基于路网环分布的隐私保护近邻查询方法

倪巍伟 冯志刚 闫 冬
(东南大学计算机科学与工程学院 南京 211189)
(东南大学计算机网络和信息集成教育部重点实验室 南京 211189)

摘 要 位置服务在方便人们生活的同时,也带来了隐私安全问题,路网环境对移动对象运动模式的限制使得位置隐私保护问题更为复杂.基于空间混淆的现有路网隐私保护近邻查询方法存在对重放攻击的抵御能力较弱,位置泛化与查询处理效率低,以及处理效率与位置保护安全强度不可调节,难以支持个性化隐私保护查询的不足.针对上述问题,引入路网环分布概念并设计生成给定路网环分布的算法,在匿名服务器端离线生成路网环分布;通过设置基于路网环分布的子网扩张结束条件,提升所提隐匿环机制对重放攻击的抵御能力;借助子网扩张结束条件调控隐匿环子网规模实现对隐匿环生成效率、近邻查询效率与位置隐私保护强度的调节.进一步,结合隐匿环的组成结构,提出可以有效降低搜索起始边规模的剪枝方法,提升基于隐匿环的近邻 POI 查询效率.理论分析和实验结果表明,所提方法能有效抵御重放攻击,提升位置泛化处理与近邻查询处理的时效,在此基础上能够兼顾对查询发起者位置信息保护效果、位置泛化处理与近邻查询时效的调节.

关键词 位置隐私保护;隐匿环;路网环分布;偏好调控
中图法分类号 TP311 **DOI 号** 10.11897/SP.J.1016.2020.01385

Location Privacy Preserving Nearest Neighbor Query Method Based on Circle Distribution on Road Networks

NI Wei-Wei FENG Zhi-Gang YAN-Dong
(*Department of Computer Science and Engineering, Southeast University, Nanjing* 211189)
(*Key Laboratory of Computer Network and Information Integration in Southeast University, Ministry of Education, Nanjing* 211189)

Abstract Location-based services (LBS) facilitate our daily life, nevertheless, they also bring privacy protection problem. The location and identity of the query initiator may be exploited by the attacker based on special query modes and other query related information. The limitation of mobile object's motion mode on road network makes location privacy protection more complex in face of location based nearest neighbor query. Road network spatial cloaking is the representative technology to realize location privacy protection in road network environment. Namely, location of the query initiator, as well as the query content, are expanded to a road subnet satisfying specific privacy protection constraints in an anonymous way, usually done on trusted third party servers. The trusted anonymous server submits the subnet instead of query initiator's real location to LBS server. The LBS server receives the subnet to finish the network expansion search and returns the candidate query results to the query initiator. Existing methods have the following shortcomings. They can not effectively resist replay attacks, there is a risk of location privacy leakage, and the anonymizing cost at the trusted third party is high. Besides, query methods lack attention to the internal relationship of each section in the anonymized subnet, and perform indifferent processing on the boundary nodes of the anonymized subnet, which results in waste of

收稿日期:2018-05-23;在线发布日期:2019-07-12.本课题得到国家自然科学基金(61772131,61370077)资助.**倪巍伟**,博士,教授,博士生导师,中国计算机学会(CCF)会员,主要研究领域为数据挖掘、数据隐私保护.E-mail:wni@seu.edu.cn.**冯志刚**,硕士研究生,主要研究方向为数据隐私保护.**闫 冬**,硕士研究生,主要研究方向为数据隐私保护.

computing resources. Query processing method lacks the regulation mechanism to anonymization performance, query efficiency and privacy preserving intensity, which is difficult to meet the personalized query needs of queriers. Concerning aforementioned problems, the solution of PNQCD algorithm is proposed, which realizes privacy-preserving k nearest neighbor querying on road networks leveraging cloaking circle-based cloaking technology. It adopts the idea of circle based road subnet, called cloaking circle, to act as anonymized road subnet. The concept of circle probability distribution of the road network is introduced, and its generating method is proposed to support anonymized subnet construction work on the trusted anonymous server. The defensive ability against replay attacks and the efficiency of location anonymizing processing are improved from the following folds. Firstly, the circle probability distribution of the network is pre-generated on the trusted third party anonymous server. On this basis, the ending conditions of cloaking circle-based anonymized subnet construction is devised leveraging circle probability distribution scheme to complete cloaking circle construction in advance. As a result, this schema facilitates the tuning ability to anonymization performance, nearest neighbor query efficiency, as well as defense effect on replay attack. On the LBS server side, pruning strategy and search edge sieving strategy are designed in the nearest neighbor querying process according to the ring structure of anonymous subnet. It can effectively reduce the cost of query processing on the server side and improve the query efficiency. Theoretical analysis and experimental results demonstrate that the proposed method can effectively resist replay attacks and has good efficiency of anonymity and query processing. Furthermore, the anonymization performance, query performance and privacy preserving intensity can be regulated according to user's preferences.

Keywords location privacy preserving; cloaking circle; circle distribution on road networks; preferences regulation

1 引 言

位置服务(Location Based Services,LBS)在给人们生活带来便利的同时,也产生了位置隐私泄露问题.近邻查询是 LBS 服务的重要查询模式,近邻兴趣点(Point of Interest,POI)查询中,查询者需将自身位置提交给服务提供方,而位置蕴含用户身份、行为模式等敏感信息,将位置信息共享给位置服务提供方带来隐私泄露风险.近年来,路网环境保护位置隐私近邻查询以其贴近现实应用的原因得到了持续的关注[1-2].

查询者在路网中需要沿路段运动,不同于非路网环境直接用欧几里德距离度量位置坐标间距,路网环境采用路网距离度量位置坐标间的距离.路网环境下,隐私保护近邻查询具有以下特点:(1)路网结构多样,近邻查询处理过程复杂;(2)查询者需沿路段移动的约束,可能为逆推目标对象位置提供了较强的线索.

已有方法多数利用空间混淆(spatial cloaking)实现路网环境兼顾位置隐私保护的近邻查询处理,位置空间混淆的原理是将精确位置坐标泛化为具有更高概括性的数据结构(路网环境即为子网).其处理流程如下:查询发起者将自身当前位置和查询请求提交可信匿名服务器,将其位置泛化为包含该位置的特定子网,匿名服务器将泛化子网与查询内容提交给位置服务器,服务器完成子网近邻查询,将查询候选解返回匿名服务器,供其筛选目标结果.近年来,在针对各类位置隐私泄露场景的位置泛化方面,已取得长足进展[3-14],通过将位置坐标泛化为特定子网实现位置隐私保护.位置服务器端采用的路网近邻查询策略主要是基于宽度优先搜索的扩张查找.针对存在的查询代价较大等问题,也提出了多种位置服务器端查询处理方法[15-19].

重放攻击(Replay Attack)是近年来攻击者侵犯位置隐私的重要形式.其模式是攻击者分析所获取泛化子网结构,对子网中有移动对象的路段采用相同泛化方法生成泛化子网,进而根据该子网和目

标子网的相似度推测目标位置位于该路段的概率，进行逆推攻击. 环结构子网是防御重放攻击的有效方法，文献[14]提出 CCF(Cloaking Circles and Forests)位置泛化方法，对环状泛化子网的组成结构进行约束，使得攻击者利用重放攻击手段推断查询者位于环内任意路段的概率相等，实现位置隐私保护.

当前，在路网环境隐私保护近邻查询方面，已有研究仍存在以下问题：

(1) 已有的针对重放攻击的位置泛化方法时间消耗较大，且仍然存在位置隐私泄露风险；

(2) 位置服务器端对泛化子网各类结点执行无差别查询处理，查询效率较低；

(3) 查询机制缺少对位置泛化与近邻查询处理效率、位置隐私保护强度的动态调节能力，对个性化查询的支持能力较差.

针对以上问题，离线构建路网的环分布信息，通过设计位置泛化结束条件调节隐匿环结构，实现对查询发起者的位置泛化与近邻 POI 查询时效、位置隐私保护效果的个性化调节；结合隐匿环组成结构，提出基于剪枝的近邻查询策略，提高位置服务器端近邻 POI 查找的效率. 基于上述思路，提出基于路网环分布的保护位置隐私 k 近邻查询方法 PNQCD (location Privacy preserving Nearest neighbor Query method based on Circle Distribution on road networks).

论文关键方法包括：

(1) 引入环分布概念，提出环分布计算方法及基于环分布的路网环境位置泛化方法；

(2) 引入终止隐匿环构建约束，实现对隐匿环构建过程的控制，通过调节隐匿环规模，支持对位置泛化与近邻查询效率、位置保护效果的动态调节；

(3) 设计适应隐匿环组成结构的查询优化策略，提高服务器端隐私保护近邻查询效率.

本文第 2 节介绍相关工作与问题描述；第 3 节介绍环分布概念以及基于环分布的隐匿环构建方法；第 4 节介绍位置服务器端的路网近邻查询方法；第 5 节对所提方法进行实验分析；最后，总结全文并展望后续工作.

2　相关工作与问题描述

2.1　相关工作

基于空间混淆的解决方法通常采用图 1 所示架构. 匿名服务器接收查询发起者的查询请求(包括查询者位置、隐私约束以及诸如近邻查询对象数 k 等查询内容)，对查询发起者位置进行泛化处理，生成符合隐私约束的特定泛化子网，将泛化子网提交位置服务器进行关于泛化子网的近邻查询，将查询结果返回匿名服务器，供其从结果中筛选出目标 POI 集合并返回给查询发起者.

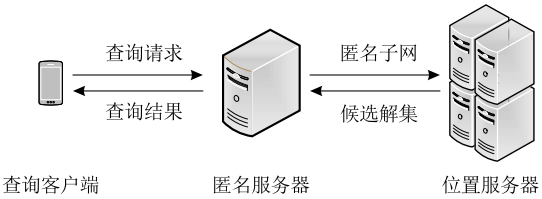

图 1　路网环境保护位置隐私查询框架

文献[20]引出相互性概念，假设对路网内某移动对象采用某泛化方法产生的泛化子网路段集与集合内任意路段按相同泛化方法生成的泛化子网路段集相同，称该泛化子网对应路段集合是符合相互性的.

文献[14]对路网环境泛化子网的结构进行限制，提出隐匿环子网结构. 若最终生成的子网为环状结构，且组成子网的路段数 l 以及位于子网内的用户数 K 符合以下约束：$l \geq \delta_l$，$K \geq \delta_k$，δ_l 与 δ_k 为隐私保护约束限制的泛化后子网包含的路段与移动对象阈值，这样的泛化子网称为隐匿环. 其构建隐匿环的主要策略是迭代寻找子网规模最小的环，选取最接近位置隐私约束条件者作为最终隐匿环；若找不到，继续扩展最小环及筛选过程.

2.2　问题描述

现有隐匿环泛化算法存在防护失效风险. 以图 2 为例，提起位置泛化请求的移动对象位于 e_8 上，其隐私保护要求为泛化子网至少由四条路段构成，且覆盖不少于 4 个移动对象，图中黑色三角描述其余移动对象位置. 假设构建了〈n_3，n_6，n_5，n_2，n_3〉隐匿环，

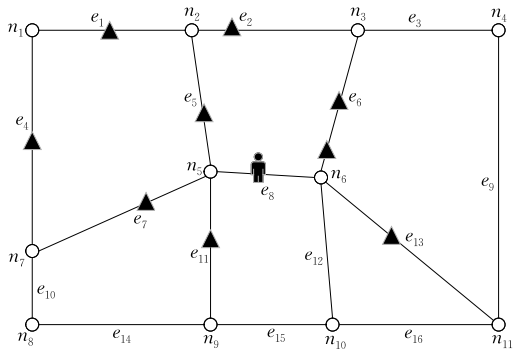

图 2　位置保护失效示例

该隐匿环由 4 条路段组成，环中包括 5 名移动对象，该隐匿环在提交位置服务器过程很容易为攻击者获取，若攻击方选取路段 e_5，采用同样泛化处理方法可构建由〈n_2，n_1，n_7，n_5，n_2〉组成的隐匿环，由于其仅包含 4 名移动对象，违反了相互性，将导致路段 e_8 上移动对象位置隐私泄露.

在匿名服务器端位置泛化处理效率方面，假设路网扩张搜索树的高度为 h，路网结点度均值为 a，搜索树最多有$(a^h-a)/a-1$ 条边需遍历. 给定路段数目约束，相比其它数据结构，环状子网的搜索树深度最大. 因而其查询搜索的复杂度也最大，存在位置泛化效率较低的问题.

位置服务器端近邻查询采用对泛化子网的结点进行扩张查询策略，已有方法不区分边界结点与子网内部结点. 相邻结点的查询结果可能有公共交集，无差异查询处理不可避免将造成计算资源浪费.

另一方面，随着保护位置隐私路网近邻查询研究不断走向实用，对个性化查询服务的要求也日益迫切. 已有的多数方法缺少对位置泛化与近邻查询效率和位置隐私保护效果动态调控能力的支持.

3 可控隐匿环位置泛化方法

目前，基于环的位置泛化在泛化处理时效和位置隐私保护效果上有缺陷，主要原因是已有泛化方法缺少对隐匿环构建过程调节功能的支持. 隐匿环结构和规模直接关系服务提供方的 k 近邻 POI 查询性能. 考虑从路网的环分布角度，通过对隐匿环构建过程施加调控，实现查询者位置泛化过程与位置服务器端隐私保护近邻查询的动态可调.

3.1 路网环分布构建方法

对给定路网，包含路网内给定路段的环是固定的，通过离线构建各路段的路网环分布，可以结合隐私约束以及路网环分布对是否仍需继续路网搜索进行预判，支持对位置泛化、近邻 POI 查询代价、查询发起者位置保护效果的调节.

定义 1. 路网环分布. 给定路网 G，G 中路段数为$|G|$，包含路段 $r(i)$的路网环中由 x 条路段组成的路网环数目为 c_i^x，则 G 中的路网环由 x 条路段构成的概率：

$$p(x)=\frac{\sum_{i=1}^{|G|}c_i^x}{\sum_{x=r_{\min}}^{r_{\max}}\sum_{i=1}^{|G|}c_i^x} \tag{1}$$

其中，$r_{\min}$ 与 $r_{\max}$ 为路网 G 中，可能构成路网环的最小路段数和最大路段数.

解析路网搜索过程，引入查询树概念，定义如下：(1) 路网查找的起始顶点为查询树的树根；(2) 查询树中某个非叶结点在路网中的邻接结点定义为该非叶结点的子女结点；(3) 查询树的叶结点是查询的终止结点；(4) 查询树从根到叶的路径上不允许出现重复结点.

生成路网环分布方法如算法 1 所示.

算法 1. 路网环分布生成方法.

输入：G，队列 Q

输出：G 的环分布 proCirD(G)

```
1.  R=null
2.  FOR G 中每条边 e
3.    以 e 的某一结点为起点 n_start，另一结点为终点 n_end；
4.    将 n_start 加入 Q；
5.    WHILE Q≠∅
6.      弹出 Q 中查询树当前处理结点的下层结点；
7.      FOR 每个结点 w
8.        IF 查询树根至 w 的路径有重复结点
9.          BREAK；
10.       ELSE
11.         IF w=n_end
12.           向 R 添加环；
13.         ELSE
14.           将 w 的邻接结点加入 Q；
15.       END IF
16.     END FOR
17.   END WHILE
18. END FOR
19. FOR i=1 to |R|
20.   m=m+R(i).key+R(i).value；
21. END FOR
22. FOR i=1 to |R|
23.   n=R(i)/m；
24. proCirD(G).push((R(i),n))；
25. END FOR
26. return proCirD(G).
```

如图 3 所示路网，其中包含路段数分别为 5 和

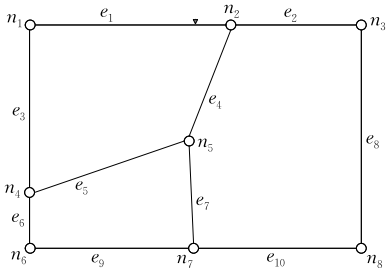

图 3 路网示例

7 的环数目分别为 4 和 17. 相应采用(路段数,分布概率)格式的环分布为(5,0.10)和(7,0.59).

3.2　基于环分布的隐匿环构建方法

组成隐匿环的路段和位于环中的移动对象数目通常用来衡量泛化子网的规模,子网规模越大,查询发起者的位置信息越安全,构建子网的泛化处理代价也越大,进而导致位置服务器端近邻 POI 查询也越复杂.

定义 2.　路段分布密度. 假设路网 G 中有路段 L 条,路网中的移动对象共 N 个,则路网 G 的路段分布密度为 $\rho=N/L$.

考虑基于组成环结构子网的路段数目和移动对象数定义隐匿环复杂度.

定义 3.　隐匿环复杂度. 假设隐匿环 C 由 l 条路段组成,路段上共有 K 个用户,C 的复杂度 $Complex(C)=\alpha l+\beta K$,$\alpha,\beta$ 为取值[0…1]间的权重因子,需满足 α 与 β 的和为 1.

对给定关于路段数目和移动对象数的隐私约束,隐匿环构建过程是否可以终止,不仅受环分布概率影响,还与局部路网内移动对象的分布特征有关,因此引入路段分布密度定义.

基于上述定义,引入构建隐匿环的结束条件.

定义 4.　结束建环约束. 隐匿环构建过程中,当已生成包含 x 条路段的环结构,且环内有 K 个移动对象时,若以下约束不成立:

$$\begin{cases}\rho(i+x)\geq\delta_k\\ \dfrac{\alpha x+\beta K-[(i+x)\alpha+\rho(i+x)\beta]}{\alpha x+\beta K}\cdot p(i+x)>\varepsilon\end{cases}\tag{2}$$

$i\in[1,\cdots,\delta_l^{\max}-x]$,$\delta_l^{\max}$ 为隐私约束限制的隐匿环所包含路段数上限,$\varepsilon\in[0,1)$,α 和 β 为隐匿环复杂度参数. 则可以终止当前隐匿环构建.

隐匿环满足上述约束表明存在比当前隐匿环复杂度更小且复杂度差距显著的隐匿环. ε 越大,结束条件越容易满足,隐匿环构建过程越可能提前终止,产生隐匿环的复杂度也往往较大;反之,ε 越小,隐匿环复杂度也往往较小.

通过性质 1,对隐匿环构建过程提前结束构建的发生概率进行分析.

性质 1.　隐匿环构建过程,假设当前生成的隐匿环由 x 条路段构成,有 K 个移动对象位于当前隐匿环内,则提前结束隐匿环构建的发生概率为 $\prod_{i=1}^{\delta_l^{\max}-x}p_i$,$p_i=P\Big(\neg\Big(\dfrac{\alpha x+\beta K-[(i+x)\alpha+\rho(i+x)\beta]}{\alpha x+\beta K}\cdot p(i+x)>\varepsilon\wedge\rho(i+x)\geq\delta_k\Big)\Big)$,$P()$为概率函数.

证明.　依据构建隐匿环结束条件,要求对任意 i(取值范围 $1,\cdots,\delta_l^{\max}-x$),不满足定义中约束条件,其物理含义是即便继续进行路网扩张搜索,也生成不了复杂度低于当前环结构复杂度,且复杂度差异不小于阈值 ε 的隐匿环,故而选择终止隐匿环构建. 对给定 i,隐匿环内用户数不小于阈值 δ_k 的约束条件为满足 $\rho(i+x)\geq\delta_k$ 关系;满足继续搜索获得的隐匿环复杂度与当前隐匿环复杂度差异足够大,则要求满足 $\dfrac{\alpha x+\beta K-[(i+x)\alpha+\rho(i+x)\beta]}{\alpha x+\beta K}\cdot p(i+x)>\varepsilon$ 条件. 相应的,对给定 i,停止隐匿环构建的概率为 $P\Big(\neg\Big(\dfrac{\alpha x+\beta K-[(i+x)\alpha+\rho(i+x)\beta]}{\alpha x+\beta K}\cdot p(i+x)>\varepsilon\wedge\rho(i+x)\geq\delta_k\Big)\Big)$.　证毕.

路网环境位置隐私保护近邻查询的场景如下:查询发起者向匿名服务器提交的位置泛化请求形如:$\{loc_X,loc_Y,\delta_l,\delta_k,\delta_l^{\max},\varepsilon\}$,其中,$(loc_X,loc_Y)$为查询者发起查询时的真实位置,$\delta_l$ 和 δ_k 表示查询者出于保护自身位置隐私考虑,期望构成隐匿环的路段数阈值,以及位于环内的移动对象数目阈值,$\delta_l^{\max}$ 为查询者对隐匿环中包含路段数的上限约束,ε 是控制隐匿环构建约束条件的阈值.

匿名服务器端,PNQCD 算法构建隐匿环的步骤如下:

(1)匿名服务器根据查询者真实位置坐标确定其所在路段,将路段的某个端点设置为扩张查找出发点,遍历路网,直到所遍历的路段构成环;

(2)若已遍历路段构成隐匿环,同时当前环结构符合提前结束隐匿环构建的约束,以当前环结构为最终生成的隐匿环,如符合条件的环不唯一,选隐匿环复杂度最低的环;

(3)若当前环是隐匿环,但不符合提前结束的约束,但该环为第一次生成的隐匿环,则存储该环,继续遍历;否则,保存该环与已保存环中复杂度低的环;

(4)若当前所构建环包含的路段数目大于 $\delta_l^{\max}$,或路网中所有路段结点都已遍历,停止遍历.

性质 2.　相较 CCF 方法,PNQCD 算法的隐匿环构建方法具有更好的重放攻击防范效果.

证明.　路网环境基于空间混淆的位置泛化所提供的保护强度主要取决于所生成子网的结构和规模(在此表现为子网包含的路段数和移动对象数,及

所定义隐匿环复杂度),通常子网复杂度越高,保护强度也越高.CCF 算法以生成满足查询者隐私约束(给定路段数和移动对象数阈值)的最小环为目标,PNQCD 算法引入参数 ε 控制隐匿环构建过程,通过提前终止环构建,牺牲环的复杂度,换取对隐匿环构建效率和服务器端基于隐匿环的查询处理效率的调控.

在提前终止隐匿环构建的场景下,PNQCD 算法生成隐匿环并非最小隐匿环,环的复杂度以及包含的路段数和移动对象数也会大于 CCF 算法生成的隐匿环,此时 PNQCD 算法能提供更好的保护效果.在未提前终止隐匿环构建的场景下,PNQCD 算法也将查询生成满足隐私约束的最小隐匿环,这时所提供保护强度与 CCF 算法相当. 证毕.

综上,基于 PNQCD 的位置泛化方法与 CCF 方法类似,属于匿名类隐私模型,能够兼容基于泛化子网中移动对象规模和路段数目的位置保护隐私约束,有效防止该类重放攻击.

4 PNQCD 算法服务器端查询处理

隐匿环中路段结点具有较好的可达性,在对路段结点进行查询时,其近邻 POI 集合与其邻接结点的近邻 POI 集合可能存在交集.

考虑基于组成隐匿环的路段结点位置特征和邻接结点近邻查询结果间的相关性,分析隐匿环中无需进行查询的结点具备的特征,进而设计剪枝策略避免无效查询对计算资源的浪费,在此基础上通过对结点查询顺序的设置,减少 PNQCD 算法服务器端的近邻查询时间.

借鉴文献[4]中所引入边界结点定义,采用 $BVs(S)$ 描述路网 G 中,子网 S 的边界结点集合.文献[4]进一步给出了对给定近邻查询与子网、近邻查询结果满足的约束关系.即对查询 q 与泛化子网 S,其近邻查询结果集是集合 $(\bigcup_{s\in S}\Omega(q,s))\cup(\bigcup_{n\in BVs(S)}\Omega(q,n))$ 的子集,其中,$\bigcup_{s\in S}\Omega(q,s)$ 为子网 S 上查询对象集合.

算法在服务器端需要对隐匿环中边界结点进行近邻 POI 搜索.考虑环结构特点使得隐匿环组成以边界结点为主,而隐匿环的环状特殊联通结构使得边界结点间具有良好的邻接关系.可以利用当前边界结点邻接路段的近邻 POI 对象集合设计预剪枝策略,进一步通过减少待搜索的起始路段来提升隐匿环的遍历查询处理效率;在此基础上,对边界结点的查询处理顺序进行优化,首先构建长度最大的邻接边界结点序列,对序列中的结点顺次遍历,以提高边界结点的处理效率.

定义 5. 强邻接路段.隐匿环中的结点 v_1 和 v_2 均为边界结点,且存在以 v_1 和 v_2 为端点的路段 $e_{v_1v_2}$,则称 $e_{v_1v_2}$ 为强邻接路段.

定义 6. 局部结果集. v_1 和 v_2 为隐匿环中的邻接边界结点,v_1 的局部结果集定义为 $R(v_2)/R(v_1,v_2)\cup R(e_{v_1v_2})$,其中 $R(v_2)$ 为 v_2 近邻查询获得的 POI 集合,$R(v_1,v_2)$ 为沿 v_1 向 v_2 搜索获得的近邻 POI 查询结果,$R(e_{v_1v_2})$ 为路段 $e_{v_1v_2}$ 上的 POI 集合.

定义 7. 强结果集.在查询边界结点 v_1 的 k 近邻 POI 过程中若其局部结果集包含与 v_1 距离递增的 m 个 POI(距离分别为 $d_1,\cdots,d_m$),v_1 沿其非强邻接路段搜索,所获 POI 与 v_1 的距离依次递增为 $d'_1,\cdots,d'_k$,则 v_1 的局部结果集为强结果集的条件是存在 $l(1\le l\le k)$:

$$\begin{cases} d'_l < d_m \\ m+l \ge k \end{cases} \tag{3}$$

对存在强结果集的边界结点 v_1,沿其非强邻接路段进行搜索不会影响其近邻 POI 查询的准确性.从而通过减少初始搜索路段,避免无效的查询处理,可有效提升近邻查询效率.

基于强结果集定义,对边界结点概念进行扩展,给出强边界结点定义.

定义 8. 强边界结点.基于隐匿环查找 k 近邻 POI 集合时,若结点 v_1 和 v_2 是由隐匿环中强邻接路段 $e_{v_1v_2}$ 连接的两个边界结点,满足 $|R(v_2)/R(v_1,v_2)\cup R(e_{v_1v_2})|=k$,称 v_2 是 v_1 的强边界结点.

对存在强边界结点的结点 v_1,通过预剪枝降低搜索树的高度,减少所需查询处理的路段,可以实现 k 近邻 POI 查询效率的进一步提升.

假设图 4 所展示近邻查询中 $k=4$,四条路段组成隐匿环为〈$n_{11},n_{10},n_3,n_4,n_{11}$〉,已知 n_3 为强边界结点,其 k 近邻已获取,由 n_3 出发沿路段 e_2、e_3 搜索获得{p_2,p_3,p_7},增加位于路段 e_7 上的 p_1,生成

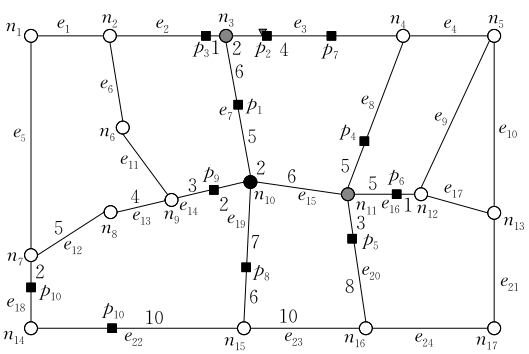

图 4 查询过程剪枝操作示例

POI 集合$\{p_2, p_3, p_7, p_1\}$，这几个 POI 与 n_{10} 的距离分别是 13、12、17 和 5，在查找结点 n_{10} 的 k 近邻 POI 时仅需沿其除路段 e_7 以外的三条邻接路段搜索，同时还可以利用 p_7 与结点 n_{10} 的距离对遍历过程进行剪枝，从而快速生成结点 n_{10} 的近邻查询结果.

PNQCD 算法对边界结点的查询处理过程如下：首先，查找获得 k 个 POI 作为初始查询结果集合，通常采用大顶堆（容量为 k）存储查询结果集合. 随后，迭代进行扩张查找，并用路网距离更近的 POI 替换查询结果集合中的 POI.

位置服务器端接受匿名服务器提交的隐匿环和近邻查询请求后，主要处理步骤：

(1) 若组成隐匿环的边界结点不存在强邻接路段，直接对其进行 k 近邻查询遍历；否则，遍历其强邻接路段，生成邻接边界结点序列.

(2) 查找获取序列中首结点的 k 近邻 POI. 后续各结点根据其前序结点是否为强边界结点处理，若前序结点是强边界结点，沿其非强邻接路段进行查询，实现剪枝；

(3) 若前序结点不满足强边界结点约束，但其局部结果集满足强结果集约束，沿其非强邻接路段进行查询，并将强结果集并入查询结果；

(4) 若前序结点不满足(2)、(3)条件，则沿强邻接路段查找获取其$(k-y)$近邻，并将结果并入沿其非强邻接路段进行查询生成的结果中，其中 y 为局部结果集内的 POI 数目；

(5) 将边界结点的近邻查询结果和隐匿环上的 POI 进行合并，作为隐匿环的候选 k 近邻 POI 集合返回.

性质 3.　假设查询树的结点平均度数和高度分别为 m 和 h，则 PNQCD 算法的查询代价为 $(m^h-m)/(m-1)[p_1+(1-p_1)p_2(1-\theta)]+(1-p_1)(1-p_2)(1-\theta)(m^{h-1}-1)$. 其中，$p_1$ 和 p_2 分别表示无邻接边界结点的出现概率和邻接边界结点非强边界结点的出现概率，θ 表示因提前剪枝而减少的搜索查询代价比例.

证明.　隐匿环中边界结点的邻接路段有以下几种情况：(1) 隐匿环中某边界结点无强邻接路段，其搜索代价为$(m^h-m)/(m-1)$；(2) 隐匿环中某边界结点有强邻接路段，且其强邻接路段连接的结点不是强边界结点，经提前剪枝后查询代价为$(1-\theta)(m^h-m)/(m-1)$；(3) 隐匿环中某边界结点有强邻接路段，且其强邻接路段连接的结点是强边界结点，经提前剪枝和减少初始搜索边后查询代价为$(1-\theta)(m^{h-1}-1)$. 故而，PNQCD 算法的平均查询代价为$(m^h-m)/(m-1)[p_1+(1-p_1)p_2(1-\theta)]+(1-p_1)(1-p_2)(1-\theta)(m^{h-1}-1)$.　证毕.

尽管 PNQCD 近邻查询复杂度和常规查询方法同为 $O(m^{h-1})$，但由于$(1-\theta)(m^h-m)/(m-1)<(m^h-m)/(m-1)$，且满足$(1-\theta)(m^{h-1}-1)<(m^h-m)/(m-1)=(m^{h-1}-1)m/(m-1)$，因此容易得出 PNQCD 算法的服务器端处理效率较常规采用的服务器端子网近邻查找方法要高.

5　实验分析

本节对所提方法的有效性进行实验分析，主要从 PNQCD 方法位置泛化效果、位置服务器端近邻查询效率、PNQCD 方法可调控性三个方面设计实验验证算法的有效性.

第一组实验对比分析 PNQCD 算法所设计位置泛化方法生成的隐匿环与 CCF 方法生成泛化子网的时效和位置隐私保护效果. 第二组实验验证 PNQCD 算法在位置服务器端的路网近邻查询性能. 第三组实验验证 PNQCD 算法对查询发起者位置泛化的时耗、基于隐匿环的服务器端 k 近邻 POI 查询效率以及位置隐私保护效果的调控效果.

5.1　实验环境

实验用计算机配置如下：CPU 2.9 GHz，4 GB 内存，操作系统：Win7. 路网数据来源于美国加利福尼亚州真实道路网络系统（http://www.cs.fsu.edu/~lifeifei/SpatialDataset.htm），覆盖 21 048 个路段交汇点和21693条公路段，POI数据集和用户位置数据随机生成. 路网环境参数和算法主要参数如表 1.

表 1　路网环境参数和算法主要参数

参数名称	参数值
移动对象数	5000
参数 $\delta_l, \delta_l, \delta_l^{\max}$	6,6,30
POI 数	20000
参数 ε	0.02
参数 α, β	0.6,0.4
近邻对象数 k	10

5.2　PNQCD 方法位置泛化效果分析

随着 δ_l 和 δ_k 的增加，泛化子网复杂度和泛化所消耗时间也逐次增加，具体变化趋势见图 5～图 8，其中泛化时间消耗为查询者向匿名服务器发起位置泛化请求到匿名服务器完成泛化子网构建的时间间隔. 由图可见，随着 δ_l 和 δ_k 的增加，PNQCD 算法与

CCF 所构建隐匿环的复杂度和泛化处理时间都变大,但 PNQCD 算法的增长幅度更大. 其原因在于 PNQCD 算法设计基于环分布的结束位置泛化约束条件,管控隐匿环构建过程,可以提前终止隐匿环的构建,提升了隐匿环构建效率也加强了隐匿环所提供的位置隐私保护强度,付出的代价则是隐匿环复杂度的增大. 泛化时间消耗方面,随着查询者对泛化子网所包含路段数和用户数约束条件的加强,为了获取满足约束条件的泛化子网,匿名服务器势必需要在更大范围内进行查找,这将直接导致泛化时间的增加. 由于 PNQCD 算法存在提前终止泛化子网(隐匿环)构建情况,其泛化效率相较 CCF 方法要高.

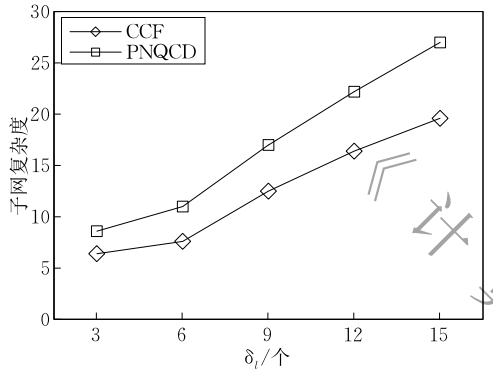

图 5 子网复杂度随 δ_l 变化趋势

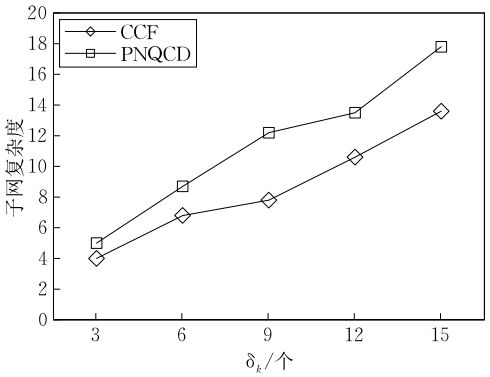

图 6 子网复杂度随 δ_k 变化趋势

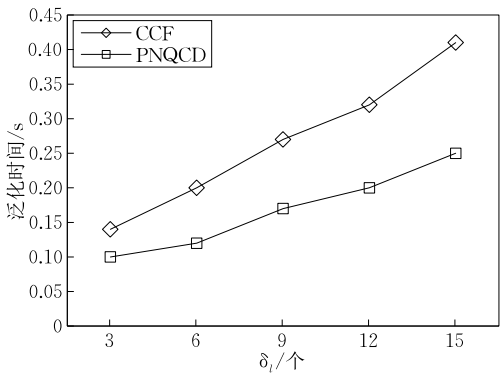

图 7 泛化时间随 δ_l 变化趋势

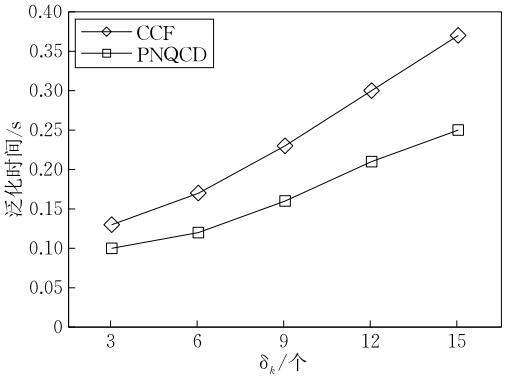

图 8 泛化时间随 δ_k 变化趋势

隐匿环构建的成功率实验见图 9 和图 10,对每组实验采用随机生成 100 个位置,发起 100 次位置泛化请求的方法统计成功率(用户数和路网边参数设置相同). 由图可知,PNQCD 方法构建隐匿环的成功率稍高于 CCF 方法,其原因在于尽管两种方法隐匿环构建成功与否都受到泛化子网内移动对象数与包含路网边数约束影响,但 CCF 方法采用先找到最小环,再拓展的策略,而 PNQCD 方法采用全局搜索,牺牲生成环的复杂度,换取子网复杂度、泛化处理时耗、以及基于隐匿环的 k 近邻查询效率可控. 泛化子网(隐匿环)复杂度越高,隐匿环构建的成功概率相对也会提高.

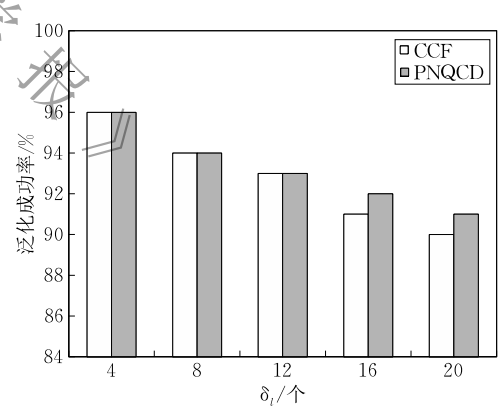

图 9 泛化成功率随 δ_l 变化趋势

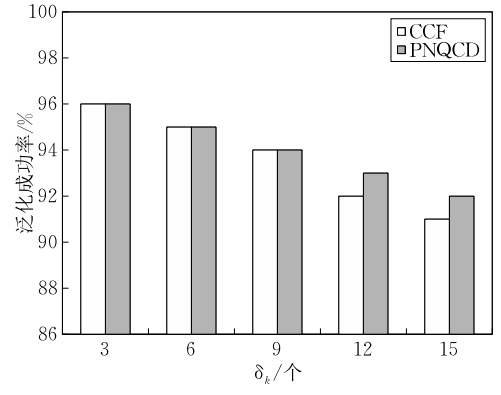

图 10 泛化成功率随 δ_k 变化趋势

5.3　近邻查询效率分析

本节对所提 PNQCD 算法在位置服务器端的近邻查询处理效率进行分析和验证，具体策略是对 PNQCD 算法以及 CCF 方法生成的相应隐匿环，分别采用常规路网扩张查询方法（采用 PSNN 方法[17]）和 PNQCD 所提位置服务器端近邻查询处理方法进行相同的近邻 POI 查询，对比各场景下的位置服务器端近邻查询效率.

图 11、图 12 为实验结果图，图中各算法采用的泛化子网构建方法以及位置服务器端近邻查询处理方法见表 2.

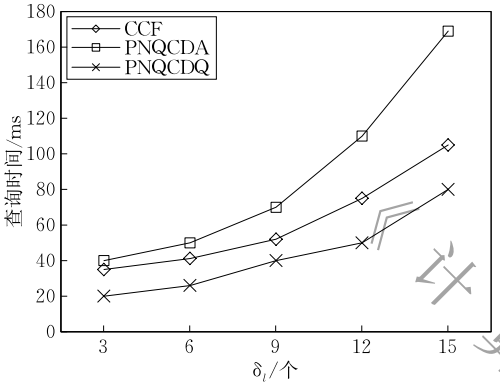

图 11　查询时间随 δ_l 的变化趋势

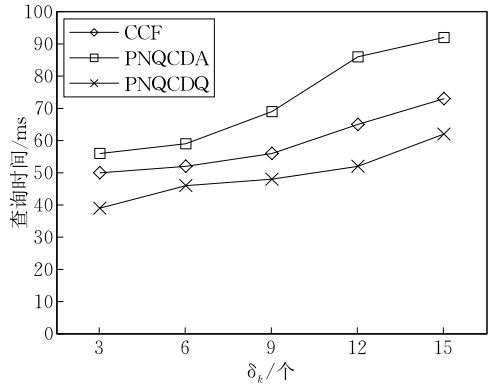

图 12　查询时间随 δ_k 的变化趋势

表 2　算法名称含义

算法名称	泛化子网生成方法	服务器端近邻查询方法
CCF	CCF 方法	PSNN 方法
PNQCDA	PNQCD 泛化方法	PSNN 方法
PNQCDQ	CCF 方法	PNQCD 服务器端查询方法

由图可知，各方法的位置服务器端查询时间均随位置泛化参数 δ_l 和 δ_k 的增加呈变大趋势，其理由是隐私约束参数控制了各位置泛化方法所生成泛化子网的复杂度，而泛化子网的复杂度直接影响位置服务器端对泛化子网的近邻查询时待处理结点的规模. 由于前文分析得出相同隐私参数约束下，PNQCD 方法生成的隐匿环复杂度通常较 CCF 方法大，图中 PNQCDA 在位置服务器端的近邻查询时间较另两方法的时间消耗要大.

此外，由图可见，对采用相同泛化子网生成方法的 PNQCDQ 方法和 CCF 方法，PNQCDQ 方法在位置服务器端的近邻查询时间消耗显著少于 CCF 方法的时间消耗. 其原因在于不同于位置服务器端通常采用的路网扩张查询方法，PNQCD 算法通过区分泛化子网边界结点，采用不同查询策略，通过预剪枝、减少搜索起始边等策略有效减少查询计算开销，提高位置服务器端近邻查询效率. 同时，由于 PNQCD 方法在位置服务器端的近邻查询时间消耗不仅与泛化子网内路段数目有关，还与泛化子网的边界结点邻接情况有关，因此其位置服务器端近邻查询时间消耗受泛化子网所包含路段数目影响相对较小，图 11 中曲线增长幅度也印证了这一结论，可以发现，PNQCDQ 方法对应曲线的增幅明显比另两个曲线小.

进一步，验证所提 PNQCD 近邻查询方法查询效率以及返回候选解规模随查询近邻数 k 的变化趋势. 主要对所提 PNQCD 算法的服务器端查询处理方法与服务器端 PSNN 处理方法的效率进行对比. PNQCD 算法在位置服务器端通过减枝与减少搜索起始边，有效降低了关于隐匿环的近邻查询时耗，也使得 PNQCD 算法查询时效受 k 值影响的敏感度变弱，如图 13 所示，随着 k 的增加，PNQCDQ 和 CCF 两种方法在位置服务器端的近邻查询处理时耗都变大，但 PNQCD 方法的查询时间增幅显著低于后者.

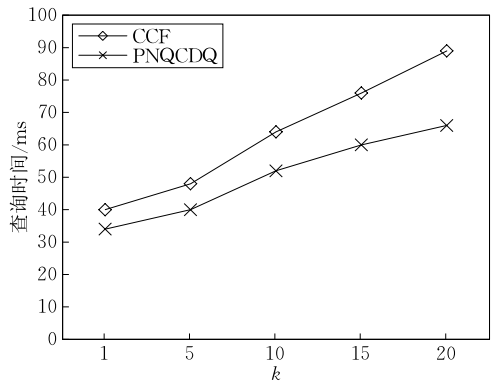

图 13　查询时间随 k 的变化趋势

基于空间混淆的保护位置查询，采用位置服务器返回包含准确查询结果的候选 POI 集合（即候选解集）方式，实现查询者准确查询结果的获取，图 14 对应 PNQCD 方法与 PSNN 方法所生成候选解规模的对比，由图易见，随着查询近邻数 k 的增加，两种方法产生候选解的规模均呈增长趋势，但总体上

PNQCD 算法的增长幅度较之 PSNN 方法更显著，其原因在于 PNQCD 存在提前结束隐匿环子网构建的情况，此时算法最终生成的环复杂度稍高，而候选解集规模与所隐匿环复杂度密切相关.

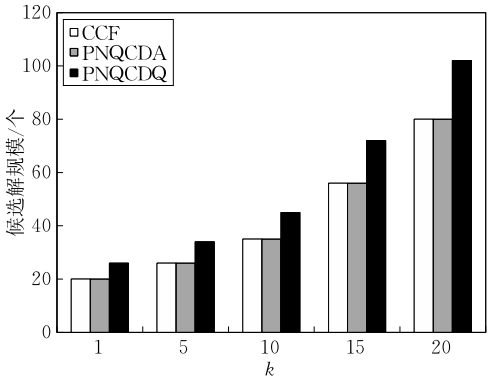

图 14 候选解规模与 k 的关系

5.4 PNQCD 算法可调控性实验

PNQCD 算法对个性化查询的支持主要表现为隐匿环构建终止条件中对参数 ε 的设置，隐匿环复杂度、构建隐匿环的时间消耗、位置服务器端 k 近邻查询时耗随 ε 增加的变化趋势见图 15～图 17. 显而易见，随着 ε 增大，隐匿环复杂度以及位置服务器端近邻查询时间消耗呈同步增加趋势，构建隐匿环的时间消耗呈减小趋势. 所定义隐匿环的复杂度通常可用来表征位置泛化的有效性，即对查询发起者实时位置保护效果的强弱，隐匿环复杂度越高，保护效果越好. 实验结果表明通过设置合适的 ε，能够根据查询发起者的偏好，实现对隐匿环位置保护强弱、隐匿环的构建时耗和算法进行 k 近邻 POI 查询处理效率的有效调节.

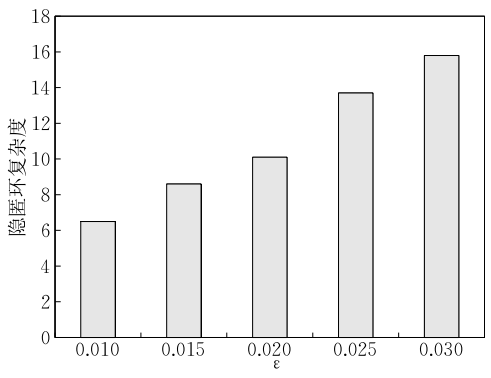

图 15 环复杂度变化趋势(ε)

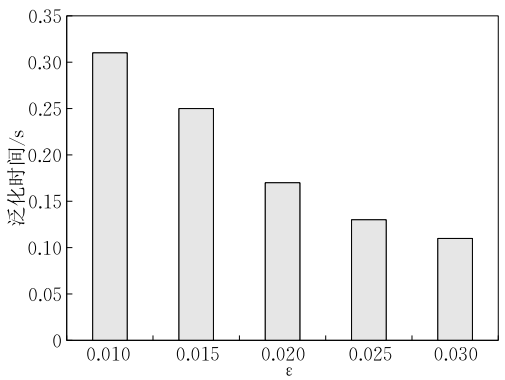

图 16 泛化时间变化趋势(ε)

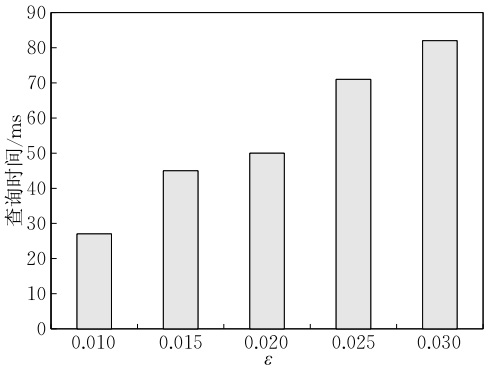

图 17 查询时间随 ε 的变化趋势

6 总结与展望

针对现有采用空间混淆的路网隐私保护查询方法存在对个性化近邻查询服务支持能力弱、处理效率低的问题和对查询发起者位置安全保护方面存在的不足，提出保护位置隐私路网 k 近邻查询方法 PNQCD，引入路网环分布概念并设计生成给定路网环分布的算法；设置基于路网环分布的结束隐匿环构建约束条件，提升所提隐匿环机制对重放攻击的抵御能力；通过对隐匿环复杂度的调控，实现关于位置泛化效果和近邻查询效率的个性化服务支持. 在此基础上解析隐匿环结点的约束关系，设计查询优化方法，提升位置服务器端基于隐匿环的近邻查询处理效率.

后续，将对有向路网环境下基于隐匿环与环分布技术的保护位置隐私近邻查询方法展开研究.

参 考 文 献

[1] Cheng R, Zhang Y, Bertino E, Prabhakar S. Preserving user location privacy in mobile data management infrastructures// Proceedings of the 6th International Workshop on Privacy Enhancing Technologies. Cambridge, UK, 2006: 393-412

[2] Chow C-Y, Mokbel M F. Enabling private continuous queries for revealed user locations//Proceedings of the 10th International Symposiums on Advances in Spatial and Temporal Databases. Boston, USA, 2007: 258-275

[3] Mouratidis K, Yiu M L. Anonymous query processing in road networks. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(1): 2-15

[4] Wang Ting, Liu Ling. Privacy-aware mobile services over road networks//Proceedings of the 35th International Conference on Very Large Data Bases. Lyon, France, 2009: 1042-1053

[5] Kim Y K, Hossain A, Hossain A A, et al. Hilbert-order based spatial cloaking algorithm in road network. Concurrency and Computation: Practice and Experience, 2013, 25(1): 143-158

[6] Fan Xinyue, Tu Jing, Ye Chaolong, et al. The research for protecting location privacy based on V-W algorithm. Eurasip Journal on Wireless Communication & Networking, 2014, 2014(1): 1-15

[7] Wang Lu, Meng Xiao-Feng. Location privacy preservation in big data era: A survey. Journal of Software, 2014, 25(4): 693-712(in Chinese)
(王璐, 孟小峰. 位置大数据隐私保护研究综述. 软件学报, 2014, 25(4): 693-712)

[8] Pan Xiao, Chen Wei-Zhang, Wu Lei, et al. Protecting personalized privacy against sensitivity homogeneity attacks over road networks in mobile services. Frontiers of Computer Science, 2016, 10(2): 370-386

[9] Liang Hui-Chao, Wang Bin, Cui Ning-Ning, et al. Privacy preserving method for point-of-interest query on road network. Journal of Software, 2018, 29(3): 703-720(in Chinese)
(梁慧超, 王斌, 崔宁宁等. 路网环境下兴趣点查询的隐私保护方法. 软件学报, 2018, 29(3): 703-720)

[10] Palanisamy B, Liu L. Attack-resilient mix-zones over road networks: Architecture and algorithms. IEEE Transactions on Mobile Computing, 2015, 14(3): 495-508

[11] Zhang Xue-Jun, Gui Xiao-Lin, Wu Zhong-Dong. Privacy preservation for location-based services: A survey. Journal of Software, 2015, 26(9): 2373-2395(in Chinese)
(张学军, 桂小林, 伍忠东. 位置服务隐私保护研究综述. 软件学报, 2015, 26(9): 2373-2395)

[12] Zhou Chang-Li, Ma Chun-Guang, Yang Song-Tao. Location privacy-preserving method for LBS continuous KNN query in road networks. Journal of Computer Research and Development, 2015, 52(11): 2628-2644(in Chinese)
(周长利, 马春光, 杨松涛. 路网环境下保护 LBS 位置隐私的连续 KNN 查询方法. 计算机研究与发展, 2015, 52(11): 2628-2644)

[13] Arain Q A, Deng Z L, Memon I, et al. Location privacy with dynamic pseudonym-based multiple mix-zones generation over road networks. Wireless Personal Communications, 2017, 95(2): 505-521

[14] Xue Jiao, Liu Xiang-Yu, Yang Xiao-Chun, Wang Bin. A location privacy preserving approach on road network. Chinese Journal of Computers, 2011, 34(5): 866-877 (in Chinese)
(薛姣, 刘向宇, 杨晓春, 王斌. 一种面向公路网络的位置隐私保护方法. 计算机学报, 2011, 34(5): 866-877)

[15] Ni Wei-Wei, Chen Xiao, Ma Zhong-Xi. Location privacy preserving k nearest neighbor query method on road network in presence of user's preferences. Chinese Journal of Computers, 2015, 38(4): 885-895(in Chinese)
(倪巍伟, 陈萧, 马中希. 支持偏好调控的路网隐私保护 k 近邻查询方法. 计算机学报, 2015, 38(4): 885-895)

[16] Papadias D, Zhang J, Mamoulis N, et al. Query processing in spatial network databases//Proceedings of the 29th International Conference on Very Large Data Bases-Volume 29. Berlin, Germany, 2003: 802-813

[17] Ku W S, Zimmermann R, Peng W C, et al. Privacy protected query processing on spatial networks//Proceedings of the 23rd IEEE International Conference on Data Engineering Workshops. Istanbul, Turkey, 2007, 2(8): 215-220

[18] Kolahdouzan M, Shahabi C. Voronoi-based k nearest neighbor search for spatial network databases//Proceedings of the 30th International Conference on Very Large Data Bases-Volume 30. Toronto, Canada, 2004: 840-851

[19] Xu J, Gao Y, Liu C, et al. Efficient route search on hierarchical dynamic road networks. Distributed & Parallel Databases, 2014, 33(2): 227-252

[20] Pan Xiao, Chen Wei-Zhang, Sun Yi-Ge, Wu Lei. Continuous queries privacy protection algorithm based on spatial-temporal similarity over road networks. Journal of Computer Research and Development, 2017, 54(9): 2093-2101(in Chinese)
(潘晓, 谌伟璋, 孙一格, 吴雷. 道路网络上基于时空相似性的连续查询隐私保护算法. 计算机研究与发展, 2017, 54(9): 2093-2101)

NI Wei-Wei, Ph. D., professor, Ph. D. supervisor. His main research interests include data mining and data privacy protection.

FENG Zhi-Gang, M. S. candidate. His main research interest is privacy preserving data application.

YAN Dong, M. S. candidate. His main research interest is privacy preserving data application.

Background

In recent years, privacy protection in location-based services becomes a hot topic in the domain of database technology. Most of current work in privacy preserving location based kNN query on road networks in common falls short in

providing sufficient location privacy protection against replay attack, as well as regulation mechanism to anonymization performance, query efficiency and privacy preserving intensity. To solve these problems, a cloaking circle-based privacy-preserving k nearest neighbor query method on road networks (PNQCD) is proposed. The anonymous circle probability distribution is pre-generated on the trusted third party and the judging condition is devised to complete anonymous subnet construction in advance. In this way, the replay attack can be efficiently resisted. Besides, the scale of the anonymous subnet can be regulated by setting appropriate judging condition, which can provide the mechanism of regulating anonymization performance, query performance and privacy. On the LBS server side, query efficiency is improved by pruning and reducing initial search edges according to the characteristic of the anonymous subnet. Theoretical analysis and experimental results demonstrate that our solution can resist replay attack effectively and provide good efficiency in location anonymization and query processing. Furthermore, the anonymization performance, query performance and privacy preserving intensity can be regulated according to user's preferences.

Our work is supported by the National Natural Science Foundation of China (Nos. 61772131 and 61370077). The project with No. 61370077 focus on the problem of privacy preference supporting in privacy preserving location-based query. Location privacy-preserving query in road network is an important research content of the project. In our previous work, we pay attention to query initiators' privacy preference supporting problem in location privacy preserving nearest neighbor query on road network. The solution can be found in Ref. [15]. Comparing with the existing method, the solution in this paper not only improves the ability of privacy protection effectively, but also accommodating query initiators' regulating ability among anonymization performance, query performance and privacy preserving intensity. It can provide more comprehensive query initiator's privacy preference support on road network from different perspective from schemes in Ref. [15].

The other project with No. 61772131 mainly solves the cooperative concealment problem of provenance and their derived data. This project devotes to compensate for current research in neglecting privacy leakage incurred by cooperative data sharing, as well as falling short in protecting privacy security in presence of provenance and its derived data be published in sequence. It is the foundation of privacy protection in location-based services.