ReverseCloak: A Reversible Multi-level Location Privacy Protection System

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

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

Aravind Kalaivanan
School of Information Science
University of Pittsburgh
Pittsburgh, PA, USA
aak77@pitt.edu

Sriram Raghunathan
School of Information Science
University of Pittsburgh
Pittsburgh, PA, USA
srr47@pitt.edu

***Abstract*—With the fast popularization of mobile devices and wireless networks, along with advances in sensing and positioning technology, we are witnessing a huge proliferation of Location-based Services (LBSs). Location anonymization refers to the process of perturbing the exact location of LBS users as a cloaking region such that a user's location becomes indistinguishable from the location of a set of other users. However, existing location anonymization techniques focus primarily on single level unidirectional anonymization, which fails to control the access to the cloaking data to let data requesters with different privileges get information with varying degrees of anonymity. In this demonstration, we present a toolkit for *ReverseCloak*, a location perturbation system to protect location privacy over road networks in a multi-level reversible manner, consisting of an 'Anonymizer' GUI to adjust the anonymization settings and visualize the multilevel cloaking regions over road network for location data owners and a 'De-anonymizer' GUI to de-anonymize the cloaking region and display the reduced region over road network for location data requesters. With the toolkit, we demonstrate the practicality and effectiveness of the *ReverseCloak* approach.**

I. Introduction

The proliferation of low-cost GPS-enabled mobile devices and the ubiquitous deployment of wireless networks drive the rapid emergence of location-based service applications. 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. 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 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 within the region. A subject is said to be location k-anonymous if her location information is indistinguishable from that of $k-1$ other users in a spatial or spatio-temporal space. 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 to the exposed location.

Unlike conventional techniques [1], [2], [4], [7] that focus on single-level unidirectional location anonymization, *ReverseCloak* [5], [6] allows selective de-anonymization of the cloaking area when suitable access credentials are provided. It supports multilevel location privacy requirements that allow different users to infer different levels of information from the same exposed location information based on their access credentials and the access privilege levels entitled to them. Our proposed approaches transform the raw location of a mobile 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 even when the adversary has complete knowledge about the location perturbation algorithm used. In this paper, we design a demonstration toolkit for *ReverseCloak*, which consists of an 'Anonymizer' and a 'De-anonymizer' GUI. The location data owners can set and adjust the anonymization parameters through the 'Anonymizer' GUI and the results of multiple cloaking regions are visualized for them. The 'De-anonymizer' GUI allows the location data requesters to fetch the access keys and the de-anonymized cloaking region can also be visualized. With the toolkit, we demonstrate the practicality and effectiveness of proposed *ReverseCloak*.

II. Overview

In this section, we first present the concept of location anonymization. Then, we define the multilevel reversible location privacy problem.

A. Location anonymization

We consider the key privacy requirement arising in a road network namely *location k-anonymity* [3], 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*

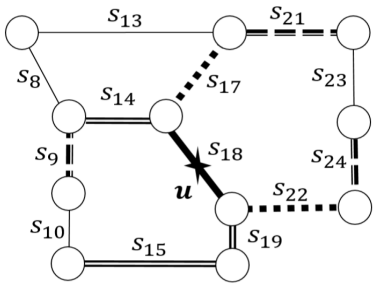

Fig. 1: Multilevel reversible location anonymization

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 [7], [9], 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).

B. Multilevel reversible location privacy

The focus in *ReverseCloak* 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 the 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. Figure 1 shows an example of a sub-graph bounded by spatial tolerance. It consists of a set of segments as the connections of adjacent junctions and a set of junctions as the intersections of segments. The segment s_{18} contains the actual user and 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.

III. ReverseCloak Approaches

In this section, we briefly present the two *ReverseCloak* algorithms proposed in [6], namely Reversible Global Expansion (RGE) and Reversible Pre-assignment-based Local Expansion (RPLE). In *ReverseCloak*, the cloaking region is formed by a set of road segments over the road network, which guarantees not only the location k-anonymization [2], but also the segment l-diversity privacy protection [9]. The anonymization and de-anonymization process are considered as a continuous selection and removal of road segments on the geographic road map respectively. To ensure that the process is reversible, each road segment is selected in a pseudo-random manner with an access key. As the location perturbation process is designed to support multilevel privacy protection, the location perturbation process employs the use of multiple access keys and each of them manages the selection and removal of road segments for one privacy level respectively. Each road segment on the map is linked to several other segments, which are located close to it. Once a road segment S is selected during anonymization, the next selected road segment is from one of its linked segments. With a certain access key, a fixed segment S' among them is deterministically selected. However, without the access key, all its linked segments would have the same probability to be selected, thus making the selection process pseudo-random and making it impossible to reverse without possessing the access key. Then, during the de-anonymization process (when a higher privileged data requester possesses the required access key), the newly selected segment S' maps to the previous road segment S using the access key. The algorithms checks which road segment is linked with S' to narrow down the options and whether segment S' can be deterministically selected with the access key if we assume a segment is S. A key challenge in *ReverseCloak* is the 'collision' issue that could happen in the de-anonymization process. That is, we may find multiple road segments that meet the conditions to be the candidate of the previously chosen road segment. *ReverseCloak* addresses this issue through two approaches. In RGE, for each road segment selection during anonymization, the links of previously selected segments are rebuilt on the fly to avoid collisions and optimize the selection based on the

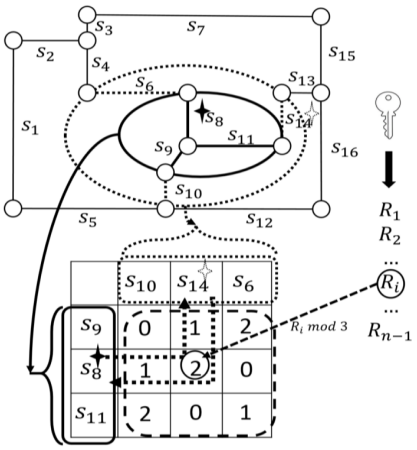

Fig. 2: Reversible global expansion

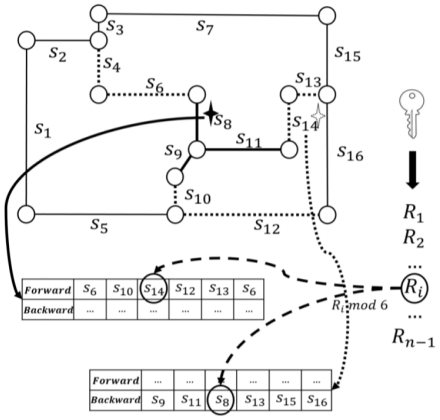

Fig. 3: Reversible pre-assignment-based local expansion

current state. In RPLE, prior to the anonymization process, all the road segments in the map are pre-assigned their links in a collision-free manner. RGE has larger anonymization runtime to build collision-free links on the fly but smaller memory requirement while RPLE has smaller anonymization runtime but requires larger memory space to store the collision-free links.

A. Reversible global expansion

We describe the process of RGE with an example presented in Figure 2. This example can be seen as either the i^{th} forward transition in the anonymization process or the $\{n-i\}^{th}$ backward transition in the de-anonymization process, where n is the length of the transition string. The solid segments within the solid ellipse form the current cloaking region, denoted by $CloakA = \{s_8, s_9, s_{11}\}$. The dotted segments within the dotted ellipse but outside the solid ellipse form the current cloaking region, denoted by $CanA = \{s_6, s_{10}, s_{14}\}$. Let us assume that the last added segment for the anonymization process is s_8, denoted by a solid star and the last removed segment for the de-anonymization process is s_{14}, denoted by a dotted star. For each addition or removal step, we assign transition values as IDs to all the possible transitions. These transition values are organized in a transition table. The structure of the table and the assignment of transition values are only related with the two sets. Therefore, for each addition step in anonymization process and its corresponding removal step in de-anonymization process, the same table can be generated. The table should contains $|CloakA|$ rows and $|CanA|$ columns. In Figure 2, s_8, s_9, s_{11} within $CloakA$ and s_6, s_{10}, s_{14} within $CanA$ 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. In the table, each transition value is assigned to one forward transition and its corresponding backward transition simultaneously so that these two transitions have the same ID. The transition value in table cell (i, j) associated with i^{th} row and j^{th} column is computed by $((i-1)+(j-1)) \ mod \ |CanA|$ so that there is no repeated transition value in each row and column if $CloakA \leq CanA$, thus no collisions [6].

Algorithm 1: RPLE(Pre-assignment)

Input : Road network graph G, original segment s_u, temporal key K_t, spatial key K_s, transition list length $\mathcal{T}$, user defined δ_k, δ_l, σ_t.
Output: forward transition table FT, backward transition table BT.

E = No. of segments in G;
Initially, the $E \times \mathcal{T}$ FT and BT are empty;
for *each segment s in G* **do**
  for $i = 1$ *to* E **do**
    Add next neighboring segment to the neighboring list NL;
  end
  for $i = 0$ *to* $E - 1$ **do**
    Potential segment $s_p = NL[i]$;
    Initialize $empFT$ and $empBT$ with size $\mathcal{T}$;
    for $j = 0$ *to* $\mathcal{T} - 1$ **do**
      if $FT[s][j]$ *is empty* **then**
        Put j to $empFT$;
      end
      if $BT[s_p][j]$ *is empty* **then**
        Put j to $empBT$;
      end
    end
    $emp = empFT \cap empBT$;
    if $emp \ != \varnothing$ **then**
      $selPosition = emp[0]$;
      $FT[s][selPosition] = s_p$;
      $BT[s_p][selPosition] = s$;
    end
  end
end

Then,the secret key is used to generate a sequence of pseudo-random numbers and each pseudo-random number controls the selection of one transition. The i^{th} pseudo-random number, denoted by R_i, is responsible for both the i^{th} forward transition and $\{n-i\}^{th}$ backward transition. Therefore, for the i^{th} forward transition and $\{n-i\}^{th}$ backward transition, the same value can be uniquely determined by the pseudo-random number and the current cloaking region. This value, called pick value, can be calculated by $p_i = R_i \ mod \ |CanA|$ and it is used to select the transition with the transition value same as the pick value. In Figure 2, if R_i is 5, p_i will be 2. For the anonymization process, since the last added segment is s_8, we find the transition value 2 in the 2^{nd} row is located in the cell $(2, 2)$, which indicates the forward transition from

s_8 to s_{14}. For the de-anonymization process, known the last removed segment s_{14}, the transition value 2 in the cell $(2, 2)$ here indicates the backward transition from s_{14} to s_8. In such a way, the forward and backward transitions are selected in a reversible manner, which guarantees the reversibility of the whole process.

B. Reversible pre-assignment-based local expansion

The reversible pre-assignment-based local expansion algorithm consists of two steps, namely pre-assignment and cloaking. Once a new graph is given, it first runs Algorithm 1 to calculate a forward transition list and a backward transition list for each segment in the graph, which are then used for anonymization and de-anonymization respectively. The assignment of transition values should be done carefully as collisions need to be avoided [6].

We present the process of RPLE with an example shown in Figure 3. In the example, once the forward transition sequence moves to the segment s_8, it searches the forward transition list of s_8 and selects one candidate s_{14} from the list as the next segment to go further. The index of s_{14} is calculated by $R_i \ mod \ 6$, where R_i is one pseudo-random number generated by the secret key and 6 is the length of the forward list. Since the selection is pseudo-randomly controlled by the secret key, once the backward transition sequence moves back to s_{14}, with the same key, it can select s_8 from backward transition list of s_{14}.

IV. Demonstration Toolkit

The demonstration toolkit designed for *ReverseCloak* is composed of an 'Anonymizer' GUI and a 'De-anonymizer' GUI. We show a screenshot of 'Anonymizer' GUI in Figure 4. The visualization of geographic maps, mobile users and cloaking regions is performed through GTMobiSim mobile trace generator [8], which is 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.

The location data owner first specifies the set of anonymization parameters, including the expected number of anonymity levels, the value of k for k-anonymization in each level, the spatial tolerance to restrict the allowed maximum area of cloaking region and the access key for each level. To make it easier for location data owners, the GUI provides the 'Auto key generation' function to automatically generate and manage access keys and the 'Default setting' function to memorize the default value of other parameters. By clicking the 'Anonymize' button, the 'Anonymizer' sends the parameters and access keys to a trusted anonymization server and visualizes the results as several colored regions on the map. If the location data owner feels dissatisfied with the results, she can adjust the parameters and re-anonymize the location data by clicking the 'Anonymize' button again. When the results are satisfactory,

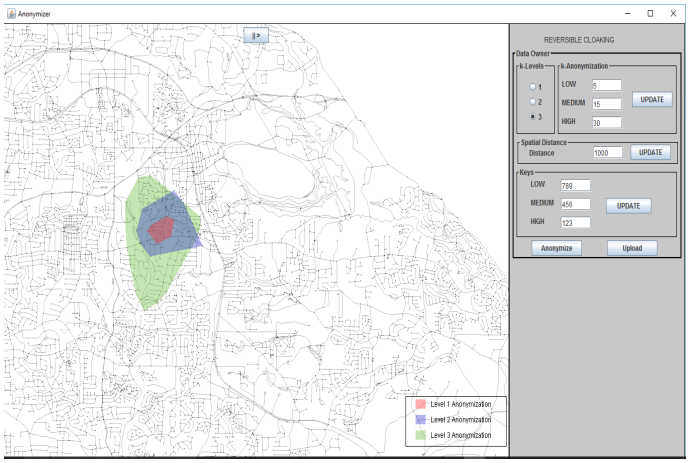

Fig. 4: Anonymizer screenshot

she can 'upload' the cloaking region to the LBS provider so that the LBS provider can serve the location data owner based on the privacy privileges and access rights. Later, the location data requesters can de-anonymize the cloaking region to get more location information about the location data owner through the 'De-anonymizer'. At the beginning, they can only see the largest cloaking region as the LBS provider. To reduce the size of the exposed cloaking region, they request the location data owners for access keys, which is managed locally by the 'Anonymizer'. The 'Anonymizer' maintains a personal access control profile, which decides the assignment of access keys based on trust degree and privileges of the location data requesters. After fetching the access keys, the location data requesters can run the de-anonymization algorithm and obtain the de-anonymized cloaking region as visualized in the 'De-anonymizer' GUI.

References

[1] 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.

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

[3] 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.

[4] 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.

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

[6] C. Li and B. Palanisamy. Reversecloak: Protecting multi-level location privacy over road networks. In *Proc. of 24th ACM International Conference on Information and Knowledge Management (CIKM)*, 2015.

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

[8] P. Pesti et al. Gtmobisim: A mobile trace generator for road networks. In *College of Computing, Georgia Inst. of Tech*, 2009.

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