控制与决策
Control and Decision

基于近似匹配的假位置k-匿名位置隐私保护方法

张永兵， 张秋余， 李宗义， 段宏湘， 张墨逸

引用本文:
张永兵, 张秋余, 李宗义, 等. 基于近似匹配的假位置k-匿名位置隐私保护方法[J]. 控制与决策, 2020, 35(1): 65-73.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2018.0783

文章编号: 1001-0920(2020)01-0065-09 DOI: 10.13195/j.kzyjc.2018.0783

基于近似匹配的假位置k-匿名位置隐私保护方法

张永兵[1,2], 张秋余[1†], 李宗义[2], 段宏湘[1], 张墨逸[1]

(1. 兰州理工大学 计算机与通信学院，兰州 730050；2. 甘肃机电职业技术学院 电气工程系，甘肃 天水 741001)

摘 要: 为了提高假位置k-匿名位置隐私保护方法中的假位置生成效率和查询服务质量,以及解决假位置生成过程中预处理复杂、没有充分考虑地理语义信息特征等问题,提出一种基于近似匹配的假位置k-匿名位置隐私保护方法. 首先,将所选区域划分为若干个正方形网格,并将各位置坐标按所在网格转换为莫顿码;然后,通过对各位置莫顿码之间的近似匹配,选取互不相邻、分布在不同网格的位置点,生成假位置候选集;最后,对候选集中位置点的地名信息进行近似匹配, 得到位置点之间的语义相似度, 并选取语义相似度最小的$k-1$个位置点作为假位置. 实验结果表明,所提出的方法在保证假位置之间物理分散性和语义多样化的同时,能够提高假位置生成效率,有效平衡隐私保护效果和查询服务质量.

关键词: 基于位置的服务；位置隐私保护；k-匿名；假位置；近似匹配；语义相似度

中图分类号: TP392 **文献标志码:** A

A k-anonymous location privacy protection method of dummy based on approximate matching

ZHANG Yong-bing[1,2], *ZHANG Qiu-yu*[1†], *LI Zong-yi*[2], *DUAN Hong-xiang*[1], *ZHANG Mo-yi*[1]

(1. School of Computer and Communication，Lanzhou University of Technology，Lanzhou 730050，China；2. Department of Electrical Engineering，Gansu Institute of Mechanical & Electrical Engineering，Tianshui 741001，China)

Abstract: In order to improve the efficiency of dummy generation and the query service quality in the k-anonymous location privacy protection method of dummy, and to solve the problem of that the preprocessing is complex and the geographic semantic features are not fully considered in dummy location generation, a k-anonymous location privacy protection method of dummy based on approximate matching is proposed. Firstly, the area is divided into several square grids, and the coordinates of each location are converted to Morton code. Then, through the approximate matching between the Morton codes of different locations, a candidate set of dummies is generated, and the locations in it are non adjacent to each other and distributed in different grids. Finally, by matching approximately geographic names information of locations, the semantic similarity between any two locations in the candidate set is obtained, and $k-1$ locations with the minimum semantic similarity are selected as dummies. Experimental results show that the method can ensure the physical dispersion and semantic diversity, and can improve the efficiency of dummy generation. At the same time, the balance between privacy protection security and query service quality is achieved.

Keywords: location-based service；location privacy protection；k-anonymous；dummy；approximate matching；semantic similarity

0 引 言

随着智能移动设备的普及和GPS的发展成熟,位置服务(location-based service, LBS)已成为为移动用户提供最有前途的服务之一[1]. LBS为人们的日常生活和社交活动带来了极大的便利,同时人们也对使用LBS时导致敏感信息泄露的问题倍加关注. 由于服务的获取需要在不同的位置服务提供商(location services provider, LSP)之间进行交互工作,用户的位置数据必须在他们之间分享. 不可信的第三方通过对以上这些位置信息进行分析对比,能较为容易地获得用户的一些重要个人信息[2]. 如:获得用户在医院附近的行为能够揭露用户的健康状况;如果分析用

收稿日期: 2018-06-07；**修回日期:** 2018-08-08.

基金项目: 国家自然科学基金项目(61363078)；甘肃省高等学校科研项目(2017B-16, 2018A-187)；模式识别国家重点实验室开放课题基金项目(201700005).

责任编委: 虞文武.

†通讯作者. E-mail: zhangqylz@163.com.

户最近的出发地点和终止地点,则可得知用户的家庭住址、工作单位和工作性质等. 一旦这些隐私信息落入非法的体系中,将严重威胁用户安全. 因此,需要保护个人位置隐私,尽量不落入商人和代理人手中.

为防止隐私信息的泄露,学者们提出了许多位置隐私保护方法,主要包括模糊方法、加密方法和基于策略的方法. 其中,模糊方法为位置隐私保护的首要选择,主要通过空间匿名或假位置的技术手段实现. 空间匿名方法通常需要借助第三方可信匿名服务器(fully-trusted third Party, TTP)[3]完成隐私保护工作. 当用户需要获得位置服务时,由TTP生成一个包括该用户位置的k-匿名区域,然后发送给LBS服务器进行查询. 在这种方法中,当匿名区域面积过大时,不仅消耗较多时间,而且会降低查询的准确性,TTP也容易成为系统瓶颈. 而假位置的方法不需要构建匿名区域,不需要TTP的协助,在移动客户端通过生成假位置实现位置匿名,能很好地弥补空间匿名方法的不足.

在基于假位置的位置隐私保护中,假位置的生成效率影响位置服务的质量,假位置与真实位置之间的不可区分性影响着隐私保护的效果. 本文充分考虑位置地理语义信息特征,给出一种基于近似匹配的假位置k-匿名位置隐私保护方法. 该方法采用空间坐标转换算法,将二维位置坐标转换为二进制莫顿码[4],通过近似匹配,选取分布在不同网格且所在网格互不相邻的位置点作为假位置候选集. 然后对候选集中位置点的地名信息,应用编辑距离[5]计算地名信息的语义相似度,选取语义相似度最小的$k-1$个位置点作为假位置. 该方法在满足语义l-多样性和物理分散性的同时,可以提高假位置生成效率,从而进一步提高位置服务质量.

1 相关工作

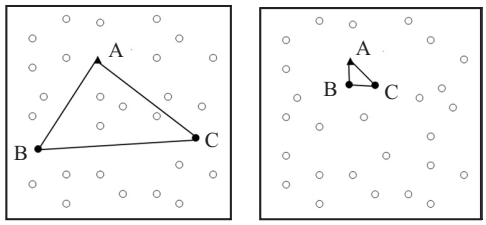
位置隐私保护方法按照体系结构主要分为两大类[6]: 基于点对点(peer-to-peer, P2P)的分布式结构[7]和基于TTP的中心服务器结构[8]. 分布式结构中,用户在客户端与近邻用户协作完成匿名,或伪造虚假位置以实现模糊化,进而实现位置隐私保护. 文献[9-11]提出了基于P2P的空间匿名方法,利用邻居节点位置信息实现k-匿名隐私保护,但忽视了邻居节点的安全问题. 基于P2P的方案具有简单灵活的优点,但增大了智能手机的各种软硬件资源开销和通信开销,并且用户时常处于移动状态[12]. 在中心服务器结构中,文献[13]应用TTP把移动端发送来的精确位置通过泛化或模糊化,实现了位置隐私保护的目的. 这种结构模式能实现良好的隐私保护效果,但对TTP需要采取安全防护措施. 文献[14]引入信息缓存机制,通过减少用户访问TTP的次数降低了信息泄露的概率,但增大了移动客户端的负担. 除此之外,Cheng等[15]提出了独立结构模式,用户根据自身能力和知识进行位置隐私保护. 这种方法结构简单,易与其他结构合并使用,但对客户端的要求较高. Li等[16]提出了多位置服务器的体系结构,根据安全要求将用户划分到不同子集,提高了位置信息的隐蔽性,适合应用在社交网络中. Mouratidis等[17]提出了基于隐私信息检索的位置隐私保护方法,采用硬盘数据检索和加密的方法实施位置隐私保护. 该方法简化了系统结构,隐私保护效果好,但增加了通信开销和硬件开销,降低了服务质量. 随着云服务技术的成熟和普及,Kim等[18]提出了可搜索加密的位置隐私保护方法,在不显示数据访问的情况下执行数据访问模式,保证了加密数据和用户查询记录的机密性,但是查询效率和查询结果的准确性有待提高.

目前,k-匿名[19]的方法仍然是位置隐私保护所采用的主流方法. k-匿名方法诞生在关系数据库中,使用泛化与模糊的技术手段对数据库中的关键属性值进行处理,使得泛化后的k条记录中,任意一条记录无法单独从中区分出来. 基于k-匿名的位置隐私保护方法主要分为空间区域匿名和假位置匿名. Gruteser等[20]通过构造k-匿名区域实现了位置隐私保护,该区域必须满足两个条件: 1) 区域面积达到一定数值; 2) 区域内必须包含k个用户. Bamba等[21]提出了网格划分的方法. 对于不同隐私度需求,提供了top-down grid cloaking和bottom-up grid两种算法可供选择使用. Xu等[22]证明了k-匿名区域的大小对查询结果准确性有很大的影响,这对匿名区域划分方法的研究提供了指导. 在此基础上,文献[23-26]提出了各种不同几何形状的匿名区域构造方法,但这些方法存在两个严重的缺陷: 1) 必须依赖于TTP,而TTP并不是绝对安全的,并且容易成为系统瓶颈; 2) 匿名区域面积的大小和查询结果精确度是一对矛盾,匿名区域面积越大,隐私保护效果越好,但查询结果精确度必然会下降.

由于空间匿名区域构造方法存在以上严重不足,假位置的方法因其不存在系统瓶颈问题、查询确准性高等优点而得到广泛研究. Kido等[27-28]最早于2005年将假位置的方法引入位置隐私保护中,不需要匿名服务器,由移动客户端生成若干假位置,然后和真实位置一起发送给LBS服务器进行查询. Lu

等[29]提出了随机化添加假位置的方法,用户能够根据自己需要在圆形或矩形区域内添加假位置实施位置隐私保护. 文献[30-32]提出了针对移动轨迹的连续查询条件下假位置隐私保护方法. Niu 等[33]考虑到掌握背景信息的敌手可能会窃取位置隐私的情况,根据熵度量选取假位置,提出了一种假位置选取(DLS)算法及其改进的 DLS 算法. 而后,Niu 等[34]提出一种基于缓存的假位置选择算法 CaDSA,将查询频率较高的位置存储在缓存中,提高了查询效率. 接着,Niu 等[35]又提出一种移动轨迹的位置隐私保护方案 DUMMY-T,旨在保护 LBS 用户隐私免遭背景攻击,先通过假位置生成(DLG)算法生成假位置,然后通过虚拟路径构造(DPC)算法生成虚假路径,从而保证了移动位置隐私安全. Sun 等[36]通过概率估计方法选取假位置,防止攻击者通过概率攻击窃取真实信息,解决了攻击者能通过分析历史数据判断出真实位置信息的问题.

上述方法都是从查询概率的角度选择假位置,没有考虑位置的地理语义信息,但攻击者可能通过地理语义信息分析判断出用户的真实位置. 图 1 中实心三角形 A 表示真实位置,空心圆点表示假位置候选集,实心圆点 B 和 C 表示选取的假位置. 图 1(a)所示的假位置选取中,A、B 和 C 三个位置点都表示医院,则攻击者通过语义分析容易判断出查询用户可能存在健康方面的问题. 图 1(b)所示的假位置选取中,所选假位置与真实位置相距太近,则攻击者通过地理距离容易找到查询用户所在的具体位置. 因此,假位置的选取应尽可能考虑位置点的地理语义信息,保证包括真实位置在内所有位置点之间的物理分散性和语义多样性,从而有效提高位置隐私保护效果.



(a) 语义分析　　(b) 地理距离

图 1　位置相似性攻击样图

针对可能遭受语义攻击的问题,Chen 等[37]提出了基于语义感知的假位置选取方法,保证了假位置点之间的物理分散性和语义多样性. 但使用欧氏距离计算两点之间的物理距离,当数据量较大时效率较低,而且通过在 WiFi APs 中构建语义树计算两个位置点之间的语义距离,增加了 WiFi APs 的负担,延长了预处理时间,因而降低了服务质量.

针对上述问题,本文提出一种近似匹配的假位置选取方法,将所选区域划分成$m \times m$个网格,计算出每个位置点所在网格的莫顿码;然后通过近似匹配,选取互不相邻、处于不同网格中的位置点作为假位置候选集,再通过其地名信息,对候选集中的位置点计算两两之间的语义相似度,选取语义相似度最小的$k-1$个位置作为假位置点. 将真实位置和$k-1$个假位置点一并发送给 LBS 服务器进行查询. 实验表明,该方法可以减轻无线 WiFi APs 的负担,简化预处理过程,在保证假位置点物理分散性和语义多样性的同时,提高假位置生成的时间效率.

2　系统模型

在基于 TTP 的中心服务器模型中,当有较多用户发起查询时,TTP 容易成为系统瓶颈,而且 TTP 也并不是绝对安全可靠的. 一旦 TTP 遭受攻击,所有位置隐私都会泄露. 因此,本文采用无 TTP 的系统模型,假位置点的生成和查询请求的发送都由移动客户端完成,系统结构模型如图 2 所示.

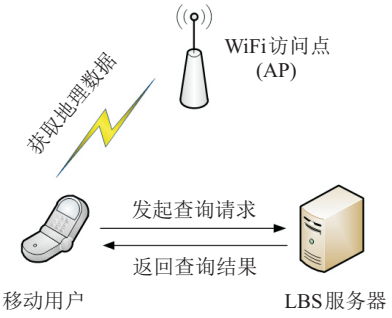


图 2　系统结构模型

根据图 2 的系统模型,移动用户通过 WiFi APs 可以获取图 3 所示的包括当前位置在内的某一区域的位置地理信息.

图 3　选取位置区域

图 4 为移动客户端将该区域划分成$m \times m$的正方形网格. 分别通过网格位置和地理语义的近似匹

配算法,选取$k-1$个互不相邻、处于不同网格、语义相似度最小的位置点作为假位置,最后将$k-1$个假位置和真实位置一起发送给LBS服务器进行查询.

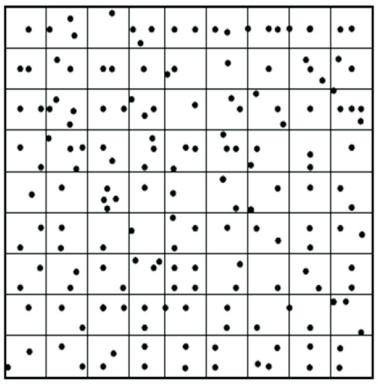


图 4 网格划分

2.1 预备知识

定义 1 令Rs表示所选取的矩形区域,Rs可定义为$\mathrm{Rs}=\{m\times m,S\}$. 其中:$m$表示在Rs区域内所划分的网格的行列数,$S$表示在Rs区域内所包含的所有位置点集.

定义 2 Rc表示每个网格所包括的区域,l_{phi}表示任意两个位置点之间的物理距离,l_{sem}表示任意两个位置点之间的语义相似度,l_{grid}表示任意两个网格之间的距离.

定义 3 S_1表示满足物理分散性的位置点候选集$\{l_1,l_2,\ldots,l_n\}$;S_2表示满足条件的假位置点集$\{l_1,l_2,\ldots,l_{k-1}\}$;假位置结果集$\mathrm{RS}_{t_i}$包括假位置集$\{l_1,l_2,\ldots,l_{k-1}\}$和真实位置$l_{\mathrm{real}}$.

定义 4 如果l_i与l_j之间的语义相似度满足:若$1-|\mathrm{SEM}_{t_i}|/C_k^2\geqslant\theta$,其中,$\mathrm{SEM}_{t_i}=\{l_{\mathrm{sem}}|l_{\mathrm{sem}}(l_i,l_j)\leqslant l\},k=\mathrm{RS}_{t_i},C_k^2$是组合公式,$l$是预设的语义多样性阈值,则称结果集$\mathrm{RS}_{t_i}$是一个$\theta$-安全集. 隐私保护的目的就是$\theta$取得最大值1,此时$l_i$与$l_j$之间的语义相似度小于等于$l$.

本文提出的假位置生成方法通过以下两个算法实现:算法1对所选区域进行网格划分,并将网格内所有位置点转换为莫顿码,通过近似匹配,选取互不相邻并处于不同网格的位置点,生成假位置候选集S_1;算法2应用编辑距离计算候选集中假位置点的语义相似度,选取其中语义相似度最小的$k-1$个位置点,生成假位置结果集S_2.

2.2 算 法 1

输入: 区域内所有位置点S,用户需求参数k;

输出: 生成一个假位置候选集S_1.

step 1: $\mathrm{Rs},S,m,(x_i,y_i)=p_i$.

step 2: 将区域Rs划分为$m\times m$的网格.

step 3: 根据网格线,沿着横坐标自左向右,在线上为1,线下为0;沿着纵坐标自上向下,在线左为1,线右为0;横坐标码值放到奇数位,纵坐标码值放到偶数位进行编码.

step 4: 重复step 3,对区域内包括真实位置在内的所有位置点进行转换,并把结果保存在双端队列D中.

step 5: 对队列D中的二进制码进行异或运算,求得的汉明距离即为各位置点之间的物理距离l_{phi}.

step 6: 通过物理距离,随机选取互不相邻并分布在不同网格中的位置点,保存在S_1中.

step 7: 生成假位置候选集S_1.

2.3 算 法 2

输入: 假位置候选集S_1,语义多样性参数l;

输出: 生成假位置结果集S_2.

step 1: 获取候选集S_1中各位置点的地名信息.

step 2: 对地名信息从前向后进行匹配,若匹配值相同,则将连续相同的若干个字符忽略,得到两个新的地名字符串A和B.

step 3: 假设A字符串包含i个字符,表示为"$A=a_1a_2a_3La_i$";B字符串包含j个字符,表示为"$B=b_1b_2b_3Lb_j$".

step 4: 构建一个$i+1$列$j+1$行的动态规划矩阵,从$D[i,j]$获得的最后一个元素取值是$\mathrm{ed}(A,B)$.

step 5: 如果$j=0$,则返回i,退出;如果$i=0$,则返回j,退出.

step 6: 将第1行初始化为$0,1,\ldots,i$;第1列初始化为$0,1,\ldots,j$.

step 7: 为矩阵中的各个元素赋值,如果$a_i=b_i$,则$D[i,j]=D[i-1,j-1]$;如果$a_i\neq b_i$,则$D[i,j]=1+\min(D[i-1,j-1],D[i-1,j],D[i,j-1])$.

step 8: 重复step 7,直到获取矩阵中所有值,最终的编辑距离为$D[i,j]$.

step 9: 通过$D[i,j]$计算相似性匹配指数$S(A,B)$,即为语义相似度.

step 10: 选取语义相似度最小的$k-1$个位置点,生成假位置结果集S_2.

2.4 算法结构与分析

2.4.1 算法1描述

获取正方形区域的位置地理信息,将该区域划分成$m\times m$的网格. 依据网格线,将所有点转换为莫顿码. 划分在同一网格中的位置点的莫顿码一定相同,因此其位置点之间的汉明距离为0. 在同一网格中只选取一个位置点,并且所选网格互不相邻,保证物理位置的分散性,如图5所示.

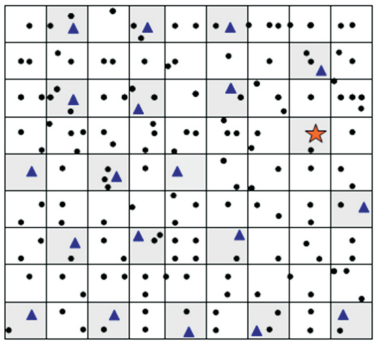


图 5　生成假位置候选集

图5中阴影网格表示选取的互不相邻网格,实心三角符号表示在该网格中选取的候选假位置,五角星表示查询用户的当前位置. 如图5所示,得到20个包含真实位置的候选集位置点.

2.4.2　算法2描述

在地名语义相似度计算中,根据中国地名的特点,首先使用“前缀词”的方法,按照正向匹配的原则,消除地名信息中的相同“前缀词”,然后对剩下的地名字符串计算其编辑距离,提高匹配效率和语义相似度的准确性. 例如“广州市第二中学”和“广州市铁一中学”,这两个地名字符串中的“广州市”3个字符对于语义相似度计算没有任何意义,还会影响计算结果的准确性.

$D[i,j]$是动态规划矩阵的编辑距离,在编辑距离的计算中,编辑操作的代价是介于[0, 1]之间的值,根据需求设定不同的值. 本文设定的值为0和1,当$a_i = b_i$时,替换的成本是0;否则,所有编辑操作的成本都是1. 下式表示计算字符串A =“第二中学”和B =“铁一中学”之间编辑距离的动态规划矩阵D:

$$D = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 2 & 3 & 4 \\ 2 & 2 & 2 & 3 & 4 \\ 3 & 3 & 3 & 2 & 3 \\ 4 & 4 & 4 & 3 & 2 \end{bmatrix}. \tag{1}$$

通过动态规划矩阵(1)计算得到字符串之间的编辑距离为$D[i,j] = D[4,4] = 2$.

按下式计算地名信息字符串之间的语义相似度:

$$S(A,B) = 1 - \frac{D[i,j]}{\max\{|A|,|B|\}} = 0.5, \tag{2}$$

其中$|A|$和$|B|$分别表示两个字符串的长度,取字符串长度最大的值参与相似性匹配指数的计算.

根据式(2)计算得到各候选位置点之间的语义相似度,然后根据下式计算得到语义相似度最小的$k-1$个位置点:

$$\mathrm{Arg}\min(S(l_i,l_j)). \tag{3}$$

2.4.3　算法分析

首先通过WiFi APs获取某一矩形区域的位置地理信息,包括地名信息和位置点的经纬度;然后将该区域划分成$m \times m$的网格,使所有位置点落在不同的网格中. 根据各位置点所在的不同网格,将位置点转换为莫顿码,对各位置点的莫顿码进行近似匹配计算,得到分布在互不相邻且处于不同网格的位置点作为假位置候选集. 对候选集中的位置点根据地名信息计算语义相似度,选取语义相似度最小的$k-1$个位置点作为假位置点.

在物理相似性计算中,将位置点地理坐标转换为莫顿码,能够提高位置点之间匹配的速度. 根据同一网格内位置点的莫顿码必然相同的特点,能够快速在每一个网格中找出一个点. 为了保证较好的分散度,位置点在互不相邻的网格中选取. 如选取当前网格中的某一点作为候选假位置,则与该网格相邻的8个网格中的位置点不再选取. 而相隔的网格数可以是一个,也可以是多个,相隔越多,分散性越好. 但必须要保证$|S_1| \geqslant 2k$,因为S_1中要有足够的位置点,可供语义相似度的计算和选取. 结果集中假位置之间的最小距离是由各自所在网格之间相隔的最少网格数决定的. 具体相隔网格数的计算可以采用以下条件确定.

假设使用$m \times m$的网格,匿名度为k,则有:

1) 当$m/k > 2$时,任意两个备选网格之间相隔$\mathrm{ceil}(m/k) - 1$个网格的距离;

2) 当$m/k = 2$时,任意两个备选网格之间相隔两个网格的距离;

3) 当$m/k < 2$时,任意两个备选网格之间相隔一个网格的距离.

在语义相似度的计算中,采用编辑距离计算地名信息中各字符串的语义相似度. 计算编辑距离时,两个字符串字符越相近,得到的编辑距离值越小,而语义相似度越大. 当两个字符串完全相同时,编辑距离值为0,计算得到的语义相似度为1.

3　实验结果与分析

实验选用谷歌地图中广州市真实位置地图数据,获取8 km×8 km矩形区域内55个位置地理信息点,对该矩形区域划分了16×16的矩形网格. 实验的主要参数k的取值为$2 \leqslant k \leqslant 30$.

实验的硬件环境为: 3.2 GHz Intel Core i5处理器,内存大小为4 GB. 操作系统为Windows 7,采用My Eclipse开发平台,以Java编程语言实现. 表1为实验默认参数配置.

表1 实验默认参数配置

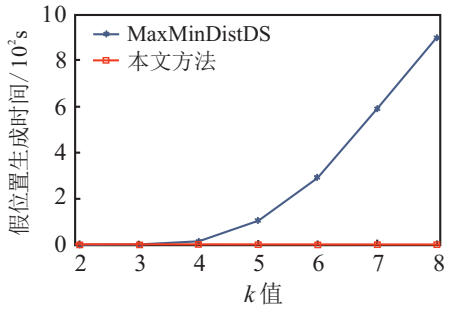
参数	值
k	$\geqslant 2$
l	0.2
网格数	16×16
位置点集	10 000
空间范围/km^2	8×8
WiFi APs覆盖范围/m	800

3.1 假位置生成效率

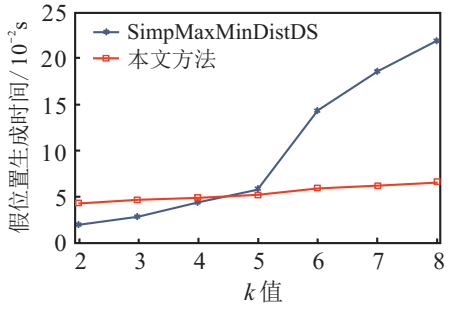
实验验证了本文方法的高效性,在考虑语义相似度的假位置选取方法中,MaxMinDistDS方法[37]、SimpMaxMinDistDS方法[37]和本文方法生成假位置平均花费的时间如表2所示. 图6为MaxMinDistDS方法、SimpMaxMinDistDS方法与本文方法假位置生成时间效率的对比结果,其中$k=2\sim 8$.

表2 假位置平均生成时间 单位:s

方法	k						
	2	3	4	5	6	7	8
MaxMinDistDS	0.07	1.69	13	106.27	295.41	592.91	899.45
SimpMaxMinDistDS	0.02	0.028	0.044	0.058	0.014 4	0.186	0.22
本文	0.043	0.047	0.049	0.052	0.059	0.062	0.066



(a) 假位置平均生成时间(1)



(b) 假位置平均生成时间(2)

图6 假位置生成时间效率对比

由图6(a)可以看出,随着k值的增加,MaxMinDistDS方法比本文方法花费的时间多很多,本文方法生成假位置的效率更高.

由图6(b)可以看出,当k值小于4时,SimpMaxMinDistDS方法比本文方法的时间效率高;当$k\geqslant 5$时,本文方法的效率更高,并且随着k值的增大,本文方法的效率优势越来越明显.

图7为本文方法与Random方法[27]、Rotation方法[30]、Footprint方法[31]、DUMMY-T方法[35]生成假位置的平均时间效率对比.

由图7可以看出,随着k值的增加,几种方法所花费的时间都在增多. 其中,本文方法和Random方法

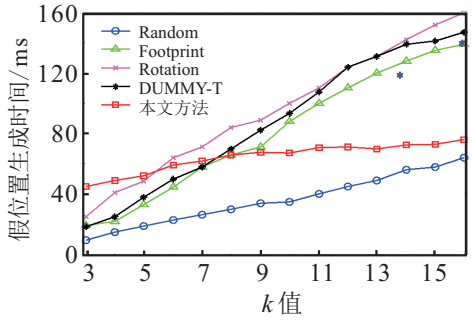


图7 假位置平均生成时间

所花费的时间比其他3种方法少,而Random方法花费的时间最少,Rotation方法花费的时间最多. 由图7可见:当$k\leqslant 5$时,本文方法花费的时间相对较多;当$k=6,7$时,本文方法与其他几种方法所花费的时间接近;而当$k\geqslant 8$时,本文方法所花费的时间比Random方法多,但比其他3种方法都少.

通过效率对比实验分析得出:当用户选取的隐私度量值k较小时,本文方法优势并不明显;但随着k值的增加,除了随机化方法,本文方法比其他方法的效率都高. 可见,在海量数据和匿名度较高的情况下,本文方法具有更高的效率,能进一步提高位置服务水平.

通过实验对比还可以看出:其他几种方法随着k值的增加,其假位置生成所花费的时间都在增加,增幅比较明显;而本文方法所花费时间的增加幅度明显较低,而且当k值越大时,本文方法时间效率高的优势更趋明显,更具有实用性.

3.2 物理分散性比较

图8为本文方法与MaxMinDistDS方法、SimpMaxMinDistDS方法假位置之间的最小距离实验对比.

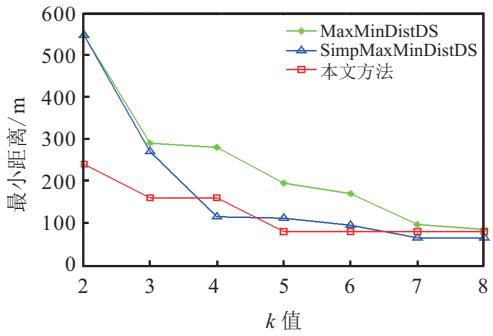


图 8　k位置点之间的最小距离

由图 8 可以看出, 几种假位置生成方法的假位置间最小距离随 k 值的变大都在减小, 但 MaxMinDist-DS 方法的最小距离值明显大于 SimpMaxMinDistDS 方法和本文方法. 但随着 k 值的增加, 三者的最小距离越来越接近.

当 $k \leqslant 8$ 时, MaxMinDistDS 方法的最小距离值最大, SimpMaxMinDistDS 方法与本文方法的最小距离比较接近; 当 $k = 2$ 时, MaxMinDistDS 方法和 SimpMaxMinDistDS 方法的最小距离均比本文方法大, 而随着 k 值的增大, 三者的差距越来越小; 当 $k = 4$ 时, 本文方法的最小距离值比 MaxMinDistDS 方法小, 但比 SimpMaxMinDistDS 方法大; 当 $k = 5, 6$ 时, 本文方法的最小距离值小于 SimpMaxMinDistDS 方法; 当 $k \geqslant 7$ 时, 本文方法的最小距离大于 Simp-MaxMinDistDS 方法, 并与 MaxMinDistDS 方法接近.

可见, 与其他两种方法相比, 本文方法生成的假位置具有较好的物理分散性.

3.3　语义多样性比较

图 9 为本文方法与 MaxMinDistDS 方法、Simp-MaxMinDistDS 方法和 DLS 方法[33] 的语义多样性对比. 通过对假位置结果集中各位置的语义多样性得到 θ-安全值.

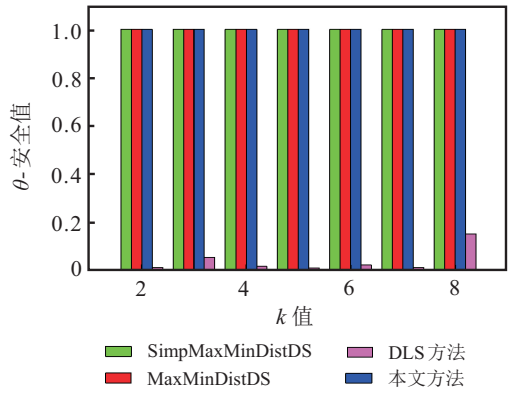


图 9　k位置的θ-安全集

由图 9 可以看出: 随着 k 值的增加, MaxMinDistDS 方法和 SimpMaxMinDistDS 方法的 θ 值基本没有发生变化, 始终无限接近于 1; 本文方法的 θ 也始终保持 1, 能够满足语义多样性的要求; 而 DLS 方法的 θ 值较小, 始终保持一个较小值. 这是因为 MaxMinDistDS 方法、SimpMaxMinDistDS 方法和本文方法都考虑了地理位置信息的语义多样性, 而 DLS 方法只考虑了假位置候选集中各位置点的查询概率, 而没有考虑各位置点可能具有相同语义信息特点的情况. 而且, 查询概率较大的位置点往往处于热点区域, 这些位置之间的语义信息非常相似, 从而具有更大的语义相似度. 因此, DLS 方法的 θ 值较小.

通过实验对比发现, 本文方法在尽可能满足位置点之间物理分散性和语义多样化的同时, 具有更高的假位置生成效率, 能有效提高位置服务质量.

4　结　论

针对目前广泛采用的基于假位置的 k-匿名位置隐私保护方法, 为了解决假位置生成效率低、预处理过程复杂、没有充分考虑地理语义信息特征等问题, 本文提出了一种基于近似匹配的假位置 k-匿名位置隐私保护方法. 该方法将空间位置坐标转换为二进制串, 通过快速匹配计算位置点之间的地理近似距离, 提高了假位置点选取效率, 并应用编辑距离计算位置点之间的语义相似度的方法, 简化了预处理过程. 实验结果显示, 该方法在满足假位置物理分散性和语义多样性的前提下, 缩短了假位置生成时间, 有效提高了位置隐私保护效果.

参考文献(References)

[1] Yu R, Kang J, Huang X, et al. MixGroup: Accumulative pseudonym exchanging for location privacy preservation in vehicular social networks[J]. IEEE Transactions on Dependable & Secure Computing, 2016, 13(1): 93-105.

[2] 张学军, 桂小林, 伍忠东. 位置服务隐私保护研究综述[J]. 软件学报, 2015, 26(9): 2373-2395.
(Zhang X J, Gui X L, Wu Z D. Privacy preservation for location-based services: A survey[J]. Journal of Software, 2015, 26(9): 2373-2395.)

[3] 倪巍伟, 马中希, 陈萧. 面向路网隐私保护连续近邻查询的安全区域构建[J]. 计算机学报, 2016, 39(3): 628-642.
(Ni W W, Ma Z X, Chen X. Safe region for privacy-preserving continuous nearest neighbor query on road networks[J]. Journal of Computer Science, 2016, 39(3): 628-642.)

[4] 易显天, 徐展, 郭承军, 等. 基于Patricia树的空间索引结构[J]. 计算机工程, 2015, 41(12): 69-74.
(Yi X T, Xu Z, Guo C J, et al. Spatial index structure

based on patricia tree[J]. Computer Engineering, 2015, 41(12): 69-74.)

[5] 朱进, 胡斌, 邵华. 基于多重运动特征的轨迹相似性度量模型[J]. 武汉大学学报: 信息科学版, 2017, 42(12): 1703-1710.
(Zhu J, Hu B, Shao H. Trajectory similarity measure based on multiple movement features[J]. Journal of Wuhan University: Information Science Edition, 2017, 42(12): 1703-1710.)

[6] Shokri R, Theodorakopoulos G, Papadimitratos P, et al. Hiding in the mobile crowd: Location privacy through collaboration[J]. IEEE Transactions on Dependable & Secure Computing, 2014, 11(3): 266-279.

[7] 徐建, 黄孝喜, 郭鸣, 等. 动态P2P网络中基于匿名链的位置隐私保护[J]. 浙江大学学报: 工学版, 2012, 46(4): 712-718.
(Xu J, Huang X X, Guo M, et al. Location privacy through anonymous chain in dynamic P2P network[J]. Journal of Zhejiang University: Engineering Edition, 2012, 46(4): 712-718.)

[8] Xu T, Cai Y. Feeling-based location privacy protection for location-based services[C]. Proceedings of the 2009 ACM Conference on Computer and Communications Security. Chicago: ACM, 2009: 348-357.

[9] Chow C Y, Mokbel M F, Liu X. A peer-to-peer spatial cloaking algorithm for anonymous location-based service[C]. Proceedings of the 14th ACM International Symposium on Geographic Information Systems. Arlington: DBLP, 2006: 171-178.

[10] 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[J]. 计算机学报, 2011, 34(10): 1976-1985.
(Huang Y, Huo Z, Meng X F. CoPrivacy: A collaborative location privacy-preserving method without cloaking region[J]. Chinese Journal of Computers, 2011, 34(10): 1976-1985.)

[11] Hwang R H, Hsueh Y L, Wu J J, et al. SocialHide: A distributed framework for location privacy protection[J]. Journal of Network & Computer Applications, 2016, 76(12): 87-100.

[12] Zhang H, Yu N, Wen Y. Mobile cloud computing based privacy protection in location-based information survey applications[J]. Security & Communication Networks, 2015, 8(6): 1006-1025.

[13] 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[J]. 计算机研究与发展, 2015, 52(11): 2628-2644.
(Zhou C L, Ma C G, Yang S T. Location privacy-preserving method for LBS continuous KNN query in road networks[J]. Journal of Computer Research and Development, 2015, 52(11): 2628-2644.)

[14] 李璐璐, 华佳烽, 万盛, 等. 基于高效信息缓存的位置隐私保护方案[J]. 通信学报, 2017, 38(6): 148-157.
(Li L L, Hua J F, Wan S, et al. Achieving efficient location privacy protection based on cache[J]. Journal on Communications, 2017, 38(6): 148-157.)

[15] Cheng R, Zhang Y, Bertino E, et al. Preserving user location privacy in mobile data management infrastructures[C]. International Workshop on Privacy Enhancing Technologies. Berlin, Heidelberg: Springer, 2006: 393-412.

[16] Li J, Yan H, Liu Z, et al. Location-sharing systems with enhanced privacy in mobile online social networks[J]. IEEE Systems Journal, 2015, 11(2): 436-448.

[17] Mouratidis K, Yiu M L. Shortest path computation with no information leakage[J]. Proceedings of the VLDB Endowment, 2012, 5(8): 692-703.

[18] Kim H I, Kim H J, Chang J W. A secure kNN query processing algorithm using homomorphic encryption on outsourced database[J]. Data & Knowledge Engineering, DOI: 10.1016/j.datak.2017.07.005.

[19] Sweeney L. k-anonymity: A model for protecting privacy[J]. International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, 2012, 10(5): 557-570.

[20] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking[C]. Proceedings of the 1st International Conference on Mobile Systems, Applications, and Services. San Francisco: DBLP, 2003: 31-42.

[21] Bamba B, Liu L, Pesti P, et al. Supporting anonymous location queries in mobile environments with privacy grid[C]. Proceedings of the 17th International World Wide Web Conference. Beijing: ACM, 2008: 237-246.

[22] Xu J, Tang X, Hu H, et al. Privacy-conscious location-based queries in mobile environments[J]. IEEE Transactions on Parallel & Distributed Systems, 2010, 21(3): 313-326.

[23] 裴卓雄, 李兴华, 刘海, 等. LBS隐私保护中基于查询范围的匿名区构造方案[J]. 通信学报, 2017, 38(9): 125-132.
(Pei Z X, Li X H, Liu H, et al. Anonymizing region construction scheme based on query range in location-based service privacy protection[J]. Journal on Communications, 2017, 38(9): 125-132.)

[24] 杨洋, 王汝传. 增强现实中基于LBS的矩形区域K-匿名位置隐私保护方法[J]. 南京师范大学学报:自然科学版, 2016, 39(4): 44-49.
(Yang Y, Wang R C. Rectangular region K-anonymity

location privacy protection based on LBS in Augmented Reality[J]. Journal of Nanjing Normal University: Natural Science, 2016, 39(4): 44-49.)

[25] Xie P. A-anonymous polygon area construction method and algorithm based on LBS privacy protecting[J]. Journal of Information & Computational Science, 2015, 12(15): 5713-5724.

[26] Yin C, Sun R, Xi J. Location privacy protection based on improved K-value method in augmented reality on mobile devices[J]. Mobile Information Systems, DOI: 10.1155/2017/7251395.

[27] Kido H, Yanagisawa Y, Satoh T. An anonymous communication technique using dummies for location-based services[C]. Proceedings of 1st International Workshop on Security, Privacy and Trust in Pervasive and Ubiquitous Computing. Santorini: IEEE, 2005: 88-97.

[28] Kido H, Yanagisawa Y, Satoh T. Protection of location privacy using dummies for location-based services[C]. Proceedings of the 21st International Conference on Data Engineering Workshops. Tokyo: IEEE, 2005: 1248-1252.

[29] Lu H, Jensen C S, Man L Y. PAD: Privacy-area aware, dummy-based location privacy in mobile services[C]. Proceedings of the 7th ACM International Workshop on Data Engineering for Wireless and Mobile Access. Vancouver: DBLP, 2008: 16-23.

[30] You T H, Peng W C, Lee W C. Protecting moving trajectories with dummies[C]. Proceedings of the 9th International Conference on Mobile Data Management. Beijing: IEEE, 2008: 278-282.

[31] Xu T, Cai Y. Exploring historical location data for anonymity preservation in location-based services[C]. Proceedings of the 27th Conference on Computer Communications. Alabama: IEEE, 2008: 547-555.

[32] 刘海, 李兴华, 王二蒙, 等. 连续服务请求下基于假位置的用户隐私增强方法[J]. 通信学报, 2016, 37(7): 140-150.
(Liu H, Li X H, Wang E M, et al. Privacy enhancing method for dummy-based privacy protection with continuous location-based service queries[J]. Journal on Communications, 2016, 37(7): 140-150.)

[33] Niu B, Li Q, Zhu X, et al. Achieving k-anonymity in privacy-aware location-based services[C]. Proceedings of the 33rd Annual IEEE International Conference on Computer Communications (INFOCOM'14). Toronto: IEEE, 2014: 754-762.

[34] Niu B, Li Q, Zhu X, et al. Enhancing privacy through caching in location-based services[C]. Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM). Hong Kong: IEEE, 2015: 1017-1025.

[35] Niu B, Gao S, Li F, et al. Protection of location privacy in continuous LBSs against adversaries with background information[C]. Proceedings of the 2016 3rd International Conference on Computing, Networking and Communications (ICNC). Sanya: IEEE, 2016: 1-6.

[36] Sun Y, Chen M, Hu L, et al. ASA: Against statistical attacks for privacy-aware users in location based service[J]. Future Generation Computer Systems, 2016, 70(5): 48-58.

[37] Chen S, Shen H. Semantic-aware dummy selection for location privacy preservation[C]. Proceedings of the 16th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. Sydney: IEEE, 2017: 752-759.

作者简介

张永兵 (1978–), 男, 讲师, 博士生, 从事网络与信息安全、隐私保护等研究, E-mail: gstszyb@163.com;

张秋余 (1966–), 男, 研究员, 博士生导师, 从事网络与信息安全、智能信息处理与模式识别、企业信息化系统与工程等研究, E-mail: zhangqylz@163.com;

李宗义 (1960–), 男, 教授, 硕士, 从事企业信息化系统与工程、先进制造等研究, E-mail: gsjdlzy@163.com;

段宏湘 (1974–), 女, 讲师, 博士生, 从事智能信息处理与模式识别、数据可视化等研究, E-mail: duanhongx@sohu.com;

张墨逸 (1985–), 女, 讲师, 博士, 从事人机交互、大数据隐私保护等研究, E-mail: zhangmoyi_1985@163.com.

(责任编辑: 李君玲)