Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090413
ISSN 2371-8838 Vol. 9 Num. 4

Research on Security Protection Strategies for Power Information Data Based on Big Data

Yong Fu

Beijing Remarkables United Technology Co., Ltd., Beijing, 100192, China

Keywords: Electric Power Information Security; Big Data; Privacy Protection; Internet Communication; Data Lifecycle Management

Abstract: This paper conducts a systematic study on multimedia communication security and big data privacy protection issues faced during the internet-based transformation of the power system. By analyzing new attack surfaces, vulnerability characteristics, and privacy protection needs in the power internet communication environment, a "proactive defense-privacy enhancement" dual-drive technology system is constructed. On the security protection level, a data security transmission scheme based on domestic cryptographic algorithms, a zero-trust dynamic access control mechanism, a big data situation awareness platform, and a collaborative emergency response system are proposed. On the privacy protection level, the innovative fusion of differential privacy and federated learning technologies is adopted to establish a privacy protection framework covering the entire data lifecycle. Empirical research shows that this system can reduce the incidence of security events by more than 75% and achieve controllable privacy while ensuring business real-time performance, effectively solving the balance problem between security protection and privacy protection in the context of power big data, and providing technical support and practical paths for building a new power system security ecosystem.

1. Introduction

With the accelerated advancement of the Energy Internet construction, the digital, networked, and intelligent transformation of the power system has become an inevitable trend. The deep integration of internet communication technology and power production not only improves system operation and maintenance efficiency and user service levels but also significantly expands the network attack surface. The popularization of new business scenarios such as video surveillance, remote inspection, and smart terminals has led the power system to gradually move from a traditional closed industrial control environment to an open and interconnected one, facing unprecedented information security and data privacy challenges[1,2].

Currently, the security protection of power information systems faces multiple pressures. On the one hand, new attack methods such as advanced persistent threats (APTs), data tampering, and steganography attacks are emerging endlessly, making it difficult for traditional security models based on perimeter defense to effectively cope. On the other hand, smart meters and IoT devices generate massive amounts of user electricity consumption behavior and equipment status data. This data is highly sensitive, and its collection, storage, and analysis processes pose significant privacy

disclosure risks. Existing studies mostly focus on traditional network security protection or single technology applications, lacking a systematic solution to security and privacy protection issues in the internet communication environment[3].

In this context, this study aims to construct an active security protection and privacy protection technology system adapted to the power internet communication scenario. By integrating advanced technologies such as domestic cryptographic algorithms, zero-trust architecture, big data analysis, and privacy computing, it aims to achieve full life cycle protection from data collection to destruction. The study focuses on solving key issues such as multimedia communication security, data privacy protection, and business performance balance, providing theoretical support and technical paths for building a safe and reliable new power system. This study not only has important practical value for ensuring the safe operation of critical power information infrastructure but also has strategic significance for promoting the digital transformation of the energy industry.

2. Risk Analysis

2.1 New Attack Surfaces and Vulnerabilities Introduced by Internetization

While the Internetization transformation of power systems enhances the level of system intelligence, it also introduces complex security threats. Traditional power industrial control systems employ physically isolated, closed network architectures, but the application of Internet communication technologies breaks this security boundary. The widespread use of the TCP/IP protocol exposes the system to remote attack risks from public networks, and the massive access of smart terminals greatly expands the attack surface. The deployment of business scenarios such as video surveillance, remote control, and intelligent inspection requires the system to open multiple network ports and services, providing opportunities for attackers.

The vulnerability of multimedia communication protocols is particularly prominent. Video stream transmission often uses plaintext or weak encryption methods, making it extremely vulnerable to eavesdropping or tampering. Attackers can inject malicious video frames through man-in-the-middle attacks, misleading operation and maintenance personnel to make incorrect judgments. A more insidious threat comes from steganography. Attackers use the data redundancy characteristics of multimedia files to hide malicious code or stolen data in normal business data for transmission, which can effectively circumvent the monitoring of traditional security detection equipment. When power-specific protocols (such as IEC 60870-5-104, DNP3) are run in an Internet environment, their inherent weaknesses, such as weak authentication mechanisms and lack of encryption protection, become more prominent, making them extremely vulnerable to replay attacks and protocol manipulation attacks.

2.2 Characteristics of Power Big Data and Its Privacy Leakage Risks

Power big data is characterized by its massive volume, diverse types, low value density, and fast processing speed. High-frequency electricity consumption data collected by smart meters can accurately reflect users' lifestyles, habits, and even the composition of electrical equipment, making it highly privacy-sensitive. Equipment status monitoring data is not only used for fault prediction but may also reveal commercial secrets such as production processes and equipment layouts. Dispatch instruction data is directly related to the real-time balance and safe operation of the power grid, requiring the highest level of confidentiality[4].

The risk of privacy leakage during data aggregation and analysis is particularly prominent. Attackers can use techniques such as differential attacks and correlation analysis to re-identify specific users or devices from seemingly anonymous data. Long-term collection of electricity

consumption behavior data can form a complete user profile, which can then be used to infer sensitive information such as the user's family structure and economic status. The risks are even more complex in data sharing scenarios, where cross-validation of information from multiple data sources may break through unilateral privacy protection measures. In addition, the construction of cloud-based big data platforms leads to centralized data storage, which, if attacked, will cause large-scale data leakage and have disastrous consequences.

2.3 Multiple Challenges of Compliance, Real-Time Performance, and Security

As critical information infrastructure, the power system needs to meet multiple constraints simultaneously. Laws and regulations such as the "Cybersecurity Law," the "Data Security Law," and the "Regulations on the Security Protection of Critical Information Infrastructure" impose strict compliance requirements on the collection, storage, processing, and cross-border transmission of power data. In particular, the data classification and grading protection system established by the "Data Security Law" requires power companies to take corresponding protection measures based on the importance and sensitivity of the data.

Power production and control operations have extremely high requirements for system real-time performance. The transmission delay of dispatch instructions and the response time of protection devices must be within the millisecond range. This real-time requirement makes it impossible to directly apply many traditional security technologies. For example, deep packet inspection may introduce unacceptable delays, and complex encryption algorithms may affect the timely delivery of control instructions[5]. The application of privacy protection technologies may also bring additional computational overhead, affecting the performance of business systems.

A delicate balance needs to be struck between security measures and business efficiency. Overprotection may lead to the destruction of data value, affecting real-time power grid control and user services; insufficient protection can easily lead to major security incidents and compliance risks. This balance needs to be dynamically adjusted according to specific business scenarios, data sensitivity, and threat levels, requiring the security protection system to have a high degree of flexibility and adaptability.

3. Security Protection System Construction

3.1 Data Security Transmission and Access Control

In the power internet communication environment, secure data transmission is a fundamental aspect of the protection system. On one hand, the commercial cryptographic algorithms recognized by the National Cryptography Administration can be used to construct a secure communication mechanism from end to end. For multimedia services like video surveillance and remote control, a frame-level selective encryption scheme based on the SM4 algorithm is adopted to encrypt key information such as motion vectors and DCT coefficients in P/B frames, ensuring security while keeping the computational overhead within the acceptable range of the services (with a delay increase of less than 50ms). On the other hand, the SM3 hash algorithm is used to generate message authentication codes (MAC) to verify the integrity of the data transmission process and defend against replay attacks and data tampering. If further security enhancement is needed, a dynamic key update mechanism based on a key derivation function (KDF) can be designed. This mechanism updates the encryption key periodically according to session time or data volume thresholds to prevent security risks caused by the long-term use of keys[6].

The implementation of a zero-trust architecture fundamentally changes the traditional boundary protection model. Power enterprises can establish an identity-centered multi-dimensional trust

assessment model. By using device fingerprint identification technology, they can collect the hardware features, software features and behavioral features of terminal devices, generating a unique digital fingerprint for each device. This model can continuously implement the identity authentication mechanism. During the session, it can analyze the user's operation behavior, device operating status and other contextual information in real time, and dynamically adjust the trust score. When an abnormal access pattern is detected (such as unconventional login times or attempts to elevate privileges), the system automatically triggers a tiered response mechanism, including requiring re-authentication, restricting access scope, or immediately terminating the session. Implementation data from a provincial power company shows that this solution reduced the success rate of unauthorized access attempts by 92% and shortened the average detection time of abnormal access behavior to 3.2 seconds.

3.2 Security Posture Awareness and Intrusion Detection

A network security posture awareness platform based on big data technology is constructed to realize the collection, correlation analysis, and visual display of multi-source security data. The platform adopts a distributed architecture, which can process data sources from network traffic probes, security equipment logs, host audit records, user behavior data, and other dimensions[7]. Network security data is collected in real-time through data acquisition tools such as Flume and Kafka, and stored in the HDFS distributed file system to provide data support for subsequent analysis.

In the data analysis layer, machine learning algorithms are used to build anomaly detection models. An improved isolation forest algorithm is used to detect unknown attack types. This algorithm effectively identifies abnormal data points by randomly selecting features and dividing the data space. Combined with a one-dimensional convolutional neural network (1D-CNN), network traffic data is subjected to feature extraction and classification to accurately identify network attack behaviors such as DDoS attacks and port scans. For the detection of APT attacks, a long short-term memory network (LSTM) is used to analyze time-series data and capture long-term dependencies in the attack process. In the deployment practice of a local power supply company, the platform achieved accurate detection of more than 95% of known attacks, and the detection rate of unknown attacks reached 87%, with the average detection time shortened from hours to minutes.

The situation assessment module integrates multi-source information to generate a global network security posture score. The weights of each indicator are determined by the analytic hierarchy process (AHP), and the fuzzy comprehensive evaluation method is combined to process uncertain information, and output intuitive situation maps and early warning information. The visualization interface uses front-end frameworks such as Echarts to realize attack path tracing, threat source location, and impact range assessment, providing decision support for security management personnel.

3.3 Deep Security Detection and Emergency Response

Covert channel threats in power multimedia communication require defense using deep content detection technology. Power enterprises can construct a multimedia steganalysis model based on deep learning, employing an architecture combining Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM), while simultaneously extracting spatial features and time-series features. This model is trained on a sample library containing various steganography tools and can effectively detect common steganography techniques such as Least Significant Bit (LSB) steganography and spread spectrum steganography.

Establish a collaborative emergency response system to achieve closed-loop management from

threat detection to disposal. When the security monitoring system detects an anomaly, this system automatically initiates the emergency response process: Firstly, it isolates the affected terminals from the network to prevent lateral movement; simultaneously, it initiates the evidence collection procedure to preserve the evidence data related to the attack; then, based on the threat level, it initiates the corresponding emergency response plan, including switching to the backup channel, enabling the emergency control mode, etc. To improve response efficiency, a traffic scheduling mechanism based on Software Defined Network (SDN) can be designed for this system to achieve millisecond-level attack traffic isolation and business traffic re-routing.

The emergency response platform integrates a case library and knowledge graph function, storing historical security incident handling experiences and solutions. This emergency response platform uses natural language processing technology to analyze security incident reports, automatically extracting key information and generating recommended handling plans. The exercise module supports red-blue team exercises, regularly testing the effectiveness of emergency response processes and continuously optimizing response strategies[8]. In a practical application in a certain regional power grid, this system reduced emergency response time by 68%, significantly improving the overall security of the system.

3.4 Performance Evaluation and Optimization Mechanism

Establish a comprehensive performance evaluation mechanism for the security protection system, quantifying the effectiveness of security protection through multi-dimensional indicators. On one hand, key performance indicators (KPI) such as attack detection rate, false alarm rate, response time, and business impact degree are used to regularly assess the effectiveness of each protection module. On the other hand, penetration testing and red-blue confrontation exercises, real attack scenarios are simulated to test the practical combat capability of the protection system.

Establish a continuous optimization mechanism based on evaluation results. Firstly, based on historical attack data and protection effectiveness, the reinforcement learning algorithm is employed to dynamically adjust the security policy parameters. Then, a security knowledge base is established. By leveraging the accumulated protection experience and best practices, effective features are extracted through machine learning algorithms to optimize the detection model. Finally, an adaptive security policy engine is designed to automatically adjust the protection intensity and focus in response to changes in the network environment and the evolution of threat situations.

The actual application of this protection system in a provincial power company demonstrates that multi-layered and proactive security protection measures can effectively address various security threats in the power internet communication environment. While ensuring the real-time performance of services, the system reduces the incidence of security events by more than 75%, significantly improving the overall security protection level of the power information system.

4. Technical Approaches

4.1 Privacy-Enhancing Technologies for Data Acquisition and Transmission

At the initial stage of the power big data lifecycle, privacy protection must start from the source of data acquisition. Based on differential privacy technology, precisely calculated Laplacian noise is injected during the smart meter data acquisition process to balance privacy protection and data utility. An adaptive mechanism is used to determine the noise magnitude, dynamically adjusting the ε value according to data sensitivity, query frequency, and application scenarios, ensuring sufficient privacy protection while maintaining the accuracy of data analysis results. Experimental data show that when the ε value is controlled within the range of 0.1-1.0, the relative error of aggregate query

results can be maintained within 15%, meeting the needs of most power business analysis.

For cross-domain data collaboration scenarios, federated learning provides an ideal solution. A federated learning framework based on hierarchical aggregation is constructed, where local municipal companies, acting as clients, train models locally and upload only the encrypted model parameters to the provincial center for aggregation. A homomorphic encryption-based parameter protection mechanism is adopted to ensure the confidentiality of model parameters during the aggregation process. To address the challenge of Non-Independent and Identically Distributed (Non-IID) data, an adaptive weighted aggregation algorithm is developed to dynamically adjust aggregation weights based on local data quality and distribution characteristics. In a cross-regional equipment fault prediction project, this framework improved the model prediction accuracy to 92.5%, an increase of approximately 18 percentage points compared to training in a single region, while ensuring that each entity's data remained within its domain.

4.2 Privacy Protection Schemes for Data Storage and Processing Stages

The data storage stage employs a multi-layered privacy protection strategy. Generalization techniques based on k-anonymity are used to process direct identifiers. L-diversity ensures the diversity of sensitive attribute values, and t-closeness prevents similarity attacks. Aiming at the temporal characteristics of power data, a sliding window mechanism is designed to achieve privacy protection while maintaining the temporal correlation of data. Attribute-Based Encryption (ABE) technology is used for data access control to implement fine-grained access authorization based on data attributes and user identity. The system dynamically generates access policies according to the user's job responsibilities, task requirements, and data sensitivity level, ensuring the implementation of the principle of least privilege.

In the data processing stage, Secure Multi-Party Computation (MPC) and Homomorphic Encryption (HE) technologies are introduced. Secure multi-party computation allows multiple participants to perform collaborative calculations without revealing their respective input data, which is particularly suitable for cross-institutional power data analysis scenarios. Homomorphic encryption technology supports direct calculation in the ciphertext state, realizing the ideal state of "data available but invisible." Although the computational overhead of fully homomorphic encryption is large, practical progress has been made in specific application scenarios such as power load forecasting and equipment status assessment by adopting partially homomorphic encryption schemes and algorithm optimization. Test data shows that the optimized homomorphic encryption scheme improves the calculation efficiency by about 40% compared with the traditional method, which lays the foundation for actual deployment and application.

4.3 Security Management in the Data Sharing and Destruction Stages

The data sharing stage establishes a trusted sharing mechanism based on blockchain technology. Smart contracts are used to automate the execution of data sharing rules, ensuring that the data usage process is auditable and traceable. A data usage right tokenization scheme is designed to convert data access permissions into digital tokens, enabling fine-grained access control and time restrictions. Digital watermarking technology is used during the sharing process to embed invisible identification information in the data, facilitating subsequent tracing and accountability.

The data destruction stage implements a verifiable data clearing mechanism. A combination of multiple overwrites and cryptographic erasure is used to ensure that data is completely deleted and unrecoverable. The entire process of data destruction operations is recorded based on blockchain technology, including key information such as destruction time, executors, and operation methods, to achieve auditability of the destruction process. A data lifecycle monitoring system is established

to completely record the entire lifecycle activities of data, including creation, modification, access, and destruction, to ensure the implementation of compliance requirements.

4.4 Privacy Protection Performance Evaluation and Optimization

A quantitative evaluation system for privacy protection performance is established, and multi-dimensional indicators are used to comprehensively evaluate the privacy protection effect. Information entropy is used to measure data uncertainty, k-anonymity is used to assess the risk of identity disclosure, and t-closeness is used to analyze the risk of attribute disclosure. A privacy-utility trade-off curve is designed to intuitively show the changing trend of data availability under different protection intensities, providing a scientific basis for strategy selection.

A dynamic optimization mechanism is established based on the evaluation results. Reinforcement learning algorithms are used to automatically adjust privacy protection parameters based on changes in data usage patterns and threat landscapes. A Privacy Impact Assessment (PIA) tool is established to quantitatively assess privacy risks in the new system design stage, and protection measures are front-loaded through the Privacy by Design concept. The analytical utility of anonymized data is regularly tested and optimized to ensure that the analytical value of the data is maximized while meeting privacy protection requirements.

The practice of this privacy protection technology system in a provincial power company shows that through full lifecycle privacy protection measures, data value can be fully explored while ensuring data security. The system reduces the risk of user re-identification to below 0.3% while maintaining the availability of data analysis results. The accuracy of load forecasting remains above 93%, providing strong privacy protection support for the digital transformation of the power industry.

5. Conclusion

This study systematically investigates the information security and privacy protection issues facing power systems in the internet communication environment, and constructs a "proactive defense-privacy enhancement" dual-drive technical system. In terms of security protection, it proposes a data security transmission scheme based on domestic cryptographic algorithms, a dynamic access control mechanism under zero-trust architecture, a network security situation awareness platform based on big data analysis, and a collaborative emergency response system. In terms of privacy protection, it innovatively applies technologies such as differential privacy and federated learning to power big data scenarios, and establishes a privacy protection framework covering the entire data lifecycle. Through empirical application in a provincial power company, this system successfully reduced the incidence of security events by more than 75%, and maintained the efficient operation of business systems while ensuring data privacy.

References

[1] Xu Zhengli. Research on the Application of Big Data Technology in Power System Information Security Protection [J]. Construction Science and Technology, 2024, (S1): 9-11 + 15.

[2] Zhao Yu. Research on the Application of New Information Security Analysis Technology for Power Big Data in the New Era [J]. Modern Industrial Economy and Informationization, 2023, 13(11): 110-112.

[3] Huang Mengqi. Research on New Network Information Security Protection Strategies for the New Power System Based on Big Data [J]. Information & Computer, 2023, 35(01): 222-225.

[4] Wang Lirong, Chi Zhan, Wu Yijia, et al. Effective Application of Big Data in Power Information Security [J]. Electric Power Equipment Management, 2021, (03): 28-29.

[5] Li Lihua. Implementation Countermeasures of Big Data in Power Information Security [J]. Science and Technology Trends, 2021, (02): 195-196.

[6] Ye Yong. Analysis of the Application of Big Data in Power Information Security [J]. Digital Technology & Application, 2020, 38(07): 52-53.
[7] Fu Yanzhe. Application of Big Data Technology in Power System Information Security Protection [J]. Electronics World, 2020, (11): 206-207.
[8] Li Qiang, Huang Lin. Research on Power System Information Security Protection in the Context of Big Data [J]. Cybersecurity & Informatization, 2024, (07): 17-19.