applied sciences

MDPI

*Article*

# A Location Privacy Preservation Method Based on Dummy Locations in Internet of Vehicles

**Xianyun Xu [1], Huifang Chen [1,2,3,*] and Lei Xie [1,4]**

[1] College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China; 21831091@zju.edu.cn (X.X.); xiel@zju.edu.cn (L.X.)
[2] Zhoushan Ocean Research Center, Zhoushan 316021, China
[3] State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China
[4] Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Hangzhou 310027, China
* Correspondence: chenhf@zju.edu.cn; Tel.: +86-571-8795-1820 (ext. 217)

**Featured Application: This work can used in location privacy preservation in internet of vehicles.**

**Abstract:** During the procedure, a location-based service (LBS) query, the real location provided by the vehicle user may results in the disclosure of vehicle location privacy. Moreover, the point of interest retrieval service requires high accuracy of location information. However, some privacy preservation methods based on anonymity or obfuscation will affect the service quality. Hence, we study the location privacy-preserving method based on dummy locations in this paper. We propose a vehicle location privacy-preservation method based on dummy locations under road restriction in Internet of vehicles (IoV). In order to improve the validity of selected dummy locations under road restriction, entropy is used to represent the degree of anonymity, and the effective distance is introduced to represent the characteristics of location distribution. We present a dummy location selection algorithm to maximize the anonymous entropy and the effective distance of candidate location set consisting of vehicle user's location and dummy locations, which ensures the uncertainty and dispersion of selected dummy locations. The proposed location privacy-preservation method does not need a trustable third-party server, and it protects the location privacy of vehicles as well as guaranteeing the LBS quality. The performance analysis and simulation results show that the proposed location privacy-preservation method can improve the validity of dummy locations and enhance the preservation of location privacy compared with other methods based on dummy locations.

**Keywords:** privacy preservation; Internet of vehicles (IoV); location-based services (LBS); location privacy; dummy location; effective distance



**Citation:** Xu, X.; Chen, H.; Xie, L. A Location Privacy Preservation Method Based on Dummy Locations in Internet of Vehicles. *Appl. Sci.* **2021**, *11*, 4594. https://doi.org/10.3390/app11104594

Academic Editor: Gianluca Lax

Received: 25 April 2021
Accepted: 14 May 2021
Published: 18 May 2021



## 1. Introduction

With the development and application of wireless networks, the vehicular ad hoc network (VANET) is becoming an important part of future intelligent transport system. It is expected to play an important role in the road safety [1], traffic management [2], information dissemination to drivers and passengers [3], and so on. With increasing number of vehicles being connected to the Internet of things, the conventional VANET is changing into the Internet of vehicles (IoV).

Moreover, the use of location-based services (LBS) application from mobile devices and applications (apps) is rapidly increasing [4]. When a user acquires the LBS, it needs to provide its location, which results in the disclosure of the location privacy. In addition, a vehicle may act as a provider of location services. For example, when a vehicle participates in a task based on swarm intelligence perception, it should expose the location privacy. Hence, the problem of privacy preservation in the LBS should be resolved [5].

To address the privacy-preserving issue, many approaches have been proposed over the past few years. Most of them are based on the location perturbation and obfuscation adopt well-known privacy metrics such as K-anonymity [6] and rely on a trusted third-party server [7,8]. However, K-anonymity privacy-preserving scheme is suitable for high vehicle density. When there are fewer vehicles, spatial anonymity may not be realized, or the anonymous area formed is too large [9]. On the other hand, for the point of interest (POI) retrieval service in IoV, the accuracy of retrieval results is related to the precision of provided location information. However, the location privacy-preservation schemes based on anonymity or obfuscation cannot guarantee the accuracy of location information, which affects the quality of LBS [10–13].

In the location privacy-preservation method based on dummy locations, a location set containing (or implied) the user's real location is provide to the LBS server. Hence, this method can ensure the accuracy of POI retrieval results [14]. At the same time, the generation of dummy locations does not need a trustable third-party server. In recent years, many location privacy-preservation methods based on dummy locations have been proposed [15–23]. However, due to the characteristics of vehicles, the location of vehicles is subject to the road distribution, many methods cannot be directly adopted in IoV.

In IoV, road information can be used to preprocess dummy locations by the LBS server. Since the enhanced dummy location selection (E-DLS) [18] algorithm does not take the road information into consideration, the validity of dummy locations cannot be guaranteed. In addition, due to the restriction of roads and roadside buildings, the distribution of dummy locations is constrained. Although dummy locations are generated combining with location semantic information [24], the required location distribution is difficult to be achieved under road constraints. Considering the geographical constraint, a method is proposed to generate and arrange dummy objects around users in a grid form [25]. However, the method ignores the history request information. Due to the shortcomings in the existing location privacy-preserving methods in IoV, we investigate the problem of location privacy preservation under road constraints in this paper.

In this paper, we propose a vehicle location privacy-preservation method based on dummy locations. In the proposed method, the dummy location selection algorithm is modified based on vehicle location features. The main contributions of this paper as follows:

- We investigate the problem of vehicle location privacy preservation in IoV and propose a vehicle location privacy-preservation method based on dummy locations.
- We define the concept of effective distance to represent the characteristics of vehicle location distribution. Moreover, we improve the dummy location selection algorithm by using anonymous entropy and effective distance.
- We analyze the performance of the proposed method in terms of security, computation overhead, and communication overhead, and conduct extensive simulations to evaluate the proposed method.

The rest of the paper is organized as follows. The related work about location privacy-preservation methods is overviewed in Section 2. In Section 3, we give some preliminaries and the problem aiming to be solved in this paper. In Section 4, we propose a vehicle location privacy-preservation method based on dummy locations. Performance analysis and simulation results are given to verify the proposed method in Sections 5 and 6, respectively. Finally, we conclude the paper in Section 7.

## 2. Related Work

The location privacy-preserving problem has been attracting wide attention from both academia and industry. This problem draws even more attention due to the booming of LBSs. Many location privacy-preservation methods have been proposed, such as K-anonymity [7–10], obfuscation [11–13], differential privacy [26,27], mixed zone [28,29], homomorphic encryption [30–32], and dummy locations [15–23]. In this work, we focus on the location privacy preservation-method based on dummy locations in IoV.

The privacy-preservation method based on dummy locations can work without a third-party server, and provides a location set containing the user's real location to guarantee the quality of LBS. Hence, this method can achieve a good tradeoff between location privacy-preservation and service quality. Sun et al. [15] proposes a privacy-preservation method based on dummy locations, where a query is submitted to an LBS provider with the actual location of a user and other dummy locations. The LBS provider searches for all the related POI locations and returns them to the user. Hence, the generation of dummy locations is the key issue in the privacy-preservation method based on dummy locations. Grid-based and circle-based algorithms for generating dummy locations are proposed to satisfy regional privacy requirements in [16]. A distributed dummy client generation method is proposed to make clients control over their privacy protection [17]. The method selects clients with movement patterns be close to user's movement pattern according to the privacy requirements. However, many dummy location generation algorithms assume that an attacker has no other background information and select locations randomly. To solve this problem, an E-DLS algorithm is proposed in [18]. In the E-DLS algorithm, dummy locations are selected to optimize the privacy-preserving effect in terms of the maximum entropy and cloaking region (CR). Based on E-DLS, Liao et al. [19] considers that when the attacker can obtain the type of service, the greedy algorithm based on entropy measurement is proposed to select the dummy locations to construct the anonymous area.

Recently, the trajectory privacy-preserving issue for continuous LBS has been becoming a hot research topic. A method named Dummy-Q is proposed for query privacy-preservation in the continuous LBS scenarios [20], where the query privacy is protected by generating dummy queries. In [21], the problem of privacy leakage under continuous LBS is studied, and a frequency-aware dummy-based method (FADBM) is proposed to ensure that dummy locations are generated around frequent areas and the time accessibility. In [22], a dummy filtering algorithm is proposed, where the spatiotemporal correlation of time-sensitive side information is used to select and generate dummy locations, and spatiotemporal correlation between locations is truncated with time accessibility and access constraints to ensure trajectory similarity. In [23], a location privacy method is proposed to prevent privacy disclosure in LBS constrained in incomplete data collection, where the anonymous candidate set is constructed with compressing sensing technology. Moreover, the differential privacy mechanism is adopted to construct the anonymous candidate set for continuous LBS.

Since the trajectory of vehicles is subject to the road distributions, the influence of road information should be considered for designing location privacy-preservation methods in IoV. Considering the geographical constraint, a method is proposed to generate and arrange dummy objects around users in a grid form [25]. However, the service request probability of locations is not considered when generating dummy locations, the effect of privacy protection is poor. Lina et al. [33] proposes a location privacy-preservation scheme based on anonymous entropy, where anonymous entropy based on location distance and request content is considered. Moreover, two algorithms are presented to select dummy users to build anonymous regions for the dense region and the sparse region, respectively. In combination with the characteristics of vehicle network, a privacy preservation algorithm converts road map into edge cluster diagram in order to hide road information and vehicle information, and constructs invisible areas based on K-anonymity and L-diversity [34]. A region-of-interest division-based algorithm is proposed to preserve the location privacy of mobile device users in location-based cyber services [24]. In this method, dummy locations are generated considering the semantic information of those locations.

We will study the location privacy-preservation based on dummy locations in IoV, in which road constraints and vehicle location characteristics are taken into consideration.

## 3. Preliminaries and Problem Formulation

In this section, we introduce some preliminaries including the system model, LBS query, service semantics, anonymous entropy, and adversary model. Then, the problem to be solved in this paper is formulated.

### *3.1. System Model*

Figure 1 illustrates the system architecture of IoV that consists of a number of intelligent vehicles with onboard unit (OBU), several roadside units (RSUs), a trusted authority (TA), and an LBS server. OBU can acquire the perceived driving information of on-board sensors, calculate, process, and store the sensed data. The communication modes in IoV, namely vehicle-to-vehicle (V2V) and vehicle-to-RSU (V2R), adopt dedicated short range communication (DSRC) technology. Through V2V communication, intelligent vehicles can not only obtain driving state information through sensors and exchange messages, but also receive and forward messages broadcasted by other vehicles. Through V2R communication, vehicles can exchange information with the RSU and access the Internet.

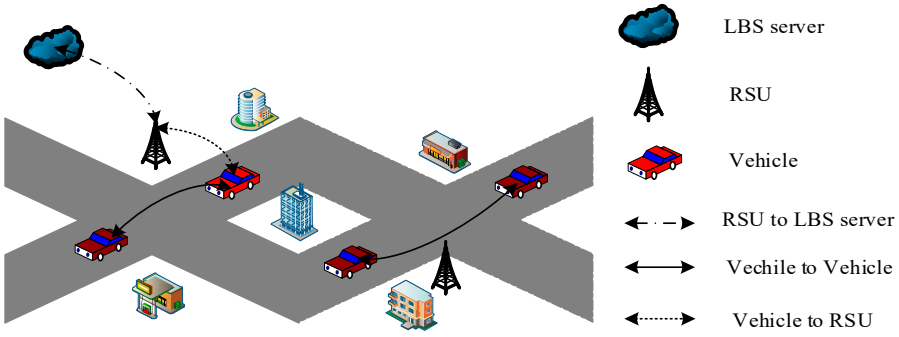


**Figure 1.** The system architecture of IoV.

### *3.2. LBS Query*

An LBS query $Lq$ is defined as $Lq = (u_{id}, \{(x, y), C, V\})$, where $u_{\mathrm{id}}$ denotes a user's identity; $(x, y)$ represents the user's location information, $x$ and $y$ represent latitude and longitude, respectively; $C$ denotes the user's query content; $V$ is the user's privacy preservation level.

However, since the LBS provider may be malicious, the user's location will be disclosed if the user directly sends $Lq$ to the LBS provider. To preserve the location privacy, dummy locations method is used to preprocess $Lq$. Hence, $Lq$ is transformed to $Lq'$ as $Lq' = (u_{\mathrm{id}}, \{(x, y), (x_1, y_1), \ldots, (x_{k-1}, y_{k-1}), C, C_1, \ldots, C_{k-1}, V\})$, where $(x_1, y_1), \ldots, (x_{k-1}, y_{k-1})$ are $k-1$ dummy locations, $C_i$ represents the query content sent at dummy location $(x_i, y_i)$, $i = 1, 2, \ldots, k-1$.

From $Lq'$, the adversary cannot determine the user's real location from $k-1$ dummy locations. By this way, the vehicle location privacy can be protected.

### *3.3. Service Semantics*

In each location, users may request entertainment, medical treatment, transportation, or other services. The service requests sent by users are closely related to their locations, and the probabilities of various services in different locations are different. Therefore, service semantics is used to represent the relationship between location and service.

Let $U$ be the number of services, $e_{i,u}$ represents the request probability of service $u$ in location $(x_i, y_i)$, $0 \leq e_{i,u} \leq 1$, $i = 0, 1, \ldots, k-1$, $u = 1, 2, \ldots, U$, and $\sum_{u=1}^{U} e_{i,u} = 1$. In this paper, the LBS server is responsible for the collection and establishment of service semantics.

### 3.4. Anonymous Entropy

It is pointed out in [17] that entropy can measure the uncertainty of target location in the location set. In this paper, we use entropy to evaluate the degree of anonymity.

Here, we consider set $\mathcal{G}$ including $k$ locations, $\mathcal{G} = \{(x_0, y_0), (x_1, y_1), \ldots, (x_{k-1}, y_{k-1})\}$. The service request probability at location $(x_i, y_i)$ is $q_i$, the candidate probability of location $(x_i, y_i)$ is $p_i$. If the vehicle user at location $(x_i, y_i)$ request service $u$, the service semantics at location $(x_i, y_i)$ is $e_{i,u}$, and the request probability of service $u$ at location $(x_i, y_i)$ is $q'_i, q'_i = q_i e_{i,u}$, $i = 0, 1, \ldots, k-1$, $u = 1, 2, \ldots, U$. Hence, the anonymous entropy is defined as

$$H = -\sum_{i=0}^{k-1} p_i log_2 p_i, \tag{1}$$

where

$$p_i = \frac{q'_i}{\sum_{i=0}^{k-1} q'_i}. \tag{2}$$

According to the mathematical property of entropy, it is required that the candidate probabilities of $k$ locations be the same to achieve the maximum entropy. That is, if $p_i = 1/k$, $i = 0, 1, \ldots, k-1$, the maximum of anonymous entropy of set $\mathcal{G}$ is $\log_2 k$.

### 3.5. Adversary Model

The goal of the adversary is to obtain sensitive information about a particular user. There are two types of adversary model, passive adversary, and active adversary.

A passive adversary can monitor and eavesdrop on wireless channels or compromise users to obtain other users' sensitive information. A passive adversary can perform eavesdropping attack to learn extra information about a user.

An active adversary can compromise the LBS server and obtain all the information known by the server.

In this work, we assume that the LBS server and RSUs are honest-but-curious, as active adversaries. Hence, the adversary can obtain global information and monitor all the LBS queries from users. In addition, the adversary knows the location privacy-preservation scheme adopted in the system. Based on the known information, the adversary tries to infer and learn other sensitive information.

### 3.6. Problem Formulation

The LBS server divides the area covered by an RSU into $I \times J$ cells as shown in Figure 2. $cell_{i,j}$ denotes the cell of row $i$ and column $j$, $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$. The location of $cell_{i,j}$ is denoted as $\boldsymbol{r}_{i,j}$, and $\boldsymbol{r}_{i,j} = (x_{i,j}, y_{i,j})$. The request probability of $cell_{i,j}$ is $q_{i,j}$, the service semantics of $cell_{i,j}$ is $e_{(i,j),u}$, and the information matrix $Q(\boldsymbol{r}, q, \boldsymbol{e})$ for each RSU can be set up.

Figure 2 shows service request probability distribution, the area is divided into $10 \times 10$ cells. The star represents the user's real location, and the triangle represents the dummy location, and the shade in each cell represents its request probability generated based on the Borlange data set [35]. The gray block represents the road, and **R** represents the location area accessible by the road.

In Figure 2a, a vehicle user randomly generates $k-1$ dummy locations in order to protect the location privacy. Then vehicle user uses the dummy locations and real location to send service request to the LBS server. In theory, the probability of exposing the user's real location can be $1/k$. However, using some auxiliary information, the LBS server can deduce the real location with a probability of $1/(k - k_d)$, where $k_d$ is the number of dummy locations be filtered out through the auxiliary information.

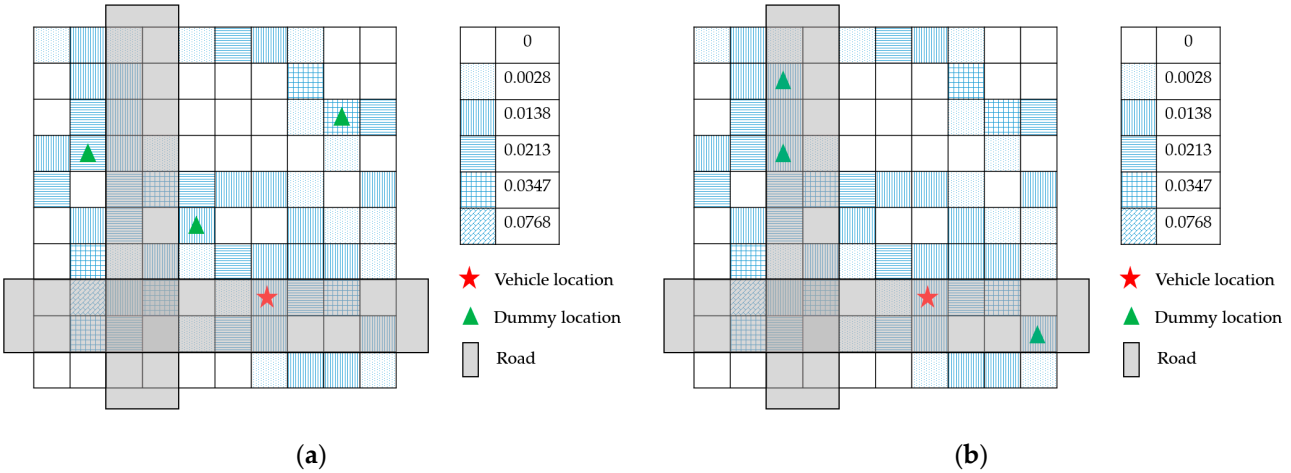


**Figure 2.** Service request probability distribution. (**a**) with random dummy location selection algorithm; (**b**) with dummy location selection algorithm under road restrictions.

In IoV, since the location of vehicles is restricted by the road, and the service request probability is used as auxiliary information, the validity of dummy locations generated with random dummy location selection algorithm in Figure 2a, E-DLS algorithm in [18] and Dest-ex algorithm in [25] is affected. The dummy locations filtered out by the LBS server, $k_{\mathrm{d}}$, increases. For example, in Figure 2a, $k = 4$, and $k_{\mathrm{d}} = 3$. Hence, the effect of privacy protection is degraded.

Therefore, to protect the location privacy of vehicles, it is necessary to ensure the validity of dummy locations generated. When road information, service request probability and service semantics are used as auxiliary information, set $\mathcal{G}$ is set up for minimizing $k_{\mathrm{d}}$. The optimization problem can be defined as

$$\begin{gathered} \min_{\mathcal{G}} k_d \\ \text{s.t.} G(r,q,e) \subset Q(r,q,e) \\ \forall r_{i,j} \in \mathcal{G}, r_{i,j} \in \mathcal{C}, r_{i,j} \in R \\ |\mathcal{G}| = k, \end{gathered} \tag{3}$$

where $G(\boldsymbol{r}, q, \boldsymbol{e})$ is the information matrix corresponding to set $\mathcal{G}$ which consists of vehicle user's location and $k-1$ dummy locations, and set $\mathcal{C}$ is the set of all locations of cells in the area covered by the RSU.

## 4. Algorithm Design

In this section, we present a location privacy-preservation method based on dummy locations in IoV, where a dummy location selection algorithm is addressed to improve the validity of dummy locations.

### *4.1. Effective Distance*

As shown in Figure 3, due to the road restrictions and roadside buildings, the distribution of vehicles is in the form of "pipeline", and the aggregation distribution may occur. Hence, the validity of dummy locations further decreases. To ensure the validity of dummy locations, it is necessary to make the location distribution be uniform and dispersed, as shown in Figure 3b.

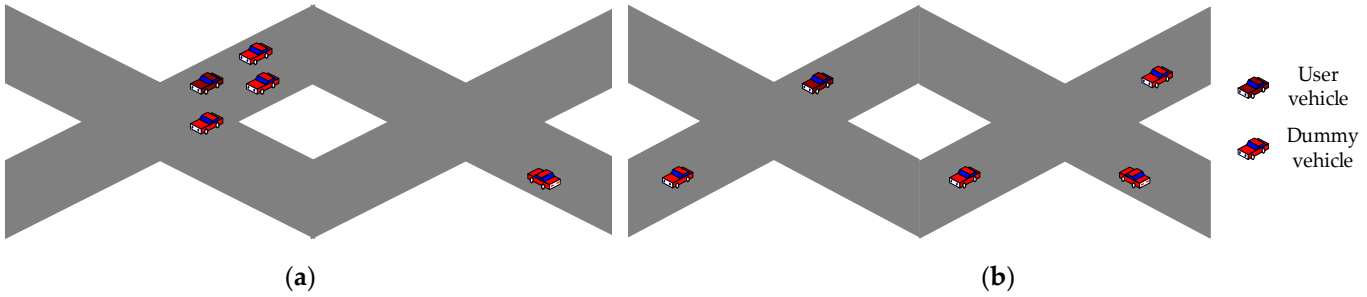


**Figure 3.** Schematic diagram of vehicle location distribution. (**a**) vehicle aggregation distribution; (**b**) vehicle dispersed distribution.

In order to make the distribution of generated dummy locations be uniform, we define the effective distance between locations as the minimum distance between the current location and other locations in a location set. That is,

$$d(\boldsymbol{r}_i) = \min_{\boldsymbol{r}_w \in \mathcal{W}, w \neq i} |\boldsymbol{r}_i, \boldsymbol{r}_w| = \min_{\boldsymbol{r}_w \in \mathcal{W}, w \neq i} \sqrt{(x_i - x_w)^2 + (y_i - y_w)^2}, \tag{4}$$

where $\mathcal{W}$ represents a location set, $\boldsymbol{r}_i$ represents location $i$ in set $\mathcal{W}$, the corresponding coordinates is $(x_i, y_i)$, $\boldsymbol{r}_w$ represents location $w$ in set $\mathcal{G}$, the corresponding coordinates is $(x_w, y_w)$, $i = 1, 2, \ldots, |\mathcal{W}|$, $w = 1, 2, \ldots, |\mathcal{W}|$, $|\mathcal{W}|$ is the number of elements in set $\mathcal{W}$, and $d(\boldsymbol{r}_i)$ is the effective distance of $\boldsymbol{r}_i$.

From Figure 4, one finds that the larger the effective distance, the greater the spacing between vehicles, and the more dispersed the distribution.

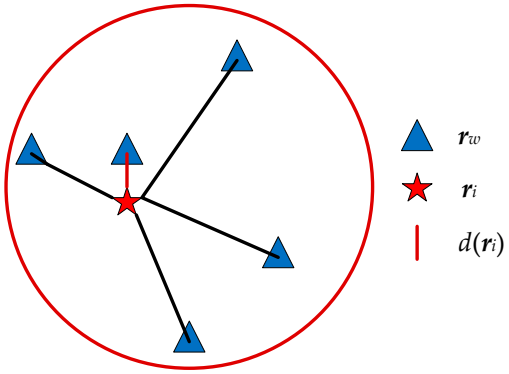


**Figure 4.** The diagram of effective distance.

### *4.2. Parameter Settings*

The location privacy protection requirement is presented by privacy protection level $V$, which indicates the success rate of location privacy protection. That is, $V = 1 - p = 1 - \frac{1}{k}$, and $V \in [0,1)$.

Privacy parameter $k$ is determined by privacy protection level $V$ set by the vehicle user. That is,

$$k = \left\lceil \frac{1}{1 - V} \right\rceil, \tag{5}$$

where $\lceil \cdot \rceil$ denotes the upper integer operation.

### *4.3. Dummy Location Selection Algorithm under Road Restriction*

In order to ensure the validity of dummy locations, two conditions should be considered simultaneously for selecting the dummy locations. One is to maximize the anonymous entropy of the candidate set. The other is to maximize the effective distance of the candidate set.

Hence, the optimization problem formulated in (3) can convert to a multiple object optimization problem as

$$\begin{gathered}\max_{\mathcal{G}}\left\{-\sum_{r_{i,j}\in\mathcal{G}} p_{i,j}\log_2 p_{i,j},\ \sum_{r_{i,j}\in\mathcal{G}} d(r_{i,j})\right\}\\ \text{s.t.}G(\boldsymbol{r},q,\boldsymbol{e})\subset Q(\boldsymbol{r},q,\boldsymbol{e})\\ \forall \boldsymbol{r}_{i,j}\in\mathcal{G},\boldsymbol{r}_{i,j}\in\mathcal{C},\boldsymbol{r}_{i,j}\in\mathbf{R}\\ |\mathcal{G}|=k,\end{gathered}\tag{6}$$

where candidate set $\mathcal{G}$ consists of the vehicle user's location and $k-1$ selected dummy locations.

Obviously, the problem formulated in (6) is difficult to resolve. Hence, we decouple the problem in (6) into two sub-problems, the anonymous entropy maximization sub-problem and the effective distance maximization sub-problem.

According to the background knowledge of the LBS server and the purpose of the dummy location selection algorithm, we give priority to the sub-problem of anonymous entropy maximization. That is,

$$\begin{gathered}\max_{\mathcal{G}'}\left\{-\sum_{r_{i,j}\in\mathcal{G}'} p_{i,j}\log_2 p_{i,j}\right\}\\ \text{s.t.}G'(\boldsymbol{r},q,\boldsymbol{e})\subset Q(\boldsymbol{r},q,\boldsymbol{e})\\ \forall \boldsymbol{r}_{i,j}\in\mathcal{G}',\boldsymbol{r}_{i,j}\in\mathcal{C},\boldsymbol{r}_{i,j}\in\mathbf{R}\\ |\mathcal{G}'|=k',\end{gathered}\tag{7}$$

where set $\mathcal{G}'$ including the vehicle user's location and $k'-1$ selected dummy locations is set up to resolve the sub-problem formulated in (7), $k'$ is the number of locations in set $\mathcal{G}'$, and $k'>k$.

According to $Q(\boldsymbol{r}, q, \boldsymbol{e})$, the vehicle user calculates the probability of service request at each location in $\mathbf{R}$, ${q'}_{(i,j),u}$, $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, $u=1,2,\ldots,U$, $cell_{i,j}\in\mathbf{R}$. According to service request probability of content $C_0$, the vehicle user selects other $k'-1$ locations whose service request probabilities are close to that of the vehicle user.

Hence, a candidate set $\mathcal{G}'$ is constructed with the vehicle user's location and $k'-1$ selected dummy locations.

Then, the sub-problem for maximizing the effective distance of the candidate set is to resolve. That is,

$$\begin{gathered}\max_{\mathcal{G}''}\left\{\sum_{r_{i,j}\in\mathcal{G}''} d(r_{i,j})\right\}\\ \text{s.t.}G''(\boldsymbol{r},q,\boldsymbol{e})\subset G'(\boldsymbol{r},q,\boldsymbol{e})\\ \forall \boldsymbol{r}_{i,j}\in\mathcal{G}'',\boldsymbol{r}_{i,j}\in\mathcal{G}'\\ |\mathcal{G}''|=k,\end{gathered}\tag{8}$$

where set $\mathcal{G}''$ including the vehicle user's location and $k-1$ selected dummy locations is set up to resolve the sub-problem formulated in (8).

To solve the sub-problem formulated in (8), the vehicle user selects $k-1$ dummy locations in a greedy manner.

Let $\boldsymbol{r}_{0,0}$ denote the location of the vehicle user. $\mathcal{G}''=\{\boldsymbol{r}_{0,0}\}$ and $\mathcal{G}''=\mathcal{G}''\setminus\{\boldsymbol{r}_{0,0}\}$. The vehicle user chooses $k-1$ locations with the maximum effective distance through $k-1$ rounds.

In the $i$th round, $i=1,2,\ldots,k-1$, the vehicle user calculates the effective distance of the location(s) in $\mathcal{G}'$ to the location(s) in $\mathcal{G}''$. If $\boldsymbol{r}_{i*,j*}=\arg\max\limits_{\boldsymbol{r}_{i,j}\in\mathcal{G}'}\left(\min\limits_{\boldsymbol{r}_{i',j'}\in\mathcal{G}''}\left|\boldsymbol{r}_{i,j},\boldsymbol{r}_{i',j'}\right|\right)$, the vehicle user puts $\boldsymbol{r}_{i*,j*}$ into set $\mathcal{G}''$ and deletes it from $\mathcal{G}''$.

Hence, set $\mathcal{G}''$ is constructed with the vehicle user's location and $k-1$ selected dummy locations.

### 4.4. A Location Privacy-Preservation Method Based on Dummy Locations under Road Restriction

The specific procedure of a location privacy-preservation method based on dummy location under road restriction can be follows:

(1) Based on the historical data of service requests, the LBS server counts the number of service requests initiated by vehicle users in each cell, and the service request probability of $cell_{i,j}$, $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$, $q_{i,j} = f_{i,j}/F$, where $f_{i,j}$ is the number of service requests initiated by vehicle users in $cell_{i,j}$, and $F$ is the number of service requests in the area. The service semantics of service $u$ is $q_{i,j} = f_{(i,j),u}/f_{i,j}$, where $f_{(i,j),u}$ is the number of requests of service $u$ initiated by vehicle users in $cell_{i,j}$, $u = 1, 2, \ldots, U$.
(2) The LBS server constructs and distributes the information matrix $Q(\boldsymbol{r}, q, \boldsymbol{e})$ within the RSU's jurisdiction to each RSU.
(3) RSU broadcasts $Q(\boldsymbol{r}, q, \boldsymbol{e})$ and $\mathbf{R}$ to users in its covered area.
(4) According to the privacy preservation level $V$, the vehicle user calculates its privacy parameter $k$ by (5).
(5) The vehicle user generates $k - 1$ dummy locations using dummy location selection algorithm under road restriction. The details are as follows:
    (5-a) Let $k' = 2k$. Within the locations in $\mathbf{R}$, other $k' - 1$ locations apart from the vehicle user's location are selected as dummy locations by solving the problem formulated in (7). Hence, a candidate set $\mathcal{G}'$ is constructed with the vehicle user's location and $k' - 1$ selected dummy locations.
    (5-b) Within set $\mathcal{G}'$, other $k - 1$ locations apart from the vehicle user's location are selected as dummy locations by solving the problem formulated in (8). Hence, set $\mathcal{G}''$ is constructed with the vehicle user's location and $k - 1$ selected dummy locations.
(6) The vehicle user generates service query $Lq'$ including locations in $\mathcal{G}''$, their corresponding service contents, and the privacy preservation level, and then, $Lq'$ is sent to the LBS server via RSU.
(7) Receiving service query $Lq'$, the LBS server retrieves service results according to $k$ locations and the corresponding service contents, and then, the LBS server returns service results to the vehicle user through RSU.
(8) The vehicle user selects the required result from service results according to its location.

## 5. Performance Analysis

In this section, the performance of the proposed location privacy-preservation method using dummy location selection algorithm under road restriction, abbreviated as RR-DLS, is analyzed.

### 5.1. Security Analysis

Since encrypt-based technologies can be easily applied to the proposed RR-DLS method, eavesdropping attack on wireless channels between users and other entities can be ignored. We focus on collusion attack and inference attack from passive and active attackers.

#### 5.1.1. Collusion Attack

Passive attackers may collude with some users to get additional information about other users or collude with the LBS server to predict sensitive information about legitimate users. If the probability of successfully guessing the real location of a vehicle user among $k$ locations in the service query does not increase with the number of collusion users, the proposed method can resist collusion attack.

We consider a situation that collusion occurs between a group of users aiming to acquire the user's real location from $k$ locations. In RR-DLS method, each user can only know the service request probability and road condition collected by itself. When eavesdropping

the service query sent to the LBS server, the attacker cannot filter out some invalid locations through additional information since $k$ locations in the service query have the same or similar service request probability and are on the road.

One extreme case for the passive adversary is that it can acquire the global information by compromising the LBS server as well as RSUs. In this case, it becomes an active adversary and can perform inference attack as discussed in the following.

#### 5.1.2. Inference Attack

The LBS server and RSUs have global information, such as information matrix $Q(\boldsymbol{r}, q, \boldsymbol{e})$, road information $\mathbf{R}$ and $k$ locations in the service query, and so on. Based on this information, the LBS server or the RSU can act as an active attacker to launch reasoning attack and acquire some sensitive information of users.

Suppose $p_G$(event) be the probability that an attacker successfully guesses that event is true. The proposed method should satisfy (9) to resist inference attack.

$$p_G(\boldsymbol{r}_{i,j} \in \mathcal{B} \Big| \mathcal{B} \cap \mathcal{G}'' \neq \varnothing) = p_G(\boldsymbol{r}_{i',j'} \in \mathcal{B} \Big| \mathcal{B} \cap \mathcal{G}'' \neq \varnothing), \boldsymbol{r}_{i,j} \in \mathcal{G}'', \boldsymbol{r}_{i',j'} \in \mathcal{G}'', \ \boldsymbol{r}_{i,j} \neq \boldsymbol{r}_{i',j'}, \tag{9}$$

where set $\mathcal{B}$ consists of the locations obtained by an attacker.

For any dummy location $\boldsymbol{r}_{i,j}$ generated by RR-DLS algorithm, the probability of $\boldsymbol{r}_{i,j}$ being guessed as the real location is

$$p_G(\boldsymbol{r}_{i,j} \in \mathcal{B} \Big| \mathcal{B} \cap \mathcal{G}'' \neq \varnothing) = \frac{p_G(\boldsymbol{r}_{i,j} \in \mathcal{B}, \ \mathcal{B} \cap \mathcal{G}'' \neq \varnothing)}{p_G(\mathcal{B} \cap \mathcal{G}'' \neq \varnothing)} = \frac{p_{i,j}}{p_G(\mathcal{B} \cap \mathcal{G}'' \neq \varnothing)}, \ \boldsymbol{r}_{i,j} \in \mathcal{G}''. \tag{10}$$

Substituting (10) into (9), we have

$$p_{i,j} \simeq p_{i',j'} \ , \boldsymbol{r}_{i,j} \in \mathcal{G}'', \boldsymbol{r}_{i',j'} \in \mathcal{G}'', \ \boldsymbol{r}_{i,j} \neq \boldsymbol{r}_{i',j'}. \tag{11}$$

The proposed dummy location selection algorithm under road restriction selects locations with the same or similar probability of service requests and service semantics. Hence, the proposed RR-DLS method satisfies the condition in (11), which means that the method can effectively resist inference attack.

### *5.2. Computation Overhead*

If RSU jurisdiction is divided into $I \times J$ cells, the number of services is $U$ and the number of results returned by the LBS server is $n$.

In the procedure of an LBS query, the vehicle user needs to generate $k$ dummy locations. First, as the vehicle user selects $2k-1$ locations based on service request probability, the computation overhead is $O(IJU)$. As the vehicle user selects dummy locations by effective distance through $k-1$ rounds. In the $i^{\text{th}}$ round, $i = 1, 2, \ldots, k-1$, the vehicle user calculates the effective distance of $2k-1-i$ locations in $\mathcal{G}'$ to $i$ locations in $\mathcal{G}''$ to update the effective distance of each location, and the location with maximum effective distance of locations in $\mathcal{G}'$ is selected. Hence, the computation overhead is $O(k^2)$. Therefore, the computation overhead of dummy location selection algorithm at the vehicle user is $O(k^2 + IJU)$.

Since RSU does not need additional computation, the computation overhead at RSU is $O(1)$.

The LBS server needs to perform service retrieval for $k-1$ dummy locations and a real location. Hence, the computation overhead at the LBS server is $O(kn)$.

### *5.3. Communication Overhead*

In the procedure of an LBS query, the vehicle user sends service query to the LBS server through RSU. The communication overhead at the vehicle user is $O(k)$.

RSU needs to broadcast the service request probability, the service semantics and other information. The communication overhead is $O(IJU)$. At the same time, RSU needs to forward the service query to the LBS server and return $kn$ service query results to the vehicle user. Therefore, the communication overhead at the RUS is $O(IJU + kn + k)$.

The LBS server needs to send the service request probability, service semantics and other information to RSU. The communication overhead is $O(IJU)$. Receiving the service query, the corresponding service results are returned to RSU. The communication cost is $O(kn)$. Therefore, the communication overhead at LBS server is $O(IJU + kn)$.

The performance of proposed RR-DLS method in terms of computation overhead and communication overhead is listed in Table 1.

**Table 1.** Performance of proposed RR-DLS method.

| Entity | Computation Overhead | Communication Overhead |
|---|---|---|
| Vehicle user | $O(k^2 + IJU)$ | $O(k)$ |
| RSU | $O(1)$ | $O(IJU + kn + k)$ |
| LBS Server | $O(kn)$ | $O(IJU + kn)$ |

## 6. Performance Evaluation and Discussion

In this section, the performance of proposed RR-DLS method is valuated. Moreover, we compare the performance of proposed RR-DLS algorithm with some existing dummy location selection algorithms, such as random dummy location selection algorithm, E-DLS algorithm in [18], Dest-ex algorithm in [25].

The simulation area and the corresponding service request probability distribution are illustrated in Figure 5, which is a region in Hangzhou with an area of 500 m × 500 m. This region is divided into 10 × 10 cells, the number of service types $U = 4$, the service request probability and service semantics are generated randomly, and orange cells represent locations that are inaccessible to the vehicle.

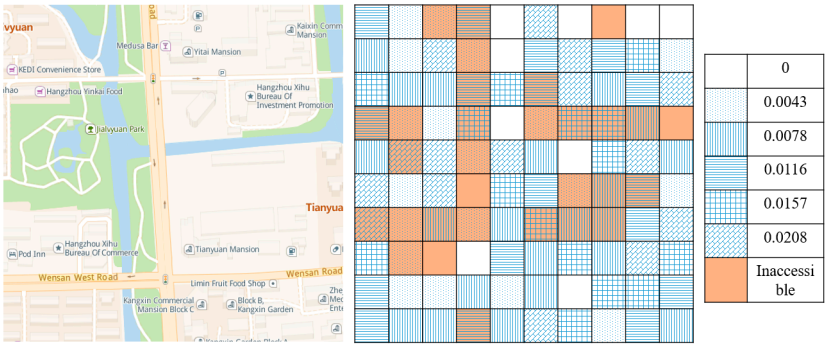


**Figure 5.** The simulation area and the corresponding service request probability distribution.

The simulation environment is Windows10, with 8 GB memory and AMD Ryzen 5 3550 H processor.

### *6.1. Computation Overhead*

Figure 6 shows the impact of privacy parameter $k$ on computation overhead in terms of execution time. From Figure 6, we observe that the computation overhead of proposed RR-DLS method is concentrated on the vehicle user side, and the execution time increases rapidly along with the increase of privacy parameter $k$. The computation overhead at RSU and the LBS server side is small. The execution time of RSU is independent of privacy parameter $k$, and the execution time of the LBS server increases linearly along with the increase of privacy parameter $k$.

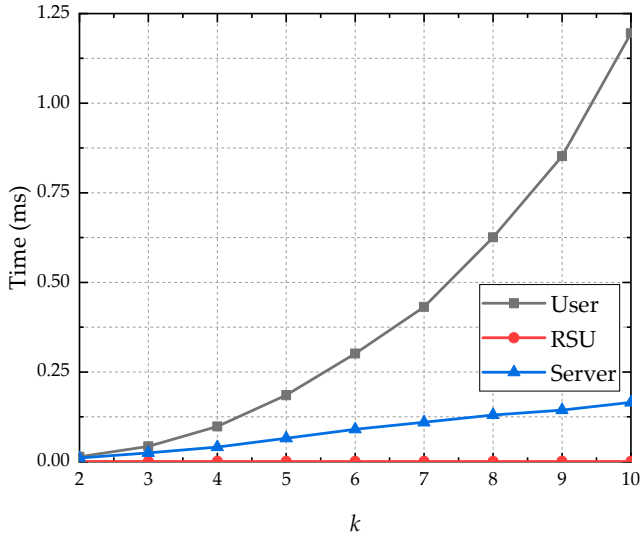


**Figure 6.** Impact of privacy parameter on execution time.

### *6.2. Communication Overhead*

Figure 7 shows the impact of privacy parameter $k$ on communication overhead in terms of data traffic. From Figure 7, we observe that the communication overhead of proposed RR-DLS method is concentrated on RSU and the LBS server, and the communication overhead at the vehicle user side is small. As privacy parameter $k$ increases, the communication overhead in terms of data traffic increases.

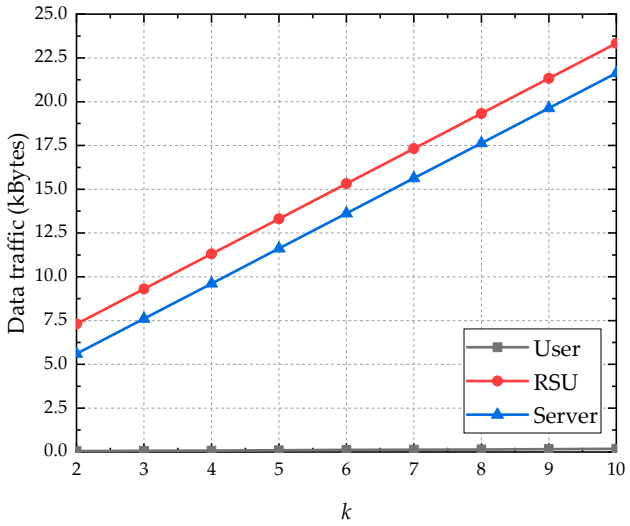


**Figure 7.** Impact of privacy parameter on data traffic.

### *6.3. Anonymous Entropy*

Figure 8 shows the anonymous entropy of four different dummy location selection algorithms, the proposed RR-DLS algorithm, random dummy location selection algorithm, E-DLS algorithm in [18], and Dest-ex algorithm in [25]. From Figure 8, we observe that the anonymous entropy of proposed RR-DLS algorithm is the largest. This is because the proposed RR-DLS algorithm can ensure the validity of dummy locations. Since Dest-ex algorithm only considers the road information, the anonymous entropy of Dest-ex algorithm is smaller than that of proposed RR-DLS algorithm, and larger than that of random dummy location selection algorithm and E-DLS algorithm. Since E-DLS algorithm selects dummy locations according to service request probability and CR, some dummy locations can be filtered using auxiliary knowledge. The anonymous entropy of E-DLS

algorithm is low. Since random selection algorithm selects dummy locations randomly, the anonymous entropy of random dummy location selection algorithm is the lowest.

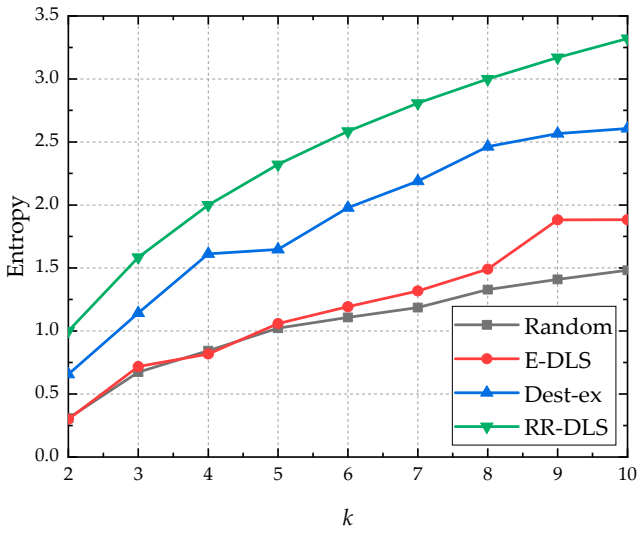


**Figure 8.** Anonymous entropy of different dummy location selections algorithms.

### *6.4. Effective Distance*

Figure 9 shows the effective distance of two different dummy location selection algorithms, the proposed RR-DLS algorithm and E-DLS algorithm in [18]. For E-DLS algorithm, the anonymous area is maximized considering the query probability. From Figure 9a, the means of effective distance of two algorithms are close. Moreover, from Figure 9b, we observe that the variance of effective distance of proposed RR-DLS algorithm is much smaller than that of E-DLS algorithm. The proposed RR-DLS algorithm can guarantee the distributed and uniform distribution of dummy locations to ensure the validity of dummy locations.

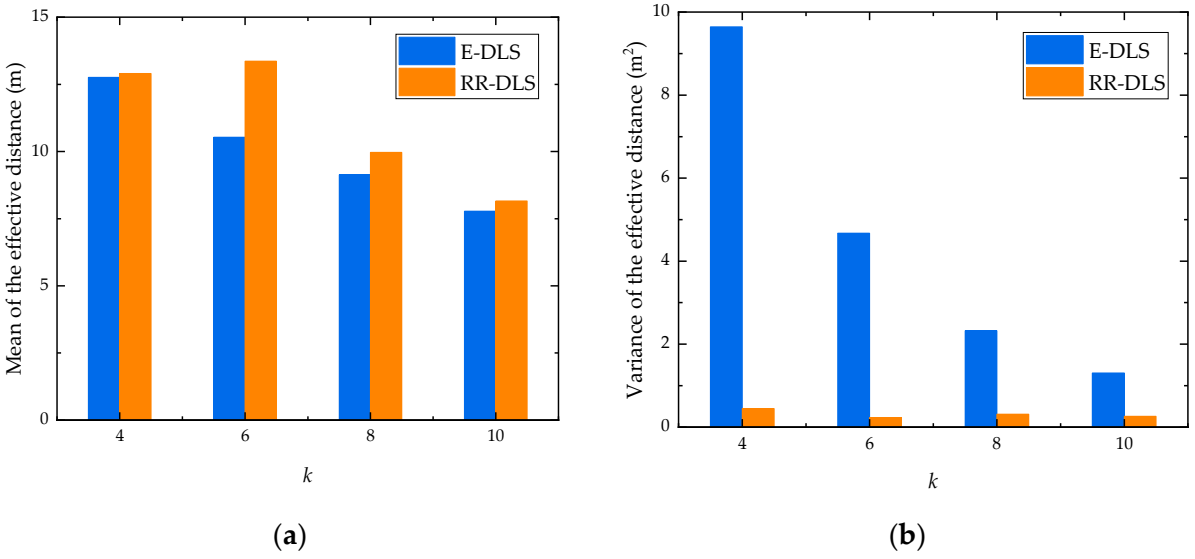


**Figure 9.** Effective distance of two different location selection algorithms. (**a**) mean; (**b**) variance.

## 7. Conclusions

In this paper, we investigated the vehicle location privacy-preserving problem in IoV and proposed a location privacy-preservation method based on dummy locations under road restriction. In the proposed RR-DLS method, the effective distance is introduced

to represent the characteristics of location distribution in order to improve the validity of dummy locations. A dummy location selection algorithm under road restriction was addressed according to anonymous entropy and effective distance. Security analysis results show that the proposed RR-DLS method can resist collusion attack and inference attack effectively. Performance analysis and simulation results show that the proposed RR-DLS method can effectively protect the vehicle location privacy and ensure the accuracy of LBS service. Furthermore, the proposed RR-DLS method increases the computation overhead at the vehicle user and communication overhead at RSU and the LBS server.

In the future, we will study the problem of vehicle trajectory privacy preservation in continuous LBS scenario.

**Author Contributions:** Conceptualization, X.X. and H.C.; methodology, X.X. and H.C.; software, X.X.; validation, X.X. and L.X.; formal analysis, X.X. and H.C.; investigation, X.X.; resources, L.X.; data curation, X.X.; writing—original draft preparation, X.X. and H.C.; writing—review and editing, H.C.; supervision, H.C.; project administration, H.C.; funding acquisition, H.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the Fund of National Natural Science Foundation of China under Grant 61671410, and the Science and Technology Department of Zhejiang Province under Grant 2018R52046 and Grant LGG18F010005.

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

## References

1. Raya, M.; Hubaux, J.-P. Securing vehicular ad hoc networks. *J. Comput. Secur.* **2007**, *15*, 39–68. [CrossRef]
2. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3848–3858. [CrossRef]
3. Qiu, H.; Qiu, M.; Lu, R. Secure V2X Communication Network based on Intelligent PKI and Edge Computing. *IEEE Netw.* **2019**, *34*, 172–178. [CrossRef]
4. Sun, G.; Sun, S.; Sun, J.; Yu, H.; Du, X.; Guizani, M. Security and privacy preservation in fog-based crowd sensing on the internet of vehicles. *J. Netw. Comput. Appl.* **2019**, *134*, 89–99. [CrossRef]
5. Gupta, R.; Rao, U.P. An Exploration to Location Based Service and Its Privacy Preserving Techniques: A Survey. *Wirel. Pers. Commun.* **2017**, *96*, 1973–2007. [CrossRef]
6. Jiang, T.; Wang, H.J.; Hu, Y.-C. Preserving location privacy in wireless lans. In *MobiSys '07: Proceedings of the 5th International Conference on Mobile Software Engineering and Systems*; Association for Computing Machinery (ACM): New York, NY, USA, 2007; pp. 246–257.
7. Sweeney, L. k-Anonymity: A Model for Protecting Privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **2002**, *10*, 557–570. [CrossRef]
8. Chow, C.-Y.; Mokbel, M.F.; Aref, W.G. Casper*: Query processing for location services without compromising privacy. *ACM Trans. Database Syst.* **2009**, *34*, 1–48. [CrossRef]
9. Liu, S.; Wang, J.H.; Wang, J.; Zhang, Q. Achieving user-defined location privacy preservation using a P2P system. *IEEE Access* **2020**, *8*, 45895–45912. [CrossRef]
10. Ji, Y.; Gui, R.; Gui, X.; Liao, D.; Lin, X. Location Privacy Protection in Online Query based-on Privacy Region Replacement. In Proceedings of the 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 6–8 January 2020; pp. 0742–0747.
11. Perazzo, P.; Skvortsov, P.; Dini, G. On Designing Resilient Location-Privacy Obfuscators. *Comput. J.* **2015**, *58*, 2649–2664. [CrossRef]
12. Kachore, V.A.; Lakshmi, J.; Nandy, S. Location Obfuscation for Location Data Privacy. In Proceedings of the 2015 IEEE World Congress on Services, New York, NY, USA, 27 June–2 July 2015; pp. 213–220.
13. Qiu, C.; Squicciarini, A.C. Location Privacy Protection in Vehicle-Based Spatial Crowdsourcing Via Geo-Indistinguishability. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 7–9 July 2019; pp. 1061–1071.
14. Parmar, D.; Rao, U.P. Towards Privacy-Preserving Dummy Generation in Location-Based Services. *Procedia Comput. Sci.* **2020**, *171*, 1323–1326. [CrossRef]
15. Sun, G.; Chang, V.; Ramachandran, M.; Sun, Z.; Li, G.; Yu, H.; Liao, D. Efficient location privacy algorithm for Internet of Things (IoT) services and applications. *J. Netw. Comput. Appl.* **2017**, *89*, 3–13. [CrossRef]
16. Lu, H.; Jensen, C.S.; Yiu, M.L. Pad: Privacy-area aware, dummy based location privacy in mobile services. In *MobiDE '08: Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access, Vancouver, BC, Canada, 13 June 2008*; Association for Computing Machinery (ACM): New York, NY, USA, 2008; pp. 16–23.

17. Liu, X.; Liu, K.; Guo, L.; Li, X.; Fang, Y. A game-theoretic approach for achieving k-anonymity in Location Based Services. In Proceedings of the IEEE INFOCOM, Turin, Italy, 14–19 April 2013; pp. 2985–2993.
18. Niu, B.; Li, Q.; Zhu, X.; Cao, G.; Li, H. Achieving k-anonymity in privacy-aware location-based services. In Proceedings of the IEEE INFOCOM 2014—IEEE Conference on Computer Communications, Toronto, ON, Canada, 27 April–2 May 2014; pp. 754–762.
19. Liao, D.; Huang, X.; Anand, V.; Sun, G.; Yu, H. k-DLCA: An efficient approach for location privacy preservation in location-based services. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6.
20. Pingley, A.; Zhang, N.; Fu, X.; Choi, H.-A.; Subramaniam, S.; Zhao, W. Protection of query privacy for continuous location based services. In Proceedings of the IEEE INFOCOM, Shanghai, China, 10–15 April 2011; pp. 1710–1718.
21. Liu, J.; Jiang, X.; Zhang, S.; Wang, H.; Dou, W. FADBM: Frequency-Aware Dummy-Based Method in Long-Term Location Privacy Protection. In Proceedings of the 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS), Tianjin, China, 4–6 December 2019; pp. 384–391.
22. Niu, J.; Zhu, X.; Shi, L.; Ma, J. Time-Aware Dummy-Based Privacy Protection for Continuous LBSs. In Proceedings of the 2019 International Conference on Networking and Network Applications (NaNA), Daegu, Korea, 10–13 October 2019; pp. 15–20.
23. Yang, X.; Gao, L.; Zheng, J.; Wei, W. Location Privacy Preservation Mechanism for Location-Based Service with Incomplete Location Data. *IEEE Access* **2020**, *8*, 95843–95854. [CrossRef]
24. Sun, G.; Cai, S.; Yu, H.; Maharjan, S.; Chang, V.; Du, X.; Guizani, M. Location Privacy Preservation for Mobile Users in Location-Based Services. *IEEE Access* **2019**, *7*, 87425–87438. [CrossRef]
25. Hara, T.; Suzuki, A.; Iwata, M.; Arase, Y.; Xie, X. Dummy-Based User Location Anonymization under Real-World Constraints. *IEEE Access* **2016**, *4*, 673–687. [CrossRef]
26. Luo, C.; Liu, X.; Xue, W.; Shen, Y.; Li, J.; Hu, W.; Liu, A.X. Predictable Privacy-Preserving Mobile Crowd Sensing: A Tale of Two Roles. *IEEE/ACM Trans. Netw.* **2019**, *27*, 361–374. [CrossRef]
27. Zhou, L.; Yu, L.; Du, S.; Zhu, H.; Chen, C. Achieving Differentially Private Location Privacy in Edge-Assistant Connected Vehicles. *IEEE Internet Things J.* **2019**, *6*, 4472–4481. [CrossRef]
28. Lin, X.; Lu, R. Pseudonym-changing strategy for location privacy. In *Vehicular Ad Hoc Network Security and Privacy*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA; John Wiley & Sons: Hoboken, NJ, USA, 2015; Volume 1, pp. 71–90.
29. Guo, N.; Ma, L.; Gao, T. Independent Mix Zone for Location Privacy in Vehicular Networks. *IEEE Access* **2018**, *6*, 16842–16850. [CrossRef]
30. Al-Anwar, A.; Shoukry, Y.; Chakraborty, S.; Balaji, B.; Martin, P.; Tabuada, P.; Srivastava, M.B. PrOLoc: Resilient localization with private observers using partial homomorphic encryption. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, Pittsburgh, PA, USA, 18–21 April 2017*; Association for Computing Machinery (ACM): New York, NY, USA, 2017; pp. 257–258.
31. Negi, D.; Ray, S.; Lu, R. Pystin: Enabling Secure LBS in Smart Cities with Privacy-Preserving Top-*k* Spatial–Textual Query. *IEEE Internet Things J.* **2019**, *6*, 7788–7799. [CrossRef]
32. Farouk, F.; Alkady, Y.; Rizk, R. Efficient Privacy-Preserving Scheme for Location Based Services in VANET System. *IEEE Access* **2020**, *8*, 60101–60116. [CrossRef]
33. Ni, L.; Tian, F.; Ni, Q.; Yan, Y.; Zhang, J. An anonymous entropy-based location privacy protection scheme in mobile social networks. *EURASIP J. Wirel. Commun. Netw.* **2019**, *2019*, 93. [CrossRef]
34. Ying, B.; Makrakis, D. Protecting Location Privacy with Clustering Anonymization in vehicular networks. In Proceedings of the 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 27 April–2 May 2014; pp. 305–310.
35. Frejinger, E. Route Choice Analysis: Data, Models, Algorithms and Applications. Ph.D. Dissertation, Linköping University, Lausanne, Sweden, 30 April 2008.