Databases for Networks

H. V. Jagadish
AT&T Bell Labs
Murray Hill, NJ 07974
jag@research.att.com

The communications industry is on the brink of a revolution, as evidenced by almost daily newspaper articles on cable TV, wireless phones, and "information highways". Just as CAD/CASE applications have had a profound effect on the database community in the 80s, so are network applications likely to in the 90s. This tutorial explores the crucial role that databases play in wide-area communications networks.

Wide area communications networks have been around for a long time. Ad hoc application-specific solutions have been adopted for data management. With rapid changes the communications industry is currently undergoing, and exponential growth in traffic volume, such "hard-wired" solutions are no longer acceptable, and generic database software is desired. However, traditional databases do not provide all the features required in networks, as we discuss below.

There are three major layers at which it is relevant to consider database needs in a large network: network operation, network management, and network services. Database requirements differ between the three, and we consider each in turn.

Physical Network Operation

For all but the simplest networks, any connection between two end-points requires the traversal of multiple links. The actual set of links used to route the connection have to be determined. At the very least this requires a look-up of "routing tables" that take an address and determine a route based on network topology. Often, these routing decisions have to take into account dynamically changing information regarding load and capacity. The time available to make a route determination is very small, in many network environments, necessitating the use of main-memory databases with data partitioned/replicated for local access.

Availability of these databases is crucial for most networks. Availability is different from reliability, which is of great importance in applications such as banking that have been investigated by the database community hitherto. High availability requires not just fast recovery schemes and hot spares, but also techniques to perform on-line code update and on-line schema evolution.

Network Management and Administration

Information regarding network usage must be recorded, both for network control and management and for billing. Raw data is typically available at a very low level, and in overwhelming quantities. Sometimes, raw data can be collected only at a higher level, thereby reducing the quantity of information somewhat. However, in the event of an error or some other problem, it is useful to be able to look at data from the lowest level. Thus, the requirement is to be able to archive overwhelming masses of data in (soft) real time, while at the same time processing the data to create summaries and to raise exceptions as required. This implies a need for incremental view maintenance to support summaries at multiple levels of aggregation, and active database support for declarative specification of data patterns of interest.

Good network management is likely to require human analysis for a long time to come. So there is a question of how best to present data in a manner that can be absorbed. It is useful to perform visualization at multiple levels of abstraction and aggregation. Good user interfaces for such databases are sorely needed.

Network Services

Information services across a network are of growing importance. Name servers and similar directory services provide information about the network itself. More generally, access may be desired to information at unidentified locations in the network, and appropriate directories, indices, or agents used to locate the information. Digital libraries, in particular, are being hotly studied at the current time. xmosaic has become a de-facto standard network navigation tool. It will not be long before third party information "finders" become commonplace. Database issues include autonomy, heterogeneity, replication, and distributed indexing.