


**Please cite the Published Version**

Zheng, Zhigao, Wang, Tao, Bashir, Ali Kashif , Alazab, Mamoun, Mumtaz, Shahid and Wang, Xiaoyan (2022) A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid. IEEE Transactions on Computers, 71 (11). pp. 2915-2926. ISSN 0018-9340

**DOI:** <https://doi.org/10.1109/TC.2021.3130402>

**Publisher:** Institute of Electrical and Electronics Engineers

**Version:** Accepted Version

**Downloaded from:** <https://e-space.mmu.ac.uk/631075/>

**Additional Information:** © 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

# A Decentralized Mechanism Based on Differential Privacy for Privacy-Preserving Computation in Smart Grid

Zhigao Zheng, *Member, IEEE*, Tao Wang <sup>†</sup>, *Member, IEEE*, Ali Kashif Bashir, *Senior Member, IEEE*, Mamoun Alazab, *Senior Member, IEEE*, Shahid Mumtaz, and Xiaoyan Wang

**Abstract**—As one of the most successful industrial realizations of Internet of Things, a smart grid is a smart IoT system that deploys widespread smart meters to capture fine-grained data on residential power usage. Unfortunately, it always suffers diverse privacy attacks, which seriously increases the risk of violating the privacy of customers. Although some solutions have been proposed to address this privacy issue, most of them mainly rely on a trusted party and focus on the sanitization of metering measurements. Moreover, these solutions are vulnerable to advanced attacks. In this paper, we propose a decentralized mechanism for privacy-preserving computation in smart grid called DDP, which leverages the differential privacy and extends the data sanitization from the value domain to the time domain. Specifically, we inject Laplace noise to the measurements at the end of each customer in a distributed manner, and then use a random permutation algorithm to shuffle the power measurement sequence, thereby enforcing differential privacy after aggregation and preventing the sensitive power usage mode information of the customers from being inferred by other parties. Extensive experiments demonstrate that DDP shows an outstanding performance in terms of privacy from the non-intrusive load monitoring (NILM) attacks and utility by using two different error analysis.

**Index Terms**—smart grid, privacy-preserving computation, random permutation, differential privacy, non-intrusive load monitoring.

The research work reported in this paper is financially supported by the National Natural Science Foundation of China (Grant No. 42001392, 61861042 and 61701453), the self-determined research funds of CCNU from the colleges' basic research and operation of MOE (No. CCNU20XJ008), the Fundamental Research Funds for the Central Universities (the China University of Geosciences (Wuhan), No. CUG190607, and Wuhan University), the Natural Science Foundation of China (No. 41571426), and Wuhan Applied Basic Research Program (No. 2017010201010114).

Zhigao Zheng is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: zhengzhigao@hust.edu.cn).

Tao Wang is with the Hubei Research Center for Educational Informationization, Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China. (e-mail: tmac@ccnu.edu.cn).

Ali Kashif Bashir is with Department of Computing and Mathematics, Manchester Metropolitan University, Manchester M15 6BH, UK. He is also with School of Electrical Engineering and Computer Science (SECS), National University of Science and Technology, Islamabad 44000, Pakistan, as an Adjunct Professor. (e-mail: dr.alikashif.b@ieee.org)

Mamoun Alazab is with College of Engineering, IT & Environment, Charles Darwin University, Casuarina, NT 0810, Australia (e-mail: alazab.m@ieee.org)

Shahid Mumtaz is with Instituto de Telecomunicacoes, Campus Universitário de Santiago, Lisboa, Portugal. (e-mail: Dr.shahid.mumtaz@ieee.org)

Xiaoyan Wang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan 430079 China (e-mail: wangxiaoyan012@ke.com.cn)

<sup>†</sup> Corresponding Author: Tao Wang

## I. INTRODUCTION

SMART grid is one of the most successful industrial realizations of advanced Internet of Things (IoT), which is also referred to as Internet of Energy [1]. In this regard, networked smart meters, which are the key elements of a smart grid, can be considered as IoT devices that autonomously provide fine-grained power consumption measurements to utility providers by their bidirectional power flow and communication capabilities. These measurements can improve the efficiency of the power grid by enabling dynamic pricing and demand response. However, such industrial IoT solution may suffer diverse cyber attacks, ranging from hardware attacks to software attacks, especially the privacy attacks that cause serious privacy threats. Privacy issues have been primary concerns of consumers in a smart grid [2]. For example, the fine-grained measurements can reveal the usage modes, personal routines, behavioral preferences, occupancy, and household financial situation of these consumers. Such privacy-related problems seriously restrict the development of smart grids.

To address these privacy issues, various privacy-preserving mechanisms for smart grid have been proposed [3]. Data anonymization and data encryption are two widespread privacy-preserving mechanisms that are designed from the company-side, wherein the utility companies and the related components are supposed to be trusted. Specifically, data anonymization [4] removes or pseudonymizes [5] the private attributes related to customer identities from the smart meter data. However, such sensitive information can be still extracted from these anonymized data by an adversary with specific auxiliary information. Data encryption [8] [25] leverages the homomorphic features of cryptographic computation to aggregate the metering measurements obtained by the gateways and the control center, so as to reduce the potential risk of privacy disclosure. However, data encryption mechanisms are limited by their relatively high computation or communication complexity.

Various privacy-preserving mechanisms have also been proposed from the customer-side. Typically, device-based load hiding mechanisms adopt a household rechargeable battery [10] [11] [12] [13] or other energy storage units [14] to hide the actual electricity consumption behaviors of customers. However, rechargeable batteries are relatively expensive, and energy storage units foremost have to respond to customer demands. Moreover, these mechanisms are tightly coupled

with the capacity and charge-discharge rate of the batteries and energy storage units, thereby impacting the privacy-preserving results. Recently, blockchain, as an emerging information networking architecture, has been applied for privacy-preserving data aggregation in smart grids [9]. Given its decentralized characteristic, a blockchain-based mechanism can preserve the meter data even if a malicious user is selected as the aggregator who tampers with these data.

In recent years, differential privacy, as a promising privacy definition, has been applied on smart grids [6] [7] [15] [16] [17] given its capability to provide a formal mathematical description and proof for smart meter privacy. Particularly, differential privacy is always combined with other privacy-preserving techniques. For instance, Bao et al. [6] proposed a secure data aggregation mechanism called DPAFT that can achieve differential privacy and fault tolerance simultaneously. Lyu et al. [15] proposed a privacy-preserving aggregation mechanism called PPFA by utilizing the Fog computing architecture and Gaussian mechanism, a basic implementation mechanism of the approximate differential privacy, thereby offering provable differential privacy guarantees for the aggregate statistic on both Fog and Cloud levels. With regard to customer-side mechanisms, researchers have attempted to enforce the battery-based load hiding (BLH) methods to ensure differential privacy [16]. Zhang et al. [17] proposed two cost-friendly differential privacy-preserving mechanisms by combining differential privacy with the BLH method, thereby providing rigorous privacy preservation and cost savings simultaneously. This paper mainly focuses on customer-side privacy-preserving solutions, which are facing the following two critical challenges:

(1) These mechanisms require a trusted party to perform data sanitization. Nevertheless, not all utility companies and other third-parties can be trusted by customers. Therefore, a decentralized or distributional solution is desired. Márk et al. [18] designed a distributed differentially private mechanism for sum queries that protects not only individual records but also the parameters of individual energy consumption modes. However, this mechanism only preserves the privacy of distribution parameters for large-scale queries.

(2) The aforementioned mechanisms generally sanitize or perturb metering measurements, which are vulnerable to filtering. Power usage information can be approximately recovered after inputting a large number of queries, thereby exposing consumers to privacy attacks. These mechanisms also have limitations related to power waveform hiding or perturbing because of capability boundaries of the energy storage units.

Motivated by the aforementioned challenges, we propose a decentralized differentially private mechanism for smart meter data called DDP, which extends the data sanitization from the value domain to the time domain. Specifically, the main contributions of this paper are described as follows:

(1) By combining the differential privacy and random permutation, we present an secure and efficient privacy-preserving mechanism, which injects Laplace noise into the measurements in a distributed manner at the end of each customer and then apply a random permutation algorithm to permute the power measurement sequence, thereby enforcing differential

privacy after aggregation and preventing the sensitive power usage mode information of customers from being inferred by other parties, such as curious participants in a smart grid and malicious adversaries.

(2) By using ECO dataset [19], we quantitatively evaluate the performance of DDP and reveal that DDP demonstrates an outstanding performance in terms of privacy from non-intrusive load monitoring (NILM) attacks and utility by using two different error analysis.

The rest of this paper is organized as follows. Section II summarizes the related works. Section III briefly revisits the preliminaries. Section IV formalizes the problem model, including the system model, adversarial model, and security features. Section V presents the DDP mechanism and analyzes its privacy, utility, and feasibility. Section VI evaluates the performance of this mechanism. Finally, Section VII concludes the paper and provides suggestions for future work.

## II. RELATED WORKS

Extensive works have been done to tackle the privacy issues in the smart grids, which can be generally classified into two categories: company-side privacy-preserving mechanisms and customer-side privacy-preserving mechanisms.

### A. Company-Side Privacy-Preserving Mechanisms

Company-side privacy-preserving mechanisms suppose that the company is trusted and leverage data anonymization and data encryption to preserve the fine-grained smart metering measurements. Data anonymization intuitively [4] removes or pseudonymizes [5] the private attributes related to customer identities from the smart meter data. However, such anonymized data can be de-anonymized by linking and inference attacks from adversaries with specific auxiliary information, and the sensitive information can be still extracted from these anonymized data. Data encryption [8] [25] generally leverages the homomorphic features of cryptographic computation to aggregate the metering measurements obtained by the gateways and the control center, so as to reduce the potential risk of privacy disclosure. Moreover, some studies combine the homomorphic encryptions with differential privacy, to achieve rigorous privacy guarantee and other purposes. Specifically, Bao et al. [6] proposed a secure data aggregation mechanism called DPAFT that uses a novel key management technique and an improved Boneh–Goh–Nissim cryptosystem, thereby achieving differential privacy and fault tolerance simultaneously. Ni et al. [7] presented DiPrism with the desirable features of data aggregation and differential privacy by employing Lifted ElGamal encryption and Laplace noise. However, those mechanisms are limited by the relatively high computation or communication complexity of homomorphic encryptions. In addition, some scholars investigate the blockchain-based privacy-preserving framework [9] and fog computing-based privacy-preserving framework [26]. Specifically, the blockchain-based privacy-preserving framework divides users into different groups, and each group has a private blockchain to record its members' data. To preserve the inner privacy within a group, pseudonyms are used to hide users' identities,

and each user may create multiple pseudonyms and associate his/her data with different pseudonyms. The fog computing-based privacy-preserving framework, for instance, PPFA[26], enables the intermediate Fog nodes to periodically collect data from nearby smart meters and accurately derive aggregate statistics as the fine-grained Fog level aggregation, and use Gaussian mechanism to distribute noise generation among parties, thus offering provable differential privacy guarantees of the aggregate statistic on both Fog level and Cloud level. However, these methods require to modify the architecture of smart grids, thus their complexity and economic cost are relatively high.

### B. Costumer-Side Privacy-Preserving Mechanisms

Costumer-side privacy-preserving mechanisms carry out data sanitization at the costumer-side. For instance, device-based load hiding mechanisms typically employ a household rechargeable battery (e.g., BLH mechanism [10], [11], [12], [13]) or energy storage units (e.g., electric vehicles [14]) to flat or randomize the meter readings by reporting the total electricity consumption from the electrical appliances and the battery, thereby hiding the actual electricity consumption behaviors of customers. For instance, Sun et al. [14] used a Markov decision process (MDP) to model the household demand and customer behavior and a Q-learning algorithm to self-adapt the control policies for energy storage units, such as electric vehicles (EVs) and heating, ventilating, and air conditioning (HVAC) systems. Additionally, BHL mechanisms are combined with differential privacy[16], [17], attempting to conduct provably privacy-preserving mechanisms and resist to non-intrusive load monitoring (NILM) attacks. For instance, Zhang et al. [17] proposed battery-based differential privacy-preserving (BDP) scheme and two cost-friendly differential privacy-preserving (CDP) schemes by extending BDP under static and dynamic pricing policies, thereby providing rigorous privacy preservation and cost savings simultaneously. However, rechargeable batteries are relatively expensive, and energy storage units foremost have to respond to customer demands. Moreover, these mechanisms are tightly coupled with the capacity and charge-discharge rate of the batteries and energy storage units.

## III. PRELIMINARIES

This section describes the notations and preliminaries, including NILM, differential privacy, and random permutation.

### A. Non-Intrusive Load Monitoring

NILM [20] is a promising energy disaggregation approach to estimate appliance-level electricity consumption from aggregated household consumption data by monitoring only one meter per household and without requiring any intrusion into the power loads. A general NILM framework has three constituent modules, namely, data acquisition, feature extraction, and load identification. Specifically, after acquiring aggregated load measurements at an adequate rate so as to identify distinctive load modes, feature extraction calculates the power metrics based on current, voltage, or waveform measurements,

and then detects events of appliance state transition. Afterwards, load identification uses supervised or unsupervised machine learning techniques to disaggregate the aggregated load measurements and to identify appliance-specific states.

Formally, given a discrete sequence of the acquired aggregate metering measurements  $\mathbf{m} = [m_1, \dots, m_t]$ , where  $t$  represents the discrete time measurements, NILM tries to determine the sequence of appliance demands  $\mathbf{v}^{(n)} = [v_1^{(n)}, \dots, v_t^{(n)}]$ . Alternatively, given the mapping relation between states and demands, this process can be represented as the determination of appliances states  $\mathbf{s}^{(n)} = [s_1^{(n)}, \dots, s_t^{(n)}]$ . The appliance states are in one-to-one correspondence with the operations (for example, 'on', 'off', or 'standby') because an operation approximately produces a constant power waveform.

Nevertheless, NILM can reveal the private information of customers, including their behavioral modes, financial situation, and daily routine information, via appliance-level energy disaggregation, thereby leading to serious safety problems. Therefore, we consider NILM as a critical part of the adversarial model.

### B. Differential Privacy

Differential privacy is motivated by the intuition that the sanitized output generated by the input of a database is approximately indistinguishable from that generated by the input of its neighbor database. A pair of datasets,  $\mathcal{D}$  and  $\mathcal{D}'$ , is called neighbor datasets *iff*  $\mathcal{D}'$  can be produced by adding, removing, or modifying exactly one tuple from  $\mathcal{D}$ .

**Definition 1** ( $\epsilon$ -differential privacy). *A sanitization mechanism  $\mathcal{M}$  satisfies  $\epsilon$ -differential privacy if it holds for any pair of neighbor datasets  $\mathcal{D}$  and  $\mathcal{D}'$  that*

$$\Pr(\mathcal{M}(\mathcal{D}) \in \mathcal{S}) \leq e^\epsilon \Pr(\mathcal{M}(\mathcal{D}') \in \mathcal{S}),$$

where  $\mathcal{S}$  denotes all possible outputs of  $\mathcal{M}$ , and  $\epsilon$  is the privacy budget that is mainly restricted by  $\mathcal{M}$ .

The inequality indicates that an adversary can possess a narrow confidence for inferring the either/or input dataset from  $\mathcal{D}$  and  $\mathcal{D}'$  (i.e., the presence or absence of exactly one tuple in the input dataset) only by observing  $\mathcal{S}$  regardless of the background knowledge of the adversary.

Differential privacy guarantees the privacy of any individual with sensitive attributes in the dataset. In practical applications,  $\epsilon$ -differential privacy is generally enforced by a fundamental mechanism (i.e., Laplace mechanism) that relies on the important parameter of  $\mathcal{L}_1$ -sensitivity.

**Definition 2** ( $\mathcal{L}_1$ -sensitivity). *Given a query function  $Q$ , its  $\mathcal{L}_1$ -sensitivity  $\Delta_Q$  is the maximum  $\mathcal{L}_1$  distance between the results of  $Q$  over any pair of neighbor datasets  $\mathcal{D}$  and  $\mathcal{D}'$  and can be expressed as follows:*

$$\Delta_Q = \max_{\mathcal{D}, \mathcal{D}'} Q(\mathcal{D}) - Q(\mathcal{D}')_1.$$

where  $\Delta_Q$  is characterized by the query function  $Q$  and its output domain rather than the input dataset  $\mathcal{D}$ .

$\mathcal{L}_1$ -sensitivity underlies the Laplace mechanism, which is formally provided by Definition 3.

**Definition 3** (Laplace mechanism). *Given dataset  $\mathcal{D}$  and query function  $Q$ , the Laplace mechanism obtains sanitized outputs  $\mathcal{S}$  by injecting the i.i.d. Laplace noise  $\mathcal{L}$  into the exact query result with a mean of 0 and scale  $\lambda = \Delta_Q/\epsilon$ . This mechanism is defined as  $\mathcal{M}_{\mathcal{L}}(\mathcal{D}) = Q(\mathcal{D}) + \mathcal{L}$ .*

Obviously, a larger sensitivity leads to a higher volume of noise. Differential privacy has two fundamental properties that also serve as core guidelines for designing differentially private mechanisms.

**Property 1** (Transformation Invariance). *Given a mechanism  $\mathcal{M}$  that satisfies  $\epsilon$ -differential privacy and a mechanism  $\mathcal{K}$  whose domain contains the range of  $\mathcal{M}$  and whose random bits are statistically independent from the random bits of  $\mathcal{M}$ , a mechanism  $\mathcal{A}(\bullet) = \mathcal{K}(\mathcal{M}(\bullet))$  also satisfies  $\epsilon$ -differential privacy.*

Obviously, the output of  $\mathcal{M}$  is the only input of  $\mathcal{K}$ , which simulates a statistical analysis by using the output of  $\mathcal{M}$ . Therefore, the property of transformation invariance guarantees that the output of  $\mathcal{M}$  and the results of the statistical analysis on this output are both secure.

**Property 2** (Convexity). *Given two mechanisms  $\mathcal{M}_1$  and  $\mathcal{M}_2$  that satisfy  $\epsilon$ -differential privacy and a parameter  $p \in [0, 1]$ , the mechanism  $\mathcal{M}^p$  that runs  $\mathcal{M}_1$  with probability  $p$  and  $\mathcal{M}_2$  with probability  $1-p$  should also satisfy  $\epsilon$ -differential privacy.*

In other words, the convexity supports a random selection of the privacy-preserving mechanism  $\mathcal{M}$  that satisfies differential privacy to inject further uncertainty into the data sanitization.

To protect the privacy of customers from adversaries, Laplace noise can be generated by each household in a distributed manner. In such case, the infinite divisibility of Laplace distribution provides a decentralized solution for enforcing  $\epsilon$ -differential privacy [7], and the Laplace distribution can be fabricated from the sum of the i.i.d. Gamma distribution.

**Definition 4** (Infinite divisibility of Laplace distribution). *Given a parameter  $\lambda$ , a random variable  $\text{Lap}(\lambda)$  is sampled from a Laplace distribution with probability density function  $f(x, \lambda) = \frac{1}{2\lambda}e^{-x/\lambda}$ . Then, the distribution  $\text{Lap}(\lambda)$  is infinitely divisible, and it holds  $\text{Lap}(\lambda) = \sum_{i=1}^n (\mathcal{G}(n, \lambda) - \mathcal{G}'(n, \lambda))$  for every integer  $n \geq 1$ , where  $\mathcal{G}(n, \lambda)$  and  $\mathcal{G}'(n, \lambda)$  are i.i.d. with PDF  $\gamma(x, n, \lambda)$ , and*

$$\gamma(x, n, \lambda) = \frac{1/\lambda^{1/n}}{\Gamma(1/n)} x^{(1-n)/n} e^{-x/\lambda}$$

for  $x \geq 0$ , where  $\Gamma(1/n)$  denotes the value of the Gamma density function at  $1/n$ .

If the number of smart meters is  $N$ , then a sanitization mechanism injects  $\mathcal{G}(N, \lambda) - \mathcal{G}'(N, \lambda)$  into the measurement  $m_i$  of the  $i$ -th smart meter before reporting. Thus, the reported electricity consumption aggregation is

$$\begin{aligned} \mathcal{M}(\mathcal{D}) &= \sum_{i=1}^N m_i + \sum_{i=1}^N (\mathcal{G}(N, \lambda) - \mathcal{G}'(N, \lambda)) \\ &= \sum_{i=1}^N m_i + \text{Lap}(\lambda), \end{aligned}$$

thereby achieving  $\epsilon$ -differential privacy.

### C. Random Permutation

Random permutation is a random ordering of a set of objects and a fundamental operation in some fields, such as coding theory and cryptography [21]. A classic algorithm of random permutation is the Fisher–Yates shuffle algorithm [22], which generates a random permutation of a finite sequence by firstly placing all elements in a hat and then continually determining the next element by randomly picking an element from the hat until no elements are left. This algorithm is unbiased and efficient, uniformly outputs the permutation results, has a time complexity that is proportional to the number of items being shuffled, and does not require any additional storage cost. Thus, the FisherYates shuffle algorithm can effectively enhance nonlinearity and uncertainty.

In our case of smart meter, the measurement sequence of customer  $i$  is denoted as  $\mathbf{m}_i = [m_1^i, \dots, m_t^i]$ , where  $T_i = [1, 2, \dots, t]$  represents the true time points of recording the power measurements. We set a time window  $T_w$ , and then randomly shuffle the measurement sequence in the window  $T_w$  to obtain a permutation  $\hat{\mathbf{m}}_i = [\hat{m}_1^i, \dots, \hat{m}_{t'}^i]$ , where  $t'$  is the time point of releasing the permutation. Then, the shuffle function can be mathematically expressed as follows:

$$\hat{m}_{t'}^i = SF(t, t') \cdot m_t^i$$

where  $SF$  represents the shuffle operation of the original measurement sequence.

Through the random shuffle, we separate the time points of measuring meter data from the time points of releasing them, thereby destroying the related attribute information about the running states of the appliances hidden in the original meter data. Accordingly, the random shuffle algorithm can mask the power usage modes of customers at the appliance-level, thereby preventing their privacy from being leaked.

## IV. PROBLEM FORMALIZATION

In this section, we state the problem by formalizing the system model and adversarial model and by identifying the security features.

### A. System Model

This paper only focuses on the privacy issue being faced by customers when their power consumption data are recorded and reported periodically or in real-time to the meter data management unit (MDMU) of the electric power company for different purposes, such as billing, operations, and value-added services. Therefore, the system model roughly consists of various customers equipped with smart meters, local aggregators, data and power communication networks, and MDMU, as illustrated in Fig.1.

A smart meter, equipped at each customer  $U_i \in \mathbb{U}$ , where  $\mathbb{U} = \{U_1, U_2, \dots, U_n\}$ , periodically records the real-time power consumption of a customer and reports the data to the local aggregator within a certain period, such as every 15 minutes. Formally, for a single customer  $U_i$ , the measurement sequence is

$$\mathbf{m}_i = [m_1^i, \dots, m_t^i].$$

Accordingly, for a smart metering network that comprise  $N$  customers, the measurement sequence is

$$\mathbf{M}_t = [\mathbf{m}_t^1, \dots, \mathbf{m}_t^N].$$

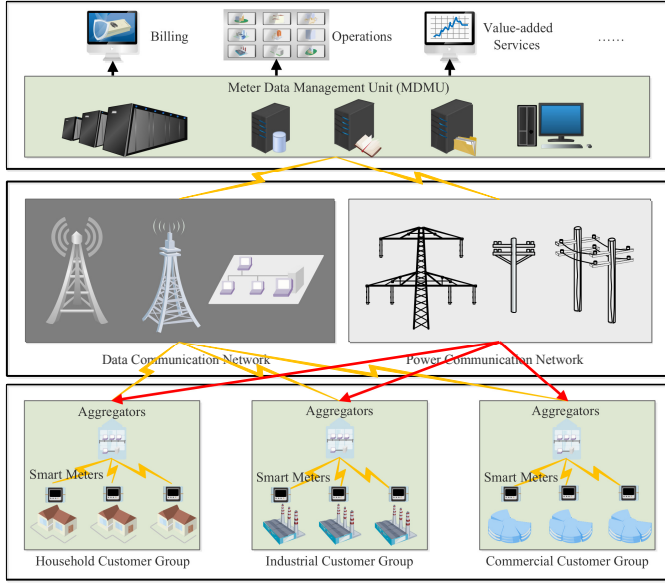


Fig. 1. Basic system model of a smart grid.

The MDMU is responsible for controlling the supply of electricity from the power plant, the distribution of electricity to users through power transmission networks, and the processing and analysis of the electricity consumption data aggregated by the aggregators, all of which ensure that the smart grid operates normally and provides various services. On the one hand, the MDMU is required to compute the specific electricity consumption and the corresponding charge under a static pricing policy or the Time-of-Use (TOU) pricing policy [23]. Given the unit electricity price at time  $t$ , denoted as  $P_t$ , the electricity charge of customer  $U_i$  in a certain period  $T_c$  is

$$C_i = \sum_{t \in T_c} P_t \times m_t^i$$

On the other hand, the MDMU schedules the electric power production and distribution and controls the price based on the aggregation result  $H$ , which is obtained by aggregating the smart meter data according to a certain rule  $R$ , that is,

$$H = \sum_{i=1}^N R \times m_t^i$$

### B. Adversarial Model

Adversaries are dishonest but non-intrusive. They may break the system rules, attempting to acquire smart meter data by eavesdropping, and inferring private information of the customers by analyzing meter data. They may also launch collusion attacks with some malicious smart grid participants to obtain some system parameters and launch differential attacks to identify sensitive information. Nevertheless, adversaries have no permission to insert, delete, or modify smart meter data.

Suppose that an adversary uses an attack algorithm  $\mathcal{A}$  (that is a NILM algorithm) to analyze the measurements  $\mathbf{m}_i$  of the

customer  $U_i$  and to subsequently learn the states of appliances in a certain period  $T_q$ . The adversarial model can then be formulated as follows:

$$\mathbf{y} = \mathcal{A}(\mathbf{m}_i)$$

$\mathbf{y}$  consist of many state labels  $y_j$  of various appliances,  $y_j = (t_j, o_j, s_j)$ , where  $t_j$  is the state transition time of appliance  $j$ ,  $o_j$  is the recognized appliance, and  $s_j$  is the running state of appliance  $j$ . Therefore, adversaries launch NILM attacks to infer the private information of customers by disaggregating the state labels.

### C. Security Features

The goal of this paper is to design a privacy-ensured mechanism that achieves the following desirable security features:

1) *Rigorous Privacy Preservation*: Neither the measurement readings of a customer nor the aggregated data would disclose any private information of customers. An adversary who is dishonest but non-intrusive cannot obtain any additional information by analyzing vari-size-grained smart meter data.

2) *Decentralization*: Customers do not fully trust the MCMU and prefer to proactively sanitize their personal meter data before the aggregation. The privacy preservation does not completely depend upon the utility company and is highly unlikely to be achieved by implementing a centralized solution.

3) *High Efficiency*: Efficiency involves computational efficiency, communication efficiency, and cost-efficiency. Large-scale smart meters have limited computing capacity, and the networks connecting meters have limited bandwidth. Therefore, the privacy-preserving mechanism must have low computational and communication complexity. Moreover, the deployment and modification for utility companies and the adoption for customers must be cost-effective to ensure a smooth implementation of the mechanism.

## V. ALGORITHM DESCRIPTION

In this section, we propose a decentralized differentially private mechanism called DDP to protect customer privacy in a decentralized and efficient manner. Specifically, DDP generates Laplace noise in a distributed manner, injects them independently into the measurements of each customer, and shuffles the noisy measurement sequence to mask the information related to the power usage modes of these customers. Consequently, the adversary cannot infer the private information of customers, and differential privacy is enforced.

### A. DDP Mechanism

In our case, three requirements must be fulfilled in designing the DDP mechanism.

- The measurement intervals of a smart meter must be successive and equal, and the time difference between the true time point of the measurement recording and the shuffled time point of the measurement releasing should be the integral multiple of the minimum clock-unit.
- The shuffle operations may impact the enforcement of the differential privacy. Therefore, a noise obey Gamma

distribution must be maintained at any time point of measurement to make sure that the sum of noise after aggregation follows a Laplace distribution and that the differential privacy is enforced.

- The measurement periods of a smart meter are always pre-set and constant, and the time window in the shuffle operations should be restrictive and within a limited range, to obtain accurate statistical results within a certain period.

The DDP can be divided into four phases, namely, initialization, noise generation, measurement shuffling, and report aggregation, and its basic algorithmic framework is shown as Fig. 2.

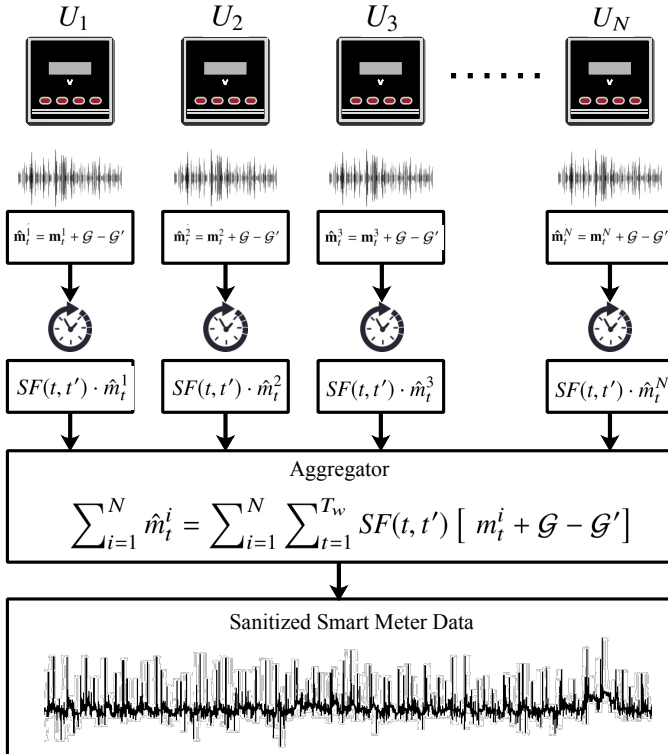


Fig. 2. Algorithmic framework of DDP.

### 1) Initialization

This step only considers the necessary parameters used in our DDP mechanism. Specifically, the whole system is bootstrapped and initialized by a trusted authority in the beginning. In the initialization, DDP sets the privacy parameter \$\epsilon\$ to 1, the number of smart meters connected to an aggregator \$N\$ to 100, the shuffling time window \$T\_w\$ to 1 min, and the interval of data report \$T\_r\$ to 15 min. The cyclic memory is also initialized. After initializing the system, the trusted authority will be offline and stay out of the subsequent data processing.

### 2) Noise generation

DDP generates two i.i.d. Gamma noises, \$\mathcal{G}(N, \lambda)\$ and \$\mathcal{G}'(N, \lambda)\$, with probability density \$\gamma(x, N, \lambda)\$, where \$\lambda = \Delta\_Q/\epsilon\$, and then obtain a Laplace noise \$\Gamma\_i = \mathcal{G}(N, \lambda) - \mathcal{G}'(N, \lambda)\$. In smart meters, \$Q\$ denotes the total electricity consumption aggregated by an aggregator, and \$\Delta\_Q\$ denotes the maximum power consumption of any customer in a constant

period. Therefore, each smart meter produces a noisy version of the measurement sequence \$\hat{\mathbf{m}}\_t^i = \mathbf{m}\_t^i + \Gamma\_i\$.

### 3) Measurement shuffling

DDP shuffles the noisy measurement sequence of each smart meter \$\hat{\mathbf{m}}\_t^i\$ in a time window \$T\_w\$ and obtains \$\hat{m}\_{t'}^i = SF(t, t') \cdot \hat{m}\_t^i\$. The shuffled measurement sequence \$\hat{\mathbf{m}}\_t^i = [\hat{m}\_{t'}^1, \dots, \hat{m}\_{t'}^N]\$ is stored in the cyclic memory of each smart meter. Then, DDP reports the shuffled data \$\hat{\mathbf{m}}\_t^i\$ to an interval \$T\_r\$.

### 4) Report aggregation

After receiving all measurements of the smart meters, the aggregator aggregates all valid measurements and obtains \$\sum\_{i=1}^N \hat{m}\_t^i = \sum\_{i=1}^N \sum\_{t=1}^{T\_w} SF(t, t') [m\_t^i + \Gamma\_i]\$. Then, the aggregator submits the aggregated smart meter data to the MDMU.

## B. Privacy Analysis

Intuitively, DDP is required to protect the working states of each appliance at any data releasing time \$t\$. If an adversary launches continuous queries in \$k\$ successive intervals, then DDP need a higher privacy budget \$\epsilon\_k = k \times \epsilon\$ to enforce \$\epsilon\$-differential privacy, which reduces the privacy guarantee. Table I presents the actual privacy budget under continuous queries in successive intervals, where the original value is \$\epsilon=1\$ and the unit interval is 10 min.

TABLE I  
PRIVACY BUDGET UNDER CONTINUOUS QUERIES IN SUCCESSIVE INTERVALS

\$N\$	\$k=1\$ (10min)	\$k=3\$ (30min)	\$k=6\$ (60min)	\$k=24\$ (4h)	\$k=48\$ (8h)	\$k=144\$ (24h)
100	0.97	2.34	3.66	9.03	14.13	23.31
200	0.91	2.15	3.25	8.27	12.55	19.72
300	0.87	2.02	2.97	7.60	11.89	17.93
400	0.83	1.95	2.83	7.33	11.37	17.12
500	0.77	1.87	2.74	7.00	10.92	16.01
600	0.71	1.84	2.68	6.87	10.55	15.36
700	0.65	1.80	2.63	6.77	10.36	14.86
800	0.58	1.77	2.58	6.60	10.25	14.59
900	0.54	1.72	2.55	6.38	10.04	13.90
1000	0.51	1.66	2.50	6.32	9.80	13.55

Obviously, with the extended observation period (i.e., \$k\$), the actual privacy budget \$\epsilon\_k\$ increases significantly. Specifically, given that \$N\$ is 100, we have \$\epsilon\_{k=1} = 0.97\$, \$\epsilon\_{k=6} = 3.66\$, and \$\epsilon\_{k=144} = 23.31\$. A larger privacy budget \$\epsilon\$ corresponds to a weaker privacy guarantee. Therefore, for the sake of maintaining the security level of \$\epsilon = 1\$, we have to reset the parameter of the Laplace noise, because the noise parameter is \$\lambda'(t) = \sum\_{i=1}^k \max(\mathbf{m}\_t^i)\$, where the initial parameter \$\lambda(t) = \Delta\_Q/\epsilon = \max(\mathbf{m}\_t^i)\$. In our DDP mechanism, instead of injecting additional noise, we use periodic random shuffling to perturb the power waveform and to mask the time sequence characteristics of the smart meter data. We adopt Pearson's correlation coefficient \$\rho\$ to measure the distinguishability between the original sequence and the shuffled sequence, that is,

$$\rho(\mathbf{m}_i, \hat{\mathbf{m}}_i) = \frac{\text{Cov}(\mathbf{m}_i, \hat{\mathbf{m}}_i)}{\text{Var}(\mathbf{m}_i)\text{Var}(\hat{\mathbf{m}}_i)}$$

As shown in Fig. 3(b), after injecting the Laplace noise distributely, the noisy smart meter data can still partially reveal the power waveform, where \$\epsilon = 1\$, \$N = 100\$, and



$\rho = 0.8955$ . After randomly shuffling in two different time windows, the power waveforms are damaged at different degrees. Specifically, in Fig. 3(c), where the time window  $T_w = 10s$ , the Pearson's correlation coefficient  $\rho = 0.7890$ . As the time window  $T_w$  extends, Fig. 3(d) shows that the damage on the power waveform becomes increasingly evident and that the correlation coefficient  $\rho$  drops to 0.4583, indicating a significant reduction of the correlation between the original and the sanitized sequences. In other words, the power usage modes of a household customer are masked significantly. Fig. 4 illustrates that with the increase of the time window, the Pearson's correlation coefficient presents a obvious decreasing trend and finally tends to flat as the time window  $T_w$  exceed 80 seconds.

However, unlike privacy analysis in the numerical domain, the security level of usage mode information in smart meter data is difficult to measure quantitatively. In the follow-up experiments in Section V, we consider NILM attacks and analyze the accuracy of recognizing the states of different appliances from aggregated smart meter data under various privacy parameters, to intuitively measure the security of the DDP mechanism.

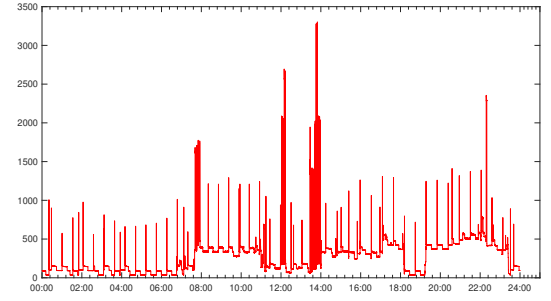
### C. Utility Analysis

We mainly consider the accumulative error of a single customer and the aggregation error of multiple customers at the exact same time to measure the utility of the DDP mechanism. The former affects the accuracy of power metering and billing, whereas the latter affects the accuracy of the aggregated data analysis of all smart meters.

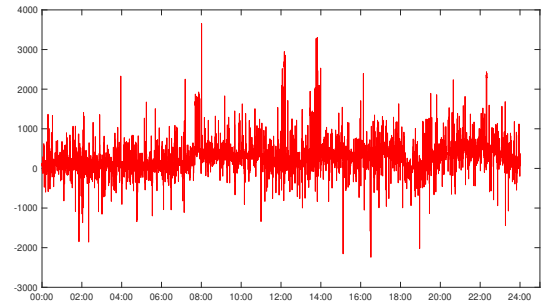
In the DDP mechanism, each smart meter generates two noise sequences sampled from two Gamma distributions,  $\mathcal{G}$  and  $\mathcal{G}'$ , and then injects  $\mathcal{G} - \mathcal{G}'$  into its reading  $\mathbf{m}_i = [m_1^i, \dots, m_t^i]$ . The noise affects the accuracy of power metering and billing within a specific cycle  $C$  (e.g., 30 days). The random shuffling operation does not introduce additional noise and has a small to minimal effects on the accumulative error during a long cycle  $C$ . Then, we can obtain the accumulative error  $e_{ac}(i)$  of customer  $i$  as follows:

$$\begin{aligned} e_{ac}(i) &\approx \frac{1}{\sum_{t=1}^T m_t^i} E \left[ \sum_{t=1}^T \hat{m}_t^i - \sum_{t=1}^T m_t^i \right] \\ &= \frac{1}{\sum_{t=1}^T m_t^i} E \left[ \sum_{t=1}^T [\mathcal{G}(N, \lambda_t) - \mathcal{G}'(N, \lambda_t)] \right] \\ &= \frac{T \cdot E[Lap(\lambda_t)]}{N \cdot \sum_{t=1}^T m_t^i} = \frac{T \cdot \lambda_t}{N \cdot \sum_{t=1}^T m_t^i} \\ &= \frac{T \cdot \Delta_{Q_t}}{N \cdot \epsilon \cdot \sum_{t=1}^T m_t^i} \end{aligned}$$

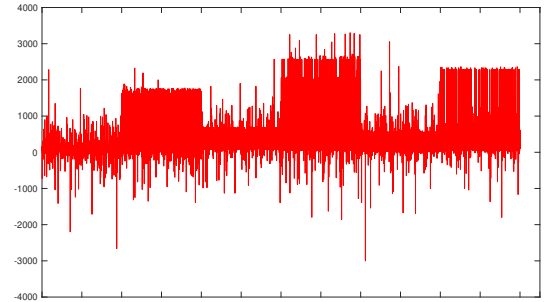
The accumulative error  $e_{ac}(i)$  is dependent on the magnitude of the injected noise and the billing cycle  $T$ . In addition, after aggregation, MDMU can obtain the sum of the total power consumption  $\sum_{i=1}^N \hat{m}_t^i$  at time point  $t$ . After the random shuffling, the measurement located originally at  $t$  moves to another time point for releasing and is denoted as  $\hat{m}_{t,out}^i$ , whereas the measurement at  $t$  that moves from another time



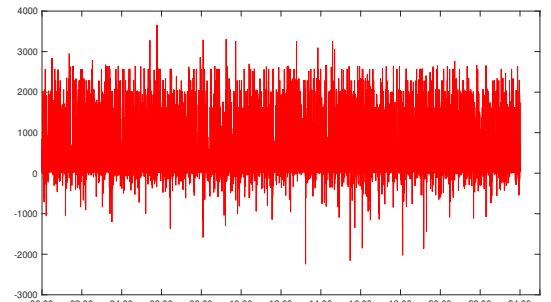
(a) Original measurement sequence



(b) Noisy sequence only with distributed Laplace noise



(c) Shuffled noisy sequence under  $T_w = 10s$



(d) Shuffled noisy sequence under  $T_w = 60s$

Fig. 3. Results of different steps in DDP



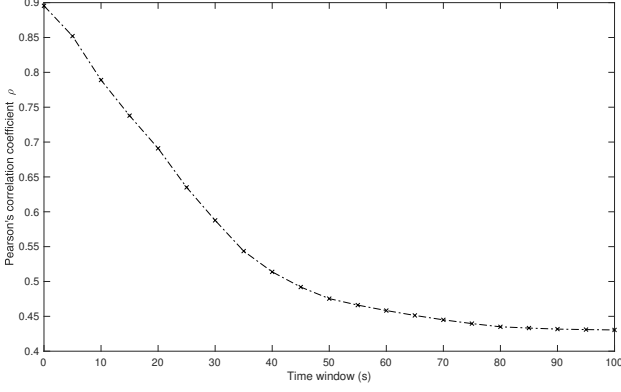


Fig. 4. Impact of time window  $T_w$  on correlation coefficient  $\rho$ .

point is denoted as  $\hat{m}_{t,in}^i$ . Then, the aggregation error  $e_{ag}(t)$  at time point  $t$  is computed as follows:

$$\begin{aligned}
 e_{ag}(t) &= \frac{1}{\sum_{i=1}^N m_t^i} E \left[ \sum_{i=1}^N \hat{m}_{t'}^i - \sum_{i=1}^N m_t^i \right] \\
 &= \frac{1}{\sum_{i=1}^N m_t^i} E \left[ \sum_{i=1}^N [\mathcal{G}(N, \lambda) - \mathcal{G}'(N, \lambda)] \right. \\
 &\quad \left. + \sum_{i=1}^N \hat{m}_{t,out}^i - \sum_{i=1}^N m_{t,in}^i \right] \\
 &= \frac{E \left[ Lap(\lambda) + \sum_{i=1}^N \hat{m}_{t,out}^i - \sum_{i=1}^N m_{t,in}^i \right]}{\sum_{i=1}^N m_t^i} \\
 &\approx \frac{\Delta Q_t}{\epsilon \cdot \sum_{i=1}^N m_t^i}
 \end{aligned}$$

The aggregation error  $e_{ag}(t)$  is mainly determined by the magnitude of the distributed Laplace noise and the difference of the measurement between  $t_{in}$  and  $t_{out}$ , where the difference is affected by the time window  $T_w$ .  $T_w$  is always much smaller than  $T_r$  (e.g.,  $T_w$  is generally set to 1 min, whereas  $T_r$  is set to 15 min), and the random shuffling operations in the range of  $T_w$  has negligible influence on the aggregated analysis at the interval of  $T_r$ . In other words, the aggregation error  $e_{ag}(t)$  is approximately determined by the magnitude of the distributed Laplace noise, which is far less than the magnitude of  $N \cdot Lap(\Delta Q/\epsilon)$  in centralized differentially private solutions. Consequently, our DDP mechanism can effectively preserve customer privacy with less errors and a high utility, and is independent of a trusted MDMU, thereby achieving a reasonable level of privacy assurance for smart meter data.

#### D. Feasibility Analysis

The feasibility analysis mainly focuses on three factors, namely robustness, complexity, and implementation cost.

Smart meters, as low-cost devices running in unprotected environments, are prone to failures [24]. If  $Y$  individual consumption reports are invalid or  $Y$  malfunctioning smart meters fails to submit their measurement, in our DDP mechanism, each one of the  $N-Y$  users injects  $\mathcal{G}(N-Y, \lambda) - \mathcal{G}'(N-Y, \lambda)$ , and the MDMU obtains the aggregated data as follows:

$$\begin{aligned}
 \sum_{i=1}^{N-Y} \hat{m}_{t'}^i &= \sum_{i=1}^{N-Y} m_{t'}^i + \sum_{i=1}^N [\mathcal{G}(N-Y, \lambda) - \mathcal{G}'(N-Y, \lambda)] \\
 &= \sum_{i=1}^{N-Y} m_{t'}^i + Lap(\Delta Q/\epsilon)
 \end{aligned}$$

Therefore, DDP still enforces  $\epsilon$ -differential privacy even if  $Y$  smart meters do not work correctly or if  $Y$  fault measurements are eliminated.

As for the computational overhead, the operations of distributed Laplace noise generation and injection and Fisher-Yates random shuffling have a low computational complexity of  $O(n)$  that is much lower than that of centralized and homomorphic encryption-based solutions. Moreover, these operations do not require additional data communications.

In addition, implementation cost is critical to a privacy-preserving mechanism because the excessive cost of smart meter modification would threaten the feasibility of DDP. A microcontroller unit (MCU) is a core component of a smart meter, which is characterized by its high performance, low capacity and accounts for the data processing, metering, storing, and exchanging processes. We take ATMELs SMART SAM4C microcontroller as an example. This device is constructed based on two 32-bit ARM Cortex-M4 RISC processors and can sufficiently to meet the demands of implementing DDP. Therefore, we can embed our DDP mechanism into the MCU to realize a privacy-aware sanitization of smart meter data. Compared with encryption-based mechanisms and device-based load hiding mechanisms, the DDP mechanism is more cost-friendly and easier to implement.

## VI. EXPERIMENTS

### A. Experimental Configurations and Datasets

We perform all experiments on a desktop PC with an Intel Quad-core i7-8700 @ 3.2 GHz CPU and 16 GB RAM. Every algorithm in each experiment is executed 100 times, and the average indicators are reported.

We use the public ECO dataset in the experiments to demonstrate the performance of DDP. This real-world dataset contains data on 6 households in Switzerland that are collected over 8 months (June 2012 to January 2013) and more than 100 million measurements reported at 1 Hz frequency during the period of deployment. Each measurement contains detailed information on voltage, current, phase shift between voltage and current, and occupancy. ECO also contains more than 650 million measurements from 45 smart plugs in total that are sampled for each appliance at 1 Hz frequency. ECO contributes to our evaluation and illustration of the effectiveness of our proposed mechanism in real-life applications.

### B. Experimental Evaluation on Privacy

We initially test the actual privacy-preserving intensity (i.e., the actual privacy budget  $\epsilon$ ) of the DDP mechanism under different scales of smart meters. We vary the initial value of  $\epsilon$  from 0.1 to 0.9 at 0.2 intervals and the scale size  $N$  from 100 to 1000. We also adopt the billing queries (that are essentially sum operations) in the ECO dataset. Table II shows the actual

TABLE II  
THE ACTUAL PRIVACY INTENSITY  $\epsilon'$  OF DDP

$N \backslash \epsilon$	0.1	0.3	0.5	0.7	0.9
100	0.093	0.291	0.494	0.699	0.874
300	0.089	0.287	0.487	0.690	0.859
500	0.086	0.281	0.483	0.686	0.849
700	0.085	0.278	0.479	0.682	0.836
900	0.080	0.277	0.475	0.674	0.804

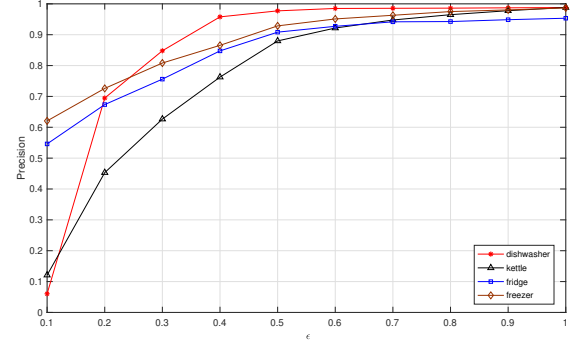
privacy intensity  $\epsilon'$  of DDP under different privacy budgets and scale sizes.

Specifically, when  $N = 100$ , the actual privacy budget  $\epsilon'$  of DDP is 0.093 given that the initial value of  $\epsilon$  is 0.1, whereas  $\epsilon'$  is 0.874 when  $\epsilon = 0.9$ , which indicates a greater privacy guarantee. In addition, as the scale size  $N$  of smart meters increases, the actual privacy budget  $\epsilon'$  decreases. For instance, given that the initial value of  $\epsilon$  is 0.1, the actual privacy budget is  $\epsilon' = 0.093$  when  $N = 100$ , and  $\epsilon'$  declines constantly from 0.089 to 0.077 when  $N$  increases from 300 to 900. Similarly, as the initial value of  $\epsilon$  is 0.5,  $\epsilon'$  declines constantly from 0.494 to 0.475 when  $N$  increases from 100 to 900. Therefore, our DDP mechanism can provide a reliable privacy guarantee and is applicable to large scale smart meters.

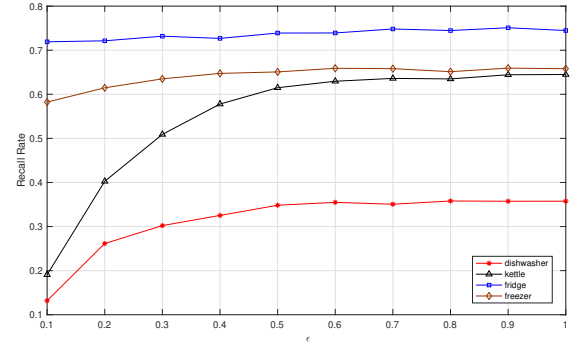
We also consider NILM attacks that use the Weiss algorithm to decompose appliance-level power load. Specifically, we select four appliances that can be optimally recognized by the Weiss algorithm, namely, dishwasher, kettle, fridge, and freezer, to evaluate the performance of the DDP mechanism in resisting NILM attacks. The dishwasher and kettle are high-power electrical appliances, whereas the fridge and freezer are low-power electrical appliances. The performance of the DDP mechanism is measured in terms of the precision, recall rate, and  $F$ -measure.

Firstly, we evaluate the recognition results of NILM attacks when only distributed Laplace noise is injected into the measurements without random shuffling operations, as illustrated in Fig. 5.

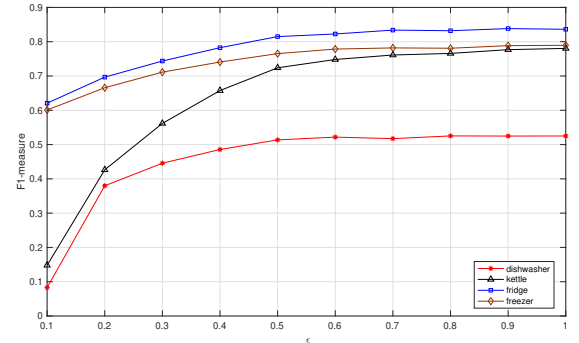
Fig. 5(a) shows that when  $\epsilon = 1$ , the recognition precisions of the NILM attacks under the proposed DDP mechanism to the dishwasher, kettle, fridge, and freezer are 0.99, 0.98, 0.95, and 0.98, respectively. However, as  $\epsilon$  decreases, the recognition precision decreases significantly. For instance, when  $\epsilon = 0.4$ , the precision values obtained for these appliances decline to 0.95, 0.76, 0.83, and 0.86, respectively, and declines further to 0.07, 0.12, 0.55, and 0.62, respectively, as  $\epsilon = 0.1$ , indicating that the suppressing effects of distributed Laplace noise on recognizing the high- and low-power electrical appliances surpass 90% and approach 40%. The similar suppressing effects can be observed for recall rate as shown in Fig. 5(b). Specifically, as  $\epsilon = 0.1$ , the recall rates of the proposed DDP mechanism for recognizing the dishwasher and kettle drop by 63.04% and 70.37%, respectively, compared with the recall rates obtained as  $\epsilon = 1$ . Although a distributed Laplace noise injection has a slight effect on low-power electrical appliances due to their special operational control, the recall rates also decrease to some degree. Meanwhile, as shown in Fig. 5(c), the  $F$ -measures of recognizing the four



(a) Precision



(b) Recall rate

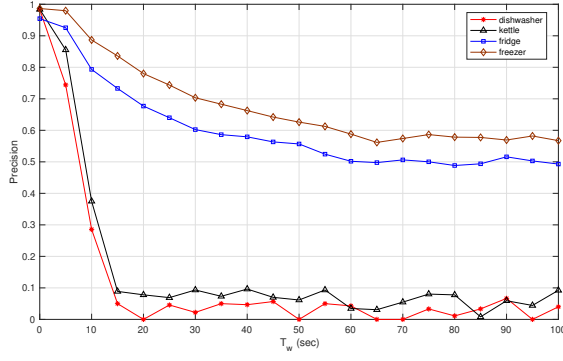


(c)  $F$ -measure

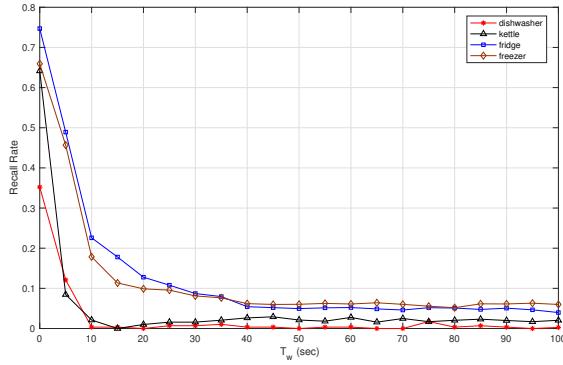
Fig. 5. Recognition results of NILM attack under only distributed Laplace noise.

appliances decrease from 0.54, 0.78, 0.85, and 0.80 to 0.08, 0.15, 0.62, and 0.59, respectively, as  $\epsilon$  decreases from 1 to 0.1. The suppressing effects on recognizing high-power electrical appliances reach 80%, whereas the effects on recognizing low-power electrical appliances are near 25%. In general, given that  $\epsilon < 0.5$ , injecting distributed Laplace noise can protect the power usage modes and suppress the recognition of high-power electrical appliances. However, a small  $\epsilon$  implies a large noise magnitude, which may introduce additional errors to the smart meter data and dramatically reduce the utility.

Then, we evaluate the recognition result of NILM attacks under our DDP mechanism, as shown in Fig. 6, where the time window  $T_w$  varies from 0 second to 100 seconds at a 5 seconds interval,  $\epsilon = 1$ , and the scale size is  $N = 100$ .



(a) Precision



(b) Recall rate

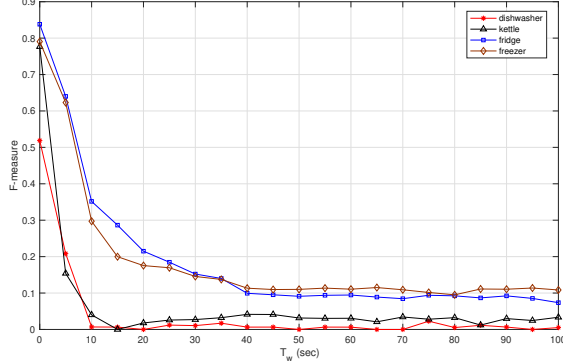
(c)  $F$ -measure

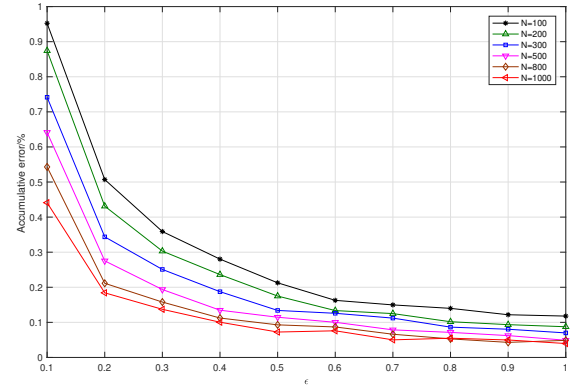
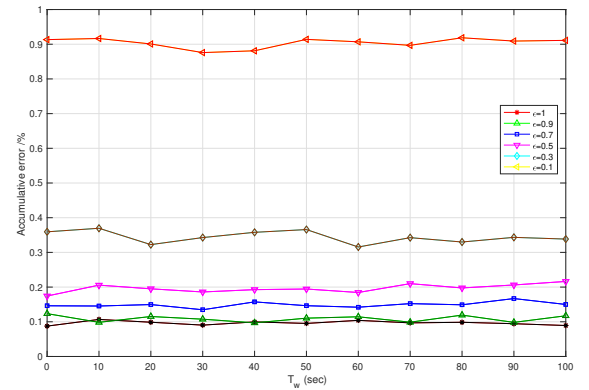
Fig. 6. Recognition results of NILM attack under DDP.

Evidently, a small shuffling time window can effectively suppress recognizing the power usage modes. When  $T_w = 0$ , the evaluation results are equivalent to those shown in Fig. 5, where the recognition precisions to dishwasher, kettle, fridge, and freezer are 0.99, 0.98, 0.95, and 0.98, respectively. When  $T_w = 10s$ , which means that the measurement sequence is randomly shuffled in a small time window, the precisions of recognizing these four appliances fall sharply to 0.28, 0.38, 0.78, and 0.87, respectively. As  $T_w$  increases to 60 s, the precisions further drop to 0.03, 0.04, 0.59, and 0.49, respectively, and the suppressing effects reach their optimum before becoming steady. The suppressing effects on recognizing high- and low-power electrical appliances approach 95% and 45%,

respectively. Fig. 6(b) illustrates that the random shuffling significantly lowers the recall rate. When  $T_w = 60s$ , the recall rates for recognizing the four appliances are 0.006, 0.02, 0.05, and 0.07, respectively, which approach 0 for the high-power electrical appliances and 0.05 for the low-power ones. Similarly, the  $F$ -measures of recognizing the four appliances at  $T_w = 60s$  are 0.006, 0.03, 0.08, and 0.11, respectively, which show evident declines compared with the results obtained at  $T_w = 0$ . Consequently, the suppressing effects of the proposed DDP mechanism on recognizing high- and low-power electrical appliances surpass 95% and reach 85%, respectively, because DDP damages the sequence correlations and hides the load characteristics of the original smart meter data, thereby preserving the privacy of customers.

### C. Experimental Evaluation on Utility

In this subsection, we measure the utility of the DDP mechanism in terms of the accumulative error  $e_{ac}$  and the aggregation error  $e_{ag}$ .

(a) Impact of  $\epsilon$  and  $N$  on  $e_{ac}$ (b) Impact of  $T_w$  on  $e_{ac}$  under  $N = 100$ Fig. 7. Average accumulative error  $e_{ac}$ .

In Fig. 7, we use the 8-month (247 days) power consumption data of houses 1 to 6, set the period of error statistics to 1 day, and calculate the average accumulative error of the multiple customers in a single day. Fig. 8(a) illustrates the changes of average accumulative error along with privacy

parameter  $\epsilon$  and scale size  $N$ . As  $N$  remains constant, a smaller  $\epsilon$  yields a higher average accumulative error, which can be inferred by the formula  $e_{ac}(i)$  in subsection IV. C. When  $N$  is fixed to 100, the average accumulative error is 0.13%, 0.21%, and 0.94%, respectively, as  $\epsilon$  is set to 1, 0.5, and 0.1, respectively, which always remain below 1%. In addition, with the increase of  $N$ , the magnitude of noise injected by a single customer decreases. Specifically, as  $\epsilon$  is fixed to 0.1, the average accumulative error reduces from 0.94% to 0.44% when  $N$  changes from 100 to 1000. Similarly, as  $\epsilon = 0.9$ , the average accumulative error reduces from 0.13% to 0.04% and approaches 0. Therefore, increasing the scale size  $N$  is beneficial in reducing the accumulative error, thereby achieving an accurate power metering and billing.

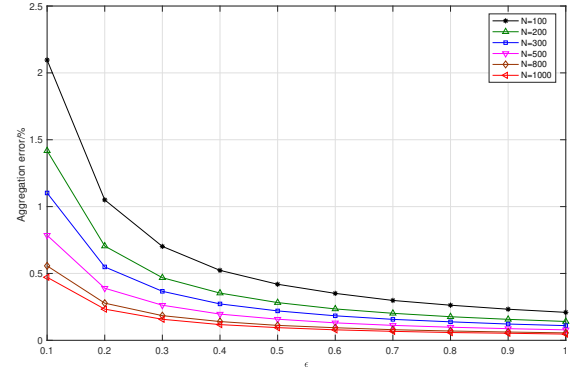
Afterwards, we fix the scale size to  $N = 100$ , and Fig. 7(b) shows that the average accumulative error is mainly correlated to the privacy parameter  $\epsilon$  and scale size  $N$ , and uncorrelated to the time window  $T_w$ , because the average accumulative error  $e_{ac}$  is caused by the noise injection, and the random shuffling realizes sequence permutation without adding any noise.

The results of the average aggregation error  $e_{ag}$  are similar to the average accumulative error, as shown in Fig. 8. As  $N$  remains constant, a smaller  $\epsilon$  leads to a higher average aggregation error. For example, when  $N$  is fixed to 100, the average aggregation error is 0.22%, 0.24%, 0.44%, and 2.1%, respectively, given that  $\epsilon$  is set to 1, 0.9, 0.5, and 0.1, respectively. The variation trend of  $e_{ag}$  is consistent with  $e_{ac}$ . Moreover, given that  $\epsilon = 0.1$ , the average aggregation error is 2.1% and 0.47%, respectively, thereby indicating that increasing the scale size of smart meters can reduce the aggregation error brought by noise injection. We can conclude that the decentralized mechanism can effectively ensure the utility while preserving the privacy of smart meter data. Fig. 8(b) also shows that the random shuffling does not cause any aggregation error. In sum, our proposed DDP mechanism can preserve the privacy of smart meter data while maintaining a high utility that can be improved by increasing scale size.

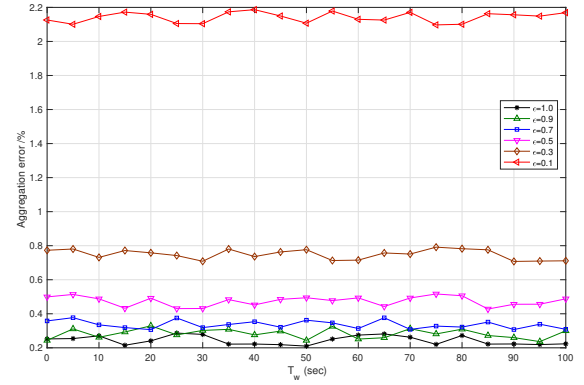
## VII. CONCLUSIONS

This study proposes DDP, a decentralized privacy-assured mechanism for smart meter data, to preserve the power measurements and hide the power usage modes. Specifically, DDP mechanism injects Laplace noise in a decentralized manner into the measurement at the end of each customer and applies a random permutation algorithm to shuffle the measurement sequence and to mask the power usage modes while enforcing  $\epsilon$ -differential privacy. Our extensive experiments demonstrate that the DDP mechanism can effectively resist NILM attacks and achieve outstanding utility in terms of accumulative and aggregation errors. This mechanism has also been proven to be applicable to large-scale smart meters.

In our future work, we will consider additional uses of smart meter data, including demand response, operations, and value-added services, and design the corresponding privacy-preserving mechanisms in a smart grid.



(a) Impact of  $\epsilon$  and  $N$  on  $e_{ag}$



(b) Impact of  $T_w$  on  $e_{ag}$  under  $N = 100$

Fig. 8. Average aggregation error  $e_{ag}$ .

## REFERENCES

- [1] Nicola Bui, Angelo P. Castellani, Paolo Casari, and Michele Zorzi, "The Internet of Energy: A Web-Enabled Smart Grid System", *IEEE Network*, vol. 26, no. 4, pp.39-45, July 2012.
- [2] Félix Gómez Mármol, Christoph Sorge, Osman Ugus, and Gregorio Martínez Pérez, "Do Not Snoop My Habits: Preserving Privacy in The Smart Grid", *IEEE Communications Magazine*, vol. 50, no. 5, pp. 166-172, May 2011.
- [3] Muhammad Rizwan Asghar, György Dán, Daniele Miorandi, and Imrich Chlamtac, "Smart Meter Data Privacy: A Survey", *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2820-2835, June 2017.
- [4] Costas Efthymiou, and Georgios Kalogridis, "Smart Grid Privacy via Anonymization of Smart Metering Data", in *Proceedings of 2010 the 1<sup>st</sup> IEEE International Conference on Smart Grid Communications*, Gaithersburg, MD, USA, IEEE, 2010, pp. 238-243.
- [5] Cristina Rottondi, Giulia Mauri, and Giacomo Verticale, "A Data Pseudonymization Protocol for Smart Grids", in *Proceedings of 2012 IEEE Conference on Green Communications (GreenCom)*, Piscataway, NJ, USA, IEEE, 2012, pp. 68-73.
- [6] Haiyong Bao and Rongxing Lu, "A New Differentially Private Data Aggregation with Fault Tolerance for Smart Grid Communications", *IEEE Internet of Things Journal*, vol. 2, no. 3, pp. 248-2493, June 2015.
- [7] Jianbing Ni, Kuan Zhang, Khalid Alharbi, Xiaodong Lin, Ning Zhang and Xuemin Shen, "Differentially Private Smart Metering with Fault Tolerance and Range-Based Filtering", *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2483-258, September 2017.
- [8] Prosanta Gope, and Biplab Sikdar, "An Efficient Data Aggregation Scheme for Privacy-Friendly Dynamic Pricing-Based Billing and Demand-Response Management in Smart Grids", *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3126-3135, August 2018.
- [9] Zhitao Guan, Guanlin Si, Xiaosong Zhang, Longfei Wu, Nadra Guizani, Xiaojiang Du, and Yinglong Ma, "Privacy-Preserving and Efficient Aggregation Based on Blockchain For Power Grid Communications in

- Smart Communities”, *IEEE Communications Magazine*, vol. 56, no. 7, pp. 82-88, July 2018.
- [10] Liehuang Zhu, Zijian Zhang, Zhan Qin, Jian Weng and Kui Ren, “Privacy Protection Using a Rechargeable Battery for Energy Consumption in Smart Grids”, *IEEE Network*, vol. 31, no. 1, 59-63, January 2017.
  - [11] Farhad Farokhi, and Henrik Sandberg, “Fisher Information as a Measure of Privacy: Preserving Privacy of Households with Smart Meters Using Batteries”, *IEEE Transactions on Smart Grid*, vol. 9, no. 5, 4726- 4734, September 2018.
  - [12] Simon Li, Ashish Khisti, and Aditya Mahajan, “Information-Theoretic Privacy for Smart Metering Systems with a Rechargeable Battery”, *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3679-3695, May 2018.
  - [13] Onur Tan, Deniz Gunduz, and H. Vincent Poor, “Increasing Smart Meter Privacy Through Energy Harvesting and Storage Devices”, *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1331-341, July 2013.
  - [14] Yanan Sun, Lutz Lampe and Vincent W. S. Wong, “Smart Meter Privacy: Exploiting the Potential of Household Energy Storage Units”, *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 69-78, February 2018.
  - [15] Lingjuan Lyu, Karthik Nandakumar, Ben Rubinstein, Jiong Jin, Justin Bedo, and Marimuthu Palaniswami, “PPFA: Privacy Preserving Fog-Enabled Aggregation in Smart Grid”, *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3733-3744, August 2018.
  - [16] Jing Zhao, Taeho Jung, Yu Wang and Xiangyang Li, “Achieving Differential Privacy of Data Disclosure in the Smart Grid”, in *Proceedings of IEEE Conference on Computer Communications (INFOCOM’14)*, Toronto, ON, Canada, 2014, pp. 504-512.
  - [17] Zijian Zhang, Zhan Qin, Liehuang Zhu, Jian Weng and Kui Ren, “Cost-Friendly Differential Privacy for Smart Meters: Exploiting the Dual Roles of the Noise”, *IEEE Transactions on Smart Grid*, vol.8, no.2, pp. 619-626, March 2017.
  - [18] Jelasity, Márk, and Kenneth P. Birman. "Distributional Differential Privacy for Large-Scale Smart Metering," in *Proceedings of the 2<sup>nd</sup> ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*, Salzburg, Austria, 2014, pp. 141-146.
  - [19] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini, “The ECO Data Set and The Performance of Non-Intrusive Load Monitoring Algorithms,” in *Proceedings of the 1<sup>st</sup> ACM Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys’14)*, Memphis, Tennessee, USA, 2014, pp. 80-89.
  - [20] Michael A. Devlin, and Barry P. Hayes, “Non-Intrusive Load Monitoring and Classification of Activities of Daily Living Using Residential Smart Meter Data,” *IEEE Transactions on Consumer Electronics*, vol. 65, no. 3, pp. 339-348, August 2019.
  - [21] Bo Zhang, Lei Yang, Kai Wang, and Yuqiang Cao, "Block Compressed Sensing Using Two-dimensional Random Permutation for Image Encryption-then-Compression Applications," in *Proceedings of the 14<sup>th</sup> IEEE International Conference on Signal Processing (ICSP’18)*, Beijing, China, 2018, pp. 312-316.
  - [22] Mazhar Tayel, George Dawood, and Hamed Shawky, "Block Cipher S-box Modification Based on Fisher-Yates Shuffle and Ikeda Map," in *Proceedings of IEEE 18<sup>th</sup> International Conference on Communication Technology (ICCT’18)*, Chongqing, China, 2018, pp. 59-64.
  - [23] Ying-Chao Hung, and George Michailidis, "Modeling and Optimization of Time-of-Use Electricity Pricing Systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4116-4127, July 2019.
  - [24] Kaiping Xue, Qingyou Yang, Shaohua Li, David S. L. Wei, Min Peng, Imran Memon, and Peilin Hong, "PPSO: A Privacy-Preserving Service Outsourcing Scheme for Real-Time Pricing Demand Response in Smart Grid," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2486-2496, April 2019.
  - [25] Joppe W. Bos, Wouter Castryck, Ilia Iliashenko, and Frederik Vercauteren, "Privacy-Friendly Forecasting for the Smart Grid Using Homomorphic Encryption and the Group Method of Data Handling", in *Proceedings of the 9<sup>th</sup> International Conference on Cryptology in Africa (AFRICACRYPT’17)*, Dakar, Senegal, 2017, pp. 184-201.
  - [26] Lyu L, Karthik N, Ben R, Jin J, Justin B, and Marimuthu P, “PPFA: Privacy Preserving Fog-Enabled Aggregation in Smart Grid”, *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3733-3744, 2018.