Query Optimization in Distributed Databases[1]

Ramzi A. Haraty and Roula Fany
Lebanese American University
P.O. Box 13-5053
Beirut, Lebanon
{rharaty@lau.edu.lb, rolafany@sodetel.net.lb}

The advent of telecommunication era and the constant development of hardware and network structures have encouraged the decentralization of data while increasing the need to access information online from different sites. Query optimization strategies aim to minimize the amount of data transferred across networks. This is achieved gradually by reducing the number of tuples transmitted from one site to another until it reaches the destination site.

Many techniques and algorithms have been proposed to optimize queries. Perhaps one of the more important algorithms is the AHY algorithm that is implemented by Apers, Hevner and Yao in 1983. This algorithm consisted of three phases:

1. Local processing: to filter unnecessary information.
2. Semi-join reduction: shipment of data from one site to another to be reduced.
3. Final assembly: at the destination site.

But somehow, their approach suffered from some problems due to the redundancy of the semi-joins. Nowadays, a new technique called PERF (Partially Encoded Record Filters), presented by Kenneth Ross in 1998 seems to bring some improvement over semi-joins. PERF joins are two-way semi-joins using a bit vector as their backward phase. Upon performing a semi-join between two sites, data is sent from site A to site B, and then a bit vector representing the matching tuples is sent back to site A. Whenever such a semi-join is needed again, data will not be resent from site A to site B but the bit vector is used, hence reducing the cost over the network.

Our research encompasses applying PERF joins to the AHY algorithm. A major improvement was noticed, leading us to further observations and studies.