KLAP for Real-World Protection of Location Privacy

Abdur R. Shahid, SCIS, FIU, Miami,FL, ashah044@fiu.edu
Niki Pissinou, SCIS, FIU, Miami, FL, pissinou@fiu.edu
S.S. Iyengar, SCIS, FIU, Miami, FL, iyengar@cis.fiu.edu
Jerry Miller, SCIS, FIU, Miami, FL, millej@fiu.edu
Ziqian Ding, Math, UM, Coral Gables, FL, z.ding@math.miami.edu
Teresita Lemus, SAS, WC, Doral, FL, tlemus@dadeschools.net

***Abstract*—In Location-Based Services (LBS), users are required to disclose their precise location information to query a service provider. An untrusted service provider can abuse those queries to infer sensitive information on a user through spatio-temporal and historical data analyses. Depicting the drawbacks of existing privacy-preserving approaches in LBS, we propose a user-centric obfuscation approach, called KLAP, based on the three fundamental obfuscation requirements: k number of locations, l-diversity, and privacy area preservation. Considering user's sensitivity to different locations and utilizing Real-Time Traffic Information (RTTI), KLAP generates a convex Concealing Region (CR) to hide user's location such that the locations, forming the CR, resemble similar sensitivity and are resilient against a wide range of inferences in spatio-temporal domain. For the first time, a novel CR pruning technique is proposed to significantly improve the delay between successive CR submissions. We carry out an experiment with a real dataset to show its effectiveness for sporadic, frequent, and continuous service use cases.**

I. Introduction

Location-Based Services (LBS) have become an integral part of our life. A typical example of LBS use is the point of interest (POI) query to obtain nearby restaurants, movies, and other consumer information on a mobile device. This interaction with the LBS poses serious threat to users' privacy, as an untrusted service provider can infer sensitive information on a user by analyzing the location-centric queries [1].

The privacy-preserving approaches proposed to protect privacy in LBS, are either centralized (require a third party) or are user-centric (work on user's device). One popular user-centric approach is *obfuscation* [2], which replaces user's location with a larger region, called Concealing Region(CR), in a LBS query. Despite a substantial amount of research, existing obfuscation approaches exhibit the following limitations. First, if a CR is generated at random, then a service provider can exclude some locations for having a low probability in being the user's location(Fig.1(a)). Second, if the time difference $(T_2 - T_0)$ between multiple queries is small enough such that the service provider can guess with high confidence that all the CRs (CR_0, CR_1, and CR_2) were actually generated from the same location, then it can find a smaller (shaded) region containing the user's location (Fig.1(b)). Third, the maximum velocity-based free space mobility-centric approaches [2] leak privacy against inferences using Real-Time Traffic Information (RTTI). For example in Fig.1(c) using RTTI, a service provider can exclude some locations from CR_0, from which it is not possible to reach any location of CR_1 in $(T_1 - T_0)$ time. Similarly, it can exclude some locations from CR_1 which are not reachable from any location of CR_0 in $(T_1 - T_0)$ time. Fourth, existing approaches cannot protect privacy against a long-term obfuscated location tracking attack. In Fig.1(d), although CR_1, CR_{11}, CR_{53}, and CR_{60} were generated on different days, through careful observation the service provider can find that all of them actually were generated for a frequently visited location and can apply a region intersection attack to get a smaller region containing that location.

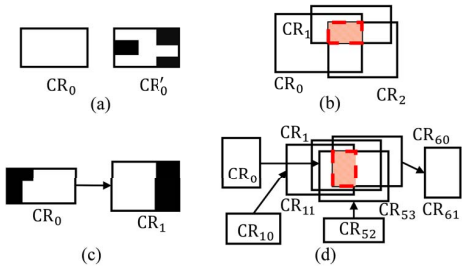

Fig. 1. Location inference attacks performed by an untrusted service provider: (a) probability distribution and personal context linking attack, (b) region intersection attack, (c) real-time traffic information (RTTI) based maximum-movement boundary attack, and (d) long-term obfuscated location tracking attack.

Contribution: Against this backdrop, we propose a user-centric obfuscation approach, called KLAP, to protect privacy against spatio-temporal and historical data analysis based inferences in LBS. We also introduce a novel CR pruning technique for consecutive queries to significantly reduce the delay compared to delay-based approach.

II. KLAP: The Proposed Approach

We consider a LBS with a set of mobile users U and a set of locations L(Fig.2). A user $u \in U$'s sensitivity to a location $l_i \in L$ is defined as $\mathcal{S}_i^u = \frac{q_i \times q_i^u}{\sum_i (q_i \times q_i^u)}$, where q_i^u and q_i are the total number of historical queries submitted by u and U from location l_i, respectively. In KLAP, users are required to mark their frequently visited locations(e.g. home, workplace, and so on). To conceal his real location

This work was supported by Army Research Office-W911NF1510572 and NSF RET SITE-1407067.

978-1-5386-7374-4/18/$31.00 ©2018 IEEE
DOI 10.1109/SERVICES.2018.00022

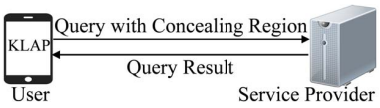

Fig. 2. System model of KLAP

with a CR, a user sets his personalized privacy setting in terms of $\langle k, l, \mathcal{A} \rangle$, referring to the required number of related locations, number of the types of the related locations, and area of the CR. A location is related to a user's real location $\mathcal{O}$, if it has similar sensitivity and is selected to form a CR. Let us consider W is a set of related locations whose convex hull constructs a CR such that, $|W| \geq k, |\text{type}(W)| \geq l, (\text{Area of the CR}) \geq \mathcal{A}$. Then the privacy level of that CR is, $\text{E}_{\text{CR}} = -\sum_i \overline{\mathcal{S}}_i \log_2 \overline{\mathcal{S}}_i$; where, $\overline{\mathcal{S}}_i = \frac{\mathcal{S}_i^u}{\sum_i \mathcal{S}_i^u}; \forall i \in W$. A CR submission is said to be delayed if the time difference between query generation and privacy-preserving query submission to the service provider is ($> \delta t$), where δt is a delay tolerant threshold. Let us assume a user submitted CR_A to conceal location $\mathcal{O}_\text{A}$ at T_A timestamp. Then he moved to a new location $\mathcal{O}_\text{B}$ in $\text{T}_\text{diff} = (\text{T}_\text{B} - \text{T}_\text{A})$ time and the RTTI-based minimum required time to submit CR_B, satisfying privacy requirements, to the LBS is T'_B. If $\text{T}'_\text{B} > \text{T}_\text{diff} + \delta t$, then we call it *delaying* the CR submission for $delay = \text{T}'_\text{B} - (\text{T}_\text{diff} + \delta t)$ amount of time. Based on privacy level and delay, we formulate our problem as a multi-objective optimization problem:

$$\text{maximize } \{\text{privacy level } \text{E}_\text{CR}, \frac{1}{e^{delay}}\} \quad (1)$$

The solution to this problem is discussed in algorithm 1.

Algorithm 1 KLAP algorithm

Require: Previous CR_A submitted at time T_A for $\mathcal{O}_\text{A}$, current time T_B, current location $\mathcal{O}_\text{B}$, set of all previously submitted CRs for frequently visited locations $(\bigcup_x \text{CR}_x)$, δt, λ, user's preference for all the locations $\mathcal{S}$, $k, l, \mathcal{A}$

Ensure: Final concealing region CR

1 **if** $\mathcal{O}_\text{B}$ is in CR_A **then** return with CR_A

2 **if** $\mathcal{O}_\text{B}$ is in CR_B, where $\text{CR}_\text{B} \in (\bigcup_x \text{CR}_x)$, **then** compute $delay$ between CR_A and CR_B. **if** $delay > 0$ and CR_A was not generated for a frequent location, **then** prune CR_A by excluding the related locations of it which are causing the delay and compute new $delay$ between pruned CR_A and CR_B. Finally, wait $delay$ amount of time and return with CR_B.

3 Select k number of random locations, including $\mathcal{O}_\text{B}$, as seeds from the region (disk centered at $\mathcal{O}_\text{B}$ with radius R $\setminus\{\text{CR}_\text{A} \cup (\bigcup_x \text{CR}_x)\}$). For each seed, $seed_i$, generate CR_i with the locations, having same sensitivity as $\mathcal{O}_\text{B}$, from the region (disk centered at $seed_i$ with radius R $\setminus\{\text{CR}_\text{A} \cup (\bigcup_x \text{CR}_x)\}$). **if** $delay > 0$ for CR_i and CR_B **then** prune CR_i and compute the $delay$ between CR_A and pruned CR_i.

4 Select all the CRs satisfying, $\text{E}_i \geq \lambda \text{E}_{(\text{CR based on seed } \mathcal{O}_\text{B})}$

5 Select the CR with minimum $delay$, wait $delay$ amount of time, and return with the selected CR.

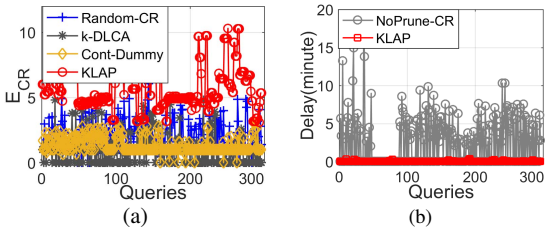

Fig. 3. Comparison: (a) privacy level E_CR and (b) delay. Parameters: $k = 9, l = 5, \alpha = 10, \lambda = 0.9, R = 1000$ meter, $\delta t = 1$ minute.

III. Performance Evaluation

We evaluate KLAP with a dataset having 227,428 check-ins of 1,083 users from 38,333 unique locations in NYC [3]. We compare KLAP with the following baseline approaches: area requirement $\mathcal{A}$ based random Rand-CR, k-DLCA [4], Cont-Dummy [5], and NoPrune-CR. Here, NoPrune-CR is a variation of KLAP without the pruning step.

We first discuss the privacy level achieved in different approaches. Results in Fig.3(a) show that KLAP can achieve the highest privacy in more than 96% of cases. The oscillating nature of KLAP's result is effected either by regions' different location density or the CR pruning step. If density is high, a generated CR may cover large amount of related locations, yielding higher privacy level. The impact of CR pruning on delay is presented in Fig. 3(b). Interestingly, we found that, with a small reduction ($\leq 10\%$) in the privacy level, this delay can be reduced drastically to (≤ 1) minute.

IV. Conclusion and Future Work

In this paper, we introduced a user-centric obfuscation approach, called KLAP, to preserve privacy from an untrusted location-based service provider. The careful design of the approach allows it to guarantee privacy against a wide array of inference attacks, including a long-term obfuscated trajectory tracking attack. Furthermore, with a novel CR pruning technique it significantly reduces the delay for successive queries in spatio-temporal domain. In the future, we aim to devise a mechanism to introduce differential privacy in our approach.

References

[1] I. Liccardi, A. Abdul-Rahman, and M. Chen. I know where you live: Inferring details of people's lives by visualizing publicly shared location data. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM, 2016.

[2] G. Ghinita, M. L. Damiani, C. Silvestri, and E. Bertino. Protecting against velocity-based, proximity-based, and external event attacks in location-centric social networks. *ACM Trans. Spatial Algorithms Syst.*, 2(2):8:1–8:36, June 2016.

[3] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu. Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1):129–142, 2015.

[4] D. Liao, X. Huang, V. Anand, G. Sun, and H. Yu. k-dlca: An efficient approach for location privacy preservation in location-based services. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2016.

[5] H. Liu, X. Li, H. Li, J. Ma, and X. Ma. Spatiotemporal correlation-aware dummy-based privacy protection scheme for location-based services. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, May 2017.