Software Engineering and Applications 软件工程与应用, 2024, 13(6), 853-863
Published Online December 2024 in Hans. https://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2024.136088

Hans汉斯

满足本地差分隐私的动态域键值数据聚合机制

王凤祥

南京航空航天大学计算机科学与技术学院，江苏 南京

收稿日期：2024年11月1日；录用日期：2024年12月20日；发布日期：2024年12月31日

摘　要

本地差分隐私具有不需要可信第三方、交互少、运行效率高等优点，近年来受到了广泛关注，而键值数据在实际生活里的各个软件中也得到了广泛的应用，键值数据的统计分析也是研究的重点。然而，现有本地差分隐私估计机制未能考虑实际应用中数据域动态变化的情况，将所有数据同等对待，这会浪费大量带宽，且导致估计结果准确度偏低的情况。针对这一问题，提出了键域独立的方法和机制，考虑到数据域动态新增和删减等情况，对不同数据按实时键域情况分批处理。理论分析证实，相对于现有的本地差分隐私机制，所提方案能够对数据域随时变化的键值数据实现基本相同的保护效果，并且在通信成本和通信效率上都有一定的优势。最后，在数据集上评估了新的方案，实验结果证明了所提的机制能够有效应对数据域随时发生变化的情况，相较于朴素地将所有数据一视同仁的情况，能够有效降低估计误差，提升数据效用。

关键词

本地差分隐私，键值数据，隐私保护，频率估计，均值估计

Dynamic Domain Key-Value Data Aggregation Mechanism Satisfying Local Differential Privacy

Fengxiang Wang

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu

Received: Nov. 1st, 2024; accepted: Dec. 20th, 2024; published: Dec. 31st, 2024

文章引用: 王凤祥. 满足本地差分隐私的动态域键值数据聚合机制[J]. 软件工程与应用, 2024, 13(6): 853-863.
DOI: 10.12677/sea.2024.136088

Abstract

Local differential privacy boasts several merits, including the absence of a trusted third party, reduced interaction, and high operational efficiency. It has garnered extensive attention in recent years. Moreover, key-value data has found wide application in real life, and the statistical analysis of key-value data constitutes a key research focus. Nevertheless, existing local differential privacy estimation mechanisms fail to take into account the situation where the data domain undergoes changes at any time in practical applications. They treat all data uniformly, which not only wastes a considerable amount of bandwidth but also leads to relatively low accuracy of estimation results. In response to this issue, a method and mechanism of key-domain independence are put forward, considering scenarios such as real-time additions and deletions of the data domain and processing different data in batches based on the real-time key-domain situation. Theoretical analysis confirms that, in contrast to the existing local differential privacy mechanisms, the proposed scheme can achieve precisely the same protective effect for key-value data with a changing data domain at any time, and has certain advantages in communication cost and communication efficiency. Finally, the new scheme was evaluated on a dataset, and the experimental results show that the proposed mechanism can effectively address the situation where the data domain changes in real time, and can effectively reduce the estimation error compared with the naive approach of treating all data equally, thereby improving data utility.

Keywords

Local Differential Privacy, Key-Value Data, Privacy Protection, Frequency Estimation, Mean Estimation

1. 引言

随着科技的飞速发展和信息技术的普及应用，人们时时刻刻都在不断地产生大量的数据信息。其中，键值数据在日常生活中有着广泛的应用场景，如购买过的物品及其价格、看过的电影及对其的评价分数、用过的手机软件及每日使用时长。通过对这些数据进行收集、记录和分析，可以挖掘出它们中的隐藏信息，对学术、工业、社会服务等多个领域都有重要意义。例如通过分析用户的购买记录，可以得出市场供给需求信息、通过分析用户的观影记录及评分，可以得出值得推荐的电影、通过分析用户手机软件使用情况，可以得出软件的受欢迎程度。但这些数据中往往包含着大量隐私信息，如购物数据会反映个人生活习惯和财务状况、电影评分数据会透露个人喜好偏向，如果直接将这些数据提供给其他人使用，不仅会对个人的人身安全、财产安全造成极大的威胁，也会使得降低用户共享数据的意愿。现如今，已有多个领域关注到了隐私保护问题的重要性，并针对特定问题提出了隐私保护的解决方案。然而，如何在保护用户隐私的前提下对数据进行收集，仍是亟待解决的问题。

目前，差分隐私[1] [2]作为一种效率高开销低的隐私保护技术，是学术界研究的热点方向之一。其中，本地差分隐私[3]于 2013 年被提出，它兼具差分隐私的优点，又摒除了对可信第三方的依赖，具有极高的实用性，在工业场景也得到了广泛的应用。谷歌公司设计了基于本地差分隐私的组件 RAPPOR [4]来采集用户行为数据，苹果公司将其应用到手机上来保护用户的使用数据。

但是，现有的本地差分隐私机制大多未考虑到，实际应用中数据域快速变化的情况，需要在用户端实现与键域无关的扰动，才能增强隐私保护的鲁棒性。以本文主要研究的键值数据的收集估计为例，现有方案主要有 PrivKVM [5]，PCKV-GRR [6]以及 PCKV-UE [6]，这几个方案在用户端扰动环节都比较依赖于键域。实际应用中，数据域的随时变动会给现有方案造成较大的通信和计算负担。例如，PrivKVM 是在整个大键域中进行随机抽样，而键域的变动会导致无效抽样；PCKV-GRR 的随机扰动概率依赖于键域大小，随机扰动的映射对象依赖于键域内容，而键域的变动会导致扰动概率变化，扰动映射的对象不满足本地差分隐私的概念；PCKV-UE 的编码方式依赖于键域，键域的变动会导致额外的通信开销或者错误的信息传输。因此，需要使用一种键域独立的键值数据扰动机制来应对实际应用中键域频繁变化的场景。

本文首先针对键域频繁变化的场景提出使用一种键域独立的本地差分隐私保护机制，该机制是一种键域独立的扰动机制，它在用户侧与键域大小和内容无关，具有较高的可用性。随后，本文将具体分析键域增加或减少情况下，各个方案对于扰动机制的处理与分析，通过对比实验验证本方案的优越性。最后，本文将在服务器端设计个性化的统计机制，使得数据的使用与分析更便捷且具有多样性。

本文的主要贡献包括 3 个方面：

1) 根据目前数据信息飞速变化的情况，定义了一个用户侧数据域独立的概念，以此应对数据域的新增和删减；

2) 基于本地差分隐私下键值数据的收集估计，对比几种机制面临键域新增或删减时，产生的问题和消耗的代价，以及实际发生时对应的策略；

3) 使用键域独立的方案，为服务器端数据收集统计设计出多样性的方案。

2. 理论基础

2.1. 相关工作

键值数据的收集与统计分析在推荐系统等领域有着极高的研究价值，但目前的相关工作都是基于固定的数据域去进行研究的。叶等人首次提出 PrivKV、PrivKVM 和 PrivKVM+，其中 PrivKVM 迭代估计平均值以保证非偏性。PrivKV 可以看作是只有一次迭代的 PrivKVM。高级版本的 PrivKVM+会选择了适当的迭代次数来平衡非偏性和通信成本。孙等人在 PrivKV 的框架下提出了另一种频率和平均值的估计器和几种机制来完成相同的任务[7]。他们还为机器学习[8] [9]中的其他复杂分析任务引入了关键值数据的条件分析。最近，顾等人提出了一种替代的键值扰动协议 PCKV。PCKV 调整了填充和采样协议，从每个用户的密钥值集中采样一个密钥值对。如果用户的键值对少于 1 个，则虚拟对将被填充，直到拥有 l 个键值对。然后，PCKV 随机选择一组键值对，通过 UE 或 GRR (即 PCKV-UE 或 PCKV-GRR)干扰它，并将其发送给数据收集器。然而，PrivKVM 需要运行多轮测试。这要求所有用户在所有过程中在线，这在许多应用程序中是不切实际的。PCKV 的两种机制也有一些其他的限制。对于 PCKV-UE，通信带宽成本与键域大小成正比，这在许多实际应用程序中可能是不可接受的。对于 PCKV-GRR，其扰动概率与键域大小有关，因此 GRR 的估计精度随着键域大小的增大而迅速降低。

2.2. 基本原理

在本地差分隐私中，用户端在本地对数据进行扰动，然后将扰动后的数据发送给服务器，服务器再利用接收到的数据计算得到所需的统计信息。由于服务器无法接触到用户的原始数据，因而无法获得用户的隐私信息。本地差分隐私的形式化定义如下。

定义 1. $\varepsilon - LDP$ 本地差分隐私[3]。给定隐私预算 $\varepsilon \in \mathbb{R}^+$，随机机制 M 满足 $\varepsilon - LDP$，当且仅当对于

任意一对输入 x、x' 和任意输出 y，输出相同 y 的概率比满足

$$\frac{\Pr\left(M\left(x\right)=y\right)}{\Pr\left(M\left(x'\right)=y\right)}\le e^{\varepsilon} \tag{1}$$

如果用户 j 根据之前的输出 $\left\{z^1,\ldots,z^{j-1}\right\}$，采用随机化机制 M^j 而不是固定通用的 M，我们称这种随机化机制为提供交互式的 $\varepsilon-LDP$。

定义 2. 交互式 $\varepsilon-LDP$ 本地差分隐私[3]。定义 D_{M^j} 为随机扰动机制 $M^j\left(j\in\left[1,n\right]\right)$ 的输出域，随机变量 z^j 是 x^j 的一个 $\varepsilon-LDP$ 视图。如果

$$\sup_{y\in D_{M^j}}\frac{\Pr\left(M^j\left(s\right)=y\mid x^j=s,z^1=z^1,\ldots,z^{-1}=z^{j-1}\right)}{\Pr\left(M^j\left(s'\right)=y\mid x^j=s',z^1=z^1,\ldots,z^{-1}=z^{j-1}\right)}\le e^{\varepsilon} \tag{2}$$

适用于所有的 $z^1\in M^1,\ldots,z^{j-1}\in M^{j-1}$，任意数据对 $s,s'\in\chi^m$，任意输出 $y\in D_{M^j}$，我们称这种机制 $\boldsymbol{M}=\left\{M^1,\ldots,M^n\right\}$ 满足交互式的 $\varepsilon-LDP$，$\varepsilon-LDP$ 的非交互版本当且仅当 $M^j\equiv M$。

3. 算法设计

3.1. 方案设计

Wheel 机制[10]是由王等人在 2020 年提出的用于集合数据频率估计的机制，而 Wheel-KV 方案是基于键值数据提出的一种新的机制，该机制对键值数据的收集与统计在效益上有较大的提升。

算法 1. Wheel-KV 用户端扰动算法。

输入：键值数据集合 $S=\left\{\langle key_1,value_1^*\rangle\ldots\langle key_m,value_m^*\rangle\right\}$，其中 $key_i\in\mathcal{K}$，$value_i^*\in\left[-1,1\right]$，随机种子为 s 的哈希函数 $h_s:key\to k:\left[0.0,1.0\right)$，隐私预算 ε，覆盖参数 $p=1\Big/\left(m\cdot\frac{e^{\epsilon}+1}{2}+2m-1\right)$。

输出：满足 $\varepsilon-LDP$ 的隐私视图 $z\in\left[0.0,1.0\right)$。

① $value\leftarrow 1\ w.p.\frac{1+value^*}{2}$ 或 $value\leftarrow -1\ w.p.\frac{1-value^*}{2}$；/* 将 $value^*\in\left[-1.0,1.0\right]$ 离散到 $value\in\left\{-1,1\right\}$ */。

② $\mathbf{k}=\left\{k_1\ldots k_m\right\}=h_s\left(\mathbf{key}\right)$；/* 将原始数据映射到 $\left[0.0,1.0\right)$ */。

③ $C_{k_i}=\left\{y\mid y\in\left[k_i,k_i+p\right)or\left[0,k_i+p-1\right),i\in\left\{1,\ldots,m\right\}\right.$。

④ $\mathcal{Q}_{z_0}\left[y\mid\left\{\langle k_1,value_1\rangle,\ldots,\langle k_m,value_m\rangle\right\}\right]=\begin{cases}\dfrac{(e^{\epsilon}+1)/2}{m\cdot p\cdot(e^{\epsilon}+1)/2+(1-m\cdot p)}, if\ y\in C_{k_i},i=\{1,\ldots,m\}\\ \dfrac{\Omega-l\cdot e^{\varepsilon}}{(1-l)\Omega}, otherwise.\end{cases}$。

⑤ $z_1=Adjust\left(z_0,value\right)$；/* 根据 value 的大小对 z_0 进行调整，得到最终输出 z_1 */。

⑥输出 z_1。

算法 2．Wheel-KV 服务器端聚合算法。

输入：用户的隐私视图和对应的随机种子 $\left(\boldsymbol{s},\boldsymbol{z}\right)=\left\{\left(s^1,z^1\right),\left(s^2,z^2\right),\ldots,\left(s^n,z^n\right)\right\}$。

输出：所有键的频率估计 $\hat{f}_k$ 与对应值的均值估计 $\hat{m}_k$。

① $n_1=\left\{0\right\}^d$，$n_2=\left\{0\right\}^d$；

② for $\left(s^j,z^j\right)\in\left(\boldsymbol{s},\boldsymbol{z}\right)$；

③ for $key_i\in\mathcal{K}$；

④ $k_i^j = h_{s^j}(key_i)$；
⑤ if $z^j - p < k_i^j \le z^j$ or $z^j - p + 1 < k_i^j$；
⑥ if $z^j - \frac{p}{2} < k_i^j \le z^j$ or $z^j - \frac{p}{2} + 1 < k_i^j$；
⑦ $n_{1i} = n_{1i} + 1$；
⑧ else $n_{2i} = n_{2i} + 1$；
⑨ end if
⑩ end if
⑪ end for
⑫ end for
⑬ for $i = 1$ to d；
⑭ $\widehat{f}_i = \frac{1}{n} \cdot \frac{n_{1i} + n_{2i} - n \cdot P_f}{P_t - P_f}$；
⑮ $\widehat{m}_i = \frac{(n_{1i} - n_{2i}) / (P_v - (1 - P_v))}{n \widehat{f}_i P_t}$；
⑯ end for
⑰ 输出 频率估计 $\hat{f}_k$ 与对应值的均值估计 $\hat{m}_k$。

3.2. 键值数据域突然变化时的键值数据收集机制对比

PrivKVM 机制：从当前键域中随机抽取一个序号，将扰动后的键值对发给服务器，扰动概率与键域大小 d 无关，且扰动对象为当前键域的所有键。

PCKV-GRR 机制：扰动概率 $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1} + d' - 1} = \frac{l(e^{\varepsilon} - 1) + 2}{l(e^{\varepsilon} - 1) + 2d'}$，$b = \frac{1}{e^{\varepsilon_1} + d' - 1} = \frac{2}{l(e^{\varepsilon} - 1) + 2d'}$，与键域大小 d 有关，且扰动对象为当前键域的所有键，输出为一个扰动后的键值对。

PCKV-UE 机制：扰动概率 $a = 0.5$，$b = \frac{1}{e^{\varepsilon_1} + 1} = \frac{2}{e^{\varepsilon} + 3}$，与键域大小 d 无关，且扰动对象为当前键域的所有键，输出为一个长度与键域大小 d 有关的键值对向量。

Wheel-KV 机制：扰动概率介于 $\frac{(e^{\epsilon} + 1)/2}{m \cdot p \cdot (e^{\epsilon} + 1)/2 + (1 - m \cdot p)} \cdot \frac{2}{e^{\epsilon} + 1}$ 和 $\frac{(e^{\epsilon} + 1)/2}{m \cdot p \cdot (e^{\epsilon} + 1)/2 + (1 - m \cdot p)} \cdot \frac{2e^{\epsilon}}{e^{\epsilon} + 1}$ 之间，与键域大小 d 无关，扰动对象为当前或更新后键域的所有键，输出为一个隐私视图实数 $z \in [0.0, 1.0)$。

键域新增时各机制对比如表 1 所示。

Table 1. Comparison of mechanisms when adding key fields
表 1. 键域新增时机制对比

机制	扰动概率	通信带宽	影响	调整通信
PrivKVM	不变	不变	无法采样新增键	需要
PCKV-GRR	偏大	不变	不满足 $\varepsilon - LDP$	需要
PCKV-UE	不变	不足	无法传输新增键的信息	需要
Wheel-KV	不变	不变	无	不需要

键域删减时各机制对比如表 2 所示。

Table 2. Comparison of mechanisms for key field deletion
表 2. 键域删减时机制对比

机制	扰动概率	通信带宽	影响	调整通信
PrivKVM	不变	不变	采样率偏低，估计的准确度偏低	需要
PCKV-GRR	偏小	不变	估计的准确度偏低	需要
PCKV-UE	不变	过大	浪费了带宽	需要
Wheel-KV	不变	不变	无	不需要

3.3. 服务器端个性化榜单设置

根据 Wheel-KV 机制在用户端与键域无关的性质，我们可以在服务器端设计出个性化的榜单设置，以提高效用。

场景 1：本地差分隐私下的新歌推荐榜

令歌曲 ID 为键值对的键，歌曲喜好程度为键值对的值，新歌曲库为全曲库的子集；发新歌时，新歌键域新增，发歌时间到达某时长后，歌曲不再计入新歌曲库，键域删减。该场景满足键域随时变化的模型，框架如图 1 所示。

Wheel-KV 机制实现：

用户端：正常扰动。

服务器端：聚合操作只对新歌键域 K_{new} 进行遍历扫描，而 K_{new} 则根据新歌的发布或过期进行新增和删减。

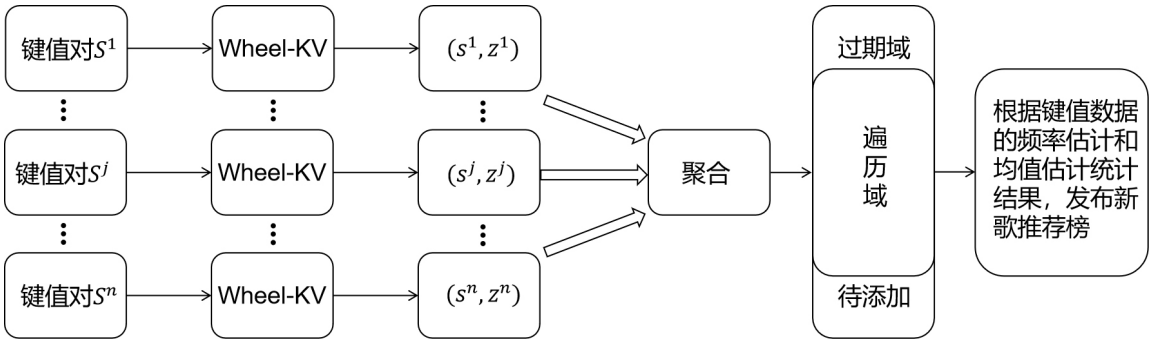


Figure 1. New song chart data collection process
图 1. 新歌榜数据收集过程

场景 2：本地差分隐私下热搜键值对的统计分析

在很多情况下，我们只对频繁项更为关注，对于全键域的统计分析往往需要耗费大量通信资源和计算资源，无法支持实时更新。因此，可以考虑在较长间隔里进行全键域的统计分析，取出前 10%的高频子键域，作为 Wheel-KV 机制遍历的键域，以此来统计前 1%的高频键值对，针对突发热词，可以人为添加进该键域。该场景满足键域随时变化的模型，框架如图 2 所示。

Wheel-KV 机制实现：

用户端：正常扰动。

服务器端：聚合操作只对一级频繁键域 K_{hot} 进行遍历扫描，而 K_{hot} 则根据人为改动进行新增和删减。

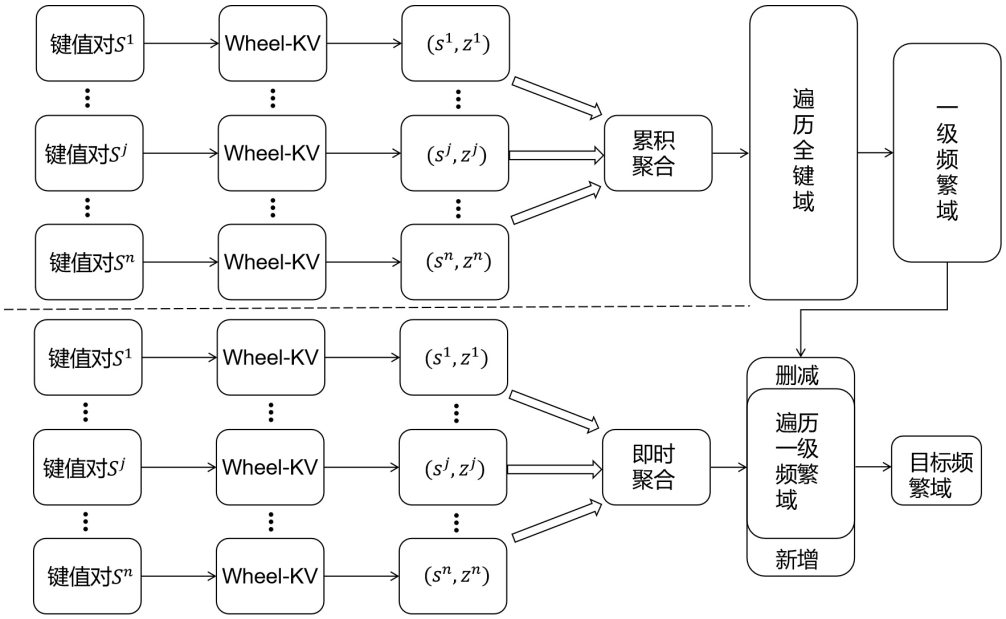


Figure 2. Real time high-frequency key value pair data collection process
图 2. 即时高频键值对数据收集过程

4. 实验分析

4.1. 评价指标

我们通过所有键或部分键之间的均方误差(MSE)来评估频率和均值估计。

$$MSE_{freq} = \frac{1}{|\chi|}\sum_{i\in\chi}\left(\hat{f}_i - \hat{f}_i^*\right)^2 \tag{3}$$

$$MSE_{mean} = \frac{1}{|\chi|}\sum_{i\in\chi}\left(\hat{m}_i - \hat{m}_i^*\right)^2 \tag{4}$$

其中：f_i^* 和 m_i^* 是真实频率和均值，$\hat{f}_i$ 和 $\hat{m}_i$ 是估计的频率和均值，χ 是键域 K 的子集，默认域为 K；我们还考虑 χ 是前 N 个频繁键的集合(如前 50 个)，因为我们通常只关心频繁键的估计结果，且不常见的键可能并没有足够的样本来获得值均值的准确估计。

4.2. 参数设置

我们在数据集上进行了实验，在数据集中，用户数为 $N = 10^5$，每个用户所拥有的 *key* 和 *key* 的值均满足均匀(或高斯)分布。高斯分布由 $\mu = 0$，$\sigma_{key} = 50$，$\sigma_{value} = 1$ 生成。表 3 总结了我们所使用的不同域大小和数据分布的数据集。

Table 3. Statistical information of experimental database
表 3. 数据集参数描述

数据集	用户数 n	键域大小 d	集合大小 m
机制对比	100,000	512	8
带宽对比	100,000	10,000	8
键域新增	10,000 → 20,000	100 → 200	4
键域删减	10,000 → 20,000	100 → 50	4

4.3. 实验结果

4.3.1. 隐私预算对 MSE 的影响

本节对比了 5 种机制在不同隐私预算 ε 下的性能差异，实验结果如图 3 所示。可以看出，随着隐私增大，5 种机制的 MSE 都在减小，相应地，频率估计和均值估计的结果会更加准确，数据效用也就越高。即 ε 越大，隐私保护程度越低，MSE 越小，数据效用越高。同时，可以看到本文使用的 WheelKV 机制具有较好的效用，在频率估计和均值估计两方面都有不错的表现。

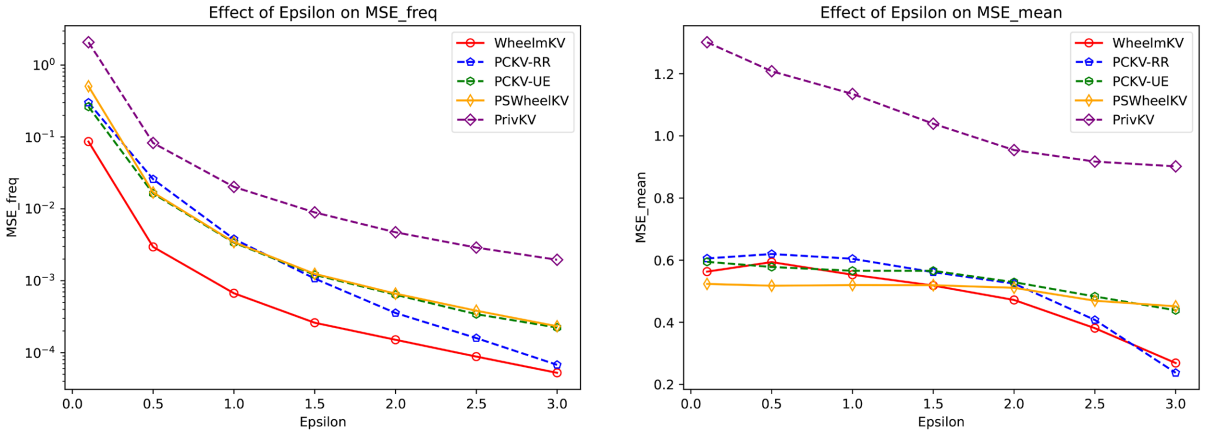


Figure 3. Effect of epsilon on MSE
图 3. 不同机制下隐私预算对均值估计 MSE 的影响

4.3.2. 不同场景下额外通信次数

正本节对比了 4 种机制在面临键域变化时，额外通信的次数，结果如表 4 所示。可以看出，当键域新增或减少时，3 种机制都需要进行额外通信，而 Wheel-KV 机制则不需要这一额外开销。

Table 4. The number of communications when the keyfield change
表 4. 不同机制下遇到键域改变时的通信次数

数据集	原始通信	键域新增	键域删减
Wheel-KV	1	0	0
PrivKVM	1	1	1
PCKV-GRR	1	1	1
PCKV-YE	1	1	1

4.3.3. 不同场景下额外通信次数通信带宽

Table 5. Communication bandwidth when the keyfield change
表 5. 不同机制下遇到键域改变时的通信带宽

数据集	原始通信	键域新增	键域删减
Wheel-KV	(O) 3	(O) 3	(O) 3
PrivKVM	(O) 13	(O) 14	(O) 12
PCKV-GRR	(O) 13	(O) 14	(O) 12
PCKV-UE	(O) 10,000	(O) 20,000	(O) 5000

本节对比了当 m 为 8， d 为 10,000 时，不同机制的通信带宽，结果如表 5 所示。可以看出，当键域新增至两倍或减少至一半时，3 种机制带宽有较大变化，且均大于 Wheel-KV 机制的带宽。

4.3.4. 对比新增键的统计准确性

在本节对比了当键域新增时，Wheel-KV 机制分批处理和朴素地将所有数据一视同仁处理的性能差异，实验结果如图 4 所示。可以看出，随着键域增大，数据量变大，2 种机制的 MSE 都在减小，但分批处理地 MSE 显然小于一致处理，数据效用更高。

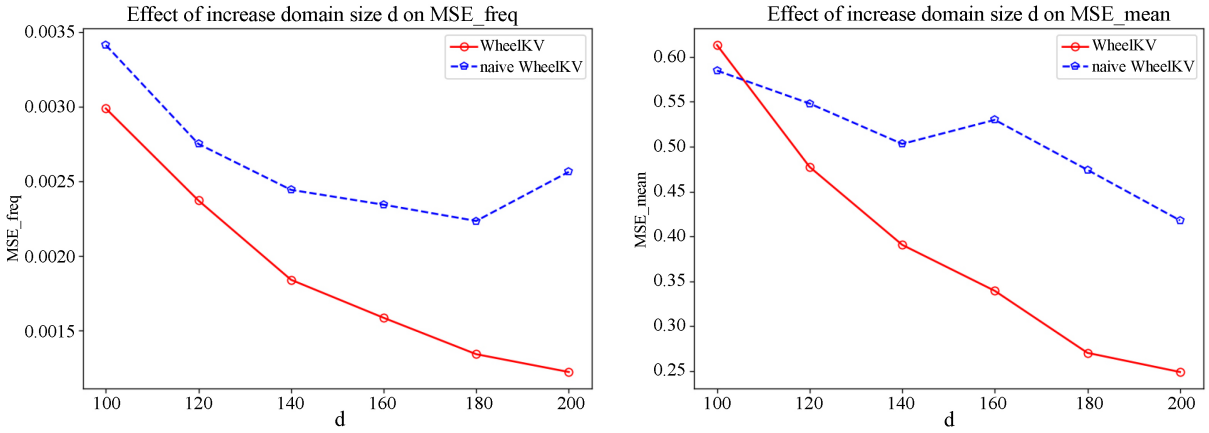


Figure 4. Effect of epsilon on MSE when adding key fields
图 4. 不同机制下键域新增时的 MSE

4.3.5. 对比删减键的统计准确性

本节对比了当键域删减时，Wheel-KV 机制分批处理和朴素地将所有数据一视同仁处理的性能差异，实验结果如图 5 所示。可以看出，随着键域减小，数据量变大，2 种机制的 MSE 都在减小，但分批处理地 MSE 显然小于一致处理，数据效用更高。

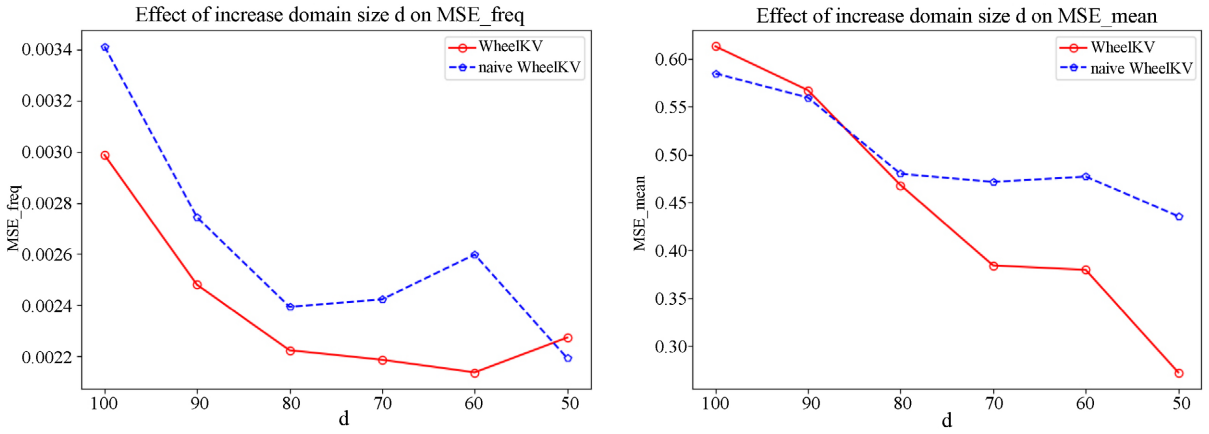
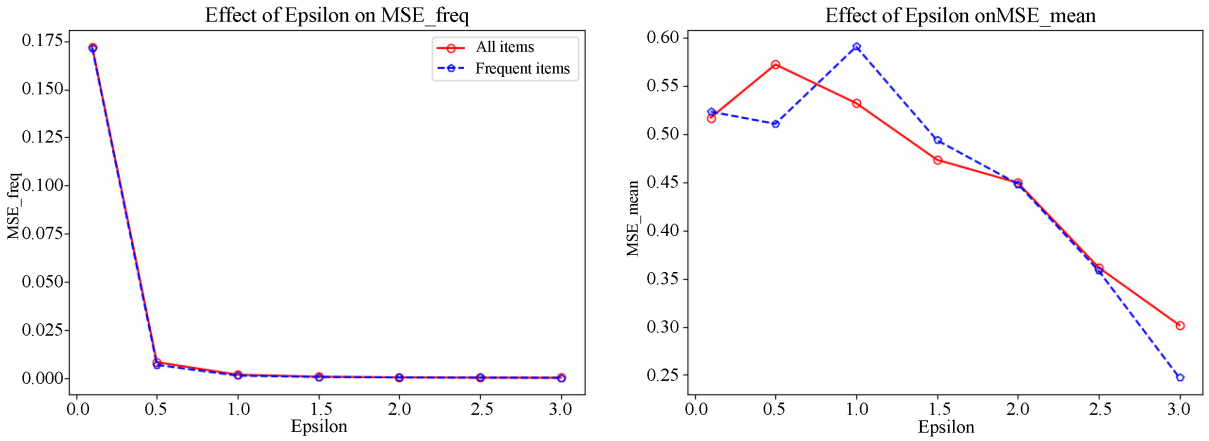


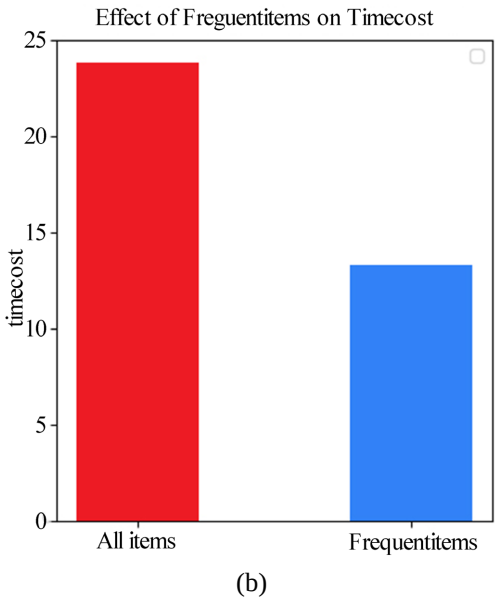
Figure 5. Effect of epsilon on MSE for key field deletion
图 5. 不同机制下键域删减时的 MSE

4.3.6. 服务器端个性化设置频繁项集

本节对比了频繁项集处理和全域处理的性能差异，实验结果如图 6 所示。可以看出，MSE 的差异不大，而频繁项集处理的耗时显著降低。



(a)



(b)

(a) 频繁项处理对 MSE 的影响；(b) 频繁项处理对 MSE 耗时的影响

Figure 6. Personalized frequent item processing
图 6. 个性化频繁项处理

5. 总结

本文首先针对键域频繁变化的场景提出使用一种键域独立的本地差分隐私保护机制，该机制是一种键域独立的扰动机制，它在用户侧与键域大小和内容无关，具有较高的可用性。本文具体分析了键域增加或减少情况下，各个方案对于扰动机制的处理与分析，通过对比实验验证本方案的优越性。本文在服务器端设计个性化的统计机制，使得数据的使用与分析更便捷且具有多样性。

未来的工作是针对本文所提的键域独立的思想，对本地差分隐私下的其他机制进行深入研究，探索实际应用中键域随时变化情况下，达到更高的数据效用的方法和机制。

基金项目

江苏省重点研发计划(产业前瞻与关键核心技术)项目(BE2022068, BE2022068-1)；
中国高校产学研创新基金——新一代信息技术创新项目课题(2023IT049)。

参考文献

[1] Dwork, C. (2008) Differential Privacy: A Survey of Results. In: *Lecture Notes in Computer Science*, Springer, 1-19. https://doi.org/10.1007/978-3-540-79228-4_1

[2] Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006) Calibrating Noise to Sensitivity in Private Data Analysis. In: *Lecture Notes in Computer Science*, Springer, 265-284. https://doi.org/10.1007/11681878_14

[3] Duchi, J.C., Jordan, M.I. and Wainwright, M.J. (2013) Local Privacy and Statistical Minimax Rates. 2013 *IEEE* 54*th Annual Symposium on Foundations of Computer Science*, Berkeley, 26-29 October 2013, 429-438. https://doi.org/10.1109/focs.2013.53

[4] Erlingsson, Ú., Pihur, V. and Korolova, A. (2014) RAPPOR. *Proceedings of the* 2014 *ACM SIGSAC Conference on Computer and Communications Security*, New York, 1054-1067, 3-7 November 2014. https://doi.org/10.1145/2660267.2660348

[5] Ye, Q., Hu, H., Meng, X., Zheng, H., Huang, K., Fang, C., *et al*. (2023) Privkvm: Revisiting Key-Value Statistics Estimation with Local Differential Privacy. *IEEE Transactions on Dependable and Secure Computing*, **20**, 17-35. https://doi.org/10.1109/tdsc.2021.3107512

[6] Gu, X.L., Li, M., Cheng, Y.Q., *et al*. (2020) PCKV: Locally Differentially Private Correlated Key-Value Data Collection with Optimized Utility. 29*th USENIX Security Symposium* (*USENIX Security* 20), Boston, 12-14 August 2020, 967-984.

[7] Sun, L., Zhao, J., Ye, X.J., *et al*. (2019) Conditional Analysis for Key-Value Data with Local Differential Privacy.

[8] 魏立斐, 陈聪聪, 张蕾, 等. 机器学习的安全问题及隐私保护[J]. 计算机研究与发展, 2020, 57(10): 2066-2085.

[9] 郭娟娟, 王琼霄, 许新, 等. 安全多方计算及其在机器学习中的应用[J]. 计算机研究与发展, 2021, 58(10): 2161-2186.

[10] Wang, S., Qian, Y., Du, J., Yang, W., Huang, L. and Xu, H. (2020) Set-Valued Data Publication with Local Privacy. *Proceedings of the VLDB Endowment*, **13**, 1234-1247. https://doi.org/10.14778/3389133.33891407