doi:10.3969/j.issn.1001-4616.2018.03.007

增强现实中基于 LBS 的双重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,南京城市职业学院工程与信息学院,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,针对这一问题,本文提出一种双重匿名方法保护用户位置隐私,该方法融合自适应 k 匿名技术和差分隐私技术,根据用户服务请求类型判断隐私等级自适应产生 k 值,然后通过差分隐私技术随机产生扰动,将扰动位置作为用户真实位置发送给服务提供商. 实验结果表明该方法提高了相对匿名度,LBS 服务质量也得到保障,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,自适应,差分隐私技术
[**中图分类号**]TP391.9　[**文献标志码**]A　[**文章编号**]1001-4616(2018)03-0042-05

Double Anonymity Location Privacy Protection Based on LBS in Augment Reality

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

(1.Institute of Engineering and Information,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,Nanjing 210003,China)

Abstract:Rapid development of location based service(LBS) and augment reality induces to enlargement application of LBS,which also brings hidden danger of disclosure of user location privacy. Here,we suggested a double anonymity privacy method to protect the user location privacy. The adaptive k-anonymity technology and differential privacy technology were combined. k value was generated self-adaptively according to privacy level which was resulted from user service request type. Disturbance location was made through differential privacy technology and sent to service producer as the real user location. Our results indicated that this method can effectively protect the user location privacy with enhanced relative anonymity and LBS service quality.

Key words:location based service,location privacy,k-anonymity technology,self-adaption,differential privacy technology

增强现实是一种允许在正常感知现实上叠加由计算机形成的图片和信息的技术[1]. LBS 是增强现实提供的常用服务之一,它可以通过用户位置信息为用户提供所需服务. LBS 推动了定位技术的发展,给人们生活带来了便利,但也威胁到个人隐私信息[2]. 攻击者能够根据用户位置信息推断出用户的运动轨迹,通过对这些轨迹的分析,攻击者还能够推断出更多的隐私信息,如家庭地址、个人健康状况或个人生活习惯等敏感信息[3]. 位置隐私的泄露严重威胁了个人隐私,为了解决这个问题,位置匿名技术得到研究者关注.

1　相关工作

现有的位置隐私保护技术可归纳为 3 类:假名法[4],对于任何 LBS 请求,用户通过一个可信的第三方生成一个可替代用户标识的信息,发送给 LBS 服务提供商从而来保护用户的隐私;基于信息检索的方

收稿日期:2018-04-16.
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

法[5],通过利用加密协议对位置数据进行加密从而保护用户的隐私;基于位置匿名的方法,将用户的位置点区域化,以一个区域来代替该位置点的查询,从而保护用户的隐私.

Gruteser 等人提出的 k-匿名方法[6],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. 例如用户在 t_1 时刻匿名区域用集合 $\{A,B,C,D\}$ 表示,那么在匿名后,攻击者获得的位置信息是一个位置集合,无法得到用户的确切位置信息,从而保护用户的位置. 但是用户在 t_2 时刻匿名区域用集合 $\{A,B,E,F\}$ 表示,在 t_3 时刻匿名区域用集合 $\{A,H,I,J\}$ 表示,那么攻击者很容易根据用户的匿名轨迹推断出用户提出的查询和所在位置.

文献中[7]提出采用信息熵来保护用户的位置信息,但是信息熵可能会导致匿名集合中用户全部汇聚于一点的情况,不能反映用户的位置分布,导致用户位置信息的泄露.

文献中[8-9]提出的差分隐私技术是将噪声加入到原始数据或数据统计结果中达到隐私保护,该方法即使插入或删除一条记录也不会影响到其他输出结果,它不考虑攻击者的背景知识,也不使用受信任的第三方(trusted third party,TTP). 噪声机制是实现差分隐私的主要技术,文献中[10]提出一个基于位置的差分隐私扩展模型,它可以生成一个匿名位置,并使其在一定范围内不会受到攻击,但是由于真实位置和匿名位置之间的距离无法预知,所以 LBS 服务质量得不到保障.

本文提出的位置隐私保护方法是融合了自适应 k 值匿名技术和差分隐私技术,首先根据用户提出的服务请求得出 k 值进行自适应 k 值匿名保护,然后利用差分隐私技术对用户真实位置进行扰动,获取一个匿名位置,将此匿名位置发送获取服务,从而保护用户位置隐私.

2　双重位置隐私保护模型

2.1　相关定义

定义 1　LBS 请求 Req,用二元组 $Req=<l,q>$表示用户在获取位置服务时向位置服务提供商发送的请求,其中,l 表示用户发出请求时所在的位置,$l=(x,y)$,q 表示查询内容,那么 $l_t=(x_t,y_t)$ 表示用户的真实位置,$l_k=(x_k,y_k)$ 表示匿名位置.

定义 2　设任意相邻的数据集 D 和 D',对于任意算法 P,其值域是 Range(P),算法 P 在相邻数据集上输出结果为 R,若 $Pr[P(D)\in R]\leqslant e^{\varepsilon}\cdot Pr[P(D')\in R]$,那么称为 P 满足 ε-差分隐私. 其中 ε 是隐私保护参数,表示隐私预算程度.

定义 3　差分隐私中拉普拉斯机制通过添加拉普拉斯分布的噪声扰动数据. 设噪声服从方差为 $\frac{\Delta Q}{\varepsilon}$,其中 Q 为查询函数,$\Delta Q=\max[Q(D)-Q(D')]$ 为输出结果变化的最大值,那么

$Pr(x,\lambda)=\frac{e^{-\frac{(x)}{\lambda}}}{2\lambda}$,其中 x 表示具体的变量,参数 $\lambda=\frac{\Delta Q}{\varepsilon}$.

定义 4　设服务提供商给定的隐私泄露阈值为 $L_{\max}$,k-匿名表的隐私泄露概率为 L,原始数据表为 S,k-匿名表为 S',被攻击方为 U,U 的隐私属性值为 T_u,U 所在的每组元组记为 MN_u,并设 $|MN_u|=h$,T_u 在 MN_u 中出现的次数记为 $|T_u|=f$,U 的连接候选集记为 C_u,并设 $|C_u|=g$. 那么,个体被攻击方 U 的隐私泄露的概率可表示为:

$$L(U)=\frac{g^h-g^{h-f}(g-1)^f}{g^h}=1-\left(1-\frac{1}{g}\right)^f.$$

定义 5　设 $L(U)=1-\left(1-\frac{1}{g}\right)^f\leqslant 1-\left(1-\frac{1}{k}\right)^l$,$1-\left(1-\frac{1}{k}\right)^l\leqslant L_{\max}$,可得 $1-\left(1-\frac{1}{k}\right)^l\geqslant 1-L_{\max}$,最终得出 $k\geqslant\frac{1}{1-(1-L_{\max})^{\frac{1}{l}}}$,其中 l 为元组中敏感属性值的重复次数.

2.2　算法描述

用户向位置服务提供商提交服务请求从而获取位置服务,该算法首先根据用户的服务请求定义隐私

等级得出 k 值,对其进行位置隐私保护处理,然后对其匿名位置通过差分隐私技术对位置进行扰动获取扰动位置[11,12],并提交匿名位置获取查询结果.

2.2.1　自适应 k 值匿名算法

使用自适应 k 值匿名算法,用户先向 LBS 提供商发送服务请求,受信任的第三方接受该服务请求,接着自适应 k 值控制器根据发送的服务请求定义隐私等级,根据隐私等级得出 k 的值,受信任的第三方根据 k 值启动相应的位置隐私保护方法进行信息隐私保护的处理,主要步骤如下:

步骤 1　用户发送服务请求,第三方接受服务请求;

步骤 2　k 值控制器根据请求内容定义隐私等级;

步骤 3　k 值控制器得出 k 值;

步骤 4　若 k 值在一定范围内,受信任的第三方联合使用 k-匿名法和假名法做隐私保护处理;

步骤 5　若 k 值超出范围,受信任的第三方使用 k-匿名法做隐私保护处理;

步骤 6　选取 k 为匿名集.

算法 1　自适应 k 值匿名算法

参数:$k_{\min}$,$k_{\max}$

输入:k

输出:k 匿名集

主要步骤:

设 k 最小取值为 $k_{\min}$,k 最大取值为 $k_{\max}$

if $k\in(k_{\min},k_{\max})$

受信任的第三方联合使用 k-匿名法和假名法做隐私保护处理;

else

受信任的第三方使用 k-匿名法做隐私保护处理;

end if

选取 k 集合为匿名集

我们发现,k-匿名方法在实现隐私保护时需要同时传输 k 个位置,其传输开销大,而且单一的 k 匿名方法并不能有效保护位置隐私,因此我们引入差分隐私扰动技术进行再次匿名,该技术已被证明可以有效地抵御具有任意背景知识的攻击者.

2.2.2　差分隐私技术的匿名算法

差分隐私技术最早用于 2006 年数据库领域[13],我们把该技术引入位置隐私保护模型,首先保护数据间的联系,降低加噪次数,提高数据处理效率;接着,只加噪访问次数高的位置,独立看待经度和纬度,将经度和纬度各自生成一个不可区分的经纬度,再将其合成为匿名点.

我们把真实位置记作 $l_t(x_t,y_t)$,匿名位置为 $l_k(x_k,y_k)$,并且 l_t 和 l_k 在参数 ε 范围内不可区分. 在自适应 k 值算法中我们得到了一个 k 匿名集,根据前面的定义,我们得出:

$$Pr(x_t\to x_p)\leqslant e^{\varepsilon}Pr(x_t\to x_p),$$

$$Pr(y_t\to y_p)\leqslant e^{\varepsilon}Pr(y_t\to y_p).$$

此时,我们引入拉普拉斯分布的噪音扰动此真实位置,得出:

$$Pr(x_t\to x_p)=\frac{e^{-\frac{|x_t-x_p|}{\lambda_x}}}{2\lambda},$$

$$Pr(y_t\to y_p)=\frac{e^{-\frac{|y_t-y_p|}{\lambda_y}}}{2\lambda}.$$

主要步骤如下:

步骤 1　根据 k 值生成参数 λ

步骤 2　判断尝试次数

步骤 3　若尝试次数小于等于最大尝试次数,则将经纬度分别匿名,得到匿名点 l_k

步骤 4　尝试次数每次加 1,返回步骤 2

步骤 5　执行步骤 3

步骤 6　当尝试次数大于最大尝试次数时,跳出算法,不再生成匿名点

算法 2　差分隐私的匿名算法

输入:l_t,k,ε,Max(m),其中 Max(m)为最大尝试次数

输出:l_k

主要步骤:

$\lambda=\lambda(k)$

count=count+1

IF count(m)≤count(Max(m))

$l_k=(\lambda,l_t,k)$

find=check(l_k,k)

ELSE

Exit

我们发现,引入差分隐私技术后,能够克服传统隐私保护模型需要背景知识假设和无法定量分析隐私保护水平的缺点,由于查询结果和输出结果中加入了噪声,使得数据失真,从而保证了最小私密性泄露和最大效用性,达到了位置隐私保护的目的[14].

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[15],数据集共包含 18 000 条轨迹,全长 2 km,实验环境为 Window OS 操作系统,4GB 内存,用户的位置平均 1~5 s 采集一次,参数 $k\in[1,50]$.

3.1　距离对 LBS 服务质量的影响

我们知道用户想要获得满意的位置服务,匿名距离和真实距离不能太远,但匿名距离和真实距离若很近,用户隐私容易暴露. 实验首先验证传统 k-匿名法、自适应 k 匿名法和本文的双重匿名法对 LBS 服务质量的影响,实验参数 $k\in[1,50]$,隐私预算参数 $\varepsilon=0.5$,距离范围 1 m 至 1 000 m,实验结果如图 1 所示. 从图 1 我们可以看出基于本文提出的位置隐私保护方法得到的 LBS 服务质量优胜于传统 k-匿名法和自适应 k 匿名法,它们的 LBS 服务准确率均随匿名距离和真实距离的延长而下降,但基于本文的方法所得到的 LBS 服务准确率最高达 90%,最低达 52%,明显优胜于其余两种方法.

3.2　相对匿名程度的比较

图 2 比较了 k-匿名法、自适应 k 匿名法和双重匿名法的相对匿名度,实验参数 $k\in[1,50]$,隐私预算参数 $\varepsilon=0.5$,在 k 值相同的情况下,双重匿名法匿名化的消息数量比 k-匿名法和自适应 k 匿名法多,因此相对匿名度也较大. 从图 2 可以看出,在 k 值相同的情况下,双重匿名法的相对匿名度比 k-匿名法和自适应 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,引入了差分隐私技术的双重匿名法较其余两种方法可以处理更多的匿名查询消息,实用性较大.

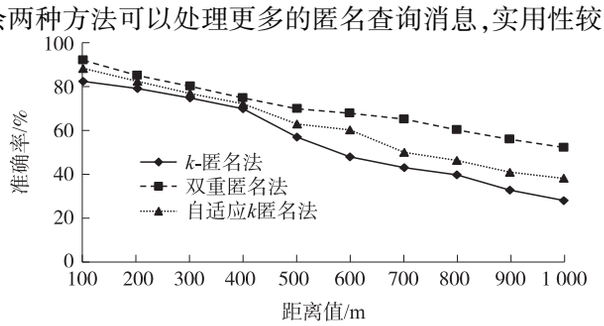

图 1　不同匿名法 LBS 服务质量比较

Fig. 1　The comparison of LBS service quality based on different anonymous methods

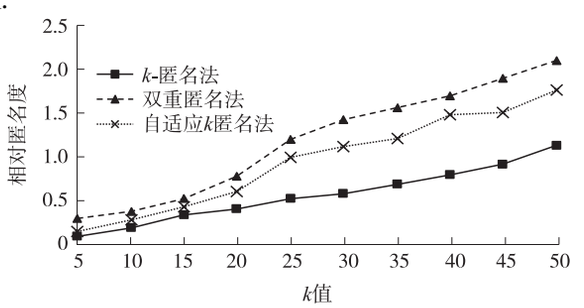

图 2　不同匿名法相对匿名度比较

Fig. 2　The comparison of relative anonymity degree based on different anonymous methods

4 结语

本文提出了双重匿名方法来保护用户的位置隐私,融合了自适应 k 值技术和差分隐私技术,并且加入扰动技术. 首先根据用户服务请求内容定义隐私等级,为用户自动分配 k 值,然后通过差分隐私技术随机产生扰动,选出匿名点作为扰动位置发送给服务提供商,从而保护用户的位置隐私. 通过实验分析,可以发现该方法能够抵御背景知识的攻击,扰动位置具有很强的隐私保护强度,并且平衡了位置隐私与服务质量.

[参考文献]

[1] HOLLERER T H,FEINER S K. Mobile augmented reality, telegeoinformatics: location-based computing and services[M]. London:Taylor and Francis Books Ltd,2004.

[2] 张学军,桂小林,伍忠东,等. 位置服务隐私保护研究综述[J]. 软件学报,2015,26(9):2373-2395.

[3] GAMBS S,KILLIJIAN M. Show me how you move and I will tell you who you are[J]. Transactions on data privacy,2010,4(2):34-41.

[4] 潘晓,肖珍,孟小峰. 位置隐私研究综述[J]. 计算机科学与探索,2007,1(3):268-281.

[5] MI Y J,SUNG J J,JAE W C. A New K-NN query processing algorithm enhancing privacy protection in location-based services[C]//IEEE First International Conference on Mobile Services. Hawaii,USA,2012:17-24.

[6] 潘晓,郝兴,孟小峰. 基于位置服务中的连续查询隐私保护研究[J],计算机研究与发展.2011,47(1):121-129.

[7] GEDIK B,LIU L. Protecting location privacy with personalized k-anonymity:architecture and algorithms[J]. IEEE Trans mobile computing,2008,9(1):1-17.

[8] DWORK C. Differential privacy,automata,languages and programming[M]. Berlin:Springer,2006:1-12.

[9] 许明艳,赵华,季新生. 位置服务隐私保护技术研究综述[J]. 信息工程大学学报,2015,16(5):543-551.

[10] ANDRES M,BORDENABEN,CHATZIKOKOLAKIS K,et al. Geo-indistinguishability:differential privacy for location-based systems[C]. Proceedings of the 2013 ACM SIGSAC conference on Computer and communications security. New York:ACM,2013:901-914.

[11] PALANISAMY B,LIU L. MOBIMIX:protecting location privacy with mix-zones over road networks[C]//Proceedings of International Conference on Data Engineering. Piscataway:IEEE,2011:494-505.

[12] 熊平,朱天清,王晓峰. 差分隐私保护及其应用[J]. 计算机学报,2014,37(1):101-122.

[13] Mc SHERRY F,TALWAR K. Mechanism design via differential privacy[C]//Proceedings of Foundations of Computer Science(FOCS).Piscata-way:IEEE Press,2007:94-103.

[14] DWORK C. The promise of differential privacy:a tutorial on algorithmic techniques[C]//Proceedings of Foundations of Computer Science.Piscata-way:IEEE Press,2011:1-2.

[15] KALNIS P,GHINITA G,MOURATIDIS K,et al.Preventing location-based identity inference in anonymous spatial queries[J]. IEEE Trans on knowledge and data engineering,2008,19(12):1719-1733.

[责任编辑:顾晓天]