

面向LBS的ADS-B报文定位数据本地化差分隐私保护方法

杨宏宇^{1,2}, 张恒鸣¹, 朱友文³, 胡慧玲⁴, 胡俊成⁵

- (1. 中国民航大学安全科学与工程学院, 天津 300300;
2. 中国民航大学计算机科学与技术学院, 天津 300300;
3. 南京航空航天大学计算机科学与技术学院, 江苏 南京 211106;
4. 民航中南地区空中交通管理局空管部, 广东 广州 510440;
5. 吉林大学计算机科学与技术学院, 吉林 长春 130012)

摘要: 由于广播式自动相关监视 (automatic dependent surveillance-broadcast, ADS-B) 报文格式完全公开且重要数据缺乏加密措施, 因此敏感定位数据面临泄露的风险。为解决不可信第三方基于位置服务 (untrusted third-party location based service, UT-LBS) 在收集和发布 ADS-B 报文定位数据时造成的隐私泄露问题, 提出一种阶段式等差随机响应 (phased equal-step randomized response, PERR) 隐私保护方法。首先, 对三维定位数据进行降维处理, 通过二进制交叉编码生成一维八进制定位信息编码。其次, 使用阶段式等差随机响应方法对一维定位信息编码进行扰动, 通过所提出的等差扰动的思想对随机响应空间进行阶梯式分组。然后, 将随机响应分成基于等差隐私的随机响应和基于等差概率的自适应随机响应空间划分两个阶段, 提高随机响应过程中数据的隐私性。最后, 通过所提出的基于阶段式概率加权的卡尔曼滤波器估算方法, 对随机响应后的隐私数据进行频率估算, 并分析证明该方法的隐私性和可行性。实验结果表明, 所提方法在实现 ADS-B 敏感定位数据隐私保护的同时, 有效减小隐私保护后数据的频数统计与原始数据的频数统计之间的误差, 保证了数据的有效性。

关键词: 本地化差分隐私; 隐私保护; 随机响应; ADS-B 报文定位数据; 基于位置服务

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2025041

Differential privacy protection method for localization of ADS-B message positioning data oriented to LBS

YANG Hongyu^{1,2}, ZHANG Hengming¹, ZHU Youwen³, HU Huiling⁴, HU Juncheng⁵

1. School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China
2. School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China
3. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
4. Air Traffic Management Department, Central and Southern Regional Air Traffic Management Bureau of CAAC, Guangzhou 510440, China

收稿日期: 2025-05-08; **修回日期:** 2025-08-02

通信作者: 杨宏宇, yang_hy2006@126.com

基金项目: 国家自然科学基金民航联合研究基金 (U2433205)

Foundation Item: Civil Aviation Joint Research Fund Project of the National Natural Science Foundation of China (U2433205)

引用格式: 杨宏宇, 张恒鸣, 朱友文, 等. 面向LBS的ADS-B报文定位数据本地化差分隐私保护方法[J]. 网络与信息安全学报, 2025, 11(4): 32-49.

Citation Format: YANG H Y, ZHANG H M, ZHU Y W, et al. Differential privacy protection method for localization of ADS-B message positioning data oriented to LBS[J]. Chinese Journal of Network and Information Security, 2025, 11(4): 32-49.

5. School of Computer Science and Technology, Jilin University, Changchun 130012, China

Abstract: Due to the completely open format of automatic dependent surveillance-broadcast (ADS-B) messages and the lack of encryption for critical data, sensitive location information was put at risk of exposure. To address the privacy leakage issues caused by untrusted third-party Location-Based Services during the collection and publication of ADS-B message location data, a phased equal-step randomized response privacy protection method was proposed. Dimensionality reduction was performed on the three-dimensional location data, and a one-dimensional octal location information code was generated through binary crossover encoding. A phased equal difference randomized response method was employed to perturb the one-dimensional location information code. The concept of equal difference perturbation was utilized to create ladder-like groupings in the randomized response space. Subsequently, the randomized response process was divided into two phases: a random response based on equal difference privacy, and an adaptive random response space partitioned by equal difference probability, thereby enhancing data privacy throughout the randomization process. Finally, a phase-based probabilistic weighted Kalman filter estimation method was proposed to estimate the frequency of the privacy data after randomization. The privacy and feasibility of this approach were demonstrated through theoretical analysis. Experimental results indicate that the proposed method effectively protects the privacy of ADS-B sensitive location data while significantly reducing the error between the frequency statistics of the privacy-protected data and the original data, thus ensuring data validity.

Keywords: local differential privacy, privacy protection, randomized response, ADS-B message positioning data, location based service

0 引言

目前广播式自动相关监视 (automatic dependent surveillance-broadcast, ADS-B) 系统已成为航空交通服务 (air traffic services, ATS) 的重要组成部分, 其中, 广播式自动相关监视发射 (automatic dependent surveillance-broadcast output, ADS-B OUT) 功能模块使用发射器在飞行器飞行期间以固定的间隔广播飞行器所处的位置、飞行高度、飞行速度、身份识别码等信息; 广播式自动相关监视接收 (automatic dependent surveillance-broadcast input, ADS-B IN) 功能模块负责接收ADS-B报文, 会对其进行处理并将ADS-B报文携带的数据信息显示在管制员的屏幕上, 用于飞行监视等应用^[1]。然而, 当前的ADS-B系统采用开放、共享的广播式架构, 其协议公开且报文未加密, 缺少必要的认证, 导致ADS-B系统面临多方面安全挑战, 如敏感定位数据易泄露、报文信息易造假、成员复杂难管理、广播报文难追溯等问题^[2-4]。本文针对敏感定位数据易泄露这一关键问题展开研究, 由于ADS-B报文格式公开, 数据缺少加密, 攻击者只需利用简单的无线电设备就可以接收并解析ADS-B广播的报文, 获取飞行器的身份、位

置、速度、高度、航向等关键运行信息, 目前, 空中交通图像可在OpenSky Network、Flightradar24网站和Flight Tracker软件上公开获取, 公众可以通过这些网站和软件观测每架配备ADS-B OUT传感器的飞行器, 公开广播的ADS-B数据信息在给研究者带来便利的同时, 也存在以下安全隐患: 攻击者可以通过基于位置服务 (location based service, LBS) 采集ADS-B数据中的定位信息, 并对特定飞行器进行跟踪^[5], 或长期收集数据并分析空中交通模式策划GPS欺骗攻击。ADS-B报文提供的精确、实时定位数据, 构成下面所述3类航空LBS的基础数据支撑: ①航班追踪与公众信息查询服务: Flight-radar24、OpenSky Network等网站利用ADS-B定位数据向公众和航空公司提供实时的航班位置、状态和轨迹信息, 是旅客行程规划、延误查询、航空公司运行以及监控的基础; ②实时空中交通监控与流量管理服务: 空中交通管理系统依靠连续的ADS-B IN数据流在屏幕上构建高保真的空中交通态势图, 用于指挥调度、冲突解脱、优化航路等操作, 敏感定位数据泄露可直接威胁飞行器飞行安全与空域飞行效率; ③空域规划与性能分析服务: 民航空管部门利用历史

ADS-B 定位数据进行空域结构评估、航线网络优化等操作,敏感定位数据的完整性和准确性直接决定分析结论的可靠性。因此,研究并解决 ADS-B 报文敏感定位数据易泄露的问题意义重大。

近年来,差分隐私(differential privacy, DP)技术在位置数据隐私保护领域取得显著进展。在动态位置保护方面,李洪涛等^[6]提出通过路网拓扑划分敏感路段隐私级别算法,结合树形结构分配隐私预算并添加拉普拉斯噪声,有效解决了连续位置关联隐私泄露问题;在连续轨迹发布方面,陈思等^[7]提出了融合时间泛化与空间分割采样模型,通过 k -means 聚类 and 双重扰动提升数据可用性,其预判机制显著减少了空轨迹的发布。孙新越等^[8]指出现有位置数据隐私保护方案大多数基于中心化差分隐私框架,面临超大规模隐私保护轨迹数据发布时数据效用过低、第三方基于位置服务隐私泄露等挑战。文献[6-7]的方案均依赖可信第三方数据聚合,而 ADS-B 数据可直接被不同航空 LBS 收集用作下游任务,存在不可信第三方泄露位置隐私的重大问题,针对上述挑战,本文提出本地化差分隐私(local differential privacy, LDP)保护方案。

目前,本地化差分隐私技术^[9-10]已成功应用于工业领域,Google、Apple 和 Microsoft 等利用 LDP 技术对用户本地数据进行隐私保护和数据聚合,但通用 LDP 方案难以满足航空定位数据隐私保护的特定需求。LDP 技术在对抗下游任务分析者、不可信的数据聚合者等具有任意背景知识的攻击者时能提供强大的隐私保护。目前,常用的随机化可聚合隐私保护序数响应(randomized aggregatable privacy-preserving ordinal response, RAPPOR)^[11]和多元随机响应(k -ary randomized response, KRR)^[12]等 LDP 方案已被扩展并应用于隐私聚合多种类型数据,如单属性单值数据、单属性集值数据、视频和图。然而,常见的 LDP 方案在聚合隐私位置数据时受到显著限制,其主要原因是编码与扰动阶段未针对不同数据进行自适应优化,且缺乏必要的数据聚合阶段,目前根据编码和扰动的方式将 LDP 方案划分为两类。

1) 直接扰动:如 KRR 无须编码可直接随机响应,实现简单且计算开销小,只适用于单属性单值和集值数据,对位置数据不进行预处理直接扰动会造成极大的信息损失。

2) 先编码后扰动:最优一元独热编码^[13](optimized unary encoding, OUE)、最优局部哈希^[13](optimal local hashing, OLH)、哈达玛随机响应^[14](hadamard random response, HR)等方案,通过对单属性单值数据进行先编码后随机响应,极大地降低通信开销,但在处理位置数据这类多属性集值数据时,缺乏对数据特征的针对性优化,导致较大的数据效用损失。

然而,将现有通用 LDP 方案直接应用于保护 ADS-B 定位数据并服务于航空 LBS 应用时存在明显的局限性。LBS 应用场景与隐私保护关联框架如图 1 所示,现有 LDP 方案难以满足航空 LBS 应用对数据效用的严苛要求,KRR 等直接扰动方案会造成明显的空间信息损失,导致数据聚合后用于生成空中交通态势图的空间密度估计产生严重误差,无法进行可靠的冲突预警或流量管理;位置数据聚合方案^[15]的分层扰动仅按固定半径划分随机响应空间,且不支持三维耦合编码,无法满足 ADS-B 报文信息中三维定位数据隐私保护和保持航空器飞行数据连续性的特殊需求;航空 LBS 需亚百米级定位精度,OUE、OLH、HR 等先编码后扰动方案频率估算精度不足,难以准确统计特定空域单元内的飞行器数量,影响空中流量统计的准确性;空间密度估计误差过大,生成的热力图模糊不清,难以识别真实的空中交通热点或异常聚集区域,严重影响空域规划分析和异常行为检测的效果。

为了保护用户位置隐私信息,研究者提出面向定位数据的 LDP 保护方案^[16-19]。Fang 等^[16]利用卷积和反卷积思想优化现有的 LDP 方案,并广泛应用到各类平滑数据集,展现出良好的数据隐私聚合效果,但在面对 ADS-B 报文数据集等非平滑数据集时处理有限,导致频率估算精度下降。Kim 等^[17]的方法在较大范围的空间场景中表现出一定的局限性,在面对 ADS-B 数据时空间密度估计的总误差较大,难以有效地

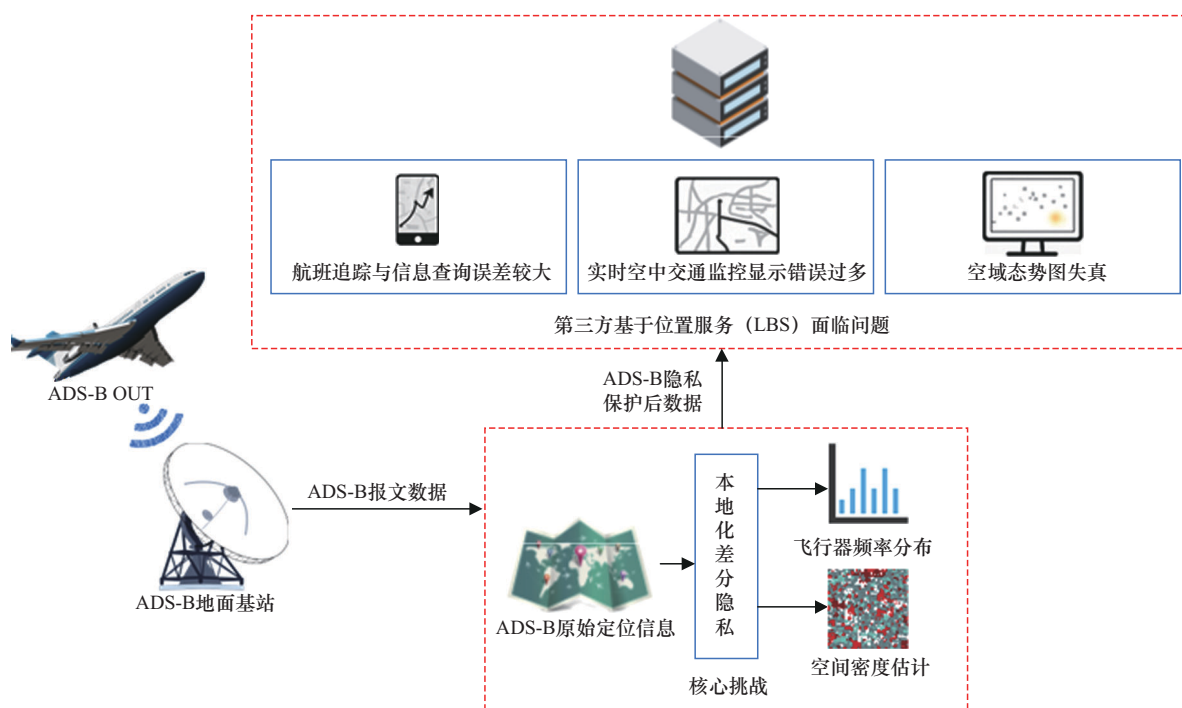


图1 LBS应用场景与隐私保护关联框架

Figure 1 LBS application scenarios and privacy protection association framework

进行空间密度估算。Zhao等^[18]和Wang等^[19]的方法在用户定位数据发布和面向LBS以及对个人位置数据实施隐私保护时性能比较突出，但仅针对二维定位数据（纬度，经度）进行了研究。为解决上述问题，满足航空领域LBS应用对隐私保护飞行器定位数据的高效用需求，本文提出一种面向LBS的ADS-B定位数据本地化差分隐私保护方法。在充分实现ADS-B敏感定位数据隐私保护的基础上，提高定位数据聚合估算效果，降低聚合后的数据频率估算结果与原始数据频率估算结果之间的误差，提高航空LBS位置查询服务的整体精度。

1 相关技术

1.1 广播式自动相关监视系统

广播式自动相关监视系统利用无线通信协议，通过数据链路广播飞机的识别和导航信息。飞行器利用导航卫星确定其位置，并将位置信息与身份等重要信息合并，通过ADS-B OUT在标准监视频道上以开放格式广播。地面、空中和空间中多个ADS-B IN收集并分析这些信息，形成完整的监视图，供地面控制员和飞行员使

用，以提高空中交通管理的效率。此外，由于其开放架构和广播性质，这些数据还可以通过互联网查看，因此ADS-B报文数据存在安全和隐私风险。

1.2 基于位置服务

LBS应用程序收集用户本地上传的原始位置数据并提供相关服务。但目前存在某些不可信第三方服务器私下聚合并分析用户敏感位置数据，从而识别用户所在的位置域并进行空间密度估计。在某些LBS应用程序中，客户端可以从服务器端查询ADS-B报文定位数据聚合后的分析结果，攻击者可以从OpenSky Network、Flightradar24等网站查询指定航班号飞行器的经纬度、海拔、航向等具体信息，并进行实时跟踪。

1.3 本地化差分隐私

本地化差分隐私是一种保护用户隐私的机制，它在数据收集的过程中，确保即使攻击者获得用户数据，也无法准确推测用户的实际信息。这种隐私保护机制特别适用于分布式环境，用户数据在上传之前进行随机化处理，从而在不依赖第三方可信数据收集者的情况下保护用户隐私。

定义1 ϵ -本地化差分隐私(ϵ -LDP)^[20]通过控

制任意两条记录输出结果的相似性来确保算法的隐私性。设存在随机化算法 Q ，对于任意输入数据 g 和 $g' \in D$ ，当且仅当任意扰动输出 $G \in \text{range } Q$ 上传至不可信第三方服务器时，满足：

$$\Pr[Q(g) = G] \leq e^\epsilon \times \Pr[Q(g') = G] \quad (1)$$

则称随机化算法 Q 满足 ϵ -本地化差分隐私， ϵ 是隐私预算，表示用户数据的隐私保护强度， ϵ 值越小，隐私保护程度越高，但相应的数据可用性就会降低。

引理 1 ϵ -本地化差分隐私具有串行组合性质，当多个 LDP 机制按串行应用时，隐私保护水平可以通过各个机制的隐私预算相加得到。即若存在两个 ϵ -LDP 机制串行应用，将得到一个 $(\epsilon_1 + \epsilon_2)$ -LDP 机制。

引理 2 ϵ -本地化差分隐私具有并行组合性质，当多个 LDP 机制并行应用于不同的数据子集时，整体算法的隐私预算 ϵ 等于单个 ϵ -LDP 机制中最大的隐私预算 $\epsilon_{\max}$ 。即若存在 k 个 ϵ -LDP 机制并行应用，其整体满足 $\epsilon_{\max}$ -LDP。

2 PERR 隐私保护方法

2.1 ADS-B 报文定位数据隐私保护框架

为了解决 ADS-B 报文敏感定位数据隐私易

泄露的问题，本文提出了面向 LBS 的 ADS-B 定位数据本地化差分隐私保护方法，其总体框架如图 2 所示，主要包括 3 类参与实体：飞行器、ADS-B 地面基站、不可信第三方服务器。

1) 飞行器：飞行器向导航卫星发送 GPS 定位信息请求，获得自身的 GPS 定位信息，然后将其位置信息与身份等重要信息合并，编码为 ADS-B 报文并广播发送。

2) ADS-B 地面基站：ADS-B 地面基站接收广播的报文，随后解码出各种信息，如航班号、经纬度、海拔等信息，并在本地服务器中对飞行器 ADS-B 报文数据中的定位信息进行编码处理，使用随机响应机制对编码后的飞行器定位信息进行随机扰动，然后地面接收器本地上传扰动后的数据编码到基于位置服务的不可信第三方服务器。

3) 不可信第三方服务器：第三方服务器端在收集到所有的 ADS-B 扰动数据后，进行数据隐私聚合并估算和更新飞行器定位信息，得到飞行器位置分布频率等分析结果，用于后续的空间密度估计、位置查询等服务。

由图 2 可知，飞行器将自身 ADS-B 报文广播发送，ADS-B 地面基站接收到解码后的 ADS-B

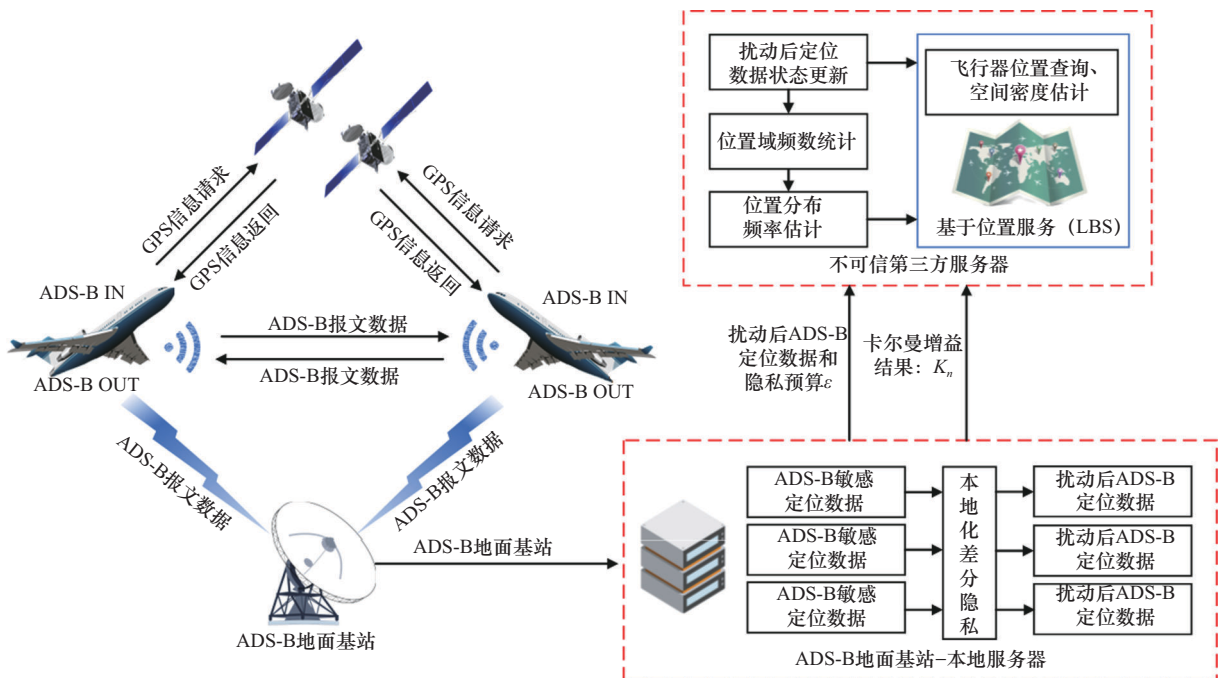


图 2 面向 LBS 的 ADS-B 定位数据本地化差分隐私保护框架

Figure 2 Local differential privacy framework for ADS-B location data protection in location-based service

报文数据，对其中的敏感定位信息进行三维定位数据编码和阶段式等差随机响应，然后将扰动后的定位信息编码和利用基于阶段式概率加权的卡尔曼滤波器估算方法得到的目标参数一并传输给不可信第三方服务器，第三方服务器端对数据进行隐私聚合并估算，最后做下游分析任务，在保证敏感数据隐私保护的同时，极大减小扰动后数据频数统计与原始数据频数统计之间的误差。上述过程无须第三方对数据进行隐私预处理，确保了ADS-B报文定位数据的隐私不受不可信第三方的威胁，且与现有的LDP模型类似，本文方法可确保数据处理过程的隐私性，防止基于任意背景知识的攻击者对ADS-B报文原始定位数据进行推断，且本文方法与降低其他类型风险的方法（如加密^[21]、身份认证^[22]和侧信道攻击防御等方法）可以结合，进一步提高了ADS-B报文数据处理过程的安全性和隐私性。

2.2 三维定位数据编码

在对ADS-B报文敏感定位数据进行随机响应前，首先针对ADS-B报文数据中的经度、纬度、海拔进行编码。

传统的随机响应机制通常只针对单属性单值和单属性集值数据，为了解决对原始ADS-B报文敏感定位数据直接随机响应产生的精度丢失问题，本文提出三维定位数据编码方法，将三维地理数据编码降维为一维地理数据编码，使其符合单属性集值特征，以减少计算开销和精度损失。

将十进制定位数据格式转换成度分秒格式，

经纬度格式为：度°分'秒"，为匹配ADS-B报文精度，本文采用国家标准《地球空间网格编码规则》^[23]中的18级剖分方案：在赤道附近，该级别单网格对应8"×8"（约247.4 m×247.4 m），并以此为粒度进行编码，设置度°位用9位二进制码表示，分'位用6位二进制码表示，秒"位用3位二进制码表示，共18位二进制码。

为保持一致性，将海拔精确到5位十进制整数，航空领域海拔变化对安全的影响更敏感，民航巡航高度万位变化代表重大高度变化（如从平流层到对流层），1 bit足以覆盖这种阶跃式变化特征，故需独立设计高位权重，设置万位用1位二进制码表示，千位用4位二进制码表示，百位用4位二进制码表示，十位和个位用9位二进制码表示，共18位二进制码。

将纬度、经度、海拔的二进制码表示为： x （纬度）、 y （经度）、 z （海拔），对其进行交叉编码为 $z|y|x$ 形式，后将 $z|y|x$ 的二进制混合编码，每三位转成八进制编码，即18位八进制编码，如图3所示，具体算法如下。

算法1 三维定位数据编码

输入 ADS-B原始定位数据 s

输出 三维定位数据编码 g

1)degrees = int(decimal_degrees)

2)minutes = int((decimal_degrees-degrees)* 64)

3)seconds = (decimal_degrees-degrees - minutes / 64)*64*64

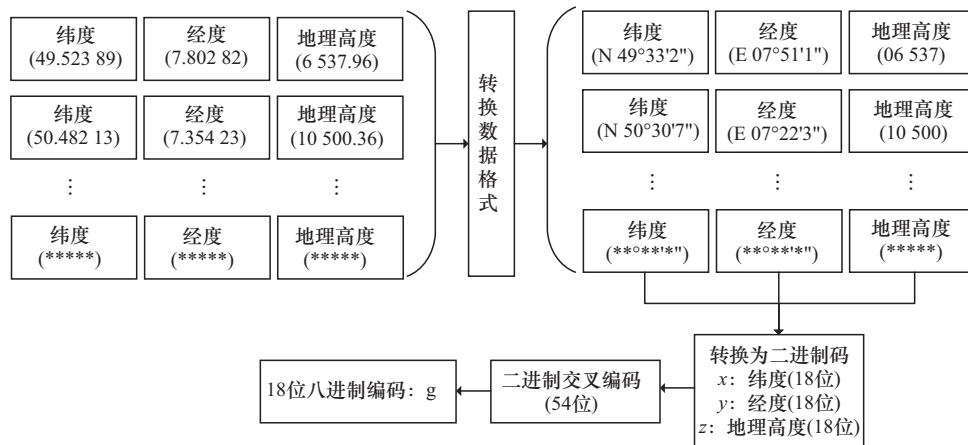


图3 三维定位数据编码方案

Figure 3 Three-dimensional positioning data encoding scheme

```

4)  $s_i = \text{degrees:02d'minutes:02d'seconds:01d''}$ 
5) RETURN  $s_i$ //将定位数据  $s$  中经纬度格式化为度分秒格式  $s_i$ 
6) 将  $s_i$  与海拔  $h$  转换为二进制格式:  $x, y, z$ 
7) result = ''
8) FOR  $i$  IN range(18):
9)   result +=  $z[i] + y[i] + x[i]$ 
10) RETURN result//将 3 个二进制字符串进行交叉编码
11)  $g = ''$ 
12) FOR  $i$  IN range(0, len(result), 3):
13)    $g += \text{str(int(binary[i:i+3], 2))}$ 
14) RETURN  $g$ //将二进制字符串转换为八进制字符串

```

算法执行示例: ADS-B 数据集中一条民航飞行器的定位数据为经度: 49.523 895 263 671 9, 纬度: 7.802 822 715 357 73, 海拔: 6 537.96。对其应用三维数据编码方案, 首先进行格式转换, 转换后数据为经度: N 49°33'2"s, 纬度: E 07°51'1"s, 海拔: 06537, 然后进行编码, 编码后数据为“004 510 627 320 423 416”(18 位八进制码)。

2.3 阶段式等差随机响应

为提高随机响应机制的隐私聚合能力, 减少定位数据处理过程中的信息损失, 本文提出一种阶段式等差随机响应 (phased equal-step randomized response, PERR) 机制。

PERR 机制分为两个阶段: 基于等差概率的自适应随机响应空间划分和基于等差隐私的随机响应。对降维后的地理数据编码进行随机响应处理, 通过设置全局敏感度实现随机响应范围的合理划分, 并利用等差思想提高随机响应的合理性, 减少定位信息精度丢失, 提高隐私聚合能力。PERR 方案如图 4 所示, 具体算法如下。

算法 2 阶段式等差随机响应

输入 隐私预算 ϵ , 全局敏感度 Δf , 三维定位数据编码 g , 等差隐私预算 $\Delta \epsilon$, 等差扰动 ΔP

输出 随机响应后编码数据 G

```

1) 根据隐私预算  $\epsilon$ 、全局敏感度  $\Delta f$ 、等差扰动  $\Delta P$  和等差隐私预算  $\Delta \epsilon$  计算出  $\epsilon_i$  和  $P_i$ 
2)  $r = \text{random.random}()$ 

```

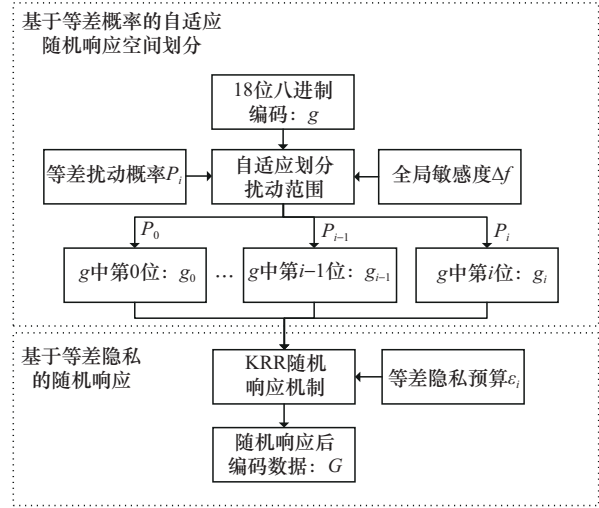


图 4 阶段式等差随机响应方案

Figure 4 Phased equal-step randomized response scheme

```

3) IF  $r < P_0$ :
4)   position = -4, epsilon = epsilon0
5) ELIF  $r < P_0 + P_1$ :
6)   position = -3, epsilon = epsilon1
7) ELIF  $r < P_0 + P_1 + P_2$ :
8)   position = -2, epsilon = epsilon2
9) ELSE: position = -1, epsilon = epsilon3//按
    概率  $P_{14} \sim P_{17}$  选择扰动位数 (第 14 位到第 17 位)
10) true_value = int( $g[\text{position}]$ )
11) e_epsilon = math.exp(epsilon)
12) 根据  $\epsilon_i$  和  $P_i$  实现阶段式等差随机响应
13)  $p = e\_epsilon / (k - 1 + e\_epsilon)$ 
14)  $q = 1 / (k - 1 + e\_epsilon)$ 
15) IF random.random() <  $p$ :
16)   RETURN true_value
17) ELSE: RETURN random.choice([i FOR i
    IN range(k) IF i != true_value])

```

18) RETURN $G = \text{KRR}(\text{true_value}, 8, \text{epsilon})$ //对选中位置的八进制码应用 KRR 随机响应

传统的随机响应方法利用不同的随机应答机制对输入数据实施隐私保护, 以概率 p 真实应答, 概率 $1-p$ 相反应答。但在面对降维后的复杂敏感定位数据编码时, 具有较大的局限性。编码的每一位对应不同的三维地理空间, 传统的随机响应方法会导致随机应答后的编码误差过大。因此, 本文将随机响应机制分为两个阶段, 有效克服了这一局限性。

(1) 基于等差概率的自适应随机响应空间划分

设位置域 D , 输入 $g \in D$ 为18位八进制一维定位数据编码, 输出 $G \in D$ 为经随机响应扰动后的18位八进制编码。

设位置查询函数 $f(g): g \rightarrow D$, 全局敏感度 $\Delta f = \max_{D1, D2 \subseteq D} ||f(D1) - f(D2)||_1$, 全局敏感度 Δf 是指ADS-B原始定位数据编码 $D1$ 与随机响应后数据编码 $D2$ 之间的最大 L_1 范数。

设扰动概率 P_i ($0 \leq i \leq 17$) 是 g 中第 i 位八进制 g_i 的选择扰动概率, 且 P_i 对应18个互斥事件。由2.2节中三维定位数据编码可知, g 中的高位变化比低位变化更接近原始定位数据编码, 与原始ADS-B报文定位数据之间的 L_1 范式距离更小, 因此, 选择高位的扰动概率 P_n 应大于选择低位的扰动概率 P_{n-1} , 且 g 越高位选择概率应越大, 所以在扰动概率之间设置相同步长。本文利用等差的思想设置等差概率 ΔP , 满足条件:

$$P_i + (n - i - 1)\Delta P = P_n \quad (2)$$

通过全局敏感度 Δf 和扰动概率 P_i 自适应划分随机响应空间, 划分方式如下:

$$G_i = \begin{cases} g_0, P_0, 10^{17} \leq \Delta f \leq 10^{18} \\ g_1, P_1, 10^{16} \leq \Delta f \leq 10^{17} \\ \dots \\ g_i, P_i, 10^{17-i} \leq \Delta f \leq 10^{18-i} \end{cases} \quad (3)$$

其中, Δf 的指数级变化对应最大 L_1 范式区间, 从而选定 g 的扰动位数范围。

如图5所示, 根据先验知识, 已知民航飞行器的平均速度为250 m/s, OpenSky Network平台提供ADS-B报文数据集中报文时间间隔为1.5s, 因此设全局敏感度 $0 \leq \Delta f \leq 10^4$ 。根据式(3)可知, g 的扰动位数范围为 $g_{14} \sim g_{17}$, 扰动概率为 $P_{14} \sim P_{17}$, 随机响应空间自适应划分为: $G_{14} \sim G_{17}$ 。其中, $G_{17} = 30 \text{ m} \times 30 \text{ m} \times 1 \text{ m}$; $G_{16} = 90 \text{ m} \times 90 \text{ m} \times 3 \text{ m} - G_{17}$; $G_{15} = 210 \text{ m} \times 210 \text{ m} \times 7 \text{ m} - (G_{17} + G_{16})$; $G_{14} = 2010 \text{ m} \times 2010 \text{ m} \times 15 \text{ m} - (G_{17} + G_{16} + G_{15})$ 。

(2) 基于等差隐私的随机响应

首先, 根据式(3)中划分后的随机响应空间 G_i 以及扰动位数 g_i , 选定隐私预算 ϵ_i 。与式(3)中利用等差的思想设置等差概率 ΔP 同理, 在隐私预算 ϵ_i 之间设置相同步长: 等差隐私预算 $\Delta \epsilon$, 且

选择的扰动位数 g_i 越高位, 随机响应结果为原始数据的概率越大, 对其分配的隐私预算 ϵ_i 也越大。以全局敏感度 $0 \leq \Delta f \leq 10^4$ 为例, 进行基于等差概率的自适应随机响应空间划分, 由式(3)可得划分后随机响应空间为 $G_{14} \sim G_{17}$, 之后对 g 进行基于等差隐私的随机响应, 其满足KRR随机响应机制:

$$K(G|g) = \frac{1}{k - 1 + e^\epsilon} \begin{cases} e^\epsilon, \text{概率为 } p, g = G \\ 1, \text{概率为 } q, g \neq G \end{cases} \quad (4)$$

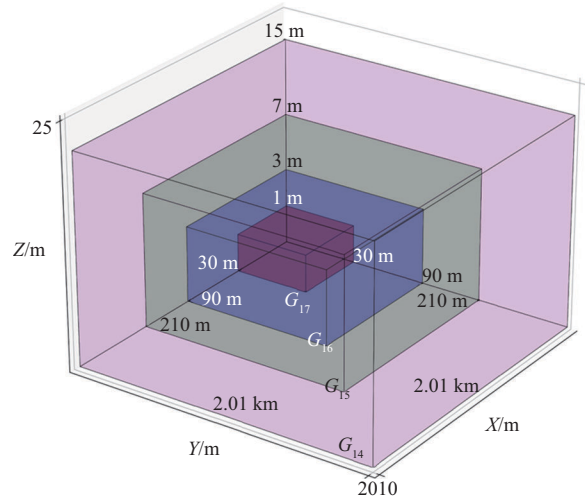


图5 分组随机响应空间范围

Figure 5 Spatial scope of grouped randomized response

综上, 本文提出一种阶段式等差随机响应方法PERR的计算过程如下:

$$Q(G|g) = \begin{cases} \frac{P_i(\cdot)e^{\epsilon_i}}{k - 1 + e^{\epsilon_i}} = p_i, g = G, G \in G_{i+1} \\ \frac{P_i(\cdot)}{k - 1 + e^{\epsilon_i}} = q_i, g \neq G, G \in G_{i+1} \end{cases} \quad (5)$$

2.4 数据聚合与频率估算

传统的随机响应频率估算方法一般只适用于平滑的数据集, 而对ADS-B报文三维定位数据进行随机响应通常包含大量过程噪声, 信息误差较大, 因此, 本文扩展了基于概率加权的卡尔曼滤波器估算方法, 对随机响应后的定位数据进行频率估算和频数统计。

通过状态转移和状态观测过程对编码与随机响应后的 G 进行最优估计, 如图6所示, 具体算法如下。

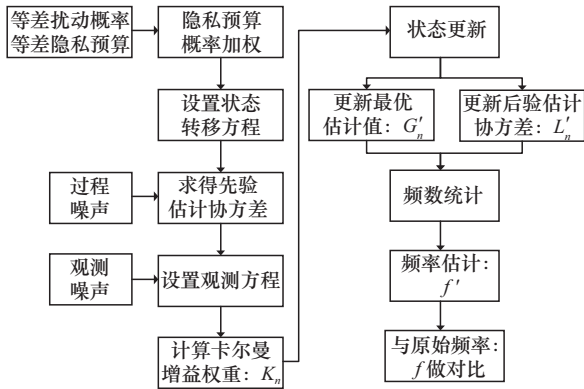


图6 数据聚合与频率估算方案

Figure 6 Data aggregation and frequency estimation scheme

算法3 基于卡尔曼滤波器的估算方法

输入 三维定位数据编码 g , 随机响应后编码数据 G , 隐私预算 ε_i , 扰动概率 P_i

输出 最优估计值 G'_n , 频率估算值 f'

```

1) N_10 = octal_to_decimal(G_n-1)
2) FOR r IN range(8):
3)   FOR delta_t IN range(8):
4)     IF (N_10 * delta_t) % m == r:
5)       RETURN delta_t // 计算  $\Delta t_i$ 
6) e_epsilon = np.exp(epsilon)
7) p = e_epsilon / (k - 1 + e_epsilon)
8) q = 1 / (k - 1 + e_epsilon) // 根据 KRR 机制求得  $p_i$  和  $q_i$ 
9) FOR i IN range(4):
10)   p, q = calculate_probabilities(epsilon[i])
11)   delta_t = calculate_delta_t(g, i + 1)
12)   A[2 * i] = P[i] * p
13)   A[2 * i + 1] = P[i] * q * delta_t
14) RETURN A // 求得状态转移矩阵 A
15) sigma_squared = (np.exp(epsilon + k - 2)) / (m * (np.exp(epsilon + k - 1) - 1) * 2) // 计算分层式方差  $\delta_i^2$ 
16) 得到过程噪声协方差 Q
17) L_n = np.dot(A, np.dot(L_n_prev, A.T)) + Q // 迭代计算协方差矩阵 L_n
18) HLn = np.dot(H, L_n)
19) HLnHT_plus_R = np.dot(HLn, H.T) + R
20) RETURN K_n = np.dot(np.dot(L_n, H.T), np.linalg.inv(HLnHT_plus_R)) // 计算卡尔曼增益 K_n
21) RETURN G'_n = G_n + K_n_1_1 * vn
  
```

$$22) L'_n = (1 - K_{n-1_1} * H) * L_n$$

23) RETURN $F' = L'_n / m$ // 计算最优估计值 G'_n 和频率估计值 f'

算法3中, 首先对原始定位数据编码 g 进行先验预测估计: 对算法2输出的 $G \in D$ 进行聚合估算, 设置状态变量 G_n , 即编码并随机响应后输出 $G \in D$, 令 $G_n = \{G \in D\}$ 。

设状态转移方程:

$$G_n = AG_{n-1} + Bu_n \quad (6)$$

其中, A 为状态转移矩阵, B 为控制矩阵, u_n 为控制输入, G_n 为 n 时刻的先验估计值, G_{n-1} 为 $n-1$ 时刻的最优估计值。

已知随机响应后定位数据编码 G 与原始定位数据编码 g , 令 g 作为 $n-1$ 时刻的最优估计 G_{n-1} , G 作为 n 时刻的先验估计值 G_n , 可得此时的 PERR 计算过程:

$$Q(G_n | G_{n-1}) = \begin{cases} \frac{P_i(\cdot) e^{\varepsilon_i}}{k - 1 + e^{\varepsilon_i}} = p_i, G_n = G_{n-1} \\ \frac{P_i(\cdot)}{k - 1 + e^{\varepsilon_i}} = q_i, G_n = \Delta t_i G_{n-1} \end{cases} \quad (7)$$

其中, $0 \leq i \leq 17$, Δt_i 为转移变量, G_i 为随机响应空间范围, $G_n \in G_i$ 。

由上述过程可计算出先验估计值 G_n :

$$G_n = \{P_i p_i, P_i q_i \Delta t_i, \dots, P_j p_j, P_j q_j \Delta t_j\}^T G_{n-1} + O u_n \quad (8)$$

O 为零矩阵, 得到先验估计值 G_n 后计算状态转移矩阵 A :

$$A = \{P_i p_i, P_i q_i \Delta t_i, \dots, P_j p_j, P_j q_j \Delta t_j\}^T \quad (9)$$

设 Δt_i 的取值范围为:

$$\Delta t_i = \left\lceil \frac{N_{10} \Delta t_i}{k^{4-i}} \right\rceil \bmod 8 \in \{0, 1, 2, 3, 4, 5, 6, 7\} \quad (10)$$

其中 $k=2^3$, N_{10} 为 G_{n-1} 的十进制码。

求得先验估计协方差 L_n :

$$L_n = AL_{n-1}A^T + Q \quad (11)$$

其中, L_{n-1} 为 $n-1$ 时刻的最优估计协方差, Q 为过程噪声 W_t 的协方差。此时 KRR 是纯 LDP 协议, 满足高斯分布, 而本文提出的 PERR 是混合分布, 即由离散型随机变量和连续型随机变量组合而成的混合概率分布事件, 可知过程噪声 W_t 的协方差 $Q = \text{COV}(W_t, W_t) = \sigma_t^2$, 根据中心极限定理 δ_i^2 近似服从于正态分布, 即过程噪声的协方差 Q 可用频率估算的近似方差 δ_{PERR}^2 表示:

$$Q = \sigma_i^2 = \sigma_{\text{PERR}}^2 |D| = \sum P_i \sigma_i^2 |D| \quad (12)$$

$$\sigma_i^2 = \frac{e^{\varepsilon_i} + k - 2}{m(e^{\varepsilon_i} + k - 1)^2} \quad (13)$$

其中, $m=|D|$, 已知 σ_{PERR}^2 在基于等差概率的自适应随机响应空间划分中扰动概率 P_i 确定的前提条件下, 仅受 ε_i 影响, 对 σ_{PERR}^2 进行概率加权:

$$\sigma_{\text{PERR}}^2 = \sum P_i \sigma_i^2 \quad (14)$$

可知过程噪声协方差 Q 与 ε_i 成反比。

设卡尔曼增益 K_n 计算过程如下:

$$K_n = L_n H^T (H L_n H^T + R)^{-1} \quad (15)$$

其中, R 为观测噪声的协方差, H 为观测矩阵, 设观测方程 Z_n 为:

$$Z_n = H G_n + v_n \quad (16)$$

其中, v_n 为观测噪声, 已知 G_n 为先验估计值, 即 G_{n-1} 随机响应前的原始近似值, 观测到的 Z_n 则为随机响应后的最终结果, 可得观测噪声 v_n 等于过程噪声 W , 并设观测矩阵 H 为单位矩阵 I 。求得 $K_n = L_n (L_n + R)^{-1}$, 同理 $R = \sigma_{\text{PERR}}^2 |D|$ 。

最终, 本地服务器上传卡尔曼增益 K_n 和先验估计值 G_n 到不可信第三方基于位置服务 (untrusted third-party location based service, UT-LBS) 进行处理。

UT-LBS 频率估算过程如下。

1) 对先验估计值 G_n 进行状态更新, 求得最优估计值 G'_n :

$$G'_n = G_n + K_n (Z_n - H G_n) \quad (17)$$

2) 更新后验估计协方差 L'_n :

$$L'_n = (I - K_n H) L_n \quad (18)$$

其中, I 为单位矩阵, L'_n 为最优估计值 G'_n 的后验估计协方差, 每次随机响应和卡尔曼迭代更新都可视为独立的伯努利试验, 即根据大数定律 L'_n 近似等于频数统计的方差 N' 。

3) 对最优估计值 G'_n 进行频率估算, 求得频率估算值 f' , 频率估算的近似方差 F' 为:

$$F' = \text{Var} f' = \frac{N'}{|D|} \quad (19)$$

2.5 隐私证明和数据效用分析

(1) 隐私证明

首先, 对 PERR 方法进行隐私分析。已知 KRR 满足纯 LDP 协议, 其符合正态分布, 根据差分隐私的串行和并行的扩展组合性质, 本文提出的 PERR 是混合分布, 即由离散型随机变量和

连续型随机变量组合而成的混合概率分布事件。PERR 满足 ε -LDP 的隐私定义, 其整体的隐私预算 $\varepsilon = \max\{\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3\}$, 证明不可信第三方基于位置服务无法从随机响应扰动后的 ADS-B 位置数据中判断出飞行器的真实位置。

以下用公式证明 PERR 方法满足 ε -本地化差分隐私。取 $\forall x, y \in D$, 对于相同的随机响应输出 G , 其概率分布如下:

$$Q(G|x) = \frac{P_i(\cdot)}{k-1+e^{\varepsilon_i}}$$

$$Q(G|y) = \frac{P_j(\cdot)}{k-1+e^{\varepsilon_j}}$$

其中, $k=2^3$, $0 < P_i < 1$, $0 < P_j < 1$, 比较两个概率分布可得:

$$\frac{Q(G|x)}{Q(G|y)} = \frac{P_i(\cdot)}{P_j(\cdot)} \cdot \frac{k-1+e^{\varepsilon_j}}{k-1+e^{\varepsilon_i}} =$$

$$\frac{P_i(\cdot)(k-1+e^{\varepsilon_j})}{P_j(\cdot)(k-1+e^{\varepsilon_i})} < \frac{(k-1+e^{\varepsilon_j})}{(k-1+e^{\varepsilon_i})}$$

已知 $\varepsilon = \max\{\varepsilon_i, \varepsilon_j\}$, 分两种情况讨论。

1) 当 $\varepsilon_i \leq \varepsilon_j$ 时:

$$\frac{(k-1+e^{\varepsilon_j})}{(k-1+e^{\varepsilon_i})} \leq e^{\varepsilon_j} = e^{\varepsilon}$$

2) 当 $\varepsilon_i > \varepsilon_j$ 时:

$$\frac{(k-1+e^{\varepsilon_j})}{(k-1+e^{\varepsilon_i})} < 1 < e^{\varepsilon_j} = e^{\varepsilon}$$

可证明 PERR 方法满足 ε -本地化差分隐私:

$$\frac{Q(G|x)}{Q(G|y)} = \frac{P_i(\cdot)}{P_j(\cdot)} \cdot \frac{k-1+e^{\varepsilon_j}}{k-1+e^{\varepsilon_i}} \leq e^{\varepsilon}$$

$$Q(G|y) = \frac{P_j(\cdot)}{k-1+e^{\varepsilon_j}}$$

(2) 数据效用分析

通过对频率估计误差的理论上限分析, 为数据效用理论提供保障。

推理过程如下, 已知卡尔曼滤波过程中过程噪声协方差矩阵 Q 为关键误差源, $Q = \sigma_i^2 = \sigma_{\text{PERR}}^2 |D| = \sum P_i \sigma_i^2 |D|$, 且 G'_n 的后验估计协方差 $L'_n = (I - K_n) L_n$ 。其中, n 为参与随机响应的独立样本数量, 频率估计误差 E 满足:

$$E = \sqrt{\text{Var}[f'(g)]} \leq \sqrt{\frac{Q}{n}} \leq$$

$$\sqrt{\frac{\sum P_i \sigma_i^2 |D|}{n}} \leq \sqrt{\frac{\sum P_i \frac{e^{\varepsilon_i + 6}}{m(e^{\varepsilon_i + 7} - 1)^2}}{n}}$$

3 实验与结果分析

实验采用的公开数据集来自 The OpenSky Network 网站, 数据集中的每条记录均包含 8 个分类属性。如表 1 所示, 本文只针对 ADS-B 数据集中的航班号、海拔以及经纬度信息实施隐私保护。首先对每条记录的航班号进行抑制, 然后采用本文提出的方法对飞行器位置信息处理。

表 1 ADS-B 数据集
Table 1 ADS-B datasets

数据集	属性	民航数量	位置分布数量
Subset-1	经纬度、海拔以及航班号	32 037	1 746
Subset-2	经纬度、海拔以及航班号	106 250	4 013
Subset-3	经纬度、海拔以及航班号	105 872	4 143
Subset-4	经纬度、海拔以及航班号	101 946	4 090

实验的硬件环境为: AMD EPYC 9654 CPU @2.40 GHz 3.70 GHz, 60 GB 内存, NVIDIA 3070 Laptop GPU。实验操作系统为: Microsoft Windows 11。实验采用 PyCharm 开发平台, 实验代码使用 Python 3.8 实现。

为了评估本文所提 PERR 对 ADS-B 位置数据隐私保护处理后的数据可用性, 实验中使用 3 个度量指标, 即平均 L_1 范式距离、KL (kullback-leibler) 散度和平均相对误差 (mean relative error, MRE)。以此衡量原始位置数据频率分布和随机响应后位置数据频率分布的分布差距, 度量指标越小, 分布差距越小, 说明经过 PERR 扰动后所得到的 ADS-B 位置数据可用性越高。

为验证本文方法的可行性与有效性, 本文设置了 3 组验证实验: ①阶段式等差策略的实验; ②定位数据频数统计与频率估算实验; ③消融实验。

3.1 阶段式等差策略的实验

为验证本文提出的阶段式等差策略能够有效提高随机响应效率, 设置隐私预算 ε 预先分配并保持不变。已知 ΔP 主要影响卡尔曼滤波误差, 因为位置域的选择决定了过程噪声, 而 $\Delta \varepsilon$ 同时影

响随机响应误差和卡尔曼滤波误差。改变 PERR 中的 ΔP 和 $\Delta \varepsilon$ 两个等差参数, 观察并分析 ADS-B 原始位置数据和随机响应 PERR 处理后的 ADS-B 位置数据之间的差异, 以 subset-1 数据集为例, 在隐私预算 ε 预先分配并保持不变的情况下, 设置全局敏感度: $0 \leq \Delta f \leq 10^4$ 。

由表 2 可知, 随着 ΔP 和 $\Delta \varepsilon$ 单调增加, L'_n 、平均 L_1 范式距离、平均 L_2 范式距离和 MRE 呈递减趋势。因为随着 ΔP 和 $\Delta \varepsilon$ 的逐渐增大, $P_{14} \sim P_{17}$ 和 $\varepsilon_{14} \sim \varepsilon_{17}$ 之间的差距随之增大, 即 ADS-B 原始位置数据在基于等差概率的自适应随机响应空间划分阶段, 被映射到 G_{17} 随机响应空间的概率 P_{17} 越大, 在基于等差隐私的随机响应阶段分配的隐私预算 ε_{17} 越大, PERR 处理后的 ADS-B 位置数据与原始位置数据之间的差距越小, 说明等差扰动方案能有效提高随机响应后数据的可用性。

由表 3 可知, 当 $\Delta \varepsilon = 0.1$ 时, 随着 ΔP 单调增加, L'_n 、平均 L_1 范式距离、平均 L_2 范式距离和 MRE 呈单调递减趋势。由表 4 可知, 当 $\Delta P = 0.1$ 时, 随着 $\Delta \varepsilon$ 单调增加, L'_n 、平均 L_1 范式距离、平均 L_2 范式距离和 MRE 呈单调递减趋势。这表明等差扰动方案中的参数 ΔP 与 $\Delta \varepsilon$ 能同时有效提高随机响应过程中数据的可用性, 减少数据处理过程中的效用损失。

在 PERR 中的 ΔP 和 $\Delta \varepsilon$ 确定的情况下, 随着整体隐私预算 ε 的增加, 各评价指标逐渐减小。原因是: 在等差扰动方案预先确定的情况下, PERR 主要受算法的整体隐私预算 ε 调控, 这表明可以通过隐私预算 ε 调整 PERR 的隐私保护程度与数据可用性之间的平衡。

3.2 定位数据频数统计与频率估算实验

为了验证本文提出的 PERR 在提高 ADS-B 定位数据频数统计与频率估算精确度和准确性方面的有效性, 首先将 ADS-B 原始位置数据频率分布和 PERR 处理后的 ADS-B 原始位置数据频率分布进行直观对比, 然后分别选取 OUE、OLH、KRR、HR、PLDP、KRR-IIW (k -ary randomized response based on improved iterative wiener) 方法^[16]与 PERR 方法进行 ADS-B 定位数据隐私保护对比实验。实验结果表明 PERR 在 ADS-B 定位数据隐私保护中有显著优势。

表 2 等差扰动对 Subset-1 作用后的评估结果
Table 2 Evaluation of Subset-1 reponse under arithmetic perturbation

隐私预算	ΔP	$\Delta \varepsilon$	L'_n	平均 L_1 范式距离	平均 L_2 范式距离	MRE
$P=1$ $\varepsilon=1$	0.050 0	0.100 0	0.000 07	1.714 8	1.699 6	0.002 4
	0.100 0	0.200 0	0.000 08	1.481 4	1.466 7	0.002 2
	0.150 0	0.300 0	0.000 08	1.215 3	1.201 3	0.002 2
$P=1$ $\varepsilon=2$	0.050 0	0.100 0	0.000 02	1.382 0	1.367 2	0.002 4
	0.100 0	0.300 0	0.000 03	1.315 9	1.301 4	0.002 4
	0.150 0	0.400 0	0.000 03	1.103 0	1.089 0	0.002 1
$P=1$ $\varepsilon=3$	0.050 0	0.500 0	0.000 01	1.350 7	1.335 9	0.002 4
	0.100 0	0.700 0	0.000 02	1.252 3	1.237 8	0.002 2
	0.150 0	0.900 0	0.000 02	1.064 1	1.050 1	0.002 2
$P=1$ $\varepsilon=4$	0.050 0	0.500 0	0.000 01	1.014 6	1.000 3	0.002 2
	0.100 0	0.750 0	0.000 01	1.001 0	1.006 7	0.002 2
	0.150 0	0.900 0	0.000 01	0.865 1	0.851 2	0.002 2
$P=1$ $\varepsilon=5$	0.050 0	0.500 0	0.000 00	0.778 1	0.764 1	0.002 2
	0.100 0	0.750 0	0.000 00	0.775 1	0.761 0	0.002 1
	0.150 0	1.000 0	0.000 00	0.765 5	0.751 6	0.002 0

表 3 基于 Subset-1 的 ΔP 敏感性分析实验结果
Table 3 Experimental results of ΔP sensitivity analysis based on Subset-1

隐私预算	ΔP	$\Delta \varepsilon$	L'_n	平均 L_1 范式距离	平均 L_2 范式距离	MRE
$P=1$ $\varepsilon=1$	0.050 0	0.100 0	0.000 07	1.714 8	1.699 6	0.002 4
	0.100 0		0.000 07	1.570 2	1.554 8	0.002 3
	0.150 0		0.000 07	1.420 6	1.411 2	0.002 3
$P=1$ $\varepsilon=2$	0.050 0	0.100 0	0.000 02	1.382 0	1.367 2	0.002 4
	0.100 0		0.000 02	1.344 6	1.341 3	0.002 4
	0.150 0		0.000 02	1.156 2	1.167 0	0.002 3

表 4 基于 Subset-1 的 $\Delta \varepsilon$ 敏感性分析实验结果
Table 4 Experimental results of $\Delta \varepsilon$ sensitivity analysis based on Subset-1

隐私预算	ΔP	$\Delta \varepsilon$	L'_n	平均 L_1 范式距离	平均 L_2 范式距离	MRE
$P=1$ $\varepsilon=1$	0.100 0	0.100 0	0.000 07	1.570 2	1.554 8	0.002 3
		0.200 0	0.000 08	1.543 2	1.488 6	0.002 2
		0.300 0	0.000 08	1.266 7	1.239 0	0.002 2
$P=1$ $\varepsilon=2$	0.100 0	0.100 0	0.000 02	1.344 6	1.341 3	0.002 4
		0.300 0	0.000 03	1.335 9	1.320 4	0.002 4
		0.400 0	0.000 03	1.123 0	1.099 0	0.002 2

由表 2 可知，隐私预算 ε 不变， ΔP 和 $\Delta \varepsilon$ 越小，随机响应后数据的可用性越低，因此选用一组代表性参数：隐私预算 ε 设为 1.0，全局敏感度设为 $0 \leq \Delta f \leq 10^4$ ， ΔP 设为 0.05， $\Delta \varepsilon$ 设为 0.1。在此参数设

置下，PERR 处理后的 ADS-B 位置数据与原始位置数据之间的差距最大，观察 ADS-B 原始位置数据频率分布和 PERR 处理后的 ADS-B 位置数据频率分布之间的变化，实验结果如图 7 所示。

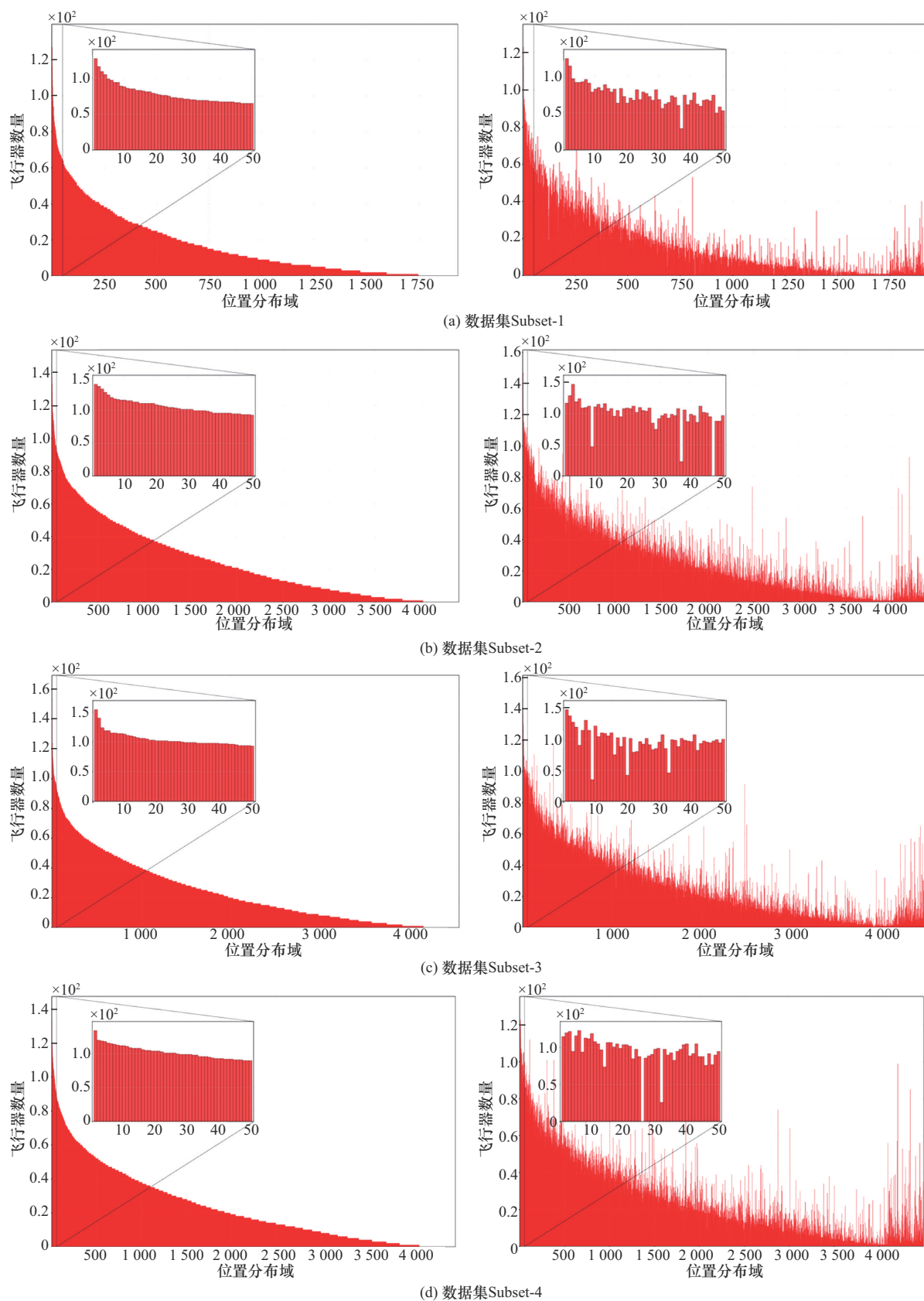


图7 不同ADS-B报文数据集中位置数据的频率分布对比

Figure 7 Comparison of frequency distributions of location data across different ADS-B message datasets

由图7可知,在算法整体隐私预算 ϵ 、等差概率 ΔP 和等差隐私预算 $\Delta\epsilon$ 预先设定的情况下,随着各个数据集中不同位置域的飞行器数量逐渐减少,ADS-B原始位置数据频率分布与PERR处理后的ADS-B位置数据频率分布之间的差异逐渐变大。说明在同一个LDP方案下的同一个位置域中飞行器数量越多,随机响应机制的影响越小、频数统计的准确率越高,LBS统计特定空域单元内的飞行器数量越精确,空中流量统计的准确性越高。

然后,在已有ADS-B报文数据集上,对比各LDP方案的评估指标:平均 L_1 范式距离、KL散度以及均方误差(mean squared error, MSE)。在现有的LDP方案设置中,通常隐私预算 ϵ 取0.5~10,然而,过大的隐私预算 ϵ 值无法有效地保护飞行器的位置隐私。因此,本文隐私预算 ϵ 取1~8。为充分评估各LDP方案对ADS-B位置数据质量的影响,各方案运行30次,取评估指标均值作对比。实验结果如图8~图10所示。

(1) 从评估指标角度进行分析

随着隐私预算 ϵ 逐渐增大, KRR、KRR-

IIW、OUE 以及 PLDP 这4个LDP方案的评估指标呈明显的下降趋势。这是因为随着差分噪声注入量逐渐减少,随机响应机制的影响也相应减小,各LDP方案的评估指标均逐渐下降并趋于稳定,即ADS-B位置数据质量逐渐提高,且当隐私预算 $\epsilon > 5$ 时,各评估指标呈明显的下降;而本文提出的PERR随机响应方案的评估指标始终趋于稳定,并明显低于其他6种LDP方案,说明本文提出的三维定位数据编码、分段式随机响应以及等差扰动方案有效解决了ADS-B报文定位数据处理过程中出现的精度丢失问题,此外,扩展基于阶段式概率加权的卡尔曼滤波器估算方法对随机响应后的数据进行频率估算,可以有效提高飞行器位置分布频率估算的精确度,即在保证数据隐私保护的前提下,PERR处理后的ADS-B位置数据质量高于其他LDP方案。同时也说明PERR可以为空中交通管理系统的实时监控和历史流量分析提供更可靠的数据基础,有助于提高空间密度估计的精度。

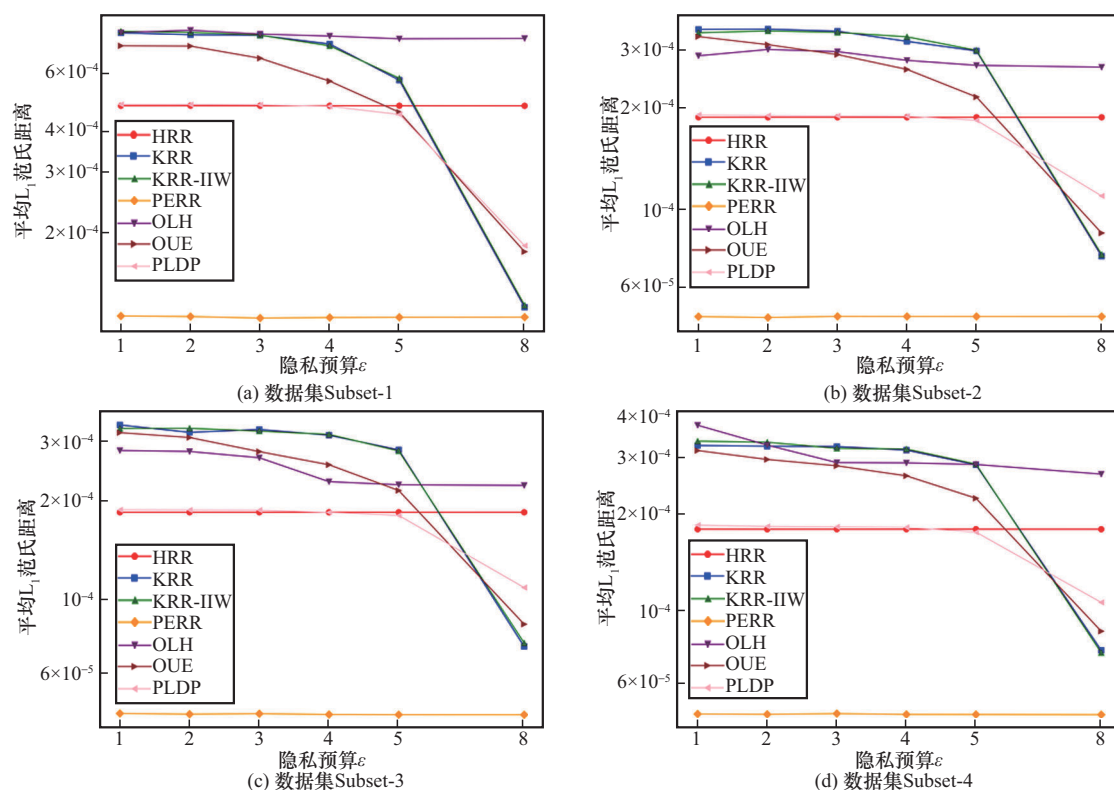


图8 不同的LDP方案在数据集上进行频率分布估计的平均 L_1 范式距离对比

Figure 8 Comparison of average L_1 norm distances for frequency distribution estimation across different LDP schemes on the datasets

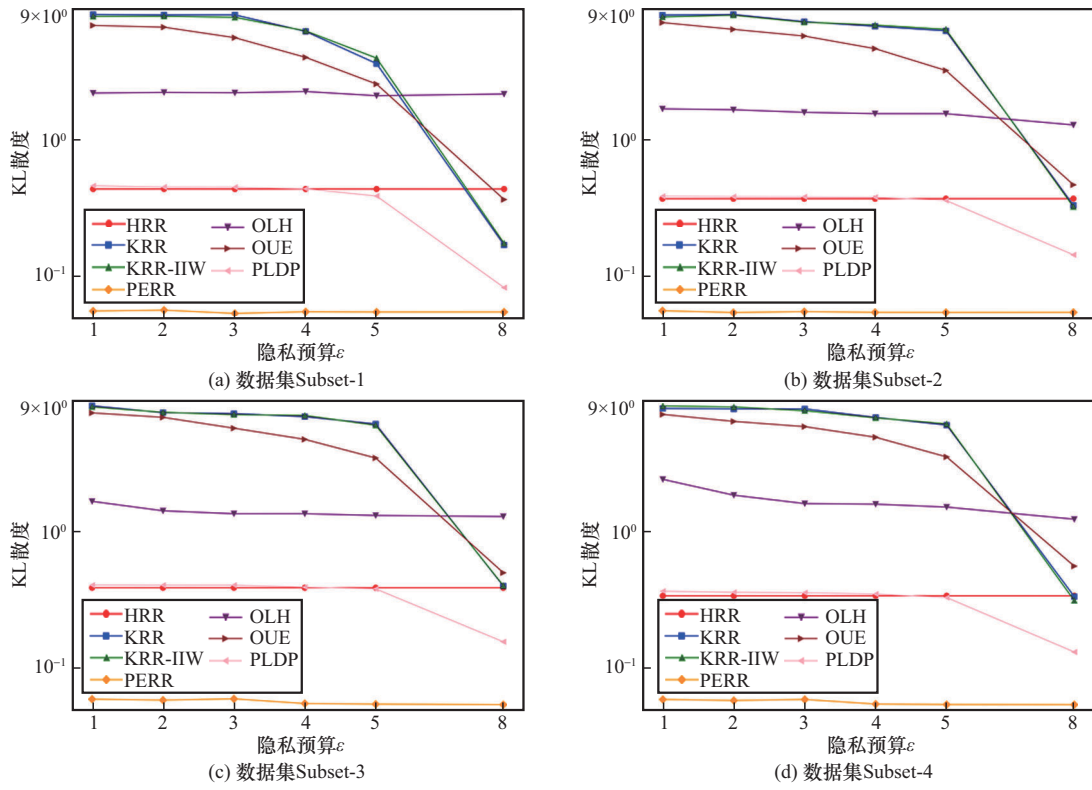


图9 不同的LDP方案在数据集上进行频率分布估计的KL散度对比

Figure 9 Comparison of KL divergence for frequency distribution estimation across different LDP schemes on the datasets

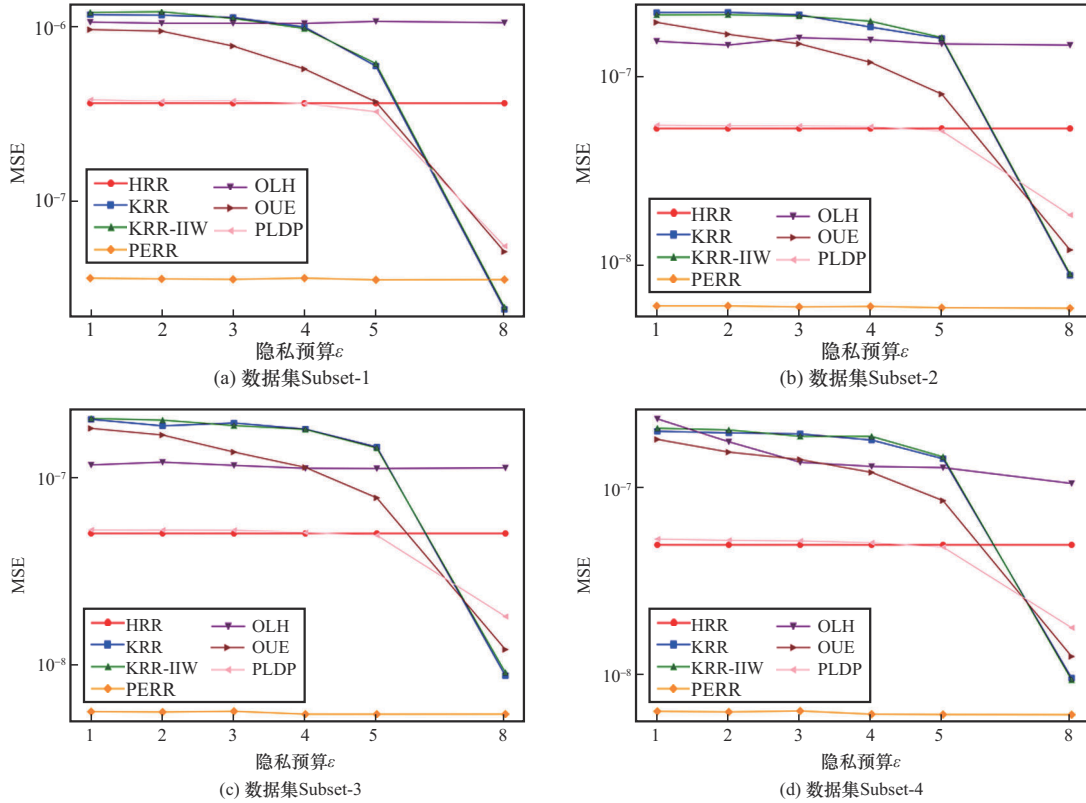


图10 不同的LDP方案在数据集上进行频率分布估计的MSE对比

Figure 10 Comparison of MSE for frequency distribution estimation across different LDP schemes on the datasets

(2) 从评估指标变化幅度进行分析

HR、OLH 以及 PERR 这 3 种 LDP 方案在隐私预算 ε 逐渐增大的情况下, 各评估指标始终趋于稳定, 变化幅度极小, 说明在隐私预算 ε 预先设置较小, 差分噪声注入量较大时, HR、OLH 和 PERR 仍能较好地输出高质量数据, 从而有效保证数据的有效性, 表明 PERR 可以有效支持多维航空定位数据 (经度、纬度、高度) 的隐私保护聚合; 相比之下, KRR、KRR-IIW、OUE 以及 PLDP 这 4 个 LDP 方案的评估指标变化趋势基本一致, 说明这些方法在处理多维属性数据时, 单纯不附加数据编码的随机响应机制以及简单数据编码的随机响应机制普遍存在数据精度丧失的问题, 导致其在输出高质量数据方面表现不足。

(3) 从数据集角度进行分析

当算法整体隐私预算 ε 相同时, 实验数据量越大, 各 LDP 方案的评估指标越小, 因为飞行器位置分布频率估算受单个位置域飞行器数据总量的影响, 数据量越多, 每个位置域的数据分布更加全面、频数统计误差越小, 说明各 LDP 方案处理后的数据可用性越高、飞行器位置分布频率估算越接近原始飞行器位置分布频率。

3.3 消融实验

由于本文方法包含多个关键步骤, 因此采用消融实验验证各个步骤的有效性。实验仅展示在 Subset-1、Subset-2 两个数据集上的实验结果。

实验 1 代表完整的 PERR 方法实验, 实验 2 代表删除阶段式等差随机响应步骤只采用 KRR 直接随机响应的 PERR 方法实验, 实验 3 代表删除基于阶段式概率加权的卡尔曼滤波器频率估算步骤的 PERR 方法实验, 实验 4 代表删除三维定位数据编码的 PERR 方法实验。表 5 和表 6 给出在两个数据集上进行两项消融实验的实验结果, 可以得出以下结论。

1) 实验 2 中, ADS-B 原始位置数据和随机响应处理后的 ADS-B 位置数据之间的平均 L1 范式距离和 MSE 均增大, 并且 KL 散度出现明显的骤增, 表明 ADS-B 原始位置数据频率分布和 KRR 直接随机响应后的 ADS-B 位置数据频率分布之间的差异过大。这说明阶段式等差随机响应能够有效地提高随机响应效率, 减少数据效用损失。

表 5 Subset-1 数据集上的消融实验

Table 5 Ablation experimental results on Subset-1 datasets

方法	平均 L1 范式距离	KL 散度	MSE
实验 1	0.000 113 44	0.056 326 98	0.000 000 03
实验 2	0.000 774 25	8.434 246 01	0.000 001 19
实验 3	1.714 892 62	0.000 567 08	2.56 279 097
实验 4	4.213 425 54	8.962 340 44	5.417 726 14

表 6 Subset-2 数据集上的消融实验

Table 6 Ablation experimental results on Subset-2 datasets

方法	平均 L1 范式距离	KL 散度	MSE
实验 1	0.000 047 77	0.055 932 32	0.000 000 06
实验 2	0.000 333 30	7.915 429 44	0.000 000 21
实验 3	1.341 027 46	0.057 184 31	2.239 876 46
实验 4	3.310 424 04	7.988 065 84	5.210 630 34

2) 实验 3 中, ADS-B 原始位置数据和阶段式等差随机响应处理后的 ADS-B 位置数据之间的 KL 散度增大, 并且平均 L1 范式距离和 MSE 骤增, 表明 ADS-B 原始位置数据与删除了基于阶段式概率加权的卡尔曼滤波器频率估算步骤的 PERR 处理后 ADS-B 位置数据之间的差异过大。这说明基于阶段式概率加权的卡尔曼滤波器频率估算步骤能够有效提高位置分布频率估算的精度, 降低了 ADS-B 位置数据的精度损失, 提高了数据的有效性。

3) 实验 4 中, ADS-B 原始位置数据和随机响应并频率估算后的 ADS-B 位置数据之间的 KL 散度、平均 L1 范式距离、MSE 均骤增, 表明 ADS-B 原始位置数据与删除了三维定位数据编码阶段的 PERR 处理后 ADS-B 位置数据之间的差异过大。这说明三维定位数据编码阶段能够有效地减少随机响应以及后续处理过程中的数据效用损失, 提高了随机响应后定位数据的精度。

4 结束语

针对 ADS-B 报文定位数据隐私易泄露的问题, 本文提出一种面向 LBS 的 ADS-B 定位数据本地化差分隐私方法, 以实现 ADS-B 敏感定位数据处理的隐私保护。该方法采用三维数据编

码方案, 针对 ADS-B 报文数据中的经度、纬度和海拔进行编码, 提出一种高效的阶段式分组随机响应方案, 有效降低了随机响应过程中定位数据的精度损失, 通过扩展基于阶段式概率加权的卡尔曼滤波器估算方法, 有效提高了对随机响应后的定位数据进行频率估算和频数统计的精度。与现有 LDP 方案相比, 本文方法在隐私保护和数据可用性之间找到了良好的平衡, 具有较好的可用性。

未来将从以下两个方面展开研究: ①在保证降低数据精度损失的同时, 确保 ADS-B 多维数据信息(速度、航向)的隐私保护; ②根据各种飞行器的平均速度和特性, 自适应划分随机响应空间。

参考文献:

- [1] AHMED H, KHAN H, KHAN M A. A survey on security and privacy of automatic dependent surveillance-broadcast (ADS-B) protocol: Challenges, potential solutions and future directions[J]. *Author Preprints*, 2023.
- [2] 丁建立, 邹云开, 王静, 等. 基于深度学习的 ADS-B 异常数据检测模型[J]. *航空学报*, 2019, 40(12): 323220.
DING J L, ZOU Y K, WANG J, et al. ADS-B anomaly data detection model based on deep learning[J]. *ACTA AERONAUTICA ET ASTRONAUTICA SINICA*, 2019, 40(12): 323220-323220.
- [3] 王文益, 古亭亭. 基于 1DCNN-BiLSTM 网络的 ADS-B 欺骗式干扰检测[J]. *信号处理*, 2021, 37(6): 984-990.
WANG W Y, GU T T. ADS-B deceptive jamming detection based on 1DCNN-BiLSTM network[J]. *Journal of Signal Processing*, 2021, 37(6): 984-990.
- [4] KACEM T, KAYA A, KECELI A S, et al. ADS-B attack classification using machine learning techniques[C]//2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops). Piscataway: IEEE Press, 2021: 7-12.
- [5] 中央电视台新闻频道. 警惕! 航空爱好者沦为“窃密志愿者”[EB]. 2024. CCTV News. Watch out! Aviation enthusiasts are becoming “volunteers for espionage” [EB]. 2024.
- [6] 李洪涛, 任晓宇, 王洁, 等. 基于差分隐私的连续位置隐私保护机制[J]. *通信学报*, 2021, 42(8): 164-175.
LI H T, REN X Y, WANG J, et al. Continuous location privacy protection mechanism based on differential privacy[J]. *Journal on Communications*, 2021, 42(8): 164-175.
- [7] 陈思, 付安民, 苏锐, 等. 基于差分隐私的轨迹隐私保护方案[J]. *通信学报*, 2021, 42(9): 54-64.
CHEN S, FU A M, SU M, et al. Trajectory privacy protection scheme based on differential privacy[J]. *Journal on Communications*, 2021, 42(9): 54-64.
- [8] 孙新越, 张伟哲, 何慧, 等. 轨迹数据的差分隐私保护方法综述[J]. *网络与信息安全学报*, 2025, 11(3): 1-18.
- [9] SUN X Y, ZHANG W Z, HE H, et al. Survey of differentially private methods for trajectory data[J]. *Chinese Journal of Network and Information Security*, 2025, 11(3): 1-18.
- [9] ARCOLEZI H H, COUCHOT J F, AL BOUNA B, et al. Improving the utility of locally differentially private protocols for longitudinal and multidimensional frequency estimates[J]. *Digital Communications and Networks*, 2024, 10(2): 369-379.
- [10] WANG H, KONG Y, HONG Y, et al. Publishing video data with indistinguishable objects[C]//Advances in Database Technology: Proceedings. International Conference on Extending Database Technology. 2020: 323.
- [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. 2014: 1054-1067.
- [12] KAIROUZ P, BONAWITZ K, RAMAGE D. Discrete distribution estimation under local privacy[C]//International Conference on Machine Learning. PMLR, 2016: 2436-2444.
- [13] WANG T, BLOCKI J, LI N, et al. Locally differentially private protocols for frequency estimation[C]//26th USENIX Security Symposium (USENIX Security 17). 2017: 729-745.
- [14] ACHARYA J, SUN Z, ZHANG H. Hadamard response: estimating distributions privately, efficiently, and with little communication[C]//The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, 2019: 1120-1129.
- [15] CHEN R, LI H, QIN A K, et al. Private spatial data aggregation in the local setting[C]//2016 Piscataway: IEEE Press 32nd International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2016: 289-300.
- [16] FANG H, CHEN L, LIU Y, et al. Locally differentially private frequency estimation based on convolution framework[C]//2023 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2023: 2208-2222.
- [17] KIM J W, JANG B. Workload-aware indoor positioning data collection via local differential privacy[J]. *IEEE Communications Letters*, 2019, 23(8): 1352-1356.
- [18] ZHAO X, LI Y, YUAN Y, et al. LDPart: effective location-record data publication via local differential privacy[J]. *IEEE Access*, 2019, 7: 31435-31445.
- [19] WANG H, HONG H, XIONG L, et al. L-srr: local differential privacy for location-based services with staircase randomized response[C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer And Communications Security. 2022: 2809-2823.
- [20] WANG T, ZHANG X, FENG J, et al. A comprehensive survey on local differential privacy toward data statistics and analysis[J]. *Sensors*, 2020, 20(24): 7030.
- [21] LI L, LU R, HUANG C. EPLQ: efficient privacy-preserving location-based query over outsourced encrypted data[J]. *IEEE Internet of Things Journal*, 2015, 3(2): 206-218.
- [22] ASARI A, ALAGHEBAND M R, BAYAT M, et al. A new provable hierarchical anonymous certificateless authentication protocol with aggregate verification in ADS-B systems[J]. *Computer Networks*,

2021, 185: 107599.

[23] 中华人民共和国自然资源部. 地球空间网格编码规则: GB/T 40087-2021[S]. 北京: 中国国家标准化管理委员会, 2021.

Ministry of Natural Resources of the People's Republic of China. Geospatial grid code rules: GB/T 40087-2021[S]. Beijing: Standardization Administration of China, 2021.

[作者简介]



杨宏宇 (1969–), 男, 吉林长春人, 博士, 中国民航大学教授、博士生导师, 主要研究方向为网络与系统安全、软件安全检测、网络安全态势感知。



张恒鸣 (1999–), 男, 河南济源人, 中国民航大学硕士生, 主要研究方向为网络信息安全。



朱友文 (1982–), 男, 安徽阜南人, 博士, 南京航空航天大学教授、博士生导师, 主要研究方向为区块链、应用密码学、低空安全、隐私计算。



胡慧玲 (1971–), 女, 四川人, 民航中南地区空中交通管理局高级工程师, 主要研究方向为空中交通管制、低空运行管理。



胡俊成 (1995–), 男, 博士, 吉林大学讲师, 主要研究方向为计算机体系结构、系统安全。