SOME ASPECTS OF INFORMATION NETWORKS

(Quelques Reflexion sur les Réseaux d'Information.)

T. RADHAKRISHNAN

Dept. of Computer Science, Concordia University
Sir George Williams Campus
Montreal H3G 1M8

ABSTRACT

A brief survey of the concepts in Information Networks are given and some of their differences from Computer Networks are considered. Two possible approaches to the development of large networks and their relative merits and demerits are discussed. The concept of viable basic subnet forms the starting point of development. This subnet is dependant upon other networks such as Computer Networks and Transportation Network. (Dans cet article, nous présentons les réseaux d'information, et nous tachons de les differencier des réseaux d'ordinateur. Nous discutons de deux méthodes de contruction de grands réseaux et nous comparons ces deux approaches. Nous présentons le principle du "sous réseaux relisable" comme point de depart de development du futur réseaux, et nous montrons sa dépendance vis-a-vis d'autres réseaux déja en opération (Réseaux de transport, réseaux d'ordinateur etc.).

I. CONCEPTS IN NETWORKS.

Network Ideas:

A network can be considered as a directed graph with several constraints upon it. Intuitively, each node of the graph represents a location such as, a library, and each edge is assigned a numerical label, called the capacity. An user at any node of the network will be able to direct an "information" request to any other node and receive a reply to his request. Hence, there must be at least one path between every pair of nodes in the network. In order to improve the reliability, there can be additional paths between them. There are different types of networks such as "cyclic network", "hierarchical network", "star network", and "decentralized network" (Schneider, 73). The above classification is based on the topological structure and reflects how the different nodes are connected to each other. Some of the design parameters which favour one type of network than others are flexibility, reliability, cost and management (Korfhage, 74).

Based on its primary function a network is termed as a Computer network, Communication network, Transportation network, Information network and so on. However, one class of networks may require the other for its existence. For example, in a computer network, communication facilities are required to exchange data, or program between computers at the selected nodes. The rules which govern the communication are known as protocals. (McKenzie, 73). A library network may use the capabilities of a computer network which inturn makes use of the communication facilities. This interaction between networks and the resulting problems can not be ignored in practical implementations.

Information Networks:

A network is called an Information network when the nodes are information centers which create and use "information". Since the term information is broad in its meaning, this definition may incorporate many network classes. There are management information networks which connect the various geographically distributed branches of business companies (Feeney, 72). A financial information network proposed by the Royal Bank of Canada is to connect its various branches across the country. The Swedish LIBRIS network integrates the research libraries in Sweden into a computerized library network. (LIBRIS, 74). In each of these examples, the "item of information" is different and loosely they are termed as information networks. In this paper, we restrict our consideration to library networks.

Employing networks to support library operations is not new and the concept has been around for years with inter library loan and centralized cataloging (Wait, 74). The library network can operate better with the computer and communication facilities. Two types of information retrieval can be identified in library operations, namely, "Reference retrieval" and "Document retrieval" (Walston, 65). For document retrieval, the transmission of full length documents over communication lines does not seem economical, with the present technology. Thus, some system design considerations are required in determining what information should be transferred between nodes in library information networks.

Why Information Networks?:

The network approach, in general, has number of advantages (CANUNET, 72). It acts as a forcing function to create some standardization for information exchange, as it happens with the MARC tapes. Rare information, available at one node can be accessed and used by all the users connected to the network. The level of inter library loan (ILL) has increased significantly in the last several years. The library networks will result in effective and economic sharing of resources (Becker, 74). It will prove to be more expensive to duplicate the specailized data banks created and maintained at various centers. An extensive survey of such data banks available at the various Canadian Universities, is given in (COMP/COMM, 72). With information networks such data banks can be shared by all centers.

Based on user needs, the documents in a library system can be categorised as follows:

a) Regular use: several copies are desirable

b) Frequent use: at least two copies are needed

c) Moderate use: one copy is sufficient

d) Infrequent use: can be borrowed/stored in microfilm

e) Very rare use: can be borrowed

It is felt that majority of a library budget is spent for cases (d) and (e) (D'Oliver, 75). The actual figures may differ from library to library, but in most cases about 90% of the user needs can be satisfied with 10% of the budget spent. The remaining user needs may be satisfied through ILL, whenever, possible. The most important factor is to decide when is it economical to borrow rather than buying a copy. This could be achieved with the use of computers in library automation.

II. TOOLS FOR NETWORK DESIGN:

Computing Facilities:

In view of the additional data processing involved due to ILL, computing facilities are indispensable. Most libraries of today, particularly university libraries, have access to such facilities in one form or the other. Actual requirement of these resources will depend on the extent of involvement of the particular information center. For ecample, one node in the network may not have elaborate data banks to be shared and most of the time be accessing information at other nodes. Then a mini-computing facility might be adaquate for all its needs. On the contrary, a large information center may have different data banks, types of query languages and many kinds of user services, etc.

Communication Facilities:

Two types of communication facilities are identified, namely, computer communication and transport communication. Details of the various communication techniques for computer networks, such as the Newhall loop and Packet switching etc, can be found in (Winkler,72). The important factors which govern the choice are cost of facilities, error control, line utlization , network response, the impact on buffering and software. In the case of document retrieval, after looking at the bibliographical reference, an user asks for the original documents in order to study them. The chief means of forwarding replies, is still by post . In many countries mail travels quickly, is uninterrupted and postage is cheap. When this method is not adequate, a suitable transport communication is necessary. This is further discussed in section IV.

User Terminals:

This forms an interface between users and the system. A simple to understand query language will enable better user interaction. Soft copy display using CRT terminals will be suitable for interactive retrieval of documents (Borko, 66). However, the hard copy output facility can not be ignored. The terminal facilities will depend on the involvement of the centre in information exchange and its budget.

Document Storage:

Each node in the network holds the frequently required documents in hard copy form. This could be printed copies or in the form of microfilms or micro-fische. Recent developments in COM, computer output microfilms, promote the automatic selection and display of such stored documents to the users. Microfilm storage and acquisition is suitable for documents of the category (d) mentioned in the last section.

Distributed Data Bases:

These are of interest to a central body. The central body may be the government, a company having branches at different nodes or a central agency interested in the information at different centers. Unlike the shared data banks, in this case, information is created, modified and used at different centers. The development of distributed data bases is involved with the Information Networks (Booth, 72). Some of their problems such as segmenting of files, access rights, protection, and file transmission, have been studied in the literature with computer Networks.

III. NETWORK DEVELOPMENT.

Top-Down Approach:

By top-down approach, it is meant that the central planning is done, policy decisions are made, a priori, and development of nodes are followed subsequently. The proposed development of the Canadian University Computer Network (CANUNET) is an example for this approach. This approach has certain advantages. A central governing body can monitor the balanced growth, impose certain standards, and plan the entire spectrum of development. Most programming projects with computers follow this top-down approach. However, there is a high development cost which is unavoidable in this case. The cost of the proposed CANUNET has been estimated to be between 4.2 and 9.5 million dollars, and the time required to be between 4.5 and 7 years.

Bottom Up Approach:

This may also be termed as the evolutionary approach. Following the pattern of inter library loans, information sharing between centers will be encouraged. An availability of computer communication between members will help this growth. For example, consider the growth of the MERIT Computer Network (Schneider, 73). This is a cooperative effort between Michigan State University, University of Michigan and Wayne State University. Such local developments evolving on the current user needs can later form subnet-works of a larger Regional or National Networks. The problem of standardization is tackled in a local scale. In this evolutionary approach, the centers exchange or share information, when it is economical to do so. When a number of such local subnetworks are being developed, there exists a possibility of duplication of effort. Combining these local subnetworks to form a Regional network might become more difficult than in the top-down approach.

There is no reason why a good combination of these two approaches be adopted. A number of local networks of "viable size" will be initially encouraged. For instance, among the university libraries in Montreal, the Information Science library at the University of Montreal, contains more periodicals, serials, technical reports than other libraries, in the area of Computers and Information Sciences. Also there exists a good computer communication facility among the universities in this city. The METANET project to interconnect computers at various, Ontario universities is a similar approach in the area of Computer Networks. When the growth of such local information networks achieve certain level, the top-down approch may take over to form Regional and National Networks. This combined approach is not entirely new. The expanded research program of the National Science Foundation to develop networks in different subject areas have been summarized in (Winkler, 72). In fact, the subnetworks organization can be done in more than one way as described below:-

a) Decentralization by function such as public libraries, university libraries, etc.

b) Decentralization by scientific sector, etc.

c) Decentralization by geographical regions.

In the bottom-up approach the growth of subnetworks is based on the needs and distinct classification as above, many not exist.

IV. SUB NETWORKS

Basic Subnet:

In the bottom approach, the smallest subnetwork (subnet) that evolves in the beginning will be called "Basic Subnet". The number of nodes in the basic subnet will depend on the mutual interest, need for exchange, technical feasibility, and economics of sharing between nodes. Whenever, possible, development of basic subnets should be encouraged to promote the evolutionary growth.

Viable Subnet:

Following the reference retrieval, there is need to obtain the full length document for user reading. There is more than one way of satisfying such user needs.

a) The required document may be transmitted over the communication line.

b) It may be sent to the borrower by post.

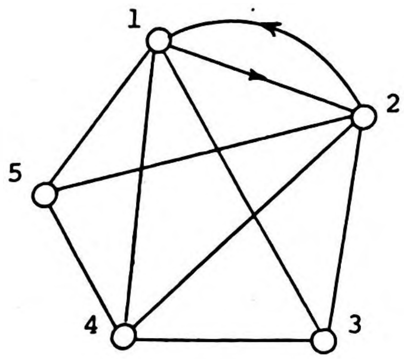
c) A special courier or shuttle service may be employed.

d) A copy of the required document in printed form or microfilm form may be acquired by the borrower.

When the postal service is not dependable, the options (c) and (d) seem attractive. Functioning of the option (c) is based on the available Transportation Network. Consider the hypothetical transportation network shown is Figure 1. It is essentially a directed graph with possible parallel edges between nodes.Each edge represents a route between two centers of the netwrok and the nodes stand for the information centers. A pair of weights, (TIME, COST) are associated with each edge that denote the time and cost of travel along that edge. In actual case these parameters are random variables with finite mean and standard deviation. Further, they may not be constant during a day. The following discussions do not account for such random behavior and time dependent variations.

Suppose the shuttle transport goes round the closed path 1-2-3-4-5 collecting documents at one node and delivering them at other nodes. Each round takes 47 units of time and costs 46 units, as per the values of Table I. This is essentially a weighted Hamiltonian circuit in the network. For economic reasons one may look for the circuit of least time or least cost. When the cost and time with different edges are random variables, their peaks may occur at different time of the day. Thus, the most desirable route will also depend upon the time of day.

The worst case waiting time, for the request just missed the shuttle, will be twice the time for each round. Let this be denoted as Δ . For a typical case, the Maximum acceptable waiting time to get the copy of a full length document, after the reference retrieval, is shown as a function of the number of requests in Figure 2. The inverse relation between them is intuitive but the exact functional relation will depend on the library service, user groups and the nature of requests. For a given value of Δ the level of user satisfaction can be obtained from the Figure 2. By minimizing Δ sufficiently, all user demands may be satisfied. This may be achieved by improving the mode of transport like the postal network in Europe or by running parallel shuttles, and so forth. Yet another way of minimizing Δ, is to use the faster telephone network and decide



Example:

1. Sir George Williams Campus
2. McGill University
3. University of Montreal
4. Loyala Campus
5. University of Quebec at Mont

Figure 1. A Hypothetical Basic Subnet

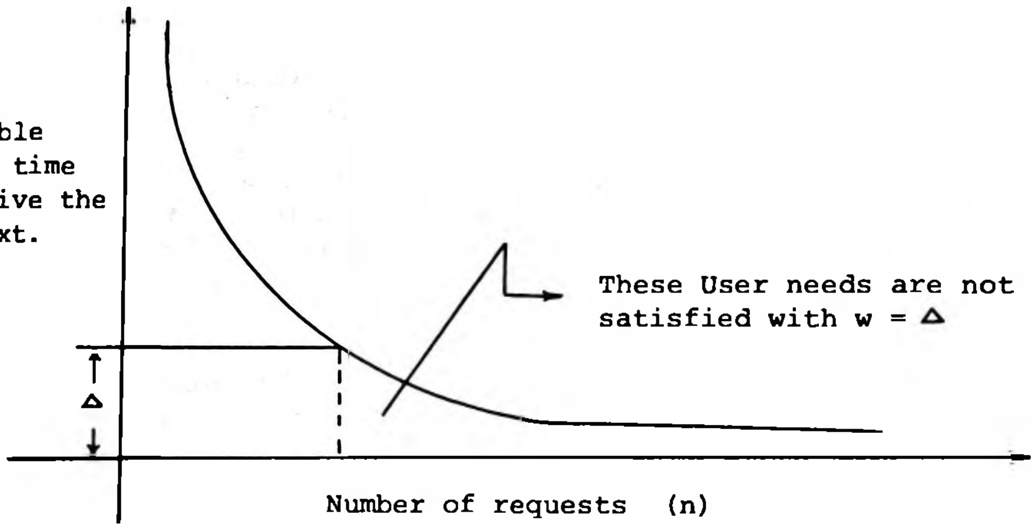


Figure 2. Acceptable Waiting Time Distribution

TABLE I

Weights of the Edges in Figure 1.

Edge	Time	Cost
1 → 2	3	3
1 - 3	10	12
1 - 4	6	5
1 - 5	8	9
2 → 1	2	3
2 - 3	15	17
2 - 4	8	8
2 - 5	4	4
3 - 4	9	7
4 - 5	12	10

which are the nodes to be visited in the current shuttle. In this way, a shuttle may avoid the nodes which do not involve any transaction. Minimizing of Δ implies the increased cost. For the given set of parameters and the trade off limits, there seems to be an optiomal condition.

Time Shared Terminals:

The on line time sharing of a central data bank is distinguished from the network nodes as follows: In the former case the information is at a central place and the sharing users do not contribute, explicitly, to the data bank. With the networks, the information required may be distributed at more than one place. The operation of CAN/OLE is an example for the shared data bank operation. A number of library networks have been planned or operated such as MEDLINE, BASIS-70, ISIS, OCLC, NELINET, LIBRIS and CAPTAIN (Kilgour, 73).

A brief overview of Information Networks and a methodology for their development have been presented. The bottom-up approach starting from a set of viable basic subnets seems to be preferable. As there are not many Nation wide Library Information Networks operational, this approach may provide opportunities to use the experience and newer technology, as they arise, in implementing the top levels in the hierarchical National Information Networks.

ACKNOWLEDGEMENT.

The author gratefully acknowledges the encouragements shown by Prof. H.S. Heaps, of the Computer Science Department at Concordia University. Thanks are due also to the graduate students Mr. Raymond Kernizan and Mr. Guy Agoston who work in this direction and discuss with the author. Their discussion provided many fruitful thoughts.

REFERENCES

Becker, K. and Hays, R.M., 1974 A System for Inter-Library Communications (SILC). NETWORK, 1(10) 11-22.

Booth, G.M., 1972 The Use of Distributed Data Bases in Information Networks. pp 371-376 in S. Winkler, ed., Computer Communications The First International Conference, Washington, D.C.

Borko, H., and Burnbaugh, H.P., 1966 Interactive Displays for Document Retrieval. Information Display, 3. 47-90.

CANUNET. 1972 A proposal for a Canadian University Computer Network. By CANUNET Advisory Committee for the Department of Communications Project by University of Quebec.

COMP COMM. 1973 Computer Communications Activities at Canadian Universities by Task force. Govt. Document. Background papers 2(3). 1-32.

D'Glier, J.H. and Delmas, B. Planning National Infrastructures for Documentation, 1975 Libraries and Archives. The Unesco Press, Paris.

Feeney, G.J. 1972 Information Networks pp. 16.4-1 in The Proceedings of the First USA-Japan Computer Conference.

Kilgour, F.G. 1973 Library Networks pp. 38-41 in Networks and Disciplines. Proceedings of the EDUCOM Fall Conference, Ann Arbor , Michigan. Oct. 1972.

Korfhage, R.R. 1974 Discrete Computational Structure Academic Press. New York.

LIBRIS. 1974 The Swdish LIBRIS Network. Network. 1(8). 9-14.

McKenzie, A.M. 1974 Some Computer network-interconnection issues. Proceedings of the National Computer Conference, 43. 857-859.

Schneider, M. 1973 A Survey on Computer Networks. Technical Report No. 177 University of Wisconsin.

Wait, D.P. 1974 Basic Design Considerations for Library Networks Network, 1(10). 15-30.

Walston, C.E. 1965 Information Retrieval. pp. 1-30 in F.L.ALT and M. Rubinoff, ed., Advances in Computers Volume 6. Academic Press. New York.

Winkler, S(ed) 1972 Computer Communication - Impacts and Implications The First International Conference, Washington, D.C.