy trading in smart grid over directed communication network," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3669–3672, Jul. 2021. doi: 10.1109/TSG.2021.3067172.

[2] K. Sharma, A. Malik, I. Batra, A. S. M. Sanwar Hosen, M. A. Latif Sarker and D. S. Han, "Technologies behind the smart grid and internet of things: A system survey," *Comput. Mater. Contin.*, vol. 75, no. 3, pp. 5049–5072, 2023. doi: 10.32604/cmc.2023.035638.

[3] A. M. Alkhiari, S. Mishra, and M. AlShehri, "Blockchain-based SQKD and IDS in edge enabled smart grid network," *Comput. Mater. Contin.*, vol. 70, no. 2, pp. 2150–2169, 2022. doi: 10.32604/cmc.2022.019562.

[4] J. Wang, L. Wu, S. Zeadally, M. K. Khan, and D. He, "Privacy-preserving data aggregation against malicious data mining attack for IoT-enabled smart grid," *ACM Trans. Sens. Netw.*, vol. 17, no. 3, pp. 1–25, Jun. 2021. doi: 10.1145/3440249.

[5] H. Wang, D. He, and S. Zhang, "Balanced anonymity and traceability for outsourcing small-scale data linear aggregation in the smart grid," *IET Inf. Secur.*, vol. 11, no. 3, pp. 131–135, May 2017. doi: 10.1049/iet-ifs.2016.0150.

[6] C. Hu, Z. Liu, R. Li, P. Hu, T. Xiang and M. Han, "Smart contract assisted privacy-preserving data aggregation and management scheme for smart grid," *IEEE Trans. Dependable Secure Comput.*, vol. 21, no. 4, pp. 1–17, Aug. 2023. doi: 10.1109/TDSC.2023.3300749.

[7] S. M. Je, H. Woo, J. Choi, S. H. Jung, and J. H. Huh, "A research trend on anonymous signature and authentication methods for privacy invasion preventability on smart grid and power plant environments," *Energies*, vol. 15, no. 12, pp. 4363–4383, Jun. 2022. doi: 10.3390/en15124363.

[8] L. Ou, Z. Qin, S. Liao, T. Li, and D. Zhang, "Singular spectrum analysis for local differential privacy of classifications in the smart grid," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5246–5255, Feb. 2020. doi: 10.1109/JIOT.2020.2977220.

[9] S. Zhao *et al.*, "Smart and practical privacy-preserving data aggregation for fog-based smart grids," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 521–536, Aug. 2020. doi: 10.1109/TIFS.2020.3014487.

[10] S. Desai, R. Alhadad, N. Chilamkurti, and A. Mahmood, "A survey of privacy preserving schemes in IoE enabled smart grid advanced metering infrastructure," *Cluster Comput.*, vol. 22, no. 1, pp. 43–69, Mar. 2019. doi: 10.1007/s10586-018-2820-9.

[11] P. O. Ajiboye, K. O. B. O. Agyekum, and E. A. Frimpong, "Privacy and security of advanced metering infrastructure (AMI) data and network: A comprehensive review," *J. Eng. Appl. Sci.*, vol. 71, no. 1, pp. 91, Apr. 2024. doi: 10.1186/s44147-024-00422-w.

[12] S. Chen, L. Yang, Y. Shi, and Q. Wang, "Blockchain-enabled secure and privacy-preserving data aggregation for fog-based ITS," *Comput. Mater. Contin.*, vol. 75, no. 2, pp. 3781–3796, Feb. 2023. doi: 10.32604/cmc.2023.036437.

[13] H. Yang, S. Liang, X. Luo, D. Tang, H. Li and X. Shen, "PIPC: Privacy- and integrity-preserving clustering analysis for load profiling in smart grids," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 10851–10861, Nov. 2021. doi: 10.1109/JIOT.2021.3125674.

[14] S. Zhang, J. Rong, and B. Wang, "A privacy protection scheme of smart meter for decentralized smart home environment based on consortium blockchain," *Int. J. Electr. Power Energy Syst.*, vol. 121, no. 10, pp. 106140, Oct. 2020. doi: 10.1016/j.ijepes.2020.106140.
[15] A. Guan and D. J. Guan, "An efficient and privacy protection communication scheme for smart grid," *IEEE Access*, vol. 8, pp. 179047–179054, Aug. 2020. doi: 10.1109/ACCESS.2020.3025788.
[16] J. Srinivas, A. K. Das, X. Li, M. K. Khan, and M. Jo, "Designing anonymous signature-based authenticated key exchange scheme for internet of things-enabled smart grid systems," *IEEE Trans. Ind. Inform.*, vol. 17, no. 7, pp. 4425–4436, Jul. 2021. doi: 10.1109/TII.2020.3011849.
[17] L. Wu, W. Zhang, and W. Zhao, "Privacy preserving data aggregation for smart grid with user anonymity and designated recipients," *Symmetry*, vol. 14, no. 5, pp. 847, Apr. 2022. doi: 10.3390/sym14050847.
[18] S. Chen, L. Yang, C. Zhao, V. Varadarajan, and K. Wang, "Double-blockchain assisted secure and anonymous data aggregation for fog-enabled smart grid," *Engineering*, vol. 8, no. 3, pp. 159–169, Jan. 2022. doi: 10.1016/j.eng.2020.06.018.
[19] K. S. Adewole and V. Torra, "DFTMicroagg: A dual-level anonymization algorithm for smart grid data," *Int. J. Inf. Secur.*, vol. 21, no. 6, pp. 1299–1321, Sep. 2022. doi: 10.1007/s10207-022-00612-8.
[20] O. R. Merad Boudia, S. M. Senouci, and M. Feham, "Elliptic curve-based secure multidimensional aggregation for smart grid communications," *IEEE Sens. J.*, vol. 17, no. 23, pp. 7750–7757, Dec. 2017. doi: 10.1109/JSEN.2017.2720458.
[21] C. Dwork, "Differential privacy," in *Int. Colloq. Autom., Lang., Program.*, Berlin, Heidelberg, 2006, pp. 1–12.
[22] J. Zhao, T. Jung, Y. Wang, and X. Li, "Achieving differential privacy of data disclosure in the smart grid," in *IEEE INFOCOM, 2014-IEEE Conf. Comput. Commun.*, Toronto, ON, Canada, 2014, pp. 504–512.
[23] M. B. Gough, S. F. Santos, T. AlSkaif, M. S. Javadi, R. Castro and J. P. S. Catalão, "Preserving privacy of smart meter data in a smart grid environment," *IEEE Trans. Ind. Inform.*, vol. 18, no. 1, pp. 707–718, Jan. 2022. doi: 10.1109/TII.2021.3074915.
[24] N. Gai, K. Xue, B. Zhu, J. Yang, J. Liu and D. He, "An efficient data aggregation scheme with local differential privacy in smart grid," *Digit. Commun. Netw.*, vol. 8, no. 3, pp. 333–342, Jun. 2022. doi: 10.1016/j.dcan.2022.01.004.
[25] Z. Zheng, T. Wang, A. K. Bashir, M. Alazab, S. Mumtaz and X. Wang, "A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid," *IEEE Trans. Comput.*, vol. 71, no. 11, pp. 2915–2926, Nov. 2022. doi: 10.1109/TC.2021.3130402.
[26] M. B. Yassein, S. Aljawarneh, E. Qawasmeh, W. Mardini, and Y. Khamayseh, "Comprehensive study of symmetric key and asymmetric key encryption algorithms," in *2017 Int. Conf. Eng. Technol. (ICET)*, Antalya, Turkey, 2017, pp. 1–7.
[27] R. Sarenche, M. Salmasizadeh, M. H. Ameri, and M. R. Aref, "A secure and privacy-preserving protocol for holding double auctions in smart grid," *Inf. Sci.*, vol. 557, no. 4, pp. 108–129, May 2021. doi: 10.1016/j.ins.2020.12.038.
[28] Y. Chen, J. F. Martínez-Ortega, P. Castillejo, and L. López, "A homomorphic-based multiple data aggregation scheme for smart grid," *IEEE Sens. J.*, vol. 19, no. 10, pp. 3921–3929, May 2019. doi: 10.1109/JSEN.2019.2895769.
[29] W. Xu, J. Sun, R. Cardell-Oliver, A. Mian, and J. B. Hong, "A privacy-preserving framework using homomorphic encryption for smart metering systems," *Sensors.*, vol. 23, no. 10, pp. 4746, May 2023. doi: 10.3390/s23104746.
[30] M. Faheem, H. Kuusniemi, B. Eltahawy, M. S. Bhutta, and B. Raza, "A lightweight smart contracts framework for blockchain-based secure communication in smart grid applications," *IET Gener., Transm. Distrib.*, vol. 18, no. 3, pp. 625–638, Jan. 2024. doi: 10.1049/gtd2.13103.
[31] Y. Wang, B. Chen, L. Li, Q. Ma, H. Li and D. He, "Efficient and secure ciphertext-policy attribute-based encryption without pairing for cloud-assisted smart grid," *IEEE Access*, vol. 8, pp. 40704–40713, 2020. doi: 10.1109/ACCESS.2020.2976746.

[32] K. Fan *et al.*, "MSIAP: A dynamic searchable encryption for privacy-protection on smart grid with cloud-edge-end," *IEEE Trans. Cloud Comput.*, vol. 11, no. 2, pp. 1170–1181, Apr.–Jun. 2023. doi: 10.1109/TCC.2021.3134015.

[33] L. Wu, S. Fu, Y. Luo, H. Yan, H. Shi and M. Xu, "A robust and lightweight privacy-preserving data aggregation scheme for smart grid," *IEEE Trans. Dependable Secure Comput.*, vol. 21, no. 1, pp. 270–283, Jan.-Feb. 2024. doi: 10.1109/TDSC.2023.3252593.

[34] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979. doi: 10.1145/359168.359176.

[35] L. J. Pang and Y. M. Wang, "A new (*t*, *n*) multi-secret sharing scheme based on Shamir's secret sharing," *Appl. Math. Comput.*, vol. 167, no. 2, pp. 840–848, Aug. 2005. doi: 10.1016/j.amc.2004.06.120.

[36] T. Tassa, "Hierarchical threshold secret sharing," *J. Cryptol.*, vol. 20, no. 2, pp. 237–264, Feb. 2007. doi: 10.1007/s00145-006-0334-8.

[37] M. Davoodi, R. Moslemi, W. Song, and J. M. Velni, "A fog-based approach to secure smart grids against data integrity attacks," in *2020 IEEE Power Energy Society Innov. Smart Grid Technol. Conf. (ISGT)*, Washington, DC, USA, 2020, pp. 1–5.

[38] M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani and N. Ghadimi, "A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future," *Elect. Power Syst. Res.*, vol. 215, pp. 108975, Feb. 2023. doi: 10.1016/j.epsr.2022.108975.

[39] G. Huang, "Missing data filling method based on linear interpolation and lightgbm," *J. Phys.: Conf. Ser.*, vol. 1754, pp. 012187, 2021. doi: 10.1088/1742-6596/1754/1/012187.

[40] S. Bissey, S. Jacques, and J. C. Le Bunetel, "The fuzzy logic method to efficiently optimize electricity consumption in individual housing," *Energies*, vol. 10, no. 11, pp. 1701, Oct. 2017. doi: 10.3390/en10111701.

[41] U. Cali, M. Kuzlu, M. Pipattanasomporn, J. Kempf, and L. Bai, "Introduction to security for smart grid systems," in *Digitalization of Power Markets and Systems Using Energy Informatics*, Springer, Cham, Sep. 2021, pp. 59–85. doi: 10.1007/978-3-030-83301-5_4.

[42] Ong, Sean, Clark, and Nathan, "Commercial and residential hourly load profiles for all TMY3 locations in the united states," 2014. Accessed: Dec. 22, 2023. [Online]. Available: http://en.openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3/

[43] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Int. conf. Theory Appl. Cryptogr. Tech.*, Berlin, Heidelberg, 1999, pp. 223–238.

[44] Y. Zhang *et al.*, "Privacy-preserving data aggregation against false data injection attacks in fog computing," *Sensors*, vol. 18, no. 8, pp. 2659–2675, Aug. 2018. doi: 10.3390/s18082659.

[45] J. Ni, K. Alharbi, X. Lin, and X. Shen, "Security-enhanced data aggregation against malicious gateways in smart grid," in *2015 IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, 2015, pp. 1–6.