Hindawi
Security and Communication Networks
Volume 2021, Article ID 6643566, 21 pages
https://doi.org/10.1155/2021/6643566

Research Article

Privacy-Preserving Publication of Time-Series Data in Smart Grid

Franklin Leukam Lako,[1,2] **Paul Lajoie-Mazenc**,[2] **and Maryline Laurent**[1]

[1]*Department of Network and Telecommunication Services, Samovar Lab, Télécom SudParis, Institut Polytechnique de Paris, 19 rue Marguerite Perey, Palaiseau 91120, France*
[2]*EDF Lab Paris-Saclay, 7 Boulevard Gaspard Monge, Palaiseau 91120, France*

Correspondence should be addressed to Franklin Leukam Lako; franklin.leukam-lako@edf.fr

Received 4 December 2020; Revised 18 February 2021; Accepted 2 March 2021; Published 26 March 2021

Academic Editor: Chalee Vorakulpipat

The collection of fine-grained consumptions of users in the smart grid enables energy suppliers and grid operators to propose new services (e.g., consumption forecasts and demand-response protocols) allowing to improve the efficiency and reliability of the grid. These services require the knowledge of aggregate consumption of users. However, an aggregate can be vulnerable to reidentification attacks which allow revealing the users' individual consumption. Revealing an aggregate data is a key privacy concern. This paper focuses on publishing an aggregate of time-series data such as fine-grained consumptions, without indirectly disclosing individual consumptions. We propose novel algorithms which guarantee differential privacy, based on the discrete Fourier transform and the discrete wavelet transform. Experimental results using real data from the Irish Commission for Regulation of Utilities (CRU) demonstrate that our algorithms achieve better utility than previously proposed algorithms.

1. Introduction

A smart city is a designation given to a city that incorporates information and communication technologies to enhance the quality and performance of urban services such as energy, transportation, and utilities in order to reduce resource consumption, wastage, and overall costs. The overarching aim of a smart city is to enhance the quality of living for its citizens through smart technology [1–3].

The smart grid is an important part of the smart city. Indeed, the smart grid allows greater penetration of highly variable renewable energy sources such as solar and wind power in the smart city.

The smart grid modernizes the traditional electricity grid by establishing a communication infrastructure in parallel to the energy delivery network. This infrastructure is used by the grid operators and suppliers to remotely collect fine-grained consumptions from household smart meters and to provide new energy services such as consumption forecasts or demand-response. These services are suitable for improving the efficiency and reliability of the grid, saving energy and, more generally, for optimizing energy usage. In particular, forecasting enables the supplier to predict future consumptions based on past aggregate data in order to improve the grid and retail operations and enhance energy trading [4], while demand-response (DR) aims to shift the users' consumption from peak to off-peak periods in order to avoid consumption peaks in the smart city.

However, aggregates are vulnerable to reidentification attacks, such as set difference attacks [5] in which two aggregates that differ by a single consumer allow learning this individual consumption. Since the individual consumption data collected by smart meters reflect the use of all electric appliances by inhabitants in a household over time and enable to deduce the behaviors, activities, age, or preferences of the inhabitants [6–11], revealing an aggregate is a key privacy concern.

Differential privacy (DP) [12] allows publishing an aggregate data while guaranteeing that an attacker does not learn any individual inputs from the aggregate. However, the noise added by DP often leads to a loss of utility. Moreover, publishing time-series data such as users' consumption, which are correlated, by using DP, results in more noise added than publishing a single aggregate for the same

privacy guarantee. Thus, disclosing time-series data leads to more loss of utility. Utility can be improved by increasing the size of the aggregate. Eibl and Engel [13] showed that for real-world smart metering, the aggregation group size must be of the order of thousands of smart meters in order to have reasonable utility. This paper shows how to obtain good utility with a group size smaller than 600. We obtain a mean relative error lower than 10% between the original data and the published one, which is considered practically suitable by energy experts for consumption forecasts.

The Laplace mechanism [14] is a popular mechanism to enable DP, by adding independent and identically distributed (IID) Laplace noise to each component of the time-series. However, adding IID noise for correlated time-series is not appropriate. In fact, an adversary can use refinement methods, such as filtering, to sanitize the IID noise and improve the probability of disclosing individual data [15, 16].

This paper focuses on disclosing an aggregate of users' consumption data without learning individual data and proposes methods with improved utility. We summarize our contributions as follows:

(i) We revisit the Fourier perturbation algorithm (FPA) [17] in order to correct some mistakes leading to poor users' privacy protection. We show that, in order to ensure the desired budget of privacy ϵ, a factor $\sqrt{2T}$ must be added to the noise, where T is the size of the time-series. However, this reduces the utility of FPA.

(ii) We propose the "clamping Fourier perturbation algorithm (CFPA)" using the clamping mechanism proposed in [18], for reducing the sensitivity, and thus the noise introduced in FPA. This new algorithm is an improvement of the Fourier perturbation algorithm (FPA). Experimental results show a utility improvement by a factor more than 6.

(iii) We also propose the "clamping wavelet perturbation algorithm" (CWPA), a similar adaptation of wavelet perturbation algorithm (WPA) [19], with a utility improvement by a factor 2.

(iv) We compare FPA, CFPA, WPA, and CWPA by analyzing their relative errors on a real dataset, and we explain why CFPA obtains the best utility.

The remainder of this paper is structured as follows. Section 2 provides an overview of the literature, while Section 3 presents preliminaries. Section 4 correctly computes the sensitivity of DFT in order to make FPA ϵ-differentially private. Section 5 details our privacy-preserving publication techniques using clamping mechanism, DFT, and DWT. Section 6 reports our experimental results. Section 7 concludes the paper.

Table 1 lists the acronyms used in this paper.

Table 1: List of acronyms.

Acronym	Meaning
CFPA	Clamping Fourier perturbation algorithm
CWPA	Clamping wavelet perturbation algorithm
DFT	Discrete Fourier transform
DP	Differential privacy
DWT	Discrete wavelet transform
FPA	Fourier perturbation algorithm
MRE	Mean relative estimation error
WPA	Wavelet perturbation algorithm

2. Related Work

Demand-response protocols [20–23], and secure aggregation protocols [24–30] aim to protect the privacy of users while supporting energy services such as demand-response, smart metering, billing, or forecasting.

In this paper, we investigate tools enabling forecasting and demand-response. In particular, we are interested in publishing an aggregate of individual consumptions, while preserving privacy.

Differential privacy (DP), introduced by Dwork in 2006, guarantees that the publication of an aggregate does not indirectly reveal the individual data [12]. Moreover, DP guarantees that two aggregates that differ by a single consumer are almost indistinguishable. DP has evolved overtime [31] and was adopted by organizations such as the US Census Bureau [32], Google [33], Apple [34], and Microsoft [35]. The Laplace mechanism [14] is a popular mechanism that allows guaranteeing DP by adding a noise drawn from the Laplace distribution $\mathscr{L}(\cdot)$ to the aggregate.

The Laplace mechanism takes as input two parameters: the privacy budget ϵ and the sensitivity of the function to publish (in our case, the sum of users' consumption). Smaller values of ϵ lead to a better protection, but add a bigger noise to the aggregate.

Utility can be improved by increasing the size of the aggregate in order that the effect of noise is small enough that the result can be utilized. Eibl and Engel [13] showed that for real-world smart metering, the aggregation group size must be of the order of thousand smart meters in order to have reasonable utility. This paper shows how to obtain good utility with a group size smaller than a thousand.

DP is typically applied to static data, i.e., to a single query. In this paper, we consider time-series consumption, which is equivalent to multiple queries on correlated data. Applying the Laplace mechanism independently to each data point of the time-series is not appropriate. Indeed, an adversary can use refinement methods, such as filtering, to sanitize the Laplace noise and improve the probability disclosing individual data [15, 16]. Thus, the data points of the time-series are correlated. The composition theorem [14] states that the privacy budget ϵ of T correlated queries adds up, i.e., setting the privacy budget for a single query to $\varepsilon_q = 0.5$, the privacy budget of $T = 48$ single queries (corresponding to a day profile with a time interval of 30 min) is $\varepsilon = 0.5 \times 48 = 24$. In order to guarantee a global privacy budget of ε, one solution is to set the privacy budget of each query to ε/T. Of course, this leads to more noise in the aggregate and a loss of utility.

One method to guarantee DP for correlated time-series data publishing consists in transforming the original correlated time-series into another representation while

maintaining its major characteristics before adding the Laplace noise. Rastogi and Nath [17] proposed the Fourier perturbation algorithm (FPA) that combines discrete Fourier transform (DFT) with DP to support time-series of count queries while not disclosing any individual data and ensuring good utility. We note that the sensitivity of count queries is 1, and the global sensitivity is T for a time-series of length T. Ács et al. [36] proposed an optimization of the FPA allowing to release histograms, where the global sensitivity is 1. They show through experimental evaluation that their scheme improves the utility of the initial FPA by a factor 10. Lyu et al. [19] applied FPA to time-series consumptions and proposed wavelet perturbation algorithm (WPA) by replacing DFT by discrete wavelet transform (DWT). The authors show through experimental results that WPA ensures better utility than FPA.

We apply these approaches to time-series of consumption data and refine them by reducing the sensitivity of the queries in order to reduce the relative error of the final result.

3. Preliminaries

3.1. System and Threat Model. The entities involved in this paper are as follows:

(i) Trustworthy homes, which smart meter (SM) enables to collect their true individual time-series consumption.

(ii) A honest aggregator, which collects users' individual consumption, and which publishes an aggregate time-series consumption of users to a forecaster in a privacy-preserving way for the forecaster not to be able to deduce any individual consumption of users.

(iii) A forecaster, which predicts future consumptions based on the aggregate consumption received in order to improve the grid and retail operations and enhance energy trading. The forecaster is considered honest-but-curious as it provides appropriate forecasts, but it may attempt to infer the users' individual consumption from the aggregate in order to deduce the behaviors, activities, age, or preferences of the inhabitants.

Figure 1 depicts the system model. In a real scenario, the aggregator can be an energy distributor, and the forecaster can be a municipality that seeks to find out the total consumption of the inhabitants of the municipality.

Considering the case where the aggregator and the forecaster belong to two entities of the same energy provider, the publication of aggregate users' consumption to forecasters in a privacy-preserving way reduces the risk of disclosing users' individual consumption. Moreover, this avoids the need for forecasters to ask for explicit consent from customers in accordance with the GDPR [37] to process their personal data.

Let N be the number of smart meters (SMs) in a district. Let $X^j = (x_1^j, x_2^j, \ldots, x_T^j)$ be the time-series of energy consumptions collected by SM j, where x_t^j is the consumption at time slot t $(t = 1, \ldots, T)$ collected by SM j $(j = 1, \ldots, N)$, with T being the time period considered. Each time-series consumption X^j is sent to an aggregator who computes the following aggregate:

$$S = (S_1, \ldots, S_T) = \left(\sum_{j=1}^{N} x_1^j, \sum_{j=1}^{N} x_2^j, \ldots, \sum_{j=1}^{N} x_T^j \right). \tag{1}$$

To reveal S to a forecaster without indirectly disclosing individual consumptions X^j $(j = 1, \ldots, N)$, the aggregator can use differential privacy (DP).

3.2. Differential Privacy. Differential privacy is a framework introduced by Dwork allowing quantifying the privacy guarantees of a request on a database [38]. This request can be the publication of a database, or a more precise one, such as "what is the sum of energy consumptions of users in this database?".

A request on databases is said to be differentially private if this request makes two similar databases indistinguishable from looking only at the output of the request. Differential privacy relies on a parameter, noted ε, called the privacy budget. The formal definition of a differentially private algorithm is given as follows.

Definition 1 (ε–differentially private). A request $\mathscr{A}: \mathscr{D} \longrightarrow \mathscr{S}$ is ε–differentially private if and only if for all databases $D_1, D_2 \in \mathscr{D}$ differing by at most one record, and for all subsets $O \subset \mathscr{S}$,

$$\Pr(\mathscr{A}(D_1) \in O) \leq \exp(\varepsilon) \Pr(\mathscr{A}(D_2) \in O). \tag{2}$$

This definition can be applied not only to requests on databases but also to any function, by considering the domain of the function as a database format.

Dwork also proposes the Laplace mechanism, which allows making any (vectors of) real-valued function ε-differentially private [38]. This mechanism relies on the notion of sensitivity of a function, which represents how a single record of the database can influence the output of the function.

Definition 2 (sensitivity). Let $f: \mathscr{D} \longrightarrow \mathbb{R}^d$ be a function; the sensitivity of f is

$$\Delta_1(f) = \max_{D_1, D_2 \in \mathscr{D} \text{ s.t. } d(D_1, D_2) \leq 1} \left\| f(D_1) - f(D_2) \right\|_1. \tag{3}$$

This sensitivity is also called L_1-sensitivity due to the L_1–norm used in its definition and is denoted by $\Delta_1(f)$. Similarly, the L_2-sensitivity used later and denoted by ε is computed using the L_2–norm (the L_1–norm and the L_2 norm of a vector $S = (s_1, \ldots, s_T)$ are respectively equal to $\|S\|_1 = \sum_{j=1}^{T} |s_j|$ and $\|S\|_2 = \sqrt{\sum_{j=1}^{T} s_j^2}$).

The Laplace mechanism consists of adding a random value to the original result of the query, where the random

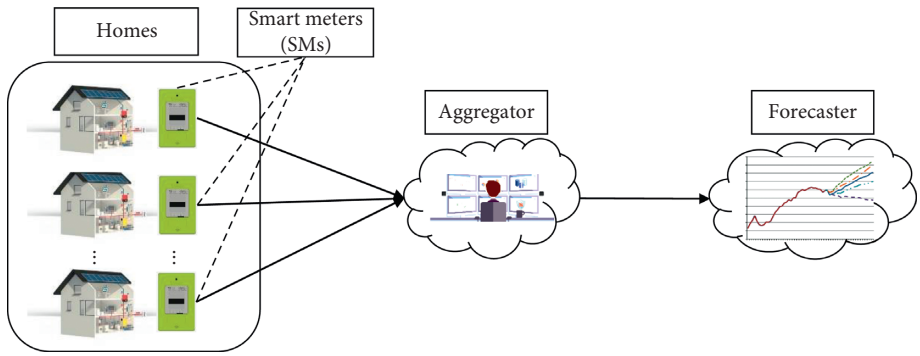

Figure 1: System model.

value follows the Laplace distribution, where the parameter depends on the chosen ϵ and on the sensitivity of the function, as follows.

Theorem 1 (Laplacian mechanism). *For all functions* $f: \mathcal{D} \longrightarrow \mathbb{R}^d$, *the algorithm* $\mathcal{A}(D) = f(D) + (\mathcal{L}_1(\Delta_1(f)/\varepsilon), \ldots, \mathcal{L}_d(\Delta_1(f)/\varepsilon))$ *is* ϵ*–differentially private, where* $\mathcal{L}(\cdot)$ *is the distribution of Laplace and* $\Delta_1(f)$ *is the sensitivity of* f.

DP introduces noise in order to guarantee privacy. This noise can decrease the utility of the function. We quantify this loss using mean relative estimation error (MRE), defined as follows.

Definition 3 (mean relative estimation error). The mean relative estimation error (MRE) between two vectors a and b of size T is $1/T \cdot \sum_{j=1}^{T} |a_j - b_j|/a_j + 12$ (we add 1 to the denominator in order to avoid dividing by zero. This definition is also used in [27]).

Consider the aggregate $S = (S_1, \ldots, S_T)$ defined in (1). Let M be the maximum consumption in the domain. One naive solution to publish S without revealing any individual consumption is to use the Laplace mechanism to add independent Laplace noise to each component of S and to release the results: $\widehat{S} = (S_1 + \mathcal{L}(M \cdot T/\varepsilon), \ldots, S_T + \mathcal{L}(M \cdot T/\varepsilon))$, where the sensitivity of the sum of time-series consumption is $M \cdot T$. However, this simple approach leads to excessive noise rendering the aggregate useless [13].

Example 1. Figures 2(a) and 2(b), respectively, show the aggregated consumption of 250 homes from December 30th, 2009, to January 5th, 2010, taken from the CER dataset [39], and its noisy version using the naively applied Laplace mechanism, with $\varepsilon = 1$ per day. Figure 2(a) shows two consumption peaks at 12 am and 6 pm which respectively correspond to lunch and dinner time. We also observe that in the night (from 12 pm to 6 am) the consumption decreases. Figure 2(b) shows that the noisy version is completely different from the original aggregate (Figure 2(a)). In this example, the MRE between the aggregate consumption and the noisy version is 141%, which is not usable. Moreover, the noisy version has inconsistent values such as negative consumptions.

Rastogi and Nath [17] introduce the Fourier perturbation algorithm (FPA) and show that is an effective tool for reducing the noise introduced by the Laplace mechanism for time-series. Section 3.3 presents the FPA. However, there are some mistakes in this version relying on the estimation of the FPA sensitivity. These mistakes are presented in Section 4, along with the corrected FPA.

Table 2 lists the symbols used in the rest of the paper.

3.3. Fourier Perturbation Algorithm. The Fourier perturbation algorithm (FPA) presented in [17, 19, 36] takes as input a time-series $S = (S_1, \ldots, S_T)$ and an integer $k \ll T$ and returns the noisy time-series $\widehat{S} = (\widehat{S}_1, \ldots, \widehat{S}_T)$, as shown in Algorithm 1.

Rastogi and Nath [17] show that FPA is ε–differentially private. However, there are some mistakes in their proof of Theorem 4.1 of [17] which justified that FPA is ε-differentially private. These mistakes rely on the estimation of the FPA sensitivity and are presented in Section 4.

3.4. Wavelet Perturbation Algorithm. By replacing the DFT with the discrete Haar wavelet transform (DWT), Lyu et al. [19] proposed the wavelet perturbation algorithm (WPA) and showed that WPA guarantees better utility than DFT. Algorithm 2 describes WPA.

Figure 3 shows the same aggregated consumption presented in Example 1 and its noisy version using WPA (Algorithm 2) with Haar wavelet, $\varepsilon = 1$ per day and $k = 5$. In Figure 3, the MRE is however higher than 10% (18%). In the noisy aggregate, the first peak of the morning is masked and the peak of the evening is truncated, as well as the trough of the night.

Theorem 2. *Wavelet perturbation algorithm (WPA) is* ε*-differentially private.*

Proof. DWT is orthonormal [40], i.e., W has the same L_2 norm as S, that is, $\Delta_2(W) = \Delta_2(S)$. Furthermore, $\Delta_2(W^k) \leq \Delta_2(S)$ (because $T - k$ DWT coefficients of W are set to 0). With the inequality of norm, $\Delta_1(W^k) \leq \sqrt{k}\Delta_2(W^k)$. Then, $\Delta_1(W^k) \leq \sqrt{k}\Delta_2(S) \leq M \cdot \sqrt{kT}$. Thus, the noise

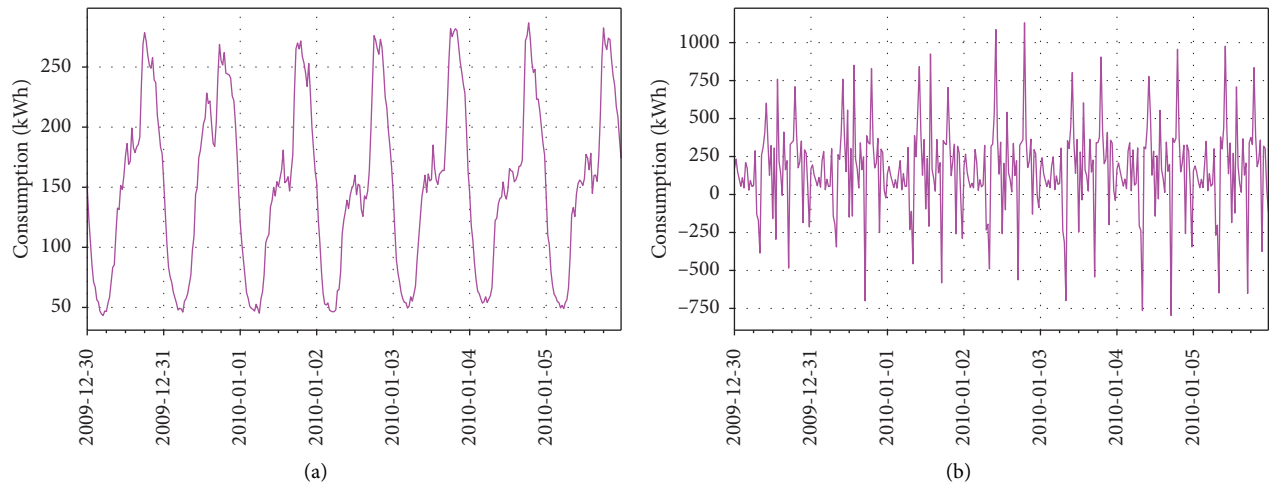

Figure 2: Aggregated time-series consumption of 250 homes from December 30th, 2009, to January 5th, 2010, taken from the CER dataset [39], and its noisy version using the naively applied Laplace mechanism, with $\varepsilon = 1$ per day. (a) Aggregated consumption. (b) Noisy version using the naive solution.

Table 2: List of symbols.

Notation	Description
N	Number of smart meters (SMs) in the district
M	Maximum consumption in the dataset
T	Time period during the collection of time-series consumption
$X^j = (x_1^j, x_2^j, \ldots, x_T^j)$	Time-series of energy consumptions collected by SM j, where x_t^j is the consumption at time slot t ($t = 1, \ldots, T$) collected by SM j ($j = 1, \ldots, N$)
$S = (S_1, \ldots, S_T)$	Sum of users' time-series consumptions to be published, where $S_t = \sum_{j=1}^{N} x_t^j$ for $t = 1, \ldots, T$
ε	Budget of privacy
$\widehat{S}$	Noisy version of S
k	Number of the first DFT or DWT coefficients conserved in the Fourier perturbation algorithm (FPA), wavelet perturbation algorithm (WPA), clamping Fourier perturbation algorithm (CFPA), and clamping wavelet perturbation algorithm (CWPA)
Δ_1	L_1-sensitivity
Δ_2	L_2-sensitivity

Inputs: $S = (S_1, \ldots, S_T)$, k, the maximum consumption M of the domain, and the privacy budget ϵ.
(1) Compute the discrete Fourier transform of S: $F = \mathrm{DFT}(S)$.
(2) Keep only the first k coefficients of F, denoted by F^k.
(3) Generate the noisy version of F^k, denoted by $\widehat{F}^k$ by adding a Laplace noise $\mathscr{L}(M\sqrt{Tk}/\varepsilon)$ to each coefficient in F^k.
(4) Pad $\widehat{F}^k$ to a T-dimensional vector, denoted by $\mathrm{PAD}^T(\widehat{F}^k)$ by appending $T - k$ zeroes.
(5) Apply the inverse DFT to $\mathrm{PAD}^T(\widehat{F}^k)$ to obtain a noisy version of S denoted by $\widehat{S}$.

Algorithm 1: Fourier perbutation algorithm [17].

Inputs: $S = (S_1, \ldots, S_T)$, k, the maximum consumption M of the domain, and the privacy budget: ε
(1) Compute the DWT coefficients of S: $W = \mathrm{DWT}(S)$.
(2) Keep only the first k coefficients of W, denoted by W^k.
(3) Generate the noisy version of W^k, denoted by $\widehat{W}^k$ by adding a Laplace noise $\mathscr{L}(M\sqrt{Tk}/\varepsilon)$ to each coefficient in W^k.
(4) Pad $\widehat{W}^k$ to a T-dimensional vector, denoted by $\mathrm{PAD}^T(\widehat{W}^k)$ by appending $T - k$ zeroes.
(5) Apply the inverse DWT to $\mathrm{PAD}^T(\widehat{W}^k)$ to obtain a noisy version of S denoted by $\widehat{S}$.

Algorithm 2: Wavelet perbutation algorithm [19].

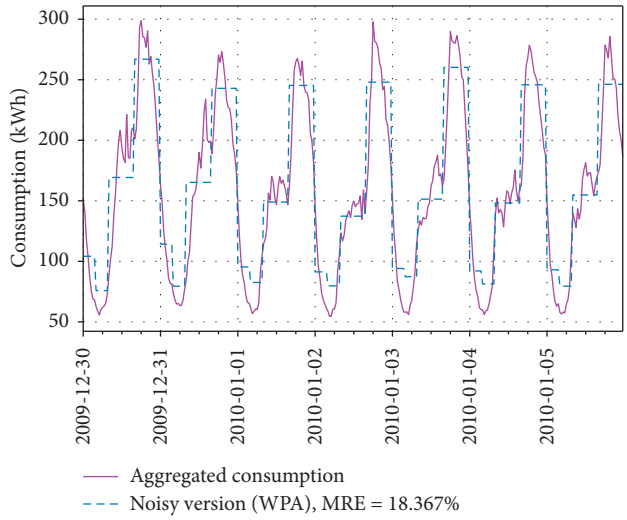

FIGURE 3: Aggregated consumption of 250 homes from 30th December 2009 to 5th January 2010 of dataset from CER [39] and its noisy version using WPA (Algorithm 2), with Haar wavelet, $\varepsilon = 1$ per day and $k = 5$.

introduced in Step 3 is justified and WPA guarantees ε–differential privacy. □

4. Correctly Estimating the Sensitivity of FPA

In [17], authors show that FPA, as described in Section 3, guarantees ε–differential privacy. The authors estimated the sensitivity of DFT to be $M\sqrt{Tk}$, while it should be $MT\sqrt{2k}$, with T being the size of the time-series and M being the maximum consumption in the domain. Thus, for a given privacy budget ε, the utility of FPA is worse than presented in [17].

This section correctly computes the sensitivity of DFT, which allows to make render FPA ε–differential private. Before that we recall the definition of DFT.

4.1. Discrete Fourier Transform (DFT). Let $S = (S_1, \ldots, S_T)$ be a time-series. DFT takes S as input and returns a time-series of T complex numbers $F = (F_1, \ldots, F_T)$ such that

$$F_k = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} S_j e^{-2\pi i (j-1)(k-1)/T} \quad \text{for } k = 1, \ldots, T, \tag{4}$$

where $i^2 = -1$. The inverse of the DFT is computed as follows:

$$S_k = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} F_j e^{-2\pi i (j-1)(k-1)/T}, \quad \text{for } k = 1, \ldots, T. \tag{5}$$

This version of the DFT is normalized, that is, $\|\mathrm{DFT}(S)\|_2 = \|S\|_2$.

DFT can be defined in other ways, for instance, the $1/\sqrt{T}$, present in both the DFT and the inverse definitions above, can be replaced by a factor 1 in the DFT and $1/T$ in the inverse DFT. In that case, the DFT is not normalized.

In [17, 19, 36], the authors use the latter version of DFT, which is not normalized. However, the sensitivity computation relies on the equality $\|\mathrm{DFT}(S)\|_2 = \|S\|_2$, while it should be $\|\mathrm{DFT}(S)\|_2 = \sqrt{T} \cdot \|S\|_2$. Thus, the correct total privacy budget is $\sqrt{T} \cdot \varepsilon$ instead of ε. This is the first mistake in this approach and can be resolved by using the normalized DFT.

Another error lies in the fact that the Laplacian mechanism is only applied to the real part of the Fourier coefficients, which are complex numbers. This mistake can be resolved by applying the Laplace mechanism to both real and imaginary parts of the Fourier coefficients.

The following section computes the sensitivity of the DFT, and thus of FPA, and takes into account those two errors.

4.2. Sensitivity of the DFT. Let DFT^k be the function which takes a time-series $S = (S_1, \ldots, S_T)$ as input and returns the first k DFT coefficients of S. This function can be seen as a DFT^k: $\mathbb{R}^T \longrightarrow \mathbb{R}^{2k}$, the function which returns the real and imaginary parts of the first k Fourier coefficients. This function is a real-valued function, we can thus use the Laplace mechanism on it. First, we need to compute the L_1-sensitivity of DFT^k.

Lemma 1. *Let DFT^k be defined as follows:*

$$\begin{aligned} &\mathrm{DFT}^k \colon \mathbb{R}^T \longrightarrow \left(\mathbb{R}^2\right)^k \\ &S \longmapsto \mathrm{DFT}^k(S) = \left((a_1, b_1), \ldots, (a_k, b_k)\right). \end{aligned} \tag{6}$$

We denote $c_j = a_j + ib_j$ the j-th coefficient of $\mathrm{DFT}(S)$, with $i^2 = -1$ and $j = 1, \ldots, k$. a_j and b_j respectively represent the real and imaginary parts of c_j.

The L_1-sensitivity of DFT^k, $\|\mathrm{DFT}^k(S)\|_1$, is $M \cdot \sqrt{2Tk}$ when the DFT is normalized (respectively, $\mathrm{MT} \cdot \sqrt{2k}$ when

the DFT is not normalized), with M as the maximum value in the dataset.

Proof. Let DFT^k be defined as in Lemma 1.

$$\begin{aligned}
\left\|\mathrm{DFT}^k(S)\right\|_1 &= \left\|(a_1, b_1, \ldots, a_k, b_k)\right\|_1 \\
&= \sum_{j=1}^{k} \left\|\left(a_j, b_j\right)\right\|_1 \leq \sqrt{2} \sum_{j=1}^{k} \left\|\left(a_j, b_j\right)\right\|_2 \text{(Minkowski inequality))} \\
&\leq \sqrt{2} \sum_{j=1}^{k} \left|c_j\right| \\
&\leq \sqrt{2}\left\|(c_1, \ldots, c_k)\right\|_1 \\
&\leq \sqrt{2}\sqrt{k}\left\|(c_1, \ldots, c_k)\right\|_2 \text{(Minkowski inequality)} \\
&\leq \sqrt{2k}\left\|(s_1, \ldots, s_T)\right\|_2 \text{ as } T \text{ Fourier coefficients have the same } L_2 \text{ norm as } S.
\end{aligned} \tag{7}$$

Then,

$$\Delta_1\left(\mathrm{DFT}^k\right) \leq \sqrt{2k}\Delta_2(S) = M\sqrt{2Tk}.$$

□

This result is true when the DFT is normalized (2) as in our case. In [17, 19, 36], the L_2 norm of Fourier coefficients equals to $\sqrt{T}$ times the L_2 norm of S (Parvesal's theorem). This result is valid when the normalized DFT (2) is used as in our case. When the DFT is not normalized, as is the case in [17, 19, 36], the sensitivity of the first k DFT coefficients should be $\Delta_1(\mathrm{DFT}^k) = \sqrt{T} \times M\sqrt{2Tk} = MT\sqrt{2k}$ instead of $(M\sqrt{Tk})$. Thus, using the normalized DFT, the function then becomes

$$\begin{aligned}
&\widetilde{\mathrm{DFT}^k}: \mathbb{R}^T \longrightarrow \left(\mathbb{R}^2\right)^k \\
&S \longmapsto \widetilde{\mathrm{DFT}}(S) = ((a_1, b_1), \ldots, (a_k, b_k)) + \left(\left(y_{1,1}, y_{1,2}\right), \ldots, \left(y_{k,1}, y_{k,2}\right)\right),
\end{aligned} \tag{8}$$

which is ε–DP, with $y_{j,\ell} = \mathscr{L}(M\sqrt{2Tk}/\varepsilon)$, for all $j = 1, \ldots, k$ and $\ell = 1, 2$.

For simplicity, in the following, we write $c_j + \mathscr{L}(M\sqrt{2Tk}/\varepsilon)$ instead of $(a_j, b_j) + (\mathscr{L}(M\sqrt{2Tk}/\varepsilon), \mathscr{L}(M\sqrt{2Tk}/\varepsilon))$, meaning that two independent Laplace noises $\mathscr{L}(M\sqrt{2Tk}/\varepsilon)$ are added to the real and imaginary parts of c_j.

Algorithm 3 shows the Fourier perturbation algorithm (FPA) revisited.

4.3. Differences between the Initial, yet Incorrect, FPA, and the Corrected FPA. For a budget of privacy ε, the differences between the initial incorrect FPA and the corrected one can be highlighted as follows:

(1) The DFT used in the initial incorrect FPA [17] is not normalized, while it is normalized in the corrected FPA. Thus, a factor $\sqrt{2T}$ is missing in the Laplace noise in Algorithm 1.

(2) In the initial incorrect FPA [17], Laplace noises are only added to the real part of the DFT coefficients, while they should be added to the real and imaginary parts of the DFT coefficients as in the corrected FPA (Algorithm 3). Thus, k imaginary coefficients are not noised in Algorithm 1.

Figure 4 shows the same aggregated consumption presented in Example 1 and its noisy version using the corrected FPA (Algorithm 3) with $\varepsilon = 1$ per day and $k = 5$. Figure 4 shows that the corrected FPA obtains a large MRE (84%), making it useless. The noisy aggregate has negative consumptions and does not contain the peaks present in the original aggregate.

For the sake of simplicity, in the following sections, we use FPA to talk about the corrected version.

5. Clamping Transform Perturbation Algorithm

The intuition behind our approach, "Clamping transform perturbation algorithm," lies in the perturbation error, caused by the Laplace mechanism, which depends on the sensitivity of the sum of consumptions. As such, by reducing the sensitivity, we expect to reduce the perturbation error.

To estimate the sensitivity of consumptions, we split our database of N users into two almost equal parts: D_1 corresponding to the consumptions of the first half of users (a training dataset) and D_2 containing the second half of users' consumptions (a validation dataset). Using D_1, we compute the distribution of users' consumptions in the frequency domain. We denote by $M = (M_1, \ldots, M_k)$ the maximum magnitude (by ignoring outliers) of the k first coefficients.

For example, using the Irish consumption database [39], the distribution of the individual consumption of the first half customers (from 1 to 1818) in the frequency domain is given in Figure 5. In Figure 5, the maximum magnitudes (rounded) of the 5 first coefficients are $M = (M_1, M_2, M_3, M_4, M_5) = (9, 4, 3, 2, 2)$.

Inputs: $S = (S_1, \ldots, S_T)$, k, the maximum consumption M of the domain, and the privacy budget ε.
(1) Compute the normalized DFT coefficients of $S : F = \mathrm{DFT}(S)$.
(2) Keep only the first k coefficients of F, denoted by F^k.
(3) Generate the noisy version of F^k, denoted by $\widehat{F}^k$ by adding a Laplace noise $\mathscr{L}(M\sqrt{2Tk}/\varepsilon)$ to each coefficient in F^k.
(4) Pad $\widehat{F}^k$ to a T-dimensional vector, denoted by $\mathrm{PAD}^T(\widehat{F}^k)$ by appending $T - k$ zeroes.
(5) Apply the inverse DFT to $\mathrm{PAD}^T(\widehat{F}^k)$ to obtain a noisy version of S denoted by $\widehat{S}$.

Algorithm 3: Fourier perturbation algorithm (FPA) revisited.

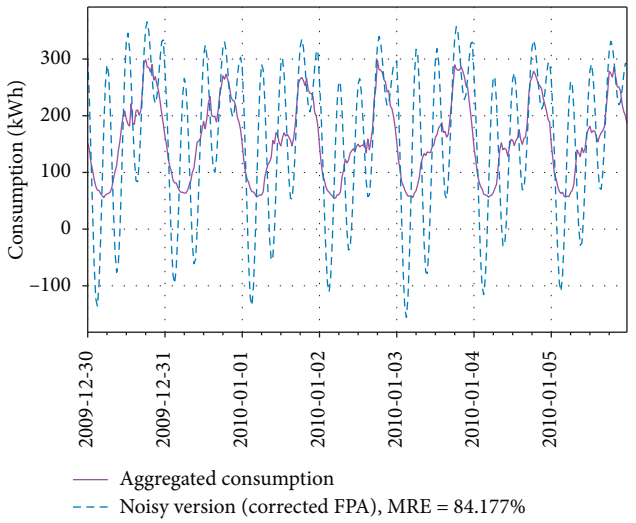

Figure 4: Aggregated consumption of 250 homes from 30th December 2009 to 5th January 2010 of dataset from CER [39] and its noisy version using the corrected FPA (Algorithm 3), with $\varepsilon = 1$ per day and $k = 5$.

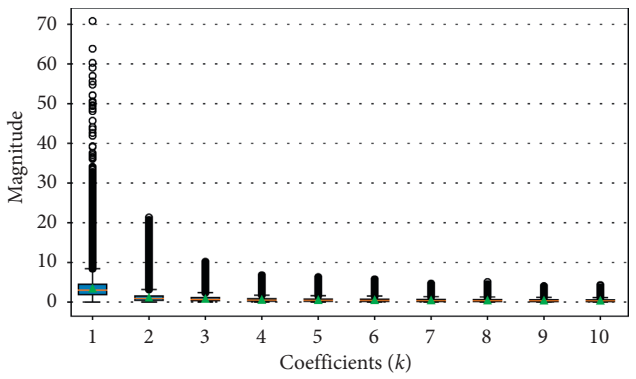

Figure 5: Distribution of users' consumptions in the frequency domain using the discrete Fourier transform (DFT).

The database D_2 is used for testing our methodology. Let $X = (X^1, \ldots, X^n)$ with $X^j = (x_1^i, \ldots, x_T^i)$ for all $j = 1, \ldots, n$ be the users' individual consumptions. To publish the sum of consumptions $S = (S_1, \ldots, S_T) = (\sum_{j=1}^n x_1^j, \ldots, \sum_{j=1}^n x_T^j)$, our methodology, which can be applied to either the Fourier transform or to wavelet transforms, is described as follows:

(1) For all individual consumptions X^j $(j = 1, \ldots, n)$, compute the corresponding magnitude in the domain of the transform and keep the first k coefficients denoted by $C^j = (C_1^j, \ldots, C_k^j)$.
(2) If the modulus of coefficient C_ℓ^j is greater than M_ℓ $(1 \le \ell \le k)$, replace C_ℓ^j with $C_\ell^j \cdot M_\ell / |C_\ell|$ so that all coefficients have a modulus smaller than M_ℓ and their phase, if the coefficient is complex, is unchanged.
(3) Compute the sum of coefficients $C = (\sum_{j=1}^n C_1^j, \ldots, \sum_{j=1}^n C_k^j)$.
(4) Add a noise following the distribution of Laplace $\mathscr{L}(\cdot)$, depending on the sensitivity of the transform, to each coefficient C_ℓ $(1 \le \ell \le k)$ of C. The result is denoted by $\widehat{C}$. We note that the Laplace noise is added to the real and imaginary parts of each coefficient when the DFT is used.
(5) Pad the vector $\widehat{C}$ by $n - k$ zeroes and compute the inverse transform to obtain the noisy version of the consumption $\widehat{S}$.

Section 5.1 presents an adaptation of this methodology using the discrete Fourier transform.

5.1. Clamping Fourier Perturbation Algorithm. This section describes the clamping Fourier perturbation algorithm (CFPA) detailed in Algorithm 4. This algorithm allows an aggregator to compute and publish an aggregate guaranteeing ε–differential privacy.

CFPA takes as inputs the individual time-series consumptions of n consumers, the maximum magnitudes of DFT coefficients of individual consumptions M (computed over database D_1), the number k of DFT coefficients to be considered, and the privacy budget ε, and it returns the noisy time-series sum of consumptions of n consumers.

Step 1, called clamping, computes the first k DFT coefficients of each individual time-series consumption. If the magnitude of a coefficient F_ℓ^j is greater than the maximum magnitude M_ℓ, then this coefficient is clamped and replaced by $F_\ell^j \cdot M_\ell / |F_\ell^j|$, in which magnitude is $|F_\ell^j| \cdot M_\ell / |F_\ell^j| = M_\ell$. Thus, for all individual consumptions X^j, the maximum magnitude of the k first DFT coefficients $F^j = (F_1^j, \ldots, F_k^j)$ is $M = (M_1, \ldots, M_k)$, i.e., the final values of the coefficients have the same phase as the initial values, but their magnitudes are bounded by $(M_1, \ldots, M_k)$.

After computing the first k DFT coefficients $F^j = (F_1^j, \ldots, F_k^j)$ of each individual time-series consumption of consumers $(j = 1, \ldots, n)$, Step 2 consists in computing the sum $(F_1, \ldots, F_k) = (\sum_{j=1}^n F_1^j, \ldots, \sum_{j=1}^n F_k^j)$ of

Inputs:
(i) Consumptions: $X = (X^1, \ldots, X^n)$ with $X^i = (x_1^i, \ldots, x_T^i)$ for all $i = 1, \ldots, n$
(ii) k
(iii) The maximum magnitudes of k first DFT coefficients: $M = (M_1, \ldots, M_k) \in \mathbb{R}_+^k$
(iv) Privacy budget: ε
(1) *Clamping*: for each individual time-series consumption X^j,
(i) compute the k first DFT coefficients of X^j: $F^j = (\mathrm{DFT}(X^j)_1, \ldots, \mathrm{DFT}(X^j)_k) = (F_1^j, \ldots, F_k^j)$
(ii) if $|F_\ell^j| > M_\ell$, then replace F_ℓ^j with $F_\ell^j \cdot M_\ell / |F_\ell^j|$ for all $\ell = 1, \ldots, k$
(2) *Laplacian mechanism*: compute the sum of noisy consumptions of each DWT coefficient: $\widehat{F}_\ell = \sum_{j=1}^n F_\ell^j + \mathscr{L}(M_\ell \sqrt{2}/\varepsilon/k)$ for all $\ell = 1, \ldots, k$. We denote $\widehat{F}^k = (\widehat{F}_1, \ldots, \widehat{F}_k)$. We note that the noise is added to the real and imaginary parts of the sum of coefficients.
(3) Pad $\widehat{F}^k$ with $T - k$ zeros; the result is denoted by $\mathrm{PAD}^T(\widehat{F}^k)$
(4) Compute the inverse DFT of $\mathrm{PAD}^T(\widehat{F}^k)$ to get the noisy sum of consumptions denoted by $\widehat{S} = (\widehat{S}_1, \ldots, \widehat{S}_T)$ of the initial sum $S = (\sum_{j=1}^n x_1^j, \ldots, \sum_{j=1}^n x_T^j)$.

Algorithm 4: Clamping Fourier perturbation algorithm (CFPA).

these coefficients using the Laplacian mechanism. The result is denoted by $\widehat{F}^k = (\widehat{F}_1, \ldots, \widehat{F}_k)$.

Finally, the noisy sum of consumptions is equal to the inverse of the noisy DFT coefficients padded with $T - k$ zeros.

Theorem 3. *Algorithm CFPA is ε-differentially private.*

Proof. To prove that Algorithm 4 is ε-differentially private, we need to prove that the sensitivity of the sum of DFT coefficients of users' individual consumptions F_1 (resp. $F_2, \ldots, F_k$) is $\sqrt{2} \cdot M_1$ (resp. $\sqrt{2} \cdot M_2, \ldots, \sqrt{2} \cdot M_k$). This is done in Lemma 2.

Then, as a Laplacian noise $\mathscr{L}(M_\ell \sqrt{2}/\varepsilon/k)$ is added to each component F_ℓ ($\ell = 1, \ldots, k$), the resulting $\widehat{F}_1$ (resp. $\widehat{F}_2, \ldots, \widehat{F}_k$) is ε/k-differentially private. Finally, the composition theorem [14] guarantees that any computation on the k components of $(\widehat{F}_1, \ldots, \widehat{F}_k)$ is ε-differentially private; thus, the inverse DFT of those coefficients is ε-DP. □

Lemma 2. *Let $F^j = (DFT(X^j)_1, \ldots, DFT(X^j)_k) = (F_1^j, \ldots, F_k^j)$ be the first k DFT coefficients of the individual consumption of consumer j ($j = 1, \ldots, n$), obtained after the clamping mechanism. The sensitivity of the sum of each DFT coefficient F_ℓ^j ($\ell = 1, \ldots, k$) of n consumers' individual consumptions is $M_\ell \cdot \sqrt{2}$.*

Proof. Let $F^j = (\mathrm{DFT}(X^j)_1, \ldots, \mathrm{DFT}(X^j)_k) = (F_1^j, \ldots, F_k^j)$. After the clamping, the magnitude of each DFT coefficient F_ℓ^j is smaller than M_ℓ for $\ell = 1, \ldots, k$, and the sensitivity of the function $f_\ell: \mathscr{D}^n \longrightarrow \mathbb{C} \equiv \mathbb{R}^2$ defined by $f_\ell: (X^1, \ldots, X^n) \longmapsto \sum_{j=1}^n F_\ell^j \equiv (\sum_{j=1}^n a_\ell^j, \sum_{j=1}^n b_\ell^j)$, with $F_\ell^j = a_\ell^j + i b_\ell^j$ being equal to

$$\begin{aligned}
\Delta_1(f_\ell) &= \max_{|a_\ell^j|, |b_\ell^j|} \left\| \left(\sum_{j=1}^n a_\ell^j, \sum_{j=1}^n b_\ell^j \right) - \left(\sum_{j=2}^n a_\ell^j, \sum_{j=2}^n b_\ell^j \right) \right\|_1 \\
&= \max \left\| \left(a_\ell^1, b_\ell^1 \right) \right\|_1 \\
&\leq \max \sqrt{2} \left\| \left(a_\ell^1, b_\ell^1 \right) \right\|_2 \quad \text{(Minkowski inequality)} \\
&= \max \sqrt{2} \sqrt{\left(a_\ell^1 \right)^2 + \left(b_\ell^1 \right)^2} \\
&= \max \sqrt{2} \left| F_\ell^1 \right| \\
&= M_\ell \sqrt{2}.
\end{aligned} \tag{9}$$

□

Thus, Lemma 2 proves Theorem 3, and algorithm CFPA guarantees ε-differential privacy.

For example, Figure 6 shows the same aggregated consumption presented in Example 1 and its noisy version using CFPA (Algorithm 4) with $\varepsilon = 1$ per day and $k = 5$. Figure 6 shows that CFPA obtains a good utility with an MRE equal to 9.7%. This good utility of CFPA can be explained by the fact that Laplace noise added in CFPA depends on the amplitude of each coefficient, while in FPA, the same noise $\mathscr{L}(M \cdot \sqrt{2Tk}/\varepsilon)$ is added to every DFT coefficients, where M is the maximum consumption in the dataset.

5.2. Clamping Wavelet Perturbation Algorithm. The clamping wavelet perturbation algorithm (CWPA), as presented in Algorithm 5, is obtained by replacing DFT with DWT in Algorithm 4. The computation of DWT is based on multiresolution analysis which determines the number of

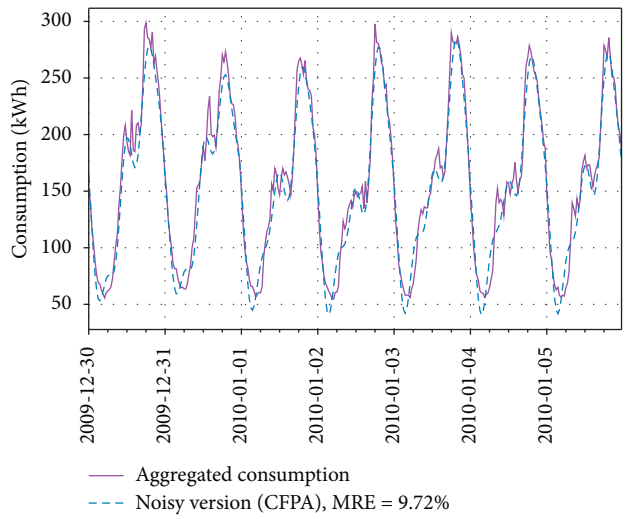

FIGURE 6: Aggregated consumption of 250 homes from 30th December 2009 to 5th January 2010 of dataset from CER [39] and its noisy version using CFPA (Algorithm 4), with $\varepsilon = 1$ per day and $k = 5$.

approximation coefficients (scaling functions) and detail coefficients (wavelet functions) [40]. DWT takes as input a time-series of length a power of 2. If the input's length is not a power of 2, we can pad it with zeroes [41].

Algorithm 5 takes as inputs the maximum magnitudes of the first k DWT coefficients which are obtained in the training process on D_1, by computing the distribution of DWT coefficients of individual consumptions.

We note that there are multiple DWTs, such as Haar, Daubechies, Symlets, and Coiflets. In this paper, we use Haar and Daubechies wavelets as shown in Section 6, because they give a low reconstruction error, as will be discussed in Section 6.1.

Theorem 4. *The clamping wavelet perturbation algorithm (CWPA), Algorithm 5, is ε-differentially private.*

Proof. The proof is similar to the one for Theorem 3. We need to prove that the sensitivity of the sum of DWT coefficients of users' individual consumptions W_1 (resp. $W_2, \ldots, W_k$) is M_1 (resp. $M_2, \ldots, M_k$). This is done in Lemma 3.

Then, as a Laplacian noise $\mathscr{L}(M_\ell/\varepsilon/k)$ is added to each component W_ℓ ($\ell = 1, \ldots, k$), the resulting $\widehat{W}_1$ (resp. $\widehat{W}_2, \ldots, \widehat{W}_k$) is ϵ/k-differentially private. Finally, the composition theorem [14] guarantees that any computation on the k components $(\widehat{W}_1, \ldots, \widehat{W}_k)$ is ϵ-differentially private; thus, the inverse DWT of those coefficients is ϵ-DP. □

Lemma 3. *Let $W^j = (DWT(X^j)_1, \ldots, DWT(X^j)_k) = (W_1^j, \ldots, W_k^j)$ be the first k DWT coefficients of the individual consumption of consumer j ($j = 1, \ldots, n$), obtained after the clamping mechanism. The sensitivity of the sum of each DWT coefficient W_ℓ^j ($\ell = 1, \ldots, k$) of n consumers' individual consumptions is M_ℓ.*

Proof. Let $W^j = (\mathrm{DWT}(X^j)_1, \ldots, \mathrm{DWT}(X^j)_k) = (W_1^j, \ldots, W_k^j)$. After the clamping, the magnitude of each DWT coefficient W_ℓ^j is smaller than M_ℓ for $\ell = 1, \ldots, k$, and the sensitivity of the function $w_\ell: \mathscr{D}^n \longrightarrow \mathbb{R}$ defined by $w_\ell: (X^1, \ldots, X^n) \longmapsto \sum_{j=1}^n W_\ell^j$ is equal to

$$\begin{aligned}\Delta_1(w_\ell) &= \max_{|W_\ell^j|} \left| \sum_{j=1}^{n} W_\ell^j - \sum_{j=2}^{n} W_\ell^j \right| \\ &= \max |W_\ell^1| \\ &= M_\ell.\end{aligned} \tag{10}$$

□

For example, Figure 7 shows the same aggregated consumption presented in Example 1 and its noisy version using CWPA (Algorithm 5) with Haar wavelet, $\varepsilon = 1$ per day and $k = 5$. However, Figure 3 shows that the MRE of CWPA is still higher than 10%. We explain this result in Section 6.

6. Experimental Results

This section compares FPA, CFPA, WPA, and CWPA and explains through experimentations why CFPA achieves a better utility than other publication techniques. After presenting the raw results, we explain them by decomposing the mean relative error into a perturbation error, caused by the clamping mechanism and the Laplace mechanism, and a reconstruction error, due to ignoring $T - k$ coefficients of the transform. The analysis of the error is thus conducted in the next two Subsections 6.1 and 6.2. Section 6.1 analyzes the reconstruction error, while Section 6.2 analyzes the perturbation one.

Conditions: the experiments rely on data originating from the Irish Commission for Energy Regulation (CER) [39]. This dataset contains real time-series consumptions. The achieved results are valid for this very specific case, for Irish consumptions with an Irish weather being never too hot or too cold. The results show that the approach is good, but will probably have to be adapted for other datasets, i.e., by computing the maximum magnitudes of the k first coefficients of the

Inputs:
(i) Consumptions: $X = (X^1, \ldots, X^n)$ with $X^j = (x_1^j, \ldots, x_T^j)$ for all $j = 1, \ldots, n$
(ii) k
(iii) The maximum magnitudes of k first DWT coefficients: $M = (M_1, \ldots, M_k) \in \mathbb{R}_+^k$
(iv) Privacy budget: ε
(1) *Clamping*: for each individual time-series consumption X^j,
(i) compute the first k DWT coefficients of X^j: $W^j = (\mathrm{DWT}(X^j)_1, \ldots, \mathrm{DWT}(X^j)_k) = (W_1^j, \ldots, W_k^j)$
(ii) if $|W_\ell^j| > M_\ell$, then replace W_ℓ^j with $W_\ell^j \cdot M_\ell / |W_j^i|$ for $\ell = 1, \ldots, k$
(2) Laplacian Mechanism: compute the sum of noisy consumptions of each DWT coefficient: $\widehat{W}_\ell = \sum_{j=1}^n W_\ell^j + \mathscr{L}(M_\ell/\varepsilon/k)$ for all $\ell = 1, \ldots, k$. We denote $\widehat{W}^k = (\widehat{W}_1, \ldots, \widehat{W}_k)$
(3) Pad $\widehat{W}^k$ with $T - k$ zeroes; the result is denoted by $\mathrm{PAD}^T(\widehat{W}^k)$
(4) Compute the inverse DWT of $\mathrm{PAD}^T(\widehat{W}^k)$ to get the noisy sum of consumptions denoted by $\widehat{S} = (\widehat{S}_1, \ldots, \widehat{S}_T)$ of the initial sum $S = (\sum_{j=1}^n x_1^j, \ldots, \sum_{j=1}^n x_T^j)$

Algorithm 5: Clamping wavelet perturbation algorithm (CWPA).

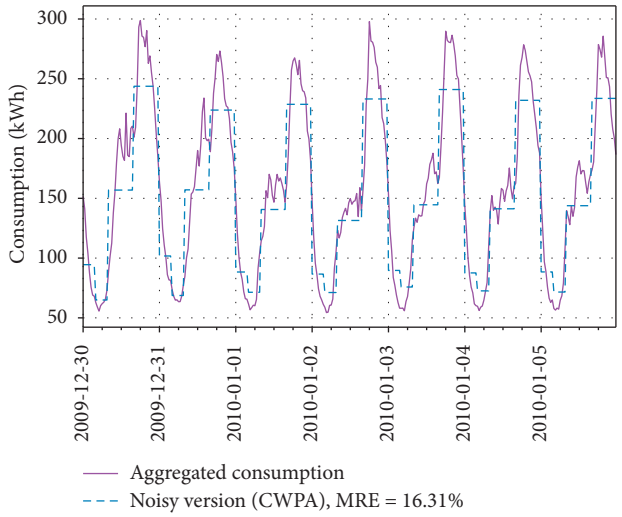

Figure 7: Aggregated consumption of 250 homes from 30th December 2009 to 5th January 2010 of dataset from CER [39] and its noisy version using CWPA (Algorithm 5), with Haar wavelet, $\varepsilon = 1$ per day and $k = 5$.

considered transform over a subpart of the dataset. Consumption data from the CER were collected every 30 minutes from 2009 to 2010 with the participation of more than 5,000 Irish homes and businesses. This experiment only considers homes. We divided the database in two parts: D_1, corresponding to the first half of consumers (1 to 1818), and D_2, corresponding to the second half (1819 to 3639). D_1 is used to calibrate the algorithms by computing the maximum magnitudes $M = (M_1, \ldots, M_k)$ of the first k coefficients in the frequency domain, and D_2 is used to test the publication techniques FPA, CFPA, WPA, and CWPA.

Notations: we note N as the number of homes or smart meters considered in the district to compute the time-series of the sum of consumptions. For each day (48 time slots), we compute the sum of consumptions of 50 different districts of N random homes, and we execute FPA, WPA, CFPA, and CWPA with privacy budget $\varepsilon \in \{1; 3\}$ for each day and $k \in \{5; 8; 12\}$. The discrete wavelet transforms used here are Haar transform (which represents the same wavelet as Daubechies with order 1, noted db1), Daubechies with order 2, and Daubechies with order 3, respectively, noted db2 and db3.

Raw results and analysis: Figures 8 and 9 show the distribution of the mean relative estimation error (MRE) according to the number of homes in the district (N from 50 to 450) and k from 5 to 12 for the budget of privacy $\varepsilon = 1$ and $\varepsilon = 3$, respectively. The boxes extend from the lower to upper quartile values of the MRE, with a line at the median and a triangle representing the mean. The whiskers extend from the box to show the range of the MRE. In order to make consumption forecasts, an MRE lower than 10% is required in practice by experts in the energy sector. In this section, an MRE of less than 10% is therefore considered useful.

In Figures 8 and 9, the first column corresponds to the comparison between the FPA and the CFPA. The other columns correspond to the comparison between the WPA and the CWPA using Haar wavelet with 2 approximation coefficients, Daubechies 2 (db2) with 5 approximation coefficients, and Daubechies 3 (db3) with 10 approximation coefficients, respectively.

Figures 8 and 9 show that CFPA has a better utility than FPA. For example, for $\varepsilon = 1$ (Figure 8), when $k = 5$ and the number of homes $N = 350$, the MRE of CFPA is 12%, while the MRE of FPA is 75%. In that configuration, the MRE of CFPA is 6.25 times lower than that of FPA. Similarly, the CWPA obtains a better utility than the WPA. For example, for $k = 5$ and the number of homes $N = 350$, the MRE of CWPA using Haar wavelet is 15%, while the MRE of the WPA is 30%. In that configuration, the MRE of CWPA is 2 times lower than that of WPA.

Generally, the larger the size of the district N, the smaller the MRE is. Similarly, the larger the budget of privacy ε, the smaller the MRE is; Figure 9 ($\varepsilon = 3$) shows a better utility than Figure 8 ($\varepsilon = 1$). However, Figures 8 and 9 show that WPA and CWPA using db3 are not useful for $k = 5$ because, as shown in Section 6.1, the reconstruction error is high (between 70% and 80%).

Figures 8 and 9 show that for larger k, the MRE of WPA and CWPA using db3 is smaller. Moreover, in Figure 9, for

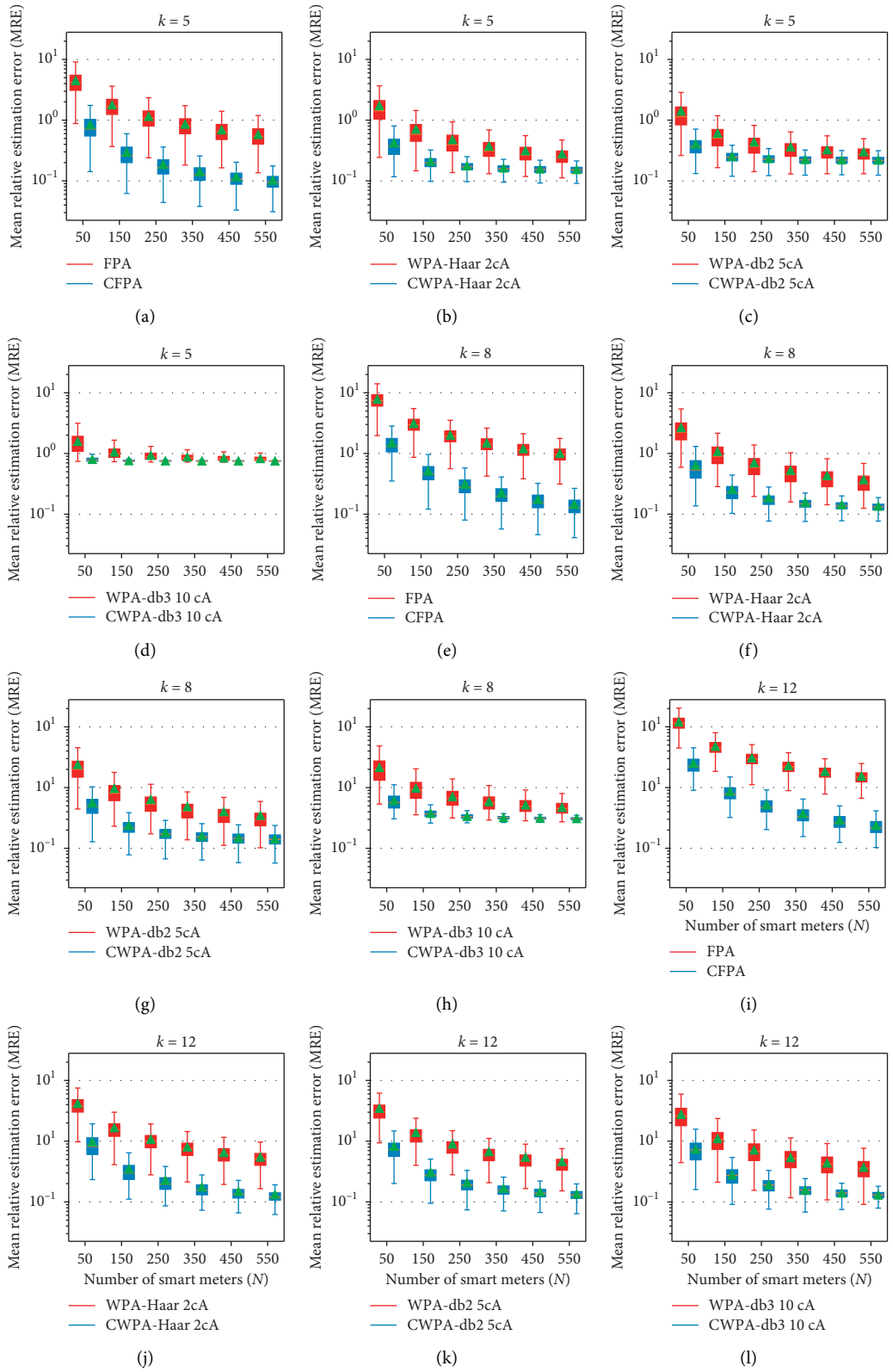

FIGURE 8: Mean relative estimation error (MRE) of FPA vs CFPA vs WPA vs CWPA, using DFT and DWT with Haar, Daubechies 2 (db2), and Daubechies 3 (db3) wavelets, according to k, and the number of smart meters (N) in the district, with $\varepsilon = 1$.

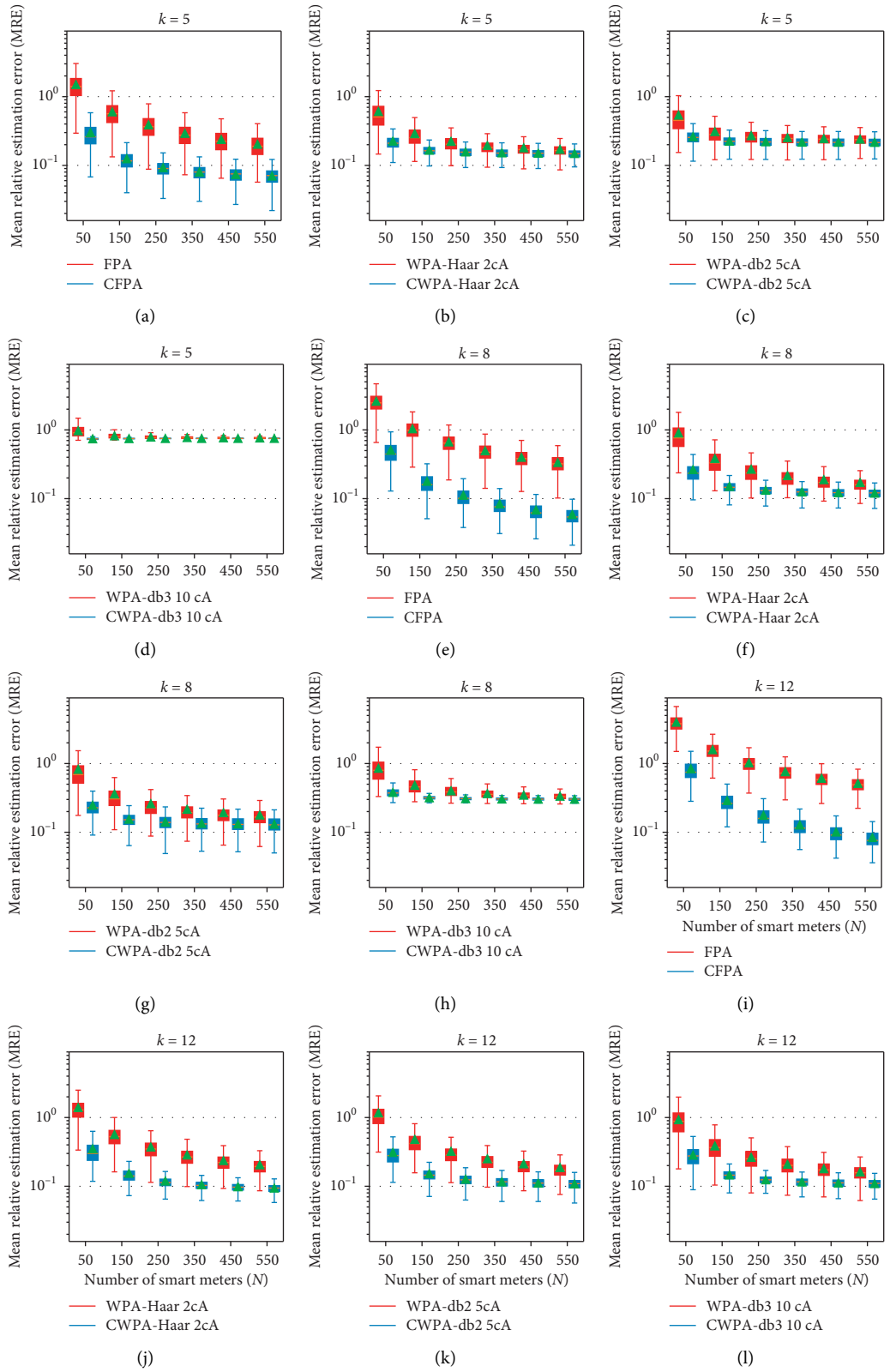

FIGURE 9: Mean relative estimation error (MRE) of FPA vs CWPA vs WPA vs CWPA with Haar, Daubechies 2 (db2), and Daubechies 3 (db3) wavelets, according to k, and the number of smart meters (N) in the district, with $\varepsilon = 3$.

$k = 12$ and when the number of homes is higher than 250 and $\varepsilon = 3$, CWPA using db3 has the median of MRE smaller than 11%. The CWPA using Haar wavelet obtains the second best utility, with the median of MRE smaller than 10% when N is greater than 250, while the CFPA gets the best utility, with the median of MRE decreasing to 5% when $N = 550$ and $k = 8$.

However, the utility of FPA and WPA decreases when k increases. This is caused by the perturbation error; indeed, the greater the k, the greater the Laplacian noise added to each coefficient is. This noise is attenuated by the clamping as shown by the CFPA. Indeed, when k goes from 5 to 8, the reconstruction error decreases and the clamping also decreases the perturbation error leading to the total error reduction. However, when k goes from 8 to 12, although the reconstruction error decreases, clamping does not reduce the perturbation error sufficiently. This explains why the MRE of CFPA is a little bigger when $k = 12$ compared to $k = 8$.

In Figures 8 and 9, we notice that the median of MRE of WPA and CWPA converge to a threshold and never goes below it. For example, for $\varepsilon = 3$ and $k = 5$, the median of MRE of WPA and CWPA using db2 converges to 23%. This is caused by the reconstruction error.

6.1. Reconstruction Error. The reconstruction error is due to considering only the k first transform coefficients, thus removing the precision brought by coefficients $(k + 1, k + 2, \ldots)$. To measure this error, a first solution consists in computing the cumulative distribution function (CDF) of the coefficients as a first assessment of the impact of the transform coefficients and, then, to get confirmation through some experimental reconstruction error measurements. Intuitively, if the CDF of some coefficients k is close to 1, it means that the coefficients after k $(k + 1, k + 2, \ldots)$ have less impact on the reconstruction, and thus, when set to zero, lead to a smaller reconstruction error.

The CDF is computed for a district of 50 homes of several transformations: discrete Fourier transform (DFT), discrete wavelet transform (DWT) using Haar, and Daubechies 2 and Daubechies 3 wavelets. The closer to 1 the cumulative distribution function at k is, the smaller the reconstruction error is. Figure 10 compares the cumulative distribution function of DFT and DWT with different wavelet transforms. This figure shows that DFT has a higher cumulative distribution than DWT for the considered range value of k $(k \leq 10)$.

In order to analyze this error more precisely, we define formally the reconstruction error below, and we then compute it experimentally.

Definition 4 (reconstruction error). Let $S = (S_1, S_2, \ldots, S_T)$ be a sum of time-series consumptions and $C = (C_1, C_2, \ldots, C_T)$ be the coefficients in the frequency domain of this time-series. We denote $\mathrm{PAD}^T(C^k) = (C_1, \ldots, C_k, 0, \ldots, 0)$ as the first k coefficients padded with zeros and $\widetilde{S} = (\widetilde{S}_1, \widetilde{S}_2, \ldots, \widetilde{S}_T)$ as the inverse of $\mathrm{PAD}^T(C^k)$ (in the time domain). The reconstruction error

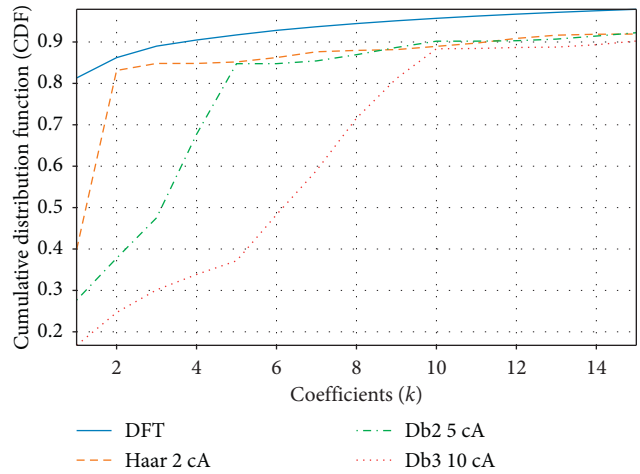

Figure 10: Comparison of cumulative distribution function of DFT and DWT with Haar, and Daubechies 2 and 3 wavelets for a district of 50 homes.

(RE) of $\mathrm{PAD}^T(C^k)$ is equal to the mean relative estimation error between S and $\widetilde{S}$ given by (we add 1 to the denominator in order to avoid the division by zero)

$$\mathrm{RE}\left(\mathrm{PAD}^T\left(C^k\right)\right) = \frac{1}{T} \cdot \sum_{j=1}^{T} \frac{\left|S_j - \widetilde{S}_j\right|}{S_j + 1}. \tag{11}$$

Figure 11 shows the reconstruction error for DFT and DWT with different wavelet transforms for a district of 50 and 450 homes. This figure shows that the DFT obtains the smallest relative error (lower than 10% when k is greater than 5) followed by Haar and Daubechies. We note that the reconstruction error of Daubechies 2 is higher than 23% when $k = 5$, which leads to a total error higher than 23% and justifies the relative error obtained in Figures 8 and 9.

Moreover, when $k = 5$, the reconstruction error of Daubechies 3 is higher than 70%, which justifies why its total error is higher than 70% when $k = 5$, according to Figures 8 and 9.

According to the database from the Irish Commission for Energy Regulation (CER) [39], the discrete Fourier transform gets the smaller reconstruction error, followed respectively by Haar (which is the same as Daubechies 1) and Daubechies 2 and Daubechies 3 wavelets.

6.2. Perturbation Error. The perturbation error is caused by the Laplace mechanism, applied on the first k transform coefficients. The higher the transform coefficients, the lower the impact of this perturbation in terms of relative error, and thus the lower the perturbation error.

We note that the amplitude of the Laplace noise introduced by the Laplace mechanism is different for CFPA and CWPA; it is $\sqrt{2}$ times greater for CFPA than for CWPA. Indeed, for all $\ell = 1, \ldots, k$, the parameter for the Laplace noise is $\mathscr{L}(M_\ell \sqrt{2}/\varepsilon/k)$ for CFPA and $\mathscr{L}(M_\ell/\varepsilon/k)$ for CWPA. Moreover, in the CFPA, $2k$ coefficients (the real and imaginary parts of the k DFT coefficients) are noisy while only k coefficients are noisy in the CWPA.

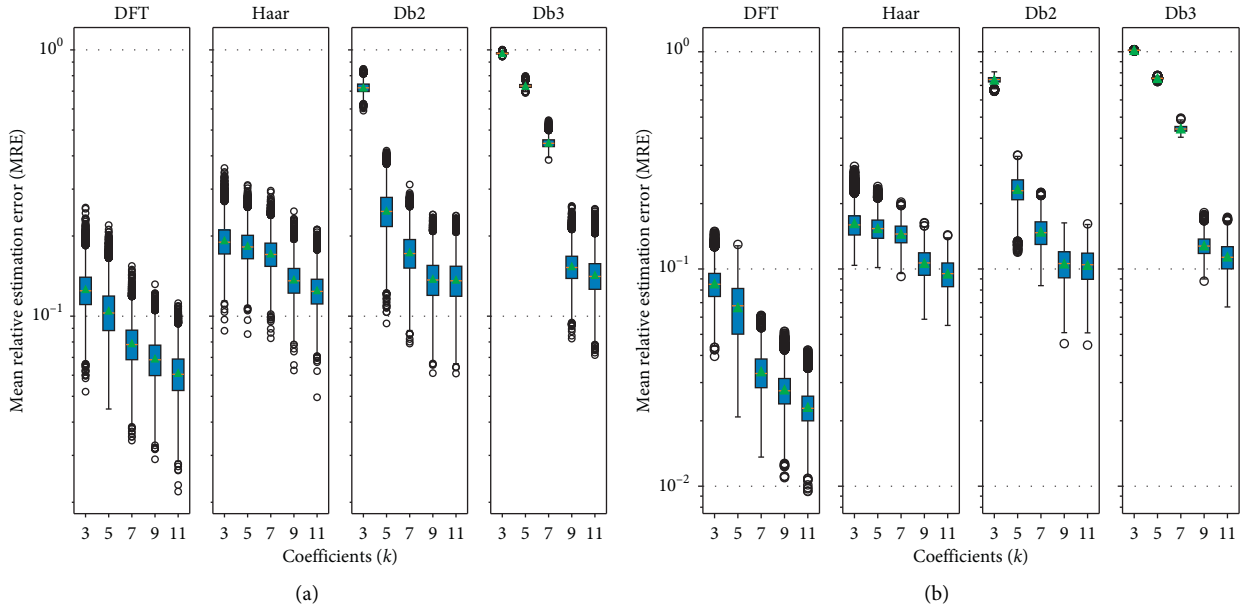

FIGURE 11: Comparison of reconstruction error of DFT and DWT with Haar amd Daubechies 2 and Daubechies 3 wavelets for a district of 50 and 450 homes. (a) $N = 50$. (b) $N = 450$.

For a district of 50 homes, we compute the distribution of the magnitude of DFT and DWT with Haar amd Daubechies 2 and 3 wavelets, and we compare their coefficient distribution median in Figure 12.

Figure 12 shows that the coefficient values vary according to the values of k and the considered transforms. For instance, when k is in the interval $[7, 10]$, Daubechies 3 obtains the highest magnitudes of coefficients, followed by Daubechies 2 and DFT.

In clamping perturbation algorithms (CFPA, CWPA), the clamping mechanism allows to add a noise proportional to the modulus of the coefficients of the considered transform (DFT, DWT). This reduces the impact of noise compared to perturbation algorithms (FPA, WPA); however, at the price of a perturbation error induced by the clamping of the coefficients. Formally, the perturbation error of clamping perturbation algorithms (CFPA, CWPA) is defined as follows:

Definition 5. Perturbation error for clamping perturbation algorithms (CFPA, CWPA).

Let $X^1, \ldots, X^N$ be the individual time-series of energy consumptions of N homes, with $X^i = (x_1^i, \ldots, x_T^i)$ for $i = 1, \ldots, N$. The sum of time-series consumptions is noted as $S = (S_1, \ldots, S_T) = (\sum_{i=1}^N x_1^i, \ldots, \sum_{i=1}^N x_T^i)$. For all $i = 1, \ldots N$, we note $\overline{C}^i = (\overline{c}_1^i, \ldots, \overline{c}_k^i, c_{k+1}^i, \ldots, c_T^i)$ as the result of the considered transform of the time-series consumption X^i whose first k coefficients $(\overline{c}_1^i, \ldots, \overline{c}_k^i)$ have been clamped. We note $M = (M_1, \ldots, M_k)$ as the maximum magnitude of the first k coefficients of the considered transform. Let $\overline{C} = (\sum_{i=1}^N \overline{c}_1^i + \mathscr{L}(\delta_1/\varepsilon/k), \ldots, \sum_{i=1}^N \overline{c}_k^i + \mathscr{L}(\delta_k/\varepsilon/k), \sum_{i=1}^N c_{k+1}^i, \ldots, \sum_{i=1}^N c_T^i)$ be the sum of coefficients of the considered transform by perturbing only the first k coefficients, with $\delta_j = M_j\sqrt{2}$ (respectively $\delta_j = M_j$) for CFPA (respectively for CWPA), for $j = 1, \ldots, k$.

Let $\overline{S} = (\overline{S_1}, \ldots, \overline{S_T})$ be the inverse transform of $\overline{C}$. The perturbation error of $\overline{C}$ equals to the mean relative estimation error (MRE) between S and $\overline{S}$, given by(we add 1 to the denominator in order to avoid the division by zero). For CWPA, the Laplace noise $\mathscr{L}(M_j\sqrt{2}/\varepsilon/k)$ must be replaced by $\mathscr{L}(M_j/\varepsilon/k)$ for $j = 1, \ldots, k$ and the DFT by the DWT,

$$\mathrm{PE}(\overline{C}) = \frac{1}{T}\sum_{\ell=1}^{T}\frac{|S_\ell - \overline{S}_\ell|}{S_\ell + 1}. \tag{12}$$

The perturbation error depends on the following parameters, k, M_j, ε, and N for $j = 1, \ldots, k$. k, M_j $(j = 1, \ldots, k)$ and ε are parameters of the Laplace distribution, so they have a direct impact on the amplitude of the added noise. Let ε and M_j be fixed; the bigger the k, the smaller the Laplace distribution parameter $\delta_k/\varepsilon/k$ is, and thus, the bigger the noise added on the k first coefficients is. This makes the perturbation error more significant. The choice of M_j is important to define the clamping threshold and it directly impacts the perturbation of the Laplace mechanism. The greater the M_j, the bigger the Laplace noise is, and thus, the more the perturbation error is. The smaller the M_j (close to zero), the less the Laplace noise is, but the more the coefficients are clamped, and thus, the more the perturbation error is. The number of homes N indirectly plays a role in the perturbation error; the larger the N, the more diluted the added noise is. This leads to decrease the perturbation error.

Figure 13 (respectively, Figure 14) shows the distribution of the perturbation error of the clamping perturbation algorithms (CFPA and CWPA) according to k, N, with $\varepsilon = 1$ (respectively, $\varepsilon = 3$).

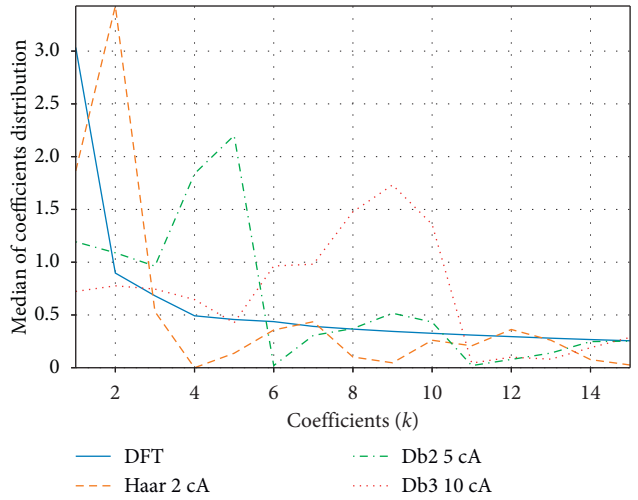

FIGURE 12: Comparison of the median of coefficients distribution of DFT and DWT with Haar and Daubechies 2 and Daubechies 3 wavelets for a district of 50 homes.

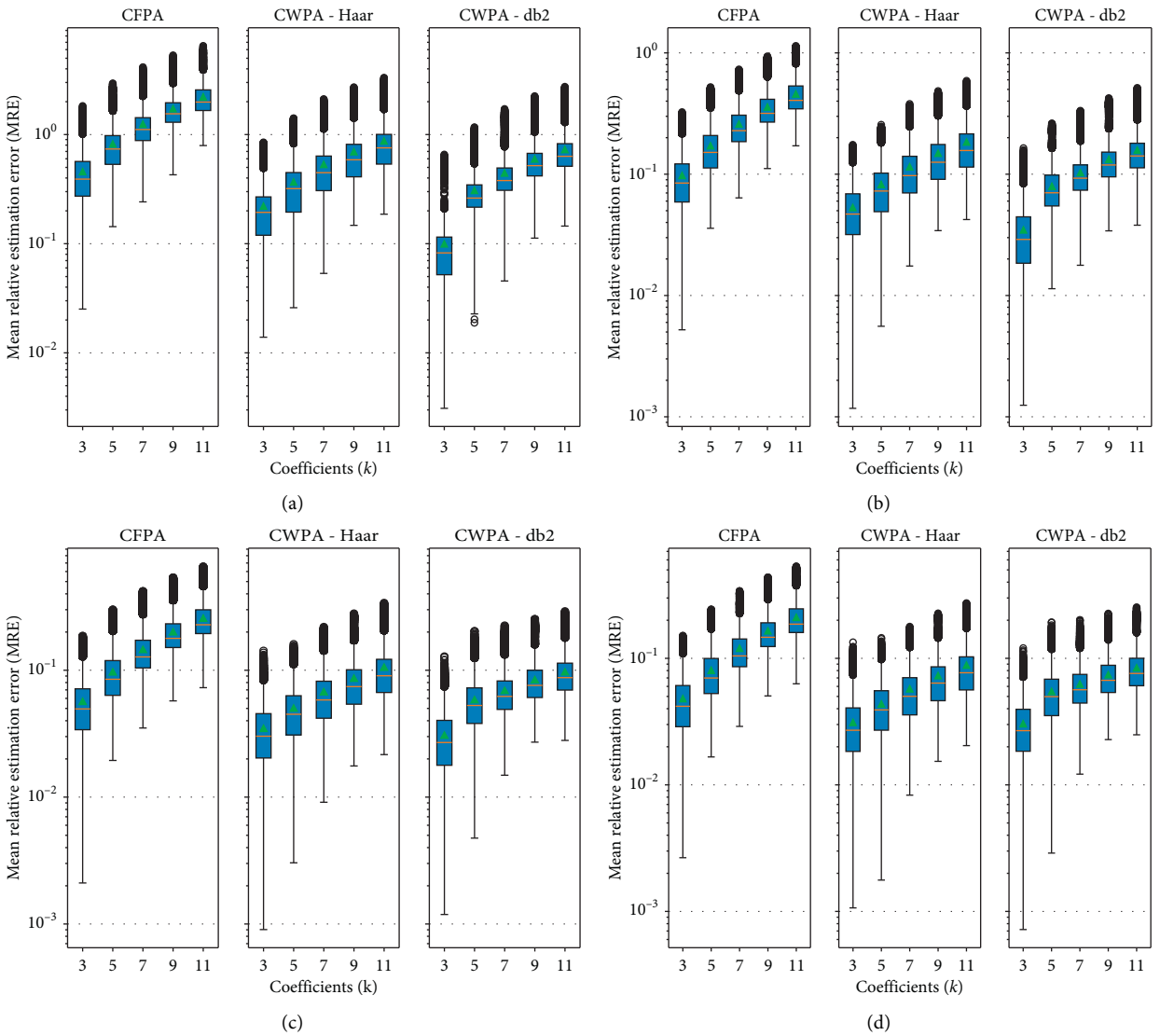

FIGURE 13: Distribution of perturbation error according to clamping perturbation algorithms CFPA and CWPA, with Haar, Daubechies 2, and Daubechies 3, and according to k and the number of homes N, for a fixed privacy budget, $\varepsilon = 1$. Note that the scales in (a)–(d) are different. (a) $N = 50$. (b) $N = 250$. (c) $N = 450$. (d). $N = 550$.

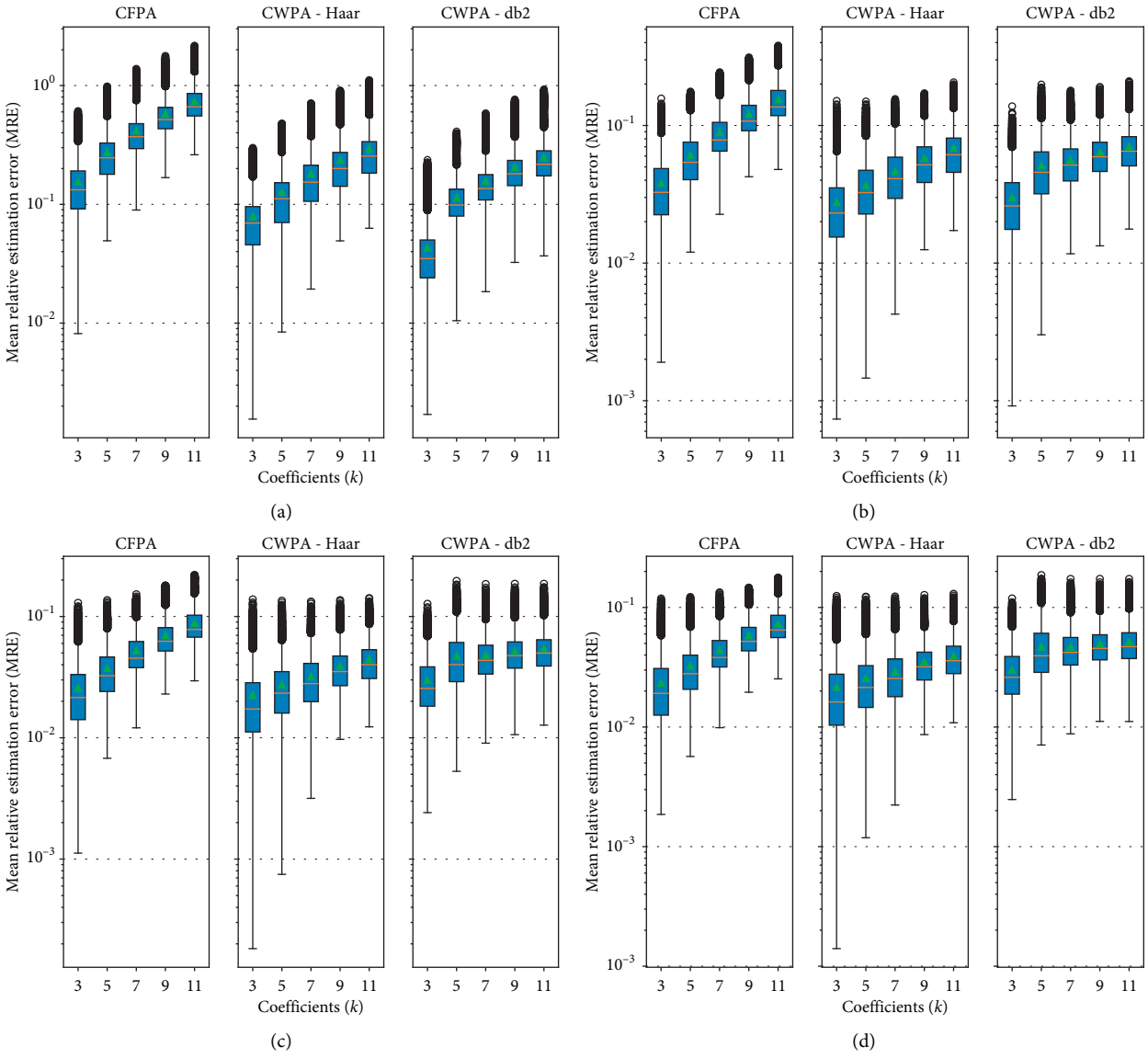

FIGURE 14: Distribution of perturbation error according to clamping perturbation algorithms CFPA and CWPA, with Haar, Daubechies 2, and Daubechies 3, and according to k and the number of homes N, for a fixed privacy budget, $\varepsilon = 3$. Note that the scales in 14(a)–14(d) are different. (a) $N = 50$ (b) $N = 250$. (c) $N = 450$. (d). $N = 550$.

Figures 13 and 14 show that the perturbation error of CFPA is higher than that of CWPA. This result is explained by

(1) The number of coefficients to be noised in CFPA is twice as many as the number of coefficients to be noised in CWPA. Indeed, in CFPA, the DFT coefficients are complex numbers, so both real and imaginary parts must be noised.

(2) The absolute value of the noise added in the CFPA is $\sqrt{2}$ times greater than that in the CWPA.

In addition, the greater the N, the more the added noise is diluted in the aggregate, causing the perturbation error to decrease. E.g., for $\varepsilon = 1$ (Figure 13), when $k = 5$, the median of the perturbation error of CFPA (respectively, CWPA with Haar) goes from 70% to 7% (respectively from 32% to 4%) when N goes from 50 to 550. Likewise, for $\varepsilon = 3$ (Figure 14), when $k = 5$, the median of perturbation error of CFPA (respectively, CWPA with Haar) goes from 25% to 2.7% (respectively from 11% to 2.1%) when N goes from 50 to 550. We notice that, the greater the N, the smaller the difference of the perturbation error between CFPA and CWPA is. This result is also true when ε increases. This can be explained by the decrease of the noise introduced on the coefficients of the transforms.

Figures 13 and 14 show that the perturbation error increases when k increases. The larger the k, the smaller the budget ε/k allocated to each coefficient is. This leads to a noise increase on each coefficient and thus on the perturbation error.

Table 3: Publishing algorithm with the smallest MRE according to the budget of privacy ε and the number of homes in the district (N).

Number of homes (N)	Budget of privacy, ε	Best algorithm	Coefficients, k	Median of MRE (%)
50	1	CWPA-Haar, db2	5	35
	3	CWPA-Haar		21
150	1	CWPA-Haar	5	19
	3	CFPA		11
250	1	CFPA	5	16
	3			8
350	1	CFPA	5	12
	3		8	7
450	1	CFPA	5	10
	3		8	6
550	1	CFPA	5	9
	3		8	5

6.3. Summary of the Experimental Results. The combination of the reconstruction error (Figure 11) and the perturbation error (Figures 13 and 14) enables to determine which transform is appropriate according to the number of homes N and the budget of privacy ε, for getting a total error as small as possible.

Lemma 4. *The mean relative error (MRE) of CFPA (respectively, CWPA) is lower than or equal to the sum of the reconstruction error and the perturbation error of CFPA (respectively, CWPA).*

Proof. The proof of the above lemma deferred to the appendix. □

Section 6.1 shows that the reconstruction error of DFT is lower than that of the considered DWT. For example, when $N = 450$ and $k = 5$, the median of the reconstruction error is 6% for DFT, while it is 13% for the Haar and Daubechies 2 transforms.

However, Section 6.2 shows that algorithms based on DFT (e.g., CFPA) have a higher perturbation error than those based on DWT (e.g., CWPA).

According to Lemma 4, the total error (MRE) is less than or equal to the sum of reconstruction error and perturbation error. Thus, if the reconstruction error or perturbation error is greater than 10%, there is a high probability that the final error will not be less than this threshold.

As the reconstruction error of the DWT is greater than 9%, there is a high probability that the final error of CWPA will not be less than this threshold, even if the Laplace noise decreases, i.e., when the number of homes N or the privacy budget ε increases. However, as the reconstruction error of the DFT is small (the median is between 2% and 3% when $k = 7, 8, 9$), then the total error of the CFPA may be lower than that of the CWPA when the impact of Laplace noise decreases. For example, the median of the perturbation error of CFPA is between 3% and 5% when $k = 7, 8, 9$, $N = 550$, and $\varepsilon = 3$. This analysis explains why, for $\varepsilon = 1$, the CWPA obtains a better utility than the CFPA when the number of homes N is less than 250. For example, when $N = 50$ and $k = 5$, the median of the perturbation error (respectively, the reconstruction error) of CFPA is 70% (respectively, 10%) against 32% (respectively, 18%) for CWPA using Haar. Thus, the median of MRE of CFPA is between 70% and 80% against 32% and 50% for CWPA.

When N is higher than 250, CFPA gets a better utility than CWPA. For example, when $N = 450$ and $k = 5$, the median of the perturbation error (respectively, the reconstruction error) of CFPA is 8.5% (respectively, 6.5%) against 4.5% (respectively, 16%) for CWPA using Haar. Thus, the median of MRE of CFPA is between 8.5% and 15% against 16% and 20.5% for CWPA.

In this use case, by comparing the different techniques for publishing time-series consumption, it appears that clamping perturbation algorithms (CFPA, CWPA) get a better utility than unbounded algorithms (FPA, WPA), which shows that the clamping mechanism reduces the total error. Furthermore, when the number of homes is greater than 250, CFPA obtains the best utility, with a mean relative error of less than 10% when $\varepsilon = 3$. When the budget of privacy $\epsilon = 1$, the mean relative error of CFPA is less than 10% for $N = 450$ homes.

The CWPA gets the best utility when the number of homes N is smaller than 150 and the budget of privacy ε is 1. This is justified by its low perturbation error.

Table 3 summarizes the publishing algorithm with the smallest MRE according to the budget of privacy ε and the number of homes in the district (N). Based on the dataset from the Irish Commission for Energy Regulation (CER) [39], Table 3 shows that the clamping Fourier perturbation algorithm (CFPA) achieves a lower MRE than the clamping wavelet perturbation algorithm (CWPA) for $N > 150$. Hence, CFPA gets a better utility than CWPA for $N > 150$.

7. Conclusion

The large deployment of smart meters provides users and suppliers with the capacity to optimize the energy consumption through forecasting and demand-response services. This paper proposes an original and efficient approach to mitigate privacy leakages of users' consumptions. This approach uses differential privacy and time-series transformations for supporting high privacy guarantees and utility. The clamping Fourier perturbation algorithm (CFPA)

we propose achieves an error 6 times lower than the Fourier perturbation algorithm (FPA). Similarly, the clamping wavelet perturbation algorithm (CWPA) achieves an error 2 times lower than the wavelet perturbation algorithm (WPA). Thanks to our algorithm, the publication of aggregate time-series consumptions is now possible while guaranteeing that the aggregate does not reveal any individual consumptions and while achieving better utility than existing algorithms. These privacy-preserving aggregate time-series consumptions can then be used as a building block, enabling services such as forecasting and demand-response, which are suitable for improving the efficiency and reliability of the electric grid.

In the future, we plan to investigate how to decentralize our clamping transform perturbation algorithm in order to resist to malicious aggregators. We plan to examine how to combine secure multiparty computation (SMC) with differential privacy (DP). SMC enables parties to compute a joint function without learning any individual inputs. SMC combined with DP could allow homes to compute and publish their aggregated consumptions without relying on an aggregator. However, SMC incurs a communication cost, which might have an impact on the running time performance.

Appendix

Proof of Lemma 4

Lemma Appendix (Lemma 4). *The mean relative error (MRE) of CFPA (respectively, CWPA) is lower than or equal to the sum of the reconstruction error and the perturbation error of CFPA (respectively, CWPA).*

Proof. Let $S = (S_1, \ldots, S_T)$ be the aggregate consumption to be published by using CFPA or CWPA. Let $C = (c_1, \ldots, c_T)$ be the coefficients of the considered transform of S. For simplicity, we consider that we use the CFPA5; we have $S = \mathrm{IDFT}(c_1, \ldots, c_T)$, where IDFT means the inverse of the DFT transform. We note $\overline{c}_j$ as the clamped coefficient of c_j for $j = 1, \ldots, k$. Let $\widehat{S} = \mathrm{IDFT}(\overline{c}_1 + \mathscr{L}(M_1\sqrt{2}/\varepsilon/k), \ldots, \overline{c}_k + \mathscr{L}(M_k\sqrt{2}/\varepsilon/k), 0, \ldots, 0)$ be the result of the aggregate consumption, where $M = (M_1, \ldots, M_k)$ is the maximum magnitude of the first k DFT coefficients. Let $d_j = \overline{c_j} - c_j$ for $j = 1, \ldots, k$:

$$\begin{aligned}
\widehat{S} &= \mathrm{IDFT}\left(\overline{c}_1 + \mathscr{L}\left(\frac{M_1\sqrt{2}}{\varepsilon/k}\right), \ldots, \overline{c}_k + \mathscr{L}\left(\frac{M_k\sqrt{2}}{\varepsilon/k}\right), 0, \ldots, 0\right) \\
&= \mathrm{IDFT}\left(d_1 + c_1 + \mathscr{L}\left(\frac{M_1\sqrt{2}}{\varepsilon/k}\right), \ldots, d_k + c_k + \mathscr{L}\left(\frac{M_k\sqrt{2}}{\varepsilon/k}\right), 0, \ldots, 0\right) \\
&= \mathrm{IDFT}(c_1, \ldots, c_T) + \mathrm{IDFT}\left(d_1 + \mathscr{L}\left(\frac{M_1\sqrt{2}}{\varepsilon/k}\right), \ldots, d_k + \mathscr{L}\left(\frac{M_k\sqrt{2}}{\varepsilon/k}\right), 0, \ldots, 0\right) \\
&\quad - \mathrm{IDFT}(0, \ldots, 0, c_{k+1}, \ldots, c_T).
\end{aligned} \tag{A.1}$$

Let $\widetilde{S} = (\widetilde{s}_1, \ldots, \widetilde{s}_T) = \mathrm{IDFT}(c_1, \ldots, c_k, 0, \ldots, 0) - \mathrm{IDFT}(c_1, \ldots, c_T) = -\mathrm{IDFT}(0, \ldots, 0, c_{k+1}, \ldots, c_T)$ corresponding to the difference between the aggregate consumption where the last $T - k$ DFT coefficients are replaced by zeros and the initial aggregate consumption (corresponding to the reconstruction error). Let $\overline{S} = (\overline{s}_1, \ldots, \overline{s}_T) = \mathrm{IDFT}(\overline{c}_1 + \mathscr{L}(M_1\sqrt{2}/\varepsilon/k), \ldots, \overline{c}_k + \mathscr{L}(M_k\sqrt{2}/\varepsilon/k), c_{k+1}, \ldots, c_T) - \mathrm{IDFT}(c_1, \ldots, c_T) = \mathrm{IDFT}(d_1 + \mathscr{L}((M_1\sqrt{2}/\varepsilon/k)), \ldots, d_k + \mathscr{L}(M_k\sqrt{2}/\varepsilon/k), 0, \ldots, 0)$ corresponding to the difference between the aggregate consumption where the first k DFT coefficients are clamped and noisy and the initial aggregate consumption (corresponding to the perturbation error). Then, we obtain $\widehat{S} = S + \overline{S} + \widetilde{S}$. Let $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ be two vectors of the same size; we note $A/B = (a_1/b_1, \ldots, a_n/b_n)$. Let $S + 1 = (s_1 + 1, \ldots, s_T + 1)$, $\widehat{S} - S/S + 1 = \overline{S}/S + 1 + \widetilde{S}/S + 1$. Then,

$$\begin{aligned}
\left\|\frac{\widehat{S} - S}{S + 1}\right\|_1 &= \left\|\frac{\overline{S}}{S + 1} + \frac{\widetilde{S}}{S + 1}\right\|_1 \\
&\leq \left\|\frac{\overline{S}}{S + 1}\right\|_1 + \left\|\frac{\widetilde{S}}{S + 1}\right\|_1.
\end{aligned} \tag{A.2}$$

Thus, the MRE of CFPA (respectively, CWPA) is lower than or equal to the sum of the reconstruction error and the perturbation error of CFPA (respectively, CWPA). □

Data Availability

The dataset used in this paper is from the Irish Commission for Energy Regulation available at https://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] V. Mishra, *An approach to recovery of critical data of smart cities using blockchain*, Ph.D. thesis, Arizona State University, Tempe, AZ, USA, 2017.

[2] C. S. Lai, Y. Jia, Z. Dong et al., "A review of technical standards for smart cities," *Clean Technologies*, vol. 2, no. 3, pp. 290–310, 2020.

[3] C. S. Lai, L. L. Lai, and Q. H. Lai, "Smart grids and big data analytics for smart cities," 2020.

[4] R. Weron, *Modeling and Forecasting Electricity Loads and-Prices: A Statistical Approach*, vol. 403, John Wiley & Sons, New York, NY, USA, 2007.

[5] T. De Souza, J. Wright, P. O'Hanlon, and I. Brown, "Set difference attacks in wireless sensor networks," in *Proceedings of the International Conference on Security and Privacy in Communication Systems*, Padua, Italy, September 2012.

[6] S. S. Clark, H. Mustafa, B. Ransford, J. Sorber, K. Fu, and W. Xu, "Current events: identifying webpages by tapping the electrical outlet," in *Proceedings of the ESORICS*, Egham, UK, September 2013.

[7] U. Greveler, P. Glösekötterz, B. Justusy, and D. Loehr, "Multimedia content identification through smart meter power usage profiles," in *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, Las Vegas, NV, USA, July 2012.

[8] M. Jawurek, F. Kerschbaum, G. Danezis, and SoK, *Privacy Technologies for Smart Grids - A Survey of Options*, Microsoft Res., Cambridge, UK, 2012.

[9] G. Bauer, K. Stockinger, and P. Lukowicz, "Recognizing the use-mode of kitchen appliances from their current consumption," *Lecture Notes in Computer Science*, vol. 9, pp. 163–176, 2009.

[10] A. Prudenzi, "A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel," *Power Engineering Society Winter Meeting*, vol. 2, 2002.

[11] G. W. Hart, "Nonintrusive appliance load data acquisition," in *Proceedings: International Load Management Conference*, Bonn, Germany, May 1985.

[12] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the Theory of Cryptography Conference*, pp. 265–284, Springer, New York, NY, USA, March 2006.

[13] G. Eibl and D. Engel, "Differential privacy for real smart metering data," *Computer Science - Research and Development*, vol. 32, no. 1-2, pp. 173–182, 2017.

[14] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, pp. 211–407, 2014.

[15] P. Barbosa, A. Brito, and H. Almeida, "A technique to provide differential privacy for appliance usage in smart metering," *Information Sciences*, vol. 371, pp. 355–367, 2016.

[16] H. Wang and Z. Xu, "Cts-dp: publishing correlated time-series data via differential privacy," *Knowledge-Based Systems*, vol. 122, pp. 167–179, 2017.

[17] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, pp. 735–746, Athens, Greece, November 2010.

[18] F. McSherry and I. Mironov, "Differentially private recommender systems: building privacy into the netflix prize contenders," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pp. 627–636, Paris, France, July 2009.

[19] L. Lyu, Y. W. Law, J. Jin, and M. Palaniswami, "Privacy-preserving aggregation of smart metering via transformation and encryption," in *Proceedings of the 2017 IEEE Trustcom/BigDataSE/ICESS*, pp. 472–479, IEEE, Sydney, Australia, August 2017.

[20] C. Rottondi, A. Barbato, L. Chen, and G. Verticale, "Enabling privacy in a distributed game-theoretical scheduling system for domestic appliances," *IEEE Transactions on Smart Grid*, vol. 8, pp. 1220–1230, 2016.

[21] C. Rottondi and G. Verticale, "Privacy-friendly appliance load scheduling in smart grids," in *Proceedings of the International Conference on Smart Grid Communications (SmartGridComm)*, Vancouver, Canada, October 2013.

[22] C. Thoma, T. Cui, and F. Franchetti, "Secure multiparty computation based privacy preserving smart metering system," in *Proceedings of the 2012 North American power symposium (NAPS)*, pp. 1–6, IEEE, Champaign, IL, USA, February 2012.

[23] F. Leukam Lako, P. Lajoie-Mazenc, and M. Laurent, "Reconciling privacy and utility for energy services–an application to demand response protocols," in *Proceedings of the 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS & PW)*, pp. 348–355, IEEE, Genoa, Italy, September 2020.

[24] M. A. Mustafa, S. Cleemput, A. Aly, and A. Abidin, "A secure and privacy-preserving protocol for smart metering operational data collection," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6481–6490, 2019.

[25] T. Dimitriou and M. K. Awad, "Secure and scalable aggregation in the smart grid resilient against malicious entities," *Ad Hoc Networks*, vol. 50, pp. 58–67, 2016.

[26] G. Danezis, C. Fournet, M. Kohlweiss, and S. Zanella-Béguelin, "Smart meter aggregation via secret-sharing," in *Proceedings of the Workshop on Smart Energy Grid Security*, Berlin Germany, November 2013.

[27] G. Ács and C. Castelluccia, "I have a DREAM! (DiffeRentially privatE smArt Metering)," in *Proceedings of the International Workshop on Information Hiding*, Prague, Czech Republic, May 2011.

[28] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart-grid," in *Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium*, Waterloo, Canada, July 2011.

[29] T. Jeske, "Privacy-preserving smart metering without a trusted-third-party," in *Proceedings of the Security and Cryptography (SECRYPT)*, Seville, Spain, August 2011.

[30] F. D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Proceedings of the International Workshop on Security and Trust Management*, Athens, Greece, September 2010.

[31] B. Pejó and D. Desfontaines, "Sok: differential privacies," 2020.

[32] S. L. Garfinkel, J. M. Abowd, and S. Powazek, "Issues encountered deploying differential privacy," in *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pp. 133–137, Toronto, Canada, October 2018.

[33] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1054–1067, Scottsdale, AZ, USA, November 2014.

[34] Differential Privacy Team, "Learning with privacy at scale," 2017.

[35] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 3571–3580, Long Beach, CA, USA, December 2017.

[36] G. Ács, C. Castelluccia, and R. Chen, "Differentially private histogram publishing through lossy compression," in *Proceedings of the 2012 IEEE 12th International Conference on*

Data Mining, pp. 1–10, IEEE, Brussels, Belgium, December 2012.

[37] "Regulation (EU) 2016/679 of the european parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation)," *Official Journal of the European Union*, vol. 119, 2016.

[38] C. Dwork, "Differential privacy: a survey of results," in *Proceedings of the International Conference on Theory and Applications of Models of Computation*, Xi'an, China, April 2008.

[39] "C for Energy Regulation (CER), CER smart metering project - electricity customer behaviour trial, 2009-2010," 2012, http://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

[40] R. Wang, *Continuous- and Discrete-Time Wavelet Transforms*, Cambridge University Press, Cambridge, UK, 2012.

[41] E. J. Stollnitz, T. D. DeRose, A. D. DeRose, and D. H. Salesin, *Wavelets for Computer Graphics: Theory and Applications*, Morgan Kaufmann, Burlington, MA, USA, 1996.