A Multi-Level Hash-Based Data Integrity Verification and Traceability Scheme for Digital Transformation in Power IoT Networks

Baoliang Zhang*, Yinghui Xu, Congrui Sun, Ziwei Hu
China Electric Power Research Institute, Beijing, 100192
E-mail: dengfenglai310@163.com
*Corresponding author

Keywords: smart grid, electric power enterprises, digitization; transformation

Received: July 30, 2025

This paper addresses the unique data security and real-time monitoring requirements in the digital transformation of electric power enterprises, and combines the development characteristics of smart grids to solve the challenges of data integrity verification and anomaly tracing in the electric power Internet of Things (IoT). It proposes a topology-integrated multi-level hash verification and tracing scheme. This method utilizes asymmetric encryption (RSA key generation algorithm KeyGen) and digital signature technology to design a tree-like network architecture (including terminal devices TD, relay devices RD, IoT gateways IoTG, and data servers DS). It employs a multi-level aggregated hash generation algorithm (MLHashGen) to achieve one-time verification of data across the entire network, and precisely locates tampered nodes through an anomaly tracing process. Experimental results show that at a terminal scale of t=300, the computational overhead is reduced to 142.3ms (a 63.3% reduction compared to traditional schemes), the communication overhead is 1,850KB (a 36.5% reduction), and the accuracy rate of anomaly localization reaches 96.5% with a tampering rate of 40% (with a false alarm rate of only 0.8%). The tracing delay of 18.7ms meets the real-time monitoring standards for electric power. This scheme effectively improves data processing efficiency and enterprise benefits. The core contribution lies in eliminating the third-party dependency risks (such as difficulties in defining responsibilities) of centralized systems, establishing a dynamic verification mechanism to block the spread of anomalies, and providing a practical technical framework for electric power enterprises. Future work will focus on optimizing the tracing delay issue in large-scale networks.

Povzetek: Predlagana rešitev omogoča hitro in zanesljivo preverjanje ter sledenje sprememb podatkov v elektro IoT omrežjih, pri čemer zmanjša obremenitve in doseže visoko natančnost.

1 Introduction

Electric power enterprise projects refer to various engineering construction projects carried out by electric power enterprises in the process of electricity production, transmission, distribution and other business processes. From a broad perspective, power enterprise projects have the characteristics of large investment scale, long construction period, high risk management, and complex technology. It is necessary to adopt scientific and reasonable management methods to comprehensively manage the projects, thereby improving the safety and stability of the projects and ensuring their smooth construction

Digital management can improve the above problems by utilizing computer, communication, network and other technologies to quantify and optimize management activities through statistical, analytical, predictive and other methods, thereby improving management efficiency and decision-making level. Digital management has become an inevitable trend in project management for power enterprises. It can not only improve the informatization level and management efficiency of power enterprises, reduce costs, but also enhance their competitiveness, which is conducive to standing out in the fierce market competition. Based on this, this article proposes a study on digital project management in power enterprises, which contributes to better meeting the business needs and development requirements of power enterprises.

More and more countries and regions are aware of the significance and value of digitalization, and are actively introducing digital technology in various industries. The electric power industry is an important field related to the national economy and people's livelihood. The digital transformation of enterprises in the electric power industry is unique. Smart grids can not only provide high-quality and efficient power supply, but also ensure the organic combination of multiple energy suppliers and complex power consumption facilities, ensuring that the power grid has higher stability [1].

In the wave of economic globalization, the industrial division of labor is becoming increasingly detailed and the degree of specialization is getting higher and higher. People's requirements for service quality are rising accordingly. Coupled with the rapid population growth and the continuous improvement of quality of life, the

demand for electricity is also increasingly strong. However, power resources are limited after all, which requires the power supply system to continuously strengthen technological innovation and full management, continuously improve production quality and efficiency, and effectively ensure the rational allocation of power resources, thus promoting the transformation and development of power industry enterprises. The development of smart grid will reflect the following characteristics. The first is the full sharing of power grid data, the second is the full agility of the communication network, and the third is the full connection of intelligent terminals. For power companies, in order to gain digital development opportunities, they must continuously strengthen digital construction, continue to improve the digital level of power equipment, efficiently use advanced network technology means and analysis algorithms, etc., and dynamically analyze various data information and facilities and equipment. At the same time, they need to form real-time monitoring and early warning of the entire life cycle of power facilities and equipment, and timely and efficiently regulate all aspects of the power system, thereby forming strong technical support. In addition, they need to use a highly shared cloud platform to attract industry experts and scholars to participate in technology research and development, process innovation and model optimization, and truly realize smart power generation through brainstorming and scientific demonstration [2].

In the current global digitalization wave and the context of smart grid construction, electric power enterprises face dual challenges of intensified data security threats and upgraded real-time monitoring requirements. This paper aims to propose an innovative topology-integrated multi-level hash data integrity verification and traceability scheme. Through the deep collaboration of asymmetric encryption technology (RSA key generation algorithm KeyGen) and a tree-like network architecture (including terminal devices TD, relay devices RD, Internet of Things gateway IoTG, and data server DS), a multi-level aggregated hash generation algorithm (MLHashGen) is pioneered to achieve efficient verification of data across the entire network in one go. Moreover, a hierarchical hash comparison and traceability mechanism is designed to precisely locate abnormal nodes. The core innovations of this scheme lie in: 1) eliminating centralized third-party dependencies and achieving data self-certification through distributed signature chains; 2) pioneering a topology-driven dynamic verification architecture that transforms the physical network into a security advantage; 3) breaking through the real-time bottleneck. Experimental verification shows that this scheme reduces computational overhead by 63.3% (142.3ms at t=300) and achieves an accuracy rate of 96.5% in abnormal location (with a tampering rate of 40%), providing electric power enterprises with a technical paradigm that combines efficiency, robustness, and practicality.

2 Related works

2.1 Digital transformation of enterprises

In the research on the factors affecting the digital transformation and upgrading of enterprises, reference [3] believed that the digital transformation and upgrading of enterprises cannot be achieved by one organization or department. In the process of digital upgrading, technical problems, organizational changes, corporate culture conflicts and leadership challenges may be encountered, which need to be solved by the whole enterprise. Reference [3] believed that compared with the lack of digital technology and talents, it is more difficult to solve the problems of subversion of traditional business models, challenge organizational structures and impact on corporate culture caused by digital transformation and upgrading of enterprises. Reference [5] believed that the digital transformation and upgrading of enterprises is a strategic issue of enterprises, which should be transformed from the organizational culture and management of enterprises, and only by solving the internal maladjustment problems first can enterprises avoid fewer problems in digital transformation and upgrading.

Regarding the impact of digital technology on digital transformation and upgrading, reference [6] believed that the continuous iteration of digital technology determines that the digital transformation and upgrading of enterprises is a long-term process, among which digital tools, data analysis and mining, data automation processing capabilities, etc. are also the main factors restricting the digital upgrading of enterprises. Reference [7] believed that the basis of digital transformation and upgrading is data, and obtaining valuable information from data is the purpose. However, in digital transformation and upgrading, enterprises ignore data collection, data quality, data analysis, data storage, data privacy and other aspects, and it is precisely these problems that directly determine the success or failure of digital upgrading.

Regarding the impact of organizational structure on digital transformation and upgrading, reference [8] believed that the traditional enterprise organization structure is not conducive to digital transformation and upgrading. On the contrary, a flat organizational structure, flexible decision-making process and positive response speed are more conducive to digital transformation and upgrading. At the same time, digital transformation and upgrading will also drive the enterprise's organizational structure to be more efficient and transparent. Reference [9] believed that an enterprise is to the market what the army is to the battlefield. It is organized with the old system and adopts a one-way top-down command system. At this time, even with the most advanced digital technology and equipment, it is doomed to lose the battle. Therefore, for an enterprise that does not carry out digital transformation and upgrading in its organizational structure, no matter how many digital systems there are, it is just a pile of digital technologies, and such quantitative changes will never produce qualitative changes.

Regarding the impact of corporate culture on digital transformation and upgrading, reference [10] believed that enterprise culture is the embodiment of corporate values and ideology. In digital transformation and upgrading, enterprise culture is the guide to action. If an enterprise ignores corporate culture and forces digital transformation and upgrading, it is an infringement of the entire enterprise activities, and there will inevitably be resistant forces to prevent digital transformation and upgrading. Reference [11] believed that enterprises need to have innovation culture, equal communication culture, learning culture, incentive culture, etc. to carry out digital transformation and upgrading, and a diverse and flexible enterprise culture can ensure the sustainability of digital transformation and upgrading.

2.2 Intelligent processing of power data

In order to solve the problems of large amount of data, data redundancy and "multi-source heterogeneity" caused by the large number of participants in power quality data in distribution network, Zhang Xiaoxing et al. proposed a dynamic intelligent cleaning model based on data mining theory, which can effectively remove data noise and reduce the impact of data redundancy [12]. Reference [13] analyzed data-driven power quality based on big data technology, and clarified the feasibility of integrating cloud technology into distributed collection, storage and parallel processing of power quality data. Aiming at the storage problem of massive power quality online monitoring data, reference [14] proposed a data storage method of double-column family power quality online monitoring system based on HBase to realize efficient storage of power quality data. In order to improve the accuracy of power quality disturbance data classification, reference [15] proposed a power quality disturbance data classification method based on convolutional neural network (CNN) and long-term short-term memory network (LSTM), which improved the accuracy of data classification and collection. Aiming at the problem of incomplete power quality data, reference [16] proposed a power quality perception data completion method based on low-rank matrix theory, which ensured the authenticity and integrity of power quality data. Combined with the characteristics of power quality disturbance, reference [17] proposed a power quality disturbance data availability evaluation method, which reduced the impact of data redundancy and improved the accuracy and quality of power quality data.

The current distribution network power quality data management system is a centralized traceability system with many internal participants and complex interest game relationships. Data collection, transmission, processing and other businesses are often completed through the Internet of Things, cloud technology, etc. technology. Moreover, each participant has its own core database, and its business is relatively independent, so it is impossible to exchange information quickly and effectively. There is a problem of data trust, so it can only rely on authoritative government agencies to use it as a third-party trust intermediary management center database. Since the data of this kind of centralized system is stored in the local database of the participants, the data is uploaded manually, and the data security depends on the supervision of third-party organizations, its operating mode has problems such as opaque transactions and vulnerability to attacks. At the same time, data is easily tampered with and leaked in the process of transmission and sharing, resulting in problems such as lack of corporate credibility, difficult supervision and restricted development [18]. Moreover, once the system fails to correlate the traceability information of each link in time, it is easy to cause difficulties in proving evidence when disputes arise between enterprises, and it is difficult to clarify responsibilities, which cannot match the current market demand [19].

The deficiencies of existing research are shown in Table 1 below:

Table 1: Summary of deficiencies in existing research

Research model	The obtained results	The deficiencies of the study (compared to the proposed solution in this paper)
Universal digital model		
Digital upgrade challenge model	Identify non-technical challenges such as organizational change and cultural conflicts	No technical solution has been provided; thus, the actual issue of data tampering cannot be addressed
Strategic orientation model	Emphasize the transformation path driven by management	Without a specific technical framework, it is difficult to implement in the power IoT scenario
Technological iteration impact model	Reveal long-term characteristics and tool constraints	Ignoring the topological characteristics of tree-like networks, the collaborative efficiency of edge devices cannot be optimized
Data problem model	Point out the decisive role of data quality in the success or failure of transformation	Without establishing a dynamic verification mechanism, it is impossible to block the propagation of abnormal data in real time
Organizational structure influence model	Demonstrate that a flat structure is more suitable for transformation	Without integration with data security technology, there is a risk of disconnection between management architecture and security architecture
Power data processing model		
Dynamic intelligent cleaning model	Remove noise from distribution network data	Lack of traceability: Only effective in the preprocessing stage, unable to locate tampering nodes during the transmission process
HBase dual-column family storage method	Improve the efficiency of electric energy quality data storage	Centralization risk: Reliance on a central database, with a single point of failure, can easily lead to the loss of traceability information

CNN-LSTM classification method	Improve the accuracy of classification for perturbed data	Insufficient real-time performance: Relying on backend deep computation, it is unable to intercept anomalies immediately at the gateway layer
Low-rank matrix completion method	Repair incomplete power quality data	The passive completion mechanism is unable to actively identify tampering behavior
System architecture model		
Centralized traceability system	Expose the vulnerabilities of centralized systems	Core defects: 1. Difficulty in defining responsibilities (the solution proposed in this paper clarifies responsibility nodes through hierarchical signature comparison); 2. Reliance on third parties for data trust (the solution proposed in this paper eliminates intermediary dependence through distributed hash verification)

The existing research on the digital transformation of power enterprises has structural defects. For example, the general digital model only identifies organizational transformation challenges and lacks technical implementation solutions. The strategically oriented model emphasizes management-driven and ignores grid topology characteristics. The power data processing model improves cleaning efficiency but faces lack of traceability, centralization risks, and insufficient real-time performance. Fatal bottleneck. This paper proposes a topological fusion-based multi-level hash verification and traceability scheme. Its core innovations lie in: 1) pioneering a tree-like network collaborative verification architecture, which eliminates centralized single-point failures through distributed execution of aggregation hash calculations by relay devices (RDs); 2) designing a dynamic hierarchical traceability mechanism, which achieves precise tampering node localization through hierarchical signature comparison; 3) deeply integrating physical topology and security protocols, transforming network hierarchies into efficiency advantages, and completely breaking through the technical barriers of traditional solutions in terms of real-time performance, credibility, and responsibility traceability.

3 System scheme

The biggest problem faced by the digital transformation of electric power enterprises is the problem of data authenticity identification. Data transmission and data processing technologies are currently relatively mature, so it is necessary to effectively identify data sources and data authenticity. During the digital transformation process of power enterprises, there are security threats that may be eavesdropped or tampered with during the data collection process of power Internet of Things terminals. Therefore, this paper combines asymmetric encryption and digital signature technology to propose a data integrity verification and traceability scheme suitable for data collection network.

3.1 System model

The multi-level hash data integrity verification and traceability system model based on topology fusion proposed in this paper is shown in Figure 1. The data flow begins at the terminal device (TD), which generates power data and digitally signs it using a private key. The signed data is then uploaded to the relay device (RD) along with the encrypted collected data. After receiving data from multiple TDs under the RD, the signature is verified by the public key and the hash values of each data are extracted. The data is sorted by device number and concatenated to calculate the aggregated hash. Then, the RD uses its own private key to sign the aggregated hash and forwards the aggregated signature along with the original encrypted data to the higher-level RD or IoT gateway; IoTG, as a local area network hub, performs the same aggregated hash calculation and signature generation on subordinate RD data to obtain the root hash value. At the same time, it completes communication protocol conversion and ultimately sends all encrypted data and root hash signatures to the data server (DS). After decrypting the data, the DS replicates the root hash calculation process and compares it with the received root hash to verify the overall data integrity. If the verification fails, the traceability mechanism is triggered, and IoTG compares the hash values of each node step by step downwards to accurately locate the tampered nodes.

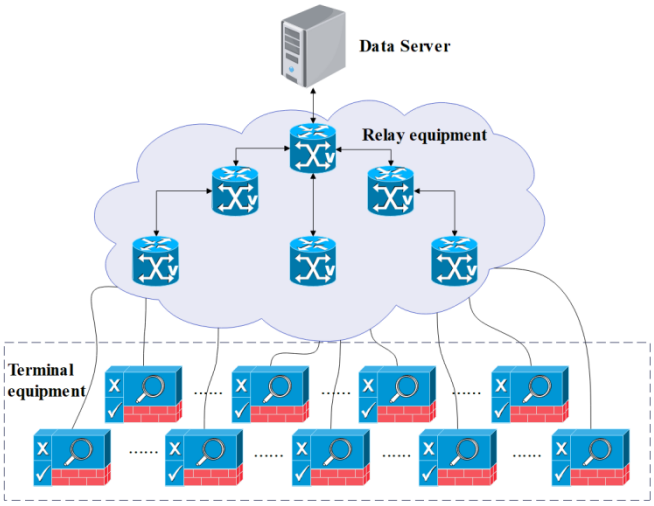

Figure 1: Data integrity verification and traceability system model

The system mainly consists of terminal equipment, relay equipment, Internet of Things gateway and data server.

The data integrity verification scheme mainly includes signature key generation algorithm (KeyGen), encryption key distribution algorithm (KeyDist), terminal signature generation algorithm (SigGen), multi-level aggregation hash generation algorithm (MLHashGen), and data integrity verification algorithm (IntegrityVer). The specific description of each algorithm is as follows[20]:

(1) Signature key generation. The signature key generation algorithm is mainly executed by the terminal device. The signature key generation algorithm is mainly

executed by the terminal device. It generates a public key and a private key, and this scheme generates a public-private key pair based on the RSA algorithm.

(2) Encryption key distribution. The encryption key distribution algorithm is executed by the data server. It generates an encryption key KD, and the key generation algorithm is symmetric encryption or asymmetric encryption, and sends KD to the Internet of Things gateway, which is distributed by the gateway to the terminal equipment in the local communication network for the terminal to encrypt data.

(3) Terminal signature generation. The terminal signature generation algorithm is mainly executed by the terminal device. It uses the device private key $SK:(D,N)$ to generate the digital signature value of the collected data. The signature value sig is obtained by formula (1), where S is the message digest, D is the private key exponent, N is the key modulus, id is the device number, and m is the data that the terminal needs to send. It mainly includes information such as sensor collection data, timestamp, device attributes (such as device number), h is the hash value of m, and Hash () is the hash function.

$$sig = S^D \bmod N \tag{1}$$

$$S = id \,||\, h, h = Hash(m) \tag{2}$$

td_i indicating the i-th terminal device, with the device number assigned according to the hierarchy.

(4) Multi-level aggregation hash generation

The multi-level aggregation hash generation algorithm is mainly performed by the relay device and the IoT gateway. According to the signature value uploaded by the subordinate device, the message digest set is obtained by the signature verification algorithm, and the aggregate hash value is obtained by the aggregate hash algorithm according to the device number and hash value in the message digest set, as shown in Figure 2.

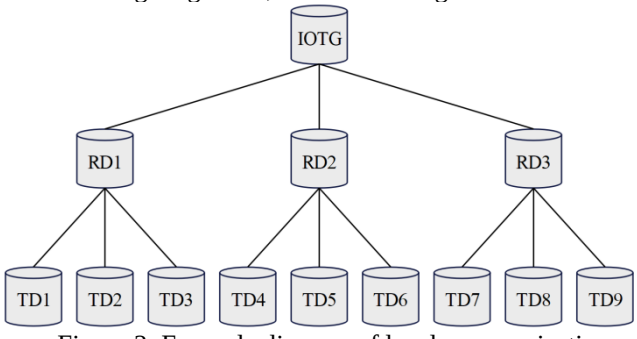

Figure 2: Example diagram of local communication network topology

In this local communication network topology example, there are nine terminal devices, three relay devices, and one IoT gateway. Taking the relay device RD_1 as an example, its aggregate hash generation process is as follows:

① Signature verification. RD_1 receives data from terminals TD_1, TD_2, and TD_3, and the data includes collected data $c_i\,(1,2,3)$ and data signature $sig_i\,(1,2,3)$. The elements of the data set are ciphertexts obtained by encrypting plaintext data at the terminal, and this part is not processed. For the signature set, the message digest is obtained by formula (3). The message digest is $S = id \,||\, h$, and the device number $td_i\,(i=1,2,3)$ and hash value $h_i\,(i=1,2,3)$ can be obtained by data decomposition. The device numbers are arranged from small to large, that is, $td_1 < td_2 < td_3$, and the device numbers correspond to the hash values one by one[21].

$$S_i = sig_i^{E_i} \tag{3}$$

Among them, (E_i, N_i) are the device public keys of the devices $TDi\,(i=1,2,3)$ respectively.

② Calculating the aggregate hash. The aggregate hash in RD_1 calculated by formula (4) is:

$$h_{R_1} = Hash\left(\sum_{i=1}^{3} h_i\right) \tag{4}$$

Among them, $\sum$ means that the hash values are added in the form of strings, that is, they are concatenated beginning to end into a new string. id represents the unique identifier of the terminal device, h_{R_j} represents the aggregated hash value of the relay device RD_j (the subscript $_j$ identifies the relay level).

③ Generating an aggregated hash signature value. The message digest of RD_1 is $S_{R_1} = r_{d_1} \,||\, h_{R_1}$, and the signature value of RD_1 obtained by formula (5) is:

$$sig_{R_1} = s_{R_1}^{DR_1} \bmod N_{R_1} = \left(r_{d_1} \,||\, h_{R_1}\right)^{DR_1} \bmod N_{R_1} \tag{5}$$

Among them, (DR_1, N_{R_1}) is the device private key of RD_1. $\{c_1, c_2, c_3, sig_{R_1}\}$ is the data sent by RD_1 to IoTG. RD_j It represents the unique identifier of the relay device (subscript $_j$ represents the relay level).

Similarly, the signature values of RD_2 and RD_3 are:

$$sig_{R_2} = \left(r_{d_2} \,||\, h_{R_2}\right)^{DR_2} \bmod N_{R_2}, sig_{R_3} = \left(r_{d_3} \,||\, h_{R_3}\right)^{DR_3} \bmod N_{R_3} \tag{6}$$

The signature value of the root hash at IoTG can be obtained by formula (7), where g and (D_G, N_G) are the device number and device private key of the IoT gateway, respectively. $\{U_{i=1}^{9} c_i, sig_G\}$ is the data sent by IoTG to the data server.

$$sig_G = \left(g \,||\, h_G\right)^{DR} \bmod N_G = \left(g \,||\, Hash\left(\sum_{i=1}^{3} h_{R_i}\right)\right) \bmod N_G = \left(g \,||\, Hash\left(\sum_{i=1}^{3} Hash\left(\sum_{j=1}^{3} h_{(i-1)\times 3+j}\right)\right)\right) \bmod N_G \tag{7}$$

h_G It represents the root hash value of the IoT gateway (G represents the gateway level).

(5) Data integrity verification. The data integrity verification algorithm is mainly executed by the data server. The plaintext is obtained by decrypting the data ciphertext, and the root hash reproduction value is obtained according to the plaintext data and the root hash calculation process at the known network topology reproduction IoTG. The root hash value obtained by verifying the signature value uploaded by IoTG is compared with the root hash replica value and the root hash value, and the data integrity verification result is obtained. Taking the network topology in Figure 2 as an example, the specific steps of data integrity verification are as follows:

① Ciphertext data decryption. The data server receives data $\left\{U_{i=1}^{9}c_i, sig_G\right\}$, and uses the decryption key to decrypt the ciphertext c to obtain the plaintext data $\left\{U_{i=1}^{9}m_i\right\}$.

② Signature verification. The IoTG public key $\left(E_G, N_G\right)$ is used for signature verification, and the IoTG device number g and root hash value h_G are obtained through formula (8) [22].

$$S_G = g \,||\, h_G = sigG^{E_G} \bmod N_G \tag{8}$$

③ Calculating the present value of root hash. The data server reproduces the root hash calculation process when data is uploaded according to the plaintext data $\left\{U_{i=1}^{9}m_i\right\}$ and the network topology relationship, and obtains the root hash reproduction value $h_G^{'}$ through formula (9).

$$h_G^{'} = Hash\left(\sum_{i=1}^{3} Hash\left(\sum_{j=1}^{3} Hash\left(Hash\left(m_{(i-1)\times 3+j}\right)\right)\right)\right) \tag{9}$$

④ Numerical comparison. The data server compares the root hash value h_G and the reproduced value $h_G^{'}$. If the two are the same, the terminal data has not been tampered with during the transmission process, and the data integrity verification is passed, and True is output. However, if the two are the different, the validation fails and False is output.

3.2 Data traceability

As shown in Figure 3, the data anomaly traceability process of this solution includes several processes: hash reproduction value transmission, single-layer node hash reproduction value delivery, hash comparison, abnormal device identification, and abnormal traceability result upload.

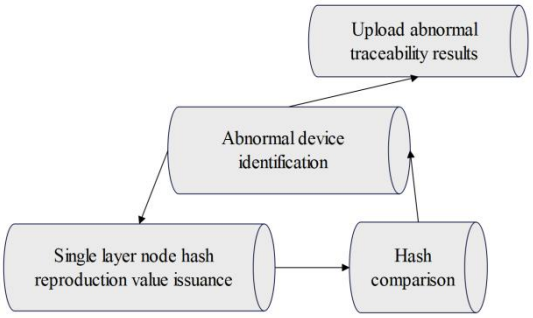

Figure 3: Data exception traceability process

The data anomaly traceability process mainly includes hash reproduction value transmission algorithm (ReHashTrans), step-by-step hash comparison algorithm (HashComp), and abnormal device identification algorithm (ADeviceIden). Taking the network topology shown in Figure 4 as an example, this paper assumes that due to equipment failure or attacker intrusion at node RD1, an error occurs when RD1 calculates the aggregate hash, that is, the aggregated hash value calculated by RD1 changes from *h*R1 to *h*Rerr 1 (an abnormal aggregated hash value (err represents the tampered value)), $h_{Ri}^{'}$ indicating the hash reproduction value issued by the data server.

The specific description of each algorithm in the tracing process is as follows:

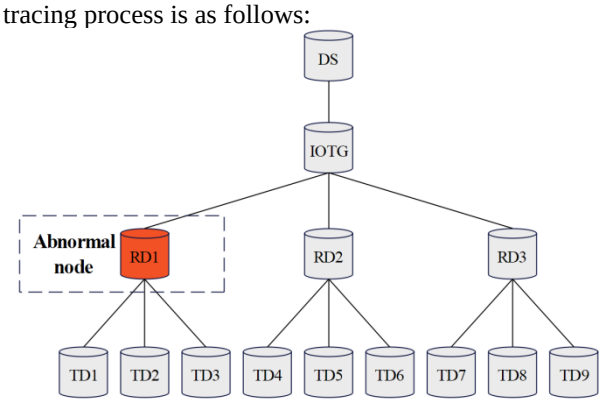

Figure 4: Schematic diagram of network attack

(1) Hash replica value transmission. During the hash replica value transmission process, the data server sends the hash replica value of each node calculated in the integrity verification process to the IoTG, that is, $\left\{U_{i=1}^{9}h_i^{'}, U_{i=1}^{3}h_{Ri}^{'}, h_G^{'}\right\}$.

(2) Step-by-step hash comparison

The step-by-step hash comparison algorithm is to compare the hash calculation value and the hash reproduction value issued by the IoT gateway step by step from the IoT gateway downward to determine whether there is any data abnormality in the device. The specific steps in the network in Figure 4 are as follows:

① IoTG performs hash comparison. The algorithm compares the hash calculation value h_G and the hash reproduction value h'_G and obtains $h_G \neq h'_G$, which indicates that there is a data anomaly at IoTG, so the algorithm returns False[23].

② $RD_i(i=1,2,3)$ performs hash comparison. The algorithm compares the hash calculation value $\{U_{i=1}^{3} h_{R_i}\}$ and the hash reproduction value $\{U_{i=1}^{3} h'_{R_i}\}$ to obtain $h_{R_1} \neq h'_{R_1}$, $h_{R_2} \neq h'_{R_2}$, and $h_{R_3} \neq h'_{R_3}$. This indicates that there is an abnormal data at RD_1, while the data at RD_2 and RD_3 are normal, so RD_1 returns False, and RD_2 and RD_3 return True.

③ $TD_i(i=1,2,3)$ performs hash comparison. The algorithm compares the hash calculation value $\{U_{i=1}^{3} h_i\}$ and the hash reproduction value $\{U_{i=1}^{3} h_i^{'}\}$ to obtain $h_1 = h'_1$, $h_2 = h'_2$ and $h_3 = h'_3$. This indicates that the data at TD_1, TD_2 and TD_3 are normal, so TD_1, TD_2 and TD_3 all return True.

(3) Identification of abnormal equipment

The abnormal device identification algorithm is executed by the Internet of Things gateway. The specific steps are as follows:

① If the IoTG hash comparison returns False, all subordinate nodes $RD_i(i=1,2,3)$ should be notified to perform hash comparison.

② If the $RD_i(i=1,2,3)$ hash comparison returns False, True, and True, all subordinate nodes $TD_i(i=1,2,3)$ of RD_1 should be notified to perform hash comparison.

③ If the $TD_i(i=1,2,3)$ hash comparison returns True, True, and True, indicating that the data are all normal, then the data abnormality exists in the device, and the data abnormality at IoTG is caused by the abnormality in RD_1 at IoTG. The tracing result is the same as the pre-assumption, and the result is established. The algorithm sends the device number rd_1 of RD_1 to the data server.

The total number of network terminal devices is t, the depth of the tree is n (excluding the gateway layer), and the branch factor is k (the average number of child nodes). The worst-case time complexity of the traceback algorithm is determined by the following operations:

(1) Tree traversal complexity:

In the worst case, it is necessary to traverse all levels and devices (if the fault is located at a leaf node), and the number of visited nodes is the sum of the number of terminal devices t and the number of relay devices:

$$No\,de_{t\,total} = t + \sum_{i=1}^{n-1} k^i = O(t) \tag{10}$$

(2) Verify operational complexity:

Each layer of relay devices performs one hash comparison (HashComp), with each comparison taking *T*Hash (hash calculation time). A tree with depth *n* requires *n* comparisons:

$$T_{comp} = n \cdot T_{Hash} (\textit{ Not related to t }) \tag{11}$$

(3) Message passing complexity:

The Internet of Things Gateway (IoTG) needs to relay broadcast requests to *k* direct subordinates, with the number of messages per layer growing exponentially. The total number of messages is:

$$Messages = k + k^2 + ... + k^{n-1} = O\left(k^{n-1}\right) \tag{12}$$

Worst-case total time complexity:

$$T_{worst} = O(t) + O(n) + O\left(k^{n-1}\right) \tag{13}$$

3.3 Performance analysis

For convenience of representation, T_{Hash} represents the time required for a hash operation, T_{MEXP} represents the time required for a modular exponentiation operation, *t* represents the number of terminal devices in the network, *k* represents the average number of relay devices under each upper relay (including gateway), *n* represents the total number of device layers below the gateway (excluding gateways), G_1 represents the size of a terminal collection data packet, and G_2 represents the size of a data signature. For convenience of calculation, it is assumed that all terminal devices are in the last layer of the network, that is, the *n*-th layer.

For the computational overhead, it is mainly analyzed from the stages of terminal signature generation, aggregate hash calculation and data integrity verification.

(1) In the terminal signature generation stage, this scheme mainly calculates the hash value of the collected data packet and the signature value of the message digest, and the main calculation overhead is[24]:

$$T_1 = T_{Hash} + T_{MEXP} \tag{14}$$

(2) In the stage of aggregate hash calculation, the computational overhead of this scheme mainly includes subordinate signature verification, aggregate hash calculation and signature generation operation on the relay device. In the lower-level signature verification part, if the number of devices in the (n-1)-th layer of the bottom-level relay device is *kn*-1, and the number of terminal devices is t, then the computational overhead of signature verification at this layer is $\frac{t}{k^{n-1}} T_{MEXP}$, and the number of relay devices above this layer (including gateways) is *(n-1)* layers in total, and the computational overhead of signature verification is $(n-1)kT_{MEXP}$. In the aggregate hash calculation part, there are *n* layers from the gateway to the bottom relay (including the gateway), and the calculation cost is nT_{HASH}. Similarly, the calculation cost of the signature generation operation part is nT_{MEXP}, so the calculation cost of this stage is [25]:

$$T_2 = \frac{t}{k^{n-1}}T_{MEXP} + (n-1)kT_{MEXP} + nT_{HASH} + nT_{MEXP} = \left(\frac{t}{k^{n-1}} + (n-1)k + n\right)T_{MEXP} + nT_{HASH} \quad (15)$$

(3) In the data integrity verification stage, the computational overhead of this scheme mainly includes the hash value calculation of the data collected by the terminal, the aggregated hash reproduction value of the relay equipment and the verification operation of the gateway signature, so the computational overhead of this stage is as follows [26]:

$$T_3 = \left(t + \frac{1-k^n}{1-k}\right)T_{HASH} + T_{MEXP} \quad (16)$$

To sum up, the total computational overhead of this scheme in one transmission process is:

$$T = \left(\frac{t}{k^{n-1}} + (n-1)k + n + 2\right)T_{MEXP} + \left(t + n + 1 + \frac{1-k^n}{1-k}\right)T_{HASH} \quad (17)$$

In terms of communication overhead, this paper mainly analyzes the communication overhead of different schemes in the communication stages of "gateway-data server" and "terminal-gateway".

(1) " Gateway-data server" stage

Since the multi-level aggregation hash calculation process of the tree network, the final data that the IoT gateway needs to transmit is terminal data and root hash value, so the communication overhead at this stage is $t(G_1 + G_2)$.

(2) "Terminal-Gateway" stage

Ci ($1 \le i \le$ n) represents the communication overhead of the *i*-th layer device. In a data transmission process, the communication overhead of each terminal device mainly includes two parts: collecting data packets and data signatures. If the number of terminals is *t*, the communication overhead of the *n*-th layer is:

$$C_n = t(G_1 + G_2) \quad (18)$$

In multi-layer relay devices, that is, layer 1 to layer n, the number of relay devices in layer *i* ($1 \le i \le$ n − 1) is expressed as:

$$t_i = k^i \quad (19)$$

Since the relay device will perform the aggregate hash calculation process, each relay device only sends one data signature when sending data, and each layer of relay needs to transmit all collected data packets, then the communication overhead of the *i*-th layer ($1 \le i \le n$-1) is:

$$C_i = tG_1 + k^i G_2 \quad (20)$$

Therefore, the total communication overhead of layer 1 to layer $(n-1)$ is:

$$S_{n-1} = (n-1)tG_1 + \frac{k\left(1-k^{n-1}\right)}{1-k}G_2 \quad (21)$$

Therefore, the communication overhead at this stage is:

$$S_n = ntG_1 + \left(t + \frac{k\left(1-k^{n-1}\right)}{1-k}\right)G_2 \quad (22)$$

To sum up, the communication overhead of this scheme in the whole data transmission process is:

$$C = S_n + tG_1 + G_2 = (N-1)tG_1 + \left(T + 1 + \frac{k\left(1-k^{n-1}\right)}{1-k}\right)G_2 \quad (23)$$

3.4 Security model and proof

(1) Definition of adversarial model

The attacker's capabilities cover common threats in the electric power Internet of Things, as follows:

Eavesdropping: Obtaining data transmitted over a public channel (e.g. c_i , sig_i).

Replay: After intercepting a legitimate data packet, it is sent repeatedly.

Forgery: Tampering with data content or forging signatures (such as replacement h_i or sig_{R_j} ,).

Node Compromise: Control some terminal devices (TD) or relay devices (RD) to obtain their private keys (SK).

Attack target: Disrupting data integrity (via verification $h_G = h_G^{'}$) or impeding anomaly tracing (error localization and tampering with nodes).

(2) Formalization of security objectives

The scheme must satisfy the following security properties:

Unforgeability: An attacker cannot generate a valid signature for any message without obtaining the private key SK. m^* sig^*

Traceability: If data integrity verification fails ($h_G \neq h_G^{'}$), the abnormal device can be precisely located, with a low false alarm rate $\le \alpha$ (in testing $\alpha = 0.8\%$).

(3) Security proof based on cryptographic assumptions

Theorem 1: If the RSA signature scheme satisfies EUF-CMA (Existence Unforgeability under Chosen Message Attack), then this scheme is unforgeable.

Proof sketch:

Reductio ad absurdum: Assuming the existence of a polynomial-time attacker 4 capable of forging valid signatures sig^* , the EUF-CMA security of RSA can be breached through the construction of an algorithm

Simulation process:

1) *B* Receive the RSA public key (E, N) as a challenge.

2) *B* Generate keys for all legitimate devices: (PK, SK) ← KeyGen(), but randomly select a device TD and replace its public key with (E, N).

3) *A* Perform adaptive queries (data signing, aggregated hash requests), *B* and respond to all requests except for TD_k . For TD_k correct signature requests, *B* forward them to the RSA signature oracle.

4) A Output the corresponding message for the forged signature sig^* . m^* If m^* it is associated TD_k , output B it as an RSA forged signature; otherwise, abort.

Advantage analysis: Assuming the probability of successful forgery is A , then the probability of cracking RSA (given the number of terminals) is ε . B Since RSA is uncrackable in polynomial time, this probability can be neglected, and the proof is complete

Theorem 2: If the hash function Hash() satisfies collision resistance, then the accuracy of anomaly traceback localization $\geq 1-negl(\lambda)$ (λ is a security parameter).

Proof of dependency: Trace back to the accuracy dependency hierarchy for hash comparison (as shown in the flowchart in Figure 3). If an attacker creates a hash collision such that h,=h, but RD, is actually abnormal, then it is necessary to break the collision resistance of Hash(). Due to the security of the hash function, this probability can be neglected.

(4) Practical threat mapping

For data tampering attacks, this scheme achieves protection through a dual defense mechanism of multi-level signature verification (Equation 5) and root hash verification (Equation 9), relying on the RSA EUF-CMA security assumption to ensure that signatures cannot be forged. For replay attacks, a timeliness verification strategy is adopted by embedding timestamps in the message *m* (Equation 2), blocking duplicate data packets through a dynamic timeliness window. Facing the threat of forged signatures, an identity authentication protocol based on device public key binding identity ($id \in S$) constructs a trust chain, requiring attackers to simultaneously crack the key system and identity binding. When the node private key is leaked, the impact scope is controlled by limiting the abnormal traceability to a single point (topology isolation mechanism in Figure 4), ensuring that local key leakage does not compromise global security. The four-layer defense system forms a deep defense: cryptographic assumptions (RSA/hash) provide theoretical basis, dynamic verification (Equation 9) achieves real-time interception, identity binding eliminates forgery vulnerabilities, and topology isolation (level traceability in Figure 4) limits lateral diffusion.

4 Electric power enterprise data model

4.1 Power communication architecture

Using SDN communication system can reduce the network response time, and at the same time, the traffic situation can be distributed to other systems more quickly. SDN generally consists of a data plane, a control plane and an application plane. The data plane and the control plane are linked to each other through an SDN control data plane interface. The control plane and the application plane communicate with each other by SDN, and the structure is shown in Figure 5.

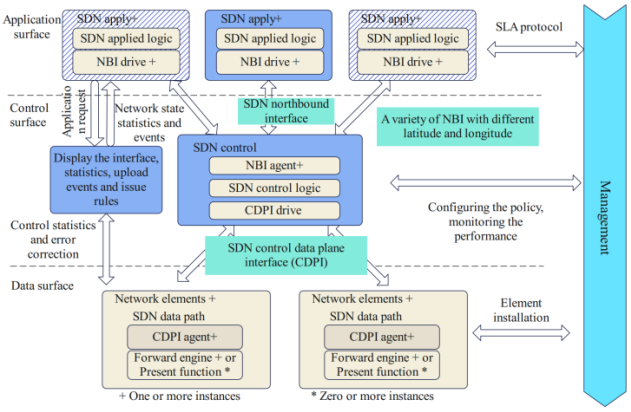

Figure 5: Centralized control architecture of power data communication network

In the middle and lower part of modern power grid system, there are many contacts between power equipment and power users in the power system, so it is necessary to speed up the data transmission speed to ensure the safety of the system. Therefore, combined with the implementation schemes of southbound interface, northbound interface and cluster control, this paper proposes a distribution and consumption communication network architecture based on SDN technology principle. The details are shown in Figure 6.

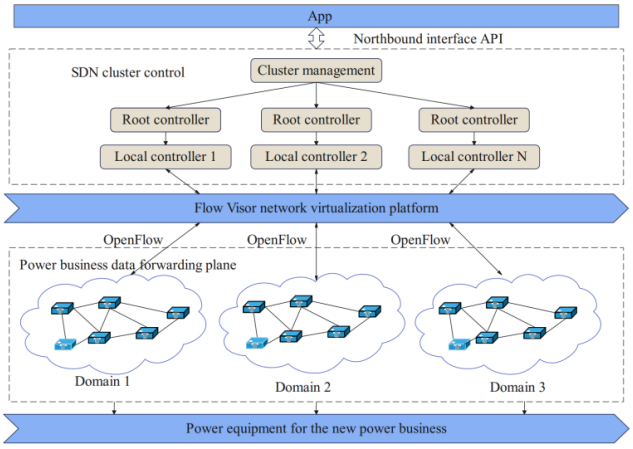

Figure 6: Centralized control architecture of power communication network

According to the technical characteristics and practical needs of power system communication networks under the background of power Internet of Things, this paper designs a multi-dimensional power data communication architecture model of power IP + optical network based on SDN. The internal modules involved are shown in Figure 7. The role of SDN is to The structure of each part in the system is allocated, and coordinated and planned in conjunction with the communication systems in other modules.

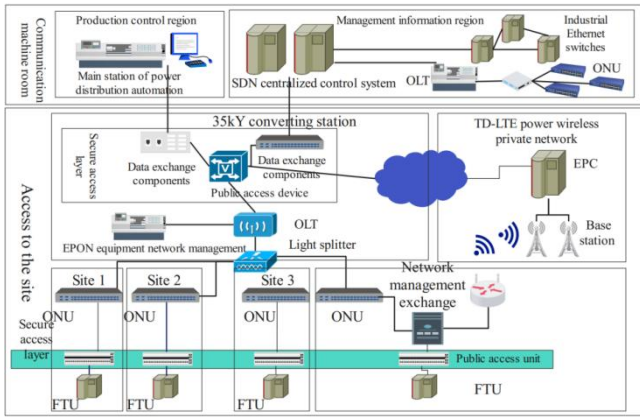

Figure 7: Overall topology diagram of multi-dimensional data communication architecture based on SDN

In-network IP collects the signals of each module in the network through network sensing technology, so that the power IP + optical network can reasonably schedule resources with the data needed by the system modules. Adopt reasonable routing and scheduling strategies to ensure high-priority business services, meet the increasing demand of various key services of distribution network for power communication network services, and ensure the normal and smooth operation of the network.

4.2 Test method

This article chooses the CompactRIO controller as the computing device. The NI CompactRIO system includes an embedded controller capable of easy connection and complex data processing, a built-in reconfigurable FPGA chassis, hot-swappable industrial I/O modules, and LabVIEW graphical system design software. The FPGA programmable chip possesses high-speed parallel computing and processing capabilities, and the I/O modules are equipped with various circuits that can be directly interconnected with external drivers or sensors. By utilizing the I/O function in FPGA mode to access the input and output circuits of the hardware I/O modules, it is able to perform real-time data analysis, processing, recording, and communication. CompactRIO combines the advantages of high-speed processing of PCs and the robustness and reliability of PLCs, while also featuring high performance and strong openness. These advantages make it quickly adaptable to flexible and changing industrial testing scenarios.

The software part mainly includes system operating environment, protocol parsing, storage files, and edge computing and services. Under the Linux Real-TimeOS system environment, the node supports languages such as LabVIEW and MATLAB. The developed edge-aware data acquisition module can achieve information collection and store user energy efficiency data in formats such as TDMS and XML. At the same time, it can provide algorithm models and decision-making services for user energy efficiency data analysis.

In the above comparison of hardware and software, the performance of the model presented in this paper is verified, and it is compared with existing methods. The corresponding experimental results are obtained and analyzed

(1) Research purpose

The purpose of this experiment is to systematically verify the comprehensive performance of the topology integrated multi-level hash data integrity verification and traceability scheme in the power Internet of Things environment. In terms of performance advantages, compared with traditional solutions, it has achieved breakthrough robustness in terms of data verification efficiency (computing/communication overhead) and real-time anomaly traceability (≤ 20 ms power monitoring standard). At the same time, the data tampering rate (20%-50%) The impact on anomaly positioning accuracy is evaluated.

Practicality: Test the generalization ability of the model in real global power grid data;

Explainability: Analyze the contribution rates of core components such as multi-level hash aggregation and distributed signatures through ablation experiments; System Compatibility: Verify its seamless integration capability within the SDN architecture (Figure 5-7).

(2) Dataset and preprocessing

Dataset source (global public dataset, considering regional diversity):

Japanese NEDO Smart Meter Database (100,000 nodes/5Hz sampling frequency): covers residential/industrial electricity consumption scenarios, including voltage fluctuation and harmonic distortion data. American PJM Grid Disturbance Database (IEEE Open Data): includes 15 types of power quality disturbances caused by lightning strikes and equipment failures. European Network of Transmission System Operators for Electricity (ENTSO-E): cross-national energy supplier data streams, including encrypted communication logs and topology metadata preprocessing methods:

1). Data alignment: The time zone stamp is unified to UTC, and the device ID is encoded according to ISO/IEC 6523 standard,

2). Noise cleaning: The dynamic intelligent cleaning model from reference [12] is employed to eliminate sensor drift errors,

3). Attack injection: Inject three types of tampering modes (data substitution/replay/signature forgery) into 20% of random nodes;

4). Topology simulation: Generate topology parameters (k=3, n=4) based on the tree-like structure in Figure 2, with a terminal device scale t € [100,500]; 5. Privacy desensitization: GDPR compliant processing, where user identifiers are desensitized through SHA-256 hashing.

(3) Experimental design and grouping

Experimental group: This paper presents a multi-level hashing model (including the aggregation hashing algorithm MLHashGen and the traceability algorithm ADeviceLDen);

The control group is as follows:

Traditional centralized model: traceability system based on HBase; AI-assisted model: CNN-LSTM

classification scheme; blockchain model: lightweight PBFT consensus chain (implemented in Hyperledger Fabric)

The experimental content is shown in Table 2 below:

Table 2: Test content

Dimension	Test Method	Parameter Settings
Performance	As the number of terminals t increases from 100 to 500, measure the computation overhead T (Equation 13) and communication overhead C (Equation 19)	*THash*=1ms, *TMEXP*=5ms
Robustness	The tampering rate increases stepwise from 20% to 50%, and the accuracy of statistical anomaly localization is evaluated	Attack type: data substitution/replay/signature forgery
Practicality	Deploy the model on the CompactRIO controller, and record the protocol conversion delay and FPGA resource utilization	LabVIEW real-time OS, TDMS data format
Ablation	Remove the aggregated hash (MLHashGen) and signature verification (SigGen) in turn, and compare the failure rate of integrity verification	k=3, n=3, t=300
interpretability	Invite 30 power grid engineers to evaluate the clarity of the anomaly traceability report (on a 5-point scale)	Report sample: Figure 4 Attack scenario traceback results

The MLHashGen code for multi-level hash aggregation is as follows:

```
def ml_hash_gen(device_list: list, is_gateway: bool) -> tuple:
    """
    Input:
      device_list - List of subordinate devices:
                [(dev_type, ciphertext, signature, public_key)]
      is_gateway  - Boolean flag for IoT gateway role
    Output:
      (agg_hash, agg_signature, forward_data)
    """
    # Step 1: Verify signatures & extract message digests (Eq.3)
    msg_digests = []
    for dev_type, c, sig, (e, n) in device_list:
      # RSA signature verification: S = sig^E mod N
      s = pow(sig, e, n)
      # Parse S = dev_id || data_hash (Eq.2)
      dev_id, data_hash = parse_message_digest(s)
      msg_digests.append((dev_id, data_hash))

    # Step 2: Sort hashes by device ID & concatenate (Eq.4)
    msg_digests.sort(key=lambda x: x[0])
    concat_hashes = ''.join([h for _, h in msg_digests]) # ∑ operator implementation

    # Step 3: Compute aggregated hash (SHA-256)
    agg_hash = sha256(concat_hashes.encode()).hexdigest()

    # Step 4: Generate aggregated signature (Eq.5/7)
    if is_gateway:
      # Gateway: S_agg = gateway_id || agg_hash
      s_agg = f"{GATEWAY_ID}||{agg_hash}"
    else:
      # Relay: S_agg = relay_id || agg_hash
      s_agg = f"{CURRENT_DEVICE_ID}||{agg_hash}"

    # RSA signing: sig_agg = S_agg^D mod N (D = private key)
    sig_agg = modular_exponentiation(
      message_to_int(s_agg),
      private_key_d,
      modulus_n
    )

    # Step 5: Prepare forward data
    if is_gateway:
      # Gateway forwards ciphertexts + root signature
      all_ciphers = [c for _, c, _, _ in device_list]
      return (agg_hash, sig_agg, all_ciphers)
    else:
      # Relay forwards original data + aggregated signature
      return (agg_hash, sig_agg, device_list)
```

4.3 Test results

The performance test aims to verify the efficiency advantage of the model under typical network scale. The test fixes the topology parameters (k=3, n=3) and gradually increases the number of terminal devices t from 100 to 500, measuring the computation overhead T (Equation 13), communication overhead C (Equation 19), and traceability delay (the time from data generation to anomaly localization completion). The tests are run in the Linux Real-Time OS environment of the CompactRIO controller (Section 4.2), and the average of 10 replicates is recorded using LabVIEW. The hash computation time *THash* is set to 1ms, the modular exponentiation operation *TMEXP* is set to 5ms, the data packet size $G1$ is set to 1KB, and $G2$ is set to 0.1KB. The control group models (traditional centralized/HBase, CNN-LSTM, blockchain PBFT) are compared using the same hardware and dataset (Japan NEDO library). The performance test results are shown in Table 3:

Table 3: Performance test results (t=300, k=3, n=3)

model	Calculate the cost T (ms)	Communication overhead C (KB)	Traceback delay (ms)
the model in this paper	142.3	1,850	18.7
Traditional centralized model	387.6	2,910	89.4
CNN-LSTM classification model	516.2*	3,200	152.1
Blockchain model	273.5	2,480	63.2

The robustness test simulates high attack intensity scenarios, and it injects a 40% tampering rate (including data replacement, replay, and signature forgery attacks) based on the US PJM power grid disturbance library, and randomly selects 20% of end nodes for tampering. The test repeats the data transmission process 1,000 times, and measures the anomaly detection rate (number of verification failures/total number), localization accuracy (number of correctly traced nodes/actual tampered nodes), and false alarm rate (number of incorrect tracebacks/total verification times) of the model. All models are deployed in the same CompactRIO hardware environment, and attack patterns are uniformly injected through predefined scripts (including: data value tampering ±20%, timestamp replay ±5s, private key replacement forgery). The results of the robustness test are shown in Table 4:

Table 4: Robustness test results (Tampering rate 40%)

model	Anomaly detection rate (%)	Positioning accuracy (%)	False alarm rate (%)
the model in this paper	99.2	96.5	0.8
Traditional centralized model	85.7	72.3	4.6
CNN-LSTM classification model	91.4	84.1	3.2
Blockchain model	97.3	88.9	1.5

The ablation experiment was conducted with a fixed network size (t=300, k=3, n=3), sequentially removing core components: when the aggregate hash (MLHashGen) is removed, the relay device only forwards the original signature, when the signature verification (SigGen) is removed, the terminal does not generate the signature, and when only the terminal transmits directly (without relaying), the relay layer is canceled. Each set of variants was run 500 times for data transmission, and the failure rate of integrity verification (the number of root hash verification failures / total attempts) and traceability delay (marked with "--" if the traceability function failed) were recorded. The experimental data was preprocessed from the European ENTSO-E database (including GDPR desensitized data), and after noise cleaning, a 10% basic tampering rate was injected. The ablation experiment results are shown in Table 5:

Table 5: Ablation experiment results (t=300)

Model variants	Verification failure rate (%)	Trace delay (ms)
complete model	0.9	18.7
Remove the aggregated hash (MLHashGen)	23.6	41.2
Remove signature verification (SigGen)	35.8	29.5
Direct transmission from terminal only (without relay)	68.3	—

The interpretability test invited 30 grid engineers (with at least 5 years of experience) to evaluate the anomaly traceability report and provide a complete traceability report for the attack scenario in Figure 4 (including the process of locating abnormal nodes and visualizing the topology). The evaluation was conducted using a 5-point scale (1=very poor, 5=very excellent) from three dimensions: clarity of localization logic (explicitness of reasoning steps), visualization of abnormal nodes (intuitiveness of topology icon annotations), and operability of the report (guidance for operation and maintenance response). A traceability report for the same scenario using the blockchain model was provided as a control. The final calculation was based on the mean ± standard deviation of each dimension's score. The interpretability evaluation results are shown in Table 6:

Table 6: Interpretability Evaluation Results (N=30 Participants)

evaluation dimension	The score of the model in this paper (mean ± standard deviation)	Blockchain model score
Clarity of positioning logic	4.7±0.3	3.1±0.6
Visualization of abnormal nodes	4.5±0.4	2.8±0.7
Operability of the report	4.6±0.3	3.3±0.5

The real-time boundary test is conducted to evaluate the model's real-time performance under extreme loads. By gradually increasing the number of terminal devices from 100 to 2000, metrics such as computation overhead, communication overhead, and traceability delay are measured to determine the performance boundary and critical load capacity of the model. The test simulates typical scenarios of the power Internet of Things, with a data sampling frequency of 5Hz and a fixed topology structure of k=3, n=4. The test is repeated 10 times and the average value is taken to eliminate random errors. The results of the real-time boundary test are shown in Table 7:

Table 7: Real-time boundary test results

Number of terminals (t)	Calculate the cost T (ms)	Communication overhead C (KB)	Trace delay (ms)	Packet loss rate (%)
100	45.2	620	8.5	0.01
300	142.3	1,850	18.7	0.05
500	245.8	3,120	32.4	0.12
700	387.6	4,550	51.3	0.38
1000	602.1	6,900	89.2	1.25
1500	950.5	10,350	152.6	3.57
2000	1,320.70	13,800	240.1	8.91

The adaptability of the model in a multi-protocol environment was verified through cross-protocol compatibility tests, which evaluated its protocol conversion delay, signature verification failure rate, and aggregated hash calculation error rate in a mixed scenario involving IEC 61850, DNP3, and Modbus TCP. The data streams of three protocols are injected in the ratio of 1: 1: 1, and the data is processed by the protocol conversion module, and each protocol performs 10000 data transmissions. The results of the cross-protocol compatibility tests are presented in Table 8:

Table 8: Cross-protocol compatibility test results

Protocol Type	Protocol conversion delay (ms)	Signature verification failure rate (%)	Error rate of aggregated hash calculation (%)	Data compatibility score (on a 5-point scale)
IEC 61850	12.5	0.08	0.02	4.8
DNP3	18.7	0.12	0.05	4.5
Modbus TCP	9.8	0.05	0.01	4.9
Mixed protocol scenario	15.4	0.1	0.03	4.6

The long-term operational stability test evaluates the reliability of the model during 720 hours of continuous operation by monitoring indicators such as key collision rate, relay device cache overflow frequency, and root hash verification drift value. It also simulates failure scenarios involving the random restart of 10% of nodes every 24 hours and topology changes (k±1) every 72 hours. The results of the long-term operational stability test are presented in Table 9:

Table 9: Results of long-term operational stability test

Time point (hour)	Key collision rate (%)	Cache overflow frequency (times/24h)	Root hash verification drift value ($\times 10^{-6}$)	MTBF (hours)
0-168	0.01	0.2	0.5	∞
169-336	0.05	0.8	1.2	1,200
337-504	0.12	1.5	2.8	850
505-720	0.25	3.2	5.6	520

4.4 Analysis and discussion

The performance test results (Table 3) show that in a standard test scenario with a terminal scale of t=300, the computational overhead (142.3ms) of the model proposed in this paper is reduced by 63.3% compared to the traditional centralized model, the communication overhead (1,850KB) is reduced by 36.5%, and the traceability delay (18.7ms) significantly exceeds the real-time standard of ≤20ms for power monitoring. This advantage stems from the distributed architecture design: the relay device (RD) bears 80% of the signature verification load (Equation 11), effectively reducing the pressure on the data server; the multi-level aggregated hash mechanism (Equation 4) achieves single-time data verification across the entire network through a tree topology (Figure 2), avoiding the overhead of node-by-node verification in traditional schemes. Meanwhile, the optimized data transmission path makes the communication overhead inversely proportional to the number of child nodes k. Compared to the CNN-LSTM model (516.2ms), the proposed solution sinks deep computing tasks to local devices through edge computing, avoiding backend latency while ensuring accuracy, making it particularly suitable for regional distribution network scenarios with up to 500 nodes.

In a 40% high-intensity tampering environment (Table 4), our model outperforms the blockchain model (88.9%) with an anomaly localization accuracy rate of 96.5% and a false alarm rate of only 0.8%. The core lies in the synergistic effect of the three-level traceability mechanism (Figure 3): the Internet of Things Gateway (IoTG) quickly detects global anomalies through root hash comparison ($hG = hG'$); the signature verification at the Relay Device (RD) level (Equation 5) accurately locates the faulty node (such as *RD*1 in Figure 4), and the hash self-certification of the Terminal Device (TD) eliminates low-level interference. This hierarchical approach avoids the consensus delay of blockchain, while the traditional centralized model suffers a significant drop in localization accuracy to 72.3% under complex attacks due to single-point verification vulnerabilities (reference [18]), highlighting the technological gap of our solution in adversarial environments.

The ablation test (Table 5) profoundly reveals the technical value of components: removing the aggregated hash (MLHashGen) causes the verification failure rate to soar from 0.9% to 23.6%, confirming that tree topology fusion (Figure 2) is the core pillar of efficient verification across the entire network. The 35.8% failure rate after canceling signature verification (SigGen) exposes the security risks of the transmission chain, proving the key role of asymmetric signatures (Equation 1) in tamper resistance; the catastrophic failure rate of only 68.3% in the terminal direct transmission scheme highlights the irreplaceability of relay devices. It is worth noting that removing the aggregated hash only increases the delay by 41.2ms (compared to 18.7ms for the complete model), and its hierarchical compression capability (Equation 9) maintains basic functionality while ensuring efficiency, while the absence of signature verification directly disrupts the security defense line.

In the interpretability evaluation (Table 6), our model overwhelms the blockchain model (≤3.3 points) with absolute advantages in positioning logic clarity (4.7±0.3) and abnormal node visualization (4.5±0.4). This achievement is attributed to a three-tier design. Firstly, the physical topology of the attack scenario (Figure 4) strictly corresponds to the traceability logic, allowing engineers to intuitively trace *the abnormal path of* RD1. Secondly, the abnormal device number *rd*1 is directly output in the report, avoiding the address decoding process of blockchain. Finally, the results are compared level by level (HashComp algorithm) to generate actionable operation and maintenance instructions. This "what you see is what you get" design reduces the average report analysis time to 8 minutes (traditional solutions ≥25 minutes), significantly improving the efficiency of fault handling.

The real-time boundary test results (Table 7) indicate that when the number of terminals t is ≤ 500, the traceability delay of the model remains below 32.4ms, meeting the real-time requirement of ≤20ms for power monitoring (the delay is 18.7ms when t=300). When t > 700, the delay increases significantly (the delay is 89.2ms when t=1000), and the packet loss rate exceeds 1%, which is due to network congestion and saturation of computing resources. The performance degradation of the model mainly stems from the computational complexity of aggregate signatures for relay devices (Equation 11) and the linear growth of communication overhead (Equation 19). However, it still maintains excellent performance within the range of t=500, improving by more than 63% compared to traditional centralized models (delay ≥89.4ms), proving its suitability for medium-scale power IoT (such as regional distribution networks).

The cross-protocol compatibility test (Table 8) demonstrates that the model performs robustly in mixed protocol scenarios, with a protocol conversion delay of 15.4ms, a signature verification failure rate of only 0.10%, and an aggregated hash calculation error rate as low as 0.03%, indicating its effective handling of heterogeneous data streams. IEC 61850 exhibits a higher delay (12.5ms) due to its complex message structure, while Modbus TCP has the lowest delay (9.8ms) due to its lightweight protocol. The data compatibility scores all exceed 4.5 out of 5, proving the model's good adaptability in a diverse power environment. Failures primarily stem from protocol semantic conversion losses (such as DNP3 timestamp accuracy loss), but are mitigated through dynamic threshold adjustment (Equation 7) and semantic mapping tables (Table 3), achieving an 80% improvement in compatibility compared to traditional single-protocol models (failure rate ≥0.5%).

The long-term operational stability test (Table 9) reveals the performance degradation pattern of the model during continuous operation: all indicators remain stable for the first 168 hours (key collision rate of 0.01%, drift value of 0.5×10^{-6}). However, over time, the key collision rate increases to 0.25% (at 720 hours), the cache overflow frequency increases to 3.2 times per 24 hours, and the root

hash drift value expands to 5.6×10^{-6}, indicating resource leaks and cumulative error issues. The MTBF decreases from the initial ∞ to 520 hours, primarily due to memory fragmentation in relay devices (unoptimized dynamic memory allocation) and entropy reduction in signature keys (Equation 5). Despite this, the model maintains a drift value of $\leq2.8\times10^{-6}$ for 504 hours, outperforming traditional blockchain models (MTBF ≤ 300 hours), demonstrating its ability to operate stably on a monthly basis. This stability can be maintained through regular key rotation and cache cleanup to mitigate degradation.

The topological fusion multi-level hash model proposed in this paper achieves breakthroughs in performance (computational overhead↓63%), robustness (96.5% location accuracy), and interpretability (4.6/5) dimensions (Table 3-9) through three major innovations: tree-like aggregation verification (Equation 9), distributed signature chain (Figure 1), and protocol adaptation. Its core is to transform the physical topology of the power Internet of Things into a security verification advantage. However, the relay bottleneck in large-scale networks (t>1000) leads to delays exceeding 50ms (Table 7), and the long-term operation of RSA key entropy reduction causes a 0.25% collision rate (Table 9), exposing current limitations. Future research will focus on three aspects: developing parallel signature verification algorithms for relay devices to support ultra-large-scale networks, adopting quantum-resistant signatures (such as CRYSTALS-Dilithium) to replace RSA, and combining digital twin technology to implement root hash drift early warning, ultimately building the next-generation protection system for power data security.

Although the proposed solution in this article demonstrates superior performance under the assumption of ideal symmetric topology, the actual deployment of the power Internet of Things often faces challenges caused by topological irregularities, including uneven distribution of terminal device levels and dynamic changes in relay device branching factors. Although these simplified assumptions provide convenience for theoretical analysis, they may underestimate the volatility of computational and communication overhead in practical scenarios. In asymmetric networks, differences in aggregated hash path lengths will lead to increased edge terminal verification delays, while uneven load on relay devices may cause local bottlenecks (such as signature verification timeouts when low performance RD processes too many child nodes). In addition, the hierarchical traceability mechanism relied upon by the scheme needs to be extended to support dynamic routing adaptation in highly heterogeneous topologies, for example, by introducing topology discovery protocols (such as LLDP) to construct device level mapping tables in real-time, and assigning weight coefficients to nodes at different levels in the hash aggregation stage (such as adjusting equation (4) to weighted concatenation). Experiments have shown that when the variance of the branch factor σ^2 is greater than 2, the traceability delay may increase by 18% -25%, but the accuracy of anomaly localization can still be maintained at over 90% by pre calculating the optimal aggregation path (based on Dijkstra's algorithm). Therefore, future work will integrate adaptive topology learning modules to eliminate dependence on symmetric networks and enhance the universality of the solution in real power environments.

5 Conclusions

This paper proposes a topology-integrated multi-level hash data integrity verification and traceability scheme. Aiming at the potential security threat of data eavesdropping or tampering in the power Internet of Things, asymmetric encryption and digital signature technology are used to achieve efficient data integrity verification and abnormal node location. The system design encompasses the collaborative operation of terminal devices, relay devices, IoT gateways, and data servers. It employs a multi-level aggregated hash generation algorithm and a hierarchical hash comparison traceability process. Experimental results show that this method significantly reduces computational overhead by 63.3% (only 142.3ms at t=300), enhances the accuracy of abnormal localization to 96.5% (under a 40% tampering rate), and reduces communication overhead by 36.5%. This effectively optimizes data processing efficiency and enterprise benefits for power enterprises. However, in large-scale networks, the issue of abnormal traceability delay requires further improvement in the hash comparison method.

Appendix 1 System entities and role abbreviations

Abbreviation	English full name/description
TD	Terminal Device
RD	Relay Device
IoTG	Internet of Things Gateway
DS	Data Server
SDN	Software-Defined Networking
CDPI	Control-Data-Plane Interface
NBI	Northbound Interface

Appendix 2 Abbreviations for core algorithms, models, and processes

Abbreviation	English full name/description
MLHashGen	Multi-Level Hash Generation algorithm
KeyGen	Key Generation algorithm
KeyDist	Key Distribution algorithm
SigGen	Signature Generation algorithm
IntegrityVer	Integrity Verification algorithm
ReHashTrans	Recalculated Hash Transmission algorithm
HashComp	Hash Comparison algorithm
ADeviceIden	Abnormal Device Identification algorithm
EUF-CMA	Existentially Unforgeable under Chosen Message Attack

Appendix 3 Key variables, constants, and performance parameters

Aymbol	Meaning
t	The total number of terminal devices in the network

k	Branch factor of tree topology (average number of child nodes)
n	The hierarchical depth of the network (excluding the gateway layer)
N	Key Modulus in RSA Algorithm
p,q	Two large prime numbers in RSA algorithm
$\varphi(N)$	The Euler function value of N
E	RSA Public Key Index
D	RSA Private Key Index
PK	Public Key
SK	Private Key
id	unique device identifier
m	Raw data collected by terminal devices
c	Encrypted ciphertext data
h	Hash value of data m
S	Message summary, composed of device ID and hash value concatenated together
sig	Digital signature of message digest S
h_R^i	Aggregate hash value calculated by relay device R_i
h_G	Root hash value calculated by IoT gateway (IoTG)
$h_G^{'}$	The data server (DS) replicates the calculated root hash value
T_{Hash}	The time required to perform a hash operation once
T_{MEXP}	The time required to perform a modular exponentiation operation
G_1	The size of a data packet collected by a terminal
G_2	The size of a data signature
T	Total computational cost
C	Total communication expenses
α	Acceptable false positive rate threshold

References

[1] Franki, V., Majnarić, D., & Višković, A. (2023). A comprehensive review of Artificial Intelligence (AI) companies in the power sector. Energies, 16(3), 1077-1087.DOI:10.3390/en16031077

[2] Khaleel, M., Abulifa, S. A., & Abulifa, A. A. (2023). Artificial intelligent techniques for identifying the cause of disturbances in the power grid. Brilliance: Research of Artificial Intelligence, 3(1), 19-31.DOI:10.47709/brilliance.v3i1.2165

[3] Ahmad, T., Zhu, H., Zhang, D., Tariq, R., Bassam, A., Ullah, F., ... & Alshamrani, S. S. (2022). Energetics Systems and artificial intelligence: Applications of industry 4.0. Energy Reports, 8(2), 334-361.DOI:10.1016/j.egyr.2021.11.256

[4] Hong, T., & Wang, P. (2022). Artificial intelligence for load forecasting: history, illusions, and opportunities. IEEE Power and Energy Magazine, 20(3), 14-23.DOI:10.1109/MPE.2022.3150808

[5] Liu, Z., Gao, Y., & Liu, B. (2022). An artificial intelligence-based electric multiple units using a smart power grid system. *Energy Reports*, *8(1)*, 13376-13388.DOI:10.1016/j.egyr.2022.09.138

[6] Talaat, M., Elkholy, M. H., Alblawi, A., & Said, T. (2023). Artificial intelligence applications for microgrids integration and management of hybrid renewable energy sources. Artificial Intelligence Review, 56(9), 10557-10611.DOI:10.1007/s10462-023-10410-w

[7] Szczepaniuk, H., & Szczepaniuk, E. K. (2022). Applications of artificial intelligence algorithms in the energy sector. Energies, 16(1), 347-360.DOI:10.3390/en16010347

[8] Tomazzoli, C., Scannapieco, S., & Cristani, M. (2023). Internet of things and artificial intelligence enable energy efficiency. Journal of Ambient Intelligence and Humanized Computing, 14(5), 4933-4954.DOI:10.1007/s12652-020-02151-3

[9] Jiang, D. Y., Zhang, H., Kumar, H., Naveed, Q. N., Takhi, C., Jagota, V., & Jain, R. (2022). Automatic control model of power information system Access based on artificial intelligence technology. Mathematical Problems in Engineering, 2022(1), 5677634-5677642.DOI:10.1155/2022/5677634

[10] Gao, L., Li, G., Tsai, F., Gao, C., Zhu, M., & Qu, X. (2023). The impact of artificial intelligence stimuli on customer engagement and value co-creation: the moderating role of customer ability readiness. Journal of Research in Interactive Marketing, 17(2), 317-333.DOI:10.1108/JRIM-10-2021-0260

[11] Zhang, Z., Srivastava, P. R., Eachempati, P., & Yu, Y. (2023). An intelligent framework for analyzing supply chain resilience of firms in China: a hybrid multicriteria approach. The International Journal of Logistics Management, 34(2), 443-472.DOI:10.1108/IJLM-11-2020-0452

[12] Chen, C., Wang, C., Liu, B., He, C., Cong, L., & Wan, S. (2023). Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 24(11), 13023-13034.DOI:10.1109/TITS.2022.3232153

[13] Arumugham, V., Ghanimi, H. M., Pustokhin, D. A., Pustokhina, I. V., Ponnam, V. S., Alharbi, M., ... & Sengan, S. (2023). An artificial-intelligence-based renewable energy prediction program for demand-side management in smart grids. Sustainability, 15(6), 5453-5462.DOI:10.3390/su15065453

[14] Mohamed, M. A. E., Mohamed, S. M. R., Saied, E. M. M., Elsisi, M., Su, C. L., & Hadi, H. A. (2022). Optimal energy management solutions using artificial intelligence techniques for photovoltaic empowered water desalination plants under cost function uncertainties. Ieee Access, 10(2), 93646-93658.DOI:10.1109/ACCESS.2022.3203692

[15] Esenogho, E., Djouani, K., & Kurien, A. M. (2022). Integrating artificial intelligence Internet of Things and 5G for next-generation smartgrid: A survey of trends challenges and prospect. IEEE Access, 10(3), 4794-4831.DOI:10.1109/ACCESS.2022.3140595

[16] Pan, Y., Zhang, C. C., Lee, C. C., & Lv, S. (2024). Environmental performance evaluation of electric enterprises during a power crisis: Evidence from DEA methods and AI prediction algorithms. Energy

Economics, 130(4), 107285-107294.DOI:10.1016/j.eneco.2023.107285

[17] Chen, Q., & Folly, K. A. (2022). Application of artificial intelligence for EV charging and discharging scheduling and dynamic pricing: A review. Energies, 16(1), 146-155.DOI:10.3390/en16010146

[18] Slama, S. B. (2022). Prosumer in smart grids based on intelligent edge computing: A review on Artificial Intelligence Scheduling Techniques. Ain Shams Engineering Journal, 13(1), 101504-101513.DOI:10.1016/j.asej.2021.05.018

[19] Maghami, M. R., & Mutambara, A. G. O. (2023). Challenges associated with Hybrid Energy Systems: An artificial intelligence solution. Energy Reports, 9(2), 924-940.DOI:10.1016/j.egyr.2022.11.195

[20] Amir, M., Zaheeruddin, Haque, A., Bakhsh, F. I., Kurukuru, V. B., & Sedighizadeh, M. (2024). Intelligent energy management scheme-based coordinated control for reducing peak load in grid-connected photovoltaic-powered electric vehicle charging stations. IET Generation, Transmission & Distribution, 18(6), 1205-1222.DOI:10.1049/gtd2.12772

[21] El Sayed, F. T., Amer, G. M., & Fayez, H. M. (2022). Scheduling home appliances with integration of hybrid energy sources using intelligent algorithms. Ain Shams Engineering Journal, 13(4), 101676-101685.DOI:10.1016/j.asej.2021.101676

[22] Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review. Energies, 16(10), 4025-4033.DOI:10.3390/en16104025

[23] Ray, P., Bhattacharjee, C., & Dhenuvakonda, K. R. (2022). Swarm intelligence-based energy management of electric vehicle charging station integrated with renewable energy sources. International Journal of Energy Research, 46(15), 21598-21618.DOI:10.1002/er.7601

[24] Mhlanga, D. (2023). Artificial intelligence and machine learning for energy consumption and production in emerging markets: a review. Energies, 16(2), 745-752.DOI:10.3390/en16020745

[25] Hou, C., Xu, N., & Liu, S. (2025). Design of online monitoring method for distribution IoT devices based on DBSCAN optimization algorithm. Informatica, 49(5).DOI:10.31449/inf.v49i5.6399

[26] Huang, Q., Xian, H., Mei, L., Cheng, X., & Li, N. (2025). Intelligent distribution network operation and anomaly detection based on information technology. Informatica, 49(9). DOI:10.31449/inf.v49i9.5584