Spatial Cloaking Algorithms for Location Privacy

Chi-Yin Chow
Department of Computer Science and Engineering
University of Minnesota, Minneapolis, MN

SYNONYMS

location cloaking; location blurring; location perturbation; location anonymization

DEFINITION

Spatial cloaking is a technique to blur a user's exact location into a spatial region in order to preserve her location privacy. The blurred spatial region must satisfy the user's specified privacy requirement. The most widely used privacy requirements are k-anonymity and minimum spatial area. The k-anonymity requirement guarantees that a user location is indistinguishable among k users. On the other hand, the minimum spatial area requirement guarantees that a user exact location must be blurred into a spatial region with an area of at least $\mathcal{A}$, such that the probability of the user being located in any point within the spatial region is $\frac{1}{\mathcal{A}}$. A user location must be blurred by a spatial cloaking algorithm either on the client side or a trusted third-party before it is submitted to a location-based database server.

HISTORICAL BACKGROUND

The emergence of the state-of-the-art location-detection devices, e.g., cellular phones, global positioning system (GPS) devices, and radio-frequency identification (RFID) chips, has resulted in a location-dependent information access paradigm, known as location-based services (LBS). In LBS, mobile users have the ability to issue snapshot or continuous queries to the location-based database server. Examples of snapshot queries include "*where is my nearest gas station*" and "*what are the restaurants within one mile of my location*", while examples of continuous queries include "*where is my nearest police car for the next one hour*" and "*continuously report the taxis within one mile of my car location*". To obtain the precise answer of these queries, the user has to continuously provide her exact location information to a database server. With untrustworthy database servers, an adversary may access sensitive information about individuals based on their location information and queries. For example, an adversary may identify a user's habits and interests by knowing the places she visits and the time of each visit.

The k-anonymity model [12, 13] has been widely used in maintaining privacy in databases [6, 8, 9, 10]. The main idea is to have each tuple in the table as k-anonymous, i.e., indistinguishable among other $k-1$ tuples. However, none of these techniques can be applied to preserve user privacy for LBS, mainly for the reason that these approaches guarantee the k-anonymity for a snapshot of the database. In LBS, the user location is continuously changing. Such dynamic behavior requires continuous maintenance of the k-anonymity model. In LBS, k-anonymity is a user specified privacy requirement which may have a different value for each user.

SCIENTIFIC FUNDAMENTALS

Spatial cloaking algorithms can be divided into two major types: k-anonymity spatial cloaking [3, 4, 5, 7, 11, 2] and uncertainty spatial cloaking [1]. k-anonymity spatial cloaking aims to blur user locations into spatial regions which satisfy the user's specified k-anonymity requirement, while uncertainty spatial cloaking aims to blur user locations into spatial regions which stratify the user's specified minimum spatial area requirement.

Adaptive interval cloaking. This approach assumes that all users have the same k-anonymity requirements [3]. For each user location update, the spatial space is recursively divided in a KD-tree-like format until a minimum k-anonymous subspace is found. Such a technique lacks scalability as it deals with each single movement of each user individually. Figure 1 depicts an example of the adaptive interval cloaking algorithm, in which the k-anonymity requirement is three. If the algorithm wants to cloak user A's location, the system space is first divided into four equal subspaces, $\langle(1,1),(2,2)\rangle$, $\langle(3,1),(4,2)\rangle$, $\langle(1,3),(2,4)\rangle$, and $\langle(3,3),(4,4)\rangle$. Since user A is located in the subspaces $\langle(1,1),(2,2)\rangle$ which contains at least k users, these subspaces are further divided into four equal subspaces, $\langle(1,1),(1,1)\rangle$, $\langle(2,1),(2,1)\rangle$, $\langle(1,2),(1,2)\rangle$, and $\langle(2,2),(2,2)\rangle$. However, the subspace containing user A does not have at least k users, so the minimum suitable subspace is $\langle(1,1),(2,2)\rangle$. Since there are three users, D, E, and F located in the cell (4,4), this cell is the cloaked spatial region of their locations.

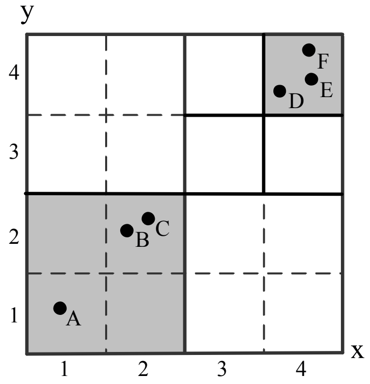
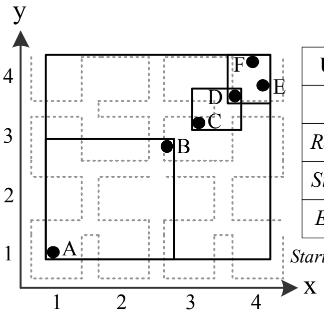


Figure 1: Adaptive interval cloaking ($k = 3$)

CliqueCloak. This algorithm assumes a different k-anonymity requirement for each user [3]. CliqueCloak constructs a graph and cloaks user locations when a set of users forms a clique in the graph. All users share the same cloaked spatial region which is a minimum bounding box covering them. Then, the cloaked spatial region is reported to a location-based database server as their locations. Users can also specify the maximum area of the cloaked region which is considered as a constraint on the clique graph, i.e., the cloaked spatial region cannot be larger than the user's specified maximum acceptable area.

***k*-area cloaking.** This scheme keeps suppressing a user location into a region which covers at least k-1 other sensitive areas, e.g., restaurants, hospital, and cinema, around the user's current sensitive area [5]. Thus, the user resident area is indistinguishable among k sensitive areas. This spatial cloaking algorithm is based on a map which is partitioned into zones, and each zone contains at least k sensitive areas. Thus, the continuous movement of users is just abstracted as moving between zones. Users can specify their own privacy requirements by generalizing personalized sensitivity maps.

Hilbert *k*-anonymizing spatial region (hilbASR). Here, users are grouped together into variant buckets based on the Hilbert ordering of user locations and their own k-anonymity requirements [7]. Using the dynamic hilbASR, the cloaked spatial regions of users A to F can be determined by using two equations $start(u)$ and $end(u)$ which are depicted in Figure 2, where $start(u)$ and $end(u)$ indicate the start and end rankings of a cloaked spatial region, respectively, u is a user identity, and the dotted line represents the Hilbert ordering.



Users	**A**	**B**	**C**	**D**	**E**	**F**
k_u	6	2	2	3	3	3
Rank(u)	0	1	2	3	4	5
Start(u)	0	0	2	3	3	3
End(u)	5	1	3	5	5	5

$Start(u) = Rank(u) - (Rank(u) \bmod k_u)$

$End(u) = Start(u) + k_u - 1$

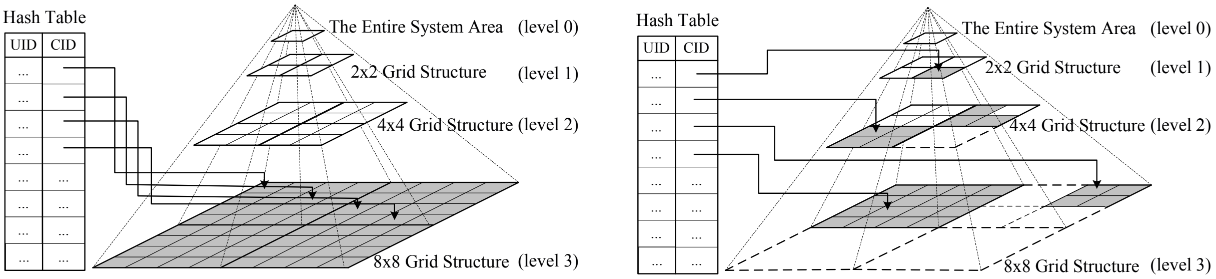
Figure 2: hilbASR

Nearest-neighbor *k*-anonymizing spatial region (nnASR). This is the randomized version of a k-nearest neighbor scheme [7]. For a user location u, the algorithm first determines a set S of k-nearest neighbors of u, including u. From S, the algorithm selects a random user u' and forms a new set S' that includes u' and the $k-1$ nearest neighbors of u'. Then, another new set S'' is formed by taking a union between S and S''. Finally, the required cloaked spatial region is the bounding rectangle or circle which cover all the users of S''.

Uncertainty. This approach proposes two uncertainty spatial cloaking schemes, *uncertainty region* and *coverage of sensitive area* [1]. The uncertainty region scheme simply blurs a user location into an uncertainty region at a particular time t, denoted as $U(t)$. The larger region size means a more strict privacy requirement. The coverage of sensitive area scheme is proposed for preserving the location privacy of users who are located in a sensitive area, e.g., hospital or home. The coverage of sensitive area for a user is defined as $Coverage = \frac{Area(sensitive\ area)}{Area(uncertainty\ region)}$.

The lower value of the coverage indicates a more strict privacy requirement.

Casper. Casper supports both the k-anonymity and minimum spatial area requirements [11]. System users can dynamically change their own privacy requirements at any instant. It proposes two grid-based pyramid structures to improve system scalability, *complete pyramid* and *incomplete pyramid.*


Hash Table
UID
CID
The Entire System Area (level 0)
2x2 Grid Structure (level 1)
4x4 Grid Structure (level 2)
8x8 Grid Structure (level 3)
Hash Table
UID
CID
The Entire System Area (level 0)
2x2 Grid Structure (level 1)
4x4 Grid Structure (level 2)
8x8 Grid Structure (level 3)

(a) Complete pyramid

(b) Incomplete pyramid

Figure 3: Grid-based pyramid data structures

Complete pyramid. Figure 3(a) depicts the complete pyramid data structure which hierarchically decomposes the spatial space into H levels where a level of height h has 4^h grid cells. The root of the pyramid is of height zero and has only one grid cell that covers the whole space. Each pyramid cell is represented as (cid, N), where *cid* is the cell identifier, and N is the number of mobile users within the cell boundaries. The pyramid structure is dynamically maintained to keep track of the current number of mobile users within each cell. In addition, the algorithm keeps track of a hash table that has one entry for each registered mobile user with the form $(uid, profile, cid)$, where *oid* is the mobile user identifier, $profile$ contains the user specified privacy requirement, and *cid* is the cell identifier in which the mobile user is located. *cid* is always in the lowest level of the pyramid (the shaded level in Figure 3(a)).

Incomplete pyramid. The main idea of the incomplete pyramid structure is that not all grid cells are appropriately maintained. The shaded cells in Figure 3(b) indicate the lowest level cells that are maintained.

Cloaking algorithm. Casper adopts a bottom-up cloaking algorithm which starts at a cell where the user is located at from the lowest maintained level, and then traverses up the pyramid structure until a cell satisfying the user specified privacy requirement is found. The resulting cell is used as the cloaked spatial region of the user location. In addition to the regular maintenance procedures as that of the basic location anonymizer, the adaptive location anonymizer is also responsible on maintaining the shape of the incomplete pyramid. Due to the highly dynamic environment, the shape of the incomplete pyramid may have frequent changes. Two main operations are identified to maintain the efficiency of the incomplete pyramid structure, namely, *cell splitting* and *cell merging.*

In the cell splitting operation, a cell *cid* at level i needs to be split into four cells at level $i+1$ if there is at least one user u in *cid* with a privacy profile that can be satisfied by some cell at level $i+1$. To maintain such criterion, Casper keeps track of the most relaxed user u_r for each cell. If a newly coming object u_{new} to the cell *cid* has more relaxed privacy requirement than u_r, the algorithm checks if splitting cell *cid* into four cells at level $i+1$ would result in having a new cell that satisfies the privacy requirements of u_{new}. If this is the case, the algorithm will split cell *cid* and distribute all its contents to the four new cells. However, if this is not the case, the algorithm just updates the information of u_r. In case one of the users leaves cell *cid*, the algorithm just update u_r if necessary.

In the cell merging operation, four cells at level i are merged into one cell at a higher level $i-1$ only if all the users in the level i cells have strict privacy requirements that cannot be satisfied within level i. To maintain this criterion, the algorithm keeps track of the most relaxed user u'_r for the four cells of level i together. If such user leaves these cells, the algorithm has to check upon all existing users and make sure that they still need cells at level i. If this is the case, the algorithm just updates the new information of u'_r. However, if there is no need for any cell at level i, the algorithm merges the four cells together into their parent cell. In the case of a new user entering cells at level i, the algorithm just updates the information of u'_r if necessary.

Peer-to-peer spatial cloaking. This algorithm also supports both the k-anonymity and minimum spatial area requirements [2]. The main idea is that before requesting any location-based service, the mobile user will form a group from her peers via single-hop and/or multi-hop communication. Then, the spatial cloaked area is computed as the region that covers the entire group of peers. Figure 4 gives an illustrative example of peer-to-peer spatial cloaking. The mobile user A wants to find her nearest gas station while being five anonymous, i.e., the user is indistinguishable among five users. Thus, the mobile user A has to look around and find four other peers to collaborate as a group. In this example, the four peers are *B*, *C*, *D*, and *E*. Then, the mobile user A cloaks her exact location into a spatial region that covers the entire group of mobile users A, B, C, D, and E. The mobile user A randomly selects one of the mobile users within the group as an *agent*. In the example given in Figure 4, the mobile user D is selected as an agent. Then, the mobile user A sends her query (i.e., what is the nearest gas station) along with her cloaked spatial region to the agent. The agent forwards the query to the location-based database server through a base station. Since the location-based database server processes the query based on the cloaked spatial region, it can only give a list of candidate answers that includes the actual answers and some false positives. After the agent receives the candidate answers, it forwards the candidate answers to the mobile user A. Finally, the mobile user A gets the actual answer by filtering out all the false positives.

KEY APPLICATIONS

Spatial cloaking techniques are mainly used to preserve location privacy, but they can be used in a variety of applications.

Location-based services

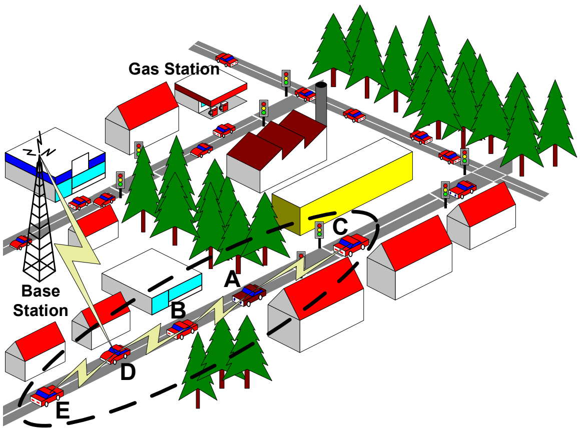


Figure 4: An example of peer-to-peer spatial cloaking

Spatial cloaking techniques have been widely adopted to blur user location information before it is submitted to the location-based database server, in order to preserve user location privacy in LBS.

Spatial database

Spatial cloaking techniques can be used to deal with some specific spatial queries. For example, given an object location, find the minimum area which covers the object and other $k-1$ objects.

Data mining

To perform data mining on spatial data, spatial cloaking techniques can be used to perturb individual location information into lower resolution to preserve their privacy.

Sensor-based monitoring system

Wireless sensor networks (WSNs) promise to have a vast significant academic and commercial impact by providing real-time and automatic data collection, monitoring applications and object positioning. Although sensor-based monitoring or positioning systems clearly offer convenience, the majority of people are not convinced to use such kinds of systems because of privacy issues. To overcome this problem, an in-network spatial cloaking algorithm can be used to blur user locations into spatial regions which satisfy user specified privacy requirements before location information is sent to a sink or base station.

FUTURE DIRECTIONS

Existing spatial cloaking algorithms have limited applicability as they are: (a) *applicable only for snapshot locations and queries.* As location-based environments are characterized by the *continuous* movements of mobile users, spatial cloaking techniques should allow continuous privacy preservation for both user locations and queries. Currently, existing spatial cloaking algorithms only support snapshot location and queries. (b) *not distinguishing between location and query privacy.* In many applications, mobile users do not mind that their exact location information is revealed, however, they would like to hide the fact that they issue some location-based queries as these queries may reveal their personal interests. So far, none of the existing spatial cloaking algorithms support such relaxed privacy notion where it is always assumed that users have to hide both their locations and the queries they issue. Examples of applications that call for such new relaxed notion of privacy include: (1) *Business operation.* A courier business company has to know the location of its employees to decide which employee is the nearest one to collect a certain package. However, the company is not allowed to keep track of the employees' behavior in terms of their location-based queries. Thus, company employees reveal their location information, but not their query information. (2) *Monitoring system.* Monitoring systems (e.g, transportation monitoring) rely on the accuracy of user locations to provide their valuable services. In order to convince users to participate in these systems, certain privacy guarantees should be imposed on their behavior through guaranteeing the privacy of their location-based queries although their locations will be revealed.

CROSS REFERENCES

1. Privacy Issues in Location-based Services
2. Privacy and Security Challenges in Geospatial Information Systems
3. Privacy Preserving of GPS Traces
4. Location Based Services: Practices and Products

RECOMMENDED READING

[1] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar (2006) Preserving User Location Privacy in Mobile Data Management Infrastructures. In *Proceedings of Privacy Enhancing Technology Workshop.*

[2] C.-Y. Chow, M. F. Mokbel, and X. Liu (2006) A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-based Services. In *Proceedings of the ACM Symposium on Advances in Geographic Information Systems, ACM GIS.*

[3] B. Gedik and L. Liu (2005) A Customizable k-Anonymity Model for Protecting Location Privacy. In *Proceeding of the International Conference on Distributed Computing Systems, ICDCS.*

[4] M. Gruteser and D. Grunwald (2003) Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services, MobiSys.*

[5] M. Gruteser and X. Liu (2004) Protecting Privacy in Continuous Location-Tracking Applications. *IEEE Security and Privacy*, 2(2):28–34.

[6] R. J. B. Jr. and R. Agrawal (2005) Data Privacy through Optimal k-Anonymization. In *Proceedings of the International Conference on Data Engineering, ICDE.*

[7] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias (2006) Preserving Anonymity in Location Based Services. *Technical Report TRB6/06, Department of Computer Science, National University of Singapore.*

[8] K. LeFevre, D. DeWitt, and R. Ramakrishnan (2006) Mondrian Multidimensional k-Anonymity. In *Proceedings of the International Conference on Data Engineering, ICDE.*

[9] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan (2005) Incognito: Efficient Full-Domain k-Anonymity. In *Proceedings of the ACM International Conference on Management of Data, SIGMOD.*

[10] A. Meyerson and R. Williams. (2004) On the Complexity of Optimal K-Anonymity. In *Proceedings of the ACM Symposium on Principles of Database Systems, PODS.*

[11] M. F. Mokbel, C.-Y. Chow, and W. G. Aref (2006) The New Casper: Query Procesing for Location Services without Compromising Privacy. In *Proceedings of the International Conference on Very Large Data Bases, VLDB.*

[12] L. Sweeney (2002) Achieving k-anonymity Privacy Protection using Generalization and Suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):571–588.

[13] L. Sweeney (2002) k-anonymity: A Model for Protecting Privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570.

Spatial Cloaking Algorithms for Location Privacy

Chi-Yin Chow
Department of Computer Science and Engineering
University of Minnesota, Minneapolis, MN 55414

SYNONYMS

location cloaking; location blurring; location perturbation; location anonymization

DEFINITION

Spatial cloaking is a technique to blur a user's exact location into a spatial region in order to preserve her location privacy. The blurred spatial region must satisfy the user's specified privacy requirement. The most widely used privacy requirements are k-anonymity and minimum spatial area. The k-anonymity requirement guarantees that a user location is indistinguishable among k users. On the other hand, the minimum spatial area requirement guarantees that a user exact location must be blurred into a spatial region with an area of at least $\mathcal{A}$, such that the probability of the user being located in any point within the spatial region is $\frac{1}{\mathcal{A}}$. A user location must be blurred by a spatial cloaking algorithm either on the client side or a trusted third-party before it is submitted to a location-based database server.

MAIN TEXT

This article surveys existing spatial cloaking techniques for preserving users' location privacy in location-based services (LBS) where users have to continuously report their locations to the database server in order to obtain the service. For example, a user asking about the nearest gas station has to report her exact location. With untrustworthy servers, reporting the location information may lead to several privacy threats. For example, an adversary may check a user's habit and interest by knowing the places she visits and the time of each visit. The key idea of a spatial cloaking algorithm is to perturb an exact user location into a spatial region that satisfies user specified privacy requirement, e.g., k-anonymity requirement guarantees that a user is indistinguishable among k users.

CROSS REFERENCES

1. Privacy Issues in Location-based Services
2. Privacy and Security Challenges in Geospatial Information Systems
3. Privacy Preserving of GPS Traces
4. Location Based Services: Practices and Products