doi:10.3969/j.issn.1001-4616.2016.04.009

增强现实中基于 LBS 的矩形区域 K-匿名位置隐私保护方法

杨　洋[1],王汝传[2,3]

(1.南京广播电视大学,南京城市职业学院信息技术系 江苏 南京 210002)
(2.南京邮电大学计算机学院,江苏 南京 210003)
(3.江苏省无线传感网络高技术研究重点实验室,江苏 南京 210003)

[摘要]　基于位置服务(LBS)和增强现实技术快速发展的同时,促进了基于位置服务的应用范围扩大,同时也带来了用户位置隐私泄露的隐患. 因此,如何确保基于位置服务中数据的安全性,成为该项技术推广应用的关键问题. 本文借助 k-匿名法,提出矩形区域 k-匿名法,将 k-匿名法的理念引入该方法中,实验结果表明该方法提高了相对匿名度和匿名区域面积,从而有效地保护了用户的位置隐私.

[关键词]　基于位置服务,位置隐私,k-匿名法,矩形区域 k-匿名法
[中图分类号]TP391.9　[文献标志码]A　[文章编号]1001-4616(2016)04-0044-06

Rectangular Region K-Anonymity Location Privacy Protection Based on LBS in Augmented Reality

Yang Yang[1],Wang Ruchuan[2,3]

(1.Department of Information Technology,Nanjing Radio and TV University,Nanjing City Vocational College,Nanjing 210002,China)
(2.College of Computer,Nanjing University of Posts and Telecommunications,Nanjing 210003,China)
(3.Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks,Jiangsu Province,Nanjing 210003,China)

Abstract:Rapid development of location based service(LBS) and augmented reality promote application of LBS,which also bring hidden danger of user location privacy disclosure. Therefore,how to ensure data security becomes key question in application of LBS. Here we introduced k-anonymity into privacy protection and proposed the rectangular region k-anonymity. Our results revealed that rectangular region k-anonymity enhanced relative anonymous degree and anonymous area,which could effectively protect user's location privacy.

Key words:LBS,location privacy,k-anonymity,rectangular region k-anonymity

增强现实是通过计算机系统将虚拟的物体或信息等实时叠加到用户周围真实的场景中,从而达到了对现实的"增强"[1]. LBS 是增强现实提供的常用服务之一,在 LBS 中移动终端用户的位置信息是通过移动的无线网络或外部定位方式来获取的,位置定位技术(Location Determination Technology,LDT)是获取用户位置信息的一种较好的技术,如 GPS、高级时差检测定位技术(Enhanced Observed Time Difference,EOTD)等,可以给出含有 X-Y 坐标的用户位置信息. 当移动终端用户向服务器发出服务请求时,用户通过位置定位技术获取位置信息并发送给服务器,服务器根据接收到的用户位置信息来处理该用户的服务请求,并将处理结果反馈给移动终端用户. LBS 服务含有多项内容,例如,紧急服务(如查找最近的医院等)、信息和娱乐服务(如查找附近最便宜的餐馆或电影院)、广告业务服务(如商家送优惠券给顾客)、定位服务(如收集亲朋好友现在的位置信息)等等. LBS 是通过用户的位置信息来增加服务价值的移动服务,随着增强现实和 LBS 的发展,LBS 的应用范围也在不断扩大,它可以为人们提供各种各样的位置服务,但是

收稿日期:2016-08-26.
基金项目:国家自然科学基金(60973139、61170065、61171053)、江苏省自然科学基金(BK2011755)、江苏省科技支撑计划项目(BE2010197、BE2010198、BE2011844、BE2011189).
通讯联系人:杨洋,副教授,研究方向:网格计算、增强现实技术、位置隐私保护等. E-mail:nj.Yangyang@163.com

这一迅猛发展的技术也带来显著的威胁,其中服务内容的泄露和位置隐私的泄露成为 LBS 用户的重要威胁. 为解决隐私泄露造成的威胁,相关研究也陆续出现:其中文献[2]介绍了 LBS 系统的关键技术及其研究进展,对现有的位置隐私保护方法进行了分析;文献[3]提出了基于位置 k-匿名的 LBS 隐私保护方法;文献[4]提出了基于敏感位置多样性的 LBS 位置隐私保护方法.

已有的位置隐私保护方法和系统从不同角度解决了 LBS 位置隐私保护的问题,但仍然存在一些难题,比如:一些隐匿区域可能存在过多的无效区域,增加了服务器的计算开销,也会导致无效的返回结果;网格状的划分不考虑地形和布局等隐私,导致位置信息的精度损失过大.

1　相关工作

现有的位置隐私保护技术研究划分为 3 类[5]. 第一类为假名技术[6],对于 LBS 请求,用户利用可信的中间件生成一个可替代用户真实身份标识的信息,发送给 LBS 提供商从而来保护用户的隐私;第二类是基于信息检索的技术[6],通过将位置信息加密从而保护用户的隐私;第三类是基于位置匿名的技术[6],将用户的位置扩大为含有该位置的一个区域,从而模糊用户的位置信息,来代替该点查询.

2003 年,Gruteser 和 Grunwald[7]引入了 k-匿名法进行隐私保护研究. k-匿名法[8-10]用于解决攻击方通过链接操作发布的数据从而推理出隐私数据泄露的问题. 用户在向 LBS 提供商提交服务请求前,先删除个人信息内容,发布的数据中存在一定数量(至少为 k 个)不可区分的个体,使得各条记录至少与数据表中其他 $k-1$ 条记录具有完全相同的准标识符属性值,从而保护个体的隐私. 但这种方法的局限性是不能抵制同质性攻击和背景知识攻击[11],攻击者很容易推断出个体相应的敏感属性数据,或者可以通过背景知识确定敏感属性数据和个体之间的对应关系,从而导致隐私泄露,而且这种方法假设 k 是一个统一的值,因此不能满足用户的个性化隐私需求.

CliqueCloak[12]提出了满足用户个性化隐私需求的位置 k-匿名方法,用户可以定义自己的隐私要求 k,可以调整其最小匿名等级及可忍受的最大时间、空间分辨率. 在该方法中,每个用户可以自定义其所需的匿名等级,通过为每个用户指定不同的敏感属性泛化约束来实现个性化匿名,减少因统一匿名化所带来的信息损失. 该方法的缺陷是 k 值不能过大,当 k 值较大时实际效果很差.

为解决 k-匿名法由于同质性攻击和背景知识攻击所带来的隐私泄露,研究者基于该领域还提出了敏感属性多样性模型. Machanavajjhala A 等[13]在 k-匿名法的基础上提出了 l-多样性模型,它通过要求匿名化的数据记录中出现频率最高的敏感属性数据的个数不大于 $1/l$,来提高敏感属性数据与其所属个体的链接难度,从而防止因敏感属性缺少多样性而可能导致的隐私泄露. 但该模型的缺陷是只适合处理分类型敏感属性数据,不适合处理数值型敏感属性数据.

从位置隐私保护方法的进展来看,位置信息的精确度引起的隐私保护安全和查询服务质量之间的矛盾,是位置服务的固有特性,如何平衡这一矛盾是很值得研究的一个问题. 本文利用 k-匿名位置隐私保护方法的思想,提出矩形区域 k-匿名法,该方法相对于网格划分方法缩小了匿名区域的面积,解决隐私保护安全和查询服务质量之间的矛盾问题. 该方法提高了位置服务的准确性和用户体验,在隐私保护方面,为用户提供更好的隐藏保护功能,实现较好的隐私保护能力.

2　矩形区域 K-匿名位置保护模型

位置信息经过数据挖掘后可以获得许多有价值的信息,比如哪个餐厅被查询的次数最多、哪里最需要医院等. 但在 LBS 的实际使用中,用户的实际需求趋向于隐匿区域和实际区域地形相吻合,比如用户希望隐藏于某条街道或者某个商场内,这种相吻合的隐匿区域最大程度地保留了位置信息的精确度,同时满足了用户的隐私要求. 本文提出的矩形区域 k-匿名位置隐私保护方法所生成的隐匿区域可以针对现有位置隐私保护方法的不足,将位置泛化,把用户所在的位置坐标泛化为一个矩形的隐匿区域,使得用户的位置很好地隐藏在这个区域中.

本文提出矩形区域 k-匿名位置隐私保护方法,假设矩形区域中的用户不互相信任,用户是采用矩形区域来代替自己的精确位置并与其他用户分享.

2.1　准备知识

为了便于叙述,我们对相关概念作如下定义:

定义 1　令 R_S 表示 1 个 S 区域,已知 R_S 是 1 个矩形. R_S 可以定义为

$$R_S=(x_{ua},y_{ua},x_{vb},y_{vb},uid),$$

其中坐标(x_{ua},y_{ua})和坐标(x_{vb},y_{vb})分别表示 S 区域的右上角点和左下角点,uid 表示用户位置的标识.

定义 2　令攻击者为 A,若用户 U 所在的区域 S 满足以下条件:(1)区域 S 至少覆盖了 $k-1$ 个其他用户;(2)$k-1$ 个其他用户的位置均匀分布,对 A 说不可区分,则区域 S 为位置隐私保护安全区域. 我们将此 k-匿名模型记作$\{U,A,S\}$.

定义 3　由定义 1 可知,S 区域面积 $S(R_S)=|x_{ua}-x_{vb}|\times|y_{ua}-y_{vb}|$,设用户可接受的最大面积为 $S_{\max}$,可接受的最小面积为 $S_{\min}$.

$S(R_s)$的值影响着用户隐私保护程度和位置服务质量. 当 $S(R_s)\leqslant S_{\min}$,匿名区域可能是用户 U 的精确位置点,攻击者 A 便可通过发送来的查询请求直接判断出用户 U 的位置;当 $S(R_s)\geqslant S_{\max}$,匿名区域过大通常使得服务结果集较大,数据库的检索兴趣点增多,LBS 称为攻击者 A 的攻击对象,查询请求的结果精确度降低,耗费用户 U 过多的时间和资源. 因此我们需设置合适的 $S_{\max}$ 和 $S_{\min}$ 来权衡隐私保护程度和位置服务质量.

定义 4　已知用户 U 的 S 区域为 R_S,$N(R_s)$表示 S 区域里所包含用户的数量,$S(R_s)$表示 S 区域的面积,其中 $N(R_s)$决定服务器的匿名程度和 $S(R_s)$大小,我们将 $N(R_s)$称为安全系数. 如在 k-匿名模型中,k 即为安全系数.

定义 5　已知用户 U 的 S 区域为 R_S,我们设在 S 区域中的用户分布密度为 ω,则

$$\omega=\frac{N(R_S)}{S(R_S)},$$

式中,$N(R_s)$表示 S 区域里所包含用户的数量,$S(R_s)$表示 S 区域的面积. 在 k-匿名模型的实际区域中 $N(R_s)\geqslant k$,因此当 $S(R_S)\in[S_{\min},S_{\max}]$时,$\omega=\dfrac{k}{S(R_s)}$;当 $S(R_s)\notin[S_{\min},S_{\max}]$时,$\omega=\dfrac{k}{\gamma\cdot S(R_s)}$,其中 γ 为面积调控系数,并且 $\gamma\in(0,1]$.

定义 6　当 $S(R_S)\in[S_{\min},S_{\max}]$,并且 ω 在此面积范围内达到最大值,则区域 S 称为最佳区域.

2.2　算法描述

该算法是基于矩形区域的,即用户都被放置到矩形区域的单元格中,我们把这个矩形区域称之为 S.

其中,d_0 代表 LBS 位置的精度,这里我们设置一个初始值为 5 m,d_1 代表 $k-1$ 所对应的两个点之间的距离,d_2 代表当前 k 所对应的两个点之间的距离. 我们将用户的真实位置和其定义的保密级别作为该算法的输入,并计算出真实位置和虚假位置之间的距离,用户通过移动终端设备加入到移动网络中,其可以通过移动网络自动搜索到相邻用户,并进行区域分享,或者替代精确位置发送给服务器提供商.

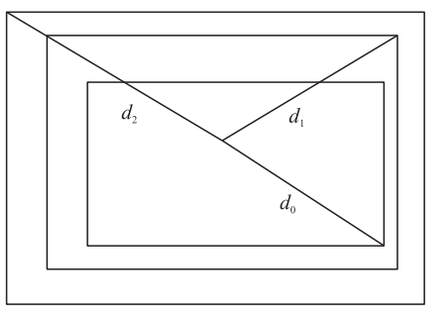

图 1　矩形区域算法框架

Fig. 1　The framework of rectangular area algorithm

生成最佳区域的算法描述如下:首先,图 2 描述算法运行的实例:

由图 2(a)可知,图中有 5 个用户,其中,用户 $k1$ 周围存在其他 4 个用户($k2$-$k5$). 这些用户通过移动网络通讯,相互不信任对方,也不知道对方的精确位置信息. 假设用户 $k1$ 收到周围其他 4 个用户分享的区域信息,如图 2(b)所示,每个区域大小和形状各不相同,并且这些区域互相覆盖重叠. 用户 $k1$ 只能从这些区域中得知其他用户大概的位置信息,并不能推断出精确的位置信息.

接着,输入用户 $k1$ 可接受的最大面积 $S_{\max}$ 和可接受的最小面积 $S_{\min}$,当 $S(R_S)\in[S_{\min},S_{\max}]$时,在区域 S 中的用户分布密度为 $\omega=\dfrac{k}{S(R_s)}$,当 $S(R_s)\notin[S_{\min},S_{\max}]$时,在区域 S 中的用户分布密度为 $\omega=\dfrac{k}{\gamma\cdot S(R_s)}$. 通过

区域 S 的 ω 值可以较好地反映出隐私保护程度和位置服务能力.

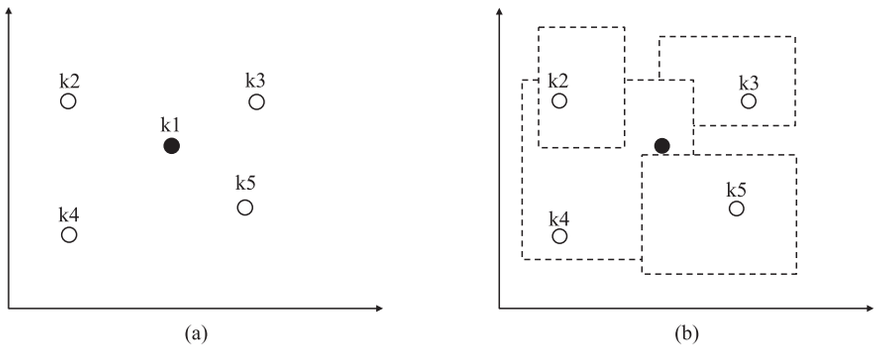

图 2　矩形区域生成算法图

Fig. 2　The figure of rectangular area generation algorithm

算法 1　生成用户最佳区域的算法

参数:已知坐标值 x_{ua},y_{ua},x_{vb},y_{vb},$S(R_S)=|x_{ua}-x_{vb}|\times|y_{ua}-y_{vb}|$,用户数量 $N(R_s)=k$.输入:$S_{\max}$,$S_{\min}$.

输出:ω 在 $S(R_s)$ 范围内达到最大值,即生成用户 U 的最佳矩形区域.

主要步骤:

Procedure:

If $S(R_S)\in[S_{\min},S_{\max}]$

Then

用户分布密度为 $\omega=\dfrac{k}{S(R_s)}$;

Else

用户分布密度为 $\omega=\dfrac{k}{\gamma\cdot S(R_s)}$.

End Procedure

此时,我们可以发现,用户 $k1$ 若提出查找距离其 d m 范围内的某超市位置,其中用户 $k1$ 的位置信息是 X,匿名服务器将用户所在矩形区域作为位置虚假区域 S,同时用户分布密度也会随 S 而变化,用户的真实位置被模糊化,攻击者即使截获位置虚假区域 S,也无法判断用户的真实位置.

算法 2　模糊用户位置坐标的算法

根据勾股定理,由两点间距离 d_2 得出三角形两条边的边长 a 和 b,其中一条边随机产生.

参数:$L_0(x_0,y_0)$ 代表用户现在的位置,$L(x,y)$ 为模糊后的用户位置,k 为用户定义的保密级别,d 为真实位置和模糊后位置的距离,m 为泛化区域标识.

输入:用户的保密级别 k,用户的位置信息 $L_0(x_0,y_0)$.

输出:模糊位置坐标.

步骤:

Procedure:

第一步:随机生成 a.

第二步:由勾股定理计算出 b.

第三步:

IF $m=0$

THEN $L(x,y)=L(x_0+a,y_0+b)$;

IF $m=1$

THEN $L(x,y)=L(x_0-a,y_0+b)$;

IF $m=2$

THEN $L(x,y)=L(x_0-a,y_0-b)$;

ELSE

$L(x,y)=L(x_0+a,y_0-b)$.

End Procedure

我们发现,用户的保密级别 k 会影响攻击者对用户所在区域的判断,并且随着 k 的增加该区域也变大,随之加大攻击者判断和攻击用户的难度.

3　实验结果及分析

本文实验采用 Thomas Brinkhoff 提出的基于网络的移动对象生成器,选用德国奥尔登堡的地图作为实验背景,生成 500 个节点分布在整个区域,实验环境为 Window XP 操作系统,内存为 2GB,参数 $k\in[1,50]$,γ 为 0.1、0.2、0.3.

3.1　匿名方法对匿名区域面积的影响

实验首先验证 k-匿名法和矩形区域 k-匿名法对匿名区域面积的影响. 参数 k 从 1 到 50,$S(R_s)$ 取值在定义的 $S_{\min}$ 和 $S_{\max}$ 之间,结果如图 3 所示. 从图 3 可以看出,随着参数 k 值的增大,匿名区域面积呈增长趋势,但是 k-匿名法的增长速度较小,矩形区域 k-匿名法呈线性增长趋势. k-匿名法的匿名区域是基于用户真实的位置信息生成的,而矩形区域 k-匿名法是基于该矩形区域中其他用户的矩形区域生成的,当 $k=50$ 时,矩形区域 k-匿名法的匿名区域面积约是 k-匿名法的 7 倍,说明在相同位置隐私需求的情况下,矩形区域 k-匿名法得到的匿名区域较大,能够为用户提供更好的位置隐私保护功能.

3.2　面积调控系数 γ 对匿名区域面积的影响

当 $S(R_S)\in[S_{\min},S_{\max}]$ 时,随着 k 值的增大,匿名区域面积随之而增长. 当 $S(R_s)\notin[S_{\min},S_{\max}]$ 时,当 γ 取不同值时,使用矩形区域 k-匿名法生成的匿名区域面积大小也不同. 随着 k 值的增大,不同 γ 值的匿名区域面积都是增长的;随着 γ 值减小,每一个匿名区域面积的增长幅度呈增长趋势,只是增长幅度不同,如图 4 所示. 对于不同的 γ 值,当用户的位置隐私需求不同时,矩形区域 k-匿名法的匿名区域面积也是不同的. 当 $k=50$ 时,$\gamma=0.1$,匿名区域面积约是 69,$\gamma=0.2$,匿名区域面积约是 22.8,$\gamma=0.3$,匿名区域面积约是 11.2. 用户可根据自己所处的区域调整 γ 值进行缓冲,最大限度实现位置隐私保护和服务质量的要求.

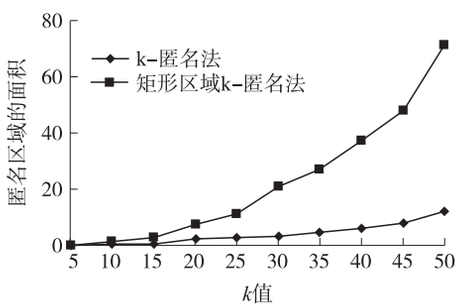

图 3　不同匿名方法的匿名面积比较

Fig. 3　The comparison of anonymous area based on different anonymous methods

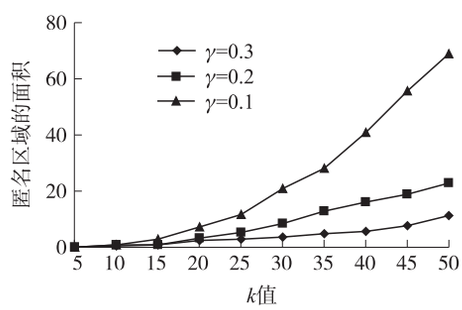

图 4　矩形区域不同 γ 值的匿名面积比较

Fig. 4　The comparison of anonymous area in rectangular region based on different γ values

3.3　相对匿名度的比较

如图 5 所示比较了 k-匿名法和矩形区域 k-匿名法的相对匿名度,在 k 值相同的情况下,矩形区域 k-匿名法匿名化的消息数量比 k-匿名法多,因此相对匿名度也较大. 从图 5 可以看出,在 k 值相同的情况下,矩形区域 k-匿名法的相对匿名度比 k-匿名法的相对匿名度大很多,说明在相同的匿名处理时间内,矩形区域 k-匿名法较 k-匿名法可以处理更多的匿名查询消息,实用性较大.

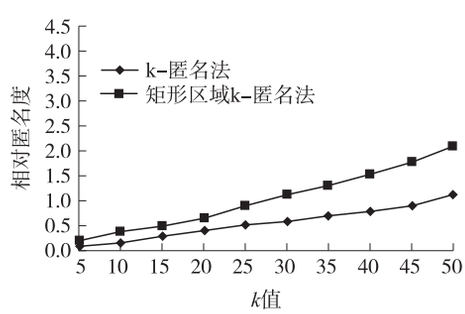

图 5　不同匿名法相对匿名度比较

Fig. 5　The comparison of relative anonymity degree based on different anonymous methods

4　结论

位置隐私保护是增强现实技术中基于 LBS 值得关注的问题之一. 本文提出矩形区域 k-匿名法的位置隐私保护解决方法,该方法让用户使用一个矩形区域来代替真实的位置信息,保证在所有场景中生成的区

域是最佳区域,达到较好的位置隐私保护效果.

[参考文献]

[1] HOLLERER T H,FEINER S K. Mobile Augmented Reality. In:Telegeoinformatics:Location-Based Computing and Services [M]. Oxford:Taylor & Francis Books Ltd.,2004.

[2] 贾金营,张凤荔. 位置隐私保护技术综述[J]. 计算机应用研究,2013,30(3):641-646.

[3] 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):2 088-2 093.

[4] 周长利,马春光,杨松涛,等. 基于敏感位置多样性的 LBS 位置隐私保护方法研究[J]. 通信学报,2015,36(4):1-12.

[5] 周傲英,杨彬,金澈清,等. 基于位置的服务:架构与进展[J]. 计算机学报,2011,34(7):1 155-1 171.

[6] 薛姣,刘向宇,杨晓春,等. 一种面向公路网络的位置隐私保护方法[J]. 计算机学报,2011,34(5):865-878.

[7] 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,2003:31-42.

[8] SAMARATI P,SWEENEY L. Generalizing data to provide anonymity when disclosing information(Abstract)[C]//Proceedings of the Seventeenth ACM Sigact-Sigmod-Sigart Symposium on Principles of Database Systems,New York,1998:188.

[9] SWEENEY L. K-anonymity:a model for protecting privacy[J]. International journal of uncertainty,fuzziness and knowledge-based systems,2002,10(5):557-570.

[10] SWEENEY L. Achieving k-Anonymity privacy protection using generalization and suppression[J]. International journal on uncertainty,fuzziness and knowledge-based systems,2002,19(5):571-588.

[11] MACHANAVAJJHALA A,GEHRKE J,KIFER D,et al. L-diversity:privacy beyond k-anonymity[C]//Proceedings of the 22nd ICDE,Atlanta,2006:24-35.

[12] CHOW C,MOKBEL M F,HE T. Tinycasper:a privacy-preserving aggregate location monitoring system in wireless sensor networks[C]//Proceedings of SIGMOD08(demo),Vancouver,Canada,2008,1 307-1 310.

[13] 韩建民,于娟,虞慧群,等. 面向数值型敏感属性的分级 l-多样性模型[J]. 计算机研究与发展,2011,48(1):147-158.

[责任编辑:顾晓天]