ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)

https://doi.org/10.17559/TV-20160304070312
Original scientific paper

Users Collaborative Mix-Zone to Resist the Query Content and Time Interval Correlation Attacks

Zhang LEI, Ma CHUN-GUANG, Yang SONG-TAO, Zheng XIAO-DONG

Abstract: In location-based services of continuous query, it is easier than snapshot to confirm whether a location belongs to a particular user, because sole location can be composed into a trajectory by profile correlation. In order to cut off the correlation and disturb the sub-trajectory, an un-detective region called mix-zone was proposed. However, at the time of this writing, the existing algorithms of this type mainly focus on the profiles of ID, passing time, transition probability, mobility patterns as well as road characteristics. In addition, there is still no standard way of coping with attacks of correlating each location by mining out query content and time interval from the sub-trajectory. To cope with such types of attack, users have to generalize their query contents and time intervals similarity. Hence, this paper first provided an attack model to simulate the adversary correlating the real location with a higher probability of query content and time interval similarity. Then a user collaboration mix-zone (CoMix) that can generalize these two types of profiles is proposed, so as to achieve location privacy. In CoMix, each user shares the common profile set to lowering the probability of success opponents to get the actual position through the correlation of location. Thirdly, entropy is utilized to measure the level of privacy preservation. At last, this paper further verifies the effectiveness and efficiency of the proposed algorithm by experimental evaluations.

Keywords: continuous location-based services; correlation attack; generalize profile; user collaborative mix-zone

1 INTRODUCTION

Nowadays, along with the development of location-based services (LBSs), the request for continuous service becomes more popular than snapshot one, such as finding the nearest gas station along this road or tell me the Italian restaurant every 5 minutes during my routing. In this type of service, users can obtain a series of responses from the LBS server, and enjoy a convenient journey or a pleasant leisure time by only submitting one query. However, as users generate a good deal of location-related data during the service, if failure to cope with, it will be possible for the adversary to compose these locations into a trajectory with time series. As trajectory contains more spatial-temporal information than discrete ones, so the adversary has the possibility to infer the regulation of a particular user, and mine out the privacy (e.g. user's home or work place). As a result, the adversary will bring some bothersome things in daily life or even some security problems, such as tracking or robbery.

In order to cope with the problem of preserving trajectory, many algorithms have been proposed in the past few years. These algorithms can be classified into two main categories: the whole trajectory disturbance [1-4] and the subsequent location generalization [5-10]. As sub-trajectories can be obtained before preserving, the adversary may infer the location correlation and identify the real trajectory from the anonymous group, which makes algorithms of whole trajectory disturbance failed. Thus, this paper mainly focuses on algorithms of subsequent location generalization. In order to generalize subsequent locations, different algorithms based on obfuscation as well as disturbance are provided. In types of obfuscation, Ma et al. [9] obfuscated the correlation with anchors, and cloaked the query content and time interval with users around the anchor's vicinity. Schlegel et al. [10] utilized a user-define privacy grid, and obfuscated the requested point of interests (PoIs) with randomly chosen cells. However, both of them assume that users are equipped with a powerful computational device and can cope with the poor quality of service.

To reduce the computational complexity and improve the service quality, algorithms based on disturbance were proposed [7, 11, 12]. In this type, algorithms usually disturb the real location by other k-1 users with the help of a trusted third party (TTP) to achieve k-anonymity [13]. Though these algorithms can provide a security service, the process of establishing a cloak region for continuous locations is extremely complex. It makes researchers turn to another scheme with a lower complexity [8]. A mix-zone can be seen as a black region with no information that can be detected from outside. After ID transformation, the adversary becomes difficult to track subsequent locations of users through ID correlation. Palanisamy et al. further optimized the mix-zone with delay-tolerant and non-rectangular structure [5, 6], and their algorithms can resist the timing and transition attacks as well as query correlation attack. However, as it can monitor the whole process of servicing, the adversary can obtain a sub-trajectory of the user before entering the mix-zone, and infer the query content and time interval from it. Then with these two profiles, the adversary can correlate locations of a particular user with the similarity comparison. Furthermore, with these identified locations, the adversary can reconstitute the whole trajectory and gain even more privacy.

The aims of this paper can be described as to resist the potential correlation attack, take advantage over the collaborative user scheme [14, 15] and achieve the query l-diversity [16] and time interval consistency. Based on these purposes, a user collaborative mix-zone (short for CoMix) is proposed. In this algorithm, query content and time interval are exchanged with other collaborative users in the same mix-zone, then each user is restricted to share the same query set and time interval. Once leaving this zone, the similarity of these two profiles will disturb the correlation between the sub-trajectory and subsequent locations. Therefore, the adversary will be difficult to confirm any location through the profile correlation attack. At last, compared with existing algorithms, our algorithm has the following features:

- No TTP is required. The whole process of achieving anonymity in mix-zone (e.g. query content and time interval transform) is implemented with collaborative users.
- We propose a scheme to resist the query content correlation attack as well as time intervals correlation attack. As the mix-zone cuts off the correlation with the shared similar profile by generalizing all users with the same query set and time interval.
- CoMix can provide a lower computational complexity and better service quality. As users no longer had to compute the nearest PoI and request services by their real locations, there is only a little computation in mobile devices and no redundant information feedback.

The rest of this paper is organized as follows. Section 2 presents the related work. Some preliminaries of our works are proposed in Section 3. Section 4 describes the process of CoMix algorithm and shows its security. Finally, evaluation results and conclusions are shown in Sections 5 and 6, respectively.

2 RELATED WORK

Location privacy is one of the most popular topics studied over the past decade, plenty of algorithms have been proposed. Based on the service type, the proposed algorithms can be roughly divided into two main categories: the snapshot location preservation as well as continuous preservation. In order to preserve the privacy in snapshot services, Gruteser et al. [13] imported *k*-anonymity from the data dissemination to obfuscate the user with at least *k*-1 users, so that the adversary can hardly identify a particular user. Gedik et al. [17] expanded the uncertainty by submitting a cloaking region. Bhuvan et al. [18], Liu et al. [16] and Wang et al. [19] considered the simplicity of *k*-anonymity is not enough and proposed the diversity of location, query and road segment respectively.

However, as TTP may become the single point of privacy preservation provider, users usually concern about whether this entity can leak the information in commercial objectives or breached by the adversary. As a result, algorithms without TTP were proposed. Based on cryptography, Ghinita et al. [20] proposed a computable private information retrieval (PIR) algorithm to provide zero information leakage retrieval. Khoshgozaran et al. [21] eased this process with a hardware-based PIR. Lien et al. [22] utilized a public-key homomorphic cryptosystem to achieve zero information leakage *k*-nearest neighbors (*k*NN) finding. Although these algorithms can provide privacy preservation services to users, they are obstructed by the complexity of computation and updating, so this paper only focuses on algorithms of user collaboration. The user collaboration scheme is firstly proposed by establishing group through P2P [23], then Chow et al. [24] remedied the drawback of the initiator usually locating in the center of collaborative users. Rebollo et al. [14] further this conception in query anonymity, and utilized entropy to select the optimal submitting user [25]. Niu et al. [15] proposed a Variance-Based Attack (VBA) to successfully attack the existing algorithms, then they proposed a random walk-based cloaking algorithm to remedy drawbacks of these algorithms.

Along with the development of LBSs technology, more and more people become willing to enjoy continuous LBSs than snapshot ones. This type of fashion makes privacy preservation of user collaboration no longer efficient, because they often submit queries with certain time delays. Thus, researchers have to go back to TTP for finding solutions and solutions of [1, 3, 4, 7, 9, 10, 26, 27] are able to be classified into two categories: whole trajectory disturbance and subsequent locations generalization. As the adversary can obtain user's sub-trajectory, algorithms of the whole trajectory disturbance are less suitable for continuous LBSs, so algorithms of subsequent locations generalization prosperous. Among these algorithms, mix-zone performs best, because mix-zone can lower the computational complexity and provide a balance between privacy and service quality.

The mix-zone can be seen as a black region with no information that can be detected by outsiders. After users' ID transformation, the adversary can no longer track subsequent locations of a particular user through ID correlation [8]. Freudiger et al. [28] introduced this scheme for location privacy in vehicular networks, and optimized it with less deployment number [29]. Le et al. [30] improved the effectiveness of mix-zone with dynamic deployment. Palanisamy et al. [5, 6, 31, 32] further this scheme with delay-tolerant and non-rectangular structure, which made the mix-zone able to resist such as timing attack, transition attack and continuous query correlation attack. Gao et al. [33] utilized this scheme in participatory sensing. Sun et al. [34] further optimized the deployment of mix-zone. However, just as Palanisamy had mentioned, there is no mix-zone completely free from the correlation query attack. To address this issue, this paper focuses on query content and time interval correlation, and utilizes a mix-zone to cut off this correlation to preserve the privacy. Different from most of current mix-zone algorithms, CoMix utilizes the collaborative user to share query content and time interval, so as to reduce the utilization of TTP to less the deployment cost.

3 PRELIMINARIES

3.1 Adversaries and Threat Model

The goal of the adversary is to obtain the sensitive information of a particular user through the whole trajectory, but this aim is obstructed by the undetectable profile transformation in mix-zone. In this zone, the adversary cannot take a statistic of how many users are in it, when and where users leave, and even how users change ID. In order to get the privacy, the adversary has to seek help from the sub-trajectory before preserved. As the sub-trajectory may contain the query content and time interval of the whole trajectory, user's subsequent locations can be identified by computing the higher probability based on the similar level between them. Formally, the adversary can compute $Pr(l \in T \mid sim(p_L{}^{(c,t)}, p_l{}^{(c,t)}))$ to guess the real location l by the similarity between sub-trajectory and l where T indicates the whole trajectory of a particular user, and the sub-trajectory L belongs to T, c and t denote the query content and time interval. When guessing the real location, the adversary can utilize the Bayes' rule to calculate the probability with the following formula.

$$\begin{aligned} & Pr(l \in T \mid sim(p_L{}^{(c,t)}, p_l{}^{(c,t)})) \\ & = \frac{Pr(l \in T) Pr(sim(p_L{}^{(c,t)}, p_l{}^{(c,t)}) \mid l \in T)}{Pr(sim(p_L{}^{(c,t)}, p_l{}^{(c,t)}))} \\ & = \frac{Pr(l \in T)}{Pr(sim(p_L{}^{(c,t)}, p_l{}^{(c,t)}))} \end{aligned} \quad (1)$$

For example, in Fig. 1, there are two trajectories of Alice and Bob. Assume that Alice queries for the nearest gas station and Bob queries for the restaurant. Although their trajectories are disturbed by mix-zone deployed in n_6, query contents of them are not changed. As the adversary may gain the sub-trajectories between n_5-n_6 and n_9-n_6, the request for gas station is denoted as triangles and the request for restaurants as pentagons may be mined out, so the real location with these contents will be correlated. The correlation of time interval is similar to this process.

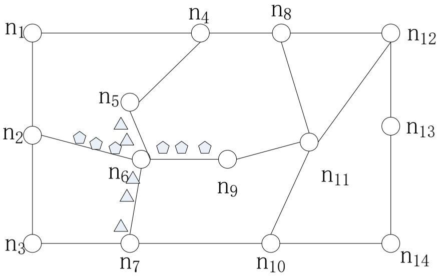

Figure 1 The query content correlation

In this work, the LBS server is directly considered as an adversary, as he is able to monitor the whole process of service, and stores all information of applicants. The LBS server can even combine a sub-trajectory with some un-preserved locations, and mine out the query content and time interval. Additionally, as the LBS server can pretend to be a user, he may also know the strategy used in the preservation algorithm. At last, collaborative users are considered as trusted entities, because the query content and time interval are all shared with each user in the collaborative users group.

3.2 Anonymity Metric

The Jaynes' rationale [35] on maximum entropy methods enables to measure the uncertainty modeled by a probability distribution by means of its Shannon entropy. In general, maximum entropy can be achieved by each user has the same probability to be treated as the real one. As the mix-zone cut off the correlation between sub-trajectory and uncertain locations, the degree of user hidden behind uncertain users can be reflected by the probability of failure to guess the real location. Thus, the maximum entropy can be used to measure anonymity degrees of users, and evaluate the optimal scheme with higher entropy.

To compute the entropy, the probability of a location belonging to a confirmed trajectory is used, and this probability is computed by the similarity of query content and time interval. In this paper, this probability is denoted as pr_i, and the sum of all probabilities pr_i is 1. Then assume there are at least k users in the set of uncertainty, the privacy level of a user hidden behind other users with the measure of entropy can be defined as

$$H(i) = -\sum_i^k pr_i \log_2 pr_i \quad (2)$$

Since higher entropy means a better privacy level, the aim of CoMix is to achieve the maximum entropy. From Eq. (2), the maximum entropy happens when all the k possible users have the same probability to be correlated as the real user. Thus, in order to resist the correlation attack, an algorithm has to let each user show the same query content and time interval.

3.3 Motivation and Basic Idea

Because the real location of the user can be guessed by the highest probability through the similarity of query content and time interval, the basic motivation of preserving user's privacy is to reduce the probability of real location been confirmed to a sub-trajectory or to reduce the similarity of each profile, but it is hard to achieve any of them. Therefore, in order to resist the correlation attack, the researcher has to seek another scheme. In the set of success probabilities, the real user is the one who has the highest probability among k users. For probabilities of the k users, they can be denoted as $Pr = \{pr_1, pr_2, \ldots pr_k\}$, and the real user is the maximum in Pr. In order to find out the real user from them, the adversary has to compute the probabilities of each user and compare them to find out the highest one. Thus, if each user has the same probability, the adversary cannot find the highest probability, which makes the result guessed by profile correlation attack to be invalid, and the user's trajectory privacy will be preserved. Based on above analysis the definition of L_k^θ privacy is proposed.

Definition 1 Let $L \in T$ denote the sub-trajectory obtained by the adversary, and $p_L{}^{(c,t)}$ denotes the profiles of query content and time interval mined from L. The scheme satisfies L_k^θ privacy if and only if for any two uncertainty locations l and l', there exist $|Pr(l) - Pr(l')| \le \theta$, where $Pr(l') = Pr(l \in T \mid sim(p_L{}^{(c,t)}, p_l{}^{(c,t)}))$, θ is the differential probability. In this paper, θ is defined as 0.

Based on definition 1, the basic idea of our solution was to generalize the correlation between sub-trajectory and subsequent locations, so that the successful probabilities of adversary guessing each user equal to each other. To this end, two basic conceptions are used. One conception is to employ a TTP, which is deployed in mix-zone to gather query content and time interval of each user, and provides the identical profile set to each user. If users leave mix-zone, they will share the same query content set and time interval. However, this conception has to confront the drawbacks of TTP (e.g. single point of failure and bottleneck of services performance). Another conception is to transmit the profile information with collaborative users, and interchange the profile with each other to achieve the profile generalization. For the sake of drawbacks in TTP, this paper mainly chooses the collaboration user scheme, and assumes that the users can communicate with each other through short-range communication technology in mix-zone, but they cannot communicate with outsiders.

For the purpose of illustration, continue to utilize the example of Alice and Bob in Fig. 2. As a subsequent location can be correlated with a confirmed sub-trajectory, the mix-zone is used to disturb the query content and time interval of each leaver. With CoMix, subsequent locations are disturbed as triangles locating in sub-trajectories between n_6-n_7 as well as n_6-n_2. In spite of gaining sub-trajectories between n_5-n_6 and n_9-n_6, the adversary utilizes the correlation attack to calculate the probabilities, but he cannot identify whether the current location belongs to Alice or Bob. The reason is that the similarity of profiles makes the probability equivalent with each user.

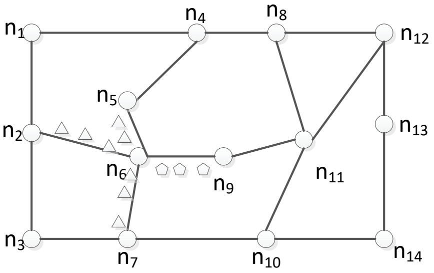

Figure 2 The result of exchanged profiles

4 USERS COLLABORATIVE MIX-ZONE

4.1 The Process of Exchanging Profiles

The process of exchanging profiles has two consecutive phases, one phase is establishing exchanging group with collaborative users, and the other is sharing profiles with them.

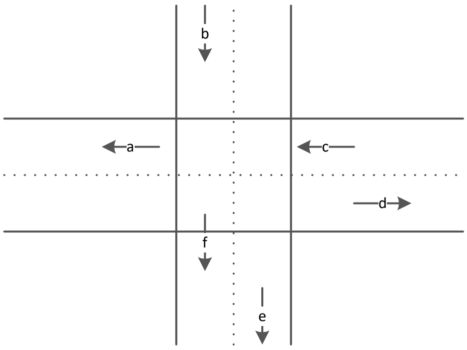

Figure 3 The exchanging group in mix-zone

1) Phase of establishing exchanging group

After entering mix-zone, the user broadcasts its request for profiles exchanging at any time before leaving. Then this user becomes a responder and he is no longer able to propose any requirement in this zone. If this request is responded by other users, this user can establish a collaborative group and exchange query content and time interval with them, so that he can achieve the profiles generalization. Otherwise, the user has to wait and respond to requirements of other users, and establishes exchanging group with them. For instance, in Fig. 3, b is a user driving in current mix-zone, and b broadcasts a request for establishing a group of exchanging. However, no user can respond to this request, as e is leaving this zone, and other users are all in its own group and cannot establish another anonymous group. Therefore, b becomes a responder of this zone, and has to wait and respond to other followers' requirements. This process is shown in Algorithm 1.

Algorithm 1 illustrates the idea of establishing a group of exchanging in details. In this algorithm, line 2 confirms the addable of this group by checking its value, and this value is used to determine whether a user is leaving this zone. In 5-11, the algorithm calculates the number of users with received requirements, and then checks this member in current group to determine whether to reject the request.

```
Algorithm 1: Establishing exchanging group
Input: u, r// A new user and its requirement
Output: G//The new exchanging group
1   each user in G' receives r;
2   if (addable==1);
3       broadcast l=1;//this user does not leave the mix-zone
4       receive all the l and computes $L=\sum_{i=1}^{G'.no} l_i$;
5   if(L==G'.no)
6   G=G'+u;
7   G.no=G'.no+1;
8   else
9   G=G';
10      addable=0;
11  end
12  end
```

2) Phase of sharing profiles

In the group of exchanging, users share their query contents and time intervals, and generate a similar query set as well as an equivalent time interval. The equivalent time is the greatest common divisor of every user's interval times. Finally, these users will have the similar query content and time interval. When leaving the mix-zone, each user requests the LBS with the similar profiles, until the requirement is terminated or the user enters into another mix-zone. If he enters another mix-zone, the user repeats above process, and sets up another group with its original query content and time interval. The process of sharing query content as well as generating interval times is shown in algorithm 2.

```
Algorithm 2: Process of sharing query content and time interval
Input: C, T//query content(bits) and time interval(seconds) set
Output: C, T
1   each user in G broadcast C and T;
2   u chooses every user in G and gains their C and T;
3   C=C+c;
4   T=gcd(C,c);//computes the greatest common divisor
5   u broadcast C and T;
6   for other users in G
7       update C and T;
8   end
```

In Algorithm 2, the group initiator first gathers other users' query contents and time intervals, then adds its own query content and computes the greatest common divisor of time intervals in line 3-4. In line 5-8, the initiator broadcasts the processing result set to this group, and everyone changes the query content and time interval, then users in this group will show a common profile when leaving this mix-zone.

4.2 Security Analyses

Security of CoMix is determined by the uncertainty of correlating a particular user. As the adversary can utilize correlation probabilities to identify subsequent locations, the main motivation is to generalize two profiles in the same group. To this end, entropy is utilized to measure the privacy level, which presents the uncertainty of correlating the sub-trajectory and subsequent locations.

The resistance to correlation attacks of CoMix can be verified by a game played between challenger *A* and user

U. In this game, *A* prepares two trajectories (T_0,T_1), and gives them to *U*. *U* randomly chooses a number $b \in \{0,1\}$ to indicate the chosen one, and generates a profiles message Msg_b with query contents and time intervals shown in these two trajectories and sends it to *A*. The challenger will win the game if *A* can output a bit *b'*=*b* by the profile correlation attack. From this game, a definition demonstrates what kind of scheme can resistant the profile correlation attack is proposed.

Definition 2 An algorithm is resistant to the profile correlation attack if for each subsequent location

$$\begin{aligned} &Pr(l_i \in T \mid sim(p_L{}^{(c,t)}, p_{l_i}{}^{(c,t)})) \\ &= Pr(l_j \in T \mid sim(p_L{}^{(c,t)}, p_{l_j}{}^{(c,t)})) \\ &\quad \forall (0 < i \neq j \leq k) \end{aligned} \tag{3}$$

Theorem 1 The CoMix is resistant to the profile correlation attack.

Proof: For an arbitrary chosen location l_i out of the mix-zone, the probability of correlating it to the sub-trajectory can be indicated as

$$\begin{aligned} &Pr(l_i \in T \mid sim(p_L{}^{(c,t)}, p_{l_i}{}^{(c,t)})) \\ &= \frac{Pr(l_i \in T)}{Pr(sim(p_L{}^{(c,t)}, p_{l_i}{}^{(c,t)}))} \end{aligned} \tag{4}$$

Similarly, for another location l_j from the same group, its probability of correlation is

$$\begin{aligned} &Pr(l_j \in T \mid sim(p_L{}^{(c,t)}, p_{l_j}{}^{(c,t)})) \\ &= \frac{Pr(l_j \in T)}{Pr(sim(p_L{}^{(c,t)}, p_{l_j}{}^{(c,t)}))} \end{aligned} \tag{5}$$

Then for the pair of locations l_i and l_j, Eq. (3) holds if

$$\begin{aligned} &Pr(sim(p_L{}^{(c,t)}, p_{l_i}{}^{(c,t)})) \\ &= Pr(sim(p_L{}^{(c,t)}, p_{l_j}{}^{(c,t)})), \\ &\quad \forall (0 < i \neq j \leq k) \end{aligned} \tag{6}$$

In the algorithm of CoMix, query content and time interval are converted into content set and the greatest common divisor of time intervals. The consistency guarantees that the user who leaves current mix-zone shows a similar profile with each other. It means that for any randomly chosen user, it will have the same probability to be correlated with the sub-trajectory by profiles similarity, and then Eq. (6) holds. Therefore, the scheme gains the maximum entropy, and the adversary has the highest uncertainty in determining which location belongs to the gained sub-trajectory.

5 EXPERIMENTAL EVALUATION

In this section, evaluations of CoMix on efficiency and effectiveness are proposed. The efficiency of performance is expressed by the process time and mix-zone size. Privacy effectiveness is depicted by entropy and the success cloak ratio.

5.1 Evaluation Criteria and Metrics

To evaluate the efficiency and effectiveness, the evaluation is implemented on a laptop with Intel Core i5 1.70 GHz CPU, 4 GB RAM memory, and Windows 7×64 ultimate operating system. Then the central part of the BerlinMOD Data Set[1] is used, and a randomly chosen intersection is used to deploy mix-zones. A set of trajectories is chosen and assumes they all passed the given mix-zone at a randomly chosen time.

The process time is determined by the time of establishing the exchanging group and sharing the similar profiles. It was affected by the user number of current mix-zone, so the evaluation of the efficiency of various algorithms has to assume that the user enters the mix-zone one after another and without leaving, so that the max time can be found in various forms of mix-zone.

In general, a larger mix-zone may provide privacy for more users, but it will expend more deployment cost. Therefore, an appropriate mix-zone size is distinctly important to the tradeoff privacy and deployment cost. To this end, the size of the mix-zone is used to evaluate the performance of three different forms of mix-zone, and the optimal one will show the best performance.

The success ratio of correlating subsequent locations depends on the disturbance of entering and leaving pairs as well as the disturbance of all leaving users. This condition makes the privacy level depend on the profile similarity between the entering and leaving users and the level of real user hidden behind others. Therefore, two different forms of entropy are used. Then combined with the correlation attack launched by passive or active adversaries, the entropy can be denoted as

$$E_{mount} = \left(E_g + E_l\right)/2 \tag{7}$$

The success cloak ratio depends on the probability of encountering number of collaborative users. If no user was encountered in mix-zone, the user cannot establish the exchanging group. Thus, the success cloak ratio is calculated by the percentage of users successfully establishing the exchanging group, and it can be computed by

$$CSR = \sum \left|U_S\right| / |S| \tag{8}$$

where *Us* is the set of users successfully establishing the exchanging group, and *S* is the set of all users passed through this zone. In addition, higher success ratio corresponds to better privacy preservation.

[1] http://dna.fernuni-hagen.de/secondo/BerlinMOD/BerlinMOD.html

5.2 Evaluation Results

In Fig. 4, process time of three different types of mix-zones is shown. Although all process times are ascending along with the increasing number of applicants, performance of shifted rectangular mix-zone is better than other two schemes, and the non-rectangular mix-zone is the worst. The reason is that with the process of establishing the exchanging group, some road segments cannot be covered by the non-rectangular mix-zone, which causes the number of users in current segments does not satisfy the anonymity degree, and users have to wait even longer for collaborative users in this type of zone. Additionally, as the inhomogeneous distribution of collaborative users, the shifted rectangular mix-zone can shift to the dense direction, and leads the initiator much easier to find collaborative users.

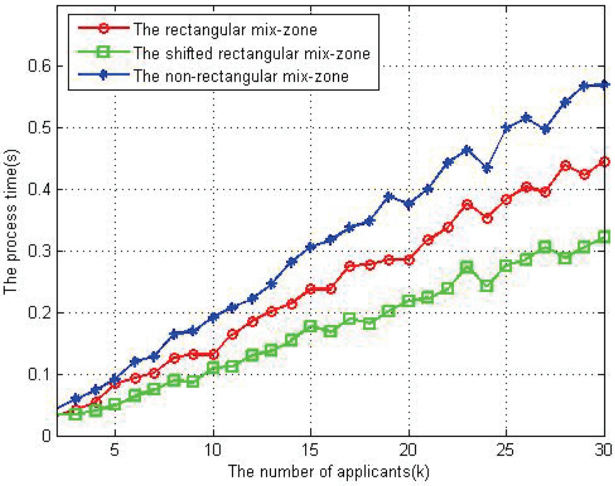

Figure 4 The process time

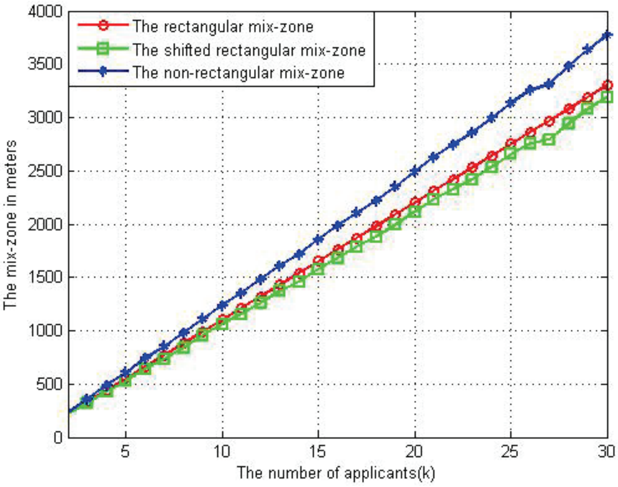

Figure 5 The mix-zone size

Fig. 5 shows the size of mix-zones followed with the increasing number of applicants. From this figure, the shifted rectangular mix-zone performs better than the other two, the rectangular one is worse only a little, but the non-rectangular one is the worst. This is because of that the mix-zone needs to provide enough space for covering more adjacent users, so the number of users in a certain period determines the size. This condition leads the failure of non-rectangular mix-zone, as the passing velocity of users changes its shape. With a higher velocity, it is harder for this zone to contain more users. Therefore, in order to provide services to all users in current region, this type of mix-zone has to be expanded even larger.

In general, entropy is used to ensure that the distribution of the mapping probabilities does not deviate much from the uniform distribution. Fig. 6 shows the pairwise entropy of users, which presents the degree of considering the profile given to a particular user. In this figure, CoMix is a bit less than the maximum entropy in theory, because this paper mainly considers the correlation of query content and time interval, and there will be other attack types utilized by the adversary.

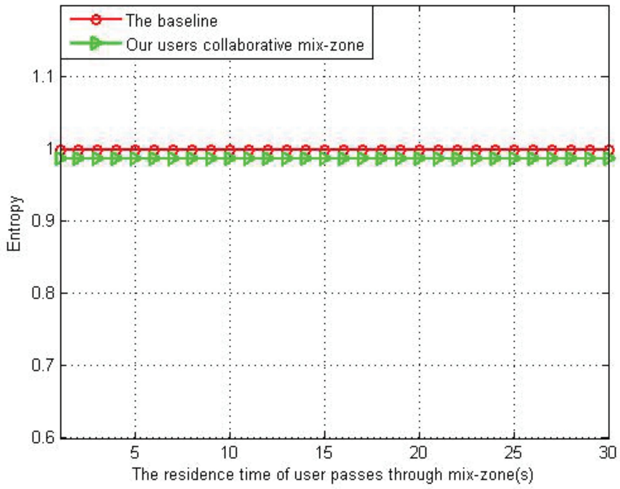

Figure 6 The pairwise entropy

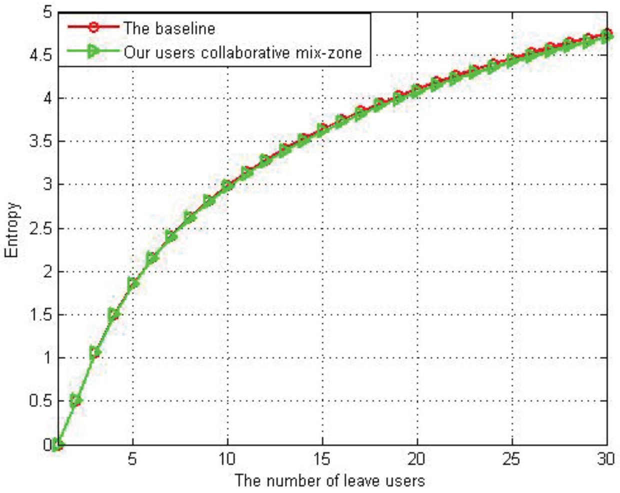

Figure 7 The entropy of subsequent locations

Fig. 7 shows the entropy of subsequent locations, it is computed with the locations of users in the same exchanging group. Generally, the entropy is increasing with the uncertainty of correlation between subsequent locations and sub-trajectory. The baseline is the highest entropy ($\log_2 k$), but our algorithm does not reach that. It is because of that the user passing through the mix-zone without encountering any collaborative user influences the entropy as the user fails to establish the exchanging group.

Fig. 8 shows the success cloak ratio of a user exchanges profile with collaborative users. In this figure, the success ratio is ascending with the increasing of resident time. This is because a longer residence time leads to a higher probability to encounter with collaborative users. Besides the above phenomenon, this figure also shows another interesting phenomenon. With the increasing number of users passing through the mix-zone, the success ratio of exchanging profiles is decreasing, and more users are passing through the lower of this ratio. The reason is that the possibility of meeting collaborative users is an independent random event and it does not depend on the residence time. Thus, more users passing through the mix-zone may bring more failure to encounter collaborative users.

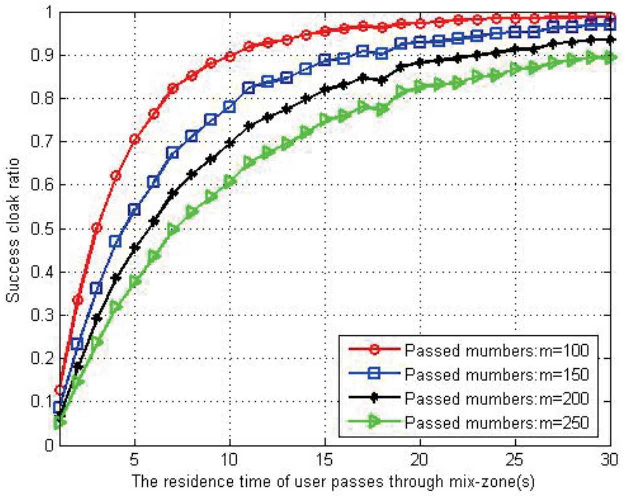

Figure 8 The success cloak ratio

6 CONCLUSION

In this paper, an algorithm to resist the profile correlation attack is proposed. As query contents and time intervals are exchanged with collaborative users in the mix-zone, each user shows a similar profile, and the similar profile makes the adversary difficult to identify a particular user by profile correlation attack. Then, the entropy and the success cloak ratio are utilized to measure the privacy level, and the security of CoMix is analyzed. Finally, evaluation results show CoMix has a better efficiency and effectiveness in preservation and performance. However, there are still some problems unsolved in our algorithm such as the adversary can correlate the subsequence locations with the undiscovered profile as well as the insufficient collaborative users will reduce the success cloak ratio.

Acknowledgments

This work was supported by National Natural Science Foundation of China No. 61472097, Specialized Research Fund for the Doctoral Program of Higher Education No. 20132304110017, Natural Science Foundation of Heilongjiang Province of China No. F2015022, University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province No. UNPYSCT-2017149 the National Undergraduate Innovation and Entrepreneurship Training Program of China No. 201810222033.

7 REFERENCES

[1] Hwang, R. H., Hsueh, Y. L., & Chung, H. W. (2014). A Novel Time-Obfuscated Algorithm for Trajectory Privacy Protection. *IEEE Transactions on Services Computing, 7*(2), 126-139. https://doi.org/10.1109/TSC.2013.55

[2] Gao, S., Ma, J., & Shi, W. (2015). LTPPM: a location and trajectory privacy protection mechanism in participatory sensing. *Wireless Communications & Mobile Computing, 15*(1), 155-169. https://doi.org/10.1002/wcm.2324

[3] Kato, R., Iwata, M., & Hara, T. (2012). A dummy-based anonymization method based on user trajectory with pauses. *International Conference on Advances in Geographic Information Systems*, 249-258. https://doi.org/10.1145/2424321.2424354

[4] Wang, Y., He, L. P., & Peng, J. (2012). Privacy Preserving for Continuous Query in Location Based Services. *International Conference on Parallel and Distributed Systems*, 213-220. https://doi.org/10.1109/ICPADS.2012.38

[5] Palanisamy, B., Liu, L., & Lee, K. (2014). Anonymizing Continuous Queries with Delay-tolerant Mix-zones over Road Networks. *Distributed & Parallel Databases, 32*(1), 91-118. https://doi.org/10.1007/s10619-013-7128-4

[6] Palanisamy, B. & Liu, L. (2015). Attack-Resilient Mix-zones over Road Networks: Architecture and Algorithms. *IEEE Transactions on Mobile Computing, 14*(3), 495-508. https://doi.org/10.1109/TMC.2014.2321747

[7] Hashem, T., Kulik, L., & Zhang, R. (2013). Countering overlapping rectangle privacy attack for moving k NN queries. *Information Systems, 38*(3), 430-453. https://doi.org/10.1016/j.is.2012.07.001

[8] Beresford, A. R. & Stajano, F. (2003). Location Privacy in Pervasive Computing. *Pervasive Computing IEEE, 2*(1), 46-55. https://doi.org/10.1109/MPRV.2003.1186725

[9] Ma, C. G., Zhou, C. L., & Yang, S. T. (2015). A voronoi-based location privacy-preserving method for continuous query in LBS. *International Journal of Distributed Sensor Networks, 2015*, 1-17. https://doi.org/10.1155/2015/326953

[10] Schlegel, R., Chow, C. Y., & Huang, Q. (2015). User-Defined Privacy Grid System for Continuous Location-Based Services. *IEEE Transactions on Mobile Computing, 14*(10), 2158-2172. https://doi.org/10.1109/TMC.2015.2388488

[11] Chow, C. Y. & Mokbel, M. F. (2007). Enabling Private Continuous Queries for Revealed User Locations. *Advances in Spatial and Temporal Databases, 2007*, 258-275. https://doi.org/10.1007/978-3-540-73540-3_15

[12] Ghinita, G., Damiani, M. L., & Silvestri, C. (2009). Preventing velocity-based linkage attacks in location-aware applications. *ACM Sigspatial International Conference on Advances in Geographic Information Systems*, 246-255. https://doi.org/10.1145/1653771.1653807

[13] Gruteser, M. & Grunwald, D. (2003). Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. *International Conference on Mobile Systems*, 31-42. https://doi.org/10.1145/1066116.1189037

[14] Rebollo-Monedero, D., Forné, J., & Solanas, A. (2010). Private location-based information retrieval through user collaboration. *Computer Communications, 33*(6), 762-774. https://doi.org/10.1016/j.comcom.2009.11.024

[15] Niu, B., Zhu, X., & Li, Q. (2014). A novel attack to spatial cloaking schemes in location-based services. *Future Generation Computer Systems*, 49, 125-132. https://doi.org/10.1016/j.future.2014.10.026

[16] Liu, F. Y., Hua, K. A., & Cai, Y. (2009). Query l-diversity in Location-Based Services. *Tenth International Conference on Mobile Data Management*, 436-442. https://doi.org/10.1109/MDM.2009.72

[17] Gedik, B. & Liu, L. (2005). Location Privacy in Mobile Systems: A Personalized Anonymization Model. *IEEE*

International Conference on Distributed Computing Systems, 620-629. https://doi.org/10.1109/ICDCS.2005.48

[18] Bamba, B., Liu, L., & Pesti, P. (2008). Supporting anonymous location queries in mobile environments with privacygrid. *International Conference on World Wide Web*, 237-246. https://doi.org/10.1145/1367497.1367531

[19] Wang, T. & Liu, L. (2009). Privacy-Aware Mobile Services over Road Networks. *Proceedings of the Vldb Endowment, 2*(1), 1042-1053. https://doi.org/10.14778/1687627.1687745

[20] Ghinita, G., Kalnis, P., & Khoshgozaran, A. (2008). Private queries in location based services: anonymizers are not necessary. *International Conference on Management of Data*, 121-132. https://doi.org/10.1145/1376616.1376631

[21] Khoshgozaran, A. & Shahabi, C. (2009). Private Information Retrieval Techniques for Enabling Location Privacy in Location-Based Services. *Privacy in Location-Based Applications*, 59-83. https://doi.org/10.1007/978-3-642-03511-1_3

[22] Lien, I. T., Lin, Y. H., & Shieh, J. R. (2013). A Novel Privacy Preserving Location-Based Service Protocol With Secret Circular Shift for K-NN Search. *IEEE Transactions on Information Forensics & Security, 8*(6), 863-873. https://doi.org/10.1109/TIFS.2013.2252011

[23] Chow, C. Y., Mokbel, M. F., & Liu, X. (2006). A peer-to-peer spatial cloaking algorithm for anonymous location-based service. *ACM International Symposium on Geographic Information Systems*, 171-178. https://doi.org/10.1145/1183471.1183500

[24] Chow, C. Y., Mokbel, M. F., & Liu, X. (2011). Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. *Geoinformatica, 15*(2), 351-380. https://doi.org/10.1007/s10707-009-0099-y

[25] Rebollo-Monedero, D., Forne, J., & Domingo-Ferrer, J. (2012). Query Profile Obfuscation by Means of Optimal Query Exchange between Users. *IEEE Transactions on Dependable & Secure Computing, 9*(5), 641-654. https://doi.org/10.1109/TDSC.2012.16

[26] Wang, Y., Xia, Y., & Hou, J. (2015). A fast privacy-preserving framework for continuous location-based queries in road networks. *Journal of Network & Computer Applications*, 53, 57-73. https://doi.org/10.1016/j.jnca.2015.01.004

[27] Rodriguezcarrion, A., Rebollomonedero, D., & Forné, J. (2015). Entropy-Based Privacy against Profiling of User Mobility. *Entropy, 17*(6), 3913-3946. https://doi.org/10.3390/e17063913

[28] Freudiger, J., Raya, M., & Félegyházi, M. (2007). Mix-Zones for Location Privacy in Vehicular Networks. *ACM Workshop on Wireless Networking for Intelligent Transportation Systems*.

[29] Freudiger, J., Shokri, R., & Hubaux, J. P. (2015). On the Optimal Placement of Mix Zones. *International Symposium on Privacy Enhancing Technologies*, 216-234.

[30] Le, Z., Ouyang, Y., & Chen, G. (2011). Dynamic mix zone: location data sanitizing in assisted environments. *Universal Access in the Information Society, 10*(2), 195-205. https://doi.org/10.1007/s10209-010-0198-4

[31] Palanisamy, B. & Liu, L. (2011). MobiMix: Protecting location privacy with mix-zones over road networks. *IEEE International Conference on Data Engineering*, 494-505. https://doi.org/10.1109/ICDE.2011.5767898

[32] Palanisamy, B. & Liu, L. (2014). Effective mix-zone anonymization techniques for mobile travelers. *Geoinformatica, 18*(1), 135-164. https://doi.org/10.1007/s10707-013-0194-y

[33] Gao, S., Ma, J., & Shi, W. (2013). TrPF: A Trajectory Privacy-Preserving Framework for Participatory Sensing. *IEEE Transactions on Information Forensics & Security, 8*(6), 874-887. https://doi.org/10.1109/TIFS.2013.2252618

[34] Sun, Y., Zhang, B., & Zhao, B. (2014). Mix-zones optimal deployment for protecting location privacy in VANET. *Peer-to-Peer Networking and Applications, 8*(6), 1-14.

[35] Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE, 70*(9), 939-952. https://doi.org/10.1109/PROC.1982.12425

Contact information:

Zhang LEI, PhD student
College of Computer Science and Technology,
Harbin Engineering University,
Harbin 150001, China
College of Information and Electronic Technology,
Jiamusi University,
Jiamusi 154007, China
8213662@163.com

Ma CHUN-GUANG, PhD
Corresponding author
College of Computer Science and Technology,
Harbin Engineering University,
Harbin 150001, China
machunguang@hrbeu.edu.cn

Yang SONG-TAO, PhD
College of Computer Science and Technology,
Harbin Engineering University,
Harbin 150001, China
College of Information and Electronic Technology,
Jiamusi University,
Jiamusi 154007, China
songtao_y@163.com

Zheng XIAO-DONG, PhD student
College of Computer Science and Technology,
Harbin Engineering University,
Harbin 150001, China
Applied Technology College, Qiqihar University,
Qiqihar 161006, China
lnxiaodong@126.com