Knowledge-Based Systems 103 (2016) 19–27

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Location privacy-preserving *k* nearest neighbor query under user's preference

Weiwei Ni*, Mingzhu Gu, Xiao Chen

School of Computer Science and Engineering, College of Software Engineering, Southeast University, Jiulonghu Campus, Jiangning District, Nanjing 211189, P.R.China

ARTICLE INFO

Article history:
Received 21 August 2014
Revised 20 March 2016
Accepted 21 March 2016
Available online 16 April 2016

Keywords:
Location based service
Location privacy model
Privacy preference
k nearest neighbor query
Farthest POI attack

ABSTRACT

Location-based services can provide users' surroundings anywhere and anytime. While this service brings convenience for users, the disclosure of user's location becomes the main concerns. Most current practices fall into *K*-anonymity model, in parallel with location cloaking. This schema commonly suffers from the following constraints. (1) *K*-anonymity cannot support users' preferential query requirements effectively. (2) location cloaking commonly assumes that there exists a trusted third party to serve as anonymizer, which is inclined to be the bottleneck of the query. Concerning these problems, a novel location privacy model (s, ε)-anonymity is devised from perspective of minimum inferred region and candidate answer region, which present location protection strength and scale of intermediate results, respectively. Particularly, user's preferential query requirements on privacy protection strength and query efficiency can be presented in a more convenient and effective way by setting parameters s and ε rather than *K*-anonymity model does. A thin server solution is developed to realize the model, which pushes most workload originated from user's preferential requirement down to client side leveraging false query technology without any trusted third parties' intervention. Furthermore, an entropy based strategy is devised to construct candidate answer region, which boosts privacy protection strength and query efficiency simultaneously. Theoretical analysis and empirical studies demonstrate our implementation delivers well trade-off among location protection, query performance and query user's privacy preference.

1. Introduction

Location-based services (LBS in short) in parallel with various applications of location-aware devices (e.g.,GPS devices) have gained tremendous popularity [1,2]. *k* nearest neighbor (*k*NN) query is an important class of LBSs, which periodically returns the *k* nearest neighbors, say point of interests (POIs in short), in relation to query user's current location. For example, a tourist may query the *k* nearest restaurants while exploring a city. While LBSs provide conventional services to query users, it threatens user's privacy as users are forced sharing their location with service provider [3,4]. Hence, how to provide location-based services while protecting user's location privacy has recently become a hot topic.

Existing solutions fall into three categories, namely, spatial cloaking [5–9] space transformation [10–13] or location obstruction [14–16]. The common of these schemes trade-offs among query performance, protection strength and query accuracy. In recent years, region cloaking [5–9,17–24,26] witnesses its wide prosperity. In detail, when a user initiates a *k*NN query, she sends her location and privacy requirement to a trusted third party instead of the service provider. At the trusted third party, user's location coordinates are replaced with a cloaking region which encloses the user and satisfies privacy requirement such as spatial *K*-anonymity (*SKA*) and the minimum inferred region (*MIR* in short) [6,9,10], namely the minimum bound of the range that the user's possible location can be derived by attackers. Subsequently, the third party submits a cloaking region based *k*NN query to the service provider and receives the returned candidate answers. Finally, the actual answer can be pinpointed by the trusted third party or the user itself. Although, region cloaking based solutions afford well protection, they suffer from the following constraints.

1. Most current practices fall into *K*-anonymity based location privacy model. *K*-anonymity based model deeply relies on users' dynamic distribution and cannot support users' preferential query requirements effectively.
2. A trusted third party is requisite for most cloaking based solutions to act as anonymizer. All users must trust the anonymizer, which becomes a single point of attack;

* Corresponding author.
E-mail addresses: wni@seu.edu.cn (W. Ni), gumingzhu@seu.edu.cn (M. Gu), chenxiao@seu.edu.cn (X. Chen).

http://dx.doi.org/10.1016/j.knosys.2016.03.016
0950-7051/

3. Complex server-side query processing is needed to determine candidate answers. This deteriorates query performance seriously.

These problems highlight the needs of designing novel location privacy model and privacy preserving scheme, which abandons brute-force enlarging way that region cloaking method adopted, as well as intervention of online trusted third party. In this paper, a location privacy model (s,ε)-anonymity is defined from view of *MIR* and *RCA*.

Existing cloaking based solutions commonly deploy the work of generating *MIR* at trusted third party and regulate the relation between *RCA* and *MIR* in a brute-force way. Our model continues to use *MIR* to present user's privacy protection requirement but generates it at client side in an implicitly user-controllable way. Besides, area ratio parameter between *RCA* and *MIR* is used to regulate user's requirement about query efficiency. The query process consists of two rounds; at the first round, a detecting query is initiated to get local POI information at server side. It generates a circle from the returned answers to cover both the targeted POIs and the user's preferred inferred region; at the second round, query users resend optional blurred regions and receive all POIs inside it. The targeted POIs can be immediately pinpointed out at client side. Further, to improve location protection strength, a rigorous initial region creation solution is proposed by initiating an extended detecting query, in parallel with an entropy based strategy to specifying center of *RCA*. Our solution can afford well location protection and good query performance, simultaneously.

Our main contributions can be summarized as follows.

1. A novel location privacy model (s, ε)-anonymity is proposed from perspective of minimum inferred region and region of candidate answers, which abandons heavy dependence on users' real-time distribution. It can incorporate privacy preference into privacy protection nearest neighbor querying well.
2. A cloaking and location obstruction based solution is devised to realize our privacy model in a thin-server way. It pushes most workload down to client side to overcome query scalability problem originated from preferences. An entropy based strategy is deployed for specifying center of *RCA* to improve location privacy protecting strength.
3. Empirical studies suggest that our location privacy model is effective and the proposed solution is highly performant.

The rest of the paper is organized as follows. Section 2 presents overview of related work. Section 3 gives the definition of our novel location privacy model and proposes a user-controllable framework *AnPNN* to realize it. Section 4 illustrates the algorithm *AnPNN* and discusses its potential risk of privacy leakage. In Section 5, to improve privacy protection strength, a rigorous version *RAPNN* is devised. Section 6 demonstrates the experimental results of our solution. Finally, Section 7 concludes and identifies research directions.

2. Related work

There has been a plethora of techniques to deal with location protection. Current practices fall into the following ways. (1) Location obstruction [14,15]. The idea is that a user first sends a query along with a false location to the server, and the server keeps sending back the list of nearest POIs to the reported false location until the received POIs satisfy user's query accuracy requirements. (2) Space transformation [10–12,25,27]. This approach converts the original location of data and queries into another space. The transformation maintains the spatial relation-ship among the data and queries to provide accuracy. (3) Spatial cloaking [5–9]. This solution embeddes a privacy-aware query processor at the database server side to deal with the cloaked spatial area received either from a querying user or from a trusted third party. Based on the above technologies, a large body of research has gone into algorithms that enforce location protection in snap-shot location-based query.

Casper [5] is the representative method of cloaking based solutions. Fig. 1 depicts *Casper's* system architecture. It has two main components, namely location anonymizer and privacy-aware query processor. Its procedure is detailed as follows. Mobile users register with Casper by a certain privacy profile that outlines the privacy requirements of each user. A user privacy profile is defined as the form (k, A_{min}), where K indicates that the user wants to be K-anonymous, while A_{min} is the minimum acceptable resolution of the cloaked spatial region. The location anonymizer receives continuous location updates from mobile users, blurs the location updates to cloaked spatial areas that match each user privacy profile (k, A_{min}), and sends the cloaked spatial areas to the location-based database server. Location anonymizer also blurs the query location information before sending a cloaked query area to the location-based database server. The privacy-aware query processor is embedded inside the location-based database server to anonymously deal with cloaked spatial areas rather than exact point locations. It returns a candidate list of answers to location anonymizer rather than the exact one.

Cloaking based solutions can provide user's preference to location privacy by transmitting a user defined profile including user location and expected minimum inferred area A_{min} to anonymizer. Anonymizer then expands user location into a cloaked region with expected area to act as finial *MIR*, and sends the region to the server for retrieving candidate answers. In this way, complex server-side query processing is needed to determine *RCA* in terms of the given *MIR*, which renders complex server-side query processing and poor scalability. In general, it provides user preference in a brute-force way at cost of complex server-side query processing and poor scalability. Besides, the K-anonymous privacy model cloaking based solutions commonly adopted suffers from the following forced requirements. (1) All users are forced to trust the third party anonymizer, which is prone to be a single point of attack; (2) A large number of cooperating, trustworthy users are needed.

Transformation-based solutions generally have at least one of the following shortcomings: (1) Fall short in offering practical

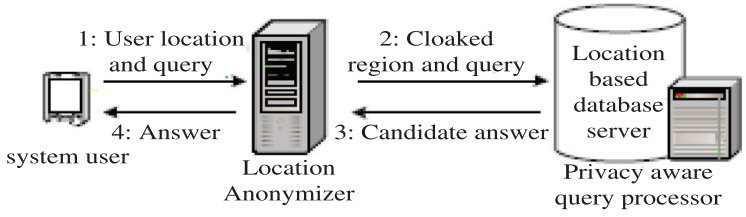

Fig. 1. *Casper*'s system architecture