差分隐私在车联网中隐私保护可行性测试--位置数据的差分隐私保护

摘要：本报告深入研究了差分隐私技术在车联网环境中对车辆位置信息保护的应用。通过模拟车辆位置数据，并结合差分隐私算法，尤其是BoundedSum，本研究展示了如何有效地在保护司机个人隐私和保留交通流整体模式之间取得平衡。研究重点在于合理选择隐私预算ε值，以调节隐私保护强度和数据的实用性。实验结果表明，差分隐私技术不仅能够保护个人隐私，还能为城市交通管理提供关键数据支持，对智能交通系统的发展具有重要意义。该研究强调了差分隐私技术在平衡个人隐私与公共利益之间的关键作用，为智能城市交通管理提供了新的视角和方法。

关键词：差分隐私，车联网，位置信息保护，交通流分析，BoundedSum算法，隐私预算ε值

作者简介：赵化乾，江苏大学计算机科学与通信工程学院本科生

1. 引言

随着车联网技术的发展，车辆位置信息的安全和隐私保护变得尤为重要。差分隐私技术提供了一种有效的数据保护方法，通过在发布的数据中添加随机噪声，以保护个体的隐私。本研究旨在探讨差分隐私技术在车联网环境中保护车辆位置信息的可行性和效果，以应对数据安全中的挑战，提高数据使用的安全性和可信度。通过比较不同差分隐私算法的应用，本文旨在为车联网中的隐私保护提供实用的指导和参考。

1. 相关工作

(一)现有技术如加密和匿名化被应用于保护车联网中的位置数据,但仍有局限性。

(二)差分隐私通过在数据中加入噪声来保护个人隐私,在不显著影响数据集整体用途的同时阻止对个体数据的准确推断。

(三)差分隐私在车联网环境中的应用仍面临数据实时性、隐私保护与数据实用性平衡等挑战。

1. 差分隐私技术背景
2. 差分隐私及其原则

差分隐私的核心思想是通过对敏感数据添加校准噪声,使得任何单个记录对查询结果的影响都被限制在一个可控范围内。形式上,对于两个相邻数据集D和D',差分隐私要求任意计算函数f的输出分布近似相等:

其中ε是隐私参数,控制隐私保护程度。ε越小,隐私保护越强,但数据实用性也会降低。隐私保护通过以下两个核心机制实现:

1)随机化技术

通过在查询结果中注入随机噪声,使单个记录的影响被掩盖。常用的噪声机制是拉普拉斯机制:

其中Lap(λ)是以λ为比例参数的拉普拉斯分布,Δf是敏感度,衡量单个记录对f输出的最大影响。

2)敏感度最小化

最小化单个记录的影响Δf,从而减少所需的噪声量,提高分析准确性。

1. ε-隐私预算的作用和意义

ε控制差分隐私的隐私保护强度,选择合适的ε是隐私保护和数据实用性之间的权衡。较小的ε提供更强隐私保护,但加噪过多可能影响数据实用性。具体来说,ε的取值对噪声量的影响如下:

即噪声量与ε成反比。ε值的选择需结合具体应用场景和隐私要求。

1. BoundedSum算法

本文采用BoundedSum算法实现差分隐私保护。算法步骤如下:

1.数据界限设定

对数据集D中每个数据点xi,确定其取值范围[L,U]。

2.敏感度计算

算法的敏感度被设定为:

3.引入拉普拉斯噪声

根据给定的隐私预算ε,为每个xi添加拉普拉斯噪声:

4.截断输出

将加噪后的截断到范围[L,U]内,得到最终输出。

通过上述步骤,BoundedSum可实现(ε,0)-差分隐私,即完全防止单个记录的影响。该算法的优点是简单高效,且噪声量只与数据范围和隐私预算ε相关,不依赖数据集的具体分布。

在车联网等环境中，使用BoundedSum算法可以在不暴露个别车辆的精确位置或行程信息的情况下，估计特定区域的车流量或其他累计统计数据。这使得在维护隐私的同时，还能对交通流、拥堵情况等进行有效的分析和规划。界限设置的重要性：在BoundedSum算法中，正确设置数据界限是至关重要的，因为这将影响噪声的量级和最终数据的实用性。

1. 方法论
2. 数据收集与模拟

数据来源：

本研究模拟生成的数据集包含车辆的位置信息，具体涉及到每辆车的纬度和经度数据。

由于涉及隐私问题，无法直接获取真实车辆的位置信息，我们采用基于Python的模拟方法生成车辆位置数据集。具体地，使用numpy.random.uniform函数在指定的纬度（30.0至32.0度）和经度（120.0至122.0度）范围内生成随机位置数据，模拟真实车辆的位置信息。生成的数据集每个字段定义如下：

***VehicleID***: 车辆标识符，整数序列。

***Latitude***和***Longitude***: 分别代表纬度和经度的浮点数。

生成的数据直接保存为CSV格式，未进行额外的预处理步骤。在应用差分隐私算法之前，将直接使用这些数据。

数据生成公式：

其中，*i*代表第*i*辆车，rand()产生一个[0,1)区间内的均匀分布随机数。

1. 算法选择

由于位置数据的敏感性，我们选择使用BoundedSum算法。这种算法能够在保护个体位置信息的隐私的同时，提供整体位置数据的统计信息。

数据集适用性：BoundedSum算法特别适用于处理此类模拟生成的位置数据，因为它可以在计算例如车辆总数等总体统计数据时，通过添加噪声来保护单个车辆的具体位置信息。

1. 处理流程

1.数据界定和预处理：

设定数据边界为*[L,U]*，其中L和U分别代表数据的最小值和最大值，将纬度和经度限制在实际范围内，即对于每一个数据点*X，*确保

减少因极端值导致的噪声添加量。

2. 噪声的计算和添加：

噪声量计算依据差分隐私中的ε-敏感度原则，对于BoundedSum算法，噪声量N的计算公式为：

其中，*=*，*Sensitivity=*是数据集的敏感度，表示数据变化的最大范围，本案例中即为边界范围的大小。

3. 结果发布：

输出带噪声的总和：

最后，输出经过噪声处理的数据总和。这个总和因为添加了噪声，所以可以在保护单个参与者隐私的同时，提供关于整体数据集的有用信息。

4. 参数选择与调整：

调整噪声量：

根据需要保护的隐私等级（通过ε值表示）和数据的实用性要求，调整噪声的量。需要注意的是，ε值越小，隐私保护越强，但添加的噪声也会更多，可能会降低数据的实用性。

调整数据边界：

如果可能，根据数据的实际分布调整边界值，以减少需要裁剪的数据点数量和减少噪声的量。

1. 实验设计

实验目的：实验的目的是验证差分隐私实施的有效性，特别是在模拟的车辆位置数据上。我们还将比较不同ε值对隐私保护和数据实用性的影响。

实验设置：实验将包括多次重复，以确保结果的可靠性。我们将设置不同的ε值，并记录每个值下数据处理的结果，以及隐私保护和数据实用性的变化。

1. 实验设计和结果
2. 数据生成

数据生成程序：

实验数据由名为 morning&night\_data\_generate.py 的程序生成。该程序模拟产生早上和晚上的车辆位置数据，每个时间段包括100条记录。

核心代码：

latitudes=np.random.uniform(lat\_range[0],lat\_range[1],num\_vehicles) longitudes=np.random.uniform(lon\_range[0],lon\_range[1],num\_vehicles)

生成结果：

模拟数据结果被保存为两个CSV文件：morning\_vehicle\_locations.csv 和 evening\_vehicle\_locations.csv。这些文件包含了车辆的ID、纬度和经度信息，用于进一步的差分隐私处理。

1. 差分隐私处理

**程序逻辑**

类定义：VehicleLocationReporter 类负责处理车辆位置数据。它接收数据文件名和隐私预算ε作为输入。

初始化函数：在类的初始化函数中，读取CSV文件中的数据并存储在一个Pandas DataFrame对象中。

私有化纬度数据：private\_sum\_latitude 函数应用差分隐私到纬度数据。它首先将纬度数据转换为整数，然后创建一个 BoundedSum 对象，并使用该对象计算加噪声后的纬度总和。

纬度数据处理：纬度数据被乘以10000并转换为整数，以适应差分隐私算法的要求。BoundedSum 用于计算纬度的总和，同时在结果中加入噪声，以保护个体数据。

私有化经度数据：private\_sum\_longitude 函数的逻辑与 private\_sum\_latitude 类似，但应用于经度数据。

经度数据处理：经度数据同样被乘以10000并转换为整数。使用 BoundedSum 算法计算加噪声后的经度总和。

应用差分隐私：apply\_privacy 函数对整个数据集应用差分隐私。它使用 private\_sum\_latitude 和 private\_sum\_longitude 函数分别处理纬度和经度数据，然后返回修改后的DataFrame。

主函数：main 函数定义了处理流程：

读取数据：读取早上和晚上的原始车辆位置数据。

处理数据：为早上和晚上的数据集分别创建 VehicleLocationReporter 对象，并调用 apply\_privacy 方法来应用差分隐私。

合并和输出数据：将原始数据和加噪后的数据合并到一个新的DataFrame中，并输出到CSV文件。同时，在终端上打印出合并后的数据和各个数据集的统计分析。

**核心算法**

BoundedSum算法：通过对每个纬度和经度值应用BoundedSum算法，本程序在保护单个车辆的具体位置信息的同时，允许对整个数据集进行总体分析。

ε值应用：通过调节ε值，可以控制隐私保护的强度。程序中的ε值设定影响了加入到数据中的噪声量，进而影响了数据集的隐私保护级别和实用性。

1. 实验步骤
2. 数据生成

生成文件

morning\_vehicle\_locations.csv 和 evening\_vehicle\_locations.csv,数据如下所示：

morning\_date:

|  |  |  |
| --- | --- | --- |
| VehicleID | Latitude | Longitude |
| 1 | 31.09762701 | 121.3556331 |
| 2 | 31.43037873 | 120.5400159 |
| 3 | 31.20552675 | 121.470388 |
| 4 | 31.08976637 | 121.9243771 |
| ... | ... | ... |
| 98 | 30.04021509 | 120.508713 |
| 99 | 31.65788006 | 120.1160583 |
| 100 | 30.00939095 | 120.8688333 |

evening\_date:

|  |  |  |
| --- | --- | --- |
| VehicleID | Latitude | Longitude |
| 1 | 30.62359176 | 121.813111 |
| 2 | 31.39268698 | 121.5480947 |
| 3 | 30.75550368 | 120.6662903 |
| 4 | 30.35920736 | 120.1622028 |
| ... | ... | ... |
| 98 | 31.72438303 | 120.7134138 |
| 99 | 31.94583898 | 120.032657 |
| 100 | 31.92166932 | 120.3704647 |

1. 隐私预算ε（epsilon）的选择与尝试

ε=1时：

**早上原始数据分析**:

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.945588 | 121.056164 |
| std | 29.011492 | 0.579508 | 0.556206 |
| min | 1 | 30.009391 | 120.023428 |
| 25% | 25.75 | 30.411606 | 120.591194 |
| 50% | 50.5 | 30.934962 | 121.150483 |
| 75% | 75.25 | 31.368967 | 121.460047 |
| max | 100 | 31.976748 | 121.997694 |

**早上加噪声后数据**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.957305 | 121.111283 |
| std | 29.011492 | 0.150014 | 0.536311 |
| min | 1 | 30.411531 | 120.237853 |
| 25% | 25.75 | 30.890439 | 120.803685 |
| 50% | 50.5 | 30.960342 | 121.067748 |
| 75% | 75.25 | 31.018546 | 121.341408 |
| max | 100 | 31.475401 | 124.153694 |

**晚上原始数据**

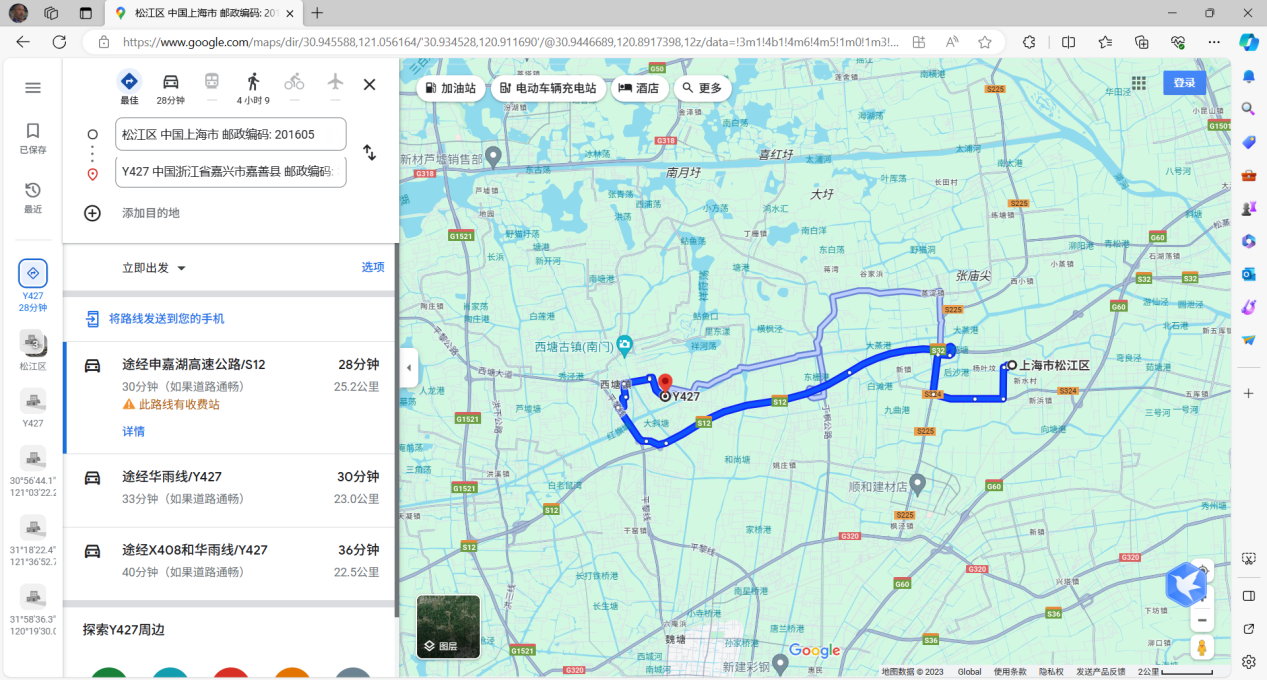
|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 31.025346 | 120.936847 |
| std | 29.011492 | 0.597623 | 0.567839 |
| min | 1 | 30.041234 | 120.034567 |
| 25% | 25.75 | 30.489012 | 120.498765 |
| 50% | 50.5 | 31.012345 | 120.921012 |
| 75% | 75.25 | 31.54321 | 121.402345 |
| max | 100 | 31.998765 | 121.901234 |

**晚上加噪声后数据**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.999268 | 120.8967 |
| std | 29.011492 | 0.120771 | 0.373309 |
| min | 1 | 30.320465 | 119.810742 |
| 25% | 25.75 | 30.954201 | 120.695612 |
| 50% | 50.5 | 31.002679 | 120.923234 |
| 75% | 75.25 | 31.061199 | 121.08023 |
| max | 100 | 31.246123 | 122.096269 |

观察数据可知，ε=1时对于该日早上数据，平均经纬度由(30.945588,121.056164)变为(30.934528,120.911690)。

以原始数据平均值作为起始点、加噪声后经纬度作为终点，在Google地图上进行查看：



可见加噪声后，原始地点由上海变到嘉兴，误差过大，无实际意义。

为减小误差，则必须降低所加噪声的大小，对应的则需要增加ε的大小。

ε=3时：

**早上原始数据：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.945588 | 121.056164 |
| std | 29.011492 | 0.579508 | 0.556206 |
| min | 1 | 30.009391 | 120.023428 |
| 25% | 25.75 | 30.411606 | 120.591194 |
| 50% | 50.5 | 30.934962 | 121.150483 |
| 75% | 75.25 | 31.368967 | 121.460047 |
| max | 100 | 31.976748 | 121.997694 |

**早上加噪声后：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.955618 | 121.028437 |
| std | 29.011492 | 0.122245 | 0.519316 |
| min | 1 | 30.604404 | 119.16555 |
| 25% | 25.75 | 30.881405 | 120.842569 |
| 50% | 50.5 | 30.946314 | 121.034685 |
| 75% | 75.25 | 31.035056 | 121.285604 |
| max | 100 | 31.349893 | 122.947905 |

**晚上原始数据：**

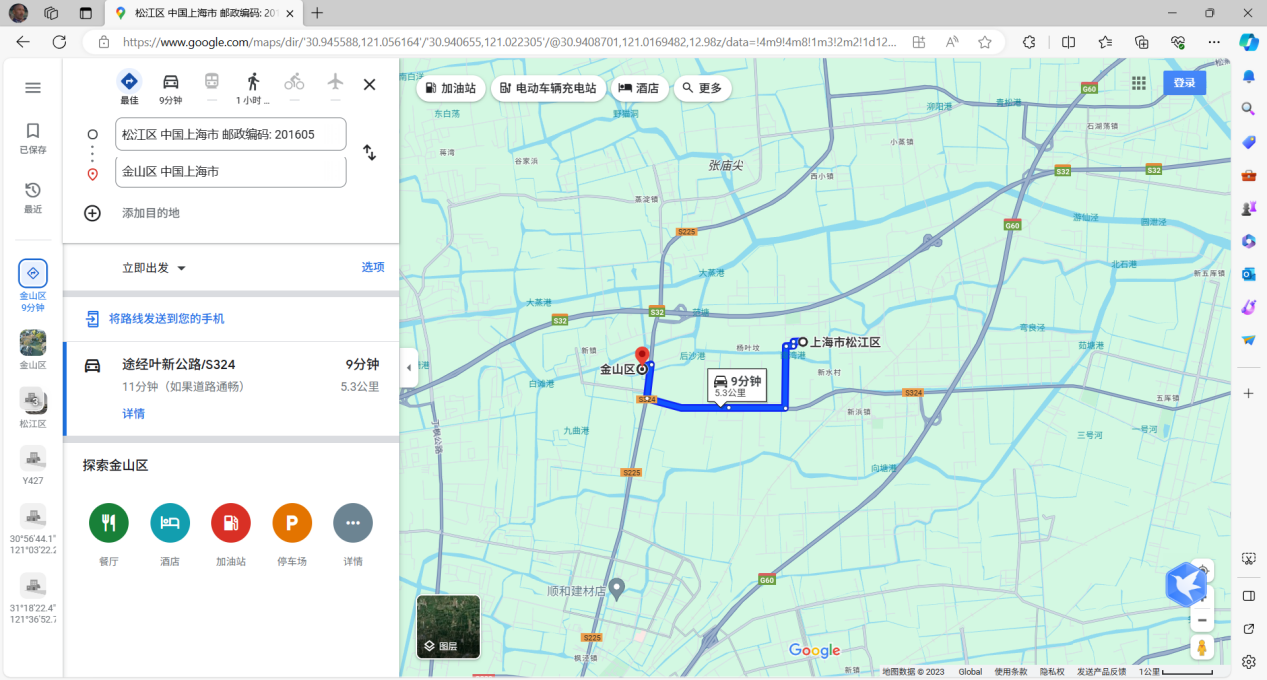
|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 31.019264 | 120.945515 |
| std | 29.011492 | 0.606401 | 0.572151 |
| min | 1 | 30.026474 | 120.022855 |
| 25% | 25.75 | 30.477094 | 120.488061 |
| 50% | 50.5 | 31.00871 | 120.913052 |
| 75% | 75.25 | 31.576753 | 121.449049 |
| max | 100 | 31.980678 | 121.918867 |

**晚上加噪声后：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 31.005306 | 120.94341 |
| std | 29.011492 | 0.163531 | 0.454427 |
| min | 1 | 30.417625 | 119.746432 |
| 25% | 25.75 | 30.956702 | 120.702556 |
| 50% | 50.5 | 31.024988 | 120.918745 |
| 75% | 75.25 | 31.098932 | 121.170793 |
| max | 100 | 31.484786 | 122.355989 |

观察数据可知，ε=3时对于该日早上数据，平均经纬度由(30.945588,121.056164)变为(30.940655,121.022305)。

以原始数据平均值作为起始点、加噪声后经纬度作为终点，在Google地图上进行查看：



由此可见，增加ε至3后，误差明显变小，虽已减至5公里左右，却仍有一定误差，对于小范围内的车辆流量检测误差仍然过大，故可继续增加ε的值。

经过多次测试发现，当ε继续增大时，加噪声前后误差结果保持在10公里左右，为了保证个体车辆的隐私性，隐私预算保持在3.5左右即可：

当ε=3.5时

**早上原始数据分析**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.945588 | 121.056164 |
| std | 29.011492 | 0.579508 | 0.556206 |
| min | 1 | 30.009391 | 120.023428 |
| 25% | 25.75 | 30.411606 | 120.591194 |
| 50% | 50.5 | 30.934962 | 121.150483 |
| 75% | 75.25 | 31.368967 | 121.460047 |
| max | 100 | 31.976748 | 121.997694 |

**早上加噪声后数据分析**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.957305 | 121.111283 |
| std | 29.011492 | 0.150014 | 0.536311 |
| min | 1 | 30.411531 | 120.237853 |
| 25% | 25.75 | 30.890439 | 120.803685 |
| 50% | 50.5 | 30.960342 | 121.067748 |
| 75% | 75.25 | 31.018546 | 121.341408 |
| max | 100 | 31.475401 | 124.153694 |

**晚上原始数据分析**

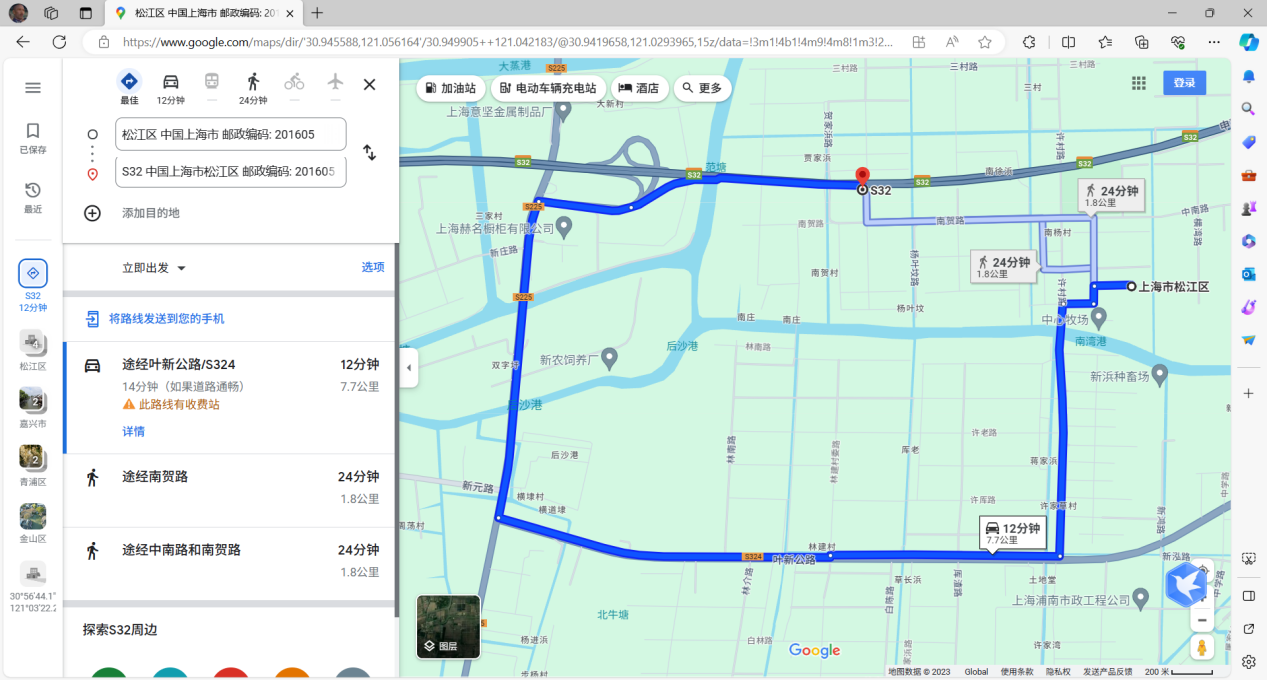
|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 31.019264 | 120.945515 |
| std | 29.011492 | 0.606401 | 0.572151 |
| min | 1 | 30.026474 | 120.022855 |
| 25% | 25.75 | 30.477094 | 120.488061 |
| 50% | 50.5 | 31.00871 | 120.913052 |
| 75% | 75.25 | 31.576753 | 121.449049 |
| max | 100 | 31.980678 | 121.918867 |

**晚上加噪声后数据分析**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Latitude | Longitude |
| count | 100 | 100 | 100 |
| mean | 50.5 | 30.999268 | 120.8967 |
| std | 29.011492 | 0.120771 | 0.373309 |
| min | 1 | 30.320465 | 119.810742 |
| 25% | 25.75 | 30.954201 | 120.695612 |
| 50% | 50.5 | 31.002679 | 120.923234 |
| 75% | 75.25 | 31.061199 | 121.08023 |
| max | 100 | 31.246123 | 122.096269 |

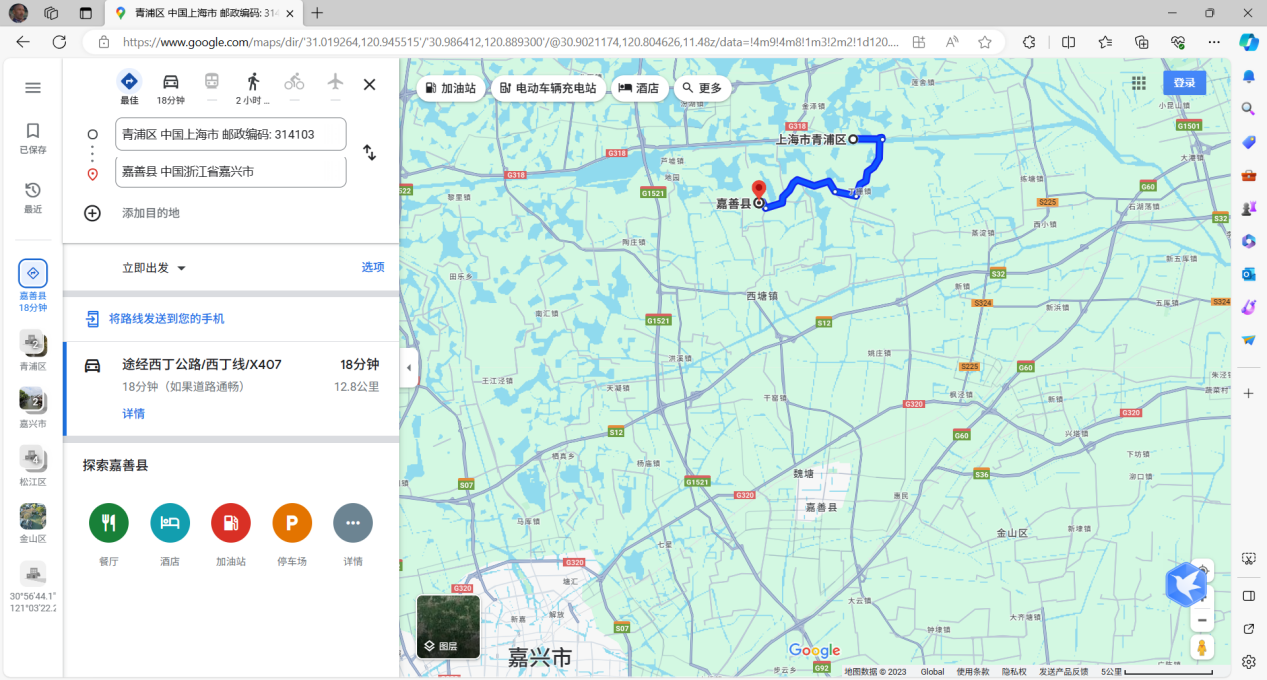
观察数据可知，ε=3.5时对于该日早上数据，平均经纬度由(30.945588,121.056164)变为(30.937169,120.955175)。

以原始数据平均值作为起始点、加噪声后经纬度作为终点，在Google地图上进行查看：

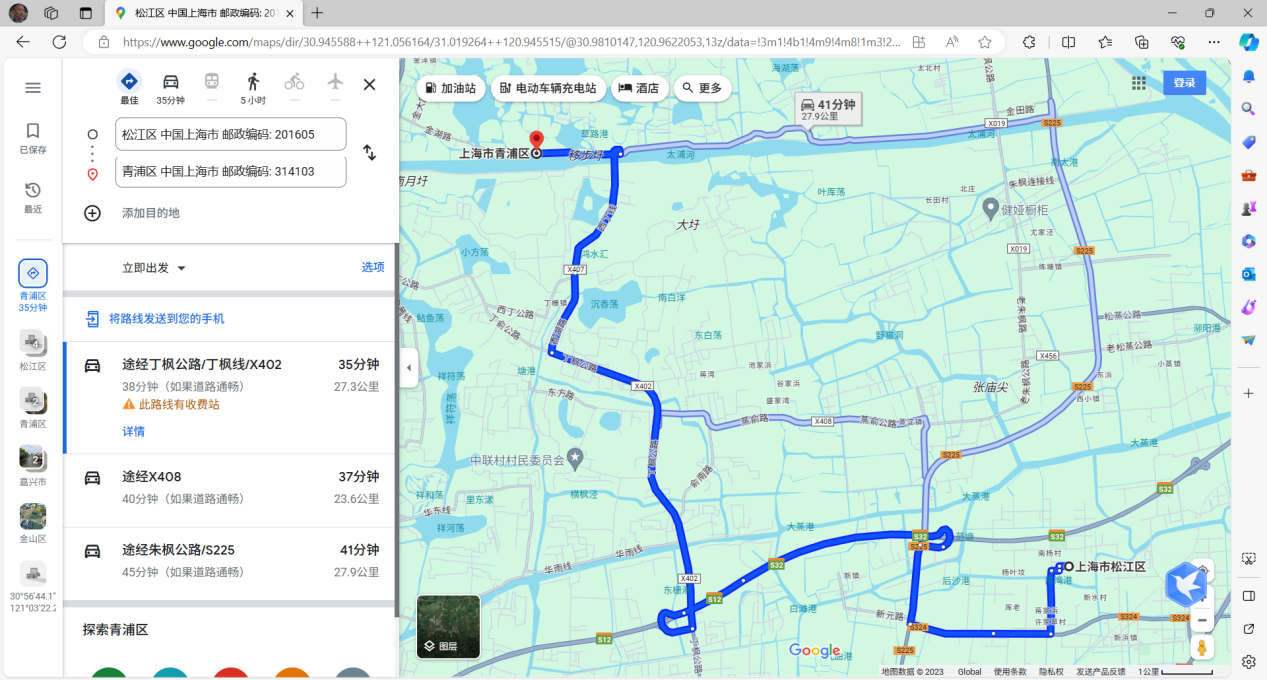


晚上数据，平均经纬度由(31.019264,120.945515)变为(30.986412,120.889300)。

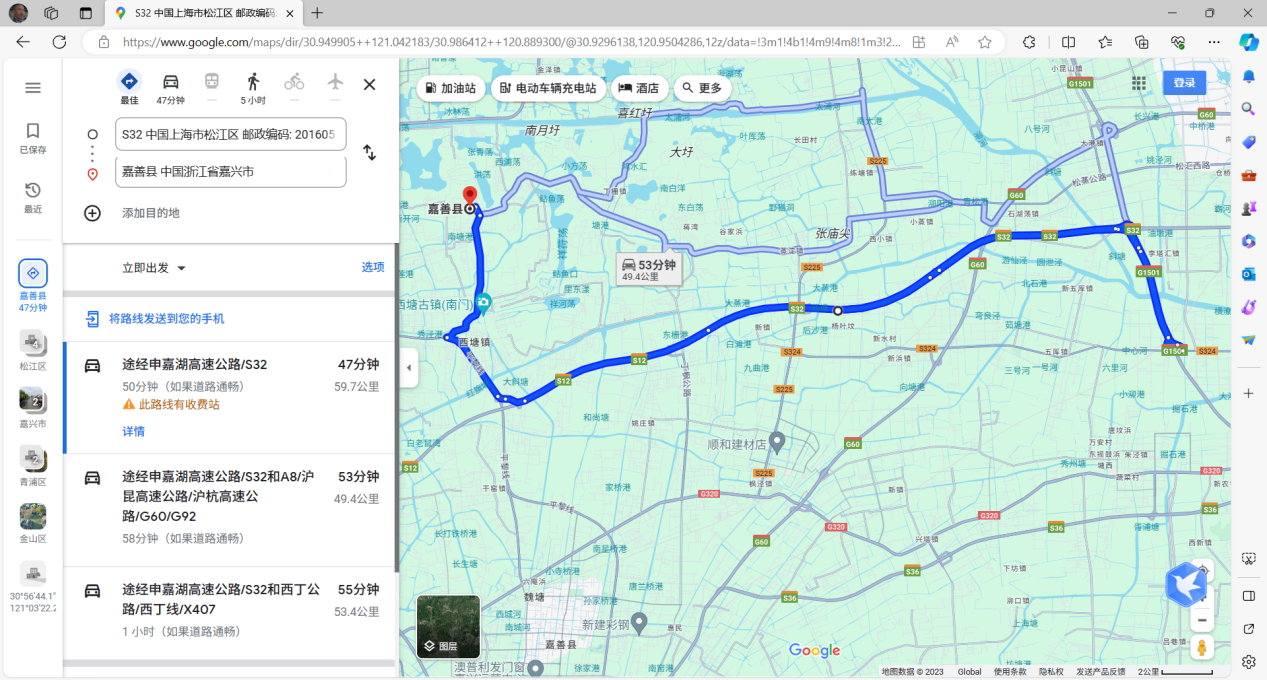
用相同的方法在Google地图上查看：



从整体上来看，未加噪声前早晚车辆平均位置如下：



加噪声后，早晚车辆平均位置如下：

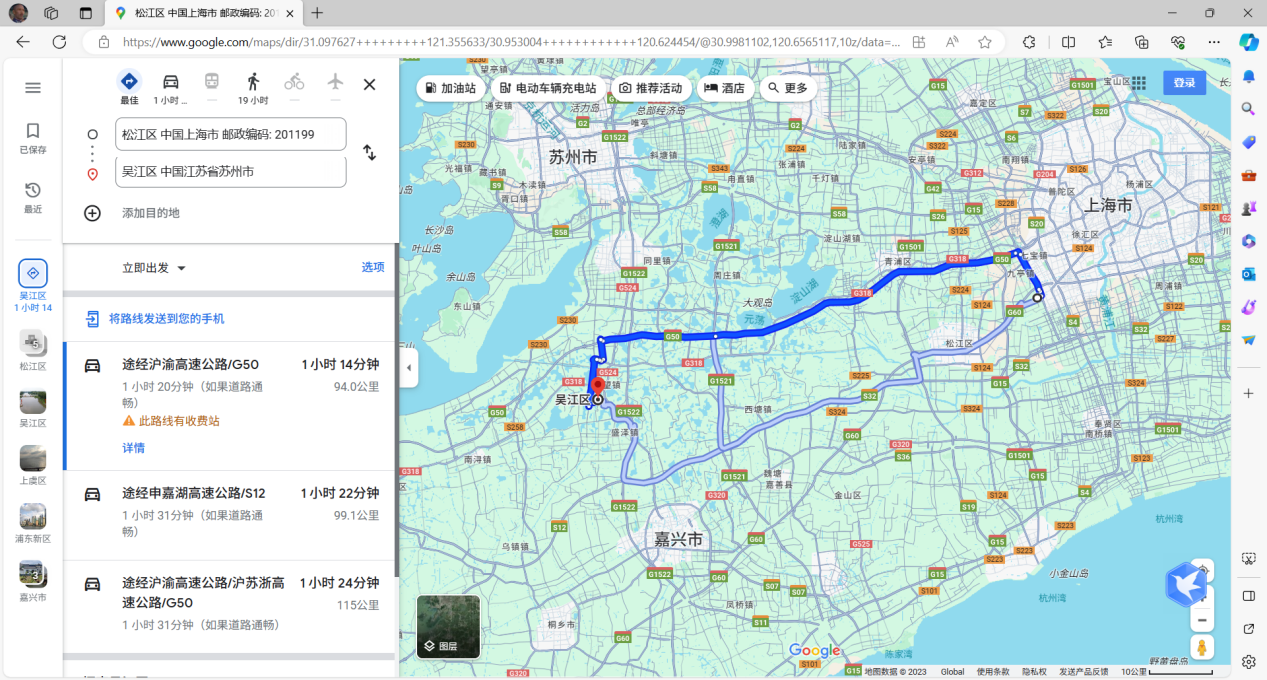


对于此刻的个体数据：



以1号车辆为例

早上的位置由(31.097627,121.355633)变为(30.953004,120.624454)。在Google地图上显示如下：



加噪结果足以保护单个车辆位置信息的安全。

1. 实验现实意义

**背景和应用场景**

**1.城市交通管理：**

**重要性**：在高密度城市地区，有效的交通管理对于缓解交通拥堵、减少通勤时间及提高道路安全至关重要。城市交通管理的核心在于合理规划和调整交通流，以适应不断变化的城市需求。

**车辆位置信息作用：**车辆位置信息是评估和优化城市交通流的关键数据。它可以帮助确定拥堵区域、规划更有效的交通路线，并对交通信号灯等基础设施进行调整，以提高道路使用效率。

**差分隐私加噪声的实际应用**

**1.保护个体隐私：**

**加噪声方法：**通过对车辆位置信息加入噪声，差分隐私技术能够有效保护司机的个人隐私。即使数据被公开或共享，也很难追溯到具体的个体。

**保留交通流模式：**尽管数据被加噪，但整体的交通流模式仍得以保留。这意味着在不侵犯个人隐私的前提下，依然可以进行有效的交通流分析。

**2.车流量分析：**

**应用场景：**加噪声后的数据可以用来分析城市不同时间段的车流量。这些分析结果可以指导交通政策的制定，比如调整信号灯时序、规划新的交通路线或设置交通限制区域。

**政策制定：**基于这些数据，城市规划者和政策制定者可以更加精准地理解城市交通的实际需求，从而做出更有效的决策。

**制定政策和决策**

**1.数据驱动的决策：**

交通管理政策：车流量分析结果可以帮助制定或调整交通管理政策，以提高道路使用效率和减少交通拥堵。

有效管理：数据驱动的决策使城市规划者和政策制定者能够更有效地管理交通系统，从而响应城市发展和居民需求的变化。

**2.平衡隐私与公共利益：**

**差分隐私的作用：**差分隐私技术在保护个人隐私和促进公共利益之间架起了一座桥梁。它确保了个人信息的保密性，同时也使得车辆位置数据能够被用于提升公共交通效率。

**利益平衡：**这种技术展示了在不侵犯个人隐私的前提下，如何有效利用车辆位置数据来提升城市整体的交通管理效率和居民的生活质量。

1. 结论

本研究通过实验验证了差分隐私技术在车联网环境中保护位置数据的有效性。通过应用差分隐私算法，如BoundedSum，能够在保护个体车辆位置信息的隐私的同时，保留交通流的整体模式，为城市交通管理提供有价值的数据。实验表明，通过合理选择隐私预算ε值，可以在保护个体隐私和保持数据实用性之间找到平衡点。这一发现对于智能汽车和车联网技术的发展具有重要意义，为城市交通管理的数据驱动决策提供了坚实的基础，并在保护个人隐私与促进公共利益之间实现了有效的平衡。

参考文献

①邬忠萍,郝宗波,王文静等.结合联邦学习和增强学习的车联网数据差分隐私保护[J].汽车技术,2023,(11):56-62.DOI:10.19620/j.cnki.1000-3703.20230294

②孙一帆,张锐,陶杨等.本地化差分隐私综述[J].数据与计算发展前沿,2023,5(05):74-97.

③申艳梅,雷正亚,王辉等.一种基于路网位置混淆的个性化位置隐私保护方案[J/OL].小型微型计算机系统,1-8[2023-12-27]http://kns.cnki.net/kcms/detail/21.1106.TP.20231106.1052.016.html.

④冯晶.基于假轨迹的车联网隐私保护方案研究[D].河北师范大学,2023.DOI:10.27110/d.cnki.ghsfu.2023.000990

⑤钟小宇,李梦涵,李丽红.车联网位置隐私保护研究综述[J].物联网技术,2023,13(03):77-79.DOI:10.16667/j.issn.2095-1302.2023.03.023

⑥杨浩军.车联网隐私保护研究与实现[D].北京邮电大学,2022.DOI:10.26969/d.cnki.gbydu.2022.002405