International Journal of Information Security and Privacy
Volume 20 • Issue 1 • January-December 2026

Personalized Local Differential Privacy Frequency Estimation Mechanisms Based on Partitioning the Domain of Real Attribute Values

Yunfei Li
https://orcid.org/0009-0006-4004-5360
Kunming University of Science and Technology, China & Yunnan University of Finance and Economics, China

Xiaodong Fu
https://orcid.org/0000-0002-1354-2293
Kunming University of Science and Technology, China

Li Liu
Kunming University of Science and Technology, China

Jiaman Ding
Kunming University of Science and Technology, China

Wei Peng
Kunming University of Science and Technology, China

Lianyin Jia
Kunming University of Science and Technology, China

Received: August 13th, 2025 | **Accepted:** February 3rd, 2026

ABSTRACT

Existing multi-domain personalized local differential privacy (MDPLDP) mechanisms, which extend attribute domains by introducing fake values, often fail to provide adequate personalized privacy protection and limit utility in frequency estimation. To address these limitations, the authors propose two novel MDPLDP mechanisms that construct multiple domains by partitioning real attribute values, support cross-domain aggregation, and flexibly accommodate diverse privacy requirements and budgets. The methods further extend to multi-dimensional frequency estimation, catering to complex user privacy preferences. Theoretical analysis and experimental results demonstrate that our mechanisms achieve substantially lower estimation error and communication overhead, while delivering over 20% average utility improvement compared to state-of-the-art methods in both single- and multi-dimensional settings.

KEYWORDS

Personalized Local Differential Privacy, Frequency Estimation, Utility Optimization, Domains of Real Attribute Values, Multi-Domains, Multi-Dimensional Data

INTRODUCTION

With the rapid development of mobile intelligence and Internet of Things devices, data collection and utilization have reached unprecedented scales. At the same time, the growing awareness of public privacy protection, along with the introduction of regulations such as the General Data Protection Regulation and the California Consumer Privacy Act, has highlighted the increasing tension between data privacy concerns and data collection practices (Arcolezi et al., 2022; Yang et al., 2023; Zhou et al., 2023). This has further deepened the demand for decentralized, lightweight, and compliant

DOI: 10.4018/IJISP.401370

privacy protection technologies, which have become key research topics in the field of big data privacy protection (Zhou et al., 2023).

Differential privacy (DP), due to its rigorous security model and efficient, low-overhead characteristics, has become one of the core technologies for ensuring privacy protection (Yang et al., 2023). Local differential privacy (LDP), as an important form of DP, not only retains the advantages of DP but also avoids the dependence on trusted third parties in centralized differential privacy (CDP), making it naturally compliant with privacy regulations and significantly enhancing the practical applicability of the technology (Yang et al., 2023). Currently, LDP has been widely applied in privacy protection for distributed data collection (Arcolezi et al., 2022; Shen et al., 2024; Yang et al., 2023; Zhou et al., 2023). However, the local data perturbation mechanism of LDP can lead to a reduction in data utility, limiting its further application (Yang et al., 2023). Additionally, users have different privacy protection needs due to factors such as profession and culture. If LDP cannot provide flexible privacy protection levels, it will face practical challenges in data collection (Bao et al., 2022; Gu et al., 2020; Li et al., 2025; Nie et al., 2018; Zhu et al., 2023).

Frequency estimation, as a fundamental operation within the LDP framework, directly impacts the performance of higher-level applications such as machine learning (Arcolezi, 2022), recommendation systems (Bao et al., 2022), frequent itemset mining (Wu et al., 2023), multidimensional data estimation (Arcolezi et al., 2021; Arcolezi et al., 2023; Qiu et al., 2023; Wang et al., 2019), and federated learning (Liu et al., 2023). Therefore, enhancing frequency estimation utility while meeting users' personalized privacy needs has become a hot topic in current LDP research (Bao et al., 2022; Gu et al., 2020; Li et al., 2025; Nie et al., 2018; Zhu et al., 2023;).

Existing solutions can be classified into three categories: (1) personalized LDP (PLDP) methods based on personalized privacy budgets (Gu et al., 2020; Nie et al., 2018; Zhu et al., 2023). These methods allocate different privacy budgets within the same domain of attribute values, defining different levels of privacy protection. By assigning higher privacy budgets to less sensitive data, they can enhance data utility while meeting personalized privacy needs. However, when the privacy budget is fixed, they still fail to fully satisfy personalized privacy preferences. (2) utility-optimized LDP (ULDP) methods based on domains of sensitive and non-sensitive attribute values (Cao et al., 2022; Murakami et al., 2019; Zhu et al., 2024). These methods reduce the domain of sensitive attribute values to improve data utility, with privacy protection levels for the domain of sensitive attribute values defined by the privacy budget. However, these methods tend to divide the domain of sensitive attribute values in a relatively simple manner, unable to meet the needs of multiple domains of sensitive attribute values and still fall short of meeting personalized privacy requirements when the privacy budget is fixed. (3) PLDP methods based on multiple domains of attribute values (Chen et al., 2019; Li et al., 2025; Shen et al., 2023). These methods partition the domain of attribute values into smaller subdomains and characterize the privacy protection strength from both the domain of attribute values and privacy budget dimensions. By reducing the domain of attribute values and setting different privacy protection levels for different subdomains, these methods can enhance data utility (Li et al., 2025). Although these methods can meet personalized privacy needs when the privacy budget is fixed, their frequency estimation often relies on introducing fake (synthetic) values to construct multiple domains, which enlarges the perturbation space and can degrade utility.

Multi-attribute data dominates modern analytics (Arcolezi et al., 2021; Arcolezi et al., 2023; Wang et al., 2019). Existing methods for LDP frequency estimation of multidimensional data primarily include: the simple privacy budget splitting (SPL) scheme, which provides privacy protection for all attributes (Arcolezi et al., 2023) but suffers from low data utility and cannot effectively improve frequency estimation accuracy; the random sampling (SMP) scheme, which improves data utility through sampling techniques (Arcolezi et al., 2023) but may potentially leak information about the attributes involved in the sampling process, thereby increasing the risk of privacy breaches; and the latest approach, the random sampling plus fake data (RS+FD) scheme (Arcolezi et al., 2021; Arcolezi et al., 2023; Wang et al., 2019), which uses fake attributes to "hide" the sampled attributes but still