计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2017.20160551
54(9):2092-2101,2017

道路网络上基于时空相似性的连续查询隐私保护算法

潘 晓[1,2] 谌伟璋[1] 孙一格[1] 吴 雷[1]
[1](石家庄铁道大学经济管理学院 石家庄 050043)
[2](河北省高校人文社会科学重点研究基地(石家庄铁道大学) 石家庄 050043)
(smallpx@stdu.edu.cn)

Continuous Queries Privacy Protection Algorithm Based on Spatial-Temporal Similarity Over Road Networks

Pan Xiao[1,2], Chen Weizhang[1], Sun Yige[1], and Wu Lei[1]
[1](*School of Economic & Management, Shijiazhuang Tiedao University, Shijiazhuang* 050043)
[2](*Key Research Base for Humanities and Social Sciences in Hebei Province (Shijiazhuang Tiedao University), Shijiazhuang* 050043)

Abstract Continuous queries are one of the most common queries in location-based services (LBSs), although particularly useful, such queries raise serious privacy concerns. However, most of the existing location cloaking approaches over road networks are only applicable for snapshots queries. If these algorithms are applied on continuous queries directly, due to continuous location frequently updated, continuous query privacy will be disclosed. Moreover, combined with the network topology and other network parameters (limited speed etc.), the attackers are knowledgeable, which can easily lead to precise location privacy disclosure. We observe that mobile objects have similar spatial and temporal features due to the existing of network topology. In order to resist continuous query attacks and location-dependent attacks simultaneously, we propose a continuous queries privacy protection algorithm based on spatial-temporal similarity over road networks. The algorithm adopts user grouping and K-sharing privacy requirement strategies to constitute cloaking user sets, which is used to resist continuous queries attack. Then, with the same premise of cloaking user sets, a continuous cloaking segment sets generating algorithm is proposed to resist location-dependent attacks, which makes a balance between location privacy and service quality. Finally, we conduct series of experiments to verify our algorithm with four evaluation measures, and the experimental results show the effectiveness of the proposed algorithm.

Key words location privacy; continuous query; road networks; location based services (LBSs); mobile computing

摘 要 连续查询作为基于位置服务中常见的服务类型之一,为人们的生活和工作带来了巨大的便利.最近几年,针对位置服务中的隐私保护引起了学术界研究者的广泛关注.然而,现有在道路网络上的位

收稿日期:2016-08-02;**修回日期**:2016-10-10
基金项目:国家自然科学基金项目(61303017,61502146);河北省自然科学基金项目(F2014210068);河北省教育厅青年基金项目(QN2016083);河北省高等学校人文社会科学研究项目(GH161079);石家庄铁道大学第四届优秀青年科学基金项目(Z661250444);河北省研究生创新资助项目(Z99910);国家级大学生创新创业训练计划项目(201510107013,201610107003)
This work was supported by the National Natural Science Foundation of China (61303017, 61502146), the Natural Science Foundation of Hebei Province (F2014210068), the Young Scholars Project of the Hebei Provincial Education Department (QN2016083), the Project of Humanities and Social Sciences for the Colleges in Hebei Province (GH161079), the Fourth Outstanding Youth Foundation of Shijiazhuang Tiedao University (Z661250444), the Graduate Innovative Foundation of Hebei Province(Z99910), and the College Innovative Training Program Foundation of China(201510107013, 201610107003).

置隐私保护工作大多针对快照查询提供隐私保护.如果直接将这些算法应用于连续查询,由于连续查询中位置频繁更新,将同时产生连续查询隐私泄露和精确位置的泄露.由于网络拓扑的存在,移动用户的运动在一段时间内具有时空相似的特点.利用连续查询用户的时空相似性,提出了一种在道路网络上基于时空相似性的连续查询隐私保护算法.通过采取分组策略构造匿名集和 K-共享机制,提出了一种启发式宽度优先用户搜索算法 HBFS 来构造匿名用户集,并提出了一种连续时刻内匿名路段集生成算法 CSGA 生成匿名路段集合,可以同时防止连续查询攻击和位置依赖攻击.最后,采用 4 个评价标准对算法进行了一系列实验,验证了算法的有效性.

关键词　位置隐私;连续查询;道路网络;基于位置服务;移动计算

中图法分类号　TP391

随着移动技术的不断发展,基于位置的服务(location based services, LBSs)已广泛应用于地理导航、社交娱乐、紧急救援、广告推广等应用,其中连续查询是 LBS 中非常常见的一种服务形式.用户在获得 LBS 位置服务带来的便利的同时,个人数据也在遭受着隐私泄露的风险[1-2].具体来讲,位置查询中包含的位置信息将帮助攻击者推理用户的工作生活方式、身体健康状况等.

现有道路网络上的位置隐私保护算法多针对快照查询提供隐私保护[3-4].其基本思想是基于位置 K 匿名和路段 L 多样性模型,将 K 个用户的确切位置隐藏于 L 个不同的路段中.连续查询作为 LBS 常见服务之一,用户需要在查询有效期内不断更新位置.如果直接将适用于快照查询的隐私保护算法应用于连续查询,将产生连续查询攻击[5].同时,攻击者结合网络拓扑、用户在路网上的最大运动速度限制以及不同时刻的匿名位置,可以推测用户的确切位置,造成位置依赖攻击[6].然而,现有在道路网络上的位置隐私保护工作无法同时防止连续查询攻击和位置依赖攻击.

文献[7]针对路网中的连续查询攻击问题,提出了一种基于 Snet 层级结构的隐私保护算法.Snet 层级结构基于用户转移概率,预先将路网划分成不同等级的路段子图.利用 Snet 层级结构可为查询用户构造具有相似运动的匿名集,从而防止连续查询攻击.然而,该方法中所使用的转移概率是基于历史数据来预测移动用户在路网中的运动情况,较难准确地构造出能够保持长期具有相似运动的匿名用户集.所以该方法难以保持较高的匿名成功率且容易出现查询代价过大的现象.

文献[8]针对路网中位置依赖攻击问题,提出了一种启发式的匿名算法.该算法以移动用户所在路网上的路段集合作为匿名区域,保证了任意连续时刻的匿名路段集合之间的网络距离不超过最大速度限制要求从而防止了位置依赖攻击.虽然该方法仅防止了路网中的位置依赖攻击,但对连续查询攻击无效.文献[8]将连续查询位置隐私概念延伸到用户的路径隐私上,针对用户的路径隐私问题提出了 M-cut 隐私需求.M-cut 隐私需求其实是一种基于 K 匿名的思想,其目的是要让用户在连续时刻内的匿名路段集能够形成 K 条完整的路径,从而不仅能够保护用户的位置隐私而且能够保护路径隐私.但是因为其匿名集不满足文献[5]提出的要求,所以该方法对防止连续查询攻击也无效.

为了同时防止连续查询攻击和位置依赖攻击,本文提出了一种道路网路上基于时空相似性的连续查询位置隐私保护算法.综合考虑用户的空间相似性和时间相似性,基于查询用户的时空相似性采取分组策略构造匿名集,实现 K-共享机制.提出了一种启发式的宽度优先用户搜索算法以防止连续查询攻击;同时,在匿名用户集不变的前提下,提出了一种连续时刻内匿名路段集生成算法以防止位置依赖攻击,在位置隐私保护与服务质量间寻找到了一种均衡.

总结本文贡献有 3 方面:

1) 提出了一种在道路网路上基于时空相似的连续查询隐私保护算法 SCPA,有效地保护移动用户连续查询隐私;

2) 针对连续查询攻击和位置依赖攻击,分别提出了启发式宽度优先用户搜索算法 HBFS 和连续时刻内匿名路段集生成算法 CSGA;

3) 进行了一系列实验,验证了算法的有效性.

1　相关工作

文献[5]针对连续查询隐私问题研究提出了一种基于组的方法,该方法是一种根据用户当前位置,找到和其邻近的用户构成匿名集,匿名集在查询有效期内保持不变.该方法因未考虑用户位置在未来

的邻近性,容易导致匿名区域过大,影响服务质量.为了解决这一问题,文献[9]提出了基于扭曲度的方法,给出了 δ_p-隐私模型和 δ_q-扭曲度模型来平衡用户隐私和服务质量,该工作不仅考虑到了移动用户连续查询隐私需求,而且还考虑了移动用户的运动速度的影响,从而最小化信息扭曲度,保证了服务质量.文献[10]将文献[5]对匿名集的要求放宽,提出了查询 m-不变性(m-invariance)模型,要求在用户查询有效期内,所有匿名集合的敏感属性交集最少有 m 个敏感属性保持不变.另外,文献[11]提出了一种基于预先划分匿名区域的连续查询位置隐私保护算法.该算法将用户的运动轨迹按时间预先分成 m 段,使得各时间段中的匿名区域和最小,从而在保证隐私安全的同时提高服务质量.然而,以上工作仅适用于自由空间.

路网中仅有的一些连续查询隐私保护工作中,Palanisamy 等人[12-14]提出了一种基于 mix-zone 的隐私保护技术,其主要思想是以路网交叉路口为中心,沿着路网出口的方向建立具有自适应特点的非矩形 mix-zone.每当用户进入到该 mix-zone 中,不作任何位置更新并更换用户进入 mix-zone 前的假名.由于用户在 mix-zone 中位置信息被隐藏同时其对应假名信息也发生了变更,增加了将同一个用户前后使用的假名关联起来的难度,从而达到了保护用户标识的目的.但因 mix-zone 匿名技术是一种不规则形状局部匿名的思想,其缺点是不能够保证查询用户的全局匿名.当用户处于不规则区域外时,需要发布确切位置.

文献[7]针对路网中的连续查询攻击问题,受自由空间中将空间划分成不同的区域的匿名处理思想启发,提出了一种基于 Snet 层级结构的隐私保护方法,将路网预先划分成不同的路段子图,每一个子图为基础匿名单元,通过历史数据预测移动用户的转移概率,从而基于 Snet 单元构造具有相似运动趋势的匿名集.该方法中,为了抵御连续查询攻击,要求连续查询有效期内的匿名用户集之间的公共用户需满足全局 K 匿名要求.另外,为了抵御查询同质性攻击,要求连续查询有效期内匿名集中的查询服务属性满足全局查询 L 多样性要求.然而,该工作仅能防止连续查询攻击,忽略了位置依赖攻击造成的位置泄露问题.

2 预备知识

2.1 路网模型

定义 1. 路网(road network).路网被形式化地表示为一个无向图 $G(V,E)$.其中,$V=\{n_1,n_2,\cdots,n_n\}$,表示路网各结点集合;E 为边集,其中的每一条边表示为$(e_{id},n_in_j,d(e),v_{\max})$,$e_{id}$ 为路段标识,n_in_j 为结点序列表示该路段是从结点 n_i 到结点 n_j($n_i,n_j\in V$),$d(e)$表示边的长度,$v_{\max}$为该路段的最大速度限制.

如果一个移动用户 u 在边 e 上,则用户 u 的位置信息被形式化地表示为 $u=\{e_{id},dist,v\}$,$dist$ 为用户 u 所在位置距其所在边的起点之间的距离,v 为用户的运动速度.路网中的 2 点 p_i 和 p_j 之间的距离为点 p_i 到点 p_j 之间的最短路径,可表示为 $SP(p_i,p_j)$.

定义 2. 边到边的网络距离[8].给定 2 条边 e_i 和 e_j,则 2 边之间的距离可表示为

$$SP(e_i,e_j)=\min\{SP(a_1,b_1)+d(e_{\min}),SP(a_1,b_2)+d(e_{\min}),SP(a_2,b_1)+d(e_{\min}),SP(a_2,b_2)+d(e_{\min})\},$$

其中,$\{a_1,a_2\}$和$\{b_1,b_2\}$分别为边 e_i 和 e_j 的 2 邻接点,$d(e_{\min})$为 $d(e_i)$和 $d(e_j)$的较小者.

以图 1 所示的路网结构为例,如边 e_{14},其括号

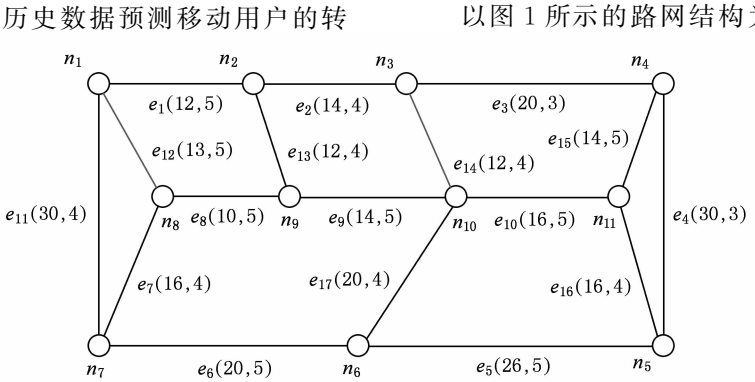

Fig. 1 The road networks

图 1 路网结构

内,12 表示的是边的长度,4 为最大速度限制. 边 e_{14} 到 e_{12} 之间的网络距离 $SP(e_{14},e_{12})=\min\{SP(n_3,n_1)+d(e_{14}),SP(n_3,n_8)+d(e_{14}),SP(n_{10},n_1)+d(e_{14}),SP(n_{10},n_8)+d(e_{14})\}=SP(n_{10},n_8)+d(e_{14})=36$.

2.2　连续查询相似度

一般情况下,连续查询可以表示为 2 种形式:

1)已知用户当前位置和查询有效期[5,7,9],如"查询从当前位置开始 5 min 内距离我最近的医院";2)已知用户在查询有效期内的路径[15],如"查询从人民大学路径北京大学、清华大学到达上地过程中距离我最近的饭店". 鉴于道路网络上用户运动在交叉路口运动的不可预知性,本文采用第 2 种形式表示道路网络上移动用户的连续查询.

定义 3. 连续查询. 一个连续查询被形式化的表示为路段结点的有序集合,即 $path_i=\{n_1^i,n_2^i,\cdots,n_m^i\}$,其中 $n_j^i\in V,1\leqslant j\leqslant m$.

为了简便,假设系统中的连续查询用户的起点和终点都为路段结点. 若连续查询用户位于路段中的某个感兴趣点上,系统会自动将其转化为最近邻路段结点.

为了防止连续查询攻击,文献[5]提出在查询有效期内保证用户在每一个时刻发布的匿名集需要包含完全相同的用户. 为了保证匿名集用户在查询有效期内具有较小的查询代价,需要将相似用户匿名在一起. 本文从空间位置和用户速度 2 个方面评价移动用户的相似性.

定义 4. 空间相似性 ζ. 已知有 2 个用户 u_i 和 u_j,其对应的连续查询运动路径分别为 $path_i$ 和 $path_j$,连续查询的空间相似性 ζ 可表示为

$$\xi(u_i,u_j)=\frac{path_i\bigcap path_j}{|\max(path_i.length,path_j.length)|},\tag{1}$$

其中,分子为路段的公共结点数,分母为路段结点数的较大值. 易知 ζ 的取值范围为[0,1],越接近 1 则表示空间越相似.

在给出查询用户的速度相似性前,先给出查询用户 u 在其运动路径上的平均速度 v_a 的计算为

$$v_a=\frac{\sum\limits_{e=n_in_{i+1},n_in_{i+1}\in path}e.v_{\max}}{|path.length|-1}\times\frac{u.v}{u.v_{\max}}.\tag{2}$$

平均速度 v_a 的计算公式表示的是:根据用户的当前运动速度来预测用户在整个连续查询运动路径上的平均速度. 其中,$u.v$ 为查询用户在当前路段上的运动速度,$u.v_{\max}$ 为当前路段的最大限制速度.

定义 5. 速度相似性 η. 已知有 2 个连续查询用户 u_i 和 u_j,$u_i.v_a$ 和 $u_j.v_a$ 为用户 u_i 和 u_j 在其运动路径上的平均运动速度. 则查询用户间的速度相似性 η 可表示为

$$\eta(u_i,u_j)=1-\frac{|u_i.v_a-u_j.v_a|}{\max(u_i.v_a,u_j.v_a)}.\tag{3}$$

定义 5 中的速度相似性 η 仅考虑速度的大小差异. 因用户的连续查询运动路径为有序结点的集合,已经反应了用户的运动方向. 易知 η 的取值范围为[0,1],越接近 1 则表示速度越相似.

定义 6. 时空相似性 δ. 已知有 2 个用户 u_i 和 u_j,以及他们的空间相似性 ζ 和速度相似性 η. 查询用户的时空相似性 δ 可表示为

$$\delta(u_i,u_j)=w_1\times\zeta+w_2\times\eta,\tag{4}$$

其中,系数 w_1 和 w_2 分别决定了参数 ζ 和 η 的权重,有 $w_1,w_2\geqslant 0$ 且 $w_1+w_2=1$.

易知,δ 具有 3 种性质:

1) $\delta(u_i,u_j)\geqslant 0$;

2) $\delta(u_i,u_j)=\delta(u_j,u_i)$;

3) $\delta(u_i,u_j)\in[0,1]$.

定义 7. (K,L)-匿名模型. 设 CS_i 为用户 u 的匿名集,其中,$CS_i=\{Cus_i,Css_i\}$,Cus_i 为匿名用户集合,Css_i 为匿名路段集合,则有 4 个条件:

1) $|Cus_i|\geqslant K$;

2) $Cus_i=Cus_1$;

3) $|Css_i|\geqslant L$;

4) $\forall u\in Cus_i$,u 发布 Css_i 作为匿名区域.

其中,条件 1 确保了匿名集用户满足位置 K 匿名模型;条件 2 保证了 CS_i 在每一个时刻发布的匿名用户集包含完全相同的用户;条件 3 保证了每个匿名路段集满足位置 L 多样性①;条件 4 确保了每个 CS_i 中的用户满足位置 K-共享性质.

3　基于时空相似性的连续查询匿名算法

本文采用中心服务器结构[6-9],由匿名服务器完成匿名工作. 3.1 节说明道路网络上基于时空相似性的连续查询隐私保护算法 SCPA 的主要思想,3.2 节介绍启发式宽度优先用户搜索算法 HBFS,

① L 可以根据不同的应用和隐私需求,设置不同的语义,如 L 个不同路段数或 L 个不同敏感度的感兴趣点等. 简单起见,本文中的 L 指的是 L 个不同路段.

3.3 节为连续时刻内匿名路段集生成算法 CSGA 及安全分析.

3.1 SCPA 主要思想

匿名服务器中的连续查询可以分为新到查询用户和活动查询用户 2 种. 新到查询用户是指刚刚提出连续查询的用户;活动查询用户是指在连续查询有效期内发生位置更新的用户.

SCPA 算法的基本思想是:若用户为新到查询用户,则基于时空相似性为查询用户分组,构造匿名用户集合 Cus,并生成在初始位置的匿名路段集合 Css_1(具体参见 3.2 节). 若用户为活动用户,根据已知的匿名集用户进行匿名位置 Css_i 的更新(具体参见 3.3 节).

算法 1. 道路网路上基于时空相似性的连续查询隐私保护算法 SCPA.

输入:连续查询的路径、匿名度需求 K、位置 L 多样性、路网 $G(V,E)$;

输出:时刻 t_i 的匿名集 $Cs_i(1\leqslant i\leqslant m)$.

① if {*Users*}是新到查询用户 then
② while true do
③ u 从 *Users* 集合中随机选取一个用户;
④ if {*Users*}不为空 then
⑤ 使用算法 2 构造匿名用户集 *Cus*;
⑥ *Users*=*Users*−*Cus*;
⑦ else
⑧ break;
⑨ end if
⑩ end while
⑪ for $Cus_i\in\{Cus\}$ do
⑫ 根据匿名用户集 *Cus*,找到 *Cus* 中用户所在路段的集合 Css_1;
⑬ while $|Css_1|<L_{\max}$ do
⑭ 选取至少有一个公共结点的路段加入 Css_1;
⑮ end while
⑯ end for
⑰ return $\{Cs_1\}$;
⑱ else
⑲ 运用算法 3 生成连续匿名路段集$\{Css_i\}$;
⑳ return $\{Cs_i\}$;
㉑ end if

算法 1 中显示了 SCPA 的具体步骤. 首先,若查询用户为新到查询用户,则应用算法 2 为新到查询用户 u 寻找合格匿名用户集 *Cus*(行⑤);然后将集合 *Cus* 中的用户从集合 *Users* 中移除,重复执行行②～⑩,直到集合 *Users* 为空;接着,根据匿名用户集 *Cus*,找到 *Cus* 中用户所在路段的集合作为匿名路段集返回(行⑪～⑯). 如果查询用户为活动查询用户,则应用算法 3 更新匿名路段集合并返回(行⑲～⑳).

3.2 启发式宽度优先用户搜素算法

在匿名初始阶段,仅有用户当前时刻的位置,所以此阶段不存在位置依赖攻击. 为了防止连续查询攻击,只需找到在查询有效期内能够保持一致的匿名用户集 *Cus*.

算法 2. 启发式宽度优先用户搜索算法 HBFS.

输入:连续查询用户 u、路网 $G(V,E)$;

输出:匿名用户集合.

① $Uset=\varnothing$;
② while true do
③ 从$u.edge$ 开始在 $G(V,E)$ 上执行宽度有限搜索;
④ if 在 $u.path.size()\times(1-\delta)$中的结点未被访问过 then
⑤ for each *user* in *edge* do
⑥ 计算 $Similarity(u,user)$;
⑦ if $Similarity\geqslant\delta$ then
⑧ $Uset$.add(*user*);
⑨ 更新 $u.K$;
⑩ end if
⑪ end for
⑫ end if
⑬ if $|Uset|\geqslant u.K \parallel times>(1-\delta)\times 10$ then
⑭ break;
⑮ else
⑯ 从用户 u 当前所在位置进行扩展遍历;
⑰ $times$++;
⑱ end if
⑲ end while
⑳ if $|Uset|\geqslant u.K$ then
㉑ return *Uset*;
㉒ else
㉓ break;
㉔ end if

基于定义 6 的时空相似性 δ,本文提出了一种启发式宽度优先用户搜素算法 HBFS. 其主要思想是从用户当前所在位置的附近,寻找与用户具有时空相似性的用户. 若能够找到满足用户 K 隐私需求

的候选匿名用户集,则将其作为匿名用户集返回;否则以宽度优先的方式从用户当前位置进行扩展,继续寻找满足要求的用户.

具体算法参见算法 2. 对于每个待匿名查询用户 u,从 u 在路网上的初始位置开始做宽度优先遍历搜索(行③). 依次计算查询用户的时空相似性 δ(行⑥),将满足要求的查询用户放入集合 $Uset$ 中(行⑦～⑨). 若能找到 K 个合格用户,则可从行②～⑲的循环跳出,否则将用户 u 从当前位置进行宽度扩展(行⑯),然后重新执行行②～⑲. 最后,判断用户集合 $Uset$ 包含的用户个数是否大于等于 K,若满足,则将 $Uset$ 返回(行⑳～㉑).

表 1 给出了 5 个新到查询用户的信息,其在路网中的分布情况如图 2 所示(实心点). 设时空相似性 δ 默认为 0.7. 随机选取 u_1,做启发式宽度优先遍历搜索用户. 可得用户 u_2 和 u_3 满足时空相似性 δ 要求,候选匿名用户集 $Uset=\{u_1,u_2,u_3\}$. 又 u_1 的隐私需求 $K=5$,将用户 u 从当前位置进行宽度扩展,再次计算,可得 u_4,u_5 也满足时空相似性 δ 要求. 最后,有 $Uset=\{u_1,u_2,u_3,u_4,u_5\}$ 作为匿名用户集返回.

Table 1　Sample of Query Users

表 1　查询用户信息

Query User	(K,L)	v	Continuous Query Path
u_1	(5,5)	4	$p_1=\{\langle n_1,n_8,n_9,n_{10},n_{11},n_4\rangle\}$
u_2	(4,4)	4	$p_2=\{\langle n_7,n_8,n_9,n_{10},n_{11},n_5\rangle\}$
u_3	(4,4)	3	$p_3=\{\langle n_1,n_2,n_3,n_{10},n_{11}\rangle\}$
u_4	(3,3)	3	$p_4=\{\langle n_8,n_9,n_{10},n_{11}\rangle\}$
u_5	(3,4)	3	$p_5=\{\langle n_7,n_6,n_5\rangle\}$

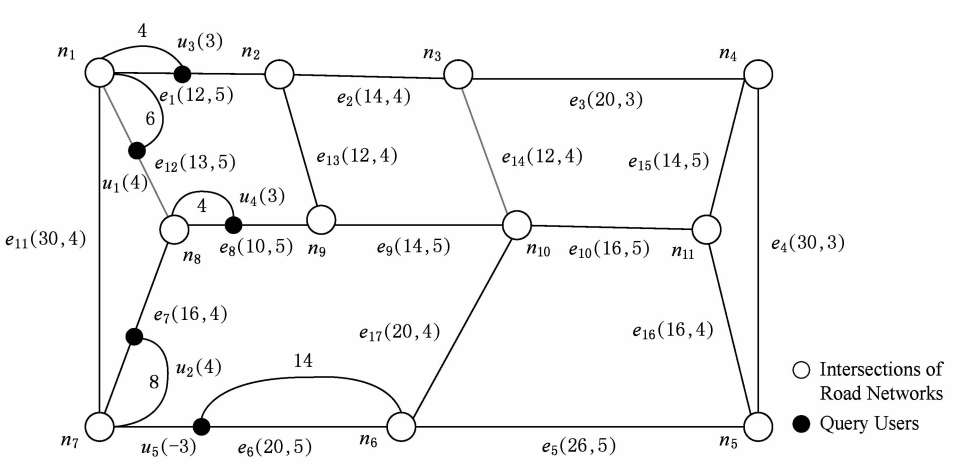

Fig. 2　Continuous queries on road networks

图 2　路网中的连续查询

3.3　连续时刻内匿名路段集生成算法

当查询用户在连续时刻内发生位置更新时,由于匿名集用户和用户所在路段已知,可根据查询用户的匿名用户集所在位置,直接采用算法 1 中(行⑪～⑮)寻找满足位置 L 多样性的匿名路段. 但是这样的做法,会造成位置依赖攻击.

定义 8. 边到边集的网络距离. 给定一条边 $edge$ 和一个边的集合 Css,边 $edge$ 到 Css 的网络距离可表示为

$$ND(edge,Css)=\max\{e\in\{Css\}\mid SP(edge,e)\},$$

直观上,边到边集的网络距离是边到边的网络距离的最大值.

定义 9. 边集到边集的网络距离[8]. 给定 2 个边集 Css_A 和 Css_B,2 边集之间的网络距离可表示为

$$ND(Css_A,Css_B)=\max\{\forall e_x\in Css_A\mid ND(e_x,Css_B)\},$$

易知,$ND(Css_A,Css_B)=ND(Css_B,Css_A)$.

为了防止道路网路上的连续查询最大速度位置依赖攻击,文献[8]证明了任意连续时刻的匿名路段集合之间的网络距离满足用户最大速度限制要求. 然而,但未考虑用户在不同路段上具有不同的限制速度的情况. 为保证匿名集的位置 K-共享属性,本文在构造匿名路段集时,采用式(5)来判断路段是否满足匿名要求:

$$ND(Css_i,Css_{i-1})\leqslant v_{\min}\Delta t,\tag{5}$$

其中,Css_i 和 Css_{i-1} 为 2 个连续时刻的匿名路段集,$v_{\min}$ 为匿名用户集 Cus 中用户的最小速度,Δt 为查询更新的间隔时间.

易知,对于任意 2 个时刻的匿名路段集,若匿名用户集中的用户能以最小速度到达最大距离位置,很明显,对于匿名集中的任意用户则都可达. 所以应用式(5)可保证匿名集满足位置 K-共享属性,从而能够有效防止位置依赖攻击.

算法 3. 连续时刻匿名路段集生成算法 CSGA.

输入：连续查询用户 u、路网 $G(V,E)$；

输出：匿名路段集合.

① Css_{i-1}=在 t_{i-1} 匿名路段集合；

② Cus=用户 u 的匿名集用户集合；

③ 找到 Cus 中用户当前时刻所在路段 Css_i；

④ if Css_i 是一个非连通图 then

⑤ 寻找连通路段加入路段集合 Css_i；

⑥ end if

⑦ while $|Css_i|<L_{\max}$ do

⑧ 选取和匿名路段集合 Css_i 具有公共结点的路段插入 Css_i；

⑨ end while

⑩ if $ND(Css_i,Css_{i-1})\geqslant v_{\min}\Delta_t$ then

⑪ $t_i=ND(Css_i,Css_{i-1})/v_{\min}+t_{i-1}$；

⑫ end if

⑬ return Css_i.

具体算法参见算法 3. 对于活动用户 u，其匿名用户集合是 Cus. 找到 Cus 中用户当前时刻所在路段(行③). 若这些路段不连通，则找连通路段加入路段集合(行④～⑥)；接着判断路段集合是否满足隐私需求 L. 若满足，则从行⑦～⑨的循环跳出；否则，优先选取和匿名路段集合具有公共结点的路段(行⑧)，直到 Css_1 满足隐私需求 L 为止. 最后，利用式(5)检查匿名路段集是否满足位置依赖攻击的速度限制要求(行⑩)，若不满足，则计算延迟时间，然后将匿名路段集 Css_i 返回(行⑪～⑬).

继续图 2 中的例子，假设查询用户 u_1 在时刻 t_2 更新查询，此时，匿名集中的用户 u_1，u_2，u_4 位于 e_{10}，u_3 位于边 e_{14}，u_5 位于边 e_5 上. 根据连续时刻内匿名路段集生成算法 CSGA，可知，首先应找到连通路段 e_{16}，接着判断用户的 L 隐私需求. u_1 所在匿名集的 L 位置多样为 5，将边 e_{17} 加入匿名路段集，接着利式(5)判断匿名路段集是否满足位置依赖攻击的速度限制要求. 最后，得到查询用户 u_1 在时刻 t_2 的匿名路段集合集 $Css_2=\{e_{10},e_{14},e_5,e_{16},e_{17}\}$.

安全分析如下：SCPA 算法生成的匿名集采用了降低位置粒度的方法保护位置隐私，同时满足位置 K 匿名模型. 位置 K 匿名模型可以保护用户标识；L 路段多样性将用户的确切位置隐藏于 L 个不同的路段中. 定义 7 中的条件 2 确保了连续查询的匿名集包含完全相同的用户，条件 4 确保在匿名集中的用户均以匿名路段集 Css 作为匿名位置，保证了位置 K-共享性，有效防止了连续查询攻击. 另外，利用式(5)保证了匿名路段集之间的网络距离满足位置依赖攻击的速度限制要求，可有效防止道路网络中的位置依赖攻击.

4 实验结果与分析

4.1 实验设置

本文比较了 SCPA 算法和 NVBA 算法. NVBA 算法为文献[8]中抵御路网中的连续查询最大速度攻击的位置匿名算法. 所有的匿名算法用 Java 实现，并运行于 2.4 GHz 处理器、2 GB 内存的 Windows 7 计算机上. 本实验使用 Thomas Brinkhoff 移动对象生产器① 生成模拟移动对象. 生成器的输入是 Oldenburg 地图，包含 6 105 个路段结点和 7 035 条边. 实验模拟生成 10 000 个模拟移动对象，设每个对象均提出连续查询，每 10 s 更新一次位置且默认包括 100 个快照位置，隐私需求 K 和 L 的默认值都为 5，时空相似性 δ 的默认值为 0.7. 表 2 中列出了实验参数具体信息.

Table 2 Default System Settings

表 2 实验参数及默认取值

Parameter	Default Settings
Number of Mobile Objects	10 000
K	5
L	5
δ	0.7
Number of Snapshots	100

4.2 评价标准

采用的评价标准包括信息熵、匿名成功率、查询处理代价、匿名时间.

1) 信息熵[16-17] 反映匿名算法为移动用户提供的保护强度. 信息熵越大，则保护强度越大.

2) 匿名成功率是指成功匿名的移动用户在所有提出查询的移动用户中所占比例. 匿名成功率越高，说明匿名算法对查询响应能力越好.

3) 匿名时间指的是一定规模移动用户的所有查询请求在多长时间内可以得到匿名处理. 这是反映匿名算法好坏的重要指标之一. 匿名时间越短越好，说明了匿名算法的高效性.

① http://www.fhoow.de/institute/iapg/personen/brinkhoff/generator/

4) 查询处理代价是指位置经过匿名处理后,在服务提供商端由于保护隐私查询处理产生的额外代价.实验利用文献[17]中的查询代价模型,使用匿名路段集中包含的路段和开放点个数评价查询代价.查询代价越小越好.

4.3　性能评估

1) 匿名度 K 的影响.该部分评价了匿名度 K 对匿名算法性能的影响.匿名度 K 从 3 增加到 15.随着匿名度 K 的增加,意味着每一个匿名集需要包含更多的用户.从图 3(a)观察到,无论在何种设置下,SCPA 的信息熵均比 NVBA 高.因为 SCPA 算法采用用户分组策略,并实现了 K-用户共享.所以相比于 NVBA 算法能够更好地抵御连续查询攻击.图 3(b)显示随着匿名需求的变化,2 个匿名算法的成功率均呈现下降的趋势.因为当 K 值越来越大时,用户的匿名度要求变的更严格,寻找满足匿名度要求的用户集将变得越来越难.但 SCPA 算法的匿名成功率要比 NVBA 的高.

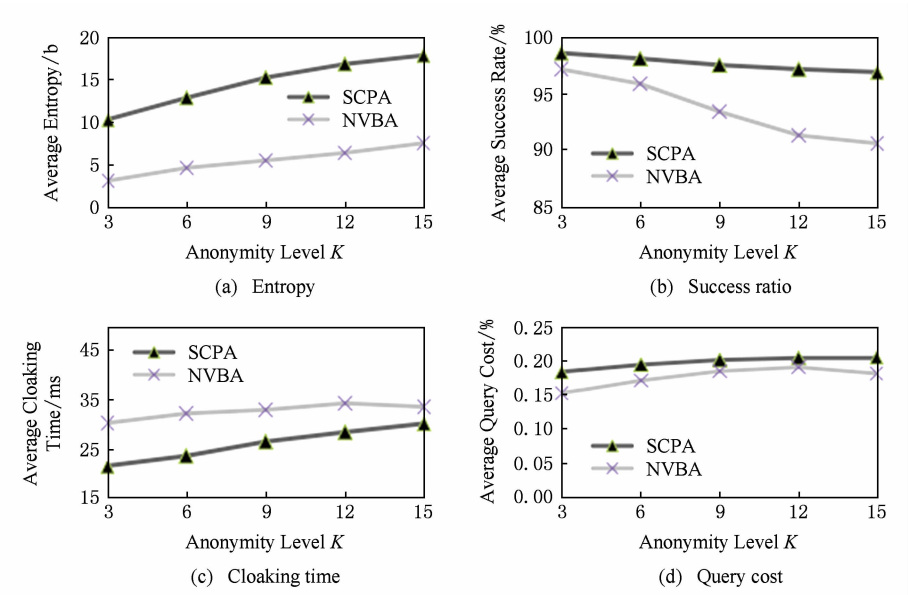

Fig. 3　Evaluation on different anonymity level K

图 3　2 种算法在不同 K 匿名度需求下的比较

比较 SCPA 与 NVBA 的平均匿名时间.图 3(c)显示,随着匿名度 K 的增加,2 种算法的匿名时间也呈上升的趋势,且 SCPA 比 NVBA 的算法效率更好.因为 SCPA 算法采用分组构造匿名集,相比之下,NVBA 需要为每一个用户重复遍历整个道路网络,将耗费更多的时间.如图 3(d)所示,2 种算法的查询代价均随着 K 的增加而增加,因为匿名集中包含了更多的路段.虽然 SCPA 比 NVBA 提供了更好的隐私保护和算法效率,但 SCPA 的查询代价较之 NVBA 高,然而,SCPA 仅比 NVBA 大约多花费了 0.03%查询代价.

2) 位置多样性 L 的影响.该部分评价匿位置多样性 L 对匿名算法性能的影响.路段差异性需求 L 增加意味着用户的位置隐私需求更加严格,L 亦从 3 增加到 15.从图 4(a)观察到,SCPA 的信息熵在所有设置上均比 NVBA 高,平均要高出 5 倍,这说明 SCPA 比 NVBA 提供了更安全的隐私保护强度.从图 4(b)中,可以看到 NVBA 的匿名成功率随着 L 的增大逐渐下降,相比 NVBA,SCPA 算法的匿名成功率要比 NVBA 的高,且 SCPA 具有更稳定的匿名成功率.从图 4(c)可以看出,对于 L 的增加,SCPA 的平均匿名时间并没有太大变化,从整体情况来看,SCPA 比 NVBA 的算法效率更好.虽然 SCPA 比 NVBA 提供了更好的隐私保护和较好的算法效率,但 SCPA 的查询代价较之 NVBA 高,从图 4(d)看出,2 种算法的查询代价均随着 L 的增加而增加,因为匿名集中包含了更多的路段.

3) 时空相似性 δ 的影响.该部分对系统参数时空相似性 δ 进行实验评估,以验证时空相似性 δ 对 SCPA 算法的性能的影响.由于 NVBA 不涉及时空相似性,所以此节仅对 SCPA 进行了实验验证.δ 值从 0.5 变化到 0.9.

从图 5(a)和 5(b)中,可以看到,在 δ 取值为 0.7 之前,信息熵和匿名成功率基本保持不变,但当 δ 大于 0.7 后信息熵和匿名成功率都逐渐下降.因为随着 δ 值变大,将很难为所有查询用户构造具有高时空

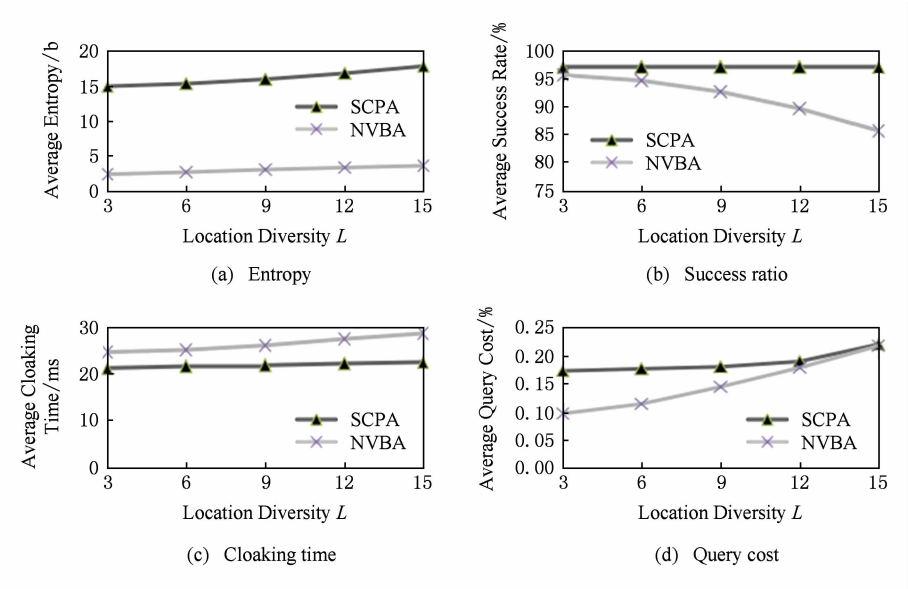

Fig. 4 Evaluation on different location diversity L

图 4 2 种算法在不同路段差异性 L 下的比较

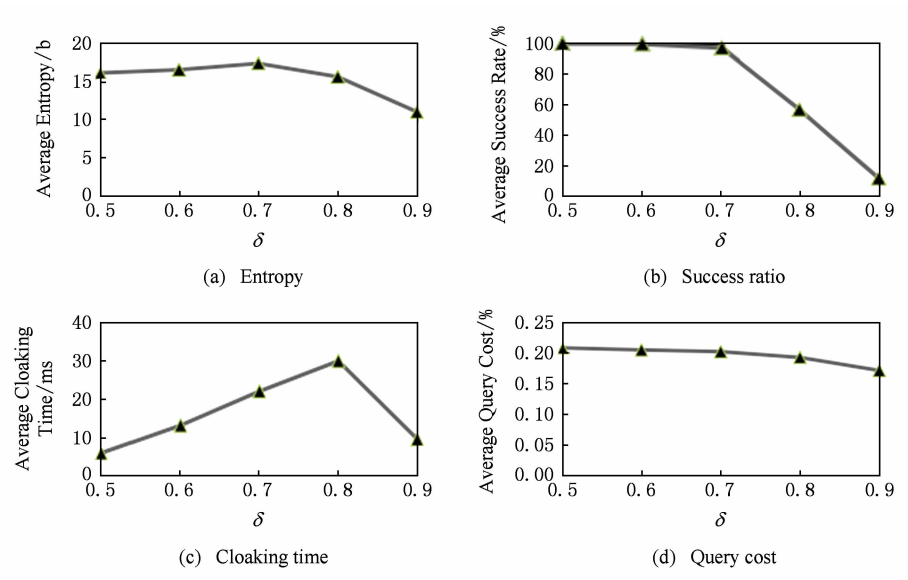

Fig. 5 Evaluation on different δ of SCPA Algorithm

图 5 SCPA 算法在不同时空相似性 δ 下的性能评估

相似性的匿名集,从而导致用户的平均信息熵和成功率都呈下降趋势. 图 5(c)显示,随着 δ 的增加,平均匿名时间先增加,在 δ 值为 0.8 时达到最大,之后急剧下降. 因为随着时空相似性 δ 的增加,在为查询用户做启发式宽度优先用户搜索时,将需要花费更多的时间. 然而,当 δ 值达到一定值时,如 0.9 时,因系统中大部分查询用户难以找到具有如此高相似的匿名用户集,所以在为用户生成匿名路段集时将花费相对较少的时间. 由图 5(d)中所示,随着 δ 的增加,查询代价逐渐递减. 因为 δ 越大,表明查询用户越相似,匿名用户在路网上将越集中,所以构造匿名路段集时所产生的开放点个数将相对减少,从而查询代价变小.

5 结 论

本文研究了一种在道路网络上基于时空相似性的连续查询隐私保护算法 SCPA. 现有在道路网络上的位置隐私保护工作大多针对快照查询提供隐私保护,当移动用户的位置发生连续更新时,容易遭受连续查询攻击和位置依赖攻击. 为了解决此问题,本文基于查询用户的时空相似性采取分组构造匿名集

实现 K-共享机制,并提出了一种启发式的宽度优先用户搜索算法和连续时刻内匿名路段集生成算法.最后通过一系列实验,验证了算法的有效性,匿名算法可保证平均 95%以上的匿名成功率,且平均匿名时间为 18 ms.

参　考　文　献

[1] Wang Lu, Meng Xiaofeng. Location privacy preservation in big data era: A survey [J]. 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[J]. 软件学报, 2014, 25(4): 693-712)

[2] Terrovitis M, Liagouris J, Mamoulis N, et al. Privacy preservation by disassociation [J]. Proceedings of the VLDB Endowment, 2012, 5(10): 944-955

[3] Chow C, Mokbel M F, Bao J, et al. Query-aware location anonymization for road networks [J]. Geoinformatical, 2010, 15(3): 571-607

[4] Wang Ting, Liu Ling. Privacy aware mobile services over road networks [C] //Proc of the 35th Int Conf on VLDB. New York: ACM, 2009: 1042-1053

[5] Chow C Y, Mokbel M F. Enabling private continuous queries for revealed user locations [C] //Proc of the 10th Int Conf on SSTD. Berlin: Springer, 2007: 258-275

[6] Xu Jianliang, Tang Xueyan, Hu Haibo, et al. Privacy-conscious location-based queries in mobile environments [J]. IEEE Trans on Parallel & Distributed Systems, 2010, 21(3): 313-326

[7] Wang Yong, Xia Yun, Hou Jie, et al. A fast privacy-preserving framework for continuous location-based queries in road networks [J]. Journal of Network & Computer Applications, 2015, 53(3): 57-73

[8] Wang Yilei, Zhou Hao, Wu Yingjie, et al. Preserving location privacy for location-based services with continuous queries on road network [C] //Proc of the 7th Int Conf on Computer Science & Education. Los Alamitos, CA: IEEE Computer Society, 2012: 822-827

[9] Pan Xiao, Meng Xiaofeng, Xu Jianliang. Distortion based anonymity for continuous queries in location based mobile services [C] //Proc of the ACM SIGSPATIAL, New York: ACM, 2009: 256-265

[10] Dewri R, Ray I, Ray I, et al. Query m-Invariance: Preventing query disclosures in continuous location-based services [C] //Proc of the 11th Int Conf on Mobile Data Management. Piscataway, NJ: IEEE, 2010: 95-104

[11] Shin H, Vaidya J, Atluri V, et al. Ensuring privacy and security for LBS through trajectory partitioning [C] //Proc of the 11th Int Conf on Mobile Data Management. Piscataway, NJ: IEEE, 2010: 224-226

[12] Palanisamy B, Liu L, Lee K, et al. Anonymizing continuous queries with delay-tolerant mix-zones over road networks [J]. Distributed & Parallel Databases, 2014, 32(1): 91-118

[13] Palanisamy B, Liu L. MobiMix: Protecting location privacy with mix-zones over road networks [C] //Proc of Int Conf on ICDE. Piscataway, NJ: IEEE, 2011: 494-505

[14] Palanisamy B, Liu Ling, Lee K, et al. Location privacy with road network mix-zones [C] //Proc of Int Conf on MSN. Piscataway, NJ: IEEE, 2012: 124-131

[15] Tao Yufei, Papadias D, Shen Qiongmao. Continuous nearest neighbor search [C] //Proc of the 28th Int Conf on VLDB. New York: ACM, 2002: 287-298

[16] Zeng Zhihao, Sun Qi, Yao Bei, et al. A virtual trajectory privacy protection method for continuous queries [J]. Science Technology and Engineering, 2014, 33(1): 93-98 (in Chinese)
(曾志浩,孙琪,姚贝,等. 一种面向连续查询的虚拟轨迹隐私保护方法[J]. 科学技术与工程, 2014, 33(1): 93-98)

[17] Pan Xiao, Chen Weizhang, Wu Lei, et al. Protecting personalized privacy against sensitivity homogeneity attacks over road networks in mobile services [J]. Frontiers of Computer Science, 2016, 10(2): 370-38

Pan Xiao, born in 1981. PhD. Associate professor at Shjiazhuang Tiedao University. Member of CCF. Her main research interests include mobile computing, social network, and privacy protection.

Chen Weizhang, born in 1992. Master. His main research interests include data mining, data management on moving objects and privacy-aware computing.

Sun Yige, born in 1996. Undergraduate. Her main research interests include data management on moving objects and privacy-aware computing.

Wu Lei, born in 1980. Master. Lecturer at Shjiazhuang Tiedao University. Member of the Soft Science Research Institute on Engineering and Construction Management in Hebei Province. His main research interests include data management on moving objects and location based social network.