An Algorithm of the Privacy Security Protection Based on Location Service in the Internet of Vehicles

Peng-Shou Xie[1,2], Tian-Xia Fu[2], and Hong-Jin Fan[2]
(Corresponding author: Tian-Xia Fu)
Research Center of Engineering and Technology for Manufacturing Informatization of Gansu Province[1]
School of Computer and Communication, Lanzhou University of Technology[2]
No. 287, Lan-Gong-Ping Road, Lanzhou 730050, China
(Email: fukaix_wang@163.com)
(Received Nov. 14, 2017; Revised and Accepted June 15, 2018; First Online Feb. 24, 2019)

Abstract

The safe and comfortable location based services are required for users in the internet of vehicles, at the same time, the privacy and confidential requirements are indispensable. Enciphering merely user′s information cannot guarantee the safety of user′s privacy, and the query itself may leak the user′s location information and identity one. Aimed at the problem of privacy security in location based service under the internet of vehicles environment, in this paper, through analysis of LBS privacy security technology, in the V2I system of internet of vehicles, combining K area with fake names anonymous technology, a kind of improved PPA-IOV privacy protection algorithm is formed. Experimental verification indicated that the algorithm performs a higher anonymous success rate for users in the car network environment and reduced the average anonymous space, thus the service quality of user′s query is improved.

Keywords: Anonymous Success Rate; Internet of Vehicles; Location Based Services; Privacy Protection

1 Introduction

With the popularization of cars and ownership increased significantly nowadays, the road conditions become more and more complex and the safety of road situation is not optimistic as well. When users in the internet of vehicles enjoy the great convenience provided by location based services (LBS), their own identity information is involved inevitably. Location coordinates and the contents queried of user privacy information are exposed to the network. Location based service is the main basis for the server to process the service request. The more accurate the location information provided by the user, the more accurate the service information returned to their after the server query processing. However, the query service request that contains the exact location information is recorded in the location server, which will undoubtedly open the door to malicious attackers to steal the user′s location privacy and query privacy. If the location server is not trusted or the communication process that between the user and the location server is unsafe, the user′s location information or query content may be stolen or disguised by the malicious attacker. The attacker deduces the user′s trajectories and status according to obtain the user′s location information, or from the user query content to infer the user to travel to the destination and so on, which will bring serious privacy threats to user [6].

In recent years, a variety of user's location privacy protection scheme has been proposed. Zhang *et al.* [16] presented a personalized LBS (P, L, K) anonymous model based on sensitivity, and on this basis, formed the privacy protection query anonymous algorithm under the use of mesh division and pseudonym users. By searching for the user′s neighborhood space iteratively, the purpose of protecting the privacy of the query was achieved. Li *et al.* [13] proposed a mobile-cloud framework, which is an active approach to eradicate the data over-collection. By putting all users′ data into a cloud, the security of users′ data can be greatly improved. Che *et al.* [4] put forwarded a location anonymity algorithm based on P2P and dynamic grids. The grid is used to divide the space provided by the LBS service provider, and dynamically adjust its size of the space according to the user′s required anonymity and the number of users.

Arain *et al.* [1] proposed Dynamic Pseudonym based multiple mix zone (DPMM) technique by analyzing limitations of existing methods related to location privacy with mix zones, such as RPCLP, EPCS and MODP, which ensures the highest level of accuracy and privacy, and addresses the issues related with existing location privacy protection techniques. Han *et al.* [7] analyzed the location K-anonymity technique, which requires that when

an user sends a LBS request, its location information is undistinguishable from other location information of at least $k-1$ users, which can effectively resist query tracking. Jin *et al.* [8] improved the positional K-anonymity algorithm based on a quad-tree-like scan using a bottom-up approach. Based on the Casper model, this method after scanning the information of the lowest-level grid, and choose to iterate upwards cells to improve spatial resolution if it satisfies the minimum anonymity requirements. Some progresses have been made in the solution to privacy protection, but there is still a problem that the contradiction between the effect of privacy protection and the accuracy of the processing results is difficult to balance. Importantly, there are relatively few researches on privacy protection methods specific to the environment of internet of vehicles.

Considering the shortcomings of the above researches, we propose a Privacy Preservation Algorithm-Internet of Vehicles (PPA-IOV) correspondingly. By the PPA-IOV method, privacy information can be well protected. Through the analysis of service structure of LBS and location privacy protection technology in internet of vehicles [12], the PPA-IOV is improved immensely by mean of combining K anonymous area and pseudonym technique in P2P network structure. The simulation experiments indicate that the PPA-IOV can get a better success rate and improves the accuracy of the service by reducing the anonymous area without exposing the exact location of the user comparing with the SCAPGID [5] algorithm and P2P-IS-CA-HL [4] algorithm.

The rest of the paper is organized as follows. The related location privacy protection technology of the algorithm and design of PPA-IOV in internet of vehicles are introduced in Section 2. The key Steps of the algorithm and the realization process in detail are introduced in Section 3. The detection performance of the proposed algorithm are analyzed and compared through two sets of simulation experiments in Section 4. Section 5 concludes the solutions.

2 Privacy Protection Algorithm - Internet of Vehicles (PPA-IOV)

2.1 Related Location Privacy Protection Technology

In LBS, the most widely used privacy protection model is the location K-anonymous model [15]. The basic idea is that the location of the mobile user is satisfied the location k-anonymity when a mobile user′s location cannot be distinguished from the location of other k-1 mobile users. There are three basic techniques for implementing location k-anonymity: Dummy location, spatial cloaking, and spatial-temporal cloaking [9].

1) Dummy location.

   Generally speaking, publishing location pseudonym information is putting the false location of service request instead of the real location. As shown in the Figure 1, the small rectangle represents the user object in the LBS. The black dot indicates the location information of the user. The hollow dot indicates the user's false location submitted to the location server. One of the important advantages of pseudonym technology is that users can generate dummies on their own without the need for any other communication protocol components. When using the pseudonym technology, the attacker does not know where the users real position is. Therefore, the farther the distance between true and false positions is, the higher the security factor is, but the worse the service quality is. It is possible to obtain the correct query result if the user object provides a false position for L1 instead of L2 in Figure 1, but it is also easier to expose the actual position of L1 than L2.

2) Spatial cloaking.

   The main idea of this approach is to replace the user's precise location information with a spatial hiding area that requires the location of the user and at least other k-1 users. For example, the real position of a user is u, and the idea of space clocking is to expand this point into a hidden range A, as shown in the dotted ellipse area in the Figure 2, use the hidden area A instead of the user′s location u, and to ensure that the probability of user occurrence on each location in the area is the same. So that the attacker can only know that the user is in the hidden area, but cannot tell the user's exact location. The size of the hidden area is proportional to the degree of privacy protection and inversely proportional to the quality of the service.

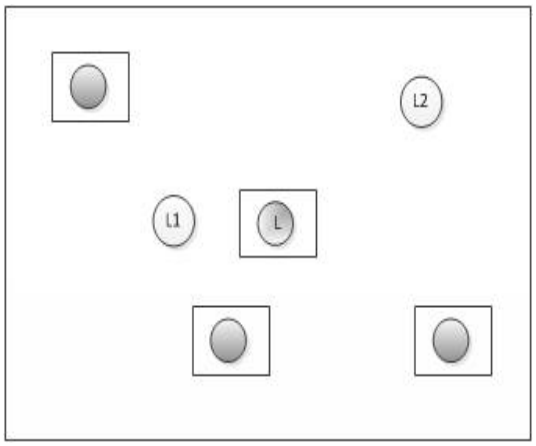


Figure 1: Diagram of the dummy location

3) Spatial-temporal cloaking.

   The basic idea of spatial-temporal cloaking is to delay the response time when constructing a hidden area of space, with a temporal and spatial hidden area

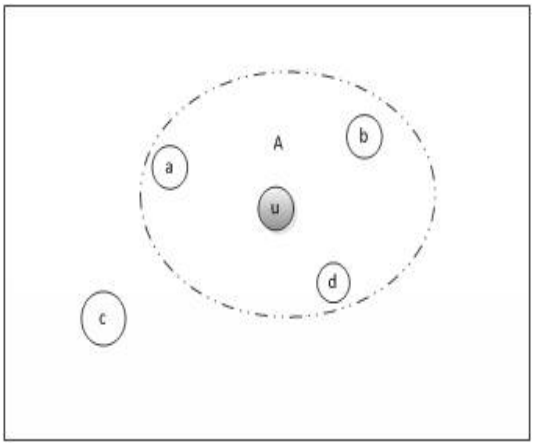


Figure 2: Diagram of the spatial cloaking

instead of the user′s precise location, which also contains the location of the user and at least other k-1 users. As a result of the delay in response time, so during this time, there will be more users and more queries. The greater the density of users, the smaller the hidden area is, and thereby improved the degree of privacy protection and the quality of service. However, apart from the shortcomings of spatial clocking, spatial-temporal cloaking technology also increases service response time.

2.2 PPA-IOV Design

In the V2I communication system of the internet of vehicles, vehicle nodes can communicate with the roadside unit (RSU) directly by using the P2P structure. The P2P structure [10] can not only use the knowledge of other nodes in the network and enrich the anonymous technology diversity when compared with the independent structure, but also eliminate the need for third-party anonymous server, to avoid the risk of information disclosure by attacked server when compared with the central structure. The RSU plays a certification and supervisory role in nodes within its coverage. In the environment of the internet of vehicles, the vehicle nodes are constantly changing and the formation of anonymous groups is transformed and updated in real time. The P2P structure relies on the collaboration of terminal users to achieve privacy protection. The nodes of the structure can communicate with each other through peer-to-peer network and establish a real-time assistance relationship. The security of a service model based on P2P structure depends on the selection of the anonymous areas. The privacy protection algorithm PPA-IOV proposed in this paper is described as follows:

1) The vehicle requests the identity authentication for RSU to obtain the node identity, and records it as V_i. The corresponding pseudonym node is generated by the pseudonym generation algorithm, denoted it as V_{pi}. The value of i is in the range of 1 to n, and n represents the number of vehicles passing through the RSU. Each vehicle node corresponds to only one corresponding vehicle pseudonym node.

2) The terminal nodes carry out peer-to-peer communication with the neighboring vehicle nodes (V_i or V_{pi}) through the ad-hoc network.

3) A certain node V_i within the signal coverage of an RSU actively forms an anonymous group with other nodes. A node can appear within multiple anonymous groups.

4) Set a fixed value K, when the number of nodes within the anonymous group reaches the K value, no new nodes will be added. Otherwise, then return to 3).

5) The K-anonymous region of the RSU is denoted as R_j, where j is in the range of 1 to m, and m represents the number of anonymous areas within an RSU, R_j including the K nodes in the anonymous region, R_j=$\{V_1, V_{p1}, V_2, V_{p3}, \ldots, V_K\}$.

6) Anonymous area R_j randomly selects a user as an agent to send the query to the RSU. The RSU receives a query from the entire anonymous area, and returns the reply to the entire anonymous area when the RSU responds the query result.

The model of the privacy protection algorithm is shown in Figure 3 below:

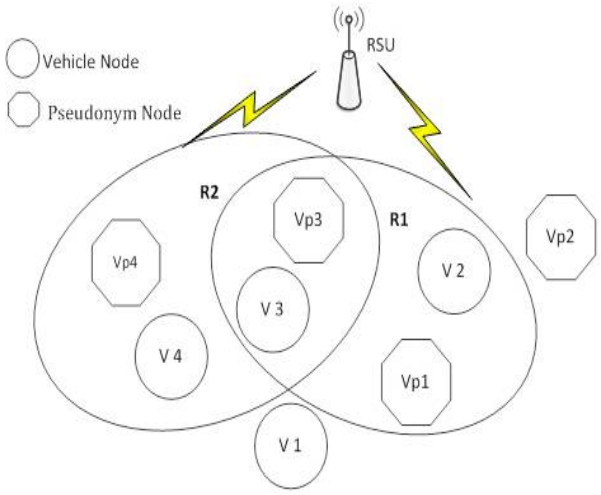


Figure 3: The model of the privacy protection algorithm

Take the Figure 3 for example, V_i (i=1, 2, 3, 4) represents the vehicle node, V_{pi} represents the corresponding pseudonym node, and R_i represents the K anonymous area within the RSU. From the figure, both node V_3 and V_{p3} are in the R_1 area, but also in the R_2 area. At the same time, V_2 in the R_1 area, but there can be no V_{p2}. In general, while the anonymous region R_j is formed, the elements in the R_j (j=1,2) region have both real vehicle nodes and fake position nodes, but the real nodes V_i

and the corresponding pseudonym nodes V_{pi} not must appear simultaneously in the R_j region. The algorithm ensures that the generated pseudonym nodes are at the same level as the real node. When the RSU returns the service response information, the node issuing the request information takes the initiative to obtain the response information, and the other nodes ignore it. Since the K-anonymous area sends out the request information as a whole, it is not easy for eavesdroppers to discern sends the service request accurate identification of the vehicle. At the same time, the use of the method can also effectively solve the difficult problem of the formation of anonymous group of vehicle nodes .

3 The Key Steps of the PPA-IOV Algorithm

3.1 Pseudonymous Generation Algorithm

In the privacy protection algorithm proposed in this paper, after obtaining the RSU authentication, the user generates a fake location node (dummy). And the two message nodes can send and receive the information normally and form an anonymous area together with the neighbor nodes, then send the message to the RSU. After receiving the service reply message sent back from the cloud server, the RSU broadcasts the response information to the anonymous area. The requesting user only needs to take the initiative to extract the necessary information. In this way, even if the RSU roadside unit stores a set of position data, it cannot distinguish the real position data from them.

As the road navigation service, the user must send the location data continuously in the LBS service of the Internet of Vehicles. Generally speaking, each object distance that can be moved is limited within a fixed time. If the dummy is randomly generated, the difference between the real location data and the fake position of the dummy node can be easily detected by the observer. In this case, the location anonymity is reduced. In order to avoid this situation, the dummy must not be completely different from the real location data. For this purpose, we first propose the following pseudonymous generation algorithm to double the pretender for real location data. First add the following assumptions:

a) All users generate the same number of virtual objects. Namely, each vehicle user generates only one virtual node.

b) In addition to location data, the user does not send other personal information.

c) The location data of the user in the process of generating the pseudonym node remains the same.

d) The user location information remains the same within the unit time. Over a period of time, it automatically re-authenticate and generates a pseudonym node.

The pseudo-code of the pseudonymous generation algorithm is shown in Algorithm 1. In this algorithm, the position of the dummy depends on the location coordinates of the real vehicle nodes at the last moment.

Algorithm 1 Pseudo-code of pseudonymous generation algorithm

Input: Positions of entities at $t-1$, m, n
Output: Positions of pseudonymous at t

```
Define a dummy structure {
    double x;  // x coordinate
    double y;  // y coordinate
    double t;  // time
}
Assignment entity[] to the Input
for i = 1; i < n; i++ do
  // rand(x,y): generate a random number between x and y
  dummie[i].x ← rand(entity[i].x-m, entity[i].x+m);
  dummie[i].y ← rand(entity[i].y-m, entity[i].y+m);
  dummie[i].t ← (entity[i].t)+1;
end for
Output the contents of the dummies[]
```

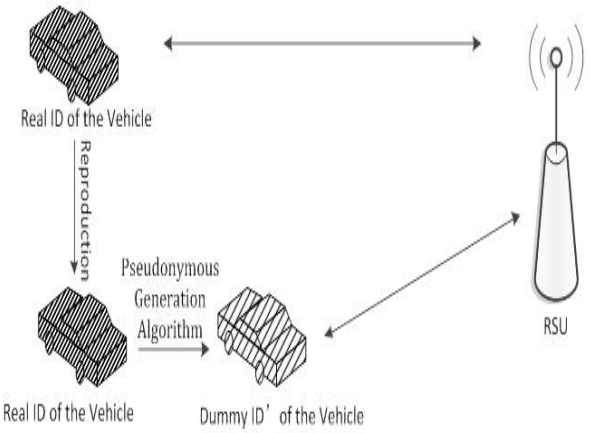


Figure 4: The model of the pseudonym generation algorithm

Set the basic information of the vehicle node: I=$\{ID$,$(x,\ y)$,t-1$\}$, where ID represents the node identification of the real vehicle, $(x,\ y)$ represents the position coordinates of the vehicle, and t represents the current time. After the pseudonym generation algorithm, the target user generates a user whose basic information is I'= $\{ID,$ random $(x,\ y),\ t\}$, where I and I' exist in the car network system at the same time, to realize the protection of double nodes of user's identities. The pseudonym generation algorithm model is shown in Figure 4 .

3.2 K-anonymous Area Generation Algorithm

As mentioned in Section 2.1 above, the most widely used privacy protection model in LBS is the location K-anonymity model. The K-anonymous location privacy protection technology that based on the generalized performances better in terms of accuracy and practicability, is a commonly used location privacy protection technology. In the process of K anonymity, Nearest Neighbor Clock [3] is the most classic anonymous area formation algorithm. As shown in Figure 5 (a), where user A performs an LBS query with an anonymous degree of 3, and the two nearest neighbors of user A are user B and user C. The anonymous area of user A is a rectangular area indicated by a dotted line in the figure. And the two nearest neighbors of user B are user C and user D, and the anonymous area of user B is the rectangular area indicated by the dotted line in the Figure 5 (b).

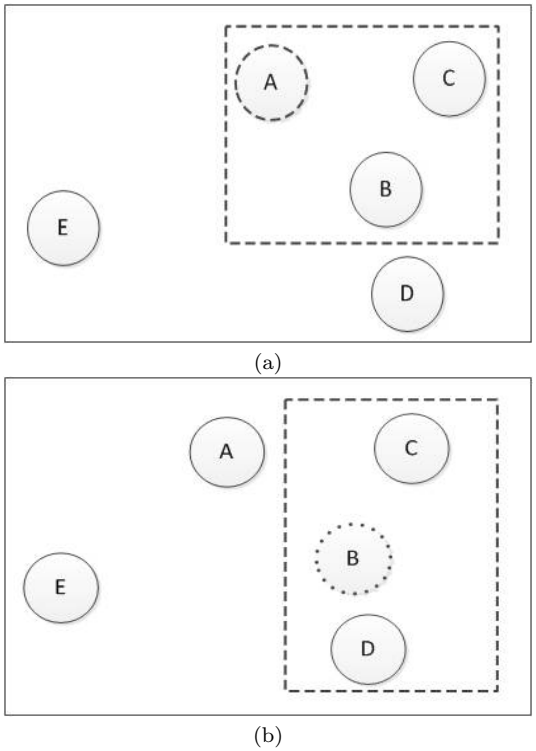


Figure 5: The nearest neighbors of anonymous area; (a) The nearest neighbors of anonymous area A; (b) The nearest neighbors of anonymous area B

The construction of anonymous areas is directly related to the accuracy of service search. If the anonymous area is too large, the service quality will be reduced. If the anonymous area is too small, the location of the user will be exposed. Associating the special trajectory of vehicle operation in the internet of vehicles, under normal circumstances, it only travels on a fixed road. The path of the vehicle is limited, and the area of the vehicle node is also limited.

The regular circular or rectangle anonymous area is not applicable under the environment of the internet of vehicles. Due to the fact that only one-way or two-way road systems are commonly used to drive vehicles on specific road sections, if the anonymous area is simply constructed in a circular shape, it will cause nearly half the invalid anonymous space in the anonymous area when the requesting node is located on the side of the road, thus causing the inaccuracy of the return of the query result. In addition, the vehicle road construction is not a regular pattern owing to the complex and changeable road terrain, but as the road condition changes constantly. So it is necessary to adopt an irregular anonymous area selection method to meet the needs of the anonymous area construction demand. For this reason, a K-anonymous area study based on boundary irregular polygons [14] is developed.

Algorithm 2 Pseudo-code of K-anonymous area generation algorithm

1: Input A_{min}, A_{max}, K
2: // **Step 1:** Peer search step
3: $List \leftarrow \{\varnothing\}$
4: U broadcast a request to the peers V_i
5: **for** $i = 1$; $i < k$; $i++$ **do**
6: check t_s of the v_i node
7: **if** $t_s \geq t - \Delta t$ **then**
8: $list \leftarrow list \sqcup$ {the received location information of node }
9: **else**
10: abandon the node
11: **end if**
12: **end for**
13: // **Step 2:** Cloaked Area step
14: $T \leftarrow$ the point that x-coordinates is the largest and smallest or the y-coordinates is the largest and smallest in List
15: $A \leftarrow$ a minimum bounding irregular polygons of all users in T
16: $S \leftarrow$ the acreage of the area A
17: **if** node is in the anonymous area A **then**
18: $List2 \leftarrow$ {the location information of node}
19: $List \leftarrow List$ - $List2$
20: **end if**
21: **while** $List = \{\varnothing\}$ **do**
22: Calculate the acreage S of A
23: **end while**
24: **if** $S < A_{min}$ **then**
25: recruit new nodes, and execute lines 6 to 23 to ensure that A satisfies the minimum area privacy requirement
26: **else**
27: **if** $S > A_{max}$ **then**
28: re-execute the Algorithm 2
29: **end if**
30: **end if**
31: **Return** A

When a user needs to query a specific service nearby,

the required anonymous requirements are first identified: the minimum area of anonymous area (A_{min}), the maximum anonymous area (A_{max}), and the anonymity degree (K). As the vehicle node is constantly moving, the longer the peer position information is the user cached, the lower the information timeliness of the nodes, the greater the offset of the position of the vehicle node, and the lower the accurate of the hidden area, so parameters Δt be used to control the obsolescence of the node caching information. And it also represents the life of the vehicle identity beacon. When $t_s \geq t - \Delta t$, the user node information is valid, where t_s is the point in time when the node caches location information and t represents the occurrence time of the query. Otherwise the requesting vehicles will automatically access to the RSU to obtain new identity authentication.

The generation of K anonymous areas in the PPA-IOV algorithm is described as follows.

Input: Set the service query to Q, Q ={ID, (x, y), v, M, t), where ID represents the node identify of the requesting vehicle, (x, y) represents the current position coordinates of the vehicle, and v represents the current velocity of the vehicle, M represents the content of query, and t represents the occurrence time of the event.

Output: Generate an irregular polygonal area that is not less than K nodes.

1) The user U first builds an anonymous area with the help of nearest neighbor anonymous algorithm to recruit neighbors. U checks the freshness of the nearest neighbor node cache information in turn. If the node within the last peer search time, that is $t_s \geq t - \Delta t$, the user U requests its neighbor to return its list and the size of the t_s. If the neighbor peer cache time information is too old, that is $t_s \leq t - \Delta t$, U will discard the node information. Through multiple operations, a result set $List$={(x_1, y_1), (x_2, y_2), $\cdots$, (x_k, y_k)}, with K node location coordinates that satisfies the K anonymity degrees is obtained.

2) Join all the points with the x-coordinates is the largest and smallest or the y-coordinates is the largest and smallest in the $List$ set into the set T. Such as T ={(x_{min}, y_a), (x_{max}, y_b) $(x_c, y_{min})(x_d, y_{max})$, $\cdots$}. The anonymous area made up of the points in the T set is recorded as A.

3) The x-coordinates is the largest and smallest or the y-coordinates is the largest and smallest of all points outside the A region are added to set T, and the anonymous region of the points in the T is recorded as a new anonymous region A. Repeatedly executed 3) until all nodes in $List$ are included in the A.

4) After completing the above steps, the regions formed in T are irregular polygonal regions with K nodes.

5) Calculate the acreage S of the polygon anonymous area A by dividing polygons into multiple triangles. Set the vertices in T to be arranged in counter clockwise order of $a_1, a_2, \cdots, a_m$. The coordinates of the vertices are in turn(x_1, y_1), (x_2, y_2), $\cdots$, (x_m, y_m), and the acreage of the anonymous area is:

$$S = \frac{1}{2}\Big(\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \cdots + \begin{vmatrix} x_{m-1} & y_{m-1} \\ x_m & y_m \end{vmatrix} + \begin{vmatrix} x_m & y_m \\ x_1 & y_1 \end{vmatrix} \Big). \quad (1)$$

6) If S<A_{min}, then add new neighbor nodes into the set $List$ and re-execute Step 2) to Step 5) until $S \geq A_{min}$. If $S > A_{max}$, then anonymous failure.

The pseudo-code of the K-anonymous area generation algorithm is shown in Algorithm 2.

After the K anonymous area is obtained, a user is randomly selected from the anonymous area as an agent for sending the query to the RSU. The RSU receives a query from the entire anonymous area and returns the reply to the entire anonymous area. Assume that the K anonymous area does not change during the time of the server returning the query result.

The K anonymous area generation algorithm makes the requesting user distribute evenly in the anonymous area, and realizes the probability that the requesting node is attacked at $1/K$. Compared with the circular anonymous area, the problem of the requesting user is located in the anonymous center, and the probability of attack is far greater than $1/K$ is well solved. In fact, in the case of the same number of nodes, the area of the irregular polygon based on the boundary is also smaller than the area of the anonymous area based on the circle or rectangle.

Figure 6 shows the idea of K anonymous area adjustment. The boundary point in $List$ is found to join the set T firstly (the boundary point is the maximum or minimum point of the x or y in the List set), in Figure (a), which constitutes a ring anonymous area. The anonymous area A contains A={1, 2, 3, 4, 5, 6, 13, 14, 15, 17, 20, 21}, $A_{outside}$={7, 8, 9, 11, 12, 16, 18, 19, 22}. The boundary point {8, 11, 7, 22} is found again outside the anonymous area A to join the set T in Figure (b), so the boundary set T and the anonymous region A are respectively: T={1, 11, 8, 6, 7, 17, 21, 22}, A={1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22}, at this time only two nodes {18,19}outside the anonymous area A . Then node 19 is incorporated into the T set, and the resulting anonymous area A contains all the nodes in the $List$. So far, the irregular K- anonymous region is formed, as shown in Figure (c).

4 Algorithm Implementation

In order to verify the effect of the algorithm, the experiment is carried out on the platform of processor Intel (R)

Table 1: The experimental configuration parameters of this paper

module category	parameters	value
traffic scene	The number of lanes	6
	The length of Road /m	1500
	The width of Lane /m	3.5
	The speed of the node /(m/s)	10
	The number of user nodes	100,200,300,400,500
network communication	The acreage of minimum anonymous area (A_{min}/m^2)	12500
	The acreage of maximum anonymous area (A_{max}/m^2)	25000
	The time for caching record lose efficacy(Δt /s)	10
	The value of K	50,60,70,80,90,100

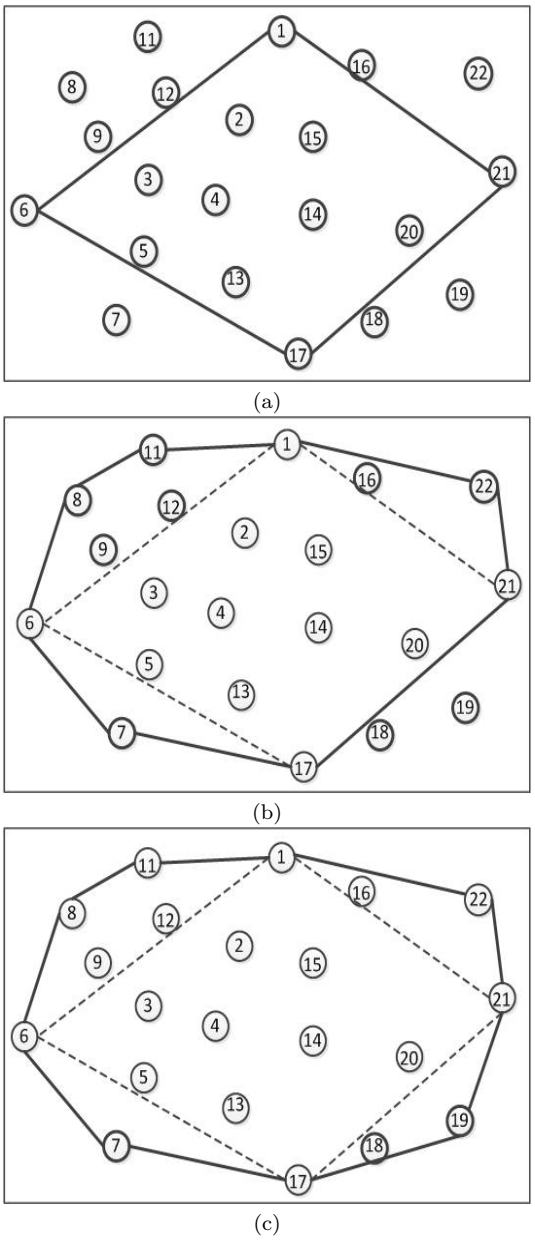


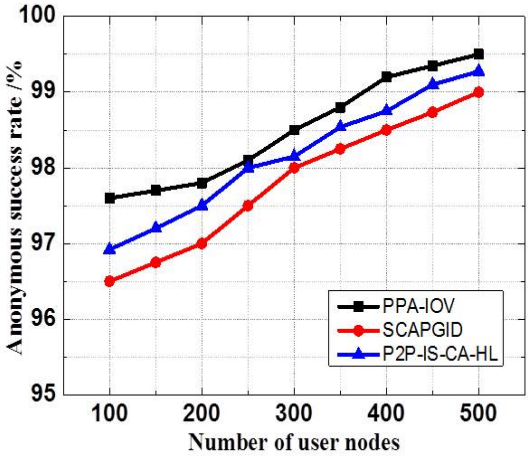
Figure 6: K anonymous regional adjustment

Core (TM) 2 Duo CPU E7500 3.0GHz with 4.0GB memory. First, the vehicle movement model is generated by simulation of urban mobility (SUMO) [2], because only the privacy security protection of nodes in the running process is discussed, so the setting of this paper experiment scenario is simple and only simulates a crossroad to a total of 6 straight lanes in two-way road. Supposing that a RSU signal is covered within 1500m∗1500m, and the number of nodes with different density is set up in the road to generate experimental scenes. Then the trace file generated in SUMO is connected to the network simulator NS-2 [11]. NS-2 reads the trace file to generate the vehicle node data. For better the result observation, the observation time was set at 200ms.

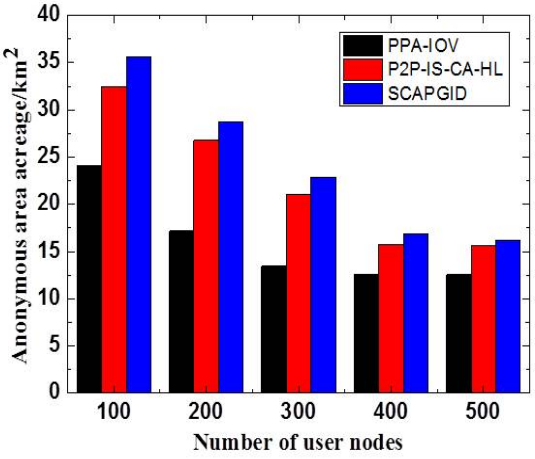
Algorithm SCAPGID divides the plane by the gird, and through the dynamic expansion of the grid and the user in the same grid to complete the location anonymous. In the P2P-IS-CA-HL Algorithm, the nodes in the anonymous area can share information and the anonymous area center can be adjusted. On each group of experiments, we were compared PPA-IOV with the SCAPGID and P2P-IS-CA-HL Algorithm varies in algorithm anonymous success rate and anonymous area size with the number of users changed within [100, 500] and the degree of anonymity changed within [50,100] respectively. The anonymous success rate indicates that the anonymity capacity of the privacy protection algorithm for the user's query request. While the smaller the anonymous area, the more accurate the quality of the query result obtained from the server.

The experimental process is divided into two parts: traffic scene and network communication. The table 1 shows the experimental configuration parameters.

Figure 7 shows the anonymous success rate and anonymous regions with the number of nodes changes respectively. It can be seen from the Figure (a) that the algorithm anonymous success rate of the three algorithms increases with the increasing of the number of nodes. Because the more nodes, the more assisted neighbor nodes in the anonymous area with the same large area, resulting in an increase in the anonymous success rate. The reason for the anonymous success rate in the PPA-IOV algorithm is relatively higher than that the SCAPGID and



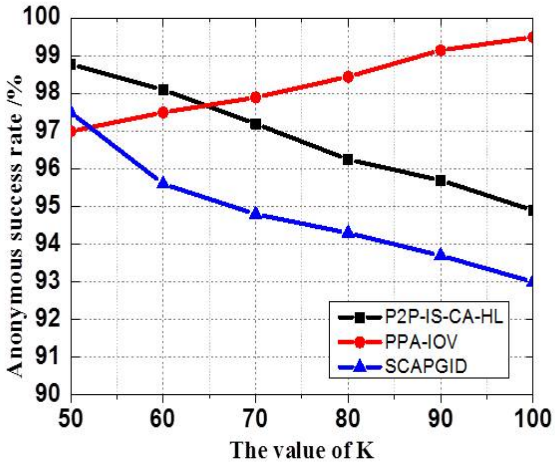
(a)



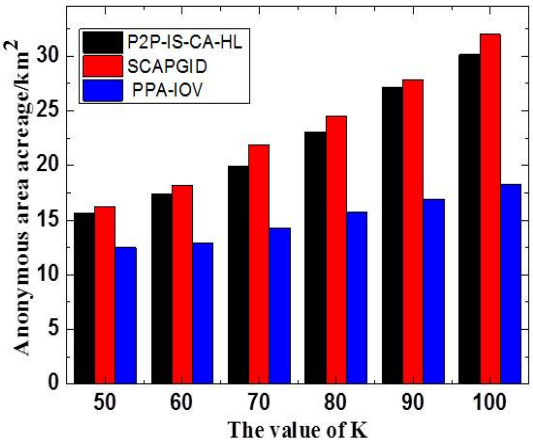
(b)

Figure 7: The anonymous success rate and anonymous area size varies with the number of user nodes. (a)The anonymous success rate varies with the number of user nodes. (b)The anonymous area size varies with the number of user nodes.

P2P-IS-CA-HL algorithm is that the pseudonym nodes in the PPA-IOV algorithm play a role and the number of nodes per unit area is larger. The Figure (b) shows that the anonymous area acreage of the three algorithms decreases as the number of users increases. In the case of fewer users, the algorithm needs to meet the degree of anonymity by enlarging the anonymous area. The SCAPGID and P2P-IS-CA-HL algorithm restrict the expansion way of the anonymous area by the idea of grid amplification and information sharing respectively, then causes the anonymous area to be far larger than the way that the PPA-IOV algorithm only expands the anonymous area through its own nodes. When the number of users reaches a certain degree and the nodes in the smallest anonymous area have already satisfied the degree of anonymity, then the anonymity area gradually stabilizes in the smallest anonymous area. Similarly, the pseudonymous node in the PPA-IOV algorithm makes the number of nodes in the entire environment higher than the other two algorithms, so it is easier to meet the anonymity in the case of a small number of users.



(a)



(b)

Figure 8: The anonymous success rate and the anonymous acreage under different k anonymity changes; (a)The anonymous success rate under different k anonymity changes; (b)The anonymous acreage under different k anonymity changes

Figure 8 depicts the variation of the anonymous success rate and the area of the anonymous area under different K anonymity levels respectively. The number of nodes at this time is set to 200. The graph displayed that the PPA-IOV algorithm increases the anonymous success rate as the increase of K anonymous, while the SCAPGID and P2P-IS-CA-HL algorithm showed an upward trend. This is because when the degree of anonymity increases, users need to recruit more neighbors help to collect enough peer

position information to fuzz its location, and these two algorithms are more likely to encounter network partitioning problem due to their own region amplification manner. Network partitioning problem is the number of users residing in the network partition is less than the required anonymous level K. Comparing to the SCAPGID and P2P-IS-CA-HL algorithm, the advantage of the PPA-IOV algorithm is the generation of pseudonym nodes and becomes the effective nodes in the anonymity area, which making the number of nodes in the whole experimental environment higher than them. Therefore, the area of anonymous acreage increases is very small with the K anonymity increasingly. While the SCAPGID and P2P-IS-CA-HL algorithm needs to accumulate the gird and share area of the anonymous regions to expand the anonymous area so that the number of nodes within it satisfies with K anonymity, which requires a larger area of anonymous area.

Furthermore, the above experiments indicated that the size of anonymous areas is not only related to the degree of anonymity K but also to the size of the smallest anonymous area. When the value of K is small, the decisive factor is the smallest anonymous region for the anonymous area. At this point, the higher node density in the region, the more nodes contained, and the greater the anonymous success rate. When the K value is larger, the nodes in the smallest anonymous area do not satisfy k anonymity, then anonymous areas spread out until the K nodes to meet, so the anonymous area formed in the environment of small node density is larger, and the success rate of anonymity increases.

5 Conclusions

This article around the theme of the terminal vehicle node privacy protection under the V2I system in the internet of vehicles, briefly describes three major privacy protection technologies at present. Then the privacy protection algorithm PPA-IOV by combining the characteristics of P2P structure is put forward. The content of the formation of pseudonyms and K anonymous areas in the algorithm are mainly introduced. Two groups of experiments via changing the number of user nodes and K anonymous respectively are performed, and the PPA-IOV algorithm is compared with the original algorithm SCAPGID and P2P-IS-CA-HL algorithm. The results showed that the algorithm proposed in this paper increases node density in the environment due to the application of pseudonym nodes, which improved the algorithm anonymous success rate. Furthermore the algorithm for getting the smaller anonymous space area and improve the quality of the query service.

Acknowledgments

This study was supported by the National Science and Technology Support Program of China (2012BAF12B19). The authors gratefully acknowledge the anonymous reviewers for their helpful comments and suggestions.

References

[1] Q. A. Arain, Z. L. Deng, I. Memon, A. Zubedi, J. C. Jiao, A. Ashraf, and M. S. Khan, "Privacy protection with dynamic pseudonym-based multiple mix-zones over road networks," *China Communications*, vol. 14, no. 4, pp. 89–100, 2017.

[2] M. Behrisch, D. Krajzewicz, and M. Weber, *Simulation of Urban Mobility*, 2014. (`http://sumo.dlr.de/2014/SUMO2014.pdf`)

[3] S. Bhattacharyya and G. Sanyal, "Feature based audio steganalysis (FAS)," *International Journal of Computer Network & Information Security*, vol. 4, no. 11, pp. 62–73, 2012.

[4] H. R. Che, Y. Z. He, and J. Q. Liu, "Spatial cloaking algorithm based on peer-to-peer and grid id," *Netinfo Security*, vol. 2015, no. 3, pp. 28–32, 2015.

[5] C. Y. Chow, M. F. Mokbel, and X. Liu, "Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments," *GeoInformatica*, vol. 15, no. 2, pp. 351–380, 2011.

[6] P. Gormley, M. Mcbride, and J. Harkin, "Internal location based services using wireless sensor networks and rfid technology," *International Journal of Computer Science & Network Security*, vol. 6, no. 4, pp. 108–113, 2006.

[7] J. M. Han, Y. Lin, J. Yu, J. Jia, and L. Q. Zheng, "Lbs privacy preservation method based on location k-anonymity," *Journal of Chinese Computer Systems*, vol. 35, no. 9, pp. 2088–2093, 2014.

[8] F. S. Jin, Z. S. Ye, and H. Song, "A similar quadtree based on location k-anonymity algorithm," *Transactions of Beijing Institute of Technology*, vol. 34, no. 1, pp. 68–71, 2014.

[9] J. Karjee and H. S. Jamadagni, "Data accuracy models under spatio-temporal correlation with adaptive strategies in wireless sensor networks," *International Journal of Network Security*, vol. 4, no. 1, pp. 2152–5064, 2013.

[10] E. Kim, J. Kim, and C.Lee, "Efficient neighbor selection through connection switching for p2p live streaming," *Journal of Ambient Intelligence & Humanized Computing*, vol. 2018, no. 1, pp. 1–11, 2018.

[11] N. Kumar, M. Kumar, and R. B. Patel, "A secure and energy efficient data dissemination protocol for wireless sensor networks," *International Journal of Network Security*, vol. 15, no. 6, pp. 490–500, 2013.

[12] C. T. Li, M. S. Hwang, and Y. P. Chu, "A secure and efficient communication scheme with authenticated key establishment and privacy preserving for vehicular ad hoc networks," *Computer Communications*, vol. 31, no. 12, pp. 2803–2814, 2008.

[13] Y. Li, W. Dai, Z. Ming, and M. Qiu, "Privacy protection for preventing data over-collection in smart

city," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1339–1350, 2016.

[14] G. A. D. Luna, P. Flocchini, N. Santoro, G. Viglietta, and M. Yamashita, "Meeting in a polygon by anonymous oblivious robots," *31st International Symposium on Distributed Computing (DISC'17)*, vol. 91, no. 14, pp. 1–15, 2017.

[15] Y. J. Wu, *Privacy Preserving Data Publishing : Models and Algorithms*. China: Tsinghua University press, 2015.

[16] F. X. Zhang and C. H. Jiang, "Research on lbs (p,l,k) model and its anonymous algorithms," *Netinfo Security*, vol. 2015, no. 11, pp. 66–70, 2015.

Biography

Peng-shou Xie He was born in Jan. 1972. He is a professor and a supervisor of master student at Lanzhou University of Technology. His major research fields include theory and technology of Internet of Things, technology and application of Internet of Manufacturing Things, theory and method of information system engineering, software theory and methodology.

Tian-xia Fu She was born in Aug. 1992. She is a master student at Lanzhou University of Technology. Her major research fields include security for privacy preservation in internet of vehicles.

Hong-jin Fan He was born in Mar. 1993. He is a master student at Lanzhou University of Technology. His major research fields include big data security in intelligent transportation.