Differentially Private Mobile Crowd Sensing Considering Sensing Errors

sensors

MDPI

Article

Differentially Private Mobile Crowd Sensing Considering Sensing Errors

Yuichi Sei [1,2,*] and Akihiko Ohsuga [1]

[1] Department of Informatics, Graduate School of Informatics and Engineering, University of Electro-Communications, Chofu, Tokyo 182-8585, Japan; ohsuga@uec.ac.jp
[2] JST, PRESTO, Kawaguchi, Saitama 332-0012, Japan
* Correspondence: seiuny@uec.ac.jp

Received: 7 April 2020; Accepted: 11 May 2020; Published: 14 May 2020

Abstract: An increasingly popular class of software known as participatory sensing, or mobile crowdsensing, is a means of collecting people's surrounding information via mobile sensing devices. To avoid potential undesired side effects of this data analysis method, such as privacy violations, considerable research has been conducted over the last decade to develop participatory sensing that looks to preserve privacy while analyzing participants' surrounding information. To protect privacy, each participant perturbs the sensed data in his or her device, then the perturbed data is reported to the data collector. The data collector estimates the true data distribution from the reported data. As long as the data contains no sensing errors, current methods can accurately evaluate the data distribution. However, there has so far been little analysis of data that contains sensing errors. A more precise analysis that maintains privacy levels can only be achieved when a variety of sensing errors are considered.

Keywords: crowdsensing; differential privacy; data mining; sensing errors

1. Introduction

Today's smartphones are powerful minicomputers that contain an impressive array of sensing components such as cameras or accelerometers, with the ability to collect and analyze users' surrounding information [1] (Figure 1). Extensive research shows that as well as through mobile phones, data is collected through different means of transportation, such as trains, cars or bicycles. Such information collection is referred to as participatory sensing or mobile crowdsensing. Many studies have been conducted on participatory sensing. For example, Bridgelall et al. proposed a system that detects anomaly locations of roadways using participatory vehicle sensors [2]. Kozu et al. developed a hazard map of bicycle accidents based on data from accelerometers of participatory smartphones [3].

Although high participation is necessary for participatory sensing to be successful, participants may be discouraged by privacy concerns or having to use extra battery power. As such, it is necessary to develop a participatory sensing method featuring both low battery power requirements and high privacy protection [4].

Several frameworks use geotagged posts of Twitter and/or Instagram [5,6]. Although Twitter and Instagram users disclose their locations intentionally, a privacy mechanism could motivate the users to share more geotagged posts.

Several privacy-preserving techniques have been proposed for participatory sensing, such as in References [7,8]. By perturbing data based on ϵ-differential privacy [9,10] privacy leakage can be controlled. Differential privacy has been used in many studies, such as References [11–13], as it is one of the strongest privacy metrics [14].

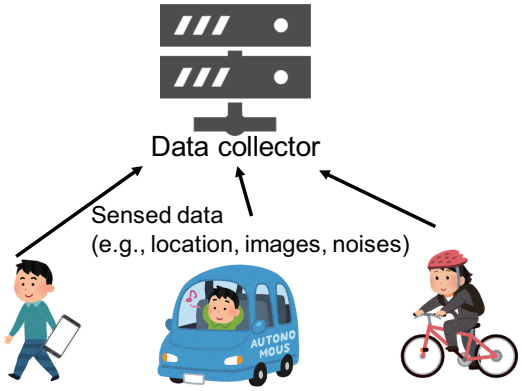


Figure 1. Participatory sensing.

It is problematic, however, that although most collected data contain sensing errors, these seem to have been overlooked in the majority of existing studies. Therefore, the methods used in existing studies reconstruct not the true values but the sensing values with sensing errors (see Table 1). As such, the accuracy of the analysis based on current methods is compromised.

Table 1. The difference between the existing methods and our method.

Methods	Output of the Method
Existing methods.	The estimated distribution of sensing data containing errors.
Our method.	The estimated distribution of true data without errors.

In this paper, we propose an architecture of privacy-preserving participatory sensing considering sensing errors. The proposed architecture consists of two parts. One is the anonymization technique on each participant's side (perturbing data with sensing errors [PDE]). Each device perturbs its sensed data and then reports the perturbed data to the data collector. Because perturbed data is reported to the data collector, the data collector cannot know the true data distribution. Therefore, the proposed architecture also provides an estimation technique, which estimates the true data distribution based on the reported data, on the data collector's side (estimating true distribution considering sensing errors [ETE]).

2. Models

We define our proposed model. This model is the same as that used in an existing study [8] except for sensing errors.

2.1. Application Model

Sensed data on participants' surrounding environment that features some sensing errors, such as their location or the radiation level, is collected on mobile phones and sent to the data collector. It is then assumed that the data collector's analysis results in an accurate data distribution (see Figure 2).

Many factors are worth considering when developing mobile crowdsensing applications, such as radiation levels, urban planning, class of vehicle (for example, whether it is a flatbed truck, taxi or ambulance), and anonymous driver monitoring, as well as more general information such as the participant's city of residence, surrounding noise levels or personal data such as age and gender [15,16].

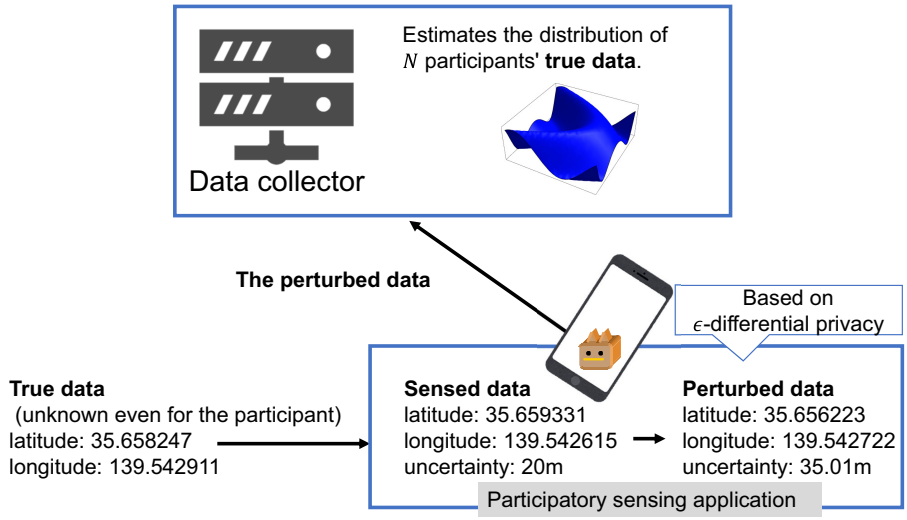


Figure 2. Overview of the proposed architecture.

There are several stages to the mobile crowdsensing application process. First, the crowdsensing application ID is determined by the data collector so that a selection of crowdsensing applications can be used simultaneously and still be easily distinguishable. Following from this, the data collector must source participants who own an electronic mobile device such as a GPS device or smartphone. Once a participant has volunteered to collaborate with the crowdsensing application, then PVE, the suggested anonymization algorithm, is applied. The final stage is for the data collector to analyze the mass of data using the ETE.

Because several studies suggest that measurement errors follow a normal distribution [17], it is used in this paper as an error model. Standard deviation is defined by the parameter σ, which typifies normal distribution. It is widely recognized that true sensing data falls within the normal distribution pattern [18–20]. Indeed, a study of 29,000 items of GPS data collected by Devon et al. [21] and real-time gesture recognition achieved by the pose tracking accuracy of the Microsoft Kinect 2 reported by Wang et al. [22] both follow normal distribution patterns.

It can thus be predicted that error probability also, for the most part, emulates a normal distribution pattern. The accuracy of a sensor is normally depicted on a data sheet shown by sensor vendors. For example, a standard deviation of a normal distribution is shown on the data sheet. If an average error is shown, we can obtain the standard deviation of the normal distribution.

Jiang et al. proposed a fault diagnosis system that took into account a measurement error problem [23]. They assumed that the measurement error usually follows a normal distribution. Wang et al. proposed a measurement system for the rotational angle of the wheel [24]. They considered sensing error analysis to be a very important problem. Location errors of an accelerometer were set to follow normal distributions in their experiments.

MPU-6000 IMU is a low-cost navigation system for ground vehicles. Gonzalez et al. [25] collected real sensing data from MPU-6000 IMU and concluded that sensing errors of accelerometers of ZMPU-6000 follow normal distributions, and sensing errors of gyroscopes of ZMPU-6000 can be modeled as pseudo normal distribution processes, although the errors do not follow a perfect, normal distribution. They also collected real sensing data of Ekinox IMU. They showed that sensing errors of accelerometers and gyroscopes followed normal distributions.

Nguyen et al. discussed how sensing location errors affects mobile, robotic, wireless sensor networks [26]. The sensing location errors were modeled to follow normal distributions in their proposed algorithm. They showed that their algorithm realized a high performance using real data sets containing sensing errors.

Similarly, a machine learning technique featuring deep neural networks has been adopted by sensing systems. Several studies based on deep neural networks reported that prediction errors

followed a normal distribution [27–29]. If several training samples can be amassed, a data collector can analyze the standard deviation of the error distribution.

Although not all sensing errors follow normal distributions, many sensing errors are considered to follow normal distributions, as described above. Our proposed method targets the situation in which sensing errors can be considered to follow normal distributions.

2.2. Motivating Example

Assume that the data collector wants to analyze the noise level in each location to tackle a plane noise problem. To increase the number of participants, the data collector wants to mitigate the privacy issues of the participants. In this case, each participant can perturb her/his location information, and then each participant reports the perturbed location information. Because the reported location information is perturbed by each participant, the data collector should reconstruct the true information. However, because existing studies did not consider the sensing errors of the true location information, the accuracy of the reconstructed location information with the data collector will decrease.

In this paper, we aim to increase the accuracy of the reconstructed information with the data collector by modeling the sensing errors at the participants' side.

2.3. Privacy Metric

Extensive research completed within publications in the data-mining field [30–32] reveals that differential privacy [33] is among the most powerful privacy measures available. The following context can be considered: an honest data holder with a database containing participants' true information is paired with a malicious data analyst desiring access to that database. Whenever the database is accessed by the analyst, noise is added to the query response based on a privacy mechanism $\mathcal{A}$. The differential privacy can be understood in the following manner, with ϵ as a positive real number:

Definition 1 (ϵ-differential privacy)**.** *Databases D and D' are neighboring databases, if they differ only in at most one record. A privacy mechanism $\mathcal{A}$ satisfies ϵ-differential privacy if and only if for any output Y, the following equation holds for all databases of D and D':*

$$P(\mathcal{A}(D) \in Y) \leq e^{\epsilon} P(\mathcal{A}(D') \in Y),$$

where $Y \subseteq Range(\mathcal{A})$.

This method can be used for privacy-preserving participatory sensing [34]:

Definition 2 (local privacy)**.** *Databases x and x' are neighboring databases with size = 1. A privacy mechanism $\mathcal{A}$ satisfies ϵ-differential privacy if and only if for any output y, the following equation holds for all databases of x and x':*

$$P(\mathcal{A}(x) = y) \leq e^{\epsilon} P(\mathcal{A}(x') = y). \quad (1)$$

2.4. Utility Metric

Data analysis can be achieved through the data collector producing data distribution, which is expressed by a (multidimensional) histogram or a cross-tabulation. To measure the difference between the original data generated distribution, information known to neither the participant nor the data collector, and the reported data generated distribution analyzed by the data collector, the utility metric Mean Squared Error (MSE) is employed.

Let N denote the number of participants, and let $H_1, H_2, \ldots, H_{b_n}$ denote each bin of a histogram of sensed data. Here, b_n represents the number of bins. Let $\S_j$ denote the number of participants whose true data is categorized to H_j, and let $\ddagger_i$ denote the number of participants categorized to H_i in the estimated histogram at the data collector.

Definition 3 (MSE). *We use the MSE between* $\ddagger_j$ *and* $\S_j$ *to quantify the utility for the estimated histogram:*

$$MSE = \frac{1}{b_n} \sum_{j=1}^{b_n} (\S_j - \ddagger_j)^2. \tag{2}$$

2.5. Problem

The objective is to ensure the ϵ-differential privacy is achieved, each sensed value is anonymized, and a (multidimensional) histogram is created, while the MSE remains minimized to retain superior quality. This is outlined below:

Problem 1. *Given a set of participants* U *(the size of* U *is* N*), their sensed data* x_i $(i = 1, \ldots, N)$*, and a privacy parameter* ϵ*, find anonymized data* y_i *satisfying* ϵ*-differential privacy for all* i*. Moreover, given the anonymized data* y_i*, find estimated data distribution* $\ddagger_i$ $(i = 1, \ldots, b_n)$ *that minimizes the MSE.*

3. Related Work

3.1. Privacy-Preserving Mobile Crowdsensing

Several privacy-preserving systems including [35–37] that are based on encryption, known as encryption schemes, can be established for this context. They assume that the data collector might be a malicious entity but that the participant fraction conspiring together with the data collector, can be no higher than the predefined value γ. Honest participants' private data could be leaked if the data collector connives with over γ% of participants. It must also be highlighted that, as demonstrated in Section 2.1, it is quite simple for data collectors to create smartphone emulators to freely connive within mobile crowdsensing situations.

One increasingly trusted system that safeguards participants' data regardless of whether data collector and $N - 1$ of N participants are conspiring together is randomized response [38]. Here, a sensed value represents a predefined category that is then substituted with a certain probability category before the data collector receives it. In this way, participants' privacy is to some extent ensured as the true data with probability p and the perturbed data with probability $1 - p$ are sent to the server. Although the data collection server cannot obtain reliable information about each participant's data, by collecting information from many participants and conducting a statistical analysis, it is possible for the data collection server to estimate the true data distribution with some degree of accuracy.

Several methods that extend randomized response have been proposed, such as [7,8,39]. In a method called S2Mb (single to randomized multiple dummies with bayes) [8], each participant selects and reports several category IDs to the data collector. By adjusting the probability of selecting the original category ID and the number of selected category IDs, S2Mb can achieve higher accuracy while maintaining the privacy protection level. S2Mb outperformed other privacy-preserving methods [8].

Task allocation is one of the main issues of mobile crowdsensing. Yang et al. proposed a privacy-preserving framework that can allocate tasks to each participant [40]. They assume that the data collector is a trusted entity, and the data collector has a signed agreement with participants. On the contrary, the data collector in our proposed method does not need such an agreement.

The protecting location privacy (PLP) framework was proposed by Ma et al. [41]. Each participant specifies her/his privacy location in advance, and the participant sends all sensing data with location information as they are, except for the data sensed in the specified privacy location. Because several data are sent to the data collector without any modification, PLP does not satisfy differential privacy. PLP uses another privacy metric named δ-privacy, and satisfying differential privacy is out of the scope of the PLP framework.

In mobile crowdsensing, there is a tradeoff between participants' privacy and data utility. Gao et al. proposed a game model that addressed this contradictory issue [42]. Their method helps to determine the value of the privacy budget ϵ of differential privacy. As noted in their paper, their method does not

care about how to add noises to the sensing data or how to conduct statistical analysis. Our proposed perturbing data with sensing errors (PDE) and estimating true distribution considering sensing errors (ETE) can be used for adding noise and conducting statistical analysis, respectively.

Huai et al. proposed a privacy-preserving aggregation framework [43]. Their method can realize high data utility while preserving privacy. They assume that the data collector might not be a trusted entity. However, if many participants collude with the data collector, an honest participant's privacy will be leaked to the data collector. Because it is difficult for the honest participant to know how many other honest participants there are, the honest participant still might have privacy concerns.

Huang et al. proposed a privacy-preserving incentive mechanism for mobile crowdsensing [44]. Their target application is a noise monitoring system that collects noise levels and corresponding location information. Because the noise level is related to the location, an attacker could estimate each participant's location using a location inference attack. Although they did not consider sensing errors, Huang et al.'s proposed mechanism satisfies differential privacy and prevents location inference attacks.

Nonetheless, sensing errors are not taken into account in the methods outlined.

3.2. Privacy Metrics

There are many privacy metrics other than differential privacy. For example, k-anonymity was originally proposed as a privacy model when publishing medical data [45], and it is used today in many studies [46,47]. k-anonymity ensures that there are k or more records that have the same quasi-identifier values so that k-anonymity can protect against "identity disclosure". For example, a method wherein a database that originally recorded ages in 1-year increments is abstracted to 30 s, 40 s, and so forth. Even in the event an attacker knows all the quasi-identifier information about a given user, because there are k or more records corresponding to that user, they cannot tell beyond a $1/k$ level of confidence which record belongs to the corresponding user. There are also k-anonymity related privacy metrics such as l-diversity [48] and t-closeness [49]. These privacy metrics are also important; however, applying our proposed model to privacy metrics other than differential privacy is out of the scope of this paper and considered for future work.

3.3. Incentive Mechanism and Trustworthiness for Mobile Crowdsensing

An incentive mechanism is a very important issue for mobile crowdsensing. If the incentive mechanism works well, it is expected that the crowdsensing system can gather many participants even if the privacy levels are relatively low. On the other hand, if there are no good incentive mechanisms, the privacy levels should be higher to recruit many participants.

Suliman et al. proposed an incentive-compatible mechanism for group recruitment [50]. They considered the greediness of participants of in-group recruitment, and the proposed mechanism can increase the quality of the collected information by selecting participants who are expected to give high-quality data at a low cost.

A reverse auction mechanism also can be used for recruiting participants. The participants bid their expected rewards, and the crowdsensing manager selects good participants. In general, the winning probability is not known to the participants. Modified reverse auction (MRA) mechanisms proposed by Saadatmand et al. provide the estimated winning probability to participants [51].

The participants can modify their bidding price to increase their probability of winning. Wu et al. proposed a modified Thompson sampling worker selection (MTS-WS) mechanism, which uses reinforcement learning to estimate each participant's data quality [52].

The prevention techniques against false data injection attacks are also important for the success of mobile crowdsensing. We can use these techniques, such as in References [53–55], to select reliable participants who contribute to maximizing the quality of mobile crowdsensing.

Zhang et al. proposed a privacy-preserving crowdsensing framework using an auction mechanism [56]. They assumed that the data collector is a trusted entity, and each participant sends her/his sensitive

data to the data collector as-is. Therefore, the privacy information of participants is known to the data collector. On the contrary, we assume that the data collection server might not be a trusted entity. Each participant's original data need not be sent to any other entities in our proposed method.

There are several important mobile crowdsensing survey articles. Capponi et al. analyzed mobile crowdsensing studies and outlined future research directions [57]. Liu et al. [58] focused on privacy and security, resource optimization, and incentive mechanisms. They argued that ensuring privacy and trustworthiness is important.

Pouryazdan et al. [59] proposed three new metrics to quantify the performance of mobile crowdsensing: platform utility, user utility, and false payments. Using these metrics, they showed that data trustworthiness and data utility could be improved by collaborative reputation scores, which are calculated based on statistical reputation scores and vote-based reputation scores.

Pouryazdan et al. [60] proposed a gamification incentive mechanism. They formulated a game theory approach and showed that their mechanism could improve data trustworthiness greatly. Moreover, the proposed mechanism could prevent the data collector from paying rewards to malicious participants.

Xiao et al. formulated the interactions between the data collector and the participants as a Stackelberg game [61]. Because the sensing accuracy determined the reward, each participant was motivated to sense highly accurate data. Deep Q-Network, a reinforcement learning algorithm with deep neural networks, was used to determine the optimal reward.

Privacy-preserving mechanisms, including our proposed method, could be combined with such incentive mechanisms to increase participants while maintaining a low cost.

Domínguez et al. [5] proposed a method that detects unusual events based on geolocated posts on Instagram. The framework uses DBSCAN, a density-based clustering algorithm that executes an outlier detection algorithm to detect unusual events. INRISCO, an incident detection platform for smart cities, was proposed by Igartua et al. [6]. INRISCO uses Twitter and Instagram posts along with the data of vehicular and mobile ad hoc networks. Although Twitter and Instagram users disclose their locations intentionally, privacy-preserving mechanisms and incentive mechanisms could motivate the users to share more geotagged posts. As a result, the ability to detect unusual events can be improved.

4. Method

4.1. Overview

We assume that sensing errors follow a probability distribution such as a normal distribution, as described in Section 2.1.

Here, there are two scenarios. In the first scenario, the standard deviation of the sensing error is not considered private information. Because the standard deviation itself does not have any sensitive meaning, this scenario is reasonable. In the second scenario, we consider the standard deviation of the sensing error to also be private information. For example, if the standard deviation is correlated with the sensing value, then the second scenario is preferred. Our proposed architecture can address both scenarios.

A differential private value can be obtained by adding Laplace noise to a target value [9]. Each participant adds a Laplace noise to the sensed value; then, the noised value is reported to the data collector. The data collector estimates the data distribution (see Figure 2) from all of the reported values. If only one person participates in the participatory sensing, then the data collector concludes that the reported value is most likely to be the real value. However, if there are many participants, the data collector can estimate a more accurate data distribution through the statistical analysis proposed in this paper.

Our main notations are summarized in Table 2.

Table 2. Notations.

N	Number of participants.
x_i	True sensing value of participant i.
y_i	Reported sensing value of participant i.
X	$\{x_1, \ldots, x_N\}$.
Y	$\{y_1, \ldots, y_N\}$.
Y_σ	Set of standard deviations of the normal distributions of sensing errors of all participants.
b_n	Number of bins of a histogram.
$maxv_{org}$	Maximum value of a sensing data.
$minv_{org}$	Minimum value of a sensing data.
$maxv_{rep}$	Maximum value of a reported data.
$minv_{rep}$	Minimum value of a reported data.
$max\sigma_{org}$	Maximum value of a standard deviation.
$min\sigma_{org}$	Minimum value of a standard deviation.
b_v	Scale factor of a Laplace noise with regard to the sensing value.
b_σ	Scale factor of a Laplace noise with regard to the standard deviation.
$\dagger_i$	Number of participants whose reported values were categorized into the ith bin.
$\ddagger_i$	Estimated number of participants whose true values were categorized into the ith bin.
$\dagger$	$\{\dagger_1, \ldots, \dagger_{b_n}\}$.
$\ddagger$	$\{\ddagger_1, \ldots, \ddagger_{b_n}\}$.

4.2. PDE for Participants

In this section, we propose an anonymization technique at each participant's side: perturbing data with sensing errors (PDE).

The data collector determines the minimum and maximum values of the sensed data for which to use differential privacy. For example, the data collector can determine whether the participant's noise volume is from 0 to 120 dB. If the sensed value is out of this range, the value is considered to be 0 (if the sensed value is less than 0) or 120 (if the sensed value is greater than 120) on the participant's device. Let $minv_{org}$ and $maxv_{org}$ represent the minimum and maximum values of the sensed values.

The value range of perturbed data is infinity because a Laplace noise is added to the sensed data. To avoid decreasing the accuracy of an estimated histogram, the data collector also determines the minimum and maximum values of the reported data with which to create a histogram. Let $minv_{rep}$ and $maxv_{rep}$ represent these values.

If the data collector considers the standard deviation of the sensing error to also be private information, then the data collector will determine the minimum and maximum values of the standard deviation. Let $min\sigma_{org}$ and $max\sigma_{org}$ represent these values.

The Laplace mechanism [9] can be used, which adds noise based on the Laplace distribution. The theorem of the Laplace mechanism for data collection is introduced.

Theorem 1 (Laplace Mechanism)**.** *A privacy mechanism $\mathcal{A}$ realizes ϵ-differential privacy if $\mathcal{A}$ adds the Laplace noise $Lap(\Delta/\epsilon)$, where Δ is the range of the target attribute's possible values, and $Lap(b)$ returns independent Laplace random variables with the scale parameter b.*

If the standard deviation is considered private information, then a Laplace noise is added to the standard deviation as well as to the sensing data.

In the second scenario, in which the standard deviation σ of the sensing error is considered private information, a Laplace noise is added to not only the sensed value x but also the value of σ. If two elements are protected by ϵ-differential privacy, we should divide the privacy budget ϵ into two elements [34].

Algorithm 1 shows the PDE algorithm.

Algorithm 1 Anonymization Algorithm.

Input: $minv_{org}$, $maxv_{org}$, $minv_{rep}$, $maxv_{rep}$, $min\sigma_{org}$, $max\sigma_{org}$, ϵ.
Output: Report value v and standard deviation σ of sensing error

1: Obtain sensed value v and standard deviation σ of sensing error
2: **if** the standard deviation is considered as private information **then**
3: $\quad \epsilon \leftarrow \epsilon/2$
4: **end if**
5: $v \leftarrow \min(\max(minv_{org}, v), maxv_{org})$ /* If v is smaller than $minv_{org}$ (or larger than $maxv_{org}$), v is set to $minv_{org}$ (or $maxv_{org}$).*/
6: $v \leftarrow v + Lap((maxv_{org} - minv_{org})/\epsilon)$ /* The global sensitivity is $maxv_{org} - minv_{org}$.*/
7: $v \leftarrow \min(\max(minv_{rep}, v), maxv_{rep})$ /* If v is smaller than $minv_{rep}$ (or larger than $maxv_{rep}$), v is set to $minv_{rep}$ (or $maxv_{rep}$).*/
8: **if** the standard deviation is considered as private information **then**
9: $\quad \sigma \leftarrow \min(\max(min\sigma_{org}, \sigma), max\sigma_{org})$ /* If σ is smaller than $min\sigma_{org}$ (or larger than $max\sigma_{org}$), σ is set to $min\sigma_{org}$ (or $max\sigma_{org}$).*/
10: $\quad \sigma \leftarrow \sigma + Lap((max\sigma_{org} - min\sigma_{org})/\epsilon)$ /* The global sensitivity is $max\sigma_{org} - min\sigma_{org}$.*/
11: **end if**
12: Report v and σ.

First, a sensing device for each participant measures target data. The device obtains the sensed value v and the standard deviation σ (Line 1). If the standard deviation is considered to be private information, the privacy budget ϵ is divided by two (Line 2).

If v is smaller than $minv_{org}$, v is set to $minv_{org}$, and if v is larger than $maxv_{org}$, v is set to $maxv_{org}$ (Line 5). Then, PDE adds a Laplace noise to v to satisfy ϵ-differential privacy (Line 6). Here, the global sensitivity is $maxv_{org} - minv_{org}$.

Finally, if the value of v with Laplace noise is smaller than $minv_{rep}$ (or larger than $maxv_{rep}$), v is set to $minv_{rep}$ (or $maxv_{rep}$) (Line 7). If the standard deviation σ is considered to be private information, PDE adds a Laplace noise to σ (Line 10).

Theorem 2. *The proposed PDE realizes ϵ-differential privacy.*

Proof. The global sensitivity Δ_v of a sensing value and the global sensitivity of the standard deviation of a sensing error Δ_σ are $(maxv_{org} - minv_{min})$ and $(max\sigma_{org} - min\sigma_{org})$, respectively. According to Theorem 1, when a Laplace noise with scale Δ_v/ϵ is added to the sensing value, we can achieve ϵ-differential privacy with regard to the sensing value. Similarly, when a Laplace noise with scale Δ_σ/ϵ is added to the standard deviation of the sensing error, we can achieve ϵ-differential privacy with regard to the standard deviation.

When we consider the standard deviation to be private information, we should achieve ϵ-differential privacy for the combination of the sensing value and the standard deviation. In this case, PDE achieves $\epsilon/2$-differential privacy with regard to the sensing value and the standard deviation, respectively. Therefore, according to Reference [34], PDE achieves ϵ-differential privacy in total. □

4.3. ETE for Estimation

In this section, we propose an estimation technique that estimates the true data distribution based on the reported data, at the data-collector side: estimating true distribution considering sensing errors (ETE).

The data collector estimates the true data's distribution, which is represented by a (multi-dimensional) histogram, from the reported data. Each true data point of each participant might be unknown to the participant.

Let $\mathfrak{F}(y; x, \theta)$ be the probability density function with regard to y, which represents the reported sensing value, where x represents the true value and θ represents the set of parameters comprising the sensing error and a Laplace noise.

Let x_i and y_i represent the true sensing value and the reported sensing value of participant i, respectively. The value y_i contains a sensing error following a normal distribution and a Laplace noise to satisfy ϵ-differential privacy. That is when the true value is x_i, the probability density with which the reported value becomes y_i is $\mathfrak{F}(y_i; x_i, \theta)$. Let X and Y represent $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_N\}$, respectively. Based on $\mathfrak{F}(y; x, \theta)$, by using Bayes' technique, we can estimate the distribution of X from Y.

Let w be the width of each bin of the histogram. The value of w is calculated by

$$w = \frac{maxv_{rep} - minv_{rep}}{b_n}, \tag{3}$$

where b_n represents the number of bins of an estimated histogram, as determined by the data collector.

The function $\mathfrak{F}(y; x, \theta)$ is a probability density function, and y is a continuous random variable. The number of samples of y is a finite set in a real situation; therefore, we approximate the probability density function as a probability mass function. The domain of y is defined as

$$\boldsymbol{V} = (minv_{rep} + w/2, minv_{rep} + 2w/2, minv_{rep} + 3w/2, \ldots, minv_{rep} + b_n * w/2). \tag{4}$$

Let $\boldsymbol{P}$ be the $b_n \times b_n$ matrix and $P(i, j)$ represent the value of $\boldsymbol{P}$ in the ith row and jth column. $P(i, j)$ represents the probability that the reported value is categorized into jth bin when the true value is categorized into ith bin.

Let $\dagger_i$ be the number of participants whose reported values are categorized into the i'th bin, and let $\ddagger_i$ be the estimated number of participants whose true values are categorized into the i'th bin. Let $\boldsymbol{\dagger}$ and $\boldsymbol{\ddagger}$ be the sets $\{\dagger_1, \ldots, \dagger_{b_n}\}$ and $\{\ddagger_1, \ldots, \ddagger_{b_n}\}$, respectively.

Based on the iterative Bayes' technique [62], we have

$$\ddagger_i \leftarrow \sum_{j=1}^{b_n} \dagger_j \frac{P(i, j)\ddagger_i}{\sum_{k=1}^{b_n} P(k, j)\ddagger_k}. \tag{5}$$

Equation (5) is repeated a sufficient number of times.

Several values of the estimated data distribution might be negative. Therefore, the data distribution should be adjusted so that all values are greater than or equal to zero. The values are perturbed based on the probability simplex algorithm [63]. Moreover, because the data collector determines the value range for sensing in advance, values that are out of range should be zero. Note that to use differential privacy as a privacy metric, we must determine the value range in advance if we use any other methods that can satisfy differential privacy. Therefore, in each iteration, for

$$i \leq \left\lceil \frac{minv_{org} - minv_{rep}}{w} \right\rceil - 1, \tag{6}$$

and

$$i \geq \left\lceil \frac{maxv_{org} - minv_{rep}}{w} \right\rceil - 1, \tag{7}$$

we set

$$\ddagger_i \leftarrow 0, \tag{8}$$

because the true values are within $minv_{org}$ and $maxv_{org}$.

Now, we describe how to obtain $\boldsymbol{P}$. Each value of $P(i, j)$ is calculated by the following equation for all values of i;

$$\begin{cases} P(i,1) = \int_{-\infty}^{minv_{rep}+w} \mathfrak{F}(y; minv_{rep} + (i-1) * w + w/2, \theta)dy \\ P(i,j) = \int_{minv_{rep}+(j-1)*w}^{minv_{rep}+j*w} \mathfrak{F}(y; minv_{rep} + (i-1) * w + w/2, \theta)dy \text{ for } j = 2, \ldots, b_n - 1 \\ P(i,b_n) = \int_{minv_{rep}+(b_n-1)*w}^{\infty} \mathfrak{F}(y; minv_{rep} + (i-1) * w + w/2, \theta)dy. \end{cases} \tag{9}$$

The function $\mathfrak{F}(y; x, \theta)$ differs for the two scenarios. First, we consider the scenario in which the standard deviation of an error distribution is not private information. That is, a Laplace noise is added to the sensed value before the value is reported to the data collector, but each participant reports the standard deviation of the sensing error as it is to the data collector. In this case, the data collector can determine the true standard deviation of the sensing error's normal distribution. Let b_v be a scale factor of a Laplace noise with regard to the sensed value. The value b_v is represented by

$$b_v = \frac{maxv_{org} - minv_{org}}{\epsilon}, \tag{10}$$

and we can consider $\theta = \{\sigma, b_v\}$.

In this case,

$$\mathfrak{F}(y; x, \theta) = \mathfrak{F}(y; x, \sigma, b_v) = \int_{-\infty}^{\infty} \mathcal{N}(t; x, \sigma) * \mathcal{L}(y; t, b_v)dt, \tag{11}$$

where $\mathcal{N}(t; x, \sigma)$ represents the probability density of t in a normal distribution with a mean of x and a standard deviation of σ, and $\mathcal{L}(y; t, b_v)$ represents the probability density of y in a Laplace distribution with a mean of t and a scale factor of b_v.

In the second scenario, where the standard deviation σ of the sensing error is considered to be private information, a Laplace noise is added to not only the sensed value x, but also the value σ, as described in Section 4.2.

Let b_v and b_σ be scale factors of a Laplace noise with regard to the sensed value and the standard deviation, respectively. The values b_v and b_σ are represented by

$$b_v = \frac{maxv_{org} - minv_{org}}{\epsilon/2}, \tag{12}$$

and

$$b_\sigma = \frac{max\sigma_{org} - min\sigma_{org}}{\epsilon/2}. \tag{13}$$

In this case, we consider $\theta = \{\sigma, b_v, b_\sigma\}$, and obtain

$$\begin{aligned} &\mathfrak{F}(y; x, \theta) = \mathfrak{F}(y; x, \sigma, b_v, b_\sigma) \\ &= \int_{-\infty}^{\infty} \int_{0}^{\infty} \mathcal{N}(t; x, u) * \mathcal{L}(y; t, b_v) * \mathcal{L}(\sigma; u, b_\sigma)dudt/V, \end{aligned} \tag{14}$$

where

$$V = \int_{-\infty}^{0} \mathcal{L}(x; \sigma, b_\sigma)dx = e^{-\sigma/b_\sigma}/2. \tag{15}$$

Figure 3 shows a high-level diagram of the estimation algorithm (ETE) and Algorithm 2 shows the details.

Because the values of $P(i,j)$ ($i = 1, \ldots, b_n$ and $j = 2, \ldots, b_n - 1$) are the same when the values $|i - j|$ are the same, we calculate only $P(1,j)$ (represented by $Q(j)$) and additional values represented by $left$ and $right$ in lines 9–13. Then, we construct $P(i,j)$ in lines 14–20.

Figure 4a represents the relationship between $P(1,j)$ and $Q(j)$. Each value of $Q(j)$ represents the area marked by the corresponding arrow. The curve line represents the $\mathfrak{F}(y; x, \theta)$. The value of x in Algorithm 2 can be arbitrary but is set to the middle of the area, represented by $Q(1)$. Because the summation of $left + right + \sum_{j=1}^{b_n} Q(j)$ is equal to one, we obtain the value of $right$ by $1 - left - \sum_{j=1}^{b_n} Q(j)$ in Line 13.

Algorithm 2 Estimation Algorithm.

Input: $Y, Y_\sigma, \epsilon, minv_{org}, maxv_{org}, minv_{rep}, maxv_{rep}, min\sigma_{rep}, max\sigma_{rep}, b_n$
Output: ‡

$\sigma_{ave} \leftarrow Average(Y_\sigma)$ /* Consider σ_{ave} is the standard deviation of each participant*/
if standard deviation is considered as private information **then**
  b_v and b_σ are calculated by Equations (12) and (13), and set $\theta = \{\sigma_{ave}, b_v, b_\sigma\}$.
else
  b_v is calculated by Equation (10), and set $\theta = \{\sigma_{ave}, b_v\}$.
end if
$w \leftarrow (maxv_{rep} - minv_{rep})/b_n$ /*w represents the width of each bin*/
$x \leftarrow$ an arbitrary real number
$left \leftarrow \int_{-\infty}^{x-w/2} \mathfrak{F}(y; x, \theta)dy$
for $j = 1, \ldots, b_n$ **do**
  $Q(j) \leftarrow \int_{x-w/2+(j-1)*w}^{x-w/2+j*w} \mathfrak{F}(y; x, \theta)dy$ /*$P(1,j) = Q(j)$ for $j = 2, \ldots, b_n - 1$. $P(1,1) = left + Q(1)$. $P(1, b_n) = Q(b_n) + right$.*/
end for
$right \leftarrow 1 - left - \sum_{j=1}^{b_n} Q(j)$
for $i = 1, \ldots, b_n$ **do**
  $P(i,1) \leftarrow left + Q(1) - \sum_{j=1}^{i-1} Q(j)$ /*Note that $\sum_{j=1}^{0} Q(j) = 0$*/
  for $j = 2, \ldots, b_n - 1$ **do**
    $P(i,j) \leftarrow Q(|i-j|+1)$
  end for
  $P(i, b_n) \leftarrow \sum_{j=b_n-i+1}^{b_n} Q(j) + right$
end for
Set $\dagger_i$ for each i based on Y.
for Repeat sufficient times **do**
  for $i = 1, \ldots, b_n$ **do**
    $d_i \leftarrow 0$
    for $j = 1, \ldots, b_n$ **do**
      $d_i \leftarrow d_i + P(k,j) * \ddagger_k$ /*Calcuation of the denominator of Equation (5)*/
    end for
  end for
  for $i = 1, \ldots, b_n$ **do**
    for $j = 1, \ldots, b_n$ **do**
      $\ddagger'_i \leftarrow \dagger_j * P(i,j)/d_j$
    end for
    $\ddagger_i \leftarrow \ddagger_i * \ddagger'_i$
  end for
  for $i = 1, \ldots, b_n$ **do**
    if $i \leq \lceil \frac{minv_{org} - minv_{rep}}{w} \rceil - 1$ OR $\lceil \frac{maxv_{org} - minv_{rep}}{w} \rceil - 1 \leq i$ **then**
      $\ddagger_i \leftarrow 0$
    end if
  end for
end for

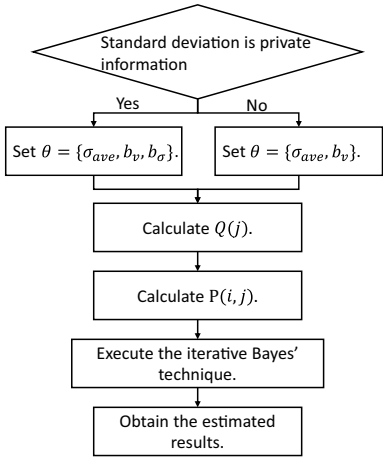


Figure 3. A high-level diagram of the estimation algorithm.

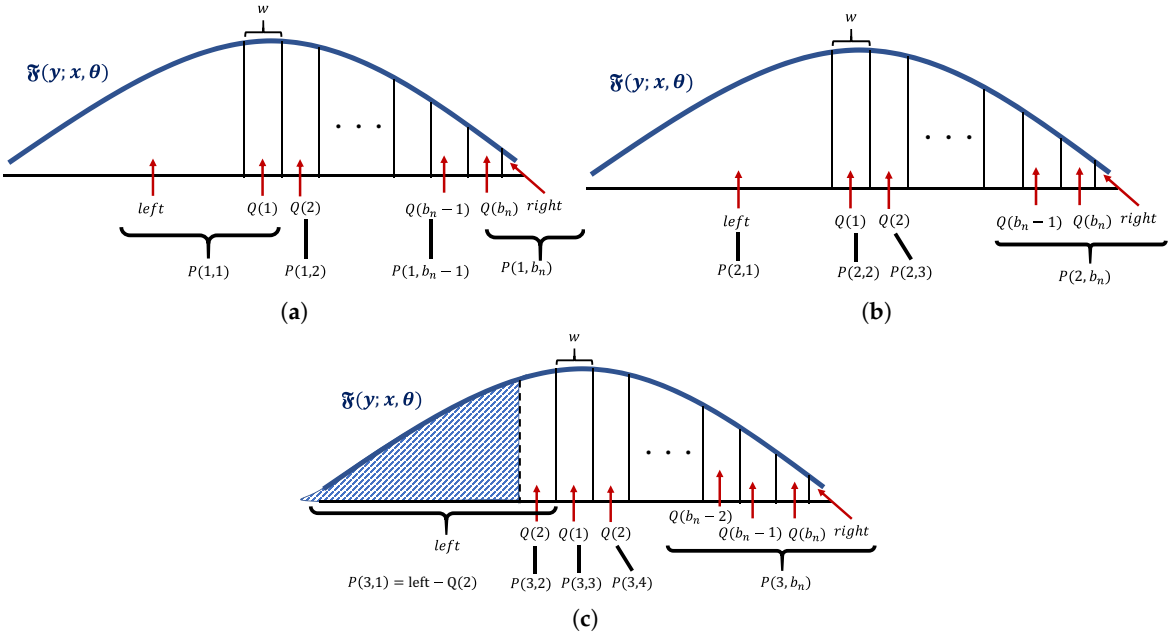


Figure 4. Relationship between $Q(j)$ and $P(i, j)$. (**a**) Relationship between $Q(j)$ and $P(1, j)$. (**b**) Relationship between $Q(j)$ and $P(2, j)$. (**c**) Relationship between $Q(j)$ and $P(3, j)$. $P(3, 1)$ represents the shaded area.

Figure 4b,c represent the relationship between $P(2, j)$ and $Q(j)$ and the relationship between $P(3, j)$ and $Q(j)$, respectively. As increases i, $P(i, 1)$ decreases, and $P(i, b_n)$ increase.

Lines 23–34 show the iterative Bayes' technique. Line 25–27 calculates the value of the denominator of Equation (5). Lines 30–32 calculate the fraction of Equation (5). Finally, the summation of Equation (5) is calculated by Line 33.

Lines 35–39 show the process of Equations (6)–(8).

5. Evaluation

Our proposed architecture models sensing errors. If we do not consider the sensing errors, then we consider that only a Laplace noise is added to the true data, even if the sensed data differs from the true data in a real situation. To verify the usefulness of considering the sensing errors, we developed a method of considering only the Laplace noise. We refer to this method as the Laplace mechanism. In this section, we compare our proposal with the Laplace mechanism and with S2Mb, which is described in Section 3. The Laplace mechanism, S2Mb, and the proposed method all use iterative

Bayes' technique. We set the iteration times as the best values for each method, for each simulation, within 100,000 iterations.

The source code for the proposed architecture can be obtained from https://uecdisk.cc.uec.ac.jp/index.php/s/WfIyH8hRMhoF01R. This source code consists of the server (data collector) program and the client (participant) program.

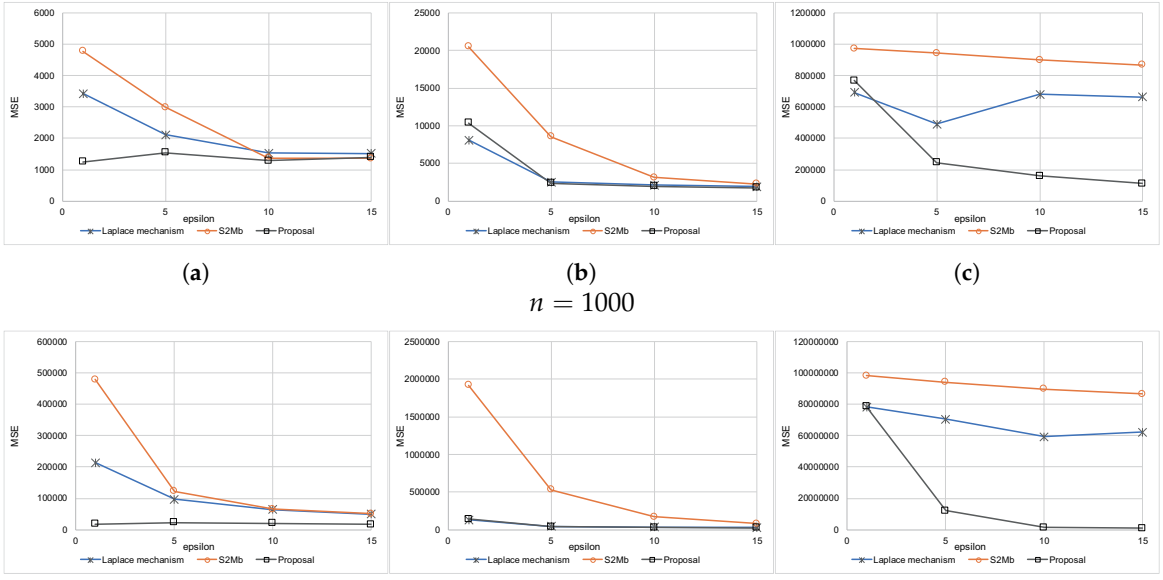
Apple's deployment ensures that ϵ is equal to 1 or 2 per each datum [64], and that the total privacy loss is 16 per day. An Apple differential privacy team set $\epsilon = 2, 4, 8$ for its evaluations [65]. Based on these settings, ϵ is set in the range 1–15 in the experiments.

5.1. Evaluation of Synthetic Data

First, we evaluated the MSE using synthetic datasets. We conducted experiments using several distributions to determine how different data distributions would affect the results. We used three distributions: normal, uniform, and peak. In the uniform distribution, all values of $\S_i$ were set to the same value. In the normal distribution, the values of $\S_i$ followed a normal distribution. In the peak distribution, all of the participants had the same true value.

Every setting was executed 10 times. The average results are shown in Figure 5 for when the standard deviation of sensing errors is not considered private information. Because the MSEs measure the difference between the true number of people and the estimated number of people within each bin, the MSEs become larger as the number of participants N becomes larger. A large value of ϵ means a low privacy-protection level. Therefore, when ϵ is large, the MSEs tend to become small for all methods. Figure 6 represents the experimental results when the standard deviation of the sensing errors is considered private information. Because the standard deviation should be protected in the same way as the sensed values in this situation, the MSEs are larger than those of the results in Figure 5. In all of the settings, the MSEs of our proposed architecture were the smallest among the three methods.

We measured the calculation time at the data collection server's side. All of the experiments were conducted on a desktop PC with an Intel i7-4770 CPU and 16 GB of RAM. The average calculation time was less than 1 s for the Laplace mechanism and for S2Mb. Our proposed ETE required 14.7 s for each simulation, on average. Although the calculation time of the proposed method is longer than those of the other methods, we believe that the time does not greatly impact the data analysis because gathering participants takes a much longer time (for example, a few days).



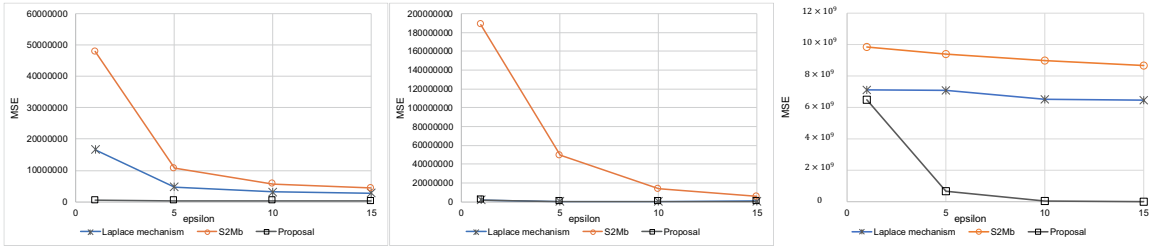
(**a**) (**b**) (**c**)

$n = 1000$

(**d**) (**e**) (**f**)

$n = 10{,}000$

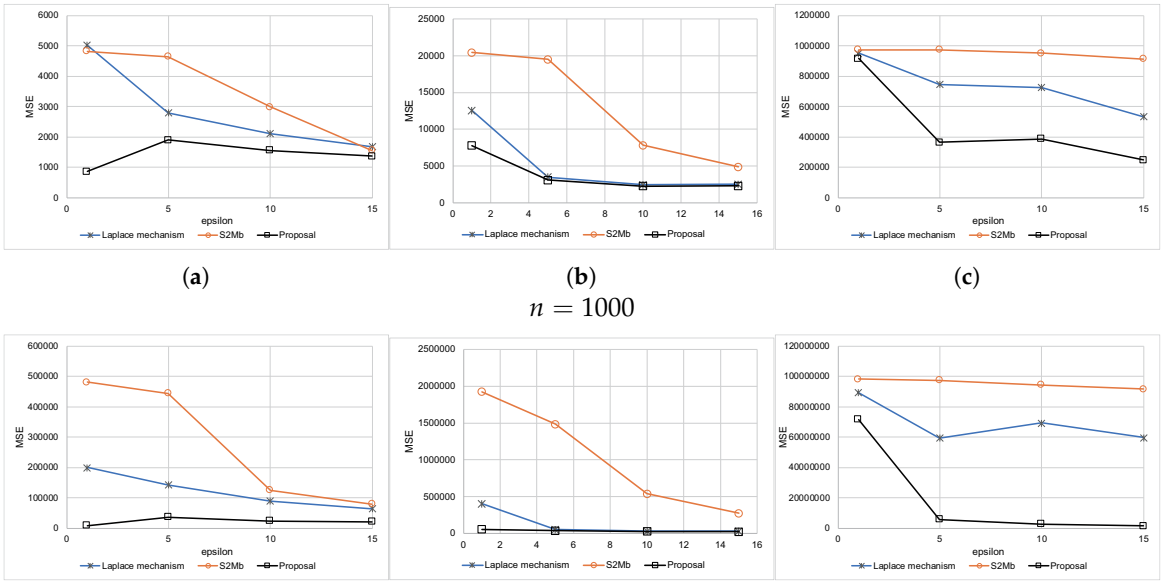
Figure 5. *Cont.*



(**g**) (**h**) (**i**)

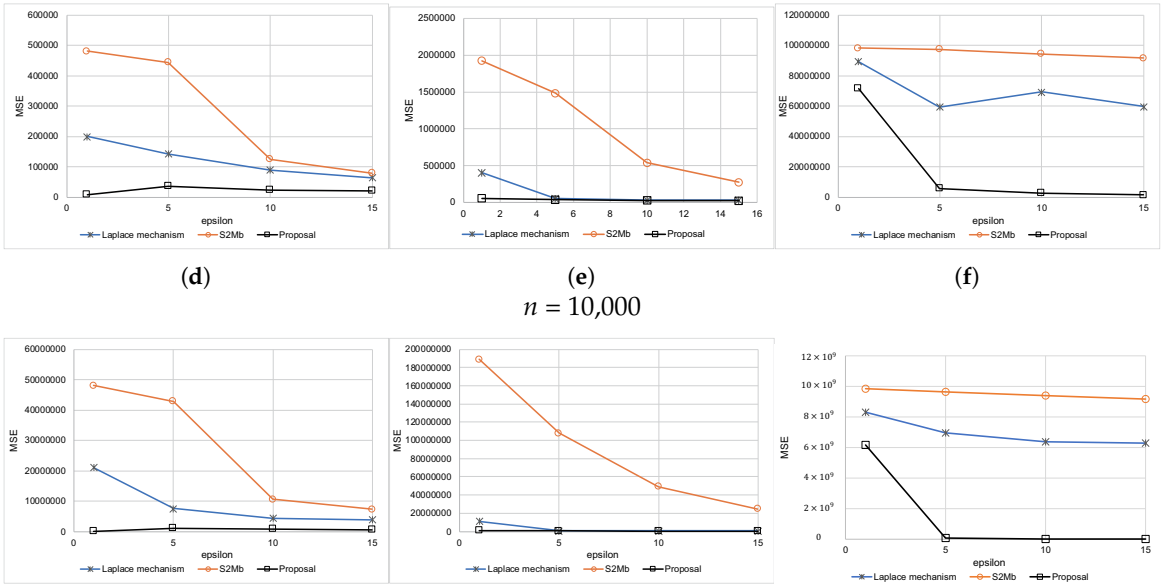
$n = 100{,}000$

Figure 5. Results of synthetic data (the standard deviation is not private information). (**a**) Uniform. (**b**) Normal. (**c**) Peak. (**d**) Uniform. (**e**) Normal. (**f**) Peak. (**g**) Uniform. (**h**) Normal. (**i**) Peak.



(**a**) (**b**) (**c**)

$n = 1000$



(**d**) (**e**) (**f**)

$n = 10{,}000$

(**g**) (**h**) (**i**)

$n = 100{,}000$

Figure 6. Results of synthetic data (the standard deviation is private information). (**a**) Uniform. (**b**) Normal. (**c**) Peak. (**d**) Uniform. (**e**) Normal. (**f**) Peak. (**g**) Uniform. (**h**) Normal. (**i**) Peak.

5.2. Evaluation of Real Data

5.2.1. Location Data

We implemented our proposed PDE as a smartphone application for Android to obtain real sensing data with sensing errors and to verify the algorithm's feasibility.

Operating systems such as iOS and Android express location by latitude, longitude, and uncertainty (https://developer.apple.com/documentation/corelocation/cllocation [Accessed on 26 March 2020],

https://developer.android.com/reference/android/location/Location [Accessed on 26 March 2020]). Uncertainty means a radius of a circle centered at the location's latitude and longitude, and the true location is inside the circle with 68% probability. In a normal distribution, 68% of the data fall within one standard deviation from the mean.

The smartphone was located in the same place and sensed its location along with its uncertainty 200 times. In this experiment, we considered that 200 different people were in the same place. The true distribution of locations is shown in Figure 7. The smartphone reported its differential private location and uncertainty to the data-collection server. We evaluated the MSEs of each method. Figure 8 represents the results. The MSEs of our proposed method were much smaller than those of the other methods.

Figures 9 and 10 show the example results of the histograms generated with the Laplace mechanism, S2Mb, and the proposed method. The standard deviation of the sensing errors was considered private information in Figure 10. The histograms of Figures 9c and 10c, which were generated by our proposed architecture, are similar to the true histogram (Figure 7). However, the histograms generated by the Laplace mechanism and S2Mb (Figures 9a,b and 10a,b) are very different from the true histogram.

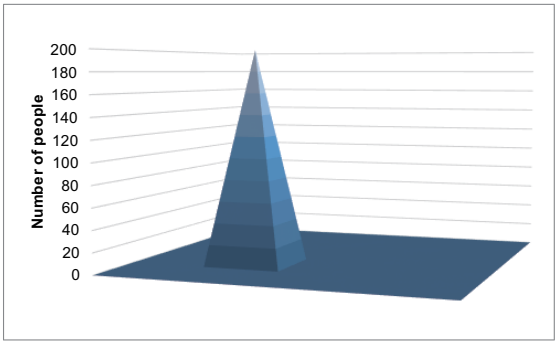
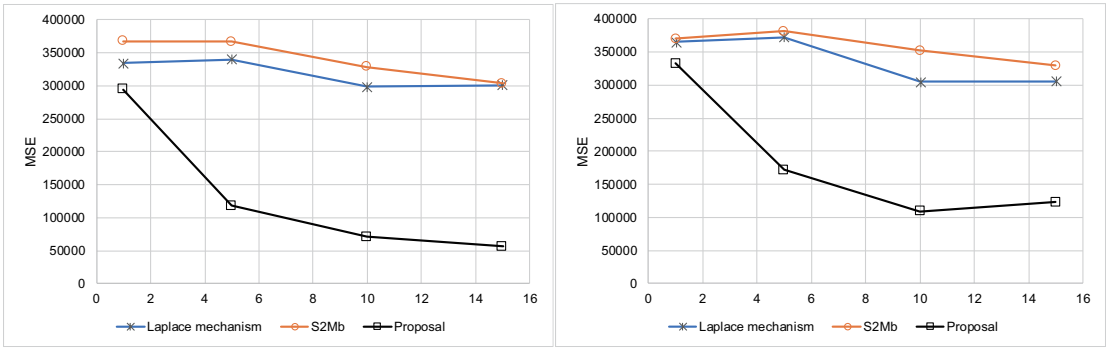
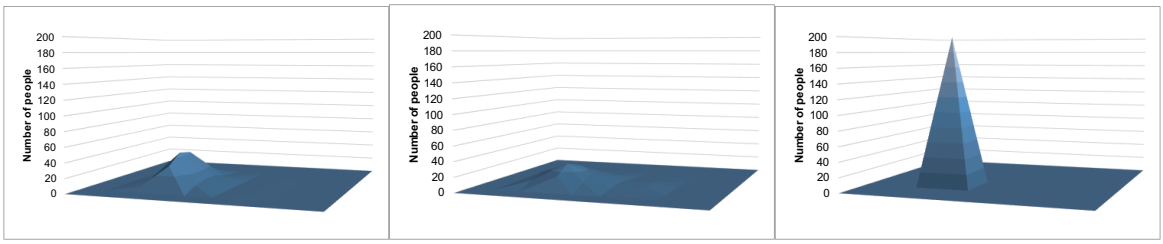


Figure 7. The distribution of the participants' true locations.



(**a**) (**b**)

Figure 8. Results summary of the location data. (**a**) Standard deviation is not private information. (**b**) Standard deviation is not private information.



(**a**) (**b**) (**c**)

Figure 9. Example results of the location data (the standard deviation is not private information). (**a**) Laplace mechanism. (**b**) S2Mb. (**c**) Proposal.

Furthermore, because some of the participants were concerned about battery consumption [66], we measured the calculation time needed for sensing the GPS and generating differential private data. The smartphone used in this experiment was a SH-M09 with a Snapdragon 845 CPU and 4 GB of RAM. The application was developed with Java. The average time spent for 10 simulations was 100.6 ms. Our PDE is efficient for smartphones, and participants do not need to worry about their smartphones' battery life.

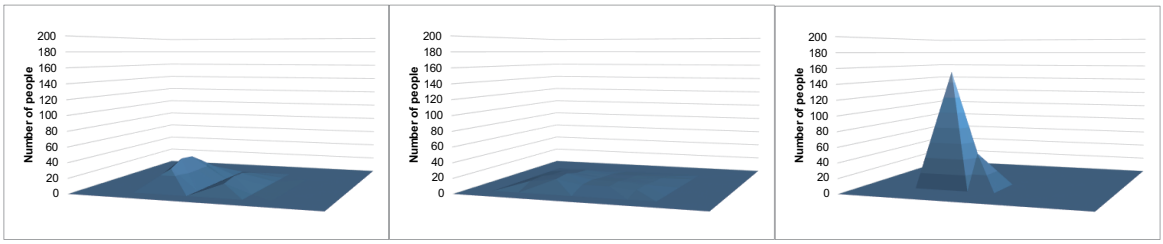


Figure 10. Example results of the location data (the standard deviation is private information). (**a**) Laplace mechanism. (**b**) S2Mb. (**c**) Proposal.

5.2.2. Deep Neural Network's Output Data

Crowdsensing might collect an output of a machine learning model, such as deep neural networks (DNNs). For example, each participant's device can recognize his/her activity from an accelerometer, magnetometer, and gyroscope [67,68] and recognize surrounding people's age from pictures [69,70]. Surrounding information, such as how many people there are and how old they are, is useful to analyze for a pandemic such as the coronavirus pandemic. For example, age is an important factor for COVID-19 [71,72].

The estimated values from deep neural networks might include estimation errors, and researchers such as [27–29] have reported that such estimation errors followed a normal distribution. Several machine-learning models can obtain the probability distribution of a model's estimated value. For example, the age-estimation model [73] outputs the probability for a person being each age (e.g., the probability of being 1 year old is 0.01%, the probability of being 2 years old is 0.05%, . . ., the probability of being 33 years old is 32.3%, . . .). We developed a deep neural network model that estimates a person's age from a picture, based on Reference [73].

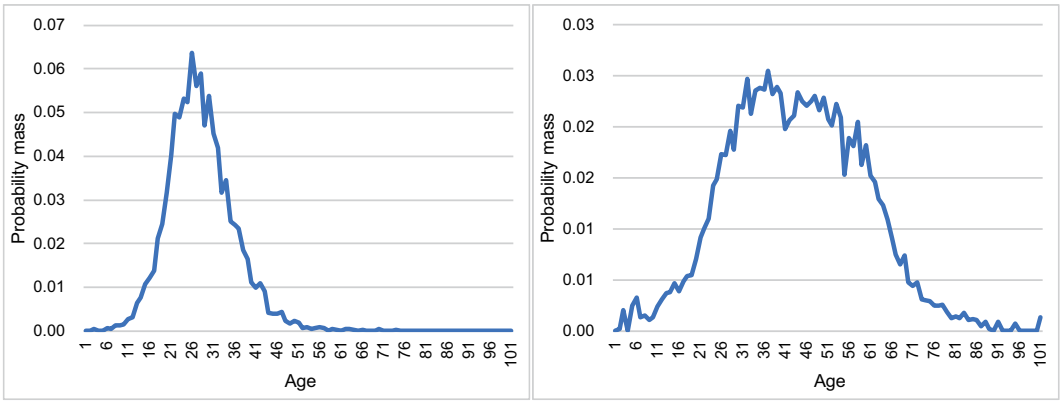
We assume that it does not make a big difference if the participants report sensing data or an estimated age value. This is because the estimation error of deep neural networks can be considered to follow normal distributions, much as how sensing errors follow normal distributions. We consider that not all estimation errors of deep neural networks follow normal distributions. However, several estimation errors of deep neural networks follow normal distributions, and our proposed method targets such deep neural networks. To confirm that our proposed method can be used for outputs of deep neural networks, this experiment has been conducted.

Table 3 shows the architecture of the deep neural network model we constructed. All of the activation functions of layers are rectified linear units (ReLUs [74]). The loss function was the softmax function. Because our aim is not to increase the accuracy of the deep neural network itself, the accuracy might be increased by tuning architecture or parameters.

We assumed that a crowdsensing application for each smartphone would estimate the surrounding person's age. Because the model outputs the probability distribution of age, our PDE can calculate the standard deviation of errors at each device. Figure 11 represents the probability distributions of age, which were obtained from the trained deep neural network model. These distributions can be considered as normal distributions.

Table 3. Architecture of a deep neural network used in the experiment.

Layer ID	Description of Each Layer
1	Input Layer
2	Convolutional Layer
3	Convolutional Layer
4	Max Pooling Layer
5	Convolutional Layer
6	Convolutional Layer
7	Max Pooling Layer
8	Convolutional Layer
9	Convolutional Layer
10	Convolutional Layer
11	Max Pooling Layer
12	Convolutional Layer
13	Convolutional Layer
14	Convolutional Layer
15	Max Pooling Layer
16	Convolutional Layer
17	Convolutional Layer
18	Convolutional Layer
19	Max Pooling Layer
20	Convolutional Layer
21	Convolutional Layer
22	Fully Connected Layer
23	Output Layer

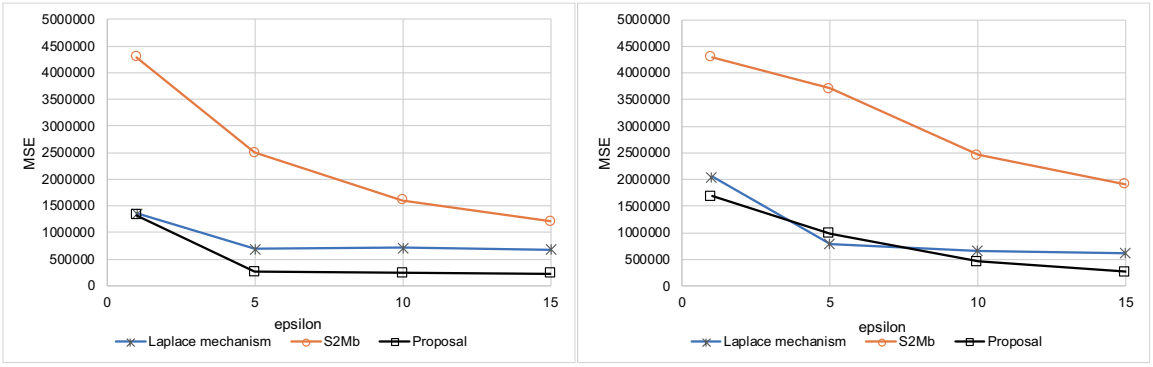


(**a**) (**b**)

Figure 11. Examples of DNN's output (probability distribution). (**a**) Example 1. (**b**) Example 2.

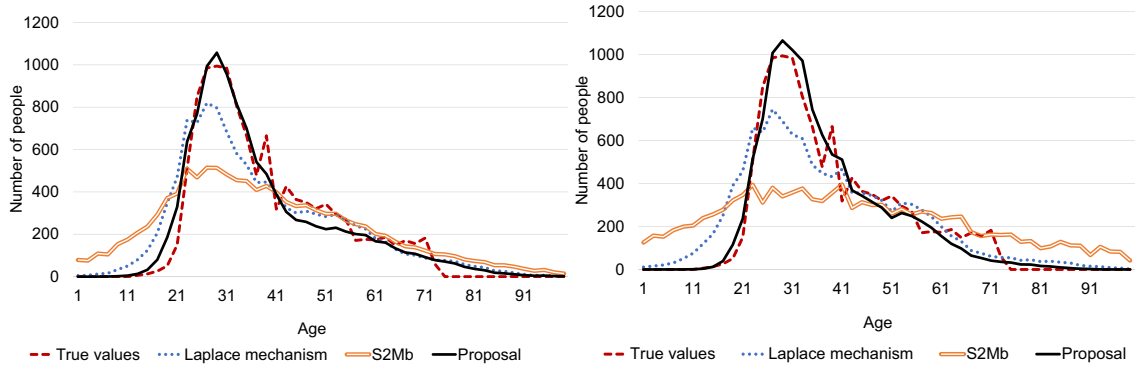
We used the WIKI dataset, which consists of 22,578 instances (1 GB) (https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/static/wiki_crop.tar [Accessed on 26 March 2020]). Fifty percent of the dataset was used for our prediction task, that is, we assumed that 11,289 people were the participants. The data collector estimated the true age distribution from the reports. Because each picture in WIKI dataset is labeled true age, we can evaluate the performance of Laplace mechanisms, S2Mb, and the proposal.

Figure 12 summarizes the results of this experiment. In both scenarios, the MSEs of the proposal were smaller than those of the other methods in almost all settings. The true and estimated data distributions are shown in Figure 13. The line of the proposal fits the true values' line in Figure 13a,b.



(**a**) (**b**)

Figure 12. Summary results of age estimation. (**a**) Standard deviation is not private information. (**b**) Standard deviation is private information.



(**a**) (**b**)

Figure 13. Example results of age estimation. (**a**) Standard deviation is not private information. (**b**) Standard deviation is private information.

6. Discussion

In this paper, we assume that the sensing campaigns assign a single sensing task for simple discussion. However, our method can also easily be used for multiple tasks.

Assume that there are two tasks. For example, the first task is collecting a noise, and the second task is collecting humidity. In this case, we assume that the aim of the data collector is to create a 3D histogram (Figure 14).

Each participant perturbs the two values separately by our proposed PDE method. Then, each participant reports the resulted values and the standard deviations to the data collector. The data collector constructs $P_1(i_1, j_1)$ for the first task (noise sensing) and $P_2(i_2, j_2)$ for the second task (humidity sensing) separately (Lines 1–20 in Algorithm 2). Here, $P_1(i_1, j_1)$ represents the probability that the reported value of the first task is categorized into j_1th bin when the true value of the first task is categorized into i_1th bin in the first dimension. In the example in Figure 14, $P_1(1, 2)$ represents the probability that the reported value of the noise is "Noise 2" when the true value of the noise is "Noise 1".

Assume that the number of bins for the first task is b_{n1}, and the number of bins for the second task is b_{n2}. In the example in Figure 14, $b_{n1} = 4$ and $b_{n2} = 5$. The data collector constructs $P_{1,2}([i_1, i_2], [j_1, j_2])$ for $i_1, j_1 = 1, \ldots, b_{n1}$ and $i_2, j_2 = 1, \ldots, b_{n2}$, which represents that the reported values of the first and second tasks are categorized into j_1th and j_2th bins, respectively, while the true values of the first and second tasks are categorized into i_1th and i_2th bins, respectively. In the example in Figure 14, $P_{1,2}([1, 3], [2, 1])$ represents the probability that the reported values are "Noise 2" and "Humidity 1," while the true values are "Noise 1" and "Humidity 3".

Because each sensed value is perturbed separately, we can calculate $P_1,2([i_1,i_2],[j_1,j_2]) = P_1(i_1,j_1)$* $P_2(i_2,j_2)$. Then, the data collector executes the iterative Bayes' technique using $P_1,2([i_1,i_2],[j_1,j_2])$ (Lines 21–40 in Algorithm 2). Finally, the data collector obtains each estimated number of people in each two-dimensional bin (Figure 14).

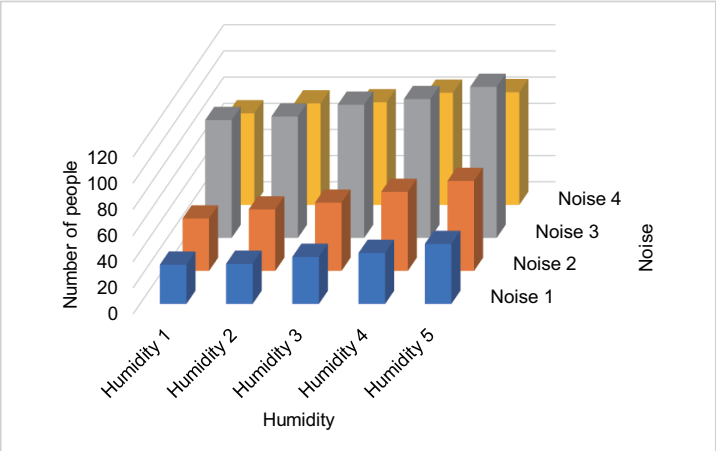


Figure 14. An example of a histogram created by two tasks.

7. Conclusions and Future Work

Participatory sensing is growing in popularity. Differential privacy can protect a user's privacy by adding noise to a target value that must be protected. However, in participatory sensing scenarios, the target value contains sensing errors. Because existing studies do not consider the sensing errors, the accuracy of the data analysis decreases when the sensing data contain errors. In this paper, therefore, the proposed architecture can address the noise added to the sensed value. The true data might be unknown to the participants; however, our proposal estimated the participants' true data distribution with higher accuracy than existing methods by modeling the sensing error.

The proposed architecture consists of two parts. One is the anonymization technique for each participant's side (PDE). Each device perturbs its sensed data and then reports the perturbed data to the data collector. The proposed architecture also provides an estimation technique, which estimates the true data distribution based on the reported data for the data collector's side (ETE). We have proved that the PDE satisfies differential privacy. We showed that the accuracy of ETE outperformed existing studies in our experiments. Further, the calculation time of PDE with a normal smartphone was less than 1 s. Therefore, participants do not need to worry about the battery life of their smartphones.

In this paper, we target numerical data with regard to sensing data. Moreover, images can be directly sent to the data collector. In recent years, several methods of protecting images based on differential privacy have been proposed [75]. We will apply our proposal to such data in our future work.

Author Contributions: Conceptualization, Y.S. and A.O.; methodology, Y.S. and A.O.; software, Y.S.; validation, Y.S.; formal analysis, Y.S.; investigation, A.O.; resources, Y.S.; data curation, Y.S.; writing—original draft preparation, Y.S.; writing—review and editing, Y.S. and A.O.; visualization, Y.S. and A.O.; supervision, A.O.; project administration, Y.S. and A.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI Grant Numbers JP17H04705, JP18H03229, JP18H03340, JP18K19835, JP19K12107, JP19H04113. This work was supported by JST, PRESTO Grant Number JPMJPR1934.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Miluzzo, E.; Lane, N.D.; Fodor, K.; Peterson, R.; Lu, H.; Musolesi, M.; Eisenman, S.B.; Zheng, X.; Campbell, A.T. Sensing meets mobile social networks. In Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, Raleigh, NC, USA, 5–7 November 2008; pp. 337–350.
2. Bridgelall, R.; Tolliver, D. Accuracy Enhancement of Anomaly Localization with Participatory Sensing Vehicles. *Sensors* **2020**, *20*, 409. [CrossRef] [PubMed]
3. Kozu, R.; Kawamura, T.; Egami, S.; Sei, Y.; Tahara, Y.; Ohsuga, A. User participatory construction of open hazard data for preventing bicycle accidents. In Proceedings of the Joint International Semantic Technology Conference (JIST), Gold Coast, Australia, 10–12 November 2017; pp. 289–303. [CrossRef]
4. Khoi, N.; Casteleyn, S.; Moradi, M.; Pebesma, E. Do Monetary Incentives Influence Users' Behavior in Participatory Sensing? *Sensors* **2018**, *18*, 1426. [CrossRef] [PubMed]
5. Domínguez, D.R.; Díaz Redondo, R.P.; Vilas, A.F.; Khalifa, M.B. Sensing the city with Instagram: Clustering geolocated data for outlier detection. *Expert Syst. Appl.* **2017**, *78*, 319–333. [CrossRef]
6. Igartua, M.A.; Almenares, F.; Redondo, R.P.D.; Martin, M.I.; Forne, J.; Campo, C.; Fernandez, A.; De la Cruz, L.J.; Garcia-Rubio, C.; Marin, A.; et al. INRISCO: INcident monitoRing In Smart COmmunities. *IEEE Access* **2020**, *8*, 72435–72460. [CrossRef]
7. Kairouz, P.; Bonawitz, K.; Ramage, D. Discrete Distribution Estimation under Local Privacy. In Proceedings of the ICML, New York, NY, USA, 19–24 June 2016; pp. 2436–2444.
8. Sei, Y.; Ohsuga, A. Differential Private Data Collection and Analysis Based on Randomized Multiple Dummies for Untrusted Mobile Crowdsensing. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 926–939. [CrossRef]
9. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating Noise to Sensitivity in Private Data Analysis. In Proceedings of the Theory of Cryptography (TCC), New York, NY, USA, 4–7 March 2006; pp. 265–284.
10. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
11. Liu, F. Generalized Gaussian Mechanism for Differential Privacy. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 747–756. [CrossRef]
12. Ren, X.; Yu, C.M.; Yu, W.; Yang, S.; Yang, X.; McCann, J.A.; Yu, P.S. LoPub: High-dimensional crowdsourced data publication with local differential privacy. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2151–2166. [CrossRef]
13. Phan, N.; Wu, X.; Hu, H.; Dou, D. Adaptive Laplace Mechanism: Differential Privacy Preservation in Deep Learning. In Proceedings of the IEEE ICDM, New Orleans, LA, USA, 18–21 November 2017; pp. 385–394.
14. Zhang, X.; Chen, R.; Xu, J.; Meng, X.; Xie, Y. Towards Accurate Histogram Publication under Differential Privacy. In Proceedings of the SIAM SDM Workshop on Data Mining for Medicine and Healthcare, Philadelphia, PA, USA, 24–26 April 2014; pp. 587–595.
15. Chen, P.T.; Chen, F.; Qian, Z. Road Traffic Congestion Monitoring in Social Media with Hinge-Loss Markov Random Fields. In Proceedings of the IEEE ICDM, Shenzhen, China, 14–17 December 2014; pp. 80–89.
16. Schubert, E.; Zimek, A.; Kriegel, H.P. Generalized Outlier Detection with Flexible Kernel Density Estimates. In Proceedings of the SIAM SDM, Philadelphia, PA, USA, 22–29 January 2014; pp. 542–550.
17. Lyon, A. Why are Normal Distributions Normal? *Br. J. Philos. Sci.* **2014**, *65*, 621–649. [CrossRef]
18. Peng, R.; Sichitiu, M.L. Angle of arrival localization for wireless sensor networks. In Proceedings of the IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON), Santa Clara, CA, USA, 28 September 2006; pp. 374–382. [CrossRef]
19. Floris, I.; Calderón, P.A.; Sales, S.; Adam, J.M. Effects of core position uncertainty on optical shape sensor accuracy. *Meas. J. Int. Meas. Confed.* **2019**, *139*, 21–33. [CrossRef]
20. Burguera, A.; González, Y.; Oliver, G. Sonar sensor models and their application to mobile robot localization. *Sensors* **2009**, *9*, 10217–10243. [CrossRef] [PubMed]
21. Devon, D.; Holzer, T.; Sarkani, S. Minimizing uncertainty and improving accuracy when fusing multiple stationary GPS receivers. In Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, San Diego, CA, USA, 14–16 September 2015; pp. 83–88. [CrossRef]
22. Wang, Q.; Kurillo, G.; Ofli, F.; Bajcsy, R. Evaluation of pose tracking accuracy in the first and second generations of microsoft Kinect. In Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), Dallas, TX, USA, 21–23 October 2015; pp. 380–389. [CrossRef]

23. Li, J. Permissible Area Analyses of Measurement Errors with Required Fault Diagnosability Performance. *Sensors* **2019**, *19*, 4880. [CrossRef]
24. Wang, D.; Chen, S.; Li, X.; Zhang, W.; Jin, H. Research on Rotational Angle Measurement for the Smart Wheel Force Sensor. *Sensors* **2020**, *20*, 1037. [CrossRef] [PubMed]
25. Gonzalez, R.; Dabove, P. Performance Assessment of an Ultra Low-Cost Inertial Measurement Unit for Ground Vehicle Navigation. *Sensors* **2019**, *19*, 3865. [CrossRef]
26. Nguyen, L.V.; Kodagoda, S.; Ranasinghe, R.; Dissanayake, G. Adaptive Placement for Mobile Sensors in Spatial Prediction under Locational Errors. *IEEE Sens. J.* **2017**, *17*, 794–802. [CrossRef]
27. Pereira Barbeiro, P.N.; Krstulovic, J.; Teixeira, H.; Pereira, J.; Soares, F.J.; Iria, J.P. State estimation in distribution smart grids using autoencoders. In Proceedings of the IEEE International Power Engineering and Optimization Conference (PEOCO), Langkawi, Malaysia, 24–25 March 2014; pp. 358–363. [CrossRef]
28. Wang, Q.; Ye, L.; Luo, H.; Men, A.; Zhao, F.; Huang, Y. Pedestrian Stride-Length Estimation Based on LSTM and Denoising Autoencoders. *Sensors* **2019**, *19*, 840. [CrossRef]
29. Uss, M.; Vozel, B.; Lukin, V.; Chehdi, K. Efficient Discrimination and Localization of Multimodal Remote Sensing Images Using CNN-Based Prediction of Localization Uncertainty. *Remote Sens.* **2020**, *12*, 703. [CrossRef]
30. Bhaskar, R.; Laxman, S.; Smith, A.; Thakurta, A. Discovering frequent patterns in sensitive data. In Proceedings of the ACM KDD, Washington, DC, USA, 24–28 July 2010; pp. 503–512.
31. Chen, R.; Fung, B.C.; Desai, B.C.; Sossou, N.M. Differentially private transit data publication. In Proceedings of the ACM KDD, Beijing, China, 12–16 August 2012; pp. 213–221.
32. Ren, H.; Li, H.; Liang, X.; He, S.; Dai, Y.; Zhao, L. Privacy-Enhanced and Multifunctional Health Data Aggregation under Differential Privacy Guarantees. *Sensors* **2016**, *16*, 1463. [CrossRef]
33. Dwork, C. Differential Privacy. In Proceedings of the ICALP, Venice, Italy, 10–14 July 2006; pp. 1–12.
34. Kasiviswanathan, S.P.; Lee, H.K.; Nissim, K.; Raskhodnikova, S.; Smith, A. What Can We Learn Privately? *SIAM J. Comput.* **2013**, *40*, 793–826. [CrossRef]
35. Li, Q.; Cao, G.; Porta, T.F. Efficient and privacy-aware data aggregation in mobile sensing. *IEEE Trans. Dependable Secur. Comput.* **2014**, *11*, 115–129. [CrossRef]
36. Lu, R.; Liang, X.; Li, X.; Lin, X.; Shen, X.S. EPPA: An Efficient and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications. *IEEE Trans. Parallel Distrib. Syst.* **2012**, *23*, 1621–1631.
37. Shen, X.; Zhu, L.; Xu, C.; Sharif, K.; Lu, R. A privacy-preserving data aggregation scheme for dynamic groups in fog computing. *Inf. Sci.* **2020**, *514*, 118–130. [CrossRef]
38. Agrawal, S.; Haritsa, J. A Framework for High-Accuracy Privacy-Preserving Mining. In Proceedings of the IEEE ICDE, Tokyo, Japan, 5–8 April 2005; pp. 193–204.
39. Agrawal, S.; Haritsa, J.R.; Prakash, B.A. FRAPP: A framework for high-accuracy privacy-preserving mining. *Data Min. Knowl. Discov.* **2008**, *18*, 101–139. [CrossRef]
40. Yang, M.; Zhu, T.; Xiang, Y.; Zhou, W. Density-Based Location Preservation for Mobile Crowdsensing with Differential Privacy. *IEEE Access* **2018**, *6*, 14779–14789. [CrossRef]
41. Ma, Q.; Zhang, S.; Zhu, T.; Liu, K.; Zhang, L.; He, W.; Liu, Y. PLP: Protecting Location Privacy Against Correlation Analyze Attack in Crowdsensing. *IEEE Trans. Mob. Comput.* **2017**, *16*, 2588–2598. [CrossRef]
42. Gao, H.; Xu, H.; Zhang, L.; Zhou, X. A Differential Game Model for Data Utility and Privacy-Preserving in Mobile Crowdsensing. *IEEE Access* **2019**, *7*, 128526–128533. [CrossRef]
43. Huai, M.; Huang, L.; Sun, Y.E.; Yang, W. Efficient Privacy-Preserving Aggregation for Mobile Crowdsensing. In Proceedings of the IEEE Big Data and Cloud Computing, Dalian, China, 26–28 August 2015; pp. 275–280.
44. Huang, P.; Zhang, X.; Guo, L.; Li, M. Incentivizing Crowdsensing-based Noise Monitoring with Differentially-Private Locations. *IEEE Trans. Mob. Comput.* **2019**, 1–14. [CrossRef]
45. Sweeney, L. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. Uncertainty Fuzziness Knowl. Based Syst.* **2002**, *10*, 571–588. [CrossRef]
46. Li, X.; Miao, M.; Liu, H.; Ma, J.; Li, K.C. An incentive mechanism for K-anonymity in LBS privacy protection based on credit mechanism. *Soft Comput.* **2017**, *21*, 3907–3917. [CrossRef]
47. Oganian, A.; Domingo-Ferrer, J. Local synthesis for disclosure limitation that satisfies probabilistic k-anonymity criterion. *Trans. Data Priv.* **2017**, *10*, 61–81.
48. Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M.; Kifer, D.; Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. *ACM TKDD* **2007**, *1*, 3. [CrossRef]

49. Li, N.; Li, T.; Venkatasubramanian, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In Proceedings of the IEEE ICDE, Istanbul, Turkey, 15–20 April 2007; pp. 106–115.
50. Suliman, A.; Otrok, H.; Mizouni, R.; Singh, S.; Ouali, A. A greedy-proof incentive-compatible mechanism for group recruitment in mobile crowd sensing. *Future Gener. Comput. Syst.* **2019**, *101*, 1158–1167. [CrossRef]
51. Saadatmand, S.; Kanhere, S.S. MRA: A modified reverse auction based framework for incentive mechanisms in mobile crowdsensing systems. *Comput. Commun.* **2019**, *145*, 137–145. [CrossRef]
52. Wu, Y.; Li, F.; Ma, L.; Xie, Y.; Li, T.; Wang, Y. A Context-Aware Multiarmed Bandit Incentive Mechanism for Mobile Crowd Sensing Systems. *IEEE Internet Things J.* **2019**, *6*, 7648–7658. [CrossRef]
53. Abououf, M.; Otrok, H.; Singh, S.; Mizouni, R.; Ouali, A. A Misbehaving-Proof Game Theoretical Selection Approach for Mobile Crowd Sourcing. *IEEE Access* **2020**, *8*, 58730–58741. [CrossRef]
54. Zhou, T.; Cai, Z.; Wu, K.; Chen, Y.; Xu, M. FIDC: A framework for improving data credibility in mobile crowdsensing. *Comput. Netw.* **2017**, *120*, 157–169. [CrossRef]
55. Xie, K.; Li, X.; Wang, X.; Xie, G.; Xie, D.; Li, Z.; Wen, J.; Diao, Z. Quick and Accurate False Data Detection in Mobile Crowd Sensing. In Proceedings of the IEEE INFOCOM, Paris, France, 29 April–2 May 2019; pp. 2215–2223. [CrossRef]
56. Zhang, M.; Yang, L.; Gong, X.; Zhang, J. Privacy-Preserving Crowdsensing: Privacy Valuation, Network Effect, and Profit Maximization. In Proceedings of the IEEE GLOBECOM, Washington, DC, USA, 4–8 December 2016; pp. 1–6.
57. Capponi, A.; Fiandrino, C.; Kantarci, B.; Foschini, L.; Kliazovich, D.; Bouvry, P. A Survey on Mobile Crowdsensing Systems: Challenges, Solutions, and Opportunities. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2419–2465. [CrossRef]
58. Liu, Y.; Kong, L.; Chen, G. Data-Oriented Mobile Crowdsensing: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2849–2885. [CrossRef]
59. Pouryazdan, M.; Kantarci, B.; Soyata, T.; Foschini, L.; Song, H. Quantifying user reputation scores, data trustworthiness, and user incentives in mobile crowd-sensing. *IEEE Access* **2017**, *5*, 1382–1397. [CrossRef]
60. Pouryazdan, M.; Fiandrino, C.; Kantarci, B.; Soyata, T.; Kliazovich, D.; Bouvry, P. Intelligent Gaming for Mobile Crowd-Sensing Participants to Acquire Trustworthy Big Data in the Internet of Things. *IEEE Access* **2017**, *5*, 22209–22223. [CrossRef]
61. Xiao, L.; Li, Y.; Han, G.; Dai, H.; Poor, H.V. A Secure Mobile Crowdsensing Game with Deep Reinforcement Learning. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 35–47. [CrossRef]
62. Agrawal, R.; Srikant, R.; Thomas, D. Privacy preserving OLAP. In Proceedings of the ACM SIGMOD, Baltimore, MA, USA, 14–16 June 2005; pp. 251–262.
63. Wang, W.; Carreira-Perpiñán, M.A. Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. *arXiv* **2013**, arXiv:1309.1541v1.
64. Tang, J.; Korolova, A.; Bai, X.; Wang, X.; Wang, X. Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12. *arXiv* **2017**, arXiv:1709.02753.
65. Differential Privacy Team Apple. Learning with Privacy at Scale. *Apple Mach. Learn. J.* **2017**, *1*, 1–25.
66. Wang, L.; Zhang, D.; Yan, Z.; Xiong, H.; Xie, B. EffSense: A Novel Mobile Crowd-Sensing Framework for Energy-Efficient and Cost-Effective Data Uploading. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *45*, 1549–1563. [CrossRef]
67. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [CrossRef]
68. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional Neural Networks for human activity recognition using mobile sensors. In Proceedings of the EAI International Conference on Mobile Computing, Applications and Services (MobiCASE), Berlin, Germany, 12–13 November 2015; pp. 197–205.
69. Shen, W.; Guo, Y.; Wang, Y.; Zhao, K.; Wang, B.; Yuille, A. Deep Regression Forests for Age Estimation. In Proceedings of the CVPR, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2304–2313.
70. Chen, S.; Zhang, C.; Dong, M.; Le, J.; Rao, M. Using Ranking-CNN for Age Estimation. In Proceedings of the CVPR, Honolulu, HI, USA, 22–25 July 2017; pp. 5183–5192.
71. Lai, C.C.; Shih, T.P.; Ko, W.C.; Tang, H.J.; Hsueh, P.R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob. Agents* **2020**. [CrossRef]

72. Wu, J.T.; Leung, K.; Bushman, M.; Kishore, N.; Niehus, R.; de Salazar, P.M.; Cowling, B.J.; Lipsitch, M.; Leung, G.M. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **2020**, 1–5. [CrossRef]
73. Rothe, R.; Timofte, R.; Van Gool, L. DEX: Deep EXpectation of apparent age from a single image. In Proceedings of the ICCV Workshops, Santiago, Chile, 13–16 December 2015; pp. 10–15.
74. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the ICML, Haifa, Israel, 21–24 June 2010; pp. 807–814.
75. Fan, L. Image pixelization with differential privacy. In Proceedings of the IFIP Annual Conference on Data and Applications Security and Privacy (DBSec), Bergamo, Italy, 16–18 July 2018; pp. 148–162. [CrossRef]