Efficient Protocols for Heavy Hitter Identification with Local Differential Privacy

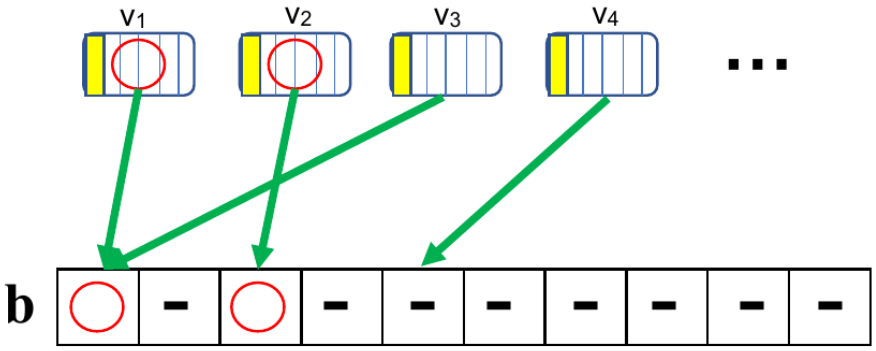
Dan ZHAO, Suyun ZHAO, Hong CHEN, Ruixuan LIU, Cuiping LI, Wenjuan LIANG

Frontiers of Computer Science, DOI: 10.1007/s11704-021-0412-y

Problems & Ideas

- Problems of identifying heavy hitters in large domains under local differential privacy framework.
  - excessive grouping
  - privacy budget allocation
  - large domain with variable length
  - high error

  Ideas: Mixed independent channel for each group
  - flexible splitting
  - false positive rate with interaction

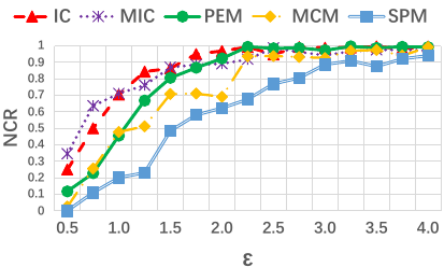


The red circle represents the top-*k* values. We only need to ensure that the red circles are detected in each group.

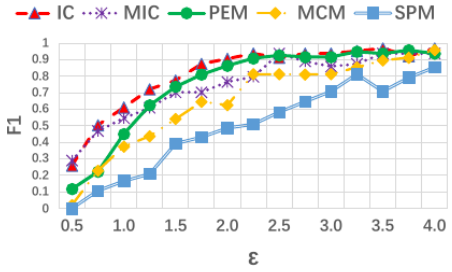
Main Contributions

- We adopted false positive rate with interaction to reduce identification error. Because a value that was not a top-*k* heavy hitter can be detected after iteration with tiny probability. Furthermore, we designed a flexible method to split the long domain into several sub-domains. This splitting strategy was flexible to specific dataset so that the collected data had a better availability.
- We proposed an efficient protocol for heavy hitter, called IC, using flexible splitting and FPR with interaction, which not only reduces the communication and computation cost, but also obtains an accurate estimation. We then proposed the parallel version of IC, namely MIC, which sacrificed some accuracy to achieve the lowest communication cost using the Walsh-Hadamard matrix among all the protocols.
- Numerical experiments demonstrate that IC has better data availability than existing protocols while MIC performs well under a small privacy budget.

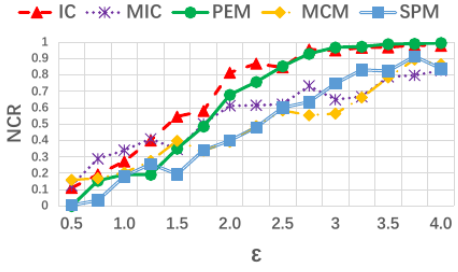
- **Evaluation of different protocols with *k*=16**



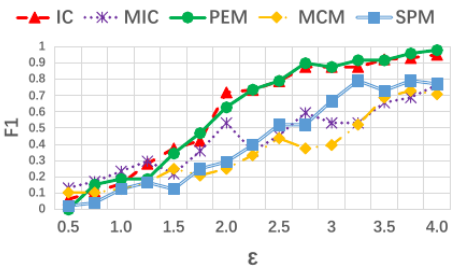
(a) URL, NCR, vary ε



(b) URL, F1, vary ε



(c) AOL, NCR, vary ε



(d) AOL, F1, vary ε