Privacy Preserving User Energy Consumption Profiling: From Theory to Application

Chenbei Lu, *Graduate Student Member, IEEE*, Jingshi Cui, *Member, IEEE*, Haoxiang Wang, *Graduate Student Member, IEEE*, Hongyu Yi, and Chenye Wu, *Member, IEEE*

***Abstract*—The smart grid benefits and suffers from smart meter data. Proper use of massive data can improve energy services but may raise privacy concerns. For example, user energy consumption profiling, a classic method, can identify energy consumption patterns based on the collected load profiles from users. Thus, the privacy of these individual load profiles needs to be protected. However, most of the existing works focus on data transmission and calculation privacy, and often require additional computation, communication, or platform construction costs. In contrast, noise-injection-based data source privacy-protecting works can avoid such additional costs and provide theoretical differential privacy (DP) guarantee. This paper theoretically analyzes noise-injection-based user profiling mechanisms in terms of both privacy protection and accuracy. Specifically, we establish the privacy-accuracy trade-off. We then propose an optimal user energy consumption pattern estimation method for heterogeneous noise-injection-based data. Finally, we design a valid information ratio-based pricing scheme for noisy data that is independent of downstream tasks and easy to implement. Numerical studies based on field data confirm the effectiveness of our theoretical results.**

***Index Terms*—Privacy-preserving, user profiling, differential privacy.**

I. Introduction

THE WIDESPREAD deployment of smart meters in the residential sector generates massive amounts of real-time data that are essential for power distribution and demand-side management. These data have greatly improved consumer-oriented decision-making, such as demand response [1], electricity theft detection [2], and clustering-based user profiling [3].

Manuscript received 10 March 2023; revised 26 June 2023 and 18 August 2023; accepted 2 September 2023. Date of publication 14 September 2023; date of current version 21 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 72271213, and in part by the Shenzhen Science and Technology Program under Grant JCYJ20220530143800001, Grant RCYX20221008092927070, and Grant ZDSYS20220606100601002. Paper no. TSG-00353-2023. *(Corresponding author: Chenye Wu.)*

Chenbei Lu is with the Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China.

Jingshi Cui is with the Department of Control Science and Intelligence Engineering, Nanjing University, Nanjing 210093, Jiangsu, China.

Haoxiang Wang is with the Department of Automation, Tsinghua University, Beijing 100084, China.

Hongyu Yi and Chenye Wu are with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, Guangdong, China (e-mail: chenyewu@yeah.net).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2023.3315690.

Digital Object Identifier 10.1109/TSG.2023.3315690

Clustering-based user profiling is one of the most impactful applications that can extract energy consumption patterns from users' load profile data. These extracted patterns further facilitate the provision of customized services. While these customized services can enhance economic efficiency, the clustering-based user profiling approach necessitates a substantial amount of smart meter data from individuals, thereby giving rise to privacy concerns. To tackle this challenge, most privacy-preserving methods, like federated learning [4] and secure multi-party computation [5], focus on protecting privacy for data transmission and calculation. Although these approaches do not compromise user profiling accuracy, they are often highly customized with additional computation, communication, or hardware burdens.

Noise-injection-based privacy mechanisms [6], on the other hand, only add noise to the original data and are therefore additional resource-free, though they imprecisely profile users. Fig. 1 illustrates the noise injection mechanisms for user load profiles. Specifically, directly adopting original load profiles for user profiling will suffer from privacy leakage. The privacy mechanisms inject noises into the load profiles to provide a privacy guarantee. However, different scales of noise provide different levels of privacy guarantee, and also lead to differentiated impreciseness in user profiling. For example, a small noise injection has little impact on the user profiling accuracy, but the privacy guarantee level is also low. On the contrary, a large injected noise achieves a higher level of privacy guarantee at the cost of reduced user profiling accuracy. How to trade off such user profiling accuracy and the privacy protection level is an urgent and important issue in practice, which influences both the willingness of end users' data sharing, and the effectiveness of user profiling.

To quantify such induced impreciseness, this paper theoretically analyzes how the noise-injection-based mechanisms affect clustering-based user profiling accuracy and provides privacy-accuracy trade-off results. Moreover, we facilitate a privacy-preserving data market selling load profiles with heterogeneous privacy mechanisms. From the perspective of data demanders, we propose a systematic approach to optimally utilize the heterogeneous data in the market for user profiling. On the other hand, from the market operator's perspective, we suggest a valid information ratio-based pricing scheme for the data commodities in the market, which is both task- and data-independent and easy to implement.

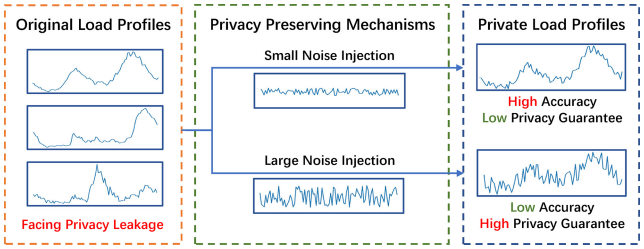


Fig. 1. Privacy Preserving for User Load Profiles.

A. Related Works

Clustering-based user profiling is a well-investigated research topic in the electricity sector. Just to name a few, McLoughlin et al. [7] combine three clustering methods: *k*-means, *k*-medoid and self-organizing maps to obtain the best profiling results based on smart meter data. Wang et al. [8] design a clustering model of consumption behavior dynamics based on fast search and density peak detection. Haben et al. [9] propose a finite mixture model-based clustering exploiting relevant energy consumption attributes of key time periods. Zhang et al. [10] design a stability index for choosing clustering algorithms and a priority index for determining the cluster's priority rank to improve the clustering performance. Kwac et al. [11] design a household electricity segmentation method to identify policy-relevant consumption behavior. Teeraratkul et al. [12] propose a shape-based load cluster method utilizing dynamic time warping to capture hidden patterns in the regular consumer behavior. Huang et al. [13] propose a federated shift-invariant dictionary learning clustering approach to enable distributed and computationally efficient user profiling. Different advanced clustering algorithms have been adopted to capture users' consumption patterns, and now the clustered user profiles have been widely applied to improve the effectiveness of power system operation from many aspects, such as electricity demand forecast [14], operation of energy sharing [15], electric vehicle charging [16], electricity price design [17], etc. In contrast, we focus on the privacy protection for user profiling.

With the growing concern over user privacy leakage, various approaches are proposed to guarantee users' data privacy and security in the electricity sector. For example, Halder et al. [18] design algorithms to enable privacy-preserving thermal inertial load management for the load serving entity. Li et al. [19] propose a privacy-preserving multi-subset data aggregation in the smart grid to ensure users' privacy. Abdallah and Shen [20] design a lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for the smart grid. Baza et al. [21] design a privacy-preserving charging-station-to-vehicle and vehicle-to-vehicle energy trading scheme to protect the privacy of EV charging information using blockchain technology. Wan et al. [22] explore the privacy-preserving fair exchange scheme for vehicle-to-grid also based on blockchain technology. Lee and Choi [23] propose a privacy-preserving energy management method in smart electric vehicle charging stations based on federated reinforcement learning. Except for the above works about general privacy preservation in the power system, only a few works specifically investigate privacy protection for user profiling. Wang et al. [3] design a federated learning approach for electricity consumption pattern extraction in a distributed way to protect user privacy. Jia et al. [24] propose a privacy-preserving distributed clustering-based user profiling based on accelerated average consensus. These two closely related works mainly focus on privacy protection during the user profile calculation process. We further this line of research by considering the noise-injection-based privacy protection for the data source of user profiling.

Most existing works focus on privacy-preserving computational procedures utilizing blockchain, homomorphic encryption, and accelerated average consensus. Although these approaches barely affect calculation accuracy, they often incur additional computational costs, agent communication costs, or platform construction costs (like blockchain systems). More importantly, these methods are often tailored to a specific task and often cannot be applied to other tasks. Privacy preservation for data sources—by directly changing the original data (i.e., by injecting noises)—is simpler and more intuitive than data transmission and computation privacy protection. It avoids additional costs and can be directly applied to different computation scenarios, but may introduce computational inaccuracy. Differential privacy (DP) [25] measures how well these methods protect privacy. The smart grid community has adopted this notion for power line obfuscation [26], non-intrusive load monitoring [27], time series data protection [28], load profile data synthesis [29] and storage control [30]. In practice, DP can be viewed as a trade-off between privacy guarantee and data accuracy. Soria-Comas et al. [31] replace conventional DP and design the individual differential privacy to guarantee the privacy for individuals and provide more accurate data. Inspired by the idea of the trade-off between privacy and accuracy, we further this line of research by providing a theoretical DP guarantee for noise-injection-based mechanisms to conduct user profiling, which helps quantify how these mechanisms affect user profiling performance.

Data valuation and pricing are enablers for the power grid digitalization, recently receiving significant attention. Just to name a few, Wang et al. [32] evaluate the value of information by improving the efficiency of solar power plant operation. Wang et al. [33] also measure the value of wind power forecasting information for improving economic dispatch performance. Yu et al. [34] design an information market framework and propose an information valuation model to help price photovoltaic-related data in power system operation problems. These works mainly focus on analyzing the value of data in improving the effectiveness of a specific power grid task. In contrast, we consider pricing the user load profile data with privacy guarantees to improve a class of downstream tasks. With the increasing public privacy concern, some recent works also examine the value of data with privacy guarantees. Ghosh and Roth [35] initiate the study of auction markets for private data in two cases: when the data analyst has a fixed accuracy goal and has a fixed budget. The results are extended and generalized by exploring approximately optimal data pricing schemes [36], [37], [38], which can be viewed as

a trade-off between performance and computational efficiency. In contrast to this line of research focusing on auction, we seek to facilitate the privacy-preserving data market selling heterogeneous noisy load profiles from both data pricing and data utilization perspectives.

B. Our Contributions

Our major contributions can be summarized as follows:

- *Privacy-Accuracy Trade-off for DP User Profiling Center Estimation:* We theoretically analyze the privacy-accuracy trade-off between the injected noise to load profiles and the estimation accuracy of the user profile center based on the resulting noisy data, which provides valuable guidelines for deciding the noise-injection level to user profiles.
- *Optimal User Profile Estimation With Noisy Data:* We design the optimal user profile estimation approach for minimizing variance and tail probability based on data perturbed by heterogeneous noise-injection-based mechanisms.
- *Price Design for DP Noisy Load Profiles:* We propose a valid information ratio-based price scheme for DP noisy load profiles that is easy to implement. We also address the advantages of our designed pricing scheme in terms of task independency and data independency.

Our paper proceeds as follows: Section II introduces user load profile privacy-preserving mechanisms. Section III theoretically describes the tension between privacy protection level and user profile estimation accuracy. Using DP load profiles with heterogeneous mechanisms, Section IV designs the optimal user profile estimation approach to minimize the estimation error. Section V proposes the pricing scheme for DP noisy data. Section VI conducts the numerical study. Finally, Section VII concludes our paper. All the necessary proofs are deferred to the Appendix.

II. Privacy-Preserving for Load Profiles

In this section, we introduce two widely adopted noise-injection-based mechanisms, i.e., the Laplace mechanism and the Gaussian mechanism, to protect the privacy of user load profiles. We then theoretically characterize the differential privacy level that these mechanisms can guarantee.

Specifically, a user load profile is denoted by

$$\boldsymbol{d} = (d_1, d_2, \ldots, d_T) \in \mathbb{R}^T, \tag{1}$$

where d_t represents the user's energy consumption at time t, and T denotes the length of the profile. To protect the privacy of load profile $\boldsymbol{d}$, we consider injecting different kinds of noises into the original load profile.

Mathematically, we denote a noise-injection-based mechanism as $\mathscr{B}: \mathbb{R}^T \rightarrow \mathbb{R}^T$, which is a mapping function from a T-dimensional vector (the original load profile) to a T-dimensional vector (the noisy load profile).

For the convenience of subsequent definitions of DP, we first define the distance metric function $dis(\boldsymbol{d}_1, \boldsymbol{d}_2)$ measuring the similarity between two load profiles $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$. Generally, the distance metric function *dis* can be the l_1 norm, i.e., $dis(\boldsymbol{d}_1, \boldsymbol{d}_2) = \|\boldsymbol{d}_1 - \boldsymbol{d}_2\|_1$, or the l_2 norm, i.e., $dis(\boldsymbol{d}_1, \boldsymbol{d}_2) = \|\boldsymbol{d}_1 - \boldsymbol{d}_2\|_2$.

We follow the classical choice [6] to define the distance metrics function *dis* separately for the Laplace mechanism and the Gaussian mechanism in the subsequent parts. Based on the distance metric function, we can define the neighbor profiles:

Definition 1 (Neighbor Profiles): For a predefined threshold $\Delta > 0$ and a given distance metric function *dis*, if two load profiles $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^T$ satisfy:

$$dis(\boldsymbol{d}_1, \boldsymbol{d}_2) \leq \Delta,$$

then $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are neighbor profiles.

In this definition, $\Delta > 0$ is a predefined sensitivity parameter. In our context, we follow the convention [6] and take $\Delta = 1$.

A. Laplace Mechanism With ϵ-DP Guarantee

We first introduce the Laplace mechanism, which injects the Laplacian noise into the load profiles. Before diving into the details of this mechanism, we introduce an important notion, ϵ-differential privacy (ϵ-DP) [25], which can rigorously characterize the level of privacy protection.

The ϵ-DP describes the inability to differentiate two similar datasets (user load profiles in our setting). Mathematically, consider a mapping function $\mathscr{B}(\boldsymbol{d})$ from a load profile $\boldsymbol{d}$ to $\mathbb{R}^T$, the ϵ-DP is defined as:

Definition 2 (ϵ-DP [25]): For any two neighbor load profiles $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^T$ with distance metric function $dis(\boldsymbol{d}_1, \boldsymbol{d}_2)$ being the l_1-norm, and any subset $Y \subseteq \mathbb{R}^T$, if there exists a constant $\epsilon > 0$, such that the mapping function $\mathscr{B}: \mathbb{R}^T \rightarrow \mathbb{R}^T$ satisfies:

$$\frac{\mathbf{Pr}\left[\mathscr{B}(\boldsymbol{d}_1) \subseteq Y\right]}{\mathbf{Pr}\left[\mathscr{B}(\boldsymbol{d}_2) \subseteq Y\right]} \leq e^{\epsilon}, \tag{2}$$

where $\mathbf{Pr}[\cdot]$ represents the probability of an event, then the mapping function (mechanism) $\mathscr{B}$ achieves ϵ-DP.

The parameter ϵ characterizes the privacy protection level of a mechanism. In the definition of ϵ-DP, a smaller ϵ indicates a lower probability of differentiating two similar load profiles. Intuitively, it means that it is more difficult to identify a user through the load profiles, which guarantees a higher privacy protection level.

We provide a more specific interpretation of ϵ as follows. For any two neighbor load profiles with noise-injection-based mechanisms under ϵ-DP guarantee, the maximal probability P of successfully distinguishing each other is as follows:

$$P = \frac{e^{\epsilon}}{1 + e^{\epsilon}}. \tag{3}$$

This probability P is within $[\frac{1}{2}, 1)$. Specifically, as ϵ increases, the probability P approaches 1, which indicates a very high probability of accurately distinguishing each other, and thus a low privacy protection level. In contrast, when ϵ is very small (approaching 0), the probability P approaches $\frac{1}{2}$, which is close to the random guess, indicating a very high privacy protection level.

The Laplace mechanism is a standard approach to achieve ϵ-DP by injecting the Laplacian noise into the original load

profiles. We provide the details of the Laplace mechanism below, and advise readers who are already familiar with the Laplace mechanism to skip this subsection and proceed directly to Section II-B.

Lemma 1 (Laplace Mechanism, [6, Th. 3.6]): For any given load profile $\boldsymbol{d} \in \mathbb{R}^T$, if a mechanism $\mathscr{B}$ satisfies:

$$\mathscr{B}(\boldsymbol{d}) = \boldsymbol{d} + \mathcal{L}(\lambda), \tag{4}$$

where $\mathcal{L}(\lambda) = (X_1, X_2, \ldots, X_T) \in \mathbb{R}^T$ is a T-dimensional vector with each entry X_t, $t \in \{1, 2, \ldots, T\}$, being *i.i.d.* Laplacian random variable, and the probability density function (*pdf*) $h_L(x)$ of each X_t is characterized by parameter $\lambda > 0$ as follows:

$$h_L(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \tag{5}$$

then $\mathscr{B}$ achieves $\frac{1}{\lambda}$-DP.[1]

This theorem indicates that, with a larger λ, the Laplacian noise becomes larger, offering a privacy protection guarantee with a higher level.

Based on this theorem, we submit that, to achieve a desired privacy protection level ϵ, the parameter λ of the Laplacian noise should satisfy $\lambda = \frac{1}{\epsilon}$.

For the convenience of mathematical expression, we use $\mathscr{B}_L(\lambda)$ to denote a Laplace mechanism with parameter λ.

B. Gaussian Mechanism With (ϵ,δ)-DP Guarantee

In practice, pure ϵ-DP in Definition 2 is often too strict, and only very limited mechanisms can achieve ϵ-DP. A more general differential privacy notion is (ϵ, δ)-differential privacy ((ϵ, δ)-DP) with the following definition:

Definition 3 ((ϵ, δ)-DP [25]): For any two neighbor load profiles $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^T$ with distance metric function $dis(\boldsymbol{d}_1, \boldsymbol{d}_2)$ being the l_2 norm, and any subset $Y \subseteq \mathbb{R}^T$, if there exists constants ϵ, $\delta > 0$, such that the mapping function $\mathscr{B} : \mathbb{R}^T \to \mathbb{R}^T$ satisfies:

$$\mathbf{Pr}\big[\mathscr{B}(\boldsymbol{d}_1) \subseteq Y\big] \leq e^{\epsilon} \mathbf{Pr}\big[\mathscr{B}(\boldsymbol{d}_2) \subseteq Y\big] + \delta, \tag{6}$$

where $\mathbf{Pr}[\cdot]$ represents the probability of an event, then the mechanism $\mathscr{B}$ achieves (ϵ, δ)-DP.

Compared with the definition of ϵ-DP, the requirement of (ϵ, δ)-DP is relatively lower with a relax ratio δ. A smaller δ indicates a stronger privacy guarantee, and when $\delta = 0$, the (ϵ, δ)-DP is specified to pure ϵ-DP. The definition of (ϵ, δ)-DP enables more privacy protection mechanisms, e.g., the Gaussian mechanism. We provide the details of the Gaussian mechanism below, and again advise readers who are already familiar with the Gaussian mechanism to skip this subsection and proceed directly to Section II-C.

Lemma 2 (Gaussian Mechanism, [6, Th. A.1]): For any given load profile $\boldsymbol{d} \in \mathbb{R}^T$, if a mechanism $\mathscr{B}$ satisfies:

$$\mathscr{B}(\boldsymbol{d}) = \boldsymbol{d} + \boldsymbol{G}(\sigma), \tag{7}$$

[1]Actually, the specific form of privacy protection level ϵ is $\frac{\Delta}{\lambda}$ and is correlated to the sensitivity factor Δ. For the simplicity of expression, we adopt the common choices [6] to set $\Delta = 1$. Note that the value of Δ is a constant and does not affect the analysis in this paper.

TABLE I
APPLICATION SCENARIOS OF NOISE-INJECTION-BASED MECHANISM

	More Applicable	Less Applicable
Scenarios	Demand Response [40], DSO Economic Operation [41], User Profiling [8], Retailer Price Design [42]	Power Billing [43], Load Forecasting [44], Electricity Theft Detection [45]

where $\boldsymbol{G}(\sigma) = (X_1, X_2, \ldots, X_T) \in \mathbb{R}^T$ is a T-dimensional vector with each entry X_t, $t \in \{1, 2, \ldots, T\}$, being *i.i.d.* zero-mean Gaussian random variable, and the probability density function (*pdf*) $h_N(x)$ of each X_t is characterized by parameter $\sigma > 0$ as follows:

$$h_N(x) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{8}$$

then $\mathscr{B}$ achieves $\sqrt{1.25/\delta}\sigma^{-1}, \delta)$-DP for any $\delta > 0$.

Note that, δ is a predefined parameter according to our need. A larger σ leads to a smaller $\sigma^{-1}\sqrt{1.25/\delta}$, indicating a higher privacy protection level.

Based on this theorem, we can derive that, to achieve an (ϵ, δ)-DP, the injected Gaussian noise should be with a standard deviation $\sigma = \epsilon^{-1}\sqrt{1.25/\delta}$.

For the convenience of mathematical expression, we use $\mathscr{B}_G(\sigma)$ to denote a Gaussian mechanism with standard deviation parameter σ.

Remark: Note that, in practice, our designed noise-injection mechanism is conducted when the user load profile data are shared with the third-party data demanders (instead of the data collection stage). It guarantees the user privacy protection during all the subsequent processes, including data transmission, storage and computation. Since the data are noise-free during collection, our noise-injection-based mechanism does not influence the original smart meter services for end users, like power billing.

C. Application Scenarios

Although with DP guarantees, the noise-injection-based mechanisms will introduce certain noises to the original load profiles, which is unsuitable for scenarios with paramount data accuracy requirements. We summarize the applicable and inapplicable application scenarios of the noise-injection-based mechanism in Table I. The noise-injection-based mechanism has more potential for applications where accuracy is not paramount, and sacrificing a small amount of accuracy for these scenarios is welcome in exchange for a high level of privacy protection [39]. These scenarios include demand response, economic operation of distribution system operators (DSOs), user profiling, customized price design for end users, etc. In contrast, for those scenarios with strong data accuracy requirements, like power billing, load forecasting, and electricity theft detection, customized privacy protection approaches with perfect accuracy are more favorable.

III. Privacy-Accuracy Trade-Off for User Profiling

Although privacy-preserving mechanisms can provide certain DP guarantees, the injected noise will inevitably affect the accuracy of user load profiling. In this section, we theoretically characterize such impacts of privacy-preserving mechanisms on user profiling accuracy. And then, we construct the key results of the privacy-accuracy trade-off for different mechanisms.

A. Clustering With Multiple Noisy Data

Consider N original load profiles $\mathcal{D} = \{\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)}\}$ that belong to a single user type. We assume all load profiles in $\mathcal{D}$ are randomly and independently drawn from the same distribution. Each single profile is denoted by $\boldsymbol{d}^{(i)} = (d_1^{(i)}, \ldots, d_T^{(i)})$. Now we consider the scenario that all the N load profiles are protected by the noise-injection-based mechanisms. Specifically, the noisy load profiles satisfy:

$$\tilde{\boldsymbol{d}}^{(i)} = \boldsymbol{d}^{(i)} + \boldsymbol{\eta}^{(i)}, \forall i \in \{1, 2, \ldots, N\}, \tag{9}$$

where $\boldsymbol{\eta}^{(i)}$ denotes the vector of injected random noise, i.e., $\boldsymbol{\eta}^{(i)} = (\eta_1^{(i)}, \ldots, \eta_T^{(i)})$. All the injected noises (i.e., $\eta_t^{(i)}, i = 1, \ldots, N, t = 1, \ldots, T$) for different load profiles and different t are *i.i.d.* random variables.

With these noisy load profiles, we seek to estimate the typical pattern of this user type, i.e., the cluster center. For most clustering algorithms [46], the cluster center is estimated by the average of all data belonging to this cluster. That is, the estimated cluster center $\hat{\boldsymbol{s}} = (\hat{s}_1, \ldots, \hat{s}_T)$ is calculated as follows:

$$\hat{\boldsymbol{s}} = \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{d}}^{(i)} = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{d}^{(i)} + \boldsymbol{\eta}^{(i)}\right). \tag{10}$$

Now we evaluate the accuracy of the estimated cluster center $\hat{\boldsymbol{s}}$ under different noise-injection-based mechanisms.

B. Cluster With Gaussian Mechanism

We first analyze the case with the Gaussian mechanism $\mathscr{B}_G(\sigma)$. The following fact can be derived by standard mathematical manipulation:

Fact 1: Given N randomly and independently sampled load profiles $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)} \in \mathbb{R}^T$ where $\boldsymbol{d}^{(i)} = (d_1^{(i)}, \ldots, d_T^{(i)})$, if all the data are protected by the Gaussian mechanism $\mathscr{B}_G(\sigma)$, then the mean and variance of the estimated cluster center $\hat{\boldsymbol{s}} = (\hat{s}_1, \ldots, \hat{s}_T)$ satisfy:

$$\mathbb{E}\left[\hat{s}_t\right] = \mu_t, \forall t \in \mathcal{T}, \tag{11}$$

$$\mathbf{Var}\left(\hat{s}_t\right) = \frac{1}{N}\left(\mathbf{Var}(d_t) + \sigma^2\right), \forall t \in \mathcal{T}, \tag{12}$$

where $\boldsymbol{d} = (d_1, \ldots, d_T)$ denotes the random variable characterizing the distribution of load profile samples (i.e., $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)}$) with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)$, the set $\mathcal{T} \equiv \{1, 2, \ldots, T\}$, and $\mathbf{Var}(d_t)$ denotes the variance of entry d_t.

This fact indicates that $\hat{\boldsymbol{s}}$ is an unbiased estimation to the true cluster center $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)$, but with an estimation variance. With more load profiles for estimation (larger N), the estimation variance can be effectively reduced. A smaller Gaussian noise (smaller σ^2) also contributes to a smaller estimation variance.

Except for the moment statistics, to better evaluate the estimation $\hat{\boldsymbol{s}}$, a commonly adopted metric is the tail probability [47], i.e.,

$$\mathbf{Pr}\left[||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k\right]. \tag{13}$$

The tail probability characterizes the chance that the estimation $\hat{\boldsymbol{s}}$ deviates from the mean $\boldsymbol{\mu}$ by more than a certain range k, which is an effective metric for describing the risk of large estimation error.

Before analyzing the tail probability for the estimation $\hat{\boldsymbol{s}}$, we first introduce a useful characterization as follows:

Definition 4 (Sub-Gaussian Random Variable): A univariate random variable X is a sub-Gaussian random variable if there exists a positive constant K, such that the tail probability of X satisfies:

$$\mathbf{Pr}[|X - \mathbb{E}[X]| \geq z] \leq 2\exp\left(-z^2/K^2\right), \forall z \geq 0. \tag{14}$$

Intuitively, the *pdf* of a sub-Gaussian variable has a "light tail". When the deviation z increases, the corresponding probability decreases rapidly at the rate of at least $O(e^{-z^2})$. A typical sub-Gaussian distribution is the Gaussian distribution.

We assume d_t is a sub-Gaussian random variable for each t. This is a practical assumption since $\boldsymbol{d}$ denotes the distribution of user load profiles in a cluster. Therefore, load profiles highly deviating from the cluster center will be divided into the other clusters with high probability in practice. Further, the physical power limit also makes $\boldsymbol{d}$ bounded, i.e., $\exists M > 0$ such that $\sup \|\boldsymbol{d}\|_1 \leq M$. And bounded random variables are all sub-Gaussian random variables [47]. In addition, we can easily verify that the injected Gaussian noises are sub-Gaussian.

Based on the sub-Gaussian properties of both the original load profiles and the injected noises, we can derive the following fact by characterizing the tail probability of $\hat{\boldsymbol{s}}$:

Fact 2: Given N randomly and independently sampled load profiles $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)} \in \mathbb{R}^T$ where $\boldsymbol{d}^{(i)} = (d_1^{(i)}, \ldots, d_T^{(i)})$, if all the data are protected by the Gaussian mechanism $\mathscr{B}_G(\sigma)$, then for any estimation error range $k > 0$, the tail probabilities of the estimated cluster center $\hat{\boldsymbol{s}} = (\hat{s}_1, \ldots, \hat{s}_T)$ satisfy:

$$\mathbf{Pr}\left[|\hat{s}_t - \mu_t| \geq k\right] \leq 2\exp\left(-\frac{N}{2}\frac{k^2}{\sigma^2 + \tilde{\sigma}_t^2}\right), \forall t \in \mathcal{T}, \tag{15}$$

$$\mathbf{Pr}\left[\frac{1}{T}||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k\right] \leq 2\sum_{t=1}^{T}\exp\left(-\frac{N}{2}\frac{k^2}{\sigma^2 + \tilde{\sigma}_t^2}\right), \tag{16}$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$, the parameter $\tilde{\sigma}_t^2$ is the proxy variance [47] of d_t defined as follows:

$$\tilde{\sigma}_t = \arg\min_{v} v : \mathbb{E}\left[e^{\lambda(d_t - \mu_t)}\right] \leq e^{\frac{\lambda^2 v^2}{2}}, \forall \lambda \in \mathbb{R}, \tag{17}$$

and $\boldsymbol{d} = (d_1, \ldots, d_T)$ denotes the random variable characterizing the distribution of load profile samples (i.e., $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)}$) with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)$.

This fact indicates that, when the required estimation error range k linearly increases, the tail probability decreases rapidly at the rate of $O(\exp(-k^2))$. Further, with a larger sample size N, the tail probability decreases exponentially

fast. Additionally, a larger Gaussian noise (larger σ) also contributes to a larger tail probability.

Based on the analysis of the tail probability, we can summarize the following theorem to characterize the trade-off between the privacy protection level and the estimation accuracy of the cluster center:

Theorem 1 (Privacy-Accuracy Trade-off for Gaussian Mechanism): For N randomly and independently sampled load profiles $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)} \in \mathbb{R}^T$ and any given parameters δ, $\sigma > 0$, a Gaussian mechanism $\mathscr{B}_G(\sigma)$ achieves $(\sigma^{-1}\sqrt{1.25\delta^{-1}}, \delta)$-DP, and leads to the tail probability of the cluster center estimation error at the rate of $O(T\exp(-N\sigma^{-2}))$.

We can observe that, the parameter σ bridges privacy and accuracy. Gaussian mechanism with a larger σ achieves a higher privacy protection level of $O(\sigma^{-1})$, but also brings a larger estimation error of $O(\exp(-\sigma^{-2}))$. Note that, these results can be easily extended to the other noise-injection-based mechanisms which only require the noise distributions to be sub-Gaussian.

Another important issue that we care about is the minimum required amount of data to achieve the desired estimation accuracy. Based on Fact 2, we can derive the following theorem:

Theorem 2 (Sample Complexity of Gaussian Mechanism): Given the Gaussian mechanism $\mathscr{B}_G(\sigma)$, to guarantee the estimation error smaller than k with probability $1-\tau$, the required amount of load profiles N satisfies:

$$N \geq 2\ln\left(\frac{2T}{\tau}\right)\frac{\max_{t\in\mathcal{T}}\tilde{\sigma}_t^2 + \sigma^2}{k^2}, \tag{18}$$

where T is the length of load profiles, the set $\mathcal{T} \equiv \{1, 2, \ldots, T\}$, and $\tilde{\sigma}_t^2$ denotes the proxy variance of load profiles for entry t.

This theorem indicates that when the length of load profile T increases, the amount of the required load profiles increases in $O(\ln T)$. To reduce the tail probability τ, the amount of required load profiles should increase in $O(\ln(\tau^{-1}))$. Further, when the variance σ^2 of the Gaussian noise increases, the amount of required load profiles increases linearly.

C. Cluster With Laplace Mechanism

Now we consider the Laplace mechanism $\mathscr{B}_L(\lambda)$. Similarly, we can characterize the moment statistics for the cluster center estimation under the Laplace mechanism:

Fact 3: Given N randomly and independently sampled load profiles $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)} \in \mathbb{R}^T$ where $\boldsymbol{d}^{(i)} = (d_1^{(i)}, \ldots, d_T^{(i)})$, if all the data are protected by the Laplace mechanism $\mathscr{B}(\lambda)$, then the mean and variance of the estimated cluster center $\hat{\boldsymbol{s}} = (\hat{s}_1, \ldots, \hat{s}_T)$ satisfy:

$$\mathbb{E}\big[\hat{s}_t\big] = \mu_t, \quad \forall t \in \mathcal{T}, \tag{19}$$

$$\mathbf{Var}\big(\hat{s}_t\big) = \frac{1}{N}\Big(\mathbf{Var}(d_t) + 2\lambda^2\Big), \quad \forall t \in \mathcal{T}, \tag{20}$$

where $\boldsymbol{d} = (d_1, \ldots, d_T)$ denotes the random variable characterizing the distribution of load profile samples (i.e., $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)}$) with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)$, the set $\mathcal{T} \equiv \{1, 2, \ldots, T\}$, and $\mathbf{Var}(d_t)$ denotes the variance of entry d_t.

We can observe that the estimation is also unbiased, and with an estimation variance linearly increases in λ^2.

The key difference between the Laplace mechanism and the Gaussian mechanism comes from the tail probability. Intuitively, the Laplacian distribution is not sub-Gaussian, with a heavier tail decreasing in $O(e^{-z})$ instead of $O(e^{-z^2})$. We can take advantage of this property and derive the tail probability of the estimation as follows:

Fact 4: Given N randomly and independently sampled load profiles $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)} \in \mathbb{R}^T$ where $\boldsymbol{d}^{(i)} = (d_1^{(i)}, \ldots, d_T^{(i)})$, if all the data are protected by the Laplace mechanism $\mathscr{B}_L(\lambda)$, then for any estimation error range $k > 0$, the tail probabilities of the estimated cluster center $\hat{\boldsymbol{s}} = (\hat{s}_1, \ldots, \hat{s}_T)$ satisfy:

$$\mathbf{Pr}\big[|\hat{s}_t - \mu_t| \geq k\big] \leq 2\exp\left(-\frac{N}{2}R_t\right), \forall t \in \mathcal{T}, \tag{21}$$

$$\mathbf{Pr}\left[\frac{1}{T}||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k\right] \leq 2\sum_{t=1}^{T}\exp\left(-\frac{N}{2}R_t\right), \tag{22}$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$, R_t is defined as:

$$R_t = \min\left(\frac{k^2}{4\lambda^2 + \tilde{\sigma}_t^2}, \frac{k}{\sqrt{2}\lambda}\right), \forall t \in \mathcal{T}, \tag{23}$$

and $\tilde{\sigma}_t^2$ denotes the proxy variance of load profiles for entry t.

This result is in a similar form to Fact 2: the tail probability decreases in $O(e^{-N})$ with the increasing amount of load profiles. Also, a larger λ will increase the tail probability. Based on these results, we can summarize the following theorems:

Theorem 3 (Privacy-Accuracy Trade-off for Laplace Mechanism): For N randomly and independently sampled load profiles $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(N)} \in \mathbb{R}^T$ and any given parameters $\lambda, k > 0$, a Laplace mechanism $\mathscr{B}_L(\lambda)$ achieves $\frac{1}{\lambda}$-DP, and leads to the tail probability of the cluster center estimation error at the rate of $O(T\exp(-N(\lambda k + k^2)\lambda^{-2}))$.

We can also derive the sample complexity result with the Laplace mechanism to characterize the required data amount to achieve the desired estimation accuracy:

Theorem 4 (Sample Complexity of Laplace Mechanism): Given the Laplace mechanism $\mathscr{B}_L(\lambda)$, to guarantee the cluster center estimation error smaller than k with probability $1-\tau$, the required amount of load profiles N satisfies:

$$N \geq 2\ln\left(\frac{2T}{\tau}\right)\frac{1}{\min\left(\max_{t\in\mathcal{T}}\frac{k^2}{4\lambda^2+\tilde{\sigma}_t^2}, \frac{k}{\sqrt{2}\lambda}\right)}, \tag{24}$$

where T is the length of load profiles, the set $\mathcal{T} \equiv \{1, 2, \ldots, T\}$, and $\tilde{\sigma}_t^2$ denotes the proxy variance of load profiles for entry t.

This theorem reveals a similar relationship between N and T as Theorem 2. We can also observe that, when the injected Laplacian noise and the Gaussian noise have the same variance, i.e., $\sigma^2 = 2\lambda^2$, the tail probability with the Laplace mechanism is larger than that with the Gaussian mechanism. This coincides with our intuition. Since the tail of the Laplace distribution is heavier than that of the Gaussian distribution (e^{-z} v.s. e^{-z^2}), the tail probability of Laplace mechanism-based load profiles is larger. Note that, these results can be extended to the other noise-injection-based

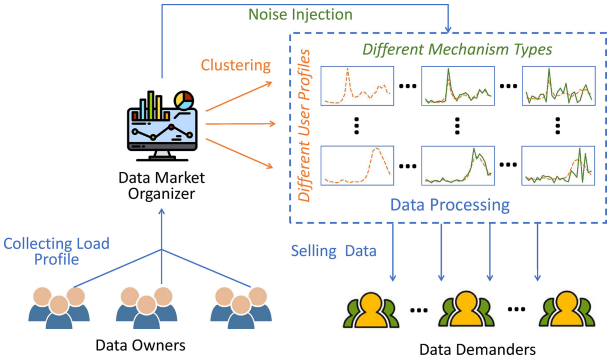


Fig. 2. Privacy-Preserving Data Market.

mechanisms with sub-exponential noise distribution [47] by minor modifications.

The analysis in this section assumes that all load profiles are protected by the same mechanism. In the next section, we will consider a more practical scenario where a data demander obtains the load profiles with heterogeneous noise-injection-based mechanisms from the data market.

IV. Optimal User Profiling With Noisy Data

In this section, we first introduce the data market for noisy load profile data trading. And then, we propose the optimal cluster center estimation approach with heterogeneous data from the data market.

A. Privacy-Preserving Data Market

Consider a data market that treats users' load profile data as commodities, which is illustrated in Fig. 2. Initially, the data market organizer collects load profiles from data owners. Then, the data are clustered into different groups characterizing different energy consumption patterns.[2] After that, the organizer injects different noises into the load profiles for privacy protection. Finally, different processed data commodities are sold to the data demanders.

Specifically, the data commodities are diverse in two dimensions, i.e., the belonging patterns and the injected noises. Denote $\mathcal{P}$ as the set of patterns. For each pattern $p \in \mathcal{P}$, the corresponding load profiles are injected by $K+1$ types of noises, i.e., type 0, 1,..., K.[3] For convenience, the type 0 noise is the zero noise, but the corresponding noise-free data cannot be directly sold as commodities in the market due to privacy concerns. We assume the type k noise mechanism injected to pattern p data is Gaussian mechanism $\mathscr{B}_G(\sigma_k)$.

[2]The clustering process is time-consuming when the data volume is very large. However, this process can be accelerated from various aspects. For example, we can first sample some data and then conduct the clustering algorithm on the sampled data. When the sample size is large enough, the clustering result can be rather accurate. Also, the clustering process can be accelerated by the decentralized implementation. We can also apply the most advanced clustering algorithm (e.g., [48], [49]) to accelerate this process, alleviating the computational burden.

[3]The key advantage of the designed privacy-preserving data market is its ability to sell heterogeneous data commodities with different levels of noises and prices, which can meet the requirement of data demanders with diverse downstream tasks.

Now consider a data demander that seeks to estimate the typical energy consumption profile of pattern p users. To achieve this target, the data demander could purchase several pattern p data with different injected noises from the data market. With $K+1$ types of available pattern p data commodities for sale, the data demander purchases N_k pieces of data with type k noise for each k. We denote the total amount of bought data as $N = \sum_{k=0}^{K} N_k$. With these data, the data demander seeks to estimate the consumption profile of pattern p users as accurately as possible.

Remark: Note that, although we mainly focus on the cluster-center estimation, in practice, due to the diversified downstream tasks that the data demanders need to accomplish, different data demanders are interested in not only the center information of user profile, but also different types of statistic information about users at both group level (group variance, correlations, distribution, etc.) and individual level (statistic features of a single piece of user load profile data). Therefore, by purchasing the data, data demanders can conduct personalized data processing based on load profiles to obtain all the statistics they want.

Essentially, the data market organizer is a trusted center (trusted server), which is a common assumption in smart gird privacy protection [50] [51] [52], and this trusted center can be the power system operator in practice. A trusted center naturally has access to the raw data and is obliged to prevent privacy leakage from all third parties. We also note that the market can also be implemented in a decentralized manner to protect user privacy better. Specifically, the clustering algorithm can be conducted [3] in a distributed manner to divide end users into different clusters, relaxing the requirement for the trusted center. Then, when a data demander requests the data from a specific cluster, the market organizer randomly chooses a load profile from a random user in this cluster. After that, the market organizer injects noises into the data and sells the noisy data to the data demander.

B. Sample Average Estimation

We first consider the simplest and most frequently adopted approach to obtain the cluster center estimation, i.e., the sample average. This approach averages all load profiles to produce the estimation. Specifically, the estimation of pattern p's cluster center $\hat{\boldsymbol{s}}$ satisfies:

$$\hat{\boldsymbol{s}} = \frac{1}{N} \sum_{k=0}^{K} \sum\nolimits_{i \in \mathcal{S}_k} \tilde{\boldsymbol{d}}^{(i)}, \tag{25}$$

where $\mathcal{S}_k$ denotes the set of indices of load profiles belonging to pattern p with mechanism type k. $\tilde{\boldsymbol{d}}^{(i)}$ denotes the noisy load profile satisfying:

$$\tilde{\boldsymbol{d}}^{(i)} = \boldsymbol{d}^{(i)} + \boldsymbol{\eta}^{(i)}, \forall i \in \mathcal{S}_k, \tag{26}$$

where $\boldsymbol{\eta}^{(i)}$ denotes the injected Gaussian noise.

Next, we analyze the estimation performance of $\hat{\boldsymbol{s}}$. Standard manipulations similar to Fact 1 yield the following results:

Fact 5: Given $K+1$ types of load profiles, i.e., type 0, 1,..., K of a specific pattern p, if the amount and injected noise mechanism for type k data are N_k and $\mathscr{B}_G(\sigma_k)$, respectively,

then the mean and variance of sample average estimation $\hat{\boldsymbol{s}} = (\hat{s}_1, \ldots, \hat{s}_T)$ satisfy:

$$\mathbb{E}\big[\hat{s}_t\big] = \mu_t, \forall t \in \mathcal{T}, \tag{27}$$

$$\mathbf{Var}\big(\hat{s}_t\big) = \frac{1}{N}\Big(\sum\nolimits_{k=0}^{K} \alpha_k\Big(\sigma_k^2 + \mathbf{Var}(d_t)\Big)\Big), \forall t \in \mathcal{T}, \tag{28}$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$ and $\alpha_k = \frac{N_k}{\sum_{k=1}^{N} N_k}$. The parameter μ_t denotes the expectation of d_t, i.e., the real pattern to estimate, and $\mathbf{Var}(d_t)$ denotes the variance of d_t.

This result directly indicates that the sample average estimation is unbiased, and the estimation variance decreases in $O(\frac{1}{N})$ with the total amount of data N. Further, we can observe that the right-hand-side term inside the braces in Eq. (28) is the weighted average variance, in which the weight α_k is proportional to the corresponding data amount with type k noise.

However, the sample average estimation will not necessarily yield the minimum variance with the given data. We illustrate this issue with a simple example: consider there are two noisy load profiles A and B with variances of 1 and 10 each. When we only use A for estimation, the estimation variance is direct 1. After including load profile B, the estimation error becomes $\frac{1}{2}(\frac{1}{2} \times 1 + \frac{1}{2} \times 10) = 2.75$, which is worse than before.

More specifically, the estimation variance will increase when including a new load profile with variance σ^2 satisfying:

$$\sigma^2 \geq \frac{2\sum_k N_k\big(\sigma_k^2 + \mathbf{Var}(d_t)\big)}{\sum_k N_k}. \tag{29}$$

That is, σ^2 is more than double of the weighted average data variance. This condition can be derived by checking the derivatives of $\mathbf{Var}(\hat{s}_t)$ with respect to the number of samples N_k. It indicates that the load profiles with large noises (about twice of average) contribute negatively to the estimation.

To improve the sample average estimation, in the subsequent analysis, we propose the optimal weighted average approach to minimize the estimation variance and the tail probability, respectively.

C. Optimal Variance-Minimization Estimation

Intuitively, the load profiles with small noises will contribute more to estimation than those with large noises. Therefore, naively averaging different data is not good enough. In contrast, offering them different weights and conducting the weighted average may yield a better estimation.

This inspires us to conduct the following weighted average estimation for different time slot t separately:

$$\hat{s}_t = \frac{1}{\sum_{k=0}^{K} w_{t,k} N_k} \sum_{k=0}^{K} \sum_{i \in \mathcal{S}_k} w_{t,k} \tilde{d}_t^{(i)}, \forall t \in \mathcal{T}, \tag{30}$$

where $w_{t,k}$ denotes the weight factor to be designed for data with type k noise at time t, and $\mathcal{T} \equiv \{1, 2, \ldots, T\}$. Compared with Eq. (25), the weighted estimation assigns different weights $w_{t,k}$ to samples with different noises.

With all weights $w_{t,k} \geq 0$, $\hat{s}_t$ is apparently an unbiased estimation. The variance satisfies:

$$\mathbf{Var}\big(\hat{s}_t\big) = \frac{\sum_{k=0}^{K} N_k w_{t,k}^2\big(\sigma_k^2 + \mathbf{Var}(d_t)\big)}{\Big(\sum_{k=0}^{K} w_{t,k} N_k\Big)^2}, \forall t \in \mathcal{T}. \tag{31}$$

This result can be derived by checking the definition of $\hat{s}_t$ in Eq. (30) and further standard mathematical manipulations. Now we can choose the optimal weights $w_{t,k}$ to minimize the variance $\mathbf{Var}(\hat{s}_t)$:

Theorem 5: Given $K+1$ types of load profiles, i.e., type $0, 1, \ldots, K$ of a specific pattern p, if the amount and injected noise mechanism for type k data are N_k and $\mathscr{B}_G(\sigma_k)$, respectively, then the optimal weights $\{w_{t,k}, t = 1, 2, \ldots, T, k = 0, 1, \ldots, K\}$ for minimizing the estimation variance satisfy:

$$w_{t,k} = \frac{1}{\mathbf{Var}(d_t) + \sigma_k^2}. \tag{32}$$

Moreover, with the optimal weights, the minimal variance satisfies:

$$\mathbf{Var}\big(\hat{s}_t\big) = \frac{1}{N} \frac{\sum_{k=0}^{K} N_k}{\sum_{k=0}^{K} \frac{N_k}{\mathbf{Var}(d_t) + \sigma_k^2}}, \forall t \in \mathcal{T}, \tag{33}$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$, and $\boldsymbol{d} = (d_1, \ldots, d_T)$ denotes the random variable characterizing the distribution of load profile samples with variance $\mathbf{Var}(d_t)$ for each entry t.

It is an interesting result, because the optimal weights are only related to $\mathbf{Var}(d_t)$ and the variance of the injected noise. Load profiles with a larger noise variance σ_k^2 lead to a smaller weight $w_{t,k}$. And they are independent of the amount of data N_k. Further, the optimal variance in Eq. (33) is essentially the weighted harmonic mean, which is always smaller than the weighted average in Eq. (28).

D. Optimal Tail-Minimization Estimation

In the previous analysis, we proposed the optimal weighted average approach to obtain the optimal estimation with minimal variance. Now we consider designing the optimal estimation with minimal tail probability.

Again, consider the following weighted average estimation, which is with the same form as Eq. (30):

$$\hat{s}_t = \frac{1}{\sum_{k=0}^{K} w_{t,k} N_k} \sum_{k=0}^{K} \sum_{i \in \mathcal{S}_k} w_{t,k} \tilde{d}_t^{(i)}, \forall t \in \mathcal{T}. \tag{34}$$

Applying the results of Fact 2, we can derive the tail probability as follows:

$$\mathbf{Pr}\big[|\hat{s}_t - \mu_t| \geq z\big] \leq 2\exp\left(\frac{-2\Big(\sum_{k=0}^{K} w_{t,k} N_k\Big)^2 z^2}{\sum_k N_k w_{t,k}^2\big(\sigma_k^2 + \tilde{\sigma}_t^2\big)}\right), \forall t \in \mathcal{T},$$

where $\tilde{\sigma}_t^2$ is the proxy variance of d_t. Similarly, we can choose the optimal weights $w_{t,k}$ to minimize the right-hand-side bound, which yields the following:

Theorem 6: Given $K+1$ types of load profiles, i.e., type $0, 1, \ldots, K$, the amount and injected noise for type k data are N_k and $\mathscr{B}_G(\sigma_k)$, respectively. The optimal weights $\{w_{t,k}, t =$

$1, 2, \ldots, T, k = 0, 1, \ldots, K\}$ for minimizing the estimation tail probability satisfy:

$$w_{t,k} = \frac{1}{\tilde{\sigma}_t^2 + \sigma_k^2}, \tag{35}$$

where $\tilde{\sigma}_t^2$ denotes the proxy variance of load profiles for entry t.

This result can be derived following the same routine as the proof for Theorem 5. Compared with the optimal weights for variance minimization, the only difference is that the variance $\mathbf{Var}(d_t)$ is replaced with the proxy variance $\tilde{\sigma}_t^2$. Note that, the analysis with the Laplace mechanism can be derived similarly.

V. Data Pricing for DP User Profiles

In this section, we design a pricing scheme for noisy user load profiles under different mechanisms. Specifically, we propose a valid information ratio-based practical pricing approach.

A. Assumptions for Differentially Private Data Pricing

Essentially, the data price can be revealed by the value of data in improving the performance of downstream tasks [53]. Data with less noise enable more accurate estimation of user profiles, and more accurate user profiles can further improve the economic revenue for different tasks, like demand response (DR) and load forecasting. Before designing the pricing scheme, we first make an assumption on the tasks:

Assumption 1: For downstream task $\mathcal{Q}$, the task utility function $J^{\mathcal{Q}}$ is a monotonic decreasing function of the user profiling estimation variance $\mathbf{Var}(\hat{\boldsymbol{s}}) = \sum_{t=1}^{T} \mathbf{Var}(\hat{s}_t)$, denoted by $J^{\mathcal{Q}}(\mathbf{Var}(\hat{\boldsymbol{s}}))$.

This assumption indicates that larger estimation error of user profiling leads to less task utility, which is consistent with our intuition. DR is an effective demand side management method based on user profiles, and we now provide an example of DR to justify the assumption. Specifically, we consider two variants of task utility:

- *Accuracy of Potential Load Level*: For DR, we are often concerned about the potential load level [40], which characterizes users' capacities of peak load shaving. Specifically, the potential load level PLL_t at time t may be of the following form:

$$PLL_t = d_t - d_t^{\text{inflx}}, \tag{36}$$

where d_t^{inflx} denotes the inflexible load at time t and is assumed to be known. The potential load level's estimation accuracy can be considered as the task utility. When we adopt the mean squared error (MSE)[4] to characterize the estimation accuracy, the task utility becomes,

$$J^{\mathcal{Q}} = -\mathbf{MSE}(PLL), \tag{37}$$

where $PLL = (PLL_1, PLL_2, \ldots, PLL_T)$. With the estimation $\hat{s}_t$ for energy consumption d_t, we can easily derive the following:

$$J^{\mathcal{Q}} = -\mathbf{MSE}(PLL) = -\mathbf{Var}(\hat{\boldsymbol{s}}), \tag{38}$$

which indicates the task utility is linear in $\mathbf{Var}(\hat{\boldsymbol{s}})$.

- *Cost of Economic Dispatch*: The shifted load of DR can contribute to reducing the cost of economic dispatch. We also adopt economic dispatch cost with DR to characterize the task utility:

$$J^{\mathcal{Q}} = -\sum_{t=1}^{T} \mathbb{E}\Big(a(D_t - PLL_t)^2 + b(D_t - PLL_t)\Big),$$

where D_t is the aggregated load at time t, a and b are the quadratic and linear cost coefficients for power generation. Injecting PLL's definition yields:

$$J^{\mathcal{Q}} = -a\mathbf{Var}(\hat{\boldsymbol{s}}) + H, \tag{39}$$

where H is a constant independent of the estimation $\hat{\boldsymbol{s}}$. We can observe that $J^{\mathcal{Q}}$ is also a function of $\mathbf{Var}(\hat{\boldsymbol{s}})$.

Further, since the load profile $\hat{\boldsymbol{s}}$ is estimated from dataset $\mathcal{D}$, we use $J^{\mathcal{Q}}(\mathbf{Var}(\hat{\boldsymbol{s}})|\mathcal{D})$ to demonstrate such a correlation. Specifically, $\mathcal{D}$ includes the data with different noise-injection-based mechanisms, i.e.,

$$\mathcal{D} = \{\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K\}, \tag{40}$$

and $\mathcal{D}_k$ denotes the data from group k protected by homogeneous mechanism with variance σ_k^2. The data amount of $\mathcal{D}_k$ is $N_k \geq 0$. Also, $\mathcal{D}_0$ is defined as the noise-free dataset.

The estimated user profiling can be applied to enhance the performances of different tasks $\mathcal{Q}$, so the overall revenue $J^*(\mathbf{Var}(\hat{\boldsymbol{s}})|\mathcal{D})$ across all tasks satisfies:

$$J^*\big(\mathbf{Var}(\hat{\boldsymbol{s}})|\mathcal{D}\big) = \sum\nolimits_{\mathcal{Q}} \beta_{\mathcal{Q}} J^{\mathcal{Q}}\big(\mathbf{Var}(\hat{\boldsymbol{s}})|\mathcal{D}\big), \tag{41}$$

where $\beta_{\mathcal{Q}}$ denotes the ratio indicating the importance of task $\mathcal{Q}$. For simplicity, we use $J^*(\mathcal{D})$ to represent $J^*(\mathbf{Var}(\hat{\boldsymbol{s}})|\mathcal{D})$.

Remark: Note that, we use the user profile $\hat{s}_t$ to represent the actual energy consumption d_t. This is because user profiles are utilized to capture the typical energy consumption patterns of end users, and are often directly utilized to represent d_t for customized services like DR [54] in practice. For DR program designers, they often do not have exact energy consumption information about the users in advance, and need to complete the design solely based on users' historical energy consumption curves. Therefore, using the user profile $\hat{s}_t$ to represent energy consumption d_t for guiding the DR program is a common choice with acceptable performance. Although there exists a certain gap between d_t and $\hat{s}_t$, we can often reduce such a gap by increasing the number of target clusters output by the clustering algorithm, which can further improve the accuracy and benefit customized services.

B. Valid Information Ratio-Based Pricing

Economically, the value of data can be characterized by the marginal revenue improvement [55], i.e., by incorporating one more piece of data. Therefore, the marginal revenue v_k by including type k data can be calculated as follows:

$$v_k = \frac{\partial J^*(\mathcal{D})}{\partial N_k}, \forall k \in \mathcal{K}, \tag{42}$$

[4] Adopting the other error metrics like mean average error (MAE) won't fundamentally influence our subsequent analysis, since we can bridge different error metrics (sometimes approximately). Hence, if MSE is a function of the variance $\mathbf{Var}(\hat{\boldsymbol{s}})$, so do the other error metrics.

where $\mathcal{K} \equiv \{0, 1, 2, \ldots, K\}$ represents the set of different data types.

The marginal revenue improvement can help derive the prices of data with different noises. Specifically, the price of a type k load profile can be calculated by:

$$C_k = C_0 \cdot \frac{v_k}{v_0} = C_0 \cdot \frac{\partial J^*(\mathcal{D})}{\partial N_k} \left(\frac{\partial J^*(\mathcal{D})}{\partial N_0} \right)^{-1}, \forall k \in \mathcal{K}, \quad (43)$$

where C_0 denotes the data price for a conventional noise-free load profile (group 0).

Before deriving the specific form of C_k, we first provide a lemma to offer the intuitions of pricing. Intuitively, including a new load profile with larger noise contributes less to reduce the estimation variance $\mathbf{Var}(\boldsymbol{s})$. The following lemma helps us to quantify such contributions by the notion of valid information ratio:

Lemma 3: To achieve a desired estimation variance σ^2_{des} at time t, the data amounts N_k of different groups should satisfy:

$$\sum\nolimits_{k \in \mathcal{K}} N_k r_{t,k} = \frac{\mathbf{Var}(d_t)}{\sigma^2_{\text{des}}}, \quad \forall t \in \mathcal{T}, \quad (44)$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$ and $\mathcal{K} \equiv \{0, 1, 2, \ldots, K\}$. The parameter $\boldsymbol{d} = (d_1, \ldots, d_T)$ denotes the random variable characterizing the distribution of load profile samples with variance $\mathbf{Var}(d_t)$ for entry t.

We also term $r_{t,k}$ the valid information ratio for type k data at time t with the following mathematical definition:

$$r_{t,k} = \frac{\mathbf{Var}(d_t)}{\mathbf{Var}(d_t) + \sigma_k^2}, \ \forall t \in \mathcal{T}, \forall k \in \mathcal{K}. \quad (45)$$

This is an important result indicating the relationship between the required data amount N_k and the valid information ratio $r_{t,k}$. We can observe that the left-hand-side term of Eq. (44) can be seen as the aggregate contribution of different groups to achieve the desired estimation variance. And the contribution of group k data is the product of data amount N_k and the valid information ratio $r_{t,k}$. In other words, the valid information ratio $r_{t,k}$ characterizes the marginal contribution of type k data.

The valid information ratio $r_{t,k}$ exactly equals the ratio between the noise-free data variance and noisy data variance. A larger injected noise σ_k^2 leads to a smaller $r_{t,k}$ and contributes less to the estimation. Based on this lemma, we can finally derive the following theorem of pricing:

Theorem 7: For any task $\mathcal{Q}$ and the corresponding overall revenue function $\bar{J}^*(\mathcal{D})$, the price C_k for type k data is independent of the dataset $\mathcal{D}$ and satisfies:

$$C_k = \frac{\sum_{t=1}^{T} r_{t,k}}{T} C_0, \forall k \in \mathcal{K}, \quad (46)$$

where $\mathcal{K} \equiv \{0, 1, 2, \ldots, K\}$ represents the set of different data types, the parameter T denotes the length of load profiles, and $r_{t,k}$ denotes the valid information ratio of type k data at time t.

This is a surprisingly simple pricing scheme. The term $\frac{1}{T}\sum_{t=1}^{T} r_{t,k}$ can be seen as the discount ratio due to the noise of type k data, which equals the average of valid information ratios $r_{t,k}$ across all T time slots. The parameter C_0 denotes the noise-free data price, and the pricing for noise-free data has been extensively studied [49]. Various pricing approaches have been proposed and empirically adopted, like fixed flat-rate pricing, usage-based pricing, application and content-based pricing [56], etc. In practice, we can follow these classical pricing approaches to determine the noise-free data price C_0. It is worth remarking that this pricing scheme has the following two nice properties:

- Independence of Task $\mathcal{Q}$: The pricing scheme does not include additional task-related information, like $\beta_{\mathcal{Q}}$ and $J^{\mathcal{Q}}$, which are difficult to obtain in practice.
- Independence of data amount N: Although $\mathcal{D}$ is included in the price definition in Eq. (43), only variance terms are retained in the final pricing scheme, without including the data amount N. This enables uniform pricing for data demanders owning different amounts of data.

In general, the marginal value of data depends on the information already available. With more available information, the marginal value of new data decreases. However, note that our pricing is not exact pricing. It is more like a promotion that investigates how the good should be discounted due to the injected noise. A key property of our pricing is that, the discount ratio is not related to the available information, and it is a constant.

Note that, Assumption 1 can be relaxed to the tail-sensitive tasks, i.e., the revenue $J^{\mathcal{Q}}$ is a function of the tail probability as $J^{\mathcal{Q}}(\mathbf{Pr}[\frac{1}{T}||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k]|\mathcal{D})$. It will lead to a similar pricing scheme, and the only difference is that the definition of valid information ratio is slightly changed to $r_{t,k} = \frac{\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2 + \sigma_k^2}$, where $\tilde{\sigma}_t^2$ denotes the proxy variance of load profiles for entry t.

VI. Numerical Study

In this section, we empirically illustrate the estimation performance of user profiling with privacy-preserving mechanisms. We also demonstrate the high consistency of our theoretical results.

In the experiments, we adopt the user electricity consumption data in Pecan Street [57] with the resolution of 15 minutes from January 1 to December 31, 2018. The dataset includes 8, 360 daily energy consumption profiles of individual residential users.

A. Clustering With Privacy-Preserving Mechanisms

We first simulate the data market organizer to conduct the k-means clustering algorithm and divide the load profiles into 24 groups. Fig. 3 characterizes the typical user energy consumption patterns and the corresponding proportions. We can observe that most patterns are diversified.

Consider a data demander to estimate the user pattern based on limited samples from the data market. Fig. 4 illustrates the estimation results with different numbers of samples and various privacy-preserving mechanisms. We use the Laplace mechanism with $\lambda = 1$. We can observe that in Fig. 4(a), when only 10 samples are utilized, the estimation error is considerablely large. In contrast, with more samples being utilized, the estimation error is significantly reduced in Fig. 4(b). Considering the original load profiles are protected by the Gaussian mechanism in Fig. 4(c), the estimation

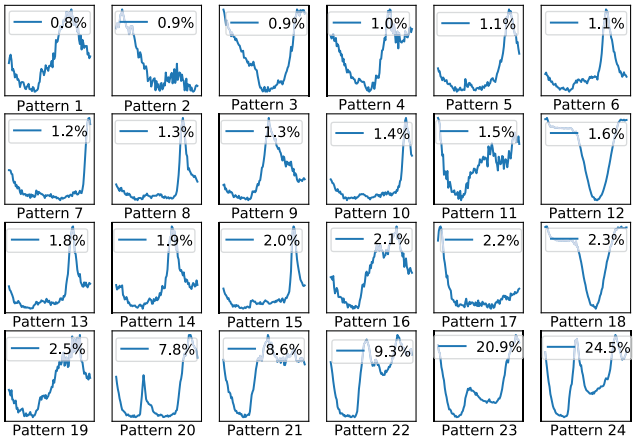
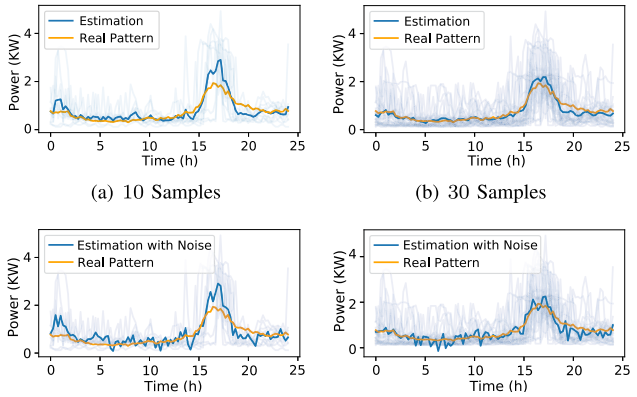


Fig. 3. Real Clustered User Profiles: (p_j:x_j) means pattern j and its proportion.



(a) 10 Samples (b) 30 Samples

(c) 10 Samples with DP Mechanism (d) 30 Samples with DP Mechanism

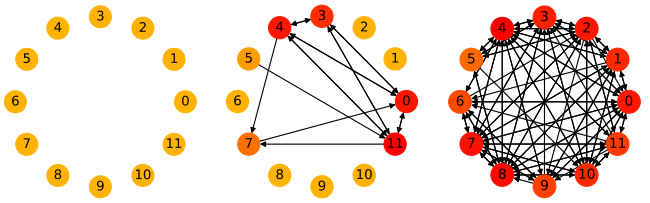
Fig. 4. Cluster with Limited Samples and DP Mechanisms.

TABLE II
PERFORMANCE OF PRIVACY-PRESERVING CLUSTERING

Sample Size	Noise-Free		Noise-Injection	
	MAE	MSE	MAE	MSE
10 samples	0.1857	0.0736	0.2709	0.1255
30 samples	**0.1050**	**0.0210**	**0.1644**	**0.0450**

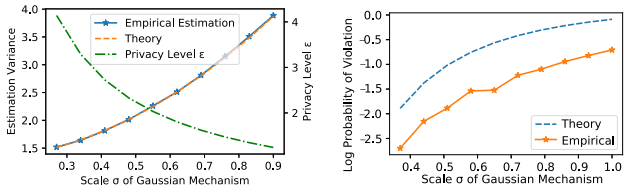
error increases compared with the noise-free case in Fig. 4(a). And more importantly, the user energy consumption patterns are certainly influenced due to the DP mechanisms. But with more samples, such an estimation error can be well contained, as illustrated in Fig. 4(d). Table II summarizes the estimation performance of different cases with two metrics: the mean absolute error (MAE) and mean squared error (MSE). Specifically, for both the noise-free and noise-injection cases, more samples (only 20 additional samples) can reduce the MAE and MSE by around 40% and 70%, respectively. In contrast, after applying the noise-injection mechanism, both the MAE and the MSE of estimation are enlarged, which is the cost to provide a certain DP guarantee.

To understand how the DP guarantee works, we provide an example of confounding among users with the privacy-preserving mechanisms in Fig. 5. Consider there are 12 users with different load profiles. When their load profiles are

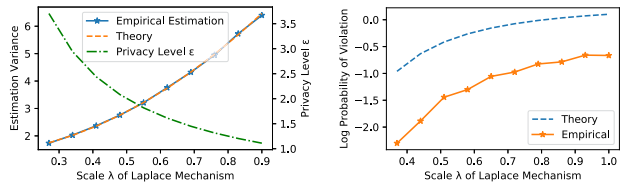


(a) $\lambda = 0, \epsilon = \infty$ (b) $\lambda = 0.8, \epsilon = 1.25$ (c) $\lambda = 1.6, \epsilon = 0.625$

Fig. 5. Privacy Guarantee: Confounding between Users.



(a) Variance with Gaussian Mechanism (b) Tail Probability with Gaussian Mechanism



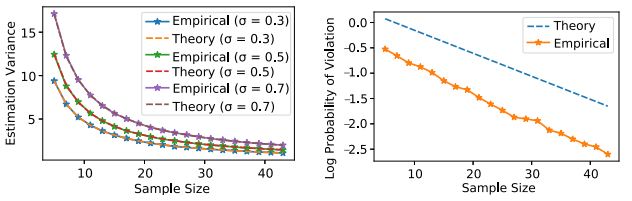
(c) Variance with Laplace Mechanism (d) Tail Probability with Laplace Mechanism

Fig. 6. Estimation Performance with Gaussian and Laplace Mechanisms.

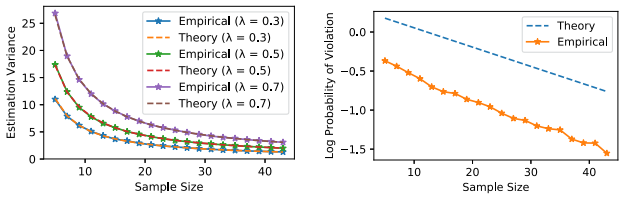
noise-free, every user can be distinguished from the others (i.e., classified into distinct clusters), which is illustrated in Fig. 5(a). After the profiles are protected by the Laplace mechanism with $\lambda = 0.8$, the privacy protection level becomes 1.25. We can observe that some users can be mistakenly identified as other types of users (like user 3 and user 4). When the magnitude λ in the Laplace mechanism increases to 1.6, the privacy protection level further improves to 0.625. Fig. 5(c) shows that almost all users are mistakenly identified as the other types of users.

B. Performance Evaluation

Now we analyze the relationship between the estimation accuracy and the magnitude of injected noise for the two mechanisms. Fig. 6 characterizes the impact of the Gaussian and Laplace mechanisms on estimation variance and tail probability. Specifically, a larger scale parameter (i.e., σ of the Gaussian mechanism and λ of the Laplace mechanism) leads to a more rapid increasing rate of the empirical estimation variance, illustrated in Figs. 6(a) and 6(c). This observation is highly consistent with our theory. For the impact on tail probability, Figs. 6(b) and 6(d) show that the tail probability increases with the growing scale parameter. A certain gap exists between the theoretical and the empirical results. This is because our result doesn't require the exact distribution of $\boldsymbol{d}$. However, the trends of the empirical and theoretical results are quite aligned with each other, which indicates that the theoretical result can fit the empirical curve well when multiplying by an empirically decided constant.

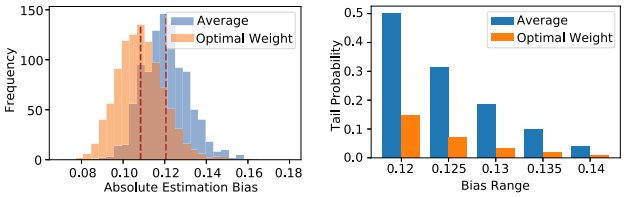


(a) Variance with Gaussian Mechanism (b) Tail Probability with Gaussian Mechanism

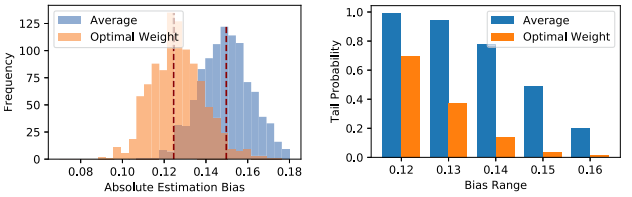


(c) Variance with Laplace Mechanism (d) Tail Probability with Laplace Mechanism

Fig. 7. Value of Data with Gaussian and Laplace Mechanisms.



(a) Distribution of Bias with Gaussian Mechanism (b) Tail Probability of Bias with Gaussian Mechanism



(c) Distribution of Bias with Laplace Mechanism (d) Tail Probability of Bias with Laplace Mechanism

Fig. 8. Performance of Optimal Weighted Clustering.

Fig. 7 reveals the value of data to different DP mechanisms. In Figs. 7(a) and 7(c), we can observe that a large sample size can significantly reduce the estimation variance for both Gaussian and Laplace mechanisms, which fits the theory well. Meanwhile, in Figs. 7(b) and 7(d), the log-scale tail probability linearly decreases with the sample size, which is also consistent with our theory on the order.

We further verify the performance improvement of the optimal weighted approach. Fig. 8 illustrates the improvement for the optimal weighted approach compared with the sample average approach. Specifically, Fig. 8(a) shows the distribution of the estimation bias under Gaussian mechanism, the optimal weighted approach can reduce the estimation bias by 10.2%. Further, for the tail probability under Gaussian mechanism, Fig. 8(b) indicates that the optimal weighted approach can significantly reduce the probability that a large estimation bias occurs, which demonstrates the effectiveness of the optimal weighted approach. Fig. 8(c) further shows that with the Laplace mechanism, the optimal weighted approach can effectively reduce the estimation bias by 15.5%. Fig. 8(d) reveals

TABLE III
PRICE MENU FOR PRIVATE DATA WITH GAUSSIAN MECHANISM ($\delta = 1$)

Privacy Level ϵ	∞	10	5	2	1	0.5
Noise Level σ	0	0.11	0.22	0.56	1.1	2.2
Data Price ($)	1.00	0.94	0.82	0.48	0.22	0.07

TABLE IV
PRICE MENU FOR PRIVATE DATA WITH LAPLACE MECHANISM

Privacy Level ϵ	∞	10	5	2	1	0.5
Noise Level λ	0	0.1	0.2	0.5	1.0	2.0
Data Price ($)	1.00	0.91	0.75	0.39	0.15	0.05

that the optimal weighted approach outperforms the average estimation for different bias ranges. When the allowed bias becomes larger, the effectiveness of our weighted approach is even more highlighted.

We provide price menus in Tables III and IV to illustrate the impact of the Gaussian and Laplace mechanisms on the price. Specifically, suppose the price of a noise-free load profile is \$1, which cannot guarantee any privacy requirement ($\epsilon = \infty$). When we require a better privacy level ϵ, the magnitude parameters (again, σ of the Gaussian noise and λ of the Laplace noise) become large, and the corresponding prices decrease at similar rates. When the required privacy level becomes 0.5, the corresponding noisy data only have $5\% - 7\%$ of the original values.

VII. CONCLUSION

In this paper, we first provide the theoretical trade-offs between the DP privacy protection level and user pattern estimation accuracy for clustering-based electricity user profiling, which provide valuable guidelines for choosing the noise-injection level for user profiling. We further implement the privacy-preserving data market, selling heterogeneous load profiles from both data utilization and data pricing perspectives. For data utilization, we propose a variance- and tail-minimization user pattern estimation approach with data protected by heterogeneous privacy-preserving mechanisms. For data pricing, we propose a valid information ratio-based price scheme for noisy load profiles.

Our work can be extended in various interesting ways. First, our analysis mainly targets at cluster center estimation with multiple samples belonging to the cluster. It is interesting to extend the current analysis to the case with samples from different clusters. Except for this, investigating the exact data pricing for noise-free data in the context of user profiling is an interesting extension. Also, it is worth considering extending our analysis to other tasks beyond user profiling.

For practical implementation, our framework may face difficulties in various aspects. For example, from the data privacy aspect, we need to guarantee the information security for the market organizer who collects all the raw data. From the market implementation aspect, a reasonable and incentive revenue-sharing mechanism is required to allocate the revenue of selling data among the market organizer and all

data providers. Designing solutions to these practical problems is crucial to the successful deployment of our proposed mechanism.

APPENDIX

A. Proof for Fact 2

We first prove the results for $\mathbf{Pr}[|\hat{s}_t - \mu_t| \geq k]$, and then combine the results for different t to derive the results for $\mathbf{Pr}[\frac{1}{T}||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k]$. We define the proxy variance [47] $\tilde{\sigma}_t^2$ for variable d_t as:

$$\tilde{\sigma}_t = \arg\min_v : \mathbb{E}\left[e^{\lambda(d_t-\mu_t)}\right] \leq e^{\frac{\lambda^2 v^2}{2}}, \forall \lambda \in \mathbb{R}. \tag{47}$$

For the Gaussian noise $\eta \sim \mathcal{N}(0, \sigma^2)$, its moment-generating function is as follows:

$$\mathbb{E}\left[e^{\lambda\eta}\right] = e^{\frac{\lambda^2\sigma^2}{2}}. \tag{48}$$

Hence, the proxy variance of η is σ^2. Combining Eqs. (48) and (47) yields the following result:

$$\mathbb{E}\left[e^{\lambda(\eta+d_t-\mu_t)}\right] \leq e^{\frac{\lambda^2(\tilde{\sigma}_t^2+\sigma^2)}{2}}, \forall \lambda \in \mathbb{R}. \tag{49}$$

The desired probability $\mathbf{Pr}[|\hat{s}_t - \mu_t| \geq k]$ satisfies:

$$\begin{aligned}&\mathbf{Pr}\left[|\hat{s}_t - \mu_t| \geq k\right]\\ &= \mathbf{Pr}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\left(d_t^{(i)} - \mu_t + \eta_t^{(i)}\right)\right| \geq k\right].\end{aligned} \tag{50}$$

Since $d_t^{(i)}$ and $\eta^{(i)}$ are all *i.i.d.*, with a given positive k, the inequality (15) in terms of $\mathbf{Pr}[|\hat{s}_t - \mu_t| \geq k]$ can be derived by applying the Hoeffding's inequality [47] to Eq. (50).

Further, the desired probability $\mathbf{Pr}[\frac{1}{T}||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k]$ can be transformed into the following form based on the union bound:

$$\begin{aligned}\mathbf{Pr}\left[\frac{1}{T}||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k\right] &= \mathbf{Pr}\left[\frac{1}{T}\sum_{t=1}^{T}|\hat{s}_t - \mu_t| \geq k\right]\\ &= 1 - \mathbf{Pr}\left[\sum_{t=1}^{T}|\hat{s}_t - \mu_t| < kT\right]\\ &\leq 1 - \prod_{t=1}^{T}\mathbf{Pr}\left[|\hat{s}_t - \mu_t| < k\right]\\ &= 1 - \prod_{t=1}^{T}\left(1 - \mathbf{Pr}\left[|\hat{s}_t - \mu_t| \geq k\right]\right).\end{aligned} \tag{51}$$

We next want to prove the following inequality:

$$1 - \prod_{t=1}^{T}\left(1 - \mathbf{Pr}\left[|\hat{s}_t - \mu_t| \geq k\right]\right) \leq \sum_{t=1}^{T}\mathbf{Pr}\left[|\hat{s}_t - \mu_t| \geq k\right]. \tag{52}$$

Denoting $1-(1-\mathbf{Pr}[|\hat{s}_t-\mu_t| \geq k])$ as a_t, the inequality (52) can be transformed into:

$$1 - \prod_{t=1}^{T} a_t \leq \sum_{t=1}^{T}(1 - a_t), \tag{53}$$

where $0 \leq a_t \leq 1, \forall t$.

Mathematical manipulation yields:

$$\sum_{t=1}^{T} a_t - \prod_{t=1}^{T} a_t \leq T - 1. \tag{54}$$

Denoting $\sum_{t=1}^{T} a_t - \prod_{t=1}^{T} a_t$ as a function $f(a_1, a_2, \ldots, a_T)$ with respect to variables $a_1, a_2, \ldots, a_T$, we can easily verify that:

$$\frac{\partial f(a_1, a_2, \ldots, a_T)}{a_t} = 1 - \prod_{i\neq t} a_i \geq 0, \forall t \in \mathcal{T}, \tag{55}$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$.

Therefore, $f(a_1, a_2, \ldots, a_T)$ is minimized when $a_t = 1$ for all t. Since $f(1, 1, \ldots, 1) = T - 1$, it directly indicates the inequality of Eq. (54) holds. Combining Eqs. (52), (51) and Eq. (50) yields our result. ■

B. Proof for Fact 4

For the Laplace noise η_t with the *pdf* $h_L(x)$ satisfying:

$$h_L(x) = \frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}}, \tag{56}$$

we can derive the moment-generating function of the Laplacian noise η_t as follows:

$$\mathbb{E}\left[e^{p\eta_t}\right] = \frac{1}{1 - p^2\lambda^2}, \forall t \in \mathcal{T}, \tag{57}$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$.

We further prove that η_t is sub-Exponential (SE) [47] with parameters $(4\lambda^2, \sqrt{2}\lambda)$, denoted by $\mathrm{SE}(4\lambda^2, \sqrt{2}\lambda)$ in short. Specifically, we need to prove the following condition:

$$\mathbb{E}\left[e^{p\eta_t}\right] = \frac{1}{1 - p^2\lambda^2} \leq e^{\frac{p^2(4\lambda^2)}{2}}, \forall 0 \leq p \leq \frac{1}{\sqrt{2}\lambda}. \tag{58}$$

Denote function $g(p)$ with variable p as follows:

$$g(p) := e^{\frac{p^2(2\lambda^2)}{2}} - \frac{1}{1 - p^2\lambda^2}. \tag{59}$$

It is apparent that $g(0) = 0$. The derivative of $g(p)$ satisfies:

$$g'(p) = 2\lambda^2 p\left(2e^{2\lambda^2p^2} - \frac{1}{\left(1 - \lambda^2p^2\right)^2}\right). \tag{60}$$

We use z to represent $p^2\lambda^2$:

$$2e^{2z} - \frac{1}{(1 - z)^2} \geq 0. \tag{61}$$

It is straightforward to check that this inequality holds for all $z \leq \frac{1}{2}$, i.e., $p \leq \frac{1}{\sqrt{2}\lambda}$. Also, by the definition of proxy variance in Eq. (17), we know d_t is also sub-Exponential with parameters $\mathrm{SE}(\tilde{\sigma}_t^2, \infty)$. By the additivity of sub-Exponential variables, $\eta_t + d_t$ is also sub-Exponential with parameters $\mathrm{SE}(\tilde{\sigma}_t^2 + 4\lambda^2, \sqrt{2}\lambda)$.

Therefore, for any given positive k, the desired probability $\mathbf{Pr}[|\hat{s}_t - \mu_t| \geq k]$ satisfies:

$$\begin{aligned}&\mathbf{Pr}\left[|\hat{s}_t - \mu_t| \geq k\right]\\ &= \mathbf{Pr}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\left(d_t^{(i)} - \mu_t + \eta_t^{(i)}\right)\right| \geq k\right], \forall t \in \mathcal{T}.\end{aligned} \tag{62}$$

Since $d_t^{(i)}$ and $\eta_t^{(i)}$ are both *i.i.d.*, applying sub-Exponential version of Bernstein's inequality [47] for the tail probability yields the result.

The probability $\mathbf{Pr}[\frac{1}{T}||\hat{\boldsymbol{s}} - \boldsymbol{\mu}||_1 \geq k]$ can be further derived following the same routine in the proof for Fact 2. This completes our proof. ■

C. Proof for Theorem 5

We denote the variance $\mathbf{Var}(\hat{s}_t)$ as a function y with variables $w_{t,k}$ for all t and k. The first-order optimality condition yields that, for all $t \in \mathcal{T}$ and $k \in \mathcal{K}$:

$$\sum\nolimits_i N_i w_{t,i}\Big(\sigma_k^2 + \mathbf{Var}(d_t)\Big)w_{t,k} = \sum\nolimits_i N_i w_{t,i}^2\Big(\sigma_i^2 + \mathbf{Var}(d_t)\Big),$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$ and $\mathcal{K} \equiv \{0, 1, 2, \ldots, K\}$.

By simplification, we can derive the following conditions:

$$w_{t,k}\Big(\sigma_k^2 + \mathbf{Var}(d_t)\Big) = \frac{\sum_i N_i w_{t,i}^2\big(\sigma_i^2 + \mathbf{Var}(d_t)\big)}{\sum_i N_i w_{t,i}}. \quad (63)$$

It indicates that $w_{t,k}(\sigma_k^2 + \mathbf{Var}(d_t))$ is constant for all k. Letting the constant be 1 directly yields our result.

Further, we can check the second-order optimality conditions hold under the optimal condition in Eq. (63). This immediately indicates the uniqueness of the solution, which completes our proof. ■

D. Proof for Lemma 3

According to Theorem 5, the specific form of the optimal variance $\mathbf{Var}(\hat{s}_t)$ given $\mathcal{D}$ satisfies:

$$\mathbf{Var}\big(\hat{s}_t\big) = \frac{1}{N}\frac{\sum_{k=0}^{K} N_k}{\sum_{k=0}^{K}\frac{N_k}{\mathbf{Var}(d_t)+\sigma_k^2}}, \forall t \in \mathcal{T}, \quad (64)$$

where $\mathcal{T} \equiv \{1, 2, \ldots, T\}$.

We define the valid information ratio $r_{t,k}$ as:

$$r_{t,k} = \frac{\mathbf{Var}(d_t)}{\mathbf{Var}(d_t) + \sigma_k^2}. \quad (65)$$

Plugging the definition of the valid information ratio $r_{t,k}$ into Eq. (64) yields the following:

$$\mathbf{Var}\big(\hat{s}_t\big) = \frac{1}{\sum_{k=0}^{K}\frac{N_k r_{t,k}}{\mathbf{Var}(d_t)}}, \forall t \in \mathcal{T}. \quad (66)$$

Letting $\mathbf{Var}(\hat{s}_t)$ to be the desired variance level σ_{des}^2 directly yields our results. ■

E. Proof for Theorem 7

The price of differentially private data of type k has the following definition:

$$C_k = C_0 \cdot \frac{\partial J^*(\mathcal{D})}{\partial N_k}\left(\frac{\partial J^*(\mathcal{D})}{\partial N_0}\right)^{-1}. \quad (67)$$

The data price formulas can be reformulated into the following form by the chain rule:

$$\frac{\partial J^*(\mathcal{D})}{\partial N_k} = \sum_{\mathcal{Q}} \beta_{\mathcal{Q}} \frac{\partial J^{\mathcal{Q}}(\mathcal{D})}{\partial \mathbf{Var}\big(\hat{\boldsymbol{s}}\big)} \sum_{t=1}^{T} \frac{\partial \mathbf{Var}\big(\hat{\boldsymbol{s}}\big)}{\partial \mathbf{Var}\big(\hat{s}_t\big)} \cdot \frac{\partial \mathbf{Var}\big(\hat{s}_t\big)}{\partial N_k}, \quad (68)$$

$$\frac{\partial J^*(\mathcal{D})}{\partial N_0} = \sum_{\mathcal{Q}} \beta_{\mathcal{Q}} \frac{\partial J^{\mathcal{Q}}(\mathcal{D})}{\partial \mathbf{Var}\big(\hat{\boldsymbol{s}}\big)} \sum_{t=1}^{T} \frac{\partial \mathbf{Var}\big(\hat{\boldsymbol{s}}\big)}{\partial \mathbf{Var}\big(\hat{s}_t\big)} \cdot \frac{\partial \mathbf{Var}\big(\hat{s}_t\big)}{\partial N_0}. \quad (69)$$

Further, taking derivatives on Eq. (44) in Lemma 3 yields the following condition:

$$\frac{\partial \mathbf{Var}\big(\hat{s}_t\big)}{\partial N_k} = r_{t,k} \cdot \frac{\partial \mathbf{Var}\big(\hat{s}_t\big)}{\partial N_0}. \quad (70)$$

Given the fact that $\mathbf{Var}(\hat{\boldsymbol{s}}) = \sum_{t=1}^{T} \mathbf{Var}(\hat{s}_t)$, combining conditions in Eqs. (70), (68) and (69) yields our results. ■

REFERENCES

[1] F. Wang et al., "Household profile identification for behavioral demand response: A semi-supervised learning approach using smart meter data," *Energy*, vol. 238, Jan. 2022, Art. no. 121728.

[2] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in ami using customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2016.

[3] Y. Wang, M. Jia, N. Gao, L. V. Krannichfeldt, M. Sun, and G. Hug, "Federated clustering for electricity consumption pattern extraction," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2425–2439, May 2022.

[4] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[5] R. Canetti, U. Feige, O. Goldreich, and M. Naor, "Adaptively secure multi-party computation," in *Proc. 28th Annu. ACM Symp. Theory Comput.*, 1996, pp. 639–648.

[6] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends® Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[7] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Appl. Energy*, vol. 141, pp. 190–199, Mar. 2015.

[8] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016.

[9] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016.

[10] T. Zhang, G. Zhang, J. Lu, X. Feng, and W. Yang, "A new index and classification approach for load pattern analysis of large electricity customers," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 153–160, Feb. 2012.

[11] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 420–430, Jan. 2014.

[12] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-based approach to household electric load curve clustering and prediction," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5196–5206, Sep. 2018.

[13] Q. Huang, W. Jiang, J. Shi, C. Wu, D. Wang, and Z. Han, "Federated shift-invariant dictionary learning enabled distributed user profiling," *IEEE Trans. Power Syst.*, early access, Jul. 19, 2023, doi: 10.1109/TPWRS.2023.3296976.

[14] M. Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014.

[15] N. Liu et al., "Online energy sharing for nanogrid clusters: A Lyapunov optimization approach," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4624–4636, Sep. 2018.

[16] C. Lu, J. Wu, J. Cui, Y. Xu, C. Wu, and M. C. Gonzalez, "Deadline differentiated dynamic EV charging price menu design," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 502–516, Jan. 2023.

[17] J. Yang, J. Zhao, F. Wen, and Z. Dong, "A model of customizing electricity retail prices based on load profile clustering analysis," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3374–3386, May 2019.

[18] A. Halder, X. Geng, P. Kumar, and L. Xie, "Architecture and algorithms for privacy preserving thermal inertial load management by a load serving entity," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3275–3286, Jul. 2017.

[19] S. Li, K. Xue, Q. Yang, and P. Hong, "PPMA: Privacy-preserving multisubset data aggregation in smart grid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 462–471, Feb. 2018.
[20] A. Abdallah and X. S. Shen, "A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 396–405, Jan. 2018.
[21] M. Baza et al., "Privacy-preserving blockchain-based energy trading schemes for electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9369–9384, Sep. 2021.
[22] Z. Wan, T. Zhang, W. Liu, M. Wang, and L. Zhu, "Decentralized privacy-preserving fair exchange scheme for V2G based on blockchain," *IEEE Trans. Depend. Secure Comput.*, vol. 19, no. 4, pp. 2442–2456, Jul./Aug. 2022.
[23] S. Lee and D.-H. Choi, "Dynamic pricing and energy management for profit maximization in multiple smart electric vehicle charging stations: A privacy-preserving deep reinforcement learning approach," *Appl. Energy*, vol. 304, Dec. 2021, Art. no. 117754.
[24] M. Jia, Y. Wang, C. Shen, and G. Hug, "Privacy-preserving distributed clustering for electrical load profiling," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1429–1444, Mar. 2021.
[25] C. Dwork, F. McSherry, K. Nissim, and A. D. Smith, "Calibrating noise to sensitivity in private data analysis," *J. Privacy Confidential.*, vol. 7, no. 3, pp. 17–51, 2016.
[26] F. Fioretto, T. W. Mak, and P. Van Hentenryck, "Differential privacy for power grid obfuscation," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1356–1366, Mar. 2020.
[27] H. Wang, J. Zhang, C. Lu, and C. Wu, "Privacy preserving in non-intrusive load monitoring: A differential privacy perspective," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2529–2543, May 2021.
[28] H. Wang and C. Wu, "Privacy preservation for time series data in the electricity sector," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 3136–3149, Jul. 2023.
[29] J. Huang, Q. Huang, G. Mou, and C. Wu, "DPWGAN: High-quality load profiles synthesis with differential privacy guarantees," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 3283–3295, Jul. 2023.
[30] R. R. Avula, T. J. Oechtering, and D. Månsson, "Privacy-preserving smart meter control strategy including energy storage losses," in *Proc. IEEE PES Innovat. Smart Grid Technol. Conf. Europe*, 2018, pp. 1–6.
[31] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, and D. Megías, "Individual differential privacy: A utility-preserving formulation of differential privacy guarantees," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 6, pp. 1418–1429, Jun. 2017.
[32] B. Wang, Q. Guo, T. Yang, and H. Sun, "Evaluation of information value for solar power plants in market environment," in *Proc. IEEE 4th Conf. Energy Internet Energy Syst. Integr.* , 2020, pp. 3574–3580.
[33] B. Wang, Q. Guo, T. Yang, and B. Wen, "Value evaluation of wind power forecasting information for economic dispatch," in *Proc. IEEE 12th PES Asia-Pac. Power Energy Eng. Conf.*, 2020, pp. 1–5.
[34] M. Yu et al., "Pricing information in smart grids: A quality-based data valuation paradigm," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3735–3747, Sep. 2022.
[35] A. Ghosh and A. Roth, "Selling privacy at auction," *Games Econ. Behav.*, vol. 91, pp. 334–346, May 2015.
[36] L. Fleischer and Y. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data," in *Proc. 13th ACM Conf. Electron. Commer.*, 2012, pp. 568–585.
[37] P. Dandekar, N. Fawaz, and S. Ioannidis, "Privacy auctions for recommender systems," *ACM Trans. Econ. Comput.*, vol. 2, no. 3, pp. 1–22, 2014.
[38] A. Ghosh, K. Ligett, A. Roth, and G. Schoenebeck, "Buying private data without verification," in *Proc. ACM Conf. Econ. Comput.*, 2014, pp. 931–948.
[39] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.
[40] P. Siano, "Demand response and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 461–478, Feb. 2014.
[41] B. Hashemi, M. Shahabi, and P. Teimourzadeh-Baboli, "Stochastic-based optimal charging strategy for plug-in electric vehicles aggregator under incentive and regulatory policies of DSO," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3234–3245, Apr. 2019.
[42] D. T. Nguyen, H. T. Nguyen, and L. B. Le, "Dynamic pricing design for demand response integration in power distribution networks," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3457–3472, Sep. 2016.
[43] D. Vieira, R. A. Shayani, and M. A. G. de Oliveira, "Reactive power billing under nonsinusoidal conditions for low-voltage systems," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 2004–2011, Aug. 2017.
[44] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, "Probabilistic individual load forecasting using pinball loss guided LSTM," *Appl. Energy*, vol. 235, pp. 10–20, Feb. 2019.
[45] K. Zheng, Q. Chen, Y. Wang, C. Kang, and Q. Xia, "A novel combined data-driven approach for electricity theft detection," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1809–1819, Mar. 2019.
[46] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.
[47] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*, vol. 47. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[48] S. Guha, R. Rastogi, and K. Shim, "Cure: An efficient clustering algorithm for large databases," *Inf. Syst.*, vol. 26, no. 1, pp. 35–58, 2001.
[49] S. Pei, F. Nie, R. Wang, and X. Li, "Efficient clustering based on a unified view of k-means and ratio-cut," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 14855–14866.
[50] S. Tan, D. De, W.-Z. Song, J. Yang, and S. K. Das, "Survey of security advances in smart grid: A data driven approach," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 397–422, 1st Quart., 2016.
[51] A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 99–107, Jun. 2010.
[52] V. Odelu, A. K. Das, M. Wazid, and M. Conti, "Provably secure authenticated key agreement scheme for smart grid," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1900–1910, May 2018.
[53] J. Pei, "A survey on data pricing: From economics to data science," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4586–4608, Oct. 2022.
[54] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Sci. Technol.*, vol. 20, no. 2, pp. 117–129, Apr. 2015.
[55] J. Cochrane, *Asset Pricing: Revised Edition*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[56] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "A survey of smart data pricing: Past proposals, current plans, and future trends," *ACM Comput. Surveys*, vol. 46, no. 2, pp. 1–37, 2013.
[57] Pecan Street. "Historical load data." Accessed: Feb. 1, 2023. [Online]. Available: https://www.pecanstreet.org/dataport/

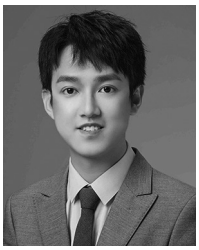
Chenbei Lu (Graduate Student Member, IEEE) received the bachelor's degree from the School of Software Engineering, Huazhong University of Science and Technology. He is currently pursuing the Ph.D. degree in computer science and technology with the Institute for Interdisciplinary Information Sciences, Tsinghua University, advised by Prof. C. Wu. He is currently working on the optimal design and operation of power systems. He has been awarded the National Scholarship in 2017 and Excellent Graduate of Huazhong University of Science and Technology in 2020.

Jingshi Cui (Member, IEEE) received the Ph.D. degree from the Institute for Interdisciplinary Information Science, Tsinghua University in June 2023.

She is an Associate Research Fellow with the Department of Control Science and Intelligence Engineering, Nanjing University. Her current research interests include economic design for electricity market and optimal dispatch for power system.

Haoxiang Wang (Graduate Student Member, IEEE) received the bachelor's degree from the Department of Energy and Power Engineering, Tsinghua University, and the master's degree from the IIIS, Tsinghua University, where he is currently pursuing the Ph.D. degree with the Department of Automation.

He is currently working on privacy preservation and stochastic analysis. He has been awarded Excellent Comprehensive Scholarship of Tsinghua University in 2018.



Hongyu Yi received the bachelor's degree from the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, where he is currently pursuing the master's degree, advised by Prof. C. Wu. His research interests include control in power systems, online algorithms, and online learning.

Chenye Wu (Member, IEEE) received the Ph.D. degree from the IIIS, Tsinghua University in July 2013.

He is an Assistant Professor with th School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen (CUHK Shenzhen). Before joining CUHK Shenzhen, he was an Assistant Professor with the Institute for Interdisciplinary Information Sciences, Tsinghua University. He worked with ETH Zurich as a wiss. Mitarbeiter (Research Scientist), working with Prof. G. Hug in 2016. Before that, Prof. K. Poolla and Prof. P. Varaiya hosted him as a Postdoctoral Researcher with the University of California at Berkeley, for two years. From 2013 to 2014, he spent one year with Carnegie Mellon University as a Postdoctoral Fellow, hosted by Prof. G. Hug and Prof. S. Kar. He is currently working on economic analysis, optimal control, and operation of power systems. His Ph.D. advisor is Prof. A. Yao, the laureate of the A.M. Turing Award in the year of 2000. He was the Best Paper Award co-recipients of IEEE SmartGridComm 2012, IEEE PES General Meeting 2013, and IEEE PES General Meeting 2020.