一种新型基于位置服务的隐私保护方案

李 兰[1] 张才宝[1] 奚舒舒[1] 马鸿洋[2]
[1](青岛理工大学信息与控制工程学院 山东 青岛 266000)
[2](青岛理工大学理学院 山东 青岛 266033)

摘 要 基于位置服务的移动社交网络迅速发展,为用户提供了诸多便利。使用这种增值服务时,用户必须向位置服务提供商提供当前位置信息与请求内容,这一过程难免会造成隐私信息泄露,给用户的人身和财产安全造成威胁。因此提出一种新型基于位置服务的隐私保护方案,利用基于球树的匿名区域构造算法构造匿名区域,根据距离与请求内容的权重计算方法,选择熵最大的用户组,保证区域内用户分布均匀性和请求内容多样性。实验结果表明,该算法构造的匿名区域面积与类四叉树算法相比减少 30%,能有效保护用户的隐私安全,降低 LBS 服务器的开销。

关键词 基于位置服务 位置隐私保护 k-匿名 球树

中图分类号 TP309 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2022.12.050

A NOVEL PRIVACY PROTECTION SCHEME BASED ON LOCATION SERVICE

Li Lan[1] Zhang Caibao[1] Xi Shushu[1] Ma Hongyang[2]
[1](*School of Information and Control Engineering*, *Qingdao University of Technology*, *Qingdao 266000*, *Shandong*, *China*)
[2](*School of Science*, *Qingdao University of Technology*, *Qingdao 266033*, *Shandong*, *China*)

Abstract The rapid development of mobile social networks based on location services provides great convenience to users. When using this value-added service, users must provide location information and requested content to the location service provider. This process will inevitably cause the leakage of private information and threaten the safety of users' personal and property. Therefore, this paper proposes a novel privacy protection scheme based on location service. It used an anonymous area construction algorithm based on a ball tree to build an anonymous area. According to the distance and requested weight calculation method, the user group with the largest entropy was selected to ensure uniform distribution of users in the area and ensure the diverse request content. The experimental results show that compared with the Quadtree-like algorithm, the anonymous area constructed by this algorithm is reduced by 30%, which can effectively protect the privacy of users and reduce the overhead of LBS server.

Keywords Location-based service Location privacy protection K-anonymity Ball tree

0 引 言

在基于位置服务[1-2](Location-Based Service,LBS)的移动社交网络[3](Mobile Social Networks,MSNs)中,用户通过位置定位设备查询附近的兴趣点(Point of Interest,POI)[4],来满足生活和工作上的需求[5-6]。然而,用户必须主动提供位置信息和查询内容才能使用这种服务,当包含用户隐私信息的日志文件被攻击者窃取后,用户的职业、政治观点和行为模式等很容易被推断出来[7-8]。因此,当务之急是要有效地保护用户的隐私安全。

区域 k-匿名技术虽然可以将用户被识别的概率降低到 $1/k$,但这种技术存在部分问题。首先,构建的匿

收稿日期:2020-05-29。国家自然科学基金项目(11975132);山东省高等教育科技计划项目(J18KZ012)。**李兰**,教授,主研领域:数据挖掘,图像处理。**张才宝**,硕士生。**奚舒舒**,硕士生。**马鸿洋**,教授。

名区域中除查询用户外,其他用户可能出现分布集中的情况,并可能集中分布在查询用户附近,易造成用户位置隐私泄露。其次,匿名区域的构造过程中可能产生冗余空间。最后,匿名区域内的查询请求类型可能过于单一,易造成查询信息泄露。

基于上述原因,本文提出一种新型的基于位置服务的隐私保护方案。首先利用基于球树[9]的匿名区域构造算法(Balltree-Based Anonymous Region Construction Algorithm,BT-RCA)搜索邻居用户,与其他搜索算法相比,球树算法可以提高搜索邻居用户的效率。此外,综合考虑了用户组中距离权重与请求内容权重,在构建的多个用户组中,选择熵最大的一组,有效地保护了移动用户的位置信息和查询内容。最后,通过安全性能分析,进一步验证了该方法在隐私保护方面的有效性。

1 相关工作

位置隐私保护方法根据标准不同可以分为多种类型。本文通过对国内外研究现状的分析,将位置隐私保护方法分为三类:空间隐匿法[10-12]、虚拟定位法[13-15]、基于原语的密码学方法[16-18]。

1.1 空间隐匿法

Gruteser 等[10]提出基于四叉树结构的 Interval Cloak 算法,该算法结合 k-anonymity[11]思想,将用户的具体位置信息替代为一个包含至少 k 个用户节点位置信息的区域向 LBS 服务器请求查询,将用户被成功推断的概率降低到 $1/k$。

在文献[10]基础上,Mokbel 等[12]提出匿名区域构建算法,该算法以 Casper 模型为基础,利用匿名服务器管理空间索引信息,使得到的矩形匿名空间较 Interval Cloak 算法更小,提高了算法性能。但当用户数量很少时,Casper 算法会因为一直找不到足够的邻居用户导致匿名区域构建失败。

1.2 虚拟定位法

Hong 等[13]提出一种将用户真实位置替换为附近路标或相近位置的算法。Kido 等[14]在文献[13]的基础上,提出一种匿名通信技术,可以生成多个虚拟位置,并将其与用户的位置信息一并发送给 LBS 服务器,更好地隐藏用户的位置信息。

Wu 等[15]提出了多目标优化算法,综合考虑了查询概率和匿名区域的面积,通过生成的 $k-1$ 个假位置来实现 k-匿名。该算法降低了虚拟位置被过滤的可能性,但算法计算量较大,对 MSN 用户设备要求较高。

1.3 基于原语的密码学方法

基于原语的密码学方法[16-18]通过对用户与 LBS 服务器的交互信息加密实现隐私保护目的,可以提供很好的安全性,但用户与 LBS 服务器通信中计算开销很大,因此这些方案对 MSN 用户隐私保护可行性较差。

在本文中,BT-RCA 利用球树作为空间索引结构,减少搜索邻居用户的时间,保证匿名区域生成过程中不产生冗余空间,并结合用户的距离权重与请求内容权重,有效提高了服务质量。

2 BT-RCA 模型

2.1 基本概念

定义 1 距离度量:用户 u_i 和 u_j 之间的距离度量(本文使用欧几里得距离),定义为:

$$dis(u_i,u_j)=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \tag{1}$$

式中:x、y 分别表示用户位置的经纬度信息。

定义 2 区域划分:假设 Reg 是一个半径为 r 的区域。

如果 Reg 中的用户与真实用户 u_{real} 的距离度量小于 r,则称这些用户为 u_{real} 的 dis_r-邻域用户,表示为:

$$N_{dis\_r}(u_{\text{real}})=\{u_i\in Reg \mid dis(u_{\text{real}},u_i)\leqslant r\} \tag{2}$$

如果 Reg 中用户的查询请求内容与真实用户 u_{real} 的查询内容不同,则称这些用户为 u_{real} 的 dis_c-邻域用户,表示为:

$$N_{dis\_c}(u_{\text{real}})=\{u_i\in Reg \mid boolean(u_{\text{real}},u_i)=\text{False}\} \tag{3}$$

式中:$boolean()$ 是判断请求内容是否相同的函数。

定义 3 距离权重:用 α_i 表示 dis_r-邻域用户 u_i 的距离权重,定义为:

$$\alpha_i=\frac{dis(u_{\text{real}},u_i)}{\sum_{j=1}^{3k}dis(u_{\text{real}},u_j)} \quad i=1,2,\cdots,3k \tag{4}$$

用户 u_i 在用户组 U_t 中的距离权重定义为:

$$\alpha_{ti}=\frac{dis(u_{\text{real}},u_i)}{\sum_{j=1}^{k}dis(u_{\text{real}},u_j)}\times\alpha_i \quad i=1,2,\cdots,k \tag{5}$$

与文献[19]相比,本文不仅考虑到 u_i 在整个邻居用户中的权重,更考虑到 u_i 在固定用户组中的权重。

因此用户组 U_t 基于距离的熵为:

$$H(U_t)=-\sum_{i}^{k-1}\alpha_{ti}\log_2\alpha_{ti} \tag{6}$$

定义 4 内容权重:用户组 U_t 的请求内容的权重

用 β_t 表示,定义为:

$$\beta_t = \frac{\sum boolean(u_i^c, u_j^c)}{2k} \quad i,j=1,2,\cdots,k,i\neq j \quad (7)$$

式中:u_i^c 和 u_j^c 分别表示用户 u_i 和用户 u_j 的请求内容。

由此可得用户组 U_t 的熵为:

$$HA(T) = H(U_t) + \beta_t \quad (8)$$

最后在构建的候选用户组中选择熵最大的一组。

$$U_{\max} = \underset{n\in\{1,2,\cdots,m\}}{\operatorname{argmax}} \{HA_n\} \quad (9)$$

2.2　系统模型

本文系统架构如图 1 所示,包括 MSN 用户、匿名服务器及 LBS 服务器,并假设匿名服务器是可信的。

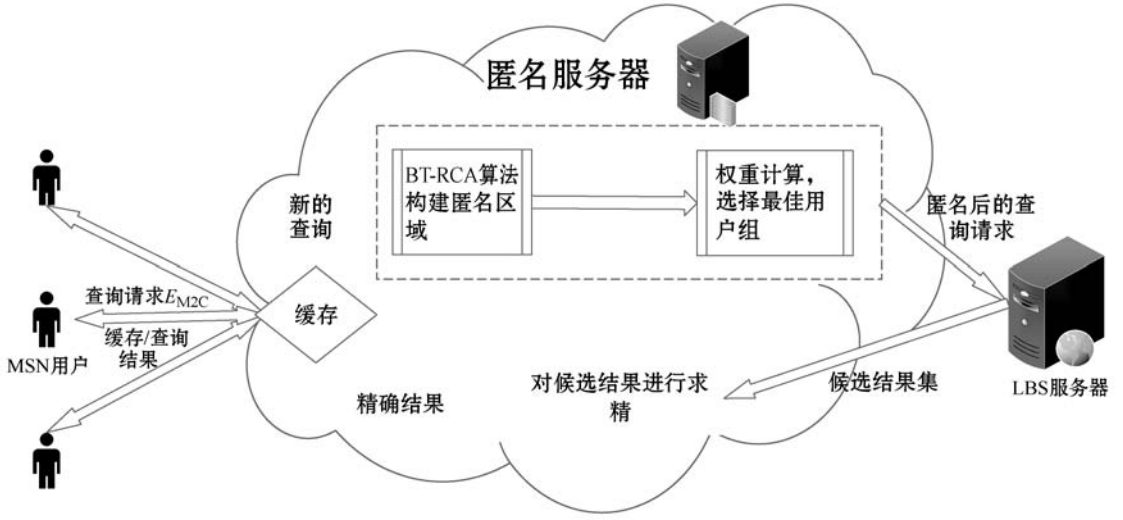

图 1　系统架构

(1) MSN 用户:MSN 用户需要向匿名服务器发送一条查询请求:

$$E_{M2C} = \{ID, A, k, c, m, l, Cac, MinC\} \quad (10)$$

式中:ID 表示用户身份;A 表示可接受的匿名区域的最小范围;k 表示区域匿名度;c 表示查询内容;m 表示构建候选区域的轮次;l 表示用户位置(x,y);Cac 表示查询结果是否需要缓存;$MinC$ 表示用户组请求类型多样性阈值。

(2) 匿名服务器:主要包括匿名模块、候选结果处理模块和缓存处理模块。

收到用户查询请求后,匿名模块根据请求内容构造球树,完成最近邻用户搜索,并用 BT-RCA 进行匿名性处理。

LBS 服务器的查询结果到达后,候选结果处理模块会对查询结果进行选择,然后将精确结果返回给用户。

为了减轻 LBS 开销,提出匿名服务器的缓存方案,在查询请求 E_{M2C} 中加入参数 Cac,其中 $Cac = \{0, 1\}$。当 $Cac = 0$ 时,表示用户不再需要本次查询结果,结果返回给用户后,匿名服务器便将之丢弃;否则,就将结果缓存。下次查询来临时,匿名服务器先将缓存反馈给用户,用户检查反馈结果,若对结果不满意,则重新发送新的查询。

(3) LBS 服务器:LBS 服务器收到来自匿名服务器的请求内容后,开始进行查询,然后将结果发送回去。

2.3　球　树

2.3.1　球树的建立

构造球树的具体流程为:

(1) 先构建一个可以将所有样本数据包含进去的超球体。

(2) 在球中选择一个点 A,A 点满足到球心 O 的距离大于球内其他任何点到点 O 的距离,再从球内选择一个离 A 点距离最远的点 B,然后将球中剩余点以距离最近为原则分配到 A、B 上,当所有数据点都正好包含于聚类时,逐个计算聚类的中心和半径。这样,便得到了两个子超球体。

(3) 将得到的每个子超球体均递归执行上述步骤(2),直至球树构建完成。

2.3.2　球树最近邻搜索

球树搜索目标点的最近邻方法如下:

(1) 从根节点开始贯穿整棵树查找包含目标点所在的叶子节点,并找出球中距离目标点最邻近的点,此时便可以得出目标点与它的最近邻点的上限的值 max,下一步就是对兄弟节点进行检查,如果 max 与兄弟节点的半径的和小于目标点与兄弟节点中心的距离,则该兄弟节点中不会存在更近的点;否则,必须对兄弟节点下的子树进行检查。

(2) 为了搜索最小邻近的值,当兄弟节点的检查

结束后,还需要向父节点进行回溯,直至到达根节点,这时最终的搜索结果就是最小邻近值。

2.4 BT-RCA

为了对用户的位置和查询内容等隐私信息进行有效保护,本文提出 BT-RCA,算法的步骤如下:

(1) 用球树搜索算法找到距离 u_{real} 最近的 $3k$ 个邻居用户,存储在长度为 $3k$ 的队列中。

(2) 球树搜索完成后,如果邻居用户达到 $3k$ 个,则从 $3k$ 用户中随机选取 $k-1$ 个用户与真实用户构成候选用户组,循环执行 m 次。

(3) 对组内的距离权重与内容权重进行计算,并在 m 个用户组中选择熵最大的一组。

(4) 如果邻居用户数小于 $3k$ 或组内请求内容多样性小于 $MinC$,提醒用户重新输入。

BT-RCA 伪代码描述如算法 1 所示。

算法 1 BT-RCA

输入:真实用户 u_{real},匿名度 k,执行轮次 m,请求内容阈值 $MinC$。

输出:匿名区域。

1. 初始化队列 q,并设置 $|q|=3k$;
2. 使用球树算法搜索 u_{real} 的最近邻用户;
3. **if** 最近邻用户数量小于 $3k$ **then**
4. 重新设置 k;
5. **else**
6. 将最近邻用户设为 $3k$ 并存入队列 q;
7. **for** $i=1$ **to** m **do**
8. 随机从 $3k$ 用户中选择 $k-1$ 个用户与 u_{real} 组成用户组 U;
9. 计算 $HA(T)$;
10. **end for**
11. 从 m 个用户组选择熵最大的一组 $U_{\max}$;
12. **if** $|N_{dis_c}(u_{\text{real}})|<MinC$ **then**
13. 重新设置 $MinC$;
14. **else**
15. 返回 $U_{\max}$;
16. **end if**
17. **end if**

3 安全性分析

目前存在的攻击方式主要为合谋攻击和推理攻击,两种攻击均无法对本文方案造成威胁,现安全性分析如下。

3.1 抗用户合谋攻击

BT-RCA 在 $3k$ 个邻居用户中随机构造用户组 $U=\{U_1,U_2,\cdots,U_m\}$,每个用户组包括随机 $k-1$ 个邻居用户与真实用户 u_{real},且考虑到用户组的距离权重与请求内容权重,因此组内的任一用户 A 无法判断真实用户 u_{real} 的位置。$P(X+A)$ 表示攻击者 X 与用户 A 合谋时推断真实用户的概率,有:

$$P(X+A)=\frac{p_{\text{real}}}{\sum_{i=1}^{k}p_i}=\frac{1}{k} \tag{11}$$

式中:p_{real} 表示真实用户 u_{real} 被识别的概率;p_i 表示组内任一用户被识别的概率。

攻击者又与用户 B 合谋,因为用户 A 与用户 B 没有联系,所以推断成功的概率为:

$$P(X+B)=P(X+A)=\frac{p_{\text{real}}}{\sum_{i=1}^{k}p_i}=\frac{1}{k} \tag{12}$$

因此 BT-RCA 可以成功抵抗用户合谋攻击。

3.2 抗 LBS 推理攻击

连续查询时,LBS 服务器中有真实用户的查询记录,记为:

$$QueryRec=\{\hat{U},POIs\} \tag{13}$$

式中:$\hat{U}=\{U_{\max1},U_{\max2},\cdots,U_{\max n}\}$ 表示历次查询的用户组的集合;$POIs=\{P_{\text{id}},P_{\text{na}},P_{\text{Inf}},P_{\text{Loc}}\}$ 表示历次查询的兴趣点信息。P_{id} 表示兴趣点序号;P_{na} 表示兴趣点名称;P_{Inf} 表示兴趣点详细信息;P_{Loc} 表示兴趣点的位置信息。

BT-RCA 随机选取邻居用户构造用户组 $U=\{U_1,U_2,\cdots,U_m\}$,并选择熵最大的一组 $U_{\max}$,且在 $\hat{U}$ 中 $U_{\max i}$ 与 $U_{\max j}$ 没有固定联系,因此 LBS 服务器根据已有信息不能推断真实用户的位置。

匿名服务器利用缓存提供服务,这一过程中,用户与 LBS 服务器没有联系,因而 LBS 服务器中的历史离散位置难以关联,为推断用户真实位置方面增加了难度。可以说,缓存的使用成功降低了 LBS 服务器推断真实用户的概率,有效保护了用户的位置隐私。

4 实验仿真

4.1 实验环境描述

算法采用 Python 编程语言实现,实验环境为 2.4 GHz 的双核 CPU,8 GB 内存,操作系统是 Windows 10,在 Thomas Brinkhoff[20] 上进行仿真实验。在 Oldenburg 地图中部,取大约 4 km × 4 km 区域位置数据,区域划分块数为 1 600 块,其中 20 个 POIs 是随机生成的,用户数量由参数控制。将 BT-RCA 与 K-DDCA[19] 和类四叉树算法[21] 进行比较。

4.2　仿真结果分析

4.2.1　匿名区域面积与 k 值的关系

空间分辨率[21]指满足 MSN 用户的匿名要求的最小空间面积与匿名算法最终构建得到的匿名区域面积的比值。

由图 2 和图 3 可以看出,三种算法的匿名区域面积随 k 的增加逐渐增加,空间分辨率逐渐减小。但类四叉树算法使用四叉树作为存储结构,没有充分考虑邻居用户之间的位置关系,k 由 24 变化到 30 的过程中,该算法产生的匿名区域面积过大,算法的空间分辨率不理想。K-DDCA 使用 kd 树作为存储结构,虽然 k 值固定时构造的匿名区域面积最小,但在隐私保护中更大的匿名面积意味着更好的隐私保护效果,该算法构造匿名区域面积过小,无法有效保护用户隐私安全。

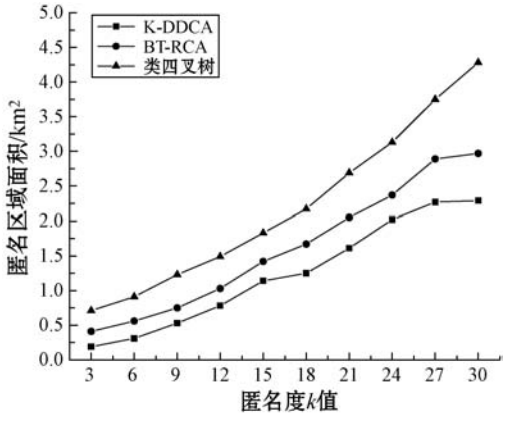

图 2　匿名区域面积与 k 值的关系

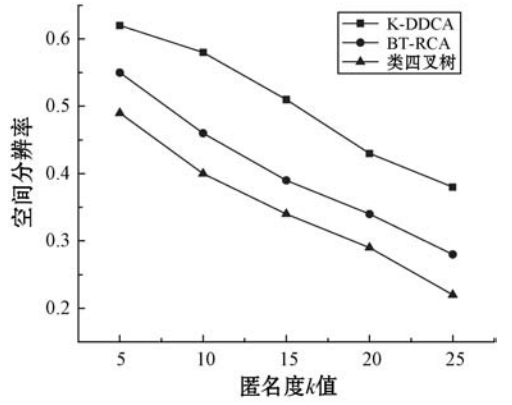

图 3　空间分辨率与 k 值的关系

BT-RCA 使用球树进行存储,构造匿名区域的过程中不会产生冗余空间,可以避免不必要的查找。当 k 由 20 增长到 25 的过程中,空间分辨率变化很小,性能较优。该算法构造的匿名区域面积不会太大或者太小,既保证了效率,又避免用户隐私信息的泄露。

4.2.2　匿名区域面积与用户数量的关系

图 4 表示 k 分别为 10 和 15 时,类四叉树算法与 BT-RCA 构造匿名域面积与用户数量的关系。匿名区域面积随用户数量增加而减小,当用户数量大于 1 000 时,类四叉树算法与 BT-RCA 构建的匿名区域面积都逐渐稳定。而 BT-RCA 利用球树模型,构建匿名区域时不会产生冗余,所以构建的区域面积要比类四叉树算法小,避免不必要的消耗,性能更加优良。

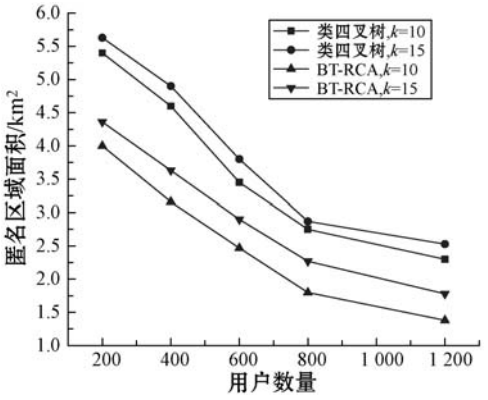

图 4　匿名区域面积与用户数量的关系

4.2.3　熵与 k 值的关系

图 5 表示以熵的形式来表示不同匿名区域构造方案的隐私级别。随着 k 的增加,类四叉树算法、K-DDCA 和 BT-RCA 的匿名性也随之增加,但 BT-RCA 性能最优。因为 BT-RCA 和 K-DDCA 形成匿名区域过程中没有产生冗余空间,而类四叉树算法则达不到这种效果,攻击者可以根据已有知识排除大量邻居用户,因此在一般情况下,类四叉树算法并不能达到真正的 k 匿名。与本文算法相比,由于 K-DDCA 构造的匿名区域过小,因此无法达到最好的隐私保护效果。

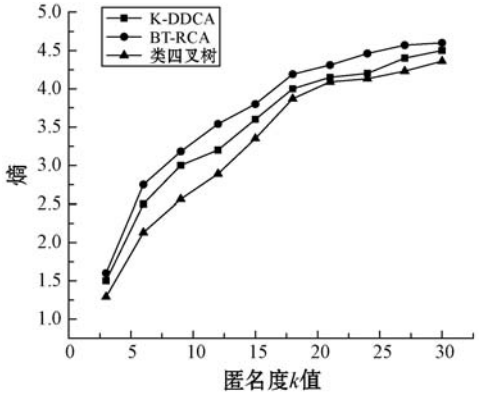

图 5　熵与 k 值的关系

5　结　语

以球树作为存储结构,本文提出新型的基于位置服务的隐私保护方案。利用 BT-RCA 构造匿名区域,综合考虑用户组中的距离权重与请求内容权重,在 m 个候选用户组中选择熵最大的一组,保证匿名区域中用户分布均匀性和请求内容多样性。最后利用 Thomas Brinkhoff 进行仿真,验证了算法在用户隐私保护方面的有效性。

算法在用户数量少时效果不佳,以后会考虑如何在用户稀少地区优化该算法。且算法以匿名服务器可信任为基础,因此以后会进行一些有效的策略来约束匿名服务器。

参 考 文 献

[1] 倪巍伟,李灵奇,刘家强. 基于 Voronoi-R* 的隐私保护路网 k 近邻查询方法[J]. 软件学报,2019,30(12):3782 - 3797.

[2] 徐启元,陈珍萍,付保川,等. 基于差分隐私的混合位置隐私保护[J]. 计算机应用与软件,2019,36(6):296 - 301.

[3] 罗恩韬,王国军,刘琴,等. 移动社交网络中矩阵混淆加密交友隐私保护策略[J]. 软件学报,2019,30(12):3798 - 3814.

[4] 侯士江,刘国华,候英. 路网环境下基于星图的位置隐私保护技术研究[J]. 计算机工程与科学,2015,37(8):1465 - 1471.

[5] 王丹,龙士工. 权重社交网络隐私保护中的差分隐私算法[J]. 计算机工程,2019,45(4):114 - 118.

[6] Primault V, Boutet A, Mokhtar S B, et al. Adaptive location privacy with ALP[C]//2016 IEEE 35th Symposium on Reliable Distributed Systems(SRDS),2016:269 - 278.

[7] 吴振强,胡静,田堉攀,等. 社交网络下的不确定图隐私保护算法[J]. 软件学报,2019,30(4):1106 - 1120.

[8] 张青云,张兴,李万杰,等. 基于差分隐私保护的兴趣点推荐算法设计[J]. 计算机应用与软件,2019,36(9):243 - 248,269.

[9] Dolatshah M, Hadian A, Minaei-Bidgoli B. Ball * -tree: Efficient spatial indexing for constrained nearest-neighbor searchin metric spaces[EB]. arXiv:1511.00628,2015.

[10] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//1st international conference on Mobile systems, applications and services,2003: 31 - 42.

[11] Sweeney L. k-anonymity: A model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,2002,10(5): 557 - 570.

[12] Mokbel M F, Chow C Y, Aref W G. The new casper: Query processing for location services without compromising privacy[C]//32nd International Conference on Very Large Data Bases,2006: 763 - 774.

[13] Hong J I, Landay J A. An architecture for privacy-sensitive ubiquitous computing[C]//2nd International Conference on Mobile Systems,applications,and services,2004:177 - 189.

[14] Kido H, Yanagisawa Y, Satoh T. An anonymous communication technique using dummies for location-based services[C]//International Conference on Pervasive Services,2005: 88 - 97.

[15] Wu D, Zhang Y, Liu Y. Dummy location selection scheme for K-Anonymity in location based services[C]//2017 IEEE Trustcom/BigDataSE/ICESS,2017:441 - 448.

[16] Jiang R, Lu R, Choo K K R. Achieving high performance and privacy-preserving query over encrypted multidimensional big metering data[J]. Future Generation Computer Systems,2018,78:392 - 401.

[17] Zheng X, Cai Z, Li J, et al. Location-privacy-aware review publication mechanism for local business service systems[C]//IEEE Conference on Computer Communications,2017: 1 - 9.

[18] Zheng X, Cai Z, Li Y. Data linkage in smart internet of things systems: A consideration from a privacy perspective[J]. IEEE Communications Magazine,2018,56(9):55 - 61.

[19] Ni L, Tian F, Ni Q, et al. An anonymous entropy-based location privacy protection scheme in mobile social networks[J]. EURASIP Journal on Wireless Communications and Networking,2019(1): 93.

[20] Brinkhoff T. A framework for generating network-based moving objects[J]. Geoinformatica,2002,6(2):153 - 180.

[21] 金福生,叶子石,宋红. 一种基于类四叉树的位置 K-匿名算法[J]. 北京理工大学学报,2014,34(1):68 - 71,76.

(上接第 277 页)

[11] Zhou C, Sun C, Liu Z, et al. A C-LSTM neural network for text classification[J]. Computer Science,2015,1(4):39 - 44.

[12] 朱海麒. 基于深度学习的运维数据异常检测研究[D]. 哈尔滨:哈尔滨工业大学,2019.

[13] Hochreiter S, Schmidhuber J. Long Short-Term memory[J]. Neural Computation,1997,9(8):1735 - 1780.

[14] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[EB]. arXiv:1803.01271,2018.

[15] Tan Z, Pan P. Network fault prediction based on CNN-LSTM hybrid neural network[C]//2019 International Conference on Communications, Information System and Computer Engineering(CISCE),2019.

[16] Bhunia A K, Konwer A, Bhunia A K, et al. Script identification in natural scene image and video frames using an attention based Convolutional-LSTM network[J]. Pattern Recognition,2019,85:172 - 184.

[17] Karim F, Majumdar S, Darabi H, et al. LSTM fully convolutional networks for time series classification[J]. IEEE Access,2017,6:1662 - 1669.

[18] Yahoo[EB/OL]. [2021 - 06 - 10]. https://webscope. sandbox. yahoo. com/.

[19] NAB[EB/OL]. [2021 - 06 - 10]. https://github. com/numenta/NAB.

[20] Catania C A, Bromberg F. An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection[M]. Pergamon Press,2012.