COMPUTER NETWORKS

IN EDUCATION

CONTROL DATA
CORPORATION

Preface

The pages which follow have been written for a broadly diversified audience. It is unlikely that any single reader will have occasion to use them all. On the other hand, it is to be hoped that each of the various sections will be of value to one or more segments of the equally various groups who become concerned with educational computer networks.

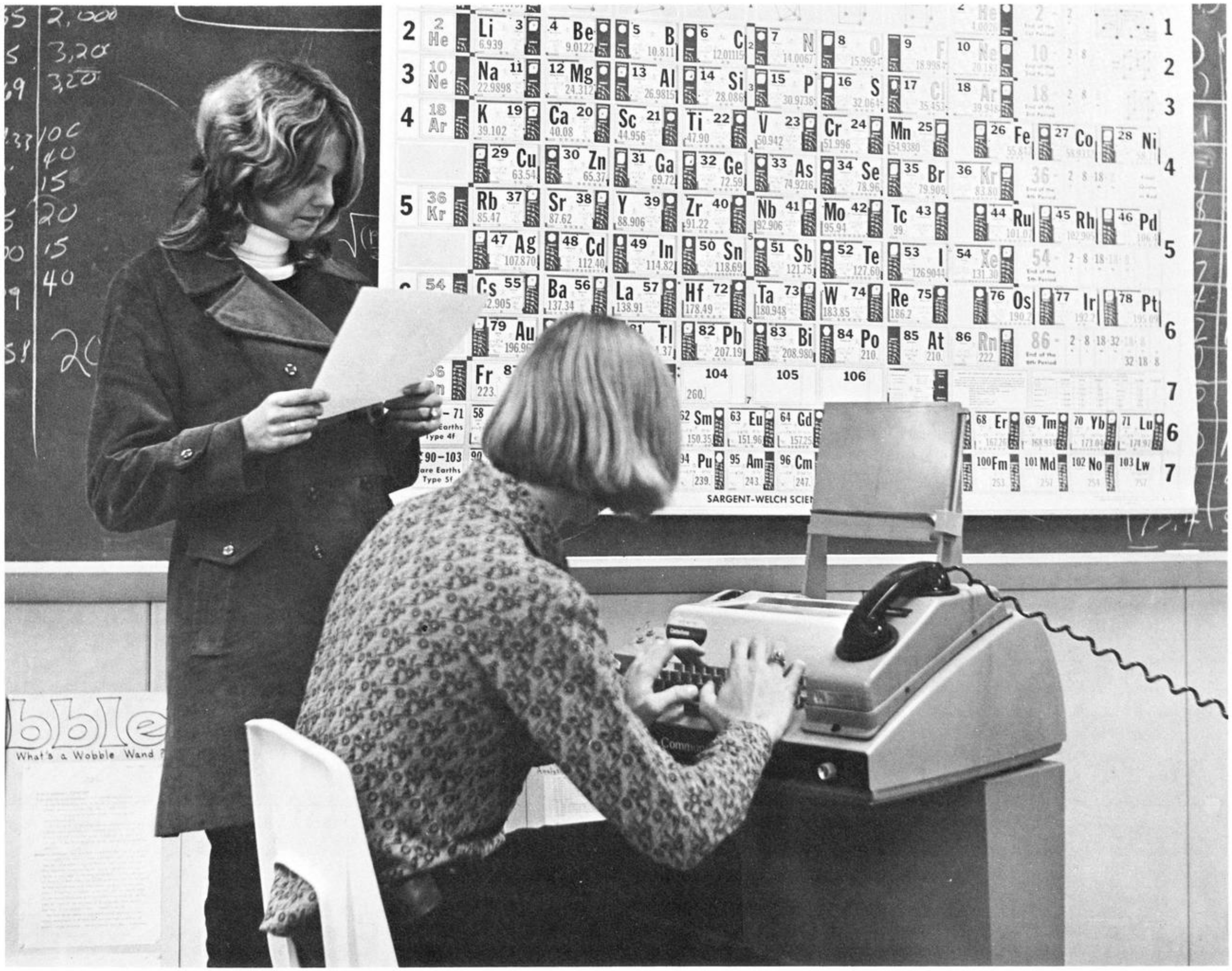
For the university administrator or state legislator who is not a data-processing specialist but with whom rest decisions, often the final decisions concerning establishment of educational computer networks, the opening sections can prove valuable. Here he or she may become familiar with various users and types of processing handled by such networks; with terms used in discussing such networks; and with the principles of distributed processing which underlie network-establishment.

For the already competent specialist in the use or in the analysis/design of such networks, it will be later sections which can prove useful, especially the technical and financial considerations offered in Sections VII and VIII. For this material and permission to reprint it, we are gratefully indebted to the Office of Computer Coordination of the Council of Ontario Universities, Toronto, Ontario, Canada. Any further quotation or reprinting of this material should carry proper and sufficient acknowledgement and be submitted to the Office at 102 Bloor Street West, Toronto 181.

•

TABLE OF CONTENTS

COMPUTER NETWORKS IN EDUCATION



I BACKGROUND

Computing in the late 1950's was a new technology being introduced to the education community by a small number of people. Through their work, horizons gradually broadened and computer technology became established as a powerful educational tool with great potential.

During the early years of educational computing, it was not recognized that this was the beginning of a potentially exponential growth in computer use. As with many innovations, people were encouraged to make use of computers through the provision of a subsidized resource which was often funded, in whole or in part, by national grants.

During the last ten years, educational institutions have developed their own computing facilities as rapidly as possible in order to serve their individual needs. They have recognized the need to acquaint their students with computers and to teach them to apply the computer in their science and business courses. They have provided computers and computer-related courses in their curricula and have satisfied this need to varying degrees. The problem has been to make their computer facilities available or accessible to all students requiring their use for class work or research projects.

For administrative data processing, many educational institutions have had separate dedicated computer systems. Almost all of these administrative applications systems were developed independently on an as-able basis and many of them are no longer adequate to handle the volume of data resulting from the growth of the institutions. Also, some of the older systems do not provide such facilities as direct-access storage and interactive communications which are needed to collect and analyze timely data for running a modern educational institution and to develop the on-line data-processing applications now being contemplated.

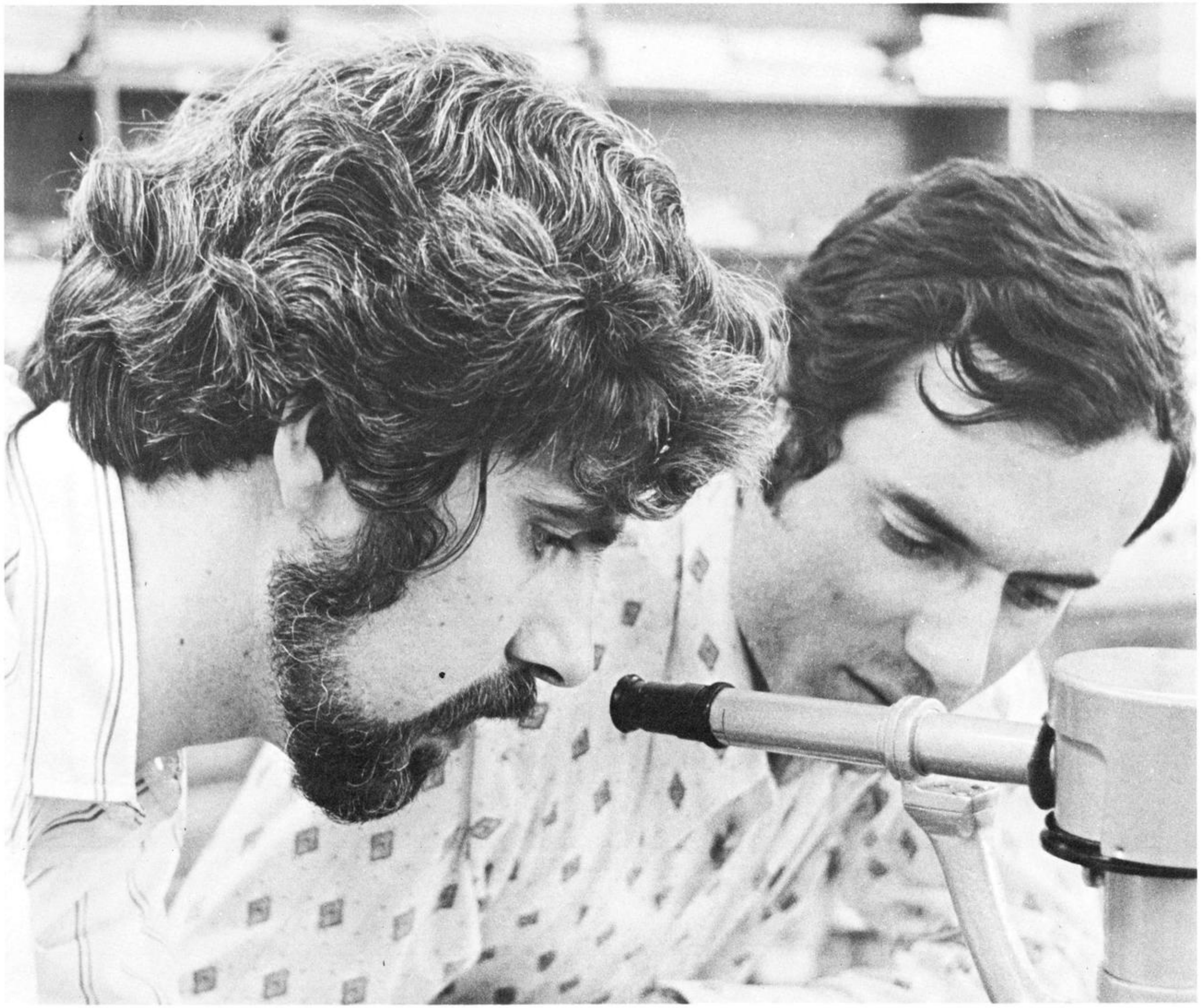
In the past, computing requirements were relatively simple; batch-processing techniques were sufficient to fill the needs of departments that desired processing service. The environment was well-structured and additional needs were met by unstructured growth in computing capability without reasonable attention to long-range goals.

The seventies will see an increasing number of students completing their secondary education and going on to further their education. A baccalaureate degree may no longer be sufficient education to acquire an entry into commerce and industry. The implications of this are clear — a larger student-age population and a longer duration in educational pursuits.

Junior colleges and state colleges will grow in enrollments and new ones will develop. Present large state institutions offering doctoral degrees will not grow in enrollment but will instead cut down on undergraduate enrollments and significantly increase graduate enrollments. This will cause a shift toward a system of local and state schools that will feed students to the larger graduate centers. In addition, the junior colleges and state colleges will require expanded curriculums so students may progress from one school to the next without loss of credits and have terminal employment capability.

The data processing situation described previously is now reversed; the problem is that of fulfilling the unstructured computing needs of many departments with the absolute necessity for structured growth to control costs.

The evolution of separate administrative and research-instructional data-processing centers within individual universities has produced cost inefficiencies and inequitable distribution of resources. In the final analysis, individual universities and colleges suffer the consequences of the excess costs simply because their funds are not buying an optimum level of computer power. They are, however, in an excellent position to improve the computational power and sophistication available to them at reasonable cost increments by implementing coordinated plans to share resources.



II THE USER COMMUNITY

In college and university communities, with their shared objectives for acquiring, preserving, and disseminating knowledge, data-processing communities have emerged with shared concerns in the use of computers to implement achievement of those objectives. A major reason for this emergence is that the needs of any individual for data processing cannot be served independently from the needs of others.

Processing activities of the user community fall under four major headings: A) Instruction and research in the use of computers; B) Applying the computers and learning obtained under A) to students' classroom problems and to research which can be expedited by application of data-processing techniques; C) Use of computers by instructors as an aid to teaching and to the management of coursework; and D) Administrative data processing. These headings break down further into the following categories:

1. *Introductory instruction for students who require only a basic comprehension of computing.*

Survey courses are intended to give students a proper appreciation of the benefits and dangers of the spreading use of computers in business, government and education. Introductory courses are designed to develop basic skills of the student in using the computer. This instruction activity is often expressed in computer usage as "student jobs." These student jobs are class assignment problems and typically are processed from batch terminals or local card readers. They often deal with a subset of a total given capability. For example, the FORTRAN and COBOL languages at the full level of implementation (American National Standards Institute) are too large and complicated for effective beginning-student use; so subsets of these have been implemented as separate packages, with emphasis on fast processing and meaningful error diagnostics for quick debugging and resubmission of the job until it is successfully executed.

These types of compilers are made available for student use in local and remote batch mode and are also supported by typewriter-keyboard terminals used in a timesharing mode. In this mode, BASIC and APL (A Programming Language) are effective. BASIC may be used where the primary emphasis is on learning to use a computer. APL, on the other hand, goes beyond learning to use the computer and is widely recognized as having great interactive and problem-solving capabilities.

2. *Computing instruction that will permit students to use the computer in their academic work and upon graduation as a part of their professional skill.*

Service courses are designed to train people from various disciplines in the latest computer techniques so they may effectively use the computer in their work. This category of instructional use is presently undergoing perhaps the greatest growth of all categories.

Programming languages such as FORTRAN and COBOL are used extensively, at their full implementation. Instruction is also provided in the use of discipline-oriented application packages.

The importance of the application package is very great. This is especially true for research and institutional applications. A social science package, for example, would handle census data, political or economic surveys, as well as statistical programs to enable the user to obtain averages, variances, correlations and regressions. This means that the student trained in its use will be able to concentrate on the social science rather than on the drudgery of data insertion and statistical calculation. He will be able to do original research with bona-fide statistical variables and data taken from non-trivial, real-world situations.

The use of application packages gives the student insight into many phenomena and also permits him to design and study structures of all types. Packages of this kind take years to develop and require experts knowledgeable in the relevant disciplines to maintain them. Such packages cannot easily be transferred from one computer even to another large one. Even the largest centers do not have the staff or the space to store and maintain all the packages of interest to students and their faculties, even if they could be transferred.

3. *Computing for student and faculty research projects (Development).*

This activity is generally composed of work related to theses. The prime concern is the use of the computer for processing data concerned with particular academic disciplines.

In the past, much of the work centered on statistical computation, engineering design and general mathematics. Newer work is now being done and includes:

- Studies in the humanities
- Chemical hypotheses formulation
- Linguistics
- History
- Music
- Artificial Intelligence

Research activities often generate a group of programs oriented toward a particular discipline; e.g., Social Sciences, SPSS (Statistical Package for Social Sciences). Generation of such programs is often the work of the principal investigators of *early projects.*

When developed, they are generally made available at little or no cost to students and other researchers at other schools through user program libraries.

4. *Use of the Computer as a Teaching Device.*

The principal activities in this category are Computer-Managed Instruction (CMI) and Computer-Assisted Instruction (CAI).

Computer-Managed Instruction (CMI)

In Computer-Managed Instruction the actual process of instruction is performed not by the computer but by teachers or conventional instructional aids. The computer is used primarily for test-scoring and record-keeping.

Each student progresses at his own pace through the use of textbooks, supplementary readings, videotapes, films, records and individual conferences with his instructor. Although the teacher is available, an emphasis is placed on the individual student's use of instructional materials on his own initiative.

Computer-Managed Instruction promises several major benefits. And since the computer is used only for examination and prescription, the costs of computer support are not great.

Computer-Assisted Instruction (CAI)

In CAI, the computer plays an important role in the instructional process and serves as a mechanism for problem solution as well as a medium to transfer knowledge. So far, however, CAI techniques are not widely used. Cost effectiveness has not been attained and course material is scarce.

It is envisioned that the use of CAI will become widespread as more cost-effective systems become available and the development of course material accelerates.

CAI is taking three approaches to producing better user costs:

- Time-rental on a large computer, with low-cost communications and special terminals.
- Ancillary CAI with more standard communications and terminals on a system acquired primarily for other reasons.
- Mini-based systems wherein course content and the number of terminals are limited. Costs are reduced by use of a small computer as well as by the use of standard terminals.

5. *Computer Science.*

This activity is — on many campuses — separate and distinct from other research activities and often has full department status. The newly emerged Computer Science department is concerned with instruction about and research on computers well beyond that discussed in 1) and 2) above. Service Courses in Computer Science require convenient and reliable access to non-disruptive, computational service. This need is usually met by the service available from the campus center. To satisfy the need for professional courses, students should be able to modify and design software at all levels and have "hands-on" experience with hardware.

The history of Computer Science on campus paralleled for a long time significant developments as shown in Figure 1. The emergence of separate Computer Science Departments has created the demand for Computer Science laboratories which ideally would be equipped with hardware having overall characteristics that are as conducive to experiments in system architecture and in data compression.

6. *Administrative Processing.*

Administrative processing activity relates to the following broad activities: personnel, student records, curriculum and facilities. In the past, administrative activities, when implemented on a computer, often resulted in using the clerical abilities of the computer to perform faster and more efficiently the tasks that were previously done manually.

Until recently, there have been very few cooperative efforts that might help to reduce the costs of developing effective

HISTORY AND EVOLUTION OF MAJOR INTEREST AREAS

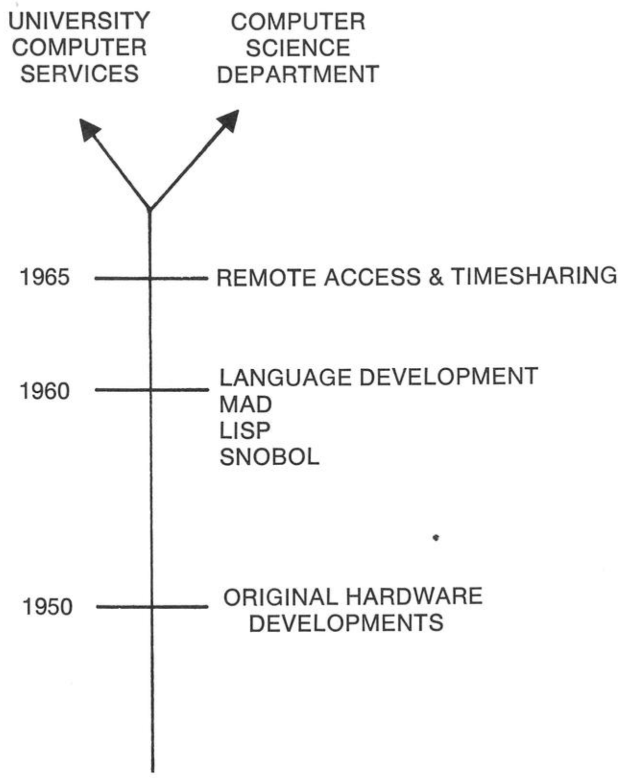


FIGURE 1.

administrative computer application programs. Also, administrative methods have differed among colleges — even within the same educational system — to the point where major changes in practices were necessary to permit adopting common processing policies. In addition, vendors were unable to supply the standardized data-management software required for development of advanced on-line management systems. Now, the efforts of WICHE/NCHEMS* are serving to standardize many of the file formats, and vendors have, off-the-shelf, the required data-management software to enhance any cooperative efforts toward developing on-line administrative systems.

*Western Interstate Commission for Higher Education/National Center for Higher Education Management Systems.

7. *Libraries.*

Conceptually, there are two separate and distinct parts to the library activity. The first is the area of circulation control and the second area is cataloging and searching. Great economies may be realized from implementation of these two problems in a computer-based system. A prime example of progress to date is the Ohio College Library Center System.

III NETWORKS

Computer technology has expanded rapidly and expenditures on computing services in educational institutions have expanded in concert. The use of computers is still developing and as their use further expands, there will be increased competition for funds which cannot be expected to grow at the rate experienced during the 1960's.

Thus there are great pressures building for a better utilization of computing resources and for exploiting the economies of scale that exist in the computer field. All of this should be and can be achieved without sacrificing quality and with assurance that users will get the services they need irrespective of their location.

Integration of the two technologies of computers and communications has now reached the point where computers of varying sizes have been linked together by communications lines of varying speeds, forming computer networks. Via remote terminals, users may have access to one computer which can process the job or can forward it to another computer in the network if the job requirements so dictate. It will soon become impossible to think of a major computer system which is not part of a combined computer/communications network.

Traditionally, computer/communications networks have been classified according to their topology; however, to understand the ways computers are used in networks, four broad methods of application processing are identified. This imposes a degree of commonality on the elements that may make up a given network.

1. *Utility Method* — computer applications and data are dynamically transferred among processing systems in a media-linked network.

Figure 1 illustrates the concept of transferring work between systems:

Utility Example: A user may wish to add capacity to relieve overflow from his on-site system. The load splitting will vary dynamically in situations that:

a. Relieve temporary overflow conditions which may arise from time to time; or

b. Provide temporary backup for systems within the network.

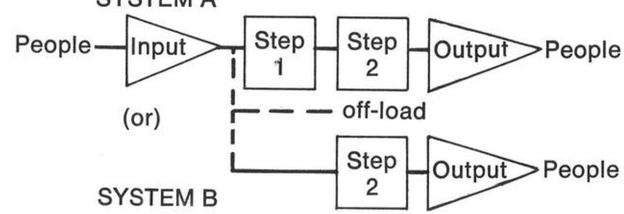


Figure 1

2. *Task Method* — a single application where steps in the processing sequence are pre-assigned to separate processors with special characteristics suitable to the steps assigned to them.

Figure 2 shows a typical application or sequential process, spanning two or more systems:

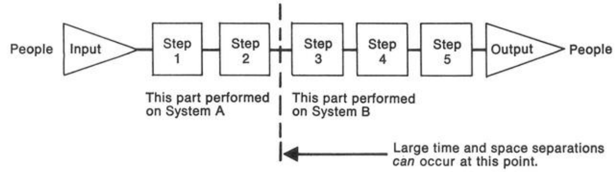


Figure 2.

Task Example: An example would be a student-record-keeping application where the on-site processor accepts transactions, performs editing and preliminary processing and forwards them to a centralized processor for updating, consolidation and summary reporting. In this example, the task workload may be systematically distributed so that the editing and verifying of the transaction are performed on-site while the central file update and reporting tasks are performed on the centralized computer(s). In this case, tasks are distributed for reasons of reliability and efficiency.

3. *Functional Method* – separate applications on each system in a physically linked network. Figure 3 shows that the systems are physically linked, as when a computer is used as a terminal to another computer, but no processing is performed by the terminal computer:

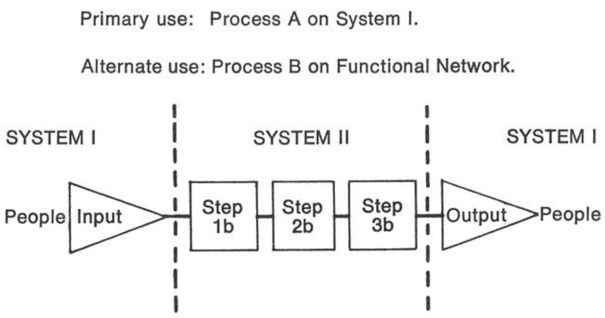


Figure 3

Functional Example: Users often have a requirement for specialized processing. An on-site, clerically-oriented or record-keeping computer can be used as a terminal providing access to a service center for special applications.

4. *Information Service Method* – a user inquiry or transaction is routed from point of origin to whatever computer in a network stores a desired data base. Figure 4 indicates that the method is a composite of functional (communications and media-transformation services without processing) and task (distributed processing of a single job/request) methods.

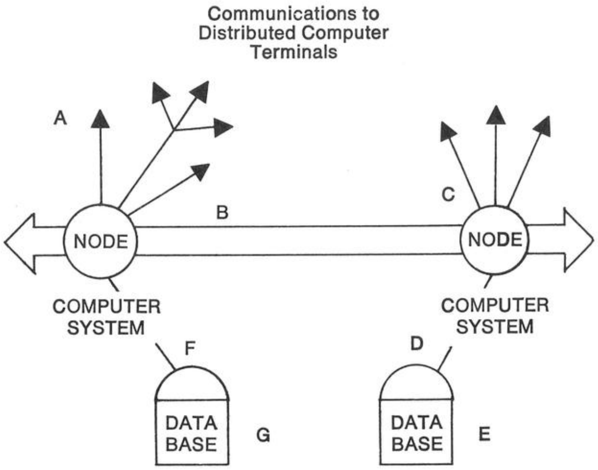


Figure 4

Information Service Example: Requests from Terminal A can probe data base E by the path A-B-C-D-E or Data Base G by the path A-B-F-G. Although primary activity is within a region (ABFG) there is provision for access between regions.

COMPUTER NETWORK CONFIGURATIONS

Regional (Radial) System

A radial computing system is one which may serve a number of terminal users directly or through multiplexer and/or concentrators but is not itself connected to a computing system or a network of systems. Figure 5 is a diagram of such a system.

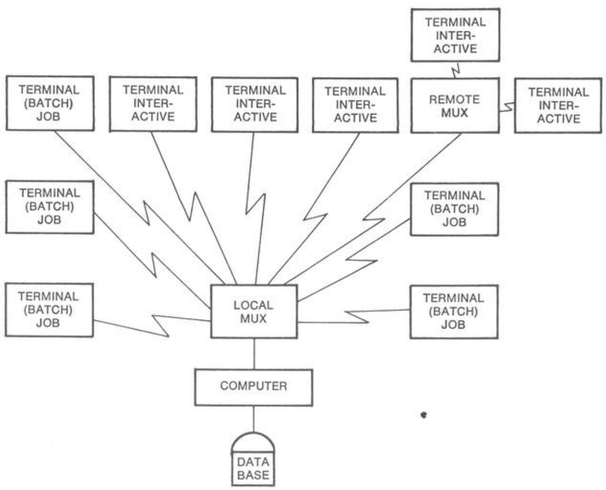


Figure 5 RADIAL SYSTEM

Each person at a terminal may submit both his job and data for processing by the computer system. If the user has a large amount of data, the terminal user may desire to permanently store it on the data base located at the computer system site. He then may merely submit his job to generate, modify or retrieve the data.

If the person at the terminal has a frequently used job, he may desire to store it along with his data file on the data base rather than continually transmit it across the communications line. He may then use a less expensive, slower-speed interactive terminal to update both his jobs and data files before submitting his jobs for execution. After his job has been executed (and filed), he may retrieve the results at will from his data file or job output file. Interactive terminals are especially useful whenever a small amount of data must be transmitted between the terminal user and computer system.

If many interactive terminals are located close together and it is desired to reduce communication-line costs, it may be worthwhile considering the use of a remote multiplexer. The remote multiplexer interfaces many terminals on tributary lines but requires only one trunk line connected to the distant computer system.

BACKUP SYSTEM

A backup system is one which may be used whenever a primary system is not available.

Each person at a terminal may have several jobs that must be processed within a specific time frame. The primary computer system, however, may not be available for two reasons:

a. Lack of resources
b. System outage

Computer systems are configured with finite amounts of resources. There are limited amounts of memory, processing power, storage and terminal ports available at any instant.

If a job requires more resources than currently available, he must either wait until they become available or segment his job in such a way that it is able to operate within the constraint of the available resources.

It is possible to overcome this situation in a more timely manner. The terminal user may select to use a backup system which has the required resources. He may do this in one of two ways. He may connect his terminal directly to a backup system via a communication line or may request his primary computer system to divert his job and data to the backup system. In the second situation the primary computer system appears as a batch-job terminal to the backup system.

In a situation where the primary computer system is not operational, or will not have the required resources available until after an extended period of time, the user must connect his terminal directly to the backup system. He must transmit both his job and data file across the communication line to the backup system or he may anticipate the situation by electing to store his job and data permanently on his backup system as well as his primary system.

The use of a backup system thus ensures the processing of jobs in a timely manner but requires several operational steps on the part of the user.

NETWORK (GENERALIZED)

A network is the interconnection of two or more nodes in order to pass data or information between them. The terminals and computer systems are connected to the nodes and are thereby able to communicate with each other.

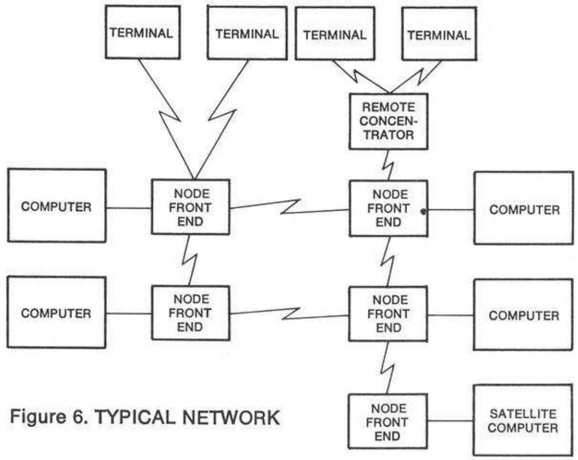


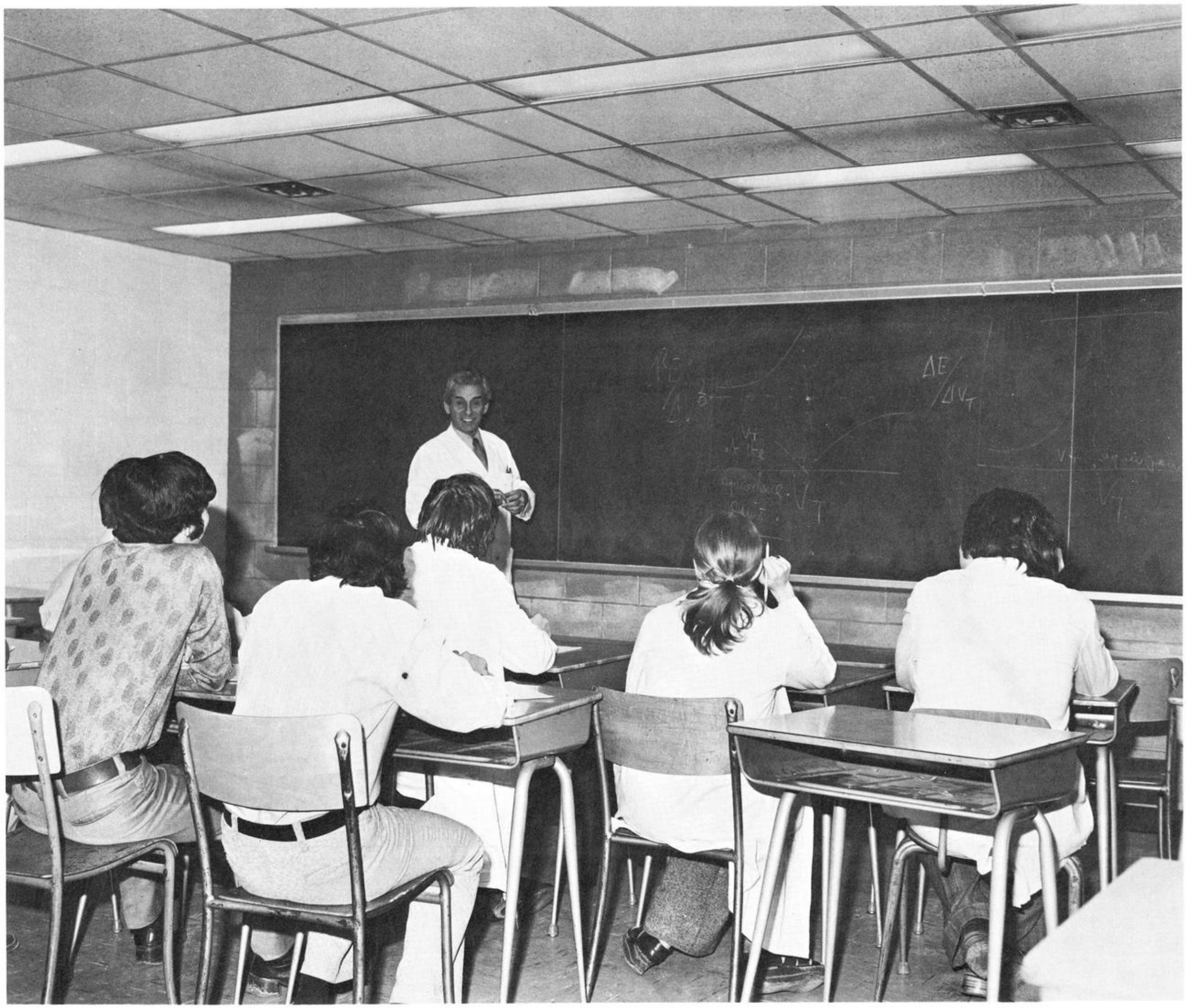
Figure 6. TYPICAL NETWORK

EDUCATIONAL NETWORKS

Ideally, an educational computing network is conceived as a community of computers sharing calculation power, data bases and application packages. It would link together computer resources so that the user has access to the proper resources for performing his work at maximum efficiency.

The network will overlay current in-place systems and provide a service by coupling these systems. Users of this service will have certain minimum requirements. Briefly, in approximate order of significance, these requirements are:

- *Availability* – although this attribute is inversely related to our growing sophistication, a high degree of processing availablity is an absolute requirement for a successful computer network.
- *Super-computing* – a powerful processing resource implies one or more large-scale systems with a large amount of core memory, on-line data storage and high-speed computational capability.
- *Responsiveness* – the system must have the capacity to react to a user's request in the time frame the user expects.
- *Ease of Use* – the characteristics of the user's interface are many: simplicity of job and terminal control languages, applications-oriented terminal, tutorial software that helps the user as he uses it, discipline or applications-oriented languages.
- *Data Base* – computers may be considered to be information-centered or computation-centered. Today's information-centered software lags the computation-oriented software in development yet the growth potential for the computer as an information-centered resource is much greater than as a computation-centered resource. In any case, adequate storage and adequate accessibility of its contents for a fully effective data base is one of the greatest needs in network services. This is especially true for the education network.
- *Compatibility* – although several unlike systems may exist within the computer network, files and user interfaces should be compatible.
- *Uniformity/Consistency* – although a user can access the network via remote batch, time-sharing or transactions, he should not be required to maintain separate copies of his data base. Further, system commands and software tools should be uniform across the three modes of access, and consistent over time. It should be possible to add new features or technological refinements without obsoleting existing usage or applications.



IV DISTRIBUTED-PROCESSING COMPUTER MODES

An effective approach to computer networking is "distributed processing." The term has been defined as "allocation of processing functions to system elements which can perform those functions most effectively and economically."

Distributed processing first manifested itself in the various peripheral elements of any computer system: punched-card-handling equipment, magnetic tape transports and printers for relaying data to and from the systems' central processors. Then, as central processing units (CPU's) grew in speed and capacity, peripheral processors were added to distribute various pre-processing functions which had previously been handled by the CPU. This freed the CPU's for concentration of their billionths-of-a-second computations and manipulations of data.

While this distribution of functions was evolving around the center of processing activity, the periphery was also growing in complexity. Terminals were developed which permitted introducing data and instructions into the system from locations far removed from the operator's console and from peripheral equipment in the computer room. And all along the way, the necessary software and applications programs were being written as communications links between man and this increasingly sophisticated hardware.

Most recently, what started as autonomous "systems" of one or more processors with associated input/output peripherals and terminals have been linked to other such autonomous systems to form "networks." Here the concept of distributed processing is still at work. As a result of distributing processing among various systems in the network, the linked systems are able to handle assignments more effectively and/or more economically than they could when working as separate autonomies.

In a network, concentration is still focused on assigning every job or every element within a job to that hardware-software element or combination of elements which can accomplish it fastest, most accurately and at least cost. Space, time and computer technology all enter into this concentration.

For instance, it is often most effective and economical to store specialized software and related data in strategic geographical locations. Moreover, at a time when a needed facility in one location is busy, it is often effective and economical to call upon a comparable facility in another place for immediate execution of a user's need. And as for computer technology, within limits of space and communication costs it is most generally most economical to have a sizeable assignment processed by the fastest and most powerful computer available.

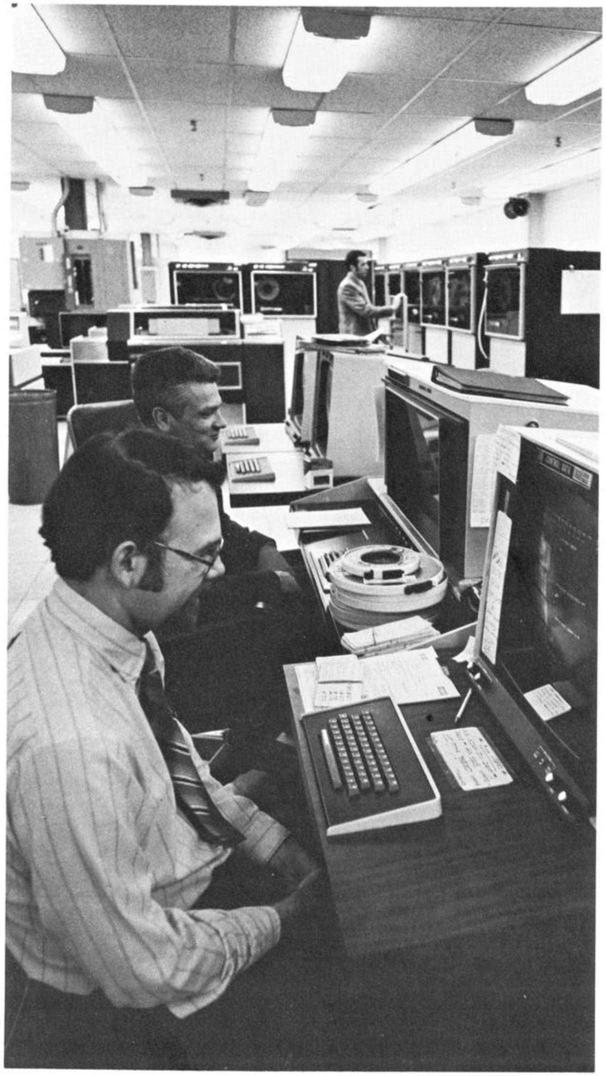
As the reader can see, there is such a multiplicity of factors in the weighing of alternatives here that a computer itself is best equipped to deal with them. Evidences of this are visible in a software system provided by Control Data to radio and television systems for weighing the various factors in programming their own networks and in the communication-computers which have been added to computer networks for policing the networks' traffic.

The distributed processing approach to data processing, however, has evolved from the combining of readily understandable, traditional processing methods into a single, ever-more-inclusive methodology. Three processing methods or modes, for example, that have long been available to the educational use of computer systems and now of computer networks are called "local batch," "remote batch" and "interactive." Actually, the terms apply not to processing itself but to methods of submitting jobs to the CPU for processing.

1. *Local batch* – Initially, most computers were capable of efficient operations only in a "local batch" environment, with jobs literally handed in at a computer room and processed in a queue, like customers in a bakery, with each customer's job or order handled in sequence.

Where few departments used the computer, this was successful and efficient. In the administrative area these early systems functioned simply as faster replacements for earlier "unit-record equipment" and, as such, were completely dedicated to administrative processing.

2. *Remote batch* – The addition of remote terminals which could prepare data and transmit it to the computer room, ready for processing, provided a much greater degree of accessibility to the computer; however, a major constraint was often imposed on both local and remote batch users in the area of magnetic-core-storage availability. The requirement in many systems for fixed core partitions proved to be a significant deterrent to effective use of the system for many potential users.



3. *Interactive processing* – The advent of interactive processing in which a user and computer work out problems together, step by step, instead of as a pre-batched program, provided a further expansion of computer accessibility for a large number of users. In the interactive environment time-sharing became, at last, a working reality, with numerous users profiting by the computer's speed in apparently simultaneous operations. The level of technical sophistication available at the time, however, was such that the practical application of either time sharing or CAI on any sizable basis required the use of functionally dedicated computing systems. As such, economic realities generally dictated the denial of this valuable tool to many who had need for it.

Each of these methods, was, of course, a major improvement or addition to its predecessor(s). Each, however, also has inherent advantages which precludes its successor(s) from displacing its use within the computing community. What was obviously needed was a computer system incorporating the best features of each of the traditional methods while at the same time attaining a new level of cost effectiveness. Once again distributed processing provided the answer.

When this approach is applied to a network, it means that various levels of processing power exist within the network and that the processing power of each level can be made available to any user in the network. As a result each processing function can be performed at the level most cost-effective for each user's application. The amount of processing power to be provided at each level is dependent upon the priority and emphasis that is placed upon such design criteria as processing functions required at each level, cost-effectiveness desired, and geographic location of systems in the network.

V CONSIDERATIONS FOR NETWORK PLANNING

Before determining how a data-processing network is to be structured, the objectives of the network must be clearly defined. Obviously, we at Control Data are not initially in a position to know what all of the specific objectives are. However, we can assume that reducing overall computing expenditures and making greater computing capability accessible to more users in educational institutions are two of the more important objectives.

Variables which should be considered in developing a network:

1. *Concepts Underlying Uses of the Computers at the Institutions Involved.*

For example, the concept that each undergraduate should have computer experience.

2. *Size of User Population.*

The current size and anticipated growth of the institutions involved.

3. *Cost Reduction.*

It is important to remember that networking should show distinct advantages either in cost reduction (immediate or future) or increased services to warrant its consideration.

4. *Functions.*

The specific functions of the network as a whole, should be defined. Specifically, the need for batch processing must be identified as well as the need for time-sharing. The types of user problems to be handled must be identified as to teaching, scientific computing (number crunching) vs. social science or business-type, administrative work, etc., along with the particular mix of these applications at the institution. From these data will emerge specifications for core storage requirements, disk storage, CPU speeds and types of local computation power needed.

5. *Geography.*

The distances between institutions govern line costs. In addition, early network builders and users found that heavy personal communications among them is essential. Thus telephone communications between center staffs and travel costs for frequent meetings must be considered. In many cases, "circuit riders" have been deemed essential. These are persons who go from user to user helping them resolve their network-related problems.

6. *User History and Expectations.*

The network should not impose restrictions on present users and, ideally, should provide more service than they expect.

7. *Resource Sharing vs. Load Sharing.*

Resource sharing refers to networks in which the basic motivation is to provide a greater degree or breadth of service to users. It permits multiple access to unique and specialized data bases. However, it is not necessarily aimed at cost reduction, although this often is a side benefit. *Load Sharing,* on the other hand, has the basic motivation of handling a volume of work at the least cost.

8. *Autonomy.*

The question of autonomy for each institution is a problem that needs to be addressed very carefully. Since most institutions that now offer computing capability have their own computer systems, they will be reluctant to give up their own systems unless they are very dissatisfied with their current capabilities. Yet, in order for the network to function properly, there must be at least a minimum level of cooperation between all institutions in the network. Recognition of this will be necessary if the network is to meet its objective of providing more and better computing service to a wider range of users. Centralization of service cannot be equated with centralization of authority and decisions.

9. *General.*

When establishing guidelines for the levels of processing power to be located at the different institutions, it should be noted that technical advisors are required at all levels for the normal operation of the network; and to keep qualified personnel of this type requires a certain level of capability for development work to provide personal motivation. Advisory personnel at the terminal level must be involved in development projects within the network and have access to computing capability to provide sufficient personal motivation in their profession or else they will leave.

The network should also be designed so that the large systems in the resource centers can be upgraded or replaced as needed, by larger systems without impacting the local users. Ideally this upgrade should be invisible to the local user. This becomes a very significant consideration for the future growth of the network. If the network cannot grow easily at the top, individual smaller institutions will supply their own small systems, thus destroying the network concept and losing the cost-effectiveness advantage.

All of these factors or considerations will influence the final network configuration. The amount of processing power to be provided at each location or level will depend upon the priority and emphasis that are placed on each of the above considerations as well as others.

Figure No. 1 indicates the wide range of options that is possible in a distributed-processing network. From a functional point of view, the different levels of systems might be used in an educational network as follows:

The remote terminals are input/output devices for instructional-type student jobs, for example; these may be remote batch terminals or interactive terminals connected via multiplexers or mini-computers to larger systems in the network and would provide processing capabilities to remote users. The mini- and small computers would serve as concentrators performing such functions as job routing, message switching and utility functions (card-to-tape, tape-to-print, etc.). These could have CRT's and limited disk storage.

Small to medium computers would provide specialized functions such as graphics, for example, as well as processing a limited number of student jobs. The medium to large computer systems are the point at which centralization takes place. These systems would provide facilities for complete student-job processing, complete administrative processing, and limited research. They would be located at large campuses and would provide computing capability for other smaller institutions as well as for the campus on which they are located. These large systems would be connected to each other by wide-band lines so that load leveling between them can occur to provide optimum use of the systems.

Above the medium to large computer systems, there might be a very large-scale system for large-scale research projects and any academic training that could not be made available on a lower level. This might well be a state computer utility which is made available to larger computers users at individual campuses. It may be advantageous initially to consider a commercial vendor for this service. Purchasing or leasing such a system can be deferred until use justifies it.

In summary, the distributed-processing approach to a network provides a great deal of flexibility, with the final network configuration depending upon the network objectives.

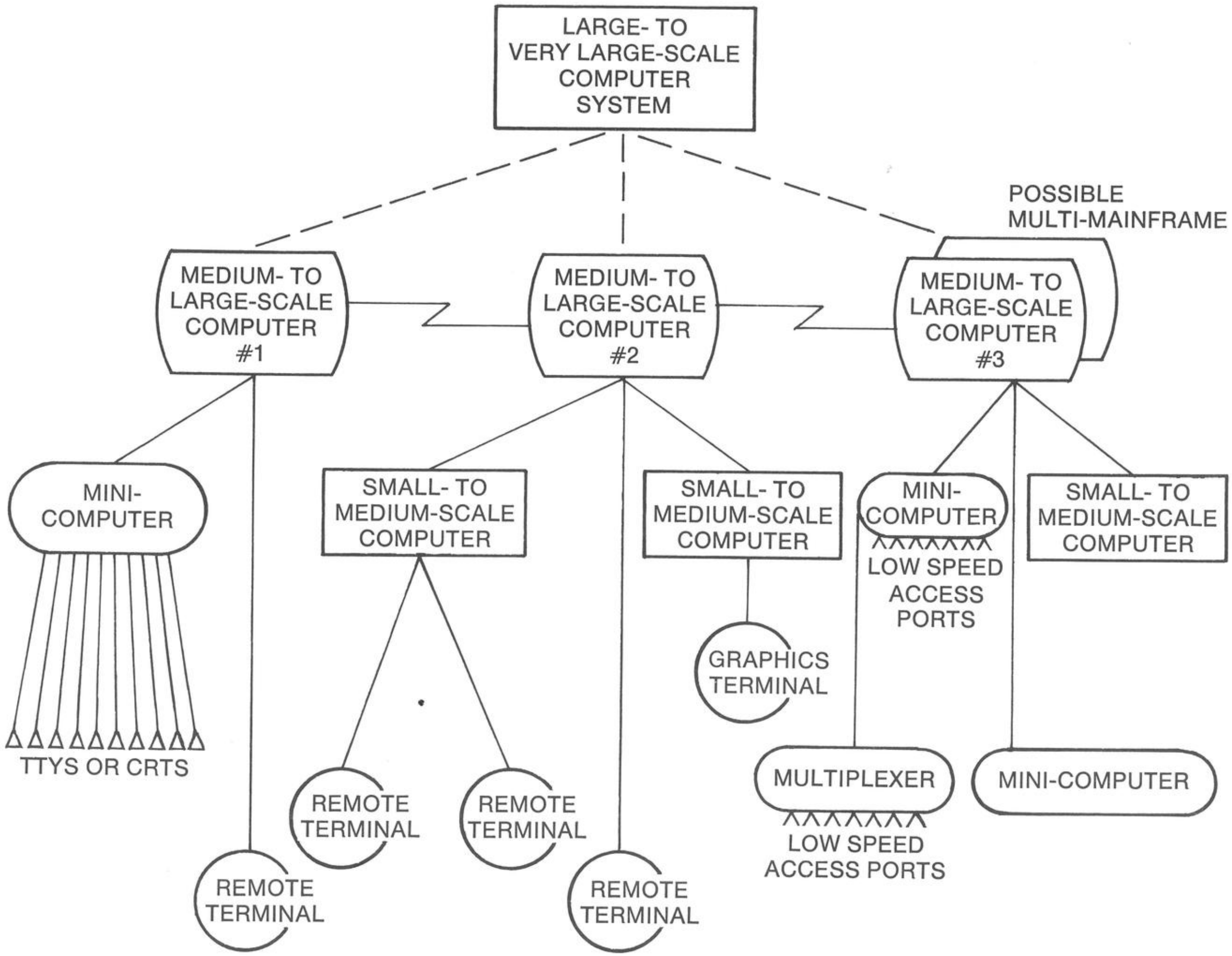
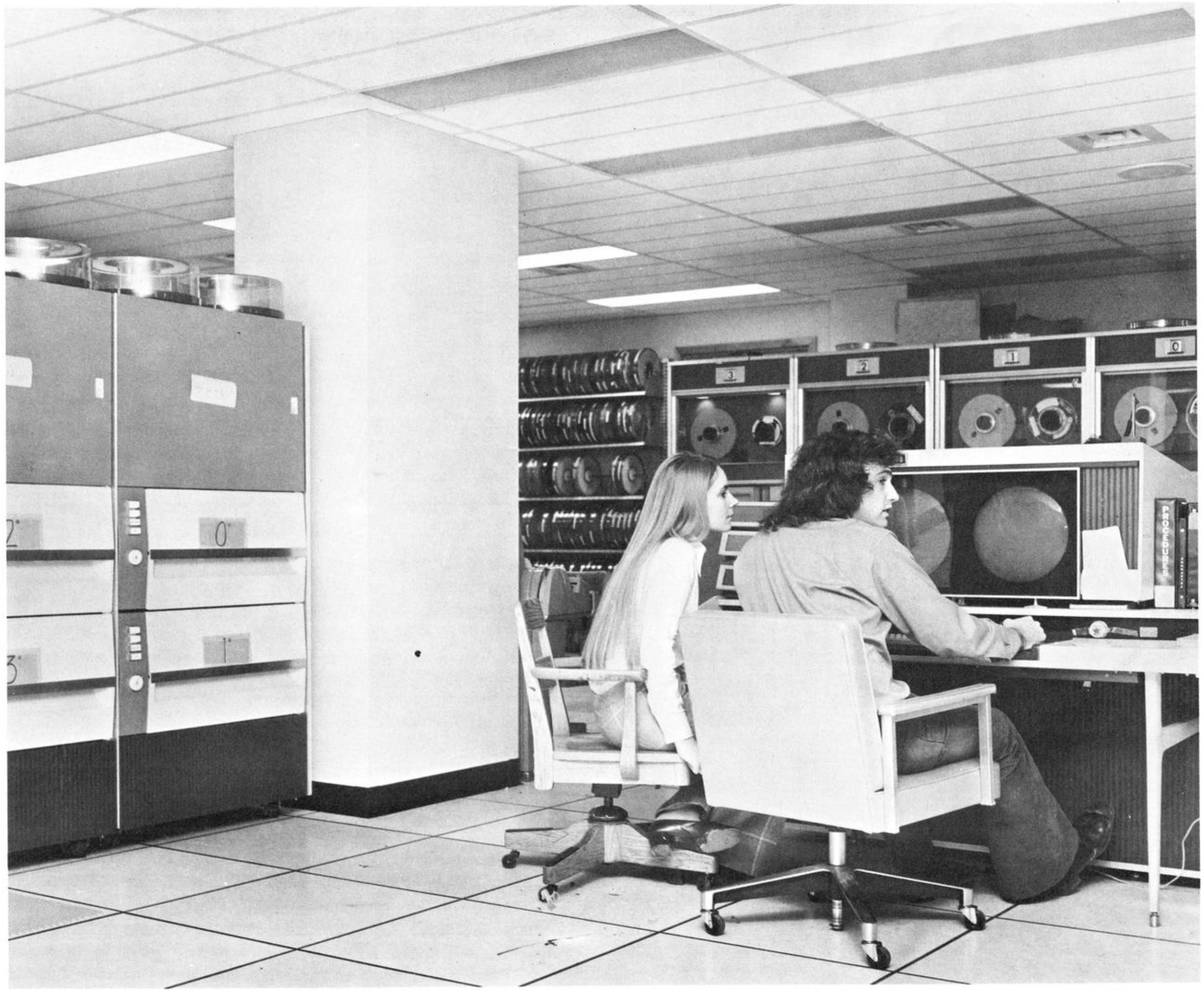


FIGURE 1 **DISTRIBUTED PROCESSING NETWORK**


3
2
1
0
PROCEDURES

VI CENTRALIZATION VS. DECENTRALIZATION OF NETWORK SERVICES

There are several key points in evolving a recommendation for the degree of centralization of a computer network. First, the "degree of centralization" may be different for each of the primary application areas: Instruction and Research, Administration, and Library. Second, each of these application areas should be considered with respect to the three primary modes of operation: local batch, remote batch job entry and interactive timesharing.

In the instructional and research areas, the support requirements will necessarily vary directly with curriculum and disciplinary emphasis of the individual campuses. A gross measure of instructional need may be derived from campus enrollments and instructional support plans keyed to the Higher Education Information Survey classification.

In the Administrative and Library areas, the data bases, applications and reporting requirements, once defined, will proceed relatively unchanged and will vary between campuses primarily with respect to volume.

For discussion of a general network configuration philosophy, the nature of work to be done in the three primary application areas is reviewed:

1. *Administrative.*

Historically — and predictably — the administrative workload tends to be relatively unchanged with respect to application and processing; i.e., admissions, registration, etc., and most must be run on a cyclic basis); however, each of the jobs in this area tends to be complex. Relatively small but significant abnormalities often develop each time a specific job is run. This is due primarily to two things: 1) changes in operating procedure on the campus and 2) data combinations and alternatives which had not previously been considered. In other words, the administrative application tends to be a time-consuming and complex effort which requires a great deal of local attention.

Further, a majority of the administrative files which are maintained and accessed within a network are pertinent only to the local campus. It is noted that a data base can be created and a program placed in a library at a given location which is remote from the individual campus. It should also be noted, however, that if the majority of the data originates at the local campus and if the majority of the resulting output is to be used at the local campus — and no other campus — then the work may best be accomplished at the local campus. This saves transmission time and places the workload in the hands of those who must eventually try to solve whatever problems exist and tolerate those which cannot be solved.

For these reasons, many of the administrative processing applications could be done at the local level in a campus processing center. The exceptions to this will be those applications which are extremely large, standardized, and/or time-consuming or which require a consolidation of data from the various campuses; e.g., common admissions.

2. *Library.*

It is desirable that each campus have access to the data available to all other campuses in order to establish the availability of and to take advantage of specific material; a decentralized approach would be restrictive, cumbersome, and redundant with respect to data management.

The problem of circulation control should not be directly affected by the location of the data base. By proper planning, relatively inexpensive data-capture methods can be initially employed and can later be evolved into an on-line system. The ultimate system would provide for immediate and possibly automatic check-in and check-out of material and might even include a security system. The location of the data base, however, should be inconsequential to the users.

3. *Instruction and Research.*

Since the basic purpose of an educational computing network is centered on instruction, the applications in this area will be diverse and numerous and will require a great quantity and variation of computer power and capability.

The general philosophy surrounding questions of centralization vs. decentralization within this broad application area is a function of the nature and volume of the work to be done.

The total workload will in some way be divided among the three basic computing modes: local batch, remote batch and conversational timesharing. Also, the total workload can be divided into four major categories. These are:

Use of the computer: 1) As a computational tool
2) For coursework and related research
3) As a learning tool
4) As a study in itself

None of these categories should be tied directly to a given computing mode to the exclusion of all others. Timesharing, however, is rapidly becoming recognized — due to its convenience — as a valuable tool in all areas of instruction and research. Control Data has found that the best cost/performance ratios in timesharing can be obtained through the economies of scale inherent in a large-scale computer system. It should be further noted that maximum cost-effectiveness is achieved before the maximum number of terminals per system is reached. The following graph shows what Control Data has found to be the per terminal cost using a large-scale computer system for a range of from 128 to 512 active terminals. Note the tendency for cost to stabilize at a relatively high number of terminals:

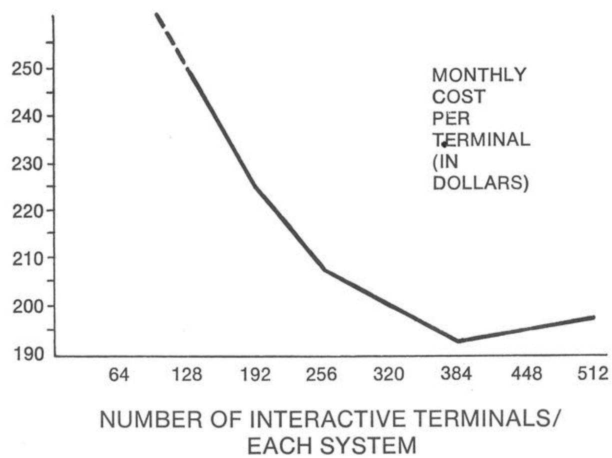


Figure 1

The same general curve results when the cost per terminal hour is plotted as a function of number of terminals on a single system.

A single, large system driving approximately 400 terminals would probably become saturated in a short period of time. As such, multiple but identical systems sharing a common data base would be required. Large systems are more advantageous for timesharing also because they have memories large enough to allow use of such necessary tools of education as the larger simulation or statistical packages. An additional but not so obvious cost saving associated with a larger system as opposed to many smaller ones is the singularity of program and system maintenance. The one exception is in the area of computer science. The student in this area needs the capabilities of the large system but also requires the ability to experiment with and modify systems in ways which would prove to be occasionally catastrophic for the other users of a large one.

The use of batch-job input for instructional applications is not decreasing. Computer-center application libraries are rapidly being expanded with additional instructional applications which run in the batch mode and which are being acquired from outside sources. It is observed that these programs — whether called or submitted — rarely request or get

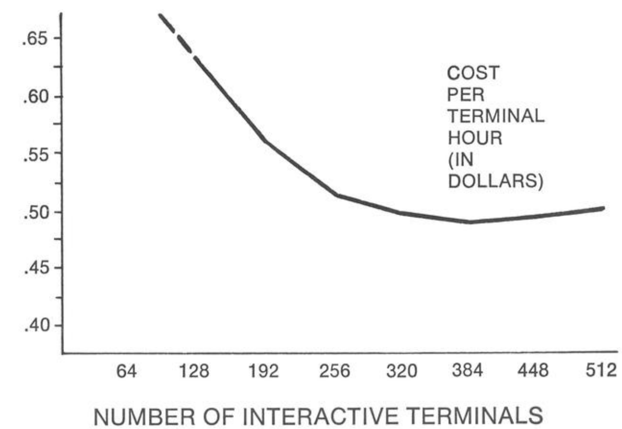


Figure 2

CAMPUS NETWORK

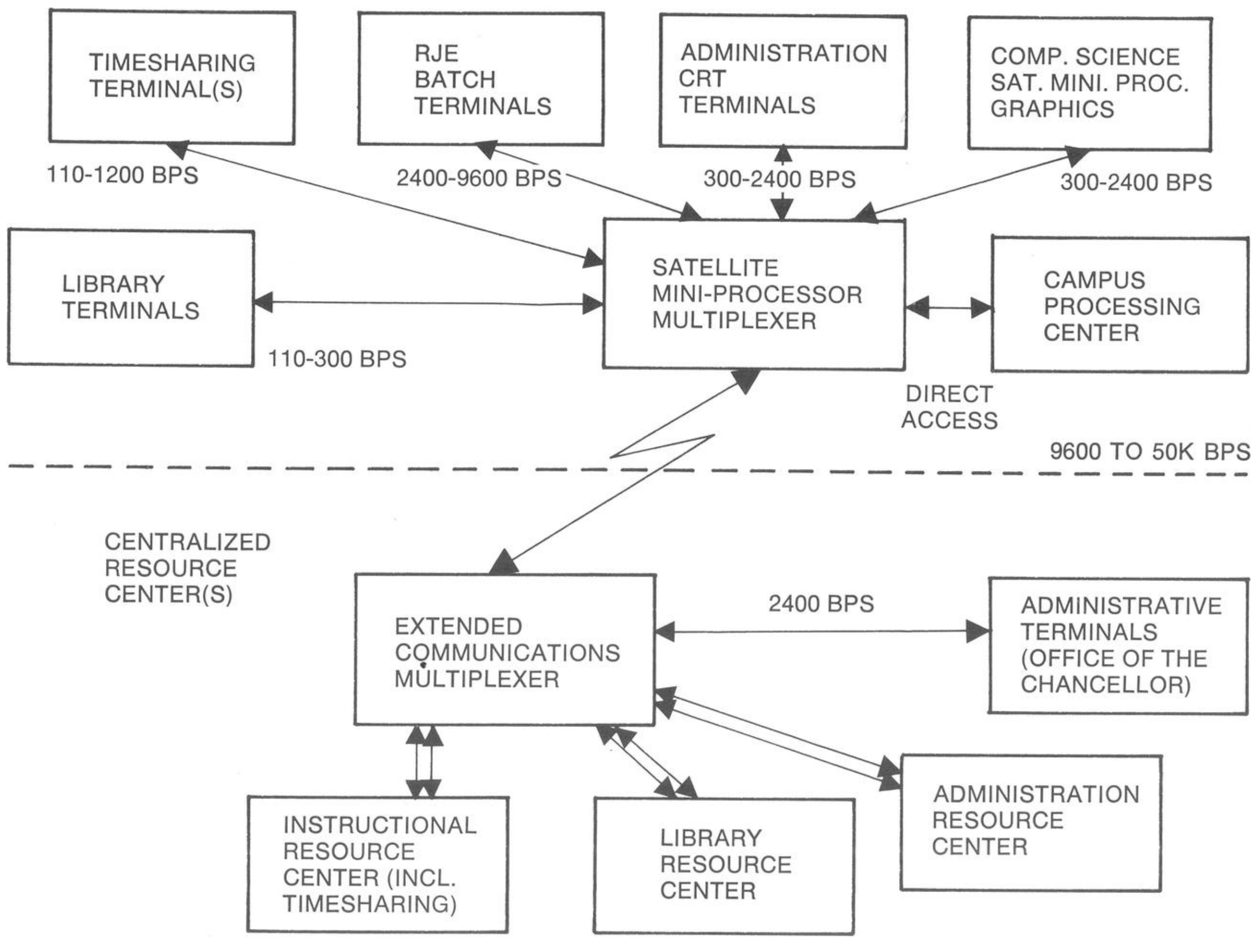


Figure 3

direct attention from computer-center personnel while being run. The location of the computer therefore, on which this type of job is run is of little significance to the user provided that he receives his results within a time period that he considers satisfactory.

All of these factors support a plan for a network which is both centralized and distributed according to application specialization, computer processing mode and individual institution facility. Figure 3 shows a pictorial concept of such a network.

The campus equipment would consist of a resource center capable of handling most local functions, comprising a variety of slow, medium and high-speed remote job-entry terminals and a variety of time-sharing terminals with speeds of up to 2400 BPS, connected to an on-campus Satellite Mini-Processor (SMP) which would control all communication with the network of specialized service centers and similar facilities.

The campus processing center would have an I/O channel data link to the SMP and from there to the special services network and would itself have adequate file-storage capability to handle all files pertinent to the local campus. It would also have terminal-handling capability for remote but on-campus interaction with those files. In place of a remote job entry (RJE) terminal, the Computer Science Department may require one or more SMP's which could simulate an RJE terminal and which could connect to the central campus SMP. The Computer Science Department SMP's would be used for a variety of purposes relating to the many varied areas of coursework and research in this discipline. Ideally it would be capable of interfacing a variety of special devices such as Analog to Digital subsystem controllers, graphics controllers, etc.

The remote job entry terminals would probably contain no stand-alone compute capability. They would consist of various combinations of slow, medium and high-speed card readers, printers, punches and magnetic tapes. Those departments which can justify dedicated terminals for planned coursework should have them. For the remainder of the users, a "time-sharing laboratory" may be better. Resources should be shared among departments as required. For instance, a graphics terminal used by Computer Science may be shared with Physics, Geography, etc.

The economies realized by the use of centralized but distributed network services, together with increased maximum resource capability, should improve the quality of support services in the form of skilled practitioners in computer applications. Thus, what is normally a duplication of effort in academic areas would probably decrease. Increased savings should occur especially in the area of program-library development and maintenance and a diminished need for multiple purchase of certain programs.

As previously discussed, there are trends towards centralization. These trends have in the past been hampered in varying degrees by the inadequacies of existing communication networks.

Several vendors and the Federal Communications Commission have recently taken actions to relieve the user of these known inadequacies.

These new network offerings, with data transmission as a prime concern, will lend very positive attributes towards centralization of large-scale data files while allowing for distribution processing of those jobs/files unique to any campus facility submitting a job request.

Other Concepts

Consideration should be given to establishment of decentralized laboratories only if this can be accomplished without sacrificing capabilities available through the network concept.

Examples might include:

- Creative graphics (as opposed to specific disciplines which may be better serviced by computation capabilities of a larger resource).
- Computer science ("hands on" devices)
- CAI research (involving special terminals)

In each case, laboratories should be available to all on-campus disciplines with special requirements. Also, each terminal device (Graphics/CRT/TTY) should be capable of accessing the network (general) as well as on-campus laboratory special-purpose utilization.

Perhaps at least one remote-job terminal may be placed on each campus in every major area. Campus geography — as well as centers of student activity and degree of computer involvement by discipline — should dictate the physical locations, quantity of terminals and I/O speeds to best balance the loads.

Data Acquisition and Analysis Applications

Within this category are such applications as instrument sensing and controlling, process sensing and controlling, analog-digital conversion, etc.

The campus computer facility should include the data acquisition components necessary to support these applications and their variations. The hardware components should be supported by user-oriented software enhanced by programming aids and languages to facilitate project implementation. Standardized hardware and software products should be available to obtain applicable economies and should have sufficient modularity in designs to provide adequate flexibility for dynamic applications and future growth.

Medical/Hospital Applications

Although the field of hospital/medical computer applications is still in a developmental phase, the accelerating cost of patient care and associated support activities and the increasing demand for health services are creating significant pressures to utilize fully the potential efficiencies and capacities of automation. Accordingly, although many primary functions of a medical center are already automated, many more will be in the future.

For the campus computer facility to support the needs of its Student Health Center and/or Medical Center, it should include capabilities, either actual or potential, in the following application areas:

- Interpretation of electrocardiograms
- Administrative and accounting data processing
- Storage, retrieval and communication of patient medical files
- Patient screening and admissions
- On-line patient monitoring within intensive care and surgical wards
- On-line sensing and record-keeping of clinical laboratory instruments and equipment
- Training of nurses and technicians

Control Data is involved in medical centers in the administrative environment as well as in monitoring special applications.

VII TECHNICAL DESIGN CONSIDERATIONS

A pilot program of network development would cover from one to three years. But it is likely to be five years before the full capability resulting from such a development would be available to the user. In fact, it is difficult to adequately define beforehand the "full" capability that will evolve within the time specified. This is not because of any technological issues but because of the very real and practical problems encountered in introducing a new operating environment to the educational communities involved in the network. User acceptance and local autonomy will be major areas of concern.

In view of the extended time period, it is important to identify the things which impact today's design. First and foremost, many of the sources of computer resources for the network will probably already be distributed within the boundaries of the proposed network. It is doubtful if much consolidation of those existing resources would occur initially. The emergence of the network will affect the growth pattern of each member institution's computational capability even though geographic distribution is likely to remain. A factor in the evolving distribution of resources is the concept of having discipline and/or technique specialization centers develop within the network.

The most common mode of communication between computers and remote users is by use of common-carrier facilities. In spite of many comments to the contrary, the cost of communication facilities to the end user is not decreasing; nevertheless, the initial network development may well be to effect an upgraded, switched network-like communication facility. Here, a design decision to separate communication functions from application-processing functions is appropriate. The evident benefits in doing this approximate those resulting from the analogous "front-end" computers that are a vital component of large timesharing system configurations and handle most – if not all – of the communications processing. A front-end mini-computer complex, i.e., a Node Control Unit (NCU), may be utilized as the basis for the communication subsystem. The diagram at right illustrates the concept.

It is probable that some of the network host sites may have teleprocessing devices already. It is considered desirable to emulate these devices.

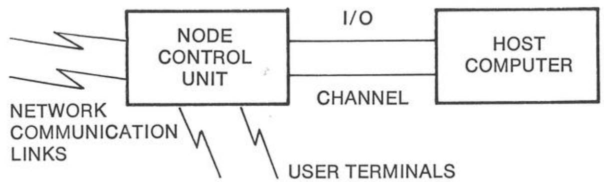


Figure 1 NODE CONTROL UNIT

A broad network-development plan might have the following form:

Stage 1

Installation of NCU's at selected sites within the network. It is not necessary to install the NCU's at all sites at this time. This development is designed to stabilize operation of the NCU and to implement the terminal interfaces, Host-NCU linkage, emulation capability and port control functions.

Stage 2

This development approximates the present state-of-the-art netting activities that are concerned with the objectives of resource sharing and load equalization. The objective is to provide a network communications facility which is transparent to the user and to the host operating system. Host operating systems, to the extent possible, will remain unchanged. The communication subsystem is to have a transmission band width and response time that meet projected needs and provide for an upgraded, switched network facility.

Stage 2, as defined, implies the following capabilities:

- terminal support facilities provided by the NCU
- short response time as related to message transmission and NCU processing time
- established levels of protocol for communication between nodes, users and hosts
- sufficient network functions to permit traffic accounting and remote port control
- operating-system changes minimized
- message routing
- terminal translation mechanisms
- installation of remaining, originally specified NCU's

Stage 3

The objective of Stage 3 development is to work towards optimizing the network. Inter-translation facilities which have not been previously implemented should be implemented for both interactive (e.g., TTY37 and IBM 2741) and batch-type (e.g., CDC 200 User and IBM 360/20) terminals. Emphasis will be on implementation of more general translation facilities and the evolution of communication protocols that are better suited to the network. Network-performance measurement facilities are implied.

Stage 4

Emphasis will be on development of comprehensive netting functions which include but are not limited to:

- introduction of unique hardware and/or software for network users
- a network language designed to facilitate use of network resources
- distributed data-base methodology
- inter-process communication

A MODEL OF A COMPUTER NETWORK

To better illustrate the complexities and challenges of achieving a useful computer network a brief description of the various design aspects of such a network will be outlined.

As a base for discussion consider the model portrayed in Figure 2 as a base for future network development.

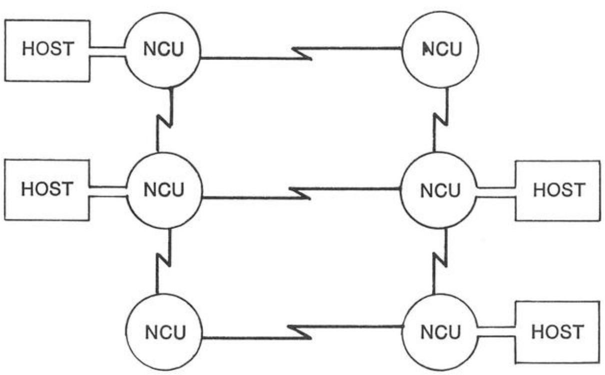


Figure 2. BASIC NETWORK MODEL

This model defines a possible form of a heterogeneous computer network. In the diagram, host computers are not identified by type but are classed simply as "hosts" for the purpose of discussion. The use of store-and-forward "packet" switching techniques is implied by the diagram. The diagram also reflects overconnection of nodes. Such overconnection permits optimum network accessibility to users and reflects concern for sensitivity to network loading and node failures. By definition, the design permits a large user community on acquisition to of an NCU access the network independent of a host.

The objectives of the above design may be stated as being:

- to enable a distributed user community, by means of terminals in the batch and interactive mode, to access distributed and diverse computational resources composed of hardware, software and application packages
- to make the network transparent to the users and the large-scale host processors
- to ensure proper response time and transmission bandwidth for proper utilization of "state-of-the-art" terminals
- to maximize reliability
- to permit host upgrade or replacement to occur which in turn permits the network to move into the future
- to provide for eventual inter-process communications, distributed data bases and data base transfers
- to enable users to use the network without knowing the command languages of all the hosts
- to provide widespread access to specialized local facilities or unique devices.

There are many network considerations which affect the design. Some of them are to:

- permit transparency of the network to the hosts and terminals
- allow "foreign" terminals to be attached to hosts
- provide access for both asynchronous and synchronous terminals
- allow a site without a host to connect to the network
- provide a developmental mechanism for the implementation of various levels of a network access language.
- allow parallel operation of a presently installed local environment during network development activity
- make the host computer relatively free of communications functions and minimize degradation in reliability of the hosts
- allow mixing of communication facilities
- allow some variation in communication facilities between

NCU's and terminal types
- allow access to remote hosts (restricted in the early stages) to be similar to local access
- obtain leverage from existing network activities
- allow performance measurement to improve network operations
- allow use of presently installed communications equipment
- develop effective network protocols within the context of a global design.

The Node Control Unit

The Node Control Unit (NCU) would assume all responsibility for network functions and partial responsibility for terminal handling. It is appropriate that the NCU handle the switching task, too. The NCU is the network-access facility and therefore must handle three types of information streams as shown in Figure 3. When a store-and-forward switching system is employed, the additional traffic stream is generated.

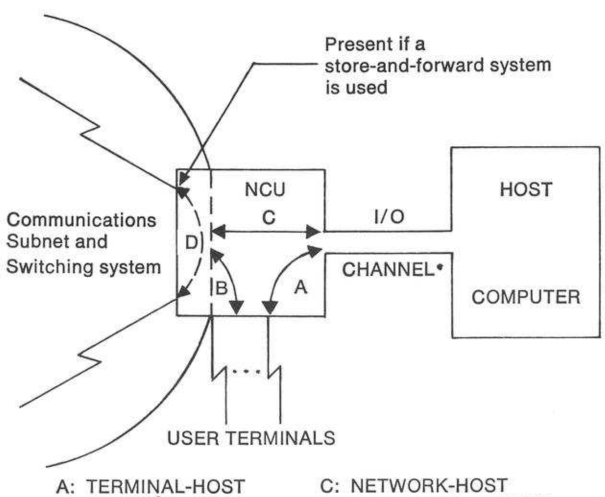


Figure 3 NCU INFORMATION STREAMS

A description of the NCU function for each of the various information streams follows:

A. *Terminal-Host*

The NCU functions as a front-end communications controller on the local host-computer system. The NCU will make "non-standard" terminals appear to the host system as one of the acceptable types of terminals.

B. *Terminal-Network*

The NCU functions as a remote terminal multiplexer. Additionally, the NCU must multiplex messages to many destinations, possibly over several communication links.

C. *Network-Host*

The NCU functions as a front-end communications controller on the host system where the user terminals are connected via a special non-standard terminal, i.e., another NCU functioning as a remote multiplexer.

D. *Network-Network*

The NCU performs the switching function. A store-and-forward technique is implied. The NCU receives a message on one of the network communication links and routes it on toward its destination via one of the other network communication links.

NCU-HOST Communications Goals

The goals in the area of NCU-HOST Communications are:
- to provide widespread access to unique local facilities
- to permit normal use of the host systems by minimizing host disruption during network development
- to provide greater flexibility in configuration of the teleprocessing facility
- to isolate the host from network communications activity
- to minimize effect on future network developments
- to have multiple concurrent communication paths between a host and the network

Requirements and Means of Implementation

These goals result in a set of requirements for the Host-NCU Network complex:
- to attach the network to a host via an "intelligent" front-end controller
- to minimize operating system changes in initial phases of the network

• to ensure that error rates in communication between a host and the network are consistent with those for normal peripheral equipment communications
• to ensure that data rates between a host and the network be limited only by the I/O channel and be capable of full duplex operation as appropriate
• to ensure that there be multiple, bi-directional, pseudo-ports between the NCU and the host operating system. They are pseudo-ports in contrast to the physical ports for terminal handling.

The methods suggested to accomplish the goals and satisfy the requirements are:

• to emulate the host's standard communications controller in initial network implementations
• to support terminals on the NCU
• to use standard peripheral interfacing techniques appropriate to the particular host. The interface should be designed so that it may connect to the differing classes of I/O facilities. This is more easily and cheaply incorporated as an initial design than as a retrofitted modification.

Emulation of the teleprocessing environment does not necessarily infer a terminal-supporting NCU. It does, however, mean that links to the network appear as terminal ports to the operating system. A terminal-supporting NCU is deemed desirable so that:

• access to the network is independent of the status of the local host
• a site without a host may access the network
• additional terminal-handling facilities can be provided on hosts that are deficient in this area
• a computing center may replace its present teleprocessing equipment or choose to add ports to the NCU

Design Constraints

The ability to provide a teleprocessing emulation facility for a host – using an intelligent front end – has been demonstrated by a number of independent manufacturers. What has not been done is to map the various host communication protocols into a common network protocol. Although this is undoubtedly possible, it may be costly in terms of NCU memory and processing time. A relaxation from the requirement to emulate would result in a more easily implemented and efficient host-network access method. However, terminal and protocol translation must be performed to provide a transparent network access facility.

Since the NCU is to be connected to a host, the possibility of host-to-host communication exists. Some operating systems already support host-to-host communication through their teleprocessing devices. It has been indicated that there will be two teleprocessing devices – the standard one and the NCU; however, with the standard operating system, not all hosts may be able to support more than one teleprocessing device concurrently.

Terminal Characterization

The requirement for characterization is twofold:
• to provide a means of identification
• to provide criteria for translation

Thus it is necessary to identify those fundamental properties which cannot be ignored when communicating with a terminal and determine the variables associated with them. Considering the interactive (generally asynchronous) terminal types, the fundamentals are:

• character set
• characters per line
• bits per character
• number of stop bits used
• bit sequence within a character
• parity bit and type
• bit rate
• line discipline; e.g., the manner in which output is printed or input data is gathered (full- or half-duplex use of "line-break" signal)
• function-delay characteristics (high incidence of idiosyncrasies in this area)
• format effectors (tabs, newline, form feed)

For the batch (synchronous) type we have a similar set of fundamentals:

• compression techniques
• communications protocol (line discipline)
• data-blocking scheme
• bits per character
• bit rate
• character set
• synchronizing code
• bit sequence within a character
• error-recovery philosophy

In addition to the fundamental properties there are often additional features such as reader-punch units, two-color ribbon, half-line motions. Recognition of these features will aid in providing an interface matched more closely to user needs.

The Central Switch Network (see fig. 4)

In this network concept, a number of host computer systems are linked individually to a communications controller system referred to as the central switch. None of the host computers communicate directly with each other — all communications, from a host's point of view, are directed toward the central switch.

The "host" elements may be homogeneous, although, more commonly, the network is comprised of heterogeneous hosts. Each host system may have its own terminal devices, mass-storage media and peripheral equipment. Most importantly, however, each host system may be individually selected to meet the processing requirements of its primary user.

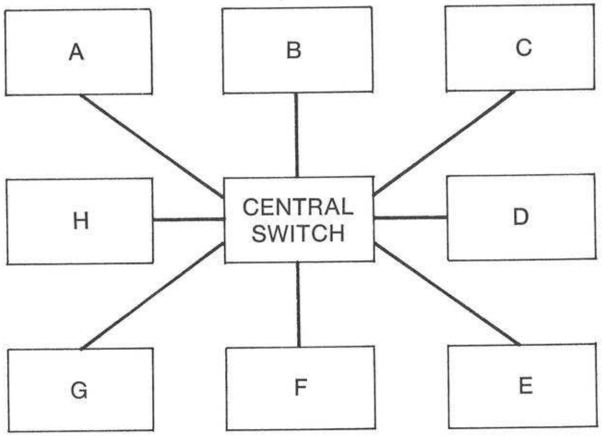
The central switch itself is simply a communications controller. It possesses no peripheral equipment of its own and performs no processing functions. Its job is merely to accept data from a processing "sending" host and transmit this same data to a "receiving" host. The host computers themselves are responsible for the proper encoding and decoding of data to conform with their own protocol requirements.

The central switch concept is one which can be implemented very economically to provide a powerful resource-sharing capability among a number of existing, heterogeneous computer systems. Communications costs are minimized by the radial nature of the host-central switch network, and the simplicity of functions required at the central switch can be readily fulfilled by small, somtimes "mini" computer systems.

In a network of homogeneous systems, the central switch provides an effective mechanism for load-leveling. But the primary importance of the concept is in the provision of resource-sharing and its numerous advantages. Users of each host element are able to use any of the applications programs that are available on any of the other host elements in the network. This capability obviates the necessity for the duplication of applications development efforts. Host elements can be programmed to perform those functions for which they are optimally suited, and the users of other host systems can use these resources through the central switch. In essence, the users of any host computer actually have access to the applications libraries of all other host systems in the network.

In large, sophisticated versions of the central switch concept, data in common use among several host elements can be controlled and stored by one of the host elements — thus avoiding the cost of multiple, redundant storage.

The central switch network is also an ideal solution to the problem of providing users with a computer system to handle their primary processing requirements without concerning them about their secondary — or less justifiable — needs. Secondary tasks can be transmitted through the central switch to a suitable system.



Schematic Diagram of a Central Switch Network. A through H represent "Host" systems.

Figure 4

Individual Sites

It is envisioned that network computing will provide impetus for continued growth of present large centers towards what might be termed "cluster centers" composed of multiple, large-scale computers with shared files and common input queues.

Such centers will become true information-processing centers of high order. At the same time, remote satellite centers will begin to emerge. These smaller centers will be user-oriented with disks, cassette tapes, plotters, etc., and will not require any special facilities such as air conditioning. Emphasis will be on tailoring the large network service to the needs of the local user.

COMMUNICATIONS

One of the major areas of concern in planning a network is provision of common-carrier facilities. The following information briefly describes some of the capabilities presently available:

Line Types – The types of lines in use and their abbreviations are

- FD = Full-Period (Voice-Grade) Data Line
- FW = Full-Period (Wide-Band) Data Line
- FDA = Full-Period (Voice-Grade) Data Line, Alternating Voice Channels
- FWA = Full-Period (Wide-Band) Data Line, Alternating 12-Voice Channels
- FXD = Full-Period (Voice-Grade) Foreign Exchange, Alternating Data Channel

Line, Voice-Grade – Any telephone line operating at less than 40,000 bps (bits per second) but more than 300 bps.

Speed levels currently available are:

9600 bps
7200 bps
5400 bps
4800 bps
3600 bps
2400 bps
2000 bps
1200 bps
300 bps
150 bps
110 bps
75 bps

There may be three types of voice-grade lines: "switched" lines which are essentially like residential phones in that they operate by dialing through the public switched network; "full-period" lines, which are fixed point-to-point lines operated for only one terminal; "foreign-exchange" lines, which appear as "full-period" to a network node but are open to the public at the terminal end which is in some other location – which would normally require long-distance toll charges.

Line, Wide-Band – Any telephone line operating at 19,200 bps or higher.

Multiplexer – A multiplexer is a computer system or device which serves the sole function of handling a number of communications lines on a polling or interrupt basis. It is connected directly on the inboard side to a network node, and on the outboard side to lines, another multiplexer, or a number of multiplexers which are connected by lines to terminals.

VIII FINANCIAL CONSIDERATIONS

The financial problems which must be resolved to develop a network are of great consequence. A good financial plan will serve to identify them and encourage their early resolution. The development of a financial plan may be one of the more frustrating aspects of a network project. Ideally, a network would be designed to meet its user's needs and, at the same time, be simple enough to be economically viable.

The resource-sharing contemplated as a user benefit of evolving networks tends to redistribute the cost of computing and, as a result, a new mechanism is required to distribute funding. The relationships between the user and the computer center are expanded. The user of computing resources now has more to choose from; i.e., the aggregate of network resources from multiple suppliers. The center director for the local system now becomes, in effect, a retailer or broker of services provided by other centers as well as his own. The impact of this on financial planning cannot be overstated.

The following information is presented as background in developing financial understanding in matters relative to providing computer services.

BUDGETING

Budgeting is the process by which financial resources are allocated to all departments of the university. Some of these resources are used by departments to obtain computing services and result in the expenditure of funds in support of the campus computing facility. Ideally, these funds would be "tied" in the sense that they could not be spent on anything except computing.

"Tied" funds may be of two kinds: "tied-local" and "tied-global." "Tied-local" funds can only be used to purchase services from the on-campus computer facility. "Tied-global" funds may be spent at service agencies of the user's choice. This could be the local computing center or an off-campus source of computing.

"Tied" funds should be distinguished from "discretionary" funds which can be spent on things other than computing.

The expenditure of "tied-local" funds will be uneven over a period of a year. As a result, a budget deficit may occur because of the incurring of expenses by the computing center without having sufficient business to recover. Further, it appears reasonable that eventually the buying power of users would be the dominant factor in the economic system of computing services. Thus the "tied-local" funds would exceed substantially the centrally-provided computing center budget.

COSTING OF COMPUTING SERVICES

The purposes of costing are

- to relate the actual cost in a given period to the computing services provided during that period
- to provide the university with information on the relative "success" of its computing services
- to provide a basis for pricing of computing services
- to allow the university to allocate its resources between the computing cost center and other cost centers in the university according to priorities.

A cost center is a segment of the university, clearly defined by the management of the university to be responsible in administering funds allocated in the budgeting process. There may be more than one computing cost center. Costs are of two types – direct and indirect:

Direct costs may be defined as money which must be spent to acquire, produce or effect the computing services provided and whose expenditure does not acquire, produce or effect any other product or service. They are clearly identifiable as pertaining to the provision of computing services.

Indirect costs are defined as money which must be spent to support the provision of computing services but whose expenditure, in total, also supports other programs and services in the university. Indirect costs such as space, light and purchasing services emanate as charges from other cost centers in the university and are allocated as costs to the computing cost center on the basis of a cost-allocation policy exercised by management.

Costs which are applicable to a computing cost center may be grouped into nine main categories as follows:

- salaries
- benefits
- equipment

- software
- supplies
- utilities
- services
- financial
- other

The charging of depreciation calculated on the basis of a computing cost center's capital assets may be both necessary and desirable. It is desirable because of the obvious cost disparities that would arise in "rented" vs. "purchased" situations — which would become important when considering inter-university sales and also, for example, when examining the relative positions of subcenters within a university.

Of more importance, however, is the necessity to take proper account of depreciation in light of the current trend to provide computing from operating funds. Unless depreciation is written — and ultimately, funded — a university will not be able to provide for the replacement or upgrading of computer facilities without seriously distorting a single year's operations.

A computer manufacturer will offer a component or system for purchase at a price equal to the present value of expected future net rental charges calculated by using a discount rate appropriate to the risk involved.

The ratio of purchase/rent can be considered an approximation of the manufacturer's estimate of the economic life of the system.

In general, this may be expected to be *less* than the useful life of the system to the user. The calculation of the *useful life* of the equipment is necessary to provide insight into the length of time for depreciating the system.

To illustrate the significance of the purchase/rent ratio for the useful life of the equipment, let us assume that a manufacturer has established a purchase/rent ratio of 48 and a required rate of return of ten percent per annum.

The monthly rent figure is related to the purchase price as follows:

$$R = \frac{1}{m}C + \frac{r}{12}\frac{C}{2}$$

Where R is monthly rent, C is the purchase price, m is the anticipated life, r is the rate of return per annum and $\frac{C}{2}$ is the average amount of capital tied up.

For $\frac{C}{R} = 48$, and $r = 10\%$, solving for m gives an anticipated life of 60 months.

PRICING OF COMPUTING SERVICES

The purposes of a pricing policy may be defined as follows:

- recovery of costs
- control of demand
- smoothing of peak loads
- providing a reflection of the relative costs of different services

When capacity exceeds demand, it may be appropriate to have a pricing mechanism which will encourage the purchase of the output, so that demand will increase and capacity will not be wasted. Purchasers of services provided under such a condition should know that the price for the services reflects the fact of temporary excess capacity and that they should not expect the price for those services to remain the same when capacity is exceeded by demand. In a like manner, when demand exceeds capacity, pricing may be used to allocate computing resources.

With regard to reflection of the relative costs of different services, a pricing mechanism may

- serve to educate the user
- respond to a user's needs for priority
- respond to a user's needs for special equipment or software
- reflect the cost to the user of services which could be provided by outside suppliers
- permit recovery of funds from granting agencies.

Four of the main types of pricing schemes, identified below in relation to how they meet the purposes of pricing, are:

Single-Price Scheme

Every completed job is priced at the same rate.

Variable-Price Scheme

Prices change based on time of day, week, month or year.

Multiple-Input Queue

Prices are dependent on which queue the job is placed in. In advanced schemes, the user may buy a position in the queue he selects.

Right-of-Access Scheme

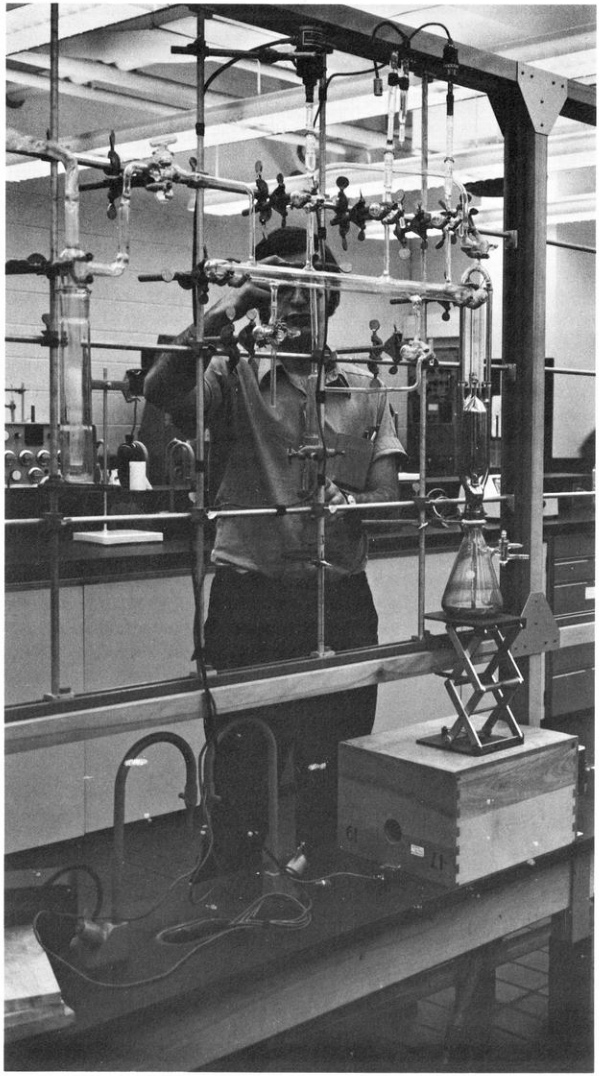
Users purchase rights-of-access to a service which perishes on a cyclic basis, e.g., every week. Priorities at any time in the cycle are based on the relative proportions of actual use to purchased rights-of-access.

A case may be made for the pricing of each computer service individually; however, *the price for a particular service may not relate to the cost of providing that service.* There are services provided for which accurate cost figures can be obtained, but there are other services where accurate costing is beyond the precision of available means of measurement.

Thus, what is costed should be measurable and should be worth measuring.

Going further into the question of "pricing by service," it is assumed that the price for a particular service need not necessarily recover the cost for that service. Cost recovery should be based on revenues accruing from all of the charging mechanisms. As has been discussed before, the price for a given service may control demand, possibly by creating a demand through low prices or inhibiting demand through high prices and the price for a given service may be established with a view to smoothing of peak loads.

All subsidies or grants should be identified explicitly irrespective of the way in which they are distributed, either to a central facility or to an individual user.



IX GOVERNING A DISTRIBUTED-PROCESSING NETWORK

Although the planning of a network is an extremely large undertaking, it is just the first step. The implementation and governing of the network are on-going tasks which mean success or failure in meeting planned objectives. As with any cooperative effort, especially one of the magnitude of a distributed processing computer network, good communication at all levels is of utmost importance. Each level of management must be open and responsive to all other levels both in proposed requirements and in proposed solutions to fulfill those requirements. At all times, management must keep the objectives of the overall utility in mind.

The management of CYBERNET® Services Division of Control Data Corporation, for example, is structured on a formal basis. Each functional management level has both specific limits within which it works as well as responsibilities which contribute to the overall network. Each level meets on a regular basis, with agenda items provided to all representatives prior to the actual meeting. If questions concerning overall benefits of any operational policy or service to be offered arise, the CYBERNET Planning Council becomes the final authority.

The management structure of the CYBERNET Network which has developed over the past ten years may well be applied generally to other networks offering similar services.

Four different groups or levels of management are necessary to govern a network such as has been discussed: 1) an overall planning council, 2) regional management, 3) local management and 4) systems and software-standardization review boards. To describe more completely the function and objectives of each level, a brief description follows:

- The Overall Planning Council establishes objectives of the utility as well as total utility planning and capability. The Planning Council becomes the final authority on all matters affecting the network. Standardization in equipment, levels of participation, staffing, terminal compatibility, etc., are examples of standards which should be established by the Planning Council.
- Regional Boards should be established to coordinate activities within their designated geographic areas which pertain to the overall network. These will discourage both redundant software development and the over-buying of computer hardware. It is also their responsibility to educate and assist all participating institutions in utilizing both the regional and total utility network. In this way, each unit can obtain the maximum benefits of a utility from both an administrative and instructional viewpoint.
- Local or Campus Boards become responsible for disseminating information on all three levels of capability: 1) total utility, 2) regional and 3) local. The local board must also insure that local use adheres to standards and limits established by the overall Planning Council. Individual-campus computer requirements will vary greatly both in size and types of hardware and software. The size and make-up of each local board accordingly will vary, depending upon the level of participation in the total network.
- Design Review Boards recommend areas of standardization to the overall Planning Council. These should include operating systems, levels of compilers, communication software, etc. Even though different types of equipment may be involved, certain standards can be established to assure maximum versatility and utilization of educational computer facilities. The responsibility of these boards is great and it is important that the members be well-qualified to make the technical decisions of the magnitudes that affect an entire utility.

In this structure, each level is responsible to a larger number of students and institutions to provide the highest level of exposure, education and sophistication that an educational, distributed-processing computer network offers.

X ADVANTAGES OF NETWORKS

By sharing a network, users can reduce their own computing costs, as they are minimizing the need for duplicate equipment, and through load-leveling within the network, the overbuy problem is reduced.

In addition, duplicate staffs to operate and maintain equipment are minimized. Likewise complete duplicate staffs to develop and maintain system and applications software are no longer required. All of these reasons can result in significant savings.

This reduction in expenditures does not, however, necessarily involve a corresponding reduction in computer service. On the contrary, the power and capabilities of the large-scale computing facilities, along with a more sophisticated and broader range of software, now become available to small institutions as well as to large ones. Centralizing or regionalizing computer services can result in economies of scale: larger machines at the central or regional sites provide equivalent computation at much less cost.

Perhaps of greater long-term importance is the fact that *centers of specialization are encouraged to develop.* The economies of scale may not be feasible for many smaller institutions. In a network, however, users of a specialized service which cannot be provided economically on their local campus could purchase it via the network from a center which specializes in that service.

With the network approach it is possible to achieve better overall utilization of available resources. This is especially true when administrative, research, and academic computing are combined on the network.

Depending on the student population serviced, network hubs may support a particular activity; e.g., time-sharing, research or administrative processing. In the past, attempts to use common hardware resulted in friction because the resources available did not allow the needs of administrative data processing and research-academic computing to be simultaneously satisfied. In the last few years, however, computer manufacturers have developed hardware and software which makes a common computer facility feasible and especially advantageous from the standpoint of equipment utilization.

Other advantages also result from total centralization or centralization on a regional basis with the network approach. A larger institution or regionalized center can pay better salaries and at the same time spend fewer dollars for computing personnel. It is possible to have better-trained people, provide more strict continuity of functions, develop specialization groups, enforce standard procedures, control documentation, control operation for satisfying network or region-wide goals and objectives; jointly attack problems, train managers through formal and on-the-job training; attract more prospective computing professionals; and combine administrative, research and instructional data-processing support functions and staffs into one organization. These advantages can be achieved in a larger institution or regionalized center but it is hard to evaluate them on an economic basis. In any case, these advantages will improve the overall computing service available to its users.

The foregoing pages, in pursuing various aspects of the planning and use of computer networks for educational institutions, may at times have given pause to the reader at the complexity of the subject. Granted, an effective computer network does not emerge overnight.

Providing a college, a university, or a state college-university system with such a network requires leadership, dedication, and a committed cooperation on the part of numerous members of the academic community. However, as with any large-scale effort to reorganize and improve the effectiveness of a major phase of academic activity, the complexity of the problem, viewed from a distance or in the abstract, is appreciably reduced when attention is brought to bear on its specifics and they begin to be dealt with in order.

The rewards, moreover, are great. Rewards not only of economy but also of notably improved data-processing capabilities and services. There is encouragement to be gained from the fact that the rewards are in direct proportion and often in geometric proportion to the size of the task.

Finally, there is the example of institutions which have already established such networks. Not only do the benefits and advantages they enjoy as results from their pioneering activities serve as actual examples of the value of such efforts, but their experience also serves as an aid to other institutions following their example. They have set up guideposts for even more successful networks to come.