94-MRV

Managing Network Resources for Efficient, Reliable Information Systems

Naphtali Rishe

High Performance Database Research Center
School of Computer Science, Florida International University
University Park, Miami, FL 33199; E-mail: rishen@fiu.edu

Abstract. Recent advances in public network technology and emerging database interface standards enable the implementation of a new type of distributed database systems and create demand for improved query distribution and query optimization algorithms.

Keywords: distributed databases, heterogeneous databases, query distribution, query optimization, public networks, bandwidth on demand, application manager, interface standards, ODBC, ISDN.

Introduction

Current technology offers the possibility of connecting local computer systems with remote servers via high speed telephone lines. In addition to that, direct access to Intelligent Networks allows for centralized communication management. A network coordinator node may communicate with every other node in the network, using the Intelligent Network, to serialize and optimize data flow over data lines, e.g. ISDN. Moreover, this coordinator node may establish the necessary telephone connections between nodes and manage system resources such as data line allocation. Without loss of generality, we will refer to ISDN telephone lines in our considerations, where any similar type of data line could be utilized.

These new technological possibilities are of particular interest for distributed database system implementation. High bandwidth data exchange is possible, and the central coordinator node may take responsibility for enforcing the required consistency constraints present in database environments.

A research topic that immediately arises is cost minimization for ISDN usage. Even if policies for charging telephone communications may vary, it is a fair assumption that cost will be proportional to the duration of an ISDN connection, and thus, to the volume of data exchanged. Previous research on distributed database systems has produced results that can be utilized for this purpose. Nevertheless, Intelligent Network and ISDN usage requires current knowledge to be adapted and extended.

Distributed query processing

Many authors like, [Bernstein&*al.*-81], have addressed the problem of query processing in a distributed database environment. Several papers have polished ideas such as local and remote semi-join operations [Segev-86], data duplication [Yu&*al.*-87], heuristics for query graph optimization [Shasha&Wang-91], etc. Progress has been achieved in determining optimal solutions for restricted instances of the general problem, such as in [Hevner&Yao-79], [Chiu&Ho-80], [Sugihara-83], [Chiu&*al.*-84], [Chen&Li-85], [Shasha&Wang-91]. Nevertheless, research work has often been more successful in assessing the inherent difficulties of the subject. The generalized query distribution optimization is NP-hard [Hevner-79].

The general approach when dealing with NP-hard problems is to seek heuristic methods for determining good approximate, sub-optimal, solutions. In the case of query distribution optimization, research has often determined that even simplified and restricted formulations of the generalized problem still belong to the NP class [Segev-83], [Chen&Li-89], [Shasha&Wang-91].

Unfortunately, many of the heuristics proposed for query distribution optimization have only been analyzed on severely restricted instances of the problem. Moreover, analysis of results for many of the proposed methods was frequently based on simplifying assumptions. It is not clear if the same methods will yield satisfactory performance and adequacy when implemented in commercial database environments.

The utilization of ISDN lines increases the complexity of the query distribution problem in many ways. Unlike a traditional network, data lines are established as data needs to be transferred. Bandwidth between sites varies as a function of the number of available lines at a particular moment. We believe that known heuristics should be adapted, and new algorithms developed, with the aid of simulation and prototyping tools, for faster validation of the theoretical approach.

A distributed database built on top of local database systems

In a distributed database system connected via ISDN lines, query requests may be directed to a central node for execution coordination. Nevertheless, local processing at sites should be allowed without interference of that central node, to provide a way of reducing the impact of failure or unavailability of communication lines. In this scheme, a failure of the central node would not preclude local processing. Another desirable goal would be to allow each site's administrator to select the database management system that is more appropriate for its needs.

Recent efforts in industry have provided the necessary standards for interfacing different DBMSs. The Call Level Interface, defined by the X/Open and SQL Access Group, provides the necessary flexibility and scope needed for dynamic query definition. Microsoft's Open Database Connectivity (ODBC) implements an interface paradigm illustrated in Figure 1.

A reasonable approach is to utilize the characteristics offered by ODBC to build a distributed database environment on top of local database systems. Among other advantages, this approach allows the connection of existing local database systems to a distributed environment with relatively little cost.

Figure 2 shows a possible architecture for such a distributed system. The central node runs an Application Manager Interface Program (AMIP) and each site runs instances of a Local Interface Program (LIP). The AMIP communicates with LIP instances via the Intelligent Network, and LIP instances share intermediate and final query results through ISDN lines.

Conclusion

The current state of telecommunication technology provides a potential for developing efficient distributed heterogeneous database systems, utilizing public data networks such as ISDN. Emerging technology and standards, such as ODBC, make the development of such systems cost effective and their operation flexible.

Acknowledgment

I am grateful to Leonardo Haddad Loureiro for extensive help. Our research is sponsored by NASA, DoD BMDO&ARO, DoI, NATO, Florida DoC, and private industry.

References

[Chen&Li-85] A. L. Chen and V. O. Li, "An optimal algorithm for processing distributed star queries." *IEEE Transactions on Software Engineering,* Vol. SE-11, No. 10, Oct. 1985. (pp. 1097-1107)

[Chen&Li-89] S. J. Chen and V. O. Li, "Optimizing joins in fragmented database systems on a broadcast local network." *IEEE Transactions on Software Engineering,* Vol. 15, No. 1, Jan 1989. (pp. 26-38)

[Chiu&Ho-80] D. M. Chiu and Y. C. Ho, "A method for interpreting tree queries into optimal semi-join expressions." *Proceedings of the ACM SIGMOD International Conference on Management of Data,* 1980.

[Chiu&*al.*-84] D. M. Chiu, P. A. Bernstein and Y. C. Ho, "Optimizing chain queries in a distributed database system." *SIAM Journal on Computing,* Vol. 13, No. 1, Feb. 1984 (pp. 116-134)

[Hevner-79] A. R. Hevner, "The optimization of query processing on distributed database systems.", *Ph.D. dissertation, Purdue University,* 1979.

[Hevner&Yao-79] A. R. Hevner and S. B. Yao, "Query processing in distributed databases." *IEEE Transactions on Software Engineering,* Vol. SE-5, May 1979. (pp. 177-187)

[Segev-83] A. Segev, "Optimizing 2-way joins in fragmented database systems." *Working Paper MS-2, School of Administration, Univ. of California, Berkeley,* 1983.

[Segev-86] A. Segev, "Optimization of join operations in horizontally partitioned database systems." *ACM Transactions on Database Systems,* Vol. 11, No. 1, March 1986. (pp. 48-80)

[Shasha&Wang-91] D. Shasha and T. Wang, "Optimizing equijoin queries in distributed databases where relations are hash partitioned." *ACM Transactions on Database Systems,* Vol. 16, No. 2, June 1991. (pp. 279-308)

[Sugihara-83] K. Sugihara, "Optimization algorithms for processing simple queries in star networks.", *Proceedings of the IEEE COMPSAC,* Nov. 1983. (pp. 547-554)

[Yu&*al.*-87] "Algorithms to process distributed queries in fast local networks." *IEEE Transactions on Computers,* Vol. 10, Oct. 1987. (pp. 1153-1164)

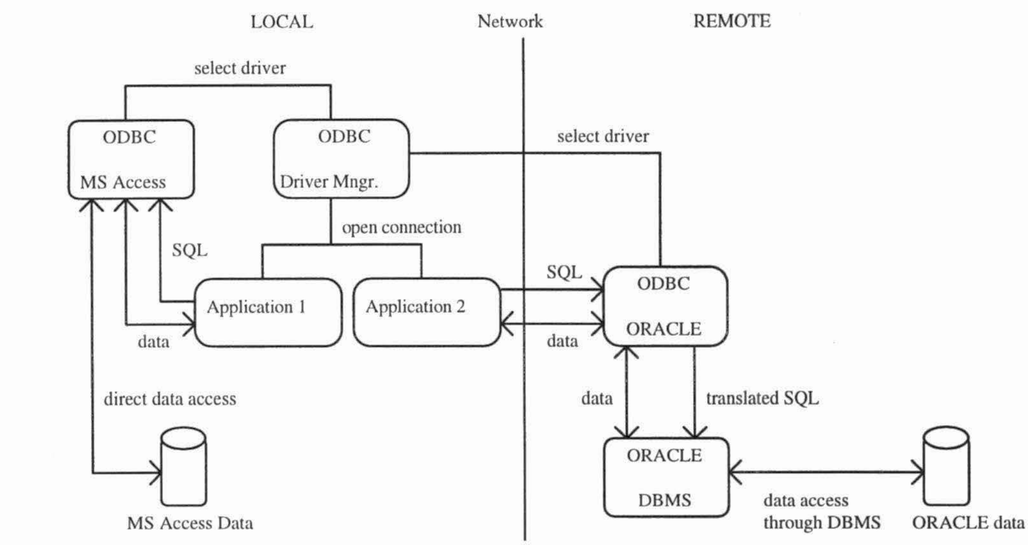

Figure 1. ODBC Paradigm

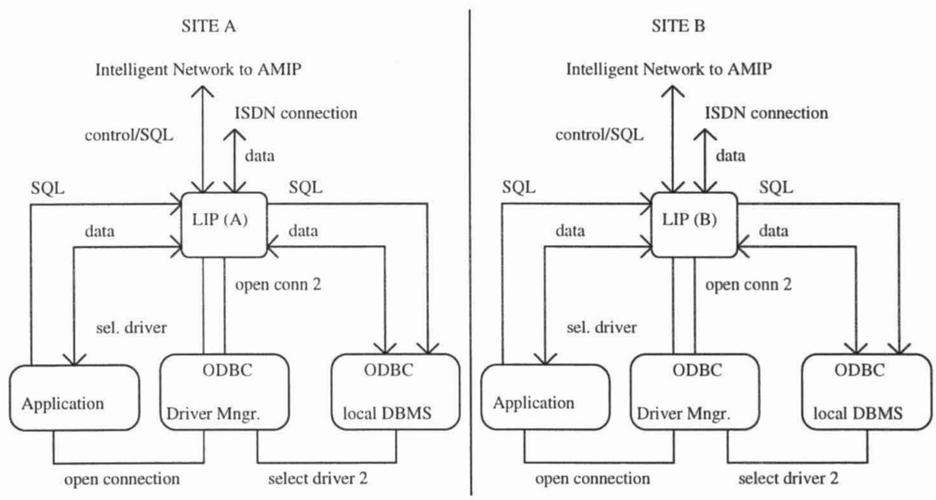

Figure 2. LIP as the ODBC driver for the distributed database. Notice that the two sites have identical LIPs and architectures.