DOI：10.7652/xjtuxb202001002

抵御背景知识推理攻击的服务相似性位置 k 匿名隐私保护方法

闫光辉，刘婷，张学军，蔡刚，何福存，李俊成
（兰州交通大学电子与信息工程学院，730070，兰州）

摘要：针对已有位置 k 匿名方法不能很好地权衡位置隐私和服务可用性之间的关系、易于遭受背景知识推理攻击、资源开销大的问题，提出了一种抵御推理攻击的服务相似性位置 k 匿名隐私保护方法。通过引入服务相似性生成标签相似地图，并根据服务相似度将用户所在分区与其他分区合并，生成满足用户服务质量需求的匿名候选区。为抵御推理攻击，在选择 k 匿名集时使用位置熵对攻击者的背景知识进行量化以生成熵最大的 k 匿名集。在生成 k 匿名集时，采用贪心策略并在匿名集随机选取一个位置点请求服务以降低资源开销。安全性分析和实验结果表明：相较于已有位置 k 匿名方法，在相同条件下，所提方法的隐私保护度和服务可用性平均提高了48.49%和60.01%，时间开销和通信开销平均降低了86.12%和66.57%。所提方法在隐私保护度、服务可用性和资源开销之间取得了更好的权衡，且可有效抵御背景知识推理攻击。

关键词：位置隐私；服务相似性；位置熵；背景知识；k 匿名；贪心策略
中图分类号：TP309 **文献标志码**：A **文章编号**：0253-987X(2020)01-0008-11

Service Similarity Location k Anonymity Privacy Protection Scheme against Background Knowledge Inference Attacks

YAN Guanghui, LIU Ting, ZHANG Xuejun, CAI Gang, HE Fucun, LI Juncheng
(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Aiming at the problem that existing location k anonymity methods cannot provide good trade-offs between location privacy and location based service (LBS) utility, which are vulnerable to background knowledge inference attacks and leading to an overhead of resources cost, a service similarity location k anonymity privacy protection scheme against background knowledge inference attacks is proposed. Introducing definition service similarity to generate a label similar map and merging the user partition with other partitions according to service similarity, an anonymous candidate region that satisfies the user service quality requirement is generated. To resist the inference attacks, the attacker background knowledge is quantified by location entropy as selecting k anonymous set to generate the k anonymous set with the largest entropy value. During generating k anonymous set, a greedy strategy is adopted. To reduce the over-expenditure of resources, a location point in the k anonymous set is randomly selected to request services. The security analysis and experimental results show that compared with the existing location k anonymity methods, the privacy protection and LBS utility of the proposed method increase by

收稿日期：2019-05-20。 作者简介：闫光辉(1970—)，男，教授，博士生导师；张学军(通信作者)，男，副教授。 基金项目：国家自然科学基金资助项目(61762058,61662066)；兰州交通大学"百名青年优秀人才培养计划"基金资助项目。
网络出版时间：2019-09-19 网络出版地址：http://kns.cnki.net/kcms/detail/61.1069.T.20190919.0910.002.html

48.49% and 60.01% on average, respectively, and the time overhead and communication overhead reduced by 86.12% and 66.57% on average, respectively, under the same conditions. Therefore, the better trade-offs among privacy protection, LBS utility and resource cost are realized by this scheme, which effectively resists background knowledge inference attacks.

Keywords: location privacy; service similarity; location entropy; background knowledge; k anonymity; greedy strategy

随着无线通信和移动定位技术的快速发展,越来越多的移动设备具备 GPS 或北斗定位功能,使得位置服务(LBS)几乎被广泛应用在社会生活的各个领域,极大地方便了用户的出行和社交生活,因此 LBS 已被认为是能够为移动用户提供的最有前途的服务之一[1-2]。但是,恶意的服务提供商对用户特殊的个人隐私位置数据的揭露可能会直接导致用户兴趣爱好、家庭住址、健康状况等其他个人隐私的泄露。这些隐私数据的滥用,会严重危及用户的生命财产安全,进而挫伤用户使用 LBS 应用的积极性,同时也会给 LBS 服务提供商带来严重的经济、信誉损失,阻碍 LBS 市场的健康发展[2]。为此,研究者们相继提出了众多位置隐私保护方案。

位置 k 匿名技术由 Marco Gruteser 于 2003 年提出[3],相较于已有的加密保护法[4]、差分隐私保护法[5],基于位置 k 匿名技术设计的位置隐私保护法[6-8]更加简单易行和便捷高效,是目前位置隐私保护研究领域使用最广泛的位置隐私保护技术。位置 k 匿名技术的核心思想是将用户位置隐藏在其他至少 $k-1$ 个位置中,使得攻击者难以分辨出用户真实位置,从而保护用户位置隐私。k 为隐私保护度度量参数,k 越大,隐私保护度就越强。

Gruteser 最先使用四叉树搜索算法将用户位置泛化成满足 k 匿名的匿名区域,虽然该方法能够快速实现隐私保护,但过剩匿名区增加了资源开销[3]。文献[6]采用随机策略在地图上选择生成 k 匿名集,解决了匿名区过剩问题,但是隐私保护度和服务质量低。在此基础上,Zhu 等结合缓存机制设计了 MobileCache 协作系统,进一步降低了资源开销[7]。叶阿勇等利用服务相似性提高了匿名区的生成效率和查询请求的服务质量,但是难以抵御背景知识推理攻击[8-9]。为解决面临大量用户请求时服务器端匿名效率低下的问题,文献[10]采用 Voronoi 图划分法,结合中心服务器组织用户实现了协作匿名,但是该方法对用户运动状态的预估不够精细,且频繁的协作匿名增加了用户端的资源开销。吴荻等利用无线网中用户连接在一定圆形区域内的特点,在无线网覆盖区域选择 k 个虚假位置作为匿名集[11]。

虽然这些方法都对经典的 k 匿名方法做出了不同程度的改进,但是都假设攻击者没有与用户相关的背景知识,因此不能很好地抵御背景知识推理攻击。为此,Niu 等结合背景信息提出了一种 k 匿名位置选择算法,在全局范围内选择与用户位置更相似的位置点生成 k 匿名集,提高了隐私保护度[12]。Zhang 等提出了抵御关联匹配攻击的 Cor-k 算法,用户在匿名区内寻找满足节点相关性阈值的位置生成匿名集[13]。He 等提出了基于空间分集的隐私保护方法,用户可利用空间多样性的分集保护位置隐私[14]。虽然这些方法考虑了一定的背景信息,但是对隐私保护度的度量过于单一。为此,将位置熵度量[15]应用在位置隐私保护方法的设计中,以提高隐私保护度[16-18],但已有方法没能在位置隐私和服务可用性之间取得更好的权衡,且资源开销大。

针对这些问题,本文提出了一种能抵御背景知识推理攻击的服务相似性位置 k 匿名隐私保护方法。该方法首先根据服务相似性生成能够保证服务质量的匿名候选区;然后充分考虑攻击者背景知识,采用贪心策略在匿名候选区中选择生成隐私保护度最强的 k 匿名集;最后在既能保证服务质量又能保证隐私保护度的 k 匿名集中,随机选择一个位置点代替用户真实位置请求查询。该方法可有效抵御背景知识推理攻击,在位置隐私、服务可用性和资源开销之间取得了更好的权衡。

1　准备工作

1.1　系统模型

本文系统模型采用集中式隐私保护架构,即使用一个可信第三方(TTP)来实现和执行位置隐私保护机制,TTP 的高性能处理能力能够帮助用户快速实现位置隐私保护。

具体的系统架构图如图 1 所示。LBS 服务器拥有所有地图数据以及地图中每个位置单元的历史查询数,因此它可根据服务相似性定义,预先生成一个包含历史查询数的标签相似地图,然后发送给

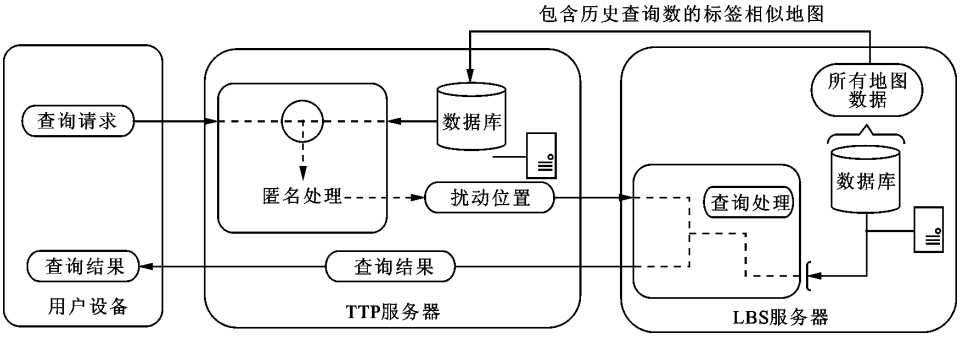


图 1 系统模型

TTP。标签相似地图中包含了区域间的服务相似度,可帮助 TTP 生成能够保证查询结果精确性的匿名候选区。当 TTP 收到用户的查询请求后,首先根据本文提出的位置隐私保护方法对用户的查询请求进行匿名处理,然后将生成的扰动位置发送给 LBS 服务器进行查询处理,最后 LBS 服务器将查询结果返还给 TTP,由 TTP 将查询结果发送给用户,完成查询服务。

1.2 基本概念

1.2.1 服务相似性 用户在获取 LBS 服务时发现,相近位置的查询结果具有一定的相似性,位置之间越相近,查询结果也更相似,甚至完全相同,即具有服务相似性[9]。图 2 是服务相似性示意图,用户 U_A 和用户 U_B 分别在位置 A、B 处,在查询距离自己最近的肯德基(KFC)时,得到相同的查询结果 KFC-1,说明位置 A、B 具有很强的服务相似性。

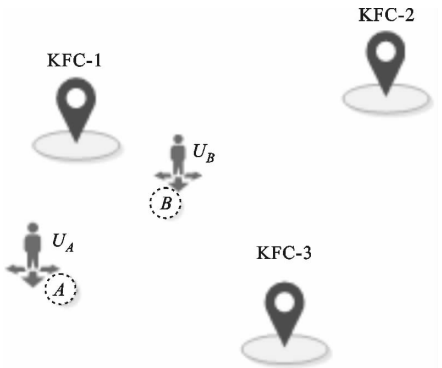


图 2 服务相似性示意图

根据查询结果的相似性来度量两位置间的服务相似度 ρ,公式为

$$\rho = S(A,B) = S((x_A,y_A),(x_B,y_B)) = \frac{F_w(x_A,y_A) \cap F_w(x_B,y_B)}{w} \tag{1}$$

式中:$S(\cdot)$为服务相似度计算函数;$F_w(x,y)$表示对位置坐标(x,y)处查询结果排序后的前 w 个查询结果集;w 为查询结果数;$0 \leqslant \rho \leqslant 1$。查询函数和排序规则均由 LBS 服务器定义,本文默认使用欧氏距离排序。

1.2.2 背景知识 本文背景知识为用户在特定位置的查询概率,即某个位置单元中提交某类查询请求的历史查询概率。在实际查询中,不同位置查询不同内容的概率是不同的,例如人们大概率会在市中心查找商场、电影院等,而很少在此查询加油站等。因此,每个位置在过去都有发出某类查询请求的历史查询概率。攻击者可通过获取这类背景信息推断出匿名集中用户的真实位置等用户敏感信息。

为此,本文在设计位置隐私保护方法时,考虑查询概率背景信息以抵御推理攻击。根据 LBS 服务器中的历史查询数来量化查询概率。为方便处理,将地图划分为 $n \times n$ 的网格,每个网格代表一个大小相等的位置单元。每个位置单元都有对应的历史查询概率,公式为

$$P_{hi} = \frac{N_{qi}}{N_q},\ i = 1,2,3,\cdots,n^2 \tag{2}$$

式中:P_{hi}为位置单元 i 的历史查询概率,$\sum_{i=1}^{n^2} P_{hi} = 1$;$N_{qi}$为位置单元 i 中某类查询的历史查询数;N_q为地图上所有位置单元的历史查询总数。

1.2.3 位置熵 使用位置熵来度量隐私保护强度,熵越大,攻击者从匿名集中识别出用户位置的不确定性就越大。因此,在匿名候选区中选择生成熵最大的 k 匿名集来提高隐私保护度。位置熵的公式为

$$E = -\sum_{j=1}^{k} p_j \mathrm{lb} p_j \tag{3}$$

式中:E 为位置熵;p_j 为匿名集中位置单元的查询概率,公式为

$$p_j = \frac{P_{hj}}{\sum_{j=1}^{k} P_{hj}} \tag{4}$$

其中 P_{hj}为匿名集中第 j 个位置单元的查询概率。

当匿名集中 k 个位置单元都具有相同的识别概

率 $1/k$ 时,最大熵为 $E_{\max}=\mathrm{lb}k$。

2　位置 k 匿名隐私保护策略

2.1　标签相似地图的生成

为方便处理,将地图划分成一个 $n\times n$ 的网格地图 M,M 中的每个网格代表一个位置单元 m_{ij},$M=\{m_{ij}\mid i,j=1,2,3,\cdots,n\}$。LBS 服务器首先根据查询函数 F 和已知的兴趣点位置信息计算每个位置单元的查询结果,并使用排序函数得到每个位置单元的前 w 个查询结果,默认查询距离用户最近的前 w 个兴趣点。然后,根据式(1)计算每个位置单元间的服务相似度 ρ,将 $\rho=1$ 的位置单元合并成同一分区,并给该分区的位置打上相同标签。

由此,LBS 服务器根据服务相似性将地图 M 划分为包含不同分区的标签相似地图,记为 T_{SM},$T_{\mathrm{SM}}=\{z_1,z_2,z_3,\cdots,z_q\}$,$1\leqslant q\leqslant n^2$,$T_{\mathrm{SM}}$ 中的标签为每个分区的区号。同时,将每个位置单元的历史查询数也附在 T_{SM} 上,生成包含历史查询数的 T_{SM}。图 3 是包含 5 个兴趣点的 5×5 的网格地图 M。

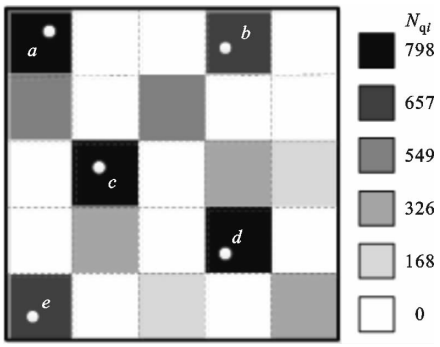


图 3　包含 5 个兴趣点的网格地图

以图 3 为例,生成包含历史查询数的 T_{SM} 的步骤如下。

(1)生成查询结果表。设 $w=3$,根据函数 $F_w(x,y)$ 计算出每个位置单元的查询结果,得到与位置单元一一对应的查询结果表,如表 1 所示。

表 1　查询结果表

i	查询结果				
	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
1	$\{a,c,b\}$	$\{a,b,c\}$	$\{b,a,c\}$	$\{b,c,a\}$	$\{b,d,c\}$
2	$\{a,c,e\}$	$\{c,a,b\}$	$\{b,c,a\}$	$\{b,d,c\}$	$\{b,d,c\}$
3	$\{c,a,e\}$	$\{c,a,d\}$	$\{c,d,b\}$	$\{d,b,c\}$	$\{d,b,c\}$
4	$\{e,c,a\}$	$\{c,e,d\}$	$\{d,c,e\}$	$\{d,c,b\}$	$\{d,b,c\}$
5	$\{e,c,d\}$	$\{e,c,d\}$	$\{d,e,c\}$	$\{d,c,e\}$	$\{d,c,e\}$

(2)计算服务相似度 ρ。根据表 1 的查询结果和式(1),计算出每两个位置单元间的 ρ。

(3)生成 T_{SM}。给 $\rho=1$ 的位置单元打上相同标签合并成同一分区,得到包含有不同分区的 T_{SM}。

(4)生成包含历史查询数的 T_{SM}。LBS 服务器统计每个位置单元的历史查询数 N_{qi},并附在 T_{SM} 上生成包含历史查询数的 T_{SM},如图 4 所示。

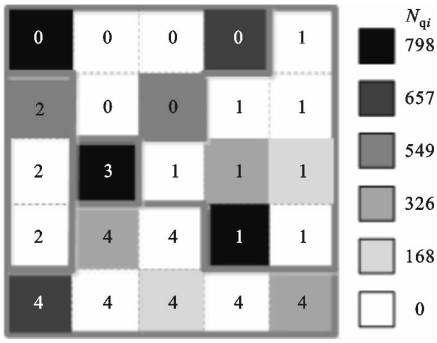


图 4　包含历史查询数的标签相似地图

(5)生成分区相似度表 T。根据表 1 的查询结果和式(1)计算每个分区间的 ρ,生成不同分区服务相似度表 T,如表 2 所示。

表 2　不同分区服务相似度表

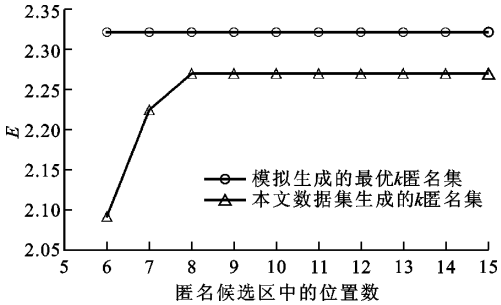
分区	ρ				
	z_1	z_2	z_3	z_4	z_5
z_1	1.00	0.67	0.67	0.67	0.33
z_2	0.67	1.00	0.33	0.67	0.67
z_3	0.67	0.33	1.00	0.67	0.67
z_4	0.67	0.67	0.67	1.00	0.67
z_5	0.33	0.67	0.67	0.67	1.00

从图 3 可以看出,地图 M 中包含 a、b、c、d、e 共 5 个兴趣点,并且每个位置单元都有相应的历史查询数。表 1 显示了每个位置单元的查询结果。据此,根据(2)~(4)步可以生成如图 4 所示的包含历史查询数的标签相似地图 T_{SM},地图中共有 5 个分区 z_1~z_5,分别用 0~4 的分区号标识。TTP 可以根据此地图上的 N_{qi} 利用式(2)计算得到每个位置单元的历史查询概率 P_{hi},并结合用户精确位置确定用户在 T_{SM} 中的分区,之后利用表 2 中不同分区的服务相似度构建位置隐私保护方法。

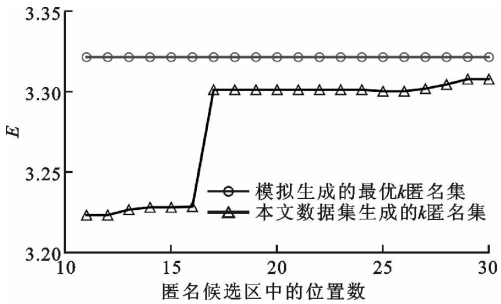
2.2　位置 k 匿名隐私保护方法

2.2.1　匿名候选区生成参数选取　为了提高隐私保护度,已有的位置隐私保护方法(如文献[12]中的方法)在地图的全局范围内选取与用户位置相似的位置点生成匿名集,忽略了对服务质量的考量。为

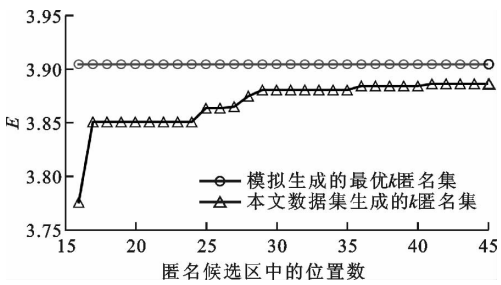
了更好地保证用户获取 LBS 服务的可用性，本文根据服务相似性生成能够保证服务质量的匿名候选区 A_{CR}，结合攻击者背景信息在该区域内选取满足最大熵的 k 匿名集 A_S，从而确保隐私保护度。为了保证在生成的 A_{CR} 中选取到熵最大的 k 匿名集，设计了实验对 A_{CR} 的生成参数进行验证选择。实验条件如下：为实现 k 匿名，需保证 A_{CR} 中的位置数大于 k；A_{CR} 中包含的位置数要能够保证生成熵最大的 k 匿名集。实验结果如图 5 所示。



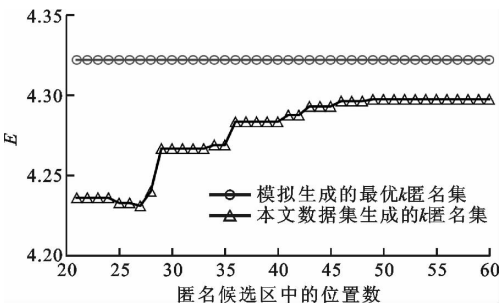
(a)$k=5$



(b)$k=10$



(c)$k=15$



(d)$k=20$

图 5 匿名候选区位置数与 k 匿名集熵的关系

该实验考察了 $k=5,10,15,20$ 时生成的 A_{CR} 中的位置数与 k 匿名集熵之间的关系，A_{CR} 中的位置数变化范围为 $(k+1)\sim 3k$。在实验中：模拟生成的最优 k 匿名集是指匿名集中位置点历史查询概率都相等，能得到最大熵的匿名集；本文生成的 k 匿名集是在本文使用的数据集上生成的 A_{CR} 中选择生成的熵最大的 k 匿名集。

分析图 5 发现，随着 A_{CR} 中位置数的增加，k 匿名集的熵也不断增大，但都在位置数达到 $2k-2$ 时逼近最优 k 匿名集的熵，并保持稳定。这是因为当 A_{CR} 中的位置数过少时，与用户位置查询概率相近的位置点较少，不足以生成熵最大的 k 匿名集。实验结果表明，A_{CR} 包含的位置数大于等于 $2k-2$ 时，即可生成熵接近于最优熵的 k 匿名集，并且每种情况生成的 k 匿名集的最大熵与最优 k 匿名集的熵仅差 0.02～0.05。虽然 A_{CR} 中的位置数在 $2k-2$ 的基础上不断增加会生成熵更加接近、甚至等于最优熵的 k 匿名集，但同时也会增大计算开销，且熵的增加空间较小。因此，为了更好地保证位置隐私保护方法各方面的性能，将 A_{CR} 的生成条件设置为 A_{CR} 中的位置数大于等于 $2k+1$。

2.2.2 匿名候选区生成算法 用户在请求查询时，需要向 TTP 发送包含自己精确位置、查询内容以及隐私参数 k 等的查询请求。当 TTP 接收到用户的查询请求后，根据获取到的用户精确位置等执行相应的位置隐私保护机制。

为更好地保证位置隐私保护方法的位置服务可用性，根据服务相似性生成能够保证服务质量的匿名候选区 A_{CR}。匿名候选区生成算法的伪代码如下。

1. TTP 结合 T_{SM} 上的历史查询数，根据式(2)计算得到每个位置单元的 P_{hi}
2. 根据用户位置 L_u 在 T_{SM} 上确定 z_u
3. 根据 T 将其他分区按照与 z_u 的服务相似度进行从大到小的排序
4. $A_{CR}\leftarrow z_u$
5. WHILE A_{CR} 中的位置数 $L_{count}(A_{CR})<2k+1$
6. 根据第 3 步排序后的分区顺序依次取出分区 z_i
7. IF $T(z_u,z_i)\geqslant\rho$
8. $A_{CR}\leftarrow z_i$
9. END IF
10. END WHILE
11. RETURN 匿名候选区 A_{CR}

在 A_{CR} 生成算法中，TTP 首先根据用户位置 L_u 在 T_{SM} 上确定用户所在分区 z_u，然后根据分区相似

度表 T 将其他分区按照与 z_u 的服务相似度进行降序排序。初始化 A_{CR},将 z_u 作为初始 A_{CR}。当 A_{CR} 中的位置数不足 $2k+1$ 时,根据第 3 步得到的分区排序序列依次取出其他分区 z_i,当 z_i 与 z_u 的服务相似度 ρ 大于等于设定的 ρ 时,将其放入 A_{CR} 中,直到满足第 5 步的条件,生成能够保证特定服务质量的 A_{CR}。

2.2.3　*扰动位置生成算法*　在匿名候选区生成算法的基础上,充分考虑攻击者的背景知识,生成能够抵御背景知识推理攻击的熵最大的最优 k 匿名集,使隐私保护方法既能保证服务质量又能保证隐私保护度。扰动位置生成算法的伪代码如下。

1. 在 A_{CR} 中取出用户位置 L_u
2. $A_S \leftarrow L_u$
3. FOR ($i=1; i \leqslant k-1; i++$) DO
4. $v \leftarrow 0$
5. 初始化最大熵 $E_{max} \leftarrow 0$
6. FOR ($j=1; j \leqslant$ sizeof(A_{CR}); $j++$) DO
7. $A_S \leftarrow A_{CR}[j] \cup A_S$
8. 根据式(4)和 P_{hi} 计算 A_S 中的归一化查询概率 p_i
9. 根据式(3)和 p_i 计算 $E(A_S)$
10. IF $E(A_S) > E_{max}$
11. $E_{max} \leftarrow E(A_S)$
12. $v \leftarrow j$
13. END IF
14. $A_S \leftarrow A_S / A_{CR}[j]$
15. END FOR
16. $A_S \leftarrow A_{CR}[v]$
17. $A_{CR} \leftarrow A_{CR} / A_{CR}[v]$
18. END FOR
19. $\forall L_p \in A_S$
20. RETURN 扰动位置 L_p

扰动位置生成算法在生成 k 匿名集时,首先将用户位置放入匿名集 A_S 中,然后以用户位置查询概率为基准,使用贪心策略将 A_{CR} 中与用户位置查询概率最相近的位置依次放入 A_S 中,生成最优 k 匿名集,最后在该 k 匿名集中随机选取一个扰动位置 L_p 发送给 LBS 服务器进行查询,降低通信开销。贪心策略的使用降低了算法时间复杂度,通过分析可知,扰动位置生成算法的时间复杂度仅为 $O(k^2)$。由于 A_{CR} 可以保证用户服务的可用性,因此生成的 k 匿名集既能保证服务质量又能保证隐私保护度。

2.3　安全性分析

攻击者的攻击目标是获取特定用户的真实位置,进而推断出个人敏感信息,可将攻击者根据执行手段分为被动攻击者和主动攻击者。任何通过窃听、监听实体之间信道来了解特定用户额外敏感信息,从而危害用户安全的攻击者都被视为被动攻击者。为了获取其他用户更多敏感信息而攻击和破坏 LBS 服务器的攻击者则被视为主动攻击者,这类攻击者可以获取 LBS 服务器上的全局信息,包括特定用户的历史查询数据和当前查询,并知道 LBS 系统中使用的位置隐私保护机制。为了方便起见,将 LBS 服务器视为主动攻击者,它可根据已知的全局信息推断和学习所有用户的敏感信息。

公钥基础设施技术可以保证无线信道传输内容的安全性,因此本文不考虑无线信道窃听的被动攻击,只关注主动攻击引起的推理攻击。推理攻击是指攻击者根据获取到的 k 匿名集或扰动位置信息,利用自己掌握的背景信息等推理出用户位置隐私的一种攻击方式。

LBS 服务器知道所有的地图数据、兴趣点分布、生成的包含历史查询数的 T_{SM}、TTP 提交的扰动位置 L_p、系统使用的位置隐私保护机制和相关参数。当 LBS 服务器获取到扰动位置 L_p 后,试图根据掌握的背景信息结合位置隐私保护机制推断出用户真实位置,但这将失败。这是因为本文提出的位置隐私保护方法在生成 k 匿名集时,考虑了攻击者背景信息,在冗余的匿名候选区 A_{CR} 中选择生成了熵最大的 k 匿名集,匿名集中所包含的位置点历史查询概率都很相近,攻击者只能从这 k 个位置中随机猜测用户位置,猜中的概率不会大于 $1/k$,即攻击者很难根据掌握的背景信息推断出 k 匿名集中用户真实位置,因此推理攻击将会失效。本文方法能够很好地抵御具有背景信息的推理攻击。

3　实验与性能评估

采用北美道路路网 NA 数据集[19]来验证本文所提方法的性能。NA 路网数据集包含 178 512 个兴趣点,如图 6 所示,图中 x、y 坐标分别是经度、纬度转换成的直角坐标。所有算法使用 Python 3.5 编程实现,操作系统为 Windows 10 专业版,计算机 CPU 型号为 Intel i7,内存容量为 64 GB。

3.1　数据集处理

LBS 服务器可以从历史查询日志中获取位置单元的历史查询数,直观上看,用户偏向于在兴趣点

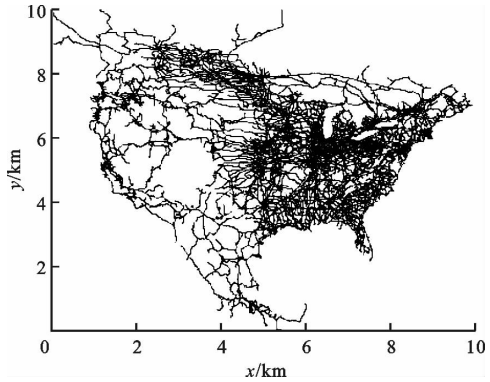


图 6 NA 数据集[19]

集中的地方请求查询,因此在获取不到历史查询日志时,可以使用该位置兴趣点数近似代替历史查询数[20]。

实验过程中,使用 NA 数据集中的兴趣点数模拟位置单元的历史查询数。将 NA 数据集划分成 100×100 的网格地图,每个位置单元的兴趣点数为该位置单元的历史查询数 N_{qi}。根据式(2)计算出每个位置单元相应的历史查询概率 P_{hi},将 P_{hi} 作为先验概率在生成的网格地图中放置 1 000 个兴趣点,每个位置单元生成 1 个兴趣点。

3.2 性能度量指标

采用 4 个度量指标来评估本文方法的性能。

(1)隐私保护度度量指标。使用位置熵 E 度量相应位置隐私保护方法的隐私保护度。

(2)位置服务可用性度量指标。使用查询精确度 Q 度量位置服务可用性,计算式为

$$Q=\frac{|R(L_{\mathrm{u}})\cap R(A_{\mathrm{S}})|}{|R(A_{\mathrm{S}})|}=|R(L_{\mathrm{u}})\cap (R(L_1)\cup R(L_2)\cup R(L_3)\cup\cdots\cup R(L_k))|/|R(L_1)\cup R(L_2)\cup R(L_3)\cup\cdots\cup R(L_k)| \tag{5}$$

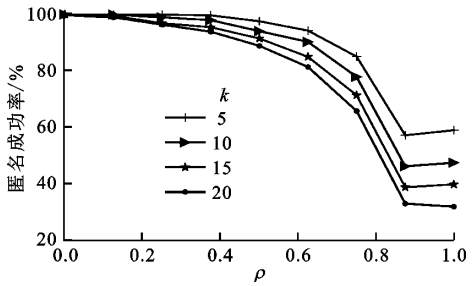
式中:$R(A_{\mathrm{S}})$ 为 A_{S} 的查询结果集;$R(L_{\mathrm{u}})$ 为用户精确位置查询结果;Q 为 $R(A_{\mathrm{S}})$ 中包含 $R(L_{\mathrm{u}})$ 的比例,Q 越大则表示查询结果越精确。

(3)时间开销度量指标。使用位置隐私保护方法完成匿名所用的时间来度量时间开销,单位为 s。

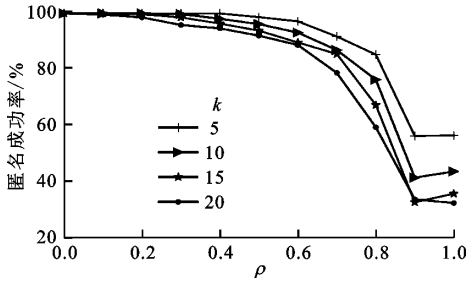
(4)通信开销度量指标。通信开销一般指与 LBS 服务器进行通信的数据量,因此使用 k 匿名集或扰动位置获取到的查询结果数来度量通信开销。

3.3 性能评估实验

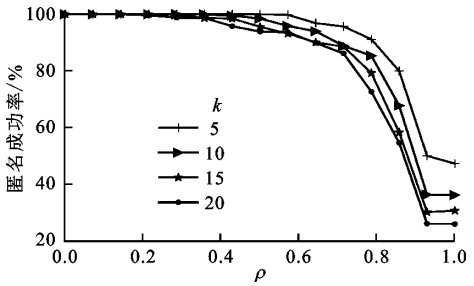
3.3.1 *服务相似度 ρ 与匿名成功率的关系* 在生成 A_{CR} 时,每次将与用户所在分区 z_u 的服务相似度 ρ 最大的分区 z_i 合并在 A_{CR} 中。为了在用户位置隐私和服务可用性之间取得更好的权衡,研究了在不同 w 条件下,生成 A_{CR} 的过程中 ρ 与匿名成功率之间的关系,实验结果如图 7 所示。



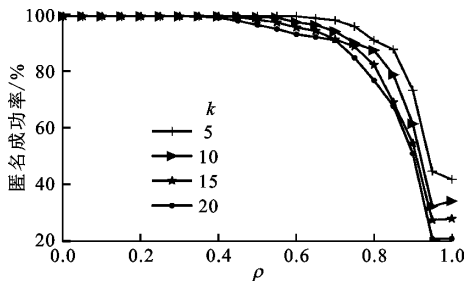
(a)$w=8$



(b)$w=10$



(c)$w=14$



(d)$w=20$

图 7 服务相似度 ρ 与匿名成功率的关系

从图 7 可以看出,当 ρ 增大 k 不变或 k 增大 ρ 不变时,匿名成功率都不断减小。这是因为与用户所在分区 ρ 越大的分区位置点越少,就越难以满足 A_{CR} 的生成条件;k 越大,生成 A_{CR} 时对位置点的需求也越大,因此匿名成功率也越小。但是,在相同 k 和 ρ 下,随着查询结果数 w 的增大,匿名成功率不

断增大。这是由于 w 越大,用户查询的搜索范围也越大,使得对分区和服务相似度的划分也更加细致,因此在相同的 k 和 ρ 下可以获得更高的匿名成功率。

进一步分析发现,当 ρ 低于 0.5 时,在不同的 w 和 k 下,匿名成功率都能达到 85%以上。但是,过低的服务质量对用户而言是没有意义的,因此考察 ρ 在 0.5 以上的匿名成功率。当 ρ 达到 0.9～1.0 时,匿名成功率最小,这是因为此时与用户位置查询结果相同的位置单元较少,难以达到匿名需求。当 $0.5\leqslant\rho\leqslant0.8$ 时,匿名成功率可以达到一个较高水平,并且当 $w>10$ 时,匿名成功率可以达到平均至少 80%,所以为了更好地保护用户位置隐私,LBS 服务器可以在生成 T_{SM} 时选择较大的 w,以确保位置隐私保护方法的有效性。

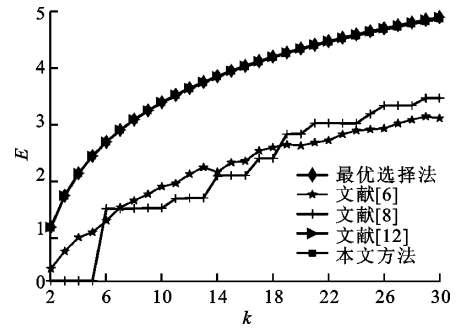
根据以上研究分析发现,当 $0.5\leqslant\rho\leqslant0.8$ 时,本文位置隐私保护方法可以在用户位置隐私和服务可用性之间取得较好的权衡,用户也可以根据自己对位置隐私和服务质量的需求选择更适合自己的 k 和 ρ。由于在 $w=20$、$\rho=0.5$ 时,服务质量和匿名成功率之间能够取得很好的平衡,因此下面所有实验都是在此参数条件下进行的。

3.3.2　位置隐私保护方法性能对比　为了更好地验证本文方法的性能,与最优选择法、文献[6][8][12]中的方法进行对比。如果生成的 k 匿名集中位置点的查询概率都相等,则 k 匿名集熵最大,隐私保护度最强。据此,本文将生成这种最优匿名集的方法定义为最优选择法,以此为基准来衡量其他方法的隐私保护强度。

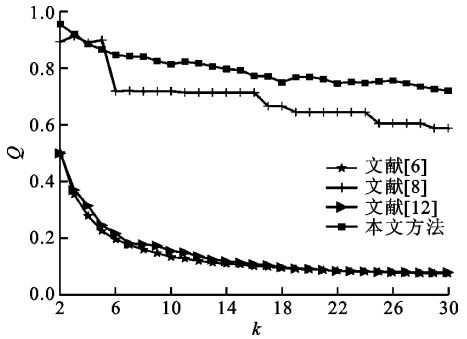
考察在相同隐私保护参数 $k(2\leqslant k\leqslant30)$ 下,各方法在隐私保护度、服务可用性、时间开销和通信开销 4 个方面的性能。每组实验的结果都是在统计 100 次实验结果后求得的平均值。

文献[8]是目前最新的基于服务相似性定义设计的位置隐私保护方法,本文与之进行对比。由于在实际情况中,真实用户的分布是不均匀的,因此根据位置单元中真实用户数作为扩充匿名区条件的方法会导致隐私保护性能的不稳定。用户分布过于集中或分散都会降低隐私保护方法性能,甚至导致匿名失败。因此,在本文实验中,分别在用户分布不均匀和用户均匀分布这两种场景下,对各方法的性能进行对比分析。

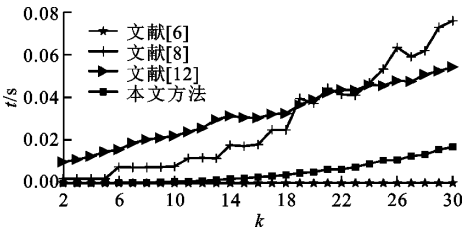
用户分布不均匀场景下的实验结果如图 8 所示,该场景是给每个位置单元随机生成 0～3 个真实用户。



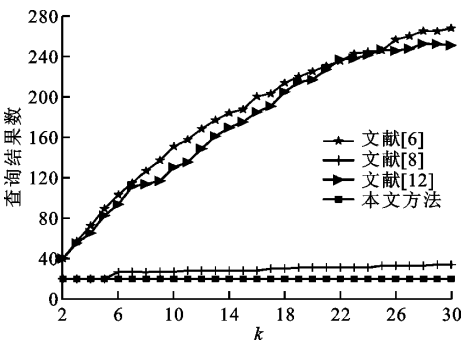
(a)隐私保护度



(b)位置服务可用性



(c)时间开销



(d)通信开销

图 8　用户分布不均匀场景下的隐私保护方法性能对比

从图 8a 可以看出,随着 k 的增大,各方法生成的 k 匿名集的熵也在不断增大,即隐私保护度增强。其中,最优选择法的隐私保护度最强,文献[12]的熵与最优选择法最接近,本文方法的熵也仅比最优选择法平均低 0.5%,说明本文方法可以提供很高的隐私保护度。这是因为本文方法在匿名时充分考虑

了攻击者的背景信息，在冗余的 A_{CR} 中选择生成了熵最大的最优的 k 匿名集。由于文献[6]仅在全局范围内随机选择匿名位置完成匿名，没有充分考虑背景信息，所以导致其隐私保护度比最优选择法平均低 47.45%。因为用户分布不均匀，所以文献[8]生成的匿名区在一定的 k 范围内是不变的，且也没有充分考虑攻击者背景信息，造成了隐私保护度低。在用户分布不均匀场景下，本文方法的隐私保护强度比文献[6]和文献[8]平均高 46.1%和 50.88%。

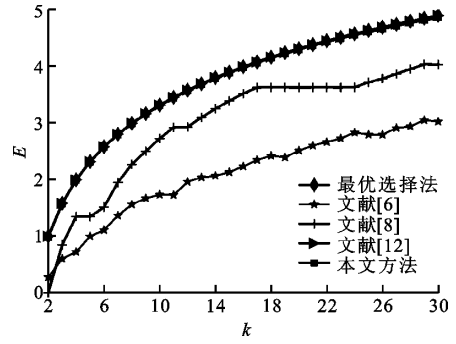
从图 8b 可以看出，随着 k 的增大，Q 会降低。本文方法在生成的既能保证服务质量又能保证隐私保护度的 k 匿名集中选择扰动位置进行查询，因此匿名后可以获得很高的 Q。文献[8]根据位置单元中的用户数生成匿名集，所以匿名集中位置点距离较近，使得该方法的 Q 比文献[6]和[12]的高。文献[6]和[12]在全局范围内选取匿名集，因此它们的 Q 比文献[8]的分别平均低 81.11%和 79.86%。本文方法的 Q 比文献[8]的平均高 14.83%，比文献[6]和[12]的分别平均高 83.16%和 82.03%。

随着 k 的增大，文献[8]要不断扩大范围搜索用户来满足匿名条件，所以从图 8c 可以看出，该方法的时间开销增长迅速，且存在突增的情况。但是，其他方法匿名时不用考虑用户分布情况，所以时间开销增长比较平缓。

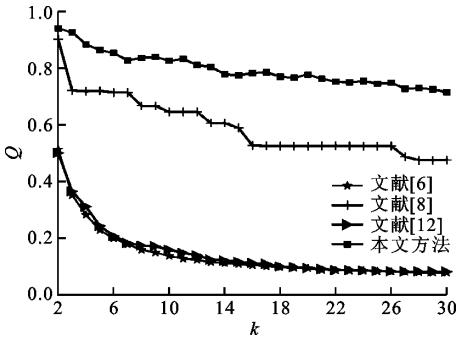
图 8d 显示了在 $2\leqslant k\leqslant 30$ 时各方法的通信开销，可以看到，在全局范围内选择生成 k 匿名集的文献[6]和[12]方法没有很好地兼顾服务质量，具有很大的通信开销。当 $k=30$ 时，它们的查询结果数达到了 275，极大的通信开销严重降低了位置隐私保护方法的性能。文献[8]的匿名集中位置点距离较近，通信开销较小，但是随着 k 的增大也呈现出逐渐增大的趋势。由于本文方法只提交了一个扰动位置进行查询，因此其通信开销最小。

用户均匀分布场景下的实验结果如图 9 所示。从图 9a 可以看出，即使用户均匀分布，文献[8]依然不能提供很高的隐私保护度，虽然较文献[6]的隐私保护度高了 44.65%，但依然比本文方法低 28.33%。而且，从图 9b 可以看出，在该场景中，由于用户均匀分布导致文献[8]的匿名范围变大，所以查询精确度 Q 也随之降低，比本文方法低了 36.21%。图 9d 则显示了在该场景下，本文方法的通信开销依然最低。

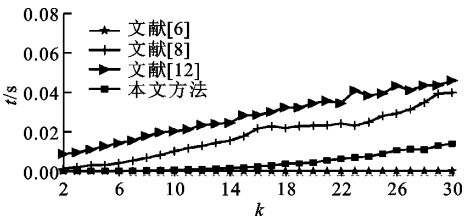
从图 8c 和 9c 可以看出，在时间开销方面，采用全局穷举策略实现匿名的文献[12]的时间开销远高于其他方法，而本文采用贪心策略在范围较小的



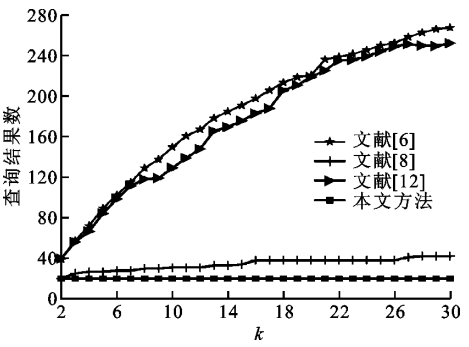
(a)隐私保护度



(b)位置服务可用性



(c)时间开销



(d)通信开销

图 9 用户均匀分布场景下的隐私保护方法性能对比

A_{CR} 中完成匿名，因此时间开销比文献[8]和[12]分别平均低 83.64%和 88.6%。由于文献[6]只采用了随机选择方案生成匿名集，没有复杂的条件约束，所以时间开销最低。

根据以上分析可知，与已有方法相比，本文方法在用户位置隐私和服务可用性之间取得了更好的权

衡,并有效降低了时间开销和通信开销。此外,本文方法在生成 k 匿名集时充分考虑了攻击者的背景信息,能够有效抵御具有背景信息的推理攻击,提供了更强的隐私保护度。

4　结　论

针对已有位置 k 匿名方法不能很好地权衡用户位置隐私和位置服务可用性之间的关系、在遭遇背景知识推理攻击时存在隐私泄露风险且资源开销大的问题,提出了一种新的基于服务相似性的位置 k 匿名隐私保护方法,本文的主要结论如下:

(1)引入服务相似性生成了标签相似地图,并根据服务相似度将用户所在分区与其他分区合并,生成了能够保证服务质量的匿名候选区;

(2)研究了在不同查询结果数条件下服务相似度和匿名成功率之间的关系,构建了考虑位置隐私、服务质量和资源开销的统一优化模型;

(3)采用位置熵对攻击者背景知识进行量化,生成了能够抵御背景知识推理攻击的 k 匿名集;

(4)采用贪心策略在能够保证服务质量的匿名候选区中选择生成熵最大的最优 k 匿名集,并随机选择一个位置点请求服务,既保证了服务质量和隐私保护度,又有效降低了资源开销。

安全性分析和实验结果表明,本文方法能有效抵御背景知识推理攻击,且与已有方法相比,隐私保护度和服务质量平均提高了 48.49%和 60.01%,时间开销和通信开销平均降低了 86.12%和 66.57%,在用户位置隐私、位置服务可用性和资源开销之间取得了更好的权衡。

参考文献:

[1] 张学军,桂小林,伍忠东. 位置服务隐私保护研究综述 [J]. 软件学报,2015,26(9):2373-2395.
ZHANG Xuejun, GUI Xiaolin, WU Zhongdong. Privacy preservation for location-based services: a survey [J]. Journal of Software, 2015, 26(9): 2373-2395.

[2] 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.
WAN Sheng, LI Fenghua, NIU Ben, et al. Research progress on location privacy-preserving techniques [J]. Journal on Communications, 2016, 37(12): 124-141.

[3] 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. New York, USA: ACM, 2003: 31-42.

[4] YI X, PAULET R, BERTINO E, et al. Practical k nearest neighbor queries with location privacy [C] // 2014 IEEE 30th International Conference on Data Engineering. Piscataway, NJ, USA: IEEE, 2014: 640-651.

[5] DWORK C, ROTH A. The algorithmic foundations of differential privacy [J]. Foundations and Trends® in Theoretical Computer Science, 2014, 9(3/4): 211-407.

[6] KIDO H, YANAGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services [C] // Proceedings of the International Conference on Pervasive Services. Piscataway, NJ, USA: IEEE, 2005: 88-97.

[7] ZHU Xiaoyan, CHI Haotian, NIU Ben, et al. MobiCache: when k-anonymity meets cache [C] // Proceedings of Global Communications Conference. Piscataway, NJ, USA: IEEE, 2013: 820-825.

[8] 叶阿勇,李亚成,马建峰,等. 基于服务相似性的 k-匿名位置隐私保护方法 [J]. 通信学报,2014,35(11):162-169.
YE Ayong, LI Yacheng, MA Jianfeng, et al. Location privacy-preserving method of k-anonymous based on service similarity [J]. Journal on Communications, 2014, 35(11): 162-169.

[9] DEWRI R, THURIMELLA R. Exploiting service similarity for privacy in location-based search queries [J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(2): 374-383.

[10] 马春光,周长利,杨松涛,等. 基于 Voronoi 图预划分的 LBS 位置隐私保护方法 [J]. 通信学报,2015,36(5):5-16.
MA Chunguang, ZHOU Changli, YANG Songtao, et al. Location privacy-preserving method in LBS based on Voronoi division [J]. Journal on Communications, 2015, 36(5): 5-16.

[11] 吴荻,张玉,刘银龙,等. 面向基于位置服务的一种改进型隐私感知虚假位置选择机制 [J]. 信息安全学报,2018,3(2):62-75.
WU Di, ZHANG Yu, LIU Yinlong, et al. An improved privacy-aware dummy location selection scheme in LBSs [J]. Journal of Cyber Security, 2018, 3(2): 62-75.

[12] NIU Ben, LI Qinghua, ZHU Xiaoyan, et al. Achieving k-anonymity in privacy-aware location-based services [C] // Proceedings of the 33th IEEE Conference on

Computer Communications. Piscataway, NJ, USA: IEEE, 2014: 754-762.

[13] ZHANG Lei, YU Lili, LI Jing, et al. Location privacy protection algorithm based on correlation coefficient [C]// Proceedings of the 4th International Conference on Control, Automation and Robotics (ICCAR). Piscataway, NJ, USA: IEEE, 2018: 332-335.

[14] HE Xiaofan, JIN Richeng, DAI Huaiyu. Leveraging spatial diversity for privacy-aware location-based services in mobile networks [J]. IEEE Transactions on Information Forensics and Security, 2018, 13(6): 1524-1534.

[15] JIANG T, WANG H J, HU Y C. Preserving location privacy in wireless LANs [C]// Proceedings of the 5th International Conference on Mobile Systems, Applications and Services. New York, USA: ACM, 2007: 246-257.

[16] SHARMA V, SHEN C C. Evaluation of an entropy-based *k*-anonymity model for location based services [C]// 2015 International Conference on Computing, Networking and Communications. Piscataway, NJ, USA: IEEE, 2015: 374-378.

[17] 冯霞, 刘亚伟. 基于联合熵隐私保护的自适应动态 Mix-zone 方案 [J]. 通信学报, 2018, 39(3): 76-85.
FENG Xia, LIU Yawei. Dynamic mix-zone scheme with joint-entropy based metric for privacy-preserving in IoV [J]. Journal on Communications, 2018, 39(3): 76-85.

[18] NI Lina, TIAN Fulong, NI Qinghang, et al. An anonymous entropy-based location privacy protection scheme in mobile social networks [J]. EURASIP Journal on Wireless Communications and Networking, 2019, 2019(1): 93.

[19] 张学军, 桂小林, 蒋精华. 用户为中心的差分扰动位置隐私保护方法 [J]. 西安交通大学学报, 2016, 50(12): 79-86.
ZHANG Xuejun, GUI Xiaolin, JIANG Jinghua. A user-centric location privacy-preserving method with differential perturbation for location-based services [J]. Journal of Xi'an Jiaotong University, 2016, 50(12): 79-86.

[20] PINGLEY A, ZHANG N, FU X W, et al. Protection of query privacy for continuous location based services [C]// Proceedings of the Computer and Communications Societies. Piscataway, NJ, USA: IEEE, 2011: 1710-1718.

[本刊相关文献链接]

滕晓宇,桂小林,戴慧珺,等. 乱序局部敏感哈希音频零水印方案. 2019,53(9):110-119. [doi:10.7652/xjtuxb201909015]

彭振龙,桂小林. 采用社会网络能力发现的众包服务节点选择算法. 2019,53(11):148-155. [doi:10.7652/xjtuxb201911021]

蒋精华,桂小林,郑宜峰,等. 支持加密搜索的定向息票安全投放框架. 2017,51(2):13-19. [doi:10.7652/xjtuxb201702003]

张磊,马春光,杨松涛,等. 面向路网环境速度预测攻击的隐私保护. 2017,51(2):27-32. [doi:10.7652/xjtuxb201702005]

张文东,桂小林,蔡宁超,等. 群智感知中采用节点社会属性的亲密度量化方法. 2017,51(6):73-78. [doi:10.7652/xjtuxb201706012]

边根庆,翟红,邵必林. 一种采用云模型的同驻虚拟机侧通道攻击威胁度量方法. 2016,50(4):21-27. [doi:10.7652/xjtuxb201604004]

李刘强,桂小林,安健,等. 采用模糊层次聚类的社会网络重叠社区检测算法. 2015,49(2):6-13. [doi:10.7652/xjtuxb201502002]

杨攀,桂小林,安健,等. 利用贝叶斯原理在隐私保护数据上进行分类的方法. 2015,49(4):46-52. [doi:10.7652/xjtuxb201504008]

张学军,桂小林,冯志超,等. 位置服务中的查询隐私度量框架研究. 2014,48(2):8-13. [doi:10.7652/xjtuxb201402002]

贾晓琳,葛建超,綦艳丽,等. 一种 Gaussian-Hermite 矩分块重构的抗几何攻击水印算法. 2014,48(04):20-25. [doi:10.7652/xjtuxb201404004]

杨建伟,桂小林,安健,等. 一种信任关系网络中的社团结构检测算法. 2014,48(12):80-86. [doi:10.7652/xjtuxb201412013]

夏秦,王志文,卢柯. 入侵检测系统利用信息熵检测网络攻击的方法. 2013,47(2):14-19. [doi:10.7652/xjtuxb201302003]

庄威,桂小林,林建材,等. 云环境下基于多属性层次分析的虚拟机部署与调度策略. 2013,47(2):28-32. [doi:10.7652/xjtuxb201302005]

李清华,康海燕,苑晓姣,等. 个性化搜索中用户兴趣模型匿名化研究. 2013,47(4):131-136. [doi:10.7652/xjtuxb201304022]

(编辑 陶晴)