ReverseCloak: Protecting Multi-level Location Privacy over Road Networks

Chao Li
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA, USA
chl205@pitt.edu

Balaji Palanisamy
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA, USA
bpalan@pitt.edu

ABSTRACT

With advances in sensing and positioning technology, fueled by the ubiquitous deployment of wireless networks, location-aware computing has become a fundamental model for offering a wide range of life enhancing services. However, the ability to locate users and mobile objects opens doors for new threats - the intrusion of location privacy. Location anonymization refers to the process of perturbing the exact location of users as a cloaking region such that a user's location becomes indistinguishable from the location of a set of other users. A fundamental limitation of existing location anonymization techniques is that location information once perturbed to provide a certain anonymity level cannot be reversed to reduce anonymity or the degree of perturbation. This is especially a serious limiting factor in multi-level privacy-controlled scenarios where different users of the location information have different levels of access. This paper presents ReverseCloak, a new class of reversible location cloaking mechanisms that effectively support multi-level location privacy, allowing selective de-anonymization of the cloaking region to reduce the granularity of the perturbed location when suitable access credentials are provided. We evaluate the ReverseCloak techniques through extensive experiments on realistic road network traces generated by GT-MobiSim. Our experiments show that the proposed techniques are efficient, scalable and provide the required level of privacy.

Categories and Subject Descriptors

H.2.7 [**Database Management**]: Database Administration—*Security, integrity, and protection*; H.2.8 [**Database Management**]: Database Applications—*Spatial databases and GIS*

General Terms

Algorithms, Design.

CIKM'15, October 19–23, 2015, Melbourne, Australia.
© 2015 ACM. ISBN 978-1-4503-3794-6/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2806416.2806437.

Keywords

Location privacy, multilevel privacy, reversible cloaking algorithm, k-anonymity, road network.

1. INTRODUCTION

The proliferation of low-cost GPS-enabled mobile devices and the ubiquitous deployment of wireless networks drive the rapid emergence of mobile technology to satisfy the growing demands for location-based service applications. Examples of location-based applications include searching nearest points of interest *("where is the nearest gas station to my current location?")*, spatial alerts *("Remind me when I drive close to the grocery store.")*, location-based social networking *("Tell my friends where I am.")*. The market of location-based systems is predicted to be $3.3 billion in 2013 [23] and is expected to grow further to sustain the growth in LBS and mobile applications markets [18]. While location-based services find numerous potential benefits, they also open new doors for privacy threats. For example, through statistical analysis of usual haunts of the users, an attacker can speculate about user's private information, such as hobbies, living habits, health status and so on. In the worst cases, disclosure of location information can be even life-threatening [25]. Such privacy threats directly affect people's attitude towards location-based applications as people become more conscious and aware of the potential privacy risks associated with it [16].

Location privacy is a system-level capability of location-based systems, which controls the access to location information at different spatial and temporal granularity instead of completely stopping access. Location anonymization refers to the process of perturbing user location information such that it masks the exact location of the user using a cloaked region. A subject is said to be location k-anonymous if her location information is indistinguishable from that of $k-1$ other users . However, a fundamental limitation of all existing location privacy protection schemes is that location information once perturbed to provide a certain anonymity level cannot be reversed to reduce anonymity or the degree of perturbation. This is especially a serious limiting factor in multi-level privacy controlled scenarios where different users of the location information have different levels of access on the exposed location. For instance if Alice is concerned of her location privacy, she might decide to expose her location with a certain privacy level with one location-based service provider. However, she may wish to give some other location-based service providers access to a reduced anonymity level as she may trust those providers more than the others. Also, Alice may want to give access to her exact location to some providers who are most trustworthy. In existing schemes, location information once lost during perturbation cannot be restored and therefore, becomes unavailable when more privileged users try to access finer information, resulting in a loss of utility.

There are several location anonymization techniques [3, 10, 11, 14, 19, 28] proposed in the literature to tackle the location privacy problem. Most of these techniques are developed as unidirectional location perturbation schemes that can only perturb in an irreversible manner without being able to de-anonymize when users with higher privilege have access to more fine granular information. This paper presents ReverseCloak, a new class of reversible location cloaking mechanisms that allows selective de-anonymization of the cloaking area when suitable access credentials are provided. A key objective of our work is to support multilevel location privacy requirements that allow different users to infer different levels of information from the same exposed location based on their access credentials and the access privilege levels entitled to them. Our proposed approaches transform the raw point location of a user into a cloaked location region such that finer location information can be obtained through careful de-anonymization using a shared secret anonymization key. However, without the secret key, the cloaked region preserves strong privacy properties, allowing no additional information to be inferred. To the best of our knowledge, the techniques proposed in this paper are the first set of location privacy protection schemes aimed at supporting multi-level privacy/utility controls through novel development of secure reversible location cloaking techniques.

The rest of the paper is organized as follows: Section 2 provides a background and an overview of the multi-level reversible location anonymization problem. In Section 3, we discuss two reversible multi-level location anonymization schemes namely reversible global expansion and reversible pre-assignment-based local expansion techniques that support multi-level location privacy. In Section 4, we present the analysis of our experiments on realistic road network traces generated using GTMobiSim. We discuss related work in Section 5 and we conclude in Section 6.

2. CONCEPTS AND MODELS

In this section, we first describe the road network model used to capture the mobility features of mobile users and then introduce the concept of location cloaking. After that, we define the multilevel reversible location privacy problem and discuss two effective attack models used to evaluate the attack resilience of the proposed solutions.

2.1 Road network model

We model the road network as a graph $G = (V_G, E_G)$, where V_G represents the set of junctions and E_G represents the set of road segments. An example is shown in Figure 1, which consists of 23 junctions and 33 road segments. A junction is defined as the crossover point of any two roads or the end of a road. A road segment is defined as the direct road connecting any two adjacent junctions. Each segment is uniquely determined by the two junctions associated with it while each junction is associated with one or more adjacent road segments. The number of segments adjacent to a road junction, j_i defines the degree of the junction, denoted by $d(j_i)$. Based on that, each junction can be classified into three categories namely an intersection junction ($d \geqslant 3$), an intermediate junction ($d = 2$) and an end junction ($d = 1$). In the road network, each mobile user is assumed to move along the segments and change direction only at junctions. A user interested in sharing her anonymized location information forwards her true location with the anonymization requirements and spatial information to a trusted anonymization which then transforms this raw point location into a perturbed spatial region that meets the required privacy levels.

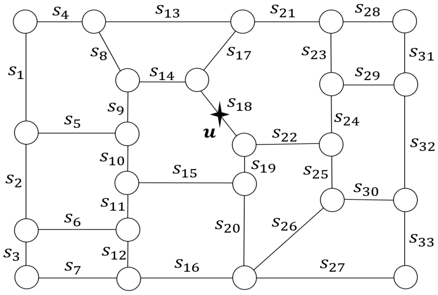


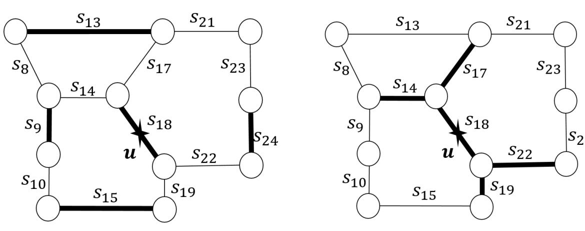
Figure 1: Road network model

2.2 Location anonymization

We consider the key privacy requirement arising in a road network namely *location k-anonymity*, which ensures that the exposed location of a user is indistinguishable from a set of other users on the road network.

Definition 1 (*Location k-anonymity*). *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.*

In a personalized location privacy model, for each location anonymization request, the level of k-anonymity is decided by the user in a customizable manner. Also, in order to bound the size of the cloaking region that has a direct influence on the performance of the anonymous query processing technique [19, 27], a customizable maximum spatial resolution level, denoted by σ_s is specified. These two parameters together define the user-defined privacy profile: (δ_k, σ_s). In the past, several cloaking models have been proposed for location anonymization. In this paper, we discuss two broad class of techniques namely (i) random sampling and (ii) road-network-based expansion as candidate approaches for location cloaking over road networks.



(a) Random sampling (b) Road network expansion

Figure 2: Two typical anonymizatioin models

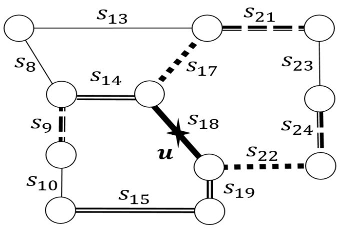


Figure 3: Multilevel reversible location anonymization

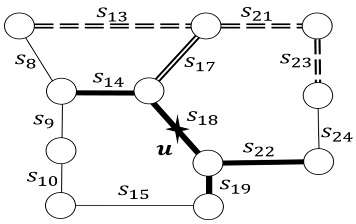


Figure 4: Replay attack

Random sampling: Given the anonymization request and the user-defined profile, the random sampling approach first chooses the road segment containing the actual user. It then randomly adds segments within the bounded area restricted by σ_s into the cloaking region until the requirements of δ_k is met. In the example shown in Figure 2(a), we find that the entire region is a part of the road network shown in Figure 1. It represents the region restricted by the spatial tolerance level σ_s. To meet the requirement of δ_k, in addition to the segment containing the actual user, the algorithm randomly adds four additional segments to form the cloaked region. While the random sampling approach attempts to ensure higher randomness in the cloaking process, it is however shown that it is not desirable from an anonymous query processing perspective as the discrete segments added in the random sampling process are not amenable for scalable and efficient query processing [27]. From a query processing perspective, a well-connected cloaked region over the network is more efficient as the query processing complexity is simplified. This motivates us to the road-network-based expansion techniques, which are described next.

Road-network-based expansion: In the road-network-based expansion approach, instead of randomly choosing segments from the bounded region in a discrete manner, the segments are chosen continuously based on an expansion scheme. The expansion begins from the segment containing the actual user and randomly expands such that each newly added segment is adjacent to at least one other segment in the currently formed cloaking region. Figure 2(b) shows an example of a road-network-based cloaking where the chosen segments, satisfying δ_k, form a tightly connected structure than the one generated by random sampling. Several road network-based cloaking schemes have been proposed in the past [27, 29] and they vary in terms of how randomly the expansion process proceeds to reach the required anonymity levels.

As it can be noted, most existing work on location privacy have focused on achieving variations of location k-anonymity and in such schemes, anonymity level once achieved in the cloaked region cannot be reduced when users with higher privileges need to access fine granular information. Therefore, in multi-level access controlled scenarios, such irreversible approaches can protect location privacy at at-most one privacy/utility level.

2.3 Multilevel reversible location privacy

The focus in this paper is developing a new class of reversible cloaking techniques, which can support multi-level location privacy in access controlled scenarios. In such cases, the location privacy of users is protected under multiple privacy levels, with higher anonymity levels for users with lower privileges and lower privacy levels for users with higher privileges.

In the multi-level reversible location privacy framework, a trusted anonymizer obtains the raw location information from the mobile clients with the user-defined profile. However, with the multi-level privacy model, the user-defined profile consists of the privacy requirements for each privacy level, L^i, except L^0 referring to a cloaking region with only the segment of actual user. Accordingly, the user-defined privacy profile is represented by (δ_k^i, σ_s^i), where $1 \leq i \leq N-1$ and N denotes the number of privacy levels. In addition, each privacy level, L^i is associated with a shared secret key, Key^i, which is used to drive anonymization process for that privacy level. Therefore, with access to the anonymization key of a particular privacy level, users of the cloaked location can selectively de-anonymize the cloaked region to reduce privacy levels to obtain finer location information. A detailed example of a four level case is shown in Figure 3. The segment s_{18} contains the actual user belongs to level, L^0. Using the anonymization key Key^1, $\{s_{17}, s_{22}\}$ are added to reach the privacy level, δ_k^1 of L^1. Then, Key^2 is used further to extend the cloaking region to meet δ_k^2 of level L^2 by adding segments $\{s_{14}, s_{15}, s_{19}\}$. Finally, $\{s_9, s_{21}, s_{24}\}$ are added using the anonymization key, Key^3 to reach the highest privacy level, L^3.

Later, when the cloaked location information needs to be reduced in privacy levels, it can be done using the anonymization keys. For instance, for accessing the information at the lower privilege level, L^2, Key^3 can be used to exactly identify and remove the segments $\{s_9, s_{21}, s_{24}\}$ from the cloaking region to reduce to the cloaked region corresponding to level, L^2. Similarly, using both Key^3 and Key^2, the segments $\{s_9, s_{21}, s_{24}, s_{14}, s_{15}, s_{19}\}$ can be identified and removed from the cloaking region to reduce to level, L^1. Therefore, by merely managing the shared anonymization keys among the location users at different privilege levels, the whole process protects location privacy under multiple discrete levels as customized in the user-defined privacy profile.

2.4 Replay attack

The security and privacy strength of any location privacy mechanism comes from the attacker's inability to infer the exact user location within the cloaked location. Specifically, the attacker is interested in analyzing the associativity of the actual user with each road segment in the cloaked location. Therefore, a segment with higher associativity is more likely to be the segment of the actual user. From the attacker's perspective, the attack is most ineffective when the associativity for the segments follows a uniform distribution. This leads to the highest uncertainty in identifying the true location of the user. In this paper, we use the replay attack

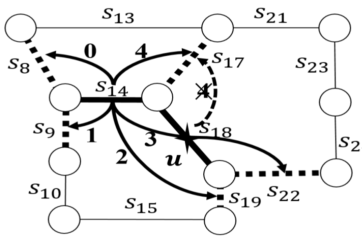


Figure 5: Forward transition collision

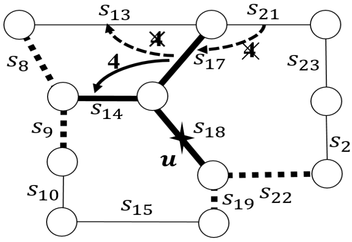


Figure 6: Backward transition collision

to evaluate the attack-resilience of the proposed cloaking schemes. Similar to the adversary models in [3, 19, 27], we primarily focus on snapshot exposure of location information for supporting snapshot location-based queries. For continuous location-based queries, with the additional ability to combine and correlate information from the location exposure of multiple snapshot instances, the adversary's chances of inferring the true location can be increased [9, 22]. While addressing such query-correlation attacks is a promising direction of future work, the scope of the replay attack model considered in our work is limited to snapshot queries.

In the replay attack, each segment within the cloaking region is iteratively considered to be the segment of the actual user to compute the associativity for all the segments. If the number of segments shared by the replayed cloaking region generated from a segment, s_i, and the real cloaking region generated from the real start segment is N_i, which is 1, (s_{17}), in the example of Figure 4. The associativity, A_i of s_i, can be calculated as $A_i = \frac{N_i}{\sum N_i}$. After obtaining A_i for all the segments within the cloaking region, the uncertainty of the attacker can be quantified by Entropy [24] measured as $E = -\sum A_i \log A_i$. The Entropy is a measure of the amount of information required to break the anonymity provided by the system. Therefore, the larger is the entropy, the higher is the uncertainty of the attacker and the scheme is more attack-resilient.

In the next section, we present our proposed reversible cloaking mechanisms that support efficient multi-level location privacy over road networks.

3. REVERSIBLE LOCATION CLOAKING

In this section, we present the proposed ReverseCloak cloaking mechanisms that support efficient multi-level location privacy over road networks. Before we present the details of the proposed approach, we discuss why conventional location cloaking techniques that perform road network-based expansion are not amenable for reversibility. We then discuss the key ideas behind the proposed reversible cloaking approach and discuss the challenges and the solution techniques for achieving reversible location cloaking over road networks.

In road network-based location cloaking schemes, a road segment is considered as the basic unit of expansion. At any point in the expansion process, the set of road segments selected to be part of the cloaking region forms the cloaking set. Therefore, in the beginning of the expansion process, the cloaking set contains only the segment containing the actual user. As the expansion process proceeds, the segments are added to the cloaking set until the required privacy levels are met. In the de-anonymization process, upon providing the anonymization key, the segments in the cloaking set are removed in the reverse order to reduce the anonymity level. At the end of the process, the segment containing the actual user is obtained.

The conventional road-network-based expansion cloaking mechanisms are designed to expand either randomly or semi-randomly based on additional factors such as network distance and the number of users present in the added segment. For example, in Figure 5, we find that the actual user who sends query is located in segment s_{18} and together with the second chosen segment s_{14} form the current cloaking region. At this phase of the cloaking process, the five segments, denoted with dotted lines are the candidates of the next segment to be added to the cloaking region. Since these segments are adjacent to at least one segment within the current cloaking region, these five segments form the candidate set.

For conventional expansion-based cloaking algorithms, the newly added segment within the candidate set may be selected either randomly or semi-randomly based on additional factors. However, in both the cases, we note that it is impossible for the privileged users to de-anonymize it as the cloaking region inherently protects against such de-anonymization attempts by the adversary due to the randomness. In order to support effective multi-level privacy protection, an ideal situation is when the cloaked region provides the highest uncertainty to an attacker attempting to de-anonymize it but has no uncertainty to users who possess the privileges to de-anonymize it. We achieve this property by devising our road network-based expansion process to be driven through an anonymization key to handle the randomness. In our approach, the cloaking process expands in a pseudo-random fashion by adding road segments into the cloaking region until the privacy requirements are met.

We refer to each addition of a new segment in anonymization process as a forward transition from a segment within the current cloaking set to a segment in the current candidate set. Similarly, during the de-anonymization process, the removal of each segment in the cloaked region represents a backward transition, which is essentially an inversion of the forward transition in the anonymization process. In this way, the relationships between any two successively selected segments are established. Therefore, the anonymization and de-anonymization can be considered as a series of forward and backward transitions respectively. For two segments s_i and s_j, the forward transition from s_i to s_j is denoted by $ft\{s_i \rightarrow s_j\}$ and the backward transition from s_j to s_i is denoted by $bt\{s_j \rightarrow s_i\}$. They together form a transition pair. To reverse the anonymization process, the goal is to make the transition chain reversible so that the forward and backward transitions in pair are correlated one by one. In the proposed approach, we use a secret key to generate the transition chain in a pseudo-random manner. Specifically, the anonymization process generates a stream of pseudo-random numbers using the anonymization key as the seed value. In each expansion step, the candidate set contains a certain number of candidate segments and each of them corresponds to a candidate forward transition pointing to them from the segment selected in the last expansion step. Then, each of them is assigned a unique transition value to be used for this step. Among these possible candidate forward transitions, the algorithm pseudo-randomly picks a unique transition based on a *pick value*, determined by the pseudo-random integer and the size of the candidate set. This pick value enables to choose a unique candidate forward transition based on the transition value. In the example shown in Figure 5, the candidate expansion region contains 5 segments and the transition values 0 to 4 are assigned to the five possible transitions. Here the pick value is generated as $R_i mod\ 5$, where R_i is the i^{th} pseudo-random number in the stream. The pick value uniquely selects the transition with a matched transition value.

In contrast, in the de-anonymization process, uniquely identifying the exact backward transition is more challenging. Even though the pick value of every step can be calculated using the secret key in the de-anonymization process, the assignment of transition values for forward and backward transitions should be done very carefully. Otherwise collision of transition values may occur for either forward transitions (Figure 5) or backward transitions (Figure 6), which may result in incorrect selection of backward transitions. In the example shown in Figure 5, we find that when s_{17} is removed in the de-anonymization process, the transition values from the segments within the cloaking region to segment, s_{17} should be checked. If the pick value here is 4, for instance, we find that the transition corresponding to $ft\{s_{14} \rightarrow s_{17}\}$ matches the pick value, 4. While this indicates that s_{14} is a potential previous segment, we notice that a collision may happen if another transition such as value of $ft\{s_{18} \rightarrow s_{17}\}$ also has a transition value matching the pick value, 4, in this case. Therefore, the de-anonymization can no longer uniquely identify the backward transitions and de-anonymize the cloaked location.

In general, there are two ways to handle collisions in a reversible cloaking scheme. A simple and straight-forward approach is to simply record information about the collisions as additional metadata which can then be used to aid the de-anonymization process to enable skipping the corresponding colliding transitions [17]. However, from a location cloaking mechanism design standpoint, it is more desirable to have a collision-free anonymization approach that guarantees that the de-anonymization process is collision-free. Such a scheme can operate without any additional metadata. In this paper, we devise collision-free reversible cloaking mechanisms based on two different approaches. In the first approach, collisions are prevented by carefully assigning the transition values to the forward transitions $ft\{s_i \rightarrow s_j\}$ in a dynamic on-the-fly manner during anonymization. We propose the reversible dynamic global expansion scheme based on this (Section 3.1). In the second approach, both the backward and forward transitions are assigned transition values and we ensure that all backward and forward transition values are unique. In the example shown in Figure 6, the backward transitions are also assigned unique transition values. Since the pick values for a transition pair are same, we need to ensure that their transition values are also the same. Therefore, for instance if the pick value is 4 and if the transition value of $bt\{s_{17} \rightarrow s_{14}\}$ is also 4, then de-anonymization can proceed successfully without collision as only s_{14} matches the backward transition value uniquely. However, since 4 is uniquely assigned among the backward transitions from s_{17} and the transition values for a transition pair are same, we note here the same transition value, 4 in this example, cannot be assigned to any other forward transitions to s_{17} or backward transitions from s_{17} (denoted by the dotted arrow lines). Therefore, to determine the transition value for a certain transition, the restrictions from other transitions need be carefully taken into account. We also note that this assignment of unique forward and backward transitions can be done *apriori* for the entire road network in a statically pre-assigned manner. We develop reversible pre-assigned local expansion scheme (Section 3.2) based on this approach.

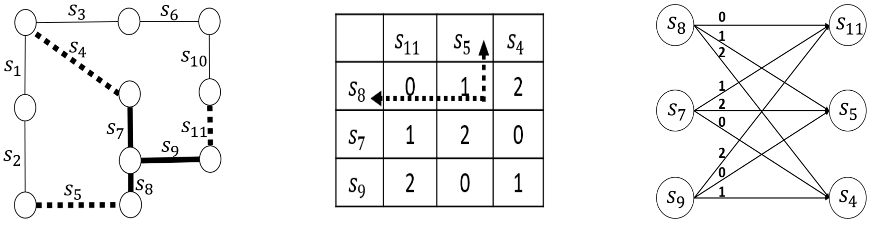
3.1 Reversible global expansion

The goal of the reversible global expansion cloaking mechanism is to perturb the raw point location of the actual user into a cloaked location region on the road network in a reversible manner such that the key privacy requirement namely location k-anonymity is met. The cloaking algorithm starts from the segment containing the actual user. For each transition, a transition matrix, containing transition values for all possible transitions between the current cloaking region and the candidate expansion region are generated. These transition values are carefully assigned to make sure that there are no repeated values in each row and column so that no collisions occur during de-anonymization. For each forward transition in the anonymization process, a pick value generated by the pseudo-random number uniquely determines the forward transition.. During de-anonymization, the same pick value can be used to find the forward transition uniquely.

Next, we introduce the mechanism to assign the transition values in a collision-free manner using the notion of n-hop neighboring segment on a road network.

Definition 2 (*n-hop neighboring segment*). *Two segments on a road network are n-hop neighboring segments of each other if there is a path between them on the road network passing through n junctions. Therefore, their topological distance on the road network is n hops.*

The proximity between road segments is defined based on their topological distance and therefore all n-hop neighboring segments of a segment form its n-hop neighboring set. Since there may be multiple road network paths between two segments, a segment may belong to multiple n-hop neighboring sets of another segment. The n-hop neighboring set for a single segment s_i is represented as N_i^n and the n-hop neighboring set for a cloaking segment set $C = \{s_a, s_b, s_c\}$ is defined as the n-hop global neighboring set, represented as $G_C^n = \{s_i \mid s_i \in N_a^n \bigcup N_b^n \bigcup N_c^n,\ s_i \notin C\}$. Therefore, the transition matrix used in this algorithm is formed by C and its global neighboring set G_C^n. Segments from global neighboring set with fewer hops are put into the global neighboring set in order until the number of segments in the global neighboring set is same as that of C, $|G_C^n| = |C|$. For each transition, a matrix $\mathbb{M}$ with X columns and Y rows is built,



	s_{11}	s_5	s_4
s_8	0	1	2
s_7	1	2	0
s_9	2	0	1

(a) Model (b) Transition matrix (c) Transition graph

Figure 7: Reversible global expansion

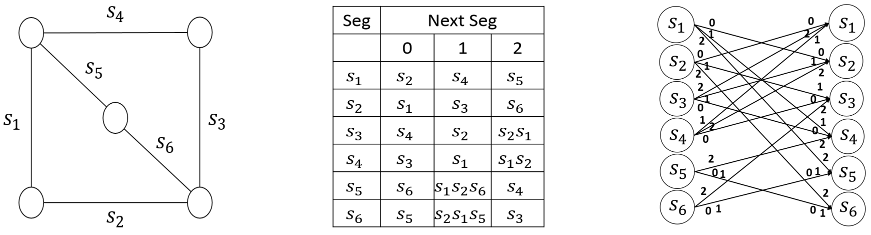
where $X = |G_C^n|$ and $Y = |C|$. Based on the segment length, the elements of G_C^n and C are sorted and mapped to columns and rows respectively. Here, if the first column and first row of $\mathbb{M}$ are numbered zero, then transition values can be calculated by $\mathbb{M}(x, y) = (x + y)\ mod\ X$, where x and y are the number of columns and rows respectively. Once the transition matrix is dynamically formed, the pick value, denoted by p, (calculated as $p = R\ mod\ X$, where R stands for the pseudo-random number), uniquely picks the forward transition among the set of candidate transitions for the currently formed cloaking region.

An example of the global expansion scheme is presented in Figure 7(a). We assume that the actual user asks for three privacy levels with (δ_k^1) and (δ_k^2) for L^1 and L^2 respectively. For this example, we assume that the currently formed cloaking region is $C = \{s_7, s_8, s_9\}$ and it is close to (δ_k^1), with $G_C^n = \{s_4, s_5, s_{11}\}$. The size of G_C^n can be larger for large n, but we just need the first three in this case to guarantee $|G_C^n| = |C|$. To select the next segment from G_C^n with the pick value generated using the key for level 1, L^1, a transition matrix is built as discussed above. The transition matrix is shown in Figure 7(b) and the corresponding transition graph is shown in Figure 7(c), where the arrow lines stand for forward transitions and the values represent the transition values. For this example, if the last selected segment is s_8, since $X = 3$ and $y = 0$, the pick value will be computed as $p = (R\ mod\ 3)$ and for $R = 7$, we will get $p = 1$. Therefore, based on the graph, the transition from s_8 with value 1 points to s_5. After adding the segment corresponding to this forward transition, the privacy requirement for L^1 may be met, however, the privacy requirements for L^2 may require adding more segments. In this case, the anonymization process changes the anonymization key to Key^2 and the process continues to pseudo-randomly expand until (δ_k^2) is met. In the de-anonymization process, it is straight-forward to see that users can de-anonymize the cloaking region if they have access to Key^2 in order to access the location information entitled to users at level, L^1. By removing the last selected segment from the current cloaking region, the remaining cloaking region can be used to build the same transition matrix and transition graph and since no collision exists in this approach, the backward transition can be uniquely determined by finding the forward transition that shares the same transition value as the generated pick value.

3.2 Reversible pre-assignment-based local expansion

Similar to the global expansion scheme, the reversible pre-assignment-based local expansion algorithm also starts the expansion from the segment containing the actual user. It keeps adding new segments until the privacy requirements of the user-defined profile is satisfied. As mentioned before, the transition values in the pre-assignment-based approach are assigned to both forward and backward transitions. However, unlike the reversible global expansion algorithm, where the forward transitions represent segments within the currently formed cloaking region to segments within the candidate expansion region, the forward transitions in the pre-assignment-based approach represent transitions from the previously added segment to a set of candidate segments for expansion. The candidate segments for expansion in the global scheme are neighboring segments of the entire currently formed cloaking region, however, the candidates for the local expansion scheme include only the neighboring segments of the previously added segment in the cloaking area.

The transition pre-assignment algorithm first determines the number of candidate transition segments for each transition. The algorithm ensures that for all the forward and backward transitions, this number is the same in order to ensure that the overall number of forward transitions is equal to that of backward transitions. For example, if the number of candidate transition segments is chosen as 3, the required transition values assigned to candidates should be 0, 1 and 2. Therefore, for each segment within the map, the three transition values need be carefully assigned to avoid collisions as discussed before. To build the transition graph, the algorithm establishes three tables namely a neighboring table, an encoding table and a decoding table. The encoding and decoding tables contain the transition values for the forward and backward transitions and the neighboring table contains the road network adjacency information for all the segments in the road network. For each segment s_i, its neighboring segments are recorded as a row in the order of topological distance. That is, the entries start from N_i^1 and end to N_i^m where m is the allowed maximum number of hops in a single expansion step. In this algorithm, the encoding table and decoding table are established together progressively to guarantee collision-free transitions for both forward and backward transitions. Concretely, the encoding and decoding tables are built by following three rules: (i) the candidates for each transition can be selected in a flexible manner, however, if a neighboring segment of s_i cannot be added into its encoding table due to collisions, the next neighboring segment should be considered. Given that the neighboring segment set is built based on the n-hop neighboring segment sets, the encoding and decoding tables bounded by the number of candidates can be always fully established, (ii) the transition values of any transition pair



(a) Model

Seg	Next Seg		
	0	1	2
s_1	s_2	s_4	s_5
s_2	s_1	s_3	s_6
s_3	s_4	s_2	$s_2 s_1$
s_4	s_3	s_1	$s_1 s_2$
s_5	s_6	$s_1 s_2 s_6$	s_4
s_6	s_5	$s_2 s_1 s_5$	s_3

(b) Encoding table

(c) Transition graph

Figure 8: Reversible pre-assignment-based local expansion

should be equal. For example, if s_j is added into the encoding table of s_i, then the transition value of $ft\{s_i \rightarrow s_j\}$, which is 0, should be same for s_i as well in the decoding table of s_j in order to guarantee that the transition value of $bt\{s_j \rightarrow s_i\}$ is 0. (iii) Once a certain transition value, v is assigned to $bt\{s_j \rightarrow s_i\}$, all the forward transitions from other segments except s_i to s_j cannot be assigned the same transition value, v. The encoding and decoding tables are gradually filled, for each neighboring segment, the encoding table shows a set of the remaining positions, denoted by E, and the decoding table also provides a set of remaining positions, denoted by D. Only if $E \cap D \neq \emptyset$, the neighboring segments are chosen to update the two tables.

An example with 6 segments is shown in Figure 8(a) where the number of candidates for each transition is set to 3. The corresponding encoding table is shown in Figure 8(b), which can be used to easily deduce the decoding table as the decoding table is a matrix transposition of the encoding table. Here, the candidate with one segment belongs to one-hop neighboring set N_i^1 while the candidates with two and three segments belong to N_i^2 and N_i^3 respectively. The final transition graph generated from the encoding and decoding tables is shown in Figure 8(c) where the transition values on the left stand for the forward transitions and the ones on the right represent the backward transition values. Once the pre-assigned transition graph is established, for any future cloaking requests, the algorithm proceeds to generate pick values pseudo-randomly and chooses the transitions that are guaranteed to be collision-free based on pre-assignment. Similarly, the de-anonymization process becomes straightforward, involving the removal of the road segments based on the backward transitions from this pre-assigned collision-free transition graph.

4. EXPERIMENTAL EVALUATION

In this section, we experimentally evaluate the performance and privacy offered by the proposed ReverseCloak algorithms. Before reporting our results, we first briefly describe the experimental setup.

4.1 Experimental setup

To simulate and compare different anonymization schemes, we use GTMobiSim mobile trace generator for road network [13]. Our experiments were designed based on a real road network map of northwest part of Atlanta, involving 6979 junctions and 9187 segments, obtained from maps of National Mapping Division of the USGS. There are 10,000 cars randomly generated along the roads based on Gaussian distribution. Once a car is generated, the associated destination is also randomly chosen and the route selection is based on shortest path routing.

In our experiments, four different anonymization schemes are implemented: Random Sampling (RS), Star-based road-network expansion (SE)[27], a candidate representative of existing road network-based expansion schemes, Reversible Global Expansion (RGE) and Reversible Pre-assignment-based Local Expansion (RPLE). The first two algorithms (RS and SE) are irreversible while the two ReverseCloak algorithms proposed in this paper (RGE and RPLE) are reversible and support multi-level privacy control. All the schemes are implemented in Java with the help of GTMobiSim.

4.2 Experimental results

Our experimental evaluation consists of three parts. First, we evaluate the performance of the selected cloaking algorithms by measuring anonymization time, de-anonymization time, relative spatial resolution and success rate. We collect and compare the results by varying user-defined parameters, δ_k and σ_s. Then, we evaluate the performance of the algorithms under multilevel privacy scenarios. Finally, we evaluate the effectiveness of these algorithms in terms of resilience to replay attack by measuring entropy towards varying δ_k and σ_s. Our results show that the proposed algorithms have good resilience towards replay attacks and effectively support multilevel privacy requirements while still maintaining good performance and efficiency.

4.2.1 Varying User-defined k-anonymity

This set of experiments evaluates the performance of the algorithms by varying the anonymity level δ_k as

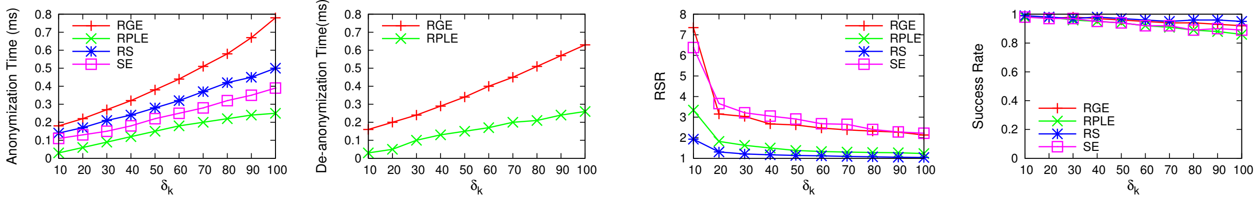
$$\delta_k = 10i \ for \ i = 1, 2...10$$

Here, the spatial tolerance, d, is set as a function of the anonymity level, δ_k such that

$$d = 400\sqrt{i} \ for \ i = 1, 2...10$$

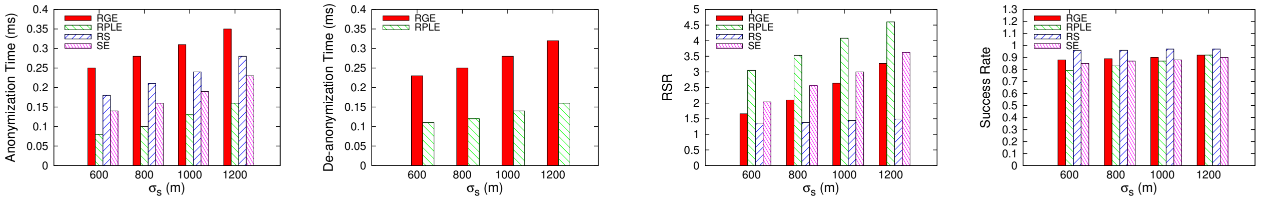
where the unit is meter(m). Therefore, the maximum allowable special region is a circular region with the user's actual location as the center and the spatial tolerance, d as the radius. We also set 5% standard deviation for each d. For this experiment, we consider only two privacy levels and for the multi-level reversible techniques, this privacy requirement represents the privacy levels of the least and most privileged users.

We compare the average anonymization time for the various approaches in Figure 9(a). In Figure 9(a), we find that RPLE is fastest in the anonymization phase among all the compared techniques. The reason is that the assignment of



(a) Anonymization Time (b) De-anonymization Time (c) Relative Spatial Resolution (d) Success Rate

Figure 9: Performance with Varying Anonymity Level



(a) Anonymization Time (b) De-anonymization Time (c) Relative Spatial Resolution (d) Success Rate

Figure 10: Performance with Varying Spatial Tolerance

transition values in the RPLE scheme has been done *apriori* and at the time location cloaking, the transition graph is directly looked up as compared to dynamically computing it on the fly in the RGE approach. Also, for all the algorithms, the anonymization time is longer for larger δ_k as stricter privacy requirements result in cloaking areas with more segments and it therefore requires addition of more segments into the cloaking area.

Figure 9(b) shows the impact of varying δ_k on the de-anonymization time. Since only ReverseCloak algorithms can perform de-anonymization of the cloaked region, RS and SE are not considered for this experiment. For both RGE and RPLE, the variation trends of de-anonymization time are similar as the anonymization time in Figure 9(a) as the computational complexity of the de-anonymization process is similar to that of the anonymization process. In both anonymization and de-anonymization phases, RPLE is faster than RGE because RPLE prevents collision in a *apriori* manner through its intelligent pre-assignment of forward and backward transitions while RGE prevents collision by dynamically assigning the transition values during location cloaking.

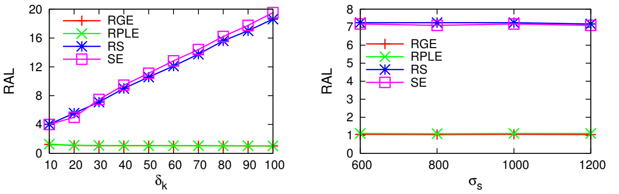
Figure 9(c) displays the impact of changing the anonymity level, δ_k on relative spatial resolution (RSR) which is defined the ratio of the size of the obtained cloaking area to size of the maximum allowable spatial area, specified by the spatial tolerance level, d. Here, RS has the lowest relative spatial resolution (RSR) as its candidate expansion region covers all the segments within the maximum allowable spatial area, thus making the size of the cloaking area close to the maximum spatial area even when δ_k is small. We also find that the relative spatial resolution of SE and RGE is larger than RPLE as the cloaking segments in SE and RGE are selected from a global neighboring segment set, providing a tighter structure as compared to a local neighboring set in the RPLE approach.

In Figure 9(d), we compare the success rate of anonymization process with varying δ_k value. The success rate represents the fraction of the cases where the cloaking algorithm is able to provide a cloaking region meeting the privacy requirements in terms of δ_k. We find that all the algorithms have a high success rate indicating that most of the anonymization requests are cloaked successfully to meet the privacy requirements. We also find that for all the schemes, the success rate decreases slowly when δ_k becomes very large. This is because a larger δ_k requires a larger cloaking area, which is harder to be satisfied by a given spatial tolerance. However, we note that even for higher anonymity levels, such as $\delta_k = 100$, the success rates of both RGE and RPLE are high and are close to 90%. RS keeps the highest success rate here as its failure occurs only when the total number of users within the maximum spatial area is smaller than δ_k. In fact, the success rate of the RS scheme defines the theoretical maximum success rate of the cloaking process for the given anonymization requests. We also note that SE and RGE have slightly higher success rate than RPLE as their cloaking regions have higher density and smaller size, thus being easier to meet the spatial tolerance requirement.

4.2.2 Varying User-defined Spatial Tolerance

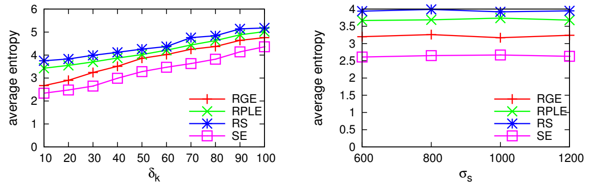
To test the impact of spatial tolerance, the anonymity level δ_k is set to be 30 with standard deviation of 10. The spatial tolerance values, σ_s, are chosen as 600m, 800m, 1000m and 1200m as the mean values with 5% standard deviation.

Figure 10(a) shows that all the algorithms experience an increase in cloaking time as σ_s increases. With an increase in the maximum spatial region size, the expansion process in the cloaking schemes have more candidate segments for expansion, causing an increase in the anonymization time. Figure 10(b) indicates that the de-anonymization time for the reversible algorithms follow a similar trend as the anonymization time. In Figure 10(c), we find that all techniques except RS have higher relative spatial resolution as the RS scheme attempts to distribute the cloaking segments throughout the entire maximum spatial cloaking area. In Figure 10(d), we measure the influence of spatial tolerance on the obtained success rate. We find that the success rate of all the algo-



(a) RAL - δ_k (b) RAL - σ_s

Figure 11: Multilevel Privacy Protection



(a) Resilience of Varying δ_k (b) Resilience of Varying σ_s

Figure 12: Repaly Attack

rithms grows with increase in σ_s. The increments for RS and RGE are not very significant as their initial success rate for 600m spatial tolerance itself is higher. We again note here that the success rate of the RS approach is a theoretical maximum as the RS scheme would fail only when there are less than δ_k users within the maximum allowable spatial region. However, for the other approaches, we find that an increase in spatial tolerance has a slight increase in the success rate. Higher spatial tolerance provides a larger candidate expansion region and permits larger cloaking area, so the privacy requirements of a query is much easier to be achieved. Besides that, since the failures mostly occur where the density of mobile users is low, with higher spatial tolerance, part of the request that were dropped with lower spatial tolerance may now become successful.

4.2.3 Multilevel Privacy Protection

To evaluate the multilevel location privacy performance, relative anonymity level (RAL) is measured. The relative anonymity in a multi-level privacy scenario refers to the ratio of the obtained anonymity to the anonymity level entitled to a given access privilege level. Figure 11(a) and Figure 11(b) show the average relative anonymity of the cloaking techniques with respect to δ_k and σ_s . In Figure 11(b), the requirement of δ_k is set as 30. In both Figure 11(a) and Figure 11(b), we assume that there are six access privilege levels with uniformly distributed requirement of δ_k. The users with lowest privilege have access to the cloaking region with the maximum δ_k while users with the highest privilege have access to the location with the lowest anonymity. A good cloaking algorithm should ensure that the RAL is close to one for all access privilege levels, thereby providing exactly the anonymity level required for that level. We notice that existing irreversible algorithms result in very large RAL due to their lack of reversibility. In contrast, ReverseCloak algorithms achieves a RAL close to one for all privilege levels ensuring that users in each privilege level obtain the information at the level of anonynmity and accuracy entitled to them, thus protecting the multilevel privacy successfully.

4.2.4 Attack Resilience

This set of experiments evaluate the effectiveness of the algorithms in terms of their resilience to replay attacks. For replay attack, average information entropy is calculated as the metric to evaluate the uncertainty of the attacker: $E = -\sum A_i \log A_i$, where A_i is the associativity for each segment. Here, higher entropy means higher randomness and higher uncertainty for the attacker in inferring the true location of the user, thus leaking out less information and providing better privacy protection. Figure 12(a) shows average entropy of the replay attack with varying δ_k. Here, RS keeps the highest entropy for the whole range as its cloaking process is completely random and is therefore the most resilient to replay attacks. We also find that the entropy of SE is high as the road network expansion has a good degree of randomness, however, both RGE and RPLE provide even higher entropy, indicating that the randomness of the expansion process in RGE and RPLE is much higher than that of the SE approach. In the RGE and RPLE schemes, the randomness of both these schemes are only a little lower than the RS scheme. Figure 12(b) shows the average entropy with varying σ_s. Here, the anonymity requirement, δ_k is set as 30. Compared with the previous case, the entropy for all the algorithms is nearly constant suggesting that larger spatial tolerance does not have impact on the attacker's uncertainty. Our final observation here is that in general, higher δ_k provide higher entropy for all the algorithms, illustrating that a cloaking area with more segments provides better protection.

5. RELATED WORK

Location privacy has been an active area of research in the past. Broadly, location privacy protection mechanism can be classified into policy-based protection techniques and inference prevention-based techniques. Policy-based schemes give users permission to define privacy rules according to the service request, thus getting users' active participation. The inference-prevention schemes are more focused on prevention by protectively processing and perturbing the location information prior to disclosure. The latter can be further broken down into location data perturbation techniques represented by [1, 10, 19, 28] and trajectory inference prevention techniques represented by [5, 4, 21, 22, 20]. Location data perturbation schemes consists of perturbation through dummies [15], spatial location cloaking [3, 7, 10, 11, 14, 19, 28] and encryption-based techniques [1]. Recent work has studied the location privacy problem by perturbing the location information based on differential privacy constraints prior to disclosure [2, 6]. By replacing the real location information with dummies, location privacy can be protected, but the safety and service quality is dependent on the distance between the two positions. Encryption-based techniques provide strong confidentiality properties, however, an encrypted location has very little utility compared to a perturbed location which can be effectively used for query processing. In the past, there has been many works related to spatial location cloaking. To proactively protect user's location privacy, k-anonymity, which was proposed for sensitive data protection [26], was applied to protect location privacy in the context of location-aware systems [12]. Since then, the techniques related to spatial cloaking has been developing rapidly. *CliqueCloak* algorithm proposed in 2004 considered the individual user's personalized privacy requirement for the first time [10]. A grid-based cloaking framework, *Casper* further extended this model with a privacy-aware query processor [19]. Subsequently, a directed-graph based cloaking algorithm was proposed to improve the success rate of anonymization [28] and the Hilbert Cloak algorithm uses a Hilbert curve to fill the whole area and track users [11]. While these techniques were designed for mobile users traveling on Euclidean space, recent work has considered the location cloaking problem under a constrained road network model [8, 27, 30]. As we can observe, most existing location privacy protection mechanisms have focused on developing unidirectional location perturbation approaches that does not allow fine granular information to be inferred even when some users have the privileges to access it. The approach presented in [17] introduces a de-anonymizable location perturbation scheme by recording some additional information as metadata during the anonymization process. However, such an approach requires the overhead of maintaining and managing the additional metadata, without which, the perturbed data cannot be reduced in privacy levels. In comparison, to the best of our knowledge, the work presented in this paper is the first comprehensive set of location privacy protection mechanisms aimed at providing a multi-level reversible location privacy model to support privacy-preserving exposure of location information in access controlled environments.

6. CONCLUSION

In this paper, we presented ReverseCloak, a new class of reversible location privacy protection mechanisms for supporting multi-level privacy requirements in access controlled environments. We argue that conventional location privacy protection techniques are not inherently designed to support reversible privacy and we proposed two reversible location cloaking mechanisms namely (i) reversible dynamic global expansion scheme and (ii) reversible pre-assignment-based local expansion technique that effectively support multi-level privacy, allowing users with higher privileges to obtain finer information through reduced anonymity levels. Extensive experiments conducted on GTMobiSim show that the proposed techniques are efficient, scalable and achieve the multilevel functionality through their reversibility property. Our ongoing work is focused on applying the principles and techniques developed in this work to protect multi-level privacy for continuous location-based services that require continuous exposure of location information as compared to one-time exposure in the case of snapshot queries.

7. REFERENCES

[1] S.I. Ahamed, M.M. Haque, and C.S. Hasan. A novel location privacy framework without trusted third party based on location anonymity prediction. *ACM SIGAPP Applied Computing Review*, 12(1):24–34, 2012.

[2] M.E. Andres, N.E. Bordenabe, K. Chatzikokolakis, et al. Geo-indistinguishability: differential privacy for location-based systems. In *20th ACM SIGSAC conference on Computer and communications security*, pages 901–914, 2014.

[3] B. Bamba, L. Liu, P. Pesti, et al. Supporting anonymous location queries in mobile environments with privacygrid. In *17th international conference on World Wide Web*, pages 237–246, 2008.

[4] A. Beresford and S. Frank. Mix zones: User privacy in location-aware services. 2004.

[5] A. Beresford and F. Stajano. Location privacy in pervasive computing. In *Pervasive Computing*, pages 46–55, 2003.

[6] N.E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Optimal geo-indistinguishable mechanisms for location privacy. In *21th ACM SIGSAC Conference on Computer and Communications Security*, pages 251–262, 2014.

[7] R. Cheng, Y. Zhang, E. Bertino, et al. Preserving user location privacy in mobile data management infrastructures. In *Privacy Enhancing Technologies*, pages 393–412, 2006.

[8] H.J. Cho, S.J. Kwon, R. Jin, et al. A privacy-aware monitoring algorithm for moving k-nearest neighbor queries in road networks. *Distributed and Parallel Databases*, pages 1–34, 2014.

[9] C.F. Chow and M.F. Mokbel. Enabling private continuous queries for revealed user locations. *Advances in Spatial and Temporal Databases*, pages 258–275, 2007.

[10] B. Gedik and L. Liu. A customizable k-anonymity model for protecting location privacy. 2004.

[11] G. Ghinita, P. Kalnis, and S. Skiadopoulos. Prive: anonymous location-based queries in distributed mobile systems. In *16th international conference on World Wide Web*, pages 371–380, 2007.

[12] M. Gruteser, D. Grunwald, and X. Liu. Anonymous usage of location-based services through spatial and temporal cloaking. In *1st international conference on Mobile systems, applications and services*, pages 31–42, 2003.

[13] GTMobiSim. `https://code.google.com/p/gt-mobisim/`.

[14] P. Kalnis, G. Ghinita, K. Mouratidis, et al. Preventing location-based identity inference in anonymous spatial queries. *IEEE Transactions on Knowledge and Data Engineering*, 19(12):1719–1733, 2007.

[15] H. Kido, Y. Yanagisawa, and T. Satoh. Protection of location privacy using dummies for location-based services. In *25th International Conference on Distributed Computing Systems*, pages 1248–1248, 2005.

[16] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

[17] C. Li and B. Palanisamy. De-anonymizable location cloaking for privacy-controlled mobile systems. In *9th International Conference on Network and System Security*, 2015, in press.

[18] Marketsandmarkets. Location based services market worldwide forecasts and analysis (2014 - 2019). 2014.

[19] M.F. Mokbel, C.Y. Chow, and W.G. Aref. The new casper:query processing for location services without compromising privacy. *VLDB Endowment*, pages 763–774, 2006.

[20] B. Palanisamy and L. Liu. Mobimix: Protecting location privacy with mix-zones over road networks. In *27th International Conference on Data Engineering*, pages 494–505, 2011.

[21] B. Palanisamy and L. Liu. Attack-resilient mix-zones over road networks: architecture and algorithms. *IEEE Transactions on Mobile Computing*, 14(3):495–508, 2014.

[22] 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.

[23] ABI Research. Location-based mobile social networking. 2008.

[24] R. Shokri, G. Theodorakopoulos, J.Y. Le Boudec, et al. Quantifying location privacy. In *32nd IEEE Symposium on Security and Privacy*, pages 247–262, 2011.

[25] K.O. Stalker. Victims should check for GPS. *Associated Press*, 2003.

[26] L. Sweeney. A model for protecting privacy. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, pages 557–570, 2002.

[27] T. Wang, L. Liu, and P. Pesti. Privacy-aware mobile services over road networks. *VLDB Endowment*, 2(1):1042–1053, 2009.

[28] Z. Xiao, X. Meng, and J. Xu. Quality aware privacy protection for location-based services. *Advances in Databases: Concepts, Systems and Applications*, pages 434–446, 2007.

[29] J. Xue, X. Liu, X. Yang, et al. Protecting location privacy using cloaking subgraphs on road network. In *11th Web Information Systems and Applications Conference*, pages 65–68, 2010.

[30] B. Ying and D. Makrakis. Protecting location privacy with clustering anonymization in vehicular networks. In *Computer Communications Workshops*, pages 305–310, 2014.