Protecting Moving Trajectories with Dummies

Tun-Hao You Wen-Chih Peng
National Chiao Tung University
Hsinchu, Taiwan, ROC
{thyou, wcpeng}@cs.nctu.edu.tw

Wang-Chien Lee
The Pennsylvania State University
State College, PA 16801, USA
wlee@cse.psu.edu

***Abstract*—Dummy-based anonymization techniques for protecting location privacy of mobile users have been proposed in the literature. By generating dummies that move in human-like trajectories, [8] shows that location privacy of mobile users can be preserved. However, by monitoring long-term movement patterns of users, the trajectories of mobile users can still be exposed. We argue that, once the trajectory of a user is identified, locations of the user is exposed. Thus, it's critical to protect the moving trajectories of mobile users in order to preserve user location privacy. We propose two schemes that generate consistent movement patterns in a long run. Guided by three parameters in user specified privacy profile, namely, *short-term disclosure*, *long-term disclosure* and *distance deviation*, the proposed schemes derive movement trajectories for dummies. A preliminary performance study shows that our approach is more effective than existing work in protecting moving trajectories of mobile users and their location privacy.**

***Keywords* —Location privacy, user movement patterns, location-based services.**

I. Introduction

Location-based services (LBSs) have emerged as one of the killer applications for mobile computing and wireless data services. These LBSs are critical to public safety, transportation, emergency response, and disaster management, while providing great market values to companies and industries. Due to the unrestricted mobility of users in the mobile computing environments, users are often interested in acquiring information or services related to their current locations. Thus, very frequently locations information of users are submitted along with queries to the LBS servers. Examples of such queries include finding the nearest restaurants to a user (k nearest neighbor query) and finding ATMs within 500 meters from a user's current location (range query). While LBSs have shown to be valuable to users' daily life, on the other hand, they also expose extraordinary threats to user privacy. If not well protected, the location information of users may be misused by some untrustworthy service providers or stolen by hackers. Once the location information is exposed, adversaries may dig for cues to invades user privacy. Obviously, it is important to protect location privacy.

Recently, the problem of location privacy preserving has received growing interests from the research community [1, 2, 7, 8, 9]. These studies aim at protecting exact location information of users from the potential abuse of LBS providers and hackers. Two primary approaches have been considered, including 1) *trusted anonymizer* based approach; and 2) *client* based approach. In the former, users submit their queries to the LBSs via a trusted server (which is different from the LBS server), such as a base station in the cellular networks. This trusted anonymizer transforms the exact locations of a number of users into a *cloaked spatial area* in accordance with privacy requirements set by users in order to obtain data or services from the LBSs. 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 its true location to the LBSs, clients may obtain redundant information or services corresponding to the submitted locations [8]. Unwanted information is then filtered locally to obtain the final answers. In both approaches, the true location of a user is either 1) not distinguishable from 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 some issues faced in the client based approach.

Motivation and Problems. Without relying on a trusted server, generating fake user locations (called *dummies*[1]) for location-dependent queries has been shown to be an effective way to preserve location privacy [8]. In addition to generate dummies based on the user locations, this prior work proposes to generate dummies based on realistic user movements. However, it does not consider a well-recognized observation, i.e., moving behaviors of users usually follow certain patterns [10, 11]. Thus, adversaries may be able to employ data mining techniques to discover movement patterns of users and distinguish trajectories of true users from the dummies. Fig. 1 shows some examples that illustrate our discussions. In the figure, the solid line denotes the moving trajectory of a true user (denoted as T) and the dotted lines are generated trajectories of dummies (denoted as $d1$ and $d2$). Since true users usually exhibit certain human moving behavior, one is able to identify the solid line as a true user based on the typical moving behavior of humans (as shown in Fig. 1(a)). Thus, it's important to generate dummy trajectories based on human moving 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 be collected to filter inconsistent trajectories. For example, comparing the current trajectories (in Fig. 1(c)) and trajectories

[1]We follow the terminology used in [8] to name the fake user locations as *dummy locations* and *dummies* in short.

collected in a different day (e.g., Fig. 1(b)), one can tell T is the true trajectory of user. Once the moving trajectory of true user is identified, locations (i.e., not only the current location but also the past locations) of the user is disclosed. Thus, it's important to generate dummies that not only demonstrate moving behavior of users but also follow consistent, long-term movement patterns.

Given that the adversaries obtain a set of trajectories, they will have difficulty determining the true trajectory of a user if users generate dummies 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 the adversaries. For example, in Fig. 1(c), three trajectories are collected and thus the disclosure is $\frac{1}{3}$. To reduce the disclosure, a naive approach is to simply increase the number of dummies, which however incurs overhead in terms of query message length and thus communication and client processing costs. Thus, in this paper, we propose to generate *intersecting dummy trajectories* aiming at increasing the number of possible trajectories from the adversaries' perspective and thus decreasing disclosure of the user trajectory.

Nevertheless, an issue exists with this intersecting trajectories. When the generated trajectories are too close to the true trajectory, the locations of a user may still be exposed, e.g., Fig. 1(d) shows an example where the user's moving trajectory (the shadowed path) can be identified. Thus, our design of dummy generation schemes also take the factor of *distance 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 2). We propose two schemes, namely, *random pattern* and *rotation pattern*, to generate dummy trajectories based on the privacy profile. A preliminary performance study shows that by generating dummies based on moving patterns, our schemes perform better than the existing techniques.

Related Work. A significant amount of research efforts have been put forth on location privacy. Generally speaking, one could protect either user identification or location to guarantee location privacy. Specifically, the authors in [1, 2] propose mixed zone to protect user identification. A number of research works are performed to protect location information of users [7, 8, 9]. With trusted servers, the authors in [4, 5, 6] propose a cloaking algorithm to blur the resolution of location information in spatial and temporal dimensions. Based on k-anonymity, the authors in [3] devises a personalized and customized k-anonymity model. Without trusted servers, the authors in [8] propose an algorithm to generate dummy movement similar to true user movements. While our work is also based on dummies, we address two new issues, i.e., long-term location privacy and protection of user trajectories.

Organization. The rest of this paper is organized as follows. Section 2 and Section 3 present preliminaries and our proposal of dummy trajectory generation schemes, respectively. Section 4 shows our performance study. Section 5 concludes this paper.

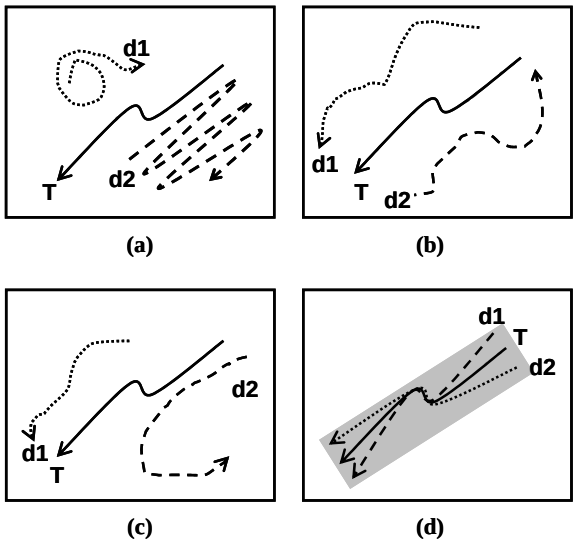

Fig. 1. Moving Trajectories of User and Dummies.

II. Preliminaries

We assume no trusted server available for location anonymization. Wireless networks are only responsible for communication and will not reveal locations of mobile users. Mobile clients are location-aware (via GPS or network based positioning techniques). To simplify our discussions, we assume users are free to move in the space, which is divided into grid cells. Each grid cell has a cell identifier (x, y) indicating that this cell is located at the x column and the y row of the space. A query message issued by a mobile user U_i to a LBS server is defined as $M = \{uid, \langle L_i^t, L_{d1}^t, L_{d2}^t ... L_{dn}^t \rangle\}$, where uid is a user identifier and L_i^t) is the true user location and $L_{d1}^t, ..., L_{dn}^t$ are n dummy locations, respectively, at time slot t. Therefore, given m consecutive queries, the trajectory of U_i is $\{L_i^1, L_i^2, ..., L_i^m\}$, while the trajectory of dummy x is $\{L_{dx}^1, L_{dx}^2, ..., L_{dx}^m\}$). Here, L_i^j (and L_{dx}^m, 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 $\mathrm{P}_i = \{PL_{i,}^1, PL_i^2, ..., PL_i^m\}$, where PL_i^j is the location of mobile user U_i at the jth time slot (note that the length of trajectories is m).

Users may set up their privacy profile, which is specified by the following three parameters:

1) **Short-term Disclosure (SD)**: This parameter specifies 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 requirement for protecting the user trajectory. Given n trajectories, among which k trajectories have intersected with other trajectories and $(n - k)$ trajectories do not

Time slot	1	2	3	4	5	6
True user	(7,3)	(6,3)	(5,2)	(4,2)	(3,2)	(2,2)
Dummy X	(6,1)	(6,2)	(5,2)	(5,3)	(6,4)	(7,4)
Dummy Y	(1,4)	(2,4)	(3,4)	(3,3)	(3,2)	(3,1)
$\|D_i\|$	3	3	2	3	2	3
Distance deviation	4.2	2.6	1.4	1.4	1.8	3.4

TABLE I
PRIVACY MEASUREMENT OF DUMMY TRAJECTORIES.

have any intersection. 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. Here, 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 distance difference among trajectories of dummies and the user. As a result, dst of mobile user U_i is formulated as $\frac{1}{m}*\frac{1}{n}*\sum_{k=1}^{n}\sum_{j=1}^{m} dist(PL_{i,}^{j} L_{dk}^{j})$, where $dist$ is distance between the true user location and dummy locations in unit of cell size.

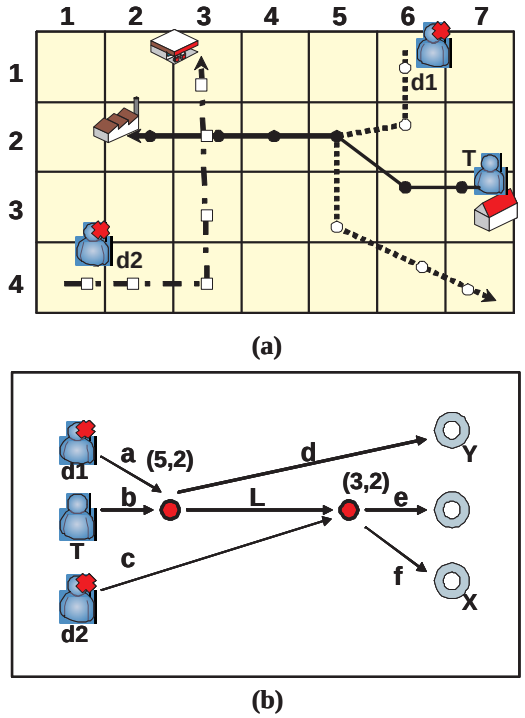

Fig. 2. Trajectories with Intersections.

Fig. 2 shows an example of generated dummy trajectories with intersections, while Table I shows the trajectories as well as the number of current locations (D_i) and distance deviation at different time slots. Thus, we can derive $SD = \frac{1}{6}(\frac{1}{3} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3}) = \frac{7}{18}$. Furthermore, for each time slot, we could derive distance differences between dummy and true trajectories to obtain the average distance deviation as 2.47. To facilitate the derivation of total possible trajectories, Fig. 2(a) is transformed into Fig. 2(b), where intersection points (i.e., cell (5, 2) and (3, 2)) are marked. Since these three trajectories have two intersection points, it can be verified that we have 8 possible trajectories (i.e., ad, aLe, aLf, bd, bLe, bLf, ce, cf). As such, we could have long-term disclosure $LD = \frac{1}{8+(3-3)} = \frac{1}{8}$.

III. GENERATING DUMMIES WITH PATTERNS

Given a privacy profile, our goal is to generate dummy trajectories that satisfy the user privacy profile. In this Section, we propose two schemes, namely, *random* and *rotation* pattern schemes, to generate dummies that exhibit long-term user movement patterns.

A. Random Pattern Scheme

In this scheme, the starting point and the destination of a dummy are first selected. Then, the grid cells between the starting point and the destination are determined based on the speed of a dummy and three movement types, including horizontal movement, vertical movement, and both. In this scheme, a dummy will move randomly from the starting point towards the destination. This naive scheme demonstrates that even after a long term observation, it's difficult for adversaries to identify true user since dummies also exhibit long-term, consistent movement patterns. However, without taking into account factors such as distance deviation, this scheme simply include more dummies when the privacy requirements are not satisfied,

B. Rotation Pattern Scheme

The main idea behind this scheme is to have some intersections between trajectories of dummies and the user. Given a user trajectory, we generate a new trajectory for a dummy by rotating the known user trajectory. Clearly, the rotation point of user trajectory is an intersection point. 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 is to first derive 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 smaller than what specified are selected as dummy trajectories. With 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 deviation (i.e., dst), we consider both of the *rotation angle* and the *rotation point* within a true user trajectory, which have a great impact on the distance deviation. To simplify the derivation of distance deviation, assume that we have a true user trajectory in Fig. 3(a), where the distance between two consecutive movements is d, the rotation point is the location

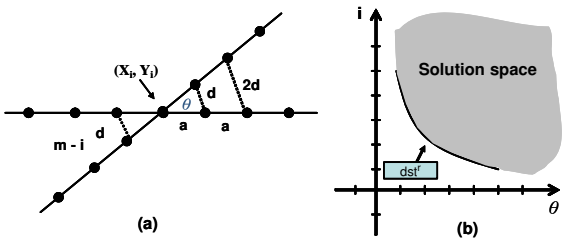

Fig. 3. Solution space for distance deviation.

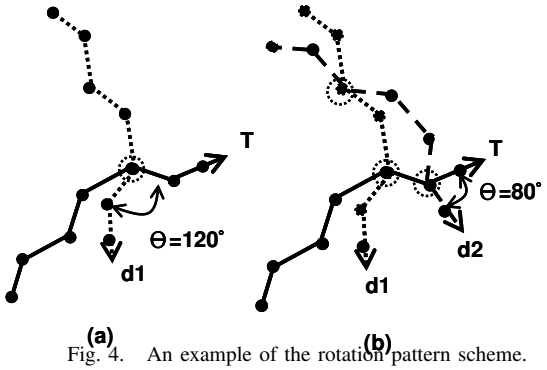

Fig. 4. An example of the rotation pattern scheme.

at the ith time slot in a true user trajectory, denoted as (X_i, Y_i), 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 $d = \sqrt{2}|a|\sqrt{1-\cos\theta}$. Hence, we could derive the distance deviation of these two trajectories as follows:

$$\begin{aligned} dst^r &= \frac{1}{m}((d+2d+...+id)+(d+2d+...+(m-i)d)) \\ &= \sqrt{2}|a|\sqrt{1-\cos\theta} * (\sum_{j=0}^{i} j + \sum_{j=0}^{m-i} j) \end{aligned}$$

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 the distance deviation. Assume that we have n dummies 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 distance deviation (i.e., dst). Thus, we have the following formula:

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

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

θ	i	SD	LD
120	5	56.25%	25%*
50	3	56.25%	25%
180	1	56.25%	25%

(a). Solution space when n=0

θ	i	SD	LD
170	8	37.5%	16.67%
120	7	37.5%	12.5%
80	6	39.6%	8.33%*

(b). Solution space when n=1

TABLE II
SOLUTION SPACES.

For example, consider a true trajectory (the line marked with T) in Fig. 4(a) and a user privacy profile (i.e., $SD = 40\%$, $LD = 10\%$, $dst = 2.1$). Initially, there is no dummy (i.e., $n = 0$) and $dst_0 = 0$. As such, we could have $dst^r \geq (0+1) * 2.1$. Table II(a) show some selected possible solution space when the number of dummy is 0. In Table II(a), the solution (i.e., $(120^{\mathbf{o}}, 5)$) is selected and then n is increased to 1. The value of dst_1 is updated accordingly. However, since disclosures are still larger than the required values (i.e., $56.25\% \geq 40\%$ and $25\% \geq 10\%$), we should add one more dummy to reduce the disclosures. Following the same procedure, we have $dst^r \geq (1+1) * 2.1 - 1 * 2.8$ and Table II(b) is the solution space when the number of dummy is one. From Table II(b), one could select $(80^{\mathbf{o}}, 6)$ since the corresponding disclosures is smaller than the required values. Hence, Fig. 4(b) shows the final dummy trajectories.

IV. PERFORMANCE STUDY

In this section, we evaluate the performance of our proposed schemes. We first describe the simulation model and then show the experimental results.

A. Simulation Model

In our simulation, the space is divided into 50*50 grid cells. Assume that the number of time slots is 20 and that there exists a moving pattern for each user, i.e., the pattern has a starting point and a destination point. Then, those grid cells between the starting point and the destination are selected based on the nature of movements, i.e., the next move is a neighboring cell of the current location. Three movement types are the horizontal movement, the vertical movement, and both. To emphasize the privacy threat of long-term observation, we implement the prior work in [8] (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 explore data mining techniques [11] to discover movement patterns of users.

B. Experimental Results

We first investigate the impact of movement patterns. Suppose a privacy profile is set to $SD = 20\%, LD = 10\%$,

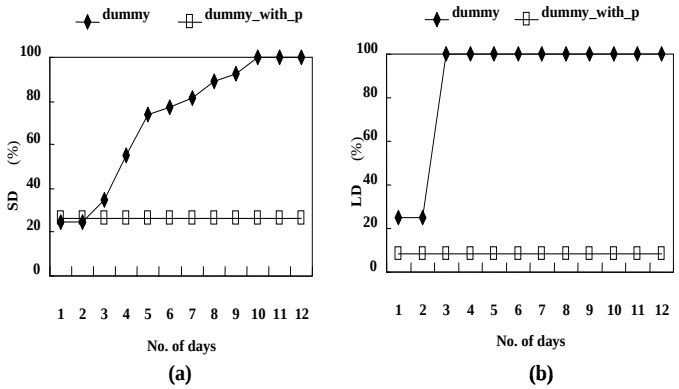

Fig. 5. Comparison of Dummy-based Schemes.

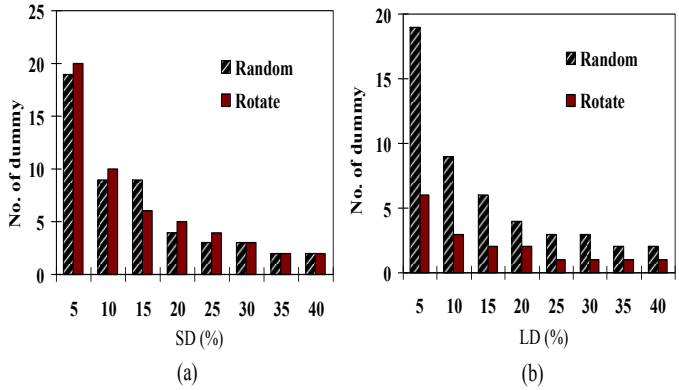

Fig. 6. Comparison of Random and Rotate Pattern Schemes

and $dst = 2.8$. We compare our rotation pattern scheme with the dummy scheme. Fig. 5 shows the experimental result, where dummy (respectively, dummy_with_p) refers the scheme of generating dummies without (respectively, with) patterns. In Fig.5(a), it can be seen that when the amount of data collected increases with the time, both SD and LD of the dummy scheme increase. This agrees with our claim that long-term privacy threat exists if dummies do not follow long-term, consistent movement patterns. Once collected a sufficient amount of data, the true user trajectory is exposed (i.e., resulting in 100% disclosures in term of SD and LD). On the other hand, our scheme remains within the specified disclosures (i.e., $SD = 20\%, LD = 10\%$), showing that generating dummies with patterns could preserve both short-term and long-term location privacy.

Next, the performance of the proposed random pattern and rotation pattern schemes (denoted as Random and Rotate, respectively) are compared. As mentioned earlier, when the privacy requirements are not satisfied, additional dummies are included. However, a larger number of dummies increases query message length, leading to a considerable cost in communication and client processing. Thus, one should use as few dummies as possible to satisfy user privacy requirements. The performance of Random and Rotate with the value of SD varied is shown in Fig. 6(a), where $LD = 50\%$ and $dst = 2.8$. Since SD is related to short-term disclosure, both scheme Random and Rotate use almost the same number of dummies to meet the requirement of SD. Furthermore, an experiment that varies LD is conducted with $SD = 50\%$ and $dst = 2.8$. Fig. 6(b) shows the experimental result. It can be seen that Rotate uses a smaller number of dummies than Random. By intersecting trajectories, Rotate is able to increase the number of possible trajectories. Hence, Rotate only needs a smaller number of dummies than Random to meet the privacy requirement.

V. Conclusion

We observed that existing works using dummies to protect location privacy are still exposed to privacy threat in a long run. By exploring data mining techniques, adversaries may be able to determine user movement patterns, thereby invading user location privacy. To deal with this problem, we proposed two schemes to derive dummy trajectories. Specifically, random pattern scheme randomly generates dummies with consistent movement patterns, while the rotation pattern explores the idea of creating intersections among moving trajectories. Our preliminary performance study shows that by generating dummies with movement patterns, our proposal outperforms the existing dummy-based scheme for protecting trajectory and locations of mobile users.

References

[1] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[2] David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–90, 1981.

[3] Bugra Gedik and Ling Liu. Location privacy in mobile systems: A personalized anonymization model. In *Proc. of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS)*, pages 620–629, 2005.

[4] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proc. of the First International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 31–42, 2003.

[5] Marco Gruteser and Dirk Grunwald. A methodological assessment of location privacy risks in wireless hotspot networks. In *Proc. of the First International Conference on Security in Pervasive Computing (SPC)*, volume 2802, pages 10–24, 2003.

[6] Marco Gruteser and Dirk Grunwald. Enhancing location privacy in wireless lan through disposable interface identifiers: A quantitative analysis. *ACM Mobile Networks and Applications (MONET)*, 10(3):315–325, 2005.

[7] Jason I. Hong and James A. Landay. An Architecture for Privacy-Sensitive Ubiquitous Computing. In *Proc. of the Second International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 177–189, 2004.

[8] Hidetoshi Kido, Yutaka Yanagisawa, and Tetsuji Satoh. An Anonymous Communication Technique using Dummies for Location-based Services. In *Proc. of the Second International Conference on Pervasive Services (ICPS)*, pages 88–97, 2005.

[9] Mohamed F. Mokbel, Chi-Yin Chow, and Walid G. Aref. The New Casper: Query Processing for Location Services without Compromising Privacy. In *Proc. of the 32nd International Conference on Very Large Data Bases (VLDB)*, pages 763 – 774, 2006.

[10] Wen-Chih Peng and Ming-Syan Chen. Developing data allocation schemes by incremental mining of user moving patterns in a mobile computing system. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 15(1):70–85, 2003.

[11] Wen-Chih Peng, Yu-Zen Ko, and Wang-Chien Lee. On mining moving patterns for object tracking sensor networks. In *Proc. of the 7th International Conference on Mobile Data Management (MDM)*, pages 41–44, 2006.