electronics

MDPI

Article

Hierarchical-Based Dynamic Scenario-Adaptive Risk Assessment for Power Data Lifecycle

Yubo Song [1,2,*], Shuai Jiang [1,2], Qiuhong Shan [1,2], Yixin Yang [1,2], Yue Yu [1,2], Wen Shen [3] and Qian Guo [3]

[1] School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China; 220224992@seu.edu.cn (S.J.); qiuhshan@seu.edu.cn (Q.S.); seu_yangyx@seu.edu.cn (Y.Y.); 220224906@seu.edu.cn (Y.Y.)
[2] Purple Mountain Laboratories, Nanjing 211189, China
[3] Power Grid Smart Grid Research Institute Co., Ltd., Nanjing 210003, China; shenwen@geiri.sgcc.com.cn (W.S.); guoqian@geiri.sgcc.com.cn (Q.G.)
* Correspondence: songyubo@seu.edu.cn

Abstract: In an era marked by rapid advancements in information technology, the task of risk assessment for data security within the complex infrastructure of the power grid has become increasingly vital. This paper introduces a novel methodology for dynamic, scenario-adaptive risk assessment, specifically designed to address the entire lifecycle of power data. Integrating hierarchical analysis with fuzzy comprehensive evaluation, our approach provides a flexible and robust framework for assessing and managing risks in various scenarios. This method enables the generation of adaptive weight matrices and precise risk level determinations, ensuring a detailed and responsive analysis of data security at each lifecycle stage. In our study, we applied predictive analytics and anomaly detection to conduct a thorough examination of diverse data scenarios within the power grid, aiming to proactively identify and mitigate potential security threats. The results of this research demonstrate a significant enhancement in the effectiveness of risk detection and management, leading to improved data protection and operational efficiency. This study contributes a scalable, adaptable model for data security risk assessment, meeting the challenges of big data and complex information systems in the power sector.

Keywords: data security protection; full data lifecycle; scenario-based requirements analysis; risk assessment; analytic hierarchy process

check for updates

Citation: Song, Y.; Jiang, S.; Shan, Q.; Yang, Y.; Yu, Y.; Shen, W.; Guo, Q. Hierarchical-Based Dynamic Scenario-Adaptive Risk Assessment for Power Data Lifecycle. *Electronics* **2024**, *13*, 631. https://doi.org/10.3390/electronics13030631

Academic Editor: Andrei Kelarev

Received: 9 January 2024
Revised: 29 January 2024
Accepted: 30 January 2024
Published: 2 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the dynamic landscape of information technology and the data-driven ethos of modern business, the role of commercial data in strategic decision-making has become indispensable. The State Grid Corporation, a pivotal player in China's energy sector, has amassed an extensive array of valuable data through diverse operational activities. These data, which include insights into customer power purchasing behaviours, usage patterns, and comprehensive user profiles, hold immense potential for optimising marketing strategies and improving service quality.

The challenge for the power grid, however, lies in harnessing the full value of these data while safeguarding their security during the data openness process. The transition towards greater data openness exposes the power grid to a plethora of new security risks and threats [1]. It is imperative, therefore, to develop a robust and scenario-specific approach to data security that covers the entire lifecycle of the data.

To this end, various enterprises and organisations have been proactive in establishing effective data security frameworks. Microsoft, for instance, with its "DGPC Framework", emphasises the critical aspects of privacy, confidentiality, and compliance in data security [2]. The "Data Security Governance White Paper" by China's Data Security Governance Committee, published in 2018, lays down foundational principles and a framework

for data security, underscoring the importance of hierarchical classification and scenario-based security measures [3]. Huawei's "Cloud Data Security White Paper" from 2021 further explores the future trajectory of data security, highlighting the need for strengthening basic industry capabilities and the enforcement of legal frameworks for data protection [4].

In parallel, the advancements in artificial intelligence (AI) have revolutionised various sectors, with particularly impactful strides in machine learning and its application to data security. This intersection of AI and security has birthed sophisticated methodologies capable of pre-empting and neutralising threats in increasingly complex data environments. Research in this domain has extensively utilised deep learning algorithms, a subset of machine learning, for their unparalleled proficiency in pattern recognition and anomaly detection. These algorithms have shown remarkable effectiveness in identifying subtle and complex anomalous patterns within network traffic, significantly enhancing threat detection capabilities [5,6]. By analysing vast datasets and learning from them, deep learning systems can detect deviations from normal behaviour, flagging potential security breaches with higher accuracy and speed than traditional methods.

Moreover, other studies in the field have shifted focus towards the intelligent generalisation of data, a critical aspect in maintaining privacy and confidentiality. Techniques such as data anonymisation, which involves the alteration of personal data so that individuals cannot be readily identified, are at the forefront of these efforts [7]. These techniques employ sophisticated algorithms that balance the need for data utility with privacy requirements. For instance, machine learning algorithms can be trained to selectively blur or modify specific data elements, ensuring that the data remain useful for analysis while significantly reducing the risk of exposing sensitive information. This approach is particularly vital in scenarios where data are shared across different departments within an organisation or with external partners. The lifecycle of open data, stretching from its collection to eventual destruction, is fraught with distinct security risks at each stage [8]. In the collection phase, the power grid faces the challenges of ensuring data authenticity and protecting against malicious data injections. During transmission, the focus shifts to securing data against theft or tampering through robust encryption and authentication protocols [9]. The storage phase requires rigorous access controls and encryption techniques to prevent unauthorised data breaches. In the data processing and sharing phase, desensitisation and anonymisation become critical for maintaining user privacy. Finally, the destruction phase mandates secure deletion practices to guarantee the irreversible elimination of data [10].

In previous research in the field of data security, scholars have focused on specific aspects of data so as to improve the security of data in these aspects. The use of cryptography, especially in data transmission, provides a strong guarantee to ensure the confidentiality of data in network dissemination [11]. The design and implementation of access control mechanisms enable data to be effectively protected during the storage and processing stages [12]. In addition, encryption methods in data storage also mitigate the security risks in the storage process to a certain extent. However, although these studies have made significant progress in data security in specific segments, the comprehensive protection of the data lifecycle still faces a series of challenges. For example, in the process of data delivery, especially when crossing different platforms and organisational boundaries, data may be affected by unknown threats, resulting in a non-negligible risk of information leakage. At the same time, with the widespread application of cloud computing, edge computing, and other emerging technologies, the security requirements for the data processing stage are becoming increasingly complex.

This paper aims to fill this research gap by proposing a comprehensive full data lifecycle protection framework to cope with increasingly complex and diverse security threats. By deeply analysing the process of data collection, storage, processing, and sharing, we aim to develop a comprehensive set of security policies and techniques to ensure the security and privacy of data throughout its lifecycle. The approach proposed in this paper not only bridges the gap between existing research, but also focuses on the security needs

of future data applications and provides new ideas for solving the data security problem on a global scale.

In addressing these multifaceted challenges, this paper seeks to achieve the following:

1. Thoroughly investigate the spectrum of security needs throughout the entire lifecycle of data within the power grid, especially focusing on scenario-specific disclosures and their inherent security challenges.
2. Develop a comprehensive framework for data lifecycle security protection, meticulously analysing security requirements at each stage of data openness and usage.
3. Introduce an innovative, holistic approach to data security risk assessment, blending hierarchical analysis with fuzzy comprehensive evaluation. This methodology is designed to dynamically adapt to different scenarios, thereby enhancing the precision and effectiveness of security level assessments.

2. Data Lifecycle Protection Requirements

Data openness is a process that involves multiple phases and activities, from data collection to data destruction, each with different objectives, activities, and security needs [13]. In this section, this paper provides an overview of the typical lifecycle stages of data openness and the scenarios designed to better understand the full lifecycle security requirements of scenario-based data openness.

The full data lifecycle protection scenarios and requirements are shown in Figure 1.

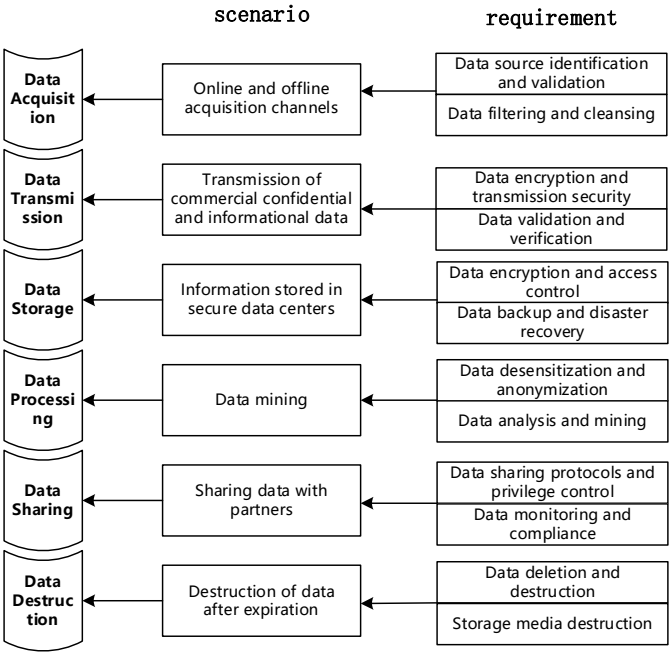


Figure 1. Data lifecycle protection scenarios and requirements.

2.1. Data Acquisition Phase

The data collection phase is the beginning of the open data lifecycle in which data are collected and captured from different sources. In smart grid rollout scenarios, data collection is the collection of user electricity consumption data and related information through devices such as smart meters and sensors. These data contain important information about the user's electricity consumption, energy demand, and other important information. In this phase, key activities and processes include the following:

Data source identification and validation: identify reliable data sources and verify their legitimacy and credibility.

Data filtering and cleaning: cleaning and pre-processing of collected data to remove invalid or redundant data and ensure data quality.

It is important to protect the integrity and accuracy of data during the data collection phase to avoid malicious data injection and quality issues.

2.2. Data Transmission Phase

The data transmission phase involves the process of transferring the collected data from the data source to the target system or platform. In the power grid business scenario, data transmission involves transmitting the collected electricity consumption data to the smart grid system or data centre through encrypted protocols and secure transmission channels to ensure the confidentiality and integrity of the data during transmission [14]. Key activities and processes in this phase include the following:

Data encryption and transmission security: using encryption protocols and secure transmission channels to ensure the confidentiality and integrity of data during transmission.

Data validation and verification: ensuring the integrity and correctness of the transmitted data in the target system and preventing data tampering or loss.

Secure transmission of data and integrity verification are key measures to protect data from unauthorised access and tampering during the data transmission phase.

2.3. Data Storage Phase

The data storage phase involves storing the data in the storage device or database of the target system or platform. The power grid stores electricity consumption data and related information collected from smart meters, sensors, and other devices in a secure data centre or cloud platform to ensure the confidentiality and reliability of the data. The following key activities and processes need to be considered at this stage:

Data encryption and access control: encrypt and protect the stored data and use appropriate access control mechanisms to restrict access to the data.

Data Backup and Disaster Tolerance: develop a data backup strategy to ensure the reliability and recoverability of the data, as well as the ability to recover the data in the event of a disaster.

Security measures at the data storage stage are designed to prevent unauthorised access, data leakage, and data loss.

2.4. Data Processing Phase

The data processing stage involves analysing, handling, and processing the stored data to extract valuable information and insights. In power demand forecasting and optimisation scenarios, the power grid may use data mining and machine learning techniques to process historical power consumption data to forecast future power demand and formulate optimisation plans accordingly. In smart grid rollout scenarios, the power grid may use real-time data processing techniques to monitor the operating status of the power system and make quick responses and adjustments. Key activities and processes include the following:

Data desensitisation and anonymisation: desensitising sensitive information to protect user privacy.

Data analysis and mining: applying data analysis techniques and algorithms to extract knowledge and insights from data.

During the data processing phase, privacy protection and data anonymisation are key elements to ensure data security and compliance with regulatory requirements.

2.5. Data Sharing Phase

The data sharing phase involves sharing the processed data with partners, stakeholders, or the public. In an electricity supply contracting scenario, the national grid may share electricity consumption data with companies to provide customised electricity consumption plans and energy saving advice. Key activities and processes include the following:

Data sharing agreements and permission controls: developing appropriate data sharing agreements and ensuring that partners or stakeholders only have access to the data they need.

Data monitoring and compliance: the data sharing process is monitored and audited to ensure data compliance and security.

Security measures at the data sharing stage are designed to prevent unauthorised access and misuse of data and ensure data compliance.

2.6. Data Destruction Phase

The data destruction phase involves the secure deletion of data that are no longer needed and the secure destruction of storage media [15]. At the end of the data development lifecycle, when the data are no longer needed, the power grid needs to perform data destruction to ensure that sensitive information is not misused or leaked. Under compliance requirements, StateNet may establish a data retention period and securely destroy or permanently delete the data after they have expired. Key activities and processes include the following:

Data deletion and destruction: secure data deletion methods are adopted to ensure that data is not recoverable.

Storage media destruction: secure destruction of storage media to prevent data leakage.

Security measures during the data destruction phase are designed to ensure that data cannot be accessed or misused in a malicious manner.

3. Analysis of Data Lifecycle Protection Processes

In the electric power industry, data security is an important issue, because the data involve the privacy of customers, the operating status of power equipment, and the transaction information of the power market. Through the study of the needs and scenarios of the whole lifecycle of data openness, the security risks in the process of data openness to the public are challenging, and in order to cope with this series of problems, it is necessary to protect the overall process, and that products and services businesses carry out security reviews and security protections to protect the security and privacy of data openness to the public [16].

3.1. Overall Process of Data Lifecycle Protection

In order to protect the security and integrity of data, a series of protection processes need to be carried out, and the following will introduce the specific contents and steps of the electric power data security protection process [17].

As depicted in Figure 2, the protective process initiates by meticulously planning and designing data, followed by classifying and grading them in accordance with relevant state regulations. Subsequently, the data classified based on their grading are collected, encompassing both customer authorisation information and power equipment data acquired through collection devices. To safeguard the collected power equipment data against theft or tampering during transmission, encryption technology is employed.

Concurrently, a credit assessment is conducted on customers who access the service, and a dedicated customer database is established to house customer information along with the respective data classifications. Advancing to the data processing stage, the procedure encompasses compliance checks, the identification of sensitive information, data desensitisation, and the application of digital watermarks, ensuring standardised processing and secure protection of sensitive and private data.

In the event of a security incident during data processing, the system transitions into the security incident processing stage, implementing tailored countermeasures based on the specific incident. Special attention is given to securing data exchanges according to specific categories. Any abnormal security events during data exchange trigger the data variance tracking and processing stage, where corresponding countermeasures are enacted

to address specific abnormal behaviours. Once anomalies are resolved, an appropriate approach is chosen for data application.

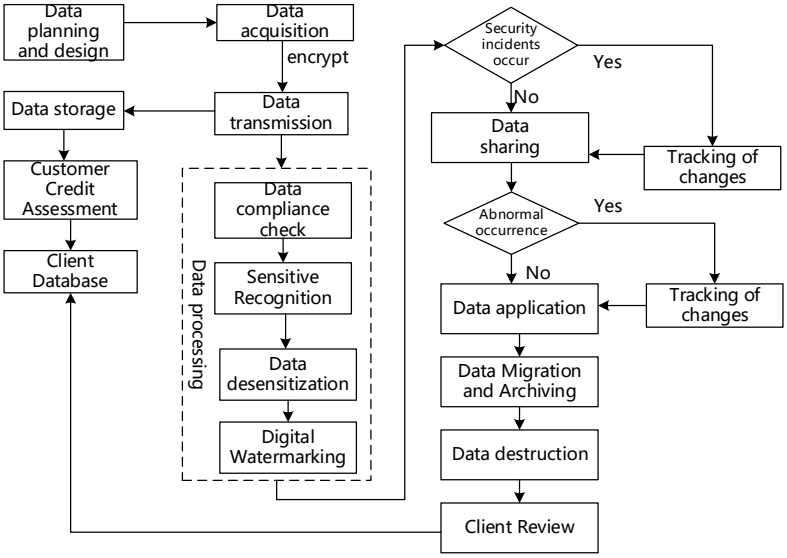


Figure 2. Data lifecycle protection flowchart.

Upon completion of the data application, adherence to designated protocols guides the decisions on data migration, archiving, and destruction. Subsequently, customer follow-up visits are conducted to gather insights and address any issues arising from the product application process, thereby facilitating updates to the customer database.

3.2. Data Services Workflow

Securing data is a crucial measure for upholding the quality and value of data, serving as a fundamental requirement to protect the legitimate rights and interests of both data owners and users. To achieve robust data security, specific processes must be meticulously followed to ensure effective management and protection throughout the entire data lifecycle.

As illustrated in Figure 3, the product planning stage marks the inception of determining data requirements and solutions. During this phase, the demand initiator engages in market research, formulates a comprehensive business plan/technical specification, and standardises all facets of the data. Multiple departments within the power grid scrutinise these documents to ascertain alignment with the organisation's data security policies and standards. Upon successful review, a contract is signed between the demand initiator and relevant departments, clearly outlining the rights, obligations, and measures for data security protection.

The product construction stage signifies the realisation of identified data requirements and solutions. Here, the demand contractor initiates the account application process and implements data traceability, meticulously recording a variety of information related to the data. Concurrently, the demand initiator seeks approval for data usage, applies for data access from relevant departments, and provides detailed information about the data. Power grid departments evaluate the data usage application, conducting a thorough review and authorisation in adherence with data security policies and standards. Only after successful review can the demand initiator access the provided data and proceed with product construction as outlined in the contractual agreement. Throughout the product construction, security testing and operation and maintenance audits are conducted by the department to verify and uphold product security. Prior to product delivery, a confidentiality agreement is signed between the demand initiator and the department, outlining the confidentiality obligations and responsibilities of both parties regarding the data and the product.

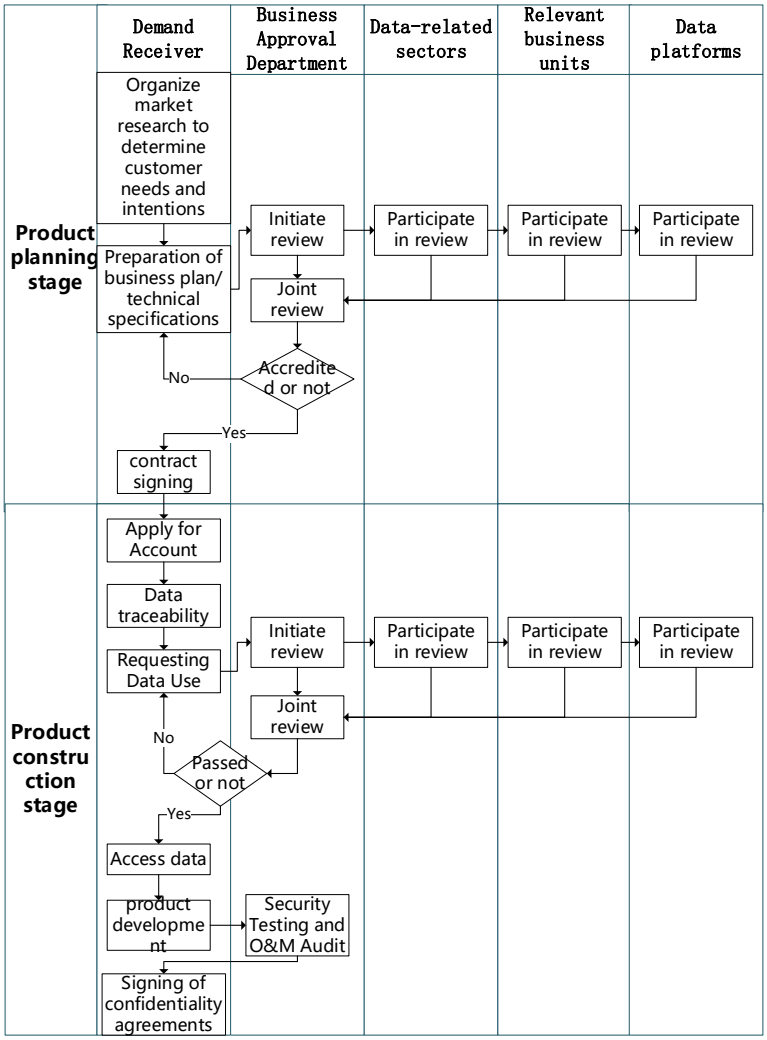


Figure 3. Open data services workflow diagram.

4. Risk Detection Programme for Full Lifecycle Protection of Data

In order to more accurately assess the security level of the whole lifecycle of data, this topic proposes a security risk assessment method based on the hierarchical analysis method and the fuzzy comprehensive evaluation method [18]. The hierarchical analysis method is used to dynamically generate the weight matrix, and then the fuzzy comprehensive evaluation method is used to determine the risk level results.

4.1. Hierarchical Modelling Constructs for Full Lifecycle Data Security Risks

The Analytic Hierarchy Process (AHP) is an assessment method that combines quantitative and qualitative methods. Its premise is to provide a systematic, hierarchical, structured analysis method and thinking mode, which is widely used in the analysis of complex problems that are not easy to quantify as variables [19]. The main algorithmic idea is to continuously decompose and refine a complex problem to form a hierarchical structure. The main steps are the following:

(1) Establishing a hierarchical structural model to stratify and refine the various factors of a complex problem

Generally speaking, the hierarchical analysis method is divided into three layers: the objective layer, the guideline layer, and the programme layer. The upper and lower elements are the guidelines needed to accomplish the goal, and the bottom programme layer is the programme needed for the goal of the upper layer, which can also be understood as the lower layer serving the upper layer [20].

Combined with the risk calculation model proposed in this paper, the factor set of the risk value of the target layer is divided into two layers, and the specific model is shown in Figure 4.

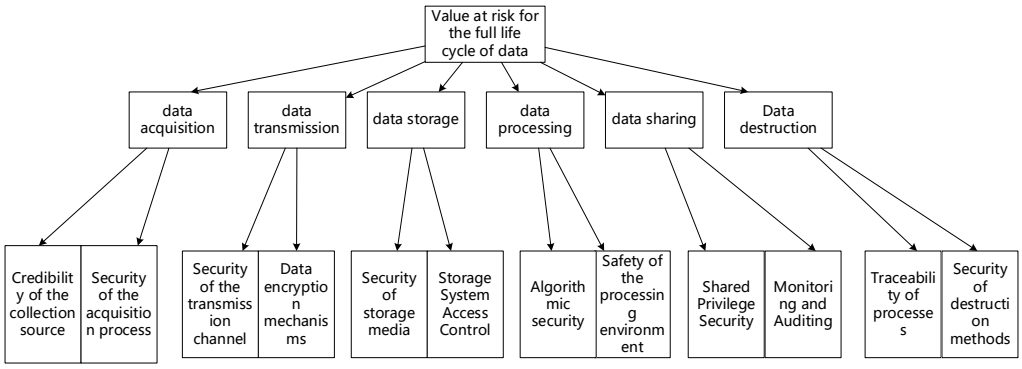


Figure 4. Structural modelling of value-at-risk hierarchical factors.

The first tier of factors is

$$\mathrm{U} = \{U_1, U_2, \ldots, U_6\} = \{Data\ acquisition, Data\ transfer, Data\ storage, Data\ processing,\ Data\ sharing, Data\ destruction\}$$

The second tier of factors is U = {U_{i1}, U_{i2}}, where U_{i1} and U_{i2} are the requirements corresponding to each phase.

(2) Constructing a comparative judgement matrix

Each layer is associated with the layer above it, and the matrix is constructed using the relative importance of the elements in the layer before they are associated with the layer above. Suppose that the bottom level of the hierarchical model is level B, and the level above level B is level C. The judgement matrix of the factors $B_1, B_2, \ldots, B_n$ in level B with upper layer C_k is

$$D = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

where a_{ij} indicates the importance of B_i in relation to B_j in relation to C_k. The value of a_{ij} in the hierarchical analysis has been determined by experts on the basis of information and mathematical algorithms.

Firstly, a decision-making expert group composed of experts in relevant fields assigns values to each risk factor of the security risk of distribution in the IoT. The judgement matrix is constructed based on the given judgement criteria starting from level 2 of the hierarchy diagram according to the results of the two-by-two comparison of the experts' scores on the importance of the indicators at each level [21].

$$A = \begin{bmatrix} 1 & \frac{\omega_1}{\omega_2} & \cdots & \frac{\omega_1}{\omega_n} \\ \frac{\omega_2}{\omega_1} & 1 & \cdots & \frac{\omega_2}{\omega_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\omega_n}{\omega_1} & \frac{\omega_n}{\omega_2} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{2}$$

The judgement matrix A is an $n \times n$ square matrix with diagonal value of 1. It satisfies $a_{ij} \times a_{ij} = 1, i \neq j,\ \ i, j = 1, 2, 3, \ldots, n, a_{ij} > 0$, and a_{ij} is the ratio of the relative weights of the two factors i and j.

For the n-order judgement matrix, only n(n-1)/2 elements need to give the value; there will be estimation error in the judgement of the factors of complex things, so it is necessary to carry out a consistency test in order to avoid a large error.

(3) Mathematically derive the relative weights of the factors to determine the impact of the bottom element on the top and the severity of the risk present.

First, calculate the product M_i of the elements of each row of the matrix A

$$M_i = \prod_{k=1}^{n} a_{kj} \tag{3}$$

where $k = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$.

Square M_i to obtain $W_i' = n\sqrt{M_i}$, where $i = 1, 2, \ldots, n$. Normalise the vector W_i':$W_i = \frac{W_i'}{\sum_{j=1}^{n} W_j'}$, where $i = 1, 2, \ldots, n$. Obtain the eigenvector $W = (w_1, w_2, w_3, \ldots, w_n)^T$ corresponding to the largest eigenvalue λ_{max}, where $W_i = \frac{W_i'}{\sum_{i=1}^{n} W_i'}$ is the relative weight of the element being compared with respect to the criterion layer.

Calculate the maximum eigenvalue λ_{max} as $AW = \lambda_{max}W$ so we have $\lambda_{max} = \sum_{k=1}^{n} \frac{(AW)_k}{nW_k}$ where $k = 1, 2, \ldots, n$.

Calculation of the consistency indicator CI:

$$CI = \frac{\lambda_m - n}{n - 1} \tag{4}$$

(4) Testing consistency indicators

Calculation of the consistency ratio:

$$CR = \frac{CI}{RI} \tag{5}$$

In Equation (4), n represents the order of the judgement matrix A. The larger the value of CI is, the worse the consistency of the judgement matrix A is. At this time, it is necessary to adjust the value of the judgement matrix elements. Equation (5) defines the consistency ratio CR, and RI represents the average random consistency index, whose value has been calculated and analysed by the proposer of the hierarchical analysis method to provide reference for users.

4.2. Weight-Based Fuzzy Comprehensive Evaluation Analysis

After calculating the weights of the factors involved in the risk using the hierarchical analysis method, the fuzzy comprehensive evaluation method can be used to calculate the risk value [22]. The fuzzy evaluation method is a method of making predictions and evaluations based on the theory of fuzzy mathematics, and its algorithms are similar to people's thought patterns, describing objects in terms of the degree.

The principles of the fuzzy evaluation method are as follows:

(1) Establish a rubric set

This is a subset of the possible levels of safety for each factor. This is generally categorised into five levels, i.e., $V = \{V_1, V_2, V_3, V_4, V_5\}$, which denote {very safe, safe, relatively safe, relatively dangerous, and dangerous}, respectively [23]. A number of experts (typically 20 to 40) are then asked to evaluate and assign values to each factor separately. Suppose, $V = \{V_1, V_2, V_3, V_4, V_5\} = \{10, 5, 5, 5, 10, 0\}$, which means that, for a certain factor, 10 out of a total of 30 experts consider it to be "very safe", 5 out of 30 consider it to be "safe", 5 think it is "relatively safe", 10 think it is "relatively dangerous", and 0 think it is "dangerous". Then, we can obtain that the degree of affiliation of this factor in the level of "very safe" is 1/3, in the level of "safe" is 1/6, in the level of "relatively safe" is 1/6, in the level of "relatively dangerous" is 1/3, and 0 for "dangerous".

When experts make assignments, they only need to assign values to the elements represented by the leaf nodes (nodes with no children) in the hierarchical model, and then calculate the values of the nodes in the previous level [24]. This is repeated until the value at the top level is obtained, i.e., the final evaluation value.

(2) Fuzzy comprehensive evaluation

The affiliation matrix R_i for each sublevel factor U_{ij} of U_i can be obtained by first evaluating each factor with no subfactors (analogous to a leaf node in a tree) starting from the bottom level, and finding the affiliation vector $r_{ij} = (r_{ij1}, r_{ij2}, \ldots, r_{ij5})$ of the factor U_{ij} with respect to all the ratings in its set of ratings.

$$R_i = \begin{pmatrix} r_{i1} \\ r_{i2} \\ \vdots \\ r_{im} \end{pmatrix} \tag{6}$$

According to the fuzzy evaluation model formula

$$B = W \circ R \tag{7}$$

where W is the weights of the factors, and "$\circ$" denotes the fuzzy operator, matrix multiplication is used here. R denotes the affiliation matrix.

Then, starting from the bottom layer, the fuzzy evaluation B_i of the layer is obtained according to the weights of the factors W_i and the corresponding affiliation matrix R_i. After normalising B_i, the fuzzy evaluation of the layer above it is calculated using the obtained value until the fuzzy evaluation B of the top layer is obtained [25].

(3) Calculate the value at risk

The formula for calculating the value at risk in the algorithm is the following:

$$F = \sum_{i=1}^{5} b_i' \times v_i \tag{8}$$

where F is the value at risk; v_i denotes the expert's comment; and b_i' denotes the final evaluation value obtained in the fuzzy evaluation.

This project is dedicated to proposing an innovative approach to the security risk assessment of the full lifecycle of data, which combines the hierarchical analysis method and the fuzzy comprehensive evaluation method. First, the weight matrix is dynamically generated by the hierarchical analysis method to more accurately assess the importance and influencing factors of each stage in the full data lifecycle. Subsequently, the fuzzy comprehensive evaluation method is used to comprehensively assess the risks in order to determine the final security risk level. This comprehensive approach not only considers the dimensions of data security, but also fully draws on the flexibility of fuzzy comprehensive evaluation to provide a more comprehensive and accurate means of analysing security assessment over the full lifecycle.

We must recognise some limitations, including subjectivity and dependency, consistency issues, and challenges to real-time performance. Firstly, subjectivity and dependency are significant limitations of this approach. The subjective judgment of decision-makers may lead to inconsistent evaluation results, especially in risk assessments involving professional knowledge and experience. Differences in subjective perceptions of risk factors among decision-makers may lead to biases in evaluation results, thereby affecting the reliability of decision-making. In addition, the challenge of real-time performance is also an aspect that we need to pay attention to. AHP and fuzzy comprehensive evaluation methods may require considerable computational time, which may become impractical in situations where immediate decision-making is required. In some emergency situations, ensuring fast and accurate risk assessment may become more important. However, this method still has potential for application and can be considered for fields such as project management decision-making, supply chain management, and environmental risk assessment in order to more comprehensively solve multi-criteria decision-making problems. In practical applications, methods can be flexibly adjusted according to specific situations to better meet the decision-making needs of different fields.

5. Simulation Experiment

In order to verify the effectiveness of the method in this paper, a simulated experimental environment was built to carry out experimental tests. The experimental environment simulates a unit's computer network, which includes an intranet and an extranet. The intranet faces the co-operative units and connects to the backbone network, which is used for marketing, financial statements, administrative business, and resource sharing; the extranet provides services to the users.

In the simulation experiments, the attacking host installs some brute force attack tools, including Nmap, SSH, and tfn2k DDoS. By scanning the -O parameter on the network, it achieves access to the internal information of the enterprise network. The SSH brute force attack tool is used to find out the SSH vulnerabilities present in the system and launch a malicious attack in order to obtain user information inside the system. The data generated during the attack are captured by the Snort Intrusion Detection System. The experiment was conducted a total of 10 times, the whole attack time lasted for 1 h, and the performance of the method of this paper was tested with constantly changing numbers of attacks.

We divided risk events into five levels based on their probability of occurrence:

(1) Level I (85~100%);
(2) Level II (70~84%);
(3) Level III (60~69%);
(4) Level IV (50~59%);
(5) Level V (Less than 50%).

As shown in Table 1, we conducted 10 experiments, and the test data covered four different risk levels. Through experimental testing of our proposed method, we observed that the risk values obtained in these 10 experiments were consistent with the actual results. This consistent observation further confirms the significant effectiveness of our proposed method in risk assessment. It is worth emphasising that our method performs well in handling different risk levels, demonstrating its universality and robustness in risk prediction and assessment. This series of experimental results clearly reveals the reliability and accuracy of our method, providing sufficient support for its potential value in practical applications. Overall, our method successfully meets the challenge of risk level diversity and provides a robust and reliable solution for risk management in practical scenarios.

Table 1. The performance of this paper's methodological security risk assessment.

Number of Experiments/Session	Probability of Occurrence of Real Risk/%	True Risk Level	Assessing Risk Level
1	77	II	II
2	67	III	III
3	90	I	I
4	55	IV	IV
5	78	II	II
6	89	I	I
7	64	III	III
8	51	IV	IV
9	95	I	I
10	75	II	II

The experimental results show that this paper's method achieves accurate results in the risk assessment of full-lifecycle security protection of data, this paper's method successfully identifies and responds to potential security threats, and all attack scenarios show good results, providing a reliable means of protection for data security. This further validates the effectiveness of this paper's method in practical applications and provides strong support for full-lifecycle data security management.

6. Conclusions

This paper presents a comprehensive examination of scenario-based protection analysis throughout the entire lifecycle of data. It delves into the security requirements and corresponding protection solutions for each phase of the data's lifecycle. In the section addressing the protection requirements across the data lifecycle, the paper meticulously analyses the data collection, transmission, storage, processing, sharing, and destruction stages. This scrutiny unveils the challenges in meeting scenario-specific requirements for ensuring data security.

The section focusing on the analysis of the full data lifecycle openness process introduces an overarching security protection process and the workflow of data services. This strategic approach aims to address and overcome security challenges effectively. Furthermore, this paper proposes a comprehensive risk detection scheme for protecting the complete lifecycle of data openness. The suggested method involves a hierarchical analysis and fuzzy comprehensive evaluation approach, providing a more accurate assessment of data security levels. Experimental results demonstrate an enhanced security detection efficacy.

Through this research, it becomes evident that data security is an ongoing evolutionary process, necessitating diverse technical means for support. Recognising that data security involves not only technical considerations but also regulatory compliance and business requirements, this paper advocates for a holistic protection system. The scenario-based protection analysis framework proposed in this paper serves as a valuable exploration and practical guide for data security protection. Its significance lies in ensuring data security and privacy protection, aligning with regulatory compliance and business needs.

Author Contributions: Conceptualisation, Y.S. and S.J.; methodology, Y.S.; software, S.J.; validation, Y.S., Q.S. and Y.S.; formal analysis, Q.S.; investigation, Q.G.; resources, Y.S.; data curation, Y.Y. (Yixin Yang); writing—original draft preparation, S.J.; writing—review and editing, Y.S.; visualisation, Q.S.; supervision, Y.Y. (Yue Yu); project administration, W.S.; funding acquisition, Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Science and Technology Programme of the power grid Corporation Headquarters. (Research on unsupervised learning-based data risk prevention and control technology for marketing online cooperation channel business, 5400-202318221A-1-1-ZN).

Data Availability Statement: All data are unavailable due to privacy restrictions.

Acknowledgments: We express our heartfelt gratitude to the reviewers and editors for their meticulous work.

Conflicts of Interest: Authors Wen Shen and Qian Guo were employed by the company Power Grid Smart Grid Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Xie, F.; Lin, H.; Lin, Y.; Zhang, X.; Liu, F. A lifecycle security solution for sensitive data. *Telecommun. Eng. Stand.* **2020**, *33*, 61–67.
2. Gartner. How to Use the Data Security goverznance Framework [EB/OL]. 15 August 2019. Available online: https://www.gartner.com/en/documents/3873369/how-to-use-the-data-security-governance-framework (accessed on 2 January 2024).
3. Yang, N. Summary Version of Data Security Governance White Paper. *Cyber Secur. Informatiz.* **2018**, *7*, 22–26.
4. Huawei Cloud, Huawei Cloud Data Security White Paper. Available online: https://www.huaweicloud.com/special/baipishusjaq.html (accessed on 30 October 2022).
5. Zhang, G.; Qiu, X.; Gao, Y. Software defined security architecture with deep learning-based network anomaly detection module. In Proceedings of the 11th IEEE Int Conf on Communication Software and Networks (ICCSN), Chongqing, China, 12–15 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 784–788.
6. Wu, Z.; Liang, C.; Li, Y. Intrusion detection method based on deep learning. In Proceedings of the IEEE Int Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, Haikou, China, 5–7 July 2019; IEEE: Piscataway, NJ, USA, 2022; pp. 445–452.

7. Indhumathi, R.; Devi, S.S. Healthcare Cramér generative adversarial network (HCGAN). *Distrib. Parallel Databases* **2022**, *40*, 657–673. [CrossRef]
8. Peng, C.; Jin, L.; Li, Z.; Xing, S.; Zhang, H. Data security risk analysis for the whole data life cycle. *Digit. Commun. World* **2022**, *2*, 99–101.
9. Yang, P.; Xiong, N.; Ren, J. Data security and privacy protection for cloud storage: A survey. *IEEE Access* **2020**, *8*, 131723–131740. [CrossRef]
10. Knowen, C.; Ronoh, L. Prioritizing Personal Data Protection in Insurance Organizations: A Review. *J. Inf. Secur. Cybercrimes Res.* **2023**, *6*, 40–56. [CrossRef]
11. Pan, J.; Wang, J.; Ma, Y. Smart Grid System Security Protection Authentication Encryption Technology. *Inf. Syst. Eng.* **2023**, *1*, 64–66.
12. Peng, J.; Gao, J. Research on Computer Network Information Security and Protection Strategies. *Comput. Digit. Eng.* **2011**, *39*, 121–124+178.
13. Ge, Y.; Xu, W.; Zhang, L. Integrative security protection system for full life cycle of big data based on SM crypto algorithm. In Proceedings of the International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2021), Sanya, China, 19–20 November 2021; Volume 12156, pp. 396–402.
14. Shah, S.I.H.; Peristeras, V.; Magnisalis, I. DaLiF: A data lifecycle framework for data-driven governments. *J. Big Data* **2021**, *8*, 1–44. [CrossRef]
15. Hansen, E.R.; Lissandrini, M.; Ghose, A.; Løkke, S.; Thomsen, C. Transparent integration and sharing of life cycle sustainability data with provenance. In Proceedings of the International Semantic Web Conference, Athens, Greece, 2–6 November 2020; Springer International Publishing: Cham, Switzerland, 2020; pp. 378–394.
16. Zhou, F.; Qing, C.; Pei, X.B. Data security protection of electricity marketing information system. *Comput. Mod.* **2019**, *3*, 111.
17. Wang, C.; Song, L. Exploration of data security protection system for industrial internet. *Inf. Commun. Technol. Policy* **2020**, *4*, 7–11.
18. Liu, Y.; Eckert, C.M.; Earl, C. A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Syst. Appl.* **2020**, *161*, 113738. [CrossRef]
19. Malinin, A.; Gales, M. Reverse kl-divergence training of prior networks: Improved uncertainty and adversarial robustness. *Adv. Neural Inf. Process. Syst.* **2019**, *2*, 32.
20. Lyu, H.M.; Sun, W.J.; Shen, S.L.; Zhou, A.N. Risk assessment using a new consulting process in fuzzy AHP. *J. Constr. Eng. Manag.* **2020**, *146*, 04019112. [CrossRef]
21. Lin, X. Design of network security risk detection system based on N-gram algorithm. *Inf. Comput. (Theor. Ed.)* **2023**, *35*, 215–217.
22. Li, N. A fuzzy evaluation model of college English teaching quality based on analytic hierarchy process. *Int. J. Emerg. Technol. Learn. (Ijet)* **2021**, *16*, 17–30. [CrossRef]
23. Kong, D. Research on network security situation assessment technology based on fuzzy evaluation method. *J. Phys. Conf. Ser.* **2021**, *1883*, 012108. [CrossRef]
24. Pamucar, D.; Ecer, F. Prioritizing the weights of the evaluation criteria under fuzziness: The fuzzy full consistency method–FUCOM-F. *Facta Univ. Ser. Mech. Eng.* **2020**, *18*, 419–437. [CrossRef]
25. Zhou, B.; Chen, J.; Wu, Q.; Pamučar, D.; Wang, W.; Zhou, L. Risk priority evaluation of power transformer parts based on hybrid FMEA framework under hesitant fuzzy environment. *Facta Univ. Ser. Mech. Eng.* **2022**, *20*, 399–420.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.