

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 9 月 10 日 (10.09.2020)



(10) 国际公布号
WO 2020/177484 A1

(51) 国际专利分类号:
H04W 4/029 (2018.01)

(21) 国际申请号: PCT/CN2020/070657

(22) 国际申请日: 2020 年 1 月 7 日 (07.01.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910156339.3 2019年3月1日 (01.03.2019) CN

(71) 申请人: 华南理工大学 (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) [CN/CN]; 中国广东省广州市天河区五山路 381 号, Guangdong 510640 (CN)。

(72) 发明人: 唐韶华 (TANG, Shaohua); 中国广东省广州市天河区五山路 381 号, Guangdong 510640 (CN)。彭帆 (PENG, Fan); 中国广东省广州市天河区五山路 381 号, Guangdong 510640 (CN)。

(74) 代理人: 广州市华学知识产权代理有限公司 (GUANGZHOU HUAXUE INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省广州市天河区五山路 381 号物资大楼首层, Guangdong 510640 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,

(54) Title: LOCALIZED DIFFERENCE PRIVACY URBAN SANITATION DATA REPORT AND PRIVACY CALCULATION METHOD

(54) 发明名称: 一种本地化差分隐私城市环卫数据报告和隐私计算方法

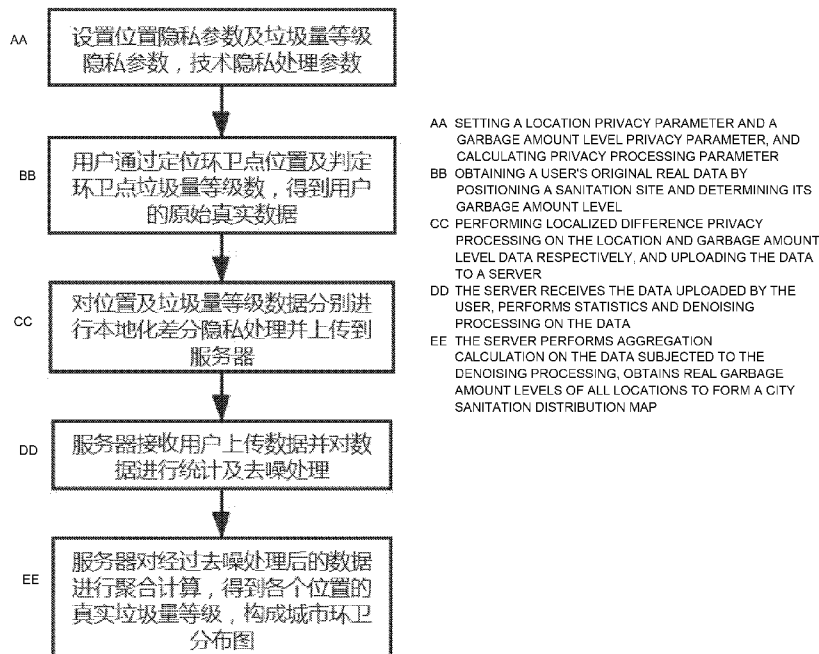


图 1

(57) Abstract: The present invention discloses a localized difference privacy urban sanitation data report and a privacy calculation method, the specific steps comprise: setting a location privacy parameter and a garbage amount level privacy parameter, and calculating a privacy processing parameter; obtaining a user's original real data by positioning a sanitation site and determining its garbage amount level; performing localized difference privacy processing on the location and garbage amount level data respectively, and uploading the data to a server; the server receives the data uploaded by the user, performs statistics and denoising processing on the data; the server

LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

— 发明人资格 (细则4.17(iv))

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

performs aggregation calculation on the data subjected to the denoising processing, obtains real garbage amount levels of all locations to form a city sanitation distribution map. The present invention can, under the premise of protecting the location and data privacy of all data collectors, achieve the goal of urban sanitation dynamic distribution on the basis of mobile group perception.

(57) 摘要: 本发明公开了一种本地化差分隐私城市环卫数据报告和隐私计算方法, 具体步骤包括: 设置位置隐私参数及垃圾量等级隐私参数, 计算隐私处理参数; 用户通过定位环卫点位置及判定环卫点垃圾量等级数, 得到用户的原始真实数据; 对位置及垃圾量等级数据分别进行本地化差分隐私处理并上传到服务器; 服务器接收用户上传数据并对数据进行统计及去噪处理; 服务器对经过去噪处理后的数据进行聚合计算, 得到各个位置的真实垃圾量等级, 构成城市环卫分布图。本发明在达到在保护所有数据收集者位置和数据隐私的前提下, 能够实现基于移动群体感知的城市环卫动态分布的目标。

一种本地化差分隐私城市环卫数据报告和隐私计算方法

技术领域

本发明涉及信息安全隐私保护领域，尤其涉及一种本地化差分隐私城市环卫数据报告和隐私计算方法。

背景技术

随着传感器精度的不断提升以及移动智能设备的普及，以智能手机为代表的移动设备能随时随地地感知人和周边的环境信息，使人们可以及时地获取其所处环境的状况，更好地为人们及社会提供所需的服务，促进移动群体感知领域不断向前发展。移动群体感知为人们提供了一种新的感知世界的方式，能够使任何人都能通过移动终端执行感知任务参与到感知过程，并为人们提供各种服务。

移动群体的应用中，感知参与用户的隐私能否得到保证是关系用户是否愿意参与感知任务的关键因素。传统加密方式存在计算开销大、基于匿名的各种隐私保护方法不能抵抗背景知识攻击等不足之处，中心化差分隐私方法需要基于第三方数据收集者诚实可信，但是不会窃取或泄露用户的敏感信息的假设在实际应用中并不现实。而本地化差分隐私技术在继承中心化差分隐私技术定量化定义隐私攻击的基础上，将隐私保护的处理过程让每个用户独立完成，能够实现更彻底的隐私保护。

随机响应技术是本地化差分隐私保护技术的主流扰动机制，随机响应需要的样本数据量巨大，在大部分的移动群体感知场景当中，参与用户的数据量非常有限，因此如何将随机响应技术应用在中小样本移动群体感知场景有待进一步的研究。

发明内容

本发明的目的在于克服现有技术的不足，提供一种基于随机响应实现的本地化差分隐私城市环卫数据报告和隐私计算方法。本发明在移动群体感知领域应用随机响应技术进行报告收集，对隐私数据进行聚合计算分析，得到隐私属性的联合分布，实现了对城市环卫分布的监控。

本发明的目的能够通过以下技术方案实现：

一种本地化差分隐私城市环卫数据报告和隐私计算方法，具体步骤包括：

设置位置隐私参数及垃圾量等级隐私参数，计算隐私处理参数；

用户通过定位环卫点位置及判定环卫点垃圾量等级数，得到用户的原始真实数据；

对位置及垃圾量等级数据分别进行本地化差分隐私处理并上传到服务器；

服务器接收用户上传数据并对数据进行统计及去噪处理；

服务器对经过去噪处理后的数据进行聚合计算，得到各个位置的真实垃圾量等级，构成城市环卫分布图。

具体地，所述设置位置隐私参数及垃圾量等级隐私参数，计算隐私处理参数的步骤中，设定位置隐私保护参数 ϵ_1 和垃圾量等级隐私参数 ϵ_2 ，并将参数公布给每个用户。

通过布隆过滤器假阳率 k ，感知值集合 M 计算布隆过滤器长度 l 和哈希函数个数 h ，计算公式为：

$$l = \frac{\ln(1/k)}{(\ln 2)^2} |M|$$

$$h = \frac{l}{|M|} \ln 2 = \frac{\ln(1/k)}{\ln 2}$$

根据隐私参数 ϵ_1 、 ϵ_2 、环卫点总数 F 及哈希函数个数 h ，计算隐私处理参数：

r 表示位置报告数，其计算公式为：

$$r = \max\left(\left\lfloor \frac{F}{(1 + e^{\epsilon_1})} \right\rfloor, 1\right),$$

p 表示真实位置选中概率，其计算公式为：

$$p = \frac{e^{\epsilon_1} \cdot r}{F - r + e^{\epsilon_1} \cdot r},$$

f 表示位翻转概率，其计算公式为：

$$f = \frac{2}{e^{\frac{\epsilon_2}{2h}} + 1}.$$

具体地，所述用户通过定位环卫点位置及判定环卫点垃圾量等级数，得到用户的原始真实数据的步骤中，用户通过携带的移动智能设备定位环卫点位置，并通过位置映射表得到该环卫点的编号 i 。所述位置映射表由服务器提前为每个环卫点的位置编号得到。用户通过自行判定，得到环卫点垃圾量的等级数 $level$ ，编号和垃圾等级数即为用户的原始真实数据 $\langle i, level \rangle$ 。

具体地，所述对位置及垃圾量等级数据分别进行本地化差分隐私处理的步骤中，用户在本地进行本地化差分隐私处理，处理步骤为：

初始化位置集合 $L = \emptyset$ 和上传结果集合 $R = \emptyset$ ；

生成一个 $(0,1)$ 之间的随机数 $rand$ ，结合隐私处理参数，对位置数据进行本地化差分隐私处理，使其满足位置隐私保护参数 ϵ_1 ；

根据本地化差分隐私处理后得到的位置集合 L ，对垃圾量等级数据进行本地化差分隐私处理，计算用户的上传结果集 R 。

更进一步地，所述生成一个 $(0,1)$ 之间的随机数 $rand$ ，结合隐私处理参

数，对位置数据进行本地化差分隐私处理的步骤中，具体处理方法为：

对随机数rand及真实位置选中概率p的大小关系进行判断：

如果随机数rand小于p，则位置集合为真实位置i与除去i之外的位置集合中任意选择 $r - 1$ 个位置的并集，表示为：

$$L = \{i\} \cup \text{getRandomElements}(F \setminus \{i\}, r - 1)$$

其中，函数 $\text{getRandomElements}(A, b)$ 表示返回从集合A中随机选择b个元素的结果；

如果随机数rand大于p，则位置集合为除去i之外的位置集合中任意选择r个位置的并集，表示为：

$$L = \text{getRandomElements}(F, r)$$

综上，随机数rand与真实位置选中概率p的大小判断关系的表示方式为：

$$L = \begin{cases} \{i\} \cup \text{getRandomElements}(F \setminus \{i\}, r - 1), & \text{rand} < p \\ \text{getRandomElements}(F, r), & \text{rand} \geq p \end{cases}$$

此处理方法能够保证用户的位置数据满足 ϵ_1 -差分隐私。

更进一步地，所述根据本地化差分隐私处理后得到的位置集合L，对垃圾量等级数据进行本地化差分隐私处理，计算用户的上传结果集R的步骤中，如果用户的真实位置i包含在L中，真实位置垃圾量的对应项为真实垃圾量等级 $level_i$ 。此时对于位置集合L中的其他位置元素t，其垃圾量等级对应项则在垃圾量等级值集合中等概率随机选择一个等级 $level'_t$ ，随后对每一个垃圾量等级进行处理，具体处理方法为：

采用h个哈希函数将垃圾量等级 $level_i$ 或 $level'_t$ 映射到长度为l的布隆过滤器，得到一个长度为l的向量 S_t ；

对于 S_t 中的每一位 $S_t[j]$ ，运行一个随机数生成函数得到一个(0,1)之间的

随机数 q ，如果 $q < 1 - f/2$ ，第 j 位保持不变，如果 $q > 1 - f/2$ ，对第 j 位进行翻转，即原始值为0时，翻转为1；原始值为1时，翻转为0；得到 S_t' ；

综上，对垃圾量等级数据进行处理的方式为：

$$S_t[j]' = \begin{cases} S_t[j], & q < 1 - f/2 \\ \neg S_t[j], & q \geq 1 - f/2 \end{cases}$$

此处理方式保证用户的垃圾量等级感知数据满足 ϵ_2 -差分隐私。

将每个位置 i 和对应的处理过后的垃圾等级值 S_i' 组合成对 (i, S_i') ，加入到上传结果集合 R ，得到最后的上传结果集合 R 。

具体地，所述服务器接收用户上传数据并对数据进行统计及去噪处理的步骤中，所述统计及去噪处理方法为：

混合所有用户上传数据 Z ，提取上传报告中所有的位置数据，计算每个位置的真实报告个数 E_i ，计算公式为：

$$q = (r - p)/(F - 1)$$

$$E_i = w_i$$

$$L_i = w_i/(pE_i + q(rN - E_i))$$

$$V = \sum_i L_i$$

$$E_i = E_i(pL_i + q(V - L_i))$$

$$E_i = E_i/r$$

其中， Z 表示用户上传数据， F 表示环卫点总数， p 表示真实位置选择概率， r 表示位置报告数， N 表示用户数量， w_i 表示 Z 中包含位置 i 的个数， $i = 1, \dots, F$ ；

每个上传位置 i ，服务器筛选出所有位置 i 的报告对数据集 D_i 。假设 D_i 总共有 N_i 条数据项，真实报告个数为 E_i ，则噪声报告个数为 $N_i - E_i$ ，对垃圾

量等级数据集 S'_i 的每一位做如下统计：计算每一位为“1”的总数 $X_i[j]$ ， $X_i[j]$ 表示第 i 个位置的第 j 位， $0 \leq j < l$ ，组成一个向量 X_i ， X_i 表示编号为 i 的位置收到的每一位为“1”的计数；

X_i 的来源由四部分组成，如表 1 所示。

表 1

真实数据 E_i	原始值为“1”	$Z_i(1 - f/2)$
	翻转后为“1”	$(E_i - Z_i)f/2$
噪声数据 $N_i - E_i$	原始值为“1”	$U_i(1 - f/2)$
	翻转后为“1”	$(N_i - E_i - U_i)f/2$

在真实数据集中，假设原始垃圾量等级经过布隆过滤器后，总共有 Z_i 个“1”，其有 $(1 - f/2)$ 的概率保持为“1”， $E_i - Z_i$ 个“0”，其有 $f/2$ 的概率翻转为“1”；同理，假设噪声数据的原始垃圾量等级经过布隆过滤器后共有 U_i 个“1”，其有 $(1 - f/2)$ 的概率保持为“1”； $N_i - E_i - U_i$ 个“0”，其有 $f/2$ 的概率翻转为“1”。

因此下列等式成立：

$$Z_i \left(1 - \frac{f}{2}\right) + (E_i - Z_i)f/2 + U_i(1 - f/2) + (N_i - E_i - U_i)f/2 = X_i$$

由于加入的噪声呈均匀分布，上式中的 U_i 能够通过计算得到，因此，上式可推导得：

$$Z_i = \left[X_i - \frac{(N_i - E_i)f}{2} - U_i(1 - f) - E_i \frac{f}{2} \right] / (1 - f)$$

由此求得原始真实的垃圾量等级在各个位上的“1”的数组成的向量 Z_i 。

由于步骤 1 的频率估计存在一定的误差，数据生成和处理过程也有一

定的随机性， Z_i 计算的值也存在一定的误差，因此需要一定的分析方法来得到位置*i*的真实垃圾量等级。

具体地，所述服务器对经过去噪处理后的数据进行聚合计算，得到各个位置的真实垃圾量等级，构成城市环卫分布图步骤中，将垃圾量等级的所有*w*个候选值通过布隆过滤器的*h*个哈希函数处理后，组成一个*w*·*l*的矩阵作为特征矩阵*Q*，经过去噪处理得到的所有 $Z_i[j](0 \leq j < l)$ 组成长度为*l*的向量 Z_i ，求解线性回归问题 $Z_i = A \cdot Q$ ，向量 $A(1 \cdot w)$ 中最大值所对应位置的垃圾量等级即为位置*i*的真实垃圾量等级。

通过上述步骤能够计算出每个位置的真实垃圾量等级，从而得到城市垃圾量分布图。

本发明相较于现有技术，具有以下的有益效果：

1.本发明相较于基于物联网技术的智能垃圾桶，大大节省了硬件设施的部署成本、网络连接的通信开销和设备运维的人力资源，充分节约了社会资源；

2.本发明不依赖可信的第三方服务器，数据完全由用户在移动端处理，更彻底地保证了用户的数据隐私安全；

3.本发明在保证了聚合准确性的同时，大大减少了用户参与人数的最低限额，是现实可行的方案；

4.本发明在提升速度同时保持了方案的安全性。

附图说明

图 1 是一种本地化差分隐私城市环卫数据报告和隐私计算方法的流程图。

图 2 是一种本地化差分隐私城市环卫数据报告和隐私计算方法的结构

示意图。

图 3 是本发明实施例中垃圾量等级随机扰动的样例图。

具体实施方式

下面结合实施例及附图对本发明作进一步详细的描述，但本发明的实施方式不限于此。

实施例

如图 1、2 所示分别为一种本地化差分隐私城市环卫数据报告和隐私计算方法的流程图及结构图，所述方法的具体步骤包括：

设置位置隐私参数及垃圾量等级隐私参数，计算隐私处理参数；

用户通过定位环卫点位置及判定环卫点垃圾量等级数，得到用户的原始真实数据；

对位置及垃圾量等级数据分别进行本地化差分隐私处理并上传到服务器；

服务器接收用户上传数据并对数据进行统计及去噪处理；

服务器对经过去噪处理后的数据进行聚合计算，得到各个位置的真实垃圾量等级，构成城市环卫分布图。

在本实施例中，假设城市某区域环卫点个数总共为 50 个，垃圾量的等级有 3 个等级，分别为 High、Middle 和 Low，设置位置的隐私参数为 $\epsilon_1 = 1.6$ ，垃圾量等级的隐私参数 $\epsilon_2 = 1.9$ ，布隆过滤器假阳率为 0.1，则计算得到最优布隆过滤器长度 $l = \frac{\ln(\frac{1}{k})}{(\ln 2)^2} |M| = \frac{\ln 10}{(\ln 2)^2} \times 3 = 14$ ，哈希函数个数为

$$h = \frac{l}{|M|} \ln 2 = 3, \text{ 隐私处理参数 } r = \max\left(\left\lfloor \frac{F}{(1+e^{\epsilon_1})} \right\rfloor, 1\right) = 8, \quad p = \frac{e^{\epsilon_1 \cdot r}}{F - r + e^{\epsilon_1 \cdot r}} = 0.486, f = \frac{2}{e^{2h} + 1} = 0.84。$$

假设一个感知用户经过一个环卫点时，要对该环卫点的垃圾量进行报告：首先通过移动设备的 GPS 传感器对当前环卫点位置进行定位，然后通过提前设置的位置-编号表匹配到该环卫点的编号，如该环卫点的编号为 12，用户通过对垃圾量的观察给定相应的垃圾量等级，如判定该环卫点的真实垃圾量等级为 *Middle*，数据{12, *Middle*}即为该感知参与用户的真实感知数据。

根据上述计算的参数，用户在本地运行一个随机数生成函数生成一个随机数 $\text{rand} = 0.67$ ，由于 $0.67 > 0.486$ ，该用户的位置集合为从集合 {1, 2, ..., 11, 13 ..., 50} 中随机选择 8 个位置作为上传位置，由于真实感知位置不在位置集合中，所以所有 8 个位置对应的垃圾量等级都从垃圾量等级集合中随机抽取决定。假设经过 8 次随机化选择后得到的数据集合为：
{< 1, *Middle* >, < 5, *High* >, < 9, *Low* >, < 14, *Middle* >, < 21, *Low* >, < 23, *High* >, < 36, *Low* >, < 41, *Middle* >}。

假设垃圾量等级经过布隆过滤器后得到对应的向量分别为 {High: 1000001001000, Middle: 0100000000011, Low: 1000001000001}。集合中对应的各个垃圾量等级经过布隆过滤器后的每一位通过生成的随机数和参数 $1 - \frac{f}{2} = 0.58$ 比较，小于 0.58 保持不变，大于 0.58 进行翻转，得到最终上传的集合为：

{< 1, 0101010011011 >, < 5, 1011100100110 >, < 9, 0111000101011 >, < 14, 1101000101011 >, < 21, 1010111101010 >, < 23, 0110101101010 >, < 36, 1101010100111 >, < 41, 0011011011010 >}

垃圾量等级随机扰动的样例图如图 3 所示。

当服务器接收到所有感知用户的上传值后，对所有的位置项进行概率

统计，能大致估计每个位置真实的报告个数，如服务器总共收到 2000 个用户发来的数据，其中位置数据共有 $2000 \times 8 = 16000$ 个，估计出真实位置为“1”的用户数为 50 个，实际收到的所有报告中包含位置为“1”的用户数为 300 个，由此推断所有为“1”的数据项中大约有 250 个数据项是噪声位置数据，噪声位置数据项的垃圾量等级是随机均匀选取的，由此可估算出每个垃圾量等级的噪声数据大约有 $250/3$ 个，记为 $Noise = (noise) \cdot 3$ 。

将垃圾量等级对应的原始布隆过滤器的值组合成一个矩阵 Q ，表示为：

$$Q = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

将每一个位置的垃圾量等级数据筛选出来单独分析，如所有位置为“1”的垃圾量等级项，表示形式为：

第0位	第1位	...	第 l 位
1	0	1	0
...	...	...	...
0	1	1	0

以列为单位进行求和，得到一个长度为 l 的向量 $(sum_1, \dots, sum_l)$ ，记为 X ，根据方案中计算得到的噪声估计量，计算每个噪声等级原始为“1”的个数向量 U ， $U = Noise \cdot A$ ，根据统计计算的真实估计值、噪声值和计算公式，计算各个位原始为“1”的个数，为长度为 l 的向量 Z 。

求解线性回归方程问题 $Z = A \cdot Q$ ，向量 A 中最大值对应的垃圾量等级即为该位置的估计真实垃圾量等级。

所有位置的垃圾量等级能够通过上述过程依次求出，从而得到城市该区域的垃圾量等级分布图。

上述实施例为本发明较佳的实施方式，但本发明的实施方式并不受上述实施例的限制，其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化，均应为等效的置换方式，都包含在本发明的保护范围之内。

权 利 要 求 书

1、一种本地化差分隐私城市环卫数据报告和隐私计算方法，其特征在于，具体步骤包括：

设置位置隐私参数及垃圾量等级隐私参数，计算隐私处理参数；

用户通过定位环卫点位置及判定环卫点垃圾量等级数，得到用户的原始真实数据；

对位置及垃圾量等级数据分别进行本地化差分隐私处理并上传到服务器；

服务器接收用户上传数据并对数据进行统计及去噪处理；

服务器对经过去噪处理后的数据进行聚合计算，得到各个位置的真实垃圾量等级，构成城市环卫分布图。

2、根据权利要求 1 所述的一种本地化差分隐私城市环卫数据报告和隐私计算方法，其特征在于，所述设置位置隐私参数及垃圾量等级隐私参数步骤中，设定位置隐私保护参数 ϵ_1 和垃圾量等级隐私参数 ϵ_2 ，将参数公布给每个用户；

通过布隆过滤器假阳率 k ，感知值集合 M 计算布隆过滤器长度 l 和哈希函数个数 h ，计算公式为：

$$l = \frac{\ln(1/k)}{(\ln 2)^2} |M|$$

$$h = \frac{l}{|M|} \ln 2 = \frac{\ln(1/k)}{\ln 2}$$

根据隐私参数 ϵ_1 、 ϵ_2 、环卫点总数 F 及哈希函数个数 h ，计算隐私处理参数，计算公式为：

$$r = \max\left(\left\lfloor \frac{F}{(1 + e^{\epsilon_1})} \right\rfloor, 1\right)$$

$$p = \frac{e^{\epsilon_1 \cdot r}}{F - r + e^{\epsilon_1 \cdot r}}$$

$$f = \frac{2}{\frac{\epsilon_2}{e^{2h}} + 1}$$

其中, r 表示位置报告数, p 表示真实位置选中概率, f 表示位翻转概率。

3、根据权利要求 1 所述的一种本地化差分隐私城市环卫数据报告和隐私计算方法, 其特征在于, 所述用户通过定位环卫点位置及判定环卫点垃圾量等级数, 得到用户的原始真实数据的步骤中, 用户通过携带的移动智能设备定位环卫点位置, 并通过映射表得到该环卫点的编号 i ; 用户通过自行判定, 环卫点垃圾量的等级数 $level$, 编号和垃圾等级数即为用户的原始真实数据 $\langle i, level \rangle$ 。

4、根据权利要求 1 所述的一种本地化差分隐私城市环卫数据报告和隐私计算方法, 其特征在于, 所述对位置及垃圾量等级数据分别进行本地化差分隐私处理的步骤中, 用户在本地进行本地化差分隐私处理, 处理步骤为:

初始化位置集合 $L = \emptyset$ 和上传结果集合 $R = \emptyset$;

生成一个 $(0,1)$ 之间的随机数 $rand$, 结合隐私处理参数, 对位置数据进行本地化差分隐私处理, 使其满足位置隐私保护参数 ϵ_1 ;

根据本地化差分隐私处理后得到的位置集合 L , 对垃圾量等级数据进行本地化差分隐私处理, 计算用户的上传结果集 R 。

5、根据权利要求 4 所述的一种本地化差分隐私城市环卫数据报告和隐私计算方法, 其特征在于, 所述生成一个 $(0,1)$ 之间的随机数 $rand$, 结合隐私处理参数, 对位置数据进行本地化差分隐私处理的步骤中, 具体处理方法为:

对随机数 $rand$ 及真实位置选中概率 p 的大小关系进行判断:

如果随机数 $rand$ 小于 p , 则位置集合为真实位置 i 与除去 i 之外的位置集合中任意选择 $r - 1$ 个位置的并集, 表示为:

$$L = \{i\} \cup getRandomElements(F \setminus \{i\}, r - 1)$$

其中, 函数 $getRandomElements(A, b)$ 表示返回从集合 A 中随机选择 b 个元素的结果;

如果随机数rand大于p，则位置集合为除去i之外的位置集合中任意选择r个位置的并集，表示为：

$$L = getRandomElements(F, r)$$

综上表示为：

$$L = \begin{cases} \{i\} \cup getRandomElements(F \setminus \{i\}, r - 1), & \text{rand} < p \\ getRandomElements(F, r), & \text{rand} \geq p \end{cases}.$$

6、根据权利要求4所述的一种本地化差分隐私城市环卫数据报告和隐私计算方法，其特征在于，所述根据本地化差分隐私处理后得到的位置集合L，对垃圾量等级数据进行本地化差分隐私处理，计算用户的上传结果集R的步骤中，如果用户的真实位置i包含在L中，真实位置垃圾量的对应项为真实垃圾量等级 $level_i$ ；此时对于位置集合L中的其他位置元素t，其垃圾量等级对应项则在垃圾量等级值集合中等概率随机选择一个等级 $level'_t$ ，随后对每一个垃圾量等级进行处理，具体处理方法为：

采用h个哈希函数将垃圾量等级 $level_i$ 或 $level'_t$ 映射到长度为l的布隆过滤器，得到一个长度为l的向量 S_t ；

对于 S_t 中的每一位 $S_t[j]$ ，运行一个随机数生成函数得到一个(0,1)之间的随机数q，如果 $q < 1 - f/2$ ，第j位保持不变，如果 $q > 1 - f/2$ ，对第j位进行翻转，即原始值为0时，翻转为1，原始值为1时，翻转为0，得到 S'_t ；

综上表示为：

$$S_t[j]' = \begin{cases} S_t[j], & q < 1 - f/2 \\ \neg S_t[j], & q \geq 1 - f/2 \end{cases}$$

将每个位置i和对应的处理过后的垃圾等级值 S'_i 组合成对 (i, S'_i) ，加入到上传结果集合R，得到最后的上传结果集合R。

7、根据权利要求1所述的一种本地化差分隐私城市环卫数据报告和隐私计算方法，其特征在于，所述服务器接收用户上传数据并对数据进行统计及去噪处理步骤中，统计及去噪处理方法为：

混合所有用户上传数据Z，提取上传报告中所有的位置数据，估计每

个位置的真实报告个数 E_i ;

对于每个上传位置 i , 服务器筛选出所有位置 i 的报告对数据集 D_i ;

假设 D_i 总共有 N_i 条数据项, 真实报告个数为 E_i , 则噪声报告个数为 $N_i - E_i$, 对垃圾量等级数据集 S'_i 的每一位做如下统计: 计算每一位为“1”的总数 $X_i[j]$, $X_i[j]$ 表示第 i 个位置的第 j 位, $0 \leq j < l$, 组成一个向量 X_i , X_i 表示编号为 i 的位置收到的每一位为“1”的计数;

在真实数据集中, 假设原始垃圾量等级经过布隆过滤器后, 总共有 Z_i 个“1”, 其有 $(1 - f/2)$ 的概率保持为“1”, $E_i - Z_i$ 个“0”, 其有 $f/2$ 的概率翻转为“1”; 同理, 假设噪声数据的原始垃圾量等级经过布隆过滤器后共有 U_i 个“1”, 其有 $(1 - f/2)$ 的概率保持为“1”; $N_i - E_i - U_i$ 个“0”, 共有 $f/2$ 的概率翻转为“1”;

因此下列等式成立:

$$Z_i \left(1 - \frac{f}{2}\right) + (E_i - Z_i)f/2 + U_i(1 - f/2) + (N_i - E_i - U_i)f/2 = X_i$$

由于加入的噪声呈均匀分布, 上式中的 U_i 能够通过计算得到, 因此, 上式推导得:

$$Z_i = \left[X_i - \frac{(N_i - E_i)f}{2} - U_i(1 - f) - E_i \frac{f}{2} \right] / (1 - f)$$

由此求得原始真实的垃圾量等级在各个位上的“1”的个数组成的向量 Z_i 。

8、根据权利要求1所述的一种本地化差分隐私城市环卫数据报告和隐私计算方法, 其特征在于, 所述服务器对经过去噪处理后的数据进行聚合计算, 得到各个位置的真实垃圾量等级, 构成城市环卫分布图步骤中, 将垃圾量等级的所有 w 个候选值通过布隆过滤器的 h 个哈希函数处理后, 组成一个 $w \cdot l$ 的矩阵作为特征矩阵 Q , 经过去噪处理得到的所有 $Z_i[j]$, 组成长度为 l 的向量 Z_i , 求解线性回归问题 $Z_i = A \cdot Q$, 向量 $A(1 \cdot w)$ 中最大值所对应位置的垃圾量等级即为位置 i 的真实垃圾量等级。

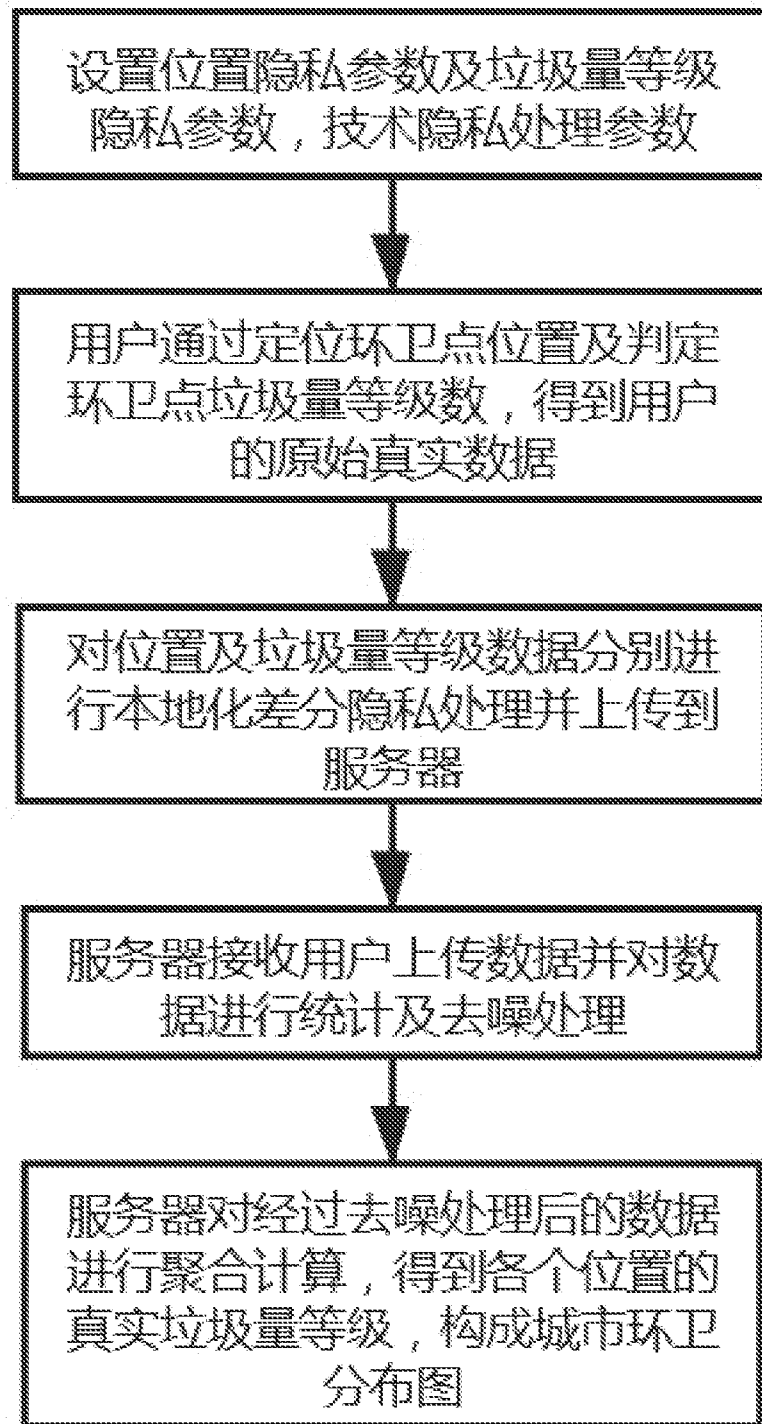


图 1

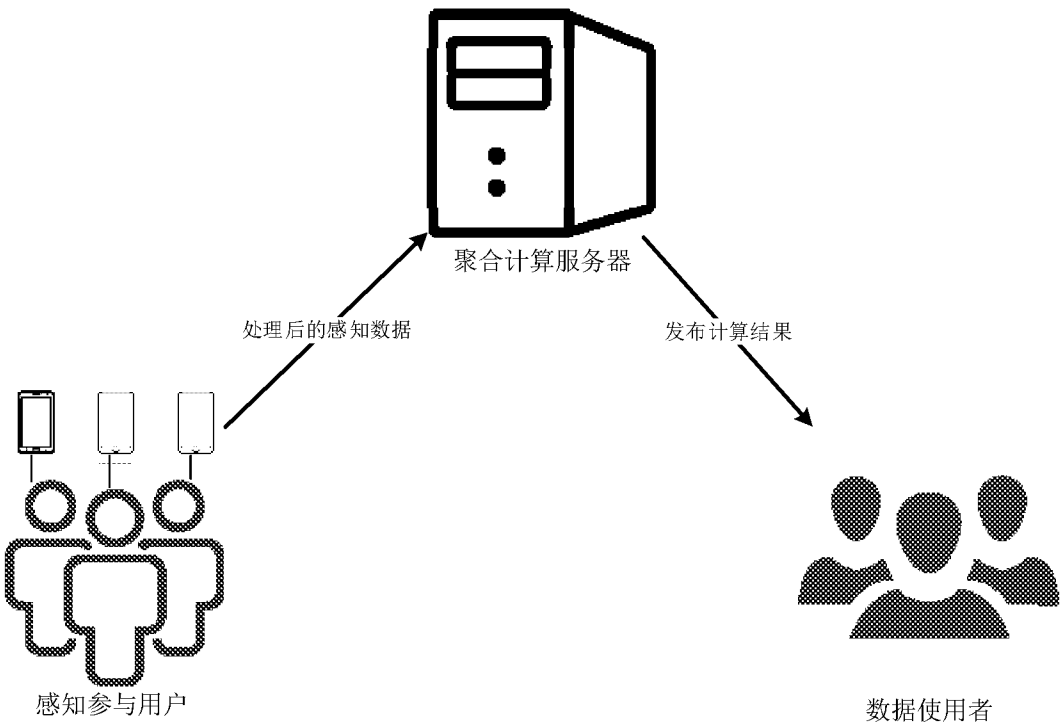


图 2

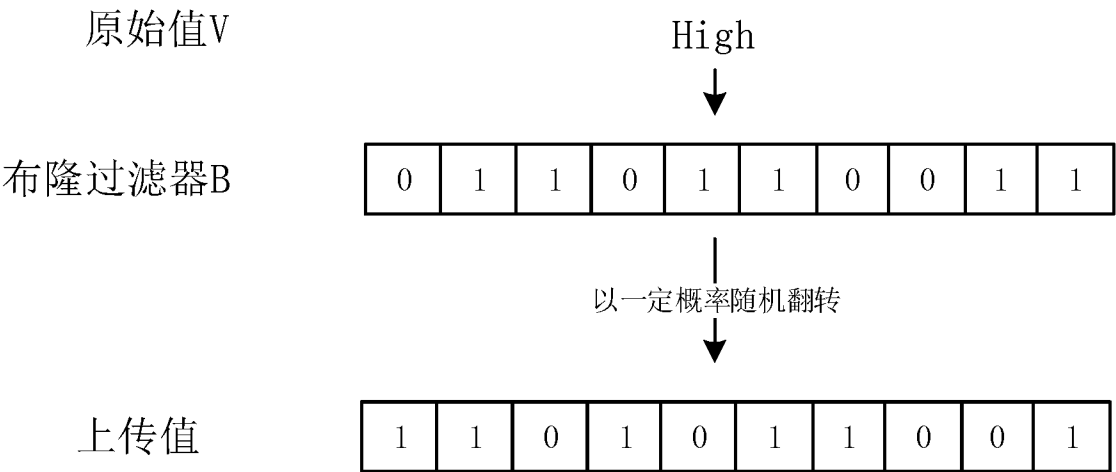


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070657

A. CLASSIFICATION OF SUBJECT MATTER

H04W 4/029(2018.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04W G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC, IEEE: 差分, 隐私, 保护, 本地, 分布, 局部, 局地, 随机, 扰动, 噪音, privacy, differen+, local, distribute, radom+, response, noise, scrambling

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110022531 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 16 July 2019 (2019-07-16) claims 1-8	1-8
X	高志强等 (GAO, Zhiqiang et al.). "满足本地差分隐私的位置数据采集方案 (Collection Scheme of Location Data based on Local Differential Privacy)" 清华大学学报 (自然科学版) (Journal of Tsinghua University (Science and Technology)), 31 January 2019 (2019-01-31), pages 23-27	1
A	CN 109299436 A (BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 01 February 2019 (2019-02-01) entire document	1-8
A	CN 107871087 A (GUANGXI NORMAL UNIVERSITY) 03 April 2018 (2018-04-03) entire document	1-8
A	CN 108256000 A (WUHAN UNIVERSITY) 06 July 2018 (2018-07-06) entire document	1-8
A	US 2018173894 A1 (SAP SE) 21 June 2018 (2018-06-21) entire document	1-8



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

21 March 2020

Date of mailing of the international search report

10 April 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070657

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	高志强等 (GAO, Zhiqiang et al.). "本地差分隐私保护及其应用 (Local Differential Privacy Protection and Its Applications)" <i>计算机工程与科学 (Computer Engineering & Science)</i> , 30 June 2018 (2018-06-30), pages 1029-1036	1-8

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/070657

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110022531	A	16 July 2019	None			
CN	109299436	A	01 February 2019	None			
CN	107871087	A	03 April 2018	None			
CN	108256000	A	06 July 2018	None			
US	2018173894	A1	21 June 2018	CN	108229181	A	29 June 2018
				EP	3340101	A1	27 June 2018

国际检索报告

国际申请号

PCT/CN2020/070657

A. 主题的分类 H04W 4/029 (2018.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																										
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) H04W G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, CNKI, WPI, EPDOC, IEEE: 差分, 隐私, 保护, 本地, 分布, 局部, 局地, 随机, 扰动, 噪音, privacy, differen+, local, distribute, radom+, response, noise, scrambling																										
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110022531 A (华南理工大学) 2019年 7月 16日 (2019 - 07 - 16) 权利要求1-8</td> <td>1-8</td> </tr> <tr> <td>X</td> <td>高志强等. “满足本地差分隐私的位置数据采集方案” 清华大学学报 (自然科学版), 2019年 1月 31日 (2019 - 01 - 31), 第23-27页</td> <td>1</td> </tr> <tr> <td>A</td> <td>CN 109299436 A (北京邮电大学) 2019年 2月 1日 (2019 - 02 - 01) 全文</td> <td>1-8</td> </tr> <tr> <td>A</td> <td>CN 107871087 A (广西师范大学) 2018年 4月 3日 (2018 - 04 - 03) 全文</td> <td>1-8</td> </tr> <tr> <td>A</td> <td>CN 108256000 A (武汉大学) 2018年 7月 6日 (2018 - 07 - 06) 全文</td> <td>1-8</td> </tr> <tr> <td>A</td> <td>US 2018173894 A1 (SAP SE) 2018年 6月 21日 (2018 - 06 - 21) 全文</td> <td>1-8</td> </tr> <tr> <td>A</td> <td>高志强等. “本地差分隐私保护及其应用” 计算机工程与科学, 2018年 6月 30日 (2018 - 06 - 30), 第1029-1036页</td> <td>1-8</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110022531 A (华南理工大学) 2019年 7月 16日 (2019 - 07 - 16) 权利要求1-8	1-8	X	高志强等. “满足本地差分隐私的位置数据采集方案” 清华大学学报 (自然科学版), 2019年 1月 31日 (2019 - 01 - 31), 第23-27页	1	A	CN 109299436 A (北京邮电大学) 2019年 2月 1日 (2019 - 02 - 01) 全文	1-8	A	CN 107871087 A (广西师范大学) 2018年 4月 3日 (2018 - 04 - 03) 全文	1-8	A	CN 108256000 A (武汉大学) 2018年 7月 6日 (2018 - 07 - 06) 全文	1-8	A	US 2018173894 A1 (SAP SE) 2018年 6月 21日 (2018 - 06 - 21) 全文	1-8	A	高志强等. “本地差分隐私保护及其应用” 计算机工程与科学, 2018年 6月 30日 (2018 - 06 - 30), 第1029-1036页	1-8
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
PX	CN 110022531 A (华南理工大学) 2019年 7月 16日 (2019 - 07 - 16) 权利要求1-8	1-8																								
X	高志强等. “满足本地差分隐私的位置数据采集方案” 清华大学学报 (自然科学版), 2019年 1月 31日 (2019 - 01 - 31), 第23-27页	1																								
A	CN 109299436 A (北京邮电大学) 2019年 2月 1日 (2019 - 02 - 01) 全文	1-8																								
A	CN 107871087 A (广西师范大学) 2018年 4月 3日 (2018 - 04 - 03) 全文	1-8																								
A	CN 108256000 A (武汉大学) 2018年 7月 6日 (2018 - 07 - 06) 全文	1-8																								
A	US 2018173894 A1 (SAP SE) 2018年 6月 21日 (2018 - 06 - 21) 全文	1-8																								
A	高志强等. “本地差分隐私保护及其应用” 计算机工程与科学, 2018年 6月 30日 (2018 - 06 - 30), 第1029-1036页	1-8																								
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																										
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																									
国际检索实际完成的日期 2020年 3月 21日		国际检索报告邮寄日期 2020年 4月 10日																								
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 闫洪波 电话号码 86-(10)-53961740																								

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/070657

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110022531	A	2019年 7月 16日	无	
CN	109299436	A	2019年 2月 1日	无	
CN	107871087	A	2018年 4月 3日	无	
CN	108256000	A	2018年 7月 6日	无	
US	2018173894	A1	2018年 6月 21日	CN 108229181 A	2018年 6月 29日
				EP 3340101 A1	2018年 6月 27日