软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2020,31(2):471−492 [doi: 10.13328/j.cnki.jos.005679]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

保护位置隐私和查询内容隐私的路网 *K* 近邻查询方法*

周长利[1], 陈永红[1], 田 晖[1], 蔡绍滨[1,2]

[1](华侨大学 计算机科学与技术学院,福建 厦门 361021)

[2](哈尔滨工程大学 计算机科学与技术学院,黑龙江 哈尔滨 150001)

通讯作者: 周长利, E-mail: zhouchangli666@163.com

摘 要: 位置隐私和查询内容隐私是LBS兴趣点(point of interest,简称POI)查询服务中需要保护的两个重要内容,同时,在路网连续查询过程中,位置频繁变化会给 LBS 服务器带来巨大的查询处理负担,如何在保护用户隐私的同时,高效地获取精确查询结果,是目前研究的难题.以私有信息检索中除用户自身外其他实体均不可信的思想为基本假设,基于Paillier密码系统的同态特性,提出了无需用户提供真实位置及查询内容的*K*近邻兴趣点查询方法,实现了对用户位置、查询内容隐私的保护及兴趣点的精确检索;同时,以路网顶点为生成元组织兴趣点分布信息,进一步解决了高强度密码方案在路网连续查询中因用户位置变化频繁导致的实用效率低的问题,减少了用户的查询次数,并能确保查询结果的准确性.最后从准确性、安全性及查询效率方面对本方法进行了分析,并通过仿真实验验证了理论分析结果的正确性.

关键词: 基于位置的服务;隐私保护;连续 *K* 近邻查询;私有信息检索

中图法分类号: TP309

中文引用格式: 周长利,陈永红,田晖,蔡绍滨.保护位置隐私和查询内容隐私的路网 *K* 近邻查询方法.软件学报,2020,31(2):471−492. http://www.jos.org.cn/1000-9825/5679.htm

英文引用格式: Zhou CL, Chen YH, Tian H, Cai SB. Location privacy and query privacy preserving method for *K*-nearest neighbor query in road networks. Ruan Jian Xue Bao/Journal of Software, 2020,31(2):471−492 (in Chinese). http://www.jos.org.cn/1000-9825/5679.htm

Location Privacy and Query Privacy Preserving Method for *K*-nearest Neighbor Query in Road Networks

ZHOU Chang-Li[1], CHEN Yong-Hong[1], TIAN Hui[1], CAI Shao-Bin[1,2]

[1](School of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

[2](School of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: Location privacy and query content privacy are both critical elements in LBS querying for points of interest (POIs). For continuous queries in road networks, frequent changes of a user's location bring huge burden of query processing to LBS server, how to release a user's privacy information as little as possible, and obtain accurate query results efficiently are still great challenges in current researches. Taking the idea of private information retrieval (PIR), i.e. no trusted entities except the user himself, as a basic assumption, a privacy-preserving method is put forward based on homomorphic properties of Paillier cryptosystem, which the user does not need to provide his actual location or query content to LBS server in *K* nearest neighbor POIs query, it achieves privacy preservation in LBS and accurate retrieval of POIs. Meanwhile, takeing the vertexes in road networks as generating elements to organize the distribution

* 基金项目: 国家自然科学基金(61802134, 61872154, 61472097, 61370007, U1536115, U1405254); 数据挖掘与智能推荐福建省高校重点实验室开放课题(DM201905); 华侨大学科研基金(15BS412)

Foundation item: National Natural Science Foundation of China (61802134, 61872154, 61472097, 61370007, U1536115, U1405254); Open Fund of Key Laboratory of Data Mining and Intelligent Recommendation, Fujian Province University (DM201905); Scientific Research Funds of Huaqiao University (15BS412)

收稿时间: 2017-07-13; 修改时间: 2017-09-09, 2018-08-29; 采用时间: 2018-10-16

information of POIs, the inefficient problem is further solved in most cryptographic query schemes, which is caused by frequent location changes in continuous query, the proposed method significantly reduces the frequency of initiating queries to LBS server without decreasing the query accuracy. Finally, the proposed method is analyzed from the aspects of accuracy, security, and efficiency, extensive experiments verify the effectiveness.

Key words: location-based service; privacy preservation; continuous *K*-nearest neighbor query; private information retrieval

基于位置的服务(location based service,简称 LBS)[1,2]成为移动智能终端中最广泛的应用服务形式,现有多数移动终端软件在安装和使用时都会默认读取用户位置数据,以便为用户提供高质量的服务.基于位置的 K 近邻(K-nearest neighbor,简称 KNN)兴趣点(point of interest,简称 POI)查询是 LBS 中应用最为广泛的一种服务形式[3],用户通过提供自己的真实位置和查询内容给位置服务器,获取服务器返回的距离该用户最近的 K 个兴趣点[4,5],如"查询距离我当前位置最近的5家医院".然而,直接提交这样的查询请求,会带来严重的用户位置隐私和查询内容隐私泄漏问题,攻击者可以通过二者并结合其他的背景知识,推断出用户的其他敏感信息,如家庭住址、习惯爱好、健康情况及社会关系等[1].随着使用智能终端用户数量的不断增多,LBS 中的隐私泄漏问题也受到越来越多用户的关注,用户在享受高质量的位置服务时,其隐私信息必须得到有效的保护[6,7].

LBS 查询中的隐私保护问题可以分成位置隐私保护和查询内容隐私保护两类[8].一方面,由于用户真实位置的时空关联特性,将自身真实位置发送给 LBS 服务器会带来直接位置隐私泄漏,模糊泛化的方法可用来保护用户的位置隐私[1,6,8].常用的位置隐私保护方法可分为构造匿名区(cloaking region,简称 CR)[1,9,10]和生成假位置(dummy)[11−16]两类.然而,匿名区方法在连续查询时存在难以构造的问题,而假位置方法多数情况需要利用多个虚假位置模糊真实位置,因此会带来额外的查询处理负担.

另一方面,现有的研究大多会考虑如何保护用户的位置隐私,而忽略了对查询内容的保护,查询内容同样可作为推断用户真实身份的直接关联条件.最好的隐私保护方式是用户不对外释放任何信息,然而这与获取服务相矛盾.在通常情况下,用户要获取服务就必须以释放自身信息为代价,如何在不知晓用户查询内容的条件下为其提供精确的服务结果看似难以实现.私有信息检索(private information retrieval,简称 PIR)[3,8,17,18]技术为上述问题提供了良好的解决途径,然而这种方法大多基于高强度密码系统设计,存在不适用于位置变换频繁的连续查询、无法区分兴趣点类型、多面向欧氏空间设计实用性差等缺陷[3];同时,现有多数 PIR 协议依然要依靠一个可信的安全处理器[18,19],对 LBS 数据库及用户真实位置预处理,其可信性依然基于假设,存在更多不安全因素.

最后,从实体架构和实体可信性来看,现有 LBS 隐私保护研究大多采用有中心服务器架构[1,6,8],负责模糊泛化用户隐私数据、代理用户发起查询请求,并假设中心服务器不会泄漏用户隐私.然而,用户将涉及隐私的数据交给任何第二方都存在隐私泄漏的可能,因此这种安全性假设并不符合实际情况.

针对上述问题,本文研究内容及贡献如下.

(1) 针对路网连续查询中的位置和查询内容隐私保护问题,基于 Paillier 密码系统的同态加密属性,提出了无需用户提供真实位置和查询内容的连续 K 近邻兴趣点查询方法,主要包括用户端基于同态加密的秘密查询请求生成、服务端在密文查询请求上基于同态特性的密文查询结果生成、用户端同态解密并计算精确结果这 3 个阶段.

(2) 针对 Paillier 等高强度密码系统中普遍存在的处理开销大、无法适用于连续查询等问题,基于锚点技术和以路网顶点为生成元的兴趣点分布信息组织结构,解决了高强度密码算法在面对连续查询时因用户位置变化频繁导致的实用性低的问题,有效降低了查询次数,同时能够确保查询准确率.

(3) 本方法基于"用户-LBS 服务器"两方实体架构,即除了用户自身外,假设其他实体都存在隐私泄露可能,实体可信性假设更符合实际;性能分析部分对本方法的查询服务质量和安全性进行了分析,并提出"ε_p-隐私确保模型"和"δ_q-质量确保模型"来确定隐私保护与服务质量之间的均衡点.大量的仿真实验验证了性能分析中所得出结论的正确性.

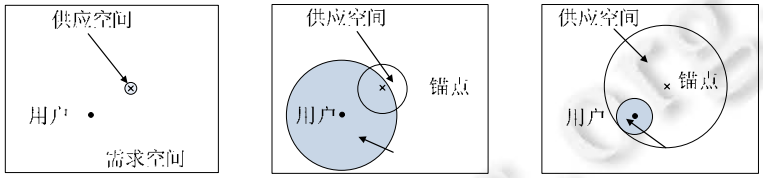
1 相关工作

LBS 兴趣点查询中的隐私保护问题主要可以分为两类:位置隐私保护和查询内容隐私保护[3,6,8].LBS 查询

请求可形式化为 $Q=\langle u_k,loc,time,K,C\rangle$,其中每个元素分别代表用户身份标识、位置、查询时间、近邻兴趣点查询个数及查询内容,其中,身份标识的保护可用匿名通信 Tor[20]或假名变换[21]的方法来实现,这里不做讨论.当用户提交查询请求时,既不愿意提供自己的真实位置数据,也不愿意提供查询请求内容,例如用户提交查询请求“查询距离我当前位置最近的 K 个酒吧”,位置和查询内容都可能作为推断出用户身份、生活习惯及经济状况等隐私信息的条件[1].从信息熵的角度看,用户释放的自身信息越少,攻击者推断其隐私信息的可能性就越低.因此,用户在提交查询请求时,尽量少提供或不提供精确的位置和查询内容,同时希望获得精确的查询结果.位置隐私保护通常采用模糊泛化的方法实现,对查询请求内容的保护 PIR 是目前行之有效的保护方法之一.

位置隐私保护方法通常采用构造匿名区和生成假位置两类[8].构造匿名区的方法是构造包含用户在内的 k–1 个其他用户的区域,以该区域替代自己的真实位置发起查询,以实现 k 匿名(k-anonymity)的方式确保攻击者无法确定 k 个用户中哪个是发起查询的真实用户.这种方法虽然较为常用,但 LBS 服务器必须具备处理非精确地理坐标查询的能力,且在用户移动连续查询时难以构造包含多数初始用户的连续匿名框[8].假位置方法是用户一次提交 1 个[13−16]或多个[11,12]虚假的位置用来替代自己的真实位置,使攻击者难以确定用户的真实位置,然而提交多个虚假位置会给 LBS 服务器带来额外的负担.但假位置却具有构造灵活、易处理、查询精度高等优势,适用于路网动态用户的连续查询,但需要解决假位置替代用户真实位置后如何确保查询准确率的问题.

作为假位置方法的一种,锚点[13−16,22]技术是以一个假位置(锚点)替代用户真实位置发起查询请求,用户以锚点为起点采用增量近邻查询的方式逐步扩大获取近邻兴趣点候选集,并通过两个被称为供应空间和需求空间的大小变化来控制查询开始和结束,如图 1 所示.用户根据 LBS 服务器返回的结果候选集计算出距离自己最近的 K 个兴趣点.该方法(SpaceTwist)由 Yiu 等人[13]首次提出,在无需提供用户真实位置的条件下,实现了精确的 KNN 兴趣点查询.但该方法也存在未实现 k 匿名、锚点随机生成等缺陷[15,16,22,23].虽然 Yiu 等人提出过改进方法[14],国内外学者也针对上述问题分别提出了 KAWCR[15]、Coprivacy[16]、HINN[22,23]等改进方案,但这些方法仍基于欧氏空间设计,在实际路网环境中并不适用[24−26].而且锚点位置随机生成,并没有考虑锚点位置的选取对查询效率的影响.LBS 服务器在面临频繁变换的锚点位置时,查询效率显著降低.2015 年,我们面向路网环境提出了利用锚点实现兴趣点 KNN 查询的解决方案[24],在方法的实用性、查询效率及位置隐私保护等方面取得了一定的进展.但没有解决查询内容隐私保护问题,仍存在查询内容关联攻击的可能.但锚点技术在构造灵活性、易处理、易共享、便于用来组织兴趣点分布信息等方面仍然具有显著优势.本文拟采用锚点技术,在先前研究的基础上提出一种路网环境中无需提供用户真实位置和查询内容的精确 KNN 兴趣点查询方法.



(a) 初始阶段 (b) 查询阶段 (c) 结束阶段

Fig.1 Workflow diagram of SpaceTwist

图 1 SpaceTwist 工作示意图

私有信息检索(PIR)[4,5,8,18]是查询内容隐私保护最有效的方法之一.该方法以 LBS 服务器为不可信实体,用户通过向 LBS 服务器提交一个经过加密处理后的查询请求,服务器在不知晓用户真实请求内容的情况下,依然能够为用户返回查询结果,并且服务器无法确定为用户提供了哪个具体结果,实现了秘密获取兴趣点查询结果的目标[4,5].但这种方法存在计算开销较大、连续查询时效率较低及大多基于欧氏空间设计等缺陷[3,8],并且很多情况下要依赖一个可信设备来对数据库和用户的真实位置信息进行预处理,存在诸多不安全因素[18,19].特别是在实际路网环境下,行进中的用户会多次以不同位置发起相同的兴趣点连续查询请求,频繁的查询请求会给 LBS 服务器带来巨大的处理负担,因此需要改进.文献[25–27]为本文提供了路网查询方法参考.

近年来,我们借鉴私有信息检索的思想提出了保护查询内容的盲查询协议[28].该方法基于伪随机置换方法,具有良好的查询效率.但该方法依然不适用于路网环境,且索引结构不够高效.针对此问题,我们进一步提出了改进方案,但依然要依靠可信中心服务器[29].Yi 等人[3]针对 PIR 在查询协议复杂、连续查询及查询粒度等方面的缺陷,借鉴 Paillier-RSA 密码系统,提出了一个保护查询内容的兴趣点秘密检索协议.然而,该协议仍然基于欧氏空间设计,将用户位置限定在某个网格空间内,存在严重位置隐私泄漏风险.在查询过程中,由于用网格代替用户位置查询,会带来查询结果不精确或查询范围大等问题;并且,同态加解密处理在连续查询中存在实际效率低的问题.

综上,我们针对 LBS 查询中位置隐私和查询内容隐私保护问题,提出了用户无需提供真实位置和查询内容的路网 K 近邻兴趣点精确查询方法.该方法通过路网锚点共享实现位置隐私保护,并基于锚点组织兴趣点分布信息,面向现有私有信息检索方法中普遍存在的无法区分兴趣点类型、连续查询效率差、多基于欧氏空间设计等问题,基于 Paillier 密码系统[30]实现了保护查询内容隐私的 K 近邻兴趣点秘密检索方法.该方法无需引入可信的中间实体,最终实现对用户位置隐私和查询内容隐私保护的目标.

2 预备知识

本节首先介绍 Paillier 密码系统,然后定义系统架构并解释所涉及的概念,最后是对路网模型的描述.

2.1 Paillier公钥密码系统

Paillier 公钥密码系统由密钥生成、加密和解密这 3 个过程组成.

- 密钥生成过程:用户随机选取两个不同的大素数 p,q 满足 $gcd(pq,(p-1)(q-1))=1$,计算 $N=pq$,选取一个整数元 g,满足 $g\in\mathbb{Z}^*_{N^2}$,然后公布公钥 $pk=(N,g)$,保存私钥 $sk=(p,q)$作为秘密.
- 加密过程:给定用户公钥 pk,可以用来加密一个消息 m,其中,m 是一个不超过 N 的正整数,从集合 $\mathbb{Z}^*_{N^2}$ 中随机选取整数 r,并通过 Enc 计算出 m 的密文 c:

$$c=Enc(m,pk)=g^m r^N(\mathrm{mod}\ N^2) \tag{1}$$

- 解密过程:对密文 c 的解密可以利用私钥通过如下计算过程 Dec 实现,其中,$\lambda=lcm(p-1,q-1)$.

$$m=Dec(c,sk)=\frac{(c^\lambda(\mathrm{mod}\ N^2-1))/N}{(g^\lambda(\mathrm{mod}\ N^2-1))/N}(\mathrm{mod}\ N) \tag{2}$$

Paillier 是一种随机非对称加密算法,对于任意元素 $m_1,m_2,a\in\mathbb{Z}^*_N$,满足如下同态特性:

$$Enc(m_1)Enc(m_2)=Enc(m_1+m_2) \tag{3}$$

$$Enc(m_1)^a=Enc(am_1) \tag{4}$$

2.2 系统架构及相关定义

本文采用“用户-LBS 服务器”两方实体架构,除自身外,用户认为任何实体都存在隐私泄漏的可能.定义如下.

定义 1(架构内实体集 *E*). 架构内存在两类实体,用 2 元组 $E=(U,LBS)$表示,其中,U 表示用户集合,用户 $u_k\in U$ 携带具有 GPS 等感知功能模块的移动智能终端;LBS 表示位置服务器集合.

为了解决单一 LBS 服务器在面对大量用户服务请求时的瓶颈问题,一定数量的 LBS 服务器部署在路网环境中,每个 LBS 服务器负责为本区域内的用户提供兴趣点查询服务.在服务开始前,先为用户提供负责区域内兴趣点分布情况索引,用户据此提出查询请求.用户认为,LBS 服务商同样存在隐私泄漏的可能.

定义 2(不可追踪性). 不可追踪性是指攻击者通过用户查询请求信息并结合掌握的背景知识分析出来用户连续位置(轨迹)隐私信息的概率很低.

不可追踪性主要是针对用户位置隐私而言.背景知识由攻击者掌握,用户很难知晓攻击者掌握什么程度的知识信息,而用户可以在查询请求中通过少释放涉及个人位置等信息的方式来降低被追踪的可能.

定义 3(不可关联性). 不可关联性是指查询内容 C 与用户集合 U 中的任意用户 u_k 相互关联的概率很低,即

查询内容无法映射到某个具体用户身份标识上.

不可关联性主要是针对用户的查询内容隐私而言,用户连续发出相同查询请求内容依然可以作为关联条件,攻击用户深入隐私信息.与假名变换和位置模糊泛化不同,查询内容很难变换、模糊泛化,因此对查询内容的保护较难实现,需要借助秘密信息检索技术.

定义 4(私有信息检索). 假设除用户自身外,其他实体均不可信,用户秘密向 LBS 服务器检索其第 i 条记录的过程中,LBS 服务器无法知晓用户的查询内容,并依然可以满足用户查询需求,则称实现了私有信息检索或秘密检索.本文中实现兴趣点的秘密检索,实际是指在保护用户隐私的条件下获取查询结果的过程.

定义 5(位置隐私). 对于用户 u_k 发起的任何查询请求 $Q=\langle u_k',loc,time,K,C\rangle$,如果存在方法 f,使得:

$$f:(k_{adv},\langle u_k',loc,time,K,C\rangle)\rightarrow u_k \tag{5}$$

成立,且当 loc 为用户真实位置坐标时,则称 loc 为位置隐私.其中,k_{adv} 为攻击者掌握的背景知识集;u_k 表示用户真实身份标识;u_k' 表示查询中使用的假名,每次查询均变换不同假名,用户生成的假名之间无关联;loc 是经过去模糊化后得到的真实位置坐标,其余为辅助输入,是否需要辅助输入取决于攻击者对方法 f 的构造能力;f 可以是攻击者自运行的算法,也可以是与其他实体交互的协议或策略等.

位置隐私泄漏以攻击者获取用户真实位置为前提,并且要实现从该位置到某一具体用户身份的唯一映射.用户查询内容隐私及其他个人隐私均可通过类似方法定义.可见,不实现身份映射的任何隐私信息对于攻击者来说没有任何意义.用户公布的任何涉及个人隐私的数据均存在用来推断其真实身份的可能,隐私的私密性决定了任何第二方实体均是不可信的.隐私保护的最有效方法就是什么也不公布,但这也给获取服务带来了难题.

2.3 路网模型

行驶在路网中的用户,行驶方向和行驶距离受路网约束,如图 2 所示,有向路网图相关概念定义如下.

定义 6(有向路网图). 有向路网图可用二元组 $digraph=(Vex,Edg)$表示,其中,Vex 表示顶点集合,$v_i\in Vex$ 表示某个路网顶点,任意一条路段两端顶点称为邻接顶点;Edg 表示有向边集合,$\overrightarrow{v_mv_n}\in Edg$ 表示路网中的某条有向路段(简称路段),该路段表示从顶点 v_m 的一条出边或 v_n 的一条入边,行进在路段 $\overrightarrow{v_mv_n}$ 上的某个用户 u_k 或该路段上的兴趣点 p_i 可以表示为 $u_k\in\overrightarrow{v_mv_n}$ 或 $p_i\in\overrightarrow{v_mv_n}$,$v_n$ 称为用户 u_k 所在路段的正方向顶点.

定义 6 中的路段方向也用来表示用户的行进方向,同时也表明兴趣点位于哪一个路段.如图 3 所示,$p_3\in\overrightarrow{v_5v_9}$ 与 $p_5\in\overrightarrow{v_9v_5}$ 表示分别位于用户 u_k 行驶方向左、右两侧的兴趣点.

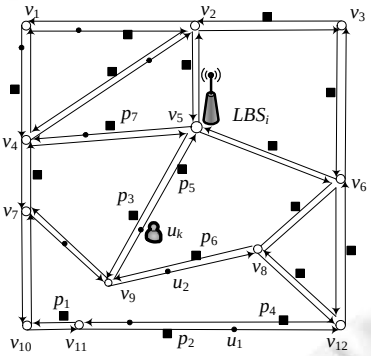


Fig.2 Road network model

图 2 路网模型

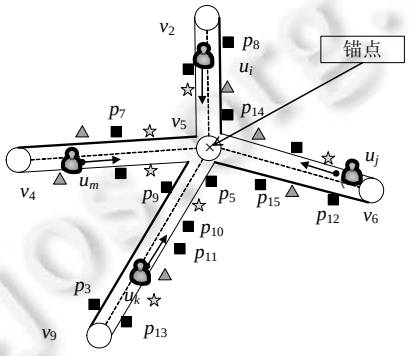


Fig.3 Local description of road network

图 3 路网局部详细描述

定义 7(路网距离). 路网边上任意两点 v_i 与 p_i 最短行驶路径长度 $dist_r(v_i,p_i)$称为路网距离.如图 3 所示,$dist_r(v_5,p_5)=dist_e(v_5,v_9)+dist_e(v_9,p_5)$,其中,表达式 $dist_e(v_5,v_9)$表示两点之间的欧氏距离,并假设用户均沿路段行进,不存在跳跃路段障碍行进可能.

由定义 7 可知,$dist_r(v_i,p_i)$与 $dist_r(p_i,v_i)$可能并不相等,且欧氏距离近的点可能路网距离较远.如图 3 所示,顶点

v_5 到兴趣点 p_5 的路网距离比到 p_3 的更大,而 v_5 到 p_5 的欧氏距离却相对更小.

定义 8(锚点). 即被用户选择替换自己真实位置发起查询的路网顶点,非顶点位置不可选作锚点.

用户通常选择所在路段正方向顶点作为查询锚点,同一锚点可以被不同路段上的用户复用.如图 3 所示,用户 u_i,u_j,u_k 及 u_m 均可以选择 v_5 作为锚点,当这些用户共用同一个锚点时,初步实现了位置匿名,当路网上的所有路网用户均用锚点发起查询时,从攻击者角度看,全部查询均从固定位置发出,起到了模糊泛化用户真实位置的作用,进一步增加了攻击者关联用户连续位置的难度.从用户角度看,锚点位置并非随机生成而是固定的,相对于其他使用随机锚点的查询方法,这种固定锚点有利于提高查询结果的复用率.

通过路网定义,我们可以发现,用户在行进过程中总会经过不同的路网顶点,用户要达到其他路段的某个兴趣点,必须先到达当前路段所指向顶点,并从该顶点出发到达目标兴趣点.由此可见,兴趣点是依据路段分布的,以路网顶点为基本单位组织路网兴趣点分布信息,便于查找 K 近邻兴趣点;同时,具有同一指向顶点的不同路段上的用户,以同一路网顶点作为锚点发起查询时,攻击者无法确定哪个用户处在哪条路段发起了查询,可以实现对用户位置隐私的保护.

3 路网连续查询中的隐私保护方法

本节分为 3 部分:首先对保护隐私查询流程 3 个阶段进行概述;然后阐述基于路网顶点的兴趣点分布信息组织结构,该结构为下文兴趣点秘密检索提供了实现基础;最后阐述基于 Paillier 密码系统的秘密检索详细过程.

3.1 隐私保护流程概述

如图 4 所示,为了保护用户位置及查询内容隐私,基于 Paillier 的兴趣点秘密检索流程简要描述如下.

① 用户 u_k 依据兴趣点分布索引,以所在路段正方向顶点 v_n 作为锚点向 LBS 服务器发起兴趣点查询请求 $Q=\{u_k',loc_{v_n},time,K,C\}$,$Q$ 中各个元素依次表示用户身份标识(通常使用假名)、第 i 个锚点位置(路网顶点)、查询时间戳、K 近邻兴趣点查询个数及经过后文算法 1 中 Paillier 同态加密处理生成的密文查询内容 C.LBS 服务器无法得知 C 的真实内容,因此无法确定用户查询哪个类型的兴趣点,即假设在 LBS 端存在 m 个类型的兴趣点,其中某个兴趣点类型 $t(t=1,2,\ldots,m)$是用户查询的目标兴趣点,经过后文算法 1 的处理,实现对其秘密标识,而 LBS 服务器并不知道哪个类型兴趣点是经过秘密标识的.在连续查询中,用户每次发起查询都会变换假名且 K 也会不同,以避免由身份标识及 K 等相同属性关联带来对用户连续位置隐私的推断攻击.

② 基于 Paillier 的同态特性,LBS 服务器在密文 Q 上运行后文算法 2,以锚点 loc_{v_n} 为起点,依次在数据库中提取全部 m 个类型兴趣点的 K 近邻查询结果 $(P_1^K,P_2^K,\ldots,P_m^K)$,并利用同态属性将无区别的 m 个兴趣点查询结果作为输入累乘处理,得出一个密文处理结果 R.LBS 无法根据 R 推断出用户目标兴趣点类型,并将其返回给用户.

③ 用户根据返回的密文查询结果 R,通过计算公式并运行后文算法 3 计算得出其所需的精确 K 近邻查询结果.

在连续查询过程中,用户始终无需提供自己的真实位置及查询内容,实现了对位置及查询内容隐私保护,并且能够确保查询结果的准确性,最终实现了对目标兴趣点的秘密检索(PIR).详细实现过程在第 3.3 节中说明.

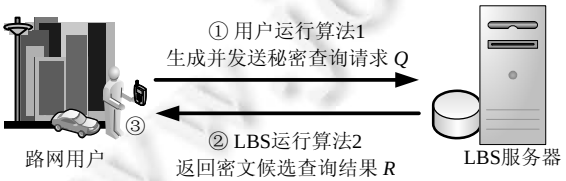


Fig.4 Workflow diagram of POIs query

图 4 兴趣点查询流程图

3.2 基于路网顶点的兴趣点分布结构设计

为了方便用户利用锚点查询,就必须在 LBS 服务端设计适用于路网顶点发起查询的兴趣点分布情况组织方式.如前所述,可采用以路网顶点为基本单位来组织兴趣点分布情况信息.如图 3 所示,以图 3 中顶点 v_5 为例,从 v_5 出发可到达不同类型的兴趣点,图 3 以不同形状的点表示不同类型兴趣点.

定义 9(顶点直达兴趣点). 从某个顶点出发,不经过其他顶点达到的兴趣点,称为顶点直达兴趣点.图 3 中,顶点 v_5 的黑色正方形直达兴趣点按照路网距离由近到远依次有$\{p_{14},p_9,p_{15},p_7,p_8,p_{12},p_3\}$,其他同类黑色正方形兴趣点不属于顶点 v_5 的直达兴趣点,如$\{p_{11},p_{10},p_5\}$等,属于其他顶点的直达兴趣点.

由此,可以根据顶点直达兴趣点的定义,以路网空间内所有顶点为基础,构造 LBS 端的路网兴趣点分布情况信息表,见表 1.其中,第 1 行每个列名依次表示路网顶点的位置坐标(即用户查询时所用的锚点)、邻接顶点集合、不同类型兴趣点的标号、兴趣点类型名、$K_{\max}$ 个此类近邻兴趣点坐标集合和每个此类兴趣点的详细描述信息.每个锚点中,不同类型兴趣点标号唯一.$K_{\max}$ 表示从该顶点出发此类型的近邻兴趣点个数的一个上限值,由 LBS 服务商预先存储在数据库中,并定期更新.该集合中的兴趣点可能分布在不同路段上,用户发起 K 近邻查询时满足 $K\leqslant K_{\max}$,并且集合中所有兴趣点按照从该顶点出发的距离递增排序,兴趣点描述信息列中的每条详细信息按相同顺序依次排列.本文采用分布式数据库,每个 LBS 服务器只负责管理所在区域内的兴趣点分布表,并响应区域内用户的查询请求,不同区域内的兴趣点分布表不同.这种分布式结构有利于提高查询效率.

Table 1 Structure of POIs distribution in LBS

表 1 LBS 端兴趣点分布结构

路网顶点(锚点)	邻接顶点	兴趣点标号	兴趣点类型	$K_{\max}$ 近邻兴趣点集	兴趣点描述信息
v_1 (anchor 1)	$\{v_2,v_4,v_5,v_7\}$	1	Hospital	$\{loc_1,loc_2,...,loc_{K_{\max}}\}$	$\{\inf_1,\inf_2,...,\inf_{K_{\max}}\}$
		2	Bank	$\{loc_1,loc_2,...,loc_{K_{\max}}\}$	$\{\inf_1,\inf_2,...,\inf_{K_{\max}}\}$
		3	Hotel	$\{loc_1,loc_2,...,loc_{K_{\max}}\}$	$\{\inf_1,\inf_2,...,\inf_{K_{\max}}\}$
		4	Gas station	$\{loc_1,loc_2,...,loc_{K_{\max}}\}$	$\{\inf_1,\inf_2,...,\inf_{K_{\max}}\}$
v_2 (anchor 2)	$\{v_1,v_3,v_6\}$	1	Hospital	$\{loc_1,loc_2,...,loc_{K_{\max}}\}$	$\{\inf_1,\inf_2,...,\inf_{K_{\max}}\}$
		2	Bank	$\{loc_1,loc_2,...,loc_{K_{\max}}\}$	$\{\inf_1,\inf_2,...,\inf_{K_{\max}}\}$
		3	Hotel	$\{loc_1,loc_2,...,loc_{K_{\max}}\}$	$\{\inf_1,\inf_2,...,\inf_{K_{\max}}\}$
…	…	…	…	…	…

部署在各个区域内的不同 LBS 服务器以路网顶点为基本元素,掌握自己所负责区域内的全部路网顶点的兴趣点分布情况,并由表 1 生成兴趣点分布情况索引表,见表 2.该表在查询开始前公布给用户.用户首先根据索引表确定所在路段,如计算用户位置到两个邻接顶点之间距离的方法等[24,31],以确保后续对锚点的有效选择,此类定位方法较多,此处不再赘述.

Table 2 Distribution index of POIs in LBS

表 2 LBS 端兴趣点分布索引

路网顶点(锚点)	邻接顶点	兴趣点标号	兴趣点类型
v_1 (anchor 1)	$\{v_2,v_4,v_5,v_7\}$	1	Hospital
		2	Bank
		3	Hotel
		4	Gas station
v_2 (anchor 2)	$\{v_1,v_3,v_6\}$	1	Hospital
		2	Bank
		3	Hotel
…	…	…	…

索引表 2 提供了用户所在位置一定范围内各类兴趣点沿路网分布情况,是一种分布式结构化的广播索引,其中,路网顶点和邻接顶点用于帮助用户确定所在路段.兴趣点类型表示从该对应顶点出发可以到达区域内哪些类型的兴趣点,兴趣点标号唯一标识一类兴趣点.用户通过同态加密对标号进行处理,生成对某类型兴趣点的秘密检索请求.该表周期广播给 LBS 服务器负责区域内的用户.用户在确定自身所在路段后,首先选取当前所在

路段正方向顶点 v_n 作为锚点 anchor n,然后通过后文算法 1 对该锚点下所有的兴趣点类型标号做同态加密处理,并秘密标记其感兴趣的目标兴趣点类型,最终生成查询请求 Q 发送给 LBS 服务器.而 LBS 服务器无法依据 Q 知晓用户查询了哪类兴趣点,但依然可以据此生成密文查询结果返回给用户.因此,用户需要首先获取索引表 2.

3.3 保护位置和查询内容隐私的*K*近邻查询方法

由用户发起的 K 近邻兴趣点秘密检索流程可以概括为用户端查询请求生成、服务端对查询请求的处理及用户端对查询结果求精这 3 个过程,依次对应后文算法 1~算法 3.用户基于同态加密 *Enc* 执行后文算法 1,生成秘密检索请求 Q;LBS 服务器基于同态属性对密文 Q 执行后文算法 2,生成密文查询结果 R;用户依据相应公式解密查询结果,并依据自身实时位置运行后文算法 3,生成路网环境下的精确 K 近邻查询结果.

(1) 秘密检索请求的生成、处理

LBS 服务器针对不同顶点将各类兴趣点进行分类处理,将从某个顶点 v_n 出发的全部 m 个类型的兴趣点依次标号,如表 1 第 3 列所示,并以二进制数的形式存储,用户通过标号来表示自己的目标查询兴趣点.用户在收到 LBS 服务器发来的区域内兴趣点分布情况索引表 2 后,根据索引表兴趣点类型标号生成的查询请求可表示为 $Q=\langle u_k', loc_{v_n}, time, K, C\rangle$.为了避免自身查询内容隐私泄漏,其中的 C 是用户对目标兴趣点类型经过 Paillier 同态加密处理后的密文查询内容,处理过程如算法 1 所示,确保 LBS 服务器不知道其查询兴趣点类型(即查询内容隐私).其中,用户针对自己的目标兴趣点类型标号 $t\in(1,2,\ldots,m)$,获取从所在路段正方向顶点 v_n 出发的 $K_{\max}$ 个近邻兴趣点结果,可表示为 $P_{v_n,t}^{K_{\max}}$, 该类型兴趣点的 $K_{\max}$ 个结果按照到顶点 v_i 的路网距离递增排序,而用户发起某个类型 t 的 $K(K\leqslant K_{\max})$近邻兴趣点查询可表示为 $P_{v_n,t}^{K}$ .LBS 服务器端也需在查询自己的数据库表时执行特定的算法,使 LBS 服务器既能根据密文查询内容 C 返回给用户其所需的兴趣点查询结果,同时其自身又无法确定返回给用户什么样的查询结果.处理过程如后文算法 2 所示.

算法 1. 用户端密文查询请求生成算法.

1. 输入:用户目标兴趣点类型(如医院、银行等),近邻兴趣点查询个数 K.
2. 输出:用户查询请求 Q、加密算法 *Enc* 的公私钥对(pk_{pail}, sk_{pail}).
3. 用户获取 LBS 服务器公布的索引表 $IndexTable_{lbsi}$
4. 确定自己所在路段 $u_k\in\overrightarrow{v_mv_n}$,正前方顶点为 v_n
5. 依据索引表 $IndexTable_{lbsi}$ 确定顶点 v_n 中的目标兴趣点类型的标号为 $t\in(1,2,\ldots,m)$
6. 随机选取两个大素数 p,q,使得 $N_{pail}=pq$
7. 依据公式(1)生成 Paillier 加密算法 *Enc* 的公私钥对 $pk_{pail}=\{N_{pail},g\}, sk_{pail}=\{p,q\}$
8. 选取两个不相等的整数α,β
9. 在 $IndexTable_{lbsi}$ 中提取所在路段正前方顶点 v_n 中的全部 m 个兴趣点类型标号
10. **for** 每个兴趣点类型标号 $x\in(1,2,\ldots,m)$
11. 生成随机数 $r_x\in\mathbb{Z}_{N_{pail}^2}^*$
12. **if** $x=t$ **then** //类型标号是用户的目标兴趣点
13. $c_x=Enc(\alpha, pk_{pail})=g^{\alpha}r_x^{N_{pail}}(\bmod N_{pail}^2)$
14. **else** $c_x=Enc(\beta, pk_{pail})=g^{\beta}r_x^{N_{pail}}(\bmod N_{pail}^2)$
15. **end if**
16. **end for**
17. 将上面生成的每个 c_x 表示为密文查询内容集 $C=(c_1,c_2,\ldots,c_m)$
18. 选取所在路段正前方顶点 v_n 作为锚点,选取假名 u_k' ,生成查询请求 $Q=\langle u_k', loc_{v_n}, time, K, C\rangle$
19. 返回查询请求 Q 及 Paillier 加密算法 Enc 的公私钥对(pk_{pail}, sk_{pail})给用户
20. **END**.

算法 1 中,用户需要执行 m 次 Pailiier 加密过程,其计算复杂度可以用 $O(m)$表示.由算法 1 第 13 行可见,用户在生成查询请求过程中,用α对自己目标类型兴趣点标号 t 进行了秘密标识,其他类型兴趣点类型标号均用β来运算处理.由于算法 1 对每个类型标识符在加密处理时均选取不同时随机数 r_x,使加密结果未表现出任何特征,因此,LBS 服务器及其他潜在攻击者无法通过密文查询请求确定用户真实查询请求.由此,用户根据算法 1 生成的密文查询内容集 C,并选取当前所在路段正方向顶点 v_n 作为查询锚点,向 LBS 发起查询请求 $Q=\langle u_k', loc_{v_n}, time, K, C\rangle$,由于 C 是经过算法 1 处理后的密文查询内容,因此 LBS 服务器并不知晓用户对 m 种类型兴趣点中的哪个类型感兴趣,然而 LBS 服务器如何实现在不知晓用户目标兴趣点类型的条件下为用户提供密文查询结果 R,即 LBS 服务器也不知道提供了哪类兴趣点的查询结果.因此,需要依据查询请求及 Paillier 的同态特性进一步做特殊处理.LBS 服务器处理过程如算法 2 所示.

算法 2. LBS 服务器端查询请求处理算法.

1. 输入:用户查询请求 $Q=\langle u_k', loc_{v_n}, time, K, C\rangle$ 、LBS 服务器端路网顶点兴趣点存储表 $Table_{lbsi}$.
2. 输出:密文查询结果 R.
3. LBS 服务器依据 Q 确定顶点 v_n
4. **for** $Table_{lbsi}$ 中顶点 v_n 的每个类型兴趣点数据集合 $P_{v_n,t}^{K_{\max}}, t=1,2,...,m$
5. 提取前 K 个兴趣点集 $P_{v_n,t}^{K}$,用 w_t 表示每个 $P_{v_n,t}^{K}$
6. 按序将每个 w_t 存储在集合 W 中
7. **end for**
8. 提取 Q 中的集合 C
9. **for** 每个 $c_x \in C$ 及每个 $w_t \in W$,其中,$x=t=1,2,\ldots,m$
10. 累乘计算 $R=\prod_{x=1}^{m} c_x^{w_x} (\mathrm{mod}\, N_{pail}^2)$
11. **end for**
12. 返回密文查询结果 R 给用户
13. **END**.

LBS 服务端收到用户查询请求后执行算法 2.由算法 2 第 9 行、第 10 行可知,LBS 服务器提取查询内容 C 中的每个兴趣点类型标号密文 c_x,并结合自己数据库表中提取表 1 中该顶点 v_n 的 m 个类型的兴趣点 K 近邻结果(主要为表 1 中的最后两列),基于同态属性累乘计算出一个结果密文结果 R,并将其返回给用户 u_k.该算法的计算复杂度为 $O(m^2)$.用户需要根据自己掌握的私钥,通过公式(2)描述的 Paillier 解密算法 Dec 解密计算出自己想要的类型为 t 的 K 近邻目标兴趣点信息 $P_{v_n,t}^{K}$,如式(6)所示,从而实现目标兴趣点的秘密检索.

$$P_{v_n,t}^{K}=Dec(R, sk_{pail}) \tag{6}$$

然而,该结果仅是以用户所在路段正方向顶点为起点获取的 K 近邻查询结果,而由于用户行进在路段的某个位置上,该查询结果并非基于用户当前位置获得的最终 K 近邻结果.为了确保查询的准确性,用户需要进一步结合自身当前位置对查询结果进行求精,计算出距离自身最近的 K 个目标兴趣点集合 $P_{u_k,t}^{K}$,而不是 $P_{v_n,t}^{K}$,求取精确的查询结果如后文算法 3 所示.

(2) 精确查询结果生成

为了实现位置隐私保护效率和实用性,根据当前路段正方向顶点必达性的特点,如图 3 所示,路网用户 $u_k \in \overrightarrow{v_9v_5}$ 可以选取当前路段指向顶点 v_5 作为锚点,替代自身位置发起查询.在连续查询中,用户无需频繁提交实时变化的位置发起连续查询,用户仅需以当前路段正方向顶点 v_5 为锚点发起 1 次查询即可.用户根据表 1“$K_{\max}$ 近邻兴趣点集”列中提供的 $K \leqslant K_{\max}$ 近邻信息,获取 K 近邻兴趣点查询结果.同时,用户可以依据索引表 2 提供的邻接顶点信息,以较高概率判断出自己当前所在路段的顶点和即将进入的下一路段顶点[24],即使没有进入预测的路段,依然可以根据刚刚获取的结果,并结合新路段信息获取精确查询结果.当用户进入下一路段 $\overrightarrow{v_5v_2}$ 时,以新的路

段指向顶点 v_5 发起新一次连续查询即可.当路网用户均以锚点发起查询时,攻击者无法判断用户所在该顶点入边具体路段,实现了用户位置隐私保护.因此,以路网顶点作为查询锚点的方法不仅可以保护用户的位置隐私,同时还有效降低了连续查询过程中的查询次数,提高了查询效率及查询准确性.

然而,从查询准确度来看,用户在第 1 次发起查询时,如图 3 所示,如果以当前顶点 v_5 发起实心正方形类兴趣点的 K 近邻查询结果并不准确.这是因为用户 u_k 前进的正方向上依然有 3 个同类兴趣点$\{p_{11},p_{10},p_5\}$距离用户更近,而根据定义 9 及锚点查询方式可知,从锚点 v_5 出发的 K 近邻查询结果中,这 3 个兴趣点很可能并不是距离用户路网距离最近的 3 个顶点.为了解决这个问题,如果用户采用直接获取自己当前路段兴趣点的方式,会泄漏用户所在路段的隐私信息.为此,在起始阶段,用户需要以路段 $\overrightarrow{v_9v_5}$ 的起始点 v_9 发起一次 K 近邻查询,确定自己前方目标兴趣点的个数 X,然后结合自身当前位置和以锚点 v_5 为出发点的$(K–X)$近邻查询结果,获得精确完整的 K 近邻查询结果,避免因兴趣点疏漏而造成的查询不精确的问题,K 近邻兴趣点精确结果计算算法如下.

算法 3. 用户端 K 近邻兴趣点精确结果计算算法.

1. 输入:用户 $u_k \in \overrightarrow{v_mv_n}$ 自身当前位置坐标 loc_{u_k},当前路段两个端点坐标 loc_{v_m},loc_{v_n}.
2. 输出:从用户当前位置出发的精确 K 近邻查询结果.
3. **if** 缓存中不存在以 v_m 为锚点的目标兴趣点 K 近邻结果
4. 调用算法 1,以 loc_{v_m} 为锚点发起目标兴趣点 K 近邻查询,获取目标兴趣点类型 t 的结果集合 $P^K_{v_m,t}$
5. **end if**
6. 获取当前位置 loc_{u_k}
7. **while** ($(loc_{u_k} - loc'_{u_k}) \geqslant \zeta_{\max}$) // loc'_{u_k} 表示上一次获取的位置,$\xi_{\max}$ 为某一距离阈值
8. 计算当前位置到所在路段指向顶点路网距离 $dist_r(loc_{u_k},v_n)$
9. **for** 每个兴趣点 $p_i \in P^K_{v_m,t}$
10. **if** $dist_r(p_i,v_n) \leqslant dist_r(loc_{u_k},v_n)$ //判断用户前目标兴趣点
11. 将兴趣点 p_i 放入集合 $P^X_{u_k}$
12. **else** 丢弃兴趣点 p_i
13. **end if**
14. **end for**
15. 计算集合 $P^X_{u_k}$ 的势 $X = |P^X_{u_k}|$
16. 调用算法 1,以 loc_{v_n} 为锚点发起目标兴趣点$(K–X)$近邻查询,获取结果集合 $P^{(K-X)}_{v_n}$
17. 返回精确 K 近邻查询结果集 $P^K_{u_k} = P^X_{u_k} \cup P^{(K-X)}_{v_n}$
18. **end while**
19. **END**.

算法 3 第 7 行用来判断用户位置变化情况,当交通拥塞时(连续位置更新变化较小),则无需不断重复计算 K 近邻兴趣点;当用户位置出现较大变化时(如前进了一定的距离、拥塞缓解或通过了路口信号灯等),再次开启连续查询.第 9 行~第 14 行将用户当前位置前的目标兴趣点保留下来,丢弃其他兴趣点,用户确定了前方兴趣点数量后,以前方顶点为锚点发起$(K–X)$近邻查询.由于用户每次前方兴趣点数量不同,所以,可能用户每次以锚点发起的近邻查询 K 值会有不同.当用户进入到下一路段后,依然可以依据上一次路网顶点为锚点的查询的结果,运行算法 3 得出当前所在路段中用户位置前方的兴趣点个数$(K–X)$,再次发起新查询,减少查询次数.由于用户首次查询需要计算所在路段兴趣点并调用算法 1 生成查询请求,因此算法 3 的计算复杂度为 $O(m+k)$.

4 性能分析

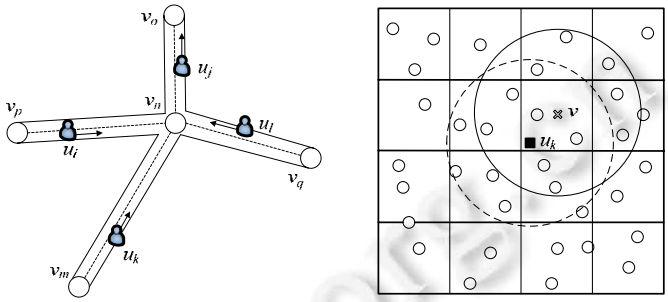
本节将从准确性、安全性和工作效率这 3 个方面对所提出方法的性能做出分析.通过分析可以看出,本文

所提出的方法能够在确保用户位置隐私和查询内容隐私的基础上,确保较高的查询准确度.

4.1 准确性

基于位置的服务不应因为隐私保护需求而降低服务质量,用户体验较差的服务很难获得认可.本文提出,既不提供真实位置信息,也不提供查询内容明文,确保查询结果准确性是首要问题.

一方面,在不提供用户精确位置的条件下,我们基于路网兴趣点沿路段分布及其可达性的特点,设计了基于路网顶点的兴趣点分布情况组织结构.如图 5(a)、图 5(b)所示.



(a) 路网空间兴趣点查询 (b) 欧氏空间兴趣点查询

Fig.5 Comparison of querying with an anchor and a cloaking region

图 5 利用锚点和利用匿名框查询比对

为了保护位置隐私而将用户所在网格作为匿名框查询获得的查询结果(以 v 为圆心的实线圆内的兴趣点),与用户真实位置获得的查询结果(以 u_k 为圆心的虚线圆内的兴趣点)具有明显的差异,这种基于欧氏空间设计的查询方法在建筑体等障碍物阻挡的情况下,可能也不是距离用户最近的兴趣点.因此匿名框方法在实际查询中,准确性相对较低.本方法考虑了路网限制因素,某个路段上的用户 $u_k \in \overrightarrow{v_m v_n}$ 如果想到达目标兴趣点,必然会首先到达当前所在路段正方向顶点 v_n,并以此为出发点到达 K 近邻兴趣点,因此,以路网顶点为锚点的查询在查询准确度方面具有显著的优势.考虑到用户当前路段 $\overrightarrow{u_k v_n}$ 上依然存在目标兴趣点的可能,算法 3 将精确的用户 K 近邻兴趣点集 $P_{u_k}^K$ 分成两个集合分布逐步查询 $P_{u_k}^X \cup P_{v_n}^{(K-X)}$, 其中, $P_{u_k}^K$ 根据用户上一次的依据顶点 v_m 发起的查询结果中通过计算获取, $P_{v_n}^{(K-X)}$ 以为锚点 v_n 的新一次查询中获得,用户通过计算获取距离自身真实位置最近的 K 近邻兴趣点集,确保查询结果的准确性.当用户驶入新路段后 $u_k \in \overrightarrow{v_n v_o}$,则以新路段正方向顶点 v_o 发起新的查询.由此可见,基于路网顶点的查询方式不仅可以解决传统欧氏空间 K 近邻查询中存在的因障碍物阻挡等因素带来的查询结果与实际情况不符的问题,还能确保连续查询过程中的结果准确性.

另一方面,在用户不提供真实查询内容的条件下,为了确保用户能够获取目标类型兴趣点查询结果,我们基于 Paillier 密码系统给出了秘密信息检索交互过程.由于兴趣点分布信息是依据路网顶点组织的,并且每个类型兴趣点的 $K_{\max}$ 近邻查询结果上限值在 LBS 服务器端已经预置好,在秘密检索过程中,任意兴趣点类型 t 的 K 近邻($K \leqslant K_{\max}$)查询结果明文 $P_{v_n,t}^K$, 用户均可以通过 $P_{v_n,t}^K = Dec(R, sk_{pail})$ 解密获得.这是因为算法 1、算法 2 和公式(6)中的密文查询处理过程等价于 Paillier 算法对目标类型兴趣点 c_t 的一次加解密过程,即有:

$$R = \prod_{x=1}^{m} c_x^{w_x} = g^{P_{v_n,t}^K} \left(\prod_{x=1}^{m} r_x^{P_{v_n,t}^K} \right)^{N_{pail}} (\mathrm{mod}\, N_{pail}^2) \tag{7}$$

成立.因此,根据 Paillier 的同态属性,用户可以结合私钥 sk_{pail} 解密获得目标类型 K 近邻兴趣点查询结果明文.因此,本方法可以确保对任意顶点和任意兴趣点类型的秘密检索结果解密的准确性.

4.2 安全性

由定义 5 可知,在一次完整的连续查询中,安全性是指用户在使用不同路网顶点发起查询过程中,方案中不

存在可以推断出用户真实身份的关联项,用户每次查询可表示为 $Q=\langle u_k',loc_{v_n},time,K,C\rangle$.这是用户唯一向外界提供的信息,其中任意两次查询之间假名 u_k' 无关联,时间 *time* 和查询兴趣点数量 *K* 用来作为关联项的概率较低,用户位置和查询内容是本方案中可以关联多次查询并推断用户真实身份隐私关键因素.即整体方案安全性取决于用户位置不可追踪性和内容不可关联性的实现.下面通过信息熵证明连续查询中位置的不可追踪性,通过语义安全性证明连续查询中查询内容的不可关联性,从而说明本方案在连续查询中的整体安全性.

(1) 不可追踪性

如前所述,行驶在路网中的用户不断发起 LBS 兴趣点查询服务,如果为了保护位置隐私而采用构造匿名框的方法,由于用户移动的特点,初始构造的匿名框在连续查询中会因为用户的离开而失效.为了不被唯一锁定,用户需要维持初始匿名框中的绝大多数用户仍在后续构造匿名框中,然而由于用户移动方向及速度的不同,上述目标很难实现,因此在连续查询中,匿名框在用户共用实现位置隐私保护方面效果不佳.针对这个问题,当路网内所有用户均利用所在路段正方向顶点作为锚点查询时,如图 5(a)所示,不同路段上的用户 u_k,u_i,u_l 共用同一个锚点 v_n,相当于构造了一个匿名度 k=3 的匿名框,实现了一次 k 匿名,其中,某个用户 u_k 的实际位置可能位于以顶点 v_n 为入边顶点的任意 4 条路段上.用户到达下一路段后,用新的顶点作为锚点发起查询,依然可以和其他用户共用下一路口锚点,实现连续查询中的位置隐私保护.由于路网锚点不是临时构造的,类似于预先部署很多可以供用户共用的假位置,从全局看,路网内所用用户似乎都在同一个匿名区内,因此相对于临时构造的匿名框,无需考虑保持初始多数用户的问题.

同时,由于用户在查询过程中查询内容 *C* 是不可见的,并且算法 2 中查询内容密文每次均选取不同的随机数 r_x 生成,因此由查询内容关联带来的连续位置隐私泄漏概率几乎为 0.并且根据算法 3,每次 *K* 近邻查询结果 $P_{u_k}^K$ 的计算由并集 $P_{u_k}^X \cup P_{v_n}^{(K-X)}$ 生成,并集中的每次用户当前路段前方兴趣点集不确定,因此用户每次进入新路段后发起新查询的 *K* 值都会不同,这样利用兴趣点查询值 *K* 关联带来的连续位置隐私泄漏概率较低.由定义 2 可知,在用户提交的查询请求 $Q=\langle u_k',loc_{v_n},time,K,C\rangle$ 中,由位置 loc_{v_n} 、查询兴趣点数量 *K*、查询内容 *C* 关联带来的连续位置隐私泄漏概率较低,用户可以采用匿名通信或变换假名的方式解决由身份标识 u_k' 带来的关联隐私泄漏问题,因此,用户每次发起的查询请求所使用的位置坐标可以看作是相互独立的.这样,每次查询都可能是以 loc_{v_n} 为正方向顶点路段上的 *n* 个用户发起的查询,每个用户是真实查询用户的概率为 $p(u_i)$=1/*n*,单次查询的信息熵为

$$H(Q)=-\sum_{i=1}^{n}p(u_i)\log_2 p(u_i)=\log_2 n \tag{8}$$

此时,单次查询的信息熵最大.如前所述,因为用户用不同锚点发起的 *M* 次连续查询请求都是相互独立的,即有 $p(Q_1,Q_2,\ldots,Q_M)=p(Q_1)p(Q_2)\ldots p(Q_M)$成立,此时,*M* 次查询的联合信息熵为

$$H(Q_1,Q_2,...,Q_M)=-\sum_{i=1}^{M}H(Q_i)=M\log_2 n \tag{9}$$

即某用户连续查询中每次查询 Q_i 无显著可用来关联连续多个位置的关联数据项.因此,本方法在抵抗查询项关联带来的连续位置隐私泄漏问题上具有较好的安全性,即能够保证较强的不可追踪性.

(2) 不可关联性

在一次查询请求中,身份标识、位置、时间及兴趣点查询数量 *K* 都可以采用明文替换泛化的方法实现.由于LBS需要根据查询内容为用户提供服务,查询内容无法实现明文泛化,用户需要明确自己的查询内容(即用户对哪一类兴趣点感兴趣).因此,本文采用秘密检索技术来实现查询内容的不可关联性.

本文采用 Paillier 密码系统来实现秘密检索,Paillier 是一个满足语义安全[32]的密码系统,即计算能力有限的攻击者无法从密文和公钥中得到明文的重要信息,语义安全性是本文提出不可关联性的充分条件.本文的语义安全性可以采用 LBS 服务器和挑战者(用户)如下博弈过程描述:给定锚点 loc_{v_n} 和 *m* 个类型兴趣点的 *K* 近邻查询结果,博弈过程如下.

① 对于任意给定的锚点 loc_{v_n},LBS 服务器选择两个不同类型的 *K* 近邻兴趣点结果 x_0,x_1,并将其发送给挑

战者.

② 挑战者随机选择一个比特 $b\in\{0,1\}$,执行算法 1 生成查询请求 Q,并将 Q 及 Paillier 公钥 pk_{pail} 发送给 LBS 服务器.

③ LBS 服务器根据用户返回查询请求及公钥通过执行多项式有限次数的计算,得出一个比特 $b'\in\{0,1\}$,对挑战者选择的比特 b 进行猜测.

通过执行上述步骤,如果 LBS 服务器最终得出的 b'等于 b,则 LBS 服务器攻击成功;否则失败.由此可定义 LBS 服务器在博弈中的概率优势为

$$Prob(b=b'\,|\,(Q,pk_{pail}))=\frac{1}{2}\pm Adv(\lambda) \tag{10}$$

其中,λ为安全参数.如果上述公式中 $Adv(\lambda)$是一个可忽略的函数,对于一个能力为概率多项式时间有界的攻击者,在上述博弈中的成功概率不高于 $1/2\pm Adv(\lambda)$,则称本文所提出的秘密检索方案是语义安全(IND-CPA)[32]的,即 LBS 服务器无法知晓用户查询的兴趣点类型.

这是由于,如果 LBS 服务器猜测结果是 b=1,Q 是用户真实查询,则 LBS 服务器能够以 $1/2\pm Adv(\lambda)$的概率博弈获胜,其中,$Adv(\lambda)$是一个不可忽略的函数;否则,当 b=0 时,由算法 1 计算出的查询请求 Q 与用户身份标识相互独立,LBS 服务器仅能以 1/2 的概率获胜.由此,LBS 服务器猜测出真实结果的概率为 $1/2+1/2+1/2\times(1/2+Adv(\lambda))=1/2+Adv(\lambda)/2$.由此可见,LBS 服务器并不具备不可忽略的优势,处理的查询请求 Q 具有语义安全性,LBS 服务器无法获知其感兴趣的兴趣点类型.由此,当语义安全性获得保证时,LBS 服务器由于无法获取其密文,密文查询请求与用户真实身份相互无关联,因此可以实现定义 3 的不可关联性和定义 4 的秘密信息检索.

综上,通过确保不可追踪性和不可关联性,攻击者很难构造出有效的方法 f,使得公式(5)成立,进而推断出用户的位置和查询内容隐私.因此,整个方案可以确保在连续查询中用户的位置隐私和查询内容隐私安全.

4.3 查询效率

本文提出了基于路网顶点的兴趣点分布信息组织方法,面向查询效率提高和用户隐私保护两个问题,在查询次数和缩小查询范围两个方面具有显著优势.

(1) 查询次数少

首先对比匿名框的查询方法.如图 5(b)所示,处在某个网格内的用户如果以当前网格为匿名框发起查询,则查询结果以当前网格中心 v 为出发点获取 K 近邻查询结果.然而用户 u_k 在该网格边缘,为了确保查询准确性,还需要对多个邻近网格进行查询,增加了查询次数,会带来额外的计算、通信开销.本方法每次只发起 1 次查询即可获得精确的结果,显著降低了 LBS 服务端的计算开销和通信开销.亦无需为了保护隐私提交多个假位置或多个虚假查询内容.本文虽使用密码学方法实现隐私保护,但在实际应用中,特别是在兴趣点较多时,扩展性良好.

其次,在连续查询中,路网锚点仅需在每次进入新路段时提交一次 K 近邻查询请求即可,用户可以根据自身不断变化的位置并结合上一路段查询的结果,自行计算出距离自己最近的 K 近邻兴趣点集合 $P_{u_k}^K$,不必将变化的位置频繁地发送给 LBS 服务器,给 LBS 服务器带来巨大的处理负担,相应的通信开销和计算开销会减少.

最后,对比以往提交一个假位置(锚点)的查询方法,由于锚点选取随机,因此很难以锚点为基础组织兴趣点分布信息.本方法选取的路网顶点是固定锚点,即所有用户在查询过程中使用的全部锚点位置总是不变化,这样的好处就是,下次有新用户以该路网顶点为锚点进行查询时可以利用之前用户的查询结果.且有如下定理成立.

定理 1(路网固定锚点查询优势). 初始状态下,在路网兴趣点分布信息数据库的缓存空间一定时,路网内用户利用固定锚点发起的查询越多,后续用户越能以较高的概率快速获取目标查询结果.

证明:LBS 服务器端数据库缓存容量 c 一定,每次利用随机锚点查询获取的最新结果被存储在该缓存中,假设某个初始状态下缓存中有此类查询 a 个,利用固定锚点查询的结果存储了 b 个,且 $a+b=c$,在某个 LBS 服务器负责的区域内固定锚点的数量是固定的.在每次到达新的查询请求中,利用随机锚点发起查询的概率为 p,固定锚点查询的概率为 q,假设每次查询都能够命中,则有 $p+q=1$.由于缓存容量固定,每增加一个固定锚点查询则随机锚点查询相应减少一个,反之亦然,则固定锚点查询充满数据库缓存的事件可以看作状态空间为$\{0,1,2,\ldots,c\}$

的查询数量质点随机游走问题,即质点从初始状态 a 出发到达状态 0 先于到达状态 c 的概率.假设质点从状态 i 出发到达状态 0 先于到达状态 c 的概率用 x_i 表示,则依据全概公式有:$x_i=x_{i+1}\cdot p+x_{i-1}\cdot q$,其中,$p+q=1$.若有 $u=q/p$,则可得差分方程:

$$d_x=x_i-x_{i+1}=ud_{x-1}=u^id_0 \tag{11}$$

由假设可知 $x_0=1$,$x_c=0$,则有:

$$x_0-x_c=\sum_{x=0}^{c-1}(x_i-x_{i+1})=\sum_{x=0}^{c-1}u^id_0=\frac{1-u^c}{1-u}d_0=1 \tag{12}$$

所以有 $d_0=\frac{1-u}{1-u^c}$,进而有:

$$x_i=x_i-x_c=\sum_{k=i}^{c-1}(x_k-x_{k+1})=\sum_{k=i}^{c-1}u^kd_0=u^i(1+u+...+u^{c-i-1})d_0=\frac{u^i-u^c}{1-u}d_0=\frac{u^i-u^c}{1-r^c} \tag{13}$$

因此,LBS 数据库中充满固定锚点查询的概率 x_a 为

$$x_a=\frac{u^a-u^c}{1-u^c}=\frac{\left(\frac{q}{p}\right)^a-\left(\frac{q}{p}\right)^c}{1-\left(\frac{q}{p}\right)^c} \tag{14}$$

由随机锚点和固定锚点的选取方式可知,由于 LBS 服务器区域内锚点位置和数量一定,路网内全体用户每次利用固定锚点发起查询时,总能以相对较高的概率遇到重复锚点,而随机选取的锚点由于位置坐标选取随机、重复性较小.因此,从用户感觉出发,选取固定锚点发起的每次查询能够在缓存中找到重复结果的可能要远大于选取随机锚点的可能,查找速率更高.因此,用户利用固定锚点提出查询的概率要高于随机锚点查询的概率,即有 $q>>p$.因此在 c 固定的条件下,a 越小,x_a 概率越高.这样,每次查询后就会有更多基于固定锚点的查询被 LBS 服务器暂时缓存下来,这样,缓存容量 c 中就含有更多的固定锚点查询结果.由于区域内固定锚点数量有限,最终后续用户就越能以较高的概率在快速获取目标查询结果. □

由此可见,用户利用固定锚点比利用随机锚点的数据库查询次数少很多.利用随机锚点的每一次查询几乎都是新查询,都要进行一次数据库查询,而利用固定锚点可以促使一些常用锚点查询结果保留在缓存中,便于该锚点结果再次被查询时提高查询效率.假设 LBS 端一次缓存查询耗时为 t,而一次数据库查询为 $nt(n>1)$,一段时间内路网用户发起的查询请求数量为 h,其中缓存命中比率为 r_c,则用户完全利用固定锚点查询消耗的时间为 $r_c\cdot ht+(1-r_c)\cdot hnt$.而在无重复命中的随机锚点查询每次都可以看作是一次数据库查询,其消耗时间为 hnt,则利用固定锚点查询与随机锚点查询消耗的时间比 r 可表示为

$$r=\frac{r_c\cdot ht+(1-r_c)hnt}{hnt}=\frac{r_c+(1-r_c)n}{n}=\frac{r_c(1-n)+n}{n} \tag{15}$$

由此可知,当 $n(n>1)$固定时,缓存命中比率 r_c 越高,固定锚点查询与随机锚点查询消耗时间之比 r 就越低.在区域内固定锚点数量一定且不过多的条件下,固定锚点查询越多,越会提升缓存的命中率,即 LBS 端数据库查询次数少,特别是在兴趣点类型较多导致同态加解密处理时延长情况下,可显著提高查询效率.

(2) 查询范围小

为了保护位置隐私和查询精确性,对比匿名框方法,一方面为了保护隐私需要构建包含用户在内的匿名框,这种方法不仅需要计算匿名框内的兴趣点,还需要计算匿名框外的兴趣点,这是为了在保护用户位置隐私的同时,还要确保查询结果精确所需付出的计算代价.另一方面,为了保护查询内容隐私而进行秘密检索必然会带来计算开销的增加.在 Yi 等人[3]采用的匿名框为基础的秘密检索中,不仅要计算用户当前所在网格的 K 近邻兴趣点结果,还要同时计算邻近网格(甚至匿名区内所有网格)的 K 近邻兴趣点结果.这是因为用户实际位置可能距离邻近网格更近一些,这会带来计算开销的成倍增加,而本方法采用路网顶点的兴趣点组织方法.这种方法可以确保用户每次只需发起一次正方向顶点为锚点的秘密 K 近邻查询,而不必额外计算从其他顶点出发的 K 近邻兴趣点,因此查询范围更减小、更精准,使得计算开销与通信量相应地减小,同时确保较高的隐私保护强度.

综上,本文提出的方法能够为用户提供较强的位置隐私保护和查询内容隐私保护,路网顶点不仅可以作为兴趣点分布信息的组织基础来提高查询效率,还可以作为锚点保护用户位置隐私,便于查询隐私保护的处理.

4.4 服务质量与安全性的均衡

引入锚点会带来查询准确率和效率的提升,但也会带来一定程度的隐私泄漏.这是因为采用泛化位置精度的方法将可以实现位置隐私的保护.路段上的用户采用锚点代替真实位置,实际将自己的位置泛化到了整个路段上,与路段上其他用户共同构成了一个匿名集,类似于构成了一个路段形状的匿名框.显然,用户在越短的路段上,可能包含的用户越少,此时,用户位置泛化范围越小,用户很容易在进入到该路段后发起查询因而被迅速识别锁定.可见,锚点的分布决定了用户隐私与服务质量之间的权衡,即锚点稀疏时,用户真实位置不确定性大,但是基于稀疏锚点的兴趣点数据量大,服务质量不高;反之,锚点稠密,用户真实位置易暴露,但是服务质量高.因此,需要分析锚点分布情况对隐私保护和服务质量两个矛盾因素影响的分析,找到二者的均衡点.

对锚点分布情况的衡量,可以等价量化为全网路段平均长度,即锚点分布稠密或稀疏则路段平均长度小或大.为了达到保护用户隐私的目标,可以在初始构造路网图时,通过剪枝的方法合并一些较短的路段,提高位置的泛化程度.然而剪枝合并路段需要去掉一些路段顶点,用户在经过这些路段顶点时不会发起查询,使得该顶点可到达的目标兴趣点被忽略,查询精确度和效率会下降.因此,查询服务质量和隐私保护需要均衡考虑.我们借鉴并修改"δ_p-δ_q"平衡确定方法[33],使其可用来分析本文中基于路网锚点的隐私保护和服务质量的均衡情况.本文采用泛化度来评价位置保护程度及数据可用性.

定义 10(位置泛化度). 假设某用户 $u_k \in \overrightarrow{v_m v_n}$ 发起以路网顶点 v_n 为锚点的查询请求 Q,此时用户位置被泛化到整个路段 $\overrightarrow{v_m v_n}$ 上,则位置泛化度 $Generalization(\cdot)$定义为

$$Generalization(Q,\overrightarrow{v_m v_n})=\frac{S_{\overrightarrow{v_m v_n}}}{\left(\sum S_{\overrightarrow{v_i v_j}}/A\right)}=S_{\overrightarrow{v_m v_n}}/S \tag{16}$$

其中, $S_{\overrightarrow{v_m v_n}}$ 表示路段长度, $\sum S_{\overrightarrow{v_i v_j}}/A$ 用来计算全路网路段平均长度 S.在用户进入下一路段前的查询有效期 $T_{\exp}$ 内,位置信息的总泛化度为 $\int_{T_s}^{T_{\exp}} Generalization(Q,\overrightarrow{v_m v_n})\mathrm{d}t$, T_s 为用户进入路段 $\overrightarrow{v_m v_n}$ 的时间.为了保证用户不在较短的路段发起查询,确保用户位置隐私不被迅速锁定泄漏,定义 ε_p-隐私确保模型如下.

定义 11(ε_p-隐私确保模型). 设 $Set=\{\overrightarrow{v_n v_x} \mid x \in N^*\}$ 表示以 v_n 为起始点的路段集合,ε_p 是用户指定的路网中路段最小粒度长,则有:

$$\forall time \in [T_s,T_{\exp}],\ \min(S_{\overrightarrow{v_n v_x} \in Set}) \geqslant \varepsilon_p \times S \tag{17}$$

成立,则称该路段满足 ε_p-隐私确保模型.

路段长度不仅要满足最低的隐私确保需求,位置泛化度不能过高,即路段不能过分剪枝合并,且在一些较长的路段上(如高架路)上需要添加路口信息,确保兴趣点丢失率不过高,由此定义 δ_q-质量确保模型如下.

定义 12(δ_q-质量确保模型). 假设用户能够忍受的最差查询质量为 δ_q,对于

$$\forall time \in [T_s,T_{\exp}],\ \forall u_k \in \overrightarrow{v_m v_n},\ Generalization(Q_{uk},\overrightarrow{v_m v_n}) \geqslant \delta_q \tag{18}$$

成立,则称该路段满足 δ_q-质量确保模型.

可见,在用户分布相对均匀的条件下,用户所在路段长度在同时满足 ε_p-隐私确保模型和 δ_q-质量确保模型时,用户即可用该路段正方向顶点发起查询,既能实现较好的查询质量,又能保证查询隐私安全,实现位置隐私保护和服务质量间的平衡.

5 实 验

基于上一节性能分析的结果,本节主要讨论在用户数量、兴趣点查询数量、位置变换频率等因素影响下,用户基于本方法发起 LBS 查询时的准确性、安全性及查询效率等指标的变化情况.实验在模拟数据集上进行,通过影响因素变化和同类方法比较的方式,对本方法的性能进行了比对验证.

5.1 实验环境配置

本实验提出的所有算法均采用 JAVA 语言在 Windonws 7 操作系统中移动终端模拟器上实现,运行硬件环境为 3.2GHz Intel Core i5 四核处理器,内存大小为 8GB.路网地图采用美国国家地理勘探局(USGS)[34]提供的佐治亚州亚特兰大市路网图,其中,被选作固定锚点的路网顶点约为顶点总数的 70%.由于该路网数据具有简洁、易处理的特点,本文默认采用该地理数据进行绝大多数实验.同时,为了进一步对比说明本方法在实际路网的效率,本文增加两个同样较为常用的真实路网数据集 California Map(CA)和 New York Map(NY)[35]来对比测试本方法.该实验数据集仅在后文测试平均处理时间(average processing time,简称 APT)的图 15 和图 16 中使用.3 种类型路网数据集见表 3.

Table 3 Datasets of road networks

表 3 路网数据集

数据集	边数量	顶点数量	POIs 数量
USGS(默认)	9 286	6 927	30 000
CA	47 185	20 997	84 000
NY	56 263	14 890	60 000

另外,本文采用 Thomas Brinkhoff 路网数据生成器[36]生成用户位置及连续查询中的轨迹数据,用户随机均匀分布在每条路网上,起点及终点随机选择.实验的网络通信带宽为 3Mbps,路网兴趣点分布索引大小为 25.6KB.参数配置情况见表 4.

Table 4 Parameters configuration

表 4 参数配置

参数名	取值范围	默认值
全局移动用户总数 U	$50000 \leqslant U \leqslant 300000$	100 000
用户查询兴趣点数 K	$5 \leqslant K \leqslant 60$	20
用户位置更新频率 f	$5 \leqslant f \leqslant 30$	15s
缓存中兴趣点保留时间 T_k	$1 \leqslant T_k \leqslant 12$	6h
固定锚点查询比率 r_{anchor}	$20\% \leqslant r_{anchor} \leqslant 100\%$	100%
平均路段长度 S	$200 \leqslant S \leqslant 2000$	1000m

5.2 准确率及带宽消耗

如图 4 描述的查询处理过程,查询准确率是指发起总数为 n 次的查询请求所获得的结果中,准确的查询结果数量与查询总数的比率,如图 6、图 7 所示.

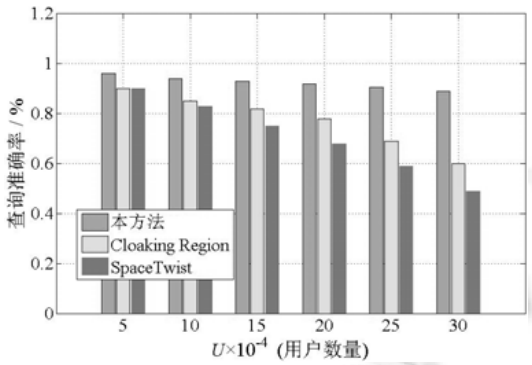


Fig.6 Query accuracy (U varies)

图 6 查询准确率(U 变化)

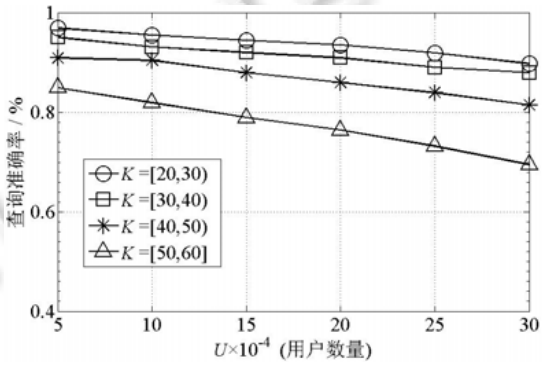


Fig.7 Query accuracy (U, K vary)

图 7 查询准确率(U,K 变化)

在图 6 中,当用户数量变化时,由于处理负担增加,3 种方法的准确率均表现出了下降的趋势,本方法的查询准确率总体保持不低于 90%.这是由于本方法考虑了路网兴趣点的分布特点建立了兴趣点索引,因此可以较好

地保证查询准确率.而后文图 9 显示,本方法能够保持一个较为合理的带宽消耗.

在 Cloaking Region[3]方法中,由于查询在欧氏空间进行,在障碍物阻挡的实际情况下,导致查询准确率相对较低.该方法采用匿名框的查询方法,使得查询处理过程中存在较近兴趣点没有被查出的情况,因此查询准确率在用户数量增大的过程中下降明显.SpaceTwist[13]方法在查询过程中存在查询不均衡问题[22],因此查询准确率较低.该方法不仅存在上述欧氏空间查询导致的不准确问题,而且由于锚点位置是随机选取的,因此在用户数量增长的过程中,LBS 服务器面临查询请求不断增大的处理瓶颈问题,导致准确率显著下降.

当 K 值变化时,也会导致查询准确率变化.如图 7 所示,本文提出的方法在 K 增大时,由于 LBS 服务器端需要计算更多的兴趣点结果,导致服务延迟,因此查询准确率会相应地降低.

同时,在路网内兴趣点分布均匀的情况下,平均路段长度 S 的变化也会影响查询质量.当相邻路网顶点距离较小时,我们会采取剪枝等方法去掉一些出边或入边较少的顶点,使全网平均路段长度变长,以避免过于发起查询带来位置隐私遭到推断锁定的问题,但这会带来查询准确率的下降.如图 8 所示,当平均路段长度增加时,由于一些顶点被裁剪掉,因此由该顶点直接连通的兴趣点也会相应地被丢弃,造成查询过程中兴趣点丢失率的上升,进而带来准确率下降的问题.对于路段过短带来的不安全因素,我们将在下一节中讨论.

另一方面,用户发起 K 近邻兴趣点查询时,需要请求某一锚点上全部 m 个类别的 POI 数据(这里假定每个锚点均有 m 类 POI 数据),因此理论上用户需要从一个锚点处获得 $K\times m$ 个 POI 数据.如果是首次查询用户,需要从路段两端锚点 v_m,v_n 获取总数为 $2\times K\times m$ 个 POI 数据,因此会带来较大的带宽资源消耗.一般地,如果用 K_0 来表示某个初始状态下查询兴趣点个数,R_{BW}表示带宽利用率,则用户查询兴趣点数量 K 与带宽利用率 R_{BW}之间的关系如公式(19)所示,二者的变化关系如图 9 中三角形标识 A 线所示.可见,带宽利用率理论上会随用户兴趣点查询数量的增长而快速增长.

$$K=\frac{K_0}{1-R_{BW}} \tag{19}$$

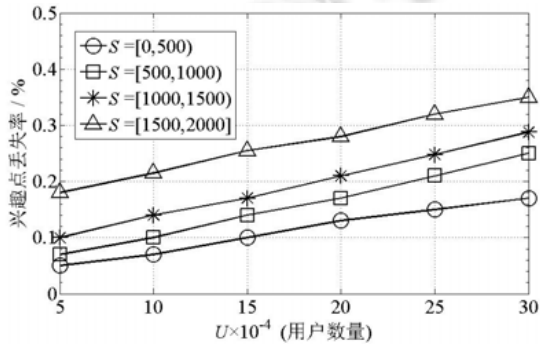


Fig.8 Loss rate of POIs (U, S vary)

图 8 兴趣点丢失率(U,S 变化)

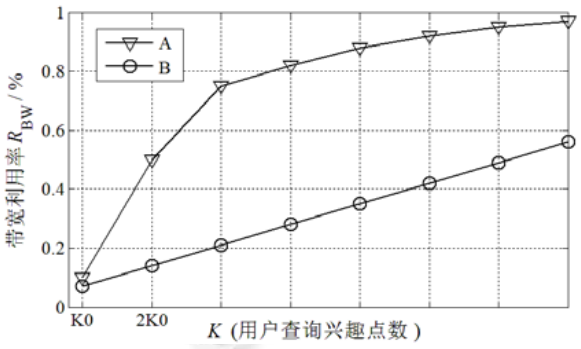


Fig.9 Use rate of bandwidth (K varies)

图 9 带宽利用率(K 变化)

然而,本文不采用将全部 m 个兴趣点查询结果一起发送回来的方式来达到隐藏真实查询类型的目标,而是通过算法 2 第 9 行、第 10 行的描述将该顶点下各类型兴趣点进行累乘计算,得出 1 个乘积结果后返回给用户.因此,在用户查询 K 值不断增大的过程中,带宽消耗呈现线性增长,如图 9 中圆形标识 B 线所示.而 K 增大却会给服务端带来计算开销的迅速增大,不会带来公式(19)中难以承受的带宽迅速增加,而通过提升固定锚点使用率,增加服务器端兴趣点缓存时长,可以帮助降低服务端计算开销,从而提升整体查询效率.我们将在第 5.4 节中讨论该问题.

5.3 安全性

假设通常情况下,当每个路网顶点 v_n 的平均入边数量为 4,且一次连续查询经过的路网顶点数不超过 5 个时,以某顶点 v_n 发起的查询请求的用户位于该顶点的任意入边上的概率均相等,因此,由其查询请求 Q 中的元素

通过追踪关联,推断出用户所在具体入边的自信息量为$-\log_2(1/4)$=2bit,由第 4.2 节不可追踪性分析可知,在连续查询中,每次锚点查询相互独立,一次连续查询结束后自信息量至少为 5×2bit=10bit.以该信息量作为隐私泄露界限,如果用户在一次连续查询过程中的自信息量低于 10bit,则可看作存在隐私泄漏可能.本文比对包括本文提出方法在内的 3 种方法中大于 10bit 的连续查询请求占全部查询请求的比率(*QRIE*>10bit).如图 10 所示.

如第 4.4 节分析所述,通过剪枝的方法提高平均路段长度可以带来查询效率的降低,但是随着平均路段长度 *S* 的增加,用户的匿名集变大,泛化度增高.如图 11 所示,随着用 *U* 和 *S* 的增长,存在隐私泄漏风险的查询比率逐步降低,*U* 的增长起到了增大匿名集的作用,*S* 的增长起到了扩大泛化用户位置的作用,二者共同作用提高了隐私保护度.可见,当平均路段长度较小,即路段中存在较多短路段时,很多用户在进入较短路段后必须立即用前方顶点发起查询,因此很容易遭受查询时序推断攻击,极端情况下,某个时间间隔内只有 1 个用户进入该路段,就被迅速锁定,造成位置隐私泄漏.但本文采用了同态加密处理查询内容,不会造成查询内容等隐私的泄漏.

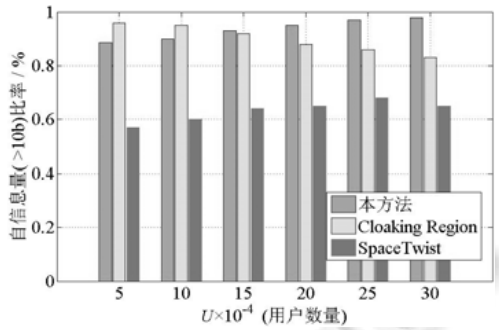


Fig.10 Query rate satisfies *QRIE*>10bit (*U* varies)

图 10 满足 *QRIE*>10bit 查询比率(*U* 变化)

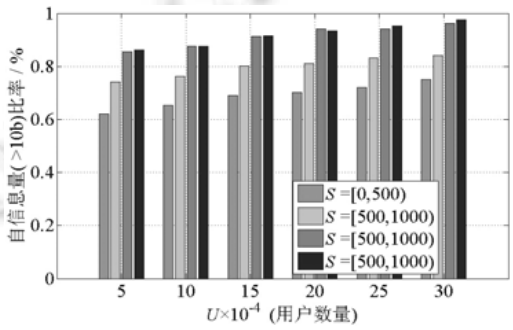


Fig.11 Query rate satisfies *QRIE*>10bit (*U, S* vary)

图 11 满足 *QRIE*>10bit 查询比率(*U,S* 变化)

本文提出的方法在初始阶段 *QRIE*>10bit 的比率略低于 Cloaking Region,这是由于该方法不仅对查询内容进行了秘密检索处理,还对位置数据进行了相同处理,因此用户被追踪关联的概率较低,但相应的处理时延也会增加(详见后文第 5.4 节实验).然而本方法随着用户数量的增加,比率逐渐升高.这是因为随着更多用户的加入,本方法锚点的共用程度更高,匿名性更好,因此被追踪的可能性降低,使得安全查询的比率升高.而 SpaceTwist 方法由于仅对用户的位置进行保护,存在查询内容关联推断风险,安全的查询比例相对较低.

5.4 查询效率

本文主要从平均处理时间和平均数据包量两个指标来评测本方法的查询效率.通过第 4.3 节性能分析可知,影响两个指标的因素较多.如图 12~图 19 所示,本节主要围绕这两个指标进行实验分析.

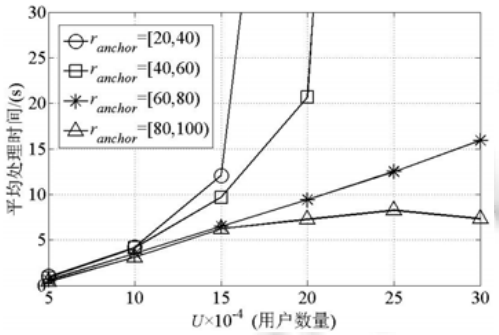


Fig.12 APT (*U*, r_{anchor} vary)

图 12 平均处理时间(*U*,r_{anchor} 变化)

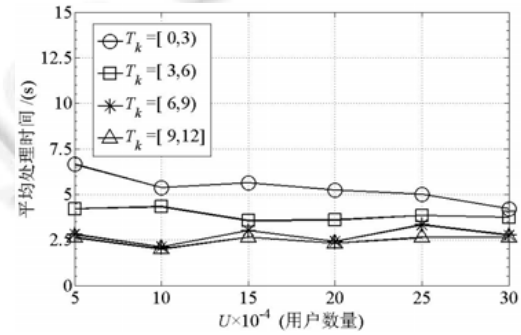


Fig.13 APT (*U*, T_k vary)

图 13 平均处理时间(*U*,T_k 变化)

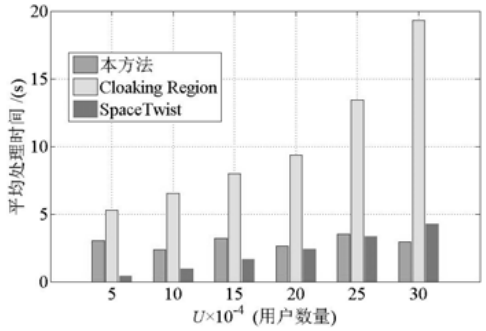


Fig.14 Comparison of APT (U varies)

图 14 平均处理时间比较(U 变化)

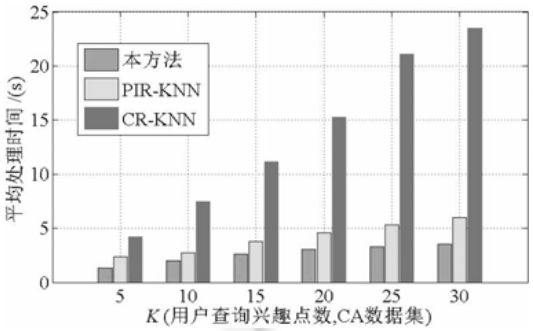


Fig.15 Comparison of APT (K varies)

图 15 平均处理时间比较(K 变化)

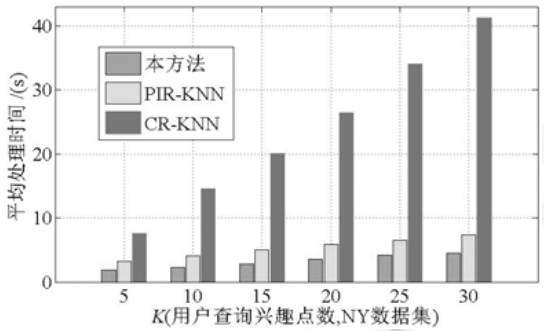


Fig.16 Comparison of APT (K varies)

图 16 平均处理时间比较(K 变化)

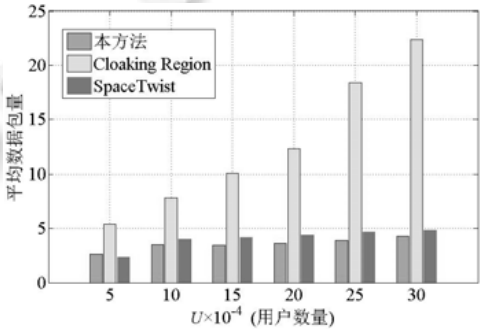


Fig.17 Comparison of AP (U varies)

图 17 平均数据包量比较(U 变化)

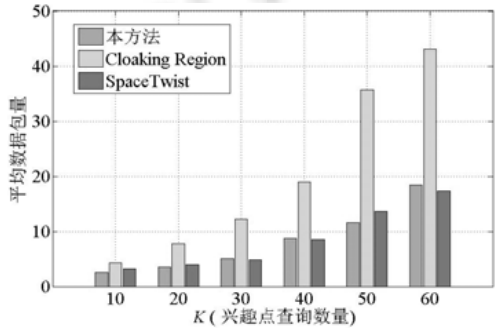


Fig.18 Comparison of average packets (K varies)

图 18 平均数据包量比较(K 变化)

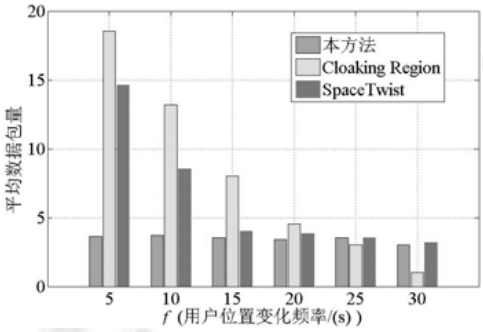


Fig.19 Comparison of average packets (f varies)

图 19 平均数据包量比较(f 变化)

LBS 服务端的处理速度直接影响用户的服务质量。在实际的查询中,LBS 服务器需要面对大量用户的服务请求,当用户总量增加时,每个 LBS 服务器管辖区域内查询用户数量增加,这使得平均处理时间增加.如第 4.3 节分析所述，当有更多用户使用固定锚点查询时,查询效率显著提高.在图 12 中,当固定锚点查询占比 r_{anchor} 不断增加时,LBS 服务器的处理时间明显表现出降低趋势.当占比范围在 80%~100%时,随着用户数量的增加,甚至出现了处理时间下降的情况.这是由于当固定锚点数量不多时,用户每次查询都能以较高的概率在 LBS 数据库缓存找到查询结果,而不必耗时地进行数据查询,使得 LBS 服务器的处理时间显著降低.

兴趣点查询结果在LBS服务端缓存中保留时间 T_k 的长短也是影响平均处理时间的因素之一.如图13所示,当保留时间增长时,可以有更多的兴趣点查询结果被保留在缓存中,以此提高新查询在缓存中的命中率.当命中

率升高时,会减少数据库查询次数,因此有如图 14 所示的平均处理时间的减少.另外,我们还分别选取了两个路网环境下保护用户隐私的查询方法 PRI-KNN[26]和 CR-KNN[27]来对比分析本方法在处理时间方法的优势.类似地,本方法也分别在上述两种方法采用的两类真实路网数据集 CA 和 NY 上进行.如图 15 所示,由于 PIR-KNN 方法的查询方式决定用户在查询前首先要确定其所在的 Voronoi 单元格,而本方法采用的路网和兴趣点组织结构无需此过程,且查询过程中无需借助协处理器,因此,虽然同样在路网环境下都采用了 PIR 的查询处理方法,但本方法所需处理时间稍短.当 K 增加时,由于 CR-KNN 方法需要在更广范围内持续搜寻大量兴趣点信息,因此其平均处理时间明显高于两外两种方法.同样,如图 16 所示,在 NY 数据集上,由于路段和顶点数量分布更为密集,因此处理时间相对于 CA 数据集更长,但总体趋势与图 15 类似,本方法依然能够保持较好的处理效率.

同时,与其他经典方法相对比,如图 17、图 18 所示.在图 17 中,由于 Cloaking Region 需要搜索邻近多个网格内的兴趣点信息,并且该方法采用了较为复杂的秘密信息检索方法,对位置和查询内容均做处理,因此平均处理时间随用户数量增加而显著增长.本方法虽然也采用了秘密信息检索处理方式,但由于兴趣点的组织形式影响,使得本方法在查询兴趣点时无需额外搜索任何邻近区域的兴趣点,极大地减小了 LBS 服务端的处理负担,因此本方法的处理时间增长并不明显;并且由于使用固定锚点查询提高了 LBS 数据库缓存命中率,使得平均处理时间短暂出现了下降.SpaceTwist 方法由于不采用秘密检索方法,只利用一个随机锚点发起查询,由于随机锚点重复查询率较低,因此处理时间随用户数量的增加而增加,但相对于秘密检索方法的处理时间要少很多.因此,本方法能够以接近随机锚点的处理时间实现兴趣点信息的秘密检索.

对于平均数据包量,如图 17 所示,本方法的平均数据包量随用户数量的增加保持相对稳定;而 Cloaking Region 方法由于需要检索邻近网格内的兴趣点来确保精确的查询结果,并将大量候选结果全部发送给用户,因此会带来通信数据包量的显著增加.SpaceTwist 同样是利用锚点查询,虽然会返回一定量的候选结果,但在查询过程中由于不采用私有信息检索方式,因此隐私平均数据包量略低于本文方法.该方法在 LBS 服务器处理不出现明显瓶颈的条件下,随用户数量的增加,平均数据包数量表现出缓慢增长.

当兴趣点查询数量 K 增加时,3 种方法的平均数据包量均随之增加,如图 18 所示.由于本文方法采用了基于路网顶点的兴趣点分布信息组织方法,使得在 K 增加时,数据包量的增加并不明显;然而,Cloaking Region 方法在 K 增大时却需要搜索更多的邻近区域,并将该候选结果发送给用户,因此平均数据包量增加明显;SpaceTwist 方法的平均数据包量与本方法接近.

在连续查询中,用户位置更新频率 f 也是影响数据包量的重要因素之一,如图 19 所示.本方法受位置变换频率的影响较小,平均数据包量变化不大.这是因为本文采用锚点查询,仅需根据当前路段正方向顶点发起一次查询所获得的结果即可确保用户在该路段时无需再次发起查询,因此极大地降低了连续查询中位置更新时的查询次数.而 Cloaking Region 方法由于需要查询邻近网格的兴趣点情况,因此在位置更新频率从 5s/次升高到 30s/次时,相应的数据包量会逐渐减少.这是由于相应的查询次数降低的原因.SpaceTwist 方法数据包量的消涨情况与 Cloaking Region 方法接近.

6 结论及展望

本文提出了一种路网环境下能够同时保护用户位置隐私和查询内容隐私的 K 近邻兴趣点查询方法.该方法以路网顶点为基础组织路网兴趣点分布信息.这种组织方法便于用户以路网顶点为锚点发起兴趣点查询保护用户位置隐私;同时,该组织方法便于用户秘密检索目标兴趣点信息.进而,本文基于 Palillier 密码系统给出了秘密检索的过程.最后,我们对所提出方法的准确性、安全性及查询效率进行了分析,并对其进行了有针对性的实验.性能分析和实验结果表明,本方法具有良好的安全性能和查询效率.然而,本方法也存在一些不足,如对用户速度、兴趣点查询数量、路段剪枝方法、交通拥塞情况等其他时空影响因素考虑不够全面.我们将在今后的工作中针对此类问题继续展开研究.

References:

[1] Zhang XJ, Gui XL, Wu ZD. Privacy preservation for location-based services: A survey. Ruan Jian Xue Bao/Journal of Software, 2015,26(9):2373−2395 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4857. htm [doi: 10.13328/j.cnki.jos.004857]

[2] Zhou AY, Yang B, Jin CQ, Ma Q. Location-based services: Architecture and progress. Chinese Journal of Computers, 2011,34(7): 1155−1171 (in Chinese with English abstract).

[3] Yi X, Paulet R, Bertino E, *et al*. Practical approximate *k* nearest neighbor queries with location and query privacy. IEEE Trans. on Knowledge and Data Engineering, 2016,28(6):1546−1559.

[4] Ni W, Gu M, Chen X. Location privacy-preserving *k* nearest neighbor query under user's preference. Knowledge-Based Systems, 2016,103:19−27.

[5] Kim HI, Kim HJ, Chang JW. A kNN query processing algorithm using a tree index structure on the encrypted database. In: Proc. of the 2016 Int'l Conf. on Big Data and Smart Computing (BigComp). IEEE, 2016. 93−100.

[6] Wernke M, Skvortsov P, Dürr F, *et al*. A classification of location privacy attacks and approaches. Personal and Ubiquitous Computing, 2014,18(1):163−175.

[7] Wang YH, Zhang HL, Yu XZ. Research on location privacy in mobile Internet. Journal on Communications, 2015,36(9):230−243 (in Chinese with English abstract).

[8] Ghinita G. Privacy for location-based services. Synthesis Lectures on Information Security, Privacy, & Trust, 2013,4(1):1−85.

[9] Pan X, Xu J, Meng X. Protecting location privacy against location-dependent attacks in mobile services. IEEE Trans. on Knowledge and Data Engineering, 2012,24(8):1506−1519.

[10] Gao S, Ma JF, Yao QS, Sun C. Towards cooperation location privacy-preserving group nearest neighbor queries in LBS. Journal on Communications, 2015,36(3):146−154 (in Chinese with English abstract).

[11] Niu B, Zhang Z, Li X, *et al*. Privacy-area aware dummy generation algorithms for location-based services. In: Proc. of the 2014 IEEE Int'l Conf. on Communications (ICC). IEEE, 2014. 957−962.

[12] Niu B, Zhu X, Chi H, *et al*. 3PLUS: Privacy-preserving pseudo-location updating system in location-based services. In: Proc. of the 2013 IEEE Wireless Communications and Networking Conf. (WCNC). IEEE, 2013. 4564−4569.

[13] Yiu ML, Jensen CS, Huang X, *et al*. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In: Proc. of the IEEE 2008 24th Int'l Conf. on Data Engineering (ICDE 2008). IEEE, 2008. 366−375.

[14] Yiu ML, Jensen CS, Møller J, *et al*. Design and analysis of a ranking approach to private location-based services. ACM Trans. on Database Systems (TODS), 2011,36(2):Article No.10.

[15] Gong Z, Sun GZ, Xie X. Protecting privacy in location-based services using *k*-anonymity without cloaked region. In: Proc. of the 2010 11th Int'l Conf. on Mobile Data Management (MDM). IEEE, 2010. 366−371.

[16] Huang Y, Huo Z, Meng XF. CoPrivacy: A collaborative location privacy-preserving method without cloaking region. Chinese Journal of Computers, 2011,34(10):1976−1985 (in Chinese with English abstract).

[17] Papadopoulos S, Bakiras S, Papadias D. Nearest neighbor search with strong location privacy. Proc. of the VLDB Endowment, 2010,3(1-2):619−629.

[18] Mouratidis K, Yiu ML. Shortest path computation with no information leakage. Proc. of the VLDB Endowment, 2012,5(8): 692−703.

[19] Khoshgozaran A, Shahabi C, Shirani-Mehr H. Location privacy: Going beyond *K*-anonymity, cloaking and anonymizers. Knowledge and Information Systems, 2011,26(3):435−465.

[20] Sakai K, Sun MT, Ku WS, *et al*. An analysis of onion-based anonymous routing for delay tolerant networks. In: Proc. of the 36th IEEE Int'l Conf. on Distributed Computing Systems (ICDCS 2016). 2016. 609−618.

[21] Beresford AR, Stajano F. Mix zones: User privacy in location-aware services. In: Proc. of the PerCom Workshops. 2004. 127−131.

[22] Zhou CL, Ma CG, Yang ST. Research of LBS privacy preserving based on sensitive location diversity. Journal on Communications, 2015,36(4):125−136 (in Chinese with English abstract).

[23] Ma CG, Zhou CL, Yang ST. Location privacy-preserving method in LBS based on Voronoi division. Journal on Communications, 2015,36(5):1−12 (in Chinese with English abstract).

[24] Zhou CL, Ma CG, Yang ST. Location privacy-preserving method for LBS Continuous *K*NN query in road networks. Journal of Computer Research and Development, 2015,52(11):2628−2644 (in Chinese with English abstract).

[25] Zheng B, Zheng K, Xiao X, *et al*. Keyword-aware continuous *k*NN query on road networks. In: Proc. of the 2016 IEEE 32nd Int'l Conf. on Data Engineering (ICDE). IEEE, 2016. 871−882.

[26] Wang L, Ma R, Meng X. Evaluating *k* nearest neighbor query on road networks with no information leakage. In: Proc. of the Int'l Conf. on Web Information Systems Engineering. Springer Int'l Publishing, 2015. 508−521.

[27] Um JH, Kim YK, Lee HJ, *et al*. *k*-nearest neighbor query processing algorithm for cloaking regions towards user privacy protection in location-based services. Journal of Systems Architecture, 2012,58(9):354−371.

[28] Yang ST, Ma CG, Zhou CL. LBS-oriented location privacy protection model and scheme. Journal on Communications, 2014,35(8): 116−124 (in Chinese with English abstract).

[29] Zhou CL, Tian H, Ma CG, *et al*. Research on LBS privacy preservation based on pseudorandom permutation in road network. Journal on Communications, 2017,38(6):19−29 (in Chinese with English abstract).

[30] Paillier P. Public-key cryptosystems based on composite degree residuosity classes. In: Proc. of the Int'l Conf. on the Theory and Applications of Cryptographic Techniques. Berlin, Heidelberg: Springer-Verlag, 1999. 223−238.

[31] Feng J, Zhang LX, Lu JM, Wang C. Review on moving objects query techniques in road network environment. Ruan Jian Xue Bao/ Journal of Software, 2017,28(6):1606−1628 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5254.htm [doi: 10.13328/j.cnki.jos.005254]

[32] Goldwasser S, Micali S. Probabilistic encryption. Journal of Computer and System Sciences, 1984,28(2):270−299.

[33] Pan X, Meng X, Xu J. Distortion-based anonymity for continuous queries in location-based mobile services. In: Proc. of the 17th ACM SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems. ACM Press, 2009. 256−265.

[34] USGS. Maps, data publications. 2017. http://www.usgs.gov

[35] Openstreetmap. Maps, data publications. 2017. http://www.openstreetmap.org

[36] Brinkhoff T. Network-based generator of moving objects. 2005. http://iapg.jade-hs.de/personen/brinkhoff

附中文参考文献:

[1] 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. http://www.jos.org.cn/1000-9825/4857.htm [doi: 10.13328/j.cnki.jos.004857]

[2] 周傲英,杨彬,金澈清,马强.基于位置的服务:架构与进展.计算机学报,2011,34(7):1155−1171.

[7] 王宇航,张宏莉,余翔湛.移动互联网中的位置隐私保护研究.通信学报,2015,36(9):230−243.

[10] 高胜,马建峰,姚青松,孙聪.LBS 中面向协同位置隐私保护的群组最近邻查询.通信学报,2015,36(3):146−154.

[16] 黄毅,霍峥,孟小峰.CoPrivacy:一种用户协作无匿名区域的位置隐私保护方法.计算机学报,2011,34(10):1976−1985.

[22] 周长利,马春光,杨松涛.基于敏感位置多样性的 LBS 位置隐私保护方法研究.通信学报,2015,36(4):125−136.

[23] 马春光,周长利,杨松涛,赵蕴龙.基于 Voronoi 图预划分的 LBS 位置隐私保护方法.通信学报,2015,36(5):1−12.

[24] 周长利,马春光,杨松涛.路网环境下保护 LBS 位置隐私的连续 *K*NN 查询方法.计算机研究与发展,2015,52(11):2628−2644.

[28] 杨松涛,马春光,周长利.面向 LBS 的隐私保护模型及方案.通信学报,2014,35(8):116−124.

[29] 周长利,田晖,马春光,等.路网环境下基于伪随机置换的 LBS 隐私保护方法研究.通信学报,2017,38(6):19−29.

[31] 冯钧,张立霞,陆佳民,王冲.路网环境下的移动对象查询技术研究综述.软件学报,2017,28(6):1606−1628. http://www.jos.org.cn/1000-9825/5254.htm [doi: 10.13328/j.cnki.jos.005254]

周长利(1985－),男,黑龙江哈尔滨人,博士,讲师,CCF 专业会员,主要研究领域为位置隐私保护,网络与信息安全.

田晖(1982－),男,博士,教授,CCF 高级会员,主要研究领域为网络和信息安全,云计算安全,多媒体内容安全和数字隐写.

陈永红(1974－),男,博士,教授,主要研究领域为入侵检测系统,数字水印.

蔡绍滨(1973－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为水下无线传感器网络,网络与信息安全.