On Cloaking Sensitive Pattern Sets for Long-Term LBS Applications

Jian-Kai Song, Bo-Wen Duan, Hsu-Heng Chou, and Hsiao-Ping Tsai
National Chung Hsing University, Taiwan, R.O.C.
hptsai@nchu.edu.tw

***Abstract*—To protect privacy for long-term LBS users, we propose a pattern-based privacy preserving LBS system that incorporate a privacy preserving server to handle the privacy protection affairs for all registered users. We formulate the p3 problem for long-term LBS and propose the APP algorithm that iteratively identify sensitive patterns and replace a sensitive pattern with an SPSC region. To prevent adversaries culling out user's patterns, for a sensitive pattern, we define a Secure Pattern-Safe-Cloak (SPSC) region that is a rectangle area containing at least other $l-1$ companion patterns to conceal the sensitive one. The experimental results show that the p3 problem is very common and the proposed the system is useful and feasible.**

I. INTRODUCTION

There are many LBS applications frequently used in our daily life, like Navigator, TripAdvisor [2], and Map My Friends [1]. While using the LBS applications, user's locations are sent to the LBS providers in order to acquire information. However, LBS applications have privacy leakage issues since the LBS providers may not be trustworthy and the locations of LBS users can be observed by an adversary. To protect privacy of LBS users, previous privacy preserving techniques mostly consider protecting a prompt location or consecutive locations in a short period, e.g., location cloaking [3], mix-zone[4], and dummy trajectories [5]. Some consider the privacy issues on query contents [10]-[11]. Also, [7]-[9] focus on anonymizing data before publishing in an off-line fashion. Observing that patterns are frequent subsequences or motifs that represent the characteristics in original movement trajectories [6]. They are easily observed and frequently used to profile a user in many applications. In case that location data of a user are long-term accumulated, users' patterns are likely to be discovered and used to re-identify a user. However, few in the literature consider the problem of protecting sensitive patterns for on-line LBS applications

From the perspective of privacy preserving, a pattern owned by a smaller set of users is more specific and sensitive. Thus, individual patterns are of different sensitivity because patterns can be shared by different numbers of users. Inherent from the k-anonymity model, a pattern is owned by less than k users is considered sensitive. Conforming to the l-diversity model, we define the Secure Pattern-safe Cloak (SPSC) region to conceal a sensitive pattern. Moreover, we find that multiple insensitive patterns together can become sensitive, i.e., a user can be re-identified if he has a specific set of patterns that are insensitive individually. Since it is not easy to obtain all patterns of a user, as a compromise, we formulate the Pattern-set Privacy Protection (p3) problem is to cloak individual user's patterns so that the pattern sets with size $\leq M$ observed are all K-anonymous. To tackle the problem, we propose the Apriori-based Pattern Set Protecting (APP) algorithm that iteratively exams for sensitive pattern sets and computes SPSC regions to cloak specific patterns. In the proposed privacy preserving LBS system, a privacy preserving (PP) server is incorporated to handle the privacy protection affairs for all registered users. It conducts the APP algorithm in an off-line manner periodically and provide the sensitivity information to LBS users. It also provides indirect LBS query to protect location privacy for new LSB users. To validate the effects of the proposed APP algorithms and the impact factors that affect the SPSC regions, we conduct experiments with the Geolife datasets. The experimental results show that the p3 problem is very common and the proposed APP algorithm can efficient identify sensitive pattern sets and find SPSC regions with a reasonable area to conceal specific patterns. The KNN query results by using the SPSC regions still retain good data utility even when the dataset contains 182 users, i.e., the proposed the system is useful and feasible.

II. THE PROPOSED PRIVACY PRESERVING LBS SYSTEM

As our previous work [13], we assume that the LBS providers may not be trustworthy and incorporate a trusted PP server in the privacy preserving LBS system to handle the privacy protection affairs. As shown in Fig. 1, the PP server acts as an agent to conduct indirect queries for new LBS users meanwhile collecting their location data. It can preserve location privacy for new users by using traditional location cloaking techniques like [3]-[4]. After a certain amount of location data is accumulated, it mines the movement patterns in an off-line manner based on the location sequences and uses an R-tree data structure to store the patterns. Next, it identifies sensitive pattern sets for individual users and computes cloak regions for concealing specific patterns. After that, it pushes a sensitivity table that contains information about an individual user' sensitive patterns as well as their cloak regions back to the client device.

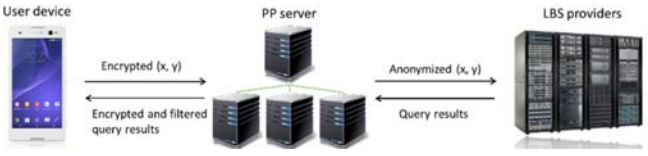

Fig. 1 The privacy preserving LBS system in indirect query mode.

To protect personal location privacy, a new LBS user first registers with the PP server and enters indirect LBS query mode. The user's queries are forwarded the PP servers and the PP cloak the queries and filter query results for the users. As shown in Fig. 2, while the sensitivity table is available, the user device can conduct LBS in direct query mode to alleviate the workload

of the PP server. The user device can continue to send location information to the PP server meanwhile the PP server can push the updated sensitivity table back to the client end. As more users register in the system, the PP server has more information to identify a pattern's sensitivity such that the cloak region of a sensitive pattern may shrink and become insensitive. During the direct query mode as shown in Fig. 3(a), the LBS user first looks up the table to figure out whether a query location is sensitive. If a location is sensitive, it makes a query by using the cloak region (like R_D in Fig. 3 (a)); otherwise, it queries the LBS provider directly.

Fig. 2 The privacy preserving LBS system in direct query mode.

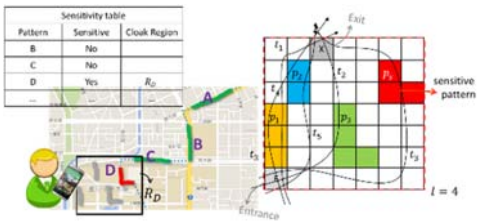

Fig. 3 (a) Using scenario of the privacy-aware LBS: Queries issued from inside region R_D can be carried out by using range query or indirectly through the PP server. (b) Example of a SPSC region where p_t is a specific pattern to conceal.

As [13], to prevent adversaries culling out some patterns with some simple clues like direction and spatial connectivity, we define the secure pattern-safe-cloak (SPSC) region of a pattern p_t as a minimal rectangle area that covers p_t and at least other $l-1$ companion patterns. A companion pattern is a pattern that is, to a certain degree, different from p_t and there exists at least γ trajectories entering the SPSC region from the same grid as that of p_t, passing the companion pattern, and exiting the SPSC region from the grid as p_t, i.e., companion patterns share the same entry and exit with p_t regarding the SPSC region. To confuse adversaries, p_t and the l patterns should be mutually dissimilar and we use Jaccard coefficient as the metric to measure the dissimilarity. Fig. 3 (b) shows an example of a SPSC region to conceal p_t.

Fig. 4 An original pattern table and its anonymized results.

To tackle the p3 problem, we propose the APP algorithm that exams combinations of patterns for sensitive one iteratively. It is an Apriori-like algorithm and the difference is that we look for low frequency pattern sets and try to wipe them out by computing a SPSC region to substitute partial patterns of a sensitive pattern set as shown in Fig. 4. Specifically, it first figures out sensitive pattern sets. For a sensitive pattern set, it searches the R-tree for the near-by patterns that together forms a SPSC region to cloak a pattern or a subset of the pattern set and chooses the one with the smallest region.

III. Experimental Results

To study the p3 problem and the performance of our approach, we conduct experiments with the Geolife dataset [12]. Fig. 5 (a) and (b) show that more than 68% patterns are contained by a single user and there are many sensitive pattern sets even when the size of pattern sets is 2. Our LBS App randomly chooses a location within a SPSC region to conduct LBS query and a large cloak region generally results in a lower data utility. Fig. 5(c) and (d) show the average size of SPSC regions and the data utility of the query results. The KNN query results by using the SPSC regions still retain good data utility for data types like tourist attractions or subway stations.

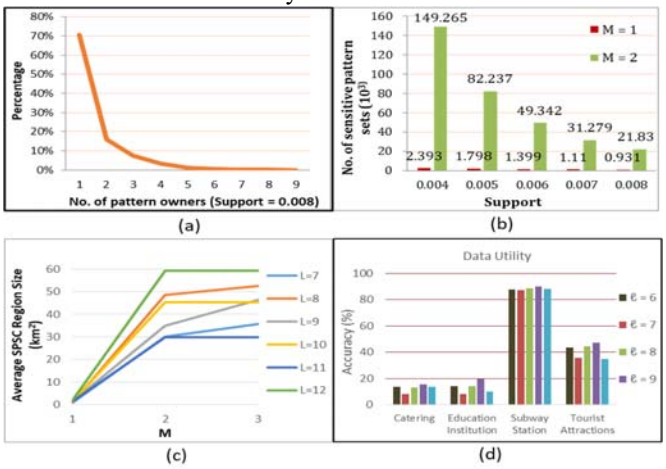

Fig. 5 Experimental results (grid size: $10^4\ m^2$, default support $= 0.008$, k $= 2$)

Acknowledgment

The work was supported in part by the MOST of Taiwan, R.O.C., under Contracts 106-2218-E-018 -003 and 106-3114-E-005-002.

Reference

[1] http://play.google.com/store/apps/details?id=com.vizapps.mapmyfriends.
[2] http://play.google.com/store/apps/details?id=com.tripadvisor.tripadvisor
[3] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking," MobiSys, 2003
[4] J Freudiger, M Raya, M Félegyházi, P Papadimitratos, and J. Hubaux, "mix-zones for location privacy in vehicular networks,",Win-ITS, 2007
[5] Wen-Chih Peng ; Wang-Chien Lee, "Protecting Moving Trajectories with Dummies," MDM, 2007
[6] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, " Trajectory Pattern Mining," ACM SIGKDD, pages 330-339, 2007.
[7] R. G. Pensa, A. Monreale, F. Pinelli and D. Pedreschi, "Pattern-Presertving k-Anonymization of Sequences and its Application to Mobility Data Mining," PiLBA, 2008
[8] G. Andrienko, N. Andrienko, and F. Giannotti, "Movement Data Anonymity through Generalization," ACM SPRINGL, 2009.
[9] Faisal Shahzad, Sohail Asghar, and Khalid Usmani, " A Fuzzy Based Scheme for Sanitizing Sensitive Sequential Patterns," the Int'l Arab J. of Information Technology, Vol. 12, No.1, 2015
[10] A. Pingley, N. Zhang, X. Fu, H.-A.Choi, S.S. Subramaniam, W. Zhao, "Protection of query privacy for continuous location based services," IEEE INFOCOM, vol., no., pp.1710-1718, 2011
[11] F. Olumofin, I. Goldberg, P. K. Tysowski, U. Hengartner, "Achieving efficient query privacy for location based services," PETS, 2010
[12] Y. Zheng, L.Wang, R. Zhang, X. Xie, W.-Y. Ma, "GeoLife: Managing and Understanding Your Past Life over Maps," MDM, p.211-212, 2008
[13] Feng-Yuan Tai, Jian-Kai Song, Yi-Chen Tsai and Hsia-Ping Tsai. "Cloaking Sensitive Patterns To Preserve Location Privacy For LBS Applications," IEEE ICCE- Taiwan, 2016.