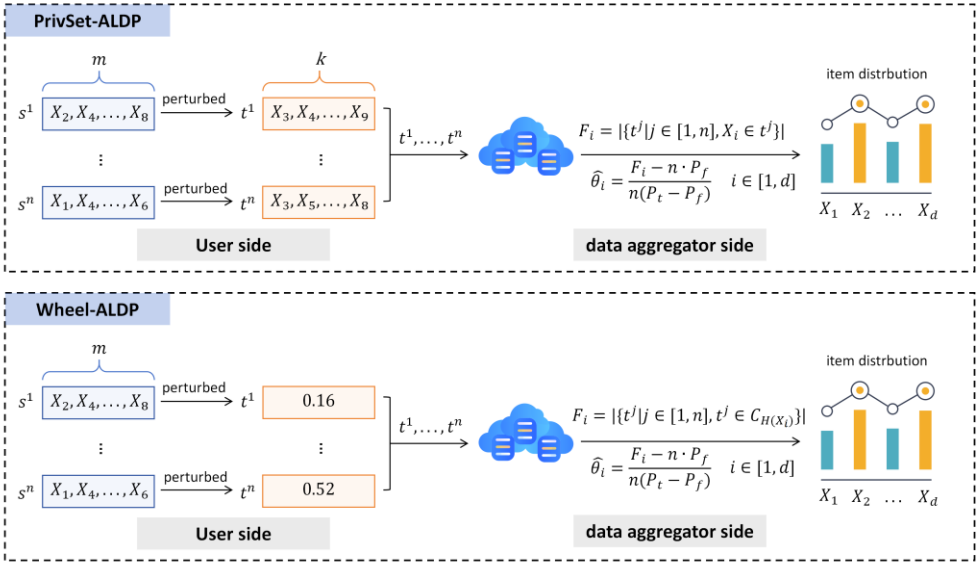
(ϵ, δ)-Local Differential Privacy Mechanisms for Set-Valued Data Analysis

Bing CHEN, Youwen ZHU

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50469-y

Problems & Ideas

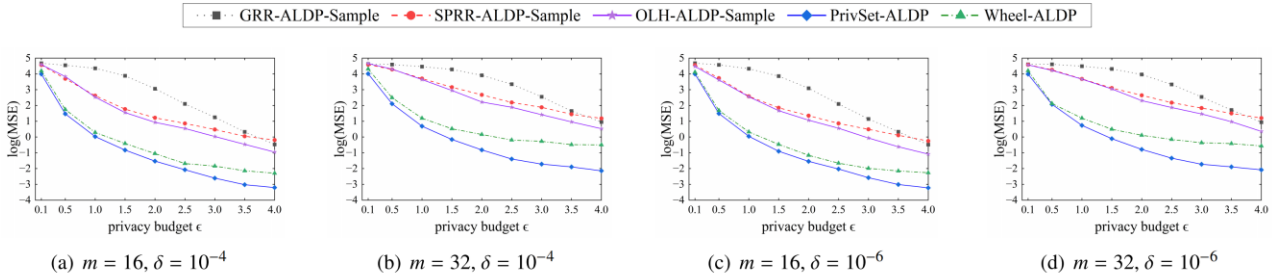
- Problems of set-valued data analysis under (ϵ, δ)-LDP:
  - Existing (ϵ, δ)-LDP approaches focus on frequency estimation of categorical data and mean estimation of numerical data, which are notsuitable for set-valued data.
  - The random-sampling based approaches achieve low utility due to information loss inherent in dimensionality reduction process.
- Ideas: Two new (ϵ, δ)-LDP mechanisms, PrivSet-ALDP and Wheel-ALDP, with higher data utility in frequency estimation of set-valued data.



The overall flow of PrivSet-ALDP and Wheel-ALDP. Each user independently randomizes their true data to perturbed data through PrivSet-ALDP or Wheel-ALDP mechanism. Then, each user sends the perturbed data instead of raw data to the data aggregator. Finally, the data aggregator estimates the item distribution based on all collected perturbed data.

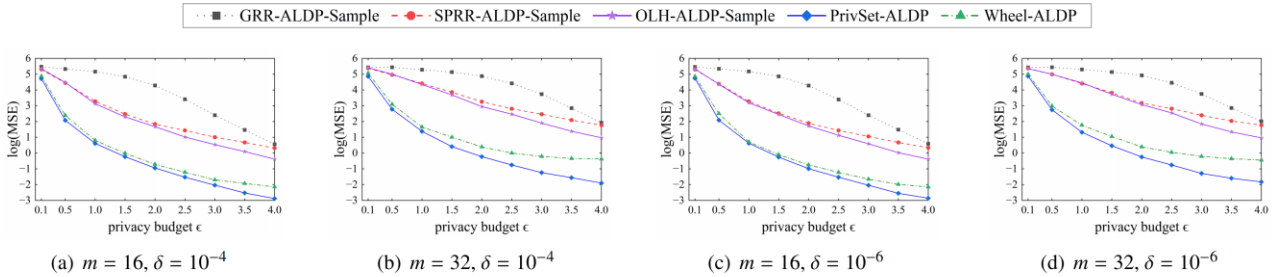
Main Contributions

- Contributions:
  - Two novel (ϵ, δ)-LDP mechanisms, PrivSet-ALDP and Wheel-ALDP, are proposed to support frequency estimation task over set-valued data;
  - The privacy analysis and utility analysis of PrivSet-ALDP and Wheel-ALDP mechanisms;
  - Extensive experiments on synthetic datasets confirm that PrivSet-ALDP and Wheel-ALDP mechanisms achieve better utility compared to the existing methods.



(a) $m = 16, \delta = 10^{-4}$ (b) $m = 32, \delta = 10^{-4}$ (c) $m = 16, \delta = 10^{-6}$ (d) $m = 32, \delta = 10^{-6}$

Fig. 1 MSE of frequency estimation on $n = 10\,000$ users, domain size $d = 256$



(a) $m = 16, \delta = 10^{-4}$ (b) $m = 32, \delta = 10^{-4}$ (c) $m = 16, \delta = 10^{-6}$ (d) $m = 32, \delta = 10^{-6}$

Fig. 2 MSE of frequency estimation on $n = 10\,000$ users, domain size $d = 512$