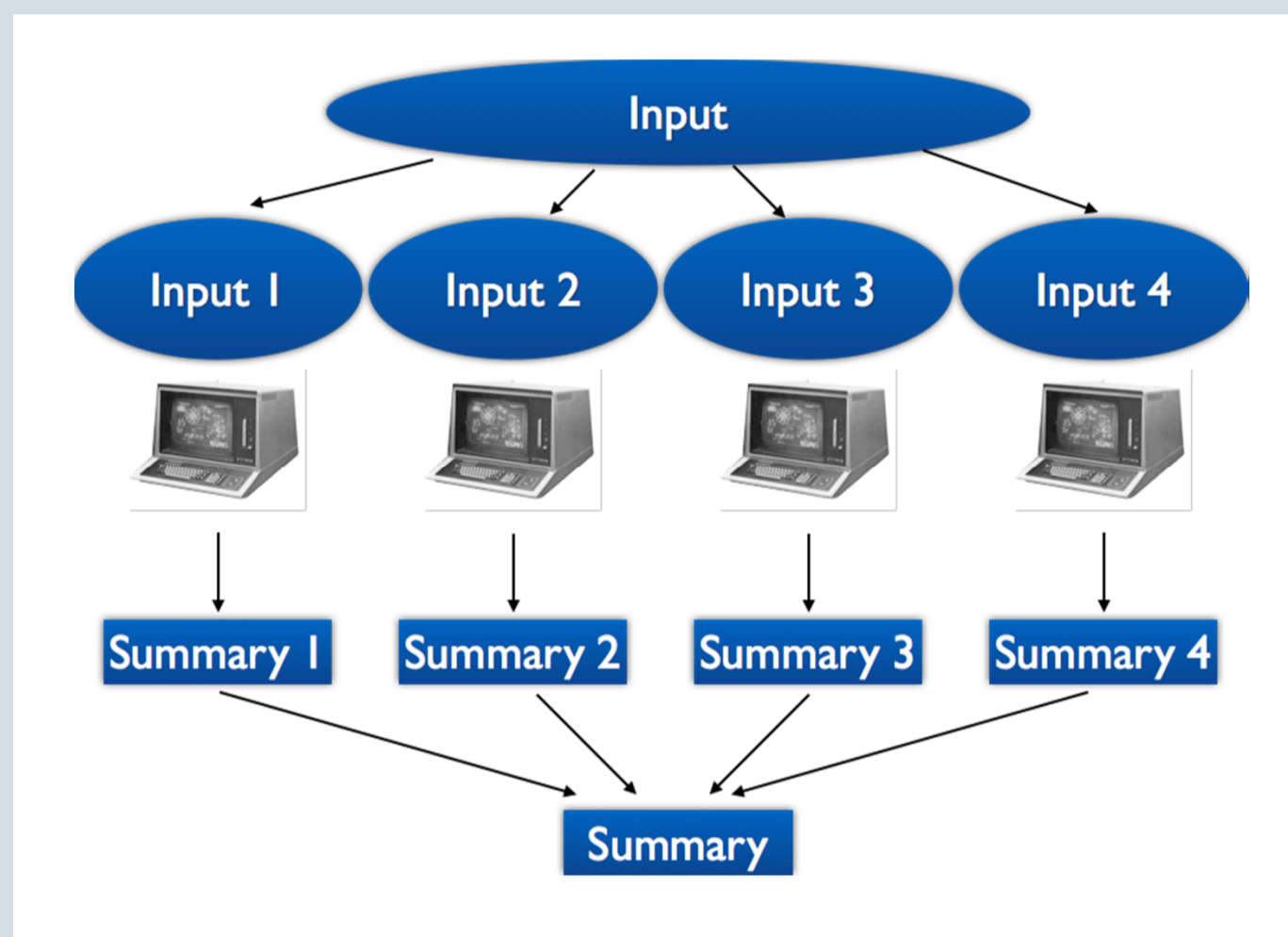


Mergeable Summaries

Introduction

Motivation – Distributed Computing

- Input data is broken into an arbitrary number of pieces
- Each piece is potentially handled by a different machine
- Summaries are combined together to answer queries on original input



Motivations – In-network Aggregation

- Nodes in a sensor network organize themselves into a routing tree
- Each sensor holds some data
- The goal of data aggregation is to compute a summary of all the data

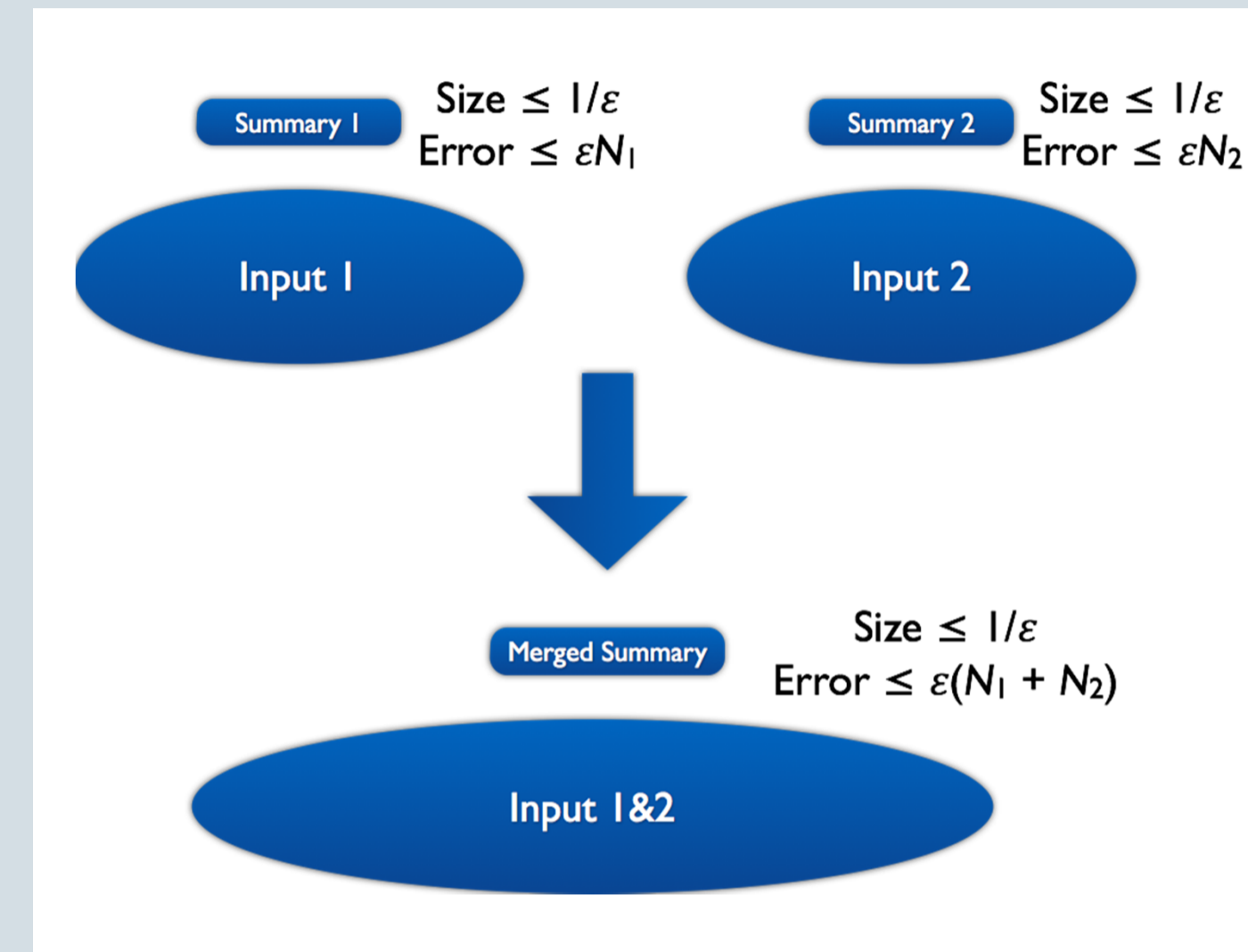


Previous results

To obtain any overall guarantee, it is necessary to have a bound on the number of rounds of merging operations in advance so that the error parameter ϵ can be scaled down accordingly.

Solutions: Mergeable Summaries

A summary is **mergeable**, if error and space does not increase after the merge

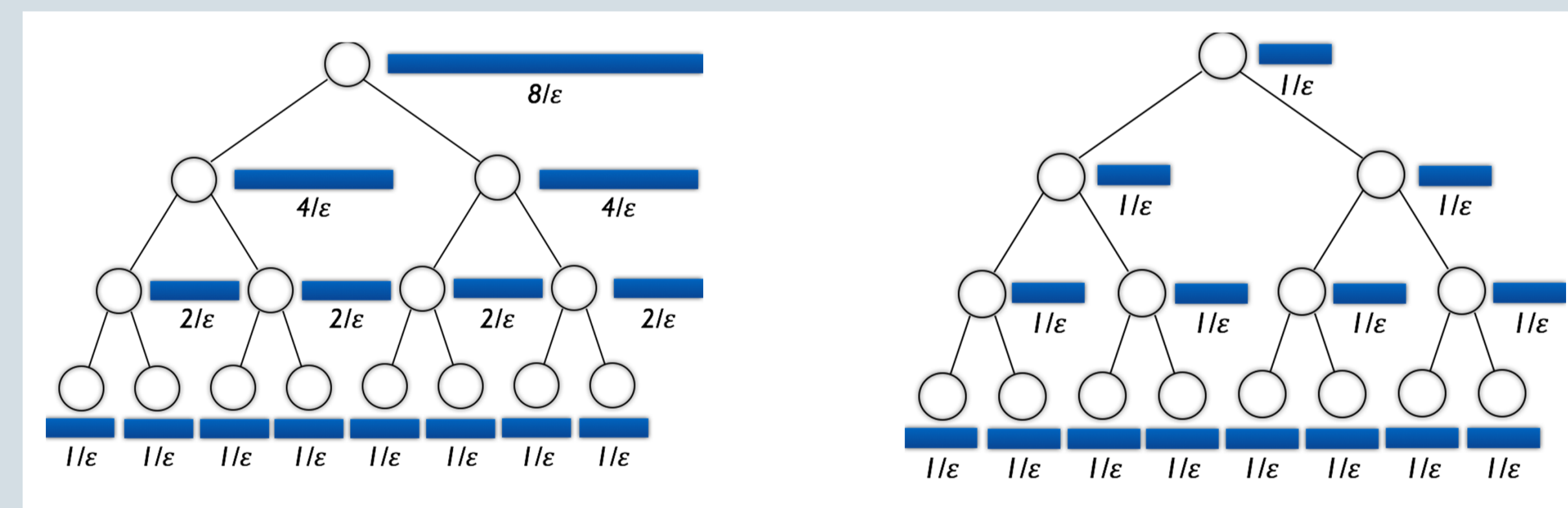


Highlight 1: Maximize Battery Life

- Communication is the major source of battery drain
- Maximize battery life by minimizing summary size

Highlight 2: Balance Power Consumption

- If summary sizes varies on different sensors → Unbalance power consumption over network
- In many cases, life of the network depends on the worst case battery life
- Minimize worst-case power consumption at any node, maximize network lifetime



Highlight 3: Topology Independent

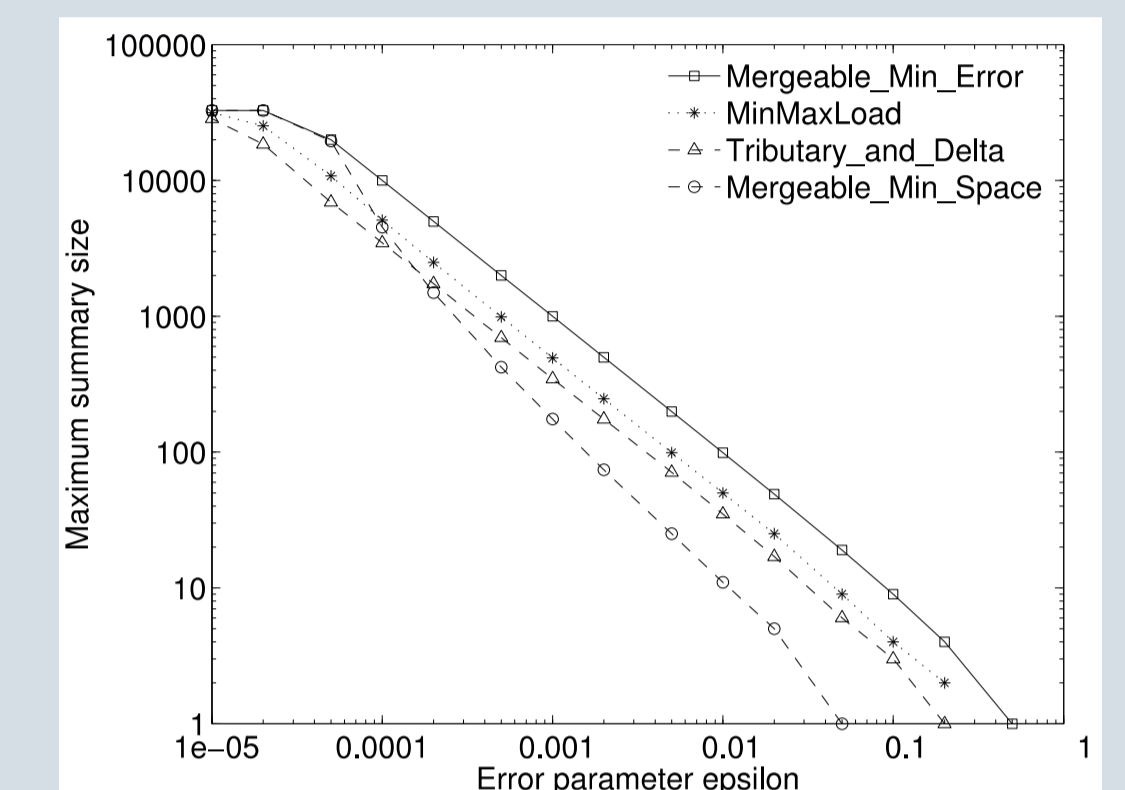
- Works when the number of merges is not prespecified
- Algorithms adapt to network changes

Theoretical Bounds

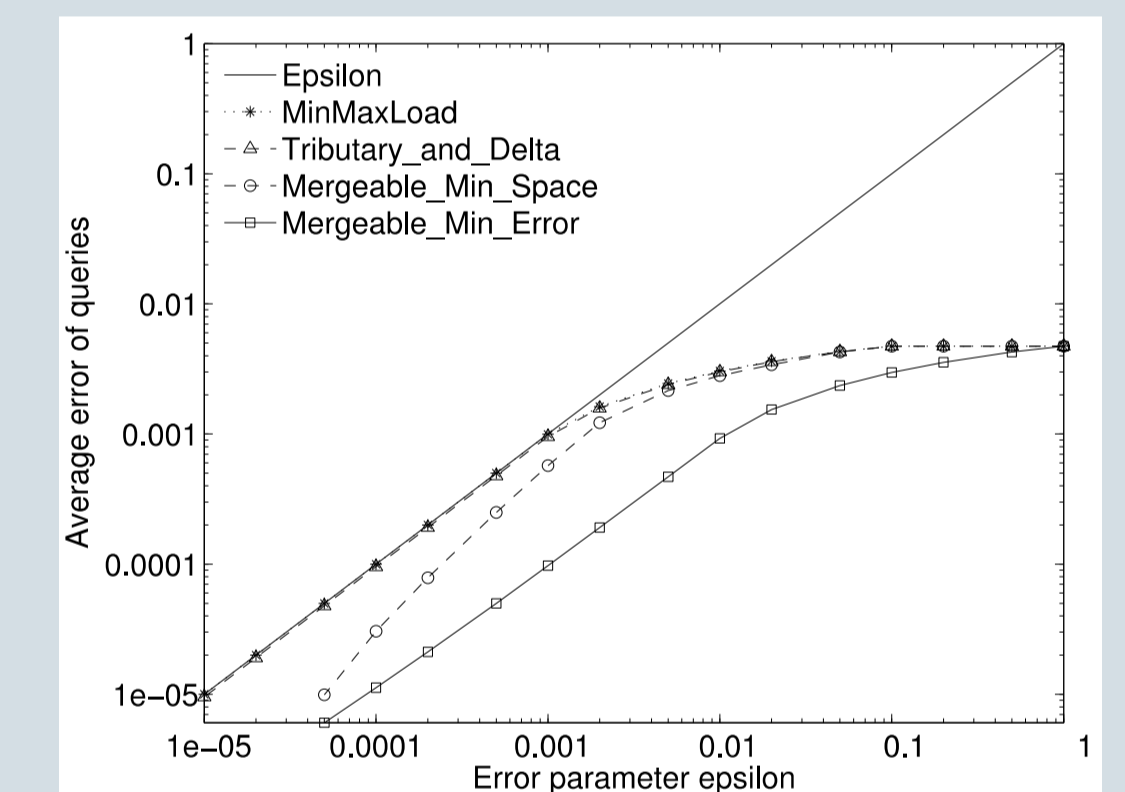
- Mergeable Heavy Hitter summaries: $O(1/\epsilon)$
- Randomized mergeable quantile summary: $O(1/\epsilon \log^{1.5} 1/\epsilon)$
- Mergeable ϵ -approximation for orthogonal rectangles: $O(1/\epsilon \log^{2d+1.5} 1/\epsilon)$
- Mergeable ϵ -kernel: $O(1/\epsilon^{(d-1)/2})$

Experiment

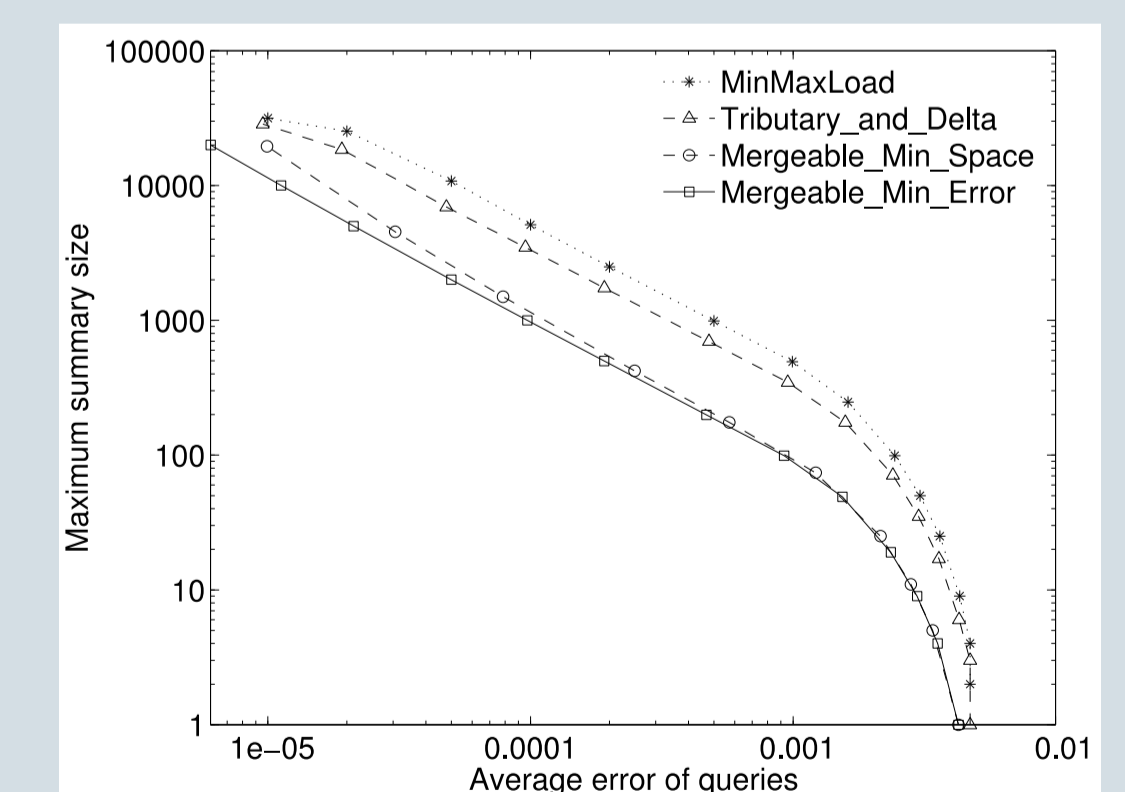
- Data set:** synthetic data following Zipf distribution
- Sensor network:** Randomly generated



Error parameter ϵ vs. summary sizes



Error parameter ϵ vs. actual error



Actual error vs. summary size

References

- [1] Pankaj K. Agarwal, Graham Cormode, Zengfeng Huang, Jeff M. Phillips, Zhewei Wei and Ke Yi. *Mergeable summaries*. ACM Transactions on Database Systems, 2013.