基于 SpaceTwist 的 K-匿名增量近邻查询位置隐私保护算法*

胡德敏[a,b], 郑 霞[a]
(上海理工大学 a. 光电信息与计算机工程学院; b. 计算机软件技术研究所, 上海 200093)

摘 要: 随着移动网络的持续进步,基于位置的服务在日常生活中被广泛应用,同时位置隐私保护也成为广大用户所关注的焦点。基于 SpaceTwist 算法和 K-匿名算法,结合路网环境提出一种新的位置隐私保护方法。该方法摆脱第三方可信匿名器,采用客户—服务器体系结构,根据用户的位置隐私需求结合用户所在路网环境设计出用户端匿名区生成算法,并且保证 K-匿名。用户端以该匿名区请求基于位置的服务,服务器根据用户请求返回检索点并满足用户期望的 K 近邻结果。根据不同的路网环境和用户隐私需求进行大量实验,证明该算法在满足用户基于位置服务需求的同时提高了对用户位置隐私的保护。

关键词: SpaceTwist 算法; K-匿名; 基于位置服务; 路网环境; 位置隐私; k 近邻结果
中图分类号: TP309.2,TP301.6 **文献标志码**: A **文章编号**: 1001-3695(2016)08-2402-03
doi:10.3969/j.issn.1001-3695.2016.08.035

SpaceTwist-based K-anonymity incremental nearest neighbor query algorithm for location privacy protection

Hu Demin[a,b], Zheng Xia[a]
(*a. School of Optical-Electrical & Computer Engineering, b. Institute of Computer Technology, University of Shanghai for Science & Technology, Shanghai* 200093, *China*)

Abstract: With continued advances in mobile Internet, the Location-based services are widely used in our daily life, more and more users focus on the location privacy protection. This paper proposed a new location privacy protection scheme which combined with the network environment and based on the SpaceTwist algorithm and K-anonymity algorithm. The approach prefers to client-server structure rather than trusted third party server and designed the anonymity algorithm in user end according to user's request of location privacy and the network environment, which ensures the K-anonymity. The mobile end users can request the location-based services in its anonymous area, and the server will return the access points to meet user's expect K-value . Finally, the proposed approach was evaluated by a large number of experiments according to different network environment together with different users privacy needs ,the results demonstrate this algorithm meets the user's requests of location-based services and at the same time it improves the protection of user's location privacy.

Key words: SpaceTwist algorithm; K-anonymous; location-base service; road network environment; location privacy; K-nearest neighbors result

0 引言

随着全球定位技术和无线通信技术的进步给基于位置的服务(location based service,LBS)创造了更大的发展空间。基于位置的服务是指通过无线电通信网络(如 GSM 网、CDMA 网)和外部定位方式(如 GPS)获取移动端用户的位置信息,并根据用户的位置提供相关的服务。通常情况下,用户是非常享用 LBS 提供的便携服务。例如:一个移动端用户在并不太熟悉的环境下寻找附近的银行、餐厅、宾馆等等。LBS 可以根据请求者的位置信息和请求内容,非常快捷地帮助用户解决这些难题。但前提是用户必须向 LBS 服务器提供自己精确的位置信息,但在交换服务过程中用户位置隐私极其容易受到严重威胁。

目前位置隐私越来越受到人们的关注,随即位置隐私保护方法也得到广泛研究。常见的位置隐私保护方法主要有[1]假位置(dummy)、时空伪装(spatial temporal cloaking)、空间加密(space encryption)。假位置是指用户在向 LBS 请求服务时,发送的不是用户的真实位置,而是以一个代理用户的位置请求 LBS 服务。服务质量与代理用户的位置和用户真实位置之间的距离成反比;时空伪装方法[2]是指以一个匿名区域来替代用户的真实位置;空间加密指请求 LBS 服务器之前采用某种加密协议对用户的真实位置加密。

1 相关工作

在用户位置隐私保护中分为三种体系结构[3]:可信第三

收稿日期: 2015-06-02; **修回日期**: 2015-07-19 **基金项目**: 国家自然科学基金资助项目(61170277,61472256);上海市教委科研创新重点项目(12zz137);上海市一流学科建设项目(S1201YLXK)

作者简介:胡德敏(1963-),男,上海杨浦人,副教授,博士,主要研究方向为计算机网络、分布式计算、云计算;郑霞(1990-),女,山东莱西人,~~硕士研究生~~,主要研究方向为位置隐私保护(332940966@qq.com).

方体系结构、客户—服务器体系结构、分布式体系结构。其中 K-匿名算法[4]采用可信第三方体系结构,主要技术是时空伪装,用户向 LBS 服务器请求服务时确保用户真实位置与其他 $K-1$ 个用户位置是不可区分的,用户的匿名区是在第三方可信匿名器形成,并由匿名器发送给 LBS 服务器请求服务。

文献[5,6]采用分布式体系结构,文献[5]提出了基于分层的希伯特的 peer-to-peer 结构,该方法基于希尔伯特曲线进行分组,并且让每一个组的第一个用户为代理发起请求;文献[6]采用加密技术,只让用户知道是否满足 K-匿名,而不知 K 的具体大小,从而决定是否发送请求。

相对于 K-匿名算法采用第三方体系结构,SpaceTwist 算法[7]摆脱第三方可信匿名器,采用客户机—服务器结构,算法的基本思想如下:移动端用户随机选择用户真实位置附近的一个锚点替代用户真实位置向 LBS 服务器发送请求。初始化需求空间和供应空间,需求空间是以用户的真实位置为圆心,以 LBS 服务器返回检索点中距用户真实位置最近的检索点到用户真实位置的距离(初始值为∞)为半径的圆。供应空间为以锚点为圆心,以 LBS 服务器返回的检索点距锚点最远的距离为半径(初始值为 0)的圆。用户端以锚点的位置向 LBS 服务器发送请求,LBS 服务器根据锚点位置返回查询结果,用户端根据 LBS 服务器返回的检索点更新需求空间和供应空间,然后判断供应空间是否完全覆盖需求空间。如果是,则服务器停止检索,否则 LBS 服务器继续检索,直至供应空间完全覆盖需求空间为止。

K-匿名算法是用户将自己的精确位置发送给第三方可信匿名器,根据用户需求在匿名器中形成匿名区。但在实际生活中,并不可能存在完全可靠的第三方匿名器,一旦第三方可信匿名器被黑客攻击,用户的位置隐私将会受到严重的威胁。SpaceTwist 算法虽然摆脱了第三方可信匿名器,提高了用户位置隐私保护,但不能保证 K-匿名。该算法是在用户真实位置附近随机寻找一个锚点代替用户发送请求,但用户所在地区的用户密度是不确定的,如果用户所在地点恰巧有且只有一个用户(除代替用户真实位置的锚点),那攻击者知道用户的真实位置的概率为 1/2,用户位置隐私很容易泄露。LSB 服务器返回结果的数量与查询点到用户位置的距离有关,距离越远时,供应范围会随之增加,返回的检索点也会越来越多,如果查询点距离用户很近时,LBS 服务器返回的检索点不能满足用户期望的 K 近邻结果。

为了解决上述问题,本文参照路网密度,结合用户所在地区人口密度实际情况,根据用户的位置隐私需求对用户真实位置形成 K-匿名区,以该匿名区发送到 LBS 服务器请求服务,LBS 服务器根据用户的请求返回检索点,直至返回的检索满足用户期望的 K 近邻结果。这样不仅提高了用户位置隐私保护而且满足了用户期望的 K 近邻结果。通常路网密集区域,用户的流量比较大,反之,用户流量比较小。目前 LBS 分为快照 LBS 和连续 LBS,本文只考虑基于快照 LBS 的用户位置隐私保护。

2　基于 SpaceTwist 的 K-匿名增量近邻查询算法

2.1　相关定义

定义 1(用户位置 Location)　用户在某一时刻的位置可以用 Location = $\{X,Y,T\}$ 表示,其中 X 代表经度,Y 代表纬度,T 表示用户所处当前位置的时间。

定义 2(位置隐私需求 Qc)　用户位置隐私需求 $Qc = \{\text{id}, \text{Location}, K, A_{\min}, \}$ 其中 id 是用户的唯一标志符(其中真实位置只有用户端自己知道);K 代表用户自定义的位置隐私保护程度,指用户端形成的匿名区最少包含用户真实位置和 $K-1$ 个其他用户位置;$A_{\min}$ 指最小匿名区的面积,是为了保证当用户所在路网密集处,K-匿名所形成的匿名区非常小时,以 $A_{\min}$ 作为最小匿名区保护用户位置隐私。

定义 3(查询请求 Qa)请求 LBS 服务器 $Qa = \{\text{id}, \text{K-ASR}, \text{content}, \text{目标数 } K_{\text{result}}\}$ 其中 id 是该用户请求的唯一标志符合,K-ASR 是指用户端根据匿名区生成算法形成的 K-匿名区(具体算法见 2.3 节),目标数 K_{result} 指用户期望的 K 近邻查询结果。

定义 4(路网粒度 λ)　设 G 为平面空间区域中的任意一个子空间区域,L 为 G 的周长,S 为 G 内道路总长度,则定义 $\lambda = S/L$ 为 G 的路网粒度。

2.2　系统结构

K-匿名算法是移动端以一定的频率将自己的真实位置传递给第三方可信匿名器,第三方可信匿名器根据用户的真实位置搜索 K 个用户的位置(其中一个是用户的真实位置),构成匿名区发送给 LBS 服务器请求服务[8]。但在实际生活中,并不可能存在完全可靠的第三方匿名器,一旦第三方可信匿名器被攻击,用户的位置隐私就会被泄露[7]。为了防止这种情况,本文采用客户—服务器体系结构,用户不直接发送自己的精确位置到 LBS 服务器,而是根据自己所在路网环境和位置隐私需求形成匿名区(具体算法见第 2.3 节)发送给 LBS 服务器,使用户的位置隐私保护得到进一步提升。

本文提出的算法是基于客户—服务器体系结构,整个体系结构主要包括移动设备、通信网络、服务与内容提供商(LBS 服务器)。具体如图 1 所示,其中移动终端具备基本的定位功能,例如:手机 GPS、车载 GPS 等。

2.3　用户端匿名区的生成算法

SpaceTwist 算法是在用户真实位置附近随机寻找一个锚点代替用户发送请求,但用户所在地区的用户密度是不确定的,如果用户所在地点恰巧有且只有一个用户(除代替用户真实位置的锚点),那攻击者很容易获取用户位置。为了避免上述情况,用户端匿名区生成算法结合路网环境根据用户隐私需求形成的 K-匿名区。

1)移动端用户获取自己当前时刻的真实位置 Location = $\{X,Y,T\}$。

2)用户端根据自己所在经纬度,确定用户所在城市,该城市即是用户所在区域 G。

3)用户在请求 LBS 服务之前,根据自己所在区域 G,向 LBS 服务器请求获取该区域的分割结果,LBS 服务器以十字递进形式对 G 分割,将用户所在区域 G 分割成大小相同的正方形,其中每个单元格的边长不大于给定的 λ。每个被分割的单元格以四叉树的形式保存,四叉树的根节点代表用户所有区 G,每个非叶子节点都有四个孩子。每个节点对应一个单元格。每个节点数据域中存储对应的单元格信息 $\text{Cell}_{\text{info}} = [\text{id}, \text{中心点 } O, \text{单元格的边长 } L_{\text{cell}}, \text{该单元格内道路的长度 } R_L]$。同时存储每个单元格对应的用户数量 $\text{Mount}_{\text{user}} = \{\text{id}, \text{time}, \text{mount}\}$。其中 id 用来标志对应的单元格。mount 代表 time 这个时间点时,当前单元格中用户的数量,每个单元格中的用户数量信息以一定的频率实时更新。

4)用户所在的单元格的外接圆为匿名区 K-ASR。查看该匿名区的用户数是否不小于 K,如果不小于 K,则以该单元格的外接圆作为用户请求 LBS 服务的匿名区。如果用户所在单元格的用户数小于 K 时,取该单元格的四周单元格中移动对象最少的单元格。然后根据该单元格中的路网粒度 λ 和移动对象数,估算出用户所在单元格的四周单元格每个单位面积拥有的最低移动对象数即人口密度因子 α,然后匿名区根据 K 减去当前单元格的移动对象,扩大相应的面积。最后查看匿名区是否满足用户要求的最小匿名区 A_{min}。如果不满足,则扩大匿名区的半径,直至匿名区不小于 A_{min}。

2.4 基于 SpaceTwist 的 K-匿名增量近邻查询算法

输入:匿名区的中心坐标 O 及半径 L,查询内容 content,目标数 K_{result}。

输出:查询结果集 result。

```
1. result←new max-heap{p,dist(O,p)};
2. R←L;//需求区半径
3. r←0;//供应区半径
4. Count=0;//初始化 Count 的值
5. Send an INN query with Qa={id,K-ASR,content,目标数 K_result};
6. While (R+dist(q,p)>r) do
7.     s←get the next pack of point from LBS service;
8.     r←max_{p∈s} dist(q,p);
9.     Update supplement;
10.    Update count;
11.        If(dist(q,p)>=dist(q,O)){
12.                Check the count;
13.                If(count >= K_result){
14.                    r←dist(q,p);
15.            Insert into result {p,dist(O,p)};}
16.             Else {
17.                    Expand the radius R;}
18. Terminate the INN query at the service;
19. Return result;
```

基于 SpaceTwist 的 K-匿名增量近邻查询算法是移动端用户首先获取自己的位置 Location(X,Y,T),然后根据用户端匿名区生成算法,形成 K-匿名区 K-ASR,以 K-ASR 作为用户位置向 LBS 服务器发出请求 $Qa=\{$id, K-ASR, content, 目标数 $K_{result}\}$,初始化的需求区是以匿名区的中心点 O 为圆心,L 为半径的圆。以用户真实位置 q 为圆心 r←0 为半径的圆为供应区。同时用初始化一个变量 count(初始值为 0)用来记录供应区检索点的数量。服务器根据用户请求的内容,LBS 服务器把检索到的检索点进行打包返回。客户端根据 LBS 服务器返回的检索点$\{p_1,p_2,\cdots,p_n\}$,计算目前检索到的检索点到用户真实位置 q 最远的距离为最新供应区半径,然后更新供应区。服务器继续以增量检索的方式返回,直至供应区完全覆盖需求区,然后查看 count 是否不小于目标数 K_{result},如果 count 小于目标数 K_{result},则用户端通知 LBS 服务器扩大匿名区半径,LBS 服务器继续检索,直至供应区完全覆盖需求区并且检索点的数量不小于 K_{result}。

3 实验结果与分析

本次实验采用 Thomas Brinkhoff 路网数据生成器[9]并分别在 Oldenburg 城市的交通路网和 San Joaquin 城市的交通路网生成移动对象数据。为了对比不同路网密度下对匿名区的影响,实验分别以 Oldenburg 城市的交通路网(以下简称 OB)和 San Joaquin 城市的交通路网(以下简称 SJ)分别生成移动对象。验证本文方法有效性的实验参数如表 1 所示。

表 1 在不同路网密度下实验参数对比

实验参数	OB	SJ
区域宽度/m	23 572	656 570
区域高度/m	26 915	716 904
路网密度 λ	139.3	17.6
生成移动对象个数	18 331	43 826

为了实验容易实施,假定移动对象和兴趣点与道路长度成正比。随机选择 200 个用户的位置为查询用户的位置,每个对象的隐私需求随机生成,其中 $K\in[5,10]$,$A_{min}=0.03$ Km^2,在不同路网密度下,计算用户匿名区的大小。图 2 显示的是不同路网密度下生成的匿名区,由图 2 可以看出用户匿名区的大小与路网密度之间的关系,路网密度越大时,用户的匿名区越小;否则,反之。

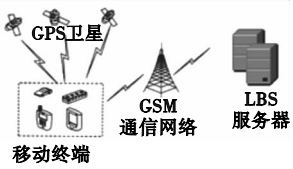

图1 体系结构图

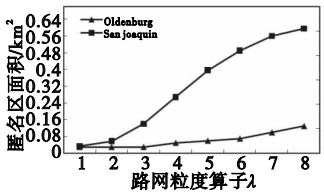

图2 不同路网粒度形成的匿名区对比

本文算法相对于 SpaceTwist 算法增加了匿名区的形成以及增加一个 count 变量统计供应区中检索点的个数是否满足用户期望的 K 近邻查询结果,该方法不仅实现了 K-匿名保护用户的位置隐私,并且满足了用户期望的 K 近邻结果。为了检验该方法的有效性,本次实验使用相同的数据集合(基于 Oldenburg 城市的交通路网产生的移动对象),实验将 SpaceTwist 算法查询点与用户位置之间的距离(单位为 m)分别为:50、100、200、300、400。分别通过查询准确率和响应时间两个方面与 Spac-Twist 算法进行比较。其中查询准确率是指算法匿名区中用户期待的 K 近邻结果占最终 LBS 服务器返回结果的比例。本文算法与 SpaceTwist 算法查询准确度对比如图 3 所示。

由图 3 可以看出,本文算法与 SpaceTwist 算法相比,准确度有明显的提高。本文算法的查询准确率基本维持在 86% ~ 93%,而 SpaceTwist 算法随着查询点距离用户真实位置的距离的增加,准确率呈现稳步上升稳定的趋势,但仍然未达到本文算法的平均准确率。

响应时间是指用户发送查询请求到 LBS 服务器,LBS 服务器返回的检索结果满足用户需求的时间。对比结果如图 4 所示。

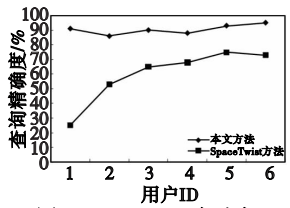

图3 SpaceTwist方法与本文方法查询准确度对比

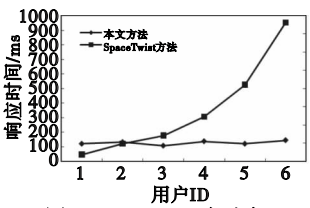

图4 SpaceTwist方法与本文方法响应时间对比

由图 4 可以看出,SpaceTwist 方法查询点距用户位置越远时,供应范围会随之增加,返回的检索点也会越来越多,以至于响应时间也会越来越长。本文方法相对于 SpaceTwist 方法增加了用户端匿名区的形成和检查 LBS 服务器返回的检索点是否满足用户期望的 K_{result} 近邻结果,虽然比 SpaceTist (下转第 页)

(上接第 页)算法当选择锚点距用户真实位置很近时的响应时间长,但响应时间相对稳定,并都在用户可接受的范围内。

通过上述查询准确率和响应时间两个方面的对比,可以看出本文算法与 SpaceTwist 算法相比不仅响应时间相对稳定,而且可以有效地提高查询准确率,从而提高了用户的查询服务质量的保障。

4 结束语

本文算法相对于 K-匿名算法,不再采用第三方可信匿名器,而是采用客户—服务器体系结构来防止黑客通过攻击第三方可信匿名器获取用户位置信息。与 SpaceTwist 算法相比,并不是随机选择用户附近的一个锚点作为代理用户,而是根据用户所在路网环境以及用户位置隐私需求在客户端形成 K-匿名区,使用户位置隐私保护得到进一步提高,并且 LBS 服务器返回结果满足用户期望的 K 近邻查询结果。基于个性化隐私保护[10]是本文继续研究的方向。

参考文献:

[1] Pan Xiao, Xiao Zhen, Meng Xiaofeng. Survey of location privacy-preserving[J]. Journal of Computer Science and Frontiers, 2007, 1(3): 268-281.

[2] 韩建民,林瑜,于娟,等. 基于位置 K-匿名的 LBS 隐私保护方法的研究[J]. 小型微型计算机系统, 2014, 35(9): 2088.

[3] Gruteser M, Grunw ald D. Anonymous usage of location-based privacy using dummies for location-based service[C]//Proc of the 21st International Conference on Mobile Systems, Application and Servi-ces. [S. l.]: IEEE Press, 2003: 31-42.

[4] Gruteser M, Grunwald D. Anonymous Usage of location-based services through spatial a temporal cloaking[C]//Proc of International Conference on Mobile Systems, Applications, and Services. New York: ACM Press, 2003: 163-168.

[5] Konjevod G, Richa A W, Xia Donglin. On sampling in higher-dimensional peer-to-peer system[M]//Theoretical Informatics. Berlin Heridelberg: Springer, 2006: 641-652.

[6] Takabi H, Joshi J B D, Karimi H A. A collaborative K-anonymity approach for location privacy in locat-ion privacy in location-based services[C]//Proc of the 5th International Conference Application and Worksharing. [S. l.]: IEEE Press, 2009: 1-9.

[7] Yiu M L, Jensen C S, Huang Xuegang, *et al*. SpaceTwist: managing the trade-offs among location privacy, query performance, and query accuracy in mobile services[C]//Proc of the 24th IEEE International Conference on Data Engineering. [S. l.]: IEEE Press, 2008: 366-375.

[8] Wang Song, Wang X S. AnonTwist: nearest neighbor querying with both location privacy and K-anonymity for mobile users[C]//Proc of the 10th International Conference on Mobile Data Management, Services and Middleware, [S. l.]: IEEE Press, 2009: 443-448.

[9] Brinkhoff T. A framework for generating network-based moving object [J]. GeoInformation, 2002, 6(2): 153-180.

[10] Gedik B, Liu Ling. Protecting location privacy with personalized K-anonymity: architecture and algorithms[J]. IEEE Trans on Mobile Computing, 2008, 7(1): 1-18.