Semantic Location Privacy Protection Based on Privacy Preference for Road Network

Yonglu Wang[1,2], Kaizhong Zuo[1,2(✉)], Rui Liu[1,2], and Liangmin Guo[1,2]

[1] School of Computer and Information, Anhui Normal University, Wuhu 241002, Anhui, China
zuokz@ahnu.edu.cn

[2] Anhui Provincial Key Laboratory of Network and Information Security, Anhui Normal University, Wuhu 241002, Anhui, China

Abstract. In recent years, with the popularization of mobile intelligent terminals, location-based services (LBS) have been widely used. When users enjoy the convenience of LBS, they also face with the risk of leakage of location privacy. Therefore, it is very important to provide effective privacy protection service during the use of LBS. Previous methods of location privacy protection cannot meet the requirements of users for location privacy protection and service qualities. For this reason, an adjustable semantic location privacy protection scheme is presented in this paper. According to the road network, this scheme introduces the privacy tolerance and the deficient number to select adjacent road segment to satisfy users' requirements. Experimental results show that the proposed scheme supports users' privacy preference for location privacy protection and the required quality of service, and fully considers the user's personalized privacy requirements.

Keywords: Location-based service · Privacy preference · Privacy preservation · Road network · Semantic location

1 Introduction

With the rapid development of wireless communication technology and positioning technology, it has promoted the wide application of location-based services (LBS) [1–3]. In order to get services in LBS, the user has to share its current location information, such as querying the nearest hospital. However, the location information can be stolen by the attackers, then more private information maybe leaked by the mining methods [4]. Therefore, the personal privacy information in LBS should be protected.

The location privacy protection method for road network is mainly based on *K*-anonymity and *L*-diversity [5–9], which is an anonymous region contains both *K* users and *L* road segments. However, this method doesn't consider the influence of semantic information of the location. Based on this, Li et al. [10] divided the city map based on the Voronoi diagram, and the anonymous region is constructed according to the semantic location popularity of the Voronoi cell. Xu et al. [11] proposed an incremental query optimization method based on local optimization and global

J. Vaidya et al. (Eds.): CSS 2019, LNCS 11983, pp. 330–342, 2019.
https://doi.org/10.1007/978-3-030-37352-8_30

optimization, using the deficient popularity to build an anonymous region. Li et al. [12] converted the road network into road segment clustering map, balancing quality of service and privacy requirements by constructing the anonymous region with different semantic location types. Chen et al. [13] introduces the regional popularity and combines with the user-defined sensitivity to calculate the privacy of adjacent road segments, a privacy protection algorithm based on location semantic is designed. Lv et al. [14] obtained the adjacent candidate road segments efficiently until the privacy requirements are satisfied. Chen et al. [15] considered the semantic information of the dummy location and enhanced the privacy protection by constructing a location semantic tree. Xu et al. [16] proposed a sensitivity-fade algorithm, which uses the semantic location influence vector to select the road segment to construct the anonymous region and improve privacy protection.

However, none of these methods consider most the user's privacy preference for location privacy protection and quality of service. Therefore, an adjustable semantic location privacy protection scheme for road network is proposed in this paper.

2 Preliminaries

2.1 Related Definition

In this paper, the road network is denoted as an undirected connectivity diagram $G = (V, E)$, $E = \{e_1, e_2, \ldots, e_m\}$ denotes the road segments in the road network, each road segment $e_i = \{eid, v_s, v_e\}$ is an edge in the road network, with *eid* is the road segment number, v_s and v_e respectively denote the starting and the end point of the road segment. $V = \{v_1, v_2, \ldots, v_n\}$ denotes the intersection of road segment. Anonymous region *CR* is comprised of multiple adjacent road segments $Edges = \{e_1, e_2, \ldots, e_i\}$, multiple users $Users = \{u_1, u_2, \ldots, u_j\}$ and multiple semantic locations $Locs = \{loc_1, loc_2, \ldots, loc_k\}$ on the road segments, in which the number of road segments *Edges* and users *Users* should satisfy the personalized privacy requirement of users.

Definition 1 (*Semantic location*). $loc = \{lid, eid, (x, y), tp\}$ denotes the semantic location in the road network, with *lid* is the number of the semantic location, *eid* is the number of the road where the semantic location is located, (x, y) is the coordinate of the semantic location, and *tp* is the type of the semantic location. The type of semantic location is divided into n types in total, and $Type = \{tp_1, tp_2, \ldots, tp_n\}$ is the set of n semantic location types.

Definition 2 (*Semantic location popularity*). It is used to describe the popularity of a semantic location type in the road network. For each semantic location type $tp_i \in Type$ set a popularity p_{tp_i}. The set $Pop = \{p_{tp_1}, p_{tp_2}, \ldots, p_{tp_n}\}$ indicates the popularity of all semantic location.

Definition 3 (*Semantic location sensitivity*). It is used to describe the sensitivity of a semantic location type in the road network. Each user sets a sensitivity s_{tp_i} for each semantic location type $tp_i \in Type$ according to their own circumstances. The set

$S_u = \{s_{tp_1}, s_{tp_2}, \ldots, s_{tp_n}\}$ is the set of sensitivity of all semantic location types relative to user u.

Based on definition 2 and definition 3, the popularity and the sensitivity of the *CR* can be defined as follow:

Definition 4 (*Regional popularity*). The popularity $Popular_{CR}$ of the *CR*,

$$Popular_{CR} = \sum_{i=1}^{|Type|} \frac{|CR.Locs.tp = tp_i|}{|CR.Locs|} p_{tp_i} \tag{1}$$

Definition 5 (*Regional sensitivity*). The sensitivity $Sens_{CR}$ of the *CR*,

$$Sens_{CR} = \sum_{i=1}^{|Type|} \frac{|CR.Locs.tp = tp_i|}{|CR.Locs|} s_{tp_i} \tag{2}$$

$|Type|$ in formulas (1) and (2) is the total number of semantic location types contained in the *CR*; and $|CR.Locs|$ is the number of semantic locations contained in the *CR*.

Definition 6 (*Regional privacy*). The regional privacy RP_{CR} of the *CR*,

$$RP_{CR} = \frac{Popular_{CR}}{Sens_{CR}} \tag{3}$$

Definition 7 (*Privacy tolerance*). It is used to describe the importance of the user's privacy information. Assuming that all adjacent road segments of the CR are the set $NearEdges = \{e_1, e_2, \ldots, e_n\}$, and the regional privacy is formed by adding road segments in *NearEdges* to the *CR* one by one, which is recorded as the set RP_{set}.

$$\forall rp_i \in RP_{set}, \delta_i = \frac{rp_{\max} - rp_i}{rp_{\max} - rp_{\min}}, \delta_i \in [0, 1] \tag{4}$$

$rp_{\max}$ is the maximum of RP_{set}, $rp_{\min}$ is the minimum of RP_{set}. Privacy tolerance reflects the user's choice of location privacy protection and the quality of service based on subjective intent, which is in a range of values [0,1]. The more attention is paid to the location privacy protection, the privacy tolerance is lower; otherwise, the quality of service is more important.

Definition 8 (*Privacy requirement*). For a user u that makes a query, his privacy requirements are expressed in $PR(K,L,\delta,S)$. In this case, K denotes the user-defined lowest number of anonymous users; L denotes the user-defined lowest number of road segments; δ denotes the user-defined highest value of privacy tolerance; and S is user-defined sensitivity of a group of different semantic location types.

Definition 9 (*Deficient number).* For a user u, it is used to describing how many anonymous users are still missing from the CR to satisfy $u.PR.K$,

$$dn = u.PR.K - |CR.Users| \tag{5}$$

$|CR.Users|$ is the number of anonymous users in the CR.

2.2 System Model

This paper is based on the central server architecture (Fig. 1), a trusted third anonymous server that exists in the client and the location server. Users send their locations, inquiry contents, and privacy requirements to anonymous servers. The anonymous server sends the users' locations after the privacy preference selection module and the anonymous module to the LBS server. The location server queries the candidate results and returns it to the anonymous server, the anonymous server analyzes the query candidate set, and returns the screening valid results to the requester. Besides, the anonymous server needs to store the city map information and the semantic location information.

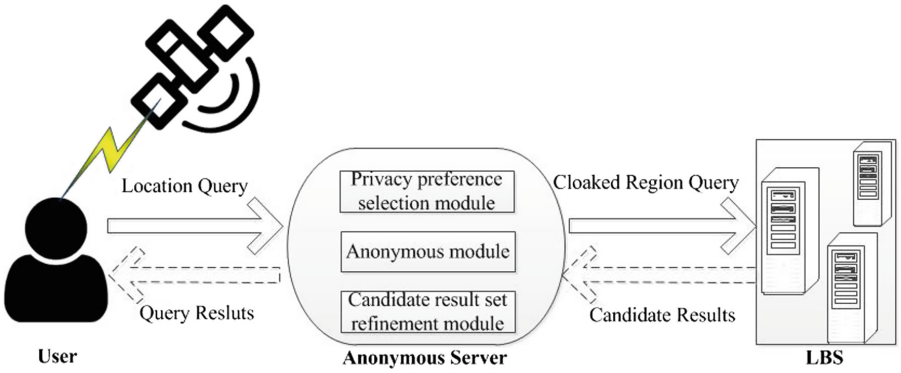


Fig. 1. Central Server Architecture

3 Adjustable Semantic Location Privacy Protection Scheme

This paper assumes that third-party servers(anonymous servers) are trustworthy to users. In order to achieve user's privacy preference for location privacy protection and the quality of service in the road network, this paper designs an adjustable semantic location privacy protection scheme. The algorithm flow as shown in Fig. 2.

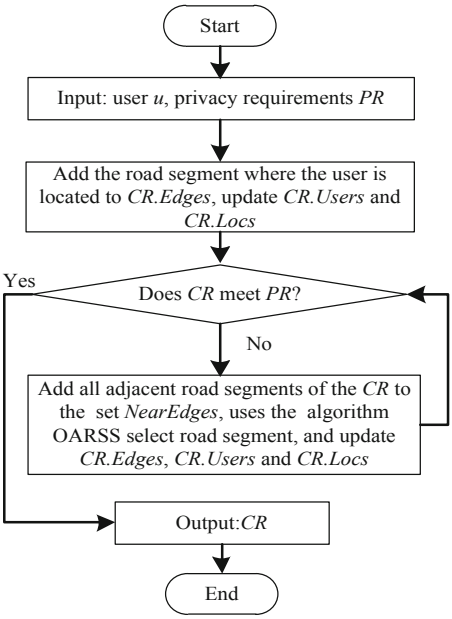


Fig. 2. Algorithm Flow

Algorithm 1 is an adjustable semantic location privacy protection scheme(ASLPP). The algorithm's input parameters include user u and privacy requirements PR. The concrete steps are as follows:

(1) Initialize the input parameters (line 1);
(2) Add the road segment where the user is located to the *CR.Edges* (line 2–3), and update *CR.Users* and *CR.Locs* (line 4);
(3) Determine whether the *CR* meets the user privacy requirements (line 5), If the requirements are satisfied, return *CR* (line 13); otherwise, execute step 4;
(4) Execute the algorithm OARSS, add the result to the *CR* and execute step 3 (line 5–12).

The pseudo-code of the algorithm is as follows:

Algorithm 1 Adjustable Semantic Location Privacy Protection Scheme

Input: user u, privacy requirements PR
Output: CR
1) $CR = \varnothing$;
2) find the road segment e where u is located;
3) $CR.Edges = CR.Edges \bigcup e$;
4) update $CR.Users$ and $CR.Locs$;
5) while $|CR.Users| < u.PR.K$ or $|CR.Edges| < u.PR.L$
6) $dn(CR) = u.PR.K - |CR.Users|$;
7) $NearEdge_{set} = Findedges(CR)$;//find all adjacent road segments of the CR
8) $BestEdge = OARSS(CR, NearEdge_{set}, dn(CR), u.PR.\delta, u.PR.S)$;
9) $CR.edges = CR.edges \bigcup BestEdge$;
10) update $CR.Users$ and $CR.Locs$;
11) $NearEdge_{set} = \varnothing$;
12) end while
13) return CR;

Algorithm 2 is the optimal road segment selection algorithm(OARSS). The input parameters of the algorithm are the anonymous region CRS, the deficient number dn, the privacy tolerance δ, the set of road segments $NearEdges$ and the set of sensitivity SS. The concrete steps are as follows:

(1) Initialize the input parameters (line 1–4);
(2) Calculate RP after adding the set of road segments and add it to the set RP_{set} (line 5–8), find the maximum and minimum of RP_{set} (line 9–10).
(3) Calculate the privacy tolerance of each road segment in the set $NearEdge$, and take the road segment of less than or equal to δ as the set $Cadedges_{set}$ (line 11–17).
(4) Add a road segment greater than or equal to dn in the $Cadedges_{set}$ to the set $DPEdge1_{set}$, and add a road segment smaller than dn to the set $DPEdge2_{set}$ (line 18–24).
(5) If $DPEdge1_{set}$ is not empty, return to the road segment of the minimum of anonymous users. Otherwise, return to the road segment of $DPEdge2_{set}$ with the maximum of anonymous users (line 25–30).

The pseudo-code of the algorithm is as follows:

Algorithm 2 Optimal Adjacent Road Segment Selection Algorithm

Input: anonymous region *CRS*, adjacent road segments set *NearEdges*, deficient number *dn*, privacy tolerance δ, the sensitivity set *SS*

Output: *AEdge*

1) $AEdge = \varnothing$;
2) $Cadedges_{set} = \varnothing, DPEdge1_{set} = \varnothing, DPEdge2_{set} = \varnothing$;
3) *MAX=0, MIN=0*;
4) $RP_{set} = \varnothing$;
5) for each *edge* in *NearEdges* do
6) $RP_{set} = RP_{set} \bigcup (RP = Popular_{(CRS \cup edge)} / Sens_{(CRS \cup edge)})$;
7) *CRS*.remove(*edge*);
8) end for
9) $MAX = \{RP \mid \max\{RP \in RP_{set}\}\}$;
10) $MIN = \{RP \mid \min\{RP \in RP_{set}\}\}$;
11) for each *edge* in *NearEdges* do
12) $RP = Popular_{(CRS \cup edge)} / Sens_{(CRS \cup edge)}$;
13) *CRS*.remove(*edge*);
14) if $\frac{MAX - RP}{MAX - MIN} \le \delta$ then
15) $Cadedges_{set} = Cadedges_{set} \bigcup edge$;
16) end if
17) end for
18) for each *edge* in $Cadedges_{set}$ do
19) if $(NumUser(edge) - dn) \ge 0$ then
20) $DPEdge1_{set} = DPEdge1_{set} \bigcup edge$;
21) else
22) $DPEdge2_{set} = DPEdge2_{set} \bigcup edge$;
23) end if
24) end for
25) if $DPEdge1_{set}.size() > 0$ then
26) $AEdge = \{edge \mid \min_{edge \in DPEdge1_{set}} \{NumUser(edge)\}\}$;
27) else
28) $AEdge = \{edge \mid \max_{edge \in DPEdge2_{set}} \{NumUser(edge)\}\}$;
29) end if
30) return *AEdge*;

4 Experiment and Analysis

4.1 Experiment Data Sets and Parameter Settings

The environment of the experiment is Intel Core(TM) 2 CPU @ 2.83 GHz; 2 GB RAM; the operating system is Microsoft Windows 7 Professional, and the algorithm is written in Java based on MyEclipse environment.

The experimental data is based on the Oldenburg map of Germany, which includes 6105 vertices and 7035 edges. There are 1 0000 uniform distribution users obtained from Brinkhoff based network mobile object generator [17] by introducing the highway network of Germany Oldenburg city into Brinkhoff generator. The users are distributed on the road segments, and a specific semantic information definition is labeled on the attribute of location type in location data generated by Brinkhoff generator, including 4 types of semantic location (hospital, bar, shopping mall, and school). All experimental parameters in the experiment set are shown in Table 1.

Table 1. Parameter settings

Parameter	Defaults	Evaluation rang
The number of users	10000	
K	25	[15,35]
L	6	[3, 15]
$L_{\max}$	20	
The number of semantic locations	10000	
The number of users that request services	1000	
δ	0.7	[0.1,1]

The experiment randomly selects 1000 users who request service to simulate experiments. For the convenience of calculation, the hypothetical popularity for the type of semantic locations is as following: {hospital:0.3, bar:0, shopping mall:0.4, school:0.3}. Considering the time complexity and the quality of service, the maximum number of road segments L_{max} is set in the experiment.

4.2 Analysis of Experimental Results

The experiment compares and evaluates ASLPP with LSBASC proposed in literature [13] and Enhance-LSBASC proposed in literature [14] from the aspects of anonymous success rate, average anonymous execution time, average number of semantic locations and regional privacy.

(1) Influence of δ

Figure 3 depicts the influence of δ on the algorithm ASLPP. Since the algorithm LSBASC and Enhance-LSBASC don't consider privacy tolerance δ, only the algorithm ASLPP is experimentally verified when K = 25, L = 6, L_{max}= 20 and δ changes from 0.1 to 1. It can be seen from Fig. 3 that the anonymous success rate of the algorithm

ASLPP is increasing, but the number of semantic locations, average anonymous execution time, and regional privacy are decreasing, especially after δ is greater than 0.7, the trend of change is more obvious. This is because when the δ increases, more road segments meet the privacy tolerance δ. The algorithm ASLPP selects the road segment according to the deficient number, reduces the number of road segments that need to be added. Therefore, the average anonymous execution time and the number of semantic locations are reduced, the anonymous success rate is improved. However, the larger of δ, the road segment selected by the algorithm ASLPP contains more sensitive semantic locations of the user, which makes the regional privacy decrease. It can be seen that the algorithm ASLPP can effectively implement the privacy preference selection of the user's location privacy protection and the quality of service by adjusting δ.

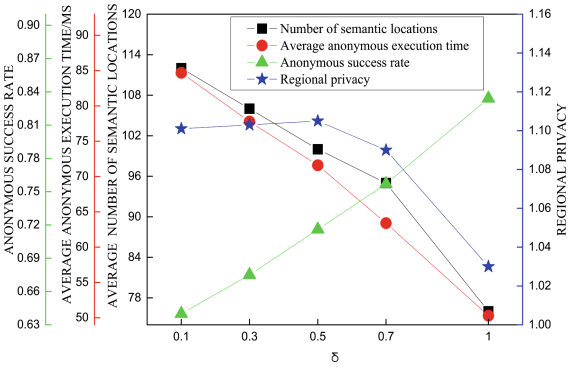
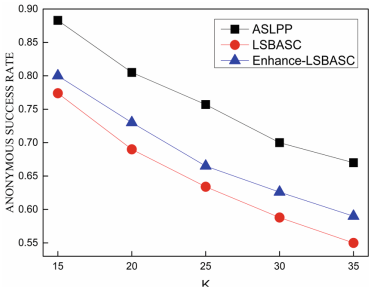


Fig. 3. Anonymous success rate, average number of semantic locations, average anonymous execution time, regional privacy and δ

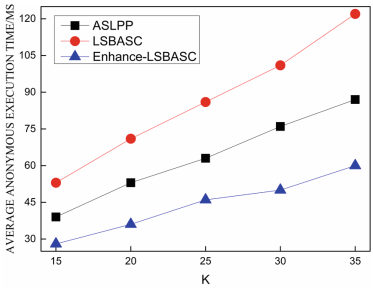
(2) Influence of *K*

Figure 4 depicts the influence of K on the algorithm ASLPP, LSBASC, and Enhance-LSBASC when $L = 6$, $\delta = 0.7$, $L_{max} = 20$ and K changes from 15 to 35. In Fig. 4(a), the anonymous success rate of the three algorithms is decreasing and the algorithm ASLPP is higher than that of the algorithm LSBASC and Enhance-LSBASC. This is because the algorithm LSBASC chooses the best one to join the anonymous set each time, while the algorithm Enhance-LSBASC selects the optimal road segment set to join the current anonymous set each time. When the number of added road segments reaches the upper limit of the road segment tolerance L_{max}, the number of anonymous users can't meet the privacy requirements, resulting in anonymous failure. The algorithm ASLPP selects the optimal road segment by the deficient number. In Fig. 4(b), the average anonymous execution time of the three algorithms is increasing, but the algorithm ASLPP is lower than that of the algorithm LSBASC and higher than that of Enhance-LSBASC. This is because when the K increases, more road segments need to be added to meet the user's privacy requirements.

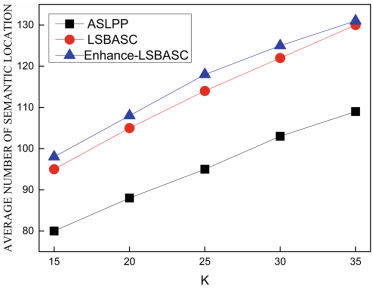
In Fig. 4(c), the number of semantic locations of the three algorithms increases, but algorithm ASLPP is lower than that of the algorithm LSBASC and Enhance-LSBASC. This is because algorithm ASLPP selects adjacent road segment based on the deficient



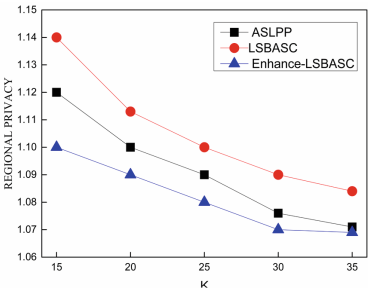
(a) Comparison of anonymous success rate



(b) Comparison of execution time



(c) Comparison of number of semantic locations



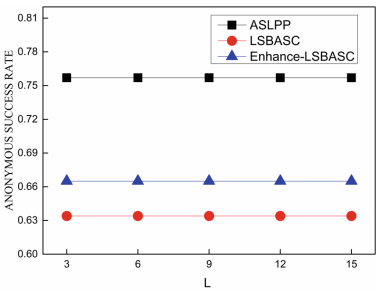
(d) Comparison of regional privacy

Fig. 4. Change of K

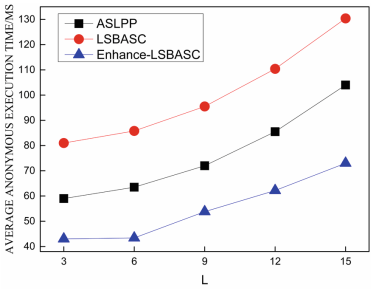
number, and reduces the number of road segments that need to be added. In Fig. 4(d), the regional privacy of the three algorithms is decreasing. This is because the anonymous region contains more relatively sensitive semantic locations of the user, which makes the regional privacy decrease. The algorithm ASLPP is lower than that of the algorithm LSBASC and higher than the algorithm Enhance-LSBASC. However, the algorithm ASLPP only reduces regional privacy by about 1% compared to the algorithm LSBASC.

(3) Influence of L

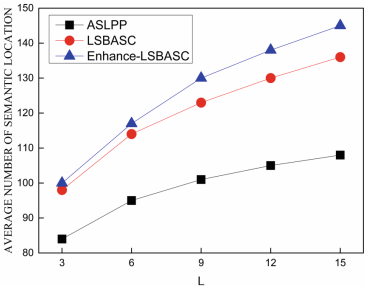
Figure 5 depicts the influence of L on the algorithm ASLPP, LSBASC, and Enhance-LSBASC when $K = 25$, $\delta = 0.7$, $L_{max} = 20$, and L changes from 3 to 15. In Fig. 5(a), the anonymous success rate of the three algorithms is not affected by L, and the algorithm ASLPP is higher than that of the algorithm LSBASC and Enhance-LSBASC. In Fig. 5(b), the average anonymous execution time of the three algorithms is increasing, but the algorithm ASLPP is lower than that of the algorithm LSBASC and higher than that of the algorithm Enhance-LSBASC. This is because when the L increases, the number of road segments in the anonymous region must reach a certain number. The algorithm ASLPP and LSBASC choose the best one each time, while the algorithm Enhance-LSBASC selects the optimal road segments set each time.



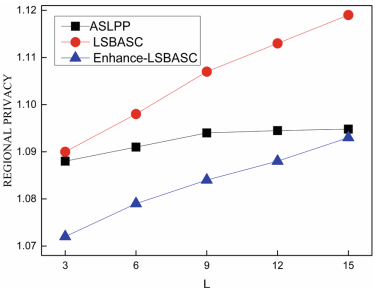
(a) Comparison of anonymous success rate



(b) Comparison of execution time



(c) Comparison of number of semantic locations



(d) Comparison of regional privacy

Fig. 5. Change of L

In Fig. 5(c), the number of semantic locations of the three algorithms is increasing, but the algorithm ASLPP is lower than that of the algorithm LSBASC and Enhance-LSBASC. This is because the algorithm ASLPP selects the road segment by the deficient number, reducing the number of road segments so that the number of semantic locations is decreasing. In Fig. 5(d), the regional privacy of the three algorithms is increasing, but the algorithm ASLPP is lower than that of the algorithm LSBASC and higher than that of the algorithm Enhance-LSBASC. This is because when the L increases, the number of road segments in the anonymous region is increasing, making the number of semantic locations are less sensitive to the increasing number of users.

5 Conclusion

In this paper, based on the existing semantic location privacy protection methods without the consideration of the user's privacy preference, we propose an adjustable semantic location privacy protection scheme for road network. The scheme selects the adjacent road segment through privacy tolerance and deficient number. Under the premise of fully satisfying personalized privacy requirements, the privacy preference of location privacy protection and service quality is realized. Finally, simulation experiments show that the proposed scheme supports the user's privacy preference.

Acknowledgement. This paper is supported by the *Natural Science Foundation of China* through projects 61672039 and 61370050, by the *Anhui Natural Science Foundation* through project 1508085QF133.

References

1. Wang, B., Yang, X., Wang, G., et al.: Energy efficient approximate self adaptive data collection in wireless sensor networks. Front. Comput. Sci. **10**(5), 936–950 (2016)
2. Sun, Y., Chen, M., Hu, L., et al.: ASA: against statistical attacks for privacy-aware users in location based service. Future Gen. Comput. Syst. **70**(4), 48–58 (2017)
3. Zhang, X.J., Gui, X.L., Wu, Z.D.: Privacy preservation for location-based services: a survey. J. Softw. **26**(9), 2373–2395 (2015)
4. Mingjie, M.A., Yuejin, D.U., Fenghua, L.I., Jiawen, L.: Review of semantic-based privacy-preserving approaches in LBS. Chin. J. Netw. Inf. Secur. **2**(12), 1–11 (2016)
5. Wang, Y., Xia, Y., Hou, J., et al.: A fast privacy-preserving framework for continuous location-based queries in road networks. J. Netw. Comput. Appl. **53**(1), 57–73 (2015)
6. Xinghua, L., Ermeng, W., Weidong, Y., et al.: DALP: a demand-aware location privacy protection scheme in continuous location-based services. Concurr. Comput. Pract. Exp. **28** (4), 1219–1236 (2016)
7. Cui, N., Yang, X., Wang, B.: A novel spatial cloaking scheme using hierarchical hilbert curve for location-based services. In: Cui, B., Zhang, N., Xu, J., Lian, X., Liu, D. (eds.) WAIM 2016, Part II. LNCS, vol. 9659, pp. 15–27. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39958-4_2
8. Niu, B., Li, Q., Zhu, X., et al.: Achieving k-anonymity in privacy-aware location-based services. J. Graph Algorithms Appl. **20**(2), 363–410 (2016)
9. Xiao, P., Weizhang, C., Yige, S., Lei, W.: Continuous queries privacy protection algorithm based on spatial-temporal similarity over road networks. J. Comput. Res. Dev. **54**(9), 2092–2101 (2017)
10. Li, M., Qin, Z., Wang, C.: Sensitive semantics-aware personality cloaking on road-network environment. Int. J. Secur. Appl. **8**(1), 133–146 (2014)
11. Xu, M., Xu, H., Xu, C.: Personalized semantic location privacy preservation algorithm based on query processing cost optimization. In: Wang, G., Atiquzzaman, M., Yan, Z., Choo, K.-K.R. (eds.) SpaCCS 2017. LNCS, vol. 10656, pp. 153–168. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-72389-1_14
12. Li, Y., Yuan, Y., Wang, G., Chen, L., Li, J.: Semantic-aware location privacy preservation on road networks. In: Navathe, S.B., Wu, W., Shekhar, S., Du, X., Wang, X.S., Xiong, H. (eds.) DASFAA 2016, Part II. LNCS, vol. 9643, pp. 314–331. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32049-6_20
13. Chen, H., Qin, X.: Location-semantic-based location privacy protection for road network. J. Commun. **37**(8), 67–76 (2016)
14. Lv, X., Shi, H., Wang, A., et al.: Semantic-based customizable location privacy protection scheme. In: International Symposium on Distributed Computing and Applications for Business Engineering and Science, pp. 148–154. IEEE Computer Society (2018)
15. Chen, S., Shen, H.: Semantic-aware dummy selection for location privacy preservation. In: Trustcom/bigdatase/ispa, pp. 752–759. IEEE (2017)

16. Xu, H., Zheng, Y., Zeng, J., Xu, C.: Location-semantic aware privacy protection algorithms for location-based services. In: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, pp. 1219–1224 (2018)
17. Brinkhoff, T.: A framework for generating network-based moving objects. Geoinformatica **6** (2), 153–180 (2002)