Differential Private Motion Sensor and Wasted Energy in Building Energy Management System

SIRAVIT KWANKAJORNKEAT AND CHAODIT ASWAKUL, (Member, IEEE)
Department of Electrical Engineering, Faculty of Engineering, Wireless Network and Future Internet Research Unit, Chulalongkorn University, Bangkok 10330, Thailand
Corresponding author: Chaodit Aswakul (chaodit.a@chula.ac.th)
This work was supported by the TEIN*cc Asi@Connect’s Data-Centric IoT-Cloud Service Platform for Smart Communities (IoTcloudServe@TEIN) Project through the umbrella of Smart-Energy@Chula demonstration site under Contract ACA 2016/376-562 Asi@Connect-18-061.

ABSTRACT Building energy management system (BEMS) uses various sensors and smart meters to detect power consumption and user movement within buildings. By applying data integration, raw BEMS data can be employed for multiple purposes. For example, power consumption data can be matched with motion sensor data and electric appliance details in any staff room to calculate wasted energy and continually predict an individual electric appliance status. Therefore, disclosing these data severely violates local electricity users’ privacy. To resolve data privacy concerns using differential privacy, the traditional Laplacian mechanism cannot be used to detect overloads and the standard coin-flipping algorithm gives a high percentage error of wasted energy calculation. Thus, this paper proposes a modified truncated Laplacian noise addition and random motion state processes to hide the actual electrical use and movement patterns. Based on a 30-day sample period of actual BEMS, both privacy and usability of the proposed processes have been compared with the traditional methods in terms of the privacy loss parameter, the reported error of energy consumption, the correctness of overload detection, and the error of the wasted energy calculation. Based on the obtainable effectiveness, these proposed processes are recommended for further extension to implement at BEMS data gateways in real-time.

INDEX TERMS Data integration, data privacy, differential privacy, energy consumption, energy management, motion sensor.

I. INTRODUCTION
A building energy management system (BEMS) deploys many sensors [1] and smart meters [2] that send real-time data to a data storage. The storage collects the data at specified time intervals (e.g., 1 minute and 1 hour). The data can be utilized for load monitoring, air-conditioner control, and weekly pattern data analysis. A BEMS usually consists of gateways to collect data from sensors and smart meters and make decisions about releasing data to consumers, such as data analytic research teams or other data user counterparts. For example, a demonstration of BEMS in Chulalongkorn University (CU-BEMS) runs such operations of data collection and processing at the time granularity of 1 minute [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Alon Kuperman.

The CU-BEMS also has a test facility for renewable energy, such as solar cells and wind turbines, to demonstrate the renewable energy potential. For sensor data requests, uniform resource identifiers (URIs) with specified sensor types, dates, times, and locations must be sent to the storage to query sensor data of an area, and the real data will be returned. Like other BEMS studies, CU-BEMS analytics can then be utilized to raise public awareness of electricity use by presenting the analyzed raw data in various ways [4], such as an interactive visualization energy game [5] and its advanced version with IDaaS architecture [6]. For another example, if no motion has been detected for at least 15 time points (minutes) and the smart meter in that area measures at least 15 W (watt) of average power consumed during those time points, then wasted energy is said to occur [7]. Such incidences result in wasted energy could be up to 30% of the annual electricity...
bill [7], and they are caused by preventable scenarios, such as people taking breaks without turning off unnecessary air-conditioning systems or lights.

However, opting for data usability alone without careful data release governance causes easily violated individual privacy. There are many conventional ways to protect individual local electricity users’ privacy, where BEMS sensors must continuously carry out their monitoring tasks. Original encryption had worked well until machine learning algorithms became available to analyze encrypted data [8]. Additionally, such an encryption method becomes inappropriate for public data sharing since the encrypted data must be decrypted before sharing with the public [9]. Also, cryptographic approaches become irrelevant for data consumers that are intended by design to see unencrypted data. Likewise, anonymized data can be matched with non-anonymized data to reidentify an individual [9]. This technique is referred to as the linking attack [10]. Moreover, encryption and anonymization techniques are computationally complex, so that the system slows. Hence, such conventional methods are no longer viable. The challenge of securely releasing data without jeopardizing privacy is prominent, especially when BEMS relies on addressing that could be linked to the physical world. For the CU-BEMS context, as an example that is commonly found in a BEMS, full name details of the staff occupying each room are publicly accessible, e.g., via the campus websites. Assume that staff room details are matched with the data released from BEMS. In this case, data requesters may know enough even at the level of an individual movement from electricity usage patterns or directly from motion sensors. To worsen matters, criminals who obtain actual data can easily choose when to commit criminal acts if the BEMS is a part of the enterprise or governmental facility.

In recent solutions to resolve privacy concerns, we have chosen differential privacy due to its simplicity, controllable privacy, aggregated original data persistence [9], and scalability [11]. Currently, differential privacy processes have been applied to data in many sectors, such as personal information in organizations [12], private data sharing for collaborative filtering purposes [13], and other infrastructures, including energy systems [9]. The main concept of differential privacy is to apply necessary noise addition, e.g., Laplacian, exponential, and Gaussian noise patterns [9], or randomization like the coin-flipping example in [14], to statistical and real-time individual private data before being released for public access [12]. In this research, data types such as energy consumption and local electricity movements are considered in privacy processing since these readings refer to the activities of electricity users in an area. However, the renewable energy sources in CU-BEMS are not consumed by staff rooms, which are privacy concern areas, so differential privacy processes are not needed to apply with them. The challenge of differential privacy is to balance privacy and usability of data [9]. The conclusions of using randomized data should be correct or in the acceptable error range of a system [11], [14]. Motivated by this challenge, we propose an overload detectable truncated Laplacian mechanism and a conditional random motion states algorithm with a low percentage error of wasted energy calculation as the main contributions of this paper. In addition, the proposed truncated Laplacian mechanism can be applied to renewable energy sources, such as solar cells and wind turbines, if they are designed to be consumed by staff rooms with arising privacy concerns in the future.

In the remainder of this paper, related literature and research gaps are mentioned in Section II. Some basic concepts of this paper are mentioned in Section III. The overload detectable truncated Laplacian mechanism is mentioned in Section IV. The random algorithm to hide the actual time-series data of the motion sensor and wasted energy in an area is mentioned in Section V. The privacy and usability of the proposed processes are discussed in Section VI. This paper is concluded in Section VII.

II. RELATED WORK

There have been many differential privacy solutions for energy systems in the past. The traditional Laplacian noise addition was usually applied in many studies, e.g., random Laplacian noise for privacy-preserved smart metering in [11], differential privacy-based distributed load balancing for smart grid in [15], non-intrusive load monitoring in a compressive sensing framework in [16], differentially private data clustering for intelligent electrical IoT in [17], and deep learning non-intrusive load monitoring in [18]. Some energy systems use Gaussian noise for data obfuscation. For example, a perturbation method using Gaussian noise was combined with encryption [19] to prevent filtering and true value attacks. Additionally, geometric noise addition can be applied with plug readings [20].

The typical constraint of the randomization mechanism in a BEMS is to hide the actual time-series data of electrical use in the queried area. Apart from other studies, the additional conditions of this research are to detect overload and calculate wasted energy in an area. For overload detection, traditional Laplacian and Gaussian noises cannot be applied to actual power usage since their distribution range is infinite. Noise-added power usage may exceed the power rating of all electric appliances in the queried area, although the actual power usage is less than the power rating. Therefore, an intended truncated noise addition is required for overload detection.

There are some studies on energy systems using truncated Laplacian noise addition techniques. For instance, truncated Laplacian noise, normalized by adding a constant parameter, was applied to preserve the privacy of capacity-limited rechargeable batteries [21]. Another example of battery-based differential privacy preservation, truncated Laplacian noise, normalized by scaling a parameter, was applied to rechargeable batteries to obfuscate smart meter readings [22]. These noise patterns can be modified for overload detection purposes in a BEMS. However, the accuracy of the aggregated noise-added energy consumption values is not guaranteed due to non-zero mean noise addition.
To apply differential privacy with a constrained system, feasible distribution ranges of the randomization mechanism must be specified [23]. However, some infeasible ranges are allowed for noise addition. For example, negative noise-added power consumption values are allowed in [11], [24]. The truncated Laplacian noise pattern in [25]–[27] is symmetric, so regression analysis [28] like the Bayesian inference attack [8] can be easily executed since the distribution ranges of different noise-added values differ. Additionally, symmetric truncated Laplacian noise gives a high probability that actual data will be leaked [29]. The truncated Laplacian noise pattern in [30] becomes difficult to apply with energy systems since its distribution range is [0, 1].

Therefore, unlike the past studies, this paper proposes to apply asymmetric two-side scaled truncated Laplacian noise to the original time-series electricity consumption data. This paper evaluates this noise addition process for non-intrusive load monitoring and overload detection. In addition, to the best of our knowledge, there is still no wasted energy based differentially private algorithm for motion sensor data. The challenge of generating a random algorithm for motion sensor data is to minimize the percentage error of the wasted energy calculation. Furthermore, since motion sensor data in an area used to calculate wasted energy refer to the presence and absence of people in that area, actual sensor data must not be shown. For example, consider the case in which the actual occurrence of wasted energy time is shown, which indicates that nobody (no motion detected) is in the area in that duration. Information hiding must be carefully crafted both in the value of time-series data and the indexing time, where interesting events occur. Therefore, apart from other energy systems, this paper presents an algorithm to solve this challenge by applying time shifting to a randomization algorithm that is adapted from a coin-flipping example in [14]. With the proposed algorithm, wasted energy with a low percentage error can be displayed to raise public awareness of electricity use while mitigating the risk of privacy leakage.

III. PRELIMINARY

In this section, we introduce the concepts of differential privacy and time-series data.

A. DIFFERENTIAL PRIVACY

Differential privacy [9], [12] aims to protect an individual private data by applying randomization processes to the statistical and real-time data that can be used to identify an individual [9]. However, non-private personal information may be revealed [12]. In terms of usability, the conclusions of using randomized data should be correct or in the acceptable error range of the system [11], [14].

A database in differential privacy is defined in terms of a vector. Each element in the vector is the number of database records belonging to a data type [14]. Let \( U = \{\text{datatype}_1, \ldots, \text{datatype}_m\} \) be a database universe, \(|U|\) be the number of data types in \( U \) (in this case, \(|U| = m\)), \( \alpha = [\alpha_1 \ldots \alpha_m] \in (\mathbb{R}_0^+)^{[U]} \) be a database record where \( \mathbb{R}_0^+ = \{0, 1, 2, 3, \ldots\} \), \( \alpha_i \) be a number of \( i^{th} \) type (\( \text{datatype}_i \)) in database \( \alpha \), and \( ||\alpha||_1 = \sum_{i=1}^{[U]} |\alpha_i| \) be a norm \( l_1 \) of \( \alpha \).

The distance between two databases is the sum of the differences in the number of records of each data type between any two isometric vectors, that is, the database distance between two isometric vectors representing the databases, \( \alpha, \gamma \in (\mathbb{R}_0^+)^{[U]} \), is [14]

\[
||\alpha - \gamma||_1 = \sum_{i=1}^{[U]} |\alpha_i - \gamma_i|
\]

For example, if \( \alpha = [2 \ 3 \ 4 \ 4] \) and \( \gamma = [2 \ 3 \ 3 \ 5] \) then \( ||\alpha - \gamma||_1 = |2 - 2| + |3 - 3| + |4 - 3| + |4 - 5| = 2 \).

In the case that \( ||\alpha - \gamma||_1 \leq 1 \), if \( \mathcal{R} \) is an \((\epsilon, \eta)\)-differentially private randomization process, then for all \( \mathcal{V} \subseteq \text{Range}(\mathcal{R}) \), the following probability condition must be satisfied:

\[
P(\mathcal{R}(\alpha) \in \mathcal{V}) \leq e^\epsilon P(\mathcal{R}(\gamma) \in \mathcal{V}) + \eta
\]

Therefore, the privacy loss parameter of the process \( \mathcal{R} \) in a subset \( \mathcal{V} \) of \( \text{Range}(\mathcal{R}) \), \( L_{\mathcal{R}}(\mathcal{V}) \), satisfies the condition that

\[
P \left( \frac{L_{\mathcal{R}}(\mathcal{V})}{\epsilon} \geq 1 - \eta \right) \geq 1 - \eta
\]

is the privacy loss boundary does not exceed \( \epsilon \) with probability \( 1 - \eta \) [14] where \( \eta \) is the probability that actual data will be leaked [29] by using the process \( \mathcal{R} \).

In practice, the privacy loss boundary ranges from 0.01 \( \leq \epsilon \leq 7 \) [29] and \( \eta \ll 1/|U| \) are utilized since \( \eta \approx 1/|U| \) permits a few of the actual records to be published [14]. Consider a universe of \( m \) data types (\( |U| = m \)); if \( \eta \approx 1/|U| = 1/m \) then the number of the actual records released to the public is expected to be \((m)(1/m) = 1\) record. For the privacy context, the smaller \( \epsilon \) is, the higher the privacy. However, privacy should be balanced with usability.

B. TIME-SERIES DATA

For any sensor’s time-series data, let \( t \in \{0, 1, \ldots, 1439\} \) be a time point (minutes) after midnight of each considered day. A negative time point refers to the previous day’s time point. For example, time point \( t = -60 \) refers to the 1380\(^{th}\) time point (or 23:00) of the day before the considered day. A time point with a value exceeding 1439 refers to the next day’s time point, e.g., time point \( t = 1799 \) refers to the 360\(^{th}\) time point (or 05:59) of the day after the considered day.

IV. DIFFERENTIAL PRIVACY FORMULATION FOR ELECTRICITY CONSUMPTION

This section focuses on applying differential privacy, our proposed truncated Laplacian mechanism, with electrical use data. This proposed mechanism will be compared and contrasted with the traditional Laplacian mechanism in terms of privacy and usability in Section VI. The notations of this section are summarized in Table 1.
Apart from $x_t$, if $y_t$ is another electrical use vector where $||x_t - y_t||_1 \leq 1$ and all $S \subseteq \text{Range}(A)$ then $A$ is $(\epsilon, \delta)$-differentially private when \[ P(A(x_t) \in S) \leq e^\epsilon P(A(y_t) \in S) + \delta \] (3)

By using generated intended noise, the released noise-added power consumption data should be practicable for purposes that do not violate privacy. The total energy usage computed from the noise-added data on a period must be similar to the actual value, e.g., error lower than 2.5% for Thailand’s Provincial Electricity Authority billing constraints. Additionally, the data could also be made usable for other purposes, for example, load prediction and other abnormal electrical use detection [11].

In the following, we first discuss why using traditional noise distributions here would incur undesired consequences. Firstly, since the Laplacian distribution range is $(-\infty, \infty)$, applying this noise addition may cause false-positive and false-negative overload detection. If the actual power usage at time $t$, denoted by $\mu_t$, is less than $P_r$, but the input noise is so high that the noise-added power usage exceeds $P_r$, then this situation causes false-positive overload detection (refer to Fig. 1). On the other hand, if the actual power usage at that time exceeds $P_r$, but the noise-added power usage is lower than $P_r$, then this situation leads to false-negative overload detection. Secondly, for symmetric truncated Laplacian noise patterns, different distribution ranges of different noise-added values yield high privacy loss (refer to Fig. 2).

Due to these drawbacks foreseen if the symmetric truncated Laplacian noises are here used, therefore, the distribution range of noise-added data should be truncated and asymmetric. To detect electrical overload, intended noise-added power usage values under normal loadings must still be lower than the power rating of all electrical appliances in the queried area $P_r$. However, power usage values exceeding $P_r$ will be shown without any noise addition. The maximum of the intended noise distribution range is $P_r$, which means the released noise-added data distribution range is $[u_0, P_r]$, where $u_0$ is set so that the noise addition mechanism’s privacy loss parameter is in the range $[0.01, 7]$ [29]. Thus, modified truncated Laplacian noise is proposed in this paper to elide infeasible ranges [23] that affect typical constraints of the systems. However, infeasible ranges that do not affect the constraints of the systems are still acceptable. In this research, $u_0$ is set to $-P_r$, which means $\text{Range}(A) = [-P_r, P_r]$, where the infeasible range $[-P_r, 0)$ is acceptable for this

### TABLE 1. Differential private electricity consumption model notations in queried area.

| Notation | Description |
|----------|-------------|
| $P(E)$  | Probability that an event $E$ occurs |
| $x_t = [x_{1,t} \ldots x_{m,t}]$ | Electrical use vector in the queried area at time $t$. |
| $x_{i,t}$ | Status of $i^{th}$ appliance at time $t$. Its value can be either 0 (the $i^{th}$ appliance is not being used at time $t$) or 1 (the $i^{th}$ appliance is being used at time $t$). |
| $P_i$   | Actual power usage of $i^{th}$ appliance (in kW unit) |
| $f(x_t) = \sum_{i=1}^{m} x_{i,t}P_i$ | Query function to return the actual power usage of any electrical use vector $x_t$ in the queried area [31]. |
| $\mathcal{A}$ | Differential private randomization process. |
| $\text{Range}(\mathcal{A})$ | Set of all possible values randomized by using the process $\mathcal{A}$ |
| $\mathcal{L}_A^{(S)}$ | Privacy loss parameter of the process $\mathcal{A}$ in a subset $S$ of $\text{Range}(A)$ |
| $\mathcal{W}$ | Any intended zero-mean noise |
| $\mathcal{A}(x_t) = f(x_t) + \mathcal{W}$ | $(\epsilon, \delta)$ noise addition mechanism [14] to hide the actual power usage in the queried area [31]. The condition of this mechanism is further defined in (3). |
| $\epsilon$ | Maximum value of $\mathcal{L}_A^{(S)}$ where $S \subseteq \text{Range}(A)$ [14] |
| $\delta$ | Probability of actual data being leaked [29] by using the mechanism $\mathcal{A}$ |
| $P_r$  | Sum of all electric appliances’ power rating in the queried area |

VOLUME 10, 2022
research because only aggregated energy consumption values of periods are released accurately for any data analytic to protect electricity users’ privacy. In contrast, if the actual power usage value of the queried area is revealed, the privacy of the electricity users in that area will be violated.

For each time point with an overload, the actual power usage will be released without any noise addition to show the occurrence of overload. However, this paper proposes that the following modified truncated Laplace noise is added to the data for other time points without the overload. In particular, the following modified truncated Laplacian noise is added to the electricity users in that area will be violated.

The probability density function (pdf) normalization of (4) requires that

$$\int_{-\mathcal{P}_r}^{\mathcal{P}_r} \frac{1}{b} e^{(u-\mu_t)/b} du = \int_{-\mathcal{P}_r}^{\mathcal{P}_r} \frac{1}{b} e^{-(u-\mu_t)/b} du + \int_{\mu_t}^{\mathcal{P}_r} \frac{1}{b} e^{-(u-\mu_t)/b} du$$

$$= r e^{(u-\mu_t)/b} - \frac{\rho}{b} e^{-(u-\mu_t)/b} \bigg|^{-\mathcal{P}_r}_{\mathcal{P}_r}$$

$$= r (1 - e^{-(\mathcal{P}_r+\mu_t)/b}) + \rho (1 - e^{-(\mathcal{P}_r-\mu_t)/b}) = 1$$

Further, we propose that the expected value of the distribution in (4) is herein set equal to the actual data value $\mu_t$:

$$E[U] = \int_{-\mathcal{P}_r}^{\mathcal{P}_r} u \frac{1}{b} e^{(u-\mu_t)/b} du$$

$$= \int_{-\mathcal{P}_r}^{\mathcal{P}_r} \frac{1}{b} u e^{(u-\mu_t)/b} du + \int_{\mu_t}^{\mathcal{P}_r} \frac{1}{b} u e^{-(u-\mu_t)/b} du$$

$$= \frac{r}{b} \left( bue^{(u-\mu_t)/b} \bigg|_{-\mathcal{P}_r}^{\mathcal{P}_r} - \frac{\rho}{b} e^{-(u-\mu_t)/b} \bigg|_{\mathcal{P}_r}^{\mu_t} \right)$$

$$+ \frac{\rho}{b} \left( -bue^{-(u-\mu_t)/b} \bigg|_{-\mathcal{P}_r}^{\mathcal{P}_r} + b \frac{1}{b} e^{-(u-\mu_t)/b} \bigg|_{\mathcal{P}_r}^{\mu_t} \right)$$

$$= r \left( u e^{(u-\mu_t)/b} \bigg|_{-\mathcal{P}_r}^{\mathcal{P}_r} - \frac{\rho}{b} e^{-(u-\mu_t)/b} \bigg|_{\mathcal{P}_r}^{\mu_t} \right)$$

$$+ \rho \left( -ue^{-(u-\mu_t)/b} \bigg|_{\mathcal{P}_r}^{\mu_t} + b e^{-(u-\mu_t)/b} \bigg|_{\mathcal{P}_r}^{\mu_t} \right)$$

$$= r \left( \mu_t + (\mathcal{P}_r + b) e^{-(\mathcal{P}_r+\mu_t)/b} - b \right)$$

$$+ \rho \left( \mu_t - (\mathcal{P}_r + b) e^{-(\mathcal{P}_r-\mu_t)/b} + b \right)$$

$$= \mu_t$$

To solve for $r$ and $\rho$ using (5) and (6), let

$$c_1 = 1 - e^{-(\mathcal{P}_r+\mu_t)/b}$$

$$c_2 = 1 - e^{-(\mathcal{P}_r-\mu_t)/b}$$

$$c_3 = \mu_t + (\mathcal{P}_r + b) e^{-(\mathcal{P}_r+\mu_t)/b} - b$$

$$c_4 = \mu_t - (\mathcal{P}_r + b) e^{-(\mathcal{P}_r-\mu_t)/b} + b$$

Rewrite (5) and (6) as

$$c_1 r + c_2 \rho = 1$$

$$c_3 r + c_4 \rho = \mu_t$$

With algebraic derivations, the parameters $r$ and $\rho$ can be expressed as

$$r = \frac{c_2 \mu_t - c_4}{c_2 c_3 - c_1 c_4}$$

$$\rho = \frac{c_3 - c_1 \mu_t}{c_2 c_3 - c_1 c_4}$$
Let \( \nu \) be a uniformly distributed random value in the interval (0, 1) and, by using the inverse transform method, the noise-added power usage at time \( t \) is

\[
\mathcal{A}(x_t) = \begin{cases} 
\mu_t + b \ln \left( \frac{\nu + r - rc_1}{r} \right); & 0 < \nu < rc_1 \\
\mu_t + b \ln \left( \frac{\varphi}{\varphi + rc_1 - \nu} \right); & rc_1 \leq \nu < 1 
\end{cases}
\]

(15)

To obtain the privacy loss parameter as parameterized by \( \epsilon \), let \( \mu_{x,t} \) and \( \mu_{y,t} \) be the actual power usage at time \( t \) of the electrical use vectors \( x_t \) and \( y_t \), respectively. The distributions of \( \mathcal{A}(x_t) \) and \( \mathcal{A}(y_t) \) that satisfy the privacy loss parameter (from (5)) and expected value (from (6)) conditions are

\[
h_x(u) = \begin{cases} 
\frac{r_x}{b} e^{(u-\mu_{x,t})/b}; & \mathcal{P}_r \leq u < \mu_{x,t} \\
\frac{\varphi_x}{b} e^{-(u-\mu_{x,t})/b}; & \mu_{x,t} \leq u \leq \mathcal{P}_r 
\end{cases}
\]

(16)

and

\[
h_y(u) = \begin{cases} 
\frac{r_y}{b} e^{(u-\mu_{y,t})/b}; & \mathcal{P}_r \leq u < \mu_{y,t} \\
\frac{\varphi_y}{b} e^{-(u-\mu_{y,t})/b}; & \mu_{y,t} \leq u \leq \mathcal{P}_r 
\end{cases}
\]

(17)

respectively.

From (16) and (17), suppose that \( ||x_t - y_t||_1 = 1 \), and let \( S = du \) be any narrow interval in \([-\mathcal{P}_r, \mathcal{P}_r]\) and \( \mu_{x,t} > \mu_{y,t} \). Hence, \( \epsilon \) can be calculated by

\[
L^S_A = \ln \left( \frac{P(\mathcal{A}(x_t) \in S)}{P(\mathcal{A}(y_t) \in S)} \right) = \ln \left( \frac{h_x(u)}{h_y(u)} \right) du = \ln \left( \frac{h_x(u)}{h_y(u)} \right) \frac{\mu_{y,t} - \mu_{x,t}}{b} + \ln \left( \frac{r_x}{r_y} \right); \quad -\mathcal{P}_r \leq u < \mu_{y,t} \\

= \ln \left( \frac{\mu_{x,t} - \mu_{y,t}}{b} + \ln \left( \frac{r_x}{r_y} \right); \quad \mu_{y,t} \leq u < \mu_{x,t} \\

= \ln \left( \frac{\mu_{x,t} - \mu_{y,t}}{b} + \ln \left( \frac{\varphi_x}{\varphi_y} \right); \quad \mu_{x,t} \leq u \leq \mathcal{P}_r 
\right)
\]

(18)

The variables \( r_x, r_y, \varphi_x, \) and \( \varphi_y \) are presented in (13) and (14) by changing \( \mu_y \) to \( \mu_{x,t} \) and \( \mu_{y,t} \). The remaining degree-of-freedom parameter, \( b \), is numerically tried out with values in the range [0.001, 10] using a step size of 0.001 and all possible \( (x_t, y_t) \) pairs, where \( ||x_t - y_t||_1 = 1 \), to minimize the privacy loss parameter of a proposed mechanism in (18). Once the appropriate value of \( b \) is obtained, this value will be applied without recalculation. The maximum absolute value of (18) is \( \epsilon \) because \( L^S_A \leq \epsilon \) as required by differential privacy.

Assume that for each time interval \( [t, t+1) \), the power usage is at the same level during the interval. The energy consumption (in kWh units) in each minute is the product of the power usage (in kW units) at time \( t \) and the interval size (which is 1/60 hours). Therefore, the actual and noise-added energy consumption in any time period, \([t_{first}, t_{last}]\) where \( t_{last} > t_{first} \), can be calculated by

\[
\mathcal{E}_c = \sum_{t=t_{first}}^{t_{last}-1} f(x_t) / 60 
\]

and

\[
\mathcal{E}_{c, noise} = \sum_{t=t_{first}}^{t_{last}-1} A(x_t) / 60 
\]

(19)

(20)

respectively.

V. DIFFERENTIAL PRIVACY FORMULATION FOR MOTION SENSOR

For motion sensor privacy, if the actual movement data are revealed, data requesters will know the actual motion in the queried area all the time, including the case that the actual occurrence time of wasted energy is shown. Hence, differential privacy and time-shifting are applied with motion sensor data to hide the actual movement of local electricity users while maintaining the accuracy of wasted energy calculation. Table 2 summarizes all main notations in this section.

The energy consumption in an area is defined as wasted energy if a smart meter measures at least 15 W of average power consumption in that area where the motion sensor states have continuously been 0 for at least 15 minutes [7]. For each of these minutes, wasted power usage is the power consumption at that time. For all the other minutes, the wasted power usage is 0. This definition of wasted energy [7] is motivated by the circumstance of area users in which they forget to turn off air-conditioners in their area when they leave the area for a prolonged period. A past study revealed that such wasted energy could account for up to 30% of the total building energy consumption, which explains why it is essential to allow this metric to be extractable from the BEMS data release to improve building user engagement in raising public awareness without harming privacy.

The differential private randomization algorithm formulation must also be carefully designed to ensure an acceptable percentage error of wasted energy calculation in maintaining the usefulness of released noisy data. Information hiding with such randomization must be carefully crafted both in the value of time-series data and the indexing time when interesting events would occur. In this case, such interesting events are the time points when wasted energy occurs.

Let \( g[t] \) be another motion state vector, where \( ||s[t] - g[t]||_1 = 1 \) (i.e., if \( s[t] = 0 \), \( g[t] = 1 \), and vice versa), and all \( B \subseteq R^\{M\} = \{0, 1\} \). The process \( M = (\xi, 0) \)-differentially private when [14]

\[
P(M(s[t]) \in B) \leq e^\xi P(M(g[t]) \in B) 
\]

(21)
TABLE 2. Differential private motion sensor model notations in queried area.

| Notation | Description |
|----------|-------------|
| \( \tau = [... r[0] r[1] r[2] ...] \) | Actual motion state vector of the queried area by combining the states of all motion sensors in the queried area with 'or' logic at every time point |
| \( r[t] \) | Actual motion state at time \( t \) in the queried area. Its value can be either 0 (no motion detected) or 1 (motion detected). |
| \( s \) | Time-shifted motion state vector by letting \( s[t] = r(t - t_s) \) where \( t_s \) represents time shifting value, e.g., \( t_s = 15 \) minutes |
| \( \mathcal{M} \) | \( (\xi, 0) \) differential privacy algorithm [14] to hide the motion sensor data in the queried area. The condition of this mechanism is further defined in (21). |
| \( \nu \) | Conditional randomized motion state vector using algorithm \( \mathcal{M} \) |
| \( \xi \) | Maximum value of \( \max\{\mathcal{M}(B)\} \) where \( B \subseteq \text{Range}\{\mathcal{M}\} = \{0, 1\} \) [14] |
| \( \mathcal{P}_w \) | Actual wasted power vector which contains \( \mathcal{P}_w[t] \) value in the queried area at every time point. The \( \mathcal{P}_w[t] = f(x_t) \) (see Table 1) if wasted energy occurs, otherwise \( \mathcal{P}_w[t] = 0 \). |
| \( \mathcal{P}_{w,alg} \) | Wasted power vector of the queried area based on a motion state vector randomized by an algorithm. It contains \( \mathcal{P}_{w,alg}[t] \) value at every time point. The \( \mathcal{P}_{w,alg}[t] = \mathcal{A}(x_t) \) (see Table 1) if wasted energy occurs, otherwise \( \mathcal{P}_{w,alg}[t] = 0 \). |
| \( \text{Rand}(0, 1) \) | Uniformly distributed random value in the interval (0, 1) |
| \( N(t - t_w, t] \) | Number of time points with no motion detected in \( t \) during time period \( [t - t_w + 1, t] \) where \( t_w \) is time window |
| \( n(t - t_w, t] \) | Number of time points with wasted energy in the queried area during time period \( [t - t_w + 1, t] \) |

In this paper, we propose Algorithm 1 to find \( v[t] = \mathcal{M}(s[t]) \) that satisfies the required probability condition in (21). Our algorithm is adapted from a coin-flipping example in [14].

Algorithm 1 Differential Private Motion State

For any time point \( t \) with a probability \( p \) defined in (27):

```plaintext
if (1st Rand(0, 1) \leq p or \mathcal{P}_w[t - t_s] > 0) then
    v[t] \leftarrow s[t]
else
    if (2nd Rand(0, 1) \leq p) then
        v[t] \leftarrow 0
    else
        v[t] \leftarrow 1
end if
end if
```

The difference between Algorithm 1 and the standard coin-flipping example in [13] pertains to the condition of the if statement and the value being released. By using Algorithm 1, both the probabilistic condition and wasted power condition are employed in the if statement. Thus, the probability parameter \( p \) is time-varying and the occurrence time of wasted energy is shifted from the actual time for approximately \( t_s \) time points. However, the standard coin-flipping example in [13] uses only the probabilistic condition \( 1^{st} \text{Rand}(0, 1) \leq p \), where \( p \) is set to 0.5, in the if statement.

From the wasted energy definition, the 0 state in a motion state vector at a time point does not always infer the occurrence of wasted energy unless the motion states have continuously been 0 for at least 15 minutes. Therefore, the time-shifted motion state \( s \) must be shifted for at least 15 minutes to determine whether wasted energy will occur.

The privacy loss parameter of using a coin-flipping algorithm in [14] is \( \ln(3) \). To obtain the parameter \( p \) of Algorithm 1, let \( B \) in (21) to be \( \{s[r]\} \), then

\[
P[v[t] \in \{s[r]\}] = P[v[t] = s[r]] \\
= P[\mathcal{M}(0)] \\
= P[\mathcal{M}(1) = 1] \tag{22}
\]

The probability that wasted energy occurs at time \( t \) is assumed to be calculated by using the statistical data in time period \( [t - t_w + 1, t] \). The events \( \mathcal{P}_w[t] > 0 \) and \( v[t] = 0 \) are dependent. Therefore, the probability that wasted energy occurs at
time \( t \) if \( v[t] = 0 \) is

\[
P[\mathcal{P}_W[t] > 0 \mid v[t] = 0] = \begin{cases} 0; & N(t - s_5 - t_w, t - s_5) = 0 \\ n(t - t_w, t); & \text{otherwise} \end{cases} \tag{23}
\]

and from \( s[r] = v[t - t_1] \) in Table 2, \( P[\mathcal{P}_W[t - t_1] > 0 \mid v[t - t_1] = 0] \) can be rewritten as:

\[
P[\mathcal{P}_W[t - t_1] > 0 \mid s[r] = 0] = \begin{cases} 0; & N(t - s_5 - t_w, t - s_5) = 0 \\ n(t - s_5 - t_w, t - s_1); & \text{otherwise} \end{cases} \tag{24}
\]

The events \( 1^{st} \text{Rand}(0, 1) \) and \( 2^{nd} \text{Rand}(0, 1) \) are independent. These events are also independent of the events that \( \mathcal{P}_W[t - t_1] > 0 \) and \( s[r] = 0 \), so:

\[
P[\mathcal{M}(0) = 0] = P[\mathcal{P}_W[t - t_1] > 0 \mid s[r] = 0] = 1 - P[1^{st} \text{Rand}(0, 1) = 0] + P[1^{st} \text{Rand}(0, 1) > 0] \tag{25}
\]

and

\[
P[\mathcal{M}(1) = 1] = P[1^{st} \text{Rand}(0, 1) \leq p] + P[1^{st} \text{Rand}(0, 1) > p] \tag{26}
\]

With constraint in (22), combining (25) and (26) results in

\[
p = \frac{1}{2} - \frac{n(t - s_5 - t_w, t - s_5)}{2N(t - s_5 - t_w, t - s_5)} = 0.5(1 - P[\mathcal{P}_W[t - t_1] > 0 \mid s[r] = 0]) \tag{27}
\]

and

\[
P[v[r] = s[r]] = \frac{3(N(t - s_5 - t_w, t - s_1))^2 + (n(t - s_5 - t_w, t - s_5))^2}{4(N(t - s_5 - t_w, t - s_5))^2} \tag{28}
\]

Therefore, \( P[v[r] \neq s[r]] \) is equal to \( P[\mathcal{M}(0) = 1] = 1 - P[\mathcal{M}(0) = 0] = 1 - P[\mathcal{M}(1) = 1] = P[\mathcal{M}(1) = 0] \). The privacy loss parameter is:

\[
\mathcal{L}_M^{(B)} = \ln \left( \frac{P[\mathcal{M}(0) = 0]}{P[\mathcal{M}(0) = 1]} \right) = \ln \frac{n(t - t_w, t)}{N(t - t_w, t)}, \quad n(t - t_w, t) \tag{29}
\]

The maximum absolute of the privacy loss parameter \( \mathcal{L}_M^{(B)} \) in (29) is \( \xi \). Therefore, if \( \xi \) is not too large, then the returned values of \( s[r] \) and \( g[r] \) randomized by the algorithm \( \mathcal{M} \) will fall in any subset \( \mathcal{B} \) with the similar probabilities. In that case, it becomes expectedly difficult to identify that each returned value is a randomized query of \( s[r] \) or \( g[r] \) and hence the privacy is protected.

By using the same assumption as (19) and (20), the actual energy wasted in any time period, \( [t_{first}, t_{last}] \) where \( t_{last} > t_{first} \), can be calculated by

\[
\mathcal{E}_w = \sum_{t=t_{first}}^{t_{last}-1} \frac{\mathcal{P}_w[t]}{60} \tag{30}
\]

and the energy wasted in that time period based on the random motion states is defined in Algorithm 2 and can be calculated by

\[
\mathcal{E}_{w,alg} = \sum_{t=t_{first}}^{t_{last}-1} \frac{\mathcal{P}_{w,alg}[t]}{60} \tag{31}
\]

The actual wasted power consumption can also be defined by using Algorithm 2 and changing \( v \) to \( r \), \( \mathcal{A}(x_t) \) to \( f(x_t) \), and \( \mathcal{P}_{w,alg} \) to \( \mathcal{P}_w \).

**Algorithm 2 Wasted Energy Based on Random Motion States**

For any time point \( t \):

Let \( t_1 \) be the last time point with state 1

/* Assume no energy wasted at first */

\( \mathcal{P}_{w,alg}[t] \leftarrow 0 \)

if \( (v[t] = 0 \text{ and } v[t - 1] = 1) \) then

/* Assign new \( t_1 */

\( t_1 \leftarrow t - 1 \)

end if

if \( \text{for all states have been } 0 \text{ since time point } t_1 + 1 \) and

\( t - t_1 \geq 15 \text{ and } \sum_{k=t_1+1}^{t} \mathcal{A}(a_k) \geq 0.015(t - t_1) \) then

/* \( t - t_1 \geq 15 \) refers to the time constraint of wasted energy; nobody was in the queried area for at least 15 minutes. */

/* 0.015(t - t_1) refers to the average power constraint (in kW unit) of wasted energy; the average power usage of the queried area in that duration is at least 0.015 kW. */

for \( (t_1 + 1 \leq k \leq t) \) do

\( \mathcal{P}_{w,alg}[k] \leftarrow \mathcal{A}(a_k) \) (see Table 2)

end for

end if
VI. RESULTS AND DISCUSSIONS

From [3], consider an example office area with approximately $4 \times 15 \text{ m}^2$ equipped with 2 air-conditioners (with power ratings of 4.4 and 3.5 kW) and 3 light zones (with power ratings of 0.7, 0.3, and 0.4 kW), as shown in Fig. 3. The sum of the power ratings of all electric appliances in the room is 9.3 kW. For the actual power consumption and motion sensor data of the 30-day sample period as shown in Figs. 4 and 5, the actual wasted energy data are plotted in Fig. 6. There were 155 minutes with the overload shown in Fig. 4. These overloads could occur upon the transient surge of power required at each initial starting time of air-conditioner compressors and electric appliance deterioration caused by a lack of maintenance or usual hardware degradation over the equipment lifetime.

FIGURE 3. Schematic of office area used in experimental case study.

FIGURE 4. Actual energy consumption data with a total of 1564.019 in the 30-day sample period.

FIGURE 5. Actual motion sensor data of the 30-day sample period.

Consider the 4th day of the 30-day sample period as an example for further discussion in Sections VI-B and VI-C. The actual power consumption, motion state and wasted energy data of the office on the considered day are shown in Figs. 7, 8, and 9, respectively. The electrical usage period (office hours) is from 08:00 to 20:00. The wasted energy occurred from 16:16 to 16:38.

A. TRADITIONAL LAPLACIAN MECHANISM’S PRIVACY AND USABILITY VALIDATION

Let $U \in \mathbb{R}$ (set of real numbers) be a random variable by adding Laplacian noise to the actual processed value of an electrical use vector. The probability density function (pdf)
The parameter $\mu_t$ represents the actual power usage and $\kappa \in \mathbb{R}^+$ is the distribution parameter.

The privacy loss parameter of the traditional Laplacian mechanism is [14]:

$$\epsilon = \frac{\Delta f}{\kappa}$$

(33)

and the appropriate $\epsilon$ proposed by [32] is

$$\epsilon = \frac{\Delta h}{\Delta v} \ln \left( \frac{(m-1)\rho}{1-\rho} \right)$$

(34)

To approximately calculate the privacy loss parameter in the case of the office area investigated in this research, let

- $x_{1,t}$ be a status of 4.4 kW air-conditioner at time $t$
- $x_{2,t}$ be a status of 3.5 kW air-conditioner at time $t$
- $x_{3,t}$ be a status of 0.7 kW light zone at time $t$
- $x_{4,t}$ be a status of 0.3 kW light zone at time $t$
- $x_{5,t}$ be a status of 0.4 kW light zone at time $t$
- $y_t$ be an electrical use vector at time $t$

such that $||x_t - y_t||_1 = 1$

Assume the maximum probability that an electric appliance will be identified as being used at time $t$ (being identified as present in the electrical use vector) is $\rho = 1/3$ according to [32]. From (33), the parameter $\Delta f \approx 4.4$ since the maximum power usage difference between electrical use vectors $x_t$ and $y_t$ occurs when the 4.4 kW air-conditioner usage statuses of these vectors are different, which means $x_{1,t} = 1$ and $y_{1,t} = 0$, while other appliance usage statuses of both vectors are the same. The parameter $\Delta h \approx 4.4$ for the same reason. The parameter $\Delta v \approx 4.1$ since the different power usage between the appliance with the highest wattage (4.4 kW) and the appliance with the lowest wattage (0.3 kW) is approximately 4.1 kW. Thus, the privacy loss boundary of this case is:

$$\epsilon = \frac{4.4}{4.1} \ln \left( \frac{(5 - 1)(1/3)}{1 - 1/3} \right) = 0.74387$$

(35)

By applying the traditional Laplacian noise addition to the original power usage data of the 30-day sample period, the result of applying that noise with appropriate privacy loss calculated by using (34) is shown in Fig. 10. The result of applying the low privacy loss ($\epsilon = 0.01$) traditional Laplacian noise to the original power usage data of the 30-day sample period is shown in Fig. 11. The parameter $\kappa = \Delta f / \epsilon = 440$ in this case. Thus, the distribution variance of this case is $Var_t = 2\kappa^2 = 387200$. The result of applying the high privacy loss ($\epsilon = 7$) traditional Laplacian noise to the original power usage data of the 30-day sample period is shown in Fig. 12. The parameter $\kappa = \Delta f / \epsilon = 0.62857$ in this case. Thus, the distribution variance of this case is $Var_t = 2\kappa^2 = 0.79020$.

In terms of the noise addition mechanism usability, the percentage error of energy consumption can be calculated by

$$Err_{E} = \left| \frac{E_{c,mec} - E_{c}}{E_{c}} \right| \times 100\%$$

(36)

where $E_{c}$ is defined in (19) and $E_{c,mec}$ is energy consumption after applying a noise-addition mechanism.

The percentage error values of energy consumption in the 30-day sample period after applying the privacy-usability balanced, low privacy loss, and high privacy loss cases of traditional Laplacian noise addition mechanisms are 1.156%, 86.367%, and 0.00147%, respectively. The percentage error of the energy consumption calculation after applying the privacy-usability balanced case of traditional Laplacian noise addition is acceptable (lower than 2.5%) for billing purposes while the actual energy consumption data are hidden. Thus, the privacy of electricity users in this queried area is protected. However, the usability of overload detection remains
The queried area is rarely protected. Although electricity users’ privacy in the queried area is protected after applying the low privacy loss case of traditional Laplacian noise addition, the percentage error of energy consumption calculation is not acceptable (higher than 2.5%) for billing purposes. This high percentage error is due to a low variance of random values. However, the noise addition mechanism with privacy-usability balanced case are $N_{fp} = 7418$ and $N_{fn} = 59$. There are quite a few false-detection values compared to the other cases since the probability density of the distribution is greater than 0 in a narrow range caused by a low $\kappa$ value (refer to (32)). The overload detection correctness of the case with low privacy loss is 0.01425. For the last main case, the high privacy loss traditional Laplacian, the number of false-positives and false-negatives in overload detection are $N_{fp} = 457$ and $N_{fn} = 59$. There are quite a few false-detection values compared to the other cases since the probability density of the distribution is greater than 0 in a narrow range caused by a low $\kappa$ value (refer to (32)). The overload detection correctness of the case with high privacy loss is 0.37530. Due to the overload detection capability of the traditional Laplacian noise addition, this mechanism cannot be employed for the constraint of overload detection in this research. This mechanism will not be mentioned further in this paper.

### B. TRUNCATED LAPLACIAN MECHANISM’S PRIVACY AND USABILITY VALIDATION

For electrical use vectors, by trying out $b$ value in the range of [0.001, 10] using a step size of 0.01, the maximum $|L_A(\vec{S})|$ of all possible $(x_i, y_i)$ pairs, where $||x_i - y_i||_1 = 1$ (refer to (18)), is the lowest at $b = 2.458$, as shown in Fig. 13. At this $b$ value, the $(x_i, y_i)$ pair that the power consumption values of the vectors are $f(x_i) = \mu_{x,t} = 9.0$ and $f(y_i) = \mu_{y,t} = 4.6$, gives the highest $|L_A(\vec{S})|$ since the electrical use vector $[1 1 1 1 1]$ usually incurs an overload. All the overloads are shown without any noise-addition. As in this research case, the ratio between $h_y(u)$ and $h_x(u)$ in Fig. 14 is the lowest at $u = \mu_{x,t}$, and its absolute natural logarithm value, which is the privacy loss parameter, is greater than that of the highest ratio between $h_y(u)$ and $h_x(u)$. Thus, the privacy loss boundary is calculated as the absolute natural logarithm of the ratio between the probability density values in the purple circles. If the electrical use vector $[1 1 1 1 1]$ does not usually incur an overload, the parameter $P_f$ can be set slightly greater than the sum of the power ratings of all electric appliances in the queried area. In practice, the range of $\varepsilon$ is [0.01, 7] [29]. Therefore, by numerical experiments, $b$ value lower than 0.834 cannot be applied in this case. The parameter $b = 2.458$ is chosen since its value minimizes $\epsilon$. In addition, the probability density distribution with that $b$ value incurs a low percentage error of energy consumption calculation in (36), while the regularly distributed probability
density caused by a high $b$ value tends to produce a higher percentage error of that calculation because of the higher variance of random values.

From Fig. 13, the best selection of the $b$ value gives the minimum $\varepsilon$ at 4.30036, so the privacy loss parameter boundary in this case is:

$$\left| L_{A}^{(S)} \right| \leq \varepsilon = 4.30036$$  \hspace{1cm} (38)

The results of adding the proposed truncated Laplacian noise to the original energy consumption data of the 30-day sample period (Fig. 4) and the considered day (Fig. 7) using (15) are shown in Figs. 15 and 16, respectively. From (36), the percentage error values of energy consumption in the 30-day sample period and the considered day after applying the noise addition mechanism are 0.054% and 1.957%, respectively. These energy consumption values after applying the noise addition mechanism are similar to the actual energy consumption values because the expected value of the noise-added power consumption at any time point is equal to the value of the actual power being used at each time point.

The overload detection correctness of the mechanism is 1 because the randomization has been designed by constraining on $N_{fp} = N_{fn} = 0$. Moreover, this mechanism can be employed for purposes that are based on time-aggregated values of energy consumption.

The probability that actual data will be leaked [29], $\delta$, (refer to Table 1) of the modified mechanism in this research is $155/43200 = 0.00359$ because 155 overloads are detected (which must be shown to inform electrical consumers) based on the original data of 43200 power usage values. This obtainable level of $\delta$ is acceptable in practice [14]. However, the limitation of applying this noise-addition mechanism is the uncontrollable $\delta$. In the extreme case of all-time overload, the actual power consumption data cannot be hidden using the proposed mechanism. Nevertheless, such extreme cases are rare for, once they occur, the responsible engineers must be urgently notified to provide overload investigation and correction.
C. RANDOM MOTION STATES ALGORITHM’S PRIVACY AND USABILITY VALIDATION

Similar to Section VI-B, considering the 4th day of the 30-day sample period, the parameter $\xi$ (refer to Table 2) of Algorithm 1 depends on $N(t - t_s - t_w, t - t_s)$ and $n(t - t_s - t_w, t - t_s)$ in (29), as shown in Fig. 17. The measured wasted energy and error of that measurement (calculated by using (39)) on the considered day based on the noise-added energy consumption data shown in Fig. 16 and different motion state vectors, with varying time shifting values, are shown in Figs. 18 and 19, respectively.

An example of motion states randomized by using Algorithm 1 with $t_s = 60$ is shown in Fig. 20. These motion states obviously differ from the actual states in Fig. 8, so the actual movement of electricity users in the queried area is hidden. The energy wasted on the considered day based on the noise-added energy consumption data in Fig. 16 and the motion states in Fig. 20 is shown in Fig. 21.

The result of wasted energy, based on the noise-added energy consumption of the 30-day sample period in Fig. 15, using the motion states randomized by Algorithm 1 is shown in Fig. 22.

The motion states randomized by using the coin-flipping algorithm in [14] are shown in Fig. 23. These motion states obviously differ from the actual states in Fig. 8, so the actual movement of electricity users in the queried area is hidden. However, the accuracy of the wasted energy calculation is very low. The privacy loss parameter of this algorithm is $\ln(3)$. By using the noise-added energy consumption result shown in Fig. 16, wasted energy based on the randomized motion states in Fig. 23 is shown in Fig. 24.

The result of wasted energy, based on the noise-added energy consumption of the 30-day sample period in Fig. 15, using the motion states randomized by the coin-flipping algorithm in [14] is shown in Fig. 25.

In terms of privacy, if the actual energy consumption data are revealed, people who receive the data can only guess about an individual movement. However, assume that actual motion sensor data are revealed. In that case, people who receive these data know the actual individual movement, which means that the privacy of people using that queried area is jeopardized. Therefore, the random algorithm of motion sensor data is more important than the noise addition mechanism of energy consumption data. The privacy loss parameter boundary of the proposed algorithm shown in Fig. 17 is $1.12500 > \xi > 1.10000 > \ln(3) = 1.09861$, which means that the privacy loss of the proposed algorithm is...
FIGURE 21. Wasted energy data, based on the noise-added energy consumption data in Fig. 16 and motion states in Fig. 20, with a total of 2.188 kWh on the considered day.

FIGURE 22. Wasted energy data, based on the noise-added energy consumption data in Fig. 15 and motion states randomized by using Algorithm 1 with time shifting value $t_s = 60$ minutes, with a total of 190.398 kWh in the 30-day sample period.

FIGURE 23. Randomized motion states on the considered day using the coin-flipping algorithm in [14].

slightly greater than the privacy loss of the coin-flipping algorithm in [14]. This algorithm also shifts the wasted energy occurrence time because the occurrence of wasted energy means that nobody is in the queried area in that duration. Since the privacy loss parameter boundary of the modified truncated Laplacian noise addition mechanism from (38) is $\varepsilon = 4.30036$, the privacy loss parameters of the proposed processes are acceptable [29].

In terms of the usability of the random motion state algorithm, the percentage error of wasted energy can be calculated by

$$Err_{wasted} = \left| \frac{E_{w,alg} - E_w}{E_w} \right| \times 100\%$$  \hspace{1cm} (39)

where $E_w$ and $E_{w,alg}$ are defined in (30) and (31), respectively.

From (39), the percentage error values of wasted energy on the considered day using Algorithm 1 and the coin-flipping algorithm in [14] are 4.042% and 100%, respectively. The percentage error values of wasted energy in the 30-day sample period using Algorithm 1 and the coin-flipping algorithm in [14] are 3.962% and 97.988%, respectively. Thus, the proposed conditional randomization is better than the standard coin-flipping algorithm in [14].

Since wasted energy usually occurs in the middle of the office hour with similar power consumption for hours, by shifting the wasted energy occurrence time to
approximately 60 time points (an hour), the wasted energy calculation using Algorithm 2 is still similar to the actual value as shown in Figs. 21 and 22. In addition, by using Algorithm 1, the number of time points with wasted energy in the office hour duration should be similar to the actual value since the motion states are conditionally randomized with an appropriate time-shifting value. This value depends on the presence or absence of electricity users who have been to the queried area in the considered duration (e.g., 1 day and 1 month), which means that the power consumption defined as wasted energy in both durations is similar. Moreover, the probability of at least 15 consecutive 0 states randomized by the random algorithm. For any time duration out of the office hour, the actual motion states are 0, so at least 15 consecutive 0 states may occur. However, the actual power usage is nearly 0 kW. Applying the modified truncated Laplacian noise to the original power consumption data has been found to incur an insignificant amount of wasted energy in that duration. Hence, this finding suggests the practicality of the proposed algorithm. Nevertheless, the actual motion sensor data cannot be hidden using the proposed algorithm in the extreme case of all-time energy wasted due to the condition to show shifted occurrence time of wasted energy. This condition is the limitation of the algorithm.

VII. CONCLUSION

Based on the differential privacy framework, this research has investigated comparatively various noise-addition mechanisms to apply for the BEMS dataset. Although the traditional Laplacian noise-addition mechanism with an appropriate privacy loss parameter can hide the actual power consumption data, it fails to detect overloads. The distribution range \((-\infty, \infty)\) or the distribution symmetry of traditional Laplacian noise give false positives and negatives in overload detection. Additionally, an excessive low privacy loss case incurs a high percentage error of the energy consumption calculation, while an excessive high privacy loss case can hardly hide the data.

Due to these limitations of existing traditional Laplacian noises, this research has proposed a modified truncated Laplacian noise addition mechanism that can hide the collected actual time-series of electrical appliance usage patterns effectively while maintaining the accuracy of power usage overload detection. Based on the reported 30-day actual BEMS dataset, this mechanism’s privacy loss parameter boundary is \(4.30036\). The percentage error of the noise-added energy consumption in the 30-day sample period is lower than 2.5% for the standard billing constraint. The \(F_1\)-score overload detection correctness of the proposed truncated Laplacian noise addition is 1, i.e., no overload detection errors at all by design. These results confirm that the mechanism is practicable for billing computation, overload detection, and other purposes based on time-aggregated energy consumption values. Moreover, a random motion state algorithm is proposed to hide the actual movement in the queried area. This algorithm’s privacy loss parameter boundary is in the range of \([1.1000, 1.12500]\), which varies with regard to tunable time-shifting values, 15 to 150 minutes in this research. With an optimal time-shifting value, the percentage error of the wasted energy calculation is lower than 5%. In future work, this time-shifting value can periodically (e.g., daily and weekly) be randomized within the optimal range based on the electrical use pattern to increase individual privacy. Additionally, based on the potential indicated by the results in this research, these data perturbation processes can be further extended to operate in real-time at the BEMS data gateways.

REFERENCES

[1] T. Inthasuth and C. Aswakul, “Development and reliability testing of IEEE1888 gateway for ZigBee wireless sensor network in Chulalongkorn University’s building energy management system,” in Proc. 5th Int. Collaboration Symp. Inf., Prod. Syst., Nov. 2014, pp. 1–5.
[2] D. H. Le and W. Pora, “Development of smart meter for building energy management system based on the IEEE 1888 standard with Wi-Fi communication,” in Proc. Int. Conf. Electron., Inf. Commun. (ICEIC), Jan. 2014, pp. 1–2.
[3] M. Pipattanasomporn, G. Chittalai, J. Songsiri, C. Aswakul, W. Pora, S. Suwankanlin, K. Audomvongserer, and N. Hooncharoem, “CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets,” Scientific Data, vol. 7, no. 1, pp. 241–254, Jul. 2020, doi: 10.1038/s41597-020-00582-3.
[4] R. Thirasa, M. Homchan, S. Kwankajornkeat, and C. Aswakul, “Development of IoTcloudServer@TEIN smart-energy@Chula service gateway: Case study of secured on-demand building energy management system data platform using NETPIE,” in Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia), Jun. 2019, pp. 96–99.
[5] P. Khawsa-Ard and C. Aswakul, “Application of simple computer board game with gesture sensor input for increasing awareness in electrical energy consumption,” in Proc. 29th Int. Tech. Conf. Circuit/Syst. Comput. Commun., Jul. 2014, pp. 1–3.
[6] P. Khawsa-Ard and C. Aswakul, “IEEE1888 interactive display as a service (IDaaS): Example in building energy management system,” in Proc. IEEE 39th Annu. Comput. Softw. Appl. Conf., vol. 3, Jul. 2015, pp. 517–522.
[7] K. Chooputtipong and C. Aswakul, “Development of data analytic program for building energy management system with wasted energy analysis using motion sensor,” in Proc. 39th Elect. Eng. Conf. (ECON), Chonburi, Thailand, Nov. 2016, pp. 1–4.
[8] J. Liu, C. Zhang, and Y. Fang, “EPIC: A differential privacy framework to defend smart Homes against internet traffic analysis,” IEEE Internet Things J., vol. 5, no. 2, pp. 1206–1217, Apr. 2018.
[9] M. U. Hassan, M. H. Rehmani, and J. Chen, “Differential privacy techniques for cyber physical systems: A survey,” IEEE Commun. Surveys Tuts., vol. 22, no. 1, pp. 746–789, 1st Quart., 2020.
[10] T. Wang, Z. Zheng, M. H. Rehmani, S. Yao, and Z. Hao, “Privacy preservation in big data from the communication perspective—A survey,” IEEE Commun. Surveys Tuts., vol. 21, no. 1, pp. 753–778, 1st Quart., 2019.
[11] P. Barbosa, A. Brito, and H. Almeida, “A technique to provide differential privacy for appliance usage in smart metering,” Inf. Sci., vol. 370–371, pp. 355–367, Nov. 2016.
[12] A. Wood, M. Altman, A. Bembenek, M. Bun, M. Gaboardi, J. Honaker, K. Nissim, D. O’Brien, T. Steinke, and S. Vadhan, “Differential privacy: A primer for a non-technical audience,” SSRN Electron. J., vol. 21, no. 1, pp. 209–276, Jan. 2018.
[13] J. Li, J.-J. Yang, Y. Zhao, B. Liu, M. Zhou, J. Bi, and Q. Wang, “Enforcing differential privacy for shared collaborative filtering,” IEEE Access, vol. 5, pp. 35–49, 2017.
[14] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” Found. Trends Theor. Comput. Sci., vol. 9, nos. 3–4, pp. 211–407, 2014.

[15] X. Liao, P. Srinivasan, D. Formby, and R. A. Beyah, “Di-PriDA: Differentially private distributed load balancing control for the smart grid,” IEEE Trans. Dependable Secure Comput., vol. 16, no. 6, pp. 1026–1039, Nov. 2019.

[16] H. Wang, J. Zhang, C. Lu, and C. Wu, “Privacy preserving in non-intrusive load monitoring: A differential privacy perspective,” IEEE Trans. Smart Grid, vol. 12, no. 3, pp. 2529–2543, May 2021.

[17] J. Xiong, J. Ren, L. Chen, Z. Yao, M. Lin, D. Wu, and B. Niu, “Enhancing privacy and availability for data clustering in intelligent electrical service of IoT,” IEEE Internet Things J., vol. 6, no. 2, pp. 1530–1540, Apr. 2018.

[18] X.-Y. Zhang, S. Kuenzel, J.-R. Córdoba-Pachón, and C. Watkins, “Privacy-functionality trade-off: A privacy-preserving multi-channel smart metering system,” Energies, vol. 13, no. 12, p. 3221, Jun. 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/12/3221

[19] U. B. Baloglu and Y. Demir, “Lightweight privacy-preserving data aggregation scheme for smart grid metering infrastructure protection,” Int. J. Crit. Infrastruct. Protection, vol. 22, pp. 16–24, Sep. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874548217301105

[20] A. Gohar, F. Shafik, F. Duerr, K. Rothermel, and A. ElMougy, “Privacy-preservation mechanisms for smart energy metering devices based on differential privacy,” in Proc. IEEE Wireless Commun. Netw. Conf. Workshop (WCNCW), Apr. 2019, pp. 1–6.

[21] Z. Zhang, Z. Qin, L. Zhu, W. Jiang, C. Xu, and A. K. Ren, “Toward practical differential privacy in smart grid with capacity-limited rechargeable batteries,” 2015, arXiv:1507.03000.

[22] Z. Zhang, Z. Qin, L. Zhu, J. Weng, and K. Ren, “Cost-friendly differential privacy for smart meters: Exploiting the dual roles of the noise,” IEEE Trans. Smart Grid, vol. 8, no. 2, pp. 619–626, Mar. 2017.

[23] W. Lee Croft, J.-R. Sack, and W. Shi, “Differentiable privacy via a truncated and normalized Laplace mechanism,” 2019, arXiv:1911.06062.

[24] P. Barbosa, A. Brito, H. Almeida, and S. Clauß, “Lightweight privacy for smart metering data by adding noise,” in Proc. 29th Annu. ACM Symp. Appl. Comput., Mar. 2014, pp. 531–538.

[25] Q. Geng, W. Ding, R. Guo, and S. Kumar, “Privacy and utility tradeoff in approximate differential privacy,” 2018, arXiv:1810.00877.

[26] N. Holohan, S. Antonatos, S. Braghin, and P. Mac Aonghusa, “The bounded Laplace mechanism in differential privacy,” 2018, arXiv:1808.10410.

[27] M. Bun, C. Dwork, G. N. Rothblum, and T. Steinke, “Composable and versatile privacy via truncated CDP,” in Proc. 50th Annu. ACM SIGACT Symp. Theory Comput., Jun. 2018, pp. 74–86.

[28] X. Fang, F. Yu, G. Yang, and Y. Qu, “Regression analysis with differential privacy preserving,” IEEE Access, vol. 7, pp. 129353–129361, 2019.

[29] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayanan, B. C. Pierce, and A. Roth, “Differential privacy: An economic method for choosing epsilon,” in Proc. IEEE 27th Comput. Secur. Found. Symp., Jul. 2014, pp. 398–410.

[30] N. Fernandes, A. McIver, and C. Morgan, “The Laplace mechanism has optimal utility for differential privacy over continuous queries,” 2021, arXiv:2105.07176.

[31] J. Soria-Comas and J. Domingo-Ferrer, “Optimal data-independent noise for differential privacy,” Inf. Sci., vol. 250, pp. 200–214, Nov. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025513005094

[32] J. Lee and C. Clifton, “How much is enough? Choosing ε for differential privacy,” in Information Security, vol. 7001. Berlin, Germany: Springer, 2011, pp. 325–340.