symmetry

MDPI

Article

A Multi-Source Big Data Security System of Power Monitoring Network Based on Adaptive Combined Public Key Algorithm

Chengzhi Jiang [1,*], Chuanfeng Huang [1], Qiwei Huang [1] and Jian Shi [2]

[1] Nanjing Institute of Technology, Nanjing 211167, China; hcfnjit@njit.edu.cn (C.H.); j00000003238@njit.edu.cn (Q.H.)
[2] Nanjing Zhiqing Information Technology Co., Ltd., Nanjing 210008, China; jian.shi@dataoriental.com
* Correspondence: jcz@njit.edu.cn

Abstract: The multi-source data collected by the power Internet of Things (IoT) provide the data foundation for the power big data analysis. Due to the limited computational capability and large amount of data collection terminals in power IoT, the traditional security mechanism has to be adapted to such an environment. In order to ensure the security of multi-source data in the power monitoring networks, a security system for multi-source big data in power monitoring networks based on the adaptive combined public key algorithm and an identity-based public key authentication protocol is proposed. Based on elliptic curve cryptography and combined public key authentication, the mapping value of user identification information is used to combine the information in a public and private key factor matrix to obtain the corresponding user key pair. The adaptive key fragment and combination method are designed so that the keys are generated while the status of terminals and key generation service is sensed. An identification-based public key authentication protocol is proposed for the power monitoring system where the authentication process is described step by step. Experiments are established to validate the efficiency and effectiveness of the proposed system. The results show that the proposed model demonstrates satisfying performance in key update rate, key generation quantity, data authentication time, and data security. Finally, the proposed model is experimentally implemented in a substation power IoT environment where the application architecture and security mechanism are described. The security evaluation of the experimental implementation shows that the proposed model can resist a series of attacks such as counterfeiting terminal, data eavesdropping, and tampering.

Keywords: combined public key algorithm; elliptic curve cryptography; power monitoring network; multi-source big data; data encryption

Citation: Jiang, C.; Huang, C.; Huang, Q.; Shi, J. A Multi-Source Big Data Security System of Power Monitoring Network Based on Adaptive Combined Public Key Algorithm. *Symmetry* **2021**, *13*, 1718. https://doi.org/10.3390/sym13091718

Academic Editors: Xin Luo and Di Wu

Received: 12 August 2021
Accepted: 15 September 2021
Published: 16 September 2021

1. Introduction

In recent years, with high-level development at the economic level promoting the transformation of the traditional power grid, the smart grid has been produced, which integrates the electricity network, telecommunication infrastructure, and information technology [1]. Smart grid is an advanced digital infrastructure with two-way information communication, equipment control, and power distribution capabilities. The power monitoring network in the smart grid needs the development of new smart sensing systems with improved characteristics, including embedded data processing, remotely controllable, etc. [2]. Hence, wireless sensor network (WSN) and IoT (Internet of Things) technologies are gradually applied for monitoring power transmission lines, substation electrical equipment, and distribution site operations in smart grid [3]. An IoT-assisted power monitoring system comprises various sensors, communication components, intermediate data storage, and a remote monitoring center [4]. The massive quantity of data collected by power IoT systems is not only of large volume but is also more diversified compared with the traditional grid data, and its sources and distribution are more extensive [5]. Advances in

the IoT technologies of the smart grid have led that smart grid data demonstrating typical features of big data such as the abundant amount of data generated, the diversity of data sources and the high velocity of data generation [6,7]. To handle the smart grid with data characterized by big data features, ref. [8] proposed an architecture using random matrix theory to achieve anomaly detection. The security-related data generated by electrical devices, network devices, control centers, and substation field buses in smart grid can be collected to support the security awareness based on big data analysis [9]. The massive information of the power system obtained by the power monitoring network is the data basis and premise for reasonable control and the optimal dispatching of the power grid [10]. When the data have errors due to power monitoring network failures, channel transmission packet loss, and delays, etc., the analyzed control and scheduling instructions may mislead and affect the dispatchers to make wrong decisions [11,12]. Therefore, the security of these data plays an important role in the operation stability of the smart grid.

Due to the widespread applications of IoT technologies in the smart grid and power big data, these increased interconnections not only enhance the status awareness ability of power systems but also provide potential attack vectors in power systems, leading to new security challenges [13]. The IoT security risks can be from physical attacks, network attacks, software attacks, and encryption attacks such as malicious node injection, unauthorized access, and node tampering [14]. The security of IoT and big data in the smart grid has been attracting many researchers in recent years. Reference [15] proposed a Q-learning-based system that can be used to preserve the privacy of electricity data in the IoT-enabled smart grid. In that system, the massive data collected from electricity regions was sent to the control center, and a secret-sharing protocol was applied between the control center and edge servers to output the privacy-preserving final decisions. Reference [5] reviewed the energy big data security studies from the perspective of data, summarizing the energy big data security from the aspects of big data-oriented cryptosystems, privacy-preserving applications, and anomaly detection. Reference [16] proposed a security solution based on identity-based encryption and signature for a big data management framework in smart grid. Reference [17] reviewed the cyber threats and countermeasures in IoT and big data-enabled smart grid where countermeasure strategies are categorized into protection and detection. Among the protection methods, cryptography was considered as an effective way for protecting data from security breaches where different encryption and authentication algorithms were listed. Reference [18] proposed a recognition algorithm to eliminate the impact of abnormal electrical load data on the accuracy of load forecasting. It can be seen that data security is continuously emphasized in the security of the smart grid which should guarantee the confidentiality, integrity, availability and the authenticity of data.

Data security measures may need to consider the confidentiality and integrity of terminal data and the credibility of terminals in smart grid networks [19]. Efforts were made to improve the data transmission efficiency and data security in the smart grid. In reference [20], an efficient data transmission algorithm for the power grid was proposed. By sending data between different nodes, an energy model was constructed. By counting the energy of the nodes sending data, the grid data can be efficiently transmitted. The algorithm extends the survival time of nodes and has higher performance. In reference [21], a layered security architecture was proposed for mobile terminals in power information systems. To protect the data transmission of terminals, the authors proposed a hybrid encryption method combined with public key infrastructure (PKI) technology. To balance efficiency and security, reference [22] proposed a smart meter scheme to prevent pollution attacks from misbehaving collectors. Through this scheme, secure data aggregation and dynamic billing were simultaneously achieved. It can be seen that the above studies mainly focused on prolonging the survival time of the data node to improve the transmission efficiency [20] or on the confidentiality of the terminals' data transmission [21], or proposed a method to solve a specific security problem for smart meters [22]. Considering the realistic implementations in some scenarios of the State Grid Corporation of China (SGCC)

that lack the security protection of massive terminals via wireless network or low-key generation and distribution rate, we are motivated to try to find a more general, practical and easy-to-implement method that could solve the data security problem of massive connecting terminals/sensors, resisting the primary threats outlined in reference [14] and considering the security in reference [19] mentioned above.

This paper proposes a system of multi-source big data security based on the combined public key algorithm in the power monitoring network. The system consists of a server-side security device that is installed with a security service based on adaptive combined public key (CPK) algorithms and security agents that are installed on terminals/nodes in the power monitoring networks of the smart grid. By implementing the proposed system, we can achieve the end-to-end security and efficient data transmission. Our contributions can be summarized as follows:

1. Inspired by the realistic implementations of SGCC (lack of efficient security countermeasures for massive IoT terminals/nodes) and the security threats faced by IoT and big data in smart grid, we propose a more efficient, effective, and easy-to-implement security system to provide end-to-end security, so that the risks such as malicious node injection, unauthorized access, and node tampering [14] are significantly lowered;
2. To increase the efficiency of the security system, we upgraded the CPK algorithm with the adaptive key fragment and combination method, so that the key generation and updating process is adapted according to the number of connecting terminals;
3. To eliminate the negative impact of malicious terminals, we propose an identity-based public key authentication protocol. The lightweight protocol can achieve the efficient secure access of massive terminals;
4. To verify the easy-to-implement property of the proposed system, the system was experimentally implemented in a substation scenario where the full functions were tested. It shows that the system can be easily implemented with minor changes to the existing network.

The remainder of the paper is organized as follows. The preliminary CPK algorithm is described in Section 2. Section 3 describes our proposed system, including the adaptive method and authentication protocol. Section 4 establishes experiments to demonstrate the performance and effectiveness of the proposed system. In Section 5, the proposed system is experimentally implemented in a substation scenario to test its full functions. Section 6 gives the conclusions of our paper.

2. Combined Public Key Algorithm

CPK is a public key system based on the combination, which combines key production and key management, and can meet the requirements of identification authentication [23]. A CPK cryptosystem is based on the user identification information. It uses information mapping technology to map user identification and then uses the mapping value to combine the information in the public–private key factor matrix until the corresponding key information is obtained. CPK is based on elliptic curve cryptography (ECC). Its generation does not rely on the participation of a trusted third party (e.g., certificate authority).

2.1. Elliptic Curve Cryptography

ECC uses a short key length to achieve a strong security strength, and CPK authentication is implemented based on the ECC algorithm. Since ECC with a short key length can achieve the security strength of other known key algorithms with a long key length (e.g., RSA), ECC technology has broad application prospects such as identity authentication and digital signature [24].

An elliptic curve $E(K)$ is a curve that satisfies the Weierstrass equation, including all solutions of the function plus the so-called infinity point as shown in Equation (1):

$$E(K): Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6 \quad (1)$$

Figure 1 shows the addition of two different points $P(x_1, y_1)$ and $Q(x_2, y_2)$ on the elliptic curve. Point P and Q are connected and prolonged to intersect with the curve $E(K)$ at R'. Then, a line is made through R' parallel to the Y axis and intersects with curve $E(K)$ at $R(x_3, y_3)$. R is the sum of the two points P and Q, denoted as $R = P + Q$.

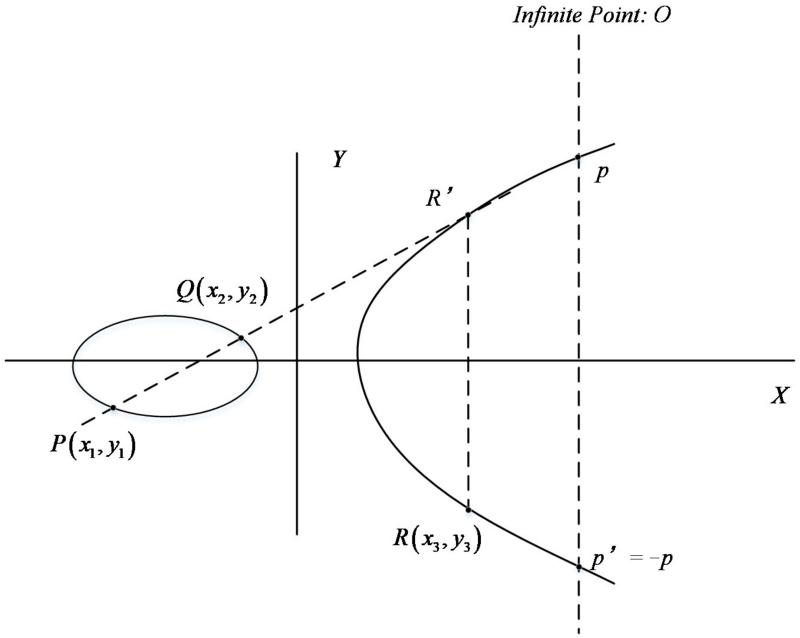


Figure 1. Double point operation on elliptic curve.

When P and Q coincide, a tangent line through point P can be made which intersects with curve $E(K)$ at one point. Then, a line parallel to the Y axis can be made through that point, which intersects with curve $E(K)$ at a point $R(x_3, y_3)$, denoted as $R = P + P = 2P$, this kind of drawing method is called the elliptic curve double point operation [25]. The addition of multiple coincident points P is denoted as: $Q = nP$. It can also be seen from Figure 1 that the infinite point O intersects with the curve at point p', and makes a line through p' parallel to the Y axis to intersect with the curve at p. Then, we have $O + p = p$, $p' = -p$, where O is called zero element and p' is the inverse element of p.

The above curve is discussed in the real number domain. However, the elliptic curves in the discussed cryptography are basically in the finite field F_p which consists of p elements where p is a prime number. Then, the elliptic curve equation can be expressed by Equation (2) where $a, b \in F_p$:

$$E(K)_{a,b} : Y^2 = X^3 + aX + b(\mathrm{mod}p) \tag{2}$$

The corresponding curve discriminant is shown in Equation (3):

$$\Delta = -16\left(4a^3 + 27b^2\right) \tag{3}$$

The following multiplication operation definition can be obtained from the double point operation in the elliptic curve: in the finite field F_p, G and Q are two points on the elliptic curve $E(K)$, where G is the base point. There is a point $k \in [1, n-1]$, so that $k \cdot G = Q$. The security of the ECC is guaranteed by the following mathematical problem called the discrete logarithm problem. It is an easier process to solve for $k \cdot G = Q$ given a base point G on the elliptic curve and an integer k. However, given a base point G and known Q, it is hard to find k.

When using the digital signatures based on discrete logarithms, the confidentiality of k is very important. The cryptographic algorithm is so far one of the most important

and fundamental algorithms. Through this discrete logarithm problem, processes such as encryption and signature are implemented [26].

2.2. CPK System

Based on ECC, CPK generates a secure elliptic curve, uses this secure elliptic curve to generate a key matrix that is used to generate key pairs required by the secure communications between nodes in the network.

The CPK system adopts the elliptic curve of Equation (2) on the finite field F_p, which is defined by the parameter $T = (a, b, G, n, p)$, where a, b is the coefficients, $a, b, x, y \in p$, G is the base point on the curve, n is the order of G. All the multiple points of the base point G form the subgroup S. The elements in subgroup S are all multiple points $kG, k = (1, 2, 3, \ldots, n)$ of G, and $S = \{G, 2G, 3G, \ldots, nG\} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is obtained.

The key generation center is responsible for constructing the matrix. First, according to the ECC principle, the base point G and the multiple point S are selected to construct the matrix on the finite field F_p. The key matrix generation process is shown in Figure 2.

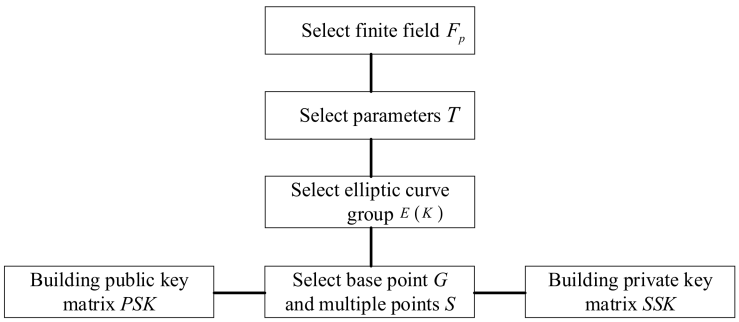


Figure 2. Construction of key matrix.

Based on the given parameter $T = (a, b, G, n, p)$, a public and a private key matrix can be constructed. The public key matrix's dimension is $m \times h$, and the elements are $X_{i,j}(1 \le i \le m, 1 \le j \le h)$ where $X_{i,j} = (x_{ij}, y_{ij}) \in S$. The public key matrix is denoted as PSK. The elements in the private key matrix are denoted as r_{ij}, and the private key matrix is denoted as SSK.

A public key and private key pair can be expressed as

$$r_{ij}G = X_{i,j} = (x_{ij}, y_{ij})(1 \le r_{ij} \le (n-1)) \tag{4}$$

The implementation process is shown in Figure 3. Firstly, the user identity information was used to generate the seed matrix through the hash function operation [27] and the row mapping algorithm. Suppose the row and column coordinates of the identification mapping value are $(i_1, j_1), (i_2, j_2), \cdots, (i_t, j_t)$, and the obtained public key is shown in Equation (5):

$$PK = (x_{i_1,j_1}, y_{i_1,j_1}) + (x_{i_2,j_2}, y_{i_2,j_2}) + \cdots + (x_{i_t,j_t}, y_{i_t,j_t}) \tag{5}$$

The obtained private key is shown in Equation (6):

$$SK = (r_{i_1,j_1} + r_{i_2,j_2} + \cdots + r_{i_t,j_t}) \bmod n \tag{6}$$

Hence, the generated key pair is shown in Equation (7):

$$\begin{aligned} PK &= (x_{i_1,j_1}, y_{i_1,j_1}) + (x_{i_2,j_2}, y_{i_2,j_2}) + \cdots + (x_{i_t,j_t}, y_{i_t,j_t}) \\ &= r_{i_1,j_1}G + r_{i_2,j_2}G + \cdots + r_{i_t,j_t}G = SK \times G \end{aligned} \tag{7}$$

SK is kept by the CPK key management center, and PK is open to users.

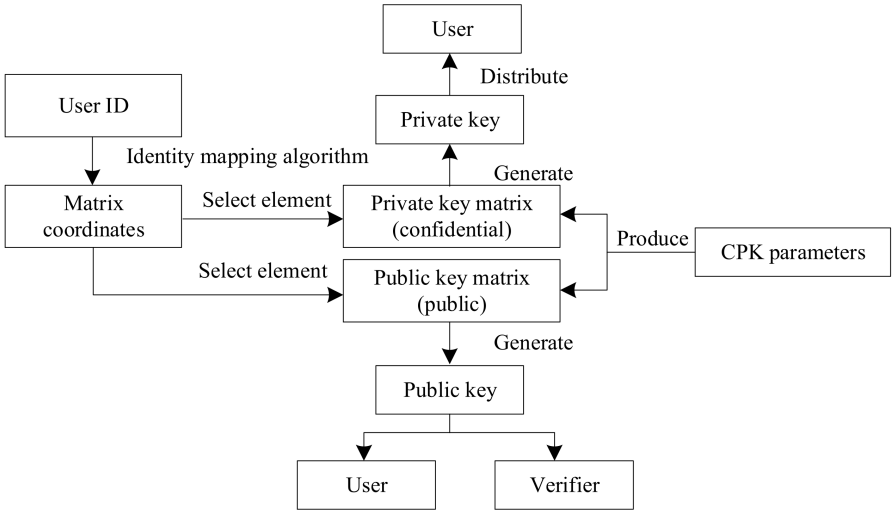


Figure 3. Process of key generation and distribution.

In the CPK-based identity authentication system, centralized generation, and the decentralized storage of keys is adopted [28]. The CPK system structure is shown in Figure 4.

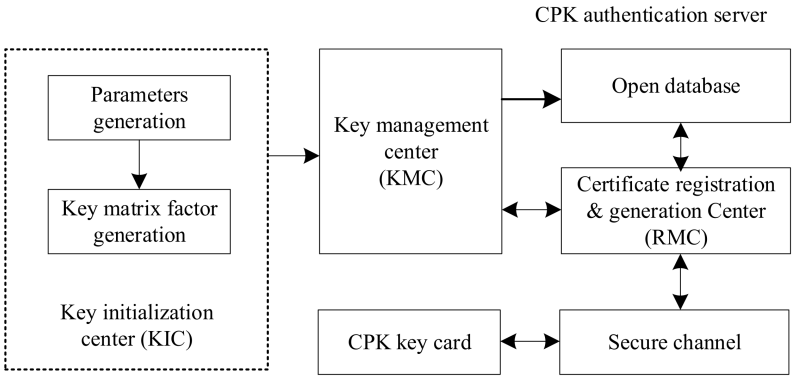


Figure 4. CPK authentication system modules.

Key initialization center (KIC): the parameter generation module generates the relevant parameters of the key matrix factor and transmits them to the key matrix factor generation module. In order to ensure the security of the system, KIC parameters, and algorithms are sent to KMC through a secure channel.

Key management center (KMC): it is responsible for generating the public and private key matrices. According to the parameters sent by KIC, the matrix is generated and sent to the public database, which is convenient for users to query. When KMC receives the application from the certificate registration generation system (RMC), it generates the ID certificate according to the user's ID and sends it to RMC after signing it with the algorithm.

Certificate registration and generation center (RMC): it is mainly responsible for certificate application, distribution, and other work. It accepts the user's application and then applies for the public key and private key from KMC, writes the private key into the CPK key card, protects it with the random number and distributes it to the user. The public key factor matrix is publicly accessible to the user, and the user can generate the public key according to the elliptic curve [29]. The ID certificate in the CPK key is an important structure to implement the system, in which the identity authentication of the terminal and the level division of the authority access are implemented based on the ID certificate.

Open database: it is mainly responsible for the management of accessible information database, and it is the interface for other modules of the system to access data.

CPK key card: the server authenticates the user's identity through the CPK key card. The user can use the private key safely stored in the key card and call the cryptographic algorithm in the CPK key card to verify the information transmitted by the server through the elliptic curve digital signature algorithm (ECDSA).

Interface: the API interface related to the encryption algorithm is provided for the application layer in the operating system SDK, and the interface provides the required function for the application that needs to call the encryption algorithm. The digital signature, encryption, and decryption of data can be implemented through this interface.

Secure channel: the secure communication channel between the terminal and the server.

3. Security System of Multi-Source Big Data in Power Monitoring Network

3.1. System Architecture

As shown in Figure 5, the system consists of a server-side security service and a monitoring terminal-side security agent for each monitoring terminal in smart grid networks. The server-side service includes the preliminary CPK module introduced in Section 2, the adaptive service module that adjusts the parameters of CPK based on the sensed environment, the encryption module that provides secure data transmission, the identity-based public key authentication service module that provides terminal identity verification and the communication module. Accordingly, the terminal-side agent is comprised of secure data storage that ensures the security of intermediate data, key-related information, authentication module, encryption module, and communication module.

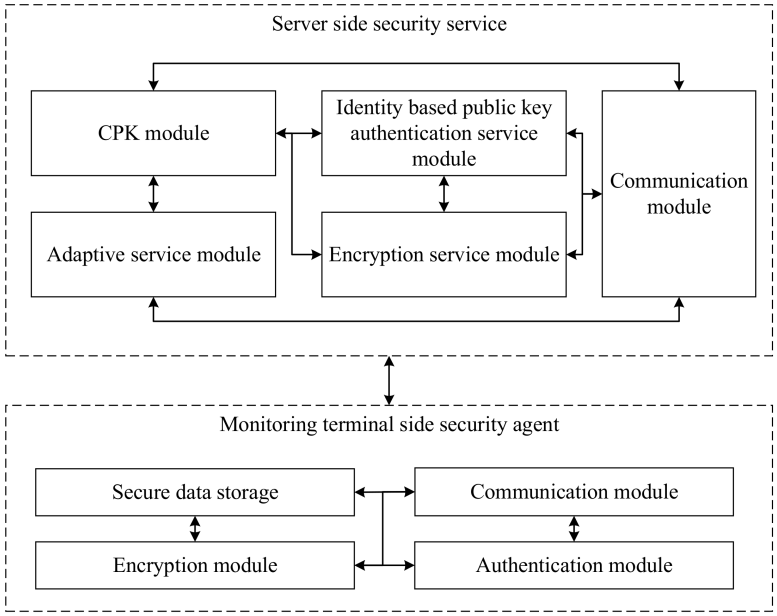


Figure 5. Security system architecture.

The working procedure of the system during the power monitoring process can be summarized as follows:

(1) The terminal agent sends the terminal's identification information to the CPK module via the communication module. If the CPK module approves the application, it generates a key pair of that terminal and sends it to the terminal as well as the public key matrix.

(2) While the terminal agent applies for access to a specific business application, the authentication process begins between the server-side authentication service module and terminal side authentication module. The details of the authentication process will be discussed in Section 3.2.

(3) Once the terminal identity is verified, the negotiated symmetric key can be used in the following encrypted data transmission between server-side the encryption service

module and the terminal-side encryption module. The encryption algorithm may use the (Advanced Encryption Standard (AES) or SM1, SM4 that are published by the State Cryptography Administration of China.

(4) When the data transmission is done, the server sider security service will decrypt the data and send them to the intranet application servers.

(5) During the above steps, the server-side adaptive service module monitors the status of connecting terminals and the current performance of CPK, and adjusts the parameters of the CPK module if necessary. The details of the adaptive service module will be discussed in Section 3.1.

3.2. Adaptive Key Fragment and Combination Method

Based on performing identity authentication through public and private key pairs generated by an identity-based cryptographic system to ensure the secure access of multi-source big data in the power monitoring network, an adaptive key slicing and combination method is proposed where the initial key is used as the seed key. Through the process of partition and combination, it is made up for the defects of low-key generation and distribution rate. The components of the adaptive service module are shown in Figure 5.

As shown in Figure 6, by monitoring the key generation center via communications between the key generation rate monitoring module, a key resources monitoring module in adaptive service module, and an adaptation assisting module in the CPK module, the generation rate of keys is identified that varies by time and the total key resources are obtained at each time. Meanwhile, the number of terminals in this area is detected and the key update speed is measured via the terminal status detection module. By comparing the current key-related performance of the CPK module and the required key-related performance of connecting terminals, the key fragmentation and combination mode is adjusted via communications between the key generation adjustment module of the adaptive service module and adaptation assisting module in CPK module in order to fulfill the performance demands of multi-source big data encryption in the power monitoring network.

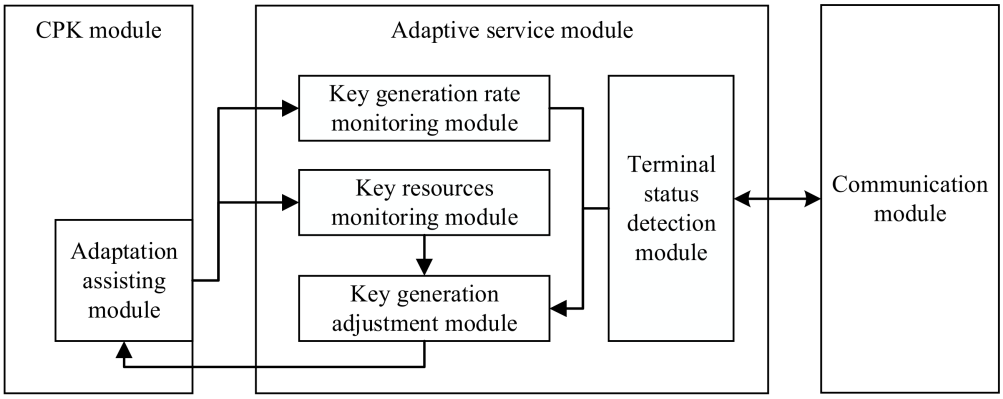
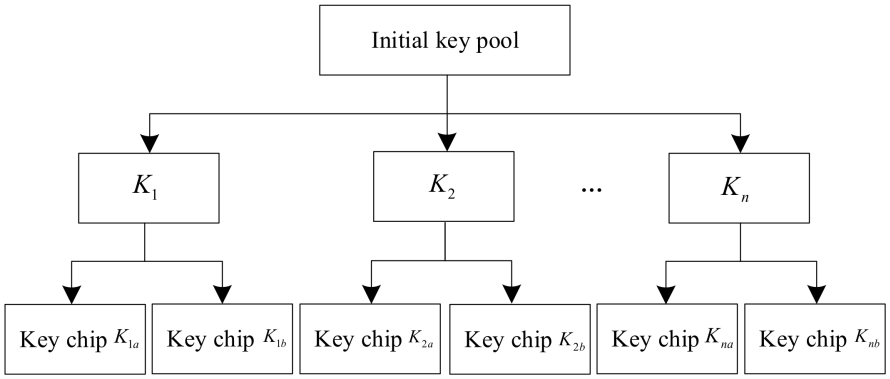
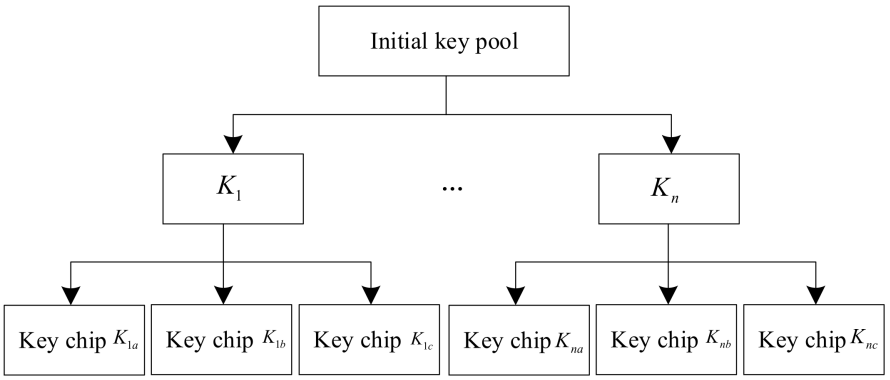


Figure 6. Components of the adaptive service module.

The method of key fragmentation is to divide the initial key pool into m fragment groups, called m sub-keys, each of which includes n keys, so that $L = m \times n$ sub-key resources are formed after fragmentation. Then, the sub-keys are further used to generate the corresponding keys. The key fragmentation method can adopt a two-segment or multi-segment method. Two-segment and three-segment methods are taken as examples whose processes are shown in Figure 7.



(a) Two segmentation mode



(b) Three segmentation mode

Figure 7. Demonstration of two-segment and three-segment modes.

Firstly, the number of key generations is compared with the number of keys required, and then the number of key segments is determined. The relationship between the different partition numbers and the corresponding total number of key generations is:

$$K_Z = C_{m\times n}^{m} = C_L^m \tag{8}$$

if a two-fragment mode is adopted where the key pool has n keys. After the two fragments, a total of $2n$ sub-keys are formed. The set of sub-keys is shown in Equation (9):

$$R = \{K_{1a}, K_{1b}, K_{2a}, K_{2b}, \ldots, K_{na}, K_{nb}\} \tag{9}$$

Considering the combination of keys, C_{2n}^2 combined keys can be generated.

If the three-segment mode is taken, a total of $3n$ sub-keys are formed after fragmentation, and C_{3n}^3 combined keys can be generated by considering the combination of keys. The number of key segments can be adjusted according to the number of terminals.

The next step is to combine the sub-key resources to generate keys:

$$T_K = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1j} \\ r_{21} & r_{22} & \cdots & r_{2j} \\ \cdots & \cdots & \cdots & \cdots \\ r_{i1} & r_{i2} & \cdots & r_{ij} \end{bmatrix} \tag{10}$$

As shown in Equation (10), the rows of the matrix represent the key variables of the combined key, and the columns of the matrix represent the number of segments of the combined key. Let the matrix be a key factor matrix with i rows and j columns, where

$i = m \times n$, $j = m$, and i represents the number of sub-keys and j represents the number of stages of the combined key operation.

The key slices are extracted and combined to generate the corresponding keys. Based on the terminal identification information, the key factors are selected by the sub-key set in some columns of the key factor matrix as K_m:

$$K_m = (k_1, k_2, \ldots, k_L) \begin{pmatrix} r_{1m} \\ r_{2m} \\ \cdots \\ r_{Lm} \end{pmatrix} \tag{11}$$

Then, the new key is generated as

$$K_{ID} = (K_1 + K_2 + \ldots + K_m) \mathrm{mod} n \tag{12}$$

Using the key combination method presented above, the public key is obtained from the public key matrix PSK and the private key is obtained from the private key matrix SSK to facilitate the multi-source data encryption in the power monitoring network, fulfilling the needs of the terminals for the large capacity key, and ensuring the security of multi-source big data in the power monitoring network.

3.3. Identity-Based Public Key Authentication Protocol

Aiming at the security requirements of rejecting malicious nodes in the smart grid, we propose an identity-based public key authentication protocol that combines CPK-based authentication and the public key encryption scheme and is integrated into the server-side and terminal-side authentication modules. The private key of the system is distributed and managed by a special private key generator. The secure access of multi-source big data in the power monitoring network is guaranteed by public key authentication based on identity information. Once the terminal applies the data transmission to the application server, both sides verify their identities using the proposed authentication protocol. Compared with PKI technology which requires the full online participation of a trusted third party, the proposed authentication protocol can be more efficient [30].

The process of the authentication protocol is described as follows: (where the "client" refers to the terminal-side security agent, and the "server" is the server-side security service).

(1) The client randomly generates a 128-bit seed, namely *rand*, and uses the current time to generate *timestamp*. In plaintext $M = rand + timestamp$, the hash value of *rand* and *timestamp* is calculated to obtain $H(M)$, and the client uses the private key to sign $H(M)$ to obtain $Sign_A(H(M))$. The identification A of the client, the plaintext information M, and the signature $Sign_A(H(M))$ are taken as the information to be sent for the client. The client uses the public key matrix disclosed by CPK to map the server's unique identifier to calculate the server's public key PK_B. The client uses PK_B to encrypt the information that needs to be sent, and the cipher text C_1 is expressed as shown in Equation (13):

$$C_1 = E_{PK_B}(A + M + Sign_A(H(M))) \tag{13}$$

The client sends C_1 to the server through the secure socket layer.

(2) After the server receives the cipher text C_1, it decrypts it with its private key SK_B, and obtains:

$$D_{SK_B}(C_1) = (A + M + Sign_A(H(M))) \tag{14}$$

According to the public key matrix of the client's identifier A and CPK, the public key PK_A is obtained after mapping, and the public key can be used to verify whether the signature $Sign_A(H(M))$ of A is consistent with the received data. If it succeeds, it passes the signature verification and confirms that the information comes from the client. The

server verifies the *timestamp*. If the verification succeeds, the server reverses the random number *rand* bit by bit to obtain *rand_s* and uses the current server time to generate the timestamp *timestamp_s*, and combines *rand_s* and *timestamp_s* to obtain *M_S*, which is the plaintext information for server. The server uses the hash algorithm to calculate $H(M\_S)$, and uses its private key SK_B to sign $H(M\_S)$ to obtain $Sign_B(H(M\_S))$. Finally, the server uses PK_A to encrypt $M\_S + Sign_B(H(M\_S))$ to obtain the cipher text C_2, as shown in Equation (15), and sends it to the client:

$$C_2 = E_{PK_A}(M\_S + Sign_B(H(M\_S))) \tag{15}$$

(3) After receiving the cipher text C_2, the client uses their private key SK_A to decrypt it to obtain $M\_S + Sign_B(H(M\_S))$. Then, the client uses the server public key PK_B to verify whether the signatures $Sign_B(H(M\_S))$ and $H(M\_S)$ match. If it matches, the verification is successful. Otherwise, it fails, and the customer will be prompted with a warning that the authentication has failed. After passing the verification, the client gets *rand_s* and *timestamp_s*, and reverses *rand_s* by bit to get *rand*. Additionally, check the timestamp *timestamp_s*, judge the time of this session through the timestamp, and enter the next step within a reasonable range.

(4) The client extracts *rand_s* and *timestamp_s* to obtain $H(M\_S)$, and then uses a fixed key *Key* negotiated in advance to encrypt $H(M\_S)$ to obtain the session key K_S, which is expressed as

$$K_S = E_{Key}(H(M\_S)) \tag{16}$$

After obtaining the session key, the client uses the session key to encrypt *rand_s*, and obtain the cipher text information C_3 expressed as

$$C_3 = E_{K_S}(rand\_s) \tag{17}$$

Then, the cipher text information C_3 is sent to the server.

(5) Negotiation of the session secret key. The server obtains the hash value $H(M\_S)$ of *rand_s* and *timestamp_s*, encrypts $H(M\_S)$ with the symmetric key *Key* negotiated in advance, obtains the session key K_S, which is used to decrypt the cipher text C_3. The server matches its *rand_s* with the decrypted information. If they agree, the key negotiation is successful. Both parties use K_S as the session key for this communication. The server uses K_S to encrypt $rand\_s + 1$ to obtain C_4 as shown in Equation (18) and returns it to client:

$$C_4 = E_{K_S}(rand\_s + 1) \tag{18}$$

(6) After receiving C_4, the client uses the session key K_S of this communication to decrypt, and compares the result with $rand\_s + 1$. If it matches, the key negotiation is successful. The communication parties use $rand\_s + 1$ as the starting number of the communication packet to avoid replay attacks.

4. Experimental Results and Analysis

The proposed system is developed and tested in the CentOS Linux 8.0 environment and Docker engine on the server side and various Linux editions on the terminal side. The development environment is the Visual Studio Code and MATLAB. Run the key management center module to construct the basic private key matrix, the auxiliary private and auxiliary public key matrix according to the elliptic curve parameters. The management center retains the basic private and the auxiliary private key matrix and discloses the auxiliary public key matrix to the Web service. Considering the research background of reference [21] that is similar to ours and the purpose to demonstrate the effectiveness of the proposed adaptation mechanism, the experiments are established using our system, our system without the adaptive service module and method in reference [21]. The experiments of Sections 4.1–4.3 aimed to verify the performance of the proposed system where

the authentication time [31], key generation quantity [32], and key update rate [33] are tested and compared. The experiments in Section 4.4 were established to test the security strength [30] of the proposed system.

4.1. Data Authentication Time Analysis

The number of simulated terminals is set to 500, 1000, 1500, 2000, and 2500, respectively. The data authentication time under different methods is calculated as the average authentication time of terminals over three repeated simulations in this experiment. The results are shown in Figure 8.

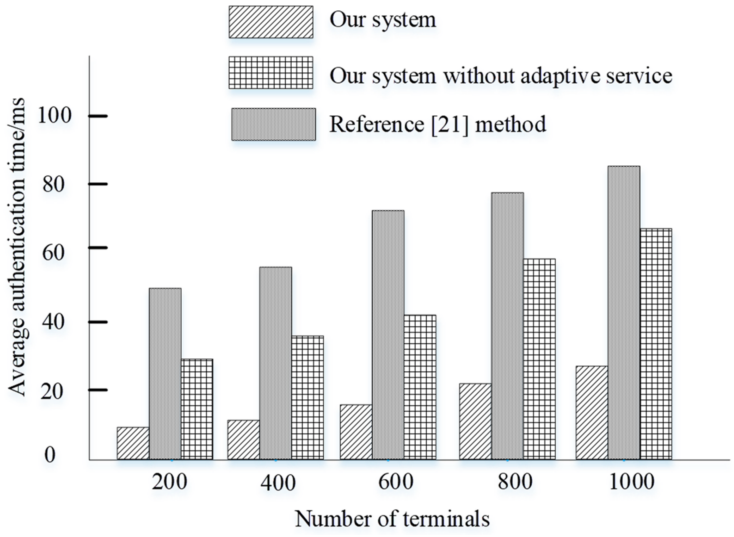


Figure 8. Comparison results of data authentication time.

As shown in Figure 8, it can be seen that with the increase in terminals, the average data authentication time of the three methods increases. Since the authentication time mainly relies on the computational capability of terminals and the authentication server, the communication channel quality and the authentication protocol, increased terminals may correspondingly increase the computational burden on both sides, so that the authentication time of both methods raises. It can also be seen that our system demonstrates a shorter authentication time than the other two methods under different terminal amounts. This might be due to the authentication protocol difference between the two methods. Reference [21] uses the PKI mechanism which involves the certification of a third-party authority while our system does not need that, so that the authentication process of our system requires fewer computational steps. In addition, it shows that the adaptive service may assist in reducing the authentication time.

4.2. Key Generation Quantity Analysis

The number of simulated terminals was set to 5000. The results of the number of key generations of different methods under different simulation periods are as shown in Figure 9.

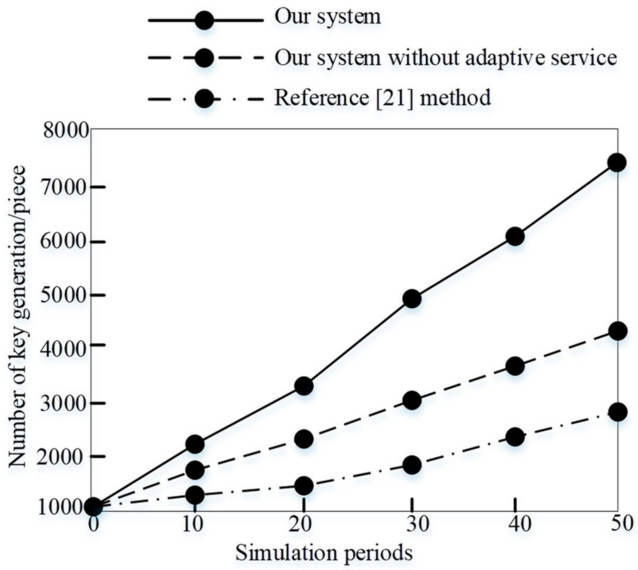


Figure 9. Results of key generation quantity.

It can be seen from Figure 8 that when the simulation period proceeds, the generated key increases to fulfill the communication demands of the terminals. When the simulation is run for 50 periods, the number of keys generated by the method reference [21] is 2900, the number of keys generated by our system without an adaptive service is 4200, and the number of keys generated by our system is 7400. It also can be seen that the number of keys generated by our system increases faster than the other two methods, so that the keys generated by our system can fulfill the demands of the terminal keys in power IoT in a shorter time. This may be due to the adaptive CPK method in our system where the requirements of terminals can be acquired in time and the hash function is used to map the key chip ID, and then extracts the corresponding key factors to combine them adaptively.

4.3. Key Update Rate Analysis

Several power monitoring network terminals, most of which are sensors in the power Internet of Things, were taken for this experiment. The number of simulated terminals is set to 500, 1000, 1500, 2000, and 2500, respectively. The key update rates of different methods are compared, as shown in Figure 10.

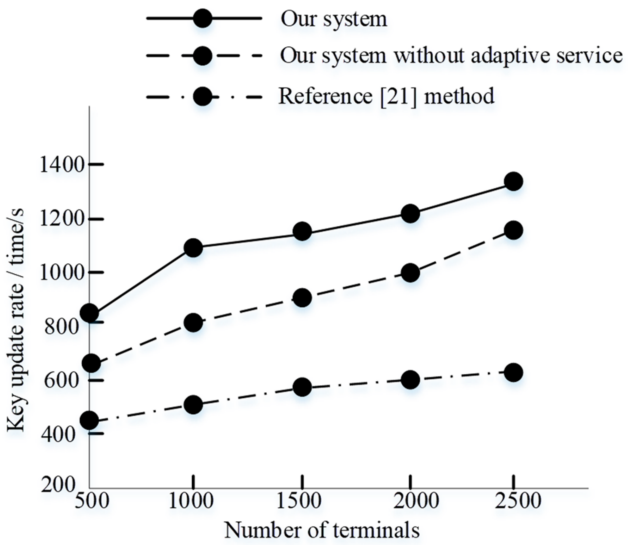


Figure 10. Results of key update rates.

It can be seen from Figure 8 that with the increase in the number of power monitoring network terminals, the key update rate increases in all three methods. When the number of power monitoring network terminals is 2500, the key update rate of reference [21] is 610 times/s, the key update rate of our system without adaptive service is 1190 times/s while the key update rate of our system is 1350 times/s. It can be seen that our system demonstrates the highest rate, since the initial key is used as seed key to generate terminal keys where adaptive key fragment and combination is adopted to facilitate the efficient key generation of a large number of terminals. The acceleration of the terminal key pair generation process and the following authentication will lead to the faster generation of negotiated symmetric keys that are used in secure communications between the server and terminals.

4.4. Data Security Analysis

The data security is tested and evaluated by simulated attacks under different methods in this experiment. The evaluation results are shown in Table 1.

Table 1. Results of data security evaluation of different methods.

Security Test	Our System	Our System without Adaptive Service	Reference [21]
Anti-forgery	✓	✓	✓
Anti-eavesdropping	✓	✓	✓
Anti-tampering	✓	✓	✓
Anti-repudiation	✓	✓	✓
Anti-interference	✓	✓	✓

It can be seen from Table 1 that our system may resist the attacks of the forgery of terminals, eavesdropping, tampering, repudiation, and interference. It can also be seen that the adaptive service does not affect the security strength of our system. The first four attacks were simulated by the test suite and the results show that those attacks can be resisted by the security mechanism of authentication and data encryption in our system. The fifth security was evaluated under the strong electromagnetic field environment of a substation and the results show that all methods remain functional. The implementation of our system in a substation scenario will be discussed in detail in Section 5.

5. Implementation of the Proposed System in Power System

The proposed security model was experimentally applied in a power IoT scenario in a substation of SGCC.

As shown in Figure 11, the implemented security system is comprised of a security server that includes a security gateway and a key management center, and agents installed on terminals. The proposed security system is implemented at the border of the intranet region to ensure the end-to-end security of the communications with terminals. The implementation of the security system can be easily done with minor network configuration changes. The key management center retains the primary private key, the private key factor matrix and public key factor matrix and is responsible for the terminal registration and key generation. The security gateway is a special type of terminal that integrates the access control mechanism and allows only legitimate terminals to connect, recording the terminal access history and authorizing the terminals to access specific power applications. Agents are installed on terminals that are responsible for establishing the secure connections with security gateway to ensure the data transmission security. The communication channel between terminals and the security gateway is compatible with power wireless networks such as LoRa and NB-IoT.

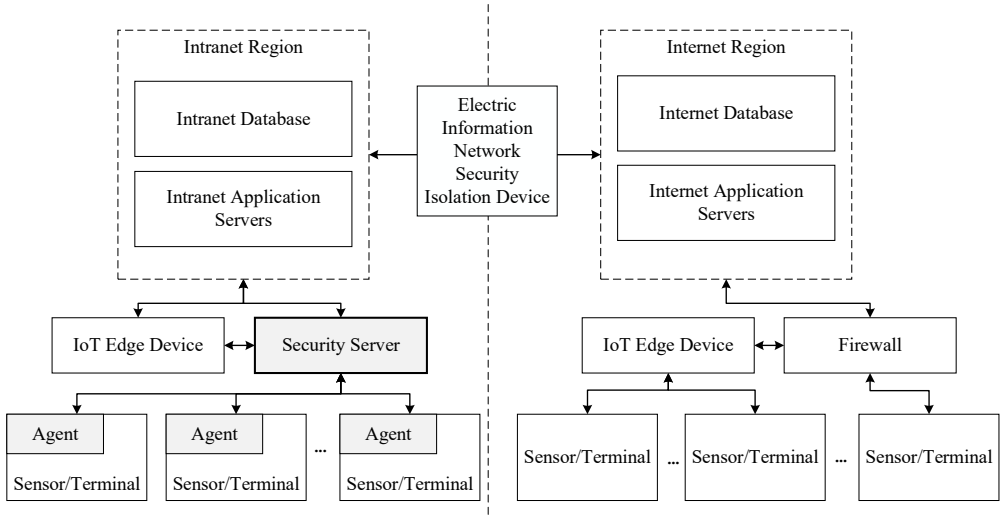


Figure 11. Implementation in power system.

As shown in Figure 12, the security gateway registers it on the key management center and obtains its private key and public key matrix. When a terminal connects to the security gateway for the first time, it sends its identification to the gateway and obtains the public key of the gateway and public key matrix. Then, it applies for registration to the gateway that consequently forwards the application to the key management center. If the registration application is disapproved, the gateway will disconnect the terminal and record this connection—including the terminal's identification. Otherwise, the key pair is generated by the key management center and forwarded to the terminal. Secondly, the terminal applies for authentication to the gateway. While the authentication process in the previous section is completed, the gateway will allocate the specific application access authority to the terminal according to predefined rules, then the terminal and gateway will start encrypted data transmission using the negotiated session key.

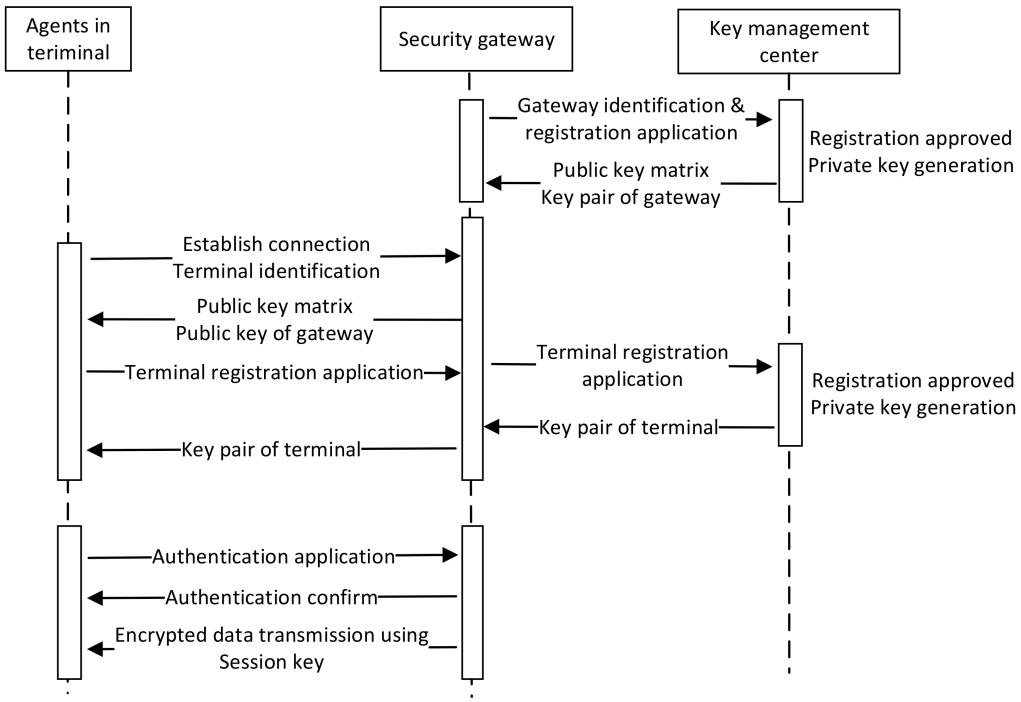


Figure 12. Working process of the experimental system.

Before the application of the proposed system, the power IoT terminals are directly connected to IoT edge devices or power applications through power wireless private network or public network so that the intranet servers have been facing the security risks of malicious terminals, the leakage of transmitting data, and so on. After the experimental application of the proposed system, a series of simulated attacks are carried out to test the security of the system including malicious terminal, eavesdropping, and tampering. The records of the security gateway show that the malicious (counterfeiting) terminal is rejected. The network packets are captured to demonstrate the encrypted data so that the data are protected. Finally, the test results show that the implementation of the proposed system can resist a series of attacks while the transmission performance is not compromised.

6. Conclusions

A security system of multi-source big data in power monitoring networks is proposed based on combined public key algorithm where an adaptive key fragment and combination method and an identity-based authentication protocol are described in detail. The proposed security system demonstrates the advantages in the aspects of the key update rate, key generation quantity, data authentication time, and implementation complexity while maintaining strong security strength in the multi-source big data environment of the power monitoring network. However, the stability of the communication between the client and the server needs further research. Compared with the actual environment, the test environment of the system in this paper has some differences in the number of concurrent users and access environment. How to improve the concurrent mechanism of terminal access is a problem that needs to be further solved. The big data features and associated advanced processing algorithms in the smart grid also need to be further studied such as incomplete big data [34], the prediction of missing data in big data [35], and deep learning in big data [36]. Meanwhile, future research may also pay attention to the combination of the combined public key algorithm with other popular security mechanisms in order to improve the adaptability of the proposed model for more application scenarios.

Author Contributions: Conceptualization, C.H.; methodology, C.J.; software, J.S.; validation, C.J., Q.H.; writing—original draft preparation, C.J.; writing—review and editing, C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the High Level Introduction of Talent Research Start-Up Fund of Nanjing Institute of Technology, grant number YKJ201989. Major Project of Philosophy and Social Science Research in Universities of Jiangsu Province Education Department, grant number 2020SJZDA069. Innovation Fund Major Project of Nanjing Institute of Technology, grant number CKJA201706.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Ganguly, P.; Nasipuri, M.; Dutta, S. Challenges of the Existing Security Measures Deployed in the Smart Grid Framework. In Proceedings of the 2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 12–14 August 2019; pp. 1–5.
2. Morello, R.; De Capua, C.; Fulco, G.; Mukhopadhyay, S.C. A Smart Power Meter to Monitor Energy Flow in Smart Grids: The Role of Advanced Sensing and IoT in the Electric Grid of the Future. *IEEE Sens. J.* **2017**, *17*, 7828–7837. [CrossRef]
3. Liu, J.; Zhao, Z.; Ji, J.; Hu, M. Research and application of wireless sensor network technology in power transmission and distribution system. *Intell. Converg. Netw.* **2020**, *1*, 199–220. [CrossRef]
4. Khan, F.; Siddiqui, M.A.B.; Rehman, A.U.; Khan, J.; Asad, M.T.S.A.; Asad, A. IoT Based Power Monitoring System for Smart Grid Applications. In Proceedings of the 2020 International Conference on Engineering and Emerging Technologies (ICEET), Lahore, Pakistan, 22–23 February 2020; pp. 1–5.

5. Hu, J.; Vasilakos, A.V. Energy Big Data Analytics and Security: Challenges and Opportunities. *IEEE Trans. Smart Grid* **2016**, *7*, 2423–2436. [CrossRef]
6. Huang, C.; Huang, Q.; Wang, D. Stochastic Configuration Networks Based Adaptive Storage Replica Management for Power Big Data Processing. *IEEE Trans. Ind. Inform.* **2020**, *16*, 373–383. [CrossRef]
7. Shobol, A.; Ali, M.H.; Wadi, M.; TüR, M.R. Overview of Big Data in Smart Grid. In Proceedings of the 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA), Brasov, Romania, 3–6 November 2019; pp. 1022–1025.
8. He, X.; Ai, Q.; Qiu, R.C.; Huang, W.; Piao, L.; Liu, H. A Big Data Architecture Design for Smart Grids Based on Random Matrix Theory. *IEEE Trans. Smart Grid* **2017**, *8*, 674–686. [CrossRef]
9. Wu, J.; Ota, K.; Dong, M.; Li, J.; Wang, H. Big Data Analysis-Based Security Situational Awareness for Smart Grid. *IEEE Trans. Big Data* **2018**, *4*, 408–417. [CrossRef]
10. Zhao, J.; Kamwa, I. Guest Editorial: Next Generation of Synchrophasor-based Power System Monitoring, Operation and Control. *IET Gener. Transm. Distrib.* **2020**, *14*, 3943–3944.
11. Liu, H.; Wang, Y.; Chen, W.G. Anomaly detection for condition monitoring data using auxiliary feature vector and density-based clustering. *IET Gener. Transm. Distrib.* **2020**, *14*, 108–118. [CrossRef]
12. Koziel, S.; Hilber, P.; Westerlund, P.; Shayesteh, E. Investments in data quality: Evaluating impacts of faulty data on asset management in power systems. *Appl. Energy* **2021**, *281*, 116057. [CrossRef]
13. Smith, M.D.; Paté-Cornell, M.E. Cyber Risk Analysis for a Smart Grid: How Smart is Smart Enough? A Multiarmed Bandit Approach to Cyber Security Investment. *IEEE Trans. Eng. Manag.* **2018**, *65*, 434–447. [CrossRef]
14. Andrea, I.; Chrysostomou, C.; Hadjichristofi, G. Internet of Things: Security vulnerabilities and challenges. In Proceedings of the 2015 IEEE Symposium on Computers and Communication (ISCC), Larnaca, Cyprus, 6–9 July 2015; pp. 180–187.
15. Wang, Z.; Liu, Y.; Ma, Z.; Liu, X.; Ma, J. LiPSG: Lightweight Privacy-Preserving Q-Learning-Based Energy Management for the IoT-Enabled Smart Grid. *IEEE Internet Things J.* **2020**, *7*, 3935–3947. [CrossRef]
16. Baek, J.; Vu, Q.H.; Liu, J.K.; Huang, X.; Xiang, Y. A Secure Cloud Computing Based Framework for Big Data Information Management of Smart Grid. *IEEE Trans. Cloud Comput.* **2015**, *3*, 233–244. [CrossRef]
17. Hossain, E.; Khan, I.; Un-Noor, F.; Sikander, S.S.; Sunny, M.S.H. Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review. *IEEE Access* **2019**, *7*, 13960–13988. [CrossRef]
18. Deng, S.; Chen, F.; Dong, X.; Gao, G.; Wu, X. Short-term load forecasting by using improved GEP and abnormal load recognition. *ACM Trans. Internet Technol.* **2021**, *21*, 1–28.
19. Jiang, Y.; Zhang, Y.; Xu, A.; Kuang, X.; Meng, J.; Chu, H. An Overview: Data Security Mechanism of Power Terminal in Edge Computing. In Proceedings of the 2020 IEEE International Conference on Energy Internet (ICEI), Sydney, NSW, Australia, 24–28 August 2020; pp. 22–27.
20. Chen, X.; Liang, W.; Zhou, X.; Jiang, D.; Kui, X.; Li, K. An Efficient Transmission Algorithm for Power Grid Data Suitable for Autonomous Multi-Robot Systems. *Inf. Sci.* **2021**, *572*, 543–557. [CrossRef]
21. Liu, R.; Zheng, Y.; Yang, Y.; Chao, Y.; Li, Y.; Yan, Y. Research on Secure Access Technology of Electric Power Wireless Private Network Based on Hybrid Encryption. In Proceedings of the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 18–20 June 2021; pp. 69–74.
22. Ni, J.; Zhang, K.; Lin, X.; Shen, X.S. Balancing Security and Efficiency for Smart Metering Against Misbehaving Collectors. *IEEE Trans. Smart Grid* **2019**, *10*, 1225–1236. [CrossRef]
23. Hong, J.; Liu, B.; Sun, Q.; Li, F. A combined public-key scheme in the case of attribute-based for wireless body area networks. *Wirel. Netw.* **2019**, *25*, 845–859. [CrossRef]
24. Zhang, F.; Zhang, Z.; Guan, P. ECC2: Error Correcting Code and Elliptic Curve based Cryptosystem. *Inf. Sci.* **2020**, *526*, 301–320. [CrossRef]
25. Cowan, A. The distribution of multiples of real points on an elliptic curve. *J. Number Theory* **2020**, *211*, 530–544. [CrossRef]
26. Jin, H.; Lee, H. Solving discrete logarithm problems faster with the aid of pre-computation. *Discret. Appl. Math.* **2019**, *267*, 93–119.
27. Do, T.T.; Le, K.; Hoang, T.; Lea, H.; Cheung, N.M. Simultaneous Feature Aggregating and Hashing for Compact Binary Code Learning. *IEEE Trans. Image Process.* **2019**, *28*, 4954–4969. [CrossRef]
28. Zhang, S.; Chen, Z.K.; Shi, R.H.; Liang, F.Y. A novel quantum identity authentication based on Bell states. *Int. J. Theor. Phys.* **2020**, *59*, 236–249. [CrossRef]
29. Zia, M.; Ali, R. Cryptanalysis and improvement of blind signcryption scheme based on elliptic curve. *Electron. Lett.* **2019**, *55*, 457–459. [CrossRef]
30. Cui, Y.; Yao, Y.; Xu, G.N. Research of Ubiquitous Power Internet of Things Security Authentication Method Based on CPK and RIFD. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; pp. 1519–1523.
31. Chen, D.; Zhang, N.; Cheng, N.; Zhang, K.; Qin, Z.; Shen, X. Physical Layer based Message Authentication with Secure Channel Codes. *IEEE Trans. Dependable Secur. Comput.* **2020**, *17*, 1079–1093. [CrossRef]
32. Yang, C.; Zhang, H.; Su, J. Quantum key distribution network: Optimal secret-key-aware routing method for trust relaying. *China Commun.* **2018**, *15*, 33–45. [CrossRef]
33. Xia, Z.; Zhou, H.; Gu, K.; Yin, B.; Zeng, Y.; Xu, M. Secure Session Key Management Scheme for Meter-Reading System Based on LoRa Technology. *IEEE Access* **2018**, *6*, 75015–75024. [CrossRef]

34. Wu, D.; Luo, X.; Shang, M.; He, Y.; Wang, G.; Zhou, M. A Deep Latent Factor Model for High-Dimensional and Sparse Matrices in Recommender Systems. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 4285–4296. [CrossRef]
35. Wu, D.; Shang, M.; Luo, X.; Wang, Z. An L_1-and-L_2-Norm-Oriented Latent Factor Model for Recommender Systems. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**. [CrossRef]
36. Luo, X.; Zhou, M.; Li, S.; Wu, D.; Liu, Z.; Shang, M. Algorithms of Unconstrained Non-Negative Latent Factor Analysis for Recommender Systems. *IEEE Trans. Big Data* **2021**, *7*, 227–240. [CrossRef]