

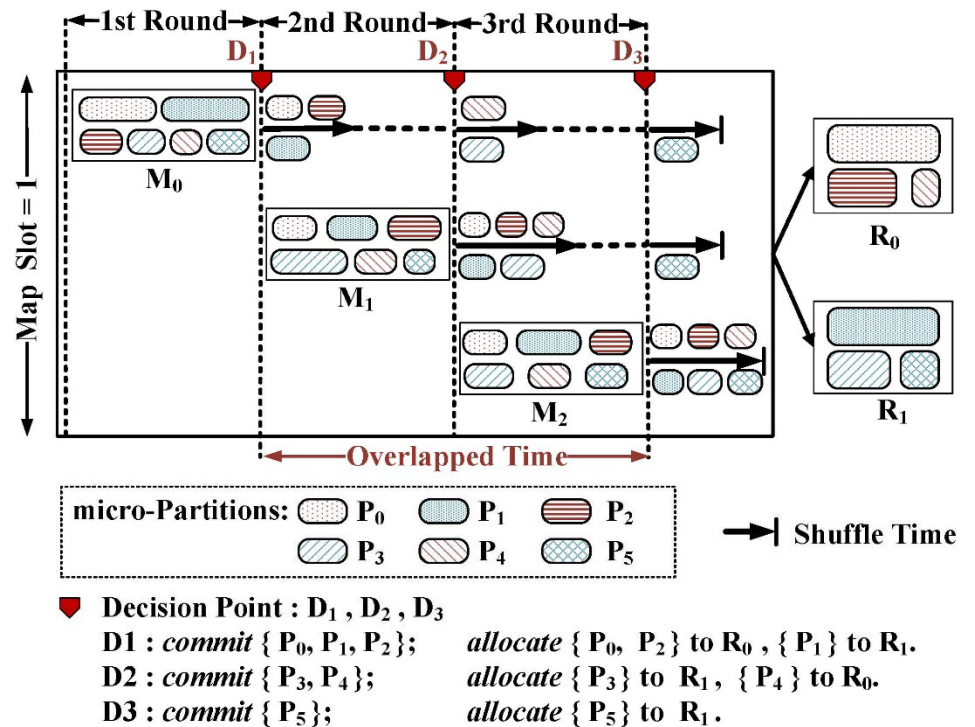
Reducing partition skew on Mapreduce: an incremental allocation approach

Zhuo WANG , Qun CHEN, Bo SUO, Wei PAN, ZhanHuai LI

Frontiers of Computer Science, DOI: [10.1007/s11704-018-6586-2](https://doi.org/10.1007/s11704-018-6586-2)

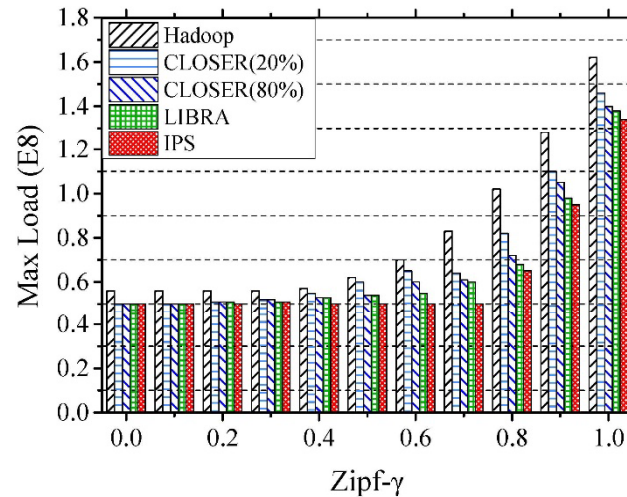
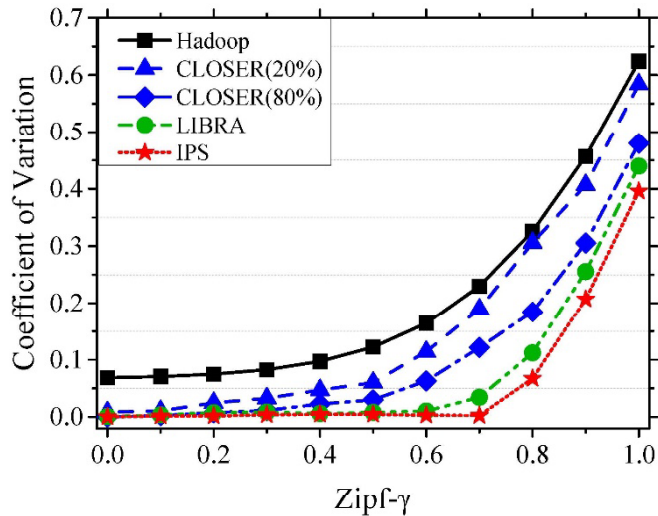
Problems & Ideas

- Problems of reducing partition skew among reducers without detailed distribution information on mapped data
 - get distribution by pre-sampling
 - slow shuffle start for distribution statistics garnering
- Ideas: **(IPS)** An incremental data allocation approach to reduce partition skew
 - a Markov Decision Process(MDP) model to optimize the problem of multiple-round micro-partition scheduling
 - a Min-Max programming model to handle the allocation

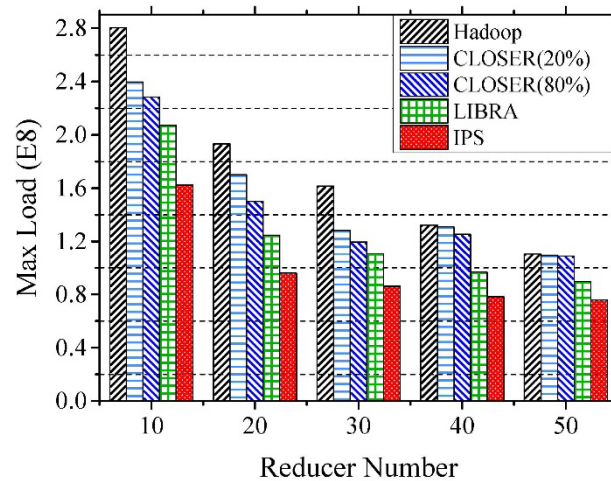
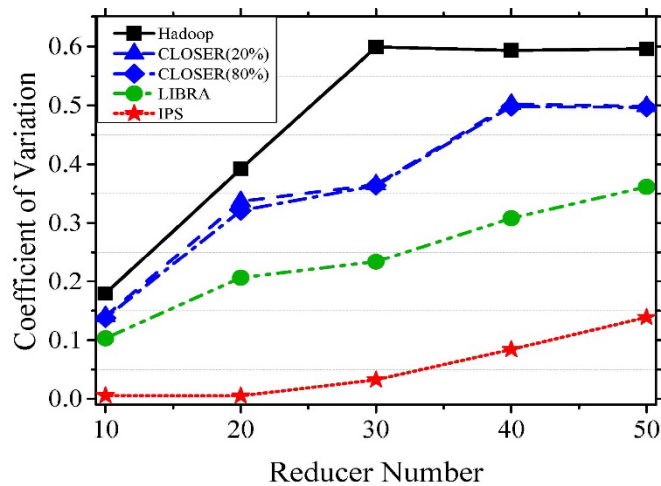


Incremental Allocation Approach

Main Contributions



- The Evaluation Results on Zif p- γ



- The Evaluation Results on real-data