Computer Fraud and Security
ISSN (online): 1873-7056

An Innovative Anti-Tampering Encryption Method for Enhancing Privacy Protection in Power Big Data

Guangqian Lu[1], Xiaowen Zeng[1], and Zhiyu Zhao[1]
[1]Information Center, Yunnan Power Grid Co., Ltd, Yunnan, China.

Abstract:
Power big data encompasses vast amounts of private information. Traditional encryption algorithms often exhibit longer computation times and greater resource demands when handling large-scale data. This not only heightens the risk of private information decryption but also diminishes encryption efficiency. To address these challenges, we propose a novel tamper-proof encryption method for power big data based on national security algorithms. Initially, we conduct an analysis of the composition, operational principles, and advantages of various national security algorithms. Subsequently, we select the Advanced Encryption Standard (AES) algorithm and enhance its performance through expansion and optimization. Finally, the improved AES algorithm is employed to achieve tamper-proof encryption of power big data privacy information, utilizing operations such as round transformation, byte substitution, and row matrix manipulation. Experimental results demonstrate that the proposed encryption method significantly reduces the likelihood of private information decryption while enhancing encryption efficiency. These findings underscore the effectiveness of our approach in achieving robust tamper-proof encryption for power big data.

Keywords: National Security Algorithm; Power big data; Privacy information; Anti tampering encryption

1 INTRODUCTION

As power systems undergo digitization and intelligent processing, the volume of power big data expands exponentially. This data encompasses various facets of power generation, transmission, and distribution, including sensitive user information such as energy consumption patterns and user behavior [1]. However, the proliferation of power big data also amplifies concerns regarding security and privacy. Particularly during data transmission and storage, there is a heightened risk of tampering and theft. Hence, it is imperative to investigate tamper-proof encryption methods to safeguard the privacy of power big data [2-3].

Power big data contains sensitive information about users, such as household electricity consumption, consumption habits, and so on. By studying tamper-proof encryption methods for private information, users' personal privacy can be effectively protected and user information cannot be leaked or abused [4]. The accuracy of power big data is crucial for the stable operation and decision-making of the power system. The adoption of tamper-proof encryption can effectively prevent data from being tampered with illegally during transmission and storage, and ensure data integrity [5]. As one of the key infrastructures, the safety of power system is very important. The study of tamper-proof encryption methods can improve the overall security of the power system, prevent malicious attackers from tampering with and destroying power big data, and ensure the normal operation of the power system [6]. Smart grid is an important direction for the future development of power system, and power big data is the basis for realizing smart grid. Research on tamper-proof encryption of private information can provide reliable data protection means for smart grid and promote the construction and application of smart grid.

To tackle the pressing challenges associated with securing power big data, we propose a tamper-proof encryption method rooted in national security algorithms. Our approach entails a systematic analysis of the composition, operational principles, and merits of various national security algorithms. This preliminary investigation provides crucial insights into selecting the most suitable algorithm for our purpose. Subsequently, we opt for the widely recognized Advanced Encryption Standard (AES) algorithm and embark on an endeavor to augment its efficacy through expansion and optimization strategies. This process involves refining the algorithm's underlying mechanisms to better align with the unique characteristics and demands of power big data. Specifically, we focus on enhancing AES's performance by refining its core components, such as round transformation, byte substitution, and row matrix manipulation. Through these enhancements, we aim to fortify AES's capabilities in safeguarding

the privacy of power big data against unauthorized access and manipulation. Moreover, we conduct a comprehensive evaluation of our proposed encryption method using rigorous experimental methodologies. By subjecting the method to various test scenarios and performance benchmarks, we aim to assess its effectiveness in mitigating the risks associated with data decryption and enhancing overall encryption efficiency. Our experimental results not only validate the efficacy of the proposed method but also provide valuable insights into its practical applicability and scalability.

The remainder of this article is organized as follows. We review the related work in Section 2. Section 3 describes the methods proposed in this paper. Section 4 reports the experimental results. Finally, we conclude the paper in Section 5.

2 RELATED WORK

In recent years, research in power data encryption techniques has seen significant advancements, driven by the increasing digitization of power systems and the critical need to secure sensitive energy-related information. These advancements aim to address the unique challenges posed by the large-scale collection, transmission, and storage of power data while ensuring confidentiality, integrity, and availability.

Ding et al. [7] introduced a power big data encryption technique employing an enhanced Blowfish algorithm, which employs the Blowfish algorithm for encrypting power big data and employs a 512-bit grouping processing approach. They devised a dynamic and static ciphertext knowledge proof mechanism by integrating the MD5 algorithm with the Schnorr identity authentication protocol. This mechanism facilitates encryption and decryption of power big data using the enhanced Blowfish algorithm. The authors implemented regular key updates triggered by a timer to ensure security, initiated key change events, and executed the data key update process using the Blowfish algorithm. Hash algorithms were employed to detect key changes, thereby preventing eavesdropping and data tampering during transmission. Moreover, additional identity information transmission was ensured for information data by encrypting the improved MD5 algorithm key with the Blowfish algorithm key, thereby enhancing the security of power big data. Li et al. [8] proposed a data encryption methodology based on the double hook function. This technique involves creating vertical lines on the asymptote and parallel lines on the X-axis, with the plaintext value corresponding to the number of times the vertical and parallel lines intersect alternately. The X or Y value of the final intersection point serves as the corresponding ciphertext. Furthermore, the authors designed symmetric and asymmetric encryption algorithms based on two characteristic parameters of the double hook function and the selection of the base point. They explored the potential of lattice encryption using any double hook curve function as the lattice basis to form a nonlinear lattice space. Zhao et al. [9] introduced a power big data encryption approach based on homomorphic encryption. This method involves collecting and transmitting power big data hourly, initially encrypting the data using a password encoder, and temporarily storing the encrypted data in a data aggregator. A key generator is utilized to generate a dedicated key for received power data, and homomorphic encryption technology is employed for power data encryption control. Lastly, Wang et al. [10] proposed a power big data encryption method based on deep learning. Their approach involves acquiring the requisite initial research data according to data extraction criteria, decrypting the data ciphertext based on deep learning parameters to obtain data on the degree of privacy, and training the decryption key to complete the de-privacy process of power big data. They developed a data encryption model at the data aggregation center, mined data features, cleansed them, selected sample specification data, searched for their feature target parameters, and completed the encryption of power big data.

Overall, the research progress in power data encryption techniques underscores the importance of ongoing innovation and collaboration in addressing the evolving security needs of modern energy systems. However, challenges persist in the domain of power data encryption. These include scalability concerns, particularly in large-scale power systems where massive volumes of data are generated continuously. Additionally, ensuring interoperability and compatibility across diverse power system architectures and communication protocols remains a challenge for encryption techniques.

3 METHOD OF POWER BIG DATA PRIVACY INFORMATION WITH NATIONAL SECURITY ALGORITHM

In this section, we present the anti-tampering encryption method for data privacy information based on National Security Algorithm of China. Firstly, we introduce the Selection and optimization of national security algorithms. Then, we provide the Analysis of anti-tamper mechanism. Lastly, we introduce the power big data privacy information encryption processing.

It is worth noting that the improved AES algorithm has demonstrated outstanding innovative advantages in the field of power big data privacy protection, significantly enhancing encryption strength and processing efficiency through a series of careful optimization measures. This algorithm not only enhances the flexibility of key management, supports dynamic updates and hierarchical management, but also introduces strategies to resist side channel attacks, effectively defending against attacks that analyze keys through physical characteristics. In addition, the adaptive tamper detection mechanism ensures real-time integrity monitoring and rapid response of data. The improved AES algorithm provides good scalability and adapts to future development needs while being compatible with existing power system infrastructure. It constructs a multi-level security protection system, combining mechanisms such as data encryption, access control, and identity authentication, providing comprehensive and solid guarantees for the security of power big data. This series of advantages make the improved AES algorithm a key technical support for the power industry in the face of increasingly severe network security challenges.

3.1 Selection and optimization of national security algorithms

Number equations consecutively with equation numbers in The State Cryptography Administration of China formulates the State Secret Algorithm as a standardized system of cryptographic algorithms aimed at safeguarding national information security. These algorithms encompass various categories, including symmetric cryptographic algorithms, asymmetric cryptographic algorithms, digital signature algorithms, and cryptographic hash algorithms. Notably, SM1 serves as a domestic symmetric cryptographic algorithm utilized for data encryption and decryption, while SM2 functions as a domestic asymmetric cryptographic algorithm enabling key exchange, digital signature, and public key encryption. Additionally, SM3 serves as a domestic cryptographic hash algorithm, facilitating digital digest generation and integrity verification. Complementing these, SM4 operates as a domestic block cipher algorithm designed for data encryption and decryption. These national secret algorithms find widespread application in pivotal sectors such as government agencies, military operations, and financial industries, garnering international recognition and adoption. Engineered to deliver high security, efficiency, and reliability, they not only meet stringent national security standards but also furnish indispensable cryptographic tools for relevant industries and organizations. The SM1 symmetric encryption algorithm, notable for its comparable strength to AES, requires interface calls to encryption chips due to undisclosed implementation principles. With a key length of up to 128 bits and inherent algorithmic robustness, SM1 ensures communication security.

(1) SM2 elliptic curve cryptography algorithm

Elliptic curve public key cryptography operates on elliptic curves over finite fields, leveraging specific curve properties. Firstly, elliptic curves over finite fields form finite commutative groups under the point addition operation, with their order akin to the scale of the base field. Secondly, akin to exponentiation in finite field multiplicative groups, the operation of multiple points on elliptic curves constitutes a monomial function. The problem of solving multiples in this operation, given multiple points and base points, is termed the elliptic curve discrete logarithm problem. As various public-key cryptography algorithms rely on distinct mathematical problems, their security strengths vary. Consequently, to attain equivalent security levels, the key lengths required for different algorithms must differ. This distinction in security strength is typically denoted by computational security indicators L:

$$L_p(v,b) = o\left(e^{b(\ln p)^v (\ln\ln p)^{1-v}}\right) \tag{1}$$

In the formula, e represents the base of the natural logarithm, p represents the size of the algebraic structure in the cryptosystem, b represents the normal number, and v represents a constant in the interval [0,1].

- When $v=0$, the cryptosystem is polynomial time complex;

- When $0<v<1$, the cryptosystem is subexponential in time complexity;
- When $v=1$, the cryptosystem is exponentially time complex.

For the RSA algorithm based on IFP, the fastest algorithm is the number field sieve method. The constant v in the security formula is 1/3, and the parameter p is the product of two large prime numbers used in RSA. For elliptic curve public key cryptography algorithm, Pollard's rho algorithm is generally used to solve it, and the constant v in the security formula is 1, and the parameter p is the order n of the elliptic curve. In other words, for the discrete logarithm problem of the general elliptic curve algorithm, there is only the solution problem of exponential computational complexity [15].

SM2 elliptic curve algorithm has better repudiation resistance compared with the current commonly used elliptic curve algorithm. At the same time, because there is a random number K in the process of key negotiation in SM2 algorithm, the attacker needs to solve the discrete logarithm problem of two elliptic curves continuously without knowing K and the private key d_A, and the time complexity required is self-evident. In addition, because the SM2 elliptic curve algorithm standard requires the use of SM3 hash algorithm, the security of the latter has been proved to be higher than MD5, SHA-1, etc., so the SM2 elliptic curve algorithm is better than other commonly used elliptic curve algorithms in terms of security.

(2) Hash cipher algorithm

Hash cryptography algorithm is also called hash cryptography algorithm or hash cryptography algorithm, which can take any length string as the input algorithm, and output a fixed length string. When the structure of the algorithm is sufficiently sophisticated and complex, it can basically be achieved that there is a unique and determined output corresponding to any input, and conversely, the output can also uniquely identify the input.

Generally, the security performance of the hash algorithm can be evaluated from two perspectives, as follows:

- Calculation irreversibility: given the hash value of the input message , it is computationally impossible to get the original message ;
- Collision resistance: given a message , finding the message makes computationally difficult.

At present, in the field of network security, the more commonly used hash algorithms mainly include MD5, SHA-1, etc., all of which adopt Merkle-Damgard structure or derivative structure, and have good computational irreversibility and collision resistance. The SM3 algorithm also uses the Merkle-Damgard structure, but has a more complex compression function than SHA-256.

(3) Block cipher algorithm

In packet encryption , the plaintext message is divided into groups of fixed length, and then the plaintext group is converted into a ciphertext group of equal length under the action of key , that is, . If a cipher block has a block length of bits, it is called an -bit cipher algorithm. At present, the more widely used password block is generally 64 bits or 128 bits, and the key length is generally 128 bits, 192 bits or 256 bits. Now the cipher block is generally an iterative structure, and its iterative function is called the round function. The master key generates a series of subkeys through the key generation algorithm, and adds these subkeys to the round function [16].

In general, the analysis of a cryptosystem is to distinguish the cryptosystem from the random function. Generally, the following three indicators are used to measure the efficiency of an attack.

- Time complexity: refers to the time required to attack a system, usually measured by the number of encryption and decryption or memory access;
- Space complexity: refers to the storage space required to attack a cryptographic algorithm, usually measured in bytes or words;
- Data complexity: refers to the plaintext and its corresponding ciphertext required to attack a password algorithm.

The SM4 block cipher algorithm standard stipulates that the decryption algorithm has the same structure as the encryption algorithm, only the round key is used in the opposite order, the decryption round key is the reverse

order of the encryption round key, the algorithm block length is 128 bits, and the key length is 128 bits. If the exhaustive attack method needs 2128 operations, which is much larger than the 256 operations of the DES algorithm, it takes a long time to crack the exhaustive attack method, which is not realistic at present. In addition, the encryption algorithm and key extension algorithm of SM4 algorithm adopt 32-round nonlinear iterative structure. At the same time, the S-box specification of SM4 block cipher algorithm is 8-bit input and 8-bit output. Compared with the 6-bit input and 6-bit output of S-box, it has better nonlinear properties and can effectively resist differential attacks [17].

After analyzing the state secret algorithm and anti-tamper mechanism, the SM4 block cipher algorithm in the state secret algorithm is used to encrypt the power big data privacy information. The SM4 block cipher algorithm has been confirmed as a strong encryption algorithm after rigorous cryptographic analysis and review. In the process of encrypting privacy information of power big data, security is crucial. SM4 uses 128-bit key length, which provides a large key space, making brute force cracking of encrypted data extremely difficult. In addition, SM4 also applies techniques such as multi-round iteration and confusion replacement, which further increases the difficulty of attackers to crack the ciphertext. The amount of data in the power industry is often very large, so performance becomes an important factor when encrypting big data privacy information. As a lightweight algorithm, SM4 has high efficiency in encryption and decryption. It can be implemented through hardware and software, and provides the ability to encrypt and decrypt quickly while maintaining a high level of security. This enables SM4 to guarantee efficient and stable performance when processing power big data privacy information.

(4) Optimization of AES algorithm

In the SM4 algorithm, the AES algorithm provides strong security, by using 128-bit, 192-bit or 256-bit key length, can provide a larger key space and higher security, and effectively resist cryptanalysis attacks. Therefore, AES algorithm is used for tamper-proof encryption of power big data privacy information.

Although the sequence constructed by one-dimensional Logistic chaotic mapping has good randomness and divergence, it is still slightly insufficient for the demand of encryption, so the AES algorithm is optimized by using two-dimensional Logistic mapping. The key expansion is optimized, and its dynamic equation can be expressed as:

$$\begin{cases} x_{n+1} = \mu\lambda_1 * x_n * (1 - x_n) + \gamma * y_n \\ y_{n+1} = \mu\lambda_2 * y_n * (1 - y_n) + \gamma * x_n \end{cases} \tag{2}$$

After many tests and adjustments, the parameters of chaotic state by two-dimensional Logistic mapping are determined, where $\mu = 0.4$, $\gamma = 0.1$, $\lambda_1 = \lambda_2 = 0.83$.

Due to the high security of the execution process and most encryption steps of the AES algorithm, optimization and improvement were carried out in the generation of its seed key and extension key in this study. To improve the problem of fixed seed key and extended key space in the chaotic mapping system. The specific optimization and improvement process is shown below.

Firstly, two initial passwords $PW1$ and $PW2$ need to be inputted, with lengths of L_1 and L_2. As the length of the password may not match the three key modes used by AES in terms of length, it is necessary to calculate the difference between L_1 and L_2 and the AES key length separately to obtain $DV1$ and $DV2$. Then, take their cubic values and assign them to A_1 and A_2. Calculate formula (3) to obtain N_1 and N_2 respectively:

$$N_i = (A_i = \mathrm{mod}100000) / 100000 \tag{3}$$

Assign N_1 and N_2 to x_0 and y_0 as the initial values in formula (2), and execute n cycles through this formula. Multiply the calculated x_n and y_n by 100000 each, and then combine the results in reverse order to form a string of 12 characters. Take $n/8$ values in order to suppleWment the input $PW1$, so that it meets the AES algorithm's requirements for key length, which is the first seed key $key1$; Similarly, $key2$ can also be calculated using this method.

In order to increase the complexity of the sequence, it is necessary to reorganize the extended key by substituting L_1 and L_2 as initial values x_0 and y_0 into the iterative equation, and setting the remaining parameters as:

$$\begin{cases} x_{n+1} = 0.4\lambda_1 * x_n * (1 - x_n) + 0.1 * y_n \\ y_{n+1} = 0.4\lambda_2 * y_n * (1 - y_n) + 0.1 * x_n \end{cases} \quad (4)$$

Select five and six unique non-repeating numbers from the results of two rounds of iteration, each ranging from 1 to 11. Utilize these numbers as serial numbers to identify the corresponding keys from the original extended key set. Subsequently, employ interpolation to reconstruct a new set of eleven rounds of keys. The process is illustrated in Figure 1.

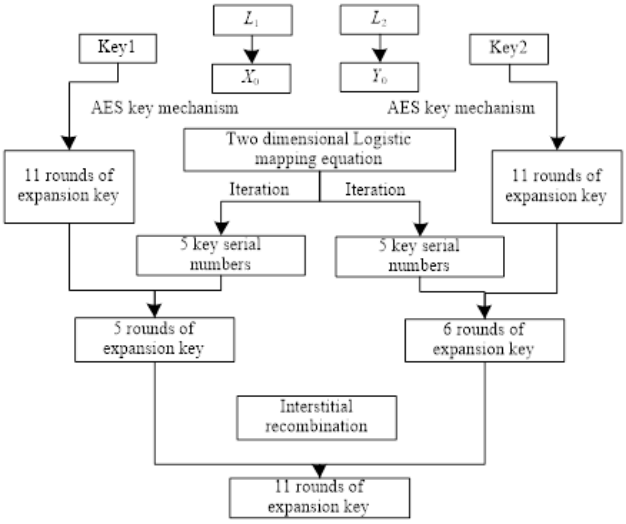


Fig.1 Sequence Reorganization Algorithm

Key extension is the core content of this improved scheme. According to the AES algorithm mechanism, this link should be composed of 11 rounds of iterative calculation. This section takes the first round as an example to introduce the design idea of this algorithm.

Step 1: Take out the generated in the construction of the seed key and the first round of the 11 rounds of the recombination key set calculated in the recombination sequence, and convert both of them into 128-bit binary representation, denoted as and respectively.

Step 2: Add 0 to the end of and to form a 129-bit binary string, which is converted into a decimal number in a group of three digits, and a total of 43 integers with the value range of [0,7] are obtained in two groups.

Step 3: Construct two one-dimensional integer arrays and with 48 elements. Incorporate the two groups of numbers obtained in step (2) into the corresponding array and fill the insufficient elements with zeros.

Step 4: The 6 elements in each array are divided into groups, a total of 8×2 groups, each group is converted to the corresponding decimal in order, and stored in and respectively [18-20].

Step 5: Take out the elements in the above two arrays according to their subscript sizes and form 8 groups, denoted as , and substitute them into the two-dimensional Logistic mapping equation described in formula (4). After iterative operation of U turns, 8 groups of corresponding outputs are obtained, denoted as and .

Step 6: Take and in order of six decimal places, and ensure that the mean of each digit is within [0,7].

Step 7: Convert each digit in step 6 into a 3-bit binary number to get two sets of 18-bit binary numbers.

Step 8: Repeat steps 5 to 7 to convert all decimal numbers in and into binary strings, totaling 8 groups.

Step 9: Perform the XOR operation on each set of two strings to obtain a binary with a length of 144 bits. The first 128 bits are the first round of keys in the new extended key set.

Step 10: Take out the second round key of the reorganized key set in step 1, and use as the initial key, and repeat the algorithm to get the second round key of the extended key set.

The subsequent iteration process is all executed according to the above algorithm, in which is used for odd rounds, is used for even rounds, and all 11 rounds of extended keys can be reconstructed. Thus, the optimization and improvement of the key extension mechanism in the AES algorithm has been completed, and the difficulty factor of decoding the AES algorithm has been reliably improved, and the security of the privacy information of the power big data has been increased.

3.2 Analysis of anti-tamper mechanism

The aim of implementing an anti-tampering mechanism for safeguarding the privacy of power big data is to prevent unauthorized access, tampering, or misuse of personal information contained within. Such mechanisms are designed to uphold data confidentiality, ensure integrity verification, thwart unauthorized access and data disclosure, and provide robust security, compatibility, and versatility. Common anti-tampering measures include:

(1) Data encryption: Utilizing robust cryptographic algorithms and secure key management mechanisms to encrypt power big data, ensuring that only authorized individuals can decrypt and access it, thereby safeguarding data confidentiality during transmission and storage.

(2) Access control: Implementing stringent rights management systems to restrict access to specific data solely to authorized personnel. Techniques such as identity authentication, access authorization, and role management are employed to ensure legitimate and necessary data access.

(3) Secure transmission: Employing secure communication protocols like HTTPS, VPN, etc., to securely transmit power big data, mitigating the risks of interception or tampering during transmission.

(4) Log auditing: Establishing comprehensive logging and auditing mechanisms to monitor and track access and operation activities concerning power big data, promptly identifying anomalies and taking appropriate remedial actions.

(5) Data backup and recovery: Regularly backing up data and ensuring the secure storage of backup copies, facilitating timely data recovery in case of tampering or damage to the original data.

(6) Anonymization: Employing anonymization techniques to obfuscate sensitive personal information and privacy data, thereby reducing the risk of data leakage by separating information tied to specific personal identities.

(7) Security review: Conducting routine security assessments and vulnerability scans to identify and patch security vulnerabilities in power big data systems, mitigating the risks of hacker attacks and unauthorized access. Failure to adequately protect data security in power application platforms may expose the data to interception, tampering, or forgery during transmission. The topology diagram illustrating the interaction of power big data is depicted in Figure 2.

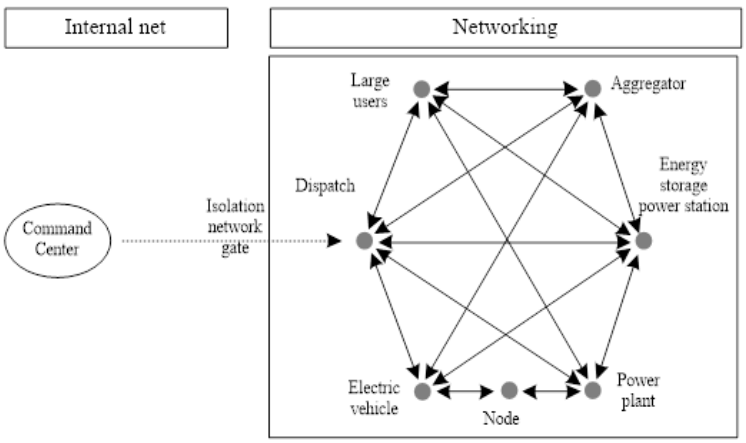


Fig.2 Topology structure of power big data interaction

To safeguard the privacy of power big data and mitigate tampering risks, encryption technology assumes a pivotal role, offering numerous advantages. Firstly, encryption ensures data confidentiality by restricting access solely to authorized individuals, thereby preventing unauthorized parties from obtaining sensitive information. Secondly, encryption facilitates data integrity verification, enabling detection of tampering as decrypted data will not match the original data if tampered with. Moreover, encryption effectively prevents unauthorized access and data leakage, rendering stolen data indecipherable to unauthorized parties. Encryption algorithms typically boast robust security features capable of withstanding large-scale attacks, and encryption technology finds widespread application across diverse fields, owing to its compatibility, ease of management, and implementation simplicity.

3.3 Power big data privacy information encryption processing

(1) Privacy information clustering of power big data

With the continuous expansion of the scale of power grid, the amount of data increases rapidly. Clustering and mining of private information from power big data can quickly encrypt it, improving the quality and efficiency of encryption. K-means clustering algorithm is one of the most common clustering partitioning algorithms, which has been effectively applied in power big data processing. K-means algorithm is a typical distance-based clustering algorithm, which determines the similarity of objects by calculating the distance between objects, and the similarity between objects is inversely proportional to the distance. A cluster is a set of highly similar data that is combined together. K-means clustering algorithm divides N data objects into k clusters to achieve the effect that the similarity between data objects in the same cluster is relatively high, while the similarity between data objects in clusters is very low.

The experimental flow of K-means clustering algorithm is as follows:

(1) From the given N data objects, k data objects are selected as the initial center of mass by random algorithm.

(2) Calculate the distance between the remaining N data objects and k centroids, and assign the data objects to the nearest centroid cluster.

(3) Calculate the mean in each cluster as the temporary center of mass of this cluster.

(4) Repeat the two steps (2) and (3) until the new center of mass is the same as the original center of mass or within a certain range, exit the algorithm.

(5) Generally, the distance between two data objects can be calculated using the average error criterion, as shown in formula (5) :

$$V = \sum_{i=1}^{m} \sum_{x_j} \left| x_j - M_i \right|^2 \tag{5}$$

In the formula, V represents the sum of squared errors of privacy information in all power big data, x_j represents the J th data object in the data object, and M_i represents the i th center of mass.

During the execution of the K-means algorithm, when mapping power big data objects to designated centroids, each data object operates independently without engaging in information exchange. To assign a data object to its corresponding centroid, the object computes its distance to each centroid individually, subsequently identifying the centroid cluster to which it belongs by determining the minimum distance via a sorting algorithm. Importantly, this process occurs independently for each data object, with no direct interaction between the computation of centroids for preceding data objects and subsequent ones.

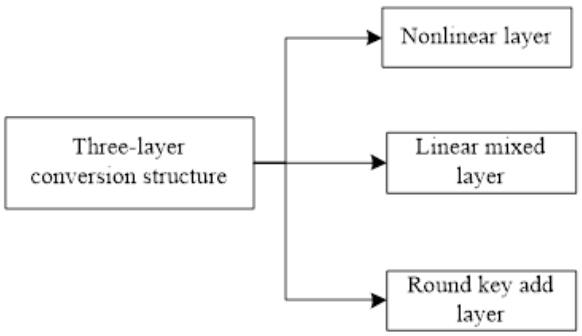


Fig.3 Three-layer structure diagram of the conversion network

(2) Tamper-proof encryption based on improved AES algorithm

The AES algorithm allows for specifying both the packet length and key length as 128 bits, with options for key lengths of 192 bits and 256 bits, offering enhanced security compared to the DES algorithm with its fixed key length of 56 bits. The enhanced AES algorithm incorporates a conversion network, with each round undergoing three layers of transformation, as depicted in Figure 3.
In Figure 3, each layer is:
(1) Nonlinear layer: that is, the process of byte replacement, which is formed by the juxtaposition of 16 S-boxes, and its role is mainly to confuse bytes inside.

(2) Linear mixing layer: This layer is transformed by row displacement and column confusion, and finally ensures that the plaintext has the characteristics of overall confusion and high diffusion after multiple rounds of transformation.

(3) Round key adding layer: The sub-key matrix and the plaintext intermediate state matrix are performed by XOR operation.

The length of the key can be 128 bits, 192 bits or 256 bits. The operation process of these three lengths in the algorithm is similar, but the number of cycles is different, and the corresponding 128-bit cycle is 10 rounds, 192-bit cycle is 12 rounds, and 256-bit cycle is 14 rounds. In this study, we consider the case where both the packet length and the key length are 128 bits.

The encryption process of the improved AES algorithm is shown in Figure 4.

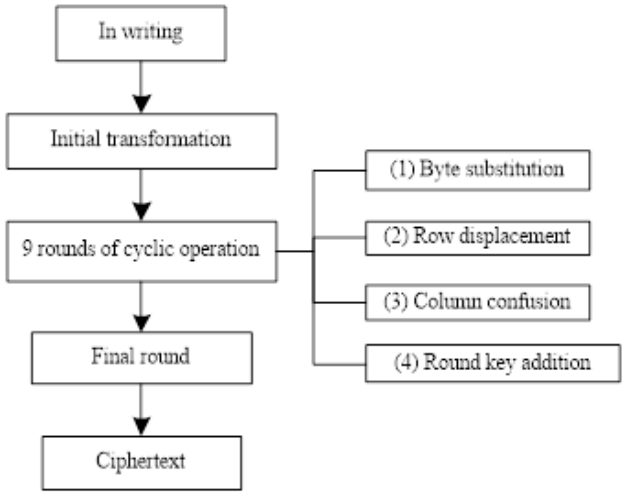
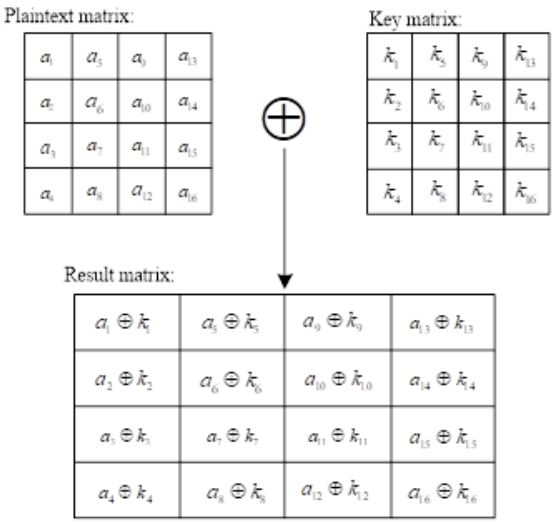


Fig. 4 Encryption flow of improved AES algorithm

In the improved AES algorithm encryption process, both the plaintext data and the key are organized as 4×4 matrices, with bytes arranged from top to bottom and left to right.

- Initial transformation: In the process of initial transformation, the plaintext block and the initial subkey block matrix are confused by the bit operation, and the processing result of the initial transformation operation is obtained. The process of initial transformation is shown in Figure 5.
- Byte substitution: Serving as the first step in the repeat round operation, byte substitution maps input data from one byte to another using an S-BOX. In the initial round, it replaces the 4×4 matrix from the initial transformation with new values obtained by table lookup, resulting in a new state matrix. This process is nonlinear, and its inverse is utilized in decryption to map the state matrix back to byte data. Figure 6 shows the byte substitution process.
- Row displacement: This operation enhances the diffusion of the improved AES algorithm by rotating each row of the result matrix leftward after byte substitution. The rotation rule cycles as follows: the first row remains unchanged, the second row shifts one byte left, the third row shifts two bytes left, and the fourth row shifts three bytes left. The row displacement diagram is shown in Figure 7.
- Column obfuscation: This step involves multiplying the state matrix left by a given 4×4 matrix.

- Round key addition: Implemented via XOR operation between the input matrix and the key matrix, round key addition is a crucial operation requiring keys during both encryption and decryption processes. The 128-



bit key undergoes 11 rounds of key changes throughout the encryption process. The diagram of wheel key addition is shown in Figure 8.

Fig. 5 Initial transformation process

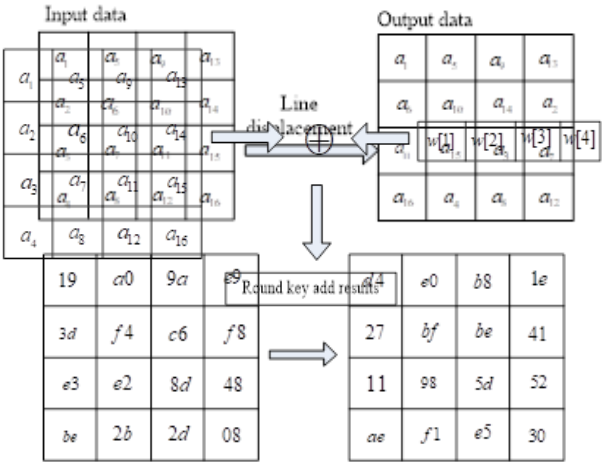
Fig. 6 Byte substitution diagram

Fig. 7 Displacement diagram of row

Fig. 8 Schematic diagram of round key addition

4 EXPERIMENT RESULTS

In this section, we conduct extensive experiments to evaluate the efficacy of the tamper-proof encryption method proposed for safeguarding the privacy of power big data.



4.1 Sample data

This experimental test necessitates a substantial dataset of power big data privacy information, thus emphasizing the collection of experimental sample data. Power big data privacy information encompasses various categories:

- Personal identifiable information: including names, contact details (phone numbers, email addresses), and residential addresses.
- Account information: comprising power supplier account details and electricity bills.
- Electricity information: encompassing statistics on electricity consumption, trend analysis, and information related to electricity equipment.

- Equipment details: covering smart meter or related equipment numbers, models, and installation locations.
- Geographical location data: such as real-time power consumption areas and power addresses of users.
- Usage patterns: including time periods of electricity consumption, peak and off-peak electricity consumption distribution, and frequency and duration of electrical appliance usage.

The experimental data is sourced from permissible privacy information within the power enterprise's big data of China, with a dataset size of 10GB.

4.2 Experimental scheme

In order to improve encryption efficiency, a national encryption algorithm suitable for the characteristics of power big data should be selected, and parallel processing and algorithm optimization techniques should be used to accelerate the encryption process. In terms of managing spatial complexity, data compression and effective key management strategies are used to reduce storage requirements, while considering adopting lightweight encryption schemes to adapt to limited storage resources. To address the complexity of data, simplify data structures, select appropriate encryption modes, and implement dynamic encryption strategy adjustments to reduce algorithm complexity while maintaining its adaptability and efficiency. By integrating these strategies, an efficient encryption scheme can be designed that ensures both data security and system resource utilization, while ensuring regular evaluation and optimization to address evolving threats and business needs.

Taking the deciphering probability, robustness and encryption time of power big data privacy information as indicators, the proposed method is compared with the reference [9] method and the reference [10] method.

The decoding probability of power big data privacy information refers to the probability of unauthorized users successfully obtaining privacy information when attempting to decrypt encrypted power data. This probability reflects the security of the encryption method. Usually, a secure encryption method should make the decoding probability close to zero, making it almost impossible for unauthorized users to decrypt:

$$\eta = \frac{\beta_1}{\beta_2} \tag{6}$$

In the formula, β_1 represents the decoded power privacy information, and β_2 represents the total amount of power privacy information.

Robustness: The tamper-proof encryption performance of power big data privacy information is inseparably related to the encryption algorithm, and the stability of the encryption algorithm can continue to ensure the security of power big data privacy information. Its calculation formula is as follows:

$$Q = \frac{\sum_{i}^{t} D_i}{t} \tag{7}$$

In the formula, D_i represents the number of tampering attacks and t represents the number of tests.

Encryption time: The encryption time refers to the time required to encrypt the privacy information of power big data. The shorter the encryption time, the higher the efficiency of the encryption method.

4.3 Analysis of experimental results

(1) Probability of deciphering privacy information

The assessment of the encryption method's security against adversarial attacks involves verifying the probability of decoding private information. This evaluation helps determine whether the method sufficiently safeguards power big data privacy information from unauthorized access and tampering. The probability outcomes regarding privacy information decoding for the three methods are depicted in Figure 9.

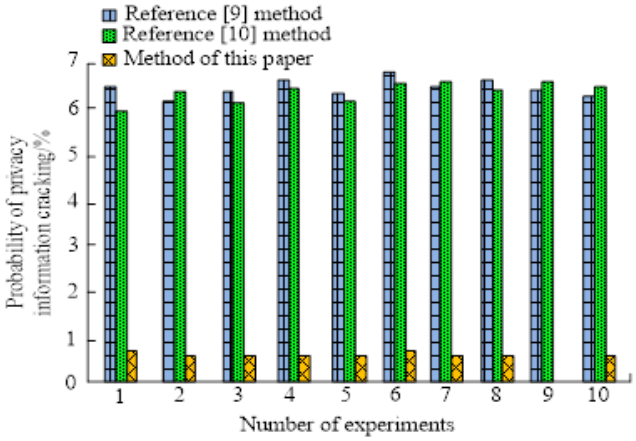


Fig.9 Decoding probability of privacy information

This method successfully reduces the probability of decoding big data privacy information to an unprecedented low level through exquisite algorithm design, and its probability value is strictly controlled below 1%. This achievement not only reflects the efficiency of the method, but also its enormous potential in practical applications. In contrast, traditional methods such as encryption techniques described in references [9] and [10], after being encrypted, still maintain a high decryption probability of 6%. This number clearly reveals the vulnerability of traditional methods in the face of modern complex network attacks, as well as their limitations in protecting sensitive data. The research results not only emphasize the robustness of the proposed method in protecting privacy information of power big data, but also highlight its superior tamper resistance performance. Against the backdrop of increasingly severe network security threats, this approach undoubtedly provides a solid defense line for the power industry, ensuring the security and integrity of critical data. By reducing the decoding probability, our method has set a new benchmark for privacy protection of power big data, providing valuable reference and inspiration for future research and practice.

(2) Encryption robustness results

By conducting tests and assessments on the resilience of encryption methods, their capacity to withstand various attack techniques, such as password cracking, analysis attacks, and man-in-the-middle attacks, can be gauged. It's crucial to verify the encryption method's performance stability, ensuring it maintains efficiency and speed while handling substantial data volumes without compromising security. The robustness results of the three methods are presented in Table 1.

Table 1 Robustness results of the three methods

Number of experiments	Encryption robustness		
	Method of this paper	Reference [9] method	Reference [10] method
2	98.25	81.35	78.25
4	98.36	81.02	77.24
6	98.35	79.09	76.01
8	98.99	78.21	75.03
10	98.56	78.15	75.01

Upon analyzing the data in the table, it's evident that the proposed method consistently exhibits the highest encryption robustness across all tests. The relatively high scores indicate its effectiveness in thwarting diverse attacks. In comparison to the methods referenced in [9] and [10], the proposed method demonstrates significantly greater encryption robustness. This underscores the method's reliability and security in safeguarding the privacy information of power big data from tampering.

(3) Encryption time result

Analyzing encryption time enables the assessment of an encryption method's practical efficiency. Given that power big data typically encompasses substantial private information, the speed of encryption directly influences the method's overall responsiveness. Verification of encryption time helps ascertain whether the method is efficient enough to complete data encryption within a reasonable timeframe. The encryption time results of the three methods are depicted in Figure 10.

Figure 10 illustrates that, with equivalent data volumes, the proposed method markedly reduces encryption time compared to those referenced in [9] and [10]. Specifically, when processing 10GB of data, the proposed method encrypts in less than 300 seconds, whereas the other methods exceed 600 seconds. This underscores the high encryption efficiency of the method proposed in this paper.

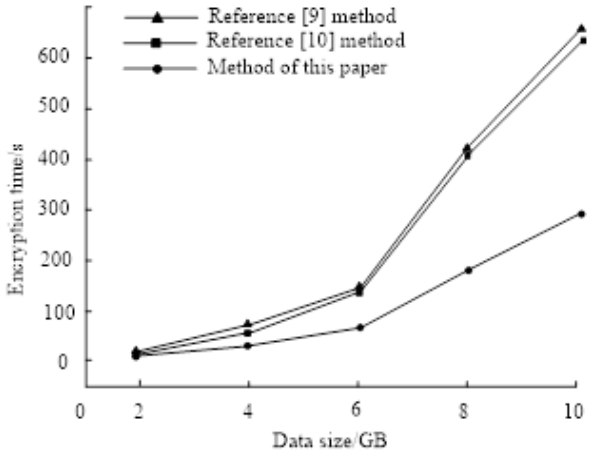


Fig. 10 Encryption time

5 CONCLUSION

With the rapid expansion of power big data, safeguarding its privacy and preventing tampering has emerged as a critical imperative. The development of a tamper-proof encryption method for power big data privacy information, based on state secret algorithms, holds paramount importance in ensuring data security. This study aims to enhance the confidentiality, integrity, and reliability of power big data by employing national secret algorithms for encryption. Through a comprehensive examination of state secret algorithms, this paper proposes an encryption scheme tailored to protect the privacy information within power big data. Leveraging the inherent characteristics and advantages of state secret algorithms, the proposed scheme effectively guards against unauthorized access and tampering attempts. Experimental findings corroborate the outstanding performance of the encryption method based on state secret algorithms, showcasing remarkable robustness, security, and computational efficiency.

ACKNOWLEDGEMENT

This research was supported by the Yunnan Power Grid Technology Project under Grant 059300202303030105000001.

REFRENCES

[1] Zhang Jia, Lu Xinghua. Simulation of Big Data Attribute Authorization Encryption Algorithm Based on Link Weights[J]. Computer Simulation,2023,40(05):295-298.
[2] Ibrahem M I , Nabil M , Fouda M M , et al. Efficient Privacy-Preserving Electricity Theft Detection With Dynamic Billing and Load Monitoring for AMI Networks[J].Institute of Electrical and Electronics Engineers (IEEE), 2021,8(2):1243-1258.
[3] Gaybullaev T , Kwon H Y , Kim T ,et al. Efficient and Privacy-Preserving Energy Trading on Blockchain Using Dual Binary Encoding for Inner Product Encryption[J].Sensors, 2021, 21(6):2024-2031.
[4] Ranjith R , Shalini E . Location Based Services with Data Examination and Restoration[J].IOSR journal of computer engineering, 2021,23(2):47-51.
[5] Liu S , Liu Y , Liu W , et al. A certificateless multi-dimensional data aggregation scheme for smart grid[J].J. Syst. Archit. 2023, 140:1028-1037.

[6] Varkuti K S , Manideep G . A Novel Architectural Design of Light Weight Modified Advanced Encryption Standard for low power and high speed applications[J]. High Technology Letters, 2021, 27(2):360-370.
[7] Ding Jinduo, Wang Xiaoxiang, Jiang Yurong, et al. Substation integrated information data encryption method based on improved Blowfish algorithm[J]. Microelectronics & Computer,2023,40(8):87-93.
[8] Li Hongwei, Pan Zhiyuan, Huang Jijie. Research on a Data Encryption Algorithm Based on Double Hook Function[J]. Computer Technology and Development,2022,32(06):120-125.
[9] Zhao Ying, Fan Xingyuan. Design of Anti-leakage Encryption System for Power Metering Data Based on Homomorphic Encryption[J]. Computing Technology and Automation,2023,42(03):118-123.
[10] Wang Linxin, Luo Shigang, Li Shulin, et al. De privacy and data encryption of smart grid big data based on deep learning[J]. Electronic Design Engineering,2021,29(03):175-178+183.
[11] Khan H M , Khan A , Jabeen F ,et al. Fog-enabled secure multiparty computation based aggregation scheme in smart grid[J]. Computers & Electrical Engineering, 2021, 94(12):1073-1085.
[12] Zekaj B , Jusufi A , Imeri-Jusufi B . Using incomplete polynomial functions of the odd degree n and their inverses for data encryption and decryption[J]. IFAC-PapersOnLine, 2022, 48(15):641-648.
[13] Manjula P , Priya S B . Intelligent Chimp Metaheuristics Optimization with Data Encryption Protocol for WSN[J]. Computers, Materials and Continua (Tech Science Press), 2022,32(1):326-332.
[14] Shi J , Yu Q , Yu Y , et al. Privacy protection in social applications: A ciphertext policy attribute-based encryption with keyword search[J]. International journal of intelligent systems, 2022, 37(12):12152-12168.
[15] Arulananth T S , Baskar M , Anbarasu V , et al. Multi party secure data access management in cloud using user centric block chain data encryption[J]. Pattern recognition letters, 2021, 152(32):295-301.
[16] Viswanath G , Krishna P V . Hybrid encryption framework for securing big data storage in multi-cloud environment[J]. Evolutionary Intelligence, 2021, 14(2):691-698.
[17] Zhe J , Kai Z , Liangliang W , et al. Forward Secure Public-key Authenticated Encryption with Conjunctive Keyword Search[J]. The Computer Journal, 2022, 66(9):2265-2278.
[18] Hu S , Wang X , He H , et al. Complex and flexible data access policy in attribute-based encryption[J]. Journal of supercomputing, 2022, 78(1):1010-1029.
[19] Mittal S , Ramkumar K R .Research perspectives on fully homomorphic encryption models for cloud sector[J]. Journal of Computer Security, 2021, 29(3):359-359.
[20] Zhang Q , Zhao Z . Distributed storage scheme for encryption speech data based on blockchain and IPFS[J]. Journal of supercomputing, 2023, 79(1):897-923.