•信息工程•

DOI:10.15961/j.jsuese.201801051

基于局部差分隐私的电动汽车充电位置隐私汇聚

熊星星[1],刘树波[1*],李　丹[1,2],李永凯[1],王　俊[3]

(1.武汉大学 计算机学院,湖北 武汉 430072;2.湖北省水利水电科学研究院,湖北 武汉 430070;
3.中南民族大学 计算机学院,湖北 武汉 430074)

摘　要:电动汽车频繁接入充电桩充电而产生的位置数据对优化充电桩布置、指导电力调度具有重要意义。然而充电位置数据对于汽车用户来说属于隐私信息。为防止汽车用户的隐私泄露,亟需探索研究隐私汇聚充电位置数据的方法。采用局部差分隐私技术保护电动汽车充电位置数据,通过引入贝叶斯随机多伪隐私算法设计一种基于分区的隐私保护充电位置数据汇聚方法。该方法利用贝叶斯随机多伪隐私算法设计了一个用于本地化扰动充电位置数据的局部混淆算法,然后,结合随机多伪算法的重构算法设计了满足稀疏、样本量小等特点的充电位置数据的隐私汇聚方法。同时,在保证隐私保护水平的前提下,通过对位置域进行划分以缩小隐私位置域,进一步提高汇聚结果的可用性。对所设计方法的隐私性进行分析。最后,在正态分布、均匀分布、峰值分布和随机分布4种不同的合成数据集以及公开的Gowalla数据集上进行验证。实验结果表明:在相同隐私水平的条件下,所设计的方法在可用性方面优于基于随机映射矩阵的隐私汇聚方法。

关键词:电动汽车;充电位置;局部差分隐私;隐私保护

中图分类号:TP309.2　　**文献标志码**:A　　**文章编号**:2096-3246(2019)02-0137-07

Private Electric Vehicle Charging Location Aggregation Based on Local Differential Privacy

XIONG Xingxing[1], *LIU Shubo*[1*], *LI Dan*[1,2], *LI Yongkai*[1], *WANG Jun*[3]

(1.School of Computer Sci., Wuhan Univ., Wuhan 430072, China; 2.Hubei Water Resources Research Inst., Wuhan 430070, China;
3.College of Computer Sci., South–Central Univ. for Nationalities, Wuhan 430074, China)

Abstract: The charging location data generated by electric vehicles frequently accessing charging piles for charging are of great significance for optimizing the arrangement of charging piles and guiding the electric power dispatching. However, charging location data are private information for vehicle users. In order to prevent the leakage of the privacy of these users, it is urgent to explore a way of private charging location data aggregation. Therefore, a local differential privacy technology is adopted to preserve the charging location data of electric vehicles. A partition-based privacy preservation charging location data aggregation method is proposed by introducing Bayesian randomized multiple dummies algorithm. The method employs the Bayesian randomized multiple dummies algorithm to design a local obfuscation algorithm for locally perturbing a vehicle's charging location. Then, the private location aggregation method for charging location data with the characteristics of sparseness and small size samples is designed by combining reconstruction algorithm of the randomized multiple dummies algorithm. At the same time, under the premise of ensuring the level of privacy preservation, the whole location domain is divided to narrow the privacy location domain, thereby further improving the utility of aggregation result. The privacy analysis of the proposed method is given. Finally, experimental results on four different synthetic datasets, namely, uniform distribution, normal distribution, peak distribution and random distribution, as well as the public Gowalla dataset are carried out. The experimental results show that the proposed method is superior to the existing randomized projection matrix based private aggregation method in terms of utility under the same privacy level.

Key words: electric vehicles; charging location; local differential privacy; privacy preservation

收稿日期:2018 – 09 – 24

基金项目:国家自然科学基金资助项目(61872431);湖北省技术创新专项资助(2018AAA046);武汉市应用基础研究计划资助项目(2017060201010162)

作者简介:熊星星(1989—),男,博士生. 研究方向:隐私保护、机器学习. E-mail:xiong_xx@whu.edu.cn

* 通信联系人 E-mail:liu.shubo@whu.edu.cn

网络出版时间:2019 – 03 – 13 10 : 55 : 28　　**网络出版地址**:http://kns.cnki.net/kcms/detail/51.1773.TB.20190312.1455.004.html

随着新能源技术的快速发展，电动汽车（electric vehicles，EV）逐渐普及。由于受车载电池技术的限制，EV需频繁地访问充电桩进行充电，在与充电桩进行交互的过程中，数据汇聚器将收集EV的充电位置数据，这些数据可为第三方科研机构和企业提供数据服务。例如，充电服务提供商和国家电网可根据充电位置数据分别对充电桩（站）的布置和输电线路进行优化。EV充电的系统模型[1]包含3个组件：EV、充电桩/站、数据汇聚器，如图1所示。其中：EV是插入式EV（PEV）或插入式混合EV（PHEV）；充电桩是EV充电系统中与EV交互的前端接口；数据汇聚器用于收集EV相关数据，如EV用户的身份和位置信息等。值得注意的是，在所研究模型中，充电桩是可信的，只具有充电和标志充电位置的功能。EV与充电桩交互所产生位置信息是直接或通过充电桩代理传输至数据汇聚器。然而，数据汇聚器可能完全不可信，可能从EV的充电位置数据关联出车主的日常活动轨迹，这将导致充电位置的隐私泄露。为更好地保护EV的安全和隐私，许多学者进行了相关研究[2–4]。

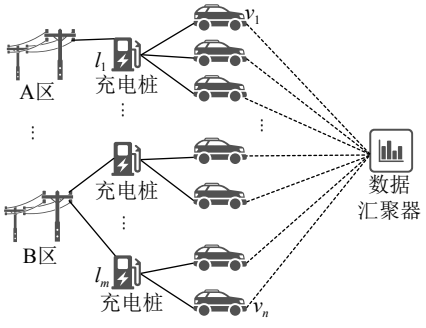


图 1　EV充电系统模型

Fig. 1　Electric vehicle charging system model

Yang等[5]首次阐述了V2G（vehicle-to-grid）的隐私保护问题，为保护EV用户的位置和身份信息，提出了基于身份部分盲签名的隐私保护数据汇聚方法。Jiang等[6]提出了一种针对V2G的关联性可控的群签名的隐私保护数据汇聚方法，抵御关联分析引起的敏感信息泄露。上述基于密码学技术的数据隐私汇聚方法，需要对原始数据进行加密，而后在不可信数据汇聚器上进行解密。虽然能够有效保护数据隐私，但运算代价过高。

Han等[7]针对分布式EV相关数据发布过程中的隐私泄露问题，提出了一种基于差分隐私保护技术的发布算法。Han等[8]利用差分隐私防止EV用户的恶意参与，同时，获得近似真实性以实现EV充电的优化调度。上述研究都是基于中心化差分隐私模型解决EV用户的隐私泄露问题，需假设数据汇聚器半可信（诚实但好奇），并不适用于数据汇聚器完全不可信的场景。

局部差分隐私是一种新颖的、完全分布式的隐私保护模型[9]。Kasiviswanathan等[10]首次提出局部差分隐私的概念，研究局部模型下的可解的学习问题。Erlingsson等[11]最早将局部差分隐私用于解决实际问题，所提出的RAPPOR方法利用随机响应和布隆过滤器技术隐私收集Chrome浏览器的用户设置的统计数据。Bassily和Smith等[12]针对大敏感属性域（大小为k）情形，设计了一个基于随机映射矩阵（random projection matrix，RPM）的局部随机器，其思想是利用随机映射矩阵$\boldsymbol{\Phi}_{m\times k}$实现降维处理（$k \gg m$），以在更低维上实现局部差分隐私算法，可有效降低通信（传输1 bit）和运算（低维）开销。然而，随机映射矩阵方法所带来的噪声比较高，导致数据可用性差。因此，对小属性域的情形，Nguyên等[13]提出了构建随机映射矩阵$\boldsymbol{\Phi}_{k\times k}$以避免降维引入的噪声。Chen等[14]提出一种针对空间位置数据的基于局部差分隐私的隐私计数估计汇聚方法（private count estimation method，PCEM），该方法是基于随机映射矩阵[12]方法实现的，适用于隐私域较大的情况。

Sei等[15]提出一种基于贝叶斯多伪的局部差分隐私算法。该算法是在重构过程中通过增加数据样本量以提高数据可用性。根据大数定律，样本量越大，其统计概率越接近真实值。在实际情况中，充电桩的数量比EV的数量少得多，因此不需要对充电桩位置进行降维处理。此外，EV的充电频率并不高，充电时产生的位置数据在一段时间内（如一个月）较稀疏。

针对EV充电时的隐私位置汇聚问题，作者考虑到EV访问一充电桩（站）进行充电时，不希望暴露真实的充电位置给不可信的数据汇聚器。而是发送伪装的充电位置给汇聚器。汇聚器在收集到全部的位置数据后，却能够重构访问充电桩的EV数量的统计分布。也就是说，不可信的汇聚器在不知道EV真实充电位置（为保护EV用户的位置隐私）的前提下，仍能够近似统计每个充电桩（位置）为多少辆EV提供过充电服务。

为实现这一目标，作者引入一种基于贝叶斯的随机多伪局部差分隐私算法（S2Mb）[15]，设计一种保护EV充电位置的隐私位置汇聚方法（PLAM）。为提高方法的可用性，通过缩小隐私位置域，提出基于分区的隐私位置汇聚优化方法。作者对所设计的方法的隐私性进行了分析。同时，在合成和真实数据集上与基于随机映射矩阵的隐私空间数据汇聚方法（PCEM）进行对比实验，结果表明所设计的PLAM方法在可用性方面优于PCEM方法。

1　问题描述和隐私保护目标

1.1　问题描述

充电桩（站）的位置是公开的。EV访问充电桩充电（充电事件）而产生的位置数据，如果直接被不可信的数据汇聚器收集，则EV的充电位置隐私可能被泄露。问题描述：假定位置域$L=\{l_1,l_2,\cdots,l_k\}$和电动汽车域$V=\{v_1,v_2,\cdots,v_n\}$，其中，L表示全部充电桩的位置，V表示访问充电桩进行充电的全部EV。同时，每辆EV充电时会有一个隐私的充电位置$l_i\in L$、一个隐私预算ε_i和一个隐私位置域τ_i。隐私保护目标：在不知道单辆EV的充电位置的前提下，不可信的数据汇聚器也能获得访问充电桩的EV数量的近似统计分布。

1.2　局部差分隐私

局部差分隐私（local differential privacy, LDP）[10]的形式化定义如下：

定义1　随机算法A满足ε–LDP，对任意输入值$l,l'\in L$和输出域$O\subseteq Range(A)$，满足$\Pr[A(l)\in O]\le e^{\varepsilon}\cdot\Pr[A(l')\in O]$，其中概率是服从关于$A$的0–1分布。

LDP模型是假定数据汇聚者不可信且具有任意背景知识，则在用户侧对数据进行扰动后发送给数据汇聚者，使得扰动结果似是而非。虽然LDP具有严格、可证明的隐私保护特性，但其缺点在于可用性比较差。通常情况下，输入域范围$|L|$比较大，因此在LDP条件下实现合理的隐私性和可用性平衡具有很大挑战。

随机响应（randomized response, RR）技术[16]是实现局部差分隐私的基本方法。该技术最初用于问卷调查的研究，不直接回复敏感问卷信息，而回复似是而非的结果，使收集者无法判定真实信息，但能够得到精确的统计结果。随机响应实现LDP的基本过程：$L=\{l_1,l_2,\cdots,l_k\}$表示可能的敏感充电位置，k表示所有充电位置的数量。某一v_i真位置（true location）是l_i，假设它以概率p向中心汇聚器发送位置l_i；$l_r(r\ne i)$是l_i的伪位置（disguised location），则以$1-p$的概率发送。$p_{r,i}$表示l_i随机生成l_r的概率，位置概率矩阵$\boldsymbol{M}$如下：

$$\boldsymbol{M}=\begin{bmatrix} p_{1,1} & \cdots & p_{1,k} \\ \vdots & & \vdots \\ p_{k,1} & \cdots & p_{k,k} \end{bmatrix} \tag{1}$$

x_i表示报告真位置l_i的参与者的真实数量，y_i表示报告伪位置l_r的参与者数量。$\tilde{x}_i$表示报告真位置l_i的参与者估计量，$\tilde{x}_i$可通过式（2）计算得到：

$$\tilde{\boldsymbol{X}}=\boldsymbol{M}^{-1}\boldsymbol{Y} \tag{2}$$

式中，$\tilde{\boldsymbol{X}}=(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_i,\cdots,\tilde{x}_k)^{\mathrm{T}}$，$\boldsymbol{Y}=(y_1,y_2,\cdots,y_i,\cdots,y_k)^{\mathrm{T}}$，$\boldsymbol{M}^{-1}$是$\boldsymbol{M}$的逆矩阵。值得注意的是，LDP所固有的特点是可抵御除自身之外的合谋攻击。

2　隐私保护充电位置汇聚方法

2.1　局部混淆算法

为了保护充电位置，EV一旦接入充电桩进行充电，则启动局部混淆算法（local obfuscation algorithm, LOA）。

算法1　局部混淆算法（LOA）

输入：v_i真实的充电位置l_i，位置域τ，参数s和隐私参数ε_i；

输出：v_i的混淆位置集R_i。

1. 初始化空集$R_i(R_i\subset\tau,|R_i|=s)$；

2. 随机混淆l_i：

$$R_i=\begin{cases}\{l_i\}\cup RandElement(\tau\backslash\{l_i\},s-1), & p;\\ RandElement(\tau\backslash\{l_i\},s), & (1-p)。\end{cases}$$

其中，$p=\dfrac{e^{\varepsilon_i}s}{|\tau|-s+e^{\varepsilon_i}s}$。

3. return R_i

算法1的具体过程如下：首先，v_i初始化一空集R_i。接着，v_i以偏概率p扔一硬币。如果硬币是正面（概率为p），则添加v_i的真位置l_i项和$s-1$个从集合$\tau\backslash\{l_i\}$均匀抽取的伪位置项至集合R_i；如果硬币是反面（概率为$1-p$），则添加s个从$\tau\backslash\{l_i\}$均匀抽取的伪位置项至集合R_i。最后，v_i将集合R_i发送给汇聚器。其中，函数$RandElement(S,s)$返回从集合S中均匀随机抽取的s个元素。

2.2　隐私位置汇聚方法

隐私位置汇聚方法（private location aggregation method, PLAM）的目的是重构访问充电位置的EV数量分布。具体来说，数据汇聚器收集到隐私位置域τ内所有EV的充电位置信息后，执行汇聚算法。令N表示全部参与局部混淆充电位置的EV数量。假设同一隐私位置域内EV用户隐私预算ε相同。在每个报告集合中，含真位置的概率为p，含伪位置的概率为q。汇聚器计算含每个位置（$k=1,2,\cdots,|\tau|$）的报告集R_i的数量w_k，计算w_k如式（3）所示：

$$w_k=\sum_{i\in V}I(R_i,k),I(R_i,k)=\begin{cases}1,k\in R_i;\\ 0,k\notin R_i\end{cases} \tag{3}$$

算法2　隐私位置汇聚方法（PLAM）

输入：汽车充电位置$\{l_i\in\tau\subseteq L:1\le i\le n\}$，隐私位置域大小$|\tau|$，隐私预算$\varepsilon$；

输出：充电位置的EV数量分布$\tilde{x}_k(k=1,2,\cdots,|\tau|)$。

1. 汇聚器计算

$$p=e^{\varepsilon}s/(|\tau|-s+e^{\varepsilon}s),s=\max((|\tau|/1+e^{\varepsilon}),1);$$

2. 汇聚器计算$q=(s-p)/(|\tau|-1)$;
3. for 电动汽车$v_i \in V$ do
4. 汇聚器发送参数s给v_i;
5. v_i发送$R_i=LOA(l_i,\tau,s,\varepsilon)$至汇聚器;
6. for 位置$l_k \in \tau$ do
7. 汇聚器计算w_k;
8. 汇聚器初始化$\tilde{x}_k=w_k$;
9. while(1) //迭代循环
10. for $k=1,2,\cdots,|\tau|$ do
11. 汇聚器计算$L_k=w_k/(p\tilde{x}_k+q(sN-\tilde{x}_k))$;
12. end for
13. 汇聚器计算$Z=\sum_k L_k$;
14. for $k=1,2,\cdots,|\tau|$ do
15. 汇聚器计算
$\tilde{x}_k[t+1]=\tilde{x}_k[t](pL_k+q(Z-L_k))$;
16. 汇聚器计算
$SumErr+=|\tilde{x}_k[t+1]-\tilde{x}_k[t]|$;
17. end for
18. if ($SumErr<ThreshHold$) break
19. end while
20. for $k=1,2,\cdots,|\tau|$ do
21. 汇聚器计算$\tilde{x}_k=\tilde{x}_k/s$;
22. end for
23. return $\tilde{x}_k(k=1,2,\cdots,|\tau|)$

算法2的过程如下：首先，计算隐私位置域内的真位置的概率p、伪位置的概率q以及报告集合长度s这3个参数[15]。隐私位置域τ内的全部EV的充电位置报告集合被汇聚器收集之后，根据式（3）计算每隐私位置域内的包含l_k位置的扰动位置报告的数量w_k，而后利用贝叶斯技术计算每个位置l_k的充电EV数量的估计值。步骤15中，$\tilde{x}_k[t]$表示迭代第t次的$\tilde{x}_k$的值。终止条件在步骤18，取决于前后两次迭代的误差和$SumErr$。如果误差和大于阈值，则继续迭代；否则，终止迭代。因此每次的迭代次数是不确定的。通常来说，迭代阈值与隐私域大小有关，隐私域越大，阈值则设定较大；反之，设定较小，两者是成正比的。

2.3 基于分区的隐私位置汇聚优化方法

隐私位置汇聚方法的汇聚结果可用性取决于隐私位置域的大小。小域随机化方法[17]和个性化局部差分隐私[14]的思想都是在更小的局部随机域（而不是整个范围域）中执行随机响应算法。直观地说，在相同隐私预算和隐私算法条件下，相比于全局随机域，局部随机域内执行随机算法的统计结果的可用性更高。结合EV充电系统的真实场景，作者考虑以一供电台区作为一个隐私位置域对整个区域进行分区，可进一步提高数据可用性。倘若隐私位置域是跨分区的，这将可能导致可用性变差。例如，假设分区A和B合并作为隐私位置域，如果一辆EV访问分区A的充电桩，执行局部混淆算法将可能被扰动到分区B区。此外，隐私位置域越大，则可用性越差，进而影响两个区域的电网优化。因此，利用供电台区对整个位置域进行划分，将单个台区内的全部充电位置作为隐私位置汇聚方法的隐私位置域，以提升汇聚结果的可用性。

算法3 隐私位置汇聚优化方法

输入：全部参与充电EV，整个位置域D；

输出：访问全部充电位置的EV数量。

1. 汇聚器将D划分为多个分区D_i;
2. for 分区$D_i \in D$ do
3. 汇聚器在分区D_i上执行PLAM;
4. end for
5. 汇聚器统计访问D中每个位置的EV数量$\{\tilde{x}_l : l \in L\}$;
6. return $\{\tilde{x}_l : l \in L\}$

算法3过程如下：首先，将充电位置依据供电台区将整个区域D的充电桩位置进行划分，每个分区内的全部充电EV都分配相同的隐私预算ε_i。然后，对每个分区D_i应用PLAM方法，以获得分区内的所有充电位置的EV数量近似精确估计。由于每分区彼此独立，PLAM的汇聚结果互不影响。最后，获得全部充电位置域内的EV数量的近似分布。

2.4 隐私性分析

隐私位置汇聚方法的隐私性关键在于局部混淆算法（LOA）。如果本地化执行的LOA满足可证明的局部差分隐私，则该方法也满足差分隐私。LOA是基于随机响应的差分隐私方法，具有随机响应的似是而非特性。对LOA的隐私性进行分析，即证明LOA算法是否满足ε_i–LDP。

证明：从局部差分隐私定义的角度来说。对于任意的两个充电位置l和l'，如果参数s和p满足，则需证明：

$$\frac{\Pr[A(\tau_i,l_i,\varepsilon_i)=R_i]}{\Pr[A(\tau_i,l_i',\varepsilon_i)=R_i]} \le e^{\varepsilon_i} \tag{4}$$

在LOA中，输出集合含真位置的概率为P_t和不含真位置的概率为P_f，分别是：

$$P_t = p\times\frac{1}{C_{|\tau|-1}^{s-1}} = p\times\frac{(s-1)!(|\tau|-s)!}{(|\tau|-1)!} \tag{5}$$

$$P_f = (1-p)\times\frac{1}{C_{|\tau|-1}^{s}} = (1-p)\times\frac{s!(|\tau|-s-1)!}{(|\tau|-1)!} \tag{6}$$

式（4）中，左边可能结果为$\frac{P_t}{P_f}$、$\frac{P_f}{P_t}$和1。由于

$$\frac{P_t}{P_f}=p\cdot\frac{\frac{(s-1)!(|\tau|-s)!}{(|\tau|-1)!}}{1-p}\cdot\frac{s!(|\tau|-s-1)!}{(|\tau|-1)!}=\frac{p}{1-p}\cdot\frac{|\tau|-s}{s} \tag{7}$$

文献[15]中给出了满足ε_i–LDP差分隐私条件下的最优参数为：

$$s=\max\left(\frac{|\tau|}{1+e^{\varepsilon}},1\right),\quad p=\frac{e^{\varepsilon}s}{|\tau|-s+e^{\varepsilon}s} \tag{8}$$

因此，由式（7）、（8），可得 $\frac{P_t}{P_f}=\frac{p}{1-p}\cdot\frac{|\tau|-s}{s}=e^{\varepsilon_i}$，同理可得$\frac{P_f}{P_t}=\frac{1}{e^{\varepsilon_i}}$，因此，式（4）左侧可能的取值为$e^{\varepsilon_i}$、$\frac{1}{e^{\varepsilon_i}}$和1。

由于$\varepsilon_i>0$，所以，$\max\left(\frac{P_t}{P_f},\frac{P_f}{P_t},1\right)\le e^{\varepsilon_i}$。

证毕。

3　实验评估

实验硬件环境：Intel Core CPU i3–4160，内存8 GB；实验软件环境：Window10+Anaconda3。实验中的数据包含合成数据集和真实数据集。以最近的基于随机映射矩阵的PCEM方法[14]作为对比方法，由于分区后充电位置隐私域比较小以及降维会带来更大噪声，因此采用$k\times k$随机映射矩阵[13]。可用性的两个评价指标分别为*MSE*和*JSD*（JS-divergence），定义分别为式（9）和（10）：

$$MSE=\frac{1}{|\tau|}\sum_{i=1}^{|\tau|}\left(\frac{x_i}{N}-\frac{\tilde{x}_i}{N}\right)^2 \tag{9}$$

$$JSD=\frac{KL(P_X\|R)+KL\left(\tilde{P}_X\|R\right)}{2} \tag{10}$$

其中，

$$KL(P\|R)=\sum_i P(i)\ln\frac{P(i)}{Q(i)},R=\frac{P_X+\tilde{P}_X}{2}。$$

在不同分布的合成数据上对比两种方法，同时，两种方法在真实数据上也是基于相同分区进行对比的。

3.1　合成数据评估

合成数据是基于4种不同分布随机生成的数据，分别是正态分布、均匀分布、峰值分布和随机分布。隐私预算设置在0.1～1.0范围之间。样本量N=1 000，位置域K=10。PCEM和PLAM方法分别在上述4种类型数据上进行实验，运行10次取平均值。

合成数据上的方法评估，如图2所示。结果表明：PLAM方法的*MSE*和*JSD*在4种合成数据上整体优于PCEM方法。然而，部分合成数据上PLAM的指标并不随着隐私预算的增加而降低。这是因为含真位置的随机输出集合的概率p不仅与隐私预算ε有关，还与该集合的长度s有关。并且ε的变化，也会影响s，从式（8）中可得到解释。所以ε与p不成正比。同时，在某些合成数据上，PCEM的误差出现波动，甚至略高于PLAM。这是因为在不同合成数据上PLAM生成的不同随机映射矩阵引入噪声量的差异所致。此外，从图2（c）和（g）的峰值分布可以看到，在隐私预算小（$\varepsilon<0.35$）的情况下，隐私预算引入的噪声与随机映射矩阵引入的噪声相互抵销，导致PCEM的可用性优于PLAM。

但从整体上来说，图2表明PLAM的可用性比PCEM的高，这是因为随机映射矩阵引入的噪声过大，导致扰动后汇聚结果的可用性差。相比PCEM，PLAM能够实现更好的可用性。因此，作者选择PLAM作为EV充电位置隐私汇聚方法是合理的。

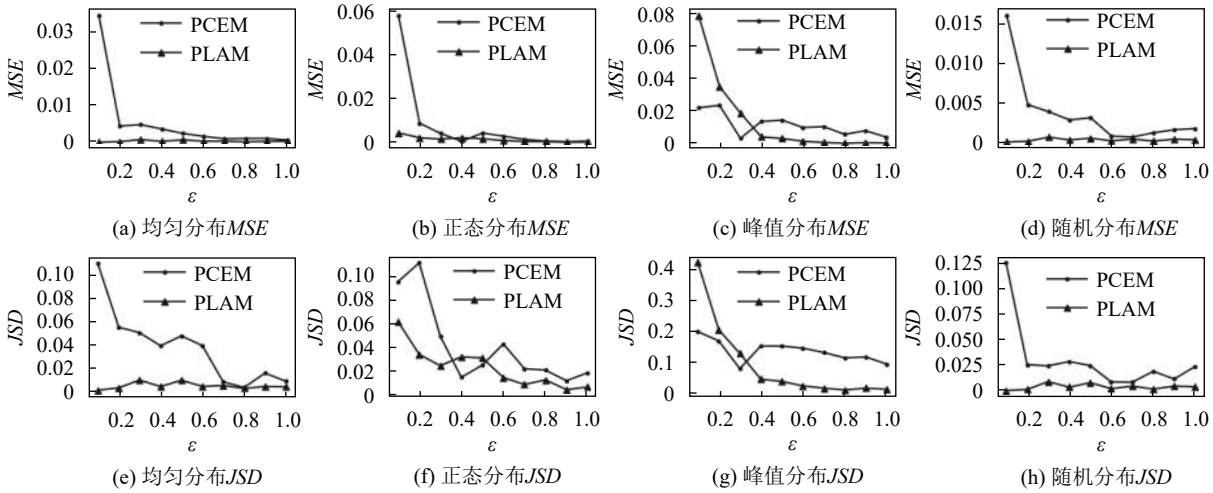
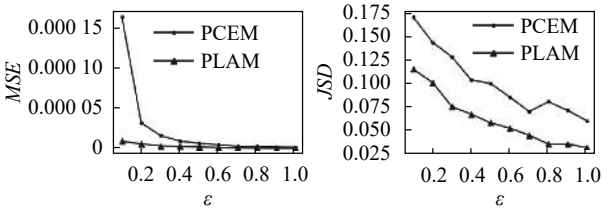


图 2　不同合成数据上的汇聚结果对比

Fig. 2　Comparison of aggregation results of synthetic data with different distributions

3.2 公开数据评估

签到数据集与EV访问充电桩的数据极为相似。因此，采用Gowalla网站提供的用户签到数据集作为公开数据集，选择拉斯维加斯城市主城区（北纬35.95°～36.35°，西经度115.00°～115.35°）作为实验数据集，其中，包含17 644条签到记录。然后，以0.01°的经纬度间隔对该区域进行地理格网划分。假设将每个区块看作一个供电台区，其内的全部签到位置可看作是充电桩的充电位置。由于该数据集内包含大量的稀疏签到数据（单个位置被访问不超过两次），而实际充电桩的位置是频繁被访问的。因此，需对数据集进行预处理，选择签到用户数量高于4的签到位置和位置数量超过2的隐私域中的数据作为实验数据。在公开数据集上，评价*MSE*和*JSD*两个可用性指标，如图3所示，结果表明 PLAM方法在可用性方面优于PCEM方法。



(a) *MSE* 指标的对比　　(b) *JSD* 指标的对比

图 3　公开数据下的汇聚结果对比

Fig. 3　Comparison of aggregation results in public dataset

为比较隐私预算对可用性影响，设置3组不同的隐私预算为$E_{\text{low}}=\{0.25,0.50,0.75\}$、$E_{\text{mid}}=\{0.5,0.75,1.0\}$和$E_{\text{high}}=\{0.75,1.0,1.25\}$的实验。同时，设定同一隐私位置域内的全部用户（EV）的隐私预算都相同。每组实验时，每个隐私位置域的所有用户都随机从对应隐私级别中选择的隐私预算执行PLAM。其中，无预算（Real）表示真实计数。EV数量排名Top 8的位置ID在不同隐私预算下的汇聚结果，如图4所示。横坐标表示数量排名前8的位置ID，纵坐标表示EV数量。

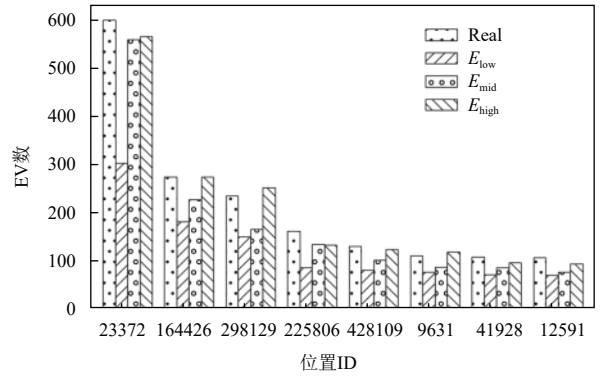


图 4　不同隐私预算下的汇聚结果

Fig. 4　Aggregation results in different privacy budget

图4的结果表明：在高隐私预算下，汇聚结果最接近真实计数；在低隐私预算下，由于引入的噪声比较大，造成汇聚结果的可用性低。

4 结束语

将局部差分隐私技术用于对EV的充电位置数据的保护，以抵御不可信的数据汇聚器泄露EV用户的充电位置隐私。通过引入基于贝叶斯的随机多伪隐私算法，设计了一种基于分区的隐私保护充电位置汇聚方法；并通过划分位置域，缩小隐私位置域大小，提高汇聚结果的可用性。对所设计方法的隐私性给出了较为详细的证明过程。实验评估结果表明：所设计的方法相比于基于随机映射矩阵的隐私汇聚方法，汇聚结果的可用性更高。

本研究针对记录型充电位置数据，更为实际的是研究EV充电轨迹的隐私保护问题，后续工作将引入局部差分隐私解决这一问题。

参考文献：

[1] Han Wenlin,Xiao Yang.Privacy preservation for V2G networks in smart grid:A survey[J].Computer Communications,2016,91:17–28.

[2] Green R C,Wang Lingfeng,Alam M.The impact of plug-in hybrid electric vehicles on distribution networks:A review and outlook[J].Renewable and Sustainable Energy Reviews,2011,15(1):544–553.

[3] Stegelmann M,Kesdogan D.Location privacy for vehicle-to-grid interaction through battery management[C]//Proceedings of the IEEE 9th International Conference on Information Technology:New Generations.Las Vegas:IEEE,2012:373–378.

[4] Liu J K,Susilo W,Yuen T H,et al.Efficient privacy-preserving charging station reservation system for electric vehicles[J].The Computer Journal,2016,59(7):1040–1053.

[5] Yang Zhenyu,Yu Shucheng,Lou Wenjing,et al.P2:Privacy-preserving communication and precise reward architecture for V2G networks in smart grid[J].IEEE Transactions on Smart Grid,2011,2(4):697–706.

[6] Jiang Rong,Lu Rongxing,Lai Chengzhe,et al.A Secure communication protocol with privacy-preserving monitoring and controllable linkability for V2G[C]//Proceedings of the 1st International Conference on Data Science in Cyberspace.Changsha:IEEE,2017:567–572.

[7] Han Shuo,Topcu U,Pappas G J.Differentially private distributed protocol for electric vehicle charging[C]//Proceed-

ings of the 52nd Annual Allerton Conference on Communication,Control,and Computing.Monticello:IEEE,2014:242–249.

[8] Han Shuo,Topcu U,Pappas G J.An approximately truthful mechanism for electric vehicle charging via joint differential privacy[C]//Proceedings of the 2015 American Control Conference.Chicago:IEEE,2015:2469–2475.

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

[10] Kasiviswanathan S P,Lee H K,Nissim K,et al.What can we learn privately?[J].SIAM Journal on Computing,2011,40(3):793–826.

[11] Erlingsson Ú,Pihur V,Korolova A.Rappor:Randomized aggregatable privacy-preserving ordinal response[C]//Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.Scottsdale:ACM,2014:1054–1067.

[12] Bassily R,Smith A.Local,private,efficient protocols for succinct histograms[C]//Proceedings of the 47th annual ACM Symposium on Theory of Computing.Portland:ACM,2015:127–135.

[13] Nguyên T T,Xiao Xiaokui,Yang Yin,et al.Collecting and analyzing data from smart device users with local differential privacy[EB/OL].(2016–06–16)[2017–12–07].https://arxiv.org/abs/1606.05053.

[14] Chen Rui,Li Haoran,Qin A K,et al.Private spatial data aggregation in the local setting[C]//Proceedings of the IEEE 32nd International Conference on Data Engineering.Helsinki:IEEE,2016:289–300.

[15] Sei Y,Ohsuga A.Differential private data collection and analysis based on randomized multiple dummies for untrusted mobile crowdsensing[J].IEEE Transactions on Information Forensics and Security,2017,12(4):926–939.

[16] Warner S L.Randomized response:A survey technique for eliminating evasive answer bias[J].Journal of the American Statistical Association,1965,60(309):63–69.

[17] Chaytor R,Wang Ke.Small domain randomization:Same privacy,more utility[J].Proceedings of the VLDB Endowment,2010,3(1/2):608–618.

（编辑　赵　婧）

引用格式：Xiong Xingxing,Liu Shubo,Li Dan,et al.Private electric vehicle charging location aggregation based on local differential privacy[J].Advanced Engineering Sciences,2019,51(2):137–143.[熊星星,刘树波,李丹,等.基于局部差分隐私的电动汽车充电位置隐私汇聚[J].工程科学与技术,2019,51(2):137–143.]