SECLUSION SAFEGUARDING AND DATA SHIELDING LOCATION BASED QUERIES

P Panindra Kumar[1] C.Madhuri Yashoda[2]

[1]*M.Tech Dept of CSE, B.I.T.S College,Affiliated to JNTUA, AP, India*
[2]*Assistant Professor, Dept of CSE,B.I.T.S College,Affiliated to JNTUA, AP, India.*

***Abstract*—The growing trend of embedding positioning capabilities (e.g., GPS) in mobile devices facilitates the widespread use of Location Based Services. For such applications to succeed, privacy and privacy are essential. Existing privacy enhancing techniques rely on encryption to safeguard communiqué channels, and on pseudonyms to protect user identities. Nevertheless, the query contents may disclose the physical location of the user. In this paper, we present a structure for preventing location based identity inference of users who issue spatial queries to Location Based Services. We propose transformations based on the well-established K-anonymity concept to compute exact answers for range and nearest neighbour search, without revealing the query source. Our methods optimize the entire process of anonymizing the needs and processing the transformed spatial queries. Extensive experimental studies suggest that the proposed techniques are applicable to real-life scenarios with numerous mobile users.**

Index Terms— Privacy, Anonymity, Location Based Services, Spatial Databases, Mobile Systems.

I. INTRODUCTION

The status of mobile devices with localisation chips and ubiquitous access to Internet give rise to a large number of location based services (LBS). Consider a user who wants to know where the nearest gas station is. He sends a query to a location-based service provider (LBSP) using his smart-phone with his location attached. The LBSP then processes the query and responds with results. Location-based queries lead to privacy concerns especially in cases when LBSPs are not trusted. Attackers can cooperate with LBSPs and have access to users' location-related queries. The amount and risk of information leakage from LBS queries have been discussed, for example, in [7].The analysis mainly focused on information leakage from locations. However, query content itself is also a source of users' privacy leakage. For instance, a query about casinos implies theissuer's gambling habit which the issuer wants to keep secret. Thus besides location privacy, the anonymity of issuers with respect to queries is also important in privacy preservation. Intuitively, query privacy is the ability to prevent other parties to learn the issuers of queries. One way to protect query privacy is to anonymise queries by removing users' identities. However, this does not suffice when considering locations which can helpreveal users'identities, since attackers can acquire users' locations through a number of ways, e.g., triangulating mobile phones' signals and localising users' access points to Internet. Sometimes, public information such as home addresses and yellow pages can also help obtain users' positions. In the last few years, k-anonymity [2] has been widely used and investigated in the literature on releasing micro data, e.g., medical records. A common assumption for k-anonymity is that all users have the same probability to issue queries. In other words, a uniform probability distribution is assumed over users with respect to sending any query, which is often not realistic especially when attackers gain more information about the users. Given a specific query, certain users tend to be more likely to issue it when compared to others. For instance, users who love movies are more possible to search for nearing cinemas. For any user in a generalised area satisfying k-anonymity, the probability to be the issuer is no longer 1 k in such situations. The case can be worse especially for those users who are more likely than others. Suppose a k-anonymised region of a query from a young person for searching clubs at midnight. If there are only two young people in the generalised region, then they are more likely to be taken as the candidates for the issuer from attackers' view than other users in this region. Therefore, k-anonymity is not a sufficient metric to describe users' privacy requirements when taking into account user profiles, which was addressed first by Shin et al. [2]. Nowadays, the popularity of social networks and more exposure of people's information on Internet provide attackers sources to gather enough background knowledge to obtain user profiles. Besides passive attacks in which attackers simply observe the connection between users, attackers can also perform active attacks,e.g., by creating new accounts so as to identify users even in an anonymised social network [1]. Wu et al. give a literature study on the existing attacks to obtain users' profiles [3]. Therefore, it is a new challenge to measure and protect users' query privacy in LBSs with the assumption that attackers have the knowledge of user profiles.Our contributions. In this paper, we extend k-anonymity and propose new metrics to correctly measure users' query privacy in the context of LBSs, which enable users to specify their query privacy requirements in different ways.Furthermore, we design new generation algorithms to compute anonymising spatial regions according to users' privacy requirements. Through experiments, we show that our algorithms are efficient enough tomeet users' demands on real-timeresponsesand generate regions satisfying privacy requirements. We also show the different strengths of our metrics which help users choose the correct requirements to achieve a balance between privacy and the quality of service delivered by the LBSP.

II RELATED WORK

We give a brief literature study on measuring anonymity and on query privacy metrics with focus on k-anonymity. Then we summarise existing region generalisation algorithms. 2.1 Anonymity metrics In the literature, various ways to measure anonymity have been proposed. Chaum [6] uses the size of an anonymity set to indicate the degree of anonymity provided by a network based on Dining Cryptographers. An anonymity set is defined as the set of users who could have sent a particular message as observed by attackers. Berthold et al. [3] define the degree of anonymity as log N, where N is the number of users. Reiter and Rubin define the degree of anonymity as the probability that an attacker can assign to a user of being the original sender of a message. They introduce metrics like beyond suspicion, probable innocence and possible innocence. Serjantov and Danezis [2] define an anonymity metric. based on entropy and a similar metric is given by Díaz et al. which is normalised by the number of users. Zhu and Bettati propose a definition of anonymity based on mutual information. The notion relative entropy is used by Deng et al. to measure anonymity. Different information-theoretic approaches based on Kullback-Leider distance and min-entropy are proposed to define information leakage or the capacity of noisy channels.

2.1 Query privacy metrics

The concept of k-anonymity was originally proposed by Samarati and Sweeney in the field of database privacy [24]. The main idea of k-anonymity is to guarantee that a database entry's identifier is indistinguishable from other k−1 entries. However, this method does not work in all cases. For instance, the fact that an HIV carrier is hidden in k carriers does not help protecting his infection of the virus. Further research has been done to fix this problem [18]. In the context of privacy in LBSs, k-anonymity is first introduced by Gruteser and Grunwald [15]. It aims to protect two types of privacy – location privacy and query privacy. The former means that given a published query, attackers cannot learn the issuer' exact position while the latter enforces the unlink ability between the issuer and the query. Because of its simplicity, k-anonymity has been studied andrefined in many ways. For instance, Tan et al. define information leakage to measure the amount of revealed location information in spatial cloaking, which quantifies the balance between privacy and performance [32]. Xue et al. [34] introduce the concept of location diversity to ensure generalised regions to contain at least ℓ semantic locations (e.g., schools, hospitals). Deeper understanding of k-anonymity reveals its drawbacks in preserving users' location privacy. Shokri et al. analyse the effectiveness of k-anonymity in protecting location privacy in different scenarios in terms of adversaries' background information [3], i.e., real-time location information, statistical information and no information. Based on the analysis, they conclude that cloaking (e.g., k-anonymity) is effective for protecting query privacy but not location privacy. They also show its flaws which the adversary can exploit to infer users' current locations. In this paper, we focus on protecting query privacy using cloaking with the assumption that the adversary learns users' real-time locations. Recently, Shokri et al. design a tool Location-Privacy Meter that measures location privacy of mobile users in different attack scenarios .Their work assumes that attackers can utilise user profiles (e.g., mobility patterns) extracted from uses' sample traces to infer the ownership of collected traces. It is in spirit close to our work. They use the incorrectness of attackers' conclusions on users' positions drawn from observations as the privacy metric. In this paper, we focus on users' query privacy with regards to an individual query rather than query histories. Moreover, we make use of users' static and public personal information, such as professions and jobs as user profiles. Considering information such as mobility patterns and query histories is part of our future work. The work by Shin et al. [26] is most closely related. They describe user profiles using a set of attributes whose domains are discretised into disjoint values. User profiles are represented by profile vectors with a bit for each value. Shin et al. propose three new metrics based on k-anonymity by restricting different levels of similarity between profiles of users in generalised regions. This is analogous to our notion of k-approximate beyond suspicion which will be discussed in Sect. 4. Compared to Shin et al.'s work [26], we define a more comprehensive set of metrics that can measure query privacy from different perspectives and develop corresponding generalisation algorithms. 2.3 Area generalisation algorithms The first generalisation algorithm called IntervalCloaking is designed by Gruteser and Grunwald [15]. Their idea is to partition a region into quadrants with equal area. If the quadrant where the issuer is located contains less than k users, then the original region is returned. Otherwise, the quadrant with the issuer is taken as input for the next iteration. The algorithm CliqueCloak [14] is proposed by Gedik and Liu in which regions are generalised based on the users who have issued queries rather than all potential issuers. The major improvement is that this algorithm enables users to specify their personal privacy requirements by choosing different values for k. Mokbel et al. [21, 8] design the algorithm Casper which employs a quadtree to store the two-dimensional space. The root node represents the whole area and each of other nodes represents a quadrant region of its parent node. The generalisation algorithm starts from the leaf node which containsthe issuer and iteratively traverses backwards to the root until a region with morethan k users is found.Another algorithm nnASR [16] simply finds the nearest k users to the issuer and returns the region containing these users as the anonymising spatial region. The above algorithms suffer from a particular attack called "outlier problem" [2], where attackers have the generalisation algorithms and users' spatial distribution as part of their knowledge. Intuitively, this happens when some users in a generalised region do not have the same region returned by the algorithm as the issuer. Thus, these users can be removed from the anonymity set, resulting in a set with less than k users. Hence, an algorithm against this attack needs to ensure that for each user in the anonymity set it always returns the same region. Kalnis et al. design the first algorithm called hilbASR that does not have the outlier problem [16]. The algorithm exploits the Hilbert space filling curve to store users in a total order based on their locations. The curve is then partitioned into blocks with k users. The block with the issuer is returned as the generalised region. Mascetti et al. propose two algorithms, dichotomicPoints and grid, which are also secure against the outlier problem [20]. The former iteratively partitions the region into two blocks until less than 2k users are located in the region while the latter draws a grid over the two-dimensional space so that each cell contains k users and returns the cell with the issuer. Because of the simplicity of implementation and the relatively smaller area of the generalised regions, we adopt and extend these two algorithms in our algorithm design. The area of generalised regions is usually used to measure the quality of service responded by LBSPs, as smaller regions lead to more accurate query results and less communication overhead.

III. THE ANONYMIZER

Assumptions and Goals of Spatial Anonymization The anonymizer is a trusted server, which collects the current location of users and anonymizes their queries. Each query has a required degree of anonymity K, which ranges between 1 (no privacy requirements) and the user cardinality (maximum privacy). We assume that an attacker has complete knowledge of (i) all the ASRs ever received at the LBS, (ii) the cloaking algorithm used by the anonymizer, and (iii) the locations of all users. The first assumption states that either the LBS is not trusted (e.g., a commercial service that collects unauthorized information about its clients for unsolicited advertisements), or the communication channel between the anonymizer and the LBS is not secure. The second assumption is common in the security literature since the data privacy algorithms are usually public. The third assumption is motivated by the fact that users may often (or always) issue queries from the same locations (home, office), which may be easily identified through public databases, telephone directories, etc. Furthermore, they may reveal their locations by issuing queries without privacy requirements. In scenarios with highly mobile users, the attacker may not be able to learn exact user locations. However, one can argue that in these cases spatial K-anonymity is not important, because (i) the user ids are removed by the anonymizer anyway, and (ii) a query at a random position does not necessarily reveal information about the identity of the corresponding user. However, in practice, a determined attacker may be able to acquire (through triangulation, public databases, physical observation, etc.) the locations of at least a few users in the vicinity of the targeted victim. Similar to existing work on LBS query privacy [10], [15], [23] we focus on snapshot queries, where the attacker uses current data, but not historical information about movement and behavior patterns of particular clients (e.g., a user often asking a particular query at a certain location or time). This assumption is reasonable in practice, because if a client obtains the items of interest (e.g., the closest restaurant), it is unlikely to ask the same query from the same location again in the future. We also assume that the attacker does not have a priori knowledge of the user query frequencies (i.e., a query may originate from any user with equal probability). Furthermore, the value of K is not subject to attacks since it is transferred from the client to the anonymizer through a secure channel. Given a query, the anonymizer removes the user id, applies cloaking to hide the user's location through an ASR, and forwards the ASR to the LBS. The cloaking algorithm is said to preserve spatial K-anonymity, if the probability of the attacker pinpointing the query source under the above assumptions does not exceed 1/K. Note that simply generating an ASR that includes K users is not sufficient for spatial K-anonymity. Consider for instance, aanalgorithm, called Center Cloak (CC) in the sequel, which given a query from U, finds his K-1 closest users, and sets the ASR as the minimum bounding rectangle (MBR) or circle (MBC) that encloses them. In fact, a similar technique is proposed in[10] for anonymization in peer-to-peer systems, i.e., the K-ASR contains the query issuing peer and its K-1 nearest nodes. CC is likely to disclose the location of U under the center-of-ASR attack. Specifically, let indexU be the position of U in the sequence of users enclosed by the K-ASR, sorted in ascending order of their distance from the center of the K-ASR; for example, if indexU = 1, then U is the closest user to the center. The center-of-ASR attack is successful if P[indexU = 1] > 1/K, i.e., if the probability of U being the closest user to the center exceeds 1/K. Figure 5 shows the distribution of the positions of U inside an MBR enclosing its 9 NNs (for details of the experimental setting, see Section V). In most cases, U is close to the center of the 10-ASR (i.e., P[indexU = 1] > 1/10). Hence, an attacker with knowledge of the cloaking algorithm (assumption ii) may easily pinpoint U as the query source. Note that, since the MBR may enclose more than 10 users it is possible to get P[indexU = i] > 0 for i > 10. The dashed line in the graph corresponds to the "flat" index distribution obtained by an ideal anonymization technique, which would always generate 10-ASRs with exactly 10 users.

IV CONCLUSION

In this paper, we consider a powerful attacker who can obtain user profiles and has access to users' real-time positions in the context of LBSs. Assuming this stronger attacker model, we propose new metrics to correctly measure users' query privacy in LBSs, including k-ABS, α-USI, β-EBA and γ-MIA. For information theory based metrics, the determination of users' specified values is not intuitive. However, users can use other metrics as references. For instance, k-anonymity corresponds to log k-EBA when the distribution for users to issue a query is (close to) uniform. Spatial generalisation algorithms are developed to compute regions satisfying user's privacy requirements specified in the proposed metrics. Extensive experiments show our metrics are effective in balancing privacy and quality of service in LBSs and the algorithms are efficient to meet the requirement of real-time responses. Our metrics are not exhaustive, and there exist other ways to express query privacy. For instance, we can use min-entropy to express information leakage [31] in a way analogous to mutual information: $I\infty(X; Y) = H\infty(X) - H\infty(X \mid Y)$. Intuitively, it measure the amount of min-entropy reduced after the attacker has observed a generalised query. It is very interesting to study differential privacy [12] to see how it can be adopted for LBS scenarios. In future, we want to develop an application for an LBS, making use of the proposed metrics to protect users' query privacy. This can lead us to a better understanding of privacy challenges in more realistic situations. The implementation of our algorithms can also be improved as well, e.g., using a better clustering algorithm for kABS. Another interesting direction is to study a stronger attacker model, where the attacker, for instance, can have access to mobility patterns of users.

REFERENCES

[1] M. Bellare and S. Micali, "Non-interactive oblivious transfer and applications," in Proc. CRYPTO, 1990, pp. 547–557.

[2] A. Beresford and F. Stajano, "Location privacy in pervasive computing," IEEE Pervasive Comput., vol. 2, no. 1, pp. 46–55, Jan.–Mar. 2003.

[4] C. Bettini, X. Wang, and S. Jajodia, "Protecting privacy against location-based personal identification," in Proc. 2nd VDLB Int. Conf. SDM, W. Jonker and M. Petkovic, Eds., Trondheim, Norway, 2005, pp. 185–199, LNCS 3674.

[5] X. Chen and J. Pang, "Measuring query privacy in location-based services," in Proc. 2nd ACM CODASPY, San Antonio, TX, USA, 2012, pp. 49–60.

[6] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," J. ACM, vol. 45, no. 6, pp. 965–981, 1998.

[7] M. Damiani, E. Bertino, and C. Silvestri, "The PROBE framework for the personalized cloaking of private locations," Trans. Data Privacy, vol. 3, no. 2, pp. 123–148, 2010.

[8] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," in Proc. 3rd Int. Conf. Pervasive Comput., H. Gellersen, R. Want, and A. Schmidt, Eds., 2005, pp. 243–251, LNCS 3468.

[9] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," IEEE Trans. Inform. Theory, vol. 31, no. 4, pp. 469–472, Jul. 1985.

[10] B. Gedik and L. Liu, "Location privacy in mobile systems: A personalized anonymization model," in Proc. ICDCS, Columbus, OH, USA, 2005, pp. 620–629.

[11] C. Gentry and Z. Ramzan, "Single-database private information retrieval with constant communication rate," in Proc. ICALP, L. Caires, G. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, Eds., Lisbon, Portugal, 2005, pp. 803–815, LNCS 3580.

[12] G. Ghinita, P. Kalnis, M. Kantarcioglu, and E. Bertino, "A hybrid technique for private location-based queries with database protection," in Proc. Adv. Spatial Temporal Databases, N. Mamoulis, T. Seidl, T. Pedersen, K. Torp, and I. Assent, Eds., Aalborg, Denmark, 2009, pp. 98–116, LNCS 5644.