Dummy-Based Schemes for Protecting Movement Trajectories

PO-RUEY LEI[1], WEN-CHIH PENG[2], ING-JIUNN SU[1] AND CHIEN-PING CHANG[3]
[1]*Chung Cheng Institute of Technology*
National Defense University
Taoyuan, 334 Taiwan
[2]*Department of Computer Science*
National Chiao Tung University
Hsinchu, 300 Taiwan
[3]*Department of Computer Science and Information Engineering*
Ching Yun University
Chungli, 320 Taiwan

Dummy-based anonymization techniques for protecting the location privacy of mobile users have appeared in the literature. By generating dummies that move in human-like trajectories, this approach shows that the location privacy of mobile users can be preserved. However, the trajectories of mobile users can still be exposed by monitoring the long-term movement patterns of users. We argue that, once the trajectory of a user is identified, the locations of the user are exposed. Thus, it is critical to protect the movement trajectories of mobile users in order to preserve user location privacy. We propose two schemes that generate consistent movement patterns in the long run. Guided by three parameters in a user specified privacy profile, namely, *short-term disclosure*, *long-term disclosure* and *distance deviation*, the proposed schemes derive movement trajectories for dummies. The experimental results show that our proposed schemes are more effective than existing works in protecting movement trajectories.

Keywords: location privacy, user movement patterns, location-based services, trajectory pattern protection, dummy-based anonymization

1. INTRODUCTION

Location-based services (LBSs) have emerged as one of the killer applications for mobile computing and wireless data services. Due to the ubiquitous mobility of users in the mobile computing environments, users are often interested in acquiring information or services related to their locations. Upon requesting a service, the location information of a user is submitted along with the query to a LBS provider. These location-dependent queries may disclose sensitive and private information about users by cross-referencing their location with information of queried location-based data. If not well protected, the location information of users may be misused by some untrustworthy service providers or attackers. Obviously, it is important to protect location privacy in LBSs.

The problem of location privacy preservation has received growing interests in the research community [5, 10, 11, 14]. These studies aim at protecting the exact location information of users from the potential abuse of LBS providers and hackers. Two primary approaches have been considered: (1) the trusted anonymizer based approach; and (2) the client based approach. In the former, users submit their queries to the LBSs via a third-party trusted server, called a trusted anonymizer. This trusted anonymizer trans-

Received April 7, 2010; revised December 28, 2010; accepted February 11, 2011.
Communicated by Vincent S. Tseng.

forms the exact locations of a number of users into a cloaked spatial area as a query to obtain data or services from the LBSs [5, 14]. The second approach assumes no trusted server. Thus, clients are responsible for anonymizing their own location information before transmitting queries to the LBS servers. By issuing several fake locations along with their true location to the LBSs, clients may obtain redundant information or services corresponding to the submitted locations while preserving their location information [10, 11]. Unwanted information is filtered locally to obtain the final query results. In both approaches, the true location of a user is either (1) not distinguishable from those of other users (the trusted anonymizer based approach), or (2) not distinguishable from the fake locations (the client based approach). Since a trusted server is not always available, in this paper, we tackle the issues faced in the client based approach.

Without relying on a trusted server, generating fake user locations (called dummies1) for location-dependent queries has been shown to be an effective way of preserving location privacy [10]. The authors in [10] propose generating dummies whose movements are consecutive, reflecting realistic user movements. However, the prior work in [10] does not consider a well-recognized observation, *i.e.*, the movement behaviors of users usually follow certain patterns [15, 16]. Consequently, adversaries may discover the movement patterns of users and distinguish the trajectories of true users from the dummies. Fig. 1 shows an example that illustrates the problem we discussed. In the figure, the solid line denotes the movement trajectory of a true user (denoted as *T*) while the dotted lines are the generated trajectories of dummies (denoted as *d*1 and *d*2). Since true users usually exhibit certain human movement behavior, one is able to identify the solid line as a true user based on the typical movement behavior of humans (as shown in Fig. 1 (a)). Thus, it is important to generate dummy trajectories based on human movement behavior (as shown in Fig. 1 (b)). Even though this effort may reduce the chance of the true moving trajectory being identified, a long-term movement pattern can still be collected to filter inconsistent trajectories. For example, comparing the current trajectories (in Fig. 1 (c)) with trajectories collected on a different day (*e.g.*, Fig. 1 (b)), one can tell that *T* is the true trajectory of the user. Once the true movement trajectory of the user is identified, the locations (*i.e.*, not only the current location but also the past locations) of the user are disclosed. Thus, we propose that dummy generation not only demonstrates the movement behavior of the users but also follows consistent, long-term movement patterns.

Given that adversaries can obtain a set of trajectories, they may have difficulty determining the true trajectory of a user if dummies are generated following certain movement patterns. However, the user trajectory is still disclosed to a certain degree. Therefore, we use disclosure to denote the probability that the user trajectory may be correctly identified by adversaries. For example, in Fig. 1 (c), three trajectories are collected and thus the disclosure is 1/3. To reduce the disclosure, a naive approach is to simply increase the number of dummies. With the increase in the number of dummies, the number of query results which tends to increase may incur communication and client processing costs. Thus, in this paper, we propose some efficient schemes to generate as few dummy trajectories as possible while still decreasing the chance of disclosure of the user trajectory.

Nevertheless, an issue exists with regards to these intersecting trajectories. Consider the example in Fig. 1 (d); when the generated dummy trajectories are too close to the true trajectory, a user's movement trajectory (the shadowed path) may still be exposed and identified. Thus, our proposed schemes for dummy generation take the factor of distance

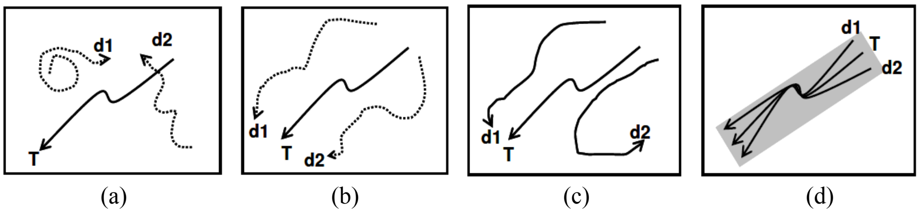


Fig. 1. Moving trajectories of user and dummies.

deviation among trajectories into consideration. Our approach is to allow users to set up their privacy profile in terms of disclosures (both short-term and long-term) and distance deviation (more details to be discussed in section 3). We propose two schemes, namely, the random pattern scheme and the intersection-based pattern schemes, to generate dummy trajectories based on a user's privacy profile. We conduct our experiments on both a real dataset and a synthetic dataset. The experimental results show that our schemes perform better than the existing techniques.

The rest of this paper is organized as follows. Related works are presented in section 2. Section 3 provides preliminaries, including assumption and user profile. Section 4 presents our proposed schemes for generating dummies. Section 5 is devoted to the experimental results. Finally, section 6 concludes this paper.

2. RELATED WORKS

A significant amount of research efforts have been elaborated on location privacy. These studies have worked on finding the solutions to allow the mobile users to request LBSs without disclosing their privacy. As described before, according to the architecture of location privacy, these techniques can be further classified into two categories: the trusted anonymizer based approach and the client based approach. With trusted anonymity servers, many studies have proposed location cloaking algorithms [2, 7, 18] to reduce the resolution of location information along spatial and temporal dimensions by blurring the users' locations into cloaking regions. By applying k-anonymity methods to location cloaking, the server computes a cloaking region that contains the client user and $(k - 1)$ other users, and then uses the cloaking region instead of the client user location to request the LBS. Based on k-anonymity, the authors in [4, 5] devised a personalized and customized k-anonymity model that assumes a different k-anonymity requirement for each user. A framework, called Casper, was developed in [14], where a grid-based pyramid structure is implemented to index user locations for both k-anonymity and granularity metrics. Moreover, privacy-aware query processing is developed when cloaked spatial areas are used as query predicates. The drawback of the trusted anonymizer approach is that the anonymizer must be trustworthy to all users and reliable against a single point of failure. In other words, such a trusted server may not always be available.

Without assuming trusted servers, some client based privacy protection techniques have been proposed. The frameworks based on cryptographic techniques [6, 9] develop a spatial transformation to provide location privacy. User locations are encrypted and are only known by the clients, but this approach could incur additional computation and communication costs. The dummy based location privacy is proposed by [10-12]. In the proposed techniques, a user sends the true user location mixed with fake locations, *i.e.* dummies, to the LBS provider. The user extracts the necessary information from the reply message while the service provider can only learn vague details of the user location. However, the problem of the existing dummy based approach, which is also the main problem associated with the other above works, is that they only consider the snapshot scenario. Adversaries can comprehend the user movements when tracking data during long-term observations. The sequence of location data can be used to discover a rough trajectory by applying data mining techniques. Although the authors in [13, 17] brought the concept of trajectory protection under historical attack, similar to the problem our research addresses, they have to promise that the LBS provider is reliable and secure because of their cloaking approach.

Our research addresses location privacy issues from the aspect of long-term observation, and is essentially a dummy based approach to get rid of the trusted server. Since users usually follow certain movement patterns, we intend to generate dummies whose trajectories have consistent and long-term movement patterns. Furthermore, to protect user movement trajectories, a new kind of user privacy profile is proposed. Based on the user privacy profile, we develop some efficient schemes to generate dummies to fulfill the user privacy profiles. These features differentiate our paper from the existing works.

3. PRELIMINARIES

In this section, some assumptions and notations are presented first. Finally, a new kind of user privacy profile is presented.

3.1 Assumptions and Notations

We assume that no trusted server is available for location anonymization. Upon receiving a user query, a mobile client U_i sends the query to the LBSs through a secured connection that will not get hijacked. A query message issued by a mobile user U_i to a LBS server at time slot t is defined as $M = \{uid, \langle L^t_i, L^t_{d1}, L^t_{d2}, \ldots, L^t_{dn}\rangle, Q\}$, where uid is the pseudonym user identification, L^t_i is the true user location, $L^t_{d1}, L^t_{d2}, \ldots, L^t_{dn}$ are n dummy locations, respectively, and Q is the location-dependent query issued. Therefore, given m consecutive queries, the trajectory of mobile user U_i is $\{L^1_i, L^2_i, \ldots, L^m_i\}$; while the trajectory of dummy d_x is $\{L^1_{dx}, L^2_{dx}, \ldots, L^m_{dx}\}$. Here, L^j_i (and L^j_{dx}, respectively) denotes the location of user U_i (and dummy d_x, respectively) at the jth time slot. Denote a trajectory of mobile user U_i as $T_i = \{TL^1_i, L^2_i, \ldots, L^m_i\}$, where TL^j_i is the location in a trajectory T of mobile user U_i at the jth time slot.

3.2 User Privacy Profile

Our proposal of dummy pattern generation allows the users to set up their privacy profiles to fulfill the requirements to protect their location and trajectory privacy. We have to consider the following three requirements to enhance privacy preservation before explaining how the proposed schemes work to achieve the user location privacy by dummy pattern generation:

Requirement 1 Given the current locations of a true user and dummies, the probability of successfully identifying the user's true location is smaller than a threshold value specified.

Requirement 2 Given a true user trajectory and n dummies, the probability of successfully identifying the true trajectory is smaller than a threshold value specified.

Requirement 3 Given a true user trajectory and n dummies, the average of distance difference, which is denoted as distance deviation among the trajectories of a true user and dummies must be larger than a threshold value specified.

Based on the above requirements, the disclosure probability of a user's location or trajectory can be reduced. An adversary may not be able to distinguish the user's trajectory or locations from dummies. As such, both location and trajectory privacy can be preserved. Therefore, users may set up their privacy profile, which is specified by the following three parameters corresponding to the requirements:

1. **Short-term Disclosure (SD):** This parameter specifies the requirement for protecting the current user location. Thus, given a set of current locations (including true and dummy locations), SD is the probability of successfully identifying the true user location, *i.e.*, $SD = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{|D_i|}$, where m is the number of time slots in a trajectory, D_i is the set of true and dummy locations at the ith time slot, and $|D_i|$ is the size of D_i.
2. **Long-term Disclosure (LD):** This parameter specifies the requirement for protecting the user trajectory. Given n trajectories, among which k trajectories have intersected with other trajectories and $(n - k)$ trajectories do not have any intersections. Thus, for those $(n - k)$ trajectories, we have exactly $(n - k)$ possible trajectories. For those k trajectories, we may enumerate all possible trajectories by exhaustively traversing intersections from the start point of each trajectory to the end point. In order not to distract readers from the main theme of this paper, we simply denote the number of possible trajectories among k trajectories as T_k. Consequently, we have LD as $\frac{1}{T_k + (n-k)}$.
3. **Distance Deviation (dst):** The distance deviation (dst) is the average of the distance difference among the trajectories of the dummies and the user. As a result, the dst of mobile user U_i is formulated as $\frac{1}{m}\times\frac{1}{n}\times\sum_{k=1}^{n}\times\sum_{j=1}^{m} dist(TL_i^j)$, where $dist$ is the average distance between the true user location and the dummy locations in the unit of grid cell size.

4. DUMMY-BASED SCHEMES FOR GENERATING DUMMIES WITH MOVEMENT PATTERNS

Given a privacy profile, our goal is to generate dummies with long term, consistent movement patterns (*i.e.* dummy trajectories) that satisfy the user privacy requirements. Explicitly, given a frequent movement trajectory of a user, we propose two dummy generation schemes to generate dummy trajectories, namely the *random pattern scheme* and the *intersection pattern-based schemes*.

4.1 Random Pattern Scheme

To generate dummies with movement patterns, a dummy with its own starting point and destination should first be determined. After that, a movement trajectory between the starting point and the destination is randomly derived. Specifically, a movement trajectory consists of a sequence of grid cells that are decided by the speed of a dummy and three movement types, where the movement types are horizontal movement, vertical movement, and both horizontal and vertical movement. Due to the movement nature of humans, the velocity of dummies should be bounded and smaller than user maximum moving velocity. Once a movement trajectory between the starting point and the destination is derived, the dummy will frequently follow the user movement trajectory. As such, even after a long term observation, it is difficult for adversaries to discern the true user and the dummy trajectories since the dummies also exhibit long term and consistent movement patterns. As pointed out earlier, the generation of dummies should satisfy the user privacy profile. Thus, we will evaluate whether the number of dummies with their frequent movement trajectories is sufficient for user privacy profiles or not. If the user privacy profile is not satisfied, more dummies are simply added in this scheme. Hence, in the following scheme, we will judiciously derive dummy trajectories to fulfill user privacy profiles with the purpose of minimizing the number of dummies.

4.2 Intersection Pattern-based Scheme

The main idea of this scheme is to minimize the number of dummies while still guaranteeing the user privacy requirements. The concept is to have some intersections among the trajectories of the dummies and a user. Clearly, increasing the number of intersections is beneficial to both short-term and long-term disclosures. We thus propose two intersection-based dummy generation methods, *i.e.*, rotation dummy generation and *K*-intersect dummy generation.

4.2.1 Rotation dummy generation

Given a user trajectory, we generate a new trajectory for a dummy by rotating the known user trajectory. Clearly, the rotation point of the true user trajectory is an intersection point. Fig. 2 shows an example of generating dummies by rotating the true user trajectory, where the dotted trajectory is the dummy trajectory.

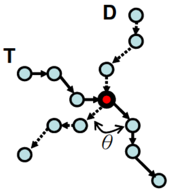


Fig. 2. An example of generating dummy trajectories by rotating a true user trajectory.

In this rotation pattern scheme, generated dummy trajectories should fulfill the privacy profile of the user. Since there are three requirements in privacy profiles, our approach first derives the solution space for the requirement of distance derivation. Then, within this solution space, we obtain the short-term and long-term disclosures (*i.e.*, *SD* and *LD*). The trajectories with disclosures that are smaller than what specified in privacy profile are selected as the dummy trajectories. With the proper selection of dummy trajectories, we can minimize the number of dummies so as to satisfy the user privacy requirements. In order to derive the solution space for the distance derivation (*i.e.*, *dst*); both the rotation angle and the rotation point within a true user trajectory have a great impact on the distance deviation.

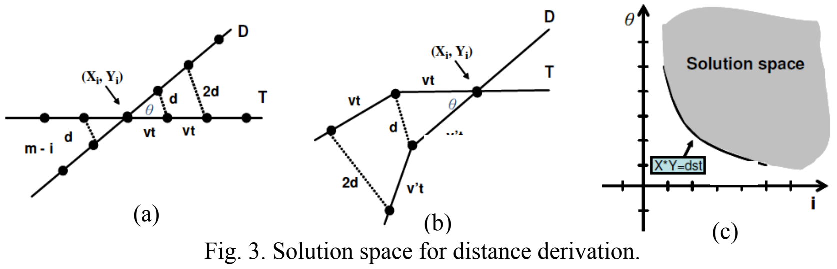


Fig. 3. Solution space for distance derivation.

To simplify the derivation of the distance deviation, assume that we have a true user trajectory *T* intersected with a dummy trajectory *D*, which is generated by *T*, in Fig. 3 (a), where the distance between two consecutive movements is *vt* under the assumption that the user moving speed is *v* and the length of a time slot is *t*. The rotation point is the location (X_i, Y_i) at the *i*th time slot in a true user trajectory, and the rotation angle is *θ*. *d* is the distance difference between the location of a true user and that of a dummy at the (*i* + 1)th time slot.

According to the cosine theorem, we have $d = \sqrt{2}|vt|\sqrt{1-\cos\theta}$. Hence, we could derive the distance deviation of these two trajectories as follows,

$$dst^r = \frac{1}{m} \times ((d + \ldots + id) + (d + \ldots + (m-i)d)) \tag{1}$$

$$= \sqrt{2}|vt|\sqrt{1-\cos\theta} \times (\sum_{j=0}^{i} j + \sum_{j=0}^{m-i} j). \tag{2}$$

From the above derivation, we could conclude that both the rotation angle (*i.e.*, *θ*) and the rotation point (*i.e.*, *i*) are important to distance deviation. Furthermore, assume that we have *n* dummy trajectories and the distance deviation of *n* dummy trajectories is dst_n. If one dummy is added into the set of *n* dummies, the (*n* + 1) dummies should be larger or equal to the requirement of the distance deviation (*i.e.*, *dst*). Thus, we have the following formula:

$$\frac{n}{n+1} dst_n + \frac{1}{n+1} dst^r \geq dst$$

$$\Rightarrow dst^{r'} \geq (n+1) \times dst - n \times (dst_n). \tag{3}$$

Consequently, when one dummy is added into the current set of dummies, this dummy should have a constraint on $dst^{r'} \geq (n + 1) \times dst - n \times (dst_n)$. Therefore, we could have the solution space shown in Fig. 3 (c). For each point (expressed by (θ, i)) in the solution space, we should calculate the corresponding disclosure and then select the solution point with minimal disclosures. If the disclosures are still larger than the required disclosures, one should repeat the above procedure to add one additional dummy until all of the requirements in the privacy profile are satisfied.

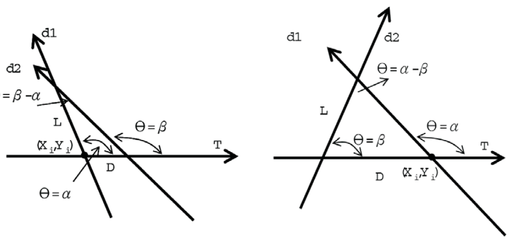
Note that with more intersections among trajectories, lower disclosures are expected. Thus, we consider letting newly added dummies have more intersections with the existing trajectories, including a true user trajectory and dummy trajectories. For example, in Fig. 4, dummy $d2$ is added and there are already two trajectories (*i.e.*, one true trajectory T and one dummy trajectory $d1$). The added dummy $d2$ has intersections not only with the true trajectory T but also with another dummy trajectory $d1$. Now, we have to determine some constraints on the rotation angle between T and $d2$ (*i.e.*, β) to increase the number of inter-sections among trajectories. Suppose that the rotation angle between T and $d1$ is α and the intersection of trajectory i and j is denoted as $I_{i,j}$. In Fig. 4 (a), the distance between $I_{T,d1}$ and $I_{d1,d2}$ is L, and the distance between $I_{T,d1}$ and $I_{T,d2}$ is D. By exploring the sine theorem, we have the following formula:

$$\frac{D}{\sin(\beta - \alpha)} = \frac{L}{\sin(180° - \beta)}.$$

Obviously, the rotation angle (*i.e.*, β) has two cases to be considered – $d1$ intersects with T before $d2$ intersects with T ($\beta \geq \alpha$), or after $d2$ intersects with T ($\beta < \alpha$) as shown in Figs. 4 (a) and (b), respectively. Let m be the total time slot and i be the intersection of $d1$ and T. If $d1$ and $d2$ intersect, we could have $(m - i) \geq L$. Thus, we have the following formulas:

$$\text{If } \beta \geq \alpha, \frac{m-i}{D} \geq \frac{L}{D} = \frac{\sin(180° - \beta)}{\sin(\beta - \alpha)} = \frac{\sin \beta}{\sin(\beta - \alpha)}, \tag{4}$$

$$\text{If } \beta < \alpha\ , \frac{m-i}{D} \geq \frac{L}{D} = \frac{\sin(180° - \alpha)}{\sin(\alpha - \beta)} = \frac{\sin \alpha}{\sin(\alpha - \beta)}. \tag{5}$$



(a) (b)

Fig. 4. Generating more intersections among dummy trajectories.

From the above derivation, we can further reduce the candidate sets of rotation point i and rotation angle β for dummy $d2$. Consequently, when one dummy trajectory is added to a set of trajectories by rotation generation, we will select one existing dummy trajectory and use the above formulas to determine the rotation point and the rotation angle from the solution spaces of distance derivation.

4.2.2 *K*-intersected dummy generation

In the rotation dummy generation scheme, each dummy trajectory has only one intersection point with the true user trajectory. To reduce the disclosure of a true user trajectory, the number of dummy trajectories may need to be increased to satisfy the requirements of the user privacy profile. Intuitively, when the number of intersection points increases, more possible trajectories are generated and then the query cost is also increased. Therefore, we propose the K-intersected dummy generation scheme to reduce the number of dummy trajectories but increase the number of intersected points. In this scheme, the number of intersection points between a true trajectory and a dummy trajectory is K, where K is a parameter specified by a user.

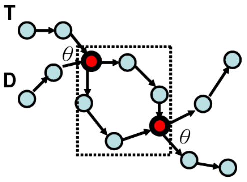


Fig. 5. An example of K-intersected pattern scheme with $K = 2$.

To understand the idea behind K-intersected dummy generation, we consider the example in Fig. 5, where K is set to 2. The bold-circle points are the intersection points that selected from a true user trajectory T at first. Obviously, some considerations should be taken into account when determining these intersection points. For example, the intersection points should not be those sensitive or important places, which may easily disclose user privacy [3]. The selection of intersection points is not the main theme of this paper so that it is beyond the scope of our discussion. Next, a dummy trajectory D is generated by composing two sets of sub-dummy trajectories. One is the sub-dummy trajectory between K intersection points, *i.e.* the sub-dummy trajectory within the dotted square. The other is the set of sub-dummy trajectories which do not contain intersection points, *i.e.* the sub-dummy trajectories which are not included in the dotted square. In order to generate the sub-dummy trajectory between K intersection points, we borrow the concept of random dummy generation in that the starting point and the end point are those consecutive intersection points and then the trajectory is obtained by random movements from the starting point to the end point. For the generation of sub-dummy trajectories which do not contain intersection points, we explore rotation dummy generation. Denote the intersection points as $C = C_1, C_2, \ldots, C_k$, where C_i is the ith intersection point. The intersection points C_1 and C_k are defined as rotation points. Similar to rotation dummy generation, we have to determine the solution space of distance derivation first. Donate that dst_k is the distance deviation among intersection points.

Suppose that a is the length of a trajectory between the intersection points C_1 and C_k and dst^r is the solution space of the distance derivation for the dummy trajectories which do not contain intersection points. Since distance deviation of a dummy trajectory should fulfill the requirement (*i.e.*, *dst*) defined in the user privacy profile, we have the following derivation:

$$\frac{(m-a)\times dst^r + a\times dst_k}{m} \geq dst$$
$$\Rightarrow dst^r \geq \frac{m\times dst - a\times dst_k}{m-a}. \tag{6}$$

After deriving the solution space for distance derivation, one can select the rotation angle which satisfies the requirements of the user privacy profile. Note that if the discourses do not fit the requirements of the user privacy profile, one would need to add one more dummy by repeating the above procedure.

5. PERFORMANCE STUDY

In this section, extensive experiments are conducted to evaluate the performance of our proposed schemes. We first describe the simulation model and then show the experimental results.

5.1 Simulation Model

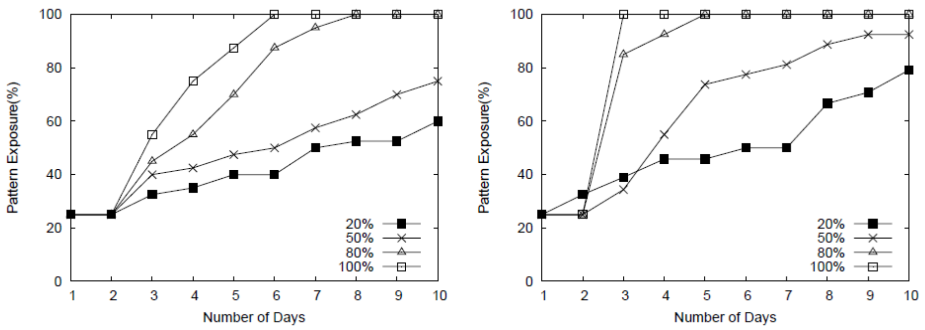
We use both real and synthetic datasets to evaluate our proposed schemes. The real dataset is derived from the INFATI project [8], a set of trajectories that are collected from a set of cars equipped with GPS receivers in the city of Aalborg. In order to investigate our proposed schemes, we also use synthetic datasets generated by a trajectory generator. Specifically, the space consists of 50 × 50 grid cells whose size is 10m × 10m in our simulation. To emphasize the privacy threat of long-term observation, we implement the prior work in [10, 11], called the dummy scheme. Suppose that adversaries are able to collect the query log in which the movements of dummies and true users are recorded. Adversaries may employ data mining techniques [16] to discover the movement patterns of users.

5.2 Experimental Result

In section 5.2.1, we verify that long-term disclosure is a location privacy threat, and investigate the impact of long-term disclosure. Sensitive analysis of generating dummy trajectories is conducted in section 5.2.2.

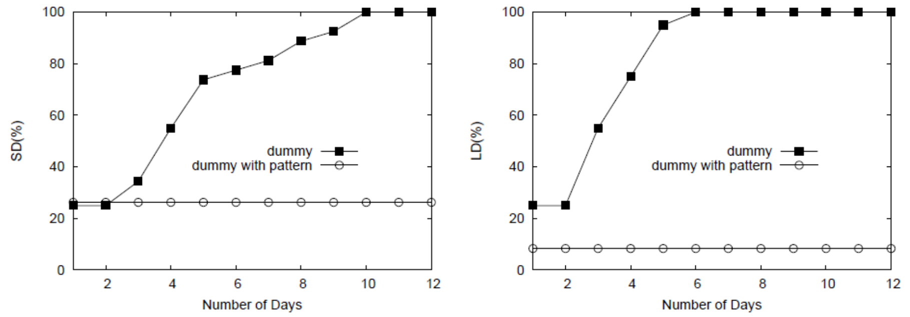
5.2.1 Impact of long-term disclosure

We first verify our claim that long-term disclosure is a location privacy threat for dummy techniques. The fake user locations generated by the dummy scheme and our proposed scheme (*i.e.*, the rotation dummy scheme) are mixed with true user locations into requests. The query frequency is varied from 20% to 100%, where the query frequency is the inverse of the query time interval. The sequential pattern mining technique [1] is implemented to discover user trajectory patterns. As seen in Fig. 6, the experiment results show two important observations. First, the pattern exposure grows when the query frequency increases. Second, as the number of collected query logs increases with time (the number of days), the pattern exposure grows. These observations are attributed to the fact that adversaries can collect a sufficient number of query logs by monitoring long-term movements, and then user movement patterns can be discovered even if the location information includes dummies.



(a) INFATI dataset. (b) Synthetic data.

Fig. 6. Pattern exposure with various query frequencies.



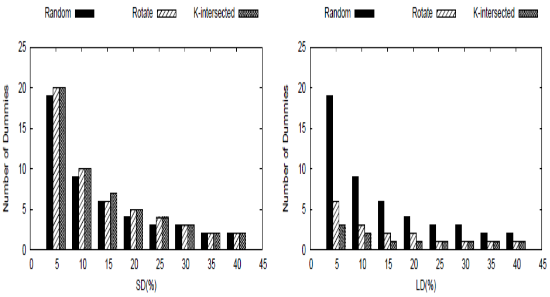
(a) Short term disclosure. (b) Long term disclosure.

Fig. 7. The performance comparison of dummy without movement patterns and dummy with movement patterns.

We next investigate the impact of movement. Suppose a privacy profile is set to *SD* = 20%, *LD* = 10%; and *dst* = 2.8. We compare our rotation pattern scheme with the dummy scheme. Fig. 7 shows the experimental result that when the amount of collected data increases with the time, both the *SD* and *LD* of the dummy scheme increase. This agrees with our claim that a long-term privacy threat exists if dummies do not follow long-term, consistent movement patterns. Once a sufficient amount of data is collected, the true user trajectory is exposed by using the existing dummy method, resulting in 100% disclosure in terms of *SD* and *LD*. On the contrary, our scheme remains within the specified disclosures (*i.e.*, *SD* = 20%, *LD* = 10%) and indicates that generating dummies with patterns could prevent both short-term and long-term location privacy.

5.2.2 Sensitivity analysis of dummy generation schemes

Next, the performance of our proposed schemes, that is, the random pattern, rotation pattern and *K*-intersected pattern schemes (denoted as Random, Rotate and *K*-intersected, respectively) are compared. As mentioned earlier, when the privacy requirements are not satisfied, additional dummies are included. However, a large number of dummies increases the query message length, leading to a considerable cost in communication and client processing. Thus, one should use as few dummies as possible to satisfy the user privacy profiles. The performance of the Random, Rotate and *K*-intersected with the value of *SD* varied is shown in Fig. 8 (a), where *LD* = 50%, *dst* = 2.8 and *K* is set to 2. Since *SD* is related to short-term disclosure, these schemes use almost the same number of dummies to fulfill the requirement of *SD*. Furthermore, the experiment of varying *LD* is conducted while *SD* = 50% and *dst* = 2.8. It can be seen in Fig. 8 (b) that both Rotate and *K*-intersected use a smaller number of dummies to meet the *LD* requirement than Random does. By intersecting trajectories, Rotate and *K*-intersected are able to increase the number of possible trajectories. Hence, Rotate and *K*-intersected only need a smaller number of dummies than Random does to meet the privacy requirements.



(a) Short term disclosure. (b) Long term disclosure.

Fig. 8. The performance of schemes random, rotate and *K*-intersected.

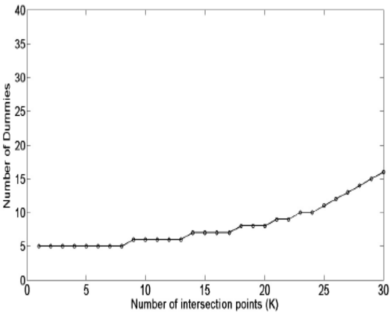
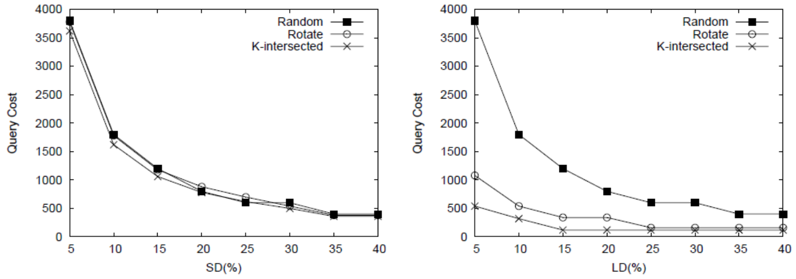


Fig. 9. The performance of *K*-intersected with various *K*.

We also investigate how the performance of the *K*-intersected scheme is affected by various *K*. Given a privacy profile (*SD* = 20%, *LD* = 10% and *dst* = 2.8), the performance of the *K*-intersected scheme varied with *K* as shown in Fig. 9. Although lower disclosures are expected when more intersections exist among trajectories, the number of dummies still grows slowly as the number of intersections (*K*) increases. More intersections among trajectories may enumerate more possible trajectories and result in lower *LD*. On the other hand, the more intersections among trajectories may reduce the number of dummy locations at a certain time slot and result in higher *SD*. Thus, we have to add more dummies to satisfy the *SD* requirement while the number of intersections (*K*) increases.



(a) Short term disclosure with cache. (b) Long term disclosure with cache.

Fig. 10. The query cost of scheme Random, Rotation and *K*-intersected.

To highlight that the benefit of increasing the number of intersection points is able to reduce the query cost, we conduct the query cost evaluation for Random, Rotate and *K*-intersected schemes. Assume that the query result is set to 10. The query costs of these proposed schemes are shown in Figs. 10 (a) and (b). In Fig. 10 (a), the query cost decreases as the value of *SD* increases. With a larger *SD*, the number of dummies is increasing. Consequently, caching more query results can be used in the future. As mentioned before, given a *SD*, the number of dummies generated from these three schemes is almost the same. Thus, the query costs of these three schemes are very close. Fig. 10 (b) shows the query costs of the three schemes with the value of *LD* varied. As can be seen in Fig. 10 (b), the query cost also decreases as *LD* increases. This is because more dummies are required to satisfy a larger value of *LD*. Moreover, for the same value of *LD*, both Rotate and *K*-intersected performs better than Random. The reason is that more intersection points are able to increase the cache utilization, which significantly reduces the query cost.

6. CONCLUSION

We observed that existing works using dummies to protect location privacy are still exposed to privacy threats in the long run. By exploring data mining techniques, adversaries may be able to discover user movement patterns and then invade user location privacy. To deal with this problem, we proposed several schemes to derive dummy trajectories, namely, the random scheme and the intersection schemes. Specifically, the random pattern scheme randomly generates dummies with consistent movement patterns, while the rotation pattern scheme and *K*-intersected dummy generation scheme explore the idea of creating intersections among trajectories. The performance analysis shows that by generating dummies with movement patterns, our proposal outperforms existing dummy-based schemes for protecting the trajectories and locations of mobile users.

REFERENCES

1. R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the 11th IEEE International Conference on Data Engineering*, 1995, pp. 3-14.

2. B. Bamba, L. Liu, P. Pesti, and T. Wang, "Supporting anonymous location queries in mobile environments with privacy grid," in *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 237-246.
3. R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving user location privacy in mobile data management infrastructures," in *Proceedings of the 6th International Workshop on Privacy Enhancing Technologies*, 2006, pp. 393-412.
4. B. Gedik and L. Liu, "A customizable *k*-anonymity model for protecting location privacy," in *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems*, 2005, pp. 620-629.
5. B. Gedik and L. Liu, "Location privacy in mobile systems: A personalized anonymization model," in *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems*, 2005, pp. 620-629.
6. G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K. L. Tan, "Private queries in location based services: Anonymizers are not necessary," in *Proceedings of ACM SIGMOD International Conference on Management of Data*, 2008, pp. 121-132.
7. G. Ghinita, P. Kalnis, and S. Skiadopoulos, "PRIVE: Anonymous location-based queries in distributed mobile systems," in *Proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 371-380.
8. C. Jensen, H. Lahrmann, S. Pakalnis, and J. Runge, "The INFATI data," Technical Report No. 79, Department of Computer Science, Aalborg University, 2004.
9. A. Khoshgozaran and C. Shahabi, "Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy," in *Proceedings of the 10th International Conference on Advances in Spatial and Temporal Databases*, 2007, pp. 239-257.
10. H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proceedings of the 2nd IEEE International Conference on Pervasive Services*, 2005, pp. 88-97.
11. H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Proceedings of the 21st International Conference on Data Engineering Workshops*, 2005, pp. 1248.
12. H. Lu, C. S. Jensen, and M. L. Yiu, "PAD: Privacy-area aware, dummy-based location privacy in mobile services," in *Proceedings of the 7th ACM International Workshop on Data Engineering for Wireless and Mobile Access*, 2008, pp. 16-23.
13. S. Mascetti, C. Bettini, X. S. Wang, D. Freni, and S. Jajodia, "ProvidentHider: An algorithm to preserve historical *k*-anonymity in LBS," in *Proceedings of the 10th International Conference on Mobile Data Management*, 2009, pp. 443-448.
14. M. F. Mokbel, C. Y. Chow, and W. G. Aref, "The new casper: Query processing for location services without compromising privacy," in *Proceedings of the 32nd International Conference on Very Large Data Bases*, 2006, pp. 763-774.
15. W. C. Peng and M. S. Chen, "Developing data allocation schemes by incremental mining of user moving patterns in a mobile computing system," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, 2003, pp. 70-85.
16. W. C. Peng, Y. Z. Ko, and W. C. Lee, "On mining moving patterns for object tracking sensor networks," in *Proceedings of the 7th International Conference on Mobile Data Management*, 2006, pp. 41-44.
17. T. Xu and Y. Cai, "Exploring historical location data for anonymity preservation in

location-based services," in *Proceeding of the 27th Conference on Computer Communications*, 2008, pp. 547-555.

18. G. Zhong and U. Hengartner, "A distributed *k*-anonymity protocol for location privacy," in *Proceedings of the 7th Annual IEEE International Conference on Pervasive Computing and Communications*, 2009, pp. 1-10.

Po-Ruey Lei (雷伯瑞) received the M.S. degree in Electrical Engineering from University of Southern California, USA, in 2000. He is currently pursuing the Ph.D. degree in the Department of Electrical and Electronic Engineering, National Defense University, Taiwan. His research interests include trajectory pattern mining and data management for sensor networks.

Wen-Chih Peng (彭文志) received the B.S. and M.S. degrees from the National Chiao Tung University, Taiwan, in 1995 and 1997, respectively, and the Ph.D. degree in Electrical Engineering from the National Taiwan University, Taiwan, in 2001. Currently, he is an Associate Professor at the department of Computer Science, National Chiao Tung University, Taiwan. Prior to joining the department of Computer Science and Information Engineering, National Chiao Tung University, he was mainly involved in the projects related to mobile computing, data broadcasting and network data management. Dr. Peng serves as PC members in several prestigious conferences, such as IEEE International Conference on Data Engineering (ICDE), Pacific Asia Knowledge Discovering and Mining (PAKDD) and Mobile Data Management (MDM). His research interests include mobile computing, network data management and data mining. He is a member of IEEE.

Ing-Jiunn Su (蘇英俊) was born in Chiayi, Taiwan, in 1962. He received the B.S. Degree in Communication Engineering from National Chiao Tung University, Hsinchu, and the M.S. and Ph.D. Degrees in Electrical Engineering from National Taiwan University, Taipei, Taiwan, in 1985, 1987, and 1998, respectively. He joined the Industrial Technology Research Institute in 1989, where he worked in communication networks. Currently he is an Associate Professor in the Department of Electrical and Electronic Engineering, Chung Cheng Institute of Technology, National Defense University, Taoyuan, Taiwan. His research interests include in wireless communications, communication signal processing, and spread spectrum techniques.

Chien-Ping Chang (張劍平) received the B.S. degree in Electrical Engineering from Chung Cheng Institute of Technology in 1986, and the Ph.D. degree in Computer and Information Science from National Chiao Tung University, Taiwan, in 1998. He is currently a professor in the Department of Computer Science and Information Engineering, Ching Yun University, Taiwan. His research interests include parallel computing, interconnection networks, graph theory, image processing, and data hiding.