# De-anonymizable Location Cloaking for Privacy-controlled Mobile Systems

Chao Li, Balaji Palanisamy

University of Pittsburgh
{chl205,bpalan}@pitt.edu

**Abstract.** The rapid technology upgrades of mobile devices and the popularity of wireless networks significantly drive the emergence and development of Location-based Services (LBSs), thus greatly expanding the business of online services and enriching the user experience. However, the personal location data shared with the service providers also leave hidden risks on location privacy. Location anonymization techniques transform the exact location of a user into a cloaking area by including the locations of multiple users in the exposed area such that the exposed location is indistinguishable from that of the other users. However in such schemes, location information once perturbed cannot be recovered from the cloaking region and as a result, users of the location cannot obtain fine granular information even when they have access to it. In this paper, we propose Dynamic Reversible Cloaking (DRC) a new de-anonymziable location cloaking mechanism that allows to restore the actual location from the perturbed information through the use of an anonymization key. Extensive experiments using realistic road network traces show that the proposed scheme is efficient, effective and scalable.

## 1 Introduction

With the popularity of mobile positioning devices and the wide emergence of location-based services (LBSs), we are witnessing a rapid development of mobile location-based applications. Through location-aware techniques (e.g. GPS, wireless access point), users can acquire personalized services based on their current location. Examples of such services include weather forecast, traffic condition updates, location-based travel services and emergency care. The richness and diversity of location-based services has dramatically improved the quality of life for people. However, these services and benefits also come with a hidden cost: the intrusion of location privacy. For example, knowing the haunts of users during the day time and at nights, an attacker can infer a user's social activities, religious beliefs and political views. Moreover, with the advent of big data and big data analytics, the risk of disclosing location information is further exacerbated as an adversary can correlate the exposed location with information from various data sources to infer more accurate and fine-grained information about individuals.

Various techniques have been proposed to achieve location privacy protection in mobile location-based system. Location anonymization refers to the process of transforming the exact location of a user to a cloaking area by including the location of multiple other users. A user is considered to be location k-anonymous if and only if her location information is indistinguishable from that of at least $k-1$ other users. However in such schemes, location information once perturbed cannot be recovered from the cloaking region and as a result, users of the location cannot obtain fine granular information even when they have access to it. In several access controlled scenarios, such as when some users of the location have more privileges than the others, it may be desirable to obtain fine granular location information from the exposed perturbed location when the data user has access to the finer granular location information.

During the last several years, many location anonymization techniques[1, 6, 8, 10, 13, 20] have been proposed. Most of them were developed as unidirectional cloaking techniques without considering the ability to de-anonymize the perturbed data. In this paper, we propose Dynamic Reversible Cloaking (DRC) a new de-anonymziable location cloaking mechanism that allows to restore the actual location from the perturbed information through the use of an anonymization key. Our proposed mechanism uses an anonymization secret key to uniquely generate a cloaking region which allows the original location data to be restored from the cloaked data using the secret key, However, without the secret key, the original data cannot be inferred even when the adversary has complete knowledge of the cloaking mechanism.

We organize the rest of the paper as follows: In Section 2, we present the required background. In Section 3, we introduce the proposed Dynamic Reversible Cloaking. In Section 4, we analyze the experiment results. We discuss related work in Section 5 and we conclude in Section 6.

## 2 Overview of Concepts and Framework

In this section, we discuss the road network model used for the cloaking process and introduce the conventional location cloaking mechanisms to protect location privacy over road networks. We then define the de-anonymizable location privacy problem and introduce the evaluation metrics used in this paper.

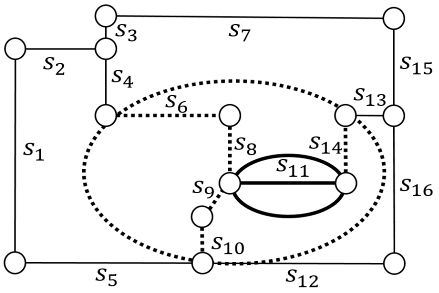


Fig. 1: Road network model with de-anonymizable privacy controlled framework

A road network can be modeled as a graph $G = (V_G, E_G)$, consisting of a set of junctions, $V_G$, and a set of road segments, $E_G$. Figure 1 shows an example with 14 junctions and 16 segments respectively. In our work, junctions are defined as the crossover or end points of roads while segments are defined as the direct roads between adjacent junctions. We assume that all mobile users move along the segments. The raw location information of the mobile users is sent with their customizable privacy requirements to a trusted third-party location anonymizer, which transforms the raw location into a cloaking region preserving the required privacy properties, which may then be shared with other untrusted location-based service providers.

The location information of a user is said to be *k-anonymous* if the location information is indistinguishable from the location information of at least k-1 other users. Obviously, the larger the value $k$, the better is the privacy protection that can be achieved. However, large cloaking regions typically provide lower utility as it increases the complexity of the operations involved in obtaining an accurate answer to the location-based query [13]. Therefore, to bound the size of the cloaking region, the customizable requirements contain not only the privacy requirement, denoted by $\delta_k$, but also the maximum spatial resolution level, denoted by $\sigma_s$ [13, 19], which defines the maximum acceptable size of the cloaking region. The customizable privacy requirements with both the two parameters are organized as the user-defined privacy profile: $(\delta_k, \sigma_s)$.

In the past, several models have been proposed for location anonymization. Random sampling is a basic location cloaking technique that picks segments from the whole graph one by one in a random manner while road-network-based expansion selects adjacent segments of the cloaking region to make the structure of cloaking region tighter [19]. However, to the best of our knowledge, in most existing location privacy-preserving mechanisms, the original exact location information once perturbed in the cloaking scheme cannot be restored to infer finer location information when users have the required access privileges. The focus of our work in this paper is developing a new de-anonymizable cloaking technique that can support privacy control in access controlled scenarios. In such cases, the location privacy of users is protected while allowing only the parties with the required permission to access finer information.

In the proposed de-anonymizable location privacy model, only the data users who possess the permission to access finer information get the secret key to de-anonymize the data. Such data users can de-anonymize the perturbed location using the secret key. A detailed example is shown in Figure 1. The segment $s_{11}$ contains the actual user. Using the secret key, $\{s_6, s_8, s_9, s_{10}, s_{14}\}$ are added to reach the privacy level $\delta_k$. To de-anonymize the cloaking region, the same secret key is used to exactly identify and remove the segments $\{s_6, s_8, s_9, s_{10}, s_{14}\}$ from the cloaking region, thus reducing the perturbed location to the actual segment of the user.

To evaluate the performance of de-anonymizable cloaking scheme, multiple metrics are required. Four metrics are used in this paper.

*Anonymization Time:* The time required to cloak the location information of the user. A shorter anonymization time indicates higher effectiveness.

*De-anonymization Time:* The time required to de-anonymize the cloaking area to get the exact location information of the user. Like anonymization time, a shorter de-anonymization time indicates higher effectiveness.

*Relative Spatial Resolution (RSR):* This metric reflects the relationship between the maximum spatial area defined by $\sigma_s$ and the cloaking area. Specifically, the maximum spatial area is a rectangular area centering on the user. Its lateral and vertical lengths are provided by $\sigma_s = \{ML_l, ML_v\}$. For the cloaking space, its area is also abstracted as a rectangle. The lateral and vertical distances between each pair of segments within the cloaking space are calculated and the largest two values, expressed as $\{CL_l, CL_v\}$, are considered as the lateral and vertical lengths of the cloaking rectangle. Therefore, considering a set of LBS requests with N elements, the RSR is defined as:

$$RSR = \frac{1}{N} \sum_{N} \sqrt{\frac{ML_l \times ML_v}{CL_l \times CL_v}}$$

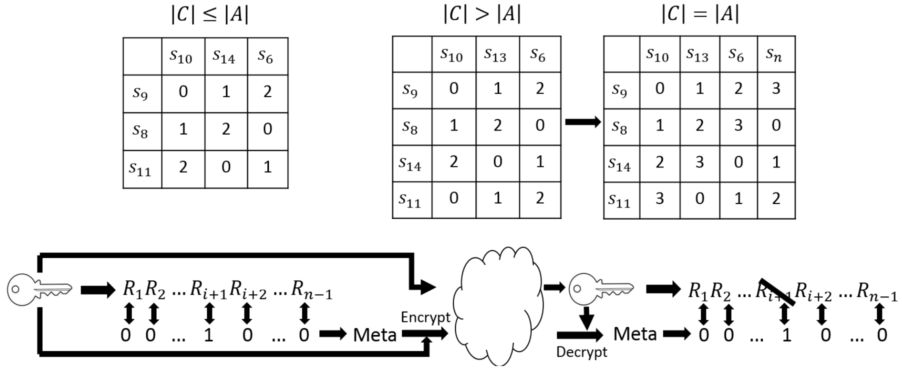
Since the RSR is the square root of the ratio between maximum spatial area and cloaking area, a higher value of RSR indicates a smaller cloaking space required to satisfy the $\delta_k$, meaning higher effectiveness.

*Success Rate:* This metric represents the rate of successful cloaking of the requests. For a set $\mathbb{Q}$ of LBS requests, the cloaking area of each query $q$ is represented by $C_q = f(q)$. A parameter $S$ is 1 if the process is successful and 0 if the process is failed. The success rate can be defined as:

$$Success\ Rate = \frac{|\{C_q \mid C_q = f(q), q \in \mathbb{Q}, S = 1\}|}{|\mathbb{Q}|}$$

## 3 De-anonymizable Location Cloaking

In this section, we present the proposed dynamic reversible cloaking algorithm that forms cloaked location regions containing tightly structured segments meeting the required $k$ anonymity level. In this scheme, the anonymization and de-anonymization processes can be considered as two inverse transition strings controlled by the secret key. Specifically, during the two processes, the exchange of segments can be seen as forward and backward transitions between two set of segments, namely the set of segments within the cloaking region and the set of adjacent segments of the cloaking region. Therefore, the two processes are sequences of continuous $n - 1$ forward and backward transitions respectively, which are the inverse of each other. Since there are multiple transition choices, a secret key is used as the transition controller. With the key, both the forward and backward transition strings are determinate and reversible. Without the key, the two transition strings become random and irreversible.



$|C| \leq |A|$

| | $s_{10}$ | $s_{14}$ | $s_6$ |
|---|---|---|---|
| $s_9$ | 0 | 1 | 2 |
| $s_8$ | 1 | 2 | 0 |
| $s_{11}$ | 2 | 0 | 1 |

$|C| > |A|$

| | $s_{10}$ | $s_{13}$ | $s_6$ |
|---|---|---|---|
| $s_9$ | 0 | 1 | 2 |
| $s_8$ | 1 | 2 | 0 |
| $s_{14}$ | 2 | 0 | 1 |
| $s_{11}$ | 0 | 1 | 2 |

$|C| = |A|$

| | $s_{10}$ | $s_{13}$ | $s_6$ | $s_n$ |
|---|---|---|---|---|
| $s_9$ | 0 | 1 | 2 | 3 |
| $s_8$ | 1 | 2 | 3 | 0 |
| $s_{14}$ | 2 | 3 | 0 | 1 |
| $s_{11}$ | 3 | 0 | 1 | 2 |

$R_1 R_2 \ldots R_{i+1} R_{i+2} \ldots R_{n-1}$
0 0 ... 1 0 ... 0
Meta
Encrypt
Decrypt
Meta
$R_1 R_2 \ldots \cancel{R_{i+1}} R_{i+2} \ldots R_{n-1}$
0 0 ... 1 0 ... 0

Fig. 2: Dynamic reversible cloaking

We describe the process with an example presented in Figure 2. For each transition, the number of cloaked segments $|C|$ and adjacent segments $|A|$ are found. For example, suppose in Figure 1, we have $C = \{s_8, s_9, s_{11}\}$ and $A = \{s_6, s_{10}, s_{14}\}$ so that

$|C| \leq |A|$. For each addition or removal step, we assign IDs to all the potential forward and backward transitions between the two sets to distinguish them. These IDs, called transition IDs (TIDs), are organized in a transition table. In Figure 2, $s_8, s_9, s_{11}$ within $C$ and $s_6, s_{10}, s_{14}$ within $A$ are mapped to the three rows and three columns respectively in the order of segment length so that the shortest segments are mapped to the $1^{st}$ row and $1^{st}$ column. The TID in table cell $(i, j)$ associated with $i^{th}$ row and $j^{th}$ column is computed by $((i-1)+(j-1))\ mod\ |A|$ to make sure no same value is generated in the same row or column. The anonymization key is used to generate a sequence of pseudo-random numbers. The $i^{th}$ pseudo-random number uniquely determines a value for both the $i^{th}$ forward transition and $\{n-i\}^{th}$ backward transition. This value, called picked TID, can be calculated by $p_i = R_i\ mod\ |A|$ and it is used to select the transition with the TID value same as the picked TID.

However, in the case $|C| > |A|$, since the number of potential backward transitions is larger than the number of available TIDs, the same TID may be assigned to multiple backward transitions, called a collision. Once a collision occurs, the key may fail to distinguish the transitions with same TID and select the backward transition unmatched with the forward transition. In general, collisions during cloaking expansion can be dealt with two approaches. By carefully managing the assignment of TIDs before the anonymization, collisions can be eliminated so that the de-anonymization process can automatically run in a collision-free manner [12]. In this paper, we adopt a different scheme based on using metadata information for collision-resolution. In this approach, the collisions are not eliminated during cloaking expansion but recorded as part of the metadata, which is then used in the de-anonymization process to resolve the collisions. Suppose in Figure 1, we get $C = \{s_8, s_9, s_{11}, s_{14}\}$ and $A = \{s_6, s_{10}, s_{13}\}$. After establishing the table, same TIDs are seen in each column, as shown in Figure 2. To handle such scenarios, in the cloaking process, additional segments, called *null segments* are added to the set $A$ to make $|C| = |A|$. The null segments are not real segments that can be found in the graph and they are conceptual segments used to deal with the collisions during transition. In this example, one null segment is set, denoted by $s_n$. By adding this null segment, we get $|C| = |A| = 4$, and therefore the transitions become free of collisions. However, if the null segment is picked as the next added segment during the anonymization process, that pseudo-random number should be skipped. The information of this skipped pseudo-random numbers is recorded as metadata, which is a binary stream matched with the pseudo-random stream. Each bit is 0 for non-skipped pseudo-random numbers and 1 for skipped ones. After anonymization, the metadata is encrypted by the secret key and shared with the key to the user of the location data. The user uses the metadata to identify and remove the skipped pseudo-random numbers to do a collision-free de-anonymization.

In the next section, we present our experimental results to evaluate the performance and effectiveness of the proposed scheme.

## 4 Experimental Evaluation

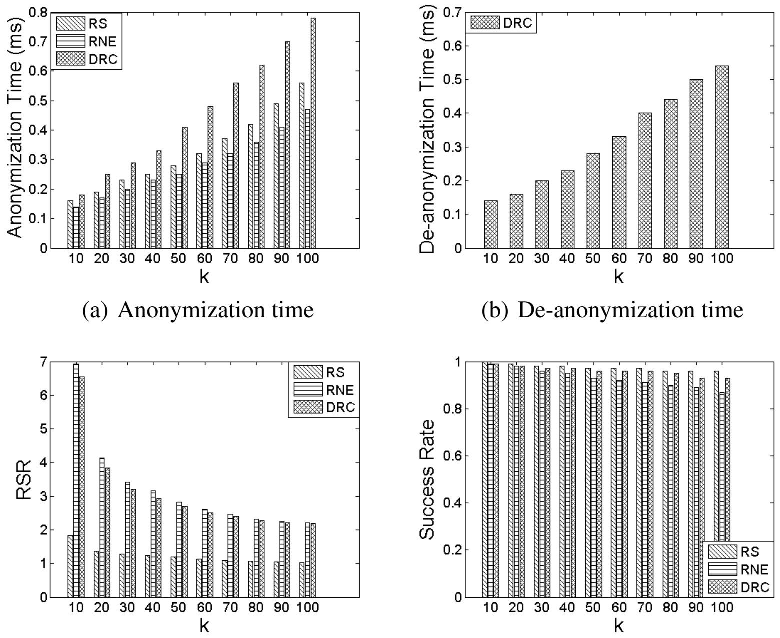
In this section, we first briefly describe the experimental setup and then present our experimental results to evaluate the dynamic reversible cloaking algorithm.

### 4.1 Experimental setup

In our experiments, we use the GTMobiSim mobile trace generator [9] to generate a realistic road network trace on the map of northwest part of Atlanta with 6979 junctions and 9187 segments. 10000 cars are randomly generated along the roads, which then move to random destinations through shortest paths on the road network. We implement three different location anonymization schemes namely random sampling (RS), road-network expansion (RNE) represented by the XStar technique in [19] and our proposed dynamic reversible cloaking (DRC).

### 4.2 Experimental results

To evaluate the three schemes, we measure the anonymization and de-anonymization time, relative spatial resolution and success rate. Our results show that the proposed dynamic reversible cloaking algorithm is effective and scalable.



(a) Anonymization time

(b) De-anonymization time

(c) Relative spatial resolution

(d) Success rate

Fig. 3: Performance with varying anonymity level

We evaluate the effectiveness of the algorithms by varying the k-anonymity level from 10 to 100. Figure 3(a) presents the results of the anonymization time. Higher anonymity level results in longer anonymization time because it requires more segments to be put into the cloaking region. Since the DRC scheme is reversible and generates a stream of pseudo-random values to drive the anonymization process, its anonymization time is higher than the other schemes. However, it continues to have only linearly increasing anonymization time even for higher anonymity levels, thus showing that the scheme is effective and scalable. Figure 3(b) shows values of de-anonymization time for the DRC approach. Here we consider only the DRC scheme as only the DRC scheme can perform de-anonymization. The basic variation trend of the de-anonymization time is similar as the anonymization time. However, the de-anonymization time for all anonymity levels is shorter than the corresponding anonymization time as the anonymization process dynamically adds the null transition to avoid the collisions during de-anonymization. Also, since metadata helps record the collisions, the de-anonymization process directly resolves the collision and it results in a shorter de-anonymization time. In Figure 3(c), the relationship between anonymity level and relative spatial resolution (RSR) is given. Instead of picking segments from the adjacent segment set, RS randomly selects segments from the whole area defined by the maximum spatial resolution, which gives larger cloaking region, compared with RNE and DRC. Also, we see that the relative spatial resolution of DRC is very close to RNE, especially for higher anonymity levels. Both of the two algorithms pick segments from the adjacent segments of cloaking region, so the structure of the cloaking region is tighter. In Figure 3(d), success rate for the three schemes with varying anonymity level is measured. Among the three schemes, RS always gets the highest success rate. This is an inherent feature of RS because it succeeds in all cases unless the number of users within the whole area defined by maximum spatial resolution cannot satisfy the required anonymity level. Therefore, the success rate of RS can be seen as the upper bound of all the anonymimization schemes. Compared with RNE, the success rate of DRC is closer to the performance of RS. Though the success rate for all the three schemes slightly decreases for increased anonymity level, the overall success rate of the proposed scheme is significantly high. It can be seen that even for the highest anonymity level, the success rate of RNE is higher than 80% and that of DRC is higher than 92%.

## 5 Related Work

As location privacy is gaining more attention, several research efforts on location privacy were made in recent years. Various models of privacy protection systems have been proposed to support privacy-preserving and efficient data communication between mobile clients and servers, including client/server models [4], trusted third party models [5, 13, 19] and distributed models [7, 8]. While the client/server model is simpler to implement, due to the lack of global knowledge in client side, the protection cost is in general higher with lower protection quality. For the distributed model, a decentralized cooperative p2p network is deployed among clients, which requires high overhead to support the infrastructure of communication and movement of clients. In contrast, a trusted third party anonymizer model, such as the one used in our work yields good performance in both query processing quality and computation cost. Various privacy protection algorithms proposed for data privacy have been adopted for protecting location privacy of mobile users. The types of privacy protection algorithms include anonymization [1, 5, 8, 10, 13, 20], data suppression [18], trajectory inference prevention [2, 3, 14–16] and encryption [11]. While most of the existing schemes are aimed at preventing the adversary from distinguishing the location of a given user from that of other users, their perturbation techniques are mostly unidirectional and lack the ability to de-anonymize the perturbed information even when a user accessing the information has suitable credentials for obtaining finer information. ReverseCloak algorithms proposed in [12] provide support for multi-level privacy control with the ability to reduce the granularity of the perturbed location based on access credentials. However, in contrast to this proposed work, the approach in ReverseCloak is to perform the location cloaking in a collision-free manner by avoiding the possible segment expansions that may lead to collisions. The cloaking algorithm proposed in this paper takes an alternate approach of allowing collisions to happen during the cloaking expansion process and resolves them with the help of metadata information during the de-anonymization process. While avoiding collisions during cloaking expansion avoids the overhead of metadata management, the approach of collision resolution using metadata can be more efficient in terms of lower anonymization and de-anonymization time overhead.

## 6 Conclusion

In this paper, we present a de-anonymizable location cloaking scheme for protecting location privacy in mobile computing system. Unlike existing location cloaking techniques which are developed as unidirectional location perturbation algorithms, the dynamic reversible cloaking algorithm proposed in this paper can restore the original location information from the perturbed cloaking region when suitable access credentials are provided. Our experiments based on GTMobisim show that the proposed cloaking scheme is efficient and scalable. In our future work, we plan to apply the reversible cloaking algorithm developed in this work to protect continuous location-based queries which require continuous exposure of location information leading to possible correlation attacks.

## References

1. B. Bamba, L. Liu, P. Pesti, *et al*. Supporting anonymous location queries in mobile environments with privacygrid. *Proceedings of the 17th international conference on World Wide Web*, ACM, 237-246, 2008.
2. A. Beresford and F. Stajano. Location Privacy in Pervasive Computing. in *Pervasive Computing*, 46-55, 2003.
3. A. Beresford, S. Frank. Mix zones: User privacy in location-aware services, 2004.
4. R. Cheng,Y. Zhang, E. Bertino, *et al*. User Location Privacy in Mobile Management Infrastructures. In *Proc. of Privacy Enhancing Technology Workshop(PET'06)*, 2006.
5. B. Gedik and L. Liu. Location Privacy in Mobile Systems: A Personalized Anonymization Model. In *25th IEEE International Conference* on. IEEE, 620-629, 2005.
6. B. Gedik, L. Liu. A customizable k-anonymity model for protecting location privacy, 2004.
7. G. Ghinita,P. Kalnis, S. Skiadopoulos. MOBIHIDE: a mobilea peer-to-peer system for anonymous location-based queries. *Advances in Spatial and Temporal Databases*, Springer Berlin Heidelberg, 221-238, 2007.
8. G. Ghinita, P. Kalnis, S. Skiadopoulos. PRIVE: anonymous location-based queries in distributed mobile systems. *Proceedings of the 16th international conference on World Wide Web*, ACM, 19(12): 371-380, 2007.
9. GTMobiSim. `https://code.google.com/p/gt-mobisim/`.
10. P. Kalnis, G. Ghinita, K. Mouratidis, *et al*. Preventing location-based identity inference in anonymous spatial queries. *Knowledge and Data Engineering*, IEEE Transactions on, 19(12): 1719-1733, 2007.
11. A. Khoshgozaran and C. Shahabi. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy. In *Proc. of SSTD*, 2007.
12. C. Li, B. Palanisamy. ReverseCloak: Protecting Multi-level Location Privacy over Road Networks. in *Proc. of 24th ACM International Conference on Information and Knowledge Management*, 2015, in press.
13. M. F. Mokbel, C. Y. Chow, and W. 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*, 2006.
14. B. Palanisamy and L. Liu. Attack-resilient mix-zones over road networks: architecture and algorithms. *IEEE Transactions on Mobile Computing*, 14(3), 495-508, 2015.
15. B. Palanisamy, L. Liu, K. Lee, *et al*. Anonymizing Continuous Queries with Delay-tolerant Mix-zones on Road Networks. *Distributed and Parallel Databases*, 32(1), 91-118, 2014.
16. B. Palanisamy, L. Liu. Mobimix: Protecting location privacy with mix-zones over road networks. in *27th International Conference on Data Engineering*, 494-505, 2011.
17. L. Sweeney. A model for protecting privacy. in *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 557-570, 2002.
18. M. Terrovitis, N. KMamoulis. Privacy preservation in the publication of trajectories. In *Proc. of 9th International Conference on Mobile Data Management*, IEEE, 65-72, 2008.
19. T. Wang, L. Liu, P. Pesti. Privacy-aware mobile services over road networks. *Proceedings of the VLDB Endowment*, 2(1): 1042-1053, 2009.
20. Z. Xiao, X. Meng, J. Xu. Quality aware privacy protection for location-based services. *Advances in Databases: Concepts, Systems and Applications*, Springer Berlin Heidelberg, 434-446, 2007.