一种用户自我感知的位置隐私保护算法

叶吉祥 曹文慧
(长沙理工大学计算机与通信工程学院 湖南 长沙 410076)

摘 要 基于位置服务(Location based service,LBS)带来生活便利的同时也存在着潜在的隐患。针对该问题,提出一种用户自我感知的位置隐私保护算法。用户向 LBS 服务器发送请求,自我感知周围其他真实用户的存在与分布情况并形成匿名区域,将匿名区域随机划分为几个子区域,一同将位置信息发送至服务器。仿真结果表明,该方法能够有效提高用户匿名质量,降低合谋攻击成功率,降低通信开销,确保用户位置隐私安全。

关键词 LBS 位置隐私 自我感知 区域半径 区域划分

中图分类号 TP3 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2020.09.051

A LOCATION PRIVACY PROTECTION ALGORITHM BASED ON USER SELF-PERCEPTION

Ye Jixiang Cao Wenhui
(*School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410076, Hunan, China*)

Abstract Although Location based service (LBS) brings convenience in life, it also has potential hidden dangers. Aiming at this problem, this paper proposes a user self-perception location privacy protection algorithm. The user sent a request to the LBS server to self-perceive the existence and distribution of other real users around and formed an anonymous area. The anonymous area was randomly divided into several sub-areas, and the location information was sent to the server together. The simulation results show that the proposed method can effectively improve the anonymity quality of user, reduce the success rate of collusion attacks, reduce communication overhead, and ensure the privacy of user location.

Keywords LBS Location privacy Self-perception Regional radius Regional division

0 引 言

日常生活中,用户可以通过第三方软件使用基于位置服务(Location based service,LBS)[1]来获取导航服务、紧急救援服务、娱乐信息服务等[2]。在此过程中,如果攻击者盗取用户的位置信息或暴露给不可信任的第三方软件,必然造成严重影响。因此,在享有 LBS 的同时保证用户的位置隐私成为当前一个研究热点。

1 研究概述

位置隐私的概念最早由 Beresford 等[3]提出,自此,国内外许多学者提出了一系列的解决方案。目前,位置隐私保护措施主要有三类[4]:(1) 位置模糊保护法,主要包括位置的偏移、构造假位置等;(2) 引入密码学的加密方法,对用户位置进行加密后发送;(3) 制定隐私保护策略,主要针对服务商进行规范和约束。Gruteser 等[5]将 K-匿名[6]技术用于位置请求服务的隐私保护中从而达到保护目的。文献[7-10]将 K-匿名算法通过构造不同的几何形状,产生不同的匿名区域来完成保护。Kim 等[11]在未知的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,但是使用 TTP 结构不能保证中间匿名服务器绝对的安全,且容易造成系统堵塞,必然会产生新的问题。

在位置隐私保护的过程中,不仅要保护用户的位置信息,还要防止用户其他私人信息(姓名、爱好等)泄露。周长利等[12]通过构造真实用户与邻居用户的共同特征来形成匿名区域,采取几何形心作为基准进行查询,达到保护的目的。文献[13-15]根据用户与

收稿日期:2019-05-31。叶吉祥,教授,主研领域:数据隐私,语音情感处理。曹文慧,硕士生。

高频兴趣点将全局地图的位置进行单元区别划分,用户可根据网格单元中兴趣点的分布获取周围具体各项兴趣点的分布情况,保证匿名区域的多样性。

上述大多数文献均在理想环境下,以 K-匿名方法作为收集方式,达到匿名效果。在人迹稀疏的情况下,通过假位置方法正好弥补了这一不足,但由用户根据自身的隐私需求构造假位置,并将这些假位置与真实位置一起发送到服务器,使得攻击者无法区分用户的真实位置信息。在生成的假位置的过程中,由于无法判定地形(如湖面、山脉)等因素,因此会影响假位置的可靠性。

针对上述问题,本文提出一种用户自我感知的位置隐私保护算法(简称 USA)。用户自我感知周围邻居用户的分布密度情况,排除地形原因并形成匿名区域,随后将匿名区域呈多个矩形划分,最后按所分区域一同将请求内容发送至服务器。与现有方法相比,本文方法能够提高用户位置匿名性、可靠性,降低合谋攻击成功率。

2　系统模型

2.1　系统结构

位置隐私保护方法目前适用于两种系统结构:中心服务器结构和基于 P2P 结构。

中心服务器结构中,在移动用户和位置服务提供商(LSP)之间引入一个可信任的匿名器作为中间体,将用户位置信息通过匿名服务器模糊,最后发送到 LBS 服务器完成请求。P2P 结构由移动用户组成与 LBS 服务器,搭载 P2P 协议通过单跳或多跳通信产生可靠的匿名区域,各区域间用户之间相互合作完成保护。

为了方便用户感知有效的邻居用户位置信息,本文采用 P2P 结构的 LBS 系统,系统模型如图 1 所示。由于本文主要研究的是用户的位置隐私保护,所以假设在用户之间、用户与服务器之间的通信都是安全的。

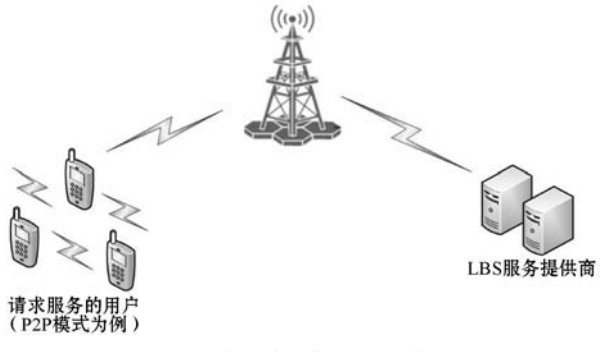

图 1　位置隐私保护系统模型

2.2　基本思路

如图 2 所示,USA 算法主要分为四个步骤:(1) 请求使用 LBS 服务的用户在有限跳数下感知自身周围邻居用户的分布密度;(2) 将感知到所有邻居用户所在位置的整体区域划分为多个矩形子区域;(3) 在每个矩形子区域中添加伪用户来均衡矩形内的用户稀疏分布;(4) 用户、邻居用户和伪用户共同将请求 LBS 的用户的内容发送至服务器,等待回应,当服务器回应给用户时,对返回结果筛选求精即可得到当前位置信息。

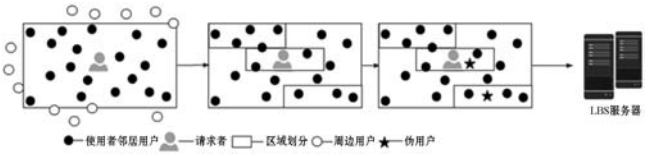

图 2　USA 算法步骤演示

3　算法实现

3.1　用户感知

在连通空间 C 中,对于用户 u,周围的邻居用户都有特定的用户群密度 ρ_u,称为周边用户平均密度,即周围感知用户个数为 $n(\rho_u, h)$,当 $h=1$ 时,即一跳所感应的平均用户数量($h \geqslant 1$),ρ_u 以增加 h 的情况下在自身的 ρ_u 上进行迭代更新,此处使用极大似然估计的原理来估计平均用户数量。

$$\rho_u = \frac{1}{n}\left(\sum_{i=1}^{n}\rho_i + C_u\right) \tag{1}$$

式中:C_u 表示第一跳周边用户的总数量。

由于通信传输时并不是在理想环境下的传输,因此,必须考虑非理想情况下能感知到的用户总数,因此,加入损耗因子 μ,计算方法为:

$$\mu = \sum^{n}\prod^{H}\frac{1}{n_h} \tag{2}$$

因此:

$$\rho_u = \mu\left(\sum_{i=1}^{n}\rho_i + C_u\right) \tag{3}$$

显然,h 越大,用户的数量虽然增多,但通信链路增多,通信开销增大,且 LBS 的准确度降低。

邻居用户感知算法中,用户 u 在初始化后,在规定的 t 周期内检测 ρ_u,当在检测周期内发现 ρ_u 发生改变或有新的邻居用户加入时,则向“L”发送携带自身 ρ_u 和邻居位置信息;收集完成“L”集合并重新检测当前邻居用户的个数;读取每条“L”的位置信息并更新 ρ_u。具体算法如下:

输入:邻居的“L”位置信息集合。

输出:ρ_u 以及“L”的信息。

1. 设置 ρ_u 的初始值 P、时间周期 t
2. 初始化邻居用户的集合且为空
3. WHILE(t 周期)
4. IF(ρ_u 有变化)
5. 产生并发送“L”位置信息
6. END IF
7. 收到“L”的位置信息
8. $P_u \leftarrow P(u,1)$;
9. WHILE(“L”中的每一个位置信息)
10. 读取每个邻居的 ρ_u
11. ρ_u 更新
12. END WHILE
13. IF(ρ_u 发生变化)
14 $\rho_u \leftarrow (\Sigma \rho_u + P_u)/|P_u|$
15. END IF
16. END WHILE

3.2 矩形区域划分

在区域划分这一过程中,主要会经历以下步骤:

(1) 用户 u 发出位置请求时,使用上述位置感知算法之后感知周围的用户,且收集到至少 $K-1$ 个用户节点信息。

(2) 将这 $K-1$ 个用户节点开始模糊化去除,使得以用户为中心的整体区域去重化偏移,提高用户位置安全性。

(3) 在满足 K 匿名的情况下,将用户及感应到的邻居用户形成一个区域,随后将生成的区域划分成多个矩形子区域。

(4) 为了使每个矩形子区域内的用户分布均衡,不失重,在完成矩形子区域划分之后,通过随即添加伪用户来调节,提高用户位置的模糊性和自我匿名的能力区域划分的算法如下:

输入:邻居用户节点集合 U,用户初始位置 Loc,搜寻节点数 N_u,划分子区域数目 n,匿名需求 K。

输出:n 个匿名区域 $C_i(1 \leqslant i)$。

1. IF($N_u < K-1$)
2. End if
3. 把 Loc 添加到集合 U 中
4. 多于节点数目 $N' = N_u - K + 1$,IF($N' = 0$),跳至第 7 步
5. 将 U 中节点横纵坐标按随机方向排序求出最大与最小的 N'个节点,并排序节点集合 U_x 与 U_y
6. 去除 U_x 中 q(q 为随机数,且 $0 \leqslant q \leqslant N$)个节点与 U_y 中前($N'-q$)个节点,并将去除的节点放入多余节点集合 U_d。若 U_d 中已经包括去除的节点,则该集合多取一次节点放入 U_d 集合中,直到 U_d 里无重复的节点,且个数为 N',$U = U - U_d$[16]
7. U 为随机选取中的节点,依据节点坐标方向以及坐标值大小进行排序,同时得到排序节点集 U'
8. 用户余数 $r = K\%n$,平均用户个数 $m = K/n$
9. 将集合 U'依据顺序进行划分为 n 个集合,选择一个集合包含 $m+r$ 个用户,另 $n-1$ 个集合则包含 m 个用户
10. 在 n 个节点集中,计算出每个集合的最小外接矩阵 $\boldsymbol{C}$
11. 返回 n 个子匿名区域 $C_1, C_2, \cdots, C_n$

4 性能分析

4.1 匿名成功率

在 P2P 通信结构下,由于用户可以与用户直接传送信息,没有中间服务器工作的影响和其他的物理损耗,因此,匿名成功率的本质即可表示为通过查询用户本身成功匿名的数目与进行总查询的用户数目之间的比值[16],具体见公式:

$$SR = N_{\text{success}}/N_{\text{total}} \tag{4}$$

4.2 合谋攻击成功率

在完成匿名的用户区域中,如果攻击型用户与用户 u 在同一个子区域,此时会增加成功被攻击的机率。假设区域中有 k 个用户,m 个攻击型用户,矩形子匿名区域有 w 个用户,真实用户在第 j 个子匿名区域,即用户 u 在查询区域中暴露的概率等于所有子匿名区域受到攻击的概率之积[15]。公式如下:

$$P_{\text{Q}}(u) = \prod_{i=1}^{j-1} \frac{m_i}{k_i} \times \prod_{i=j+1}^{w} \frac{m_i}{k_i} \times \frac{m_j}{k_{j-1}} \tag{5}$$

5 仿真与分析

5.1 实验环境

本文仿真采用 Java 语言实现,实验数据来自德国 Oldenburg 为主的交通路网[17],仿真数据使用参数如表 1 所示。在此基础上对 USA 算法的性能进行仿真,挑选和 USA 算法在相同空间里和网络环境下的区域相似性划分算法(ASDA)和基于用户均衡性划分算法(UUDA)以及 P2P 经典方法,分别在不同场景下进行性能比较。

表 1 仿真参数

参数	默认值
跳数 h	13
区域大小/km^2	23.572 × 26.915
K 值	40
子区域个数	1
恶意用户比例	0.4
用户数	2 000
搜索半径/km	200

将整体区域划分多个子区域是 USA 算法的关键,其性能决定了 USA 算法对用户保护的力度。在固定的用户基数内,划分的矩形子区域的个数不同所带来的影响也不同。如图 3 所示,当用户范围在 2 000 人时,随着划分的子区域个数的增加,用户被攻击成功的概率就逐步降低,但是无限划分子区域将带来额外的通信开销,故本文所用的子区域个数均为 4。

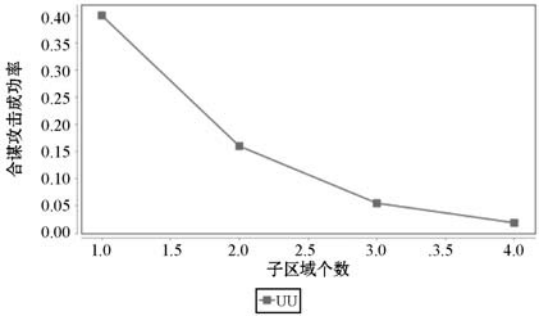

图 3　合谋攻击成功率受子区域个数的影响

5.2　算法性能分析

图 4 为 P2P 经典算法和 ASDA 算法与 USA 算法进行比较的示意图。对于 P2P 经典算法来说,没有子区域的划分,用户直接通过 K 值的大小在整体范围把请求的内容发送至服务器,造成被攻击的概率大幅度提高,而 ASDA 算法和 USA 算法在划分子区域的基础上,用户被攻击的概率明显降低。由于 USA 算法提前感知用户周围用户分布密度,继而通过添加伪用户的调和,进一步降低了被攻击的概率。

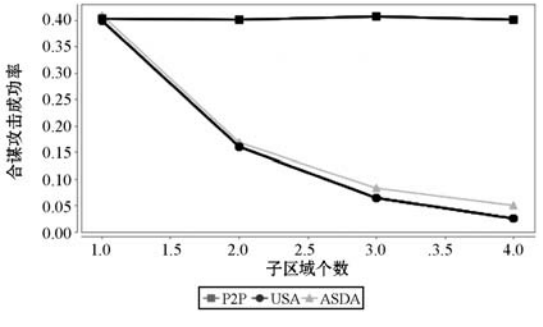

图 4　划分子区域对合谋攻击率的影响

图 5 为相同的环境下在不同算法中,用户数对平均匿名时间的影响。可以看出,在经典 P2P 算法当中,用户所需要的平均匿名时间是最少的,ASDA 算法次之。尽管如此,经典 P2P 算法在安全性上对比时间上微小的毫秒差距用户是可以忍受的,而 UUDA 算法和 USA 算法虽然平均匿名时间比经典 P2P 算法高,但安全性大大提高。由于 ASDA 算法中没有对匿名区域内的用户进行失重调整,因此其平均匿名时间较低。而 USA 算法与 UUDA 算法在匿名区域中都使用了添加伪用户的过程。在平均匿名时间的性能上,USA 算法略低于 UUDA 算法,提高了用户使用位置服务的时间,同时也提高了位置隐私保护方法的质量。

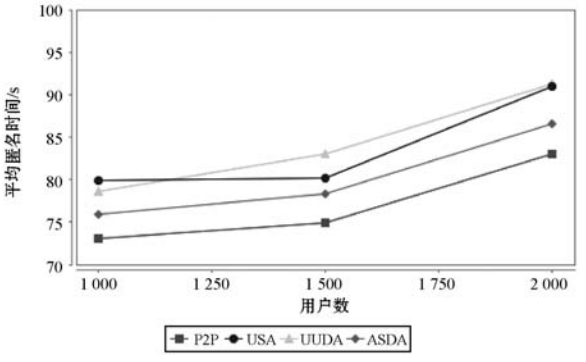

图 5　用户对平均匿名时间的影响

图 6 为在 USA 算法中不同的 K 值对于匿名成功率的影响的对比图。随着用户数量的增多,匿名成功率整体上涨。当 K 值越小,用户的数量越多时,匿名成功率越大;当 K 值越大,用户数量越少,匿名成功率越小,可能会导致匿名失败。因此,在一定范围内选择合适的 K 值是至关重要的。

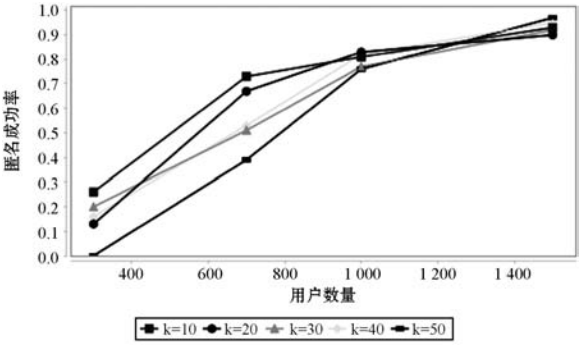

图 6　K 值对匿名成功率的影响图

6　结　语

针对 LBS 中存在的位置隐私泄露问题,提出了一种用户自我感知的位置隐私保护算法,通过用户提前感知自身周围用户分布疏密的情况,排除因地形因素影响位置隐私保护方法效果不佳的原因,再使用区域划分的方法主动降低被攻击型用户攻击的概率,在一定程度上加强保护用户位置隐私的力度。下一步将通过把用户位置隐私保护的方法放在不同的生活场景下进行深入探索,比如社交网络等,同时也会融入密钥保护等手段与增强隐私安全度结合来提高位置隐私的保护效率。

参　考　文　献

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

[2] Grissa M, Yavuz A, Hamdaoui B. Preserving the location

privacy of secondary users in cooperative spectrum sensing [J]. IEEE Transactions on Information Forensics and Security,2017,12(2):418 - 431.

[3] Beresford A R, Stajano F. Location privacy in pervasive computing[J]. IEEE Pervasive Computing, 2003, 2(1): 46 - 55.

[4] 万盛,李凤华,牛犇,等. 位置隐私保护技术研究进展[J]. 通信学报,2016,37(12):124 - 141.

[5] 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. ACM Press,2003:31 - 42.

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

[7] Zhang Y, Tong W, Zhong S. On designing satisfaction raio-aw are truthful incentive mechanisms for K-anonymity location privacy[J]. IEEE Transactions on Information Forensics and Security,2016,11(11):2528 - 2541.

[8] 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.

[9] 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,2017(12):1 - 7.

[10] 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,2016,39(4):44 - 49.

[11] Kim H I, Kim H J, Chang J W. A secure KNN query processing algorithm using homomorphic encryption on outsourced database[J]. Data and Knowledge Engineering, 2019,123:101602.

[12] 周长利,马春光,杨松涛. 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):125 - 136.

[13] 周长利,马春光,杨松涛,等. 一种保护隐私的 LBS 近邻兴趣点低通信查询方法[J]. 四川大学学报(工程科学版),2015,47(3):114 - 122.

[14] 周长利,马春光,李增鹏. 一种保护用户隐私的路网兴趣点 KNN 查询方法[J]. 计算机应用研究,2016(1):262 - 265.

[15] 张勇. 基于 P2P 的位置隐私保护算法的研究[D]. 广州:华南理工大学,2017.

[16] Zhu S Z, Huang L, Zhou C L, et al. Anonymous box KNN query method based on distribution of interest points[J]. Journal of Electronics, Computers, 2016, 44(10): 2423 - 2431.

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

(上接第 252 页)

[11] Ou M, Cui P. Asymmetric transitivity preserving graph embedding[C]//Proceedings of 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2016:1105 - 1114.

[12] Wang D, Cui P, Zhu W. Structural deep network embedding [C]//Proceedings of 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016:1225 - 1234.

[13] Yuan S, Wu X, Xiang Y. SNE: Signed network embedding [C]//Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 2017: 183 - 195.

[14] Islam M, Prakash B, Ramakrishnan N. SIGNet: scalable embeddings for signed networks[C]//Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer,2018:157 - 169.

[15] Wang S, Tang J, Aggarwal C, et al. Signed network embedding in social media[C]//Proceedings of SIAM International Conference on Data Mining,2017:327 - 335.

[16] Wang H, Zhang F, Hou M, et al. SHINE: Signed heterogeneous information network embedding for sentiment link prediction[C]//Proceedings of 11th ACM International Conference on Web Search and Data Mining. ACM, 2018: 592 - 600.

[17] Wang S, Aggarwal C, Tang J, et al. Attributed signed network embedding [C]//Proceedings of 2017 International Conference on Information and Knowledge Management. ACM, 2017:137 - 146.

[18] Kim J, Park H, Lee J, et al. SIDE: representation learning in signed directed networks[C]//Proceedings of 27th International Conference on World Wide Web. ACM,2018.

[19] Gilbert E. Predicting tie strength in a new medium[C]// Proceedings of Computer Supported Cooperative Work,2012.

[20] Xiang R, Neville J, Rogati M. Modeling relationship strength in online social networks[C]//Proceedings of 19th International Conference on World Wide Web. ACM,2010.

[21] Kumar S, Spezzano F, Subrahmanian V S, et al. Bitcoin Alpha trust weighted signed network[DS/OL]. [2019 - 07 - 10]. http://snap. stanford. edu/data/soc-sign-bitcoin-alpha. html.

[22] Kumar S, Spezzano F, Subrahmanian V S, et al. Bitcoin OTC trust weighted signed network[DS/OL]. [2019 - 07 - 10]. http://snap. stanford. edu/data/soc-sign-bitcoin-otc. html.