Privacy Preserving Nearest Neighbor Search based on Topologies in Cellular Networks

Mahdi Daghmehchi-Firoozjaei
Department of Computer Science and Engineering
Sungkyunkwan University
Suwon, Korea
mdaghmechi@skku.ac.kr

Jaegwan Yu
Department of Electronic and Information Engineering
Korea University
Saejong, Korea
compaso@korea.ac.kr

Hyoungshick Kim
Department of Computer Science and Engineering
Sungkyunkwan University
Suwon, Korea
hyoung@skku.ac.kr

***Abstract*—As the popularity of location-based services (LBSes) is increasing, the location privacy has become a main concern. Among the rich collection of location privacy techniques, the spatial cloaking is one of the most popular techniques. In this paper, we propose a new spatial cloaking technique to hide a user's location under a cloaking of the serving base station (SeNB) and anonymize SeNB with a group of dummy locations in the neighboring group of another base station as central eNB (CeNB). Unlike the most existing approaches for selecting a dummy location, such as the center of a virtual circle, we select a properly chosen dummy location from real locations of eNBs to minimize side information for an adversary. Our experimental results show that the proposed scheme can achieve a reasonable degree of accuracy (>96%) for nearest neighbor services while providing a high level of location privacy.**

Keywords—Location-based service (LBS); spatial cloaking; anonymity; dummy location; eNode B (eNB)

I. INTRODUCTION

Because of the popularity of smart phones with positioning capabilities and global trend toward online services (e.g. social networks), Location-based services (LBSes) are getting more used these days. By these services, users can obtain useful information about their surroundings (e.g. nearest neighbor query) while their privacy would be under some violations if there is no privacy protection. Exposing user's accurate position or interested activities by an untrusted service provider (SP) can be mentioned as some examples of this privacy violence. Therefore location privacy will be one of the key issues to deploy LBS.

Generally, spatial obfuscation/cloaking and user anonymity are common approaches for privacy protection. Location obfuscation techniques have generally tried to decrease the precision of user position so that attackers can only retrieve coarse-grained position information. Despite a spatial cloaking technique is to blur a user's exact location into a cloaked area that satisfies the user's location privacy requirements, but it cannot protect the user and may lead to unauthorized revealing of information about its activities and interests [1]. In anonymity approaches, like *k-anonymity*, a group of users (*k* users) that are indistinguishable from each other are collected to form an anonymity set. In this situation, an adversary cannot identify the real user out of this set from the other dummy locations. Basically, *k-anonymity* approach needs a trusted third-party (TTP) server to select *k-1* users to perform anonymity. Therefore, it is questionable whether the assumption of a centralized location anonymizer (TTP) is realistic [2]. Moreover, selecting dummy locations to achieve *k-anonymity* without any explosion of side information to the adversary is another challenge. Traditionally, dummy locations are generated randomly [3][4] but these generated locations usually cause side information. For example, in carelessly dummy generating, may some locations fall at some unlikely locations such as lakes, rugged mountains, or illogical places in time and can be easily filtered out by the adversary [5].

In order to overcome mentioned limitations, we propose a method of location privacy based on spatial cloaking in the cellular networks. Basically, instead of user's location, we use the location of its serving base station, i.e. serving eNB (SeNB), the connection node of the LTE network and user, to create a cloaking area. In the next phase, we design an algorithm to choose some eNBs in the neighboring of SeNB as proper dummy locations. We present a novel method such that in response to every point of interest (POI) query, SP replies a group of POIs. These points are close to each eNB in the neighboring group of a central eNB (CeNB). The CeNB is an eNB which is selected by user instead of its real location. Therefore, we can attain anonymity without TTP involving.

There is a collection of related work that try to protect users' privacy. In order to achieve privacy by a *k-anonymity* approach and to avoid trustworthy limitation, generating dummy location without TTP involvement same as our method, was studied in [6] [7] and [8]. In a spatial cloaking approach, since the user himself defines obfuscation area some near POIs may be excluded, some papers like [9] try to propose solutions to optimize cloaking area. In [10] an approach called *SpaceTwist* is proposed to answer *k*-nearest-neighbor query in which instead of user's position, a fake location (anchor) is selected to query and SP returns points incrementally in ascending order of their distances from the anchor. User can choose the better one among the suggested points while through higher query and communication costs. However, in our work we initially use the same idea as using fake location instead of user's precise

location, we extend our protecting solution beyond this concept and create a *k*-cluster group of special locations of cellular network as dummy locations to prepare an anonymity set for this fake location. Whereas, communication cost is blamed as drawback of *SpaceTwist*, in our model by using features of cellular network, the location privacy is achieved through lower query and communication costs and acceptable precision.

II. Model Description

Nowadays, because of professional activities of LBS providers and required technology and software, these service providers are not necessarily a part of the network operators [5]. Therefore, we assume the network operators and service providers are separate entities. Every user has an identity for a service which has no relationship to its network identification and accordingly the SP has no view of its reality. In the cellular networks, every user is covered and is surrounded by multiple base stations, eNBs [11]. The positions of eNBs are fixed and there are enough number of them in the vicinity of a user where can be seen as landmarks. Based on this property, we use these positions as Basic Locations (BL).

The basic idea of our model is selecting *m* positions which not only have the proper relation to the user, but also completely environ it. We suggest eNB as BL because:

- They are distributed everywhere in the network according to density of population and traffic.
- The adversary has no chance to detect and filter out some improperly selected dummy locations which may fall at some unlikely locations such as lakes.
- Although the locations of eNBs are fixed, because of their similarity and abundance we can select *m* number of them properly to decrease the detecting chance for adversary.

In this scheme, we assume SP as the adversary which is able to obtain information and monitor queries sent by users. So he knows the history of the user's query and tries to learn the user's sensitive information by some methods like map matching.

III. Dummy Locations Selection

Basically, diverse position of eNBs provides better privacy for a user, but SP as adversary completely monitors locations selected by the user and can detect its movement and preferences. To minimize information leakage available to SP, we define a weight index plan to classify eNBs and BLs selection. These weight indexes are set up according to eNBs' location in mobile user's point of view. The eNB that serves this user, SeNB, has the least weight index and is located in the layer, the neighboring eNBs of SeNB are located in the second layer and third layer consists of other stations which are not the neighbor of SeNB and their weight indexes increase according to their layers. Therefore, stations in the third layer have the highest weight index and it is better to be located in the vicinity of the second layer.

Beside to SeNB, we select *m-1* eNBs, which are related to SeNB, as dummy locations. According to this plan, a neighbor node of SeNB which is closer to user more than other neighbors is selected as central eNB (CeNB). Then neighbors of CeNB which consists of SeNB are used as dummy locations.

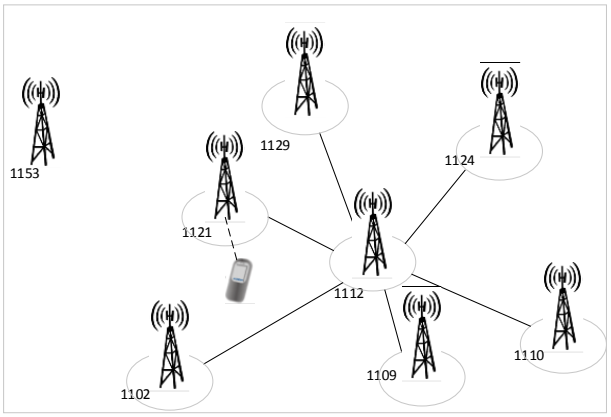

Figure 1. Selection of dummy locations

Figure 1 shows an example of dummy locations: the node eNB-1112, as CeNB, and its neighbors (include SeNB, eNB-1121). In this example, the user selects eNB-1112 as CeNB because in the user's view the received signal of this station has the biggest strength after SeNB's signal.

IV. Privacy Analysis

On behalf of the mobile user's location, we use the SeNB to create an obfuscation area. Furthermore, by selecting a fake location (CeNB) and its neighbors as dummy locations, we set an anonymous group to conceal the identity of SeNB. To analyze the produced location privacy protection, we define two concepts: privacy value and anonymity degree.

A. Privacy Value

Basically, the proposed model is a spatial obfuscation approach which benefits dummy locations to achieve anonymity for a cloaking area of SeNB on behalf of the user. To analyze, we use the concept of *privacy value* defined by Yiu et al. [10] as the average distance from a location (uniformly distributed) in the cloaking area to the user's actual location. Based on this definition, if q be the user's location and φ be a cloaking area the privacy value $\gamma(q,\varphi)$ is:

$$\gamma(q,\varphi) = \frac{\int_{z\in\varphi} distance(z,q)dz}{\int_{z\in\varphi} dz} \quad (1)$$

This value is based on the distance between the user's location and other locations in the cloaking area. Therefore, the high value shows most adversary's picked locations in the cloaking area are far from user [10].

B. Anonymity degree

In the next step by creating an anonymous set, we try to hide the identity of SeNB. Based on the definition of anonymity defined by Pfitzmann and Kohntopp in [12], anonymity is the state of being not identifiable within a set of subjects, the anonymity set. If we look every eNB in the dummy set as an information point, the anonymity of the eNB in this model can be measured by information entropy of all BLs. We use *X* which denotes the anonymous model and *H(X)* as its entropy value.

Suppose p_i is the probability of identifying the i[th] BL as real SeNB, so:

$$H(X) = -\sum_{i=1}^{m}(p_i log_2 p_i) \quad (2)$$

The maximum entropy, H_M, for the actual size of our dummy set, *m* is:

$$H_M = log_2(m) \quad (3)$$

This amount for maximum entropy will be achieved when the query behavior of all eNBs looks same in SP's view, so every station has a possibility of *1/m* to be identified as SeNB.

The information which SP can achieve with the attack can be expressed as $H_M - H(X)$ and by dividing by H_M it will be normalized. We use the "*anonymity degree*" provided by the system which defined by Diaz et al. [13] as:

$$d = 1 - \frac{H_M - H(X)}{H_M} = \frac{H(X)}{H_M} \quad (4)$$

The amount of *d* shows the information leakage of the model. The amount of *d* will be minimum (*d=0*) if an eNB in the dummy set appears as being SeNB with probability *1* and it is maximum if all nodes have the same probability of being SeNB.

V. Evaluation

In order to perform, we used datasets of base stations of cellular network provided by *OpenMobileNetwork* in the city of Berlin, Germany. A dataset of real locations of 153 base stations on *T-Mobile* network, a dataset of neighboring groups of this operator and to find POI locations behalf SP, we used *Google API*. To perform the model, we consider a user with a smart phone which equipped with GPS receiver to calculate its position and a map to match suggested POIs with its location. To respond a POI query, by receiving user's request which consists of the cell ID of CeNB, SP extracts CeNB's neighbors (and their locations) from the dataset, neighboring group table, and based on these locations sends back a set of locations of POI in the vicinity of each BL.

A. POI Positioning Accuracy

Basically, in the *nearest neighbor* service for a user, its location is used as a center of a circle with a defined radius and *n* nearest POIs will be suggested, e.g. *Google Places API*. To analyze the accuracy, we compare our model to *Google Places API* model which POIs are found in the area centered by user's location. Similar to this model, in our model user chooses the best one among suggested POIs. In this analysis model if m denotes the real number of POIs nearby the user (*n nearest* service) and n denotes the number of these POIs which are suggested by SP in a selecting model, the accuracy of this selecting model is:

$$A = \frac{n}{m} \times 100 \quad (5)$$

In the Figures 2, accuracies of POI positioning of our model based on the user's preference number of POI (*k*) and the number of POI near to each BL (*k'*). According to these results, by selecting 10 POIs near to each BL (*k'=10*) we can achieve the perfect accuracy for almost all the user's choices.

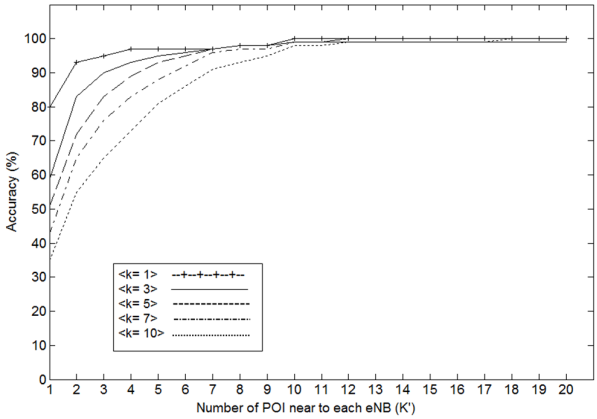

Figure 2. Positioning accuracy based on the number of POI

Although the bigger *k'* leads more accuracy of POI positioning, but it imposes users to download a bigger size of message and accordingly increases communication cost. The communication cost directly relates to message size which should be downloaded by a user during a typical query. The size of the message varies depending on the number of POI in the proximity of BLs (*k'*) and definitely on the number of BLs in the neighboring group of CeNB. The simulation results show that the selection of *k'=8* with the message size of 467 Byte and the accuracy of 96.68% of POI positioning is a reasonable choice.

B. Privacy Value

As mentioned in Section IV, we use the concept of privacy value for this evaluation. Unlike the standard cloaking model which uses circular cloaking, the cloaking area in our framework is a polygon which the number of its vertexes and its area depend on the number of eNB and the network structure change in every neighboring group. Since the cloaking region, φ, does not have certain equation and form in the general case, we use the *Monte Carlo* method for approximating distance to the user's location in φ and performed these two models over 150 users from different neighboring groups.

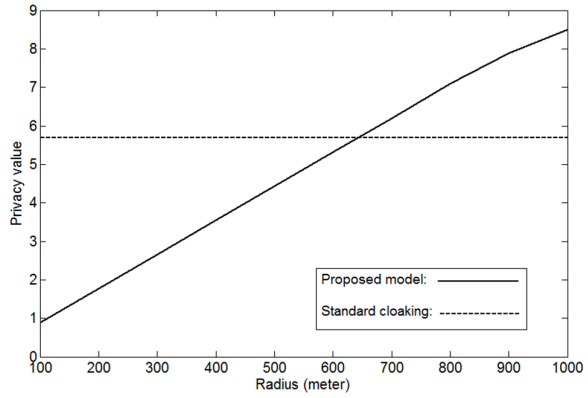

Figure 3. Privacy value of our model in comparison to standard cloaking model

In the Figure 3, the average amount of privacy value achieved by our model is compared to the privacy value in the standard cloaking model. In comparison to the standard model, the average amount of 5.70 (1/m) for privacy value obtained by

our model is comparable to privacy of a standard model with the radius of almost 650m which offers good safety guard for users.

C. Anonymity Degree

As mentioned before, our method benefits a set of *m* related dummy locations as a neighboring group to make an anonymous set. Therefore, that SP faces averagely six locations (expect central eNB) possibilities for SeNB which have the same query probability. According to our adversary model, we suppose in the worst case the SP with the aiding of some attacks like map matching, detects BLs of layer-3 in each group. Since SP knows that CeNB cannot be SeNB, he has the chance of $\frac{1}{m-1}$ to detect SeNB in a neighboring group of *m* members. Basically, the number of eNB in the neighboring group has a direct proportion to the anonymity degree of this model.

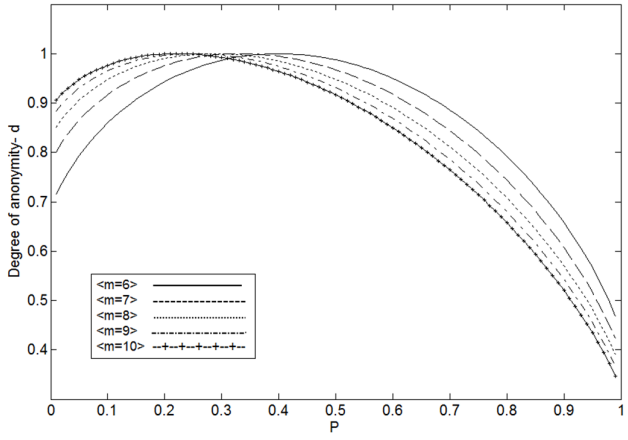

Figure 4. The degree of anonymity of SeNB for different members (*m*) neighboring groups

In the neighboring group of each CeNB, the number of members varies depending on network planning and structure. Hence, we face different members of neighboring groups (m=6, 7, 8, 9 and 10) in our used dataset of base location.

Figure 4 illustrates the anonymity degree of SeNB in the neighboring groups of CeNB with different members. Regarding to this, the bigger member of the neighboring group, the more increase in the degree of anonymity when SP has a little knowledge about BLs of layer-3 and the more decrease in the degree of anonymity when SP can detect BLs of this layer. As depicted in this figure, the degree of anonymity of SeNB does not drop to zero even though SP detects BLs of layer-3, because he has to find it among two or more (depends on neighboring members) same choice.

VI. Conclusion

In this paper, we proposed a framework to protect user privacy for LBSes by utilizing the basic geographical features of cellular networks. Based on these features, the cloaking area of SeNB is used to conceal the user's location and we hide the identity of SeNB in a group of dummy locations in the neighboring group of CeNB. The evaluation shows that the privacy can be guaranteed with a cloaked region. Obviously, even if the adversary can filter out BLs of layer-3 and find SeNB with a high probability.

The evaluation shows that, the proposed model provides a remarkable amount of anonymity degree and privacy value to providing strong privacy guarantees. Furthermore, with the average message size of 467 Bytes and the precision more than 96% for POI positioning, this model achieves location privacy for users with the balance of accuracy and cost.

References

[1] P.A. Kirpekar, "Review of Location Privacy Protection against Location-Dependent Attacks in Mobile Services," International Journal of Computer Science and Management Research, eETECME, vol. 13, 103-105, 2013.

[2] M. Wernke, P. Skvortsov, F. Durr, and K. Rothermel, "A Classification of Location Privacy Attacks and Approaches," Personal and Ubiquitous Computing, Springer-Verlag, vol. 18, issue 1, 163-175, 2014.

[3] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in Privacy-Aware Location-Based Services," In proceeding of the 33rd Conference on Computer Communications, IEEE INFOCOM 2014, 754-762, 2014.

[4] M. F. Mokbel, C.Y. Chow, and W.G. Aref, "The new Casper: query processing for location services without compromising privacy," In proceeding of the 32th international conference on Very Large Data Bases, ACM VLDB 2006, 763-774, 2006.

[5] U. Hengartner, "Hiding Location Information from Location-Based Services," In proceeding of International conference on Mobile Data Management, 268-272, 2007.

[6] E. Novak, and Q. Li, "Near-Pri: Private, Proximity Based Location Sharing," In proceeding of the 33rd Conference on Computer Communications, IEEE INFOCOM 2014, 37-45, 2014.

[7] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," In proceeding of IEEE International Conference on Pervasive Services, ICPS '05, 88-97, 2005.

[8] P. Shankar, V. Ganapathy, and L. Iftode, "Privately querying location-based services with sybilquery," In proceeding of the 11th International Conference on Ubiquitous Computing, UbiComp 2009, 31-40, 2009.

[9] H. Kim, "A Spatial Cloaking Framework Based on Range Search for Nearest Neighboring Search," Data Privacy Management and Autonomous Spontaneous Security Lecture Notes in Computer Science, Springer-Verlag, vol. 5939, 93-105, 2010.

[10] K.L. Yiu, C.S. Jensen, J. Moller, and H. Lu, "Design and analysis of a ranking approach to private location-based services," ACM Transactions on Database Systems- TODS 36(2), vol. 36, 1-42, 2011.

[11] A. Roessler, "Cell search and cell selection in UMTS LTE," Rohde & Schwart, Application Note, 9.2009-1MA150_0E, 2009.

[12] A. Pfitzmann, and M. Kohntopp, "Anonymity, Unobservability and Pseudonymity- A Proposal for Terminology," Hannes Federath (Ed.), Designing Privacy Enhancing Technologies, Lecture Notes in Computer Science, Springer-Verlag, vol. 2009, 1-9, 2001.

[13] C. Diaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," In proceeding of the 2nd International workshop on Privacy Enhancing Technologies, PET 2002, 54-68, 2002.