EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

The research for protecting location privacy based on V-W algorithm

Xinyue Fan, Jing Tu, Chaolong Ye and Fei Zhou*

Abstract

With the development of mobile internet, protecting location privacy has already been an important issue. Based on previous studies and the drawback of traditional algorithms, the paper proposes a novel algorithm for protecting location privacy. The algorithm is based on the voronoi map of a road network, considers the problem of side-weight inference, and utilizes the information entropy as metrics. Meanwhile, the algorithm can defense the attack of side-weight inference and replay. Lastly, we verified the algorithm based on the real data of the road network. Results of experiences show that the improved algorithm has the better performance on some key performance metrics.

Keywords: Location privacy; Voronoi map; Road network; Information entropy

1 Introduction

With the rapid development of mobile network, the leakage of location privacy has become a problem that cannot be ignored. So lots of researchers have paid more attention on it. So far, researches on protecting location privacy mainly focus on two aspects: The infrastructure of a location privacy protection system and the method of protecting location privacy. Now, the mainstream architectures of a location privacy protection system are usually divided into three categories: non-cooperative architecture, distributed peer-to-peer architecture, and centralized architecture [1].

The existing methods about location privacy protection mainly include false-name location privacy protection, landmark location privacy protection, false-address location privacy protection, and spatial anonymity location privacy protection. False-name location privacy protection replaces the real identity of a user by a false name so that attackers cannot obtain the real query source from location-based services (LBS) severs [2]. Landmark location privacy protection only utilizes some landmark position within a certain range instead of a user's real position to interact with LBS, so attackers cannot identify the real position of users [3]. Through the further research on landmark method, the current method of protecting location privacy based on a landmark not only uses landmark position to replace the real position of a target, but also utilizes some algorithms to give a false position to replace the real position of users. For instance, literature [4] adopts an incremental nearest neighbor query algorithm to realize location privacy protection. In literature [5], an anonymous regional transformation algorithm is adopted. In the method of false-address location privacy protection, a position information set, that includes the real position of a user and a sequence of a false position (dubbed as dummy), is sent to LBS, so attackers have no way to distinguish the real position information from the received information [6,7]. Spatial-anonymity location privacy protection is also a very interesting method. An anonymous server makes user information anonymous to obtain a fuzzy space or user set and then uses the fuzzy space or user set as a requester body to interact with an LBS server. So far, K-Anonymous algorithm proposed by Marco Gruteser is the most classic in spatial-anonymity location privacy protection [8]. Based on K-Anonymous algorithm, some scholars propose several improved algorithms aimed at different performance demands. For instance, literature [9,10] propose personalized K-Anonymous algorithm based on personalized demands of users. The algorithm can adjust the requested k value for users according to their security demand. To solve the problem of a low anonymous success rate

*Correspondence: zhoufei@cqupt.edu.cn
Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongwen Road, 400065 Chongqing, China

of a traditional algorithm, Xiao et al. designed an efficient directed graph based on choking algorithms [11]. The Casper algorithm in [12] mainly considers a large-scale requested condition and personalized anonymous demand.

However, the current researches mostly focus on Euclidean geometric space. In fact, a road network environment limits the user's activity in most cases. Meanwhile, it is very meaningful to make further research on location privacy protection based on the road network. In [13], Rubner gives a method that transforms Euclidean geometric space into the road network environment. Literature [14] gives an X-STAR algorithm. The algorithm regards road crossing point as a node and conducts an anonymous process to users. It can simultaneously make an anonymous set satisfy two conditions, k-anonymous of location and l-diversity of road. According to the road network environment, an anonymous ring, anonymous tree, or anonymous cellular can also be adopted to realize location privacy protection [15,16]. In addition, Zhao et al. give an anonymous method based on a voronoi map. The anonymous method can be fulfilled in a v zone after satisfying the condition of K-Anonymous and l-diversity. The above algorithms have their own advantages, but their shortcomings are also obvious. For example, the algorithm in [13] has too low security. The anonymous success rate in [14] is too low, and its computation complexity is too high. In [15], a too large anonymous ring may result in the serious degradation of QOS. Algorithms in [16,17] do not consider the attack of side-weight inference, so their security is not enough. Therefore, in this paper, we consider the problem of side-weight inference and utilize the information entropy as metrics, then propose a novel algorithm based on voronoi-weight (V-W) for protecting location privacy.

The remainder of this paper is organized as follows. We analyze the related issues of a traditional algorithm and determine the design target of our algorithm in Section 2. The complete design is given in Section 3. Section 4 presents the performance evaluation result of our proposed algorithm. Finally, the paper is concluded in Section 5.

2 Problem proposed

2.1 The comparison of algorithms

As far as we are concerned, due to a simple architecture, centralized architecture is the most popular. In this architecture, the interaction between mobile terminal and anonymous severs is encrypted to guarantee the security of user requests. The data interaction between anonymous severs and an LBS server adopts plaintext transmission to save system resource.

In this paper, we mainly consider the shortcoming of traditional anonymous ring algorithms and anonymous cellular algorithms and improve the existing location privacy algorithm based on the road network. The advantages of an anonymous ring algorithm is that an attacker starts searching from any path by using the anonymous ring algorithm; the final output of the anonymous ring or anonymous tree are the same. So the algorithm has good anti-replay attack capability. However, the anonymous ring algorithm only considers if there are mobile users at two sides of a ring and ignores the distributed probability of users. If the user's distribution of each side is not the same, the user is easily attacked by side-weight inference. In addition, if the ring or road is too long, the output of anonymous road sets would cover a too large region and result in a too heavy computational burden.

Different from the anonymous ring algorithm, anonymous cellular algorithm does not seek an anonymous ring as an anonymous road set and utilize the vertex, which degree is larger than the default threshold, to construct an anonymous cellular to guarantee the l-diversity of the road. The output of the algorithm is the user set that satisfies k-anonymous and l-diversity of the road. The method dividing the road network into anonymous cellular and not utilizing the whole road largely reduces the coverage of the anonymous set and system overhead. Simultaneously, selecting the right vertex can make an anonymous area satisfy the l-diversity of road and improve the anonymous success rate. The ability of anti-replay attack can be improved by randomly selecting a user within an adjacent road to join the anonymous set. However, because the cellular adopts round topology, it cannot cover the whole road network completely. If the user is located in the area where cellular does not cover, its position would be leaked, such as U_6 in Figure 1. And the cellular usually selects the half length of the longest road as radius, so it would produce lots of overlap regions and reduce anonymous efficiency. For instance, U_{18} and U_{21} belong to two cellulars simultaneously, so cellular selecting is a prerequisite to an anonymous process. Additionally, the algorithm has no restraint for a path where the user in the anonymous set lies, so it is easy to be attacked by side-weight inference. For instance, in the n_{12}-centered cellular, if U_{14} requests location services with the l-diversity demand of a road section ($l = 2$), the user in the anonymous set may come from the n_{12} n_{13} road section or n_{12} n_{15} road section. Because of the probability that the user lies in n_{12} n_{13} is higher than that in n_{12} n_{15}, an attack may inference the real position of the user and result in user privacy leakage.

In general, aimed at the drawback of traditional algorithm, we summarize the following problem:

1) Assuming that an attacker already intercepted anonymous servers' requested information from an

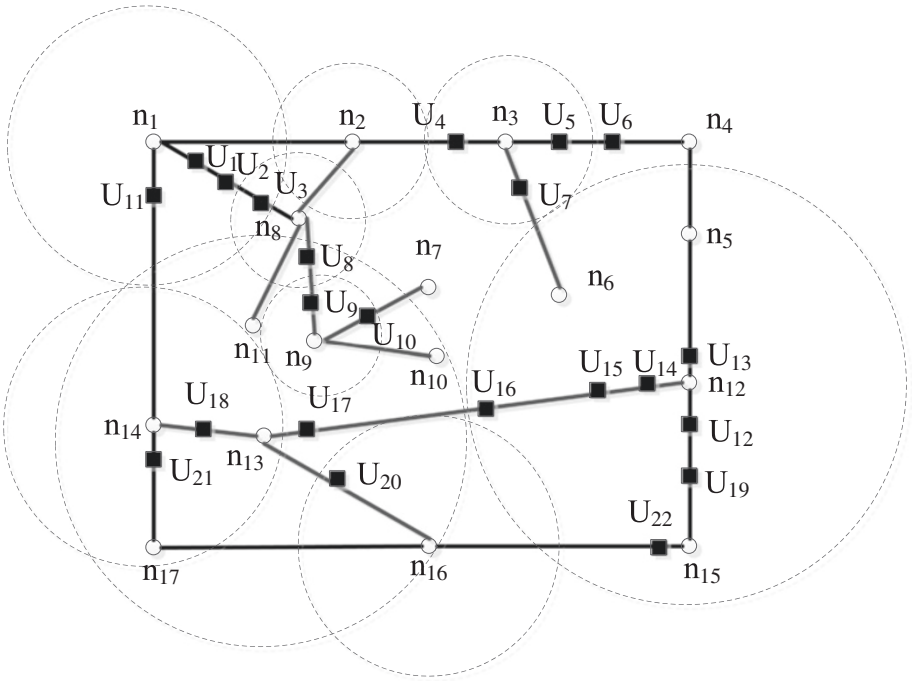

Figure 1 The model of anonymous cellular network. Select the vertex whose degree is larger than the default threshold as center point and select the half length of the longest road as radius to draw a circle. All of the circles are anonymous cellular zones, and the number of roads is greater than or equal to the degree of vertex in every zone.

LBS server, how to guarantee that the attacker cannot obtain the correct location information?

2) How to guarantee that the server can defense the replay attack and side-weight inference attack effectively?

2.2 The design target of algorithm

Based on the above problem, we give a novel location privacy protection algorithm based on the road network. The design targets of the algorithm are as follows:

Target 1 The output anonymous set of algorithm satisfies the k-anonymous demand. Given that an attacker obtain the requested information that an anonymous server submitted to an LBS server, the request does not contain the real position information of a user but contain the anonymous road section where the user lies. Because there are k_i mobile users in a set at least, the probability that the attacker can identify the position of the user is not higher than $1/k_i$, that is $k \geq k_i$, where k is the number of users in an anonymous road set, k_i denotes as the number of the user requests.

Target 2 The output anonymous road set satisfies the l-diversity demand of the road. Assuming that an attacker intercepts the request and obtains the anonymous road set. However, the set at least contains l_i path, so the probability that the attacker can identify the correct path of a user from the set is not higher than $1/l_i$, that is $l \geq l_i$, where l is the number of roads in the anonymous road set, and l_i denotes as the l-diversity demand of U_i.

Target 3 The algorithm has a good ability to defense the replay attack.

Target 4 The algorithm has a good ability to defense the side-weight inference attack.

Replay attack and side-weight inference attack are the most common attack methods to location privacy and also pose the greatest threat on location privacy. Therefore, the defense abilities against these threats are the main design target of our algorithm.

Commonly, for a more deep analysis of a replay attack and side-weight inference attack, we often assume that an attacker already obtained the request information submitted by an anonymous server, having the same knowledge of the road network with the anonymous server and knowing the anonymous algorithm run by the anonymous server. A replay attack is when an attacker determines the path where a user lies by a re-run anonymous algorithm. The method firstly utilizes anonymous road set S to

determine all possible roads, $S = \{e_1, e_2, \cdots, e_i, \cdots, e_n\}$, and respectively assumes the requested user U_k to lie upon road e_i, where $i = 1, 2, \cdots, n$, then utilizes the anonymous algorithm to process each assumption. Finally, we can obtain anonymous road set S_i and determine the same road between S and S_i by $S \cap S_i$. If $N(S \cap S_i)$ denotes the number of the same road, the ratio n_i can be obtained by Equation 1.

$$n_i = \frac{N(S \cap S_i)}{N(S)} \tag{1}$$

So, we also derive the replay probability p_{ic} that user U_k lies upon road section e_i, that is:

$$p_{ic} = \frac{n_i}{(n_1 + n_2 + \cdots + n_i + \cdots + n_n)} \tag{2}$$

After attackers compute the probability of each assumption, they commonly regard the road with the maximum replay probability as the targeted road that a user lies in, that is:

$$e_{U_k} = \max\{e_{p_{1c}}, e_{p_{2c}}, \cdots, e_{p_{nc}}\} \tag{3}$$

As for a side-weight inference attack, the method utilizes the uneven distribution in road to determine the correct road section where a user lies in. Similarly, we also assume the attack probability in the same anonymous road set is equal to $1/k$, so the uneven distribution of the user in each road makes the probability that the attacker successfully inferred, also called the probability of side-weight inference p_{ib}, not be $1/l$ anymore, but the ratio between the number of user in ith road W_{e_i} and the total number of user in a set, that is:

$$p_{ib} = \frac{W_{e_i}}{(W_{e_1} + W_{e_2} + \cdots W_{e_i} + \cdots + W_{e_n})} \tag{4}$$

3 Algorithm implementation

The V-W algorithm proposed in this paper is based on centralized architecture. Figure 2 gives the whole work flow.

Before the user first submits the location service request to the anonymous server, registration information should be submitted to the anonymous server; the format is $[ID, Loc(x, y), 0]$, where ID represents the user identity,

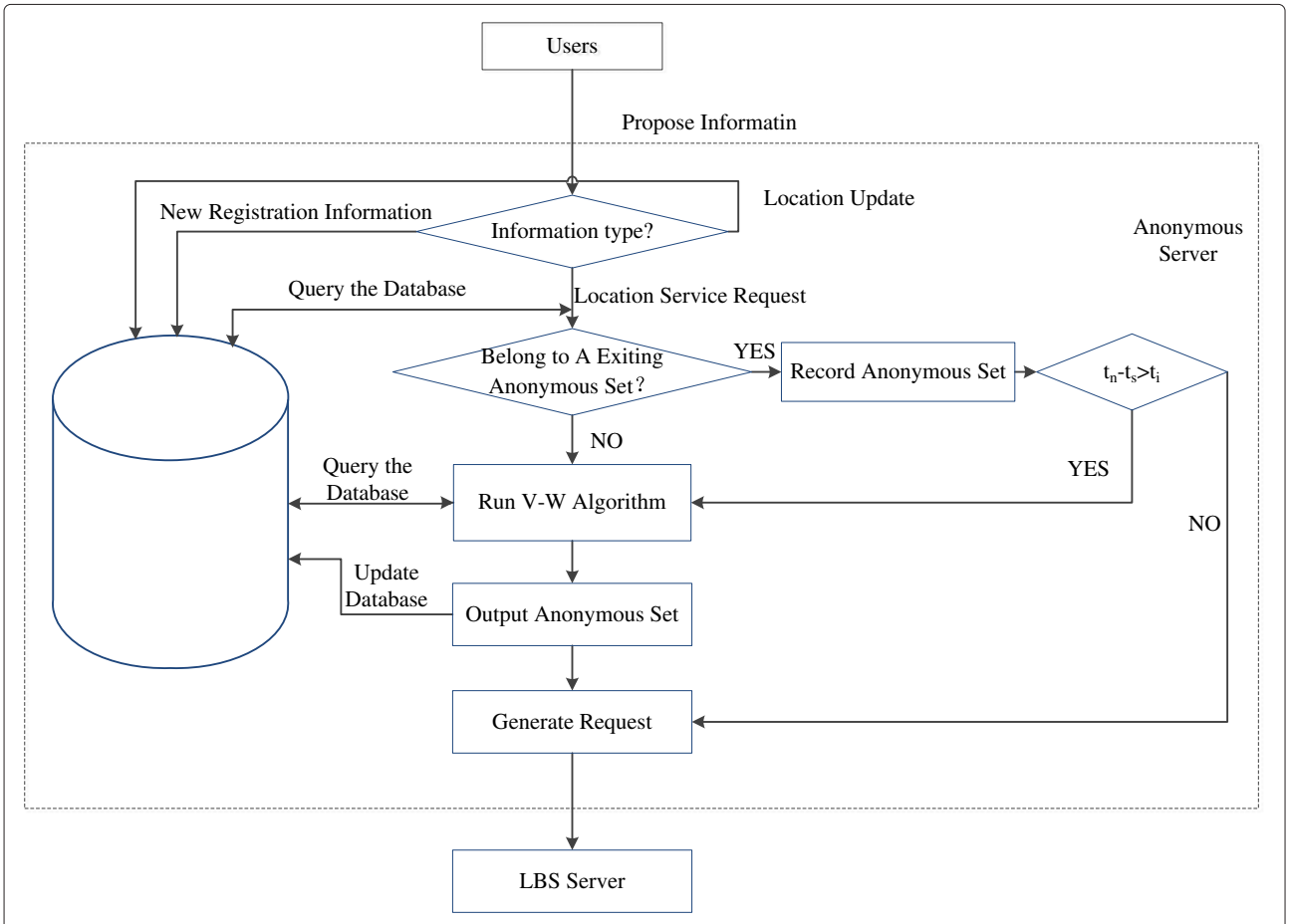

Figure 2 The flow of an anonymous request. The users submit the location service request to third party anonymous servers through the IPsec encryption tunnel. With the operation of V-M algorithm in the anonymous servers, the servers submit the anonymous location service request to LBS.

$Loc(x, y)$ represents the real user's location information, and 0 means that this information is the registered information. Another kind of information that the user submits to the anonymous server represents location update; the format is $[ID, Loc(x, y), 2]$, where ID represents the user identity, $Loc(x, y)$ represents the user's new location information, 2 means that this information represent the location update. The user needs to submit the information to update the user location information in the server regularly and thus ensure the quality of user service.

In Figure 2, to judge whether the user belongs to an effective anonymous road set, we should consider whether the set's time is effective, that is $t_n - t_s > t_i$, where t_n is the current time, t_s is the generation time of an anonymous set, and t_i is the user's tolerance time. If the set's time is not effective, the V-W anonymous algorithm is returned. Therefore, it would update the anonymous road set of the corresponding user and generate the requested information submitted to LBS. However, if the set's time is effective, the request information can be generated directly by using the anonymous set.

The key step of an anonymous system is the design of V-W algorithm. The main task of the algorithm is the three-stage search for an anonymous road set. The first stage, pre-processing stage, adopts a voronoi map to divide the road network [17]. The second stage, that named as road search stage, mainly implements the search in the output road set of the pre-processing stage. The third stage is extended search stage. If the second stage cannot find the targeted road, the extended search must be implemented in the neighboring area.

Based on the anonymous cellular algorithm, the simplest way to guarantee the diversity of a road is by dividing the areas of the road network, which ensures each area includes l road at least. So the pre-processing stage must implement area dividing to the road network. In the V-W algorithm, the road network can be regarded as an undirected graph $G(V, E)$, which is also called as a road network map and is combined with line set $E = \{e_1, e_2, \cdots, e_n\}$ and node set $V = \{v_1, v_2, \cdots, v_n\}$. The voronoi map is also called as Tyson polygon and adopted in GIS field.

A voronoi map consists of serials of continuous polygons whose edges are the midnormals of consecutive points on the plane. The center of each voronoi polygon area is also the endpoints of the perpendicular bisector, as shown in Figure 3, such as n_8 and n_1, the edges of the voronoi polygon area are also the perpendicular bisectors of the connectors between the center and the consecutive points. The edges of the voronoi polygon which centers on n_8 are formed by the midnormals of n_8 and n_1, n_2 and n_9.

The V-W algorithm adopts the voronoi map to divide the road network. Hence, it not only ensures each user to link the corresponding voronoi polygon, but also reduces the search coverage for finding user U_k effectively and improves the quality of service. For convenience of description, we denote $V(V, E)$ as the corresponding voronoi map. Combining the real situation of the road network environment and the user demand

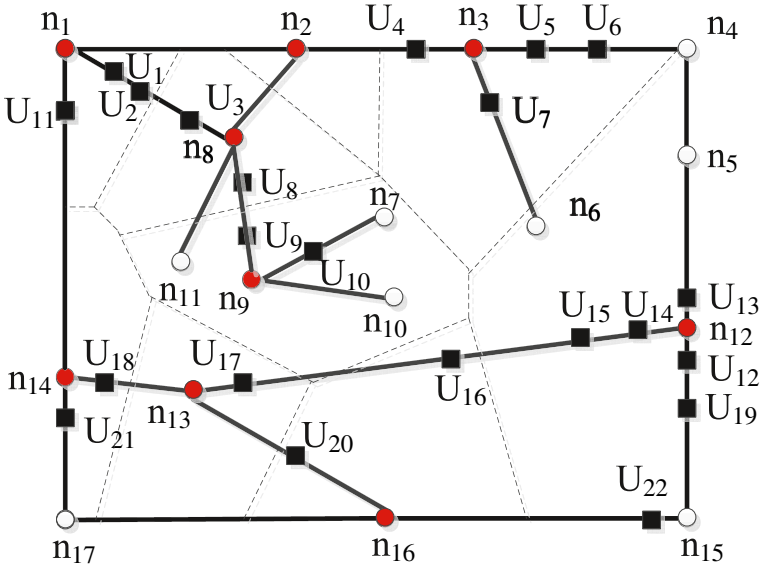

Figure 3 The voronoi map about a simple road model. The red nodes represent the nodes whose degree must be greater than 3 in the road network, at least three roads in each voronoi polygon can be guaranteed by the voronoi map which is drawn at the center of these nodes.

of road diversity, V-W algorithm selects the suited vector (V, E), which node metric must be greater than 3 in $G(V, E)$.

The node metric is defined as the number of road that crosses through the related node. So each voronoi polygon at least contains three roads. Figure 3 represents the voronoi map about a simple road model.

In the V-W algorithm, we regard a voronoi polygon as a v zone. We would implement the search for the related v zone in the second and third stages. Hence, in the pre-possessing stage, we need find the corresponding $V(V, E)$ and map it into the related $G(V, E)$. If so, we can determine the road contained in each v zone and map the user position into the corresponding $G(V, E)$ and $V(V, E)$ while the system receives the user's registration information and updated location information. In other words, we can locate the user to the corresponding v zone and the corresponding road. For instance, in Figure 3, if User U_{13} requests, the system can find the v zone where the user lies according to the user information and find the corresponding road $V_i = \{n_4n_5, n_5n_{12}, n_{12}n_{13}, n_{12}n_{15}, n_{15}n_{16}\}$. After road mapping, we may adopt a quad-tree way to code the corresponding v zone in order to search for the neighboring v zone in the third stage.

After the road dividing and information mapping in the first stage, we would implement a search for the anonymous road set in the corresponding v zone where U_k lies. In view of too many users in the real situation, we would adopt a special search method, described in Figure 4. The search algorithm is as follows:

Step I Locate the v zone where the user lies and sort all roads in v zone by its own weight. The road weight w_e is the number of users in the targeted road. For example, the result of road sorting in Figure 4 is as follows:

$$e_1(7) \rightarrow e_4(6) \rightarrow e_5(6) \rightarrow e_2(5) \rightarrow e_6(5) \rightarrow e_3(3) \rightarrow e_7(2)$$

where the value within brackets $(\cdot)$ is road weight.

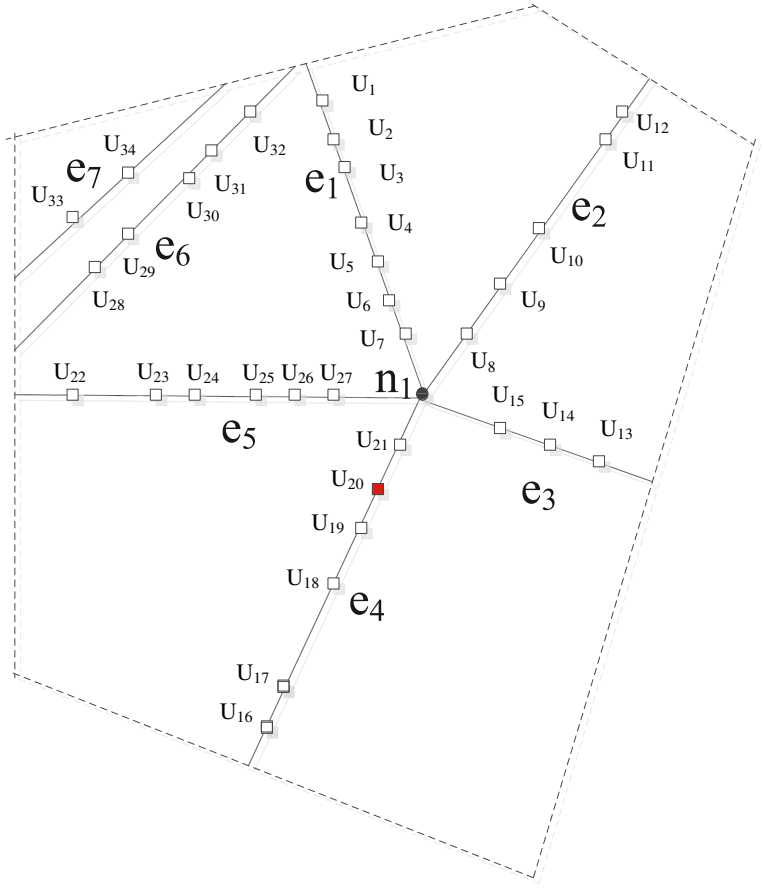

Figure 4 The v zone map. In the V-M algorithm, we regard a voronoi polygon as a v zone. The algorithm takes v zone as search unit. In the v zone map, the road weight is the number of users in the targeted road.

Step II Locate the road where the user lies. For instance, if user U_{20} requests, the road where the user lies is e_4 and is joined to quasi-anonymous set S', $S' = \{e_4\}$. Then, select the maximum k and l of all users in road and assign these values to the anonymous demand of system, k_s and l_s. For convenience of description, let us assume that the anonymous demand of user U_{20} is $k_{20} = 10$, $l_{20} = 3$ while the anonymous demand of user U_{19} is $k_{19} = 12$, $l_{19} = 3$, so the system should assign k_s and l_s, $k_s = 12$, $l_s = 3$.

Step III Select the road randomly from $l + \partial$ roads those that are adjacent to the targeted road to join quasi-anonymous set S', where l corresponds to the number of roads which are adjacent to the road where the user lies in, ∂ corresponds to random factor. Commonly, the system assigns ∂ as a certain value by default and guarantees the randomness of the anonymous road set that the user selected. Therefore, it can defense the replay attack. For instance, we randomly select six roads from the neighboring areas of e_4, $(e_1, e_4, e_5, e_2, e_6, e_3)$, to join S'. Then we modify k_s and l_s of the system to the maximum demand of all users in S', that is $k_s = \max\{k_i\}$, $l_s = \max\{l_i\}$. When the selected road e_2 is joined to the quasi-anonymous set, we can obtain $S' = \{e_4, e_2\}$. Then we need to compare the k_i and l_i of all users who are in e_4 and e_2. If only U_{10} in S' satisfies $l_i = 4 > l_s$, the system would update l_s and select a road from the updated road within $l_s + \partial$ area, l_s represents the number of roads which are adjacent to the road where the user lies after updated quasi-anonymous set S', ∂ corresponds to random factor. After updated the anonymous demand of system, system need meet the conditions, $n_k \geq k_s$ and $n_l \geq l_s$ where n_k is the number of user and n_l is the number of road. If not, system need select road again from the remaining road within $l_s + \partial$. The above process would be repeated until the conditions, $n_k \geq k_s$ and $n_l \geq l_s$, are met. If the algorithm cannot satisfy $n_k \geq k_s$ and $n_l \geq l_s$, even though the number of the roads of all the candidates sets an increase, the system would return the anonymous failure information. In view of the introduced random value, the anonymous process has the following three cases:

i) $l_s + \partial \leq n_v$. Where n_v corresponds to the total number of road in the v zone. In this case, as the above example shows, the system may select a road from $l_s + \partial$ area directly.

ii) $l_s + \partial > n_v$ and $l_s < n_v$. When the conditions are met, we may ignore ∂ and randomly select the road to join S' from all the roads in the v zone. In other words, we need not extend the v zone.

iii) $l_s > n_v$. When the condition is met, we need extend the v zone by combining the neighboring v zone, then repeat the procession of case i) or ii) according to the updated l_s. In addition, if $k_s > v_k$, we also need to extend the v zone. v_k corresponds to the total number of users in the v zone.

Step IV After obtaining the suited anonymous set S' that satisfies $n_k \geq k_s$ and $n_l \geq l_s$, we need to determine if S' met the conditions that can defend the attack of side-weight inference. The condition is as follows.

$$\max(p_{ib}) \leq \delta, i \in (1, n_l) \tag{5}$$

where p_{ib} is the probability of side-weight inference and δ is the experience value (here, $\delta = 0.5$), which denotes the probability threshold of side-weight inference. If the probability of inference is greater than the threshold, it represents that the position privacy of the user may be attacked by side-weight inference.

In Equation 5, the side-weight inference probability of all roads would be computed. If the probability of all roads is less than the threshold, the quasi-anonymous set S' would be regarded as anonymous set S. If the situation that a certain probability is greater than the threshold δ happened, we need to modify the set S' by adding a new road from $l_s + \partial$ roads until Equation 5 is satisfied. If all roads in $l_s + \partial$ roads are already added to the set S' completely, Equation 5 can still not be satisfied. We would select the new road from the v zone until Equation 5 is met. However, when we run out of all roads in the v zone, Equation 5 still cannot be satisfied, the system would return the anonymous failure information. For convenience, we may assume that quasi-anonymous set $S' = \{e_4, e_5, e_2, e_3\}$ of user U_{20} can be obtained by Step III, we have, $\max(p_{ib}) = p_{4b} < \delta$. Hence, the system would regard the output S' as the anonymous set of user U_{20} and update all users' anonymous set to the output S'. The anonymous condition of all users within the set would be satisfied. At the same time, if another user submits a request within the time tolerance $t_n - t_s \leq t_i$, the system also regards the set S' as the anonymous set of the requester.

The method can make the system avoid repeated computing for other users within the tolerance time and reduce the processing time and overhead of the system. On the other hand, even if an attacker obtains the anonymous set, because all users use the same set, it can confuse the attacker and enhance the ability of an anti-replay attack of the system.

The third stage of the algorithm needs to realize the extension of the v zone when case iii) in the second stage is meet. The system utilizes the linear quad-tree to find the neighboring v zone of the current v zone. Then, the system would combine two zones and conduct the search process in stage 2 repeatedly. Figure 5 shows the whole flow of the algorithm.

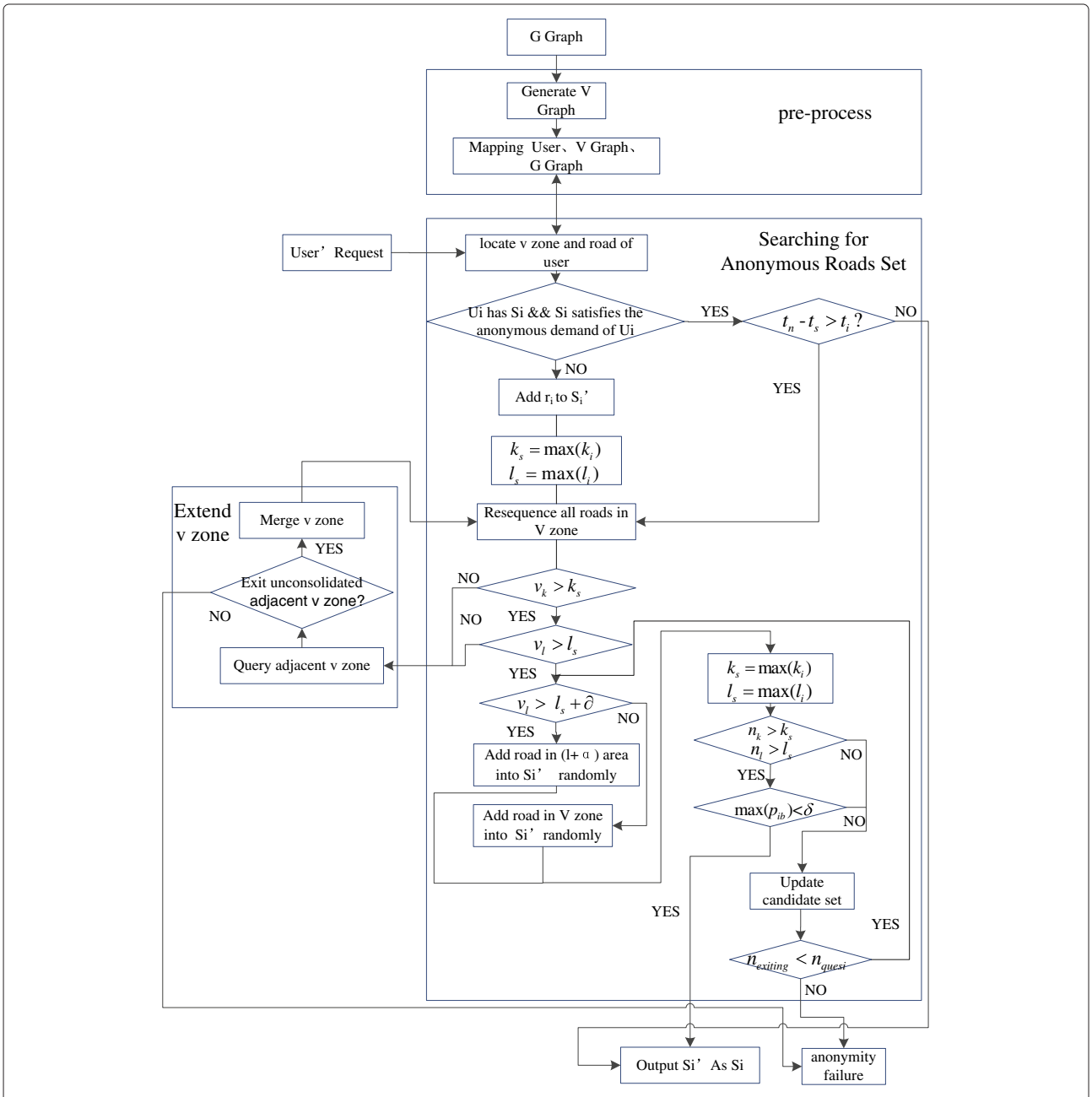

Figure 5 The flow of the V-W algorithm. The V-M algorithm is made up of three stages. The first stage, pre-processing stage, adopts the voronoi map to divide the road network. The second stage, road search stage, mainly implements the search in the output road set of the pre-processing stage. The third stage, extended search stage, is if the second stage cannot find the targeted road, the extended search must be implemented in the neighboring area.

In the algorithm, the core code about a search step is as follows:

Algorithm 1 The search steps of the V-W algorithm

Input:

U_i, v_i, r_i

Output:

S_i

```
if (U_i ∈ S_i&&S_i satisfies the anonymous demand of U_i) then
  if (t_n − t_s > t_i) then
    print S_i
    return
  end if
else
  Add r_i into S'_i; // S'_i is quasi-anonymous set
  Assign k_s and l_s by k_s = max(k_i), l_s = max(l_i);
  Resequence all roads in v_i;
  if (v_k < k_s&&v_l < l_s) then
    if (v_l > l + ∂) then
      Select l + ∂ roads as candidate set;
      Add road in (l + ∂) area into S'_i randomly;
    else
      Select all roads in v zone as quasi-set;
      Add road in quasi-set into S'_i randomly;
    end if
    Update k_s and l_s by k_s = max(k_i), l_s = max(l_i)
    if (n_s > k_s&&n_l < l_s) then
      if (max(p_ib) < δ) then
        print S'_i as S_i
        return
      end if
    end if
    update S'_i
    if (n_obtained < n_candiated) then
      loop
        11
      end loop
    else
      print Output anonymity failure
      return
    end if
  else
    Extend v zone
    loop
      9
    end loop
  end if
end if
```

4 Simulation and analysis

4.1 Experimental environment and data

Table 1 gives the parameter of the environment, as follows:

Table 1 Test platform parameters

Category	Parameter
Hardware platform	CPU Intel Core T660, main frequency 2.2 GHz
Software platform	Eclips
Operation system	Windows XP professional
Database	Oracle
Coding language	Java

We used the PC to imitate the anonymous server to evaluate the performance of the proposed V-M algorithm.

To obtain the data of the road network, this paper applies the tools, network-based generator of moving objects, of Thomas Brinkhoff. We select the part of the map of Chongqing City of China. The road network is shown in Figure 6. In this test, we only consider a snapshot on the database instead of considering the continuous attack, so the user data within a single time interval is our concern. Here, we set the number of users within a single time interval as 10,000, the anonymous demand of a user's request is 3 to 15; the diversity demand of the road is 3 to 15.

Figure 7 produced in pre-process stage is the voronoi map of Chongqing City.

4.2 Simulation result and analysis

4.2.1 The success rate of anonymity

The success rate of anonymity represents the search probability that an anonymous sever searches the user successfully when the user requests a positioning service. It mainly evaluates the operating efficiency of the system. The higher the success rate of anonymity, the better the system performance. Equation 6 is the representation of the success rate of anonymity. Where N_s denotes the number of users whose name are protected successfully. N_t is the total number of requesters for an anonymous demand.

$$R_k = \frac{N_s}{N_t} \tag{6}$$

Figure 8a shows the change trend of the success rate with the change of the average demand of k anonymity. From Figure 8a, we can conclude that the success rate of anonymity would decline slightly with the increase of the demand of k anonymity. The reason is that the increase of k would result in the increase of probability of $n_{\text{obtained}} \geq n_{\text{candidate}}$, that is, all the candidate roads cannot satisfy the demand of k anonymity, so the system would output the failure information. Figure 8b shows the similar change trend of the success rate of anonymity with the change of l-diversity of the road. Simultaneously, we also conclude that the efficiency of the algorithm proposed in this paper

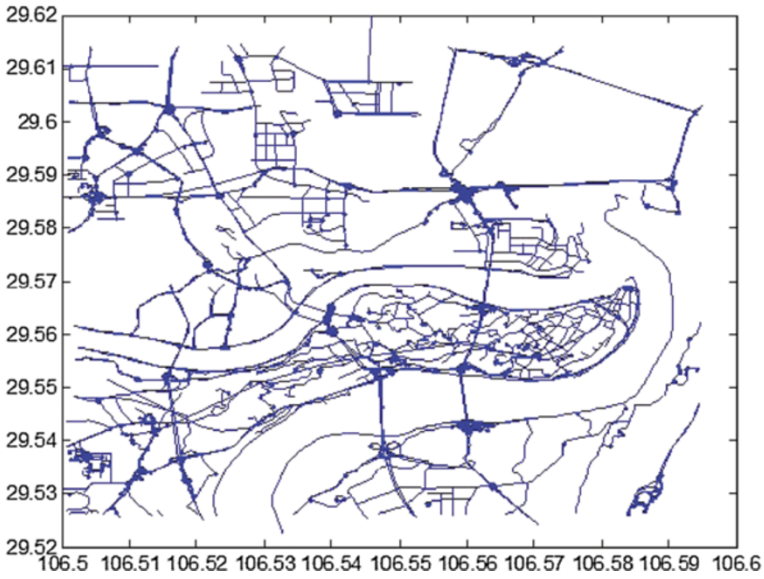

Figure 6 The road network of Chongqing. The road network of the city is comprised of 7,320 nodes and 8,130 sides. The average metric of a road node is 3.1.

is lower than that of the anonymous ring algorithm and anonymous cellular algorithm. The reason is that anonymous ring algorithm and anonymous cellular algorithm do not consider the side-weight attack, so some sets that cannot defend the side-weight attack are also regarded as the successful cases. Therefore, the algorithm in this paper achieves the higher security at the cost of the declination of success rate of anonymity. The proposed V-W algorithm and the other algorithms cannot guarantee a one hundred percent success rate, so sometimes anonymous failure is acceptable in practice.

4.2.2 The average time of anonymity

The average time of anonymity is also one of the important indexes. It indicates the average time consumed of one times anonymity. The less the average time, the better the system performance. We can obtain the average anonymous time by Equation 7, where T_C denotes the

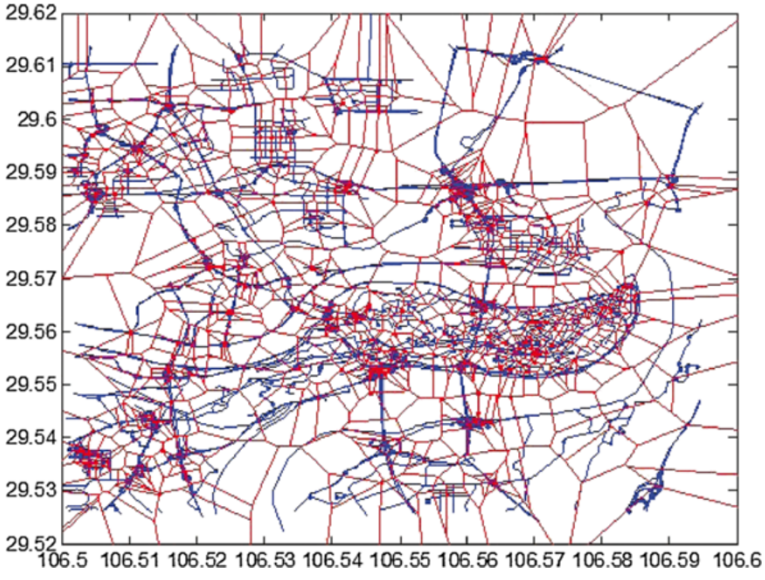

Figure 7 The corresponding voronoi map of Chongqing City. In this voronoi map, the node with a metric greater than 3 is selected as the central node of v zone, that is, each v zone at least contains 3 roads.

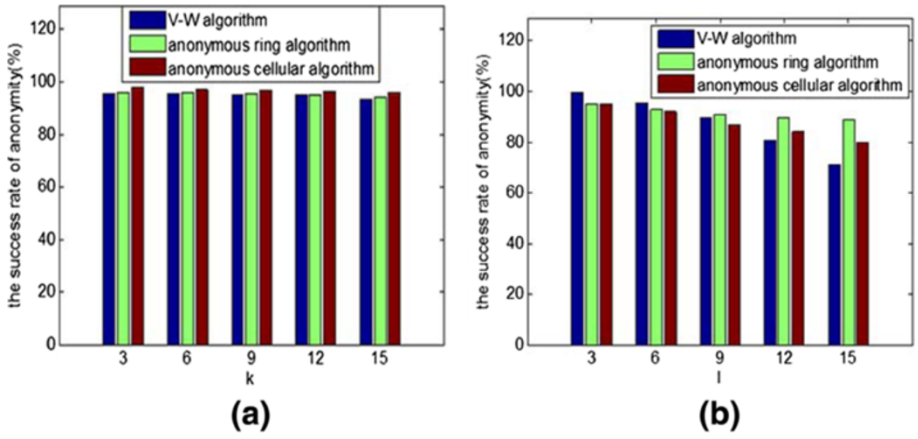

Figure 8 The success rate comparison of anonymity. From **(a)**, we can conclude that the success rate of anonymity would decline slightly with the increase of the demand of k anonymity. **(b)** The figure shows the similar change trend of the success rate of anonymity with the change of l-diversity of road.

average time of anonymity, t_{ci} is the time consumption of the ith user's anonymity, and N corresponds to the total number of anonymous user.

$$T_C = \frac{1}{N}\sum_{i=1}^{N} t_{ci} \tag{7}$$

From Figure 9a, we can see that the time consumption would increase with the increase of k values. If the demand of k anonymity increases, the system would cost more time to search. Especially, when the zone cannot satisfy the demand of anonymity, the system needs to extend the search area. Similarly, we can obtain the same conclusion to the l-diversity of the road. Comparing with the anonymous ring algorithm and anonymous cellular algorithm, the V-W algorithm in this paper has the smaller time consumption. This is because the average time of anonymity does not contain the time of pre-process in this paper. However, the pre-process stage should be a part of the whole anonymous process. Although the pre-process will consume a part of the time in practice, the voronoi map which is generated after the metric is confirmed and can be used many times. In other words, if the voronoi map already exists in practice, the system will enter into the anonymous search stage without the pre-process stage. Another reason is that the system need not compute the anonymous set repeatedly within the time tolerance and return the existing set when the user in the same road set requests.

4.2.3 *The relative anonymous rate*

The relative anonymous rate is also an important norm. It can be divided into k-relative anonymous rate RAL_k and l-relative anonymous rate RAL_l. RAL_k represents the

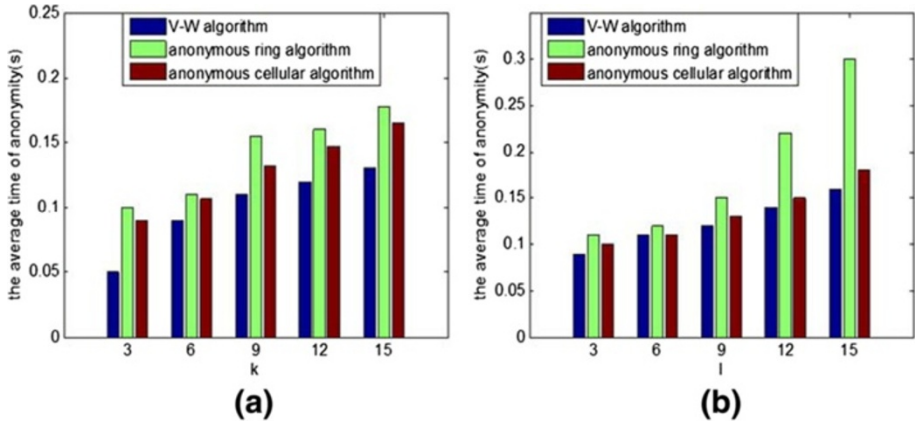

Figure 9 The average anonymous time comparison. From **(a)**, the average anonymous time would increase with the increase of k values. **(b)** The figure shows the similar change trend of the average anonymous time with the change of l-diversity of road.

average value of the ratio between the number of user n_{ik} in the anonymous set and the k-anonymous demand of the user, as shown in Equation 8. The larger the RAL_k, the higher the user's security. The calculation of RAL_l is the same with RAL_k.

$$\mathrm{RAL}_k = \frac{1}{N}\sum_{i=1}^{N}\frac{n_{ik}}{k_i} \tag{8}$$

In Figure 10a, with the increase of the average demand of k-anonymity, RAL_k would decrease; however, the total number of users in the set would not increase largely. The result of Figure 10b is similar with that of Figure 10a, so the result analysis is not repeated again. From Figure 10c, it gives the change trend of RAL_k with the change of l. Because there are so many users in the road, the increase of l would result in the increase of k drastically, then cause the increase of RAL_k. It can provide the better protection for users. In Figure 10d, when l increases, RAL_l would decrease slightly. This is because the algorithm in this paper need to consider the attack of side-weight inference. If the requirements of the side-weight inference cannot meet, the quasi-anonymous set would be enlarged by adding a road. It would cause RAL_l larger. However, if the demand of l-diversity is too large, the requirement of the side-weight inference would be met. So RAL_l would

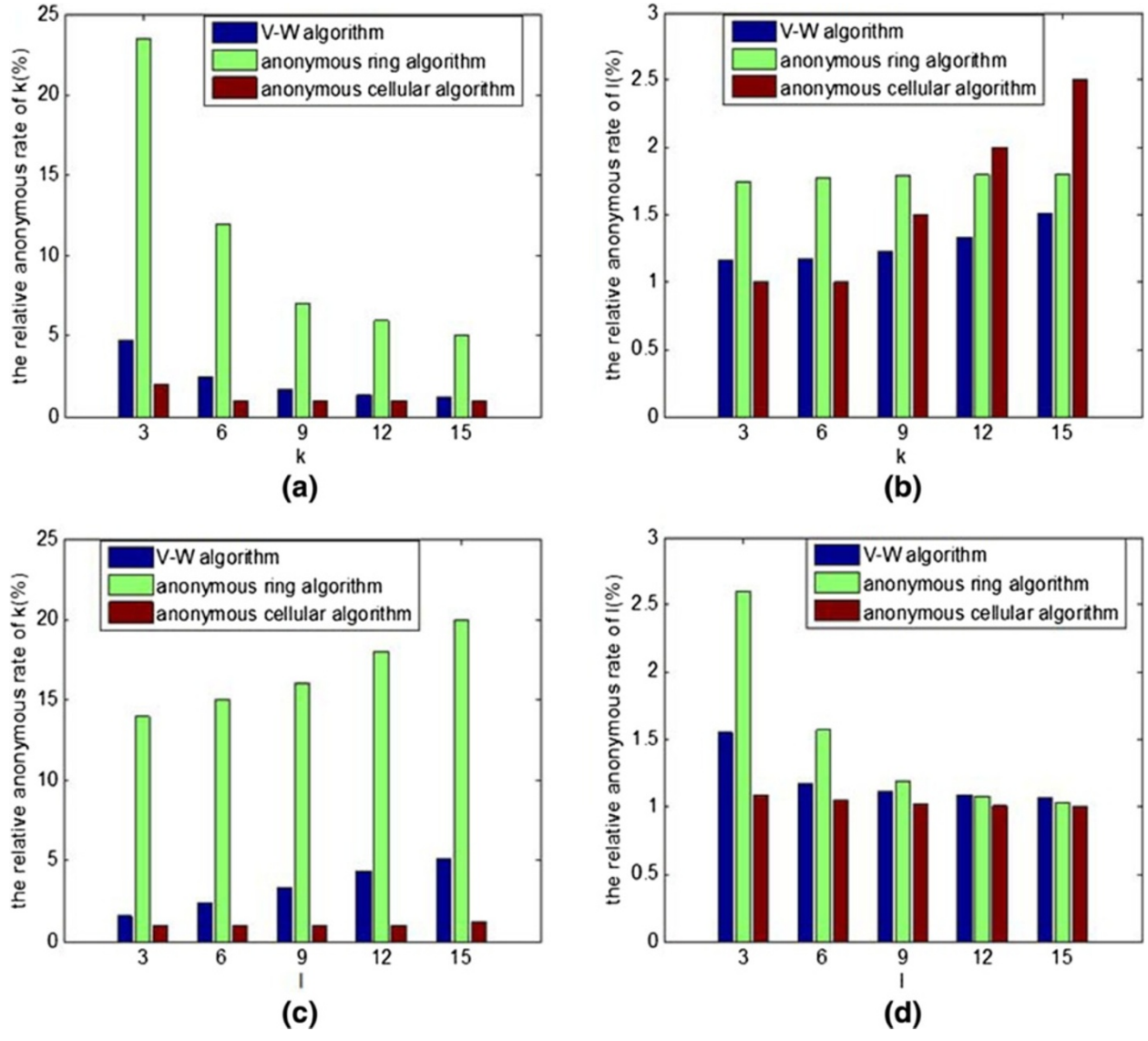

Figure 10 The system relative anonymity comparison. From **(a)**, the RAL_k would decrease with the increase of k values. From **(b)**, the RAL_l would increase with the increase of k values. From **(c)**, the RAL_k would increase with the increase of l values. From **(d)**, the RAL_l would decrease with the increase of l values.

decrease. In the V-W algorithm, in the same situation, RAL_k is greater than the value of the anonymous ring algorithm and anonymous cellular algorithm. The reason is that k and l selected in this algorithm are the largest demand in the anonymous set, which causes the increase of the anonymous rate. Simultaneously, because the number of roads contained in anonymous ring is usually greater than the demand of l-diversity, RAL_l is less than that of the anonymous ring algorithm in the same situation.

4.2.4 The average entropy of a replay attack

The average entropy of a replay attack usually represents the ability of an anti-replay attack and is one of parameters that evaluate the system security. An attacker infers the probability p_{ic} that a user laid into a certain road by a replay attack, as shown in Equation 2. The available information entropy H_{ci} usually represents the uncertain metric of a replay attack. Obviously, the higher the entropy of replay attack, the larger the uncertain metric that the attacker infers the location of the user. Hence, our algorithm can provide the stronger ability of an anti-replay attack. Equation 9 is expressed as the average entropy of a replay attack.

$$H_c = \frac{1}{N} \sum \sum_{i=1}^{i=n_l} -(p_{ic} \ \log \ p_{ic}) \tag{9}$$

Figure 11 gives the analysis of the average entropy of a replay attack. From Figure 11a, we can see that the increase of k has little influence on the average entropy H_c. Because the replay attack is used to infer the location of users and k has little influence on the number of roads in anonymity, so H_c has little increase with the increase of the value of l. From Figure 11b, we can conclude that H_c would increase with the increase of l. This is because the more number of roads would result in the larger uncertain metric that the attacker infers the location of the users, that is, the decrease of p_{ic} results in the increase of H_c. Simultaneously, the average entropy H_c of a replay attack is greater than that of the anonymous ring algorithm and anonymous cellular algorithm. Because three algorithms all considered the anti-replay attack, all of them have a high average entropy. On the other side, we introduce the candidate extended factor in the V-M algorithm to increase the ability to an anti-replay attack.

4.2.5 The average entropy of a side-weight inference attack

Side-weight inference is a common attack method. Its average entropy can represent the defense ability of a side-weight attack and is also an important parameter to evaluate the performance of system security. An attacker may obtain the probability p_{ib} at a certain road that a user lies in by utilizing side-weight inference, so the available information entropy H_b can represent the uncertain metric that the attacker can infer the correct road where the user lies. We can obtain H_{bi} based on Equation 10. Obviously, the higher the average entropy H_b, the more difficult the side-weight inference. So the defense ability of attack is stronger.

$$H_b = \frac{1}{N} \sum \sum_{i=1}^{i=n_l} -(p_{ib} \ \log \ p_{ib}) \tag{10}$$

Figure 12 gives the analysis of the average entropy of side-weight inference. From Figure 12a, we can see that the increase of k has little influence on the average entropy H_b. In this paper, we select the maximum k and l as anonymous conditions, so the anonymous users that satisfy the conditions are so many that the variation of k influence on H_b is little. From Figure 12b, we can conclude that H_b would increase with the increase of l. This is because the more the number of roads that an anonymous user is maybe lying in, the less the inferable

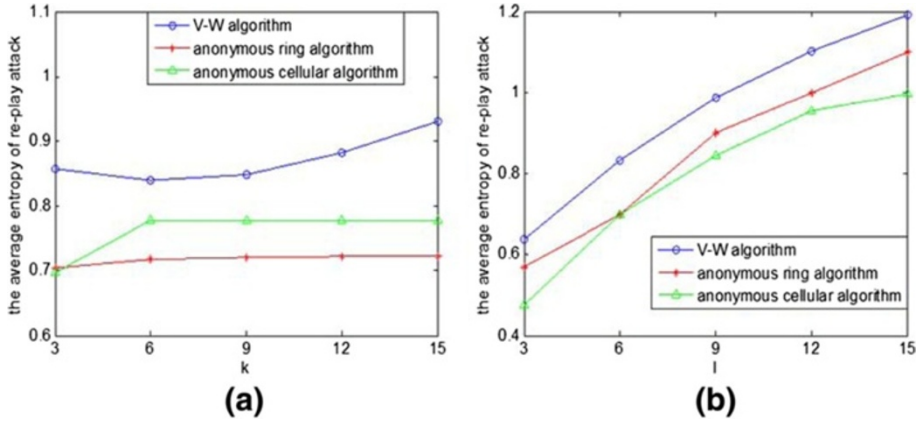

Figure 11 The average information entropy comparison. From **(a)**, the increase of k has little influence on the average entropy H_c. From **(b)**, the H_c would increase with the increase of l values.

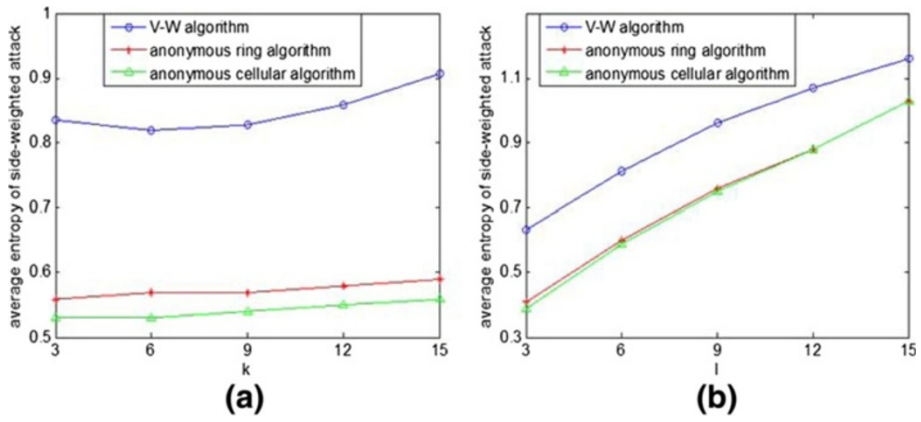

Figure 12 The average entropy comparison of side-weight attack. From **(a)**, the increase of k has little influence on the average entropy H_b. From **(b)**, the H_b would increase with the increase of l values.

probability that the user maybe lying in, that is, increasing the denominator of Equation 4 results in the increase of H_b. Simultaneously, the average entropy H_b of side-weight inference is larger than that of the anonymous ring algorithm and cellular algorithm. The reason is that the anonymous ring algorithm and cellular algorithm do not consider the uniform distribution of the user in the road, so it is easy to be attacked by side-weight inference. However, in this paper, the V-W algorithm regards the attack probability of side-weight inference as a requirement to system in the process of anonymity. So the average entropy of the V-W algorithm is higher than that of other algorithms. In other words, we obtain the higher security of the system by lowering the success rate of anonymity.

Lastly, based on the above analysis, the algorithm proposed in this paper can meet our design target given in chapter III. The algorithm not only satisfies the demand of k-anonymity and l-diversity, but also has the higher the average entropy of replay attack and side-weight inference. These provide the better ability on protecting location privacy for users.

5 Conclusions

This paper firstly introduces the existing method on location privacy protection. Secondly, we analyze the drawback of traditional algorithms and give the design target of our algorithm. Lastly, we proposed and analyzed the V-W algorithm, then compared it with traditional algorithms. The results can show that our algorithm can provide the better performance on location privacy protection.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61471077, 61301126), Fundamental and Frontier Research Project of Chongqing (cstc2013jcyjA40034, cstc2013jcyjA40041, cstc2013jcyjA40032), Science and Technology Project of Chongqing Municipal Education Commission (KJ1400413, KJ130528), Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), and Special Fund of Chongqing Key Laboratory (CSTC).

Received: 25 June 2014 Accepted: 19 October 2014
Published: 28 November 2014

References

1. R Cheng, Y Zhang, E Bertino, S Prabhakar, in *Proceedings of Privacy Enhancing Technologies: 28-30 June 2006; Cambridge*, ed. by G Danezis, Golle P. Preserving user location privacy in mobile data management infrastructures (Springer, Berlin Heidelberg, 2006), pp. 393–412
2. A Aryan, S Singh, in *Proceedings of the 2010 International Conference on Computer and Communication Technology: 17-19 September 2010; Allahabad*. Protecting location privacy in augmented reality using k-anonymization and pseudo-id (IEEE, Piscataway, 2010), pp. 119–124
3. JI Hong, JA Landay, in *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services: 6-9 June 2004; Boston*. An architecture for privacy-sensitive ubiquitous computing (ACM, New York, 2004), pp. 177–189
4. ML Yiu, CS Jensen, XG Huang, H Lu, in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering: 7-12 April 2008; Cancun*. SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services (ACM, New York, 2008), pp. 366–375
5. YF Xiao, HY Xu, A location privacy protection method based on anonymous region transformation. Comput. Eng. **39**(1), 157–163 (2013)
6. H Lu, CS Jensen, ML Yiu, in *Proceedings of the 7th ACM International Workshop on Data Engineering for Wireless and Mobile Access: 9-12 June 2008; Vancouver*. Pad: privacy-area aware, dummy-based location privacy in mobile services (ACM, New York, 2008), pp. 16–23
7. H Kido, Y Yanagisawa, T Satoh, in *Proceedings of the 2nd International Conference on Pervasive Services: 11-14 July 2005; Santorini*. An anonymous communication technique using dummies for location-based services (IEEE, Piscataway, 2005), pp. 88–97
8. M Gruteser, D Grunwald, in *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services: 5-8 May 2003; San Francisco*. Anonymous usage of location-based services through spatial and temporal cloaking (ACM, New York, 2003), pp. 31–42
9. B Gedik, L Liu, in *Proceedings of the International Conference on Distributed Computing Systems: 6-10 June 2005; Columbus*. A customizable k-anonymity model for protecting location privacy (Georgia Institute of Technology, Georgia, 2005), pp. 1–12
10. B Gedik, L Liu, Protecting location privacy with personalized k-anonymity: architecture and algorithms. Mobile Comput. **7**(1), 1–18 (2008)

11. Z Xiao, XF Meng, JL Xu, Quality aware privacy protection for location-based services. Adv. Database: Concepts Syst. Appl. **10**(33), 434–446 (2007)
12. MF Mokbel, CY Chow, WG Aref, in *Proceedings of the 32nd International Conference on Very Large Data Bases: 12-15 September 2006; Seoul*, ed. by U Dayal, KY Whang, D Lomet, G Alonso, G Lohman, M Kersten, SK Cha, and YK Kim. The new casper:query processing for location services without compromising privacy (ACM, New York, 2006), pp. 736–774
13. Y Rubner, C Tomasi, LJ Guibas, The earth mover's distance as a metric for image retrieval. Comput. Vis. **40**(2), 99–121 (2000)
14. T Wang, L Liu, in *Proceeding of the 35th International Conference on Vary Large Data Bases: 24-28 August 2009; Lyon*. Privacy-aware mobile services over road networks (ACM, New York, 2009), pp. 1042–1053
15. J Xue, XY Liu, XC Yang, B Wang, A location privacy preserving approach on road network. Chin. J. Comput. **34**(5), 865–878 (2011)
16. J Xu, M Xu, X Lin, N Zheng, Location privacy protection through anonymous cells in road network. J. Zhe Jiang University (Engineering science). **45**(3), 429–434 (2011)
17. P Zhao, CG Ma, XB Gao, W Zhu, Protecting location privacy with voronoi diagram over road networks. Comput. Sci. **40**(7), 116–120 (2013)

doi:10.1186/1687-1499-2014-202
Cite this article as: Fan *et al.*: The research for protecting location privacy based on V-W algorithm. *EURASIP Journal on Wireless Communications and Networking* 2014 **2014**:202.