# Collecting and Analyzing Key-value Data under Shuffled Differential Privacy

Ning WANG, Wei ZHENG, Zhigang WANG, Zhiqiang WEI, Ge YU, Peng TANG, Ge YU

Problems & Ideas

- Problems of collecting key-value data under SDP:
  - The inherent correlation between the key and value sets barriers to the analysis of privacy benefits.
  - Existing approach takes method used in set-valued data, which is not suitable for key-value data.

- Ideas: A SDP framework using PCKV-UE as the basic component to perturb the original data, which sets multiple Shufflers for each key.
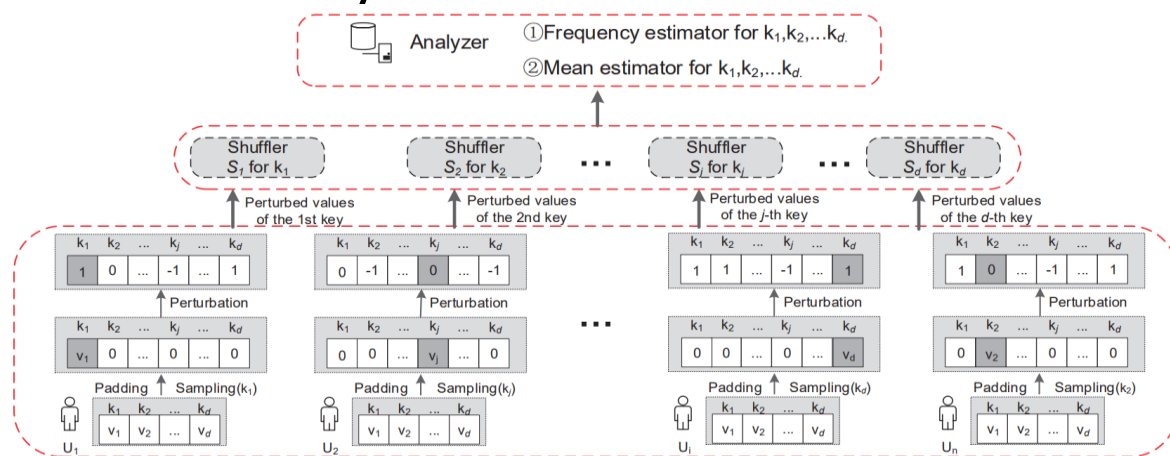
The proposed framework: Users firstly perturb their raw data on local side using PCKV-UE(including Padding&Sampling operation and the core perturbing procedure). Then users send perturbed data to shufflers prepared for each key. Finally the Analyzer receives the shuffled data and performs frequency estimation and mean estimation.

Main Contributions

- Contributions:
  - A new framework named SDPKV is proposed for collecting key-value data under SDP model, which provides strict privacy protection for users, while accurate statistics including frequency for key and mean for value;
  - The analysis of privacy benefits based on the relationship between hockey-stick divergence and differential privacy;
  - Extensive experiments confirm that SDPKV achieves considerably higher result utility compared to the existing methods.

(a) Jdata   (b) MovieData
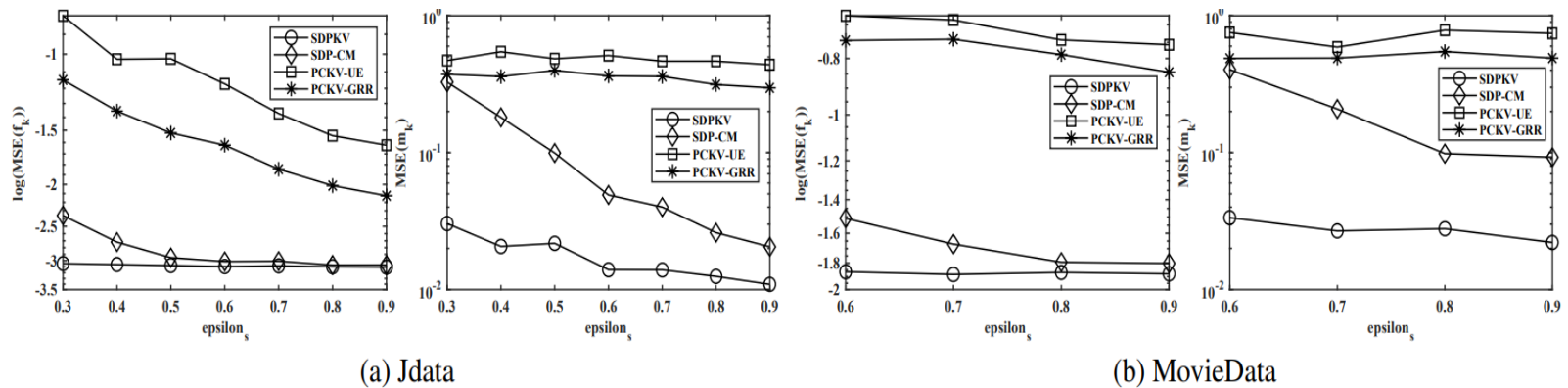
The MSEs of frequency and mean estimation on two real-word datasets. SPDKV outperforms other methods under all given privacy budget.