Privacy-preserving nearest neighbor queries using geographical features of cellular networks

Mahdi Daghmehchi Firoozjaei, Jaegwan Yu, Hyoungeek Choi, Hyoungshick Kim*

College of Information & Communication Engineering, Sungkyunkwan University, Suwon, Republic of Korea

A R T I C L E   I N F O

Article history:
Received 17 December 2015
Revised 19 July 2016
Accepted 17 September 2016
Available online 29 September 2016

Keywords:
Location-based service (LBS)
Location privacy
Spatial cloaking
Dummy location
Nearest neighbor

A B S T R A C T

Although location-based services (LBSes), such as nearest neighbor query, have become popular, privacy remains a challenging issue for users. Many privacy preserving techniques have been proposed, but their complexity, insufficiency, and time consumption make them unattractive to users, who prefer accuracy and quickness. To address this limitation, we introduce a framework to protect user privacy for nearest neighbor queries by utilizing the basic geographical features of cellular networks. In the proposed framework, we provide two layers of spatial anonymity such that the user’s location is not directly provided to a location service provider. Based on the features of the cellular network (e.g., LTE) at the first layer, the user’s location is kept hidden under the cloaking of the base station (eNB) that provides a network connectivity to serve the user (SeNB). At the second layer, we anonymize SeNB in a group of dummy locations neighboring a central eNB (CeNB), all of which have the same query probability. Unlike most existing approaches with fake dummy locations, the proposed framework depends on real locations of eNBS to minimize the likelihood that side information might be exposed to an adversary. Moreover, our model is motivated by the practicality of employing the ubiquity of cellular networks and their geographical features. The simulation results show that the proposed scheme can achieve a decent degree of accuracy (> 98%) while providing strong privacy guarantees.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Currently, the global trend for cellular networks to support smartphones with positioning capability has produced a huge surge of location-based services (LBSes) in almost all social and business sectors. Although LBSes have created new business opportunities, there are some important concerns regarding threats to user privacy. For instance, in the point-of-interest (POI) service, a query to nearby hospitals not only releases the user’s location but also his sensitive information (e.g., health status). Furthermore, an untrustworthy service provider (SP) might reveal a user’s precise position, which could be abused for criminal activities like stalking, robbery, or theft [1,2]. Because the use of an LBS creates privacy concerns regarding user location, location privacy is a key issue to deploying an LBS.

Generally, location privacy techniques have different structures, depending on the protected information and the attack model. Protection techniques have included spatial obfuscation/cloaking, k-anonymity, position dummies, mix zones, and encryption [2–4]. Some protection techniques require a trusted third party (TTP) (e.g., k-anonymity), while others can be performed directly by users (e.g., spatial obfuscation).

Location obfuscation techniques (e.g., spatial cloaking) are common methods to protect the user location. Basically, these approaches aim to decrease the precision of user location so that attackers can only retrieve coarse-grained information. Using this approach, the attack model must include a map and guesses the user’s position by analyzing the cloaking areas of the user’s consecutive queries and excluding non-reachable areas from the obfuscated area. This map matching is an attack against the spatial-cloaking approach and poses a serious threat to location privacy [2,3]. Furthermore, to obtain higher quality of positioning, the degree of cloaking (and so location privacy) must be reduced [5].

Selecting dummy locations is another challenge for spatial cloaking and position dummies approaches. By monitoring the queries of a user, an adversary can potentially obtain some side information such as the user’s query probability related to location and time and information related to the semantics of the query. Traditionally, dummy locations are generated randomly [3,6] or using a virtual circle or grid model [7], but the generated locations usually release side information. For example, careless dummy
generation may cause some dummy locations to fall at unlikely locations, such as lakes, rugged mountains, or illogical places in time and can be easily filtered out by the adversary [5].

To overcome these limitations, we propose a method of location privacy that employs the benefits of the spatial cloaking approach in mobile cellular networks. To implement this model, we use an LTE network as a case study, but it is not limited to LTE and can be generally used in any networks which benefit cellular technology. In a cellular network (e.g., GSM, UMTS and LTE), system information is broadcasting periodically by each base station (e.g., eNodeB (eNB) in LTE). For instance, to initial access to LTE, a mobile device (or cell phone) starts the cell search and cell selection procedures to synchronize the network and select a serving eNB (SeNB) after power-up [8]. After these steps, the mobile device obtains the needed system information, such as cell ID, cell bandwidth, neighbor cells information, etc [9]. To support users’ mobility in the cellular network, a group of eNBs in the vicinity of every eNB is defined as a neighboring group to handover the mobile connections. This information is publicly available for users in the network. Based on this property, we define our model based on the SeNB and its neighboring group.

In the proposed approach, eNBs, as landmarks of the cellular network, are used to create a spatial cloaking. To extend the cloaking area, we designed an algorithm to choose neighboring base stations as appropriate dummy locations to create an anonymity set. We present a novel method in which, for every nearest neighbor query, especially for POI services, an eNB is mentioned by the user instead of its location. In the response, the SP replies a group of suggested POIs related to each eNB in the neighboring group of the mentioned eNB. Consequently, user’s location privacy can be protected against an adversary that has side information, e.g., an untrusted SP who can monitor the user’s queries. The main contributions of this paper are summarized as follows:

- We propose a framework to protect user privacy for nearest neighbor queries by providing a reasonable cloaking region with eNBs as dummy nodes. Those locations are appropriate in a cellular network because they are well distributed and sufficient in number to cover any area to represent a user’s position without exposing information about the user’s location to an adversary.
- To analyze our model’s applicability, we implement the proposed framework and evaluate its performance with a real dataset of base stations in the city of Berlin and query results using Google APIs.

The rest of this paper is organized as follows. In Section 2, we briefly describe related work. In Section 3, the preliminaries of our method are explained. In Section 4, we present methods to select base locations and the conditions to prevent dissemination of side information, and in Section 5, we analyze the anonymity provided by the proposed scheme. The implementation of the proposed model and its results are discussed in Section 6. In Section 7, we provide our conclusions and discuss future work.

2. Related work

Privacy issues in LBSes have been studied (e.g., location privacy in social networks [10,11] and privacy in cellular networks [12,13]). Generally, user privacy is defined as a triple set of spatial information (location), identity, and temporal information (time) [2]. Depending on privacy preference, several methods have been proposed, each with its own applicability and limitations. Although all of these methods aim for effective functioning of LBSes, they gradually erode service quality [14].

To achieve privacy using a k-anonymity approach, dummy locations have been chosen in the presence of a TTP so as to release the smallest amount of side information to an advisory (e.g. [3,7]). To avoid the trustworthiness limitation, other studies investigated generation of dummy locations without TTP involvement, as in our method [14–16]. Others [17,18] implemented a decentralized approach like peer-to-peer communication or a Hilbert curve to perform user anonymizing. In some concepts, our model is similar to the method presented in [19], in which users measure the distance between the current position and the positions of other users based on the measurable WiFi signal strength to form an obfuscation area using a k-cluster set.

In a spatial cloaking approach, the accuracy of positioning is decreased by the user to achieve privacy without a TTP [2]. Since the user defines the obfuscation area, some nearby POIs may be excluded. Therefore, some papers, like [20], have proposed solutions to optimize cloaking and search for the nearest neighbor. In [21], an approach called SpaceTwist was proposed to answer a k-nearest-neighbor query in which, instead of the user’s position, a fake location called an anchor is selected for the query, and the SP incrementally returns interest points in ascending order of distances from the anchor. The user then calculates the query results based on his/her precise position and the received data. The user chooses the best of the suggested points although incurring high query and communication costs. In our work, we initially use the SpaceTwist idea of a fake location instead of the user’s precise location. We then extend our protection solution to create a k-cluster group of special locations in the cellular network as dummy locations to prepare an anonymity set for this fake location. The main novel aspects of our model are utilizing the geographical properties of the cellular network and creating spatial cloaking and dummy locations to achieve anonymity, such that the identity of the fake locations become anonymous to the SP. Whereas communication cost is a drawback of SpaceTwist, our model achieves location privacy with lower query and communication costs and acceptable precision.

Mobile cellular networks have some unique properties that are beneficial for location concepts and services. For instance, location-based encryption, or geo-encryption, is performed using a GSM cellular network’s positioning features in order to improve security [22]. Similarly, our model avoids any changes to the network infrastructure and uses unique geographical properties of the eNB as widespread landmarks that can be considered to be fixed, along with related basic locations, to achieve location privacy at a reasonable cost.

3. Preliminaries

In this section, we introduce the basic concepts and motivation of our scheme and describe our model. Then, we define the adversary model and its ability to obtain user information.

3.1. Basic concepts

Generally, the adversary can obtain two types of side information, partial information, which is directly related to the user, and global information, which is related to the service and query history. Logically, to minimize discovery of partial information, the user should be aware of global information and select some real or dummy locations that have no clear relationship to his/her location [3]. Practically, however, the historical knowledge is not easily available for an ordinary user; however, based on the used services and the selected dummy locations, acceptable anonymity is achievable. Thus, we create a cluster of m real locations in the vicinity of the user in the cellular network and query POIs for them.

To achieve anonymity, every user has an identity for a service that has no relationship to its network identification. Accordingly, the SP has no information regarding the user’s real position. To
conceal the location information of the user, we select \( m \) proper neighboring locations (e.g., proper distance from the user in the service area or surrounding the user) and use their locations as the anonymous group. Therefore, the adversary does not receive the user’s exact location and must use side information to estimate it. We aim to select \( m \) locations with the same properties to avoid difference among the query probabilities. In this situation, the attacker cannot detect and filter out unrelated locations but can only relay other information. This will be discussed in Section 5.

3.2. Motivation and model description

By introducing the global positioning system (GPS) technology into cell phone, LBSes are becoming more popular. In this paper, we assume that network operators and service providers are independent network entities. In this model, a network operator provides users with cellular services which are independent from location-based applications requiring a client’s location information at the application layer. In many real-world LBSes (e.g., Google Now, Google Maps, Foursquare), our assumption is valid; service providers provide their own services based on users’ GPS and local access point (AP) information rather than cellular technology [5]. If this assumption is not valid, the security of the proposed framework is not guaranteed. We note that the user’s location information can be leaked in cases the location service provider and telecommunication operators collude.

In practice, the nature of cellular communication causes network operators to be aware of all users’ approximate locations in the network, through location update or time advance (TA) update progress [9]. On the other hand, the service provider typically manages user accounts at the application layer. Based on this, we consider two separate IDs for each user: network ID (e.g., IMSI for the cellular network) and service ID (e.g., Google account for Google Maps). The network ID is employed by the network operator for all activities related to the user. Therefore, the user queries any service in the application layer, and the service provider serves the information based on the user’s service ID, which has no link to the network ID or the user’s location. For each LBS, a user has a different service ID which is individually managed by each service provider.

In cellular networks, the operator seeks to completely cover the network area, so each mobile user can access multiple base stations, eNBs [23]. The positions of eNBs are fixed and there is enough number of them in the vicinity of a user where can be seen as landmarks. We use these positions as basic locations and use them to query POI service.

Our model selects \( m \) positions with the proper relationships to the user’s position and that completely surround it. To perform this, a cluster of eNBs is selected to query POI service. We suggest eNBs because:

- They are ubiquitous in the network and usually are distributed according to densities of population and traffic.
- The adversary has no chance to detect and filter out any improperly selected dummy locations that may fall at unlikely locations such as lakes.
- Although the locations of eNBs are fixed and can be easily detected, there are enough of them that we can properly select \( m \) of them to decrease the chance of detection by an adversary.

In the location selection process, the attacker can analyze semantic information about the locations of the cluster nodes to obtain side information and determine node diversity. However, diverse position information provides better privacy than homogeneous position information [2], so a group of well-related locations should be chosen. In this scheme, the users are equipped with GPS positioning and maps that can match suggested POIs from the SP to choose the points in their best interests.

The major technical contributions of this method are as follows:

- To guarantee location privacy, we never use the user’s exact location. Instead, we use the location of SeNB, and choose proper eNBs in the anonymity group enquiring the user in order to create a proper cloaking area.
- To prevent side information leakage, we provide an algorithm to select the members of the anonymity group. A neighbor eNB of SeNB is selected as a central node, and the neighbor nodes of this node (consists of SeNB) compose the anonymity group. To prevent leakage, we set conditions to select the central node as will be described in Section 4.

A major feature of this method is that it uses the available features of cellular networks to achieve location privacy.

3.3. Adversary model

Obtaining the user’s location is the main goal of the adversary in our model. In this scheme, we assume the SP is the adversary and is able to obtain global information and monitor queries sent by users. The SP (adversary) also knows the location privacy protection mechanism used by the user and tries to obtain the user’s location using the following methods.

Knowing the positions of the eNBs in the LTE network, the adversary can use map matching to match these locations on the map to reduce the obfuscation area introduced by the user. We assume the adversary employs the shrink region attack [24] to discover the position of the user. The adversary monitors statistical information about the user’s mobility and consecutive updates or queries and the corresponding members of the eNB set. If the members of the set change, an attacker can follow the user’s direction of motion and can infer the location of the sender of the update or query. Furthermore, the attacker can obtain other sensitive information about the user like travel velocity. Because the success of the adversary depends on statistical information, computing the cloaking region based on actual locations does not necessarily improve the user’s location privacy and is neither efficient nor effective [25].

4. Location selection methods

Selecting proper dummy locations is important to minimize information leakage to the SP. To this end, we classify the eNBs from the viewpoint of the base station that serves the user, SeNB, into two groups: neighbor and neighbor-of-neighbor. For instance, in the example shown in Fig. 1, a mobile user is served by eNB-1121 (as SeNB) and eNBs: 1153, 1102, 1112, and 1129 are the neighbors of SeNB; and the rest are neighbors of neighbors.

In the selection of \( m \) eNBs near the mobile user to act as dummy locations, the following questions must be answered. To
best cover the location of the user, which eNBs should be selected? SeNB should be selected to have accurate POIs, but what about the other $m-1$ locations? If SeNB and its neighboring nodes are selected as dummy locations, the SP can easily detect SeNB as the central node and determine the user’s presence area. Therefore, it is better to select a neighbor node as a fake location and its neighbors, which include SeNB, as dummy locations. We call this fake location the central eNB (CeNB) because the other dummy locations are in its neighbor group. Thus, we must determine which node is the best choice for CeNB.

4.1. Selecting CeNB

By selecting a neighbor of SeNB as CeNB, it is not easy for a SP to detect the SeNB. In the view of the SP, all the neighbors of CeNB have the same probability to be the SeNB. Furthermore, the selected nodes consist of eNBs from both groups, neighbor and neighbor-of-neighbor. Adding the neighbor-of-neighbor nodes expands the cloaking area and improves the privacy but is not helpful for positioning accuracy. The accuracy of the model depends on SeNB and its neighbors because the mobile user is surrounded by these nodes.

Our experience shows that the best choice for CeNB is a neighbor of SeNB that has the largest number of common neighbors with SeNB or has the highest signal strength for the user. This selection increases the accuracy and provides the best environment for the user with the best choices of nearby POIs. Therefore, we select the eNB with the strongest signal after that of SeNB in the receiving signal strength list to be CeNB.

Fig. 2 illustrates our model’s node selection for the previous example. In the receiving signal strength list of the mobile user, eNB-1112 (−85 dBm) is located in the second position after serving cell; eNB-1121 (−75 dBm), so this node is selected as the fake location (CeNB), and its ID is sent to the SP. In the next step, the SP extracts members of the group neighboring CeNB and replies the POIs near these locations (seven dummy locations). In this dummy location set (Fig. 2), other than SeNB-1121, there are nodes from the neighbor group (eNBs: 1102, 1112, and 1129) and nodes from the neighbor-of-neighbor group (eNBs: 1109, 1110, and 1124).

5. Threat analysis

To hide the mobile user’s location, we use SeNB to create an obfuscation area. Furthermore, by selecting a fake location (CeNB) and some other dummy locations, we create an anonymity group to conceal the identity of SeNB. To shrink the cloaking region, the SP tries to identify SeNB, which is the nearest location to the user. To this end, the SP must find SeNB from $m-1$ equally probable positions in the dummy set.

5.1. Privacy analysis- privacy value

To analyze the produced privacy protection, we use the concept of privacy value defined by Yi et al. [21] as the average distance from uniformly distributed locations in a cloaking area to the user’s actual location. Based on this definition, if $q$ is the user’s location and $\varphi$ is a cloaking area, the privacy value $\gamma(q, \varphi)$ is:

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

(1)

This value is based on the distances between the user’s location and the other locations (shown by $z$) in the cloaking area. Therefore, a high value indicates that most of the adversary’s selected locations in the cloaking area are far from the user [21]. The cloaking region is used to normalize the privacy value, and relatively large areas lead to higher values.

5.2. Privacy analysis- anonymity degree

In the next step, we create an anonymous set to try to hide the identity of SeNB. Based on the definition presented by Pfitzmann and Kohntopp in [26], anonymity is the state of being unidentifiable within a set of subjects called the anonymous set. A sender is identifiable when information that can be linked to the sender is obtained. The maximum degree of anonymity is achieved when an attacker sees all subjects in the anonymous set as having an equal probability of being the originator of a message [27].

If the query probabilities of the BSs are the same, the adversary sees all nodes as having an equal probability of being SeNB, the user’s serving node. Accordingly, the SP has to guess the locations from among a set of $m$ locations, each with probability $1/m$.

To ensure both accuracy and privacy, proper locations must be chosen from the both groups of neighbor and neighbor-of-neighbor. Their selections represent a tradeoff between accuracy and privacy regarding user position and the distribution of eNBs. If we view every eNB in the dummy set as an information point, the anonymity of the eNB in this model can be measured based on the information entropy of all eNBs. We use $X$ to denote the anonymity model and $H(X)$ as its entropy value. Suppose $p_i$ is the probability of identifying the $i^{th}$ eNB as the true SeNB; then,

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

(2)

When $H_M$ is the maximum entropy of the system we want to measure for the size of our dummy set, $m$:

$$H_M = \log_2(m)$$

(3)

This maximum entropy is achieved when the query behaviors of all eNBs look the same to the SP. So every station has the probability $1/m$ of being identified as SeNB.

The information that the SP can achieve with an attack can be expressed as $H_M - H(X)$, which is normalized by dividing by $H_M$.

The information that the SP has achieved for all eNBs is equal to the information we want to achieve for this system. We define the anonymity degree for the system presented by Diaz et al. [28] as:

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

(4)

The value of $d$ indicates the information leakage in the model. We consider the SP to be an active attacker who monitors all queries in the network and who can make use of some conditions, such as wrongly selecting SeNB as the central node or an inappropriate geographical distribution of eNBs’ position in the dummy set, to recognize SeNB. The value of $d$ will be minimal ($d=0$) if an eNB in the dummy set appears to be SeNB with probability $1$, whereas if all nodes have the same probability of being SeNB ($p_i = 1/m$), $d$ will have its maximal value ($d=1$).
5.2.1. Anonymity degree of a proposed system

As an example of the anonymity degree of a system, consider the model shown in Fig. 2. In this model, a user with a defined service ID queries POIs for a group of locations consisting of seven eNBs \((m = 7)\). The adversary tries to find SeNB using map matching and the shrink region attack and by assigning a probability to each location. Due to the SP ability to monitor user’s query history, he is able to assign probabilities to each eNB location. Let \(p_i\) \((i = 1 \text{ to } 7)\) be the probability assigned to these seven locations. In this case:

\[
\sum_{i=1}^{7} (p_i) = 1
\]

The maximum entropy of this model is:

\[
H_M = \log_2(7)
\]  

(6)

Based on the query history of the user, the SP knows some information related to the eNBs, which helps in the assignment of a probability to each station. For instance, sequential queries sent by a user may consist either of the locations of eNBs that are mentioned more often than others or of the locations of eNBs that have been rarely mentioned. Furthermore, some eNBs, depending on their positions, have fluctuating traffic that may become very low at some hours; for example, the traffic of an eNB in a subway station at midnight.

The selection of these kinds of eNB assists the SP to assign the nodes with the lowest probability. In this case, the SP has two groups of locations, one with nodes mentioned frequently, which can be deduced to belong to the neighbor group, and the other with rarely mentioned nodes. The following example shows the influence of this improper selection on the anonymity degree of the model.

Consider a neighboring group that consists of seven eNBs and an SP, who by supervising the query history of the user, suspects certain nodes to be members of the neighbor-of-neighbor group (because of their rare appearance). By considering the probability \(p\) that a group of eNBs includes SeNB, the SP classifies the nodes into two groups, with probabilities of \(p\) and \((1 − p)\), respectively. The nodes that belong to the same group are seen as having the same probability, and the SP assumes a uniform distribution such that \(p_i = p/n\), where \(n\) is the number of eNBs in this group, for each eNB in the first group, which is suspected to include SeNB, and \(p_i = (1 − p)/(7 − n)\) for each member of the second group. So the degree of anonymity is:

\[
d = \frac{(\sum_{i=1}^{n} \frac{p_i \log_2 \frac{p_i}{n}} + \sum_{j=1}^{7-n} \frac{(1-p) \log_2 \frac{1-p}{7-n}}}{\log_2(7)}
\]

(7)

The variation in anonymity degree with \(p\) and \(n\) is illustrated in Fig. 3. The value of \(n\) is varied from 1 to 6 to show all possibilities of the model from worst to best case. As shown in the figure, the number of eNBs that are rarely selected in the user’s query \((7 − n)\) and the ability of the SP to detect this kind of node \((1 − p)\) are two factors affecting the degree of anonymity.

Comparing these curves shows that the number of eNBs that guarantees a degree of anonymity higher than 0.6 is \(n = 3\) or 4. This is true even if the SP definitely detects the eNBs of the neighbor-of-neighbor group. This value of \(n\) achieves the highest anonymity \((d = 1)\) if the SP achieves a detection probability of \(p = 0.3 \sim 0.6\). Definitely, anonymity degree depends on the number of members of the anonymity set \((n)\) and detection probability \((p)\). Based on this, the following condition can guarantee the reasonable anonymity degree more than 0.5:

\[
n \geq (m - 1)/2
\]

(8)

where \(n\) indicates the number of neighbors of SeNB in \(m\) dummy locations. Accordingly, to have an acceptable anonymity degree, the minimum value of \(m\) is six.

![Fig. 3. Variation of degree of anonymity with \(p\) and \(n\), where \(n\) is the number of eNBs in the neighbor-of-neighbor group.](image)

6. Experiments

To evaluate and scrutinize the performance and security of this model, we use a location dataset of base stations of a cellular network provided by OpenMobileNetwork in the city of Berlin, Germany. OpenMobileNetwork is an open solution providing an approximated and semantically enriched mobile network and WiFi access point topology data based on the principles of linked data. Since the quality of the estimated network topology is crucial when providing services, OpenMobileNetwork provides acceptable base station positions from crowd-sourced data in comparison to real base-station locations. The processed data in this dataset are structured according to a comprehensive ontology in order to provide information about base station and WiFi access point locations, their coverage areas, neighboring cell relations, and dynamic data, such as traffic, service usage, and number of users [29].

6.1. Model requirements

In our model, the user uses a smartphone equipped with a GPS receiver to calculate his/her position and a map to match suggested POIs. In the server side, we used three separate datasets: real locations of 153 base stations on a T-Mobile network in Berlin, neighboring groups of this operator, and POI locations found using Google API.

Unlike [21], instead of one fake location, to achieve better privacy, we use an \(m - \text{cluster}\) location set not only to provide location ambiguity, but also to cover the user and increase the accuracy of the query. As an example, Fig. 4 illustrates positions of cafes as POIs in the vicinity of the user (indicated by •). To request a POI, the user selects an eNB as CeNB (described in Section 4) and sends its ID to the SP. The SP retrieves the neighboring group of the CeNB and the members’ locations, respectively, from the neighboring group and the base station location datasets. Finally, the SP retrieves POI locations in the vicinity of each location and replies to the user. To this end, we exploited the GoogleAPI to find the cafes.
In the next step, using its own positioning ability, the user calculates the cafe nearest to his/her location. In this stage, the accuracy of positioning depends on the user’s positioning ability.

6.2. Performance evaluation

To evaluate the performance of the model, we focus the evaluation on these concepts:

- POI positioning accuracy
- User location privacy
- Communication cost

To determine the optimum case, we examine several user situations. The following subsections explain these in detail.

6.2.1. POI positioning accuracy in comparison to Google places API

To protect the user’s location privacy, we conceal his/her location using eNBs in its vicinity. How then can we achieve the best accuracy as the SP suggests POIs based on eNBs? In the nearest neighbor service for a user, the user’s location is used as the center of a circle with a defined radius, and n nearest POIs are suggested, e.g., using Google Places API. On this basis, we should determine the proper radius for positioning POIs that are near the user. To this end, we evaluate three solutions of positioning:

1. n nearest points
2. fixed radius
3. variable radius

In the first solution based on each eNB, the SP replies n nearby POIs, where n is determined by the user. In this way, the user can control the data size to be downloaded. In the second solution, the SP suggests POIs within a fixed radius of each eNB. In the third solution, the radius is not fixed but is dynamically chosen by the user. Fig. 5 shows the concept of POI positioning based on eNBs in a cellular network.

To analyze the accuracy, we compare the selection solutions in our model with the Google Places API model in which POIs are found in the area centered at the user’s location. Similar to this model, the user in our model chooses the best of the suggested POIs. In our analysis model, if m denotes the true number of POIs near the user (n nearest service) and \( \tilde{m} \) denotes the number of these POIs suggested by the SP in the selection model, the accuracy of this selection model is:

\[
A = \frac{\tilde{m}}{m} \times 100 \quad (9)
\]

In Figs. 6 and 7, the accuracy of the positioning solutions of fixed radius and n nearest POIs are illustrated.

According to these results, the fixed radius solution with a radius of 800 m achieves an accuracy of 98.57%, and the solution of n nearest POIs with \( k' = 8 \) (\( k' \): the number of POIs near each eNB) achieves an accuracy of 97%. These examples were selected to perform the comparison because of their acceptable accuracy for all values of k (the number of POIs chosen by the user) in the nearest neighbor query service. The positioning accuracies of these three solutions are compared in Table 1.

Due to traffic intensity and network design, the distances between eNBs vary differently in downtown and suburban districts,
so some POIs may be iterated for some eNBs. In this case, in the mass area, the user selects the proper radius depending on the distances from SeNB and central eNB. In this way, using the dynamic radius model, the user can manage this distance. However, generally, the user does not have knowledge about the cellular network structure and can only measure the distances to the eNBs that receive the user’s signals. To this end and to determine the optimum radius without releasing side information, we consider a surrounding triangle with vertices of SeNB, CeNB, and the eNB that has the third highest signal strength in the user’s beacon. As shown in Fig. 8, the average distance to the vertices is used as the dynamic radius.

In this way, the user can balance communication cost and accuracy, while achieving reasonable accuracy. Theoretically, the dynamic radius model is better from the perspective of communication cost; however, according to the results provided in Table 1, this positioning model has the lowest accuracy.

6.3. Location privacy of the user

To analyze the privacy protection, we address it as two separate concepts of evaluation: spatial cloaking and anonymity.

6.3.1. Privacy value in comparison to the standard cloaking model

First, we evaluate the privacy of the proposed model and compare it with the privacy of the standard cloaking model introduced by Ardagna et al. [4], which uses circular cloaking area instead of the user’s precise location. As described in Section 5.1, we use the concept of privacy value for this comparison. Unlike the standard cloaking model which uses the circular cloaking area, we used the polygon-type cloaking area to improve the measurement accuracy where a polygon consists of eNBs and its network structure can be changed depending on its neighboring cells. To get some insight into this concept, Fig. 13 shows the effects of geographic properties of network structure on the privacy value. The density of base stations and their distances generate different cloaking area and provide different privacy values with the network density. Fig. 9 shows Cell-4723 and Cell-64360 from T-Mobile and their neighboring groups in the city of Berlin. Here, Cell-4723 (at the intersection of Altonaer Straße and Lessingstraße roads) and Cell-64360 (in the vicinity of the Bundeswehrkrankenhaus Berlin hospital) are used to represent the low- and high-density regions, respectively. For the cloaking area of Cell-4723’s neighboring group, the mean privacy value is 7.48 whereas that value drops to 5.65 for the cloaking area of Cell-64360. Therefore, in the dense area (Cell-64360) where base stations are located close together, the cloaking area is relatively smaller and privacy value could be decreased compared with wider cloaking area (Cell-4723).

Since the cloaking region $\varphi$ has no specific equation or form in the general case, we use the Monte Carlo method to approximate the distance to the user’s location in $\varphi$ and employ these two models over 150 users from different neighboring groups. Fig. 10 shows the histogram of privacy values achieved by our model over these users. In Fig. 11, the average privacy value achieved by our model is compared to that achieved by the standard cloaking model. The
average privacy value of 5.70 obtained by our model is comparable to the privacy value of the standard model at a radius of almost 650 m, which offers good security for users.

6.3.2. Anonymity degree

We measure the anonymity provided by using the anonymity degree, introduced in Section 5.2. As described there, our method employs a set of \( m \) related dummy locations as a neighbor group to form an anonymous set. Therefore, the SP faces on average six locations (except the CeNB) as possibilities for an SeNB with the same query probability. Based on the query history of the user, even if the SP detects the eNBs as being part of the neighbor-of-neighbor group and filters them out, the SP cannot detect the location of the user because the user is under the cloaking of SeNB with the privacy degree determined in the previous section.

Because they attain the same spatial ambiguity, all three methods of POI positioning examined in this work have the same privacy degree. According to our adversary model, we suppose in the worst case that the SP, employing map matching and shrink region attacks, detects the eNBs of the neighbor-of-neighbor group in each set. Since the SP knows that CeNB cannot be SeNB, there is a probability of \( 1/(m-1) \) to detect SeNB in a CeNB neighbor group with \( m \) members. The number of eNBs in the CeNB neighbor group is in direct proportion to the anonymity degree achieved with this model. For Cell-4723 having four neighbor cells in Fig. 13, the SP has a higher chance to find SeNB compared with Cell-64360 having ten neighbor cells \( (m = 10) \). Therefore, in the neighbor group of each CeNB, the number of members depends on the network topology and structure. Hence, we use neighbor groups with different numbers of members \( (m = 6, 7, 8, 9, \) and 10 \) in the dataset of base locations.

Fig. 12 illustrates the anonymity degree of SeNB in the neighbor groups of CeNB with different numbers of members. For larger numbers of members in the neighbor group, the degree of anonymity is more highly increased when the SP has less knowledge about the members of the neighbor-of-neighbor group and is more highly decreased when the SP can detect the eNBs of this group. As shown in the figure, the degree of anonymity of SeNB does not decrease to zero even when the SP detects the eNBs of the neighbor-of-neighbor group, because the SP must still identify the SeNB among two or more similar choices, depending on the neighbor group members.

6.4. Communication cost

The communication cost directly relates to the size of a message downloaded by a user during a typical POI query. During each query, the SP sends one or multiple messages to the user, depending on the data size. In our simulation, the response message to a query consists of names of cafes, their location information (latitude and longitude), and description. The size of the message varies depending on the number of cafes in the proximity of the eNBs, especially on the number of eNBs in the neighbor group of CeNB. It should be noted that the number of POIs and the number of eNBs in CeNB’s neighbor group have direct effects on the data size.

Fig. 13 shows the message sizes of fixed radius and \( n \) nearest POIs positioning methods; the dynamic radius model has an average message size of 1168.8 Bytes. The message sizes of these three POI positioning methods are compared in Table 2. The message of the \( n \) nearest POIs positioning method with 467 Bytes has the smallest size among these methods.

7. Analysis and conclusions

In this paper, we proposed a framework to protect user’s location privacy for LBSes by utilizing the basic geographical features of cellular networks. Based on these features, we use the cloaking area of SeNB to conceal the user’s location, and we hide the
identity of SeNB in a group of dummy locations in the neighbor group of CeNB.

The evaluation showed that user privacy can be guaranteed using a cloaked region. Even if the adversary can filter out the eNBs of the neighbor-of-neighbor group and has a high probability of finding SeNB, our approach achieves an anonymity degree greater than 0.3 in the worst case. Our experiences show that, although a cluster group with a larger number of nodes has a wider cloaking area, the model will be more vulnerable if the adversary obtains clues (from side information) to filter out the members of the neighbor-of-neighbor group.

To implement the proposed concept, we tested three POI positioning solutions that achieve the same privacy protection but have distinguishable differences in positioning accuracy and cost. With respect to the precision of the POI positioning, the solution of fixed radius positioning with a radius of 800 m achieved a greater than 98% accuracy and demonstrated better performance than the n nearest POIs solution, which achieved almost 97% accuracy, and the dynamic radius solution with almost 54% accuracy. However, the user pays the cost for the accuracy achieved with the fixed radius solution by being required to download a larger file. We observed that the message size of the fixed radius solution is almost 14 times larger than the n nearest POIs message and five times larger than the dynamic radius message. We conclude that the n nearest POIs solution of positioning, with an average message size of 467 Bytes and precision of almost 97%, is the best solution for POI positioning. This solution achieves location privacy for users while balancing accuracy and cost.

In the proposed model, we tried to exploit the publicly available features of cellular network without imposing any changes. For doing this, network operators offer either APIs or the information about their cellular network topologies so that a service provider can obtain the neighboring cells of a given base station. Also, the proposed framework may inherently incur an expensive communication overhead.

The proposed framework may inherently incur an expensive communication overhead because the response size could be quite large while the request size is typically very small (i.e., POI and CeNB). That is, users can often be burdened with a high communication cost compared with conventional POI services. Moreover, considering our assumption that network operators and service providers are independent network entities, network operators offer either APIs or the information about their cellular network topologies so that a service provider can obtain the neighboring cells of a given base station. This may require some degree of cooperation between network operators and service providers.

In further development of our proposed model, we will expand this framework for users’ location privacy protection to cases of emergencies. While respecting users’ location privacy, we will develop an approach to appropriately release users’ location information to disaster management agencies in disaster situations.

Acknowledgments

This work was supported in part by the NRF Korea (No. 2014R1A1A1003707), the ITRC (IITP-2016-R0992-16-1006), and ICT R&D program (No. B0717-16-0116). Authors would like to thank all the anonymous reviewers for their valuable feedback.

References