Communication Lower Bounds for Parallel Query Processing under Data Distribution Policies

Sampriti Roy [1]

[1] Vrije Universiteit Brussel, Belgium
Sampriti.Roy@vub.be

In distributed databases, data is stored across multiple servers, and evaluating a query often requires communication between them. The amount of communication depends not only on the query and the size of the input relations, but also on how the data is initially distributed. Existing lower bounds for one-round parallel query processing, particularly those of [1] typically assume that different input relations are stored on separate servers. In practice, however, multiple relations may be colocated on the same server. A distribution policy decides which relations are stored on the same server. Relations stored together can be processed locally before exchanging data with other servers.

In this work, we investigate how data distribution policies affect the communication lower bounds for conjunctive query evaluation in the one-round Massively Parallel Computation (MPC) model. We first identify conditions under which existing lower bounds continue to apply. Our analysis covers both skew-free databases and databases with known heavy hitters. We then study relaxations of these conditions, allowing partial colocation. Finally, we consider more general distribution policies, where several relations are stored together and processed jointly before communication. We study a component-based formulation that groups colocated relations and examines their combined contribution to query evaluation. This setting raises new challenges because a server can exploit information from multiple relations simultaneously. We investigate how this joint information affects communication lower bounds and how fractional packing techniques can be adapted to account for colocation.

References

[1] Paul Beame, Paraschos Koutris, and Dan Suciu. Skew in parallel query processing. In *Proceedings of the 33rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '14, page 212–223, New York, NY, USA, 2014. Association for Computing Machinery.