energies

MDPI

Article

The Location Privacy Protection of Electric Vehicles with Differential Privacy in V2G Networks

Yuancheng Li [1,*], Pan Zhang [1,2] and Yimeng Wang [1]

[1] School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China; pan-zhang@sgcc.com.cn (P.Z.); 1162227078@ncepu.edu.cn (Y.W.)
[2] State Grid Information & Telecommunication Branch, Beijing 100761, China
* Correspondence: yuancheng@ncepu.edu.cn

Received: 10 September 2018; Accepted: 28 September 2018; Published: 1 October 2018

Abstract: Vehicle-to-grid (V2G) is an important component of smart grids and plays a significant role in improving grid stability, reducing energy consumption and generating cost. However, while electric vehicles are being charged, it is possible to expose the location and movement trajectories of the electric vehicles, thereby triggering a series of privacy and security issues. In response to this problem, we propose a new quadtree-based spatial decomposition algorithm to protect the location privacy of electric vehicles. First of all, we use a random sampling algorithm, which is based on differential privacy, to obtain enough spatial data to achieve the balance between large-scale spatial data and the amount of noise. Secondly, in order to overcome the shortcomings of using tree height to control Laplacian noise in the quadtree, we use sparse vector technology to control the noise added to the tree nodes. Finally, according to the vehicle-to-grid network structure in the smart grid, we propose a location privacy protection model based on distributed differential privacy technology for EVs in vehicle-to-grid networks. We demonstrate application of the proposed model in real spatial data and show that it can achieve the best effect on the security of the algorithm and the availability of data.

Keywords: electric vehicle (EV); location privacy protection; differential privacy; random sampling algorithm; sparse vector technology; vehicle to grid (V2G)

1. Introduction

Vehicle-to-grid (V2G) is an important sub-system of smart grids. With the characteristics of electric vehicles charging and discharging, a vehicle-to-grid network can help grid load to "peak-fill", improve grid stability, reduce energy consumption and reduce power generation cost [1]. Vehicle-to-networks is also suitable for some small-scale energy management systems [2]. As we all know, the range of charging pile locations can affect the degree of participation. In the case of high participation [3], the total cost of the energy system will decrease [4], and electricity prices will fall [5]. However, electric vehicle-to-grid accelerates transformer aging [6]. Therefore, vehicle-to-grid is a double-edged sword [7]. At the same time, the V2G network has also introduced new privacy issues, such as the user's home address, place of work, place of entertainment, and places frequented, which may be reflected in the charging history. Leaking location information has a negative impact on users, being harassed by location-based spam, unscrupulous merchants selling location-related products or services to users without permission [8]. In addition, the leaked location information may also expose the user's health status, religious beliefs, personal preferences, social relations and other private information [9]. For example, the stay period and frequency of visits of an electric car at a hospital may expose the user's health condition. More seriously, location information may also cause security problems, which might be used by criminals, allowing users to be tracked, looted, and even suffer from personal

attacks [10]. Therefore, the privacy of EV location is very important to the normal operation of smart grid, the safety of electric vehicle users and the popularization of electric vehicles. Therefore, it is important to study the privacy protection of the EV location in vehicle-to-grid networks.

The control center of the power grid optimizes the charging and discharging of the electric vehicle by monitoring the position of the electric vehicle. When performing location data query and access, the spatial search tree or grid structure that meets the requirements is usually established based on the spatial segmentation technique. There is a lot of research work based on the privacy protection scheme proposed by the spatial segmentation technology, such as adding noise to the established spatial search tree or network unit, and disturbing the location of the individual. As the degree of noise is added, the privacy protection of individual locations is better, but the accuracy of search and query is also reduced. At the same time, most privacy protection algorithms based on spatial segmentation technology are mostly affected by data distribution. When the amount of data is relatively large, or the data skew is serious, the accuracy and privacy protection effect of the algorithm are very limited. Among the problems we studied, the amount of location data in the V2G network is huge, and the original spatial segmentation algorithm does not have a good effect on accuracy and data availability. Therefore, in this paper, we proposed a spatial data privacy protection algorithm for V2G networks, adding noise that satisfies differential privacy in the spatial segmentation algorithm, ensuring the security of individual location data while keeping data query and access with good precision.

In summary, we make the following contributions in this paper:

(1) We propose a new random sampling algorithm based on differential privacy, which can obtain enough samples to deal with large-scale spatial data.
(2) We propose a new quadtree-based spatial decomposition algorithm, and use sparse vector technology to control the depth of the tree to achieve independent tree-depth noise control and better protect the privacy of the location data.
(3) We propose the vehicle-to-grid location data protection model based on differential privacy to realize the privacy protection of EVs locations in vehicle-to-grid networks.
(4) We conduct experiments on actual EVs locations data to prove our proposed method and to achieve the best effect on the security of algorithm and availability of data.

The rest of our paper is organized as follows: in Section 2 we introduce the related work of privacy protection of vehicle-to-grid in a smart grid; in Section 3 we introduce the network structure of vehicle-to-grid in a smart grid; in Section 4 we propose the spatial data decomposition method with differential privacy; in Section 5 we implement experiments to verify the validity of the algorithm; and in Section 6 we conclude our paper.

2. Related Work

In recent years, many researchers have proposed some protocols to protect the privacy of electric vehicles. Based on the characteristics of vehicle-to-grid networks, Yang et al. first proposed the privacy issue of electric vehicle users, and proposed a protection privacy communication with an accurate reward system structure [11]. In this architecture, the user's "permission" generated by the ID-based restricted partially blind signature technology can access the V2G network anonymously, so as to protect the user's identity and location privacy. Each time the user sells electricity, he can obtain the "reward" signature, and according to this "reward", the corresponding reward can be obtained anonymously. However, it has the problem of key escrow in this scheme [12], which proposes a new scheme using the restricted partial blind signature technique under the setting of certificate-free public key. Reference [13] protects the location privacy of charging users by constantly changing their fake identities to ensure that electric vehicles can change different fake identities in different parking lots. A secure electric vehicle payment system was proposed to support two-way anonymous payment while still paying the right fee or getting paid [14]. Their scheme can guarantee anonymity, while supporting the function of tracking, fraud prevention and arbitration. Reference [15]

analyzed the impact of honest but curious aggregators on the privacy of EV users and proposed a model to prevent aggregators from tracking users by reducing the amount of data transferred. Reference [16] studied the privacy problem when users use advanced measurement systems to participate in vehicle-to-grid at home, and proposed the vehicle-to-grid architecture to hide the user's pseudo-identity information in K gateways to realize the privacy protection of user data. These methods all use pseudo-identities [13–16] for privacy protection. In this paper, we proposed a location privacy protection algorithm based on differential privacy. We obtained good effects only by processing location data. Unlike pseudo-identities technology, we do not need extra computational overhead when calculating pseudo-identities.

On the other hand, many researches on spatial decomposition technology are based on differential privacy to protect the location of spatial data privacy. Spatial decomposition technology commonly uses indexing technology, such as grid structure and tree structure. The HKD-tree proposed in Reference [17] is an early representative of data-independent decomposition. This method divides the data by grid and adds noise to each grid cell, which utilizes KD-tree for indexing. However, this algorithm is valid only if the data distribution is balanced. The uniform-grid (UG) model in Reference [18] uses a well-distributed grid to decompose two-dimensional spatial data and adds noise to each of the cells. Although UG can be more reasonable to set the granularity of the division, it does not consider skew and sparseness of data distribution. If a cell is especially sparse, even a count of zero, it will result in excessive noise error; on the other hand, if a cell is especially dense, it cannot be completely divided and will result in an excessive assumption error. The DP-where method in Reference [19] also uses a well-distributed grid to decompose the working position and family position of a moving crowd, but the disadvantages of this method are similar to that of UG. For the lack of UG method, adaptive-grid (AG) model was proposed [18] according to the different granularity of the high-level division unit, to divide the spatial data adaptively top-down. Although AG can set the granularity of spatial data adaptively according to the data sparseness, it does not give the corresponding heuristic rules to distinguish the boundaries between the dense and sparse data. In addition, this method does not consider the actual distribution of the original data either. Reference [20] proposed the AG method adapted for spatial decomposition, and then utilizing the Laplace mechanism to protect the worker's position information. A complete quadtree is used to decompose the two-dimensional spatial data top-down in Reference [21]. The complete quadtree needs to satisfy all leaf-to-root paths with the same length, and all intermediate nodes have the same fan-out. In order to improve the decomposition accuracy, quadpost uses the geometric distribution technology to divide the privacy cost and post-processes the noise by the least squares unbiased estimation. The advantage of this method is that the privacy budget can be rationally distributed, and the noise error is low. The disadvantage is that the depth of the tree is used to control the noise value. If the depth of the tree is relatively large, the noise added in each layer is especially high. Accordingly, the final query accuracy will be low. In addition, this method does not consider the original data distribution, and the uniform hypothesis error is relatively high. Quadtree and Kalman filtering were used to decompose dynamic spatial data in Reference [22]. It utilizes a heuristic threshold to judge whether each partitioned unit is sparse or dense. If the unit is still dense, it will be partitioned continually. The disadvantages of Reference [22] are similar to those of Reference [21], depending on the depth of tree to control the noise value. Compared to the first two methods, [23] combined complete quadtree to partition spatial data, responding to range queries by releasing leaf node noise counts and non-leaf node domain information. This method does not rely on the tree depth. It reduces the noise through the offset value of the node count, and then uses a noise constant to determine whether to divide the node. Meanwhile, this method uses the sparse vector technique [24,25] to calculate the node decomposition threshold. The b-ary tree is used to partition the data levelly in Reference [26]. The noise is used to perturb the counts in each node, and the statistical information of each layer is published as a histogram. However, this method also uses tree depth to control noise. Reference [27] uses b-ary tree to decompose the data as well. The method discusses the relationship between tree depth, tree

fan-out and data dimension, and post-processes the query results. Reference [28] used the sampling method to process the spatial data, and then divided the spatial counts into groups of the same size, adding noise to the mean of each group. However, the final accuracy of the method was relatively low. DP-tree method using embedded trees was proposed to decompose multi-dimensional spatial data and supports range count queries [29]. However, this method uses tree depth to control noise, which is easily affected by tree fan-out. The decomposition method in References [30–32] considers the actual distribution of the underlying spatial data, and divides data according to the actual position of the spatial data points. However, these algorithms must be carried out under the protection of differential privacy, otherwise it will reveal the privacy of the underlying data.

In conclusion, most of the spatial segmentation are affected by the actual data distribution. The tree depth is usually used to control the Laplacian noise level, which leads to high computation overhead and low availability. These methods do not consider well how to balance the noise error and the uniform hypothesis error. Some decomposition methods, although taking into account the above two kinds of error balance, did not consider how to use heuristic rules to adaptively set the equalization parameters. When the count of spatial data reaches millions, these methods usually cannot obtain accurate results. Although the above methods are able to give rigorous data availability theory bounds, they did not perform well in data availability and efficiency on the actual data. In the vehicle-to-grid network, electric vehicle location data is millions-level. The methods of spatial decomposition protect the location privacy, usually leading to the availability of location data being especially low. In this paper, we adopt the sampling algorithm based on differential privacy to achieve the balance between large-scale spatial data and noise volume. To overcome the shortcomings of Laplacian noise controlled by tree-depth in quadtree, we utilize sparse vector techniques to control when to partition the tree node. Based on the vehicle-to-grid network structure in the smart grid, we propose a location privacy protection model for electric vehicles in vehicle-to-grid networks adopting distributed differential privacy technology.

3. Network Structure of V2G in Smart Grid

Vehicle-to-grid is a system that serves the energy interaction between electric vehicles and the grid. Electric vehicles want to be able to get power when the grid load is at its nadir, and feed power back to the grid when the grid load is at its peak [1]. At the same time, it hopes that the electric grid will feed the electric energy back when the load is at its peak. When the electric vehicle and the grid are in an energy interaction, they must establish real-time communication for the transmission of relevant information, such as the status of the electric vehicle and the load of the electric network [10]. Therefore, the main activity in vehicle-to-grid is actually the two-way interaction related to energy and information between EVs and the grid [8].

The vehicle-to-grid system is mainly concentrated in the distribution domain. In the vehicle-to-grid network, a large number of electric vehicles, charging stations and parking lots jointly construct a bidirectional power and communication network through a power distribution network and a communication network, as shown in Figure 1.

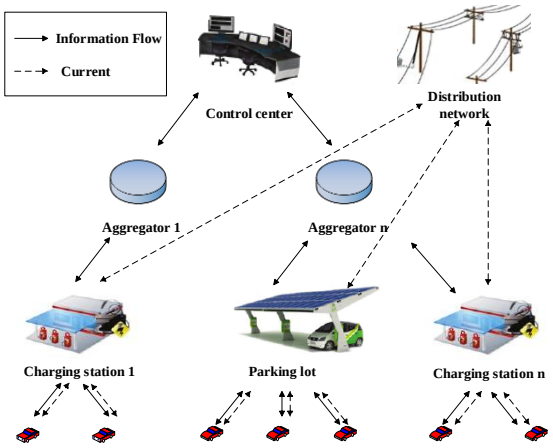


Figure 1. Network structure of V2G in smart grid.

(1) Control center: Control center is the most important component of the smart grid, solving the dispatching and control problems of electric vehicles after they are connected to the grid. It is an indispensable "brain" of the grid operation.

(2) Aggregators: On the one hand, aggregators can receive vehicle-to-grid service requests from the smart grid control center to provide feedback-related information to the smart grid. On the other hand, aggregators can gather vehicle-to-grid business services from the smart grid control center after aggregating the information of multiple EVs, and then completing the subsequent related resource scheduling.

(3) Distribution network: The distribution network is composed of overhead lines, cables, towers, distribution transformers, isolation switch, reactive power compensators and some ancillary facilities. It plays an important role in the distribution of power in the power grid and distributes the electric energy to the electric vehicles in the vehicle to grid network.

(4) Charging station & charging parking lot: Charging stations and charging parking lots provide electric vehicles with supplementary electric energy, in which there are many charging piles. The input end is directly connected with the AC grid, and the output end is equipped with a charging plug for charging electric vehicles.

(5) Electric vehicle: Electric vehicles are powered by on-board power and are equipped with on-board battery packs. The batteries of a large number of EVs form a distributed, mobile power warehouse that can be used to help the grid "fill the valley" (electric vehicles charge at night) during down periods and "cut the peak" (electric cars discharge during the day) during peak periods.

The vehicle-to-grid system brings great economic benefits, social benefits and ecological benefits to people, meanwhile, it also has the potential to leak users' privacy. In order to meet the requirements of power load adjustment (usually several million kilowatt hours), the vehicle to grid system must ensure that a sufficient number of EVs are provided as energy storage resources within a given period of time. Therefore, as shown in Figure 1, a certain number of EVs must be aggregated through an aggregator, and monitor the related information of the EVs, such as the location of EVs, the state of charge of the batteries, the expected departure time, and the real-time capacity of chargeable and dischargeable, etc. so that the control center can optimally schedule the load requirements of the grid on the basis of the EVs charge and discharge.

At the same time, the aggregators transfer the collected information to the control center. At this point, if the original location data is uploaded, the control center can trace the user's whereabouts and analyze the user's privacy information. In this paper, we mainly discuss the issue of privacy protection of user location in vehicle-to-grid networks in this situation.

4. Location Privacy Protection Algorithm with Differential Privacy

According to the characteristics of EV location data in V2G network of smart grid and the shortage of existing spatial decomposition algorithms, we propose a spatial data decomposition algorithm with privacy protection in this chapter. Based on this, we propose a location privacy protection model of EVs in the V2G network.

4.1. Data Preprocessing with Differential Privacy

Existing methods of spatial decomposition often deal with small-scale spatial data. However, spatial data in V2G are usually large-scale and skewed. This often results in tree-based decomposition methods that cannot be implemented or the availability of final query or analysis results are very low. Therefore, how to decompose large-scale and skewed spatial data is a very big challenge. Therefore, we take as many samples as possible with the sampling technique that satisfies the differential privacy and spatially partition the samples for solving the problem.

4.1.1. Differential Privacy

Differential privacy means that one queries two different data sets with only one record different; if query results are almost identical, the attacker cannot obtain the data of the individual by analyzing the query results. This can achieve privacy protection. Assuming there are two datasets with only one record different, the ratio of probabilities that query results on both datasets is close to 1, which achieves differential privacy protection.

Definition 1. *An algorithm A satisfies ε-differential privacy if, for any two neighboring datasets D and D′ and for any possible output O of A, where Pr*[•] *denotes the probability of an event*.

$$\ln\left(\frac{Pr(A(D)=O)}{Pr[A(D')=O]}\right) \leq \varepsilon \tag{1}$$

Definition 2. *Let f be a function that maps a dataset D into a vector of real numbers. The global sensitivity of f is defined as*

$$S(f) = \max_{D,D'} \| f(D) - f(D') \|_1 \tag{2}$$

where D and D' are any two neighboring datasets, and $\| \cdot \|_1$ denotes the L_1 norm.

Lemma 1. *Let* $A_1, \ldots, A_k$ *be k algorithms, such that* A_i *satisfies* ε_i*-differential privacy* ($i \in [1, k]$). *Then, for the same dataset, the sequential composition* $(A_1, \ldots, A_k)$ *satisfies* $(\sum_{i=1}^{k} \varepsilon_i)$*-differential privacy*.

Lemma 2. *Let* $A_1, \ldots, A_k$ *be k algorithms, such that* A_i *satisfies* ε_i-differential *privacy* ($i \in [1, k]$). *Then, for the different datasets, the sequential composition* $(A_1, \ldots, A_k)$ *satisfies* $(\max \varepsilon_i)$-differential privacy.

Theorem 1. *Let A satisfies ε-differential on dataset D. If take sample from D to get D′ with probability* γ, *algorithm A satisfies* $\ln(1 + \gamma(e^{\varepsilon} - 1))$*-differential on dataset D′*.

Theorem 2. *An algorithm* A *satisfies ε-differential privacy if,*

$$A(D, d_i) = \left\{ d_i : \left| \Pr[d_i \in \Omega] \propto \exp\left(\frac{\varepsilon u(D, d_i)}{2\Delta u}\right) \right. \right\} \tag{3}$$

where Δu is the global sensitivity of $u(D, d_i)$, which is a scoring function, d_i is the output from the output domain Ω.

4.1.2. Bernoulli Random Sampling Algorithm Based on Differential Privacy

For the problem of large-scale spatial data, our proposed decomposition method tries to extract sufficient data as the decomposition data under the conditions of differential privacy. The Bernoulli random sampling algorithm that satisfies differential privacy will be introduced in detail as Algorithm 1.

Algorithm 1 Random Sampling Algorithm with Differential Privacy (D, ε)

1 Obtain spatial data sample $\hat{D} = \left(\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_m\right)$ after implementing multiple Bernoulli experiments with probability γ;
2 Calculate $\varepsilon_\gamma = \ln(\gamma + e^\varepsilon - 1) - \ln\gamma$ on the basis of Theorem 1.

Firstly, we determine the sampling probability γ, and then make the Bernoulli experiment with γ on D. If the experiment is successful, obtain the spatial sample, otherwise, abandon the sample. Finally, calculate the privacy cost ε_γ required for the entire space decomposition. The key of the process is how to make the sampling process to meet the differential privacy. Since $\varepsilon_\gamma = \ln(\gamma + e^\varepsilon - 1) - \ln\gamma$, we bring ε_γ into $\ln(1 + \gamma(e^\varepsilon - 1))$, and obtain $\ln\left(1 + \gamma\left(e^{\ln(\gamma + e^\varepsilon - 1) - \ln\gamma} - 1\right)\right) = \varepsilon$. So we can prove the proposed sampling process satisfies *ε-differential* privacy.

4.2. Spatial Decomposition Algorithm BQ-Tree

By studying the existing spatial decomposition algorithms, we know that the existing algorithms do not work well in dealing with millions of spatial data. In this part, we combine the proposed random sampling algorithm with the quadtree algorithm, and propose a new spatial decomposition algorithm BQ-tree. It can overcome the problem that the traditional quadtree algorithm cannot deal with a huge number of spatial data. We also prove that the BQ-tree algorithm satisfies the differential privacy.

4.2.1. The Quadtree Algorithm

The specific algorithm of quadtree is as Algorithm 2. It has four input parameters, which are: (1) a dataset D of spatial data distributed in a multidimensional domain Ω, (2) the Laplacian noise of size ε added to the tree, (3) the threshold θ of splitting node in the tree, (4) the threshold h of the maximum height of the tree. The algorithm returns a quadtree, each node contains two parts of information, namely the sub-domain corresponding to v, and the value of the number of spatial data in the sub-domain which is added noise. At the same time, the depth of v is defined as the maximum distance between v and the root node. It is recorded as depth *(v)*.

4.2.2. The BQ-Tree Algorithm

Compared with the traditional quadtree algorithm, the proposed algorithm first initializes the quadtree based on the sampled dataset, and calculates the size of added noise. BQ-tree specific establishment process is as Algorithm 2:

Algorithm 2 BQ-Tree (D, ε, θ, h)

1 Compute ε_γ and $\hat{D}$ on basis of Algorithm 1;
2 initialize a quadtree T with a root node v_1 on dataset $\hat{D}$, and mark v_1 as unvisited;
3 **while** there exists an unvisited node v **do**
4 mark v as visited;
5 compute the number c(v) of data points that are contained in $dom(v)$;
6 compute a noisy version of c(v): $\hat{c}(v) = c(v) + Lap(\varepsilon)$;
7 if $\hat{c}(v) > \theta$ **and** $depth(v) < h - 1$ **then**
8 split v, and add its children to T;
9 mark the children of v as unvisited;
10 **return T**

The algorithm starts by computing the privacy cost ε_γ and sample dataset $\hat{D}$ by using Algorithm 1 proposed in Section 4.1.2. Next, the quadtree is initialized on $\hat{D}$ and the root node is set to unvisited. The subsequent part of the algorithm consists of a number of iterations. In each iteration, we examine if there are unvisited nodes in the tree. If such v exists, we mark the nodes as visited state, calculate the number of space points, and add ε_γ noise to the number. After that, we split v if the following two conditions simultaneously hold. One of the conditions is that the number $\hat{c}(v)$ is greater than the threshold of decomposition θ and the other is the height of the tree is less than the height threshold h of the tree. If both of the above conditions are met, then we generate v's children and insert them into T as unvisited nodes. Finally, when all node have been visited, we return the quadtree.

According to Section 4.1.2, we know that the sampling process satisfies the differential privacy. To prove that the overall BQ-tree algorithm satisfies the differential privacy, we only need to prove that step 6 in Algorithm 2 satisfies the differential privacy. Step 6 adds a noise of $Lap(\varepsilon_\gamma)$ size to each node because the count up to h nodes is affected when adding or removing a data point in D. Combining the differential privacy Lemma 1 and Lemma 2, Step 6 satisfies ε_γ-*differential* privacy. Then, according to Theorem 1, the proposed algorithm satisfies ε-*difference* privacy.

4.3. Spatial Decomposition Algorithm BQ-Tss

The existing tree-based spatial decomposition methods usually adopt tree depth to control Laplacian noise. However, it is very difficult to set a proper tree depth. If artificially directly adjust the depth of the tree, the adjustment process will violate the differential privacy and thus the sensitive information in the spatial data cannot be protected. If we can add noise for the nodes in the tree without depending on the depth of the tree to control the noise, this will control the added noise better. In this section, we use sparse vector techniques to set the decomposition conditions of the nodes in the tree to solve the problem of tree-depth dependent noise control.

4.3.1. Sparse Vector Technology

Sparse vector technique is commonly used to respond to a limited number of count queries greater than a certain threshold. The technique consists of two main steps: One is to find a suitable threshold θ and obtain $\widetilde{\theta}$ after adding noise; the other one is to obtain $\widetilde{c}(v)$ after adding noise to each query result $c(v)$ and compare it with the noise threshold. One of comparison results is to output $\widetilde{c}(v)$ if = $c(v) + Lap\left(\frac{2}{\varepsilon_1}\right) \geq \hat{\theta}$, otherwise an identifier $\perp$ is output. The specific application of sparse vector technology in our work is shown as Equation (4):

$$\hat{c}(v) = \begin{cases} c(v) + Lap\left(\frac{2}{\varepsilon_1}\right) if c(v) + Lap\left(\frac{2}{\varepsilon_1}\right) \geq \hat{\theta} \\ \perp others \end{cases} \tag{4}$$

4.3.2. The BQ-Tss Algorithm

In order to overcome the shortcomings of the original algorithm, which control noise dependent on the depth of tree, we combine the sparse vector technique with the algorithm shown in Section 4.2.2, and propose a new spatial decomposition algorithm BQ-Tss. The specific algorithm is shown as Algorithm 3.

Algorithm 3 BQ-Tss (D, ε, θ)

Compute ε_γ and $\hat{D}$ on basis of Algorithm 1;
initialize a quadtree T with a root node v_1 on dataset $\hat{D}$, and mark v_1 as unvisited;
while there exists an unvisited node v **do**
  mark v as visited;
  compute the number c(v) of data points that are contained in $dom(v)$;
  compute noise threshold $\hat{\theta} = \theta + Lap\left(\frac{2}{\varepsilon_\gamma}\right)$;
  compute a noisy version of c(v): $\hat{c}(v) = c(v) + Lap\left(\frac{2}{\varepsilon_\gamma}\right)$;
  if $\hat{c}(v) > \hat{\theta}$ **and** v is not the leaf **then**
    split v, and add its children to T;
    mark the children of v as unvisited;
return T

In the vehicle-to-grid network, there exists large-scale skewed data, which most existing spatial decomposition algorithms cannot handle. In BQ-Tss, we use the privacy-based random sampling algorithm, which is proposed in Section 4.1, to solve this problem. We use random sampling to obtain enough electric vehicles' position data from raw data, which represents the overall data distribution. The selected data can be used to replace raw data. Then, we perform spatial decomposition on the selected data.

Tree-based spatial decomposition algorithms usually control the Laplace noise by the depth of the tree, which is very difficult to determine. In BQ-Tss, we use a different way to control the size of noise. The method is based on sparse vector technology. When adding noise to the nodes of tree, it can provide the appropriate size of noise without depending on the depth of the tree. Therefore, compared with other spatial segmentation algorithms, BQ-Tss can handle large-scale skew data in V2G, and it can ensure the addition of appropriate scale noise, while realizing the protection of location data privacy.

According to Algorithm 3, the BQ-Tree satisfies the differential privacy. The newly proposed algorithm uses extra privacy cost only on the SVT step. Therefore, BQ-Tss can be inferred to satisfy ε-*differential* privacy as long as it is proved that SVT satisfies ε_γ-*differential* privacy.

In BQ-Tss algorithm, we use SVT technology to determine whether the tree node should be divided, and this seems likely to judge yes or no. Therefore, to prove conveniently, we use the binary vector $V = \langle x_1, x_2, \ldots, x_t \rangle$ to record whether the nodes should be divided. If $\hat{c}(v) > \hat{\theta}$, then $x_i = 1$, that represents node v_i is divided; otherwise $x_i = 0$, which represents node v_i is not divided and v_i is a leaf node. Given the two adjacent spatial datasets D and D', $\Pr_1(v)$ and $\Pr_2(v)$ denote the probabilities of SVT acting on D and D′ with the output of V, respectively. Let $x^{<i}$ denote the first $(i-1)$ responses in vector V. We can conclude the distribution of $\hat{c}(v_i)$ and $\hat{\theta}$ satisfies Laplace distribution based on the Equation (5) shown as follows:

$$\frac{\Pr_1(v)}{\Pr_2(v)} = \frac{\prod_{i=1}^{t} \Pr_1\left(x_i = 1 or 0 \middle| x^{<i}\right)}{\prod_{i=1}^{t} \Pr_2\left(x_i = 1 or 0 \middle| x^{<i}\right)} = \prod \frac{\Pr_1\left(x_i = 1 \middle| x^{<i}\right)}{\Pr_2\left(x_i = 1 \middle| x^{<i}\right)} \times \prod \frac{\Pr_1\left(x_i = 0 \middle| x^{<i}\right)}{\Pr_2\left(x_i = 0 \middle| x^{<i}\right)} \tag{5}$$

We suppose that x is a decomposition threshold on D, let $H_i(x)$ denote the probability of $x_i = 1$ on D, and $H_i'(x)$ denote the probability of $x_i = 1$ on D'. Therefore, $H_i(x)$ can be described as the following Equation (6), $Lap(\lambda)$ represents the independent noise generated by the Laplace distribution.

$$H_i(x) = \Pr\left(x_i = 1 \middle| x^{<i}\right) = \Pr\left(\hat{c}(v_i) \geq x \middle| x^{<i}\right) = \Pr\left(Lap(\lambda) + c(v_i) \geq x \middle| x^{<i}\right) \tag{6}$$

Based on the Laplace distribution, $\lambda = \frac{2}{\varepsilon_1}$ and $f(y:u,\lambda) = \frac{1}{2\lambda}\exp\left(-\frac{|y-u|}{\lambda}\right)$, $H_i(x)$ can be represented as Equation (7):

$$H_i(x) = \int_x^{\infty} f\left(y : c(v_i), \frac{2}{\varepsilon_1}\right) dy \tag{7}$$

According to the global sensitivity Definition 2, the sensitivity of the count $c(v)$ is 1, that is $\Delta = 1$. Let $c(v)$ and $c'(v)$ denote the counts of nodes v_i on D and D' respectively, and $c'(v) = c(v) + 1$. Therefore, $\mu = \lambda + 1$. $H_i(x)$ can be expressed as the following Equation (8):

$$H_i(x) = \int_{x+1}^{\infty} f\left(u : c(v_i) + 1, \frac{2}{\varepsilon_1}\right) d\mu = \int_{x+1}^{\infty} f\left(u : c'(v_i) + 1, \frac{2}{\varepsilon_1}\right) d\mu = H'_i(x+1) \tag{8}$$

And $\prod \Pr_1\left(x_i = 1 \middle| x^{<i}\right)$ can be expressed as the following Equation (9):

$$\begin{array}{l} \prod \Pr_1\left(x_i = 1 \middle| x^{<i}\right) \\ = \int_{-\infty}^{\infty} Pr(x) \prod H_i(x) dx \\ = \int_{-\infty}^{\infty} Pr(x) \prod H'_i(x+1) dx \\ \leq \exp\left(\frac{\varepsilon_1}{2}\right) \int_{-\infty}^{\infty} Pr(x+1) \prod H'_i(x+1) dx \\ = \exp\left(\frac{\varepsilon_1}{2}\right) \int_{-\infty}^{\infty} Pr(x) \prod H'_i(x) dx \\ = \exp\left(\frac{\varepsilon_1}{2}\right) \prod \Pr_2\left(x_i = 1 \middle| x^{<i}\right) \end{array} \tag{9}$$

Similarly, we can obtain Equation (10) as follows:

$$\prod \Pr_1\left(x_i = 0 \middle| x^{<i}\right) \leq \exp\left(\frac{\varepsilon_1}{2}\right) \prod \Pr_2\left(x_i = 0 \middle| x^{<i}\right) \tag{10}$$

And the Equation (11) as follows:

$$\frac{\Pr_1(V)}{\Pr_2(V)} \leq \exp(\varepsilon_1) \tag{11}$$

So, we can obtain Equation (12) as follows:

$$\Pr_1(V) \leq \exp(\varepsilon_1) \times \Pr_2(V) \tag{12}$$

As we can see from the Definition 1 of differential privacy and Equation (12), SVT operation satisfies ε_γ-*differential* privacy. Therefore, BQ-Tss satisfies ε-*difference* privacy.

4.4. Location Protection Model in V2G Network

In the V2G network, a new network equipment is added between the electric vehicle and the smart grid control system, commonly referred to as an aggregator. V2G manager through the aggregator to monitor the location of the EV, charging status and other related information so that the manager can optimally schedule EVs' charge and discharge according to the load requirements of the grid. Each aggregator's connected charging equipment and electric vehicles can compose a small independent network. Therefore, if we can ensure the privacy of electric vehicles in each independent network is secure, the location privacy of electric vehicles in the whole V2G network can be guaranteed. In this section, we apply the distributed differential privacy technology to the location privacy protection algorithm proposed in Section 4.3. Combining with the specific V2G network structure, we propose a EVs location privacy protection model with differential privacy in the V2G network.

4.4.1. Distributed Differential Privacy

Theorem 3. *Let* $\varepsilon \in (0,1)$ *be arbitrary. For* $c^2 > 2\ln\left(\frac{1.25}{\delta}\right)$, *the gaussian mechanism with parameter* $\delta \geq \frac{c\Delta_2 f}{\varepsilon}$ is (ε, δ)*-differentially private*.

Assuming that all parties in a distributed structure are honest, if each party generates gaussian distribution noise with a standard deviation of $\delta_i = \frac{\delta}{\sqrt{t}}$ and t represents the number of participants who generate the noise, then the overall standard deviation of noise is σ, that is to say, the distributed structure satisfies the differential privacy of (ε, δ) as a whole.

4.4.2. EVs' Location Privacy Protection Algorithm with Differential Privacy

In the V2G network architecture proposed above, we assume that each aggregator is trustworthy. If we assign a privacy budget of $\delta_i = \frac{\delta}{\sqrt{t}}$ to each aggregator and t represents the number of aggregators, the overall V2G network location protection algorithm satisfies the differential privacy of (ε, δ).

We assign the appropriate privacy budget to each aggregator network based on distributed differential privacy. Combining the BQ-Tss algorithm proposed in Section 4.3, we propose EVs location protection algorithm with differential privacy in V2G network. The specific location protection algorithm is shown in Figure 2.

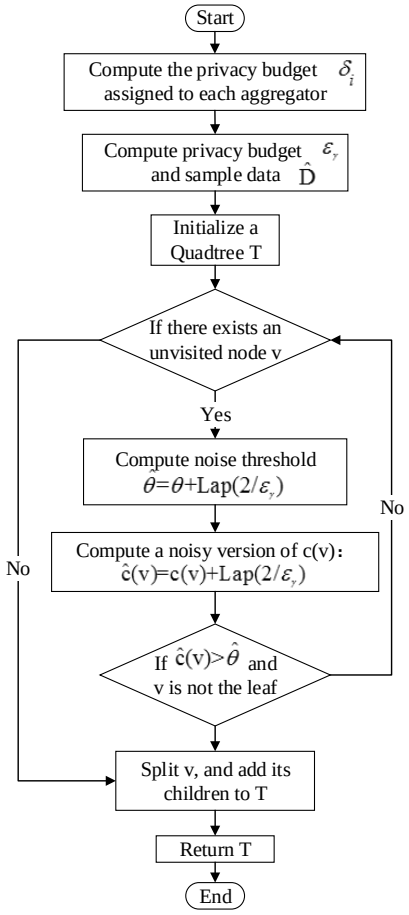


Figure 2. EVs Location Privacy Protection Algorithm with Differential Privacy.

5. Experimental Results and Analysis

Our experiments were conducted on an 8-core Intel i7-3612 CPU (2.10 GHz), 8G RAM, Win7 system platform and all algorithms were implemented in Python. The experiment used collected 1500

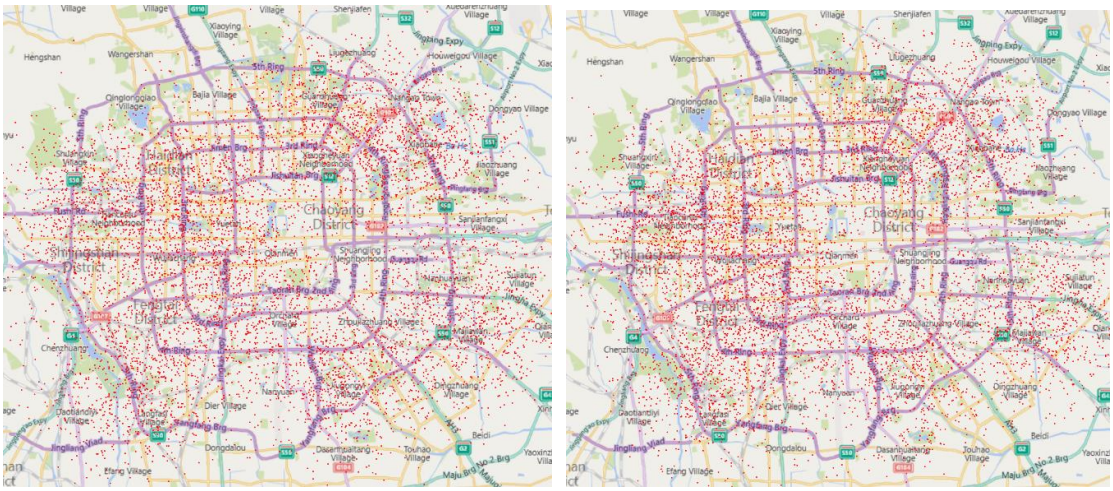
public charging posts in Beijing and the charging position of 100,000 electric vehicles in a certain week, with a total of 600,000 messages.

The relative error (RE) is used to measure the data availability of quardtree, BQ-tree, DP-tree and BQ-Tss algorithms. The relative error is calculated using Equation (13):

$$RE\left(\widetilde{Q}(D)\right) = \frac{\left|\widetilde{Q}(D) - Q(D)\right|}{\max\{Q(D), \Delta\}} \tag{13}$$

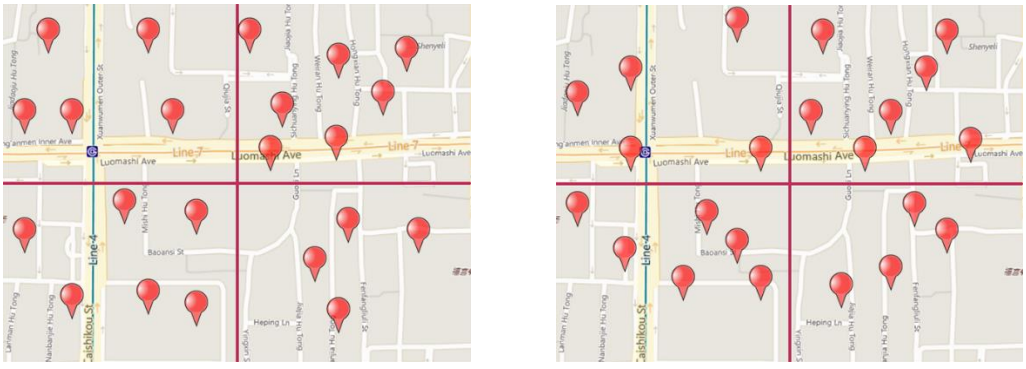
where $Q(D)$ denotes the true range query result on D and $\widetilde{Q}(D)$ denotes the noise result of the range query on D. Δ is a smoothing factor whose value is 0.1% of the experimental dataset size.

In this experiment, we set Bernoulli random sampling probability as 1% and privacy budget parameters ε values of 0.1, 0.3, 0.5, 0.7, 1.0 and 1.2. When ε is 0.5, we extract 16,000 position information from the data set. The overall position maps of the electric vehicle, before and after noise addition, are shown in Figure 3a,b, respectively. The local position maps are shown in Figure 4a,b, respectively.



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

Figure 3. The overall position maps of EVs. (**a**) before noise addition; (**b**) after noise addition.



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

Figure 4. The local position maps of EVs. (**a**) before noise addition; (**b**) after noise addition.

We can see from Figure 3a,b that, after adding noise, the change of the overall position distribution is slight. It indicates that the position data after noise addition is relatively accurate. If we use position data in Figure 3b instead of that in Figure 3a, the query of data can still achieve high accuracy. To better illustrate this point, we have partially enlarged the overall image to generate Figure 4a,b. In Figure 4a,b, we can clearly see that each position coordinate has a slight change after adding noise. Applying the spatial decomposition algorithm, we divide the small area into four parts. In each part, the number

of electric vehicles in Figure 4b is approximately equal to the number in Figure 4a. Therefore, when querying the position data in Figure 3b, the accuracy remains very high.

We set the range of query Q to cover 1%, 3%, 5%, 7%, 10% and 12% of the data set respectively, and 5000 queries are randomly generated within each query range.

We compare our BQ-Tss algorithm with quardtree, BQ-tree, and DP-tree algorithms. When the fixed-range query is fixed, we obtain the relative error of four different algorithms by changing the privacy budget ε value, and then compare the data availability of the each algorithm.

The figure above shows the query range results of EVs location dataset. As can be seen from Figures 5–7, when the query range is fixed to 1%, 5%, 10%, ε changes from 0.1 to 1.2, the accuracy of the proposed algorithm is almost three times that of the BQ-tree algorithm, and is nearly 10 times more than the DP-tree and quardtree algorithms. Especially when the query range is 1%, it is nearly 13 times more than the quardtree and DP-tree algorithm. From this we can conclude that as the range of query increases, the query accuracy of the algorithm is improved. When the query range is fixed, the query accuracy of the algorithm decreases with the increasing of the privacy budget. Under the same condition, the accuracy of our proposed algorithm is obviously better than that of quardtree and DP-tree, and is also superior to BQ-tree. When the query range is fixed at 10%, the algorithm can achieve an accuracy of 99.97%, which shows that the proposed algorithm is suitable for a big data environment and can guarantee the data availability under the condition of protecting the privacy of the location data.

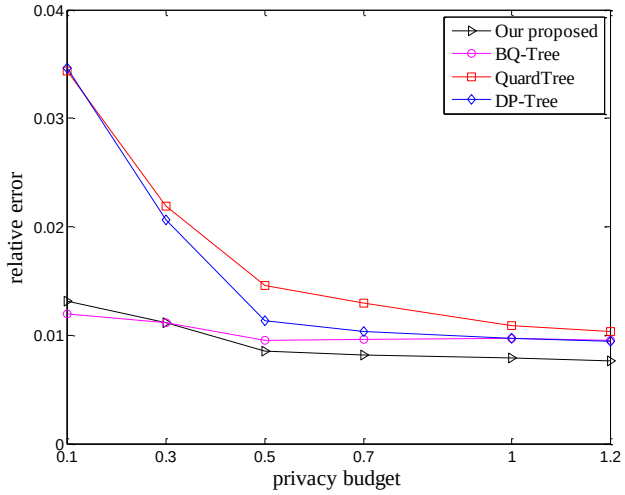


Figure 5. Query range Q = 1%.

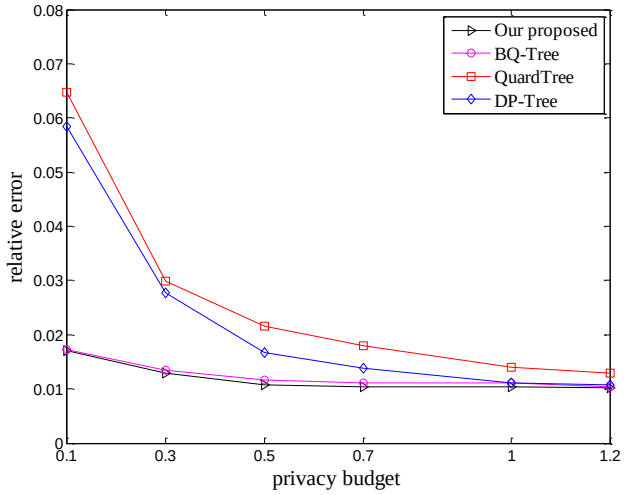


Figure 6. Query range Q = 5%.

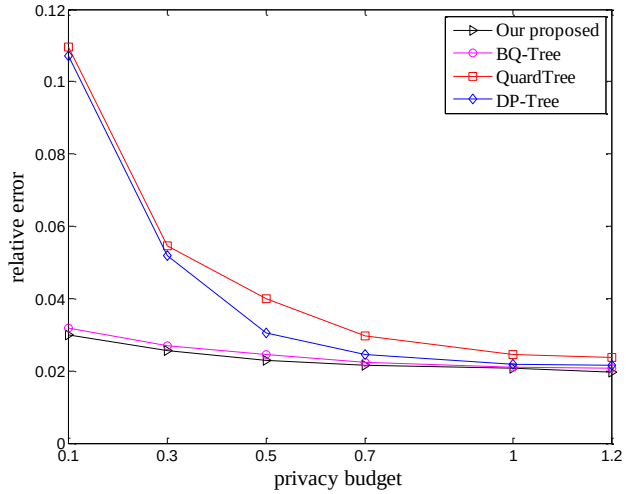


Figure 7. Query range Q = 10%.

Figures 8–10 shows that when the query range is changed from 1% to 12% and the ε is fixed to 0.1, 0.5, and 1.0 respectively, the query accuracy of BQ-Tss is nearly twice that of BQ-tree, which is nearly 12 times that of the quardtree and DP-tree algorithm. In addition, as shown in Figures 6–8, the BQ-tree algorithm is also superior to quardtree and DP-tree. From this, we can conclude that when the privacy budget is fixed, the accuracy of the algorithm decreases as the query scope increases. The accuracy of the proposed algorithm is obviously superior to quardtree, DP-tree and BQ-tree when the privacy budget is same. When the privacy budget is fixed at 1.0, the accuracy rate of 99.93% can be achieved, which shows that the sampling technique we use can avoid the skew problem of data distribution well. We also avoid multi-level decompose privacy cost by using SVT technology, which is obviously better than other similar algorithms. This can also prove, even in the case of unbalanced data distribution, that our method can still achieve high query accuracy.

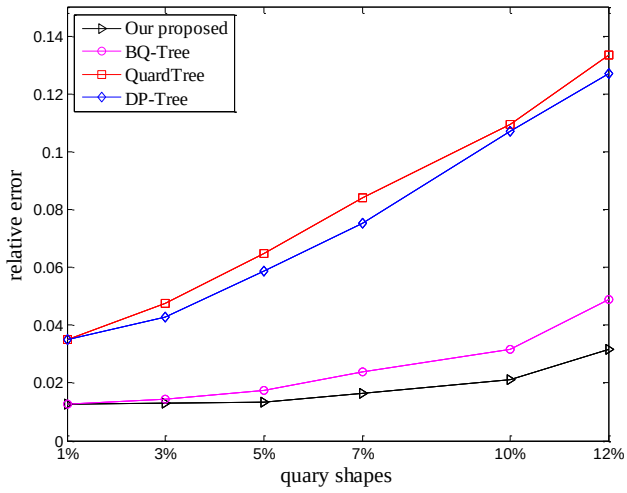


Figure 8. Privacy budget ε = 0.1.

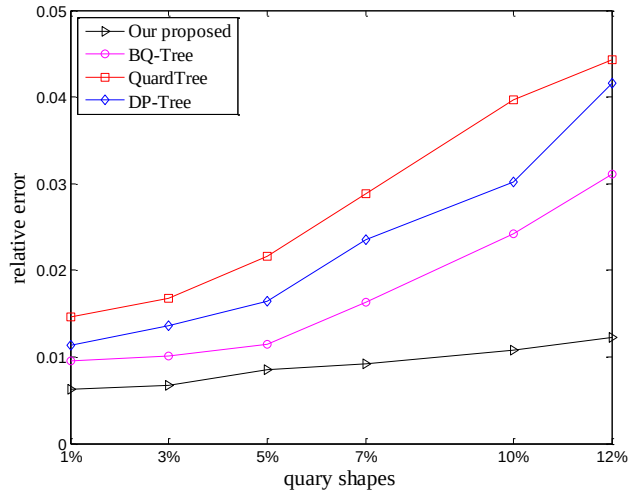


Figure 9. Privacy budget $\varepsilon = 0.5$.

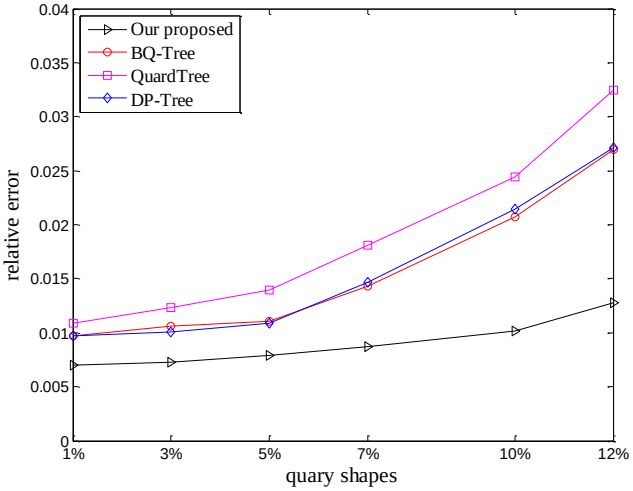


Figure 10. Privacy budget $\varepsilon = 1.0$.

6. Conclusions

The privacy-protected spatial decomposition algorithm for V2G networks needs to be able to handle large-scale skewed data in V2G, and add the appropriate noise to nodes in the tree without relying on the depth of the tree. In order to meet these requirements, we propose a spatial decomposition algorithm called BQ-Tss. Compared with other existing spatial decomposition algorithms, BQ-Tss algorithm achieves higher query accuracy. When adding the same noise, the error rate of BQ-tree is twice that of BQ-Tss, and quardTree and DP-tree are 10 times that of BQ-Tss. Therefore, our proposed algorithm can deal with large-scale skewed data better and get a higher precision query result. Under the slight noise of different scales, the accuracy of our query can be kept at around 99.9%. The proposed algorithm is based on the quad-tree spatial segmentation algorithm, including a random sampling algorithm with privacy protection and a method for segmenting nodes in a tree based on a sparse vector technique. Our proposed algorithm can overcome the shortcomings of the existing spatial segmentation algorithm and is more suitable for V2G networks. Therefore, our proposed algorithm can guarantee the normal control of the V2G network by the grid control center, while ensuring the privacy of the location data.

Author Contributions: Conceptualization, Y.L.; Methodology, Y.L., Y.W. and P.Z; Validation, Y.W. and P.Z.; Formal Analysis, Y.W. and P.Z.; Investigation, Y.W. and P.Z; Resources, P.Z.; Data Curation, Y.W.; Writing-Original Draft Preparation, Y.M.; Writing-Review & Editing, Y.L.; Visualization, Y.W.; Supervision, P.Z.; Project Administration, Y.L.

Funding: This research was supported by the Fundamental Research Funds for the Central Universities (2018ZD06).

Acknowledgments: This work was supported by the Fundamental Research Funds for the Central Universities (2018ZD06).

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

References

1. Harighi, T.; Bayindir, R.; Padmanaban, S.; Mihet-Popa, L.; Hossain, E. An overview of energy scenarios, storage systems and the infrastructure for vehicle-to-grid technology. *Energies* **2018**, *11*, 2174. [CrossRef]
2. Rottondi, C.; Fontana, S.; Verticale, G. Enabling privacy in vehicle-to-grid interactions for battery recharging. *Energies* **2014**, *7*, 2780–2798. [CrossRef]
3. Geske, J.; Schumann, D. Willing to participate in vehicle-to-grid (V2G)? Why not! *Energy Policy* **2018**, *120*, 392–401. [CrossRef]
4. Child, M.; Nordling, A.; Breyer, C. The impacts of high V2G participation in a 100% renewable Åland energy system. *Energies* **2018**, *11*, 2206. [CrossRef]
5. Nefedov, E.; Sierla, S.; Vyatkin, V. Internet of energy approach for sustainable use of electric vehicles as energy storage of prosumer buildings. *Energies* **2018**, *11*, 2165. [CrossRef]
6. Paterakis, N.G.; Pappi, I.N.; Erdinç, O.; Godina, R.; Rodrigues, E.M.G.; Catalão, J.P.S. Consideration of the impacts of a smart neighborhood load on transformer aging. *IEEE Trans. Smart Grid* **2016**, *7*, 2793–2802. [CrossRef]
7. Blasius, E. Possible role of power-to-vehicle and vehicle-to-grid as storages and flexible loads in the German 110 kV distribution grid. *Front. Energy* **2017**, *11*, 146–154. [CrossRef]
8. Aziz, M.; Oda, T.; Mitani, T.; Watanabe, Y.; Kashiwagi, T. Utilization of electric vehicles and their used batteries for peak-load shifting. *Energies* **2015**, *8*, 3720–3738. [CrossRef]
9. Zhang, Y.; Li, J.; Zheng, D.; Li, P.; Tian, Y. Privacy-preserving communication and power injection over vehicle networks and 5G smart grid slice. *J. Netw. Comput. Appl.* **2018**, *122*, 50–60. [CrossRef]
10. Roman, L.F.A.; Gondim, P.R.L.; Lloret, J. Pairing-based authentication protocol for V2G networks in smart grid. *Ad Hoc Netw.* **2018**. [CrossRef]
11. Yang, Z.Y.; Yu, S.C.; Lou, W.J.; Liu, C. P^2: Privacy-Preserving communication and precise reward architecture for V2G networks in smart grid. *IEEE Trans. Smart Grid* **2011**, *2*, 697–706. [CrossRef]
12. Tseng, H.R. A secure and privacy-preserving communication protocol for V2G networks. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 1–4 April 2012.
13. Nicanfar, H.; Hosseininezhad, S.; Talebifard, P.; Leung, V.C.M. Robust privacy-preserving authentication scheme for communication between electric vehicle as power energy storage and power stations. In Proceedings of the 2013 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Turin, Italy, 14–19 April 2013.
14. Au, M.H.; Liu, J.K.; Fang, J.; Jiang, Z.L.; Susilo, W.; Zhou, J. A new payment system for enhancing location privacy of electric vehicles. *IEEE Trans. Veh. Technol.* **2014**, *63*, 3–18. [CrossRef]
15. Stegelmann, M.; Kesdogan, D. Location privacy for vehicle-to-grid interaction through battery management. In Proceedings of the 2012 Ninth International Conference on Information Technology—New Generations, Las Vegas, NV, USA, 16–18 April 2012.
16. Stegelmann, M.; Kesdogan, D. V2GPriv: Vehicle-to-Grid privacy in the smart grid. In Proceedings of the 4th International Symposium, CSS 2012, Melbourne, Australia, 12–13 December 2012; Cyberspace Safety and Security, Xiang, Y., Lopez, J., Kuo, C.-C.J., Zhou, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2012.
17. Xiao, Y.; Xiong, L.; Yuan, C. Differentially private data release through multidimensional partitioning. In Proceedings of the 7th VLDB Workshop, Workshop on Secure Data Management, Singapore, 17 September 2010; Secure Data Management, Jonker, W., Petković, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2010.
18. Qardaji, W.; Yang, W.; Li, N. Differentially private grids for geospatial data. In Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE), Brisbane, Australia, 8–12 April 2013.

19. Mir, D.J.; Isaacman, S.; Cáceres, R.; Martonosi, M.; Wright, R.N. DP-WHERE: Differentially private modeling of human mobility. In Proceedings of the 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA, 6–9 October 2013.
20. To, H.; Ghinita, G.; Shahabi, C. A framework for protecting worker location privacy in spatial crowdsourcing. *Proc. VLDB Endow.* **2014**, *7*, 919–930. [CrossRef]
21. Cormode, G.; Procopiuc, C.; Srivastava, D.; Shen, E.; Yu, T. Differentially private spatial decompositions. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, Washington, DC, USA, 1–5 April 2012.
22. Fan, L.; Bonomi, L.; Xiong, L.; Sunderam, V. Monitoring web browsing behavior with differential privacy. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014.
23. Zhang, J.; Xiao, X.; Xie, X. PrivTree: A differentially private algorithm for hierarchical decompositions. In Proceedings of the 2016 International Conference on Management of Data, San Francisco, CA, USA, 26 June–1 July 2016.
24. Lee, J.; Clifton, C.W. Top-k frequent itemsets via differentially private FP-trees. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014.
25. Chen, R.; Xiao, Q.; Zhang, Y.; Xu, J. Differentially private high-dimensional data publication via sampling-based inference. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015.
26. Hay, M.; Rastogi, V.; Miklau, G.; Dan, S. Boosting the accuracy of differentially private histograms through consistency. *Proc. Vldb Endow.* **2010**, *3*, 1021–1032. [CrossRef]
27. Qardaji, W.; Yang, W.; Li, N. Understanding hierarchical methods for differentially private histograms. *Proc. VLDB Endow.* **2013**, *6*, 1954–1965. [CrossRef]
28. Kellaris, G.; Papadopoulos, S. Practical differential privacy via grouping and smoothing. *Proc. VLDB Endow.* **2013**, *6*, 301–312. [CrossRef]
29. Peng, S.; Yang, Y.; Zhang, Z.; Winslett, M.; Yu, Y. DP-tree: Indexing multi-dimensional data under differential privacy (abstract only). In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, Scottsdale, AZ, USA, 20–24 May 2012.
30. To, H.; Fan, L.; Shahabi, C. Differentially private H-tree. In Proceedings of the 2nd Workshop on Privacy in Geographic Information Collection and Analysis, Bellevue, WA, USA, 3–6 November 2015.
31. He, X.; Cormode, G.; Machanavajjhala, A.; Procopiuc, C.M.; Srivastava, D. DPT: Differentially private trajectory synthesis using hierarchical reference systems. *Proc. VLDB Endow.* **2015**, *8*, 1154–1165. [CrossRef]
32. Acs, G.; Castelluccia, C. A case study: Privacy preserving release of spatio-temporal density in Paris. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014.