SJMR: Parallelizing Spatial Join with MapReduce

I. Summary

This paper and just introduces a primitive algorithm called SJMR(Spatial Join with MapReduce), without spatial indexes. The following is the detailed description of SJMR algorithm.

1) Determining the partition number. The partition number should make the partitions, which are going to be merged, fit entirely in memory).
2) Map stage. Redistribute the tuples of R and S into different Reduce tasks according to spatial partitioning function. There is trade-off between Coefficient of variation and replication overhead.
3) Reduce stage Spatial join is carried out in two steps: filter step and refinement step.
    a) Filter step:
    b) Refinement step:

II. important issues

Decomposing the universe into smaller tiles makes it easier to produce a more uniform partition distribution. However, spatial objects that span tiles from multiple partitions have to be replicated in all those partitions, thereby increasing the replication overhead.

III. contents require further exploring

- tile coding method