Value-based Data Governance and Security Protection for Virtual Power Plants Aggregated by Demand-side Flexible Loads

Jiabao Li, *Student Member*, *IEEE*, Hongxun Hui, *Senior Member*, *IEEE*, Yonghua Song, *Fellow*, *IEEE*, Ye Chen, *Member*, *IEEE*, Tao Chen, *Member*, *IEEE*, and Pierluigi Siano, *Senior Member*, *IEEE*

***Abstract*—Virtual power plants (VPPs) aggregated by demand-side flexible loads have become a key mechanism for balancing supply and demand in power systems. However, compared with conventional power plants, VPPs generate vast and heterogeneous datasets that are challenging to manage and protect effectively. Existing solutions often fail to unlock the full value of these data while imposing excessive security costs. This paper proposes a value-based data governance and security protection framework tailored for VPPs aggregated by demand-side flexible loads. Within this framework, a real-time data value assessment model is developed to dynamically assess the value of demand-side flexible loaddata. Furthermore, a fine-grained data management and protection strategy is introduced to enable differentiated governance and security measures. These measures are applied across different stages of the data life cycle according to the assessed data value levels. Numerical results demonstrate that the proposed framework enhances both data protection and operational performance while reducing security costs. Moreover, it promotes data circulation and value creation, and supports the sustainable and intelligent transformation of modern power systems.**

***Index Terms*—Virtual power plant, flexible load, data valuation, data governance, data security, security protection.**

I. Introduction

As the world increasingly prioritizes sustainable energy solutions, the integration of renewable energy sources into power systems is becoming a crucial focus. However, the continuous integration of renewable energy sources has made the power supply highly intermittent [1]. The resulting volatility on both the generation and load sides has compounded the pressure on maintaining the real-time balance. The traditional operational mode is facing severe challenges in smoothing load fluctuations. Consequently, the low-carbon and stable operation of power systems urgently requires large-scale, high-quality regulation resources. The development of technologies such as the Internet of Things, big data, artificial intelligence, and 5G communications has enabled virtual power plants (VPPs) aggregated by demand-side flexible loads to provide balancing services [2]. Compared with conventional generators, demand-side flexible loads have finer granularity, greater type diversity, and more distributed deployment. In urban power systems represented by Guangzhou, Shenzhen, Hong Kong, and Macao, these demand-side flexible loads are growing rapidly. These loads have great potential to participate in the regulation and flexibility services of the power system [3], [4].

At the same time, the volume of data generated by demand-side flexible loads has grown explosively. These data offer significant value, revealing economic and social patterns and creating new opportunities for the smart transition of the power system. Integrating digital technologies with power systems allows intelligent decision-making and dynamic resource optimization. Leveraging vast, high-value datasets and advanced analytics enables more effective regulation of decentralized demand-side flexible loads. This will also enhance renewable energy integration and ensure that the power system operates safely, efficiently, and sustainably, supporting social and economic development. However, as a new factor of production, data have unique properties such as non-rivalry [5], partial excludability [6], and heterogeneity [7]. These properties challenge traditional governance and value assessment systems. Additionally, research on demand-side flexible load data is still in its early stages, which presents several issues. These include inadequate assessment

Manuscript received: August 24, 2025; revised: October 21, 2025; accepted: November 20, 2025. Date of CrossCheck: November 20, 2025. Date of online publication: XX XX, XXXX.

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515010141), in part by the National Natural Science Foundation of China (No. 52407075), in part by the Multi-year Research Grant – General Research Grant 2025 of University of Macau (No. MYRG-GRG2025-00305-IOTSC), and in part by the Science and Technology Development Fund, Macau SAR (No. 001/2024/SKL).

J. Li, H. Hui (corresponding author), and Y. Song are with the State Key Laboratory of Internet of Things for Smart City, and Department of Electrical and Computer Engineering, University of Macau, Macao 999078, China, and they are also with the University of Macau Advanced Research Institute in Hengqin, Zhuhai 519031, China (e-mail: li.jiabao@connect.um.edu.mo; hongxunhui@um.edu.mo; yhsong@um.edu.mo).

Y. Chen is with the Sate Grid Jiangsu Electric Power Company Electric Power Research Institute, Nanjing 211103, China (e-mail: joey_chenye@foxmail.com).

T. Chen is with the School of Electrical Engineering, Southeast University, Nanjing 214135, China (e-mail: taoc@seu.edu.cn).

P. Siano is with the Department of Management & Innovation Systems, University of Salerno, Fisciano 84084, Italy, and he is also with the Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg 2006, South Africa (e-mail: psiano@unisa.it).

DOI: 10.35833/MPCE.2025.000787

methods, low data utilization efficiency, and difficulties in establishing stable data trading markets. Power systems also struggle to manage large-scale and diverse data due to the absence of unified standards, which leads to redundancy and poor data quality. Finally, traditional solutions that rely on coarse-grained encryption result in an inefficient allocation of security resources and high costs.

To fully harness the regulation capability of demand-side flexible loads and enhance renewable energy integration, it is crucial to promote efficient data circulation. Achieving this requires an effective allocation of security resources and improvements in data management, while maintaining robust data protection. Therefore, this study develops a value-based data governance and security protection framework tailored for VPPs aggregated by demand-side flexible loads. This framework integrates data value assessment into the fine-grained data management and protection strategy to enable simultaneous improvements in operational efficiency and data security.

The rest of this paper is organized as follows. Section II reviews related work. Section III presents the value-based data governance and security protection framework. Section IV proposes a real-time data value assessment model. In Section V, the feasibility and effectiveness of the proposed framework are verified through experiments and analysis. Section VI summarizes the paper and discusses future work.

II. Literature Review of Related Work

In recent years, research on demand-side flexible loads has emerged as a prominent topic in international academic literature, focusing on aspects such as load modeling and prediction [8], [9] and scheduling optimization and control [10], [11]. However, research and applications concerning demand-side flexible load data remain in their infancy. This section elaborates on relevant research from three perspectives: data value assessment, data governance, and data security.

A. Data Value Assessment

With the smart transition of power systems, the volume of generated data often exceeds the system capacity for effective utilization. Limited resources necessitate filtering for high-quality data to ensure efficient operations. Accurately assessing the value of demand-side flexible load data is essential for optimizing resource allocation, enhancing load regulation accuracy, and supporting decision-making. It also contributes to the development of power data trading markets. Recently, scholars have begun exploring methods to define and assess data value using mathematical models. Reference [12] introduces a cost-value model for smart meter data in demand response systems to optimize data granularity and maximize profitability. Reference [13] defines data value based on its ability to reduce uncertainty and improve profits, using parametric and nonparametric estimation methods. References [14] and [15] propose a quality-based information value assessment method using Shannon entropy to link information value to economic benefits. Reference [16] develops an end-to-end method to assess data value from a cost perspective in multi-energy systems. Reference [17] introduces a data value assessment method considering internal characteristics such as quality and timeliness, as well as external factors like electricity prices. This method prioritizes valuable data for the efficient dispatch of demand-side flexible loads in ancillary services.

Despite these efforts, existing methods for data value assessment in the power sector remain limited in scope, as they predominantly emphasize direct economic benefits. This focus restricts their adaptability to diverse stakeholders, heterogeneous data characteristics, and varying application scenarios. Consequently, there is a clear need for more comprehensive and context-aware assessment methods that can capture the multifaceted nature of data value.

B. Data Governance

The scale and quality of demand-side flexible load data are critical for effective load regulation and for safe, low-carbon grid operations. However, the power system, still in its smart transition, struggles to manage large-scale, multi-source, and high-dimensional data [18]. Issues such as redundancy, poor quality, and high processing costs impede data mining and application. Furthermore, fragmented systems with inconsistent formats and granularity create data silos, obstructing aggregation and sharing. To address these challenges, researchers have proposed various data management methods. Reference [19] emphasizes the role of big data technologies in enhancing data management and analytics. Reference [20] proposes a sustainable energy big data curation paradigm covering the entire data life cycle. Reference [21] introduces a deep learning model to evaluate data governance, considering power data density and the impact of abnormal data on system performance.

While these studies have advanced the understanding of data sharing and trading mechanisms, they offer limited insight into the governance of demand-side flexible load data from a value perspective. Furthermore, the efficiency of data and resource utilization in power systems, especially regarding demand-side flexible loads, has received limited attention. Addressing these gaps is essential for developing governance frameworks that enhance operational efficiency while maximizing data value.

C. Data Security

Smart metering and load management generate vast volumes of sensitive data, but current security solutions remain inadequate [22]. Traditionally, power systems relied on dedicated networks and physical isolation for protection [23]. However, the integration of public communication networks with decentralized resources complicates this isolation, making data security critical for both system stability and user privacy. Data breaches and attacks can lead to significant economic and social consequences [24]. Several studies have proposed approaches to enhance data security in governance frameworks. Reference [25] proposes a cloud-based approach for big data management in smart grids, utilizing identity-based encryption. Reference [26] assesses the vulnerabilities of cyber-physical power systems, while [27] uses XGBoost to detect and correct false data. Reference [28]

presents a security management approach for the power Internet of Things, using edge-cloud collaboration technology. Reference [29] proposes an approach based on distributed data storage with homomorphic encryption to enable secure data queries. Reference [30] introduces a life cycle-based security management approach for power data. Meanwhile, [31] leverages blockchain technology to improve the security, reliability, and traceability of grid data management.

Although current security solutions protect sensitive information to some extent, they are typically coarse-grained, which leads to inefficiencies in resource allocation and increases security costs, especially in large-scale, distributed power systems. The trade-off between security strength and system performance remains inadequately addressed. This highlights the need for fine-grained, scalable, and cost-effective security solutions.

Overall, existing studies have investigated the value, governance, and security of demand-side flexible load data from various perspectives. Supplementary Material A Table SAI summarizes the reviewed literature. Despite these efforts, several critical issues remain unresolved.

1) Due to the nonrivalry of data, data do not degrade or diminish after repeated use as traditional commodities do, allowing multiple parties to utilize the same data simultaneously. Moreover, the heterogeneity of data makes the data value assessment vary significantly across different users and application scenarios. However, current data value assessment methods primarily focus on direct economic benefits, offering limited adaptability.

2) Currently, most data governance research focuses on broad categories of power-related data, with an emphasis on integrating big data technologies into the power industry. These studies primarily address data sharing and trading, with limited research dedicated to the governance of demand-side flexible load datafrom a value perspective. In addition, few studies focus on improving the efficiency of data and resource utilization in power systems, especially in relation to demand-side flexible load data.

3) Existing security solutions in power systems rely on coarse-grained strategies, which are insufficient to address the challenges posed by widely distributed entities and continuously generated large-scale data in modern power systems. This leads to inefficient allocation of security resources and excessively high security costs. As a result, there is often a trade-off between security strength and system performance, which needs to be more effectively balanced.

To bridge these research gaps, this paper proposes a value-based data governance and security protection framework tailored for VPPs aggregated by demand-side flexible loads. Within this framework, a real-time data value assessment model is developed based on multiple indicators such as data quality, timeliness, security, and application scenarios. Based on this assessment, a fine-grained data management and protection strategy is introduced, covering the entire data life cycle. The demand-side flexible load data are categorized and graded by assessed value, allowing for differentiated governance and security measures. This enables a balance between security strength and system performance while reducing security costs.

III. Value-Based Data Governance and Security Protection Framework

A. Comprehensive Framework

To establish a unified standard for managing and securing large-scale demand-side flexible load data, this subsection presents a comprehensive framework. As shown in Fig. 1, the framework is structured according to the stages of the data life cycle and is organized into three hierarchical layers. This life-cycle-aware design ensures adaptive and value-based data governance and security protection to effectively address the challenges that arise at different stages of the data life cycle.

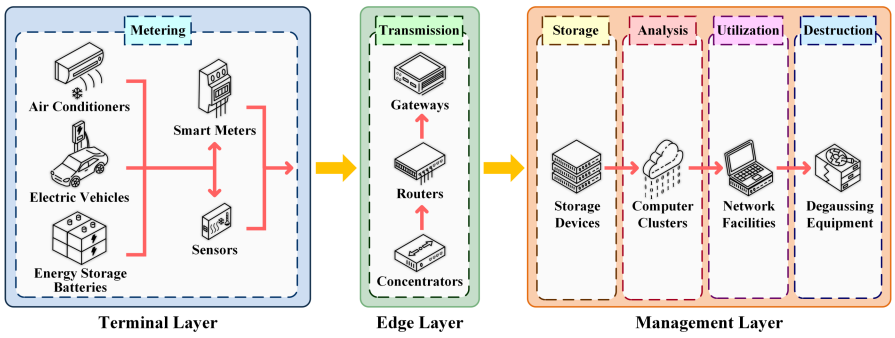

Fig. 1. Comprehensive framework for value-based data governance and security protection.

1) Terminal Layer

The terminal layer encompasses the metering stage of the data life cycle and various terminal devices on the user side. At the metering stage, various raw data such as users' electricity consumption patterns are captured in real time through metering devices. Specifically, data generated by demand-side flexible loads, including energy storage batteries, electric vehicles, and air conditioners, are collected by smart meters and other sensors. It is important to clarify that these metering devices may vary in granularity and deployment scope. The terminal layer accommodates diverse data acquisition methods, providing essential data for further analysis and serving as crucial support for the entire framework.

2) Edge Layer

The edge layer encompasses the transmission stage of the data life cycle and includes various communication devices

and network infrastructure. At this stage, data from the terminal layer are first aggregated and preliminarily processed by concentrators, which perform tasks such as data format conversion and information compression. The processed data are then securely transmitted via devices such as routers and gateways. Throughout the transmission process, encryption, authentication, and other security measures are employed to prevent data theft, tampering, or unauthorized access. These measures ensure that data are transmitted accurately and timely to their destination. Overall, the edge layer serves as a critical component in ensuring the smooth functioning of the entire framework.

3) Management Layer

The management layer encompasses the storage, analysis, utilization, and destruction stages of the data life cycle. It handles the reception and processing of data transmitted from the edge layer. In addition, anomaly detection mechanisms are essential to ensure the quality and reliability of downstream analytics and utilization. This is achieved by identifying and addressing erroneous or suspicious data entries in a timely manner.

In the data storage stage, large-capacity storage devices such as disk arrays ensure the secure and long-term retention of various types of data. Typically, historical data retention spans from several months to multiple years, depending on operational requirements and regulatory policies. Maintaining such extensive datasets enhances forecasting accuracy, enables long-term trend analysis, and supports strategic decision-making in power system operations. However, the accumulation of large datasets also presents significant computational challenges during storage and analysis. Efficient data management and scalable computing infrastructures are therefore essential to handle the increasing volume and complexity of data.

In the data analysis stage, high-performance server clusters and advanced data mining techniques enable in-depth analysis of demand-side flexible load data. This analysis helps optimize demand-side resource scheduling, accurately predict power demand trends, and support decision-making in the power system.

In the data utilization stage, secure and reliable data sharing mechanisms facilitate interoperability between participants in the power system, enhancing operational efficiency. Additionally, a data trading market can be established by implementing standard rules and pricing mechanisms for data transactions. Clear policies for data ownership and access control must be established to define access rights and usage scope and to prevent unauthorized usage or data breaches.

Finally, in the data destruction stage, data that have reached their expiration date are securely erased through physical destruction of storage media or by overwriting. Professional data destruction techniques ensure that sensitive information is fully deleted, preventing data breaches or misuse. In summary, the management layer is pivotal for enabling data circulation, maximizing data value, and ensuring the efficient and secure operation of the power system.

In summary, this framework ensures consistent quality management while enhancing security through tailored measures in each stage. By incorporating physical devices and real-world interactions, the framework accurately represents data flow, making it practical to implement. Ultimately, this structured framework improves data quality and security, leading to better decision-making, cost saving, and enhanced power system performance. Additionally, high data quality and security promote efficient data circulation and aggregation.

B. Fine-grained Data Management and Protection Strategy

To implement a fine-grained data management and protection strategy, it is essential to categorize and grade data in different stages of the data life cycle. Categorization standards can be adapted to specific needs. For example, this paper categorizes demand-side flexible load data by user type. Once categorized, the data are first assessed for their value and then further graded into three levels within each category, as illustrated in Fig. 2. This process can be formalized as:

$$\mathcal{D}_{C_j,G_k}=\left\{X_i\in\mathcal{D}|C(X_i):\mathcal{D}\to\left\{C_1,C_2,C_3\right\}\wedge G(X_i):\mathcal{D}\to\left\{G_1,G_2,G_3\right\}\right\} \tag{1}$$

where $\mathcal{D}$ is a set consisting of datasets of demand-side flexible loads, with X_i being the i^{th} dataset; $\mathcal{D}_{C_j,G_k}$ is a set consisting of datasets that are categorized into category C_j and graded at level G_k; $C(X_i)$ is a function that categorizes each dataset X_i into one of three categories: C_1 for industrial users, C_2 for commercial users, and C_3 for residential users; and $G(X_i)$ is a function that assigns a value level to each dataset: G_1 for high value, G_2 for medium value, and G_3 for low value.

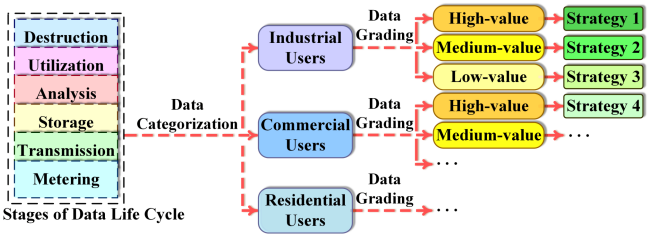

Fig. 2. Overall structure of data categorization and grading.

This data categorization and grading scheme allows for a more precise reflection of different electricity consumption patterns and enables the development of targeted security measures. Adopting differentiated governance and security measures provides a foundation for optimizing demand-side management, and ultimately improves the overall efficiency and security of the power system.

Different categories of demand-side flexible load data not only display distinct characteristics in terms of load patterns, but also vary significantly in their importance with respect to economic impact and data security. The data from industrial users generally reflect higher load levels with relatively stable power consumption curves. These data often include sensitive information related to production efficiency and equipment operation, necessitating stringent security. Commercial loads tend to fluctuate in cycles, with noticeable

variations in load curves between weekdays and weekends. Such data can provide valuable insights into business activity, customer traffic, and other economic indicators, making them particularly useful for economic analysis. In contrast, the data from residential users typically show more pronounced daily fluctuations, influenced by external factors such as weather and seasonal changes, leading to a high degree of randomness and diversity. Such data reveal household electricity consumption patterns and lifestyle habits, which are highly valuable for research but require strong privacy protection.

Based on the grade, differentiated governance and security measures are developed for demand-side flexible load data at each value level. The primary differences involve the data retention period and the strength of security measures. High-value data are retained for the longest period to enable continuous analysis and utilization. They are protected with longer key lengths and more robust cryptographic algorithms. For medium-value data, the focus is on balancing security and efficiency. Medium-length keys provide adequate protection while maintaining processing efficiency. Although these data are retained for a shorter period than high-value data, they are still long enough to meet operational needs. Low-value data are retained for the shortest period to minimize storage costs and optimize system resource usage. They are protected with lightweight cryptographic algorithms to ensure efficiency while maintaining necessary security.

In summary, this fine-grained data management and protection strategy enhances data security, optimizes system resource utilization, and reduces security costs by aligning differentiated governance and security measures with data value. Furthermore, this strategy integrates with the data life cycle, allowing for dynamic adjustments to governance and security measures in each stage. This ensures the effective management and protection of demand-side flexible load data throughout the entire life cycle.

C. Data Security Measures

To ensure the security and privacy of demand-side flexible load data throughout the entire life cycle, encryption remains the most fundamental and effective measure. By encrypting the data, unauthorized users are prevented from interpreting the original content. The advanced encryption standard (AES) is a widely adopted cryptographic algorithm, extensively applied in smart grids and other critical infrastructures [32] due to its high computational efficiency in large-scale data processing scenarios.

AES is a block cipher that operates on 128-bit plaintext blocks and supports key lengths of 128, 192, or 256 bits, corresponding to AES-128, AES-192, and AES-256. The number of encryption rounds is determined by the key length: 10 rounds for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys [33]. According to Shannon's principles of cryptographic design, AES transforms plaintext into ciphertext through a sequence of substitution-permutation operations that provide strong confusion and diffusion. It implements these operations using a structured round function comprising four core steps [34], which are described in Supplementary Material A.

Formally, AES encryption is defined as:

$$C = E_K(P) \tag{2}$$

where $P \in \{0,1\}^{128}$ is the plaintext block; $K \in \{0,1^m\}$ ($m \in \{128, 192, 256\}$) is the secret key; and E_K denotes the AES encryption function parameterized by the key K.

The corresponding decryption function is:

$$P = D_K(C)$$

where D_K is the AES decryption function parameterized by the key K.

AES provides robust resistance against known cryptanalytic techniques such as linear cryptanalysis, differential attacks, and meet-in-the-middle attacks. The selection of key length directly impacts the security level and computational overhead. Longer key lengths enhance brute-force resistance but incur additional processing cost. In practice, AES-128 is suited for high-throughput, low-latency scenarios such as real-time control commands or short-interval forecasting, where it provides sufficient security with minimal computational overhead. AES-192 offers a balance between security and efficiency, making it appropriate for medium-term data protection like daily load archiving or monthly billing. AES-256 is employed for high-value, sensitive data that require long-term storage and stringent security. Its 256-bit key provides the highest security margin, effectively mitigating risks from future computational advances and ensuring privacy compliance.

IV. Real-time Data Value Assessment Model

Accurately assessing the value of demand-side flexible load data is essential for optimizing power system resource allocation and improving operational efficiency. However, data value depends on various factors, including scale, quality, multi-source integration, and application scenario, making it difficult to capture with a single metric. This section introduces a generalized, multi-dimensional real-time data value assessment model that accounts for intrinsic, application, and security value. The model is decoupled from direct economic metrics and aims to optimize resource allocation within the value-based data governance and security protection framework. It improves data utilization efficiency and enhances adaptability across different application scenarios. Specifically, the data value can be expressed as:

$$VoD_\gamma(\boldsymbol{X}) = VoI(\boldsymbol{X}) + VoA_\gamma(\boldsymbol{X}) + VoS(\boldsymbol{X}) \tag{3}$$

where $\boldsymbol{X} = [x_1, x_2, \ldots, x_n]$ is the demand-side flexible load dataset, containing n data points $x_1, x_2, \ldots, x_n$ collected by smart meters; $VoD_\gamma(\boldsymbol{X})$ is the value of the dataset $\boldsymbol{X}$ in the application scenario γ; $VoI(\boldsymbol{X})$ is the intrinsic value of the dataset $\boldsymbol{X}$, reflecting its quality, timeliness, and other inherent characteristics that do not depend on specific application scenarios; $VoA_\gamma(\boldsymbol{X})$ is the application value of the dataset $\boldsymbol{X}$ in the application scenario γ, depending on how effectively the data support specific use cases in the scenario; and $VoS(\boldsymbol{X})$ is the security value of the dataset $\boldsymbol{X}$, accounting for the sensitivity of the data and the risks associated with potential breaches or unauthorized access.

Figure 3 illustrates the key computational steps for assessing intrinsic, application, and security values, thereby facilitating a clearer understanding of the sequential relationships among the assessment components.

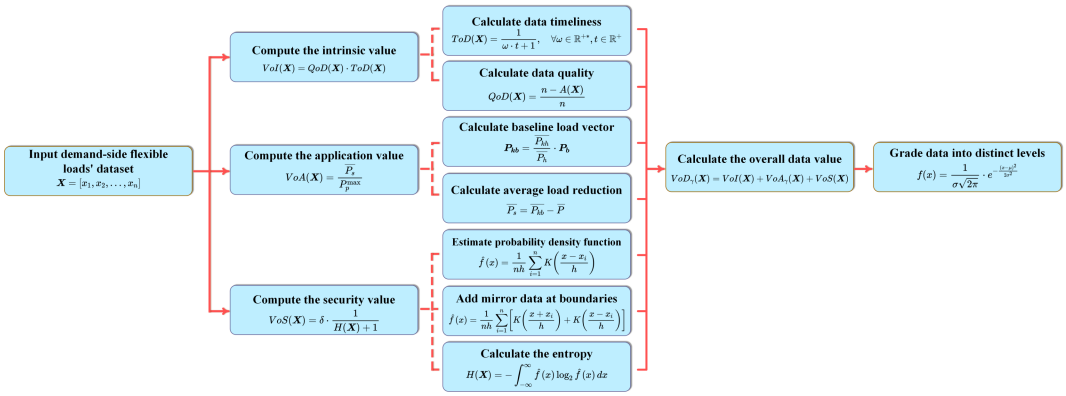

Fig. 3. Key computational steps for assessing intrinsic, application, and security values.

It is worth noting that (3) employs an equal-weighted additive structure, wherein VoI, VoA_γ, and VoS contribute equally to the overall composite value. This design choice enhances model interpretability, facilitates implementation, and provides a neutral and transparent baseline in the absence of detailed domain-specific knowledge of data value assessment. Such an approach is particularly useful in early-stage system deployments or in the design of data governance frameworks, where transparency and interpretability are critical. Moreover, as discussed in [17], equal-weighted models have demonstrated effectiveness in similar scenarios involving coordinated optimization within power-communication networks. Nevertheless, it is acknowledged that equal weighting may not be universally appropriate across all application scenarios. Hence, the equal-weighted model should be regarded as a foundational approximation that can be further refined to accommodate scenario-specific requirements.

A. Intrinsic Value

The intrinsic value of data is determined primarily by the inherent characteristics, independent of specific application scenarios. Key factors that define this intrinsic value include data quality and timeliness. Only when data possess high intrinsic value can they effectively support various applications. The intrinsic value of data can be quantified as:

$$VoI(\boldsymbol{X})=QoD(\boldsymbol{X})\cdot ToD(\boldsymbol{X}) \tag{4}$$

where $QoD(\boldsymbol{X})$ is the quality of the dataset $\boldsymbol{X}$, which reflects how accurate and complete the data are for their intended purpose; and $ToD(\boldsymbol{X})$ is the timeliness of the dataset $\boldsymbol{X}$, which reflects the relevance of the data with respect to the time it is used or analyzed.

1) Data quality. Data quality is a comprehensive concept that varies depending on the context, and there is no universal definition [35]. In this paper, data quality refers to the accuracy and completeness of data for its intended purpose. Data quality can fluctuate as it moves between devices and undergoes processing. For instance, data precision might degrade during transmission, lowering its quality. Conversely, techniques such as smoothing and interpolation can improve data quality by correcting anomalies. High-quality data ensure reliable decision-making, which is crucial for the efficient operation of the power system. In contrast, poor-quality data can result in inaccurate strategies and may compromise the grid safety and stability. The data quality can be expressed as:

$$QoD(\boldsymbol{X})=\frac{n-A(\boldsymbol{X})}{n} \tag{5}$$

where n is the number of data points in the dataset $\boldsymbol{X}$; and $A(\boldsymbol{X})$ is the number of outliers in the dataset $\boldsymbol{X}$. Outliers are defined as missing values or unreasonable deviations caused by metering or transmission errors. These outliers are detected using the standard deviation method. Specifically, the mean μ and standard deviation σ of the dataset $\boldsymbol{X}$ are first calculated. The normal range is defined as $\mu\pm k\sigma$, where k is an adjustment factor. Data points outside this range are flagged as outliers. In practical applications, k can be calibrated based on historical error statistics and validated using expert-labeled datasets.

2) Data timeliness. Data timeliness refers to the degree to which data are available and accessible at the moment they are needed [36]. Timeliness is critical to intrinsic value—data become less useful as they age, losing relevance for decision-making. Demand-side flexible load data with high timeliness, available shortly after generation, are especially valuable for making prompt and effective decisions. Conversely, outdated data, with low timeliness, may no longer accurately represent current conditions, leading to suboptimal decisions. The data timeliness can be expressed as:

$$ToD(\boldsymbol{X})=\frac{1}{\omega t+1}\quad \forall\omega\in\mathbb{R}_{>0},\forall t\in\mathbb{R}_{\geq 0} \tag{6}$$

where t is the age of the dataset $\boldsymbol{X}$; and ω is an adjustment factor that accounts for the varying decay rates of data timeliness across different situations. ω can be adjusted according to application latency requirements or inferred from historical decision performance. Specifically, let t_m represent the time when the load data are recorded, and t_c represent the time when they are assessed. The age t is the difference between these two timestamps. As time passes and t increases, data timeliness decreases, approaching a minimum value of 0. When $t_c=t_m$, the timeliness reaches its maximum value

of 1.

B. Application Value

The application value of data can vary significantly depending on the specific scenario, making it difficult to define with a single measure. To illustrate this concept, this paper uses the scenario of demand response evaluation as an example. Evaluating the impact of demand response is essential for its successful implementation. By comparing load data before and after a demand response event, the effectiveness of these measures and their impact on the stability and efficiency of the power system can be evaluated. This evaluation helps to optimize demand response strategies to increase the overall benefits to the power system. In this scenario, the application value of demand-side flexible load data can be expressed using the peak performance index. This index is defined as the ratio of the average load reduction during a demand response event to the user's maximum peak load demand [37]. A higher peak performance indicates greater enthusiasm from the user in participating in demand response, as well as higher potential for effective response. Consequently, data in such cases hold greater application value and can significantly influence the design of future demand response strategies.

Therefore, the application value of data can be expressed as:

$$VoA_{\gamma}(\boldsymbol{X}) = \frac{\overline{P_s}}{P_p^{\max}} \tag{7}$$

where $\overline{P_s}$ is the average load reduction during the demand response event based on the dataset $\boldsymbol{X}$; and $P_p^{\max}$ is the user's maximum peak load demand observed in the dataset $\boldsymbol{X}$.

To calculate $\overline{P_s}$, the baseline load must first be established. Since the baseline load represents the full daily load profile, it is treated as a vector, with each element corresponding to a specific time interval. For a given time interval *PT*, the five most recent typical days with the highest daily load are selected. The uncorrected baseline load vector $\boldsymbol{P}_b$ is then calculated by averaging the load values at each time interval across these days. However, since random factors such as climate can influence the load, the uncorrected baseline load must be adjusted. The corrected baseline load vector P_{kb} is computed as:

$$\boldsymbol{P}_{kb} = \frac{\overline{P_{kh}}}{\overline{P_h}} \boldsymbol{P}_b \tag{8}$$

where $\overline{P_{kh}}$ is the average load recorded during the same time intervals within the two hours before the demand response event; and $\overline{P_h}$ is the average historical load for the same time intervals on the days used to establish the uncorrected baseline load. The adjustment accounts for external variations, ensuring the corrected baseline load accurately reflects real conditions.

Finally, $\overline{P_s}$ can be calculated as:

$$\overline{P_s} = \overline{P_{kb}} - \overline{P} \tag{9}$$

where $\overline{P_{kb}}$ is the average corrected baseline load during the demand response period; and $\overline{P}$ is the actual average load during that period. The difference between the corrected baseline and the actual load represents the load reduction achieved through the demand response event.

C. Security Value

Information entropy is a fundamental tool for measuring data uncertainty, playing a key role in communication theory. In data security, it is commonly used to quantify privacy risks, assess the potential for data breaches, and identify emerging security threats [38]. In this paper, information entropy is applied to assess the security value of demand-side flexible load data, which refers to the required level of protection for the data. Higher entropy in the data typically indicates greater randomness in electricity consumption, reducing the likelihood of sensitive information being exposed. As a result, the data have a lower security value. Conversely, data with lower entropy suggest more predictable consumption patterns, which may potentially reveal sensitive information. These data have a higher security value and therefore require stricter security measures.

To calculate information entropy, it is essential to first approximate the probability distribution of the random variables. This paper employs kernel density estimation to estimate the probability density function:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{10}$$

where h is the bandwidth;; and $K(\cdot)$ is the kernel function.

Common kernel functions include the Gaussian kernel, uniform kernel, and triangular kernel functions. In this paper, the Gaussian kernel function (11) is chosen.

$$K(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \tag{11}$$

However, traditional kernel density estimation assumes that data distributions are smooth across the real number range. In practice, demand-side flexible load data often have boundary constraints, as they cannot be negative. To address this, mirror data are added at the boundaries using the reflection method [39] to mitigate boundary effects:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} \left[K\left(\frac{x + x_i}{h}\right) + K\left(\frac{x - x_i}{h}\right) \right] \tag{12}$$

Differential entropy as an extension of classical information entropy is used to calculate the entropy of demand-side flexible load data in this paper:

$$H(\boldsymbol{X}) = -\int_{-\infty}^{\infty} \hat{f}(x) \log_2 \hat{f}(x) \mathrm{d}x \tag{13}$$

where $H(\boldsymbol{X})$ is the entropy of the dataset $\boldsymbol{X}$.

The security value is inversely proportional to the information entropy, given by:

$$VoS(\boldsymbol{X}) = \delta \frac{1}{H(\boldsymbol{X}) + 1} \tag{14}$$

where δ is an adjustment factor used to scale the security value. The parameter δ can be calibrated based on privacy risk assessments, domain-specific regulatory requirements, and policy constraints.

Once the data value assessment is complete, the data can be further graded into different levels. Assuming that the data values follow a normal distribution within the interval

[0, 3], the probability density function is given by(15), where the mean μ is 1.5 and the standard deviation σ is 0.5.

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{15}$$

As shown in Fig. 4, data in the range of [0, 1) are graded as low-value, data in the range of [1, 2) as medium-value, and data in the range of [2, 3] as high-value.

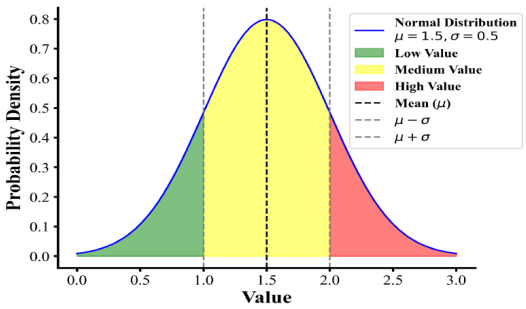

Fig. 4. Grading of load data based on value distribution.

V. Case Study

To verify the effectiveness of the proposed framework, this section presents a case study on value-based data governance and security protection for VPPs aggregated by demand-side flexible loads. The framework is specifically applied to the data storage stage, within the scenario of demand response evaluation.

A. Test System

The data used in this case study are derived from load profiles collected from a selected group of users in a Chinese city who participated in peak-load regulation. The dataset includes user load conditions on typical days during two demand response events, with data collected every 15 min. Both events ran from 14:00 to 15:00, during which the regulation focused on flexible loads. The two events occurred two months apart, with a total of 498 participants—330 in the first event and 168 in the second event. Six months after the first event, the proposed framework was applied to the stored data. Table I summarizes the key metrics of the two demand response events.

TABLE I
Key Metrics of Demand Response Events

Item	Event 1	Event 2
Number of participants	330	168
Event duration	14:00-15:00	14:00-15:00
Data collection interval	15 min	15 min
Application of proposed framework	Six months after Event 1	Four months after Event 2

B. Data Categorization

The first step, as introduced in Section III-B, is data categorization. The raw dataset is categorized into three distinct categories based on user types: industrial, commercial, and residential users. Specifically, industrial users generally correspond to large-scale facilities with relatively stable and high-volume energy consumption, such as manufacturing plants or production lines. Commercial users mainly include public buildings or enterprises with more variable load profiles influenced by business hours and occupancy patterns. Residential users represent individual households with diverse and less predictable consumption behaviors. The dataset consists of 121 industrial users, 256 commercial users, and 121 residential users.

C. Data Value Assessment

Next, commercial users are the main focus of further analysis. The demand-side flexible load data for each of the 256 commercial users are first assessed for value and then graded. The value assessment process is illustrated in Fig. 5, where the value is assessed across three dimensions: intrinsic value, application value, and security value. The assessment results for each dimension are displayed using a heat map.

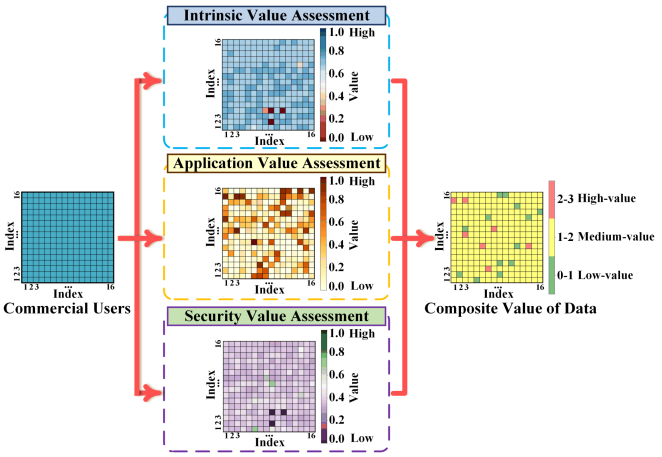

Fig. 5. Value assessment process for load data of 256 commercial users.

The heat map of intrinsic value shows that most commercial users' data have relatively good intrinsic value, while only a small portion exhibits poor value. Specifically, most light blue squares fall within the range of 0.6-0.8, indicating stable and moderately timely data. Some dark blue squares show values of 0.8-0.9, reflecting highly reliable and timely data. By contrast, several dark red squares correspond to very low values, around 0-0.2, caused by missing records or abnormal data. Orange squares, typically in the range of 0.2-0.6, indicate data with occasional outliers. This numerical distribution clearly highlights how timeliness and quality jointly determine intrinsic value.

The heat map of application value reflects the actual contribution to demand response. Darker squares correspond to values of 0.8-1.0, indicating active participation and significant potential. Conversely, lighter yellow squares with values around 0-0.3 indicate weaker contribution potential. A large portion of the squares are orange, showing values of 0.3-0.8, which suggest moderate support for demand response.

The heat map of security value shows that most data have a relatively low security value. This is mainly due to the ran-

domness of electricity consumption patterns, which reduces the likelihood that sensitive information will be exposed and implies fewer security resources are required. However, a few green squares show higher values, around 0.6-0.8, indicating users with more predictable consumption patterns. Their data could reveal sensitive information and therefore require stronger protection. Interestingly, several dark purple squares show values dropping to nearly 0, where outliers obscure the consumption pattern so strongly that the data become essentially non-informative from a security perspective.

Finally, based on the composite assessment, the value of the demand-side flexible load data for each of the 256 commercial users was determined. Specifically, 7 users provided high-value data (values in the range of $[2, 3]$), 236 users provided medium-value data (values in the range of $[1, 2)$), and 13 users provided low-value data (values in the range of $[0, 1)$). In the composite heat map, red squares represent high-value data. These data are timely, accurate, and highly contributive, but they also require significant security resources to prevent sensitive information breaches. Yellow squares represent medium-value data, which provide meaningful support for load regulation and grid stability, although they are less critical and less sensitive than high-value data. Green squares indicate low-value data. These are characterized by poor quality, limited contribution, and an inability to accurately reflect the current situation. They require only basic security resources.

D. Differentiated Governance and Security Measures

Once the data are graded by value, differentiated governance and security measures can be implemented. As previously mentioned, the primary differences involve the data retention period and the strength of security measures. AES supports flexible key lengths, enabling hierarchical protection based on the value and security requirements of the data. Figure 6 compares the security measures used in the proposed framework with those used in the traditional framework.

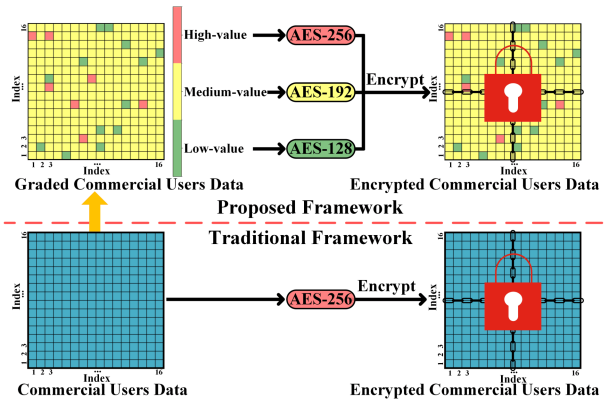

Fig. 6. Comparison of security measures between proposed and traditional frameworks.

1) Traditional framework: all data are encrypted with the highest security level (AES-256) without distinguishing the data value.

2) Proposed framework: different security levels of encryption (AES-128/192/256) are applied according to the data value to achieve hierarchical protection.

Specifically, the traditional framework often fails to recognize the varying importance and security requirements of different data. As a result, it tends to apply the strongest encryption such as AES-256 to all data indiscriminately. However, overuse of strong encryption on low-value data leads to unnecessary performance overhead. More stringent security measures also complicate management and increase security costs, ultimately reducing system efficiency. Therefore, while the traditional framework ensures security, it fails to strike a balance between protection and resource efficiency.

In the proposed framework, AES-256, AES-192, and AES-128 [40] are used to encrypt high-, medium-, and low-value data, respectively. Table II shows the comparison of AES-256, AES-192, and AES-128. AES encryption strength is determined by key length, which defines the key space size [41]. AES-128 offers a key space of 2^{128}, making brute-force attacks infeasible with current computing power. AES-192 provides a key space of 2^{192}, offering greater security. AES-256 with a key space of 2^{256} is the most secure and remains resistant to future quantum attacks due to the vast number of required attempts. From a performance perspective, the computational complexity of AES increases with key length [42]. AES-128 has 10 rounds of encryption, AES-192 has 12, and AES-256 has 14. More rounds increase encryption time and resource usage, raising security costs. Therefore, AES-128 not only meets the basic security requirements for demand-side flexible load data but is also resource-efficient and provides higher encryption speed, making it suitable for low-value data that can be destroyed after the required retention period. In contrast, AES-256 offers stronger protection but requires more resources and has lower encryption efficiency. AES-192 strikes a balance between security and performance, falling between AES-128 and AES-256. Accordingly, AES-192 is designated for medium-value data, while AES-256 is recommended for high-value data that require long-term storage and usage. This hierarchical protection ensures data security, improves resource utilization, and reduces security costs.

TABLE II
COMPARISON OF AES-256, AES-192, AND AES-128

AES	Key length (bit)	Key space	Number of rounds	Security level	Performance
AES-128	128	$2^{128} \approx 3.4 \times 10^{38}$	10	Standard	High
AES-192	192	$2^{192} \approx 6.3 \times 10^{57}$	12	Strong	Medium
AES-256	256	$2^{256} \approx 1.2 \times 10^{77}$	14	Highest	Low

To evaluate the effectiveness of the proposed framework, we compared it with a traditional framework by measuring the time and memory usage required to encrypt the commercial users' data. We also assessed the overall information entropy [43] after encryption. The experiment was conducted on a personal computer with an Intel(R) Core(TM) i7-8565U CPU @ 1.80 GHz, Intel(R) UHD Graphics 620, and NVIDIA

GeForce MX150. As shown in Table III, the overall information entropy after encryption is similar for both frameworks. This indicates that the proposed framework provides nearly the same level of data security as the traditional framework. However, the proposed framework requires 16.24% less memory and achieves a 18.60% faster execution speed compared with the traditional framework. The findings confirm that differentiated governance and security measures can effectively ensure data security and alleviate the computational burden on the system, thereby reducing unnecessary security costs and optimizing resource utilization.

Another key difference is the data retention period. Traditional framework retains all data for the maximum period, overlooking varying storage needs and raising risks. By comparison, differentiated governance and security measures ensure that sensitive data are securely destroyed after their useful life, whcih frees up storage and improves efficiency. To meet audit, compliance, or reporting requirements, the system must retain demand-side flexible load data for at least 3 years with a maximum retention period of 5 years. Figure 7 shows the comparison of the governance measures between the proposed and traditional frameworks. The gray area indicates that data have been securely destroyed and the corresponding storage resources have been released.

TABLE III
COMPARISON OF THE PERFORMANCE OF SECURITY MEASURES BETWEEN PROPOSED AND TRADITIONAL FRAMEWORKS

Framework	Overall information entropy (bit)	Execution time (s)	Memory usage (MB)
Traditional framework	7.9982	0.0629	0.2211
Proposed framework	7.9980	0.0512	0.1852

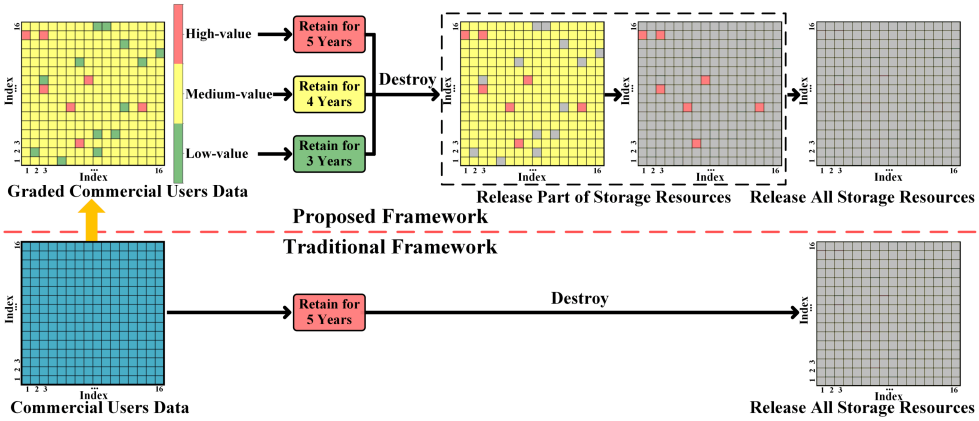

Fig. 7. Comparison of governance measuress between proposed and traditional frameworks.

1) Traditional framework: retaining all data for the maximum period of 5 years prevents the premature deletion of critical data.

2) Proposed framework: retention periods are differentiated according to data value, after which the data are securely destroyed, releasing storage space for new data. Specifically, high-value data are retained for 5 years to support long-term decision-making and trend analysis. Medium-value data are retained for 4 years to meet business and short-term analysis needs. Low-value data are retained for 3 years to comply with regulations and audit requirements.

The differentiated governance and security measures optimize storage resource utilization. Specifically, compared with the traditional framework, the proposed framework enables the early release of 5% of storage capacity by the end of the third year, an additional 92% by the end of the fourth year, and about 97% in total before the fifth year. The remaining 3% is released at the maximum retention period by the end of fifth year. In contrast, the traditional framework retains all data until the end of the maximum retention period, keeping storage resources fully occupied throughout. The comparison of the performance of governance measures between the two frameworks is summarized in Table IV.

In summary, the numerical results demonstrate that the proposed framework enhances both data protection and operational performance while reducing security costs.

TABLE IV
COMPARISON OF PERFORMANCE OF GOVERNANCE MEASURES BETWEEN PROPOSED AND TRADITIONAL FRAMEWORKS.

Time	Storage release status	
	Traditional framework	Proposed framework
The third year	No release	Release 5%
The fourth year	No release	Release 92%
The fifth year	Release 100%	Release 3%

VI. CONCLUSION

This paper proposes a value-based data governance and security protection framework tailored for VPPs aggregated by demand-side flexible loads, addressing challenges in securing large-scale data while optimizing operational performance. The key contribution is a real-time data value assessment model using multi-dimensional indicators such as data quality, timeliness, security, and application scenarios. It also introduces a fine-grained data management and protection strategy covering the entire data life cycle. Demand-side flexible load data are categorized and graded by assessed val-

ue, allowing for differentiated governance and security measures. Numerical results demonstrate the effectiveness of the proposed framework in enhancing data protection and resource utilization, striking a balance between security strength and system performance while reducing security costs.

However, the proposed real-time data value assessment model focuses on demand-side flexible load data in a single scenario and assumes static, equal-weighted contributions of value dimensions. While simplifying implementation, this may overlook scenario-specific priorities in dynamic or risk-sensitive contexts. Future work will extend the real-time data value assessment model to multiple grid scenarios and adopt adaptive weighting mechanisms based on learning or optimization techniques. Additionally, it could be improved through strategies such as periodic re-evaluation and sliding window analysis to better capture the temporal evolution of data importance. Integrating time-series models (e. g., long short-term memory (LSTM) networks or transformer architectures) would further strengthen the ability to model timeliness and track changes in data value. Furthermore, multiple adjustable parameters governing data quality range, timeliness decay, and security value scaling are currently set empirically. Automated tuning using methods such as Bayesian optimization or reinforcement learning is expected to improve the generalizability and robustness of the data value assessment across diverse scenarios.

In conclusion, the proposed framework promotes data circulation and value creation, supporting the sustainable and intelligent transformation of modern power systems.

REFERENCES

[1] S. Impram, S. Varbak Nese, and B. Oral, "Challenges of renewable energy penetration on power system flexibility: a survey," *Energy Strategy Reviews*, vol. 31, p. 100539, Sept. 2020.

[2] Y. Fu, H. Bai, Y. Cai *et al*., "Optimal configuration method of demand-side flexible resources for enhancing renewable energy integration," *Scientific Reports*, vol. 14, p. 7658, 2024.

[3] W. Wang, Y. Luo, and D. Zhao, "The power transition under the interaction of different systems: a case study of the Guangdong – Hong Kong–Macao greater bay area," *Sustainability*, vol. 15, no. 6, p. 5577, Jan. 2023.

[4] T. Xu, T. Chen, C. Gao *et al*., "Intelligent home energy management strategy with internal pricing mechanism based on multiagent artificial intelligence-of-things," *IEEE Systems Journal*, vol. 17, no. 4, pp. 6045-6056, Dec. 2023.

[5] C. I. Jones and C. Tonetti, "Nonrivalry and the economics of data," *American Economic Review*, vol. 110, no. 9, pp. 2819-2858, Sept. 2020.

[6] Y. Carriere-Swallow and V. Haksar, "The economics and implications of data: an integrated perspective," *Departmental Papers*, vol. 18, no. 12, p. 51, Sept. 2019.

[7] L. Wang, "Heterogeneous data and big data analytics," *Automatic Control and Information Sciences*, vol. 3, no. 1, pp. 8-15, Jan. 2017.

[8] C. Utama, S. Troitzsch, and J. Thakur, "Demand-side flexibility and demand-side bidding for flexible loads in air-conditioned buildings," *Applied Energy*, vol. 285, p. 116418, Mar. 2021.

[9] N. Mahdavi, J. H. Braslavsky, M. M. Seron *et al*., "Model predictive control of distributed air-conditioning loads to compensate fluctuations in solar power," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 3055-3065, Nov. 2017.

[10] H. Hui, Y. Ding, Z. Lin *et al*., "Capacity allocation and optimal control of inverter air conditioners considering area control error in multi-area power systems," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 332-345, Jan. 2020.

[11] H. Hui, Y. Ding, W. Liu *et al*., "Operating reserve evaluation of aggregated air conditioners," *Applied Energy*, vol. 196, pp. 218-228, Jun. 2017.

[12] S. Feuerriegel, P. Bodenbenner, and D. Neumann, "Value and granularity of ICT and smart meter data in demand response systems," *Energy Economics*, vol. 54, pp. 1-10, Feb. 2016.

[13] B. Wang, Q. Guo, T. Yang *et al*., "Data valuation for decision-making with uncertainty in energy transactions: a case of the two-settlement market system," *Applied Energy*, vol. 288, p. 116643, Apr. 2021.

[14] L. Chen, Z. Wu, J. Wang *et al*., "Toward future information market: an information valuation paradigm," in *Proceedings of 2021 IEEE PES General Meeting*, Washington DC, USA, Jul. 2021, pp. 1-5.

[15] M. Yu, J. Wang, J. Yan *et al*., "Pricing information in smart grids: a quality-based data valuation paradigm," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3735-3747, Sept. 2022.

[16] Y. Zhou, Q. Wen, J. Song *et al*., "Load data valuation in multi-energy systems: an end-to-end approach," *IEEE Transactions on Smart Grid*, vol. 15, no. 5, pp. 4564-4575, Sept. 2024.

[17] L. Ma, H. Hui, and Y. Song, "Data valuation-aware coordinated optimization of power-communication coupled networks considering hybrid ancillary services," *IEEE Transactions on Smart Grid*, vol. 16, no. 1, pp. 568-581, Jan. 2025.

[18] B. A. Schuelke-Leech, B. Barry, M. Muratori *et al*., "Big data issues and opportunities for electric utilities," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 937-947, Dec. 2015.

[19] H. Daki, A. E. Hannani, A. Aqqal *et al*., "Big data management in smart grid: concepts, requirements and implementation," *Journal of Big Data*, vol. 4, no. 1, p. 13, Apr. 2017.

[20] M. Zeng, Y. Xu, H. Wu *et al*., "Sustainable insights for energy big data governance in China: full life cycle curation from the ecosystem perspective," *Sustainability*, vol. 14, no. 10, p. 6013, Jan. 2022.

[21] K. Zhou, E. Meng, Q. Jin *et al*., "Evaluation of data governance effectiveness in power grid enterprises using deep neural network," *Soft Computing*, vol. 27, no. 23, pp. 18333-18351, Sept. 2023.

[22] W.-L. Chin, W. Li, and H.-H. Chen, "Energy big data security threats in IoT-based smart grid communications," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 70-75, Oct. 2017.

[23] H. Hui, Y. Ding, Q. Shi *et al*., "5G network-based Internet of Things for demand response in smart grid: a survey on application potential," *Applied Energy*, vol. 257, p. 113972, Jan. 2020.

[24] J. Hu and A. V. Vasilakos, "Energy big data analytics and security: challenges and opportunities," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423-2436, Sept. 2016.

[25] J. Baek, Q. H. Vu, J. K. Liu *et al*., "A secure cloud computing based framework for big data information management of smart grid," *IEEE Transactions on Cloud Computing*, vol. 3, no. 2, pp. 233-244, Apr. 2015.

[26] K. Pan, A. Teixeira, C. D. López *et al*., "Co-simulation for cyber security analysis: data attacks against energy management system," in *Proceedings of 2017 IEEE International Conference on Smart Grid Communications*, Dresden, Germany, Oct. 2018, pp. 253-258.

[27] A. Y. Kermani, A. Abdollahi, and M. Rashidinejad, "Cyber-secure energy and flexibility scheduling of interconnected local energy networks with introducing an XGBoost-assisted false data detection and correction method," *International Journal of Electrical Power & Energy Systems*, vol. 155, p. 109683, Jan. 2024.

[28] Y. Gong, C. Chen, B. Liu *et al*., "Research on the ubiquitous electric power Internet of Things security management based on edge-cloud computing collaboration technology," in *Proceedings of 2019 IEEE Sustainable Power and Energy Conference*, Beijing, China, Nov. 2020, pp. 1997-2002.

[29] J. Zhong and X. Xiong, "Data security storage method for power distribution Internet of Things in cyber-physical energy systems," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, p. 15, Jan. 2021.

[30] Y. Liu, T. Gao, D. Niu *et al*., "Research on collaborative governance of data security in the whole life cycle of electric power manufacturing data space," in *Proceedings of 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design*, Hangzhou, China, May 2022, pp. 119-126.

[31] D. Li and Y. Gong, "The design of power grid data management system based on blockchain technology and construction of system security evaluation model," *Energy Reports*, vol. 8, pp. 466-479, Oct. 2022.

[32] Y. Li, C. Song, J. Dong *et al*., "An efficient encryption method for smart grid data based on improved CBC mode," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 9, p. 101744, Oct. 2023.

[33] J. Nechvatal, E. Barker, L. Bassham *et al*., "Report on the develop-

ment of the advanced encryption standard (AES)," *Journal of Research of the National Institute of Standards and Technology*, vol. 106, no. 3, pp. 511-577, May-Jun. 2001.
[34] S. Murphy and M. J. B. Robshaw, "Essential algebraic structure within the AES," in *Proceedings of Advances in Cryptology*, Santa Barbara, USA, Aug. 2002, pp. 1-16.
[35] W. Chen, K. Zhou, S. Yang *et al*., "Data quality of electricity consumption data in a smart grid environment," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 98-105, Aug. 2017.
[36] B. Heinrich and M. Klier, "A novel data quality metric for timeliness considering supplemental data," in *Proceedings of 17th European Conference on Information Systems*, Verona, Italy, Jun. 2009, pp. 1-6.
[37] *Guide for Monitoring Effect and Comprehensive Benefit Evaluation of Demand Response*, China Electricity Council Std. GB/T 32 127-2015, 2015
[38] C. Peng, H. Ding, Y. Zhu *et al*., "Information entropy models and privacy metrics methods for privacy protection," *Journal of Software*, vol. 27, no. 8, pp. 1891-1903, Dec. 2016.
[39] S. Zhang, Y. Wang, Y. Zhang *et al*., "Load probability density forecasting by transforming and combining quantile forecasts," *Applied Energy*, vol. 277, p. 115600, Nov. 2020.
[40] J. Daemen, V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*, *Information Security and Cryptography*. Berlin: Springer-Verlag, 2002.
[41] A. Al-Mamun, S. S. M. Rahman, T. Ahmed Shaon *et al*., "Security analysis of AES and enhancing its security by modifying S-box with an additional byte," *International Journal of Computer Networks & Communications*, vol. 9, no. 2, pp. 69-88, Mar. 2017.
[42] C.-P. Su, T.-F. Lin, C.-T. Huang *et al*., "A high-throughput low-cost AES processor," *IEEE Communications Magazine*, vol. 41, no. 12, pp. 86-91, Dec. 2003.
[43] U. Maurer, "Information-theoretic cryptography," in *Proceedings of Advances in Cryptology*, Santa Barbara, USA, Aug. 1999, pp. 47-65.

Jiabao Li received the B.E. degree from Henan University, Kaifeng, China, in 2022, and the M.S. degree from the University of Macau, Macao, China, in 2024. His research interests include smart grid security, applied cryptography, data and privacy protection, and artificial intelligence security.

Hongxun Hui received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively. From 2018 to 2019, he was a Visiting Scholar with the Advanced Research Institute, Virginia Tech, Arlington, USA, and the CURENT Center, University of Tennessee, Knoxville, USA. He is currently an Assistant Professor with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao, China. His research interests include smart grid optimization and control, demand response, power economics, carbon market, and interdisciplinary energy-environment system.

Yonghua Song received the B.E. degree from Chengdu University of Science and Technology, Chengdu, China, in 1984, and the Ph.D. degree from China Electric Power Research Institute, Beijing, China, in 1989, both in electrical engineering. In 2004, he was elected as a Fellow of the Royal Academy of Engineering, U.K.. In 2009, he was elected as the Vice President of Chinese Society for Electrical Engineering (CSEE) and appointed as the Chairman of the International Affairs Committee of the CSEE. In 2019, he was elected as a Foreign Member of the Academia Europaea. Since 2018, he has been the Rector of the University of Macau, Macao, China, and the Director of the State Key Laboratory of Internet of Things for Smart City, University of Macau. His current research interests include smart grid, electricity economics, and operation and control of power systems.

Ye Chen received the B.S. degree from North China Electric Power University, Baoding, China, in 2015. He received the master degree from Zhejiang University, Hangzhou, China, in 2018. He then worked as an Engineer at Sate Grid Jiangsu Electric Power Company Electric Power Research Institute. His research interests include application of statistical methods in electricity load and load modeling.

Tao Chen received the Ph.D. degree in electrical engineering from University of Michigan, Dearborn, USA, in 2018. He is currently an Associate Professor in School of Electrical Engineering, Southeast University, Nanjing, China. Before joining in Southeast University, he worked as a Postdoctoral Associate in Advanced Research Institute (ARI), Virginia Tech, Washington D.C., USA, from 2018 to 2019, an Intern Engineer in Global Energy Interconnection Research Institute North America (GEIRINA), California, USA, from 2017 to 2018, and Project Researcher in Tampere University of Technology, Tampere, Finland, from 2013 to 2015. His research interests include demand side management, electricity market, and machine learning application in power systems.

Pierluigi Siano received the M.Sc. degree in electronic engineering and the Ph.D. degree in information and electrical engineering from the University of Salerno, Salerno, Italy, in 2001 and 2006, respectively. He is a Full Professor of Electrical Power Systems and Scientific Director of the Smart Grids and Smart Cities Laboratory with the Department of Management & Innovation Systems, University of Salerno. He has been the Chair of the IES TC on Smart Grids. His research interests include demand response, energy management, integration of distributed energy resources into smart grids, electricity market, and planning and management of power systems.