IEEE Power & Energy Magazine

Smart Meter Data Sharing for AI-Enhanced Energy Systems

Ruiyang Yao, Jie Song, Zengxiang Li, Han Yu, Yi Wang.

Digitization is a prevailing trend in modern energy systems. With advancements in information and communications technology (ICT), advanced metering infrastructures (AMI) such as electric meters and gas meters are developed to record fine-grained energy consumption data. As a result, an increasing amount of data can now be accessed and collected, ranging from basic load, voltages, and gas readings to tamper indication and outage records, etc. With AMI, companies and operators can collect smart meter data with high accuracy and low latency. Appropriate utilization of smart meter data can result in significant benefits. By analyzing smart meter readings using data mining techniques, various functionalities can be enhanced, including behavior modeling, load forecasting, generation forecasting, etc. The installation of smart meters also enables demand response programs where utility companies can remotely modify the energy supply to meet demand variations. The increasing availability of smart meter data also boosts the development of artificial intelligence (AI), a fast-developing technology that demands a large amount of data to effectively learn and make accurate decisions. These insights can improve performance and efficiency in the power and energy systems.

However, smart meter data are not always freely available and are usually kept and possessed by different stakeholders. Factors like concerns for sensitive (or personal) information leakage and lack of incentives stop stakeholders from willingly sharing their data with others. Consumer privacy needs to be protected from unauthorized parties, and it seems that legal enforcement and regulations will be needed to govern the process. Consequently, promoting data sharing between energy sector stakeholders requires a collaborative effort between ICT, privacy-preserving techniques, and incentive design. In this paper, we discuss various privacy-preserving techniques and incentive mechanisms, as well as several present-day case studies to illustrate the application of these approaches to improve current practices.

Section 1 Data Sharing: A Boost for Modern Smart Meter Data Analysis

Smart Meter data sharing is crucial for various general purposes, such as decarbonization and grid resilience, as acquiring more data essentially gives more information about the corresponding energy consumption patterns. A basic example of this is load forecasting for accurate demand response, where the common industrial approach is to collect history load data, geospatial data, and geological data from grid operators and weather service providers in order to train a forecasting model. Beyond the single-operator case, by collecting and jointly analyzing load data from different regions, grid operators can produce a more accurate and comprehensive load profile characterization to enhance the distribution of electricity. A similar case is in renewable generation, where wind generation data from nearby regions (or power plants) provide a foundation for more accurate AI-based generation forecasting due to their correlated weather conditions. Data sharing is unquestionably one of the most important issues and is the basis for many modern advanced analytics of energy systems.

Section 1.1 Development of Metering Infrastructures

Smart meters and smart gas meters are advanced technologies that have transformed the way energy consumption is monitored and managed, and constitute an implacable role in the digitalization of energy systems. These devices provide real-time data on electricity and gas usage, enabling consumers to track their energy consumption, optimize usage patterns, and make informed decisions to reduce costs and promote energy efficiency.

By leveraging wireless communication technologies, smart meters enable the transmission of recorded data to utility companies, eliminating the need for manual meter readings. With the aid of smart meters, it is now possible to shift consumers' energy usage to off-peak hours, reducing strain on the grid during peak demand periods. One major application of smart meters is to support the integration of renewable energy sources, such as solar panels or wind turbines. By measuring both energy consumption from the grid and energy generation from these renewable sources, operators can monitor their energy self-sufficiency and export excess energy back to the grid, thus promoting a more sustainable energy system. Similarly, by utilizing smart gas meter data, individuals and organizations can estimate future gas usage, make informed operations decisions, and identify potential saving opportunities, aligning with the decarbonization goal. Another field of important applications is the quality monitoring of the supply of blending of natural gas and hydrogen in future gas distribution grids. Ensuring the hydrogen purity of the blend is the key to addressing safety concerns in hydrogen handling and distribution.

In practice, many gas meters rely on batteries or energy harvesting, which leads to the difference in the data granularity between smart meter and gas meter data. The electric meter transmits data with a standardized resolution of 15 minutes, while gas usage is typically recorded every one or three months. Therefore, it is necessary for researchers to adopt tailored strategies for analyzing high-resolution smart meter data and low-resolution gas meter data.

Section 1.2 ICT & Cloud Infrastructures

The evolution in data sharing is particularly linked to the innovation of particular types of ICT ── cloud infrastructures. Cloud infrastructure refers to a set of hardware, software, and networking resources that are used to deliver computing services over the Internet. Instead of relying on physical servers and storage devices, cloud infrastructure enables users to access virtualized resources that can be easily scaled based on demand. On top of the digitalization of energy systems, cloud infrastructures stand out to provide a scalable, flexible, and secure platform for storing, processing, and analyzing massive amounts of smart meter data. Cloud infrastructure relies on virtualization technology, which allows multiple virtual machines or containers to run on a single physical server. This enables efficient utilization of hardware resources and allows for the quick creation and deployment of new instances for smart meter data analysis and processing. Moreover, cloud infrastructure is often built on a distributed architecture, meaning that resources are spread across multiple data centers and geographic locations, which coincides with the wide distribution of smart meters.

Section 1.3 Edge Infrastructures and Edge Intelligence

Edge infrastructure refers to computing and storage resources that are deployed at the edge of networks. By processing data locally, edge infrastructures reduce the time it takes to transmit data to the cloud, improve system reliability, and lower the risk of data breaches. Based on the foundation of edge infrastructures, edge intelligence is achieved through the implementation of AI algorithms directly at the edge devices. By leveraging edge intelligence, smart meters can be utilized to bring benefits, including real-time visibility, control, and flexibility of power consumption. Edge intelligence represents an important departure from traditional centralized settings by providing lower communicational costs, enhanced privacy, and reduced latency.

Section 1.4 Overview of Smart Meter Data Sharing

The importance and value of smart meter data in energy systems are apparent. However, there are certain challenges that hinder the realization of comprehensive data sharing despite the potential benefits. The potential issues or considerations that impact smart meter data sharing can be categorized into two perspectives: privacy and security and willingness, as depicted in Table 1. In a sharing scenario, participants can be classified as either data owners or data requesters. For example, energy consumers such as households or business can be data owners to their consumption data, while energy retailers, load aggregators, or energy service providers request such data to optimize operations. It is worth noting that some stakeholders can assume both roles, like prosumers, who are both energy consumers and producers.

Stakeholders	Privacy & Security	Willingness
Data Owner	• I need to have control over possessed data • My data needs to be protected	• I am willing to trade my data for monetary rewards/services • My data needs to be fairly valued • Other reasons
Data Requester	• My received data needs to be protected • It is legal requirement to protect private information	• I am willing to pay a price to have high quality data • I need to be fairly charged • Other reasons

Table 1 The two data barriers in smart meter data sharing

- ✓ Privacy & security is intrinsically related to privacy issues and leakage risks during the data-sharing process. Privacy relates to data owners' control over how their data are processed while security relates to how to safeguard data from malicious attackers during the whole data lifecycle. Privacy and security issues can be explained in three different ways. Firstly, the smart meter data itself usually contains sensitive information about the data owners, like the ID, living address, name, consumption patterns, etc. Data owners need to have control over who and how their data are processed, which requires anonymization or other privacy-preserving data techniques. Secondly, the transmission process needs to be safeguarded from a malicious third party, as attackers can deduce the living pattern of a particular household from the leaked data. On a higher level, if data is considered a product, leaking the data to the public reduces the value, as data products can be easily replicated and distributed without cost compared with conventional goods.

- ✓ Willingness issues relate to how the data owner is incentivized. Even though the data can be transmitted with privacy protection techniques under the regularized platform, an incentive mechanism is still needed to encourage stakeholders to willingly share high-quality data. Quality is a major concern, as the data can be easily forfeited. Rewarding counterfeiters will only discourage stakeholders from sharing the original valuable data. An evaluation strategy is needed to quantify the monetary value of a dataset. Besides data monetization, in some cases, stakeholders are willing to exchange their data for services and data, or even for free. Energy consumers may willingly exchange their data with service providers to discover energy-saving opportunities. Some grid operators are willing to share their data with the public and devote to constructing open data portals to promote research, general well-being, and gain reputation.

It is important to point out that apart from the two issues mentioned above, legal requirements such as data rights confirmation is also vital for promoting sharing. Before sharing and earning profits from data, it is important to let individual persons, companies, or facility operators have control over their private information and give consent for its usage. There are different practices around the globe that regulate the confirmation of rights and the processing of data. In Canada, the Personal Information Protection and Electronic Documents Act (PIPEDA) specifies principles for the collection, use, and disclosure of personal information, and data from metering devices also falls in the range. The EU has stringent data protection laws under the General Data Protection Regulation (GDPR), which requires organizations to obtain explicit consent from individuals before collecting, processing, or sharing their personal data. One particularly captivating example is the implementation of Common European Data Spaces, which is a framework for data sharing and collaboration across different sectors, including the energy sector, to promote the free flow of data within the EU, while also ensuring privacy and security. By establishing common standards, interoperability frameworks, and data governance mechanisms, Common European Data Space achieves sharing while remaining compliant with GDPR.

Section 2 Privacy & Security

Everything has two sides. In energy systems, digitalization and cybersecurity are precisely two sides of the same coin. While wireless technology enables convenient data transfer between two parties, it also brings the issue of privacy and security to smart meter data, which requires protection through privacy-preserving technologies and scheme design.

Section 2.1 Why is Data Security a Concern?

Data security is a major concern, due to the sensitive nature of the smart meter data. Such data is labeled with the private information of the householder, such as the house address, national ID, and name, and unprotected sharing poses risks to such sensitive information. Once a malicious attacker intercepts the data during transmission, the consequences can be devastating. Some might suggest blurring the sensitive information from the transmission and only sharing the unlabeled data, which is a method known as anonymization by shifting the data order and hiding the data label. However, statistical research pointed out that even the data itself contains sensitive information like the living and working patterns of a particular household, and statistical attacks like membership inference attacks (MIA) can be carried out to verify the existence of a particular household. Figure 1 shows a simplified data-sharing process, where data is first collected by end meters, transmitted to control centers (data owners), and shared with other stakeholders (data requesters) with the aid of cloud platforms. Different attacks can be carried out at different stages.

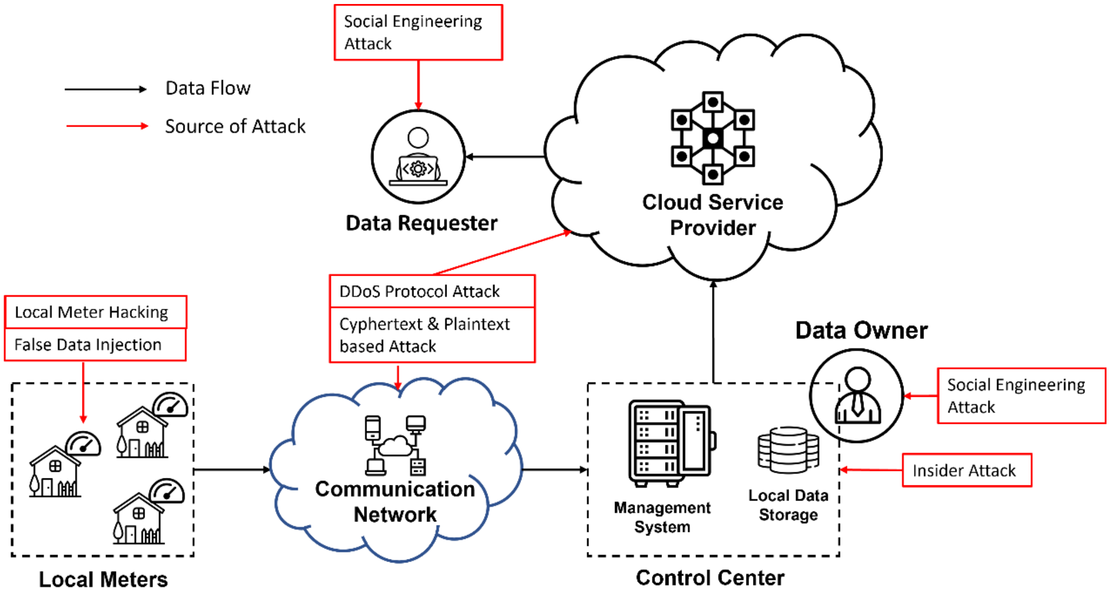
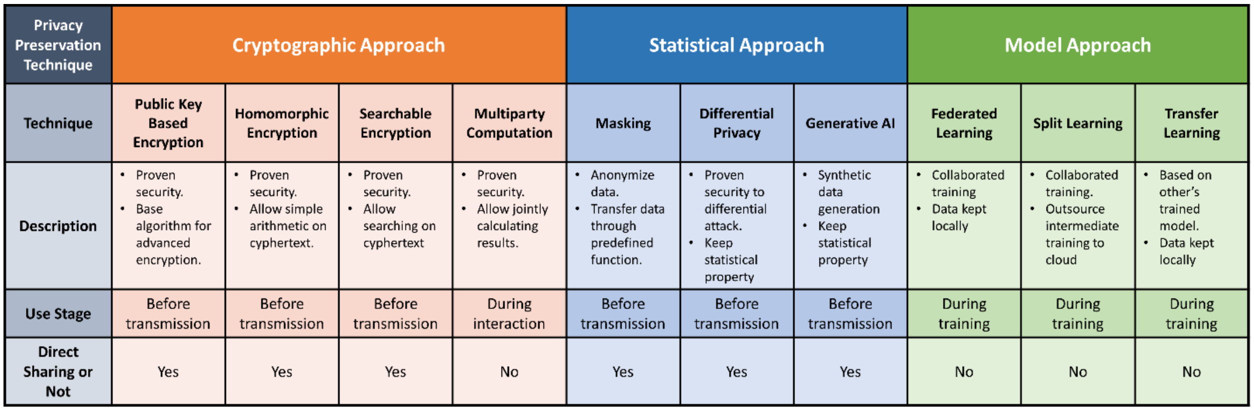


Figure 1 Flow of data and the potential attacks

Local meter hacking and false data injection can be carried out at the household's end so that meters will transmit modified false smart meter data to the operator center. This technique was once widely used to conduct electricity theft in suburban areas where meters are rarely examined. Beyond the physical level, cyphertext and plaintext-based attacks attempt to eavesdrop and intercept the transmitted data between meters and control centers, as well as between data owners and requesters, if the smart meter data is transmitted using a cloud service. DDoS protocol attacks can also be carried out to exhaust the processing ability of communication networks by using up firewall resources. Ironically, the source of data leakage is not limited to outside attacks on the networks, as reported cases of insider attacks happen in all industries beyond the energy systems. More advanced attacks, such as social engineering attack, are used to illegally obtain sensitive information from individuals in positions of trust and authority. In summary, further measures like encryption are needed to safeguard the data transmission process and allow stakeholders to reliably share the data.

Section 2.2 How to safeguard our data?

In the past decade, privacy & security issues have garnered tremendous attention, and different privacy-preserving techniques have been proposed to protect the privacy of the data through the transmission process. Three major divisions are summarized in Table 2 below, each having its own characteristics and use stage.



Privacy Preservation Technique	Cryptographic Approach				Statistical Approach			Model Approach		
Technique	Public Key Based Encryption	Homomorphic Encryption	Searchable Encryption	Multiparty Computation	Masking	Differential Privacy	Generative AI	Federated Learning	Split Learning	Transfer Learning
Description	• Proven security. • Base algorithm for advanced encryption.	• Proven security. • Allow simple arithmetic on cyphertext.	• Proven security. • Allow searching on cyphertext	• Proven security. • Allow jointly calculating results.	• Anonymize data. • Transfer data through predefined function.	• Proven security to differential attack. • Keep statistical property	• Synthetic data generation • Keep statistical property	• Collaborated training • Data kept locally	• Collaborated training. • Outsource intermediate training to cloud	• Based on other's trained model. • Data kept locally
Use Stage	Before transmission	Before transmission	Before transmission	During interaction	Before transmission	Before transmission	Before transmission	During training	During training	During training
Direct Sharing or Not	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No

Table 2 Privacy-preserving techniques for energy data sharing

Cryptographic Approaches

Encryption presents a new set of solutions to the leakage problem by converting information to cyphertext that appears to be a random, meaningless string to any unwanted third party. Encryption and cryptology form another important interdisciplinary field between mathematics and computer science. In simple words, encryption transfers the plaintext smart meter data into an encrypted form, and such form can be proven to be statistically indifferent to random strings if the third party does not have access to the key information. Encryption techniques eliminate the security issue through a mathematically rigorous framework and reduce the privacy problem into a key management problem: as long as the key is safely kept between stakeholders, then the data remains safe through the transmission process, no matter directly or through third-party cloud providers. The National Institute of Standards and Technology (US) has set out clear guidelines concerning practical and safe encryption methods, known as advanced encryption standards, which specify the block size, the complexity of the key, etc. A few variations on the encryptions exist to achieve more functionalities while keeping the data safe.

- ✓ Public key-based encryption is a type of asymmetric encryption technique that enables sharing with different entities. For example, the electricity retailer (data requester) broadcasts the public key so that any data owner who intends to transfer the data to the requester can encrypt the data according to the public key. Then, upon receipt of the encrypted data, the data requester decrypts the data locally. The public key systems solve the dilemma of traditional symmetric key encryption, which decrypts and encrypts using the same key and, therefore, is nearly impossible to scale to multiple requesters.
- ✓ Homomorphic encryption (HE) ensures the additivity of the cyphertext, sometimes even multiplicity, allowing the third parties to perform basic computations on the encrypted smart meter data. Such encryption drastically reduces the computational burden for the stakeholders in the sharing process, as encryption can be at a high cost in processing time, which is a potential issue if frequent sharing occurs. Such algebraic property showcases the effectiveness of the

application of HE in energy systems. For instance, frameworks that apply homomorphic encryption to electric meter data aggregation are proposed so that the operators will not get knowledge of the individual readings but only the aggregated consumption pattern, protecting individual privacy.

- ✓ Searchable encryption takes a big step forward by allowing users to perform searches on the encrypted dataset, but without the decryption process. Searchable encryption provides a feature that is important for the sharing platform design, by providing the possibility of uploading the data to the cloud platform in a one-time manner. Interested buyers can download the required portion without disclosing any sensitive information to the third party and with this, dramatically decreasing the communication cost.
- ✓ Secure multiparty computation achieves similar effects to homomorphic encryption by performing computation on the encrypted data through advanced cryptographic protocols, such as secret sharing, secure function evaluation, and secure multiparty protocols. By dividing data and computations across multiple parties, secure MPC ensures that no single entity has complete access to sensitive information, making it an effective solution for scenarios where privacy is paramount, such as financial transactions, healthcare data analysis, and collaborative research.

Statistical Approaches

- ✓ Masking is a technique used to ensure data privacy, in which sensitive information is obscured while the usefulness of the data is preserved. This process involves adding a layer of noise or randomization to data, making it more difficult for unauthorized parties to access or identify individual data points. Masking techniques are commonly used in data analytics, machine learning, and statistical computations, where the privacy of sensitive data must be preserved. Masking techniques are particularly useful in scenarios where data must be shared across multiple parties for analysis or decision-making, such as in medical research or financial analysis. Most publicly available datasets, like load profiles, are usually masked.
- ✓ Differential privacy (DP) is another approach that was first proposed in the field of statistics. DP refers to the inability to distinguish between the random outputs of any two adjacent datasets. A social experiment revealed that it is possible to deduce every single identity in an anonymized medical record by matching them with public datasets; this experiment disproves the effectiveness of simply masking the name information. A mathematically rigorous solution to this identity issue is the DP, which is achieved by adding a small statistical white noise to smart meter data, and the practicability of the data is unchanged. This has the potential to be combined with typical data analysis in energy systems, such as the clustering of user load profiles or the forecasting of future demand.
- ✓ Generative AI is another approach utilizing evolving AI technology to generate synthetic data that maintains the statistical properties and patterns of the original data, while also ensuring the exclusion of any personally identifiable information or sensitive attributes. This synthetic data can be safely shared without the concern of compromising individuals' private information. A classical model is a generative adversarial network, composed of a generator network to generate synthetic data samples that are similar to the original data, and a discriminator network to

distinguish between real and synthetic data. Through an iterative training process, the generator and discriminator networks compete against each other, thus improving the quality of the synthetic data generated.

Model Approaches

- ✓ Federated learning (FL) is another privacy-preserving framework that was first utilized in Google mobile app data collection. The original purpose was to generate recommendations for app users using deep generative models, which are proven to be effective in content recommendation. However, user data contains private information, just like smart meter data. FL was then introduced to allow for the training of AI models, without exchanging the local data. Note that deep learning usually requires a large amount of input data for promising model performance and FL solves the data paucity problem in a privacy-preserving way.
- ✓ Split learning is a machine learning technique that allows data to be trained without being transmitted to a central server or cloud. It works by dividing the model into two parts: one that runs on the device or edge node and another that runs on a server or cloud. The device or edge node trains the first part of the model on its local data and then sends only the updated weights to the server or cloud for aggregation with other updates. The server or cloud then trains the second part of the model on the aggregated weights and sends the updated weights back to the device or edge node, allowing the cycle to repeat. This technique enables the training of machine learning models without compromising the privacy of the data.
- ✓ Transfer learning is a machine learning technique that leverages knowledge that is gained from one task to improve the performance of another task. It involves training a model on a representative dataset and then transferring the learned features or parameters to a new model for a different but related task. In doing so, transfer learning allows the new model to benefit from the knowledge and generalizations learned from the previous task, even if the new dataset is smaller or different in nature. This approach is particularly useful when labeled data for the new task is limited or expensive to obtain.

Section 2.3 Case Study: Searchable Encryption for Smart Meter Data Sharing

Smart meter data exhibits characteristics of high volume and diverse formats, necessitating careful handling. In a typical scenario with a sampling rate of one record every 15 minutes, a large utility company deploying millions of smart meters can generate transmitted data on the scale of thousands of terabytes. Traditional local storage for smart meter data entails significant effort in terms of writing and loading when transferring or sharing the data. Establishing immobilized local storage facilities can be expensive and time-consuming, making it an impractical choice. As an alternative, cloud storage can enable efficient smart meter data sharing for utility companies. However, this approach also introduces the risk of sensitive information leakage. To mitigate this risk, it is crucial to encrypt the data before uploading it to the cloud. Conventional encryption methods often require downloading and decrypting the entire dataset to extract specific information, leading to inefficiency and unnecessary processing. To address this challenge, more advanced techniques, such as searchable encryption, can be employed. Searchable encryption allows for encrypting only the necessary portions of the data, reducing the computational burden and optimizing data access. This approach ensures that sensitive information remains protected while enabling streamlined and targeted access to the required data, improving overall data management in smart grid systems.

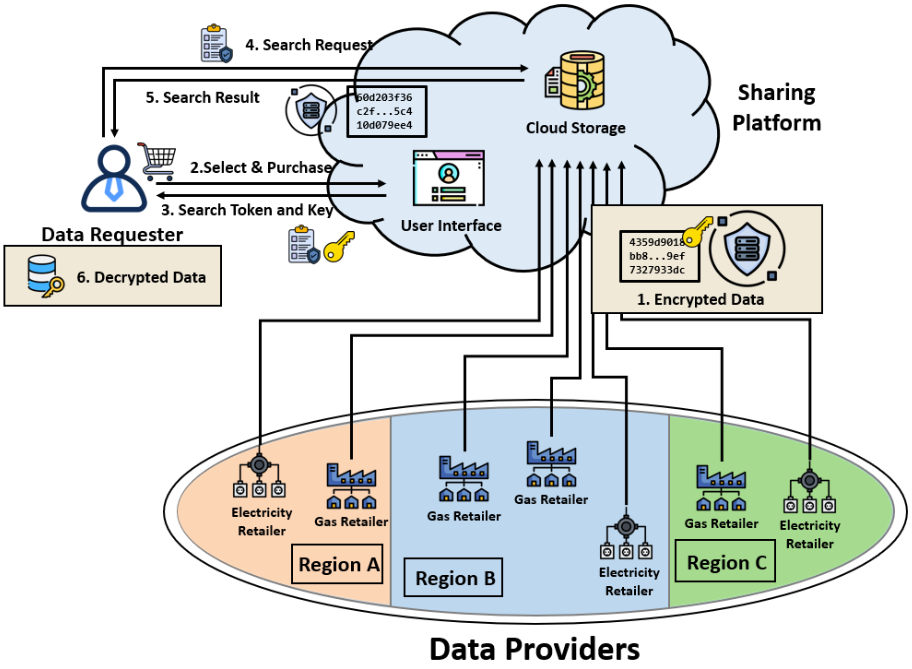


Figure 2 Searchable encryption-based data sharing framework among retailers

Figure 2 shows the implemented framework designed for smart meter data sharing among electricity and gas retailers in different regions. In step 1, the retailers encrypt their data using their key and upload it to the cloud server. The data requester, for example, the third-party developer, views the description of the data and performs transactions using the user interface provided by the platform. Then, in step 3, the platform returns the generated search token back to the data requester. The token is generated based on the searching range of the order, for example, "voltage angles", "voltage magnitude", and "2019-01-01" to "2019-09-01". The requester then sends the token to the cloud server, and the server performs an encrypted search. Then, the server returns the data corresponding to the token. No other data beyond the scope of the range, like "2020-01-01", will be returned. Finally, the data requester uses the key to decrypt the message locally. Throughout the process, the data remains encrypted for cloud platforms and is therefore deemed secure.

Section 2.4 Case Study: Federated Split Learning with Edge Computing

FL is proven to be effective by enabling implicit sharing through joint model training using smart meter data while preserving privacy by keeping the data local. While conventional establishments promote sharing among local servers, a more advanced and computationally efficient setup is proposed to implement the sharing model at end meters.

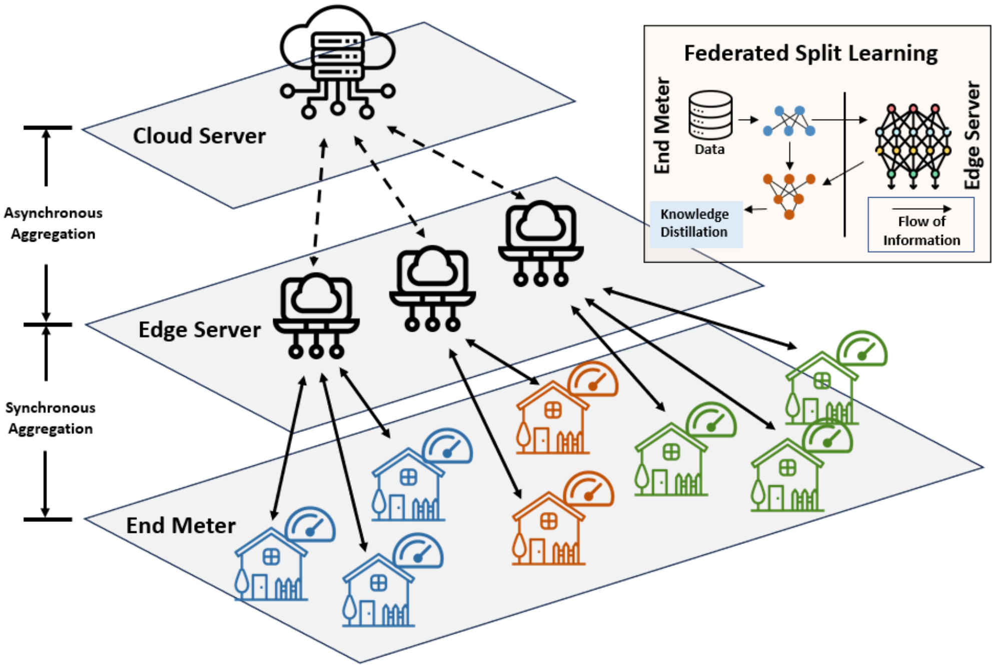


Figure 3 Federated split learning framework

Such construction is realized on local urban metering infrastructures with the framework shown in Figure 3. With this, each end meter aims to construct an artificial neural network (ANN) for load forecasting. To overcome the computational constraint in end meters and the paucity of data for individual households, FL is deployed to enable training between end meters to improve performance. Knowledge distillation is used to accommodate meters with different ANN architectures. At a higher level, edge computing is deployed by outsourcing the computationally intensive model training at the edge server without collecting and transmitting the data to the central server. Observing the restricted processing capacity at the end meter, split learning is used to further outsource the backpropagation of the model from the end meter to the edge. During each iteration, the model is initialized using the local data, then the model is sent to the edge, and the training is performed at the edge level.

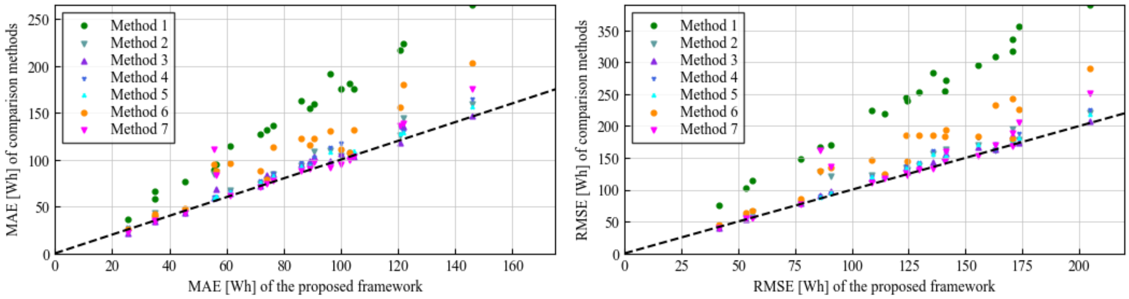


Figure 4 Performance of federated framework on different models

Evidence shows the effectiveness of the proposed federative framework. As shown in Figure 4, while each dot represents a specific model, the y-axis shows the accuracy trained under the local framework, and x-axis is trained under the FL framework. The dotted line on the diagonal indicates the equal performance line, while points above the diagonal represent an improvement in the forecasting accuracy. Different colors represent different base ANN models. Most models experience an improvement in accuracy, and it is safe to conclude that the FL is proven effective under most forecasting scenarios in our experiments. The choice of the base model is not restricted to the FL framework, and it is possible to apply FL to other machine learning methods, such as long- and short-term memory networks or tree-based methods.

Section 3 Incentivizing Smart Meter Data Sharing

Promoting and incentivizing smart meter data sharing is a complex endeavor that requires a collaborative effort between academia, government entities, and legal experts. For example, offering financial rewards or compensation can encourage participation. This could be in the form of monetary compensation, reduced energy costs, or access to exclusive services or benefits. Providing value-added services like personalized energy efficiency recommendations can also encourage consumers to share their data. On a higher level, sharing can also be promoted by creating opportunities for collaboration and knowledge sharing among stakeholders.

Section 3.1 Data Monetization

Data monetization refers to the process of generating revenue from data assets. Data monetization incentivizes smart meter data sharing by giving financial rewards to data owners. To facilitate this process, a data valuation method is needed to fairly price the smart meter data. Requesters of data need to evaluate the data before making a purchase decision to ensure it meets their requirements, and the data provider demands fair compensation and rewards. The valuation method is highly dependent on the specific use case it serves. Different purposes, such as load forecasting, deriving sensitivity indices, or optimizing energy distribution, can lead to varying monetary values for the data. Conventional data valuation strategies include a usage-based pricing strategy, which calculates prices based on the actual or anticipated usage of smart meter data products and services, making it more suitable for consumers with low data usage. Recently, more advanced pricing and valuation strategies tailored to the energy industry have been developed. One example is the evaluation in load forecasting, where the pricing of data is linked to predictive accuracy, as well as the quantity of the data. Such a strategy encourages the contribution of high-quality load data and reduces the risk of agents forfeiting smart meter data.

Based on valuation strategies, different clearing mechanism is utilized to settle the transaction of data products. A classical approach is the cooperative game theory, where the Shapley values are used to identify each data owner's contribution fairly. However, Shapley value has scalability problems when the number of participants increases and a more delicate solution is needed for that case. Another practice is non-cooperative game where data requesters and owners possess a conflict of interest. Nash equilibrium represents a condition where no participant can improve their profit by changing their strategies individually and can be used to model the market behavior of a non-cooperative game setting. However, we can't assume that all participants make pricing decisions simultaneously and act according to their own interests in the real world. Cooperative game-based market clearing would be a more sensible choice. Advances in valuation theory also bring clearing mechanisms that are based on pure statistical concepts. For example, a regression market is developed for monetary distribution where multiple agents jointly train a regression-based energy forecasting model. Such schemes rely on interpretable machine learning and cooperative game theory to fairly distribute rewards.

Section 3.2 Other Incentives

Data monetization is, indeed, an important incentive mechanism to encourage data sharing in energy systems. However, it is not the only approach. In some cases, grid operators and other stakeholders in the energy sector recognize the value of data sharing for research, innovation, and the overall improvement of the energy system. As a result, they are willing to share data for free and even construct open data portals to facilitate access to this information.

Meanwhile, data owners in energy systems are increasingly open to exchanging their data for data and services. This exchange allows them to access additional data sources and specialized services that can enhance their decision-making processes and optimize their operations. For instance, a smart home system collects real-time data on energy consumption, weather conditions, and occupancy patterns. The energy consumer can exchange this data with their utility provider, who can analyze it to provide personalized recommendations for optimizing energy usage. The utility provider, in turn, can exchange aggregated data from various consumers with energy researchers or grid operators to improve load forecasting and grid planning.

Section 3.3 Case Study: Incentive Design for Gas Usage Estimation

Gas consumption estimation and forecasting are essential for power generation and delivery operations. Accurate estimation facilitates efficient resource planning and budgeting. However, the federative nature brings challenges in quantifying the contribution of each participant, which is necessary for constructing a fair incentive mechanism. 'HiFi-gas', a hierarchical FL incentive mechanism, is proposed to quantify and distribute the monetary rewards to participants in the learning framework.

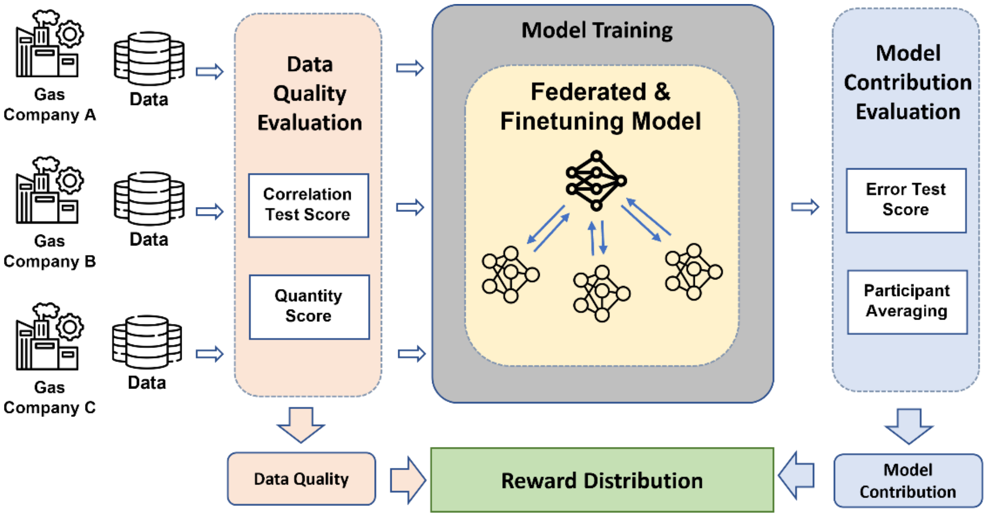


Figure 5 Incentive mechanism for federated gas usage estimation

Figure 5 shows the reward distribution flowchart of the FL process. To begin with, data quality evaluation is carried out by appealing to the correlation and quantity test scores, thus quantifying the data quality. In the next stage, the local gas company participates in the FL learning process by iteratively performing federation and fine-tuning the local model, which is a convolutional neural network. Finally, model contribution evaluation is carried out by comparing the symmetric mean absolute percentage error, and then a marginal contribution is calculated. Finally, the reward for data quality and model contribution will be specified and distributed back to the participants.

Starting in December 2022, HiFi-Gas was adopted by two major gas companies at the city level. These companies are referred to as Company A and Company B for confidentiality purposes. Company A has four affiliated heating stations, while Company B has 47 affiliated heating stations. The main goal of HiFi-Gas is to improve the accuracy of gas usage estimation models for both gas companies and their affiliated heating stations within the FL ecosystem. The results show significant improvements in the accuracy of the FL gas usage estimation model, which was collaboratively trained by Companies A and B under HiFi-Gas. For Company A and its affiliated heating stations, the average gas usage estimation accuracy has increased by 14.67%, with the model accuracy improving from 81.1% to 93.0% after the deployment of HiFi-Gas. Similarly, for Company B and its affiliated heating stations, the average gas usage estimation accuracy has improved by 10.31%, with the model accuracy increasing from 78.6% to 86.7%.

The mechanism incentivizes participating entities to carefully refine their local datasets and remove noisy and erroneous data prior to engaging in federated model training. This approach helps improve training accuracy and also leads to significant cost savings in gas procurement.

Section 3.4 Case Study: Center of Energy Big Data

To promote sharing between large stakeholders in urban energy systems, a project named Center of Energy Big Data is carried out at the pilot city. The aim is to unite the government, utility companies, and third-party developers in a unified platform. The detailed operational diagram is shown in Figure 6.

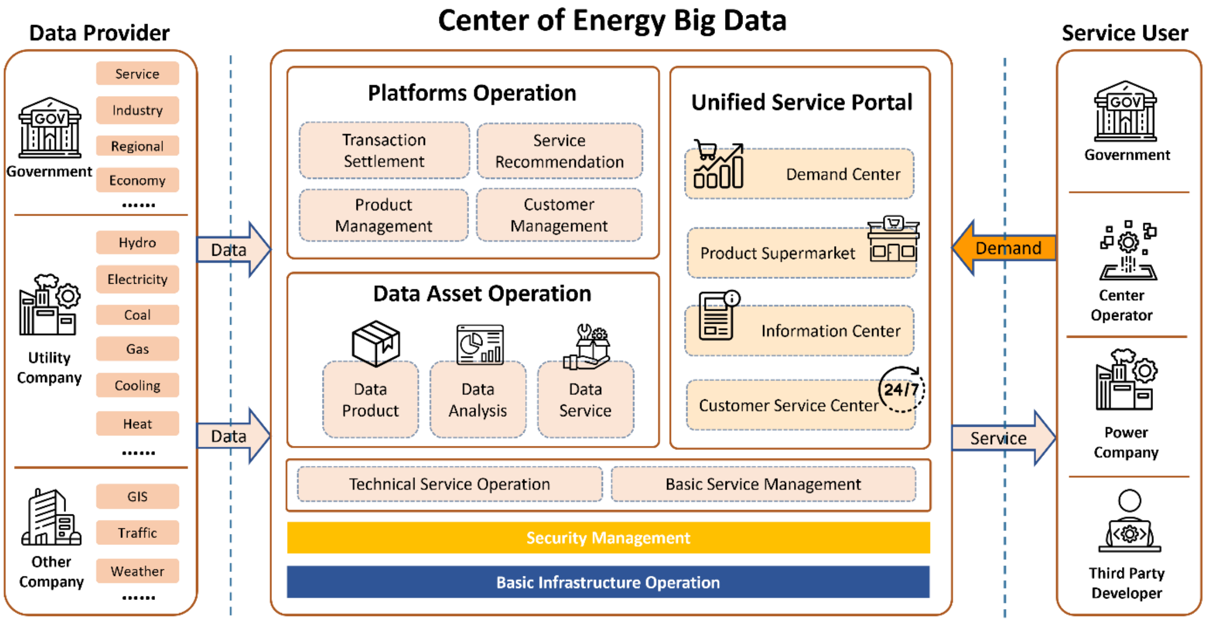


Figure 6 Center of energy big data

Basic platform operations include transaction settlement, service recommendation, product management, and customer management. Transaction settlement refers to the process of ensuring that payments are processed and settled accurately, efficiently, and securely. Service recommendation involves suggesting relevant services or data products to data requesters based on their preferences and needs. Product management involves the management of products offered on the platform and includes inventory, pricing, and availability. Customer management refers to the process of interacting with requesters, addressing their concerns, and ensuring their satisfaction with the platform's services. Basic data asset-related operations, including data product, data analysis, and data service, are carried out to help organizations leverage their data assets and gain a competitive advantage by making data-driven decisions and developing data-driven products and services. Unified service portals modularize the function into different service centers to provide a range of services.

On top of the basic operations illustrated above, security management and basic infrastructure operations are performed. Security management involves implementing measures to protect the platform and its users from cyber threats, unauthorized access, and data breaches. Basic infrastructure operations focus on managing the underlying infrastructure that supports the platform, such as servers, networks, and databases. This ensures the platform's stability, scalability, and performance, thus enabling it to handle user traffic and data both effectively and efficiently. The platform is tested on a city-wide collaboration platform, and a pilot run is carried out by a major national grid company that evaluates energy data according to the framework proposed.

Conclusions

The advancement in sensors, meters, and ICT brings a substantial amount of smart meter data for disposal. Data sharing plays an essential role in combining and harnessing such data from heterogeneous sources by empowering many advanced applications in energy systems, such as DR and grid transmission and distribution optimization. However, promoting data sharing is subject to developing privacy and security-enhancing techniques to safeguard data and constructing incentive mechanisms to increase willingness to share. This article highlights those issues by first reviewing relevant techniques, followed by detailed case studies on real-world applications. These applications promote sharing through privacy-preserving learning frameworks, reliable digital platforms, and delicate valuation and incentive mechanisms. Further investigation is necessary to establish an optimal combination of ICT, privacy-preserving techniques, and incentive design that can effectively unlock the full potential of data sharing in the energy system.

For Further Reading

- D. Qin, C. Wang, Q. Wen, W. Chen, L. Sun, and Y. Wang, "Personalized Federated DARTS for Electricity Load Forecasting of Individual Buildings," in *IEEE Transactions on Smart Grid*, vol. 14, no. 6, pp. 4888-4901, Nov. 2023, doi: 10.1109/TSG.2023.3253855.
- Li, Yang, et al. "Detection of false data injection attacks in smart grid: A secure federated deep learning approach." *IEEE Transactions on Smart Grid* 13.6 (2022): 4862-4872.
- H. Sun, X. Tang, C. Yang, Z. Yu, X. Wang, Q. Ding, Z. Li & H. Yu, "HiFi-Gas: Hierarchical Federated Learning Incentive Mechanism Enhanced Gas Usage Estimation," in *Proceedings of the 36th Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-24)*, AAAI Innovative Application of AI Award, 2024.
- Pinson, P., Han, L. & Kazempour, J. Regression markets and application to energy forecasting. *TOP* 30, 533–573 (2022). https://doi.org/10.1007/s11750-022-00631-7.
- Mazzi, Nicoló, and Pierre Pinson. "Wind power in electricity markets and the value of forecasting." *Renewable energy forecasting*. Woodhead Publishing, 2017. 259-278.
- Yu, Mingkai, et al. "Pricing information in smart grids: A quality-based data valuation paradigm." *IEEE Transactions on Smart Grid*, 13.5 (2022): 3735-3747.

Acknowledgment

This work was supported in part by the Research Grants Council of the Hong Kong SAR (HKU 27203723) and in part by the National Natural Science Foundation of China (72241420).

Biographies

Ruiyang Yao and Yi Wang are with the University of Hong Kong, China.

Jie Song is with Peking University, China.

Zengxiang Li is with the Digital Technology Research Institute ENN Group, China.

Han Yu is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore.