Location Semantics Protection Based on Bayesian Inference

Zhengang Wu[1,2(✉)], Zhong Chen[1], Jiawei Zhu[1], Huiping Sun[1], and Zhi Guan[1]

[1] Institute of Software, School of EECS, MoE Key Lab of High Confidence Software Technologies (PKU), MoE Key Lab of Network and Software Security Assurance (PKU), Peking University (PKU), Beijing, China
wuzhengang@pku.edu.cn, {chen,sunhp}@ss.pku.edu.cn, zhujw.happy@163.com, guanzhi1980@gmail.com

[2] China Academy of Information and Communications Technology, Beijing, China

Abstract. In mobile Internet, popular Location-Based Services (LBSs) recommend Point-of-Interest (POI) data according to physical positions of smartphone users. However, untrusted LBS providers can violate location privacy by analyzing user requests semantically. Therefore, this paper aims at protecting user privacy in location-based applications by evaluating disclosure risks on sensitive location semantics. First, we introduce a novel method to model location semantics for user privacy using Bayesian inference and demonstrate details of computing the semantic privacy metric. Next, we design a cloaking region construction algorithm against the leakage of sensitive location semantics. Finally, a series of experiments evaluate this solution's performance to show its availability.

Keywords: Location privacy protection · Location semantics · Bayesian inference · Spatial cloaking

1 Introduction

Location Based Services (LBSs), as the representative of context-aware services, can recommend accurate and timely information according to user locations. The wide application of LBSs (such as Check-ins, Navigation, Maps and Mobile Social Networks) is benefit from the widespread availability of wireless networks and smart devices with built-in positioning modules. However, LBS gets involved in the problematic concern about location privacy because of its operating mechanism. Generally, in popular LBS-based applications and systems, real-time user locations from the LBS clients (e.g. some specific APPs installed in smartphones) as the vital contextual information need to be reported to the corresponding LBS providers in the on-demand manner. As a result, massive user locations are readily collected by potential adversaries via some untrusted servers and connection channels in mobile Internet.

Following privacy protection of relational databases, existing techniques of location privacy preservation have aimed at constructing the cloaking region

J. Li and Y. Sun (Eds.): WAIM 2015, LNCS 9098, pp. 297–308, 2015.
DOI: 10.1007/978-3-319-21042-1_24

under general privacy metrics such as k-anonymity and l-diversity to generalize exact user locations into custom extended spatial regions. Although effectively achieving a limited guarantee for location privacy, these techniques are vulnerable to Location Semantics Attack [1]. Intuitively, for a target user of LBS, the entire or major part of a k-anonymity cloaking region may be annotated with a similar sensitive semantic label such as Cancer Treatment Hospitals, and therefore adversaries can breach his privacy by learning his poor health status with a high probability.

Contributions. This solution involves three-fold contributions. First, the proposed approach LSRG models the process which extracts sensitive semantics form user requests and measures the degree of the semantics leakage. Second, this paper introduces a spatial cloaking method for preserving sensitive semantics on user locations. Third, its performance is demonstrated experimentally under different configurations by our adjusting crucial parameters.

Outline. The rest of the article is organized as follow. The 2nd section describes background. Section 3 and Section 4 shows two major parts of this work, extracting sensitive location semantics and constructing the cloaking region respectively. And Section 5 evaluates the performance of this solution through experiments. Section 6 reviews related works and the last section makes a summary.

2 Structure and Motivation

System Description. Following the popular three-tire architecture [2] for location privacy protection, our solution runs on this middle server in Figure 1. An LBS provider holds massive Point-of-Interest (POI) records which are meaningful location points over real maps. This middle server as a Trusted-Third-Party (TTP) is deployed between mobile clients and LBS servers to protect location privacy. First, for a user, this middle server extends exact locations into cloaking regions where all POI results are ready for forwarding. Second, this middle server refines and dispatchs POIs to corresponding users.

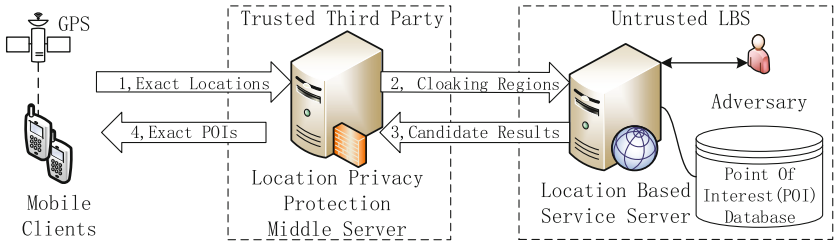


Fig. 1. System Model

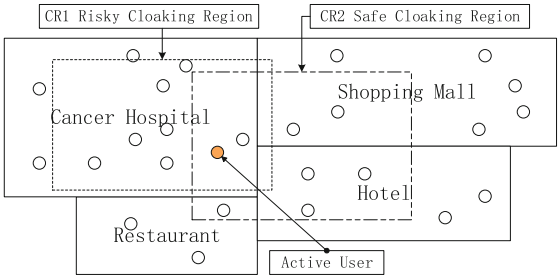


Fig. 2. Location Semantics Attack

Problem Setting and Motivation. Intuitively, the adversary analyzes published locations by mining semantic information for a target user. For example, a user Alice reports her current location coordinate *loc* to an untrusted LBS provider Malice in real activities. Next, Malice learns that a cancer hospital is located in the location point *loc* after querying public POI databases and map services such as Baidu Maps, Tencent Maps and Google Maps. Finally, Malice learns that Alice's health is poor with a high probability since some LBS requests are linked with meaningful labels.

The crux of Location Semantics Attack is that the adversary holds the public background knowledge on POI databases as same as users. In Figure 2, the cloaking region $CR1$ discloses that the active user is probably a cancer patient since all requests in $CR1$ are from cancer hospitals and the poor health status is one of his sensitive attributes. By contrast, these requests of $CR2$ are dispersed into various semantic regions such as hospitals, malls, restaurants and hotels and thus it is safer if the distribution of the adversary's guessing is uniform over these regions without additional information.

A recent work [1] has aimed at the semantic safety. The adversary may learn sensitive semantics on various locations. The majority of existing cloaking methods fail to capture the semantic risk. A cloaking region can still leak some risky semantics information in spite of satisfying the k-anonymity rule, since the major or entire part of the cloaking region which holds k users in a snapshot of LBS requests may be mapped into a risky semantic label such as infectious hospitals.

3 Evaluating Location Privacy

Generally, in client-side of LBS, when visiting LBS, a user submits a location request (U, L, T) where the users identifer U, the raw location L and the timestamp T. e.g. A request is (Alice,(116.42284,39.908063),'12:00'). In server-side of LBS, a POI entry is defined as a tuple (L, S, D). The raw location L is a pair (lng, lat) refers to the longitude lng and the latitude lat. The semantic label S refers to a meaningful brief name on this raw location L. The detail content D is a readable text to describe this raw location. e.g. a POI is ((116.42284,39.908063),

'bank', 'The Bank of China'). Thus, a location request discloses that this user may execute a personal activity about this semantic information 'bank'.

We model the causal relationship among raw locations (i.e. location coordinates), semantic labels (i.e. the meaning name of the raw location in real maps) and privacy risks (i.e. possible privacy disclosure events on special semantic labels) and naturally measure the belief of privacy risks using the probability of the privacy disclosure events on any raw locations and regions (the section/set of raw locations).

3.1 Modeling Location Semantics

As shown in Figure 3, this graphical model Location Semantics Risk Graph (in short LSRG) describes the privacy risk belief of a location request when the adversary eavesdrops this request after knowing semantic information of raw locations.

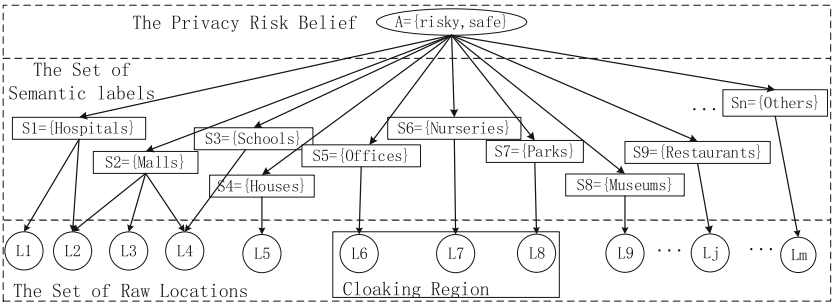


Fig. 3. Location Semantics Risk Graph

Definition 1. *Location Semantics Risk Graph is a three-tier directed acyclic graph* $\boldsymbol{G} = (V, E)$. *The node collection* V *falls into three mutually exclusive subsets, the set of raw locations* V_l, *the set of semantic labels* V_s *and the privacy risk belief* A. *A directed edge* $e =< a, b > \in E$ *refers to the dependency belief between its start point* a *and its end point* b *that is a conditional probability* $P(b|a) > 0$.

To simplify this model properly, we adopt an assumption that events of locations are independent of one another and so those of semantic labels are. i.e. There are no edges which connect two peer nodes of locations or semantic labels. Clearly, connections between locations and semantic labels have a many-to-many relationship, referring to edges from the location node set V_l to the semantics node set V_s. This is consistent with real-world experiences. A building on a location may be comprehensive with offices and shopping centers. Similarly, hospitals may be dispersed into different regions in a real city.

3.2 Inferring Privacy Risks

The binary class variable A that is the root node of LSRG denotes the probability event that $A = A_t$ if the adversary learns privacy information of users via location semantics inference and otherwise $A = A_f$. In brief, the event A_t means a risky request and A_f refers to a safe request form the perspective of user privacy.

The evidence variable O_{loc} is the observed location information (e.g. a region) which the middle server submits into an untrusted LBS server for forwarding the user request. O_{loc} actually is a cloaking region in the generalization-based location privacy protection schemes. O_{loc} is a subset of location nodes V_l. i.e. $O_{loc} \subseteq V_l$. Without cloaking, O_{loc} holds only one location that is the user's current location. But after cloaking locations, O_{loc} becomes a continuous spatial region which includes the current location.

Therefore, the posterior probability $P(A_t|O_{loc})$ refers to the conditional probability for the privacy disclosure event A_t of a request on the published location information O_{loc}. Naturally, $P(A_t|O_{loc})$ can be used to measure the privacy risk degree. The privacy risk $P(A_t|O_{loc})$ can be calculated by the Bayesian rule as follow.

$$PrivacyRisk \overset{def}{=} P(A_t|O_{loc}) = \frac{P(O_{loc}|A_t)P(A_t)}{P(O_{loc}|A_t)P(A_t) + P(O_{loc}|A_f)P(A_f)} \quad (1)$$

3.3 Estimating Parameters

Computing the posterior belief needs to obtain three prior beliefs $P(O_{loc}|A_t)$, $P(O_{loc}|A_f)$ and $P(A_t)$ which are estimated by given samples and the Maximum Likelihood Estimation (MLE). For all cloaked location-based requests, each published location information $O_{loc} \subseteq V_l$ can be decomposed into a series of basic locations, relying on specific methods of clustering or partitioning spatial data for original location coordinates. Basic locations in Figure 3 are the m-order collection of leaf nodes, $V_l = \{L_1, \cdots, L_m\}$. Therefore, for $A \in \{A_t, A_f\}$, $P(O_{loc}|A) = \sum_{l \in O_{loc}} P(l|A)$ where $P(l|A)$ is the condition probability for a basic location $l \in V_l$. Naturally, repetitive computation steps can be reduced using precalculated beliefs $P(l|A)$ of all basic locations.

The Prior Belief $P(O_{loc}|A_t)$. Without loss of generality, the adversary observes a spatial region O_{loc} which refers to a set of locations as the evidence. Specifically, throughout a middleware for location privacy protection, the cloaking region R_{ca} is actually the observed region. i.e. $O_{loc} = R_{ca}$. This belief can be calculated as follow.

$$P(O_{loc}|A_t) = \sum_{l \in O_{loc}} \sum_{s \in pa(l)} [P(l|s)P(s|A_t)] \quad (2)$$

For an untrusted LBS, the adversary's ability involves two-fold factors. The first factor is the location semantics knowledge. The adversary can access public

POI databases and thus obtains corresponding semantics information on locations. Next, the semantics risk knowledge is the other factor. The adversary's intention relies on semantics labels for learning sensitive attributes of a target user and thus different location semantics implies different risk levels for location privacy. Based on the intuitive understanding, we can estimate this belief $P(O_{loc}|A_t) = \sum_{l \in O_{loc}} P(l|A_t)$ under the LSRG model, after knowing these two factors which express as two condition probabilities $P(s|A_t)$ and $P(l|s)$ respectively where $s \in pa(l) \subseteq V_s$ and $l \in V_l$. For simplicity, $pa(l)$ denotes the set of parent nodes of the node l in LSRG.

The location semantics knowledge can be computed using $\widehat{P}(l|s) = \frac{F(l,s)}{F(s)}$ where the function $F(x)$ is the metric of the event x. We assume that for a semantics label s the adversary's attack is the spatial uniform distribution over the region of this semantics s and so the metric function $F(x)$ should be the area of the region meeting the event (l, s) or (s). i.e. $P(l|s) = \frac{Area(l,s)}{Area(s)}$. However, computation of exact areas of massive irregular regions over a real map will generally consume intensive resources since popular online map services fail to provide related data directly. As a practical alternate, we can employ the number of POI entries in the region meeting specific semantics conditions. i.e. $P(l|s) = \frac{count(l,s)}{count(s)}$. Given the POI database which the untrusted LBS holds, the function $count(s)$ counts up the number of POI entries whose semantics label is s and the function $count(l, s)$ refers to the number of POI entries whose semantics label is s and meanwhile whose location coordinates fall in the spatial cell annotated by l. For convenience, we use the pyramid structure [2] based on Quad-Tree to index POI entries in the 4^n grid and in fact the location semantics knowledge reflects the inherent feature of POI databases over real maps.

The semantics risk knowledge can be estimated using the frequency of risky events which are annotated by the semantics label $s \in V_s$ over all risky events. i.e. $\widehat{P}(s|A_t) = \frac{count(s,A_t)}{count(A_t)}$. Given a sample dataset of risky events, we can make a statistic analysis on the frequency of risky events grouped by semantic categories such as 'hospitals', 'offices' and so on. $count(s, A_t)$ adds up the number of risky events with the semantics label s and $count(A_t)$ is the total number of all risky events. e.g. 50 risky events on 'hospitals' exist in 100 risky events and thus we can learn the belief $P(s = hospitals|A_t) = 0.5$ on the semantics information 'hospatials' based on this sample. Note that, all events of a sample are classified into defined catalogs (semantics labels). i.e. Each event relates to only one label, and for all semantics labels $\sum_{s \in V_s} P(s|A_t) = 1$.

The Prior Belief $P(O_{loc}|A_f)$. The probability of safe requests on the observed region O_{loc} denotes this prior belief $P(O_{loc}|A_f)$. By collecting requests via a safe LBS, this MLE is obtained by Equation 3. Since O_{loc} is a set of basic locations on this partitioned maps, the probability of each basic location $P(l|A_f) = \frac{count(l,A_f)}{count(A_f)}$ can be calculated from the safe request sample. $count(l, A_f)$ is the number of safe requests on this basic location l and $count(A_f)$ is the total number of all requests on the safe sample dataset.

$$\widehat{P}(O_{loc}|A_f) = \frac{count(O_{loc}, A_f)}{count(A_f)} = \sum_{l \in O_{loc}} \frac{count(l, A_f)}{count(A_f)} \tag{3}$$

The Prior Belief $P(A_t)$. Intuitively, the prior belief $P(A_t)$ implies the trust status of the entire LBS system including related network connections. Given an event sample of accessing an LBS, the MLE of $P(A_t)$ can express as the frequency of past request events in Equation 4 where the class variable A_t refers to events of risky requests. By the sample, $count(A_t)$ is the number of violated request events where user sensitive information is disclosed and $count(A)$ denotes the number of all events on both risky and safe requests simply. Note that, $P(A_t) + P(A_f) = 1$.

$$\widehat{P}(A_t) = \frac{count(A_t)}{count(A)} \tag{4}$$

4 Cloaking Published Locations

This section describes a cloaking region construction method to protect location semantics. Based on the aforementioned privacy risk evaluation method, we design Algorithm 1 which can recursively construct a (k, l, t)-Secure Cloaking Region (for short, (k, l, t)-SCR) to meet three privacy requirements. First, k-anonymity[2,3] means that the cloaking region holds k different users at least. Second, l-diversity[2,4] means that the cloaking region covers l different locations (or spatial cells) at least. Third, t-safety ensures that the semantics safety of the cloaking region is larger than a threshold t. This can be defined as follow.

Definition 2. *A cloaking region O_{loc} meets t-safety if and only if its semantics safety $P(A_f|O_{loc}) = 1 - P(A_t|O_{loc}) \geq t$.*

Definition 3. *(k, l, t)-Secure Cloaking Region is a cloaking region which satisfies k-anonymity, l-diversity and t-safety.*

In the pyramid structure[2], a location point falls into a rectangular region linked with a node of the Quad-Tree. Each non-root node has only one parent node. Importantly, each non-leaf node has four child nodes like a cross and thus the non-root node has the only vertical or horizontal neighbor node in the four quadrants of the cross. For convenience, two notations $VNode$ and $HNode$ refer to the vertical neighbor and the horizontal one of $Node$ respectively.

The bottom-up Algorithm 1 can recursively create a continuous region from leaf to root along the Quad-Tree by gradually merging neighbors and check whether these candidate regions satisfy the pre-defined privacy profile. First, for k-anonymity, the region's request amount defines the anonymity degree. Here, $Node.N$ is the request amount in the region referred by $Node$. Second, for l-diversity, the region's area denoted by $Area(Node)$ measures the diversity degree. We employ the number of cells in the region $Node$ to count $Area(Node)$ since all cells occupy the same area as the basic unit of the Quad-Tree partitioned

Algorithm 1. SCR$(k, l, t, Node)$

```
1  if Area(Node) ≥ MaxArea then                          /* Restrict oversize. */
2      return CR ← ∅;                                    /* Cloaking fails. */
3  Risk ← P(A_t|O_loc = {Node}); Safety(Node) = 1 − Risk;
4  if Node.N ≥ k ∧ Area(Node) ≥ l ∧ Safty(Node) ≥ t then
5      return CR ← {Node};
6  else
7      (VNode, HNode) ← GetNeighbors(Node);
8      VN ← VNode.N + Node.N; HN ← HNode.N + Node.N;
9      if (VN ≥ k ∨ HN ≥ k) ∧ ((2 * Area(Node)) ≥ l) then
10         if (VN ≥ k ∧ HN ≥ k ∧ HN ≤ VN) ∨ VN < k then
11             CR ← {HNode, Node};
12         else
13             CR ← {VNode, Node};
14         Risk ← P(A_t|O_loc = CR); Safety(CR) = 1 − Risk;
15         if Safety(CR) ≥ t then                        /* Check safety. */
16             return CR;                                /* Return one CR */
17         else                                  /* Search its parent recursively. */
18             CR ← SCR(k, l, t, Node.ParentNode);
19     else                                      /* Search its parent recursively. */
20         CR ← SCR(k, l, t, Node.ParentNode);
```

maps. Finally, for t-safety about location semantics, the function $Safety(Node)$ refers to $1 - Pr(A_t|O_{loc} = Node)$.

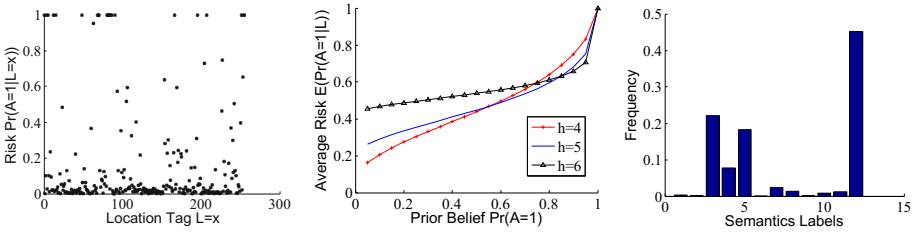
In addition, the computation of $P(A_t|O_{loc})$ can be divided into two phases for reducing its time cost since a region O_{loc} are divided into a set of distinct spatial cells and for $A \in \{A_t, A_f\}$, $P(O_{loc}|A) = \sum_{l \in O_{loc}} P(l|A)$. First, the off-line phase can calculate these prior beliefs $P(l|A)$ for each basic cell $l \in V_l$. Second, Algorithm 1 can obtain $P(A_t|O_{loc})$ with linear complexity $O(m)$ where m is the number of cells in the region O_{loc}, using the prepared prior beliefs from the off-line phase. This way can help to achieve the high processing performance on spatial cloaking and POI forwarding in the real-time LBS environment.

5 Experiments

We implement the proposed solution using JAVA and run it in the experiment platform which is a laptop with a quad-core 2.4Ghz Intel i7 CPU and 16G RAM. The experimental dataset from MNTG[5] holds trajectory data of about 1000 users who move along the real road networks of Beijing on 20 continuous timestamps (from 0 to 19) . All raw locations lie in a rectangle region about $67km^2$ and are indexed by the n-height full Quad-Tree structure [2] where 4^n leafs divide the region into 4^n cells which refer to atomic regions and the default height is 4.

To build the adversary's background knowledge, we extract the real POI dataset from a popular electronic map web site 'map.baidu.com', including about 8700 POI entries in this experimental region. Next, we explore 12 chosen semantics labels which are s_1=hospitals, s_2=nurseries, s_3=restaurant, s_4=hotels, s_5=bank, s_6=malls, s_7=offices, s_8=houses, s_9=school, s_{10}=museums, s_{11}=parks and s_{12}=others. The default privacy profile (k, l, t) is $(10, 2, 0.9)$ and additionally the default value of the total risk belief $Pr(A_t)$ is set to 0.05.

5.1 Evaluating Privacy Risks on Location Semantics



(a) Distribution of Risks (b) Risks on Grids (c) Semantics Distribution

Fig. 4. Risk Evaluation for Location Privacy

Experiments in Figure 4 explore the degree of the location privacy risks that mobile users leak their current locations to untrusted LBS servers on partitioned maps annotated by semantic information. Each location point refers to a POI record labeled by a meaningful string according to public real maps and POI databases, and therefore the leakage of location coordinates via a request leads to the leakage of the corresponding meaningful labels.

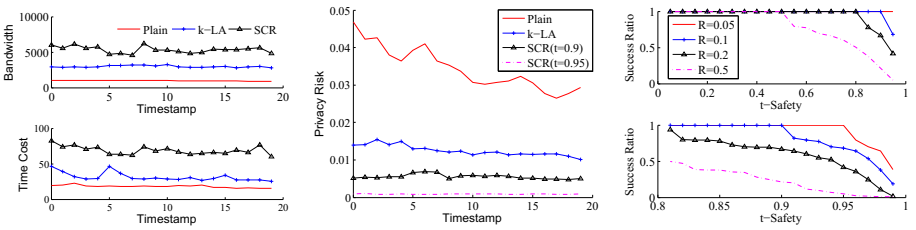
Figure 4(a) displays a distribution of the belief $Pr(A_t|O_{loc} = x)$ on $4^4 = 256$ cells in the 4-height full Quad-Tree structure. Intuitively, each location-based request involves a piece of risky semantic information. Generally, the majority of these cells have low risks for the perspective of user privacy. e.g. Mobile users visit in locations of public places like offices and malls. And there are some high-sensitive cells which refers to restricted regions such as hospitals and military areas. This distribution relies on two factors: First, the inherent semantic feature of a POI database or a real map expresses as the belief $Pr(O_{loc}|s \in V_s)$; Second, the adversary's intention refers to $Pr(s \in V_s|A_t)$.

Figure 4(b) shows the curves of average risk by adjusting the prior belief $P(A_t)$. Under the distinct values of $P(A_t) \in [0, 1]$, we count up the mathematical expectation (Average Risk) of $Pr(A_t|O_{loc} = x)$ for all cells. Three curves are under different Quad-Tree partitioning [2] configurations whose heights are 4, 5 and 6 respectively. More accurate location information (i.e. more finer granularity and higher Quad-Tree) leads to more privacy leakages and a higher

privacy risk level. There is a positive correlation between the risk at a cell $Pr(A_t|O_{loc} = x)$ and the prior belief $P(A_t)$ which refers to the estimated total risk. Specially, when $P(A_t)$ approximates 1, the privacy disclose event on any location is inevitable with the probability that is close to 1.

Figure 4(c) demonstrates the distribution of these 12 semantics labels over the POI dataset. The majority of POI entries have low risks for user privacy and by contrast POI entries with two high risk semantics labels, s_1=hospitals and s_2=kids, take over 0.32% and 0.18% respectively. Clearly, high risky POI entries are sparse in a real-world maps. As a result, (k, l, t)-SCR can be constructed with an accepted success ratio to satisfy its custom privacy conditions.

5.2 Cloaking Published Locations



(a) Complexity (b) Semantics Risk (c) Privacy Profile

Fig. 5. Performance of Location Anonymization

By comparing existing location cloaking methods, experiments in Figure 5 demonstrates that the proposed location cloaking method is feasible and practical. First, the label 'Plain' means the straightway method that the location cloaking server is only a simple proxy to forward requests from mobile clients to LBS servers by replacing an exact location with a spatial cell. Next, the label 'k-LA' is the popular location k-anonymity method (NewCasper[2]) which generalizes an extended rectangular region under the k-anonymity metric. Finally, the label 'SCR' represents our solution that can guard against the Location Semantics Attack.

Figure 5(a) compares complexity on time and communication. The straightway method 'Plain' has the lowest cost on both execution time and downloaded data amount. And SCR possesses slightly more costs for controlling privacy risks under location semantics than location k-anonymity. Thus additional costs of SCR are still affordable.

As shown in Figure 5(b), the proposed method can control privacy risks on location semantics by checking $Pr(A_t|O_{loc})$ of all cloaking regions. The straightway method labeled by 'Plain' has high risks on location semantics disclosure. Next, location k-anonymity and SCR hold similar performance of privacy preservation but SCR builds safer cloaking regions than other two methods. On two

SCR curves of t=0.9 and t=0.95, for a higher safety threshold t, this method reduces privacy risks by generalizing exact locations into larger regions.

Figure 5(c) shows effects of privacy profiles by comparing success ratios of building SCRs. For specific values of the total risk belief $P(A_t)$ denoted by R=0.05,R=0.1,R=0.2, and R=0.5, the ratios drop significantly after horizontal lines which refer to 100% cloaking success, when the required safety thresholds t increasing gradually. As a result, visiting high-risk LBSs especially, we have to trade off the required safety and the cloaking success ratio.

6 Related Works

When publishing a dataset where each object holds generally one identifier and multiple attributes, the adversary can re-identify objects because of the possible uniqueness of attribute values in spite of removing identifiers. For this, k-anonymity[3][6] ensures that at least k objects are indistinguishable in an anonymity set. l-Diversity[4] requires that the number of different attributes which each object in an anonymity set associates with is more than at l. t-Closeness [7] guarantees that an anonymity set is statistically similar under the probability metric such as Earth-Mover-Distance.

Previous techniques of location privacy protection employed two basic ideas, cryptography and anonymization. Wernke et al.[8] survey research works on attacking and protecting location privacy. Cryptography-based methods[9][10] can give strong privacy assurance but need extremely intensive resources. By comparison, location anonymization (e.g. spatial cloaking) can achieve enough privacy assurance under appropriate resources.

Location k-anonymity[11] from Gruteser et al. generalizes an exact location into a region which holds at least k requests, extended from k-anonymity. Plenty of solutions such as CliqueCloak[12], HilbertCloak[13] and NewCasper[2] have adopted location k-anonymity in the last decade. Following t-closeness[7], Lee et al. introduce a location anonymization method which constructs θ-Secure Cloaking Area[1] after extracting semantics information from staying duration. Shokri et al. introduced a Markov Chain based approach[14] to measure location privacy.

Additionally, location semantics mining is a hot topic in mobile Internet. Parent et al.[15] review various methods which model and mine semantics information on trajectory data.

7 Conclusion

This paper investigated privacy protection against Location Semantics Attacks. To solve this problematic issue, we introduce the Location Semantics Risk Graph model to evaluate privacy risks about the dependence of location coordinates and sensitive semantics information, using Bayesian inference. And next we proposed a spatial cloaking algorithm under this model. Finally, experiments demonstrate that this solution can achieve a better privacy guarantee than existing schemes.

Acknowledgments. This work is partially supported by the HGJ National Significant Science and Technology Projects under Grant No. 2012ZX01039-004-009, Key Lab of Information Network Security, Ministry of Public Security under Grant No.C11606, the National Natural Science Foundation of China under Grant No. 61170263.

References

1. Lee, B., Oh, J., Yu, H., Kim, J.: Protecting location privacy using location semantics. In: Apté, C., Ghosh, J., Smyth, P. (eds.) KDD, pp. 1289–1297. ACM (2011)
2. Mokbel, M.F., Chow, C.Y., Aref, W.G.: The new casper: query processing for location services without compromising privacy. In: Dayal, U., Whang, K.Y., Lomet, D.B., Alonso, G., Lohman, G.M., Kersten, M.L., Cha, S.K., Kim, Y.K. (eds.) VLDB, pp. 763–774. ACM (2006)
3. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **10**(5), 557–570 (2002)
4. Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M.: l-diversity: privacy beyond k-anonymity. In: Liu, L., Reuter, A., Whang, K.Y., Zhang, J. (eds.) ICDE, p. 24. IEEE Computer Society (2006)
5. Mokbel, M.F., Alarabi, L., Bao, J., Eldawy, A., Magdy, A., Sarwat, M., Waytas, E., Yackel, S.: MNTG: an extensible web-based traffic generator. In: Nascimento, M.A., Sellis, T., Cheng, R., Sander, J., Zheng, Y., Kriegel, H.-P., Renz, M., Sengstock, C. (eds.) SSTD 2013. LNCS, vol. 8098, pp. 38–55. Springer, Heidelberg (2013)
6. Sweeney, L.: Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **10**(5), 571–588 (2002)
7. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity. In: Chirkova, R., Dogac, A., Özsu, M.T., Sellis, T.K. (eds.) ICDE, pp. 106–115. IEEE (2007)
8. Wernke, M., Skvortsov, P., Dürr, F., Rothermel, K.: A classification of location privacy attacks and approaches. Personal and Ubiquitous Computing **18**(1), 163–175 (2014)
9. Papadopoulos, S., Bakiras, S., Papadias, D.: Nearest neighbor search with strong location privacy. PVLDB **3**(1), 619–629 (2010)
10. Paulet, R., Kaosar, M.G., Yi, X., Bertino, E.: Privacy-preserving and content-protecting location based queries. IEEE Trans. Knowl. Data Eng. **26**(5), 1200–1210 (2014)
11. Gruteser, M., Grunwald, D.: Anonymous usage of location-based services through spatial and temporal cloaking. In: Siewiorek, D.P. (ed.) MobiSys. USENIX (2003)
12. Gedik, B., Liu, L.: Protecting location privacy with personalized k-anonymity: Architecture and algorithms. IEEE Trans. Mob. Comput. **7**(1), 1–18 (2008)
13. Kalnis, P., Ghinita, G., Mouratidis, K., Papadias, D.: Preventing location-based identity inference in anonymous spatial queries. IEEE Trans. Knowl. Data Eng. **19**(12), 1719–1733 (2007)
14. Shokri, R., Theodorakopoulos, G., Boudec, J.Y.L., Hubaux, J.P.: Quantifying location privacy. In: IEEE Symposium on Security and Privacy, pp. 247–262. IEEE Computer Society (2011)
15. Parent, C., Spaccapietra, S., Renso, C., Andrienko, G.L., Andrienko, N.V., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., de Macêdo, J.A.F., Pelekis, N., Theodoridis, Y., Yan, Z.: Semantic trajectories modeling and analysis. ACM Comput. Surv. **45**(4), 42 (2013)